Release Statement

Gridded population estimates for 40 countries in Latin America and the Caribbean using official population estimates, Version 1.0

9 January 2023

The data were produced by WorldPop at the University of Southampton. These data include gridded population estimates, at approximately 100m resolution, for 40 countries in Latin America and the Caribbean (Appendix A). These results were created using official population estimates at the finest-available resolution provided by National Statistic Offices (NSOs) throughout the region, and built-up area, height and volume covariates produced from World Settlement Footprint 3D (WSF3D) datasets[1].

We acknowledge the contribution of WorldPop's partners, notably the United Nations Population Fund (UNFPA) Latin America and Caribbean Regional Office in supporting the collection of population and administrative boundary data, and to the German Aerospace Center (DLR) for preparing and providing built settlement data from the WSF3D framework.

Modelling work and geospatial data processing was carried out by McKeen T., Bondarenko M., Kerr D. and Sorichetta A. Esch T., Marconcini M., Zeidler J. and Palacios-Lopez D. prepared and provided the WSF3D datasets. Juran S. and Valle C. aided with population and administrative boundary data collection. Oversight was provided by Andrew J. Tatem.

These data may be distributed using a [Creative Commons Attribution 4.0 International (CC BY 4.0) License](#). Contact release@worldpop.org for more information.

The authors followed rigorous procedures designed to ensure that the used data, the applied method and thus the results are appropriate and of reasonable quality. If users encounter apparent errors or misstatements, they should contact WorldPop at release@worldpop.org. WorldPop, University of Southampton, and their sponsors offer these data on a "where is, as is" basis; do not offer an express or implied warranty of any kind; do not guarantee the quality, applicability, accuracy, reliability or completeness of any data provided; and shall not be liable for incidental, consequential, or special damages arising out of the use of any data that they offer.

CITATION

McKeen T., Bondarenko M., Kerr D., Esch T., Marconcini M., Palacios-Lopez D., Zeidler J., Juran S., Tatem A.J. and Sorichetta A. 2023. *Gridded population estimates for 40 countries in Latin America and the Caribbean using official population estimates, Version 1.0.* Worldpop, University of Southampton. doi:10.5258/SOTON/WP00755.

SOURCE DATA

- *Population:* Official population statistics at the finest level of administrative units available were provided by NSOs from the region (Appendix 1).
- *Geospatial covariate layers:* A set of ancillary geospatial layers known to be related to human population distribution were constructed (Appendix 2).
- *Settlement layers:* Notably amongst these geospatial covariate layers, high-resolution settlement data via the WSF3D project were used (Appendix 2).

RELEASE CONTENT

1. **ppp_*XXX_RF1_v1.tif**
2. **ppp_*XXX_RF2_v1.tif**
3. **ppp_*XXX_RF3_v1.tif**
4. **ppp_*XXX_RF4_v1.tif**
5. **ppp_*XXX_RF5_v1.tif**

*Spatially referenced gridded rasters are provided for 40 countries, identifiable using the ISO-3 country codes in the filenames in place of XXX above. For example, ABW refers to Aruba. See appendix 1 for the full list of countries. For each country, these data can be downloaded as a WinRAR Zip archive.

FILE DESCRIPTIONS

The spatially referenced GeoTiff grid layers introduced in this description all have a resolution of 3 arc seconds (~100m at the equator) and have a geographic spatial reference (WGS1984). All GeoTiffs have a NoData value of -99999 (pixels that do not contain any people).

ppp_*XXX_RF1_v1.tif
Each grid cell of this dataset represents the estimated population count for a specified country as calculated by RF modelling scenario 1, i.e. RF model fitted with just '*base*' covariate set, no WSF3D-derived data (Appendix 2).

ppp_*XXX_RF2_v1.tif
Each grid cell of this dataset represents the estimated population count for a specified country as calculated by RF modelling scenario 2, i.e. RF model fitted with '*base*' covariate set and WSF3D built area layer (Appendix 2).

ppp_*XXX_RF3_v1.tif
Each grid cell of this dataset represents the estimated population count for a specified country as calculated by RF modelling scenario 3, i.e. RF model fitted with '*base*' covariate set and WSF3D built height layer (Appendix 2).

ppp_*XXX_RF4_v1.tif
Each grid cell of this dataset represents the estimated population count for a specified country as calculated by RF modelling scenario 4, i.e. RF model fitted with '*base*' covariate set and WSF3D built area and built height layers (Appendix 2).

ppp_*XXX_RF5_v1.tif
Each grid cell of this dataset represents the estimated population count for a specified country as calculated by RF modelling scenario 5, i.e. RF model fitted with '*base*' covariate set and WSF3D built area, built height and volume layer (Appendix 2).

