Esri's 2016 World Population Estimate Methodology

Author: Charlie Frye

An Esri® White Paper

December 2017

Esri's World Population Estimate Methodology

Contents

I. Introduction

Since 2014, Esri has published a collection of global image services that represent the global dasymetric distribution of human population. This collection is called the World Population Estimate (WPE), and Esri's goal is to release annual estimates. The WPE image services are based on raster datasets, where the cells represent estimates of population count, population density, likelihood of settlement, and confidence scores.

Esri made improvements to the WPE model, and input data with each release, and the 2016 model represents what is expected to be stable in terms of the process and requirements for input data going forward.

Table 1: WPE Release History			
Data Represent	2013	2015	2016
Esri Released	Dec 2014	July 2016	Nov 2017
Population	✔	✔	✔
Population Density	n/a	✔	✔
Confidence	n/a	✔	✔
Settlement Likelihood	n/a	✔	✔
Cell Size*	230	162	150

*The cell size is representing the size of a cell at the equator, and the farther a cell is from the equator, the smaller the area of the earth it represents.

Figures 1 and 2 illustrate the coverage and resolution of the services. The WPE is a dasymetric surface, meaning the raster cells on land where people are not estimated to live are set to "No Data", while cells where people likely live are assigned an estimated population count.
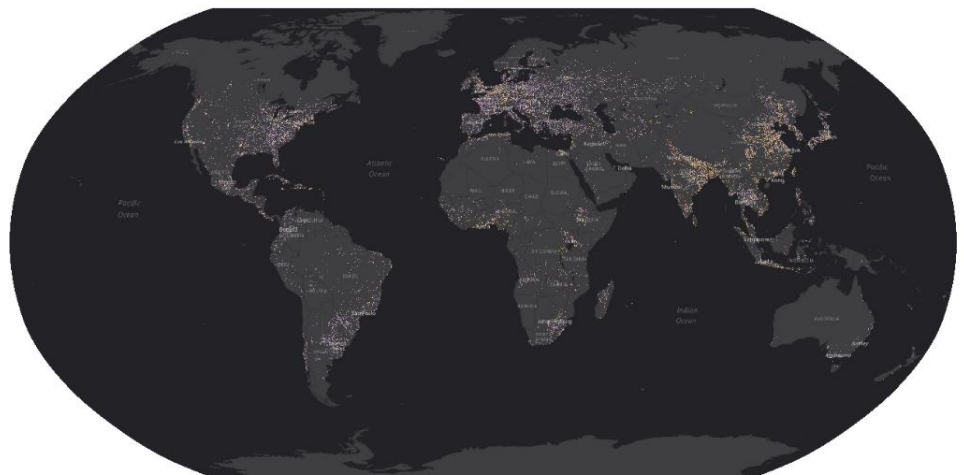
Figure 1. WPE global coverage

Figure 2. Part of Italy, featuring Rome (left) and Naples (lower right) as an example of WPE 2016 Population Density's available level of detail.

This white paper is intended to introduce the nature and purpose of the WPE services, provide recommendations to successfully use the services, and to present the methodology used to produce the WPE.

A. Overview

The model Esri uses to produce the WPE first estimates the footprint of human activity and assigns a settlement likelihood score everywhere within the footprint. Then population data in the form of census polygons is apportioned or distributed onto the footprint. This population data is from the most recently available census tabulation and population estimate geographies. The settlement likelihood footprint and population estimate data in the WPE are delivered as raster surfaces, which have been characterized by a broader community as "gridded population" datasets.

The WPE is designed to represent where people live, or "night time" population, rather than a specific time of day or "ambient" population. This corresponds to the idea of a "census de jour", which enumerates where individuals usually reside, regardless of where they are located on the day the census is taken. Most countries do not produce a de jour census, and instead produce de facto estimates of the count of people present at the time of the estimate. However, by first modeling the footprint of where people live, then re-distributing the population from these estimates, Esri is

spatially refining these estimates to locations where this population is most likely to live.

The footprint of settlement score is modeled by starting with classified land cover data. Esri reclassifies this into three classes:

1. Uninhabited Areas (base score = 0) where people generally cannot live
2. High density urban areas (base score of 200)
3. Medium density urban areas (base score of 150)
4. Agricultural areas (base score = 25)

Esri adds ancillary data to augment the base scores:

5. Road intersections (base score = 75)
6. Populated Place points from GeoNames (based score = 50)

The maximum of the values in areas 4, 5, and 6 are calculated everywhere. The model then looks for high levels of texture in Landsat8 panchromatic imagery within areas 2, 3, and 4. The texture values are converted into a texture score with a range of -20 to 272, which is added to the base score (negative results are set to zero). The result for each cell in the raster dataset is a settlement likelihood score. In 2016, it ranged from 0 to 472.

Then population data, in the form of enumeration unit polygons, are apportioned onto this footprint. The population data is sourced from censuses of various countries, and from commercial and United Nations data with population estimates as surrogates for censuses. The complete list of sources is included as an appendix to this paper.

The resolution of the WPE settlement score raster layer is 75-meters cell size, which approximately corresponds to a city block. For population and population density, the cell size is 150-meters, which is about four city blocks. The population data was redistributed, or apportioned, based on settlement score, and then aggregated by a factor of two to produce the population raster dataset. The population density raster dataset was derived from the population dataset.

The settlement score raster is not perfect. The process of modeling texture from Landsat8 imagery effectively identifies locations where the landscape has been disturbed, and at least conceptually, the idea is the higher the score, the more likely people are living in a location. However, the model

3

cannot determine which structures people are living within. Imagine a rural setting where a farm with six buildings is located. One of those buildings may be a house where people do live, it is also possible nobody lives there because the farm is either abandoned or owned by a corporation (and the farmer lives in a nearby town). There are also cases where Landsat8 imagery is too coarse, and does not contain levels of texture that correspond to closely spaced small dwelling units. The normal case scenario is a structure that is roughly at least 15x15-meters and casts a shadow into an adjoining cell that is not a structure.

Esri mitigates false-positives and false-negatives by augmenting settlement likelihood with proximity to a road intersection or GeoNames Populated place points. In particular, high quality census tabulation unit polygons filter out most false-positive settlement score cells. For instance, people do not live in airports, hotels, schools, and industrial complexes. Thus, highly precise census enumeration data, which excludes these places is highly desirable. However, in many places, such data has never been produced. Thus, the WPE is only as good as the model input data, and individual cell values of settlement likelihood or population may have a very high margin of error (MOE).

B. Esri's Purpose for Producing the WPE

The WPE services are designed to support business GIS, demographic analysis, and consumer economic analysis by providing a globally consistent gridded population dataset, and a corresponding settlement score dataset. The settlement score dataset is the basis to redistribute population, or any other statistics from census or economic activity enumeration polygons. The fundamental concept is that where people live, is where they spend their money, ergo where the WPE estimates people to live is equivalent to estimating where consumer spending originates.

C. What analyses should the WPE data be used for?

Given the purpose of supporting consumer economic analysis, Esri intends the WPE can be used in ArcGIS to undertake a wide variety of tasks. These include using the WPE data as the basis for weighted estimates and summary statistics of user-supplied demographic or economic characteristics, in any locale or area in the world, in a comparable way. Esri uses the WPE's settlement score layer as the starting point for this task in its data enrichment tools in Business Analyst, ArcGIS Online, and Insights for ArcGIS. Esri uses the WPE for creating such estimates in countries where Esri's distributors or business partners have not provided a better basis to redistribute demographic or economic characteristics.

There are additional tasks the WPE can be applied to:

- Estimate where political attitudes are rooted by estimating the distribution of party affiliations, voting frequency, poll responses by postal code, and survey responses.
- Estimate where cultural characteristics occur based on surveys of cultural topics or participation in cultural activities.
- Estimate where social values are engrained by examining the distribution of surveys and polls, including those created in Survey123 and crowdsourcing Story Maps.
- Estimate human impacts on the environment for purposes of sustainability, resilience, land management, and planning. Note that economics may only be one of several factors, and where people live is not where they work or spend significant amounts of time that may have impacts on their environment. This includes using population as a surrogate for the sources of some types of pollution, including air pollution from automobiles or byproducts of product-specific consumer spending such as batteries, high tech devices, etc.
- Estimate location of populations affected by:
  - Natural Disasters and Complex Humanitarian Emergencies (for mitigation, response, & recovery)
  - Exposure to disease (to prevent transmission)
  - Disease outbreak (to direct effective treatment efforts)
  - Access to infrastructure

  Note that the model that produces the WPE was not designed to count people, but estimate where people live. The community of people who work in these areas need data that very specifically locates people, versus data that indicates a likelihood of residence-based economic activity.

D. Quality Checks used to Certify the WPE

To understand the extent that WPE services can inform these tasks, it is useful to describe what Esri does in certifying the WPE datasets for release as web services. Esri uses several criteria to ensure the highest possible quality exists. Those are as follows:

i. The total number of people

The raster dataset that is the basis for a given year's estimate must be reasonably close to the most recent U.N. Statistical Division's estimate.

To check the population represented by a raster dataset, Esri adds a field to the attribute table and calculated it to be the value times the cell count, and then the sum of values in the new field is equal to the total population.

ii. The proportion of people in major density categories

Esri does a series of zonal statistics where population is summed within one person (integer) intervals of population density. From this Esri can determine what percentage of people have been modeled as living in urban densities versus rural densities. Because definitions for urban vary, and often include additional criteria, such as presence of services, Esri's testing is, at best, approximate.

