EOSTAT CropMapper – Technical Report

Version: October 26, 2022

<u>AUTHORS</u>

Lorenzo De Simone, PhD

Technical Adviser (Geospatial)

Office of Chief Statistician (OCS)

Viale delle Terme di Caracalla

00153, Rome, Italy

Phone: +390657053808, ext: 53808

Lorenzo.DeSimone@fao.org

Food and Agriculture Organization of the United Nations

Silvan Ragettli, PhD

Tobias Siegfried, PhD

hydrosolutions ltd.

Winkelriedstrasse 5

CH-8006 Zurich, Switzerland

Phone: +41 43 535 05 80

ragettli@hydrosolutions.ch and www.hydrosolutions.ch

Contents

List of Figures

Supplementary Figures

List of Tables

Supplementary Tables

1 Introduction

FAO in partnership with hydrosolutions GmbH have developed a web-based solution for country-scale crop monitoring using earth observations to support National Statistics Offices (NSO's) in the production of official statistics using alternative data sources. The tool is called EOSTAT CropMapper, and has been first deployed and tested in Afghanistan in collaboration with the National Statistic and Information Authority (NSIA). The present document provides an overview of its functionalities, relevant workflows and technical background.

The aim of developing the EOSTAT CropMapper is to facilitate the use of the Earth Observations (EO) data and geospatial technology for improved agriculture monitoring and production of official agricultural statistics.

The unprecedented availability of free and open EO data stemming from the Sentinel fleet and the Landsat 8 and the long-term archives, and accessibility to free and low cost cloud computing provide the ideal conditions implementing scalable solutions that can be used in operational contexts. However, the uptake of EO data in NSO's is still limited, especially in developing countries. Only few NSO's globally are using such solution (e.g., Canada, Poland). The main reason for this is the common lack in countries of sufficient and high quality of in-situ data which is required to provide ground truth information for the training of the classification algorithms and for validation of crop maps.

In this context FAO is providing a solution to solve the issue by developing a system that is capable to map the main crop types in a given country or region, with a minimum mapping Unit of 0,01 ha. The main crop types are considered per region, which are defined by the NSO and in general are covering each a minimum area of 5 % of the annual cropland in the region and are representing a cumulated area higher than 75 % of the latter.

The solution provided by FAO can cope with the different availability of in-situ data according to three different scenarios (Figure 1), as listed below:

- **Scenario 1**: a large and accurate in-situ data is available.
  The system relies on a traditional Random Forest classifier.

- **Scenario 2**: a limited amount of in-situ data is available.
  The system relies on the use of a Dynamic Time Warping (DTW) algorithm to classify pixels into crop types based on only a few reference samples per crop type that represent the characteristic phenologies .

- **Scenario 3**: no in-situ data is available.
  The system relies on K-means clustering to map clusters of crop pixels. Subsequently the user is requested to associate each cluster to a crop label based on his expert knowledge.

The EOSTAT CropMapper allows to produce crop type maps bi-annually (summer harvest and/or autumn harvest crops). The tool has been implemented as a Google Earth Engine (GEE) app and relies on free Sentinel-1 and Sentinel-2 satellite imagery, that enable the generation of crop maps at a spatial resolution of 10 meters.

The tool assists the operator through a graphical user interface, offering access to spectral signatures and high-resolution images to steer his decisions. The tool is currently configured to operate in Afghanistan and can be re-configured to serve any other geographic areas.

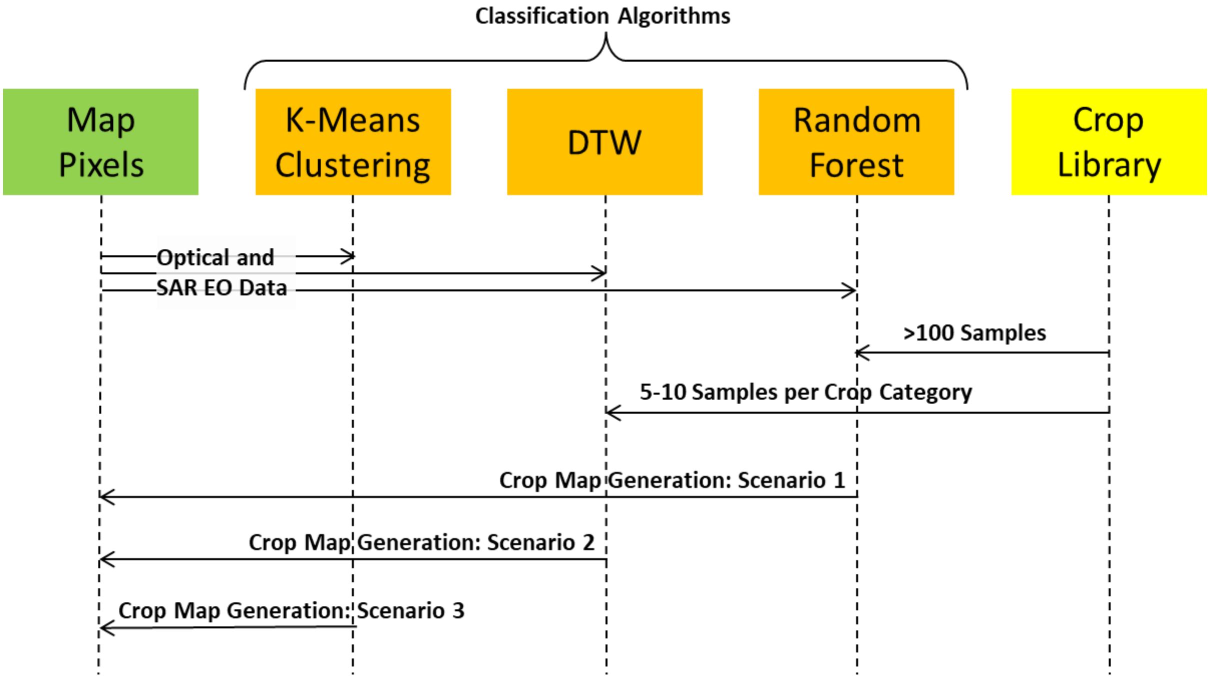

Figure 1. Sequence diagram of the three scenarios for crop map generation. The scenarios depend on the availability of crop type samples from the crop signature library.

2 EOSTAT CropMapper

The **EOSTAT CropMapper suite** is composed of **two tools** and one Crop Library

- the **Administrator tool** for trusted users to perform back-end tasks
- the **End-user tool** for viewing, analyzing, and downloading the maps.
- the **crop signatures library**

2.1 Administrator Tool

https://ocsgeospatial.users.earthengine.app/view/eostat-afghanistan-admin

The main purpose of the EOSTAT CropMapper administrator tool is the management of the built-in library of crop signatures. Trusted users can access the tool with their GEE login. New in-situ samples can be uploaded as point shapefiles to a dedicated cloud asset folder in GEE. The tool can then be used for the following main tasks:

1. Integration of new in-situ samples into the built-in library of crop signatures.
2. Validation of new in-situ samples: verification based on comparison with typical NDVI signatures of a given crop and based on up-to-date Sentinel-2 and Sentinel-1 imagery.
3. Definition and addition of new samples directly in the tool.
4. Modification of existing sample attributes and their coordinates.

5. Export of new crop maps with the selected library for selected spatial units (defined based on political boundaries and/or agro-ecological zones). Exported crop maps then become available for analysis in the front-end application.

The administrator tool is available as a GEE-App for demonstration purposes (Figure 2. The link to the app is provided above). Note that no administrator tasks can be carried out with the GEE-App version of the tool. For the modification of assets in GEE or on GCS, login into the GEE console is required (https://code.earthengine.google.com/) and a dedicated GEE script needs to be shared with the administrator. However, no programming skills are required from administrators. The interface of the tool in the GEE console is identical to the interface in the demonstration app.

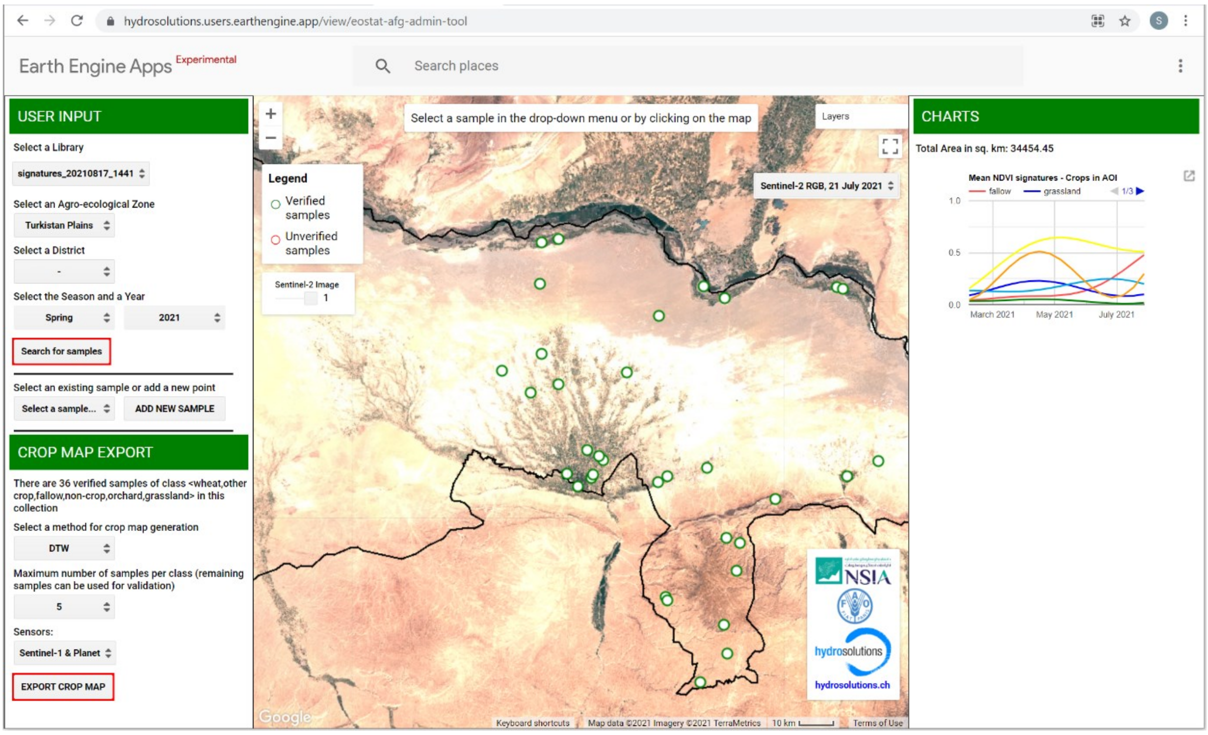

Figure 2. Screenshot of the EOSTAT CropMapper Administrator Tool.

2.2 Front-end Application

https://ocsgeospatial.users.earthengine.app/view/eostat-afghanistan

The front-end application is available to all users with access to the dedicated URL of the tool (the link to the demonstration version for Afghanistan is provided above). The main tasks that can be carried out with the tool are the following:

- Visualization of crop maps, generated from different combinations of sensors (currently Sentinel-1 & Sentinel-2) and with two available methods (DTW and Random Forest).
- Visualization of optical satellite imagery (Sentinel-2)
- Calculation of crop statistics per selected area of interest
- Downloading of crop maps to local drives
- Classification accuracy assessment

The front-end application has reading access to the library of crop signatures that was used for the generation of each crop map. This offers great flexibility with respect to the types of crops and the total number of crop classes that can be displayed and analyzed. It will be automatically noted if the administrator uses samples of an additional crop for the generation of a crop map. An additional label will be added to the interface (lower-left corner, see Figure 3) and the corresponding cropping areas will be calculated (bar-plot and table under OUTPUTS, Figure 3). Moreover, the producers and Users classification accuracy are calculated for each individual crop for which validation data points are available. For this purpose, the administrator can modify the maximum number of samples per crop class that are used by the supervised classification algorithms. Any additional samples will then be automatically available for validation in the end-user tool.

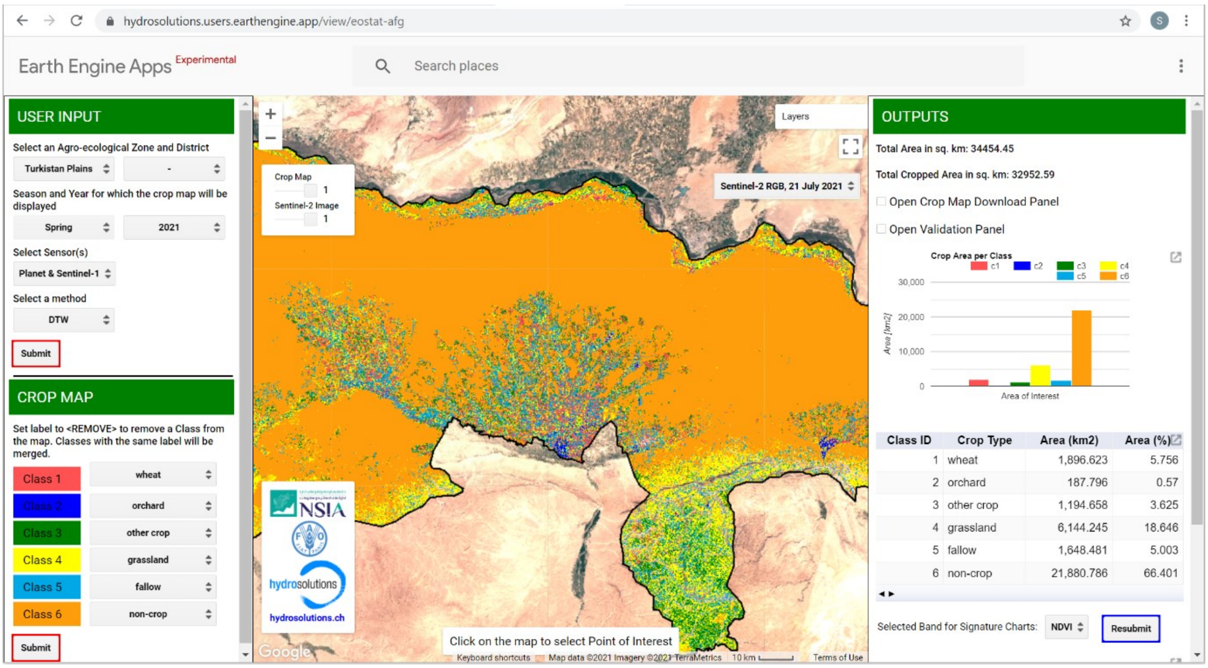

Figure 3. Screenshot of the EOSTAT CropMapper Front-end Application.

