Generating multi-period crop distribution maps for Southern Africa using a data fusion approach - Supplementary Information

Michiel van Dijk[1,2,*] Ulrike Wood-Sichra[3] Yating Ru[3] Amanda Palazzo[2]
Petr Havlik[2] Liangzhi You[3,4,*]

januari 04, 2023

[1] Wageningen Economic Research, the Hague, the Netherlands

[2] International Institute for Applied Systems Analysis, Laxenburg, Austria

[3] International Food Policy Research Institute, Washington DC, USA

[4] Macro Agriculture Research Institute, College of Economics and Management, Huazhong Agricultural University, Wuhan, China

[*] Correspondence: Michiel van Dijk <michiel.van.dijk@wur.nl>, Liangzhi You <l.you@cgiar.org>

S1 data

Crop statistics

Information on harvested area by crop at the subnational level is of key importance and greatly improves the quality of the gridded maps (Joglekar et al., 2019). Our main source of information was the global database with subnational agricultural statistics that was prepared to inform the the Spatial Production Allocation Model (SPAM, www.mapspam.info) for 2000 (SPAM2000) and 2010 (SPAM2010). The database contains production statistics (e.g. yield, harvested area and production) at three administrative area levels (ADMs): national level (ADM0), state/provincial level (ADM1) and district/department level (ADM2). For each year the subnational statistics were scaled to the FAOSTAT average national totals using a three year average (i.e. 1999-2001 and 2009-2011) to account for annual volatility. To convert harvested area into physical area, SPAM also includes crop, country and production system specific information on cropping intensity. Detailed descriptions about the sources and construction of the database can be found in You et al. (2014), Wood-Sichra et al. (2016) and Yu et al. (2020). In a few cases, we updated the SPAM subnational agricultural statistics database with recent data from AQUASTAT (FAO, 2016) on irrigated area and information derived from OpenStreetMap.

The databases that were constructed for SPAM2000 and SPAM2010 are not always directly comparable due to (1) differences in the number of crops (and crop groups) covered - SPAM2000 includes 20 crops, while the number was expanded to 42 in SPAM2010 - and (2) changes in the boundaries and definition of the ADMs over time. Fortunately, both versions of the database distinguish between the same four production systems (subsistence, rainfed low-input, rainfed high-input and irrigated).

To construct a consistent dataset with subnational statistics that can be compared over time, we started by harmonizing the crop definitions for SPAM2000 and SPAM2010, using the database for 2010 as basis. 13 out of the 20 crops in SPAM2000 could be directly linked to the crops in SPAM2010, 4 crops could be split into more detailed crops and for 3 crops we used national information from FAOSTAT. Similarly, we also harmonized the definition of the ADMs over time. It appeared that for a number of countries, there was a considerable change in the definition of the ADM2s while the definition of ADM1s was mostly stable. For this reason, we decided to only use the ADM1 statistics for the year 2000. Table S1 compares the sub-national coverage for the seven South African countries and the two relevant years and Figure S1 depicts the locations of the ADMs in the region.

Table S1: Coverage of subnational statistics by region. Crop coverage refers to the proportion of crops for which data (zero or positive values) at a certain administrative level is available. Harvest area coverage denotes the percentage of national harvested area from FAOSTAT covered by ADM1 and ADM2 statistics. For several countries and years there was no subnational information. For Angola, Mozambique and Zimbabwe, most information at the ADM2 level indicates crops were not grown in certain areas, resulting in a large crop coverage share but zero or low harvested area coverage share.

		ADM1			ADM2		
Country	Harvested area (1000 ha)	Number of ADMs	Crop coverage (%)	Harvested area coverage (%)	Number of ADMs	Crop coverage (%)	Harvested area coverage (%)
2000							
Angola	1.819	18	50	91	0	0	0
Botswana	133	10	31	50	0	0	0
Mozambique	3.936	10	46	82	0	0	0
Malawi	2.960	0	0	0	0	0	0
Namibia	363	13	43	86	0	0	0
Zambia	1.295	9	58	79	0	0	0
Zimbabwe	2.605	10	50	79	0	0	0
2010							
Angola	4.061	18	100	90	159	77	0
Botswana	194	10	50	81	0	0	0
Mozambique	6.745	12	89	97	146	89	1
Malawi	3.827	3	79	95	27	69	84
Namibia	334	13	46	84	97	38	0
Zambia	2.105	9	100	100	72	100	83
Zimbabwe	2.932	10	80	99	61	80	1

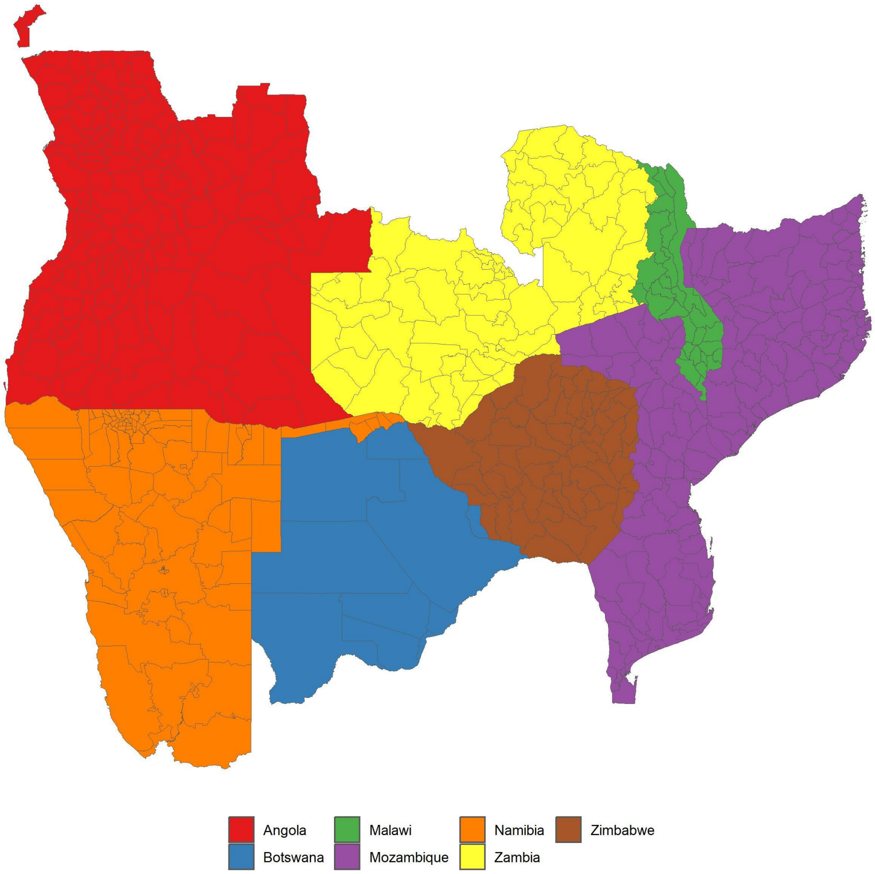


Figure S1: Administrative areas (ADMs) in Southern Africa for 2010. For all countries except Botswana, ADM level 2 statistics are presented.

Cropland extent

We combined several sources to determine the cropland extent in 2000 and 2010. For 2010, we used data from Lu et al. (2020) that was used to produce a global synergy cropland map at a resolution of 500 m x 500 m. They developed the Self-adapting statistics allocation model (SASAM) to combine and rank five different global cropland products: GlobeLand30, CCI-LC, GlobCover 2009, MODIS C5, and the Unified Cropland Layer. After harmonization of cropland classes, resolution and projection, cropland area statistics from FAOSTAT were used to rank the cropland maps and construct a scoring table that reflects the agreement among the datasets.

We used three global layers produced by SASAM: (1) a map with the median cropland per grid cell, (2) a map with the maximum cropland per grid cell, and (3) a map with the scores for each grid cell. To create the cropland extent we aggregated the maps to 30 arcsec and selected the grid cells with the highest score from the median cropland map until the subnational and national cumulative cropland area were equal or slightly larger than the statistics. In case the accumulated cropland was not sufficient to allocate the statistics, the maximum cropland layer was used (for details about the procedure see Wood-Sichra et al. (2016)).

To create the cropland extent for 2000, we followed a comparable approach as was used to construct the global synergy cropland map for 2010. Only three global products with cropland information were available for 2000: GLC2000, ESA-CCI-LC and GlobeLand30 (see Table 1 main text). As described in the main text, we also included the cropland extent for 2010 that can be derived by aggregating the crop area in all 2010 crop distribution maps. We combined the three maps for 2000 and the cropland extent for 2010 to create the median, maximum and score maps for 2000. We considered the following order of increasing importance when creating the scoring table: (1) ESA-CCI-LC, (2) GlobeLand30, (3) 2010 cropland extent and (4) GLC2000. The ESA-CGI-LC received the highest score because it was the only product with consistent information for both 2000 and 2010. GlobeLand30 was second because it has a much higher resolution (30 x 30 m) than GLC2000 (1 x 1 km) and can be considered somewhat outdated. For the same reason, we preferred the 2010 cropland extent over GLC2000. Grid cells that show cropland in all four layers receive the highest score. Figure S2 compares the cropland extent between 2000 and 2010.

