Gridded 5-arcmin, simultaneously farm-size- and crop-specific harvested area for 56 countries

Han Su[1, 2], Bárbara Willaarts [2], Diana Luna-Gonzalez [2], Maarten S. Krol [1], Rick J. Hogeboom [1 ,3]

[1] Multidisciplinary Water Management group, Faculty of Engineering Technology, University of Twente, Enschede, 7500AE, the Netherlands
[2] Water Security group, International Institute for Applied Systems Analysis (IIASA), Laxenburg, 2361, Austria
[3] Water Footprint Network, Enschede, 7522NB, the Netherlands

Correspondence to: Han Su (h.su@utwente.nl)

Abstract.

Farms are not homogeneous. Smaller farms generally have different planted crops, yields, agricultural input, and irrigations compared to larger farms. Mapping farm size could facilitate studies to quantify how water availability and climate change affect small and large farms, respectively. Given the lack of gridded farm size specific data, this study aims to develop a global gridded, simultaneously farm size- and crop-specific dataset of harvested area. We achieved it by downscaling a best-available dataset, which collected direct measurements on crop and farm size, using crop maps, cropland extent, and dominant field size distributions for 2010. Uncertainties in crop maps were explicitly considered by using two crop maps separately during downscaling. Due to data availability, our downscaled maps cover 56 countries accounting for half of the global cropland. Based on the two different crop maps, we have 5-arcmin gridded, simultaneously farm size- and crop-specific dataset of harvested areas, one for 11 farm sizes, 27 crops, and 2 farming systems and another one for 11 farm sizes, 42 crops, and 4 farming systems. The downscaled maps show major planted crops and irrigation change along with farm sizes, which support previous findings. Validations show well consistencies with observations on farm size specific oil palm from satellite images, farm size specific irrigation from household surveys, and previous studies that map farm size but are not crop-specific. We observed some uncertainties at the grid cell level and found conclusions at the country level are robust to these uncertainties including the uncertainties from the crop maps. Our downscaled maps will help to explicitly include farm size into global agriculture modeling. The source data, code, and downscaled maps are open-access and free available at https://doi.org/10.5281/zenodo.5747616 (Su et al., 2022).

1 Introduction

There are over 608 million farms around the world (Lowder et al., 2016; Lowder et al., 2021). Land and water resources are not equally distributed among these farms. More than 80% of these farms are smaller than 2 hectares and they only utilize around 20% of farmland area (Lowder et al., 2021; Bosc et al., 2013). The largest one percent of farms utilize 70% of global farmland area (Lowder et al., 2021). Smaller farms also insufficiently apply irrigation to adapt to water scarcity in low- and middle-income countries (Ricciardi et al., 2020).

In addition to water and land resources, the characteristics of agricultural production differ across farm sizes, which may be country-dependent. For example, in terms of crops, smaller farms plant more fruits, pulses, and roots and tubers while larger farms plant more vegetables, nuts, and oil crops (Ricciardi et al., 2018b; Herrero et al., 2017). In terms of agricultural practices used to increase agricultural productivity, farmers who operate smaller farms tend to increase the use of non-fixed inputs, such as fertilizers and pesticides, while larger farms tend to increase fixed inputs, such as machinery (Ren et al., 2019). Smaller farms also have a greater biodiversity on average (Ricciardi et al., 2021; Noack et al., 2021). Though whether smaller farms have a higher yield has long been debated, it appears that yield often correlates with farm size (see Rudra (1968); Savastano and Scandizzo (2017); Gollin (2019); Ricciardi et al. (2021)).

These above-mentioned characteristics stimulate studies to explicitly discern small and large scale farms in agriculture studies and map farm sizes (Meyfroidt, 2017; Riesgo et al., 2016). At the global level, mapping farm sizes can be traced back to the studies of Lowder et al. (2016), Samberg et al. (2016), and Fritz et al. (2015). Lowder et al. (2016) estimated the country level distribution of farm size, based on multiple agricultural censuses. Samberg et al. (2016) used the Mean Agricultural Area (MAA) to assign each subnational administrative unit with a farm size. This may overestimate the area of small farms because not all farms are small, even if they are in the administrative unit dominated by small farms (Ricciardi et al., 2018b). Fritz et al. (2015) developed a gridded global dominant field size map using manually labeled field size data on the satellite images and spatial interpolation. The dominant field size map was updated by Lesiv et al. (2019). When interpreting fields as farms, the small farm area will also be overestimated as large farms can include small-sized fields. Herrero et al. (2017) used the country level farm size data from Lowder et al. (2016) and Fritz et al. (2015) to develop a dominant farm size map which was later updated by Mehrabi et al. (2020) using the field size map from Lesiv et al. (2019). Given that dominant farm size only assigns one farm size to each cell (usually 10 km by 10 km), dominant farm size may over/underestimate some kinds of farm sizes when it is used to estimate the number and area distribution of different farm sizes.

In previous studies, farm size mapping is not crop-specific. One way to estimate the planted crops for different farm sizes is to overlap the farm size map with crop maps, e.g., Monfreda et al. (2008) in Samberg et al. (2016) and Mehrabi et al. (2020), Ray et al. (2013) in Herrero et al. (2017). Overlays with crop maps may lead to biases in the allocation of crop-specific cropping areas to farm sizes (Ricciardi et al., 2018b), because of differences between farm size and MAA, field sizes, and dominant farm sizes and also due to possible structural differences in crop choices between farm sizes.

One way to avoid such biases is to develop a simultaneously farm size- and crop-specific map. Ricciardi et al. (2018b); Ricciardi et al. (2018a) established an empirical global database using agriculture census and household survey that directly measured crop production or areas in combination with farm size. This dataset covers half of the global cropland, including data for 56 countries[1] – with subnational data for 46 countries. Ricciardi's dataset, however, does not have gridded maps, so it has limited capability to fulfill the needs of global climate change and water resources studies, where the hydrological model and climate models commonly use grided maps as input. Lacking gridded farm size- and crop-specific maps limits the evaluations of how water scarcity and climate change affect small and large farms, respectively.

This study aims to develop a global gridded, simultaneously farm size- and crop-specific dataset of harvested areas with additional information on farming systems. Considering the data availability, the baseline year is 2010 with data covering 56 countries. We compiled multiple datasets including cropland extent, field size distribution, as well as crop distribution and farming systems and used them to downscale the empirical farm size- and crop-specific datasets developed by Ricciardi et al. (2018b); Ricciardi et al. (2018a), from the level of administrative units into a 5 arcmin grid cell level. We also explicitly considered the uncertainties in crop distributions by using two crop maps. The resulting downscaled maps were validated with empirical data and compared with previous studies.

2 Methods

2.1 Overview

Imagine that we know the crop area of small and large farms within an administrative unit, to downscale it, if we get a high spatial resolution map of crop area, we may have some idea on where the small and large farms may locate because some crops are planted more by small farms and some crops are planted more by large farms. In addition, if we have the field size distribution within the administrative unit, we could know more about the location of small and large farms because large fields only belong to large farms and small farms can only be located in small fields. When we combine the information from the crop map and field size distribution, even though we could not precisely locate small and large farms, we can estimation their distributions in this administrative unit with some extent of uncertainties. This is how we develop the gridded, simultaneously farm size- and crop-specific dataset of harvested areas. Theoretically, we could estimate multiple distributions of small and large farms that are consistent with all the administrative level and grid cell level data. Practically, however, these distributions may not exist because of the background inconsistencies in the datasets. To deal with the background inconsistencies, we assume the best estimation of the farm size- and crop-specific distributions are the distributions that could maximize consistencies with datasets. In those cases, we tried to find multiple distributions that meet the same level of consistency with datasets and averaged the multiple distributions to get the final estimation.

1 In their paper, they claim to have data for 55 countries. In the dataset they published, it contains the 56th country, the Czech Republic.

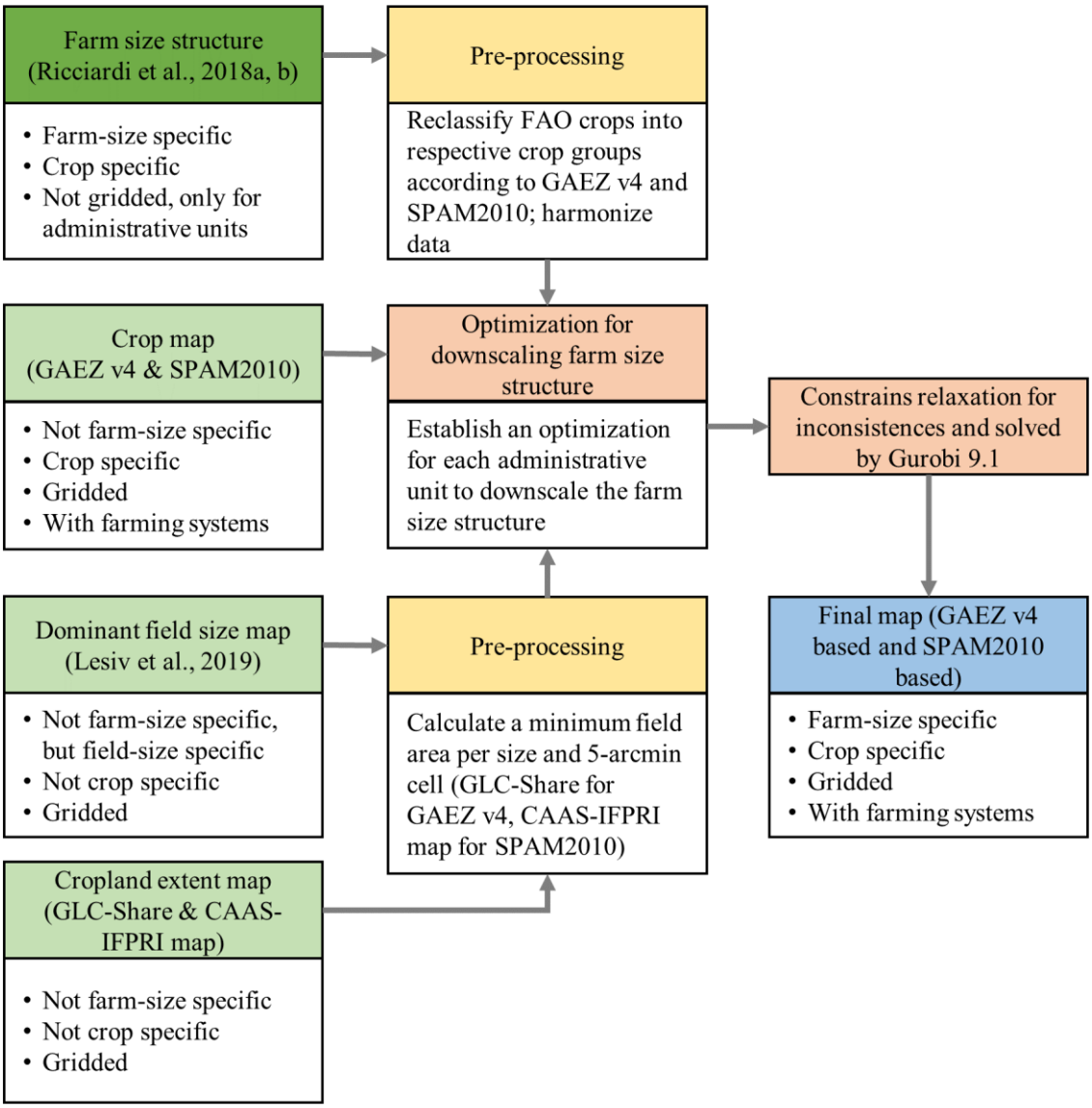

Figure 1. Diagram of map development processors.

