Documentation for the National Aggregates of Geospatial Data Collection: Population, Landscape, And Climate Estimates, Version 4 (PLACE IV)

May 2022

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

Abstract

This document outlines the basic methodology and data sets used to construct the Population, Landscape And Climate Estimates, Version 4 (PLACE IV), along with use cases, limitations, and use constraints. PLACE IV provides measures of population (head counts) and land area (square kilometers) as totals and by urban and rural designation, within multiple biophysical themes for 248 statistical areas (countries and other territories recognized by the United Nations (UN)), UN geographic regions and subregions, and World Bank economic classifications. It improves upon previous versions by providing these estimates at both the national level, and where possible, at subnational administrative level 1 for the years 2000, 2005, 2010, 2015, and 2020, and by 5-year and broad age groups for the year 2010.

Data set citation: Center for International Earth Science Information Network (CIESIN), Columbia University. 2022. National Aggregates of Geospatial Data Collection: Population, Landscape, And Climate Estimates, Version 4 (PLACE IV). Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/rht8-jv78. Accessed DAY MONTH YEAR.

Suggested citation for this document: Center for International Earth Science Information Network (CIESIN), Columbia University. 2022. Documentation for the National Aggregates of Geospatial Data Collection: Population, Landscape, And Climate Estimates, Version 4 (PLACE IV). Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/y2wq-pv06. Accessed DAY MONTH YEAR.

We appreciate feedback regarding this data set, such as suggestions, discovery of errors, difficulties in using the data, and format preferences. Please contact:

NASA Socioeconomic Data and Applications Center (SEDAC)
Center for International Earth Science Information Network (CIESIN)
Columbia University
Phone: 1 (845) 365-8920
Email: ciesin.info@ciesin.columbia.edu

Contents

I. Introduction

The *National Aggregates of Geospatial Data Collection (NAGDC)* converts geospatial data into national and subnational level data in tabular formats as a service to researchers and analysts who do not have access to geoprocessing tools. The Population, Landscape, And Climate Estimates, Version 4 (PLACE IV) data set provides measures of population and land area for 248 statistical areas (countries and other territories recognized by the United Nations (UN)). PLACE IV improves upon previous versions by providing these estimates at both the national level, and where possible, at subnational administrative level 1 for the years 2000, 2005, 2010, 2015, 2020, and by 5-year and broad age groups for the year 2010. Zonal statistics are computed to estimate the number of people (head counts) and the land area (square kilometers), as totals and by urban and rural designation, within multiple biophysical themes for each administrative unit, United Nations geographic region and subregion, and World Bank economic classification. The biophysical themes include: biomes, climate zones, coastal proximity zones, elevation zones, and population density zones. PLACE IV provides these country characteristics in tabular format so they may be used in statistical analyses. It converts geospatial data into national and subnational level data in tabular formats as a service to researchers and analysts who do not have access to geoprocessing tools.

Input data were chosen that met the following criteria:

- Global in scope (though some omit coverage for Polar Regions).
- Capable of meaningful aggregation at the national and subnational levels.
- Relevant to understanding human-environment interactions.

II. Data and Methodology

Input data

PLACE IV is produced by aggregating gridded population counts and land area values within zones representing the intersection of administrative units, urban and rural areas, and classes from a suite of biophysical themes. Additional country level attributes in the final table allow for filtering by regional geography and economic status. All source data used to represent these inputs are openly available and described below. The data source for each PLACE IV input layer is included in Table 1.

Table 1. PLACE IV input layers and sources

Global Coastline (analysis mask)	Gridded Population of the World, Version 4 (GPWv4): National Identifier Grid, Revision 11 (CIESIN, 2018c)
Variables (Values)	
Population counts	Gridded Population of the World, Version 4 (GPWv4): Population Count Adjusted to Match 2015 Revision of UN WPP Country Totals, Revision 11 (CIESIN, 2018d) Gridded Population of the World, Version 4 (GPWv4): Basic Demographic Characteristics, Revision 11 (CIESIN, 2018a)
Land area estimates	Gridded Population of the World, Version 4 (GPWv4): Land and Water Area, Revision 11 (CIESIN, 2018b)
Administrative Areas (Zones)	
Level 0 administrative units	Gridded Population of the World, Version 4 (GPWv4): National Identifier Grid, Revision 11 (CIESIN, 2018c)
Level 1 administrative units	Administrative vector boundary inputs to Gridded Population of the World, Version 4 (GPWv4): Revision 11
Rural and urban areas	GHS Settlement Model grid (GHS-SMOD R2019A) (Pesaresi et al., 2019)
Biophysical Themes (Zones)	
Biomes	Terrestrial Ecoregions of the World (Olson et al., 2001)
Climate zones	Observed and projected climate shifts 1901-2100 depicted by world maps of the Köppen-Geiger climate classification (Rubel and Kottek, 2010)
Coastal proximity zones	Shoreline from: Gridded Population of the World, Version 4 (GPWv4): National Identifier Grid, Revision 11 (CIESIN, 2018c)

Elevation zones	Multi-Error-Removed Improved-Terrain Digital Elevation Model (MERIT-DEM) (Yamazaki et al., 2017)
Population density zones	Gridded Population of the World, Version 4 (GPWv4): Population Density Adjusted to Match 2015 Revision of UN WPP Country Totals, Revision 11 (CIESIN, 2018e)
Geographic Regions and Economic Classification	
Geo Regions and Geo Subregions	United Nations Statistics Division Standard country or area codes for statistical use (UNSD M49)
Income Groups and Lending Categories	World Bank list of economies (World Bank, 2019)

Gridded Population of the World, Version 4 (GPWv4): Revision 11

The Gridded Population of the World, Version 4 (GPWv4): Revision 11 data collection (https://sedac.ciesin.columbia.edu/data/collection/gpw-v4) provided source data for several inputs to PLACE IV. GPW models the distribution of human population (counts and densities) as continuous global raster surfaces using population census tables and corresponding geographic boundaries as inputs. Version 4 inputs were sourced from the 2010 round of Population and Housing Censuses, which occurred between 2005 and 2014. The GPWv4 Revision 11 collection provides spatially disaggregated population layers that are compatible with data sets from social, economic, and Earth science disciplines, and remote sensing. These globally consistent and spatially explicit data layers are available at the native 30 arc-second resolution as well as four coarser resolutions, and are suitable for use in research, policy-making, and communications.

