



OPEN

# CROPGRIDS: a global geo-referenced dataset of 173 crops

DATA DESCRIPTOR

Fiona H. M. Tang<sup>1,2</sup>, Thu Ha Nguyen<sup>3</sup>, Giulia Conchedda<sup>4</sup>, Leon Casse<sup>3,4</sup>,  
Francesco N. Tubiello<sup>4</sup> & Federico Maggi<sup>3,5</sup> ✉

CROPGRIDS is a comprehensive global geo-referenced dataset providing area information for 173 crops for the year 2020, at a resolution of 0.05° (about 5.6 km at the equator). It represents a major update of the Monfreda *et al.* (2008) dataset (hereafter MRF), the most widely used geospatial dataset previously available, covering 175 crops with reference year 2000 at 10 km spatial resolution. CROPGRIDS builds on information originally provided in MRF and expands it using 27 selected published gridded datasets, subnational data of 52 countries obtained from National Statistical Offices, and the 2020 national-level statistics from FAOSTAT, providing more recent harvested and crop (physical) areas for 173 crops at regional, national, and global levels. The CROPGRIDS data advance the current state of knowledge on the spatial distribution of crops, providing useful inputs for modelling studies and sustainability analyses relevant to national and international processes.

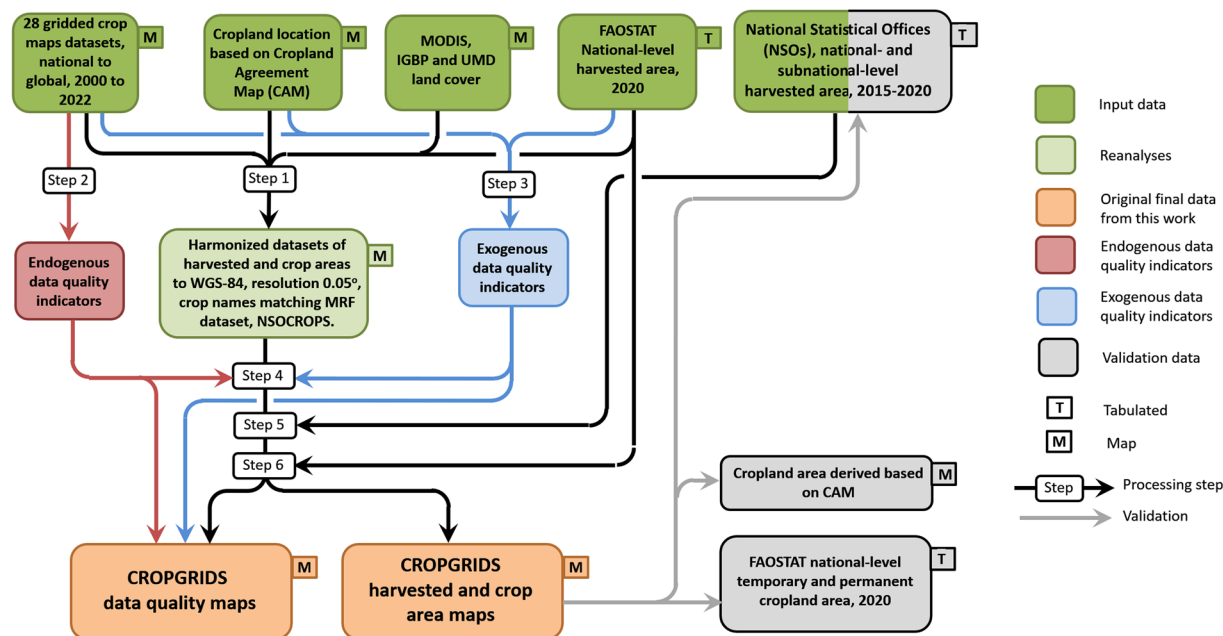
## Background & Summary

Detailed global geospatial information on the distribution of crop types over time is required to understand planetary boundaries and support decision-making at all scales, from land use change dynamics to the impacts of agricultural inputs on the environment. Geo-referenced crop information is particularly valuable for improving reporting and monitoring progress at sub-national scales under the Sustainable Development Goals (SDG), in particular Goal 2 indicators on the productivity and sustainability of agriculture<sup>1</sup>.

The most comprehensive geospatial product available today, covering 175 crops at a resolution of about 10 km globally<sup>2</sup>—henceforth referred to herein as MRF from the initials of the authors—provides however dated information, limited to the year 2000, whereas significant changes in cropland extent have been documented over the past twenty years<sup>3,4</sup>. MRF was created by spatially disaggregating official national and sub-national harvested area information obtained from various sources, over a gridded cropland map derived from remote sensing. It has since been used in several published studies, most notably for assessing planetary boundaries with respect to food and agriculture<sup>5</sup>. Several crop type mapping efforts were made since the production of MRF (see ref. <sup>6</sup> for a comprehensive review). More recently, important initiatives such as those promoted by the European Space Agency (ESA)<sup>7,8</sup>, by the USA National Aeronautics and Space Administration (NASA)<sup>9</sup>, and by the G20 Ministers of Agriculture were launched and are already contributing considerable new information<sup>10–13</sup>. However, none of these efforts has matched the original MRF scope and crop coverage, so much so that many global assessments of agricultural impacts have continued to use MRF as a reference<sup>14–17</sup>.

To update the MRF information, we produced CROPGRIDS, a new global gridded harvested and crop (physical) area geospatial dataset for 173 crops for the year 2020. CROPGRIDS was produced using a similar approach as in ref. <sup>13</sup> with MRF data used as starting point and updated through hybridisation of more recent information, by merging all available, published and gridded datasets for periods more recent than 2000 and using a set of endogenous and exogenous data quality indicators, within a multi-criteria ranking scheme, to determine best-fit data by crop type and country. For some crop types and countries where gridded data more recent than 2000 were not available, we spatialized recent subnational data obtained from National Statistical Offices (NSOs) following a similar algorithm used in MRF but with a new cropland agreement map circa 2020<sup>18</sup> as the cropland mask. The resulting CROPGRIDS is a novel synthesis of the most recently available information on harvested

<sup>1</sup>School of Environmental and Rural Science, University of New England, Armidale, New South Wales, 2351, Australia. <sup>2</sup>Department of Civil Engineering, Monash University, Clayton, 3800, Victoria, Australia. <sup>3</sup>Environmental Engineering, School of Civil Engineering, The University of Sydney, Sydney, New South Wales, Australia. <sup>4</sup>Statistics Division, Food and Agriculture Organization of the United Nations, Viale delle Terme di Caracalla, Rome, 00153, Italy. <sup>5</sup>Sydney Institute of Agriculture, The University of Sydney, Sydney, NSW, 2006, Australia. ✉e-mail: [federico.maggi@sydney.edu.au](mailto:federico.maggi@sydney.edu.au)



**Fig. 1** Workflow of the development of CROPGRIDS. Step 1: Input data harmonization; Step 2: computation of endogenous data quality indicators; Step 3: computation of exogenous data quality indicators; Step 4: assemblage of global maps; Step 5: gap filling of crop geographic distribution; and Step 6: data adjustment.

and crop area maps for 173 crops, at a global spatial resolution of  $0.05^\circ$  (approximately 5.6 km at the equator). Crop type name, harvested area and crop area definitions used in CROPGRIDS are aligned to the relevant FAO commodities and land use definitions<sup>19</sup>.

## Methods

The development of CROPGRIDS involves several steps that were carried out either sequentially or in parallel (Fig. 1) as follows: Step 1) input data harmonization; Step 2) computation of endogenous data quality indicators; Step 3) computation of exogenous data quality indicators; Step 4) assemblage of global maps; Step 5) gap filling of crop geographic distribution; and Step 6) data adjustment to FAOSTAT. These steps are described in detail in the next sections. While Step 1) to 3) are mostly data curation and pre-calculations for later steps, Step 4), at the core of the workflow, was achieved through a multi-criteria ranking scheme designed using the endogenous and exogenous data quality indicators to select, for countries and territories for which data were available from multiple input datasets, the one dataset best describing a specific crop. Similarly important, Steps 5) and 6) were next used to update missing information using existing independent subnational statistics and adjust the assembled data maps to the FAOSTAT reference year 2020.

**Input data.** We conducted a search for published peer-reviewed datasets providing geo-referenced crop-specific information, including, by grid cell: amount of harvested area (*HA*); amount of crop area (*CA*); fractional crop area (*f*, *i.e.*, proportion of the grid cell area occupied by crop type); or binary values (*w*), specifying whether a grid cell was cultivated with a specific crop type or not. The following four criteria were applied for inclusion of a dataset: (1) reference year later than 2000; (2) at least one crop species also present in MRF; (3) geospatial coverage for at least one country (complete national extent); and (4) spatial resolution at least  $0.083^\circ$  (about 10 km at the equator). Based on these criteria, we created a library of 28 datasets, including 14 national, 8 multinational/continental, and 6 global datasets (Table 1). Amongst the selected datasets, two provided both *HA* and *CA*, two provided *HA*, three provided *f*, and 21 provided *w* (see details in Table 1). The information collected to build CROPGRIDS spanned the period 2000–2021, with 25 out of the 28 input datasets referring to the period 2015–2021. These datasets provided diverse variables and had different resolution from the target resolution of  $0.05^\circ$  per grid cell for CROPGRIDS. Hence, a number of steps were undertaken to harmonize the input datasets as described in the section below.

Additionally, we used the following datasets for data processing: a cropland agreement map (CAM) circa year 2020 at 30 m resolution<sup>18</sup>; the MODIS land use maps for year 2020 at 500 m resolution<sup>20</sup>; the FAO Global Administrative Unit Layers (GAUL) dataset<sup>21</sup> (FAO, 2015); subnational statistics from various NSOs (see Supplementary Table S1) and FAOSTAT national statistics of harvested area<sup>19</sup>. CAM provides geospatial statistics of cropland areas, generated based on six open-access high-resolution remote sensing products<sup>18</sup>.