The map development involved pre-processing of multiple datasets, establishing optimization for downscaling, and constraints relaxation and solving optimization problems (Fig. 1). The pre-processing included two parts: reclassifying crops to accommodate differences in crop classification used in the underlying datasets and harmonizing Ricciardi's dataset and converting the dominant field size map into a minimum field area per size and 5-arcmin grid cell (Sect. 2.2). The downscaling was achieved by maximizing consistencies with multiple datasets that provide information on the location of each farm size and planted crops. Specifically, we established an optimization for each administrative unit (Sect. 2.3) and solved it via constraints relaxations (Sect. 2.4). Priorities in achieving consistency with the various underlying datasets were considered

during these processes (Sect. 2.3 and 2.4). The spatial crop distribution affects both crop location and farm size location during downscaling and is usually uncertain. To consider the uncertainties in crop maps, we used two crop maps to develop two alternative versions of the final downscaled map separately.

2.2 Datasets and pre-processing

All the datasets used in this study can be found in Table 1. Ricciardi's dataset provides the farm size- and crop-specific cropping area for 56 countries at the administrative unit level ([S1] for the list of 56 countries). The eleven farm sizes in this dataset are based on the classification from the World Census of Agriculture (WCA) (FAO, 2020b; Ricciardi et al., 2018a): 0–1 ha, 1–2 ha, 2–5 ha, 5–10 ha, 10–20 ha, 20–50 ha, 50–100 ha, 100–200 ha, 200–500 ha, 500–1000 ha, and >1000 ha. The cropping area in this dataset means either crop area, planted area, harvested area, or cultivated area. Because the data quality varies from country to country and this dataset was not harmonized over time, we chose to downscale its crop-specific farm size structure, i.e., the crop-specific percentage of area per farm size, instead of the area. The crop-specific harvested area is from two separate crop maps: GAEZ v4 (Fischer et al., 2021; FAO and IIASA, 2021) and SPAM2010 (Yu et al., 2020). These are the only two crop maps containing harvested area of tens crops for the year 2010 at 5 arcmin spatial resolution (Kim et al., 2021). GAEZ v4 and SPAM2010 have their own crop classification systems ([S2, S3] for details). GAEZ v4 distinguishes two farming systems: irrigated and rainfed. SPAM2010 further distinguishes rainfed into low- and high-input rainfed and subsistence rainfed. The dominant field size distribution (Lesiv et al., 2019) indicates where larger farms may locate. It provides spatial distribution for five field sizes: < 0.64 ha, 0.64–2.56 ha, 2.56–16 ha, 16–100 ha, and >100 ha. For pre-processing, cropland extent maps were also included.

Table 1. Datasets that were used to develop the gridded, farm size specific, and crop-specific dataset of harvested area.

Dataset	Indicator	Spatial coverage and resolution	Time	Crop	Note
Ricciardi et al. (2018b); Ricciardi et al. (2018a)	Farm size structure*	56 countries; (sub)national administrative unit	Varies from 2001 to 2015	154 FAO crops	11 farm sizes
GAEZ v4 (Fischer et al., 2021; FAO and IIASA, 2021)	Harvested area	Global; gridded, 5 arcmin (10 km)	2010	27 GAEZ crops**	2 farming systems (irrigated and rainfed)
SPAM2010 (Yu et al., 2020)	Harvested area	Global; gridded, 5 arcmin (10 km)	2010	42 SPAM crops	4 farming systems (irrigated, low- and high-input rainfed and subsistence rainfed)

Dominant field size distribution (Lesiv et al., 2019)	Dominant field size	Global; gridded, 30 arcsec (1 km)	Varies from 2000 to 2017	Not crop-specific	5 field sizes
GLC-Share (Latham et al., 2014)	Cropland extent	Global; gridded, 30 arcsec (1 km)	Around 2010	Not crop-specific	The based map of GAEZ v4
CAAS-IFPRI cropland extent map (Lu et al., 2020)	Cropland extent	Global; gridded, 15 arcsec (0.5 km)	2010	Not crop-specific	The base map of SPAM2010

* Here we mean the crop-specific percentage of harvested area per farm size within an administrative unit

** The 27th crop is Fruits and Nuts which is not listed in the document but available in their dataset

To pre-process Ricciardi's dataset, we first reclassified the FAO crops in this dataset into 27 GAEZ crops and 42 SPAM crops, respectively. Detailed criteria can be found in [S2, S3]. We used the cropping area to get the crop-specific farm size structure. In this dataset, the cropping area is crop-specific and includes four items: crop area, planted area, harvested area, and cultivated area. These variables were identified by Ricciardi's dataset from the local agriculture census. There is no worldwide standard definition for these items (FAO, 2015). Local agriculture censuses have their own preference to use one of them for specific crops. In generally, planted area is used for temporary crops; cultivated area for temporary crops and permanent crops; crop area for temporary crops, permanent crops, fallow, meadows, and pastures; harvested area is the cultivated area excluding the area destroyed by natural disasters or other reasons (FAO, 2020a, 2015). In terms of data availability, one or two items are available for most countries. To harmonize data, when more than one item is available, we used the item with a larger overall area (after crop reclassification) to estimate farm size structure because larger overall area means more farm size classes have available data in most cases. If none of the four items were available, we used crop production data provided by Ricciardi's dataset to get the crop-specific farm size structure. In this case, we assumed constant yield across farm sizes.

We also converted the 1 * 1 km dominant field size map into a minimum field area per size and 5-arcmin cell during pre-processing. We interpreted *dominant field size* as a field of that size accounting for at least 50% of cropland in the cell. For each field size, we calculated the minimum field area by using the 50% of cropland extent that is dominated by the respective field size. We then summed and scaled the minimum field area to cover all croplands of 5-arcmin cells. To keep cropland extent consistent with crop map during downscaling, GLC-Share is used when the crop map is GAEZ v4; CAAS-IFPRI cropland extent map is used when the crop map is SPAM2010. The minimum field area of size 16–100 ha is 120 ha in the cell #23 which means, for example, farms larger than 16 ha should occupy at least 120 ha in the cell #23.

2.3 Optimization for downscaling

For each administrative unit defined in Ricciardi's dataset, we established the following optimization problem for downscaling:

Sets:

c, Crops

f, Farm size, labelled by the lower bound of the eleven farm sizes

e, Field size, labelled by the lower bound of the five field sizes

s, Farming system

a, Administrative unit

g, Grid cell

Parameters:

$ha.R_{c,f,a}$, Crop-specific farm size structure, percentage of the harvested area of farm size f that plant crop c in the administrative unit a, from Ricciardi's dataset

$ha.S_{c,s,g}$, Harvested area of crop c under farming system s at grid cell g, from crop map

$ha.L_{e,g}$, Minimum field area of field size e at grid cell g, from dominant field size map and crop extent map

p_f, The minimum farm area of farm size f in any gird cell when the farm size f exists; it is the lower bound of the farm size class f

l, Elastic factor

Variables:

$ha_{c,f,s,g}$ Harvested area of crop c, farm size f, farming system s at grid cell g, estimated by this study

Objective function:

Since we aim to downscale Ricciardi's dataset, we wanted to maximize consistencies with Ricciardi's dataset when constraints allow:

$$min \sum_{c,f} abs\left(ha.R_{c,f,a} \sum_{s,g \in a} ha.S_{c,s,g} - \sum_{s,g \in a} ha_{c,f,s,g}\right) \quad (1)$$

Constraints:

The first constraint ensures consistencies with crop map: the total harvested area per crop per farming system per grid cell in our map equals the harvested area per crop per farming system per grid cell in the crop map.

$$\sum_{f} ha_{c,f,s,g} = ha.S_{c,s,g}, \forall c, s, g \quad (2)$$

The second constraint ensures minimum consistencies with Ricciardi's dataset. The relative difference in farm size structure between our estimation and Ricciardi's dataset would be less than 10%. This ensures that we do not diverge far from Ricciardi's dataset when other constraints are hard to meet. In this case, we would relax other constraints to ensure these minimum consistencies with Ricciardi's dataset. The arbitrary 10% relative difference considers timestamp differences in Ricciardi's dataset and overall uncertainties underlying each of the datasets.

$$90\% * ha.R_{c,f,a} \sum_{s,g \in a} ha.S_{c,s,g} \leq \sum_{s,g \in a} ha_{c,f,s,g} \leq 110\% * ha.R_{c,f,a} \sum_{s,g \in a} ha.S_{c,s,g}, \forall c, f \quad (3)$$

Third, we also applied a minimum allocated area for each farm size at each grid cell. The minimum allocated area is not necessarily required by the definition of farm size since the farm size is defined based on the total operated or cultivated area that does not need to be a single crop area and single farming system. It is still reasonable to include it because we applied it at the 5-arcmin (~10 km) grid cell level. Considering the uncertainties in these constraints and inconsistencies among datasets, we consider this constraint in a hard form and soft form. We used hard form by default. We consider relaxing these constraints using the soft form when the optimization is infeasible (see Sect. 2.4). The soft form does not require the minimum allocation area for each farming system.

Hard form

$$ha_{c,f,s,g} \geq p_f, \forall c, f, s, g, if\ ha_{c,f,s,g} > 0 \quad (4)$$

Soft form

$$\sum_{s} ha_{c,f,s,g} \geq l \times p_f, \forall c, f, g, if\ ha_{c,f,s,g} > 0 \quad (5)$$

Fourth, we applied a minimum area constraint for some farm sizes according to the dominant field size distribution. This constraint follows the logic that a field could only belong to an equal or larger size of farm. We assumed a linear distribution of area within each farm size to accommodate the different classifications of size in farms and fields.

