https://doi.org/10.5194/essd-2023-536
Preprint. Discussion started: 6 February 2024
© Author(s) 2024. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

GMIE-100: a global maximum irrigation extent and irrigation type dataset derived through irrigation performance during drought stress and machine learning method

Fuyou Tian[1], Bingfang Wu[1,2,*], Hongwei Zeng[1,2], Miao Zhang[1], Weiwei Zhu[1], Nana Yan[1], Yuming Lu[1,2]
Yifan Li [3,1]

[1]State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China;
[2]University of Chinese Academy of Sciences, Beijing 100049, China;
[3]School of Computer Science, China University of Geosciences, Wuhan 430078, China

Correspondence to: Bingfang Wu (wubf@aircas.ac.cn)

Abstract. Irrigation stands as the primary sector of human water consumption and plays a pivotal role in enhancing crop yields and mitigating drought effects. The precise distribution of irrigation is crucial for effective water resource management and the assessment of food security. However, the existing global irrigated cropland map is characterized by a coarse resolution, typically around 10 kilometers, and is often not regularly updated. In our study, we present a robust methodology that leverages irrigation performance during drought stress as an indicator of crop productivity and water consumption to identify global irrigated cropland. Within each irrigation mapping zone (IMZ), we identified the dry months occurring during the growing season from 2017 to 2019 or the driest months from 2010 to 2019. To delineate irrigated cropland, we utilized collected samples to calculate normalized difference vegetation index (NDVI) thresholds for the dry months of 2017 to 2019 and the NDVI deviation from the ten-year average for the driest month. By combining the results with the higher accuracy between these two methods, we generated the Global Maximum Irrigation Extent dataset at 100-meter resolution (GMIE-100), achieving an overall accuracy of 83.6%. GMIE-100 reveals that the maximum extent of irrigated cropland encompasses 403.17 million hectares, accounting for 23.4% of the global cropland. Concentrated in fertile plains and regions adjacent to major rivers, the largest irrigated cropland area is found in India, followed by China, the United States, and Pakistan, ranking 2nd to 4th, respectively. Importantly, the spatial resolution of GMIE-100, at 100 meters, surpasses that of the dominant irrigation map, offering more detailed information essential for supporting estimates of agricultural water use and regional food security assessments. Furthermore, with the help of the DL method, the global central pivot irrigation system (CPIS) was identified using Pivot-Net. We found that there are 11.5 million hectares of CPIS, accounting for about 2.9% of total irrigated cropland. In Namibia, the US, Saudi Arabia, South Africa, Canada, and Zambia, the CPIS proportion was larger than 10%. To our best knowledge, this is the first effort to identify irrigation methods globally. The GMIE-100 dataset containing both or irrigated extent and CPIS distribution is accessible on Harvard Dataverse at: https://doi.org/10.7910/DVN/HKBAQQ (Tian et al., 2023a).

https://doi.org/10.5194/essd-2023-536
Preprint. Discussion started: 6 February 2024
© Author(s) 2024. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

1. Introduction

Irrigation plays a pivotal role in mitigating the impacts of drought events (Wang et al., 2021; Wu et al., 2022). With the intensification of climate change leading to more frequent droughts and heatwaves, irrigation emerges as an effective strategy to counter these extreme events and bolster the resilience of agricultural systems (Mcdermid et al., 2023). However, it's crucial to acknowledge that irrigation represents a significant human intervention in the global water cycle, accounting for 67% of global freshwater withdrawal and 87% of total water consumption (Wu et al., 2022). Therefore, the accurate information pertaining to irrigation is of paramount importance, serving both crop monitoring and water resource management purposes (Wu et al., 2023b). However, the highest available resolution for existing irrigation maps remains confined to a range of 500 meters to 10 kilometers. This limitation falls far short of adequately supporting crop condition monitoring and sustainable water resource management at the sub-basin level (Zhang et al., 2022b; Xie and Lark, 2021).

Traditionally, two primary methods have been employed for generating gridded irrigation maps. The first method involves the allocation of statistical data using specific indicators such as land cover area, peak NDVI values, and irrigation potential indices (Zhu et al., 2014; Pervez and Brown, 2010; Zajac et al., 2022). Notably, the Food and Agriculture Organization (FAO) utilized this approach to produce the Global Map of Irrigation Area (FAO-GMIA) from 1995 to 2005 at a 10-kilometer resolution, a renowned irrigation map widely applied in global water resource management (Siebert et al., 2015). At the national scale, several irrigation maps for China have been proposed with resolutions ranging from 500 to 1000 meters, primarily utilizing data from the Chinese Statistical Yearbook (Zhu et al., 2014; Zhang et al., 2022a). For the United States, Pervez and Brown, 2010, developed an Irrigated Agriculture Dataset for the US (MIrAD-US) with a resolution of 250 meters. Zajac et al., 2022, produced the European Irrigation Map for the year 2010 (EIM2010), albeit with a coarser 10-kilometer×10-kilometer resolution. It is important to note that the accuracy of irrigated cropland maps generated through these methods heavily relies on the representativeness of the spatial allocation indicators and the precision of the statistical data. The indicators used to allocate irrigation area to each grid often fail to capture the precise distribution of irrigated cropland, especially in humid regions (Pervez and Brown, 2010). Consequently, achieving higher-resolution irrigation maps using this approach can be challenging. Furthermore, due to variations in terrain types and irrigation techniques, census data may underestimate the actual irrigation area (Zhang et al., 2022b). Complicating matters further, data from different departments may exhibit discrepancies owing to differing statistical criteria. For example, in 2010, the reported irrigation area in California differed by more than 10% between the US Geological Survey and the state's Department of Water Resources (Meier et al., 2018).

Scholars have sought to independently derive irrigated cropland through spectral signatures of irrigated croplands(Thenkabail et al., 2009; Salmon et al., 2015). It has been verified that the peak values in the time-series vegetation index can serve as indicators of crop water stress, biomass, and chlorophyll content. Given that irrigated crops typically exhibit reduced water stress and elevated chlorophyll content, disparities in peak vegetation index values can be harnessed to differentiate between irrigated and rain-fed croplands. Commonly employed vegetation indices for this approach encompass the Normalized Difference Vegetation Index (NDVI), Greenness Index (GI), Land Surface Water Index (LSWI), Chlorophyll

https://doi.org/10.5194/essd-2023-536
Preprint. Discussion started: 6 February 2024
© Author(s) 2024. CC BY 4.0 License.

Vegetation Index (GCVI), Enhanced Vegetation Index (EVI), and others (Shahriar Pervez et al., 2014; Lu et al., 2021; Chen et al., 2018; Xiang et al., 2019; Dela Torre et al., 2021). The discrimination between irrigated and rain-fed croplands is typically accomplished through thresholding or decision tree classification, relying on the selected vegetation indices. Nevertheless, it's important to note that vegetation indices may not entirely capture crop water stress, leading to subtle differences in peak vegetation indices and complicating the mapping of large-scale irrigated farmland.

To enhance the distinction of irrigated cropland, supervised classification models incorporate climate variables and environmental factors such as precipitation, temperature, surface temperature, and terrain (Salmon et al., 2015). For instance, Thenkabail et al. (2009) utilized a set of factors including AVHRR vegetation index time series, precipitation data, elevation information, and vegetation cover maps as inputs to a decision tree classifier, resulting in the creation of the first global irrigation area map (IWMI-GIAM) at a 10-kilometer resolution based on remote sensing data. Salmon et al. (2015) employed MODIS vegetation indices and 19 climate variables to produce the Global Rainfed and Irrigated Cropland map (GRIPC-500) for the year 2005 at a resolution of 500 meters.

In recent years, the mapping of irrigated croplands at the national and regional scales has seen significant advancements due to the availability of extensive meteorological and remote sensing data stored in Google Earth Engine (Zhang et al., 2022b; Deines et al., 2019; Xie et al., 2019; Xie and Lark, 2021). Xie et al. (2021) developed a random forest model incorporating a wide array of variables, including environmental factors (precipitation, Palmer drought severity index, soil moisture, aridity index, land surface and air temperature), vegetation indices (NDVI, NDWI, GCVI, WGI, and AGI), and ground irrigation samples. This model achieved an impressive 30-meter resolution irrigation dataset for the United States (LANID). Subsequently, Zhang et al. (2022a) applied this methodology to generate an irrigated cropland map for China from 2000 to 2019, offering a resolution of 500 meters (IrriMap_CN). In the same year, Zhang et al. (2022c) enhanced the resolution of the irrigation cropland distribution map for China to 250 meters. However, this method heavily relies on samples, and the spatial representativeness of these irrigation and rainfed samples directly influences the accuracy of the results (Zhang et al., 2022b). Collecting ground sample points is a labor-intensive and time-consuming process, and ensuring their spatial representativeness across larger areas, including at a global scale, poses considerable challenges (Zhang et al., 2022a; Zhang et al., 2022d).

Despite the existence of various irrigation maps at global and national scales, many of them suffer from either very low spatial resolution or outdated information, as outlined in Table 1 (Dari et al., 2023). While some high-resolution irrigation maps are annually updated, they are typically applicable only at the national level (Zhang et al., 2022b; Xie et al., 2021). In essence, the challenge of generating a higher-resolution and up-to-date global irrigated cropland map using supervised methods persists.

An additional significant issue is the phenomenon of "mixed pixels" in MODIS data, which is particularly pronounced in regions with fragmented croplands, such as Southern China and farmlands in Africa, where agricultural fields are often smaller than one MODIS pixel (0.25 hectares) (Zhang et al., 2022c). Consequently, there is an urgent need for a global irrigation map with higher resolution to support both water resource management and food security assessments.

https://doi.org/10.5194/essd-2023-536
Preprint. Discussion started: 6 February 2024
© Author(s) 2024. CC BY 4.0 License.

Taking inspiration from the fundamental purpose of irrigation, which is to alleviate the impact of drought, we have introduced the Global Maximum Irrigated Extent with 100-meter resolution (GMIE-100) dataset. This dataset leverages irrigation performance during periods of drought stress. When drought conditions prevail, disparities in crop conditions, as indicated by the peak values of NDVI, become more pronounced between irrigated and rainfed farmlands. This amplification enables precise identification of irrigated farmland across most regions, while also reducing the quantity of required training samples (Wu et al., 2023a).

Furthermore, considerable variations in irrigation efficiency are apparent among different irrigation types, with central pivot irrigation systems (CPIS) emerging as the predominant global sprinkler irrigation method, demonstrating an efficiency exceeding 80%. In contrast, gravity flowing irrigation methods, while widespread, exhibit a comparatively lower efficiency, approximately 60%(Waller and Yitayew, 2016). Despite the important role of irrigation in agriculture, there is a few research dedicated to the remote sensing identification of various irrigation types, indicating a notable gap in scientific exploration. Noteworthy is the unique circular configuration of CPIS, facilitating their visual interpretation from satellite imagery, presenting an avenue for enhanced monitoring and analysis through remote sensing technologies. The advent of deep learning (DL) has opened avenues for the classification of irrigation method based on distinctive spatial patterns, such as CPIS. This study used Pivot-Net, a shape attention neural network designed for CPIS identification in satellite imagery and generate global CPIS dataset (GCPIS) to estimated irrigation method proportion for CPIS.

Table 1 List of some of existed irrigation map for global or national scale.

Dataset	Coverage	Spatial Resolution	Time	Method summary	Reference
Global Irrigated Area Map (IWMI-GIAM)	Global	10km	2000,	Using decision tree classifier with vegetation index &environmental data as input	(Thenkabail et al., 2009)
Global Map of Irrigation Area (FAO-GMIA)	Global	10km	1995/2000/2005	allocate census data using landcover area	(Siebert et al., 2015)
Global Rainfed, Irrigated and Paddy Croplands (GRIPC-500)	Global	500 m	Single map 2005	Climate variables and environmental factors was included in a Decision tree classifier	(Salmon et al., 2015)
Global Food-Support Analysis Data (GFSAD)	Global	1km	2010	This was created using multiple input data including satellite data, climatic and census data.	(Thenkabail et al., 2012)
Landsat-derived Global Rainfed and Irrigated-Cropland Product at nominal 30m of the World (USGS-LGRIP30)	Global	30m	2015	Landsat-derived global rainfed and irrigated cropland product within cropland extent	(Teluguntla et al., 2023)
Landsat-based Irrigation Dataset (LANID)	US	30m	1997-2017	random forest model based on environmental variables & vegetation indices	(Xie et al., 2021; Xie et al., 2019; Xie and Lark, 2021)

https://doi.org/10.5194/essd-2023-536
Preprint. Discussion started: 6 February 2024
© Author(s) 2024. CC BY 4.0 License.

Annual irrigation maps across China (IrriMap_CN)	China	500m	2000-2019	Random forest with remote sensing index and environmental index	(Zhang et al., 2022b)
Remotely sensed high resolution irrigated area in India	India	250m	2000-2015	NDVI series in decision tree method	(Ambika et al., 2016)

2. Material and method

Taking inspiration from the fundamental purpose of irrigation, our aim is to identify periods of drought stress to accentuate the disparities in crop conditions between irrigated and rainfed croplands. We initiated this process by utilizing the sixty-five Monitoring and Reporting Units (MRUs) established by CropWatch (Wu et al., 2015; Gommes et al., 2016). These MRUs, which take into account factors like crop types, agricultural potential, and environmental conditions, served as the basis for further division of global cropland into 110 irrigation mapping zones (IMZs). The first-level 65 agroecological zones offer a fundamental global overview. To address limitations in depicting water stress and irrigation within zones, a more detailed classification was introduced, creating second-level agroecological zones based on arid indices, water availability, soil types, and landforms. Ultimately, we utilized the 110 IMZs as the foundational units for determining the specific timing of drought stress, as illustrated in Figure 1. This comprehensive approach allows us to capture and amplify the distinctions in crop conditions between irrigated and rainfed croplands.

https://doi.org/10.5194/essd-2023-536
Preprint. Discussion started: 6 February 2024
© Author(s) 2024. CC BY 4.0 License.

