European Commission

JRC SCIENTIFIC INFORMATION SYSTEMS AND DATABASES

Description of the GHS Urban Centre Database 2015

Public Release 2019
Version 1.0

Florczyk, A.J., Melchiorri, M., Corbane, C., Schiavina, M., Maffenini, M., Pesaresi, M., Politis, P., Sabo, S., Freire, S., Ehrlich, D., Kemper, T., Tommasi, P., Airaghi, D., Zanchetta, L.

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 - DG Joint Research Centre
Space, Security and Migration Directorate
Disaster Risk Management Unit E.1
Email: thomas.kemper@jrc.ec.europa.eu
Tel.: +39 0332 78 55 76

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

JRC115586

PDF ISBN 978-92-79-99753-2 doi:10.2760/037310

Luxembourg: Publications Office of the European Union, 2019

How to cite this report: Florczyk, A.J., Melchiorri, M., Corbane, C., Schiavina, M., Maffenini, M., Pesaresi, M., Politis, P., Sabo, S., Freire, S., Ehrlich, D., Kemper, T., Tommasi, P., Airaghi, D. and L. Zanchetta, *Description of the GHS Urban Centre Database 2015, Public Release 2019, Version 1.0*, Publications Office of the European Union, Luxembourg, 2019, ISBN 978-92-79-99753-2, doi:10.2760/037310, JRC115586.

Contents

Acknowledgements

This Database is the product of collaboration between the whole Global Human Settlement Layer (GHSL) team of the European Commission's Joint Research Centre, Directorate E "Space, Security and Migration", Disaster Risk Management Unit (JRC.E.1) and experts from other groups in the E1 unit, in other JRC directorates, in the European Commission as well as from other institutions.

We are grateful to our collaborators from JRC and other institutions that were essential for the successful completion of the Urban Centre Database (R2019A) V1.0, and in particular:

— Elisabetta Vignati, Monica Crippa, and Diego Guizzardi, of the EC JRC Directorate C "Energy, Transport and Climate", Air and Climate Unit (JRC.C.5), for their extremely valuable help in integrating and understanding the data from the JRC Emissions Database for Global Atmospheric Research (EDGAR);

— Luca Montanarella and Panos Panagos of the EC JRC Directorate D "Sustainable Resources", Land Resource Unit (JRC.D.3), for their timely and prodigious support in interpreting the data collected from the Harmonized World Soil Database

— Peter Salomon and the JRC Global Flood Awareness System team (GloFAS) of the EC JRC Directorate E "Space, Security and Migration", Disaster Risk Management Unit (JRC.E.1), for their precious help provided in integrating the hydrological data included in this database;

— Alessandro Dosio, of the EC JRC Directorate E "Space, Security and Migration", Disaster Risk Management Unit (JRC.E.1) for the valuable help provided for apprehending the secrets of the Heatwave Magnitude Index data integrated in this database;

— Gustavo Naumann of the EC JRC Directorate E "Space, Security and Migration", Disaster Risk Management Unit (JRC.E.1) for the great help provided in integrating the data on annual precipitation and temperature from the CRU TS v. 4.02 gridded time-series dataset;

— Luca Vernaccini, supporting the Disaster Risk Management Knowledge Centre of the EC JRC, for the help provided in integrating the data from the UN Global Assessment report (GAR);

— Alexander Mackie and Ivan Hascic, of the Environmental Performance Information Division, in the Environmental Directorate of the Organisation for Economic Co-operation and Development (OECD), for their help in integrating the $PM_{2.5}$ concentration data used in this databse;

— John Schneider and Marc Pagani of the Global Earthquake Model (GEM) Initiative for the support in the accessing the new Global Seismic map integrated in this database;

Authors

Aneta J. Florczyk[a], Michele Melchiorri[b], Christina Corbane[a], Marcello Schiavina[a], Luca Maffenini[c], Martino Pesaresi[a], Panagiotis Politis[d], Filip Sabo[e], Sergio Freire[a], Daniele Ehrlich[a], Thomas Kemper[a], Pierpaolo Tommasi[f], Donato Airaghi[f], Luigi Zanchetta[a]

[a]European Commission, Joint Research Centre, Disaster Risk Management Unit, [b]Piksel S.r.l., Milano, Italy, [c]GFT Italia S.r.l., Milano, Italy; [d]Arhs Developments S.A., Luxembourg; [e]Arhs Developments Italia S.r.l., Milano, Italy; [f]Engineering Ingegneria Informatica S.p.A., Rome, Italy

Abstract

The Global Human Settlement Layer Urban Centres Database (GHS-UCDB) is the most complete database on cities to date, publicly released as an open and free dataset - GHS STAT UCDB2015MT GLOBE R2019A V1.0. The database represents the global status on Urban Centres in 2015 by offering cities location, their extent (surface, shape), and describing each city with a set of geographical, socio-economic and environmental attributes, many of them going back 25 or even 40 years in time. Urban Centres are defined in a consistent way across geographical locations and over time, applying the *"Global Definition of Cities and Settlements"* developed by the European Union to the Global Human Settlement Layer Built-up (GHS-BUILT) areas and Population (GHS-POP) grids.

This report contains the description of the dimensions and the derived attributes that characterise the Urban Centres in the database. The document includes notes about methodology and sources. The GHS-UCDB contains information for more than 10,000 Urban Centres and it is the baseline data of the analytical results presented in the Atlas of the Human Planet 2018.

1 Introduction

Cities and urban areas are today home to more than half of the world's seven billion people. The latest urbanization trends indicate that an additional three billion people will be living in urban areas by 2050 (UNDESA, 2018b), increasing the urban share of the world's population even more. However, the current era of rapid urbanization has been marred with inadequacy of capacity and sometimes resources to match urban development needs. Moreover, there is a gap in the global monitoring of the urbanization process in all its dimensions. Until today, there is no globally harmonised definition of cities and settlements, defining their size and boundaries.

This report presents the characteristics and contents of the Global Human Settlement Layer Urban Centre Database (GHS-UCDB), released as open and free dataset entitled "GHS Urban Centre Database 2015, multitemporal and multidimensional attributes, R2019A"[1] (Florczyk et al., 2019). The database is built upon the Degree of Urbanisation (Dijkstra & Poelman 2014), a definition used to outline the spatial extent of cities and settlements, to create the first global, harmonized, consistent database of Urban Centres. The database represents the global status on Urban Centres in 2015 by offering cities location, their extent (surface, shape), and describing each city with a number of geographic, socio-economic and environmental attributes, many of them going back 25 or even 40 years in time.

The GHS-UCDB is based on the Global Human Settlement Layer (GHSL) data. The GHSL project produces new, global, spatial information, evidence-based analytics and knowledge describing the human presence on the planet based mainly on two quantitative factors: i) the spatial distribution (density) of built-up structures, and ii) the spatial distribution (density) of resident population. Both factors are observed in the long-term temporal domain and per uniform surface units in order to support trends and indicators for monitoring the implementation of international framework agreements. The GHSL uses various input data including global, multi-temporal archives of fine-scale satellite imagery, census data, and volunteered geographic information. The satellite archives and available census data allow generating information layers for four epochs: 1975, 1990, 2000, and 2015.

The GHS-UCDB is presented first by presenting the principles and the fundamentals of the GHSL framework (Section 2), explaining the definition adopted to classify Urban Centres (section 3). Section 4 describing the thematic dimensions, and the derived attributes.

GHS Urban Centre Database 2015, multitemporal and multidimensional attributes, R2019A

Florczyk, A.J., Corbane, C., Schiavina, M., Pesaresi, M., Maffenini, L., Melchiorri, M., Politis, P., Sabo, F., Freire, S., Ehrlich, D., Kemper, T., Tommasi, P., Airaghi, D. and L. Zanchetta. 2019. GHS Urban Centre Database 2015, multitemporal and multidimensional attributes, R2019A. European Commission, Joint Research Centre (JRC) [Dataset] PID: http://data.europa.eu/89h/53473144-b88c-44bc-b4a3-4583ed1f547e

The Atlas of the Human Planet 2018

European Commission, Joint Research Centre, Atlas of the Human Planet 2018 – A World of Cities, EUR 29497 EN, European Commission, Luxembourg, 2018, ISBN 978-92-79-98185-2, doi:10.2760/124503, JRC114316.

http://dx.doi.org/10.2760/124503

The Atlas of the Human Planet 2018 (European Commission, Joint Research Centre, 2018) and this report mutually reinforce. While the first proposes key messages, narratives, and research perspectives derived from the GHS-UCDB, this report contain

[1] Dataset ID: GHS STAT UCDB2015MT GLOBE R2019A V1.0

the information required for a fair and informed use of the database. The GHS-UCDB is an ultimate example of the reach of harmonised data integration, serving the GEO (Group on Earth Observations) Human Planet Initiative[2]. The initiative maximises the use of (big) open data through bringing EO data into the socio-economic and other domains. By developing a new generation of measurements and information products, the initiative provides new scientific evidence and a comprehensive understanding of the human presence on the planet that can support global policy processes with agreed, actionable and goal-driven metrics.

[2] https://www.earthobservations.org/activity.php?id=119

2 The Global Human Settlement Layer

The GHSL project produces new, global, spatial information, evidence-based analytics and knowledge describing the human presence on the planet based mainly on two quantitative factors: i) the spatial distribution (density) of built-up structures, and ii) the spatial distribution (density) of resident population. Both factors are observed in the long-term temporal domain and per uniform surface units in order to support trends and indicators for monitoring the implementation of international framework agreements. The GHSL uses various input data including global, multi-temporal archives of fine-scale satellite imagery, census data, and volunteered geographic information. The satellite archives and available census data allow generating information layers for four epochs: 1975, 1990, 2000, and 2015. The GHSL uses satellite remote sensing as a primary source of information to delineate and size the physical extent of human settlements from large megacities to villages and towns. In the GHSL framework, the physical extent of the human settlement as collected by the satellite sensor is called "built-up area". This GHSL built-up area grid (GHS-BUILT) – is produced by application of automatic supervised classification data processing methods to global streams of open decametric-resolution satellite imagery collected by the Landsat and Sentinel missions.

The GHSL project produces also global population density grids (GHS-POP), by combining the GHS-BUILT with census data through spatial modelling techniques. The data integration process foresees the downscaling of the information from the national census district level to a regular, finer-scaled, gridded built up density information layer. The result is a population density information layer available at a 250x250 m² and 1x1 km² grid scale. The combination of GHS-BUILT and GHS-POP is used to produce the GHS settlement model building on the "Degree of Urbanization" concept (Dijkstra & Poelman 2014) applied to the GHSL data (GHS-SMOD). The GHS-SMOD is implemented at the 1 km scale and, in the version used here, distinguishes between three main typologies of human settlements, based on population density cut-off values: "Urban Centres", "Urban Clusters" and "Rural Settlements". The Urban Centres, which represent the most densely inhabited part of human settlements, are analysed in more detail in this atlas. They are extracted from the GHS-SMOD of the epoch 2015 and include more than 10,000 individual cities. Each Urban Centre is characterised by a number of variables describing the geography and the environment of the place as well as socio-economic parameters and the potential exposure of an Urban Centre to natural disasters. The high-level process of information extraction and aggregation of the information in the GHS-UCDB is schematised in Figure 1. Conceptual schema of the GHSL input data, processing and products.

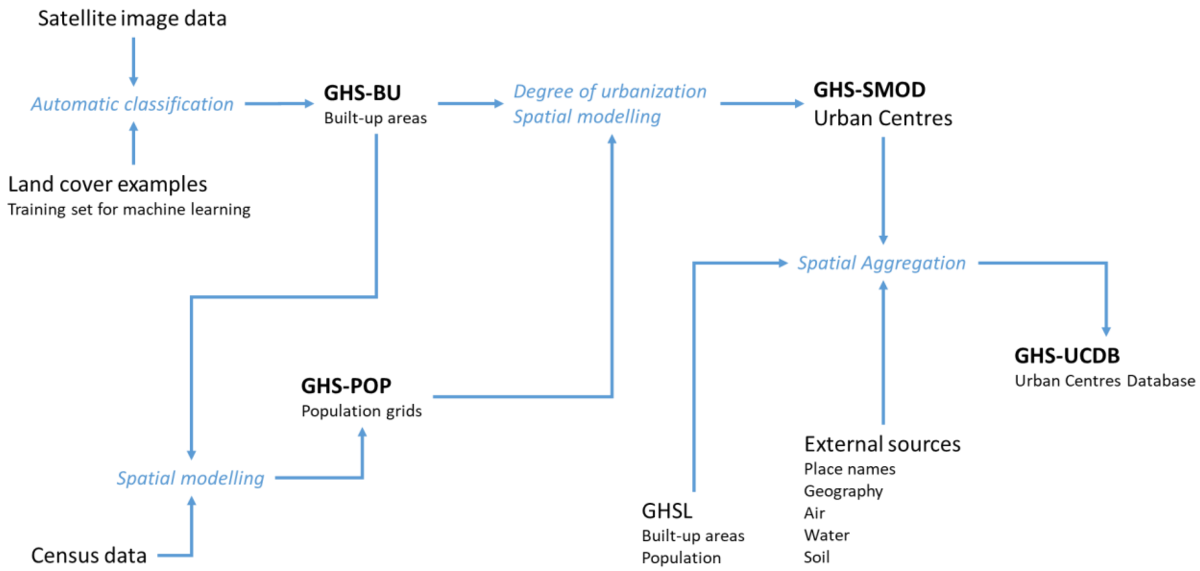

Figure 1. Conceptual schema of the GHSL input data, processing and products.

2.1 Fundamentals

The GHSL consists of three main information components hierarchically placed at three different levels of abstraction: built-up areas, population and the settlement model grids.

At the base of the hierarchy - including the most spatially accurate and the least abstract information level - we have a layer collecting concrete evidences about the human presence on the planetary surface as seen from global Earth Observation systems. In the GHSL paradigm, the fundamental link between Earth Observation sensor data and the human presence is the observable presence of built-up structures or buildings. From the GHSL perspective, the "building" makes the physical part of the human settlement fabric or spatial extension that is observable and measurable using the available global sensors. At this basic level the GHSL reports about *built-up areas*, as *areas (spatial units) where buildings can be found* (Pesaresi et al., 2016). The concept of "buildings" formalized by the GHSL are *enclosed constructions above ground which are intended or used for the shelter of humans, animals, things or for the production of economic goods and that refer to any structure constructed or erected on its site* (Pesaresi et al., 2013). This abstraction is very similar to the standard topographic definition of the "building" class as compiled in the INSPIRE directive[3], except that the condition of the *permanency of the structure* it is not in the GHSL definition. The GHSL definition of built-up also to includes refugee camps, informal settlements, slums and other temporary settlements and shelters.

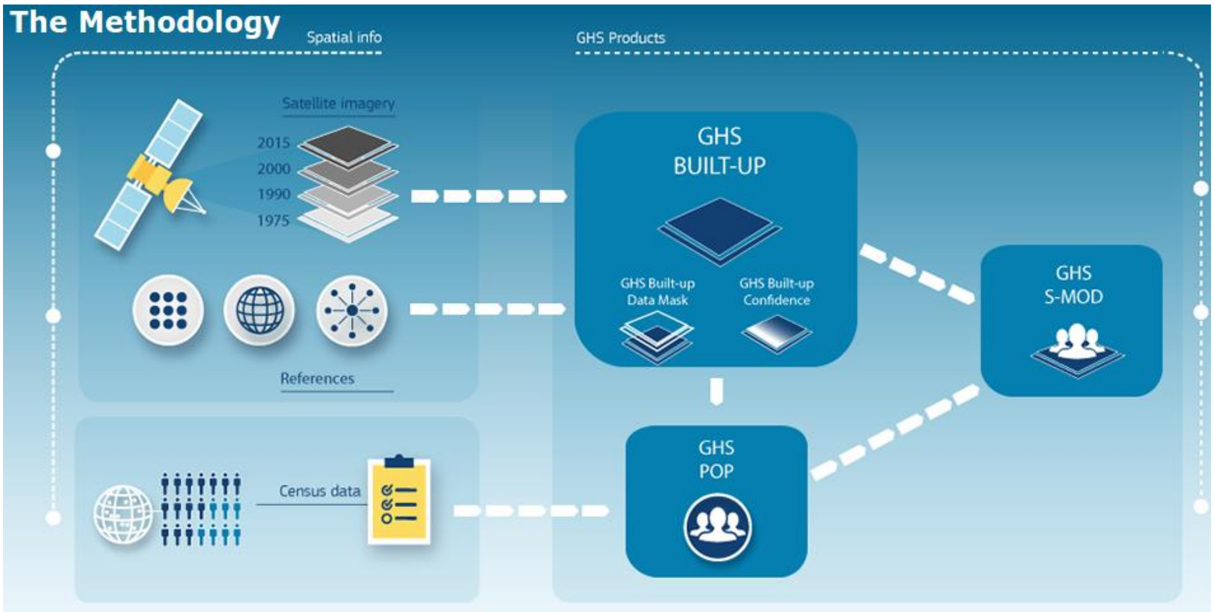

Figure 2. Overview on the main data components in the GHSL framework.

The intermediate abstraction information layer of the GHSL is the population grid that is produced in an in-between spatial resolution. This information layer is derived from the combination of global collections of national population census data and global built-up areas as extracted from Earth Observation data analytics (Figure 2). In the approach taken by the GHSL, the population data collected by national censuses with heterogeneous criteria and heterogeneous update time are harmonized in the space and time domains in to the GHS-POP grids, by systematic and consistent application of the same set of data interpolation and spatial disaggregation methods to the best available global spatial baseline data (Freire_et al., 2016)

The top abstraction information layer of the GHSL it is the settlement model classification grid. It is provided with the least spatial detail (1 km) by combining the two less-abstract and more-spatially-detailed built-up and population grids, GHS-BUILT and GHS-POP,

[3] INSPIRE Infrastructure for Spatial Information in Europe D2.8.III.2 Data Specification on Buildings – Draft Technical Guidelines http://inspire.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_BU_v3.0rc3.pdf

respectively. The GHS-SMOD model implemented by the GHSL it is consistent with the "Degree of Urbanisation" (DEGURBA) model adopted by EUROSTAT[4]. It discriminates 3 settlement class abstractions: 1) Cities, 2) Towns and suburbs and 3) Rural areas. The discrimination is based on the population density in the square kilometre grid[5], total settlement population and other spatial generalization parameters.

In the GHSL paradigm, the base layer GHS-BUILT it is designed to be the most stable against different visions and approaches, while GHS-SMOD is the most abstract and as such exposed to conceptual changes and alternative problem settings proposed by the different stakeholders involved in the post-2015 international framework processes. The modular hierarchical abstraction schema used in the GHSL design allows to protect the investment made in the global, fine-scale information gathering from perturbations on the abstract classification schema that may be introduced by different decision-makers involved in the process and potentially producing different problem setting and abstractions. On the other side, the modular hierarchical abstraction schema facilitates the test of alternative abstract models on the same agreed information baseline, facilitating the discussion and the comparison of the results also between international stakeholders not necessary sharing the same high abstraction definitions.

The following sections help the reader to understand fundamental concepts of GHSL and its data (Figure 3). Section 2.1.1 with extraction of information from satellite imagery and built-up definition. Section 2.1.2 second paragraph explore the process allows to combine built-up grids with census data to produce the population grids. Section 2.1.3 shows with simple images, and example of three GHSL datasets (GHS-BUILT, GHS-POP and GHS-SMOD) for the city of Madrid, Spain. Section 3 introduces the Degree of Urbanisation in the frame of the work on global definition, and presents the details of the Urban Centre delineation.

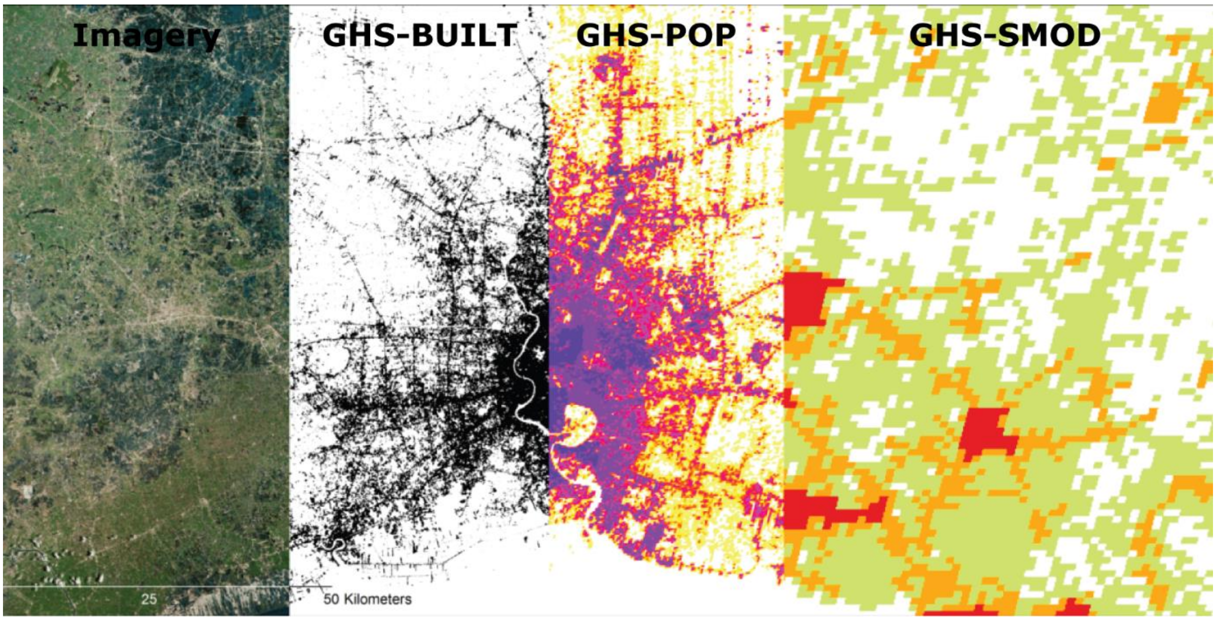

Figure 3. Transition from imagery to built-up areas extraction (GHS-BUILT), population modelling (GHS-POP), and settlements classification (GHS-SMOD), examples in the area of Bangkok (Thailand).

[4] http://ec.europa.eu/eurostat/web/degree-of-urbanisation/overview

[5] densely, intermediate density and thinly populated areas

2.1.1 From Earth's surface to built-up area

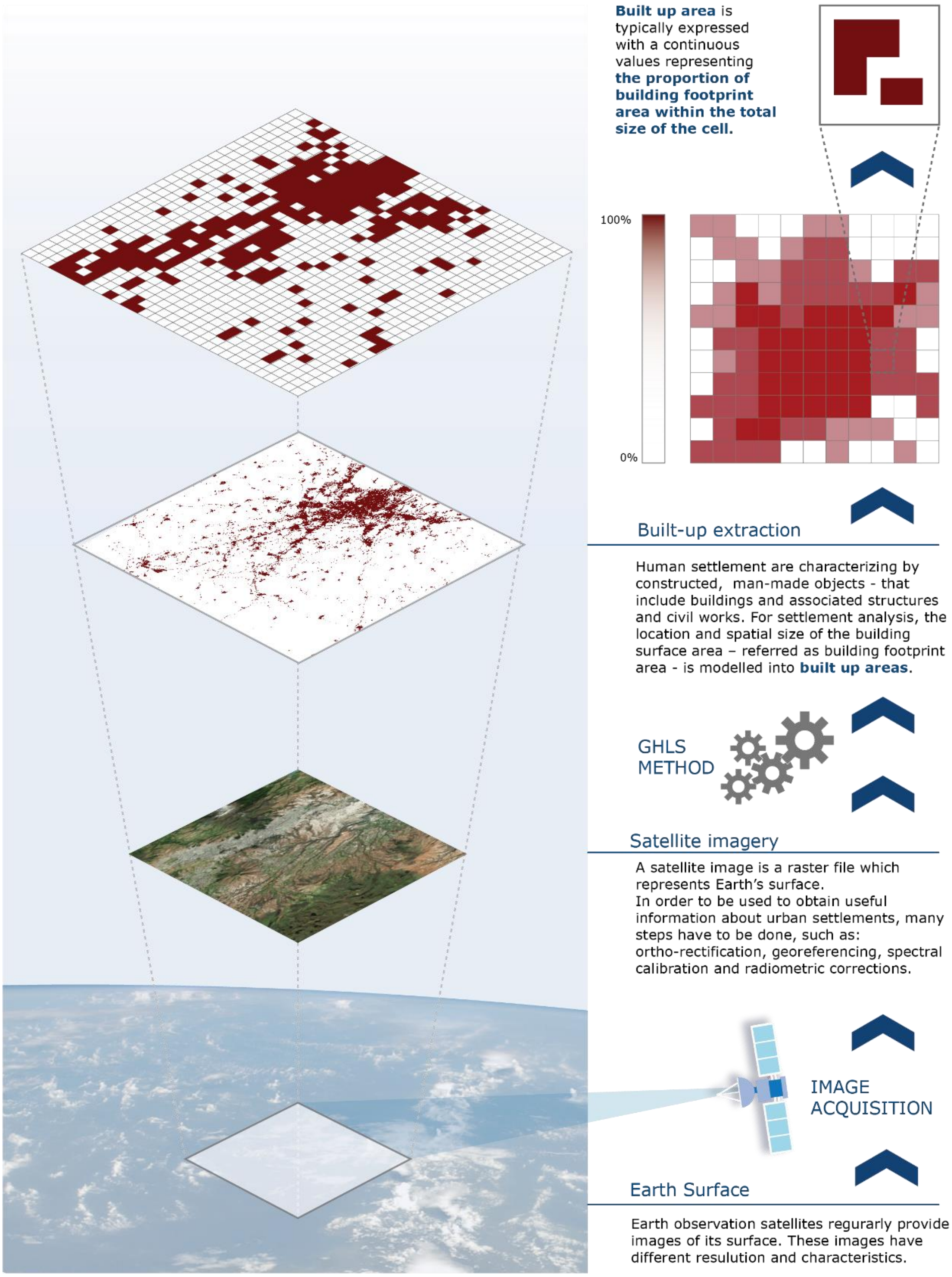

Figure 4. The information extraction process from the satellite images of the earth surface (bottom) to the built-up area extraction (middle) to the aggregated built-up area density (top).