Given the area of field larger than 100 ha, for farms larger than 100 ha:

$$\sum_{c,s,f \geq 100} ha_{c,f,s,g} \geq ha.L_{100,g}, \forall g \quad (6)$$

Given the area of field larger than 16 ha, for farms larger than 10 ha:

$$\sum_{c,s,f \geq 20} ha_{c,f,s,g} + \frac{20-16}{20-10} \sum_{c,s} ha_{c,10,s,g} \geq ha.L_{100,g} + ha.L_{16,g}, \forall g \quad (7)$$

Given the area of field larger than 2.56 ha, for farms larger than 2 ha:

$$\sum_{c,s,f \geq 5} ha_{c,f,s,g} + \frac{5-2.56}{5-2} \sum_{c,s} ha_{c,2,s,g} \geq ha.L_{100,g} + ha.L_{16,g} + ha.L_{2.56,g}, \forall g \quad (8)$$

Given the area of field larger than 0.64 ha, for all farms:

$$\sum_{c,s,f \geq 1} ha_{c,f,s,g} + \frac{1-0.64}{1-0} \sum_{c,s} ha_{c,0,s,g} \geq ha.L_{100,g} + ha.L_{16,g} + ha.L_{2.56,g} + ha.L_{0.64,g}, \forall g \quad (9)$$

Last but not least, we have non-negative area constraints:

$$ha_{c,f,s,g} \geq 0, \forall c, f, s, g \quad (10)$$

2.4 Constraints relaxation and solving procedures

When the above optimization (Eq. (1)–(10)) is infeasible due to the inconsistencies among datasets, we first replaced the hard form of minimum allocated area for each farm size (Eq. (4)) with soft form (Eq. (5)) and tried the elastic factor with the following values in order: 1, 1/2, 1/4, 1/8, 1/16, 1/32, 1/64, and 0. If it was still infeasible, we relaxed the minimum area constraint required by the dominant field size distribution by removing the constraints from large to small farms until the optimization was feasible. Relaxing the minimum area constraint does not happen often during downscaling.

Whenever the above optimization becomes feasible, the optimization does not necessarily yield a unique global optimum. We calculated up to 80 (sub)optimal solutions with the same level of consistencies and averaged these solutions to get the final one. This helps us to avoid potential bias of single optimal solutions. There may be still bias on the final averaged solution because the number and quality of solutions depend on the searching process of solving the toolbox.

Each optimization problem was solved by Gurobi v9.1 using the dual simplex method with a time limit of 150 seconds. Gurobi v9.1 is a fast commercial optimization solver (Gurobi Optimization, 2021). Most of the optimization problems in this study could be solved within 60 seconds with the optimal solutions. For the administrative units containing more than 300 5-arcmin grid cells, the optimization problem becomes extremely large posing a great challenge for the solver. The number of decision variables would be more than half-million. In this case, we applied a two-tiered optimization. We first randomly divided all grid cells into several groups. Each group includes around 100 grid cells (for Russia, it was 200 to keep the number of groups below 300). We first solved the optimization problem at the group level. Then, we solved the cell level optimization for each group. Of 3421 administrative units, 244 units need to be dealt with in this way – they cover 89.4% of grid cells in this study. The whole computation was performed on a desktop computer (Intel(R) Core(TM) i7-8700 CPU @ 3.20GHz, RAM 16 GB) taking 9 days.

Finally, we masked the farm size of crops as unknown if these crops are not covered by Ricciardi's dataset. For these crops, the optimization could estimate their farm size components, but the uncertainties are significantly larger than those covered by Ricciardi's dataset.

2.5 Validation of downscaled maps and comparison with previous studies mapping farm sizes

The ideal way to validate our downscaled simultaneously farm size- and crop-specific dataset is to compare it with observations. However, most of the available datasets are not farm size specific. This creates challenges for validating our downscaled maps for all crops and farming systems. We searched for validation datasets that are global focused, farm size specific with additional information on crop or farming systems. Limited by data availability, we were able to validate our downscaled maps with two empirical datasets and we compared them with previous studies to assess the reliability of our downscaled maps. More validations are expected when more validation datasets become available.

For the first validation with empirical datasets, we compared our downscaled map with Descals et al. (2020) on oil palm map. Descals et al. (2020) developed a global gridded farm size specific oil palm map using deep learning and satellite images for

2019 (Fig. A1). With satellite images, they classified oil palm areas into small farms and large farms based on landscape features. In order to interpret this size classification, we adopted the definition of small oil palm farms by Indonesia (the world's largest palm oil producer and exporter) and used 25 ha as the threshold for the two scales (Descals et al., 2020). The validation was in five countries because only the five countries are covered by both our dataset and validation dataset(Fig. A1). The crop *Oil palm* in GAEZ v4 and SPAM2010 based map was used for validation separately. We calculated the Pearson correlation coefficient between our downscaled map and Descals et al. (2020) at grid cell level on three spatial scales using spatial moving average, 5 arcmin, 15 arcmin, and 25 arcmin.

For the second validation with empirical datasets, we compared our downscaled maps with farm size specific irrigation at the country level using the FAO RuLIS (Rural Livelihoods Information System) database (FAO, 2021). Eleven of 56 countries' micro-level household survey data around 2010 are available [S4]. Based on the household surveys, we calculated the percentage of the total irrigated area (irrigated area divided by cultivated area) for each farm size (classified by crop area) where at least 5 survey samples are available. We calculated the correlations between our estimations and household surveys. This validation considers farm size specific farming systems, with data aggregated over crops.

We also compared our downscaled map with previous studies, Lowder et al. (2016) and Mehrabi et al. (2020), which mapped the geographic distribution of farm sizes but were not crop-specific and not farm-system specific. Lowder et al. (2016) provides the percentage of harvested area operated by each farm size at the country level. Mehrabi's dataset keeps the same farm size distribution as Lowder's dataset at the country level but provides the dominant farm size per 5-arcmin grid cell. We calculated the dominant farm size from our downscaled map with the farm size that operates the largest total harvested area per grid cell for GAEZ based downscaled map and SPAM based downscaled map, respectively. The comparison was pixel-to-pixel by counting the number of cells that have similar, larger, and smaller dominant farm size in our maps compared with Mehrabi's dataset. Similar dominant farm size means the farm size in our downscaled map are the same or next to the farm size in Mehrabi's dataset.

3 Results and analysis

3.1 The crop type and farm size

With the crop map from GAEZ v4 (SPAM2010), we identified the 5-arcmin gridded harvested area for 56 countries, 11 farm sizes, 27 crops (42 crops for SPAM based map), and 2 farming systems (4 farming systems for SPAM based map). One example can be found in Fig. 2, where we illustrate the harvested area of rainfed maize belonging to two farm sizes (2-5 ha and 500-1000 ha). Overall, our results show the preference for crop groups for eleven farm sizes ([S2] for the crop group of 27 GAEZ crops). As farm size increases, oil crops and fodder crops become more popular; fruits and nuts, pulses, and roots and tubers become less popular (Fig. 3(a)). Larger farms (>20 ha) dominate the planting of fodder crops, sugar crops and oil crops; smaller farms (< 20 ha) dominate the planting of vegetables, stimulates, roots and tubers, pulses, fruits, nuts, and cotton (Fig. 3(b)). The SPAM based map shows comparable results (Fig. A2 and [S3]). These results are consistent with our datasets

Ricciardi et al. (2018b) and previous studies Herrero et al. (2017), which indicate that the optimization resulted in modest remaining inconsistency.

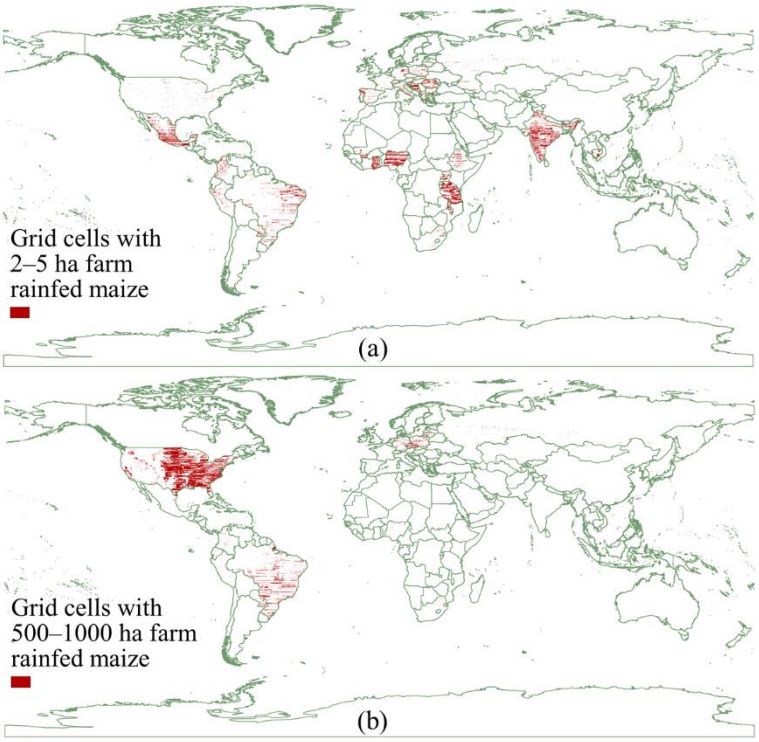

Figure 2. The gird cells with a harvested area of rainfed maize belonging to the farm size 2–5 ha (a) and farm size 500–1000 ha (b), according to the GAEZ based downscaled map.

