



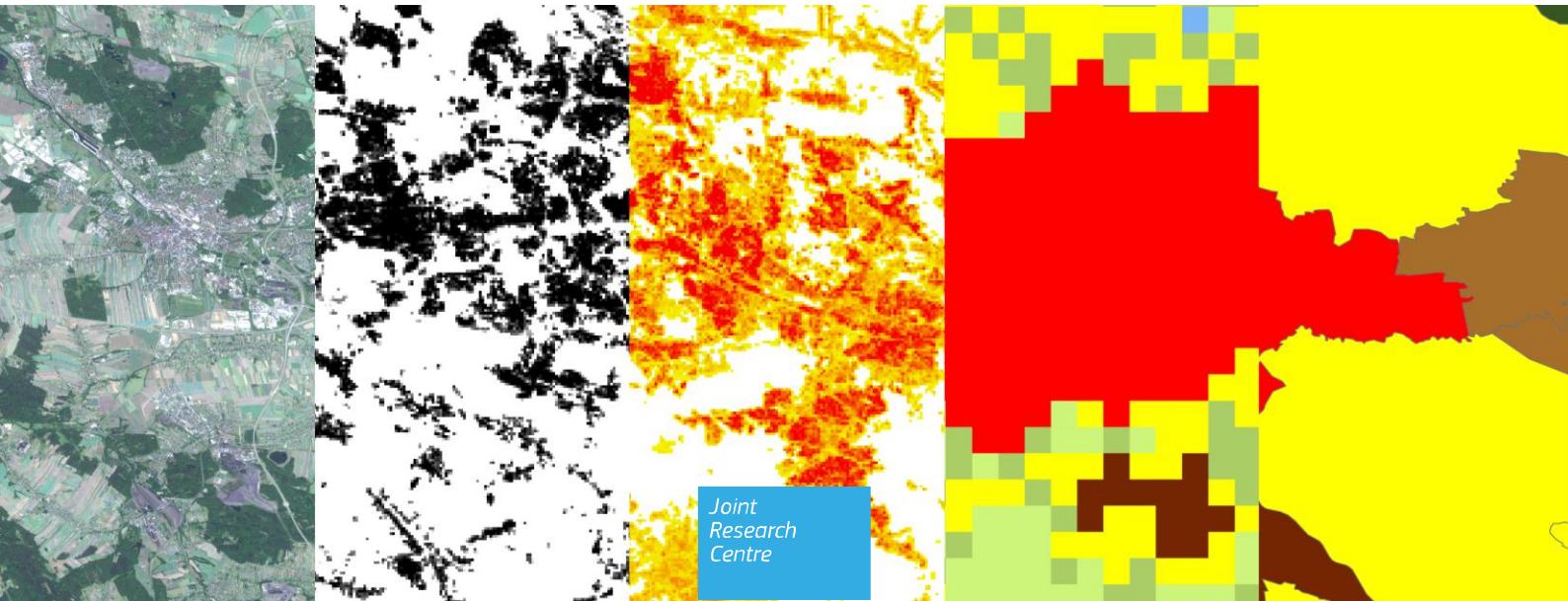
JRC SCIENTIFIC INFORMATION SYSTEMS AND DATABASES REPORT

GHS Data Package 2022

*Public release
GHS P2022*

Schiavina M., Melchiorri M., Pesaresi M., Politis P., Freire S., Maffenini L., Florio P., Ehrlich D., Goch K., Tommasi P., Kemper T.

2022



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. For information on the methodology and quality underlying the data used in this publication for which the source is neither Eurostat nor other Commission services, users should contact the referenced source. The designations employed and the presentation of material on the maps do not imply the expression of any opinion whatsoever on the part of the European Union concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

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@ec.europa.eu

Tel.: +39 0332 78 5576

GHSL project: JRC-GHSL@ec.europa.eu

GHSL Data: JRC-GHSL-DATA@ec.europa.eu

EU Science Hub

<https://ec.europa.eu/jrc>

JRC 129516

PDF ISBN 978-92-76-53071-8 doi:10.2760/19817

Print ISBN 978-92-76-53070-1 doi:10.2760/526478

Luxembourg: Publications Office of the European Union, 2022

© European Union, 2022



The reuse policy of the European Commission is implemented by the Commission Decision 2011/833/EU of 12 December 2011 on the reuse of Commission documents (OJ L 330, 14.12.2011, p. 39). Except otherwise noted, the reuse of this document is authorised under the Creative Commons Attribution 4.0 International (CC BY 4.0) licence (<https://creativecommons.org/licenses/by/4.0/>). This means that reuse is allowed provided appropriate credit is given and any changes are indicated. For any use or reproduction of photos or other material that is not owned by the EU, permission must be sought directly from the copyright holders.

All content © European Union, 2022

How to cite this report: Schiavina M., Melchiorri M., Pesaresi M., Politis P., Freire S., Maffenini L., Florio P., Ehrlich D., Goch K., Tommasi P., Kemper T., GHSL Data Package 2022, Publications Office of the European Union, Luxembourg, 2022, ISBN 978-92-76-53071-8, doi:10.2760/19817, JRC 129516

Contents

Authors.....	1
Abstract	2
1 Introduction.....	3
1.1 Overview.....	3
1.2 Rationale	4
1.3 History and Versioning	4
1.4 Main Characteristics	5
2 Products.....	7
2.1 GHS-BUILT-S R2022A - GHS built-up surface grid, derived from Sentinel-2 composite and Landsat, multitemporal (1975-2030).....	7
2.1.1 Definitions	8
2.1.2 Expected Errors	13
2.1.3 Improvements compared to the previous release	16
2.1.4 Input Data.....	18
2.1.5 Technical Details	19
2.1.6 Summary statistics.....	20
2.1.7 How to cite	21
2.2 GHS-BUILT-H R2022A - GHS building height, derived from AW3D30, SRTM30, and Sentinel-2 composite (2018).....	23
2.2.1 Definitions	23
2.2.2 Input data	24
2.2.3 Expected errors	27
2.2.4 Technical Details	28
2.2.5 How to cite	28
2.3 GHS-BUILT-V R2022A - GHS built-up volume grids derived from joint assessment of Sentinel-2, Landsat, and global DEM data, for 1975-2030 (5yrs interval)	30
2.3.1 Input Data.....	30
2.3.2 Technical Details	30
2.3.3 Summary statistics.....	33
2.3.4 How to cite	34
2.4 GHS-BUILT-C R2022A - GHS Settlement Characteristics, derived from Sentinel-2 composite (2018) and other GHS R2022A data.....	35
2.4.1 Input data	42
2.4.2 Technical Details	42
2.4.3 How to cite	43
2.5 GHS-LAND R2022A - GHS land surface grid, derived from Sentinel-2 composite (2018), Landsat Panchromatic data (2014), and OSM data	44
2.5.1 Input data	45

2.5.2	Technical Details	45
2.5.3	How to cite	45
2.6	GHS-POP R2022A - GHS population grid multitemporal (1975-2030)	46
2.6.1	Improvements compared to the previous release	47
2.6.2	Input Data.....	48
2.6.3	Technical Details	49
2.6.4	Summary statistics.....	49
2.6.5	How to cite	50
2.7	GHS-SMOD R2022A - GHS settlement layers, application of the Degree of Urbanisation methodology (stage I) to GHS-POP R2022A and GHS-BUILT-S R2022A, multitemporal (1975-2030).....	51
2.7.1	Improvements compared to the previous release	52
2.7.2	GHS-SMOD classification rules	52
2.7.3	GHS-SMOD spatial entities naming.....	55
2.7.4	GHS-SMOD L2 grid and L1 aggregation	55
2.7.5	Input Data.....	59
2.7.6	Technical Details	59
2.7.6.1	GHS-SMOD raster grid.....	59
2.7.6.2	GHS-SMOD urban centre entities.....	59
2.7.6.3	GHS-SMOD dense urban cluster entities.....	60
2.7.6.4	GHS-SMOD semi-dense urban cluster entities.....	60
2.7.7	Summary statistics.....	61
2.7.8	How to cite	63
2.8	GHS-DUC R2022A - GHS Degree of Urbanisation Classification, application of the Degree of Urbanisation methodology (stage II) to GADM 3.6 layer, multitemporal (1975-2030)	64
2.8.1	Improvements compared to previous release	65
2.8.2	GHSL Territorial Units Classification.....	65
2.8.2.1	Territorial units classification Level 1	65
2.8.2.2	Territorial units classification Level 2	66
2.8.2.3	Classification workflow	67
2.9	A consistent nomenclature for the Degree of Urbanisation.....	67
2.9.1	How to use the statistics tables	68
2.9.2	Input Data.....	68
2.9.3	Technical Details	69
2.9.3.1	GHS-DUC Summary Statistics Table.....	69
2.9.3.2	GHS-DUC Admin Classification layers.....	70
2.9.4	How to cite	76
2.10	GHS-BUILT-LAU2STAT R2022A - GHS built-up surface statistics in European LAU, multitemporal (1975-2020)	77
2.10.1	Input data	77
2.10.2	Technical Details	77

2.10.3 How to cite	77
2.11 GHS-SDATA R2022A - GHSL data supporting the production of R2022A Data Package (GHS-BUILT, GHS-POP).....	78
2.11.1 Technical Details.....	78
2.11.2 How to cite	80
References.....	81
List of figures	83
List of tables	85

Authors

Schiavina M., Melchiorri M., Pesaresi M., Politis P., Freire S., Maffenini L., Florio P., Ehrlich D., Goch K., Tommasi P., Kemper T.

Abstract

The Global Human Settlement Layer (GHSL) produces new global spatial information, evidence-based analytics and knowledge describing the human presence on planet Earth. It operates in a fully open and free data and methods access policy. The knowledge generated with the GHSL is supporting the definition, the public discussion and the implementation of European policies and the monitoring of international frameworks such as the 2030 Development Agenda. The GHSL are the core data set of the Exposure Mapping Component under the Copernicus Emergency Management Service. This release is the first official contribution of GHSL to the Copernicus services. GHSL data continue to support the GEO Human Planet Initiative (HPI) that is committed to developing a new generation of measurements and information products providing new scientific evidence and a comprehensive understanding of the human presence on the planet and that can support global policy processes with agreed, actionable and goal-driven metrics. The Human Planet Initiative relies on a core set of partners committed in coordinating the production of the global settlement spatial baseline data.

This document describes the public release of the GHSL Data Package 2022 (GHS P2022). The release provides improved built-up (including surface, volume and height) and population products as well as a new settlement model and classification of administrative and territorial units according to the Degree of Urbanisation.

Prior to cite this report, please access the updated version available at:

http://ghsl.jrc.ec.europa.eu/documents/GHSL_Data_Package_2022.pdf

1 Introduction

1.1 Overview

The Global Human Settlement Layer (GHSL) project produces global spatial information, evidence-based analytics, and knowledge describing the human presence on the planet. The GHSL relies on the design and implementation of new spatial data mining technologies that allow automatic processing, data analytics and knowledge extraction from large amounts of heterogeneous data including global, fine-scale satellite image data streams, census data, and crowd sourced or volunteered geographic information sources.

This document accompanies the public release of the GHSL Data Package 2022 (GHS P2022) and describes its contents.

Each product is named according to the following convention:

GHS-<name>_<spatial extent>_<release>

For example, a product name GHS-BUILT-V_GLOBE_R2022A indicates the built-up volume (BUILT-V) produced globally in the release R2022A.

Each product can be made by one or more datasets and layers. A layer is named with a unique identifier according to the following convention:

GHS-<name>_<Epoch&Model>_<spatialExtent>_<release>_<projection>_<resolution>_<version>

For example, a layer name GHS_BUILT_V_P2030MED_GLOBE_R2022A_54009_100_V1_0 indicates the built-up volume (GHS_BUILT_V) in the predicted future epoch 2030 under the median scenario (P2030MED), included in the release R2022A, in World Mollweide projection (ESRI:54009) at 100m of spatial resolution, version 1.0

The GHSL Data Package 2022 contains the following products:

GHS-BUILT-S R2022A - GHS built-up surface grid, derived from Sentinel-2 composite (2018) and Landsat, multitemporal (1975-2030)

GHS-BUILT-H R2022A - GHS building height, derived from AW3D30, SRTM30, and Sentinel-2 composite (2018)

GHS-BUILT-V R2022A - GHS built-up volume grids derived from joint assessment of Sentinel-2, Landsat, and global DEM data, for 1975-2030 (5 years interval)

GHS-BUILT-C R2022A - GHS Settlement Characteristics, derived from Sentinel-2 composite (2018) and other GHS R2022A data

GHS-LAND R2022A - Land fraction as derived from Sentinel-2 image composite (2018) and OSM data

GHS-POP R2022A - GHS population grid multitemporal (1975-2030)

GHS-SMOD R2022A - GHS settlement layers, application of the Degree of Urbanisation methodology (stage I) to GHS-POP R2022A and GHS-BUILT-S R2022A, multitemporal (1975-2030)

GHS-DUC R2022A - GHS Degree of Urbanisation Classification, application of the Degree of Urbanisation methodology (stage II) to GADM 3.6 layer, multitemporal (1975-2030)

GHS-SDATA R2022A - GHS release R2022A supporting data

GHS-BUILT-LAU2STAT R2022A - GHS built-up surface statistics in European LAU2, multitemporal (1975-2020)

1.2 Rationale

Open data and free access are core principles of GHSL (Melchiorri et al., 2019). They are aligned with the Directive on the re-use of public sector information (Directive 2003/98/EC¹). The free and open access policy facilitates the information sharing and collective knowledge building, thus contributing to a democratisation of the information production.

The GHSL Data Package 2022 contains the new GHSL data produced at the European Commission Directorate General Joint Research Centre in the Directorate for Space, Security and Migration in the Disaster Risk Management Unit (E.1) in the period 2020-2022.

1.3 History and Versioning

Previous GHSL releases relied on the processing of Landsat imagery for producing the GHS-BUILT information layer. The Landsat satellite platform collects Earth observation data since the beginning of the civilian space programs in the 1970s. In January 2008 Barbara Ryan, the Associate Director for Geography at the U.S. Geological Survey, and Michael Freilich, NASA's Director of the Earth Science Division, signed off a Landsat Data Distribution Policy that made Landsat images free to the public. The USGS announced the free-and-open data policy on April 21, 2008. The Global Land Survey (GLS) data sets were created as a collaboration between NASA and the USGS from 2009 through 2011. GLS datasets allowed scientists and data users to have access to a consistent, terrain corrected, coordinated collection of data. Each of these collections were created using the primary Landsat sensor in use at the time for each collection epoch. Early global experiments on the GHS-BUILT production by the European Commission's Joint Research Centre (JRC) date back to 2014, using in input GLS1975, GLS1990, GLS2000, and a collection of Landsat8 imagery of the year 2014 autonomously selected and downloaded by the JRC from the USGS portal. These data constitute the first set of data evidences used to support the GHSL epochs 1975, 1990, 2000, and 2014 (Pesaresi, Ehrlich, et al., 2016).

Copernicus, previously known as GMES (Global Monitoring for Environment and Security), is the European Union's Earth observation programme. It relies as well on a free-and-open data access policy. Sentinel-1 is the first of the Copernicus Programme satellite constellation using active radar sensor technology. The first satellite, Sentinel-1A, was launched on 3 April 2014, and Sentinel-1B was launched on 25 April 2016. In December 2016 the JRC successfully completed the first experiment of Sentinel-1 global data processing in the frame of the Global Human Settlement Layer (GHSL) project (Corbane et al., 2017). Sentinel-2 (S2) is an Earth observation mission from the Copernicus Programme that systematically acquires optical imagery at high spatial resolution (10 m to 60 m) over land and coastal waters. The launch of the first satellite, Sentinel-2A, occurred 23 June 2015. Sentinel-2B was launched on 7 March 2017. In 2018, the JRC produced the first Sentinel-2 cloud-free global pixel based image composite from L1C data for the period 2017-2018 for public use, leveraging on the Google Earth Engine platform (Corbane, Politis, et al., 2020). In October 2020, the JRC successfully completed the first public release of global built-up areas assessment from these Sentinel-2, 10m-resolution Copernicus imagery data (Corbane, Syrris, et al., 2020).

In 2016, the first public GHSL Data Package was released (GHS P2016). It consisted in several multi-temporal and multi-resolution products, including built-up area grids (GHS-BUILT), population grids (GHS-POP), settlement model (GHS-SMOD) and selected quality grids. The first GHS-BUILT product included in this release was generated from Landsat image data using the URBAN class generated from MODIS 500m-resolution data as learning set (Pesaresi, Ehrlich, et al., 2016), (Schneider et al., 2010). The population grids (GHS_POP_GPW41MT_GLOBE_R2016A) were produced in collaboration with Columbia University, Center for International Earth Science Information Network (CIESIN) in 2015. The GHS-SMOD grids (GHS_SMOD_POP_GLOBE_R2016A) present an implementation of the Degree of Urbanization (DEGURBA) model using as input the population grid cells.

In 2018, the second multi-temporal GHS-BUILT issue was released (GHS R2018) re-processing the same Landsat images, but with an improved learning set obtained by the introduction of the built-up areas collected from the classification of Sentinel-1 radar imagery at 20m-resolution (Corbane et al., 2017). In the GHS P2019 release downstream GHS-POP and GHS-SMOD data were distributed.

In 2020 a new single-epoch GHS-BUILT data was released (GHS R2020) by processing of 10m-resolution Sentinel-2 imagery data of year 2018 and by introducing as improvement of the learning set new data on

¹ <http://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX:32003L0098>

building footprints made available from open efforts of Microsoft and Facebook on classification of VHR image data (Corbane, Syrris, et al., 2020).

This new release GHS P2022 builds upon the past experience and data with two main objectives: A) augment the thematic contents of the built-up areas information; and B) reconnect and rebuild the historical series of the GHSL information rooting on the Landsat imagery with the new data coming from the Sentinel mission at higher spatial resolution. Several innovative aspects are included in the new GHS-BUILT of the GHS P2022, they are summarized below:

1. a new sub-pixel built-up surface fraction estimates at 10m-resolution
2. a new Boolean classification of the built-up surfaces in residential (RES) vs. non-residential (NRES) semantic abstractions at 10m-resolution
3. an average building height estimates at 100m-resolution
4. a spatio-temporal interpolation of the built-up surface information at 100m-resolution with equal time (5-years) intervals from 1975 to 2030. Future epochs 2025 and 2030 are predicted from past trends by application of linear (LIN), second-order-polynomial (PLY), and median (MED) scenarios.

The new GHS-POP layers also include several improvements deriving both from the new information available in built-up surface and from the use of a new population temporal model:

1. spatial resolution of 100m at 5-year intervals between 1975 and 2030.
2. allocation of population according to the presence of residential (RES) and non-residential (NRES) built-up
3. country total population time series aligned to the latest UN World Population Prospects 2019 (United Nations, Department of Economic and Social Affairs, Population Division, 2019)
4. temporal population estimates anchored to the UN “urban agglomeration” population time series of the latest UN World Urbanization Prospects 2018 (United Nations, Department of Economic and Social Affairs, Population Division, 2018)

The settlement classification layer (GHS-SMOD) benefits from improvements in built-up surface and population grid and it is based on the final specifications of the Degree of Urbanisation (stage I) (European Commission & Statistical Office of the European Union, 2021).

This new GHS-SMOD layer and the new GHS-POP grids are used to update the GADM 3.6 territorial units classification (GHS-DUC) according to the stage II of the Degree of Urbanisation (European Commission & Statistical Office of the European Union, 2021).

As in all previous releases, the GHS P2022 is available at the GHSL download portal (<https://ghsl.jrc.ec.europa.eu/download.php>) and as GHSL collection in JRC Open Data Repository (<http://data.jrc.ec.europa.eu/collection/ghsl>). The current data release contains the most up-to-date products and datasets, therefore all previous releases and versions shall be treated as obsolete data.

1.4 Main Characteristics

In order to facilitate data analytics - as it was done in previous issues - the release includes a set of multi-resolution products created by aggregation of the main products. Additionally, the density grids are produced in an equal-area projection as grids of 100 m and 1 km spatial resolution. For example, the multi-temporal population grids were produced in grids of 100 m spatial resolution, and they were then aggregated to 1 km².

The differences between the products in the previous GHS P2019 and those in the current GHS P2022 release are substantial. They include new and more precise 10m-res sub-pixel-fraction built-up surface estimations, new semantics as RES vs. NRES, the building height information, and new seamless interpolated grids at 100m-resolution with equal time intervals of 5 years from 1975 to 2030.

Moreover, an improved approach for production of population grids was applied. Total population time series by country is aligned to the latest UN World Population Prospects 2019 (United Nations, Department of Economic and Social Affairs, Population Division, 2019).The local (i.e. at census polygon level) temporal population estimates are derived with a new model that takes into account the UN ‘city’ population time series of the latest UN World Urbanization Prospects 2018 (United Nations, Department of Economic and Social Affairs, Population Division, 2018). Finally, the population distribution takes advantage of the residential (RES)

vs. non-residential (NRES) semantic abstractions by weighting the population dasymetric downscaling according to the presence of residential (RES) and non-residential (NRES) built-up surface.

The subsections of the Section 2 introduce briefly each product (including more details on differences with the corresponding past version). Dedicated reports are under preparation.

Terms of Use

The data in this data package are provided free-of-charge © European Union, 2022. Reuse is authorised, provided the source is acknowledged. The reuse policy of the European Commission is implemented by a Decision of 12 December 2011 (2011/833/EU). For any inquiry related to the use of these data please contact the GHSL data producer team at the electronic mail address:

JRC-GHSL-DATA@ec.europa.eu

Disclaimer: The JRC data are provided "as is" and "as available" in conformity with the JRC [Data Policy](#)² and the [Commission Decision on reuse of Commission documents](#) (2011/833/EU). Although the JRC guarantees its best effort in assuring quality when publishing these data, it provides them without any warranty of any kind, either express or implied, including, but not limited to, any implied warranty against infringement of third parties' property rights, or merchantability, integration, satisfactory quality and fitness for a particular purpose. The JRC has no obligation to provide technical support or remedies for the data. The JRC does not represent or warrant that the data will be error free or uninterrupted, or that all non-conformities can or will be corrected, or that any data are accurate or complete, or that they are of a satisfactory technical or scientific quality. The JRC or as the case may be the European Commission shall not be held liable for any direct or indirect, incidental, consequential or other damages, including but not limited to the loss of data, loss of profits, or any other financial loss arising from the use of the JRC data, or inability to use them, even if the JRC is notified of the possibility of such damages.

Prior to cite this report, please access the updated version available at:

http://ghsl.jrc.ec.europa.eu/documents/GHSL_Data_Package_2022.pdf

¹JRC Data Policy <https://doi.org/10.2788/607378>

2 Products

2.1 GHS-BUILT-S R2022A - GHS built-up surface grid, derived from Sentinel-2 composite and Landsat, multitemporal (1975-2030)

The GHS-BUILT-S spatial raster dataset depicts the distribution of the built-up (BU) surfaces estimates between 1975 and 2030 in 5 years intervals and two functional use components a) the total BU surface and b) the non-residential (NRES) BU surface. The data is made by spatial-temporal interpolation of five observed collections of multiple-sensor, multiple-platform satellite imageries: Landsat (MSS, TM, ETM sensor) supports the 1975, 1990, 2000, and 2014 epochs, while Sentinel-2 (S2) composite (GHS-composite-S2 R2020A) supports the 2018 epoch.

The sub-pixel built-up fraction (BUFRAC) estimate at 10m resolution is produced from the 10m-resolution Sentinel-2 image composite, using as learning set a composite of data from GHS-BUILT-S2 R2020A, Facebook, Microsoft, and Open Street Map (OSM) building delineation. The adopted inferential engine is a multiple-quantization-minimal-support (MQMS) generalization of the symbolic machine learning (SML) approach (Pesaresi, Syrris, et al., 2016). The SML for the classification of the Sentinel-2 image data uses in input both radiometric and multi-scale morphological image descriptors (Pesaresi, Corbane, et al., 2016) also including functional RES vs. NRES automatic delineation of the built-up areas. In particular, the multiscale decomposition of the image information it is supported by the characteristic-saliency-levelling (CSL) model (Pesaresi et al., 2012) from generalization of the image segmentation based on the derivative of the morphological profile (DMP) (Pesaresi & Benediktsson, 2001). The multi-scale CSL it is solved by computational efficient approach (Ouzounis et al., 2012). The inference is computed locally in data tiles of 100x100 km size.

The non-residential (NRES) domain is predicted from S2 image data by observation of radiometric, textural, and morphological features in a multi-faceted image processing framework merging global unsupervised rule-based reasoning and inductive locally-adaptive methods leveraging on per-pixel spectral indexes, multiscale textural fields assessments, and object-oriented shape analysis. Textural analysis is performed by multi-scale anisotropic rotation invariant contrast measurements done at increasing displacement vector of the co-occurrence matrix selecting the areas where contrast of large objects dominate the textural contrast generated by smaller image structures (Gueguen et al., 2012) (Pesaresi et al., 2008). The connected component (“object”) image analysis is solved by a segmentation of salient image structure based on the watershed of the inverse of the saliency layer in the multiscale image decomposition model introduced as “characteristics-saliency-levelling” CSL (Pesaresi et al., 2012).

As in all the previous GHSL releases (Corbane et al., 2019; Pesaresi, Ehrlich, et al., 2016), the multi-temporal (MT) process works stepwise from recent epochs to past epochs, cancelling the BU information if the decision is supported by empirical evidences from satellite data of the specific epoch. By definition, the process can only decrease the amount of built-up surface going from recent to past epochs. In this release, the same logic it is applied to the data segmented using CSL approach: salient spatial units are delineated by the watershed of the inverse of the continuous BUFRAC function at 10m resolution. This is done in order to increase the robustness of the change detected by the system, vs. the changing sensor resolution of the supporting image data in the various epochs. The surface of the image segments set the minimal mapping unit of the change detected in the GHSL system. The average size of the image segments with the sum of BUFRAC>0 is of 58 pixels of 10m resolution, corresponding to 5800 square meters of surface. The size of the minimal mapping unit is comparable with the surface of pixel from the satellite data of the lowest resolution from datasets used to support the multi-temporal assessment of the GHSL. This was the MSS sensor of the Landsat platform supporting the 1975 epoch, recording data with a native spatial resolution of 57 x 79 meters, corresponding to a surface of 4503 square meters. For each evaluated epoch and for each available Landsat scene, the probability that any specific sensor sample (pixel) can be associated to the foreground (BU) vs. the background (NBU) information semantic is evaluated, by observing the statistical association between the combinations of the quantized reflectance values with the training set data. This inferential process is solved by multiple-quantization-minimal-support (MQMS) generalization of the symbolic machine learning (SML) approach (Pesaresi, Syrris, et al., 2016). The final decision on cancelling or preserving the image segment with BU information is solved by evaluating the cumulative probability of each segment to be classified as BU or NBU using information from all the available image data observations in each specific epoch.

The predicted spatio-temporal interpolation of the built-up surface information for future epochs 2025 and 2030 is solved by rank-optimal spatial allocation using explanatory variables related to the landscape (slope, elevation, distance to water, and distance to vegetation) and related to the observed dynamic of BU surfaces

in the past epoch. Between observed epochs, at each 5-years interval not corresponding to an observed epoch, a linear spatial-temporal interpolation it is applied. The values of the pixels are interpolated linearly, and the spatial evolution of the built domain defined as the BUFRAC>0 it is solved by rank-optimal spatial allocation as for the future epochs.

2.1.1 Definitions

The building

Since the initial concept (Pesaresi et al., 2013) the definition adopted by the GHSL is the same of the INSPIRE “building” abstraction (<https://inspire.ec.europa.eu/id/document/tg/bu>), limited to the above-ground case, and without the “permanent” characterization of the built-up structures, allowing to be inclusive to temporary settlements as associated to slums, rapid migratory patterns, or displaced people because of natural disasters or crisis . “...Buildings are constructions above (and/or underground) which are intended or used for the shelter of humans, animals, things, the production of economic goods or the delivery of services and that refer to any structure (permanently) constructed or erected on its site...” . In short, and taking in to account the remote-sensing technology characteristics and limitations we can summarize the implicit GHSL abstraction of the “building” as: “any roofed structure erected above ground for any use”.

The built-up surface (BUSURF)

The built-up surface is the gross surface (including the thickness of the walls) bounded by the building wall perimeter with a spatial generalization matching the 1:10K topographic map specifications, that it also informally called “building footprint”.

The built-up fraction (BUFRAC)

The built-up fraction (BUFRAC) is the share of the raster sample (pixel) surface that is covered by the built-up surface.

The residential (RES) domain

The RES domain is defined as the built-up surface dedicated prevalently for residential use. The residential use is defined as from INSPIRE: “...Areas used dominantly for housing of people. The forms of housing vary significantly between, and through, residential areas. These areas include single family housing, multi-family residential, or mobile homes in cities, towns and rural districts if they are not linked to primary production. It permits high density land use and low density uses. This class also includes residential areas mixed with other non-conflicting uses and other residential areas...” (https://inspire.ec.europa.eu/codelist/HILUCSValue/5_ResidentialUse)

The “non-residential domain” (NRES)

The “non-residential domain” (NRES) is defined as the domain of the BUFRAC>0 complement of the RES domain. This can be worded also as “any built-up surface not included in the RES class”

Examples:

Let assume a 100m resolution raster grid with a 100x100 = 10,000 square meters of surface per spatial sample (pixel) of this grid. Moreover, let be the built-up surface predicted at the sample X of this grid is BUSURFx = 750 square meters.

The corresponding built-up fraction estimate will be : BUFRACx = 750/10,000 = 0.075

Let assume in the sample x of 100x100m resolution the total BUSURFx = 4380 square meters, while the NRES BUSURFx = 850 square meters. Then the residential built-up surface RES BUSURFx can be predicted as RES BUSURFx = 4380 – 850 = 3530 square meters.



Figure 1 - Santiago de Chile: comparison of the built-up surfaces as assessed by the previous GHS_BUILT_LDSMT_GLOBE_R2018A for the epoch 2014 from Landsat image data with a Boolean 30m-resolution method (upper), vs the new GHS-BUILT-S_GLOBE_R2022A for the epoch 2018 from Sentinel-2 image data with a continuous 10m-resolution method (lower).

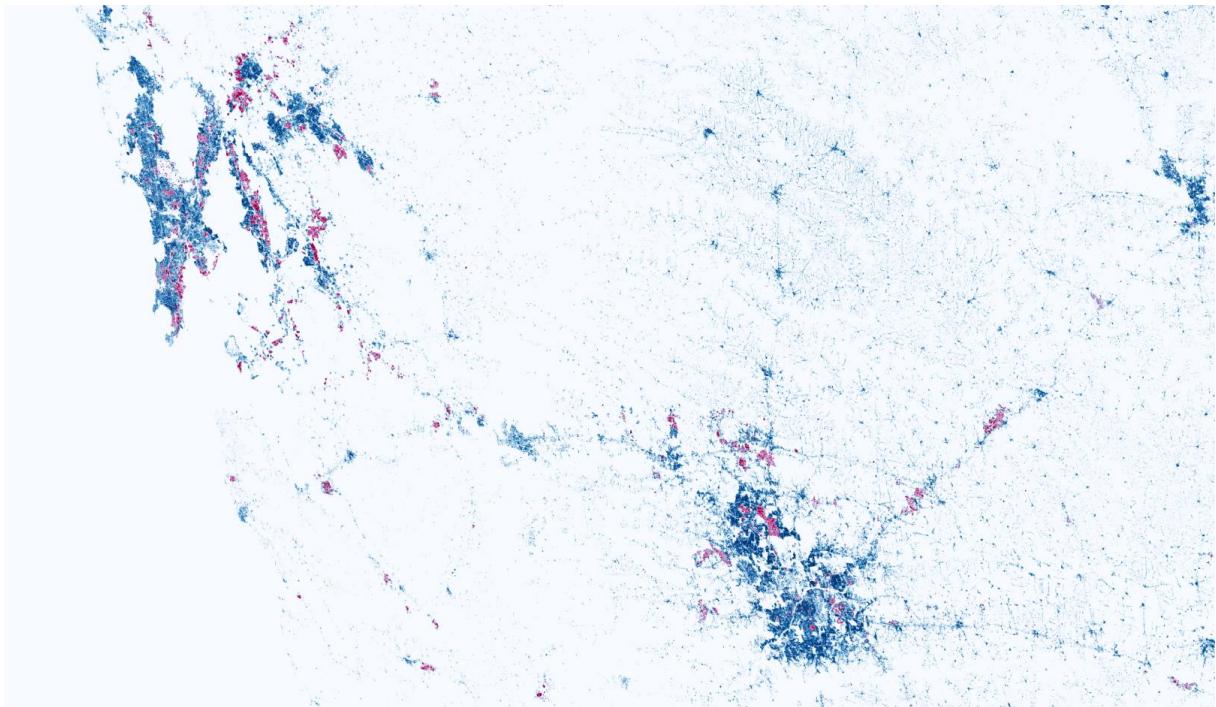


Figure 2 - Mumbai-Pune (India): residential (RES) and non-residential (NRES) components of the built-surfaces estimated for the GHSL 2020 epoch. RES and NRES are represented with blue and magenta, respectively.