2.1.2 From Built-up area to population grid

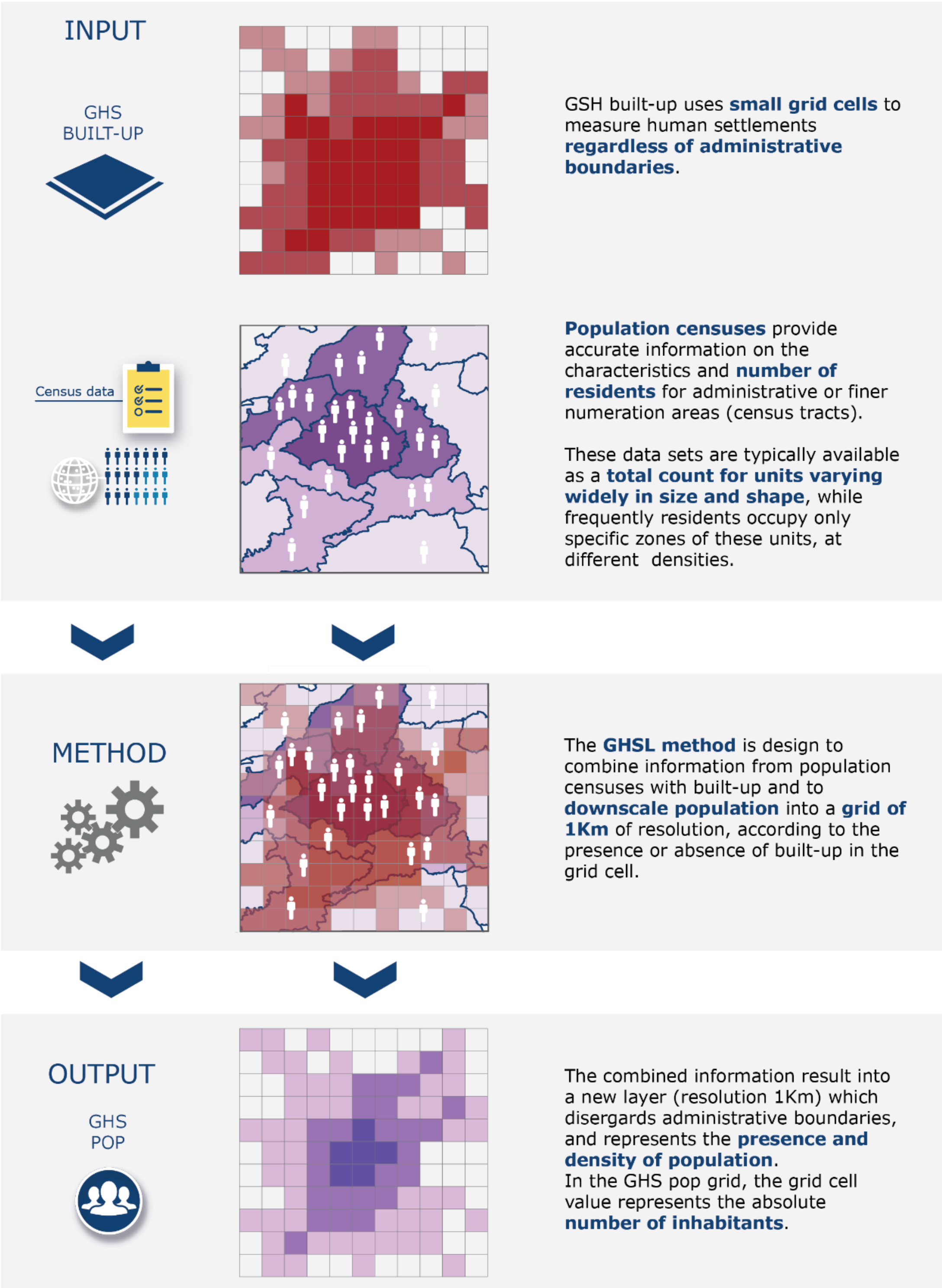

Figure 5. Illustration of the combination of GHS-BUILT with the census data to produce a regular fine scale grid of population density.

2.1.3 An example from the city of Madrid, Spain

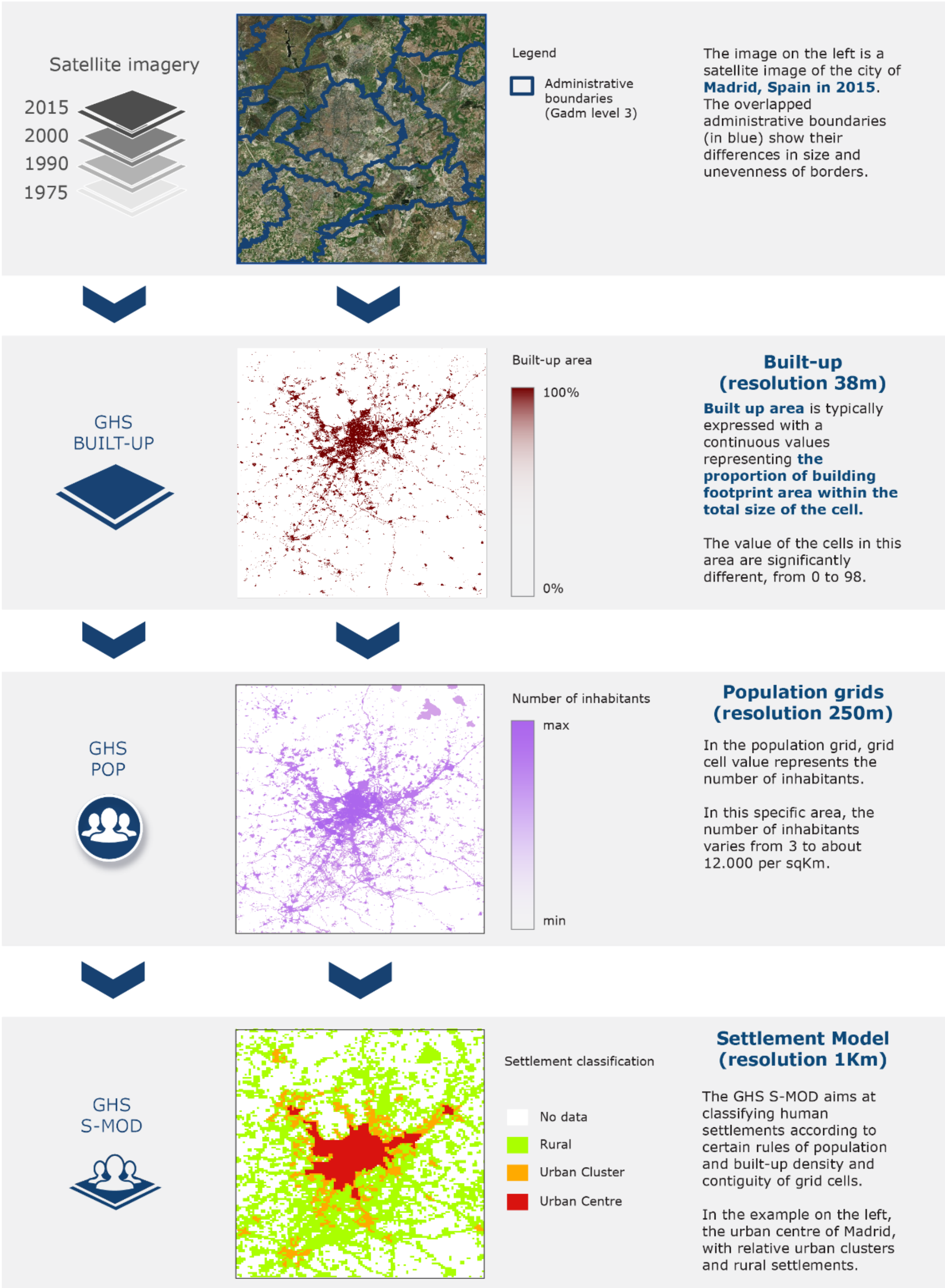

Figure 6. The GHS-BUILT and GHS-POP are combined to classify the gird cells into rural and urban areas.

3 Urban Centres

In this Section, we introduce the Degree of Urbanisation in the frame of the work on the global harmonised definition of cities and settlements, and present the details of the Urban Centre delineation.

3.1 Global definition

The Department of Economic and Social Affairs of the United Nations (UN) reports in the 2018 Revision of the World Urbanisation Prospect: *"globally, more people live in urban areas than in rural areas, with 55 % of the world's population residing in urban areas in 2018"* (UNDESA, 2018b). However, the UN acknowledges that the analysis relies on the data produced by national sources, which reflect the definitions and criteria established by national authorities. Therefore, the criteria used to identify urban areas vary from country to country and may not be consistent even between different data sources within a given country. About half of the definitions described in the report's methodological annex, include a minimum population size, either exclusively or in combination with other indicators or criteria. A specific size threshold is mentioned for 100 countries. Of these, the vast majority (85%) use a threshold of 5,000 or less. The most popular thresholds are 5,000 with 27 countries and 2,000 with 24 countries. Japan and China are outliers with thresholds that are ten to twenty times higher, respectively 50,000 and 100,000.

To address this issue, the European Union, the Organisation for Economic Cooperation and Development (OECD) and the World Bank launched a voluntary commitment to develop a global, people-based definition of cities and settlements during the third United Nations Conference on Housing and Sustainable Urban Development (Habitat III) in October 2016. Since then Food and Agriculture Organisation (FAO) and UN-Habitat have also joined this commitment. FAO is the 'custodian' UN agency for 21 SDG indicators for which a harmonised definition of rural areas is needed. UN-Habitat is the 'custodian' UN agency for 8 SDG indicators, for which a harmonised of city definition is needed. These definitions are designed to support the New Urban Agenda[6], the global strategy to improve agricultural and rural statistics (GSARS)[7] and the monitoring of Sustainable Development Goals (SDG). Agreeing the first global consistent definition of cities and rural areas can also help many other policies and research areas.

3.2 Model Description

The "Urban Centres" (UC) using the Degree of Urbanisation are defined as: "high-density clusters of contiguous grid cells of 1 km^2 with a density of at least 1500 inhabitants per km^2 and a minimum population of 50000" (Dijkstra & Poelman, 2014) (Figure 7). The UC as implemented in the current GHSL settlement model (SMOD) formulation is defined as: *"the spatially-generalized high-density clusters of contiguous grid cells of 1 km^2 with a density of at least 1,500 inhabitants per km2 of land surface or at least 50% built-up surface share per km2 of land surface, and a minimum population of 50,000."* The original degree of urbanisation was based on the population size, density and contiguity of local administrative units level 2 (LAU2). However, this method is based on LAU2s, which vary considerably in area size; hence the results are distorted and reduce the comparability between countries with large LAU2s and small LAU2s. This well-known problem, known as Modifiable Area unit problem (Gehlke & Biehl, 1934), is solved by using the GHSL population grid (or any other high-resolution population grid).

[6] http://nua.unhabitat.org/
[7] http://gsars.org/en/

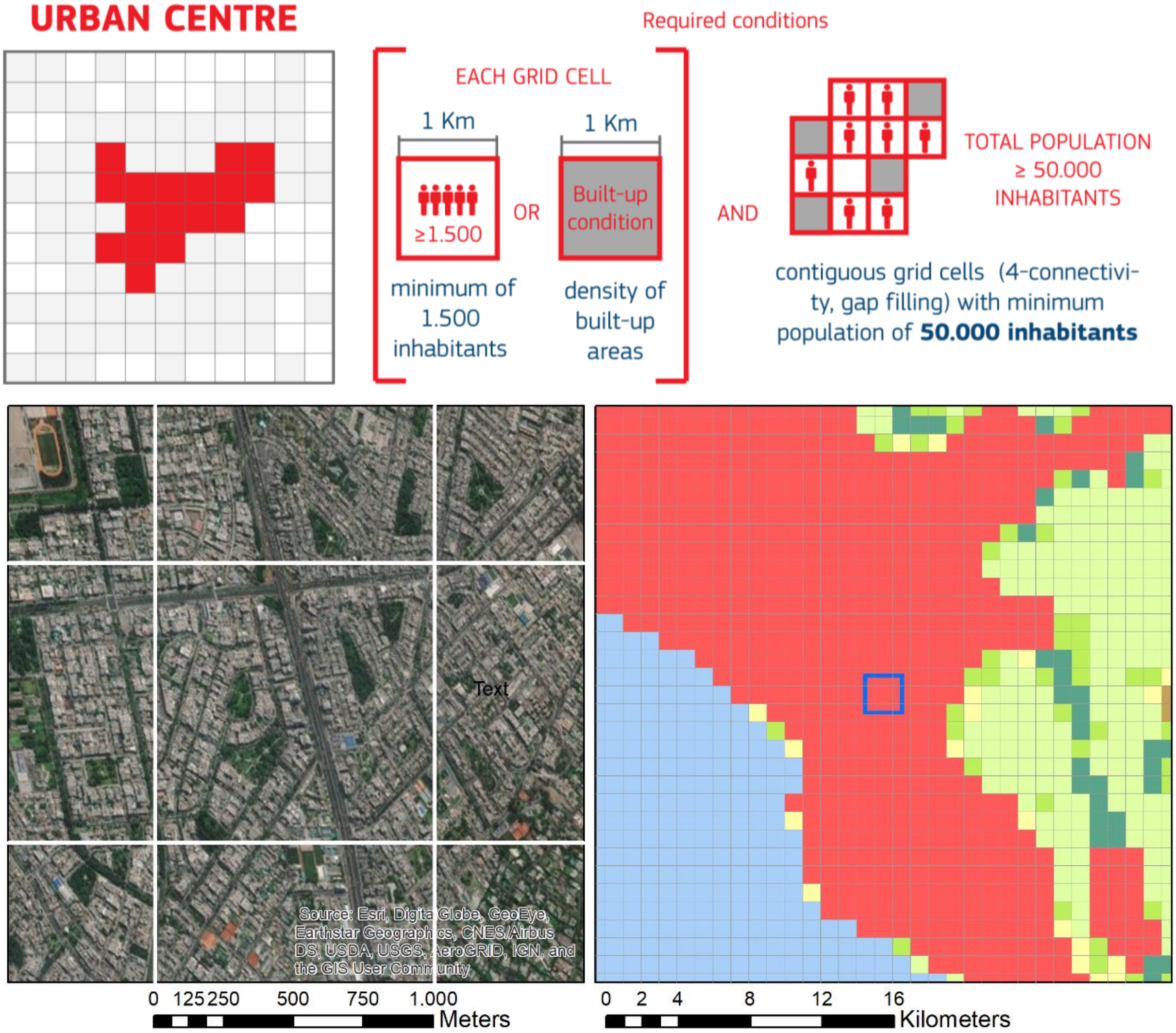

Figure 7. Schema representing an abstract Urban Centre, grid cell criteria, and population size threshold; and example of Lima HDC (area represented with imagery outlined in blue).

This is illustrated for the city of Cork, Ireland (Figure 8). Cork lies at the mouth of river lee into Lough Mahon. The land use is characterised by a typical mix of residential areas and some patches of industrial/commercial units. According to the land use map the city is well confined and is surrounded by agricultural areas and semi-natural vegetation (top left). The highest population densities are in the Urban Centre and the residential areas, dropping significantly in the agricultural areas (top right). The Degree of Urbanisation grid (bottom right) clearly reflects this. It maps very well the dense Urban Centre of Cork and classifies the less dense suburbs as urban clusters connected with the Urban Centre or as individual town. When comparing this with the Degree of Urbanisation applied to the LAU2 (bottom-left), the Urban Centres match well. However, the surrounding LAU2s are all classified as towns and suburbs, despite the fact that only a small part of the LAU2 is occupied by settlements. This problem aggravates, when working at a global scale, where administrative layers are often available only at district level or worse.

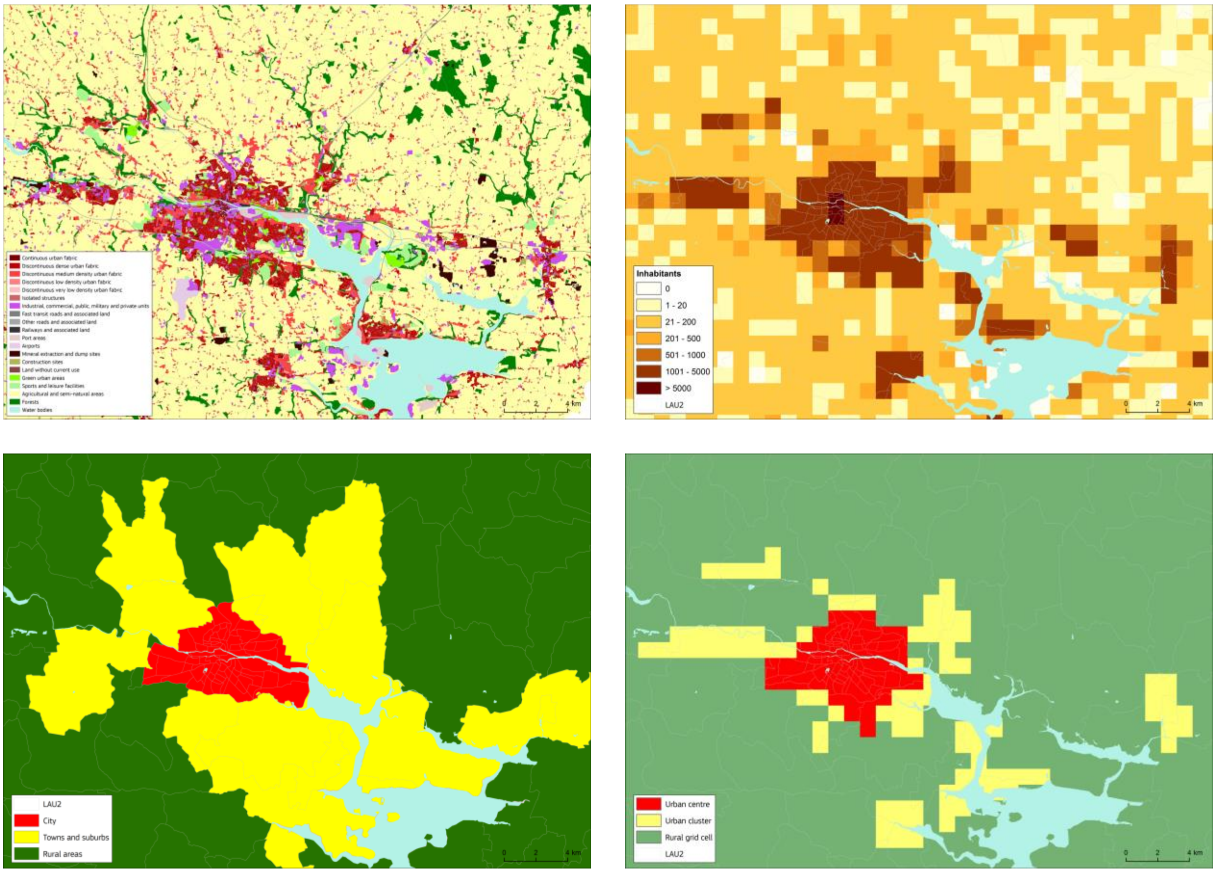

Figure 8. City of Cork (Ireland). Land use map based on Urban Atlas[8] (top left); GEOSTAT population grid[9] (top right); Degree of Urbanisation applied to the GEOSTAT grid (bottom right); Degree of Urbanisation applied to the LAU2 (bottom left).

3.3 Urban Centres Delineation

Be X a squared grid of cells x with uniform surface of 1 km^2 representing the global Earth surface with a World Mollweide pseudo cylindrical geographical projection. Be POP_x the estimated amount of resident population in the cell x by the GHSL, and be BU_x the estimated share of built-up surface in the cell x by the GHSL by observing 30-m-res satellite measurements. Be $LAND_x$ the estimated share of permanent land surface in the cell x by a GHSL processing of the JRC Global Surface Water (GSW) input data. $LAND_x$ was estimated as $LAND_x = 1 - w_x^{0.8}$, with $w_x^{0.8}$ the share of surface water occurring for more than 80% of the 30-m-resolution satellite measurements of the past 20 years (15 days interval) in the cell *x.*

Given the above, the densities of population and built-up areas per land surface are calculated as follows:

$$POP_x' = \frac{POP_x}{LAND_x}$$

$$BU_x' = \frac{BU_x}{LAND_x}$$

Consequently, the support set of the High Density Cluster (HDC) spatial domain is determined as follows:

[8] https://land.copernicus.eu/local/urban-atlas
[9] https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/population-distribution-demography/

$$HDC_x^{supp} = \{x : POP_x' > 1500 \cup BU_x' > 0.5\}$$

The HDC_x^c spatial clusters are determined by the application of the "*contiguous grid cells of 1km²*" criteria of the root definition, by assuming 4-connectivity rule on the grid X representing the HDC_x^{supp} set. 4-connected samples are neighbours to every sample that touches one of their edges. These samples are connected horizontally and vertically. The 4-conn rule it is showed in the schema below. Respect to the sample X, the cells in the horizontal and vertical directions and one step of displacement are considered adjacent. The cells in the diagonal along the grid are not considered adjacent.

	not	yes	not	
	yes	X	yes	
	not	yes	not	

The population size of each cluster *c* of the HDC_x^c is calculate as

$$POP_x^c = \sum POP_x \cap HDC_x^c$$

Finally, the HDC_x^c clusters are selected so that $POP_x^c > 50000$

$$HDC_x^{50K} = HDC_x^c, c : POP_x^c > 50000$$

Subsequently, the HDC_x^{50K} individual clusters that passed the above test are processed by a spatial generalization procedure *G* including an iterative local union-majority filter (also called "smoothing") until idempotence is reached, followed by a gap-filling step. That gap-filling step is filling all the holes remaining after the smoothing, and having an area less than 15 km². The local union-majority filter applied in the generalization *G* has a kernel *K* of 3x3 spatial units in the grid X, corresponding to a surface of 9 km².

At each iteration, the union-majority filter Φ it is defined as follows:

$$\Phi_x = \begin{cases} x = true \rightarrow \Phi_x = true \\ x = false \rightarrow \begin{cases} \sum_{x \in K} true_x > \frac{\hat{n}^K}{2} \rightarrow \Phi_x = true \\ \sum_{x \in K} false_x > \frac{\hat{n}^K}{2} \rightarrow \Phi_x = false \end{cases} \end{cases}$$

With $\frac{\hat{n}^K}{2}$ being the half of the number of samples included in the kernel *K* considered in the spatial filtering. In the specific case, $\frac{\hat{n}^K}{2} = \frac{3x3}{2} = 4.5$

The iterative local union-majority filter Φ it is applied individually to each HDC, testing at each iteration that the total number of HDCs determined before the filtering process it is maintained constant. Consequently, avoiding the merging of two distinct HDCs because of the spatial filtering process.

Figure 9 presents the baseline data and the Urban Centre delineation result on an example from the Northern Region of China. The whole process can be summarised in seven major steps, which is represented as a sequence of Figures in Annex 1.

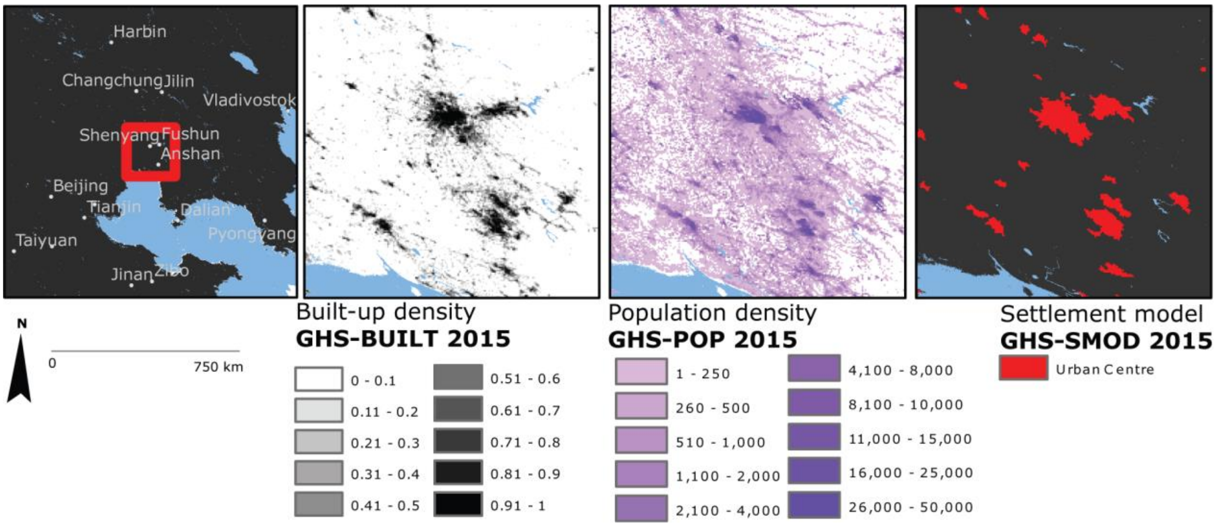

Figure 9. Example at 1km resolution of GHSL baseline data (GHS-BU, GHS-POP) for the delineation of Urban Centres in the Northern Region of China.

4 Urban Centres Database

The Urban Centres Database (GHS-UCDB) builds on the improved GHS-BUILT and GHS-POP grids published as the Community Release in 2018 (Florczyk et al., 2018). The Urban Centres are spatially delimited by applying the "degree of urbanization" model. The GHS-UCDB (Figure 10) was generated by spatial integration of the Urban Centres with the GHSL data and with other sources related to five main thematic areas: geography, socio-economic, environment, Disaster Risk Reduction, and Sustainable Development Goals (Table 1). The principal methods used to derive the attributes to characterise Urban Centres through geospatial processing is displayed in Figure 11. An additional list of attribute complements the one by thematic area (General information and, Multitemporal Urban Centre spatial domain). It includes mainly classification and reference attributes (such as name, latitude-longitude, etc.). Figure 12 shows a regional distribution of the Urban Centres in the database.

Table 1. Overview of dimensions and variables in the Urban Centre Database 2015, and their temporal extent.

Dimensions	Variable	Temporal coverage			
		1975	1990	2000	2015
General information	Identification				
	Extension				
	Location				
	Name				
Multitemporal Urban Centre spatial domain	Number				
	Area				
Geography	Elevation				
	Biome				
	Climate	1986-2010			
	Soil				
	River basin				
	Temperature				
	Precipitation				
Socio-economic	Resident population				
	Built-up area				
	Night time light				
	GDP				
	Development				2018
	Accessibility & remoteness				
Environment	Greenness				
	Pollutants' emission				
	Pollutants' concentration				
DRR	Floods (exposure)				
	Earthquake (hazard estimate)				
	Storm surge (exposure)				
	Maximum magnitude of the heatwaves	1980-2010			
SDG	Land Use Efficiency (11.3.1)	1990–2015			
	Open spaces (11.7.1 –proxy)				

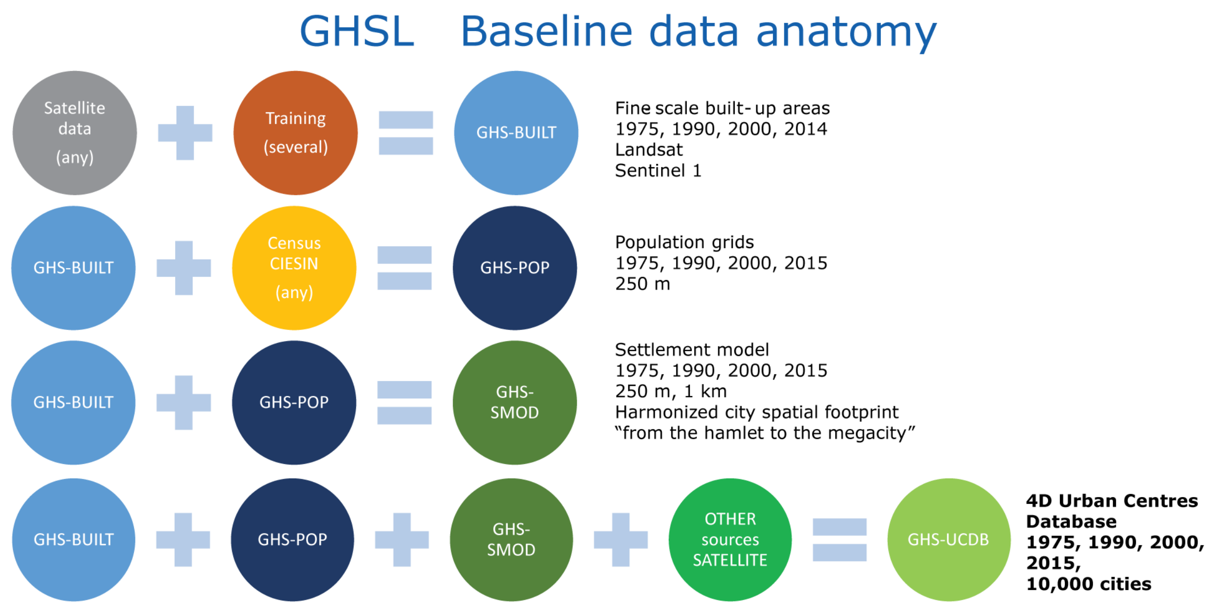

Figure 10. Anatomy of the key information components and data processing/integration used to create the GHS-UCDB

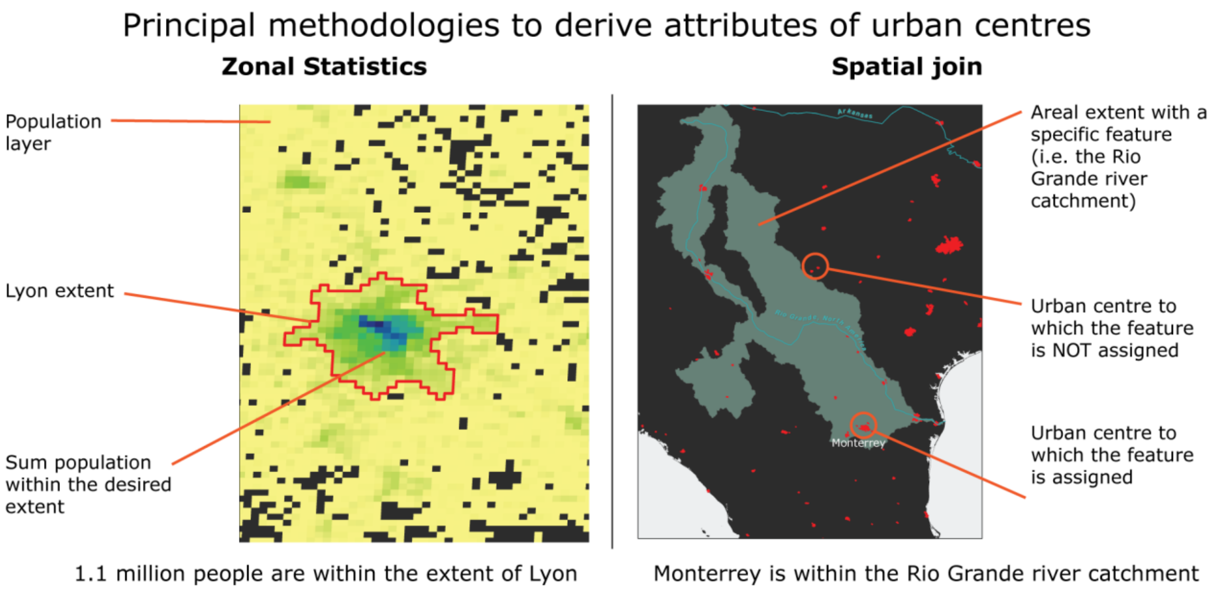

Figure 11. Overview of the main geospatial operations applied to derive attributes of Urban Centres.

