PRIO-GRID v.2.0 Codebook

Andreas Forø Tollefsen, Karim Bahgat, Jonas Nordkvelle and Halvard Buhaug

June 2, 2016

Version Information: This codebook describes the content and development of the PRIO-GRID version 2.0.

Citation: PRIO-GRID is a unique data framework developed at PRIO and made freely available to all interested users. Whenever using the PRIO-GRID data frame or parts of its content, please cite:

> Tollefsen, Andreas Forø; Håvard Strand & Halvard Buhaug (2012) PRIO-GRID: A unified spatial data structure. *Journal of Peace Research*, 49(2): 363-374. doi: 10.1177/0022343311431287

The article is *open access* and freely available for download at the journal's web page.

PRIO-GRID consists of data from multiple third-party sources. Hence, users are requested to cite the original source for each variable used in their work, in addition to citing the PRIO-GRID article. See variable descriptions in this codebook for each variable for the correct citation).

If you want to cite this codebook, please use:

> Tollefsen, Andreas Forø, Karim Bahgat, Jonas Nordkvelle and Halvard Buhaug (2015). *PRIO-GRID v.2.0 Codebook*. Peace Research Institute Oslo.

Funding: PRIO-GRID was initiated by Andreas Forø Tollefsen and Halvard Buhaug in 2008. The project was an integral part of the Advanced Conflict Data Catalogue (ACDC) project (2011-2013), led by Håvard Strand, and funded by the Research Council of Norway. The upgrade to PRIO-GRID version 2.0 has been funded by the Research Council of Norway and the European Research Council through separate research project grants (#240315-F10 and #648291, respectively), led by Halvard Buhaug (PRIO).

Acknowledgements: We thank numerous colleagues at PRIO, Uppsala University, ENCoRe Cost Network, ETH Zürich, University of Colorado Boulder, CIESIN Columbia University, participants at FOSS4G 2011 conference in Denver, and users around the world for crucial feedback during the development of the PRIO-GRID project. Gerdis Wischnath, Johan Dittrich Hallberg, and Nils Weidmann provided important input on the initial version.

We also appreciate the cooperation with providers of the source data that are integrated into the PRIO-GRID framework. Please do cite the sources as described in relation to each variable.

Questions and Support: The PRIO-GRID website can be found at: grid.prio.org. The website provides an interactive visualization of the PRIO-GRID variables and allows users to download variables of interest as spreadsheets or shapefiles through the data portal.

Questions and comments should be addressed to Andreas Forø Tollefsen: andreas@prio.org.

Chapter 1

Introduction

This document describes the development and content of the PRIO-GRID dataset, a standardized spatial grid structure with global coverage at a resolution of 0.5 x 0.5 decimal degrees. See Tollefsen, Strand & Buhaug (2012) for additional information on the background, motivation, and application of PRIO-GRID.

PRIO-GRID consists of four components. The first is the tabular dataset, containing spatially disaggregated data at the grid cell level. While these tables do not contain the geometries per se, they can be represented and visualized using the PRIO-GRID geographic information systems (GIS) shapefiles, which contain the polygon grid and the corresponding cell centroids. Two .csv files are available for download; one .csv table for static variables and one .csv table for time-varying variables where the grid has one realization per calendar year. The content of the .csv files depends on the variables selected for download through the data portal.

The second component includes open-source replication scripts that were used to generate the PRIO-GRID dataset, publically available through the GitHub-repository at github.com/prio-data/priogrid. These files facilitate replication, modification, and extension of the original files, including joining of additional geo-referenced data, should the user wish to do so.

The third component is the documentation, consisting of the journal article presenting PRIO-GRID (Tollefsen, Strand & Buhaug 2012), this codebook, and the instructions for how to use the replication files at GitHub.

The fourth component is a shapefile of the grid-cells. The shapefile is a 0.5 x 0.5 decimal degree grid system of the world. It contains both land and sea grids. When merged with the data, you can use this shapefile to plot data on a map.

PRIO-GRID is a versioned dataset, meaning that changes to the data are released with new version numbers. Higher version numbers indicate more recent data. All files, scripts, and documentation should reflect these version changes.

Changes in version 2.0

PRIO-GRID version 2.0 introduces several updates, changes, and new additions since version 1.01. PRIO-GRID 2.0 extends the temporal coverage until 2014, providing one annual grid representation of the globe for each year, 1946-2014. While previous PRIO-GRID versions only contained grid cells occupied by independent states as defined by the Gleditsch & Ward

system membership list, version 2.0 contains all terrestrial cells, in total 64,818 per grid, for all years regardless of the political status of the territory. For grid cells covering non-independent territory, the country code (gwno) will be missing. This is done to facilitate inclusion of time-series data that are not associated with a country, notably climate statistics.

List of changes

- The data structure has been refactored for easier use. Version 2.0 consists of one static table, one yearly table, and one shapefile with the corresponding geometries. The content of the tables will be determined by the user when generating the data through the data portal.
- Duplicates of cell-year observations (previously done to permit info on overlapping ethnic groups) have been removed. Instead, link tables, such as GeoEPR2PRIO-GRID, are provided as an extension.
- A new interactive data portal has been created where PRIO-GRID data can be visualized, queried, and downloaded.
- The development process has been made fully automated and replicable in a downloadable package, along with instructions for how to use. This includes converting from a mixed use of Python and SQL scripts to nearly pure SQL scripts, for less dependencies and easier replication. Any replication will require a PostGIS compatible database. The scripts used to create PRIO-GRID have been made available as open-source files through the GitHub data repository.
- In addition to general updating of time-varying variables to cover more recent years and minor adjustments in some variable operationalizations (see data description below for details), several new indicators have been added, and some have been removed. More specifically:
  - The Conflict Sites and onset data featured in v.1.01 are not included in v.2.0 since the underlying Conflict Site dataset has not been updated after 2008. Please refer back to v.1.01 for these data (merge using gid-year). Note that the UCDP Georeferenced Event Data, UCDP conflict polygons, and the ACLED datasets provide links to PRIO-GRID cell IDs for each event and can thus be easily imported into the grid.
  - All distance measurements are now spherical distances, rather than geometric distances.
  - Rather than listing all GeoEPR groups within a cell, v.2.0 includes a count of excluded [Original data] groups within a gid-year. A new link table, GeoEPR2PRIO-GRID, is available on the PRIO-GRID web as an extension for users who want to import additional information from the EPR family datasets.
  - Globcover landuse coverages are now separated into variables (urban_gc [Original data], agri_gc [Original data], forest_gc [Original data], shrub_gc [Original data], herb_gc [Original data], aquaveg_gc [Original data], barren_gc [Original data], water_gc [Original data]), rather than using the combination of **lclass** and **lclasspct** in PRIO-GRID v.1.01. Users wishing to aggregate their own combinations of landuse types should consult the old v.1.01.
  - Data on irrigation (irrig_ [Original data]) is now based on a new data source that also captures changes over time.
  - Precipitation data (prec_gpcc [Original data] and prec_gpcp [Original data]) are now derived from two alternative sources, GPCC and GPCP, rather than the University of Delaware (NOAA 2011) data provided through PRIO-GRID v.1.01.