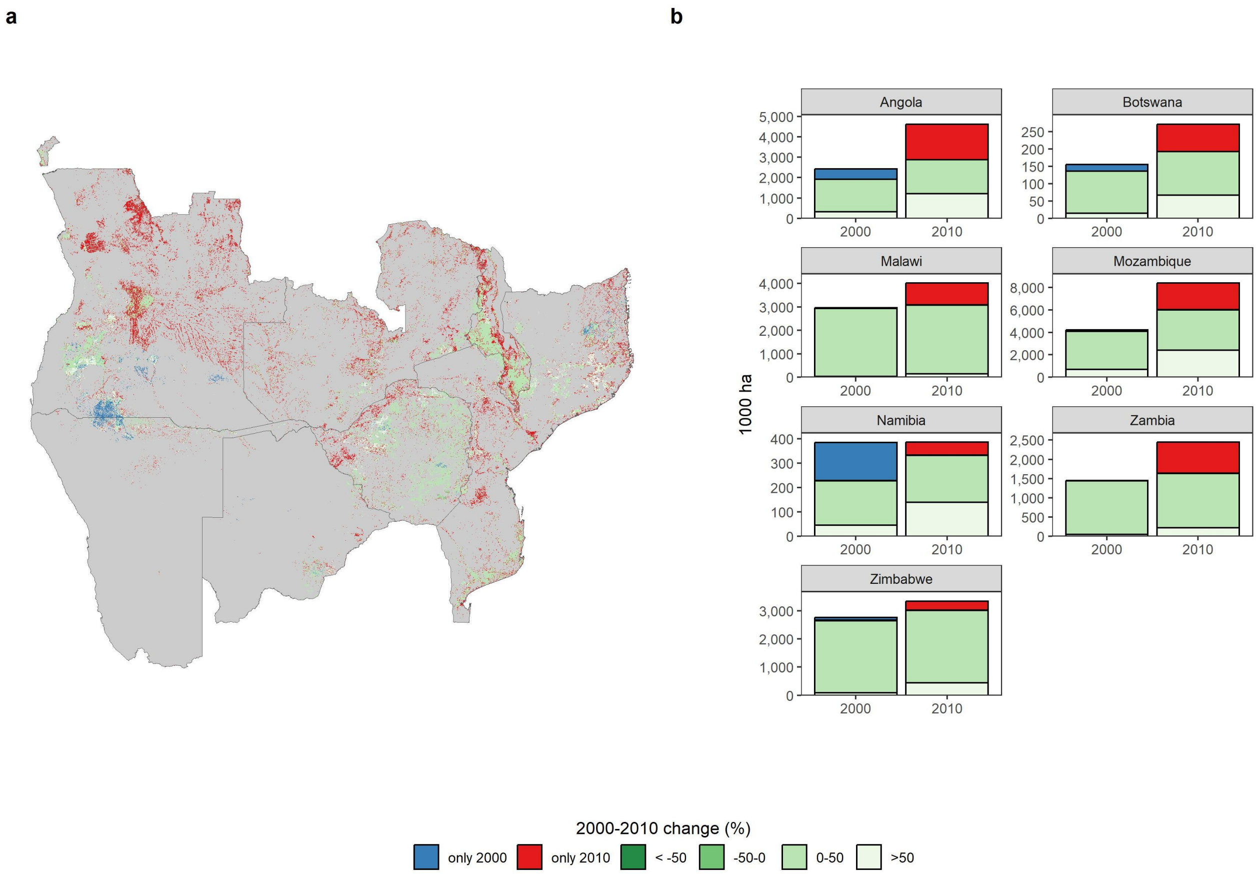


Figure S2: Comparison of the cropland exent between 2000 and 2010. a Cropland exent used to allocate crop statistics for 2000 and 2010. b Comparison of cropland between countries and over time. Differences between periods are caused by changes in crop area within grid cells (green bars) and changes in the number of cropland grid cells that are identified in each period (red and blue bars).

Irrigated area

To allocate the irrigated crops, we created a new irrigated area map by combining a number of products. As main source of information, we used the Global Irrigated Areas (GIA) map (Meier et al., 2018), which depicts irrigated areas for the world around the period 2005. The GIA combines normalized difference vegetation index (NDVI) maps, crop suitability data and information on the location of areas equipped for irrigation to create an irrigated area map with a resolution of 30 arcsec. In comparison to the Global Map of Irrigated Areas (GMIA) (Siebert et al., 2013), which has a resolution of 5 arcmin and is used as irrigation layer by SPAM, the GMIA shows 18% more irrigated area globally.

To create irrigation maps for 2000 and 2010, we combined and extended the GIA using two additional sources of information. First, we collected additional information on the location of irrigation schemes from regional irrigation surveys. World Bank (2010) provides an in-depth analysis of irrigated areas in the Zambezi basin, which overlaps with the seven target countries of our study. For Malawi, we also added data from the National Irrigation Master Plan (irrigation et al., 2015), which presents a geo-referenced inventory of irrigation systems in Malawi for the period around 2000-2010. Second, we compared the data from the irrigation surveys, the GIA and the GMIA with the OSM farmland polygons (i.e. including those for which we could not identify the crop and estate name). In many cases there was considerable overlap, which made it possible to match an irrigation scheme and/or GIA and GMIA grid cell with an OSM polygon. On the basis of this, we made the assumption that any OSM farmland polygon that overlaps with a GIA grid cell or the location of an irrigation scheme can be considered as irrigated area. Finally, the irrigation map was created by merging the GIA with the OSM irrigated farmland polygons and overlaying the resulting map with the 2000 and 2010 synergy cropland maps to scale the cropland area to the respective years. For a few countries, the area in the irrigated area map was lower than the irrigated area in the country statistics. In these cases we expanded the irrigation map by adding grid cells from the GMIA (which is much coarser than the GIA) that have at least 10% overlap with the OSM polygons. If that was still not enough we added the remaining GMIA grid cells to the irrigation map. Figure S3 depicts the irrigation map for 2010.

