USAID
FROM THE AMERICAN PEOPLE

The Geospatial Covariate Datasets Manual

The Demographic and Health Surveys Program

Second Edition

The DHS Program Geospatial Covariate Datasets Manual

Second Edition

Benjamin Mayala[1]
Thomas D. Fish[1]
David Eitelberg[2]
Trinadh Dontamsetti[1]

ICF
Rockville, Maryland, USA

September 2018

[1] The DHS Program, ICF
[2] The DHS Program, Blue Raster

Corresponding author: Benjamin Mayala, International Health and Development, ICF, 530 Gaither Road, Suite 500, Rockville, MD 20850, USA; phone: 301-572-0507; email: Ben.Mayala@icf.com

Acknowledgement: The authors would like to thank Trevor Croft and Clara R. Burgert for their work to get the first version of this dataset published.

This publication and dataset were developed with support provided by the United States Agency for International Development (USAID) through The Demographic and Health Surveys Program (#AID-OAA-C-13-00095). The views expressed are those of the authors and do not necessarily reflect the views of USAID or the United States government.

The DHS Program assists countries worldwide in the collection and use of data to monitor and evaluate population, health, and nutrition programs. Information on The DHS Program may be obtained from ICF, 530 Gaither Road, Suite 500, Rockville MD, 20850, USA; telephone: 301-407-6500; fax: 301407-6501; e-mail: info@DHSprogram.com; website: www.DHSprogram.com.

Recommended Citation:

Mayala, Benjamin, Thomas D. Fish, David Eitelberg, and Trinadh Dontamsetti. 2018. *The DHS Program Geospatial Covariate Datasets Manual* (Second Edition). Rockville, Maryland, USA: ICF.

CONTENTS

1 Rationale **4**

2 DHS Clusters **4**

3 Extraction Methods **4**

3.1 Neighborhood Calculations using Raster Data 5

3.2 Distance Calculations using Vector Data 6

4 Notes about the Data Description **7**

5 Data Description **8**

References **46**

1 RATIONALE

The DHS Program routinely collects GPS location data of surveyed clusters (DHS, MIS, and AIS surveys). These data are processed and made available upon request for download through The DHS Program website following the application of geospatial displacement on the GPS cluster data to protect the confidentiality of respondents.

The data are utilized by thousands in academia (students and researchers), government (researchers, decision makers, program planners, and implementing agencies) and the private sector. The nature of these requests made it apparent that The DHS Program's data are often analyzed in conjunction with geospatial covariates to determine the impact of location on health outcomes. However, these covariate data often come from multiple different sources at different levels of national coverage and with varying quality, making it difficult for researchers to link The DHS Program's data to these covariates and conduct analyses. To address this, The DHS Program Geospatial Team endeavored to prepare and make freely available a set of standardized files of the most commonly used geospatial covariates which can easily be linked to DHS datasets without the need for the GPS data itself, increasing accessibility to those with little or no GIS experience.

2 DHS CLUSTERS

For all of the extractions, we used the publicly available cluster locations published by The DHS Program. The GPS location of the center of each cluster is recorded during either the fieldwork or listing stage of the survey. Those locations are processed to verify they are within the correct administrative units. To protect the confidentiality of our respondents the locations are displaced – sometimes called 'geo-masked' or 'geo-scrambled'. Each of the clusters was displaced from the actual location by up to 2 kilometers (for urban points) and 10 kilometers (for rural points). An effort is made to ensure that during the displacement procedure the point does not move between large administrative units. More information about the displacement procedure used can be found in Burgert et al. (2013).

3 EXTRACTION METHODS

The covariate variables came from two types of data: raster and vector. Raster data, such as images and modeled surfaces, rely on pixels or cells to convey their data values. On the other hand, vector data, such as points, lines, and polygons, show the discrete location or boundary of a feature. Because of the differences in the data types, the methods needed to extract meaningful values varied. The conceptual framework, Figure 1, provides an overview of the extraction process that was undertaken using the following steps:

Step 1 - Geospatial covariate layers (i.e. modeled surfaces) that are relevant to The DHS Program indicators were acquired from publicly available remote sensing and modeling sources. GPS coordinates representing the location of a survey cluster were obtained from The DHS program. In addition to modeled surfaces, we also included vector (polygon and line) data, which were obtained from various publicly available sources.

Step 2 - Mosaicing (if needed) was undertaken to spatially mosaic (stitch together) modeled surfaces delivered in tiled format. We then masked out oceans and large lakes from the modeled surfaces to remove null and skewing values.

Step 3 - Raster and vector datasets were imported and linked to GPS using a standalone Python programming language script and ArcGIS, respectively.

Step 4 - Finally, we extracted the values at the point.

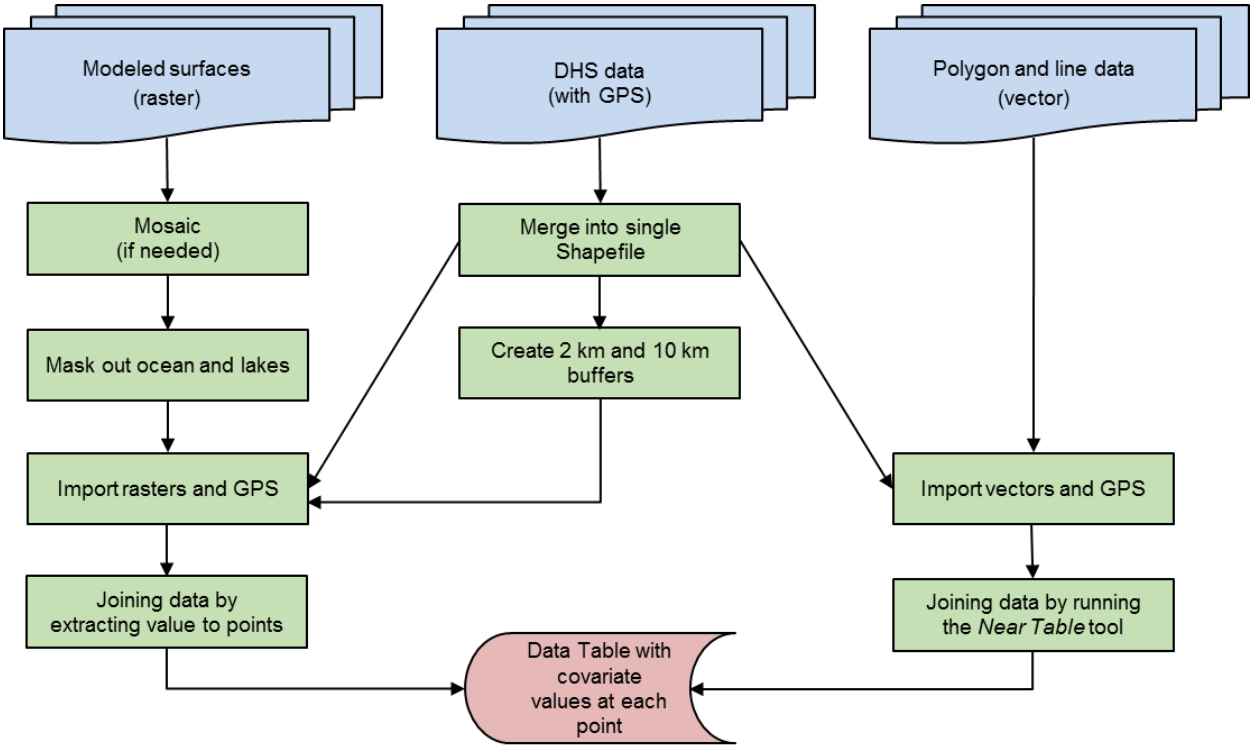

Figure 1: Conceptual framework of the covariate extraction procuress.

If the data were provided as a raster, we used a neighborhood calculation. For data in a vector format, we used a distance measure. The procedures that we used attempted to compensate as much as possible for the uncertainty of the cluster locations, but the uncertainty adds an element of error to all extracted values.

3.1 Neighborhood Calculations using Raster Data

The neighborhood calculations were done using several Python scripts that moved data through the process and did the actual extractions. Instead of writing a zonal statistics algorithm ourselves, we relied on the *rasterstats* package's version of the algorithm (Perry 2016).

First, a circular buffer was drawn around each point. For all of the covariates, the buffers had a radius of either 2 kilometers for urban points or 10 kilometers for rural points. This was done to compensate for the geographic displacement of points and to account for the varying pixel size of data sets. Figure 2 depicts an example of a 10 kilometer buffer around a cluster from the 2009 Guyana DHS survey, overlaid on a raster used to extract rainfall measurements.

