Release Statement

South Sudan 2019 gridded population estimates from census projections adjusted for displacement, version 1.0

26 March 2020

Original Release: 6 December 2019

These data were produced by the WorldPop Research Group at the University of Southampton. This work was funded by the Bill and Melinda Gates Foundation (BMGF) and the United Kingdom's Department for International Development (OPP1182408). The primary intended use of these data was aiding the BMGF field teams. The modelling work was led by Claire Dooley with support from Chris Jochem and oversight by WorldPop director Andy Tatem and GRID3 lead Attila Lazar. The support of the whole WorldPop group is acknowledged, as well as the our GRID3 partners (UNFPA, Columbia University and Flowminder). We thank the teams at IOM and ACLED for their excellent work in collecting data and making it freely available. This work was supported with funding from the Bill & Melinda Gates Foundation (BMGF) and the United Kingdom's Department for International Development (DFID).

DATA DESCRIPTION

This dataset provides population estimates for each settled 100m grid square in South Sudan. The grid square values were derived using the National Bureau of Statistics' 2019 population projection estimates that were adjusted to account for displacement of people. The locations people have been displaced to were directly obtained from IOM's Displacement Tracking Matrix (DTM). The locations people have been displaced from were derived using DTM and the Armed Conflict Locations and Events Database (ACLED). Numbers of displaced people per location were calculated using recorded numbers of international refugees and internally displaced persons.

CITATION

WorldPop (School of Geography and Environmental Science, University of Southampton). 2020. South Sudan 2019 gridded population estimates from census projections adjusted for displacement, version 1.0. https://dx.doi.org/10.5258/SOTON/WP00659

RELEASE CONTENT

1. SSD_population_v1_0_gridded.zip

FILE DESCRIPTIONS

SSD_population_v1_0_gridded.zip

This zip file contains two files:

SSD_population_v1_0_gridded.tif

This geotiff raster contains estimates of total population size for each approximately 100 m grid cell (0.0008333 decimal degrees grid) across South Sudan. The values were derived from the National Bureau of Statistics' 2019 population projection estimates adjusted to account for recorded displacement of people. Note: This raster is accompanied by one ancillary file that contains metadata (SSD_population_v1_0_gridded.tif.xml).

SSD_population _v1_0_ stateIDs.tif

This geotiff raster contains state IDs for each populated pixel in SSD_population_v1_0_gridded.tif. States are Central Equatoria (ID = 101), Eastern Equatoria (102), Jonglei (103), Lakes (104), Northern Bahr el Ghazal (105), Unity (106), Upper Nile (107), Warrap (108), Western Bahr el Ghazal (109), Western Equatoria (110). Note: This raster is accompanied by one ancillary file that contains metadata (SSD_population_v1_0_stateIDs.tif.xml).

RELEASE HISTORY

Version 1.0 (6 December 2019)

- Original release of the South Sudan adjusted population projections dataset

ASSUMPTIONS AND LIMITATIONS

Limited data on fertility and mortality exists. This makes it difficult to evaluate the accuracy of NBS's population projection or to produce alternative projected estimates. The main goal of our population dataset is to provide better spatial distribution of population than currently exists and so the uncertainty in national and state level totals will hopefully not impact the use of our dataset.

The IOM-DTM data covered 444 out of the total 470 payams, therefore IDPs present in the payams not assessed in DTM round 5 are not considered. From the initial release of DTM's round 6 data these payams seem to have been covered. Once more detailed data for round 6 are released we will be able to include them in our analysis and update the population estimates accordingly.

The IOM-DTM GPS locations may not always be accurate. While they usually correspond to a region below payam level, e.g. town, hamlet or village, this may not always be the case. Where this issue exists the IDPs may be concentrated in the incorrect pixels. However, if these population estimates are aggregated for larger spatial units the total number should include the IDPs correctly and should not be impacted by GPS dislocation.

It is likely that the ACLED database misses conflict events. If ACLED consistently excludes conflict that causes displacement in certain locations or regions this will impact the estimated number of people displaced from that location, and ultimately result in the final population estimates being potentially higher than in reality.

