https://doi.org/10.5194/essd-2021-86
Preprint. Discussion started: 30 June 2021
© Author(s) 2021. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

GCI30: a global dataset of 30-m cropping intensity using multisource remote sensing imagery

Miao Zhang[1], Bingfang Wu[1,2] *, Hongwei Zeng[1,2], Guojin He[1,2], Chong Liu[3] *, Shiqi Tao[4], Qi Zhang[5,6], Mohsen Nabil[1,2,7], Fuyou Tian[1,2], José Bofana[1,2,8], Awetahegn Niguse Beyene[1,2,9], Abdelrazek Elnashar[1,2,10], Nana Yan[1], Zhengdong Wang[1,2]

1 State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, PR China
2 University of Chinese Academy of Sciences, Beijing 100049, PR China
3 School of Geospatial Engineering and Science, Sun Yat-Sen University, Guangzhou, 510275, PR China
4 Graduate School of Geography, Clark University, Worcester, MA 01610, USA
5 Department of Earth and Environment, Boston University, Boston, MA 02215, USA
6 Frederick S. Pardee Center for the Study of Longer-Range Future, Frederick S. Pardee School of Global Studies, Boston University, Boston, MA 02215, USA
7 Division of Agriculture Applications, Soils, and Marine (AASMD), National Authority for Remote Sensing & Space Sciences (NARSS), Cairo, New Nozha, Alf Maskan,1564, Egypt
8 Center for Agricultural and Sustainable Development Research (CIADS), Catholic University of Mozambique-Faculty of Agricultural Sciences, Cuamba 3305, Mozambique
9 Tigray Agricultural Research Institute, P.O. Box 492, Mekelle 251, Ethiopia
10 Department of Natural Resources, Faculty of African Postgraduate Studies, Cairo University, Giza 12613, Egypt

Correspondence to: Bingfang Wu (wubf@aircas.ac.cn), Chong Liu (liuc@mail.sysu.edu.cn)

Abstract. The global distribution of cropping intensity (CI) is essential to our understanding of agricultural land use management on Earth. Optical remote sensing has revolutionized our ability to map CI over large areas in a repeated and cost-efficient manner. Previous studies have mainly focused on investigating the spatiotemporal patterns of CI ranging from regions to the entire globe with the use of coarse-resolution data, which are inadequate for characterizing farming practices within heterogeneous landscapes. To fill this knowledge gap, in this study, we utilized multiple satellite data to develop a global, spatially continuous CI map dataset at 30-m resolution (GCI30). Accuracy assessments indicated that GCI30 exhibited high agreement with visually interpreted validation samples and *in situ* observations from the PhenoCam network. We carried out both statistical and spatial comparisons of GCI30 with existing global CI estimates. Based on GCI30, we estimated that the global average annual CI during 2016–2018 was 1.05, which is close to the mean (1.04) and median (1.13) CI values of the existing six estimates, although the spatial resolution and temporal coverage vary significantly among products. A spatial comparison with two other satellite based land surface phenology products further suggested that GCI30 was not only capable of capturing the overall pattern of global CI but also provided many spatial details. GCI30 indicated that single cropping was the primary agricultural system on Earth, accounting for 81.57% (12.28 million km^2) of the world's cropland extent. Multiple-cropping systems, on the other hand, were commonly observed in South America and Asia. We found large variations across countries and agroecological zones, reflecting the joint control of natural and anthropogenic drivers on regulating cropping practices. As the first global coverage, fine-resolution CI product, GCI30 can facilitate ongoing efforts to achieve sustainable development goals (SDGs) by improving food production while minimizing environmental impacts. The data are available on Harvard Dataverse: https://doi.org/10.7910/DVN/86M4PO (Zhang et al, 2020).

https://doi.org/10.5194/essd-2021-86
Preprint. Discussion started: 30 June 2021
© Author(s) 2021. CC BY 4.0 License.

1 Introduction

The interrelated targets of zero hunger, no poverty, and promoting sustainable agriculture have been collectively recognized as the core sustainable development goals (SDGs) by the United Nations (UN 2015; Wu et al. 2017; Whitcraft et al. 2019; Hinz et al. 2020). However, 750 million people are currently exposed to severe food insecurity, and the COVID-19 pandemic may have added approximately 100 million people to the total undernourished population in 2020 (FAO et al. 2020). Projections have further demonstrated that from 2010 to 2050, the world's agricultural production must increase by 70–110% to meet the demands caused by increasing populations and changing diets (Tilman et al. 2011). However, intensified agricultural activities have many ripple effects on terrestrial ecosystems, including forest degradation (Morton et al. 2006; Zeng et al. 2018), soil pollution (Lal 2002; Jankowski et al. 2018), and changes in carbon/water flux seasonality (Gray et al. 2014; Hao et al. 2015), which in turn damage the welfare of human society (Qi et al. 2020). To meet the critical human needs for food security and environmental sustainability, it is of major scientific significance to better understand how existing agricultural land resources are utilized, both locally and globally.

Cropping intensity (CI), defined as the number of crop planting and harvesting cycle(s) within a full year (Gray et al. 2014; Liu et al. 2020), offers a measure of cropland utilization that has profound implications for closing food production gaps and agricultural intensification (Challinor et al. 2015; Ding et al. 2016; Wu et al. 2018; Waha et al. 2020). CI also plays an essential role in crop modelling that assesses grain yield (Becker and Johnson 2001), soil quality (Sherrod et al. 2003), and the impacts of climate change (Pielke et al. 2007; Challinor et al. 2015). Given its importance, it is necessary to accurately estimate CI to improve the management of agricultural activities as well as their interactions with other physical components of the Earth system. Before the advent of remote sensing, information about CI could be estimated only based on limited agricultural census data, but these data are often outdated and variable in accuracy (Liu et al. 2020). Remote sensing has revolutionized our ability to estimate CI, especially at continental to global scales (Liu et al. 2020). The presence of crop growth and senescence phenology constitutes the most characteristic temporal feature of agricultural practices, and numerous attempts have been made to link high temporal frequency vegetation index time series to CI identification. Yan et al. (2014; 2019) developed a growing season peak detection-based algorithm and used it to monitor the CI spatiotemporal change in China. A similar approach was adopted by Kotsuki and Tanaka (2015) to derive a global crop calendar dataset containing CI metrics. Despite their prominent contributions to cropland intensification assessments, most existing CI products have coarse spatial resolutions, giving rise to the common presence of mixed pixels that can lead to a decreased CI mapping accuracy. To alleviate this issue, in recent years, fine-resolution optical satellite sensors, such as Landsat and Sentinel-2, have been employed to extract CI information. For example, Jain et al. (2013) found that fine-resolution satellite imagery can more accurately depict the CI pattern in smallholder agriculture regions than can coarse-resolution satellite data. Hao et al. (2019) also reported an improved performance of CI identification using harmonized Landsat Sentinel-2 (HLS) data.

It is becoming increasingly clear that a global, fine-resolution CI product is essential for monitoring the ongoing cropland intensification process on Earth. However, to the best of our knowledge, such a dataset has not yet been

https://doi.org/10.5194/essd-2021-86
Preprint. Discussion started: 30 June 2021
© Author(s) 2021. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

created, reflecting the necessity of a generalizable CI mapping framework that is representative of diverse climate zones and cropping systems. To overcome this challenge, we proposed a phenophase-based CI mapping framework in our pilot study (Liu et al. 2020) with the use of multiple satellite data and the Google Earth Engine (GEE) platform (Gorelick et al. 2017), offering both methodological and practical bases for operationalizing a global fine-resolution CI product. Taking advantage of the proposed framework, the primary goal of this research is to advance and develop a global, spatially continuous CI map at a 30-m resolution (GCI30) using the full archive of Landsat, Sentinel-2 and MODIS data from 2016 to 2018. To achieve this goal, we regenerated the global cropland extent layer and modified the CI estimation algorithm on flooded rice paddies by considering the flooding/transplanting signals. The performance of GCI30 was examined with *in situ* data as well as with two existing global land surface phenology products containing CI metrics. With a much finer spatial resolution and global coverage, GCI30 is expected to contribute to our fundamental understanding of the dynamics of the Earth's terrestrial surface as well as the human role in land modification through agricultural activities.

2 Materials and Methods

2.1 GCI30 input data

2.1.1 Cropland extent

This study integrated an ensemble of multiple land cover/cropland layer products (Figure 1) to delimit the global cropland extent while masking out irrelevant non-cropland pixels for the period of 2016–2018. Readers can refer to Table S1 for detailed information on these land cover/cropland layer products. Spatially, FROM-GLC was selected for Europe, Africa, New Zealand, the majority of Asia, and part of Latin America. GFSAD30 was selected for tropical Asian islands, including Indonesia, Malaysia and the Philippines. In addition to these two global-coverage cropland extent products, several national or regional datasets, including ChinaCover, CDL, AAFC ACI, NLCD, MapBiomass, CLUM, SERVIR and INTA, were used because they have been extensively validated by local experts and hence exhibited high accuracies of cropland mapping. Their spatial extents cover China, the contiguous U.S., Canada, Alaska, Brazil, Australia, the Lower Mekong River basin (Myanmar, Thailand, Lao, Cambodia and Vietnam), and part of Argentina, respectively.

https://doi.org/10.5194/essd-2021-86
Preprint. Discussion started: 30 June 2021
© Author(s) 2021. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

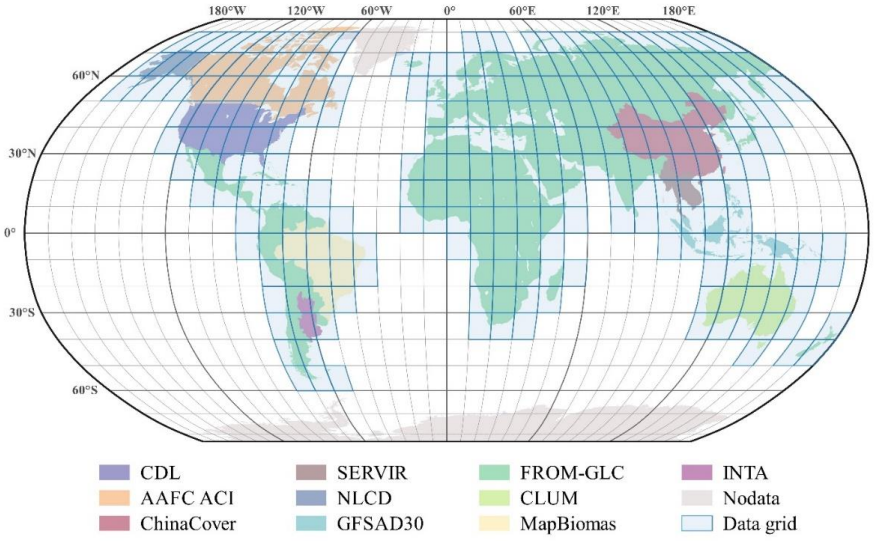

Figure 1: Spatial distribution of the land cover/cropland layer products used for the global 30-m cropland extent generation.