Because most countries' national statistical offices report census data values that differ from United Nations population estimates, GPWv4.11 population estimates are made available as either, adjusted to the UN estimates, or in their unadjusted form. For the UN adjustment, a national-level conversion factor representing the difference between the GPW estimate and the UN estimate is applied to the population values. The GPWv4.11 adjusted estimates are based on the estimates published in the 2015 Revision of the United Nations World Population Prospects (UN WPP) and were used for all population calculations in PLACE IV.

Details on the methodologies used to produce the various GPWv4.11 data sets are available in the data collection documentation (CIESIN, 2018f).

The following GPWv4.11 data sets, at 30 arc-second resolution, were used as inputs to PLACE IV:

1. *Population Counts (totals and by age and sex) and Densities*

   The UN WPP-Adjusted Population Count, v4.11 data set provided gridded population estimates for the years 2000, 2005, 2010, 2015, and 2020.

   The Basic Demographic Characteristics, v4.11 data set provided population estimates by sex as totals and for 5-year and broad age groups for year 2010.

   These two data sets provided 64 population variables, which were used as inputs to zonal statistics to estimate the variable within each administrative unit and each class from the biophysical data themes.

   The UN WPP-Adjusted Population Density, v4.11 data set consists of estimates of human population density (number of persons per square kilometer) for the years 2000, 2005, 2010, 2015, and 2020. These rasters were produced by dividing the GPWv4.11 population count rasters by the GPWv4.11 Land Area grid. The population density rasters were used to create rasters of population density zones for the years 2000, 2005, 2010, 2015, and 2020, to be used to compute zonal statistics.

2. *Land Area Estimates*

   The land area raster from the Land and Water Area, v4.11 data set provided estimates of the land area in square kilometers, excluding permanent ice and water, within each pixel. These data were used to estimate the land area within each administrative unit and each class from the biophysical data themes.

3. *National Identifier Grid*

   The National Identifier Grid, v4.11 data set includes a raster representation of nation-states in GPWv4, and a polygon layer derived from that raster. Note that these data are not official representations of country boundaries; rather, they represent the area covered by the input data. This data set served as the framework for all PLACE IV processing. It was used to create an analysis mask for raster processing, a coastline layer for generating coastal zones, and for computing zonal statistics on population and land area estimates at the Admin 0 level.

The vector administrative boundaries compiled for input to GPW Version 4, Revision 11 were used to create a raster of level 1 administrative units for use in PLACE IV. Due to country-dependent terms of use for these input boundary data sets, the level 1 raster is not publicly available.

Terrestrial Ecoregions of the World (TEOW)

The Biomes data used in PLACE IV were obtained from the World Wildlife Fund (WWF) Terrestrial Ecoregions of the World (TEOW) Version 2 data set (Olson et al., 2001). The TEOW data set is a global shapefile, which depicts the world's 825 ecoregions delineated based on current and historical terrestrial vegetation biodiversity patterns. The ecoregions nest hierarchically into 14 biomes and 8 biogeographic realms. Together, this nested classification "provides a useful framework for conducting biogeographical or macroecological research, for identifying areas of outstanding biodiversity and conservation priority, for assessing the representation and gaps in conservation efforts worldwide, and for communicating the global distribution of natural communities on earth" (excerpted from data set abstract in the data set metadata). The data set was first published in 2001 and updated in 2004 (Version 2) according to the accompanying metadata. Table 2 describes the 14 biomes, plus zones of "Lake" and "Rock and Ice", represented in the TEOW data set.

Table 2. Descriptions of Biomes (Olson et al., 2001)

Boreal Forests/Taiga	Low annual temperatures characterize northerly latitudes; precipitation ranges from 40–100 centimeters per year and may fall mainly as snow. Soil is nutrient poor, largely due to permafrost and poor drainage.
Deserts and Xeric Shrublands	The amount of rainfall varies; generally, however, evaporation exceeds rainfall, which is usually less than 25.4 centimeters annually. Temperature variability is also diverse; the Sahara is hot all year while the Gobi has a cold winter.
Flooded Grasslands and Savannas	These areas support numerous plants and animals adapted to the unique hydrologic regimes and soil conditions, and may host large congregations of migratory and resident waterbirds. Examples include the Everglades and Pantanal.
Lakes	Lakes
Mangroves	Occurring in the waterlogged, salty soils of sheltered tropical and subtropical shores, they are subject to the twice-daily ebb and flow of tides, and seasonal weather fluctuations. They stretch from the intertidal zone up to the high-tide mark.
Mediterranean Forests, Woodlands, and Scrub	Characterized by hot and dry summers, while winters tend to be cool and moist. Although the habitat is globally rare, it features an extraordinary biodiversity. Most plants are fire adapted, and dependent on this disturbance for their persistence.
Montane Grasslands and Shrublands	High elevation (montane and alpine) grasslands and shrublands, including the puna and paramo in South America, subalpine heath in New Guinea and East Africa, and steppes of the Tibetan plateaus. They are tropical, subtropical, and temperate.