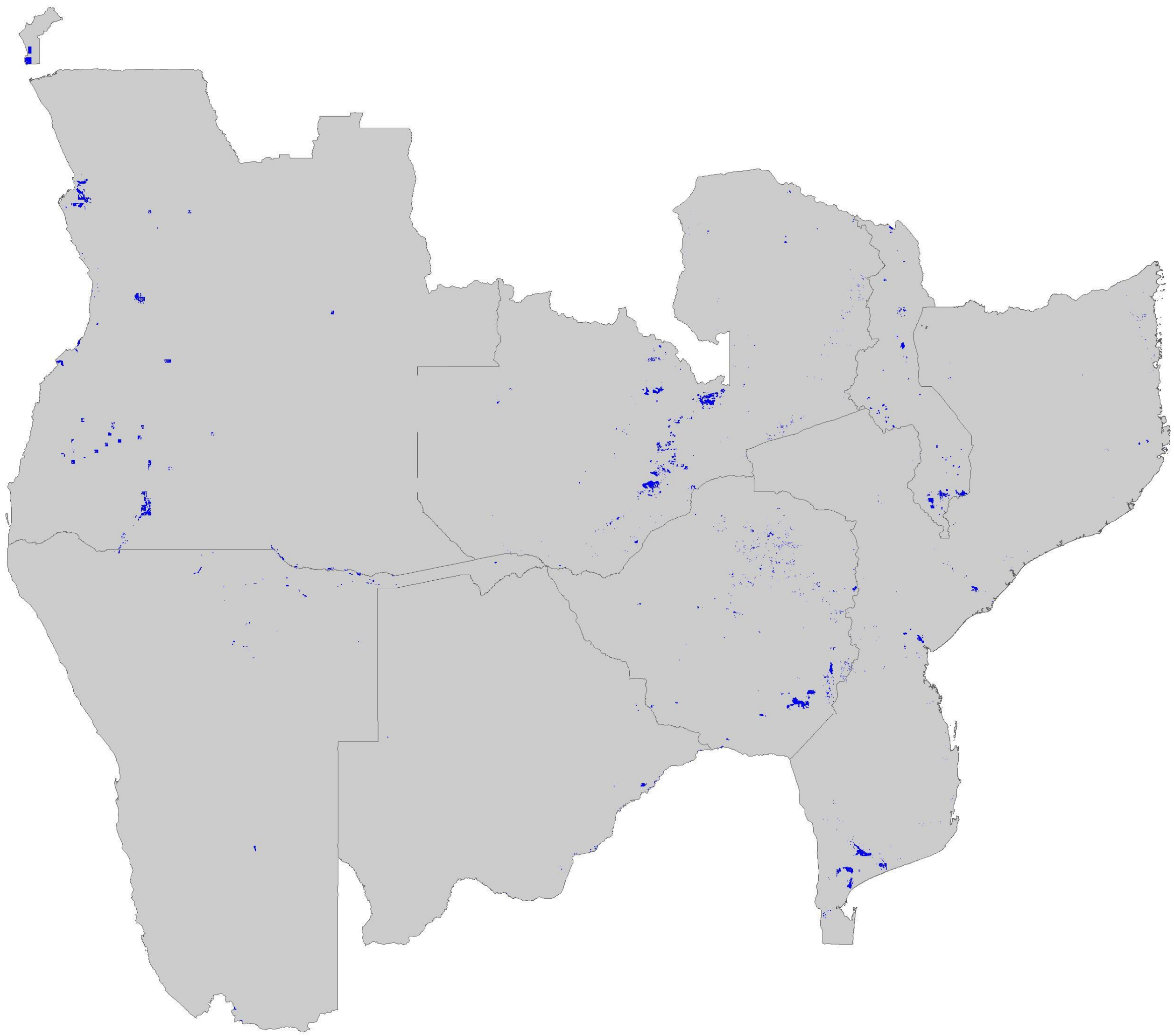


Figure S3: Irrigated area map for 2010. Blue areas indicate the irrigated area. The map was created by combining irrigated area information from OpenStreetMap, GIA [@Meier2018] and GMIA [@Siebert2013]. See text for details.

OpenStreetMap (OSM)

The following steps were used to prepare the database with the size and location of large estates (hereafter the OSM estate database) in the Southern African region:

1. We downloaded the OpenStreetMap (OSM) files for the seven African target countries from http://download.geofabrik.de/downloaded [accessed 25-06-2019], which provides daily updated data extracts from OSM, free of charge.

2. We used software developed by OSM (osmconvert64.exe and osmfilter.exe) to extract vector (polygon) information on farmland and cropland. The latter category provides information on the use of farmland for a number of selected crops (https://wiki.openstreetmap.org/wiki/Key:crop?uselang=en-US) but is not available for all farmland polygons.

3. We merged the farmland and cropland polygons, removing duplicates.

4. We combined and overlayed the OSM data with additional information from a number of sources to (1) validate the location of the large farms and (2) identify additional crops that are produced on OSM farmland polygons. A particularly useful source of information was the Malawi Irrigation Survey (irrigation et al., 2015), which contains the coordinates of a 57 estates with irrigation facilities. For a large number of estates, it also presents information on the main crop that is produced. To further expand the OSM estate database, we cross-referenced it with lists of farms mentioned in sector reports (Chinsinga, 2017; e.g. World Bank, 2010) and company websites.

5. We selected only polygons with a total area of 50 ha or larger, which is the standard threshold in the literature to select commercial farms (e.g. Deininger et al., 2016). The result is a database including 390 polygons with information on the location, main crop and name of the estate.

To assess the coverage of the OSM estate database, we compared the size of the estates with the FAOSTAT crop statistics. To estimate estate size, we overlayed the OSM polygons with the 2010 global synergy cropland map (Lu et al., 2020) and aggregated the cropland area under the estate polygons. A comparison with total harvested area per crop from FAOSTAT shows that our database includes many sugar plantations and tea estates. This is not surprising as sugar and tea are key cash crops in the region, which are mostly produced by large commercial farms. The coverage for other crops is much lower as these include food crops that predominantly grown by smallholders.

We used two strategies to validate the OSM estate database. First we compared the OSM polygons with information on Google Earth, which includes historical satellite imagery for our base year 2010. Figure 2 in the main text compares the OSM estates and the Google Earth image for two selected estates, showing strong resemblance between the two sources. Second, we compared the derived size of the estates for 2010 with comparable data from other sources, in particular World Bank (2020) and irrigation et al. (2015), which information could be linked to our list of estates (Figure S4).

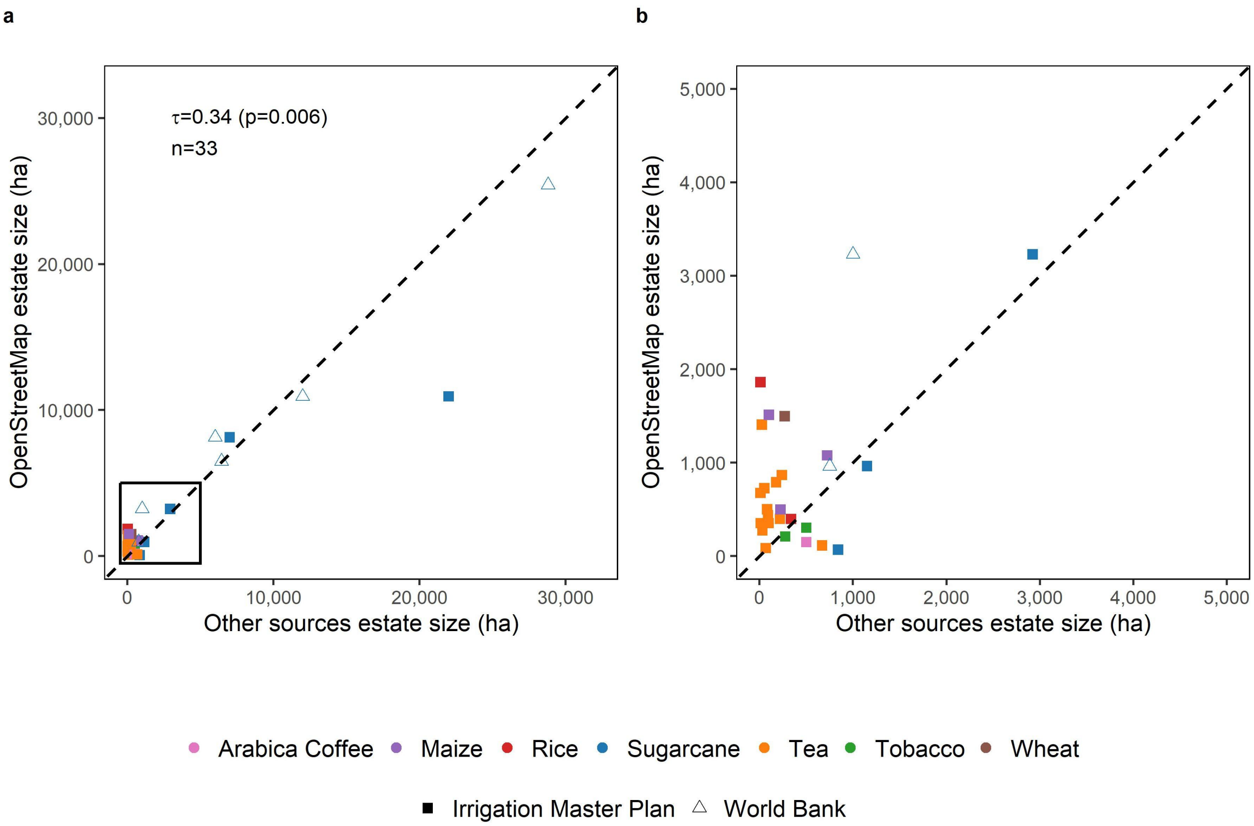


Figure S4: Comparison of the OSM estate database with similar information from other sources. a Comparison of all 33 estates for which additional information is available. b Comparison of estates with a size smaller than 5,000 ha, corresponding with the demarcated plot area in a. The correlation coefficient (τ), p-value (p), and sample size (n) are presented. Dashed lines indicate the 1:1 lines. Data from other sources was taken from World Bank (2020) and irrigation et al. (2015), which present data on the size of irrigated areas, including the name of the estate, around 2010.

Biophysical suitability and potential yield

Spatially explicit information on biophysical suitability and potential yield is taken from IIASA et al. (2012). The International Institute for Applied Systems Analysis (IIASA) in collaboration with the FAO, developed the global agro-ecological zones (GAEZ) methodology to assess the biophysical limitations and potentials for a large number of crops across three production systems: rainfed–low input/subsistence, rainfed–high input and Irrigated-high input. The first class is used for both the low-input and subsistence system in the model. GAEZ presents spatially explicit information on the biophysical suitability (on a scale from 1 to 100) and potential yield (in t/ha) for each production system. As an example, Figure S5 and Figure S6 show the suitability and potential yield for rainfed subsistence/low-input maize production in the Southern African region, respectively.