- Temperature data (temp [Original data]) are now derived from GHCN/CAMS (Fan et.al. 2008), rather than the University of Delaware (NOAA 2011) data provided through PRIO-GRID v.1.01.
- A number of new drought measures have been added.
- Data on the location of diamond, petroleum, gems, gold, and drugs deposits have been added.
- Data on child malnutrition has been added (CIESIN CMR).
- Satellite Nightlight emission data have been added (DMSP-OLS).
- Crop and landuse data from MIRCA2000 and ISAM-HYDE have been added.
- Population data from HYDE have been added to supplement the GPW data.

The development of PRIO-GRID

PRIO-GRID is generated in a relational database management system (RDBMS); PostgreSQL with the spatial PostGIS extension supplying the geometric functionality of the Structured Query Language (SQL) database. PRIO-GRID is released with a 0.5 x 0.5 decimal degree cell resolution. This corresponds to a cell of roughly 55 x 55 kilometers at the Equator (3025 square kilometers area). Cell area decreases at higher latitudes due to the curvature of the earth.

The grid structure is defined by a south-western starting point defined by x and y coordinates (90S and 180W) and represented using the WGS84 geographic reference system. The cell identifier starts at 1 at the south-western corner (column 1 and row 1) and increases by 1 for each column, until reaching 720 (column 720 and row 1). The cell identifier then starts at the next row and begins at 721 (column 1 row 2). The full grid at 0.5 x 0.5 degrees resolution contains 259,200 cells (720 x 360). A majority of these cells cover water and other uninhabited areas (notably the Arctic and Antarctica) and are of little relevance in most applications. To limit file size, the released PRIO-GRID only includes terrestrial grid cells (64,818) although the full grid is maintained and is available on request. The current version of PRIO-GRID consists of one grid per calendar year for the period 1946–2014.

The remaining sections of the codebook contain a brief presentation of all variables in the PRIO-GRID files and how they were imported and modified to fit into the PRIO-GRID data structure.

The grid reference file

The grid reference file contains information about the PRIO-GRID spatial data structure. This file is provided in the ESRI shapefile format, where each cell is represented by a rectangular vector geometry in addition to a shapefile containing the centroid point. Variables from the static and temporal files can be visualized and analyzed by merging data to the shapefile via the grid identifier (see below).

Adding additional data using the provided shapefile

In addition to the data available in the tables explained above, we provide a shapefile with the cell geometry that make it possible for users to add their own data. This file may be used in a

GIS software to extract, join or overlay with other spatial data. The shapefile may be joined to the various attribute tables using the gid variable.

Using the replication scripts

PRIO-GRID aims to be transparent and is fully replicable with a set of automated script. The necessary files and instructions can be found at the GitHub repository at github.com/prio-data/priogrid.

Chapter 2

Overview of Included Data

This section presents the data available through the PRIO-GRID, and descriptions of the variable names. Below each source of data is a reference to the appropriate citation for each data source. Please cite the original source in addition to the JPR article and this codebook whenever using PRIO-GRID.

Grid Cell Identifiers

gid is the grid cell identifier, a unique id code for each cell in the grid. Since we only include the terrestrial cells from the full grid, the gid starts at 49182 and ends at 249344. See The development of PRIO-GRID for explanation of the grid structure.

col denotes column number for the grid cell. Column 1 is the westernmost column in the grid, between 180 and 179.5 decimal degrees W. With one column per half degree, there are 720 columns in PRIO-GRID.

row denotes the row number for the grid cell. Row 1 is the southernmost row (between 90 and 89.5 degrees S) and row 360 is the northernmost row in the full grid in the underlying data.

xcoord denotes the longitude coordinate (decimal degrees) for the centroid of the grid cell. Negative coordinates are located west of the Prime Meridian (Greenwich) at 0 degrees longitude.

ycoord denotes the latitude coordinate (decimal degrees) for the centroid of the grid cell. Negative coordinates are located south of the Equator at 0 degrees latitude.

The Static Table

The static table contains a grid cell identifier (gid). This means that gid constitutes the unique identifier in the static file. The PRIO-GRID static table contains observations of all terrestrial grid cells (based on cShapes, thus excluding Antarctica, Greenland, and several smaller island states). In total, the table contains 64,818 observations.

gid is the grid cell identifier, a unique id code for each cell in the grid.

landarea gives the total area covered by land in the grid cell in square kilometers as defined by the CShapes dataset. Hence, we exclude Antarctica, Greenland, and several smaller island states. Areas are calculated assuming that the earth in an oblate spheroid (WGS 84).

Please cite:

> Weidmann, Nils B; Doreen Kuse & Kristian Skrede Gleditsch (2010) The geography of the international system: The CShapes Dataset. *International Interactions*, 36(1): 86-106.

Accessibility variables

ttime_ [**Original data**] is an estimate of the travel time to the nearest major city, derived from a global high-resolution raster map of accessibility developed for the EU. The original indicator is a result of network analysis using a combination of several sources, most collected between 1990 and 2005. The original pixel value is the estimated travel time in minutes by land transportation from the pixel to the nearest major city with more than 50 000 inhabitants.

- **ttime_mean** gives the average travel time within each cell.
- **ttime_sd** gives the standard deviation of original pixel values within each cell.
- **ttime_min** gives the minimum original pixel value within each cell.
- **ttime_max** gives the maximum original pixel value within each cell.

Please cite:

> Uchida, Hirotsugu and Nelson, Andrew (2009). Agglomeration Index: Towards a New Measure of Urban Concentration. *Background paper for the World Bank's World Development Report 2009*.

mountain_mean [Original data] measures the proportion of mountainous terrain within the cell based on elevation, slope and local elevation range, taken from a high-resolution mountain raster developed for UNEP's Mountain Watch Report. The original pixel values are binary, capturing whether the pixel is a mountain pixel or not based on the seven different categories of mountainous terrain in the report.

Please cite:

> Blyth, Simon, Brian Groombridge, Igor Lysenko, Lera Miles, and Adrian Newton (2002). *Mountain Watch: environmental change & sustainable development in mountains.* UNEP-WCMC Biodiversity Series 12. ISBN: 1-899628-20-7

Socioeconomic variables

imr_ [Original data] measures infant mortality rate, based on raster data from the SEDAC Global Poverty Mapping project. The original pixel value is the number of children per 10,000 live births that die before reaching their first birthday. This indicator is a snapshot for the year 2000 only.

- **imr_mean** gives the average infant mortality rate within the grid cell.
- **imr_sd** gives the standard deviation of original pixel values within each cell.
- **imr_min** gives the minimum of original pixel values within each cell.
- **imr_max** gives the maximum of original pixel values within each cell.

Please cite:

> Storeygard, Adam; Deborah Balk, Marc Levy & Glenn Deane (2008) The global distribution of infant mortality: A subnational spatial view. *Population, Space and Place*, 14(3):209-229.
>
> Center for International Earth Science Information Network - CIESIN - Columbia University. 2005. *Poverty Mapping Project: Global Subnational Infant Mortality Rates*. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). doi:10.7927/H4PZ56R2. Accessed 19.05.2006.

cmr_ [Original data] measures the prevalence of child malnutrition, based on raster data from the SEDAC Global Poverty Mapping project. The original pixel value is the percent of children under the age of 5 that are malnutritioned. This indicator is a snapshot for the year 2000 only.

- **cmr_mean** gives the average prevalence of child malnutrition within the grid cell.
- **cmr_sd** gives the standard deviation of original pixel values within each cell.
- **cmr_min** gives the minimum of original pixel values within each cell.
- **cmr_max** gives the maximum of original pixel values within each cell.