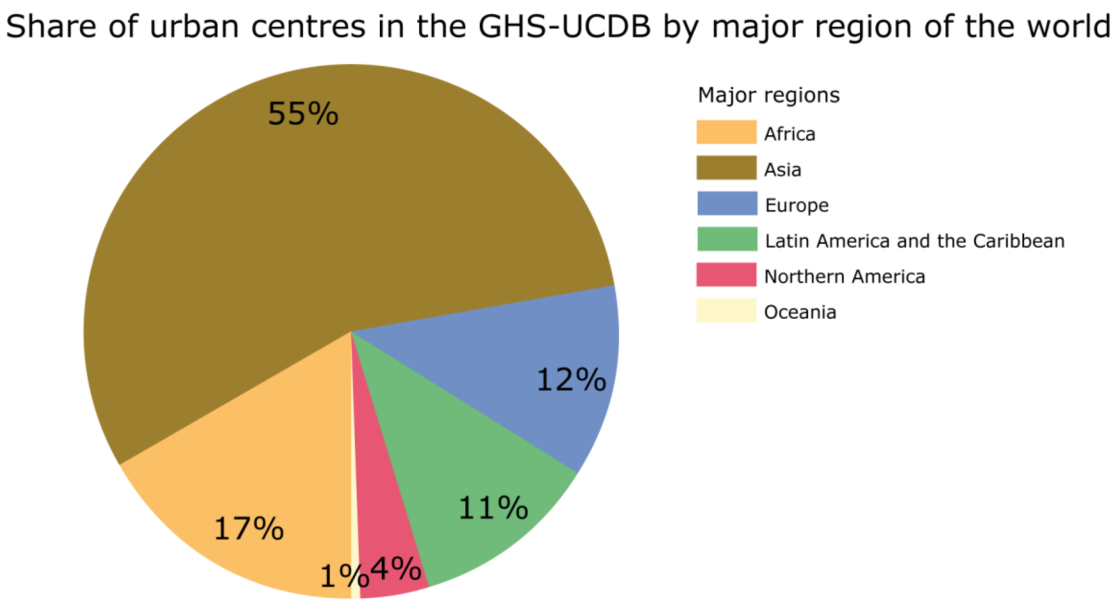

Figure 12. GHS-UCDB regional coverage, share of centres by major region of the world.

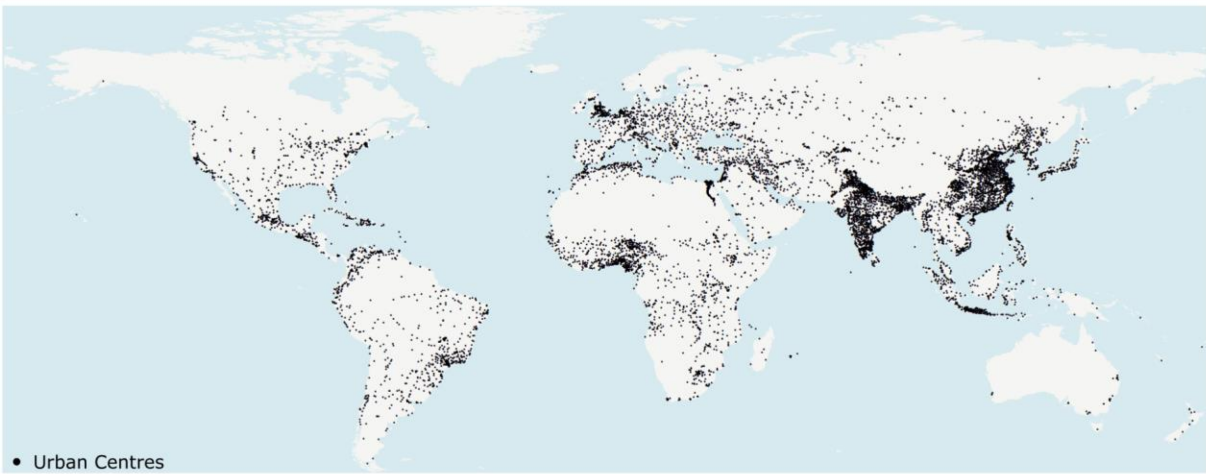

Figure 13. Status on spatial distribution of Urban Centres in 2015.

Urban Centre Database describes more than 10,000 Urban Centres delineated and uniquely identified from the GHS-SMOD grid of epoch 2015. Therefore, the database represents the global status on Urban Centres in 2015 (Figure 13. Status on spatial distribution of Urban Centres in 2015.). The thematic dimensions of this urban space are described via a set of variables, of which some are multitemporal (when available).

4.1 Dimensions

The Urban Centre Database describes several dimension of the delineated urban space. These dimensions are organised in the following categories: *general characteristics, multitemporal Urban Centre domain*, *geography*, *environment*, *socio-economy*, *disaster risk reduction*, and *sustainable development goals*. Detailed list of attributes of each dimension is enumerated in the Section 4.2.

4.1.1 General characteristics

General characteristics of the Urban Centre gathers basic information about the delineated urban space. This database gathers the following general information:

- ***Control codes*** (Urban Centre unique ID and quality control code),
- ***Urban Centre Extension*** (area, bounding box)
- ***Location*** (centroid, country or countries and geographical region where the Urban Centre is located),
- ***Naming*** (assigned name and / or list of names).

4.1.2 Mutlitemporal Urban Centre Domain

This dimension gather basic information related to the mutitemporal characteristics of Urban Centres. It gathers results from the analysis of the multitemporal GHS settlement model grids (GHS-SMOD) for epochs 1975, 1990 and 2000.

The global definition of cities and settlement is applied to the multitemporal GHSL data GHS-BUILT and GHS-POP grids of epochs 1975, 1990 and 2000 (published as the Community Release in 2018 (Florczyk et al., 2018)) per each epoch separately, resulting in three GHS-SMOD grids: GHS-SMOD 1975, GHS-SMOD 1990 and GHS-SMOD 2000. Each Urban Centre described in this database (i.e., Urban Centre of 2015) is characterised by *number* of different Urban Centres and their *total sum of area* per each past epoch (1975, 1990 and 2000), calculated for each Urban Centre by intersecting its polygon (i.e., spatial domain) and the three GHS-SMOD grids.

4.1.3 Geography

Geographic conditions have long enabled the emergence and expansion of cites and Urban Centres, but also affect and constraint their growth and development. The global set of Urban Centres, as mapped by GHSL, is intersected with a set of relevant geographic variables for which global geospatial datasets were available, but novel attributes were also computed. Enriching the Urban Centres database with these physiographic attributes enables analysis of the interplay of cities with their geographic setting. The geographic dimension of the GHS-UCDB includes:

- **Biome;**
- **Soil;**
- **Elevation;**
- **Climate;**
- **Temperature;**
- **Precipitation;**
- **River Basins.**

Biomes are the most basic units that ecologists use to describe global patterns of ecosystem form, process, and biodiversity. Interestingly, existing descriptions of biome systems mostly ignore human influence or use a limited number of anthropogenic ecosystem classes (Ellis, 2018, Ellis & Ramankutty, 2008). Without entering into the on-going discussions in ecology, this section uses the Terrestrial Ecosystems of the World map, which contains a distinct assemblage of natural communities sharing a large majority of species, dynamics, and environmental conditions, to map each Urban Centre into one of its biomes. Urbanisation is reported as one of the most important threats to biodiversity worldwide as urban areas may threaten ecosystems through direct habitat conversion (Clergeau et al., 1998, Blair, 1999, McKinney, 2002). Also, high concentrations of human population has various indirect effects, which include freshwater contamination, waste generation, resource use or habitat fragmentation (Mikusinski & Angelstam, 1998).

The rapid urbanisation and population growth described by the GHSL data have a strong impact on *soils*. In the process of urbanisation many soils are permanently sealed or considerably altered due to excavation, mixing and compacting. This changes the capacity of food production and other diverse ecological services. It consequently increases the risk of potential floods and water scarcity, endangers biodiversity, and leads to environmental change on a larger scale. Soil type attribute allows to analyse, which are the soil types that are mostly affected by urbanisation.

Elevation affects humans in direct and indirect way. The direct effect is on human health. In fact, over the 2500 m altitude that the lower oxygen content may be not suited to everybody, in addition, altitude brings harsher climatic conditions. The indirect effect relate to the ruggedness that comes with living in elevated areas that makes , it makes it more prone to natural hazards, especially landslides, flash floods, but also earthquakes. Elevation and ruggedness also hampers accessibility of Urban Centres that are more difficult to get to and for which goods and transports are more costly to deliver. A direct effect of elevation in low elevated coastal areas is the increase risk to coastal flooding due to sea level surge and also climate induced sea level rise.

Human settlements interact with the *climate* conditions in which they lay (Landsberg, 1976). Climate characteristics are principally important for weather related disasters (i.e. floods, cyclones), energy use, for the adoption of specific building construction techniques and materials, and for health (i.e. air pollution).

Global *temperature* is a popular metric for summarizing the state of global climate. The Intergovernmental Panel on Climate Change in its most recent report (AR5) in 2013 stated: *'Warming of the climate system is unequivocal, and since the 1950s, many of the observed changes are unprecedented over decades to millennia. The atmosphere and ocean have warmed, the amounts of snow and ice have diminished, sea level has risen, and the concentrations of greenhouse gases have increased'* (IPCC, 2014). According to (Morice et al., 2012), the period 2001-2010 (0.49°C above the 1961-90 average) was 0.21°C warmer than the 1991-2000 decade (0.28°C above the 1961-90 average).

Among the effects of global warming is the increase in atmospheric evaporative demand, which intensifies the hydrological cycle, resulting in more intense and frequent storms, but also contributing to drying over some land areas. Increasing global temperatures are very likely to lead to changes in *precipitation* pattern, due to changes in atmospheric circulation. Overall, global land precipitation has increased by about 2% since the beginning of the 20th century (Jones & Hulme, 1996). The increase is statistically significant, though neither spatially nor temporally uniform (Doherty et al., 1999).

Water is an essential element in the Earth's system. It is critical for socio-economic development, healthy ecosystems and for human survival itself. At the same time, the excess of water (in the form of inundation and flooding) poses a threat to socio-economic development. Water and water management in all its forms is at the core of any sustainable development. However, population growth, agricultural intensification, urbanization and industrial production are putting pressure on fresh water resources. The Sustainable Development Goals have recognized this issue and dedicate Goal 6 to 'Ensure availability and sustainable management of water and sanitation for all'. Therefore, the attribute on *major river basin,* in which the Urban Centres is located, is introduced in the database.

4.1.4 Socio-economy

The Urban Centre variables directly linked to the socio-economic development discussed are:

— **Resident population 1975-1990-2000-2015;**
— **Built-up surface 1975-1990-2000-2015;**
— **Night time light emissions 2015;**
— **Income class 2018;**
— **Development group 2018;**
— **Gross domestic product; and**
— **Travel time to the capital 2015.**

The multitemporal information on the total resident population and built-up surface are used to evaluate the rate of development, and are directly linked with land consumption. Also, the harmonised grid with both information allows assessing the ratio of people per built-up area unit (for example, km^2), how much of constructed area is accounted per person, and how these relations changes over time. These measures might be useful as a proxy of the quality of life measure.

Another variable estimates amount of night light emitted by given urban area. The Night Light Emission data recorded by satellite platforms have been introduced in a number of application areas (Elvidge et al., 2007, Elvidge et al., 2017). In particular, they have been prosed for global urban delineation (Elvidge et al., 2010), for the production of spatially explicit measure of human development (Elvidge et al., 2012), as proxy measure of human well-being (Ghosh et al., 2013), and for post-conflict humanitarian needs assessment (Corbane et al., 2016).

Another indicator of development status of an area might be the Gross Domestic Product (GDP) at Purchasing Power Parity (PPP), which measures the monetary value of goods and services produced in a given period of time. Therefore, the GDP (PPP) variable is introduced in the database.

In order to align with the commonly used metrics on human development, the database incorporates the classification of Urban Centre according the *income class* and *development group* of UN World Urban Prospect 2018 (UNDESA, 2018a), derived from the classification of the country the Urban Centre belongs to.

An important measure of accessibility and remoteness of an Urban Centre is expressed via the *travel time to country capital*. This assessment represents the travel distance to reach the country capital Urban Centre from each Urban Centre considered. In other words, it represents the distance from each person (living in the Urban Centre) to the central administration of the country.

4.1.5 Environment

Urban centres comprise less than 1% of the Earth's surface, but there is an extraordinary concentration of population, industry and energy use, leading often to massive local pollution and environmental degradation. Urban environmental problems are mostly related to pollution of soil, water and air through traffic, industrial production, and inadequate wastewater/solid waste management. It leads to loss of green and natural spaces, and urban sprawl. All these problems are particularly serious in developing countries and countries with economic transition, where there is often a conflict between the short-term economic plan and the protection of the environment. Cities consume much of the world's energy and account significantly to global CO_2 emissions. In particular, in developing countries, cities are faced with the worst urban air pollution in the world, which occurs as a result of rapid industrialization and increased motorized traffic.

Many of the above-mentioned environmental issues are only collected at the local (city) level and are not available in a globally harmonized manner. The Urban Centre Database 2015 tries to overcome these limitations by providing information about the environmental status of the Urban Centres in the world in terms of the:

— **Urban green;**	— **PM2.5** ○ **concentration and** ○ **emission;** — **C02** ○ **emission**

These variables are linked to the SDG's (indicator 11.7.1 for the urban green[10] and 11.6.2 for the air pollution[11]).

The presence of green spaces within Urban Centres has been recognized as an essential component of the urban environment (Lee et al., 2015) . Green spaces in cities are mostly composed of semi-natural vegetation cover, e.g. street trees, lawns, parks, gardens, forests, green roofs (Gan et al., 2014). Improving availability of green spaces in cities is considered in the United Nations Sustainable Development Goals (SDGs), specifically in target 11.7, which aims to achieve the following: *"By 2030, provide universal access to safe, inclusive and accessible, green and public spaces, in particular for women and children, older persons and persons with disabilities"* (United Nations, 2015).

Fine Particulate matter ($PM_{2.5}$) is of natural (i.e. sand and dust) or of anthropogenic source (i.e. combustion residuals), and its concentration is of high concern especially in urban agglomerations that concentrate many people and develop fast (Chan and Yao 2008). *"Fine particulate matter ($PM_{2.5}$) is responsible for significant negative impacts on human health[12]"* (Directive 2008/50/EC of The European Parliament and Of The Council). $PM_{2.5}$ is the air pollutant that poses the greatest risk to health globally, affecting more people than any other pollutant and chronic exposure to it considerably increases the risk of respiratory and cardiovascular diseases in particular (WHO, 2018).

Carbon dioxide (CO_2) is the primary greenhouse gas contributing to global warming. It is naturally present in the atmosphere as part of the Earth's carbon cycle. However, human activities are adding more CO_2 to the atmosphere. In addition, human activities influence the ability of natural sinks, like forests and soils, to remove CO_2 from the atmosphere. While CO_2 emissions come from a variety of natural sources, human-related emissions are responsible for the increase that has occurred in the atmosphere since the industrial revolution. The industrial activities have raised atmospheric carbon dioxide levels from 280 parts per million to 400 parts per million in the last 150 years (IPCC, 2014).

4.1.6 Disaster risk reduction

Natural hazardous events – those that release high energy or that impact human nutrition and health – are part of the Earth system processes of Planet Earth. Atmospheric circulation may generate strong winds and hurricanes and associated storm surges in coastal area. High intensity precipitation may cause flash floods and inundation. Plate tectonics that continuously shape the topography of Planet Earth generate volcanic eruptions and earthquakes; and tsunamis originate when seismic shaking occurs in the proximity of low lying coastal areas. Droughts the impact life supporting system of humanity and in heat waves impacts health condition directly.

When high-energy events are released over or in the proximity of populated areas, disaster may unfolds. The only option we have is to try to estimate the frequency and intensity with which these events may affect a given place so that we may have the option to mitigate, avoid – when possible by re-locating – or prepare for. Climate change

[10] https://unstats.un.org/sdgs/metadata/files/Metadata-11-07-01.pdf
[11] https://unstats.un.org/sdgs/metadata/files/Metadata-11-06-02.pdf
[12] PM2,5' shall mean particulate matter which passes through a size-selective inlet as defined in the reference method for the sampling and measurement of PM2,5, EN 14907, with a 50 % efficiency cut-off at 2,5 μm aerodynamic diameter;

may have also changed temperature in many cities generating hazardous living conditions.

Cities are most at risk because they concentrate high population and infrastructural assets as built-up areas. In fact, due to natural population growth and urbanization, many cities are increasing in size and density. Risk of damage to hazardous events can be reduced, but not eliminated. The pre-condition to reduce risk and create resilient city to natural hazards is to quantify the exposure to hazards.

In this database, four dimensions of disaster risk reduction are gathered, namely:

— **Flood exposure**
— **Storm surge exposure**
— **Earthquake hazard estimate**
— **Heatwave index.**

Riverine Floods – hereafter referred as *floods* - affect Urban Centres across the globe. In fact, human settlements are often located in the proximity of rivers and in the flat fertile low-lying terrain that is the preferred geographical areas for humans to live in. Flooding is the most recurrent and damaging disaster type and most of the countries of the world have experienced damaging floods and will have to respond to floods in the future also due to the increasing population and infrastructure in flood prone areas.

The low-lying costal areas can be flooded by *storm surges* generated by the air pressure of tropical storms and hurricanes on the ocean surfaces. This natural phenomena can affect a number of cities located in low-lying coastlines.

Earthquake is a major devastating hazard that strikes unexpectedly and damages Urban Centres in seismic prone areas of the world. The estimation of exposure in Urban Centres is related to seismic shaking maps and information layers. Those seismic maps – based on probabilistic assessment of the frequency and magnitude of the shaking intensity - continue to be improved with the accumulation of data and knowledge

One of the most severe effects of global warming is the increase in the frequency and intensity of extreme events such as *heatwaves* (Seneviratne et al., 2016). The severe, extreme and exceptional heatwaves that occurred over the Balkans in 2007, France in 2003 or Russia in 2010 are associated with increased mortality and reduced labour productivity (Dosio et al., 2018). Russo et al. (2015) proposed an index, namely the Heatwave Magnitude Index (HWMId) to take into account both heatwave duration and intensity. The Heatwave Magnitude Index was successfully applied to classify observed heatwaves that occurred globally in the period 1980-2010 (Zampieri et al., 2016).

4.1.7 SDG

With the unanimous adoption of the United Nations (UN) General Assembly resolution 70/1 "Transforming our World: the 2030 Agenda for Sustainable Development" Member States agreed upon a framework of 17 Sustainable Development Goals (SDG) to guide societal development. The action plan, building on the experience of the Millennium Development Goals intertwines aspirational goals with an ambitious monitoring framework composed of 169 targets to monitor progress made in meeting the SDGs. However, the capacity to monitor such progress is entangled by the lack of data and statistical capacity to support the monitoring framework (UN Statistical Commission, 2017). Alternative and innovative sources of data, especially derived Earth Observation (EO) offer significant information, and especially data to support the SDG reporting (Anderson et al., 2017, Paganini & Petiteville, 2018, United Nations, 2015).

According to the United Nations Department of Economic and Social Affairs, the human society is predominantly urban as more than half of global population lives in cities (UNDESA, 2008). The relevance of urban areas is also recognized in the 2030 Development Agenda, which devoted to urban areas a specific Goal, SDG 11 that aspires to "Make cities and human settlements inclusive, safe, resilient and sustainable". Many SDG 11 indicators require fine scale local data that are to be sourced locally, making it more difficult to reach adequate data availability – especially in countries in transitions and data-poor territories. Against this condition, remote sensing and EO are capable to collect information, at a large scale, at high degree of spatial resolution, repeatedly over time, and over wide geographical areas serving multiple applications (Donaldson & Storeygard, 2016, Zell et al., 2012), especially in the SDG framework (Paganini & Petiteville, 2018, GEO, 2017, Noort, 2017), or for generic urban development indicators (Chrysoulakis et al., 2014).

The 232 individual indicators of the 2030 Development Agenda monitoring framework require local yet globally consistent, multi-temporal data. The GHSL maps human settlements and produces fine scale built-up areas and population density grids.

The database integrates core GHSL information on built-up areas and population over time with additional information to characterise Urban Centres. This database can be applied in support the SDG framework. In the Urban Centre database there are a specific attributes about:

— **Land Use Efficiency** Indicator SDG 11.3.1

— Two proxy indicators for SDG 11.7.1:

- **Share of open spaces**
- **Share of Urban Centre population living in areas with high presence of green**

SDG 11.3.1 is classified by the Inter-Agency Expert Group on SDG Indicators as a Tier II indicator (meaning an indicator is conceptually clear and with a methodology for its monitoring, but for which data are not regularly produced or available). SDG 11.7.1 is instead classified as Tier III indicator, meaning "No internationally established methodology or standards are yet available for the indicator, but methodology/standards are being (or will be) developed or tested"[13].

For the estimation of the Land Use Efficiency indicator, we adopted the extent of built-up areas as the input data for land consumption, and population as input for demographic change, and applied the internationally agreed methodology[14]. To complement UN-Habitat reporting on SDG 11 at the High Level Political Forum in 2018 (United Nations, 2018), this database offers a ready-to-use indicator for the circa 10,000 Urban Centres in this open data dataset.

SDG 11.7.1 does not have an established monitoring framework, yet with the Urban Centre data, it is possible to propose a characterisation of Urban Centres based on the presence of greenness (NDVI) and of open spaces. Section 4.2.7 proposes methods to calculate two proxy indicators to estimate *"Average share of the built-up area of cities that is open space"*[15]. With remote sensing it is however not possible to fully align with the aspirations of the indicator formulation, that adds a disaggregation *"for public use for all, by sex, age and persons with disabilities"*. Therefore, the proposed estimates may be regarded as proxies to support the ongoing discussion for a viable method to monitor this indicator.

Green spaces, that may be approximated by the presence of greenness, have many functions that can moderate the climate change impact and help prevent diseases and thus alleviate public health expenses in a context of aging societies (Ngom et al., 2016). The World Health Organization (WHO) suggests that green spaces with a minimum size of one ha and a maximum distance of 300 m to people's residence should be used as threshold values for accessibility (Annerstedt van den Bosch et al., 2016). The European Environment Agency (EEA, 1995) recommends that people should have access to green spaces within 15 min. walking distance (that is 1.61 km considering an average walking

[13] https://unstats.un.org/sdgs/iaeg-sdgs/tier-classification/
[14] https://unstats.un.org/sdgs/metadata/files/Metadata-11-03-01.pdf
[15] https://unstats.un.org/sdgs/metadata/files/Metadata-11-07-01.pdf

speed). Despite this average, access to urban green spaces is not sufficient, because the spatial distribution may result in a significant bias towards certain locations and hence social groups (Le Texier et al., 2018). Besides, many researchers now argue on the appropriate walking distance to consider for a global scale analysis. Until today, there is still no methodological consensus about how to conceptualize access and measure the provision of urban green spaces.

Open spaces can be of different nature and may include beaches, parks, playing fields and also green spaces. The landscape of urban open spaces can range from playing fields to highly maintained environments to relatively natural landscapes. One definition holds that, *"As the counterpart of development, urban open space is a natural and cultural resource, synonymous with neither 'unused land' nor 'park and recreation areas." Another is "Open space is land and/or water area with its surface open to the sky, consciously acquired or publicly regulated to serve conservation and urban shaping function in addition to providing recreational opportunities" (Myers, 1975).* In almost all instances, the space referred to by the term is, in fact, **green space**. However, there are examples of open space which, though not green (e.g. beaches) that are still considered as open spaces.

4.2 Attributes

The dimensions of the urban space are materialised through variables, some of them multitemporal. This section provides information on the data sources and methodological details.

4.2.1 General characteristics

The *control codes* include *unique identifier* of the Urban Centre, and the *quality control code*. The Urban Centre identifier allows to identify uniquely n Urban Centre across the related resources.

4.2.1.1 Quality control

The quality control code is the result of the quality check procedure performed over the full dataset of Urban Centres, which relays on visual assessment. The assessment was performed in two steps: automatic and manual. The automatic quality exploits the results from the visual assessment performed over the Urban Centres dataset derived from the SMOD data released in 2016. Each Urban Centre was visually validated using a VHR map (Google Maps[16] or Bing Maps[17]), and assessed as true positive if within the polygon of the Urban Centre a high density settlement is present (rule: at least 0,5 km^2 of very densely built-up area). As a result, a positive spatial domain and a negative spatial domain were identified. These two datasets were crossed with the updated Urban Centre dataset described in this document (derived from the recent SMOD, planned for release in 2019). The Urban Centres that overlap with the positive domain were marked as the true positive samples. The remaining Urban Centres (the ones that overlap with the negative domain, or both, negative and positive domains) were assessed visually by group of experts (at least three evaluations per sample) using the same rules as in 2016. Additionally, all Urban Centres of one kilometre square are defined as uncertain. The assessment results are as follows:

- True positives: there is a high density settlement present;
- False positive: there is no presence of a high density settlement;
- Uncertainty: the expert was not sure or there was disagreement between experts.

[16] https://www.google.com/maps
[17] https://www.bing.com/maps

The attributes related to *control codes* are:

- **ID_HDC_G0**: unique ID of the Urban Centre
- **QA2_1V**: quality code (0 – false positive, 1 – true positive, >1 uncertain).

