# Documentation for the Global Urban Heat Island (UHI) Data Set, 2013

**September 2016**

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

## Abstract

This document presents the development of the Global Urban Heat Island (UHI) Data Set, 2013. The Introduction describes the motivation for producing the UHI data set, and summarizes the approach taken. Details of the input data, processing steps, and final distributed data set are covered in the Data and Methodology, and Data Set Description sections. Additional sections of this documentation describe potential use cases, limitations, and use constraints.

## Data set citation:

Center for International Earth Science Information Network (CIESIN), Columbia University. 2016. Global Urban Heat Island (UHI) Data Set, 2013. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). http://dx.doi.org/10.7927/H4H70CRF. Accessed DAY MONTH YEAR.

## Suggested citation for this document:

Center for International Earth Science Information Network (CIESIN), Columbia University. Documentation for the Global Urban Heat Island (UHI) Data Set, 2013. Palisades NY: NASA Socioeconomic Data and Applications Center (SEDAC). http://doi.org/10.7927/H44M92HC. 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

Please address comments to SEDAC User Services
http://sedac.uservoice.com/knowledgebase/topics/75786

# Contents

# I. Introduction

The idea for a Global Urban Heat Island (UHI) Data Set arose from a joint Technical Interchange Meeting of the Socioeconomic Data and Applications Center (SEDAC) and Land Processes Distributed Active Archive Center (LP DAAC) in May 2013. The urban heat island (UHI) effect represents the relatively higher temperatures found in urban areas compared to surrounding rural areas due to higher proportions of impervious surfaces and the release of waste heat from vehicles and heating and cooling systems. Paved surfaces and built structures tend to absorb shortwave radiation from the sun and release long-wave radiation after a lag of a few hours. For this data set, UHI is defined as the temperature differential between an urban area and a buffer of 10km surrounding the urban area based on land surface temperature. The basic approach was to use LP DAAC's MODIS data on land surface temperature in conjunction with SEDAC's Global Rural-Urban Mapping Project (GRUMP) urban extents data (CIESIN, 2015) to estimate the average daytime maximum and average nighttime minimum land surface temperatures within the urban extents during the highest temperature period of the northern and southern hemisphere summers. The same average was calculated for a 10km buffer surrounding the urban extents. Nighttime was included in addition to daytime because the health impacts of the urban heat island effect are often most pronounced at night.

We used the maximum daytime and minimum nighttime land surface temperatures (LSTs) extracted from MODIS LST 8-day composites (NASA LP DAAC, 2002b) acquired during a 40-day time-span in July-August 2013 in the northern hemisphere and January-February 2013 in the southern hemisphere to produce global grids of summer daytime maximum temperature and summer nighttime minimum temperature. These grids, which comprise the Global Summer Land Surface Temperature (LST) Grids, 2013 data set (CIESIN, 2016), were then used as inputs to zonal statistics to calculate the average summer daytime maximum temperature and average summer nighttime minimum temperature within the GRUMP urban extents and within the 10km buffer around those extents. Thus, these data represent the zonal (urban and buffer) averages of the maximum temperature recorded for each 1-km pixel over 40 days (five MODIS 8-day composites) that fall during the hottest months of summer (or the locally defined hot season). From these outputs we calculated the temperature differential between the urban extents and the buffers. Note that we experimented with different buffer distances and settled on 10km because larger buffer distances result in intersections between the buffers and other nearby urban areas. This may account for why some smaller urban areas adjacent to large agglomerations actually have negative UHIs, meaning surface temperatures are lower in the urban footprint than in the surroundings.

