 doi:10.20944/preprints202208.0112.v1

Data Descriptor

Collecting and Disseminating Ground Truth Data for Generation of Agriculture Statistics

Dan Tran-Thanh [1], Frank Greco Yrle [1,*], Chamaka Karunanayake [1], Kavinda Gunasekara [1], Sangita Dubey [2], Anthony Dean Burgard [2], and Manzul Kumar Hazarika [1]

[1] Asian Institute of Technology, Pathum Thani 12120, Thailand; ttdan@ait.ac.th (D.T.T.), kavinda@ait.ac.th (K.G.), chamakakarunanayake@gmail.com (C.K.), manzul@ait.ac.th (M.K.H.)
[2] Asian FAO Regional Office for Asia and the Pacific, Bangkok, 10200 Thailand; Sangita.Dubey@fao.org (D.S.), Anthony.Burgard@fao.org (A.D.B.)
[*] Correspondence: frankyrle@ait.ac.th

Abstract: Over the last few years, Earth Observation (EO) data has shifted towards increased use to produce official statistics, particularly in the agriculture sector. National statistics offices worldwide, including in Asia and the Pacific, are expanding their use of EO data to produce agricultural statistics such as crop classification, yield estimation, irrigation mapping, and crop loss estimation. The advances in image classification, such as pixel-based and phenology-based classifications, and machine learning create new opportunities for researchers to analyze EO data applied to agriculture statistics. However, it requires the ground truth (GT) data because classification result mainly depends on the quality of GT. Therefore, in this study, we introduced a random sampling approach to design and collect GT data using EO imagery and ancillary data. As a result of data collection, GT data improve the algorithms and validates classification results. Nevertheless, despite the importance of GT data, they are rarely disseminated as a data product in themselves. Thus, this results in an untapped opportunity to share GT data as a global public good, and improved use of survey and census data as a source of GT data.

Dataset: https://data.mendeley.com/datasets/xcyfctdg4v/1

Dataset License: Creative Commons Attribution 4.0 International.

Keywords: ground truth data; drone; mobile application; windshield survey; sample design; crop mapping; agriculture statistics; data dissemination; earth observation data; spatial database.

1. Summary (required)

Earth Observation (EO) plays a significant role in producing and helping official statistics complement traditional socio-economic and environmental data sources [1]. Since 1970, modern digital remote sensing has been used for agriculture, starting with a study of a rampant corn blight in the United States [2]. Soon after, the National Aeronautics and Space Administration (NASA) adopted the Multi-Spectral Scanner sensor (MSS) in 1972, marking the dawn of the Landsat program, one of the longest-running EO data collection platforms. The ongoing advancement of EO data, such as processing capability, availability, and quality (spatial and temporal resolution), makes it an attractive option for extracting crop information for agricultural management. In addition, agriculture remains one of the most researched EO data applications, with exponential growth occurring since 2013 [3]. Some of the most notable agricultural applications include crop areas and monitoring [4-8], yield estimates [9-11], irrigation mapping [12-14], and crop loss assessment [15-16].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 August 2022 doi:10.20944/preprints202208.0112.v1

Image classification is one of the most widely used remote sensing techniques for agriculture monitoring. Classification algorithms classically used in remote sensing include pixel-based unsupervised and supervised approaches. Output from both approaches is a single-layer thematic image with discrete classes representing features from the input raster data. Modern machine learning-based approaches, such as random forest [17-19], and deep neural network structures (convolutional neural network [CNN], deep neural network [DNN], recurrent neural network [RNN]) [20-22], are widely used for image classification. Besides, the satellite imagery's spatial and temporal resolution should always be considered when performing image classification for agriculture mapping. For example, Asia has typically small agricultural holdings with an average size of 1 ha, with an average of 3.2 parcels per holding. Considering the 1ha parcel size, GEOGLAM [23] recommends an EO data spatial resolution of 5-10m.

There is a problem when using EO data for agriculture applications is GT data. Whether traditional or modern methods are used for classification, the quality of GT data is paramount. GT data is in-situ data that verify, typically using sampling techniques, the commodities and land areas identified or estimated by the image processing algorithms. If GT data is not representative and accurate, the classification results will not be accurate for further analysis. Therefore, this study aims to share our experience in collecting and using GT data for the generation of agriculture statistics, and exploring opportunities to improve GT data collection, use, and exchange.

2. Data Description (required)

Because of agriculture statistics field application, agricultural land is mainly focused on and observed with three interested crops, including maize, sugarcane, and rice, in this GT database. This database was collected at field scale by our field survey experts in Nakhon Sawan province, Thailand, in 2022. As the fieldwork result, 156 points are collected (**Table 1**). Therein, there were 90 GT points visited in the field using the global positioning system (GPS) waypoints (therein 12 points of trees landcover class were collected as GT data (non-crop areas) for producing crop maps with the Sen2Agri platform [24]) and 66 GT points were also collected with the windshield survey data collection approach. Of the 156 points, 144 belonged to three crop classes: maize (31 points, 21.5%), sugarcane (28 points, 19.4%), and rice (66 points, 45.8%). Moreover, cassava is also observed when conducting fieldwork with 31 points, accounting for 21.5% of total collected GT points.

Table 1. Summary of GT data collection points (for crops only).

Crop Type	Points Visited	Windshield Survey Points*	Total	Percentage (%)
Sugarcane	18	10	28	19.4
Maize	11	20	31	21.5
Rice	39	27	66	45.8
Cassava	10	09	19	13.2
Total	78	66	144	100.0

*Crop fields estimated to be 1ha or larger were recorded for their crop type and location in the QField mobile application using the untether functionality. These points were entered while in the van enroute to GT point locations. The points from the windshield survey will be used as additional training data for the Sen2Agri platform.

According to GT database format, our database contains 156 records (as of June 2022) from visiting points in the fieldwork, is a geographic layer in shapefile format (point), and the geographic coordinate system is GCS_WGS_1984. In addition, each record corresponds to a point format with 10 attributes (**Table 2**). The spatial distribution and statistic results of GT points are presented in **Figure 1**.

Table 2. Description of point's attribute.

doi:10.20944/preprints202208.0112.v1

Name of Attribute	Type of Data	Description
ID	Numeric	Feature unique ID
Shape	Text	Format of the feature such as point, line, and polygon
LC	Text	Type of land cover of the point
Crop	Text	Type of crop of the point such as cassava, maize, rice, etc.
GT_point	Text	Name of GT point when collecting on the field
comments	Text	Condition of the particular crop on the field during the survey
Date	Date	Fieldwork date of GT data collection
X_coord	Numeric	X coordination of the point
Y_coord	Numeric	Y coordination of the point
intercrop	Text	Presence or absence of intercropping (single crop, mixed crop, etc.)