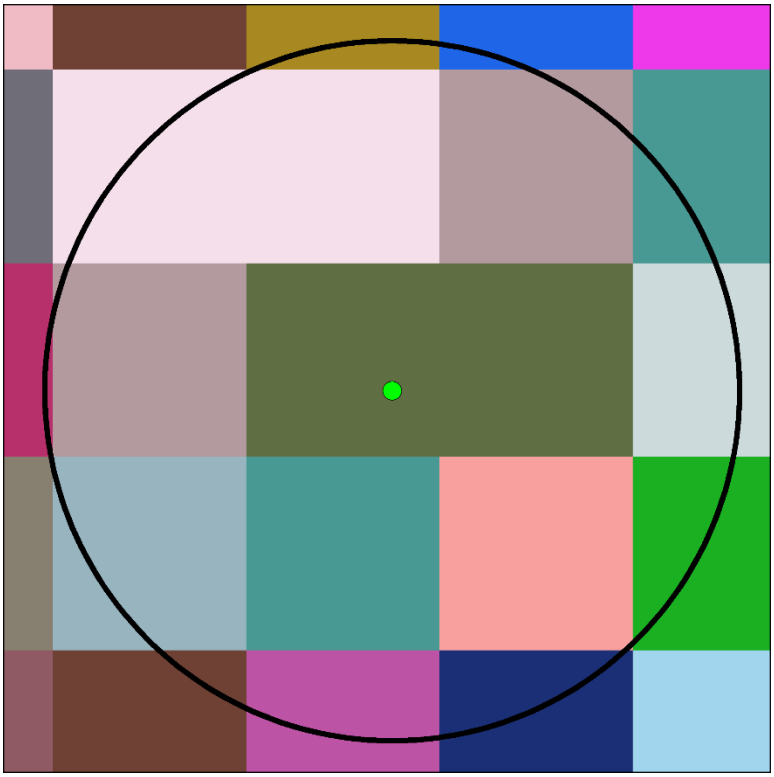

Figure 2: A 10-kilometer buffer around a DHS cluster location

All raster cells with centroids falling within these buffers were used in the raster extraction calculation. Any raster cell whose centroid did not fall within these buffers was not used for the calculations. The zonal statistics algorithm can output a number of summary statistics including sum, count, and mean. If this process failed to return a value, the value of the cell directly underneath the cluster location was used. These steps were performed in sequence for all GPS points in the input DHS dataset.

3.2 Distance Calculations using Vector Data

The distance from the DHS points to protected areas, international boundaries, lakes, or the coastline was measured using the *Near Table* tool in ArcMap (ESRI 2017). The tool calculates the geodesic distance between each DHS point and the nearest boundary of a selected polygon

or line feature class. The distance calculated was appended as an attribute to the DHS points' attribute table, which we then joined with all the data from the raster extraction activity (Figure 1).

4 NOTES ABOUT THE DATA DESCRIPTION

- Covariates that are presented for different years (e.g. 2000, 2005, 2010, 2015) are referred to as YEAR instead of listing out the individual years; likewise covariates presented for different months are referred to as MONTH.
- If the data for a covariate is not available, its field will be "-9999".
- All calculations between angular units and distance units are performed at the equator with 1 decimal degree = 110.567 kilometers.

5 DATA DESCRIPTION

Column Name: All Population Count YEAR

Derived Data Set: WorldPop

Derived Data Set Cell Size: 0.000833333 decimal degrees (~100 m)

Summary Statistic: Sum

Year: The closest national census to 2000, 2005, 2010, or 2015

Units: Number of people

Description:

The count of individuals living within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS survey cluster at the time of measurement (year).

An understanding of the numbers, characteristics and locations of human populations underpins operational work, policy analyses and scientific development globally across multiple sectors. Most methods for estimating population rely on census data. However, in most countries population censuses are conducted every 10 years at best and can be longer in many low-income countries (Alegana et al. 2015). Thus, data from the census can often be outdated, unreliable, and have low granularity (Tatem 2014). New data sources and recent methodological advances made by the WorldPop project (University of Southampton) now provide high resolution, open, and contemporary data on human population, allowing accurate measurement of local population characteristics across national and regional scales.

Citation:

Gaughan, Andrea E., Forrest R. Stevens, Catherine Linard, Peng Jia, and Andrew J. Tatem. 2013. "High Resolution Population Distribution Maps for Southeast Asia in 2010 and 2015." *PLOS ONE* 8 (2):e55882. http://10.1371/journal.pone.0055882.

Linard, Catherine, Marius Gilbert, Robert W. Snow, Abdisalan M. Noor, and Andrew J. Tatem. 2012. "Population Distribution, Settlement Patterns and Accessibility across Africa in 2010." *PLOS ONE* 7 (2):e31743. http://doi.org/10.1371/journal.pone.0031743.

Sorichetta, Alessandro, Graeme M. Hornby, Forrest R. Stevens, Andrea E. Gaughan, Catherine Linard, Andrew J. Tatem. 2015. “High-resolution gridded population datasets for Latin America and the Caribbean in 2010, 2015, and 2020” *Scientific Data* 2. http://doi.org/10.1038/sdata.2015.45

WorldPop. 2016. “Africa Continental Population Datasets (2000 - 2020).” Accessed August 21, 2017. http://doi.org/10.5258/SOTON/WP00004.

WorldPop. 2016. “Asia Continental Population Datasets (2000 – 2020).” Accessed August 21, 2017. http://doi.org/10.5258/SOTON/WP00013.

WorldPop. 2016. “Latin America and the Caribbean Continental Population Datasets (2000 - 2020).” Accessed August 21, 2017. http://doi.org/10.5258/SOTON/WP00138.

Column Name: Annual Precipitation YEAR

Derived Data Set: CRU TS v. 4.01

Derived Data Set Cell Size: 0.5 decimal degrees (~55 km)

Summary Statistic: Mean

Year: 2000, 2005, 2010, 2015

Units: Millimeters

Description:

The average precipitation measured within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS survey cluster in a given year.

The Climate Research Unit (CRU) of the University of East Anglia, UK, produces a range of global climate time series gridded data which are derived from meteorological stations across the world's land areas. The datasets are provided on high resolution (0.5 x 0.5 degrees) grids over the period 1901- 2016 (Harris and Jones 2017; Harris et al. 2014). The main purpose for which this dataset was constructed was to provide modelers (climate and environmental) with some of the inputs they require to run their models. These data have found extensive use in numerous geospatial modeling and mapping studies (Caminade et al. 2014; Chabot-Couture, Nigmatulina, and Eckhoff 2014; Graetz et al. 2018; Osgood-Zimmerman et al. 2018; Rogers and Randolph 2006).

Citation:

Harris, Ian, P. D. Jones, Timothy J. Osborn, David Lister. 2014. "Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset". *International Journal of Climatology* 43: 623-642. http://dx.doi.org/10.1002/joc.3711

Climate Research Unit. 2017. "CRU TS v. 4.01". Accessed August 21, 2018. http://doi.org/10/gcmcz3

Column Name: Aridity YEAR

Derived Data Set: CRU TS v. 4.01

Derived Data Set Cell Size: 0.5 decimal degrees (~55 km)

Summary Statistic: Mean

Year: 2000, 2005, 2010, 2015

Units: Index between 0 (most arid) and 300 (most wet).

Description:

The dataset represents the average yearly precipitation divided by average yearly potential evapotranspiration, an aridity index defined by the United Nations Environmental Programme (UNEP).

Aridity Index (AI), which is defined as the ratio of annual precipitation to annual potential evapotranspiration, is a key parameter in drought characterization (Salem 1989; Tereshchenko et al. 2015). Investigating variation of aridity and the role of climate variables are thus of great significance for managing agricultural water resources and maintaining regional ecosystem stability. If the aridity index becomes higher than normal in a region, the climate tends to suffer from drought and water resource shortages (Allen et al. 1998; Li et al. 2017).

The global aridity that was used in the previous extraction of covariates (First Edition) was modeled using data available from the WorldClim Global Climate Data (Hijmans and Graham 2006) as input. This is one grid layer representing the annual average over the 1950 – 2000 period.

We have updated the aridity covariate for the period of 2000, 2005, 2010 and 2015 using high resolution grids obtained from the CRU datasets (Harris and Jones 2017; Harris et al. 2014). We used precipitation and potential evapotranspiration datasets to quantify the aridity index covariate following methods described in detail by Arnold (1997).

Citation:

Harris, Ian, P. D. Jones, Timothy J. Osborn, David Lister. 2014. “Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset”. *International Journal of Climatology* 43: 623-642. http://dx.doi.org/10.1002/joc.3711

Climate Research Unit. 2017. “CRU TS v. 4.01”. Accessed August 21, 2018. http://doi.org/10/gcmcz3

Column Name: BUILT Population YEAR

Derived Data Set: GHS built-up grid, derived from Landsat, multitemporal (1975, 1990, 2000, 2014)

Derived Data Set Cell Size: 1 km

Summary Statistic: Mode

Year: 1990, 2000, or 2014

Units: Built-up index between 0.00 (extremely rural) and 1.00 (extremely urban)

Description:

An index ranging from 0.00 (extremely rural) to 1.00 (extremely urban) for the area within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS survey cluster location.

This is a high-resolution dataset on built-up presence derived from remote sensing image collections (e.g. Landsat imagery). The data is obtained from the Global Human Settlement Layer (GHSL) project, which produces global spatial information about the human presence on the planet over time. GHSL technology relies on automatic analysis of satellite imagery to produce fine-scale maps quantifying built-up structures in terms of their location and density. The image processing technology exploits structure (texture, morphology, and pattern) as key information, outputting a texture-derived “built-up presence index”. The distribution of built-up areas is expressed as their proportion (ratio) of occupied footprint in each cell (Pesaresi et al. 2016).

Citation:

Pesaresi, Martino, Daniele Ehrlich, Aneta Florczyk, Sergio Freire, Andreea Julea, Thomas Kemper, Pierre Soille, Vasileios Syrris. 2015. “GHS built-up grid, derived from Landsat, multitemporal (1975, 1990, 2000, 2014).” Accessed August 21, 2017. http://data.europa.eu/89h/jrc-ghsl-ghs_built_ldsmt_globe_r2015b.