**Input data harmonization (Step 1).** We first determined *HA* and *CA* for each crop type by grid cell for each input datasets listed in Table 1 as follows.

	Acronym	Description	Reference
1	MRF	Global gridded <i>HA</i> [ha] for 175 crops at a resolution of 0.0833 degree (~10 km at the equator) in 2000.	2
2	SPAM	Global gridded <i>HA</i> and <i>CA</i> [ha] for 38 crops at a resolution of 0.0833 degree (~10 km at the equator) in 2010. Only 30 crops considered for this work.	29
3	GAEZ + 2015	Global gridded <i>HA</i> [ha] for 26 crops at a resolution of 0.0833 degree (~10 km at the equator) in 2015. Only 20 crops considered.	30
4	GEOGLAM	Global gridded <i>f</i> [%] for 4 crops at a resolution of 0.05 degree (~5.6 km at the equator) in 2020.	13,31
5	OIPA	Global gridded <i>w</i> [-] for oil palm at a resolution of 0.0000898 degree (~0.01 km at the equator) in 2019.	32
6	RAP	Global gridded <i>w</i> [-] for rapeseed at a resolution of 0.0000898 degree (~0.01 km at the equator) in 2019.	33
7	EU	Gridded <i>w</i> [-] for 28 countries in EU for 17 crops at a resolution of 0.0000898 degree (~0.01 km at the equator) in 2018. Only 12 crops considered.	34
8	SPAMAF	Gridded <i>HA</i> and <i>CA</i> [ha] for Africa for 42 crops at a resolution of 0.0833 degree (~10 km at the equator) in 2017. Only 34 crops considered.	35
9	AFCAS	Gridded <i>f</i> [ha km <sup>-2</sup> ] for Africa for cassava at a resolution of 0.00833 degree (~1 km at the equator) in 2014.	36
10	SASOY	Gridded <i>w</i> [-] for South America for soybean at a resolution of 0.00025 degree (~0.03 km at the equator) in 2018.	37
11	MYSTHA	Gridded <i>w</i> [-] for Malaysia, Indonesia, and Thailand for oil palm at a resolution of 0.0002695 degree (~0.03 km at the equator) in 2017.	38
12	ASIARICE	Gridded <i>w</i> [-] and cropping intensity for 21 countries in Asian monsoon region for rice at a resolution of 0.0045 degree (~0.5 km at the equator) in 2020.	39
13	CIVGHA	Gridded <i>w</i> [-] for Cote d'Ivoire and Ghana for cocoa at a resolution of 0.0000898 degree (~0.01 km at the equator) in 2019.	40
14	UZBTJK	Gridded <i>w</i> [-] for Uzbekistan and Tajikistan for 38 crops distributed as shapefile at a resolution of 0.0001 degree (~0.01 km at the equator) in 2015 to 2018. Only 20 crops considered.	41
15	USA	Gridded <i>w</i> [-] for USA for 105 crops at a resolution of 0.0000898 degree (~0.01 km at the equator) in 2021. Only 64 crops considered.	42
16	CA	Gridded <i>w</i> [-] for Canada for 52 crops at a resolution of 0.00027 degree (~0.03 km at the equator) in 2021. Only 31 crops considered.	43
17	AFG	Gridded <i>w</i> [-] for Afghanistan for 6 crops at a resolution of 0.0000898 degree (~0.01 km at the equator) in 2020. Only 3 crops considered.	44
18	DEU	Gridded <i>w</i> [-] for Germany for 24 crops at a resolution of 0.0000898 degree (~0.01 km at the equator) in 2019. Only 15 crops considered.	45
19	CHNWH	Gridded <i>w</i> [-] for China for winter wheat at a resolution of 0.0003 degree (~0.03 km at the equator) for 2018.	46
20	CHNMZ	Gridded <i>w</i> [-] for China for maize at a resolution of 0.005 degree (~0.56 km at the equator) in 2017.	47
21	CHNMZWHRI	Gridded <i>w</i> [-] of single, double, triple cropping for China for rice, maize, and wheat at a resolution of 0.005 degree (~0.56 km at the equator) in 2020.	48
22	BGDRICE	Gridded <i>w</i> [-] of 3 growing seasons for Bangladesh for rice at a resolution of 0.0000898 degree (~0.01 km at the equator) in 2017.	49
23	BRA	Gridded <i>w</i> [-] for Brazil for sugarcane at a resolution of 0.0003 degree (~0.03 km at the equator) in 2019.	50
24	SEN	Gridded <i>w</i> [-] for Senegal for 22 crops at a resolution of 0.00009 degree (~0.01 km at the equator) in 2018. Only 17 crops considered.	51
25	AU	Gridded <i>f</i> [-] for Australia for 25 crops at a resolution of 0.0833 degree (~10 km at the equator) in 2015. Only 6 crops considered.	52
26	FR	Gridded <i>w</i> [-] for France for 11 crops at a resolution of 0.0001 degree (~0.01 km at the equator) in 2021. Only 5 crops considered.	53
27	JP	Gridded <i>w</i> [-] for Japan for rice at a resolution of 0.0000833 degree (~0.01 km at the equator) in 2020.	54
28	CHNMZSOY	Gridded <i>w</i> [-] for China for maize and soybean at a resolution of 0.0000833 degree (~0.01 km at the equator) in 2019.	55

**Table 1.** CROPGRIDS input datasets. Variables provided in datasets are: *HA*, harvested area; *CA*, crop (physical) area; *f*, fractional crop area; and *w*, binary value for a crop existence in a grid cell. Aggregated crops that cannot be matched against the crop list in MRF were excluded. The full list of crop name matching is provided in Supplementary Table S2.

When only *HA* data are given (i.e., MRF and GAEZ + 2015, Table 1), we imputed  $CA = HA$  for permanent crops and  $CA = \min\{HA, LA, CAM95\}$  for temporary crops (see list of permanent and temporary crops in Supplementary Table S3), with *LA* being the area of land available for cropping, i.e., grid cell area (*GA*) excluding water bodies, wetlands, urban and built-up lands, permanent snow and ice, and barren land following land use classification in MODIS<sup>20</sup>, and *CAM95* being the 95<sup>th</sup> percentile surface area calculated from CAM. When only *f* was provided (i.e., GEOGLAM, AFCAS, and AU), we calculated crop area as  $CA = f \times GA$  and imputed  $HA = CA$ . Next, the computed and original georeferenced maps of *HA* and *CA* in MRF, SPAM, GAEZ + 2015, and SPAMAF datasets were harmonized to a common spatial resolution of 0.05° (approximately 5.6 km at the equator) using the *imresize* function<sup>22</sup> in Matlab with bilinear interpolation and pynophylactic methods to ensure areal conservation, and a bounding box of -180° to 180° longitude and -90° to 90° latitude using the WGS-84 coordinate system (World Geodetic System 1984).

When only *w* was provided, we first derived corresponding *f* values and then made the same imputations as above. Specifically, since datasets providing *w* for individual crops had typically high spatial resolution (ranging

10–550 m at the equator), we performed pixel counting of  $w$  values to derive  $f$  values at the required 5.6 km resolution (i.e.,  $0.05^\circ$  per grid cell). Additionally, for datasets providing  $w$  values of a specific crop over multiple growing seasons  $s$ , the annual  $CA$  of that crop was computed as  $CA = \max\{CA_1, CA_2, \dots, CA_s\}$  across the seasons  $s$ ; while  $HA$  was computed as  $HA = \sum_{i=1}^s CA_i$ . Alternatively, when the geo-referenced cropping intensity  $CI$  was provided (i.e., ASIARICE, Table 1), then we calculated  $HA = CA \times CI$ .

Finally, we set a threshold for  $CA$  and  $HA$  values, i.e., both were set to zero whenever  $CA < 100 \text{ m}^2$ . This lower bound corresponds to the finest spatial resolutions of all input datasets and was set to prevent from accounting of unrealistically small agriculture parcels. Consistency diagnostics checked that  $CA \leq HA$ ,  $CA \leq LA \leq GA$ , and  $CI \leq 3$  (i.e.,  $CI$  commonly less than  $3^{23}$ ) were satisfied in all grid cells for individual crops.

In building CROPGRIDS, we also harmonized crop names in the input datasets, including performing aggregations where needed, to correspond to the crop names in MRF, thus ensuring internal consistency and alignment with FAO crop classifications following the Indicative Crop Classification (ICC) of the World Programme for the Census of Agriculture<sup>24</sup> (Supplementary Table S2).

**Compute endogenous data quality indicators (Step 2).** Endogenous data quality indicators assessed both quantitative and qualitative features of a dataset that do not depend on external information. These endogenous features included: synchrony ( $Q_y$ ), administration ( $Q_a$ ), data source ( $Q_s$ ), validation ( $Q_v$ ), resolution ( $Q_r$ ), maturity ( $Q_m$ ), and type of dispatch ( $Q_d$ ). All endogenous features were assigned an indicator value ranging 0–1, with the end points corresponding to the lowest and highest quality, respectively. Endogenous features were not expressed as geo-referenced maps, but rather we used them to tag individual input datasets regardless of crop type.