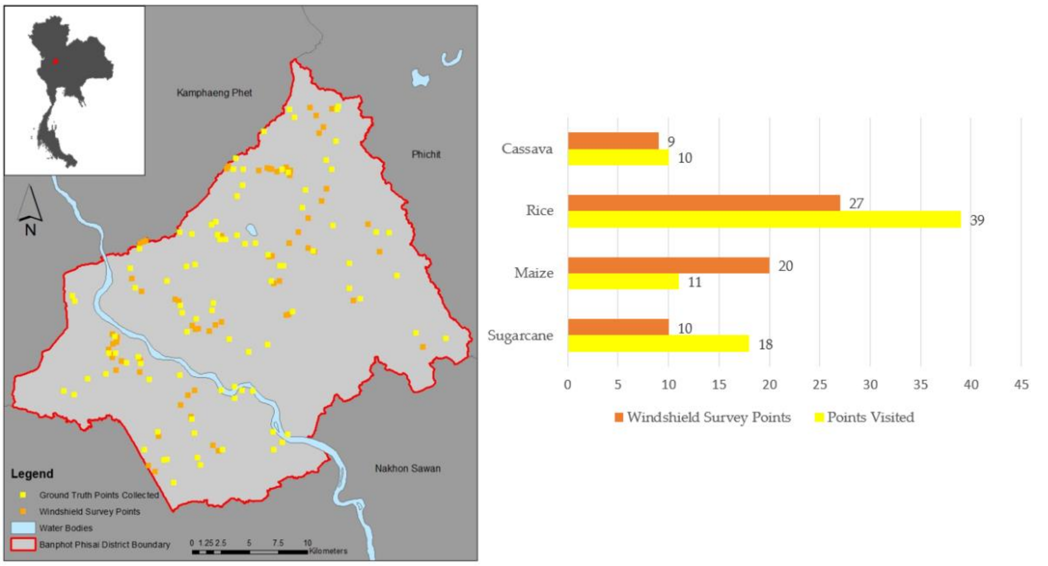

Figure 1. Map of Banphot Phisai district, Nakhon Sawan province, Thailand with GT data collection points. Yellow points are GT points visited and orange points are GT windshield survey points.

3. Methods (required)

3.1. Crop of Interest

Through the Office of Agricultural Economics (OAE) [25], we learned that they could use assistance in identifying a few major crops in Thailand with remote sensing techniques. OAE primarily used Sentinel-2 and Landsat 8 multispectral EO data for crop type mapping, which was both limited during the rainy season due to persistent cloud cover. Based on that, three primary crops were selected for this study: rice, maize, and sugarcane. Mapping results and GT data will be made widely available online.

3.2. Study Area Selection Process

Published statistics from the OAE [25] and the Office of Cane and Sugar Board (OCSB) [26] were used to gain an understanding of the distribution of the three crops of interest. From here, it was able to understand the distribution of crops at province and

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 August 2022 doi:10.20944/preprints202208.0112.v1

district levels. Districts with the largest agriculture area for all three selected crops were selected for this study. In some instances, there were districts that ranked high for one or two of the crops of interest but had especially low representation of the third crop. Based on the data from OAE and OCSB (**Table 3**), we selected Nakhon Sawan province (Banphot Phisai district) and Phitsanulok province (Phrom Phiram district), hereafter referred to as the study area. However, as of the time of writing this paper, data in Phitsanulok has not yet been collected but it will be added to our GT database later. Thus, GT data collection was conducted in Banphot Phisai district, Nakhon Sawan province, starting on 30 May and ending on 3 June 2022. **Figure 2** shows the location of the field survey area.

Table 3. Top 10 provinces with most selected crops planted by area in 2019.

No.	Province Name	Crop Type Area (km²)						Total Agriculture Area (km²)
		Sugarcane	%	Rice	%	Maize	%	
1	Nakhon Ratchasima	927.0	6.9	5761.4	43.2	1,064.9	8.0	13354.0
2	Phetchabun	795.0	0.1	1,987.6	38.1	142.1	2.7	5219.0
3	Chaiyaphum	814.0	17.7	2,903.3	54.6	135.6	2.6	5317.0
4	Lopburi	899.0	14.8	1252.6	35.7	630.2	17.9	3507.0
5	Nakhon Sawan	1,240.0	0.3	856.9	12.3	125.9	1.8	6986.0
6	Kamphaeng Phet	1,252.0	18.2	1,909.9	39.03	084.6	1.7	4893.0
7	Ubon Ratchathani	6.2	15.3	6513.3	76.4	116.8	1.4	8525.0
8	Si Sa Ket	17.5	15.2	4,816.6	74.3	50.9	0.8	6483.0
9	Khon Kaen	990.6	25.6	3,817.1	57.0	8.0	0.1	6692.0
10	Udon Thani	1,119.0	19.1	3,302.0	53.7	4.7	0.1	6148.0

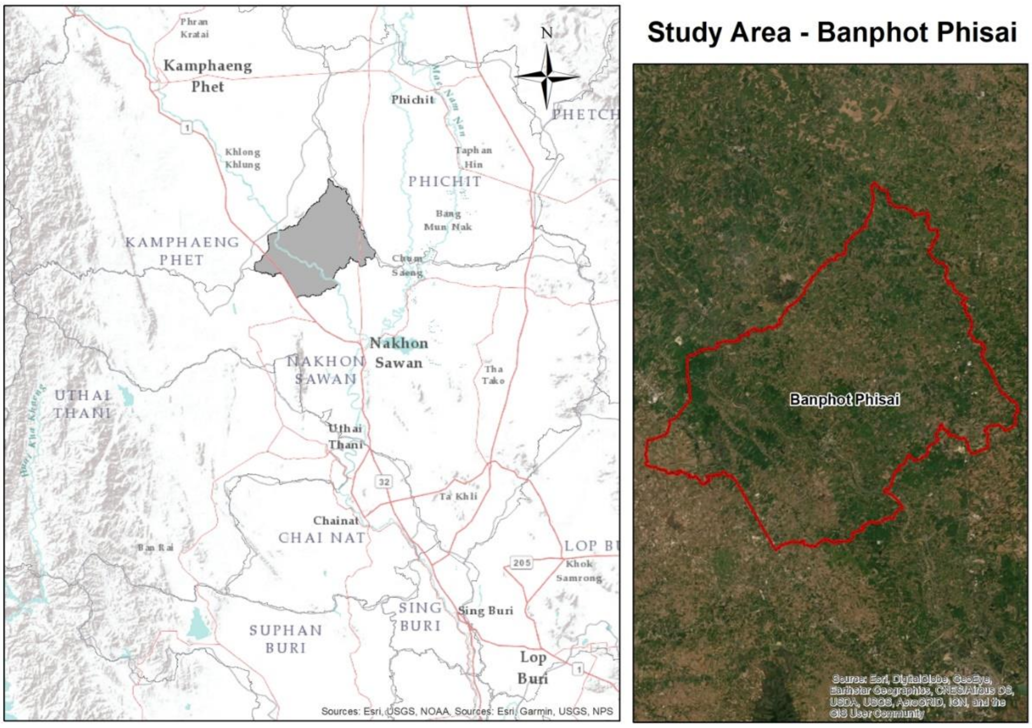

Figure 2. Map of the field survey area – Banphot Phisai, Nakhon Sawan province.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 August 2022 doi:10.20944/preprints202208.0112.v1

3.3. Sample Design

The sample design is essential in designing sampling for the field data collection. In this fieldwork, we used the random sampling method to identify survey locations with supporting data such as satellite imagery (Landsat 8 imagery, 30m spatial resolution) and ancillary data (road, administrative boundary). Sentinel-2 imagery with 10m spatial resolution was not selected due to a lack of data availability and cloud cover during the rainy season. There were two main steps to identify sample locations (points of data collection): (i) to use EO data to perform a binary classification to determine areas growing crops and non-crop growing areas, then identify crop types of areas growing crops using temporal NDVI profiles' plots; and (ii) to select data collection points using random sampling approach.

3.3.1. Crop Areas Mapping with Time-series NDVI of Landsat 8 Imagery

Landsat 8 imagery was downloaded from the USGS Earth Explorer data portal [27]. The website provides archive satellite imagery for various purposes such as land use mapping, change detection, and so on. We downloaded ten images in 2021 (January, February, March, May, June, August, October, November, and December) in good condition (clear and had less cloud cover) for producing a crop area map.

