

Documentation for the  
Global Rural-Urban Mapping Project, Version 1 (GRUMPv1): Urban  
Extent Polygons, Revision 01

**July 2017**

Center for International Earth Science Information Network (CIESIN),  
Columbia University

**Abstract**

This document outlines the basic methodology and data sets used to construct the GRUMPv1 Urban Extent Polygons, v1.01 (1995), along with use cases, limitations, and use constraints.

**Data set citation:**

Center for International Earth Science Information Network (CIESIN), Columbia University, CUNY Institute for Demographic Research (CIDR), International Food Policy Research Institute (IFPRI), The World Bank, and Centro Internacional de Agricultura Tropical (CIAT). 2017. Global Rural-Urban Mapping Project, Version 1 (GRUMPv1): Urban Extent Polygons, Revision 01. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). <https://doi.org/10.7927/H4Z31WKF>. Accessed DAY MONTH YEAR.

This is a revision of the Global Rural-Urban Mapping Project, Version 1 (GRUMPv1): Urban Extent Polygons (Internal Working Version) which has not been released.

**Suggested citation for this document:**

Center for International Earth Science Information Network (CIESIN), Columbia University. 2017. Documentation for the Global Rural-Urban Mapping Project, Version 1 (GRUMPv1): Urban Extent Polygons, Revision 01. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). <https://doi.org/10.7927/H4C53HSR>. 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](mailto:ciesin.info@ciesin.columbia.edu)

## Contents

I.	Introduction.....	2
II.	Data and Methodology.....	3
III.	Data Set Description(s).....	5
IV.	How to Use the Data.....	6
V.	Potential Use Cases.....	6
VI.	Limitations.....	6
VII.	Acknowledgments.....	6
VIII.	Disclaimer.....	7
IX.	Use Constraints.....	7
X.	Recommended Citation(s).....	7
XI.	Source Code.....	7
XII.	References.....	8
XIII.	Documentation Copyright and License.....	9
	Appendix 1. Data Revision History.....	10
	Appendix 2. Contributing Authors & Documentation Revision History.....	10

## I. Introduction

The primary output of the Global Rural Urban Mapping Project, Version 1 (GRUMPv1) are a series of grids representing estimated population counts and density for the years 1990, 1995, and 2000 at 30 arc-second (1km) spatial resolution. GRUMPv1 represents a spatial reallocation of the Gridded Population of the World, Version 3 (GPWv3) population distribution in a way that moved populations from rural to urban portions of administrative units. An important secondary output of GRUMPv1 was an urban extents grid that was used in the reallocation process. To define the spatial extent of urban areas, the authors used a combination of nighttime lights from the Defense Meteorological Satellite Program Optical Line Scan (DMSP-OLS) instrument, and, where lights were too dim (e.g., for smaller towns or for regions with lower lighting levels), buffered settlement points. Buffering, in this case, is a geospatial operation in which a circle of a certain radius is drawn around a point. The radius of the buffer was drawn in proportion to the estimated population size associated with a settlement point.

This Revision 01 (i.e. v1.01) includes new settlements and represents the first time that SEDAC has released polygons (in Esri shapefile format) with the settlement name (or name of the largest city in the case of multi-city agglomerations). The shapefile consists of polygons defined by the extent of the nighttime lights and approximated urban extents (circles) based on buffered settlement points. This version includes some new urban extents identified using the UN extended city population list or by correcting the georeference of some previously mis-referenced settlements (see separate documentation for Global Rural-Urban Mapping Project, Version 1 (GRUMPv1): Settlement Points, Revision 01 for the data and methods). The Urban Extent Polygons (Internal Working Version) were produced by the Columbia University Center for International Earth Science Information Network (CIESIN) in collaboration with the International Food Policy Research Institute (IFPRI), The World Bank, and Centro Internacional de Agricultura Tropical (CIAT). This revision was produced by the Columbia University Center for International Earth Science Information Network (CIESIN) in Collaboration with CUNY

Institute for Demographic Research (CIDR), with funding from the U.S. National Institutes of Health (NIH).