At the upper end of these various definitions for urban density, Esri tested 2,500 persons per square kilometer, which produced 55.47% of people lived at urban densities in 2013, 76.2% of people lived at urban densities in 2015, and 53.56% of people lived at urban densities in 2016. This compares well to the United Nations estimate of 54% and the World Bank's estimate of 53.9% of people living within urban areas in 2015.

On the lower end of these definitions for urban density is 500 persons per square kilometer, which is based on the United States Census Bureau's definition, which also presumes a set of services typically found in highly developed population centers to exist. At this level, Esri's 2013 estimate had 72.96%, in 2015 it was 94.8%, and in 2016 it was 83.9%. The expectation here is that it is commonly reported that approximately 80% of people live in urban settings.

The differences between these estimates is due to three factors. First is a significant evolution in the settlement score model between 2013 and 2015. Second, from 2015 to 2016 the quality of the input data was greatly improved. Third, the production process for the 2016 estimate shifted the processing away from 0.5 degree tiles, which caused the introduction of statistical artifacts, to a workflow using countries as processing units.

iii. Analysis of Input Datasets

The model Esri uses to produce the WPE services uses the input datasets in novel ways and often shows their strengths and weaknesses. The quality of these datasets varies geographically because they are independent, in purpose, topic, and measuring or recording techniques. In some locales, the combination of these datasets works to strengthen the resulting settlement score surface, and in other places, an error in one source, usually of omission, significantly and detrimentally impacts the quality for that location.

The cell size for the WPE rasters is 150-meters, and that means there are billions of locales. Esri's model for producing the WPE settlement score fully depends on the quality of the input datasets. Each has issues that can affect the quality, and Esri has isolated the most significant of these and compiled a confidence score dataset. However, there are issues that cannot be accounted for systematically, including the following:

- Errors in classified land cover: Most classified landcover raster products make claims of accuracy along the lines of 60% of the cells are correctly classified and up to 85% are classified correctly within one class. The latter meaning grassland might be misclassified as cropland, or vice versa. Therefore, it is reasonable to expect that 15% of the cells are misclassified, and Esri estimates that as many as one third of those, or 5% overall detrimentally affect the WPE model.
- Errors in road intersection data: Missing road intersections or old roads that correspond to abandoned settlements will cause the WPE model to respectively under- or over-estimate population for a locale.
- Age of Census Data: Old census data allocates people where they do not belong when roads in abandoned settlements are used; and new census data can be forced to over-allocate people when new roads are not included in the sources Esri uses for roads.
- Quality of location and currency of GeoNames settlement points: As with roads, GeoNames includes populated place points that may not be updated to indicate a settlement has been abandoned.

II.   Recommended Steps to Get Started

The WPE image services are designed to be used online for mapping and visualization. These services are also designed to support analysis (geoprocessing and raster functions) in ArcGIS Desktop applications. Mapping and visualization tasks include drawing the WPE layers with default colors or changing the colors to use any of the supported methods for symbolizing raster data. The most common use case is to use the population density layer in lieu of a population density choropleth layer. This use is cartographically similar to a dot density map. The population density layer focuses map reader's attention on the locations where people live in the greatest density. This pattern of mapping enhances the ability of many of the basemaps in ArcGIS to contextualize the settings for human population.

A. General Considerations for Raster Analysis

The keys to successful analysis with the WPE image services are first expecting that image services can be used just like any other raster dataset as the source for a raster layer, and second, that population data values are uniquely affected by raster geoprocessing tools.

Image services are created from mosaic datasets, which in turn are created from one or more raster data sources. Therefore, image service layers in ArcGIS are very much like raster layers that are based on mosaic datasets, in that they ultimately provide access to raster data, no matter where the data are stored.

In any analysis of raster data, there are important characteristics of the raster datasets to know and manage. These include the following:

- Coordinate system: The 2013 and 2015 WPE services were produced using WGS_1984, but published as image services using Web Mercator Auxiliary Sphere. This kept these layers from being used for analysis because projecting from WGS_1984 to Web Mercator changed the cell values. This is because the projection process required image resampling to be applied. Thus, the 2013 and 2015 services should not be used in analysis workflows in ArcGIS Desktop and ArcGIS Pro. ***The 2016 WPE services are set to use WGS_1984 which is the same coordinate system as the underlying raster datasets. Thus, there is no error introduced when creating raster layers from these services.***
- Cell size (a.k.a. resolution):
    - Population: 0.0013474728 deg. (~150-m) at equator.
    - Density: 0.0013474728 deg. (~150-m) at equator.
    - Settlement: 0.00067373639 deg. (~75-m) at equator.
    - Confidence: 0.0038112113 deg. (~450-m) at equator.
- Origin (coordinate) of the raster dataset's extent: If the origins of one or more rasters do not match, or the cell sizes do not have an integer common denominator, one will need to be chosen as the snap grid, which will be specified in the geoprocessing environment settings. It may also be necessary to aggregate or resample one or more of the rasters, or shift the cells using the raster Clip tool, but specifying a new shifted extent.
- Bit depth (8-bit, 16-bit, floating point, etc.): Determines the ultimate potential value range of the numbers in the raster's value field. Not managing this can result in inadvertently re-scaling values or truncating the value range.

- Statistic type of the values in the Value field (nominal, ordinal, interval, or ratio): Some geoprocessing tools, such as the Project Raster, and Resample tools, need to redistribute the cells, and the chosen resampling will impact the statistics of the output raster. There are several resampling methods, including nearest neighbor, bilinear, cubic convolution, and majority. ***Only nearest neighbor resampling is recommended for use on the WPE services***.
- Continuous surface, versus level of discreteness: The more discrete the footprint of a raster dataset's data (versus NoData) cells, the more likely resampling (above) will affect the processing of this raster.
- No Data value: Mainly used for cartographic purposes to allow NoData cells to be not drawn (fully transparent), or represented with a specific neutral color.
- Format (.tif, Geodatabase raster, etc.): Certain formats such as JPG represent the possibility that the raster data is not original, or has been converted to RGB rather than containing the original data values.
- Whether pyramids (raster data) overviews (mosaic datasets and image services) exist: When doing analysis, it is vital not to include overview layers, as these data usually have uncontrolled statistical skews. Set definition queries to exclude overview rasters from layers based on mosaic datasets or image services.

B. Common Geoprocessing Tasks

Depending on the type of analysis you want to do, it may or may not be necessary to project the WPE data. Here are recommendations for coordinate systems to use:

i. Apportion more detailed population data using the settlement score raster.

Project your population polygon data to WGS_1984, to match the WPE Settlement Score raster, before proceeding with apportionment (Esri's method is described later in this paper).

ii. Zonal (Statistics) or Aggregation

No change in coordinate system. Ideally project the zones (polygon data) or raster to WGS_1984 and snap to the WPE layer. The following zonal analyses or aggregation methods are common:

- Confidence: Mean
- Settlement Score: Sum or Mean
- Population: Sum

- Population Density: Mean

iii. Proximity

Project to Equidistant Cylindrical. Projecting from WGS_1984 to Equidistant Cylindrical causes negligible loss in total population, ranging from 0.007% to 0.15% per cell, depending on distance from the equator.

iv. Area

Project to Mollweide. Projecting from WGS_1984 to Mollweide causes some loss in total population, ranging from 0.78% to 30% per cell, depending on the distance from the equator. Unfortunately, this is the best option for projecting raster.

However, recent advances in hardware, particularly SSDs (Solid State Drives) allow for conversion of rasters to polygons. As polygons, an attribute for geodesic area can be calculated using the Add Geometry Attributes tool. This should be considered the preferred option, if accuracy of area calculation is critical to the integrity of the analysis.

C. Raster Functions and Raster Function Chains

When the WPE image services are added to ArcGIS Pro, as Image Service Layers, they can be used in raster function chains, where raster functions can be applied to modify the output values of the layer. These function chains can be saved and shared as part of raster analytics workflows.

D. About the colors used to portray the WPE services

The 2015 WPE data are presented using the following legend:
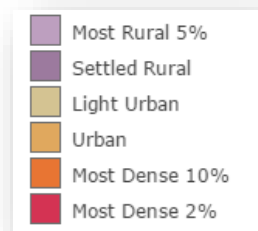
Figure 3. World Population Estimated Density Layer's Legend

The cut points for these population density classes are:

- Most Rural 5%: 100 persons/sqkm or less
- Settled Rural: 400 persons/sqkm
- Light Urban: 1,500 persons/sqkm
- Urban: 2,500 persons/sqkm
- Most Dense 10%: 16,979 persons/sqkm

- Most Dense 2%: 29,000 persons/sqkm or more

Figure 4 illustrates another way to visualize the population and population density using these colors, where the distribution of the world's population is displayed, radiating outward from the center starting at 1 person per square kilometer and turning clockwise toward the highest densities. Note the threshold of 200 for rural is shown in lieu of the most rural 5% threshold used in the web services.
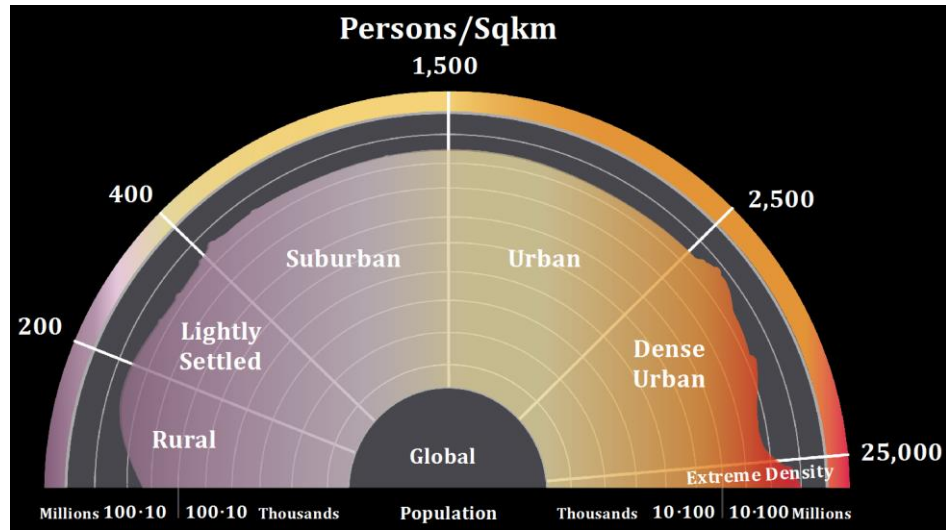
Figure 4. World's population by estimated density graph.