$Q_y$  described the level of synchrony between the year of reference  $Y_r$  of a dataset and the year of reference of CROPGRIDS, which was set to 2020. Specifically, datasets with  $Y_r$  departing from 2020 were assigned a lower rank than those in 2020 as

$$Q_y = \begin{cases} \frac{Y_r - 2000}{2020 - 2000} & \text{if } Y_r < 2020 \\ 1 & \text{if } Y_r = 2020 \\ 1 - \frac{Y_r - 2020}{2040 - 2020} & \text{if } Y_r > 2020 \end{cases} \quad (1)$$

$Q_a$  described the administrative domain of a dataset (i.e., national to global). A national dataset was assigned a higher  $Q_a$  value than global datasets, under our assumption that national datasets are constructed using better information from direct local knowledge.  $Q_a$  was defined as:

$$Q_a = \begin{cases} 1 & \text{if national or multi-national} \\ 0.5 & \text{if global.} \end{cases} \quad (2)$$

$Q_s$  described the primary data source used to develop a dataset. We assumed that datasets developed using survey data (i.e., field survey and censuses) have higher quality than those based on satellite imagery, with datasets constructed using modelling techniques having the lowest quality. We used  $Q_s$  to also account for hybrid methods, assigning in such cases intermediate quality scores, as follows:

$$Q_s = \begin{cases} 1 & \text{survey, satellite, model integration} \\ 0.8 & \text{survey and satellite integration} \\ 0.7 & \text{survey and model integration} \\ 0.5 & \text{satellite and model integration} \\ 0.5 & \text{survey only} \\ 0.3 & \text{satellite only} \\ 0.2 & \text{model only} \end{cases} \quad (3)$$

$Q_v$  was used to rank the level of validation of a dataset, against ground truth, users' feedback, statistical data, satellite images or other sources. We ranked the validation level from high to low based on the presence of field observations, the number of sources used for validation, and the separation between calibration and validation sets.  $Q_v$  was defined as:

$$Q_v = \begin{cases} 1, & \text{if validated using groundtruth data with sound statistical approaches} \\ 0.5, & \text{if validated using statistics, users' feedback or satellite images} \\ 0, & \text{if no attempt of validation.} \end{cases} \quad (4)$$

$Q_r$  described the spatial resolution  $r$  of a dataset. A higher rank was given to a dataset with finer resolution:

$$Q_r = 1 - \frac{r - r_{min}}{r_{max} - r_{min}}, \quad (5)$$

Dataset	Synchrony $Q_y$	Administration $Q_a$	Source $Q_s$	Validation $Q_v$	Resolution $Q_r$	Maturity $Q_m$	Dispatch $Q_d$	$\sum_k Q_k$
<sup>(a)</sup> MRF	0	0.5	0.5	0.5	0	0	0.5	2
SPAM	0.667	0.5	0.7	0.5	0	0.5	0.5	3.367
GAEZ + 2015	1	0.5	0.5	0.5	0	0	0.5	3
GEOGLAM	0.867	0.5	0.5	0.5	0.400	0	0.5	3.267
OIPA	1	0.5	0.5	1	1.000	0	0.5	4.5
RAP	1	0.5	0.3	0.5	1.000	0	0.5	3.8
EU	1	1	0.8	1	1.000	0	0.5	5.3
SPAMAF	1	1	0.7	0	0.004	0.5	0.5	3.704
AFCAS	0.933	1	0.5	0	0.901	0	0.5	3.834
SASOY	0.933	1	0.8	1	0.998	0	0.5	5.231
MYSTHA	1	1	0.3	0.5	0.998	0	0.5	4.298
ASARICE	1	1	0.3	0.5	0.947	0	0.5	4.247
CIVGHA	1	1	0.8	1	1.000	0	0.5	5.3
UZBTJK	1	1	0.5	1	1.000	0	0.5	5
USA	0.933	1	0.8	1	0.998	1	1	6.731
CA	1	1	0.8	1	0.998	1	1	6.798
AFG	1	1	0.8	1	1.000	0	0.5	5.3
DEU	1	1	1	0.5	1.000	0	0.5	5
CHNWH	1	1	0.8	1	0.997	0	0.5	5.297
CHNMZ	1	1	0.3	1	0.937	0	0.5	4.737
CHNMZWHRI	1	1	0.3	1	0.941	0	0.5	4.741
BGDRICE	1	1	0.8	1	1.000	0	0.5	5.3
BRA	1	1	0.8	0.5	0.998	0	0.5	4.798
SEN	1	1	0.8	1	1.000	0	0.5	5.3
AU	0.733	1	0.7	0	0.975	0.5	1	4.908
FR	0.933	1	0.3	1	1.000	1	0.5	5.733
JP	1	1	0.8	1	1.000	0.5	1	6.3
CHNMZSOY	0.95	1	0.8	1	1.000	0	0.5	5.25

**Table 2.** Endogenous dataset quality indicators of all input datasets. <sup>(a)</sup>Endogenous qualities in MRF datasets are used for overall data quality but not for multi-criteria selection ranking (Step 4) when other datasets are available for the construction of CROPGRIDS.

where  $r_{min} = 0.0000833^\circ$  and  $r_{max} = 0.0833^\circ$  were the finest and coarsest resolutions across input datasets.

$Q_m$  was used to assess the level of maturity of a dataset, depending on the frequency of revisions, updates, or releases:

$$Q_m = \begin{cases} 1 & \text{if annual} \\ 0.5 & \text{if every some years} \\ 0 & \text{if never.} \end{cases} \quad (6)$$

$Q_d$  was used to assess the level of officiality, i.e., whether a dataset was the result of an official government or non-government dispatch, assuming that official government dispatches have higher reliability than those conducted by non-government entities. It was defined as

$$Q_d = \begin{cases} 1 & \text{if government} \\ 0.5 & \text{if non-government.} \end{cases} \quad (7)$$

All endogenous data quality indicators values are reported in Table 2 below.

**Compute exogenous data quality indicators (Step 3).** Exogenous data quality indicators were defined to describe the quality of a dataset against independent external information. They included  $Q_{CAM}$ , comparison against the cropland agreement map (CAM)<sup>18</sup>, and  $Q_{FAO}$ , comparison against FAOSTAT harvested area<sup>19</sup> in the year 2020. Unlike the endogenous indicators, exogenous data quality indicators were evaluated for each input dataset by crop and country.

Specifically,  $Q_{CAM}$  was used to measure the level of agreement of the crop spatial distribution in a dataset against CAM, which is a 2020-updated, statistically robust cropland mask integrating six independent cropland data products. We first converted the CA maps of each dataset and the cropland area map of CAM into binary maps, where a grid cell was assigned a value of one for non-zero crop area or zero otherwise. We then calculated  $Q_{CAM}$  for crop  $i$  in country  $j$  as:



$$Q_{CAM,i,j} = \frac{N_{overlap}(i,j)}{N_{CA}(i,j)}, \quad (8)$$

where  $N_{CA}(i,j)$  is the number of grid cells identified as crop  $i$  in country  $j$  in a given dataset and  $N_{overlap}$  is the number of grid cells where both CAM and the given dataset have non-zero values.

$Q_{FAO}$  was used to measure the relative error of the input dataset crop harvested area against FAOSTAT<sup>19</sup>. For crop  $i$  in country  $j$ ,  $Q_{FAO,i,j}$  was defined as:

$$Q_{FAO,i,j} = 1 - \min \left\{ 1, \frac{|HA(i,j) - HA_{FAO}(i,j)|}{HA_{FAO}(i,j)} \right\}. \quad (9)$$

where  $HA(i,j)$  is the total harvested area of crop  $i$  in country  $j$  in a dataset, and  $HA_{FAO}$  is the corresponding FAOSTAT value for the year 2020.  $Q_{FAO}$  ranges between 0 and 1, with  $Q_{FAO} = 1$  representing a perfect match against FAOSTAT. For specific crops where some countries and territories were not included in FAOSTAT, we set  $Q_{FAO} = 0$ .

**Assemblage of global harvested and crop area maps (Step 4).** Assemblage of geo-referenced harvested and crop area maps for individual crops and countries was conducted along two alternative pathways of availability: (1) only MRF data is available; or (2) multiple input datasets are available. In the first case, we proceeded to Step 5 (described later). In the second case, we used the multi-criteria ranking scheme based on endogenous and exogenous data quality indicators described above to select and use data from the dataset with the highest combined quality scores,  $Q_{k,i,j}$ , defined in relation to input dataset  $k$  for crop  $i$  in country  $j$  as:

$$Q_{k,i,j} = \frac{1}{3} \times \frac{(Q_y + Q_a + Q_s + Q_v + Q_r + Q_m + Q_d)_{k,i,j}}{7} + \frac{Q_{CAMk,i,j}}{3} + \frac{Q_{FAOk,i,j}}{3}. \quad (10)$$

The best-fit datasets  $k_{best}$  for crop  $i$  in country  $j$  are provided in Supplementary Table S3. In this case, the MRF dataset was excluded from the ranking. Hence, if only one dataset other than MRF is available, it will be automatically selected as the best dataset.

For each crop, we then compiled an *Arlecchino* map (mosaic) of  $HA$  and  $CA$  from best-fit datasets into one global map including all countries. The result of the multi-criteria analysis was that 27 out of the 28 geo-referenced datasets were included in CROPGRIDS (MYSTHA was not selected).