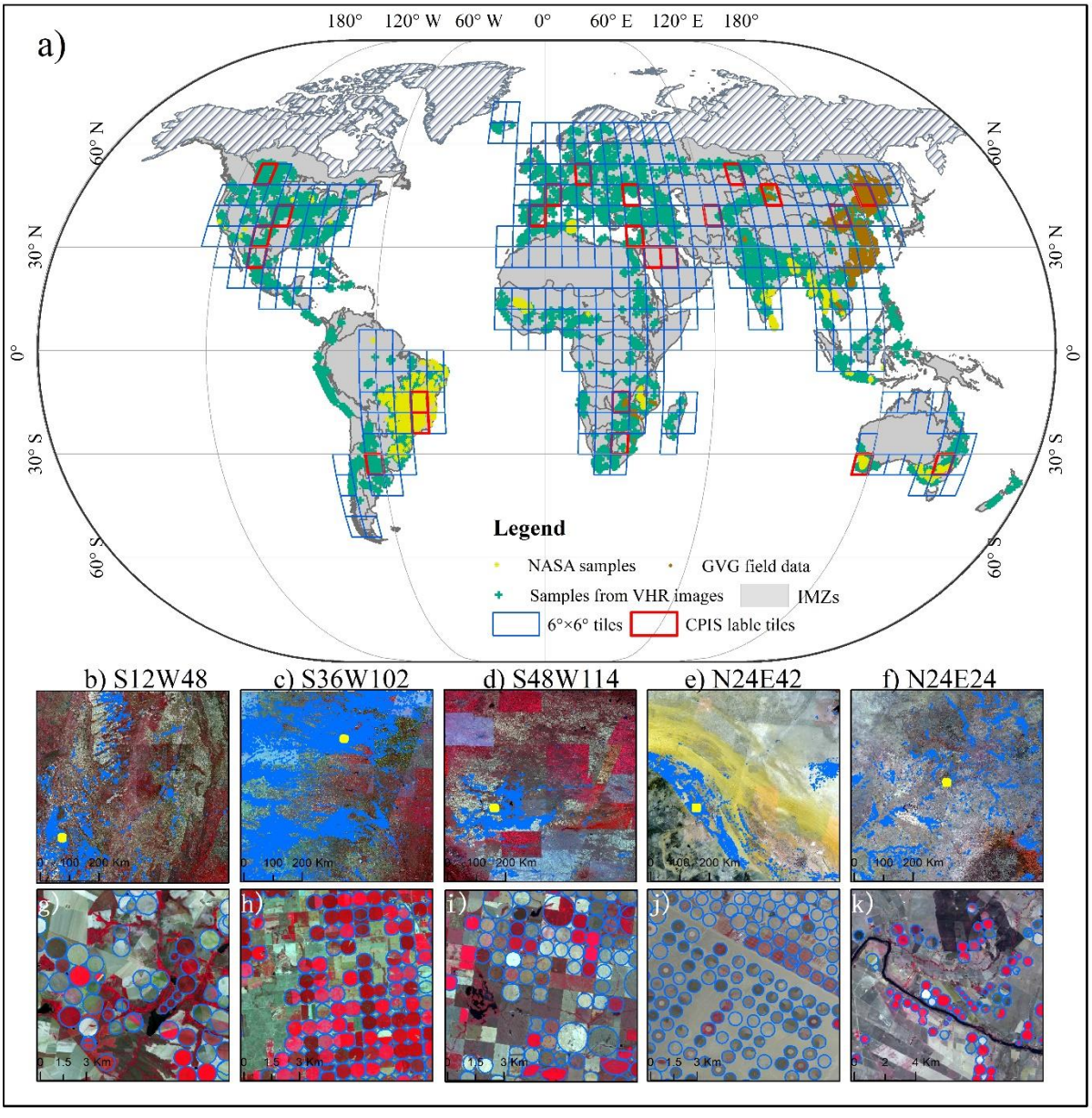


Figure 1 Samples of irrigated, rainfed and CPIS form multi-sources and mapping unit for irrigation mapping and CPIS identification. a) is the distribution of irrigation mapping zones and irrigated & rainfed cropland sample samples. b-f) show 5 of annotated tiles for CPIS label and images. The name of b-f) was the coordinate of lower left corner point of each tile. g-k are detail map of CPIS label. Its location is shown in b-f) as yellow rectangle. The background images of b-k are Landsat-8 data.

The general framework for detecting drought stress and evaluating crop conditions in irrigated and rainfed cropland is illustrated in Figure 2. Within each irrigation mapping zone (IMZ), the dry months occurring during the growing season and

https://doi.org/10.5194/essd-2023-536
Preprint. Discussion started: 6 February 2024
© Author(s) 2024. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

the driest months in the driest year was identified. To distinguish irrigated cropland, multi-sources samples were used to determine NDVI thresholds for the dry months and the NDVI deviation from the ten-year average for the driest month. These metrics were used as proxies for assessing the disparities in crop conditions between irrigated and rainfed croplands. The final result was determined based on the method that yielded the highest level of accuracy.

Then with support of DL model, a CPIS identification model focus on its circle shape, was trained and applied to the entire word and generate Global CPIS distribution data. The extent of CPIS was recognized as irrigation extent to update the global irrigation extent. Final, we estimated irrigation type proportion of CPIS within irrigated cropland.

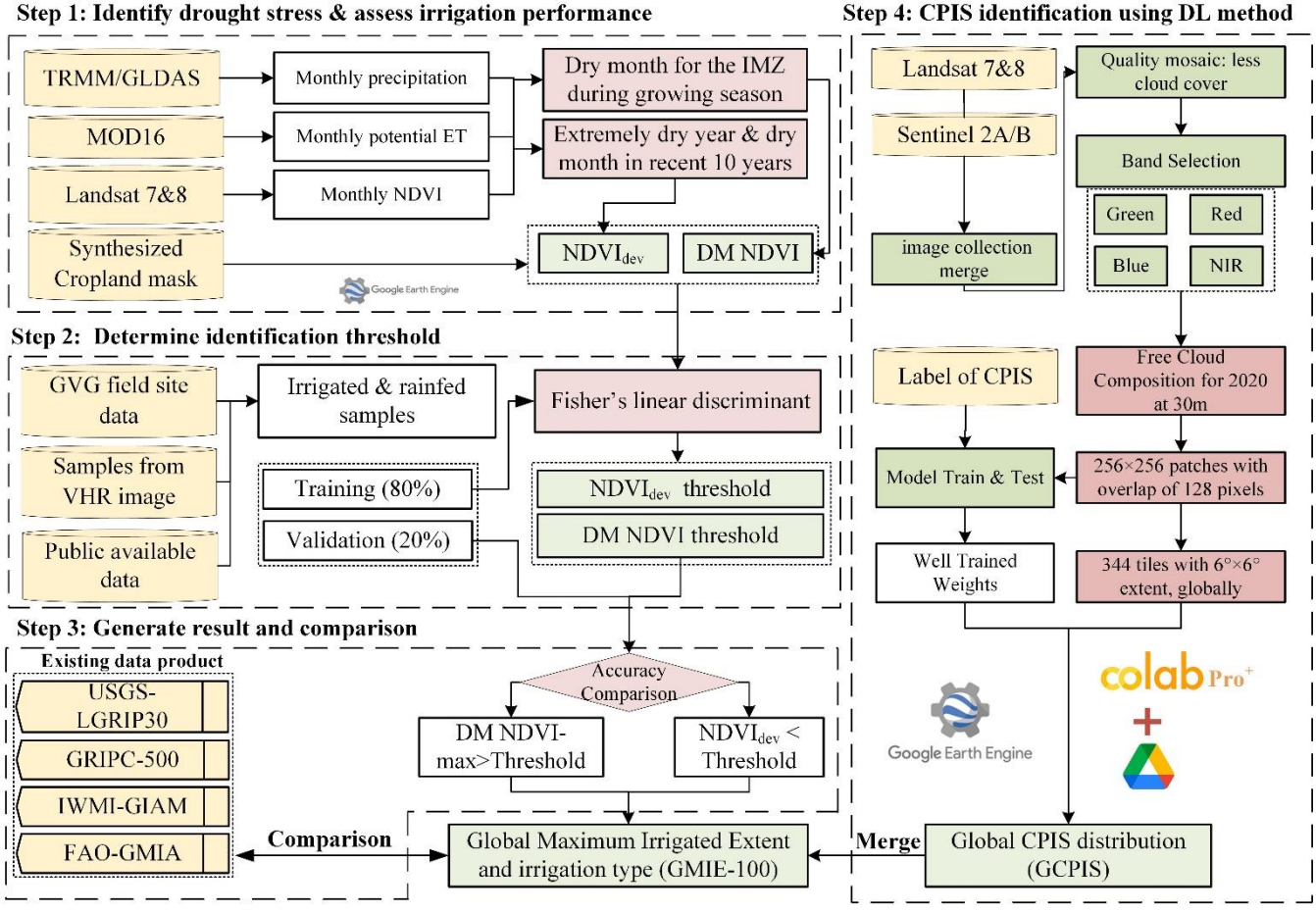


Figure 2 Flow chart of GMIE-100 with typical irrigation type of CPIS

https://doi.org/10.5194/essd-2023-536
Preprint. Discussion started: 6 February 2024
© Author(s) 2024. CC BY 4.0 License.


2.1 Input data

In this research, the distribution of rainfall on a global scale plays a pivotal role in determining the necessity for crop irrigation. The study focused on a ten-year period spanning from 2010 to 2019 and aimed to identify the driest year within this timeframe. To accomplish this, two distinct sources of precipitation data were utilized: a) Tropical Rainfall Measuring Mission (TRMM) Data: The TRMM collection TRMM/3B43V7, which provides monthly precipitation estimates, was employed for the geographical area ranging from 50°S to 50°N. This data source offers insights into precipitation patterns within this specific region. b) Global Land Data Assimilation System (GLDAS) Data: For areas outside the 50°S to 50°N range, precipitation data from the Global Land Data Assimilation System (GLDAS) were utilized. GLDAS provides information on precipitation in regions beyond the tropical band.

Additionally, the research relied on an evapotranspiration product known as MOD16A2.006, as introduced by Mu et al. in 2013. This product serves the purpose of determining the water surplus during the driest months within each IMZ. The MOD16A2.006 dataset is characterized by an 8-day composite timeframe and a pixel resolution of 500 meters. It is derived from the Penman–Monteith equation and incorporates daily meteorological reanalysis data as well as remotely sensed data products from MODIS. This comprehensive dataset aids in the assessment of water availability and evapotranspiration dynamics during critical dry periods.

NDVI data at 30-meter spatial resolution from Landsat sensors Thematic Mapper (TM), Enhanced Thematic Mapper Plus (ETM+), and Thermal Infrared Sensor (OLI-TIRS) onboard Landsat-5, Landsat-7, and Landsat-8 were utilized in Google Earth Engine (GEE) (Gorelick et al., 2017) to differentiate irrigated and non-irrigated areas across various IMZs during a specific time period.

2.2 Sample Data

Acquiring irrigation samples on a global scale presents a huge challenge, characterized by significant labor and cost requirements, primarily attributable to the extensive geographic scope. When conducting a global evaluation of cropland classification into irrigated and non-irrigated categories, a notable limitation emerges - the absence of a single dataset capable of furnishing adequately representative samples. The scarcity of irrigation datasets tailored to specific crop types hinders precise differentiation between irrigated and non-irrigated croplands. In the majority of countries, excepting India, China, and Pakistan, the area allocated to irrigated croplands constitutes a relatively minor fraction of the total cultivated expanse. This paucity of representation poses challenges in amassing a substantial sample size suitable for classification purposes. Contemporary irrigation maps are often afflicted by coarse spatial resolutions, curtailing their efficacy in generating precise samples for classification endeavors. To surmount these impediments and establish a robust sample dataset, an integrative methodology was employed. This approach entailed the fusion of data originating from three independent sources, facilitating a more comprehensive and accurate appraisal of global irrigated and non-irrigated croplands.

https://doi.org/10.5194/essd-2023-536
Preprint. Discussion started: 6 February 2024
© Author(s) 2024. CC BY 4.0 License.


The first source involves field data points collected using the GVG (GPS, Video, GIS) application in China (surveyed from 2010 to 2019), Cambodia (in 2019), Ethiopia (from 2018 to 2019), Zambia (from 2016 to 2019), Mozambique (from 2016 to 2019), and Zimbabwe (from 2016 to 2019). This application serves as a comprehensive field data collection system that integrates GPS for precise positioning, a video for capturing geo-tagged photographs, and a GIS system for managing geographic information. By conducting observations of irrigation infrastructure, including irrigation canals, reservoirs, lakes, rivers, and irrigation wells, and through interactions with farmers, we were able to determine the irrigation types in the fields. The collected dataset comprises a total of 78,338 sample points, including 36,809 rainfed samples and 41,529 irrigation samples, with the majority of these points located in China, totalling 72,224 points.

The second data source consists of validation points collected as part of the Global Food Security Analysis Data 30 (GFSAD30) project, which is made available to the public through the website https://croplands.org/app/data/search. This project is a collaborative effort involving the United States Geological Survey (USGS), various universities, research institutions, and companies like Google. These sample points were collected or derived as part of the project's objective to support global food security analysis at a 30-meter spatial resolution. Some of the sample points were gathered through field surveys conducted using mobile applications. Others were derived by interpreting remote sensing imagery, such as MODIS and Landsat TM data, crop-specific thematic maps, foundational geographic data (e.g., road networks), and other geospatial information (e.g., elevation data layers). The dataset encompasses a total of 17,076 sample points, comprising 3,000 rainfed points and 14,076 irrigated points. The majority of these points are located in Brazil (13,368), Australia (2,192), Thailand (393), and Tunisia (389).

The third supplementary data source involved the acquisition of samples through visual interpretation of Very High-Resolution (VHR) images available in Google Earth Engine (GEE). Irrigated points were selected based on identifiable irrigation infrastructure, including: 1) Central pivot irrigation system, which are easy to identify due to their shapes.2) Clearly Visible Irrigation Systems: Irrigation systems that were clearly visible on VHR images. 3) Rain-Deficient Cultivated Areas: Areas classified as cropland with insufficient rainfall but exhibiting NDVI values indicating vegetation presence and annual growth rings. 4)High Vegetation Signals During Dry Seasons: Areas displaying elevated vegetation signals during dry seasons. The United Nations Food and Agriculture Organization's Global Map of Irrigation Areas (FAO GMIA) (Siebert et al., 2013) and the World Heritage Irrigation Structures (WHIS) list (https://www.icid.org/icid_his1.php#HIS) were used as reference sources. FAO GMIA's Irrigation Areas of Interest (AEI) and WHIS listings were consulted to identify irrigation areas. Rainfed irrigation points were selected based on FAO GMIA's criteria. If a region lacked any irrigation infrastructure and the AEI value from FAO GMIA was zero, the area was classified as a rainfed irrigation sample.

