The Hidden Dimensions of Poverty Dataset

A subnational dataset with global coverage linking environment and natural Resources to measures of economic wellbeing, including poverty

A collaboration between:

Table of Contents

1. Background of the Hidden Dimensions of Poverty Dataset

Purpose of the dataset. The Hidden Dimensions of Poverty Dataset (HDD) is a unique and global geospatial dataset linking environment and natural resource (ENR) measures to poverty and other human development indicators at the subnational level. The HDD was created to uncover the spatial linkages between ENR and poverty at the subnational scale and shed some light on the correlations between the degradation of the environment and poverty (as well as other measures of human development) at the global level.

Type of indicators. The ENR indicators contain measures pertaining to the Green, Brown, and Blue business lines of the World Bank's Environment, Natural Resources and Blue Economy Global Practice (ENB GP), as well as geographical variables. The poverty measures contain monetary and asset-based poverty measures, as well as measures closely correlated with poverty such as children's health and GDP. The next section, section 2, provides a detailed overview of the indicators created for the HDD dataset.

Spatial scale and size of the nested dataset. There are two main different subnational units (aside from the national unit) of analysis for the global dataset: The district level (administrative unit 2 level), and the province level (administrative unit 1 level). Depending on the country, the admin-1 level may refer to a province, a state, or a department and an admin-2 level may refer to either a district, or a municipality. For consistency, this note will from here on refer to an admin-2 unit as a district and an admin-1 unit as a province. The global district level dataset contains 46,311 observations and the global province level dataset contains 3,609 observations.[1] The HDD is a nested dataset, meaning that the three levels of spatial aggregation are related based on administrative hierarchy; this is true not only for the GIS dataset, but also for the Stata and Excel files. In other words, each district relates to the unique province within which it falls, and each province relates to a unique country. See Appendix A for an elaboration on the nested structure of the dataset.

Time dimension of the dataset. HDD is mainly a cross-sectoral dataset including measures of poverty, which are lined up in the first version of HDD to the year 2013 (see Section 3 for details). For most countries, poverty measures are not dynamic i.e. it is measured at one point in time only, and changes in poverty cannot be evaluated. Similarly, the alternative measures of poverty such as the Multidimensional Poverty Index (MPI) and the Demographic and Health Surveys (DHS) indicators do not have any time dimension. On the ENR side, some measures are only stock measures from one point in time, while others have a trend dimension). We have

[1] This is the maximum amount of records available in our dataset. The actual observations vary from indicator to indicator, and from country to country. There are several indicators such as air pollution and GDP for which we have full coverage, but there are others such as monetary poverty, for which we do not.

panel data for vegetation index, forest cover loss, net primary productivity, PM2.5 air pollution and land cover.

Administrative Boundaries. The sub-national dataset largely follows Global Administrative Database (GADM) administrative boundaries, with the exception of the monetary poverty dataset, which follows a hybrid of several administrative boundaries (see Section 3). The socioeconomic and environmental science community is generally using two administrative boundary datasets, which are GADM and Global Administrative Unit Layers (GAUL). For Admin-1 level, **t**he GADM was chosen over GAUL because of the presence of textual IDs and the unique identification or HASC codes in the GADM shape file, which is essential for joining tabular data to spatial data. In addition to this advantage, for some countries GADM represents more recently updated boundaries than GAUL. For Admin-2 level, GADM was used where national sources were lacking spatial boundary data. With the help of colleagues in DECDG, we are producing a key that will allow for easier conversion between GADM and GAUL administrative boundary codes.

HDD software. The HDD is stored in Stata as well as GIS data formats, and can be readily used with these two software packages for statistical and graphical analysis. Other software packages such as R, SPSS and QGIS also easily read and load the dataset.

2. The HDD dimensions and indicators

HDD captures the poverty and human development dimension with an array of measurement and relates them to a suite of different environment and natural resource (ENR) measures at the subnational level. With a global scale and subnational units of analysis, HDD has more than 100 indicators and 3,000 observations at the admin-1 level. Tables 1 to 7 below contain an overview of all poverty, human development, and environmental measures included in the HDD dataset. The tables contain a detailed description of the respective indicators used and information about their unit of analysis for which it is available (some data is only available on the admin1 level, other data also on the admin2 level). We created two different versions of admin1 and admin2 level subnational boundaries which represent the delineations for our subnational data. We compute the ENR measures corresponding to this amalgamate of country-specific subnational boundaries.

Figure 1: The subnational measures of poverty

Indicator Category	Indicator definitions	Unit of Analysis Admin 1 = province Admin 2 = district	Country Coverage	Source
Monetary Poverty	Share of people below the four poverty lines $1.90, $3.20, $5.50 and $21.70 (2011 int'l PPP conversion factors)	Admin 1	146 countries (In GMD, there are 1,616 observations in 122 countries at the admin-1 level; at admin-2 level, there are 100 observations in 9 countries	WBG
	Number of persons below the four poverty lines $1.90, $3.20, $5.50 and $21.70 (2011 int'l PPP conversion factors)			
Multidimensional Poverty	Multidimensional Poverty Index = headcount ratio * intensity of deprivation, 0 to 1	Admin 1	66 countries	OPHI/UNDP
	% population in multidimensional poverty			
	Intensity of deprivation among the poor, average % of weighted deprivations			
Vulnerability	% population that are destitute			
	% of MPI poor who are destitute, % of population			
	Inequality among the poor, 0 to 1			
	% population vulnerable to poverty (intensity between 20-33%)			
	% population in severe poverty (intensity > 50%)			
	Number of MPI poor people, thousands			
Health	Child mortality			
	Nutrition deprivation			
Education	Schooling deprivation			
	Child school attendance deprivation			
Living standards	Electricity deprivation			
	Improved sanitation deprivation			
	Drinking water deprivation			
	Flooring deprivation			
	Cooking fuel deprivation			
	Asset ownership deprivation			

Figure 2 The subnational measures of socioeconomic indicators

Indicator Category	Indicator definitions	Unit of Analysis Admin 1 = province Admin 2 = district	Country Coverage	Source
GDP	Estimate of gross domestic product (gridded)	Admin 1 & 2	Global	UNEP GRID
Population	Estimate of number of people per grid cell	Admin 1 & 2	Global	LandScan
Health	Infant mortality rate	Admin 1 & 2	81 countries	CIESIN
	Infant, under-5, and maternal mortality rates	Admin 1		DHS
	Children received 8 basic vaccinations			
	Children stunted			
	Children wasted			
	Children underweight			
	Total fertility rate, 15-49			
	problems in accessing health care			
	prevalence of malaria in children			
	Women: Ever experienced physical violence since age 15			
	Children age 5-14 involved in child labor			
Standard of Living	% households with electricity			
	Population living in households using an improved water source			
	Population living in households with water 30 minutes or longer away round trip			
	Population living in households with improved, non-shared toilet facilities			
	Households using solid fuel for cooking			
	Households owning agricultural land			
	Households owning farm animals			
	Wealth index Gini coefficient			
	Land ownership: women and men, alone and jointly			
Education	% women and men who are literate			
	Gross secondary school attendance rate: female & male			

Figure 3 The subnational measures of ENR: Green

Indicator	Indicator definitions	Unit of Analysis Admin 1 = province Admin 2 = district	Country Coverage	Spatial Resolution	Year	Source
Forest coverage and loss	Total area of district Forest Cover (2000) Area Forest Cover Loss (2000-2012) Area Forest Cover Loss Area, Average Annual (2000-2012) Forest Cover Percentage (2000) Forest Cover Loss Percentage (2000-2012) Forest Cover Loss Percentage, Average Annual (2000-2012)	Admin 1 & 2	Global	30m x 30m	2000 - 2012	Hansen et al., 2013
Soil degradation	Global yearly erosion rates according to the RUSLE model	Admin 1 & 2	Global	5 arcmin	Modeled based on data from various years	Naipal 2015
Vegetation Index	NDVI Normalized Difference Vegetation Index	Admin 1 & 2	Global	0.1 degrees	2001 (2000-2002), 2005 (2003-2007), 2010 (2008-2012), 2015 (2013-2016)	NEO (MODIS/Terra)
Net Primary Productivity	NPP Net Primary Productivity	Admin 1 & 2	Global	0.1 degrees	2001 (2000-2002), 2005 (2003-2007), 2010 (2008-2012), 2015 (2013-2016)	NEO (MODIS/Terra)
Biodiversity	Area of biodiversity hotspot Share of area that is a biodiversity hotspot	Admin 1 & 2	Global	shapefile	2011	Biodiversity Hot Spots (2011): IBAT and Conservation International
Bird areas	Endemic bird area size Share of area that is an endemic bird area	Admin 1 & 2	Global	shapefile	1998	Endemics Bird Areas (1987-1998): IBAT
Protected areas	Protected area size Share of area that is a protected area	Admin 1 & 2	70 countries	shapefile	2014	Protected Areas (2014): World Database on Protected Areas (WDPA) (April 2014)

Figure 4 The subnational measures of ENR: Brown

Indicator	Indicator definitions	Unit of Analysis Admin 1 = province Admin 2 = district	Country Coverage	Spatial Resolution	Year	Source
Air pollution (AP)	PM2.5 air pollution, 2010, satellite	Admin 1 & 2	Global	0.1 degrees x 0.1 degrees	2000 - 2013	CIESIN, 2015
	PM2.5 air pollution, all years, satellite			0.1 degrees x 0.1 degrees	2000 - 2013	
	PM2.5 air pollution, 2013			0.1 degrees x 0.1 degrees	2013	Brauer et al., 2016
	PM2.5 air pollution, all composition, 2013			0.01 degrees x 0.01 degrees	2013	van Donkelaar et al., 2016
	PM2.5 air pollution, all composition, pov. year			0.01 degrees x 0.01 degrees	2000-2013	
	PM2.5 air pollution, dust- and salt-removed, 2013			0.01 degrees x 0.01 degrees	2013	
	PM2.5 air pollution, dust- and salt-removed, pov. year			0.01 degrees x 0.01 degrees	2000-2013	
	PM2.5 air pollution in cities (from monitoring stations)			Point data	2016	WHO AAP (2016)
	PM2.5 air pollution, emissions			0.5 decimal degrees x 0.5 decimal degrees	2010	
	PM10 air pollution			0.5 decimal degrees x 0.5 decimal degrees	2010	
	Black Carbon			0.5 decimal degrees x 0.5 decimal degrees	2010	
	NO x			0.5 decimal degrees x 0.5 decimal degrees	2010	
	Mercury			0.5 decimal degrees x 0.5 decimal degrees	2000	

Figure 5 The subnational measures of ENR: Blue

Indicator	Indicator definitions	Unit of Analysis Admin 1 = province Admin 2 = district	Country Coverage	Spatial Resolution	Year	Source
Reefs at Risk (REEF)	Area of district's shoreline classified as low, medium, high & very high threat Share of district's shoreline classified as very high threat level Share of district's shoreline classified as very high & high threat level	Admin 1 & 2	70 countries	500m	2011	World Resources Institute (WRI), Reefs at Risk Revisited, 2011
Degradation of Marine Ecosystems (COAST)	CI: Ocean Health Index/Cumulative Impact Score	Admin 1 & 2	Global	1km^2	Based on data from various years	Halpern et al (2015)
	Land-based Fertilization			modeled 1km^2	2007-2010	
	Land-based Pesticides			modeled 1km^2	2007-2010	
	Land-based Inorganic Pollution			modeled 1km^2	2000-2001	
	Land-based Human Pollution			modeled 1km^2	2011	
	Land-based Night Lights			1km^2	2007-2009	
	Fishing (Demersal Destructive)			half-degree	2011	
	Fishing (Demersal Nondestructive High Bycatch)			half-degree	2011	
	Fishing (Demersal Nondestructive Low Bycatch)			half-degree	2011	
	Fishing (Pelagic High Bycatch)			half-degree	2011	
	Fishing (Pelagic Low Bycatch)			half-degree	2011	
	Fishing (Artisinal)			modeled 1km^2	2006	
	Climate Change Sea Surface Temperature (SST)			~21km^2	2005-2010	
	Climate Change UV			1 degree	2008-2012	
	Climate Change ocean Acidification			1 degree	1870 vs. 2000-2009	
	Climate Change Sea Level Rise (SLR)			modeled 0.25 degree	1992-2012	
	Ocean Shipping			0.1 degree	2003-2011	
	Ocean Invasive Species			modeled 1km^2	2011	
	OPOLL: Ocean Pollution (pressure)-combination of commercial shipping & invasive species			modeled 1km^2	2003-2011	
	Ocean Oil Rigs			1km^2	2007-2010	
	Climate Change Sea Level Trend (SLT)			0.25 degree	1993-2016	Nicholls and Cazenave
Water stress	Groundwater stress (ratio of groundwater withdrawal relative to recharge rate over a given aquifer)	Admin 1 & 2	Global	Polygons	1958-2000	T. Gleeson, Y. Wada, M.F. Bierkens, and L.P. van Beek (2012)
	Flood occurrence (number of floods recorded)			Flood extent polygons (multiple scales)	1985-October 2011	G.R. Brakenridge, Dartmouth Flood Observatory, University of Colorado (2011)
	Drough severity (average length x dryness of droughts)			1 degree	1901-2008	J. Sheffield and E.F. Wood (2008)
	Threatened Amphibian			Polygons	2010	IUCN (2010)
Mangroves	Global distribution of modelled mangrove biomass	Admin 1 & 2	Global	Polygons	2014	Hutchison J, Manica A, Swetnam R, Balmford A, Spalding M (2014) Predicting global patterns in mangrove forest biomass. Conservation Letters 7(3): 233–240. doi: http://dx.doi.org/10.1111/conl.12060; Data URL: http://data.unepwcmc.org/datasets/39

Figure 6 The subnational measures of Geography and Climate

Indicator	Indicator definitions	Unit of Analysis Admin 1 = province Admin 2 = district	Country Coverage	Spatial Resolution	Year	Source
Land cover	Land cover types: cropland, rainfed; cropland, irrigated; mosaic cropland/vegetation; mosaic vegetation/ cropland; tree broadleaved evergreen; tree broadleaved deciduous; tree needleleaved evergreen; tree needleleaved deciduous; tree mixed leaved type; mosaic tree, shrub/hc; mosaic hc/tree, shrub; shrubland; grassland; lichens and mosses; sparse vegetation; tree flooded, fresh water; tree flooded, saline water; shrub or herbaceous flooded; urban areas; bare areas; water bodies; permanent snow and ice	Admin 1 & 2	Global	300m x 300m	2000 (1998-2002), 2005 (2003-2007), 2010 (2008-2012)	CCI
	Land Cover Total Area					
Grid cell ruggedness	Ruggedness index	Admin 1 & 2	Global	30 x 30 arcsecond grid	1999	Nunn and Puga, 2012
Less Favored Agricultural Land (LFAL)	LFAL % of total area	Admin 1 & 2	Global	0.1 degree	2000/2005	Barbier & Hochard (2014); Ramankutty et al. (2008); Siebert et al. (2006); FAO Global Agro-Ecological Zones Data Portal version 3; Sebastian (2007)
	LFAL % of Agricultural Land in area					
	MFAL % of total area					
	MFAL % of Agricultural Land in area					
	Agricultural extent % of total area					
Area size	Size of the administrative boundary	Admin 1 & 2	Global	NA	NA	GADM
Market Access	Travel time to major city	Admin 1 & 2	Global	0.01 degree	2000	Nelson & Uchida, 2008
Road density	Road Density, total	Admin 1 & 2	Global	10km	2013	PBL
	Road Density, highways					
	Road Density, primary					
	Road Density, secondary					
	Road Density, tertiary					
	Road Density, urban/residential					
Built-up area	Built-up area %			2.8 arc seconds (75-84m)	2011	Global Urban Footprint (GUF); DLR 2016
Temperature	Temperature mean (of decade)	Admin 1 & 2	Global	0.5 degree	1901-2015	CRU Climate Research Unit
Rainfall	Precipitation total (of decade)					

Figure 7 The subnational measures of Mining and Energy

Indicator	Indicator definitions	Unit of Analysis Admin 1 = province Admin 2 = district	Country Coverage	Spatial Resolution	Year	Source
Diamond deposits	Known locations of diamond deposits	Admin 1 &2	Global	Point data	2005	PRIO (2005)
Onshore oil and gas	Known locations of onshore oil and gas deposits	Admin 1 &2	Global	Point data	2005	PRIO (2005)
Power plants	Locations of power plants	Admin 1 &2	Global	Point data	2017	WRI and Google Earth Outreach (2017)

3. Subnational monetary poverty measures

In this section, we provide background to subnational poverty measurement, and describe in detail how we create the global subnational monetary poverty map.

The World Bank possesses a significant collection of official household survey data, covering nearly all developing countries in the world. To make better use of the available data and allow for automated analysis across many districts, the Global Monitoring Database (GMD) was developed. The GMD combines household survey data for key indicators, including official measures of poverty (at national and subnational levels) along with individual and household characteristics. These data provide crucial inputs for the bank to assess the global development processes towards the twin goals and help to improve the quality of Bank's diagnostics and policy advice.

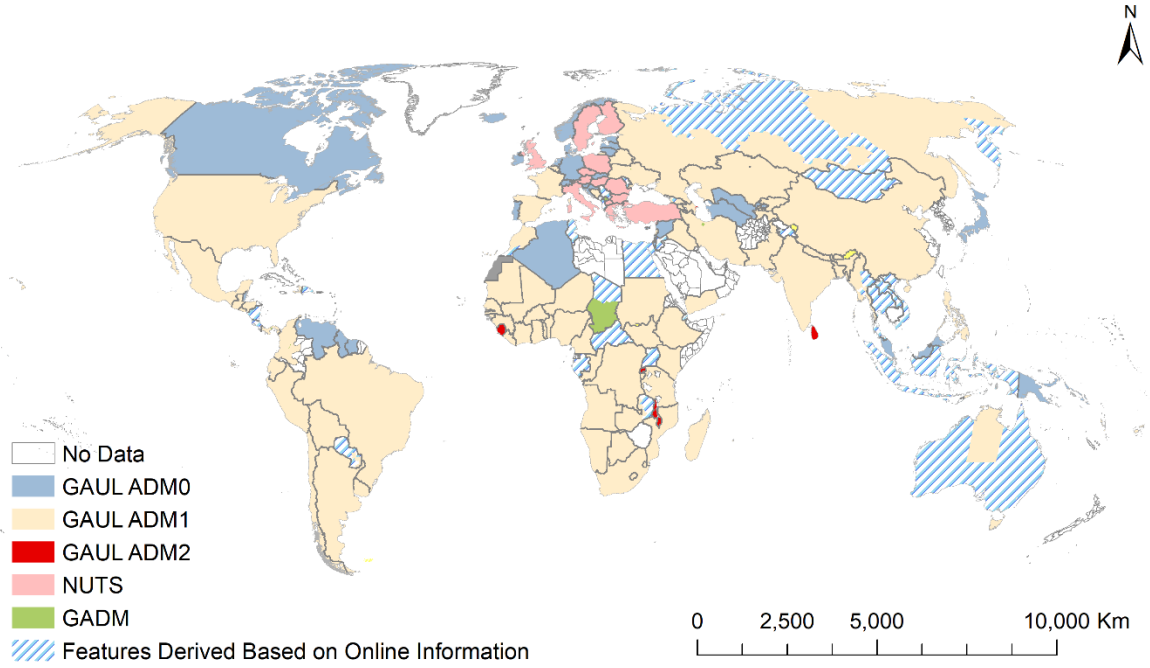
But up to now, the use of the geographic information and geo-reference of the survey data in poverty analysis has not been well explored in the bank. Related poverty mapping work, such as the small area estimations, has been done for a few regions (e.g. South Asia) (World Bank, 2016) and countries (e.g. Bhutan) (World Bank, 2010), but the efforts so far remain partial and disconnected from each other. To fill the gap, we collaborated with colleagues in the Poverty Global Practice, who led an effort to develop a global geodatabase (cross-comparable map) of poverty.

The geodatabase, built using ESRI ArcGIS, Stata, and Microsoft Excel software, was designed to be a single, comprehensive repository of spatial and non-spatial information. The global cross-comparable map draws mainly on the Global Monitoring Database. The poverty data measured at the new 2011 international purchasing power parity conversion factors (PPPs) are produced at four poverty lines, at $1.90, $3.20, $5.50 and $21.70. A detailed description on the choice of these poverty lines suggested by the World Bank was provided by Ferreira et al. (2015). Datasets are stored as features (shape files) and relational tables, with a unique identifier for districts linking the two. This code was produced based on administrative units of the poverty data and is stored in the geodatabase feature class.

The administrative units of the poverty data differ across countries. To ensure the consistency between spatial boundaries across countries, it was imperative to use the same administrative boundaries. The spatial feature class of the geodatabase was established based on the United Nations FAO's Global Administrative Unit Layers (GAUL). The GAUL set contains three administrative levels, states (ADM0), districts (ADM1), and sub-districts (ADM2). But boundaries in GAUL are not always available to map the survey data which are only representative at their respective administrative units. New features were created by merging or dissolving parts of the existing units according to the survey derived administrative units. In the geodatabase, there are 41 countries with poverty data at national level. At the admin-1 level, there are 1,616 observations in 122 countries. At the admin-2 level, we have 100 observations in

9 countries. A total of 267 observations in 54 countries were linked to the created features in the geodatabase (Figure 1).

Figure 8 The administrative units of the global poverty map



The temporally adjusted poverty estimations are another spotlight feature of the geodatabase. To gain the knowledge of poverty situations at a specific time, the best possible way is to use surveys conducted in the year closest to the chosen time, which is year 2013 in this case. The availability of survey data, however, differs across countries, resulting in poverty estimations in various years. A "line-up" process was conducted to bring poverty indicators from all available countries at subnational levels to the 2013 values. This global line-up approach has never been conducted at levels lower than the national. The estimated lined up estimations would provide a 2013 global subnational poverty map and allow readers and users of the geodatabase to compare districts across countries.