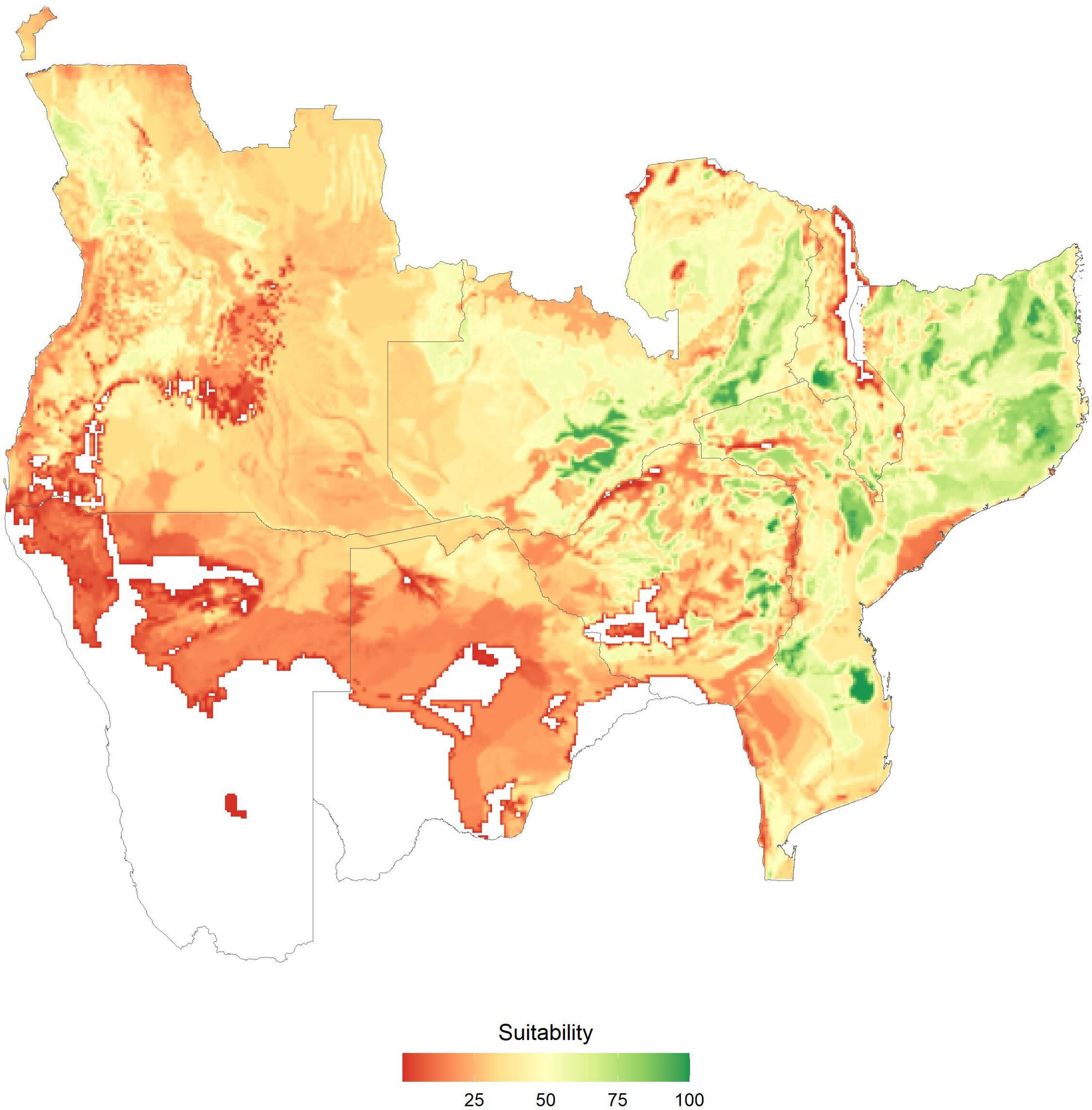


Figure S5: Biophysical suitability for rainfed subsistence/low-input maize. Original maps were resampled from 5 arcmin to 30 arcsec to facilitate comparison with other input maps. Source @IIASA2012.

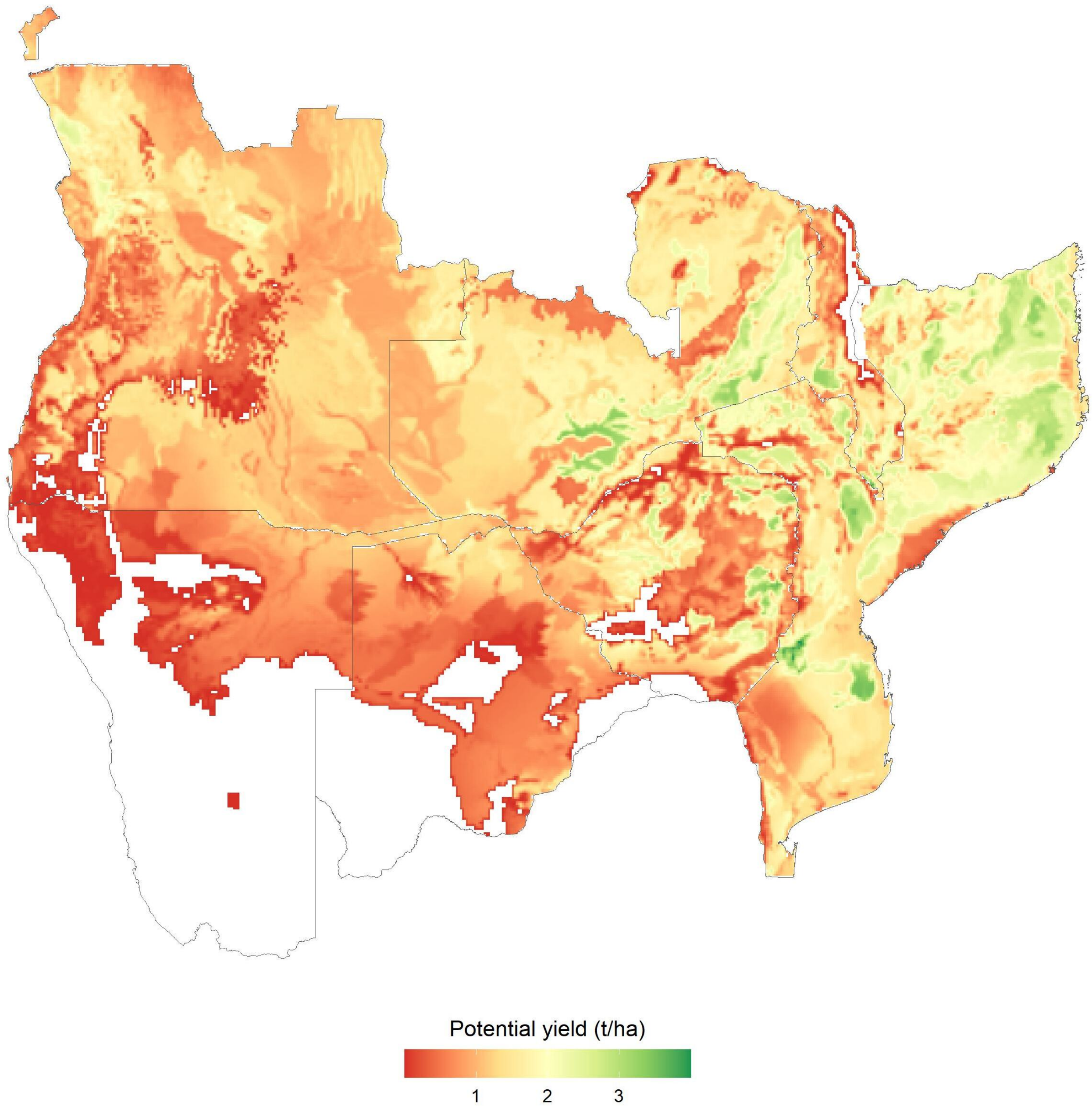


Figure S6: Potetial yield for rainfed subsistence/low-input maize. Original maps were resampled from 5 arcmin to 30 arcsec to facilitate comparison with other input maps. Source @IIASA2012.

Accessibility

In contrast to SPAM, which uses population density (Wood-Sichra et al., 2016), we applied recently updated maps on travel time as a proxy for accessibility. We argue that farmers that belong to the high-input and irrigated systems sell (most of) their products on the market. Hence, these systems will be located in areas that are close to the market or are close to main roads that provide easy access to markets. As a proxy for access to markets and quality of road infrastructure we used two global maps that measure accessibility to high-density urban centers as measured by travel time at a resolution of 1 $\times$ 1 km. The first map (Nelson, 2008) depicts the global road infrastructure in 2000, while the second map (Weiss et al., 2018) contains information for the year 2015. Although both maps depict travel time and are related, they cannot be regarded as proper time-series because the maps were created with a different definition of a city. Nelson (2008) used travel time to the nearest city with 50,000 or more population, while Weiss et al. (2018) used the urban areas in the Global Human Settlement Grid of high-density land cover to locate cities. The most recent product also includes minor roads (e.g., unpaved rural roads), which were only marginally covered by the older map. We used the 2000 and 2010 global travel time maps as an indicator for accessibility for the 2000 and 2010 crop distribution maps, respectively.