E. General Considerations for working with Gridded Population

The following topics may be relevant when working with the WPE layers.

i. Resampling due to Projection and Aggregation

Raster analysis in ArcGIS with dasymetric (highly discrete) raster population data requires some special considerations. First and foremost is whether WPE rasters will need to be projected. Several common analysis scenarios were presented earlier (II.B.), and some involved projecting the WPE rasters. Projection and aggregation of raster datasets is fundamentally a form of modifiable areal unit problem (MAUP), meaning that the summary statistics of the result will differ from the input.

The root cause of the MAUP is the potential for redistribution, elimination, or duplication of values as part of projection or aggregation algorithms. The reason these changes occur is because both processes change the number of rows and columns, and potentially the footprint of data versus NoData cells in the resulting raster. ArcGIS uses a process called resampling to manage the impact these changes. Resampling almost always occurs when

projecting raster data (there are only a few cases where the input and output coordinate systems have exactly the same cell counts and aspect ratio). Two factors exacerbate the degree to which resampling can mitigate the impacts of projection or aggregation. The first is how radical the shape change is between the footprint of the input and the footprint of the output raster; the greater the change the higher the risk to the raster values. Second is the complexity of the footprint of the data values versus the NoData values in discrete (dasymetric) rasters; the more complex the shape of the footprint the less likely resampling can manage the values during the transformation. In Table 2, the color of the text represents the risk to the integrity and quality of the data values when projected or aggregated. Green text is low risk, orange is moderate risk, and red is not just high risk, but also high impact to the values.

Table 2: Impact of Resampling on Raster Data				
	Nominal	Ordinal	Interval	Ratio
Continuous Raster Surface	Example: Land cover. Usually insignificant redistribution of histogram.	Example: Hazard rating. Usually insignificant redistribution of histogram.	Example: Precipitation (mm). **Significant changes in summary statistics possible**	Example: Elevation. **Possible loss or duplication of extreme/outlier values**
Discrete/ Dasymetric Raster Surface	Example: Habitat. Usually insignificant change in cell count. S**ingle isolated cells may disappear.**	Example: Population density classes (rural vs. urban). Usually insignificant change in cell count. **Single isolated cells may disappear.**	Example: Population counts. **Significant changes in summary statistics are highly likely. Single isolated cells may disappear, and others may be duplicated.**	Example: Population Density. **Possible loss or duplication of extreme/outlier values. Single isolated cells may disappear.**

ii. Expectations for Conformance to Real-World Population

Because the WPE datasets represent models, based on several independently compiled inputs, the likelihood of a given cell being assigned a population that matches reality for that location is low. It should be expected that the population assigned to a given cell is close to the actual value, though the margin of error for a single raster cell's population will be quite high. The margin of error for the sum of one hundred contiguous cells will go down dramatically, and for several thousand contiguous cells the margin of error should effectively be minimized at around +/- 4.0%.

The figure of 4.0% is an estimate based on two factors. First is the amount of normal turnover (in migration and out migration) within a place and the rate of natural increase (RNI, births minus deaths). Second is the age of

census data combined with the age of the model input datasets. If all inputs and data were current and concurrent, an even lower margin of error could be achieved.

iii. Error, Uncertainty, and Confidence

The WPE services may be used for a variety of spatial analyses, the most common of which would be to provide an estimate of the number of people living inside a given region or within a distance of a given location such as an event, an earthquake epicenter, or along a given segment of coastline. However, qualifying that estimate, in terms of how accurate it might be, given the input data to the WPE model, is not simple.

For instance, the local variation and accuracy of all the input datasets, would need to be estimated, along with the potential model's processes to introduce artifacts. Realistically, not all of these factors can be fully accounted for. However, Esri has produced a confidence surface that accounts for most of them. The 2015 and 2016 versions of the WPE have a confidence level service, which can be found online.

This service represents confidence on the basis of a one to five rating, where one is the least confident and five the most confident. This can be used to either find the average level of confidence for the footprint by taking the mean confidence score for a region, or deriving a function that produces a margin of error for the population estimate of each cell within an area. Note that the larger the region analyzed, the more accurate the estimate is likely to be, as the region may eventually include entire census reporting units, which would represent the highest level of confidence possible. Thus, this discussion is mostly about how much confidence to have in the areas of the world where excellent census data does not exist, and when the population of a relatively small area is being estimated.

iv. Population Count Values

In the 2015 WPE population layer, the values estimate the count of people living within the area represented by that cell. The values range from 1 to being capped at 32,767 in the 162-meter resolution data. Generally, the cells with values of 4 or less are considered very unreliable, and cells with values over 2,500 is similarly unreliable, because people do not live at such high densities (equivalent to a density of over 35,000 persons per square kilometer).

The process of allocating people to sparsely populated areas arbitrarily must choose from among many locations to allocate a few people. Allocating people in densely populated areas that also have complex edges due to

13

parks, airports, coastlines, etc. will often force too many people into a given cell.

As the estimated population for a given cell becomes higher, there is generally higher reliability. However, at the very top of the range, there are also unreliable cells. See the upper tail in figures 5 and 6. This is due to spatial (horizontal) inaccuracies between the polygons reporting, usually urban populations, which can be displaced by a few meters, and in some cases, up to a kilometer. The result is a high number of people allocated to a small amount of footprint. For instance, the upper value of 32,767 persons in a 162m cell, implies a population density that physically cannot exist. Esri has been working to progressively eliminate these issues. Conventional wisdom dictates any value above 35,000 persons per square kilometer square is unreliable unless a very tall building is located within that square. As with the very low values, the total number of people within two kilometers of this location is likely quite accurate.
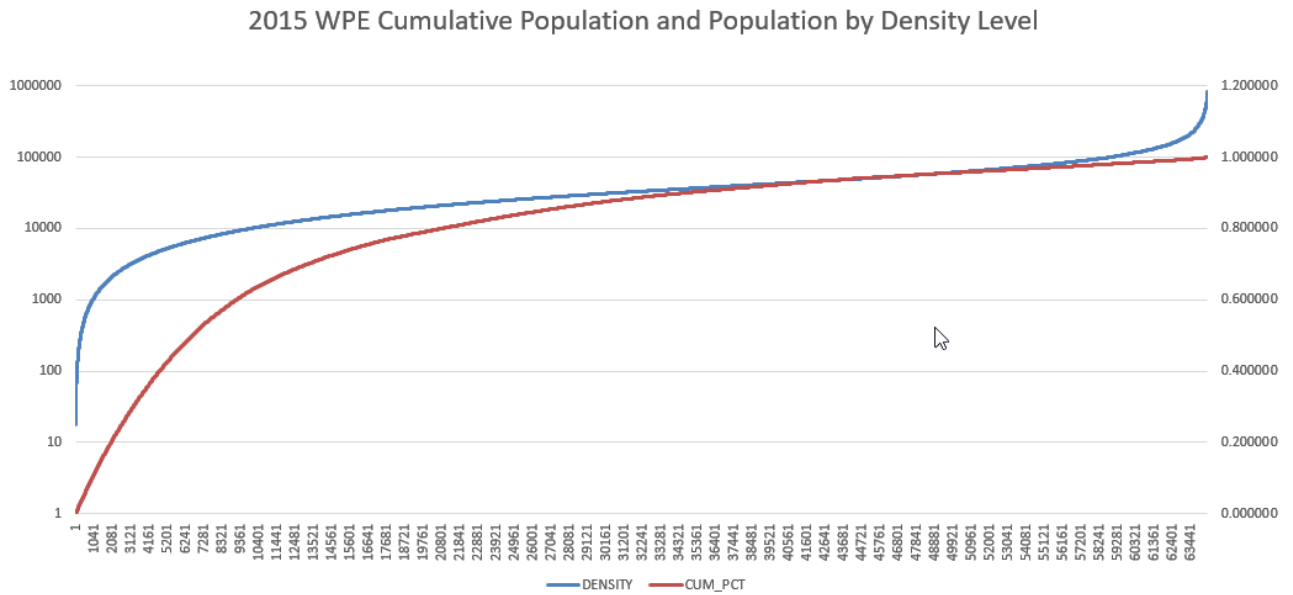
Figure 5. Plot comparing the cumulative percent of total population in blue by population density level in red. The purpose is to note the inflection points in the blue curve occurs in 2015 at around 62,000 persons per square kilometer, indicating a much higher number of density spikes due to horizontal mismatches. The inflection point for these upper values can be seen in in the lower right of Figure 4.
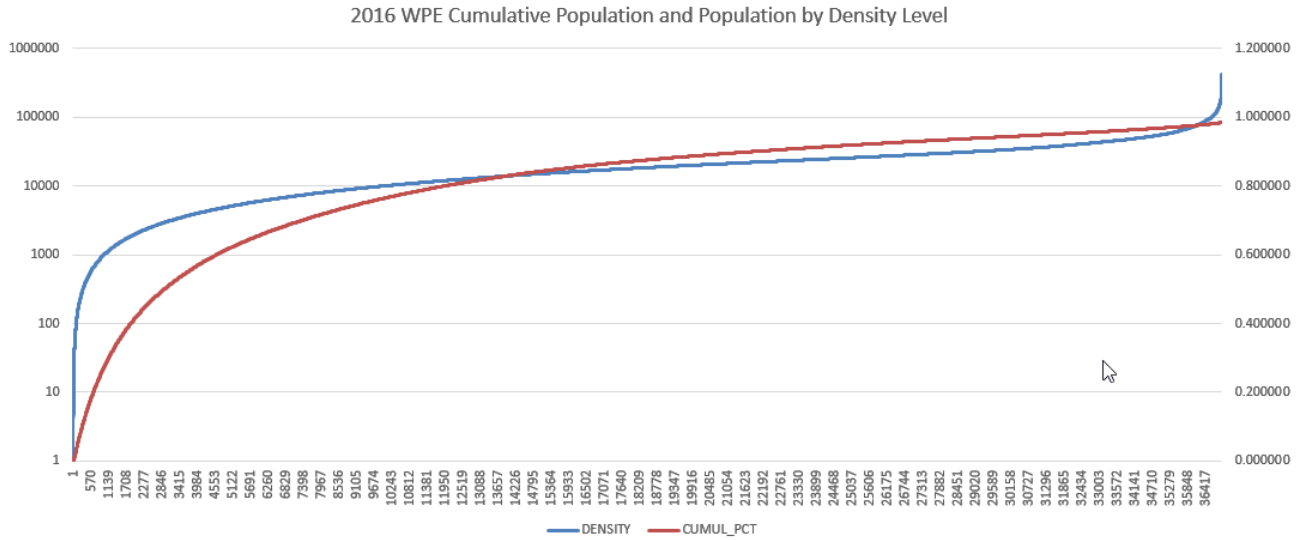
Figure 6. Plot comparing the cumulative percent of total population in blue by population density level in red. The purpose is to note the inflection points in the blue curve occurs in 2016 at around 36,000 persons per square kilometer, indicating a general improvement in the occurrence of density spikes due to horizontal mismatches.