Rock and Ice	Rock and Ice
Temperate Broadleaf and Mixed Forests	Temperate forests experience a wide range of variability in temperature and precipitation. In regions where rainfall is broadly distributed throughout the year, deciduous trees mix with species of evergreens.
Temperate Coniferous Forests	Temperate evergreen forests are found in areas with warm summers and cool winters, and vary enormously in their kinds of plant life. In some, needleleaf trees dominate, while others are home primarily to broadleaf evergreen trees, or a mix of both.
Temperate Grasslands, Savannas, and Shrublands	Prairies in North America, pampas in South America, veld in Southern Africa, and steppe in Asia. These regions are devoid of trees, except for riparian or gallery forests around streams and rivers. Diverse floral communities and large grazing mammals.
Tropical and Subtropical Coniferous Forests	Found mainly in North and Central America, these tropical regions experience low levels of precipitation and moderate variability in temperature. They have diverse species of conifers and a thick, closed canopy.
Tropical and Subtropical Dry Broadleaf Forests	These forests occur in climates that are warm year-round. Although they may receive several hundred centimeters of rain per year, they deal with long dry seasons, which last several months. Deciduous trees predominate these forests.
Tropical and Subtropical Grasslands, Savannas, and Shrublands	Characterized by rainfall levels between 90-150 centimeters per year, which is not enough rain to support extensive tree cover. Grasses and large mammals dominate.
Tropical and Subtropical Moist Broadleaf Forests	Generally found in large, discontinuous patches between the Tropics of Cancer and Capricorn; characterized by low variability in temperature and > 200 centimeters annual rainfall. The forest is dominated by semi-evergreen and evergreen deciduous tree species.
Tundra	A treeless polar desert found in the high latitudes in the polar regions, primarily in Alaska, Canada, Russia, Greenland, Iceland, and Scandinavia. The region's long, dry winters feature months of total darkness and extremely frigid temperatures.

Observed and projected climate shifts 1901-2100 depicted by world maps of the Köppen-Geiger climate classification

Köppen-Geiger Climate Classification maps of the world, acquired from the Vienna Institute of Veterinary Public Health (Rubel and Kottek, 2010), were selected to represent

global climatological regions based on observed climate data for the period 1976-2000, as well as projected climate shifts for four time periods. The classification system is based on annual and monthly averages of temperature and precipitation ranges. For observed data, two separate data sets were used; the Climatic Research Unit (CRU) of the University of East Anglia (Mitchell and Jones, 2005) for temperature, and the Global Precipitation Climatology Centre (GPCC) Full Data Reanalysis Version 4 for 1901 to 2007 (Fuchs, 2008) for precipitation. For projected data, the Tyndall Centre for Climate Change Research data set, TYN SC 2.03 (Mitchell et al., 2004) was used. From here, data was averaged over periods of 25 years. These results show ensemble-means runs against 5 Global Climate Model (GCM) projections illustrating 4 emissions scenarios described by the Intergovernmental Panel on Climate Change (IPCC) (Nakicenovic et al., 2000) for the periods 2001–2025, 2026–2050, 2051–2075, and 2076–2100.

Climate zones are grouped into five super categories, based on general annual distributions of temperature and rainfall. Tropical systems are coded as “A”, Dry systems as “B”, Temperate systems as “C”, Cold systems as “D” and Polar systems as “E”. Each of these five main classes contains combinations of subclass identifiers based on seasonality, precipitation and temperature patterns. These classifications are reflected in climate zone abbreviations, which are composed of three letters (Table 3). Although there are 34 possible three-letter combinations, three of these (Csd, Cwd, and Cfd) cannot be realized as defined by their temperature requirements. Of the remaining 31 possible classes, one (Dsd) never occurs in the projected climate shift maps.

The IPCC emissions scenarios under which the projected climate shifts are modeled are described in the Special Report on Emissions Scenarios (Nakicenovic et al., 2000) and are summarized below:

- A1 (FI: fossil fuel intensive): rapid economic growth; launch of new/efficient technologies; economic, social, and cultural homogenization; fossil fuel intensive energy sources
- A2: heterogeneous world with focus on national self-reliance, regionally oriented economic development, and increasing global population.
- B1: convergent world as in A1, but with shift towards service and information economy, reduction in material intensity, and launch of clean technologies; focus on global solutions.
- B2: characterized by local and regional solutions for economic, social, and environmental sustainability, and less rapid, more diverse technological change.

The Rubel and Kottek climate classification map covers global land areas and was acquired as global layers in ASCII format at 0.5 degree resolution for each time period and IPCC emission scenario, and the observed climate data for the period 1976-2000. Climate classes are abbreviated according to the climate classification established by Köppen and Geiger and illustrated in Table 3.

A more current climate classification data set based on the newer Representative Concentration Pathway (RCP) scenarios is available (https://essd.copernicus.org/preprints/essd-2021-53/), but the data have not yet been peer-reviewed so the Rubel and Kottek data were retained for PLACE IV.

Table 3: The Köppen Climate Classification System (FAO, 2006a and 2006b)

First letter: *Main climate*	A	Tropical: Temperature of the coldest month: > 18°C. This is the climate where the most water- and heat-demanding crops (e.g. for instance oil palm and rubber) are grown. The climate is also ideal for yams, cassava, maize, rice, bananas, and sugarcane.
	B	Dry: Arid regions where annual evaporation exceeds annual precipitation. Even the wettest variants of this climate are characterized by a marked dry season. The climate is, therefore, mostly unsuitable for the crops that require year-round moisture. The main crops are usually millet, sorghum and groundnuts. Sunshine is usually high, which leads to high productivity where a sufficiently long rainy season or irrigation ensures a sufficient water supply; rice, sugarcane and maize are also common crops under this climate.
	C	Temperate: Average temperature of the coldest month < 18°C and > -3°C, and average temperature of warmest month > 10°C. The main crops are the temperate cereals such as wheat, barley and Irish (white) potatoes. An important variant of this climate is the Mediterranean climate, characterized by the olive tree, which is also very suitable for grapes.
	D	Cold: Average temperature of the warmest month > 10°C and that of coldest month < -3°C. This climate grows essentially the same crops as the temperate climate, but seasons tend to be shorter and limited at the beginning and end by frost.
	E	Polar: Average temperature of the warmest month < 10°C. No crops are grown under this climate.
Second letter: *Precipitation*	W	Desert
	S	Steppe
	f	Fully humid: precipitation of the driest month > 60mm
	s	Summer dry: precipitation of the driest month in summer is < 40mm and less than a third of the precipitation of the wettest month in winter

	w	Winter dry: precipitation in the driest month in winter is < .10 of the precipitation in the driest month in summer
	m	Monsoonal
Third letter: *Temperature*	h	Hot: mean annual temperature > 18°C
	k	Cold: mean annual temperature < 18°C
	a	Hot summer: average temperature of the hottest month ≥ 22°C
	b	Warm summer: not (a); number of months warmer than 10°C is ≥ 4
	c	Cool summer: not (a), (b), or (d); number of months warmer than 10°C is both ≥ 1 and < 4
	d	Very cold winter: not (a) or (b); average temperature of the coldest month is < -38°C
	F	Polar frost: temperature of the hottest month is ≤ 0°C
	T	Polar tundra: temperature of the hottest month is > 0°C