2.1.2 Satellite images and vegetation indices

All available images of top-of-atmosphere (TOA) reflectance from Landsat-7 ETM+, Landsat-8 OLI and Sentinel-2 MSI during 2016–2018 were used for global CI mapping via the GEE platform. Invalid observations, including clouds, cloud shadows, snow and saturated values, were identified and masked by the function of the mask (Fmask) algorithm (Zhu and Woodcock 2012; Qiu et al. 2019). To overcome the multi-sensor mismatch issue, we adopted an inter-calibration approach, which converted Sentinel-2 MSI and Landsat-8 OLI TOA reflectance data to the Landsat-7 ETM+ standard (Chastain et al. 2019). The calibrated images were used to composite the 16-day TOA reflectance time series of each pixel. Based on the harmonized TOA reflectance composite, the following vegetation indices were calculated:

$$\mathrm{NDVI} = \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + \rho_{RED}}$$

$$\mathrm{EVI} = 2.5 \times \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + 6 \times \rho_{RED} - 7.5 \times \rho_{BLUE} + 1}$$

$$\mathrm{LSWI} = \frac{\rho_{NIR} - \rho_{SWIR}}{\rho_{NIR} + \rho_{SWIR}}$$

where ρ_{BLUE}, ρ_{RED}, ρ_{NIR}, and ρ_{SWIR} are the TOA reflectance values of the blue, red, near-infrared, and shortwave-infrared bands, respectively. We also used the MOD13Q1 NDVI/EVI product and MOD09A1-derived LSWI in our study to fill data gaps caused by the vacancy of cloud-free Landsat/Sentinel-2 observations. After gap-filling, a

https://doi.org/10.5194/essd-2021-86
Preprint. Discussion started: 30 June 2021
© Author(s) 2021. CC BY 4.0 License.

weighted Whittaker smoother (Kong et al. 2019) was adopted to generate a dataset of smoothed, seamless image time series of vegetation indices at a resolution of 30 m.

2.2 Reference samples

The validation of the global CI product requires carefully constructed reference samples (Kontgis et al. 2015; Li et al. 2014; Liu et al. 2020). In this study, we constructed two independent reference datasets (termed RDsat and RDsite hereafter) (Figure S1) to evaluate the GCI30 performance. The first dataset, RDsat, was generated based on a visual interpretation of satellite time series via the Geo-Wiki platform (http://www.geo-wiki.org/). Based on the global segmentation of agroecological zones (Gommes et al. 2016; 2017, termed AEZs here after) (Table S2), we applied a stratified sampling approach to ensure that RDsat was geographically representative across the globe. We divided all 65 AEZs into four categories based on their cropland proportions: VL (cropland proportion < 4%), L (4% ≤ cropland proportion < 15%), M (15% ≤ cropland proportion < 40%) and H (cropland proportion ≥ 40%). For each category, 1000 points were randomly collected only within the cropland extent, and their phenological cycles during the period of 2016–2018 were visually counted. We kept only well-interpreted points with high-level confidence, which led to 3744 sample records. The second dataset, RDsite, was derived from PhenoCam dataset v2.0 (Richardson et al. 2018; Richardson et al. 2018; Seyednasrollah et al. 2019), which has been widely used as a robust *in situ* reference for remotely sensed phenology metric validation. Globally, there are 115 PhenoCam sites on cropland, and a total of 40 sites were collected after removing those with data records of less than one year (Table S3). For each selected site, we used the green chromatic coordinate (GCC) index and *in situ* phenology camera image time series for cropping cycle number identification. It should be noted that not all the selected PhenoCam sites precisely covered a period matching 2016-2018. For instance, the site with ID of 'usof6' in Table S3 provided measurements from May 2018 to September 2020, which was out of the study period used for our GCI30 product. However, to make full use of these measurements, we implemented our approach of CI identification by aligning the study period with the period containing the measurements at each of these sites. Thus, some CI identification covered longer periods than the three-year length (2016-2018).

2.3 Global cropping intensity mapping method

2.3.1 CI mapping on non-flooded croplands

We adopted the phenophase-based approach (Liu et al. 2020) to map CI in the non-flooded cropland (relative to the flooded rice paddy). The methodology can be divided into three main steps, including 1) the construction of NDVI time-series for cropland pixels, 2) the identification of phenological phases of crops, and 3) the calculation of the cropping cycles. A descriptive summary of each step is provided as follows. First, NDVI time series based on satellite data from Landsat, Sentinel-2 and MODIS were fused with the observations aligned with TOA reflectance of Landsat-7 ETM+ in a 16-day interval. Gap filling and linear models were performed to take into account the lack of sufficient observations as well as the mismatch of spatial resolutions among different sensors (Liu et al. 2020). Then, given the smoothed, seamless NDVI time series, transitioning points defined as the 50% of the NDVI amplitude (i.e., difference between minimum and maximum values) were detected and used to separate the entire NDVI time series trajectory

https://doi.org/10.5194/essd-2021-86
Preprint. Discussion started: 30 June 2021
© Author(s) 2021. CC BY 4.0 License.

into staggered segments of growing or non-growing periods. The transitioning points were labelled as the mid-greenup and mid-greendown points, which together can define the growing and non-growing periods of the crop phenophase. Finally, we enumerated the total numbers of the mid-greenup (N_{up}) and mid-greendown transitioning points (N_{down}), respectively for each pixel, and took the lower value as the number of the potential cropping cycles (N_{pc}) of that pixel.

$$N_{pc} = \min\{N_{up}, N_{down}\}$$

We set a lower limit of the growing period to 48 days to remove the false cropping cycle(s) (N_{fc}) caused by NDVI time series outliers (Yan et al. 2019). Since the whole study period covers three years from 2016 to 2018, the actual CI can be derived as:

$$\mathrm{CI} = \frac{N_{pc} - N_{fc}}{3}$$

It should be noted that using the binary phenophase profile itself is not effective enough for identifying the continuous cropping type because such a cropping type exhibits a lower degree of seasonality. Therefore, for each cropland pixel, we calculated the coefficient of variation (the ratio of the standard deviation to the mean, termed CV hereafter) of the NDVI time series and adopted a threshold method to determine whether the pixel belonged to the continuous cropping type (low CV value). Specifically, within each AEZ, half of the RDsat samples labelled continuous cropping (if they existed) were adopted to obtain the CV threshold. This method generated an independent continuous cropping type layer, which was integrated with the initially derived CI result. A more detailed description of the CI mapping framework can be found in Liu et al. (2020).

2.3.2 CI mapping on flooded rice paddies

Flooded rice paddy, which accounts for 12% of the global cropland area but feed approximately half of the population (FAOSTAT 2010; Ding et al. 2020), bear special mention in our study because it supports the only staple grains that need to be transplanted (Dong and Xiao 2016), resulting in a relatively short non-growing period that may be mistakenly missed when using the abovementioned approach. This issue becomes more prominent in areas with high cloud cover (e.g., Monsoon Asia). Therefore, we modified our approach in flooded rice paddy areas by considering the influence of the "flooding/transplanting signal" on the created phenophase profile (Figure 2). Similar to the approach for non-flooded croplands, the NDVI time series trajectory was used to generate an initial phenophase profile (Figure 2a). Then, within each identified growing season, the flooding/transplanting signals were detected and recognized based on the following criteria: LSWI > EVI or LSWI > NDVI (Figure 2b), indicating that the water signal dominates the pixel spectral performance (Xiao et al. 2005; Dong et al. 2015; 2016). We regard the "flooding/transplanting" period as a non-growing phenophase. Thus, the initial phenophase profile can be divided into two segments accordingly (Figure 2c), reflecting the reality of double-season rice planting cycles. Finally, the CI information was determined by enumerating the transition points between different cropping cycles. Due to data limitations, this specific CI identification approach was applied only in southern China (AEZs C33, C37, C40, C41,

https://doi.org/10.5194/essd-2021-86
Preprint. Discussion started: 30 June 2021
© Author(s) 2021. CC BY 4.0 License.

and C42) and the Lower Mekong River basin, where the paddy rice type was included in the land cover type scheme (derived from ChinaCover and SERVIR, respectively).

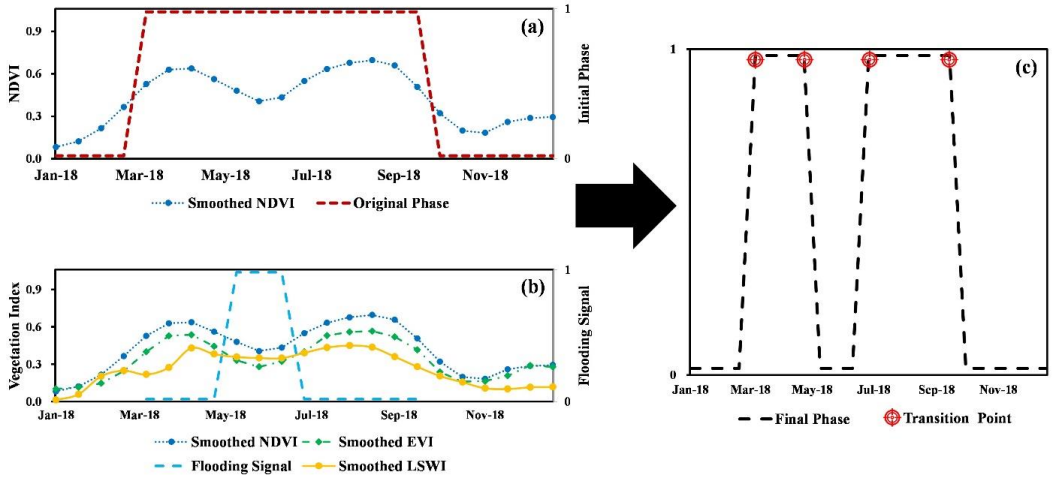

Figure 2: Illustration of the specific CI identification for a flooded rice paddy pixel.