Figure 1 illustrates a total of 115,379 sample points. 80% of this dataset, or 92,303 points (comprising 37,650 rainfed and 54,653 irrigated points) was employed for training or calibrating the threshold. The remaining 20%, or 23,076 points (comprising 10,892 rainfed cropland points and 12,184 irrigated points), was used for result validation.

https://doi.org/10.5194/essd-2023-536
Preprint. Discussion started: 6 February 2024
© Author(s) 2024. CC BY 4.0 License.

2.3 Land cover and cropland datasets

In this research, we delineated irrigated croplands within the extent of cropland. The definition of cropland was same as the Joint Experiment of Crop Assessment and Monitoring (JECAM) network for Group on Earth Observations Global Agricultural Monitoring Initiative (GEOGALM), which define the land used for seasonal crops (sowed/planted and harvested at least once within the 12 months) such as cereals, root and tuber crops, oil crops, as well as economically significant crops like sugar, vegetables, and cotton(Waldner et al., 2016). Additionally, land occupied by greenhouses was considered as cropland. To achieve comprehensive global cropland coverage, the synthesized data was from 16 recent national and regional datasets spanning the years 2015-2019, which were supplemented by two global satellite-derived land cover datasets, as list in Table 2. In this study, all land cover classes that met the cropland definition were consolidated into a single category labeled as "cropland." On the other hand, various non-vegetation land cover classes (e.g., urban or water) and vegetated classes (e.g., forest or grasslands), including agricultural categories (e.g., permanent crops, cultivated rangeland, and grassland), were amalgamated into one class as "non-cropland." The cropland mask at 30-meter resolution could be obtained from International Research Center of Big Data for Sustainable Development Goals via https://data.casearth.cn/thematic/cbas_2022/158. The overall accuracy of this cropland was 89.4%. Meanwhile, this mask is also employed to map global crop intensity (Zhang et al., 2021).

2.4 irrigation mapping method

2.4.1 Identifying the dry months and dry years

The cumulative yearly and monthly rainfall (P) was calculated from the TRMM dataset for all IMZs in Google Earth Engine (GEE) spanning the years 2010-2019. Simultaneously, monthly potential evapotranspiration (PET) for the same time was derived from the MOD16A2.006 product in GEE. The monthly water surplus (P - PET) was established as the difference between monthly rainfall and monthly potential evapotranspiration.

Within the growing seasons of 2017-2019, we identified the dry months by pinpointing the lowest differences between monthly rainfall and PET. Additionally, we determined the driest year from the period 2010-2019 based on the lowest annual rainfall, and the corresponding driest month was identified as the month with the lowest P-PET value during the driest year within the growing season.

2.4.2 Identifying thresholds of NDVI and NDVI deviation

Irrigated cropland is characterized as cropland subjected to human interventions and equipped with irrigation infrastructure, including systems like canals and central pivot systems (Wu et al., 2023a). The specific threshold for distinguishing between irrigated and non-irrigated cropland differs from IMZs. The threshold for each IMZ was determined by training samples, through visual interpretation on very high-resolution image from Google Earth.

https://doi.org/10.5194/essd-2023-536
Preprint. Discussion started: 6 February 2024
© Author(s) 2024. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

For each IMZ, the maximum NDVI was calculated within cropland extent during the dry month (NDVImax-DM) using Landsat-8 images in Google Earth Engine to detect vegetation signals. In regions where regular irrigation is necessary, irrigated cropland could be mapped annually. However, to avoid missing fallow land based on the result of single year, the irrigated lands was considered as the collection of irrigated croplands identified through the NDVI threshold over a three-year period from 2017 to 2019.

For regions with ample rainfall, drought stress may not be a concern. Hence, satellite data spanning the 2010-2019 period was utilized to identify the crop condition during extremely drought event within ten years. The NDVI deviation ($NDVI_{dev}$) was calculated for the driest month of driest year across 2010-2019 at cropland pixels, following these formulas:

$$NDVI_{dev} = \frac{\mathrm{NDVI}_{\mathrm{max-DriestM}} - 10\mathrm{YNDVI}_{DM}}{10\mathrm{YNDVI}_{DM}} \quad (1)$$

where $\mathrm{NDVI}_{\text{max-DriestM}}$ is the maximum NDVI value in the driest month over 10 years, and $10\mathrm{YNDVI}_{DM}$ is the monthly average NDVI in the same month.

For each IMZ, the midpoint value for a cropland pixel was determined from the irrigated and non-irrigated training points using Fisher's linear discriminant (Duda et al., 2012):

$$\mathrm{Nmidpoint} = \frac{N_{irrigated} + N_{nonirrigated}}{2} \quad (2)$$

where $N_{irrgated}$ and $N_{non-irrgated}$ represent the mean value of NDVI or $\mathrm{NDVI}_{\mathrm{dev}}$ in irrigated and non-irrigated points, respectively.

For each IMZ, the Nmidpoint, which serves as the threshold value, of NDVI value and $\mathrm{NDVI}_{\mathrm{dev}}$ was computed using irrigated and rainfed samples. Subsequently, pixels exhibiting NDVI exceeding their specific threshold values for dry months or $\mathrm{NDVI}_{\mathrm{dev}}$ less than the threshold during the driest month of driest year were designated as irrigated, otherwise pixels falling were classified as non-irrigated.

The final threshold value was determined by selecting the NDVI or $\mathrm{NDVI}_{\mathrm{dev}}$ threshold that yielded the highest overall accuracy in distinguishing irrigated cropland in validation samples. Subsequently, the chosen threshold value for either NDVI or $\mathrm{NDVI}_{\mathrm{dev}}$ of the IMZ was applied to the respective pixels, which were accepted as the final results. If maximum NDVI value in dry month archived higher accuracy for identifying irrigated cropland, this region usually needs regular irrigation, labeled as region needs irrigation regular (RIR). Otherwise, the region only needs irrigation occasionally for some years, labeled as RIO.

https://doi.org/10.5194/essd-2023-536
Preprint. Discussion started: 6 February 2024
© Author(s) 2024. CC BY 4.0 License.

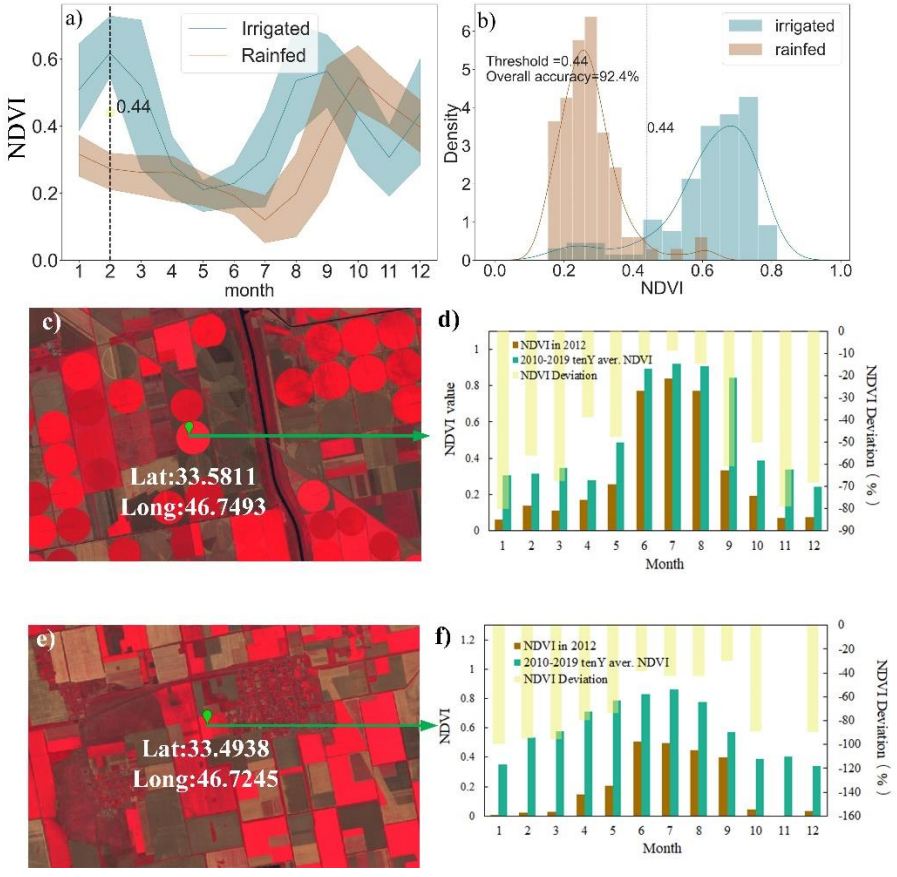


Figure 3. NDVI profile in 2017 (a) and NDVI histogram in February 2017 (b) (Pakistan IMZ C48 as an example); Example of monthly NDVI in extremely dry year (2012), ten-year average NDVI, and $NDVI_{dev}$ for typical central pivot irrigated cropland (c, d) and rainfed cropland (e, f) in south of Ukraine (IMZ C58) as an example. The background images of c and e are Landsat-8 data. Credit of c and e is @U.S. Geological Survey

Taking IMZ C48, primarily situated in Pakistan, as an exemple, Figure 4a illustrates the monthly NDVI profile for the year 2017 within Pakistan (IMZ C48, South Asia Punjab to Gujarat). It is evident that the discrepancy in NDVI values between irrigated and non-irrigated crops remains marginal for the majority of the months in 2017. However, in February 2017, during a period of drought stress characterized by a meagre precipitation of 4.4 mm or a precipitation-to-evapotranspiration ratio of 0.02, the disparity in NDVI values becomes notably more pronounced and distinguishable. Consequently, the optimal NDVI threshold of 0.44 was ascertained as the most suitable for discriminating irrigated from non-irrigated regions, as depicted in Figure 4b.

For the region need irrigation occasionally (RIO), IMZ C58 was chosen as an example. Figure 3d &f shows the monthly NDVI profiled for extremely drought year of 2012, ten-year average NDVI value, and NDVI deviation of extremely drought year to ten-year average. The comparison reveals that rainfed cropland exhibits more substantial fluctuations in NDVI than irrigated cropland. Consequently, the $NDVI_{dev}$ (NDVI deviation) during severe drought or extremely arid conditions was

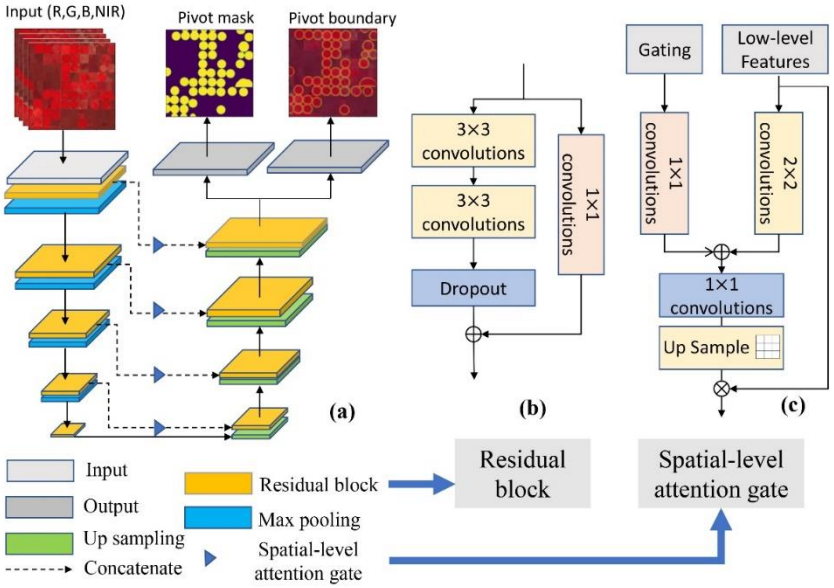
https://doi.org/10.5194/essd-2023-536
Preprint. Discussion started: 6 February 2024
© Author(s) 2024. CC BY 4.0 License.

employed to differentiate irrigated cropland from other categories. The $NDVI_{dev}$ midpoint was established as 0.12 following equation (2).

By amalgamating these two categories of irrigated cropland, we have introduced a comprehensive global irrigation map. For further detail information, you can refer to (Wu et al., 2023a). Originally, the Global Maximum Irrigated Extent (GMIE) dataset was established at a 30-meter resolution, featuring a binary classification into irrigated and rainfed cropland. This resolution was determined by the availability of cropland masks and NDVI data, both of which are at a 30-meter scale. But the irrigation extent maybe varied due to crop rotation and fallow cropland, which can be distinctly observed at 30-meter resolution and impact the extent of irrigated cropland. We calculated the irrigated cropland proportion within 100m ×100m to reduce these effects. The GMIE-100 dataset ranges from 0 to 1, with a no-data value set at -99.

2.5 irrigation method identification method

Motivated by the spatial attention gate, four attention blocks have been incorporated into the connections between down-sampling and up-samplings within the U-Net architecture (Figure 4). The Pivot-Net includes four spatial attention gates to effectively capture information pertaining to the round shape of CPIS. To enhance model comprehension of shape-related intermediate representations during boundary detection and segmentation tasks, a multi-task learning approach was employed in training the model. This approach encompasses pixel-wise segmentation and boundary prediction as integral components of the Pivot-Net's learning objectives. This method was successfully applied in identifying CPIS for the whole US (Tian et al., 2023b).



https://doi.org/10.5194/essd-2023-536
Preprint. Discussion started: 6 February 2024
© Author(s) 2024. CC BY 4.0 License.

Figure 4 Architecture of the shape-attention Pivot-Net (Tian et al., 2023b).

We generated composite, cloud-free satellite data by utilizing optical images from Sentinel-2 and Landsat-8 for each tile within the Google Earth Engine (GEE) from March to August 2020. All exported data from GEE were stored in Google Drive. The world was divided into 345 6°×6°tiles and 23 of them was that annotated manually (Figure 1). 80% of all the CPIS labels or 9140 patches with 256×256 pixels were used for training the model, and rest 20% CPIS labels or 2284 patches was used for accuracy validation.

Subsequently, we transferred the trained model, stored on a local high-performance computer, to Google Drive. Employing the robust computational capabilities of Google Colab Pro+ (https://colab.research.google.com/), which seamlessly accesses satellite data in Google Drive, we applied the well-trained Pivot-Net model across all tiles. The satellite data was partitioned into 256×256 patches with a 128-pixel overlap (Stride = 128 pixels). The final prediction was determined by selecting the maximum prediction probability within the overlap region.