Please cite:

> Center for International Earth Science Information Network - CIESIN - Columbia University. 2005. *Poverty Mapping Project: Global Subnational Prevalence of Child Malnutrition*. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). doi:10.7927/H4K64G12. Accessed 13.08.2015.

Resource variables

petroleum_s [Original data] is a dummy variable for whether onshore petroleum deposits have been found within the given grid cell, based on the Petroleum Dataset v.1.2. This variable only codes those petroleum deposits that do not have a known discovery or start of production year. For a complete picture, these data must therefore be combined with the petroleum_y [Original data] data.

Please cite:

> Lujala, Päivi, Jan Ketil Rød & Nadia Thieme, 2007. Fighting over Oil: Introducing A New Dataset. *Conflict Management and Peace Science*, 24(3), 239-256.

diamsec_s [Original data] is a dummy variable for whether secondary (alluvial) diamond deposits have been found within the given grid cell, based on the Diamond Resources dataset v1a. This variable only codes those deposits that do not have a known discovery or start of production year. For a complete picture, these data must therefore be combined with the diamsec_y [Original data] data.

Please cite:

> Gilmore, Elisabeth, Nils Petter Gleditsch, Päivi Lujala & Jan Ketil Rød, 2005. Conflict Diamonds: A New Dataset, *Conflict Management and Peace Science* 22(3): 257–292
>
> Lujala, Päivi, Nils Petter Gleditsch & Elisabeth Gilmore, 2005. A Diamond Curse? Civil War and a Lootable Resource. *Journal of Conflict Resolution*, 49(4): 538–562.

diamprim_s [Original data] is a dummy variable for whether primary (kimberlite) diamond deposits have been found within the given grid cell, based on the Diamond Resources dataset v1a. This variable only codes those deposits that do not have a known discovery or start of production year. For a complete picture, these data must therefore be combined with the diamprim_y [Original data] data.

Please cite the same source as diamsec_s [Original data].

goldplacer_s [Original data] is a dummy variable for whether placer gold deposits have been found within the given grid cell, based on the GOLDATA_L subset of the GOLDDATA v1.2. This variable only codes those deposits that do not have a known discovery or start of production year. For a complete picture, these data must therefore be combined with the goldplacer_y [Original data] data.

Please cite:

> Balestri, Sara, 2015. GOLDATA: The Gold deposits dataset codebook, Version 1.2. *UCSC-Cognitive Science and Communication Research Centre* WP 02/15, Milan. doi:10.13140/RG.2.1.1730.8648
>
> Balestri, Sara, 2012. Gold and civil conflict intensity: evidence from a spatially disaggregated analysis, Peace Economics. *Peace Science and Public Policy*, 18(3): 1-17. doi:10.1515/peps-2012-0012.

goldsurface_s [Original data] is a dummy variable for whether surface gold deposits have been found within the given grid cell, based on the GOLDATA_S subset of the GOLDDATA v1.2. Surface gold deposits are defined as deposits that are located near the surface but "do not hold enough information to be properly defined as lootable [placer gold]". This variable only codes those deposits that do not have a known discovery or start of production year. For a complete picture, these data must therefore be combined with the goldsurface_y [Original data] data.

Please cite the same source as goldplacer_s [Original data].

goldvein_s [Original data] is a dummy variable for whether vein gold deposits have been found within the given grid cell, based on the GOLDATA_NL subset of the GOLDDATA v1.2. This variable only codes those deposits that do not have a known discovery or start of production year. For a complete picture, these data must therefore be combined with the goldvein_y [Original data] data.

Please cite the same source as goldplacer_s [Original data].

gem_s [Original data] is a dummy variable for whether gem deposits have been found within the given grid cell, based on the GEMDATA dataset. This variable only codes those deposits that do not have a known discovery or start of production year. For a complete picture, these data must therefore be combined with the gem_y [Original data] data.

Please cite:

> Lujala, Päivi 2009. Deadly Combat over Natural Resources: Gems, Petroleum, Drugs, and the Severity of Armed Civil Conflict. *Journal of Conflict Resolution*, 53(1): 50-71.

Landuse variables

urban_gc [Original data] measures the coverage of urban areas in each cell, based on the Globcover 2009 dataset v.2.3. To compute urban_gc [Original data] we follow the FAO land cover classification system used by Globcover and aggregate to the category "Artificial areas" (Landuse class 190). The value indicates the percentage area of the cell covered by urban area. This indicator is a snapshot for the year 2009 only.

Please cite:

> Bontemps, Sophie; Pierre Defourny & Eric Van Bogaert (2009) Globcover 2009. Products Description and Validation Report. *European Space Agency*. (http://due.esrin.esa.int/files/GLOBCOVER2009_Validation_Report_2.2.pdf).

agri_gc [Original data] measures the coverage of agricultural areas in each cell, extracted from the Globcover 2009 dataset v.2.3. To compute agri_gc [Original data] we follow the FAO land cover classification system used by Globcover and aggregate to the category "Cultivated terrestrial areas and managed lands" (landuse classes 11, 14, 20, 30). The value indicates the percentage area of the cell covered by agricultural area. This indicator is a snapshot for the year 2009 only.

Please cite:

Bontemps, Sophie; Pierre Defourny & Eric Van Bogaert (2009) Globcover 2009. Products Description and Validation Report. *European Space Agency*. (http://due.esrin.esa.int/files/GLOBCOVER2009_Validation_Report_2.2.pdf).

forest_gc [Original data] measures the coverage of forest areas in each cell, extracted from the Globcover 2009 dataset v.2.3. To compute forest_gc [Original data] we follow the FAO land cover classification system used by Globcover and aggregate to the category "Woody - trees"(landuse classes 40, 50, 60, 70, 80, 90, 100, 110, 120). The value indicates the percentage area of the cell covered by forested area. This indicator is a snapshot for the year 2009 only.

Please cite:

Bontemps, Sophie; Pierre Defourny & Eric Van Bogaert (2009) Globcover 2009. Products Description and Validation Report. *European Space Agency*. (http://due.esrin.esa.int/files/GLOBCOVER2009_Validation_Report_2.2.pdf).

shrub_gc [Original data] measures the coverage of shrubland in each cell, extracted from the Globcover 2009 dataset v.2.3. To compute shrub_gc [Original data] we follow the FAO land cover classification system used by Globcover and aggregate to the category "Shrub" (landuse class 130). The value indicates the percentage area of the cell covered by shrubland. This indicator is a snapshot for the year 2009 only.

Please cite:

Bontemps, Sophie; Pierre Defourny & Eric Van Bogaert (2009) Globcover 2009. Products Description and Validation Report. *European Space Agency*. (http://due.esrin.esa.int/files/GLOBCOVER2009_Validation_Report_2.2.pdf).

herb_gc [Original data] measures the coverage of herbaceous vegetation and lichens/mosses in each cell, extracted from the Globcover 2009 dataset v.2.3. To compute herb_gc [Original data] we follow the FAO land cover classification system used by Globcover and aggregate to the category "Herbaceous" (landuse class 140). The value indicates the percentage area of the cell covered by herbaceous vegetation and lichens/mosses. This indicator is a snapshot for the year 2009 only.