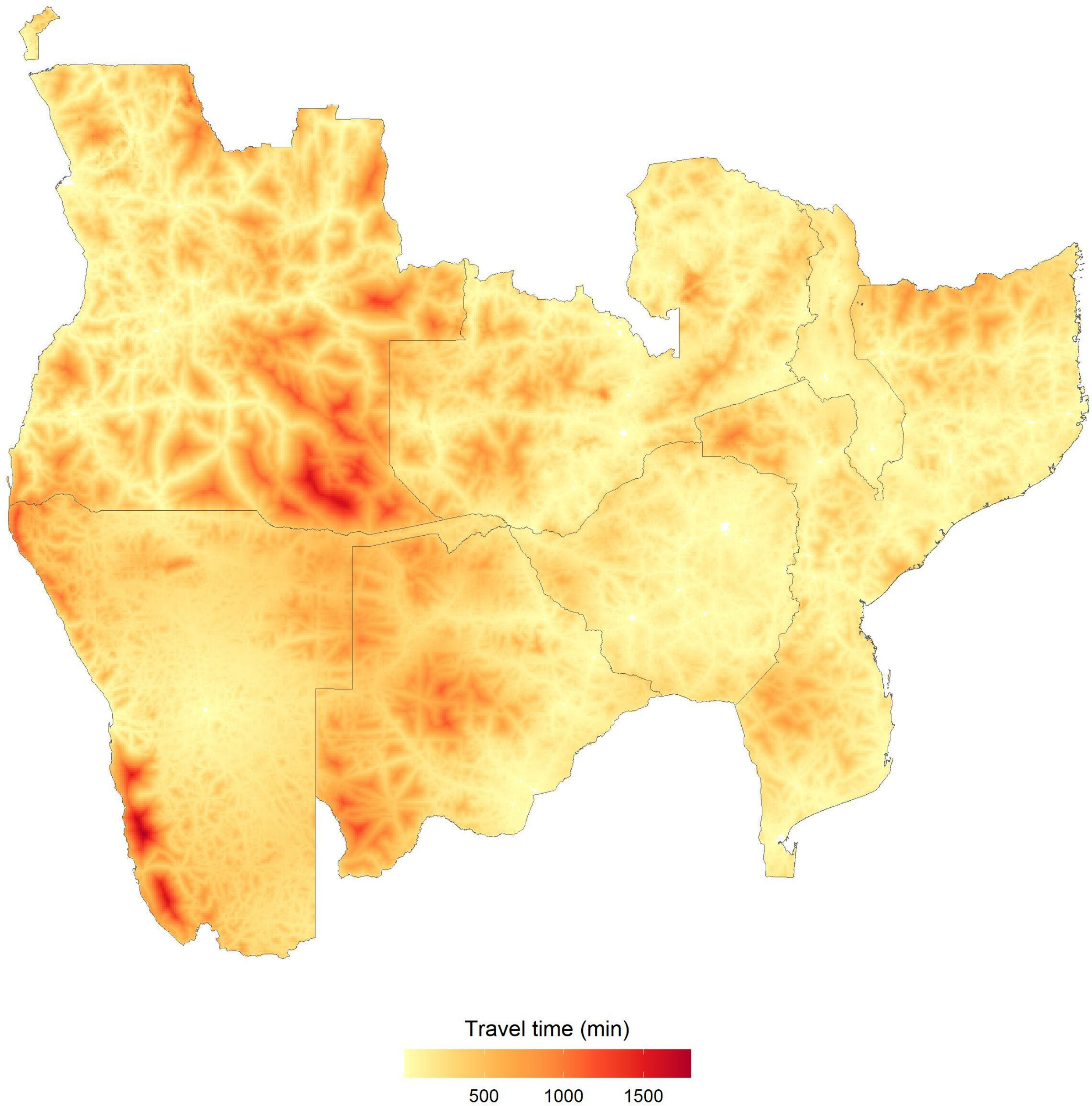


Figure S7: Travel time in 2015. Source: @Weiss2018.

Rural population density

We used the WorldPop (Tatem, 2017) database as a our primary source of information for national population density. WorldPop combines a random forest model with census data to generate a gridded prediction of population density at ~100 m spatial resolution (Stevens et al., 2015). Data is available for at a number of spatial resolutions and various years. We used the global maps at 1 x 1 km and for the period 2000 and 2010 as input. We preferred the WorldPop data, over the SEDAC population data that is used by SPAM because it is available at much higher resolution and available for multiple and more recent years.

To identify rural population areas, and similar to SPAM, we used the urban extent from GRUMP (CIESIN et al., 2011). Urban areas, were identified by analyzing nighttime lights in combination with information on settlement points. Rural population was selected by removing grid cells that are located within the urban extent polygons.

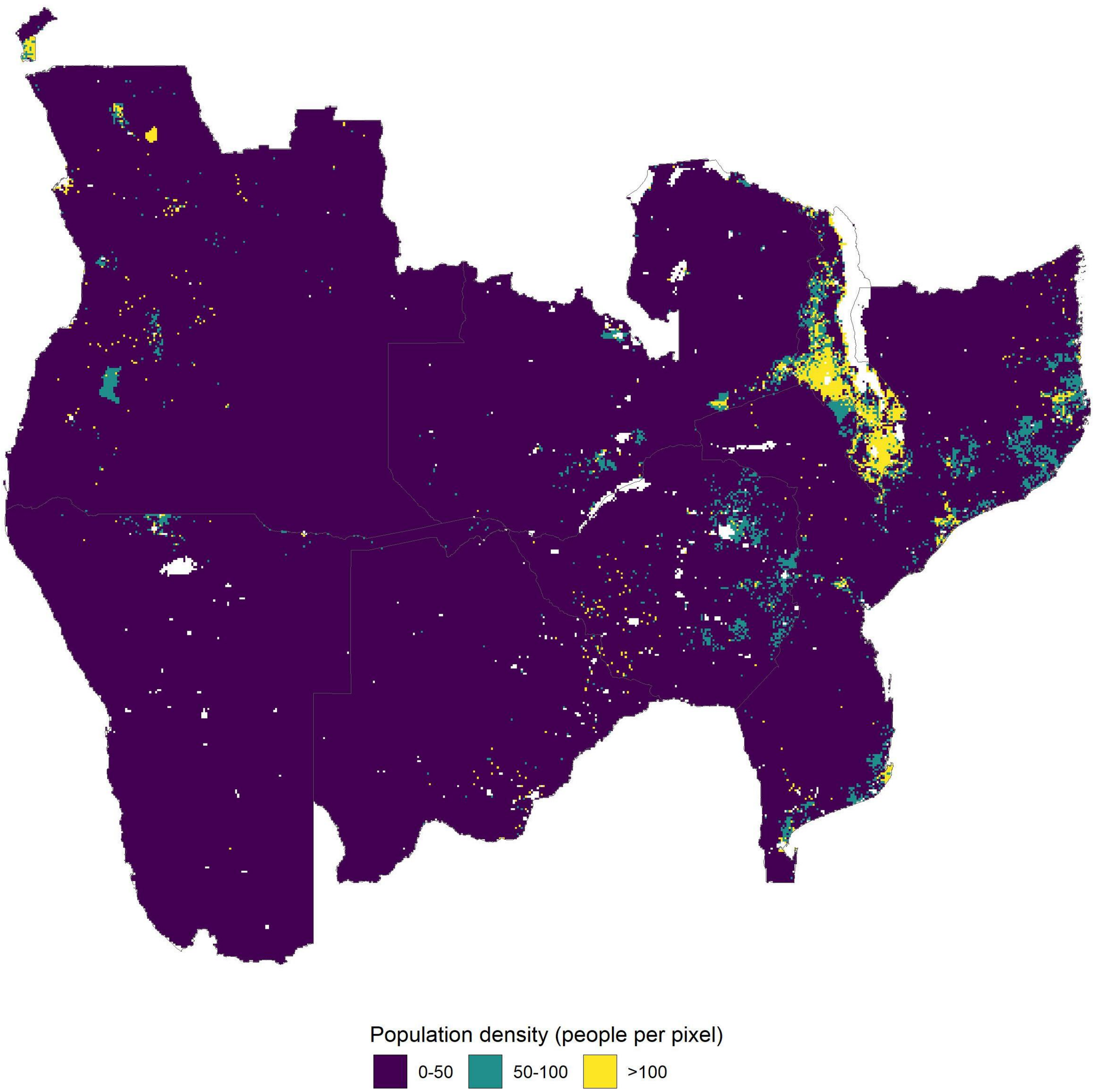


Figure S8: Rural population density in 2010. Source: @Tatem2017 and @CIESIN2011

Crop price

To calculate the potential revenue of a crop at the grid cell level, we followed the SPAM approach (Wood-Sichra et al., 2016) and multiplied the potential yield from IIASA et al. (2012) with crop prices in USD from FAO (2019). Similar to harvested area information, we took the average over three years to smooth annual fluctuations. Due to problems of missing data, we used average prices for the African continent.