For more information on the global poverty geodatabase, please refer to the working paper titled Global Subnational Poverty: An illustration of the Global Geodatabase of Household Surveys (forthcoming, 2019).

4. Alternative poverty indicators and other human development measures

4.A. Multidimensional Poverty Index (MPI)

We also included other measures of poverty such as the subnational Multidimensional Poverty Index (MPI), as well as measures that are highly correlated with poverty such as infant mortality and birthweight. We included other measures of human development such as subnational GDP and population. For these other measures of poverty and more broadly human development, we were able to compile data for many more countries, albeit sometimes at the detriment of fineness of the boundary observations. The alternative measures of poverty came in tabular format originally and we georeferenced them and mapped them in GIS.

What ensues in this section are descriptions of the alternative poverty measures and human development measures:

Measures and their definitions:

Variable name	Description
Multidimensional Poverty Index, MPI	Multidimensional Poverty Index = Headcount ratio * Intensity of deprivation, 0 to 1
MPI headcount ratio	Population in multidimensional poverty, % Population
Intensity of deprivations among the poor	Intensity of deprivation among the poor, Average % of weighted deprivations
Destitutes	% of Population that are Destitute
Proportion of MPI poor who are destitutes	% of MPI poor who are destitutes, % of population
Inequality among the poor	Inequality among the poor, 0 to 1
Population vulnerable to poverty	Population vulnerable to poverty (intensity between 20-33%), % pop
Population in severe poverty	Population in severe poverty (intensity > 50%), %pop
Number of MPI poor people	Number of MPI poor people, thousands

Explanation:

The multidimensional poverty index (MPI) is an alternative index (an alternative to the monetary poverty measure) measuring global poverty. The MPI complements income-based poverty measures by capturing the deprivations that each person experiences with respect to education, health and living standards.

The MPI is comprised of 10 sub-indicators, pertaining to three equally weighted dimensions: health, education and living standards. These dimensions are the same as those used in the

Human Development Index (HDI), and the MPI was modelled after the HDI. A person is identified as multi-dimensionally poor (or 'MPI poor') if they are deprived in at least one of the three dimensions. The MPI is calculated by multiplying the incidence of poverty (the percentage of people identified as MPI poor or *headcount ratio*) by the average intensity of poverty across the poor (*Average deprivation share*). For a more detailed description of the MPI, see the Oxford – OPHI website (http://www.ophi.org.uk/policy/multidimensional-poverty-index/) and the UNDP website (http://hdr.undp.org/en/content/multidimensional-poverty-index-mpi). For a brief glimpse into the dimensions and indicators of the MPI see table 6.

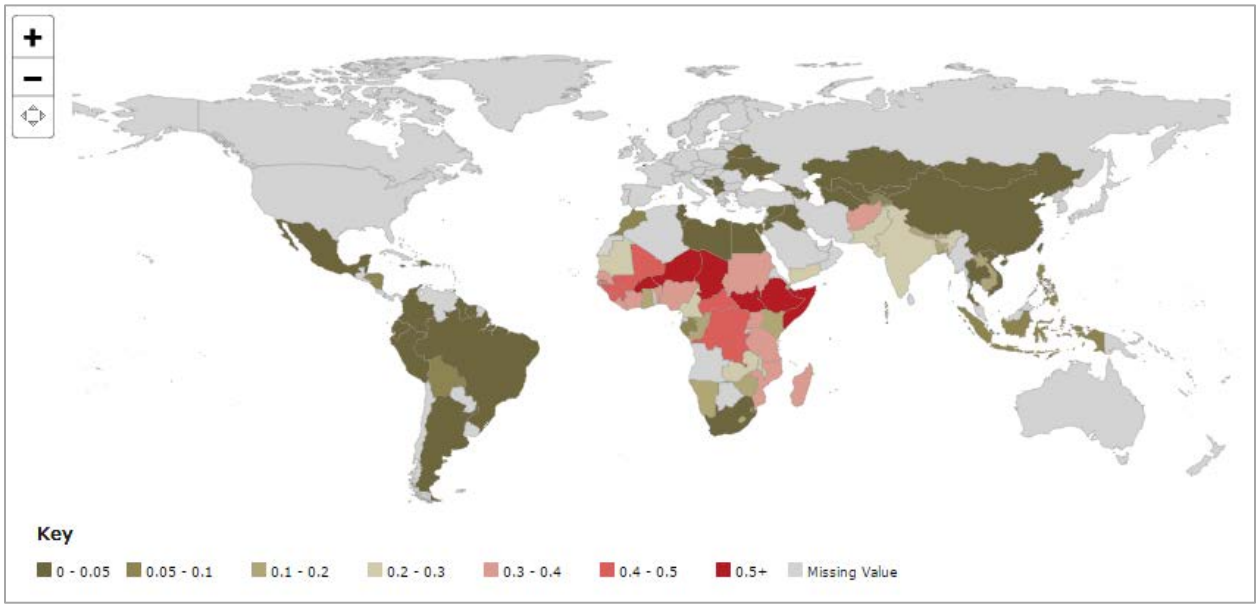
Figure 9 Summary of MPI dimensions, indicators, thresholds and relative weights.

Dimension	Indicator	Deprived if…	Related to…	Relative Weight
Education	Years of Schooling	No household member aged 10 years or older has completed five years of schooling.	MDG2	1/6
	Child School Attendance	Any school-aged child is not attending school up to the age they'd finish class 8.*	MDG2	1/6
Health	Child Mortality	Any child has died in the household.	MDG4	1/6
	Nutrition	Any adult under 70 years of age or any child for whom there is nutritional information is malnourished.*	MDG1	1/6
Living Standard	Electricity	The household has no electricity.		1/18
	Improved Sanitation	The household's sanitation facility is not improved (according to MDG guidelines), or it is improved but shared with other households.**	MDG7	1/18
	Safe Drinking Water	The household does not have access to safe drinking water (according to MDG guidelines) or safe drinking water is at least a 30-minute walk from home roundtrip.***	MDG7	1/18
	Flooring	The household has a dirt, sand, dung or other (unspecified) type of floor.		1/18
	Cooking Fuel	The household cooks with dung, wood, charcoal or other solid fuels	MDG7	1/18
	Assets ownership	The household does not own more than one radio, TV, telephone, bike, motorbike or refrigerator and does not own a car or truck.	MDG7	1/18

Source: Alkire and Santos (2010, 2014).

The MPI has two particular advantages over the monetary poverty measures. First it has a broader global coverage (see figure 2), including for example also countries such as Brazil, China, and South Africa. Second, the MPI measures are comparable cross-country, which the current measures of monetary poverty are not.

Figure 10 Global MPI country coverage



Source: Global MPI Interactive Databank, OPHI.

The Global MPI was developed in 2010 by OPHI at the University of Oxford and the United Nations Development Programme for its flagship Human Development Report. The Global MPI uses DHS, MICS, and national survey data, from 2005-2014, as the raw input data.

Sources:

Alkire, S. and Santos, M.E., 2010. Multidimensional poverty index: 2010 data.

Alkire, S. and Santos, M.E., 2014. Measuring acute poverty in the developing world: Robustness and scope of the multidimensional poverty index. World Development, 59, pp.251-274.

Alkire, Sabina, Adriana Conconi, Gisela Robles, José M. Roche, María Emma Santos, Suman Seth and Ana Vaz. 2015. "The Global Multidimensional Poverty Index (MPI): 5-year methodological note". *The Oxford Poverty and Human Development Initiative (OPHI).*

United Nations Development Program. 2015. "Human Development Reports: Multidimensional Poverty Index (MPI)". http://hdr.undp.org/en/content/multidimensional-poverty-index-mpi.

Oxford Poverty and Human Development Initiative (OPHI). 2016. http://www.ophi.org.uk/multidimensional-poverty-index/.

4.B. Sociodemographic Variables

i. Population

Measures and their definitions:

Estimate of number of people per grid cell	2012 global population distribution (raw, LandScan grid)
Population count, UN-adjusted	Estimates of population count for the years 2000, 2005, 2010, 2015, and 2020, based on national censuses and population registers, adjusted to match United Nations country totals.

Explanation:

Gridded Population of the World, Version 4 (GPWv4) dataset estimates human population count, based on counts consistent with national censuses and population registers with respect to relative spatial distribution, but adjusted to match the 2015 Revision of UN World Population Prospects country totals for the years 2000, 2005, 2010, 2015, and 2020. It employs a proportional allocation gridding algorithm, using12.5 million national and sub-national administrative units, with a resolution of 30 arc-second (~1 km) grid cells. For more information, see here.

Using Geographic Information System and Remote Sensing, Oak Ridge National Laboratory's LandScan estimates global population distribution. The data is available at approximately 1 km resolution (30 arc-seconds) and represents an ambient population (average over 24 hours). The dataset uses a multi-variable dasymetric modeling approach to disaggregate census counts within an administrative boundary. For more information, see here.

Sources:

Oak Ridge National Laboratory. LandScan 2012.

Center for International Earth Science Information Network - CIESIN - Columbia University. 2016. Gridded Population of the World, Version 4 (GPWv4): Population Density Adjusted to Match 2015 Revision UN WPP Country Totals. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). http://dx.doi.org/10.7927/H4HX19NJ.

ii. Demographic and Health Surveys (DHS) Data

As Kofi Annan stated, "the biggest enemy of health in the developing world is poverty." Inadequate sanitation, healthcare, nutrition, and many other factors result in a perpetual feedback loop of poverty and health, where a lack of resources leads to poor health conditions and insufficient access to health services, which in turn perpetuate poverty. Because of this relationship, health variables are crucial pieces to consider when understanding poverty, and vice versa.

Child health data is gathered from The Spatial Data Repository of the Demographic and Health Surveys (DHS) Program, except where noted. This repository provides subnational data on indicators

related to health and human development. Most indicator data gathered from the DHS are presented as percentages of a studied population, with median ages of marriage and breastfeeding as exceptions. DHS surveys are conducted in order to be nationally representative, and generally survey 5,000-30,000 households per country. Women age 15-49 in all surveyed households are eligible to participate, while a sub-sample of men age 15-54 (or 59) are also eligible to participate. Spatial data for DHS surveys is available at the Administration 1 level of our analysis.

The DHS Program has subnational data for the indicators included in HDD for 81 countries, for survey years 1987-2015, and is focused largely on the SSA region (see Table 1). Data is downloaded as a shapefile from the Spatial Data Repository, which provides geographically linked health and demographic data from The DHS Program and the U.S. Census Bureau for mapping in a geographic information system (GIS).

Figure 11 Countries and Regional Distribution of DHS Spatial Data

EAP	ECA	LAC	MENA	SA	SSA
Cambodia	Albania	Bolivia	Egypt	Bangladesh	Angola
Indonesia	Armenia	Brazil	Jordan	India	Benin
Philippines	Azerbaijan	Colombia	Morocco	Maldives	Burkina Faso
Thailand	Kazakhstan	Dominican Rep.	Tunisia	Nepal	Burundi
Timor-Leste	Kyrgyz Rep.	Ecuador	Yemen	Pakistan	Cameroon
Vietnam	Moldova	Guatemala		Sri Lanka	Central African Rep.
	Tajikistan	Guyana			Chad
	Turkey	Haiti			Comoros
	Turkmenistan	Honduras			Congo
	Ukraine	Mexico			Congo, Dem. Rep.
	Uzbekistan	Nicaragua			Cote d'Ivoire
		Paraguay			Eritrea
		Peru			Ethiopia
					Gabon
					Gambia
					Ghana
					Guinea
					Kenya

					Lesotho
					Liberia
					Madagascar
					Malawi
					Mali
					Mauritania
					Mozambique
					Namibia
					Niger
					Nigeria
					Rwanda
					Sao Tome & Principe
					Senegal
					Sierra Leone
					South Africa
					Sudan
					Swaziland
					Tanzania
					Togo
					Uganda
					Zambia
					Zimbabwe
6	**11**	**13**	**5**	**6**	**40**

a. Child Health

Measures and their definitions:

Infant mortality rate[1]	Probability of dying before the first birthday (in the period 1-60 months [1-120 months for background characteristics] preceding the survey) per 1,000 live births; CIESIN: The number of children who die before their first birthday for every 1,000 live births for the year 2000
Under-five mortality	DHS: Probability of dying before the fifth birthday (in the period 1-60 months

rate	[1-120 months for background characteristics] preceding the survey) per 1,000 live births
Children stunted	Percentage of children stunted (-2 SD) [WHO standard]
Children underweight	Percentage of children underweight (-2 SD) [WHO standard]
Children wasted	Percentage of children wasted (-2 SD) [WHO standard]
Births in a health facility	Percentage of live births in the three/five years preceding the survey delivered at a health facility
Breastfeeding duration	Median duration of exclusive breastfeeding (months)
Childhood vaccination rate	Percentage of children 12-23 months who had received all vaccinations
Children with ORS/RHS diarrhea treatment	Percentage of children born in the three/five years preceding the survey with diarrhea in the two weeks preceding the survey who received either oral rehydration salts or recommended home fluids
Children's bednet use	Percentage of children under 5 who slept under an insecticide treated net (ITN) the night before the survey
Women's marriage age	Median age at first marriage among women age 25-49

[1] Infant mortality figures are included from two data sources in the HDD: the Demographic and Health Surveys Program (DHS, 2016) and the Center for International Earth Science Information Network (CIESIN, 2005).

Explanation:

Infant and child mortality rates are indicators of premature death, which are often linked with infectious disease and neonatal complication (WHO 2015), and are presented as rates per 1,000 live births. These mortality rates are estimated through a synthetic cohort life table approach, estimating mortality probability based on real cohorts for small age groups. Deaths of infants and under-five children (0-12 months and 0-59 months, respectively) are estimated for specific age ranges and time periods, and then standardized by the number of live births at the beginning of the time period.

Data on the rate of births in a health facility, childhood vaccination rates, children with diarrhea, and children's use of bednets provide information related to the prevention and prevalence of major causes of infant and child death. Births in a health facility are calculated based on the birth location reported for live births to interviewed women in the three or five years preceding the survey (varying based on survey). Childhood vaccination is reported by respondents or acquired from vaccination cards and the rate is calculated based on number of living children in the population. Children's diarrheal treatment was calculated for children receiving oral rehydration salts (ORS) packets,

recommended home fluids (RHF), increased fluids, or a combination of these treatments. Insecticide-treated bednet use is included in DHS as a malaria-related indicator.

Rates of children that are stunted, underweight, or wasted show lack of access to nutrition and basic requirements for a healthy life. Children under five are considered in this data based on stunting (height for age z-score is two standard deviations below the mean), wasting (weight for age z-score is two standard deviations below the mean), and underweight (weight for age z-score is two standard deviations below the mean).

Two additional variables are included due to their impacts on childhood development. Mean duration of breastfeeding has short- and long-term impacts on child development and wellbeing, including potential impacts on diarrhea and respiratory infection prevalence (Horta & Victora, 2013a, b). Mean duration of exclusive breastfeeding is calculated using the sum of smoothed proportions of age groups breastfeeding for children born 0-35 months before the survey.

The median age of marriage for women is included as a child health measure due to the ongoing threat of child marriage, which disproportionately impacts girls and women worldwide and preempts lifelong consequences to education, childbirth, and health access (UNICEF 2014). Age of first marriage is considered by DHS as the age when a woman began living with their partner.

Sources:

Horta, B. L. & Victora, C.G. 2013. Short-Term Effects of Breastfeeding: A Systematic Review on the Benefits of Breastfeeding on Diarrhoea and Pneumonia Mortality. WHO. http://apps.who.int/iris/bitstream/10665/95585/1/9789241506120_eng.pdf?ua=1

Horta, B. L. & Victora, C.G. 2013. Long-Term Effects of Breastfeeding: A Systematic Review. WHO. http://apps.who.int/iris/bitstream/10665/79198/1/9789241505307_eng.pdf?ua=1

Rutstein, S.O. & Rojas, G. (2006). Guide to DHS Statistics (English). Demographic and Health Surveys & ORC Macro. Produced for review by the United States Agency for International Development (USAID), prepared by MEASURE DHS/ICF International. http://dhsprogram.com/pubs/pdf/DHSG1/Guide_to_DHS_Statistics_29Oct2012_DHSG1.pdf

UNICEF (2014). Ending Child Marriage: Progress and prospects, UNICEF, New York, 2014. https://www.unicef.org/media/files/Child_Marriage_Report_7_17_LR..pdf

UNICEF (2015). Levels & Trends in Child Mortality. http://www.childmortality.org/files_v20/download/IGME%20Report%202015_9_3%20LR%20Web.pdf

Data:
Spatial Data Repository, the Demographic and Health Surveys Program (DHS). ICF International. Funded by the United States Agency for International Development (USAID). Available from spatialdata.dhsprogram.com. [Accessed 12 August 2016]

b. Adult Health

Measures and their definitions:

Fertility rate	Total fertility rate for the three years preceding the survey
Any contraception	Percentage of currently married/in union women currently using any method of contraception
Modern contraception	Percentage of currently married/in union women currently using any modern method of contraception
Age at first sexual intercourse (women)	Median age at first sexual intercourse among women age 25-49
Unmet family planning need	Percentage of currently married/in union women with an unmet need for family planning

Explanation:
DHS data on adult health used in the HDD provide information mainly related to reproduction and HIV prevalence, which are intertwined with factors of development. Fertility rates, for example, have been found to decrease poverty when lowered, allowing for more public support per capita for education and health at the macro level and fewer users of household resources at the micro level (Sinding, 2009). In the DHS, total fertility rate refers to the average number of births a cohort of women would have before age 50 if they gave birth at the current fertility rates for their specific age. This data is presented as the average number of births per woman.

Fertility rates are affected by use of contraception and other family planning strategies. Contraceptive prevalence rates (CPR) are calculated for the population of women age 15-49 who are currently married, in a consensual union, or unmarried and sexually active. Modern methods of contraception include the lactational amenorrhea method, various methods of female and male sterilization, the use of female and male condoms, diaphragms, foams, jellies, and oral, intrauterine, injectable, and implanted contraceptives. Women who are considered to use any method of contraception are those who answered that they use a specific technique or tool to delay or avoid pregnancy. The percentage of married women with an unmet family planning need is calculated based on women that are fecund, not currently using contraception, and either do not want more children, wish to wait at least two years to have their next child, are unsure whether to have another child, want a child but feel it would be problematic in the near future, or are pregnant or recently gave birth and said it was mistimed or unwanted. Additionally, women that are pregnant or recently gave birth despite using contraception are included in this measure.

While studies of the connection between poverty and early sexual behavior have been somewhat inclusive or mixed, previously studied have found evidence for higher vulnerability to sexually transmitted infections in poor women due earlier sexual activity and lower prevalence of condom use (Madise et al., 2007). Age at first sexual intercourse is reported in DHS data as the age when a woman began living with a husband or consensual partner.

Sources:

Madise, N., Zulu, E., & Ciera, J. (2007). Is poverty a driver for risky sexual behaviour? Evidence from national surveys of adolescents in four African countries: original research article. African journal of reproductive health, 11(3), 83-98.

Rutstein, S.O. & Rojas, G. (2006). Guide to DHS Statistics (English). Demographic and Health Surveys & ORC Macro. Produced for review by the United States Agency for International Development (USAID), prepared by MEASURE DHS/ICF International. http://dhsprogram.com/pubs/pdf/DHSG1/Guide_to_DHS_Statistics_29Oct2012_DHSG1.pdf

Sinding, S. W. (2009). Population, poverty and economic development. Philosophical Transactions of the Royal Society of London B: Biological Sciences, 364(1532), 3023-3030. http://rstb.royalsocietypublishing.org/content/364/1532/3023

Data:

Spatial Data Repository, the Demographic and Health Surveys Program (DHS). ICF International. Funded by the United States Agency for International Development (USAID). Available from spatialdata.dhsprogram.com. [Accessed 12 August 2016]

c. Education

Measures and their definitions:

Women's literacy	Percentage of women who are literate
Women with secondary education	Percentage of women with secondary or higher education

Explanation:

Various aspects of education indicate opportunity, particularly for groups in vulnerable positions. Post-elementary education has been shown to have positive impacts on wellbeing through aspects such as individual earnings, absolute and relative poverty, infant mortality, and life expectancy (Tilak, 2007), although increased education alone does not always translate to poverty alleviation (Wedgwood, 2007). Gender equality is a particularly important aspect of educational development, thus women's literacy and secondary school achievement are considered from the DHS data in the HDD. Literacy is defined in the DHS as whether a respondent is able to read a sentence provided by the interviewer, or whether they attended secondary school or higher.

Sources:

Demographic and Health Surveys. Description of the Demographic and Health Surveys Individual Recode Data File: DHS VI. (2013). Version 1.0. Produced for review by the United States Agency for International Development (USAID), prepared by MEASURE DHS/ICF International. http://dhsprogram.com/pubs/pdf/DHSG4/Recode6_DHS_22March2013_DHSG4.pdf

Rutstein, S.O. & Rojas, G. (2006). Guide to DHS Statistics (English). Demographic and Health Surveys & ORC Macro. Produced for review by the United States Agency for International Development (USAID), prepared by MEASURE DHS/ICF International. http://dhsprogram.com/pubs/pdf/DHSG1/Guide_to_DHS_Statistics_29Oct2012_DHSG1.pdf

Tilak, J. B. (2007). Post-elementary education, poverty and development in India. *International Journal of Educational Development*, 27(4), 435-445.

Wedgwood, R. (2007). Education and poverty reduction in Tanzania. *International Journal of Educational Development*, 27(4), 383-396.

