Documentation for the Global Urban Polygons and Points Dataset (GUPPD), Version 1

July 2024

Center for International Earth Science Information Network (CIESIN), Columbia University
Joint Research Centre (JRC), European Commission

Abstract

The Global Urban Polygons and Points Dataset (GUPPD), Version 1 is a global data set of 123,034 urban settlements with place names and population for the years 1975-2030 in five-year increments. The data set builds on and expands the European Commission, Joint Research Centre's (JRC) 2015 Global Human Settlement (GHS) Urban Centre Database (UCDB). The JRC Settlement Model (GHS-SMOD) data set includes a hierarchy of urban settlements, from urban centre (level 30), to dense urban cluster (level 23), to semi-dense urban cluster (level 22). The UCDB only includes level 30, whereas the GUPPDv1 adds levels 23 and 22, and uses open data sources to both check and validate the names that JRC assigned to its UCDB polygons and to label the newly added settlements. The methodology described in this documentation was able to consistently label a greater percentage of UCDB polygons than were previously labeled by JRC.

Data set citation: Center for International Earth Science Information Network (CIESIN), Columbia University, and Joint Research Centre (JRC), European Commission. 2024. Global Urban Polygons and Points Dataset (GUPPD), Version 1. Palisades, New York: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/brq1-xc29. Accessed DAY MONTH YEAR.

Suggested citation for this document: Center for International Earth Science Information Network (CIESIN), Columbia University, and Joint Research Centre (JRC), European Commission. 2024. Documentation for the Global Urban Polygons and Points Dataset (GUPPD), Version 1. Palisades, New York: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/89j7-zw60. Accessed DAY MONTH YEAR.

We appreciate feedback regarding this data set, such as suggestions, discovery of errors, difficulties in using the data, and format preferences. Please contact:

NASA Socioeconomic Data and Applications Center (SEDAC)
Center for International Earth Science Information Network (CIESIN)
Columbia University
Phone: 1 (845) 365-8920
Email: ciesin.info@ciesin.columbia.edu

Contents

I. Introduction

A well labeled geospatial urban points and polygon data set is tremendously useful for a range of applications, as shown by the wide array of uses of the Global Rural-Urban Mapping Project's (GRUMP) Urban Extent Polygons, v1.02 (Balk et al., 2006, CIESIN et al., 2021). The Global Urban Polygons and Points Dataset (GUPPD), Version 1 (GUPPDv1) aims to build upon the European Commission Joint Research Centre's (JRC) Global Human Settlement Model (GHS-SMOD). This model uses population density and land cover data in order to classify the world along the urban-rural continuum (Florczyk et al., 2019a).[1] One advantage of GHS-SMOD is that it implements the Degree of Urbanisation (DEGURBA), which is a classification system supported by the UN for delineating cities (Dijkstra et al., 2021).[2] GHS-SMOD categorizes grid cells into one of eight different classifications: urban centre, dense urban cluster, semi-dense urban cluster, suburban or peri-urban, rural cluster, low density rural, very low density rural, and water. Urban centre and urban cluster grid cells can be converted into polygons, representing the boundaries of urban entities around the world with associated attributes such as name (or names) and associated population size.

In 2015, JRC used GHS-SMOD to produce the Urban Centre Database (UCDB). This database attempted to label all of the urban centres captured by SMOD in order to serve as a city boundaries data set computed with a standardized approach (Florczyk et al., 2019b). The UCDB provides boundaries, names, and key statistics for more than 10,000 urban centres

[1] https://ghsl.jrc.ec.europa.eu/documents/GHSL_Data_Package_2019.pdf
[2] https://ghsl.jrc.ec.europa.eu/degurba.php

around the world (Florczyk et al., 2019b).[3] In collaboration with JRC, CIESIN aimed to expand upon and validate the UCDB naming conventions. The GUPPDv1 data set provides settlement names for urban centres, dense urban clusters, and semi-dense urban clusters. These lower level clusters often represent smaller urban settlements like towns, bringing the total number of urban settlements covered to 123,058. GUPPDv1 provides names, population (for the period 1975-2030 in five-year increments), area, and boundaries for all polygons included in the data set. The polygons were also converted to points, representing centroids of the associated urban settlement extents, with the same attributes. The methodology used to create GUPPDv1 shares commonalities with the methodology used to produce the UCDB. However, modifications such as the use of gridded population data and name transliteration were implemented in order to improve upon both the usefulness and accuracy of the resulting data set. In order to support NASA's mission towards open science, all of the input data sets used in the GUPPDv1 methodology are open data. For example, settlement names were derived from one of two open data sources: GeoNames or OpenStreetMap (OSM). This data set is aligned with the goals of the GEO Human Planet Initiative and supports Sustainable Development Goal (SDG) monitoring, the New Urban Agenda, and the Sendai Framework for Disaster Risk Reduction.