- Calculating Normalize Different Vegetation Index (NDVI) for each image using Red and Near-Infrared (NIR) bands and then creating the time-series Landsat 8 NDVI with ten bands (ten images). The formula to calculate NDVI [28] is as follows:

$$NDVI = \frac{NIR - Red}{NIR + Red} \tag{1}$$

  Therein, NIR is band 5, and Red is band 4 of the Landsat 8 image. The output values of NDVI are between -1 and 1. Low NDVI values indicate a low density of green vegetation, and high NDVI values indicate a high density of green vegetation.
- Time-series NDVI is classified to produce crop areas map using ISODATA unsupervised classification algorithm [29-30].
- A clustering algorithm was then applied to the crop area map to understand the unique classes of crops found in the field survey area using temporal NDVI profiles's plots [31]. Based on clustering output and image interpretation, three main classes were identified: agriculture (1-4), water (1-2), and built-up area. The crops type map presents in **Figure 3**.

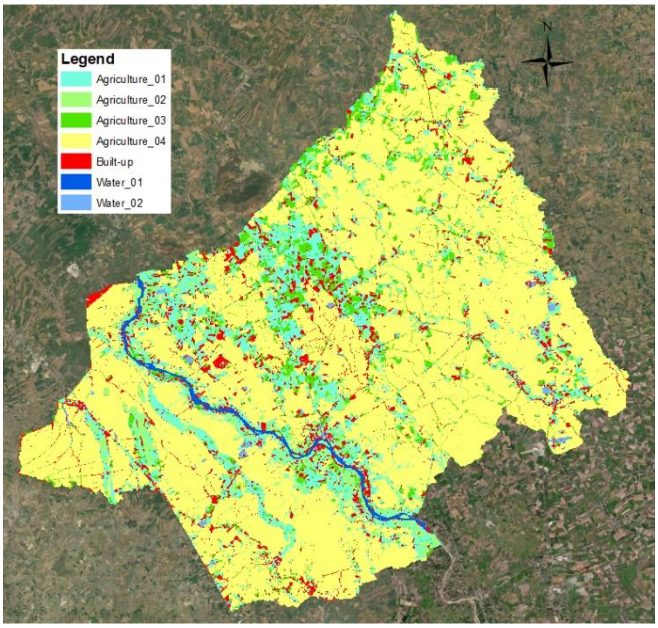

Figure 3. Crops type map in Banphot Phisai in 2021. The map is produced using the time series of NDVI with the ISODATA unsupervised classification algorithm.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 August 2022 doi:10.20944/preprints202208.0112.v1

3.3.2. Data Collection Points Selection

- A grid-based sampling scheme for the spatial design of the samples is designed [32]. Due to the spatial resolution of the Landsat image, a 1x1 km grid was created using a "Create Fishnet" tool in ArcMap [33] with the district administrative boundary map. As a result, 795 grids are created **(Figure 4A)**.
- The agriculture area (the result of section 3.3.1) within a grid is calculated. If the area is greater than 0.5 km², consider it a GT data collection point; otherwise, it will be removed (**Figure 4B**).

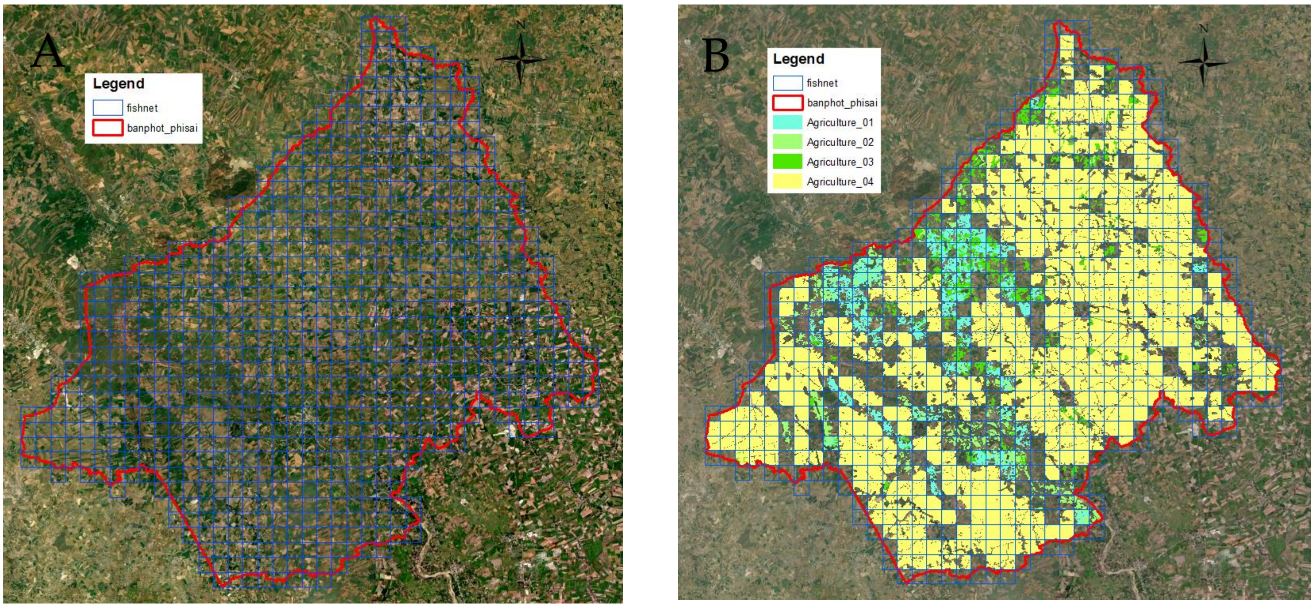

Figure 4. A: Creating a 1x1 km grid using Create Fishnet tool, and B: Selecting grids with a total area of agriculture within a grid greater than 0.5 km².

- Around 10% (79 grids) of the total grids above are selected using the "Subset Features" tool in ArcMap [34] to select random grids for data collection (**Figure 5**). In addition to selected grids, it was converted to point using ArcMap's "Feature to Point" tool [35]. Then, all points were loaded into the Google Maps – My Maps [36] application to navigate to the desired data collection location. All three data collection forms (section **3.4**) were used one after the other.

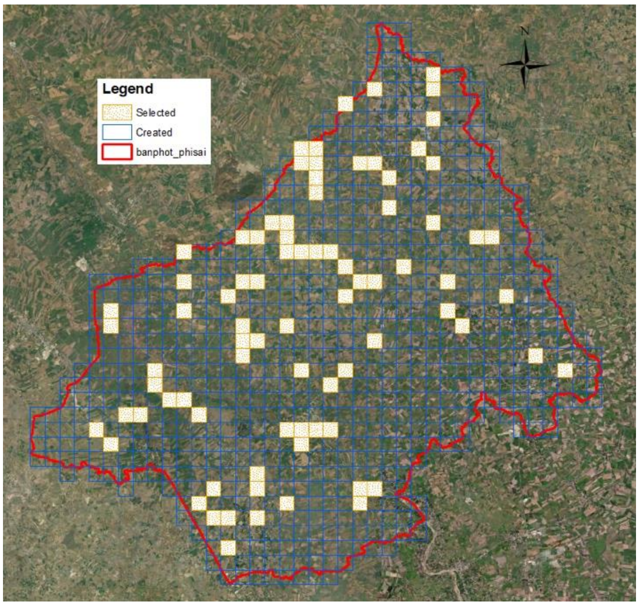

Figure 5. Data collection points selection in Banphot Phisai district. Yellow grids are selected grids using the "Subset Features" tool in ArcMap.