Fitness score

The constraint for the subsistence production system and the fitness score for low-input, high-input and irrigated production systems are constructed by combining one or more indicators for rural population density, biophysical suitability, accessibility and potential revenue. In order to make the data comparable, all indicators were scaled between 0 and 1 using the using min-max normalization before further processing:

$$X'_{ijl} = \frac{X_{ijl} - min(X_{ijl})}{max(X_{ijl}) - min(X_{ijl})} \quad (1)$$

where X_{ijl} is the one of the indicators to construct the fitness score, defined for each grid cell i, crop j and production system l, which is converted to X'_{ijl}.

Post-processing

The output of the updated SPAM model consists of regional maps depicting the physical area distribution of at maximum 42 crops across four production systems (in practice not all four systems and crops are observed in a country). Using the same procedure as implemented by SPAM (Wood-Sichra et al., 2016), we also created maps for harvested area, crop yield and production. More specifically, the following approach was used:

1. As a consequence of the optimization process, it is possible that a very small crop area is allocated to a grid cell. We removed all areas that were smaller than 0.0001 ha = 1 m2.
2. To account for the slight reduction in total harvested area for each crop, we scaled the area in all grid cells to ensure the total is equal to the FAOSTAT total harvested area for each crop for the period 2009-2011.
3. We used the information on cropping intensity to prepare harvested area maps for each crop and production system combination.
4. We combined crop specific data from the SPAM2010 database on (1) the ratio of low-input and high-input rainfed yield and (b) the ratio of rainfed and irrigated yield, with the production system shares

and average yield information from FAOSTAT to derive the crop yield for each production system. The derived yields were subsequently spatially distributed using relative potential yield values from GAEZ (IIASA et al., 2012).

5. We prepared production maps by multiplying harvested area with crop yield.

Figure S9 depicts the harvested area, crop yield and production for the four maize production systems.

Figure S9: a Harvested area, b crop yield and c production for the maize subsistence farming system

S2 Solving the model

The General Algebraic Modeling System (GAMS) Release 25.0 (GAMS Development Corporation, 2019) was used to solve the model. As we combined information from many different sources, we often encountered inconsistencies in the data, for example because the available cropland in a certain administrative area is lower than what is needed to allocate the subnational statistics. This implies the model cannot be solved. Another potential problem is the assumption that crops that belong to the subsistence production system are allocated in proportion to the rural population density. This constraint will almost never be within the bounds of the model solution that are set by the other constraints and therefore it likely to result in infeasible model.

To deal with these inconsistencies we added slack variables to the cropland availability (iii), the subnational statistics (iv), the irrigation (v) and the subsistence production system (vi) constraints. A slack variable transforms an inequality constraint in an equality constraint (Boyd et al., 2004) by adding 'slack'. This makes it possible to solve the model even though it is technically infeasible. To capture the interaction and trade-offs between the different slack variables (e.g. a lower slack for irrigated area may result in a higher slack for available cropland) we used different weights (i.e. penalties). Relative high weights are used for the cropland availability and irrigated area slack variables to ensure they are minimized. We prefer low or zero slack for these variables to make sure they are close to the national values in FAOSTAT and other national-level databases. Relative low weights are used for the subnational statistics and subsistence production system slack variables

If the model is solved with a zero sum of slacks the input data can be considered as consistent and there was no need to make adjustments. Positive slacks point at inconsistencies in the data and show were problems occur. Often slack variables are near zero, signaling differences in aggregate variables due to rounding or scaling of raw data, which we ignore. In case of large slack variables, we inspected the data for common errors (e.g. data entry errors in the subnational statistics) and, if needed, corrected them. If there were still problems, we adjusted the data adopting a hierarchy of 'credibility' in the following decreasing order of importance:

1. National agricultural statistics
2. OpenStreetMap
3. Cropland
4. Irrigated area
5. Subnational agricultural statistics

To maintain the link with FAOSTAT, the national agricultural statistics were never changed. As discussed before OpenStreetMap (OSM) information was scaled to the year 2000 and 2010 using the cropland extent for these two periods. In case the OSM information was larger than the are indicated by the (sub)national statistics, we decreased it accordingly. If there were issues related to cropland and irrigated area (i.e. not enough area), values were scaled upwards so that they matched with the (sub)national statistics. We ignored the slack variable for the subnational statistics constraint. Adjustments to the subnational data (e.g. scaling to FAOSTAT data and combining a variety of sources) potentially introduced inconsistencies at province or district level, which are impossible to solve completely. Wood-Sichra et al. (2016) presents more details about the various procedures to prepare the data.

S3 Validation details

For Malawi, we used the 2010-2011 Third Integrated Household Survey (IHS3) implemented by the National Statistical Office and supported by World Bank Living Standards Measurement Study - Integrated Surveys on Agriculture (LSMS-ISA). The IHS contains data on 12.271 households distributed over 768 survey locations. The data can be downloaded from the LSMS-ISA website: http://surveys.worldbank.org/lsms/programs/integrated-surveys-agriculture-ISA [accessed 10-03-2020]. For Zambia, we used the RALS2012, which surveys 442 communities and a total of 8,840 households. The data is available on request from IAPRI: http://www.iapri.org.zm/surveys [accessed 10-03-2020]. We used the following approach to compare the household survey data with the crop distribution maps:

1. We mapped the crop names in the household surveys to the 42 SPAM crops.
2. We reprojected the geocoded data from the household surveys and the crop distribution maps to an equal area projection that is appropriate for Malawi and Zambia (EPSG:32736).
3. We removed all plots with a distance to the farm larger than 1 kilometer (Malawi only) as plots located far from the survey location/household will introduce a bias.
4. We overlayed the household observations with the cropland extent and removed a small number of values that were located outside the cropland area.
5. We removed all crop household observations that were located in administrative units where the SPAM subnational statistics database indicates this crop was not produced. For these observations, the model will never be able to provide accurate results as they reflect an inconsistency between data sources.
6. We compared each crop distribution map (combining the subsistence and low-input rainfed production systems) with the location of the same crop as presented in the household surveys. We used buffers of 5 kilometer and 10 kilometer around the household survey coordinates to account for the offset in location and the possibility that a field is located away from the main location of the household. Overlap between the buffered location and the crop distribution map is considered as a true positive, i.e. a correct prediction of the location of the crop. We calculated the resemblance statistic as the share of true positives in the number of survey observations.
7. We calculated the same statistic for a counterfactual in which the crop distribution maps are replaced by 1,000 randomly selected locations drawn from within the cropland extent. Administrative regions where crops are not produced according to the subnational statistics are excluded from sampling. We calculated a 95% binomial proportion confidence interval for the sample to indicate the probability of success.