2.3 Crop Library

The core element of the EOSTAT CropMapper is the crop library. This contains the pheno-spectral characteristics of crops and is used as a reference by the Random Forest and DTW classifiers in Scenario 1 and 2 (Figure 1). Such characteristics are provided by the multitemporal optical and SAR EO data provided by various satellites (see Table 3).

To generate labelled crop maps, it is recommended to gather at least five georeferenced samples per crop type, year, and agro-ecological zone available. Such a collection of spatial data points and their attributes (Table 1) needs to be provided by the user. To achieve high accuracy in crop type classification, it is of paramount importance that the in-situ data are gathered with very high geospatial accuracy and correct identification in the field.

All management tasks related to this library of crop signatures can then be carried out within the administrator tool. Administrators are required to verify all uploaded in-situ samples in this tool by visually inspecting their location on optical satellite images and by comparing their NDVI signature to the average signature of a given class (see workflow depicted in Figure 4). Finally, the administrator tool is also used to trigger the generation of new crop maps when the crop library is ready. Administrators can choose among options to decide which EO data sources should be considered for crop map generation and can modify the time window used for the analysis.

In the context of the pilot project deployed in Afghanistan, the crop library has been built using the in-situ data provided by NSIA for wheat, maize and cotton.

Table 1. Summary of attributes required for each in-situ sample. More details are provided in the EOSTAT CropMapper User Guide (FAO and Hydrosolutions Ltd., 2022).

Attribute	Description	Attribute Type
Crop Name	Crop type class attribute	String
Year	Year of observation	Number
Period	Validity period of observation	Pre-defined string

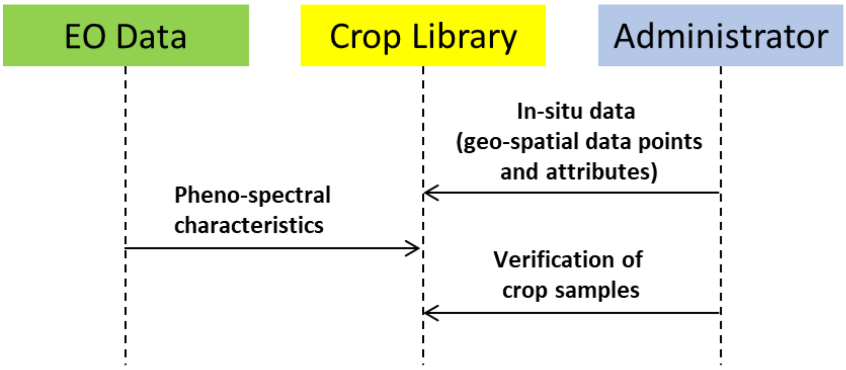

Figure 4. Sequence diagram for the constitution of a crop library.

3 Technical Background

3.1 App Architecture

The tool is has been developed using a Google cloud technology stack as shown in the architecture diagram (Figure 5). Image processing and geo-spatial analysis tasks are carried out in Google Earth Engine (Gorelick et al., 2017). The two browser applications (front-end application and administrator tool) have been developed as Google Earth Engine Apps. They are connected to each other (and to the GEE remote sensing archive) via the GEE Javascript API. A Google Cloud Storage (GCS) account guarantees sufficient file storage capacity. The export of the crop maps to GCS is triggered via the administrator tool, while the front-end application accesses the stored maps for the purpose of visualization and further processing. The file size of individual crop maps can be larger than 100 MB, but GCS offers practically unlimited storage capacity. The crop library is directly saved as an asset on the FAO EOSTAT account on GEE. Both applications are connected with the crop library, but only the administrator tool has permission to execute modifications.

Public access can be allowed to the front-end application because the app does not have permission to make changes to any assets saved on GEE or GCS. Full access to the administrator tool, however, should be restricted to trusted logged-in users. It is therefore necessary to run this browser application via the GEE Code Editor.

If commercial images (i.e., PlanetScope) are used for crop map generation (as it is the case for the Afghanistan example), these images are also saved on GCS. We use Google Compute Engine for the automated download of weekly PlanetScope mosaics (Figure 5). The purpose of each component of the app's architecture is further summarized in

Table 2.

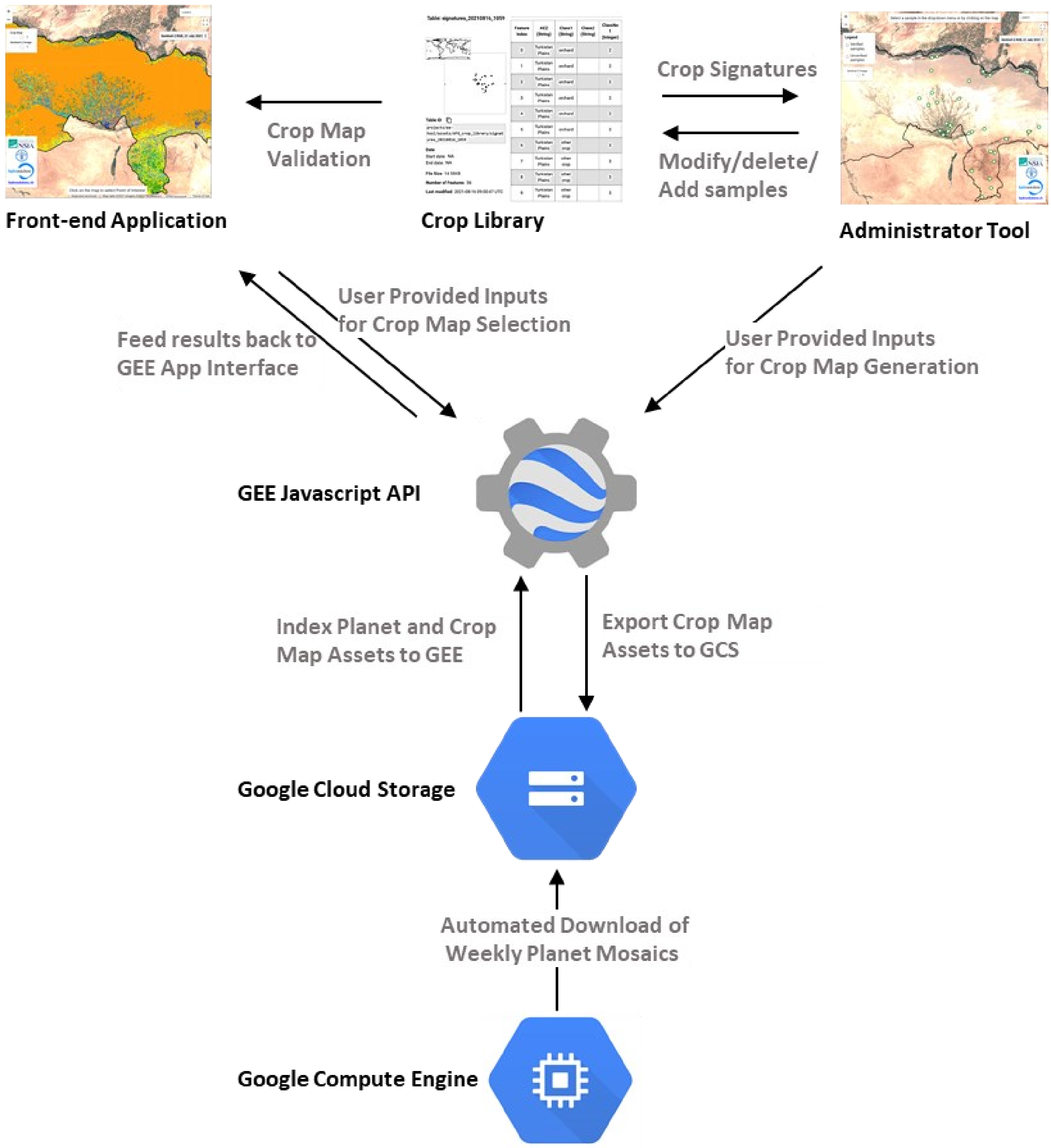

Figure 5. Architecture diagram of the EOSTAT CropMapper Afghanistan Application

Table 2. Description of each component of the app's architecture.

Component	Type	Purpose(s)
Front-end Application	Browser application	- Visualizing crop maps based on user queries - Crop area calculations - Crop map download - Crop map classification accuracy assessment.
Administrator Tool	Browser application run via the GEE Code Editor	- Crop library management - Generation of new crop maps
Crop Library	Database	Storing information on in-situ samples that are required for the training of supervised classification algorithms (DTW and Random Forest)
GEE Javascript API	Application Programming Interface	Software interface connecting the browser applications with Google Cloud Storage
Google Cloud Storage	Online file storage web service	Storing and accessing data on Google Cloud Platform infrastructure
Google Compute Engine	Service component of Google Cloud Platform	Run computationally expensive tasks on the Google infrastructure

3.2 Sensors

EOSTAT Cropmapper allows the user to use a combination of open and free EO data (optical and SAR) and of commercial products. The following (combination of) sensors are currently available for selection in the EOSTAT CropMapper Afghanistan tool:

1. Sentinel-1 and Sentinel-2
2. Sentinel-2
3. Sentinel-1 and Planet

The choice of sensors affects the resulting crop map spatial resolution (4.7 m if Planet data are used, 10 m otherwise). Using only visible and near infrared remote sensing observations (Sentinel-2) instead of combining them with radar remote sensing data (Sentinel-1) decreases the processing time required for the generation of crop maps. Option 2 is therefore a suitable choice for running a relatively quick analysis. However, the fusion of optical and radar data remarkably improve the results of classification, according to most sources (Orynbaikyzy et al., 2019). The option that leads to the longest processing times is option 3 above (Sentinel-1 and Planet), but produces maps with the highest level of detail. Up to 24 hours are required to export a crop map for an area of approximately 30,000 km^2 in 4.7 meters resolution. This means that after triggering the export of the map in the administrator tool it takes up to 24 hours until the map becomes available in the end-user tool.

Option 3 is currently only available for mapping summer harvest crops of the year 2021. PlanetScope data (optical images with a 4.7 m resolution) are partially available for Afghanistan due to an arrangement between Planet Labs and the Afghan National Statistic and Information Authority (NSIA). With Planet data crop maps of 4.7-meter spatial resolution can be generated. However, the agreement with NSIA has ended with the Taliban take-over. For any time before January 2021 and after August 2021 no Planet imagery is available through the EOSTAT CropMapper Afghanistan.

A multi-band approach is used for the crop type classification, which means that several bands and layers from each sensor are considered for training ('Used Bands' in Table 3)(Csillik et al., 2019)(Csillik et al., 2019)(Csillik et al., 2019) (Csillik et al., 2019). The normalized difference vegetation index (NDVI) expresses the differential reflection of green vegetation in the visible and near-infrared portions of the spectrum. NDVI is therefore calculated from the Red and NIR bands. It is used because of its well-known power in identifying crop classes.

Table 3. Summary of satellite sensors used by the EOSTAT CropMapper Afghanistan

Sensor	Type	Bands (total)	Used Bands	Temporal Resolution	Spatial Resolution (used bands)
Sentinel-1	Radar	VV and VH	VV and VH	One image every 6 days	10 m
Sentinel-2	Optical	13 spectral bands	Green, Blue, SWIR1, SWIR2, NDVI	One image every 5 days	10 m
Planet	Optical	4 spectral bands	Green, Blue, NDVI	One image every 8 days	4.7 m

3.3 Agricultural Seasons and Image Preprocessing

Fundamental for the use of EO data for crop type mapping is the knowledge of the crop calendar, as it is necessary to ensure that the satellite images used for the assessment are acquired during the appropriate time window.

The period of assessment can be adjusted according to user needs. It is possible to either only map summer or autumn harvest crops, respectively, or to map both at the same time. The choice depends on the validity period of the available in-situ data (Table 1). For Afghanistan, the following validity periods are currently allowed:

1. Summer harvest crop season
2. Autumn harvest crop season
3. Summer & Autumn harvest crop seasons

In the administrator tool, the user can specify the validity period of the desired crop map and trigger the generation of the map if the corresponding in-situ samples are available. The agricultural season is then stored as an attribute of the map and can be queried in the front-end application.

In Afghanistan, harvesting of the winter wheat and barley starts in May in the eastern part of the country and is completed as late as August in the mountainous areas[1]. The harvesting dates may shift from year to year due to meteorological factors. Users can therefore specify the main harvest month of summer or autumn harvest crops, respectively, prior to the generation of crop maps in the administrator tool. The algorithm will then automatically consider a time period of four months for

[1] https://www.fao.org/giews/countrybrief/country.jsp?code=AFG

training, whereas the main harvest month is the last month of this period. If summer and autumn harvest crops are mapped at the same time, the algorithm will automatically consider all months March to October, independently of the main harvest months.

The remote sensing images are aggregated to temporal mosaics to reduce the computational burden of the crop map generation, and to generate a harmonized time series of temporal composites. Cloud coverage areas may still exist in mosaics. Missing data in mosaics are therefore filled by interpolating the values in time. All available imagery is aggregated into 8 time-steps (15-day intervals for mapping summer harvest or autumn harvest crops, respectively, and 30-day intervals for mapping both summer and autumn harvest at the same time). Aggregation is limited to 8 time-steps because further increasing the number of time-steps may lead to the exceedance of the user memory limits in GEE.

3.4 Classification Algorithms

The following three crop type classification algorithms are available in the tool:

1. Time-constrained Dynamic Time Warping (DTW; Csillik et al., 2019).
2. Random Forest supervised classification (Gorelick et al., 2017).
3. Unsupervised classification based on harmonic regression (Wang et al., 2019).