Please cite:

Bontemps, Sophie; Pierre Defourny & Eric Van Bogaert (2009) Globcover 2009. Products Description and Validation Report. *European Space Agency*. (http://due.esrin.esa.int/files/GLOBCOVER2009_Validation_Report_2.2.pdf).

aquaveg_gc [Original data] measures the coverage of aquatic vegetation in each cell, extracted from the Globcover 2009 dataset v.2.3. To compute aquaveg_gc [Original data] we follow the FAO land cover classification system used by Globcover and aggregate to the category "Natural and seminatural aquatic vegetation" (landuse classes 150, 160, 170, 180). The value indicates the percentage area of the cell covered by aquatic vegetation. This indicator is a snapshot for the year 2009 only.

Please cite:

Bontemps, Sophie; Pierre Defourny & Eric Van Bogaert (2009) Globcover 2009. Products Description and Validation Report. *European Space Agency*. (http://due.esrin.esa.int/files/GLOBCOVER2009_Validation_Report_2.2.pdf).

barren_gc [Original data] measures the coverage of barren areas in each cell, extracted from the Globcover 2009 dataset v.2.3. To compute barren_gc [Original data] we follow but deviate slightly from the FAO land cover classification system used by Globcover by aggregating to the category "Barren" and also including the "Permanent snow and ice" class (landuse classes 200, 220). The value indicates the percentage area of the cell covered by barren area. This indicator is a snapshot for the year 2009 only.

Please cite:

Bontemps, Sophie; Pierre Defourny & Eric Van Bogaert (2009) Globcover 2009. Products Description and Validation Report. *European Space Agency*. (http://due.esrin.esa.int/files/GLOBCOVER2009_Validation_Report_2.2.pdf).

water_gc [Original data] measures the coverage of water areas in each cell, extracted from the Globcover 2009 dataset v.2.3. To compute water_gc [Original data] we follow but deviate slightly from the FAO land cover classification system used by Globcover and aggregate only to the "Natural/Artificial water bodies" class excluding the "Permanent snow and ice" class (landuse class 2010). The value indicates the percentage area of the cell covered by water area. This indicator is a snapshot for the year 2009 only.

Please cite:

Bontemps, Sophie; Pierre Defourny & Eric Van Bogaert (2009) Globcover 2009. Products Description and Validation Report. *European Space Agency*. (http://due.esrin.esa.int/files/GLOBCOVER2009_Validation_Report_2.2.pdf).

maincrop [Original data] indicates the main crop code for the cell, based on the Cropping Periods List data from the MIRCA2000 dataset v.1.1. The main crop is determined as the subcrop with the highest harvested area for each cell. Crop codes 26 and below are irrigated crops, while 27 and up are rainfed crops. Data is only a snapshot for the year 2000.

Please cite:

Portmann, Felix T., Stefan Siebert & Petra Döll (2010): MIRCA2000 – Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling, *Global Biogeochemical Cycles*, 24, GB 1011, doi:10.1029/2008GB003435.

harvarea [Original data] is the sum of the harvested area (given in hectares) for the cell's main crop determined in the maincrop [Original data] variable, based on the Cropping Periods List data from the MIRCA2000 dataset v.1.1.

Please cite the same source as maincrop [Original data].

Climate variables

rainseas [**Original data**] gives the initial month of the rainy season in the cell (values 1-12), defined as the first of the three consecutive months during a normal year with the highest total amount of rainfall, calculated on the basis of the prec_gpcc [Original data] variable for the 1946-2013 period.

Please cite:

> Schneider, Udo, Andreas Becker, Peter Finger, Anja Meyer-Christoffer, Bruno Rudolf and Markus Ziese (2015). *GPCC Full Data Reanalysis Version 7.0 at 0.5°: Monthly Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historic Data*. doi:10.5676/DWD_GPCC/FD_M_V7_050

growstart [**Original data**] provides the starting month of the growing season for the cell's main crop determined in the maincrop [Original data] variable, values 1-12, based on the Cropping Periods List data from the MIRCA2000 dataset v.1.1.

Please cite:

> Portmann, Felix T., Stefan Siebert & Petra Döll (2010): MIRCA2000 – Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling. *Global Biogeochemical Cycles*, 24, GB 1011, doi:10.1029/2008GB003435.

growend [**Original data**] provides the final month of the growing season for the cell's main crop determined in the maincrop [Original data] variable, values 1-12, based on the Cropping Periods List data from the MIRCA2000 dataset v.1.1.

Please cite the same source as growstart [Original data].

The Temporal Table

In addition to the grid identifier, the temporal data table also includes a year variable. This means that gid + year create a unique identifier in the time-series data. The PRIO-GRID v.2.0 temporal table contains yearly observations of all terrestrial grid cells (excluding Antarctica and Greenland) for all calendar years between 1946 and 2014. In total, the table contains 64,818 cells x 69 years = 4,472,442 observations (cell years) in total. However, all variables are not available for all years.

gid is the grid cell identifier, a unique id code for each cell in the grid.

year gives the calendar year of observation.

Country variables

gwno [**Original data**] denotes the numerical country code for the country to which the cell is allocated, based on the Gleditsch & Ward system membership list and cShapes geometries. Each cell is assigned to one and only one country in each yearly file. To determine country ownership, PRIO-GRID draws on the cShapes dataset v.0.4-2, last modified 22 Mar 2015. Grid cells that fall completely within the territory of an independent state are assigned the corresponding Gleditsch & Ward country code (gwno). The country code reflects the status as of 31 December of each year, which means that in the case of territorial transfer (e.g., from East Pakistan to Bangladesh in 1971), a cell is given the country code that applies to the status at the end of the year, 31 December. Grid cells that cover the territory of two or more independent states (i.e., the cell intersects with multiple country polygons) are assigned to the country that covers the largest share of the cell's area. Note that while all terrestrial cells are included in the yearly file, country codes are assigned to cells only in those years that the host country is a member of the Gleditsch & Ward international system. Missing values imply non-independent territory.

Please cite:

> Gleditsch, Kristian Skrede & Michael D. Ward (1999) Interstate system membership: A revised list of the independent states since 1816. *International Interactions*, 25: 393-413.

> Weidmann, Nils B., Doreen Kuse & Kristian Skrede Gleditsch (2010) The geography of the international system: The CShapes Dataset. *International Interactions*, 36(1): 86-106.

gwarea gives the land area in square kilometers of the grid cell belonging to the allocated country for that year, based on cShapes. Areas are calculated assuming that the earth in an oblate spheroid (WGS 84).

Please cite:

> Weidmann, Nils B., Doreen Kuse & Kristian Skrede Gleditsch (2010) The geography of the international system: The CShapes Dataset. *International Interactions*, 36(1): 86-106.

bdist1 gives the spherical distance in kilometer from the cell centroid to the border of the nearest land-contiguous neighboring country, based on country border data using cShapes v.0.4-2. This implies that cells in e.g. Northern Denmark are measured to the border to Germany even if the straight-line distance to Norway (across international waters) is shorter. Cells belonging to island states with no contiguous neighboring country (e.g., New Zealand) are coded as missing.

Please cite:

> Weidmann, Nils B., Doreen Kuse & Kristian Skrede Gleditsch (2010) The geography of the international system: The CShapes Dataset. *International Interactions*, 36(1): 86-106.

bdist2 gives the spherical distance (in kilometer) from the cell centroid to the border of the nearest neighboring country, regardless of whether the nearest country is located across international waters. Hence, for cells belonging to island states (e.g. New Zealand), bdist2 gives the shortest distance to the nearest land territory of another state.

