

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

**March 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 Settlement Points, v1.01 (1990, 1995, 2000), 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): Settlement Points, Revision 01. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). <http://dx.doi.org/10.7927/H4BC3WG1>. Accessed DAY MONTH YEAR.

This is an updated version of the Global Rural-Urban Mapping Project, Version 1 (GRUMPv1): Settlement Points, which was released in 2011. <http://dx.doi.org/10.7927/H4M906KR>.

**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): Settlement Points, Revision 01. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). <http://dx.doi.org/10.7927/H4GX48H5>. 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).....	4
IV.	How to Use the Data.....	6
V.	Potential Use Cases.....	6
VI.	Limitations.....	6
VII.	Acknowledgments.....	6
VIII.	Disclaimer.....	6
IX.	Use Constraints.....	7
X.	Recommended Citation(s).....	7
XI.	Source Code.....	7
XII.	References.....	7
XIII.	Documentation Copyright and License.....	9
	Appendix 1. Data Revision History.....	9
	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. To define the spatial extent of urban areas, GRUMPv1 used a combination of night-time 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 population size associated with a settlement point, and each settlement point in the Settlement Points Version 1 data set has data on the estimated population for the years 1990, 1995 and 2000.

This Settlement Points, Revision 01 data set is an update to Global Rural-Urban Mapping Project, Version 1 (GRUMPv1) Settlement Points used in the original population reallocation. This Revision 01 (i.e. v1.01) includes improved geospatial location (latitude and longitude coordinates) for selected settlements, as well as new georeferenced settlements. The total number of settlements included in Revision 01 is 70,629 as compared to 67,129 for Version 1, a difference of about 3,500 settlements. Note that for these newly added settlements, we did not develop time series population values for 1990, 1995 and 2000. Instead we created two new fields, Pop and Year, in which the population point estimate for the available year is included (the field is coded -999 for missing values). The intent of this data set is to include information on all settlements

globally, but owing to data gaps, it is important to note that this data set represents a non-comprehensive “best available” list of settlement locations and populations. The likelihood that a settlement is included increases with population size, since larger settlements tend to be better documented. So users who wish, for example, to tally the population of all settlements above a certain population size are advised to pick a sufficiently high threshold (e.g., greater than 100,000).

The Version 1 data set was 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 updated version was produced by the Columbia University Center for International Earth Science Information Network (CIESIN) in collaboration with the CUNY Institute for Demographic Research (CIDR) and was funded by the 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 input data sets used in producing the settlement points data set are:

- a) population counts
- b) 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).

### **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 Settlements 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.

Population figures were not necessarily available for all the years included in our time series. If population counts were available for only one time slice, e.g., 1990, population figures for 1995 and 2000 were projected using UN population projections, based on calculated national exponential growth. A similar model was used to backcast population numbers where they were only available for later time periods. If population counts were available for more than one year, the estimated population was calculated using intrinsic growth rates. Population was estimated for more than one target year (1990, 1995, and 2000).

This updated version is the by-product of an effort to identify large settlements (of over 100,000 people) which were missing in GRUMP v1. These are located by searching for urban extents defined by night-time lights that had no associated settlement point(s). In addition, new settlements were added to the data set based on the UN extended settlement database. Some additional settlements were discovered during data processing. The type of settlement discovery is included in field named “Post\_Hoc” (see section II below for a description of variables). We also identified and corrected the coordinates of settlements that were incorrectly georeferenced in Version 1 of the data set. These records are identified in the field named “Intentiona”.

As validation is only applied to a selected number of points, we expect that there may be other cities in the data set that have incorrect spatial location.

### **III. Data Set Description(s)**

#### **Data set description:**

This updated version of the GRUMPv1 Settlement Points data set includes new settlements using the UN extended city population list or by correcting the georeferencing of some previously miss-referenced settlements (see Section II).

Version 1 of the GRUMP Settlement Points included 67,129 points. In Revision 01, there are 70,629 settlement points.

#### **Data set web page:**

<http://sedac.ciesin.columbia.edu/data/set/grump-v1-settlement-points-rev01>

#### **Data set format:**

The data are available in shapefile and csv formats as downloadable zip files.