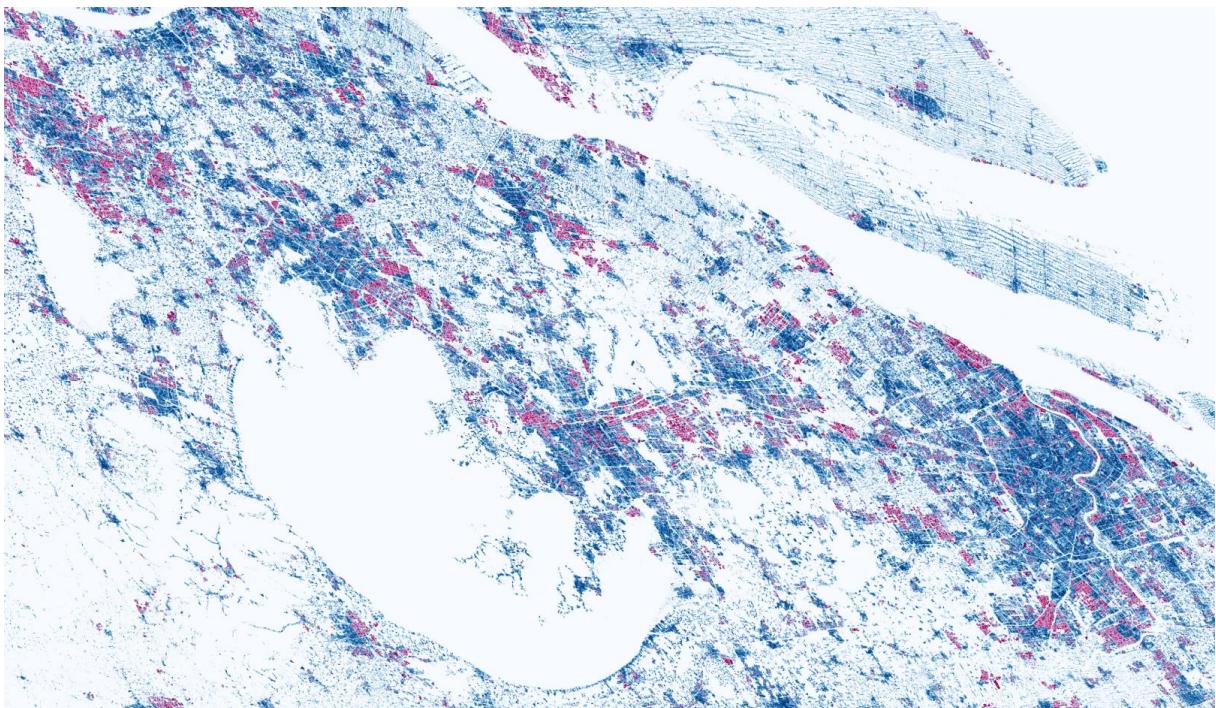


Figure 3 - Shanghai-Changzhou (China): residential (RES) and non-residential (NRES) components of the built-surfaces estimated for the GHSL 2020 epoch. RES and NRES are represented with blue and magenta, respectively.

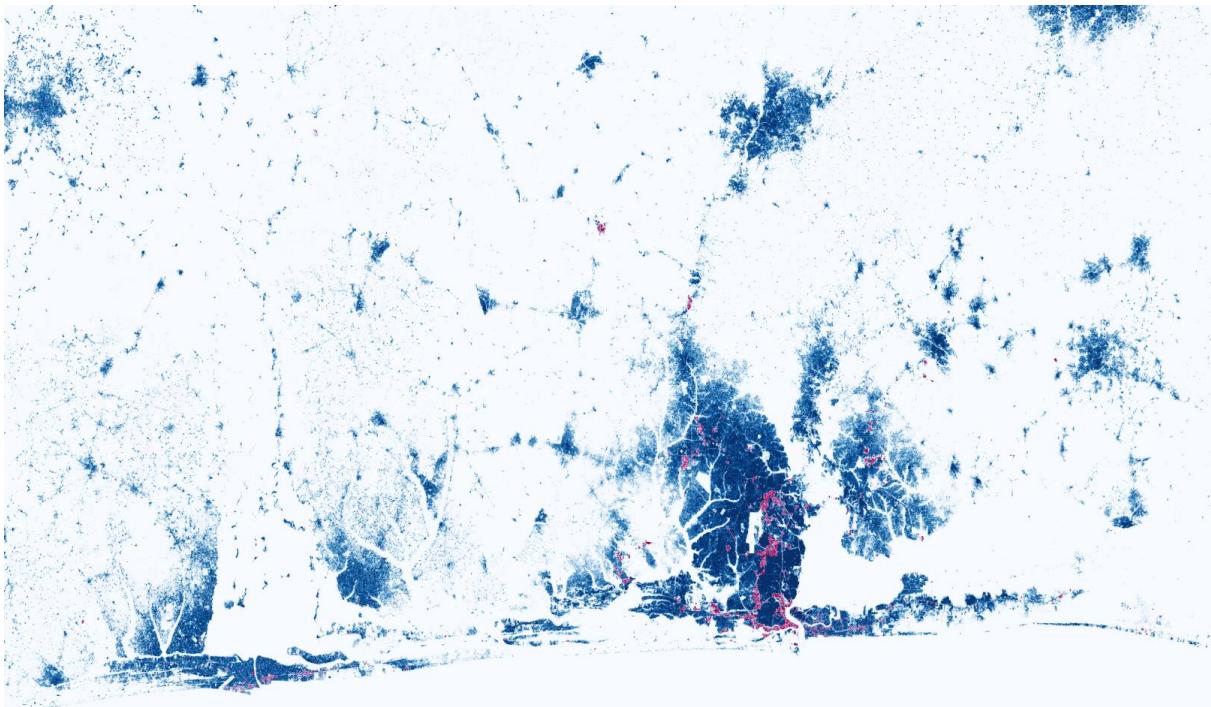


Figure 4 - Lagos-Porto Novo-Abeokuta (Nigeria): residential (RES) and non-residential (NRES) components of the built-surfaces estimated for the GHSL 2020 epoch. RES and NRES are represented with blue and magenta, respectively.

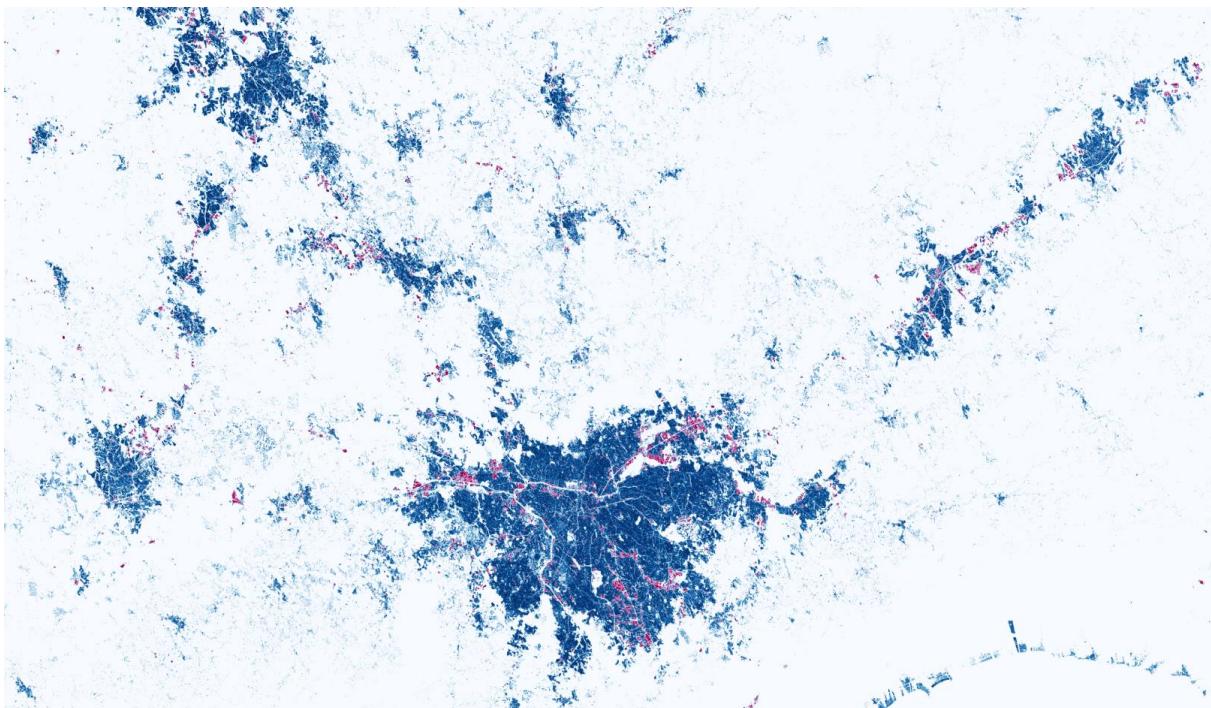


Figure 5 - Sao Paulo- Campinas - Sao Jose dos Campos (Brazil): residential (RES) and non-residential (NRES) components of the built-surfaces estimated for the GHSL 2020 epoch. RES and NRES are represented with blue and magenta, respectively.

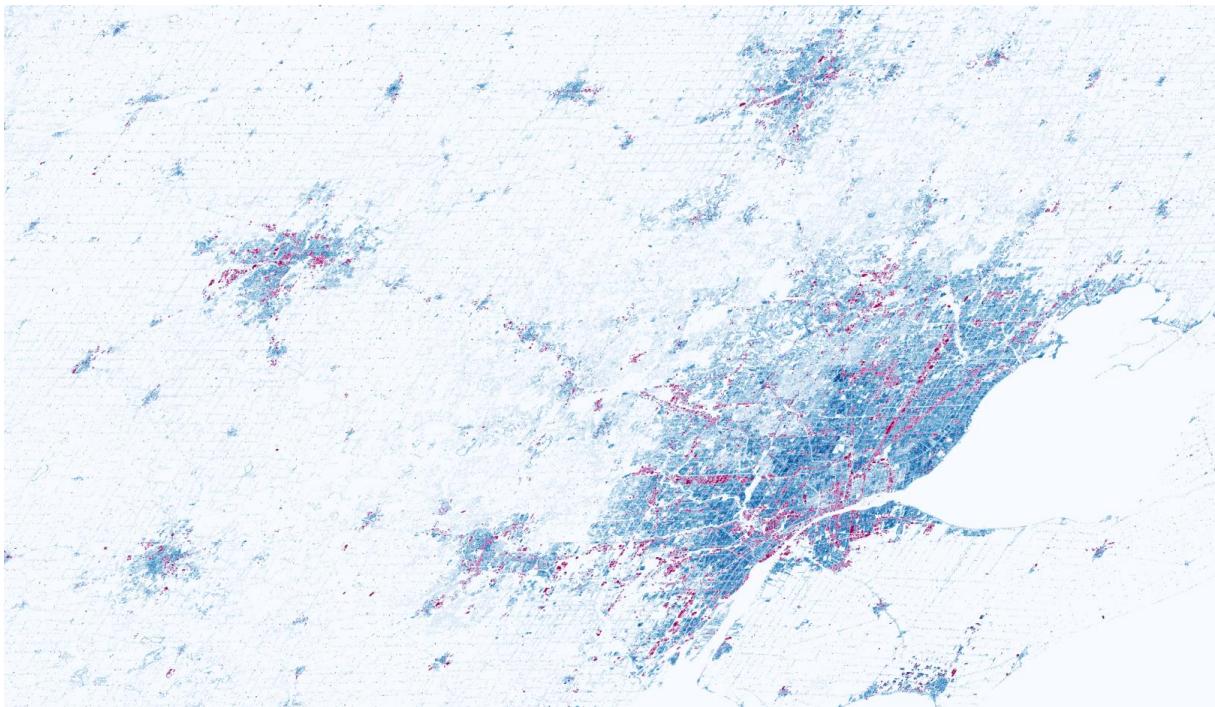


Figure 6 - Detroit-Lansing-Flint (United States): residential (RES) and non-residential (NRES) components of the built-surfaces estimated for the GHSL 2020 epoch. RES and NRES are represented with blue and magenta, respectively.

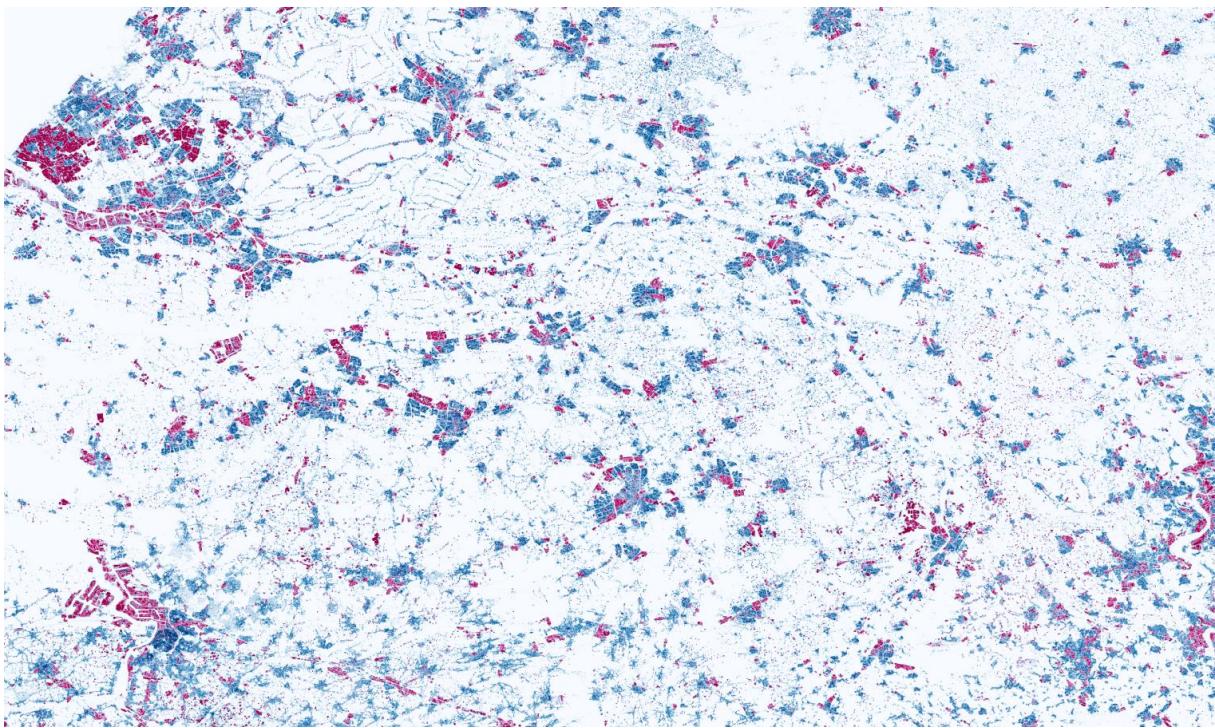


Figure 7 - The Hague - Rotterdam- Antwerp (The Netherlands): residential (RES) and non-residential (NRES) components of the built-surfaces estimated for the GHSL 2020 epoch. RES and NRES are represented with blue and magenta, respectively.

2.1.2 Expected Errors

The estimation of the GHS-BUILT-S error is currently ongoing, and will be delivered in peer-reviewed publication possibly in 2023-2024.

Preliminary expected error scores in prediction of the built-up surface fraction (BUFRAC) were estimated vs. observed built-up surfaces from building footprints available in vector format at scale 1:10K. The test set is made by the global collection of building footprints used for QC during the GHSL production.

Table 1 shows the expected errors of the new GHS-BUILT-S R2022A release at 10m resolution stratified by class of the Copernicus Global Land cover at 100m resolution (Buchhorn et al., 2020).

Table 2 shows the expected errors of the new GHS-BUILT-S R2022A release at 100m resolution as compared to the previous GHSL release made from Boolean classification of Landsat data (GHS-BUILT R2018A) and as compared to the predictions included in the continuous “urban cover fraction” (UCF) of the Copernicus Global Land cover at 100m resolution (Buchhorn et al., 2020). The test data are subdivided in two geographical strata (“non-US” and “US”) in order to control the performances of the model in different settlement pattern conditions.

In both cases, the errors of the predicted built-up surface fractions (BUFRAC) are estimated vs. observed built-up surface fraction as obtained by rasterization of building footprints available in vector format at scale 1:10K. The test set is made by the global collection of building footprints used for QC during the GHSL production.

The capacity of GHS-BUILT-S R2022A to predict the built-up share (quasi continuous, 64 levels) in a globally-representative set of almost 50,000 test cases of 80x80 meter size visually inspected with a Boolean interpretation schema at 10m resolution (See et al., 2022), yield a Pearson Correlation Coefficient of the linear least square regression equal to 0.81363. To be noticed that the correlation is systematically decreased by the fact that the reference data is not spatially aligned with the GHSL data, and by the fact that the GHSL uses a continuous classification schema of the 10m-res raster samples, while the reference data applies a human interpretation schema based on Boolean classification.

Table 3 shows the amount of total built-up surfaces and NRES built-up surfaces assessed by the GHS-BUILT-S R2022A data (epoch 2020) stratified by land use classes in United States (NLUD³), from (Theobald, 2014), and Europe (CLC⁴), ordered by the GHSL NRES surface share. The table shows the empirical association between the NRES class and the Land Use classes. To be noticed that the measured association is only indicative and systematically decreased by the fact that the GHSL data is derived from 10m-resolution imagery and consequently has a much higher spatial precision as compared to the land use data used as reference, which is defined with a minimal mapping unit in the order of hectares.

³ Land Use Classification and Map for the US http://csp-inc.org/public/NLUD2010_20140326.zip

⁴ Corine Land Cover European seamless 100m raster database (Version 20b2)
<https://land.copernicus.eu/pan-european/corine-land-cover/clc2018?tab=metadata>

Table 1 – Expected error scores in prediction of the built-up surface fraction (BUFRAC) at 10m resolution in the new GHS-BUILT-S R2022A release.

CODE	LABEL	RMSE	MAE	NSAMPLES
0	Not Classified	0.122	0.070	7,509
20	Shrubs	0.120	0.056	555,261,719
30	Herbaceous vegetation	0.136	0.060	1,310,171,720
40	Cultivated and managed vegetation agriculture cropland	0.071	0.020	914,235,915
50	Urban / built up	0.296	0.218	337,089,799
60	Bare soil or sparse vegetation	0.192	0.111	123,557,016
70	Snow and Ice	0.001	0.000	13,447,153
80	Permanent water bodies	0.028	0.005	116,262,034
90	Herbaceous wetland	0.062	0.019	63,299,127
100	Moss and lichen	0.006	0.001	347,875
111	Closed forest, evergreen needle leaf	0.071	0.025	573,176,106
112	Closed forest, evergreen, broad leaf	0.012	0.002	575,661,481
113	Closed forest, deciduous needle leaf	0.069	0.018	19,610
114	Closed forest, deciduous broad leaf	0.035	0.007	751,956,669
115	Closed forest, mixed	0.027	0.006	145,625,305
116	Closed forest, unknown	0.071	0.026	142,841,787
121	Open forest, evergreen needle leaf	0.101	0.042	111,872,259
122	Open forest, evergreen broad leaf	0.023	0.005	2,599,647
123	Open forest, deciduous needle leaf	0.000	0.000	181
124	Open forest, deciduous broad leaf	0.067	0.021	168,233,648
125	Open forest, mixed	0.048	0.012	1,299,363
126	Open forest, unknown	0.107	0.043	839,029,693
200	Open sea	0.072	0.025	144,012,925
Total		0.075	0.034	6,890,008,541

Table 2 – Expected error scores in prediction of the built-up surface fraction (BUFRAC) at the aggregated 100m and 1km resolution.

	non-US			US			TOTAL		
	RMSE	MAE	NSAMPLES	RMSE	MAE	NSAMPLES	RMSE	MAE	NSAMPLES
Prediction at 100m resolution									
GHS-BUILT-S R2022A	0.062	0.035	32418503	0.062	0.035	35886736	0.062	0.035	68305239
GHS-BUILT R2018A	0.194	0.129	32418503	0.308	0.213	35886736	0.258	0.177	68305239
Copernicus Global Land Service UCF	0.255	0.173	32418503	0.285	0.195	35886736	0.272	0.186	68305239
Prediction at 1km resolution									
GHS-BUILT-S R2022A	0.036	0.026	301940	0.038	0.030	329121	0.037	0.028	631061
GHS-BUILT R2018A	0.144	0.110	301940	0.259	0.209	329121	0.213	0.170	631061
Copernicus Global Land Service UCF	0.195	0.148	301940	0.242	0.193	329121	0.223	0.175	631061

Table 3 – The amount of total built-up surfaces and NRES built-up surfaces assessed in the GHS-BUILT-S R2022A data stratified by land use classes in United States (NLUD) and Europe (CLC).

LUSOURC	LUCLASS	LABEL1	LABEL2	LABEL3	BUTOT m2	BUNRES m2	NRES_shai
NLUD	251	Built-up	Transportation	Airports (developed)	371014506	253357093	68.29%
CLC	121	Artificial surfaces	Industrial, commercial and transport units	Industrial or commercial units	7649839186	5056391057	66.10%
CLC	123	Artificial surfaces	Industrial, commercial and transport units	Port areas	327183968	215970214	66.01%
NLUD	231	Built-up	Industrial	Factory, plant	1619490370	1045588694	64.56%
NLUD	223	Built-up	Commercial	Entertainment (stadiums, amusement, etc.)	53632078	33581582	62.61%
NLUD	221	Built-up	Commercial	Office	1614791936	978439733	60.59%
NLUD	222	Built-up	Commercial	Retail/shopping centers	1742663387	1039508683	59.65%
CLC	124	Artificial surfaces	Industrial, commercial and transport units	Airports	170180711	98086496	57.64%
NLUD	233	Built-up	Industrial	Confined animal feeding	88703844	50398369	56.82%
NLUD	249	Built-up	Institutional	Prison/penitentiary	16152586	8924402	55.25%
NLUD	241	Built-up	Institutional	Schools (dev)	225846287	123608473	54.73%
NLUD	242	Built-up	Institutional	Schools (undeveloped)	598016804	326285150	54.56%
NLUD	243	Built-up	Institutional	Medical (hospitals, nursing home, etc.)	367951278	196239687	53.33%
NLUD	244	Built-up	Institutional	Government/public	86726621	44112906	50.86%
CLC	212	Agricultural areas	Arable land	Permanently irrigated land	622458355	268234717	43.09%
CLC	132	Artificial surfaces	Mine, dump and construction sites	Dump sites	40734981	12866115	31.58%
NLUD	261	Built-up	Transportation	Rural buildings, cemetery	1242246217	391906221	31.55%
NLUD	341	Production	Timber	Timber harvest	165945723	46319990	27.91%
CLC	122	Artificial surfaces	Industrial, commercial and transport units	Road and rail networks and associated land	224903444	57729502	25.67%
CLC	133	Artificial surfaces	Mine, dump and construction sites	Construction sites	94362978	22856256	24.22%
NLUD	255	Built-up	Transportation	Undeveloped	173761477	39555968	22.76%
CLC	213	Agricultural areas	Arable land	Rice fields	32328402	6089237	18.84%
CLC	131	Artificial surfaces	Mine, dump and construction sites	Mineral extraction sites	142861969	24980523	17.49%
NLUD	330	Production	Mining	Mining strip mines, quarries, gravel pits	8185807	1420084	17.35%
NLUD	245	Built-up	Institutional	Military/DOD/DOE (dev)	104554407	17490907	16.73%
NLUD	246	Built-up	Institutional	Military/DOD (training)	217411481	31391901	14.44%
CLC	211	Agricultural areas	Arable land	Non-irrigated arable land	7385826240	982409064	13.30%
CLC	141	Artificial surfaces	Artificial, non-agricultural vegetated areas	Green urban areas	165087965	20670163	12.52%
NLUD	252	Built-up	Transportation	Highways, railways	1042336050	121981507	11.70%
CLC	111	Artificial surfaces	Urban fabric	Continuous urban fabric	2160143750	194234820	8.99%
CLC	222	Agricultural areas	Permanent crops	Fruit trees and berry plantations	398088530	33459968	8.41%
NLUD	411	Recreation	Developed park	Urban park	202120970	16703263	8.26%
CLC	142	Artificial surfaces	Artificial, non-agricultural vegetated areas	Sport and leisure facilities	586211113	45131809	7.70%
CLC	231	Agricultural areas	Pastures	Pastures	4571276116	337410401	7.38%
NLUD	310	Production	General	General agricultural	158011938	11520002	7.29%
NLUD	313	Production	Cropland	Orchards	71453036	4800693	6.72%
CLC	112	Artificial surfaces	Urban fabric	Discontinuous urban fabric	30206040355	1990977548	6.59%
NLUD	314	Production	Cropland	Sod & switch grass	14887041	933461	6.27%
NLUD	311	Production	Cropland	Cropland/row crops	3107157034	190204508	6.12%
CLC	242	Agricultural areas	Heterogeneous agricultural areas	Complex cultivation patterns	4975196695	283937849	5.71%
NLUD	213	Built-up	Residential	Suburban (1–2.5 ac)	9335940633	520592660	5.58%
CLC	243	Agricultural areas	Heterogeneous agricultural areas	Land principally occupied by agriculture, with si	2121377478	109108914	5.14%
NLUD	321	Production	Rangeland	Grazed	3116664551	160096140	5.14%
CLC	221	Agricultural areas	Permanent crops	Vineyards	391114177	19693622	5.04%
NLUD	214	Built-up	Residential	Exurban (2.5–10 ac)	8778472493	408859021	4.66%
NLUD	211	Built-up	Residential	Dense urban (>0.1 ac)	1461479361	64398593	4.41%
CLC	223	Agricultural areas	Permanent crops	Olive groves	436848677	17939775	4.11%
NLUD	415	Recreation	Developed park	Resort/ski area	43570356	1739805	3.99%
NLUD	410	Recreation	Undifferentiated park	General park	201511693	7926588	3.93%
NLUD	421	Recreation	Natural park	Natural park	144612371	5448426	3.77%
NLUD	417	Recreation	Developed park	Campground/ranger station	20170842	710508	3.52%
CLC	244	Agricultural areas	Heterogeneous agricultural areas	Agro-forestry areas	44227686	1378654	3.12%
CLC	241	Agricultural areas	Heterogeneous agricultural areas	Annual crops associated with permanent crops	149655325	4474819	2.99%
NLUD	412	Recreation	Developed park	Golf course	202243104	5893655	2.91%
NLUD	312	Production	Cropland	Pastureland	3456950217	91632160	2.65%
NLUD	212	Built-up	Residential	Urban (0.1–1)	16435713363	399274360	2.43%
NLUD	215	Built-up	Residential	Rural (10–40 ac)	5350324389	111117959	2.08%
NLUD	422	Recreation	Natural park	Designated recreation area	203351	2066	1.02%

2.1.3 Improvements compared to the previous release

Improved input

Improved input data used in this release includes: improved satellite imagery and new prior knowledge and learning set.

Regarding improved satellite imagery we have included a more populated historical Landsat data series, a new epoch 2018 from Sentinel-2 with 10m spatial resolution vs. previous 30m Landsat data.

Regarding new prior knowledge and learning set we have included new BU labels as i) the GHS-BUILT-S2 R2020A⁵ that is a probability to the BU class grid derived from Sentinel-2 global image composite for reference year 2018 using Convolutional Neural Networks (GHS-S2Net; Corbane, Syrris, et al., 2020) and ii) the buildings derived from VHR imagery by Microsoft⁶ and Facebook⁷ open efforts, the new BU change map as included in the GHS-BUILT R2018A, GHS built-up grid, derived from Landsat, multitemporal (1975-1990-2000-2014), and Land Use and other information included in National land use data (US NLCD⁸, EU CORINE⁹) and Volunteered geographical information by Open Street Map (OSM)¹⁰ on LANDUSE, ROADS, RIVER, STREAMS.

List of Countries where Microsoft building footprint data were available during the GHSL production: USA, Canada, Australia, Uganda, Tanzania, Nigeria, Kenya, Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, Guyana, Paraguay, Peru, Uruguay, Venezuela.

List of Countries where high resolution settlement layer (HRSI) data from Facebook were available during the GHSL production: Albania, Algeria, American Samoa, Andorra, Angola, Anguilla and Barbuda, Argentina, Aruba, Australia, Austria, Bahamas, Bahrain, Bangladesh, Barbados, Belarus, Belgium, Belize, Benin, Bhutan, Bolivia, Bosnia and Herzegovina, Botswana, Brazil, British Virgin Islands, Brunei, Bulgaria, Burkina Faso, Burundi, Cabo Verde, Cambodia, Cameroon, Cayman Islands, Central African Republic, Chad, Chile, Colombia, Comoros, Congo, Cook Island, Costa Rica, Cote d'Ivoire, Croatia, Czechia, Democratic Republic of the Congo, Djibouti, Dominica, Dominican Republic, Ecuador, Egypt, El Salvador, Equatorial Guinea, Eritrea, Estonia, Federated States of Micronesia, Fiji, France, French Guiana, French Polynesia, Gabon, Gambia, Gibraltar, Georgia, Germany, Ghana, Greece, Grenada, Guadeloupe, Guam, Guatemala, Guinea, Guinea Bissau, Guyana, Haiti, Honduras, China Hong Kong Special Administrative Region, Hungary, Iceland, Indonesia, Iraq, Ireland, Italy, Jamaica, Japan, Jordan, Kazakhstan, Kenya, Kingdom of Eswatini, Kiribati, Kuwait, Kyrgyzstan, Laos, Latvia, Lebanon, Lesotho, Liberia, Libya, Liechtenstein, Lithuania, Luxembourg, China Macao Special Administrative Region, Madagascar, Malawi, Malaysia, Maldives, Mali, Malta, Marshall Islands, Mauritania, Mauritius, Mayotte, Mexico, Moldova, Monaco, Mongolia, Montserrat, Mozambique, Namibia, Naura, Nepal, Netherlands, New Caledonia, New Zealand, Nicaragua, Niger, Nigeria, The former Yugoslav Republic of Macedonia, Northern Mariana Islands, Oman, Palau, Panama, Papua New Guinea, Paraguay, Peru, Philippines, Poland, Portugal, Puerto Rico, Qatar, Reunion, Romania, Rwanda, Saint Kitts and Nevis, Saint Lucia, Saint Vincent and the Grenadines, Samoa, San Marino, Sao Tome and Principe, Saudi Arabia, Senegal, Serbia, Seychelles, Sierra Leone, Singapore, Slovakia, Slovenia, Solomon Islands, South Africa, South Korea, SouthAsia AS42, SouthAsia AS43, SouthAsia AS44, SouthAsia AS47, Spain, Sri Lanka, Suriname, Switzerland, Taiwan, Tajikistan, Thailand, Timor Leste, Togo, Tonga, Trinidad and Tobago, Tunisia, Turkmenistan, Turks and Caicos, Tuvalu, Uganda, United Arab Emirates, United Kingdom, United Republic of Tanzania, Uruguay, US Virgin Islands, United States of America, Uzbekistan, Vanuatu, Vietnam, Wallis and Futuna Islands, Zambia, Zimbabwe

Improved Output

Several improvements regarding the assessment of the built-up surfaces are included in this new release, as compared to the previous GHSL data. They may be summarized in the following points:

The prediction of the built-up abstraction class is continuous at the 10m resolution (BUFRAC), vs. the previous GHSL releases that were based on Boolean predictions at 30m resolution. This fact augments the precision of the predicted amount of built-up surfaces at the generalized grid level (100m 1km), at the city level and at the national level.

⁵ <https://ghsl.jrc.ec.europa.eu/download.php?ds=buS2>

⁶ <https://github.com/microsoft/USBuildingFootprints>

⁷ <https://research.fb.com/downloads/high-resolution-settlement-layer-hrsi/>

⁸ <https://www.mrlc.gov/data>

⁹ <https://land.copernicus.eu/pan-european/corine-land-cover/clc2018>

¹⁰ <https://www.openstreetmap.org/>

The new GHSL release includes a new classification of the non-residential (NRES) built-up surfaces that was not available in the previous releases. This fact will improve the usability of the new GHSL data in applications requiring a functional classification of the built environment.

The new GHSL release produces data at equal time interval in the time range 1975-2030 by a spatial-temporal interpolation process, while previous GHSL releases were reporting data in arbitrary points in time where satellite images were available. This fact will improve the usability of the new GHSL data in trend and projection analysis requiring consistent time intervals (see section 2.1).

Table 4 - Summary of the characteristics of the new GHS-BUILT data vs. the previous releases

Data Characteristics	GHS-BUILT R2018A	GHS-BUILT-S R2022A
Definition of the built-up class abstraction	INSPIRE “BUILDING” roofed structure above ground for any use	INSPIRE “BUILDING” roofed structure above ground for any use
Definition of the NRES vs. RES class abstraction	Not available	Derived from INSPIRE “residential” use definition
Built-up surface : class	Boolean	Continuous
Built-up surface : RES vs. NRES class	Not available	Boolean
Built-up surface : spatial resolution	30m	10m
Built-up surface : observed epochs	4 (1975,1990,2000,2014)	5 (1975,1990,2000,2014,2018)
Built-up surface : change map	Pixel-based	Segment-based
Building height : measurement	Not available	Continuous
Building height : spatial resolution	Not available	100m
Building height : observed epoch	Not available	2018
Built-up surface : spatial resolution of the generalized grids	250m 1km	100m 1km
Built-up volume : spatial resolution of the generalized grids	Not available	100m 1km
Equal-time-interval spatial-temporal interpolated grids of built-up surfaces and volume	Not available	12 epochs (1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010, 2015, 2020, 2025, 2030)
Future grids projections	Not available	2025 and 2030

2.1.4 Input Data

Remotely sensed image data

The remotely sensed image data supporting this GHSL release are collected by the Landsat and the Sentinel platforms, organized in five epochs: 1975, 1990, 2000, 2014, and 2018.

The Landsat image data used in input include 35479 individual scenes organized in four epochs 1975, 1990, 2000, and 2014. The average absolute time tolerance of the image data collection time vs. the nominal time barycentre of the epoch is 2.0, 2.2, 1.2, and 0.8 years for the 1975, 1990, 2000, and 2014 epochs, respectively. The aggregated time precision of all the data in the four epochs is of 1.5 years. The empirical time barycentre for the epochs 1975, 1990, 2000, and 2014 is the year 1975.1, 1989.4, 2000.8, and 2013.2, respectively.

Table 5 - Summary of the Landsat Image data used in input

Row Labels	Count of YEAR	Average of YEAR	Average of ABS_TimeTolerance
1975	7355	1975.1	2.0
L1	4403	1973.4	1.6
L2	1880	1976.8	1.8
L3	1072	1979.3	4.3
1990	8011	1989.4	2.2
L4	258	1983.5	6.5
L5	7475	1989.4	2.0
L6	278	1994.3	4.3
2000	9774	2000.8	1.2
L5	459	2004.7	5.7
L6	8827	2000.4	0.7
L7	488	2005.5	5.5
2014	10339	2013.2	0.8
L5	638	2009.6	4.4
L7	259	2009.5	4.5
L8	9442	2013.5	0.5
Grand Total	35479	1996.5	1.5

The epoch 2018 is made by the GHS_composite_S2_L1C_2017-2018_GLOBE_R2020A¹¹ that corresponds to global cloud-free pixel based composite created from the Sentinel-2 data archive (Level L1C) available in Google Earth Engine¹² for the period January 2017 - December 2018.

High-level semantic abstraction data

Several high-level semantic data are used in the process with the function of prior knowledge supporting the various phases of data classification concerning the BU surface fraction estimates, the historical change detection in the BU surfaces, and the land use classification of RES vs. NRES BU surfaces.