It should be noted that given these localized spikes in population it is still possible that the sum of a population for a larger region that includes such spikes is likely to still be accurate, it is just that the population was distributed to a footprint in a skewed way.

v.   Minimum Size of Area to analyze

What is the smallest number of cells that the WPE's model can reliably assign a population estimate to?

It depends on where that area is located. If the average confidence level (described in section II.E.iii) is good (confidence value = 1), it is recommended to use at least an 8x8 area of cells (~1 square kilometer). For areas of lower confidence, it is recommended to multiply this minimum area by ten for confidence level 2, by twenty for confidence level 3, and multiply by 50 and 100 for confidence levels 4 and 5. Thus, an area that is on average a confidence level of 3 should be at least 20 sqkm in area, and an area having an average confidence of 4 should be at least 50 sqkm. If populations in areas with lower confidence are consistent with current local knowledge, then smaller areas may be used.

III. Method

This section describes the sequence of processing steps, and expectations for each step, used to produce the WPE layers. The description herein is intended to facilitate replication of the method and collaboration to

improve the method. Note that occasionally, in the interest of brevity, the description herein may not follow the exact sequence in the actual workflow, and these exceptions will be noted.

A. Assign Base Scores from Classified Landcover and Ancillary Data

The process to create the WPE begins with assigning base scores to the portions of the landscape based on the likelihood of people living there given the landcover class. Several sources of information are used for this purpose:

- Landcover: MDA's BaseVue which is a global 30-meter resolution classified land cover product using an Anderson-style classification. BaseVue is also created from Landsat8 imagery.
- Populated place points: GeoNames has gazetteer points for populated places, which are useful to locate known locations of settlement that are not identified in the landcover data. The Populated Place Points are buffered at a distance of 1-km (radius), and the resulting buffer polygons are rasterized to 30-meter resolution and snapped to the Landsat8 Panchromatic imagery.
- Road intersections: Road intersections are derived from HERE and OpenStreetMap (OSM), and represent the idea that people usually live within a short distance of roads. Road intersections can be efficiently generated in ArcGIS by creating a geometric network from the road line features in a geodatabase—one byproduct of the network is a "Junctions" feature class which can be copied and used separately. The points are rasterized at 150-meter resolution and snapped to the Landsat8 Panchromatic imagery.

The base settlement scores are calculated as follows:

1. Create two mosaic datasets containing the MDA land cover rasters and assign remap functions to each as follows:
    a. Urban:
        i. 21 = 200 (High Density Urban)
        ii. 22 = 150 (Medium Density Urban)
        iii. 0-20 = 0
    b. Agriculture:
        i. 1-5 = 25 (Forest, Scrub, Grassland)
        ii. 7-9 = 25 (Agriculture)
        iii. 15-16 = 25 (Mixed forest [U.S. only])
        iv. 5-7 = 0 (Barren)
        v. 9-15 = 0 (Water, Snow, Wetlands and Clouds)

vi.   16-22 = 0 (Unused or Urban)

All subsequent steps are performed for each country or sub country processing unit.

2. Use Extract by Mask for Urban, Agriculture, Road Intersections, GeoNames, Landsat Band 8 (panchromatic) and Landsat8 QA.
3. Use Cell Statistics Maximum between Agriculture, Urban, Road Intersections and Populated Place Points to produce the base settlement scores, which have values ranging from 0 to 200.

Figure 6. Shows BaseVue 2013 imagery on the left, and the results of Step 1 on the right, where the urban locations are shown in darker red, agriculture in light red, and areas where people do not live are white.

B. Landsat8 Panchromatic Texture Score

The base scores provide a high certainty for where people live in the urban areas, but within the agricultural areas many locals were included where people obviously did not live, i.e., in the middle of fields and orchards. Therefore, Esri needed a method to filter out the locations where people obviously did not live from within any areas with a base score less than 100.

i.   Initial Selection of Landsat8 Scenes

The location of Landsat scenes is consistent, and there are just over 8,900 scene locations (identified as path / row combinations) covering the earth's inhabited land area. New scenes are added every sixteen days for these locations. The locations for scenes are based on a fixed combination of path and row coordinates. More information is available at:

- https://aws.amazon.com/public-data-sets/landsat/
- http://landsat.usgs.gov/l8handbook_section4.php

In 2016, Esri created a custom Landsat8 mosaic dataset to produce the texture score for the WPE. This mosaic selected a Landsat8 scene at each path / row location based on the following criteria, in order of priority:

17

- Lowest cloud cover
- Calendar year of 2016
- In temperate areas leaf off imagery was preferred and found by restricting the selection of scenes to a narrow set of months.
- In extreme latitudes, summer, leaf on imagery was preferred because snow tends to cover structures and settlement, reducing the detectable texture.

Esri obtained scene metadata from the AWS volume to create a score used to choose which of the scenes to use. It was found that the cloud cover metadata was unreliable in about one percent of the scenes, resulting in a manual process to verify the scenes as being best suited for deriving texture. Additionally, about fifty path / row locations did not contain even relatively cloud-free scenes within the 2016 calendar year, and earlier scenes were substituted.

For the 2013 and 2015 WPE layers, the processing is accomplished using a tile-based workflow with 0.5 degree tiles for the land areas of the world. This was done because large raster datasets (large column or row counts) processed very slowly. Cell count and processing time have a geometric linear relationship.

Using these 0.5 degree tiles as units of work resulted in the introduction of statistical artifacts due to arbitrary small sample sizes when a tile happened to include only a small portion of land. Thus, for the 2016 WPE layers, the processing was changed to countries for small countries, and groups of admin level 1 areas for larger countries such as Russia, Brazil, etc. This level of process, along with TOA correction served to eliminate statistical artifacts that were produced in Esri's earlier workflows for producing the WPE settlement score.

ii. QA Band Cloud and Snow Masking Processing

In 2013 and 2015 Esri presumed that extremely high reflectance values in the panchromatic imagery represented clouds or snow, and did not include them in the analysis to find texture. Visual analysis showed this was not reliable, in that many clouds were still included in the texture analysis.

In 2016, Esri used the USGS's quality assessment (QA) band, for each scene available on the AWS site. This band contains masks for clouds and snow. Visual analysis found this was a better solution, but there were still many errors, and a better solution is still needed.

The following steps pick up with how the cloud mask was incorporated into the process:

4. Use Set Null to set any cell in the base settlement score raster to zero, as no meaningful texture can be derived for these locations.

iii. Deriving the Texture Score

Esri developed an algorithm to identify significant texture in panchromatic Landsat8 imagery as the basis to differentiate likely settlement within these areas. The analog for this algorithm is the idea that a building's roof would have high reflectance, and the shadow cast by the building in an adjacent cell would have a low value. Thus, texture representing settlement would have at least two pairs of such cells within a small (5x5 cell) neighborhood, which is roughly equivalent to one city block.

Identifying such pairings of values in raster data is not new. Such cell pairings have long been considered as "features" within raster data, and algorithms to find or count such features started in 1973 with Haralick's algorithm for identifying textural features within grayscale imagery. The drawback to feature identification and counting algorithms is that they are computationally intensive.