## II. Data and Methodology

The basic methodology for GRUMPv1 is outlined in the introduction. For a full description of data and methodologies, users are encouraged to read Balk *et al.* (2006).

### Input data

The shapefile consists of polygons based on nighttime lights and approximated urban extents (circles) from buffered settlement points. The input data sets used in producing the urban extents data set are:

- a) population counts (persons)
- b) settlement points
- c) nighttime lights
- d) buffered settlement points

For population counts, city population data was collected from multiple sources. These include national censuses, the City Population database (undated), and World Gazetteer (web site no longer exists).

For settlement points, we obtained spatial locations (latitude and longitude) for cities from multiple sources, including the Digital Chart of the World, City Population database, World Gazetteer, and Falling Rain (undated).

For nighttime lights, the most commonly used moderate-resolution data to indicate urban areas at a global scale are the nighttime lights data derived from DMSP-OLS (Elvidge *et al.* 1997a, 1997b). These data indicate the stable sources of light produced by human nocturnal illumination sources (i.e., street lights, domestic lighting, industrial lights, and commercial lighting and signage) and permanent fires (such as gas flares). Though several nighttime lights data sets exist, only the 1994/95 stable city lights data product was cleaned to remove oil and gas flares and other lighting sources not associated with human settlements. This data set is seen as a proxy for urban areas. Nevertheless, it may be subject to bias associated with levels of economic activity—i.e. poorer, dimly-lit cities may exist but their extent may be under-represented spatially as compared to wealthier well-lit cities of equal geographic size. There may be cultural or policy factors that also affect the intensity and extent of lighting.

Finally, for settlements that were not associated with nighttime lights, areal extents for urban areas were estimated through the buffering of settlement points based on the observed relationship between population size and areal extents for the points with known parameters.

## **Methods**

The settlement points data set contains estimates of populated places: names, population, and geographic coordinates, collected from national statistical offices and publicly available databases. In many instances, three different data sources were required for a given place (source information for each datum is retained in the disseminated GRUMPv1 Settlement Points data set). When it was evident how the population estimate associated with each point was classified – city proper, agglomeration, etc. – such classifications are noted in the data set.

Next, the settlement points were spatially joined with the nighttime lights generated urban extents and the population assigned, or summed, if more than one point was found. The name of the most populous place within the buffer was assigned to the urban extent polygon.

For points without polygons (i.e., insufficiently lit to be represented by nightlights), buffers were created. Areal extents for urban areas are estimated based on a relationship between population size and areal extents for the points with known parameters. This relationship is derived from a logarithmic regression that predicts the expected geographic size of a place, given its population size, minimum number of observations or by combining data in regional groupings supplied by the UN Statistics Division. Based on these estimated area values, a circular urban area, centered on the known points' spatial location was created. The resulting polygons were then added to the existing lights-based ones to create a complete urban extent coverage.

The resulting data set includes some urban extents that are not associated with any known settlement points, and thus don't have a name or a population count. In these cases, we were not able to locate an urban settlement within the urban extent.

This revision includes some new urban extents identified using the UN extended city population list or by correcting the georeference of some previously miss-referenced settlements. However, the validation is only applied to a selected number of urban extents, therefore it is possible that some cities in the data set may have incorrect spatial location.

### III. Data Set Description(s)

#### Data set description:

The data set consists of polygons representing urban extents with population figures for 1990, 1995, and 2000. The population associated with each urban extent was estimated for 1990, 1995, and 2000. The population of added settlements in this revision is included in the POP\_V1\_01 field, and the year of respective population is included in the YEAR\_V1\_01 column. These settlements do not include populations for the reference years 1990, 1995 and 2000.