4.2.1.2 Extension

The *Urban Centre Extension* is a bounding box of the Urban Centre polygon, derived using Python (GeoPandas 0.4.0)[18]. The following attributes declares the *Urban Centre Extension*:

- **BBX_LATMN –** latitude of the low left corner of the bounding box;
- **BBX_LONMN -** longitude of the low left corner of the bounding box;
- **BBX_LATMX -** latitude of the top right corner of the bounding box;
- **BBX_LONMX -** longitude of the top right corner of the bounding box.

4.2.1.3 Location

The *Location* of the Urban Centre is described by centroid, country or countries identification, and geographical region where the Urban Centre is located. The geometric centroid of the Urban Centre shape is derived using Python (GeoPandas 0.4.0)[19]. In case the centroid is outside the Urban Centre polygon, the "Feature to Point" tool of ArcGIS 10.2 is used (with "inside" option activated)[19]. The Country is derived by overlaying the Urban Centre polygon with the GADM v2.8 (GADM, 2018), which is the base to provide the names and ISO 3166-1 alpha-3 (ISO 3) codes of the countries. There are Urban Centres that are cross border, i.e., they extends over more than one country.

The attributes related to *Location* are:

- **GCPNT_LAT:** Latitude of the geometric centroid;
- **GCPNT_LON:** Longitude of the geometric centroid;
- **XBRDR**: cross border (the value 1 indicates if it is a cross border Urban Centre, while value 0 is assigned to Urban Centres which whole polygon lays within borders of only one country);
- **XCTR_NBR**: number of the countries with which the Urban Centre polygon crosses;
- **XC_ISO_LST**: list of ISO-3 codes of the countries with which the Urban Centre polygon crosses (separation char: ';');
- **CTR_MN_ISO**: the ISO 3 code of the main country, i.e., the country within which borders the majority of the area of the Urban Centre is located;
- **CTR_MN_NM**: the name of the main country, i.e., the country within which borders the majority of the area of the Urban Centre is located;
- **GRGN_L1**: Major Geographical Region (UNDESA, 2018b), according to the classification of the main country;
- **GRGN_L2**: Geographical Region (UNDESA, 2018b), according to the classification of the main country.

4.2.1.4 Name

The name(s) assigned to the Urban Centres are calculated using available open source placename databases. These names shall be used only for visualisation purpose and shall

[18] http://geopandas.org/
[19] http://pro.arcgis.com/en/pro-app/tool-reference/data-management/feature-to-point.htm

not be understand as an official position of EC in any case. The following placename databases are considered:

- WUP300k: World's Cities in 2018 form the World Urban Prospect 2018 (WUP 2018) gathers the following attributes of a city/agglomeration: a name, a centroid, and multitemporal information on the population (UNDESA, 2018a). Of the 1,860 cities with at least 300,000 inhabitants in 2018 included in this dataset, 55% follow the "urban agglomeration" statistical concept, 35% follow the "city proper" concept and the remaining 10% refer to "metropolitan areas";
- GRUMP-SP: Global Rural-Urban Mapping Project (GRUMP), v1, Settlement Points, v1 (1990, 1995, 2000) (CIESIN et al., 2011, Balk et al., 2006), that gathers the named city points, and population counts if available;
- NE: Natural Earth Populated Places, v4.1.0 (Patterson and Kelso, 2018);
- GN: GeoNames (GeoNames, 2018), a geographical database covers all countries and contains over eleven million placenames that are available for download free of charge. Currently, it is the most compressive dataset of named places available as open and free.

First, the input sources are pre-processed (e.g., names are converted to ASCI format), and the country codes are harmonized to the ISO 3 codes used by Urban Centre database. Per each Urban Centre polygon, the naming algorithm uses a decision tree that scans the input databases to identify the point(s) located within the polygon (with permitted spatial uncertainty: 1 km). A trust value is assigned to each input datasets in the following order (1 – the highest value of trust): WUP300K (1), GRUMP-SP (2), NE (3), GN (4). The name scanner scans the input sources in the order of the trust value, and if it finds at least one place name, the scanning procedure stops. The resulted name list is ordered by population value as reported by input source, and the first name becomes the main name. Therefore, the result is strongly influenced by the quality of the input sources, i.e., name and the declared population counts. A cross border Urban Centre is treated as a special case. The algorithm tries to find at least one name per each country from the list of countries (XC_ISO_LST). The main name is a composed name: main placename per each country. Also in this case, the assigned population counts decide on the order of the names. As a result, there are more than 1,400 Urban Centres for which the algorithm is not able to assign a name (considering all countries listed in the country list). Finally, for a subset of more than 300 Urban Centres (with valid value in the quality attribute), the names are assigned manually by visual inspection using existing online services (OpenStreetMap[20], Google Maps or Bing Maps).

The attributes related to *naming* are:

- **UC_NM_MN**: the main name of the Urban Centre (the country ISO 3 is declared within '[]', to support the cross border entities);
- **UC_NM_LST**: full list of assigned names of the Urban Centre (the country ISO 3 is declared within '[]', to support the cross border entities);
- **UC_NM_SRC**: source of the list of names per each country (WUP, GRUMP, NE, GN, WM – online mapping services, OTHER – web user feedback and other manual revisions).

4.2.2 Mutlitemporal Urban Centre Domain

Multitemporal Urban Centre definitions gathers results from the multitemporal settlement model, i.e., application of the global definition per epoch 1975, 1990 and 2000. The baseline datasets (GHS-POP and GHS-BUILT grids) belongs to the Community Release, 2018 (Florczyk et al., 2018). Per each Urban Centre (epoch 2015), the following

[20] https://www.openstreetmap.org/

attributes reports number of different Urban Centres (and their total sum or area) within the polygon of the Urban Centre 2015:

- **H00_NBR:** number of separate Urban Centres delineated in the 2000 falling within the polygon of the Urban Centre 2015;
- **H00_AREA**: overall area of Urban Centres delineated in the 2000 falling within the polygon of the Urban Centre 2015;
- **H90_NBR:** number of separate Urban Centres delineated in the 1990 falling within the polygon of the Urban Centre 2015;
- **H90_AREA**: overall area of Urban Centres delineated in the 1990 falling within the polygon of the Urban Centre 2015;
- **H75_NBR:** number of separate Urban Centres delineated in the 1975 falling within the polygon of the Urban Centre 2015;
- **H75_AREA**: overall area of Urban Centres delineated in the 1975 falling within the polygon of the Urban Centre 2015;

4.2.3 Geography

Urban centres are characterised by several physiographic attributes.

4.2.3.1 Biome

One of the attributes is *biome* derived from the Terrestrial Ecoregions of the World (TEOW) (Olson et al., 2001), a biogeographic regionalization of the Earth's terrestrial biodiversity. The biogeographic units are ecoregions defined as relatively large units of land or water. An ecoregion contains a distinct assemblage of natural communities sharing a large majority of species, dynamics, and environmental conditions. TEOW delineates 867 ecoregions grouped into 14 biomes and eight biogeographic realms, and is released as a vector dataset. The attribute is produced using a spatial analysis (i.e., geometric intersection), and as the result, one or more biomes are associated with each Urban Centre (i.e., all biomes intersecting the Urban Centre polygon are listed in the order of the decreasing intersecting area).

- **E_BM_NM_LST**: semi-colon separated list of names of biome classes, intersecting with the spatial domain of the Urban Centre;

4.2.3.2 Soil group

The s*oil group* identifies the soil main group on which the urban expansion is developing. This class is derived from the Harmonized World Soil Database v1.2 (Fischer et al., 2008), a 30 arc-sec raster database (approx. 1 km at the equator), with over 15 000 different soil mapping units that combines existing regional and national updates of soil information worldwide. This dataset gathers soil properties, in terms of soil units and the characterization of selected soil parameters (organic Carbon, pH, water storage capacity, soil depth, cation exchange capacity of the soil and the clay fraction, total exchangeable nutrients, lime and gypsum contents, sodium exchange percentage, salinity, textural class and granulometry). Each Urban Centre has one or more soil groups associated via a spatial analysis. First, the global grid is aligned with the World Mollweide grid at 1 km^2, and then per each Urban Centre the corresponding classes are extracted (in the order of the decreasing area).

- **E_SL_LST**: semi-colon separated list of names of soil groups, intersecting with the spatial domain of the Urban Centre;

4.2.3.3 Elevation

The baseline data used to derive the *elevation* attribute in the database is the ALOS World 3D - 30m (AW3D30), a dataset produced by JAXA Earth Observation Research

Center with a horizontal resolution of 1 arc-sec raster (approx. 30 m at the equator) (Tadono et al., 2014, EORC & JAXA, 2017). This dataset has been generated from the DSM dataset (5 m) of the precise global digital 3D map "ALOS World 3D" (AW3D), produced through the use of 3 million scene archives acquired by the PRISM panchromatic stereo mapping sensor on the Advanced Land Observing Satellite "DAICHI" (ALOS) operated from 2006 to 2011. The *elevation* attribute is computed as the average altitude within the spatial extent of each Urban Centre, expressed in metres above sea level (MASL). The attributes are estimated using the Google Earth Engine (GEE) platform (Gorelick et al., 2011) and the ready-to-use JAXA_ALOS_AW3D30_V1_1[21] dataset.

— **EL_AV_ALS**: the average elevation estimated within the spatial domain of the Urban Centre, and expressed in metres above sea level (MASL);

4.2.3.4 Climate classification

Köppen climate classification (later updated by Geiger) is a vegetation-based, empirical climate classification system. Its aim was to define rule-based climatic boundaries, which correspond to those of the vegetation zones (biomes). The Köppen-Geiger world map on climate classification based on datasets from the Climatic Research Unit (CRU) of the University of East Anglia and the Global Precipitation Climatology Centre (GPCC) at the German Weather Service is valid for the second half of the 20 century (Kottek et al., 2016). The *climate typology* attribute gathered in the Urban Centre database is derived from the updated Köppen-Geiger climate classification map representative for the more recent 25-year period 1986-2010 (Rubel et al., 2017), and available as a 5 arc-sec raster (approx. 150 m at the equator). In the GHS-UCDB, each Urban Centre has one or more climate classes associated via a spatial analysis. First, the global grid is aligned with the World Mollweide grid at 1 km^2, and then per each Urban Centre the corresponding classes are extracted (in the order of the decreasing area).

— **E_KG_NM_LST**: semi-colon separated list of names of Köppen-Geiger climate classes, intersecting with the spatial domain of the Urban Centre;

4.2.3.5 Temperature and precipitation

The CRU TS v. 4.02 gridded time-series dataset (http://www.cru.uea.ac.uk/data) is used to derive the attributes on *temperature* and *precipitation* (Harris et al., 2014). These data are based on observations from ground stations combined with interpolation at a coarse resolution aiming for global coverage. Therefore, they do not (yet) consider localized city effects such as urban heat island that may modify intra-city observed and perceived temperatures. *Average temperatures* per Urban Centre were calculated for three time intervals centred on the years 1990, 2000 and 2015 as follows: for 1990, the interval spans from 1988 to 1991; for 2000, the interval spans from 1999 to 2002; for 2015, the interval spans from 2012 to 2015. The intervals were chosen in order to match the dates of the Landsat data collections used to derive the multi-temporal built-up areas (GHS-BUILT) and to reduce inter-annual and seasonal variability that may affect the change analysis. Similarly, the *average precipitations* were calculated for the Urban Centres in the period 1990-2015.

— **E_WR_T_90**: average temperature calculated from annual average estimates for time interval centred on the year 1990 (the interval spans from 1988 to 1991) within the spatial domain of the Urban Centre, and expressed in Celsius degrees (°C);

— **E_WR_T_00**: average temperature calculated from annual average estimates for time interval centred on the year 2000, (the interval spans from 1999 to 2002) within the spatial domain of the Urban Centre, and expressed in Celsius degrees (°C);

[21] https://developers.google.com/earth-engine/datasets/catalog/JAXA_ALOS_AW3D30_V1_1

— **E_WR_T_14**: average temperature calculated from annual average estimates for time interval centred on the year 2015 (the interval spans from 2012 to 2015) within the spatial domain of the Urban Centre, and expressed in Celsius degrees (°C);

— **E_WR_P_90**: average precipitations calculated from annual average estimates for time interval centred on the year 1990 (the interval spans from 1988 to 1991) within the spatial domain of the Urban Centre; and expressed in millimetres (mm), the amount of rain per square meter in one hour);

— **E_WR_P_00**: average precipitations calculated from annual average estimates for time interval centred on the year 2000 (the interval spans from 1999 to 2002) within the spatial domain of the Urban Centre; and expressed in millimetres (mm), the amount of rain per square meter in one hour);

— **E_WR_P_14**: average precipitations calculated from annual average estimates for time interval centred on the year 2015 (the interval spans from 2012 to 2015) within the spatial domain of the Urban Centre; and expressed in millimetres (mm), the amount of rain per square meter in one hour).

4.2.3.6 Major river basin

The global dataset of Major River Basins (MRB) used in this study consists in 405 basins across the globe, and is available in a vector format (GRDC, 2007). Based on the concept that a river basin covers all area of land that drains to the point of the lowest elevation, the delineation procedure used to produce MRB starts at the rivers outlet (pour point), which coincides with the location of the confluence with a river, the mouth into an ocean or an endorheic sink. These "pour points" represent the locations above which the drainage basin is derived from the flow direction grid. The flow direction data set used is the HYDRO1k Elevation Derivative Database[22] (Danielson, 1996). The names of the polygons were assigned using several world atlases as references. Each Urban Centre has one or more major river basins associated via spatial analysis of the vector datasets. First, the MRB dataset is projected to the World Mollweide, and then per each Urban Centre the intersecting basins are extracted (in the order of the decreasing area).

— **E_RB_NM_LST**: semi-colon separated list of major river basins names, intersecting with the spatial domain of the Urban Centre;

4.2.4 Socio-economy

The resident population and the built-up surface data are derived from the GHS-POP and GHS-BUILT products at the different epochs (1975, 1990, 2000, and 2015).

4.2.4.1 Built-up areas

The GHS-BUILT is generated by spatial aggregation (upscaling) of the information collected at various decametric spatial resolution satellite image data records (10-15-30-80 metres) available in the different GHSL epochs and different satellite platforms (Pesaresi et al., 2016, Corbane et al., 2017, Corbane et al., 2018a), to the 250x250 metres and 1x1 kilometer resolution grids, aggregated separately per each epoch.

— **B15** – total built-up area in 2015 calculated within the spatial domain of the Urban Centre of 2015, expressed in square kilometres;
— **B00** – total built-up area in 2000 calculated within the spatial domain of the Urban Centre of 2015, expressed in square kilometres;
— **B90** – total built-up area in 1990 calculated within the spatial domain of the Urban Centre of 2015, expressed in square kilometres;

[22] except to Australia, for which the following dataset is used: Australia's River Basins 1997 for Australia http://www.ga.gov.au/nmd/products/thematic/basins.htm

— **B75** – total built-up area in 1975 calculated within the spatial domain of the Urban Centre of 2015, expressed in square kilometres;

4.2.4.2 Population

The GHS-POP is generated by spatial disaggregation (downscaling) of census spatial data to 250x250 metres resolution grid, using GHS-POP as principal spatial covariate (Freire et al., 2016). Per each Urban Centre a set of multi-temporal attributes is calculated by intersecting the Urban Centre polygon with the multitemporal built-up area and population grids (Florczyk et al., 2018) at 1 km resolution.

— **P15** – total population in 2015 calculated within the spatial domain of the Urban Centre of 2015, expressed in number of people;
— **P00** – total population in 2000 calculated within the spatial domain of the Urban Centre of 2015, expressed in number of people;
— **P90** – total population in 1990 calculated within the spatial domain of the Urban Centre of 2015, expressed in number of people;
— **P75** – total population in 1975 calculated within the spatial domain of the Urban Centre of 2015, expressed in number of people;

4.2.4.3 Built-up areas per capita

By combination of the attributes reporting the total population, built-up area, the built-up areas per capita are calculated:

— **BUCAP15** – amount of the built-up area per person in 2015 calculated within the spatial domain of the Urban Centre of 2015, expressed in square meters per person;
— **BUCAP00** – amount of the built-up area per person in 2000 calculated within the spatial domain of the Urban Centre of 2015, expressed in square meters per person;
— **BUCAP90** – amount of the built-up area per person in 1990 calculated within the spatial domain of the Urban Centre of 2015, expressed in square meters per person;
— **BUCAP75** – amount of the built-up area per person in 1975 calculated within the spatial domain of the Urban Centre of 2015, expressed in square meters per person;

4.2.4.4 Night time light

The Night Light Emission data integrated in the current GHS-UCDB are the *Version 1 VIIRS (Visible Infrared Imaging Radiometer Suite) DNB (Day/Night Band) Night-time Lights Composites suite*[23] produced by the Earth Observations Group at NOAA/NCEI. These grids span the globe from 75N latitude to 65S and have a resolution of 15 arc-sec (approx. 500 m at the equator). The attribute is derived from the "vcm-orm-ntl" (VIIRS Cloud Mask - Outlier Removed - Night-time Lights) layer of reference year 2015, showing the cloud-free average radiance emitted, expressed as nano-watt per steradian per square centimetre (nW cm^{-2} sr^{-1}) with outlier removal process to filter out fires and other ephemeral lights. First, the layer is projected to the 1 km^2 resolution grid by warping the layer to Mollweide projection through oversampling (50 m) with nearest neighbour and aggregating to 1 km (with mean function). Finally, zonal statistics within Urban Centres polygons are calculated. The attribute released in the database is:

— **NTL_AV** – mean night time light emission calculated within the Urban Centre spatial domain, expressed in nano-watt per steradian per square centimetre.

[23] https://www.ngdc.noaa.gov/eog/viirs/download_dnb_composites.html#NTL_2015

4.2.4.5 *Income class and development group*

The income class (IC) and development group (DEV) used in the GHS-UCDB follows the World Urbanization Prospects 2018 (WUP 2018) (UNDESA, 2018b), and therefore are derived from the country-level classifications. The Urban Centres have been associated to a single Country (attribute *CTR_MN_ISO*) by a spatial join. In case of cross-country UCs, the Country showing the majority of resident people in the Urban Centre was selected for the join in the final database. As a result, the IC and DEV values are assigned according the country classification in the WUP 2018.

The income class schema and development group schema are structured in four and three income groups, respectively (Table 2), and the *basic socio-economic classification* attributes are as follow:

- **INCM_CMI**: UN income class;
- **DEV_CMI**: UN development class.

Table 2. The income class and development group schemas.

Income classes			Development Groups	
Name	**Acronym**		**Name**	**Acronym**
High Income Countries	HIC		More Developed Regions	MDR
Upper-middle Income Countries	UMIC		Less developed regions, excluding least developed countries	LCD
Lower-middle Income Countries	LMIC		Least developed countries	LDCL
Low Income Countries	LIC			

4.2.4.6 *Gross domestic product*

The GDP estimates reported in the database are calculated using the global data on total annual GDP (PPP), available at 30 arc-sec resolution (approx. 60 km at the equator) for three epochs: 1990, 2000 and 2015 (Kummu et al., 2018). These global grids were harmonised with the Urban Centre grid (i.e., projected with resampling to 1x1 kilometre grid in World Mollweide projection). Per each Urban Centre several statistics were calculated from the GDP values within the Urban Centre spatial domain. The resulted attributes are:

- **GDP90_SM**: sum of the GDP PPP values for year 1990 within the Urban Centre 2015, expressed in US dollars (2007);
- **GDP00_SM**: sum of the GDP PPP values for year 2000 within the Urban Centre 2015, expressed in US dollars (2007);
- **GDP15_SM**: sum of the GDP PPP values for year 2015 within the Urban Centre 2015, expressed in US dollars (2007).

4.2.4.7 *Travel time to country capital*

The global dataset Global Friction Surface 2015, v1.0, available as 30-arc sec resolution grid, was produced through a collaboration between the University of Oxford Malaria Atlas Project (MAP)[24], Google, the European Union Joint Research Centre (JRC), and the

[24] Downloaded from https://map.ox.ac.uk/wp-content/uploads/accessibility/friction_surface_2015_v1.0.zip

University of Twente, Netherlands (Weiss et al., 2018). Many different datasets were used to produce this product, including roads (OpenStreetMap and Google roads datasets), railways, rivers, lakes, oceans, topographic conditions (slope and elevation), landcover types, and national borders. This global friction surface enumerates land-based travel speed for all land pixels between 85 degrees north and 60 degrees south for a nominal year 2015. This grid is used to estimate the travel time to country capital, which represents the travel distance to reach the country capital Urban Centre from each Urban Centre considered (flight connections not taken into account). Per each country, the Urban Centre of the capital city is identified, and the distances between the capital and the other Urban Centres of the country are calculated. If the country capital is not an Urban Centre, the travel time is unknown; therefore, the NoData value (i.e., 'NAN') is assigned to Urban Centres of given country. The distance is estimated between borders of the Urban Centres. The resulted attribute is:

— **TT2CC**: travel time to country capital, expressed in minutes.

4.2.5 Environment

4.2.5.1 Greenness

The data on green spaces (also called greenness) within Urban Centres has been produced by analysing Landsat annual Top-of-Atmosphere (TOA) reflectance composites available as collections in the Google Earth Engine (GEE) platform for the period 1990-2015. These composites are created by considering the highest value of the Normalized Difference Vegetation Index (NDVI) as the composite value (i.e. greenest pixel).

Changes in the *amount of greenness* within cities are assessed for the reference periods (centred on *1990, 2000* and *2015*), and estimated within the respective built-up areas only. For each of the 10,323 Urban Centres, the average of all greenest pixels located within the built-up area of the Urban Centre is calculated for three time intervals centred on the years 1990, 2000 and 2015 as follows:

for 1990, the interval spans from 1988-01-01 to 1991-12-30;
for 2000, the interval spans from 1999-01-01 to 2002-12-30;
for 2015, the interval spans from 2012-01-01 to 2015-12-30.

These intervals were chosen in order to match the dates of the Landsat data collections used to derive the multi-temporal built-up areas (GHS-BUILT) and to mitigate inter-annual variability and seasonal anomalies that may affect the greenness change analysis. The detailed methodology on multitemporal assessment of greenness with the built-up areas is described in Corbane et al., 2018b.

— **E_GR_AV90**: average greenness estimated for 1990 located in the built-up area of epoch 1990, and calculated within the spatial domain of the Urban Centre of 2015. The values are expressed in unit less measures in range -1:1;

— **E_GR_AV00**: average greenness estimated for 2000 located in the built-up area of epoch 2000, and calculated within the spatial domain of the Urban Centre of 2015. The values are expressed in unit less measures in range -1:1;

— **E_GR_AV14**: average greenness estimated for 2014 located in the built-up area of epoch 2014, and calculated within the spatial domain of the Urban Centre of 2015. The values are expressed in unit less measures in range -1:1;

4.2.5.2 Greenness extent

The areas of three levels of greenness are estimated for each of the reference epoch within the boundaries delimited by the Urban Centres. The continuous greenness values per each epoch are classified into one of the three classes as follows:

Low green for greenness < 0.1 : corresponding to barren rock, sand, snow or impervious surfaces (e.g. built-up areas)

Medium green for 0.2 <Greenness <0.5: corresponding to shrubs or agriculture
High green for 0.6 <Greenness < 0.9: corresponding to dense vegetation (e.g. forest, gardens, etc.).

The areas of each of the three classes are estimated for each epoch within the boundaries delimited by the Urban Centres. The resulted attributes are as follows:

— **E_GR_AH90**: total area of the high green (corresponding to dense vegetation) estimated for 1990, located within the area of the Urban Centre of 2015, and expressed in square kilometres;

— **E_GR_AM90**: total area of the medium green (corresponding to shrubs or agriculture) estimated for 1990, located the area of the Urban Centre of 2015, and expressed in square kilometres;

— **E_GR_AL90**: total area of the low green (corresponding to barren rock, sand, snow or impervious surfaces) estimated for 1990, located the area of the Urban Centre of 2015, and expressed in square kilometres;

— **E_GR_AT90**: total area of greenness estimated for 1990, located the area of the Urban Centre of 2015, and expressed in square kilometres;

— **E_GR_AH00**: total area of the high green (corresponding to dense vegetation) estimated for 2000, located within the area of the Urban Centre of 2015, and expressed in square kilometres;

— **E_GR_AM00**: total area of the medium green (corresponding to shrubs or agriculture) estimated for 2000, located the area of the Urban Centre of 2015, and expressed in square kilometres;

— **E_GR_AL00**: total area of the low green (corresponding to barren rock, sand, snow or impervious surfaces) estimated for 2000, located the area of the Urban Centre of 2015, and expressed in square kilometres;

— **E_GR_AT00**: total area of greenness estimated for 2000, located the area of the Urban Centre of 2015, and expressed in square kilometres;

— **E_GR_AH14**: total area of the high green (corresponding to dense vegetation) estimated for 2015, located within the area of the Urban Centre of 2015, and expressed in square kilometres;

— **E_GR_AM14**: total area of the medium green (corresponding to shrubs or agriculture) estimated for 2015, located the area of the Urban Centre of 2015, and expressed in square kilometres;

— **E_GR_AL14**: total area of the low green (corresponding to barren rock, sand, snow or impervious surfaces) estimated for 2015, located the area of the Urban Centre of 2015, and expressed in square kilometres;

— **E_GR_AT14**: total area of greenness estimated for 2015, located the area of the Urban Centre of 2015, and expressed in square kilometres.