#### **Data set downloads:**

grump-v1-settlement-points-rev01-shp.zip  
grump-v1-settlement-points-rev01-csv.zip

**Codebook:**

<b>Field Name</b>	<b>Field Description</b>
OBJECTID	ID number
UNSD	UNSD country numerical code
Country	Country name
Latlongid	Settlement ID number
Latitude	The latitude of the point
Longitude	The longitude of the point
Pop	Population collected for the added point based on work done between GRUMPv1 and GRUMPv1 Revision 01
Year	Year for the collected population for the added point (city) based on work done between GRUMPv1 and GRUMPv1 Revision 01
Urborrur	Type of settlement, urban or rural
Schnm	Name of the settlement, caps and concatenated.
Name1	Name of the settlement
Name2	Alternative settlement name
Name3	Alternative settlement name
Forgnm	Foreign spelling of the settlement name
Schadnm	Administrative division in which the settlement is located, caps and concatenated (level one division).
Adnm1	Administrative division in which the settlement is located (level one division)
Adnm2	Administrative division in which the settlement is located (level two division)
Type	Type of the settlement (city, locality, etc.)
Srctyp	Source of settlement type (census, gazetteer)
Coordsrce	Source for coordinates (i.e. latitude and longitude)
Datsrc	Data source
Locndatsrc	Location of the data source (web address)
Notes	Notes regarding the settlement collected
Post_Hoc	Includes the type of settlement discovery for the added points in GRUMPv1 Revision 01
Intentiona	The settlements with corrected georeferencing in GRUMPv1 Revision 01
ISO3	Three letter country code
ES90POP	Estimated 1990 population, calculated using the UN population projections or intrinsic growth rates
ES95POP	Estimated 1995 population, calculated using the UN population projections or intrinsic growth rates
ES00POP	Estimated 2000 population, calculated using the UN population projections or intrinsic growth rates
INSGRUSED	Dummy variable, representing whether intrinsic growth rates was used in calculating the estimates population for 1990, 1995, or 2000.
Continent	Name of the continent

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

The data are released at the global level. Users can subset the data based on the country name or other characteristics included in the Codebook (shapefile attribute table).

## **V. Potential Use Cases**

Settlement points data are important framework data set for mapping urban populations globally, and can be used as centroids for buffers, and to assess road connectivity, among other applications. The data can be subsetted by population size to map settlements of different sizes; for map display, settlement size can also be used for graduated symbols.

## **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 updated version 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 updated version was funded under a grant from the United States 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): Settlement Points, Revision 01. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). <http://dx.doi.org/10.7927/H4BC3WG1>. Accessed DAY MONTH YEAR.

### **Scientific publication:**

Balk, D.L., 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. [http://dx.doi.org/10.1016/S0065-308X\(05\)62004-0](http://dx.doi.org/10.1016/S0065-308X(05)62004-0).

## **XI. Source Code**

Not applicable

## **XII. References**

Angel, S., S. C. Sheppard, and D.L. Civco. 2005. *The Dynamics of Global Urban Expansion*. Washington, D.C.: World Bank Transport and Urban Development Department.

- Balk, D., F. Pozzi, G. Yetman, U. Deichmann, and A. Nelson. 2005. “*The Distribution of People and the Dimension of Place: Methodologies to Improve the Global Estimation of Urban Extents.*” Paper presented at the International Society for Photogrammetry and Remote Sensing Proceedings of the Urban Remote Sensing Conference, March 14-16, Tempe, AZ.
- Balk, D. and G. Yetman. 2004. “*The Global Distribution of Population: Evaluating the gains in resolution refinement.*” Working paper Center for International Earth Science Information Network (CIESIN). [http://sedac.ciesin.columbia.edu/downloads/docs/gpw-v3/gpw3\\_documentation\\_final.pdf](http://sedac.ciesin.columbia.edu/downloads/docs/gpw-v3/gpw3_documentation_final.pdf) (accessed February 1, 2008).
- Chan, K.W. and Y. Hu. 2003. “Urbanization in China in the 1990’s: New Definition, Different Series, and Revised Trends.” *China Review* 3(2): 49-71.
- Chan, K.W. 2007. “Misconceptions and Complexities in the Study of China's Cities: Definitions, Statistics, and Implications.” *Eurasian Geography and Economics* 48(4): 383-412.
- City Population. *undated*. <http://www.citypopulation.de> (accessed June 2015).
- Dobson, J.E., E.A. Bright, P.R. Coleman, R.C. Durfee and B.A. Worley. 2002. “LandScan: A global population database for estimating populations at risk.” *Photogrammetric Engineering and Remote Sensing* 66(7): 849-857.
- Elvidge, C.D., K.E. Baugh, E.A. Kihn, H.W. Kroehl, and E.R. Davis. 1997a. “Mapping City Lights with Nighttime Data from the DMSP Operational Linescan System.” *Photogrammetric Engineering and Remote Sensing* 63(6): 727-734.
- Falling Rain. *undated*. <http://www.fallingrain.com/credits.html> (accessed June 2015).
- Reallocation (GPW-UR). <http://sedac.ciesin.columbia.edu/gpw> (accessed February 1, 2008).
- McGranahan, G., D. Balk, and B. Anderson. 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.
- Montgomery, M. R. 2008. “The Urban Transformation of the Developing World.” *Science* 319(5864): 761-764.
- Montgomery, M. R., Richard Stren, Barney Cohen, and Holly E. Reed (ed.). 2003. *Cities Transformed: Demographic Change and its Implication in the Developing World*. Washington D.C.: National Academy Press.
- Montgomery, M. R. and Deborah Balk. Forthcoming. 2008. “The Urban Transition in Developing Countries: Demography Meets Geography.” *Global Urbanization in*