Data:

Spatial Data Repository, the Demographic and Health Surveys Program (DHS). ICF International. Funded by the United States Agency for International Development (USAID). Available from spatialdata.dhsprogram.com. [Accessed 12 August 2016]

d. Standard of Living

Measures and their definitions:

Electricity	Percentage of households with electricity

Explanation:

Electrification is considered a driver for economic growth (see, e.g. Apergis & Payne, 2009), and access to electricity is considered a standard step in development, thus, DHS data on electrification at the household level is included in the HDD.

Sources:

Apergis, N., & Payne, J. E. (2009). Energy consumption and economic growth in Central America: Evidence from a panel cointegration and error correction model. *Energy Economics*, 31(2), 211–216. https://doi.org/10.1016/j.eneco.2008.09.002

Demographic and Health Surveys. Description of the Demographic and Health Surveys Individual Recode Data File: DHS VI. (2013). Version 1.0. Produced for review by the United States Agency for International Development (USAID), prepared by MEASURE DHS/ICF International. http://dhsprogram.com/pubs/pdf/DHSG4/Recode6_DHS_22March2013_DHSG4.pdf

Data:
Spatial Data Repository, the Demographic and Health Surveys Program (DHS). ICF International. Funded by the United States Agency for International Development (USAID). Available from spatialdata.dhsprogram.com. [Accessed 12 August 2016]

iii. CIESIN Infant Mortality Data

Explanation:
The Global Subnational Infant Mortality Rates consists of estimates of infant mortality rates for the year 2000. The infant mortality rate is defined as the number of children who die before their first birthday for every 1,000 live births. This data set is produced by the Columbia University Center for International Earth Science Information Network (CIESIN). For more information, see here.

Data:
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). http://dx.doi.org/10.7927/H4PZ56R2.

iv. Gross Domestic Product (GDP)

Explanation:
In the gridded global GDP dataset, sub-national Gross Regional Product (GRP) and national Gross Domestic Product (GDP) data are allocated to 30 arc second (approximately 1km) grid cells in proportion to the population residing in that cell. The method distinguishes between rural and urban population, assuming the latter to have a higher GDP per capita. The gridded GDP dataset is particularly useful for this project as its organization in grid cells allow us to carry out spatial analyses using GDP without having to change the spatial resolution of the dataset.
Input data are from 1) a global time-series dataset of GDP, with subnational gross regional product (GRP) for 74 countries, compiled by the World Bank Development Economics Research Group (DECRG). 2) Gridded population projections for the year 2009, based on a population grid for the year 2005 provided by LandScanTM Global Population Database (Oak Ridge, TN: Oak Ridge National Laboratory). This dataset has been extrapolated up to year 2010 by UNEP/GRID-Geneva.

As a note of caution, cell level anomalies may occur due to poor alignment of multiple input data sources, and it is strongly recommended that users attempt to verify information, or consult original sources, in order to determine suitability for a particular application.

The data is downloaded as raster GeoTiffs.

Data Source:

UNEP-GRID and UNISDR. 2013. “Global Risk Data Platform: GDP”. http://preview.grid.unep.ch/index.php?preview=home&lang=eng.

Credits for GIS processing: World Bank DECRG, Washington, DC, and extrapolation: UNEP/GRID-Geneva.

5. Environment and Natural Resources (ENR) indicators

The description of the ENR measures follows the World Bank, ENB GP business lines, i.e. Green, Brown and Blue. We also add the geography category.

There are no country limitations with any of our ENR measures and in fact we have full coverage for those as most of them are recorded by satellites, which are available everywhere and do not have any country-based sampling bias.

5.A. Green indicators

i. **Forest cover and forest cover loss**

Measures and their definitions:

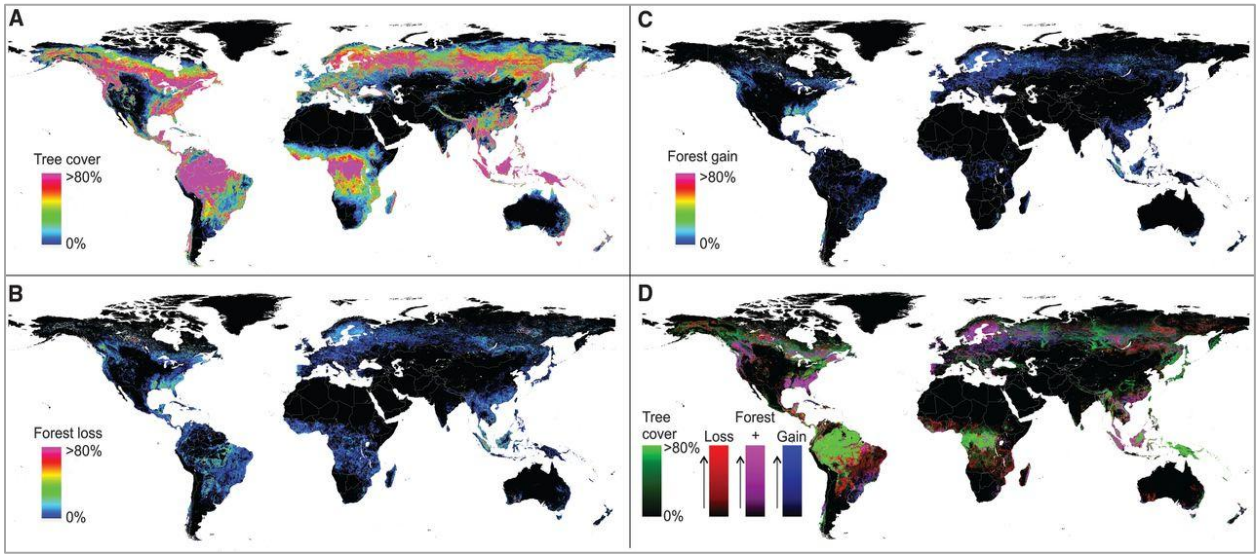
Variable name	Description
Tree Cover (2000) Area	Tree Cover in District (2000), Area (km2)
Forest Cover Loss (2000 - 2012) Area	Forest Cover Loss within District (2000 - 2012), Area (km2)
Forest Cover Loss Area, Average Annual (2000-2012)	Forest Cover Loss, Average Annual, in District (2000 - 2012), Area (km2)
Forest Cover Percentage (2000)	Forest Cover in District (2000), %
Forest Cover Loss Percentage (2000-2012)	Forest Cover Loss within District (2000-2012), %
Forest Cover Loss Percentage, Average Annual (2000-2012)	Forest Cover Loss, Average Annual, in District (2000-2012), %

Explanation (adapted from Hansen et al., 2016: see here):

The tree cover change data used in this project includes tree cover in 2000, tree cover loss 2000-2012 and tree cover gain 2000-2012. This dataset is a collaboration between the GLAD (Global Land Analysis & Discovery) lab at the University of Maryland, Google, USGS, and NASA. This dataset measures areas of tree cover loss across all global land (except Antarctica and other Arctic islands) at approximately 30 × 30 meter resolution. The data were generated using multispectral satellite imagery from the Landsat 5 thematic mapper (TM), the Landsat 7 thematic mapper plus (ETM+), and the Landsat 8 Operational Land Imager (OLI) sensors. Over 1 million satellite images were processed and analyzed, including over 600,000 Landsat 7 images for the

2000-2012 interval and approximately 400,000 Landsat 5, 7, and 8 images for updates for the 2011-2014 interval. For our research purposes, the 2000-2012 data is used for consistency of methods.

Figure 12(A) Tree cover, (B) forest loss, and (C) forest gain. A color composite of tree cover in green, forest loss in red, forest gain in blue, and forest loss and gain in magenta is shown in (D), with loss and gain enhanced for improved visualization. satellite imagery and methodology.



Note: All map layers have been resampled for display purposes from the 30-m observation scale to a 0.05° geographic grid.
Source: Hansen et al, *Science* 2016.

Tree cover (2000, Hansen/UMD/Google/USGS/NASA)
"Tree cover in the year 2000 is defined as canopy closure for all vegetation taller than 5m in height, encoded as a percentage per output grid cell, in the range 0–100 and at 30 x 30 meter resolution. Tree cover canopy (TCC) density is the estimated percent of a pixel that was covered by tree canopy. For tree cover data, pixels with greater than 30% TCC density are classified as tree cover."

""Tree cover" is the biophysical presence of trees and may take the form of natural forests or plantations existing over a range of canopy densities."

Description from Hansen et al (2013).

Tree cover loss (annual, 30m, global, Hansen/UMD/Google/USGS/NASA)
"Tree cover loss is defined as "stand replacement disturbance," or the complete removal of tree

cover canopy at the Landsat pixel scale. Tree cover loss during the period 2000–2012 is the change from a forest to non-forest state and encoded as either 1 (loss) or 0 (no loss). For the tree cover loss data, TCC density corresponds to the density of tree cover before loss occurred. A 30% threshold set as the minimum TCC density means that tree cover loss is picked up in pixels that have an original tree cover density greater than 30%. This pixel is classified as 1 (loss).

Tree cover loss may be the result of human activities, including forestry practices such as timber harvesting or deforestation (the conversion of natural forest to other land uses), as well as natural causes such as disease or storm damage. Fire is another widespread cause of tree cover loss, and can be either natural or human-induced. As such, "loss" does not equate to deforestation."

From Hansen et al (2013).

Note of Caution on Limitations

Due to different definitions of forest and methodologies to measure forest cover and forest cover loss, we should apply caution when using this dataset. The Hansen et al. / Global Forest Watch (GFW) dataset does not always show consistent findings when compared with some national forest statistics from the FAO's Forest Resources Assessment (FRA) and country government reports. For instance, some countries where the FRA reported to have stable or increasing forest area were reported to be losing forest cover by GFW, including Australia, Canada, China, Finland, Gabon, India, Malaysia, Russia, Sweden and the United States. In addition, Hansen's forest cover loss estimate for Indonesia from 2009-2012 is almost triple that of the Indonesian Ministry of Forestry's national deforestation estimates (17,000 square kilometers/year in GFW compared to 6,500 square kilometers/year). Hence, is important to use this dataset with caution and note that tree cover does not always denote forest areas and may include plantations. Meanwhile, forest cover loss does not always denote deforestation but may also include harvesting, disease or storm damage.

Omitting tree cover gain

While there is data available for tree cover gain, we decided not to use this dataset for this project for a variety of reasons. Tree cover gain captures a set minimum tree cover canopy greater than 50%, meaning that a pixel with 50% or less original tree cover would have more than 50% of its area to be classified as "1" or as forest. Firstly, this 50% threshold of tree cover canopy density is relatively arbitrary and does not correspond with the 30% threshold defined for classifying tree cover loss. Secondly, tree cover gain may indicate several potential activities, including natural forest growth or the crop rotation cycle of tree plantations. Thirdly, the accuracy of the tree cover gain dataset in considerably lower than that of tree cover loss, as seen in the table below.

Figure 13 Accuracy of tree cover loss and gain datasets.

	LOSS		GAIN	
Biome	False Positives	False Negatives	False Positives	False Negatives
Global	**13.0 percent**	**12.2 percent**	**23.6 percent**	**26.1 percent**
Tropical	13.0 percent	16.9 percent	18.1 percent	52.0 percent
Subtropical	20.7 percent	20.6 percent	14.5 percent	17.6 percent
Temperate	11.8 percent	6.1 percent	38.0 percent	23.5 percent
Boreal	12.0 percent	6.1 percent	23.3 percent	1.6 percent

Source: Global Forest Watch, 2015.

Accuracy

For tree cover loss, the authors evaluated the overall prevalence of false positives (commission errors) in this data at 13%, and the prevalence of false negatives (omission errors) at 12%, though the accuracy varies by biome and thus may be higher or lower in any particular location. The model often misses disturbances in smallholder landscapes, resulting in lower accuracy of the data in sub-Saharan Africa, where this type of disturbance is more common. The authors are 75 percent confident that the loss occurred within the stated year, and 97 percent confident that it occurred within a year before or after. Users of the data can smooth out such uncertainty by examining the average over multiple years.

Due to different research methodologies and dates of content, tree cover, loss, and gain data sets cannot be compared accurately against each other. Accordingly, “net” loss cannot be calculated by subtracting figures for tree cover gain from tree cover loss, and current (post-2000) tree cover cannot be determined by subtracting figures for annual tree cover loss from year 2000 tree cover.

Data

This global dataset is divided into 10x10 degree tiles, consisting of seven files per tile. All files contain unsigned 8-bit values and have a spatial resolution of 1 arc-second per pixel, or approximately 30 meters per pixel at the equator.

Calculations

- Conversion of shapefiles into fusion tables for use in Google Earth Engine API
- Calculating area of forest cover and forest cover loss using Google Earth Engine
- Processing zonal statistics of raster by geometry in feature collection using the “Reduce Region” command
- Obtain % forest cover loss = Area of forest cover loss/ Total area of forest within feature (p1, p2, or g2)

Sources:

Center for International Forestry Research. 2 Dec, 2013. "Use Hansen high-res forest cover maps wisely, experts say". CIFOR Blog: http://blog.cifor.org/27435/use-hansen-high-res-forest-cover-maps-wisely-experts-say?fnl=en.

Global Forest Watch. 2016. World Resources Institute. http://www.globalforestwatch.org/map/3/15.00/27.00/ALL/grayscale/loss,forestgain?tab=analysis-tab&begin=2001-01-01&end=2015-01-01&threshold=30.

Hansen, M. C., P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, and J. R. G. Townshend. 2013. "High-Resolution Global Maps of 21st-Century Forest Cover Change." *Science* 342 (15 November): 850–53. Data available on-line from:http://earthenginepartners.appspot.com/science-2013-global-forest.

Holmgren, Peter. 7 April, 2015. "Watchers of forests – what news from above?" CIFOR blog: http://blog.cifor.org/27686/watchers-of-forests-what-news-from-above?fnl=en.

Holmgren, Peter. 5 Oct 2015. "Can we trust country-level data from global forest assessments?" CIFOR blog: http://blog.cifor.org/34669/can-we-trust-country-level-data-from-global-forest-assessments?fnl=en.

Weisse, Mikaela and Rachael Petersen. 17 Dec 2015. "How accurate is accurate enough? Examining the GLAD global tree cover change data", part 1 and 2. Global Forest Watch blog: http://blog.globalforestwatch.org/2015/12/how-accurate-is-accurate-enough-examining-the-glad-global-tree-cover-change-data-part-1/.

ii. Vegetation Index (NDVI)

Normalized Difference Vegetation Index (NDVI) is a measure of the "greenness," the relative density and health of vegetation, of the earth's surface. This index shows where and how much green leaf vegetation was growing during a certain period. The data were produced and collected by the Moderate resolution Imaging Spectroradiometer (MODIS) aboard NASA's terra satellite. Two different wavebands, red and near-infrared, are combined to enhance the vegetation signal from measures spectral responses. Reflected red energy decreases with plant development due to chlorophyll absorption within actively photosynthetic leaves, whereas reflected near-infrared energy will increase with plant development through scatter processes (reflection and transmission) in healthy, turgid leaves. The NDVI is a global measure of "greenness" since it combines these two bands into an equation. The equation combining the bands of visible red (VIS) and near-infrared (NIR) that produces the NDVI is given by:

$$NDVI = \frac{NIR - VIS}{NIR + VIS}$$

MODIS provides monthly NDVI maps on a global scale of different resolutions. The resolution image used as an input in our report is $10km^2$. The values range from -1 and +1. Values greater than .1 generally denote increasing degrees in the greenness and intensity of vegetation. Values between 0 and .1 are commonly characteristic of rocks and bare soil, and values less than 0 sometimes indicate clouds, rain, and snow. The dataset was originally saved as eight byte per pixel values, between 0 and 255. These raw values area easily converted to NDVI pixel values using the following formula:

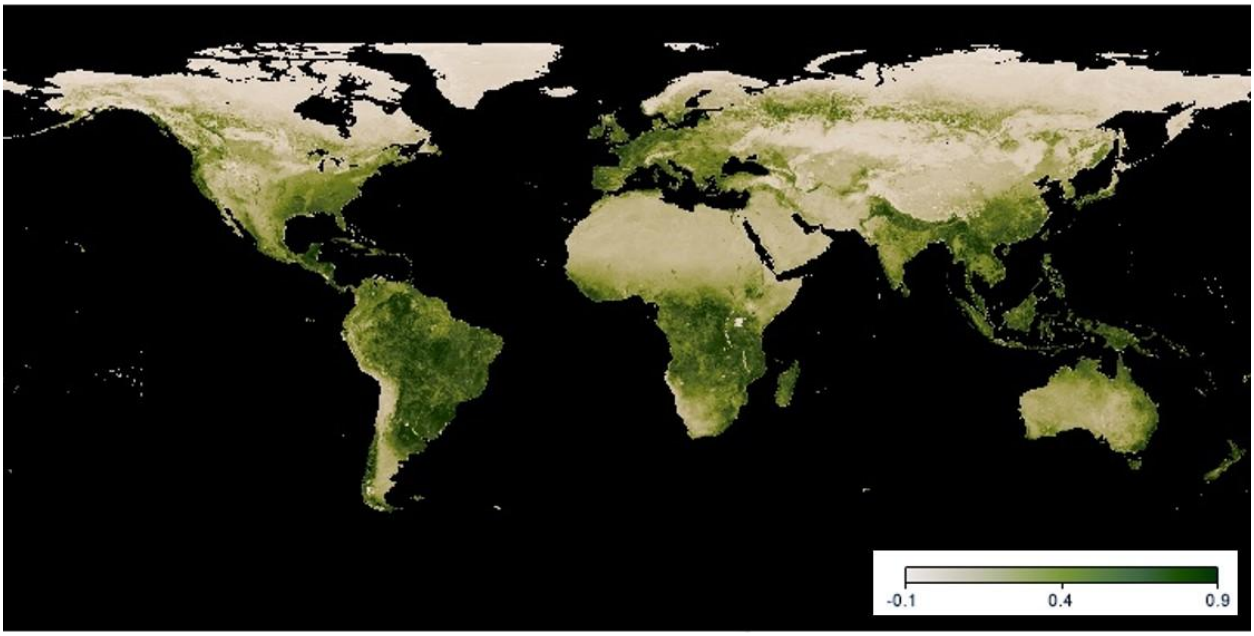
NDVI = (0.004*raw) - 0.1[2]

This implies that the converted values range between -0.096 (raw=1) and 0.92 (raw=255).

Calculation:

In our calculations, the NDVI for each pixel is averaged across all grid cells within each geographical boundary to obtain the mean NDVI of each area (admin 1 or admin 2). In addition to the mean NDVI of each area, the minimum, maximum, standard deviation, range and sum is calculated.

Figure 14 Global NDVI pixel from NEO (Nasa Earth Observation), February 2016



Sources:

NEO (Nasa Earth Observation), 2017. See here.

[2] Source: http://landval.gsfc.nasa.gov/pdf/SPOT_VGT_NDVI_subsets.pdf

iii. Net Primary Productivity

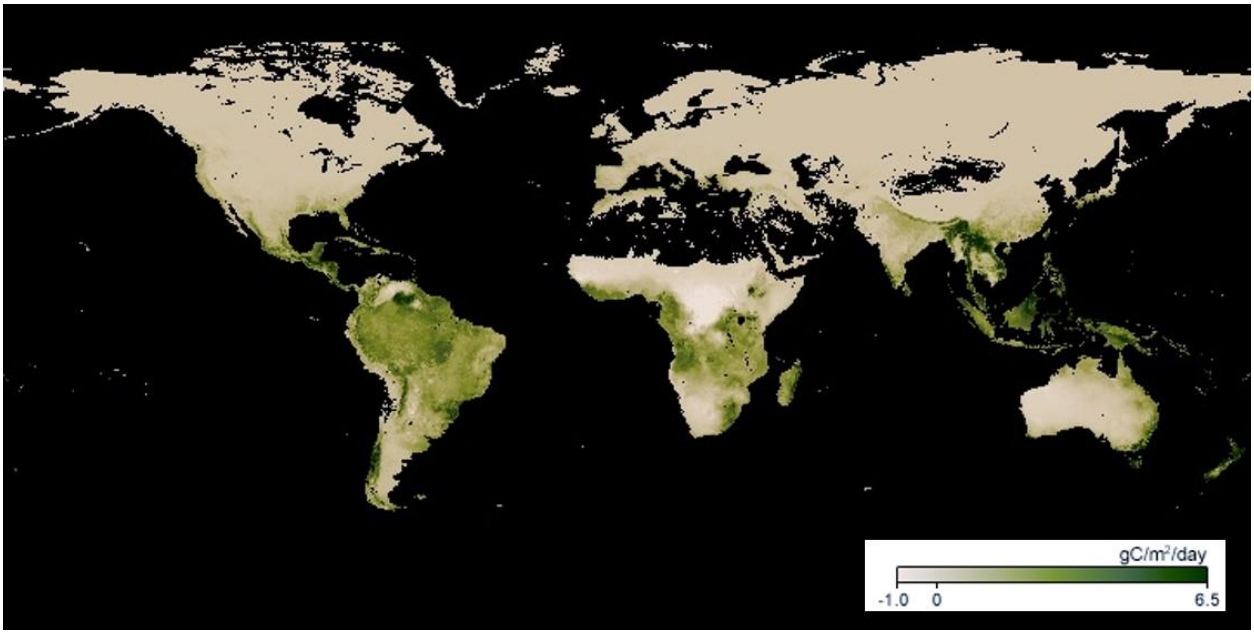
Net primary productivity (NPP) is the total amount of carbon dioxide taken in by plants. It is measured as the difference between how much carbon dioxide is taken in by plants compared to how much is put out by them. Carbon dioxide contribute to warming of the planet and it is therefore important to track plant productivity to understand where carbon dioxide comes from and where it goes. The unit of NPP is grams of carbon per m^2 per day.