4.2.5.3 PM2.5 and CO2 emissions

The data on PM2.5 and CO2 emissions are derived from the European Commission's in-house Emissions Database for Global Atmospheric Research (EDGAR v4.3.2), which estimates anthropogenic greenhouse gas and particulate air pollutant emissions for the years 1970 to 2012 (Crippa et al., 2018). The calculation of the emissions includes all human activities, except large scale biomass burning and land use, land-use change, and forestry. The results are comparable between countries thanks to the bottom-up compilation methodology of sector-specific emissions applied consistently for all world countries. The sectors definition uses the following IPCC 1996 codes: *energy* - Power

Industry (IPCC 1A1a), *residential*: Energy for buildings (IPCC 1A4), waste (IPCC 6), *industry*: Oil refineries and Transformation industry (IPCC 1A1b, 1A1c), Combustion for manufacturing (IPCC 1A2), Fuel exploitation (IPCC 1B), Industrial Processes (IPCC 2), Solvents and products use (IPCC 3), *transport* - Transport (IPCC 1A3), and *agriculture*: Agriculture (IPCC 4). Fossil CO2 emissions are included in the latest version of EDGAR, EDGARv5.0, giving an overview of country-by-country fossil CO2 emissions (Muntean et al., 2018). Here, the CO2 emissions are estimated from usage of fossil (i.e., non-short-cycle-organic) and bio (i.e., short-cycle-organic) fuels. The attributes are produced for reference epoch 1975, 1990, 2000 and 2012, and are as follows:

— **E_EC2E_E75**: total emission of CO2 from the energy sector, using non-short-cycle-organic fuels in 1975, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
— **E_EC2E_E90**: total emission of CO2 from the energy sector, using non-short-cycle-organic fuels in 1990, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
— **E_EC2E_E00**: total emission of CO2 from the energy sector, using non-organic fuels in 2000, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
— **E_EC2E_E12**: total emission of CO2 from the energy sector, using non-short-cycle-organic fuels in 2012, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
— **E_EC2E_R75**: total emission of CO2 from the residential sector, using non-short-cycle-organic fuels in 1975, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
— **E_EC2E_R90**: total emission of CO2 from the residential sector, using non-short-cycle-organic fuels in 1990, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
— **E_EC2E_R00**: total emission of CO2 from the residential sector, using non-short-cycle-organic fuels in 2000, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
— **E_EC2E_R12**: total emission of CO2 from the residential sector, using non-short-cycle-organic fuels in 2012, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
— **E_EC2E_I75**: total emission of CO2 from the industry sector, using non-short-cycle-organic fuels in 1975, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
— **E_EC2E_I90**: total emission of CO2 from the industry sector, using non-short-cycle-organic fuels in 1990, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
— **E_EC2E_I00**: total emission of CO2 from the industry sector, using non-short-cycle-organic fuels in 2000, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
— **E_EC2E_I12**: total emission of CO2 from the industry sector, using non-short-cycle-organic fuels in 2012, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
— **E_EC2E_T75**: total emission of CO2 from the transport sector, using non-short-cycle-organic fuels in 1975, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
— **E_EC2E_T90**: total emission of CO2 from the transport sector, using non-short-cycle-organic fuels in 1990, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;

- **E_EC2E_T00**: total emission of CO_2 from the transport sector, using non-short-cycle-organic fuels in 2000, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EC2E_T12**: total emission of CO_2 from the transport sector, using non-short-cycle-organic fuels in 2012, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EC2E_A75**: total emission of CO_2 from the agriculture sector, using non-short-cycle-organic fuels in 1975, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EC2E_A90**: total emission of CO_2 from the agriculture sector, using non-short-cycle-organic fuels in 1990, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EC2E_A00**: total emission of CO_2 from the agriculture sector, using non-short-cycle-organic fuels in 2000, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EC2E_A12**: total emission of CO_2 from the agriculture sector, using non-short-cycle-organic fuels in 2012, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EC2O_E75**: total emission of CO_2 from the energy sector, using short-cycle-organic fuels in 1975, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EC2O_E90**: total emission of CO_2 from the energy sector, using short-cycle-organic fuels in 1990, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EC2O_E00**: total emission of CO_2 from the energy sector, using non-organic fuels in 2000, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EC2O_E12**: total emission of CO_2 from the energy sector, using short-cycle-organic fuels in 2012, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EC2O_R75**: total emission of CO_2 from the residential sector, using short-cycle-organic fuels in 1975, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EC2O_R90**: total emission of CO_2 from the residential sector, using short-cycle-organic fuels in 1990, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EC2O_R00**: total emission of CO_2 from the residential sector, using short-cycle-organic fuels in 2000, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EC2O_R12**: total emission of CO_2 from the residential sector, using short-cycle-organic fuels in 2012, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EC2O_I75**: total emission of CO_2 from the industry sector, using short-cycle-organic fuels in 1975, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EC2O_I90**: total emission of CO_2 from the industry sector, using short-cycle-organic fuels in 1990, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EC2O_I00**: total emission of CO_2 from the industry sector, using short-cycle-organic fuels in 2000, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;

- **E_EC2O_I12**: total emission of CO2 from the industry sector, using short-cycle-organic fuels in 2012, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EC2O_T75**: total emission of CO2 from the transport sector, using short-cycle-organic fuels in 1975, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EC2O_T90**: total emission of CO2 from the transport sector, using short-cycle-organic fuels in 1990, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EC2O_T00**: total emission of CO2 from the transport sector, using short-cycle-organic fuels in 2000, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EC2O_T12**: total emission of CO2 from the transport sector, using short-cycle-organic fuels in 2012, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EC2O_A75**: total emission of CO2 from the agriculture sector, using short-cycle-organic fuels in 1975, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EC2O_A90**: total emission of CO2 from the agriculture sector, using short-cycle-organic fuels in 1990, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EC2O_A00**: total emission of CO2 from the agriculture sector, using short-cycle-organic fuels in 2000, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EC2O_A12**: total emission of CO2 from the agriculture sector, using short-cycle-organic fuels in 2012, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EPM2_E75**: total emission of PM2.5 from the energy sector in 1975, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EPM2_E90**: total emission of PM2.5 from the energy sector in 1990, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EPM2_E00**: total emission of PM2.5 from the energy sector, using non-organic fuels in 2000, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EPM2_E12**: total emission of PM2.5 from the energy sector in 2012, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EPM2_R75**: total emission of PM2.5 from the residential sector in 1975, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EPM2_R90**: total emission of PM2.5 from the residential sector in 1990, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EPM2_R00**: total emission of PM2.5 from the residential sector in 2000, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EPM2_R12**: total emission of PM2.5 from the residential sector in 2012, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EPM2_I75**: total emission of PM2.5 from the industry sector in 1975, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;

- **E_EPM2_I90**: total emission of PM2.5 from the industry sector in 1990, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EPM2_I00**: total emission of PM2.5 from the industry sector in 2000, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EPM2_I12**: total emission of PM2.5 from the industry sector in 2012, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EPM2_T75**: total emission of PM2.5 from the transport sector in 1975, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EPM2_T90**: total emission of PM2.5 from the transport sector in 1990, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EPM2_T00**: total emission of PM2.5 from the transport sector in 2000, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EPM2_T12**: total emission of PM2.5 from the transport sector in 2012, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EPM2_A75**: total emission of PM2.5 from the agriculture sector in 1975, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EPM2_A90**: total emission of PM2.5 from the agriculture sector in 1990, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EPM2_A00**: total emission of PM2.5 from the agriculture sector in 2000, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;
- **E_EPM2_A12**: total emission of PM2.5 from the agriculture sector in 2012, calculated over the Urban Centre spatial domain of 2015, and expressed in tonnes per year;

4.2.5.4 PM2.5 concentration

The estimates of PM2.5 concentration are based on the Global Burden of Disease (GBD) 2017 data on ambient air pollution from 1990 to 2015. The GBD estimates are derived at an approximate 11x11 kilometres (at the equator) resolution by integrating satellite observations (Van Donkelaar, 2018), chemical transport models and measurements from ground monitoring station networks. The accuracy of these exposure estimates varies considerably by location. Inaccuracy is particularly high in areas with few monitoring stations and in areas with very high concentrations. Accuracy is generally good in regions with dense monitoring station networks (such as most advanced economies) (Shaddick et al., 2018). The assessment follows the OECD methodology[25,26], and the resulted attributes are produced for reference epochs 2000-2005-2010-2014 (as an average value from five-year period centred on the reference epoch):

- **E_CPM2_T00**: a total concertation of PM2.5 for reference epoch 2000, calculated over the Urban Centre spatial domain of 2015, and expressed in µg/m^3;
- **E_CPM2_T05**: a total concertation of PM2.5 for reference epoch, calculated over the Urban Centre spatial domain of 2015, and expressed in µg/m^3;

[25] https://www.stateofglobalair.org/data/methods

[26] http://stats.oecd.org/wbos/fileview2.aspx?IDFile=28707511-43bc-4a03-a6bb-b3a32eea4f8b

— **E_CPM2_T10**: a total concertation of PM2.5 for reference epoch 2010, calculated over the Urban Centre spatial domain of 2015, and expressed in μg/m^3;
— **E_CPM2_T14**: a total concertation of PM2.5 for reference epoch 2014, calculated over the Urban Centre spatial domain of 2015, and expressed in μg/m^3.

4.2.6 Disaster Risk Reduction

In this database, four dimensions of disaster risk reduction are gathered, namely *flood exposure, storm surge exposure*, *earthquake hazard estimate*, and *maximum magnitude of the heatwaves*. The produced attributes are results of the crossing of the spatial information of Urban Centres with the hazardous information contained in four global hazards maps. The resulting maps of exposure of Urban Centres to hazard are the best global knowledge for that given hazard at the time of data preparation and integration process that is derived from open source data.

4.2.6.1 Flood

The hazard data for flood exposure are derived from the *Flood hazard map of the World* (Dottori et al., 2016a). These dataset is based on streamflow data from the European and Global Flood Awareness System (EFAS and GloFAS) and has been computed using two-dimensional hydrodynamic models. There are several hazards maps produced with different return period (10, 20, 50, 100, 200 and 500-RP). These maps can be used to assess flood exposure and risk; however, they are not official flood hazard maps. For purpose of this analysis, the 100-RP has been used (Dottori et al., 2016b). 100-RP is the return period used for the preparation of the flood hazard and risk maps, set forth in Article 6 of the European Flood Directive (European Parliament and the Council of the European Union 2007, para. 7). The resolution of the dataset is 30 arc-sec (approx. 1 km at the equator), and the cell values indicate water depth (in m). First, the hazard map is aligned with the World Mollweide 1 km^2 grid, and then it is overlaid with the Urban Centre layer to delineate the potentially exposed areas within each Urban Centre spatial domain. The derived hazard map represents areas flooded with 1 cm or more (that includes all affected area, independently by the height of water), per each cell of 1x1 km grid in World Mollweide projection. The limit of the minimum water depth in the flood map was set at 1 cm as it was in other previous experiments (Alfieri et al., 2017). This exposed areas are intersected with the multitemporal population and built-up area grids (GHS-POP and GHS-BUILT) for all available epochs (1975-1990-2000-2015). The resulted attributes are:

— **EX_FD_AREA**: total area of the Urban Centre of 2015 potentially exposed to floods, expressed in squares kilometres;

— **EX_FD_B75**: built-up area exposure to floods in 1975, i.e., the total built-up area in 1975 of the Urban Centre (of 2015) potentially exposed to floods, expressed in squares kilometres;

— **EX_FD_B90**: built-up area exposure to floods in 1990, i.e., the total built-up area in 1990 of the Urban Centre (of 2015) potentially exposed to floods, expressed in squares kilometres;

— **EX_FD_B00**: built-up area exposure to floods in 2000, i.e., the total built-up area in 2000 of the Urban Centre (of 2015) potentially exposed to floods, expressed in squares kilometres;

— **EX_FD_B15**: built-up area exposure to floods in 2015, i.e., the total built-up area in 2015 of the Urban Centre (of 2015) potentially exposed to floods, expressed in squares kilometres;

— **EX_FD_P75**: population exposure to floods in 1975, i.e., total population in 1975 within the Urban Centre (of 2015) potentially exposed to floods, expressed in number of people;

— **EX_FD_P90**: population exposure to floods in 1990, i.e., total population in 1990 within the Urban Centre (of 2015) potentially exposed to floods, expressed in number of people;

— **EX_FD_P00**: population exposure to floods in 2000, i.e., total population in 2000 within the Urban Centre (of 2015) potentially exposed to floods, expressed in number of people;

— **EX_FD_P15**: population exposure to floods in 2015, i.e., total population in 2015 within the Urban Centre (of 2015) potentially exposed to floods, expressed in number of people;

4.2.6.2 Storm surge

The exposure of Urban Centres to storm surge is calculated based on a hazard return period of 250 years. The storm surge hazard map is derived from the (1) global GAR 15 dataset of the wave high and (2) the DEM data, i.e., the Shuttle Radar Topography Mission v4.1 (Jarvis & Guevara, 2008), by applying the same methodology used in Pesaresi et al. 2017, as a global application of the one described in Hoque & Khan, 1997. The storm surge attributes are result of the intersection of the hazard layer and the spatial extent of the Urban Centres. The attributes are the following ones:

— **EX_SS_AREA**: total area of the Urban Centre of 2015 potentially exposed to storm surges, expressed in squares kilometres;

— **EX_SS_B75**: built-up area exposure to storm surges in 1975, i.e., the total built-up area in 1975 of the Urban Centre (of 2015) potentially exposed to storm surges, expressed in squares kilometres;

— **EX_SS_B90**: built-up area exposure to storm surges in 1990, i.e., the total built-up area in 1990 of the Urban Centre (of 2015) potentially exposed to storm surges, expressed in squares kilometres;

— **EX_SS_B00**: built-up area exposure to storm surges in 2000, i.e., the total built-up area in 2000 of the Urban Centre (of 2015) potentially exposed to storm surges, expressed in squares kilometres;

— **EX_SS_B15**: built-up area exposure to storm surges in 2015, i.e., the total built-up area in 2015 of the Urban Centre (of 2015) potentially exposed to storm surges, expressed in squares kilometres;

— **EX_SS_P75**: population exposure to storm surges in 1975, i.e., total population in 1975 within the Urban Centre (of 2015) potentially exposed to storm surges, expressed in number of people;

— **EX_SS_P90**: population exposure to storm surges in 1990, i.e., total population in 1990 within the Urban Centre (of 2015) potentially exposed to storm surges, expressed in number of people;

— **EX_SS_P00**: population exposure to storm surges in 2000, i.e., total population in 2000 within the Urban Centre (of 2015) potentially exposed to storm surges, expressed in number of people;

— **EX_SS_P15**: population exposure to storm surges in 2015, i.e., total population in 2015 within the Urban Centre (of 2015) potentially exposed to storm surges, expressed in number of

4.2.6.3 Earthquakes

The assessment of potential exposure of Urban Centres to earthquake is the result of collaboration with the Global Earthquake Model (GEM) team[27]. The GEM Global Seismic Hazard Map (version 2018.1) depicts the geographic distribution of the Peak Ground

[27] https://www.globalquakemodel.org/

Acceleration (PGA) with a 10% probability of being exceeded in 50 years (Pagani et al., 2018). This map was created by collating maps computed using national and regional probabilistic seismic hazard models, and it is the most up to date global seismic hazard layer available to date. The released layer cover entire globe; however, at the time of the GHS-UCDB production (Jan-Aug 2018) data for some areas were not available yet (South East Asia and California) or in a preliminary version, which is flagged by the control value. Additionally, the PGA estimates are converted to Modified Mercalli Intensity (MMI) scale using the methodology described in Wald et al. (1990). The related attributes are:

— **EX_EQ19PGA**: PGA estimate of the seismic risk, expressed in peak ground acceleration (g) with a 10% probability of being exceeded in 50 years;
— **EX_EQ19MMI**: classification of the seismic risk expressed in MMI scale;
— **EX_EQ19_Q**: a control flag informing about data availability and quality (*available*, *missing* – value not available, *imprecise* – not reliable estimate).

4.2.6.4 Heatwave

Russo et al. (2015) designed the HWMId to take into account both heatwave duration and intensity. HWMId is defined as the maximum magnitude of the heatwaves occurring in a year, where a heatwave is defined as the periods of at least three consecutive days with maximum temperature above the calendar 90^{th} percentile centred on a 31 day window reference period. Yearly gridded data Heatwave Magnitude Index on a 0.5 x 0.5 degree grid were analysed for the period 1980-2010 and used for calculating the HWMId in the 30 years period. The derived attribute is:

— **EX_HW_IDX**: the maximum magnitude of the heatwaves in the period 1980-2010.

4.2.7 SDG

The database include attributes about SDG 11, and in particular on 11.3 and 11.7

4.2.7.1 Land Use Efficiency –11.3.1

Land Use Efficiency (LUE) is the indicator internationally agreed to monitor the "ratio of land consumption growth rate to population growth rate" in the framework of SDG 11[28]. The indicator estimates the interdependence between spatial expansion of Urban Centres and demographic change that takes place in them. In the Urban Centre database the LUE indicator is computed measuring the expansion of the built-up areas in each Urban Centre between 1990 and 2015, and the changes in population counts in the same period (Melchiorri et al., 2018). GHS-UCDB data offers an opportunity to support the SDG 11.3.1 with a baseline information. At present, the indicator is a Tier II one (as a globally agreed methodology to estimate exists, but data are not regularly produced). The attribute on LUE is:

— **SDG_LUE9015**: land use efficiency between 1990 and 2015, estimated for the area of the Urban Centre of 2015.

4.2.7.2 Access to green –11.7.1

Access to green spaces is measured using a proxy metric, *"generalised potential access to green areas"*, which is based on the calculation of the amount of people living in areas of high green, at the generalisation scale of the spatial data used for the assessment, and regardless its usability. The metric builds on the greenness metric derived from remote sensing Landsat imagery and described in details in Section 4.2.5 (p.31) and in

[28] https://unstats.un.org/sdgs/metadata/files/Metadata-11-03-01.pdf

(Corbane et al., 2018b). Per each Urban Centre, the area of high green in 2015 is delineated (see Section 4.2.5 for details), and then intersected with the population grid of epoch 2015. Finally, the share of the Urban Centre population living in this area is estimated from the total population of the Urban Centre.

Owing to the variety of types of open spaces, we measured the surface of open spaces within Urban Centres as the union of the areas of non-built-up surfaces and high green surfaces. The principle is illustrated in Figure 14 for the city centre of Paris with the area of non-built-up surfaces shown in grey in Figure 14-a, the area of High Green surfaces shown in green in Figure 14-b and the proxy area of open spaces shown in red in Figure 14-c.

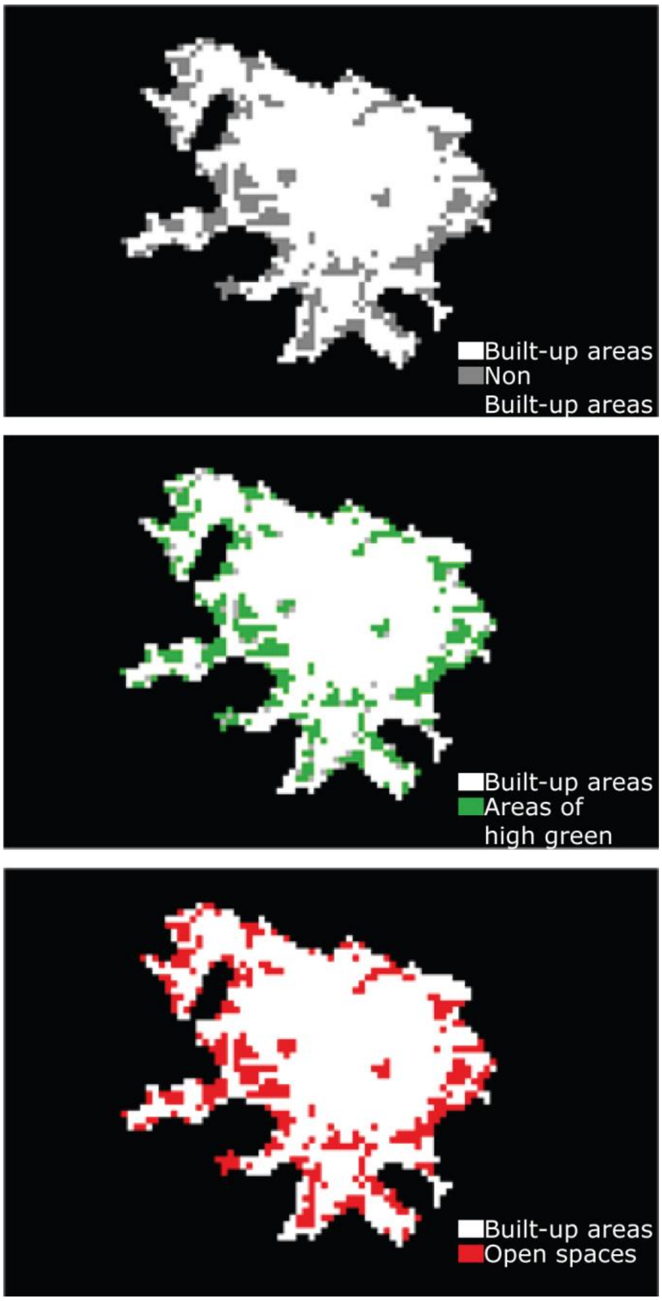

Figure 14. Illustration of the method used to estimate the area of open spaces based on areas of non-built-up surfaces and high green surfaces within the Urban Centre of Paris.

The produced attributes are:

— **SDG_A2G14**: share of population living in the high green area in 2015 in the Urban Centre of 2015, and having value in range 0-1;
— **SDG_OS15MX**: percentage of the open spaces within the area of the Urban Centre of 2015, and having value in range 0-100.

5 References

Alfieri, L., Bisselink, B., Dottori, F., Naumann, G., de Roo, A., Salamon, P., Wyser, K. and L. Feyen. 2017. Global projections of river flood risk in a warmer world. Earth's Future, 5(2):171-182. https://doi.org/10.1002/2016EF000485.

Anderson, K., Ryan, B., Sonntag, W., Kavvada, A. and L. Friedl. 2017. Earth Observation in Service of the 2030 Agenda for Sustainable Development. *Geo-Spatial Information Science*, 20(2):77–96. https://doi.org/10.1080/10095020.2017.1333230.

Annerstedt van den Bosch, M., Mudu, P., Uscila, V., Barrdahl, M., Kulinkina, A., Staatsen, B., Swart, W., Kruize, H., Zurlyte, I. and A.I. Egorov. 2016. Development of an urban green space indicator and the public health rationale. *Scandinavian Journal of Public Health*, 44:159–167. https://doi.org/10.1177/1403494815615444.

Balk, D.L., Deichmann, U., Yetman, G., Pozzi, F., Hay, S.I. and A. Nelson. 2006. Determining Global Population Distribution: Methods, Applications and Data. *Advances in Parasitology*, 62:119-156. https://doi.org/10.1016/S0065-308X(05)62004-0.

Center For International Earth Science Information Network-CIESIN-Columbia University; International Food Policy Research Institute-IFPRI; The World Bank; Centro Internacional De Agricultura Tropical-CIAT. 2011. Global Rural-Urban Mapping Project, Version 1 (GRUMPv1): Settlement Points. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H4M906KR.

Chan, C.K. and X. Yao. 2008. Air Pollution in Mega Cities in China. *Atmospheric Environment,* 42(1):1–42. https://doi.org/10.1016/j.atmosenv.2007.09.003.

Chrysoulakis, N., Feigenwinter, C., Triantakonstantis, D., Penyevskiy, I., Tal, A., Parlow, E., Fleishman, G., Düzgün, S., Esch, T. and M. Marconcini. 2014. A Conceptual List of Indicators for Urban Planning and Management Based on Earth Observation. *ISPRS International Journal of Geo-Information,* 3(3):980–1002. https://doi.org/10.3390/ijgi3030980.

Corbane, C., G. Lemoine, M. Pesaresi, T. Kemper, F. Sabo, S. Ferri, and V. Syrris. 2018. Enhanced Automatic Detection of Human Settlements Using Sentinel-1 Interferometric Coherence. *International Journal of Remote Sensing,* 39(3):842–53. https://doi.org/10.1080/01431161.2017.1392642.

Corbane, C., Pesaresi, M., Politis, P., Florczyk, J.A., Melchiorri, M., F, S., Schiavina M., Ehrlich Daniele, Naumann Gustavo, and Kemper Thomas. 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*, October, 1–18. https://doi.org/10.1080/17538947.2018.1530311.

Corbane, C., Kemper, T., Freire, S., Louvrier, C. and M. Pesaresi. 2016. Monitoring the Syrian Humanitarian Crisis with the JRC's Global Human Settlement Layer and Night-Time Satellite Data. EUR 27933. Luxembourg: Publications Office of the European Union. https://doi.org/10.2788/297909.

Corbane, C., Pesaresi, M., Politis, P., Syrris, V., Florczyk, A.J., Soille, P., Maffenini, L., Burger, A., Vasilev, V., Rodriguez, D., Sabo, F., Dijkstra, L. and T. Kemper. 2017. Big Earth Data Analytics on Sentinel-1 and Landsat Imagery in Support to Global Human Settlements Mapping. *Big Earth Data* 1 (1–2):118–44. https://doi.org/10.1080/20964471.2017.1397899.

Crippa, M., Guizzardi, D., Muntean, M., Schaaf, E., Dentener, F., van Aardenne, J.A., Monni, S., Doering, U., Olivier, J., Pagliari, V. and G. Janssens-Maenhout. 2018. Gridded Emissions of Air Pollutants for the Period 1970–2012 within EDGAR v4.3.2. *Earth System Science Data* 10(4):1987–2013. https://doi.org/10.5194/essd-10-1987-2018.

Danielson, J.J.. 1996. Delineation of drainage basins from 1 km African digital elevation data. In: Pecora Thirteen, Human Interactions with the Environment - Perspectives from Space, Sioux Falls, South Dakota, August 20-22, 1996.

Dijkstra, L. and H. Poelman. 2014. A Harmonised Definition of Cities and Rural Areas: The New Degree of Urbanisation. Working Papers, Regional Working Paper 2014. http://ec.europa.eu/regional_policy/en/information/publications/working-papers/2014/a-harmonised-definition-of-cities-and-rural-areas-the-new-degree-of-urbanisation.

Doherty, R.M., Hulme, M. and C.G. Jones. 1999. A Gridded Reconstruction of Land and Ocean Precipitation for the Extended Tropics from 1974 to 1994. *International Journal of Climatology* 19(2):119–42. https://doi.org/10.1002/(SICI)1097-0088(199902)19:2<119::AID-JOC358>3.0.CO;2-X.

Donaldson, D., and A. Storeygard. 2016. The View from Above: Applications of Satellite Data in Economics. *Journal of Economic Perspectives* 30(4):171–98. https://doi.org/10.1257/jep.30.4.171.

Dottori, F., Salamon, P., Bianchi, A., Alfieri, L., Hirpa, F.A. and L. Feyen. 2016a. Development and Evaluation of a Framework for Global Flood Hazard Mapping. Advances in Water Resources 94:87–102.

Dottori, F., Alfieri, L., Salamon, P., Bianchi, A., Feyen, L. and F. Hirpa. 2016b. Flood hazard map of the World - 100-year return period. European Commission, Joint Research Centre (JRC) [Dataset] PID: http://data.europa.eu/89h/jrc-floods-floodmapgl_rp100y-tif.

European Commission, Joint Research Centre. 2018. Atlas of the Human Planet 2018 – A World of Cities, EUR 29497 EN, European Commission, Luxembourg, 2018, ISBN 978-92-79-98185-2, https://doi.org/10.2760/124503, JRC114316.

Florczyk A.J., Ehrlich D., Corban C., Freire S., Kemper T., Melchiorri M., Pesaresi M., Politis P., Schiavina M. and L. Zanchetta. 2018. Community pre-Release of GHS Data Package (GHS CR2018) in support to the GEO Human Planet Initiative, EUR 29466 EN, Publications Office of the European Union, Luxembourg. https://doi.org/10.2760/777868.

Florczyk, A.J., Corbane, C., Schiavina, M., Pesaresi, M., Maffenini, L., Melchiorri, M., Politis, P., Sabo, F., Freire, S., Ehrlich, D., Kemper, T., Tommasi, P., Airaghi, D. and L. Zanchetta. 2019. GHS Urban Centre Database 2015, multitemporal and multidimensional attributes, R2019A. European Commission, Joint Research Centre (JRC) [Dataset] PID: http://data.europa.eu/89h/53473144-b88c-44bc-b4a3-4583ed1f547e

Dosio, A., Mentaschi, L., Fischer, E.M. and K. Wyser. 2018. Extreme Heat Waves under 1.5 °C and 2 °C Global Warming. *Environmental Research Letters* 13 (5):054006. https://doi.org/10.1088/1748-9326/aab827.

GEO. 2017. Earth Observations in Supports of the 2030 Agenda for Sustainable Development, JAXA. Online (last access on 23/01/2019): https://www.earthobservations.org/documents/publications/201703_geo_eo_for_2030_agenda.pdf.

Geonames. 2018. GeoNames Free Gazetteer Data. Online (last access on 23/01/2018) http://download.geonames.org/.

Ellis, E.C. 2018. *Anthropocene: A Very Short Introduction*. Vol. 1. Oxford University Press. https://doi.org/10.1093/actrade/9780198792987.001.0001.

Ellis, E.C. and N. Ramankutty. 2008. Putting People in the Map: Anthropogenic Biomes of the World. *Frontiers in Ecology and the Environment* 6 (8):439–47. https://doi.org/10.1890/070062.

Elvidge, C. D., K. E. Baugh, S. J. Anderson, P. C. Sutton, and T. Ghosh. 2012. The Night Light Development Index (NLDI): A Spatially Explicit Measure of Human Development from Satellite Data. *Social Geography* 7:23–35.

Elvidge, C.D., Baugh, K., Zhizhin, M., Hsu, F.C. and T. Ghosh. 2017. VIIRS Night-Time Lights. *International Journal of Remote Sensing* 38 (21):5860–79. https://doi.org/10.1080/01431161.2017.1342050.

Elvidge, C.D., Safran, J., Tuttle, B., Sutton, P., Cinzano, P., Pettit, D., Arvesen, J. and C. Small. 2007. Potential for Global Mapping of Development via a Nightsat Mission. *GeoJournal* 69 (1–2):45–53. https://doi.org/10.1007/s10708-007-9104-x.