3. Result and Discussion

3.1 Spatial pattern of irrigated cropland and GCPIS

The spatial distribution of GMIE-100 is depicted in Figure 5. GMIE-100 reveals that the maximum extent of irrigated cropland was 403.17 million hectares (Mha), which accounts for 23.4% of the global cropland, equivalent to 1,724.08 Mha. This figure surpasses the total area equipped for irrigation reported by FAOSTAT for 2000–2008 (307.60 Mha) (Siebert et al., 2013) and closely aligns with the irrigated area estimated by IWMI–GIAM(406.40 Mha, representing for 19.5 % of global cropland in 2000) (Thenkabail et al., 2009). India (94.85 Mha, representing 50.4% of cropland) has the largest area of irrigated cropland in the world, with China (85.16 Mha, 50.0% of cropland) and Pakistan (18.04 Mha, 80.2% of cropland) ranking 2nd and 4th respectively. In addition, the United States (26.54 Mha, 15.5% of cropland) ranks 3rd globally in terms of irrigated cropland. For the rest countries, the irrigated cropland is less than 10 million hectares.

The irrigated cropland is notably concentrated in regions characterized by expansive plains and proximity to rivers. These flat and river-proximate areas are well-suited for irrigation due to easy access to water resources. In fact, a substantial portion of the global irrigated cropland, encompassing 224 million hectares, or 55.6% of the total irrigated cropland, is situated in such plain regions. Prominent examples include the Ganges Plain, the Indus Plain, and the North China Plain, all of which host significant expanses of irrigated cropland. Nevertheless, there exist areas where the proportion of irrigated land remains low, despite their close proximity to water sources, For instance, regions like the Danube estuary in Romania exhibit irrigation proportion of 3.65%, despite experiencing high annual food production variability (Wriedt et al., 2009) . Similarly, the Zambezi basin, encompassing countries like Zambia (4.1%) and Mozambique (4.2%), struggles with food insecurity despite its access to water resources.

https://doi.org/10.5194/essd-2023-536
Preprint. Discussion started: 6 February 2024
© Author(s) 2024. CC BY 4.0 License.

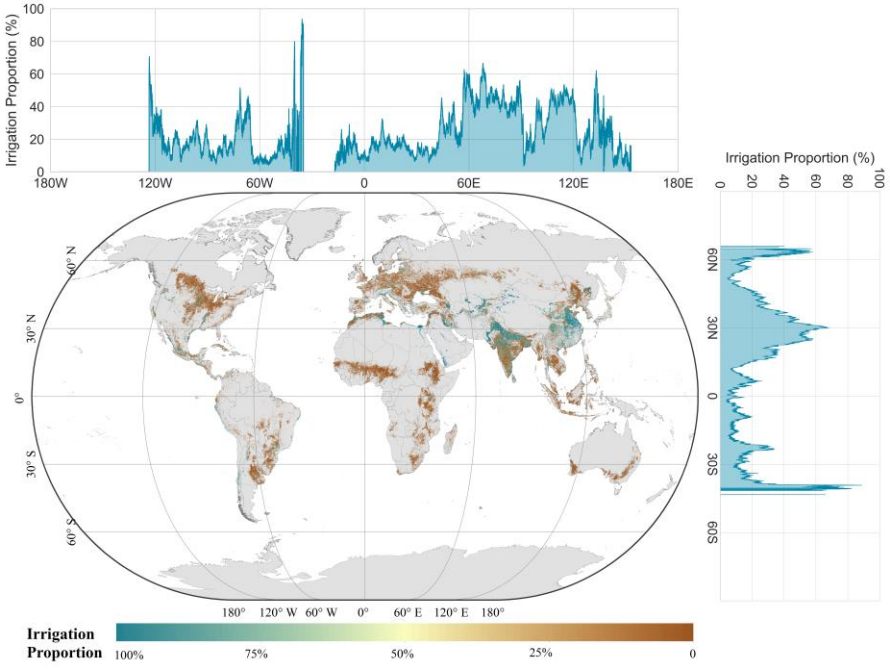


Figure 5 Global dataset of 100m resolution irrigated cropland proportion.

Apart from the plains, oases within arid zones represent another significant category of regions with extensive irrigated cropland. These areas are distinctive for their limited precipitation but abundant sunlight and heat resources. In oases, the availability of irrigation is crucial for crop survival. Approximately 31 million hectares of irrigated cropland are situated within arid zone oases, constituting 7.7% of the total irrigated cropland. Well-known oasis agricultural regions across the world include the Nile basins and the delta region in Egypt, the California Valley in the USA, and Xinjiang in China. These areas thrive on irrigation to make productive use of the scarce water resources amid arid conditions.

The distribution of irrigated cropland exhibits distinct patterns when examined along both latitude and longitude. Along the latitude axis, we observe exceptionally high irrigation proportions around the 30°N latitude line, which encompasses regions along the lower Yangtze River, Ganges River, Indus River, and Nile River. These river basins are characterized by dense concentrations of irrigated cropland, owing to the availability of water resources from these major river systems. On the other hand, when assessing irrigation proportions along the longitude axis, we note elevated levels of irrigation between 60°E and 120°E. This longitudinal span encompasses prominent regions such as the Indus-Ganges Plain and the North China Plain, which are renowned for their high levels of irrigated agriculture.

For the CPIS in the world, the spatial pattern was depicted in Figure 6. The total area of CPIS was estimated as 115,192.2 km^2, comprise 2.9% of total irrigated area. While the area in Chen's research is 107,232.8km2 (Chen et al., 2023) in global arid regions. The CPIS was mainly distributed in the High Plain Aquifer (HPA), including north of Texas, Kansas and Nebraska, south part of Brazil, South Africa, and middle east region. Along the longitude, the proportion of CPIS was high from 90°W

https://doi.org/10.5194/essd-2023-536
Preprint. Discussion started: 6 February 2024
© Author(s) 2024. CC BY 4.0 License.

to 120°W, which matches the range of HPA, while the CPIS proportion was relative apparent between 30°N to 60° N along latitude.

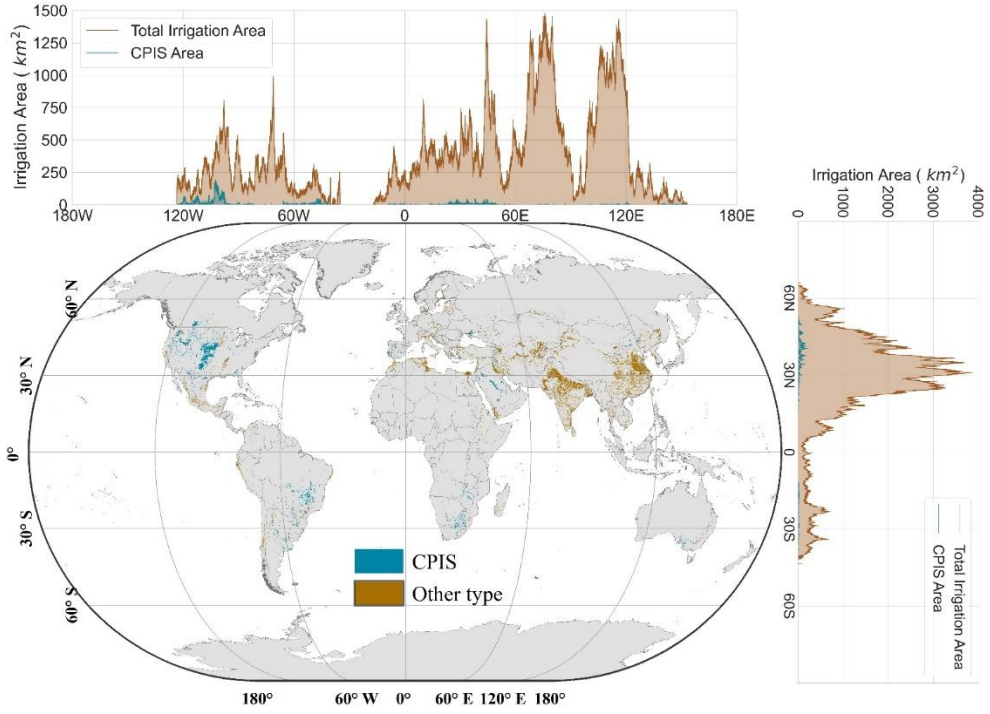


Figure 6 The distribution of irrigation types within irrigation extent

The distribution of irrigated cropland and CPIS proportion across the six continents is depicted in Figure 7a. Asia leads with the highest irrigated area, covering 273.79 million hectares (Mha), or with an irrigation proportion of 39.3%. North America follows with 16.9%, South America with 15.5%, Europe with 10.6%, Africa with 9.6%, and Oceania with 9.2%. As for irrigation method, the proportion of CPIS was highest in North America with CPIS accounting 13.8% of total irrigated areas, followed by South America of 5.0% and Oceania of 2.9%.

In Figure 7b, we summarized the irrigation and CPIS proportions across different climate zones. We used global aridity index and criterion in literature to classify the climate zone (Zomer et al., 2022).The irrigation proportion experiences a significant decrease, plummeting from 91.8% in hyper-arid zones to 20.7% in semi-humid zones. It then exhibits a slight increase to 21.4% in humid zones. These variations in irrigation proportions correspond to the distinct water availability and climatic conditions in these regions. As for the irrigation method, the CPIS proportion was highest in hyper-arid region of 5.7%, followed by 3.9% in semi-arid region.

https://doi.org/10.5194/essd-2023-536
Preprint. Discussion started: 6 February 2024
© Author(s) 2024. CC BY 4.0 License.

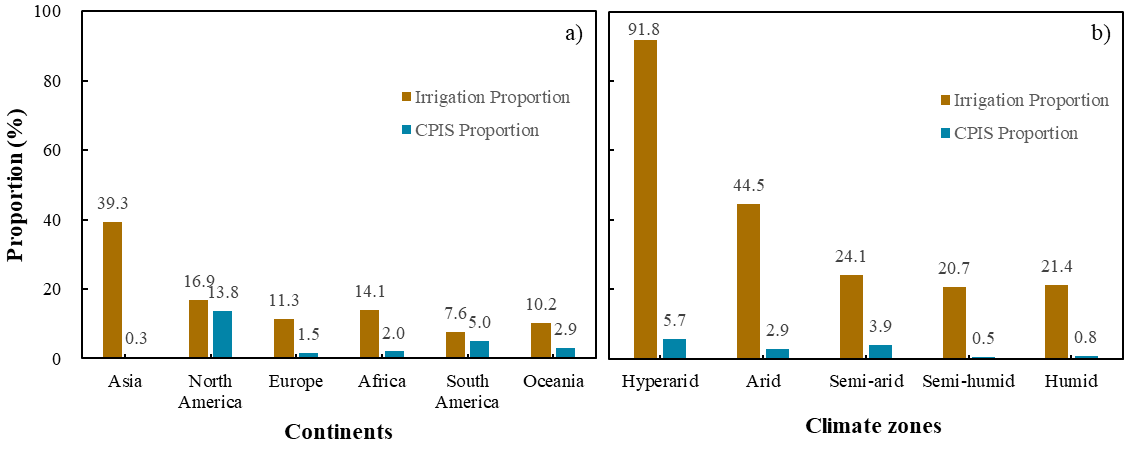


Figure 7 The irrigation proportion and CPIS proportion of total irrigated area for continents (a) and climate zones(b)

Figure 8a illustrates the irrigation proportion for each country. Notably, the irrigation proportion exhibits higher values along the geographical expanse stretching from North Africa through West Asia, South Asia, and into East Asia. In Figure 8b, the irrigation proportions are presented for each IMZ. The spatial distribution aligns with the pattern depicted in Figure 8a. Several countries in West Asia and North Africa, including Oman, Saudi Arabia, Qatar, and Egypt, boast irrigation proportions of 100%. Additionally, three countries surpass an irrigation proportion of 80%, namely Turkmenistan (89.4%), Uzbekistan (81.3%), and Pakistan (80.4%). Among all the AEZs, Gansu-Xinjiang in China have the highest irrigation proportion at 100.0%, followed by Central Northern Andes (96.2%), Old World Deserts (90.5%), Southern Himalayas in India (84.0%), Semi-Arid Southern Cone (82.9%), and China Lower Yangtze (80.8%).

Figure 8c and d are CPIS proportion for each country and IMZ, respectively. CPIS is mainly concentrated in countries with intensified agricultural regions and extreme arid zones such as the Middle East. The highest CPIS proportion was in Namibia of 23.4%, followed by the US with 20.33%, Saudi Arabia of 16.3%, south Africa of 15.7%, Canada of 12.6%, Zambia of 12.5%, Gaza Strip of 12.2% and Brazil of 9.6%. As for the IMZs, the CPIS proportion was most distinguish in Amazon (C24) of 81.2%, North of High Plain (C12-4) of 42.5%, South Zambia(C09-3) of 41.6%, America Northwestern great plains(C12-3) of 36.0%, Western of Mongolia(C47) of 25.0%, British Columbia To Colorado(C11) of 24.2%, America cotton belt_Mexican coastal plain(C14-1) of 22.8%, SW Mexico and N. Mexico highlands (C18) of 21.4%.

https://doi.org/10.5194/essd-2023-536
Preprint. Discussion started: 6 February 2024
© Author(s) 2024. CC BY 4.0 License.