**Data gap filling with NSOs (Step 5).** In the first case of Step 4 when MRF is the only dataset available for a specific country-crop pair, we used NSOs subnational data to update  $HA$  (see list of available NSO in Supplementary Table S1). Specifically, we spatially disaggregated the subnational-level data following the same principle used in MRF<sup>2</sup>, i.e., by spreading tabulated harvested areas over an assigned agricultural region. For a given crop  $i$  in subnational unit  $j$ , we iteratively calculated the harvested area in each grid cell  $g$  belonging to subnational unit  $j$  using the CAM percentile  $p$  as,

$$HA_{i,g \in j,p} = NSO_{i,j} \frac{CAM_{g \in j,p}}{\sum_{g \in j} CAM_{g,p}} \quad (11)$$

with  $NSO_{i,j}$  being the tabulated NSOs values and  $CAM_{g,p}$  being the total cropland area in a grid cell  $g$  reported by CAM crop mask at  $p$  percentile. Note that  $HA$  in Eq. (11) is not uniformly distributed over the subnational unit  $j$ , but it follows the distribution of the percentile maps in CAM. We used up to 5 percentiles (2.5%, 5%, 10%, 25% and 50%), which correspond to a sequence of maps ranging from the smallest to the median agricultural land area, respectively. We stopped the iteration over the percentiles  $p$  when we first found some grid cells in which  $HA$  fell below the lower bound of 100 m<sup>2</sup>. In those grid cells,  $HA$  was adjusted to 100 m<sup>2</sup>, consistent with the lower bound used in Step 1. The spatialized harvested areas for crop  $i$  in subnational unit  $j$  ( $HA_{i,j,g}$ ) were hence determined. Finally, we calculated  $CA_{i,g} = \min\{HA_{i,g}, LA_g, CAM95_g\}$ , and we updated  $HA_{i,g} = CA_{i,g}$  (i.e., we assume  $CI = 1$ ). When NSOs subnational data are not available for a specific country-crop pair, we repeated the spatial information of MRF.

Note that for all datasets, including NSO-updated data, the quality calculated in Eq. (10) is outputted and distributed with this data product in the form of maps (see Table 3). The endogenous quality indicators for NSOs datasets are reported in Supplementary Table S4. However, specifically for NSOs, data quality in Eq. (10) excluded  $Q_{CAM}$  as CAM was used for spatialization.

**Data adjustment to FAOSTAT 2020 (Step 6).** After the assemblage of georeferenced maps,  $HA$  of individual crop types in individual countries were scaled to the corresponding country data in FAOSTAT in year 2020<sup>19</sup>. Scaling was performed with an iterative scheme minimizing the distance in harvested area  $HA$  from FAOSTAT with two constraints - the lower bound (100 m<sup>2</sup>) and the upper bound  $LA$ . Adjustment of  $CA$  was conducted simultaneously to  $HA$  by retaining the crop intensity ratio  $CI$  in any specific grid cell. When scaled values of  $\sum CA$  exceeded  $LA$  (upper bound), the excess crop area was redistributed uniformly to all other grid cells within that country where each of the individual crops are present. Similarly, when the individual scaled values of  $HA$  became smaller than 100 m<sup>2</sup> (lower bound), excess area was uniformly removed from all other grid cells within that country containing that crop. We limited the number of iterations to 60 or when the total crop area adjusted in all crop types in a country was less than 0.5% different than in FAOSTAT in an iteration or when the

Folder Name	File Name	Description	Variable		
			Name	Description	Unit
CROPGRIDSv1.08_NC_maps.zip	CROPGRIDSv1.08_YYYY.nc	Contains globally gridded data for crop YYYY.	harvarea	Harvested area	hectares
			croparea	Crop (physical) area	hectares
qual			Data quality	—	
set			Best-fit dataset used	—	
	Countries_2018.nc	Contains GAUL <sup>21</sup> country mask (level 0) used in this study.	country	Country name and code	—
CROPGRIDSv1.08_PNG_maps.zip	CROPGRIDSv1.08_YYYY.png	Images of harvested and crop areas of crop YYYY, and corresponding data quality and best-fit datasets.			
	Table_CROPGRIDS1.08_COU.xlsx	Excel table of areas for each 173 crops in each country.	Harvested area	Harvested area	hectares
			Crop (physical) area	Crop (physical) area	hectares
CODES.zip	MAIN.m, Aggregate_REMAP.m, Correct_REMAP.m, BUILD_PATCH.m, BUILD_PATCH_fun10.m, ADDNSO_RAW.m, CORRECT_RAWNSO.m, ADJUST_CORRECTED.m	MATLAB scripts used to construct CROPGRIDS (Steps 1 to 6 in Fig. 1).			
	CROPS_matching_v17.xlsx	Excel table of crop name aggregation and matching.			
	DATASET_qualities_v10.xlsx	Excel table showing the endogenous qualities of gridded input datasets.			
	DATASET_NSQ_qualities_v1.xlsx	Excel table showing the endogenous qualities of NSOs data.			
	FAOSTAT_ALL_2020_edited.xlsx	National-level harvested areas for year 2020 obtained from FAOSTAT.			

**Table 3.** CROPGRIDS data distribution files and variables. All files are publicly available from *figshare* repository<sup>25</sup>.

relative change in total crop area was smaller than 0.01% in an iteration. These thresholds used to limit adjustment iterations were empirically chosen to ensure a balance between data quality and computational efficiency. Adjustment of CROPGRIDS to FAOSTAT was only carried out for all crops and countries that existed in both datasets (153 crops and 185 countries). For a given country-crop pair, a same adjustment factor was applied to all the grid cells containing that crop in that country. Only less than 7% of country-crop pairs required adjustment by a factor smaller than 0.1 or greater than 10 (Supplementary Figure S1). Specifically, the median adjustment factors are not significantly different across crops except for blueberry, mushrooms and triticale, which also show the largest spread across countries (Supplementary Figure S2). Similarly, the median adjustment factors of individual crops are not significantly different across countries, with a few countries showing a large spread (i.e., Afghanistan, Algeria, Iraq, Malta and Qatar, Supplementary Figure S3).

In this step, we simultaneously checked and verified again that the sum of *CA* across all crops is always smaller than or equal to *LA* in each grid cell and that, for individual crops,  $CA \leq HA$ , and  $CI = HA/PA \leq 3$  were always satisfied.

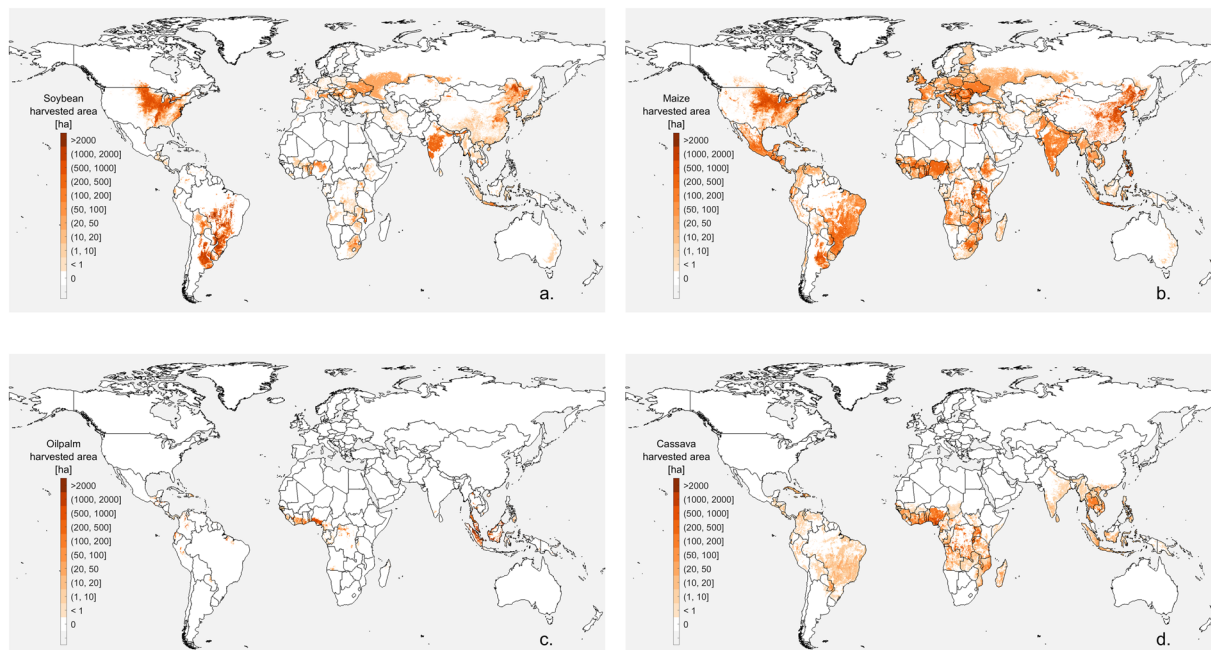
We presented examples of harvested area maps for the top four crops experiencing major changes since 2000, i.e., oil palm, soybean, cassava, and maize (Fig. 2).

### Data Records

CROPGRIDS dataset distributes global georeferenced maps of harvested and crop (physical) areas and corresponding data quality for 173 crops (refer to Supplementary Table S3 for the list of crops, where coir and gum originally available in MRF are not included in CROPGRIDS) for the year 2020 at a resolution of 0.05° (~5.6 km at the equator) with a bounding box of −180° to 180° longitude and −90° to 90° latitude using the WGS-84 coordinate system. The georeferenced maps are distributed as NetCDF files, which also provides detailed legend of values. This dataset is available for public download from the *figshare* repository<sup>25</sup> at <https://doi.org/10.6084/m9.figshare.22491997>. The files included in this distribution are described in Table 3.

### Technical Validation

The majority of the datasets listed in Table 1 included certain level of validation, which are summarized in Supplementary Table S5. Among the 28 input datasets, 17 included validation against ground truth data. In this work, we independently evaluated data on *HA* and *CA* in CROPGRIDS against (1) official national and subnational statistics of crop-specific harvested area (full references provided in Supplementary Table S1); (2) FAOSTAT land areas under temporary and permanent crops by country<sup>26</sup>; and (3) the grid cell-level crop-land areas calculated from CAM<sup>4,18</sup>. In addition to independent validations, we conducted uncertainty analysis to test the robustness of the multi-criteria selection ranking and provided comparison of the national-level crop specific *HA* in CROPGRIDS against corresponding FAOSTAT data for 2020, which was used to construct CROPGRIDS.