Users should note that these population estimates do not consider seasonal migration which has a potentially large impact on population distribution during the flooding season depending on the extent of flooding.

SOURCE DATA

The key datasets used to produce the population raster are:

- NBS 2019 state-level population projections. These were calculated by NBS using the state-level 2008 census data and fertility and mortality rates. We use NBS's 2019 state-level totals found in tables 8, 15, 22, 29, 36, 43, 50, 57, 64, 71 of the NBS (2015) Population Projections document [1]. Note that these state totals sum to 12,732,455, which differs slightly from the 2019 national total of 12,778,250 reported in table 1 of the same document. We assume this difference comes from the national total being calculated using national averages of fertility and mortality rates.
- UNHCR's South Sudanese total refugees [2]. Total South Sudanese Refugees reported as of 31st October 2019 was 2,204,277, located in Central African Republic (2,218), Democratic Republic of Congo (98,691), Ethiopia (318,909), Kenya (119,445), Sudan (810,155) and Uganda (854,859).
- IOM's Displacement Tracking Matrix round 5 [7]. This data included estimated number of IDPs per location of assessment along with other attributes such as state and county of origin. The total number of IDPs reported was 1,420,189. The following two excel spreadsheets were used:
    - "IOM DTM Mobility Tracking Round 5 Baseline Locations 20190610_with comparison R4 and R5_0" excel file, 'Locations R5' tab.
    - "IOM DTM Mobility Tracking Round 5_Baseline dataset_complete" excel file, 'MT_R5_Dataset_Baseline_complete' tab.
- ACLED conflict location data [9]. This data included all publicly reported conflict between January 2011 and October 2019. We used location, year and number of

fatalities to create the geospatial covariates used in the disaggregation of people displaced from conflict zone (step 4 of method).

- Building Footprints for South Sudan [8]. The building footprints data was processed to produce ~100m x 100m binary layer of settled area.
- Geospatial covariates, processed for South Sudan [10, 11, 12]. These were distance to water, distance to built-up areas, distance to road, slope, elevation and enhanced vegetation index.

METHODS OVERVIEW

We carried out a multi-step approach for producing these population estimates. The main steps were:

1) Adjusting the 2019 projected state-level population estimates calculated by NBS [1] to account for displacement between states and into neighbouring countries. We did this using the International Organisation for Migration's Displacement Tracking Matrix (IOM-DTM) Round 5 Mobility Tracking data released in June 2019 [7] and the United Nations High Commissioner for Refugees' (UNHCR) data that reports number of South Sudanese refugees as of October 2019 [2,3]. See table in appendix for adjusted state totals.
2) Disaggregation of these state totals into the ~100m x 100m pixels that are 'settled' according to the building footprints dataset [8]. Disaggregation was carried out using random forest methodology and a set of geospatial covariates commonly used to disaggregate population density [4, 10, 11, 12]. In this step, disaggregation did not consider variables relating to conflict. Instead, we created an adjustment layer (steps 3-5) to modify this initial state-level disaggregation layer.
3) Creation of a spatial layer for current locations of Internally Displaced Persons (IDPs). To do this, we used GPS locations of IDPs included in the IOM-DTM round 5 data. We created a 1km buffer around each point and split up each location's total number of IDPs equally across all settled pixels within the buffer. One addition step was taken to manually add the locations of three camps (Ajio, Kerwa and Logo) in Liwolo payam (in Kajo-Keji county) as the three were listed as one record with one GPS location in the DTM data. The estimated total number of IDPs for the three camps was split between the three camps according to their relative sizes reported in 2017 by IOM [5].
4) Creation of a spatial layer for original location of currently displaced people. For origin locations of internally displaced persons we used the IOM-DTM round 5 data which includes partial origin data at county level. For origin of international refugees we used the proportion of current returnees returning to each county and estimated the county of origin for current refugees by multiplying these county proportions by the total number of current international refugees. These numbers of refugees per

county of origin will potentially include large error if there are still certain areas that refugees are not returning to because of insecurity and/or other reasons. While this is almost certainly the case, our approach is objective and based on the best available data. To create the spatial layer for number of currently displaced people per origin location, we disaggregated the number of currently displaced people per county of origin across pixels within counties using the random forest methodology. The geospatial covariates used in this model all related to conflict locations reported in the Armed Conflict Locations and Events database [9] - examples include frequency of conflict and distance to conflict over set time periods.