METHODS OVERVIEW

Pre-processing
Population counts and administrative boundaries were manually checked and harmonised to ensure matching records. Covariate layers were resampled to 3 arc-second resolution and nibbled (i.e. cells with no data are replaced with the values of the nearest neighbours) to match the provided administrative boundaries.

Random forest modelling
The population modelling approach used the Random Forest (RF)-based dasymetric mapping approach[2] implemented in the popRF 'R' package[3] based on Breiman's[4] algorithm.

For several countries, the available population count data do not have sufficient administrative units (>25 administrative units) to implement the RF methodology effectively. Therefore, in these cases the RF model was applied by grouping together countries or using another country to 'train' the RF-model (Appendix 1).

REFERENCES

1. Esch, T. *et al.* World Settlement Footprint 3D-A first three-dimensional survey of the global building stock. *Remote Sens. Environ.* **270,** 112877, https://doi.org/10.1016/j.rse.2021.112877 (2022).

2. Stevens, F. R., Gaughan, A. E., Linard, C. & Tatem, A. J. Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data. *PLoS ONE* **10,** e0107042, https://doi.org/10.1371/journal.pone.0107042 (2007).

3. Bondarenko, M., Nieves, J.J., Forrest, R.S., Andrea, E.G., Jochem, C., Kerr, D. & Sorichetta, A. popRF: Random Forest-informed Population Disaggregation R package, _Comprehensive R Archive Network (CRAN)_, https://cran.rproject.org/package=popRF (2021).

4. Breiman, L. Random forests. *Mach. Learn.* **45,** 5-32, https://doi.org/10.1023/A:1010933404324 (2001).

APPENDIX 1

Table of sourced population data used for random forest modelling.

ISO-3 country code	Country name	Total population	No. of units	Unit level	Modelled with	Dataset Year
ABW	Aruba	112,683	55	2		2020
AIA	Anguilla	13,572	9	0	Grouped Islands	2011
ARG	Argentina	45,808,456	525	2		2021
ATG	Antigua and Barbuda	84,816	8	1	Grouped Islands	2019
BHS	Bahamas	391,476	32	1	Grouped Islands	2019
BLZ	Belize	322,454	6	1	GTM	2020
BMU	Bermuda	63,779	11	2	Grouped Islands	2020
BOL	Bolivia	11,841,955	9	1	PER, PRY	2021
BRA	Brazil	211,755,692	5,570	2		2020
BRB	Barbados	226,193	11	1	Grouped Islands	2010
CHL	Chile	19,678,363	346	3		2021
COL	Colombia	50,372,424	1,122	2		2020
CRI	Costa Rica	5,163,021	478	3		2021
CUB	Cuba	11,193,470	168	2		2015
CUW	Curaçao	165,983	65	1	Grouped Islands	2020
CYM	Cayman Islands	69,914	6	1	Grouped Islands	2019
DMA	Dominica	69,325	10	1	Grouped Islands	2011
DOM	Dominican Republic	10,448,499	155	3		2020
ECU	Ecuador	17,510,643	25	1	COL	2020
GRD	Grenada	114,000	7	1	Grouped Islands	2020
GTM	Guatemala	17,109,746	340	2		2021
GUY	Guyana	756,237	10	1	SUR	2021
HND	Honduras	9,302,282	298	2		2020
JAM	Jamaica	2,697,983	14	1	DOM	2019
KNA	Saint Kitts and Nevis	46,325	14	1	Grouped Islands	2020
LCA	Saint Lucia	178,696	10	1	Grouped Islands	2018
MEX	Mexico	128,972,439	2,457	2		2021
MSR	Montserrat	4,566	39	EA	Grouped Islands	2021
NIC	Nicaragua	6,595,672	153	2		2020
PAN	Panama	4,337,406	13	1	HND, SLV	2021
PER	Peru	29,381,884	1,873	3		2021
PRY	Paraguay	7,252,669	250	2		2020