BU class abstraction labels : i) the GHS-BUILT-S2 R2020A¹³ that is a probability to the BU class grid derived from Sentinel-2 global image composite for reference year 2018 using Convolutional Neural Networks (GHS-S2Net) and ii) the buildings derived from VHR imagery by Microsoft¹⁴ and Facebook¹⁵ open efforts

Multi-temporal assessments: the BU change map as included in the GHS-BUILT R2018A, GHS built-up grid, derived from Landsat, multitemporal (1975-1990-2000-2014)

Land Use and other: information included in National land use data (US NLUD¹⁶, EU CORINE¹⁷) and Volunteered geographical information by Open Street Map (OSM)¹⁸ on LANDUSE, ROADS, RIVER, STREAMS

¹¹ <https://ghsl.jrc.ec.europa.eu/download.php?ds=compositeS2>

¹² https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2

¹³ <https://ghsl.jrc.ec.europa.eu/download.php?ds=buS2>

¹⁴ <https://github.com/microsoft/USBuildingFootprints>

¹⁵ <https://research.fb.com/downloads/high-resolution-settlement-layer-hrsl/>

¹⁶ <https://www.mrlc.gov/data>

¹⁷ <https://land.copernicus.eu/pan-european/corine-land-cover/clc2018>

¹⁸ <https://www.openstreetmap.org/>

2.1.5 Technical Details

Author: Pesaresi, Martino; Politis Panagiotis

Product name: GHS-BUILT-S_GLOBE_R2022A

Spatial extent: Global

Temporal extent: from 1975 to 2030, 5 years interval

*Coordinate Systems**: World Mollweide (ESRI:54009)

*Spatial resolution available**: 10 m, 100 m, and 1 km

*Encoding**: integer (Byte, UInt16, UInt32), unit: built square meters in the grid cell

Data organisation ()*: Global VRT file (10 m) or GeoTIFF file (100m, 1km) with overview images (OVR). Data tiles of 1000x1000 km size in GeoTIFF format (10 m, 100 m, 1 km)

Table 6 outlines the technical characteristics of the datasets released in this data package.

(*) product dependent, see Table 2.

Disclaimer: the re-projection of the World Mollweide version of the GHS-BUILT-S_GLOBE_R2022A to coordinate systems requires specific technical knowledge. No responsibility is taken for workflows developed independently by users.

Table 6 - Technical details of the datasets in GHS-BUILT-S_GLOBE_R2022A

GHS-BUILT-S_GLOBE_R2022A		
ID	Description	Resolution (projection)
GHS_BUILT_S_E<epoch>_GLOBE_R2022A_54009_<res>_V1_0	BU surface <epoch> 1975-2020; <res> 100, 1000 Encoding: UInt16 (100 m), UInt32 (1 km) Values range: 0-10000 (100 m), 0-1000000 (1 km) NoData: 65535 (100 m), 4294967295 (1 km)	100 m, 1 km World Mollweide (ESRI:54009)
GHS_BUILT_S_P<epoch&model>_GLOBE_R2022A_54009_<res>_V1_0	BU surface projections <epoch> 2025-2030; <model> LIN-PLY-MED; <res> 100, 1000 Encoding: UInt16 (100 m), UInt32 (1 km) Values range: 0-10000 (100 m), 0-1000000 (1 km) NoData: 65535 (100 m), 4294967295 (1 km)	100 m, 1 km World Mollweide (ESRI:54009)
GHS_BUILT_S_E2018_GLOBE_R2022A_54009_10_V1_0	BUFRAC at 10m spatial resolution for E2018 Encoding: Byte Values range: 0-100 NoData: 255	10 m World Mollweide (ESRI:54009)

GHS-BUILT-S_NRES_GLOBE_R2022A		
ID	Description	Resolution (projection)
GHS_BUILT_S_NRES_E<epoch>_GLOBE_R2022A_54009_<res>_V1_0	<p>Non-residential BU surface <epoch> 1975-2020;; <res> 100m, 1000</p> <p>Encoding: UInt16 (100 m), UInt32 (1 km)</p> <p>Values range: 0-10000 (100 m), 0-1000000 (1 km)</p> <p>NoData: 65535 (100 m), 4294967295 (1 km)</p>	100 m, 1 km World Mollweide (ESRI:54009)
GHS_BUILT_S_NRES_P<epoch&model>_GLOBE_R2022A_54009_<res>_V1_0	<p>Non-residential BU surface projections <epoch> 2025-2030; <model> LIN-PLY-MED; <res> 100, 1000</p> <p>Encoding: UInt16 (100 m), UInt32 (1 km)</p> <p>Values range: 0-10000 (100 m), 0-1000000 (1 km)</p> <p>NoData: 65535 (100 m), 4294967295 (1 km)</p>	100 m, 1 km World Mollweide (ESRI:54009)
GHS_BUILT_S_NRES_E2018_GLOBE_R2022A_54009_10_V1_0	<p>NRES 10m (boolean) at 10m spatial resolution for E2018</p> <p>Encoding: Byte</p> <p>Values range: 0:non-NRES ; 1:NRES</p> <p>NoData: 255</p>	10 m World Mollweide (ESRI:54009)

2.1.6 Summary statistics

Table 7 - Summary statistics of total built-up surface in millions of square meters, by continent. Years 2025 and 2030 are reported for the linear (LIN), second-order polynomial (PLY), and median (MED) future projection scenarios

Values	Latin America and the Caribbean							Grand Total
	Africa	Asia	Europe	North America	Oceania			
Sum of E1975	19233	80887	37338	20499	27907	3525	189388	
Sum of E1980	20082	85067	38218	21137	28760	3599	196864	
Sum of E1985	21342	91343	39476	22045	30047	3706	207960	
Sum of E1990	22958	99486	41045	23190	31727	3846	222252	
Sum of E1995	25044	107274	43314	24876	34056	3993	238558	
Sum of E2000	28450	119689	46940	27527	38356	4253	265216	
Sum of E2005	32584	131232	51589	29649	41569	4499	291122	
Sum of E2010	39476	148659	58553	33031	46910	4908	331537	
Sum of E2015	49039	172406	68268	37560	54690	5488	387451	
Sum of E2020	63414	209098	84472	44012	68521	6459	475977	
Sum of P2025_LIN	70048	223594	90408	47309	73389	6789	511536	
Sum of P2030_LIN	74342	233175	94673	49683	76889	7051	535814	
Sum of P2025_POLY	78027	238947	98048	51642	79760	7278	553702	
Sum of P2030_POLY	90470	262318	108769	58185	88834	8013	616588	
Sum of P2025_MEDIAN	74238	231682	94434	49538	76725	7040	533657	
Sum of P2030_MEDIAN	82245	247870	101943	53974	82999	7537	576568	

Table 8 - Summary statistics of non-residential (NRES) built-up surface in millions of square meters, by continent. Years 2025 and 2030 are reported for the linear (LIN), second-order polynomial (PLY), and median (MED) future projection scenarios

Values	Africa	Asia	Europe	Latin America and the Caribbean	Northern America	Oceania	Grand Total
Sum of E1975	397	3308	6540	780	4372	276	15673
Sum of E1980	416	3471	6715	801	4481	284	16168
Sum of E1985	455	3849	7060	846	4713	299	17222
Sum of E1990	520	4504	7599	921	5106	323	18973
Sum of E1995	539	4846	7913	964	5367	332	19961
Sum of E2000	588	5819	8671	1073	6117	351	22620
Sum of E2005	604	6121	8991	1112	6254	359	23442
Sum of E2010	635	6793	9575	1189	6531	374	25097
Sum of E2015	690	8065	10540	1319	7001	397	28013
Sum of E2020	782	10014	11833	1482	7579	425	32116
Sum of P2025_LIN	801	10344	12044	1513	7689	430	32821
Sum of P2030_LIN	817	10628	12238	1539	7790	434	33446
Sum of P2025_POLY	807	10614	12282	1535	7759	432	33429
Sum of P2030_POLY	827	11115	12708	1580	7901	438	34570
Sum of P2025_MEDIAN	804	10481	12163	1524	7725	431	33127
Sum of P2030_MEDIAN	822	10878	12474	1560	7847	436	34017

2.1.7 How to cite

Dataset:

Pesaresi, Martino; Politis, Panagiotis (2022): GHS-BUILT-S R2022A - GHS built-up surface grid, derived from Sentinel-2 composite and Landsat, multitemporal (1975-2030). European Commission, Joint Research Centre (JRC) [Dataset] DOI:10.2905/D07D81B4-7680-4D28-B896-583745C27085 PID: http://data.europa.eu/89h/d07d81b4-7680-4d28-b896-583745c27085

Concept & Methodology:

Schiavina M., Melchiorri M., Pesaresi M., Politis P., Freire S., Maffenini L., Florio P., Ehrlich D., Goch K., Tommasi P., Kemper T., GHSL Data Package 2022, Publications Office of the European Union, Luxembourg, 2022, ISBN 978-92-76-53071-8, doi:10.2760/19817, JRC 129516

Essential methodological background:

Pesaresi M, Corban C, Julea A, Florczyk A, Syrris V, Soille P. Assessment of the Added-Value of Sentinel-2 for Detecting Built-up Areas. Remote Sensing 8 (4); 2016. p. 299. JRC99996

Pesaresi M; Syrris V; Julea A. A New Method for Earth Observation Data Analytics Based on Symbolic Machine Learning. Remote Sensing 8 (5); 2016. p. 399. JRC99747

Pesaresi M. Global fine-scale information layers: the need of a paradigm shift. In: Soille P, Marchetti PG, editors. Proceedings of the 2014 conference on Big Data from Space (BiDS'14). Luxembourg (Luxembourg): Publications Office of the European Union; 2014. p. 8-11. JRC92345

Pesaresi M, Ouzounis G, Gueguen L. A new compact representation of morphological profiles: report on first massive VHR image processing at the JRC. In Conference Proceedings: Sylvia S. Shen, Paul E. Lewis, editors. Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVIII. Vol. 8390. SPIE; 2012. JRC70542

Gueguen L, Soille P, Pesaresi M. A New Built-Up Presence Index Based On Density of Corners. In Conference Proceedings: Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International. Piscataway (USA): IEEE; 2012. p. 5398-5401. JRC68582

Ouzounis G, Pesaresi M, Soille P. Differential area profiles: decomposition properties and efficient computation. IEEE Transactions on Pattern Analysis and Machine Intelligence 34 (8); 2012. p. 1533-1548. JRC59388

Pesaresi M, Gerhardinger A, Kayitakire F. A Robust Built-up Area Presence Index by Anisotropic Rotation-invariant Textural Measure. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 1 (3); 2008. p. 180-192. JRC37845

Pesaresi M, Benediktsson J. A New Approach for the Morphological Segmentation of High-Resolution Satellite Imagery.. IEEE Transactions on Geoscience and RS 39 (2); 2001. JRC19264

2.2 GHS-BUILT-H R2022A - GHS building height, derived from AW3D30, SRTM30, and Sentinel-2 composite (2018)

The spatial raster dataset depicts the distribution of the building heights generalized at the resolution of 100m, and referred to the year 2018. The input data used to predict the building heights are the ALOS Global Digital Surface Model "ALOS World 3D - 30m (AW3D30)¹⁹, the NASA Shuttle Radar Topographic Mission data - 30m (SRTM30)²⁰, and the Sentinel-2 global pixel based image composite from L1C data for the period 2017-2018²¹ that is the support of the year 2018 in this release.

The first global attempt to produce an estimate of the building heights was done in the so called "GHSL_LABEL" product as described in (Pesaresi, Ehrlich, et al., 2016) and tested in (Bechtel et al., 2018) in support to the urban local climate zones (LCZ) taxonomy. The use of the global DEM for prediction of building heights using linear regression techniques was introduced in (Pesaresi et al., 2021). In the application of those techniques included in the GHS-BUILT-H R2022A, the building heights are first predicted from the filtering of a composite of the AW3D30 and SRTM30 global DEMs, by generalization of the linear regression to a multiple-objective case targeting different estimates (ANBH, AGBH, and VOL) that are linked by analytical relations. They independently estimate the same variable (ANBH) from different filtering of the same DEMs and different regression coefficients, in conjunction with the BUSURF included in the GHS-BUILT-S R2022A data. The different independent estimates of the ANBH generated by the global DEMs are composed by minimization of the expected error, and they are used to estimate the linear regression coefficient predicting the ANBH from the density of shadow markers as observed in 10m-resolution Sentinel-2 image data, of each specific data tile (100x100km). These last ANBH predictions from the S2 data are used to update the final ANBH prediction included in this GHSL release.

2.2.1 Definitions

The built-up volume (BUVOL) is the volume of the built space above ground, expressed in cubic meters

Be S_x , the surface of the grid sample x

The Average of the Net Building Height of the sample x ($ANBH_x$) is defined as:

$$ANBH_x = BUVOL_x / BUSURFx$$

$$BUVOL_x = BUSURFx * ANBH_x$$

The Average of the Gross Building Height of the sample x ($AGBH_x$) is defined as:

$$AGBH_x = BUVOL_x / S_x$$

$$BUVOL_x = S_x * AGBH_x$$

Developing the above, the $BUFRAC_x$ can be derived as

$$AGBH_x / ANBH_x = BUSURFx / S_x = BUFRAC_x$$

Examples

Let assume a sample x of a grid with 100m of spatial resolution, then $S_x = 10,000$ square meters. The built-up surface in this sample it is predicted as $BUSURFx = 750$ square meters, corresponding to a $BUFRAC_x = 0.075$.

Moreover, in the same sample the $ANBH_x$ is predicted as $ANBH_x = 11.5$ meters.

The total built-up volume in this sample will be $BUVOL_x = BUSURFx * ANBH_x = 750 * 11.5 = 8625$ cubic meters.

In the same sample x, the $AGBH_x$ will be predicted as 0.8625 meters.

The total built-up volume in this sample will be $BUVOL_x = S_x * AGBH_x = 10,000 * 0.865 = 8625$ cubic meters

While the ratio $AGBH_x/ANBH_x = 0.075$, corresponds to the $BUFRAC_x$ in the same sample x

¹⁹ https://www.eorc.jaxa.jp/ALOS/en/dataset/aw3d30/aw3d30_e.htm

²⁰ https://cmr.earthdata.nasa.gov/search/concepts/C1000000240-LPDAAC_ECS.html

²¹ GHS-composite-S2 R2020A <https://ghsl.jrc.ec.europa.eu/download.php?ds=compositeS2>

2.2.2 Input data

ALOS Global Digital Surface Model "ALOS World 3D - 30m" (AW3D30) , the NASA Shuttle Radar Topographic Mission data - 30m (SRTM30) , and the Sentinel-2 global pixel based image composite from L1C data for the period 2017-2018.

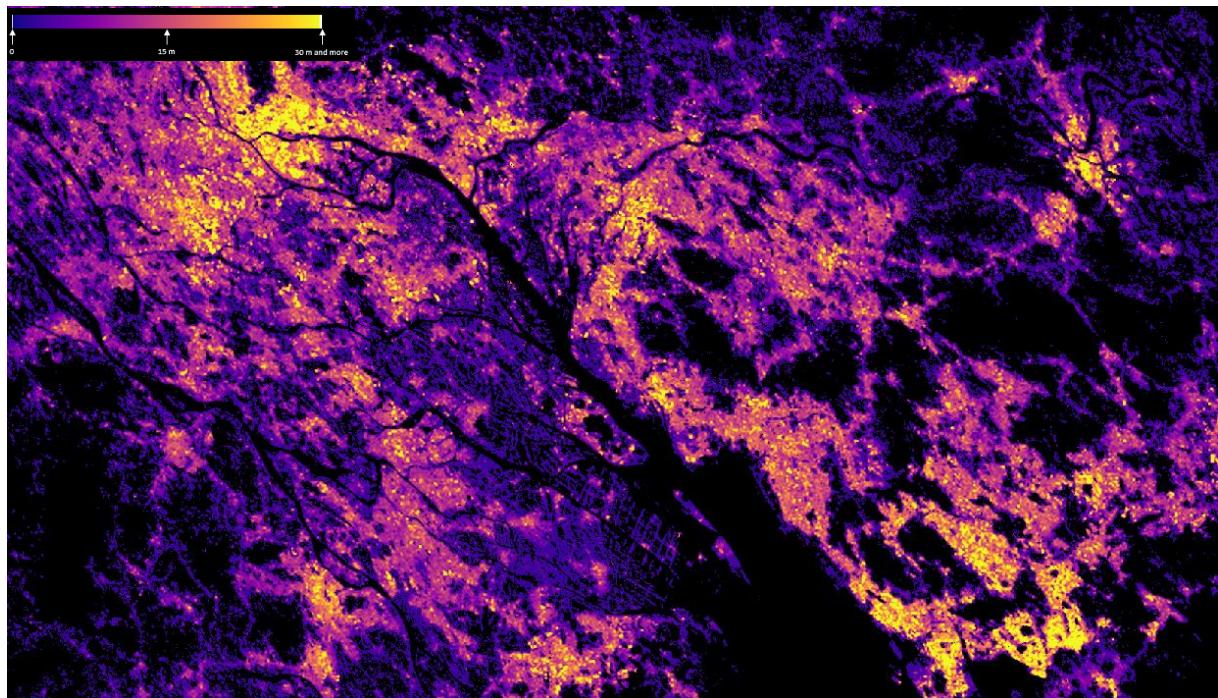


Figure 8 – Average building height (ANBH 100m) estimates in Guangzhou - Shenzhen (China).

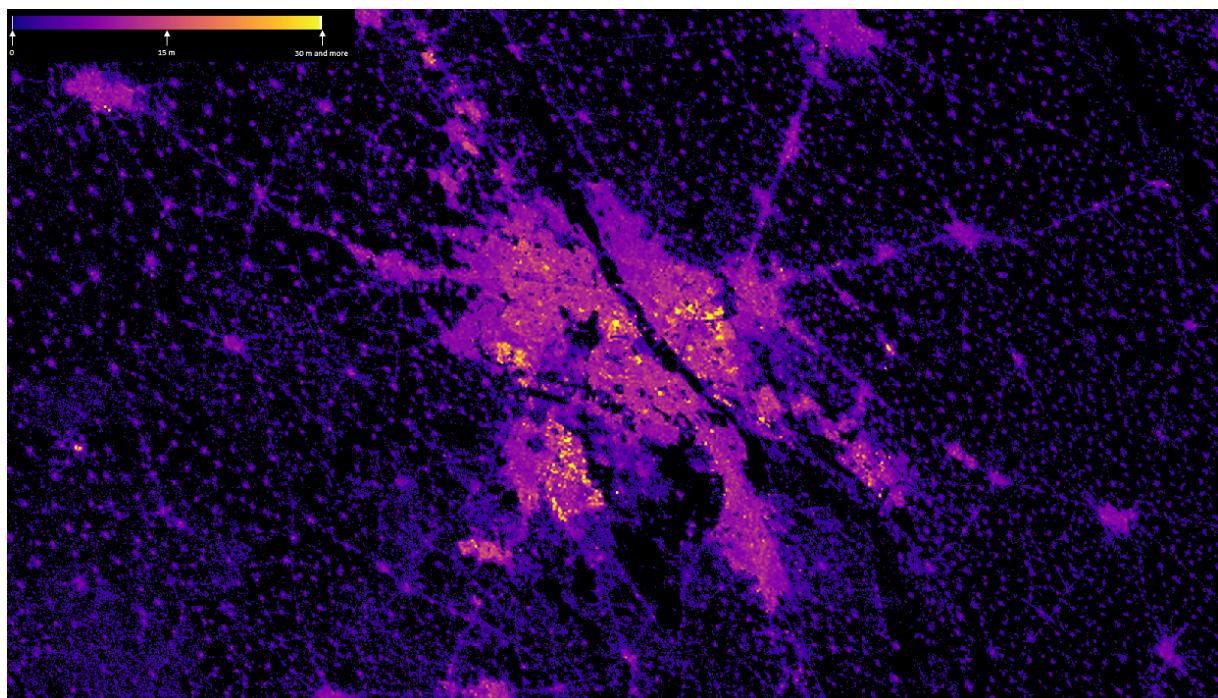


Figure 9 - Average building height (ANBH 100m) estimates in Delhi (India).

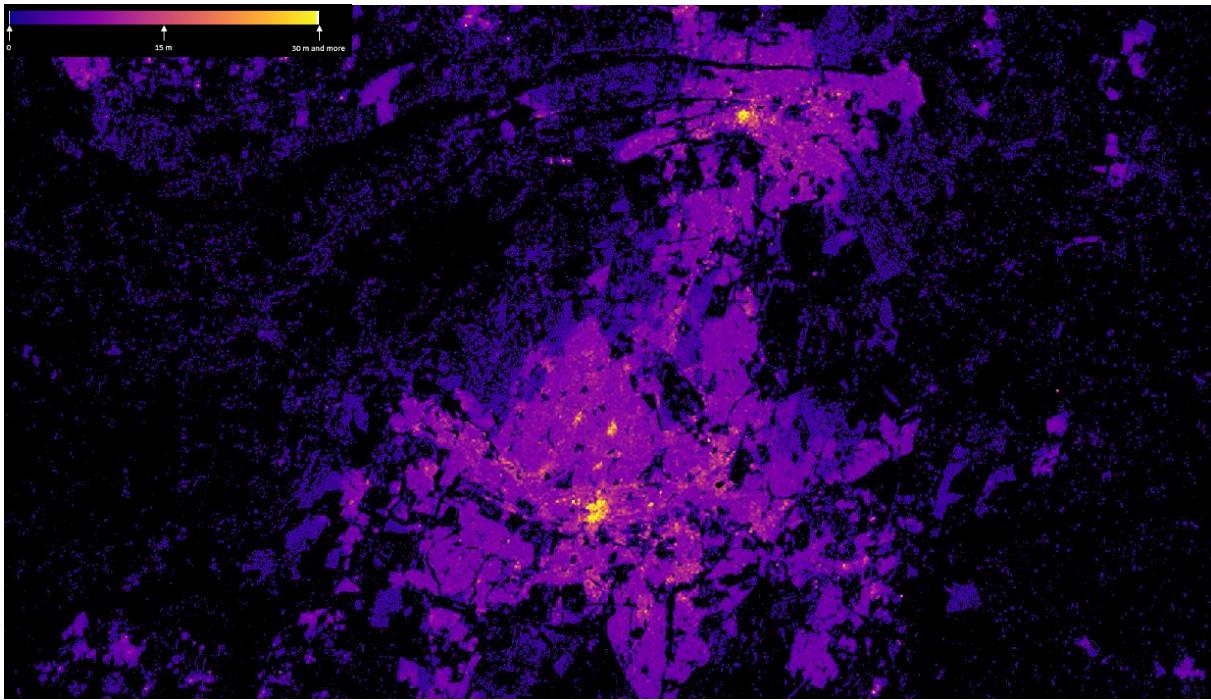


Figure 10 - Average building height (ANBH 100m) estimates in Pretoria – Johannesburg (South Africa).

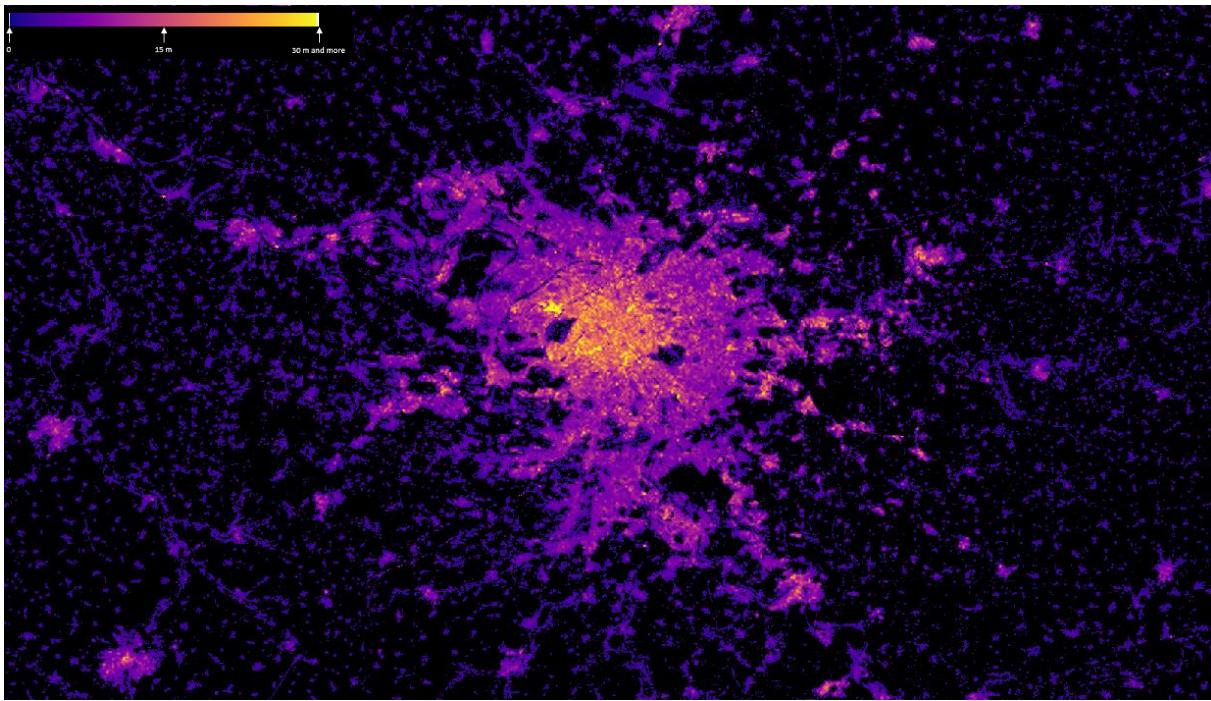


Figure 11 - Average building height (ANBH 100m) estimates in Paris (France).

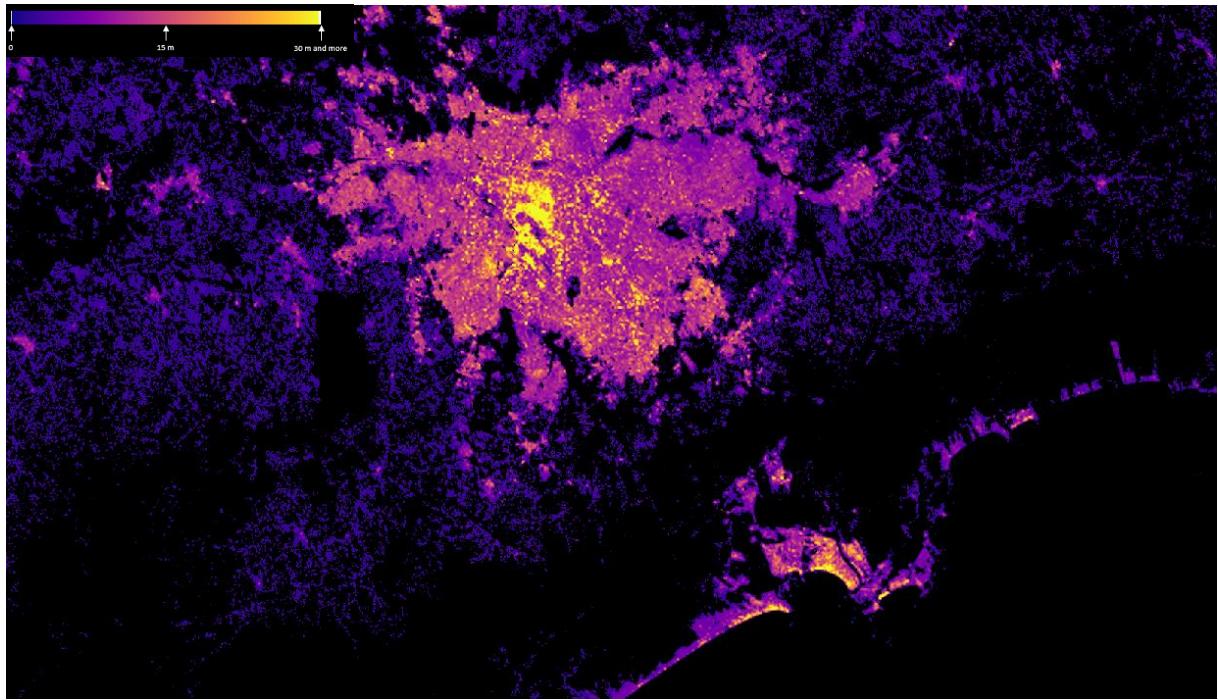


Figure 12 - Average building height (ANBH 100m) estimates in São Paulo – Santos (Brazil).

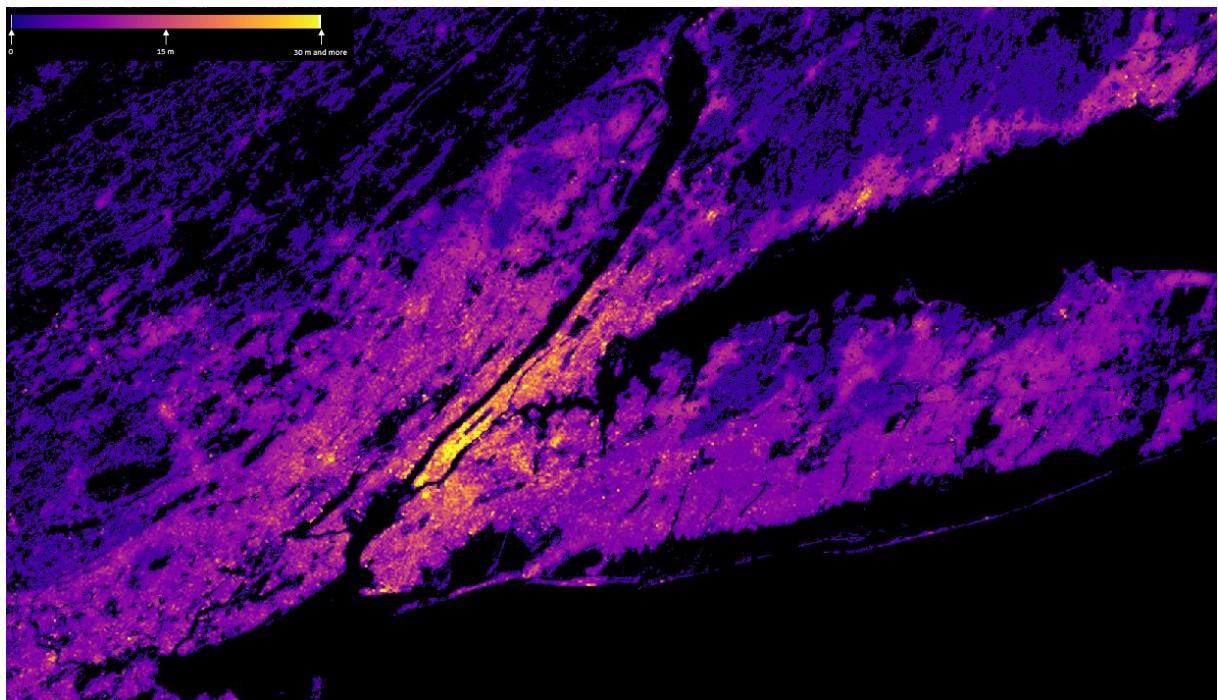


Figure 13 - Average building height (ANBH 100m) estimates in New York (United States).

2.2.3 Expected errors

The estimation of the GHS-BUILT-H error is currently ongoing, and will be delivered in peer-reviewed publication possibly in 2023-2024. A comparison with independent measurements not used in the model development and collected in 38 functional urban areas in the frame of the Copernicus Land Monitoring Service²², Urban Atlas, Building Height 2012, set the expected errors in predicting the ANBH included in this release as 1.97m of Mean Absolute Error (MAE) and 3.55m of Root Mean Square Error (RMSE), with a standard deviation of 0.87 and 1.25 meters, respectively.

Table 9 – Errors of ANBH 100m in predicting the Copernicus Building Height generalized at the same spatial resolution

TEST CASE	Observed ANBH max	Observed ANBH average	Observed vs. Predicted MAE	Observed vs. Predicted RMSE	Number of Samples
IS001L1_REYKJAVIK_UA2012_DHM_v010	43.17	5.58	0.24	1.00	103182
ME001L1_PODGORICA_UA2012_DHM_v010	54.48	5.79	0.33	1.22	137558
XK001L1_PRISTINA_UA2012_DHM_v010	43.00	7.00	0.59	1.87	51457
RS001L1_BEOGRAD_UA2012_DHM_v010	168.91	6.85	0.69	2.18	177645
AL001L1_TIRANA_UA2012_DHM_v010	64.53	6.24	0.72	1.78	147009
NO001L2_OSLO_UA2012_DHM_v010	55.60	6.81	1.02	2.30	49706
SI001L1_LJUBLJANA_UA2012_DHM_v010	42.90	6.33	1.05	2.19	26719
BA001L1_SARAJEVO_UA2012_DHM_v010	82.18	6.55	1.37	2.74	33597
SE001L1_STOCKHOLM_UA2012_DHM_v010	62.81	7.51	1.39	2.97	131914
SK001L1_BRATISLAVA_UA2012_DHM_v010	77.00	8.52	1.39	3.29	36005
HR001L2_GRAD_ZAGREB_UA2012_DHM_v010	66.61	7.84	1.43	2.85	63172
IE001L1_DUBLIN_UA2012_DHM_v010	43.98	6.63	1.43	2.48	92095
F1001L2_HELSINKI_UA2012_DHM_v010	50.34	7.33	1.65	2.98	78704
CY001L1_LEFKOSIA_UA2012_DHM_v010	25.79	7.33	1.65	2.68	20169
BG001L2_SOFRIDA_UA2012_DHM_v010	119.71	9.59	1.67	3.54	44083
CH004L1_BERN_UA2012_DHM_v010	70.00	8.17	1.79	3.66	13714
DK001L2_KOBENHAVN_UA2012_DHM_v020	124.43	7.14	1.82	3.23	58052
UK001L2_LONDON_UA2012_DHM_v020	131.60	7.66	2.07	3.36	182840
LT001L1_VILNIUS_UA2012_DHM_v010	165.00	8.14	2.08	4.07	38955
MK001L1_SKOPJE_UA2012_DHM_v010	96.58	8.28	2.10	4.11	19196
HU001L2_BUDAPEST_UA2012_DHM_v010	60.19	8.86	2.13	3.37	51940
IT001L2_ROMA_UA2012_DHM_v010	85.57	10.36	2.24	4.09	127649
DE001L1_BERLIN_UA2012_DHM_v020	77.00	9.42	2.26	3.95	111079
AT001L2_WIEN_UA2012_DHM_v010	151.97	9.79	2.29	3.90	42863
EL001L1_ATHINA_UA2012_DHM_v020	68.60	9.96	2.32	3.53	61187
LV001L0_RIGA_UA2012_DHM_v010	122.13	9.61	2.37	4.79	29770
ES001L2_MADRID_UA2012_DHM_v020	264.00	12.24	2.39	4.70	87854
CZ001L1_PRAHA_UA2012_DHM_v010	57.00	9.76	2.41	4.05	48665
NL002L2_AMSTERDAM_UA2012_DHM_v020	120.00	9.08	2.44	4.31	39733
EE001L1_TALLINN_UA2012_DHM_v010	112.20	8.32	2.49	4.33	15452
LU001L1_LUXEMBOURG_UA2012_DHM_v010	51.71	9.46	2.64	4.12	5253
PT001L2 LISBOA_UA2012_DHM_v010	78.27	8.39	2.71	4.23	70512
PL001L2_WARSZAWA_UA2012_DHM_v010	113.91	9.01	2.82	4.31	55279
FR001L1_PARIS_UA2012_DHM_v010	174.22	9.71	2.85	4.57	134717
MT001L1_VALLETTA_UA2012_DHM_v010	39.16	8.27	2.97	4.10	7200
BE001L2_BRUXELLES_BRUSSEL_UA2012_DHM_v010	86.94	10.48	3.17	4.82	20449
RO001L1_BUCURESTI_UA2012_DHM_v010	209.00	8.67	3.84	5.73	27442
TR001L1_ANKARA_UA2012_DHM_v010	162.65	12.06	3.92	7.57	72799
SUM					2515615
Average			1.97	3.55	
Standard Deviation			0.87	1.25	

²² <https://land.copernicus.eu/local/urban-atlas/building-height-2012>

2.2.4 Technical Details

Author: Pesaresi, Martino; Politis, Panagiotis

Product name: GHS-BUILT-H_GLOBE_R2022A

Spatial extent: Global

Temporal extent: 2018

Coordinate Systems: World Mollweide (ESRI:54009)

Spatial resolution available: 100m

Encoding: Values are expressed as decimals (Float32) reporting about the average height of the built surfaces in meters.

