UCL Université catholique de Louvain

Faculté des Bioingénieurs

Near-real-time crop mapping in Ningxia (China) using Sen2-Agri system: Contributions of Sentinel-2 and Gaofen-1 satellite data

Présenté par Mathilde De Vroey

Promoteur(s) : Prof. Pierre Defourny (UCL/ELI/ELIE)

Prof. Jinlong Fan (CMA/NSMC)

Lecteurs : Prof. Claude Bragard (UCL/ELI/ELIM)

Dr. Sophie Bontemps (UCL/ELI/ELIE)

Mémoire de fin d'études présenté en vue de l'obtention

du diplôme de **Bioingénieur : sciences et technologies de l'environnement**

AGRO LOUVAIN

Année académique 2017-2018

Acknowledgements

This master thesis was carried out under the supervision of Professor Pierre Defourny from the Earth and Life Institute, UCL and Professor Jinlong Fan from China Meteorological Center (CMA). I would like to thank Professor Defourny for the advice, the support and the guidance he provided throughout the year. I thank Professor Fan for welcoming me in China and providing support and advice during my stay. I also thank Professor Fan's colleague, Dr. Zhang Xiao Li for organising the field campaigns in Ningxia and his students Xu Qi and Li QiLiang for their hard work collecting field data, for welcoming me and for helping me every day in China. I would like to thank Nicolas Bellemans for his availability and his precious help and advice concerning Sen2-Agri. Finally, I thank my parents for their constant support and I thank my friends and amazing flat mates from the Oenokot for making this last year unforgettable.

Table of contents

List of figures

List of tables

Acronyms and abbreviations

CART	Classification and Regression Tree
CCI	Climate Change Initiative
CHEOS	China High-resolution Earth Observation System
CNSA	China National Space Administration
DDD	Dry and Desertified District
DT	Decision Trees
FAS	Foreign Agricultural Services
GF-1	Gaofen-1
GLAM	Global Agricultural Monitoring
JECAM	Joint Experiment of Crop Assessment and Monitoring
LAI	Leaf Area Index
LC	Land Cover
LUCAS	Land Use and Coverage Area Frame Survey
MACCS	Multi-sensor Atmospheric Correction and Cloud Screening
MARS	Monitoring Agricultural Resources
MLC	Maximum Likelihood Classifiers
MLHD	Mountainous and Loess Hilly District
MOST	Chinese Ministery of Science and Technologies
MRS	Multi-resolution Segmentation
MS	Multi-Spectral
NDVI	Normalised Difference Vegetation Index
NIR	Near Infra-red
NN	Neural Networks
NOAA	National Oceanic and Atmospheric Administration
OA	Overall Accuracy
OB	Object-Based
PB	Pixel-Based
PSP	Phenological Sequential Patterns
RF	Random Forest
S2A	Sentinel-2 A
S2B	Sentinel-2 B
SAVI	Soil Adjusted Vegetation Index
Sen2-Agri	Sentinel-2 for Agriculture
SVM	Support Vector Machines
TWDTW	Time Weighted Dynamic Time Warping
USDA	US Department of Agriculture
WFMS	Wide Field Multi-Spectral
YERID	Yellow River Irrigated District

Introduction

The world population is growing continuously, increasing the pressure on our planet's ecosystems. Food production is one of the major components of this pressure. In this context agricultural monitoring has become more and more essential. One of the key tools of agricultural monitoring is satellite remote sensing, which has known tremendous developments over the last decades. The increasing quality of embarked sensors, the growing data storage and processing capacities and the ongoing improvements in terms of data exploitation and interpretation are of great benefit for the field.

Agricultural monitoring is essential for an adequate management of food production and distribution. Being able to estimate cropland areas and predict yearly yields is crucial for decision makers around the globe. Crop land and crop type classification, using remote sensing time series form an important tool, among others, to deliver such information. Satellites are a valuable resource for obtaining information on a large scale and with a high revisit frequency, which are two important requirements for agricultural monitoring. Moreover, the timeliness of the information is another crucial concern. The earliest accurate information can be delivered, the most useful it is for decision makers and politics.

A first main challenge of crop mapping is to differentiate annual cropland from other land uses and especially from other green land covers. Working with time series instead of single date images allows to extract temporal features, which are a great asset for classifying cropland and identifying different crop types. Another great issue is the spatial resolution. Often, sensors with a high revisit frequency and a large swath width offer a coarser spatial resolution. Using such sensors to map heterogenous agricultural areas is problematic when the field areas are inferior to the sensor's pixel size, causing mixed pixels.

A large variety of crop mapping methods, at different scales and showing various levels of accuracy can be found in the literature. From the first use of satellites for agricultural monitoring in the 1970's to the latest research, crop mapping strategies have evolved tremendously. Methods have been adapted continuously, to the increasing performances of embarked sensors and to increasing computational capacities. Nowadays, several international initiatives are set up to provide local and global agricultural monitoring. One of the most important efforts has been the Global Agricultural Monitoring (GLAM) Project developed by NASA, UMD (University of Maryland) and USDA (Becker-Reshef *et al.,* 2010).

Recently, the launch of ESA's Sentinel-2 satellites has brought new perspectives for large scale agricultural monitoring. With its swath width of 290 km and 10m spatial resolution, combined with its 5 days revisit frequency, the constellation of two satellites presents a great potential for large scale crop mapping with high precision and accuracy.

This master thesis was initiated in the context of the Dragon 4 cooperation between ESA and the Chinese Ministry of Science and Technology (MOST). This Sino-European program gathers Chinese and European scientists in the field of remote sensing and aims at sharing and combining data and knowledge to trigger progress. On the European side, the recently launched Sentinel-2 satellites provide unprecedented monitoring capacities in terms of spatial resolution, swath width and revisit frequency. The Sentinel-2 for Agriculture (Sen2-Agri) system aims at

fully exploiting those capacities, by providing four relevant earth observation products for agricultural monitoring. The system downloads Sentinel-2 imagery and can provide a monthly cloud free image composite, a binary cropland mask differentiating cropland from other land covers, a crop type map, classifying the main crops of a region and a vegetation status map with LAI or NDVI values (Inglada *et al.*, 2015; Matton *et al.*, 2015; Valero *et al.*, 2016). The Chinese Gaofen-1 satellite presents technical capacities similar to Sentinel-2. Gaofen-1 imagery could potentially complement Sentinel-2 time series and thereby increase the data frequency over time and provide more accurate products.

The overall objective of the thesis is to evaluate the crop mapping capacities of Sentinel-2 and Gaofen-1 (a similar Chinese satellite). A first goal was therefore to analyse and compare both satellites in terms of data quality and assess their compatibility. Secondly, a specific focus was set on the performances of the Sentinel-2 for Agriculture system for crop mapping an irrigated agricultural area in the Ningxia Hui Autonomous region in China. This objective was reached through experiments on the cropland mask product, with or without *in situ* data and on the crop type map at the end of the season and along the growing season. The sensitivity of both products to satellite imagery, training data and the classification algorithm was assessed. Moreover, particular attention was given to the validation data and methods.

This master thesis is written in eight chapters. First of all, the state of the art gives a summary of satellite-based agricultural monitoring, describing the evolutions in the field, from the earliest research to the latest methods. The chapter covers several remote sensing classification methods and the use of remote sensing time series for crop mapping. A particular focus is then set on the Sen2-Agri system. Finally, the two concerned sensors are presented. The specific objectives of the thesis are presented in a second chapter.

The third chapter gives a brief presentation of the Ningxia Hui Autonomous region in China, which is the study area selected for this work. In a fourth chapter, Sentinel-2 and Gaofen-1 are evaluated in terms of data quality and performances and their compatibility is assessed. Chapter five and six are dedicated to crop mapping experiments with the Sen2-Agri system in the 2017 growing season. Chapter five presents the data sources, the detailed workflows of Sen2-Agri for crop mapping and the different experiments that were carried out. Chapter 6 then describes and discusses the results of those crop mapping experiments.

Chapter seven describes the near-real-time crop mapping performed for the 2018 growing season, focussing on the field campaign sampling strategy and then presenting and discussing the results. The final chapter draws conclusions from all previous results and introduces future prospects.

1. State of the art

1.1. Satellite-based agricultural monitoring

In a context of demographic growth and increasing pressure on our planet's ecosystems, management of food production is a crucial concern. With the ever-increasing importance of food security, agricultural monitoring has become a driver of several international efforts. Agricultural monitoring can be defined as analysing qualitative and quantitative variations of crop over time and space in order to forecast crop supplies, assess plant damage, map soil productivity and calculate agricultural statistics. One of the major challenges, specific to agricultural monitoring is timeliness. As stated by the Food and Agriculture Organization (FAO) "The need for timeliness is a major factor underlying agricultural statistics and associated monitoring systems—information is worth little if it becomes available too late."

Therefore, satellite remote sensing is a key tool for large scale agricultural monitoring. Thanks to the latest and still ongoing progress in this field, crop maps are more and more precise and accurate. The first studies on satellite-based agricultural monitoring where carried out in the 1970s, as a reaction to some unanticipated food shortage events (Valero *et al.* 2016). The US Department of Agriculture (USDA) and the National Aeronautics and Space Administration (NASA) started a partnership with the National Oceanic and Atmospheric Administration (NOAA) to improve crop forecasting methods, using remote sensed data (Becker-Reshef *et al.,* 2010). From there, satellite-based agricultural monitoring has been the subject of many researches, aiming to develop cost-effective systems to fully exploit available data and deliver timely and accurate products. The major challenges researchers have faced in the development of such systems are the spatial resolution (Löw and Duveiller*,* 2014) and most of all the complexity and diversity of global crop systems and agricultural landscapes.

The high revisit frequency needed to obtain timely information on crop production often limits the spatial resolution of the data. This is problematic for per-field crop type mapping, in landscapes where fields are smaller than the pixel size, causing mixed pixels. Löw and Duveiller (2014) intend to define the optimal pixel size for crop identification, exploring the trade-off between pixel size and pixel purity. They find that it strongly depends on the local agricultural landscape and that it is impossible to determine a "one-size-fits-all" solution. Their results show that sensors like MODIS (250m) can help determine the major crop types in an area and sensors with higher spatial resolution, like Landsat 8 (30m) are best suited for object-based crop type classification. The most recently launched satellites, like Sentinel-2, combine a 5 days revisit frequency and sufficient spatial resolution (20m) for a field-level crop identification, potentially solving this resolution issue.

The greater challenge of remotely sensed agricultural information is due to the complexity and diversity of agricultural landscapes and crop systems around the globe. Selecting the appropriate satellite data, collecting relevant ground truth data, applying the right pre- and post-

processing steps and following the best suited methodology is a complex task for each region and cropping system. Moreover, the transferability of a methodology from one agricultural region to another is rarely easy. Consequently, agricultural monitoring processing chains often need to be locally adapted (Davidson *et al.*, 2017). More recent studies however tackle that issue and suggest globally applicable methods (Inglada *et al.*, 2015; Matton *et al.*, 2015).