Elvidge, C.D., Sutton, P.C., Tuttle, B.T., Ghosh, T. and K.E. Baugh. 2010. Global Urban Mapping Based on Nighttime Lights. *Global Mapping of Human Settlements, Taylor and Francis, London*, 129–44.

Fischer, G., Nachtergaele, F., Prieler, S., van Velthuizen, H.T., Verelst, L. and D. Wiberg. 2008. Global Agro-Ecological Zones Assessment for Agriculture (GAEZ 2008). IIASA, Laxenburg, Austria and FAO, Rome, Italy.

EORC and JAXA. 2017. ALOS Global Digital Surface Model (DSM) ALOS World 3D-30m (AW3D30) Dataset. Product Format Description Version 1.1. Online (last access on 23/01/2019): https://www.eorc.jaxa.jp/ALOS/en/aw3d30/aw3d30v11_format_e.pdf.

Freire, S., Doxsey-Whitfield, E., MacManus, K., Mills, J. and M. Pesaresi. 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.

GADM. 2018. GADM data. Online (last access on 23/01/2019) https://gadm.org/

Gan, M., Deng, J., Zheng, X., Hong, Y. and K. Wang. 2014. Monitoring Urban Greenness Dynamics Using Multiple Endmember Spectral Mixture Analysis. Edited by Clinton N. Jenkins. *PLoS ONE* 9 (11):e112202. https://doi.org/10.1371/journal.pone.0112202.

Gehlke, C. E. and K. Biehl. 1934. Certain Effects of Grouping Upon the Size of the Correlation Coefficient in Census Tract Material. *Journal of the American Statistical Association* 29 (185):169. https://doi.org/10.2307/2277827.

Ghosh, T., Anderson, S., Elvidge, C. and P. Sutton. 2013. Using Nighttime Satellite Imagery as a Proxy Measure of Human Well-Being. *Sustainability* 5 (12):4988–5019. https://doi.org/10.3390/su5124988.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D. and R. Moore. 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone, Remote Sensing of Environment, 202:18-27, DOI:10.1016/j.rse.2017.06.031.

GRDC. 2007. Major River Basins of the World / Global Runoff Data Centre. Koblenz, Germany: Federal Institute of Hydrology (BfG). Online (last access on 23/01/2019): https://www.bafg.de/GRDC/EN/02_srvcs/22_gslrs/221_MRB/riverbasins.html?nn=201570.

Harris, I., Jones, P.D., Osborn, T.J. and D.H. Lister. 2014. Updated High-Resolution Grids of Monthly Climatic Observations - the CRU TS3.10 Dataset: UPDATED HIGH-

RESOLUTION GRIDS OF MONTHLY CLIMATIC OBSERVATIONS. *International Journal of Climatology,* 34(3):623–42. https://doi.org/10.1002/joc.3711.

Hoque, M.M.A. and S.A.M. Khan. 1997. Storm Surge Flooding in Chittagong City and Associated Risks. In Proc. of the Conference held at Anaheim, California, June 1996, IAHS Publ. 239:115-122, 1997.

IPCC. 2014. Climate Change 2014: Synthesis Report. Contribution of Working Papers I, II, and III to the Fifth Assessment Report of the Intergovernmantal Panel on Climate Change. Geneva: IPCC.

Jarvis, A., Reuter, H.I. and E. Guevara. 2008. Hole-Filled SRTM for the Globe Version 4. Available from the CGIAR-CSI SRTM 90m Database. http://srtm.csi.cgiar.org.

Jones, P.D., and M. Hulme. 1996. CALCULATING REGIONAL CLIMATIC TIME SERIES FOR TEMPERATURE AND PRECIPITATION: METHODS AND ILLUSTRATIONS. *International Journal of Climatology,* 16(4):361–77. https://doi.org/10.1002/(SICI)1097-0088(199604)16:4<361::AID-JOC53>3.0.CO;2-F.

Kottek, M., Grieser, J., Beck, C., Rudolf, B. and F. Rubel. 2006. World Map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift*, 15(3):259-263. https://doi.org/10.1127/0941-2948/2006/0130.

Kummu, M., Taka, M. and J.H.A. Guillaume. 2018. Data from: Gridded Global Datasets for Gross Domestic Product and Human Development Index over 1990-2015. Dryad Digital Repository. https://doi.org/10.5061/dryad.dk1j0.

Landsberg, H.E. 1976. WEATHER, CLIMATE AND HUMAN SETTLEMENTS. 448. World Meteorological Organization - SPECIAL ENVIRONMENTAL REPORT 7.

Le Texier, M., Schiel, K. and G. Caruso. 2018. The provision of urban green space and its accessibility: Spatial data effects in Brussels. PLOS ONE, 13(10):e0204684. https://doi.org/10.1371/journal.pone.0204684.

Lee, A., Jordan, H. and J. Horsley. 2015. Value of Urban Green Spaces in Promoting Healthy Living and Wellbeing: Prospects for Planning. *Risk Management and Healthcare Policy*, 8:131—137. https://doi.org/10.2147/RMHP.S61654.

Melchiorri, M., Pesaresi, M., Florczyk, A.J., Corbane, C. and T. Kemper. 2018. Principles and Applications of the Global Human Settlement Layer as Baseline for the Land Use Efficiency Indicator –SDG 11.3.1. Preprints 2018, 2018100085 (doi: 10.20944/preprints201810.0085.v1).

Morice, C.P., Kennedy, J.J., Rayner, N.A. and P.D. Jones. 2012. Quantifying Uncertainties in Global and Regional Temperature Change Using an Ensemble of Observational Estimates: The HadCRUT4 Data Set: THE HADCRUT4 DATASET. *Journal of Geophysical Research: Atmospheres*, 117(D8). https://doi.org/10.1029/2011JD017187.

Muntean, M., Guizzardi, D., Schaaf, E., Crippa, M., Solazzo, E., Oliver, J.G.J. and E. Vignati. 2018. CO2 Emissions of All World Countries. *Publications Office of the European Union*. https://doi.org/10.2760/30158.

Myers, M.. 1975. Decision Making in Allocating Metropolitan Open Space: State of the Art. *Transactions of the Kansas Academy of Science (1903-)*, 78(3/4):149-153. https://doi.org/10.2307/3627339.

Ngom, R., Gosselin, P. and C. Blais. 2016. Reduction of disparities in access to green spaces: Their geographic insertion and recreational functions matter. *Applied Geography*, 66:35–51. https://doi.org/10.1016/j.apgeog.2015.11.008

Noort, Mark. 2017. Earth Observation and Sustainable Development Goals in the Netherlands Towards More Synergetic Use of Earth Observation: An Exploratory Study. Ministry of Foreign Affairs of the Netherlands.

Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V.N., Underwood, E.C., D'amico, J.A., Itoua, I., Strand, H.E., Morrison, J.C., Loucks, C.J., Allnutt, T.F., Ricketts, T.H., Kura, Y., Lamoreux, J.F., Wettengel, W.W., Hedao, P. and K.R. Kassem . 2001. Terrestrial Ecoregions of the World: A New Map of Life on Earth. *BioScience,* 51(11):933. https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2.

Paganini, M. and I. Petiteville. 2018. *Satellite Earth Observations in Support of the Sustainable Development Goals*. Vol. Special 2018 Edition. The CEOS Earth Observation Handbook. CEOS - ESA.

Pagani, M., Garcia-Pelaez, J., Gee, R., Johnson, K. , Poggi, V., Styron, R., Weatherill, G., Simionato, M., Viganò, D., Danciu, L. and D. Monelli. 2018. Global Earthquake Model (GEM) Seismic Hazard Map (version 2018.1 - December 2018), https://doi.org/10.13117/GEM-GLOBAL-SEISMIC-HAZARD-MAP-2018.1.

Patterson, T. and N.V. Kelso. 2018. Natural Earth Populated Places v4.1.0. Online (last access on 23/01/2019) https://www.naturalearthdata.com/downloads/10m-cultural-vectors/10m-populated-places/.

Pekel, J.-F., Cottam, A., Gorelick, N. and A.S. Belward. 2016. High-resolution mapping of global surface water and its long-term changes. *Nature,* 540:418-422. https://doi.org/10.1038/nature20584.

Pesaresi, M., Huadong, G., Blaes, X., Ehrlich, D., Ferri, S., Gueguen, L., Halkia, M., Kauffmann, M. and T. Kemper. 2013. A Global Human Settlement Layer From Optical HR/VHR RS Data: Concept and First Results. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 6(5):2102–31. https://doi.org/10.1109/JSTARS.2013.2271445.

Pesaresi, M., Ehrlich, D., Ferri, S., Florczyk, A.J., Freire, S., Halkia, S., Julea, A., Kemper, T., Soille, P. and V. Syrris. 2016. *Operating Procedure for the Production of the Global Human Settlement Layer from Landsat Data of the Epochs 1975, 1990, 2000, and 2014*. Publications Office of the European Union. http://publications.jrc.ec.europa.eu/repository/handle/111111111/40182.

Pesaresi, M., Ehrlich, D., Kemper, T., Siragusa, A., Florczyk, A.J., Freire, S. and C. Corbane. 2017. Atlas of the Human Planet 2017: Global Exposure to Natural Hazards, EUR 28556 EN, https://doi.org/10.2760/19837.

Rubel, F., Brugger, K., Haslinger, K. and I. Auer. 2017. The Climate of the European Alps: Shift of Very High Resolution Köppen-Geiger Climate Zones 1800–2100. *Meteorologische Zeitschrift,* 26(2):115–25. https://doi.org/10.1127/metz/2016/0816.

Russo, S., Sillmann, J. and E.M. Fischer. 2015. Top Ten European Heatwaves since 1950 and Their Occurrence in the Coming Decades. *Environmental Research Letters,* 10(12):124003. https://doi.org/10.1088/1748-9326/10/12/124003.

Seneviratne, S.I., Donat, M.G., Pitman, A.J., Knutti, R. and R.L. Wilby. 2016. Allowable CO2 Emissions Based on Regional and Impact-Related Climate Targets. *Nature,* 529:477-483.

Shaddick, G., Thomas, M.L., Amini, H., Broday, D., Cohen, A., Frostad, J., Green, A., Gumy, S., Liu, Y., Martin, R.V., Pruss-Ustun, A., Simpson, D., van Donkelaar, A. and M. Brauer. 2018. Data Integration for the Assessment of Population Exposure to Ambient Air Pollution for Global Burden of Disease Assessment. *Environmental Science & Technology,* 52(16):9069–78. https://doi.org/10.1021/acs.est.8b02864.

EEA, 1995. Europe’s environment: the Dobříš assessment. Stanners, D.A. and P. Bourdeau (eds.), European Environment Agency, Copenhagen.

Tadono T., Ishida H., Oda F., Naito S., Minakawa K. and H. Iwamoto. 2014. Precise Global DEM Generation By ALOS PRISM, ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, II-4, 71-76, 2014. https://doi.org/10.5194/isprsannals-II-4-71-2014.

UN Statistical Commission. 2017. Work of the Statistical Commission Pertaining to the 2030 Agenda for Sustainable Development. United Nations. http://ggim.un.org/documents/A_RES_71_313.pdf.

UNDESA. 2008. World Urbanization Prospects The 2007 Revision. United Nations.

———. 2018a. The World's Cities in 2018 Data Booklet ST/ESA/ SER.A/417.

———. 2018b. World Population Prospects The 2018 Revision. United Nations.

United Nations. 2015. Transforming Our World: The 2030 Agenda for Sustainable Development. United Nations. A/RES/70/1. United Nations, General Assembly.

United Nations. 2018. Tracking Progress Towards Inclusive, Safe, Resilient and Sustainable Cities and Human Settlements SDG 11 SYNTHESIS REPORT HIGH LEVEL POLITICAL FORUM 2018. United Nations, Human Settlements Programme. http://uis.unesco.org/sites/default/files/documents/sdg11-synthesis-report-2018-en.pdf.Van Donkelaar, A. 2018. Global Annual PM2.5 Grids from MODIS, MISR and SeaWiFS Aerosol Optical Depth (AOD) with GWR, 1998-2016. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H4ZK5DQS.

Wald, D.J., Quitoriano, V., Heaton, T.H. and H. Kanamori. 1999. Relationships between Peak Ground Acceleration, Peak Ground Velocity, and Modified Mercalli Intensity in California. *Earthquake Spectra*, 15(3):557-564. https://doi.org/10.1193/1.1586058.

Weiss, D.J., , Nelson, A., Gibson, H.S., Temperley, W., Peedell, S., Lieber, A., Hancher, M., Poyart, E., Belchior, S., Fullman, N., Mappin, B., Dalrymple, U., Rozier, J., Lucas, T.C.D., Howes, R.E., Tusting, L.S., Kang, S.Y., Cameron, E., Bisanzio, D., Battle, K.E., Bhatt, S. and P.W. Gething. 2018. A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature*, 553:333-36. https://doi.org/10.1038/nature25181.

WHO. 2018. Ambient (Outdoor) Air Quality and Health -Factsheet. https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health.

Zampieri, M., Russo, S., di Sabatino, S., Michetti, M., Scoccimarro, E. and S. Gualdi. 2016. Global Assessment of Heat Wave Magnitudes from 1901 to 2010 and Implications for the River Discharge of the Alps. *Science of The Total Environment*, 571:1330–39. https://doi.org/10.1016/j.scitotenv.2016.07.008.

Zell, E., Huff, A.K., Carpenter, A.T. and L.A. Friedl. 2012. A User-Driven Approach to Determining Critical Earth Observation Priorities for Societal Benefit. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 5(6):1594–1602. https://doi.org/10.1109/JSTARS.2012.2199467.

List of abbreviations and definitions

ALOS – Advanced Land Observation Satellite

BUILT – Built-up

CIESIN – Center for International Earth Science Information Network

CO_2 – Carbon dioxide

DESA – Department of Economic and Social Affairs

DNB – Day/Night Band

DRR – Disaster Risk Reduction

DSM – Digital Surface Model

EC – European Commission

EDGAR – Emissions Database for Global Atmospheric Research

EO – Earth Observation

EPSG – European Petroleum Survey Group

ESA – European Space Agency

EU – European Union

EUMETSAT – European Organisation for the Exploitation of Meteorological Satellites

FAO – Food and Agriculture Organization of the United Nations

GAR – Global Assessment Report

GBD - Global Burden of Disease

GDP – Gross Domestic Product

GEE – Google Earth Engine

GEM – Global Earthquake Model

GEO – Group on Earth Observation

GHS-BUILT – Global Human Settlement data on built-up surfaces

GHSL – Global Human Settlement Layer

GHS-POP – Global Human Settlement data on resident population

GHS-SMOD – Global Human Settlement data on rural/urban classification

GHS-UCDB – Global Human Settlement data on Urban Centres

GloFAS – Global Flood Awareness System

GPW – Gridded Population of the World

GSARS – Global Strategy to improve Agricultural and Rural Statistics

HABITAT – United Nations Human Settlements Programme

HDC – High Density Clusters

HWMId – HeatWave Magnitude Index

JRC – Joint Research Centre

LUE – Land Use Efficiency

MASL – Metres Above Sea Level

MMI – Mercalli Modified Intensity scale

MRB – Major River Basins

NCEI – National Centers for Environmental Information

NDVI – Normalized differential vegetation index

NOAA – National Oceanic and Atmospheric Administration

OECD – Organisation for Economic Co-operation and Development

PGA - Peak Ground Acceleration

$PM_{2.5}$ – Fine Particulate Matter

SDG – Sustainable Development Goals

SMOD – Settlement MODel

TOA – Top-of-Atmosphere

UC – Urban Centre

UCDB – Urban Centre Database

VHR – Very High Resolution

VIIRS – Visible Infrared Imaging Radiometer Suite

WGS 84 – World Geodetic System 1984

WHO – World Health Organization

WUP – World Urbanization Prospect

List of figures

List of tables

Annexes

Annex 1. Urban Centre spatial domain: graphical explanation

This Annex the summarises the process of defining the spatial domain of the Urban Centre, described in Section 3.3, by means of visual representations of grid values. In order to simplify the explanation, POP_x and BU_x are used in input, instead than POP'_x and BU'_x.

There are the following steps presented graphically:

A) Load the $\boldsymbol{POPx}$ input grid

B) Load the $\mathbf{BU_x}$ input grid

C) Select all the x samples where $\mathbf{POP_x}$ > 1500

D) Select all the x samples where $\mathbf{BU_x}$ > 0.5

E) The support set $\mathbf{HDC_x^{supp}}$ as union of the outputs at the steps C and D

F) Identify the 4-conn clusters $\mathbf{HDC_x^c}$ from the of $\mathbf{HDC_x^{supp}}$ output at step E

G) Identify the clusters having more than 50000 population size

Step A: Load the $POPx$ input grid

Row	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	c11	c12	c13	c14	c15	c16	c17	c18	c19	c20	c21	c22	c23	c24	c25	c26	c27	c28	c29	c30
r 1	186	270	150	51	284	156	91	144	285	294	219	294	97	252	138	268	35	48	66	134	226	144	254	114	290	244	231	241	236	73
r 2	276	140	281	246	179	249	146	200	284	252	238	254	138	209	291	0	38	43	217	129	188	249	112	283	123	219	131	288	272	272
r 3	190	147	231	861	254	154	103	64	244	127	474	280	202	143	101	0	12	15	149	212	249	92	512	1089	963	97	206	85	101	84
r 4	69	198	150	1403	279	253	268	243	129	254	131	1465	211	229	262	282	0	13	18	15	281	142	230	292	212	1321	126	118	166	154
r 5	106	221	145	972	654	56	74	76	300	255	143	920	204	138	273	253	29	21	28	43	12	131	166	132	245	998	68	298	69	189
r 6	254	62	95	1001	744	923	267	220	223	92	280	138	1355	95	263	187	171	151	232	18	49	25	155	231	286	1445	189	284	158	140
r 7	286	164	218	81	1274	306	210	248	264	76	160	218	1201	61	253	242	120	107	78	226	8865	160	200	220	68	905	1119	136	517	1277
r 8	84	137	168	124	375	714	172	69	75	82	134	144	1265	178	284	190	129	252	159	287	9725	213	203	70	116	1355	1314	156	1163	1449
r 9	206	111	57	297	251	1113	1064	683	254	86	161	215	113	524	995	260	1019	1316	340	153	4309	779	612	984	566	1143	780	78	758	1171
r10	77	152	196	107	195	655	502	424	844	80	288	249	105	207	1384	79	634	1313	1412	975	6849	914	1437	1182	488	1446	805	233	692	681
r11	196	872	356	775	362	350	715	220	373	0	0	0	0	0	284	799	213	529	509	1401	11223	734	486	1075	571	787	393	1421	1450	856
r12	1423	918	998	654	479	1437	1176	90	881	0	0	0	0	0	258	1018	824	474	881	1048	4126	1356	15541	1048	683	1330	1283	1181	476	579
r13	817	590	6424	1003	968	1003	754	1063	1047	0	0	0	0	0	188	1188	424	997	1458	435	7815	12185	11363	8743	752	585	4549	600	8253	6697
r14	1346	9609	15339	5720	468	779	1488	0	0	0	0	0	196	1159	874	595	1328	744	548	3505	9454	14709	8032	9041	1495	16193	2447	6076	4997	4634
r15	13093	6459	3250	5070	725	1292	303	1206	0	0	0	0	1236	507	729	1148	360	997	16318	10270	2052	422	7581	10308	2887	14622	9771	6845	9635	1380
r16	330	1636	14968	15445	7015	1240	1224	1287	1024	3892	0	1393	550	1306	507	1174	462	374	10103	3562	12803	720	577	396	1101	1015	956	894	778	700
r17	929	1335	1255	6713	15470	1416	944	959	1195	14225	13029	14589	1084	3837	15165	11959	2048	0	0	0	13863	16256	2939	1138	499	4104	14690	10294	1450	687
r18	852	678	1280	1156	573	2036	1351	620	756	1156	548	3944	873	5430	6724	10341	9150	0	0	0	4293	4943	0	658	1447	12636	373	1047	738	1274
r19	943	197	1497	1062	1395	1337	4955	1249	702	583	1024	13717	15165	15612	3683	872	3420	6865	4464	0	0	0	0	926	1260	2636	1058	381	993	383
r20	133	144	262	178	629	1303	328	13279	3554	8585	8939	14181	10378	8515	11868	591	4624	2846	0	0	0	0	0	12004	6604	14580	962	1361	1157	787
r21	299	70	1120	904	1324	775	1485	1226	878	10914	4852	15539	8266	10931	340	914	3681	11987	0	0	4734	12288	6699	5475	5869	1405	1287	640	794	974
r22	234	1353	305	748	1296	869	1174	860	1391	928	7742	6944	12955	1867	4120	4493	6190	0	0	0	0	1597	15544	9849	13207	4059	1449	319	1237	712
r23	191	90	1404	751	1436	377	888	951	623	652	1157	13159	16365	15116	7148	6725	13487	0	0	0	0	9645	15815	1864	5758	12625	1462	1207	760	1166
r24	75	1257	759	549	339	819	738	427	4790	7445	10490	15860	3192	5304	13072	6235	6564	9897	0	0	3333	4116	16398	7214	11681	12987	1677	5839	383	525
r25	76	792	1142	9264	971	1288	1366	781	5589	4835	9468	13964	10529	11813	7172	9618	1774	16425	0	2330	7396	1428	884	6704	6937	2282	311	1035	953	783
r26	122	1126	646	15353	4104	657	580	339	1126	1399	743	2256	5521	4456	8477	8102	4443	456	0	10789	7443	418	651	8371	7881	11756	16500	4244	816	392
r27	240	1068	563	5447	8861	8008	1285	800	409	997	687	14106	7407	8035	1515	12673	849	875	3302	3814	13354	1878	11091	7099	6411	5642	15233	4585	799	851
r28	86	106	1423	2392	3281	4328	931	725	1322	12661	14374	5899	11248	4909	1769	994	834	983	511	15850	5669	15358	6869	7295	4944	11780	5927	756	1486	1421
r29	52	428	771	4043	415	1363	15533	8488	6498	5452	16495	11567	10128	981	1467	861	759	481	895	543	770	14524	8624	3956	10082	1111	584	537	1405	704
r30	1137	942	375	8730	2831	2634	1086	12951	704	407	1120	2366	10708	1299	1233	12750	6992	1037	705	956	828	6071	3404	8666	11083	488	450	828	969	1447
r31	1192	727	1245	11089	4132	867	948	1261	606	1241	1190	13690	13605	591	540	919	10246	461	518	1117	462	4350	3284	1432	1174	713	1063	1234	263	168
r32	1340	616	606	12080	13655	383	726	478	1352	1122	1206	14922	10481	1422	962	1246	1415	639	1404	333	993	14032	6620	4393	6419	720	1182	1425	131	95
r33	1316	1132	7923	9717	11723	3113	421	10185	476	1397	1201	11583	11171	431	1216	1252	1497	1011	604	3378	7845	4279	10691	15751	14699	578	1254	900	977	954
r34	325	1379	10668	1510	5017	8068	11303	1752	631	997	1404	14300	12496	15421	9741	13098	876	357	979	5477	3976	1426	13034	5911	3975	16440	3843	10944	883	1049
r35	714	524	16234	8754	6200	1147	6362	16234	1109	934	1259	12047	3710	9555	13846	13602	568	1149	10531	7365	1383	533	1200	14295	5565	2153	4883	855	1242	911
r36	869	1290	5792	16016	2005	5386	12666	14178	910	498	321	699	5030	2201	2761	14572	1431	766	13642	786	959	1442	1285	618	7241	4613	8809	501	1047	882
r37	635	1405	10275	494	7987	6124	7534	14505	11064	1200	712	1467	4215	1503	7329	14550	905	532	13963	3621	744	457	546	1096	15020	6569	308	855	190	164
r38	706	1257	8772	993	15649	14153	4517	4370	10720	357	1091	1241	3717	12592	13321	6388	1244	463	10382	830	1099	949	1247	421	9660	1261	1324	1192	97	273
r39	174	452	857	521	1293	1085	11759	4507	10860	533	632	1477	569	450	1236	1439	1499	586	13558	3191	1417	738	12436	3853	3792	856	1124	302	141	184
r40	291	1134	968	979	0	0	0	4096	308	1127	371	1438	1390	936	15103	16163	8988	13596	11094	1884	4460	770	13509	8704	11177	1094	1059	243	79	173
r41	77	162	187	131	0	0	0	1075	877	575	1249	486	1281	495	1051	906	340	1247	579	1347	2262	751	6872	953	12404	1308	840	254	289	298
r42	113	196	133	91	0	0	0	558	364	1496	72	1308	662	1047	988	656	389	1478	796	665	12269	1141	1453	539	3872	1326	1389	202	141	94
r43	232	233	65	209	198	91	323	850	971	565	156	472	1480	327	1120	581	1069	178	317	1118	335	664	1467	1289	1357	866	1192	425	102	105
r44	1061	260	94	247	292	73	626	656	576	156	234	522	433	739	1365	332	1246	195	1463	1107	1203	1129	1097	1162	1062	488	471	474	78	228
r45	44	730	798	199	107	72	159	259	248	147	269	135	402	884	1217	278	242	262	173	92	77	281	559	751	1308	1048	555	877	1035	914
r46	14	35	18	33	730	135	130	241	168	141	80	273	666	611	1032	502	266	133	130	220	118	186	255	119	172	1411	1381	812	571	1023
r47	12	0	38	27	1061	69	200	230	211	64	152	254	159	262	870	1436	85	264	152	119	298	194	206	255	1458	997	512	583	978	1075
r48	18	14	27	30	7	2	211	282	90	281	66	287	118	54	1179	983	188	140	205	257	8644	109	81	206	1042	385	11586	1925	1489	821
r49	36	45	42	18	44	48	5	84	226	284	194	110	197	130	157	1496	163	235	204	955	2162	261	177	141	861	772	16389	5721	1216	875
r50	21	48	4	31	17	24	34	24	366	584	1281	195	119	213	68	283	122	143	257	830	59	115	282	153	468	749	786	1193	1430	1500