3.4. Methodology

This study was carried out with the combination of three approaches to GT data collection, including (i) manual data collection using a handheld GPS unit and photographs; (ii) data collection using QField mobile application [37] and drone; and (iii) data collection using Open Foris Collect Mobile application [38]. Those two mobile applications are selected after our desk review of current mobile applications for fieldwork. Details are in **Appendix A**.

The field visit was planned such that initially, the southern part of the district is covered and then gradually moves towards the northern part of Banphot Phisai. The first 1.5 days were used to collect data in the southern part of the district; then, the last 2.5 days were used for data collection in the central and northern sections of the district.

3.4.1. Manual Data Collection

- A reference data sheet was used to record all the field observations at GT sites (**Figure 5A, Appendix B**).
- The date, time, location ID (sample number) and the waypoint number were recorded on the data sheet.
- After reaching the desired location, the coordinates of the point were measured by using the handheld GPS unit (**Figure 5B & 5C**) and were noted down in the data sheet.
- The North direction was identified by using a compass (**Figure 5D**).
- Photographs were captured based on five main criteria: in North, East, South, and West direction; and at the location of the GT point (**Figure 6**).
- All photographs were captured with geotagging enabled.
- The photograph's photo number and direction were separately noted in the reference datasheet.
- The land cover and the crop type were recorded in the reference datasheet.

  Any remarkable visual observations of the field were noted down under the section "Field Notes" in the reference datasheet (**Figure 7**).

Figure 5. Manual data collection using GPS: A. Reference data sheet; B. Handheld GPS device; C. Using mapping grade GPS device; and D. Finding the North direction.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 August 2022 doi:10.20944/preprints202208.0112.v1

Figure 6. Capturing photographs of crops in different directions.

Figure 7. Filling out the reference datasheet.

3.4.2. Data Collection Using QField Mobile Application

QField is an open-source mobile data collection application from the makers of QGIS. It allows an extension of an existing QGIS project to be loaded to the mobile device and to bring relevant layers from the project to the field. In addition, the QField app can toggle raster and vector layers on or off to improve loading in the case of base maps or dense vector layers that obstruct other data layers. For GT data collection purposes, it is recommended to include offline shapefiles that can be edited in the QField app from the field. Vector entries from the field can be collected by tethering to the location of the mobile device or untethered to record the location away from a mobile device.

For the ground truthing data collection task in Banphot Phisai, an offline vector point layer was included in the QField project to record GT data. Features are saved following their creation in the QField application. In addition to visiting the randomly generated GT points, we also used the untether capability of QField to add additional points for crop fields that we passed by in what is referred to as the windshield survey. There were 66 GT points obtained in the windshield survey.

 doi:10.20944/preprints202208.0112.v1

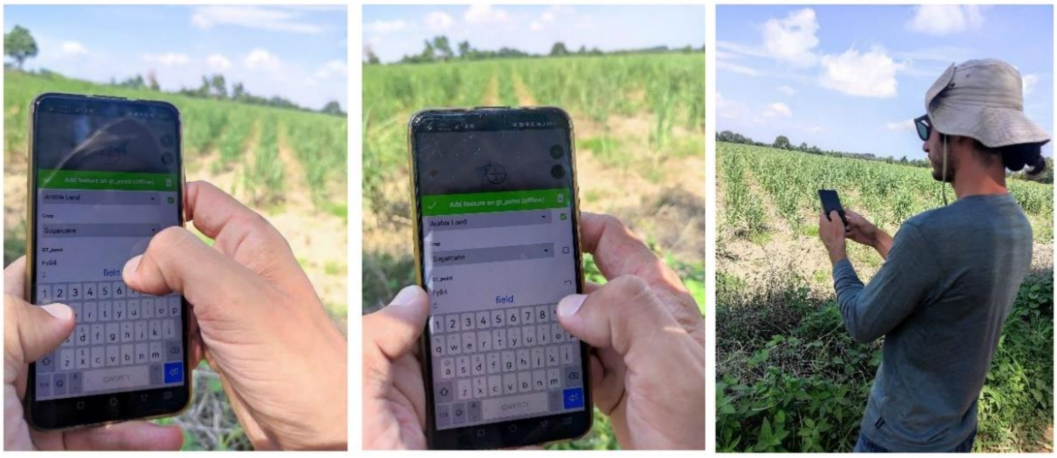

Figure 8. Field data collection with QField mobile application.

Besides, there were a few GT points which were inaccessible on foot. Lack of access was due to the environment, including dense vegetation blocking access, and dense tree line or vegetation with no suitable road for van access to. In these cases, we used the drone (DJI Mavic Pro 2) to remotely access these sites and obtain very high-resolution images, as seen in **Figure 9** below. The QField app's untether functionality was used to record the GT data remotely for the two inaccessible sites. This feature is not available in Open Foris Collect Mobile.

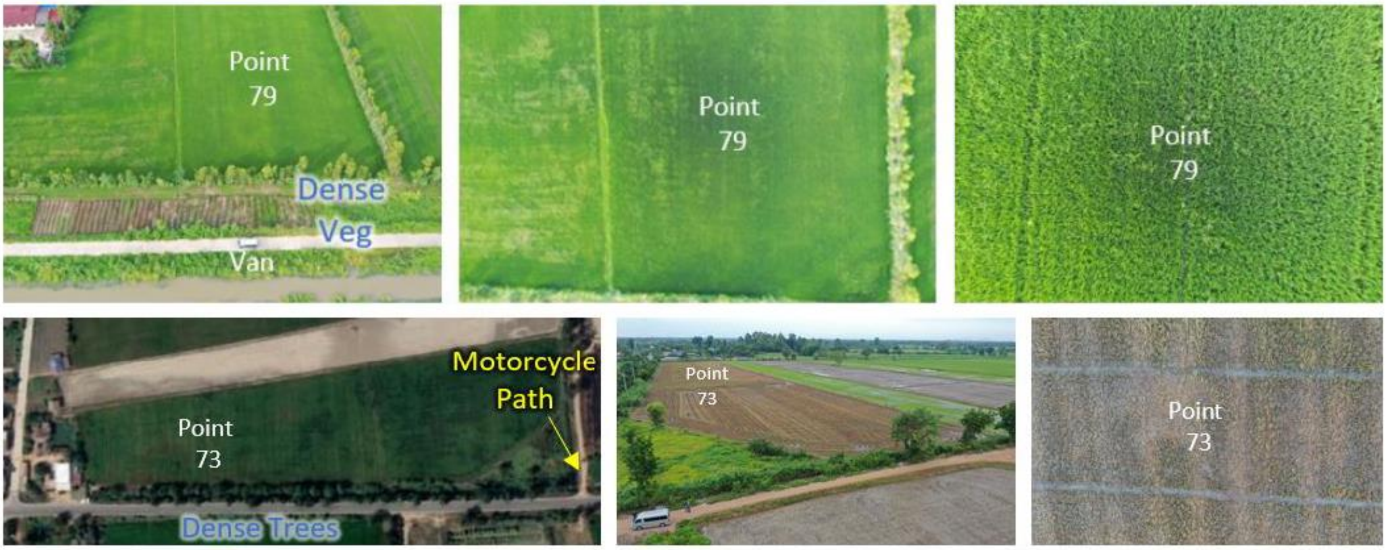

Figure 9. Remotely accessing GT data collection points with drone.

In addition, DJI Mavic Pro 2 drone was also flown at four flying heights: 90m (2.5cm/px), 60m (1.6cm/px), 30m (0.8cm/px), 10m (0.3cm/px) at each GT point location to obtain very high-resolution images of the GT location (**Figure 10**). Very high-resolution drone images are being used to relate to high-resolution satellite imagery. The final flying height of 10m was obtained to represent a near-field level view. Based on this result, for further use, it will be used as a reference to generate GT points using very high-resolution satellite imagery.