Multi-Error-Removed Improved-Terrain Digital Elevation Model (MERIT-DEM)

The Multi-Error-Removed Improved-Terrain Digital Elevation Model (MERIT-DEM), a high accuracy global DEM with open licensing at 3 arc-second resolution (~90 m at the equator), was used to create twelve elevation zones to compute statistics of population and land area. Although there are a variety of possible global digital elevation models (e.g., TanDEM-X, CoastalDEM, JAXA ALOS, SRTM, others), MERIT-DEM is one of the few that has open licensing while also producing high accuracy in statistical analysis (Hawker et al., 2019).

Global Human Settlement Layer Settlement Model (GHS-SMOD)

The GHS Settlement Model grid (GHS-SMOD R2019A), part of the Global Human Settlement Layer data set produced by the Joint Research Council of the European Commission, is the source of the urban and rural zones used to compute urban and rural population and land estimates for PLACE IV. SMOD uses the ‘degree of urbanisation’

(DEGURBA) method described by EUROSTAT (https://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Degree_of_urbanisation), to delineate and classify settlements based on population size, population density, built-up areas, and geographic contiguity. The data set is provided at Level 2 (L2) with eight classes that span the rural to urban continuum including a class for water. These L2 classes may be aggregated to urban and rural as per the data set documentation (Florczyk et al., 2019).

Geo Regions and Geo Subregions

Geo Regions are the geographical regions used by the United Nations Statistics Division (UNSD) in its publications and databases. They refer to the UN's macro geographical regions, and correspond as closely as possible to continents. Geo Subregions are more detailed geographical regions, which are grouped into the macro geographical units, called Geo Regions. Each country is shown in only one region and one subregion. From the UN: "The assignment of countries or areas to specific groupings is for statistical convenience and does not imply any assumption regarding political or other affiliation of countries or territories by the United Nations." Geo Regions and Subregions country classification data were acquired in the downloadable version of the UNSD "Standard Country or Area Codes for Statistical Use," originally published as Series M, No. 49 and now commonly referred to as the M49 standard (UNSD M49). PLACE IV includes the UNSD Geo Region and Subregion country classification to allow users to filter by regional geography.

Economic Classifications: Income Groups and Lending Categories

The World Bank classifies economies (i.e., member countries and territories) to facilitate the aggregation, grouping, and comparison of statistical data, and for the presentation of key statistics. PLACE IV includes the country classification by income groups and operational lending categories of the World Bank Group to allow users to filter by economic status.

For operational and analytical purposes, economies are divided among income groups according to Gross National Income (GNI) per capita, calculated using the World Bank Atlas method (World Bank). The classification is updated annually based on the previous year's GNI. For the period July 2019 to June 2020, income groups according to 2018 GNI per capita are: low income, $1,025 or less; lower middle income, $1,026 - $3,995; upper middle income, $3,996 - $12,375; and high income, $12,375 or more.

Economies are divided into IDA, IBRD, and Blend countries, so-called lending categories, based on the operational policies of the World Bank. International Development Association (IDA) countries are those with per capita incomes of $1,175 or less that lack the financial ability to borrow from the International Bank for Reconstruction and Development (IBRD). IDA credits are deeply concessional—interest-free loans and grants for programs aimed at boosting economic growth and improving

living conditions. IBRD loans are non-concessional. Blend countries are eligible for IDA credits because of their low per capita incomes but are also eligible for IBRD because they are financially creditworthy.

The economies country classification was downloaded as an Excel file from the World Bank website (World Bank, 2019).

Methods

Preprocessing of Input Data

All spatial data were projected to the WGS84 Geographic Coordinate System and preprocessed to match the extent and cell size of the 30 arc-second (approximately 1 km at the equator) resolution of the GPWv4.11 population, land area, and national identifier grid data sets used as the value and administrative zone inputs to PLACE IV. In cases where the extent of the data in the input data set fell short of the GPWv4.11 extent (i.e., the coastline), a workflow involving the Esri ArcGIS Nibble tool was used to fill in the gaps. The Nibble tool replaces the value of raster cells defined by a mask, with the value of their nearest neighbor. Cells in the input raster that are outside the mask area are not nibbled, i.e., the input values are retained in the output raster. Details of the methods used to preprocess each input data set follow.

Variables (Values)

The sources of the population count and land area estimates used as value inputs (see Table 1) to PLACE IV are all data sets from the GPWv4 Revision 11 data collection. All data sets in the collection are harmonized and therefore required no preprocessing.

Analysis Mask and Administrative Units (Zones)

1. *Analysis Mask*

   A binary analysis mask for all PLACE IV data pre-processing was created from the GPWv4.11 National Identifier Grid data set by setting all NoData cell values (NoData = -32768) to Null and all other values to 1. This mask was used to preprocess input data to new rasters that match the GPWv4.11 extent and cell size.

2. *Administrative Units (Level 0 and Level 1)*

   The nation-state representations in the GPWv4.11 National Identifier Grid serve as the level 0 administrative units for PLACE IV, and therefore no preprocessing was required. The vector administrative boundaries compiled for input to GPW Version 4, Revision 11 were used to create a raster of level 1 administrative units for use in PLACE IV.