Esri's algorithm moves in a novel direction from this earlier work. Instead of counting features, Esri uses the sum of ranges within a 5x5 neighborhood to produce a score. When that score is higher than the mean plus the standard deviation of the entire country's dataset, then it is considered to be high enough to likely represent structures. In the internal review of this method at Esri, staff statisticians were consulted and they noted the similarities between this algorithm and those used in terrain analysis.

The panchromatic band (Band 8) from Landsat8 is stored as 16-bit unsigned integer where values range from 0 (No Data), then skip 1 to 5,499, then continue from 5,500 to 65,535, representing reflectance levels.

5. Input the panchromatic image produced in step 2 into the Focal Statistics tool to calculate the Range (max – min) of cell values within a 5x5 cell moving neighborhood area window (NAW). This indicates how much local variance of cells values exists. See Figures 7 and 8.
6. Input the result of step 5 into the Focal Statistics tool which calculates the Sum of cell values within a 5x5 cell moving NAW. This produces a sum of ranges. The high values indicate a high local variety that it is not due to one-cell anomalies. See Figures 7 and 8.
7. Get the mean and the standard deviation of the output of step 6 (Sum of ranges) of the country. Add them together to produce the threshold level for meaningful texture.

19

8. Filter out locations in the base settlement score (step 3) that have values of less than 100, and a texture score less than the threshold from step 7. Use the con tool to accomplish this in a single processing step.
9. Scale the texture score values by taking the natural log, then subtracting 11 and multiplying by 50. This produces a range of about -50 to 275.

The initial texture score can be expressed as follows:

$$T_c = \sum_{i=c}^{25} (C_{Max} - C_{Min})_{\ n=5x5}$$

$$T_{Settlement} = T_c > (T\mu_s + T\sigma_s)$$

Where: T = Texture
c = cell location
n = neighborhood area window
s = Each country or sub-country processing unit
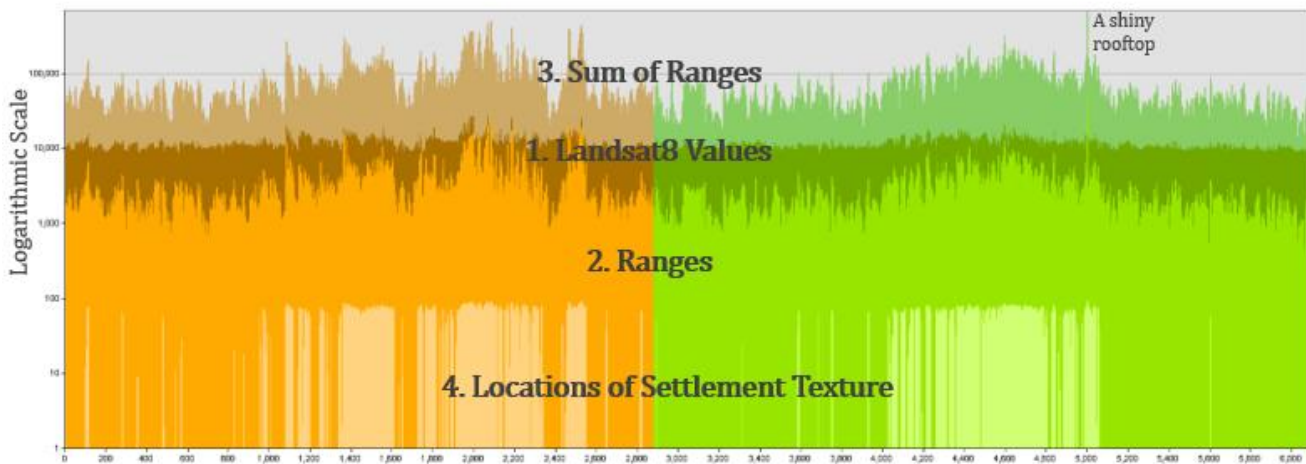
Figure 7. A graph of profile paths shown in Figure 8. This illustrates which cells are selected as settlement, which can be verified by looking along the profile lines (from left to right).

Steps 5-9 identifies texture we most associate with human settlement, but also find other circumstances with the same statistical profile. Some settlement texture is missed due to the coarse cell size of 15-meters. Figure 9 depicts several prototypical texture scenarios representing texture values in adjacent raster cells in the panchromatic Landsat8 imagery.

One issue that will be tested for the 2016 process is the mosaic dataset that references the panchromatic Landsat8 scenes will have a correction for top of atmosphere radiance. The theory is doing so will better normalize the values on a per scene basis, allowing global thresholds (versus the tile's statistics).
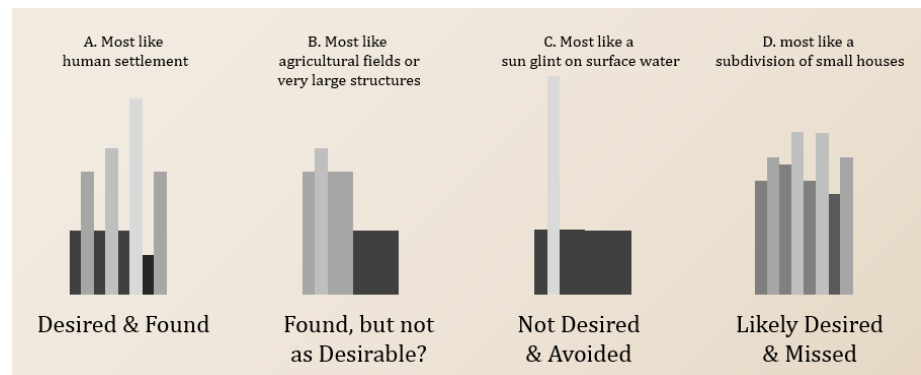
is missed because the local range is too low.

21

C. Final Settlement Score

The final footprint score is accomplished in two stages. The first combines the Landsat8 texture score to the land cover score as follows:

10. Add the output of Step 9 to the output of Step 3 to produce an initial settlement score. This is a 15-meter cell size raster.
11. Aggregate the result of step 10 by a factor of 5 using the Maximum option to produce the final 75-meter resolution settlement score raster.

D. Redistribution of Population

To create the WPE's population estimate raster, population data stored as polygons representing enumeration or tabulation units, is redistributed based on the footprint and values of the settlement score raster. There is one overarching rule in this process, which is population must be allocated as "integer people. It is possible that some population polygons do not contain a settlement point, either because they are too small, or because no meaningful settlement was found within that region of the settlement score raster.

12. Determine whether there are any polygons that are smaller than a 75-meter raster cell (0.00000045 Decimal Degrees). If so, follow the steps under 12a. if not skip to step 13.
    a. Using Feature to Point, obtain a centroid for each of the small polygons, while retaining the population attribute.
    b. Convert the centroids to raster using the result of Step 11 as the cell size and snap raster. This creates a floating point raster, which should be converted to Integer.
    c. Use Raster to Polygon to convert the result of 12b to polygons that are now large enough work with.
    d. Use the Update tool to merge the results of 12c into the original polygon dataset. This will replace the original small polygons with the larger, while updating the topology. So far we have no examples of two tiny polygons being closer than 75 meters to each other, which would break this process, causing one to be eliminated.
13. Convert the population polygons with population greater than zero to a raster such that the value for the resulting raster dataset is the polygon's object ID, which will serve as unique identifier (to ensure that different polygons with the same population are not processed in an aggregate fashion).

14. Convert the population polygons with population greater than zero to a raster such that the value for the resulting raster dataset is the population of the polygon.
15. Run zonal statistics where the zones dataset is the result of step 13, and the input dataset is the result of Step 11 (settlement score) using the sum option to obtain the sum of settlement score for each population polygon ID.
16. Use the Divide tool to divide the result of step 15 by the result of 14 to obtain the ratio of each population polygons population to sum of settlement.
17. Use the Times tool to multiply the result of Step 16 by the result of Step 11, then use the Minus tool to subtract 0.5155, Then use the RoundUp tool, and convert the result to Integer—which is the redistributed population
18. Aggregate the final result of step 17 by a factor of 2 to produce the final WPE Population Raster. This produces a 150 meter raster dataset, which when used as the basis for an image service allows an area the size of Africa to be made available for analysis without causing undo impacts to the server's CPU, Disk, or Memory consumption.

The above steps are predicated on a, thus far, undescribed population polygon layer. This layer is an amalgamation of many sources, with the intent of using the most current publicly available or available at low cost census or census surrogate data applied to the finest levels of geography possible. One of the goals of each new version of the WPE is to improve upon this by adding more polygons representing progressively finer tabulation geographies. The current population polygons dataset contains slightly over 2.1 million polygons from the following sources:

- United States: U.S. Census Block Group with Esri's current year estimate
- Canada: Environics Analytics at the Dissemination Area (DA) level.
- Michael Bauer Research GmbH: 130 countries at Admin level 3 (county), admin level 4 (city/town), or 5-digit postal code for the most recent year of estimate.
- United Nations most recent estimate in all other cases usually at admin level 3 though some are 2 (state).

E. Deriving Population Density
Population Density was derived using the following steps:

19. Convert the result of Step 18 to polygon

20. Using the Add Geometry Attributes tool, calculate a new double field with the geodesic area of each polygon in units of square kilometers.
21. Calculate a new double field for the population density by dividing population by the result of step 20.
22. Convert the result to a raster using the result of Step 18 as the cell size and snap raster.

F. Confidence Surface

The 2016 WPE Confidence surface based on the average of several factors:

- Reliability of Census information: This includes the age of the most recent census, the age of any subsequent estimate that was used, and the type of census. The reliability scores are based on the [United Nations Statistics Division's UNData program](#)'s characterizations.
- Ratio of the area of the population polygon to the number of people,
- Complexity the footprint of settlement score relative to NoData and zero population cells.

i. Population to Area Confidence Scoring

Table 3. provides the confidence scores used for the combination of polygon area and population.