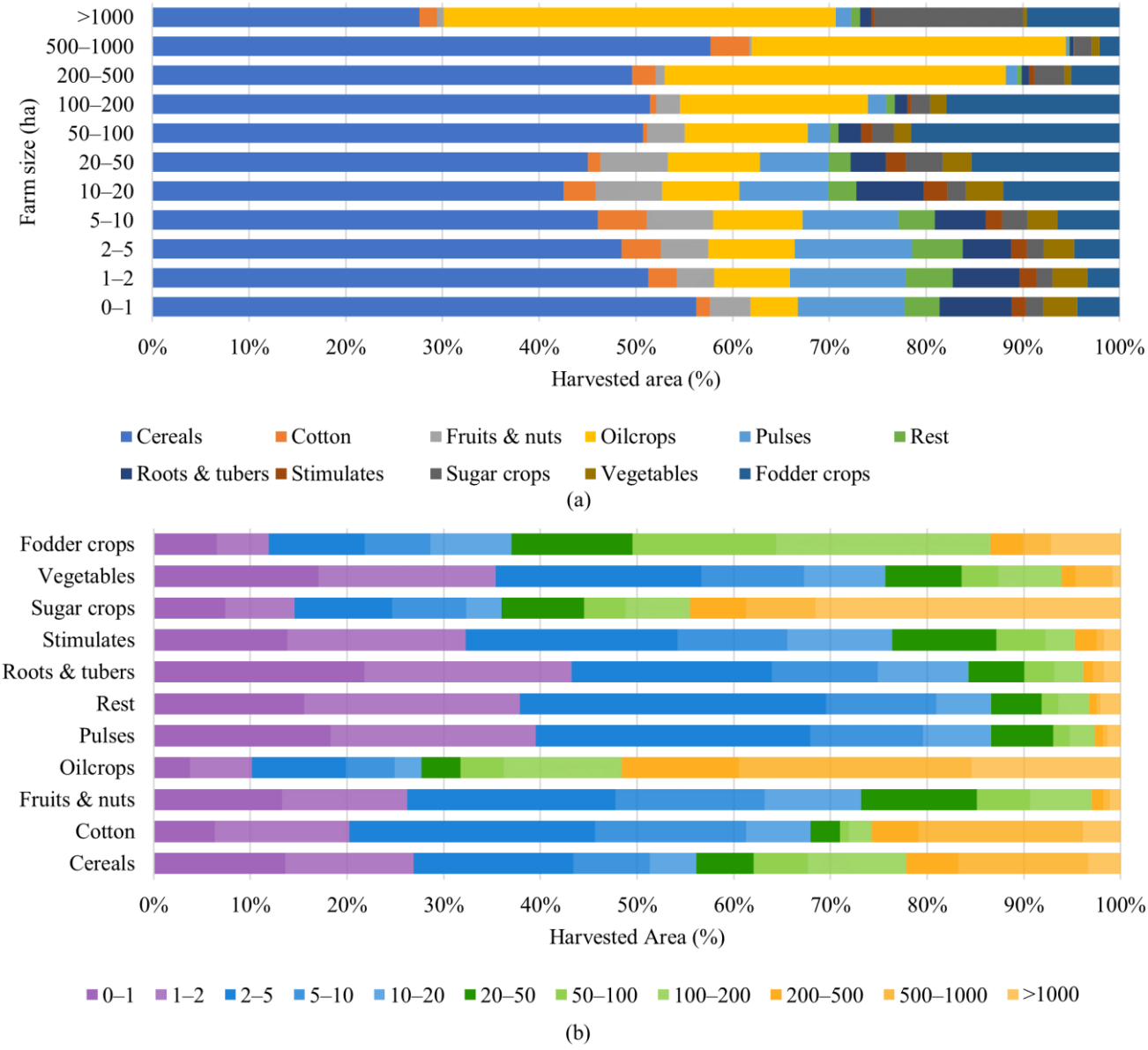

Figure 3. Harvested area of crop groups within each farm size (a) and harvested area of crop groups by farm size (b) according to GAEZ based downscaled map.

3.2 Farming systems and farm size

Comparing between irrigated and rainfed harvested area, overall, our results show that small farms irrigate a larger share of their area than large farms (Fig. 4(a)), which supports the observations of Ricciardi et al. (2020). Plausible thresholds to differentiate small and large farms can be country specific and range from 1–42 ha for most countries (Khalil et al., 2017; FAO, 2017, 2019b). With any threshold within this range, our dataset supports previous observations. The same observation

can also be found in the SPAM based downscaled map (Fig. A3). The overall observations may not hold for some countries (Sect. 3.4 for further details). The overall higher irrigation of small farms may be because most of small farms are in the severe water scarce regions (Fig. 4(b)). Here, to get water scarcity information, we overlapped our downscaled map with the annual average blue water scarcity map where water scarcity is classified as four levels: low, moderate, significant, and severe water scarcity (Mekonnen and Hoekstra, 2016; Hoekstra et al., 2012). It remains unknown whether small farms adapt to water scarcity via irrigation or irrigation of small farms increase water scarcity (Grafton et al., 2018). Another explanation for the overall higher irrigation of small farms is the country coverage. In our dataset, a large number of small farms are from Asia. Previous studies show, on average, an independent of regional water scarcity, with the percentage of irrigated area in Asian small farms being high: over 50% when water is scarce and over 20% when water is not scarce (Ricciardi et al., 2020). This percentage is much higher than that in Europe, Central Asia, Latin America, and Sub-Saharan Africa (Ricciardi et al., 2020). Thus, the overall portion of irrigated areas in small farms is high.

With water scarcity (moderate, significant, and severe), we observed that large farms irrigate to a larger extent than small farms when water is scarce, which still supports the observations of Ricciardi et al. (2020) with most thresholds to differentiate small and large farms within 1–42 ha. This observation does not depend on the relatively low irrigation extent of >1000 ha farm size since the farm size >1000 ha only contributes to less than 4.5% of water scarce area of large farms. The same observation can also be found in the SPAM based downscaled map (Fig. A3). The reason is that the water scarce area of the >1000 ha farm size is mainly contributed by limited crops from a few regions in our dataset. In this case, the characteristics of these crops and regions have more impact on the overall relationship between water scarcity and irrigation. For example, sugarcane in São Paulo, Brazil, is one of the main contributors to the significant and severe water scarce area of >1000 ha farm size. However, water scarcity is not present all year round. The level of water scarcity is low from January to June, which is the tillering phase for sugarcane. During the dry season, sugarcane is usually harvested, during which moisture in sugarcane is relatively low and the sugar is highly concentrated (Kavats et al., 2020). This may help to explain why the large farms in this area are rainfed even though under a certain level of water scarcity. Note, the main aim of Fig. 4 and Fig. A3 is to compare our dataset with previous observations instead of drawing conclusions on irrigation levels for specific farm sizes, which may need further investigation on influencing factors and uncertainties.

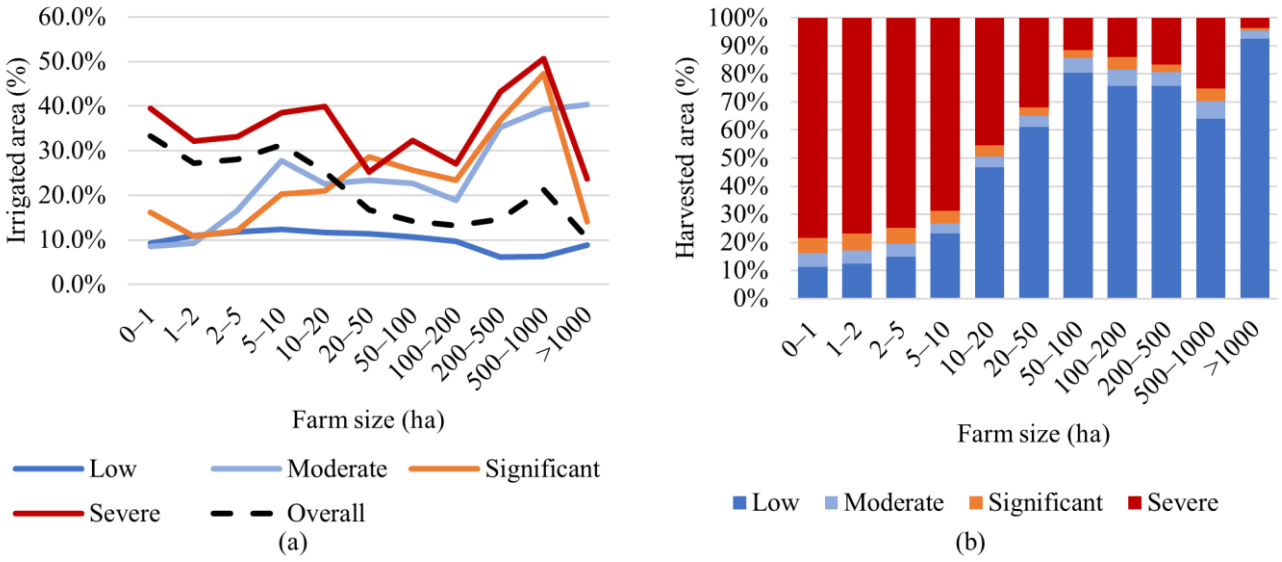

Figure 4. The percentage of the irrigated area by farm size under each water scarcity level (a) and levels of water scarcity within each farm size (b) according to GAEZ based downscaled map.

SPAM2010 further divided the rainfed farming system into low- and high-input rainfed and subsistence rainfed. With SPAM based downscaled map, our dataset indicates the subsistence and low-input rainfed farming system is mainly operated at smaller farms, but the smaller farms do not exclusively consist of subsistence and low-input rainfed farming system: they also operate a sizable portion of the irrigated and high-input rainfed area (Fig. 5). Similarly, the main type of farming system of larger farms is high-input rainfed, but the high-input rainfed is far from being limited to larger farms (Fig. 5).

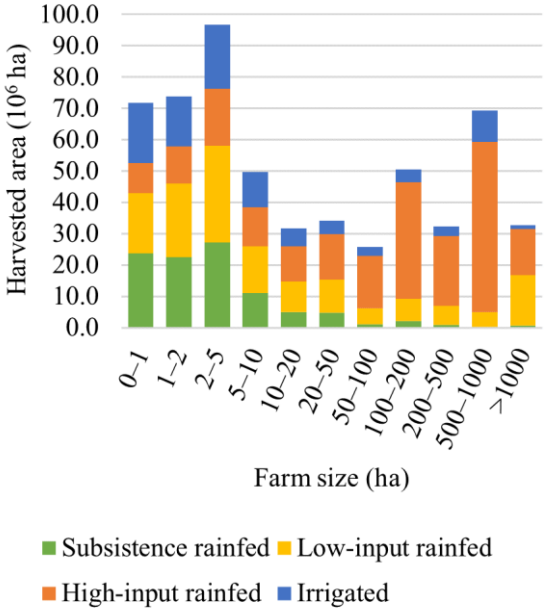

Figure 5. The distribution of irrigated, low- and high-input rainfed, and subsistence rainfed farming systems within each farm size according to the SPAM based downscaled map

3.3 Validated with farm size specific oil palm from satellite images

Validations with farm size specific oil palm data show a significant positive correlation between our downscaled maps and the validation dataset on oil palm from satellite images in most countries for both small and large farms (Table 2). At larger spatial scales, the correlation becomes stronger. This means the spatial distributions of oil palm harvested area in our downscaled maps and Descals et al. (2020) are similar. Besides the threshold of 25 ha for small and large farms, we also tried 10 ha and 50 ha as thresholds and conducted the same comparison. We found the above conclusions on oil palm comparison are not sensitive to the choice of threshold.

Still, there are some differences especially in the case of Costa Rica and the United Republic of Tanzania. Part of the above differences results from the inconsistencies between the crop maps we used and validation dataset. We compared all farms area between crop maps and validation dataset, i.e., the total area of small and large farms (Table 2). We noticed that if the cropland location in crop maps differs from the validation map (not significant positive correlation), the farm size specific validation will be poor. This means that our estimations are limited by the accuracies of farm location in crop maps. The differences between validations results for the GAEZ based map and the SPAM based map can also be attributed to the same reason, the differences in farm location between GAEZ v4 and SPAM2010.

Table 2. Pearson correlation coefficient of the harvested area between oil palms from satellite images Descals et al. (2020) and GAEZ based downscaled map and SPAM based downscaled map respectively for small farms, large farms and all farms. Since all farms results do not distinguish farm size, they indicate the differences in oil palm spatial distribution between Descals et al. (2020) and crop map datasets (GAEZ v4 and SPAM2010).