II. Data and Methodology

This data set was produced in four steps, and input data and methods are described in each step. In the first step, the spatial extent of urban areas was defined and produced a polygon (vector) data set. In the second and third steps, names of settlements were added (step 2) and population for the 1975-2030 period in five-year increments (step 3). In a fourth step, centroids for the urban settlements were derived, retaining the attributes added in steps 2 and 3. These are each, described in turn.

II.1 Urban Area Definition

A first step in the production of this data set was to define the extent of the urban areas. The baseline urban area data set used for this product is the GHS-SMOD 2019 Release, which is available from the NASA Socioeconomic Data and Applications Center (JRC and CIESIN, 2021). This is a version of the GHS-SMOD produced by Pesaresi et al., (2019) that is reprojected from Mollweide Equal Area into the Geographic Coordinate System (WGS84) at 9 arc-second and 30 arc-second horizontal resolutions in order to support integration with a variety of global raster data sets. The 30 arc-second version was reprojected. GHS-SMOD uses a combination of GHS-POP (a population grid) and a GHS-BUILT (a built-up area grid) to classify the world's land surface into seven categories, from urban centres grid cells (the most densely populated urban areas) to very low density rural grid cells.[4] In this way, the degree of urbanization classifies the rural-urban continuum rather than a simple binary classification of urban/rural. The rasters representing urban centres (code = 30), dense urban

---

[3] https://ghsl.jrc.ec.europa.eu/ucdb2018Overview.php

[4] For a basic description of the GHS-SMOD classes and overall classification system, visit https://ghsl.jrc.ec.europa.eu/degurbaDefinitions.php. For more details, see European Commission (2023).

clusters (code = 23), and semi-dense urban clusters (code = 22) were previously transformed into polygons by JRC and sent to CIESIN.

II.2 Adding Names to Polygons

Given that the settlement polygons are derived from a combination of remote sensing derived built-up area mapping and population density, naming the settlements is particularly challenging. This is because settlements do not conform to administrative boundaries of cities and towns. For larger agglomerations, what GHS-SMOD terms "urban centres", the challenge is magnified, because these cities often include multiple cities and towns. An example can be found in Figure 1, which shows the urban centre in red for the San Francisco, California region. This urban centre includes San Francisco, Oakland, San Jose, Fremont, and many smaller cities such as San Mateo, Redwood City, and Mountain View. The challenge is to find the name that represents the settlement with the largest population, or which might intuitively be considered the most important (in this case San Francisco historically is the most important of the cities in the agglomeration).