Step B: Load the BU_x input grid

Row	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	c11	c12	c13	c14	c15	c16	c17	c18	c19	c20	c21	c22	c23	c24	c25	c26	c27	c28	c29	c30
r 1	0.009	0.013	0.008	0.003	0.014	0.008	0.005	0.007	0.014	0.015	0.011	0.015	0.005	0.013	0.007	0.013	0.002	0.002	0.003	0.007	0.011	0.007	0.013	0.006	0.014	0.012	0.012	0.012	0.012	0.004
r 2	0.014	0.007	0.014	0.012	0.009	0.012	0.007	0.010	0.014	0.013	0.012	0.013	0.007	0.010	0.015	0.510	0.002	0.002	0.011	0.006	0.009	0.012	0.006	0.014	0.006	0.011	0.007	0.014	0.014	0.014
r 3	0.009	0.007	0.012	0.043	0.013	0.008	0.005	0.003	0.012	0.006	0.024	0.014	0.010	0.007	0.005	0.656	0.001	0.001	0.007	0.011	0.012	0.005	0.026	0.054	0.048	0.005	0.010	0.004	0.005	0.004
r 4	0.003	0.010	0.007	0.070	0.014	0.013	0.013	0.012	0.006	0.013	0.007	0.073	0.011	0.011	0.013	0.014	0.000	0.001	0.001	0.001	0.014	0.007	0.011	0.015	0.011	0.066	0.006	0.006	0.008	0.008
r 5	0.005	0.011	0.007	0.049	0.033	0.003	0.004	0.004	0.015	0.013	0.007	0.046	0.010	0.007	0.014	0.013	0.001	0.001	0.001	0.002	0.001	0.007	0.008	0.007	0.012	0.050	0.003	0.015	0.003	0.009
r 6	0.013	0.003	0.005	0.050	0.037	0.046	0.013	0.011	0.011	0.005	0.014	0.007	0.068	0.005	0.013	0.009	0.009	0.008	0.012	0.001	0.002	0.001	0.008	0.012	0.014	0.072	0.009	0.014	0.008	0.007
r 7	0.014	0.008	0.011	0.004	0.064	0.015	0.011	0.012	0.013	0.004	0.008	0.011	0.060	0.003	0.013	0.012	0.006	0.005	0.004	0.011	0.443	0.008	0.010	0.011	0.003	0.045	0.056	0.007	0.026	0.064
r 8	0.004	0.007	0.008	0.006	0.019	0.036	0.009	0.003	0.004	0.004	0.007	0.007	0.063	0.009	0.014	0.010	0.006	0.013	0.008	0.014	0.486	0.011	0.010	0.003	0.006	0.068	0.066	0.008	0.058	0.072
r 9	0.010	0.006	0.003	0.015	0.013	0.056	0.053	0.034	0.013	0.004	0.008	0.011	0.006	0.026	0.050	0.013	0.051	0.066	0.017	0.008	0.215	0.039	0.031	0.049	0.028	0.057	0.039	0.004	0.038	0.059
r10	0.004	0.008	0.010	0.005	0.010	0.033	0.025	0.021	0.042	0.004	0.014	0.012	0.005	0.010	0.069	0.004	0.032	0.066	0.071	0.049	0.342	0.046	0.072	0.059	0.024	0.072	0.040	0.012	0.035	0.034
r11	0.010	0.044	0.018	0.039	0.018	0.018	0.036	0.011	0.019	0.984	0.918	0.689	0.561	0.900	0.014	0.040	0.011	0.026	0.025	0.070	0.561	0.037	0.024	0.054	0.029	0.039	0.020	0.071	0.073	0.043
r12	0.071	0.046	0.050	0.033	0.024	0.072	0.059	0.005	0.044	0.684	0.913	0.594	0.692	0.647	0.013	0.051	0.041	0.024	0.044	0.052	0.206	0.068	0.777	0.052	0.034	0.067	0.064	0.059	0.024	0.029
r13	0.041	0.030	0.321	0.050	0.048	0.050	0.038	0.053	0.052	0.849	0.555	0.876	0.572	0.574	0.009	0.059	0.021	0.050	0.073	0.022	0.391	0.609	0.568	0.437	0.038	0.029	0.227	0.030	0.413	0.335
r14	0.067	0.480	0.767	0.286	0.023	0.039	0.074	0.576	0.549	0.903	0.804	0.852	0.010	0.058	0.044	0.030	0.066	0.037	0.027	0.175	0.473	0.735	0.402	0.452	0.075	0.810	0.122	0.304	0.250	0.232
r15	0.655	0.323	0.162	0.253	0.036	0.065	0.015	0.060	0.655	0.574	0.877	0.665	0.062	0.025	0.036	0.057	0.018	0.050	0.816	0.514	0.103	0.021	0.379	0.515	0.144	0.731	0.489	0.342	0.482	0.069
r16	0.016	0.082	0.748	0.772	0.351	0.062	0.061	0.064	0.051	0.195	0.626	0.070	0.027	0.065	0.025	0.059	0.023	0.019	0.505	0.178	0.640	0.036	0.029	0.020	0.055	0.051	0.048	0.045	0.039	0.035
r17	0.046	0.067	0.063	0.336	0.774	0.071	0.047	0.048	0.060	0.711	0.651	0.729	0.054	0.192	0.758	0.598	0.102	0.508	0.762	0.666	0.693	0.813	0.147	0.057	0.025	0.205	0.734	0.515	0.073	0.034
r18	0.043	0.034	0.064	0.058	0.029	0.102	0.068	0.031	0.038	0.058	0.027	0.197	0.044	0.271	0.336	0.517	0.458	0.970	0.849	0.735	0.215	0.247	0.982	0.033	0.072	0.632	0.019	0.052	0.037	0.064
r19	0.047	0.010	0.075	0.053	0.070	0.067	0.248	0.062	0.035	0.029	0.051	0.686	0.758	0.781	0.184	0.044	0.171	0.343	0.223	0.955	0.520	0.527	0.944	0.046	0.063	0.132	0.053	0.019	0.050	0.019
r20	0.007	0.007	0.013	0.009	0.031	0.065	0.016	0.664	0.178	0.429	0.447	0.709	0.519	0.426	0.593	0.030	0.231	0.142	0.946	0.988	0.851	0.665	0.933	0.600	0.330	0.729	0.048	0.068	0.058	0.039
r21	0.015	0.003	0.056	0.045	0.066	0.039	0.074	0.061	0.044	0.546	0.243	0.777	0.413	0.547	0.017	0.046	0.184	0.599	0.621	0.597	0.237	0.614	0.335	0.274	0.293	0.070	0.064	0.032	0.040	0.049
r22	0.012	0.068	0.015	0.037	0.065	0.043	0.059	0.043	0.070	0.046	0.387	0.347	0.648	0.093	0.206	0.225	0.309	0.867	0.553	0.555	0.697	0.080	0.777	0.492	0.660	0.203	0.072	0.016	0.062	0.036
r23	0.010	0.005	0.070	0.038	0.072	0.019	0.044	0.048	0.031	0.033	0.058	0.658	0.818	0.756	0.357	0.336	0.674	0.692	0.926	0.999	0.614	0.482	0.791	0.093	0.288	0.631	0.073	0.060	0.038	0.058
r24	0.004	0.063	0.038	0.027	0.017	0.041	0.037	0.021	0.240	0.372	0.525	0.793	0.160	0.265	0.654	0.312	0.328	0.495	0.538	0.541	0.167	0.206	0.820	0.361	0.584	0.649	0.084	0.292	0.019	0.026
r25	0.004	0.040	0.057	0.463	0.049	0.064	0.068	0.039	0.279	0.242	0.473	0.698	0.526	0.591	0.359	0.481	0.089	0.821	0.625	0.117	0.370	0.071	0.044	0.335	0.347	0.114	0.016	0.052	0.048	0.039
r26	0.006	0.056	0.032	0.768	0.205	0.033	0.029	0.017	0.056	0.070	0.037	0.113	0.276	0.223	0.424	0.405	0.222	0.023	0.892	0.539	0.372	0.021	0.033	0.419	0.394	0.588	0.825	0.212	0.041	0.020
r27	0.012	0.053	0.028	0.272	0.443	0.400	0.064	0.040	0.020	0.050	0.034	0.705	0.370	0.402	0.076	0.634	0.042	0.044	0.165	0.191	0.668	0.094	0.555	0.355	0.321	0.282	0.762	0.229	0.040	0.043
r28	0.004	0.005	0.071	0.120	0.164	0.216	0.047	0.036	0.066	0.633	0.719	0.295	0.562	0.245	0.088	0.050	0.042	0.049	0.026	0.792	0.283	0.768	0.343	0.365	0.247	0.589	0.296	0.038	0.074	0.071
r29	0.003	0.021	0.039	0.202	0.021	0.068	0.777	0.424	0.325	0.273	0.825	0.578	0.506	0.049	0.073	0.043	0.038	0.024	0.045	0.027	0.039	0.726	0.431	0.198	0.504	0.056	0.029	0.027	0.070	0.035
r30	0.057	0.047	0.019	0.437	0.142	0.132	0.054	0.648	0.035	0.020	0.056	0.118	0.535	0.065	0.062	0.637	0.350	0.052	0.035	0.048	0.041	0.304	0.170	0.433	0.554	0.024	0.023	0.041	0.048	0.072
r31	0.060	0.036	0.062	0.554	0.207	0.043	0.047	0.063	0.030	0.062	0.060	0.684	0.680	0.030	0.027	0.046	0.512	0.023	0.026	0.056	0.023	0.218	0.164	0.072	0.059	0.036	0.053	0.062	0.013	0.008
r32	0.067	0.031	0.030	0.604	0.683	0.019	0.036	0.024	0.068	0.056	0.060	0.746	0.524	0.071	0.048	0.062	0.071	0.032	0.070	0.017	0.050	0.702	0.331	0.220	0.321	0.036	0.059	0.071	0.007	0.005
r33	0.066	0.057	0.396	0.486	0.586	0.156	0.021	0.509	0.024	0.070	0.060	0.579	0.559	0.022	0.061	0.063	0.075	0.051	0.030	0.169	0.392	0.214	0.535	0.788	0.735	0.029	0.063	0.045	0.049	0.048
r34	0.016	0.069	0.533	0.075	0.251	0.403	0.565	0.088	0.032	0.050	0.070	0.715	0.625	0.771	0.487	0.655	0.044	0.018	0.049	0.274	0.199	0.071	0.652	0.296	0.199	0.822	0.192	0.547	0.044	0.052
r35	0.036	0.026	0.812	0.438	0.310	0.057	0.318	0.812	0.055	0.047	0.063	0.602	0.185	0.478	0.692	0.680	0.028	0.057	0.527	0.368	0.069	0.027	0.060	0.715	0.278	0.108	0.244	0.043	0.062	0.046
r36	0.043	0.064	0.290	0.801	0.100	0.269	0.633	0.709	0.045	0.025	0.016	0.035	0.251	0.110	0.138	0.729	0.072	0.038	0.682	0.039	0.048	0.072	0.064	0.031	0.362	0.231	0.440	0.025	0.052	0.044
r37	0.032	0.070	0.514	0.025	0.399	0.306	0.377	0.725	0.553	0.060	0.036	0.073	0.211	0.075	0.366	0.727	0.045	0.027	0.698	0.181	0.037	0.023	0.027	0.055	0.751	0.328	0.015	0.043	0.009	0.008
r38	0.035	0.063	0.439	0.050	0.782	0.708	0.226	0.218	0.536	0.018	0.055	0.062	0.186	0.630	0.666	0.319	0.062	0.023	0.519	0.041	0.055	0.047	0.062	0.021	0.483	0.063	0.066	0.060	0.005	0.014
r39	0.009	0.023	0.043	0.026	0.065	0.054	0.588	0.225	0.543	0.027	0.032	0.074	0.028	0.022	0.062	0.072	0.075	0.029	0.678	0.160	0.071	0.037	0.622	0.193	0.190	0.043	0.056	0.015	0.007	0.009
r40	0.015	0.057	0.048	0.049	0.867	0.900	0.625	0.205	0.015	0.056	0.019	0.072	0.069	0.047	0.755	0.808	0.449	0.680	0.555	0.094	0.223	0.039	0.675	0.435	0.559	0.055	0.053	0.012	0.004	0.009
r41	0.004	0.008	0.009	0.007	0.742	0.996	0.556	0.054	0.044	0.029	0.062	0.024	0.064	0.025	0.053	0.045	0.017	0.062	0.029	0.067	0.113	0.038	0.344	0.048	0.620	0.065	0.042	0.013	0.014	0.015
r42	0.006	0.010	0.007	0.005	0.678	0.803	0.537	0.028	0.018	0.075	0.004	0.065	0.033	0.052	0.049	0.033	0.019	0.074	0.040	0.033	0.613	0.057	0.073	0.027	0.194	0.066	0.069	0.010	0.007	0.005
r43	0.012	0.012	0.003	0.010	0.010	0.005	0.016	0.042	0.049	0.028	0.008	0.024	0.074	0.016	0.056	0.029	0.053	0.009	0.016	0.056	0.017	0.033	0.073	0.064	0.068	0.043	0.060	0.021	0.005	0.005
r44	0.053	0.013	0.005	0.012	0.015	0.004	0.031	0.033	0.029	0.008	0.012	0.026	0.022	0.037	0.068	0.017	0.062	0.010	0.073	0.055	0.060	0.056	0.055	0.058	0.053	0.024	0.024	0.024	0.004	0.011
r45	0.002	0.036	0.040	0.010	0.005	0.004	0.008	0.013	0.012	0.007	0.013	0.007	0.020	0.044	0.061	0.014	0.012	0.013	0.009	0.005	0.004	0.014	0.028	0.038	0.065	0.052	0.028	0.044	0.052	0.046
r46	0.001	0.002	0.001	0.002	0.036	0.007	0.007	0.012	0.008	0.007	0.004	0.014	0.033	0.031	0.052	0.025	0.013	0.007	0.006	0.011	0.006	0.009	0.013	0.006	0.009	0.071	0.069	0.041	0.029	0.051
r47	0.001	0.000	0.002	0.001	0.053	0.003	0.010	0.012	0.011	0.003	0.008	0.013	0.008	0.013	0.043	0.072	0.004	0.013	0.008	0.006	0.015	0.010	0.010	0.013	0.073	0.050	0.026	0.029	0.049	0.054
r48	0.001	0.001	0.001	0.002	0.000	0.000	0.011	0.014	0.005	0.014	0.003	0.014	0.006	0.003	0.059	0.049	0.009	0.007	0.010	0.013	0.432	0.005	0.004	0.010	0.052	0.019	0.579	0.096	0.074	0.041
r49	0.002	0.002	0.002	0.001	0.002	0.002	0.000	0.004	0.011	0.014	0.010	0.006	0.010	0.007	0.008	0.075	0.008	0.012	0.010	0.048	0.108	0.013	0.009	0.007	0.043	0.039	0.819	0.286	0.061	0.044
r50	0.001	0.002	0.000	0.002	0.001	0.001	0.002	0.001	0.018	0.029	0.064	0.010	0.006	0.011	0.003	0.014	0.006	0.007	0.013	0.042	0.003	0.006	0.014	0.008	0.023	0.037	0.039	0.060	0.071	0.075

Step C: Select all the x samples where $POP_x > 1500$

Row	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	c11	c12	c13	c14	c15	c16	c17	c18	c19	c20	c21	c22	c23	c24	c25	c26	c27	c28	c29	c30
r 1																														
r 2																														
r 3																														
r 4																														
r 5																														
r 6																														
r 7																					1									
r 8																					1									
r 9																					1									
r10																					1									
r11																					1									
r12																					1		1							
r13			1																		1	1	1	1			1		1	1
r14		1	1	1																1	1	1	1	1		1	1	1	1	1
r15	1	1	1	1															1	1	1		1	1	1	1	1	1	1	
r16		1	1	1	1					1									1	1	1									
r17				1	1					1	1	1		1	1	1	1				1	1	1			1	1	1		
r18						1						1		1	1	1	1				1	1				1				
r19							1					1	1	1	1		1	1	1							1				
r20								1	1	1	1	1	1	1	1		1	1						1	1	1				
r21										1	1	1	1	1			1	1			1	1	1	1	1					
r22											1	1	1	1	1	1	1					1	1	1	1	1				
r23												1	1	1	1	1	1					1	1	1	1	1				
r24									1	1	1	1	1	1	1	1	1	1			1	1	1	1	1	1	1	1		
r25				1					1	1	1	1	1	1	1	1	1	1		1	1			1	1	1				
r26				1	1							1	1	1	1	1	1			1	1			1	1	1	1	1		
r27				1	1	1						1	1	1	1	1			1	1	1	1	1	1	1	1	1	1		
r28				1	1	1				1	1	1	1	1	1					1	1	1	1	1	1	1	1			
r29				1			1	1	1	1	1	1	1									1	1	1	1					
r30				1	1	1		1				1	1			1	1					1	1	1	1					
r31				1	1							1	1				1					1	1							
r32				1	1							1	1									1	1	1	1					
r33			1	1	1	1		1				1	1							1	1	1	1	1	1					
r34			1	1	1	1	1	1				1	1	1	1	1				1	1		1	1	1	1	1	1		
r35			1	1	1		1	1				1	1	1	1	1			1	1				1	1	1	1			
r36			1	1	1	1	1	1					1	1	1	1			1						1	1	1			
r37			1		1	1	1	1	1				1	1	1	1			1	1					1	1				
r38			1		1	1	1	1	1				1	1	1	1			1						1					
r39							1	1	1										1	1			1	1	1					
r40								1							1	1	1	1	1	1	1		1	1	1					
r41																					1		1		1					
r42																					1				1					
r43																														
r44																														
r45																														
r46																														
r47																														
r48																					1						1	1		
r49																					1						1	1		
r50																														

Step D: Select all the x samples where $\mathrm{BU_x}$ > 0.5

Row	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	c11	c12	c13	c14	c15	c16	c17	c18	c19	c20	c21	c22	c23	c24	c25	c26	c27	c28	c29	c30
r 1																														
r 2																1														
r 3																1														
r 4																														
r 5																														
r 6																														
r 7																														
r 8																														
r 9																														
r10																														
r11										1	1	1	1	1							1									
r12										1	1	1	1	1									1							
r13										1	1	1	1	1								1	1							
r14			1					1	1	1	1	1										1				1				
r15	1								1	1	1	1							1	1				1		1				
r16			1	1							1								1		1									
r17					1					1	1	1			1	1		1	1	1	1	1					1	1		
r18																1		1	1	1			1			1				
r19												1	1	1						1	1	1	1							
r20								1				1	1		1				1	1	1	1	1	1		1				
r21										1		1		1				1	1	1		1								
r22													1					1	1	1	1		1		1					
r23												1	1	1			1	1	1	1	1		1			1				
r24											1	1			1				1	1			1		1	1				
r25												1	1	1				1	1											
r26				1															1	1						1	1			
r27												1				1					1		1				1			
r28										1	1		1							1		1				1				
r29							1				1	1	1									1			1					
r30								1					1			1									1					
r31				1								1	1				1													
r32				1	1							1	1									1								
r33					1			1				1	1										1	1	1					
r34			1				1					1	1	1		1							1			1		1		
r35			1					1				1			1	1			1					1						
r36				1			1	1								1			1											
r37			1					1	1							1			1						1					
r38					1	1			1					1	1				1											
r39							1		1										1				1							
r40					1	1	1								1	1		1	1				1		1					
r41					1	1	1																		1					
r42					1	1	1														1									
r43																														
r44																														
r45																														
r46																														
r47																														
r48																											1			
r49																											1			
r50																														

Step E: The support set HDC_x^{supp} as union of the outputs at the steps C and D

Row	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	c11	c12	c13	c14	c15	c16	c17	c18	c19	c20	c21	c22	c23	c24	c25	c26	c27	c28	c29	c30
r 1																														
r 2																1														
r 3																1														
r 4																														
r 5																														
r 6																														
r 7																					1									
r 8																					1									
r 9																					1									
r10																					1									
r11										1	1	1	1	1							1									
r12										1	1	1	1	1							1		1							
r13			1							1	1	1	1	1							1	1	1	1			1		1	1
r14		1	1	1				1	1	1	1	1								1	1	1	1	1		1	1	1	1	1
r15	1	1	1	1					1	1	1	1							1	1	1		1	1	1	1	1	1	1	
r16		1	1	1	1					1	1								1	1	1									
r17				1	1					1	1	1		1	1	1	1	1	1	1	1	1	1			1	1	1		
r18						1						1		1	1	1	1	1	1	1	1	1	1			1				
r19							1					1	1	1	1		1	1	1	1	1	1	1			1				
r20								1	1	1	1	1	1	1	1		1	1	1	1	1	1	1	1	1	1				
r21										1	1	1	1	1			1	1	1	1	1	1	1	1	1					
r22											1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1				
r23												1	1	1	1	1	1	1	1	1	1	1	1	1	1	1				
r24									1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		
r25				1					1	1	1	1	1	1	1	1	1	1	1	1	1			1	1	1				
r26				1	1							1	1	1	1	1	1		1	1	1			1	1	1	1	1		
r27				1	1	1						1	1	1	1	1			1	1	1	1	1	1	1	1	1	1		
r28				1	1	1				1	1	1	1	1	1					1	1	1	1	1	1	1	1			
r29				1			1	1	1	1	1	1	1									1	1	1	1					
r30				1	1	1		1				1	1			1	1					1	1	1	1					
r31				1	1							1	1				1					1	1							
r32				1	1							1	1									1	1	1	1					
r33			1	1	1	1		1				1	1							1	1	1	1	1	1					
r34			1	1	1	1	1	1				1	1	1	1	1				1	1		1	1	1	1	1	1		
r35			1	1	1		1	1				1	1	1	1	1			1	1				1	1	1	1			
r36			1	1	1	1	1	1					1	1	1	1			1						1	1	1			
r37			1		1	1	1	1	1				1	1	1	1			1	1					1	1				
r38			1		1	1	1	1	1				1	1	1	1			1						1					
r39							1	1	1										1	1			1	1	1					
r40					1	1	1	1							1	1	1	1	1	1	1		1	1	1					
r41					1	1	1														1		1		1					
r42					1	1	1														1				1					
r43																														
r44																														
r45																														
r46																														
r47																														
r48																					1						1	1		
r49																					1						1	1		
r50																														

Step F: Identify the 4-conn clusters HDC_x^c from the of HDC_x^{supp} output at step E

Row	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	c11	c12	c13	c14	c15	c16	c17	c18	c19	c20	c21	c22	c23	c24	c25	c26	c27	c28	c29	c30
r 1																														
r 2																6														
r 3																6														
r 4																														
r 5																														
r 6																														
r 7																					5									
r 8																					5									
r 9																					5									
r10																					5									
r11										5	5	5	5	5							5									
r12										5	5	5	5	5							5		5							
r13			1							5	5	5	5	5							5	5	5	5			5		5	5
r14		1	1	1				5	5	5	5	5								5	5	5	5	5		5	5	5	5	5
r15	1	1	1	1					5	5	5	5							5	5	5		5	5	5	5	5	5	5	
r16		1	1	1	1					5	5								5	5	5									
r17				1	1					5	5	5		5	5	5	5	5	5	5	5	5	5			5	5	5		
r18						3						5		5	5	5	5	5	5	5	5	5	5			5				
r19							4					5	5	5	5		5	5	5	5	5	5	5			5				
r20								5	5	5	5	5	5	5	5		5	5	5	5	5	5	5	5	5	5				
r21										5	5	5	5	5			5	5	5	5	5	5	5	5	5					
r22											5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5				
r23												5	5	5	5	5	5	5	5	5	5	5	5	5	5	5				
r24									5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5		
r25				2					5	5	5	5	5	5	5	5	5	5	5	5	5			5	5	5				
r26				2	2							5	5	5	5	5	5		5	5	5			5	5	5	5	5		
r27				2	2	2						5	5	5	5	5			5	5	5	5	5	5	5	5	5	5		
r28				2	2	2				5	5	5	5	5	5					5	5	5	5	5	5	5	5			
r29				2			5	5	5	5	5	5	5									5	5	5	5					
r30				2	2	2		5				5	5			7	7					5	5	5	5					
r31				2	2							5	5				7					5	5							
r32				2	2							5	5									5	5	5	5					
r33			2	2	2	2		2				5	5							5	5	5	5	5	5					
r34			2	2	2	2	2	2				5	5	5	5	5				5	5		5	5	5	5	5	5		
r35			2	2	2		2	2				5	5	5	5	5			5	5				5	5	5	5			
r36			2	2	2	2	2	2					5	5	5	5			5						5	5	5			
r37			2		2	2	2	2	2				5	5	5	5			5	5					5	5				
r38			2		2	2	2	2	2				5	5	5	5			5						5					
r39							2	2	2										5	5			5	5	5					
r40					2	2	2	2							5	5	5	5	5	5	5		5	5	5					
r41					2	2	2														5		5		5					
r42					2	2	2														5				5					
r43																														
r44																														
r45																														
r46																														
r47																														
r48																					8						9	9		
r49																					8						9	9		
r50																														

Step G: Identify the clusters having more than 50000 population size

Row	Area	PopSize	PopSize >50000
cluster1	14	126211	TRUE
cluster2	64	457930	TRUE
cluster3	1	2036	FALSE
cluster4	1	4955	FALSE
cluster5	347	2536216	TRUE
cluster6	2	0	FALSE
cluster7	3	29988	FALSE
cluster8	2	10806	FALSE
cluster9	4	35621	FALSE

Annex 2. GHS-UCDB Cookbook

All attributes are produced for Urban Centre spatial domain, i.e., per each Urban Centre spatial domain, as delineated for epoch 2015. The attributes are categorised by dimensions.

Variables and attributes describing general characteristics of the Urban Centres.

Variable	Attribute	Column(s)	Metric	Source	Note	Temporal coverage			
						1975	1990	2000	2015
Control code	Unique ID	ID_HDC_G0		JRC					
	Quality Code	QA2_1V		JRC	Class (1)				
Extension	Area	AREA	km^2	Derived					
	Bounding Box (WGS 84)	BBX_LATMN, BBX_LONMN, BBX_LATMX, BBX_LONMX	°	Derived	Decimal Degrees				
Location	Geometric Centroid (WGS 84)	GCPNT_LAT, GCPNT_LON	°	Derived	Decimal Degrees				
	Main Country Identification: name	CTR_MN_NM		(GADM, 2018)					
	Main Country Identification: ISO 3	CTR_MN_ISO		(GADM, 2018)					
	Cross border flag	XBRDR		JRC	Boolean				
	Number of intersected countries	XCTR_NBR		JRC	Number of entities.				
	List of intersected countries: names	XC_NM_LST		(GADM, 2018)					
	List of intersected countries: ISO 3 codes	XC_ISO_LST		(GADM, 2018)					
	Major Geographical Region	GRGN_L1		UN WUP 2018	Class (2)				
	Geographical Region	GRGN_L2		UN WUP 2018	Class (3)				
Name	Name of the Urban Centre	UC_NM_MN		(UNDESA, 2018a, UNDESA, 2018b), (CIESIN et al. 2011, Balk et al., 2006), (Patterson & Kelso, 2018), (GeoNames, 2018), *other*[29]					
	List of names	UC_NM_LST							
	Source of the names	UC_NM_SRC			Class (4)				

[29] See the schema for more details.

Variables and attributes describing multitemporal Urban Centre domain.

Variable	Attribute	Column(s)	Metric	Source	Note	Temporal coverage 1975	1990	2000	2015
Number of Urban Centres in the past	Number of Urban Centres in 1975	H75_NBR		(Florczyk et al., 2018)	Number of entities				
	Number of Urban Centres in 1990	H90_NBR			Number of entities				
	Number of Urban Centres in 2000	H00_NBR			Number of entities				
Total area of Urban Centres in the past	Total area of Urban Centres in 1975	H75_AREA	km^2						
	Total area of Urban Centres in 1990	H90_AREA	km^2						
	Total area of Urban Centres in 2000	H00_AREA	km^2						

Variables and attributes describing geography of Urban Centres.

Variable	Attribute	Column(s)	Metric	Source	Note	Temporal coverage 1975	1990	2000	2015
Biome	Biome type(s)	E_BM_NM_LST		(Olson et al. 2001)	Class (5)				
Soil	Soil group(s)	E_SL_LST		(Fischer et al. 2008)	Class (6)				
Elevation	Average Elevation	EL_AV_ALS	m	(EORC & JAXA, 2017)	Meters above sea level (MASL)				
Climate	Climate class(es)	E_KG_NM_LST		(Rubel et al., 2017)	Class (7)	1986-2010			
River basin	Major river basin(s)	E_RB_NM_LST		(GRDC, 2007)	Class (8)				
Precipitation	Average precipitation for epoch 1990	E_WR_P_90	mm	(Harris et al., 2014)					
	Average precipitation for epoch 2000	E_WR_P_00	mm						
	Average precipitation for epoch 2014	E_WR_P_14	mm						
Temperature	Average temperature for epoch 1990	E_WR_T_90	°C						
	Average temperature for epoch 2000	E_WR_T_00	°C						
	Average temperature for epoch 2014	E_WR_T_14	°C						

Variables and attributes describing socio-economic characteristics of Urban Centres.