**Fig. 2** Harvested area maps in CROPGRIDS for the top four crops experiencing the largest expansion since 2000. (a) Soybean, (b) maize, (c) oil palm, and (d) cassava.

**Validation of CROPGRIDS with official national and subnational data.** We compiled a library of independent datasets of national and subnational harvested area by crop from 36 NSOs (Supplementary Table S1), covering 71 countries and territories and 861 subnational units. Of these, 40 countries and territories reported subnational-level data and 35 reported more than 20 crops each, resulting in a total of 1,852 points available at national-level and 12,149 points at subnational-level. Among the subnational-level data points, 4,832 points were used to construct CROPGRIDS, leaving 7,317 points available for independent validation. In total, evaluations of 106 crop data were conducted against these independent crop statistics from NSOs.

We matched and aggregated crop types in each NSO dataset to match those reported in CROPGRIDS. We used the GAUL<sup>21</sup> dataset (level 1) to identify subnational units and perform relevant aggregations from pixel level to administrative level 1. The calculations were conducted for 106 crops and were quantified using the coefficient of determination  $R^2$  (analogous to Nash–Sutcliffe efficiency, NSE<sup>27</sup>) and normalized root mean squared errors (NRMSE) as

$$R_i^2 = 1 - \frac{\sum_{(j)} [HA_{NSO}(i, j) - HA(i, j)]^2}{\sum_{(j)} [HA_{NSO}(i, j) - \overline{HA}_{NSO}(i, j)]^2}, \quad (12)$$

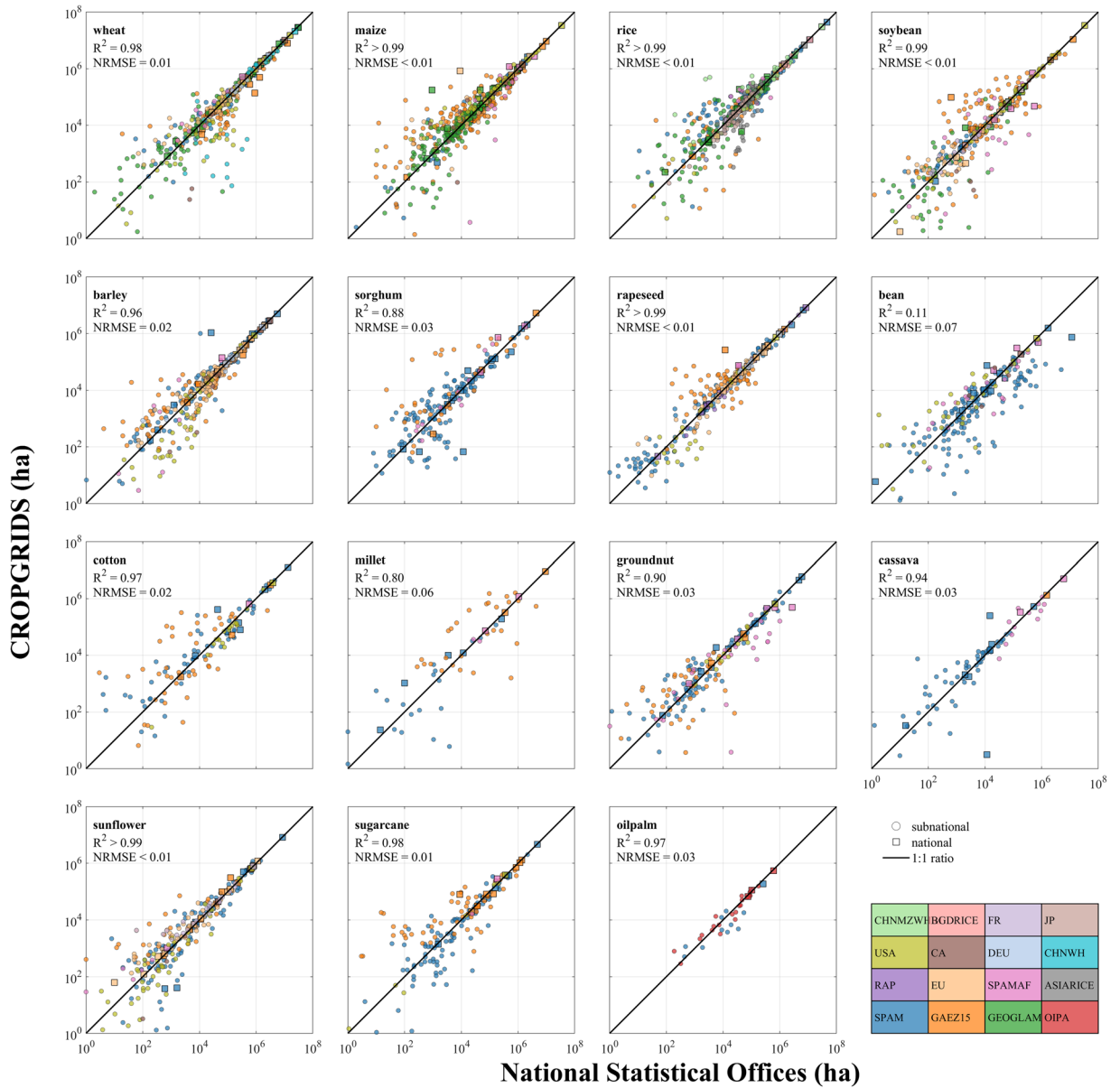
$$NRMSE_i = \frac{\sqrt{\frac{\sum_{(j)} [HA_{NSO}(i, j) - HA(i, j)]^2}{n}}}{[HA_{NSO, max}(i) - HA_{NSO, min}(i)]}, \quad (13)$$

where  $HA(i, j)$  and  $HA_{NSO}(i, j)$  are the harvested area of crop  $i$  in administrative unit  $j$  reported by CROPGRIDS and NSOs, respectively,  $\overline{HA}_{NSO}$  is the average of all NSOs data points,  $HA_{NSO, max}$  and  $HA_{NSO, min}$  are the corresponding maximum and minimum crop harvested areas of NSOs, and  $n$  is the number of data points.

Among the 106 crops suitable for comparison, the harvested area of 81 crops in CROPGRIDS agreed relatively well with data from NSOs ( $R^2 > 0.5$ ,  $NRMSE < 0.2$ , Fig. 3 and Supplementary Figure S4). Specifically, the comparisons for important crops such as wheat, maize, rice, soybean, barley, rapeseed, cotton, cassava, sunflower, sugarcane, and oil palm had  $R^2 > 0.95$  and  $NRMSE \leq 0.05$ , showing very good agreement with officially reported national and subnational statistics (Fig. 3).

**Validation of CROPGRIDS with FAOSTAT land area under temporary and permanent crops.** The crop (physical) area in CROPGRIDS refers to FAO land use classes ‘temporary’ or ‘permanent’ crops, depending on crop type<sup>28</sup>. Here, we compared crop areas with FAOSTAT land areas under temporary and permanent crops for 2020 in more than 180 countries. We first classified the 173 crop types included in CROPGRIDS into temporary and permanent crops following the ICC classification<sup>24</sup> (see Supplementary Table S3 for details). We used the GAUL<sup>21</sup> dataset (level 0) to identify country boundaries and perform relevant aggregations from pixel level to national level. The goodness of comparison was evaluated using  $R^2$  and NRMSE as



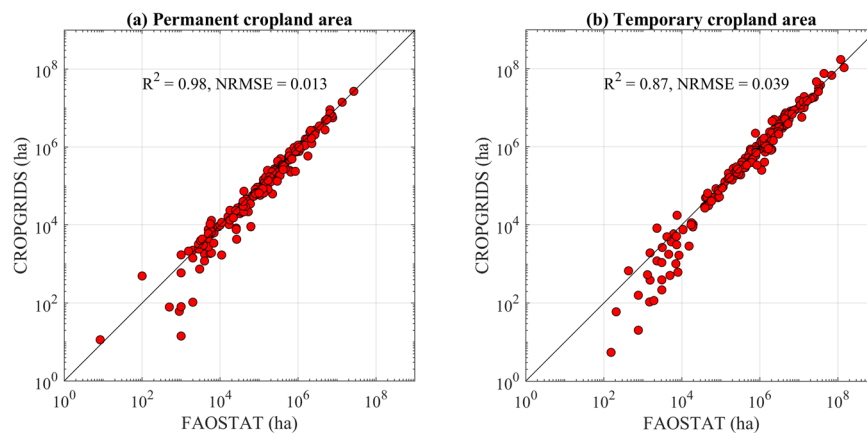


**Fig. 3** Validation of crop harvested areas in CROPGRIDS against data from National Statistical Offices at national and subnational levels. The colours of the markers refer to the georeferenced datasets selected to use in CROPGRIDS. “Squared” markers represent national-level data, while “circled” markers represent subnational-level data. This figure shows only the validation for the top 15 crops with the largest global harvested area. The validations for the other 91 crops are shown in Supplementary Figure S4.

$$R_i^2 = 1 - \frac{\sum_{(j)} [CA_{FAO}(i,j) - CA(i,j)]^2}{\sum_{(j)} [CA_{FAO}(i,j) - \overline{CA_{FAO}}(i,j)]^2}, \tag{14}$$

$$NRMSE_i = \frac{\sqrt{\frac{\sum_{(j)} [CA_{FAO}(i,j) - CA(i,j)]^2}{n}}}{[CA_{FAO,max}(i) - CA_{FAO,min}(i)]}, \tag{15}$$

where  $CA(i,j)$  and  $CA_{FAO}(i,j)$  are the crop area of either temporary or permanent crops (indicated as  $i$ ) in country  $j$  reported by CROPGRIDS and FAOSTAT, respectively,  $\overline{CA_{FAO}}$  is the average of all FAOSTAT data points,  $CA_{FAO,max}$  and  $CA_{FAO,min}$  are the corresponding maximum and minimum temporary or permanent crop areas of FAOSTAT, and  $n$  is the number of data points.