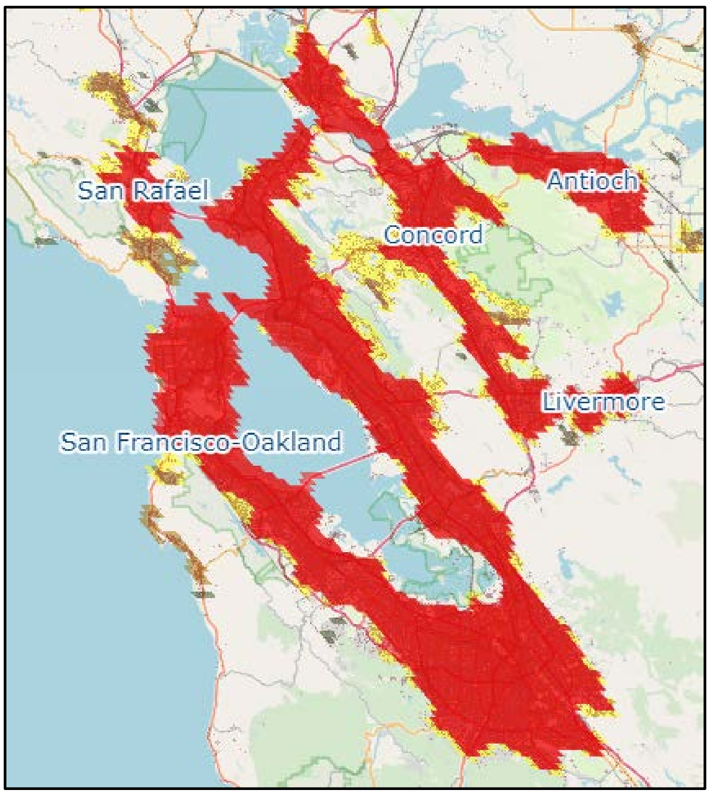

Figure 1. San Francisco urban agglomeration according to GHS-SMOD

To add settlement names to polygons, two data sources, OpenStreetMap (OSM) and GeoNames, were used. For every polygon, an attempt was made to include the name of the largest city or town, along with a concatenated set of names that may be associated with urban centres.

All spatial data were projected to the WGS84 Geographic Coordinate System before being used in any processing. The implemented methodology was automated using a series of

Python scripts (Python Software Foundation, 2022a; Python Software Foundation, 2022b), and Esri ArcPy site package modules (Esri, 2022). All input data sets used in the creation of the GUPPDv1 data set collection are considered open data.

Data from OpenStreetMap[5] (OSM) and GeoNames[6] were collected and used to assign names to polygons in GUPPDv1 (OSM Contributors, 2022). OSM data were downloaded in bulk by key from https://download.osmdata.xyz on 31 March 2022 (note that this web site is no longer operational). The OSM key GeoPackages downloaded included Place, Addresses, and Railways. All three of these GeoPackages were previously assessed and believed to contain viable candidate points for labeling settlement polygons. GeoNames data were downloaded per country and then merged into a global data set.

Preprocessing
Using national boundaries provided by the Global Administrative Areas (GADM) database v3.6[7], OSM and GeoNames points were extracted per country. The Select Layer By Location ArcGIS Pro tool was used to extract points that intersected with the given national boundary. Points were extracted from the OSM Place, Addresses, and Railways GeoPackages. GeoNames points were classified into two hierarchical classes. “Geonames1” included points tagged PPL, PPLA, PPLA2, PPLA3, PPLA4, PPLA5, PPLC, PPLF, PPLL, PPLR, PPLS, and PPLX. All of these tags pertain to populated places and are more conventional tags for a settlement. “Geonames2” included points tagged LCTY, ADM1, ADM2, ADM3, ADM4, ADM5, TERR, ZN, ADMD, AREA, CMP, CMPRF, RSTN, and MSSN. These tags are less conventional but still may provide viable labels for a settlement, so they were classified as a secondary tier. OSM and GeoNames points were then merged into national candidate points feature classes for each country and stored in individual file Geodatabases labeled by ISO code. The Add XY Coordinates ArcGIS Pro tool was used to add latitude and longitude coordinates for each point in a country’s candidate point feature class. In addition, the Extract Values to Points ArcGIS Pro Spatial Analyst tool was used to assign population values to each candidate point based on the 1 km resolution 2015 Global Human Settlement Layer Population Grid (GHS-POP). Lastly, the ArcGIS Pro Delete Identical tool was used to remove any duplicates from each country’s candidate points feature class. Separately, Urban Centre Database (UCDB) polygons were extracted by ISO code and stored in individual file Geodatabases. The process was repeated for SMOD class 23 and 22 polygons. These polygons represented the settlement extents that needed to be labeled.

Labeling
The labeling process consisted of two phases of spatial joins. A primary spatial join was conducted to join all candidate points that intersected with a given settlement. These candidate points were then scored, based on their associated tags (Table 1). The scoring system was devised by first researching the definitions of each tag according to OSM and GeoNames. The tags were then ranked according to associated settlement size and relevancy.