Please cite:

> Weidmann, Nils B., Doreen Kuse & Kristian Skrede Gleditsch (2010) The geography of the international system: The CShapes Dataset. *International Interactions*, 36(1): 86-106.

bdist3 gives the spherical distance (in kilometer) from the cell centroid to the territorial outline of the country the cell belongs to. For cells located along a coast and for cells of island states (e.g. New Zealand), bdist3 measures the shortest straight-line distance to international waters. By definition, bdist3 can never have higher values than the two other border distance indicators and for 44 % of the cell years all three border distance estimates are identical.

Please cite:

> Weidmann, Nils B., Doreen Kuse & Kristian Skrede Gleditsch (2010) The geography of the international system: The CShapes Dataset. *International Interactions*, 36(1): 86-106.

capdist gives the spherical distance in kilometers from the cell centroid to the national capital city in the corresponding country, based on coordinate pairs of capital cities derived from the cShapes dataset v.0.4-2. It captures changes over time wherever relevant. Figure 3 visualizes these straight-line distances.

Please cite:

> Weidmann, Nils B., Doreen Kuse & Kristian Skrede Gleditsch (2010) The geography of the international system: The CShapes Dataset. *International Interactions*, 36(1): 86-106.

Socioeconomic variables

pop_gpw_ [**Original data**] measures population size, taken from the Gridded Population of the World version 3. Population estimates are available for 1990, 1995, 2000, and 2005. The original pixel value is number of persons.

- **pop_gpw_sum** gives the sum of pixel values (number of persons) within the grid cell. To obtain population density estimates, this variable can be divided by landarea in the static table.
- **pop_gpw_sd** gives the standard deviation of original pixel values within each cell.
- **pop_gpw_min** gives the minimum of original pixel values within each cell.
- **pop_gpw_max** gives the maximum of original pixel values within each cell.

Please cite:

Center for International Earth Science Information Network (CIESIN) and Centro Internacional de Agricultura Tropical (CIAT) (2005). *Gridded Population of the World, Version 3 (GPWv3): Population Count Grid*. Palisades, NY. doi:10.7927/H4639MPP. Accessed 03.06.2013.

pop_hyd_ [**Original data**] measures the population size for each populated cell in the grid, taken from the History Database of the Global Environment (HYDE) version 3.1. Population estimates are available for 1950, 1960, 1970, 1980, 1990, 2000, and 2005. The original pixel value is number of persons.

- **pop_hyd_sum** gives the sum of pixel values (number of persons) within the grid cell. To obtain population density estimates, this variable can be divided bylandarea in the static table.
- **pop_hyd_sd** gives the standard deviation of original pixel values within each cell.
- **pop_hyd_min** gives the minimum of original pixel values within each cell.
- **pop_hyd_max** gives the maximum of original pixel values within each cell.

Please cite:

Klein Goldewijk, K. , A. Beusen, M. de Vos and G. van Drecht (2011). The HYDE 3.1 spatially explicit database of human induced land use change over the past 12,000 years. *Global Ecology and Biogeography*, 20(1): 73-86. doi: 10.1111/j.1466-8238.2010.00587.x.

Klein Goldewijk, K. , A. Beusen, and P. Janssen (2010). Long term dynamic modeling of global population and built-up area in a spatially explicit way, HYDE 3 .1. *The Holocene*, 20(4):565-573. doi:10.1177/0959683609356587

excluded [**Original data**] counts the number of excluded groups (discriminated or powerless) as defined in the GeoEPR/EPR data on the status and location of politically relevant ethnic groups settled in the grid cell for the given year, derived from the GeoEPR/EPR 2014 update 2 dataset.

Please cite:

Vogt, Manuel, Nils-Christian Bormann, Seraina Rüegger, Lars-Erik Cederman, Philipp Hunziker, and Luc Girardin. 2015. "Integrating Data on Ethnicity, Geography, and Conflict: The Ethnic Power Relations Dataset Family." *Journal of Conflict Resolution*, 59(7), 1327-1342. doi:10.1177/0022002715591215

gcp_mer [**Original data**] indicates the gross cell product, measured in USD, based on the G-Econ dataset v4.0, last modified May 2011. The original G-Econ data represent the total economic activity at a 1x1 degree resolution, so when assigning this to PRIO-GRID we distribute the total value across the number of contained PRIO-GRID land cells. In border areas, the G-Econ 1x1 degree cells might overlap with PRIO-GRID cells allocated to a neighboring country. To minimize bias, PRIO-GRID only extracts G-Econ data for cells that have the same country code as the G-Econ cell represents. This variable is only available for five-year intervals since 1990.

Note: The user should be aware of the following special case. Differing definitions used by G-Econ and PRIO-GRID over the territorial border between Libya and Chad across all years resulted in no G-Econ cell being matched with PRIO-GRID's country definitions, leaving a small strip of cells in the border region with missing GCP data.

Please cite:

> Nordhaus, William D. (2006) Geography and macroeconomics: New data and new findings. *Proceedings of the National Academy of Sciences of the USA*, 103(10): 3510-3517.

gcp_ppp [**Original data**] indicates the gross cell product, measured in USD using purchasing-power-parity, based on the G-Econ dataset v4.0, last modified May 2011. Else similar to gcp_mer [Original data], but uses USD at purchasing-power-parity which corrects for each currency's purchasing power. This variable is only available for 1990, 1995, 2000, and 2005.

Please cite:

> Nordhaus, William D. (2006) Geography and macroeconomics: New data and new findings. *Proceedings of the National Academy of Sciences of the USA*, 103(10): 3510-3517.

gcp_qual [**Original data**] indicates the quality of the GCP values, based on the G-Econ dataset v4.0, last modified May 2011. Quality is a measure of the quality of the economic data. Quality = 1 for countries for which the data are consistent, but it does not capture the quality of the underlying country statistics. In general, quality < 1 indicates that there are major inconsistencies in one of the underlying data inputs into GCP. See the G-Econ definition table, available at http://gecon.yale.edu/.

Please cite:

> Nordhaus, William D. (2006) Geography and macroeconomics: New data and new findings. *Proceedings of the National Academy of Sciences of the USA*, 103(10): 3510-3517.

Resource variables

petroleum_y [**Original data**] is a dummy variable for whether onshore petroleum deposits have been found within the given grid cell for any given year, based on the Petroleum Dataset v.1.2. This variable only codes those petroleum deposits that have a known discovery or start of production year. For a complete picture, these data must therefore be combined with the petroleum_s [Original data] data.

Please cite:

> Lujala, Päivi, Jan Ketil Rød & Nadia Thieme, 2007. Fighting over Oil: Introducing A New Dataset. *Conflict Management and Peace Science*, 24(3), 239-256.

diamsec_y [Original data] is a dummy variable for whether secondary (alluvial) diamond deposits have been found within the given grid cell for any given year, based on the Diamond Resources dataset v1a. This variable only codes those deposits that have a known discovery or start of production year. For a complete picture, these data must therefore be combined with the diamsec_s [Original data] data.