# II. Data and Methodology

## Input data

The UHI data set was produced using two data sets:

(1) The Global Summer Land Surface Temperature (LST) Grids, 2013: daytime maximum temperature and nighttime minimum temperature grids (CIESIN, 2016), and
(2) SEDAC's Global Rural-Urban Mapping Project, Version 1 (GRUMPv1): Urban Extent Polygons (internal working version). (CIESIN, 2015)

We produced the Global Summer Land Surface Temperature Grids, 2013 data set using the Aqua Level-3 Moderate Resolution Imaging Spectroradiometer (MODIS) Version 5 global daytime and nighttime Land Surface Temperature (LST) 8-day composite data (MYD11A2) (NASA LP DAAC, 2002b) distributed by NASA's Land Processes Distributed Active Archive Center (LP DAAC), and acquired via MRTWeb 2.0 (NASA LP DAAC, 2009). The 8-day data are composed from the daily 1-kilometer LST product (MYD11A1) (NASA LP DAAC, 2002a) and represent the average values of clear-sky LSTs during an 8-day period.

The MODIS MYD11A2 data product uses a Sinusoidal grid tiling system with tiles covering an area 10 degrees by 10 degrees at the equator (Figure 1). The tile coordinate system starts at (0,0) (horizontal tile number *h*, vertical tile number *v*) in the upper left corner and proceeds right (horizontal) and downward (vertical). The tile in the bottom right corner is (35,17).

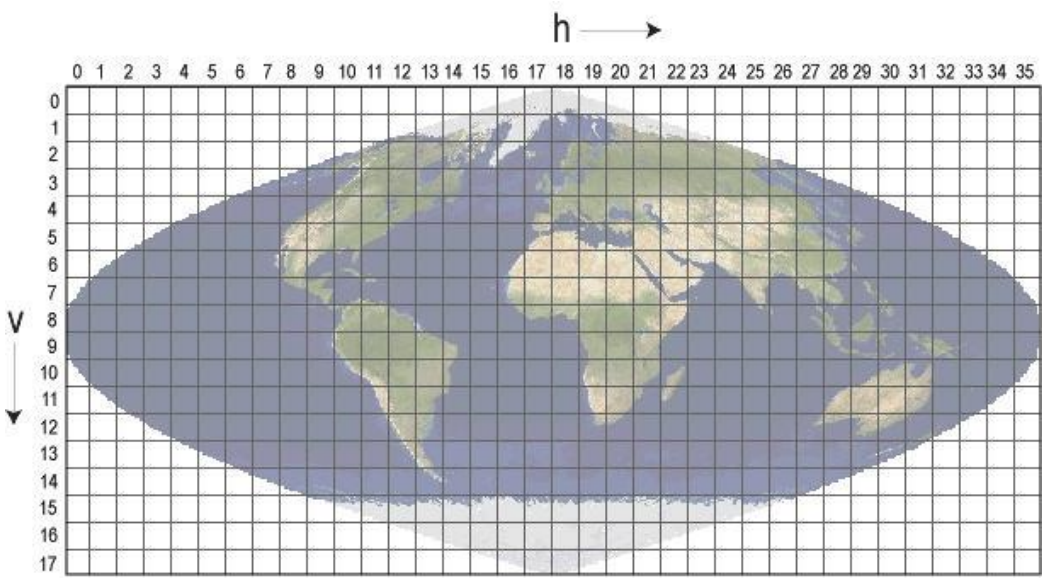

Figure 1. MODIS MYD11A2 Sinusoidal grid tiling system.
*Source: https://lpdaac.usgs.gov/dataset_discovery/modis*

The GRUMPv1 Urban Extent Polygons data set represents urban areas circa 1995, and was used as input to the urban extent grids that are currently disseminated by SEDAC (CIESIN, IFPRI, and CIAT, 2011). The data set is derived from nighttime lights, buffers of settlement points (where lights are not detected), settlement points, and their associated population count estimates for 1990, 1995 and 2000. The version used for the development of the UHI data set had some improvements made. Firstly, it includes only the urban extents identified by name and associated with at least one settlement point. This is a subset of v1, which includes many urban extents for which we couldn't find an associated settlement point. Secondly, it includes some new urban extents identified using the UN extended city population list or by correcting the georeference of some previously miss-referenced settlements. These changes resulted in some urban extents losing settlements, and some urban extents gaining settlements. In both cases, population values of the extent polygons were not adjusted, therefore the population of the extent is overestimated in the former case and underestimated in the latter case.