Some of the largest international efforts delivered in the last decade for agricultural monitoring where GEOGLAM, Monitoring Agricultural Resources (MARS), Foreign Agricultural Services (FAS) and CropWatch, all aiming at "strengthening the international community's capacity to produce and disseminate relevant, timely and accurate agricultural information" (Valero *et al.* 2016). In prevision of the launch of Sentinel-2 satellites, ESA has funded the development of Sentinel-2 for Agriculture (Sen2-Agri), a project designed to automatically deliver near-real-time agricultural monitoring products from Sentinel-2 time series. The main challenge of Sen2-Agri project is transferability between different agricultural regions with various cropping systems and physical environments.

1.1.1. Remote sensing classification methods: past and still ongoing progress

Over the past decades, several kinds of methods have been developed and evaluated for classifying land cover and more specifically crop types. As mentioned before, methods were continuously adapted and refined, due to the diversity of agricultural landscapes, but also with the ever-increasing amount and quality of available data. Classification methods vary from unsupervised to supervised and from parametric to non-parametric, they can be pixel-based or object-based and several methods can be combined to form ensemble classifiers. This section gives an overview of different kinds of methods and develops the most recent and promising ones.

From un-supervised to supervised techniques

Un-supervised classification or clustering requires no *a priori* sample ground truth data, as the algorithm iteratively gathers similar pixels in clusters. The post-classification merging and splitting of clusters and the labelling, using land surface information is the limiting step. Automatic unsupervised clustering was the most used option in previous decade. For example, Mayaux *et al.* (2004) delivered a land cover map of the African continent, using data from several satellites and processing it through an unsupervised clustering algorithm producing 100 land cover classes. They labelled the classes afterwards, using about 50 vegetation maps, which is a time-consuming process. In general, partitioning algorithms are the most used unsupervised classifiers. The less time-consuming k-means and ISODATA algorithms are preferred for large datasets (Gómez *et al.*, 2016).

As an alternative to those methods, supervised classification techniques have gained in popularity and have widely proven to be satisfactory for large-area land cover mapping (Foerster *et al.*, 2012; Matton *et al.*, 2015; Xiong *et al.*, 2017; *etc.*). Supervised algorithms require more user's input, as they are based on *a priori* training data, assigned to classes. Here

the limiting step and largest source of error are the sampling of ground truth (*in situ*) data which needs to be representative of all land cover classes.

Parametric or non-parametric supervised classifiers

Supervised classifiers can be further divided into parametric and non-parametric algorithms. The first category implies Maximum Likelihood Classifiers (MLC) and discriminant analysis. Amongst them, MLC has shown to be the most accurate one (Castillejo-González *et al.*, 2009; Yang *et al.*, 2011). Satisfactory results were obtained using MLC in different contexts (Radoux *et al.*, 2014; Jia *et al*. 2014), with overall accuracies (OA) ranging between 80% and 90%. But most studies state their results could be improved by post-classification refinements or by using more advanced, non-parametric algorithms. In general, parametric supervised classifiers encounter limitations when working on large and complex areas and datasets with many features (Gómez *et al.*, 2016), because they assume the distribution (e.g. Gaussian) of the data. Non-parametric supervised classifiers or Machine Learning classifiers (Decision Trees (DT), Neural Networks (NN) or Support Vector Machine (SVM)) are generally better suited for processing such datasets. Each of these methods have their strengths and weaknesses which can vary from one scenario to another.

DT are the simplest non-parametric algorithm and therefore the least time-consuming and the easiest to interpret. They are based on recursive binary partitions, according to a set of optimized rules. Homer *et al.* (2004) successfully developed the 2001 National Land-Cover Database for the United States (85% OA), using DT to classify Landsat 5 and 7 imagery. They state decision tree classification offers an efficient, robust method to classify large quantities of information in documentable form and allows export of mutually exclusive rules generated by the classification into generic textual rule sets allowing users access to classification parameters. DT are however sensitive to noise, they tend to overfit and require large training samples.

NN and SVM are way more complex and computationally intense than DT, but they perform more accurately with very large dimension datasets. A drawback of those algorithms is the need for input parameters and their "black box" character, making it impossible for the user to know the decision rules. An important advantage however is the lower dependence on training data. NN and SVM work well with small training datasets and are resistant to missing data. (Gómez *et al.*, 2016) Several researches have shown the efficiency of NN (Furby *et al.*, 2008; Bagan *et al.*, 2005) and SVM (Carrão *et al.*, 2008; Dash *et al.*, 2007) with 0.91 and 0.88 OA respectively.

Machine learning ensemble classifiers

It is possible to combine machine learning classifiers to simultaneously benefit from their strengths and further improve the accuracy of a classification. This obviously makes the algorithms even more complex and the decision rules unknown to the users.

Kotsiantis and Pintelas (2007) define ensemble classifiers as "a set of classifiers whose individual decisions are combined in some way (typically by weighted or unweighted voting) to classify new examples". They state bagging and boosting are the most popular ensemble classifiers in general, boosting being considered stronger on noise-free data and bagging more

robust in noisy datasets. In a bagged ensemble, multiple bootstrapped training sets are generated from one given set. Those randomly selected sets are put through base model learning algorithms. Each of the yielded base models will classify new elements, which will be attributed to the class with the maximum votes. The boosting ensemble classifier also uses several subsets of training data with a base model learning algorithm. The difference with bagging is the weight boosting assigns to the different training data, depending on how well they were classified in a previous base model. The boosting ensemble works iteratively and assigns a heavier weight to those elements that were misclassified in the previous model, increasing their chances to be resampled. When studying the combination of bagging and boosting, using a voting methodology, Kotsiantis and Pintelas (2007) show an accuracy increase of 9-16% compared to tested base classifiers.

More recently, another extremely popular machine learning ensemble classifier, especially for crop type mapping is Random Forest (RF). As proposed by Breiman (1999), RF is a combination of decision trees where each classifier is generated using a random subset of training data. In each tree, nodes are split according to a subset of features randomly chosen from the overall predictor variables. Commonly, Classification and Regression Tree (CART) method is used to form the trees and the splits are done according to GINI index, which is a measure of the class homogeneity (Ok *et al.*, 2012). Each tree casts a unit vote for a class for each element (e.g. pixels), which will then be assigned to the most popular class in the forest. According to Breiman (1999), the generalization error always converges when the number of trees increases and the method is not sensitive to overfitting, which is one of the major advantages of RF relative to other machine learning algorithms. Another considerable asset of RF classifier is the limited amount of user-defined parameters. When working with RF only two parameters need to be optimized, namely the amount of trees to be grown (Ntrees) and the number of features used at each node (Mtry).

With the increasing popularity of the RF classifier, diverse studies have focused on this method, comparing it to other classifiers and assessing its sensitivity and robustness. Pal (2007) compared the results obtained with RF and SVM, using Landsat-7 data. He concluded that RF achieved similar OA to SVM, but the significant advantages offered by RF, for example handling categorical data, unbalanced data and data with missing values, makes it a more interesting classification algorithm. Moreover, RF delivers information about the variable importance and presents a way to detect outliers in a dataset. Pal (2007) also showed RF classifier is almost insensitive to overfitting, obtaining 88.37% and 88.02% OA for 100 trees and 1200 trees respectively.

RF largely outperformed MLC as well, yielding an OA of 85.89% compared to 77.96% with MLC (Ok *et al.*, 2012). This study using SPOT5 data, also assessed the object-based approach relative to the more traditional pixel-based one, assigning a class label of the majority class computed to each parcel. The object-based approach significantly increased the accuracy from 76.15% to 85.89% with RF and from 72.55% to 77.96% with MLC.

In their review of applications and future directions of random forest in remote sensing, Belgiu and Drăguţ (2016) draw several conclusions. First of all, they point out the efficiency of RF treating high dimensional datasets. This is mainly possible thanks to information on variable importance (VI) provided by the RF algorithm, allowing to select the most significative variables in a large set. The VI is particularly interesting for treating hyperspectral imagery, which are challenging because of their very high dimensionality, and when dealing with multi-source data, RF allows to select only the most relevant datasets, reducing the computational burden. The review also tries to assess the sensitivity of RF to training sample and data dimension but finds contradictory results and recommends analysing the sensitivity when using RF in future researches. They propose some feature selection methods that can be incorporated in an RF classification to further reduce the number of variables and recommend an iterative backward feature elimination procedure to reduce the number of less relevant variables until the internal accuracy no longer varies. Belgiu and Drăguţ (2016) mention many studies comparing RF to other machine learning classifiers and conclude RF performs better with multi-dimensional data, like hyperspectral or multi-source data, that it requires very few parameter settings and needs a shorter computational time.

Random forest classification has been used in many other studies. Waldner *et al.* (2015) successfully used RF for a land cover and crop type classification based on biophysical variables using multi-sensor high-resolution time series.

Pixel-based or object-based classification

In addition to the choice of the classification algorithm, the classification unit also needs to be determined. Traditionally, classifications are more often pixel-based (PB), meaning each pixel will be labelled independently. In order to reduce a certain "salt and pepper" effect, object-based (OB) classifications are interesting. This is particularly interesting for crop type mapping using high-resolution data. Pixels can be gathered into objects (plots in the case of crop mapping) before classification or it can be done in a post-classification filtering procedure. Some studies, like Waldner *et al.* (2015), compare both approaches and find OB classification yields better results. Others find PB produces a higher OA than OB with a pre-classification filtering approach, but the post-classification OB filtering delivers a better visual effect, its only limitation being detection of smaller fields (Valero *et al.*, 2015). This could however be resolved by higher resolution data like Sentinel-2 (10 m).

1.1.2. Crop mapping using remote sensing time series

In most land cover products, local, regional or global, cropland is not considered as a single land cover class, but is part of a mosaic class together with natural vegetation. Actually, cropland shouldn't be considered as a land cover class but rather as a land use category (Valero *et al.*, 2015). In order to produce a crop type map, it is helpful to first create a binary cropland mask, to distinguish cropland from other land use categories. Once that is achieved, different crop types can be identified and classified using above mentioned techniques. It is however important to consider the specific aspects of crop type mapping. First of all, the spectral signatures of some crops are very close, making them hard to discriminate using single date

images. The seasonal characteristics or the temporal profile of crop growth can be exploited to differentiate crop types.

Temporal series approaches have been demonstrated as superior to single-date methods for a range of applications (Gómez *et al.*, 2016), including crop mapping (Osman *et al.*, 2015; Xiong *et al.*, 2017). Despite the temporal character of the data employed, classifiers used in operational mapping are those typically used for single date spectral data classification. However, implementing time series data may alter the advantages of certain methods (Gómez *et al.*, 2016). It is thereby interesting to investigate alternative methods, specifically designed for crop mapping, using time series data to integrate the phenological characteristics of crops.

Early researches

Jakubauskas *et al.* (2002) describe the application of harmonic analysis to a single year (1992) of NOAA-Advanced Very High Resolution Radiometer (AVHRR) data, of 26 periods, for identification of several common crop types occurring within the western Great Plains. They use a biweekly cloud-free composite NDVI image and decompose it in a Fourier analysis. The phase and amplitude of the first two harmonic terms as well as the additive term are used as features for discriminant analysis. This early application of remote sensing time series analysis delivered relatively poor results, with a 52% OA, majorly due to the coarse 1 km spatial resolution of the input data.