**Fig. 4** Validation of permanent (a) and temporary (b) crop (physical) area in CROPGRIDS against FAOSTAT of year 2020. Each circle representing one country.

In CROPGRIDS, the 2020 world total permanent crop area was 167 million ha, consistent with but approximately 8% lower than the 181 million ha reported by FAOSTAT. At national-level, the permanent crop areas determined from CROPGRIDS matched well with values reported by FAOSTAT with  $R^2 = 0.98$  and  $NRMSE = 0.01$  (Fig. 4a). Additionally, temporary crops in CROPGRIDS covered 1.22 billion ha of global cropland area, and overestimated by approximately 13% the temporary crop area reported in FAOSTAT for 2020, which is 1.08 billion ha. The comparison of temporary crop areas at national-level showed a relatively good match to FAOSTAT data, with  $R^2 = 0.87$  and  $NRMSE = 0.04$  (Fig. 4b). The overall overestimation of temporary crop area by CROPGRIDS may arise from multiple cropping of different crops, i.e., we may have counted the cropland area more than once if the same piece of land was cultivated with more than one type of crop in a year.

**Comparison of CROPGRIDS crop area with CAM.** We next validated the CA of all crops included in CROPGRIDS geo-spatially against the cropland area in CAM. Firstly, we calculated the sum of the CA of all crops in each grid cell  $g$  in CROPGRIDS,  $CA_{TOT}(g)$ . Next, for each grid cell  $g$ , we calculated the 5<sup>th</sup>,  $CA_{CAM}^{5th}(g)$ , and 95<sup>th</sup>,  $CA_{CAM}^{95th}(g)$ , percentile cropland area based on CAM dataset<sup>18</sup>. For each grid cell, we then determined if  $CA_{TOT}(g)$  calculated based on CROPGRIDS falls within  $CA_{CAM}^{5th}(g)$  and  $CA_{CAM}^{95th}(g)$ .

About 97% of grid cells identified as cropland (i.e., total CA across all crops in a grid cell  $>0$ ) in CROPGRIDS were also identified as cropland in CAM. Globally, the crop areas of about 93% of grid cells identified as cropland in CROPGRIDS fall within  $CA_{CAM}^{5th}$  and  $CA_{CAM}^{95th}$ , with less than 1% falling below the lower bound (Fig. 5). Those grid cells that have crop area greater than  $CA_{CAM}^{95th}$  were majorly found in African countries (e.g. Nigeria, Ghana, Cote d'Ivoire) where the six land cover layers used to build CAM are characterized by very high uncertainty.

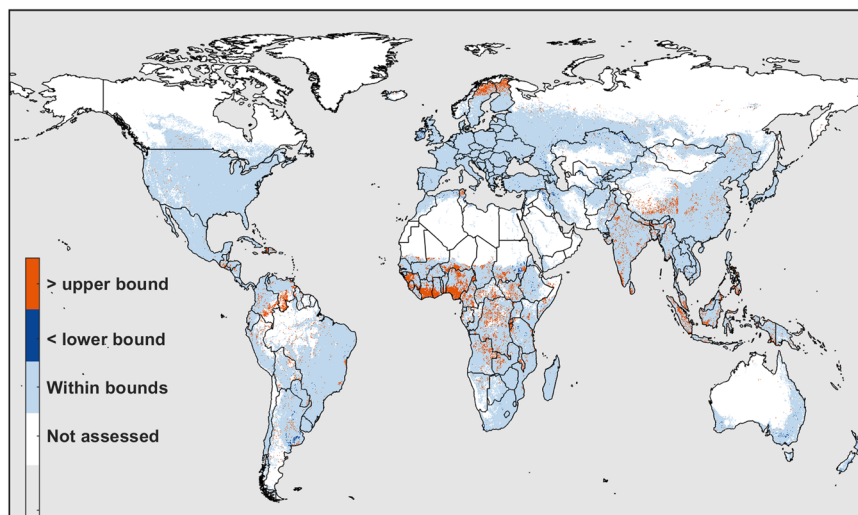
**Comparison of crop harvested area in CROPGRIDS against FAOSTAT.** We compared global-level and national-level crop-specific harvested areas of year 2020 obtained from FAOSTAT<sup>19</sup> against the corresponding values computed from CROPGRIDS. We used the GAUL<sup>21</sup> dataset (level 0) to aggregate harvested areas of each crop in CROPGRIDS from pixel level to national level. This comparison does not represent a fully independent validation as these FAOSTAT data were used in Step 6 to adjust the HA and CA values in CROPGRIDS. Rather, this comparison serves to provide an estimation of percent error ( $\% \Delta$ ) between FAOSTAT and CROPGRIDS at national-level, determined as

$$\% \Delta(i, j) = \frac{HA(i, j) - HA_{FAO}(i, j)}{HA_{FAO}(i, j)} \times 100 \quad (16)$$

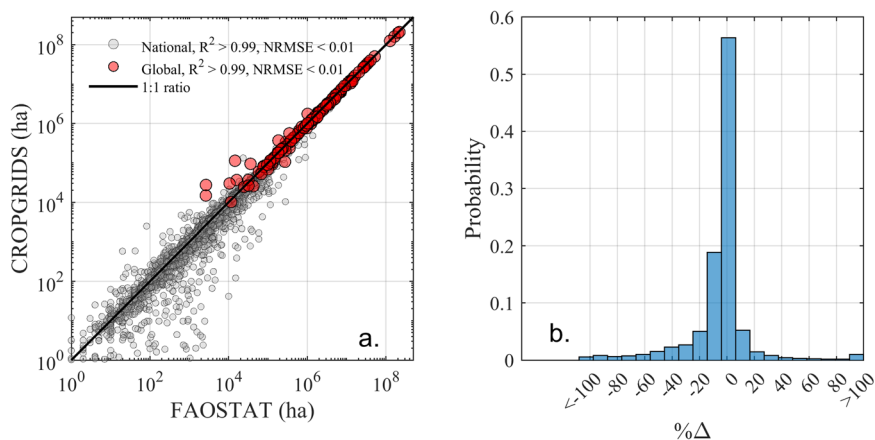
where  $HA(i, j)$  is the harvested area of crop  $i$  in country  $j$  in CROPGRIDS and  $HA_{FAO}$  is the corresponding 2020 FAOSTAT value.

The global crop-specific harvested area in CROPGRIDS matched well with those reported in FAOSTAT with an  $R^2 \approx 1.00$  and  $NRMSE < 0.01$  (Fig. 6, red markers), with 115 crops having a difference less than  $\pm 10\%$  (Supplementary Figure S5). Comparison of CROPGRIDS against national-level crop-specific harvested areas of FAOSTAT also shows good matching with an  $R^2 \approx 1.00$  and  $NRMSE < 0.01$  (Fig. 6a, grey markers). About 84% of data points (out of a total of 7,697 pairs) had differences less than  $\pm 20\%$ , while only less than 1% had a difference greater than  $\pm 100\%$  (Fig. 6b).

**Uncertainty in the multi-criteria selection ranking.** The endogenous and exogenous data quality indicators used in the multi-criteria selection were assigned meaningfully but with arbitrary values. We conducted a Monte-Carlo analysis to quantify the uncertainty associated to such arbitrary choices, by introducing random weights, in the range 0–1, to each data quality indicator, that is:  $\{w_c, w_a, w_s, w_p, w_r, w_m, w_d\}$  for endogenous and  $\{w_G, w_F\}$  for exogenous indicators—whereas we implicitly had used unity weights in Eq. (10). We extracted 10,000 values of each of the nine weights from independent Gaussian probability distribution functions with a mean



**Fig. 5** Comparison of the total crop area of all crops included in CROPGRIDS against the cropland area in CAM. The colours in the map illustrate if the total crop area estimated in CROPGRIDS in each grid cell falls within the lower (5th percentile) and upper (95th percentile) bounds of cropland area calculated from CAM.



**Fig. 6** Comparison of crop harvested area in CROPGRIDS with FAOSTAT values for 2020. (a) scatter-plot between harvested areas in CROPGRIDS against FAOSTAT, and (b) probability distribution of percent error  $\% \Delta$ . In total, there were 7,697 pairs of comparisons at national-level and 153 pairs for global crop-specific harvested areas.

equal to 1 and a standard deviation equal to 0.1 and we limited their values within the range between 0.7 and 1.3, that is three times the standard deviation. We next counted the frequency of occurrence of a selected dataset different than when using the default weight values. This uncertainty analysis was only conducted for combinations of countries and crops where more than one dataset was available.

Results of the Monte-Carlo analysis on the endogenous and exogenous characteristics of the multi-criteria selection ranking scheme suggested that the method of best-fit dataset selection was highly robust. Specifically, for the 78 crops and 187 countries with multiple datasets, the probability that the selection of the best-fit dataset would change with randomized characteristics was highly unlikely (white tiles in Supplementary Figure S6). In a minor fraction of crops and countries, the probability was greater than 10%, with only 40 out of 3352 assessed pairs of crops and countries having a probability  $\geq 40\%$  and only 3 pairs having a probability  $\geq 50\%$  (Supplementary Figure S6).