TIME-CONSTRAINED DYNAMIC TIME WARPING

In Scenario 2, where sparse in-situ data is available, it is advisable to use DTW for a supervised classification. The DTW algorithm classifies every pixel with one of the available crop labels available in the crop signature library. The selection of the label is based on the highest similarity (or lowest dissimilarity) between a query pattern and available reference patterns (Figure 6). For each pixel, the time series of pheno-spectral characteristics from the EO pre-processed data (temporal mosaics) are compared to the time series at the locations of the labelled reference data. Computing the alignment between two sequences is done recursively using the DTW matrix (Figure 6). The algorithm then picks the smallest DTW dissimilarity value between the query pattern and the available reference patterns and attributes the corresponding reference crop label to the pixel.

DTW classification has the main advantage that only a small number of training samples are required, as little as 3 samples per crop type according to Belgiu and Csillik (2018). This is a big advantage for regional and national crop type mapping, especially in countries which lack input training samples. A few clean reference samples that represent the characteristic temporal pattern of the crop type are already sufficient (Csillik et al., 2019). Of course, the positional accuracy and correctness of crop label of the training data is of paramount importance for DTW, as the algorithm is in fact very sensitive to errors in the training data.

DTW is a time-flexible method for comparing two temporal patterns by considering their temporal distortions in their alignment (Figure 7). DTW was proven to achieve better results than the Euclidean distance measure for NDVI time series clustering (Csillik et al., 2019; Zhang et al., 2014). This flexibility is desirable for crop mapping, to deal with the intra-class phenological discrepancies caused by different agricultural practices, environmental conditions, or by different weather conditions. This is beneficial also for regionally more distant comparisons, which may exhibit temporal shifts in growing patterns, while still belonging to the same crop type.

It is necessary to use time constraints for computing DTW in order to take into account the specific seasonality of different crops. For example, comparing an element of a sequence with all other elements of another sequence leads to erroneous results when aligning a winter crop with a summer crop. Applying time constraints on time warping increases the speed of processing, while providing meaningful results. A so-called time-constrained DTW implementation is therefore used in the EOSTAT CropMapper (Figure 8). The elements of two time series will be compared only if the date difference is smaller or equal to ω. We use a constraint period ω equal to 30 days, following the recommendations of Csillik et al. (2019).

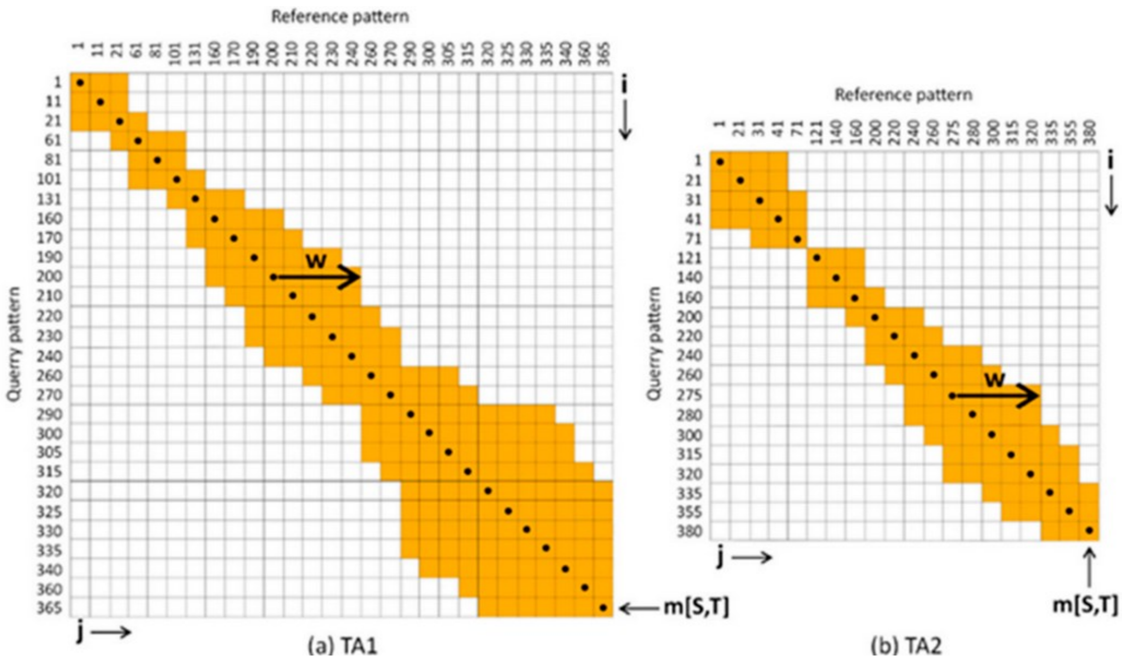

Figure 6. Computing the alignment between two sequences at hypothetical test areas (TA) TA1 andTA2. The vertical and horizontal values of the DTW matrix represent the date of an image. The alignment between two sequences is computed only for the yellow cells of the matrix, reducing the number of computations necessary (a maximum time delay, w, of 45 days is used in this example). After computing the matrix from upper left to lower right, the last element of the matrix, m[S,T], is returned, as a measure of DTW dissimilarity between the two compared sequences. Image Copyright: Csillik et al., 2019.

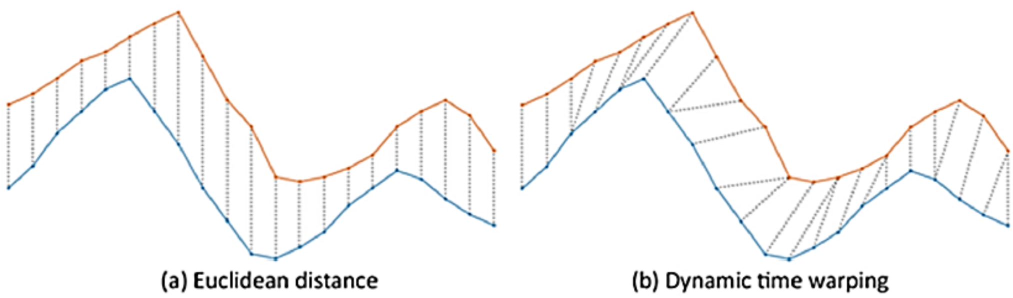

Figure 7. An illustration of two approaches for comparing two time series from Csillik et al. (2019). While Euclidean distance is time-rigid (a), the dynamic time warping (DTW) is time-flexible (b) in dealing with possible time distortion between the sequences.

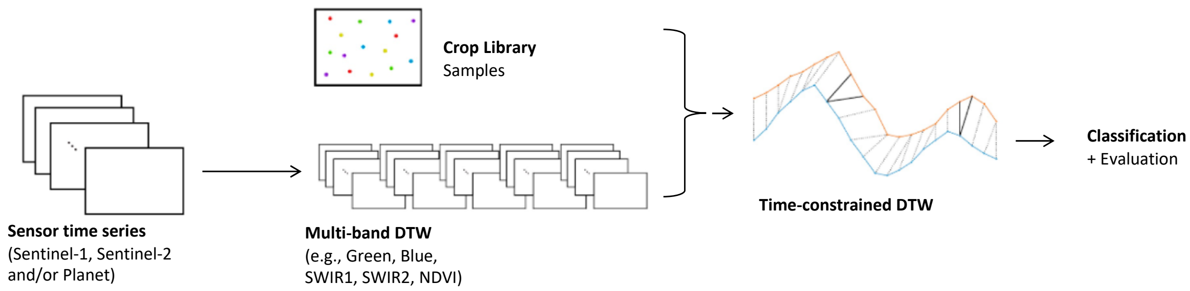

Figure 8. Workflow of time-constrained dynamic time warping (DTW) classifications (adapted from Csillik et al., 2019).

RANDOM FOREST

Random forest is an ensemble learning method for classification, that operates by constructing a multitude of decision trees based on available training data (Figure 9). In multi-band RF, the number of variables available for tree construction is *n* times *i*, where *n* is the number of bands used (Table 3) and *i* is the number of time seps. For classification, the output of the random forest is the class selected by most trees (majority voting). Using multiple deep decision trees, trained on different parts of the same training set, generates a classification model with a reduced risk of overfitting the training set. This generally increases the performance of the model, especially if a large training data set is available. The RF classifier has been extensively used to map land cover mapping and crop type mapping from Landsat images(Tatsumi et al., 2015), Sentinel 1 and 2 (Defourny et al., 2019; Orynbaikyzy et al., 2020), Sentinel -2 and the Gaofen-1 (GF-1, Chinese satellite data, (Fan et al., 2021). The main strength of the RF is that it copes well with collinearity, and is not affected by outliers, high dimensionality, and noisy features. Wang et al. (2019) confirm that random forests generalize well within regions where crop compositions and phenologies remain similar.

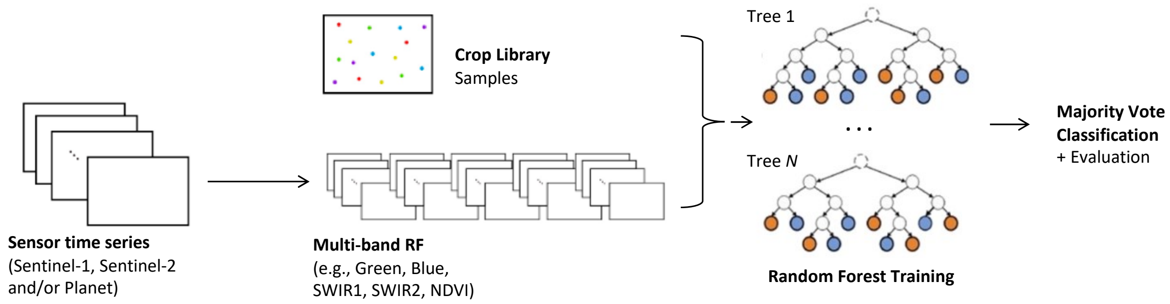

Figure 9. Workflow of the Random Forest classification.

The RF implementation in GEE is computationally very efficient and may be a better choice than DTW if the latter exceeds the user memory limits in GEE, but usually requires a larger number of training samples to reach the same classification accuracy (see Section 4 below). The most important parameter of the RF implementation in GEE is the number of decision trees to create, for which we use a value of 50 trees.

UNSUPERVISED CLASSIFICATION

Unsupervised classification can be an alternative if no in-situ samples are available at all. The approach uses harmonic regression for the pixel-wise approximation of optical remote sensing time series (Wang et al., 2019). The algorithm that is used for unsupervised classification is the k-means cluster analysis algorithm [48,49]. The algorithm is trained with 1000 pixels sampled randomly from within the area of interest. The harmonic regression coefficients of the randomly sampled points are used to cluster the pixels, i.e., to group pixels with similar coefficients into different clusters (Figure 10). Finally, all image pixels are partitioned into one of the resulting clusters. Crop labels are not provided automatically. The user is therefore requested to label the automatically identified clusters, which requires expert knowledge in interpreting characteristic NDVI signatures or familiarity with the situation on the ground. The major disadvantage of unsupervised classification is that crops are classified less consistently than by the supervised classification algorithms. The meteorological conditions, variable crop combinations or variable cropping calendars may affect the automatic clustering of pixels and therefore the classification accuracy. The applicability of unsupervised crop classification is therefore somewhat restricted to places with low crop diversity and uniform cropping calendars.

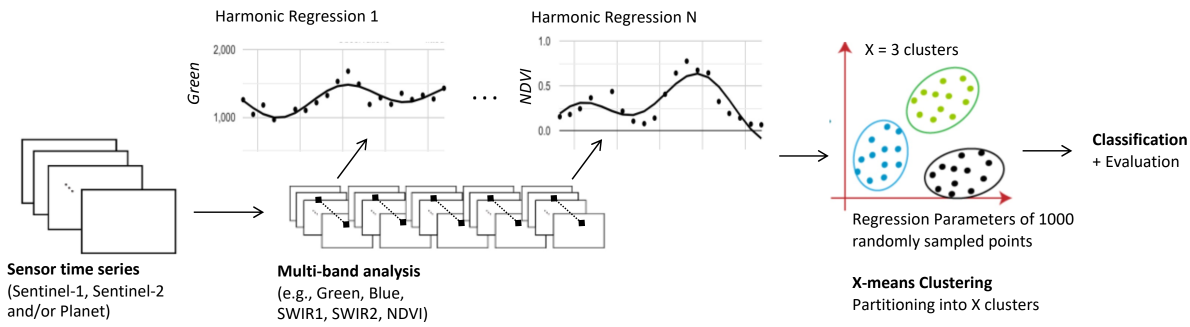

Figure 10. Workflow of unsupervised classification based on X-means Clustering.

It should be noted that unsupervised classification is currently not available as a choice for exporting crop maps via the administrator tool. Instead, crop maps based on the unsupervised classification of pixels are generated on-the-fly in the front-end application. The display of the crop maps and the generation of crop statistics will therefore take more time than the pre-computed image visualization and processing. The processing time is reduced by only considering pixels for classification with a peak-NDVI value greater or equal to 0.4.

3.5 Upload, QA/QC and Selection of Training Data

In-situ data are required to train the Random Forest and the Dynamic Time Warping algorithms to perform the supervised crop classification. The in-situ samples and their attributes (see Table 1) can be uploaded as point shapefiles to a dedicated cloud asset folder in GEE. Alternatively, it is also possible to geo-locate the samples and specify their attributes directly in the administrator tool.

To achieve high accuracy in crop type classification, it is of paramount importance that the in-situ data are gathered with very high geospatial accuracy and correct identification in the field. The quality assurance and quality control (QA/QC) of in-situ samples is therefore an important task. QA/QC is required to assure that no attributes are missing and that the crop library does not contain errors. For example, errors may occur if the coordinates of the in-situ samples are uncertain. Such positional uncertainty may result in crop samples that in fact represent the spectral signature of a road next to the field where the in-situ sample was collected. Unfortunately, such errors happen relatively frequently, and it is therefore necessary that the user verifies each in-situ sample that is used for training in the administrator tool. The administrator tool plots the points on top of high-resolution satellite imagery, which allows for visual verification. Furthermore, the spectral signature of each point can be displayed. It is therefore advised to train administrators how to interpret NDVI signatures prior to the verification of in-situ samples.

In the administrator tool, users can select the number of samples per class that will be used for training. Training points are then sampled randomly within the available verified samples. For DTW classification it is recommended to not use more than 10 samples per class. The careful verification of only few samples will lead to better classification results than choosing more but less carefully checked training samples. Random forest classification, on the other hand, benefits from using as many samples as possible. For preventing the exceedance of the user memory limits in GEE, the maximum number of training samples per class is limited to 99 samples in the case of random forest. It is recommended to use a balanced number of samples per crop class for training. For example, if only 20 samples of wheat are available, but 80 samples of cotton, the number of samples per class for training should be set to a value of 20 or lower. All verified samples that have not been used for training will be automatically available for validation once the corresponding crop map has been generated.