Column Name: Day Land Surface Temp YEAR

Derived Data Set: MYD11C3: MODIS/Aqua Land Surface Temperature and Emissivity Monthly

Derived Data Set Cell Size: .05 decimal degrees (~6 km)

Summary Statistic: Mean

Year: 2000, 2005, 2010, 2015

Units: Yearly average temperature in Degrees Celsius

Description:

The mean annual daytime land surface temperature within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS survey cluster location.

Land surface temperature (LST) estimates the temperature of the land surface (skin temperature), which is detected by satellites by looking through the atmosphere to the ground. The LST is not equivalent to near-surface air temperature measured by ground stations, and their relationship is complex from a theoretical and empirical perspective (Vancutsem et al. 2010; Yang, Cai, and Yang 2017; Zhou et al. 2012).

The global LST [day] grids represent daytime temperature in degrees Celsius at a spatial resolution of 0.05° × 0.05°. The grids are produced using Moderate Resolution Imaging Spectroradiometer (MODIS) product (MYD11C3). The MYD11C3 is a monthly composited average, derived from the daily global product (MYD11C1) and distributed by the Land Processes Distributed Archive Center (LPDAAC). The type of "surface" MODIS measures varies as a function of location. In some places, the measurement represents the skin temperature of the bare land surface. In other places, the temperature represents the skin temperature of whatever is on the land, including ice, canopy cover or human-made structures (https://neo.sci.gsfc.nasa.gov). LST surface maps can be used in a number of ways to assess surface temperature differentials as they relate to different land cover types, ecosystems, and health impacts.

Citation:

Wan, Zhengming, S. Hook, G. Hulley. 2015. “MYD11C3 MODIS/Aqua Land Surface Temperature/Emissivity Monthly L3 Global 0.05Deg CMG V006”. Accessed August 21, 2018. https://doi.org/10.5067/MODIS/MYD11C3.006

Column Name: Diurnal Temperature Range YEAR

Derived Data Set: CRU TS v. 4.01

Derived Data Set Cell Size: 0.5 decimal degrees (~55 km)

Summary Statistic: Mean

Year: 2000, 2005, 2010, 2015

Units: Degrees Celsius

Description:

The annual average of daytime temperature range for the locations within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS survey cluster location. Diurnal or daytime temperature range is calculated by subtracting the station maximum and minimum temperatures.

The Climate Research Unit (CRU) of the University of East Anglia, UK, produces a range of global climate time series gridded data, which are derived from meteorological stations across the world's land areas. The datasets are provided on high resolution (0.5 x 0.5 degrees) grids over the period 1901- 2016 (Harris and Jones 2017; Harris et al. 2014). The main purpose for which this dataset was constructed was to provide modelers (climate and environmental) with some of the inputs they require to run their models. These data have found extensive use in numerous geospatial modeling and mapping studies (Caminade et al. 2014; Chabot-Couture, Nigmatulina, and Eckhoff 2014; Graetz et al. 2018; Osgood-Zimmerman et al. 2018; Rogers and Randolph 2006).

Citation:

Harris, Ian, P. D. Jones, Timothy J. Osborn, David Lister. 2014. "Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset". *International Journal of Climatology* 43: 623-642. http://dx.doi.org/10.1002/joc.3711

Climate Research Unit. 2017. "CRU TS v. 4.01". Accessed August 21, 2018. http://doi.org/10/gcmcz3

Column Name: Drought Episodes

Derived Data Set: Global Drought Hazard Frequency and Distribution, v1 (1980–2000)

Derived Data Set Cell Size: 0.0417 decimal degrees (~5 km)

Summary Statistic: Mode

Year: Based on 1980-2000 precipitation data

Units: Individual classes between 1 (Low Drought) and 10 (High Drought)

Description:

The average number of drought episodes (categorized between 1 (low) and 10 (high)) for the areas within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS survey cluster location.

The global drought episodes dataset is produced by the International Research Institute for Climate Prediction's (IRI) Weighted Anomaly of Standardized Precipitation (WASP). Utilizing average monthly precipitation data, the WASP assesses the precipitation deficit or surplus over a three month temporal window that is weighted by the magnitude of the seasonal cyclic variation in precipitation. The three months' averages are derived from the precipitation data and the median rainfall for the 21 year period is calculated for each grid cell. Grid cells where the three month running average of precipitation is less than 1 mm per day are excluded. Drought events are identified when the magnitude of a monthly precipitation deficit is less than or equal to 50 percent of its long-term median value for three or more consecutive months. Grid cells are then divided into 10 classes having an approximately equal number of grid cells. Higher grid cell values denote higher frequencies of drought occurrences (Dilley et al. 2005).

Citation:

Center for Hazards and Risk Research - Columbia University, Center for International Earth Science Information Network - Columbia University, and International Research Institute for Climate and Society - Columbia University. 2005. “Global Drought Hazard Frequency and Distribution.” Accessed August 21, 2017. http://dx.doi.org/10.7927/H4VX0DFT.

Dilley, Maxx, Robert Chen, Uwe Deichmann, Arthur Lerner-Lam, Margaret Arnold, Jonathan Agwe, Piet Buys, Oddvar Kjekstad, Bradfield Lyon, and Gregory Yetman. 2005. *Natural Disaster Hotspots: A Global Risk Analysis*. Washington: The World Bank. http://hdl.handle.net/10986/7376.

Column Name: Enhanced Vegetation Index YEAR

Derived Data Set: Vegetation Index and Phenology (VIP) Phenology EVI-2 Yearly Global 0.05Deg CMG V004

Derived Data Set Cell Size: 0.05 decimal degrees (~5 km)

Summary Statistic: Mean

Year: 1985, 1990, 1995, 2000, 2005, 2010, or 2015

Units: Vegetation index value between 0 (least vegetation) and 10000 (Most vegetation).

Description:

The average vegetation index value within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS survey cluster at the time of measurement (year).

EVI composites are produced globally from data collected by the MODIS sensor on the Terra satellite (MOD13A3). The MODIS EVI products are computed from atmospherically corrected bi-directional surface reflectance data that have been masked for water, clouds, heavy aerosols, and cloud shadows. These data are used for global monitoring of vegetation condition.

Citation:

Kamel Didan. 2016. “NASA MEaSUREs Vegetation Index and Phenology (VIP) Phenology EVI2 Yearly Global 0.05Deg CMG”. Accessed August 21, 2017. https://doi.org/10.5067/measures/vip/vipphen_evi2.004.

Column Name: Frost Days YEAR

Derived Data Set: CRU TS v. 4.01

Derived Data Set Cell Size: 0.5 decimal degrees (~55 km)

Summary Statistic: Mean

Year: 2000, 2005, 2010, 2015

Units: Days

Description:

The average number of days in which the minimum temperatures of the location surrounding the DHS survey cluster within 2 km (urban) or 10 km (rural) buffers met the criteria to be categorized as a "frosty" day.

The Climate Research Unit (CRU) of the University of East Anglia, UK, produces a range of global climate time series gridded data, which are derived from meteorological stations across the world's land areas. The datasets are provided on high resolution (0.5 x 0.5 degrees) grids over the period 1901- 2016 (Harris and Jones 2017; Harris et al. 2014). The main purpose for which this dataset was constructed was to provide modellers (climate and environmental) with some of the inputs they require to run their models. These data have found extensive use in numerous geospatial modeling and mapping studies (Caminade et al. 2014; Chabot-Couture, Nigmatulina, and Eckhoff 2014; Graetz et al. 2018; Osgood-Zimmerman et al. 2018; Rogers and Randolph 2006).

Frost days is a synthetic measurement that is based off of the minimum temperature. The full formula to calculate the number of days can be found in the cited Harris et al. (2014) or in New, Hulme, and Jones (2000).

Citation:

Harris, Ian, P. D. Jones, Timothy J. Osborn, David Lister. 2014. "Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset". *International Journal of Climatology* 43: 623-642. http://dx.doi.org/10.1002/joc.3711

Climate Research Unit. 2017. "CRU TS v. 4.01". Accessed August 21, 2018. http://doi.org/10/gcmcz3

Column Name: Global Human Footprint

Derived Data Set: Global Human Footprint (Geographic), v2 (1995–2004)

Derived Data Set Cell Size: 0.008333 decimal degrees (~1 km)

Summary Statistic: Mean

Year: Based on 1995-2004 data

Units: Global human footprint index between 0 (extremely rural) and 100 (extremely urban)

Description:

The average of an index between 0 (extremely rural) and 100 (extremely urban) for the location within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS survey cluster.

The Global Human Footprint Dataset is the Human Influence Index (HII) normalized by biome and realm, developed by the Last of the Wild Project (LWP-2). The human influence index is a global dataset given at spatial resolution of 1 x 1 km grid cells, which is created from nine global data layers covering human population pressure (population density), human land use and infrastructure (built-up areas, nighttime lights, land use/land cover), and human access (coastlines, roads, railroads, navigable rivers).

Citation:

Wildlife Conservation Society, and Center for International Earth Science Information Network - Columbia University. 2005. “Global Human Footprint Dataset.” Accessed August 21, 2017. http://dx.doi.org/10.7927/H4M61H5F.