3. *Urban and Rural Zones*

The GHS-SMOD data set is provided in World Mollweide projection at 1 km resolution with eight classes (i.e., L2) including a class for water (code 10) which includes both oceans and inland water areas. In order to prepare the urban and rural layer, the data set was projected to WGS84 at 30 arc-second resolution. Next, since oceans are not part of the PLACE framework, but inland water bodies are, the SMOD water class had to be disaggregated into ocean and inland water classes so that ocean areas could be excluded. Several steps were necessary in order to accomplish this.

a. A raster of land and water areas was produced by aggregating all SMOD classes other than water into a single land class.

b. The land and water raster was converted to a single part polygon geodatabase feature class.

c. The ocean feature was selected and converted to raster.

d. The ocean raster was then reclassified to a binary raster (oceans=1, land=0)

e. The binary raster and the original SMOD layer were used as input to the Esri ArcGIS Set NULL tool to set all ocean cells in the SMOD layer to Null.

With the ocean pixels now set to null, the eight L2 classes of the GHS-SMOD data set were aggregated as outlined in Florczyk et al. (2019) to produce two classes: Urban and Rural (Table 4). The new raster was Nibbled using the workflow presented above under Biomes.

Table 4. GHS-SMOD L2 re-classed to urban and rural.

L2 code	**L2 class**	**Urban/Rural**
30	Urban Centre	Urban
23	Dense Urban Cluster	
22	Semi-dense Urban Cluster	
21	Suburban or Peri-urban	
13	Rural Cluster	Rural

12	Low Density Rural	
11	Very Low Density Rural	
10	Water	

Biophysical Themes (Zones)

1. *Biomes*

The World Wildlife Fund (WWF) Terrestrial Ecoregions of the World (TEOW) data set was created to be used at the scale of 1:1 million, and is distributed in vector format. The 825 ecoregion polygons in the data set were dissolved to create an Esri geodatabase feature class of 14 biomes, plus zones of "Lake" and "Rock and Ice". The resulting feature class was converted to raster format at 30 arc-second resolution. The raster was then “Nibbled” to conform to the GPWv4.11 extent using the following workflow:

1. A conditional statement was used to replace all Null cells in the raster with an integer dummy value outside of the value range of the data set. This step was necessary because in the version of the tool used, NoData cells could not be nibbled.

2. The analysis mask was used with the Esri ArcGIS Extract Values by Mask tool to extract the values from the output raster from step 1 that fell within the mask extent, resulting in a new raster of the same extent and cell size as GPWv4.11. This raster is the input to Nibble.

3. In order to define the area of the output raster from step 2 that should be Nibbled, a raster mask was created by setting to Null all the cells with the dummy value in the output from step 2.

4. Finally, the Nibble tool was used to replace cells in the raster output from step 2 that fell within the Null mask created in step 3, replacing the dummy values with the value of the nearest neighbor while keeping all other data values.

2. *Climate Zones*

Köppen-Geiger Climate Classification maps were acquired from the Vienna Institute of Veterinary Public Health, as 0.5 degree ASCII grids in a Geographic projection representing 30 climate zones. Each of the ASCII files was ingested to create a point feature class, which was then converted to an Esri raster data set at 0.5 degree resolution. All rasters were Nibbled using the workflow presented above under Biomes.

3. Coastal Proximity Zones

The coastal proximity zones produced for use in PLACE IV are regions within a specified distance (in kilometers) of an ocean coastline. These zones were created by generating buffers around a linear coastline layer and then clipping the buffers to the coastline so that they included only land area. Inland water bodies were removed from the coastline layer prior to generating buffers since inland water bodies are not at risk of storm surge or sea level rise.

To create the buffers, the GPWv4.11 National Identifier Grid polygon layer was converted to a line feature class. The inland Caspian Sea was deleted from the feature class. The coastline layer was then projected to Mollweide. Four geodesic buffers were created based on 5, 10, 100, and 200 kilometers from the coastline and dissolved into multipart features.

A polygon feature class representing all non-ocean areas (i.e., land and inland water areas) was created from the analysis mask for use in clipping the buffers. The analysis mask which represents the GPWv4.11 land extent has a gap for the inland Caspian Sea. Therefore, the following preprocessing of the analysis mask was necessary to fill in the Caspian Sea and produce the clipping layer:

1. Esri ArcGIS IsNull was applied to assign all oceans and seas (i.e., areas of NoData) a value of 1, and all land areas a value of 0.
2. The resulting binary raster was converted to a polygon feature class and the land features were exported to a new feature class.
3. In the land feature class, the gap for the Caspian Sea was filled in with a polygon feature.
4. All land features were dissolved to a multipart polygon representing all land and inland water bodies.

The four buffers were then clipped to the ocean coastline using the land polygon feature class to produce the four coastal zone layers.

Zones (kilometers): within 5, within 10, within 100, within 200.

4. Elevation Zones

The Elevation zones were obtained by converting the 3 arc-second MERIT-DEM to 30 arc-second resolution and then reclassifying ranges of land elevation values into twelve thematic elevation zones.

Zones (meters): < 5, 5-10, 10-25, 25-50, 50-100, 100-200, 200-400, 400-800, 800-1500, 1500-3000, 3000-5000, > 5000

5. *Population Density Zones*

Each raster of the GPWv4.11 UN WPP-Adjusted Population Density data set was reclassified into twelve thematic zones based on ranges of population density values, resulting in rasters of population density zones for the years 2000, 2005, 2010, 2015, and 2020.

Zones (persons per square kilometer): 0, 0–2, 2–5, 5–10, 10–15, 15–50, 50–100, 100–500, 500–1000, 1000–10000, 10000–50000, > 50000

Geographic Regions and Economic Classifications

The UNSD geographic regions and World Bank economies classification data sets are both tabular datasets. These tables were matched to the GPW National Identifier Grid v4.11 polygon layer (CIESIN, 2018c) attribute table based on ISO3 code. Then the relevant attributes (Geo Region, Geo Subregion, Income Group, and Lending Category) were joined to the final PLACE IV table, also based on the ISO3 code.