CROP TYPE DISSIMILARITY SCORES

The EOSTAT CropMapper Administrator Tool allows an automatic pre-screening of samples from individual crop classes based on Crop Type Dissimilarity Scores (CTDS). This option facilitates the identification of in-situ samples with wrong labels and/or wrong geo-tags. The pre-screening makes use of the DTW dissimilarity scores calculated between the signatures of individual samples and a reference signature. The reference pattern (Figure 6) is provided by the mean signature of all samples from a given crop type and area of interest (considering all Sentinel-2 bands and VIs listed in Table 3). Automatic pre-screening is therefore only a valid option for crop classes that are relatively homogeneous in their expected signature (unlike a class ‘Other crops’ for example). If this is the case, high dissimilarity scores between the reference pattern and individual samples indicate possible misclassifications.

Figure 11 shows an example where a high CTDS value allows identifying an in-situ Maize sample that mistakenly points to built-up area next to the actual cropping area. A 10 m GPS horizontal positioning error results in in erroneously pinpointing a Maize sample into a built-up area. Consequently, if such a sample was used for training of the DTW, all built-up area pixels might end up being labelled as Maize. Figure 12 pinpoints to the location of the in-situ Maize sample with the lowest CTDS value in comparison to all other Maize samples in the same area of interest. Such samples with low CTDS values can be marked as ‘verified’ in the administrator tool and will thereafter be considered as a candidate for the training of the supervised classification algorithms. The user of the administrator tool has also the option to automatically mark as ‘verified’ all samples with a CTDS that is lower than the CTDS of 50% of all samples from a given crop type and area of interest (Q50).

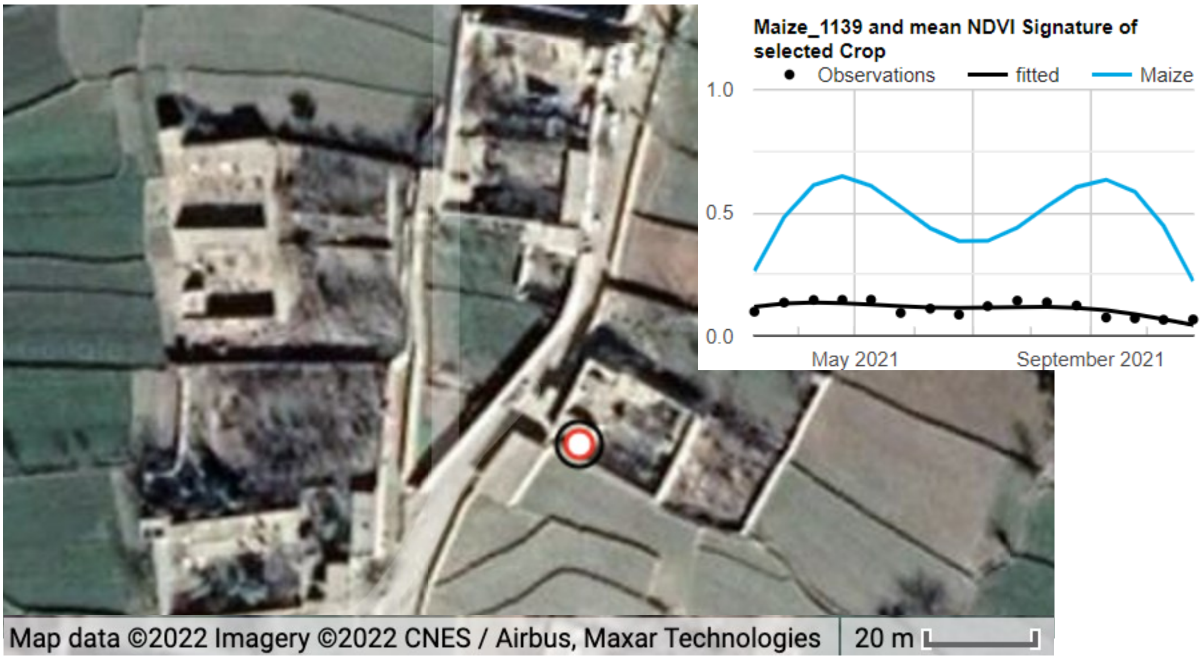

Figure 11. In-situ sample with the highest dissimilarity score for Maize in the AEZ Eastern Mountains and Foothills (red circle). The inset figure shows the NDVI signature of this location (black points and line) and the mean NDVI signature of all Maize samples in the AEZ (blue line). Print screens from the EOSTAT CropMapper Administrator Tool Afghanistan.

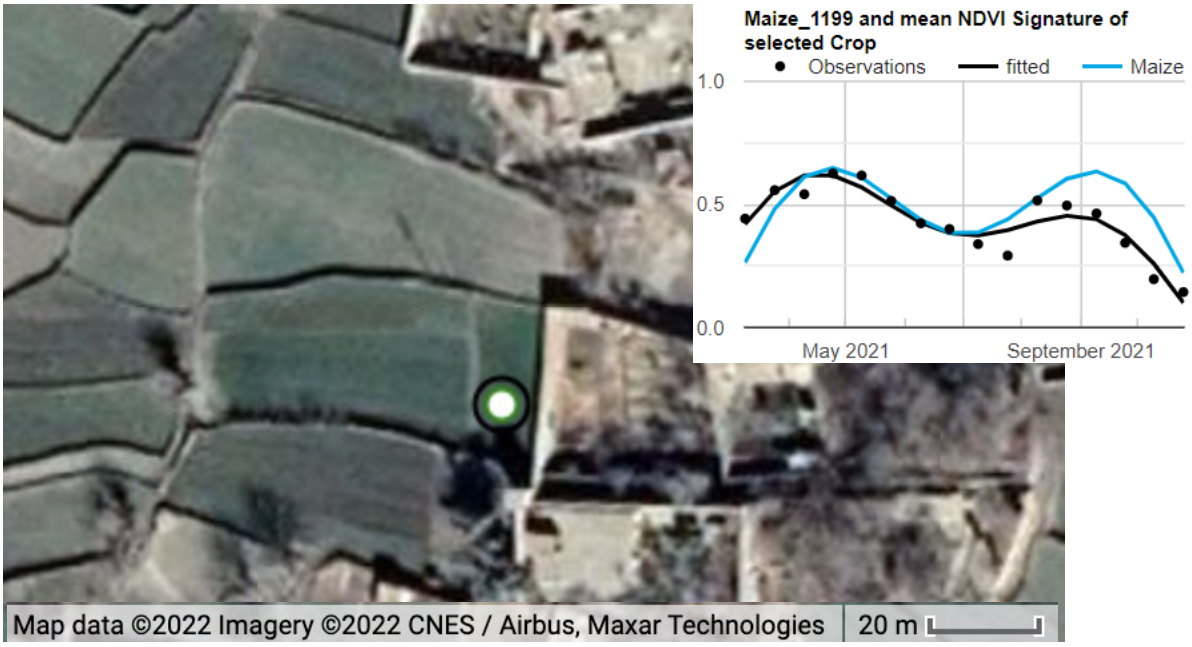

Figure 12. In-situ sample with the lowest dissimilarity score for Maize in the AEZ Eastern Mountains and Foothills (green circle). The inset figure shows the NDVI signature of this location (black points and line) and the mean NDVI signature of all Maize samples in the AEZ (blue line). Print screens from the EOSTAT CropMapper Administrator Tool Afghanistan.

3.6 Use of Masks

In the current version of the EOSTAT CropMapper Afghanistan, all pixels within a given area of interest are attributed a class label, except for pixels at elevations above 3600m or slopes steeper than 30°.

Furthermore, we use the Copernicus Global Land Cover Layers[2] to mask out built-up areas ('urban-coverfraction' greater than 20%) or permanent water ('water-permanent-coverfraction' greater than 20%). An explicit crop mask is not used, because up-to-date crop masks are difficult to obtain. This means that the training data must also include samples of fallow and barren land. If natural vegetation (e.g., in wetlands) is present within an area of interest, this should also be reflected in the training data set.

3.7 Stratification

The territory of Afghanistan has been divided into strata, by intersecting the administrative boundaries (Provinces) with the agro-ecological zones (AEZ). Users can either select whole agro-ecological zones (AEZ), or the units resulting from the intersection of the AEZ with the province boundaries (hereafter called 'sub-units', Figure 13).

The purpose of stratification using agroecological zones, and more in general of stratification, is to reduce the variability in the training dataset. The variability of available in-situ data in one area will more likely reflect the variability of the phenological spectral characteristics of the pixels in the same area. Using such representative training data increases the predictive power of the model. Another advantage of working with smaller units is that the processing time for the generation of crop maps is shorter. In Afghanistan, the reference library is built per AEZ. Therefore, if users are generating crop maps for sub-units of AEZs, they have the possibility to use in-situ data located outside a given sub-unit for training, if such samples are located within the same AEZ.

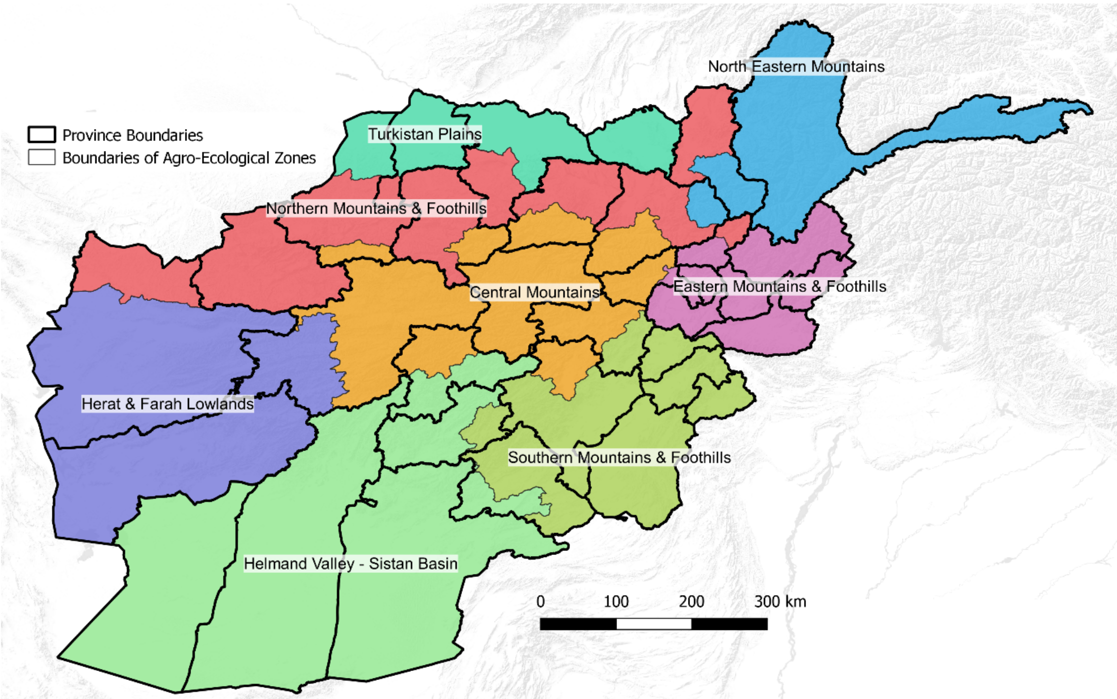

Figure 13. Provinces and agro-ecological zones (AEZ) of Afghanistan. The AEZ are shown as colored units and are labelled. The province boundaries are shown as thick black lines. The smallest spatial units available for selection in the EOSTAT CropMapper Afghanistan are the units resulting from the intersection of the AEZ with the province boundaries.

---

[2] CGLS-LC100 collection 3

4 Accuracy Assessment

This section presents an accuracy assessment of the crop type classification based on a dataset of in-situ data available for Kashkadarya Region in Uzbekistan (Figure 14). Located less than 150 km north of the Afghan border, Kashkadarya Region is climatologically and topographically similar to Afghanistan. The reference dataset (Remelgado et al., 2020) has been published in *Scientific Data*, which is a peer-reviewed, open-access journal for descriptions of scientifically valuable datasets. These ground-truth data were collected in the scope of the project Central Asia Waters (CAWa)[3] to provide consistent land cover information on crop types for efficient water management in Central Asia. The full dataset consists of 8'196 samples collected between 2015 and 2018 in several regions of Uzbekistan and Tajikistan. 2'172 samples are available for Kashkadarya, whereas they all have been collected in the year 2018.

in Kashkadarya as in the north of Afghanistan, wheat is the main staple crop. Wheat is harvested in June, allowing farmers to plant a second crop before the end of the agricultural year in October if water is available ('Double cropping'). Cotton is an important agricultural produce in Kashkadarya, and the second most frequent crop in the reference dataset. The cotton growing season extends from May to October. Furthermore, other existing crops are orchards, vineyards, and forage crops (alfalfa).

Classification maps and validation results are presented below and can also be viewed and downloaded in the sample EOSTAT CropMapper front-end application for Kashkadarya via this weblink.

Figure 14. Map of the Kashkardarya region in Uzbekistan with the crop type ground-truth data points shown as red dots, and the main water courses and water bodies in blue colour.

[3] www.cawa-project.net

4.1 Methodology

4.1.1 In-situ data QA/QC

The in-situ data are available in shapefile format as objects (polygons) drawn around the fields that were visited during the field survey in June 2018 (Remelgado et al., 2020). To each polygon, a single crop class is assigned. Training and validation data points were generated by randomly plotting points inside the polygons ensuring that they are distant at least 30 m from the border of the polygons and to each other, For crop classes where more than 50 polygons were available, only one point per polygon was generated (Figure 15). We then randomly selected 50 data points for training of the crop type classifiers. The remaining data points were used for validation (Table 4).

4.1.2 Classification

The in-situ data was used to train and test the DTW and the RF, using respectively 5 to 10 training points, and 50. The scope of this was to assess the accuracy of the two algorithms with respect to the volume of training data.