2.4 Accuracy assessment

Based on the RDsat and RDsite datasets, the accuracy assessment of GCI30 was conducted in two ways. In the first validation method, we directly evaluated the difference between the reference and estimated results. Here, the total number of cropping cycles (termed TNCC hereafter) rather than the actual CI value was used to avoid decimals. Four complementary indicators, systematic error (*SE*), mean absolute error (*MAE*), root mean square error (*RMSE*) and coefficient of determination (R^2), were calculated as follows:

$$SE = \frac{1}{N}\sum_{i=1}^{N}(\hat{f}_i - f_i)$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|\hat{f}_i - f_i|$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{f}_i - f_i)^2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(\hat{f}_i - f_i)^2}{\sum_{i=1}^{N}(\hat{f}_i - \bar{f})^2}$$

https://doi.org/10.5194/essd-2021-86
Preprint. Discussion started: 30 June 2021
© Author(s) 2021. CC BY 4.0 License.

where $\hat{f}_i$ and f_i are the estimated and reference cropping cycle number(s) for sample pixel i, respectively. N represents the number of samples, and $\bar{f}$ is the mean cropping cycle number value of all samples. In addition to directly quantifying the mapping errors, we further reclassified the GCI30 result into four categories: single cropping ($0 < CI \leq 1$), double cropping ($1 < CI \leq 2$), triple cropping ($2 < CI \leq 3$) and continuous cropping. We obtained the confusion matrix and calculated quantitative metrics, including overall accuracy (OA), kappa coefficient, producer accuracy (PA) and user accuracy (UA). Due to the limited sample sizes of RDsite, the classification-based accuracy assessment was conducted only for RDsat.

2.5 Comparison with other global products

Comparison of GCI30 with other global products or studies was conducted statistically and spatially at global scale. We first compared and evaluated the statistical differences of global CI calculated based on GCI30 with six existing statistical-based or satellite-based researches or products (Didan and Barreto 2016; Gray et al. 2019; Wu et al. 2018; Siebert et al. 2010; Ray and Foley, 2013). Then, we compared GCI30 with two existing global land surface phenology products containing CI metrics, namely, NASA's Vegetation Index and Phenology V004 with a 0.05° pixel size from 1981 to 2014 (Didan and Barreto 2016, termed VIP4 hereafter) and Moderate Resolution Imaging Spectroradiometer Land Cover Dynamics Version 6 with a 500 m pixel size from 2001 to 2018 (Gray et al. 2019, termed MDC12Q2 hereafter) to assess the spatial discrepancies between GCI30, VIP4 and MCD12Q2. To minimize uncertainty caused by temporal disagreement, we selected only the 2014 VIP4 and the 2016–2018 averaged MCD12Q2 data, within which the "Number of Seasons" layer of VIP4 and the "NumCycles" layer of MCD12Q2 were extracted for intercomparison. We upscaled GCI30 to 0.05° and 500 m using the majority algorithm to match the spatial resolution of VIP4 and MCD12Q2, respectively. The same reclassification procedure described in Section 2.4 transformed the actual CI value of GCI30 and MCD12Q2 to match the VIP4 dataset's integer value range (0, 1, 2, 3), except for the continuous cropping and non-cropland pixels that were excluded from our comparison. We generated the difference maps among GCI30, VIP4 and MCD12Q2 to understand the overall overestimation or underestimation of our mapped CI results across continents.

3 Results and Discussion

3.1 Reliability of GCI30

We examined GCI30 performance by generating a scatter plot of estimated and reference TNCC derived from RDsat and RDsite, respectively (Figure 3). In general, GCI30 could provide reliable estimation results across different agro-environmental and management conditions, with relatively small *MAE* and *RMSE* values (equal to or less than 0.4 and 0.92, respectively) using the two reference datasets. Referring to R^2, GCI30 captured over 91% of the variation in RDsite-derived TNCC and over 56% of the variation in RDsat-derived TNCC. The discrepancies between RDsat-derived metrics and RDsite-derived metrics were mainly attributed to the differences in sample size and crop planting diversity. Specifically, the network of PhenoCam spots is spatially sparse, and most cropland sites are distributed in the United States featuring single cropping systems (Figure S1). There were slight systematic underestimations

https://doi.org/10.5194/essd-2021-86
Preprint. Discussion started: 30 June 2021
© Author(s) 2021. CC BY 4.0 License.

(negative *SE*) for GCI30, with the larger bias level occurring for RDsite. This result is consistent with Liu et al. (2020), indicating an overall conservative CI estimation by GCI30. In addition, we found that larger estimation errors were commonly observed in samples with more cropping cycles. This tendency was not surprising because of the accumulated errors from every aspect of information extracted from remote sensing (Defourny et al. 2019), including data acquisition, time series modelling of vegetation indices, and phenological cycle identification. Therefore, we may expect that GCI30 faces larger challenges in terms of analysing multiple cropping systems.

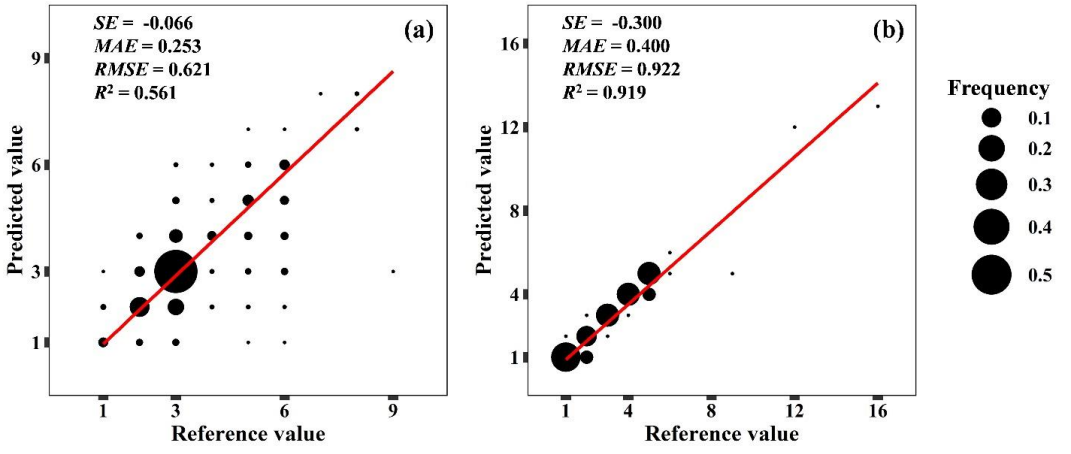

Figure 3: GCI30 accuracy assessment based on RDsat (a) and RDsite (b). The red line represents the linear fitting line with the intercept forced to 0. The frequency of a specific reference-prediction value pair is proportional to its point size. Samples identified as continuous cropping types were excluded.

Figure 4 further displays the spatial pattern of RDsat-based TNCC estimation bias across the globe. From 2016 to 2018, 79.8% of the points exhibited unbiased predictions. Among the pixels with disagreement (i.e., non-zero bias), the majority were associated with one or two cropping cycle difference(s). Overall, there were more underestimation points (12.2%) than overestimation points (8.0%). Spatially, negative biases were mainly distributed in high CI regions, including the Pampas (AEZ C26), Central Eastern Brazil (AEZ C23), the Gulf of Guinea (AEZ C03), East African Highlands (AEZ C02), South of Himalaya (AEZ C44), and Huanghuaihai Plain (AEZ C34), altogether forming a northward "underestimation belt" along the longitudinal gradient. Given the conservative CI estimation algorithm, it was somewhat unexpected to observe overestimation errors primarily concentrated in Western Europe (AEZ C60) and Ukraine to the Ural Mountains (AEZ C55), where the single cropping practice dominates due to limited hydrothermal conditions (Wu et al. 2018). The negative errors could possibly be due to the large variations in the NDVI time series during the growing season (Zhang 2015; Qiu et al. 2020), highlighting the complex suite of biotic and abiotic processes that can obscure the effectiveness of the phenophase-based CI mapping framework.

https://doi.org/10.5194/essd-2021-86
Preprint. Discussion started: 30 June 2021
© Author(s) 2021. CC BY 4.0 License.

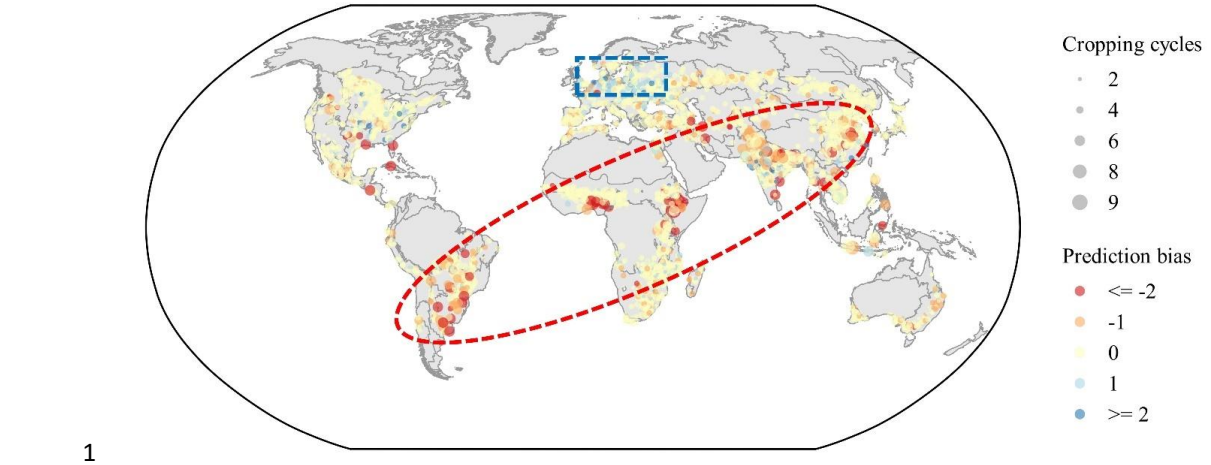

Figure 4: Spatial distribution of RDsat-based TNCC bias. The actual TNCC is proportional to its point size. Sample points identified as the continuous cropping type were excluded. The blue rectangle indicates overestimations, and the red ellipses region represents the northward "underestimation belt" along the longitudinal gradient.

