Release Statement

Gridded population estimates for Sudan using UN COD-PS estimates 2022, version 2.0

02 May 2023

These data were produced by WorldPop at the University of Southampton. These data include gridded estimates of population at approximately 100m and 1km for 2022, along with estimates of the number of people belonging to individual age-sex groups. These results were produced using subnational population estimates for Sudan in 2022 provided in the Common Operational Dataset on Population Statistics (COD-PS [1]) and built-up surfaces/volumes/height covariates extracted from GHSL datasets; GHS-BUILT-Surface epoch 2020 layer [2], combined with Digitize Africa building footprints [14], were used to delineate settled areas. The constrained top-down disaggregation method was used to produce the datasets, i.e. population was only estimated within areas classified as containing built settlement. The modelling work and geospatial data processing was led by Bondarenko M. and Leasure D.R.. Oversight was provided by Tatem A.J.

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.

CITATION

Bondarenko M., Leasure D.R., and Tatem A.J. 2023 *Gridded population estimates for Sudan using UN COD-PS estimates 2022, version 2.0*. WorldPop, University of Southampton. doi:10.5258/SOTON/WP00761.

MAIN DATA SOURCES

- GHS-BUILT-S R2022A - GHS built-up surface grid, derived from Sentinel-2 composite and Landsat, multitemporal (1975-2030) [2].
- Subnational population estimates for Sudan in 2022 provided in the Common Operational Dataset on Population Statistics (COD-PS). It is an update to an earlier COD-PS released by UNFPA [1].
- Subnational Administrative Boundaries for Sudan provided by Regional IM Working Group - Europe [3].
- Geospatial covariate layers available at WorldPop [4].
- OpenStreetMap [5] [Data file from 2023-01-17]
- WorldCover 10m 2021 v200 [6]
- VIIRS Nighttime Day/Night Band Composites Version 1 [7]. 2022 time series was extracted from Image Collections “NOAA/VIIRS/DNB/MONTHLY_V1/VCMCFG” using Google Earth Engine (GEE) platform.

- TerraClimate: Monthly Climate and Climatic Water Balance for Global Terrestrial Surfaces, University of Idaho [8]. 2022 time series was extracted from Image Collections "IDAHO_EPSCOR/TERRACLIMATE" using GEE platform.
- MOD11A2.061 Terra Land Surface Temperature and Emissivity 8-Day Global 1km [9]. 2022 time series was extracted from Image Collections "MODIS/061/MOD11A2" using GEE platform.
- Ecopia.AI and Maxar Technologies. 2020/2021. Digitize Africa data 'year 2'. http://digitizeafrica.ai

RELEASE CONTENT

1. sdn_pop_2022_100m_constrained_v2.zip
2. sdn_pop_2022_100m_unconstrained_v2.zip
3. sdn_pop_2022_1km_constrained_v2.zip
4. sdn_pop_2022_1km_unconstrained_v2.zip
5. sdn_agesex_2022_100m_constrained_v2.zip
6. sdn_agesex_2022_100m_unconstrained_v2.zip
7. sdn_agesex_2022_1km_constrained_v2.zip
8. sdn_agesex_2022_1km_unconstrained_v2.zip
9. sdn_school_age_2022_1km_constrained_v2.zip
10. sdn_pop_adm2_2022.zip
11. sdn_school_age_adm2_2022.zip

FILE DESCRIPTIONS

The projection for all GIS files is the geographic coordinate system WGS84 (World Geodetic System 1984).

sdn_pop_2022_100m_constrained_v2.zip

This geotiff raster, at a spatial resolution of 3 arc-seconds (approximately 100m at the equator), contains estimates of total population size per grid cell across Sudan. NA values represent areas that were mapped as unsettled based on the GHS settlement layer [2] and Digitize Africa layer [14]. These data are stored as floating-point numbers rather than integers to avoid rounding errors in aggregated populations for larger areas.