		Small farms			Large farms			All farms		
		5 arcmin	15 arcmin	25 arcmin	5 arcmin	15 arcmin	25 arcmin	5 arcmin	15 arcmin	25 arcmin
Colombia	GAEZ based	0.177*	0.313**	0.397**	0.112**	0.238**	0.334**	0.232**	0.374**	0.465**
	SPAM based	0.218**	0.547**	0.684**	0.385**	0.620**	0.701**	0.409**	0.652**	0.729**
Costa Rica	GAEZ based	0.086	0.183**	0.215**	-0.012	-0.074	-0.144**	0.032	0.001	-0.043
	SPAM based	0.836**	0.944**	0.971**	0.771**	0.891**	0.925**	0.877**	0.925**	0.929**
Brazil	GAEZ based	0.245**	0.396**	0.483**	0.177**	0.258**	0.271**	0.326**	0.398**	0.423**

	SPAM based	0.133**	0.190**	0.248**	0.087**	0.091**	0.084**	0.148**	0.154**	0.156**
United republic of Tanzania	GAEZ based	0.01	-0.109*	-0.202**	-0.011	-0.039	-0.063	0.022	-0.115*	-0.218**
	SPAM based	0.024	0.025	0.069				0.022	0.014	0.065
Peru	GAEZ based	0.172**	0.350**	0.438**	0.024	0.139**	0.237**	0.111**	0.263**	0.363**
	SPAM based	0.367**	0.389**	0.429**	0.141**	0.216**	0.240**	0.302**	0.395**	0.436**

* $p<0.005$

** $p<0.001$

3.4 Validated with farm size specific irrigation from household surveys

Our results also have positive correlations with household surveys in farm size specific irrigation for the GAEZ based map (Fig. 6(a)) and the SPAM based map (Fig. 6(b)) respectively. This means that our downscaled maps are consistent with validation data in terms of country level farm size specific irrigation. Detailed results show that the maps could capture the higher percentage of irrigated areas in small or large farms in most countries along the indications of household surveys [S5]. From the validations, we noticed that our downscaled maps systematically underestimate the extent of the irrigated area compared to household survey, both for the GAEZ based map and the SPAM based map. If we compare the percentages of irrigated area for all farms from the datasets, we find these underestimations are still there (Fig. 6(c), (d)). This means the underestimation may come from the different measurements of irrigated area and cultivated area in the validation dataset and datasets of the crop map.

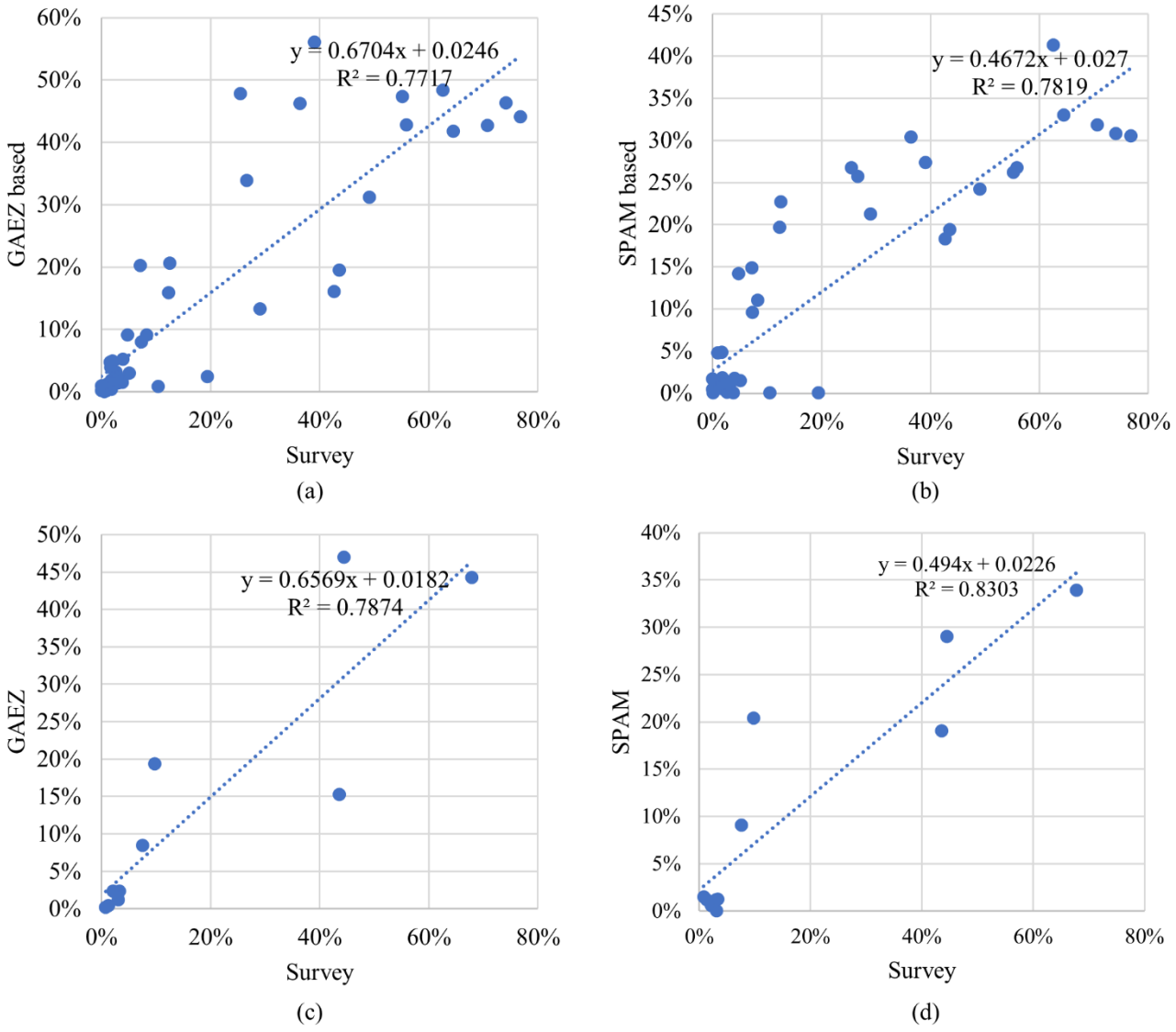

Figure 6. Correlations on the farm size specific irrigated area (% of total harvested area per farm size) between household survey and GAEZ based downscaled map (a) and SPAM based downscaled map (b) for eleven countries. The correlations on the irrigated area of all farms (% of the total harvested area) between household survey and GAEZ v4 (c) and SPAM2010 (d) are also provided.

3.5 Compared with previous studies mapping farm sizes

Compared with Lowder's dataset (Lowder et al., 2016) on the percentage of harvested area operated by each farm size, we observed positive correlations for GAEZ based map (Fig. 7(a)) and SPAM based map (Fig. 7(b)). This means at the country level, the number of farms for each farm size is similar to Lowder's dataset ([S6] for details). There are still differences between our downscaled map and Lowder's dataset. For example, Lowder's dataset estimate 78.5% of harvested area is under the farm size 50–100 ha in Bulgaria while our downscaled maps give around 5%. However, our downscaled maps estimate around 80%

of harvested are in the under farm size 100–200 ha, while Lowder's dataset gives zero. In this case, our downscaled maps are still similar to Lowder's dataset since both indicate large farms are the major farm size in the country even though it was not reflected in the correlations. These differences may be attributed to Lowder's dataset being developed for the year 2000, which is ten years earlier than our focus. Farm sizes may change during ten years in some countries. Besides reporting time, these differences may also be attributed to how different datasets harmonize farm size. The farm size classes collected from the local agriculture census usually need to be harmonized into a classification system. Different datasets may have their own choice during this process. This may lead to some differences shown in the comparison, especially when the major farm sizes are similar but not the same.

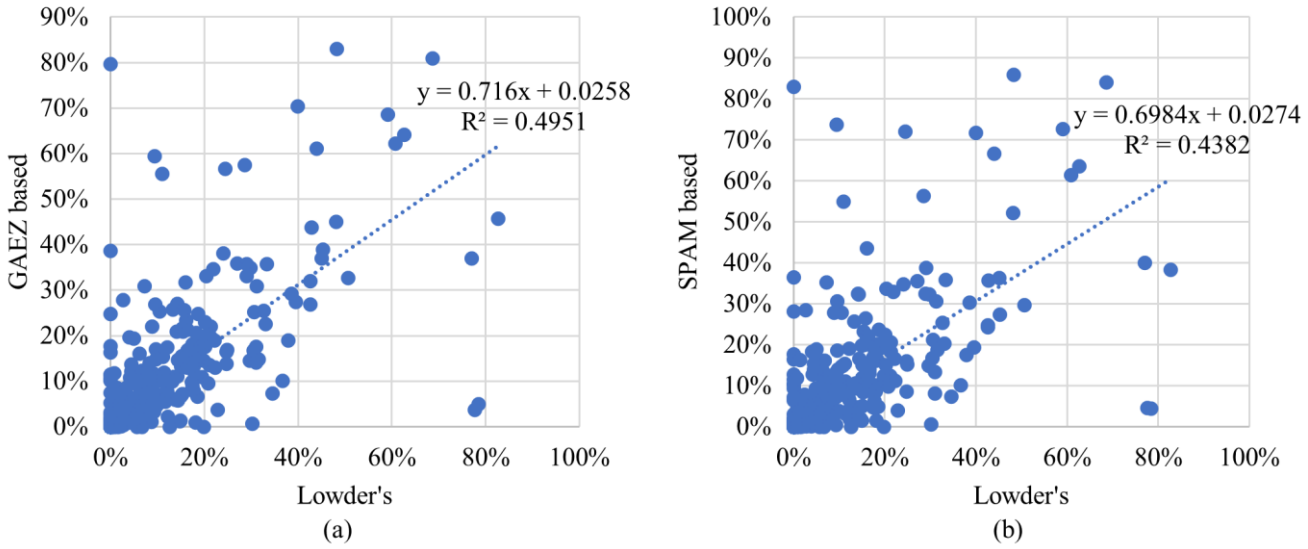

Figure 7. Correlations on the percentage of harvested area operated by each farm size between Lowder's dataset (Lowder et al., 2016) and GAEZ based downscaled map (a) and SPAM based downscaled map (b) for 37 countries and 11 farm sizes.

Compared with Mehrabi et al. (2020), the same pattern of the spatial distributions of dominant farm size could be observed in the Mehrabi's dataset (Fig. 8(a)), the GAEZ downscaled map (Fig. 8(b)), and SPAM based downscaled map (Fig. 8(c)). Overall, for GAEZ based downscaled map, 54.2% of grid cells' dominant farm sizes are similar to that in Mehrabi's dataset, 27.5% are larger, and 18.3% are smaller; for SPAM based downscaled map, 52.8% are similar, 26.0% are larger, and 21.2% are smaller ([S7] for details). These differences may be partly explained by the above comparison with Lowder's dataset since Mehrabi's dataset has the same country level farm size distribution as Lowder's dataset. Some differences could also be attributed to the comparison of dominant farm size: the dominant farm size in Mehrabi's dataset may be the second-dominant farm size in our downscaled map. The comparison of dominant farm size may magnify the difference in estimating the overall farm sizes. Since Mehrabi's dataset only include dominant farm size, it is not clear that how the difference would be estimating the overall farm sizes.