[5] https://www.openstreetmap.org/
[6] https://www.geonames.org/
[7] https://gadm.org/download_country36.html

The resulting feature class was then sorted by unique id and score. The entry per unique id with the lowest score was kept and all other rows were discarded. In an effort to further reduce ties, the remaining rows were sorted by unique id and population value. Rows with the greatest population value were kept and all other rows were discarded. Acknowledging that ties would persist, despite the use of these two filters, the ArcGIS Pro Dissolve tool was used and remaining polygons with multiple candidates had their names concatenated into one.

Next, polygons that remained unlabeled were extracted and processed through a second spatial join. This spatial join considered candidate points that were 1 km outside of the target polygon. The resulting feature class was then scored, filtered, and dissolved. The results from the primary and secondary spatial joins were merged into a completed feature class. Results with concatenated names were alphabetized and extraneous fields were removed from the completed feature class. Completed national feature classes were merged into a global data set. An almost identical procedure was used for labeling SMOD class 30, 23, and 22 polygons. The only notable difference is that SMOD class 23 and 22 polygons were clipped to Global Administrative Areas (GADM) database national boundaries instead of extracted using the Select tool. This in turn generated polygons with identical unique ids that were split along country boundaries. Post processing included removing duplicates within concatenated labels and generating new unique ids.

The settlement extent areas in square kilometers is included as an additional field in the database.

Validation

Manual and automated validation was conducted – described below – of the resulting labeled SMOD class 30, 23 and 22 settlement names. Separately, the Joint Research Centre (JRC) named these same classes using a different methodology, though only the class 30 settlement names are publicly available in the UCDB (Florczyk et al., 2019b). The results generated by CIESIN's methodologies were compared to those produced by JRC. Across each SMOD class, CIESIN's methodology was able to label more polygons (Table 2). For settlements that were successfully labeled by JRC and CIESIN, a majority of these labels matched (Table 2).

Table 1. Scoring Schema

Tag	Score
City	0.1
Town	0.2
Village	0.3
Hamlet	0.4
Suburb	0.5
Neighbourhood	0.6

Locality	0.7
Geonames1	0.8
Geonames2	0.9
Address	0.91
Railway	0.95
Named (but no tag)	0.99
Null	99

Table 2. Percent of Polygons Named by JRC vs. CIESIN Methodology

	SMOD 30	**SMOD 23**	**SMOD 22**
JRC	93%	71%	85%
CIESIN	97%	84%	96%
Matched	74%	67%	71%

For the manual validation, the ArcGIS Pro Subset Features tool was used to generate random samples of the SMOD class 30, 23 and 22 labeled feature classes. Each sample represented 1% of the total results. The accuracy of the label obtained by JRC versus CIESIN was assessed by comparing it to the Esri Topographic basemap. Settlements missing labels or missing a comparable reference on the Esri basemap from the comparison results (Table 3) were excluded. The percent of correctly labeled settlements as the total number of matches divided by the sample size minus null and no reference results were tabulated.

Table 3. Manual Validation

	SMOD 30	**SMOD 23**	**SMOD 22**
JRC	79% (n = 116)	62% (n = 398)	82% (n = 115)
CIESIN	74% (n = 120)	66% (n = 461)	87% (n = 119)

In addition to manual validation, reverse geocoding was conducted in order to further assess the quality of the GUPPDv1 data sets. We used the ArcGIS Pro Subset Features tool to generate new random samples of the SMOD class 30, 23 and 22 labeled feature classes. Each sample represented 5% of the total results. The Wikipedia python library was used to perform the reverse geocoding (i.e. *wikipedia.geosearch* function). Then the *fuzzwuzzy* python package was used to parse through the results and identify matches. All results generated were reviewed and adjusted in Excel. The percent of correctly labeled settlements was tabulated as the total number of matches divided by the sample size minus null results (Table 4.).

Table 4. Reverse Geocoding

	SMOD 30	**SMOD 23**	**SMOD 22**
JRC	72% (n = 350)	79% (n = 1157)	81% (n = 308)
CIESIN	70% (n = 359)	76% (n = 1213)	75% (n = 312)

In the last step, country names and ISO3 codes were added as an attribute to each settlement polygon. For this, the Global Administrative Areas (GADM) database v3.6 was used and spatially joined the country name to the polygon. Polygons with a total 2020 population greater than one million were manually reviewed and had their names adjusted (if necessary) based on visual comparison to the Esri Topographic basemap.