Following the reclassification procedure illustrated in Section 2.4, we derived the corresponding confusion matrix of GCI30 using RDsat samples, with the quantitative accuracy metrics shown in Table 1. We found that GCI30 had reasonable classification performances, with OA and kappa coefficients greater than 92% and 0.72, respectively. Regarding the classes of CI, single-cropping systems were associated with more robust classification results than were multiple-cropping systems. Comparatively, the single-cropping class was more subject to commission errors than omission errors (PA > UA), while the opposite tendency (PA < UA) was observed for the double- and triple-cropping classes. Continuous cropping is a fundamentally different agricultural land use management type from others. Here, we found that the continuous cropping class of GCI30 had a higher PA (93.1%) than UA (77.0%). A possible explanation for this result is likely attributed to the threshold-based method for continuous cropping identification. Some noncontinuous cropping systems may also exhibit low CV values, leading to a relatively high commission error level. Notably, although a stratified sampling strategy was conducted for creating RDsat, its sample size was still unbalanced among the different CI classes. The single-cropping class alone occupied 88.9% of the total number of samples. Therefore, future efforts of GCI30 validation need to emphasize the inclusion of more samples with multiple-cropping systems.

Table 1: Confusion matrices of GCI30.

	CI class type	Single	Double	Triple	Continuous	UA (%)
Reference	Single	3059	186			94.3
	Double	52	346	2	3	85.9
	Triple	1		6	2	66.7
	Continuous	9	9	2	67	77.0
	PA (%)	98.0	64.0	60.0	93.1	
	OA (%)	92.9				
	Kappa	0.728				

https://doi.org/10.5194/essd-2021-86
Preprint. Discussion started: 30 June 2021
© Author(s) 2021. CC BY 4.0 License.

3.2 Spatial pattern of GCI30

GCI30 provides the first spatially continuous map of global CI at a 30-m resolution (Figure 5). Based on this map, a heterogeneous pattern in CI compositions across continents was found, which are subject to varying anthropogenic and climate conditions. Overall, as expected, single cropping was the primary agricultural system on Earth, accounting for 81.57% (12.28 million km^2) of the world's cropland extent. Double cropping, on the other hand, was typically implemented in Asia, South America and the Nile River basin of Africa, together occupying 17.42% (2.62 million km^2) of global croplands. Comparatively, the proportions of triple and continuous cropping were quite small, with their distributions mainly limited to Southeast Asia. According to the area statistics at five-degree intervals, we found that the area of single cropping reached 54% or higher in all latitude and longitude zones. The double cropping distribution along latitude peaked in intervals ranging from 20°N to 40° N, which encompassed China and India, the two most populous countries in the world. Along longitude, double cropping was mainly concentrated in three zones: 55°W to 60°W, 75°E to 90°E and 100°E to 125°E. These regions are commonly characterized by warm and humid climates, except for the Nile River basin, in which highly developed irrigation infrastructures have been traditionally used to support intensive farming practices (Wu et al., 2021). Over 75% of triple and continuous cropping areas are located within tropical zones (5°S to 5°N). The tropical rainforest climate of these regions ensures sufficient water and heat supplies for crop growth throughout the year (Köppen et al. 2011).

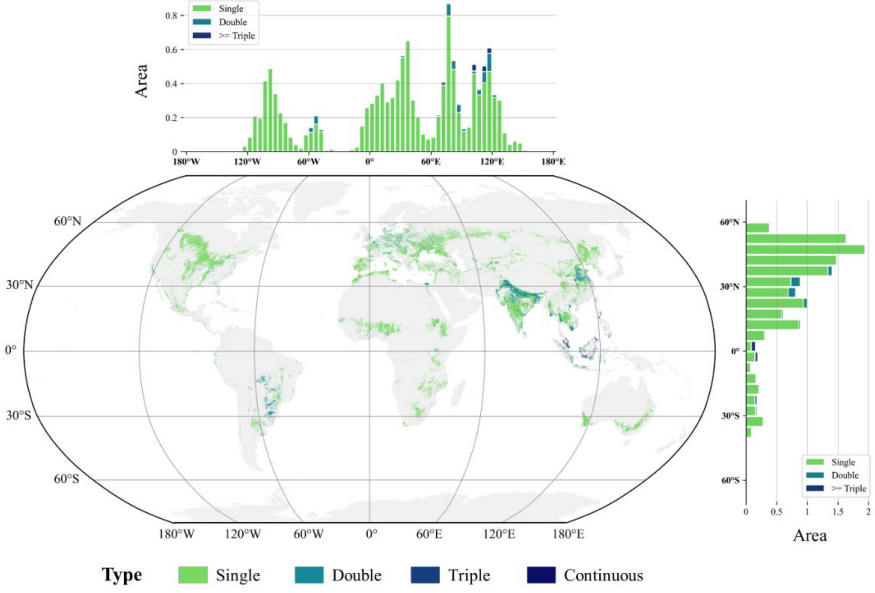

Figure 5: Geographical distribution of global CI types during 2016 to 2018 identified by GCI30. The area statistics along latitude and longitude are derived with an interval of five degrees. The area unit is million km^2.

Figure 6 displays the GCI30-based TNCC statistics at the continent level. We combined Australia and Oceania (New Zealand, Melanesia, Micronesia and Polynesia) due to the rarity of cropland on these two continents. Globally, South

https://doi.org/10.5194/essd-2021-86
Preprint. Discussion started: 30 June 2021
© Author(s) 2021. CC BY 4.0 License.

America exhibited the most intensified cropping level, followed by Asia and Europe. Specifically, the average TNCC values were 3.67, 3.38 and 3.07 for South America, Asia and Europe, respectively. South America and Asia also possessed the largest standard deviations of TNCC, indicating the inherent diversity of agricultural activities within these two continents, as weather conditions directly affect cropping practices (Iizumi et al., 2015). For example, in Asia, triple and continuous cropping systems were distributed in Southeast Asia, including Indonesia, Malaysia, southern Thailand and the Mekong River Delta in Vietnam. Double cropping was concentrated in the North China Plain, Ganga River basin and southern China, while the rest of Asia was dominated by a single cropping pattern, covering Central Asia, Northeast Asia, and southern India. Following these continents, moderate TNCC was found in North America (2.93±0.54) and Africa (2.78±0.71). Among all continents, the lowest TNCC occurred in Australia and Oceania (2.31±0.77), where arid and semiarid climate types are dominant (Köppen et al. 2011; Beck et al., 2018).

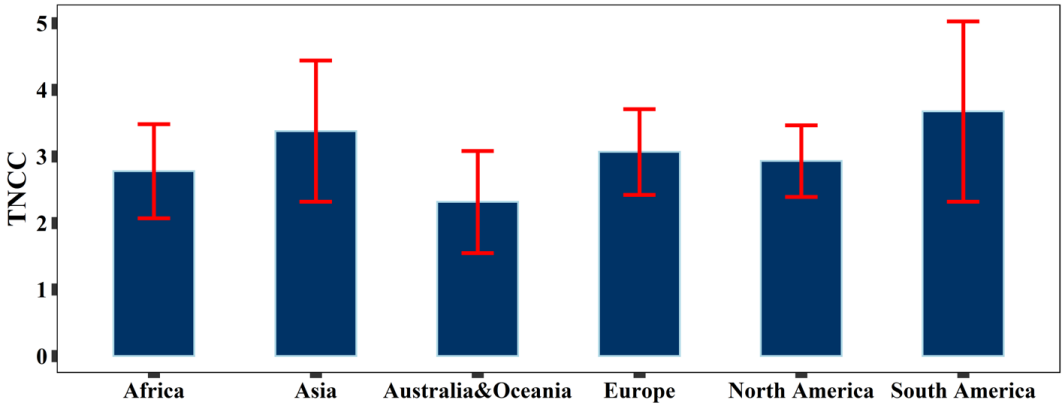

Figure 6: Statistics of GCI30-based TNCC during 2016 to 2018 at the continent level. The red line indicates the standard deviation.

At the global scale, the average CI pattern was heterogeneous across countries and AEZs (Figure 7, left panel). Countries with the highest average CI levels were commonly detected in Asia (Bangladesh, Vietnam, Philippines) and Latin America (Guyana, Paraguay, Suriname, Haiti, and Dominica). Together with Egypt, these top 10 countries exhibited TNCC values greater than 4.1 during 2016–2018. In contrast, low to moderate CI levels were typically found in high-latitude countries, such as Canada, Russia, and Mongolia. In addition to the latitude gradient, we found that the diversity of cropland management played a critical role in shaping the CI pattern. For example, some high-latitude European countries (Germany, Poland, Belarus, etc.) showed relatively high CI levels due primarily to their advanced cropland management practices (Guo, 2019). Rainfed agricultural practices lead to fewer cropping cycles in the Middle East and North African countries, except for Egypt, where most croplands are irrigated (Wu et al. 2018). Taking climate conditions into account, the heterogeneity of global CI becomes even more prominent among different AEZs. Arid regions, which cover vast areas in Africa, Australia, and Central Asia, are associated with fewer cropping cycles due to a lack of water for irrigation (Chiew et al., 2011; Guo et al., 2018) and less developed agricultural infrastructures (Mason-D'Croz et al., 2019). In contrast, intensive farming is widely distributed in humid and low-

https://doi.org/10.5194/essd-2021-86
Preprint. Discussion started: 30 June 2021
© Author(s) 2021. CC BY 4.0 License.

latitude areas such as South China and the Mekong Delta. Among all 65 AEZs, the Huanghuaihai Plain in China had the highest CI, followed by the Amazon rainforest region of South America and Taiwan Province. The lowest CI occurred in the Australian Desert, with an average TNCC less than 2.

Overall, countries and AEZs with intensive farming are more subject to internal variability, as reflected by higher standard deviations (Figure 7, right panel). Globally, there are 14 countries and 7 AEZs exhibiting standard deviations greater than 1.2, and most of them are located in South America and Asia. Regions with low CI averages but high CI standard deviations were observed only on the western coast of North America and Queensland to Victoria in Australia, where partial irrigation in the former (Xie et al., 2019) and unstable rainfall in the latter resulted in diversified cropping intensities among years (King et al., 2020). The high standard deviations in Australia and Oceania mainly resulted from the high within-country/zonal heterogeneity, which may encompass aspects including the exceptionally variable climate with the prevalence of floods and droughts (King et al., 2020) and the annual shifting of crop types, as well as cultivated and fallow lands (Song et al., 2017). In addition to these regular drivers, the political situation may cause CI spatiotemporal diversity. Notably, for instance, we found an unusually high standard deviation in Afghanistan, which was caused by both crop failure during the emergence to early development stages due to adverse weather conditions (Rousta et al., 2020) and abandoned cropland resulting from armed conflicts and refugee migrations (Iqbal et al., 2018; Galdo et al., 2020).