## Methods

### *1. Global Summer Land Surface Temperature (LST) Grids, 2013*

We defined our time period of interest as the highest temperature period of the northern and southern hemisphere summers. In the northern hemisphere this roughly corresponds to the period from July to early August, while in the southern hemisphere it corresponds to the period from January to February. We identified five MODIS LST 8-day composites with Julian day acquisition dates in a 40-day time-span in July-August (Julian days 185-224) 2013 in the northern hemisphere ($v$ = 1-8), and five 8-day composites with Julian day acquisition dates in a 40-day time-span in January-February (Julian days 001-040) 2013 in the southern hemisphere ($v$ = 9-13). We then used MRTweb to select MODIS tiles by tile location ($h,v$). MRTweb will process up to 50 images per session, so we processed the data in batches, selecting 10 adjacent tiles to mosaic and processing all five 8-day composites for that selection. For each 8-day composite, both day and night image bands were selected, projected to WGS-84 and then mosaicked with the other selected tiles to create five daytime LST mosaics and five nighttime LST mosaics. To achieve global coverage (60 S latitude to 80 N latitude), approximately 200 tiles were selected to produce 80 MRTweb mosaics (40 each of daytime and nighttime LST) for each of five Julian dates. A total of 400 mosaics were downloaded.

We then used ENVI software version 4.8 (Exelis Visual Information Solutions, 2011) to create layer stacks of the daytime and nighttime mosaic imagery for a given mosaicked MRTweb selection. This reduced the number of files we were working with from 400 mosaics for individual dates to 80 layer stacks of daytime and nighttime imagery. We prepared an *R* script (R Development Core Team, 2013) to process each layer stack. The first part of the *R* script converted the data from Kelvin to degrees Celsius and recoded unrealistic values to NoData. Then we mosaicked the scenes into global files in Esri ArcMap version 10.2.2 (Esri, 2014), and ran the second part of the *R* script to extract the maximum LST value in the global layer stack of five daytime images and the minimum LST value in the global layer stack of five nighttime images, for each pixel location.

While conducting quality assurance during processing, we noticed that some tropical locations, especially in India and Indonesia, were missing data due to cloud cover related to monsoon periods. We decided to patch these data gaps with scenes from April-May (Julian days: 97-129) for the northern hemisphere, and from December 2013-January 2014 (Julian days: 337-001) for the southern hemisphere. We re-ran the script on these files, and the outputs were then mosaicked to the global files such that only the pixels with missing values were updated with values from the supplementary files. Some data gaps remain in areas where data were insufficient (e.g., Central Africa).

The final outputs prepared for distribution as the Global Summer Land Surface Temperature (LST) Grids, 2013 data set (CIESIN, 2016) are global grids of summer daytime maximum temperature and summer nighttime minimum temperature and regional subsets of the global grids, all in GeoTIFF format. All grids are in degrees Celsius at a spatial resolution of 30 arc-seconds (~1km).

## *2. Global Urban Heat Island (UHI) Data Set, 2013*

To produce the UHI data set, we started with the updated version of the GRUMP Urban Extent Polygons data set described above and generated 10-kilometer buffers around the extents to represent rural areas. We then used the ArcGIS *Zonal Statistics as Table* tool (Esri, 2014) to summarize the daytime maximum land surface temperature and nighttime minimum land surface temperature within these *urban extents* and *urban extent buffers*.

The ArcGIS *Zonal Statistics as Table* tool was unable to generate statistics for the overlapping features which occur in the urban extent buffers layer. Therefore, we wrote a Python script (Python Software Foundation, 2014a and 2014b) using Esri's ArcPy site package Spatial Analyst module (Esri, 2014) to run the tool iteratively for each of the polygons in the urban extents layer and each of the polygons in the urban extent buffers layer for both the LST summer daytime maximum and summer nighttime minimum temperature grids.