II.3 Adding Population Time Series

To add the population data to the settlement polygons, the settlement polygons were overlaid on a time series population grid from the GHS-POP R2019 Release (Freire et al., 2016), and zonal statistics were produced in ArcGIS Pro 3.1.2 that summed the population in each polygon for the following years using GHS-POP R2019: 1975, 1990, 2000 and 2015 (labeled as GHSPOP_R19_YEAR). These reflect populations within urban definitions provided by GHS-SMOD R2019 benchmarked to circa 2015, so to be clear, the urban extent does not vary over the time periods for which the population data were summarized. In order to extend the time series, zonal stats were produced using the same methods but drawing the population data from GHS-POP R2023 (Schiavina et al., 2023) for the years 1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010, 2015, 2020, 2025, and 2030 (labeled as GHSPOP_R23_YEAR). Note that there are limitations to using GHS-POP R2023 release data in conjunction with GHS-SMOD R2019. For GHS-SMOD R2019, the delineation is based on a different set of population and built up area layers compared to GHS-POP R2023. i.e. spatial units are defined based on a set of inputs - both remote sensing and population - that are not the same as R2019. This means the urban delineations used are not fully consistent with those of GHS-POP R2023. Please note that GHS-POP R2019 data used to develop GUPPDv1 was downloaded at 9 arc-second resolution, whereas GHS-POP R2023 data was downloaded at 3 arc-second resolution.

II.4 Creating the Points Data Set

In order to create a points version of the data set, the centroids were derived for the urban settlements polygons using the ArcGIS Pro Features to Points tool. The option to ensure that points fall inside the derived polygon was selected.

II.5 Codebook

The following table includes the codebook for the fields in the GUPPDv1 data set:

Field Name	Data Type	Description
OBJECTID	Object ID	Object ID
Shape	Geometry	Shape Type
SMOD_LEVEL	Text	GHS-SMOD Class (22, 23, or 30)
SMOD_ID	Text	Unique ID
ISO	Text	Three-Letter Country Code
CNTRY_NAME	Text	Country Name
JRC_NAME_MAIN	Text	JRC Assigned Name
JRC_NAME_LIST	Text	JRC Assigned List of Names
CIESIN_NAME_ADJ	Text	CIESIN Adjusted Names (only applies to polygons with a 2020 total population greater than five million)
CIESIN_NAME	Text	CIESIN Assigned Name
CIESIN_NAME_TL	Text	CIESIN Assigned Name Transliterated
AREA_SQKM	Double	Total Area in Kilometers Squared
AREA_SQM	Double	Total Area in Meters Squared
GHSPOP_R19_1975	Float	Total R2019 GHS-POP 1975
GHSPOP_R19_1990	Float	Total R2019 GHS-POP 1990
GHSPOP_R19_2000	Float	Total R2019 GHS-POP 2000
GHSPOP_R19_2015	Float	Total R2019 GHS-POP 2015
GHSPOP_R23_1975	Float	Total R2023 GHS-POP 1975
GHSPOP_R23_1980	Float	Total R2023 GHS-POP 1980
GHSPOP_R23_1985	Float	Total R2023 GHS-POP 1985
GHSPOP_R23_1990	Float	Total R2023 GHS-POP 1990
GHSPOP_R23_1995	Float	Total R2023 GHS-POP 1995
GHSPOP_R23_2000	Float	Total R2023 GHS-POP 2000
GHSPOP_R23_2005	Float	Total R2023 GHS-POP 2005
GHSPOP_R23_2010	Float	Total R2023 GHS-POP 2010
GHSPOP_R23_2015	Float	Total R2023 GHS-POP 2015
GHSPOP_R23_2020	Float	Total R2023 GHS-POP 2020
GHSPOP_R23_2025	Float	Total R2023 GHS-POP 2025
GHSPOP_R23_2030	Float	Total R2023 GHS-POP 2030
Shape_Length	Double	Shape Length
Shape_Area	Double	Shape Area

III. Data Set Description(s)

The Global Urban Polygons and Points Dataset (GUPPD), Version 1 is a global data set of 123,034 urban settlements with place names and population for the years 1975-2030 in five-year increments.