#### Data set web page:

<http://sedac.ciesin.columbia.edu/data/set/grump-v1-urban-ext-polygons-rev01>

#### Data set format:

The data are available in shapefile and csv formats as downloadable zip files. When exporting from shapefile to table (csv), special characters in the NAME field are represented as symbols. However, the SCHNM field includes the concatenated settlement name without the symbols.

#### Data set downloads:

grump-v1-urban-ext-polygons-rev01-shp.zip  
 grump-v1-urban-ext-polygons-rev01-csv.zip

#### Codebook:

Field Name	Field Description
OBJECTID	ID number
URBID	Code of the urban extent polygons (not unique)
LIGHTDCW	Dummy variable, code 1 for polygons with source DCW
ES90POP	Estimated 1990 population, calculated based on the sum of the 1990 population for settlements (points) located within or less than 3 meters than the urban extent.
ES95POP	Estimated 1995 population, calculated based on the sum of the 1995 population for settlements (points) located within or less than 3 meters than the urban extent.
ES00POP	Estimated 2000 population, calculated based on the sum of the 2000 population for settlements (points) located within or less than 3 meters than the urban extent.
PCOUNT	Number of points within the urban extent polygon
SCHNM	Name of the Urban Extent, caps and concatenated. If the urban extents polygons include more than one settlement, the name of the urban extent is the name of the settlement with the highest population. The name of the urban extent is missing if no settlement was identified within or less than 3 meters than the urban extent.
NAME	Name of the Urban Extent. If the urban extents polygons include more than one settlement, the name of the urban extent is the name of the settlement with the highest population. The name of the urban extent is missing if no settlement was identified within or less than 3 meters than the urban extent.
SQKM_FINAL	Area of the urban extent polygon
ISO3	3 letter country code
ISOURBID	Unique code of the urban extent polygons. It is created by concatenation of country ISO code and URBID.

REMOVED_PO	Settlements removed from the urban extent polygons based on work done between GRUMPv1 and GRUMPv1 Revision 01
ADDED_POIN	Settlements added to the urban extent polygons based on work done between GRUMPv1 and GRUMPv1 Revision 01
YEAR_V1_01	Year for the collected population for the added settlements based on work done between GRUMPv1 and GRUMPv1 Revision 01
POP_V1_01	Population collected for the added point based on work done between GRUMPv1 and GRUMPv1 Revision 01
Unsdcode	UNSD country numerical code
Countryeng	Country name
Continent	Continent name

#### **IV. How to Use the Data**

The data are intended to be used in any geographic information system (GIS) or statistics software. The data set is global in coverage, and can be subsetted based on the country name or other characteristics included in the Codebook (shapefile attribute table).

#### **V. Potential Use Cases**

Urban extents are an important framework data layer. This urban extents data set can be used for any spatial operations that require urban extents, such as zonal statistics for urban and non-urban rural areas, or buffering the urban extents to obtain values inside and outside the urban area (as is done in SEDAC's Global Urban Heat Island (UHI) Data Set, 2013).

#### **VI. Limitations**

The accuracy of the geographic location of the cities depends on the quality of latitude and longitude data available from the sources. At the time of data collection (i.e. 2005 – 2010), the georeferencing sources quality was low. This revision includes corrected georeferencing for a subset of cities.

#### **VII. Acknowledgments**

GRUMPv1 was conceived by a group of like-minded researchers attempting to better understand the distribution of human population with a particular interest in being able to disentangle urban settlements from more diffuse and sparse patterns of rural settlement largely for the purpose of being able to study the human-environmental interactions at a global scale. Though there were methodological and data contributions from multiple project partners, GRUMP was primarily implemented by a single organization—the Center for International Earth Science Information Network (CIESIN), at Columbia University.

Funding for the GRUMPv1 data collection was provided by International Food Policy Research Institute (IFPRI), The World Bank, and the Centro Internacional de Agricultura Tropical (CIAT). This revision was funded under a grant from U.S. National Institutes of Health (NIH).