5) Derivation of an adjustment layer from the layers created in steps 3) and 4) that provides a measure of each pixel's relative impact from conflict compared to all other locations (pixels) within the same state.
6) Applying the adjustment layer (step 5) to the initial disaggregated totals (step 2) to produce a final adjusted population layer.

All data processing and analysis was carried out in R [6]. A list of the geospatial covariates used in the two disaggregation steps are available upon request.

WORKS CITED

[1] National Bureau of Statistics report, 2015. "Population Projections, South Sudan. From 2015 – 2020."

[2] UNHCR report, 2019. "Regional overview of South Sudanese refugee population". Downloaded from https://data2.unhcr.org/en/dataviz/62?sv=5&geo=0

[3] UNHCR dashboard, 2019. "South Sudan Situation: Population Dashboard" (as of 31 October 2019), accessed 10th November 2019. https://data2.unhcr.org/en/documents/details/72226

[4] Stevens F.R., Gaughan A.E., Linard C., Tatem A.J. 2015. "Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data." PLOS ONE 10 (2). Public Library of Science: 1–22.

[5] IOM Displacement Tracking Matrix report, 2017. "Kajo Keji, Central Equatoria". https://reliefweb.int/report/south-sudan/dtm-south-sudan-kajo-keji-central-equatoria-paper-registration-rapid-intentions

[6] R Core Team 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

DATA CITED

[7] IOM Displacement Tracking Matrix Round 5, 2019. Data downloaded from https://displacement.iom.int on 1st November 2019.

[8] Building Footprints South Sudan, Digitize Africa data © 2020 Maxar Technologies, Ecopia.AI

[9] Raleigh, Clionadh, Andrew Linke, Håvard Hegre and Joakim Karlsen, 2010. “Introducing ACLED-Armed Conflict Location and Event Data.” Journal of Peace Research 47(5) 651-660. Data downloaded from https://www.acleddata.com/data/ on 1st November 2019.

[10] Lloyd, C.T., Chamberlain, H., Kerr, D., Yetman, G., Pistolesi, L., Stevens, F.R., Gaughan, A.E., Nieves, J.J., Hornby, G., MacManus, K., Sinha, P., Bondarenko, M., Sorichetta, A. & Tatem, A.J. 2019. "Global spatio-temporally harmonised datasets for producing high-resolution gridded population distribution datasets." Big Earth Data 3 (2), 108-139, DOI: 10.1080/20964471.2019.1625151

[11] Didan, K. 2015. MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006 [Data set]. NASA EOSDIS Land Processes DAAC. https://doi.org/10.5067/MODIS/MOD13Q1.006

[12] Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. 2017. "Google Earth Engine: Planetary-scale geospatial analysis for everyone." Remote Sensing of Environment. https://doi.org/10.1016/j.rse.2017.06.031

Appendix

State	NBS 2019 population projection estimate	Adjusted 2019 population projection estimate
Upper Nile	1,494,349	959,566
Jonglei	1,999,331	1,671,141
Unity	950,922	851,194
Warrap	1,480,426	1,432,722
Northern Bahr el Ghazal	1,095,051	618,722
Western Bahr el Ghazal	508,633	497,488
Lakes	1,147,187	1,197,061
Western Equatoria	847,569	743,548
Central Equatoria	1,690,506	1,555,528
Eastern Equatoria	1,518,481	1,001,210
Total inside South Sudan	**12,732,455**	**10,528,180**
Outside country	-	2,204,277
Total	12,732,455	12,732,457