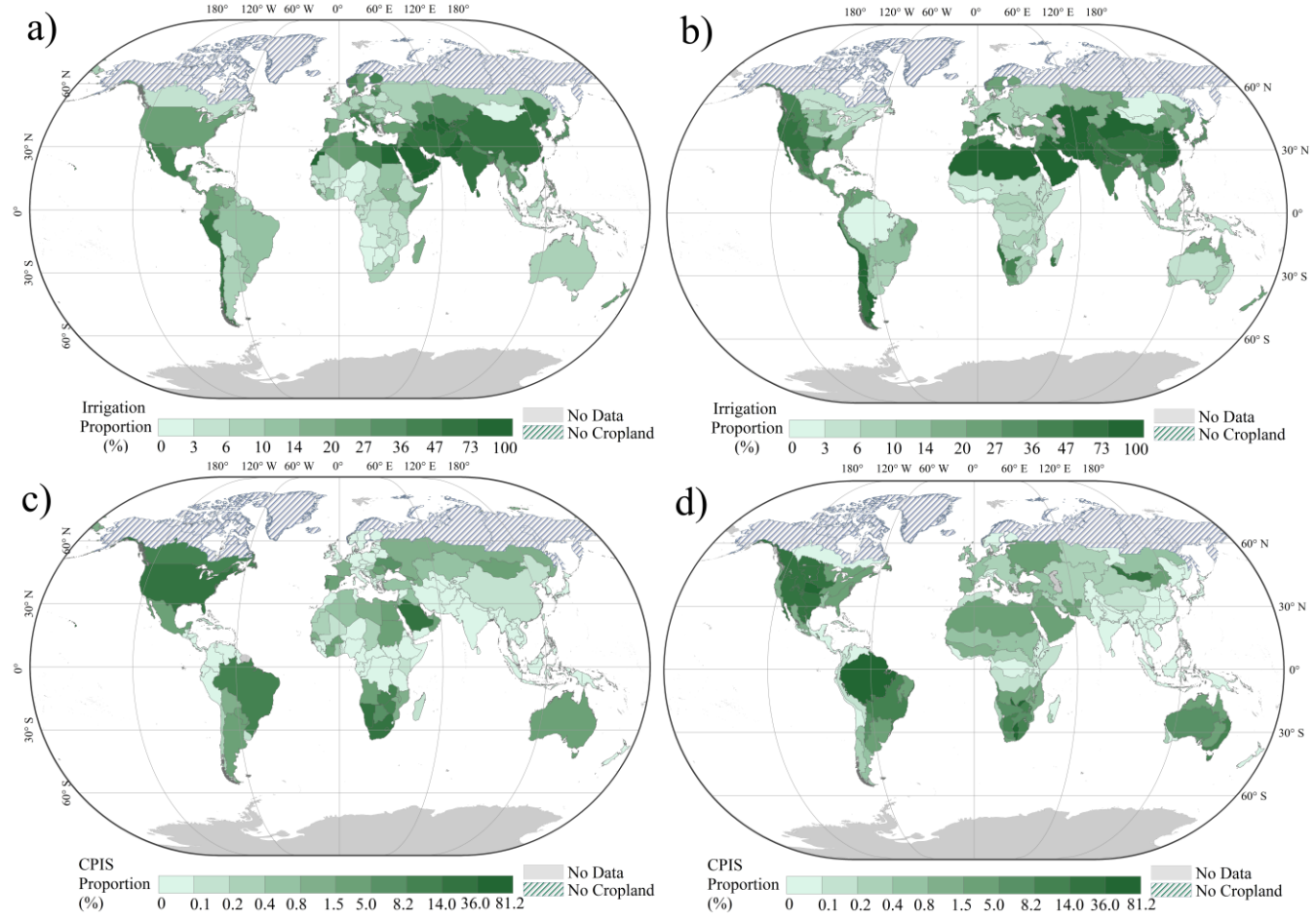


Figure 8 The irrigation proportion for each country (a) and IMZs (b) and CPIS proportion of total irrigated cropland for each country(c) and IMZs(d)

https://doi.org/10.5194/essd-2023-536
Preprint. Discussion started: 6 February 2024
© Author(s) 2024. CC BY 4.0 License.

3.2 Reliability of GMIE-100

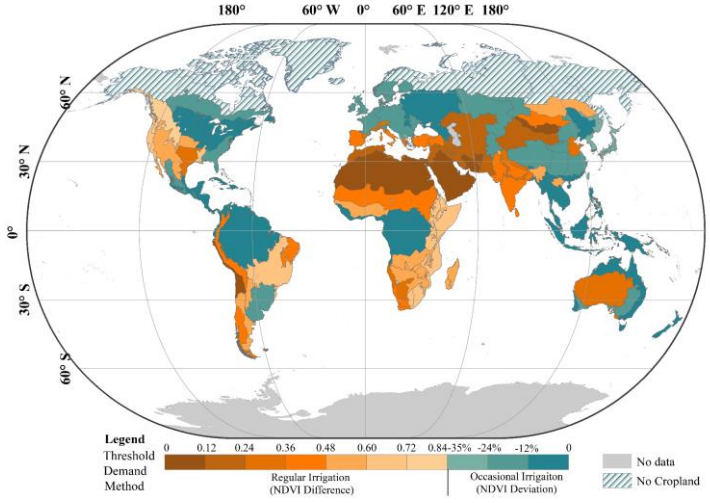


Figure 9 The threshold of NDVI difference and deviation for each IMZ

For each IMZ, the irrigation mapping method and threshold of NDVI or $NDVI_{dev}$ is shown in Figure 9. For the IMZ with regular dry season, the NDVI difference method was employed to find the amplified condition difference between irrigation and rainfed cropland. To avoid the omission of fallow land and crop rotation, the maximum NDVI in dry month during 2017-2019 was selected. The NDVI threshold for each IMZ was determined using training samples, which is ranged from 0.10 in extremely arid region, such as Old-World deserts (IMZ C64) and 0.74 in British Columbia to Colorado in North America (IMZ C11) as shown in orange series color of Figure 9. These thresholds are integral to the accurate identification of irrigated cropland within each IMZ.

For region without significant dry season, driest month of extremely dry year among 10 years (2010-2019) was selected to amplify the crop condition between irrigation and rainfed cropland. The $NDVI_{dev}$ was calculated as proxy of crop condition departure from 10-year average using collected training samples. The value was ranged from −1.0 % (Amazon, C24) and −37.0 % (C60-10, north-western Greece and southwestern Albania), as shown in blue series color of Figure 9.

https://doi.org/10.5194/essd-2023-536
Preprint. Discussion started: 6 February 2024
© Author(s) 2024. CC BY 4.0 License.

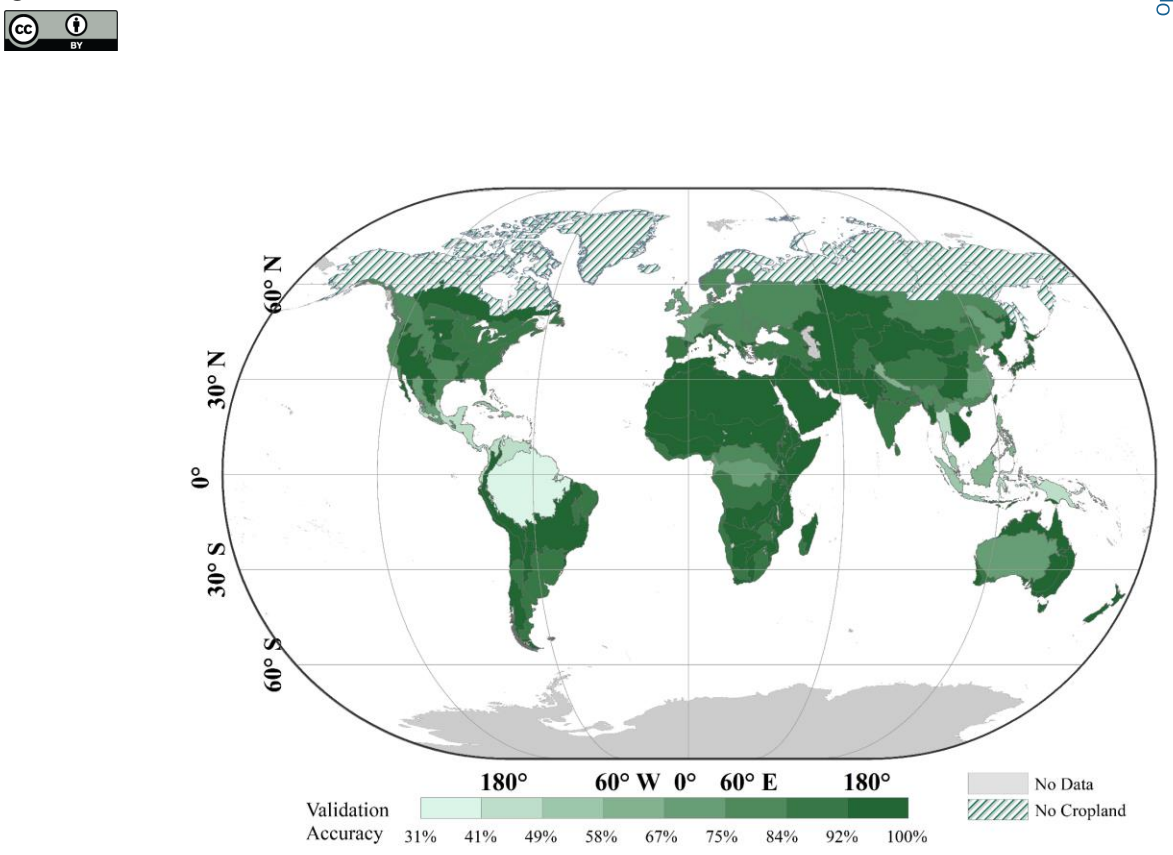


Figure 10 Training accuracy for each AEZ

Figure 10 is the training accuracy of each IMZ. The NDVI or $NDVI_{dev}$ threshold was determined using Fisher Discrimination method with 92,303 obtained samples. Then, the training accuracy was assessed, which is between 0.31% in Amazon (C24) and 100% in Western Asia(C31-2). Although the accuracy in some humid region, such as Northern South and Central America (42%) and Caribbean (49%), there is 89 IMZ with accuracy larger than 80% among 105 IMZs with cropland. The confusion matrix accuracy metrics of GMIE-100 was shown in Table 2. To validate the final accuracy of GMIE-100, the rest 20% of samples or 23,076 points was used. The overall accuracy of GMIE-100 was 83.6% with user accuracy of 86.1% and produce accuracy of 82.2%.

Table 2 Confusion matrix and accuracy assessment of GMIE-100

		Field points			
	Classes	Rainfed	Irrigation	Total	User accuracy
Predicted	Rainfed	9,270	2,170	11,440	81.0%
	Irrigated	1,622	10,014	11,636	86.1%
	Total	10,892	12,184	23,076	
	Producer accuracy	85.1%	82.2%		
Overall Accuracy:		83.6%			

https://doi.org/10.5194/essd-2023-536
Preprint. Discussion started: 6 February 2024
© Author(s) 2024. CC BY 4.0 License.

The accuracy of GMIE-100 was evaluated in 10 countries, which is presented in Figure 11, showing overall accuracy, user accuracy and producer accuracy for each country. In China, the accuracy was assessed using13,963 ground truth data from multi-year GVG data. The overall accuracy was 85.5% with produce accuracy of 86.7% and user accuracy of 83.3%. Commissions and omissions were common in the humid areas, such as Southern China, Cambodia and Myanmar. In other countries, the overall accuracy of GMIE-100 datasets was basically acceptable.

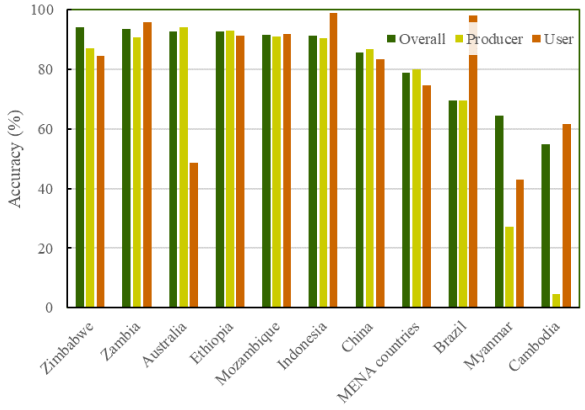


Figure 11 Accuracy for countries with GVG irrigation validation points

For the accuracy of CPIS, accuracy metrics and confusion metrics was listed in Table 3. The model achieved a high validation accuracy of 97.87%. The F1 score, which is a balance between precision and recall, is 86.87%. The Mean Intersection Over Union (IOU) is 87.25%. We visualized four patches with dense CPIS in Figure 12. Overall, the CPIS is well identified in most case.

Table 3 Confusion Matrix of GCPIS identified with Pivot-Net

		CPIS Predict		Recall
		0	1	
CPIS Label	0	119938874	735300	99.39%
	1	2077463	9303403	81.75
Precision		98.30%	92.68%	
Overall Accuracy				97.87%

https://doi.org/10.5194/essd-2023-536
Preprint. Discussion started: 6 February 2024
© Author(s) 2024. CC BY 4.0 License.

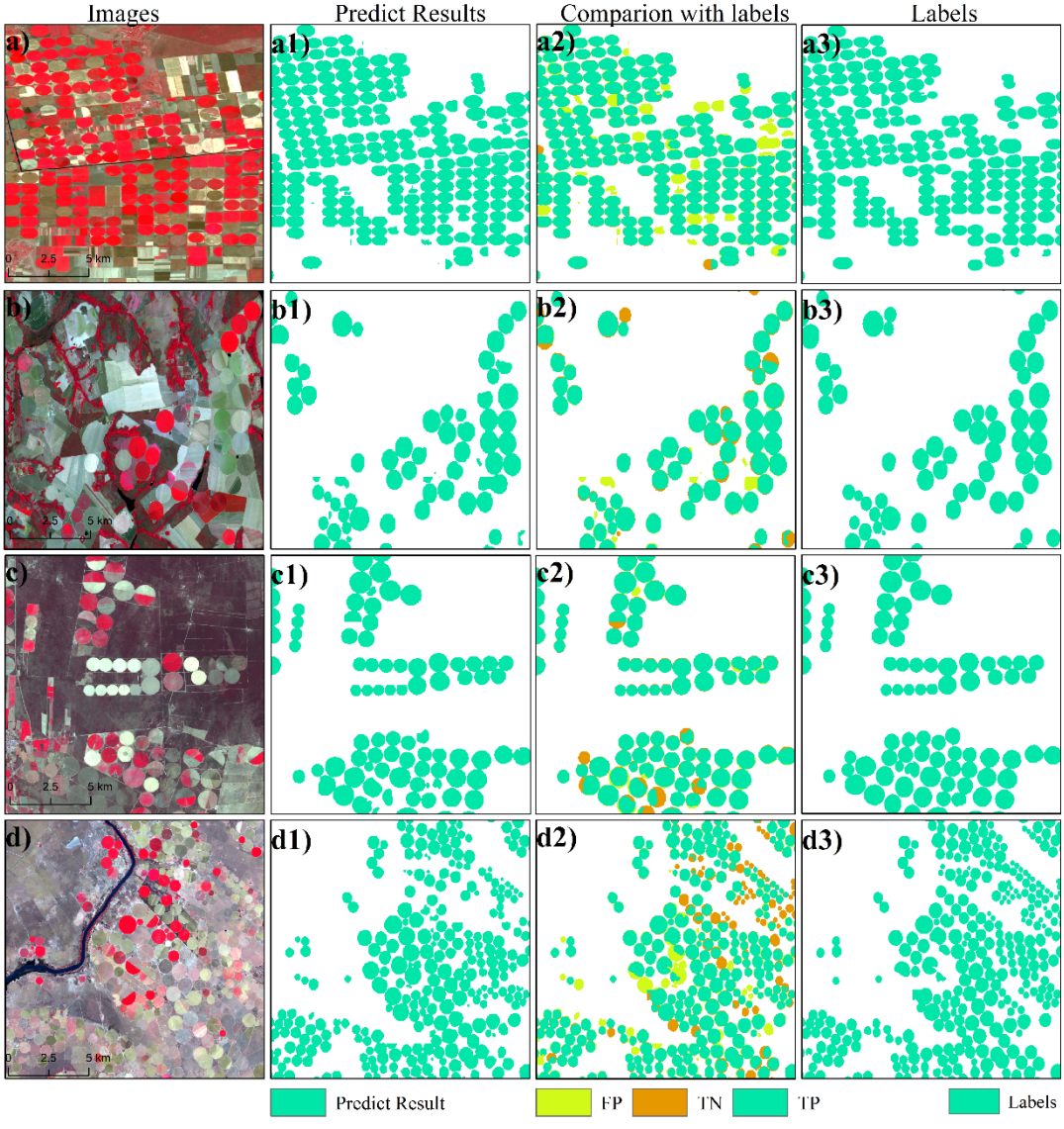


Figure 12 Accuracy assessment for CPIS identification result. a-d are the composted images; a1-d1 are the predict results of Pivot-Net; a2-d2 are the comparison between our result and labels. TP means truth positive pixels, while TN represents truth negative pixels. FP means false positive samples. a3-d3 are the labels. The central point coordinate of a-d are (33.86,46.37),(-47.34,-16.41),(-65.74,-32.03),(25.11,-28.06), respectively. The background images of a-d are Landsat-8 data. Credit of a-d is @U.S. Geological Survey

https://doi.org/10.5194/essd-2023-536
Preprint. Discussion started: 6 February 2024
© Author(s) 2024. CC BY 4.0 License.