*the 21<sup>st</sup> Century*, edited by E. Birch and S. Wachter. Philadelphia: University of Pennsylvania Press.

Thomas Brinkhoff: City Population, <http://www.citypopulation.de>

United Nations. 2006. *World Urbanization Prospects: The 2005 Revision*. New York: United Nations.

United Nations. 2008. *Principles and Recommendations for National Censuses of Population and Housing Revision 2*. New York: United Nations Department of Economic and Social Affairs.

Weeks, J. R., A. G. Hill, D. A. Stow, A. Getis, and D. Fugate. 2007. "Can You Spot a Neighborhood from the Air? Defining Neighborhood Structure in Accra, Ghana." *GeoJournal* 69(1-2): 9-22.

### **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 (<http://creativecommons.org/licenses/by/4.0/>).

### **Appendix 1. Data Revision History**

The Global Rural-Urban Mapping Project, Version 1 (GRUMPv1): Settlement Points data set was published in 2011.

Global Rural-Urban Mapping Project, Version 1 (GRUMPv1): Settlement Points, Revision 01 was published in 2017 and includes:

- improved geospatial location (latitude and longitude coordinates) for selected settlements, as well as new georeferenced settlements.
- 70,629 settlements, as compared to 67,129 for Version 1, a difference of 3,500 settlements.
- Countries (additional settlements) to the urban extents:
  - Afghanistan (4), Netherland Antilles (2), United Arab Emirates (1), Argentina (15), Antigua and Barbuda (1), Australia (7), Austria (3), Azerbaijan (13), Belgium (4), Benin (27), Bangladesh (45), Bahrain (1), Bahamas (11), Belarus (3), Brazil (145), Canada (2), Chile (8), China (453), Ivory Coast (1), Congo, Democratic Republic (6), Colombia (6), Comoros (2), Costa Rica (5), Cayman Islands (1), Commonwealth of Dominica (5), Dominican Republic (1), Algeria (3), Ecuador (2), Egypt (11), Ethiopia (2), Falkland Islands (1), France (3), Great Britain (5), Georgia (1), Ghana (3), Guatemala (3), Honduras(9), Indonesia (110), India (321), Iran (37), Iraq (14), Israel (1), Jamaica (4), Jordan (4), Japan

(42), Kazakhstan (17), Kenya (13), Cambodia (1), Saint Kitts and Nevis (2), Korea (11), Kuwait (1), Lebanon (3), Liberia (4), Libya (2), Saint Lucia (1), Sri Lanka (10), Lesotho (1), Macao (1), Morocco (14), Madagascar (1), Mexico (21), Mali(1), Myanmar (14), Northern Mariana Islands (1), Mozambique (4), Mauritania(7), Montserrat (1), Martinique (1), Malawi (6), Malaysia (17), Mayotte (1), Nigeria (39), Nepal (6), Pakistan (142), Panama (1), Peru (2), Philippines (49), Korea, Dem. People's Rep. of (4), Paraguay (3), Russia (2), Saudi Arabia (13), Serbia and Montenegro (1), Sudan (3), Senegal (2), El Salvador (5), South Africa (6), Seychelles (1), Syrian Arab Republic (22), Thailand (77), Tajikistan (7), Tunisia (6), Turkey (10), Taiwan (1), Tanzania (2), Uzbekistan (31), Saint Vincent (20), Venezuela (15), Vietnam (51), Yemen (10), South Africa (138), and Zimbabwe (1).

- Countries (removed settlements) to the urban extents:
  - Argentina (3), Algeria (1), Great Britain (1), India (18), Nigeria (1), Philippines (1), and South Africa (1).

## Appendix 2. Contributing Authors & Documentation Revision History

Revision Date	Contributors	Revisions
March 27, 2017	Valentina Mara, Alex de Sherbinin	This Revision 01 document is an update to the Version 1 documentation