Processing PLACE IV

Once all of the input data layers representing the biophysical themes were harmonized to match the GPWv4.11 extent and cell size and, if needed, reclassified into discrete rasters with the desired zones, the following workflow was applied using Esri ArcGIS geoprocessing tools accessed via custom Python scripts and the ArcPy site package:

1. The Extract By Attributes tool was used to create new rasters of each individual zone from every biophysical zone raster.
2. The Extract by Mask tool was used to extract population values from the population count rasters using the individual zone rasters as masks, resulting in unique rasters of population for each unique zone.
3. The Extract by Mask tool was used to extract land area estimates from the land area raster using the individual zone rasters as masks, resulting in unique rasters of land area for each unique zone.
4. The urban and rural zone rasters were extracted by attribute and combined with the administrative level 0 and administrative level 1 data in order to produce distinct urban and rural designations by country or level 1 unit.
5. Finally, the Zonal Statistics as Table tool was implemented considering the results from steps 2 and 3 as the *value* parameter, and the results from step 4 as the *zone* parameter in order to generate cross-tabulated estimates of urban, rural, and total land area and population within each administrative unit by each biophysical zone.

6. The results were compiled globally using python dictionaries and the ArcPy modules data access insert cursor functionality.

III. Data Set Description

Data set description:
The Population, Landscape, And Climate Estimates, Version 4 (PLACE IV) data set provides measures of population and land area for 248 statistical areas (countries and other territories recognized by the United Nations (UN)). Zonal statistics are computed to estimate the number of people (head counts) and the land area (square kilometers) as totals and by urban and rural designation, within multiple biophysical themes for each administrative unit. PLACE IV provides these country characteristics in tabular format.

Data set web page:
SEDAC URL: https://sedac.ciesin.columbia.edu/data/set/nagdc-population-landscape-climate-estimates-v4
Permanent URL: https://doi.org/10.7927/rht8-jv78

Data set format:
The data are available in Geodatabase (GDB) and Comma Separated Value (CSV[1]) formats. Each downloadable is a compressed zip file containing: 1) GDB or CSV file, 2) Readme.TXT file, and 3) data set codebook Excel file. The Readme and codebook files are also available on the Documentation web page.

Data set downloads:
- nagdc-population-landscape-climate-estimates-v4-country-xxx.zip
- nagdc-population-landscape-climate-estimates-v4-georegion-yyyyy-xxx.zip
- nagdc-population-landscape-climate-estimates-v4-geosubregion-zzzzz-xxx.zip
- nagdc-population-landscape-climate-estimates-v4-income-group-high-income-xxx.zip
- nagdc-population-landscape-climate-estimates-v4-income-group-low-income-xxx.zip
- nagdc-population-landscape-climate-estimates-v4-income-group-lower-middle-income-xxx.zip
- nagdc-population-landscape-climate-estimates-v4-income-group-not-classified-xxx.zip
- nagdc-population-landscape-climate-estimates-v4-income-group-upper-middle-income-xxx.zip
- nagdc-population-landscape-climate-estimates-v4-lending-category-blend-xxx.zip
- nagdc-population-landscape-climate-estimates-v4-lending-category-ibrd-xxx.zip

[1] Depending on your system configuration, special characters may not display correctly in MS Excel. To overcome this use Excel's data import functionality and set the code page to UTF-8.

- nagdc-population-landscape-climate-estimates-v4-lending-category-ida-xxx.zip
- nagdc-population-landscape-climate-estimates-v4-lending-category-not-classified-xxx.zip

where

- "xxx" is "gdb" or "csv"
- "country" is the 3-digit ISO code
- "yyyyy" is the georegion (i.e. africa, americas, asia, europe, oceania)
- "zzzzz" is the geosubregion (i.e. caribbean, eastern-asia, middle-africa, northern-america, south-america, western-europe, etc.)

IV. How to Use the Data

The increase in the volume of data in PLACE IV compared with PLACE III precludes the distribution of the data set as a single excel file (or even several files) with pre-established Pivot Table functionality, as was done under PLACE III. In order to distribute the PLACE IV data set in a manageable format, it is broken up into collections of tables by administrative designations: Country, Geo Region, Geo Subregion, Income Group and Lending Category. Each collection includes tables for each biophysical theme (biomes, climate zones, coastal proximity, elevation, and population density) and a table containing all the themes. Users may wish to import their desired table(s) into Microsoft Excel and create Pivot tables to filter and summarize the population and land area values by individual area designations (or combinations of administrative designations) and/or biophysical theme variables. Users who are new to Pivot Tables may wish to review the resources below or any number of resources an online search will return.

- Microsoft Support: Create a PivotTable to analyze worksheet data https://support.microsoft.com/en-us/office/create-a-pivottable-to-analyze-worksheet-data-a9a84538-bfe9-40a9-a8e9-f99134456576
- YouTube Video: Excel Pivot Tables EXPLAINED in 10 Minutes (Productivity tips included!) https://www.youtube.com/watch?v=UsdedFoTA68
- YouTube Video: Pivot Table Excel Tutorial https://www.youtube.com/watch?v=m0wI61ahfLc

V. Potential Use Cases

The goal of PLACE IV is to allow those who require tabular data on country characteristics to include these variables as elements in statistical analyses. This could include subsetting countries by certain characteristics, or controlling for factors such as percent of population living in semi-arid regions or in the coastal zone. Examples of uses of the earlier version of this data set, PLACE III, include health analyses by altitude

(Mehata et al., 2021), analyses of plastics pollution in relation to population in the coastal zone (Caldwell et al., 2020, Mancia et al., 2020), characterization of population exposure to sea level rise in small island states (Robinson, 2019), urban heat island studies focusing on those biomes with the largest cities (Chakraborthy et al., 2019), and a statistical analysis of the major factors contributing to economic growth (Martín-Retortillo et al., 2019). These same applications and more are enabled by PLACE IV.