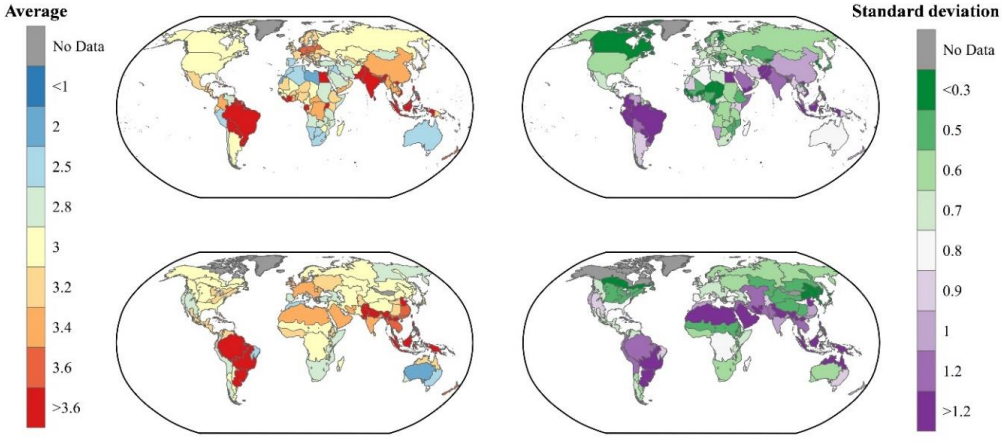

Figure 7: Average and standard deviation (std) of TNCC during 2016 to 2018 at the national and AEZ levels.

3.3 Cross comparison with other studies

Table 2 illustrates the information on the methods, input data, spatial resolution, and statistical average CI from GCI30, and estimates from other six research or products. Based on GCI30, the global average CI during 2016–2018 was 1.05 (the continuous cropping pixel excluded). Statistically, our CI is in a remarkably high agreement level with estimates based on the existing six estimates (mean CI: 1.04, median CI: 1.13), despite their significantly varying spatial

https://doi.org/10.5194/essd-2021-86
Preprint. Discussion started: 30 June 2021
© Author(s) 2021. CC BY 4.0 License.

resolution and temporal coverage. The minimum CI among the seven studies was estimated to be 0.77 per year based on the total cropland extent and the total harvested crop area reported by the agricultural statistics database of the United Nations Food and Agriculture Organization (FAO) FAOSTAT (http://www.fao.org/faostat/en/) (Siebert et al. 2010). CI estimated by GCI30 is slightly higher than that derived from the FAO statistical database. The CI of other existing products listed in table 2 ranges from 0.84, as estimated by Ray and Foley (2013), to 1.26, as evaluated by Wu et al. (2018). The statistics-based CI values estimated by FAO statistics, Siebert et al. (2010) and Ray and Foley (2013) are lower than those estimated based on remote sensing data including the GCI30 and those estimated by VIP4 and Wu et al. (2018) using AVHRR satellite observation data. The main reason is that statistics-based CI could not exclude the fallow land area as the statistical information is usually lack of statistical information on fallow land while fallow land could be easily and precisely identified using remote sensing (Zhang et al., 2014a, b) and excluded when generating satellite-based CI products. Our CI is also less than those CI products derived from AVHRR (VIP4 and Wu et al. (2018)). On the one hand, the actual harvest frequency estimated by Wu et al. (2018) might overestimate the annual harvest areas and accordingly overestimate the cropping intensity. On the other hand, GCI30 systematically underestimates the cropping intensity when the harvest window is narrow between to growing seasons as a valid phenology season should include both green-up and green-down segments based on the GCI30 algorithm (Liu et al., 2020). Interestingly, our CI is exactly the same as the global average from MCD12Q2 for the years 2016 to 2018.

Table 2: Comparison of the cropping intensity resulting from three different studies and two global products.

Existing products and studies	Methods	Input data	Temporal range	Spatial resolution	Global annual average CI
MCD12Q2	Phenometrics-based method	Time series MODIS EVI	2001 to 2018	500 m	1.05 for 2016 to 2018
VIP4	Half-maximum VI approach	Time series NDVI and EVI2 derived from both AVHRR and MODIS	1981 to 2014	0.05°	1.19 for the year 2014
Wu et al.	Peak-counting	Time series GIMMS NDVI3g	2009-2011	8 km	1.26
Siebert et al.	The ratio of harvested crop areas to total cropland area by excluding fallow land	MIRCA crop areas	2000	5 arcmin	1.13
Ray and Foley	The ratio of harvested cropland to total cropland	FAO statistics	2000-2011	National based	0.84
FAO	The ratio of harvested cropland to total cropland	FAO statistics	2010	Global based	0.77
This study	Phenophase-based approach	Time series NDVI derived from integrated dataset from Landsat 7&8, Sentinel-2 and MODIS	2016-2018	30 m	1.05

https://doi.org/10.5194/essd-2021-86
Preprint. Discussion started: 30 June 2021
© Author(s) 2021. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

We further compared the spatial pattern of the global CI difference between GCI30 and the two global land surface phenology products (MCD12Q2 and VIP4), as displayed in Figure 8. Overall, all three products revealed consistent CI estimations across continents, with zero-difference pixels reaching 79% or higher and the majority of CI differences ranging from -1 to 1. Spatially, positive CI difference values were commonly found in Southeast Asia, the Indian subcontinent and some parts of Europe. Negative CI difference values, on the other hand, were mainly detected in North and South America. There were also discrepancies when these two phenology products were used as the baselines. Referring to MDC12Q2, there were many pixels showing positive values, especially in Africa and mainland China. However, the opposite tendency was observed using VIP4, which exhibited vast negative pixel distributions in Europe and North China.

To further explore how the CI difference varied over space, we selected four 15° × 12° subregions (North America, South America, South Asia and East Asia, which are labelled A, B, C and D, respectively, in Figure 8) that were representative of the global diversity of crop species, climate types and management conditions. In general, substantial variations were detected through these spatially explicit maps. The strongest agreement between GCI30 and MCD12Q2 was found in East Asia (83% of zero values), followed by North America and South Asia, with over 75% agreement. The lowest agreement level occurred in South America, where 20% of the GCI30 estimations showed positive or negative CI differences compared to the MCD12 output. Comparatively, the significant differences and corresponding spatial distributions between the GCI30 and VIP4 outputs had a low level of agreement, although the percentages of pixels with zero difference reached 50% or higher for all subregions. Specifically, three out of the four subregions had at least one-fifth of the pixels featuring negative CI differences. The largest negative disagreement was detected in East Asia, where 41% of the total cropland area had negative values, while North America and South America also had considerable negative proportions. Finally, in South Asia, the positive and negative pixel percentages were almost equal, i.e., half and half. Neither MCD12Q2 nor VIP4 should be considered as ground truth. In fact, the reliability of these two land surface phenology products, especially VIP4, is affected by several factors, including a coarse spatial resolution, temporal mismatch and algorithm structure differences when compared to GCI30. Taking East Asia as an example, the “cross-year season cycle” phenomenon (Liu et al. 2020) caused by winter wheat planting could lead to one more “partial growing season” being detected by VIP4 (Didan and Barreto 2016), which largely explains why the CI difference between GCI30 and VIP4 shows an outstanding underestimation pattern.

https://doi.org/10.5194/essd-2021-86
Preprint. Discussion started: 30 June 2021
© Author(s) 2021. CC BY 4.0 License.

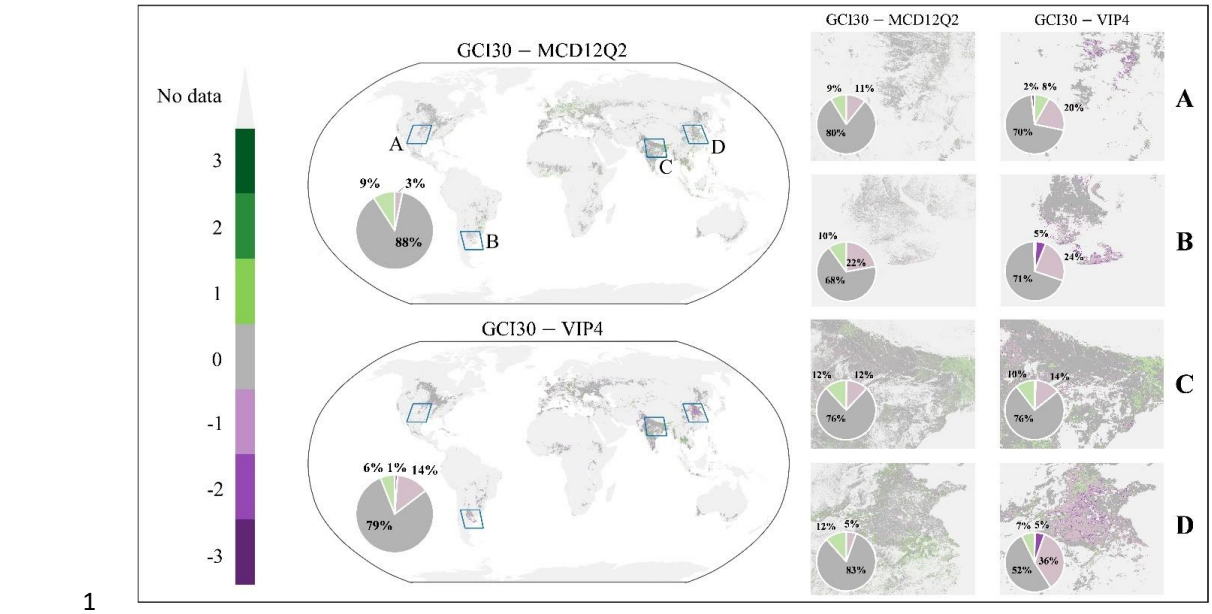

Figure 8: Spatial patterns and statistics of the CI differences among GCI30, MCD12Q2, and VIP4.