At the end of this processing, we had 4 intermediate data sets representing:

1. daytime maximum LST statistics within urban extents,
2. nighttime minimum LST statistics within urban extents,
3. daytime maximum LST statistics within buffers, and
4. nighttime minimum LST statistics within buffers.

The mean statistic from each output was joined to a copy of the urban extents data layer. During a quality review of the data, we examined all features (urban extents and urban extent buffers) for which processing produced NULL results for the mean statistic. Zonal statistics failed for two urban extents because the features were too small relative to the LST raster cell size. These features were removed from the data set. All additional NULL results were due to the urban or buffer extents overlaying NoData cells in the daytime and/or nighttime LST rasters. Seven urban extents yielded NULL results for both day and night rasters. These features were removed from the data set. For the remaining extents, if either the daytime or nighttime mean within the urban extent was non-NULL, the feature was retained.

Two additional fields were calculated: (1) the difference between the mean daytime LST within urban extents and the mean daytime LST within their associated 10km buffers, and (2) the difference between the mean nighttime LST within urban extents and the mean nighttime LST within their associated 10km buffers.

## III. Data Set Description

The Global Urban Heat Island (UHI) Data Set, 2013 estimates the summer daytime maximum and summer nighttime minimum land surface temperatures (in degrees Celsius) within urban areas and rural areas, defined as a 10km buffer around the urban area, as well as the temperature difference between urban and rural areas.

***Data set web page:***
http://sedac.ciesin.columbia.edu/data/set/sdei-global-uhi-2013

***Data set format:***
The data set is available in shapefile format or Microsoft Excel workbook. The data are stored in geographic coordinates of decimal degrees based on the World Geodetic System spheroid of 1984 (WGS84). Each downloadable is a compressed zip file which contains: 1) either the shapefile or the Microsoft Excel workbook, and 2) PDF documentation. Downloaded zipfiles need to be uncompressed in a single folder using either WinZip (Windows file compression utility) or a similar application before the file can be accessed by your GIS software package or Microsoft Excel.

***Data set downloads:***
sdei-global-uhi-2013-shp.zip
sdei-global-uhi-2013-table.zip

***Codebook:***

| Field Name | Field Description |
|---|---|
| ISOURBID | Unique ID created by concatenating ISO3 code and Urban ID number (URBID) |
| ISO3 | International Standards Organization three-letter country code |
| URBID | Urban ID number (unique within country) |
| NAME | City or urban agglomeration name |
| SCHNM | City or urban agglomeration name in CAPS, concatenated, and without accents |
| ES90POP | Estimated population for 1990 |
| ES95POP | Estimated population for 1995 |
| ES00POP | Estimated population for 2000 |
| SQKM_FINAL | Area of urban extent in square kilometers |
| URB_D_MEAN | Average summer daytime maximum land surface temperature within urban extent (degrees Celsius) |
| BUF_D_MEAN | Average summer daytime maximum land surface temperature within 10km buffer around urban extent (degrees Celsius) |

| | |
|---|---|
| D_T_DIFF | The difference (in degrees Celsius) in average summer daytime maximum land surface temperature between the urban area and buffer area (URB_D_MEAN minus BUF_D_MEAN) |
| URB_N_MEAN | Average summer nighttime minimum land surface temperature within urban extent (degrees Celsius) |
| BUF_N_MEAN | Average summer nighttime minimum land surface temperature within 10km buffer around urban extent (degrees Celsius) |
| N_T_DIFF | The difference (in degrees Celsius) in average summer nighttime minimum land surface temperature between the urban area and buffer area (URB_N_MEAN minus BUF_N_MEAN) |
| LATITUDE | Latitude of centroid of urban extent in decimal degrees |
| LONGITUDE | Longitude of centroid of urban extent in decimal degrees |

## IV. Potential Use Cases

The Global UHI data set could be used in a number of ways. For example, a public health researcher may be interested in identifying cities with pronounced UHIs for research on the health impacts of urban heat islands, or professionals working on urban adaptation for a given city may wish to see how their city compares to similar sized cities in similar climate zones or ecoregions. Integration of this data set with population data could help to quantify populations exposed to extreme heat during the summer months. The global temperature grid could serve as a baseline that will allow comparisons of urban temperatures in future years.