Spectro-temporal profiles and phenological information

The use of spectro–temporal profiles for crop identification by satellite data was first proposed in the 1980s. Profile-based crop identification is based on the fact that profiles representing a specific crop are usually more similar than profiles representing different crops (Odenweller and Johnson, 1984). More recent studies have also been focusing on the spectro-temporal profiles of crops and on their phenological features. Foerster *et al*. (2012) proposes an efficient hierarchical classification algorithm that is based on spectro–temporal profiles of crop types and accounts for weather-induced inter-annual variations in the spectro–temporal behaviour, through the use of agrometeorological information and crop cultivation data from farming companies. Their study area is the Havel River catchment in the north east of Germany. They used cultivation data of 424 reference plots, acquired over the same period as the satellite images, to develop spectro-temporal profiles for each crop and validate the final crop type map. Based on agro-meteorological data, they calculated the average day of the year of each phenological stage (e.g. onset of flowering) to obtain the average phenological pattern for each individual crop type present in the Havel catchment. They then performed classification of 35 phenologically corrected Landsat images between 1987 and 2002, using the standard deviation of NDVI temporal profiles for each type and date as classification threshold. As comparison, a traditional MLC was performed on the Landsat images as well. The OA obtained with MLC was higher than the one using the phenological patterns, 72.8% and 65.7% respectively. The writers claim the poor performance is majorly due to the low number and bad timing of the image acquisition and to the occurrence of exceptional weather conditions. Another conclusion is that the average phenological profiles developed in this study are specifically related to one

particular agricultural region and should therefore be adapted when mapping another region. They state that the increasing availability of cloud-free images will facilitate the continuous refining of the spectro–temporal profiles for common crop types and different agricultural regions under current climatic conditions and is expected to improve the classification accuracy of crop type maps using these profiles.

New studies have indeed been carried out on the use of phenological profiles for crop type mapping. Bargiel (2017) introduces a novel approach to incorporate the full knowledge of the crops' phenology using a sequence-based classification approach. This approach has already been used in the past (Julea *et al.*, 2011), but Bargiel (2017) transfers this approach to a supervised classification with the incorporation of phenological knowledge to classify crops based on phenological sequential patterns (PSP).

A region around Hanover in northern Germany was chosen as study area. Field campaigns were organised during growing seasons of 2014-2015 and 2015-2016, recording crop type and phenology (according to BBCH-scale, Meier and Bleiholder (2006)) for each field. Furthermore, 99 dual polarized Sentinel-1A images (between 13/10/14 and 8/10/16) were available for the study. Phenological stages were defined, based on the phenological appearance of the plants, so that the plants' appearance doesn't change much within a stage, but varies strongly between two stages. In this study, six phenological stages were characterized for all the crop types. For each crop type, a growing season's day of year was assigned to the phenological stages, resulting in specific phenological sequences for each crop type (Figure 1).

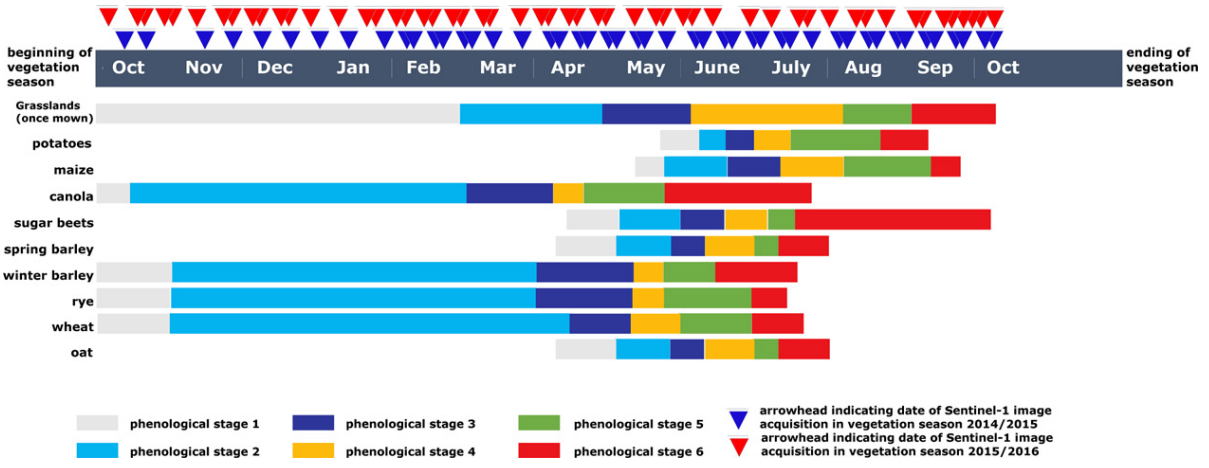

Figure 1: Phenological sequences of crops in the study area (Bargiel, 2017).

For the classification, the probability $\mathbf{p(y_s)}$, that a certain pixel belongs to a certain class or crop-type **y** at a certain phenological stage **s**, is computed for each pixel of the investigated area, using Naïve Bayes and RF. The probability $\mathbf{y_{seq}}$ of a pixel belonging to a class' phenological sequence is then calculated. The mean probability of a class is divided by the coefficient of variance **c** in order to penalize classes with a high variation within the probabilities of a sequences single stages. Moreover, to take into account the spatial context of each pixel, the probability is averaged over the eight adjacent neighbours of the pixels, resulting in Eq. 1. Each pixel is then labelled with the crop type maximizing y_{seq}.

$$y_{seq} = \frac{1}{n+1} \sum_{i=1}^{n+1} \frac{\frac{1}{k} \sum_{s=1}^{k} p(y_s)}{c} \quad \text{Eq. 1}$$

Bargiel (2017) evaluated the PSP method by comparing it to MLC and RF, without the Phenological Sequence Patterns (PSP) approach. The overall results obtained with the PSP approach were better than with MLC or RF for most crops. PSP strongly improved the accuracy for different types of cereals for example. Bargiel (2017) suggests several further researches. He recommends testing the PSP approach on Sentinel-1 data fused with Sentinel-2 data, evaluating the contribution of each phenological stage to the classification result and calculating the feature importance within each stage to limit the classification to the most important ones. He raises the issue of adapting the phenological sequences to different agricultural regions and suggests several approaches that have been studied (Conradsen *et al.*, 2016; Mascolo *et al.*, 2016; Vicente-Guijalba *et al.*, 2015). Finally, he notes PSP based classification seems to be very sensitive to subtle anomalies in certain parcels (e. g. a potato field overgrown with weeds) but claims this establishes possibilities for detecting organic farming or drought monitoring.

Time Weighted Dynamic Time Warping

Another approach for crop mapping using remote sensing time series was suggested by Belgiu and Csillik (2018). They evaluate the performance of a Time-Weighted Dynamic Time Warping (TWDTW) method, based on freely available Sentinel-2 time series, using pixels and parcels as spatial analysis units and compare it to RF classification. The data used for this research was a Sentinel-2 Level 1C (Top of Atmosphere) which was pre-processed to obtain a Level 2A (Bottom of Atmosphere) temporal image selection with less than 10% cloud coverage, with visible bands 2,3,4 and NIR (8). The TWDTW method was tested on three sites, in Romania, Italy and the USA. The image selection resulted in 13 images for Romania, 13 for Italy and 21 for the USA.

Once the images were preprocessed, NDVI time series were created and training and validation samples were generated. The samples were randomly selected across the sites using visual interpretation based on the European Land Use and Coverage Area Frame Survey (LUCAS) (for Romania and Italy) and on the CropScape-Cropland Data Layer (CDL) by the United States Department of Agriculture. For the OB classifications, they used one of the most popular segmentation algorithms which is multi-resolution segmentation (MRS). MRS is a region-growing algorithm. It starts at the pixel level and hierarchically aggregates different pixels, until it reaches a certain level of homogeneity inside an aggregation. This homogeneity level and other scale parameters are defined by the user. Belgiu and Csillik (2018) preferred an object-based over a pixel-based classification because it is less intensive and spatially more consistent. They however note the issues of under-segmentation for spatially complex landscapes and over-segmentation when fields have an important inner heterogeneity.

TWDTW involves three major steps, namely (1) creating a smoothed temporal pattern for the NDVI (8-days frequency) for the training samples and the Sentine-2 imagery, (2) applying a logistic Dynamic Time Warping, adding a time-weight and (3) classifying the raster time series.

Dynamic Time Warping compares the similarity between two time series and aligns them in the most optimal way (Rabiner and Juang, 1993; Sakoe and Chiba, 1978). The result of this DTW is a dissimilarity measure between two temporal sequences. In this case, a time weight was added to penalize larger time warps more heavily, as suggested by Maus *et al.* (2016). The TWDTW function takes each pixel's NDVI time series and analyses it in conjunction with the temporal patterns identified for training samples. The output is a raster with layers containing dissimilarity measures for each temporal pattern. The R function *twdtwClassify* generates a categorical land cover map, by associating each pixel to the class with the lowest dissimilarity value.

RF and TWDTW produced similar results (except for the USA test area, where intra-class heterogeneity was problematic for TWDTW), but when the number of training samples was reduced to 3 per class, the TWDTW clearly outperformed the RF method for all 3 areas. This shows an important advantage of TWDTW in addition to its robustness to irregular temporal sampling and to the annual changes of vegetation phenological cycles. Additionally, they observed that NDVI is a good indicator, but soil background could cause variations in the computed phenological profiles. A solution for that matter would be to use the Soil Adjusted Vegetation Index (SAVI), the Transformed SAVI or the generalized SAVI.

Given the high classification accuracy obtained without human intervention and the reduced computational time, Belgiu and Csillik (2018) concluded that the TWDTW method applied to objects could be integrated into operational programs dedicated to cropland mapping and monitoring based on satellite image time series.

1.1.3. Sentinel-2 for Agriculture system

The unprecedented capacities of Sentinel-2 satellites are an important asset in many fields, and particularly in agricultural monitoring. The 10 m spatial resolution facilitates the generation of crop maps on field level. Sentinel-2 for Agriculture is a project, funded by the European space agency and led by Université Catholique de Louvain (UCL). By developing open source time series processing chains for large-scale production, it aims to fully exploit those new Sentinel-2 capacities. The objective is to provide four near-real-time products, namely a monthly cloud-free composite, a dynamic cropland mask, a crop type map and a crop status map. Before Sentinel-2 satellites were even launched, several studies were carried out to develop, prototype and validate the methods, using simulations of Sentinel-2 time series (Inglada *et al.*, 2015; Matton *et al.*, 2015; Valero *et al.*, 2015).

A first study, Inglada *et al.* (2015), aims to assess to what extent state-of-the-art supervised classification methods can be applied to high resolution multi-temporal optical imagery to produce accurate crop type maps at the global scale, that is for different landscapes, climatic areas and crop systems. Using 12 test sites with JECAM ground truth data and simulations of Sentinel-2 time series based on SPOT 4 (Take 5) and Landsat-8 data, they evaluated several crop mapping processing chains in order to select the most accurate and robust one. The final selection of strategies included: (1) an RF classifier; (2) an SVM classifier with an RBF kernel; (3) a Dempster–Shafer fusion of the previous approaches; (4) the best classifier with a mean-

shift smoothing; (5) the best classifier with a temporal regular resampling. The quantitative assessment of the results indicated that the RF classifier operating on linearly-interpolated time series yields the best results. They conclude Sentinel-2 will certainly bring improvements in the results thanks to the enhanced spatial resolution and the increased number of spectral bands, mainly in the red-edge spectrum. It will therefore be interesting to evaluate the usefulness of object-oriented approaches, the use of textures and the use of other spectral indices.