3.4 Advantages and limitations of GCI30

As the first global 30-m CI product, GCI30 depicts the worldwide diversity of agricultural land use management in a spatially explicit manner that has not been fully revealed by existing studies or datasets. Given the CI distribution with a fine spatial resolution, GCI30 is associated with reduced uncertainties caused by the mixed pixel effect and provides a more accurate global CI estimation. In addition to the improvement of mapping accuracy, GCI30 has the potential to monitor landscape-scale cropping practices on fragmented land parcels by smallholders, which comprise over half of the rural populations in developing nations that are most vulnerable to food security and environmental challenges (Morton et al. 2006; Jain et al. 2013; Lowder et al. 2016; Liu et al. 2020). Compared with the generalizable crop phenophase pattern, the GCI30 algorithm is not only efficient in mapping the CI distribution across various AEZs but is also flexible enough to be improved with updated data inputs. For example, the Harmonized Landsat and Sentinel-2 surface reflectance dataset (Claverie et al. 2018), with a 5-day revisit interval and a 30-m pixel size, is expected to enhance the global CI mapping performance once its worldwide coverage is ready. The successful production of GCI30 on the GEE platform illustrates a paradigm of mapping farming practices that is globally consistent and locally relevant using state-of-the-art cloud computing resources (Lewis et al. 2017; Tamiminia et al. 2020; Amani et al. 2020). It inspires future global fine-scale agricultural research that was previously not applicable.

A large number of natural factors and anthropogenic drivers are related to CI at the planetary scale. Accuracy assessments show that GCI30 explained over 91% and 56% of the sample variations examined by RDsat and RDsite, respectively (Figure 3). The errors of GCI30 could be related to the uncertainties of input data and limitations of the algorithm. The reliability of the cropland extent is a major factor constraining CI mapping performance. To minimize

https://doi.org/10.5194/essd-2021-86
Preprint. Discussion started: 30 June 2021
© Author(s) 2021. CC BY 4.0 License.

this effect, we integrated an ensemble of 11 land cover or specific cropland layer products for acquiring global cropland extent at a 30-m resolution for 2016–2018. Despite the high overall accuracy of the generated cropland extent, classification errors still exist, especially in some regions of Africa and Asia where small cropland patches are mixed with other land covers (Gong et al. 2013; Xiong et al. 2017). The GCI30 algorithm depends heavily on crop phenological information derived from the time series of vegetation indices. We found that the spatial pattern of the invalid observation count of the 16-day harmonized TOA reflectance composite (Figure S2) matched well with those of RDsat sample bias in some cloudy regions, such as the Gulf of Guinea and East African Highlands (Figure 4), indicating that the performance of GCI30 may be limited in areas suffering from unfavourable weather conditions or extreme seasonal imbalances of clear observations. We further evaluated the uncertainty of the GCI30 at global scales from the perspective of data quality. We found higher uncertainty in Amazon, western Africa, South & Southeast Asia as well as south and southwest China than other regions due to the high cloudy frequency. In high latitude regions in North Hemisphere, the uncertainty is also high mainly due to the snow cover in winter but with limited impacts on GCI30 as there are limited cropland at high latitude regions. Rice paddies are fundamentally different from non-flooded croplands, which affects CI mapping performance. We designed a specific rice paddy CI identification approach by considering the influence of the "flooding/transplanting phase". While promising, its application was limited due to the lack of a specific rice paddy layer. Therefore, more improvements can be included, such as integrating SAR data time series for more accurate flood signal detection (Singha et al. 2019).

Additionally, it is noteworthy that the GCI30 product provides insight only into the current actual cropping intensity; however, it is not linked to the potential cropping cycles. To assess the CI gaps between potential and actual situations, climate models could be introduced to simulate the potential cropping cycles under long-term average weather conditions. The proposed method can be readily applied to other years to retrieve long-term CI maps, which will fill in the knowledge gaps of decades of long cropping practices and interannual variations (Iizumi et al., 2015). Such information is key to improving our understanding of the CI response to climate in a more granular manner.

4 Data availability

The GCI30 product is available on Harvard Dataverse: https://doi.org/10.7910/DVN/86M4PO (Zhang et al, 2020). It is the first 30-m resolution CI dataset covering a global extent. The GCI30 product was tiled into 504 files in GeoTIFF format with geographic projection. To be precise, the spatial resolution of the product is 0.00026949459 degrees. Each GCI30 tile encompasses an area of 10 degree × 10 degree and is named in the following format: 'Cropping_Intensity_30m_2016_2018_$regions$.tif'. The "regions" in the file name are determined as follows: N/S (Northern Hemisphere or Southern Hemisphere) followed by a two-digit latitude label of the tile's top-left corner, and E/W (Eastern Hemisphere or Western Hemisphere) followed by a three-digit longitude label of the tile's top-left corner. Each GeoTIFF file includes two layers. The first layer is the average CI during the three years from 2016 to 2018, with the noData value or masked areas assigned as -1. The valid values for the first layer are 1, 2, and 3, representing single cropping, double cropping or triple cropping, respectively. The second layer is the TNCC from 2016 to 2018 with a noData value or masked areas assigned to -1. The continuous cropping type or the number of cropping cycles

https://doi.org/10.5194/essd-2021-86
Preprint. Discussion started: 30 June 2021
© Author(s) 2021. CC BY 4.0 License.

larger than 3 per year is assigned as 127 in the abovementioned two layers. We also included a shapefile of the tiles named 'CroppingIntensity_tiles_shapefile.rar' in the repository so that users could easily find their target tiles. The GCI30 product was generated on the GEE platform using JavaScript language developed by the authors. The GEE script as well as the auxiliary data of GCI30 algorithm as an illustration for one tile is open to all potential users and available at https://code.earthengine.google.com/f23108c6c47025c4acbd90b57c0753f7.

5 Conclusions

In this study, we utilized multisource remote sensing data, including Landsat, Sentinel-2 and MODIS data, to produce the first 30-m CI map at a global scale. Based on the phenophase-based mapping framework, GCI30 identified CI by enumerating the transition points between growing and non-growing periods. To improve the CI mapping performance on flooded rice paddies, we specifically considered the influence of the "flooding/transplanting signal" on the created phenophase profile. Accuracy assessments and inter-comparisons with existing land surface phenology products suggested that GCI30 was reliable across different climate zones and cropping systems. According to GCI30, we estimated that the global average CI was 1.05 during 2016–2018. We found that single-cropping systems occupied more than 80% of the world's cropland extent, while multiple-cropping practices were more commonly observed in South America and Asia than on other continents. National and AEZ-level statistics demonstrated the joint influence of natural and anthropogenic drivers in controlling CI spatial patterns in most areas of the world. We concluded that the new GCI30 dataset provided improved estimates of global CI in a spatially explicit manner that has not been fully captured by previous studies or products and thus can serve to fill data gaps for achieving SDGs.

References

Beck, Hylke E., Niklaus E. Zimmermann, Tim R. McVicar, Noemi Vergopolan, Alexis Berg, and Eric F. Wood. "Present and future Köppen-Geiger climate classification maps at 1-km resolution." Scientific data 5, no. 1 (2018): 1-12.

Becker, M, and DE Johnson. 2001. "Cropping Intensity Effects on Upland Rice Yield and Sustainability in West Africa." *Nutrient Cycling in Agroecosystems* 59 (2): 107–17.

Bingfang, Wu, Shukri Ahmed, and He Changchui. 2017. "Shared Agronomic Information Community for the Belt and Road Initiative." Bulletin of Chinese Academy of Sciences 32 (Z1): 34–41.

Challinor, Andrew J., Ben Parkes, and Julian Ramirez-Villegas. 2015. "Crop Yield Response to Climate Change Varies with Cropping Intensity." *Global Change Biology* 21 (4): 1679–88. https://doi.org/10.1111/gcb.12808.

Chastain, Robert, Ian Housman, Joshua Goldstein, Mark Finco, and Karis Tenneson. 2019. "Empirical Cross Sensor Comparison of Sentinel-2A and 2B MSI, Landsat-8 OLI, and Landsat-7 ETM+ Top of Atmosphere Spectral Characteristics over the Conterminous United States." *Remote Sensing of Environment* 221: 274–85. https://doi.org/10.1016/j.rse.2018.11.012.

Chiew, Francis; Prosser, Ian; Post, David. On climate variability and climate change and impact on water resources. In: MODSIM 2011; 12-16 December 2011; Perth. Modelling and Simulation Society of Australia and New Zealand; 2011. 3553-3559. http://hdl.handle.net/102.100.100/102035?index=1

Claverie, Martin, Junchang Ju, Jeffrey G. Masek, Jennifer L. Dungan, Eric F. Vermote, Jean-Claude Roger, Sergii V. Skakun, and Christopher Justice. 2018. "The Harmonized Landsat and Sentinel-2 Surface Reflectance Data Set." *Remote Sensing of Environment* 219: 145–61. https://doi.org/10.1016/j.rse.2018.09.002.

Defourny, Pierre, Sophie Bontemps, Nicolas Bellemans, Cosmin Cara, Gérard Dedieu, Eric Guzzonato, Olivier Hagolle, Jordi Inglada, Laurentiu Nicola, and Thierry Rabaute. 2019. "Near Real-Time Agriculture Monitoring

https://doi.org/10.5194/essd-2021-86
Preprint. Discussion started: 30 June 2021
© Author(s) 2021. CC BY 4.0 License.

at National Scale at Parcel Resolution: Performance Assessment of the Sen2-Agri Automated System in Various Cropping Systems around the World." *Remote Sensing of Environment* 221: 551–68.

Didan, K, and A Barreto. 2016. "NASA MEaSUREs Vegetation Index and Phenology (VIP) Vegetation Indices Monthly Global 0.05 Deg CMG." *NASA EOSDIS Land Process. DAAC* 4.

Ding, Mingjun, Qian Chen, Xiangming Xiao, Liangjie Xin, Geli Zhang, and Lanhui Li. 2016. "Variation in Cropping Intensity in Northern China from 1982 to 2012 Based on GIMMS-NDVI Data." *Sustainability* 8 (11): 1123.