The data is captured with Moderate Resolution Imaging Spectroradiometer (MODIS) aboard NASA's Terra and Aqua satellites. The resolution of the imagery used for the HDD is about 10km2/pixel.

Calculation:

In our calculations, the NPP for each pixel are averaged across all grid cells within each geographical boundary to obtain the mean NPP of each area (admin 1 or admin 2). In addition to the mean NPP of each area, the minimum, maximum, standard deviation, range and sum is calculated.

Figure 15 Global NPP pixel from NEO (Nasa Earth Observation), February 2010



Sources: NEO NASA Earth Observation

Data and explanations: NEO NASA Earth Observation:
http://neo.sci.gsfc.nasa.gov/view.php?datasetId=MOD17A2_M_PSN&date=2016-04-01

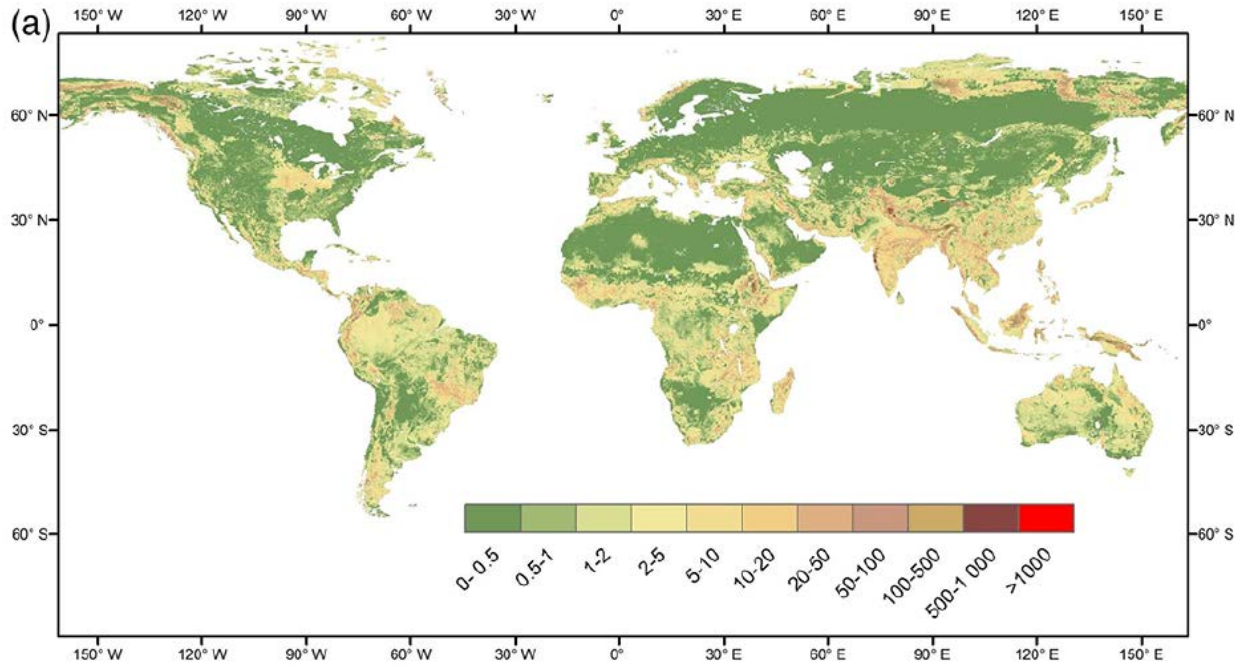
iv. Soil Degradation

Soil erosion can lead to land degradation in the form of nutrient loss, a decrease in the effective root depth, water imbalance in the root zone, productivity reduction (Yang et al., 2003). It is therefore a major threat to sustainable agriculture and food production (UNCCD, 2012; Walling, 2009). The model used to calculate soil erosion is known as the Revised Universal Soil Loss Equation (RUSLE) model (Renard et al., 1997). The RUSLE model predicts erosion rates and is a product of a rainfall erosivity factor, a slope steepness factor, a slope length factor, a soil erodibility factor, a land cover factor, and a support practice factor. Naipal et al. (2015) have derived yearly average soil erosion rates, in tons per hectar per year, by applying an adjusted RUSLE model more adapt for global analysis than the original RUSLE model. The soil erosion dataset calculated using the adjusted RUSLE model by Naipal et al (2015) is the data used for the purpose of our report.

Calculation:

In our calculations, the soil erosion for each pixel is averaged across all grid cells within each geographical boundary to obtain the mean soil erosion of each area (admin 1 or admin 2). In addition to the mean soil erosion of each area, the minimum, maximum, standard deviation, range and sum is calculated.

Figure 16 Global yearly averaged erosion rates according to the fully adjusted RUSLE model



Source: Naipal et al, 2015

Data:

Naipal, V., Reick, C., Pongratz, J. and Van Ost, K. (2015). Improving the global applicability of the RUSLE model –adjustment of the topographical and rainfall erosivity factors. *Geoscientific Model Development, 8*, 2893-2913.
Renard, K. G., Foster, G. R., Weesies, G. A., Mccool, D. K., and Yoder, D. C. (1997). Predicting Soil Erosion by Water: a Guide to Conservation Planning with the Revised Universal Soil Loss Equation (RUSLE), Agriculture Handbook 703, USDA, Washington, DC, USA.
United Nations Convention to Combat Desertification (UNCCD): Zero Net Land Degradation, Bonn, Germany, ISBN 978-92-95043-62-6, 2012.
Walling, D. E. (2009). The Impact of Global Change on Erosion and Sediment Transport by Rivers: Current Progress and Future Challenges, The United Nations World Water Assessment Programme. Scientific Paper, UNESCO, Paris, France.
Yang, D., Kanae, S., Oki, T., Koike, T., and Musiake, K. (2003). Global potential soil erosion with reference to land use and climate changes. *Hydrological Processes, 17*, 2913–2928.

v. Biodiversity Hotspots

Measures and their definition:

Variable name	Description
BIODIV_area	**Biodiversity hotspot area (km^2): Size of area that is a biodiversity hotspot**
BIODIV_pct	**Biodiversity hotspot area (%): Share of area that is a biodiversity hotspot**

A biodiversity hotspot is a place on earth that is both biologically rich but at the same time threatened by habitat loss. There are 35 regions in the world that qualify as hotspots. These regions represent 2.3% of the land on Earth's surface, but they support more than half of the world's plant species as endemics and nearly 43% of bird, mammal, reptile and amphibian species as endemics. According to Myers et al. (2000) the following criteria must be met in order for an area to be classified as a hotspot:

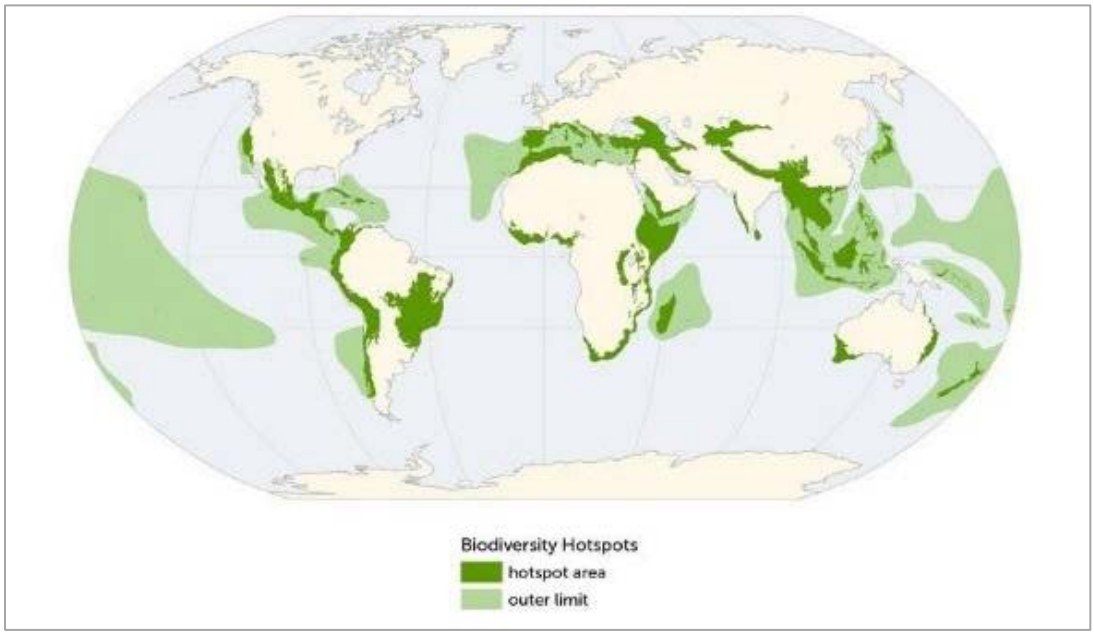
- It must have at least 1,500 vascular plants as endemics (> 0.5 percent of the world's total). A hotspot, in other words, is irreplaceable.
- It must have 30% or less of its original natural vegetation. In other words, it must be threatened.

Data & Calculation:

The data is provided as a polygon shapefile by Conservation International (2016) covering the entire globe.

- Area of 'Hotspot' in km^2 in feature
- % Area 'Hotspot' = Area of 'Hotspot' / Total Area of feature (p1, p2, g1, or g2)

Figure 17 Biodiversity Hotspots from Conservation International (2016)



Sources:

Myers, Norman, Russell A. Mittermeier2, Cristina G. Mittermeier2, Gustavo A. B. da Fonseca & Jennifer Kent (2000). Biodiversity hotspots for conservation priorities. *Nature 403*, 853-858

Conservation International (2016). http://www.conservation.org/how/pages/hotspots.aspx

Data: Biodiversity Hotspots Revisited, Conservation International, 2011. https://databasin.org/datasets/23fb5da1586141109fa6f8d45de0a260

vi. Bird Areas

Measures and their definition:

Variable name	Description
EBA_area	**Endemic Bird Area (km^2): Size of area that is an endemic bird area**
EBA_pct	**Endemic Bird Area (%): Share of area that is an endemic bird area**

Explanation:

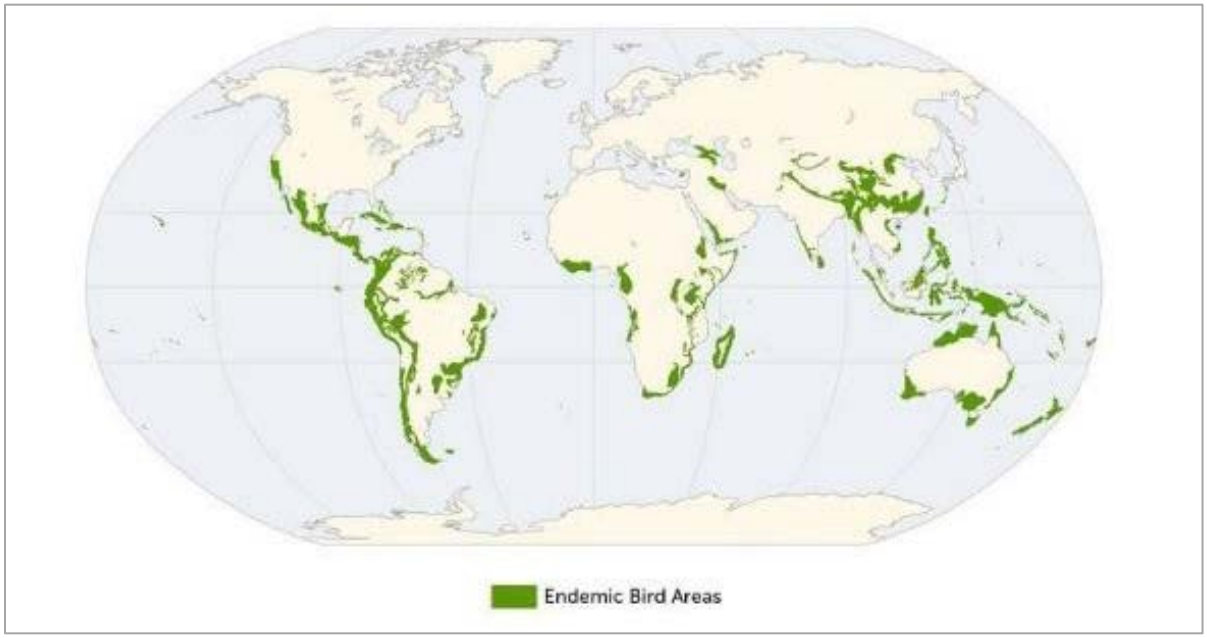
Endemic Bird Areas (EBA) are defined as critical regions of the world for the conservation of restricted-range (species with world distributions of less than 50,000km^2) bird species where the distributions of two or more restricted-range bird species overlap" (BirdLife International). Due to the small size of these bird species ranges, half of the restricted-range bird species are vulnerable to the loss or degradation of habitat while the other half is globally threatened or near

threatened. Most of the EBAs are forest covered islands or mountain ranges in the tropics or subtropics of varying sizes, from a few square kilometers to more than 100,000 km2, and they support 2 to 80 different species (BirdLife International).

Data & Calculation:
The data is provided as a polygon shapefile by the Integrated Biodiversity Assessment Tool (IBAT) covering the entire globe.

- Area of 'EBA' in km^2 in feature
- % Area of 'EBA' = Area of 'EBA' / Total Area of feature (p1, p2, g1, or g2)

Figure 18 Endemic Bird Areas from BirdLife International



Sources:
BirdLife International. http://www.birdlife.org/datazone/eba. More Info from: http://biodiversitya-z.org/content/endemic-bird-areas-eba.
Data: Integrated Biodiversity Assessment Tool (IBAT): https://www.ibat-alliance.org/ibat-ifc/login.

vii. Protected Area
Measures and their definition:

Variable name	Description
PA_area	Protected Area (km^2): Size of area that is a protected area

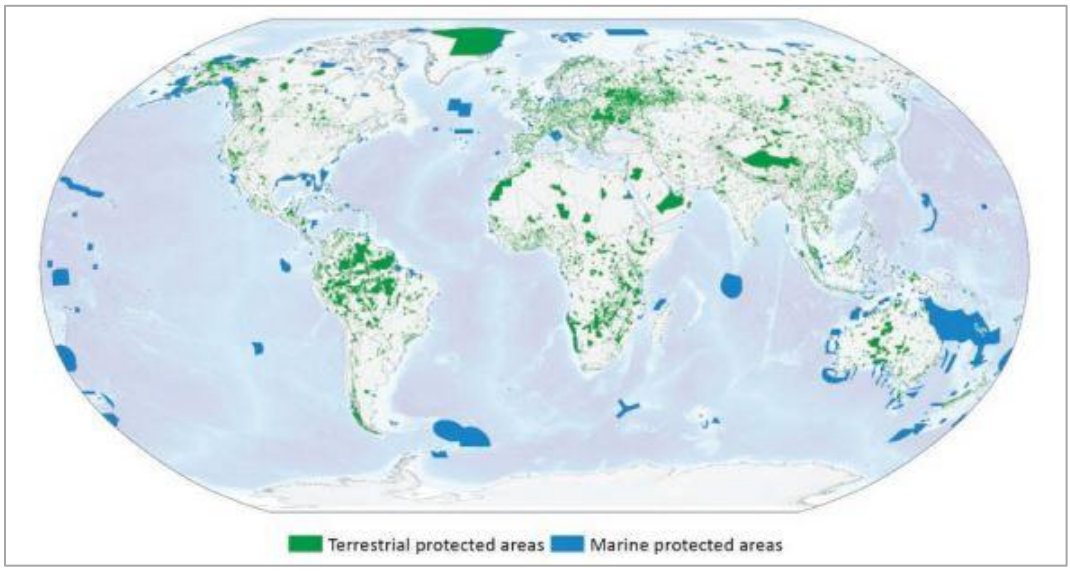
PA_pct	Protected Area (%): Share of area that is a protected area

Explanation:

The IUCN (International Union for Conservation of Nature) defines a protected area as "a clearly defined geographical space, recognized, dedicated and managed, through legal or other effective means, to achieve the long-term conservation of nature with associated ecosystem services and cultural values" (Juffe-Bignoli et al., 2014). The protected areas reflect six different management objectives: (i) strict nature reserve, (ii) wilderness area, (iii) national park, (iii) natural monument, (iv) habitat species management, (v) protected landscape/seascape and (vi) management resource protected area. Protected area coverage has been used as one of the indicators to track progress towards the Millennium Development Goals. The objectives of managing protected areas is to prevent biodiversity loss as well as maintaining food security and water supplies, strengthening climate resilience, and improving human health and well-being.

The Protected Planet Report from 2014 states that 15.4% of the world's terrestrial and inland water areas are covered by protected areas in 2014. In terms of marine and coastal areas, protected areas cover: 3.4% of the global ocean area, 8.4% of all marine areas within national jurisdiction, and 10.9% of all coastal waters are covered by protected areas. Only 0.25% of marine areas beyond national jurisdiction are within protected areas.

Figure 19 Spatial distribution of the world's protected areas.



Source: UNEP-WCMC 2014

Data & Calculation:

The data is provided as a polygon shapefile by Protected Planet covering the entire globe.

- Area of 'Protected Area' in km^2 in feature
- % Area of 'Protected Area' = Area of 'Protected Area' / Total Area of feature (p1, p2, g1, or g2)

Sources:

WDPA (World Database on Protected Areas): http://www.protectedplanet.net/about

Juffe-Bignoli, D., Burgess, N.D., Bingham, H., Belle, E.M.S., de Lima, M.G., Deguignet, M., Bertzky, B., Milam, A.N., Martinez-Lopez, J., Lewis, E., Eassom, A., Wicander, S., Geldmann, J., van Soesbergen, A., Arnell, A.P., O'Connor, B., Park, S., Shi, Y.N., Danks, F.S., MacSharry, B., Kingston, N. (2014). Protected Planet Report 2014. UNEP-WCMC: Cambridge, UK.

UNEP-WCMC. (2014). Global statistics from the World Database on Protected Areas (WDPA), August 2014. Cambridge, UK: UNEP- WCMC.

5.B. Blue indicators

i. Degradation of Marine Ecosystem

Description (adapted from Halpern et al.; see here):

Cumulative Human Impact Score

The Ocean Health Index measures the global state of the world's oceans. A Cumulative (Human) Impact Score measures the cumulative impacts to marine ecosystems globally from fishing, climate change, and ocean- and land-based measures. 19 different types of anthropogenic stress on 20 global marine ecosystem types using global scale data is used in the calculations of the index (Table 10). The stressors have been multiplied by habitat vulnerability scores to produce the Cumulative Impact Score. The data for 2013 used in our report is represented at $1km^2$ resolution converted to the Mollweide projection with a WGS84 datum. For detailed information on the individual indicators included in Table 13 below, please see here.

Figure 20 Subcomponents of the Ocean Health Index

Stressor		native resolution	Year(s)
Land-based	nutrient pollution	modeled 1km2	2007-2010
	organic pollution	modeled 1km2	2007-2010
	inorganic pollution	modeled 1km2	2000-2001
	direct human	modeled 1km2	2011

	light pollution	1km2	2007-2010
Fishing	demersal, destructive	half-degree	2011
	demersal, non-destructive, high bycatch	half-degree	2011
	demersal, non-destructive, low bycatch	half-degree	2011
	pelagic, high bycatch	half-degree	2011
	pelagic, low bycatch	half-degree	2011
	artisanal	modeled 1km2	2006
Climate change	SST anomalies	~21km2	1985-1990 vs. 2005-2010
	UV radiation	1 degree	2008-2012
	ocean acidification	1 degree	1870 vs. 2000-2009
	sea level rise	modeled 0.25 degree	1992-2012
Ocean-based	commercial shipping	0.1 degree	2003-2011
	invasive species	modeled 1km2	2011
	ocean-based pollution	modeled 1km2	
	benthic structures (oil rigs)	1km2	2007-2010

For more info about the stressors see Halpern et al (2015) –supplementary material.

Calculation:

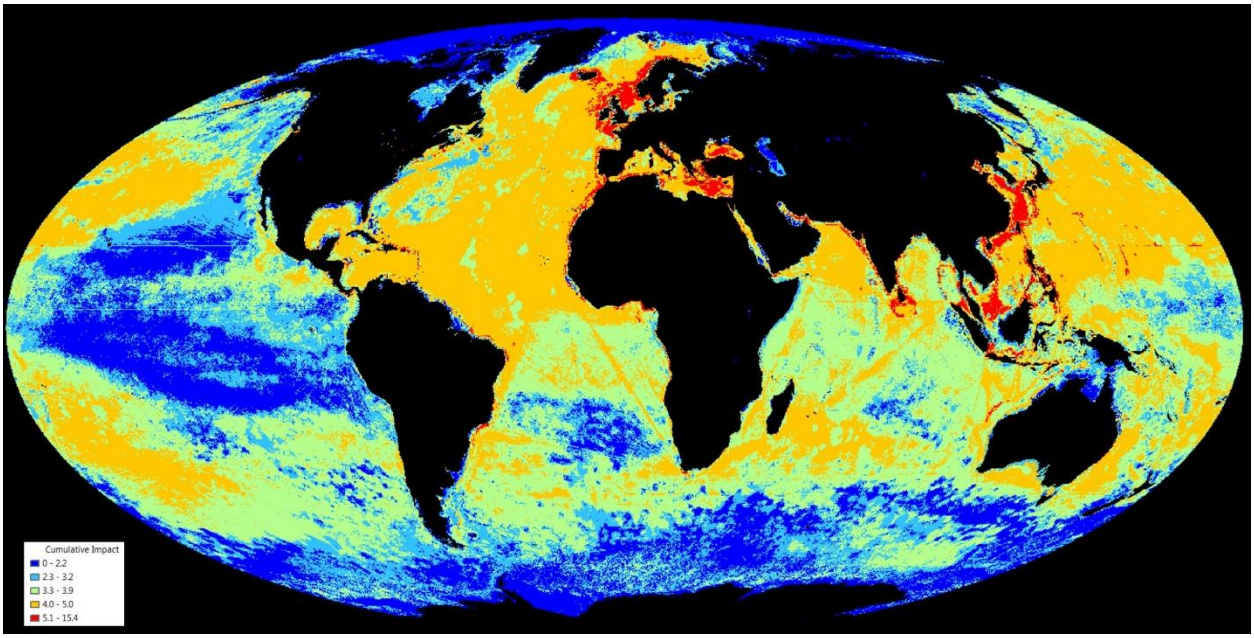
Derived by Halpern et al. (2015), the cumulative impact is the per-pixel average of the habitat vulnerability-weighted stressor intensities where weights ($\mu_{i,j}$) are determined by the vulnerability of each *i… m* habitat (*E*) to each *j… n* stressor (*D*), such that:

$$I_C = \sum_{j=1}^{n} \frac{1}{m} \sum_{i=1}^{m} D_i * E_j * \mu_{i,j}$$

The ecosystem vulnerability weights ($\mu_{i,j}$) were developed previously in Halpern et al. (2007) and can be found in the supplementary material to Halpern et al. (2015). The cumulative impact data for 2013 is provided by Halpern et al (2015) in raster format with global coverage. The following processing steps were taken to produce the final output for the HDD.