3.3 Comparison with existed irrigation dataset

3.3.1 Comparison of irrigated cropland

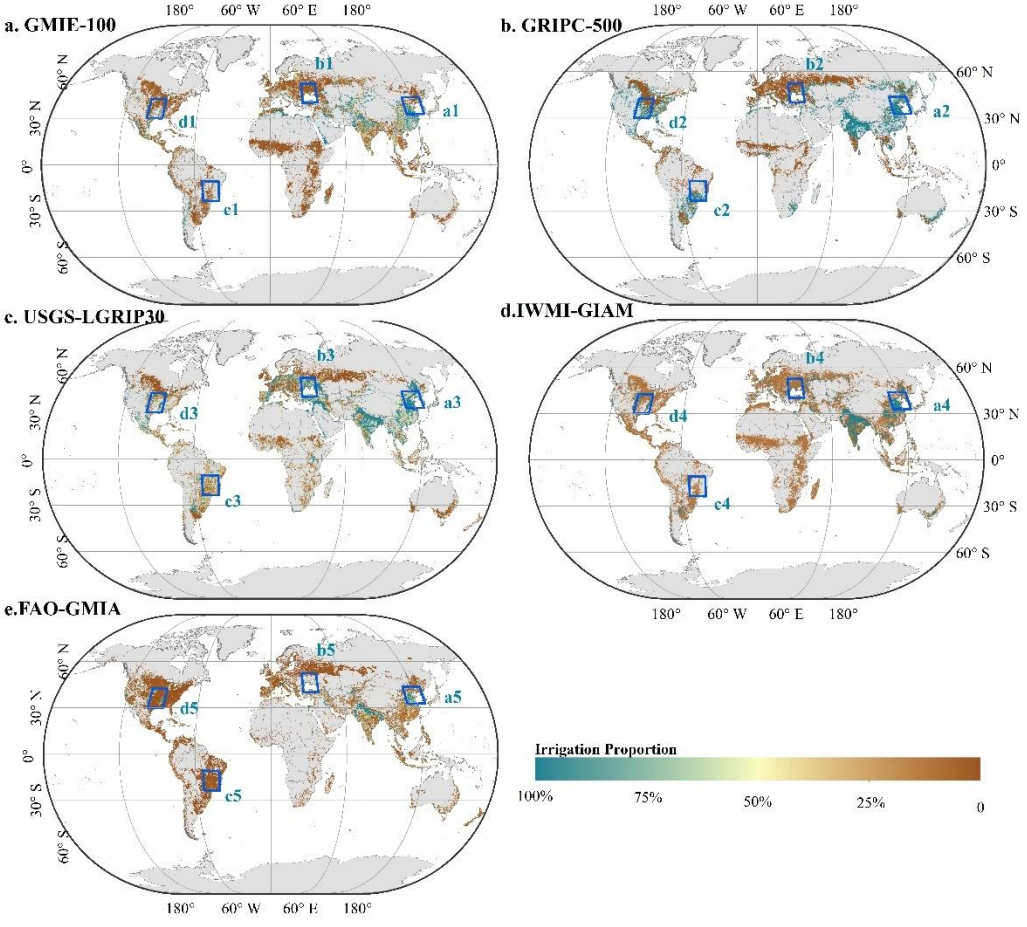


Figure 13 The comparison with exist irrigation production at 1km (GMIE-100, GRIPC-500, USGS-LGRIPC30) or 10km resolution (IWMI-GIAM and FAO-GMIA)

To compare GMIE-100 with four exist irrigation production, we downscale GMIE-100 and GRIPC-500 and USGS-LGRIP30 to 1km resolution and upscale IWMI-GIAM and FAO-GMIA to 1km resolution using bilinear interpolation method. The result was shown in Figure 13. The spatial pattern of irrigated cropland in GMIE-100 was generally coincide with other products. Irrigated cropland was most concentrated in North China Plain and Ganges & Indus River basin around the world.

Nevertheless, there were also discernible difference in the detail distribution for the patches irrigated cropland, such as the Northeast of China, Eastern European Plain, Planicie de la Plata of South America and lower Mississippi River basin (Figure 14). In Northeast of China plain, the irrigated cropland is denser in USGS-LGRIP30 and GRIPC-500 than other product. According to the census data of China, the average irrigation proportion for three province (Heilongjiang, Jilin, Liaoning

https://doi.org/10.5194/essd-2023-536
Preprint. Discussion started: 6 February 2024
© Author(s) 2024. CC BY 4.0 License.

Province) was 39.32%. According to the result in GMIE-100, the irrigation proportion was 27.45%, which is closer to the census data. For the irrigated cropland in the Eastern European Plain, USGS-LGRIP30 illustrates widely distributed irrigated cropland, which is significantly denser than what is portrayed in GMIE-100 and the other three datasets (Figure 14 b1-b4). Notably, the GRIPC-500 dataset indicates a considerable extent of irrigated cropland in the Planicie de la Plata region when compared to GMIE-100 and the other products ((Figure 14 c1-c4)). According to census data from Brazil, the reported irrigation proportion is 6%, whereas it is 58% and 72% in USGS-LGRIP30 and GRIPC-500, respectively.

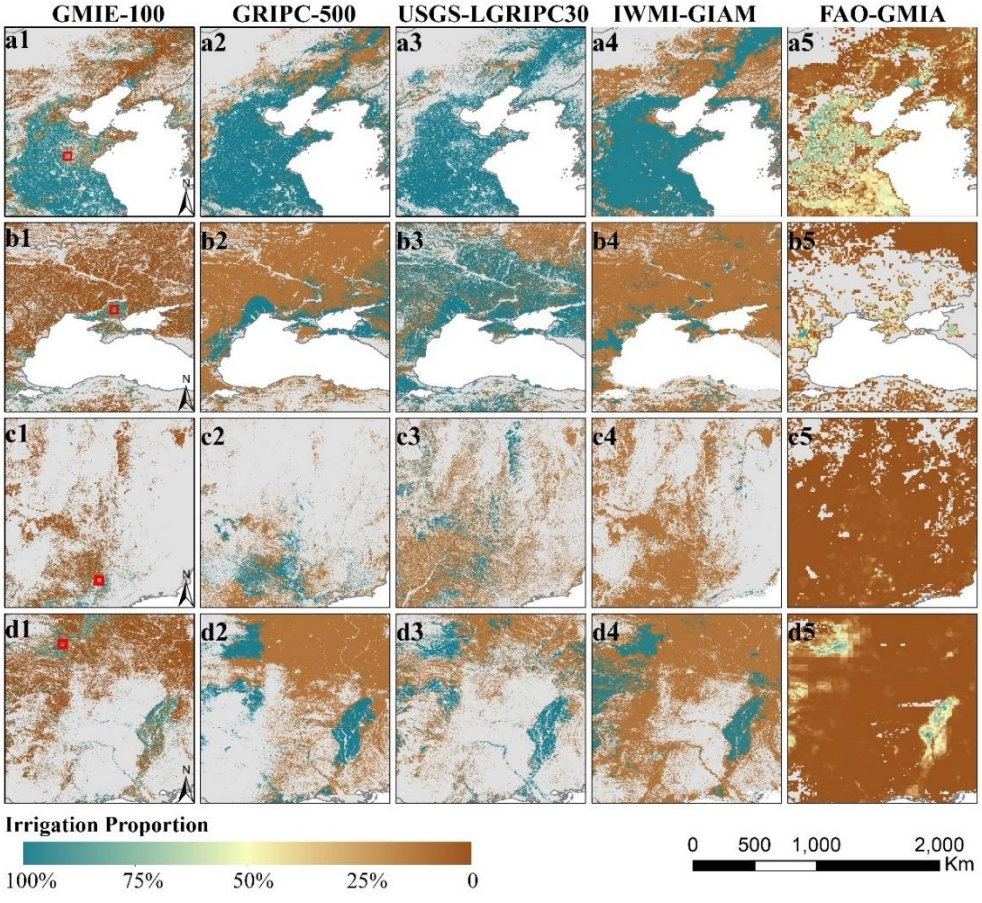
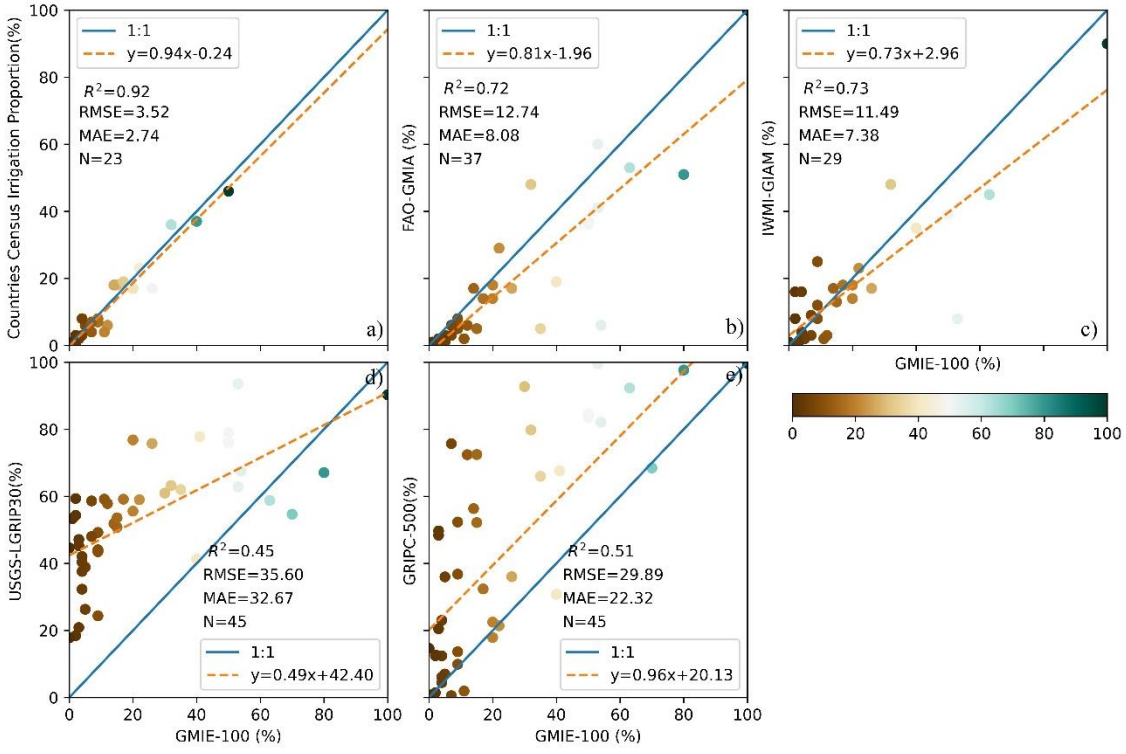


Figure 14 The comparison with exist irrigation production for hot-point region of irrigation; The corresponding location was labelled in Figure 13 with blue rectangle

To further validate the proposed GMIE-100, we compare it with national census data. The result is shown in Figure 15. To compare with exist global irrigation product, we also compare GMIE-100 with FAO-GIAM, IWMI-GMIA and USGS-LGRIP30 and GRIPC-500.The R^2 between GMIE-100 and 23 national census data was 0.92, with RMSE of 3.52 and MAE of 2.74. For FAO-GIAM and IWMI-GMIA, the R^2 with GMIE-100 was 0.72 and 0.73, respectively. The determination coefficient between USGS-LGRIP30 and GMIE-100 was only 0.45 with an RMSE of 35.6, the lowest value among these three existing irrigation products. When we compare USGS-LGRIP30 with the national census, the R^2 was only 0.25. When comparing

https://doi.org/10.5194/essd-2023-536
Preprint. Discussion started: 6 February 2024
© Author(s) 2024. CC BY 4.0 License.

GMIE-100 with GRIPC-500, the R^2 was 0.51 with an RMSE of 29.89%. The scatterplot shows that GRIPC-500 was overestimated compared to GMIE-100.



455 **Figure 15 The comparison of national irrigation proportion between GMIE-100 and national census data (a), FAO-GMIA (b), IWMI-GIAM(c), USGS-LGRIP30(d) and GRIPC-500(e).**

3.4 Advantages and limitations of GMIE-100

Using irrigation performance, we could do irrigation mapping at regular intervals. The irrigation area describes a high level of variability in irrigation water use (Puy et al., 2021; Puy et al., 2022). Thus, the changes in irrigated area could reflect
460 the variation in agricultural water use, which is important for local water resource management. Due to the lack of updated information, global maps of irrigated areas often relied on estimates from around 2000 (Nagaraj et al., 2021). For the RIR region, the irrigation map can be updated every three years by collecting the vegetation signal in each dry season. For the RIO region, the irrigation map can be updated every ten years based on crop status during extremely dry events within 10 years. Although, the irrigated cropland extent during the dry season can be identified during 2010 to 2019, our aim was to provide
465 the most up-to-date information using satellite data over the 2017-2019 period.