sdn_pop_2022_100m_unconstrained_v2.zip

This geotiff raster, at a spatial resolution of 3 arc-seconds (approximately 100m at the equator), contains estimates of total population size per grid cell across Sudan. NA values represent areas that were mapped as water. These data are stored as floating-point numbers rather than integers to avoid rounding errors in aggregated populations for larger areas.

sdn_pop_2022_1km_constrained_v2.zip

This geotiff raster, at a spatial resolution of 30 arc-seconds (approximately 1km at the equator), contains estimates of total population size per grid cell across Sudan. NA values represent areas that were mapped as unsettled based on the GHS settlement layer [2]. The dataset was produced by aggregating *sdn_pop_2022_100m_constrained_v1.tif* dataset to 1km.

sdn_pop_2022_1km_unconstrained_v2.zip

This geotiff raster, at a spatial resolution of 30 arc-seconds (approximately 1km at the equator), contains estimates of total population size per grid cell across Sudan. NA values represent areas that were mapped as water. The dataset was produced by aggregating *sdn_pop_2022_100m_unconstrained_v1_0.tif* dataset to 1km.

sdn_agesex_2022_100m_constrained_v2.zip
This zip file contains raster files in geotiff format at a spatial resolution of 3 arc-seconds (approximately 100m at the equator). Each raster provides gridded population estimates for an age-sex group of settled areas (NA represent unsettled areas). Files are labelled with either an “M” (male) or an “F” (female) followed by the age-range of the group (five year bins). For instance, “F_00_04” and “M_05_09” are population counts of under 5 year olds for females and between 5 and 9 years old for males, respectively. Eighty year olds and over are represented in the groups “F_80Plus” and “M_80Plus”. These data were produced using age-sex national proportions from COD-PS [1]. The age-sex proportions were applied to the gridded population estimates (*sdn_pop_2022_100m_constrained_v1*) to allocate the population to the different age-sex classes. While this data represents population counts, values contain decimals, i.e. fractions of people. This is because we do not estimate which grid cell each individual in a given age group occupies and are maintaining the UN COD-PS numbers. For this reason, it is advised to aggregate the rasters at a coarser scale. For example, if four grid cells next to each other have values of 0.25 this indicates that there is estimated to be 1 person of that age group somewhere in those four grid cells.

sdn_agesex_2022_100m_uconstrained_v2.zip
This zip file contains raster files in geotiff format at a spatial resolution of 3 arc-seconds (approximately 100m at the equator). Each raster provides gridded population estimates for an age-sex group of settled areas (NA represent surface waters). Files are labelled with either an “M” (male) or an “F” (female) followed by the age-range of the group (five year bins). For instance, “F_00_04” and “M_05_09” are population counts of under 5 year olds for females and between 5 and 9 years old for males, respectively. Eighty year olds and over are represented in the groups “F_80Plus” and “M_80Plus”. These data were produced using age-sex national proportions from COD-PS [1]. The age-sex proportions were applied to the gridded population estimates (*sdn_pop_2022_100m_uconstrained_v1*) to allocate the population to the different age-sex classes. While this data represents population counts, values contain decimals, i.e. fractions of people. This is because we do not estimate which grid cell each individual in a given age group occupies. For this reason, it is advised to aggregate the rasters at a coarser scale. For example, if four grid cells next to each other have values of 0.25 this indicates that there is 1 person of that age group somewhere in those four grid cells.

sdn_agesex_2022_1km_constrained_v2.zip
This zip file contains rasters in geotiff format at a spatial resolution of 30 arc-seconds (approximately 1km at the equator). Each raster provides gridded population estimates for an age-sex group of settled areas (NA represent unsettled areas). These datasets were produced by aggregating *sdn_agesex_2022_100m_constrained_v1_0* datasets respectively to 1km

sdn_agesex_2022_1km_uconstrained _v2.zip
This zip file contains raster files in geotiff format at a spatial resolution of 30 arc-seconds (approximately 1km at the equator). Each raster provides gridded population estimates for an age-sex group across Sudan (NA represent surface waters). These datasets were produced by aggregating *sdn_agesex_2022_100m_uconstrained_v1* datasets respectively to 1km.