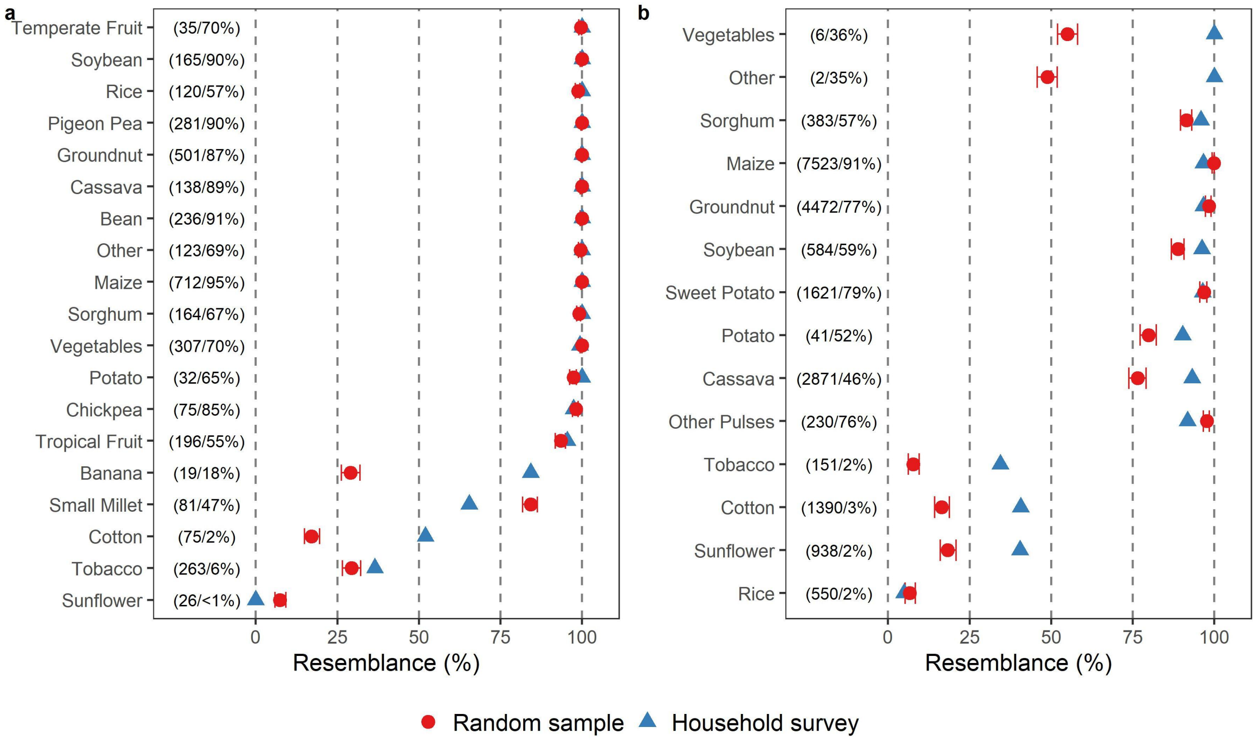
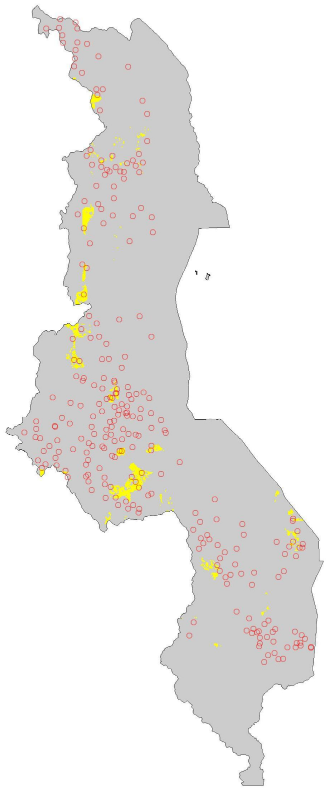


Figure S10: Model validation for 2010. a Malawi and b Zambia. The figures compare the resemblance (%) between the crop distribution maps (combining subsistence and low-input rainfed production systems) and both the location of crops from household surveys and 1,000 randomly selected locations from within the cropland extent. Survey and random sample locations are buffered using a 10 km radius to account for bias in the location of the household survey data. Error bars denote the 95% binomial proportion confidence interval. For each crop the first value on the y-axis denotes the number of locations where a crop is produced according to the household survey. The second value measures the spread of a crop. It represents the number of grid cells with allocated crop area divided by the total number of grid cells in the cropland extent. Crops are ranked using the resemblance with the household survey locations.

a



b

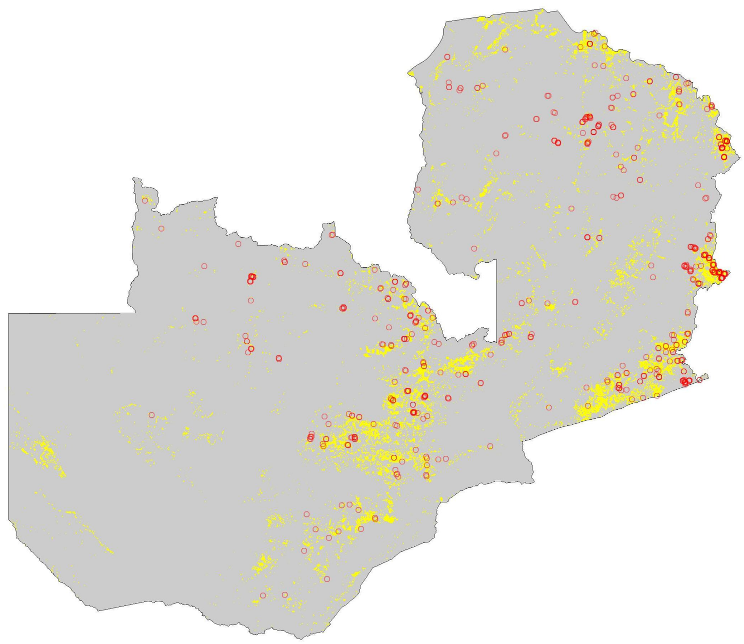


Figure S11: Comparison between crop allocation from the model and the location of crops from household survey data. a Tobacco in Malawi and b Soybeans in Zambia. Yellow area denotes the crop allocation from the model and red circles refer to the location of crops from the household surveys. Malawi and Zambia household data are taken from the 2010-2011 Third Integrated Household Survey and 2012 Rural Agricultural Livelihoods Survey, respectively.

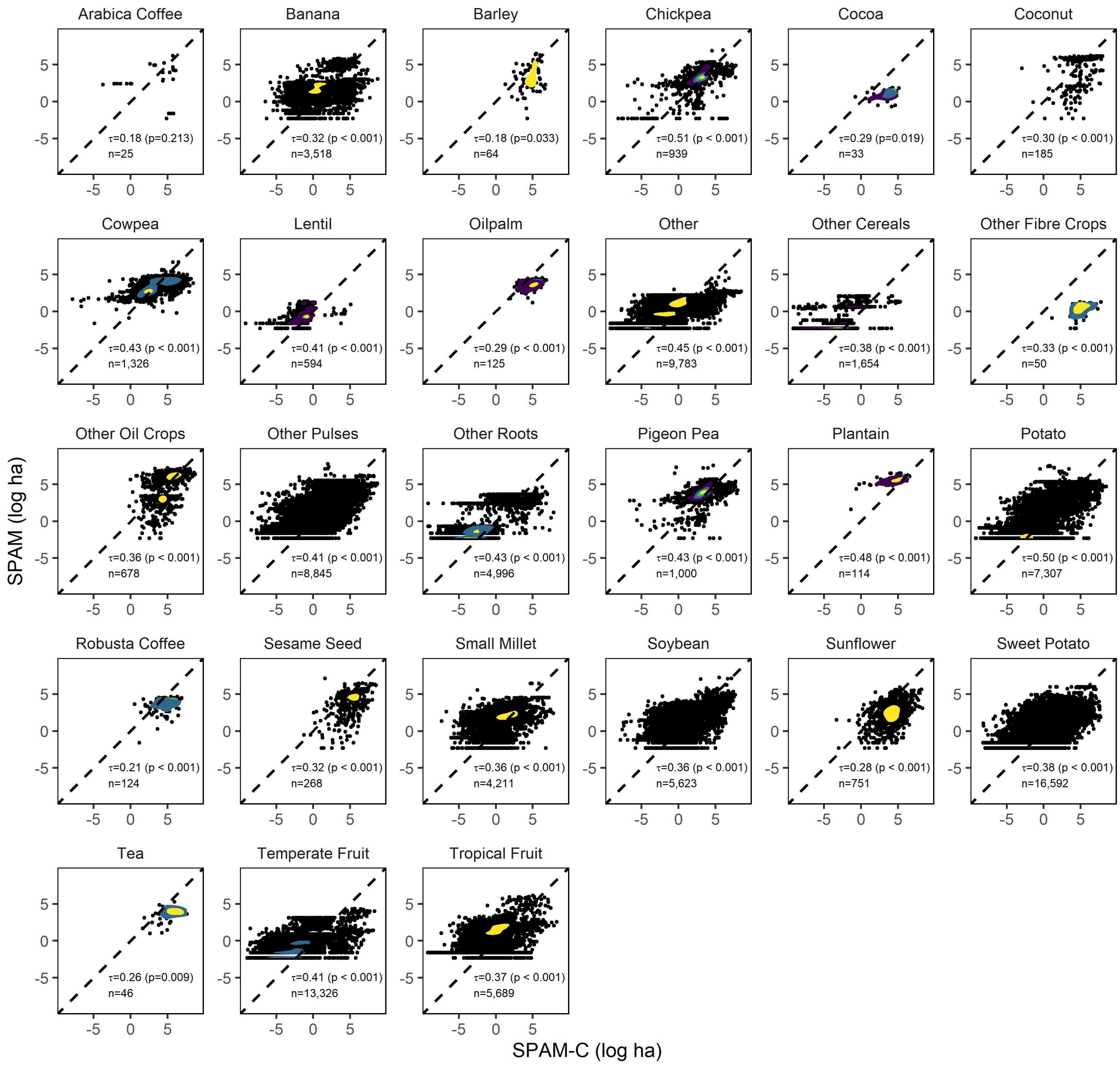


Figure S12: Comparison of physical area between the updated model and SPAM for 2010. The figure shows the results for the remaining 27 crops that are not presented in Figure 6 of the main text. The different production systems are pooled. Model results were aggregated to 5 arcmin for comparison. The colored shaded area shows the smoothed density of the grid-cell data. The correlation coefficient (τ), p-value (p), and sample size (n) are presented. Dashed lines indicate the 1:1 lines. Zero values were dropped from the analysis. To improve visualization all values are expressed in log and axis are limited in the range of -9 to 9 log ha, resulting in the removal of a small number of observations.