Data organisation ():* GeoTIFF file (100m) with overview images (OVR). Data tiles of 1000x1000 km size in GeoTIFF format (100 m)

Table 10 outlines the technical characteristics of the datasets released in this data package.

Table 10 - Technical details of the datasets in GHS-BUILT-H_GLOBE_R2022A

GHS-BUILT-H_GLOBE_R2022A		
ID	Description	Resolution (Projection/Coordinate system)
GHS_BUILT_H_AGBH_E2018_GLOBE_R2022A_54009_100_V1_0	Average of the Gross Building Height AGBH 2018 100m Encoding: Float32 NoData: 255	100 m World Mollweide (ESRI:54009)

GHS-BUILT-H_ANBH_GLOBE_R2022A		
ID	Description	Resolution (Projection/Coordinate system)
GHS_BUILT_H_ANBH_E2018_GLOBE_R2022A_54009_100_V1_0	Average of the Net Building Height ANBH 2018 100m Encoding: Float32 NoData: 255	100 m World Mollweide (ESRI:54009)

2.2.5 How to cite

Dataset:

Pesaresi, Martino; Politis, Panagiotis (2022): GHS-BUILT-H R2022A - GHS building height, derived from AW3D30, SRTM30, and Sentinel-2 composite (2018). European Commission, Joint Research Centre (JRC) [Dataset] DOI:10.2905/CE7C0310-9D5E-4AEB-B99E-4755F6062557 PID: <http://data.europa.eu/89h/ce7c0310-9d5e-4aeb-b99e-4755f6062557>

Concept & Methodology:

Pesaresi, M., Corbane, C., Ren, C., and Edward, N. (2021). Generalized Vertical Components of built-up areas from global Digital Elevation Models by multi-scale linear regression modelling. PLOS ONE 16, e0244478. <https://doi.org/10.1371/journal.pone.0244478>

Essential methodological background:

Pesaresi M, Ehrlich D, Ferri S, Florczyk A, Carneiro Freire S, Halkia S, Julea A, Kemper T, Soille P, Syrris V. Operating procedure for the production of the Global Human Settlement Layer from Landsat data of the epochs 1975, 1990, 2000, and 2014. EUR 27741. Luxembourg (Luxembourg): Publications Office of the European Union; 2016. JRC97705

Bechtel, B., Pesaresi, M., Florczyk, A. and Mills, G., Beyond built-up – information on the internal makeup of urban areas, 2017, ISBN 978-1-138-05460-8, JRC107558

2.3 GHS-BUILT-V R2022A - GHS built-up volume grids derived from joint assessment of Sentinel-2, Landsat, and global DEM data, for 1975-2030 (5yrs interval)

The spatial raster dataset depicts the distribution of built-up volumes, expressed as number of cubic meters. The data reports about the total built-up volume and the built-up volume allocated to dominant non-residential (NRES) uses. Data are spatial-temporal interpolated from 1975 to 2030 in 5 years intervals. Future predicted grids at the epochs 2025 and 2030 are provided accordingly to three scenarios extrapolating the past trends by a linear (LIN), a second-order polynomial (PLY), and a median solution (MED).

The data is made by application of the formula $BUVOLx = ANBHx * BUSURFx$, where:

the ANBHx is the GHS-BUILT-H R2022A - GHS building height, derived from AW3D30, SRTM30, and Sentinel-2 composite (2018),

the BUSURFx is the GHS-BUILT-S R2022A - GHS built-up surface grid, derived from Sentinel-2 composite and Landsat, multitemporal (1975-2030).

2.3.1 Input Data

GHS-BUILT-H R2022A - GHS building height, derived from AW3D30, SRTM30, and Sentinel-2 composite (2018)

GHS-BUILT-S R2022A - GHS built-up surface grid, derived from Sentinel-2 composite and Landsat, multitemporal (1975-2030).

2.3.2 Technical Details

Author: Pesaresi, Martino; Politis Panagiotis

Product name: GHS-BUILT-V_GLOBE_R2022A

Spatial extent: Global

Temporal extent: from 1975 to 2030, 5 years interval

*Coordinate Systems**: World Mollweide (ESRI:54009)

*Spatial resolution available**: 100 m, 1 km

*Encoding**: integers (UInt32), unit: built cubic meters in the grid cell

Data organisation ()*: Global GeoTIFF file (100m, 1km) with overview images (OVR). Data tiles of 1000x1000 km size in GeoTIFF format (100 m, 1 km)

(*) product dependent, see Table 11.

Disclaimer: the re-projection of the World Mollweide version of the GHS-BUILT-V_GLOBE_R2022A to coordinate systems requires specific technical knowledge. No responsibility is taken for workflows developed independently by users.

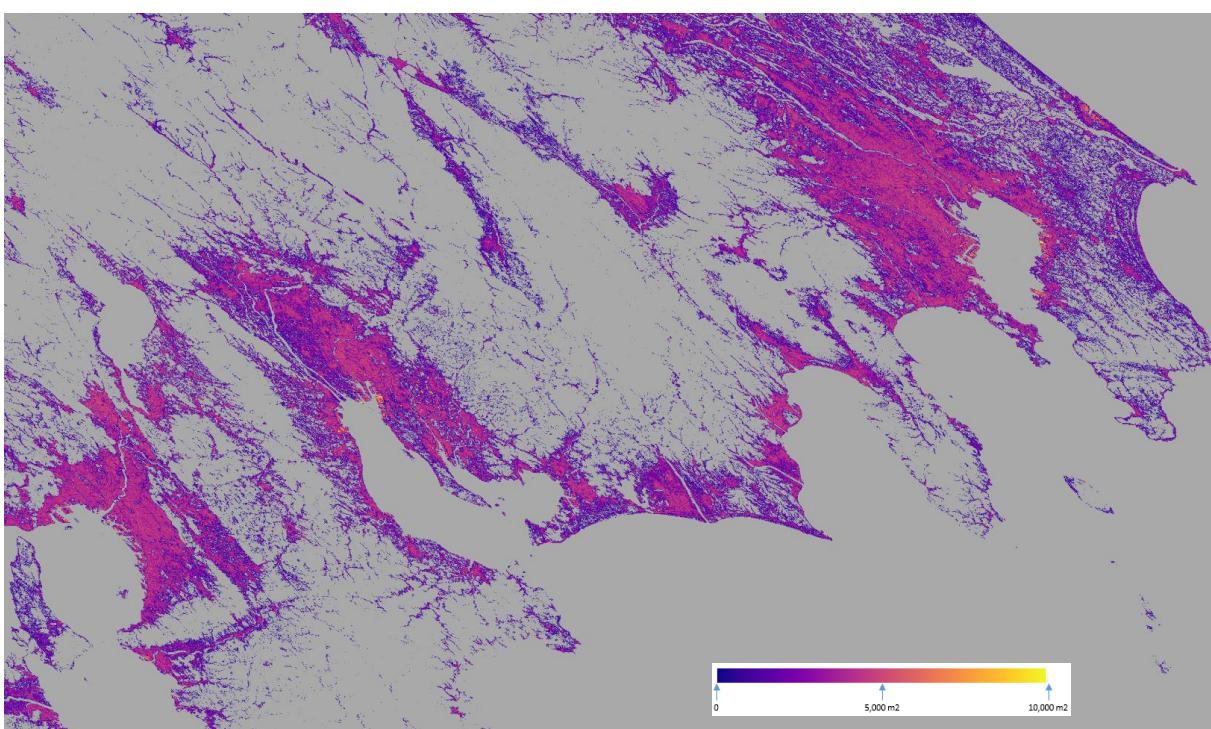
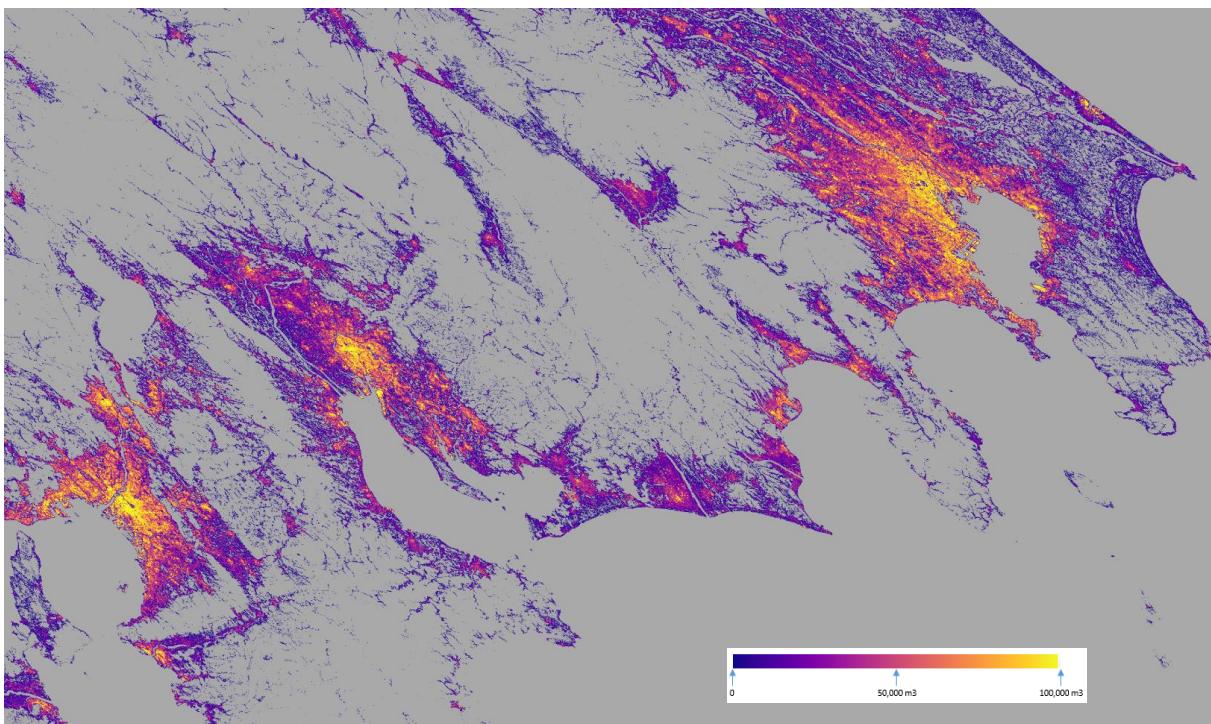


Figure 14 - Osaka-Nagoya-Tokyo (Japan): comparison between GHS-BUILT-V R2022A (above) and GHS-BUILT-S R2022A (below), year 2020.

Table 11 - Technical details of the datasets in GHS-BUILT-V_GLOBE_R2022A

GHS-BUILT-V_GLOBE_R2022A		
ID	Description	Resolution (projection)
GHS_BUILT_V_E<epoch>_GLOBE_R2022A_54009_<res>_V1_0	BU volume <epoch> 1975-2020; <res> 100, 1000 Encoding: UInt32 Values range: 0-Inf NoData: 4294967295	100 m, 1 km World Mollweide (ESRI:54009)
GHS_BUILT_V_P<epoch&model>_GLOBE_R2022A_54009_<res>_V1_0	BU volume projections <epoch> 1925-2030; <model> LIN-PLY-MED; <res> 100, 1000 Encoding: UInt32 Values range: 0-Inf NoData: 4294967295	100 m, 1 km World Mollweide (ESRI:54009)

GHS-BUILT-V_NRES_GLOBE_R2022A		
ID	Description	Resolution (projection)
GHS_BUILT_V_NRES_E<epoch>_GLOBE_R2022A_54009_<res>_V1_0	Non-residential BU volume <epoch> 1975-2020;; <res> 100, 1000 Encoding: UInt32 Values range: 0-Inf NoData: 4294967295	100 m, 1 km World Mollweide (ESRI:54009)
GHS_BUILT_V_NRES_P<epoch&model>_GLOBE_R2022A_54009_<res>_V1_0	Non-residential BU volume projections <epoch> 1925-2030; <model> LIN-PLY-MED; <res> 100, 1000 Encoding: UInt32 Values range: 0-Inf NoData: 4294967295	100 m, 1 km World Mollweide (ESRI:54009)

2.3.3 Summary statistics

Table 12 - Summary statistics of total built-up volume in millions of cubic meters, by continent. Years 2025 and 2030 are reported for the linear (LIN), second-order polynomial (PLY), and median (MED) future projection scenarios

Values	Africa	Asia	Europe	Latin America and the Caribbean	Northern America	Oceania	Grand Total
Sum of E1975	122516	587476	259204	153666	201170	25662	1349693
Sum of E1980	127929	618333	264785	158366	207203	26171	1402788
Sum of E1985	135574	662980	272507	164855	216055	26883	1478853
Sum of E1990	145078	719684	281863	172871	227367	27793	1574655
Sum of E1995	155979	768442	292799	182614	240082	28634	1668549
Sum of E2000	172583	842934	309293	196806	262414	30024	1814054
Sum of E2005	188714	895069	326903	205843	275453	31098	1923080
Sum of E2010	213460	969195	351558	218987	295447	32733	2081380
Sum of E2015	245372	1062645	383608	235164	322142	34834	2283765
Sum of E2020	285996	1180518	430521	253879	361412	37630	2549955
Sum of P2025_LIN	304093	1223316	447833	263990	375378	38567	2653177
Sum of P2030_LIN	315995	1252006	460364	271301	385168	39306	2724140
Sum of P2025_POLY	326250	1268862	470284	277321	393074	39940	2775731
Sum of P2030_POLY	361866	1341456	502899	297929	418951	42040	2965140
Sum of P2025_MEDIAN	315632	1247051	459578	270808	384612	39268	2716949
Sum of P2030_MEDIAN	338293	1296649	482064	284652	402296	40677	2844632

Table 13 - Summary statistics of non-residential (NRES) built-up volume in millions of cubic meters, by continent. Years 2025 and 2030 are reported for the linear (LIN), second-order polynomial (PLY), and median (MED) future projection scenarios

Values	Africa	Asia	Europe	Latin America and the Caribbean	Northern America	Oceania	Grand Total
Sum of E1975	3637	37654	56104	7354	38906	2629	146284
Sum of E1980	3817	39483	57508	7565	39858	2701	150933
Sum of E1985	4177	43663	60163	7987	41813	2836	160639
Sum of E1990	4750	50763	64167	8672	45025	3049	176427
Sum of E1995	4906	54236	66319	9022	46981	3120	184585
Sum of E2000	5307	63735	71119	9848	52242	3280	205531
Sum of E2005	5414	66294	72913	10130	53137	3338	211226
Sum of E2010	5607	71697	75945	10645	54855	3442	222191
Sum of E2015	5943	81310	80531	11466	57630	3601	240481
Sum of E2020	6458	93786	85657	12377	60645	3753	262676
Sum of P2025_LIN	6583	96355	86565	12569	61269	3780	267122
Sum of P2030_LIN	6679	98386	87300	12723	61773	3802	270663
Sum of P2025_POLY	6609	98172	87406	12681	61579	3788	270235
Sum of P2030_POLY	6720	101504	88904	12923	62234	3816	276101
Sum of P2025_MEDIAN	6596	97289	86990	12626	61428	3784	268713
Sum of P2030_MEDIAN	6698	99999	88109	12825	62013	3809	273453

2.3.4 How to cite

Dataset:

Pesaresi, Martino; Politis, Panagiotis (2022): GHS-BUILT-V R2022A - GHS built-up volume grids derived from joint assessment of Sentinel-2, Landsat, and global DEM data, for 1975-2030 (5yrs interval). European Commission, Joint Research Centre (JRC) [Dataset] DOI:10.2905/7A1F6B8A-D520-49A7-8F58-D5AC936B9C8A PID: http://data.europa.eu/89h/7a1f6b8a-d520-49a7-8f58-d5ac936b9c8a

Concept & Methodology:

Schiavina M, Melchiorri M, Pesaresi M, Politis P, Freire S, Maffenini L, Florio P, Ehrlich D, Goch K, Tommasi P, Kemper T, GHSL Data Package 2022, Publications Office of the European Union, Luxembourg, 2022, ISBN 978-92-76-53071-8, doi:10.2760/19817, JRC 129516

Essential methodological background:

Pesaresi, M., Corbane, C., Ren, C., and Edward, N. (2021). Generalized Vertical Components of built-up areas from global Digital Elevation Models by multi-scale linear regression modelling. PLOS ONE 16, e0244478. <https://doi.org/10.1371/journal.pone.0244478>

2.4 GHS-BUILT-C R2022A - GHS Settlement Characteristics, derived from Sentinel-2 composite (2018) and other GHS R2022A data

The GHS-BUILT-C spatial raster datasets delineate the boundaries of the human settlements at 10m resolution, and describe their inner characteristics in terms of the morphology of the built environment and the functional use.

The Morphological Settlement Zone (MSZ) delineates the spatial domain of all the human settlements at the neighbouring scale of approx. 100m, based on the spatial generalization of the built-up surface fraction (BUFRAC) function. The objective is to fill the open spaces that are surrounded by large patches of built space. MSZ, open spaces, and built spaces basic class abstractions are derived by mathematical morphology spatial filtering (opening, closing, regional maxima) from the BUFRAC function. They are further classified accordingly to the information regarding vegetation intensity (GHS-BUILT-C_VEG_GLOBE_R2022A), water surfaces (GHS_LAND_GLOBE_R2022A), road surfaces (OSM highways), functional use (GHS-BUILT-C_FUN_GLOBE_R2022A), and building height (GHS-BUILT-H_GLOBE_R2022A).

GHS_BUILT_C_MSZ_E2018_GLOBE_R2022A_54009_10_V1_0 describes the MSZ delineation and their inner characteristics

GHS_BUILT_C_FUN_E2018_GLOBE_R2022A_54009_10_V1_0 describes the RES vs. NRES classification from the Sentinel-2 image data

GHS_BUILT_C_VEG_E2018_GLOBE_R2022A_54009_10_V1_0 describes the Vegetation intensity in the MSZ

Morphological Settlement Zone delineation

First, the spatial domain it is split in two sub-domain, “open spaces” vs. “built spaces”. The “Built-spaces” set is derived from the BUFRAC continuous function, filtered for detecting salient raster samples that may summarize the function itself, also called “built-up markers” (BUMARKER). The BUMARKER set it is defined by the union of the morphological regional maxima (Vincent, 1993) of the BUFRAC function (4-conn definition), with the raster samples dominated by the built-up surface class, so that $\text{BUFRAC}_x > 0.5$. The “open space” set is defined as the logical complement of the “built space” set.

Second, the BUMARKER set it is subdivided in the “COMPACT” vs. the “SPARSE” sets based on morphological filtering. The COMPACT set is defined by the opening of the closing of the BUMARKER, with a structuring element of a disk with 5 pixel (50 meters) radius. The SPARSE set is made by the BUMARKER in the logical complement of the COMPACT set.

Third, the Morphological Settlement Zone (MSZ) it is defined as the Union of the hole-filled COMPACT set using a 4-connectivity rule and the dilation of the SPARSE set, using a minimal structure element of 2x2 (20x20m) pixels.

Fourth, the classification is applied to the samples belonging to the MSZ domain accordingly to the information regarding vegetation intensity (GHS-BUILT-C_VEG_GLOBE_R2022A), water surfaces (GHS_LAND_GLOBE_R2022A), road surfaces (OSM highways), functional use (GHS-BUILT-C_FUN_GLOBE_R2022A), and building height (GHS-BUILT-H_GLOBE_R2022A). The building height information it is downscaled to the 10m-resolution using the connected components of the BUMARKERS as spatial units.

- 01 : MSZ, open spaces, low vegetation surfaces $\text{NDVI} \leq 0.3$
- 02 : MSZ, open spaces, medium vegetation surfaces $0.3 < \text{NDVI} \leq 0.5$
- 03 : MSZ, open spaces, high vegetation surfaces $\text{NDVI} > 0.5$
- 04 : MSZ, open spaces, water surfaces $\text{LAND} < 0.5$
- 05 : MSZ, open spaces, road surfaces
- 11 : MSZ, built spaces, residential, building height $\leq 3\text{m}$
- 12 : MSZ, built spaces, residential, $3\text{m} < \text{building height} \leq 6\text{m}$
- 13 : MSZ, built spaces, residential, $6\text{m} < \text{building height} \leq 15\text{m}$
- 14 : MSZ, built spaces, residential, $15\text{m} < \text{building height} \leq 30\text{m}$
- 15 : MSZ, built spaces, residential, building height $> 30\text{m}$
- 21 : MSZ, built spaces, non-residential, building height $\leq 3\text{m}$
- 22 : MSZ, built spaces, non-residential, $3\text{m} < \text{building height} \leq 6\text{m}$
- 23 : MSZ, built spaces, non-residential, $6\text{m} < \text{building height} \leq 15\text{m}$
- 24 : MSZ, built spaces, non-residential, $15\text{m} < \text{building height} \leq 30\text{m}$
- 25 : MSZ, built spaces, non-residential, building height $> 30\text{m}$

Figure 15 - Morphological Settlement Zone (MSZ) Legend

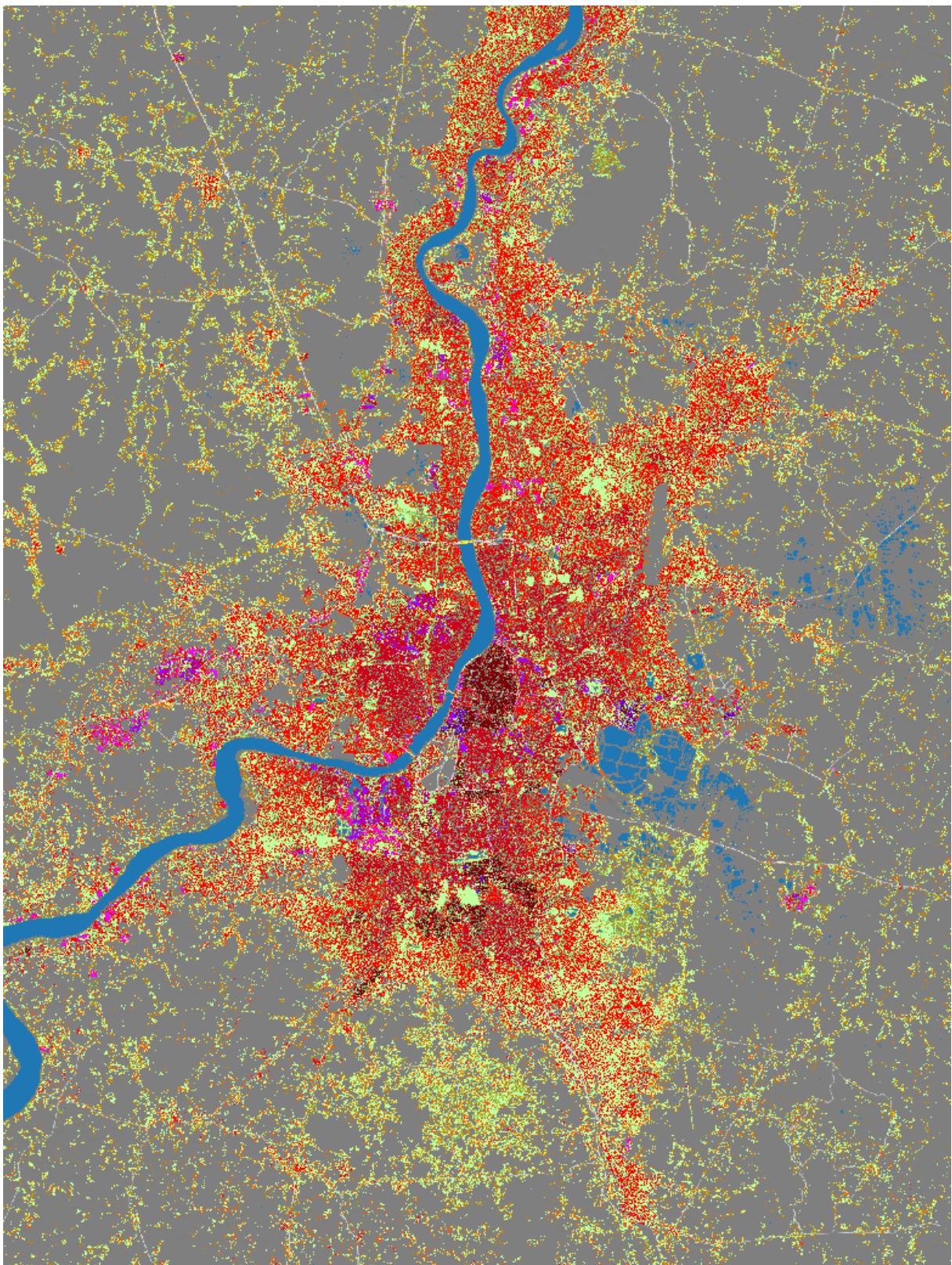


Figure 16 - Settlement Characteristics in Kolkata (India)

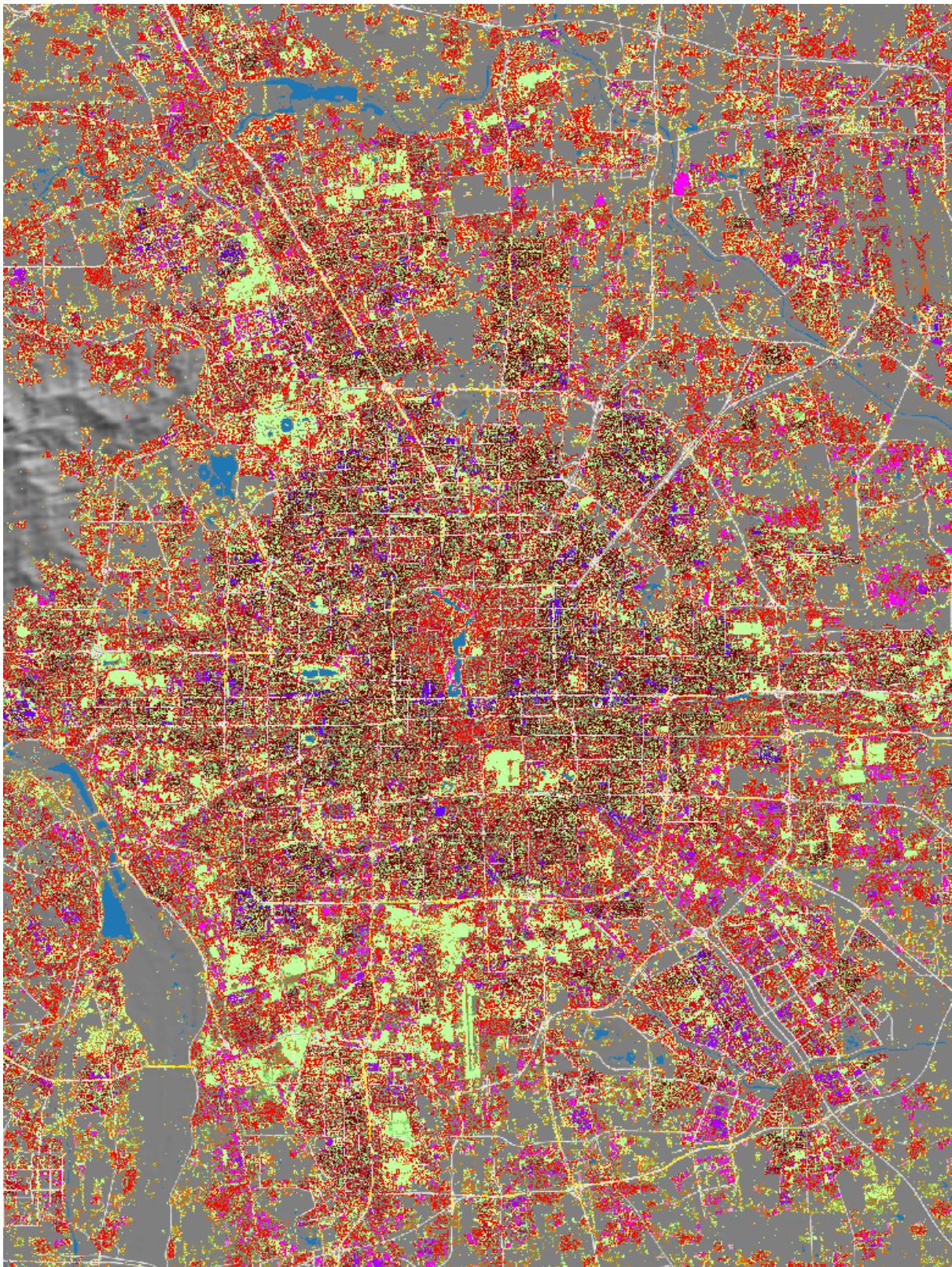


Figure 17 - Settlement Characteristics in Beijing (China)

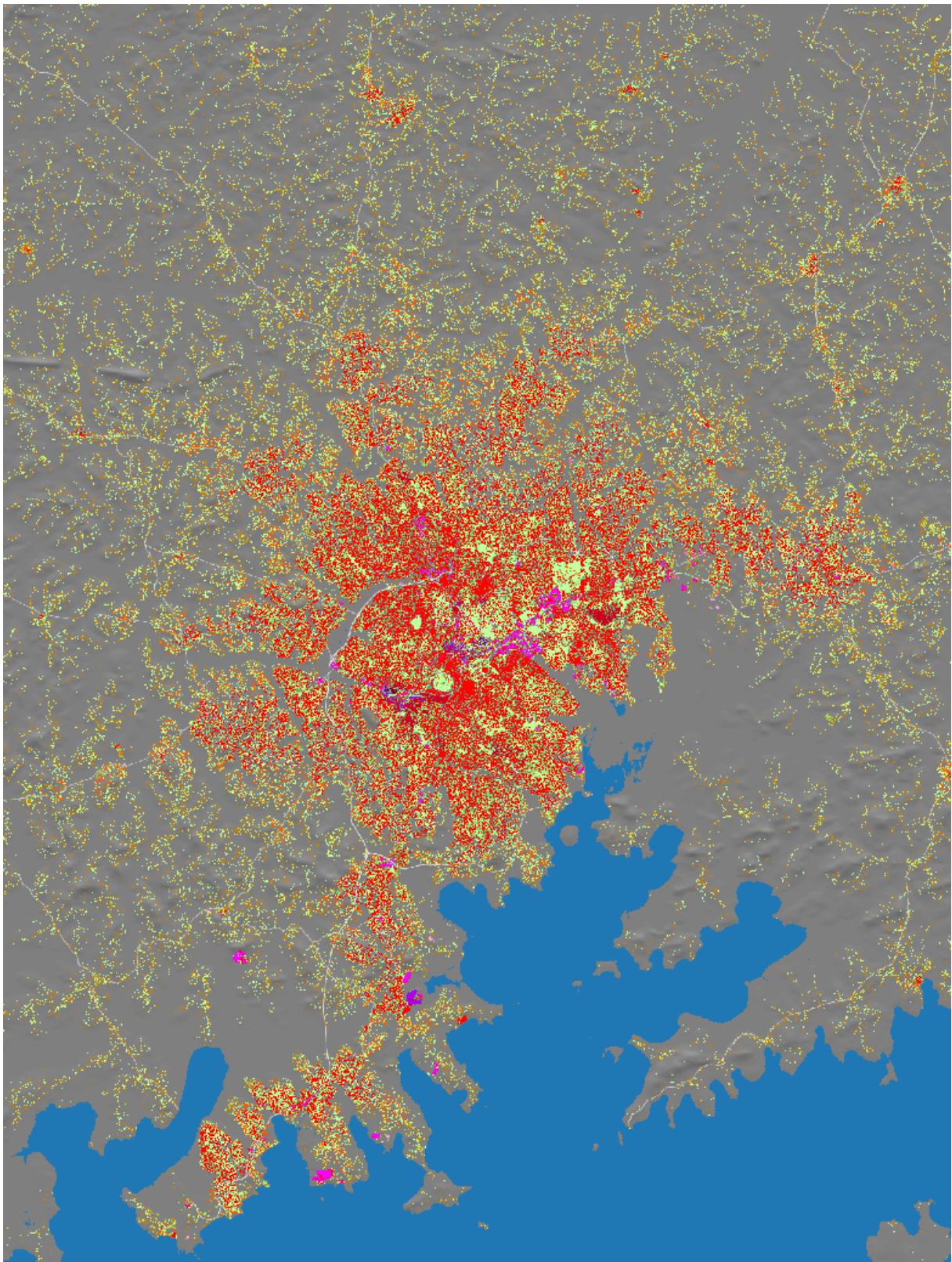


Figure 18 - Settlement Characteristics in Kampala (Uganda)