- Calculate the value of Cumulative Impact Index extending 5km from each coastal district (through buffering in a GIS software)
- Assigning calculated value to associated district

Figure 21 Cumulative Human Impact Index on marine ecosystems in 2013



Source: Halpern (2015)

Data available from: https://knb.ecoinformatics.org/#view/doi:10.5063/F15718ZN.

Halpern, B. S., Selkoe, K. A., Micheli, F. & Kappel, C. V. (2007). Evaluating and ranking the vulnerability of global marine ecosystems to anthropogenic threats. *Conservation Biology*, 21, 1301–1315.

Halpern B.S., Frazier M, Potapenko J, Casey KS, Koenig K, Longo C, Lowndes JS, Rockwood RC, Selig ER, Selkoe KA, Walbridge S. (2015). Spatial and temporal changes in cumulative human impacts on the world's ocean, Nature Communications, *6*, 2041-1723.

Halpern, B. S. et al. (2008). A global map of human impact on marine ecosystems. Science 319, 948-952.

ii. Reefs at Risk Revisited

Measures and their definitions:

Variable name	Description
Shoreline at low, medium, high	Area of district's shoreline that has been classified as low, medium, high and very high threat (km2)

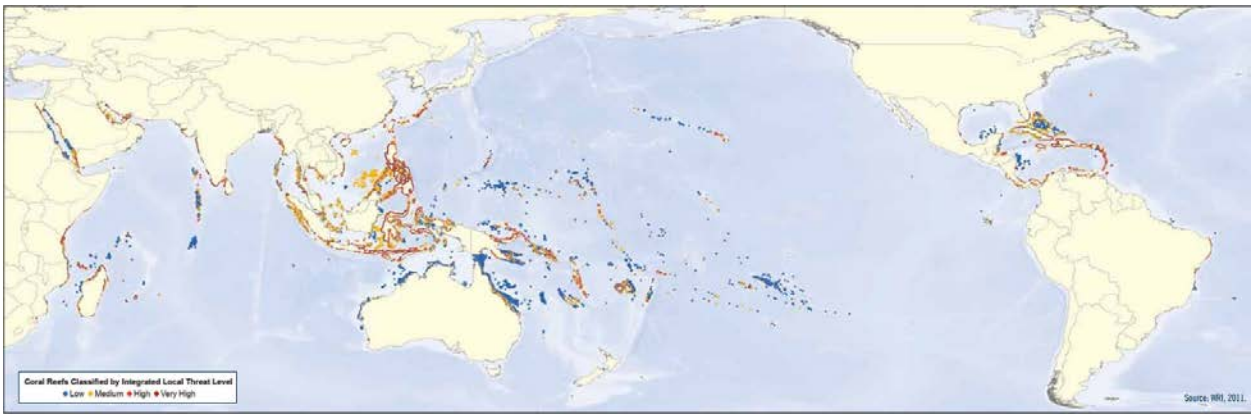
and very high risk, Area	
Share of district's shoreline that has been classified as very high threat level	Share of district's shoreline that has been classified as very high threat level (i.e. very high / (low + medium + high + very high)
Share of district's shoreline that has been classified as very high & high threat level	Share of district's shoreline that has been classified as very high & high threat level (i.e. very high + high / (low + medium + high + very high)

Explanation:
Reefs at Risk Revisited brings together data on the world's coral reefs in a global analysis designed to quantify threats and to map where reefs are at greatest risk of degradation or loss. It incorporated more than 50 data sources into the analysis—including data on bathymetry (ocean depth), land cover, population distribution and growth rate, observations of coral bleaching, and location of human infrastructure. These data were consolidated within a geographic information system (GIS), and then used to model several broad categories of threat from human activities, climate change, and ocean acidification.

Human pressures on coral reefs are categorized throughout the report as either "local" or "global" in origin. Local threats involve human activities near reefs that have a direct and relatively localized impact, while global threats affect the reef environment indirectly through the cumulative impact of human activities on the global climate and ocean chemistry. Local threats addressed in this analysis are: Coastal development, Watershed-based pollution, Marine-based pollution and damage, and Overfishing and destructive fishing. Global-level threats addressed are: thermal stress (warming sea temperatures, which can induce coral bleaching) and ocean acidification (driven by increased CO2, which can reduce coral growth rates).

Reefs at Risk Revisited is a high-resolution update of the original global analysis, Reefs at Risk: A Map-Based Indicator of Threats to the World's Coral Reefs. Reefs at Risk Revisited uses a global map of coral reefs at 500-m resolution. The dataset is downloaded as a raster.

Figure 22: Coral reefs of the world classified by threat from local activities



Source: WRI (2011).

Calculation:

- Convert all shapefiles and rasters to Lambert Azimutal Equal Area
- Convert reef threats grid into 4 shapefiles (low, medium, high and very high)
- Zonal stats using reef threats grid and "zone of influence" which is an Euclidean buffer for admin areas (p1, p2, g1, g2)
- Calculate threat level areas (low, medium, high, very high)
- Calculate ratios (very high/sum of all threat levels and very high + high/sum of all threat level) in R.

Source:

Burke, L., K. Reytar, M. Spalding and A. Perry. 2011. Reefs at Risk Revisited. World Resources Institute. http://www.wri.org/publication/reefs-risk-revisited.

iii. Water Risk Indicators

Description:

We draw from the Aqueduct dataset (2014), developed by the World Resources Institute (WRI), to integrate four key indicators of water-related risks: threatened amphibians, flood occurrence, groundwater stress, and drought severity.

Threatened amphibians

"Threatened amphibians measures the percentage of freshwater amphibian species classified by IUCN as threatened. Higher values indicate more fragile freshwater ecosystems and may be more likely to be subject to water withdrawal and discharge regulations." (Gassert, et al. 2014)

The data is from the International Union for Conservation of Nature (IUCN), and is in polygon format - each polygon representing a catchment area. The year covered in the analysis is 2010.

Flood occurrence
"Flood occurrence is the number of floods recorded from 1985 to 2011." (Gassert, et al. 2014)
The data is from the Dartmouth Flood Observatory and is in flood extent polygons.

Groundwater stress
"*Groundwater stress measures the ratio of groundwater withdrawal relative to its recharge rate over a given aquifer. Values above one indicate where unsustainable groundwater consumption could affect groundwater availability and groundwater-dependent ecosystems."* (Gassert, et al. 2014)

The data was developed by Gleeson et al. (2012) and is in polygon format – each polygon representing an aquifer.

Drought severity
"Drought severity measures the average length of droughts times the dryness of the droughts from 1901 to 2008." (Gassert, et al. 2014)

The data is from Sheffield and Wood (2007), and is provided in the same shapefile as the other water risk indicators, as compiled by World Resources Institute's Aqueduct portal.

Calculation
For the first three indicators (threatened amphibians, flood occurrence and groundwater stress), the data was converted to raster format, by assigning the corresponding polygon value to each pixel. We summarized the resulting raster datasets, as well as the drought severity grid, to our areas of interest (admin 1 and admin 2) extracting the mean value of each indicator.

Data: http://www.wri.org/our-work/project/aqueduct/

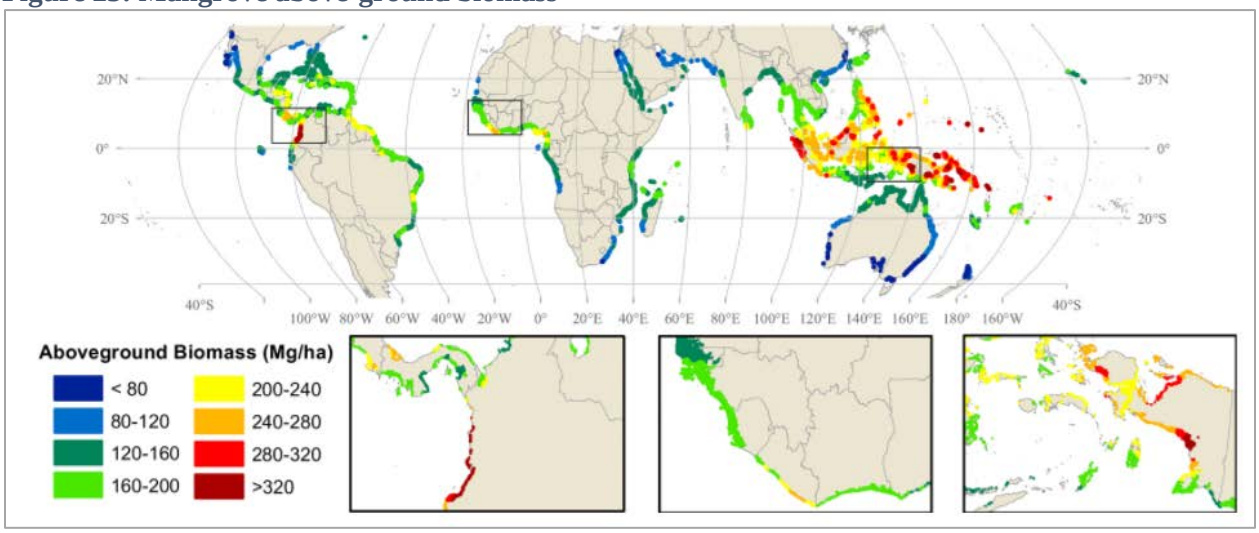
Source:
Gassert, F., M. Luck, M. Landis, P. Reig, and T. Shiao. 2014. "Aqueduct Global Maps 2.1: Constructing Decision-Relevant Global Water Risk Indicators." Working Paper. Washington, DC: World Resources Institute. Available online at http://www.wri.org/publication/aqueduct-globalmaps-21-indicators.

iv. Mangroves

Description:
Understanding spatial variation in carbon storage in natural habitats is critical for climate change mitigation efforts such as REDD. Terrestrial forests are being mapped with increasing accuracy, but the distribution of ``blue carbon'' in marine ecosystems remains poorly understood. We reviewed the literature to obtain field data on carbon storage and fluxes in mangroves world-wide. Using this material we developed a climate-based model for potential mangrove above-ground biomass (AGB) with almost four times the explanatory power of the only previous published model. From this model, we present the first ever global map of potential mangrove AGB and estimate a total global mangrove AGB of 2.83 Pg, with an average of 184.8 t ha−1. Data on other carbon stocks and fluxes confirm the importance of mangroves in carbon accounting. The map highlights the high variability in mangrove AGB and indicates areas that should be prioritized for mangrove conservation and restoration. (Hutchinson et al, 2013).

The mean aboveground mangrove biomass values were calculated for each coastal province and district where mangrove data was available.

Figure 23: Mangrove above-ground biomass



Source:
Hutchison, J., Manica, A., Swetnam, R., Balmford, A. and Spalding, M. (2014), Predicting Global Patterns in Mangrove Forest Biomass. Conservation Letters, 7: 233–240. doi:10.1111/conl.12060.

5.C. Brown indicators

i. PM2.5 Air Pollution - Concentrations

Measures and their definitions:

Variable name	Description
PM2.5 air pollution, satellite, 2010	PM2.5 air pollution, measured by satellite for 2010 (microgram/m3)
PM2.5 air pollution, satellite, all years	PM2.5 air pollution, measured by satellite for the mean of all years 2000-2013 (microgram/m3)
PM2.5 air pollution, 2013 (Brauer et al.)	PM2.5 air pollution, measured by satellite, models and monitors for 2013 (microgram/m3)
PM2.5 air pollution, all composition, 2013 (v. Donkelaar et al.)	PM2.5 air pollution, all composition, measured by satellite, models and monitors for 2013 (micrograms/m3)
PM2.5 air pollution, all composition, all years (v. Donkelaar et al.)	PM2.5 air pollution, all composition, measured by satellite, models and monitors for the year corresponding with poverty data (micrograms/m3)
PM2.5 air pollution, dust- and salt-removed, 2013 (v. Donkelaar et al.)	PM2.5 air pollution, dust- and salt-removed, measured by satellite, models and monitors, 2013 (micrograms/m3)
PM2.5 air pollution, dust- and salt-removed, all years (v. Donkelaar et al.)	PM2.5 air pollution, dust- and salt-removed, measured by satellite, models and monitors, for the year corresponding with poverty data (micrograms/m3)

Explanation:

This dataset estimates global fine particulate matter (PM2.5) concentrations using information from a combination of satellite-, simulation- and monitor-based sources. A Geographically Weighted Regression (GWR) is applied to global geophysically based satellite-derived PM2.5 estimates. Aerosol optical depth from multiple satellite products (MISR, MODIS Dark Target, MODIS and SeaWiFS Deep Blue, and MODIS MAIAC) was combined with simulation (GEOS-Chem) based upon their relative uncertainties as determined using ground-based sun photometer (AERONET) observations for 1998–2014.

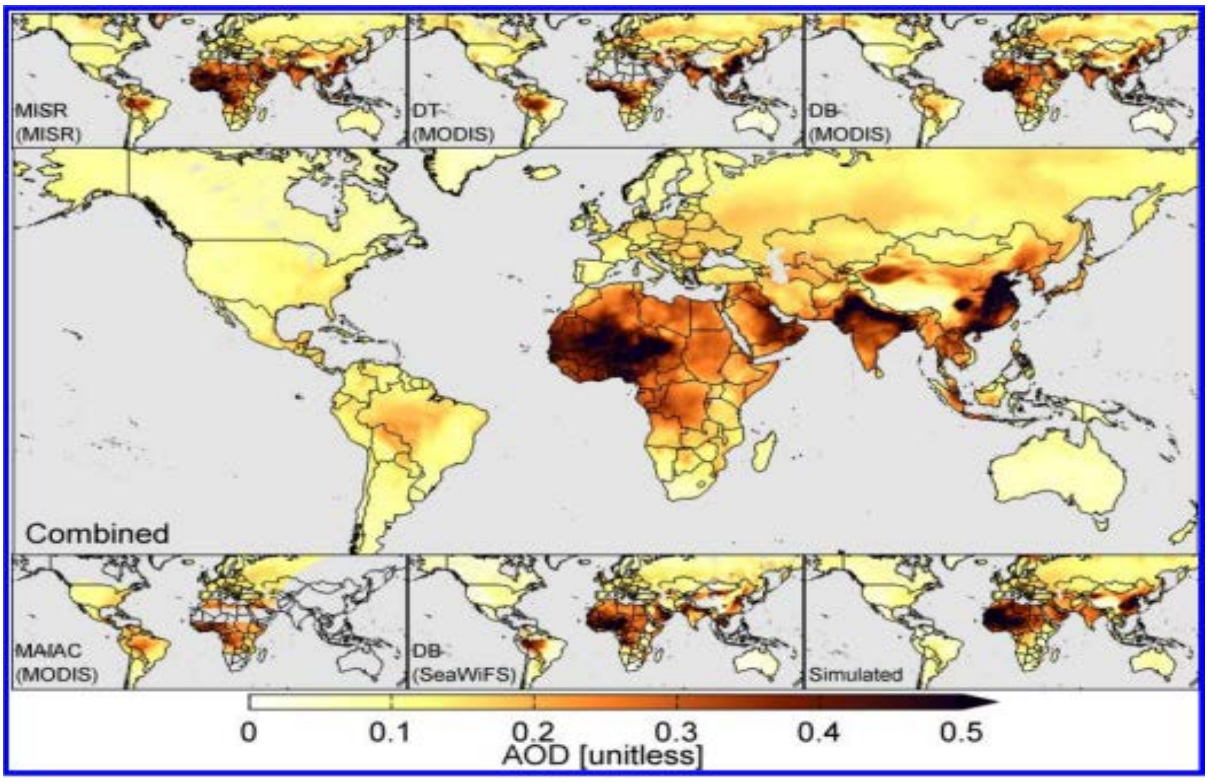
The GWR predictors included simulated aerosol composition and land use information. The resultant PM2.5 estimates were highly consistent (R2 = 0.81) with out-of-sample cross-validated PM2.5 concentrations from monitors.

The global population-weighted annual average PM2.5 concentrations were 3-fold higher than the 10 μg/m3 WHO guideline, driven by exposures in Asian and African regions. Estimates in regions with high contributions from mineral dust were associated with higher uncertainty, resulting from both sparse ground-based monitoring, and challenging conditions for retrieval and simulation. This approach demonstrates that the addition of even sparse ground-based measurements to more globally continuous PM2.5 data sources can yield valuable improvements to PM2.5 characterization on a global scale.

This data is produced for two measurements: all PM2.5 compositions as well as dust and sea-salt removed. The dust and sea-salt removed PM2.5 estimates apply simulated compositional information to the full-composition values, following van Donkelaar et al., EHP, 2015.

The data is downloaded as .ASCII raster files. Gridded files use the WGS84 projection. The 0.01° × 0.01° grid contains 12500 latitude coordinates, with centres from 54.995°S to 69.995°N, and 36000 longitude coordinates, with centres from 179.995°W to 179.995°E.

Figure 24: Mean aerosol optical depth (AOD) over land 2000-2010, by data source.



Note: The middle panel shows the combination of all data sources after calibrating with AERONET. Gray denotes missing data or water.
Source: van Donkelaar et al (2016).

Calculations:

- Conversion of .ASCII raster files into .TIFF images.
- Zonal statistics of PM2.5 air pollution as value raster and P1, P2, G1 and G2 admin levels as feature zones using ArcPy.
- Implementation of calculations for all PM2.5 compositions and dust and sea-salt removed measurements.
- Implementation of calculations for the year 2013 as well as for the year corresponding with the feature's poverty data.

Sources:

Atmospheric Composition Analysis Group. 2016. http://fizz.phys.dal.ca/~atmos/martin/?page_id=140. Dalhousie University.

Brauer, M., Freedman, G., Frostad, J., Van Donkelaar, A., Martin, R.V., Dentener, F., Dingenen, R.V., Estep, K., Amini, H., Apte, J.S. and Balakrishnan, K., 2015. "Ambient air pollution exposure estimation for the global burden of disease 2013". *Environmental science & technology*, 50(1), pp.79-88.

van Donkelaar, Aaron, Randall V. Martin, Michael Brauer, N. Christina Hsu, Ralph A. Kahn, Robert C. Levy, Alexei Lyapustin, Andrew M. Sayer, and David M. Winker. 2016. "Global Estimates of Fine Particulate Matter using a Combined Geophysical-Statistical Method with Information from Satellites, Models, and Monitors". *Environmental Science & Technology*, *50* (7), 3762-3772. DOI: 10.1021/acs.est.5b05833.

van Donkelaar, Aaron, Randall V. Martin, Michael Brauer, and Brian L. Boys. 2015. "Use of satellite observations for long-term exposure assessment of global concentrations of fine particulate matter." PhD diss., University of British Columbia.

van Donkelaar, A., R.V. Martin, M. Brauer, and B.L. Boys. 2015. "Global Annual PM2.5 Grids from MODIS, MISR and SeaWiFS Aerosol Optical Depth (AOD), 1998-2012". Palisades, NY: *NASA Socioeconomic Data and Applications Center (SEDAC).* http://dx.doi.org/10.7927/H4028PFS.

ii. PM2.5 Air Pollution - Emissions

Measure and its definition:

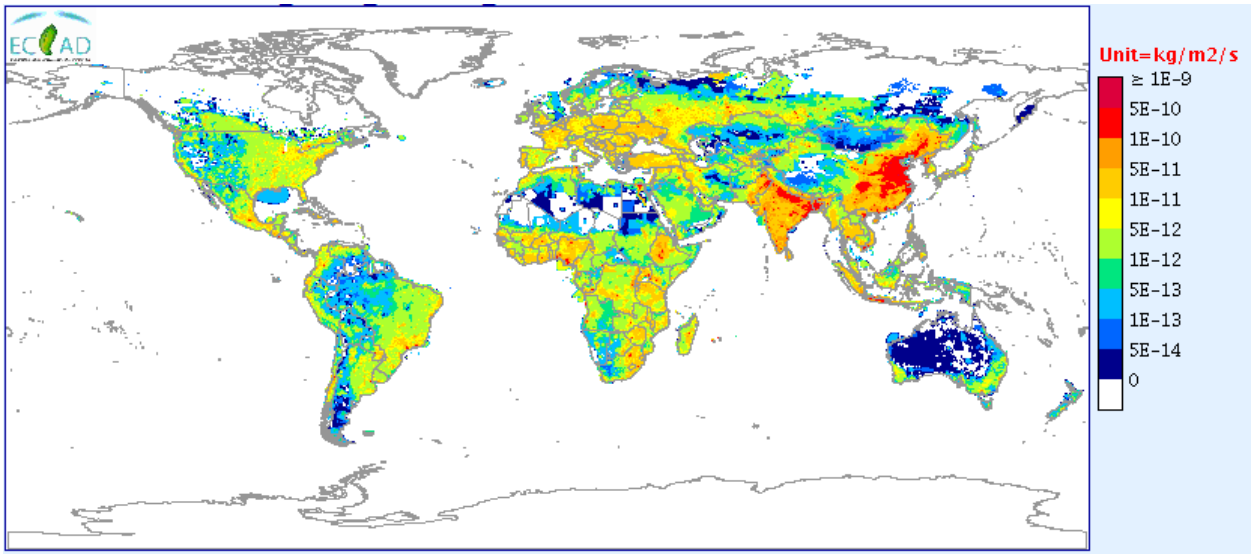
Variable name	**Description**
PM2.5 air pollution, emissions	PM2.5 emissions (ECCAD) 2010 in nanogram/m2/s

Explanation:
This dataset provides gridded global fine particulate matter less than or equal to 2.5 micrometers in diameter (PM2.5 air pollution) at 0.5 degrees spatial resolution. The data is provided as a .tiff raster.