Please cite:

> Gilmore, Elisabeth, Nils Petter Gleditsch, Päivi Lujala & Jan Ketil Rød, 2005. Conflict Diamonds: A New Dataset. *Conflict Management and Peace Science*, 22(3): 257–292
>
> Lujala, Päivi, Nils Petter Gleditsch & Elisabeth Gilmore, 2005. A Diamond Curse? Civil War and a Lootable Resource. *Journal of Conflict Resolution*, 49(4): 538–562.

diamprim_y [Original data] is a dummy variable for whether primary (kimberlite) diamond deposits have been found within the given grid cell for any given year, based on the Diamond Resources dataset v1a. This variable only codes those deposits that have a known discovery or start of production year. For a complete picture, these data must therefore be combined with the diamprim_s data.

Please cite:

> Gilmore, Elisabeth, Nils Petter Gleditsch, Päivi Lujala & Jan Ketil Rød, 2005. Conflict Diamonds: A New Dataset, *Conflict Management and Peace Science* 22(3): 257–292
>
> Lujala, Päivi, Nils Petter Gleditsch & Elisabeth Gilmore, 2005. A Diamond Curse? Civil War and a Lootable Resource. *Journal of Conflict Resolution*, 49(4): 538–562.

goldplacer_y [Original data] is a dummy variable for whether placer gold deposits have been found within the given grid cell, based on the GOLDATA_L subset of the GOLDATA dataset v1.2. This variable only codes those deposits that have a known discovery or start of production year. For a complete picture, these data must therefore be combined with the goldplacer_s [Original data] data.

Please cite:

> Balestri, Sara, 2015. GOLDATA: The Gold deposits dataset codebook, Version 1.2. *UCSC-Cognitive Science and Communication Research Centre*, WP 02/15, Milan. doi:10.13140/RG.2.1.1730.8648
>
> Balestri, Sara, 2012. Gold and civil conflict intensity: evidence from a spatially disaggregated analysis. *Peace Economics, Peace Science and Public Policy*, 18(3): 1-17. doi:10.1515/peps-2012-0012.

goldsurface_y [Original data] is a dummy variable for whether surface gold deposits defined as deposits that are located near the surface but "do not hold enough information to be properly defined as lootable" have been found within the given grid cell, based on the GOLDATA_S subset of the GOLDATA dataset v1.2. This variable only codes those deposits that have a known discovery or start of production year. For a complete picture, these data must therefore be combined with the goldsurface_s [Original data] data.

Please cite the same source as goldplacer_y [Original data].

goldvein_y [Original data] is a dummy variable for whether vein gold deposits have been found within the given grid cell, based on the GOLDATA_NL subset of the GOLDATA dataset v1.2. This variable only codes those deposits that have a known discovery or start of production year. For a complete picture, these data must therefore be combined with the goldvein_s [Original data] data.

Please cite the same source as goldplacer_y [Original data].

gem_y [Original data] is a dummy variable for whether gem deposits have been found within the given grid cell, based on the GEMDATA dataset. This variable only codes those deposits that have a known discovery or start of production year. For a complete picture, these data must therefore be combined with the gem_s [Original data] data.

Please cite:

> Lujala, Päivi 2009. Deadly Combat over Natural Resources: Gems, Petroleum, Drugs, and the Severity of Armed Civil Conflict. *Journal of Conflict Resolution*, 53(1): 50-71.

drug_y [Original data] is a dummy variable for whether large-scale drug cultivation (coca bush, opium poppy, or cannabis) is ongoing within the given grid cell, based on the DRUGDATA dataset.

Please cite:

> Buhaug, Halvard & Päivi Lujala 2005. Accounting for Scale: Measuring Geography in Quantitative Studies of Civil War. *Political Geography*, 24: 399-418.

Climate variables

prec_gpcc [Original data] gives the yearly total amount of precipitation (in millimeter) in the cell, based on monthly meteorological statistics from the Global Precipitation Climatology Centre. This indicator contains data for the years 1946-2013 in PRIO-GRID (1901/01 - 2013/12 in the original data).

Please cite:

> Schneider, Udo, Andreas Becker, Peter Finger, Anja Meyer-Christoffer, Bruno Rudolf and Markus Ziese (2015): GPCC Full Data Reanalysis Version 7.0 at 0.5°: Monthly Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historic Data. doi: 10.5676/DWD_GPCC/FD_M_V7_050

prec_gpcp [Original data] gives the yearly total amount of precipitation (in millimeter) in the cell, based on monthly meteorological statistics from the GPCP v.2.2 Combined Precipitation Data Set. Since the original data only reported the daily average for each month, we multiplied the daily average by the number of days in each month in order to obtain approximate monthly totals, from which yearly totals were estimated. This indicator contains data for the years 1979-2014.

Please cite:

> Huffman, G.J., D.T. Bolvin, R.F. Adler, 2012, last updated 2012: GPCP Version 2.2 SG Combined Precipitation Data Set. WDC-A, NCDC, Asheville, NC. Dataset accessed 26.06.2015 at ftp://precip.gsfc.nasa.gov/pub/gpcp-v2.2/psg/

Also please note when using:

> The GPCP combined precipitation data were developed and computed by the NASA/Goddard Space Flight Center's Laboratory for Atmospheres as a contribution to the GEWEX Global Precipitation Climatology Project.
>
> The GPCP data was provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at http://www.esrl.noaa.gov/psd/.

temp [Original data] gives the yearly mean temperature (in degrees Celsius) in the cell, based on monthly meteorological statistics from GHCN/CAMS, developed at the Climate Prediction Center, NOAA/National Weather Service. This indicator contains data for the years 1948-2014.

Please cite:

> Fan, Yun and Huug van den Dool (2008), A global monthly land surface air temperature analysis for 1948-present, *Journal of Geophysical Research*, 113, D01103, doi:10.1029/2007JD008470.

Also please note when using:

> The GHCN Gridded V2 data was provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at http://www.esrl.noaa.gov/psd/.

droughtstart_spi [Original data] gives the severity of drought during the first month of the cell's rainy season, as defined by the rainseas [Original data] variable. The severity value is the SPI1 value during the first month of the rainy season. The monthly SPI1 index measures deviation from long-term normal rainfall for that month. The values are standardized where deviation estimates less than 1 standard deviation indicate near normal rainfall.

We use SPI data from the International Research Institute for Climate and Society at Colombia University, as defined in Guttman (1999). The SPI values are calculated based on the CAMS_OPI precipitation dataset. This indicator contains data for the years 1979-2014.

Please cite:

Guttman, N. B., 1999: Accepting the Standardized Precipitation Index: A calculation algorithm. Journal of the American Water Resources Association, 35(2), 311-322.

McKee, Thomas B., Nolan J. Doesken, and John D. Kliest (1993) The relationship of drought frequency and duration to time scales. *In Proceedings of the 8th Conference of Applied Climatology*, 17-22 January, Anaheim, CA. American Meteorological Society, Boston, MA. 179-184.

droughtend_spi [Original data] gives the severity of drought for the entirety of the cell's rainy season, as defined by the rainseas [Original data] variable. The severity value is the SPI3 value for the last month the rainy season. For each month, the monthly SPI3 index measures deviation from long-term normal rainfall during the three preceding months. A rainy season is defined as the three consecutive months in which it on average rained the most during a year in any cell.

Please cite the same source as droughtstart_spi [Original data].

droughtyr_spi [Original data] gives the proportion of months out of 12 months that are part of the longest streak of consecutive months ending in the given year with SPI1 values below -1.5. For a year where the longest consecutive streak of months below -1.5 is three, the cell will be given a value of 3/12 = 0.25. When the longest streak starts in the previous year, it is only counted and included in the year in which the streak ended. Theoretically, the proportion can become higher than 1.