VI. Limitations

As with any effort to integrate data developed at different scales and with varying levels of accuracy, there are limits to the precision and accuracy of the PLACE IV data set. A partial list of known or potential issues include:

- The census data used to produce the GPWv4.11 population estimates are from the 2010 round of censuses (2005-2014), owing to the fact that 2020 round census data are still not widely available.
- There are inherent uncertainties in the spatial representation of population in GPW owing to the varying size of census reporting units (SDSN 2020).
- There are spatial generalizations and mismatches between many of the input data sets. Most of the spatial mismatches were addressed through spatial processing such as the Nibble workflow described above, but this results in an approximation as knowledge of the actual spatial distribution (ground truth) of the variables of interest is lacking.
- Climate zones data are based on obsolete IPCC SRES scenarios. The only available climate zone projections for the new IPCC scenarios based on Representative Concentration Pathways (RCPs) have not yet passed peer review.

VII. Acknowledgments

PLACE IV was completed by CIESIN using spatial data inputs produced by CIESIN and other parties. The PLACE IV team included Geographic Information Specialists, Tricia Chai-Onn and Linda Pistolesi, and Senior Systems Analyst/GIS Programmer, Kytt MacManus. Chai-Onn and Pistolesi were responsible for data acquisition and preprocessing. MacManus authored the Python script to automate the computation of zonal statistics and production of the final data products. John Scialdone coordinated the data set release and reviewed the data set documentation. Merlie Hansen compiled and formatted the data set metadata. Joe Schumacher prepared the data set and documentation delivery mechanism on the SEDAC website.

Funding for development and dissemination of this data set was provided under the U.S. National Aeronautics and Space Administration (NASA) contract 80GSFC18C0111 for

the continued operation of the Socioeconomic Data and Applications Center (SEDAC), which is operated by the Center for International Earth Science Information Network (CIESIN) of Columbia University.

VIII. Disclaimer

CIESIN follows procedures designed to ensure that data disseminated by CIESIN are of reasonable quality. If, despite these procedures, users encounter apparent errors or misstatements in the data, they should contact SEDAC User Services at ciesin.info@ciesin.columbia.edu. Neither CIESIN nor NASA verifies or guarantees the accuracy, reliability, or completeness of any data provided. CIESIN provides this data without warranty of any kind whatsoever, either expressed or implied. CIESIN shall not be liable for incidental, consequential, or special damages arising out of the use of any data provided by CIESIN.

IX. Use Constraints

X. Recommended Citation

Center for International Earth Science Information Network (CIESIN), Columbia University. 2022. National Aggregates of Geospatial Data Collection: Population, Landscape, And Climate Estimates, Version 4 (PLACE IV). Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/rht8-jv78. Accessed DAY MONTH YEAR.

XI. Source Code

The geoprocessing tools leveraged for this work rely on the proprietary ArcPy python module from Esri and are described in the sections above. The processing algorithm relies on subsetting, producing zonal statistics, and summarizing. These are well known processes and we therefore do not distribute source code but instead direct users to Esri documentation for the ArcPy module (https://pro.arcgis.com/en/pro-app/latest/arcpy/get-started/what-is-arcpy-.htm).

XII. References

Caldwell, J., L. F. Muff, C. K. Pham, A. Petri-Fink, B. Rothen-Rutishauser, and R. Lehner. 2020. Spatial and temporal analysis of meso- and microplastic pollution in the Ligurian and Tyrrhenian Seas. *Marine Pollution Bulletin*, 159: 111515. https://doi.org/10.1016/j.marpolbul.2020.111515.

Center for International Earth Science Information Network (CIESIN), Columbia University. 2018a. Gridded Population of the World, Version 4 (GPWv4): Basic Demographic Characteristics, Revision 11. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H46M34XX.

Center for International Earth Science Information Network (CIESIN), Columbia University. 2018b. Gridded Population of the World, Version 4 (GPWv4): Land and Water Area, Revision 11. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H4Z60M4Z.

Center for International Earth Science Information Network (CIESIN), Columbia University. 2018c. Gridded Population of the World, Version 4 (GPWv4): National Identifier Grid, Revision 11. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H4TD9VDP.

Center for International Earth Science Information Network (CIESIN), Columbia University. 2018d. Gridded Population of the World, Version 4 (GPWv4): Population Count Adjusted to Match 2015 Revision of UN WPP Country Totals, Revision 11. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H4PN93PB.

Center for International Earth Science Information Network (CIESIN), Columbia University. 2018e. Gridded Population of the World, Version 4 (GPWv4): Population Density Adjusted to Match 2015 Revision UN WPP Country Totals, Revision 11. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H4F47M65.

Center for International Earth Science Information Network (CIESIN), Columbia University. 2018f. Documentation for the Gridded Population of the World, Version 4 (GPWv4), Revision 11 Data Sets. Palisades NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H45Q4T5F.

Chakraborty, T., A. Hsu, D. Manya, and G. Sheriff. 2019. Disproportionately higher exposure to urban heat in lower-income neighborhoods: a multi-city perspective. *Environmental Research Letters*, 14(10): 105003. https://doi.org/10.1088/1748-9326/ab3b99.

FAO Sustainable Development Department (SD), Agrometeorology Group. 2006a. *Brief Guide to Köppen Climate Classification System*. https://www.fao.org/nr/climpag/climate/koeppen_en.htm. Accessed 10 September 2021 from FAO 2006b.

FAO Sustainable Development Department (SD), Agrometeorology Group. 2006b. *Global Climate Maps: Tour Guide*. https://www.fao.org/nr/climpag/climate/index_en.asp.

Florczyk A. J., C. Corbane, D. Ehrlich, S. Freire, T. Kemper, L. Maffenini. M. Melchiorri, M. Pesaresi, P. Politis, M. Schiavina, F. Sabo, and L. Zanchetta. 2019. *GHSL Data Package 2019*. EUR 29788 EN, Publications Office of the European Union, Luxembourg, ISBN 978-92-76-13186-1, JRC 117104. https://doi.org/10.2760/290498. https://publications.jrc.ec.europa.eu/repository/handle/JRC117104.