The second study, Matton *et al.* (2015), proposes and demonstrates an automated methodology for annual cropland mapping performing along the season in various agricultural systems using high spatial and temporal resolution remote sensing time series. The data in this study are the same as the previous ones. The proposed methodology includes: (1) an approach leveraging existing high-resolution baseline land cover information, if available, or a globally available baseline, if not; (2) an extraction of crop-specific spectro-temporal features targeting the most relevant reflectances to differentiate the cropland; and (3) a flexible and robust classification algorithm. The features selected for the classification are red and NIR reflectance values at the minimum NDVI stage and green, red and NIR reflectance values at the maximum NDVI stage. The algorithms assessed in this second study were (1) a K-means (unsupervised) clustering method, coupled with an automated labelling of clusters using a baseline map and (2) an iterative trimming method with an MLC. They evaluated the results, testing different baseline resolutions, comparing PB and OB approaches and assessing the performances along the season. The most accurate approach was the iterative trimming of the baseline to extract training data, necessary for the calibration of a maximum likelihood classifier. An OA of 71-99% seems compatible with the expected use of a cropland mask in an operational agricultural monitoring system. The robustness of the method showed to be quite convincing. However, the automation and genericity of the method could be further improved by a local adjustment of feature selection. The object-based approach delivered similar or slightly better results, but its added value needs to be balanced carefully with the significant computational costs. An alternative suggestion is to use objects from previous years, assuming minor annual changes. They state the performances rely heavily on the trimming process, which involves 2 major assumptions: (1) the spectral reflectance of each class fit a unimodal Gaussian distribution; (2) the quality of existing land cover data.

A third study, Valero *et al.* (2015), develops a method to create a dynamic binary cropland mask. The goal of this product is to map the cropland mask as early as possible in the season. The system requires local and *in situ* data. This cropland mask will be used as input data for the crop type map and crop status map. The study explores three strategies, namely a pixel-based approach, an object-based approach with a pre-filtering task (P-OB) and an object-based approach with a post-filtering task (OB-P). The algorithm used for the binary classification is RF. Visually and numerically P-OB gave the worst results. PB performed slightly better than OB-P numerically, especially for sites with a small number of images, significant cloud cover and a poor *in situ* data sample. However, OB-P gave better visual results, reducing the salt-and-pepper effect of PB.

Based on the reported findings, an open source system has been developed. Sentinel-2 for Agriculture system is now operational and delivers the four products for any region in the world, but it can still be improved. Further research is needed to validate or optimize the processing chains and adapt them to the diversity of agricultural landscapes and physical environments.

1.1.4. Performance metrics

As discussed in the previous sections, there are many classification methods used for crop mapping and their usage has evolved over decades. This is also true for validation methods. In order to evaluate the accuracy of a map, reference data, independent from the training dataset, is necessary to compare the classification's predictions to. The general basis for evaluation is the confusion matrix or cross-tabulation matrix, which gathers all the observed combinations of predicted and reference expressions. Table 1 shows the typical format of a confusion matrix, where n_{ij} is the number of observations categorized as i in the reference dataset and as j in the predicted crop map and J is the number of different classes.

Table 1: Typical format of a confusion matrix.

		Classified			
		$j = 1$	$j = 2$	…	$j = J$
Reference	$i = 1$	n_{11}	n_{12}		n_{1J}
	$i = 2$	n_{21}	n_{22}		n_{2J}
	…				
	$i = J$	n_{J1}	n_{J2}		n_{JJ}

From the confusion matrix, several performance metrics can be derived. The most common and simple ones are the **F1-scores** and the **overall accuracy**. The F1-score (Eq. 4) is derived from the precision and the recall for each class, which are computed through Eq.2 and 3 respectively. It gives an indication of the classification performance per class.

$$Precision = \frac{n_{ii}}{\sum_{j=1}^{J} n_{ij}} \quad \text{Eq. 2}$$

$$Recall = \frac{n_{jj}}{\sum_{i=1}^{J} n_{ij}} \quad \text{Eq. 3}$$

$$F1 - score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad \text{Eq. 4}$$

The overall accuracy gives the proportion of observations classified correctly. It can be computed through Eq. 5.

$$Overall\ Accuracy = \frac{\sum_{i=1}^{J} n_{ii}}{\sum_{i=1}^{J} \sum_{j=1}^{J} n_{ij}} \quad \text{Eq. 5}$$

Over the last decades one of the most popular metric to evaluate the performance of a classification has been **Kappa**. It is commonly attributed to Cohen (1960), but citations go back further (Galton, 1892). The standard Kappa and its variants have been used in many studies, however Pontius and Millones (2011) give a heavily critical view of the popular metric and

even strongly recommend the profession not to use Kappa anymore. They argue Kappa does not give any more relevant information than the overall accuracy to help make decisions about image classification. Kappa is often complicated to compute, difficult to understand and unpractical to interpret. Instead, they suggest two new performance metrics, namely quantity disagreement and allocation disagreement. They refer to disagreements between a reference dataset and a classified map in terms of merely quantity of pixels per class and in terms of allocation of classes respectively. In their study, they reveal four main reasons to definitively abandon Kappa in remote sensing applications and substitute them for both components of disagreement. First of all, as Kappa is a ratio, it can be undefined if the denominator is zero and in terms of interpretation, it is unclear if a large Kappa is attributable to a large numerator or a small denominator. Second, they state it is more useful to have information about two distinct components of disagreement, rather than a single statistic of agreement. By focusing on disagreement, we can try to explain the error, while agreement can always be caused by randomness. Third, and most importantly, Kappa intends to compare observed accuracy to a baseline of accuracy, which would be expected in a completely random classification. Pontius and Millones (2011) judge using randomness as a baseline for accuracy is not constructive and does not bring useful information for decision making. Last, they point out that the popularity of Kappa in remote sensing is probably attributable to its broad use in university courses concerning statistical theory. In statistical theory, comparing to randomness is important when sampling, which is completely different from crop map validation.

1.2. Recent European and Chinese satellite missions for agriculture in the optical domain

This study is carried out in the context of a Sino-European cooperation and uses optical data from two satellite missions, namely Sentinel-2 and the Chinese satellite Gaofen-1. For that reason, other satellite missions for agriculture in the optical domain, such as Landsat-8 are not described in this section. The focus is set on the theoretical description and the comparison of Sentinel-2 and Gaofen-1 satellites.

1.2.1. Sentinel-2

The Sentinel satellites are part of European Space Agency's Copernicus program. Sentinel-1a was launched back in April of 2014, followed by Sentinel-1b, on the same orbit, in April 2016. Those satellites are all-weather, day-and-night radar imaging for land and ocean on a polar orbit. Sentinel-2a was launched in June 2015 and complemented by Sentinel-2b in March 2017. This second Sentinel mission is equipped with high resolution multispectral (MS) sensors, providing information on land surface and vegetation for instance. Five other satellite pairs will complete the Sentinel constellation (Sentinel-3A and -5P are already in orbit). Sentinel-2 provides an unprecedented 10 m spatial resolution with a 5 days global revisit frequency, a 13 bands imager and a 290 km swath width.

1.2.2. Gaofen-1

Gaofen satellites are part of China National Space Administration's (CNSA) China High-resolution Earth Observation System (CHEOS) program, which aims to provide near-real-time observations for disaster prevention and relief, climate change monitoring, geographical mapping, environment and resource surveying, as well as for precision agriculture support (much like ESA's Copernicus program). Gaofen-1 carries a Panchromatic and Multi-spectral Camera (PMC) on board including panchromatic (4 bands) and multi-spectral imagers. It delivers images at a 16 m spatial resolution and has a swath width of 60 km. GF-1 is also equipped with a Wide Field MS (WFMS) imager with a swath width of 830 km, 4 bands and a 16 m spatial resolution. The PMC and the WFMS imager have a revisit frequency of 4 and 2 days respectively.

1.2.3. Comparison and compatibility

Sentinel-2 and Gaofen-1's major configurations are summarized in Table 2. Both satellites carry similar MS imagers, but Sentinel-2 provides an overall higher spectral resolution, due to its 6 red-edge bands and the Chinese satellite carries a panchromatic sensor. GF-1 has a lower spatial resolution and S-2 has a much larger swath width. Both satellites have similar (and very high) revisit frequency.

In their study for the Sen2-Agri system, Inglada *et al.* (2015) stated that the synergy between Sentinel-2 and Landsat 8 should be further exploited, namely to fill gaps in Sentinel-2 time series with Landsat 8 images. The Sentinel-2 for Agriculture system uses complementary

Landsat-8 data to deliver its products. Gaofen-1 data could potentially be used in the same way, when studying areas in China, but needs to be further analysed.

Table 2: Comparison of Sentinel-2 and Gaofen-2's parameters.

Satellite	**Sentinel-2**		**Gaofen-1**
Programme	ESA's Copernicus Program		China National Space Administration (CNSA)'s CHEOS program
Number of satellites	2 satellites (one orbit): Sentinel 2A and Sentinel 2B		1 satellite
Launch	2A: June 23, 2015	2B: Marsh 7, 2017	April 26, 2013
Sensors	Multi Spectral Imager (MSI)		2 Panchromatic imagers (PAN) and 2 MSI
Spectral resolution	13 spectral bands**:** 4 (NIR,R,G,B), 6 red-edge/shortwave IR, 3 atm correction bands		PAN, 4 MS (NIR,R,G,B)
Spatial resolution	10m, 20m, 60m		16 m
Swath width	290 km		60 km
Revisit frequency	5 days		4 days

2. Objectives

This thesis is carried out as part of the Dragon cooperation between ESA and the Chinese MOST. This Sino-European project supports cooperation between Chinese and European scientists in the field of remote sensing. In the context of the recent development of the Sentinel-2 for Agriculture system, and with the idea of jointly exploiting Chinese and European capacities for agricultural monitoring, this thesis presents two main objectives.

The first main objective of this thesis is to analyse and compare Gaofen-1 and Sentinel-2 imagery quality and assess their compatibility and the possibility of using their combined time series for crop mapping. Using Sentinel-2 and Gaofen-1 jointly would enhance the revisit frequency and generate denser time series for image classification. In order to do so, the data must be compatible in terms of spatial, geometrical, spectral and radiometric resolution.

The following objective is to assess the performances of the Sen2-Agri system for crop mapping an irrigated agricultural area. More precisely, the aim is to evaluate the quality of two of the four EO products delivered by the system, namely the cropland mask and the crop type map. First of all the sensitivity of the cropland mask product is assessed, evaluating the impact of the training dataset and of the algorithm, aiming to generate the most accurate cropland mask possible. Secondly, the capacity of accurately classifying the main and marginal crops of the study area are assessed, aiming to define the optimal characteristics of the training dataset. Another specific objective is to determine how early in the season an accurate crop type map can be generated in the Sen2-Agri system.

To sum up, this master thesis aims to give an evaluation of the current capacities for crop mapping with Sentinel-2 and Gaofen-1, more specifically with the Sentinel-2 for agriculture system, and to identify the potential means to enhance the accuracy of the crop maps.

3. Study area

The Ningxia autonomous region[1] is a 66.400 km² area in northern China, just beneath the Mongolian border (Figure 2). Amongst the 6 million people living in the region, 66,4% is rural population. In 2006, the farmland represented 1,1 million ha, including 402.000 ha of irrigated crops and 698.000 ha of rain-fed crops.