sdn_school_age_2022_1km_constrained_v2.zip
This zip file contains rasters in geotiff format at a spatial resolution of 30 arc-seconds (approximately 1km at the equator). Each raster provides gridded population estimates for children, broken down in differing school attendance age groupings (0 to 5 years old, 0 to 17 years old, 6 to 13 years old and 14 to 17 years old) for settled areas (NA represent unsettled areas). Files are labelled with either an “M” (male) or an “F” (female) or “F_M” for both male and female. These data were created by applying Sprague multipliers to “sdn_agesex_2022_1km_constrained_v2 “ raster data. Sprague multipliers are used to interpolate data and obtain population estimates by single years of age and these are then aggregated to obtain differing age groupings.

sdn_pop_adm2_2022.zip
This zip file contains the zonal statistics of total population calculated using “sdn_agesex_2022_100m_constrained_v2” and level 2 of Subnational Administrative Boundaries for Sudan provided by OCHA on HDX [15].

sdn_school_age_adm2_2022.zip
This zip file contains the zonal statistics of total population calculated using "sdn_school_age_2022_1km_constrained_v2" and level 2 of Subnational Administrative Boundaries for Sudan provided by OCHA on HDX [15].

METHODS

Pre-processing: Subnational Administrative Boundaries provided by OCHA [3] were nibbled (i.e. cells with no data are replaced with the values of the nearest neighbors) to match the WorldPop mastergrid and avoid mismatch with the WorldPop covariates.

Modelling: Building area/volume per pixel were extracted from GHS in addition to classifying pixels as residential or non-residential. The population modelling described below used the Random Forest (RF)-based dasymetric mapping approach (Stevens et al., 2015 [10]) implemented in the popRF 'R' package [11] based on the Breiman (2001) [12] algorithm.

The UN COD-PS [1] population projections do not have a sufficient number of admin units (18 regular admin 1 units) to apply the RF methodology well. Therefore admin 3 (130 admin units) projected 2020 population data from WorldPop [13] was used to train the RF model and produce an unconstrained prediction weighting layer, where all non-residential pixels were given a value of zero in the built-up covariate inputs (APPENDIX 2). An assessment of which of the ancillary data covariates were important for accurately estimating population density at the census unit level can be seen in the APPENDIX 3. The model could explain 81.7% of the population input variance (i.e. the admin 3 projected population data from WorldPop [13]). The unsettled areas of this unconstrained weighting layer were then removed by using the GHS-BUILT-Surface epoch 2020 layer containing all built-up surfaces (residential and non-residential) as a mask, thus creating a constrained weighting layer. Finally, this constrained prediction weighting layer was used for dasymetric redistribution of the UN COD-PS [1] population estimates 2022 (APPENDIX 1). The gridded population estimates were then combined with the COD-PS [1] age/sex pyramid table for Ukraine to produce gridded population estimates for females and males at regular age intervals.

WORK CITED

1. UNFPA Sudan - Subnational Population Statistics https://data.humdata.org/dataset/cod-ps-sdn
2. Pesaresi, Martino; Politis, Panagiotis (2022): GHS-BUILT-S R2022A - GHS built-up surface grid, derived from Sentinel-2 composite and Landsat, multitemporal (1975-2030). European Commission, Joint Research Centre (JRC) [Dataset] DOI:10.2905/D07D81B4-7680-4D28-B896-583745C27085 PID:http://data.europa.eu/89h/d07d81b4-7680-4d28-b896-583745c27085
3. Subnational Administrative Boundaries provided by OCHA. Sudan administrative level 1 (country) https://data.humdata.org/dataset/cod-ab-sdn
4. WorldPop (www.worldpop.org - School of Geography and Environmental Science, University of Southampton; Department of Geography and Geosciences, University of Louisville; Departement de Geographie, Universite de Namur) and Center for International Earth Science Information Network (CIESIN), Columbia University (2018). Global High Resolution Population Denominators Project - Funded by the Bill and Melinda Gates Foundation (OPP1134076).https://dx.doi.org/10.5258/SOTON/WP00644
5. OpenStreetMap contributors. (2015) Planet dump [Data file from 2023-01-17]. Retrieved from https://download.geofabrik.de/.