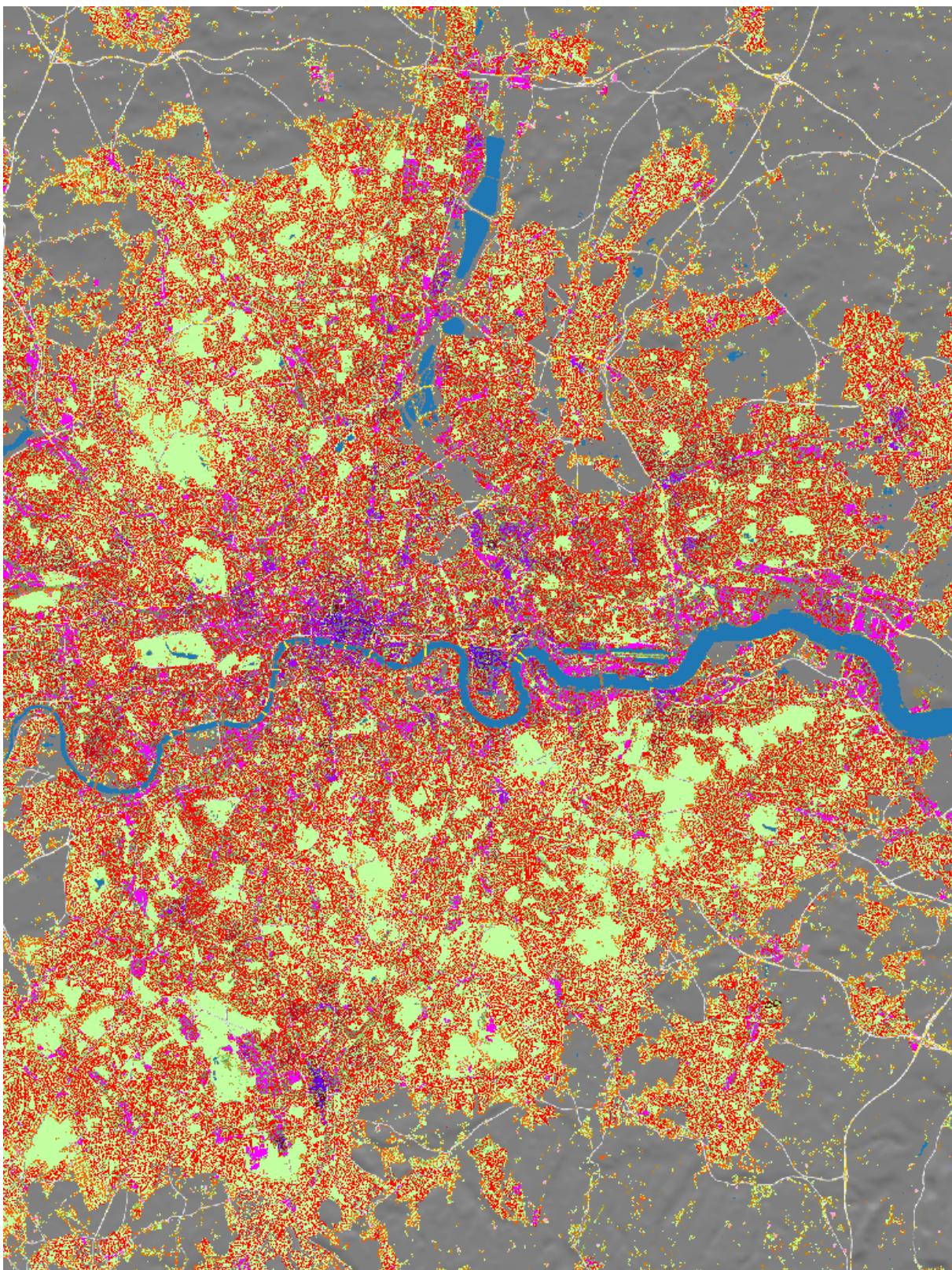


Figure 19 – Settlement Characteristics in London (United Kingdom)

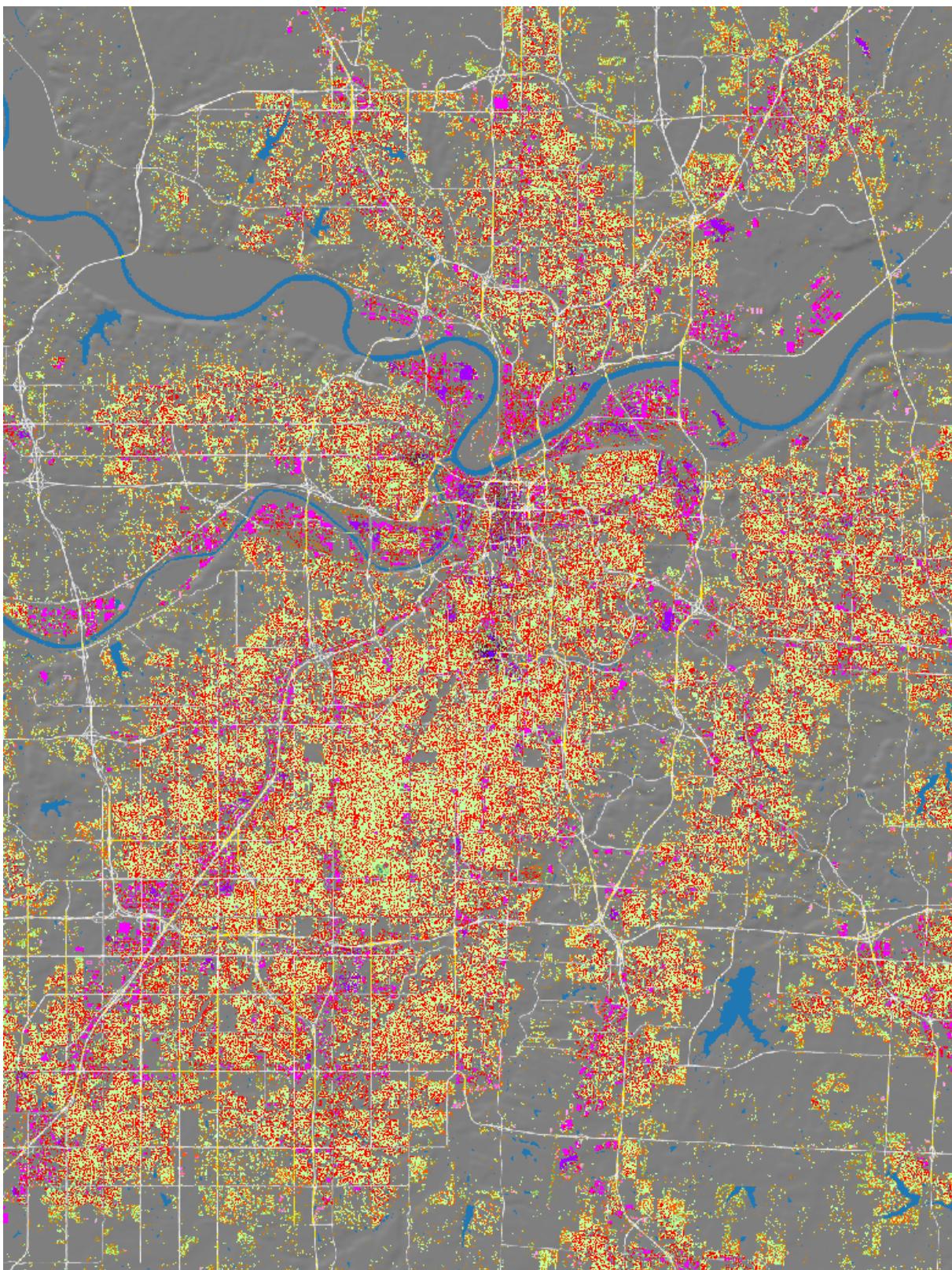


Figure 20 – Settlement Characteristics in Kansas City (United States)

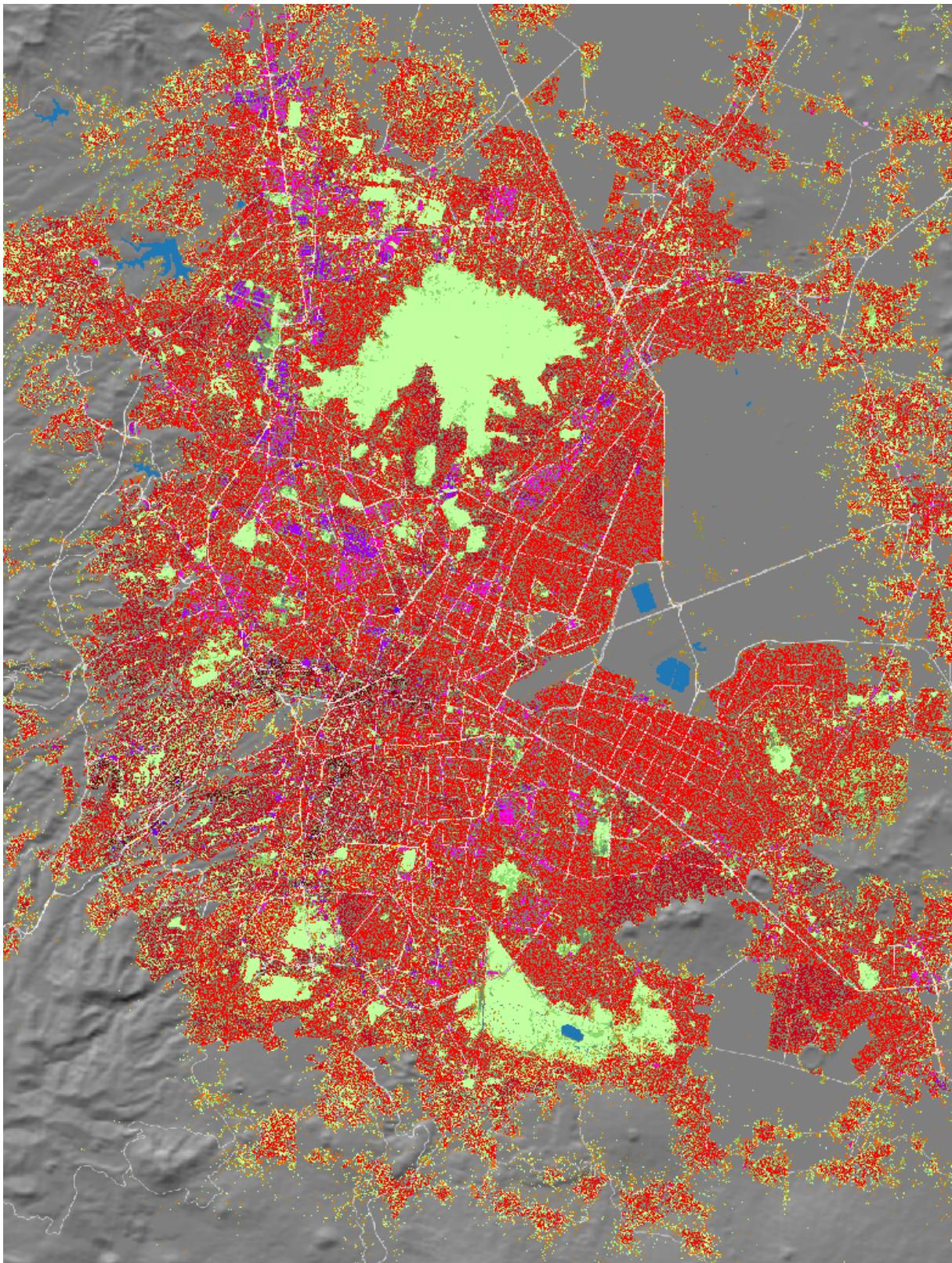


Figure 21 – Settlement Characteristics in Mexico City (Mexico)

2.4.1 Input data

Morphological filtering of the BUFRAC 10m-resolution GHS_BUILT_S_E2018_GLOBE_R2022A_54009_10_V1_0. Vegetation intensity (GHS-BUILT-C_VEG_GLOBE_R2022A), water surfaces (GHS_LAND_GLOBE_R2022A), road surfaces (OSM highways), functional use (GHS-BUILT-C_FUN_GLOBE_R2022A), and building height (GHS-BUILT-H_GLOBE_R2022A).

2.4.2 Technical Details

Author: Pesaresi, Martino; Politis Panagiotis

Product name: GHS-BUILT-C_GLOBE_R2022A

Spatial extent: Global

Temporal extent: 2018

Coordinate Systems: World Mollweide (ESRI:54009)

Spatial resolution available: 10m

Encoding: Integer (Byte)

Data organisation: Global VRT file (10 m) with overview images (OVR). Data tiles of 1000x1000 km size in GeoTIFF format (10 m)

Table 14 - Technical details of the datasets in GHS-BUILT-C_GLOBE_R2022A

GHS-BUILT-C_MSZ_GLOBE_R2022A		
ID	Description	Resolution (Projection/Coordinate system)
GHS_BUILT_C_MSZ_E2018_GLOBE_R2022A_54009_10_V1_0	Morphological Settlement Zone (MSZ) : classification by internal surface (soil sealed) and functional (BU RES/NRES) characteristics - 10m Encoding: Byte Values range: 1-25 (see Figure 15) NoData: 255	10 m World Mollweide (ESRI:54009)

GHS-BUILT-C_FUN_GLOBE_R2022A		
ID	Description	Resolution (Projection/Coordinate system)
GHS_BUILT_C_FUN_E2018_GLOBE_R2022A_54009_10_V1_0	Residential (RES) vs. non-residential (NRES) functional classification of the built spatial domain defined as BUFRAC > 0 Encoding: Byte Values range: 0 (non-BU), 1 (RES), 2 (NRES) NoData: 255	10 m World Mollweide (ESRI:54009)

GHS-BUILT-C_VEG_GLOBE_R2022A		
ID	Description	Resolution (Projection/Coordinate system)
GHS_BUILT_C_VEG_E2018_GLOBE_R2022A_54009_10_V1_0	<p>Vegetation intensity in the Morphological Settlement Zone (MSZ)</p> <p>Encoding: Byte</p> <p>Values range: 0-100</p> <p>NoData: 255</p>	<p>10 m</p> <p>World Mollweide (ESRI:54009)</p>

2.4.3 How to cite

Dataset:

Pesaresi, Martino; Politis, Panagiotis (2022): GHS-BUILT-C R2022A - GHS Settlement Characteristics, derived from Sentinel-2 composite (2018) and other GHS R2022A data. European Commission, Joint Research Centre (JRC) [Dataset] DOI:10.2905/DDE11594-2A66-4C1B-9A19-821382AED36E PID: http://data.europa.eu/89h/dde11594-2a66-4c1b-9a19-821382aed36e

Concept & Methodology:

Schiavina M., Melchiorri M., Pesaresi M., Politis P., Freire S., Maffenini L., Florio P., Ehrlich D., Goch K., Tommasi P., Kemper T., GHSL Data Package 2022, Publications Office of the European Union, Luxembourg, 2022, ISBN 978-92-76-53071-8, doi:10.2760/19817, JRC 129516

2.5 GHS-LAND R2022A - GHS land surface grid, derived from Sentinel-2 composite (2018), Landsat Panchromatic data (2014), and OSM data

The GHS-LAND spatial raster dataset depicts the distribution of the land surfaces estimates by joint assessment of satellite image data and Open Street Map (OSM) data. The main satellite data used in input is the GHS_composite_S2_L1C_2017-2018_GLOBE_R2020A that corresponds to global cloud-free pixel based composite created from the Sentinel-2 data archive (Level L1C) available in Google Earth Engine for the period January 2017 - December 2018 available at 10m resolution and reporting about the R,G,B and NIR bands of the multispectral instrument on board the Sentinel satellite platform. The gaps of these data for covering the expected global landmass were filled using the panchromatic band of the Landsat images used to support the epoch 2014 of the GHS-BUILT-S R2022A. The remaining gaps where no S2 data and no Landsat data were available, were filled by direct imposing of the OSM data if available.

The general approach taken for the production of this layer can be summarized as below

First, generate a radiometric index W_x sensitive to the presence of water. From Sentinel-2 data W_x was calculated as ratio between the luminance and the near-infrared reflectance $W_x = \text{LUM}_x / \text{NIR}_x$, assuming the luminance LUM_x as the point-wise maxima of the R,G,B reflectance values. In case of the Landsat P data, W_x it was the P band itself.

Second, look for salient thresholds in W_x based on the statistical distribution (mean, stdv) of W_x in available training sets. A composite of the OSM water bodies and the global surface water data (Pekel et al., 2016) was used as training set purposes.

Third, apply the salient cut-off-values to W_x and filter the results by morphological (linear continuity) criteria and by maximal expected slope in detected water surfaces. The Copernicus DEM - Global and European Digital Elevation Model (COP-DEM)²³ was used for the slope prediction purpose.

Fourth, thin water streams not detected by the satellite data and included in the OSM vector data were imposed to the raster grid

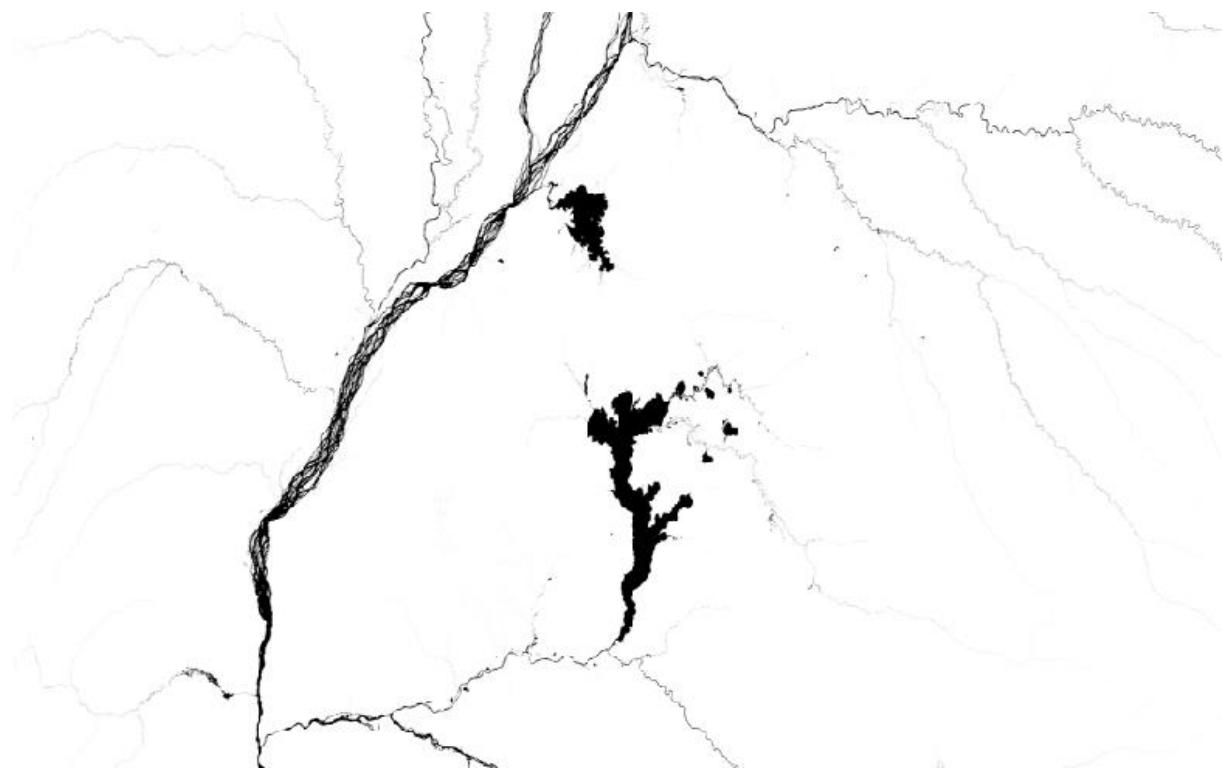


Figure 22 - GHS LAND in the Congo River

²³ <https://spacedata.copernicus.eu/web/cscda/dataset-details?articleId=394198>

2.5.1 Input data

Sentinel-2 (S2) image composite (GHS-composite-S2 R2020A) of the year 2018

Open Street Map (OSM) data on water polygons²⁴

2.5.2 Technical Details

Author: Pesaresi, Martino; Politis Panagiotis

Product name: GHS-LAND_GLOBE_R2022A

Spatial extent: Global

Temporal extent: 2018

Coordinate Systems: World Mollweide (ESRI:54009)

Spatial resolution available: 10m, 100m, 1km

Encoding: integer (Byte, UInt16, UInt32), unit: land square meters in the grid cell

Data organisation: Global VRT file (at 10 m) or GeoTIFF file (100m, 1km) with overview images (OVR). Data tiles of 1000x1000 km size in GeoTIFF format (10 m, 100 m, 1 km)

Table 15 - Technical details of the datasets in GHS-LAND_GLOBE_R2022A

GHS-LAND_GLOBE_R2022A		
ID	Description	Resolution (Projection/Coordinate system)
GHS_LAND_E2018_GLOBE_R2022A_54009_<res>_V1_0	<p>GHS land surface grid, derived from Sentinel-2 composite (2018), Landsat Panchromatic data (2014), and OSM data <res> 10, 100, 1000</p> <p>Encoding: Byte (10 m), UInt16 (100 m), UInt32 (1 km)</p> <p>Values range: 0-100 (10 m), 0-10000 (100 m), 0-1000000 (1 km)</p> <p>NoData: 255 (10 m), 65535 (100 m), 4294967295 (1 km)</p>	10 m, 100 m, 1 km World Mollweide (ESRI:54009)

2.5.3 How to cite

Dataset:

Pesaresi, Martino; Politis, Panagiotis (2022): GHS-LAND R2022A - Land fraction as derived from Sentinel-2 image composite (2018) and OSM data. European Commission, Joint Research Centre (JRC) [Dataset] DOI:10.2905/AB7AD451-5ED5-44A6-A4D0-9F7A4E848CEE PID: <http://data.europa.eu/89h/ab7ad451-5ed5-44a6-a4d0-9f7a4e848cee>

Concept & Methodology:

Schiavina M., Melchiorri M., Pesaresi M., Politis P., Freire S., Maffenini L., Florio P., Ehrlich D., Goch K., Tommasi P., Kemper T., GHSL Data Package 2022, Publications Office of the European Union, Luxembourg, 2022, ISBN 978-92-76-53071-8, doi:10.2760/19817, JRC 129516

²⁴ <https://osmdata.openstreetmap.de/data/water-polygons.html>

2.6 GHS-POP R2022A - GHS population grid multitemporal (1975-2030)

This GHS-POP spatial raster product (GHS-POP_GLOBE_R2022) depicts the distribution of human population (Figure 23), expressed as the number of people per cell. Residential population estimates at 5 years interval between 1975 and 2030 are derived from the raw global census data harmonized by CIESIN for the Gridded Population of the World, version 4.11 (GPWv4.11) at polygon level, and disaggregated from census or administrative units to grid cells, informed by the distribution, classification and density of built-up as mapped in the GHSL global layers per corresponding epoch. The disaggregation methodology is described in a peer reviewed paper (Freire et al., 2016). This is an updated and improved release of the product (GHS_POP_MT_GLOBE_R2019A) distributed within the GHSL Data Package 2019 (GHS P2019).

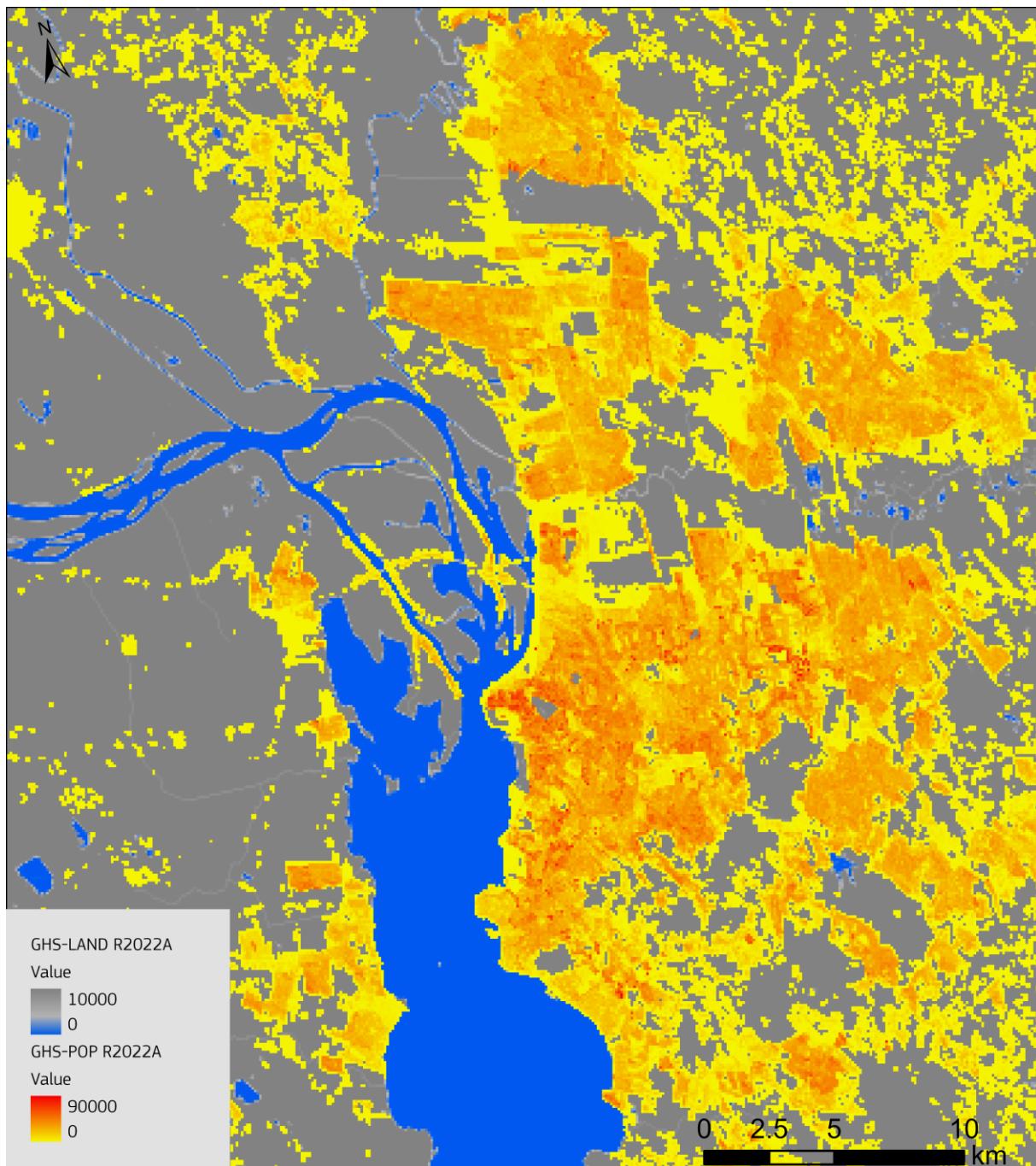


Figure 23 - GHS Population grid (GHS-POP) GHS_POP_E2020_GLOBE_R2022A_54009_100_V1_0 in Puerto Alegre (Brazil).

2.6.1 Improvements compared to the previous release

The new GHSL population distribution grids release aimed at incorporating improvements originating from input datasets, namely population estimates and built-up presence. While the disaggregation relied essentially on the same clear and simple approach, there were significant differences to the input data that had a positive effect on the final quality and accuracy of population grids, along with new approach for temporal estimation of population and systematic revision of coastlines and unpopulated areas. Here, we describe the main differences between the currently released products (GHS-POP_GLOBE_R2022A) and the previous one (GHS_POP_MT_GLOBE_R2019A). Figure 24 summarizes the main steps and outputs resulting from the workflow implemented for production of GHS-POP R2022.

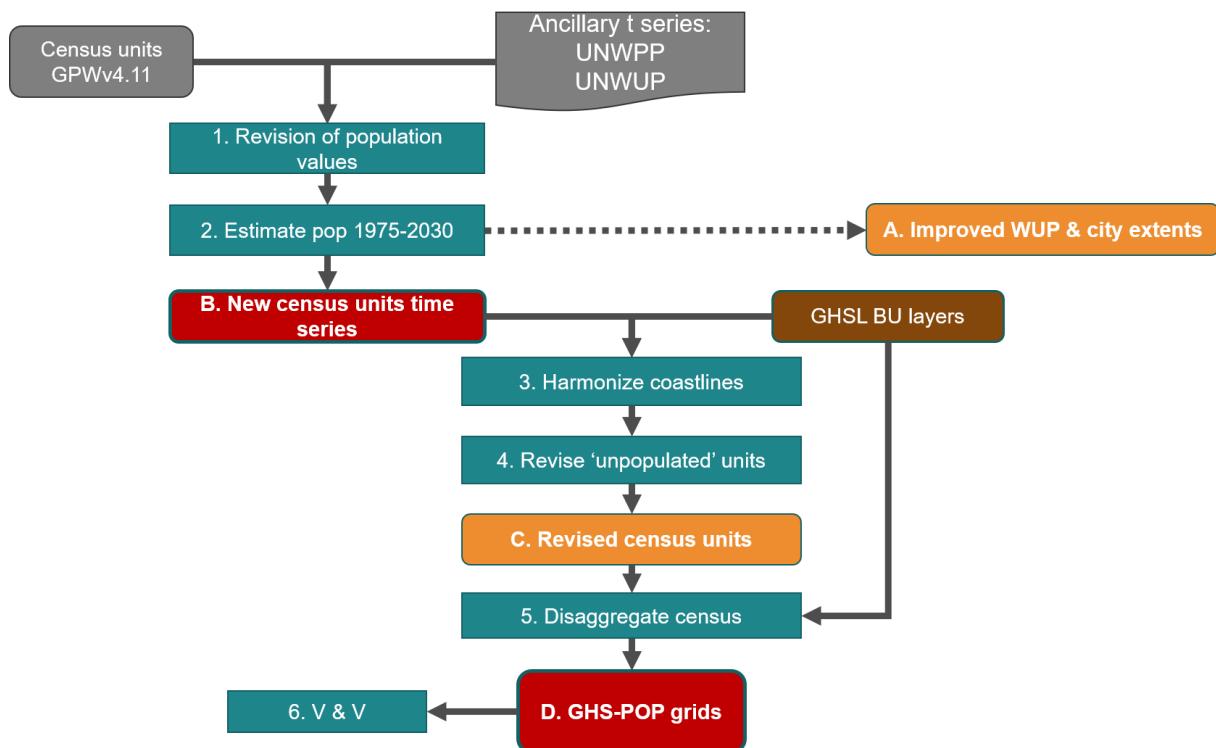


Figure 24 - Generalized workflow for producing GHS-POP R2022, with main steps (1-6) and intermediate and main outputs (A-D).

For the new GHS-POP (GHS-POP_GLOBE_R2022A), the new Sentinel/Landsat based GHS-BUILT-S (GHS-BUILT-S_GLOBE_R2022A, version 1.0) was used as target for disaggregation of population estimates. Total built-up surface (GHS-BUILT-S_GLOBE_R2022A) and non-residential built-up surface (GHS-BUILT-S_NRES_GLOBE_R2022A) were combined by subtracting a fraction of the non-residential surface according to a weight computed as the ratio of population density in fully residential and in fully non-residential units ($r=0.049151$), for all 5-year time steps between 1975 and 2030. Cells declared as "NoData" in built-up layers were treated as zero for population disaggregation.

The base source for population estimates (both census unit counts and geometries) was the raw dataset (census population at the census year and growth rates) of the Gridded Population of the World, version 4.11 (GPWv4.11), from CIESIN/SEDAC (<https://sedac.ciesin.columbia.edu/data/collection/gpw-v4/whatsnewrev11>).

The population estimates at polygon level for each mapped epoch (5-year interval between 1975 and 2030) are obtained by applying for each unit the population growth rate computed by CIESIN. The estimates are then rescaled to match (a) the total population time series at 'city' level from the extended database feeding the UN World Urbanisation Prospects 2018 (United Nations, Department of Economic and Social Affairs, Population Division, 2018), and (b) the total population time series at country level provided by the UN World Population Prospects 2019 (United Nations, Department of Economic and Social Affairs, Population Division, 2019). As the UN World Urbanisation Prospects 2018 provides the point location but not the boundaries for each 'city', the model performs an estimation of such boundaries starting from the coordinate pair for each settlement, after their location is revised. This revision was conducted using Geocoding and Reverse Geocoding, as well as visual inspection, to ensure these coordinates correspond to the centre of the urban agglomeration. The model

proceeds by aggregating administrative units (i.e. the census geometry provided by CIESIN) iteratively to match the total population of the city in the available census year. The units within the estimated boundaries are subject to the first rescaling procedure (a) at 'city' level (i.e. each UN 'city' determines the time series of the aggregated population of all units within each estimated boundary); while the other units are subject to the second procedure (b) at country level, to ensure the alignment with each country population time series estimated in UN WPP 2019 (medium variant). The set of estimated city boundaries matching UNWUP 2018 are released within this GHS P2022 data package in the GHS-SDATA product.

The previous release of the GHS-POP (R2019) already included some procedures aimed at further harmonising and critically revising the input population census data (GPW) provided by CIESIN. For the current release, those set of controls, checks and improvements were improved and expanded:

1. Revision of population values vs. an independent dataset (where available: EU)
2. Harmonisation of coastlines
3. Revision of unpopulated units

Control 1. aimed at mitigating large discrepancies between resident population counts in GPW and those of other authoritative sources. In this process, Denmark and Poland were flagged for inconsistencies between the population counts and distribution reported in GPW and those reported by ESTAT. Large discrepancies have been modified.

Coastlines are notoriously dynamic and their poor or outdated geospatial definition in census data can degrade the disaggregation of population along these critical areas. Therefore in this GHS-POP release, GPW coastlines were systematically modified to mitigate inconsistencies between the spatial extent of census data (or GADM) and built-up areas, since GHSL does not impose a coastline in mapping of these areas. This harmonisation has been carried out by extending all census units along coastlines (including those of inland water bodies) out using a 3 km buffer.

Units deemed as "unpopulated" in the GPW census data were critically assessed for presence of residential population, based on ancillary data (GHS-BUILT-S_GLOBE_R2022A) and very high-resolution imagery. Inconsistencies between census data and contradicting evidence were detected and reconciled accordingly, after confirmation.

Census units declared as "uninhabited" in GPW data but containing significant built-up surface in 2020 were selected for visual inspection with very high-resolution (VHR) imagery from web mapping services (Google maps and Bing). Those units where presence of residential built-up was confirmed by VHR imagery were selected for intervention. An automated method was applied to split and merge these polygons, based on geographical proximity, with those ones adjacent and containing resident population. This procedure was implemented while minimizing changes to source geometry, preserving the regional distribution of population, and the overall counts (Freire et al., 2018). This procedure modified 269 units in 31 countries.

GHS-POP product is produced in World Mollweide at 100 m, and then aggregated at 1 km. These two datasets are then warped to WGS 1984 coordinate system, at spatial resolutions of 3 arcsec and 30 arcsec respectively, by applying a thorough volume-preserving procedure (Maffenini et al., 2021).