The Ningxia region is divided in 3 geomorphic and economic zones: the mountainous and loess hilly district (MLHD) in the south, the centrally located dry and desertified district (DDD) and the Yellow River irrigated district (YERID) in the northern plains. Geology and climate vary tremendously, from the south to the north: the topography declines, the temperature rises and the precipitation decreases. These gradients cause 3 different agricultural landscapes. In the MLHD 70% of the cultivated land consists of slope farmlands. Although this region has the highest precipitation, the hilly landscape and the uneven temporal distribution of rainfall hinder most agricultural activities. The DDD, as its name indicates, is a very arid area, mostly covered by grassland (77% of Ningxia's grasslands). It is however suited for Yellow river irrigation, given its proximity to the water body and its flat topography. The YERID is the most important rural part of Ningxia. It represents only one third of the total farmland in the region, but accounts for two thirds of Ningxia's grain production and agricultural production value. YERID's Gross Domestic Product represents almost 90% of the region's total. Rainfall is poor in this area, but the Yellow River allows an efficient irrigation of the fields.

Since 1985 the Ningxia hui region has been self-sufficient in terms of food supply and even developed from an import province to an export province. The main crops in Ningxia are grains, oil plants, vegetables, and pasture grass. Grains account for the largest farmland surface, namely 59% of the total farmland. Those grains are mainly wheat, rice, corn and minor cereals (peas, horse beans, haricot beans, grass peas, buckwheat, glutinous millet and millet etc.). The latter are mostly produced in the southern mountainous and central arid areas. In YERID, wheat, rice and corn are by far the major crops.

This study will focus on the YERID, as it is the most important agricultural district of the whole region. The study site covers an area of 65000 km², which corresponds to six Sentinel-2 tiles (Figure 3).

[1] All information about the Ningxia Hui Autonomous Area is based on the "Situation analysis of Ningxia Province" (Li *et al.*)

Figure 2: Situation of Ningxia Hui autonomous area.
Source: http://www.chinatoday.com/city/ningxia.htm

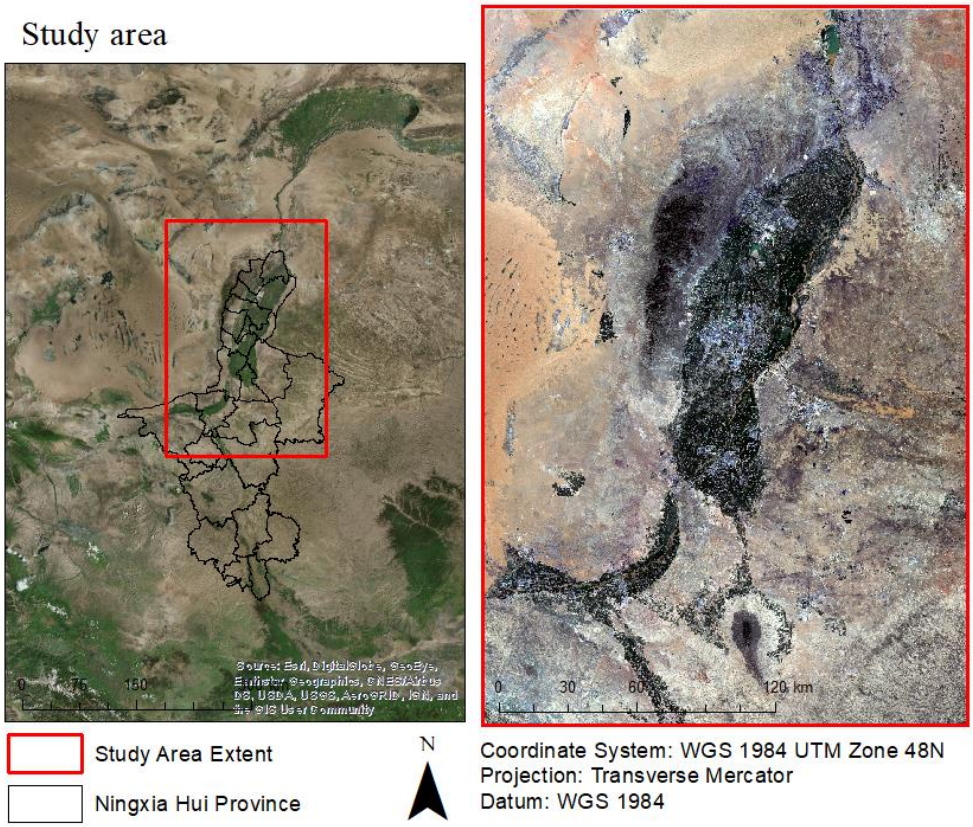

Figure 3: Study area extent.

4. Sentinel-2 and Gaofen-1 compatibility assessment

In the prospect of a compatibility study and an assessment of Gaofen-1 performances for crop mapping, a preliminary analysis of the images quality was carried out. The first part of this work thus consisted of analysing Gaofen-1 data quality in terms of spatial resolution, geometry and spectral and radiometric resolution and comparing it to Sentinel-1 images.

4.1. Satellite data acquisition

For the data quality assessment, 5 Gaofen-1 bottom of atmosphere images were acquired over the period of July and August 2017. Figure 4 illustrates the Gaofen-1 data acquisition and shows the Sentinel-2 images available over the same period. It is important to mention that Gaofen-1 data are not open access and are thereby much more complicated to acquire than Sentinel data.

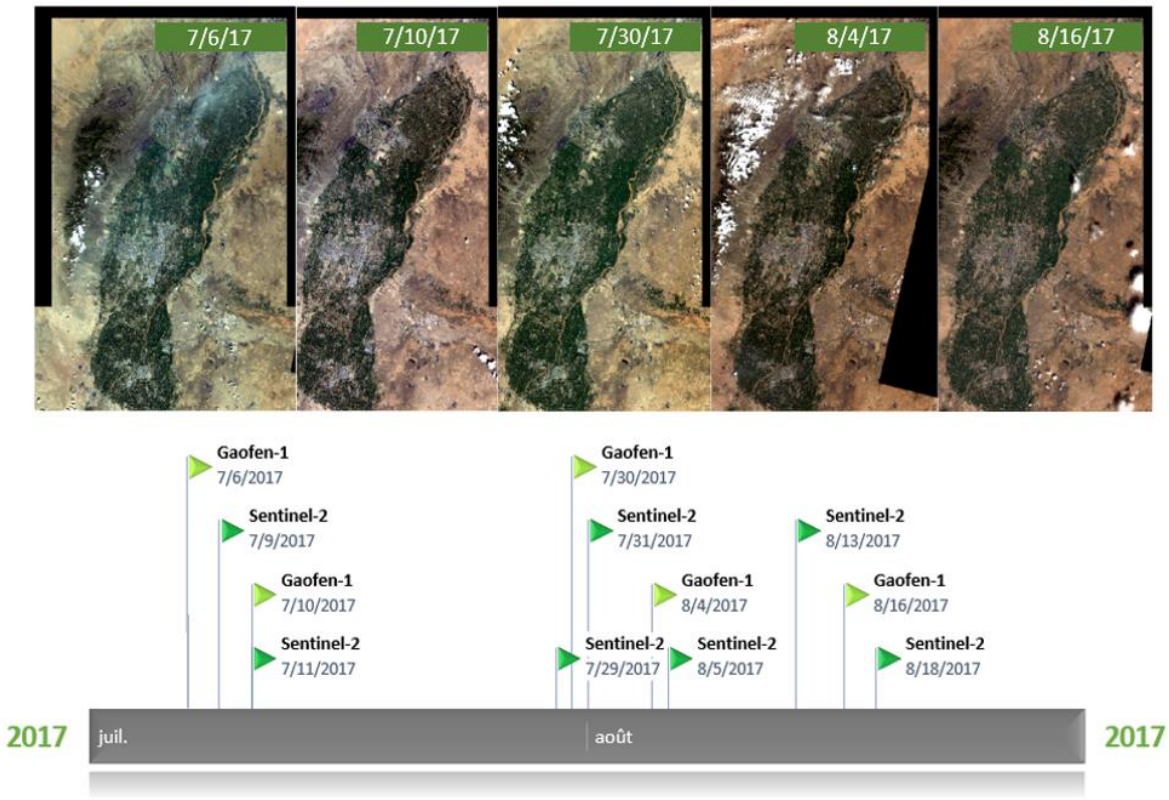

Figure 4: Gaofen-1 time series.

A bottom of atmosphere Sentinel-2 image of July 9th, 2017 was used for the comparison and the compatibility assessment. A third sensor was included for the spatial resolution assessment, namely Landsat-8, which delivers 30m pixels in the optical domain. The Landsat-8 image used in this study was acquired on July 11th, 2017.

4.2. Gaofen-1 quality assessment

4.2.1. Spatial resolution

First of all, the spatial resolution was assessed by visual analysis, comparing Sentinel-2, Gaofen-1 and Landsat-8. The spatial resolution of the Gaofen imagery is supposed to be 16 meters, but visually it seems to be closer to Lansat 8's 30 meters resolution than to Sentinel-2's 10 meters resolution (Figure 5).

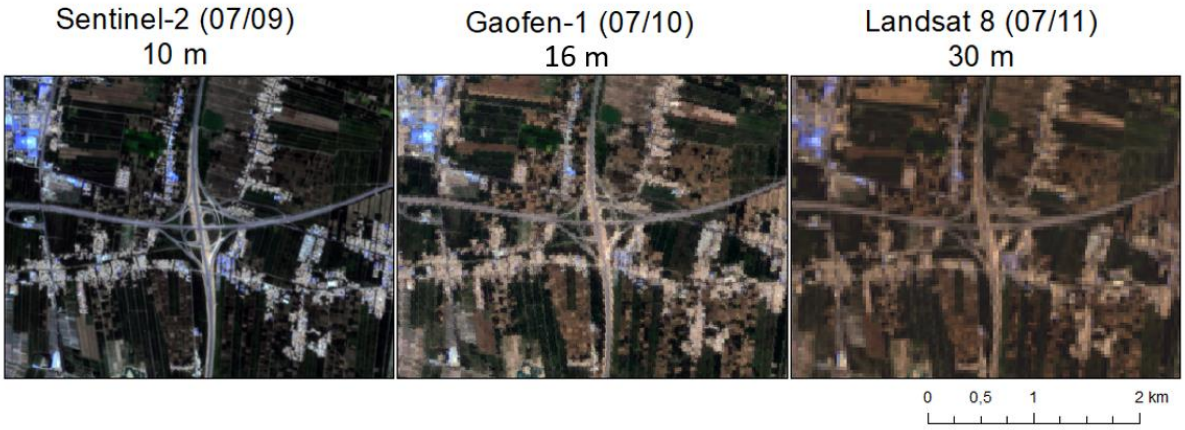

Figure 5: Illustration of the spatial resolutions of Sentinel-2 (10 m), Gaofen-1 (15 m) and Landsat 8 (30 m).