6. Zanaga, D., Van De Kerchove, R., Daems, D., De Keersmaecker, W., Brockmann, C., Kirches, G., Wevers, J., Cartus, O., Santoro, M., Fritz, S., Lesiv, M., Herold, M., Tsendbazar, N.E., Xu, P., Ramoino, F., Arino, O., 2022. ESA WorldCover 10 m 2021 v200. https://doi.org/10.5281/zenodo.7254221
7. C. D. Elvidge, K. Baugh, M. Zhizhin, F. C. Hsu, and T. Ghosh, "VIIRS night-time lights," International Journal of Remote Sensing, vol. 38, pp. 5860–5879, 2017
8. Abatzoglou, J.T., S.Z. Dobrowski, S.A. Parks, K.C. Hegewisch, 2018, Terraclimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958-2015, Scientific Data 5:170191, doi:10.1038/sdata.2017.191
9. ASTER Mount Gariwang image from 2018 was retrieved on 2023-02-01 from https://lpdaac.usgs.gov, maintained by the NASA EOSDIS Land Processes Distributed Active Archive Center (LP DAAC) at the USGS Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota. 2018, https://lpdaac.usgs.gov/resources/data-action/aster-ultimate-2018-winter-olympics-observer/.
10. Stevens FR, Gaughan AE, Linard C, Tatem AJ (2015) Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data. PLoS ONE 10(2): e0107042. https://doi.org/10.1371/journal.pone.0107042
11. Bondarenko M., Nieves J.J., Forrest R.S., Andrea E.G., Jochem C., Kerr D., and Sorichetta A. (2021): popRF: Random Forest-informed Population Disaggregation R package, _Comprehensive R Archive Network (CRAN)_,url:https://cran.rproject.org/package=popRF.
12. Breiman L. Random Forests. Mach Learn. 2001 [cited 2013 Jan 24].45(1):5–32. Available: http://link.springer.com/article/10.1023/A:1010933404324
13. WorldPop (www.worldpop.org - School of Geography and Environmental Science, University of Southampton; Department of Geography and Geosciences, University of Louisville; Departement de Geographie, Universite de Namur) and Center for International Earth Science Information Network (CIESIN), Columbia University. 2018b. Global High Resolution Population Denominators Project - Funded by The Bill and Melinda Gates Foundation (OPP1134076). Available from https://data.worldpop.org/GIS/Population/Global_2000_2020/CensusTables/
14. Ecopia.AI and Maxar Technologies. 2020/2021. Digitize Africa data 'year 2'. http://digitizeafrica.ai
15. Subnational Administrative Boundaries provided by OCHA. Sudan administrative level 2 (country) https://data.humdata.org/dataset/cod-ab-sdn

APPENDIX 1

Table of *ADM1_PCODE* Subnational Administrative Boundaries for Sudan provided by OCHA. Source https://data.humdata.org/dataset/cod-ps-sdn

ISO3	ADM1_EN	ADM1_PCODE
SDN	Khartoum	SD01
SDN	North Darfur	SD02
SDN	South Darfur	SD03
SDN	West Darfur	SD04

SDN	East Darfur	SD05
SDN	Central Darfur	SD06
SDN	South Kordofan	SD07
SDN	Blue Nile	SD08
SDN	White Nile	SD09
SDN	Red Sea	SD10
SDN	Kassala	SD11
SDN	Gedaref	SD12
SDN	North Kordofan	SD13
SDN	Sennar	SD14
SDN	Aj Jazirah	SD15
SDN	River Nile	SD16
SDN	Northern	SD17
SDN	West Kordofan	SD18

APPENDIX 2
List of covariates.