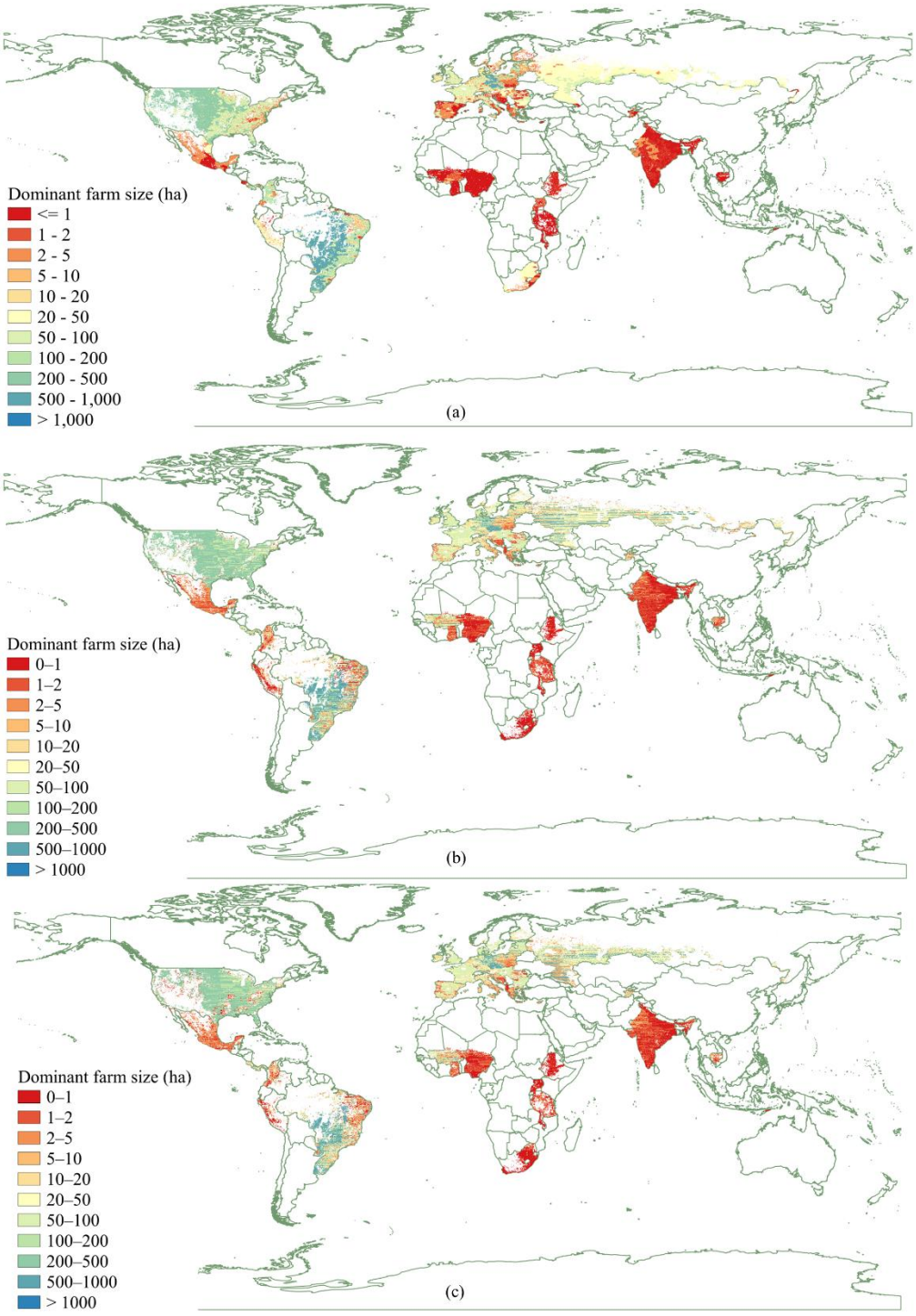

Figure 8. Dominant farm size according to Mehrabi's dataset (Mehrabi et al., 2020) (a), GAEZ based downscaled map (b) and SPAM based downscaled map (c). We only show the cells from Mehrabi's dataset where our downscaled maps have estimations.

4 Discussion

4.1 Uncertainties

We explicitly consider the uncertainties in crop maps by developing two separately downscaled maps based on two crop maps, GAEZ v4 and SPAM2010. From the results and validations, we observed some differences in the crop distribution between the two crop maps, especially at the grid cell level. This reflects the uncertainties in farmland location. It affects the spatial validations on farm size specific oil palm and the dominant farm size distribution. However, these uncertainties at the grid cell level have a limited impact on country level results and validations which can be seen from Fig. 3, Fig. 4, Fig. A2, and Fig. A3.

Uncertainty in the two crop maps is more pronounced for farming systems. From Fig. 6 and [S5] we could see the SPAM based downscaled map has a lower irrigation ratio than GAEZ based downscaled map. This is because SPAM2010 defines irrigation according to the actually irrigated area and GAEZ v4 defines irrigation by the area that is equipped with fully irrigation facilities. The lower irrigation ratio in SPAM2010 does not affect the conclusions and validations drawn from the GAEZ based map; for example, the finding of overall higher irrigation of smaller farms is robust under this uncertainty, and so is the observation on higher irrigation of larger farms under the elevated level of water scarcity.

Some uncertainties are introduced by pre-processing and constraints relaxation during the solving processes. When estimating crop-specific farm size structures using Ricciardi's datasets, around 12% of them were based on crop production instead of crop area. According to Ricciardi et al. (2018a), the introduced uncertainties are limited when using crop production. In terms of uncertainties introduced by constraints relaxation, for GAEZ (SPAM) based map, we solved 7381 (6017) optimizations. GAEZ v4 and SPAM2010 based downscaling solved different number of optimizations because of the different cropland extent which affect the number of grid cells to be allocated. Among all the optimizations, 4378 (3671) need to be relaxed using elastic factor 0.125 or smaller (Eq. (5)); 239 (203) need to be further relaxed by removing some of the minimum area constraints (Eq. (6) – (9)). Only the relaxation of minimum area constraint will introduce additional inconsistencies with the datasets used. This means the constraints relaxation introduce additional uncertainties among 3% of the total calculations. In addition, we might allocate crop area to a farm size that is not included in Ricciardi's dataset. This only happened when the crop and part of the eleven farm sizes are included in Ricciardi's dataset but meeting the minimum area constraints requires an additional farm size for the crop. In this case, the 10% relative difference with Ricciardi's dataset is still ensured for the available farm size. Only 0.1% (5.0%) of allocated area is in this case for GAEZ (SPAM) based downscaled map.

More uncertainties in the downscaled maps may come from used datasets. Since Ricciardi's dataset was not developed for 2010, farm size may change a lot in some developing countries. This put some uncertainties in our results since we relied on it to estimate farm size structure. The uncertainties in the crop map affect how we downscaled Ricciardi's dataset. Some crops can be found in Ricciardi's dataset for an administrative unit but not in crop map, or vice versa. This means that, on the one hand, 23.3% (21.6%) of the crop area in Ricciardi's dataset was not downscaled because the GAEZ v4 (SPAM2010) crop map indicates no crop. On the other hand, 17.8% (12.4%) of the harvested area in the GAEZ v4 (SPAM2010) crop map was not

allocated a farm size because Ricciardi's dataset has no relevant records. These uncertainties may have affected the allocated area in the downscaled maps, but according to validations, they are not high enough to make the downscaled maps lose the utilities. Highly accurate crop maps will reduce this part of uncertainties in the future.

Despite the uncertainties at the grid cell level, the used datasets and the downscaled maps were found to be more reliable at the country level. For example, the two crop maps were developed by downscaling the agriculture census at the (sub)national level. Collected agriculture census and social-ecological factors considered during downscaling may lead to some differences at the grid cell level in the two crop maps, while they were all adjusted to the country level data from FAOSTAT (FAO, 2019a). The dominant field size distribution is also uncertain at the grid cell level which was estimated by spatial interpolating of training samples. The uncertainty will decrease when the focus is on the regional level (Lesiv et al., 2019). Validations also show well consistencies with country level observations. Therefore, future uses of our downscaled map are more confident at the country level than grid cell level. Using GAEZ based map and SPAM based map at the same time helps to reduce uncertainties at the grid cell level.

4.2 Limitations

With the ambition to map global simultaneously farm size- and crop-specific harvested area, we were only able to cover 56 countries due to data availability, though this reflected half of the global cropland. Farm size specific data is scarce and not publicly available in some countries. The datasets we used, like Ricciardi et al. (2018b) and Lesiv et al. (2019), are the currently best-available datasets on farm or field sizes (Kim et al., 2021). Data availability is the main obstacle to creating a global map. The development of deep learning and remote sensing may help to map the global farm size- and crop-specific harvested are in another way, like the farm size specific oil palm in Descals et al. (2020). The lack of farm size training samples and enormous computational requirements are the main challenges for deep learning and remote sensing.

Our estimations are based on planted crop and harvested area, which is static for the year 2010. Farmers' choice of crop will change along with climate, market demands, and so on. Current downscaled maps could only provide a baseline for the distributions of small and large farms. It remains challenging to describe the dynamics of harvested area under changing environment.

The future updates of our downscaled maps rely on the updates of our used datasets. Fortunately, GAEZ v4, SPAM2010, and the cropland extent map have regular update plans according to their document. The dominant field size distribution was also updated since the first publication and may have more updates in the future. Ricciardi's dataset may not have updated plans but it could be updated using the data from World Programme for the Census of Agriculture (FAO, 2020b) and EUROSTAT (EUROSTAT, 2021). Any updates and extensions of Ricciardi's dataset from other data sources in the future are compliable with current model and code.

4.3 Suggestions on developing farm size- and crop-specific production dataset

Crop production of small farm is one of the main concerns of SDG 2 (Zero hunger). Developing farm size specific maps on production may be one of the applications of our dataset that directly benefits from the additional dimensionality achieved. However, compared to harvested areas, an empirical farm-size specific dataset on production or yield is even more scarce. The data on production or yield of farm sizes is available for a limited number of countries, but those countries are not always the most vulnerable in terms of food insecurity. Thus, such datasets would require estimating the production or yield based on additional models.

Current studies show the relationship between farm size and crop production or yield is complex (cf. Muyanga and Jayne (2019) and Iizumi et al. (2021)). Many factors contribute to this relationship, including but not limited to crop types, fertilizer input, climate, and soil conditions. The farm size itself does not directly affect yield, but farm size often correlates with factors that affect yield. So, estimating crop yield for different farm sizes requires first unpacking the factors that directly impact yield and correlate with farm sizes. For environmental factors like soil conditions and climate, this could be achieved by overlapping our dataset with the soil and climate database. Agricultural management and input factors, like fertilizer input, could be inferred from the agricultural production system data. Specifying agricultural management and input factors according to farming systems could help to first evaluate crop yield for different farming systems, and then allocate the yield back to farm sizes according to their proportion in each farming system. Such an approach would rely on the assumption that agricultural management practices of different farming systems do not depend on farm size. Reliable estimations of yield for different farming systems could be either derived from SPAM2010 and GAEZ v4 or based on crop modeling.