Column Name: Gross Cell Production

Derived Data Set: Geographically based Economic data

Derived Data Set Cell Size: ~5 km

Summary Statistic: Mean

Year: 2005

Units: Purchasing Power Parity (PPP) in 2005 US Dollars

Description:

The average Purchasing Power Parity (PPP) in 2005 US dollars for the 2 km (urban) or 10 km (rural) buffers surrounding the DHS survey cluster.

Most studies in economics emphasize policy factors and level or growth per capita and tend to ignore geographic factors such as climate, proximity to water, soil type, etc. (Nordhaus 2006).

The G-Econ project developed data on “gross cell product”, derived from geophysically-based datasets on economic activity. The conceptual basis of GCP is the equivalent to that of the gross domestic product (GDP), except that the geographic unit is measured at a 1-degree longitude by 1-degree latitude resolution at global scale.

Each pixel contains a gross product value calculated by accounting for economic data merged with other important demographic and geophysical data including climate (i.e. precipitation and temperature), terrain (e.g. elevation, roughness), location indicators, population, and luminosity (Nordhaus. 2006). A full description of the methods and derivation of G-Econ data can be found here https://gecon.yale.edu/sites/default/files/files/gecon_data_20051206_6.pdf. The GCP can be useful not only for economists undertaking research in spatial econometrics but for scientists looking to link satellite and other geographically based data with economic data.

Citation:

Nordhaus, William and Xi Chen. 2008. “Geographically based Economic data”. Accessed August 21, 2018. https://gecon.yale.edu/

Column Name: Growing Season Length

Derived Data Set: Length of Available Growing Period (16 classes)

Derived Data Set Cell Size: 5 Arc Minute (~0.0833333 decimal degrees; ~10 km)

Summary Statistic: Mode

Year: Based on data collected between 1961 and 1991

Units: Individual classes between 1 and 16. The values are listed below.

1: 0 days

2: 1 - 29 days

3: 30 - 59 days

4: 60 - 89 days

5: 90 - 119 days

6: 120 - 149 days

7: 150 - 179 days

8: 180 - 209 days

9: 210 - 239 days

10: 240 - 269 days

11: 270 - 299 days

12: 300 - 329 days

13: 330 - 364 days

14: < 365 days

15: 365 days

16: > 365 days

Description:

The length of the growing season in days (reported in one of 16 categories) for the area within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS survey cluster location.

It is impossible to deep link to the dataset. Searching for “growing season” and limiting results to “World” datasets should bring “Length of Available Growing Period (16 classes)”. The length of available growing period refers to the number of days within the period of temperatures above 5°C when moisture conditions are considered adequate for crop growth. Under rain-fed conditions, the beginning of the growing period is linked to the start of the rainy season. The growing period for most crops continues beyond the rainy season and, to a greater or lesser extent, crops mature on moisture stored in the soil profile.

Citation:

Food and Agriculture Organization. 2007. "Length of Available Growing Period (16 classes)." Accessed August 21, 2017. http://www.fao.org/geonetwork/srv/en/main.home.

Column Name: Irrigation

Derived Data Set: Global Map of Irrigation Areas

Derived Data Set Cell Size: 5 Arc Minute (~0.0833333 decimal degrees; ~10 km)

Summary Statistic: Mean

Year: 2005

Units: Proportion of area equipped for irrigation

Description:

The average proportion of the area within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS survey cluster location equipped for irrigation at the time of measurement.

The agriculture sector is the largest global water user, accounting for about 70% of all water withdrawn worldwide from rivers and aquifers. In several developing countries, irrigation represents up to 95% of all water withdrawn (Siebert et al. 2013). The first version of the Digital Global Map of Irrigated Areas was produced in 1999 and consisted of a raster map of percentage area that was equipped for irrigation at a spatial resolution of 50 x 50 km (Döll and Siebert 1999). However, improved modeling methods and availability of data have made it possible to increase the spatial resolution of the map to 10 x 10 km. Hence, an updated global raster map of irrigation areas was developed by combining sub-national irrigation statistics and geo-spatial information on the location and extent of irrigation schemes (Siebert et al. 2013).

Irrigation data for sub-national units (e.g. districts, counties, provinces, governorates), collected from national census surveys and reports from institutions such as the Food and Agriculture Organization (FAO), the World Bank, and other international organizations were used in the analysis. To distribute the statistics on area equipped for irrigation per subnational unit, digital spatial data layers and printed maps were used. Irrigation maps were derived from project reports, irrigation subsector studies, and books related to irrigation and drainage. These maps were digitized and compared with satellite images of various regions.

Citation:

Stefan Siebert, Verena Henrich, Karen Frenken and Jacob Burke. 2013. “Global Map of Irrigation Areas version 5”. Rheinische Friedrich-Wilhelms-University, Bonn, Germany / Food and Agriculture Organization of the United Nations, Rome, Italy. http://www.fao.org/nr/water/aquastat/irrigationmap/

Column Name: ITN Coverage YEAR

Derived Data Set: ITN coverage in Africa 2000-2015

Derived Data Set Cell Size: 0.0418 decimal degrees (~5 km)

Summary Statistic: Mean

Year: 2005, 2010, or 2015

Units: Number of people

Description:

The average number of people within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS survey cluster location who slept under an insecticide treated net the night before they were surveyed.

Data from the year 2000 were excluded from covariate extraction because all values were 0. 100% net coverage in certain areas in the 2015 data is a product of the underlying dataset and is not an error.

Citation:

Bhatt, S., D. J. Weiss, E. Cameron, D. Bisanzio, B. Mappin, U. Dalrymple, K. E. Battle, et al. 2015. "The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015." *Nature* 526 (7572):207-211. http://doi.org/10.1038/nature15535.

Malaria Atlas Project. 2015. “Insecticide treated bednet coverage.” Accessed August 21, 2017. http://www.map.ox.ac.uk/.

Column Name: Land Surface Temperature YEAR

Derived Data Set: MYD11C3: MODIS/Aqua Land Surface Temperature and Emissivity Monthly

Derived Data Set Cell Size: .05 decimal degrees (~6 km)

Summary Statistic: Mean

Year: 2000, 2005, 2010, 2015

Units: Yearly average temperature in Degrees Celsius

Description:

The average annual land surface temperature within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS survey cluster location.

Land surface temperature (LST) estimates the temperature of the land surface (skin temperature), which is detected by satellites via looking through the atmosphere to the ground. The LST is not equivalent to near-surface air temperature measured by ground stations, and their relationship is complex from a theoretical and empirical perspective (Vancutsem et al. 2010; Yang, Cai, and Yang 2017; Zhou et al. 2012).

This is a global mean land surface temperature grid derived from the MODIS daytime and nighttime measurements in degrees Celsius at a spatial resolution of 0.05° × 0.05°. See section 'Day and Night Land Surface Temp Year' for detailed descriptions.

Citation:

Wan, Zhengming, S. Hook, G. Hulley. 2015. "MYD11C3 MODIS/Aqua Land Surface Temperature/Emissivity Monthly L3 Global 0.05Deg CMG V006". Accessed August 21, 2018. https://doi.org/10.5067/MODIS/MYD11C3.006

Column Name: Livestock Cattle, Livestock Chickens, Livestock Goats, Livestock Pigs, Livestock Sheep

Derived Data Set: Gridded Livestock of the World 2

Derived Data Set Cell Size: 0.00833333 decimal degrees (~1 km)

Summary Statistic: Mean

Year: 2006

Units: Heads of Livestock per square kilometers

Description:

The average density of livestock within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS survey cluster location.

Livestock distribution surface maps were obtained from the Gridded Livestock of the World (GLW) v2.0 database produced in 2007 which provided modeled livestock densities of the world that were adjusted to match the Food and Agriculture Organization (FAO) national estimates for the reference year 2006 (Robinson et al. 2014). The surface maps were produced based on the collection of improved and detailed sub-national livestock data (derived from various national census reports and livestock surveys), new and higher resolution predictor variables, and revised modeling methods that included a more systematic assessment of the model accuracy and the representation of uncertainties associated with the predictions. More details of the modeling method can be found in Wint and Robinson (2007). We downloaded gridded livestock densities at a 1 km spatial resolution for cattle, goats, sheep, pigs and chickens.

Citation:

Robinson TP, Wint GRW, Conchedda G, Van Boeckel TP, Ercoli V, et al. 2014. “Mapping the Global Distribution of Livestock”. *PLoS ONE* 9(5): e96084. https://doi.org/10.1371/journal.pone.0096084

Column Name: Malaria Incidence YEAR

Derived Data Set: Plasmodium Falciparum Incidence Rate 2000-2015

Derived Data Set Cell Size: 0.0418 decimal degrees (~5 km)

Summary Statistic: Mean

Year: 2000, 2005, 2010, or 2015

Units: Number of people per year

Description:

The average number of people per year who show clinical symptoms of *plasmodium falciparum* malaria within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS survey cluster location.

A clinical case is defined as a malaria-attributable febrile episode (body temperature above 37.5 C), typically accompanied by headaches, nausea, excess sweating and/or fatigue, censored by a 30-day window (i.e., multiple bouts of symptoms occurring within the same 30-day period are counted as a single episode).

Citation:

Bhatt, S., D. J. Weiss, E. Cameron, D. Bisanzio, B. Mappin, U. Dalrymple, K. E. Battle, et al. 2015. "The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015." *Nature* 526 (7572):207-211. http://doi.org/10.1038/nature15535.

Malaria Atlas Project. 2015. “Plasmodium falciparum incidence rate.” Accessed August 21, 2017. http://www.map.ox.ac.uk/.