The PM2.5 measure is a part of the ECLIPSE anthropogenic emissions inventory for year 2005, 2010, 2030 and 2050, which includes global emissions from all anthropogenic sources excluding international shipping and aviation. The emission datasets do not include emissions from international shipping and aviation, biogenic VOC emissions, and forest and savannah fires (emissions from open burning of agricultural residue are included).

This global emission data set has been developed with the GAINS model [Amann et al., 2011] as part of the activities within European Commission 7 Framework funded projects ECLIPSE, PEGASOS, and activities of the UNECE Task Force on Hemispheric Transport of Air Pollutants (HTAP – http://www.htap.org). Two scenarios were developed: a baseline senario 2005-2050 (current legislation- CLE), and Maximum technically feasible reductions scenario for 2030 and 2050 (MTFR).

Figure 25: PM2.5 emission measures for 2010



Source: ECCAD (2016).

Calculations:

- Conversion of raster unit from kg/m2/s to nanogram/m2/s due to difficulty in processing extremely small values (as small as 10e-15).

- Zonal statistics of PM2.5 air pollution as value raster and P1, P2, G1 and G2 admin levels as feature zones.

Source:

Z Klimont, S. J. Smith and J Cofalar, The last decade of global anthropogenic sulfur dioxide: 2000–2011 emissions, Environmental Research Letters, 8, 014003, 2013.
Höglund-Isaksson, L.: Global anthropogenic methane emissions 2005–2030: technical mitigation potentials and costs, Atmos. Chem. Phys., 12, 9079-9096, doi:10.5194/acp-12-9079-2012, 2012.

iii. PM10 Air Pollution

Measure and its definition:

Variable name	**Description**
PM10 air pollution	PM10 (ECCAD) 2010 in nanogram/m2/s

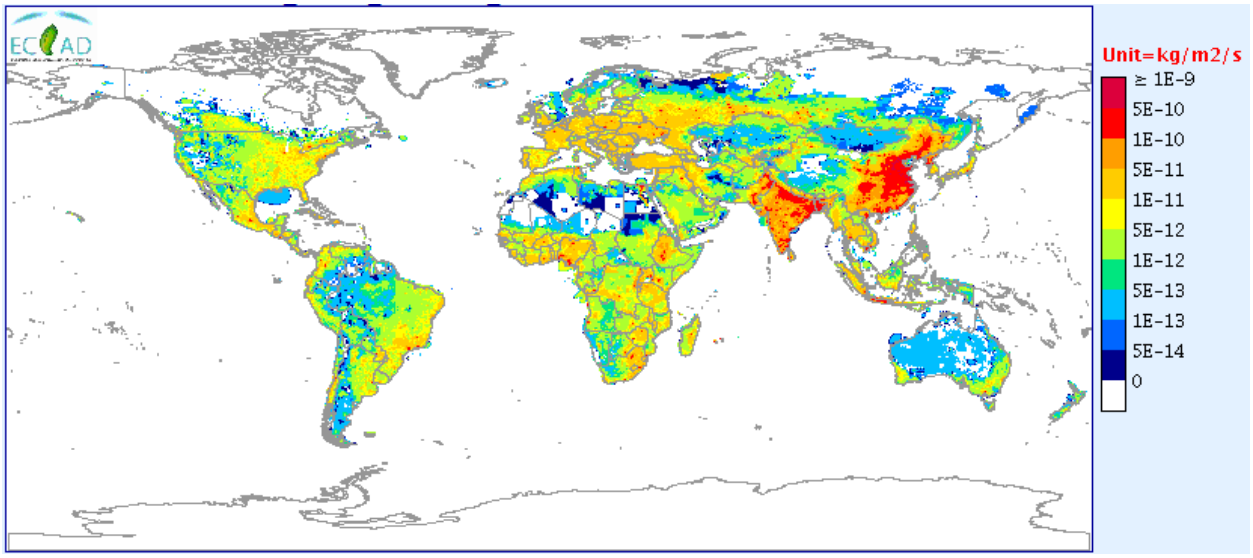
Explanation:

This dataset provides gridded global particulate matter less than or equal to 10 micrometers in diameter (PM10 air pollution) at 0.5 degrees spatial resolution. The data is provided as a raster .tiff.

The PM10 measure is a part of the ECLIPSE anthropogenic emissions inventory for year 2005, 2010, 2030 and 2050, which includes global emissions from all anthropogenic sources excluding international shipping and aviation. The developed emission datasets do not include emissions from international shipping and aviation, biogenic VOC emissions, and forest and savannah fires (emissions from open burning of agricultural residue are included).

This global emission data set has been developed with the GAINS model [Amann et al., 2011] as part of the activities within European Commission 7 Framework funded projects ECLIPSE, PEGASOS, and activities of the UNECE Task Force on Hemispheric Transport of Air Pollutants (HTAP – http://www.htap.org). Two scenarios have been developed: a baseline senario 2005-2050 (current legislation- CLE), and Maximum technically feasible reductions scenario for 2030 and 2050 (MTFR).

Figure 26: Global PM10 air pollution measures for 2010 under a baseline scenario.



Source: ECCAD (2013).

Calculations:

- Conversion of raster unit from kg/m2/s to nanogram/m2/s due to difficulty in processing extremely small values (as small as 10e-15).
- Zonal statistics of PM10 air pollution as value raster and P1, P2, G1 and G2 admin levels as feature zones.

Source:

Z Klimont, S. J. Smith and J Cofalar, The last decade of global anthropogenic sulfur dioxide: 2000–2011 emissions, Environmental Research Letters, 8, 014003, 2013.

Höglund-Isaksson, L.: Global anthropogenic methane emissions 2005–2030: technical mitigation potentials and costs, Atmos. Chem. Phys., 12, 9079-9096, doi:10.5194/acp-12-9079-2012, 2012.

iv. Black Carbon

Measure and its definition:

Variable name	**Description**
Black carbon	Black Carbon (ECCAD) 2010 in nanogram/m2/s

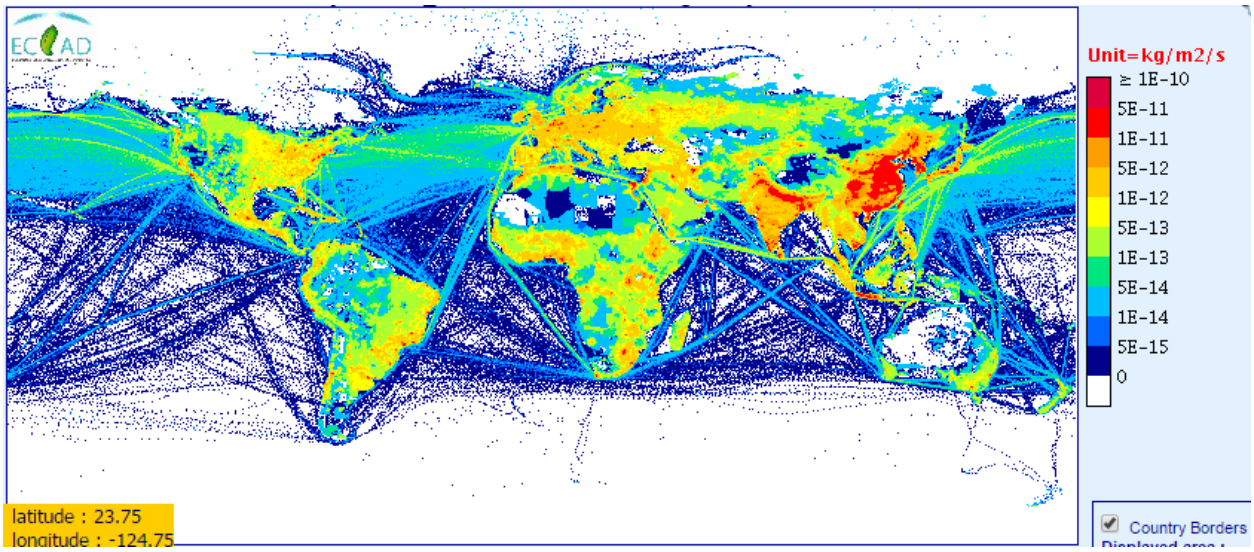
Explanation:

This dataset measures black carbon (BC) at a 0.5 degrees spatial resolution and with a unit of

nanogram/m2/s. Black carbon is the most strongly light-absorbing component of particulate matter (PM) and is formed by the incomplete combustion of fossil fuels, biofuels, and biomass. The data is provided by the MACCity (MACC/CityZEN EU projects) emissions dataset, which is derived from the ACCMIP and RCP8.5 datasets, and distributed by ECCAD. The data is downloaded as .tiff rasters.

The emissions dataset includes data on a yearly basis for the period 1960-2010 for the anthropogenic emissions, and 1960-2008 for the biomass burning emissions. The dataset includes biomass burning emissions that comprise of all emissions resulting from natural and man- made surface vegetation fires, including emissions from the combustion of soil organic matter (duff or peat). Emissions emanating from the combustion of agricultural wastes (e.g. burning rice stubble) or biofuels (e.g. cow dung, charcoal) are excluded.

Figure 27: Average black carbon measurements for 2010.



Source: ECCAD (2016).

Calculation:

- Conversion of raster unit from kg/m2/s to nanogram/m2/s due to difficulty in processing extremely small values (as small as 10e-15).
- Zonal statistics of BC as value raster and P1, P2, G1 and G2 admin levels as feature zones.

Sources:

Claire Granier et al., 2011, Evolution of anthropogenic and biomass burning emissions of air pollutants at global and regional scales during the 1980–2010 period. , DOI: 10.1007/s10584-011-0154-1 , Climate Change 109 (1-2) : 163-190.

Diehl et al., 2012, Anthropogenic, biomass burning, and volcanic emissions of black carbon, organic carbon, and SO2 from 1980 to 2010 for hindcast model experiments. , doi:10.5194/acpd-12-24895-2012 , Atmospheric Chemistry and Physics Discussion 12 : 24895-24954.
Lamarque et al., 2010, Historical (1850–2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: methodology and application , doi:10.5194/acp-10-7017-2010 , Atmospheric Chemistry and Physics 10 : 7017-7039.
van der Werf et al., 2006, Inter-annual variability in global biomass burning emissions from 1997 to 2004. doi:10.5194/acp-6-3423-2006, Climate Change 6 (Atmospheric Chemistry and Physics): 3423-3441.

v. Nitrogen Oxides (NOx)

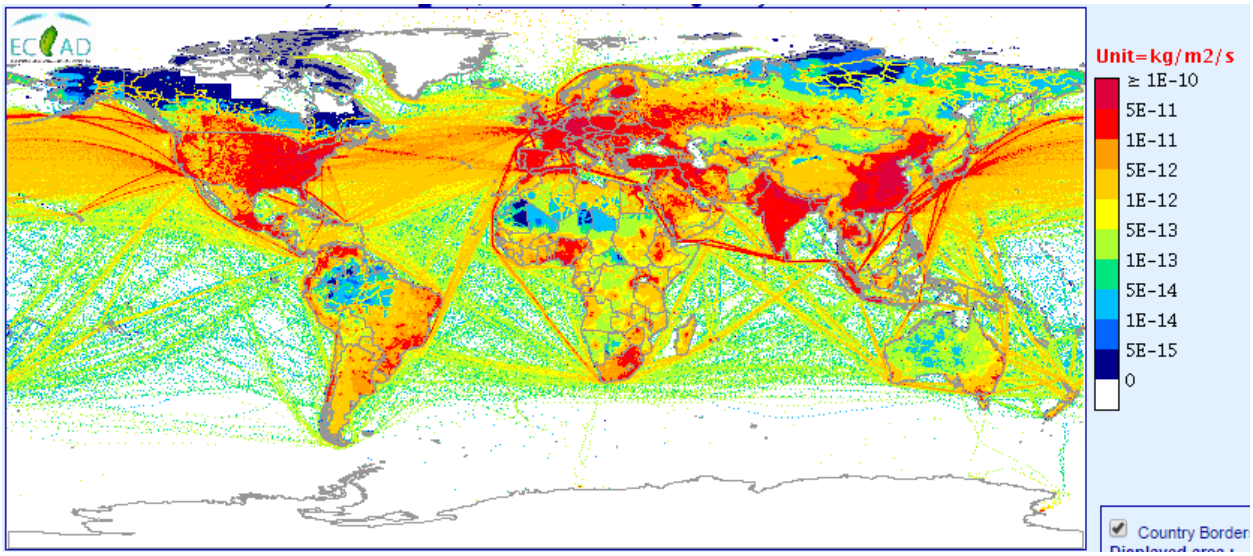
Measure and its definition:

Variable name	**Description**
NOx	Nitrogen Oxides (ECCAD) 2010 in nanogram/m2/s

Explanation:
This dataset measures nitrogen oxides (NOx) at a 0.5 degrees spatial resolution and with a unit of nanogram/m2/s. Nitrogen oxides are released into the air from motor vehicle exhaust or the burning of coal, oil, diesel fuel, and natural gas, especially from electric power plants. They are also released by industrial processes. The data is provided by the MACCity (MACC/CityZEN EU projects) emissions dataset, which is derived from the ACCMIP and RCP8.5 datasets, and distributed by ECCAD. The data is downloaded as .tiff rasters.

The emissions dataset includes data on a yearly basis for the period 1960-2010 for the anthropogenic emissions, and 1960-2008 for the biomass burning emissions. The dataset includes biomass burning emissions that comprise of all emissions resulting from natural and man- made surface vegetation fires, including emissions from the combustion of soil organic matter (duff or peat). Emissions emanating from the combustion of agricultural wastes (e.g. burning rice stubble) or biofuels (e.g. cow dung, charcoal) are excluded.

Figure 28: Average NOx measurements for 2010.



Source: ECCAD (2016).

Calculation:

- Conversion of raster unit from kg/m2/s to nanogram/m2/s due to difficulty in processing extremely small values (as small as 10e-15).
- Zonal statistics of NOx as value raster and P1, P2, G1 and G2 admin levels as feature zones.

Sources:

Claire Granier et al., 2011, Evolution of anthropogenic and biomass burning emissions of air pollutants at global and regional scales during the 1980–2010 period. , DOI: 10.1007/s10584-011-0154-1 , Climate Change 109 (1-2) : 163-190.

Diehl et al., 2012, Anthropogenic, biomass burning, and volcanic emissions of black carbon, organic carbon, and SO2 from 1980 to 2010 for hindcast model experiments. , doi:10.5194/acpd-12-24895-2012 , Atmospheric Chemistry and Physics Discussion 12 : 24895-24954.

Lamarque et al., 2010, Historical (1850–2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: methodology and application , doi:10.5194/acp-10-7017-2010 , Atmospheric Chemistry and Physics 10 : 7017-7039.

van der Werf et al., 2006, Interannual variability in global biomass burning emissions from 1997 to 2004. , doi:10.5194/acp-6-3423-2006 , Climate Chnage 6 (Atmospheric Chemistry and Physics) : 3423-3441.

vi. Mercury

Measure and its definition:

Variable name	**Description**
Mercury	Mercury, Hg (ECCAD) 2000 in nanogram/m2/s

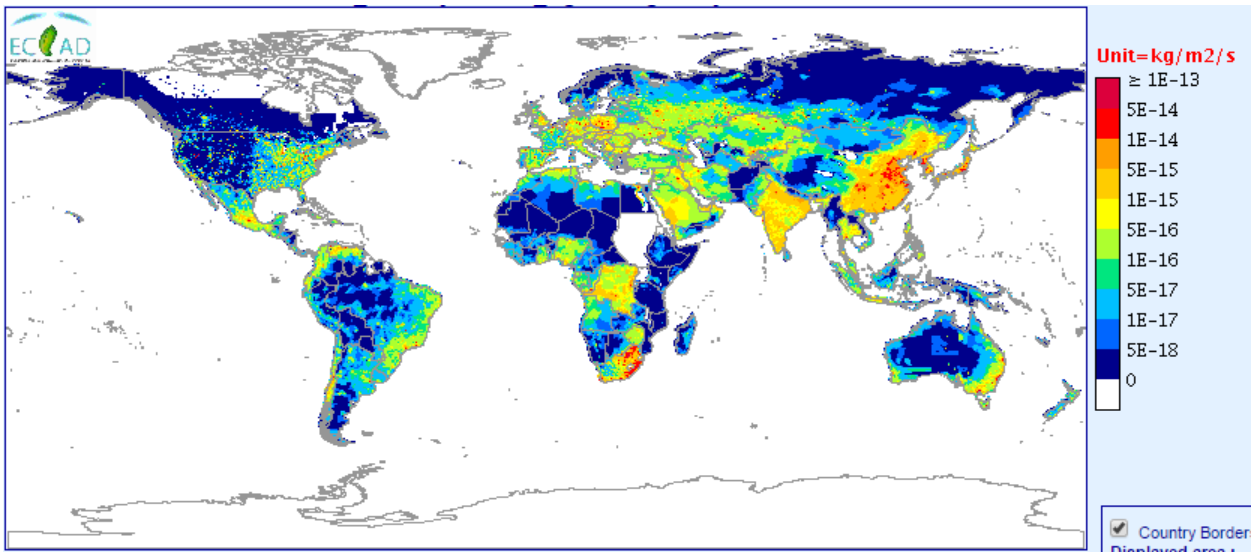
Explanation:

This dataset provides a measure of airborne mercury at a 0.5 degrees spatial resolution and at a unit of nanogram/m2/s. The largest emissions of mercury to the global atmosphere occur from combustion of fossil fuels, mainly coal in utility, industrial, and residential boilers.

The national anthropogenic mercury emission estimates were compiled at the Norwegian Institute for Air Research (NILU) by J. Pacyna and E. Pacyna. The inventories presented here are for the nominal years 1995 and 2000. For some countries with limited data, official emission estimates reported for other years (within 1-2 years of the target nominal year) have been used. In a very few cases where no new emissions estimates were available, 1995 emissions estimates for certain sectors in some countries were also used in 2000.

The methodology estimate the distribution of global mercury emission is described in Pacyna et. al (2003), with an updated methodology for the 2000 emission inventory is described by Wilson et al (2005).

Figure 29 Average mercury, Hg measures for 2010.



Source: ECCAD (2016).

Calculation:

- Conversion of raster unit from kg/m2/s to nanogram/m2/s due to difficulty in processing extremely small values (as small as 10e-15).
- Zonal statistics of Mercury, Hg as value raster and P1, P2, G1 and G2 admin levels as feature zones.

Source:

Pacyna et al., 2006, Global anthropogenic mercury emission inventory for 2000 , Atmospheric Environment 40 (24) : 4048-4063.

Wilson et al., 2006, Mapping the spatial distribution of global anthropogenic mercury atmospheric emission inventories , Atmospheric Environment 40 (24) : 4621-4632.

5.D. Geography, Climate and Mining indicators

i. Land cover

Measures and their definition:

Several measures were produced to represent the land cover types in the regions of interest as well as how the land cover has changed over time. The percentage area of each of the ten categories is calculated creating ten individual variables. Second, the dominant category of land cover within each boundary is identified in a separate variable. The data is calculated for every year, 1992 to 2015.

Variable name	Description
Cropland	Share of Cropland (%), 1992-2015
Forest	Share of Forests (%), 1992-2015
Shrubland	Share of Shrubland (%), 1992-2015
Grassland	Share of Grassland (%),1992-2015
Sparse vegetation	Share of Sparse Vegetation (%),1992-2015
Wetlands	Share of Wetlands (%),1992-2015
Urban area	Share of Urban Area (%),1992-2015
Bare area	Share of Bare Area (%),1992-2015
Water	Share of Water (%),1992-2015
Snow/ice	Share of Snow or Ice (%),1992-2015

Explanation:

The Land Cover data identifies different types of land coverage across the globe. These types have been aggregated to ten different categories: cropland, forest, shrubland, grassland, sparse vegetation, wetlands, urban areas, bare areas, water or snow and ice (summarized in Table 2

below).

The original data is provided as global GeoTiff raster files with a spatial resolution of 300m by the Land Cover (LC) project of the Climate Change Initiative (CCI) led by the European Space Agency (ESA). A surface reflectance time series served as input for generating the global land cover databases for every year, from 1992-2015 The table below summarizes the specifications of the raw data.

Table 14: Summary of data

CCI land cover map	Reference period	Spatial coverage and resolution	Data source
Land cover map	1992-2015	Global / 300m	MERIS 10-year LC map as baseline SPOT-VGT time series for updating

Source: 2017 Land Cover CCI Product User Guide Version 2.0.7.

Calculation:

The raw GeoTiff file provided by CCI has 36 different classes of land cover, which have been combined into ten categories (plus an additional category depicting tree coverage only) for the purpose of our analysis. This table displays the new categories.