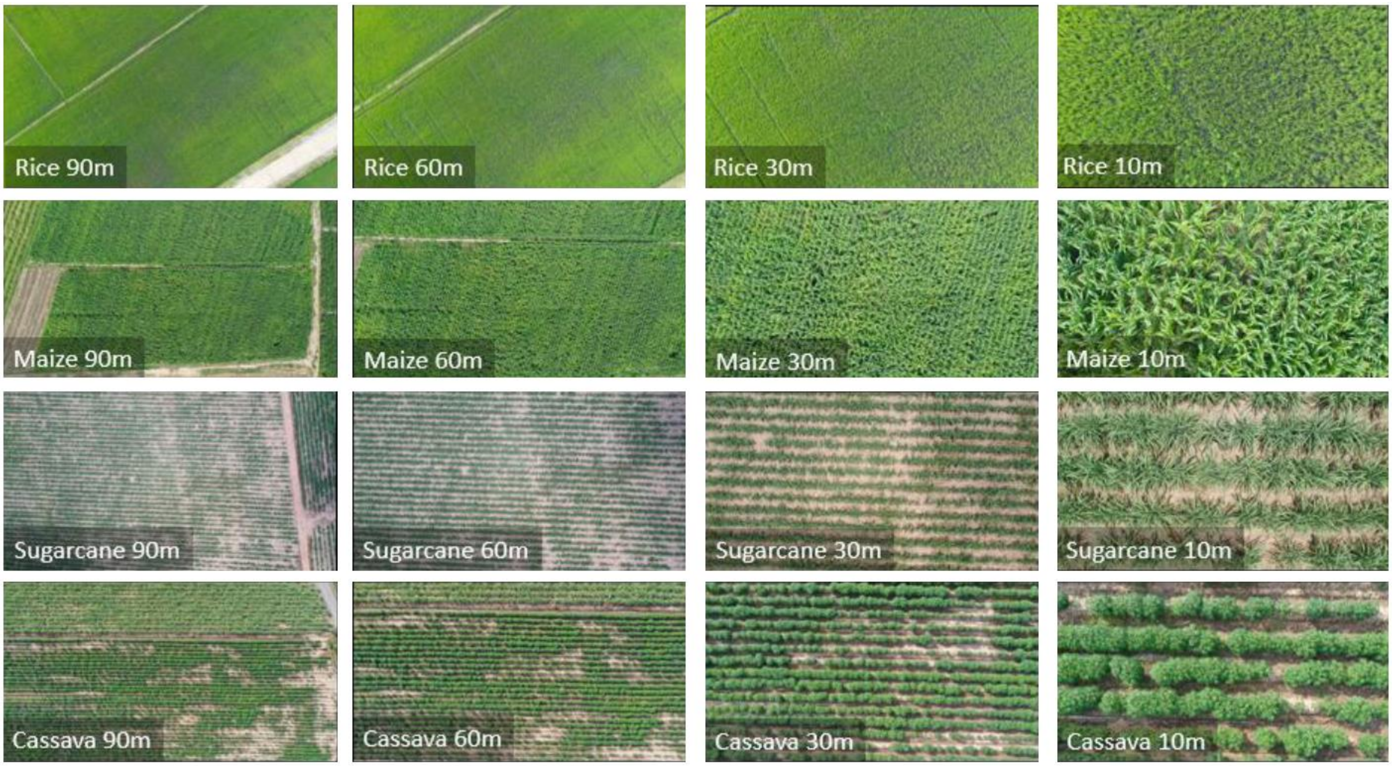

Figure 10. Very high-resolution images of crops enountered during GT data collection. Flying heights include: 90m (2.5cm/px), 60m (1.6cm/px), 30m (0.8 cm/px), 10m (0.3cm/px).

3.4.2. Data Collection Using Open Foris – Collect Mobile Application

- Initially Open Foris Collect was installed in the laptop.
- The "Survey Designer" tool in Open Foris Collect was then used to create the survey for field data collection (**Figure 11**).

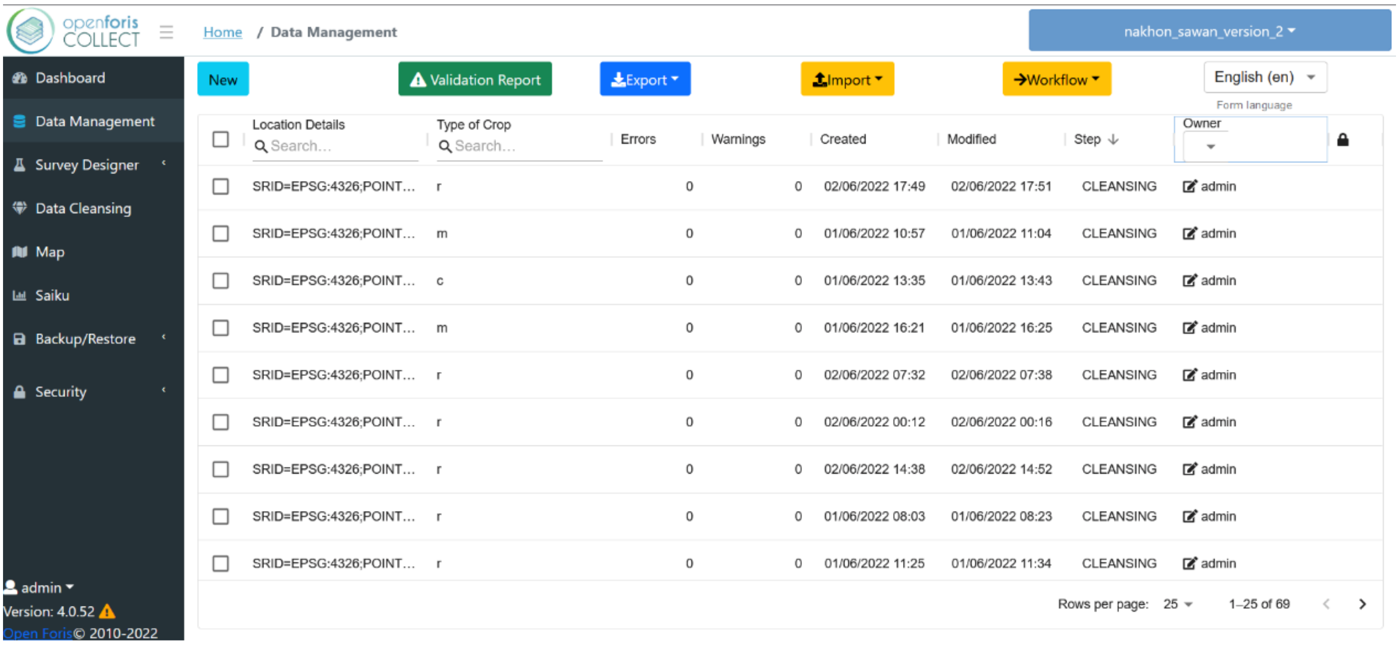

Figure 11. Open Foris Collect Data Management dashboard.

- The designed survey was then imported into Open Foris – Collect mobile application.
- The photographs in different directions were captured from the mobile phone.

- Data such as the location ID, date, time, GPS location, land cover type, and crop type were recorded in the mobile application (**Figures 12**).
- Then the photographs were imported into the mobile app.
- Finally, the field notes/remarks were recorded.