Column Name: Malaria Prevalence YEAR

Derived Data Set: Plasmodium falciparum PR2-10 2000-2015

Derived Data Set Cell Size: 0.0418 decimal degrees (~5 km)

Summary Statistic: Mean

Year: 2000, 2005, 2010, or 2015

Units: *Plasmodium* falciparum parasite rate

Description:

The average parasite rate of *plasmodium falciparum* (*Pf*PR) in children between the ages of 2 and 10 years old within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS survey cluster location.

Citation:

Bhatt, S., D. J. Weiss, E. Cameron, D. Bisanzio, B. Mappin, U. Dalrymple, K. E. Battle, et al. 2015. "The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015." *Nature* 526 (7572):207-211. http://doi.org/10.1038/nature15535.

Malaria Atlas Project. 2015. "*Plasmodium falciparum* PR_{2-10}". Accessed August 31, 2018. http://www.map.ox.ac.uk/.

Column Name: Maximum Temperature YEAR

Derived Data Set: CRU TS v. 4.01

Derived Data Set Cell Size: 0.5 decimal degrees (~55 km)

Summary Statistic: Mean

Year: 2000, 2005, 2010, 2015

Units: Degrees Celsius

Description:

The average annual maximum temperature within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS survey cluster location. The maximum temperature is calculated from the modeled mean temperature and the modeled diurnal temperature range using a formula found in Harris et al. (2014).

The Climate Research Unit (CRU) of the University of East Anglia, UK, produces a range of global climate time series gridded data, which are derived from meteorological stations across the world's land areas. The datasets are provided on high resolution (0.5 x 0.5 degrees) grids over the period 1901- 2016 (Harris and Jones 2017; Harris et al. 2014). The main purpose for which this dataset was constructed was to provide modelers (climate and environmental) with some of the inputs they require to run their models. These data have found extensive use in numerous geospatial modeling and mapping studies (Caminade et al. 2014; Chabot-Couture, Nigmatulina, and Eckhoff 2014; Graetz et al. 2018; Osgood-Zimmerman et al. 2018; Rogers and Randolph 2006).

Citation:

Harris, Ian, P. D. Jones, Timothy J. Osborn, David Lister. 2014. "Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset". *International Journal of Climatology* 43: 623-642. http://dx.doi.org/10.1002/joc.3711

Climate Research Unit. 2017. "CRU TS v. 4.01". Accessed August 21, 2018. http://doi.org/10/gcmcz3

Column Name: Mean Temperature YEAR

Derived Data Set: CRU TS v. 4.01

Derived Data Set Cell Size: 0.5 decimal degrees (~55 km)

Summary Statistic: Mean

Year: 2000, 2005, 2010, 2015

Units: Degrees Celsius

Description:

The average annual temperature within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS survey cluster location. The mean temperature is a modeled surface based on weather station data.

The Climate Research Unit (CRU) of the University of East Anglia, UK, produces a range of global climate time series gridded data, which are derived from meteorological stations across the world's land areas. The datasets are provided on high resolution (0.5 x 0.5 degrees) grids over the period 1901-2016 (Harris and Jones 2017; Harris et al. 2014). The main purpose for which this dataset was constructed was to provide modelers (climate and environmental) with some of the inputs they require to run their models. These data have found extensive use in numerous geospatial modeling and mapping studies (Caminade et al. 2014; Chabot-Couture, Nigmatulina, and Eckhoff 2014; Graetz et al. 2018; Osgood-Zimmerman et al. 2018; Rogers and Randolph 2006).

Citation:

Harris, Ian, P. D. Jones, Timothy J. Osborn, David Lister. 2014. "Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset". *International Journal of Climatology* 43: 623-642. http://dx.doi.org/10.1002/joc.3711

Climate Research Unit. 2017. "CRU TS v. 4.01". Accessed August 21, 2018. http://doi.org/10/gcmcz3

Column Name: Minimum Temperature YEAR

Derived Data Set: CRU TS v. 4.01

Derived Data Set Cell Size: 0.5 decimal degrees (~55 km)

Summary Statistic: Mean

Year: 2000, 2005, 2010, 2015

Units: Degrees Celsius

Description:

The average annual minimum temperature within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS survey cluster location. The minimum temperature is calculated from the modeled mean temperature and the modeled diurnal temperature range using a formula found in Harris et al. (2014).

The Climate Research Unit (CRU) of the University of East Anglia, UK, produces a range of global climate time series gridded data, which are derived from meteorological stations across the world's land areas. The datasets are provided on high resolution (0.5 x 0.5 degrees) grids over the period 1901- 2016 (Harris and Jones 2017; Harris et al. 2014). The main purpose for which this dataset was constructed was to provide modelers (climate and environmental) with some of the inputs they require to run their models. These data have found extensive use in numerous geospatial modeling and mapping studies (Caminade et al. 2014; Chabot-Couture, Nigmatulina, and Eckhoff 2014; Graetz et al. 2018; Osgood-Zimmerman et al. 2018; Rogers and Randolph 2006).

Citation:

Harris, Ian, P. D. Jones, Timothy J. Osborn, David Lister. 2014. "Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset". *International Journal of Climatology* 43: 623-642. http://dx.doi.org/10.1002/joc.3711

Climate Research Unit. 2017. "CRU TS v. 4.01". Accessed August 21, 2018. http://doi.org/10/gcmcz3

Column Name: Night Land Surface Temp YEAR

Derived Data Set: MYD11C3: MODIS/Aqua Land Surface Temperature and Emissivity Monthly

Derived Data Set Cell Size: .05 decimal degrees (~6 km)

Summary Statistic: Mean

Year: 2000, 2005, 2010, 2015

Units: Yearly average temperature in Degrees Celsius

Description:

The average nighttime land surface temperature within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS survey cluster location.

Land surface temperature (LST) estimates the temperature of the land surface (skin temperature), which is detected by satellites via looking through the atmosphere to the ground. The LST is not equivalent to near-surface air temperature measured by ground stations, and their relationship is complex from a theoretical and empirical perspective (Vancutsem et al. 2010; Yang, Cai, and Yang 2017; Zhou et al. 2012).

The global LST [night] grids, represent night time temperature in degree Celsius at a spatial resolution of 0.05° × 0.05°.

Land surface temperature measures the temperature of the ground (see covariate 'Day Land Surface Temp Year' for a detailed description). At night, the land surface typically cools off because it releases its warmth to the air above while no longer receiving sunlight (https://neo.sci.gsfc.nasa.gov/).

Citation:

Wan, Zhengming, S. Hook, G. Hulley. 2015. "MYD11C3 MODIS/Aqua Land Surface Temperature/Emissivity Monthly L3 Global 0.05Deg CMG V006". Accessed August 21, 2018. https://doi.org/10.5067/MODIS/MYD11C3.006

Column Name: Nightlights Composite

Derived Data Set: Version 1 VIIRS Day/Night Band Nighttime Lights

Derived Data Set Cell Size: 15 Arc Second (~0.00416667 decimal degrees; ~500 m)

Summary Statistic: Mean

Year: 2015

Units: Composite cloud-free radiance values

Description:

The average nighttime luminosity of the area within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS survey cluster location.

Nightlights time series data are annual nightlight intensity data provided as raster surfaces with near global coverage. Nightlights measure the luminosity of an area during the nighttime hours as measured by the Visible Infrared Imaging Radiometer Suite (VIIRS). These surfaces allow differentiation of regions based on the density of population and the degree of electrification of dwellings, commercial and industrial premises, and infrastructure (Sutton et al. 2001). For our extractions we used the `vcm-orm-ntl` version of the dataset which removes outliers such as flares from petroleum extraction and other short-duration lights. The background data was also shifted to 0 to account for moonlight.

Citation:

Mills, Stephen, Stephanie Weiss, and Calvin Liang. 2013. “VIIRS day/night band (DNB) stray light characterization and correction.” *Proceedings of SPIE* 8866. http://dx.doi.org/10.1117/12.2023107

National Centers for Environmental Information. 2015. “2015 VIIRS Nighttime Lights Annual Composite.” Accessed August 21, 2017. https://ngdc.noaa.gov/eog/viirs/download_dnb_composites.html.

Column Name: PET YEAR

Derived Data Set: CRU TS v. 4.01

Derived Data Set Cell Size: 0.5 decimal degrees (~55 km)

Summary Statistic: Mean

Year: 2000, 2005, 2010, 2015

Units: Millimeters per year

Description:

The average annual potential evapotranspiration (PET) within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS survey cluster location.

The Climate Research Unit (CRU) of the University of East Anglia, UK, produces a range of global climate time series gridded data, which are derived from meteorological stations across the world's land areas. The datasets are provided on high resolution (0.5 x 0.5 degrees) grids over the period 1901- 2016 (Harris and Jones 2017; Harris et al. 2014). The main purpose for which this dataset was constructed was to provide modelers (climate and environmental) with some of the inputs they require to run their models. These data have found extensive use in numerous geospatial modeling and mapping studies (Caminade et al. 2014; Chabot-Couture, Nigmatulina, and Eckhoff 2014; Graetz et al. 2018; Osgood-Zimmerman et al. 2018; Rogers and Randolph 2006).