2.6.2 Input Data

The new product GHS-BUILT-S_GLOBE_R2022A (version 1.0) was used to generate the target layers for disaggregation of population estimates by combining the GHS-BUILT-S_GLOBE_R2022A and GHS-BUILT-S_NRES_GLOBE_R2022A. The base source of population estimates for the different epochs was the raw census data with geometry of the Gridded Population of the World, version 4.11(GPWv4.11), from CIESIN/SEDAC, with some modifications as described above, along with the UN World Population Prospects 2019 and the UN World Urbanization Prospects 2018. The ESTAT LAU2 population time series was used to control the population counts at LAU2 level in EU.

2.6.3 Technical Details

Author: Marcello Schiavina, Sergio Freire, Joint Research Centre (JRC) European Commission; Kytt MacManus Columbia University, Center for International Earth Science Information Network - CIESIN.

Product name: GHS-POP_GLOBE_R2022A

Spatial extent: Global

Temporal extent: from 1975 to 2030, 5 years interval

Coordinate Systems: World Mollweide (ESRI: 54009) and WGS 1984 (EPSG: 4326)

Spatial resolutions available: 100 m, 1 km, 3 arcsec, 30 arcsec

Encoding: Population data float64 [0, ∞); population counts (ESRI: 54009; EPSG: 4326) and density (ESRI: 54009) NoData: -200

Data organisation: Global GeoTIFF file (100m, 1km; 3ss, 30ss) with overview images (OVR) and statistics (XML). Data tiles of 1000x1000 km size in GeoTIFF format (100 m, 1 km) and about 10x10 degrees size (3ss, 30ss).

Table 10 outlines the technical characteristics of the datasets released in this data package.

Table 16 - Technical details of the datasets in GHS-POP_GLOBE_R2022A

GHS-POP_GLOBE_R2022A		
ID	Description	Resolution (Projection/Coordinate system)
GHS_POP_E<epoch>_GLOBE_R2022A_<proj>_<res>_V1_0	Population density <epoch> 1975-2020; <proj> 54009, 4326; <res> 100m, 1000; 3ss 30ss Values are expressed as decimals Encoding: Float64 Values range: 0-Inf NoData [-200]	100 m, 1 km World Mollweide (ESRI:54009) 3ss, 30ss WGS84 (EPSG:4326)
GHS_POP_P<epoch>_GLOBE_R2022A_<proj>_<res>_V1_0	Population density projections <epoch> 2025-2030; <proj> 54009, 4326; <res> 100m, 1000; 3ss 30ss Values are expressed as decimals Encoding: Float64 Values range: 0-Inf NoData [-200]	100 m, 1 km World Mollweide (ESRI:54009) 3ss, 30ss WGS84 (EPSG:4326)

2.6.4 Summary statistics

Table 17 - Summary statistics of total population as obtained from the 1-km World Mollweide grid - total population adjusted to the UN WPP 2019 (United Nations, Department of Economic and Social Affairs, Population Division, 2019).

	1975	1980	1985	1990
Total Population	4,079,470,084	4,457,993,084	4,870,910,912	5,327,219,900
	1995	2000	2005	2010
	5,744,201,535	6,143,481,766	6,541,893,202	6,956,805,482
	2015	2020	2025	2030
	7,379,772,895	7,794,771,172	8,184,408,576	8,548,457,710

2.6.5 How to cite

Dataset:

Schiavina, Marcello; Freire, Sergio; MacManus, Kytt (2022): GHS-POP R2022A - GHS population grid multitemporal (1975-2030). European Commission, Joint Research Centre (JRC) [Dataset] DOI: 10.2905/D6D86A90-4351-4508-99C1-CB074B022C4A PID: <http://data.europa.eu/89h/d6d86a90-4351-4508-99c1-cb074b022c4a>

Concept & Methodology:

Freire, Sergio; MacManus, Kytt; Pesaresi, Martino; Doxsey-Whitfield, Erin; Mills, Jane (2016): Development of new open and free multi-temporal global population grids at 250 m resolution. Geospatial Data in a Changing World; Association of Geographic Information Laboratories in Europe (AGILE). AGILE 2016.

2.7 GHS-SMOD R2022A - GHS settlement layers, application of the Degree of Urbanisation methodology (stage I) to GHS-POP R2022A and GHS-BUILT-S R2022A, multitemporal (1975-2030)

The GHS Settlement Model layers (GHS-SMOD) GHS-SMOD_GLOBE_R2022A delineate and classify settlement typologies (Figure 25) via a logic of cell clusters population size, population and built-up area densities as defined by the stage I of the Degree of Urbanisation (European Commission & Statistical Office of the European Union, 2021) and recommended by the UN STAT COM. The GHS-SMOD is derived from the GHS-POP (GHS-POP_GLOBE_R2022A, version 1.0) and GHS-BUILT-S (GHS-BUILT-S_GLOBE_R2022A, version 1.0) released within this GHSL Data Package 2022 (GHS P2022).

The GHS Settlement Model layers GHS-SMOD_GLOBE_R2022A is composed by four datasets: the GHS-SMOD raster grid, the urban centre entities vector, the dense urban cluster vector and the semi-dense urban cluster vector. The first is a raster grid representing the settlement classification per grid cell and the other delineates the boundaries of settlement entities (i.e. urban centres, dense urban clusters and semi-dense urban clusters, with main attributes, in vector files.

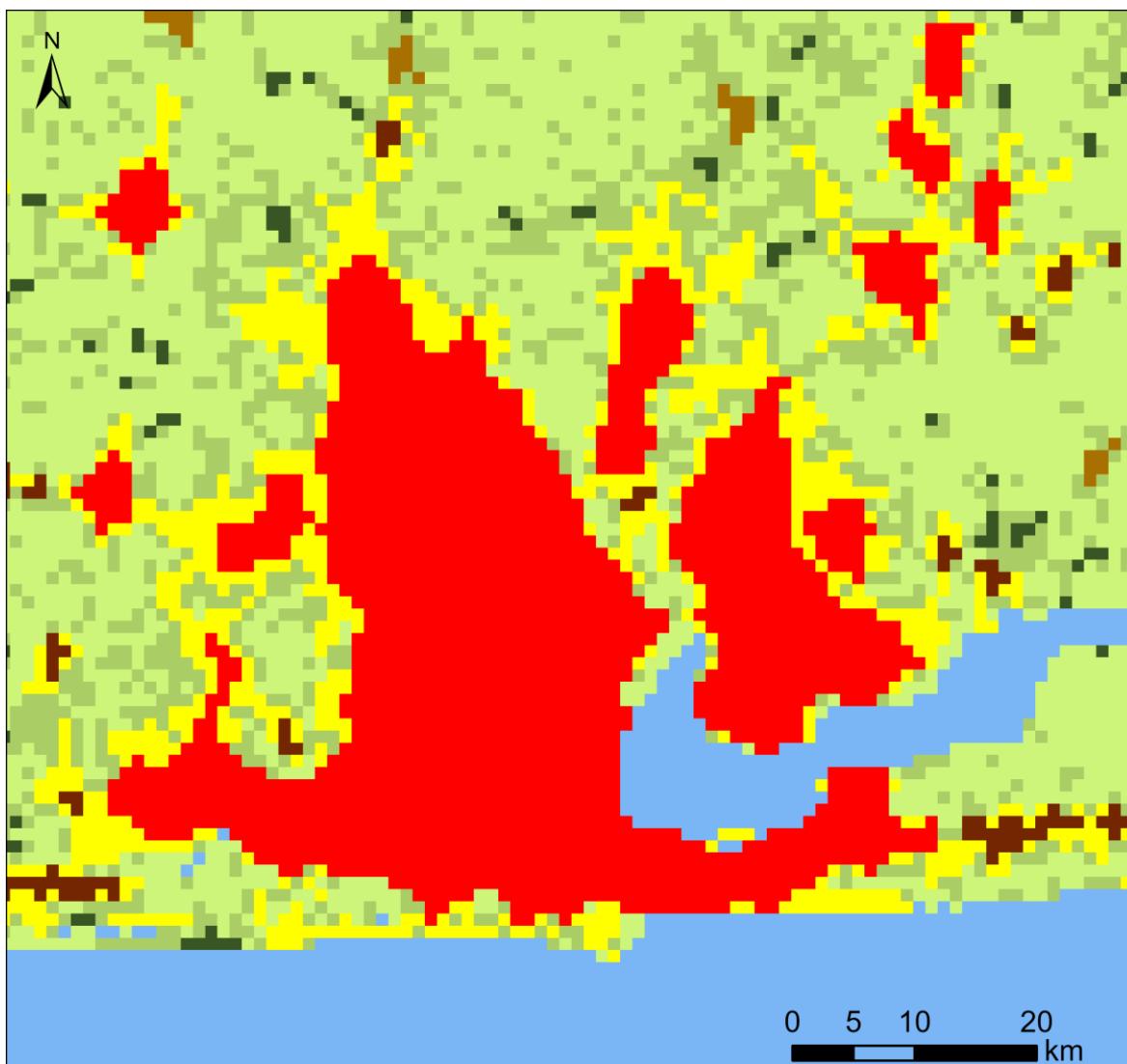


Figure 25 - GHS Settlement Model grid (GHS-SMOD) GHS-SMOD_P2020_GLOBE_R2022A_54009_1K_V1_0 displayed in the area of Lagos (Nigeria) –Legend in Table 18. The boundaries and the names shown on this map do not imply official endorsement or acceptance by the European Union © OpenStreetMap

2.7.1 Improvements compared to the previous release

The GHS-SMOD grid is an improvement of the GHS-SMOD R2019A based on the final specification of the stage I of the Degree of Urbanisation methodology. The GHS-SMOD is provided at the detailed level (Second Level - L2) and the new classification benefits of the new population and built-up surface layers (GHS-POP_GLOBE_R2022A and GHS-BUILT-S_GLOBE_R2022A).

2.7.2 GHS-SMOD classification rules

The “*Degree of Urbanisation Grid*” GHS-DUG 4 generates output grids and spatial entities by classifying 1 km² grid cells, on the basis of population density, size and contiguity, with the GHSL SMOD classification schema organized in two separate hierarchical levels.

The input for the GHSL SMOD is a 1 km² population grid (GHs-POP_GLOBE_R2022A), the built-up surface (GHS-BUILT-S_GLOBE_R2022A) and the land layer (GHS-LAND_GLOBE_R2022A). Each grid cell has the same shape and size, thereby avoiding distortions caused by using units varying in shape and size. This is a considerable advantage when compared to methods based on the population size and density of local administrative units.

At the first hierarchical level, the GHSL SMOD classifies the 1 km² grid cells by identifying the following spatial entities: a) “Urban Centre”, b) “Urban Cluster” and classifying all the other cells as “Rural Grid Cells”.

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

- “**Urban Centre**” (also “**High Density Cluster**” - **HDC**) - An urban centre consists of contiguous grid cells (4-connectivity cluster) with a density of at least 1,500 inhabitants per km² of permanent land or 50% built-up surface, and has at least 50,000 inhabitants in the cluster with smoothed boundaries (3-by-3 majority filtering) and <15 km² holes filled²⁵;
- “**Urban Cluster**” (also “**Moderate Density Cluster**” - **MDC**) - An urban cluster (or moderate density clusters) consists of contiguous grid cells (8-connectivity cluster) with a density of at least 300 inhabitants per km² of permanent land and has a population of at least 5,000 in the cluster. (The urban centres are subsets of the corresponding urban clusters).

The “**Rural grid cells**” (also “**Mostly Low Density Cells**” - **LDC**) are all the other cells that do not belong to an Urban Cluster. Most of these will have a density below 300 inhabitants per km² (grid cell); some may have a higher density, but they are not part of cluster with sufficient population to be classified as an Urban Cluster.

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

- **Class 3: “Urban Centre grid cell”**, if the cell belongs to an Urban Centre;
- **Class 2: “Urban Cluster grid cell”**, if the cell belongs to an Urban Cluster and not to an Urban Centre;
- **Class 1: “Rural grid cell”**, if the cell does not belong to an Urban Cluster.

The second hierarchical level of the GHSL SMOD is a refinement of the DEGURBA set up to identify smaller settlements. It follows the same approach based on population density, population size and contiguity with a **nested classification** into the first hierarchical level. At the second hierarchical level, the GHSL SMOD classifies the 1 km² grid cells by identifying the following spatial entities: a) “Urban Centres” as at the first level; b) “Dense Urban Cluster” and c) “Semi-dense Urban Cluster” as parts of the “Urban Cluster”, classifying all the other cells of “Urban Clusters” as “Suburban or peri-urban grid cells”; and identifying d) “Rural Cluster” within the “Rural grid cells”. All the other cells belonging to the “Rural grid cells” are classified as “Low Density grid cells” or “Very Low Density grid cells” according to their cell population (Figure 26).

²⁵ In a few countries with relatively low-density urban development and a strong separation of land use functions, the Degree of Urbanisation generates multiple urban centres for a single city. Creating urban centres using both cells with a density of at least 1,500 inhabitants and cells that are at least 50% built-up on permanent land resolves this issue (Optional flag is provided in GHS-DUG). Such highly built-up cells typically contain office parks, shopping malls, factories and transport infrastructure.

Here are reported the definition of the spatial entities at the second hierarchical:

- “**Urban Centre**” (also “**Dense, Large Settlement**”) - An urban centre consists of contiguous grid cells (4-connectivity cluster) with a density of at least 1,500 inhabitants per km² of permanent land, and has at least 50,000 inhabitants in the cluster with smoothed boundaries (3-by-3 majority filtering) and <15 km² holes filled²⁵;
- “**Dense Urban Cluster**” (also “**Dense, Medium Cluster**”) - A Dense Urban Cluster consists of contiguous cells (4-connectivity cluster) with a density of at least 1,500 inhabitants per km² of permanent land and has at least 5,000 and less than 50,000 inhabitants in the cluster;
- “**Semi-dense Urban Cluster**” (also “**Semi-dense, Medium Cluster**”) - A Semi-dense Urban Cluster consists of contiguous grid cells (8-connectivity cluster) with a density of at least 300 inhabitants per km² of permanent land, has at least 5,000 inhabitants in the cluster and is at least 3-km away from other Urban Clusters²⁶;
- “**Rural cluster**” (also “**Semi-dense, Small Cluster**”) - A Rural Cluster consists of contiguous cells (8-connectivity cluster) with a density of at least 300 inhabitants per km² of permanent land and has at least 500 and less than 5,000 inhabitants in the cluster.

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

The “**Low Density Rural grid cells**” (also “**Low Density grid cells**”) are Rural grid cells with a density of at least 50 inhabitants per km² of permanent land and are not part of a Rural Cluster.

The “**Very low density rural grid cells**” (also “**Very Low Density grid cells**”) are cells with a density of less than 50 inhabitants per km² of permanent land.

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

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

- **Class 30: “Urban Centre grid cell”**, if the cell belongs to an Urban Centre spatial entity;
- **Class 23: “Dense Urban Cluster grid cell”**, if the cell belongs to a Dense Urban Cluster spatial entity;
- **Class 22: “Semi-dense Urban Cluster grid cell”**, if the cell belongs to a Semi-dense Urban Cluster spatial entity;
- **Class 21: “Suburban or per-urban grid cell”**, if the cell belongs to an Urban Cluster cells at first hierarchical level but is not part of a Dense or Semi-dense Urban Cluster;
- **Class 13: “Rural cluster grid cell”**, if the cell belongs to a Rural Cluster spatial entity;
- **Class 12: “Low Density Rural grid cell”**, if the cell is classified as Rural grid cells at first hierarchical level, has more than 50 inhabitant and is not part of a Rural Cluster;
- **Class 11: “Very low density rural grid cell”**, if the cell is classified as Rural grid cells at first hierarchical level, has less than 50 inhabitant and is not part of a Rural Cluster;
- **Class 10: “Water grid cell”**, if the cell has 0.5 share covered by permanent surface water and is not populated nor built.

²⁵ Measured as outside a buffer of three grid cells of 1 km² around dense urban clusters and urban centres.

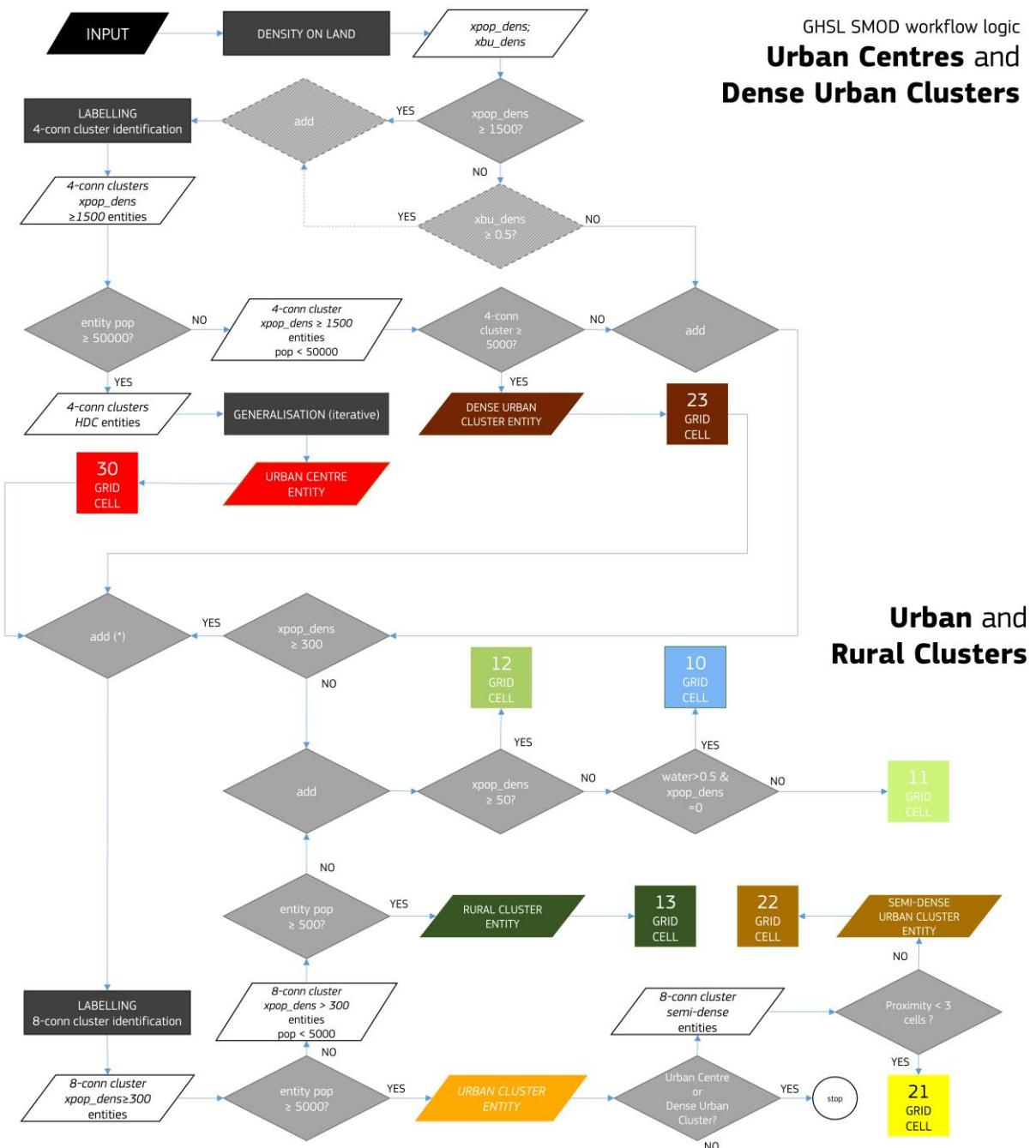


Figure 26 - Schematic overview of GHSL SMOD entities workflow logic. “xpop” represents the population abundance per grid cell; “xpop_dens” represents the population density on permanent land; “xbu” represents the built-up density per grid cell; “xbu_dens” represents the built-up density on permanent land. “DENSITY ON LAND” process fill built-up cells on water with max between 0.5 and built-up surface value and population on water with global average built-up per capita. “GENERALISATION” process performs a median filtering (3x3) for smoothing boundaries and fills gaps above 15 km². (*) this procedure of enforcement logic allows the delineation of Urban Clusters Entities which contains by definition the Urban Centres and all 2X classes. Each entity has a corresponding vector boundary.

2.7.3 GHS-SMOD spatial entities naming

The highest tier spatial entities (Urban Centre) are named using an algorithm that automatically queries the GISCO and the full OpenStreetMap datasets with the following steps:

- 1 For each node of the dataset the reference name is selected among all the available naming tag with the following priority: *name_en; name_int; name* (if in Latin characters); *name_fr; name_es; name_it; name_de; name_wiki; name;*
- 2 The algorithm filters all nodes that overlaps the extent of the spatial entity in their 3 km buffer and selects among them the nodes with the highest priority using the following tag ordering (*key:value*): *place:city; place:town; place:village; place:hamlet; place:isolated_dwelling; place:farm; place:allotments; place:borough; place:suburb; place:quarter; place:neighbourhood; place:city_block; place:plot; place:locality; place:municipality; place:civil_parish; railway:station; addr:city;*
- 3 If a single node is selected the reference name is assigned to the spatial entity as main name; if multiple nodes are selected the function ranks the nodes by *population* tag values (descending order, absence of *population* tag equal 0); if no *population* tag is present, it ranks by sum of GHS-POP in the 5 km buffer around the point (descending order). The ordered list is saved and assigned to the spatial entity; the first name is selected as main name of the spatial entity.
- 4 If any node filtered at point 2 has the *capital* tag equal to *yes, 1, 2, 3, or 4*, the reference name is marked as capital. If this differs from the main name assigned at point 3, the capital name is concatenated to the main name, between square brackets.

2.7.4 GHS-SMOD L2 grid and L1 aggregation

Settlement typologies are identified in the GHS-SMOD grid at L2 with a two digit code (30 – 23 – 22 – 21 – 13 – 12 – 11 – 10), linking to grid level and municipal level description terms (both the municipal and grid level terms are accompanied by a technical term). **Classes 30 – 23 – 22 – 21 if aggregated form the “urban domain”, 13 – 12 – 11 – 10 form the “rural domain”**. Table 20 shows the L2 grid cells population (expressed as people per square kilometre: people/km²) and built-up area (expressed as square kilometres: km²) expected characteristics in terms of min-max population and built-up density bounds. Table 19 presents the logic to define settlement typologies.

L1 classifies three settlement typologies as displayed in Table 21. Settlement typologies are identified at L1 with a single digit code (3 – 2 – 1), and grid level and municipal level terms (both the municipal and grid level are accompanied by a technical term), HDC for type 3, MDC for type 2, and LDC for type 1). Classes 3 - 2 if aggregated form the “urban domain”, 1 forms the “rural domain”. Table 22 presents the logic to define settlement typologies as described in section 2.7.2. Table 23 shows the L1 grid cells population and built-up area characteristics in terms of min-max population and built-up density bounds.

Table 18 - Settlement Model L2 nomenclature

Code	RGB	Grid level term	Spatial entity (polygon) Technical term	Other cells Technical term	Municipal level term Technical term
30	255 0 0	URBAN CENTRE GRID CELL	URBAN CENTRE (UC) <i>DENSE, LARGE CLUSTER</i>		CITY <i>LARGE SETTLEMENT</i>
23	115 38 0	DENSE URBAN CLUSTER GRID CELL	DENSE URBAN CLUSTER (DUC) <i>DENSE, MEDIUM CLUSTER</i>		DENSE TOWN <i>DENSE, MEDIUM SETTLEMENT</i>
22	168 112 0	SEMI-DENSE URBAN CLUSTER GRID CELL	SEMI-DENSE URBAN CLUSTER (SDUC) <i>SEMI-DENSE, MEDIUM CLUSTER</i>		SEMI-DENSE TOWN <i>SEMI-DENSE, MEDIUM SETTLEMENT</i>
21	255 255 0	SUBURBAN OR PERI-URBAN GRID CELL		SUBURBAN OR PERI-URBAN GRID CELLS <i>SEMI-DEMSE GRID CELLS</i>	SUBURBAN OR PERI-URBAN AREA <i>SEMI-DENSE AREA</i>
13	55 86 35	RURAL CLUSTER GRID CELL	RURAL CLUSTER (RC) <i>SEMI-DENSE, SMALL CLUSTER</i>		VILLAGE <i>SMALL SETTLEMENT</i>
12	171 205 102	LOW DENSITY RURAL GRID CELL		LOW DENSITY RURAL GRID CELLS <i>LOW DENSITY GRID CELLS</i>	DISPERSED RURAL AREA <i>LOW DENSITY AREA</i>
11	205 245 122	VERY LOW DENSITY RURAL GRID CELL		VERY LOW DENSITY RURAL GRID CELLS <i>VERY LOW DENSITY GRID CELLS</i>	MOSTLY UNINHABITED AREA <i>VERY LOW DENSITY AREA</i>
10	122 182 245	WATER GRID CELL	-	-	-

Table 19 - Settlement Model L2 synthetic explanation of logical definition and grid cell sets

Code	Logical Definition at 1 km ² grid cell	Grid cell sets used in the logical definition (shares defined on land surface)			
		P _{dens} :	P _{min} :	B _{dens} :	T _{con} :
	Local Population Density lower bound ">" (people/km ²)	Cluster Population lower bound ">" (people)	Local share of Built-up Area lower bound ">" (km ²)	Topological constraints	
30	((P _{dens} v B _{dens}) \wedge T _{con}) \wedge P _{min}) v v [iterative_median_filter(3-by-3)] v [gap_fill(<15km2)] ²⁷	1,500	50,000	0.50	4-connectivity clusters
23	((P _{dens} v B _{dens}) \wedge T _{con}) \wedge P _{min}) \wedge \neg 30	1,500	5,000	0.50	4-connectivity clusters
22	((P _{dens} \wedge T _{con_1}) \wedge P _{min}) \wedge \neg (30 v 23)) \wedge T _{con_2}	300	5,000	none	1: 8-connectivity clusters; 2: farther than 3km (beyond 3 cells buffer) from 23 or 30
21	(((P _{dens} \wedge (30 v 23)) \wedge T _{con_1}) \wedge P _{min}) \wedge \neg (30 v 23)) \wedge T _{con_2}	300	5,000	none	1: 8-connectivity clusters; 2: within 3km (within 3 cells buffer) from 23 or 30
13	((P _{dens} \wedge T _{con}) \wedge P _{min}) \wedge \neg (30 v 2X)	300	500	none	8-connectivity clusters
12	P _{dens} \wedge \neg (30 v 2X v 13)	50	none	none	none
11	T _{con} \wedge \neg (30 v 2X v 13 v 12)	none	none	none	On Land (Land>=50% v BU ²⁸ >0% v Pop>0)
10	T _{con}	none	none	none	Not on Land

²⁷ The seeds for the related spatial entity is obtained before morphological operations

²⁸ Retaining only contiguous BU at least partially on land.

Table 20 - Settlement Model L2 grid cells population and built-up area characteristics (densities on permanent land)

Code	Population		Built-up area	
	Minimum density expected (people/km ²)	Maximum density expected (people/km ²)	Minimum density expected (share)	Maximum density expected (share)
30	0	∞	0	1
23	0	50,000	0	1
22	300	5,000	0	1
21	300	5,000	0	1
13	300	5,000	0	1
12	50	500	0	1
11	0	50	0	1
10	0	0	0	0

Table 21 - Settlement Model L1 nomenclature

Code	RGB	Grid level term	Spatial entity (polygon) <i>Technical term</i>	Other cells		Municipal level term <i>Technical term</i>
				Other cells <i>Technical term</i>	Municipal level term <i>Technical term</i>	
3	255 0 0	URBAN CENTRE GRID CELL	URBAN CENTRE <i>HIGH DENSITY CLUSTER (HDC)</i>	URBAN CLUSTER <i>MODERATE DENSITY CLUSTER (MDC)</i>		CITIES <i>DENSELY POPULATED AREA</i>
2	255 170 0	URBAN CLUSTER GRID CELL	RURAL GRID CELLS <i>LOW DENSITY GRID CELL (LDC)</i>		TOWNS & SEMI-DENSE AREA <i>INTERMEDIATE DENSITY AREA</i>	RURAL AREAS <i>THINLY POPULATED AREA</i>
1	115 178 115	RURAL GRID CELL				

Table 22 - Settlement Model L1 synthetic explanation of logical definition and grid cell sets

Code	Logical Definition at 1 km ² grid cell	Grid cell sets used in the logical definition (shares defined on land surface)			
		P _{dens} :	P _{min} :	B _{dens} :	T _{con} :
3	((P _{dens} \vee B _{dens}) \wedge T _{con}) \wedge P _{min}) \vee \vee [iterative_median_filter(3-by-3)] \vee \vee [gap_fill(<15km ²)] ²⁹	1,500	50,000	0.50	4-connectivity clusters
2	P _{dens} \wedge P _{min} \wedge T _{con} \wedge \neg 3	300	5,000	none	8-connectivity clusters
1	((P _{dens} \wedge 3) \wedge T _{con}) \wedge P _{min}) \wedge \wedge \neg 3 ^s	none	none	none	none

²⁹ The seeds for the related spatial entity is obtained before morphological operations

Table 23 - Settlement Model L1 grid cells population and built-up area characteristics (densities on permanent land)

Code	Population		Built-up area	
	Minimum density expected (people/km²)	Maximum density expected (people/km²)	Minimum density expected (share)	Maximum density expected (share)
3	0	∞	0	1
2	0	50,000	0	1
1	0	5,000	0	1

2.7.5 Input Data

The input data are the multi-temporal GHS-BUILT-S, the GHS-POP and GHS-LAND grids of the GHSL Data Package 2022 (GHS P2022).

2.7.6 Technical Details

Author: Marcello Schiavina, Michele Melchiorri, Martino Pesaresi, Joint Research Centre (JRC) European Commission.

Product name: GHS-SMOD_GLOBE_R2022A

Spatial extent: Global

Temporal extent: from 1975 to 2030, 5 years interval

Coordinate System: World Mollweide (EPSG: 54009)

Spatial resolution available: 1 km

Table 24 outlines the technical characteristics of the datasets released in this data package.

2.7.6.1 GHS-SMOD raster grid

Encoding: integer16 [30 – 23 – 22 – 21 – 13 – 12 – 11 – 10], No Data: -200

Data organisation: Global GeoTIFF file with overview images (OVR) and colormap (CLR). Data tiles of 1000x1000 km size in GeoTIFF format

2.7.6.2 GHS-SMOD urban centre entities

Data organisation: GeoPackage (GPKG) database with vector layer of Urban Centre entities boundaries (polygons).

Attributes:

- ID_UC_G0: Unique Identifiers of the urban centre entity;
- Name_main: Main name assigned to the urban centre entity;
- Name_list: List of all name selected within the spatial entity;
- POP_<year>: sum of GHS-POP within the spatial entity extent for the related year;
- BU_<year>: sum of GHS-BU within the spatial entity extent for the related year

2.7.6.3 GHS-SMOD dense urban cluster entities

Data organisation: GeoPackage (GPKG) database with vector layer of dense urban cluster entities boundaries (polygons).

Attributes:

- ID_DUC_G0: Unique Identifiers of the dense urban cluster entity;
- Name_main: Main name assigned to the dense urban cluster entity;
- Name_list: List of all name selected within the spatial entity;
- POP_<year>: sum of GHS-POP within the spatial entity extent for the related year;
- BU_<year>: sum of GHS-BU within the spatial entity extent for the related year

2.7.6.4 GHS-SMOD semi-dense urban cluster entities

Data organisation: GeoPackage (GPKG) database with vector layer of semi-dense urban cluster entities boundaries (polygons).

Attributes:

- ID_SDUC_G0: Unique Identifiers of the semi-dense urban cluster entity;
- Name_main: Main name assigned to the semi-dense urban cluster entity;
- Name_list: List of all name selected within the spatial entity;
- POP_<year>: sum of GHS-POP within the spatial entity extent for the related year;
- BU_<year>: sum of GHS-BU within the spatial entity extent for the related year

Table 24 - Technical details of the datasets in GHS-SMOD_GLOBE_R2022A

GHS-SMOD_GLOBE_R2022A		
ID	Description	Resolution (Projection)
GHS_SMOD_E<epoch>_GLOBE_R2022A_54009_1K_V1_0	Settlement typology codes for <epoch> 1975-2020 (5 years interval) Encoding Int16 Values range: 10-30 NoData [-200]	1 km (World Mollweide)
GHS_SMOD_P<epoch>_GLOBE_R2022A_54009_1K_V1_0	Settlement typology codes projected for <epoch> 2025-2030 (5 years interval) Encoding Int16 Values range: 10-30 NoData [-200]	1 km (World Mollweide)
GHS_SMOD_POP2020_GLOBE_R2022A_54009_1K_labelUC_V1_0	2020 urban centre entities boundaries File format: shapefile	1 km (World Mollweide)
GHS_SMOD_POP2020_GLOBE_R2022A_54009_1K_labelDUC_V1_0	2020 dense urban cluster entities boundaries File format: shapefile	1 km (World Mollweide)
GHS_SMOD_POP2020_GLOBE_R2022A_54009_1K_labelSDUC_V1_0	2020 semi-dense urban cluster entities boundaries File format: shapefile	1 km (World Mollweide)

2.7.7 Summary statistics

Table 25 - Summary statistics of total area in square kilometres of each settlement typology at global level as obtained from the 1-km GHS-POP grids L2.

	1975	1980	1985	1990	1995	2000
30	251,001	286,014	323,944	364,913	412,478	459,512
23	211,248	229,254	249,410	272,610	288,859	300,052
22	272,645	277,094	285,330	289,173	297,325	311,834
21	920,966	1,011,709	1,109,223	1,202,087	1,307,814	1,416,550
13	633,083	657,779	684,473	706,731	731,619	752,341
12	2,949,060	3,014,438	3,091,606	3,135,228	3,326,825	3,553,497
11	140,639,331	140,400,339	140,131,875	139,905,503	139,508,390	139,081,658
	2005	2010	2015	2020	2025	2030
30	503,003	551,150	598,034	638,713	678,621	706,450
23	307,539	312,698	315,252	304,955	304,172	299,085
22	332,581	360,867	392,706	443,611	488,473	516,029
21	1,591,163	1,810,085	2,028,964	2,333,308	2,556,884	2,745,945
13	777,432	799,519	825,717	848,447	859,547	840,769
12	4,147,777	4,728,632	5,442,484	6,518,619	6,865,586	7,432,616
11	138,236,216	137,354,514	136,368,454	134,905,014	134,246,300	133,466,749

Table 26 - Summary statistics of total built-up area in square kilometres for each settlement typology at global level as obtained from the 1-km GHS-POP grids L2.