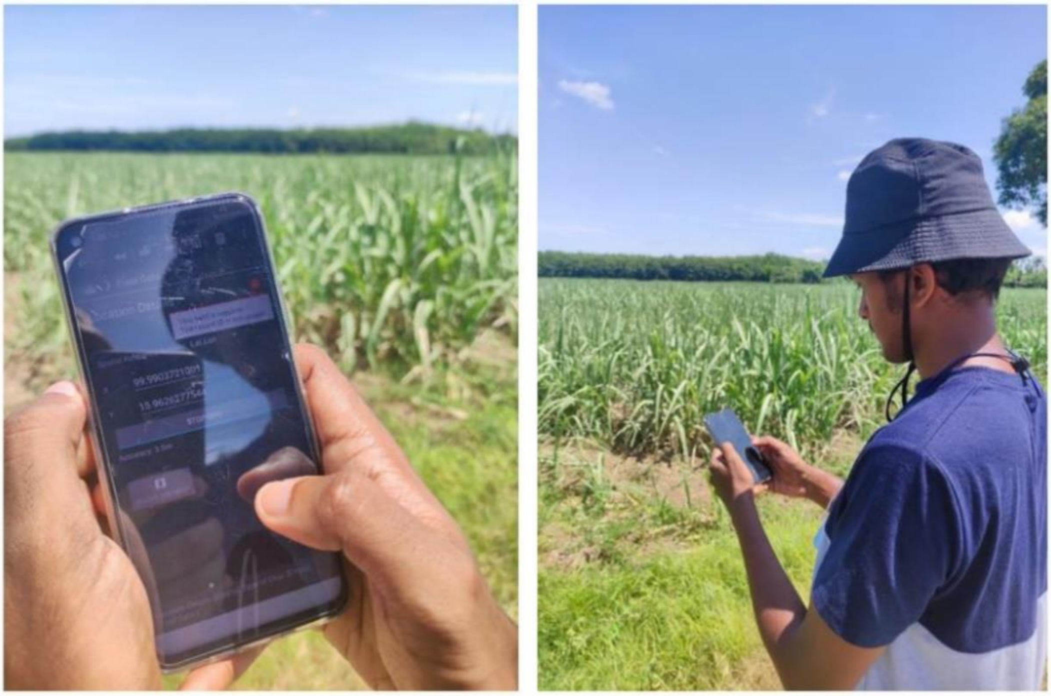

Figure 12. Using Open Foris Collect Mobile application in field.

3.5. GT Dataset Validation

Because the GT dataset is in-situ observation, it is already the GT data product. Therefore, it is impossible to validate it. Before the fieldwork, we very carefully prepared the work plan, technical guidelines, and checklist to ensure that nothing was forgotten during the fieldwork. Moreover, during the post-processing step, we carefully checked the collected GT data (waypoint collection and mobile apps) with very high-resolution satellite imagery (ArcMap base map) and the photos taken on the field to minimize errors. This step will contribute to the overall quality of the GT dataset.

4. Challenges Faced, and Steps Taken to Overcome Them

- The main challenge faced during field data collection was that it was impossible to reach out to that location. The exact location was generally in the middle of fields/crops, so it was impossible to reach the exact locations. In such cases, the handheld GPS, land photographs, and Open Foris - Collect mobile application were used from the nearest point that was accessible. In such situations, the Mavic pro drone was used to capture photographs of the exact location.
- There were incidents where the desired location was not physically accessible, and sometimes it was found inside the private property where it could not be accessed. In such cases, the point was shifted towards the nearest accessible location and was recorded as an alternative point.
- The fieldwork was obstructed by heavy rain in the late evening of day 2. To overcome the issues from rain and the time lost, field data collection was started much earlier on days 3, 4 and 5 compared to days 1 and 2.
- The Phantom 4 drone lost the connection with the controller because of the mountainous terrain in the area, and due to a lack of signal strength, the drone mission could not be completed. From the next drone mission, it was ensured that the drone

was set such that it had adequate signal strength and a good connection with the controller.
- There were occasions when road maintenance activities were being carried out; hence, even if it took more time, longer alternative roads were used to reach the desired locations.
- The drone batteries could not be used throughout the day without being charged. Therefore, steps were taken to get a power supply from the vehicle charger and to charge the drone batteries such that they would be sufficient to cover the day's work.
- Problem with exporting from QField – Data collected in the field remains saved within the mobile app but exporting the data does not update desktop QGIS with the new data collected in the field. Currently troubleshooting this issue.

Author Contributions: Conceptualization, D.T.-T. and F.G.Y.; Formal analysis, D.T.-T. and F.G.Y; Methodology, D.T.-T., F.G.Y, C.K, K.G., S.D., A.D.B. and M.K.H.; Supervision, D.T.-T., K.G. and S.D.; Visualization, D.T.-T., F.G.Y and C.K.; Writing—Original draft, D.T.-T. and F.G.Y; Writing—Review & editing, D.T.-T., F.G.Y, C.K, K.G., S.D., A.D.B. and M.K.H. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Not applicable

Data Availability Statement: the data of this study is available at https://data.mendeley.com/datasets/xcyfctdg4v/1

Conflicts of Interest: The authors declare no conflict of interest

Appendix A. Desk Review on Mobile Applications for Field Data Collection

To select a mobile application for GT data collection, we have reviewed 20 mobile applications. **Table A1** shows the details of mobile applications, including their ability to collect the GT data for this study.

Table A1. Keywords used for searching – Field data collection, Mobile Application, Free

No.	Name	Free/Paid	Supported Operating Systems	App creation date	Avail. of Map	Abil. to import photos	Abil. to record GPS data	Developer or or Organization	Additional Remarks
1	Open Data Kit [39]	Free/Paid	Android 5.0 & up	Last update on 07.06.22	No	Yes	Yes	Get ODK	Set up in web and import into mobile phone.
2	Kobo Toolbox [40]	Free	Android 5.0 & up	04.07.12	No	Yes	Yes	Funded by a combination of UN organizations, international humanitarian NGOs, and private foundations	Set up in web and import into mobile phone.
3	Open Foris -Collect [38]	Free	Android 4.1 & up	09.10.14	Yes (on Web)	Yes	Yes	Resource partners – FAO, Ministry of Foreign affairs of Finland, World Bank group etc.	Set up in web and import into mobile phone.
4	QField [37]	Free	Android 5.0 & up, IOS	15.06.15	Yes	Yes	Yes	Developed by OPENGIS.ch	Bugs while transferring field data back into laptop.
5	SW Maps [41]	Free	Android 4.4 & up	25.03.16	Yes	Yes	Yes	Offered by Softwel	Minor issues in adding photographs.
6	Fulcrum [42]	Free/Paid	Android 5.0 & up, IOS	31.07.12	Yes	Yes	Yes	Funded by Kennet and Kayne Anderson Capital Advisors, L.P.	Set up in web and import into mobile phone.
7	Collectral [43]	Free	Android 4.0 & up	02.06.18	No	Yes	No	Offered by Haykaz Kotanjyan	Create forms in Collectral.com