4.2.2. Geometry assessment

The geometry of Gaofen-1 was assessed visually as well. First of all, the coregistration between Gaofen-1 spectral bands seems to be right, as no border effects appear when zooming in. A second important aspect is the coregistration between the different Gaofen-1 images. To evaluate the inter-image coregistration, complex structures, like highway interchanges were selected and marked with a point on each image. The blue points on Figure 6 indicate the centre of the roundabout (left) or the south-eastern corner of the crossroad (right) on the 5 Gaofen-1 images. This illustrates a significant shift of one to five pixels between the Gaofen-1 images acquired at different dates. Finally, the geometrical compatibility with Sentinel-2 was assessed by marking the same reference points on a Sentinel-2 image. This revealed a geometrical discrepancy of 7-10 pixels between both satellites. On Figure 6, the red points represent position of the reference points on the Sentinel-2 image.

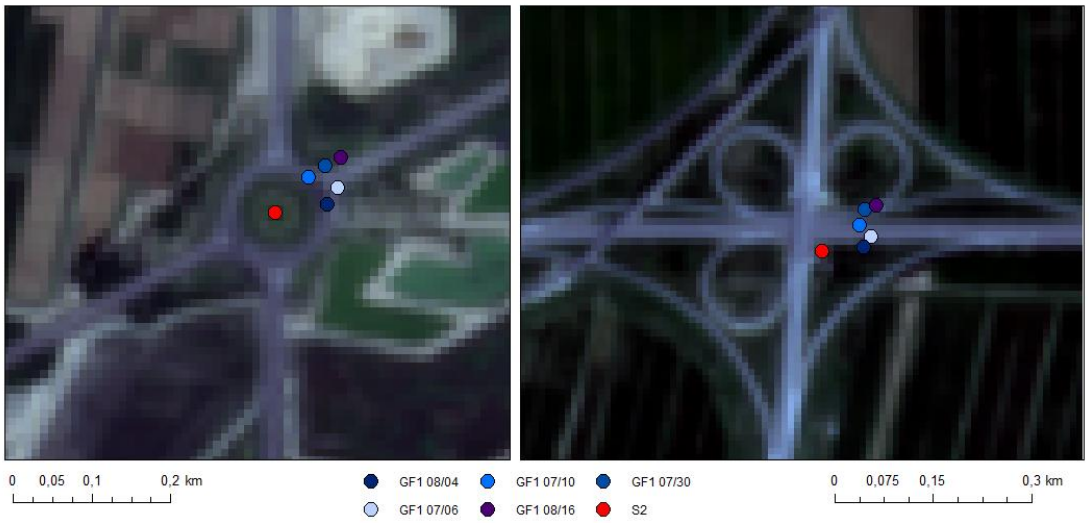

Figure 6: Gaofen images geometrical inter-date shift and shift compared to Sentinel images.

4.2.3. Spectral and radiometric resolution

In order to assess the spectral resolution, generating scatter plots and correlation matrices, a subsection of the Gaofen image was resampled to 10 m and shifted to geometrically match the Sentinel-2 image as much as possible. The subsection was chosen to cover a variety of landscapes such as crops, urban areas, bare soils and forest and grassland in the mountainous area (Figure 7). For the resampling a simple nearest neighbour method was applied. For the geometrical shift, 17 ground control points were defined, resulting in a total Root Square Error smaller than 1.

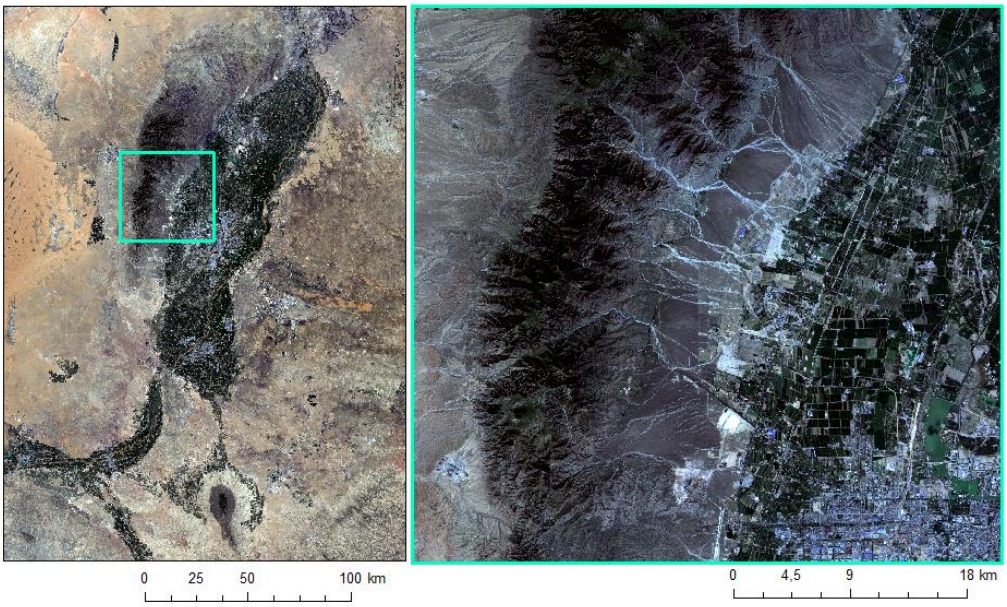

Figure 7: Subsection area defined for the assessment of the spectral and radiometric resolution of Gaofen-1 imagery, compared to Sentinel-2.

First of all, scatter plots were generated for both Sentinel-2 and Gaofen-1 images, between red (R) and near infra-red (NIR) band and between red and green (G) bands (Figure 8). The scatter plots between NIR and R and between R and G seem to have a similar general appearance for S-2 as for GF-1. The one noticeable difference is the abrupt truncation in the GF-1 scatter plots. This is probably due to prior calibration of the Gaofen-1 data.

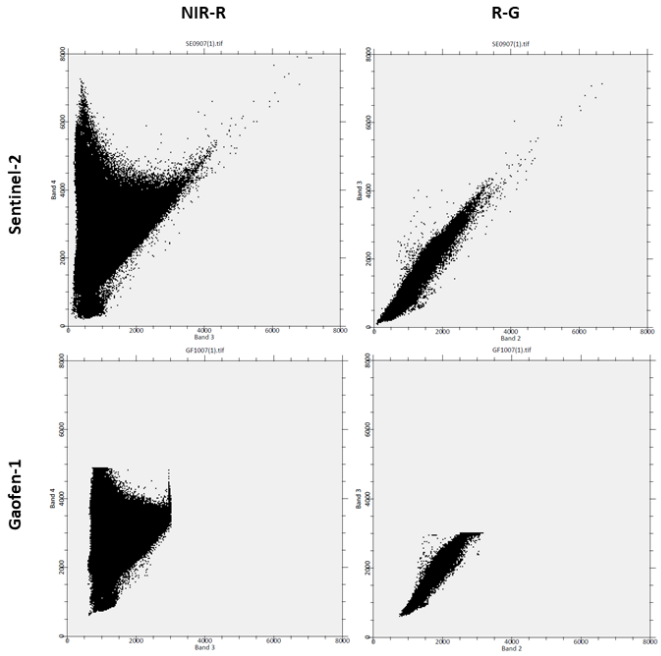

Figure 8: NIR-R and R-G scatter plots for Sentinel-2 and Gaofen-1, computed for equivalent subsections of both images.

Another particularity is the shape of the NIR-R scatter plots. The singular recess in the plots illustrates the high contrast in the landscape of the study area. It is an irrigated area inside a very arid region. Consequently, there is a very strong contrast between the spectral profiles of the irrigated agricultural area and the bare arid soil around it. Figure 9 illustrates the spatial distribution of the two parts of the NIR-R scatter plots. Crops and natural vegetation, like forest and grassland in the mountainous area reflect more in the NIR than in R, while bare soil and built-up surfaces reflect red and NIR with similar intensities.

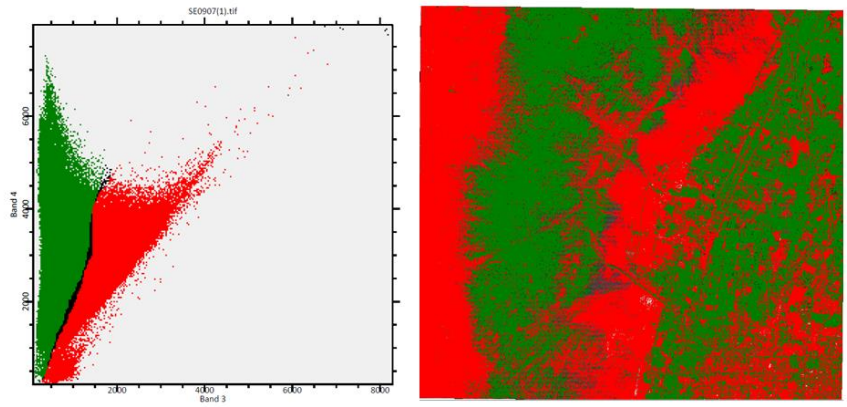

Figure 9: Spatial distribution of the NIR-R scatterplot (for Sentinel-2) over the concerned area illustrating the strong contrast between vegetation (green) and bare soil or built-up areas (red).

Secondly, the inter-band correlations for each sensor were computed for the same image subsections (Table 3). The overall tendencies are similar for both sensors. They both show the same high 98% correlations between G and B and between G and R. In terms of R-B correlation, the Chinese satellite is slightly higher (96%) than the European one (95%). There is a bigger difference in the correlations between the NIR band and the visible bands. In S-2, the NIR band is more correlated to all the visible bands than in GF-1, meaning GF-1 measures more sharply in the NIR range.

Table 3: Band correlation matrices of Sentinel-2 (left) and Gaofen-1 (right), computed for equivalent subsections of both images.

Correlation matrix S-2					*Correlation matrix GF-1*				
	Blue	Green	Red	NIR		Blue	Green	Red	NIR
Blue	1				Blue	1,00			
Green	0,98	1			Green	0,98	1,00		
Red	0,95	0,98	1		Red	0,96	0,98	1,00	
NIR	0,14	0,18	0,10	1	NIR	0,06	0,13	0,07	1,00

Finally, the correlations between same bands of both sensors are computed (

Table 4) and plotted (Figure 10). As expected from the similar inter-band correlations, the inter-sensor correlations are strong. The least correlated band is NIR (89%), which showed the most difference in terms of inter-band correlations. The scatter plots illustrate the relatively high inter-sensor correlations as well.

Table 4: inter-sensor correlations for each band.

S-2\GF-1 correlation	
Blue	0,90
Green	0,91
Red	0,93
NIR	0,89

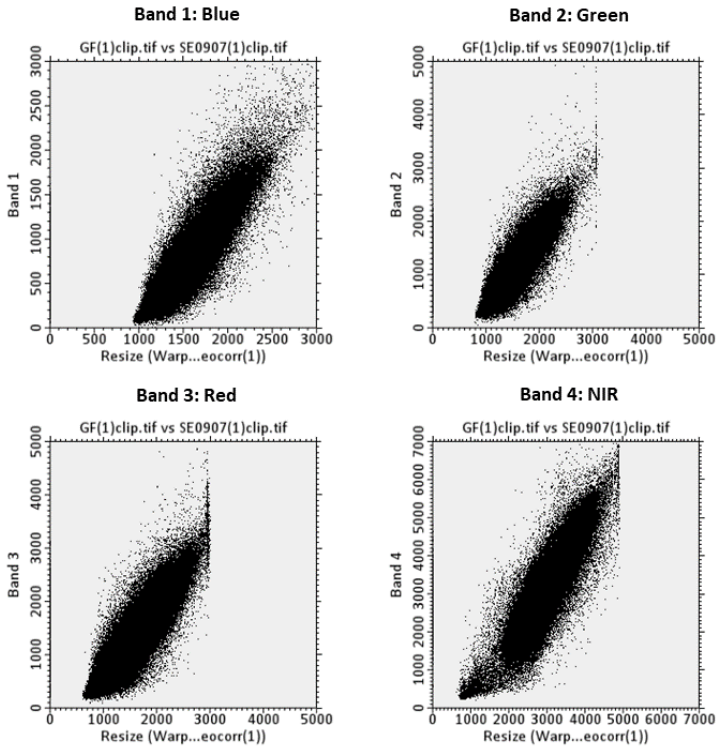

Figure 10: Scatterplots between equivalent bands of Sentinel-2 and Gaofen-1 imagery.