## 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](mailto: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

Users are free to use, copy, distribute, transmit, and adapt the work for commercial and non-commercial purposes, without restriction, as long as clear attribution of the source is provided.

## X. Recommended Citation(s)

### Data set(s):

Center for International Earth Science Information Network (CIESIN), Columbia University, CUNY Institute for Demographic Research (CIDR), International Food Policy Research Institute (IFPRI), The World Bank, and Centro Internacional de Agricultura Tropical (CIAT). 2017. Global Rural-Urban Mapping Project, Version 1 (GRUMPv1): Urban Extent Polygons, Revision 01. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). <https://doi.org/10.7927/H4Z31WKF>. Accessed DAY MONTH YEAR.

### Scientific publication:

D. L. Balk, U. Deichmann, G. Yetman, F. Pozzi, S. I. Hay, and A. Nelson, 2006. "Determining Global Population Distribution: Methods, Applications and Data," in Hay, S. I., Graham, A. J. and Rogers, D. J. (eds), *Global mapping of infectious diseases: methods, examples and emerging applications. Advances in Parasitology*, volume 62. (London: Academic Press), pp. 119-156. Available on-line: PMID: 16647969. [https://doi.org/10.1016/S0065-308X\(05\)62004-0](https://doi.org/10.1016/S0065-308X(05)62004-0).

## XI. Source Code

Not applicable

## XII. References

- Balk, D., Montgomery, M. R., McGranahan, G., Kim, D., Mara, V., Todd, M., Buettner, T., Dorelien, A. 2009. *Mapping Urban Settlements and the Risks of Climate Change in Africa, Asia and South America*. In: Population Dynamics and Climate Change, Guzman, J. M., Martine, G., McGranahan, G., Schensul, D. and Tacoli, C. (eds), pp. 80-102. IIED/UNFPA, New York.  
[https://www.unfpa.org/webdav/site/global/shared/documents/publications/2009/pop\\_dynamics\\_climate\\_change.pdf](https://www.unfpa.org/webdav/site/global/shared/documents/publications/2009/pop_dynamics_climate_change.pdf)
- Balk, D., Montgomery, M. R., McGranahan, G., Todd, M. 2009. *Understanding the Impacts of Climate Change: Linking Satellite and Other Spatial Data with Population Data*. In: Population Dynamics and Climate Change, Guzman, J. M., Martine, G., McGranahan, G., Schensul, D. and Tacoli, C. (eds), pp. 206-214. IIED/UNFPA, New York.  
[https://www.unfpa.org/webdav/site/global/shared/documents/publications/2009/pop\\_dynamics\\_climate\\_change.pdf](https://www.unfpa.org/webdav/site/global/shared/documents/publications/2009/pop_dynamics_climate_change.pdf)
- Bloom, D. E., Canning, D., Fink, G., Khanna, T. & Salyer, P. 2010. *Urban Settlement: Data, Measures, and Trends*. UNU-WIDER Working Paper 12, 23pp.  
[http://www.wider.unu.edu/publications/working-papers/2010/en\\_GB/wp2010-12/](http://www.wider.unu.edu/publications/working-papers/2010/en_GB/wp2010-12/)
- Doll, C. N. H. and Pachauri, S. 2010. Estimating rural populations without access to electricity in developing countries through night-time light satellite imagery. *Energy Policy*, 38(10): 5661-5670. <https://doi.org/10.1016/j.enpol.2010.05.014>
- Ferré, C., Ferreira, F. H. G. and Lanjouw, P. 2012. Is there a metropolitan bias? The relationship between poverty and city size in a selection of developing countries. *The World Bank Economic Review* 26(3): 351-382. <https://doi.org/10.1093/wber/lhs007>
- Linard, C., Gilbert, M., Snow, R. W., Noor, A. M. and Tatem, A. J. 2012. Population distribution, settlement patterns and accessibility across Africa in 2010. *PLoS ONE* 7(2): e31743. <https://doi.org/10.1371/journal.pone.0031743>
- McGranahan, G., Balk, D. and Anderson, B. 2006. Low coastal zone settlements. *Tiempo*, (59): 23-26. <http://www.cru.uea.ac.uk/tiempo/portal/archive/pdf/tiempo59low.pdf>
- McGranahan, G., Balk, D. and Anderson, B. 2007. The rising tide: Assessing the risks of climate change and human settlements in low elevation coastal zones. *Environment and Urbanization*, 19(1): 17-37. <http://eau.sagepub.com/cgi/content/abstract/19/1/17>
- Montgomery, M. and Balk, D. 2011. *The Urban Transition in Developing Countries: Demography Meets Geography*. Chapter 6 in: Global Urbanization, E. L. Birch & S. M. Wachter (Eds.), University of Pennsylvania, pp. 89-106.