Ding, Mingjun, Qihui Guan, Lanhui Li, Huamin Zhang, Chong Liu, and Le Zhang. 2020. "Phenology-Based Rice Paddy Mapping Using Multi-Source Satellite Imagery and a Fusion Algorithm Applied to the Poyang Lake Plain, Southern China." *Remote Sensing* 12 (6): 1022.

Dong, Jinwei, and Xiangming Xiao. 2016. "Evolution of Regional to Global Paddy Rice Mapping Methods: A Review." *ISPRS Journal of Photogrammetry and Remote Sensing* 119: 214–27.

Dong, Jinwei, Xiangming Xiao, Weili Kou, Yuanwei Qin, Geli Zhang, Li Li, Cui Jin, Yuting Zhou, Jie Wang, and Chandrashekhar Biradar. 2015. "Tracking the Dynamics of Paddy Rice Planting Area in 1986–2010 through Time Series Landsat Images and Phenology-Based Algorithms." *Remote Sensing of Environment* 160: 99–113.

Dong, Jinwei, Xiangming Xiao, Michael A Menarguez, Geli Zhang, Yuanwei Qin, David Thau, Chandrashekhar Biradar, and Berrien Moore III. 2016. "Mapping Paddy Rice Planting Area in Northeastern Asia with Landsat 8 Images, Phenology-Based Algorithm and Google Earth Engine." *Remote Sensing of Environment* 185: 142–54.

FAO, IFAD, UNICEF, WFP and WHO. 2020. The State of Food Security and Nutrition in the World 2020. Transforming food systems for affordable healthy diets. Rome, FAO. https://doi.org/10.4060/ca9692en.

FAOSTAT, 2010. Statistical Database of the Food and Agriculture Organization of the United Nations.

Galdo, Virgilio, Gladys Lopez-Acevedo, and Martin Rama. Conflict and the Composition of Economic Activity in Afghanistan. The World Bank, 2020.

Gommes, René, Bingfang Wu, Zhongyuan Li, and Hongwei Zeng. 2016. "Design and Characterization of Spatial Units for Monitoring Global Impacts of Environmental Factors on Major Crops and Food Security." *Food and Energy Security* 5 (1): 40–55.

Gommes, René, Bingfang Wu, Ning Zhang, Xueliang Feng, Hongwei Zeng, Zhongyuan Li, and Bo Chen. 2017. "CropWatch Agroclimatic Indicators (CWAIs) for Weather Impact Assessment on Global Agriculture." *International Journal of Biometeorology* 61 (2): 199–215.

Gong, Peng, Jie Wang, Le Yu, Yongchao Zhao, Yuanyuan Zhao, Lu Liang, Zhenguo Niu, et al. 2013. "Finer Resolution Observation and Monitoring of Global Land Cover: First Mapping Results with Landsat TM and ETM+ Data." *International Journal of Remote Sensing* 34 (7): 2607–54. https://doi.org/10.1080/01431161.2012.748992.

Gorelick, Noel, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, and Rebecca Moore. 2017. "Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone." *Remote Sensing of Environment* 202: 18–27.

Gray, JM, D Sulla-menashe, and MA Friedl. 2019. "User Guide to Collection 6 MODIS Land Cover Dynamics (MCD12Q2) Product.[WWW Document]. URL."

Guo, H. ed., 2019. Big Earth Data in Support of the Sustainable Development Goals: 2019. Science Press.

Guo, Hao, Anming Bao, Tie Liu, Felix Ndayisaba, Liangliang Jiang, Alishir Kurban, and Philippe De Maeyer. "Spatial and temporal characteristics of droughts in Central Asia during 1966–2015." Science of the Total Environment 624 (2018): 1523-1538.

Hao, L., G. Sun, Y. Liu, J. Wan, M. Qin, H. Qian, C. Liu, et al. 2015. "Urbanization Dramatically Altered the Water Balances of a Paddy Field-Dominated Basin in Southern China." *Hydrology and Earth System Sciences* 19 (7): 3319–31. https://doi.org/10.5194/hess-19-3319-2015.

Hao, Peng-yu, Hua-jun TANG, Zhong-xin CHEN, YU Le, and Ming-quan WU. 2019. "High Resolution Crop Intensity Mapping Using Harmonized Landsat-8 and Sentinel-2 Data." *Journal of Integrative Agriculture* 18 (12): 2883–97.

Hinz, R., T. B. Sulser, R. Huefner, D. Mason-D'Croz, S. Dunston, S. Nautiyal, C. Ringler, et al. 2020. "Agricultural Development and Land Use Change in India: A Scenario Analysis of Trade-Offs Between UN Sustainable Development Goals (SDGs)." *Earth's Future* 8 (2): e2019EF001287. https://doi.org/10.1029/2019EF001287.

Iizumi, Toshichika, and Navin Ramankutty. "How do weather and climate influence cropping area and intensity?." Global Food Security 4 (2015): 46-50.

Iqbal, Mohammad Wasim, Somchai Donjadee, Bancha Kwanyuen, and Shi-yin Liu. "Farmers' perceptions of and adaptations to drought in Herat Province, Afghanistan." Journal of Mountain Science 15, no. 8 (2018): 1741-1756.

https://doi.org/10.5194/essd-2021-86
Preprint. Discussion started: 30 June 2021
© Author(s) 2021. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

J. Gray, M. Friedl, S. Frolking, N. Ramankutty, A. Nelson, and M. K. Gumma. 2014. “Mapping Asian Cropping Intensity With MODIS.” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7 (8): 3373–79. https://doi.org/10.1109/JSTARS.2014.2344630.

Jain, Meha, Pinki Mondal, Ruth S DeFries, Christopher Small, and Gillian L Galford. 2013. “Mapping Cropping Intensity of Smallholder Farms: A Comparison of Methods Using Multiple Sensors.” *Remote Sensing of Environment* 134: 210–23.

Jankowski, KathiJo, Christopher Neill, Eric A. Davidson, Marcia N. Macedo, Ciniro Costa, Gillian L. Galford, Leonardo Maracahipes Santos, et al. 2018. “Deep Soils Modify Environmental Consequences of Increased Nitrogen Fertilizer Use in Intensifying Amazon Agriculture.” *Scientific Reports* 8 (1): 13478. https://doi.org/10.1038/s41598-018-31175-1.

King, A.D., Pitman, A.J., Henley, B.J., Ukkola, A.M. and Brown, J.R., 2020. “The role of climate variability in Australian drought”. Nature Climate Change, 10(3), pp.177-179.

Kong, Dongdong, Yongqiang Zhang, Xihui Gu, and Dagang Wang. 2019. “A Robust Method for Reconstructing Global MODIS EVI Time Series on the Google Earth Engine.” *ISPRS Journal of Photogrammetry and Remote Sensing* 155: 13–24. https://doi.org/10.1016/j.isprsjprs.2019.06.014.

Kontgis, Caitlin, Annemarie Schneider, and Mutlu Ozdogan. 2015. “Mapping Rice Paddy Extent and Intensification in the Vietnamese Mekong River Delta with Dense Time Stacks of Landsat Data.” *Remote Sensing of Environment* 169: 255–69. https://doi.org/10.1016/j.rse.2015.08.004.

Köppen, Wladimir, Esther Volken, and Stefan Brönnimann. 2011. “The Thermal Zones of the Earth According to the Duration of Hot, Moderate and Cold Periods and to the Impact of Heat on the Organic World (Translated from: Die Wärmezonen Der Erde, Nach Der Dauer Der Heissen, Gemässigten Und Kalten Zeit Und Nach Der Wirkung Der Wärme Auf Die Organische Welt Betrachtet, Meteorol Z 1884, 1, 215-226).” *Meteorologische Zeitschrift* 20 (3): 351–60.

Kotsuki, S, and K Tanaka. 2015. “SACRA–a Method for the Estimation of Global High-Resolution Crop Calendars from a Satellite-Sensed NDVI.” *Hydrology and Earth System Sciences* 19 (11): 4441–61.

Lal, R. 2002. “Soil Carbon Dynamics in Cropland and Rangeland.” *Environmental Pollution* 116 (3): 353–62. https://doi.org/10.1016/S0269-7491(01)00211-1.

Lewis, Adam, Simon Oliver, Leo Lymburner, Ben Evans, Lesley Wyborn, Norman Mueller, Gregory Raevksi, Jeremy Hooke, Rob Woodcock, and Joshua Sixsmith. 2017. “The Australian Geoscience Data Cube—Foundations and Lessons Learned.” *Remote Sensing of Environment* 202: 276–92.

Li, Le, A. Mark Friedl, Qinchuan Xin, Josh Gray, Yaozhong Pan, and Steve Frolking. 2014. “Mapping Crop Cycles in China Using MODIS-EVI Time Series.” *Remote Sensing* 6 (3). https://doi.org/10.3390/rs6032473.

Liu, Chong, Qi Zhang, Shiqi Tao, Jiaguo Qi, Mingjun Ding, Qihui Guan, Bingfang Wu, et al. 2020. “A New Framework to Map Fine Resolution Cropping Intensity across the Globe: Algorithm, Validation, and Implication.” *Remote Sensing of Environment* 251: 112095. https://doi.org/10.1016/j.rse.2020.112095.

Liu, Luo, Xiangming Xiao, Yuanwei Qin, Jie Wang, Xinliang Xu, Yueming Hu, and Zhi Qiao. 2020. “Mapping Cropping Intensity in China Using Time Series Landsat and Sentinel-2 Images and Google Earth Engine.” *Remote Sensing of Environment* 239: 111624. https://doi.org/10.1016/j.rse.2019.111624.

Lowder, Sarah K., Jakob Skoet, and Terri Raney. 2016. “The Number, Size, and Distribution of Farms, Smallholder Farms, and Family Farms Worldwide.” *World Development* 87: 16–29. https://doi.org/10.1016/j.worlddev.2015.10.041.

M. Amani, A. Ghorbanian, S. A. Ahmadi, M. Kakooei, A. Moghimi, S. M. Mirmazloumi, S. H. A. Moghaddam, et al. 2020. “Google Earth Engine Cloud Computing Platform for Remote Sensing Big Data Applications: A Comprehensive Review.” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13: 5326–50. https://doi.org/10.1109/JSTARS.2020.3021052.

Mason-D'Croz, Daniel, Timothy B. Sulser, Keith Wiebe, Mark W. Rosegrant, Sarah K. Lowder, Alejandro Nin-Pratt, Dirk Willenbockel et al. "Agricultural investments and hunger in Africa modeling potential contributions to SDG2–Zero Hunger." World development 116 (2019): 38-53.