 doi:10.20944/preprints202208.0112.v1

8	Field Task Manager [44]	Free	Android 4.0 & up	18.02.14	No	Yes	No	Offered by ConnectMyWorld Technologies Pvt.Ltd	User customization not available. Linked to web.
9	ArcGIS Field Maps [45]	Free	Android 8.0 & up, IOS	01.11.20	Yes	Yes	Yes	Esri	Require ArcGIS organizational account.
10	GeoJot+ [46]	Free Trial	Android 4.0.3 & up, IOS	05.11.12	No	Yes	Yes	Offered by GeoSpatial Experts, Inc.	Not available freely. 15-day trial free
11	GIS Mobile Data Collection [47]	Free Trial	Android 5.1 & up	18.11.11	Yes	Yes	Yes	GISCloud	14-day free trial
12	Teamscope [48]	Free	Android 5.0 & up, IOS	05.11.20	No	No	No	Teamscope B.V	Designed for clinical data collection
13	RedCap [49]	Free	Android 5.1 & up	16.03.15	No	Yes	No	Offered by the REDCap team at Vanderbilt	Designed for medical sector
14	Magpi+ [50]	Beta version	Android 4.4 & up , IOS	20.02.18	Yes	Yes	Yes	Magpi	More biased onto medical sector.
15	SurveyCTO [51]	Free/Paid	Android 4.4 & up	19.09.13	No	Yes	Yes	Dobility, Inc.	Limitations on free version.
16	Commcare [52]	Free/Paid	Android 4.1 & up	22.06.12	No	Yes	Yes	Dimagi, Inc.	Focused on medical sector. Web based creation.
17	FastField Forms [53]	Free trial	Android 5.0 & up	28.05.15	No	Yes	No	Merge Mobile, Inc.	Designed for field inspections. 14-day trial available.
18	JotForm [55]	Free/Paid	Android 5.0 & up	16.06.19	No	Yes	Yes	Jotform	Online form creator, online payments, register people etc.
19	Uinta [55]	Free beta version	Android 7.0 & up	Last updated on 01.02.22	Yes	Yes	Yes	Juniper Systems, Inc.	A mapping tool. Website does not work.
20	Mapit GIS [56]	Free/Paid	Android 4.1 & up, IOS	08.11.14	Yes	No	Yes	Mapit GIS LTD	Web based access and can be customized by the mobile app.

Appendix B. Reference Datasheet

We have developed a reference datasheet to conduct the field survey using the manual data collection method. It was used during our study in Banphot Phisai district, Nakhon Sawan province, in June 2022. The datasheet is shown in **Table A2** below:

Table A2: Reference datasheet used for data collection in this study.

 doi:10.20944/preprints202208.0112.v1

Reference Datasheet

Thailand, Nakhon Sawan Province
Banphot Phisai District

Sample Number:

Observer's Initials:

Date:
Time:

GPS Reading:
X/LON:
Y/LAT:

Zone:

Crop Type:
Rice Cassava Other
Sugarcan Maize Soybean

Ground Photo

Photo Number	Direction

Land cover:
Arable Land Permanent Crops Bare
Settlement Earth Water Trees

Field Notes:

References

1. United Nations. *The Sustainable Development Goals Report 2020*. Available online: https://www.un.org/development/desa/publications/publication/sustainable-development-goals-report-2020 (accessed on 18 May 2021).
2. MacDonald, R. A. Summary of the History of the Development of Automated Remote Sensing for Agricultural Applications. *EEE Trans. Geosci. Remote Sens.* **1984**, GE-22 6, 473-481.
3. Weiss, M.; Jacob, F.; Duveiller, G. Remote Sensing for Agricultural Applications: A Meta-review. *Remote Sens. Environ.* **2020**, 236, 19. https://doi.org/10.1016/j.rse.2019.111402
4. Wang, Y.; Zhang, Z.; Zuo, L.; Wang, X.; Zhao, X.; Sun, F. Mapping Crop Distribution Patterns and Changes in China from 2000 to 2015 by Fusing Remote-Sensing, Statistics, and Knowledge-Based Crop Phenology. *Remote Sens.* **2022**, 14, 1800. https://doi.org/10.3390/rs14081800
5. Ren, T.; Xu, H.; Cai, X.; Yu, S.; Qi, J. Smallholder Crop Type Mapping and Rotation Monitoring in Mountainous Areas with Sentinel-1/2 Imagery. *Remote Sens.* **2022**, 14, 566. https://doi.org/10.3390/rs14030566
6. Asam, S.; Gessner, U.; Almengor González, R.; Wenzl, M.; Kriese, J.; Kuenzer, C. Mapping Crop Types of Germany by Combining Temporal Statistical Metrics of Sentinel-1 and Sentinel-2 Time Series with LPIS Data. *Remote Sens.* **2022**, 14, 2981. https://doi.org/10.3390/rs14132981
7. Jayanth, J.; Aravind, R.; Amulya, C.M. Classification of Crops and Crop Rotation Using Remote Sensing and GIS-Based Approach: A Case Study of Doddakawalande Hobli, Nanjangudu Taluk. *J. Indian Soc. Remote Sens.* **2022**, 50, 197–215. https://doi.org/10.1007/s12524-020-01296-0
8. Luo, Y.; Zhang, Z.; Zhang, L.; Han, J.; Cao, J.; Zhang, J. Developing High-Resolution Crop Maps for Major Crops in the European Union Based on Transductive Transfer Learning and Limited Ground Data. *Remote Sens.* **2022**, 14, 1809. https://doi.org/10.3390/rs14081809
9. Vallentin, C., Harfenmeister, K., Itzerott, S.; Kleinschmit, B.; Conrad, C.; Spengler, D. Suitability of Satellite Remote Sensing Data for Yield Estimation in Northeast Germany. *Precis. Agric.* **2022**, 23, 52–82. https://doi.org/10.1007/s11119-021-09827-6
10. Basso, B.; Liu, L. Seasonal Crop Yield Forecast: Methods, Applications, and Accuracies. *Adv. Agron.* **2019**, 154, 201-255. https://doi.org/10.1016/bs.agron.2018.11.002
11. Prasad, K. A.; Chai, L.; Singh, P. R.; Kafatos, M. Crop Yield Estimation Model for Iowa Using Remote Sensing and Surface Parameters. *Int. J. Appl. Earth Obs. Geoinf.* **2006**, 8 1, 26-33. https://doi.org/10.1016/j.jag.2005.06.002
12. Bastiaanssen, G. M. W.; Molden, J. D.; Makin, W. I. Remote Sensing for Irrigated Agriculture: Examples from Research and Possible Applications. *Agric. Water Manag.* **2000**, 46 2, 137-155 https://doi.org/10.1016/S0378-3774(00)00080-9.
13. Ozdogan, M.; Yang, Y.; Allez, G.; Cervantes, C. Remote Sensing of Irrigated Agriculture: Opportunities and Challenges. *Remote Sens.* **2010**, 2, 2274-2304. https://doi.org/10.3390/rs2092274
14. Ambika, A., Wardlow, B. & Mishra, V. Remotely Sensed High-Resolution Irrigated Area Mapping in India for 2000 to 2015. *Sci. Data.* **2016**, 3, 160118. https://doi.org/10.1038/sdata.2016.118
15. Sawant, S.; Mohite, J.; Sakkan, M.; Pappula, S. Near Real-Time Crop Loss Estimation using Remote Sensing Observations. In Proceedings of the 8th International Conference on Agro-Geoinformatics, Istanbul, Turkey, 16-19 July 2019. doi: 10.1109/Agro-Geoinformatics.2019.8820217.
16. Rahman, M.S.; Di, L.; Yu, E. Remote Sensing Based Rapid Assessment of Flood Crop Damage Using Novel Disaster Vegetation Damage Index (DVDI). *Int. J. Disaster Risk Sci.* **2021**, 12, 90–110. https://doi.org/10.1007/s13753-020-00305-7