Data set web page:
SEDAC URL: https://sedac.ciesin.columbia.edu/data/set/urbanspatial-guppd-v1
Permanent URL: https://doi.org/10.7927/brq1-xc29

Data set format:
The data are available in Esri File Geodatabase (GDB) and GeoPackage (GPKG) formats as either polygons or points. Each downloadable is a compressed zip file, containing: 1) GDB or GPKG file, and 2) PDF documentation.

Data set downloads:
urbanspatial-guppd-v1-gdb.zip
urbanspatial-guppd-v1-gpkg.zip

IV. How to Use the Data

The polygon data can be used to generate zonal statistics using a variety of gridded data products. Points can be buffered based on population size. The tabular data underlying this vector data set can also be used for statistical analyses.

V. Potential Use Cases

This data set can be used to delineate urban areas, to mask out urban areas, and to identify the population trends over time for various urban areas. This vector data set of urban settlements with names and time series population can be used for data visualization and a variety of analytical purposes. See, for example, the use of the UCDB by Tuholske et al., (2021) to describe trends in extreme heat event frequency within urban areas globally.

VI. Limitations

This data set carries over the limitations of the GHSL products as described in their documentation, for example JRC and CIESIN (2021). It is important to recognize that the population figures for the period 2015-2030 are projections from 2010 round census figures, and thus are modeled and do not represent the observed population. Furthermore, the population figures in earlier years represent the populations in areas that are defined as urban areas as of circa 2015, and may not have been large towns, cities, or part of agglomerations in those years. This is the so-called fixed boundary approach, where the population layer varies over time but the urban definition remains static (Melchiorri et al., 2024).

VII. Acknowledgments

Funding for the development and dissemination of this data set was provided under the U.S. National Aeronautics and Space Administration (NASA) contract 80GSFC23CA001for the continued operation of the Socioeconomic Data and Applications Center (SEDAC), which is operated by the Center for International Earth Science Information Network (CIESIN) of Columbia University. Additionally, CIESIN wishes to acknowledge the collaboration with the European Commission's Joint Research Centre.

VIII. Disclaimer

CIESIN follows procedures designed to ensure that data disseminated by CIESIN are of reasonable quality. If, despite these procedures, users encounter apparent errors or misstatements in the data, they should contact SEDAC User Services at ciesin.info@ciesin.columbia.edu. Neither CIESIN nor NASA verifies or guarantees the accuracy, reliability, or completeness of any data provided. CIESIN provides this data without warranty of any kind whatsoever, either expressed or implied. CIESIN shall not be liable for incidental, consequential, or special damages arising out of the use of any data provided by CIESIN.

IX. Use Constraints

X. Recommended Citation(s)

Data set(s):

Center for International Earth Science Information Network (CIESIN), Columbia University, and Joint Research Centre (JRC), European Commission. 2024. Global Urban Polygons and Points Dataset (GUPPD), Version 1. Palisades, New York: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/brq1-xc29. Accessed DAY MONTH YEAR.

XI. Source Code

No source code is provided.

XII. References

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

Center for International Earth Science Information Network (CIESIN), Columbia University, CUNY Institute for Demographic Research (CIDR), City University of New York, International Food Policy Research Institute (IFPRI), The World Bank, and Centro Internacional de Agricultura Tropical (CIAT). 2021. Global Rural-Urban Mapping Project, Version 1 (GRUMPv1): Urban Extent Polygons, Revision 02. Palisades, New York: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/np6p-qe61.

Dijkstra, L., A. J. Florczyk, S. Freire, T. Kemper, M. Melchiorri, M. Pesaresi, and M. Schiavina. 2021. Applying the degree of urbanisation to the globe: A new harmonised definition reveals a different picture of global urbanisation. *Journal of Urban Economics*, 125, 103312. https://doi.org/10.1016/j.jue.2020.103312.

Esri. 2022. ArcGIS Pro: Release 3.0. Redlands, CA: Environmental Systems Research Institute. https://www.esri.com.

European Commission. 2023. GHSL Data Package. Publications Office of the European Union, Luxembourg, 2023, JRC133256. https://doi.org/10.2760/098587.

Florczyk, A., C. Corbane, D. Ehrlich, S. Freire, T. Kemper, L. Maffenini, M. Melchiorri, M. Pesaresi, P. Politis, M. Schiavina, F. Sabo, and L. Zanchetta. 2019a. GHSL Data Package 2019, Publications Office of the European Union, Luxembourg. https://doi.org/10.2760/0726.