Population Density (persons/sqkm)	Table 3: Area (sq.km) Classes of Population Polygon Data						
	0 to 1.5	1.5 to 5.0	5.0 to 10	10 to 100	100 to 1,000	1,000 to 10,000	over 10,000
Over 5,000	4	4	4	4	3	3	3
1,000 to 5,000	5	5	5	4	4	4	3
134 to 1,000	5	4	4	4	3	3	3
67 to 134	4	4	4	3	3	3	3
5 to 67	4	4	3	2	2	2	2
1 to 5	4	4	3	2	2	1	1
Zero	3	3	4	4	3	2	1

Table 3. Shows a matrix of confidence scores based on the population of a given reporting polygon and its area in square kilometers. The range is 5 for highest confidence and 1 for lowest confidence.

ii. Footprint Complexity Score

The footprint complexity was computed as a per cell score that rated the number and distance to no data cells within 8 kilometers. The reason complexity is a problem is due to necessary raster resampling processes that occur during the WPE projection workflow. Edges of data and NoData cause allocation errors between raster datasets and between raster and vector datasets, typically resulting in underestimation. The processing steps are as follows:

1. Project the WGS_1984 output of the 150-meter resolution aggregation to World Equidistant Cylindrical. This ensured the distances between cell centers would be comparable.
2. To speed the processing, the result of Step 1 was aggregated by a factor of 2 cells with a sum operation.
3. Reclassify the result of step 2 where values of zero or NoData are set to NoData, and all other values are set to 1.
4. Use the result of Step 3 as the input to the Euclidean Distance tool. Then divide that result by 1000, to convert it to kilometers.
5. Use the Get Raster Properties to learn the Maximum value from the result of Step 4.
6. Use the result of Step 3 as the input to Focal Statistics, set to a Circular Neighborhood with a radius of 8 cells, and a Sum operation.
7. Divide the result of Step 6 by the result of Step 4. This produces a ratio of the count of data cells within 8 kilometers to the distance to the nearest data cell
8. Reclassify the result of Step 7 based on a 5-Class Natural Breaks distribution such that the low values receive a value of 1 and the high values a value of 5.

Esri used the confidence scores as zones for the Zonal Statistics to table tool and summed the population within each confidence level.

Table 4: Summary of Confidence Levels			
Confidence Level	Count of people	% of Total population	% Urban Density
1 = Least	32,679,164	0.45%	0.85%
2	1,406,643,023	19.30%	20.12%
3	2,506,831,960	34.40%	35.74%
4	2,720,870,831	37.33%	36.96%
5 = Most	620,957,914	8.52%	6.33%

Table 4. Shows the counts, percent of total, and urban percentages at each confidence level.

Additional factors are planned for future editions of the confidence surface, including a score for the quality of the census information that provides the population figure to each populated polygon. The idea is for the best de-jure censuses to have a confidence of 5, while the countries without censuses would have a value of 1, all others may have a value of three unless the United Nations Statistical Division has a quality score to use. A second metric to include would be to factor in local variance of scores, which should be relatively low.

IV. References

The following references were used in the course of deriving and confirming the method presented herein for producing the WPE.

Azar, D Graesser, J Engstrom, R, Comenetz, J Leddy, R M Schechtman, N G and **Andrews T** 2010 Spatial Refinement of Census Population Distribution Using Remotely Sensed Estimates of Impervious Surfaces in Haiti" *International Journal of Remote Sensing* 31.21 (2010): 5635-655. Web.

Balk, D 2003 "Improving Global Population Estimates" In: Conference on Migration, Urbanization, and Health, Princeton University, September 25, 2003.

Bhaduri, B and **Bright, E** 2003 LandScan Population Projects. In: Conference on Migration, Urbanization, and Health, Princeton University, September 25, 2003.

Bhaduri, B Bright, E Coleman, P and **Urban, M L** 2007 LandScan USA: A High-resolution Geospatial and Temporal Modeling Approach for Population Distribution and Dynamics. *GeoJournal* 69.1-2 (2007): 103-17. Web.

Bielecka, E 2005 A Dasymetric Population Density Map of Poland. In: International Cartographic Conference, A Coruña, Spain, August 2005.

Cheriyadat, A Bright, E Potere, D and **Budhendra, B** 2007 Mapping of Settlements in High-resolution Satellite Imagery Using High Performance Computing. *GeoJournal* 69.1-2 (2007): 119-29. Web.

Encarnação, S, Gaudiano, M Santos, F C Tenedório, J A and **Pacheco, J M** 2012 Fractal Cartography of Urban Areas. *Scientific Reports* 2 (2012): n. Web.

Encarnação, S, Gaudiano, M Santos, F C Tenedório, J A and **Pacheco, J M** 2012, Fractal Cartography of Urban Areas: Supplementary Information. *Scientific Reports* 2 (2012): Web.

Esch, T Taubenbock, H Roth, A Heldens, W Felbier, A Theil, M Schmidt, M Muller, A A and **Dech, S** 2012 TandDEM-X Mssion- New Perspectives for the Inventory and Monitoring of Global Settlement Patterns. *Journal of Applied Remote Sensing* 6 (2012): 0617021-06170221. Web.

Esch, T Marconcini, M Felbier, A Roth, A Heldens, W Huber, M Schwinger, M Taubenbock, H Muller, A A and **Dech S** 2013 Urban Footprint

Processor—Fully Automated Processing Chain Generating Settlement Masks From Global Data of the TanDEM-X Mission. *IEEE Geoscience and Remote Sensing Letters* 10.6 (2013): 1617-621. Web.

Esch, T Marconcini, M Marmanis, D Zeidler, J Elsayed, S Muller, A A and **Dech, S** 2014 Dimensioning Urbanization- An advanced procedure for characterizing human settlement properties and patterns using spatial network analysis. *Applied Geography* 55 (2014): 212-228. Web.

Getchee Inc. 2011 *A Quick View of Grid Demographic Processing. Web*

Haralick, R M, **Shanmugam, K** and **Dinstein, I** 1973 *Textural Features for Image Classification* IEEE Transactions on Systems, Man and Cybernetics Vol. SMC-3 No. 6 Nov 1973 pp 610-621.

Kim, B Youkyung, H Yonghyun, K and **Yongil, K** 2014 Generation of Cloud-free Imagery Using Landsat-8. *Journal of the Korean Society of Surveying Geodesy Photogrammetry and Cartography* 32(2):133-142, April 2014

Linard, C and **Tatem, A J** 2012 Large-scale Spatial Population Databases in Infectious Disease Research. *International Journal of Health Geographics* 11.1 (2012): 7. Web.

Linard, C Gilbert, M and **Tatem, A J** 2012 Assessing the Use of Global Land Cover Data for Guiding Large Area Population Distribution Modelling. *GeoJournal* 76.5 (2011): 525-38. Web.

Maantay, J Maroko, A and **Herrmann, C** 2007, Mapping Population Distribution in the Urban Environment: The Cadastral-based Expert Dasymetric System (CEDS). *Cartographic and Geographic Information Science* 34.2 (2007): 77-102. Web.

Miyamota, E and **Merryman T Jr** 2005 Fast Calculation of Haralick Texture Features Unpublished Classroom Materials accessed from https://users.ece.cmu.edu/~pueschel/teaching/18-799B-CMU-spring05/material/eizan-tad.pdf

Patterson, L Urban, M Myers, A Bhaduri, B Bright, E A and **Coleman, P R** 2009 The Effects of Quality Control on Decreasing Error Propagation in the LandScan USA Population Distribution Model: A Case Study of Philadelphia County. *Transactions in GIS* 13.2 (2009): 215-28. Web.

Pesaresi, M, Blaes, X Ehrlich, D Ferri, S Gueguen, L Haag, F Halkia, M Heinzel, J Kauffmann, M Kemper, Ouzounis, T O Scavazzon, M Soille, P

Syrris, V and **Zanchetta, L** 2012 *A Global Human Settlement Layer from Optical High Resolution Imagery*. JRC Publications Repository. Publications Office of the European Union, 2012. Web. http://publications.jrc.ec.europa.eu/repository/handle/JRC77925.

Sabesan, A Abercrombie, K Ganguly, A R Budhendra, B Bright, E A and **Coleman, P R** 2007 Metrics for the Comparative Analysis of Geospatial Datasets with Applications to High-resolution Grid-based Population Data. *GeoJournal* 69.1-2 (2007): 81-91. Web.

Sleeter, R 2014 Dasymetric Mapping Techniques for the San Francisco Bay Region, California In: URISA 2014.

Tang, J 2003 Evaluating the Relationship between Urban Road Pattern and Population Using Fractal Geometry. UCGIS.org Web: http://www.ucgis.org/summer03/studentpapers/junmeitang.pdf

Tatem, A J Adamo, A Bharti, N Burgert, C R Castro, M Dorelien, A Fink, G Linard, C Mendelsohn, J Montana, L Montgomery, M R Nelson, A Noor, A M Pindolia, D Yetman, G and **Balk, D** 2012 Mapping Populations at Risk: Improving Spatial Demographic Data for Infectious Disease Modeling and Metric Derivation. *Population Health Metrics* 10.1 (2012): 8. Web.

Taubenbock, T Esch, T Felbier, A Wiesner, M Roth, A and **Dech, S** 2012 Monitoring urbanization in mega cities from space. *Remote Sensing of Environment* 117 (2012): 162-176. Web.