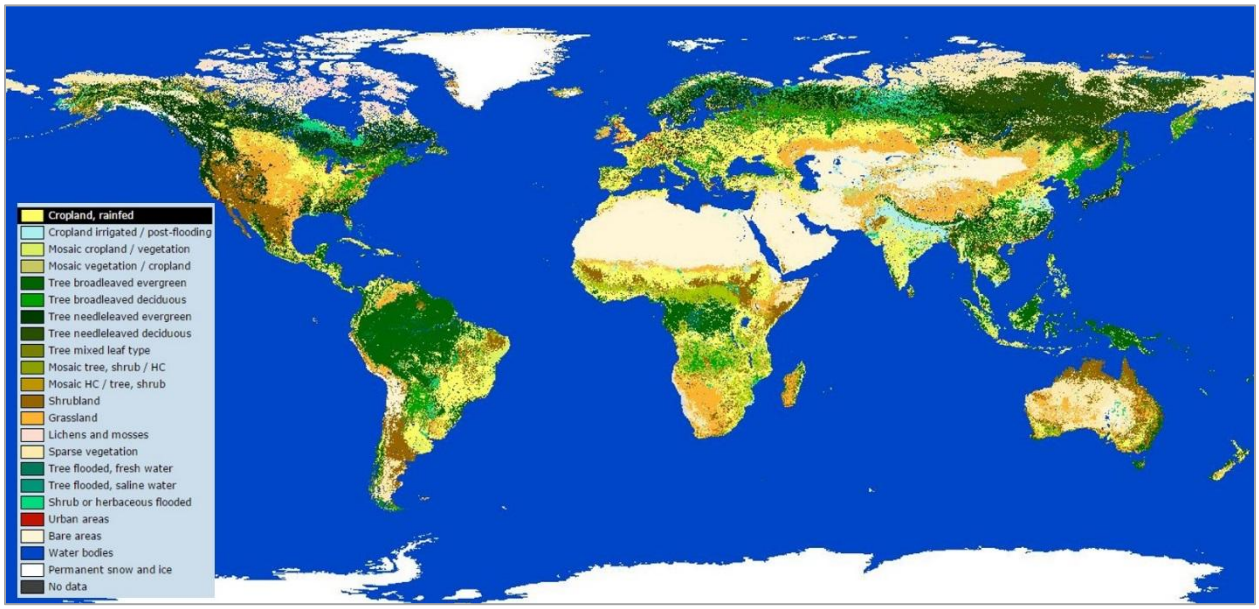
The table below illustrates the re-classification of the categories and the classes that were combined to create the tree cover variable.

Figure 30: Original land cover categories and re-classification process

ESA_Code	ESA raw data: 37 categories	Hidden Dimensions classifications
0	No data	0
10	Cropland rainfed	Cropland
11	Herbaceous cover	Cropland
12	Tree or shrub cover	Cropland
20	Cropland irrigated or post-flooding	Cropland
30	Mosaic cropland (>50%) / natural vegetation (tree shrub herbaceous cover) (<50%)	Cropland
40	Mosaic natural vegetation (tree shrub herbaceous cover) (>50%) / cropland (<50%)	Cropland
50	Tree cover broadleaved evergreen closed to open (>15%)	Forest
60	Tree cover broadleaved deciduous closed to open (>15%)	Forest
61	Tree cover broadleaved deciduous closed (>40%)	Forest

62	Tree cover broadleaved deciduous open (15-40%)	Forest
70	Tree cover needleleaved evergreen closed to open (>15%)	Forest
71	Tree cover needleleaved evergreen closed (>40%)	Forest
72	Tree cover needleleaved evergreen open (15-40%)	Forest
80	Tree cover needleleaved deciduous closed to open (>15%)	Forest
81	Tree cover needleleaved deciduous closed (>40%)	Forest
82	Tree cover needleleaved deciduous open (15-40%)	Forest
90	Tree cover mixed leaf type (broadleaved and needleleaved)	Forest
100	Mosaic tree and shrub (>50%) / herbaceous cover (<50%)	Shrubland
110	Mosaic herbaceous cover (>50%) / tree and shrub (<50%)	Shrubland
120	Shrubland	Shrubland
121	Shrubland evergreen	Shrubland
122	Shrubland deciduous	Shrubland
130	Grassland	Grassland
140	Lichens and mosses	Grassland
150	Sparse vegetation (tree shrub herbaceous cover) (<15%)	Sparse vegetation
152	Sparse shrub (<15%)	Sparse vegetation
153	Sparse herbaceous cover (<15%)	Sparse vegetation
160	Tree cover flooded fresh or brakish water	Wetlands
170	Tree cover flooded saline water	Wetlands
180	Shrub or herbaceous cover flooded fresh/saline/brakish water	Wetlands
190	Urban areas	Urban areas
200	Bare areas	Bare areas
201	Consolidated bare areas	Bare areas
202	Unconsolidated bare areas	Bare areas
210	Water bodies	Water bodies
220	Permanent snow and ice	Snow or ice

Figure 31 Land Cover in 2010 from ESA (European Space Agency)



Sources:
Data: ESA (European Space Agency): http://www.esa-landcover-cci.org/
Annual Land Cover CCI Product User Guide Version 2.0.7, 2017.

ii. Ruggedness

Explanation:
The ruggedness index is a measure of elevation differences originally calculated on the grid cell level by Nunn and Puga (2012). Rugged Terrain is an obstacle to economic development since it hinders trade and most productive activities. The source of elevation data is GTOPO30 (US Geological Survey, 1996) for which elevations are regularly spaced at 30 arc-seconds across the entire surface of the Earth on a map using a geographic projection, so the sea-level surface distance between two adjacent grid points on a meridian is half a nautical mile or, equivalently, 926 meters (Nunn and Puga, 2012). The units for the terrain ruggedness index thus correspond to these units, commonly used to measure elevation differences. The Terrain Ruggedness Index is in millimeters in the 30 arc-seconds grid (due to storage reasons).

Calculation:
The calculation of the terrain ruggedness index was originally developed by Riley, DeGloria, and Elliot (1999) to quantify topographic heterogeneity in wildlife habitats providing concealment for preys and lookout posts. The authors used the following formula, where $e_{r,c}$ denote elevation

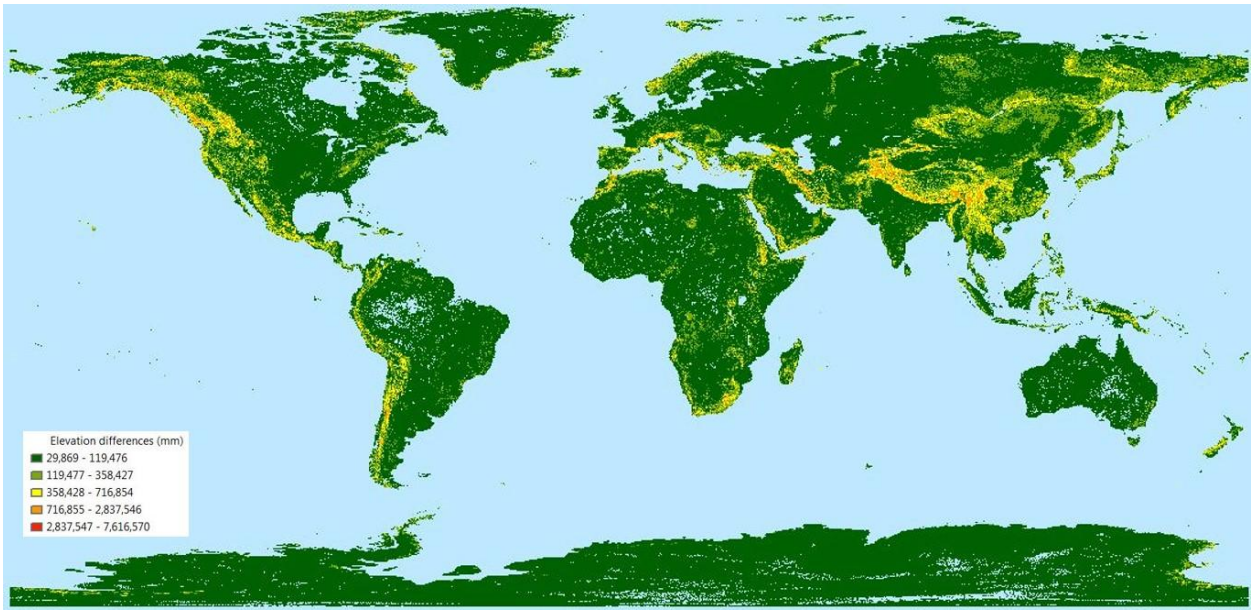
at the point located in row r and column c of a grid of elevation points:

$$\sum_{i=r-1}^{i=r+1}\sum_{j=c-1}^{j=c+1}(e_{i,j}-e_{r,c})^2$$

Nunn and Puga (2012) adapted the method by Riley et al (1999) for the entire globe but also exclude areas covered by permanent inland water area features contained.

In our calculations, the terrain ruggedness index for each point on the grid are averaged across all grid cells within each geographical boundary to obtain the ruggedness of each area (admin 1 or admin 2). In addition to the mean ruggedness index of each area, the minimum, maximum, standard deviation, range and sum is calculated.

Figure 32: Ruggedness Index



Source: Nunn et al, 2012

Sources:

Data: http://diegopuga.org/data/rugged/

Nunn, N. and Puga, P (2012). Ruggedness: The blessing of bad geography in Africa.*Review of Economics and Statistics* 94(1): 20-36.

Riley, Shawn J., Stephen D. DeGloria, and Robert Elliot. 1999. A terrain ruggedness index that quantifies topographic heterogeneity. *Intermountain Journal of Sciences* 5(1-4): 23-27.

US Geological Survey. 1996. GTOPO30. Sioux Falls, SD: United States Geological Survey Center for Earth Resources Observation and Science (EROS)

iii. Less Favored Agricultural Land

We follow the classification of less-favored agricultural land (LFAL) by Pender and Hazell (2000), World Bank (2008) and Barbier and Hochard (2014) in identifying less favored agricultural land versus favored agricultural land, with some minor adjustments and the inclusion of more recent raw data in a few cases[1]. In the classification, less-favored agricultural land (land with low biophysical potential) consists of irrigated land on terrain greater than 8 per cent median slope or poor soil quality; rainfed land with a length of growing period (LGP) of more than 120 days but either on terrain greater than 8 per cent median slope or with poor soil quality; semi-arid land (land with LGP 60-119 days); and arid land (land with LGP < 60 days). The global extent of agriculture is defined as all areas with 10 percent or greater cropland, grazing land or irrigated area net of areas with a growing period of zero days (we follow Sebastian (2006) here)[A1]. Favored agricultural land (land with high biophysical potential) is then all agricultural land that is not defined as LFAL.

Figure 33: Less favored agricultural lands

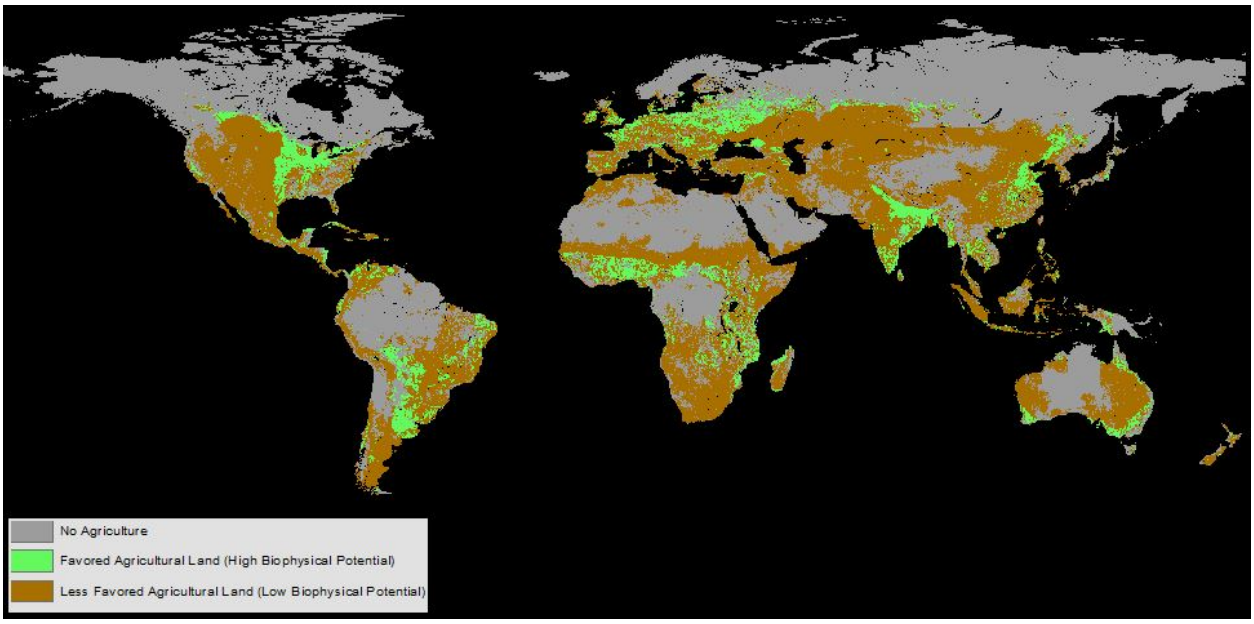


Table 16: Less favored agricultural lands

Raw Data Layers and Format	**Time**	**Sources**
Soil constraints (raster 1km resolution)	2010	FAO Global Agro-Ecological Zones Data Portal version 3.
Length of growing period (LGP) (raster 1km resolution)	2010	

Terrain data (raster 1km resolution)	2010	
Irrigated Land (raster 1km resolution)	2005	Siebert 2006 - The Global Map of Irrigated Areas (GMIA) version 5.
Cropland and Grazing Land (raster 1km resolution)	2000	Ramankutty et al. (2008) – Cropland and Pasture Area

References:

Barbier, E.B and Jacob P. Hochard, 2014. "Land Degradation, Less Favored Lands and the Rural Poor: A Spatial and Economic Analysis." A Report for the Economics of Land Degradation Initiative. Department of Economics and Finance, University of Wyoming.

FAO Global Agro-Ecological Zones Data Portal version 3. Available online: http://gaez.fao.org (Accessed 5 Jan 2017).

Pender, John and Peter Hazell. 2000. "Promoting Sustainable Development in Less-Favored Areas: Overview". Brief 1 in J. Pender and P. Hazell (eds.), Promoting Sustainable Development in Less-Favored Areas. 2020 Vision Initiative, Policy Brief Series, Focus 4. Washington, DC: International Food Policy Research Institute.

Ramankutty, N., A.T. Evan, C. Monfreda, and J.A. Foley (2008), Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000. *Global Biogeochemical Cycles* 22, GB1003, doi:10.1029/2007GB002952.

Sebastian, K. (2007). GIS/Spatial Analysis Contribution to 2008 WDR. http://siteresources.worldbank.org/INTWDR2008/Resources/2795087-1191427986785/SebastianK_ch2_GIS_input_report.pdf (Accessed on 5 January 2017).

Siebert, S.,Döll, P., Feick, S. and J. Hoogeveen. 2006. *Global map of irrigated areas version 4.0* Johann Wolfgang Goethe University, Frankfurt am Main, Germany / Food and Agriculture Organization of the United Nations, Rome, Italy".

World Bank. 2008. Word Development Report 2008: Agricultural Development. The World Bank, Washington DC.

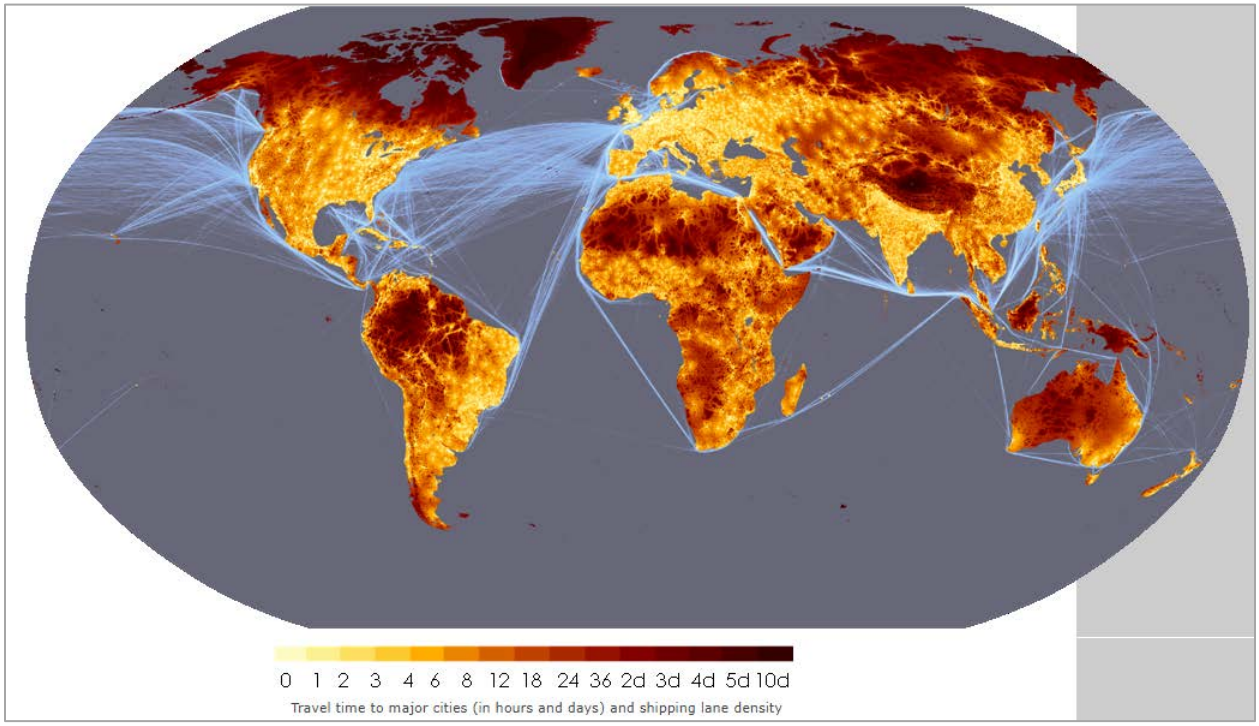
iv. Market access

Description:

Market access is an important factor for contributing to livelihoods, as well for other benefits like social interaction. The European Commission's Joint Research Centre has created a global map of travel time to major cities for the World Bank's World Development Report 2009 (Reshaping Economic Geography), which proxies accessibility to markets and other urban resources. Accessibility is defined as the travel time to a city of 50,000 or more people using road-, of -road-, or water-based travel. Factors that influence travel time include available transport networks, environmental factors like land cover and slope, and political boundaries and borders.

A cost-distance algorithm determines the "cost" of travelling between raster cells, which is generally measured in time units. The resulting raster grid represents a “friction-surface” of these costs. For more information, see here.

Figure 34: Travel time to major cities as a proxy for market access



Source: European Commission’s Joint Research Center

This 0.01-degree grid of travel time to major cities was re-scaled for HDD to reflect the admin-1 and finer-scale political units, where the mean value represents average travel time to major cities within the unit. This grid, and the related Agglomeration Index associated with it, are available only for the year 2000, and thus may not fully reflect accessibility today.

Source:

Uchida, H., and A. Nelson. 2008. “Agglomeration Index: Towards a New Measure of Urban Concentration.” Background paper for the World Bank’s World Development Report 2009. Washington, D.C.: World Bank. http://siteresources.worldbank.org/INTWDR2009/Resources/4231006-1204741572978/Hiro1.pdf.

v. Road density

Description:
The Netherlands Environmental Assessment Agency (PBL) leads the GRIP: Global Roads Inventory Project, replacing its predecessor, the VMAP0 roads database in PBL global environmental assessment models like IMAGE and GLOBIO. GRIP updated nearly 200 countries from this previous version through data collected from about 60 public sources including the United Nations, national spatial data sources, topographic agencies, non-governmental and international organizations, and volunteered geographic information. GRIP uses the UN UNSDI-T version 2 light data model.

HDD includes road density as the mean of length of road kilometers per cell, and is re-scaled from GRIP's original 10km resolution to the political boundaries of the administrative units used in HDD. Values are for the year 2013, and include total road density, as well as disaggregated densities of highways, primary roads, secondary roads, tertiary roads, and urban/residential roads.

Source:
PBL. 2014. GRIP: Global Roads Inventory Project.
http://geoservice.pbl.nl/geonetwork/srv/nl/main.home?uuid=8f1a8f7b-1474-43e1-ac4a-3d913197381a

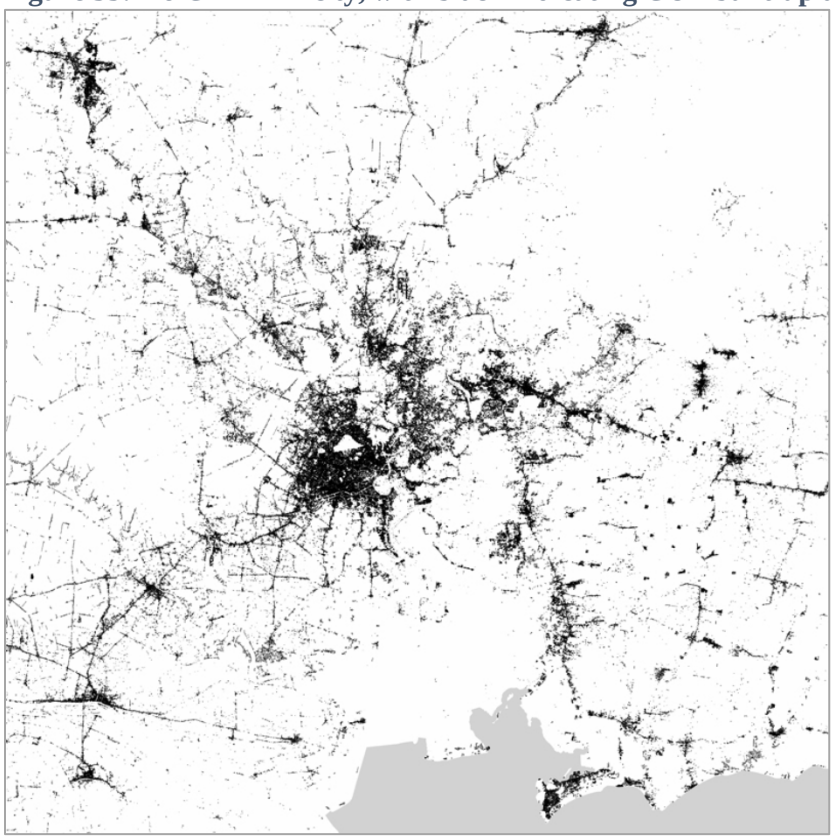
vi. Built-up areas
More than half of the world's population are in urban areas, and this figure will likely only increase. Urban areas are centers of human activity, meaning that the environmental, economic, political, societal and cultural impacts of urbanization are significant.