Periodic cropland fallowing refers to the practice of not cultivating or tilling all croplands within a single year. This approach is often employed to restore soil fertility as part of a crop rotation scheme or to prevent excess agricultural production. Utilizing the NDVI or $NDVI_{dev}$ threshold, it becomes feasible to identify and distinguish only those lands that have been actively cultivated. Subsequently, these cultivated lands can be further categorized into either irrigated or rainfed land. An area

https://doi.org/10.5194/essd-2023-536
Preprint. Discussion started: 6 February 2024
© Author(s) 2024. CC BY 4.0 License.


is designated as irrigated if it has been cultivated at least once during the driest month over a span of three years. This criterion aids in discerning areas that are actively managed for crop production from those temporarily left fallow or unplanted.

The spatial resolution of this dataset was 100 m, which is higher than that of the dominant irrigation data map. High-resolution data on irrigated cropland data is essential for quantifying agricultural water use (Wu et al., 2022). The resolution of most existing irrigation data is very coarse, varying between 500m to 10km (Xie et al., 2019). As shown in Figure 16, GRIPC-500, IWMI-GIAM and FAO-GMIA are not able to present detail information on irrigated cropland. Even the resolution of USGS-LGRIPC-30 was higher than that of GMIE-100, the later description of heterogeneous irrigated cropland distribution in the North China Plain (Figure 16 a1 and a3) and the US Plateau (Figure 16 d1 and d3) was better than the earlier one.

Furthermore, with support of DL method, we achieved the CPIS mapping around the world to investigated the irrigation method. We found that there is 11.5 Mha CPIS around the word, which comprise 2.9% of total irrigated cropland. To my best knowledge, this is the first research that mapped irrigated method of CPIS, despite Chen's research finish the CPI mapping in global arid region (Chen et al., 2023). GMIE comprise both of irrigated cropland extent and some irrigation method (CPIS) distribution with relative high resolution, which will definitely promote sub-basin water consumption and withdraw estimation for all sectors (Wu et al., 2022). Due to the variation of irrigation efficiency for different irrigation method, CPIS demonstrates an efficiency exceeding 80%, while gravity flowing irrigation methods exhibit a comparatively lower efficiency, approximately 60% (Waller and Yitayew, 2016). So, we could outlook that the irrigation efficiency may can be estimated with the component of irrigation method in the future. This could be enhance the understanding of irrigation paradox (Grafton et al., 2018), which indicating that technological advancement increase irrigation efficiency, but crop water didn't decrease.

https://doi.org/10.5194/essd-2023-536
Preprint. Discussion started: 6 February 2024
© Author(s) 2024. CC BY 4.0 License.

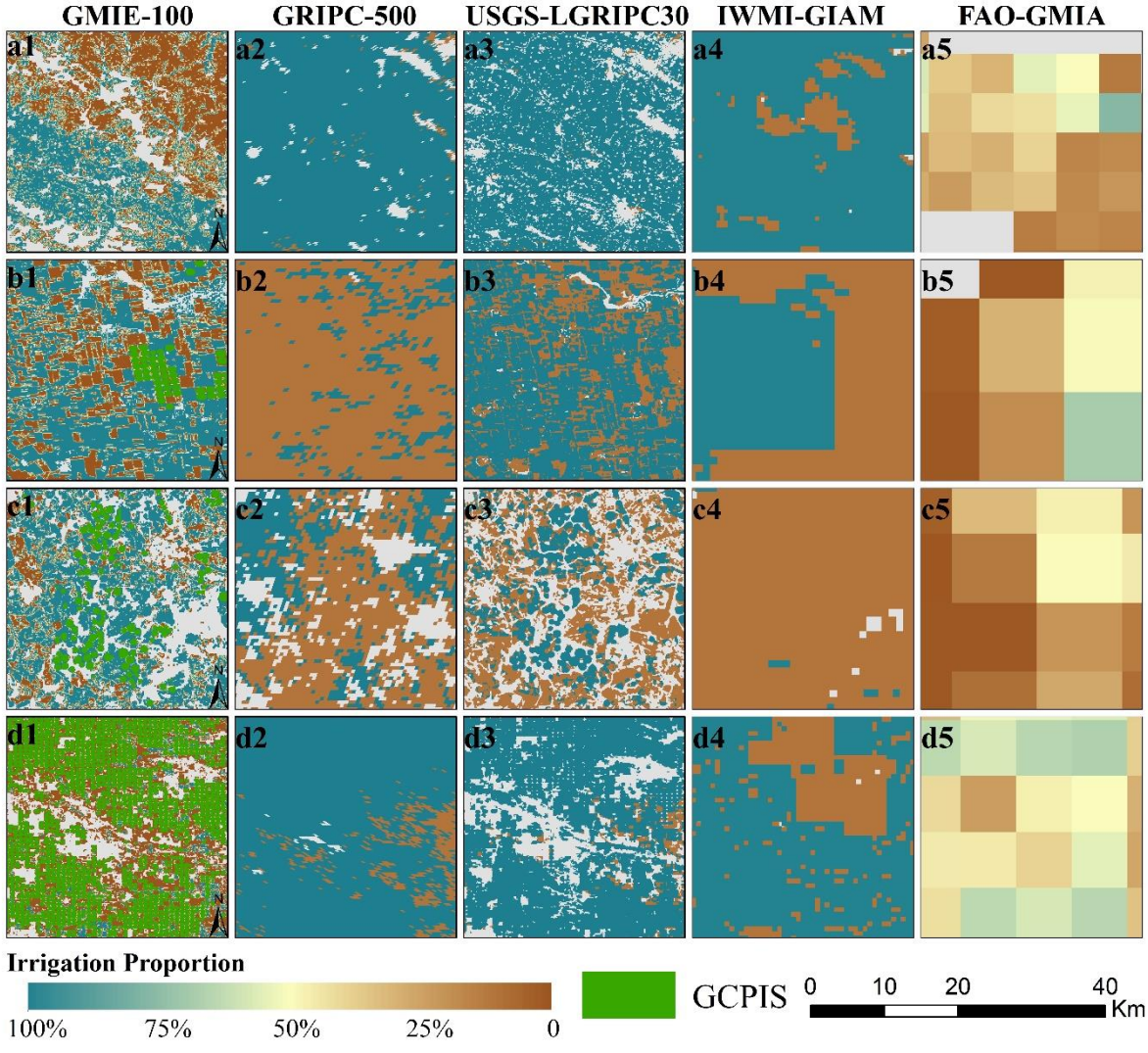


Figure 16 Comparison between GMIE-100 with exist global irrigation product on detail, and their specific location was labelled in the corresponding subfigure of Figure 14 with red rectangle.

Compared to the surveillance classification method, this method requires fewer samples. Due to lack of expertise, all spectral characteristics of irrigated farmland were studied using training samples, which will definitely increase the required sample amount. Xie's research used 20,000 samples for irrigation mapping in the United States (Xie et al., 2019). Zhang's research used approximately 100,000 samples were used to identify irrigated croplands in China (Zhang et al., 2022b). By determining the NDV difference and NDVI deviation between irrigated and rainfed cropland, the required amount of training samples could be drastically reduced. In this study, a total of 92,303 samples were involved in determining the NDVI threshold

https://doi.org/10.5194/essd-2023-536
Preprint. Discussion started: 6 February 2024
© Author(s) 2024. CC BY 4.0 License.


and NDVI deviation threshold at the global scale. Meanwhile, the training sample in China was mostly collected on site, which is more precise than visual interpretation.

Also, there is some limitation for this method. Firstly, the accuracy of GMIE-100 is extremely wet season was not so high, because the water stress is seldomly emerging. As a result, the accuracy in some southeast Asia countries need to be further improved, such as Myanmar with overall accuracy of 64.5% and Cambodia of 54.93%. Also, the representativeness of sample points can be further improved, e.g. by identifying central picot irrigation system using deep learning method (Tian et al., 2023b; Chen et al., 2023), which is common used in US, Brazil and Middle-East regions.

There are also some limitations to this method. First, the accuracy of GMIE-100 was not so high in the extremely wet season because water stress rarely occurs. As a result, the accuracy needs to be further improved in some Southeast Asian countries, such as Myanmar with an overall accuracy of 64.5% and Cambodia with 54.93%. The representativeness of sampling points can also be further improved, for example, by identifying the central picot irrigation system using the deep learning method (Tian et al., 2023b; Chen et al., 2023), which has been used in the USA, Brazil and the Middle East is commonly used.

Secondly, the cropland masks exhibited the most pronounced influence on the GMIE-100 dataset(Salmon et al., 2015; Meier et al., 2018), despite the selection of 16 distinct cropland datasets derived from country- and region-level sources as high-priority inputs. These datasets often exhibit disparities in estimating the distribution of cropland, particularly in African countries, due to the complex landscape, frequent cloud cover, and the presence of small field sizes (Nabil et al., 2020). Consequently, inaccuracies within the cropland datasets were transposed onto the GMIE-100 dataset. Nevertheless, it's important to note that these datasets remain the primary source of cost-effective and up-to-date information covering vast geographical areas.

Furthermore, although GMIE-100 provides the distribution of higher resolution of irrigated cropland, there are also mixed pixels with both of cropland or non-cropland, irrigated or rainfed cropland. Especially for the region with extremely small agricultural field (Fritz et al., 2015).

4. Conclusion

High resolution and updated irrigation map are important for tracking regional water use and food producing situation. Using irrigation performance during dry season of growing season and extremely drought event, we produced GMIE-100 at 100m with support of GEE. This study involved the division of the entire globe into 110 zones, driven by the variances in climate and phenology. In each IMZ, we identified the dry months during the growing season within the 2017-2019, or alternatively, the driest months during the most arid year from 2010-2019. To distinguish irrigated cropland, we employed 92,303 samples to determine thresholds for the NDVI during the dry months of 2017-2019 and the NDVI deviation from the ten-year average for the driest month ($NDVI_{dev}$). NDVI or $NDVI_{dev}$ threshold that achieved the higher overall accuracy was selected to distinguish irrigated and rainfed cropland. All the algorithm was conducted on GEE with code of https://code.earthengine.google.com/eaafaab35dde9bbe37f443e80c716479.

https://doi.org/10.5194/essd-2023-536
Preprint. Discussion started: 6 February 2024
© Author(s) 2024. CC BY 4.0 License.


With the support of DL method, global CPIS was identified using Pivot-Net. We found that there is 11.5 million hectares irrigated cropland using central pivot irrigation system, accounting about 2.9% of total irrigated cropland. But in Namibia, US, Saudi Arabia, south Africa, Canada and Zambia, CPIS proportion was larger than 10%. To our best knowledge, this is the first effort to identify irrigation method globally, though other type of irrigation method, such as gravity flowing, is still dominant irrigation method. But this could facilitate the irrigation efficiency estimation using different irrigation method proportions for supporting high-accuracy sub-basin scale water resource management.

Finally, the global maximum irrigation extent (GMIE-100) was produced at 100 meters. Using 23,076 points to validate the result, the overall accuracy of GMIE-100 was 83.6% but varying from different IMZs. The GMIE-100 indicates that the largest extent of irrigated cropland reached 403.17 million hectares, which accounts for 23.4% of the total global cropland. Spatially, the irrigated cropland is concentrated in the great plains and regions near the rivers. 224 million hectares irrigated cropland, accounting 55.6% of total irrigated cropland, was in plain region. The Ganges Plain, the Indus Plain and the North China Plain all have a large amount of irrigated cropland around the word. The GMIE-100 provides more detail information about irrigated and rainfed cropland, thus could better support agricultural water use estimation and regional food situation assessment.

5. Code and data availability

The data is publicly accessible through the following link: https://doi.org/10.7910/DVN/HKBAQQ (Tian et al., 2023a). The GMIE-100 dataset spans values ranging from 0 to 1, with a designated no-data value of -99. Globally, there are 67 tiles available, each with a maximum extent of 21°×21°. In cases where these tiles overlap with land, they maintain the standard extents; however, adjustments are made to the tile extents as needed to accommodate the terrestrial range. The GCPIS was stored in shapefiles formats in zip files.

Author contribution

HZ, and BW conceptualized the study. FT designed the experiments and carried out the experiments. BW and HZ were responsible for funding acquisition. MZ and WZ conducted investigation and formal analysis. FT prepared the original draft of the manuscript. FT, BW, HZ, MZ, WZ, NY ,YL, YL reviewed and edited the manuscript.

Competing interests

The authors declare that they have no conflict of interest.

https://doi.org/10.5194/essd-2023-536
Preprint. Discussion started: 6 February 2024
© Author(s) 2024. CC BY 4.0 License.

Acknowledgements

We gratefully acknowledge the support of the Google Earth Engine platform, which provided essential computational and storage resources, simplifying access to archived datasets such as TM/ETM/OLI satellite data, TRMM, GLDAS for precipitation data, and MOD16A2.006 for evapotranspiration data. These resources greatly facilitated program calculations and data retrieval. Then we thank the data provider of above-mentioned data as well as GFSAD30 team for publishing irrigated and rainfed samples. Furthermore, we would like to express our gratitude to the authors of existing irrigation datasets, namely GRIPC-500, USGS-LGRIP30, IWMI-GIAM, and FAO-GMIA, for their foundational work, which has significantly contributed to our research in this field. Their efforts have provided essential background information for our study.

Financial support

This research was supported by Natural Science Foundation of China (No. 41861144019), Agricultural Remote Sensing Innovation Team Project of AIRCAS (No. E33D0201-6), National Key Research and Development Program of China (2016YFA0600304, 2016YFA0600302).

Reference

Ambika, A. K., Wardlow, B., and Mishra, V.: Remotely sensed high resolution irrigated area mapping in India for 2000 to 2015, Sci Data, 3, 160118, 10.1038/sdata.2016.118, 2016.

Chen, F., Zhao, H., Roberts, D., Van de Voorde, T., Batelaan, O., Fan, T., and Xu, W.: Mapping center pivot irrigation systems in global arid regions using instance segmentation and analyzing their spatial relationship with freshwater resources, Remote Sens. Environ., 297, 113760, 10.1016/j.rse.2023.113760, 2023.