The main advantage of DTW as a supervised classification technique is that only a small number of well-chosen training samples are required (Belgiu and Csillik, 2018) to train the algorithm and obtain high accuracy in the crop classification (above 80%). On the other hand, Random Forest classifiers need large amount of training data. In essence, for the DTW, a small set of 5 to 10 training samples are sufficient to represent the temporal characteristic of the crop type (Csillik et al., 2019).

To evaluate the performance of the DTW algorithm, we randomly selected 5-10 points per crop class from the training data points, simulating the common case of a lack of input training samples. To assure the highest quality of the training data points for DTW, those points were carefully checked in the EOSTAT CropMapper administrator tool for consistency with the mean NDVI signature of a given crop category. If the NDVI signature of a given point was not consistent with the mean signal of a given crop class, then the point was removed from the dataset. RF classifiers have the advantage that they are not sensitive to outliers in the training data set. Additional training data points used only for the training of the RF algorithm were therefore not individually checked but it was assumed that the quality of the original dataset was sufficient for this classifier.

Table 4. Available Crop classes for Kashkadarya region, year 2018, and the respective number of geo-referenced polygons, and generated training and validation data points, respectively. The training samples for the class 'no-crop (natural vegetation)' were added to the dataset manually based on satellite image interpretation.

CROP CLASS	Available Polygons [N°]	Training data points generated [N°]	Validation data points generated [N°]
Cotton	945	50	892
Wheat	992	50	942
Wheat – other (double cropping)	18	50	30
Alfalfa	13	50	30
Vineyards	11	50	30
Orchards	85	50	33
No-crop (fallow)	100	50	47

No-crop (natural vegetation)	-	10	-
Other	7	-	-
Total	**2172**	**360**	**2004**

We trained both classifiers with monthly composites of Sentinel-1 and Sentinel-2 satellite images (sensor selection option 1, see Section 3.2) available from the period March to October 2018 (period selection option 3, see Section 3.3). We performed several rounds of training using different numbers of randomly sampled training data points (5-10 points for DTW, 10-50 points for RF). We then applied the trained algorithms to classify all pixels of Kashkadarya region at a spatial resolution of 10m, except urban areas or pixels above a certain elevation (2500 m asl) or with steep slopes (>5°). For validation we look at the classification accuracy with respect to all available validation points, as well as with respect to a subset of 30 random validation points per crop class. The subset was generated to account for the class imbalance in the validation data set (Table 4).

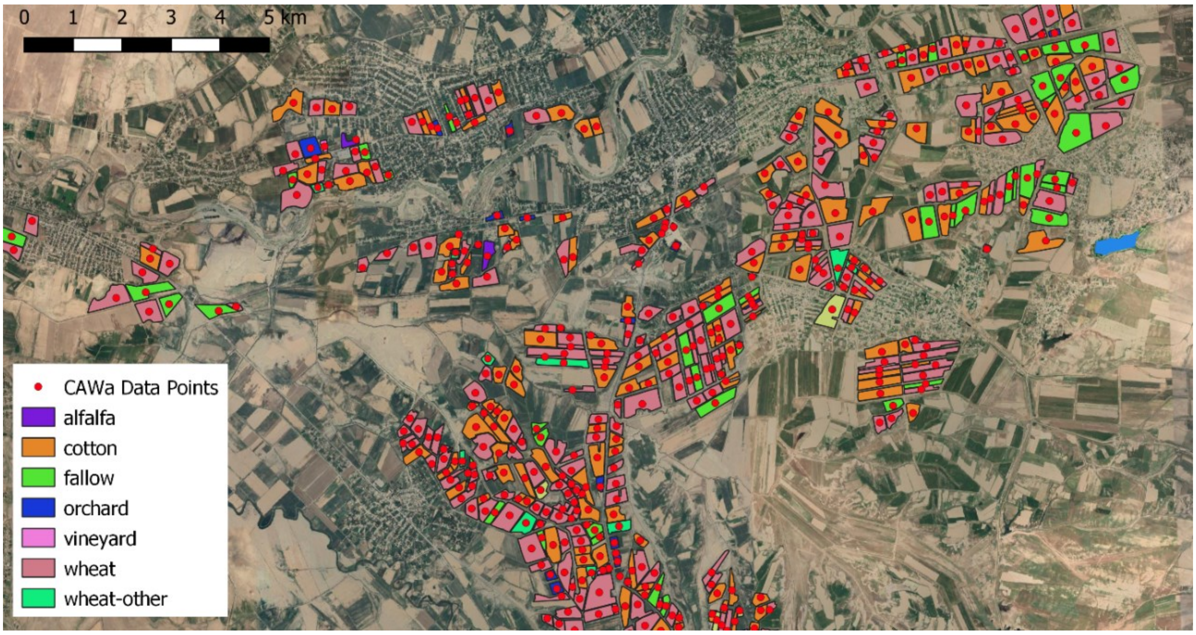

Figure 15. Detail of the CAWa ground-truth reference data available for Kashkadarya. The map shows the polygons as provided by the CAWa dataset, as well as one point per polygon generated inside each shape.

4.2 Results

Considering all 2004 available validation data points per crop class (Test A), DTW reaches an overall classification accuracy of about 84.4%. The accuracy difference resulting from using 5 or 10 training samples per crop class, respectively, is not statistically significant (Figure 16a). For the case where only 30 samples per crop class were used for validation (Test B), 10 training samples lead to a slightly higher overall accuracy (72.4%) than if only 5 training samples were used (69.1%, Figure 16b). In both tests the DTW classification accuracy is higher than the RF classification accuracy considering the same number of training samples (Test A: 1.1%-1.7% higher, Test B: 8.1%-10% higher, see Figure 16). However, the classification accuracy of the RF classifier significantly increases with the number of training samples used. With 50 training samples per crop class, the overall accuracy is 86.1% (Test A) and 76.2% (Test B), respectively (Figure 16).

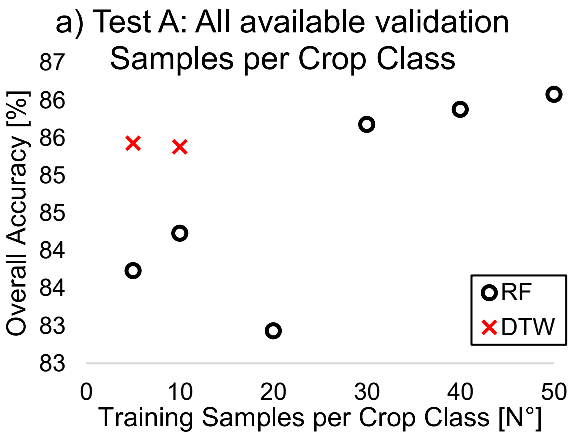

Figure 16. Overall accuracy of the DTW and RF classifiers as a function of the number of training samples used.

Table 5. Confusion matrix of the DTW classification (10 training points per crop class, Test A: all available validation points).

Predicted / ***Actual***	*wheat*	*cotton*	*alfalfa*	*orchard*	*vineyard*	*wheat-other*	*no-crop*	**Producer Accuracy**
wheat	787	7	12	23	30	48	35	**83.5%**
cotton	1	812	1	3	0	2	73	**91.0%**
alfalfa	1	0	16	2	4	1	6	**53.3%**
orchard	2	2	1	9	9	0	10	**27.3%**
vineyard	0	0	2	2	22	2	2	**73.3%**
wheat-other	2	0	0	0	0	28	0	**93.3%**
no-crop	2	3	0	4	1	0	37	**78.7%**
User Accuracy	**99.0%**	**98.5%**	**50.0%**	**20.9%**	**33.3%**	**34.6%**	**22.7%**	**Overall Accuracy 85.4%**

Table 6. Confusion matrix of the DTW classifier (10 training points per crop class, Test B: 30 validation points per crop class).

Predicted / ***Actual***	*wheat*	*cotton*	*alfalfa*	*orchard*	*vineyard*	*wheat-other*	*no-crop*	**Producer Accuracy**
wheat	27	0	1	0	1	1	0	**90.0%**
cotton	0	25	0	0	0	0	5	**83.3%**
alfalfa	1	0	16	2	4	1	6	**53.3%**
orchard	1	2	1	8	9	0	9	**26.7%**
vineyard	0	0	2	2	22	2	2	**73.3%**
wheat-other	2	0	0	0	0	28	0	**93.3%**
no-crop	1	2	0	1	0	0	26	**86.7%**
User Accuracy	**84.4%**	**86.2%**	**80.0%**	**61.5%**	**61.1%**	**87.5%**	**54.2%**	**Overall Accuracy 72.4%**

Table 7. Confusion matrix of the RF classifier (50 training points per crop class, Test A: all available validation points).

Predicted / ***Actual***	*wheat*	*cotton*	*alfalfa*	*orchard*	*vineyard*	*wheat-other*	*no-crop*	**Producer Accuracy**
wheat	782	7	35	22	23	59	14	**83.0%**
cotton	1	824	3	6	0	0	58	**92.4%**
alfalfa	1	1	21	2	2	1	2	**70.0%**
orchard	2	2	2	17	9	0	1	**51.5%**
vineyard	0	0	0	4	24	2	0	**80.0%**
wheat-other	1	0	0	0	0	29	0	**96.7%**
no-crop	2	5	6	3	3	0	28	**59.6%**
User Accuracy	**99.1%**	**98.2%**	**31.3%**	**31.5%**	**39.3%**	**31.9%**	**27.2%**	**Overall Accuracy 86.1%**

Table 8. Confusion matrix of the RF classifier (50 training points per crop class, Test B: 30 validation points per crop class).

Predicted / ***Actual***	*wheat*	*cotton*	*alfalfa*	*orchard*	*vineyard*	*wheat-other*	*no-crop*	**Producer Accuracy**
wheat	27	0	1	0	0	2	0	**90.0%**
cotton	0	27	0	0	0	0	3	**90.0%**
alfalfa	1	1	21	2	2	1	2	**70.0%**
orchard	1	2	2	15	9	0	1	**50.0%**
vineyard	0	0	0	4	24	2	0	**80.0%**
wheat-other	1	0	0	0	0	29	0	**96.7%**
no-crop	1	4	5	2	1	0	17	**56.7%**
User Accuracy	**87.1%**	**79.4%**	**72.4%**	**65.2%**	**66.7%**	**85.3%**	**73.9%**	**Overall Accuracy 76.2%**

The producer and User accuracies of individual crop classes for the tests with 10 (DTW) and 50 (RF) training samples per crop class, respectively, are presented in Tables 2-5. Test A leads to very high User accuracies for wheat and cotton (98.2%-99.1%, Table 5, Table 7). However, the other classes are much less common (Table 4), which is the main reason why the fraction of false-positive classifications for wheat and cotton is small. Test B (equal number of validation points per crop class) provides thus a more representative assessment of the User accuracies. Here the User accuracies of the DTW classification varies between 54.2% (no-crop) and 86.2% (cotton), and for the RF classification between 65.2% (orchard) and 87.1% (wheat). We obtain higher User accuracies with DTW than with RF for three out of seven classes (cotton, alfalfa and wheat-other).

The producer accuracies of Test A and DTW vary between 27.3% (orchard) and 93.3% (wheat-other). The same results for RF vary between 51.5% (orchard) and 96.7% (wheat-other). RF performs equal or better than DTW in all crop categories except no-crop (59.6% and 78.7%, respectively). However, this last class is the most important category for estimating the total crop areas because it is the most common land cover category (Table 9). According to DTW, 34.2% of the classified areas are used for crop cultivation and 48.6% according to RF. With RF we obtain a total crop area that is 28.3% larger than with DTW. About half of the difference stems from the difference in alfalfa acreages, which are almost three times larger based on the RF classifier (Table 9). Alfalfa represents only 0.6% of the reference data polygons, but 15.6% of the classified crop areas with RF (DTW: 5.1%). It is likely that both classifiers thus overestimate the total alfalfa acreages.

Table 9. Classified areas per crop category in km^2 (DTW: 10 training samples per crop category; RF: 50 training samples per crop category).

	Area [km^2]	**Area [km^2]**	**Difference**	
Crop Class	**DTW**	**RF**	**[km^2]**	**[%]**
wheat	1’446.7	1’480.1	+33.4	+2.3%
cotton	1’421.1	1’626.2	+205.1	+14.4%
alfalfa	221.2	867.8	+646.5	+292.3%
orchard	606.2	883.0	+276.8	+45.7%
vineyard	461.5	502.4	+40.9	+8.9%
wheat-other	173.9	197.1	+23.2	+13.3%
Total Crops	4’330.6	5’556.5	+1’226.0	+28.3%
No-crop	12’659.7	11’433.8	-1’226.0	-9.7%

4.3 Discussion

This accuracy assessment has demonstrated that the crop classification with DTW based on few carefully checked training samples can outperform conventional RF classification with at least two times more samples. With five times more training samples, RF outperforms DTW in terms of overall accuracy. However, also with five times more training samples RF yields low producer accuracies for the land cover category ‘no-crop’. This category is underrepresented in the ground-truth dataset used for validation but is generally the most important category for estimating the total crop acreages, because it is the most common land cover in Kashkadarya region. A low producer accuracy is the consequence of frequent false-negative classifications, and therefore usually indicates a general underestimation of a given class frequency. Indeed, the total crop area according to the RF classification is almost 30% larger than according to the DTW classification.

It is concluded from these results that DTW can lead to more accurate and more robust results than RF. The main condition for obtaining good results with DTW is a comprehensive quality assurance and quality control of the training data points. Even in a published crop type datasets such as used for this assessment we found several obvious misclassifications that could be explained with timing of the field campaign (June 2018), which was likely too early for accurate sampling of late crops, or with incorrect delineations of the field boundaries. While the full ground-truth dataset consists of 2’172 samples, we needed only 40-80 samples to train the DTW algorithm. It is understood that the quality assurance and control of such small samples sizes requires less time, and can therefore be done more thoroughly. RF is less sensitive to noise in the training data, and a large training data set can therefore compensate the mistakes in the labeling of the ground-truth data. However, large input training samples are difficult to obtain. DTW, as implemented in the EOSTAT CropMapper, is therefore a valuable alternative to RF classification.