Montgomery, M. R. 2008. The Urban transformation of the developing world. *Science*, 319(5864): 761-764. <https://doi.org/10.1126/science.1153012>

Montgomery, M. R. 2008. *The Demography of the Urban Transition: What We Know and Don't Know*. In: *The New Global Frontier: Urbanization, Poverty and Environment in the 21st Century*, G. Martine, G. McGranahan, M. R. Montgomery, and R. Fernandez-Castilla, Eds., EarthScan, pp. 17-36.

Tatem, A., Adamo, S., Bharti, N., Burgert, C., Castro, M., Dorelien, A., Fink, G., Linard, C., John, M., Montana, L., Montgomery, M., Nelson, A., Noor, A., Pindolia, D., Yetman, G. and Balk, D. 2012. Mapping populations at risk: improving spatial demographic data for infectious disease modeling and metric derivation. *Population Health Metrics* 10(8): 14pp. <https://doi.org/10.1186/1478-7954-10-8>

### **XIII. Documentation Copyright and License**

Copyright © 2017. The Trustees of Columbia University in the City of New York. This document is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>).

## Appendix 1. Data Revision History

An Internal Working Version of the Global Rural-Urban Mapping Project, Version 1 (GRUMPv1): Urban Extent Polygons was generated but not released. This internal urban extent version included 75,414 urban extents.

The Global Rural-Urban Mapping Project, Version 1 (GRUMPv1): Urban Extent Polygons, Revision 01 represents the first time that SEDAC has released polygons (in Esri shapefile format) with the settlement names (or names of the largest cities in the case of multi-city agglomerations). This version has a total of 75,445 urban extents.

It was published in 2017 and includes:

- new urban extents identified by name using the UN extended city population list or
- new urban extents identified by name by correcting the georeference of some previously miss-referenced settlements.

The number of urban extents (with respective countries) added to the data set are the following:

- Africa: Benin (23), The Democratic Republic of Congo (2), Liberia (2), Mauritania (7), Malawi (2)
- Asia: Bangladesh (4), China (1), Maldives (3), Myanmar (1), Nepal (3)
- Europe: Austria (1)
- North America: Antigua and Barbuda (1), The Bahamas (7), Cuba (28), St. Lucia (1), Martinique (1)
- South America: Argentina (3), Brazil (14), Falkland Islands (1)

Some of the urban extents from the Internal Working Version of the Global Rural-Urban Mapping Project, Version 1 (GRUMPv1): Urban Extent Polygons have been removed because of duplicate or erroneous entries. The number of urban extents (with respective countries) removed from the data set are the following:

- Africa: Liberia (1)
- Asia: Bangladesh (1), China (10), India (42), Iran (1), Sri Lanka (3)
- Europe: Hungary (1), The Netherlands (1)
- North America: Cuba (1), USA (2)
- South America: Peru (11)

## Appendix 2. Contributing Authors & Documentation Revision History

Revision Date	Contributors	Revisions
July 24, 2017	Valentina Mara, Alex de Sherbinin	This Revision 01 document is an update to the March 2017 version.