4.3. Compatibility assessment for crop mapping

Considering the results of the quality analysis, the Gaofen-1 and the Sentinel-2 sensors seem to have very similar spectral properties and the most significant differences between both satellites lie in the spatial resolution and the geometric quality. Before considering the combination of both sensors time series for crop mapping, a solution needs to be found to correct the Gaofen-1 data geometrically and make them match the Sentinel-2 data.

Once the inter-date coregistration and the geometrical discrepancy with Sentinel-2 are corrected, Gaofen-1 could potentially contribute to the quality of cropland maps in combination with Sentinel-2 data. Nevertheless, this study will not further tackle this issue and rather focus on crop mapping with Sentinel-2 data only.

5. Crop mapping (2017): data sources and methods

5.1. Data sources and pre-processing

The goal is to evaluate crop mapping performances of Sen2-Agri in the Ningxia study area for the 2017 growing season, through several experiments and analyses. Therefore, two types of data are required, namely reference training and validation data and satellite imagery. The reference data consists of ground truth data collected during field observation or by visual interpretation. The imagery used in this section is provided by Sentinel-2 exclusively.

5.1.1. Field data

The reference data needed for training and validation of the classification was collected through field observations. Reference points were gathered using GPS cameras and a Quick Photo Data Processor[2]. The method implies two major steps. First georeferenced pictures are taken along the study area's roads following a predefined itinerary. In a second phase, land cover and crop type classes are retrieved by screening the pictures through the processor. The final product of this process is a .txt file gathering all GPS points with corresponding classes, author, roadside (left or right), collecting dates and times and the corresponding image file names.

A first campaign was carried out in June 2017, providing about 1500 ground truth points with spatial reference and associated crop or other land cover classes. Those sample points are distributed over the irrigated area as shown on Figure 11. Based on those points and using Google Earth as visual reference, polygons were created, covering the plot extents, to increase the number of reference pixels per class. In addition, the *in situ* dataset was complemented by delineating additional non cropland samples to cover the whole range of landscape diversity. The dataset was randomly divided in a training and a validation subset, each containing 50% of the initial dataset. As most classes present a large number of samples, they are more or less equally distributed between the two subsets. Table 5 shows the number of polygons and estimated number of pixels per class in the training and validation datasets.

A second campaign was planned in June 2018, using the same GPS picture retrieval method. The sampling strategy however, was to be defined based on the results of the 2017 season experiments.

[2] This reference data retrieval method through GPS pictures was developed by Dr. Fan Jinlong and his students (CMA)

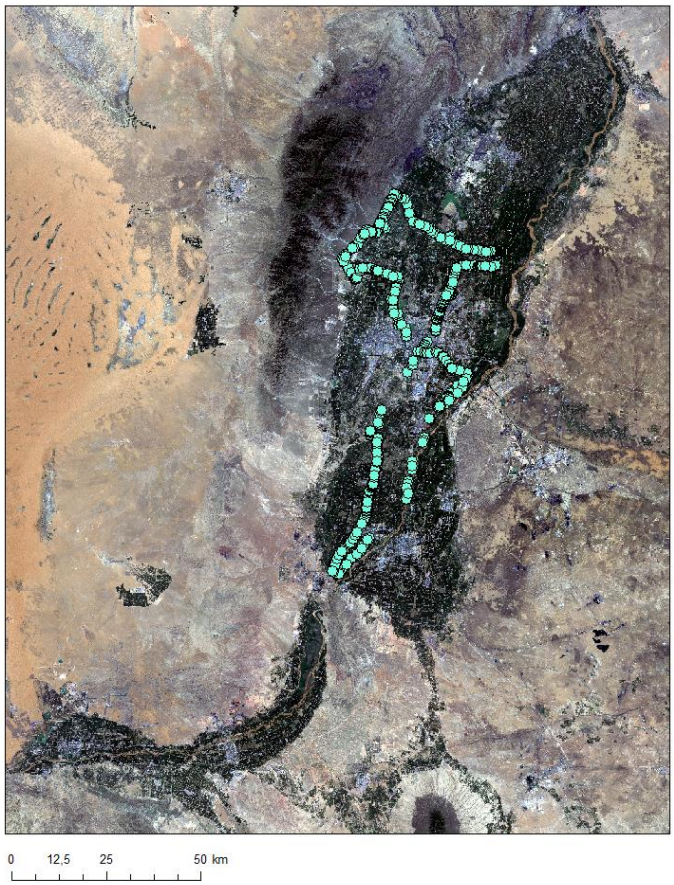

Figure 11: Field campaign sample points distribution in June 2017.

Table 5: Training and validation datasets for 2017 growing season.

		Samples		Pixels (10m)	
	Land Cover	**Training**	**Validation**	**Training**	**Validation**
Crops	Clover	19	39	433	1054
	Wheat	40	41	1241	1274
	Rice	121	111	3983	3697
	Corn	101	92	3925	3698
	Grapes	25	38	1397	2028
	Cabbage	8	12	208	336
	Tomatoes	3	4	53	70
	Watermelons	3	4	86	136
	Medlar	3	5	106	242
No crops	Plantations	67	60	2964	2273
	Grassland	6	5	7068	4700
	Forests	11	9	4054	1883
	Bare Soils	33	42	1576	5081
	Build-up	90	74	5167	6048
	Water bodies	43	35	3367	1904
	Total	**573**	**571**	**35628**	**34424**

5.1.2. Satellite imagery

Through the Sen2-Agri system, Sentinel-2 L1C (top of atmosphere) images were downloaded for the six tiles covering the study area and over the whole 2017 growing season. Using Sen2-Agri's L2A processor, a Multi-sensor Atmospheric Correction and Cloud Screening (MACCS algorithm) was performed on the top of atmosphere images, resulting in an L2A time series of 27 Sentinel-2 A (S2A) and 8 Sentinel-2 B (S2B) images between December 8, 2016 and November 1, 2017 (Figure 12). Appendix 1 shows the availability of cloud free images over the study area.

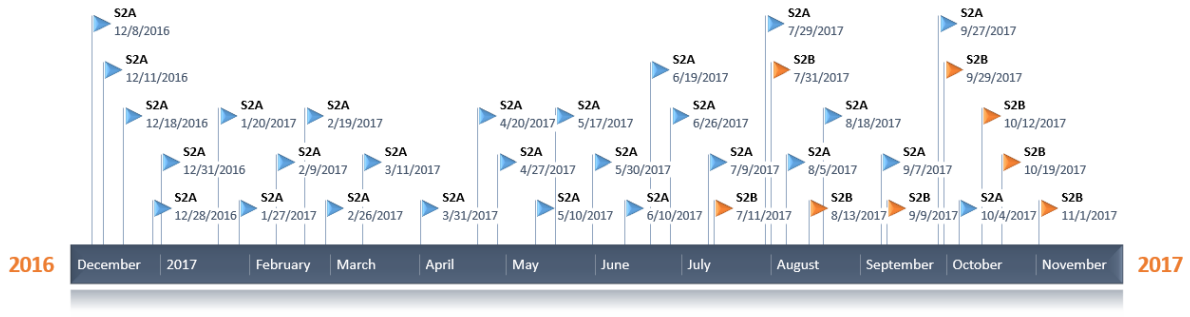

Figure 12: Sentinel-2 A and B L2A time series of successfully processed images by atmospheric correction for the 2017 growing season.

5.2. Methods

The experiments carried out on the Sen2-Agri system can be divided in two main sections. In a first series of experiments, the binary cropland mask product was evaluated and the effects of the algorithm and the training data were analysed. Secondly the accuracy of the crop type classification was assessed at the end of the season and along the season.

5.2.1. Crop mapping with Sen2-Agri

Before describing the experimental methods in more detail, it is important to outline Sen2-Agri's workflows for the cropland map and the crop type map production, as described in the Seninel-2 Agriculture Software User Manual (Udroiu *et al.*, 2017).

The cropland mask is a binary map, differentiating annual cropland areas from other areas. Annual cropland is defined as "a piece of land with a minimum area of 0,25 ha, actually sowed/planted and harvestable at least once within the year following the sowing date" (JECAM, 2016). The inputs for the cropland mask production imply Sentinel-2 L2A images (optionally complemented by atmospherically corrected Landsat-8 images) and reference data. The cropland mask production is based on a random forest classifier. The input reference data can be *in situ* data, collected and provided by the user, or a reference raster land cover map, which is the CCI Land Cover 2015 product by default (Figure 13), but can be provided by the user as well (Udroiu *et al.*, 2017). The algorithms for the cropland mask production with or without *in situ* data differ in their number of features (variables) computed to train and run the random forest classifier. The algorithm without *in situ* data is much more generic as it relies on only 5 features, while the other is trained on about 50-100 features, depending on the number of input images. As a result, the algorithm with *in situ* training data fits the reference data much more closely. The cropland mask algorithms with and without *in situ* data will further be referred to as algorithm A and B respectively.

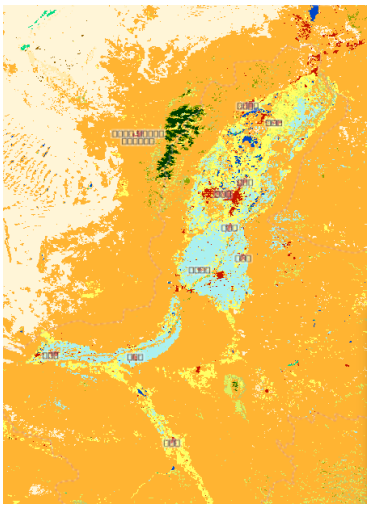

Figure 13: ESA's CCI 300m Land Cover product of 2015.

The processing chain of the crop type map is also based on a random forest classifier. The main inputs are the bottom of atmosphere Sentinel-2 images and in situ data representing all the expected crops of the study area. Moreover, the cropland mask generated for the same period is also needed as input. The L2A images are linearly interpolated to obtain smooth time series and a temporal resampling ensures a homogeneous distribution of the data over time. The features extracted for the random forest classifier are surface (TOC) reflectances, NDVI, NDWI and brightness (Euclidean norm of the surface reflectances), which were defined as most pertinent by Inglada *et al.* (2015). The classification model is then used to classify the cropland area, as defined by the cropland mask. The crop type map production workflow is schematized in Figure 14.