5 Conclusions

One of the main challenges for a successful application of the tool is to find a compromise between classification accuracy, computational burden and efforts required for in-situ sample collection. Random Forest is computationally more efficient than the DTW implementation in GEE. However, DTW allegedly performs better than Random Forest when only few geotagged in-situ samples are available. To make an optimal selection of samples, algorithms, sensors, and spatial units is therefore a major task for the user. The possibility to use the EOSTAT CropMapper for assessing the classification accuracies facilitates this task. Different settings can be tested, and the crop maps associated with the highest accuracies can be retained. The validation option in the front-end application allows to display all misclassified points with respect to the training data, which offers the user the possibility to identify the locations where more training data are required or to detect wrong labels in the training data set. The EOSTAT CropMapper therefore offers the flexibility to work with a range of different methods and to optimize its performance, in order to find an optimal solution for a given location and the available in-situ samples.

Appendix 1: EOSTAT CropMapper Application to Afghanistan

The front-end application of the EOSTAT CropMapper for Afghanistan provides already access to crop maps from every AEZ and Province of Afghanistan (Figure 13). These crop maps were generated on the basis of training data provided by a dataset of 2047 in-situ samples collected across all Agroecological Zones of Afghanistan in November and December 2021 (Figure S1). Given the bad security situation in Afghanistan over most of the year 2021 the field survey could not start before the end of November, which is after the harvest season. The in-situ samples were therefore collected by interviewing the farmers. This resulted in a large fraction of wrongly georeferenced or wrongly labelled samples. This Appendix describes how this dataset was curated within the administrator tool and applied for crop type mapping. Due to the overall lower quality of the training data, the resulting crop maps have a lower accuracy than the maps from Kashkardarya in Uzbekistan (see Section 4). Nevertheless, this application provides a first estimation of the total land area of major agricultural crops in Afghanistan.

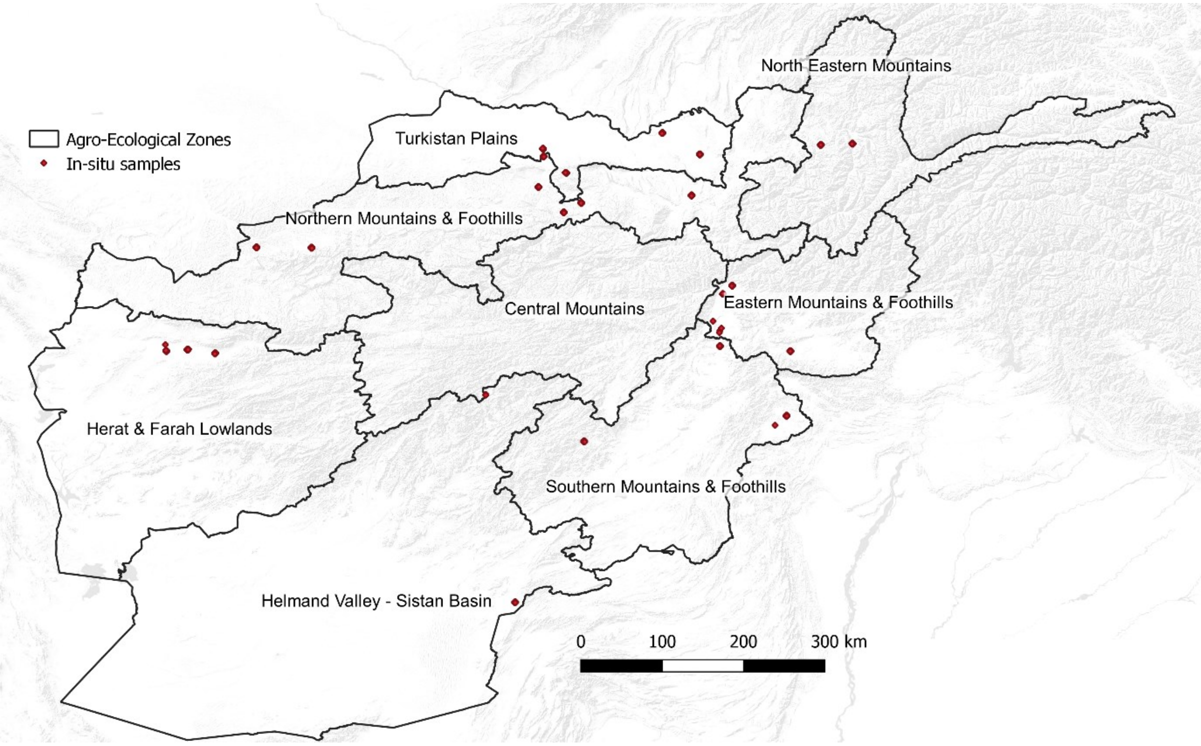

Figure S1. Locations with available in-situ samples from Afghanistan from the year 2021.

Data and Methodology

2047 samples were collected in-situ by the NSIA enumerators between the months of November and December 2021. 269 samples had to be removed due to missing geo-tags. The crop type labels and the distribution of the remaining samples per crop type are presented in Table S1.

The dataset did not include samples of natural vegetation from every AEZ. Sample points of the classes ‘Non-crop’, ‘Forest’, ‘Grassland’ or ‘Wetland’ were therefore manually added through visual interpretation of very high-resolution satellite images provided by Google Earth Engine. Samples for the classes ‘Orchard’ and ‘Vineyard’ were also added. After the removal of erroneous samples (e.g., 123 in-situ samples from the AEZ ‘Central Mountains’ pointed to only four locations) and the addition of samples the from missing categories, we obtained a dataset consisting of 1698 samples (Table S1).

As shown in Table S1, 21 crop types were sampled. However, some crop types were represented by very few in-situ samples (e.g., only one sample of type ‘Sunflower’). Such a small number of samples is not sufficient to train the machine learning classifier, and to allow discrimination from other crops. On the contrary it would lead to confusion and overall lower accuracy of the final crop type map. For this reason it was decided that any crop types having less than 4 samples per AEZ are renamed ‘Other crop’. As a final result the training dataset included overall 19 crop type classes.

Table S1. Number of samples available per class in the original dataset and in the final dataset, respectively.

Class	Original (N°)	Final (N°)
Barley	7	0
Beans	92	91
Cassava	1	0
Chick_peas	1	0
Cotton	42	41
Fallow	84	63
Fodder_crop	69	61
Forest	0	8
Grassland	234	193
Maize	79	74
Millets	4	0
Non-crop	4	44
Oats	22	12
Orchard	0	37
Other crop	259	268
Poppy	21	14
Potatoes	17	13
Rice	86	85
Rye	5	4
Soya_beans	29	28
Sunflower	1	0
Vineyard	0	4
Wetland	0	28
Wheat	721	630
Total	**1778**	**1698**

The field survey to collect the in-situ samples took place in early winter, after the harvest season. The enumerators therefore questioned farmers about the last harvested crop at each location. Unfortunately, the harvest month was not recorded, and neither the cropping intensity at a given location. This further complicates the application of the dataset within the EOSTAT CropMapper, because of very heterogeneous crop classes containing both single and double crop samples in the same class. The period of assessment had to be defined as March to October 2021, because summer and autumn harvest crops could not be distinguished from the labels of the in-situ data.

The remote sensing analysis was based on Sentinel-1 and Sentinel-2 data (see Section 3.2). We tested both supervised classification algorithms (RF and DTW, see Section 3.4). To limit the class imbalance in the training data, the maximum number of samples per crop type that are used for training of the RF classifier was set to 20. For DTW classification we used up to 10 samples per crop class.

We used the automatic Q50 CTDS pre-screening option described in Section 3.5 to select the candidates for DTW training and validation, except for the classes ‘Other crops’ and ‘Wheat’. The classes ‘Other crops’ and ‘Wheat’ were characterized by a high diversity in their crop signatures, for which the automatic pre-screening is not recommended (Section 3.5). Samples used for training from these two classes were thus verified manually one by one in the EOSTAT CropMapper Administrator Tool. For RF classification the Q50 CTDS pre-screening option was not applied, because it is supposed that this classification method is less sensitive to outliers and benefits from larger number of training samples.

To reduce the variability in the training dataset, samples are always chosen from the same AEZ where a given area of interest is located. From the AEZ ‘Central Mountains’ no in-situ samples were available, and samples from the neighbouring AEZ ‘Eastern Mountains & Foothills’ were considered instead.

Results and Discussion

In total, 447 pre-screened samples were available for validation of the DTW crop maps. For the validation of the RF crop maps, 742 samples were available. The number of samples available per AEZ and classification method varied between zero and 241 (Table S2). From the AEZ ‘Central Mountains’ and 'Helmand Valley - Sistan Basin' no validation samples were available.

Across all available validation points for Afghanistan the average classification accuracy of DTW is 61.9% (Table S3), whereas with RF we achieve an average accuracy of 54.7% (Table S4, 7.2% less than with DTW classification). In four out of six AEZs DTW overall accuracy exceeds the RF overall accuracy (Table S2). In the AEZ ‘North Eastern Mountains’ RF classification is slightly better (+3.8%). The substantially higher classification accuracy of RF in the AEZ Southern Mountains & Foothills (+19.3%) can be explained by the fact that 49 out of 51 available validation samples (96%) belong to the class ‘Grassland’, which is a class that is generally easy to classify due to the low in-class variability. The largest difference between RF and DTW performance is obtained for the AEZ Eastern Mountains & Foothills, where DTW accuracy exceeds RF accuracy by 24.2%.

Table S2. Classification accuracy per AEZ and classification method. The number of samples available for validation is indicated, as well as the crop types corresponding to the available validation samples.

AEZ	Method	Overall Accuracy	Validation Samples (N°)	Validation Crop Types
Central Mountains	-	-	0	-
Eastern Mountains & Foothills	DTW	51.3%	39	Wheat, Beans, Maize
	RF	27.1%	59	Wheat, Maize
Helmand Valley - Sistan Basin	-	-	0	-
Herat & Farah Lowlands	DTW	50.0%	44	Wheat, Grassland
	RF	42.9%	77	Wheat
North Eastern Mountains	DTW	62.5%	56	Wheat, Beans
	RF	66.3%	83	Wheat, Beans
Northern Mountains & Foothills	DTW	60.2%	128	Wheat, Fallow, Grassland, Beans, Soya beans
	RF	45.6%	241	Wheat, Fallow, Grassland, Beans, Soya beans
Southern Mountains & Foothills	DTW	69.0%	29	Wheat, Grassland
	RF	88.2%	51	Wheat, Grassland
Turkistan Plains	DTW	73.5%	151	Wheat, Grassland, Rice, Fodder crop, Cotton
	RF	69.7%	231	Wheat, Grassland, Rice, Fodder crop, Cotton

Table S3. Confusion matrix of the DTW classification. All validation points for Afghanistan. The class 'Non-crop' includes also all 'Grassland', 'Fallow' and 'Forest' points.

Predicted / ***Actual***	*Wheat*	*Rice*	*Fodder*	*Maize*	*Soya*	*Cotton*	*Beans*	*Others*	*Non-crop*	**Producer Accuracy**
Wheat	154	0	5	3	5	10	18	39	47	**54.8%**
Rice	2	31	0	0	0	0	0	0	0	**93.9%**
Fodder	2	0	3	0	0	0	0	1	2	**37.5%**
Maize	3	0	0	14	0	0	2	4	0	**60.9%**
Soya	0	0	0	0	2	0	0	0	0	**100.0%**
Cotton	2	0	0	0	0	2	0	0	1	**40.0%**
Beans	1	0	0	0	0	0	16	3	0	**80.0%**
Others	13	1	8	6	7	2	1	44	9	**48.4%**
Non-crop	1	0	4	0	0	0	0	3	67	**89.3%**
User Accuracy	**93.3%**	**100.0%**	**25.0%**	**82.4%**	**28.6%**	**16.7%**	**44.4%**	**46.8%**	**53.2%**	**Overall Accuracy 61.9%**

Table S4. Confusion matrix of the RF classification. All validation points for Afghanistan. The class 'Non-crop' includes also all 'Grassland', 'Fallow' and 'Forest' points.

Predicted / ***Actual***	*Wheat*	*Rice*	*Fodder*	*Maize*	*Soya*	*Cotton*	*Beans*	*Others*	*Non-crop*	**Producer Accuracy**
Wheat	215	9	17	5	9	45	9	115	115	**45.1%**
Rice	2	57	1	0	0	0	0	0	0	**95.0%**
Fodder	0	0	1	0	0	0	0	0	0	**100.0%**
Maize	9	0	0	12	0	0	11	3	3	**26.1%**
Soya	0	0	0	0	4	0	0	0	0	**100.0%**
Cotton	1	0	0	0	0	5	0	0	0	**83.3%**
Beans	2	0	0	0	1	0	6	19	19	**20.7%**
Others	29	0	5	4	5	3	1	74	11	**56.1%**
Non-crop	3	1	1	0	0	1	2	104	104	**87.4%**
User Accuracy	**92.7%**	**85.1%**	**5.0%**	**70.6%**	**28.6%**	**9.8%**	**21.4%**	**50.7%**	**41.3%**	**Overall Accuracy 54.7%**

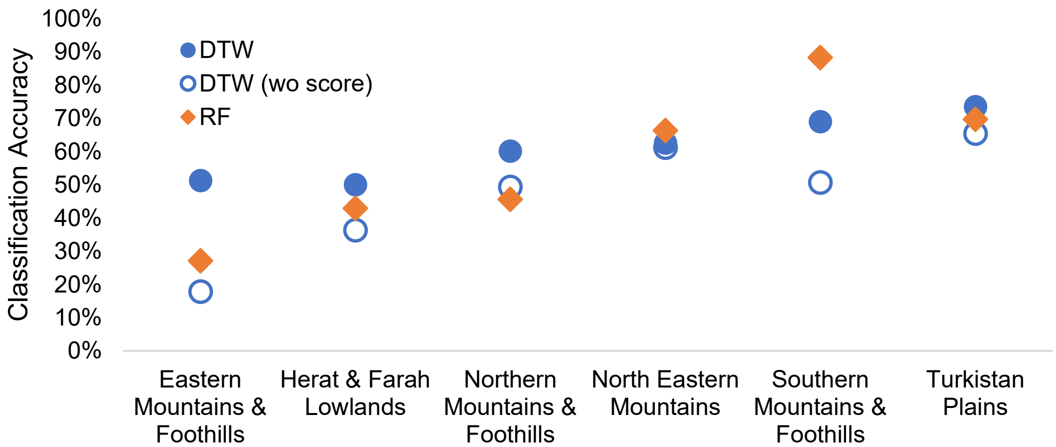

Figure S2. Classification accuracy per AEZ based on available validation samples using DTW, DTW without pre-screening of training samples ('wo score'), and RF, respectively.