**Known limitations and uncertainties.** CROPGRIDS inherits uncertainties and errors embedded in the input datasets and these uncertainties can stem from a variety of sources. Datasets constructed based on censuses surveys (e.g., MRF and SPAM) can have uncertainties stemming from the methods used to spatialize crop area statistics at administrative-level 2 and the imperfection in statistical reporting of harvested and crop areas. Datasets constructed using remote sensing approaches can suffer from the inherent uncertainties in remote sensing data, such as, atmospheric interference and limitations in spatial resolution. More generally, these datasets also carry forward uncertainties underlying in the cropland layer maps used as their input and can be limited by

the availability of ground truth data in certain regions for validation purposes. These uncertainties can propagate through the mapping process and affect the accuracy of the resulting harvested and crop area estimates in CROPGRIDS.

In addition to inherited uncertainties, the construction of CROPGRIDS also suffers from known limitations. Firstly, the imputation of *CA* for temporary crops by taking the minimum value between *HA*, *LA*, and *CAM95* as described in Methods may lead to an overestimation of *CA*, particularly for those crops that undergo multiple harvests. While we have accounted for cropping intensities greater than 1 for crops with multiple harvests (e.g., rice), we have not explicitly accounted for dual and multi-layered cropping systems when more than one crop are grown in the same cultivated area. Information about dual cropping systems across the available datasets is limited, with only the datasets for USA and Canada providing this information. Information on multi-layered cropping systems (e.g., barley below olive trees in some Mediterranean systems or coffee plantations under natural trees) is entirely lacking. The lack of information on cropping practices and irrigation management may contribute uncertainties to crop distribution mapping, potentially leading to both underestimations and overestimations in *HA* and *CA* for some countries and some crops, leaving a knowledge gap that may be filled in future releases of CROPGRIDS. In addition, we did not account explicitly for protected agriculture. The estimation of *LA* excluded urbanization area, which may encompass protected agriculture, and hence, may lead to underestimation of *HA* and *CA* of some horticultural crops. These knowledge gaps highlight the importance of expanding the spatial coverage and frequency of ground monitoring and data collection for agricultural practices to enhance crop distribution mapping.

The approach employed in developing CROPGRIDS involves the integration of crop area data from various years, with the majority of the datasets used having reference years between 2015 and 2020. However, this approach can introduce uncertainties due to potential annual variations in the crop types cultivated in specific regions, influenced by factors such as climatic suitability and market demand.

At a spatial resolution of  $0.05^\circ$ , a grid cell has a size of approximately  $5.6 \text{ km} \times 5.6 \text{ km}$ , corresponding to about 3000 ha. This leads to uncertainties in the estimated harvested and crop areas for some crops typically cultivated at smaller scales except under intensively managed systems, often monocultures. It furthermore creates uncertainty at the border between two countries and affects in particular the calculation of the exogenous data quality indicator  $Q_{FAO}$  that compares a dataset against national-level crop harvested area reported by FAOSTAT. This border effect impacts estimates mostly in small countries in two ways. The first is when a country has zero harvested and crop areas for a crop across all datasets because border grid cells fall in the neighbouring country, whereas FAOSTAT reports non-zero values. In this case, no selection is performed. The second is when, in contrast, a country has a harvested area greater than zero when grid cells of other neighbour countries fall within a country and FAOSTAT returns zero value. In this case, datasets will still be ranked and the best-fit will be selected according to other quality indicators. This known bias is difficult to detect and correct, especially for small countries, because whether a border grid cell belongs to one or another country cannot be estimated correctly at the given resolution. Specifically, this bias is scale-dependent and its occurrence decreases with increasing resolution and data quality, including of the layer of administrative boundaries used to extract country statistics. Due to constraint in spatial resolution, CROPGRIDS excludes a few small countries and territories (i.e., Falkland, Faroe Islands, French S.A.T., Heart Island, Isle of Man, Kingman Reef, Kiribati, Ma'tan al-Sarra, Mayotte, Nether. Antilles, Palau, Réunion, Saint Pierre, South Georgia, Svalbard, Virgin Islands). Greenland is also excluded, considering the small area of cultivated land.

Furthermore, uncertainties can also arise from the type of validation data, which themselves may inherit uncertainties. For example, statistics at the subnational level may be masked due to confidentiality issues, especially in cases where significant producers are located in relatively small counties or districts. As a result, the aggregated data at the national level may differ from the data available at the more detailed subnational level, thereby posing limitations in the validation process.

### Usage Notes

All georeferenced maps distributed in CROPGRIDS dataset<sup>25</sup> are formatted as standard NetCDF4 files, which can be read in various coding languages (e.g., MATLAB, Python, Julia) and software (ArcGIS, QGIS, Panoply). CROPGRIDS dataset also contains a country mask based on GAUL<sup>21</sup> dataset at administrative level 0, which can be used to aggregate pixel-level data to national-level for comparison against official statistics (e.g., FAOSTAT or other National Statistical Offices). Crop type name in CROPGRIDS follows the naming system used by FAO<sup>19</sup>, allowing direct comparison against FAOSTAT data.

### Code availability

All data processing and testing described in Methods and Technical Validation sections were conducted using MATLAB version R2021a. Main codes used to construct CROPGRIDS are distributed in the “CODES.zip” folder (Table 3) along with CROPGRIDS dataset available for public download from the *figshare* repository<sup>25</sup> at <https://doi.org/10.6084/m9.figshare.22491997>.

Received: 22 November 2023; Accepted: 10 April 2024;

Published online: 22 April 2024

### References

1. Tubiello, F. N. *et al.* Measuring Progress Towards Sustainable Agriculture FAO Statistical Working Papers Series No. 21–24. (FAO, 2021).
2. Monfreda, C., Ramankutty, N. & Foley, J. A. Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. *Global Biogeochem Cycles* **22**, GB1022, <https://doi.org/10.1029/2007gb002947> (2008).