Taubenbock, T Esch, T Roth, A and **Dech, S**, 2011 Pattern-Based Accuracy Assessment of an Urban Footprint Classification Using TerraSAR-X Data. *IEEE Geoscience and Remote Sensing Letters* 8.2 (2011): 278-282. Web.

Zhou, Y Yang, G Wang, S Wang, L Wang F and **Liu, X** 2014, A new index for mapping built-up and bare land areas from Landsat-8 OLI data. *Remote Sensing Letters*, 5:10, 862-871, DOI: 10.1080/2150704X.2014.973996

Appendix: WPE Data Sources

Country	Provider	Source
Afghanistan	U.N. Population Div.	1979 Census with 2003-2005 Household Listing
Albania	Michael Bauer Research GmbH	Based on Instat
Algeria	Michael Bauer Research GmbH	Based on Office national des statistiques
Andorra	Michael Bauer Research GmbH	Based on Departament d'Estadistica
Angola	Michael Bauer Research GmbH	Based on Instituto Nacional de Estatistica de Angola
Anguilla	U.N. Population Div.	Based on 2011 Census
Antigua and Barbuda	U.N. Population Div.	Based on 2011 Census
Argentina	Michael Bauer Research GmbH	Based on Instituto Nacional de Estadistica y Censos
Armenia	Michael Bauer Research GmbH	Based on National Statistical Service of the Republic or Armenia
Aruba	Michael Bauer Research GmbH	Michael Bauer Research GmbH
Australia	Michael Bauer Research GmbH	Based on Australian Bureau of Statistics
Austria	Michael Bauer Research GmbH	Based on Bundesanstalt Statistik Osterreich
Azerbaijan	Michael Bauer Research GmbH	Based on State Statistical Committee of Azerbaijan Republic
Bahamas	Michael Bauer Research GmbH	Michael Bauer Research GmbH
Bahrain	Michael Bauer Research GmbH	Based on Central Informatics Organisation - Statistics
Bangladesh	Michael Bauer Research GmbH	Based on Bureau of Statistics
Barbados	U.N. Population Div.	Based on 2010 Census
Belarus	Michael Bauer Research GmbH	Based on National Statistical Comittee of the Republic of Belarus
Belgium	Michael Bauer Research GmbH	Based on Direction generale Statistique et Information economique
Belize	U.N. Population Div.	2010 Census, with official estimates of the 2015 population
Benin	U.N. Population Div.	Based on 2013 Census
Bermuda	Michael Bauer Research GmbH	Based on Michael Bauer Research GmbH
Bhutan	U.N. Population Div.	Based on 2005 Census
Bolivia	Michael Bauer Research GmbH	Based on Instituto Nacional de Estadistica
Bosnia and Herzegovina	Michael Bauer Research GmbH	Based on Agencija za statistiku BiH
Botswana	Michael Bauer Research GmbH	Michael Bauer Research GmbH
Bouvet Island (uninhabited)	Esri	U.S. Central Intelligence Agency, The World Factbook
Brazil	Michael Bauer Research GmbH	Based on Instituto Brasileiro de Geografia e Estatistica (IBGE)
British Virgin Islands	U.N. Population Div.	2001 Census with official estimates through 2013
Brunei Darussalam	U.N. Population Div.	2011 Census with official estimates for 2015
Bulgaria	Michael Bauer Research GmbH	Based on National Statistical Institute
Burkina Faso	U.N. Population Div.	Based on 2010 Census
Burundi	U.N. Population Div.	Based on 2008 Census
Cabo Verde	U.N. Population Div.	2010 Census with official stimates for 2011
Cambodia	U.N. Population Div.	2008 Census with 2004 Cambodia Inter-Censal Population Survey
Cameroon	Michael Bauer Research GmbH	Michael Bauer Research GmbH
Canada	(c) Esri Canada, 2017	Statistics Canada
Cayman Islands	Michael Bauer Research GmbH	Michael Bauer Research GmbH
Central African Republic	U.N. Population Div.	Based on 2003 Census
Chad	U.N. Population Div.	Based on 2009 Census
Chile	Michael Bauer Research GmbH	Based on Instituto Nacional de Estadisticas
China	Michael Bauer Research GmbH	Based on National Bureau of Statistics of China
Christmas Island	Esri	Australia Bureau of Statistics
Cocos Islands	Esri	Australia Bureau of Statistics
Colombia	Michael Bauer Research GmbH	Based on Departamento Administrativo Nacional de Estadistica (DANE)
Comoros	U.N. Population Div.	Based on 2003 Census
Congo	U.N. Population Div.	Based on 2007 Census
Congo DRC	U.N. Population Div.	2007 Census with 2013/14 Demographic Health Survey
Cook Islands	U.N. Population Div.	Based on 2011 Census
Costa Rica	Michael Bauer Research GmbH	Based on Instituto Nacional de Estadistica y Censos
Côte d'Ivoire	Michael Bauer Research GmbH	Michael Bauer Research GmbH
Croatia	Michael Bauer Research GmbH	Based on Instat

Cuba	U.N. Population Div.	2012 Census with official estimates through 2015
Curacao	U.N. Population Div.	2011 Census with official estimates through 2016
Cyprus	Michael Bauer Research GmbH	Based on Statistical Service of Cyprus
Czech Republic	Michael Bauer Research GmbH	Based on Czech Statistical Office
Denmark	Michael Bauer Research GmbH	Based on Statistics Denmark
Djibouti	U.N. Population Div.	Based on 2009 Census
Dominica	U.N. Population Div.	Based on 2011 Census
Dominican Republic	Michael Bauer Research GmbH	Michael Bauer Research GmbH
Ecuador	Michael Bauer Research GmbH	Based on Instituto Nacional de Estatistica y Censo
Egypt	Michael Bauer Research GmbH	Michael Bauer Research GmbH
El Salvador	Michael Bauer Research GmbH	Based on Instituto Nacional de Estatistica y Censo
Equatorial Guinea	U.N. Population Div.	Based on 2002 Census
Eritrea	U.N. Population Div.	Based on 1984 Census
Estonia	Michael Bauer Research GmbH	Based on Statistics Estonia
Ethiopia	U.N. Population Div.	Based on 2007 Census
Falkland Islands	U.N. Population Div.	Based on 2012 Census
Faroe Islands	Michael Bauer Research GmbH	Based on Statistics Faroe Islands
Fiji	U.N. Population Div.	Based on 2007 Census
Finland	Michael Bauer Research GmbH	Based on Statstics Finland
France	Michael Bauer Research GmbH	Based on INSEE
French Southern Territories	Esri	U.S. Central Intelligence Agency, The World Factbook
Gabon	U.N. Population Div.	2013 Census with official estimates through 2015
Gambia	U.N. Population Div.	Based on 2013 Census
Georgia	Michael Bauer Research GmbH	Based on National Statistics Office of Georgia
Germany	Michael Bauer Research GmbH	Based on Statistisches Bundesamt
Ghana	Michael Bauer Research GmbH	Michael Bauer Research GmbH
Gibraltar	U.N. Population Div.	Based on 2012 Census
Greece	Michael Bauer Research GmbH	Based on Hellenic Statistical Authority
Greenland	Michael Bauer Research GmbH	Based on Michael Bauer Research GmbH
Grenada	U.N. Population Div.	Based on 2011 Census
Guatemala	Michael Bauer Research GmbH	Based on Instituto Nacional de Estadistica
Guernsey	Esri	U.S. Central Intelligence Agency, The World Factbook
Guinea	U.N. Population Div.	Based on 2011 Estimates
Guinea-Bissau	U.N. Population Div.	Based on 1991 Census
Guyana	U.N. Population Div.	Based on 2012 Census
Haiti	U.N. Population Div.	Based on 2003 Census
Heard Island and McDonald Islands	Esri	Commonwealth of Australia 2017
Honduras	Michael Bauer Research GmbH	Based on Instituto Nacional de Estadistica
Hungary	Michael Bauer Research GmbH	Based on Hungarian Central Statistical Office
Iceland	Michael Bauer Research GmbH	Based on Statistics Iceland
India	Michael Bauer Research GmbH	Based on Ministry of Statistics and Programme Implementation
Indonesia	Michael Bauer Research GmbH	Based on BPS- Statistics Indonesia
Iran	U.N. Population Div.	Based on 2011 Census with primary results of 2016 Census
Iraq	U.N. Population Div.	Based on 1997 Census and survey data through 2011
Ireland	Michael Bauer Research GmbH	Based on Central Statistics Office
Isle of Man	U.N. Population Div.	Based on 2011 Census
Israel	Michael Bauer Research GmbH	Based on Central Bureau of Statistics
Italy	Michael Bauer Research GmbH	Based on Italian National Institute of Statistics
Jamaica	Michael Bauer Research GmbH	Michael Bauer Research GmbH
Japan	Michael Bauer Research GmbH	Based on Statistics Bureau
Jersey	Esri	U.S. Central Intelligence Agency, The World Factbook
Jordan	Michael Bauer Research GmbH	Based on Department of Statistics
Kazakhstan	Michael Bauer Research GmbH	Based on Statistical Office
Kenya	Michael Bauer Research GmbH	Michael Bauer Research GmbH
Kiribati	U.N. Population Div.	Based on 2010 Census
Kosovo	Michael Bauer Research GmbH	Based on Statistical Office of the Republic of Serbia
Kuwait	Michael Bauer Research GmbH	Based on Central Statistical Office