Potential evapotranspiration is a synthetic measurement that was calculated using a variation of the Penman-Monteith formula. The full formula can be found in Harris et al. (2014). The formula uses the calculated or modeled numbers from the mean temperature, maximum temperature, minimum temperature, vapor pressure, and cloud cover surfaces. The number shows the number millimeters of water that would be evaporated into the air over the course of a year if there was unlimited water at the location.

Citation:

Harris, Ian, P. D. Jones, Timothy J. Osborn, David Lister. 2014. "Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset". *International Journal of Climatology* 43: 623-642. http://dx.doi.org/10.1002/joc.3711

Climate Research Unit. 2017. "CRU TS v. 4.01". Accessed August 21, 2018. http://doi.org/10/gcmcz3

Column Name: Proximity to National Borders

Derived Data Set: Large-Scale International Boundaries (LSIB)

Summary Statistic: Distance

Year: 2014

Units: Meters

Description:

The geodesic distance to the nearest international borders.

Citation:

Department of State's Office of the Geographer. 2014. "Department of State Large-Scale International Boundary (LSIB)." Via ESRI. Accessed August 21, 2017. http://www.arcgis.com/home/item.html?id=3e650cfe52b84bffacb86d028f1f0514.

Column Name: Proximity to Protected Areas

Derived Data Set: The World Database on Protected Areas

Summary Statistic: Distance

Year: 2017

Units: Meters

Description:

The geodesic distance to the nearest protected area as defined by the United Nations Environment World Conservation Monitoring Centre. Examples of protected places include national parks, national forests, and national seashores. The dataset includes both aquatic and terrestrial protected areas.

Citation:

UNEP-WCMC and IUCN. 2017. “Protected Planet: The World Database on Protected Areas (WDPA)” Accessed August 21, 2017. www.protectedplanet.net.

Column Name: Proximity to Water

Derived Data Set: GSHHG (Global Self-consistent, Hierarchical, High-resolution Geography Database)

Summary Statistic: Distance

Year: 2017

Units: Meters

Description:

The geodesic distance to either a lake or the coastline. For this extraction we used only the lakes dataset (`L2`) at full resolution and the shoreline dataset (`L1`), also at full resolution, in the GSHHG database. The datasets used were based on the World Vector Shorelines, CIA World Data Bank II, and Atlas of the Cryosphere.

Citation:

Wessel, Paul, and Walter Smith. 1996. “A Global Self-consistent, Hierarchical, High-resolution Shoreline Database” *Journal of Geophysical Research* 101: 8741-8743.

Wessel, Paul, and Walter Smith. 2017. “A Global Self-consistent, Hierarchical, High-resolution Geography Database Version 2.3.7.” Accessed August 21, 2017. http://www.soest.hawaii.edu/pwessel/gshhg/.

Column Name: Rainfall YEAR

Derived Data Set: Climate Hazards Group InfraRed Precipitation with Station data 2.0

Derived Data Set Cell Size: 0.05 decimal degrees (~5 km)

Summary Statistic: Mean

Year: 1985, 1990, 1995, 2000, 2005, 2010, 2015

Units: Millimeters per year

Description:

The average annual rainfall within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS survey cluster location.

A high quality, long-term, high-resolution rainfall dataset is key for supporting drought-related risk management, food security early warning, and other related studies (Shukla et al. 2017).

Satellite-based rainfall estimates may provide plentiful information with spatiotemporal high-resolution imagery over widespread regions where conventional rainfall data are scarce or absent (Li et al. 2010). However, previous studies have shown that the estimates have several limitations, because none of the satellite sensors directly detect rainfall and the relationship between observations and precipitation is based on one or numerous proxy variables (Toté et al. 2015; Wu et al. 2012).

A wide variety of satellite-based rainfall products derived from multiple data sources exist at present. The Climate Hazards Group InfraRed Precipitation with Stations (CHIRPS) is a relatively new rainfall product with high temporal and spatial resolution, and is based on multiple data sources (Paredes Trejo et al. 2016).

The CHIRPS dataset builds on previous approaches of interpolation techniques and high resolution, long period of record precipitation estimates based on infrared Cold Cloud Duration (CCD) observations (Funk et al. 2015). The algorithm (i) is built around a 5 km climatology that incorporates satellite information to represent sparsely gauged locations, (ii) incorporates daily, pentadal (five-day), and monthly readings from 1981 to present from CCD-based precipitation estimates, (iii) blends station data to produce a preliminary information product with a latency of about 2 days and a final product with an average latency of about 3 weeks, and (iv) uses a novel blending procedure incorporating the spatial correlation structure of CCD-estimates to assign interpolation weights (Funk et al. 2015; Paredes Trejo et al. 2016).

Citation:

Climate Hazards Group. 2017. “Climate Hazards Group InfraRed Precipitation with Station data 2.0.” Accessed August 21, 2017. http://chg.geog.ucsb.edu/data/chirps/index.html.

Funk, Chris, Pete Peterson, Martin Landsfeld, Diego Pedreros, James Verdin, Shraddhanand Shukla, Gregory Husak, et al. 2015. "The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes". *Scientific Data* 2. http://doi.org/10.1038/sdata.2015.66.

Column Name: Slope

Derived Data Set: See Description

Derived Data Set Cell Size: 30 arc seconds (~0.00833333 decimal degrees; ~1 km)

Summary Statistic: Mean

Year: 1996

Units: Degree

Description:

Slope is a measurement of how rough the terrain around a DHS cluster is. The United States Geological Survey GTOPO30 digital elevation model was processed into slope by using the *slope* tool in ArcMap 10.5.0.

Citation:

Earth Resources Observation and Science Center. 1996. “Global 30 Arc-Second Elevation (GTOPO30).” Accessed August 21, 2017. https://lta.cr.usgs.gov/GTOPO30.

Column Name: SMOD Population YEAR

Derived Data Set: GHS Settlement Model Grid

Derived Data Set Cell Size: 1 km

Summary Statistic: Mode

Year: 1990, 2000, 2015

Units: Individual classes between 1 and 3. The values are listed below:

1: rural cells

2: urban clusters

3: urban centers

Description:

The SMOD urban-rural population classification of the area within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS survey cluster location.

This is an urban-rural settlement classification model adopted by the GHSL. It is the representation of the degree of urbanization concept into the GHSL data scenario. Each grid in the GHS-SMOD has been generated by integrating the GHSL built-up areas and population grids data for reference epochs: 1975, 1990, 2000, and 2015. The degree of urbanization classification schema is a people-based definition of cities and settlements: it operates using as the main input a 1 x 1 km grid cell accounting for population at a given point in time. It discriminates the population grid cells in three main classes: 'urban centers' (cities), 'urban clusters' (towns and suburbs), and 'rural grid cells' (base). These class abstractions translate to 'high density clusters (HDC)', 'low density clusters (LDC)', and 'rural grid cells (RUR)', respectively, in the GHS-SMOD implementation (Pesaresi and Freire 2016).

Citation:

Pesaresi, Martino and Sergio Freire. 2016. “GHS Settlement grid following the REGIO model 2014 in application to GHSL Landsat and CIESIN GPW v4-multitemporal (1975-1990-2000-2015).” Accessed August 21, 2017. http://data.europa.eu/89h/jrc-ghsl-ghs_smod_pop_globe_r2016a.

Column Name: Temperature MONTH

Derived Data Set: WorldClim Version 2

Derived Data Set Cell Size: 30 arc seconds (~0.00833333 decimal degrees; ~1 km)

Summary Statistic: Mean

Year: 1970-2000

Units: Average temperature for months January to December in degrees Celsius

Description:

The average monthly temperature within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS survey cluster location.

Interpolated climate data layers are created by collecting large amounts of weather station data which are used to produce climate maps using a smoothing algorithm. A commonly used interpolated climate data source is WorldClim (Hijmans and Graham 2006). This is a set of data layers of the whole world at a 1km resolution. The variables available are monthly mean, minimum and maximum temperature, monthly precipitation and a set of bioclimatic variables. These maps were created using monthly averaged data over the period 1950-2000 from thousands of weather stations. The data were interpolated using a thin-plate smoothing spline algorithm with altitude, longitude and latitude as independent variables (Hijmans and Graham 2006; Hutchinson 1995).

Citation:

Fick, Steve and Robert Hijmans. 2016. “Worldclim 2.” Accessed August 21, 2017. http://worldclim.org/version2

Fick, Steve and Robert Hijmans. 2017. “Worldclim 2: New 1-km spatial resolution climate surfaces for global land areas.” *International Journal of Climatology*. http://doi.org/10.1002/joc.5086.

Column Name: Travel Times 2000

Derived Data Set: Estimated travel time to the nearest city of 50,000 or more people in year 2000

Derived Data Set Cell Size: 30 arc seconds (~0.00833333 decimal degrees; ~1 km)

Summary Statistic: Mean

Year: 2000

Units: Minutes

Description:

The average time (minutes) required to reach a settlement of 50,000 or more people from the area within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS survey cluster location, based on year 2000 infrastructure data.

Travel times or accessibility estimates provide a useful measure of the extent to which regions are rural or urban, as well as the degree of their connectedness to the national system of transportation. For instance, locations that are near major roads would be relatively well connected, even if they were some distance from major cities (Linard and Tatem 2012).

This estimate is based on the amount of time it takes to reach a settlement of 50,000 or more people in the year 2000.

Citation:

Nelson, Andrew. 2008. “Estimated travel time to the nearest city of 50,000 or more people in year 2000.” Accessed August 21, 2017. http://forobs.jrc.ec.europa.eu/products/gam/.