Chen, Y., Lu, D., Luo, L., Pokhrel, Y., Deb, K., Huang, J., and Ran, Y.: Detecting irrigation extent, frequency, and timing in a heterogeneous arid agricultural region using MODIS time series, Landsat imagery, and ancillary data, Remote Sens. Environ., 204, 197-211, 10.1016/j.rse.2017.10.030, 2018.

Dari, J., Brocca, L., Modanesi, S., Massari, C., Tarpanelli, A., Barbetta, S., Quast, R., Vreugdenhil, M., Freeman, V., Barella-Ortiz, A., Quintana-Seguí, P., Bretreger, D., and Volden, E.: Regional data sets of high-resolution (1 and 6 km) irrigation estimates from space, Earth System Science Data, 15, 1555-1575, 10.5194/essd-15-1555-2023, 2023.

Deines, J. M., Kendall, A. D., Crowley, M. A., Rapp, J., Cardille, J. A., and Hyndman, D. W.: Mapping three decades of annual irrigation across the US High Plains Aquifer using Landsat and Google Earth Engine, Remote Sens. Environ., 233, 111400, 10.1016/j.rse.2019.111400, 2019.

dela Torre, D. M. G., Gao, J., Macinnis-Ng, C., and Shi, Y.: Phenology-based delineation of irrigated and rain-fed paddy fields with Sentinel-2 imagery in Google Earth Engine, Geo-spatial Information Science, 24, 695-710, 10.1080/10095020.2021.1984183, 2021.

https://doi.org/10.5194/essd-2023-536
Preprint. Discussion started: 6 February 2024
© Author(s) 2024. CC BY 4.0 License.

585 Duda, R. O., Hart, P. E., and Stork, D. G.: Pattern classification, John Wiley & Sons2012.

Fritz, S., See, L., McCallum, I., You, L., Bun, A., Moltchanova, E., Duerauer, M., Albrecht, F., Schill, C., and Perger, C.: Mapping global cropland and field size, Global change biology, 21, 1980-1992, 2015.

Gommes, R., Wu, B., Li, Z., and Zeng, H.: Design and characterization of spatial units for monitoring global impacts of environmental factors on major crops and food security, Food and Energy Security, 5, 40-55, 2016.

590 Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R.: Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone, Remote Sensing of Environment, 202, 18-27, 2017.

Grafton, R. Q., Williams, J., Perry, C. J., Molle, F., Ringler, C., Steduto, P., Udall, B., Wheeler, S., Wang, Y., and Garrick, D.: The paradox of irrigation efficiency, Science, 361, 748-750, 2018.

Lu, Y., Song, W., Lü, J., Chen, M., Su, Z., Zhang, X., and Li, H.: A pixel-based spectral matching method for mapping high-
595 resolution irrigated areas using EVI time series, Remote Sensing Letters, 12, 169-178, 2021.

McDermid, S., Nocco, M., Lawston-Parker, P., Keune, J., Pokhrel, Y., Jain, M., Jägermeyr, J., Brocca, L., Massari, C., Jones, A. D., Vahmani, P., Thiery, W., Yao, Y., Bell, A., Chen, L., Dorigo, W., Hanasaki, N., Jasechko, S., Lo, M.-H., Mahmood, R., Mishra, V., Mueller, N. D., Niyogi, D., Rabin, S. S., Sloat, L., Wada, Y., Zappa, L., Chen, F., Cook, B. I., Kim, H., Lombardozzi, D., Polcher, J., Ryu, D., Santanello, J., Satoh, Y., Seneviratne, S., Singh, D., and Yokohata, T.: Irrigation in the Earth system,
600 Nature Reviews Earth & Environment, 4, 435-453, 10.1038/s43017-023-00438-5, 2023.

Meier, J., Zabel, F., and Mauser, W.: A global approach to estimate irrigated areas – a comparison between different data and statistics, Hydrol. Earth Syst. Sci., 22, 1119-1133, 10.5194/hess-22-1119-2018, 2018.

Nabil, M., Zhang, M., Bofana, J., Wu, B., Stein, A., Dong, T., Zeng, H., and Shang, J.: Assessing factors impacting the spatial discrepancy of remote sensing based cropland products: A case study in Africa, Int. J. Appl. Earth Obs. Geoinf., 85, 102010,
605 2020.

Nagaraj, D., Proust, E., Todeschini, A., Rulli, M. C., and D'Odorico, P.: A new dataset of global irrigation areas from 2001 to 2015, Adv. Water Resour., 152, 103910, 10.1016/j.advwatres.2021.103910, 2021.

Pervez, M. S. and Brown, J. F.: Mapping Irrigated Lands at 250-m Scale by Merging MODIS Data and National Agricultural Statistics, Remote Sensing, 2, 2388-2412, 10.3390/rs2102388, 2010.

610 Puy, A., Borgonovo, E., Lo Piano, S., Levin, S. A., and Saltelli, A.: Irrigated areas drive irrigation water withdrawals, Nat Commun, 12, 4525, 10.1038/s41467-021-24508-8, 2021.

Puy, A., Sheikholeslami, R., Gupta, H. V., Hall, J. W., Lankford, B., Lo Piano, S., Meier, J., Pappenberger, F., Porporato, A., Vico, G., and Saltelli, A.: The delusive accuracy of global irrigation water withdrawal estimates, Nat Commun, 13, 3183, 10.1038/s41467-022-30731-8, 2022.

615 Salmon, J. M., Friedl, M. A., Frolking, S., Wisser, D., and Douglas, E. M.: Global rain-fed, irrigated, and paddy croplands: A new high resolution map derived from remote sensing, crop inventories and climate data, Int. J. Appl. Earth Obs. Geoinf., 38, 321-334, 10.1016/j.jag.2015.01.014, 2015.

Shahriar Pervez, M., Budde, M., and Rowland, J.: Mapping irrigated areas in Afghanistan over the past decade using MODIS

© Author(s) 2024. CC BY 4.0 License.

NDVI, RSEnv, 149, 155-165, 10.1016/j.rse.2014.04.008, 2014.

Siebert, S., Henrich, V., Frenken, K., and Burke, J.: Update of the digital global map of irrigation areas to version 5, Rheinische Friedrich-Wilhelms-Universität, Bonn, Germany and Food and Agriculture Organization of the United Nations, Rome, Italy, 2013.

Siebert, S., Kummu, M., Porkka, M., Döll, P., Ramankutty, N., and Scanlon, B. R.: A global data set of the extent of irrigated land from 1900 to 2005, HESS, 19, 1521-1545, 2015.

Teluguntla, P., Thenkabail, P., Oliphant, A., Gumma, M., Aneece, I., Foley, D., and McCormick, R.: Landsat-Derived Global Rainfed and Irrigated-Cropland Product 30 m V001 [dataset], https://doi.org/10.5067/Community/LGRIP/LGRIP30.001, 2023.

Thenkabail, P. S., Knox, J. W., Ozdogan, M., Gumma, M. K., Congalton, R. G., Wu, Z., Milesi, C., Finkral, A., Marshall, M., and Mariotto, I.: Assessing future risks to agricultural productivity, water resources and food security: How can remote sensing help?, PE&RS, Photogrammetric Engineering & Remote Sensing, 78, 773-782, 2012.

Thenkabail, P. S., Biradar, C. M., Noojipady, P., Dheeravath, V., Li, Y., Velpuri, M., Gumma, M., Gangalakunta, O. R. P., Turral, H., Cai, X., Vithanage, J., Schull, M. A., and Dutta, R.: Global irrigated area map (GIAM), derived from remote sensing, for the end of the last millennium, Int. J. Remote Sens., 30, 3679-3733, 10.1080/01431160802698919, 2009.

Tian, F., Wu, B., Zeng, H., Zhang, M., Zhu, W., Yan, N., and Lu, Y.: GMIE: a global maximum irrigation extent and irrigation type dataset derived through irrigation performance during drought stress and machine learning method (V2), Harvard Dataverse [dataset], doi:10.7910/DVN/HKBAQQ, 2023a.

Tian, F., Wu, B., Zeng, H., Zhang, M., Hu, Y., Xie, Y., Wen, C., Wang, Z., Qin, X., Han, W., and Yang, H.: A Shape-attention Pivot-Net for Identifying Central Pivot Irrigation Systems from Satellite Images using a Cloud Computing Platform: An application in the contiguous US, GIScience & Remote Sensing, 10.1080/15481603.2023.2165256, 2023b.

Waldner, F., De Abelleyra, D., Verón, S. R., Zhang, M., Wu, B., Plotnikov, D., Bartalev, S., Lavreniuk, M., Skakun, S., and Kussul, N.: Towards a set of agrosystem-specific cropland mapping methods to address the global cropland diversity, Int. J. Remote Sens., 37, 3196-3231, 2016.

Waller, P. and Yitayew, M.: Center Pivot Irrigation Systems, in: Irrigation and Drainage Engineering, edited by: Waller, P., and Yitayew, M., Springer International Publishing, Cham, 209-228, 10.1007/978-3-319-05699-9_12, 2016.

Wang, X., Muller, C., Elliot, J., Mueller, N. D., Ciais, P., Jagermeyr, J., Gerber, J., Dumas, P., Wang, C., Yang, H., Li, L., Deryng, D., Folberth, C., Liu, W., Makowski, D., Olin, S., Pugh, T. A. M., Reddy, A., Schmid, E., Jeong, S., Zhou, F., and Piao, S.: Global irrigation contribution to wheat and maize yield, Nat Commun, 12, 1235, 10.1038/s41467-021-21498-5, 2021.

Wriedt, G., Der Velde, M. V., Aloe, A., and Bouraoui, F.: A European irrigation map for spatially distributed agricultural modelling, Agricultural Water Management, 96, 771-789, 2009.

Wu, B., Tian, F., Zhang, M., Piao, S., Zeng, H., Zhu, W., Liu, J., Elnashar, A., and Lu, Y.: Quantifying global agricultural water appropriation with data derived from earth observations, Journal of Cleaner Production, 358, 131891, 10.1016/j.jclepro.2022.131891, 2022.

Wu, B., Gommes, R., Zhang, M., Zeng, H., Yan, N., Zou, W., Zheng, Y., Zhang, N., Chang, S., and Xing, Q.: Global Crop

https://doi.org/10.5194/essd-2023-536
Preprint. Discussion started: 6 February 2024
© Author(s) 2024. CC BY 4.0 License.

Monitoring: A Satellite-Based Hierarchical Approach, Remote Sensing, 7, 3907-3933, 2015.

Wu, B., Tian, F., Nabil, M., Bofana, J., Lu, Y., Elnashar, A., Beyene, A. N., Zhang, M., Zeng, H., and Zhu, W.: Mapping global maximum irrigation extent at 30m resolution using the irrigation performances under drought stress, Global Environmental Change, 79, 102652, 10.1016/j.gloenvcha.2023.102652, 2023a.

Wu, B., Zhang, M., Zeng, H., Tian, F., Potgieter, A. B., Qin, X., Yan, N., Chang, S., Zhao, Y., Dong, Q., Boken, V., Plotnikov, D., Guo, H., Wu, F., Zhao, H., Deronde, B., Tits, L., and Loupian, E.: Challenges and opportunities in remote sensing-based crop monitoring: a review, Natl Sci Rev, 10, nwac290, 10.1093/nsr/nwac290, 2023b.

Xiang, K., Ma, M., Liu, W., Dong, J., Zhu, X., and Yuan, W.: Mapping Irrigated Areas of Northeast China in Comparison to Natural Vegetation, Remote Sensing, 11, 825, 10.3390/rs11070825, 2019.

Xie, Y. and Lark, T. J.: Mapping annual irrigation from Landsat imagery and environmental variables across the conterminous United States, RSEnv, 260, 112445, 10.1016/j.rse.2021.112445, 2021.

Xie, Y., Gibbs, H. K., and Lark, T. J.: Landsat-based Irrigation Dataset (LANID): 30 m resolution maps of irrigation distribution, frequency, and change for the US, 1997–2017, Earth System Science Data, 13, 5689-5710, 2021.

Xie, Y., Lark, T. J., Brown, J. F., and Gibbs, H. K.: Mapping irrigated cropland extent across the conterminous United States at 30 m resolution using a semi-automatic training approach on Google Earth Engine, Int. J. Photogramm. Remote Sens., 155, 136-149, 10.1016/j.isprsjprs.2019.07.005, 2019.

Zajac, Z., Gomez, O., Gelati, E., van der Velde, M., Bassu, S., Ceglar, A., Chukaliev, O., Panarello, L., Koeble, R., van den Berg, M., Niemeyer, S., and Fumagalli, D.: Estimation of spatial distribution of irrigated crop areas in Europe for large-scale modelling applications, Agr Water Manage, 266, 107527, 10.1016/j.agwat.2022.107527, 2022.

Zhang, C., Dong, J., and Ge, Q.: Mapping 20 years of irrigated croplands in China using MODIS and statistics and existing irrigation products, Sci Data, 9, 407, 10.1038/s41597-022-01522-z, 2022a.

Zhang, C., Dong, J., and Ge, Q.: IrriMap_CN: Annual irrigation maps across China in 2000–2019 based on satellite observations, environmental variables, and machine learning, RSEnv, 280, 113184, 10.1016/j.rse.2022.113184, 2022b.

Zhang, C., Dong, J., and Ge, Q.: Mapping 20 years of irrigated croplands in China using MODIS and statistics and existing irrigation products, Scientific Data, 9, 407, 2022c.

Zhang, L., Zhang, K., Zhu, X., Chen, H., and Wang, W.: Integrating remote sensing, irrigation suitability and statistical data for irrigated cropland mapping over mainland China, JHyd, 613, 128413, 10.1016/j.jhydrol.2022.128413, 2022d.

Zhang, M., Wu, B., Zeng, H., He, G., Liu, C., Tao, S., Zhang, Q., Nabil, M., Tian, F., and Bofana, J.: GCI30: a global dataset of 30 m cropping intensity using multisource remote sensing imagery, Earth System Science Data, 13, 4799-4817, 2021.

Zhu, X., Zhu, W., Zhang, J., and Pan, Y.: Mapping irrigated areas in China from remote sensing and statistical data, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7, 4490-4504, 2014.

Zomer, R. J., Xu, J., and Trabucco, A.: Version 3 of the Global Aridity Index and Potential Evapotranspiration Database, Sci Data, 9, 409, 10.1038/s41597-022-01493-1, 2022.