3. Potapov, P. *et al.* Global maps of cropland extent and change show accelerated cropland expansion in the twenty-first century. *Nat Food* **3**, 19–28, <https://doi.org/10.1038/s43016-021-00429-z> (2022).
4. Tubiello, F. N. *et al.* Measuring the world's cropland area. *Nat Food* **4**, 30–32, <https://doi.org/10.1038/s43016-022-00667-9> (2023).
5. Foley, J. A. *et al.* Solutions for a cultivated planet. *Nature* **478**, 337–342, <https://doi.org/10.1038/nature10452> (2011).
6. Kim, K. H., Doi, Y., Ramankutty, N. & Iizumi, T. A review of global gridded cropping system data products. *Environ Res Lett* **16**, 093005, <https://doi.org/10.1088/1748-9326/ac20f4> (2021).
7. Defourny, P. *et al.* Near real-time agriculture monitoring at national scale at parcel resolution: Performance assessment of the Sen2-Agri automated system in various cropping systems around the world. *Remote Sens Environ* **221**, 551–568, <https://doi.org/10.1016/j.rse.2018.11.007> (2019).
8. Franch, B. *et al.* Global crop calendars of maize and wheat in the framework of the WorldCereal project. *Gisci Remote Sens* **59**, 885–913, <https://doi.org/10.1080/15481603.2022.2079273> (2022).
9. Becker-Reshef, I. *et al.* in *Remote Sensing of Agriculture and Land Cover/Land Use Changes in South and Southeast Asian Countries*, 53–80 (Springer, 2022).
10. Seifert, C. A., Azzari, G. & Lobell, D. B. Satellite detection of cover crops and their effects on crop yield in the Midwestern United States. *Environ Res Lett* **13**, 064033, <https://doi.org/10.1088/1748-9326/aac4c8> (2018).
11. Azzari, G. *et al.* Satellite mapping of tillage practices in the North Central US region from 2005 to 2016. *Remote Sens Environ* **221**, 417–429, <https://doi.org/10.1016/j.rse.2018.11.010> (2019).
12. Lobell, D. B. *et al.* Eyes in the Sky, Boots on the Ground: Assessing Satellite- and Ground-Based Approaches to Crop Yield Measurement and Analysis. *Am J Agr Econ* **102**, 202–219, <https://doi.org/10.1093/ajae/aaz051> (2020).
13. Becker-Reshef, I. *et al.* Crop Type Maps for Operational Global Agricultural Monitoring. *Sci Data* **10**, 172, <https://doi.org/10.1038/s41597-023-02047-9> (2023).
14. Beyer, R. M., Hua, F. Y., Martin, P. A., Manica, A. & Rademacher, T. Relocating croplands could drastically reduce the environmental impacts of global food production. *Commun Earth Environ* **3**, 49, <https://doi.org/10.1038/s43247-022-00360-6> (2022).
15. Ortiz-Bobea, A., Ault, T. R., Carrillo, C. M., Chambers, R. G. & Lobel, D. B. Anthropogenic climate change has slowed global agricultural productivity growth. *Nat Clim Change* **11**, 306–312, <https://doi.org/10.1038/s41558-021-01000-1> (2021).
16. Tang, F. H. M., Lenzen, M., McBratney, A. & Maggi, F. Risk of pesticide pollution at the global scale. *Nat Geosci* **14**, 206–210, <https://doi.org/10.1038/s41561-021-00712-5> (2021).
17. Proctor, J., Rigden, A., Chan, D. & Huybers, P. More accurate specification of water supply shows its importance for global crop production. *Nat Food* **3**, 753–763, <https://doi.org/10.1038/s43016-022-00592-x> (2022).
18. Tubiello, F. N. *et al.* A new cropland area database by country circa 2020. *Earth Syst. Sci. Data* **15**, 4997–5015, <https://doi.org/10.5194/essd-15-4997-2023> (2023).
19. FAO. FAOSTAT Crops and livestock products database, <https://www.fao.org/faostat/en/#data/QCL> (07 November 2023).
20. Friedl, M. & Sulla-Menashe, D. in *MODIS/Terra+ Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V061* (NASA EOSDIS Land Processes Distributed Active Archive Center, 2022).
21. FAO. GAUL: Global Administrative Unit Layers 2015, *First-Level Administrative Units*, [https://developers.google.com/earth-engine/datasets/catalog/FAO\\_GAUL\\_2015\\_level1#description](https://developers.google.com/earth-engine/datasets/catalog/FAO_GAUL_2015_level1#description) (2015).
22. MathWorks. *imresize*, MATLAB, <https://www.mathworks.com/help/matlab/ref/imresize.html> (2021).
23. Zhang, M. *et al.* GCI30: a global dataset of 30 m cropping intensity using multisource remote sensing imagery. *Earth Syst Sci Data* **13**, 4799–4817, <https://doi.org/10.5194/essd-13-4799-2021> (2021).
24. FAO. World Programme for the Census of Agriculture 2020, Volume 1 Programme, concepts and definitions, FAO Statistical Development Series 15. (Rome, 2017).
25. Tang, F. H. M. *et al.* CROPGRIDS. *Figshare* <https://doi.org/10.6084/m9.figshare.22491997.v7> (2024).
26. FAO. FAOSTAT Land Use database, <https://www.fao.org/faostat/en/#data/RL> (07 November 2023).
27. Duc, L. & Sawada, Y. A signal-processing-based interpretation of the Nash-Sutcliffe efficiency. *Hydrol Earth Syst Sc* **27**, 1827–1839, <https://doi.org/10.5194/hess-27-1827-2023> (2023).
28. FAO. Land use statistics and indicators 2000–2021. Global, regional and country trends, FAOSTAT Analytical Briefs Series No. 71. (Rome, 2023).
29. Yu, Q. Y. *et al.* A cultivated planet in 2010-Part 2: The global gridded agricultural-production maps. *Earth Syst Sci Data* **12**, 3545–3572, <https://doi.org/10.5194/essd-12-3545-2020> (2020).
30. Grogan, D., Frolking, S., Wisser, D., Prusevich, A. & Glidden, S. Global gridded crop harvested area, production, yield, and monthly physical area data circa 2015. *Sci Data* **9**, 15, <https://doi.org/10.1038/s41597-021-01115-2> (2022).
31. Becker-Reshef, I. *et al.* GEOGLAM Best Available Crop Type Masks. *Zenodo* <https://doi.org/10.5281/ZENODO.6511594> (2022).
32. Descals, A. *et al.* High-resolution global map of smallholder and industrial closed-canopy oil palm plantations. *Earth Syst Sci Data* **13**, 1211–1231, <https://doi.org/10.5194/essd-13-1211-2021> (2021).
33. Han, J. C. *et al.* The RapeseedMap10 database: annual maps of rapeseed at a spatial resolution of 10m based on multi-source data. *Earth Syst Sci Data* **13**, 2857–2874, <https://doi.org/10.5194/essd-13-2857-2021> (2021).
34. dAndrimont, R. *et al.* From parcel to continental scale-A first European crop type map based on Sentinel-1 and LUCAS Copernicus in-situ observations. *Remote Sens Environ* **266**, 112708, <https://doi.org/10.1016/j.rse.2021.112708> (2021).
35. IFPRI. Spatially-disaggregated crop production statistics data in Africa South of the Sahara for 2017 (Version 3.0). *Harvard Dataverse* <https://doi.org/10.7910/DVN/FSSKBW> (2020).
36. Szyniszewska, A. M. CassavaMap, a fine-resolution disaggregation of cassava production and harvested area in Africa in 2014. *Sci Data* **7**, 159, <https://doi.org/10.1038/s41597-020-0501-z> (2020).
37. Song, X. P. *et al.* Massive soybean expansion in South America since 2000 and implications for conservation. *Nat Sustain* **4**, 784–792, <https://doi.org/10.1038/s41893-021-00729-z> (2021).
38. Danylo, O. *et al.* A map of the extent and year of detection of oil palm plantations in Indonesia, Malaysia and Thailand. *Sci Data* **8**, 96, <https://doi.org/10.1038/s41597-021-00867-1> (2021).
39. Han, J. C. *et al.* Annual paddy rice planting area and cropping intensity datasets and their dynamics in the Asian monsoon region from 2000 to 2020. *Agr Syst* **200**, 103437, <https://doi.org/10.1016/j.agsy.2022.103437> (2022).
40. Abu, I. O., Szantoi, Z., Brink, A., Robuchon, M. & Thiel, M. Detecting cocoa plantations in Cote d'Ivoire and Ghana and their implications on protected areas. *Ecol Indic* **129**, 107863, <https://doi.org/10.1016/j.ecolind.2021.107863> (2021).
41. Remelgado, R. *et al.* A crop type dataset for consistent land cover classification in Central Asia. *Sci Data* **7**, 250, <https://doi.org/10.1038/s41597-020-00591-2> (2020).
42. Boryan, C., Yang, Z. W., Mueller, R. & Craig, M. Monitoring US agriculture: the US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program. *Geocarto Int* **26**, 341–358, <https://doi.org/10.1080/10106049.2011.562309> (2011).
43. Fiset, T. *et al.* in *2013 Second International Conference on Agro-Geoinformatics* 270–274 (IEEE, 2013).
44. FAO. *Crop Type (Afghanistan - 10m - 2020) - EOSTAT, FAO Map Catalog* <https://data.apps.fao.org/catalog/iso/3cf3376c-50fa-42a6-8bfe-070a56fd2112> (2021).
45. Blickensdörfer, L. *et al.* Mapping of crop types and crop sequences with combined time series of Sentinel-1, Sentinel-2 and Landsat 8 data for Germany. *Remote Sens Environ* **269**, 112831, <https://doi.org/10.1016/j.rse.2021.112831> (2022).



46. Dong, J. *et al.* Early-season mapping of winter wheat in China based on Landsat and Sentinel images. *Earth Syst Sci Data* **12**, 3081–3095, <https://doi.org/10.5194/essd-12-3081-2020> (2020).
47. Qiu, B. W., Huang, Y. Z., Chen, C. C., Tang, Z. H. & Zou, F. L. Mapping spatiotemporal dynamics of maize in China from 2005 to 2017 through designing leaf moisture based indicator from Normalized Multi-band Drought Index. *Comput Electron Agr* **153**, 82–93, <https://doi.org/10.1016/j.compag.2018.07.039> (2018).
48. Qiu, B. W. *et al.* Maps of cropping patterns in China during 2015–2021. *Sci Data* **9**, 479, <https://doi.org/10.1038/s41597-022-01589-8> (2022).
49. Singha, M., Dong, J. W., Zhang, G. L. & Xiao, X. M. High resolution paddy rice maps in cloud-prone Bangladesh and Northeast India using Sentinel-1 data. *Sci Data* **6**, 26, <https://doi.org/10.1038/s41597-019-0036-3> (2019).
50. Zheng, Y., Luciano, A. C. D., Dong, J. & Yuan, W. P. High-resolution map of sugarcane cultivation in Brazil using a phenology-based method. *Earth Syst Sci Data* **14**, 2065–2080, <https://doi.org/10.5194/essd-14-2065-2022> (2022).
51. FAO. *Crop Type (Senegal - 10m - 2018) - EOSTAT, FAO Map Catalog*, <https://data.apps.fao.org/catalog/iso/5c377b2b-3c2e-4b70-afd7-0c80900b68bb> (2021).
52. Australian Bureau of Agricultural and Resource Economics and Sciences (ABARES). *Land use of Australia 2010-11 to 2015-16, 250m*, <https://doi.org/10.25814/7ygw-4d64> (2022).
53. Thierion, V., Vincent, A. & Valero, S. Theia OSO Land Cover Map 2021 (Version 1) Zenodo, <https://doi.org/10.5281/zenodo.6538910> (2022).
54. Japan Aerospace Exploration Agency Earth Observation Research Center (JAXA EORC). *High Resolution Land-Use and Land-Cover Map of Japan (version 21.11)* [https://www.eorc.jaxa.jp/ALOS/en/dataset/lulc/lulc\\_v2111\\_e.htm](https://www.eorc.jaxa.jp/ALOS/en/dataset/lulc/lulc_v2111_e.htm) (2021).
55. Li, H. *et al.* Development of a 10-m resolution maize and soybean map over China: Matching satellite-based crop classification with sample-based area estimation. *Remote Sens Environ* **294**, 113623 (2023).

## Acknowledgements

We thank Qazi M. Amir for helping with the survey of existing data. The research work leading to CROPGRIDS was partially supported by the United Nations, Food and Agriculture Organization under contract (UNFAO CT34329). FNT acknowledges funding by the Swiss Federal Office for Agriculture to the FAO Regular Programme under Project: “Strengthening Global Assessments of Sustainable Agriculture. Phase 2”. This project was undertaken with the assistance of resources and services from the National Computational Infrastructure (NCI), which is supported by the Australian Government via the NCMAS 2022 and Adapter 2022 schemes awarded to F.M. The authors acknowledge the Sydney Informatics Hub and the use of the University of Sydney’s high performance computing cluster, Artemis. We express our gratitude for the previous contributions made to crop type mapping, and in particular, we acknowledge the pioneering work of Monfreda *et al.*<sup>2</sup>, which laid the foundation for our current research.

## Author contributions

F.M., F.N.T., and G.C. designed and conceptualized the study; T.H.N. collected the data; L.C. provided input data for cropland agreement map; F.M. developed the workflow and conducted the construction of the dataset; T.H.N., F.H.M.T., L.C., and F.M. analysed the data; F.H.M.T. drafted the manuscript; All authors contributed to the interpretation of the results, provided in-depth advice and commented/edited the manuscript.

## Competing interests

The authors declare no competing interests. The views expressed in this work are the authors’ only and do not represent FAO’s views or policy on the subject matter.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41597-024-03247-7>.

**Correspondence** and requests for materials should be addressed to F.M.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024