DTW classification without the automatic Q50 CTDS pre-screening of in-situ samples leads to an average classification accuracy of only 47%, which is 15% less than if the training samples are selected based on basic quality criteria. The accuracy improvement which is due to the pre-screening of samples varies between 33.4% (Eastern Mountains & Foothills) and 1.3% (North Eastern Mountains), depending on the AEZ (Figure S2).

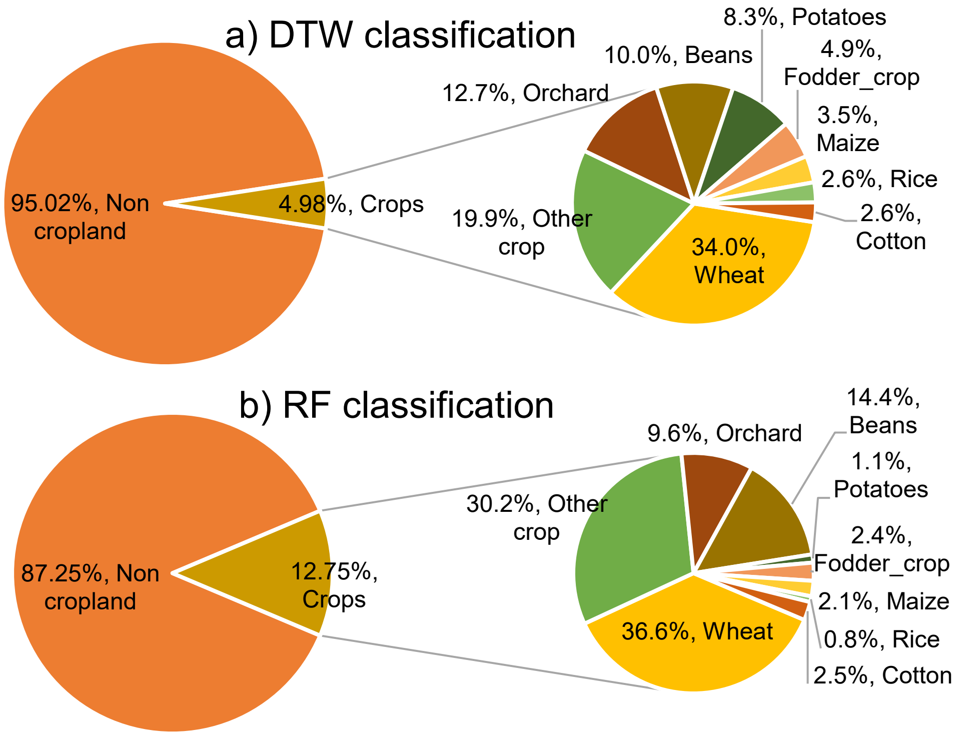

Figure S3. Fraction of cropped/non-cropped area (left) and per individual crop types (right) at the country scale.

With DTW classification (including Q50 CTDS pre-screening) we obtain a total cropped area of 3.2 million hectares. 61 million hectares (95.02%, see Figure S3a and Table S5) represent non-agricultural areas that were masked from the beginning (e.g., areas at elevations above 3600 m or with slopes steeper than 30°), or areas that obtained the class label 'Non-crop', 'Grassland', 'Fallow', 'Wetland' or 'Forest'. With RF classification we obtain a cropped area that is substantially larger (8.2 million hectares, 12.75% of the total land area of Afghanistan, Table S5), mainly due to smaller areas classified as 'Grassland' or 'Fallow' (Table S6).

The two country-scale crop maps obtained by mosaicking the AEZ crop maps per each classification method are presented in Figure S4 (DTW) and Figure S5 (RF). The crop maps are also available per AEZ and Province at their original spatial resolution (10 m) in the Web App. By zooming in and comparing the crop map to the satellite images it becomes clear that RF much overestimates the total cropped area, e.g. in Paktika, Logar, or Urozgan provinces (see also the cropland fraction difference between DTW and RF in Table S5).

Wheat is the primary crop of Afghanistan and is grown in every province of the country (Tiwari et al., 2020). In terms of total area, the crop class 'Wheat' is followed by 'Other crop' and then by Orchard, Beans and Potatoes (Figure S3a, Table S6). The country-scale wheat map obtained from the DTW outputs is presented in Figure S6. This wheat map can be compared to a similar map obtained by Tiwari et al. (2020) for the year 2017 based on Optical and SAR Time-Series Images (Figure S7). The total wheat cultivation area of the year 2017 in Afghanistan mapped by Tiwari et al. (2020) is 1.5 million hectares. This value is 36% higher than the value that we have obtained for 2021 based on DTW classification (1.1 million hectares). However, 2021 was a year characterized by a severe drought (FAO, 2021). About 560,000 ha (40%) of the wheat area in 2017 was rainfed (Tiwari et al., 2020), and therefore particularly vulnerable to a lack of rainfall in 2021. The value of 1.1 million ha of wheat area obtained by DTW classification for 2021 is therefore plausible, while the value of 3 million ha of wheat according to RF classification (Table S6) is much above the plausible range. However, official government statistics or other FAO estimates which could be used to validate the total areas of individual crops were not available for Afghanistan at the time of writing this report.

Table S5. Total cropland and non cropland area per province (DTW and RF classification).

	DTW				RF			
Province	**Cropland**		**Non cropland**		**Cropland**		**Non cropland**	
	hectares	%	hectares	%	hectares	%	hectares	%
BADAKHSHAN	211,728	4.9%	4,125,027	95.1%	457,474	10.5%	3,879,281	89.5%
BADGHIS	58,017	3.0%	1,906,983	97.0%	104,196	5.3%	1,860,804	94.7%
BAGHLAN	177,331	10.4%	1,529,944	89.6%	222,426	13.0%	1,484,849	87.0%
BALKH	172,720	9.9%	1,576,212	90.1%	173,923	9.9%	1,575,009	90.1%
BAMYAN	41,342	2.3%	1,761,259	97.7%	216,316	12.0%	1,586,285	88.0%
DAYKUNDI	71,266	5.2%	1,290,546	94.8%	279,550	20.5%	1,082,262	79.5%
FARAH	87,854	2.1%	4,002,619	97.9%	146,290	3.6%	3,944,183	96.4%
FARYAB	82,017	3.9%	1,999,115	96.1%	139,243	6.7%	1,941,889	93.3%
GHAZNI	91,177	4.2%	2,063,122	95.8%	406,494	18.9%	1,747,805	81.1%
GHOR	57,713	1.5%	3,792,273	98.5%	380,870	9.9%	3,469,116	90.1%
HELMAND	361,189	5.9%	5,713,967	94.1%	769,652	12.7%	5,305,504	87.3%
HERAT	121,387	2.0%	6,029,580	98.0%	363,381	5.9%	5,787,586	94.1%
JAWZJAN	78,094	7.4%	981,621	92.6%	45,984	4.3%	1,013,731	95.7%
KABUL	36,223	7.7%	432,493	92.3%	127,880	27.3%	340,836	72.7%
KANDAHAR	150,246	2.8%	5,287,871	97.2%	566,867	10.4%	4,871,250	89.6%
KAPISA	23,566	12.5%	164,341	87.5%	46,453	24.7%	141,454	75.3%
KHOST	112,870	27.5%	297,378	72.5%	269,447	65.7%	140,801	34.3%
KUNARHA	57,037	13.5%	364,697	86.5%	102,902	24.4%	318,832	75.6%
KUNDUZ	185,463	23.6%	600,870	76.4%	180,532	23.0%	605,801	77.0%
LAGHMAN	29,660	7.6%	360,360	92.4%	97,978	25.1%	292,042	74.9%
LOGAR	47,762	10.9%	390,759	89.1%	201,070	45.9%	237,451	54.1%
MAYDAN WODAKG	55,757	5.2%	1,024,112	94.8%	277,185	25.7%	802,684	74.3%
NANGARHAR	89,527	12.1%	647,955	87.9%	215,845	29.3%	521,637	70.7%
NIMROZ	21,369	0.5%	4,004,854	99.5%	58,865	1.5%	3,967,358	98.5%
NOORISTAN	32,270	3.4%	925,680	96.6%	73,809	7.7%	884,141	92.3%
PAKTIKA	72,622	3.8%	1,814,997	96.2%	527,708	28.0%	1,359,911	72.0%
PAKTYA	116,987	21.3%	431,101	78.7%	246,137	44.9%	301,951	55.1%
PANJSHER	7,013	1.9%	367,199	98.1%	39,407	10.5%	334,805	89.5%
PARWAN	34,512	6.2%	523,603	93.8%	112,330	20.1%	445,785	79.9%
SAMANGAN	76,644	5.7%	1,268,537	94.3%	123,062	9.1%	1,222,119	90.9%
SAREPUL	94,386	6.3%	1,402,916	93.7%	137,020	9.2%	1,360,282	90.8%
TAKHAR	203,580	16.4%	1,038,632	83.6%	222,458	17.9%	1,019,754	82.1%
UROZGAN	104,445	8.0%	1,207,103	92.0%	529,291	40.4%	782,257	59.6%
ZABUL	34,364	2.0%	1,667,501	98.0%	327,080	19.2%	1,374,785	80.8%
TOTAL	**3,198,139**	**5.0%**	**60,995,226**	**95.0%**	**8,189,126**	**12.8%**	**56,004,239**	**87.2%**

The reason why the RF classification overestimates the area under wheat is likely the strong class imbalance in the in-situ samples. 37.1% of all the available samples in the crop library represent wheat cultivation area, while the total fraction of all non-crop classes (including grassland and fallow area) is 19.8%. Almost twice as many wheat samples are therefore available than from all non-crop classes together (Table S1). Table S4 shows that non-crop samples are frequently misclassified as wheat, resulting in a particularly low Producer Accuracy of wheat with RF classification. The distribution of training samples does not reflect the true underlying distribution of wheat versus other classes, and the obtained crop area is therefore biased towards wheat (Wang et al., 2019). To improve the accuracy of the RF crop map, more 'non-crop' samples should be added to the library.

Table S6. Total area per class at the scale of the entire country (DTW and RF classification). All non cropland classes (Non-crop, Grassland, Fallow, Wetland and Forest) are uniformly presented as 'Non-crop' in Figure S3, S4 and S5.

	DTW	RF
	Cropland (hectares)	
Wheat	1,086,285	3,000,637
Other crop	635,334	2,474,009
Orchard	405,450	787,069
Beans	320,677	1,182,138
Potatoes	265,947	89,825
Fodder_crop	157,989	198,303
Maize	111,655	169,593
Rice	83,387	63,099
Cotton	81,826	203,745
Vineyard	16,830	8,272
Soya_beans	12,620	6,764
Oats	10,778	5,285
Rye	9,363	385
Total	**3,198,139**	**8,189,126**
	Non cropland (hectares)	
Non-crop	34,040,098	34,542,044
Grassland	16,353,768	13,337,940
Fallow	9,683,853	7,441,853
Wetland	815,031	530,590
Forest	102,476	151,811
Total	**60,995,226**	**56,004,239**

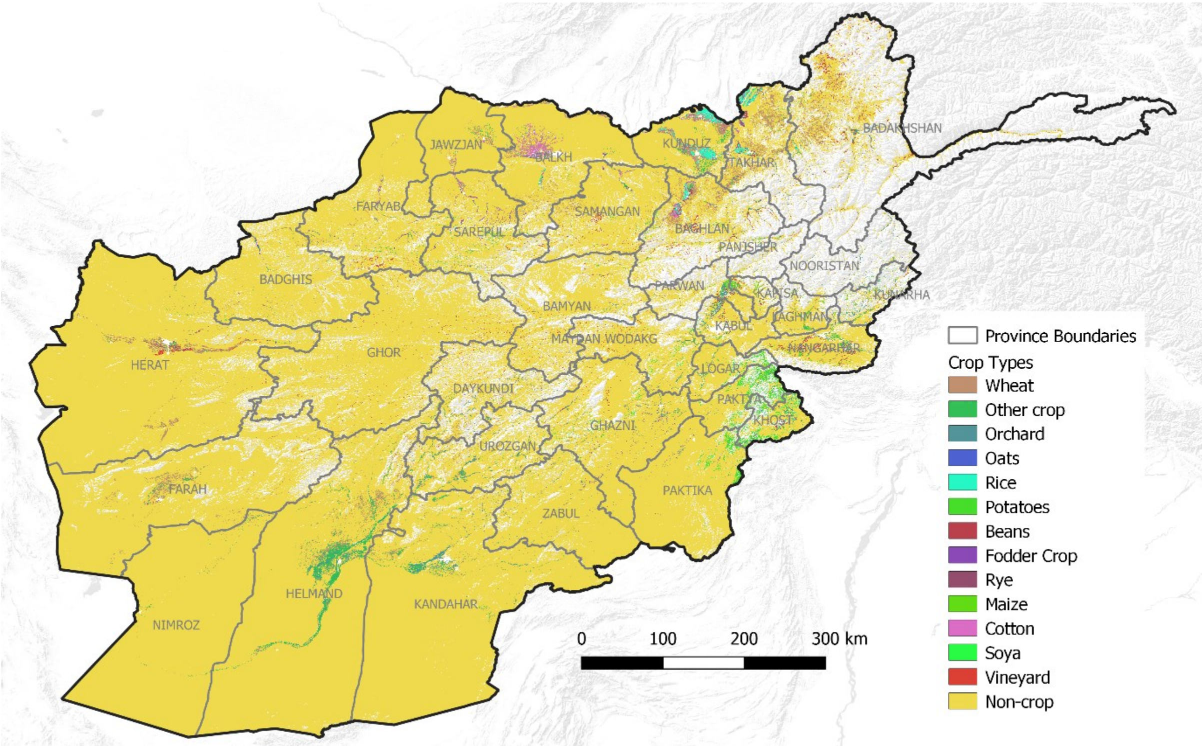

Figure S4. Crop type map for Afghanistan 2021 based on DTW classification. Areas with transparent colour represent areas that were by default considered as unsuitable for agriculture, such as areas at elevations above 3600 m or with slopes steeper than 30°.