Name of the covariate	Description
sdn_grid_100m_ccilc_dst010_2021.tif	Distance to ESA-CCI-LC Tree cover 2021
sdn_grid_100m_ccilc_dst020_2021.tif	Distance to ESA-CCI-LC Shrubland 2021
sdn_grid_100m_ccilc_dst030_2021.tif	Distance to ESA-CCI-LC Grassland 2021
sdn_grid_100m_ccilc_dst040_2021.tif	Distance to ESA-CCI-LC Cropland 2021
sdn_grid_100m_ccilc_dst060_2021.tif	Distance to ESA-CCI-LC Bare/ sparse vegetation 2021
sdn_grid_100m_ccilc_dst080_2021.tif	Distance to ESA-CCI-LC Permanent water bodies 2021
coastline_dst	Distance to coastline 2000-2020
OSM_intersections_dst	Distance to OSM major road intersections 2022
OSM_Water_dst	Distance to OSM major waterways 2022
OSM_Highway_dst	Distance to OSM major roads 2022
px_area	Grid-cell surface areas
slope	SRTM-based slope 2000 (SRTM is Shuttle Radar Topography Mission)
elevation	SRTM elevation 2000
nighttime_2022	Night-time lights 2022
Precipitation	Current average annual total precipitation 2022
Temperature	Current average annual temperature 2022
wdpa_cat1_dst_2017	Distance to IUCN strict nature reserve and wilderness area edges 2017
builtup_dist	Distance to settlement GHSL 2022

Builtup_Surfaces	GHSL built surfaces 2022
Builtup_Volume	GHSL built volume 2022
buildings_count	Counts of buildings that fall within a grid cell (Ecopia/Maxar [14])
buildings_density	Contains a measure of the number of buildings per grid cell area in square kilometres (Ecopia/Maxar [14])
buildings_total_area	Grid cell level sum of the building areas for all buildings whose centroid falls inside a grid cell (Ecopia/Maxar [14])
buildings_cv_area	Grid cell level coefficient of variation of building areas (Ecopia/Maxar [14])
buildings_total_length	Grid cell level sum of the building lengths for all buildings whose centroid falls inside a grid cell (Ecopia/Maxar [14])
buildings_cv_length	Grid cell level coefficient of variation of building lengths for all buildings whose centroid falls inside a grid cell (Ecopia/Maxar [14])

APPENDIX 3. Variable importance.

Ranking	**Covariate**	**%IncMSE**	**IncNodePurity**
1	Builtup_Volume	17.555278	56.006218
2	Builtup_Surfaces	14.950881	47.125647
3	cls_40_dst	11.321297	17.623741
4	cls_10_dst	9.481443	13.61789
5	BUILT_RES_dst	9.32451	11.182997
6	VIIRS	9.075982	18.715843
7	buildings_count	7.910745	20.242075
8	buildings_density	7.385651	15.652381
9	cls_60_dst	6.967058	12.163586
10	buildings_cv_dist	6.788468	14.602284
11	OSM_intersections_dst	6.715827	8.673185
12	coastline_dst	6.486605	3.515611
13	buildings_total_length	6.027937	15.478842
14	cls_20_dst	5.976739	7.446475

15	OSM_Highway_dst	5.708734	6.838846
16	buildings_total_area	5.554324	9.367438
17	cls_30_dst	5.170394	5.757967
18	buildings_cv_length	4.974521	5.192672
19	buildings_cv_area	4.342252	5.544589
20	cls_95_dst	3.659339	3.150699
21	Temperature	3.421206	3.019514
22	OSM_Water_dst	3.036027	5.244157
23	cls_90_dst	1.951918	4.130857
24	Precipitation	1.945718	2.161117
25	slope	1.82802	2.721751
26	elevation	-1.236691	2.006691