Fuchs, T. 2008. *GPCC's Full Data Reanalysis Version 4 for 1901-2007*. Personal Communication, Data available at https://gpcc.dwd.de.

Hawker, L., J. Neal, and P. Bates. 2019. Accuracy assessment of the TanDEM-X 90 Digital Elevation Model for selected floodplain sites, *Remote Sensing of Environment*, 232: 111319, https://doi.org/10.1016/j.rse.2019.111319.

Mancia, A., T. Chenet, G. Bono, M. L. Geraci, C. Vaccaro, C. Munari, M. Mistri, A. Cavazzini, and L. Pasti. 2020. Adverse effects of plastic ingestion on the Mediterranean small-spotted catshark (*Scyliorhinus canicula*). *Marine Environmental Research*, 155: 104876. https://doi.org/10.1016/j.marenvres.2020.104876.

Martín-Retortillo, M., and V. Pinilla. 2019. The Fundamental Causes of Economic Growth: A Comparative Analysis of the Total Factor Productivity Growth of European Agriculture, 1950-2005. *Documentos de Trabajo (DT-AEHE)*, 1912: 28.

Mehata, S., N. Shrestha, S. Ghimire, E. Atkins, D. K. Karki, and S. R. Mishra. 2021. Association of altitude and urbanization with hypertension and obesity: analysis of the Nepal Demographic and Health Survey 2016. *International Health*, 13(2): 151-160. https://doi.org/10.1093/inthealth/ihaa034.

Mitchell, T. D., and P. D. Jones. 2005. An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *International Journal of Climatology*, 25: 693-712. https://doi.org/10.1002/joc.1181.

Mitchell, T. D., T. R. Carter, P. D. Jones, M. Hulme, and M. New. 2004. *A Comprehensive Set of High-resolution Grids of Monthly Climate for Europe and the Globe: The Observed Records (1901–2000) and 16 Scenarios (2001–2100)*. Tyndall Centre of Climate Change Research, Working Paper 55.

Nakicenovic, N., J. Alcamo, A. Grubler, K. Riahi, R. A. Roehrl, H. –H. Rogner, and N. Victor. 2000. *Special Report on Emissions Scenarios (SRES), A Special Report of Working Group III of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, U.K., 599 pp. http://www.grida.no/climate/ipcc/emission/index.htm.

Olson, D. M., E. Dinerstein, E. D. Wikramanayake, N. D. Burgess, G. V. N. Powell, E. C. Underwood, J. A. D'amico, I. Itoua, H. E. Strand, J. C. Morrison, C. J. Loucks, T. F. Allnutt, T. H. Ricketts, Y. Kura, J. F. Lamoreux, W. W. Wettengel, P. Hedao, and K. R. Kassem. 2001. Terrestrial Ecoregions of the World: A New Map of Life on Earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity, *BioScience*, 51(11): 933–938. https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2.

Pesaresi, M., A. Florczyk, M. Schiavina, M. Melchiorri, and L. Maffenini. 2019. *GHS Settlement Grid, Updated and Refined REGIO Model 2014 in Application to GHS-BUILT R2018A and GHS-POP R2019A, Multitemporal (1975-1990-2000-2015), R2019A*. European Commission, Joint Research Centre (JRC) [Dataset]. https://doi.org/10.2905/42E8BE89-54FF-464E-BE7B-BF9E64DA5218. http://data.europa.eu/89h/42e8be89-54ff-464e-be7b-bf9e64da5218.

Robinson, S. -A. 2019. A commentary on national adaptation drivers: the case of small island developing states. *Climatic Change,* 154(3-4): 303-313. https://doi.org/10.1007/s10584-019-02421-w.

Rubel, F., and M. Kottek. 2010. Observed and projected climate shifts 1901-2100 depicted by world maps of the Köppen-Geiger climate classification. *Meteorologische Zeitschrift*, 19: 135-141. https://doi.org/10.1127/0941-2948/2010/0430. http://koeppen-geiger.vu-wien.ac.at/shifts.htm.

Sustainable Development Solutions Network (SDSN). 2020. *Leaving No One Off The Map: A Guide For Gridded Population Data For Sustainable Development*. New York: SDSN. https://www.popgrid.org/sites/default/files/documents/Leaving_no_one_off_the_map.pdf

UNSD M49, United Nations Statistics Division (UNSD*), Standard Country or Area Codes for Statistical Use (M49)*. Accessed 20 February 2020 from https://unstats.un.org/unsd/methodology/m49/. (Direct data set url: https://unstats.un.org/unsd/methodology/m49/overview/).
World Bank. *What is the World Bank Atlas Method?* https://datahelpdesk.worldbank.org/knowledgebase/articles/77933-what-is-the-world-bank-atlas-method. Accessed 20 February 2020.

World Bank. 2019. *World Bank Country and Lending Groups*. https://datahelpdesk.worldbank.org/knowledgebase/articles/906519, updated June 2019. Accessed 20 February 2020.

Yamazaki D., D. Ikeshima, R. Tawatari, T. Yamaguchi, F. O'Loughlin, J. C. Neal, C. C. Sampson, S. Kanae, and P. D. Bates. 2017. A high accuracy map of global terrain elevations. *Geophysical Research Letters*, 44: 5844-5853. https://doi.org/10.1002/2017GL072874.

XIII. Documentation Copyright and License

Appendix 1. Data Revision History

No revisions have been made to this data set.

Appendix 2. Contributing Authors & Documentation Revision History

Revision Date	ORCID	Contributors	Revisions
May 19, 2022	0000-0002-2980-4307 0000-0002-6416-1837 0000-0001-9484-1705 0000-0002-8875-4864	Tricia Chai-Onn, Kytt MacManus, Linda Pistolesi, Alex de Sherbinin	This document is the 1st instance of documentation.