Morton, Douglas C., Ruth S. DeFries, Yosio E. Shimabukuro, Liana O. Anderson, Egidio Arai, Fernando del Bon Espirito-Santo, Ramon Freitas, and Jeff Morisette. 2006. “Cropland Expansion Changes Deforestation Dynamics in the Southern Brazilian Amazon.” *Proceedings of the National Academy of Sciences* 103 (39): 14637. https://doi.org/10.1073/pnas.0606377103.

Pielke, Roger A., Jimmy O. Adegoke, Thomas N. Chase, Curtis H. Marshall, Toshihisa Matsui, and Dev Niyogi. 2007. “A New Paradigm for Assessing the Role of Agriculture in the Climate System and in Climate Change.” *The Contribution of Agriculture to the State of Climate* 142 (2): 234–54. https://doi.org/10.1016/j.agrformet.2006.06.012.

https://doi.org/10.5194/essd-2021-86
Preprint. Discussion started: 30 June 2021
© Author(s) 2021. CC BY 4.0 License.

Qi, J., Katic, P., Mukherji, A., Ruhweza, A., Spierenburg, M., 2020. Rethinking Global Security. In, Our Future on Earth, pp. 75.

Qiu, Shi, Zhe Zhu, and Binbin He. 2019. "Fmask 4.0: Improved Cloud and Cloud Shadow Detection in Landsats 4–8 and Sentinel-2 Imagery." *Remote Sensing of Environment* 231: 111205. https://doi.org/10.1016/j.rse.2019.05.024.

Qiu, Tong, Conghe Song, Yulong Zhang, Hongsheng Liu, and James M. Vose. 2020. "Urbanization and Climate Change Jointly Shift Land Surface Phenology in the Northern Mid-Latitude Large Cities." *Remote Sensing of Environment* 236: 111477. https://doi.org/10.1016/j.rse.2019.111477.

Richardson, Andrew D., Koen Hufkens, Tom Milliman, Donald M. Aubrecht, Min Chen, Josh M. Gray, Miriam R. Johnston, et al. 2018. "Tracking Vegetation Phenology across Diverse North American Biomes Using PhenoCam Imagery." *Scientific Data* 5 (1): 180028. https://doi.org/10.1038/sdata.2018.28.

Richardson, Andrew D., Koen Hufkens, Tom Milliman, and Steve Frolking. 2018. "Intercomparison of Phenological Transition Dates Derived from the PhenoCam Dataset V1.0 and MODIS Satellite Remote Sensing." *Scientific Reports* 8 (1): 5679. https://doi.org/10.1038/s41598-018-23804-6.

Rousta, Iman, Haraldur Olafsson, Md Moniruzzaman, Hao Zhang, Yuei-An Liou, Terence Darlington Mushore, and Amitesh Gupta. "Impacts of drought on vegetation assessed by vegetation indices and meteorological factors in Afghanistan." Remote Sensing 12, no. 15 (2020): 2433.

Seyednasrollah, Bijan, Adam M. Young, Koen Hufkens, Tom Milliman, Mark A. Friedl, Steve Frolking, and Andrew D. Richardson. 2019. "Tracking Vegetation Phenology across Diverse Biomes Using Version 2.0 of the PhenoCam Dataset." *Scientific Data* 6 (1): 222. https://doi.org/10.1038/s41597-019-0229-9.

Sherrod, Lucretia A, Gary A Peterson, Dwayne G Westfall, and Lajpat R Ahuja. 2003. "Cropping Intensity Enhances Soil Organic Carbon and Nitrogen in a No-till Agroecosystem." *Soil Science Society of America Journal* 67 (5): 1533–43.

Siebert, S., Portmann, F.T., Döll, P., 2010. Global patterns of cropland use intensity. *Remote Sens.* 2 (7), 1625–1643.

Singha, Mrinal, Jinwei Dong, Geli Zhang, and Xiangming Xiao. 2019. "High Resolution Paddy Rice Maps in Cloud-Prone Bangladesh and Northeast India Using Sentinel-1 Data." *Scientific Data* 6 (1): 26. https://doi.org/10.1038/s41597-019-0036-3.

Song, Xiao-Peng, Peter V. Potapov, Alexander Krylov, LeeAnn King, Carlos M. Di Bella, Amy Hudson, Ahmad Khan, Bernard Adusei, Stephen V. Stehman, and Matthew C. Hansen. "National-scale soybean mapping and area estimation in the United States using medium resolution satellite imagery and field survey." Remote sensing of environment 190 (2017): 383-395.

Tamiminia, Haifa, Bahram Salehi, Masoud Mahdianpari, Lindi Quackenbush, Sarina Adeli, and Brian Brisco. 2020. "Google Earth Engine for Geo-Big Data Applications: A Meta-Analysis and Systematic Review." *ISPRS Journal of Photogrammetry and Remote Sensing* 164: 152–70. https://doi.org/10.1016/j.isprsjprs.2020.04.001.

Tilman, David, Christian Balzer, Jason Hill, and Belinda L. Befort. 2011. "Global Food Demand and the Sustainable Intensification of Agriculture." *Proceedings of the National Academy of Sciences* 108 (50): 20260. https://doi.org/10.1073/pnas.1116437108.

UN, Transforming our World: The 2030 Agenda for Sustainable Development, 2015.

Waha, Katharina, Jan Philipp Dietrich, Felix T Portmann, Stefan Siebert, Philip K Thornton, Alberte Bondeau, and Mario Herrero. 2020. "Multiple Cropping Systems of the World and the Potential for Increasing Cropping Intensity." *Global Environmental Change* 64: 102131.

Whitcraft, Alyssa K., Inbal Becker-Reshef, Christopher O. Justice, Lauren Gifford, Argyro Kavvada, and Ian Jarvis. 2019. "No Pixel Left behind: Toward Integrating Earth Observations for Agriculture into the United Nations Sustainable Development Goals Framework." *Remote Sensing of Environment* 235: 111470. https://doi.org/10.1016/j.rse.2019.111470.

Wu, Wenbin, Qiangyi Yu, Liangzhi You, Kevin Chen, Huajun Tang, and Jianguo Liu. 2018. "Global Cropping Intensity Gaps: Increasing Food Production without Cropland Expansion." *Land Use Policy* 76: 515–25. https://doi.org/10.1016/j.landusepol.2018.02.032.

Xiao, Xiangming, Stephen Boles, Jiyuan Liu, Dafang Zhuang, Steve Frolking, Changsheng Li, William Salas, and Berrien Moore III. 2005. "Mapping Paddy Rice Agriculture in Southern China Using Multi-Temporal MODIS Images." *Remote Sensing of Environment* 95 (4): 480–92.

Xie, Yanhua, Tyler J. Lark, Jesslyn F. Brown, and Holly K. Gibbs. "Mapping irrigated cropland extent across the conterminous United States at 30 m resolution using a semi-automatic training approach on Google Earth Engine." ISPRS Journal of Photogrammetry and Remote Sensing 155 (2019): 136-149.

Xiong, Jun, S. Prasad Thenkabail, C. James Tilton, K. Murali Gumma, Pardhasaradhi Teluguntla, Adam Oliphant, G. Russell Congalton, Kamini Yadav, and Noel Gorelick. 2017. "Nominal 30-m Cropland Extent Map of

https://doi.org/10.5194/essd-2021-86
Preprint. Discussion started: 30 June 2021
© Author(s) 2021. CC BY 4.0 License.

Continental Africa by Integrating Pixel-Based and Object-Based Algorithms Using Sentinel-2 and Landsat-8 Data on Google Earth Engine." *Remote Sensing* 9 (10). https://doi.org/10.3390/rs9101065.

Yan, Huimin, Fang Liu, Yuanwei Qin, Zhong'en Niu, Russell Doughty, and Xiangming Xiao. 2019. "Tracking the Spatio-Temporal Change of Cropping Intensity in China during 2000–2015." *Environmental Research Letters* 14 (3): 035008. https://doi.org/10.1088/1748-9326/aaf9c7.

Yan, Huimin, Xiangming Xiao, Heqing Huang, Jiyuan Liu, Jingqing Chen, and Xuehong Bai. 2014. "Multiple Cropping Intensity in China Derived from Agro-Meteorological Observations and MODIS Data." *Chinese Geographical Science* 24 (2): 205–19. https://doi.org/10.1007/s11769-013-0637-2.

Zeng, Zhenzhong, Lyndon Estes, Alan D. Ziegler, Anping Chen, Timothy Searchinger, Fangyuan Hua, Kaiyu Guan, Attachai Jintrawet, and Eric F. Wood. 2018. "Highland Cropland Expansion and Forest Loss in Southeast Asia in the Twenty-First Century." *Nature Geoscience* 11 (8): 556–62. https://doi.org/10.1038/s41561-018-0166-9.

Zhang, Miao; Wu, Bingfang; Zeng, Hongwei; He, Guojin; Liu, Chong; Nabil, Mohsen; Tian, Fuyou; Bofana, José; Wang, Zhengdong; Yan, Nana, 2020, "GCI30: Global Cropping Intensity at 30m resolution", https://doi.org/10.7910/DVN/86M4PO, Harvard Dataverse, V1.

Zhang, Miao, Wu, Bingfang, Meng, Jihua, Dong, Taifeng and You, Xingzhi. 2014a. "Fallow land mapping for better crop monitoring in Huang-Huai-Hai Plain using HJ-1 CCD data." IOP Conference Series: Earth and Environmental Science. Vol. 17. No. 1. IOP Publishing.

Zhang, Miao, Wu, Bingfang, Yu, Mingzhao, Zou, Wentao and Zheng, Yang. 2014b. "Crop condition assessment with adjusted NDVI using the uncropped arable land ratio." *Remote sensing* 6: 5774-5794.

Zhang, Xiaoyang. 2015. "Reconstruction of a Complete Global Time Series of Daily Vegetation Index Trajectory from Long-Term AVHRR Data." *Remote Sensing of Environment* 156: 457–72. https://doi.org/10.1016/j.rse.2014.10.012.

Zhu, Zhe, and Curtis E. Woodcock. 2012. "Object-Based Cloud and Cloud Shadow Detection in Landsat Imagery." *Remote Sensing of Environment* 118: 83–94. https://doi.org/10.1016/j.rse.2011.10.028.