The "Global Urban Footprint" (GUF) data has been generated by the Urban Footprint Processor (UFP), implemented at the German Remote Sensing Data Center (DFD) of the German Aerospace Center (DLR). It shows the worldwide mapping of settlements at a spatial resolution of 0.4 arcsec (~12 m), processed with around 180 000 TerraSAR-X and TanDEM-X scenes.

The processing consists of three basic steps. First, a texture feature (speckle divergence) is extracted from all original amplitude images in order to highlight areas characterized by highly diverse and heterogeneous backscattering – a typical characteristic of built-up areas in radar imagery that results from the double bounce effects from buildings and other vertical structures in combination with extensive shadow regions. Secondly, a fully-automated classification procedure derives a binary settlement layer for each scene based on both the corresponding amplitude and texture images. Thereby, pixels exhibiting high values in the texture and the amplitude images are defined as built-up areas, whereas all remaining regions are assigned to

the class non-built-up area. The third and final step is the mosaicking and post-processing of the data, supported by a semi-automated quality assessment. More details on the GUF methodology and UFP technique are provided by the publications listed in the References section of the GUF webpage. From Global Urban Footprint (2012)

Figure 35: Ho Chi Minh city, with black indicating GUF built-up areas



Source: GUF Gallery.

Data:
Global Urban Footprint (2012). http://www.dlr.de/eoc/en/desktopdefault.aspx/tabid-11725/20508_read-47944/.

Sources:
Esch, T., Elsayed, S., Marconcini, M., Marmanis, D., Zeidler, J., Dech, S. (2014): Dimensioning the Degree of Urbanization – A Technical Framework for the Large-scale Characterization of Human Settlement Forms and Patterns based on Spatial Network Analysis. Submitted to Journal of Applied Geography.

Esch, T., Marconcini, M., Felbier, A., Roth, A., Heldens, W., Huber, M., Schwinger, M., Taubenböck, H., Müller, A., Dech, S. (2013): Urban Footprint Processor – Fully Automated

Processing Chain Settlement Masks from Global Data of the TanDEM-X Mission. IEEE Geoscience and Remote Sensing Letters, Vol. 10, No. 6, pp. 1617-1621. http://dx.doi.org/10.1109/LGRS.2013.2272953.

Esch, T., Taubenböck, H., Roth, A., Heldens, W., Felbier, A., Thiel, M., Schmidt, M., Müller, A., Dech, S. (2012): TanDEM-X Mission - New Perspectives for the Inventory and Monitoring of Global Settlement Patterns. Journal of Applied Remote Sensing, Vol. 6, No. 1, 061702 (October 04, 2012); 21 pp. http://dx.doi.org/10.1117/1.JRS.6.061702.

Esch, T., Schenk, A., Ullmann, T., Thiel, M., Roth, A. and S. Dech (2011): Characterization of Land Cover Types in TerraSAR-X Images by Combined Analysis of Speckle Statistics and Intensity Information. IEEE Transactions on Geoscience and Remote Sensing. Vol. 49, No.6, pp 1911 – 1925. http://dx.doi.org/10.1109/TGRS.2010.2091644.

Esch, T., Thiel, M., Schenk, A., Roth, A., Muller, A. and Dech, S. (2010): Delineation of Urban Footprints From TerraSAR-X Data by Analyzing Speckle Characteristics and Intensity Information. IEEE Transactions on Geoscience and Remote Sensing, Vol. 48, No. 2, pp. 905-916. http://dx.doi.org/10.1109/TGRS.2009.2037144.

v. Urban areas and definitions

Urban definitions vary from one country to another, and is rather arbitrary. While we easily accept mega cities such as Delhi to be considered urban, and similarly, isolated farming villages in sparsely-populated areas of India to be rural, the urban-rural classification is usually subjective and country-specific. For this global database, we used a global urban definition that could be applied consistently throughout all countries.

In our urban/rural definition, we define a city along the lines of the European Union's Joint Research Center. An urban cluster is defined as an area with a population greater than 5,000 inhabitants and a population density above 300 inhabitants per square kilometer (Eurostat, 2017). This is our working definition for an urban area. In addition, high-density cities, or urban centers per Eurostat nomenclature, are areas with a population greater than 50,000 inhabitants and a population density of above 1,500 inhabitants per square kilometer.

Source:

Joint Research Center, Eurostat, European Commission of the European Union (2017): Developing a global, people-based definition of cities and settlement.

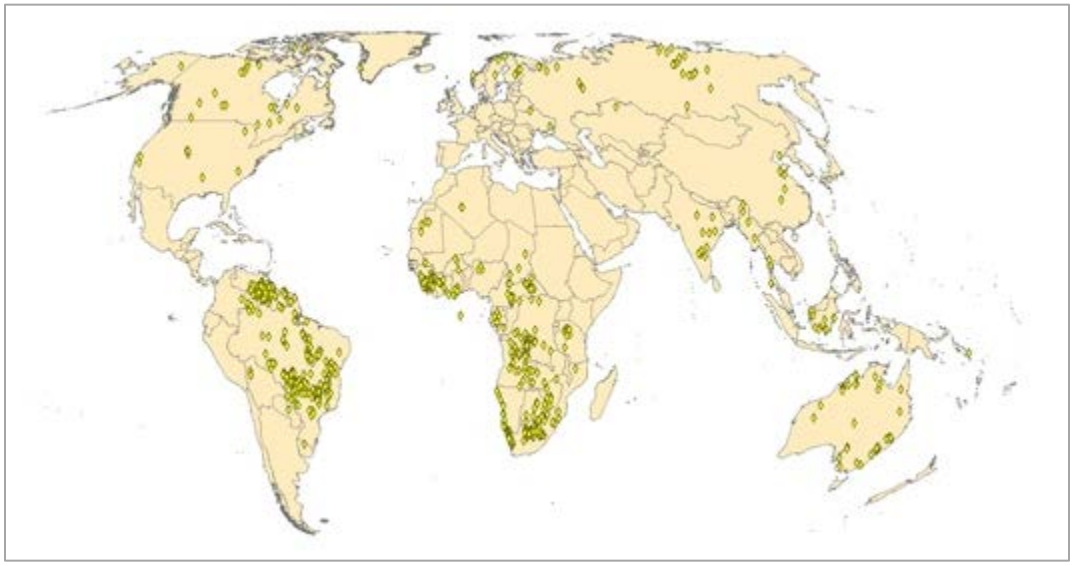
vii. Diamond deposits

The diamonds dataset offers a comprehensive list of all known diamond deposits throughout the world. Each deposit is coded with precise geographic coordinates, geological form of the

diamond, and dates of discovery and first production. The dataset is available as GIS shapefiles as well as in spreadsheet (Excel) format. See codebook for definitions and further details. (Peace Research Institute Oslo, PRIO, 2005)

For HDD, we calculated the count of diamond deposits in each admin unit in GIS.

Figure 36: Known diamond deposits



Data:
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. https://www.prio.org/Data/Geographical-and-Resource-Datasets/Diamond-Resources/.

Source:
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.

viii. Oil and gas deposits

The petroleum datasets contain information on all known oil and gas deposits throughout the world. Two datasets are available: one for on-shore deposits and another one for off-shore deposits.

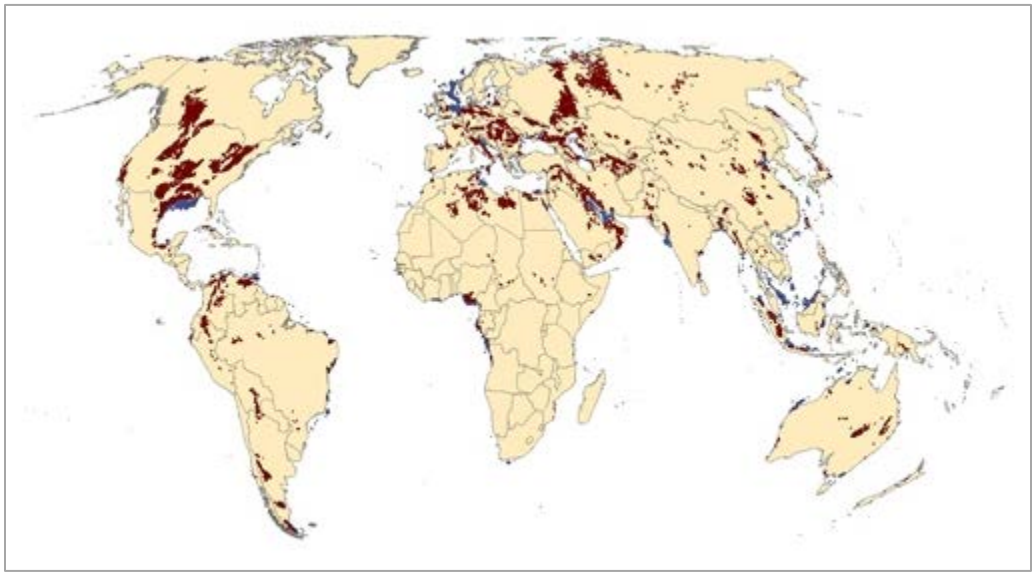
Given the geological characteristics of petroleum, the data are stored as polygons in shapefile

(GIS) format. In addition, basic data on each deposit are available as spreadsheets (Excel). Here, spatial information is limited to latitude and longitude coordinates for the center point of each deposit (polygon).

Aside from exact locational information, each deposit are registered with type of resource (oil and/or gas), discovery- and production dates (whenever known), name of petroleum basin, geographic coordinates of polygon centroid, and primary source of information.
(Peace Research Institute Oslo, PRIO, 2005)

For HDD, we only include the on-shore deposits. We calculated the number of oil and gas deposits in each administrative unit.

Figure 37: Known oil and gas deposits



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

ix. Climate - temperature and precipitation

Description:
Our temperature and precipitation data was extracted from the gridded Climate Research Unit (CRU) version 4 time-series data, developed by the Centre for Environmental Data Analysis

(CEDA). The original CRU data are monthly gridded cells (0.5-degree resolution) based on observational data from National Meteorological Services and other external agents. The precipitation variable provided is total volume of precipitation (in mm) per month, and the temperature variable is mean temperature in degrees Celsius. We collapsed both the precipitation and temperature datasets to an annual average, calculating the average value from the monthly data. The annual averages were calculated for each year between 1996 and 2015.

Calculation
Similarly to other variables, we summarize the temperature and precipitation annual averages to our regions of interest (admin 1 and admin 2) using the mean value. The resulting variables are structured as follows: *tmp_admin_year* (temperature), and *precip_admin_year* (precipitation).

Data: http://data.ceda.ac.uk/badc/cru/data/cru_ts/cru_ts_4.00/data/.

Source:
University of East Anglia Climatic Research Unit; Harris, I.C.; Jones, P.D. (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, *25 August 2017*. doi:10.5285/edf8febfdaad48abb2cbaf7d7e846a86. http://dx.doi.org/10.5285/edf8febfdaad48abb2cbaf7d7e846a86.

6. References

Elbers, Chris; Lanjouw, Jean (2000): Welfare in Village and Towns: Micro Measurement of Poverty and Inequality, Tinbergen Institute Discussion Paper, No. 00-029/2

Elbers, C., Lanjouw, J. O., & Lanjouw, P. (2002). Micro-level estimation of welfare (Vol. 2911). World Bank Publications.

Elbers, C., Lanjouw, J. O., & Lanjouw, P. (2003). Micro–level estimation of poverty and inequality. Econometrica, 71(1), 355-364.

Appendix A: Nested Dataset

To graphically illustrate the nested structure of the dataset, we plotted both state level and district level measures for the state of Karnataka in India (see figure A1) as well as the province and district level for Bangladesh and Hungary (see figure A2).

Figure A1: India

Panel A: Province observation **Panel B: District observation**

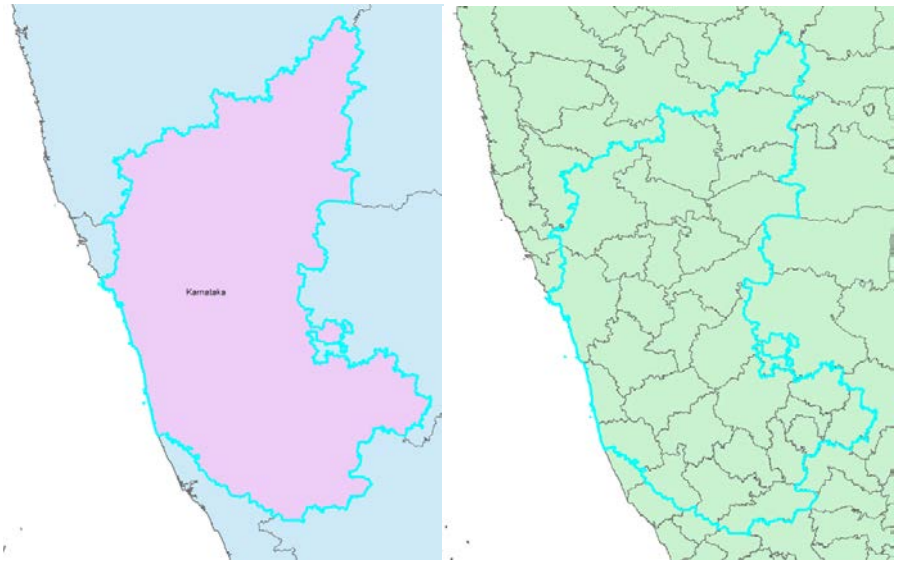
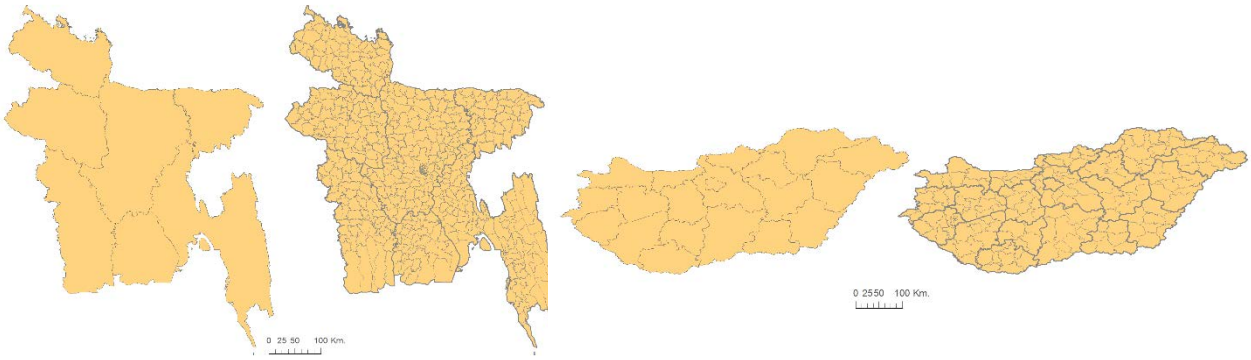


Figure A2: Example of Admni-1 (left) and Admin-2 (right) boundaries for Bangladesh and Hungary.



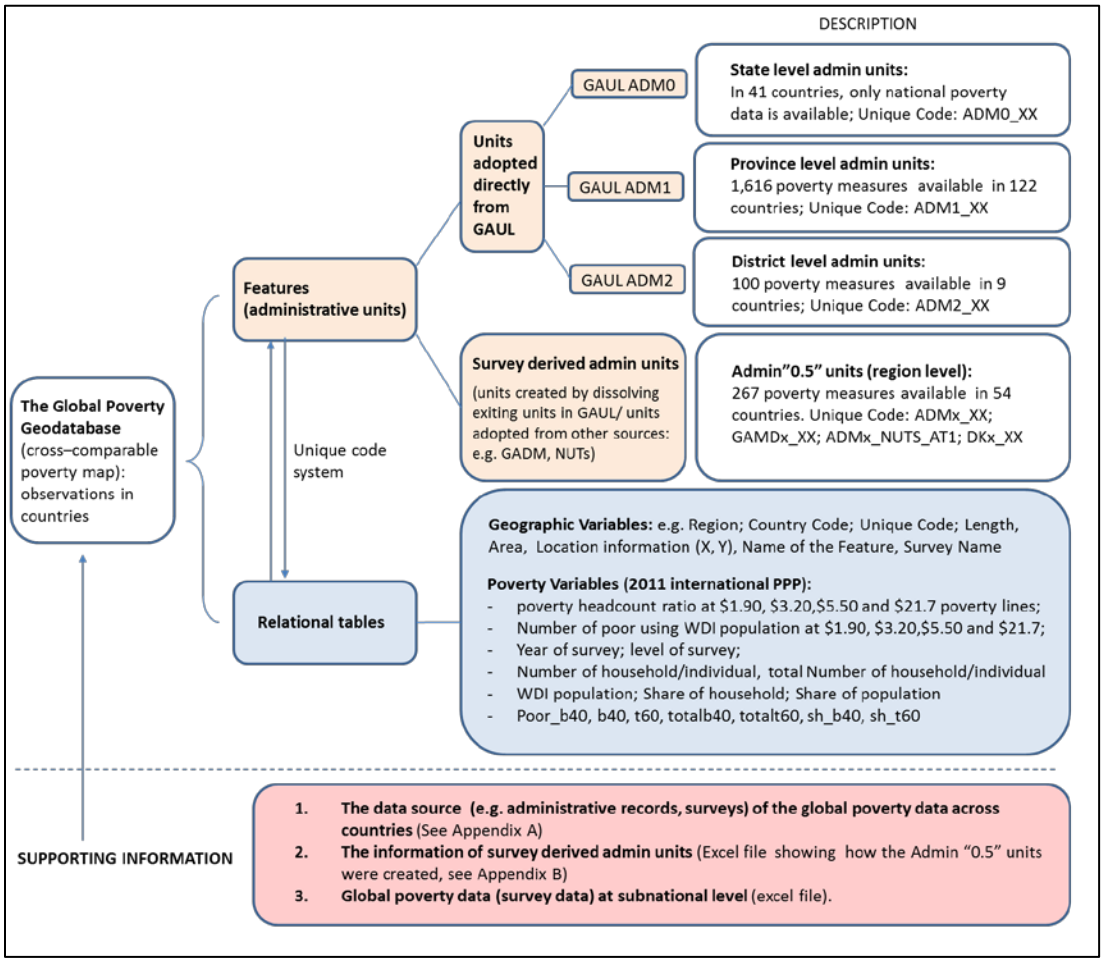
Appendix B: HDD User manual

The figure below shows the structure of the HDD. There are three subnational sets of data, all nested under the country level (admin 0) and all variables exist on all three levels (unless not available on a certain level). The ENR variables are available globally and on all admin levels, whereas the poverty variables are restricted to specific geographical admin levels and countries. The Global Administrative boundaries - GADM0 (country), GADM1 (province/state) and GADM2 (district/municipality) - are referred to as G0, G1 and G2 in the dataset respectively. Since the World Bank poverty data follows national boundaries (that sometimes correspond to GADM boundaries but not always), all ENR variables have also been aggregated to these national boundaries –referred to as PX in the dataset. The 'X' indicates that the boundaries are on different admin levels. See table 4 for a list of countries with poverty data and most detailed admin level. PX boundaries are nested within G1.

Figure A3: Nesting structure of HDD

Breakdown of poverty data, at national poverty line and international poverty line (cross-

comparable)



Working with different admin levels

One cannot simply run analysis on the master HD dataset. First, one must decide which administrative level to work with. The HDD dataset is structured in a way that makes it easy to work with one desired administrative level at a time. For example, to work with just the G1 level, the user should run the following command in Stata:

keep if gadm1_only==1

The screenshot below shows that one observation for each G1 unit is kept in the dataset.

country_name	OBJECTID_GADMadm1	name_gadm_1	FOR_G1_cover_area
Madagascar	1719	Fianarantsoa	30545.724609375
Madagascar	1720	Mahajanga	31860.01953125
Madagascar	1721	Toamasina	54164.8984375
Madagascar	1722	Toliary	22205.58203125
Malawi	1723	Balaka	27.1031265258789
Malawi	1724	Blantyre	72.6035079956055
Malawi	1725	Chikwawa	517.45947265625
Malawi	1726	Chiradzulu	20.7216529846191
Malawi	1727	Chitipa	881.554260253906
Malawi	1728	Dedza	315.525024414063

Similarly, the following command should be used if the user wants to work with G2:

keep if gadm2_only==1

Going back to the example where the user is only working with G1. After running the ‘keep if gadm1_only==1’ command, the dataset is ready for analysis. At this point, the user can also simplify the dataset and speed up processing by keeping all G1 variables (and dropping all other variables). The easiest way to do this is to type G1 in the search field in the Variables box in Stata and select and copy all variables that remain, then run the ‘keep’ command followed by all G1 variables copy-pasted. See screenshot below.