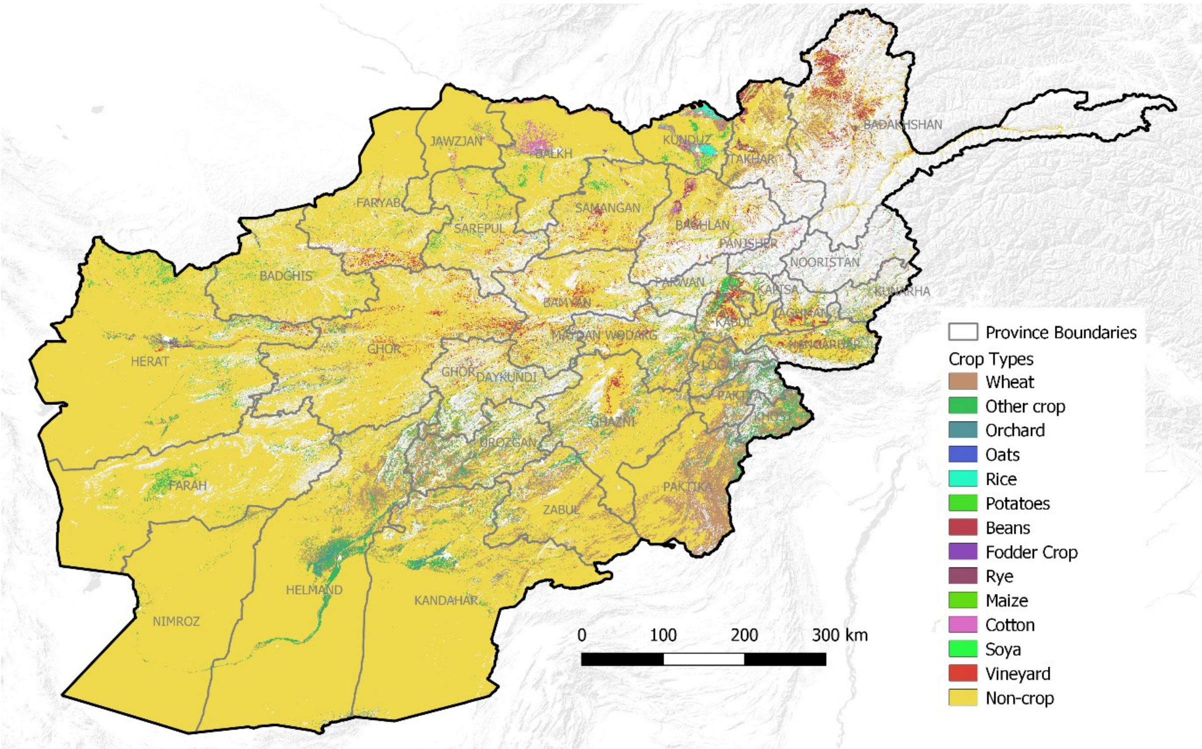

Figure S5. Crop type map for Afghanistan 2021 based on RF classification. Areas with transparent colour represent areas that were by default considered as unsuitable for agriculture, such as areas at elevations above 3600 m or with slopes steeper than 30°.

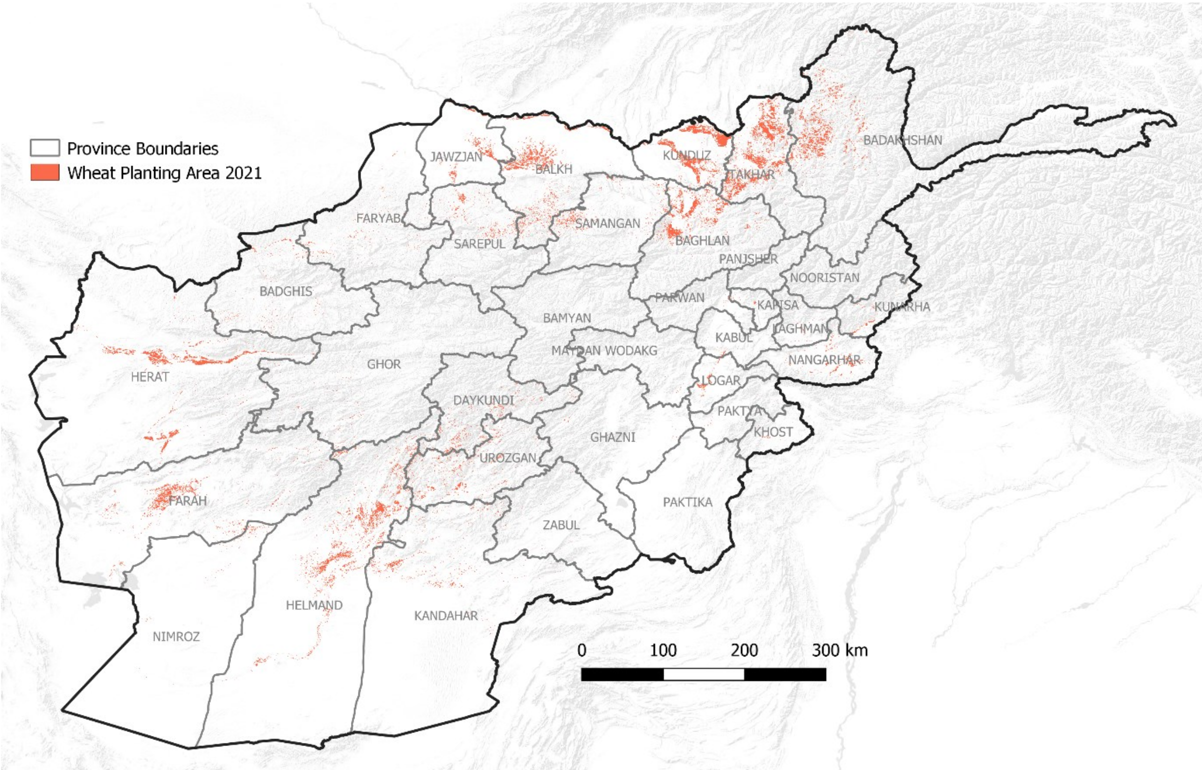

Figure S6. Wheat planting areas of the year 2021 (DTW classification).

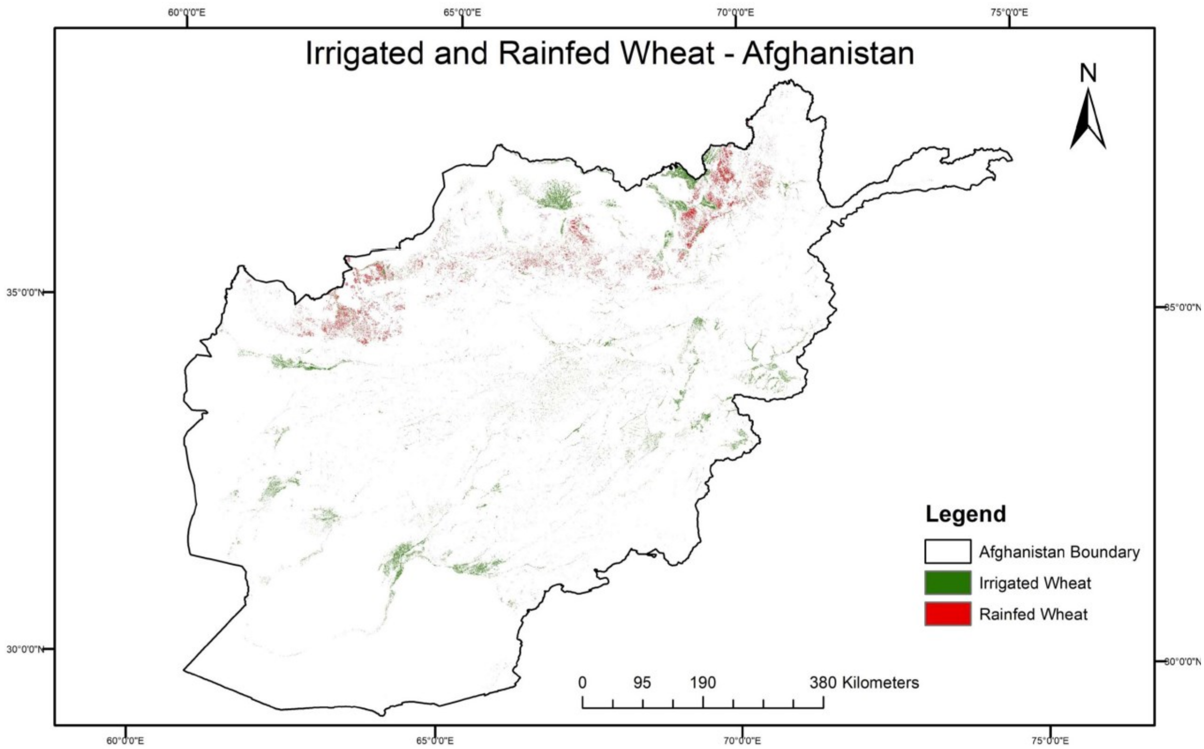

Figure S7. Irrigated and Rainfed Wheat area of Afghanistan, according to Tiwari et al., 2020*. Image Copyright:* Tiwari et al., 2020.

Conclusions

This application demonstrates the benefit of QA/QC, facilitated through the administrator tool of the EOSTAT CropMapper, with regards to resulting classification accuracies. Applying the available option for automatic quality checking of in-situ samples improved the overall classification accuracy by on average 14.3%. Furthermore, as the application to Kashkardarya region in Uzbekistan has already shown, higher classification accuracies can be achieved with DTW than with RF classification, provided that the QA/QC is carefully performed. Of course, the resulting overall classification accuracy of 61.9% is still low, but can be explained by the difficult political and security situation in Afghanistan in the year 2021, which impeded the collection of high quality field data. Note that the actual classification accuracy is likely higher than 61%, because non cropland samples are underrepresented in the validation data. The producer accuracy for no-crop samples is much higher than for most other classes (87%-89%, Table S3 and Table S4), but in relation to its actual frequency only few samples of this class are available for validation.

Overall, this application can be considered as a first attempt to obtain crop statistics for all 34 provinces of Afghanistan based on a remote sensing data analysis. The feasibility of carrying out this task with the EOSTAT CropMapper has been demonstrated. While for most crops the quantity of available in-situ samples is sufficient for DTW classification, further efforts are required to improve the quality of the samples.

6 References

Belgiu, M., Csillik, O., 2018. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. Remote Sens. Environ. 204, 509–523. https://doi.org/10.1016/j.rse.2017.10.005

Csillik, O., Belgiu, M., Asner, G.P., Kelly, M., 2019. Object-based time-constrained dynamic time warping classification of crops using Sentinel-2. Remote Sens. 11. https://doi.org/10.3390/rs11101257

Defourny, P., Bontemps, S., Bellemans, N., Cara, C., Dedieu, G., Guzzonato, E., Hagolle, O., Inglada, J., Nicola, L., Rabaute, T., Savinaud, M., Udroiu, C., Valero, S., Bégué, A., Dejoux, J.F., El Harti, A., Ezzahar, J., Kussul, N., Labbassi, K., Lebourgeois, V., Miao, Z., Newby, T., Nyamugama, A., Salh, N., Shelestov, A., Simonneaux, V., Traore, P.S., Traore, S.S., Koetz, B., 2019. Near real-time agriculture monitoring at national scale at parcel resolution: Performance assessment of the Sen2-Agri automated system in various cropping systems around the world. Remote Sens. Environ. 221, 551–568. https://doi.org/10.1016/j.rse.2018.11.007

Fan, J., Zhang, X., Zhao, C., Qin, Z., De Vroey, M., Defourny, P., 2021. Evaluation of crop type classification with different high resolution satellite data sources. Remote Sens. 13, 1–17. https://doi.org/10.3390/rs13050911

FAO, 2021. FAO Afghanistan Newsletter, 4th Quarter 2021. Kabul, Afghanistan. https://www.fao.org/3/cb8519en/cb8519en.pdf

FAO, Hydrosolutions Ltd., 2022. EOSTAT CropMapper – User Guide. Rome, Italy; Zurich, Switzerland.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Orynbaikyzy, A., Gessner, U., Conrad, C., 2019. Crop type classification using a combination of optical and radar remote sensing data: a review. Int. J. Remote Sens. 40, 6553–6595. https://doi.org/10.1080/01431161.2019.1569791

Orynbaikyzy, A., Gessner, U., Mack, B., Conrad, C., 2020. Crop type classification using fusion of sentinel-1 and sentinel-2 data: Assessing the impact of feature selection, optical data availability, and parcel sizes on the accuracies. Remote Sens. 12. https://doi.org/10.3390/RS12172779

Remelgado, R., Zaitov, S., Kenjabaev, S., Stulina, G., Sultanov, M., Ibrakhimov, M., Akhmedov, M., Dukhovny, V., Conrad, C., 2020. A crop type dataset for consistent land cover classification in Central Asia. Sci. Data 7, 1–6. https://doi.org/10.1038/s41597-020-00591-2

Tatsumi, K., Yamashiki, Y., Canales Torres, M.A., Taipe, C.L.R., 2015. Crop classification of upland fields using Random forest of time-series Landsat 7 ETM+ data. Comput. Electron. Agric. 115, 171–179. https://doi.org/10.1016/j.compag.2015.05.001

Tiwari, V., Matin, M.A., Qamer, F.M., Ellenburg, W.L., Bajracharya, B., Vadrevu, K., Rushi, B.R., Yusafi, W., 2020. Wheat Area Mapping in Afghanistan Based on Optical and SAR Time-Series Images in Google Earth Engine Cloud Environment. Front. Environ. Sci. 8. https://doi.org/10.3389/fenvs.2020.00077

Wang, S., Azzari, G., Lobell, D.B., 2019. Crop type mapping without field-level labels: Random forest transfer and unsupervised clustering techniques. Remote Sens. Environ. 222, 303–317. https://doi.org/10.1016/j.rse.2018.12.026

Zhang, Z., Tang, P., Huo, L., Zhou, Z., 2014. MODIS NDVI time series clustering under dynamic time warping. Int. J. Wavelets, Multiresolution Inf. Process. 12, 1461011. https://doi.org/10.1142/s0219691314610116