	1975	1980	1985	1990	1995	2000
30	46,538	52,653	59,941	68,368	76,678	86,554
23	23,272	25,008	27,128	29,671	31,570	34,226
22	16,447	15,887	15,662	15,457	15,508	16,310
21	44,746	47,380	50,638	54,373	58,042	64,350
13	20,181	20,411	20,829	21,429	22,250	24,079
12	31,363	29,704	28,491	27,845	28,841	32,695
11	8,217	7,227	6,721	6,618	7,253	8,708
	2005	2010	2015	2020	2025	2030
30	93,672	102,347	111,431	117,922	121,120	122,296
23	35,901	37,900	40,140	41,022	40,640	39,518
22	17,461	19,459	22,366	26,495	29,502	30,940
21	70,762	81,129	95,372	118,036	133,539	142,613
13	26,010	29,014	32,902	37,825	39,658	39,323
12	37,963	47,954	63,089	92,925	113,902	130,388
11	11,157	15,679	24,262	44,057	57,687	73,949

Table 27 - Summary statistics of total population in each settlement typology at global level as obtained from the 1-km GHS-POP grids L2.

	1975	1980	1985	1990	1995	2000
30	1,253,031,897	1,440,864,496	1,643,134,609	1,882,906,631	2,122,312,303	2,377,499,930
23	801,376,963	870,460,988	945,222,001	1,035,735,768	1,074,059,850	1,080,813,474
22	236,872,661	243,868,526	253,445,554	259,615,667	266,804,492	278,055,143
21	762,537,470	843,167,347	930,993,688	1,018,931,892	1,107,639,291	1,192,793,351
13	511,313,276	533,995,442	558,672,957	582,426,545	597,749,677	604,172,676
12	429,645,334	441,299,551	454,657,909	462,495,886	488,649,592	519,953,801
11	84,702,904	84,347,154	84,794,992	85,118,698	86,997,774	90,205,482
	2005	2010	2015	2020	2025	2030
30	2,559,846,333	2,753,716,148	2,926,774,627	3,066,558,527	3,272,997,856	3,491,602,720
23	1,054,256,307	1,018,111,687	977,074,985	891,713,774	865,287,131	841,043,193
22	292,060,851	307,567,271	324,271,019	344,576,872	367,075,439	373,797,998
21	1,321,111,003	1,475,736,001	1,627,769,674	1,818,512,092	1,972,276,372	2,087,465,801
13	605,322,160	597,857,004	592,158,635	569,223,527	553,640,665	522,933,013
12	594,359,969	672,151,134	764,220,318	907,250,249	964,501,779	1,039,682,116
11	114,950,393	131,684,408	167,527,774	196,963,755	188,658,280	191,962,602

Table 28 - Summary statistics of total area in square kilometres of each settlement typology at global level as obtained from the 1-km GHS-POP grids L1.

	1975	1980	1985	1990	1995	2000
3	211,248	229,254	249,410	272,610	288,859	300,052
2	1,826,694	1,946,582	2,079,026	2,197,991	2,336,758	2,480,725
1	143,588,391	143,414,777	143,223,481	143,040,731	142,835,215	142,635,155
	2005	2010	2015	2020	2025	2030
3	503,003	551,150	598,034	638,713	678,621	706,450
2	2,231,283	2,483,650	2,736,922	3,081,874	3,349,529	3,561,059
1	143,161,425	142,882,665	142,636,655	142,272,080	141,971,433	141,740,134

Table 29 - Summary statistics of total built-up area in square kilometres for each settlement typology at global level as obtained from the 1-km GHS-POP grids L1.

	1975	1980	1985	1990	1995	2000
3	23,272	25,008	27,128	29,671	31,570	34,226
2	81,375	83,679	87,129	91,259	95,800	104,739
1	39,580	36,931	35,212	34,463	36,094	41,402
	2005	2010	2015	2020	2025	2030
3	93,672	102,347	111,431	117,922	121,120	122,296
2	124,124	138,489	157,879	185,553	203,682	213,071
1	75,130	92,646	120,253	174,807	211,246	243,660

Table 30 - Summary statistics of total population in each settlement typology at global level as obtained from the 1-km GHS-POP grids L1.

	1975	1980	1985	1990	1995	2000
3	1,253,031,897	1,440,864,496	1,643,134,609	1,882,906,631	2,122,312,303	2,377,499,930
2	1,800,787,094	1,957,496,862	2,129,661,243	2,314,283,327	2,448,503,633	2,551,661,968
1	1,025,661,514	1,059,642,147	1,098,125,857	1,130,041,129	1,173,397,042	1,214,331,959
	2005	2010	2015	2020	2025	2030
3	2,559,846,333	2,753,716,148	2,926,774,627	3,066,558,527	3,272,997,856	3,491,602,720
2	2,667,428,160	2,801,414,959	2,929,115,678	3,054,802,738	3,204,638,943	3,302,306,992
1	1,314,632,522	1,401,692,546	1,523,906,727	1,673,437,531	1,706,800,724	1,754,577,731

2.7.8 How to cite

Dataset:

Schiavina, Marcello; Melchiorri, Michele; Pesaresi, Martino (2022): GHS-SMOD R2022A - GHS settlement layers, application of the Degree of Urbanisation methodology (stage I) to GHS-POP R2022A and GHS-BUILT-S R2022A, multitemporal (1975-2030). European Commission, Joint Research Centre (JRC) [Dataset] DOI:10.2905/4606d58a-DC08-463C-86A9-D49EF461C47F PID: <http://data.europa.eu/89h/4606d58a-dc08-463c-86a9-d49ef461c47f>

Concept & Methodology:

European Commission, and Statistical Office of the European Union, 2021. Applying the Degree of Urbanisation — A methodological manual to define cities, towns and rural areas for international comparisons — 2021 edition Publications Office of the European Union, 2021, ISBN 978-92-76-20306-3 doi: 10.2785/706535

2.8 GHS-DUC R2022A - GHS Degree of Urbanisation Classification, application of the Degree of Urbanisation methodology (stage II) to GADM 3.6 layer, multitemporal (1975-2030)

The GHS Degree of Urbanisation Classification (GHS-DUC) GHS-DUC_GLOBE_R2022A classifies all GADM 3.6³⁰ administrative units (from level 0 to level 5) by applying the stage II of the Degree of Urbanisation (European Commission & Statistical Office of the European Union, 2021) as recommended by the UN STAT COM. In total, 386,741 GADM units are classified. This is done according to a logic of majority of population (GHS-POP_GLOBE_R2022A) in unit overlaid to the settlement classification grid (GHS-SMOD_GLOBE_R2022A).

In GHS-DUC each GADM polygon (for all available GADM levels) is coded by Degree of Urbanisation Level 1 and Level 2, and has statistics on number of residents and built-up area surface. The GHS-DUC is derived from the GHS-POP (GHS-POP_GLOBE_R2022A, version 1.0) and GHS-SMOD (GHS-SMOD_GLOBE_R2022A, version 1.0) to compute population counts per epoch, with additional statistics from GHS-BUILT-S (GHS-BUILT-S_GLOBE_R2022A, version 1.0) for computing built-up area values for each epoch.

The GHS Degree of Urbanisation Classification GHS-DUC_GLOBE_R2022A is composed by:

- one summary Excel file collecting results at the finest level available per country for each epoch,
- 72 classification files of all GADM 3.6 units for each level (0-5) and each epoch (1975-2030, 5 years interval) released in CSV format.

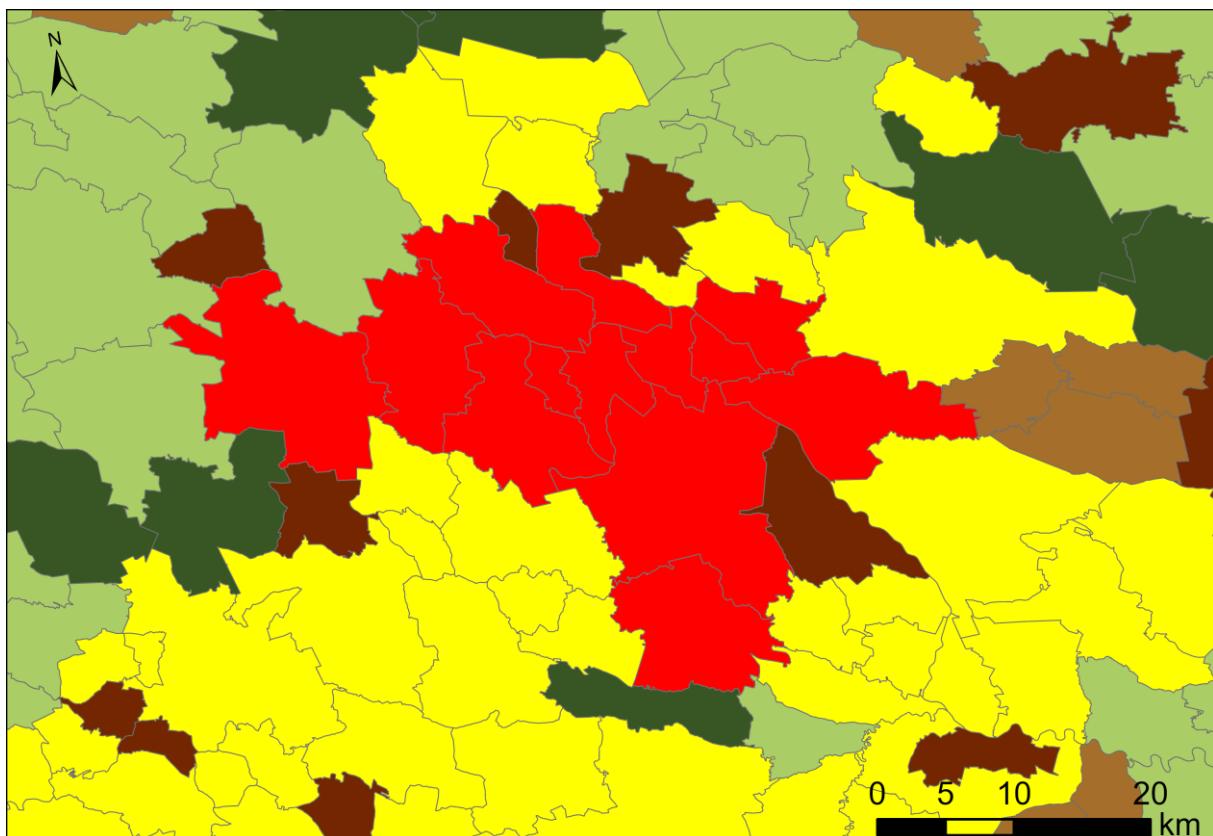


Figure 27 - GHS Degree of Urbanisation Classification (GHS-DUC) GHS-DUC_GLOBE_R2022A_V1_0_GADM36_2020_level4 joined to the GADM 3.6 level 4 layer, displayed in the area of Katowice (Poland) showing the classification of local units by Degree of Urbanisation Level 2—Legend in Table 31. The boundaries shown on this map do not imply official endorsement or acceptance by the European Union.

³⁰ <https://gadm.org/index.html>

2.8.1 Improvements compared to previous release

This classification of administrative units and territories relies on the updated settlement classifications and population grids of GHS P2022 and includes all epochs between 1975 and 2030 at 5-year interval.

2.8.2 GHSL Territorial Units Classification

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

2.8.2.1 Territorial units classification Level 1

Once all grid cells covered by GADM have been classified as urban centres, urban clusters and rural grid cells using the GHS-DUG Tool, the next step concerns overlaying these results onto local units, as follows (Figure 29):

- **Cities (or densely populated areas):** local units that have at least 50% of their population in urban centres, **code 3**.
- **Towns and semi-dense areas (or intermediate density areas):** local units that have less than 50% of their population in urban centres and no more than 50% of their population in rural grid cells, **code 2**.
- **Rural areas (or thinly populated areas):** local units that have more than 50% of their population in rural grid cells, **code 1**.

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

In some countries, not all the small spatial units contain inhabitants. To classify the spatial units without any population, the same rules should be applied to their area as are applied to their population. For example, a small unpopulated spatial unit that has more than 50 % of its area in rural grid cells is classified as a rural area.

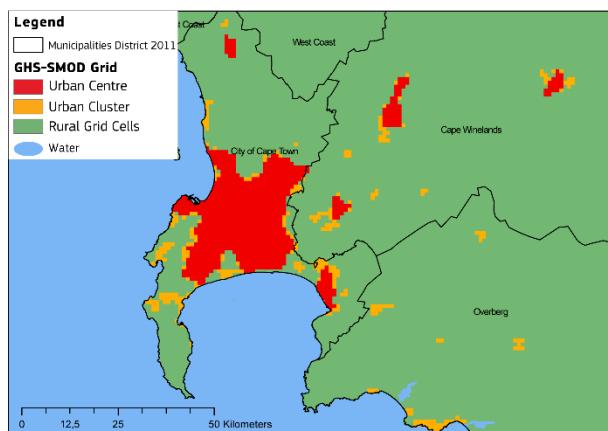


Figure 28 - Urban centre, urban cluster and rural grid cells around Cape Town, South Africa

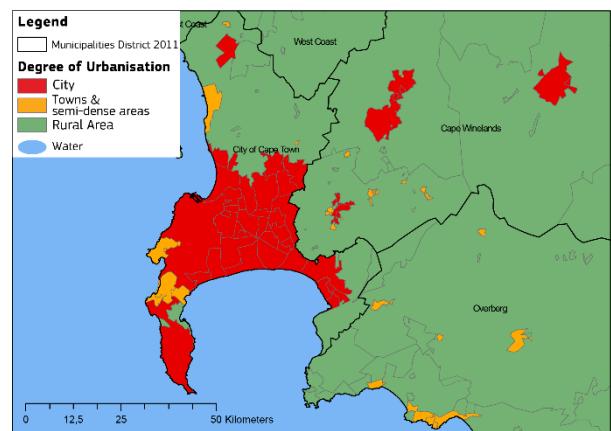


Figure 29 - City, towns & semi-dense areas and rural areas around Cape Town, South Africa (classification of Main Places units, note that Cape Peninsula is part of Cape Town Main Place)

2.8.2.2 Territorial units classification Level 2

Local units are classified as cities in identical manner to the Degree of Urbanisation level 1 (Figure 31):

- A city consists of local units that have at least 50% of their population in an urban centre, **code 30**.

Local units classified as “towns and semi-dense areas” can be divided into three classes:

- Dense Towns have a larger share of their population in dense urban clusters than in semi-dense urban clusters (i.e. are dense) and a larger share in dense plus semi-dense urban clusters than in suburban or peri-urban cells (i.e. are towns), **code 23**.
- Semi-dense Towns have a larger population share in semi-dense urban clusters than in dense urban clusters (i.e. are semi-dense) and a larger share in dense plus semi-dense urban clusters than in suburban or peri-urban cells (i.e. are towns), **code 22**.
- Suburban or peri-urban areas have a larger population share in suburban or peri-urban cells than in dense plus semi-dense urban clusters, **code 21**.

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

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

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

As for Level 1, to classify the spatial units without any population, the same rules that are applied to population are applied to area.

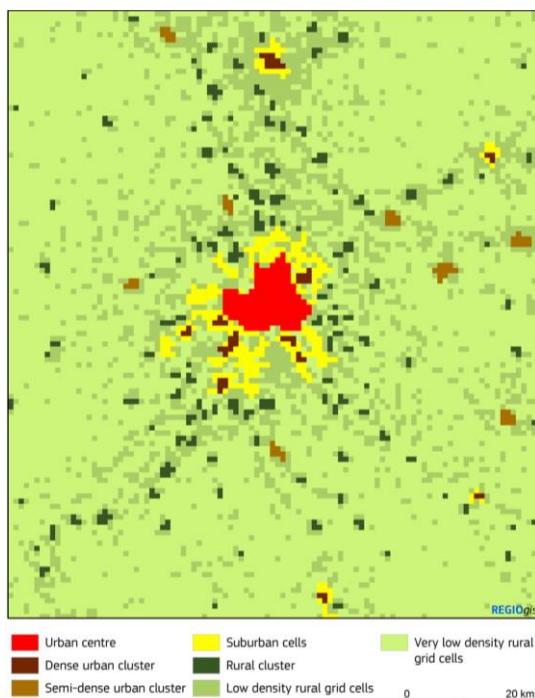


Figure 30 - Degree of urbanisation level 2 grid classification around Toulouse, France

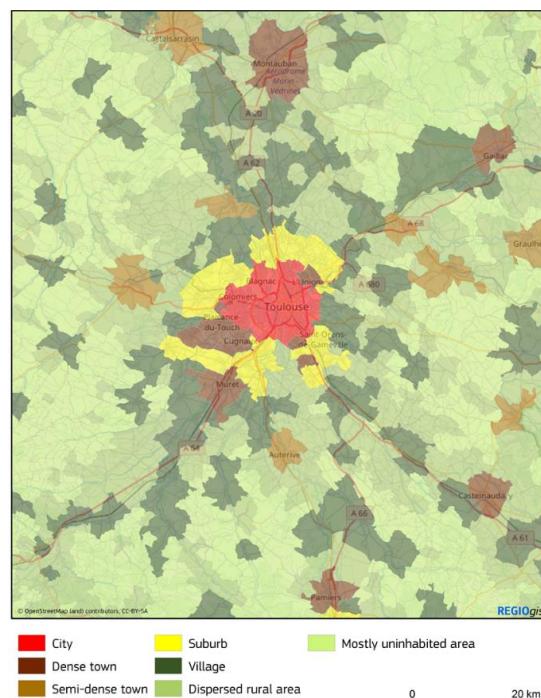


Figure 31 - Degree of urbanisation level 2 local unit classification around Toulouse, France

2.8.2.3 Classification workflow

The classification has been performed combining three inputs of different nature for each epoch and GADM level:

- two raster layers:
 - the settlement classification grid at 1 km for the processed year (GHS-SMOD);
 - the population grid at 100m resolution for the same epoch (GHS-POP);
- one vector layer of territorial units from GADM3.6³¹ at the processed level.

The procedure begins with the rasterization of the vector layer. As suggested, the unit classification works better with the smallest administrative units available, therefore these are rasterized using a resolution of 50 m, snapped to the population grid, to reduce the number of units that will not have a representation in the raster layer. The settlement classification grid (at 1 km resolution) is oversampled using nearest neighbour algorithm to align with the 50-m territorial units' raster. The population grid is also oversampled and the values are then adjusted by dividing all original cells by the oversampling ratio (e.g. from 100 m to 50 m the ratio is 4, as 4 grid cells of 50 m compose each 100 m cell) assuming a uniform distribution of population within each cell.

Once the pre-processing of the layers is completed, the algorithm computes for each unit the share of population in each class of the settlement classification grid (both at Level 1 and 2), through zonal statistics, and assigns the class of the unit accordingly (i.e. following the classification rules described in sections 2.8.2.1 and 2.8.2.2). When a unit is unpopulated it runs zonal statistics of areas per settlement classification.

Even if the working resolution is set at 50 m, it could happen that some small polygons could not be rasterised due to geospatial data processing constraints. In such cases the algorithm evaluates the class of these polygons by running separately the zonal statistics (i.e. one polygon at a time) and performing the rasterization procedure with “all touching cells” option. To avoid double counting of population, no population is assigned to such polygons, but the classification still considers population per class in the rasterised cells.

2.9 A consistent nomenclature for the Degree of Urbanisation

Two sets of terms have been developed to describe each of the classes of the Degree of Urbanisation (Table 31, Table 32). The first set uses simple and short terms such as city, town, suburb and village. The second set uses a more neutral and technical language. The second set can be helpful to avoid overlap with the terms used in the national definition.

Table 31 - Territorial Units classification L2 nomenclature

Code	RGB	Municipal level term
		<i>Technical term</i>
30	255 0 0	CITY <i>LARGE SETTLEMENT</i>
23	115 38 0	DENSE TOWN <i>DENSE, MEDIUM SETTLEMENT</i>
22	168 112 0	SEMI-DENSE TOWN <i>SEMI-DENSE, MEDIUM SETTLEMENT</i>
21	255 255 0	SUBURBS OR PERI-URBAN AREA <i>SEMI-DENSE AREA</i>
13	55 86 35	VILLAGE <i>SMALL SETTLEMENT</i>
12	171 205 102	DISPERSED RURAL AREA <i>LOW DENSITY AREA</i>
11	205 245 122	MOSTLY UNINHABITED AREA <i>VERY LOW DENSITY AREA</i>

³¹ https://gadm.org/download_world.html

Table 32 - Territorial Units classification L1 nomenclature

Code	RGB	Municipal level term
		<i>Technical term</i>
3	255 0 0	CITY <i>DENSELY POPULATED AREA</i>
2	255 170 0	TOWNS & SEMI-DENSE AREA <i>INTERMEDIATE DENSITY AREA</i>
1	115 178 115	RURAL AREA <i>THINLY POPULATED AREA</i>

2.9.1 How to use the statistics tables

The classification of territorial units by Degree of Urbanisation has two principal objectives. Primarily to relate available statistics (e.g. Demographic and Health Surveys, statistics on labour, housing, etc.) to a DEGURBA classification (e.g. urban/rural). Second, to account urban and rural populations for administrative areas. Both applications would harmonise nationally collected statistics to a common urban and rural classification of territorial units, useful for international statistical comparison as recommended by the United Nations Statistical Commission.

To relate available national statistics to the corresponding GHS-DUC level, it is important to verify in the mapping table Table 33 the relation between the territorial unit for which the statistic is available and the corresponding GADM level. Once the territorial designation in the statistic matches a GADM level, the statistic can be coded by Degree of Urbanisation joining or relating the field identifying the administrative unit in the statistics table, to the GADM level classification of the units by Degree of Urbanisation with the GHS-DUC field 'DEGURBA_L1' or 'DEGURBA_L2'. This operation can be conducted using any of the GHS-DUC tables GHS-DUC_GLOBE_R2022A_v1_0_GADM36_<year>_levelX, where <year> correspond to the epoch, and X to the GADM Level in the specific table. To account urban and rural populations for administrative areas by Degree of Urbanisation the user can refer to the product GHS-DUC_GLOBE_R2022A_v1_0.xlsx. The table lists for all GADM countries and territories, the population per unit by Degree of Urbanisation class and the corresponding share for the finest GADM level. The table also includes statistics aggregated at global level and per epoch.

The epoch refers to the underlying GHS-POP and GHS-SMOD epoch used, keeping the GADM geometry fixed (version 3.6.).

2.9.2 Input Data

The input data are the multi-temporal GHS-SMOD, GHS-POP and GHS-BUILT grids of the GHSL Data Package 2022 (GHS P2022). Territorial units are the Global ADMinstrative Map 3.6³²

³² <https://gadm.org/data.html>

2.9.3 Technical Details

Author: Marcello Schiavina, Michele Melchiorri, Sergio Freire, Joint Research Centre (JRC) European Commission.

Product name: GHS-DUC_GLOBE_R2022A

Spatial extent: Global

Temporal extent: from 1975 to 2030, 5 years interval

2.9.3.1 GHS-DUC Summary Statistics Table

Data organisation: XLSX with Global statistics of Population per DEGURBA class for each epoch, and sheets per epoch showing Territory or Country statistics of classification at the finest GADM level available.

Territory or Country statistics sheet attributes:

- GADM code: Numerical ID of territory in GADM
- GADM ISO: ISO code of territory in GADM
- GADM NAME: Territory name in GADM
- Selected GADM Level: Finest available GADM level for territory
- GADM level type: Description of the selected level according to GADM attribute "ENGTYPE" for the level (When GADM entry is missing description "N/A" is reported)
- Total Units: Total Territory units at selected GADM level
- Total Area km2: Total Area of the territory in km2
- Average Area km2: Average unit size of the territory at selected GADM level
- Share of Urban Population: Share of population in administrative units classified as Urban (Cities, Towns and Suburbs)
- DEGURBA L1 Population
 - Rural Area: Population of units classified as Rural Areas in Degree of Urbanisation level 1
 - Town & Semi-Dense area: Population of units classified as Town & Suburbs in Degree of Urbanisation level 1
 - City: Population of units classified as Cities in Degree of Urbanisation level 1
- DEGURBA L2 Population
 - Mostly uninhabited area: Population of units classified as Mostly uninhabited Areas in Degree of Urbanisation level 2
 - Rural dispersed area: Population of units classified as Rural dispersed Areas in Degree of Urbanisation level 2
 - Village: Population of units classified as Villages in Degree of Urbanisation level 2
 - Suburban or peri-urban area: Population of units classified as Suburbs or peri-urban areas in Degree of Urbanisation level 2
 - Semi-dense Town: Population of units classified as Semi-dense Towns in Degree of Urbanisation level 2
 - Dense Town: Population of units classified as Dense Towns in Degree of Urbanisation level 2
 - City: Population of units classified as Cities in Degree of Urbanisation level 2
- Total Pop: Total Territory population
- DEGURBA L1 Units
 - Rural Area: Number of units classified as Rural Areas in Degree of Urbanisation level 1

- Town & Semi-Dense area: Number of units classified as Town & Suburbs in Degree of Urbanisation level 1
 - City: Number of units classified as Cities in Degree of Urbanisation level 1
- DEGURBA L2 Units
 - Mostly uninhabited area: Number of units classified as Mostly uninhabited Areas in Degree of Urbanisation level 2
 - Rural dispersed area: Number of units classified as Rural dispersed Areas in Degree of Urbanisation level 2
 - Village: Number of units classified as Villages in Degree of Urbanisation level 2
 - Suburban or peri-urban area: Number of units classified as Suburbs or peri-urban areas in Degree of Urbanisation level 2
 - Semi-dense Town: Number of units classified as Semi-dense Towns in Degree of Urbanisation level 2
 - Dense Town: Number of units classified as Dense Towns in Degree of Urbanisation level 2
 - City: Number of units classified as Cities in Degree of Urbanisation level 2

2.9.3.2 GHS-DUC Admin Classification layers

Data organisation: CSV files to be joined to the original GADM3.6 layer at the respective level.

Attributes:

- GID_<level>: GADM 3.6 ID at current level [join filed with GADM layer at respective level]
- GID_0: GADM 3.6 ID at country or territory level0
- Tot_Pop: Total population
- UCentre_Pop: Urban Centre population
- UCluster_Pop: Urban Cluster population
- Rural_Pop: Rural Population
- UCentre_share: Share of Urban Centre Population
- UCluster_share: Share of Urban Cluster population
- Urban_share: Share of Urban Population (Urban Centre + Urban Cluster)
- Rural_share: Share of Rural Population
- DEGURBA_L1: Classification according to Degree of Urbanisation level 1
- DUC_Pop: Dense Urban Cluster Population
- SDUC_Pop: Semi-dense Urban Cluster Population
- S Urb_Pop: Suburban and peri-urban grid cells Population
- RC_Pop: Rural Cluster Population
- LDR_Pop: Low Density Rural grid cells Population
- VLDR_Pop: Very Low Density Rural grid cells Population
- DUC_share: Share of Dense Urban Cluster Population
- SDUC_share: Share of Semi-dense Urban Cluster Population
- S Urb_share: Share of Suburban and peri-urban grid cells Population
- RC_share: Share of Rural Cluster Population
- LDR_share: Share of Low Density Rural grid cells Population

- VLDR_share: Share of Very Low Density Rural grid cells Population
- DEGURBA_L2: Classification according to Degree of Urbanisation level 2
- BU_km2: Built-up area in km2

Classified GADM level types per country or territory: see Table 33

Table 34 outlines the technical characteristics of the datasets released in this data package.

Table 33 - *GADM level type per country or territory (level 0 omitted as representing the full country or territory)*

GADM ISO	GADM NAME	Level 1	Level 2	Level 3	Level 4	Level 5
AFG	Afghanistan	Province	District	-	-	-
XAD	Akrotiri and Dhekelia	Sovereign Base Area	-	-	-	-
ALA	Åland	Municipality	-	-	-	-
ALB	Albania	County	District	Bashkia	-	-
DZA	Algeria	Province	Chef-Lieu-Wilaya	-	-	-
ASM	American Samoa	District	County	Village	-	-
AND	Andorra	Parish	-	-	-	-
AGO	Angola	Province	Municipality City Council	Commune	-	-
AIA	Anguilla	-	-	-	-	-
ATA	Antarctica	-	-	-	-	-
ATG	Antigua and Barbuda	Dependency	-	-	-	-
ARG	Argentina	Province	Part	-	-	-
ARM	Armenia	Province	-	-	-	-
ABW	Aruba	-	-	-	-	-
AUS	Australia	Territory	Territory	-	-	-
AUT	Austria	State	Statutory City	Municipality	-	-
AZE	Azerbaijan	Region	District	-	-	-
BHS	Bahamas	District	-	-	-	-
BHR	Bahrain	Governorate	-	-	-	-
BGD	Bangladesh	Division	District	Upazilla	Union	-
BRB	Barbados	Parish	-	-	-	-
BLR	Belarus	Region	District	-	-	-
BEL	Belgium	Region	Capital Region	Arrondissement	Commune	-
BLZ	Belize	District	-	-	-	-
BEN	Benin	Department	Commune	-	-	-
BMU	Bermuda	Parish	-	-	-	-
BTN	Bhutan	District	Village block	-	-	-
BOL	Bolivia	Department	Province	Municipality	-	-
	Bonaire, Sint Eustatius and Saba	Municipality	-	-	-	-
BES	Eustatius and Saba	Municipality	-	-	-	-
BIH	Bosnia and Herzegovina	District	N/A	Commune	-	-
BWA	Botswana	District	Sub-district	-	-	-
BVT	Bouvet Island	-	-	-	-	-
BRA	Brazil	State	Municipality	District	-	-
IOT	British Indian Ocean Territory	-	-	-	-	-
VGB	British Virgin Islands	District	-	-	-	-
BRN	Brunei	District	Mukim	-	-	-
BGR	Bulgaria	Province	Municipality	-	-	-
BFA	Burkina Faso	Region	Province	Department	-	-
BDI	Burundi	Province	Commune	Colline	Sous Colline	-
KHM	Cambodia	Province	District	Commune	Village	-
CMR	Cameroon	Region	Department	Arrondissement	-	-
CAN	Canada	Province	Census Division	Town	-	-
CPV	Cape Verde	County	-	-	-	-
XCA	Caspian Sea	-	-	-	-	-
CYM	Cayman Islands	District	-	-	-	-
CAF	Central African Republic	Prefecture	Sub-prefecture	-	-	-
TCD	Chad	Region	Department	Sub-prefecture	-	-
CHL	Chile	Region	Province	Municipality	-	-
CHN	China	Province	Prefecture City	County City	-	-
CXR	Christmas Island	-	-	-	-	-
XCL	Clipperton Island	-	-	-	-	-

CCK	Cocos Islands	-	-	Corregimiento	-	-
COL	Colombia	Commissary	Autonomous Island	Departamento	-	-
COM	Comoros	-	-	-	-	-
COK	Cook Islands	Province	-	-	-	-
CRI	Costa Rica	Autonomous district	Province	Canton	-	-
CIV	Côte d'Ivoire	Autonomous district	Autonomous district	Department	Sub-prefecture	-
HRV	Croatia	County	N/A	-	-	-
CUB	Cuba	Province	Municipality	-	-	-
CWU	Curaçao	-	-	-	-	-
CYP	Cyprus	District	-	-	-	-
CZE	Czech Republic	Region	District	-	-	-
	Democratic Republic of the Congo	-	-	-	-	-
DNK	Denmark	Region	Municipality	-	-	-
DJI	Djibouti	Region	District	-	-	-
DMA	Dominica	Parish	-	-	-	-
DOM	Dominican Republic	Province	Municipality	-	-	-
ECU	Ecuador	Province	Canton	Cantonal Head	-	-
EGY	Egypt	Governorate	Subdivision	-	-	-
SLV	El Salvador	Department	Municipality	-	-	-
GNQ	Equatorial Guinea	Province	Districts Municipals	-	-	-
ERI	Eritrea	Region	District	-	-	-
EST	Estonia	County	Parish	Town	-	-
ETH	Ethiopia	City	Zone	District	-	-
FLK	Falkland Islands	-	-	-	-	-
FRO	Faroe Islands	Region	Commune	-	-	-
FJI	Fiji	Division	Province	-	-	-
FIN	Finland	Province	Region	Sub-Region	Municipality	-
FRA	France	Region	Department	Districts	Cantons	Commune
GUF	French Guiana	Arrondissement	Commune	-	-	-
PYF	French Polynesia	Administrative subdivisions	-	-	-	-
ATF	French Southern Territories	District	-	-	-	-
GAB	Gabon	Province	Department	-	-	-
GMB	Gambia	Independent City	District	-	-	-
GEO	Georgia	Autonomous Republic	District	-	-	-
DEU	Germany	State	District	Municipality	Town	-
GHA	Ghana	Region	District	-	-	-
GIB	Gibraltar	-	-	-	-	-
GRC	Greece	Decentralized administration	Region	Municipality	-	-
GRL	Greenland	Commune	-	-	-	-
GRD	Grenada	Dependency	-	-	-	-
GLP	Guadeloupe	District	Commune	-	-	-
GUM	Guam	Municipality	-	-	-	-
GTM	Guatemala	Department	Municipality	-	-	-
GGY	Guernsey	Parish	-	-	-	-
GIN	Guinea	Region	Prefecture	Sub-prefecture	-	-
GNB	Guinea-Bissau	Region	Sector	-	-	-
GUY	Guyana	Region	Not Classified	-	-	-
HTI	Haiti	Department	District	Commune	Sub-commune	-
HMD	Heard Island and McDonald Islands	-	-	-	-	-
HND	Honduras	Department	Municipality	-	-	-
HKG	Hong Kong	District	-	-	-	-
HUN	Hungary	County	Subregion	-	-	-
ISL	Iceland	Region	Municipality	-	-	-
IND	India	Union Territory	District	Taluk	-	-
IDN	Indonesia	Province	Regency	Sub-district	Village	-
IRN	Iran	Province	County	-	-	-
IRQ	Iraq	Province	N/A	-	-	-
IRL	Ireland	County	-	-	-	-
IMN	Isle of Man	N/A	N/A	-	-	-
ISR	Israel	District	-	-	-	-
ITA	Italy	Region	Province	Commune	-	-
JAM	Jamaica	Parish	-	-	-	-