## V. Acknowledgments

The MODIS MYD11A2 data were retrieved from MRTWeb 2.0, courtesy of the NASA EOSDIS Land Processes Distributed Active Archive Center (LP DAAC), USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota (https://lpdaac.usgs.gov/data_access).

Funding for this data set was provided under the U.S. National Aeronautics and Space Administration (NASA) contract NNG13HQ04C for 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.

## VI. 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.

## VII. 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.

## VIII. Recommended Citation(s)

### Data set:

Center for International Earth Science Information Network (CIESIN), Columbia University. 2016. Global Urban Heat Island (UHI) Data Set, 2013. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). http://dx.doi.org/10.7927/H4H70CRF. Accessed DAY MONTH YEAR.

## IX. Source Code

The *R* script used in processing the MODIS MYD11A2 dataset to produce the Global Summer Land Surface Temperature (LST) Grids, 2013 (CIESIN, 2016) has been archived with the LST Grids. The Python script used to execute zonal statistics on the LST grids, has been archived with the UHI data set. Both scripts are available upon request. Please contact SEDAC User Services at ciesin.info@ciesin.columbia.edu to request the code.

## X. References

Center for International Earth Science Information Network (CIESIN), Columbia University. 2016. *Global Summer Land Surface Temperature (LST) Grids, 2013*. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). http://dx.doi.org/10.7927/H408638T.

Center for International Earth Science Information Network (CIESIN), Columbia University. 2015. *Global Rural-Urban Mapping Project, Version 1 (GRUMPv1): Urban Extent Polygons (internal working version)*. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC).

Center for International Earth Science Information Network (CIESIN), Columbia University, International Food Policy Research Institute (IFPRI), The World Bank, and Centro Internacional de Agricultura Tropical (CIAT). 2011. *Global Rural-Urban Mapping Project, Version 1 (GRUMPv1): Urban Extents Grid*. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). http://dx.doi.org/10.7927/H4GH9FVG.

ESRI. 2014. *ArcGIS Desktop: Release 10.2.2*. Redlands, CA: Environmental Systems Research Institute. http://www.esri.com

Exelis Visual Information Solutions. 2011. *ENVI Version 4.8,* Boulder, Colorado. http://www.exelisvis.com

NASA Land Processes Distributed Active Archive Center (LP DAAC). 2002a. *MODIS MYD11A1 Version 5*. USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota (https://lpdaac.usgs.gov), https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/myd11a1

NASA Land Processes Distributed Active Archive Center (LP DAAC). 2002b. *MODIS MYD11A2 Version 5*. NASA EOSDIS Land Processes DAAC, USGS Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota (https://lpdaac.usgs.gov), accessed June 2014, at https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/myd11a2

NASA Land Processes Distributed Active Archive Center (LP DAAC). 2009. *MRTWeb 2.0 (MODIS reprojection tool web interface)*. Available at https://lpdaac.usgs.gov/data_access/mrtweb

Python Software Foundation. 2014a. *IDLE (Python's Integrated DeveLopment Environment) version 2.7.8*. Available at http://www.python.org/idle

Python Software Foundation. 2014b. *Python programming language, version 2.7.8*. http://www.python.org

R Development Core Team. 2013. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Last accessed on March 25, 2014. Available at http://www.R-project.org

## XI. Documentation Copyright and License