Column Name: Travel Times 2015

Derived Data Set: Malaria Atlas Project Accessibility to Cities

Derived Data Set Cell Size: ~1 km

Summary Statistic: Mean

Year: 2015

Units: Minutes

Description:

The average time (minutes) required to reach a high-density urban center, as defined by Pesaresi and Freire (2016), from the area within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS survey cluster location, based on year 2015 infrastructure data.

Travel times or accessibility estimates provide a useful measure of the extent to which regions are rural or urban, as well as the degree of their connectedness to the national system of transportation. For instance, locations that are near major roads would be relatively well connected, even if they were some distance from major cities (Linard and Tatem 2012).

The University of Oxford's Malaria Atlas Project (MAP), in collaboration with Google, EU Joint Research Center (JRC) and the University of Twente, has developed a new accessibility estimate (measured in travel time) for 2015 (Weiss et al. 2018). The analysis considered data on recent expansion in infrastructure networks, particularly in lower-resource settings, which are provided in unprecedented detail and precision by OpenStreetMap and Google. Using such data, the MAP developed a global accessibility estimate that quantifies travel time to cities by integrating ten global-scale surfaces that characterize factors affecting human movement rates and 13,840 high-density urban centers within an established geospatial-modelling framework.

Citation:

Weiss. Daniel J., Andy Nelson, Harry S. Gibson, Will Temperley, Stephen Peedell, Allison Lieber, Matthew Hancher, Eduardo Poyart, Simao Belchior, Nancy Fullman, Bonnie Mappin, Ursula Dalrymple, Jennifer Rozier, Timothy C.D. Lucas, Rosalind E. Howes, Lucy S. Tusting, Su Y. Kang, Ewan Cameron, Donal Bisanzio, Katherine E. Battle, Samir Bhatt, and Peter W. Gething. 2018. "A global map of travel time to cities to assess inequalities in accessibility in 2015". *Nature* 553: 333–336. http://doi.org/10.1038/nature25181

Malaria Atlas Project. 2018. "Accessibility to Cities." Accessed August 21, 2018. http://www.map.ox.ac.uk/

Column Name: U5 Population YEAR

Derived Data Set: WorldPop

Derived Data Set Cell Size: 0.000833333 decimal degrees (~100 m)

Summary Statistic: Mean

Year: 2000, 2005, 2010, 2015

Units: Number of People

Description:

The WorldPop average number of people under the age of 5 (U5) within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS survey cluster location.

An understanding of the numbers, characteristics and locations of human populations underpins operational work, policy analyses and scientific development globally across multiple sectors. Most methods for estimating population rely on census data. However, in most countries population censuses are conducted every 10 years at best and can be longer in many low-income countries (Alegana et al. 2015). Thus, data from the census can often be outdated, unreliable and have low granularity (Tatem 2014). New data sources and recent methodological advances made by the WorldPop project (University of Southampton) now provide high resolution, open, and contemporary data on human population distributions allowing accurate measurement of local population distributions, compositions, characteristics, growth and dynamics across national and regional scales. High resolution gridded population distribution estimates of the total number of people per pixel - broken down by male/female and 5-year age groupings - have been created.

Citation:

Gaughan, Andrea E., Forrest R. Stevens, Catherine Linard, Peng Jia, and Andrew J. Tatem. 2013. "High Resolution Population Distribution Maps for Southeast Asia in 2010 and 2015." *PLOS ONE* 8 (2):e55882. http://10.1371/journal.pone.0055882

WorldPop. 2016. “Africa Continental Population Datasets (2000 - 2020).” Accessed August 21, 2018. http://doi.org/10.5258/SOTON/WP00004

WorldPop. 2016. “Asia Continental Population Datasets (2000 – 2020).” Accessed August 21, 2018. http://doi.org/10.5258/SOTON/WP00013

WorldPop. 2016. “Latin America and the Caribbean Continental Population Datasets (2000 - 2020).” Accessed August 21, 2018. http://doi.org/10.5258/SOTON/WP00138

Column Name: UN Population Count YEAR

Derived Data Set: UN-Adjusted Population Count, v4 (2000, 2005, 2010, 2015, 2020)

Derived Data Set Cell Size: 30 arc seconds (~0.00833333 decimal degrees; ~1 km)

Summary Statistic: Sum

Year: 2000, 2005, 2010, 2015

Units: Number of people

Description:

The UN-adjusted count of people within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS survey cluster location.

Gridded Population of the World, Version 4 (GPWv4) consists of high resolution estimates of human population (number of persons per pixel) on a continuous global raster surface, which is consistent with national censuses and population registers, for the years 2000, 2005, 2010, and 2015. The files for this dataset were produced at a spatial resolution of 1 x 1 km.

The GPW is quantified by population data (i.e. tabular counts of population listed by administrative area) and spatially explicit administrative boundary data, and use an areal-weighting method (also known as uniform distribution or proportional allocation) to disaggregate population from census units into grid cells through the simple assumption that the population of a grid cell is an exclusive function of the land area within that pixel (Doxsey-Whitfield et al. 2015).

Citation:

Center for International Earth Science Information Network - Columbia University. 2016. "Gridded Population of the World, Version 4 (GPWv4): Population Count Adjusted to Match 2015 Revision of UN WPP Country Totals." Accessed August 21, 2017. http://dx.doi.org/10.7927/H4SF2T42.

Column Name: UN Population Density YEAR

Derived Data Set: UN-Adjusted Population Density, v4 (2000, 2005, 2010, 2015, 2020)

Derived Data Set Cell Size: 30 arc seconds (~0.00833333 decimal degrees; ~1 km)

Summary Statistic: Mean

Year: 2000, 2005, 2010, 2015

Units: Number of people per square kilometer

Description:

The average UN-adjusted population density of the area within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS survey cluster location.

Gridded Population of the World, Version 4 (GPWv4) consists of high resolution estimates of human population density (number of persons per square kilometer) based on counts consistent with national censuses and population registers, for the years 2000, 2005, 2010, and 2015. A proportional allocation gridding algorithm, utilizing approximately 13.5 million national and sub-national administrative units, was used to assign population counts to ~1 km grid cells.

The population density rasters were created by dividing the population count raster for a given target year by the land area raster.

Citation:

Center for International Earth Science Information Network - Columbia University. 2016. "Gridded Population of the World, Version 4 (GPWv4): Population Density Adjusted to Match 2015 Revision UN WPP Country Totals." Accessed August 21, 2017. http://dx.doi.org/10.7927/H4HX19NJ.

Column Name: Wet Days YEAR

Derived Data Set: CRU TS v. 4.01

Derived Data Set Cell Size: 0.5 decimal degrees (~55 km)

Summary Statistic: Mean

Year: 2000, 2005, 2010, 2015

Units: Number of Days

Description:

The average number of days receiving rainfall within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS survey cluster location.

The Climate Research Unit (CRU) of the University of East Anglia, UK, produces a range of global climate time series gridded data, which are derived from meteorological stations across the world's land areas. The datasets are provided on high resolution (0.5 x 0.5 degrees) grids over the period 1901- 2016 (Harris and Jones 2017; Harris et al. 2014). The main purpose for which this dataset was constructed was to provide modelers (climate and environmental) with some of the inputs they require to run their models. These data have found extensive use in numerous geospatial modeling and mapping studies (Caminade et al. 2014; Chabot-Couture, Nigmatulina, and Eckhoff 2014; Graetz et al. 2018; Osgood-Zimmerman et al. 2018; Rogers and Randolph 2006).

Wet days, or the number of days a year that received rainfall is a partly-synthetic measurement. It combines the number of observed days with rainfall from weather stations with the number of days that should have received rainfall using a formula found in either New, Hulme, and Jones (2000) or Harris *et al* (2014).

Citation:

Harris, Ian, P. D. Jones, Timothy J. Osborn, David Lister. 2014. "Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset". *International Journal of Climatology* 43: 623-642. http://dx.doi.org/10.1002/joc.3711

Climate Research Unit. 2017. "CRU TS v. 4.01". Accessed August 21, 2018. http://doi.org/10/gcmcz3

REFERENCES

Alegana, V. A., P. M. Atkinson, C. Pezzulo, A. Sorichetta, D. Weiss, T. Bird, E. Erbach-Schoenberg, and A. J. Tatem. 2015. "Fine Resolution Mapping of Population Age-Structures for Health and Development Applications." *J R Soc Interface* 12 (105). http://rsif.royalsocietypublishing.org/content/12/105/20150073.abstract.

Allen, R. G., L. S. Pereira, D. Raes, and M. Smith. 1998. "Crop Evapotranspiration-Guidelines for Computing Crop Water Requirements-Fao Irrigation and Drainage Paper 56." *Fao, Rome* 300 (9): D05109.

Arnold, E. 1997. *World Atlas of Desertification, Unep*. 2nd ed. London, UK: Routledge.

Burgert, C. R., J. Colston, T. Roy, and B. Zachary. 2013. *Geographic Displacement Procedure and Georeferenced Data Release Policy for the Demographic and Health Surveys*. DHS Spatial Analysis Reports No. 7. Calverton, Maryland, USA: ICF International. http://dhsprogram.com/pubs/pdf/SAR7/SAR7.pdf.