Variable	Attribute	Column(s)	Metric	Source	Note	Temporal coverage			
						1975	1990	2000	2015
Built-up surface	Total built-up area in 1975	B75	km^2	GHS_BUILT_LDS1975_GLOBE_R2018A_54009_1K_V_1_0 (Florczyk et al., 2018)					
	Total built-up area in 1990	B90	km^2	GHS_BUILT_LDS1990_GLOBE_R2018A_54009_1K_V_1_0 (Florczyk et al., 2018)					
	Total built-up area in 2000	B00	km^2	GHS_BUILT_LDS2000_GLOBE_R2018A_54009_1K_V_1_0 (Florczyk et al., 2018)					
	Total built-up area in 2015	B15	km^2	GHS_BUILT_LDS2015_GLOBE_R2018A_54009_1K_V_1_0 (Florczyk et al., 2018)					
Resident population	Total resident population in 1975	P75		GHS_POP_GPW41E1975_GLOBE_R2018A_54009_1K_V_1_0 (Florczyk et al., 2018)	Number of people				
	Total resident population in 1990	P90		GHS_POP_GPW41E1990_GLOBE_R2018A_54009_1K_V_1_0 (Florczyk et al., 2018)	Number of people				
	Total resident population in 2000	P00		GHS_POP_GPW41E2000_GLOBE_R2018A_54009_1K_V_1_0 (Florczyk et al., 2018)	Number of people				
	Total resident population in 2015	P15		GHS_POP_GPW41E2015_GLOBE_R2018A_54009_1K_V_1_0 (Florczyk et al., 2018)	Number of people				
Built-up per capita	Surface of the built-up area per person in 1975	BUCAP75	m^2 $person^{-1}$	Derived	Sq m per person				
	Surface of the built-up area per person in 1990	BUCAP90	m^2 $person^{-1}$	Derived	Sq m per person				
	Surface of the built-up area per person in 2000	BUCAP00	m^2 $person^{-1}$	Derived	Sq m per person				
	Surface of the built-up area per person in 2015	BUCAP15	m^2 $person^{-1}$	Derived	Sq m per person				

Night time light emission	Average night time light emission in 2015	NTL_AV	nW cm^{-2} sr^{-1}	(Weiss et al., 2018)	nano-watt per steradian per square centimetre				
Gross Domestic Product	Sum of GDP PPP values for year 1990	GDP90_SM	$	(Kummu et al., 2018)	USA dollar 2011				
	Sum of GDP PPP values for year 2000	GDP00_SM	$		USA dollar 2011				
	Sum of GDP PPP values for year 2015	GDP15_SM	$		USA dollar 2011				
Development Indicators	UN income class	INCM_CMI		(UNDESA, 2018b)	Class (9)				2018
	UN development group	DEV_CMI			Class (10)				2018
Accessibility & Remoteness	Travel time to country capital	TT2CC	’	(Weiss et al., 2018), JRC	minutes				

Variables and attributes describing environment of Urban Centres.

Variable	Attribute		Column(s)	Metric	Source	Note	Temporal coverage			
							1975	1990	2000	2015
Urban green	Greenness Estimate	Average greenness estimated for 1990 located in the built-up area of epoch 1990	E_GR_AV90		(Corbane et al., 2018b)	Index				
		Average greenness estimated for 2000 located in the built-up area of epoch 2000	E_GR_AV00			Index				
		Average greenness estimated for 2014 located in the built-up area of epoch 2014	E_GR_AV14			Index				
	Greenness class area	Total area of the high green estimated for 1990	E_GR_AH90	km^2	(Corbane et al., 2018b)					
		Total area of the medium green estimated for 1990	E_GR_AM90	km^2						
		Total area of the low green estimated for 1990	E_GR_AL90	km^2						
		Total area of green estimated for 1990	E_GR_AT90	km^2						
		Total area of the high green estimated for 2000	E_GR_AH00	km^2						
		Total area of the medium green estimated for 2000	E_GR_AM00	km^2						
		Total area of the low green estimated for 2000	E_GR_AL00	km^2						
		Total area of green estimated for 2000	E_GR_AT00	km^2						
		Total area of the high green estimated for 2014	E_GR_AH14	km^2						

		Total area of the medium green estimated for 2000	E_GR_AM14	km^2						
		Total area of the low green estimated for 2014	E_GR_AL14	km^2						
		Total area of green estimated for 2014	E_GR_AT14	km^2						
Emission of Pollutants	CO_2 (non-short-cycle-organic fuels)	Total emission of CO_2 from the energy sector, using non-short-cycle-organic fuels in 1975	E_EC2E_E75	t a^{-1}	(Crippa et al., 2018)	tonnes (10^3 kg) per year				
		Total emission of CO_2 from the energy sector, using non-short-cycle-organic fuels in 1990	E_EC2E_E90	t a^{-1}		tonnes (10^3 kg) per year				
		Total emission of CO_2 from the energy sector, using non-short-cycle-organic fuels in 2000	E_EC2E_E00	t a^{-1}		tonnes (10^3 kg) per year				
		Total emission of CO_2 from the energy sector, using non-short-cycle-organic fuels in 2012	E_EC2E_E12	t a^{-1}		tonnes (10^3 kg) per year				
		Total emission of CO_2 from the residential sector, using non-short-cycle-organic fuels in 1975	E_EC2E_R75	t a^{-1}		tonnes (10^3 kg) per year				
		Total emission of CO_2 from the residential sector, using non-short-cycle-organic fuels in 1990	E_EC2E_R90	t a^{-1}		tonnes (10^3 kg) per year				
		Total emission of CO_2 from the residential sector, using non-short-cycle-organic fuels in 2000	E_EC2E_R00	t a^{-1}		tonnes (10^3 kg) per year				
		Total emission of CO_2 from the residential sector, using non-short-cycle-organic fuels in 2012	E_EC2E_R12	t a^{-1}		tonnes (10^3 kg) per year				
		Total emission of CO_2 from the industry sector, using non-short-cycle-organic fuels in 1975	E_EC2E_I75	t a^{-1}		tonnes (10^3 kg) per year				
		Total emission of CO_2 from the industry sector, using non-short-cycle-organic fuels in 1990	E_EC2E_I90	t a^{-1}		tonnes (10^3 kg) per year				
		Total emission of CO_2 from the industry sector, using non-short-cycle-organic fuels in 2000	E_EC2E_I00	t a^{-1}		tonnes (10^3 kg) per year				
		Total emission of CO_2 from the industry sector, using non-short-cycle-organic fuels in 2012	E_EC2E_I12	t a^{-1}		tonnes (10^3 kg) per year				
		Total emission of CO_2 from the transport sector, using non-short-cycle-organic fuels in 1975	E_EC2E_T75	t a^{-1}		tonnes (10^3 kg) per year				

		Total emission of CO_2 from the transport sector, using non-short-cycle-organic fuels in 1990	E_EC2E_T90	t a^{-1}		tonnes (10^3 kg) per year				
		Total emission of CO_2 from the transport sector, using non-short-cycle-organic fuels in 2000	E_EC2E_T00	t a^{-1}		tonnes (10^3 kg) per year				
		Total emission of CO_2 from the transport sector, using non-short-cycle-organic fuels in 2012	E_EC2E_T12	t a^{-1}		tonnes (10^3 kg) per year				
		Total emission of CO_2 from the agriculture sector, using non-short-cycle-organic fuels in 1975	E_EC2E_A75	t a^{-1}		tonnes (10^3 kg) per year				
		Total emission of CO_2 from the agriculture sector, using non-short-cycle-organic fuels in 1990	E_EC2E_A90	t a^{-1}		tonnes (10^3 kg) per year				
		Total emission of CO_2 from the agriculture sector, using non-short-cycle-organic fuels in 2000	E_EC2E_A00	t a^{-1}		tonnes (10^3 kg) per year				
		Total emission of CO_2 from the agriculture sector, using non-short-cycle-organic fuels in 2012	E_EC2E_A12	t a^{-1}		tonnes (10^3 kg) per year				
	CO_2 (short-cycle-organic fuels)	Total emission of CO_2 from the energy sector, using short-cycle-organic fuels in 1975	E_EC2O_E75	t a^{-1}	(Crippa et al., 2018)	tonnes (10^3 kg) per year				
		Total emission of CO_2 from the energy sector, using short-cycle-organic fuels in 1990	E_EC2O_E90	t a^{-1}		tonnes (10^3 kg) per year				
		Total emission of CO_2 from the energy sector, using short-cycle-organic fuels in 2000	E_EC2O_E00	t a^{-1}		tonnes (10^3 kg) per year				
		Total emission of CO_2 from the energy sector, using short-cycle-organic fuels in 2012	E_EC2O_E12	t a^{-1}		tonnes (10^3 kg) per year				
		Total emission of CO_2 from the residential sector, using short-cycle-organic fuels in 1975	E_EC2O_R75	t a^{-1}		tonnes (10^3 kg) per year				
		Total emission of CO_2 from the residential sector, using short-cycle-organic fuels in 1990	E_EC2O_R90	t a^{-1}		tonnes (10^3 kg) per year				
		Total emission of CO_2 from the residential sector, using short-cycle-organic fuels in 2000	E_EC2O_R00	t a^{-1}		tonnes (10^3 kg) per year				
		Total emission of CO_2 from the residential sector, using short-cycle-organic fuels in 2012	E_EC2O_R12	t a^{-1}		tonnes (10^3 kg) per year				

	Total emission of CO_2 from the industry sector, using short-cycle-organic fuels in 1975	E_EC2O_I75	t a^{-1}		tonnes (10^3 kg) per year				
	Total emission of CO_2 from the industry sector, using short-cycle-organic fuels in 1990	E_EC2O_I90	t a^{-1}		tonnes (10^3 kg) per year				
	Total emission of CO_2 from the industry sector, using short-cycle-organic fuels in 2000	E_EC2O_I00	t a^{-1}		tonnes (10^3 kg) per year				
	Total emission of CO_2 from the industry sector, using short-cycle-organic fuels in 2012	E_EC2O_I12	t a^{-1}		tonnes (10^3 kg) per year				
	Total emission of CO_2 from the transport sector, using short-cycle-organic fuels in 1975	E_EC2O_T75	t a^{-1}		tonnes (10^3 kg) per year				
	Total emission of CO_2 from the transport sector, using short-cycle-organic fuels in 1990	E_EC2O_T90	t a^{-1}		tonnes (10^3 kg) per year				
	Total emission of CO_2 from the transport sector, using short-cycle-organic fuels in 2000	E_EC2O_T00	t a^{-1}		tonnes (10^3 kg) per year				
	Total emission of CO_2 from the transport sector, using short-cycle-organic fuels in 2012	E_EC2O_T12	t a^{-1}		tonnes (10^3 kg) per year				
	Total emission of CO_2 from the agriculture sector, using short-cycle-organic fuels in 1975	E_EC2O_A75	t a^{-1}		tonnes (10^3 kg) per year				
	Total emission of CO_2 from the agriculture sector, using short-cycle-organic fuels in 1990	E_EC2O_A90	t a^{-1}		tonnes (10^3 kg) per year				
	Total emission of CO_2 from the agriculture sector, using short-cycle-organic fuels in 2000	E_EC2O_A00	t a^{-1}		tonnes (10^3 kg) per year				
	Total emission of CO_2 from the agriculture sector, using short-cycle-organic fuels in 2012	E_EC2O_A12	t a^{-1}		tonnes (10^3 kg) per year				
$PM_{2.5}$	Total emission of $PM_{2.5}$ from the energy sector in 1975	E_EPM2_E75	t a^{-1}	(Crippa et al., 2018)	tonnes (10^3 kg) per year				
	Total emission of $PM_{2.5}$ from the energy sector in 1990	E_EPM2_E90	t a^{-1}		tonnes (10^3 kg) per year				
	Total emission of $PM_{2.5}$ from the energy sector in 2000	E_EPM2_E00	t a^{-1}		tonnes (10^3 kg) per year				
	Total emission of $PM_{2.5}$ from the energy sector in 2012	E_EPM2_E12	t a^{-1}		tonnes (10^3 kg) per year				
	Total emission of $PM_{2.5}$ from the	E_EPM2_R75	t a^{-1}		tonnes (10^3 kg)				

		residential sector in 1975				per year				
		Total emission of $PM_{2.5}$ from the residential sector in 1990	E_EPM2_R90	$t\,a^{-1}$		tonnes (10^3 kg) per year				
		Total emission of $PM_{2.5}$ from the residential sector in 2000	E_EPM2_R00	$t\,a^{-1}$		tonnes (10^3 kg) per year				
		Total emission of $PM_{2.5}$ from the residential sector in 2012	E_EPM2_R12	$t\,a^{-1}$		tonnes (10^3 kg) per year				
		Total emission of $PM_{2.5}$ from the industry sector in 1975	E_EPM2_I75	$t\,a^{-1}$		tonnes (10^3 kg) per year				
		Total emission of $PM_{2.5}$ from the industry sector in 1990	E_EPM2_I90	$t\,a^{-1}$		tonnes (10^3 kg) per year				
		Total emission of $PM_{2.5}$ from the industry sector in 2000	E_EPM2_I00	$t\,a^{-1}$		tonnes (10^3 kg) per year				
		Total emission of $PM_{2.5}$ from the industry sector in 2012	E_EPM2_I12	$t\,a^{-1}$		tonnes (10^3 kg) per year				
		Total emission of $PM_{2.5}$ from the transport sector in 1975	E_EPM2_T75	$t\,a^{-1}$		tonnes (10^3 kg) per year				
		Total emission of $PM_{2.5}$ from the transport sector in 1990	E_EPM2_T90	$t\,a^{-1}$		tonnes (10^3 kg) per year				
		Total emission of $PM_{2.5}$ from the transport sector in 2000	E_EPM2_T00	$t\,a^{-1}$		tonnes (10^3 kg) per year				
		Total emission of $PM_{2.5}$ from the transport sector in 2012	E_EPM2_T12	$t\,a^{-1}$		tonnes (10^3 kg) per year				
		Total emission of $PM2_{2.5}$ from the agriculture sector in 1975	E_EPM2_A75	$t\,a^{-1}$		tonnes (10^3 kg) per year				
		Total emission of $PM_{2.5}$ from the agriculture sector in 1990	E_EPM2_A90	$t\,a^{-1}$		tonnes (10^3 kg) per year				
		Total emission of $PM_{2.5}$ from the agriculture sector in 2000	E_EPM2_A00	$t\,a^{-1}$		tonnes (10^3 kg) per year				
		Total emission of $PM_{2.5}$ from the agriculture sector in 2012	E_EPM2_A12	$t\,a^{-1}$		tonnes (10^3 kg) per year				
Concertation of Pollutants	$PM_{2.5}$	Total concertation of $PM_{2.5}$ for reference epoch 2000	E_CPM2_T00	$\mu g\,m^{-3}$	GBD 2017, OECD[30,31]	micrograms per cubic meter air			2000 2005 2010 2014	
		Total concertation of $PM_{2.5}$ for reference epoch 2005	E_CPM2_T05	$\mu g\,m^{-3}$		micrograms per cubic meter air				
		Total concertation of $PM_{2.5}$ for reference epoch 2010	E_CPM2_T10	$\mu g\,m^{-3}$		micrograms per cubic meter air				
		Total concertation of $PM_{2.5}$ for reference epoch 2014	E_CPM2_T14	$\mu g\,m^{-3}$		micrograms per cubic meter air				

[30] https://www.stateofglobalair.org/data/methods

[31] http://stats.oecd.org/wbos/fileview2.aspx?IDFile=28707511-43bc-4a03-a6bb-b3a32eea4f8b

Variables and attributes describing hazards and exposure of Urban Centres (DRR).

Variable	Attribute	Column(s)	Metric	Source	Note	Temporal coverage 1975	1990	2000	2015
Flood exposure	Total surface potentially exposed to floods	EX_FD_AREA	km²	(Dottori et al., 2016a, Dottori et al., 2016b)					
	Total built-up area potentially exposed to floods in 1975	EX_FD_B75	km²	GHS_BUILT_LDS1975 _GLOBE_R2018A _54009_1K_V_1_0 (Florczyk et al., 2018)					
	Total built-up area potentially exposed to floods in 1990	EX_FD_B90	km²	GHS_BUILT_LDS1990 _GLOBE_R2018A _54009_1K_V_1_0 (Florczyk et al., 2018)					
	Total built-up area potentially exposed to floods in 2000	EX_FD_B00	km²	GHS_BUILT_LDS2000 _GLOBE_R2018A _54009_1K_V_1_0 (Florczyk et al., 2018)					
	Total built-up area potentially exposed to floods in 2015	EX_FD_B15	km²	GHS_BUILT_LDS2015 _GLOBE_R2018A _54009_1K_V_1_0 (Florczyk et al., 2018)					
	Total resident population potentially exposed to floods in 1975	EX_FD_P75		GHS_POP_GPW41E1975 _GLOBE_R2018A _54009_1K_V_1_0 (Florczyk et al., 2018)	Number of people				
	Total resident population potentially exposed to floods in 1990	EX_FD_P90		GHS_POP_GPW41E1990 _GLOBE_R2018A _54009_1K_V_1_0 (Florczyk et al., 2018)	Number of people				
	Total resident population potentially exposed to floods in 2000	EX_FD_P00		GHS_POP_GPW41E2000 _GLOBE_R2018A _54009_1K_V_1_0 (Florczyk et al., 2018)	Number of people				
	Total resident population potentially exposed to floods in 2015	EX_FD_P15		GHS_POP_GPW41E2015 _GLOBE_R2018A _54009_1K_V_1_0 (Florczyk et al., 2018)	Number of people				
Storm surge exposure	Total surface potentially exposed to storm surges	EX_SS_AREA	km²	JRC					
	Total built-up area potentially exposed to storm surges in 1975	EX_SS_B75	km²	GHS_BUILT_LDS1975 _GLOBE_R2018A _54009_1K_V_1_0 (Florczyk et al., 2018)					
	Total built-up area potentially exposed to	EX_SS_B90	km²	GHS_BUILT_LDS1990					

	storm surges in 1990			_GLOBE_R2018A _54009_1K_V_1_0 (Florczyk et al., 2018)					
	Total built-up area potentially exposed to storm surges in 2000	EX_SS_B00	km^2	GHS_BUILT_LDS2000 _GLOBE_R2018A _54009_1K_V_1_0 (Florczyk et al., 2018)					
	Total built-up area potentially exposed to storm surges in 2015	EX_SS_B15	km^2	GHS_BUILT_LDS2015 _GLOBE_R2018A _54009_1K_V_1_0 (Florczyk et al., 2018)					
	Total resident population potentially exposed to storm surges in 1975	EX_SS_P75		GHS_POP_GPW41E1975 _GLOBE_R2018A _54009_1K_V_1_0 (Florczyk et al., 2018)	Number of people				
	Total resident population potentially exposed to storm surges in 1990	EX_SS_P90		GHS_POP_GPW41E1990 _GLOBE_R2018A _54009_1K_V_1_0 (Florczyk et al., 2018)	Number of people				
	Total resident population potentially exposed to storm surges in 2000	EX_SS_P00		GHS_POP_GPW41E2000 _GLOBE_R2018A _54009_1K_V_1_0 (Florczyk et al., 2018)	Number of people				
	Total resident population potentially exposed to storm surges in 2015	EX_SS_P15		GHS_POP_GPW41E2015 _GLOBE_R2018A _54009_1K_V_1_0 (Florczyk et al., 2018)	Number of people				
Earthquake	Average peak ground acceleration (PGA) estimate of the seismic risk	EX_EQ19PGA	g	Pre-release of GEM dataset (Pagani et al., 2018), available in August 2018, missing some areas in Asia and California	Acceleration (g)				
	MMI class of the seismic risk, derived from the PGA estimate	EX_EQ19MMI			Class (11)				
	Quality control value (available, missing – value not available, imprecise – not reliable estimate)	EX_EQ19_Q			Class (12)				
Heatwave	Maximum magnitude of the heatwaves	EX_HW_IDX		JRC	Index	1980-2010			

Variables and attributes estimating SDGs of Urban Centres.

Variable	Attribute	Column(s)	Metric	Source	Note	Temporal coverage			
						1975	1990	2000	2015
Land Use Efficiency (11.3.1)	Land use efficiency 1990-2015	SDG_LUE9015		(Melchiorri et al., 2018)	Dimensionless	1990-2015			
Open spaces (11.7.1 – proxy)	Share of population living in the high green area in 2015	SDG_A2G14	share	JRC					
	Percentage of the open spaces	SDG_OS15MX	%	JRC					

(1) Schema of Quality Code attribute: 0 – invalid; 1 – valid; 2 – uncertain.

(2) Major Geographical Regions (UN): Africa; Asia; Europe; Latin America and the Caribbean; Northern America; Oceania; *Other* – not classified

(3) Geographical Regions (UN): Australia/New Zealand; Caribbean; Central America; Central Asia; Eastern Africa; Eastern Asia; Eastern Europe; Melanesia; Micronesia; Middle Africa; Northern Africa; Northern America; Northern Europe; Polynesia; South America; South-Central Asia; South-Eastern Asia; Southern Africa; Southern Asia; Southern Europe; Western Africa; Western Asia; Western Europe; *Other* – not classified

(4) Name(s) source: WUP – World's Cities in 2018 form the World Urban Prospect 2018; GRUMP – Global Rural-Urban Mapping Project (GRUMP), v1, Settlement Points, v1 (1990, 1995, 2000); NE - Natural Earth Populated Places, v4.1.0; GN – GeoNames Gazetteer, WM – online mapping services (Bing Maps, Google Maps, OpenStreetMaps); OTHER – web user feedback and other manual revisions.

(5) Biome type: Tropical and Subtropical Dry Broadleaf Forests; Mediterranean Forests, Woodlands, and Scrub; Temperate Grasslands, Savannas, and Shrublands; Deserts and Xeric Shrublands; Temperate Coniferous Forests; Tropical and Subtropical Coniferous Forests; Temperate Broadleaf and Mixed Forests; Boreal Forests/Taiga; Tropical and Subtropical Moist Broadleaf Forests; Tropical and subtropical grasslands, savannas, and shrublands; Flooded Grasslands and Savannas; Montane Grasslands and Shrubland; Mangroves; Tundra.

(6) Soil group (*other*): Acrisols; Alisols; Andosols; Anthrosols; Arenosols; Calcisols; Cambisols; Chernozems; Ferralsols; Fluvisols; Gleysols; Greyzems; Gypsisols; Histosols; Kastanozems; Leptosols; Leptosols; Lixisols; Luvisols; Nitisols; Phaeozems; Planosols; Plinthosols; Podzols; Podzoluvisols; Regosols; Solonchaks; Solonetz; Vertisols; *Rock Outcrop; Sand Dunes; Water Bodies.*

(7) Climate Classes: Desert (arid), and Cold arid; Desert (arid), and Hot arid; Mild temperate with dry summer, and Hot summer; Mild temperate with dry summer, and Warm summer; Mild temperate with dry winter, and Hot summer; Mild temperate with dry winter, and Warm summer; Mild temperate, fully humid, and Cool summer; Mild temperate, fully humid, and Hot summer; Mild temperate, fully humid, and Warm summer; Snow with dry summer, and Cool summer; Snow with dry summer, and Hot summer; Snow with dry summer, and Warm summer; Snow with dry winter, and Cool summer; Snow with dry winter, and Hot summer; Snow with dry winter, and Warm summer; Snow, fully humid, and Cool summer; Snow, fully humid, and Hot summer; Snow, fully humid, and Warm summer; Steppe (semi-arid), and Cold arid; Steppe (semi-arid), and Hot arid; Tropical monsoon; Tropical rain forest; Tropical savannah with dry summer; Tropical savannah with dry winter; Tundra.

(8) Major River Basins: Alabama River & Tombigbee; Amazonas; Amur; Apalachicola River; Aral Drainage; Armeria; Atrato; Balkhash; Bandama; Batang Hari; Batang Kuantan; Bei Jiang; Biobio; Brahmani River (Bhahmani); Brahmaputra; Brantas; Bravo; Brazos River; Buzi; Ca; Cape Fear River; Cauvery River; Cavally; Chao Phraya; Chelif; Chira; Chubut; Coco; Colorado (Argentinia); Colorado River (Caribbean Sea); Colorado River (Pacific Ocean); Columbia River; Comoe; Conception; Congo; Connecticut River; Cross; Cuanza; Cunene; Dalinghe; Damodar River; Danube; Daryacheh-Ye Orumieh; Daule & Vinces; Davo; Dead Sea; Delaware River; Dniepr; Dniestr; Don; Dong Jiang; Douro; Ebro; Elbe River; Escaut (Schelde); Esmeraldas; Fraser River; Fuchun Jiang; Fuerte; Galana; Gambia; Gamka; Ganges; Garonne; Geba; Gloma; Godavari; Grande De Matagalpa; Great Salt Lake; Grisalva; Groot- Kei; Groot-Vis; Guadalquivir; Guadiana; Han Jiang; Han-Gang (Han River); Hong(Red River); Huai He;

Huang He (Yellow River); Hudson River; Incomati; Indus; Irrawaddy; Ishikari; Issyk-Kul; James River; Kelantan; Kiso; Kitakami; Kizilirmak; Kokemaenjoki; Kouilo; Krishna; Kuban; Kura; Kymijoki; Lake Chad; Lake Mar Chiquita; Lake Titicaca; Lake Turkana; Lake Vattern; Lempa; Lena; Liao He; Limari; Limpopo; Loa; Loire; Luan He; Lurio; Mae Klong; Magdalena; Mahanadi River (Mahahadi); Mahi River; Mamberamo; Maputo; Mekong; Merrimack River; Messalo; Min Jiang; Mira; Mississippi River; Mogami; Mono; Mucuri; Murray; Naktong; Narmada; Narva; Negro (Argentinia); Negro (Uruguay); Nelson River; Neman; Neva; Niger; Nile; Northern Dvina(Severnaya Dvina); Ntem; Nyong; Ob; Oder River; Ogooue; Okavango; Orange; Orinoco; Oueme; Oulujoki; Pangani; Panuco; Papaloapan; Paraiba Do Sul; Parana; Patacua; Pee Dee River; Penner River; Po; Potomac River; Pra; Pur; Purari; Pyasina; Rajang; Rapel; Rhine; Rhone; Rio Acara; Rio Capim; Rio De Contas; Rio Doce; Rio Gurupi; Rio Itapecur; Rio Jacui; Rio Jaguaribe; Rio Mearim; Rio Paraguac; Rio Paraiba; Rio Parnaiba; Rio Prado; Rio Ribeira Do Iguape; Roanoke River; Rogue River; Rovuma; Rufiji; Ruv; Sabine River; Sacramento River; Sakarya; Salado; Salinas; Salween; San Antonio River; San Joaquin River; San Juan; San Pedro; Sanaga; Santa; Santee River; Santiago; Sao Francisco; Sassandra; Savannah River; Save; Sebo; Seine; Senegal; Sepik; Shebelle; Shinano, Chikuma; Sittang River; Solo (Bengawan Solo); Southern Bug; St.Johns River; St.Lawrence; Sungai Kajan; Sungai Kapuas; Sungai Mahakam; Suriname; Susquehanna River; Tana; Tano; Tapti River; Tarim; Tejo; Tenry; Thames; Tigris & Euphrates; Tocantins; Tone; Tranh (Nr Thu Bon); Trent; Trinity River (Texas); Tsiribihina; Tugela; Tuloma; Ulua; Ural; Uruguay; Uwimb; Vaenern-Goeta; Van Gol; Verde; Volga; Volta; Vuoksi; Waikato River; Weser; Western Dvina (Daugava); Wisla; Xi Jiang; Yangtze River (Chang Jiang); Yaqui; Yenisei; Yodo; Yongding He; Zambezi.

(9) Income classes: HIC - High Income Countries HIC; UMIC - Upper-middle Income Countries; LMIC - Lower-middle Income Countries; LIC - Low Income Countries.

(10) Development Groups: MDR - More Developed Regions; LCD - Less developed regions, excluding least developed countries; LDCL - Least developed countries.

(11) MMI classes: from 1 to 8.

(12) Quality control of the earthquake data: available; missing – value not available; imprecise – not reliable estimate

KJ-02-19-103-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

Joint Research Centre

EU Science Hub

Publications Office

doi:10.2760/037310
ISBN 978-92-79-99753-2