5 Code and data availability

The code, source data, and the simultaneously farm size- and crop-specific harvested area, including the GAEZ based downscaled map and SPAM based downscaled map, are open-access, free, and available at https://doi.org/10.5281/zenodo.5747616 (Su et al., 2022). The downscaled maps are available in *.csv files for each crop and farming system. Each *.csv file provides the grid cell index, administrative unit index, crop name, farm size, harvested area, and x and y coordinates in the projection of WGS84.

6 Conclusions

This study presents a 5-arcmin gridded simultaneously farm size- and crop-specific dataset of harvested area for 56 countries. We downscaled the best-available datasets, Ricciardi et al. (2018b) which collected direct reports of farm size and crop area, by using the latest datasets on crop-specific land use, cropland extent, and field size distribution. We explicitly addressed the uncertainty in crop maps by using two crop maps separately during downscaling. The downscaled maps are well-consistent with observations on farm size specific oil palm cultivation from satellite images and farm size specific irrigation from

household surveys. Our downscaled maps show the planted crops and irrigation differ among farm sizes which support previous findings. We observed uncertainties in the maps produced at the grid cell level but found country level conclusions to be robust to grid cell level uncertainties, including the uncertainties from crop maps.

Intended future updates will increase the spatial coverage. Our simultaneously farm size- and crop-specific dataset will facilitate studies to explicitly incorporate farm size into global agriculture, water resources, and climate change studies.

Appendices

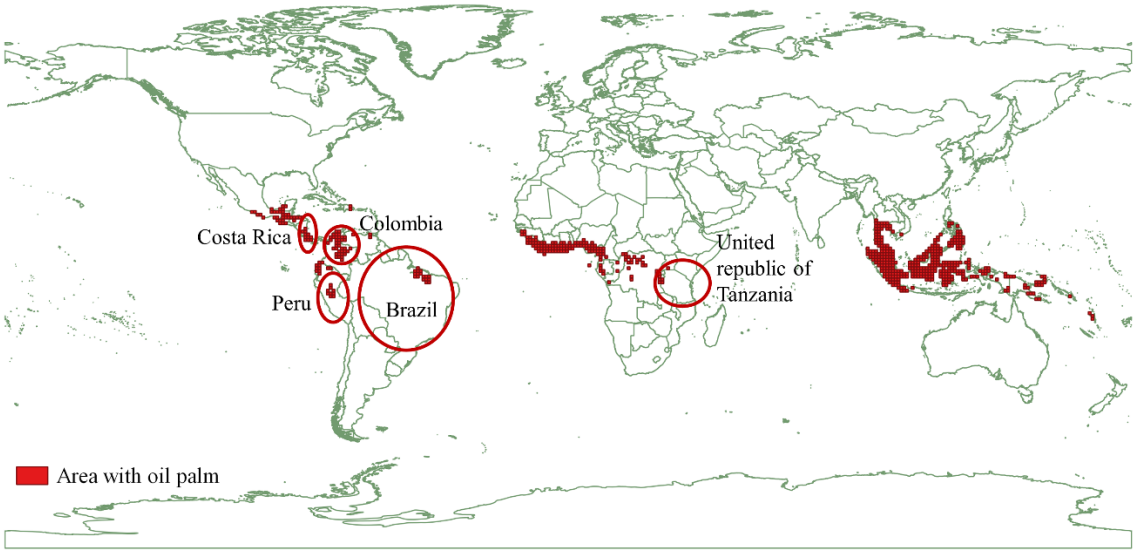

Figure A1. The global distribution of oil palms according to Descals et al. (2020) and the five countries to validate our downscaled maps.

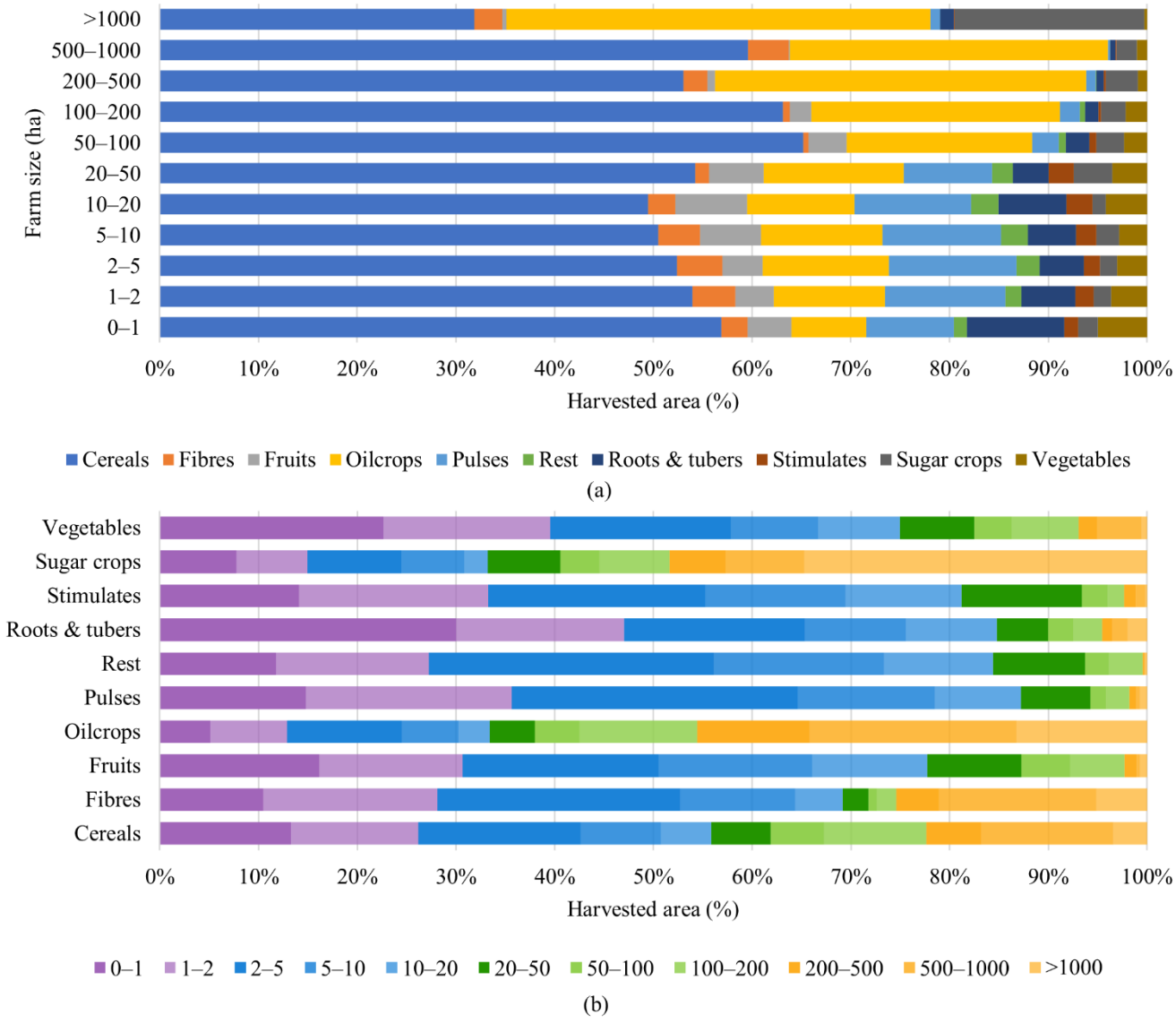

Figure A2. Harvested area of crop groups within each farm size (a) and harvested area of crop groups by farm size (b) according to SPAM based downscaled map.

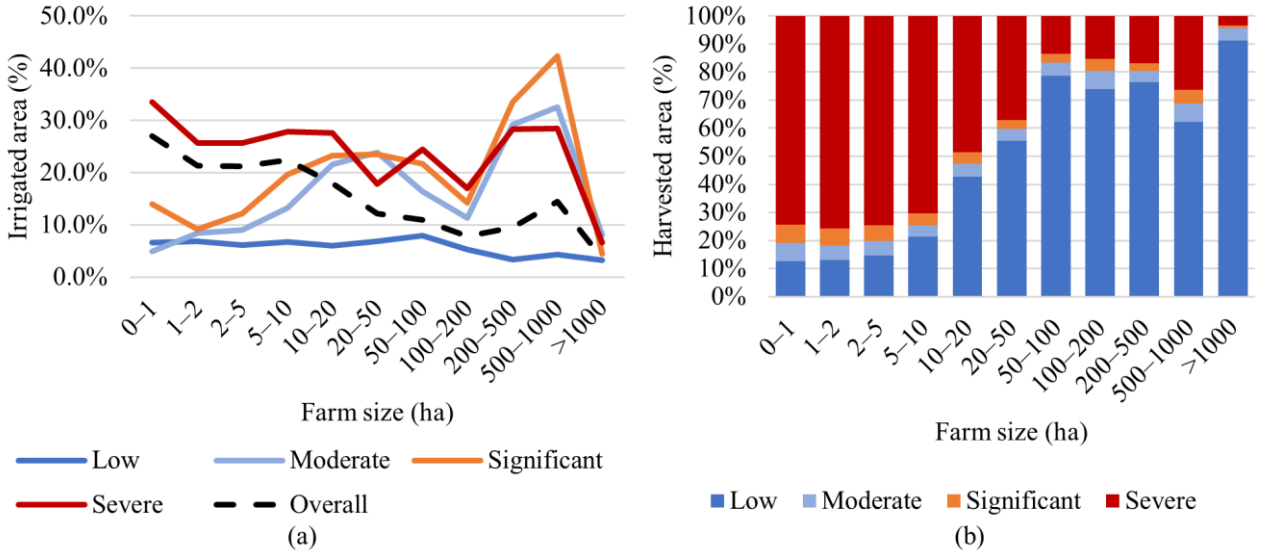

Figure A3. The percentage of the irrigated area by farm size under each water scarcity level (a) and levels of water scarcity within each farm size (b) according to SPAM based downscaled map.

Author contribution

The concept of this work originated from HS during the discussion with MSK and RJH. The study was then designed and conducted by HS under the supervision of BM and DLG with feedback from MSK and RJH. HS wrote the draft of this manuscript. HS, BM, DLG, MSK, and RJH participated in the analysis of results and revision of the manuscript.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

This study is funded in part by the European Research Council (ERC) Advanced Grant 2018 (action number 834716) and the University of Twente. Part of the research was developed in the Young Scientists Summer Program (YSSP) at the International Institute for Applied Systems Analysis, Laxenburg (Austria).