Please cite the same source as droughtstart_spi [Original data].

droughtcrop_spi [Original data] gives the proportion of months in the growing season that are part of the longest streak of consecutive months in that growing season with SPI1 values below -1.5. The growing season is the growing season for the cell's main crop, defined in the MIRCA2000 dataset v.1.1. For growing seasons that cross 1 January, we define the whole season to belong to the year in which the season ended. Thus, a year with two consecutive months below -1.5 during the growing season that started in September the previous year and ended in March in the current year, is given a value of 2/8 = 0.25. Each year only have defined one growing season.

Please cite the same source as droughtstart_spi [Original data].

droughtstart_speigdm [Original data] This variable is operationalized similarly as droughtstart_spi [Original data], only that instead of using the SPI1, it uses the Standardized Precipitation and Evapotranspiration Index SPEI1 from the SPEI Global Drought Monitor, downloaded 15 July 2015. SPEI GDM uses the GPCC 'first guess' product and GHCN/CAMS, while using the Thornthwaite potential evapotranspiration (PET) estimation.

Please cite:

Beguería, Santiago, Sergio M. Vicente-Serrano, Fergus Reig, and Borja Latorre (2014), Standardized Precipitation Evapotranspiration Index (SPEI) revisited: parameter fitting, evapotranspiration models, tools, datasets and drought monitoring. *International Journal of Climatology*, 34(10): 3001–3023. doi: 10.1002/joc.3887

droughtend_speigdm [Original data] This variable is operationalized similarly as droughtend_spi [Original data], only that instead of using the SPI-1, it uses the Standardized Precipitation and Evapotranspiration Index SPEI-3 from the SPEI Global Drought Monitor.

Please cite the same source as droughtstart_speigdm [Original data].

droughtyr_speigdm [Original data] This variable is operationalized similarly as droughtyr_spi [Original data], only that instead of using the SPI-1, it uses the Standardized Precipitation and Evapotranspiration Index SPEI-1 from the SPEI Global Drought Monitor.

Please cite the same source as droughtstart_speigdm [Original data].

droughtcrop_speigdm [Original data] This variable is operationalized similarly as droughtcrop_spi [Original data], only that instead of using the SPI-1, it uses the Standardized Precipitation and Evapotranspiration Index SPEI-1 from the SPEI Global Drought Monitor.

Please cite the same source as droughtstart_speigdm [Original data].

droughtstart_speibase [Original data] This variable is operationalized similarly as droughtstart_spi [Original data], only that instead of using the SPI-1, it uses the Standardized Precipitation and Evapotranspiration Index SPEI-1 from the SPEIbase v.2.3. SPEIbase is based on precipitation and potential evapotranspiration from the Climatic Research Unit of University of East Anglia CRU v.3.22. The PET estimation used by CRU is the Penman-Montheith method, considered better than the Thornthwaite estimation.

Please cite:

> Beguería, Santiago, Sergio M. Vicente Serrano, and Marta Angulo-Martínez (2010). A Multiscalar Global Drought Dataset: The SPEIbase: A New Gridded Product for the Analysis of Drought Variability and Impacts. *Bulletin of the American Meteorological Society*, 91 (10): 1351–1356. doi:10.1175/2010BAMS2988.1

droughtend_speibase [Original data] This variable is operationalized similarly as droughtend_spi [Original data], only that instead of using the SPI-1, it uses the Standardized Precipitation and Evapotranspiration Index SPEI-3 from the SPEIbase v.2.3. SPEIbase is based on precipitation and potential evapotranspiration from the Climatic Research Unit of University of East Anglia CRU v.3.22. The PET estimation used by CRU is the Penman-Montheith method, considered better than the Thornthwaite estimation.

Please cite the same source as droughtstart_speibase [Original data].

droughtyr_speibase [Original data] This variable is operationalized similarly as droughtyr_spi [Original data], only that instead of using the SPI-1, it uses the Standardized Precipitation and Evapotranspiration Index SPEI-1 from the SPEIbase v.2.3. SPEIbase is based on precipitation and potential evapotranspiration from the Climatic Research Unit of University of East Anglia CRU v.3.22. The PET estimation used by CRU is the Penman-Montheith method, considered better than the Thornthwaite estimation.

Please cite the same source as droughtstart_speibase [Original data].

droughtcrop_speibase [Original data] This variable is operationalized similarly as droughtcrop_spi [Original data], only that instead of using the SPI-1, it uses the Standardized Precipitation and Evapotranspiration Index SPEI-1 from the SPEIbase v.2.3. SPEIbase is based on precipitation and potential evapotranspiration from the Climatic Research Unit of University of East Anglia CRU v.3.22. The PET estimation used by CRU is the Penman-Montheith method, considered better than the Thornthwaite estimation.

Please cite the same source as droughtstart_speibase [Original data].

Landuse variables

irrig_ [Original data] measures the area equipped for irrigation within each cell (in hectares). The data is taken from the Historical Irrigation dataset v.1, which indicates pixelated data on areas equipped for irrigation across time. Specifically we used the AEI_EARTHSTAT_IR dataset, which reports irrigation based on subnational sources and Earthstat historical landuse data. In PRIO-GRID, this indicator is only available for the years 1950, 1960, 1970, 1980, 1985, 1990, 1995, 2000, and 2005.

- **irrig_sum** gives the total area (in hectares) equipped for irrigation within the grid cell.
- **irrig_sd** gives the standard deviation of original pixel values within each cell.
- **irrig_min** gives the minimum of original pixel values within each cell.
- **irrig_max** gives the maximum of original pixel values within each cell.

Please cite:

> Stefan Siebert, Matti Kummu, Miina Porkka, Petra Döll, Navin Ramankutty, Bridget R. Scanlon (2015). Historical Irrigation Dataset (HID). *MyGeoHUB*. doi:10.13019/M20599

urban_ih [Original data] gives the percentage area of the cell covered by urban area, based on ISAM-HYDE landuse data. To measure the coverage of urban areas we include the percentage urban areas in a cell extracted from the ISAM-HYDE historical landuse dataset. To compute urban_ih [Original data] we follow the land cover classification system used by ISAM-HYDE and aggregate to the category "Urban" (landuse class "Urban"). In PRIO-GRID, this indicator is available for the years 1950, 1960, 1970, 1980, 1990, 2000, and 2010.

Please cite:

> Meiyappan, Prasanth and Atul K. Jain (2012). Three distinct global estimates of historical land-cover change and land-use conversions for over 200 years. *Frontiers of Earth Science*, 6(2), 122-139. doi:10.1007/s11707-012-0314-2.

agri_ih [Original data] gives the percentage area of the cell covered by agricultural area, based on ISAM-HYDE landuse data. To measure the coverage of agricultural areas we include the percentage agricultural areas in a cell extracted from the ISAM-HYDE historical landuse dataset. To compute agri_ih [Original data] we follow the land cover classification system used by ISAM-HYDE and aggregate to the category "Total cropland" (landuse classes "C3crop", "C4crop"). In PRIO-GRID, this indicator is available for the years 1950, 1960, 1970, 1980, 1990, 2000, and 2010.