S4 Software and hardware

All data preparation was done in R (R Core Team, 2020). Cleaning, processing and visualization of data was mostly done with the Tidyverse packages (Wickham et al., 2019). GIS operations were performed using the sf (Pebesma, 2018) and raster (Hijmans, 2020) packages. Computationally intensive operations, in particular the reprojection and aggregation of high resolution raster files was done with the Geospatial Data Abstraction Library (GDAL) (GDAL OGR contributors, 2020), activated by means of the gdalUtils (Greenberg et al., 2020) package in R. We used the General Algebraic Modeling System (GAMS) Release 25.0 (GAMS Development Corporation, 2019) to solve the spatial allocation production model with the CPLEX and MOSEK solvers for the fitness score and cross-entropy models, respectively.

The model was solved on a Microsoft Windows Server 2012 R2 Standard platform with an Intel Xeon CPU E5-2650, 2.30GHZ processor and 32 GB internal memory. Solving the model takes between 37 seconds (Botswana) and 6 hours 45 min (Angola) for the fitness score model, and 49 seconds (Botswana) and 4 hours 57 (Zimbabwe) min (Figure S13), depending on the model dimensions, which in turn are determined by the combination of the number of grid cells, crop-production system combinations and number of subnational administrative regions.

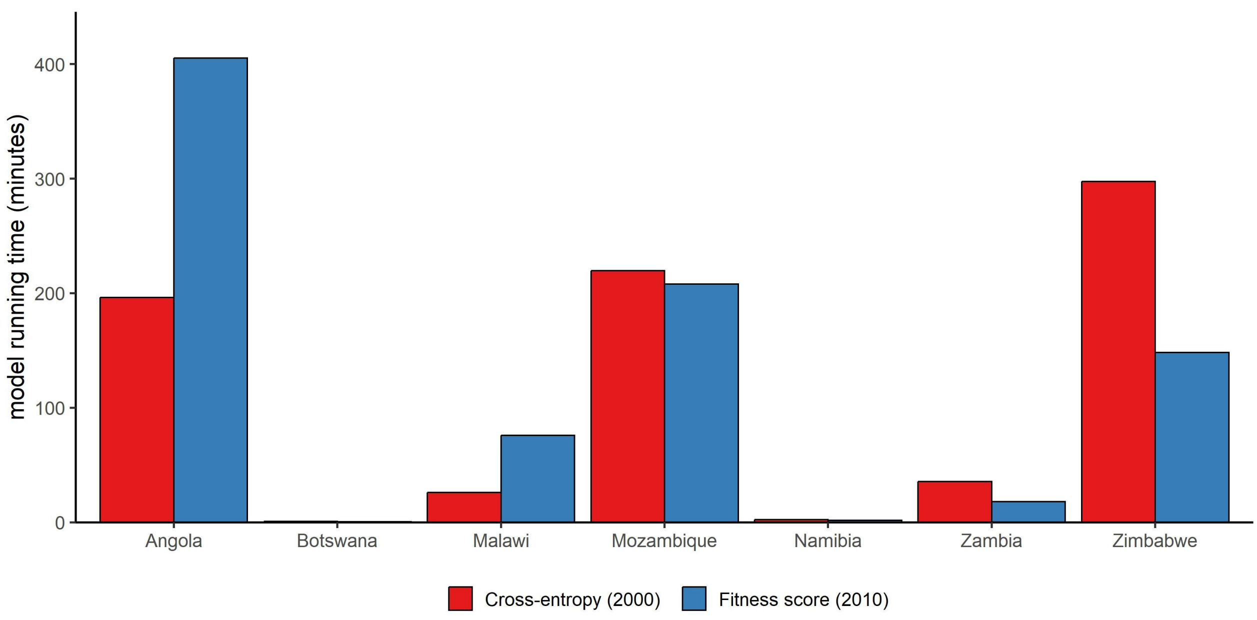


Figure S13: Running time per model and country

References

Boyd, S., & Vandenberghe, L. (2004). *Convex Optimization*. https://doi.org/10.1017/CBO9780511804441

Chinsinga, B. (2017). The green belt initiative, politics and sugar production in Malawi. *Journal of Southern African Studies*, *43*(3), 501–515. https://doi.org/10.1080/03057070.2016.1211401

CIESIN, IFPRI, World Bank, & CIAT. (2011). *The Global Rural-Urban Mapping Project, Version 1 (GRUMPv1): Urban Extents Grid*. Retrieved from https://sedac.ciesin.columbia.edu/data/collection/grump-v1

Deininger, K., & Xia, F. (2016). Quantifying Spillover Effects from Large Land-based Investment: The Case of Mozambique. *World Development*, *87*, 227–241. https://doi.org/10.1016/j.worlddev.2016.06.016

FAO. (2016). *AQUASTAT Main Database*. Retrieved from http://www.fao.org/aquastat/en/

FAO. (2019). *FAOSTAT*. Retrieved from http://www.fao.org/faostat

GAMS Development Corporation. (2019). *General Algebraic Modeling System (GAMS)*. Retrieved from www.gams.com

GDAL OGR contributors. (2020). *Geospatial Data Abstraction software Library*. Retrieved from https://gdal.org

Greenberg, J. A., & Mattiuzzi, M. (2020). *gdalUtils: Wrappers for the Geospatial Data Abstraction Library (GDAL)*. Retrieved from https://cran.r-project.org/package=gdalUtils

Hijmans, R. J. (2020). *raster: Geographic Data Analysis and Modeling*. Retrieved from https://cran.r-project.org/package=raster

IIASA, & FAO. (2012). *Global Agro-ecological Zones (GAEZv3.0)*. Laxenburg; Rome: IIASA & FAO.

irrigation, M. of agriculture, & development, water. (2015). *Malawi national irrigation master plan and investment framework: Appendix 5 Inventory of Existing Schemes*. The SMEC Group.

Joglekar, A. K. B., Wood-Sichra, U., & Pardey, P. G. (2019). Pixelating crop production: Consequences of methodological choices. *PLOS ONE*, *14*(2), e0212281. https://doi.org/10.1371/journal.pone.0212281

Lu, M., Wu, W., You, L., See, L., Fritz, S., Yu, Q., … Xue, B. (2020). A cultivated planet in 2010: 1. the global synergy cropland map. *Earth System Science Data*. https://doi.org/doi.org/10.5194/essd-2020-12

Meier, J., Zabel, F., & Mauser, W. (2018). A global approach to estimate irrigated areas – a comparison between different data and statistics. *Hydrology and Earth System Sciences*, *22*(2), 1119–1133. https://doi.org/10.5194/hess-22-1119-2018

Nelson, A. (2008). *Travel time to major cities: a global map of accessibility*. Retrieved from Global Environment Monitoring Unit, Joint Research Centre of the European Commission website: http://forobs.jrc.ec.europa.eu/products/gam/

Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, *10*(1), 439. https://doi.org/10.32614/RJ-2018-009

R Core Team. (2020). *R: A Language and Environment for Statistical Computing.* Retrieved from https://www.r-project.org/

Siebert, S., Henrich, V., Frenken, K., & Burke, J. (2013). *Global Map of Irrigation Areas version 5.* Rheinisch Friedrich-Wilhelms-University, Bonn, Germany.

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), e0107042. https://doi.org/10.1371/journal.pone.0107042

Tatem, A. J. (2017). *WorldPop, open data for spatial demography.* https://doi.org/10.1038/sdata.2017.4

Weiss, D. J., Nelson, A., Gibson, H. S., Temperley, W., Peedell, S., Lieber, A., … Gething, P. W. (2018). A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature*, *553*(7688), 333–336. https://doi.org/10.1038/nature25181

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wood-Sichra, U., Joglekar, A. B., & You, L. (2016). *Spatial Producion Allocation Model (SPAM) 2005: Technical Documentation.* International Food Policy Research (IFPRI), Washington, D.C.

World Bank. (2010). *The Zambezi river basin: A multi-sector investment opportunities analysis.* Washington, DC.

World Bank. (2020). *World Development Indicators.* Retrieved from https://databank.worldbank.org/source/world-development-indicators

You, L., Wood, S., Wood-Sichra, U., & Wu, W. (2014). Generating global crop distribution maps: From census to grid. *Agricultural Systems*, *127*, 53–60. https://doi.org/10.1016/j.agsy.2014.01.002

Yu, Q., You, L., Wood-Sichra, U., Ru, Y., Joglekar, A. K. B., Fritz, S., … Yang, P. (2020). A cultivated planet in 2010: 2. the global gridded agricultural production maps. *Earth System Science Data*. https://doi.org/https://doi.org/10.5194/essd-2020-11