17. Prodhan, F.A.; Zhang, J.; Yao, F.; Shi, L.; Pangali Sharma, T.P.; Zhang, D.; Cao, D.; Zheng, M.; Ahmed, N.; Mohana, H.P. Deep Learning for Monitoring Agricultural Drought in South Asia Using Remote Sensing Data. *Remote Sens.* **2021**, *13*, 1715. https://doi.org/10.3390/rs13091715
18. Ok, O. A.; Akar, O.; Gungor, O. Evaluation of Random Forest Method for Agricultural Crop Classification. Eur. J. Remote. Sens. **2012**, 45 1, 421-432. https://doi.org/10.5721/EuJRS20124535
19. Gumma, K. M.; Thenkabail, S. P.; Teluguntla, G. P.; Oliphant, A.; Xiong, J.; Giri, C.; Pyla, V.; Dixit, S.; Whitbread, M. A. Agricultural Cropland Extent and Areas of South Asia Derived using Landsat Satellite 30-m Time-series Big-data Using Random Forest Machine Learning Algorithms on the Google Earth Engine Cloud. *GIsci Remote Sens.* **2019**, 57 3, 302-322. https://doi.org/10.1080/15481603.2019.1690780
20. Mortensen, A. K.; Dyrmann, M.; Karstoft, H.; Jørgensen, R. N.; Gislum, R. Semantic Segmentation of Mixed Crops Using Deep Convolutional Neural Network. In Proceedings of the International Conference of Agricultural Engineering (CIGR), Aarhus, Denmark 26 - 29 June 2016.
21. Di Cicco, M.; Potena, C.; Grisetti, G.; Pretto, A. Automatic Model-Based Dataset Generation for Fast and Accurate Crop and Weeds Detection. In Proceedings of International Conference on Intelligent Robots and Systems (IROS), Vancouver, BC, Canada, 24–28 September 2017.
22. Yao, J.; Wu, J.; Xiao, C.; Zhang, Z.; Li, J. The Classification Method Study of Crops Remote Sensing with Deep Learning, Machine Learning, and Google Earth Engine. *Remote Sens.* **2022**, *14*, 2758. https://doi.org/10.3390/rs14122758
23. GEOGLAM (Group on Earth Observations Global Agricultural Monitoring Initiative). Available online: https://earthobservations.org/geoglam.php (accessed on 28 June 2022).
24. Sen2Agri. Available online: http://www.esa-sen2agri.org (accessed on 19 February 2022).
25. OAE (Office of Agricultural Economics). Available online: https://www.oae.go.th/view/1/Home/EN-US (accessed on 28 June 2022).
26. OCSB (Office of the Cane and Sugar Board). Available online: http://www.ocsb.go.th/upload/journal/fileupload/923-9200.pdf (accessed on 28 June 2022).
27. Earth explorer. The USGS Earth Explorer Data Portal. Available online: https://earthexplorer.usgs.gov/ (accessed on 27 May 2022).
28. Rouse, J.; Haas, R. H. J.; Schell, J. A.; Deering, D. W. Monitoring Vegetation Systems in the GreatPlains with ERTS. In Proceedings of the Third Symposium of ERTS. Greenbelt, Maryland, USA: NASA, 1974.
29. Ball, G. H.; Hall, D. J. *ISODATA, a novel method of data analysis and pattern classification*. Stanford Research Institute, Menlo Park, 1974, AD0699616
30. Nguyen, T. T. H.; De Bie, C. A. J. M.; Ali, A.; Smaling, E. M. A.; Chu, T.H. Mapping the Irrigated Rice Cropping Patterns of the Mekong Delta, Vietnam, through Hyper-temporal SPOT NDVI Image Analysis. *Int. J. Remote Sens.* **2012**, 33, 415–434. https://doi.org/10.1080/01431161.2010.532826
31. Bellón, B.; Bégué, A.; Lo Seen, D.; De Almeida, C.A.; Simões, M. A Remote Sensing Approach for Regional-Scale Mapping of Agricultural Land-Use Systems Based on NDVI Time Series. Remote Sens. **2017**, 9, 600. https://doi.org/10.3390/rs9060600
32. Crisp, M.P.; Jaksa, M.B.; Kuo, Y.L. Framework for the Optimisation of Site Investigations for Pile Designs in Complex Multi-Layered Soil. 2019, Res. Rep. Sch. Civil Environ. Min. Eng.
33. ESRI. Create Fishnet (Data Management)—ArcMap Documentation. Available online: https://desktop.arcgis.com/en/arcmap/latest/tools/data-management-toolbox/create-fishnet.htm (accessed on 30 June 2022).
34. ESRI. Subset Features (Geostatistical Analyst)—ArcMap. Available online: https://desktop.arcgis.com/en/arcmap/latest/tools/geostatistical-analyst-toolbox/subset-features.htm (accessed on 30 June 2022).
35. ESRI. Feature To Point (Data Management)—ArcMap Documentation. Available online: https://desktop.arcgis.com/en/arcmap/latest/tools/data-management-toolbox/feature-to-point.htm (accessed on 30 June 2022).
36. Google My Maps. Available online: https://mymaps.google.com (accessed on 30 June 2022).
37. Qfield. QField - Efficient fieldwork built for QGIS. Available online: https://qfield.org/ (accessed on 01 July 2022).
38. Open Foris Collect Mobile. Available online: https://openforis.org/ (accessed on 01 July 2022).
39. Open Data Kit. ODK - Collect data anywhere. Available online: https://getodk.org/ (accessed on 01 July 2022).
40. Kobo Toolbox. KoBoToolbox | Data Collection Tools for Challenging Environments. Available online: https://www.kobotoolbox.org/ (accessed on 01 July 2022).
41. SW Maps. SW Maps - GIS & Data Collector. Available online: https://play.google.com/store/apps/details?id=np.com.softwel.swmaps&hl=en&gl=US (accessed on 01 July 2022).
42. Fulcrum. Available online: https://www.fulcrumapp.com/ (accessed on 01 July 2022).
43. Collectral. Available online: https://collectral.com/ (accessed on 01 July 2022).
44. Field Task Manager. Available online: https://www.connectmyworld.in/mobile-data-collection-app-for-filed-staffs/ (accessed on 01 July 2022).
45. ESRI. ArcGIS Field Maps Field Apps & Mobile Data Collection. Available online: https://www.esri.com/en-us/arcgis/products/arcgis-field-maps/overview (accessed on 01 July 2022).
46. GeoJot+. GeoJot+ Field Data Collection System. Available online: http://www.geospatialexperts.com/geojot/ (accessed on 01 July 2022).

47. GIS Mobile Data Collection. Available online: https://www.giscloud.com/apps/mobile-data-collection/ (accessed on 01 July 2022).
48. Teamscope. Teamscope: Data Collection App for Research Secure, Offline. Available online: https://www.teamscopeapp.com/ (accessed on 01 July 2022).
49. RedCap. Available online: https://projectredcap.org/software/mobile-app/ (accessed on 01 July 2022).
50. Magpi+. Magpi: The Best Mobile Data Collection Software. Available online: https://www.magpi.com/ (accessed on 01 July 2022).
51. SurveyCTO. SurveyCTO: Because Your Data Is Worth It. Available online: https://www.surveycto.com/ (accessed on 01 July 2022).
52. Commcare. CommCare by Dimagi Data Collection App. Available online: https://www.dimagi.com/commcare/ (accessed on 01 July 2022).
53. FastField Forms. FastField Forms | Mobile Data Collection and Analytics. Available online: https://www.fastfieldforms.com/ (accessed on 01 July 2022).
54. JotForm. Jotform: Free Online Form Builder & Form Creator. Available online: https://www.jotform.com/ (accessed on 01 July 2022).
55. Uinta. Available online: https://play.google.com/store/apps/details?id=com.junipersys.uinta&hl=en&gl=US (accessed on 01 July 2022).
56. Mapit GIS. mapitGIS: Site Content. Available online: https://mapitgis.com (accessed on 01 July 2022).