SLV	El Salvador	6,825,935	262	2		2021
SUR	Suriname	590,100	62	2		2018
TCA	Turks and Caicos Islands	31,458	6	1	Grouped Islands	2021
TTO	Trinidad and Tobago	1,328,022	15	1	Grouped Islands	2020
URY	Uruguay	3,530,912	19	1	COL, PRY	2020
VCT	Saint Vincent and the Grenadines	110,696	13	2	Grouped Islands	2018
VEN	Venezuela	32,605,423	25	1	PRY	2020
VGB	Virgin Islands (British)	32,670	25	0	Grouped Islands	2017

APPENDIX 2

List of covariates.

Covariate name	Description
*XXX_cls_11_dst.tif	Distance to ESA-CCI-LC cultivated terrestrial areas and managed lands 2018
*XXX_cls_40_dst.tif	Distance to ESA-CCI-LC natural and semi-natural terrestrial vegetation – Woody/Trees 2018
*XXX_cls_130_dst.tif	Distance to ESA-CCI-LC natural and semi-natural terrestrial vegetation – Shrubs 2018
*XXX_cls_140_dst.tif	Distance to ESA-CCI-LC natural and semi-natural terrestrial vegetation – Herbaceous 2018
*XXX_cls_150_dst.tif	Distance to ESA-CCI-LC natural and semi-natural sparse terrestrial vegetation 2018
*XXX_cls_160_dst.tif	Distance to ESA-CCI-LC natural and semi-natural aquatic vegetation 2018
*XXX_cls_190_dst.tif	Distance to ESA-CCI-LC artificial surfaces 2018
*XXX_cls_200_dst.tif	Distance to ESA-CCI-LC bare areas 2018
*XXX_InlandWater_dst.tif	Distance to ESA-CCI-LC inland waterbodies 2000-2012
*XXX_Coastline_WPGP.tif	Coastline 2000-2012
*XXX_elevation.tif	HydroSHEDS elevation 2006
*XXX_slope.tif	HydroSHEDS-based slope 2006
*XXX_UrbanArea_dst.tif	Distance to built footprint 2021
*XXX_pedestrian_distance.tif	Distance to OSM pedestrian road intersections 2021
*XXX_residential_distance.tif	Distance to OSM residential road intersections 2021
*XXX_vehicle_distance.tif	Distance to OSM vehicle road intersections 2021
*XXX_pedestrian_roads_dst.tif	Distance to OSM pedestrian roads 2021
*XXX_residential_roads_dst.tif	Distance to OSM residential roads 2021
*XXX_vehicle_roads_dst.tif	Distance to OSM vehicle roads 2021
*XXX_OSM_Waterways_WPGP.tif	Distance to OSM major waterways 2017
*XXX_Precipitation_WPGP.tif	Average annual total precipitation 1970-2000
*XXX_Temperature_WPGP.tif	Average annual temperature 1970-2000
*XXX_VIIRS.tif	VIIRS night-time lights 2020 (Visible Infrared Imaging Radiometer Suite)
*XXX_WDPA_WPGP.tif	Distance to IUCN strict nature reserve and wilderness area edges 2017
*XXX_road_length.tif	Road length 2021
*XXX_road_density.tif	Road density 2021
*XXX_road_int_0_5_km.tif	Road intensity (0.5km bandwidth) 2021
*XXX_road_int_1_km.tif	Road intensity (1km bandwidth) 2021
*XXX_road_int_2_km.tif	Road intensity (2km bandwidth) 2021
*XXX_road_int_5_km.tif	Road intensity (5km bandwidth) 2021
*XXX_WSF3D_BuildingArea_uncnd.tif	WSF3D built area 2021
*XXX_WSF3D_Height_uncnd.tif	WSF3D built height 2021
*XXX_Volume.tif	Built volume 2021

*Spatially referenced gridded covariate layers were constructed for 40 countries, identifiable using the ISO-3 country codes in the filenames in place of XXX above. For example, ABW refers to Aruba. See appendix 1 for the full list of countries in this dataset.

Note: covariates in grey not included in '*base*' set of covariates for random forest modelling.