Florczyk A., C. Corbane, M. Schiavina, M. Pesaresi, L. Maffenini, M. Melchiorri, P. Politis, F. Sabo, S. Freire, D. Ehrlich, T. Kemper, P. Tommasi, D. Airaghi, and L. Zanchetta. 2019b. GHS Urban Centre Database 2015, multitemporal and multidimensional attributes, R2019A. European Commission, Joint Research Centre (JRC). https://doi.org/10.2905/53473144-b88c-44bc-b4a3-4583ed1f547e. https://data.jrc.ec.europa.eu/dataset/53473144-b88c-44bc-b4a3-4583ed1f547e.

Freire, S., K. MacManus, M. Pesaresi, E. Doxsey-Whitefield, and J. Mills. 2016. Development of new open and free multi-temporal global population grids at 250 m resolution. *Proc. of the 19th AGILE Conference on Geographic Information Science*, 250. https://publications.jrc.ec.europa.eu/repository/handle/JRC100523. https://www.researchgate.net/publication/304625387_Development_of_new_open_and_free_multi-temporal_global_population_grids_at_250_m_resolution.

Joint Research Centre (JRC), European Commission, and Center for International Earth Science Information Network (CIESIN), Columbia University. 2021. Global Human Settlement Layer: Population and Built-Up Estimates, and Degree of Urbanization Settlement Model Grid. Palisades, New York: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/h4154f0w.

Joint Research Centre (JRC), European Commission, and Center for International Earth Science Information Network (CIESIN), Columbia University. 2021. Documentation for the Global Human Settlement Layers (GHSL): Population and Built-Up Estimates, and Degree of Urbanization Settlement Model Grid. Palisades, New York: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/tg7r-n260.

Melchiorri, M., S. Freire, M. Schiavina, A. Florczyk, C. Corbane, L. Maffenini, and T. Kemper. 2024. The Multi-temporal and Multi-dimensional Global Urban Centre Database to Delineate and Analyse World Cities. *Scientific Data*, 11(1), 82. https://doi.org/10.1038/s41597-023-02691-1.

OpenStreetMap (OSM) Contributors. 2022. "OpenStreetMap." https://www.openstreetmap.org (accessed 2022).

Pesaresi, M., A. Florczyk, M. Schiavina, M. Melchiorri, and L. Maffenini. 2019. GHS settlement grid, updated and refined REGIO model 2014 in application to GHS-BUILT R2018A and GHS-POP R2019A, multitemporal (1975-1990-2000-2015), R2019A. European Commission, Joint Research Centre (JRC). https://doi.org/10.2905/42E8BE89-54FF-464E-BE7B-BF9E64DA5218. http://data.europa.eu/89h/42e8be89-54ff-464e-be7b-bf9e64da5218.

Python Software Foundation 2022a. IDLE (Python's Integrated DeveLopment Environment) version 3.9.1.1. https://www.python.org.

Python Software Foundation 2022b. Python programming language, version 3.9.1.1. https://www.python.org.

Schiavina, M., S. Freire, C. Alessandra, and K. MacManus. 2023. GHS-POP R2023A - GHS population grid multitemporal (1975-2030). European Commission, Joint Research Centre (JRC). https://doi.org/10.2905/2FF68A52-5B5B-4A22-8F40-C41DA8332CFE. https://data.europa.eu/89h/2ff68a52-5b5b-4a22-8f40-c41da8332cfe.

Tuholske, C., K. Caylor, C. Funk, A. Verdin, S. Sweeney, K. Grace, P. Peterson, and T. Evans. 2021. Global High Resolution Daily Extreme Urban Heat Exposure (UHE-Daily), 1983-2016. Palisades, New York: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/fq7g-ny13.

XIII. Documentation Copyright and License

Appendix 1. Data Revision History

No revisions have been made to this data set.

Appendix 2. Contributing Authors & Documentation Revision History

Revision Date	ORCID	Contributors	Revisions
July 25, 2024	0000-0002-8875-4864	James Gibson, Alex de Sherbinin	This document is the 1st instance of documentation.