JPN	Japan	Prefecture	Town	-	-	-
JEY	Jersey	Parish	-	-	-	-
JOR	Jordan	Province	Sub-Province	-	-	-
KAZ	Kazakhstan	Region	District	-	-	-
KEN	Kenya	County	Constituency	Ward	-	-
KIR	Kiribati	-	-	-	-	-
XKO	Kosovo	District	Town Municipal	-	-	-
KWT	Kuwait	Province	-	-	-	-
KGZ	Kyrgyzstan	Province	District	-	-	-
LAO	Laos	Province	District	-	-	-
LVA	Latvia	Province	District	-	-	-
LBN	Lebanon	Governorate	District	Municipality	-	-
LSO	Lesotho	District	-	-	-	-
LBR	Liberia	County	District	Clan	-	-
LBY	Libya	District	-	-	-	-
LIE	Liechtenstein	Commune	-	-	-	-
LTU	Lithuania	County	District Municipality	-	-	-
LUX	Luxembourg	District	Canton	Commune	Commune (same as level 3)	-
MAC	Macao	District	Parish	-	-	-
MKD	Macedonia	Municipality	-	-	-	-
MDG	Madagascar	Autonomous Province	Region	District	Commune	-
MWI	Malawi	District	Town	Unknown	-	-
MYS	Malaysia	State	District	-	-	-
MDV	Maldives	-	-	-	-	-
MLI	Mali	District	Circle	Arrondissement	Commune	-
MLT	Malta	Region	Local council	-	-	-
MHL	Marshall Islands	-	-	-	-	-
MTQ	Martinique	Arrondissement	Commune	-	-	-
MRT	Mauritania	Region	Department	-	-	-
MUS	Mauritius	Region	-	-	-	-
MYT	Mayotte	Commune	-	-	-	-
MEX	Mexico	State	Municipality	-	-	-
FSM	Micronesia	State	-	-	-	-
MDA	Moldova	District	-	-	-	-
MCO	Monaco	-	-	-	-	-
MNG	Mongolia	Province	Sum	-	-	-
MNE	Montenegro	Municipality	-	-	-	-
MSR	Montserrat	Parish	-	-	-	-
MAR	Morocco	Region	Province	District	Rural Commune	-
MOZ	Mozambique	Province	District	Locality	-	-
MMR	Myanmar	Division	District	Village Township	-	-
NAM	Namibia	Region	Constituency	-	-	-
NRU	Nauru	District	-	-	-	-
NPL	Nepal	Development Region	Administrative Zone	District	Village development committee	-
NLD	Netherlands	Province	Municipality	-	-	-
NCL	New Caledonia	Province	Commune	-	-	-
NZL	New Zealand	Region	District	-	-	-
NIC	Nicaragua	Autonomous Region	Municipality	-	-	-
NER	Niger	Department	Arrondissement	Commune	-	-
NGA	Nigeria	State	Local Authority	-	-	-
NIU	Niue	-	-	-	-	-
NFK	Norfolk Island	-	-	-	-	-
PRK	North Korea	Province	County	-	-	-
XNC	Northern Cyprus	District	-	-	-	-
MNP	Northern Mariana Islands	Municipality	-	-	-	-
NOR	Norway	County	Municipality	-	-	-
OMN	Oman	Region	Province	-	-	-
PAK	Pakistan	Centrally Administered Area	Division	District	-	-
PLW	Palau	State	-	-	-	-
PSE	Palestina	District	Governorate	-	-	-
PAN	Panama	Province	District	Municipality	-	-
PNG	Papua New Guinea	Autonomous Region	District	-	-	-
XPI	Paracel Islands	-	-	-	-	-
PRY	Paraguay	Department	District	-	-	-

PER	Peru	Region	Province	District	-	-
PHL	Philippines	Province	Municipality	Village	-	-
PCN	Pitcairn Islands	-	-	-	-	-
POL	Poland	Province Voivodeship	County	Municipality (urban)	-	-
PRT	Portugal	District	Municipality	Parish	-	-
PRI	Puerto Rico	Municipality	-	-	-	-
QAT	Qatar	Municipality	-	-	-	-
COG	Republic of Congo	Region	District	-	-	-
REU	Reunion	Arrondissement	Commune	-	-	-
ROU	Romania	County	Commune	-	-	-
RUS	Russia	Republic	District	N/A	-	-
RWA	Rwanda	Province	District	Sector	Cell	Village
BLM	Saint-Barthélemy	-	-	-	-	-
MAF	Saint-Martin	-	-	-	-	-
SHN	Saint Helena	Administrative Area	Administrative Area	-	-	-
KNA	Saint Kitts and Nevis	Parish	-	-	-	-
LCA	Saint Lucia	Quarter	-	-	-	-
SPM	Saint Pierre and Miquelon	Commune	-	-	-	-
VCT	Saint Vincent and the Grenadines	Parish	-	-	-	-
WSM	Samoa	District	Unknown	-	-	-
SMR	San Marino	Municipality	-	-	-	-
STP	São Tomé and Príncipe	Municipality	N/A	-	-	-
SAU	Saudi Arabia	Region	-	-	-	-
SEN	Senegal	Region	Department	Arrondissement	Commune	-
SRB	Serbia	District	Town Municipal	-	-	-
SYC	Seychelles	District	-	-	-	-
SLE	Sierra Leone	Province	District	Chiefdom	-	-
SGP	Singapore	Region	-	-	-	-
SXM	Sint Maarten	-	-	-	-	-
SVK	Slovakia	Region	District	-	-	-
SVN	Slovenia	Statistical Region	Commune Municipality	-	-	-
SLB	Solomon Islands	Province	Ward	-	-	-
SOM	Somalia	Region	District	-	-	-
ZAF	South Africa	Province	District Municipality	Local Municipality	Ward	-
SGS	South Georgia and the South Sandwich Islands	-	-	-	-	-
KOR	South Korea	Metropolitan City	District	-	-	-
SSD	South Sudan	State	District	Unknown	-	-
ESP	Spain	Autonomous Community	Province	Comarca	Municipality	-
XSP	Spratly Islands	-	-	-	-	-
LKA	Sri Lanka	District	Division	-	-	-
SDN	Sudan	State	District	Unknown	-	-
SUR	Suriname	District	Ressort	-	-	-
SJM	Svalbard and Jan Mayen	Territory	-	-	-	-
SWZ	Swaziland	District	Constituency	-	-	-
SWE	Sweden	County	Municipality	-	-	-
CHE	Switzerland	Canton	District	Municipality	-	-
SYR	Syria	Province	District	-	-	-
TWN	Taiwan	Province	County	-	-	-
TJK	Tajikistan	Region	N/A	N/A	-	-
TZA	Tanzania	Region	District	Ward	-	-
THA	Thailand	Province	District	Sub district	-	-
TLS	Timor-Leste	District	Subdistrict	Community	-	-
TGO	Togo	Region	Prefecture	-	-	-
TKL	Tokelau	Atoll	-	-	-	-
TON	Tonga	Island Group	-	-	-	-
TTO	Trinidad and Tobago	Borough	-	-	-	-
TUN	Tunisia	Governorate	Delegation	-	-	-
TUR	Turkey	Province	District	-	-	-
TKM	Turkmenistan	Province	-	-	-	-
TCA	Turks and Caicos Islands	District	-	-	-	-

TUV	Tuvalu	Town Council	-	-	-	-
UGA	Uganda	District	County	Sub-county	Parish	-
UKR	Ukraine	Region	Mis'ka Rada	-	-	-
ARE	United Arab Emirates	Emirate	Municipal Region	Municipality	-	-
GBR	United Kingdom	Kingdom	Metropolitan Borough	Metropolitan borough	-	-
USA	United States	State	County	-	-	-
UMI	United States	Island	-	-	-	-
URY	Minor Outlying Islands	Department	Poblacion	-	-	-
UZB	Uruguay	Region	District	-	-	-
VUT	Uzbekistan	Province	Area council	-	-	-
VAT	Vanuatu	-	-	-	-	-
VEN	Vatican City	State	Municipality	-	-	-
VNM	Venezuela	Province	District	Townlet	-	-
VIR	Vietnam	District	N/A	-	-	-
WLF	Virgin Islands, U.S.	Kingdom	District	-	-	-
ESH	Wallis and Futuna	Province	-	-	-	-
YEM	Western Sahara	Governorate	District	-	-	-
ZMB	Yemen	Province	District	-	-	-
ZWE	Zambia	City	District	-	-	-
	Zimbabwe					

Table 34 - Technical details of the datasets in GHS-DUC_GLOBE_R2022A

GHS-DUC_GLOBE_R2022A	
ID	Description
GHS_DUC_GLOBE_R2022A_V1_0	Global Degree of Urbanisation by epoch File format: excel table
GHS_DUC_GLOBE_R2022A_V1_0_GADM36_E<epoch>_<level>	Degree of Urbanisation of GADM36 <level> 0-5 units in epoch 1975-2020 (5 years interval) File format: comma separated values file
GHS_DUC_GLOBE_R2022A_V1_0_GADM36_P<epoch>_<level>	Degree of Urbanisation of GADM36 <level> 0-5 units in projected epoch 2025-2030 File format: comma separated values file

2.9.4 How to cite

Dataset:

Schiavina, Marcello; Melchiorri, Michele; Freire, Sergio (2022): GHS-DUC R2022A - GHS Degree of Urbanisation Classification, application of the Degree of Urbanisation methodology (stage II) to GADM 3.6 layer, multitemporal (1975-2030). European Commission, Joint Research Centre (JRC) [Dataset] DOI: 10.2905/F5224214-6B66-43DF-A9C6-CC974F17D803 PID: <http://data.europa.eu/89h/f5224214-6b66-43df-a9c6-cc974f17d803>

Concept & Methodology:

European Commission, and Statistical Office of the European Union, 2021. Applying the Degree of Urbanisation — A methodological manual to define cities, towns and rural areas for international comparisons — 2021 edition Publications Office of the European Union, 2021, ISBN 978-92-76-20306-3 doi:10.2785/706535

2.10 GHS-BUILT-LAU2STAT R2022A - GHS built-up surface statistics in European LAU, multitemporal (1975-2020)

This product contains the summary statistics of GHS-BUILT-S multi-temporal (1975-2020) at Local Administrative Unit level (LAU) from the 2020 polygon layer provided by GISCO.

For each LAU the table contains the sum of built-up surface (GHS-BUILT-S) between 1975 and 2020 in 5 years interval, expressed in square kilometres.

2.10.1 Input data

The new product GHS-BUILT-S_GLOBE_R2022A (version 1.0) was used to compute the total built-up surface present at each epoch in EU LAU. The LAU geometry is obtained from GISCO³³.

2.10.2 Technical Details

Author: Marcello Schiavina, Michele Melchiorri, Joint Research Centre (JRC) European Commission

Product name: GHS-BUILT-LAU2STAT_EUROPE_R2022A

Spatial extent: Global

Temporal extent: 1975-2020 (5 years interval)

Data organisation: Data are provided as excel table with LAU codes ("GISCO ID"), country code ("Country") and built-up in square kilometres for each epoch

Table 10 outlines the technical characteristics of the datasets released in this data package.

Table 35 - Technical details of the datasets in GHS-BUILT-LAU2STAT_EUROPE_R2022A

GHS-BUILT-LAU2STAT_EUROPE_R2022A	
ID	Description
GHS_BUILT_LAU2STAT_MT_EUROPE_R2022A_V1_0	zonal sums of built-up surface MT by LAU

2.10.3 How to cite

Dataset:

Schiavina, Marcello; Melchiorri, Michele (2022): GHS-BUILT-LAU2STAT R2022A - GHS built-up surface statistics in European LAU2, multitemporal (1975-2020). European Commission, Joint Research Centre (JRC) [Dataset] DOI:10.2905/94D62A61-25D0-42FD-9E1E-A41F877CF788 PID: <http://data.europa.eu/89h/94d62a61-25d0-42fd-9e1e-a41f877cf788>

Concept & Methodology:

Schiavina, Marcello; Melchiorri, Michele; Corbane, Christina; Freire, Sergio; Batista e Silva, Filipe (2022): Built-up areas are expanding faster than population growth: regional patterns and trajectories in Europe, Journal of Land Use Science, DOI: 10.1080/1747423X.2022.2055184

³³ <https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/lau>

2.11 GHS-SDATA R2022A - GHSL data supporting the production of R2022A Data Package (GHS-BUILT, GHS-POP)

The GHS-SDATA product contains several intermediate data supporting the production of the R2022A with the function of baseline or quality control: consequently they have a downstream effect on the characteristics of the final information included in the R2022A. They are shared for the purpose of a better understanding of the GHSL data characteristics and facilitate the transparency of the GHSL production workflow.

GHS_SDATA_LDS_QUANT and GHS_SDATA_LDS_CONF reports about the quantity of Landsat image data measurements used by the GHSL for assessing each specific epoch, and the confidence of change based on these image data, respectively, both aggregated at 100m-resolution.

GHS_SDATA_BUILT_CHANGE reports about the binary change map at 10m-resolution on the satellite-image-data-observed five epochs (1975, 1990, 2000, 2014, and 2018). This change map it is the baseline supporting the interpolated grids in equal time interval (5-years) 1975–2030 in surface and volume, which are published under the GHS-BUILT-S and GHS-BUILT-V data, respectively, at 100m-resolution.

GHS_SDATA_E2018_S2MAXNDVI reports about the maximum NDVI collected from Sentinel-2 images in 2018. These data it is used in support of the inferential engine for the non-residential (NRES) automatic classification in order to discriminate large agricultural fields with rectangular shape similar to industrial buildings, and in support of the settlement characteristics classification included in the GHS-BUILT-C product.

GHS_SDATA_WUP2018_BOUNDARIES reports about the ‘city’ boundaries as have been estimated during the GHSL production from the UN World Urbanization Prospects 2018. WUP 2018 includes population time series for cities represented by single location points (coordinates). GHSL developed an automatic approach for inferring the city boundaries and extent in WUP database that is based on iterative aggregation of administrative units adjacent to the main unit (determined by WUP city coordinates), using density and compactness criteria to reach the WUP city population data in the available census year. The procedure leverages on the GHS-SmartDissolve tool.

2.11.1 Technical Details

Author: Pesaresi, Martino; Politis, Panagiotis; Schiavina, Marcello; Freire, Sergio; Maffenini, Luca

Product name: GHS-SDATA_GLOBE_R2022A

Spatial extent: Global

Table 36 - Technical details of the datasets in GHS_BUILT_S1NODSM_GLOBE_R2018A

GHS-SDATA_LDS-QUANT_GLOBE_R2022A		
ID	Description	Resolution (projection)
GHS_SDATA_LDS_QUANT_MT_GLO_BE_R2022A_54009_100_V1_0	Amount of Landsat image data observations supporting the multi-temporal assessment in the different epochs Encoding: Byte Number of Bands: 4 (b1 : 1975, b2 : 1990, b3 : 2000, b4 : 2014) Values range: 0-254 NoData: 255	100 m World Mollweide (ESRI:54009)

GHS-SDATA_LDS-CONF_GLOBE_R2022A		
ID	Description	Resolution (projection)
GHS_SDATA_LDS_CONF_MT_GLOBE_R2022A_54009_100_V1_0	<p>Confidence on the multi-temporal change based on Landsat image data at each epoch</p> <p>Encoding: Byte</p> <p>Number of Bands: 4 (b1 : 1975, b2 : 1990, b3 :2000, b4 : 2014)</p> <p>Values range: 0-100</p> <p>NoData: 255</p>	100 m World Mollweide (ESRI:54009)

GHS-SDATA_S2MAXNDVI_GLOBE_R2022A		
ID	Description	Resolution (projection)
GHS_SDATA_E2018_S2MAXNDVI_GLOBE_R2022A_54009_10_V1_0	<p>max NDVI from S2 data composite</p> <p>Encoding: Byte</p> <p>Values range: 0-100</p> <p>NoData: 255</p>	10 m World Mollweide (ESRI:54009)

GHS-SDATA_BUILT-CHANGE_GLOBE_R2022A		
ID	Description	Resolution (projection)
GHS_SDATA_BUILT_CHANGE_MT_GLOBE_R2022A_54009_10_V1_0	<p>change map 10m 2018-2014-2000-1990-1975</p> <p>Encoding: Byte</p> <p>Values range: 0-5 (0: non built-up 1: built-up since 1975, 2: built-up from 1975 to 1990, 3: built-up from 1990 to 2000, 4: built-up from 2000 to 2014, 5: built-up from 2014 to 2018)</p> <p>NoData: 255</p>	10 m World Mollweide (ESRI:54009)

GHS-SDATA_WUP2018-BOUNDARIES_GLOBE_R2022A		
ID	Description	Resolution (projection)
GHS_SDATA_WUP2018_BOUNDARIES_MT_GLOBE_R2022A_V1_0	<p>Estimated city boundaries matching UN World Urbanization Prospects 2018</p> <p>File format: shapefile</p>	World Mollweide (ESRI:54009)

2.11.2 How to cite

Dataset:

Pesaresi, Martino; Politis, Panagiotis; Schiavina, Marcello; Freire, Sergio; Maffenini, Luca (2022): GHS-SDATA R2022A - GHS release R2022A supporting data. European Commission, Joint Research Centre (JRC) [Dataset] DOI:10.2905/3A1249B2-499A-4E0C-9DC3-518B39575448 PID: <http://data.europa.eu/89h/3a1249b2-499a-4e0c-9dc3-518b39575448>

Concept & Methodology:

Schiavina M, Melchiorri M, Pesaresi M, Politis P, Freire S, Maffenini L, Florio P, Ehrlich D, Goch K, Tommasi P, Kemper T, GHSL Data Package 2022, Publications Office of the European Union, Luxembourg, 2022, ISBN 978-92-76-53071-8, doi:10.2760/19817, JRC 129516

References

- Bechtel, B., Pesaresi, M., Florczyk, A. J., & Mills, G. (2018). Beyond built-up: The internal makeup of urban areas. *Urban Remote Sensing*.
- Buchhorn, M., Smets, B., Bertels, L., Roo, B. D., Lesiv, M., Tsendbazar, N.-E., Herold, M., & Fritz, S. (2020). *Copernicus Global Land Service: Land Cover 100m: collection 3: epoch 2019: Globe* [Data set]. Zenodo. <https://doi.org/10.5281/ZENODO.3939050>
- Corbane, C., Pesaresi, M., Kemper, T., Politis, P., Florczyk, A. J., Syrris, V., Melchiorri, M., Sabo, F., & Soille, P. (2019). Automated global delineation of human settlements from 40 years of Landsat satellite data archives. *Big Earth Data*, 3(2), 140–169. <https://doi.org/10.1080/20964471.2019.1625528>
- 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., & Kemper, T. (2017). Big earth data analytics on Sentinel-1 and Landsat imagery in support to global human settlements mapping. *Big Earth Data*, 1(1–2), 118–144. <https://doi.org/10.1080/20964471.2017.1397899>
- Corbane, C., Politis, P., Kempeneers, P., Simonetti, D., Soille, P., Burger, A., Pesaresi, M., Sabo, F., Syrris, V., & Kemper, T. (2020). A global cloud free pixel-based image composite from Sentinel-2 data. *Data in Brief*, 31, 105737. <https://doi.org/10.1016/j.dib.2020.105737>
- Corbane, C., Syrris, V., Sabo, F., Politis, P., Melchiorri, M., Pesaresi, M., Soille, P., & Kemper, T. (2020). Convolutional neural networks for global human settlements mapping from Sentinel-2 satellite imagery. *Neural Computing and Applications*. <https://doi.org/10.1007/s00521-020-05449-7>
- European Commission, & Statistical Office of the European Union. (2021). *Applying the Degree of Urbanisation – a methodological manual to define cities, towns and rural areas for international comparisons*. Publications Office of the European Union.
- Freire, S., MacManus, K., Pesaresi, M., Doxsey-Whitefield, E., & Mills, J. (2016). Development of new open and free multi-temporal global population grids at 250 m resolution. *Proc. of the 19th AGILE Conference on Geographic Information Science*, 250.
- Freire, S., Schiavina, M., Florczyk, A., MacManus, K., Pesaresi, M., Corbane, C., Bokovska, O., Mills, J., Pistolesi, L., Squires, J., & Sliuzas, R. (2018). Enhanced data and methods for improving open and free global population grids: putting ‘leaving no one behind’ into practice. *International Journal of Digital Earth*, 11(12). <https://doi.org/10.1080/17538947.2018.1548656>
- Gueguen, L., Soille, P., & Pesaresi, M. (2012). A new built-up presence index based on density of corners. *Proc. Int. Symp. on Geoscience and Remote Sensing (IGARSS)*.
- Maffenini, L., Schiavina, M., Freire, S., M. Melchiorri, & Kemper, T. (2021). *GHS-POPWARP User Guide*. Publications Office of the European Union. doi:10.2760/060678
- Melchiorri, M., Pesaresi, M., Florczyk, A. J., Corbane, C., & Kemper, T. (2019). Principles and Applications of the Global Human Settlement Layer as Baseline for the Land Use Efficiency Indicator—SDG 11.3.1. *ISPRS International Journal of Geo-Information*, 8(2), 96. <https://doi.org/10.3390/ijgi8020096>
- Ouzounis, G. K., Pesaresi, M., & Soille, P. (2012). Differential Area Profiles: decomposition properties and efficient computation. *Pattern Analysis and Machine Intelligence, IEEE Transactions On*, 34(8), 1533–1548. <https://doi.org/10.1109/TPAMI.2011.245>
- Pekel, J.-F., Cottam, A., Gorelick, N., & Belward, A. (2016). High-resolution mapping of global surface water and its long-term changes. *Nature*, 540(7633), 418–422. <https://doi.org/10.1038/nature20584>
- Pesaresi, M., & Benediktsson, J. A. (2001). A new approach for the morphological segmentation of high-resolution satellite imagery. *Geoscience and Remote Sensing, IEEE Transactions On*, 39(2), 309–320. <https://doi.org/10.1109/36.905239>
- Pesaresi, M., Corbane, C., Julea, A., Florczyk, A., Syrris, V., & Soille, P. (2016). Assessment of the Added-Value of Sentinel-2 for Detecting Built-up Areas. *Remote Sensing*, 8(4), 299. <https://doi.org/10.3390/rs8040299>

- Pesaresi, M., Corbane, C., Ren, C., & Edward, N. (2021). Generalized Vertical Components of built-up areas from global Digital Elevation Models by multi-scale linear regression modelling. *PLOS ONE*, 16(2), e0244478. <https://doi.org/10.1371/journal.pone.0244478>
- Pesaresi, M., Ehrlich, D., Ferri, S., Florczyk, A., Carneiro Freire Sergio, M., Halkia, S., Julea, A., Kemper, T., Soille, P., & Syrris, V. (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/1111111111/40182>
- Pesaresi, M., Gerhardinger, A., & Kayitakire, F. (2008). A Robust Built-Up Area Presence Index by Anisotropic Rotation-Invariant Textural Measure. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 1(3), 180–192. <https://doi.org/10.1109/JSTARS.2008.2002869>
- Pesaresi, M., Huadong, G., Blaes, X., Ehrlich, D., Ferri, S., Gueguen, L., Halkia, M., Kauffmann, M., Kemper, T., Lu, L., Marin-Herrera, M. A., Ouzounis, G. K., Scavazzon, M., Soille, P., Syrris, V., & Zanchetta, L. (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), 2102–2131. <https://doi.org/10.1109/JSTARS.2013.2271445>
- Pesaresi, M., Ouzounis, G. K., & Gueguen, L. (2012). A new compact representation of morphological profiles: Report on first massive VHR image processing at the JRC. *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVIII*, 8390, 839025.
- Pesaresi, M., Syrris, V., & Julea, A. (2016). A New Method for Earth Observation Data Analytics Based on Symbolic Machine Learning. *Remote Sensing*, 8(5), 399. <https://doi.org/10.3390/rs8050399>
- Schneider, A., Friedl, M. A., & Potere, D. (2010). Mapping global urban areas using MODIS 500-m data: New methods and datasets based on ‘urban ecoregions.’ *Remote Sensing of Environment*, 114(8), 1733–1746. <https://doi.org/10.1016/j.rse.2010.03.003>
- See, L., Georgieva, I., Duerauer, M., Kemper, T., Corbane, C., Maffenini, L., Gallego, J., Pesaresi, M., Sirbu, F., Ahmed, R., & others. (2022). A crowdsourced global data set for validating built-up surface layers. *Scientific Data*, 9(1), 1–14.
- Theobald, D. M. (2014). Development and applications of a comprehensive land use classification and map for the US. *PloS One*, 9(4), e94628.
- United Nations, Department of Economic and Social Affairs, Population Division. (2018). *World Urbanization Prospects: The 2018 Revision (ST/ESA/SER.A/420)*. United Nations.
- United Nations, Department of Economic and Social Affairs, Population Division. (2019). *World Population Prospects 2019, Online Edition*. Rev. 1.
- Vincent, L. (1993). Morphological grayscale reconstruction in image analysis: applications and efficient algorithms. *Image Processing, IEEE Transactions On*, 2(2), 176–201. <https://doi.org/10.1109/83.217222>

List of figures

Figure 1 - Santiago de Chile: comparison of the built-up surfaces as assessed by the previous GHS_BUILT_LDSMT_GLOBE_R2018A for the epoch 2014 from Landsat image data with a Boolean 30m-resolution method (upper), vs the new GHS-BUILT-S_GLOBE_R2022A for the epoch 2018 from Sentinel-2 image data with a continuous 10m-resolution method (lower).....	9
Figure 2 - Mumbai-Pune (India): residential (RES) and non-residential (NRES) components of the built-surfaces estimated for the GHSL 2020 epoch. RES and NRES are represented with blue and magenta, respectively....	10
Figure 3 - Shanghai-Changzhou (China): residential (RES) and non-residential (NRES) components of the built-surfaces estimated for the GHSL 2020 epoch. RES and NRES are represented with blue and magenta, respectively.....	10
Figure 4 - Lagos-Porto Novo-Abeokuta (Nigeria): residential (RES) and non-residential (NRES) components of the built-surfaces estimated for the GHSL 2020 epoch. RES and NRES are represented with blue and magenta, respectively.....	11
Figure 5 - Sao Paulo- Campinas - Sao Jose dos Campos (Brazil): residential (RES) and non-residential (NRES) components of the built-surfaces estimated for the GHSL 2020 epoch. RES and NRES are represented with blue and magenta, respectively.....	11
Figure 6 - Detroit-Lansing-Flint (United States): residential (RES) and non-residential (NRES) components of the built-surfaces estimated for the GHSL 2020 epoch. RES and NRES are represented with blue and magenta, respectively.....	12
Figure 7 - The Hague - Rotterdam- Antwerp (The Netherlands): residential (RES) and non-residential (NRES) components of the built-surfaces estimated for the GHSL 2020 epoch. RES and NRES are represented with blue and magenta, respectively.....	12
Figure 8 – Average building height (ANBH 100m) estimates in Guangzhou - Shenzhen (China).....	24
Figure 9 - Average building height (ANBH 100m) estimates in Delhi (India).	24
Figure 10 - Average building height (ANBH 100m) estimates in Pretoria – Johannesburg (South Africa).	25
Figure 11 - Average building height (ANBH 100m) estimates in Paris (France).	25
Figure 12 - Average building height (ANBH 100m) estimates in Sao Paulo – Santos (Brazil).	26
Figure 13 - Average building height (ANBH 100m) estimates in New York (United States).....	26
Figure 14 - Osaka-Nagoya-Tokyo (Japan): comparison between GHS-BUILT-V R2022A (above) and GHS-BUILT-S R2022A (below), year 2020.	31
Figure 15 - Morphological Settlement Zone (MSZ) Legend.....	35
Figure 16 - Settlement Characteristics in Kolkata (India)	36
Figure 17 - Settlement Characteristics in Beijing (China)	37
Figure 18 - Settlement Characteristics in Kampala (Uganda).....	38
Figure 19 – Settlement Characteristics in London (United Kingdom).....	39
Figure 20 – Settlement Characteristics in Kansas City (United States)	40
Figure 21 – Settlement Characteristics in Mexico City (Mexico)	41
Figure 22 - GHS LAND in the Congo River.....	44
Figure 23 - GHS Population grid (GHS-POP) GHS_POP_E2020_GLOBE_R2022A_54009_100_V1_0 in Puerto Alegre (Brazil).	46
Figure 24 - Generalized workflow for producing GHS-POP R2022, with main steps (1-6) and intermediate and main outputs (A-D).....	47
Figure 25 - GHS Settlement Model grid (GHS-SMOD) GHS-SMOD_P2020_GLOBE_R2022A_54009_1K_V1_0 displayed in the area of Lagos (Nigeria) –Legend in Table 18. The boundaries and the names shown on this map do not imply official endorsement or acceptance by the European Union © OpenStreetMap	51

Figure 26 - Schematic overview of GHSL SMOD entities workflow logic. “xpop” represents the population abundance per grid cell; “xpop_dens” represents the population density on permanent land; “xbu” represents the built-up density per grid cell; “xbu_dens” represents the built-up density on permanent land. “DENSITY ON LAND” process fill built-up cells on water with max between 0.5 and built-up surface value and population on water with global average built-up per capita. “GENERALISATION” process performs a median filtering (3x3) for smoothing boundaries and fills gaps above 15 km ² . (*) this procedure of enforcement logic allows the delineation of Urban Clusters Entities which contains by definition the Urban Centres and all 2X classes. Each entity has a corresponding vector boundary.....	54
Figure 27 - GHS Degree of Urbanisation Classification (GHS-DUC) GHS-DUC_GLOBE_R2022A_V1_0_GADM36_2020_level4 joined to the GADM 3.6 level 4 layer, displayed in the area of Katowice (Poland) showing the classification of local units by Degree of Urbanisation Level 2—Legend in Table 31. The boundaries shown on this map do not imply official endorsement or acceptance by the European Union.	64
Figure 28 - Urban centre, urban cluster and rural grid cells around Cape Town, South Africa	65
Figure 29 - City, towns & semi-dense areas and rural areas around Cape Town, South Africa (classification of Main Places units, note that Cape Peninsula is part of Cape Town Main Place).....	65
Figure 30 - Degree of urbanisation level 2 grid classification around Toulouse, France.....	66
Figure 31 - Degree of urbanisation level 2 local unit classification around Toulouse, France.....	66

List of tables

Table 1 – Expected error scores in prediction of the built-up surface fraction (BUFRAC) at 10m resolution in the new GHS-BUILT-S R2022A release.....	14
Table 2 – Expected error scores in prediction of the built-up surface fraction (BUFRAC) at the aggregated 100m and 1km resolution.....	14
Table 3 – The amount of total built-up surfaces and NRES built-up surfaces assessed in the GHS-BUILT-S R2022A data stratified by land use classes in United States (NLUD) and Europe (CLC).....	15
Table 4 - Summary of the characteristics of the new GHS-BUILT data vs. the previous releases.....	17
Table 5 - Summary of the Landsat Image data used in input.....	18
Table 6 - Technical details of the datasets in GHS-BUILT-S_GLOBE_R2022A.....	19
Table 7 - Summary statistics of total built-up surface in millions of square meters, by continent. Years 2025 and 2030 are reported for the linear (LIN), second-order polynomial (PLY), and median (MED) future projection scenarios.....	20
Table 8 - Summary statistics of non-residential (NRES) built-up surface in millions of square meters, by continent. Years 2025 and 2030 are reported for the linear (LIN), second-order polynomial (PLY), and median (MED) future projection scenarios.....	21
Table 9 – Errors of ANBH 100m in predicting the Copernicus Building Height generalized at the same spatial resolution.....	27
Table 10 - Technical details of the datasets in GHS-BUILT-H_GLOBE_R2022A	28
Table 11 - Technical details of the datasets in GHS-BUILT-V_GLOBE_R2022A	32
Table 12 - Summary statistics of total built-up volume in millions of cubic meters, by continent. Years 2025 and 2030 are reported for the linear (LIN), second-order polynomial (PLY), and median (MED) future projection scenarios.....	33
Table 13 - Summary statistics of non-residential (NRES) built-up volume in millions of cubic meters, by continent. Years 2025 and 2030 are reported for the linear (LIN), second-order polynomial (PLY), and median (MED) future projection scenarios	33
Table 14 - Technical details of the datasets in GHS-BUILT-C_GLOBE_R2022A.....	42
Table 15 - Technical details of the datasets in GHS-LAND_GLOBE_R2022A	45
Table 16 - Technical details of the datasets in GHS-POP_GLOBE_R2022A.....	49
Table 17 - Summary statistics of total population as obtained from the 1-km World Mollweide grid - total population adjusted to the UN WPP 2019 (United Nations, Department of Economic and Social Affairs, Population Division, 2019).	49
Table 18 - Settlement Model L2 nomenclature.....	56
Table 19 - Settlement Model L2 synthetic explanation of logical definition and grid cell sets.....	57
Table 20 - Settlement Model L2 grid cells population and built-up area characteristics (densities on permanent land)	58
Table 21 - Settlement Model L1 nomenclature.....	58
Table 22 - Settlement Model L1 synthetic explanation of logical definition and grid cell sets.....	58
Table 23 - Settlement Model L1 grid cells population and built-up area characteristics (densities on permanent land)	59
Table 24 - Technical details of the datasets in GHS-SMOD_GLOBE_R2022A.....	60
Table 25 - Summary statistics of total area in square kilometres of each settlement typology at global level as obtained from the 1-km GHS-POP grids L2.	61
Table 26 - Summary statistics of total built-up area in square kilometres for each settlement typology at global level as obtained from the 1-km GHS-POP grids L2.	61

Table 27 - Summary statistics of total population in each settlement typology at global level as obtained from the 1-km GHS-POP grids L2.....	62
Table 28 - Summary statistics of total area in square kilometres of each settlement typology at global level as obtained from the 1-km GHS-POP grids L1.....	62
Table 29 - Summary statistics of total built-up area in square kilometres for each settlement typology at global level as obtained from the 1-km GHS-POP grids L1.....	62
Table 30 - Summary statistics of total population in each settlement typology at global level as obtained from the 1-km GHS-POP grids L1.....	62
Table 31 - Territorial Units classification L2 nomenclature.....	67
Table 32 - Territorial Units classification L1 nomenclature.....	68
Table 33 - <i>GADM level type per country or territory (level 0 omitted as representing the full country or territory)</i>	71
Table 34 - Technical details of the datasets in GHS-DUC_GLOBE_R2022A.....	75
Table 35 - Technical details of the datasets in GHS-BUILT-LAU2STAT_EUROPE_R2022A.....	77
Table 36 - Technical details of the datasets in GHS_BUILT_S1NODSM_GLOBE_R2018A.....	78

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
- at the following standard number: +32 22999696, or
- by electronic mail via: https://europa.eu/european-union/contact_en

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications

You can download or order free and priced EU publications from EU Bookshop at: <https://publications.europa.eu/en/publications>. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

The European Commission's science and knowledge service

Joint Research Centre

JRC Mission

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



EU Science Hub
ec.europa.eu/jrc



@EU_ScienceHub



EU Science Hub - Joint Research Centre



EU Science, Research and Innovation



EU Science Hub



Publications Office
of the European Union

doi:10.2760/19817

ISBN 978-92-76-53071-8