Kyrgyzstan	Michael Bauer Research GmbH	Based on National Statistical Committee of Kyrgyz Republic
Lao People's Democratic Republic	U.N. Population Div.	Based on 2015 Census
Latvia	Michael Bauer Research GmbH	Based on Central Statistical Bureau
Lebanon	Michael Bauer Research GmbH	Based on Central Administration for Statistics
Lesotho	Michael Bauer Research GmbH	Michael Bauer Research GmbH
Liberia	U.N. Population Div.	Based on 2008 Census
Libya	U.N. Population Div.	Based on 2006 Census and official population estimates for 2015
Liechtenstein	Michael Bauer Research GmbH	Based on Amt fur Statistik
Lithuania	Michael Bauer Research GmbH	Based on Statistics Lithuania
Luxembourg	Michael Bauer Research GmbH	Based on Statec
Madagascar	U.N. Population Div.	Based on 1993 Census and official population estimates for 2011
Malawi	Michael Bauer Research GmbH	Michael Bauer Research GmbH
Malaysia	Michael Bauer Research GmbH	Based on Department of Statistics Malaysia
Maldives	U.N. Population Div.	Based on 2014 Census
Mali	U.N. Population Div.	Based on 2009 Census
Malta	Michael Bauer Research GmbH	Based on National Statistics Office
Marshall Islands	U.N. Population Div.	Based on 2011 Census
Mauritania	U.N. Population Div.	Based on 2013 Census
Martinique	Michael Bauer Research GmbH	Michael Bauer Research GmbH
Mauritius	Michael Bauer Research GmbH	Michael Bauer Research GmbH
Mexico	Michael Bauer Research GmbH	Michael Bauer Research GmbH
Micronesia	U.N. Population Div.	Based on 2010 Census
Moldova	Michael Bauer Research GmbH	Based on National Bureau of Statistics
Monaco	Michael Bauer Research GmbH	Based on Official Government Portal
Mongolia	Michael Bauer Research GmbH	Based on National Statistics Office
Montenegro	Michael Bauer Research GmbH	Based on Monstat
Montserrat	U.N. Population Div.	Based on 2011 Census
Morocco	Michael Bauer Research GmbH	Michael Bauer Research GmbH
Mozambique	Michael Bauer Research GmbH	Michael Bauer Research GmbH
Myanmar	U.N. Population Div.	Based on 1991 Census and official population estimates up to 2010
Namibia	Michael Bauer Research GmbH	Michael Bauer Research GmbH
Nauru	U.N. Population Div.	Based on 2011 Census
Nepal	U.N. Population Div.	Based on 2011 Census
Netherlands	Michael Bauer Research GmbH	Based on Centraal Bureau voor de Statistiek, Den Haag/Heerlen
New Zealand	Michael Bauer Research GmbH	Michael Bauer Research GmbH
New Caledonia	Michael Bauer Research GmbH	Michael Bauer Research GmbH
Nicaragua	Michael Bauer Research GmbH	Based on Instituto Nacional de Informacion de Desarrollo (INIDE)
Niger	U.N. Population Div.	Based on 2012 Census
Nigeria	U.N. Population Div.	Based on 2006 Census and 2011 MICS4 survey
Norfolk Island	Esri	Australian Bureau of Statistics
North Korea	U.N. Population Div.	Based on 2008 Census
Northern Mariana Islands	Esri	Based on 2010 Census
Norway	Michael Bauer Research GmbH	Based on Statistics Norway
Oman	Michael Bauer Research GmbH	Based on Department of Statistics Oman
Pakistan	Michael Bauer Research GmbH	Based on Federal Bureau of Statistics
Palau	U.N. Population Div.	Based on 2005 Census and de jure population in 2012 and 2015
Palestinian Territory	U.N. Population Div.	Based on official estimates from 1994 to 2012
Panama	Michael Bauer Research GmbH	Michael Bauer Research GmbH
Papua New Guinea	U.N. Population Div.	Based on Censuses from 1966 through 2011
Paraguay	Michael Bauer Research GmbH	Based on Direccion General de Estadistica, Encuestas y Censos
Peru	Michael Bauer Research GmbH	Based on Instituto Nacional de Estadistica e Informatica (INEI)
Philippines	Michael Bauer Research GmbH	Based on Philippine Statistics Authority
Pitcairn	Esri	U.S. Central Intelligence Agency, The World Factbook
Poland	Michael Bauer Research GmbH	Based on Central Statistics Office
Portugal	Michael Bauer Research GmbH	Based on Instituto Nacional de Estatistica - Portugal
Qatar	Michael Bauer Research GmbH	Based on Statistics Authority
Réunion	Michael Bauer Research GmbH	Michael Bauer Research GmbH

Romania	Michael Bauer Research GmbH	Based on National Institute of Statistics
Russian Federation	Michael Bauer Research GmbH	Based on Federal State Statistics Service
Rwanda	U.N. Population Div.	Based on 2012 Census
Saint Helena	St. Helena Statistics Office (2016)	St Helena 2016 Population & Housing Census
Saint Kitts and Nevis	U.N. Population Div.	Based on 2011 Census
Saint Lucia	U.N. Population Div.	Based on 2010 Census
Saint Vincent and the Grenadines	U.N. Population Div.	Based on 2012 Census
Samoa	U.N. Population Div.	Based on 2011 Census
San Marino	U.N. Population Div.	Based on 2010 Census and official estimates through 2015
Sao Tome and Principe	U.N. Population Div.	Based on 2012 Census
Saudi Arabia	Michael Bauer Research GmbH	Based on Ministry of Economy and Planning
Senegal	U.N. Population Div.	Based on 2013 Census
Serbia	Michael Bauer Research GmbH	Based on Statistical Office of the Republic of Serbia
Seychelles	U.N. Population Div.	Based on 2010 Census
Sierra Leone	U.N. Population Div.	Based on 2015 Census
Singapore	Michael Bauer Research GmbH	Based on Singapore Department of Statistics
Sint Maarten	U.N. Population Div.	Based on 2011 Census and 2016 official estimate
Slovakia	Michael Bauer Research GmbH	Based on Statistical Office of the Slovak Republic
Slovenia	Michael Bauer Research GmbH	Based on Statistical Office of the Republic of Slovenia
Solomon Islands	U.N. Population Div.	Based on 2009 Census
Somalia	U.N. Population Div.	Based on 1975 Census and 2014 Survey
South Africa	Michael Bauer Research GmbH	Michael Bauer Research GmbH
South Georgia	Esri	Government of South Georgia & the South Sandwich Islands
South Korea	Michael Bauer Research GmbH	Based on Statistics Korea
South Sudan	U.N. Population Div.	Based on 2008 Census (Sudan)
Spain	Michael Bauer Research GmbH	Based on National Statistics Institute, www.ine.es
Sri Lanka	Michael Bauer Research GmbH	Based on Department of Census and Statistics
Sudan	Michael Bauer Research GmbH	Michael Bauer Research GmbH
Suriname	U.N. Population Div.	Based on 2012 Census and 2013 and 2014 official estimates
Swaziland	Michael Bauer Research GmbH	Michael Bauer Research GmbH
Sweden	Michael Bauer Research GmbH	Michael Bauer Research GmbH
Switzerland	Michael Bauer Research GmbH	Based on Bundesamt fur Statistik
Syria	Michael Bauer Research GmbH	Based on 2004 Census and 2010 Estimate
Tajikistan	Michael Bauer Research GmbH	Based on State Statistical Committee of Tajikistan
Tanzania	Michael Bauer Research GmbH	Michael Bauer Research GmbH
Thailand	Michael Bauer Research GmbH	Based on National Statistics Office
Macedonia	Michael Bauer Research GmbH	Based on State Statistical Office
Timor-Leste	U.N. Population Div.	Based on 2010 Census
Togo	U.N. Population Div.	Based on 2010 Census
Tokelau	U.N. Population Div.	Based on 2011 Census
Tonga	U.N. Population Div.	Based on 2011 Census
Trinidad and Tobago	Michael Bauer Research GmbH	Michael Bauer Research GmbH
Tunisia	Michael Bauer Research GmbH	Michael Bauer Research GmbH
Turkey	Michael Bauer Research GmbH	Based on Turkish Statistical Institute
Turkmenistan	U.N. Population Div.	Based on 1989 and 1995 Censuses
Turks and Caicos Islands	U.N. Population Div.	Based on 2012 Census
Tuvalu	U.N. Population Div.	Based on 2012 Census
Uganda	Michael Bauer Research GmbH	Michael Bauer Research GmbH
Ukraine	Michael Bauer Research GmbH	Based on State Statistics Service of Ukraine
United Arab Emirates	Michael Bauer Research GmbH	Based on National Bureau of Statistics
United Kingdom	Michael Bauer Research GmbH	Based on Office for National Statistics (ONS)/UK Statistics Authority
United States	Esri	2010 Census and annual estimates through 2016
United States Minor Outlying Isls.	Esri	U.S. Central Intelligence Agency, The World Factbook
Uruguay	Michael Bauer Research GmbH	Based on Instituto Nacional de Estadistica
Uzbekistan	Michael Bauer Research GmbH	Based on The State Committee of Republic Uzbekistan on Statistics
Vanuatu	U.N. Population Div.	Based on 2009 Census
Vatican City	Esri	U.S. Central Intelligence Agency, The World Factbook

Venezuela	Michael Bauer Research GmbH	Based on Instituto Nacional de Estadistica
Vietnam	Michael Bauer Research GmbH	Based on General Statistics Office
Yemen	U.N. Population Div.	Based on 2004 Census and 2014 official population estimate
Zambia	Michael Bauer Research GmbH	Michael Bauer Research GmbH
Zimbabwe	U.N. Population Div.	Based on 2012 Census