Please cite:

Meiyappan, Prasanth and Atul K. Jain (2012). Three distinct global estimates of historical land-cover change and land-use conversions for over 200 years. *Frontiers of Earth Science*, 6(2), 122-139. doi: 10.1007/s11707-012-0314-2.

pasture_ih [Original data] gives the percentage area of the cell covered by pasture area, based on ISAM-HYDE landuse data. To measure the coverage of pasture areas we include the percentage pasture areas in a cell extracted from the ISAM-HYDE historical landuse dataset. To compute pasture_ih [Original data] we follow the land cover classification system used by ISAM-HYDE and aggregate to the category "Total pastureland" (landuse classes "C3past", "C4past"). In PRIO-GRID, this indicator is available for the years 1950, 1960, 1970, 1980, 1990, 2000, and 2010.

Please cite:

Meiyappan, Prasanth and Atul K. Jain (2012). Three distinct global estimates of historical land-cover change and land-use conversions for over 200 years. *Frontiers of Earth Science*, 6(2), 122-139. doi: 10.1007/s11707-012-0314-2.

forest_ih [Original data] gives the percentage area of the cell covered by forest area, based on ISAM-HYDE landuse data. To measure the coverage of forest areas we include the percentage forest areas in a cell extracted from the ISAM-HYDE historical landuse dataset. To compute forest_ih [Original data] we follow the land cover classification system used by ISAM-HYDE and aggregate to the category "Total forest" (landuse classes "TrpEBF", "TrpDBF", "TmpEBF", "TmpENF", "TmpDBF", "BorENF", "BorDNF", "SecTrpEBF", "SecTrpDBF", "SecTmpEBF", "SecTmpENF", "SecTmpDBF", "SecBorENF", "SecBorDNF"). In PRIO-GRID, this indicator is available for the years 1950, 1960, 1970, 1980, 1990, 2000, and 2010.

Please cite:

Meiyappan, Prasanth and Atul K. Jain (2012). Three distinct global estimates of historical land-cover change and land-use conversions for over 200 years. *Frontiers of Earth Science*, 6(2), 122-139. doi: 10.1007/s11707-012-0314-2.

grass_ih [Original data] gives the percentage area of the cell covered by grasslands, based on ISAM-HYDE landuse data. To measure the coverage of grasslands we include the percentage grassland areas in a cell extracted from the ISAM-HYDE historical landuse dataset. To compute grass_ih [Original data] we follow the land cover classification system used by ISAM-HYDE and aggregate to the category "Total grassland"(landuse classes "C3grass", "C4grass"). In PRIO-GRID, this indicator is available for the years 1950, 1960, 1970, 1980, 1990, 2000, and 2010.

Please cite:

Meiyappan, Prasanth and Atul K. Jain (2012). Three distinct global estimates of historical land-cover change and land-use conversions for over 200 years. *Frontiers of Earth Science*, 6(2), 122-139. doi: 10.1007/s11707-012-0314-2.

shrub_ih [Original data] gives the percentage area of the cell covered by shrublands, based on ISAM-HYDE landuse data. To measure the coverage of shrublands we include the percentage shrubland areas in a cell extracted from the ISAM-HYDE historical landuse dataset. To compute shrub_ih [Original data] we follow the land cover classification system used by ISAM-HYDE and aggregate to the category "Total shrubland"(landuse classes "Denseshrub", "Openshrub"). In PRIO-GRID, this indicator is available for the years 1950, 1960, 1970, 1980, 1990, 2000, and 2010.

Please cite:

> Meiyappan, Prasanth and Atul K. Jain (2012). Three distinct global estimates of historical land-cover change and land-use conversions for over 200 years. *Frontiers of Earth Science*, 6(2), 122-139. doi: 10.1007/s11707-012-0314-2.

savanna_ih [Original data] gives the percentage area of the cell covered by grasslands, based on ISAM-HYDE landuse data. To measure the coverage of savanna we include the percentage savanna areas in a cell extracted from the ISAM-HYDE historical landuse dataset. To compute savanna_ih [Original data] we follow the land cover classification system used by ISAM-HYDE and aggregate to the category "Savanna" (landuse class "Savanna"). In PRIO-GRID, this indicator is available for the years 1950, 1960, 1970, 1980, 1990, 2000, and 2010.

Please cite:

> Meiyappan, Prasanth and Atul K. Jain (2012). Three distinct global estimates of historical land-cover change and land-use conversions for over 200 years. *Frontiers of Earth Science*, 6(2), 122-139. doi: 10.1007/s11707-012-0314-2.

barren_ih [Original data] gives the percentage area of the cell covered by barren area, based on ISAM-HYDE landuse data. To measure the coverage of barren areas we include the percentage barren areas in a cell extracted from the ISAM-HYDE historical landuse dataset. To compute barren_ih [Original data] we aggregate using the following lansuse classes: "Tundra", "Desert", "PdRI". In PRIO-GRID, this indicator is available for the years 1950, 1960, 1970, 1980, 1990, 2000, and 2010.

Please cite:

> Meiyappan, Prasanth and Atul K. Jain (2012). Three distinct global estimates of historical land-cover change and land-use conversions for over 200 years. *Frontiers of Earth Science*, 6(2), 122-139. doi: 10.1007/s11707-012-0314-2.

water_ih [Original data] gives the percentage area of the cell covered by water area, based on ISAM-HYDE landuse data. To measure the coverage of water areas we include the percentage water areas in a cell extracted from the ISAM-HYDE historical landuse dataset. To compute water_ih [Original data] we aggregate using the following landuse class: "Water". In PRIO-GRID, this indicator is available for the years 1950, 1960, 1970, 1980, 1990, 2000, and 2010.

Please cite:

> Meiyappan, Prasanth and Atul K. Jain (2012). Three distinct global estimates of historical land-cover change and land-use conversions for over 200 years. *Frontiers of Earth Science*, 6(2), 122-139. doi: 10.1007/s11707-012-0314-2.

nlights_ [**Original data**] measures average nighttime light emission from the DMSP-OLS Nighttime Lights Time Series Version 4 (Average Visible, Stable Lights, & Cloud Free Coverages). We use the data gathered from the newest satellites (F10 in 1992-93, F12 in 1994-1996, and so on). These data are not calibrated for time-series analysis, but are available from 1992-2013.

- **nlights_mean** gives the mean night time lights within the grid cell.
- **nlights_sd** gives the standard deviation of original pixel values within each cell.
- **nlights_min** gives the minimum of original pixel values within each cell.
- **nlights_max** gives the maximum of original pixel values within each cell.

Please cite/note:

> Image and data processing by NOAA's National Geophysical Data Center. DMSP data collected by US Air Force Weather Agency.

nlights_calib_mean [**Original data**] measures average nighttime light emission from the DMSP-OLS Nighttime Lights Time Series Version 4 (Average Visible, Stable Lights, & Cloud Free Coverages), calibrated to account for intersatellite differences and interannual sensor decay using calibration values from Elvidge et.al. (2013). Thus, they might be more suitable for time-series analysis. Values are standardized to be between 0 and 1, where 1 is the highest observed value in the time-series, and 0 is the lowest. The times-series are available from 1992-2012.

Please cite/note:

> Elvidge, Christopher D., Feng-Chi Hsu, Kimberly E. Baugh and Tilottama Ghosh (2014). "National Trends in Satellite Observed Lighting: 1992-2012." *Global Urban Monitoring and Assessment Through Earth Observation*. Ed. Qihao Weng. CRC Press.
>
> Image and data processing by NOAA's National Geophysical Data Center. DMSP data collected by US Air Force Weather Agency.