References

Bosc, P.-M., Berdegué, J., Goïta, M., van der Ploeg, J. D., Sekine, K., and Zhang, L.: Investing in smallholder agriculture for food security, HLPE, 2013.

Descals, A., Wich, S., Meijaard, E., Gaveau, D., Peedell, S., and Szantoi, Z.: High-resolution global map of smallholder and industrial closed-canopy oil palm plantations, Earth System Science Data, 2020.
EUROSTAT: EUROSTAT, Agriculture, forestry and fisheries, Africulture, Farm structure, 2021.
FAO: World programme for the census of agriculture 2020, Volume 1, Programme, concepts and definitions, FAO UN, Rome, Italy, 2015.
FAO: Small family farms data portrait. Basic information document. Methodology and data description, FAO UN, Rome, Italy, 2017.
FAO: FAOSTAT, FAO UN, Rome, Italy, 2019a.
FAO: Methodology for computing and monitoring the Sustainable Development Goal indicators 2.3.1 and 2.3.2, FAO UN, Rome, Italy, 2019b.
FAO: RuLIS Codebook, Rural Livelihoods Information System, FAO UN, Rome, Italy, 2020a.
FAO: World Programme for the Census of Agriculture, FAO UN, Rome, Italy, 2020b.
FAO: RuLIS - Rural Livelihoods Information System, FAO UN, Rome, Italy, 2021.
FAO and IIASA: Global Agro Ecological Zones version 4 (GAEZ v4), FAO UN, Rome, Italy, 2021.
Fischer, G., Nachtergaele, F. O., van Velthuizen, H., Chiozza, F., Francheschini, G., Henry, M., Muchoney, D., and Tramberend, S.: Global Agro-ecological Zones (GAEZ v4)-Model Documentation, FAO Rome, Italy, 2021.
Fritz, S., See, L., McCallum, I., You, L., Bun, A., Moltchanova, E., Duerauer, M., Albrecht, F., Schill, C., Perger, C., Havlik, P., Mosnier, A., Thornton, P., Wood-Sichra, U., Herrero, M., Becker-Reshef, I., Justice, C., Hansen, M., Gong, P., Abdel Aziz, S., Cipriani, A., Cumani, R., Cecchi, G., Conchedda, G., Ferreira, S., Gomez, A., Haffani, M., Kayitakire, F., Malanding, J., Mueller, R., Newby, T., Nonguierma, A., Olusegun, A., Ortner, S., Rajak, D. R., Rocha, J., Schepaschenko, D., Schepaschenko, M., Terekhov, A., Tiangwa, A., Vancutsem, C., Vintrou, E., Wenbin, W., van der Velde, M., Dunwoody, A., Kraxner, F., and Obersteiner, M.: Mapping global cropland and field size, Glob Chang Biol, 21, 1980-1992, 10.1111/gcb.12838, 2015.
Gollin, D.: Farm size and productivity: Lessons from recent literature, 1-35, 2019.
Grafton, R. Q., Williams, J., Perry, C. J., Molle, F., Ringler, C., Steduto, P., Udall, B., Wheeler, S. A., Wang, Y., Garrick, D., and Allen, R. G.: The paradox of irrigation efficiency, Science, 361, 748-750, 2018.
Gurobi Optimization, L.: Gurobi Optimizer Reference Manual, 2021.
Herrero, M., Thornton, P. K., Power, B., Bogard, J. R., Remans, R., Fritz, S., Gerber, J. S., Nelson, G., See, L., Waha, K., Watson, R. A., West, P. C., Samberg, L. H., van de Steeg, J., Stephenson, E., van Wijk, M., and Havlík, P.: Farming and the geography of nutrient production for human use: a transdisciplinary analysis, The Lancet Planetary Health, 1, e33-e42, 10.1016/s2542-5196(17)30007-4, 2017.
Hoekstra, A. Y., Mekonnen, M. M., Chapagain, A. K., Mathews, R. E., and Richter, B. D.: Global monthly water scarcity: blue water footprints versus blue water availability, PLoS One, 7, e32688, 10.1371/journal.pone.0032688, 2012.
Iizumi, T., Hosokawa, N., and Wagai, R.: Soil carbon-food synergy: sizable contributions of small-scale farmers, CABI Agriculture and Bioscience, 2, 43, 10.1186/s43170-021-00063-6, 2021.
Kavats, O., Khramov, D., Sergieieva, K., and Vasyliev, V.: Monitoring of sugarcane harvest in Brazil based on optical and SAR data, Remote Sensing, 12, 1-26, 10.3390/rs12244080, 2020.
Khalil, C. A., Conforti, P., Ergin, I., and Gennari, P.: Defining small scale food producers to monitor target 2.3 of the 2030 Agenda for Sustainable Development, FAO, Rome, 2017.
Kim, K.-H., Doi, Y., Ramankutty, N., and Iizumi, T.: A review of global gridded cropping system data products, Environmental Research Letters, 16, 10.1088/1748-9326/ac20f4, 2021.
Latham, J., Cumani, R., Rosati, I., and Bloise, M.: Global land cover share (GLC-SHARE) database beta-release version 1.0-2014, FAO, Rome, Italy, 2014.
Lesiv, M., Laso Bayas, J. C., See, L., Duerauer, M., Dahlia, D., Durando, N., Hazarika, R., Kumar Sahariah, P., Vakolyuk, M., Blyshchyk, V., Bilous, A., Perez-Hoyos, A., Gengler, S., Prestele, R., Bilous, S., Akhtar, I. U. H., Singha, K., Choudhury, S. B., Chetri, T., Malek, Z., Bungnamei, K., Saikia, A., Sahariah, D., Narzary, W., Danylo, O., Sturn, T., Karner, M., McCallum, I., Schepaschenko, D., Moltchanova, E., Fraisl, D., Moorthy, I., and Fritz, S.: Estimating the global distribution of field size using crowdsourcing, Glob Chang Biol, 25, 174-186, 10.1111/gcb.14492, 2019.
Lowder, S. K., Sánchez, M. V., and Bertini, R.: Which farms feed the world and has farmland become more concentrated?, World Development, 142, 10.1016/j.worlddev.2021.105455, 2021.
Lowder, S. K., Skoet, J., and Raney, T.: The Number, Size, and Distribution of Farms, Smallholder Farms, and Family Farms Worldwide, World Development, 87, 16-29, 10.1016/j.worlddev.2015.10.041, 2016.
Lu, M., Wu, W., You, L., See, L., Fritz, S., Yu, Q., Wei, Y., Chen, D., Yang, P., and Xue, B.: A cultivated planet in 2010 – Part 1: The global synergy cropland map, Earth System Science Data, 12, 1913-1928, 10.5194/essd-12-1913-2020, 2020.
Mehrabi, Z., McDowell, M. J., Ricciardi, V., Levers, C., Martinez, J. D., Mehrabi, N., Wittman, H., Ramankutty, N., and Jarvis, A.: The global divide in data-driven farming, Nature Sustainability, 4, 154-160, 10.1038/s41893-020-00631-0, 2020.
Mekonnen, M. M. and Hoekstra, A. Y.: Four billion people facing severe water scarcity, Science Advances, 2, 10.1126/sciadv.1500323, 2016.
Meyfroidt, P.: Mapping farm size globally: benchmarking the smallholders debate, Environmental Research Letters, 12, 10.1088/1748-9326/aa5ef6, 2017.

Monfreda, C., Ramankutty, N., and Foley, J. A.: Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000, Global Biogeochemical Cycles, 22, n/a-n/a, 10.1029/2007gb002947, 2008.
Muyanga, M. and Jayne, T. S.: Revisiting the Farm Size-Productivity Relationship Based on a Relatively Wide Range of Farm Sizes: Evidence from Kenya, 101, 1140-1163, https://doi.org/10.1093/ajae/aaz003, 2019.
Noack, F., Larsen, A., Kamp, J., and Levers, C.: A bird's eye view of farm size and biodiversity: The ecological legacy of the iron curtain, American Journal of Agricultural Economics, 10.1111/ajae.12274, 2021.
Ray, D. K., Mueller, N. D., West, P. C., and Foley, J. A.: Yield trends are insufficient to double global crop production by 2050, PloS one, 8, e66428, 2013.
Ren, C., Liu, S., van Grinsven, H., Reis, S., Jin, S., Liu, H., and Gu, B.: The impact of farm size on agricultural sustainability, Journal of Cleaner Production, 220, 357-367, 10.1016/j.jclepro.2019.02.151, 2019.
Ricciardi, V., Mehrabi, Z., Wittman, H., James, D., and Ramankutty, N.: Higher yields and more biodiversity on smaller farms, Nature Sustainability, 4, 651-657, 10.1038/s41893-021-00699-2, 2021.
Ricciardi, V., Ramankutty, N., Mehrabi, Z., Jarvis, L., and Chookolingo, B.: An open-access dataset of crop production by farm size from agricultural censuses and surveys, Data Brief, 19, 1970-1988, 10.1016/j.dib.2018.06.057, 2018a.
Ricciardi, V., Ramankutty, N., Mehrabi, Z., Jarvis, L., and Chookolingo, B.: How much of the world's food do smallholders produce?, Global Food Security, 17, 64-72, 2018b.
Ricciardi, V., Wane, A., Sidhu, B. S., Godde, C., Solomon, D., McCullough, E., Diekmann, F., Porciello, J., Jain, M., and Randall, N.: A scoping review of research funding for small-scale farmers in water scarce regions, Nature Sustainability, 3, 836-844, 2020.
Riesgo, L., Louhichi, K., Gomez y Paloma, S., Hazell, P., Ricker-Gilbert, J., Wiggins, S., Sahn, D. E., and Mishra, A. K.: Food and nutrition security and role of smallholder farms: challenges and opportunities, Institute for Prospective Technological Studies; Information for Meeting Africa's Agricultural Transformation and Food Security Goals (IMAAFS),
Rudra, A.: Farm size and yield per acre, Economic Political Weekly, 1041-1044, 1968.
Samberg, L. H., Gerber, J. S., Ramankutty, N., Herrero, M., and West, P. C.: Subnational distribution of average farm size and smallholder contributions to global food production, Environmental Research Letters, 11, 10.1088/1748-9326/11/12/124010, 2016.
Savastano, S. and Scandizzo, P.: Farm Size and Productivity: A "Direct-Inverse-Direct" Relationship, 2017.
Su, H., Willaarts, B., Luna Gonzalez, D., S. Krol, M., and J. Hogeboom, R.: Gridded 5 arcmin farm-size specific and crop specific harvested area for 56 countries, https://doi.org/10.5281/zenodo.5747616, 2022.
Yu, Q., You, L., Wood-Sichra, U., Ru, Y., Joglekar, A. K. B., Fritz, S., Xiong, W., Lu, M., Wu, W., and Yang, P.: A cultivated planet in 2010 – Part 2: The global gridded agricultural-production maps, Earth System Science Data, 12, 3545-3572, 10.5194/essd-12-3545-2020, 2020.