Caminade, C., S. Kovats, J. Rocklov, A. M. Tompkins, A. P. Morse, F. J. Colón-González, H. Stenlund, P. Martens, and S. J. Lloyd. 2014. "Impact of Climate Change on Global Malaria Distribution." *Proceedings of the National Academy of Sciences* 111 (9): 3286-3291. http://www.pnas.org/content/pnas/111/9/3286.full.pdf.

Chabot-Couture, G., K. Nigmatulina, and P. Eckhoff. 2014. "An Environmental Data Set for Vector-Borne Disease Modeling and Epidemiology." *PLOS ONE* 9 (4): e94741. https://doi.org/10.1371/journal.pone.0094741.

Dilley, M., R. S. Chen, U. Deichmann, A. Lerner-Lam, M. Arnold, J. Agwe, P. Buys, et al. 2005. *Natural Disaster Hotspots: A Global Risk Analysis*. Washington, D.C.: The World Bank. http://go.worldbank.org/PT8XJZW3K0.

Döll, P., and S. Siebert. 1999. *A Digital Global Map of Irrigated Areas*. University of Kassel, Germany.

Doxsey-Whitfield, E., K. MacManus, S. B. Adamo, L. Pistolesi, J. Squires, O. Borkovska, and S. R. Baptista. 2015. "Taking Advantage of the Improved Availability of Census Data: A First Look at the Gridded Population of the World, Version 4." *Papers in Applied Geography* 1 (3): 226-234. https://doi.org/10.1080/23754931.2015.1014272.

ESRI. 2017. *Arcgis Desktop* (10.6). Environmental Systems Research Institute.

Funk, C., P. Peterson, M. Landsfeld, D. Pedreros, J. Verdin, S. Shukla, G. Husak, et al. 2015. "The Climate Hazards Infrared Precipitation with Stations - a New Environmental Record for Monitoring Extremes." *Scientific Data* 2 (150066). http://dx.doi.org/10.1038/sdata.2015.66.

Graetz, N., J. Friedman, A. Osgood-Zimmerman, R. Burstein, M. H. Biehl, C. Shields, J. F. Mosser, et al. 2018. "Mapping Local Variation in Educational Attainment across Africa." *Nature* 555: 48. http://dx.doi.org/10.1038/nature25761.

Harris, I., P. D. Jones, T. J. Osborn, and D. H. Lister. 2014. "Updated High-Resolution Grids of Monthly Climatic Observations - the Cru Ts3.10 Dataset." *International Journal of Climatology* 34: 623-642.

Harris, I. C., and P. D. Jones. 2017. *Cru Ts4.00: Climatic Research Unit (Cru) Time-Series (Ts) Version 4.00 of High-Resolution Gridded Data of Month-by-Month Variation in Climate (Jan. 1901- Dec. 2015)*. Centre for Environmental Data Analysis. http://dx.doi.org/10.5285/edf8febfdaad48abb2cbaf7d7e846a86.

Hijmans, R. J., and C. H. Graham. 2006. "The Ability of Climate Envelope Models to Predict the Effect of Climate Change on Species Distributions." *Global Change Biology* 12 (12): 2272-2281.

Hutchinson, M. F. 1995. "Interpolating Mean Rainfall Using Thin Plate Smoothing Splines." *International Journal of Geographical Information Systems* 9 (4): 385-403.

Li, J. G., H. X. Ruan, J. R. Li, and S. F. Huang. 2010. "Application of Trmm Precipitation Data in Meteorological Drought Monitoring." *Journal of China Hydrology* 30 (4): 43-46.

Li, Y., A. Feng, W. Liu, X. Ma, and G. Dong. 2017. "Variation of Aridity Index and the Role of Climate Variables in the Southwest China." *Water* 9 (10): 743.

Linard, C., and A. J. Tatem. 2012. "Large-Scale Spatial Population Databases in Infectious Disease Research." *International journal of health geographics* 11 (1): 7.

New, M., M. Hulme, and P. D. Jones. 2000. "Representing Twentieth Century Space-Time Climate Variability. Ii: Development of 1901–1996 Monthly Grids of Terrestrial Surface Climate." *Journal of Climate* 13: 2217–2238.

Nordhaus, W. D. 2006. "Geography and Macroeconomics: New Data and New Findings." *Proceedings of the National Academy of Sciences of the United States of America* 103 (10): 3510-3517. http://www.pnas.org/content/pnas/103/10/3510.full.pdf.

Osgood-Zimmerman, A., A. I. Millear, R. W. Stubbs, C. Shields, B. V. Pickering, L. Earl, N. Graetz, et al. 2018. "Mapping Child Growth Failure in Africa between 2000 and 2015." *Nature* 555: 41. http://dx.doi.org/10.1038/nature25760.

Paredes Trejo, F. J., H. A. Barbosa, M. A. Peñaloza-Murillo, M. Alejandra Moreno, and A. Farías. 2016. "Intercomparison of Improved Satellite Rainfall Estimation with Chirps Gridded Product and Rain Gauge Data over Venezuela." edited by Editor, 323-342.

Perry, M. 2016. *Rasterstats* (0.10.1). https://github.com/perrygeo/python-rasterstats/releases.

Pesaresi, M., D. Ehrlich, S. Ferri, A. Florczyk, S. Freire, M. Halkia, A. Julea, et al. 2016. "Operating Procedure for the Production of the Global Human Settlement Layer from Landsat Data of the Epochs 1975, 1990, 2000, and 2014." *Publications Office of the European Union*.

Pesaresi, M., and S. Freire. 2016. "Ghs Settlement Grid Following the Regio Model 2014 in Application to Ghsl Landsat and Ciesin Gpw V4-Multitemporal (1975-1990-2000-2015)." *European Commission, Joint Research Centre, JRC Data Catalogue*.

Robinson, T. P., G. R. W. Wint, G. Conchedda, T. P. Van Boeckel, V. Ercoli, E. Palamara, G. Cinardi, et al. 2014. "Mapping the Global Distribution of Livestock." *PLoS ONE* 9 (5): e96084. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4038494/.

Rogers, D. J., and S. E. Randolph. 2006. "Climate Change and Vector-Borne Diseases." *Advances in Parasitology* 62: 345-381.

Salem, B. B. 1989. *Arid Zone Forestry: A Guide for Field Technicians*: Food and Agriculture Organization (FAO).

Shukla, S., C. Funk, P. Peterson, A. McNally, T. Dinku, H. Barbosa, F. Paredes-Trejo, D. Pedreros, and G. Husak. 2017. "The Climate Hazards Group Infrared Precipitation with Stations (Chirps) Dataset and Its Applications in Drought Risk Management." In *EGU General Assembly Conference Abstracts, 2017,* 11498.

Siebert, S., V. Henrich, K. Frenken, and J. Burke. 2013. *Update of the Global Map of Irrigation Areas Version 5*. Rome, Italy: FAO.

Sutton, P., D. Roberts, C. Elvidge, and K. Baugh. 2001. "Census from Heaven: An Estimate of the Global Human Population Using Night-Time Satellite Imagery." *International Journal of Remote Sensing* 22 (16): 3061-3076.

Tatem, A. J. 2014. "Mapping the Denominator: Spatial Demography in the Measurement of Progress." *International Health* 6 (3): 153-155. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4161992/pdf/ihu057.pdf.

Tereshchenko, I., A. N. Zolotokrylin, E. A. Cherenkova, C. O. Monzón, L. Brito-Castillo, and T. B. Titkova. 2015. "Changes in Aridity across Mexico in the Second Half of the Twentieth

Century." *Journal of Applied Meteorology and Climatology* 54 (10): 2047-2062. https://doi.org/10.1175/JAMC-D-14-0207.1.

Toté, C., D. Patricio, H. Boogaard, R. van der Wijngaart, E. Tarnavsky, and C. Funk. 2015. "Evaluation of Satellite Rainfall Estimates for Drought and Flood Monitoring in Mozambique." *Remote Sensing* 7 (2): 1758-1776.

Vancutsem, C., P. Ceccato, T. Dinku, and S. J. Connor. 2010. "Evaluation of Modis Land Surface Temperature Data to Estimate Air Temperature in Different Ecosystems over Africa." *Remote Sensing of Environment* 114 (2): 449-465.

Weiss, D. J., A. Nelson, H. S. Gibson, W. Temperley, S. Peedell, A. Lieber, M. Hancher, et al. 2018. "A Global Map of Travel Time to Cities to Assess Inequalities in Accessibility in 2015." *Nature* 553: 333. http://dx.doi.org/10.1038/nature25181.

Wint, G. R. W., and T. P. Robinson. 2007. *Gridded Livestock of the World 2007*. Rome: FAO.

Wu, H., R. F. Adler, Y. Hong, Y. Tian, and F. Policelli. 2012. "Evaluation of Global Flood Detection Using Satellite-Based Rainfall and a Hydrologic Model." *Journal of Hydrometeorology* 13 (4): 1268-1284.

Yang, Y. Z., W. H. Cai, and J. Yang. 2017. "Evaluation of Modis Land Surface Temperature Data to Estimate near-Surface Air Temperature in Northeast China." *Remote Sensing* 9 (5): 410.

Zhou, S.-s., S.-s. Zhang, J.-j. Wang, X. Zheng, F. Huang, W.-d. Li, X. Xu, and H.-w. Zhang. 2012. "Spatial Correlation between Malaria Cases and Water-Bodies in Anopheles Sinensis Dominated Areas of Huang-Huai Plain, China." *Parasites & Vectors* 5: 106-106. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3414776/.