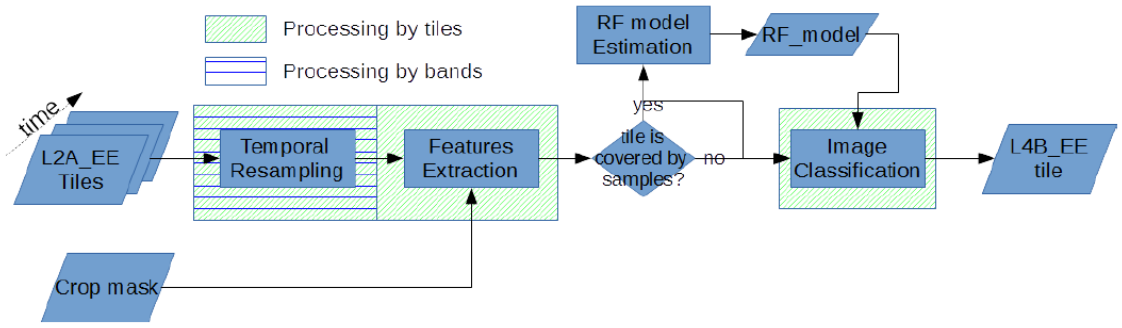

Figure 14: Sen2-Agri crop type map production (Udroiu *et al.*, 2017).

5.2.2. Crop mapping experiments

a. Binary cropland mask: quality and sensitivity analysis

The aim of the experiments on the cropland mask is to first evaluate the general quality of the binary classification with and without *in situ* data and then assess the sensitivity of the result to the algorithms and to the training data.

A first cropland mask was generated with in situ data and images of the whole growing season. The training subset of the 2017 field data (cfr. 5.1.1) was used as input, together with the 35 L2A images obtained over the season (cfr. 5.1.2), to generate a classification model through the random forest classifier (algorithm A) in Sen2-Agri. The user-defined parameters of the RF classifier were set with default values of Sen2-Agri, namely 100 trees, a maximum depth of 25 and a minimum of 25 features used at each node. This mask will be further referred to as mask A.

A second cropland mask was generated without *in situ* data, but with ESA's CCI LC 2015 product as training data source instead. The input images were the same as for mask A and the parameters of the random forest classifier were kept default as well. This second mask, generated through algorithm B, will be further referred to as mask B.

A second experiment was carried out to assess the effect of the algorithm on the resulting crop mask. Therefore, a third cropland mask was generated using the 2017 *in situ* dataset as training data in algorithm B, meant to run without *in situ* data. By comparing this mask to mask A, the effect of the different algorithms can be assessed.

In a third experiment, algorithm B was tested with different reference rasters, with varying levels of accuracy. This experiment was twofold. First mask B was used as reference raster in algorithm B. The resulting mask was then used as reference for a new mask and so on. Here the assumption was made that mask B has a higher accuracy than the CCI LC map it was based on. Considering that hypothesis, the objective was to verify if the crop mask gains in accuracy with each iteration. Secondly, the same experiment was repeated with mask A as first input raster. Practically, in order to use a cropland mask as reference raster for a new cropland mask, it was added to the CCI LC. The CCI LC was first resampled to 10m pixels. Then the pixels labelled as "Cropland, rainfed" or "Cropland, irrigated or post-flooding" were removed from the CCI

LC and the remaining pixels were used to replace the "no crop" pixels of the cropland mask. The thereby altered land cover map was used as reference raster for a new cropland mask in algorithm B. This was done iteratively, starting from mask A or mask B providing a total of 6 cropland masks, to be validated.

b. Crop type classification at the end of the growing season

A first crop type map was generated using the imagery of the whole growing season and the 2017 *in situ* training dataset, representing 9 different crops. Crop mask A was used as input for this crop type map. The values of the parameters for the random forest classifier were kept default again (100 trees, maximum depth 25 and 25 features at each node).

c. Crop mapping along the growing season.

The overall growing season in the study area runs from March to October, with certain crops such as corn being sowed slightly later than others. In order to evaluate the potential for crop mapping along the growing season with Sen2-Agri, five crop type maps were generated, using time series running from March 2017 to April, May, June, July and November of the same year. For each map, a specific cropland mask was also generated with the same dates as input. Figure 15 shows the 5 date selections for this experiment.

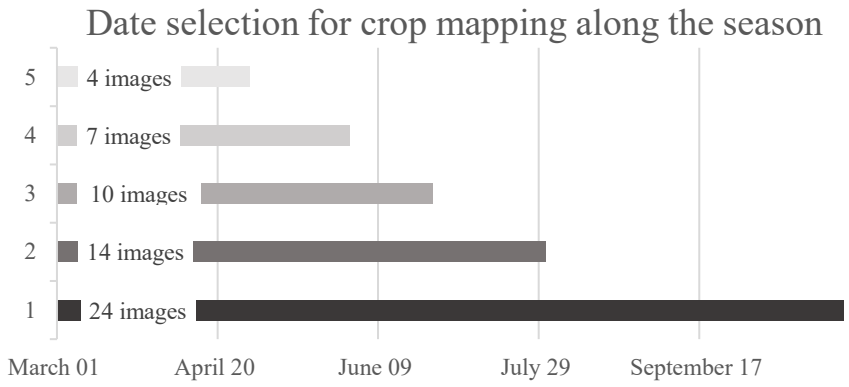

Figure 15: Date selection to evaluate the quality of crop maps along the season.

5.2.3. Validation methods

In this study the validation of the generated maps was carried out independently, outside the Sen2-Agri system, using the validation subset of the reference data (cfr. 5.1.1). To evaluate and compare the performances of the different experiences carried out with the Sen2-Agri system, a conventional validation method was used, computing performance metrics such as the F1-scores and the overall accuracy. Those metrics are based on a confusion matrix between the classified pixels and reference validation pixels. The reference pixels were obtained from the validation set of in situ polygons selected earlier. Two validation rasters were generated, one binary raster differentiating "cropland" and "no cropland" for the cropland masks validation and another featuring the different crops for the crop type map validation. The confusion matrices were computed as presented in Table 1 and the performance metrics were calculated with Eq. 1-4.

6. Crop mapping (2017): results and discussion

6.1. Cropland mask

6.1.1. Cropland mask with or without *in situ* data

Cropland mask A: algorithm A with *in situ* data

The first cropland mask was generated using Sentinel-2 imagery from the start till the end of season, implying 24 dates and with the *in situ* training dataset in algorithm A (mask A). The resulting mask is shown in Figure 16. Visually, this cropland mask seems very satisfying, as the whole irrigated agricultural area seems to be classified correctly as "crop". The main built up areas and the yellow river are clearly distinguishable as "no crop". When comparing the zoomed in section to the equivalent Sentinel-2 composite image, the smaller and more scattered built up areas, the narrow roads and even boundaries between different plots seem to be classified quite accurately as "no crop". Due to the pixel-based approach, however, a slight salt and pepper effect is almost unavoidable.

Figure 16: Mask A. Binary crop land mask generated in Sen2-Agri, with *in situ* reference data. The whole study area (left), a zoomed in section (extent in blue on the main map) as a Sentinel-2 composite of July (top right) and as the cropland mask (bottom right).

Statistically, and according to the *in situ* validation dataset, the high accuracy of this cropland mask is confirmed. As shown on Table 6, The F1-scores of "crop" and "no crop" are 92 and 94% respectively and the overall accuracy is 93%.

Table 6: Confusion matrix between cropland mask A and the *in situ* validation dataset. Precision, recall, F1-scores and overall accuracy (OA) computed based on the matrix.

Cropland Mask A		**Classified**	
		no crop	**crop**
Ref	**no crop**	13485	952
	crop	1323	18664
Precision		0,93	0,93
Recall		0,91	0,95
F1-score		**0,92**	**0,94**
OA		**0,93**	

However, in zones outside the main irrigated area in the centre, cropland seems to be misclassified as "no crop" a lot (Figure 17). This is due to the fact that the classification model tends to fit the reference data very closely, resulting in a better classification in the areas around the field campaign routes. Agricultural areas located away from the training data sampling area probably present slightly different spectro-temporal profiles, due to a climate gradient, divergent crop calendars or agricultural practices and thereby don't fit the classification model well.

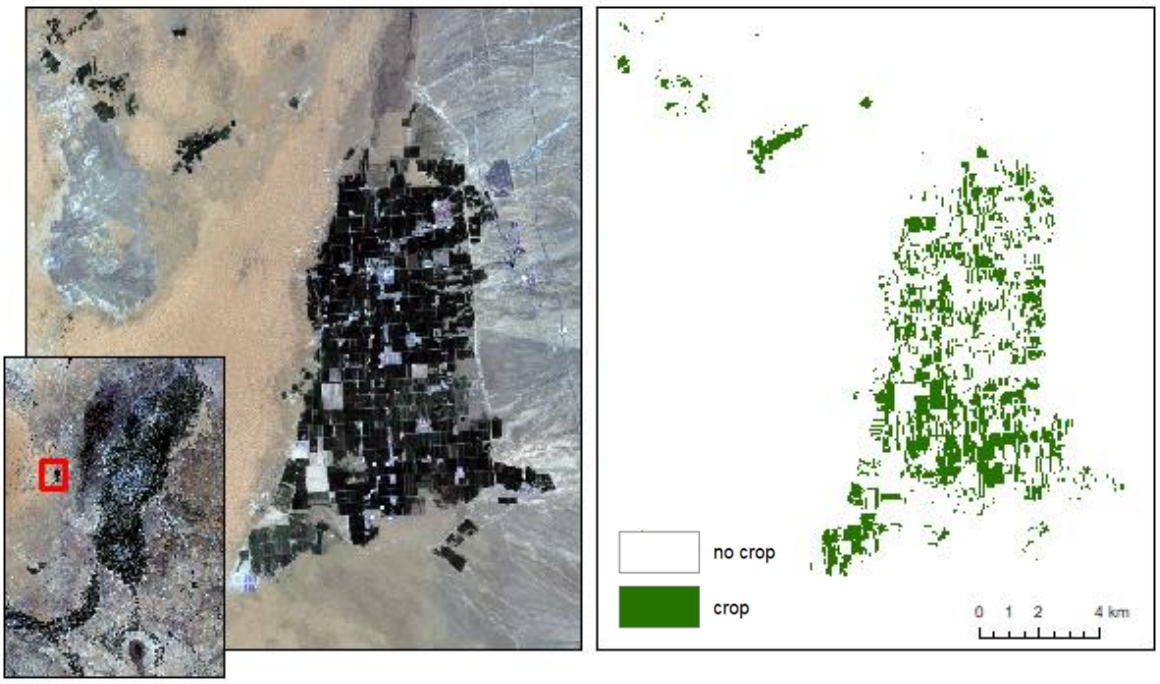

Figure 17: Close-up on a remote agricultural area, illustrating the lower accuracy of the classification in such areas.

Cropland mask B: algorithm B without *in situ* data

The aim of this experiment was to evaluate the cropland mask obtained without any *in situ* training dataset but based on ESA's CCI LC product (mask B) and to compare it to the mask obtained with *in situ* reference data (mask A).

By a first visual analysis (Figure 18), it appears mask B is less accurate in most areas. In general, mask B identifies much more cropland. Two main areas of misclassification can be identified, namely the urban areas and the mountainous area, as shown on Figure 19. Over the rest of the irrigated area, cropland seems too dense and plots are not as distinguishable as in mask A.

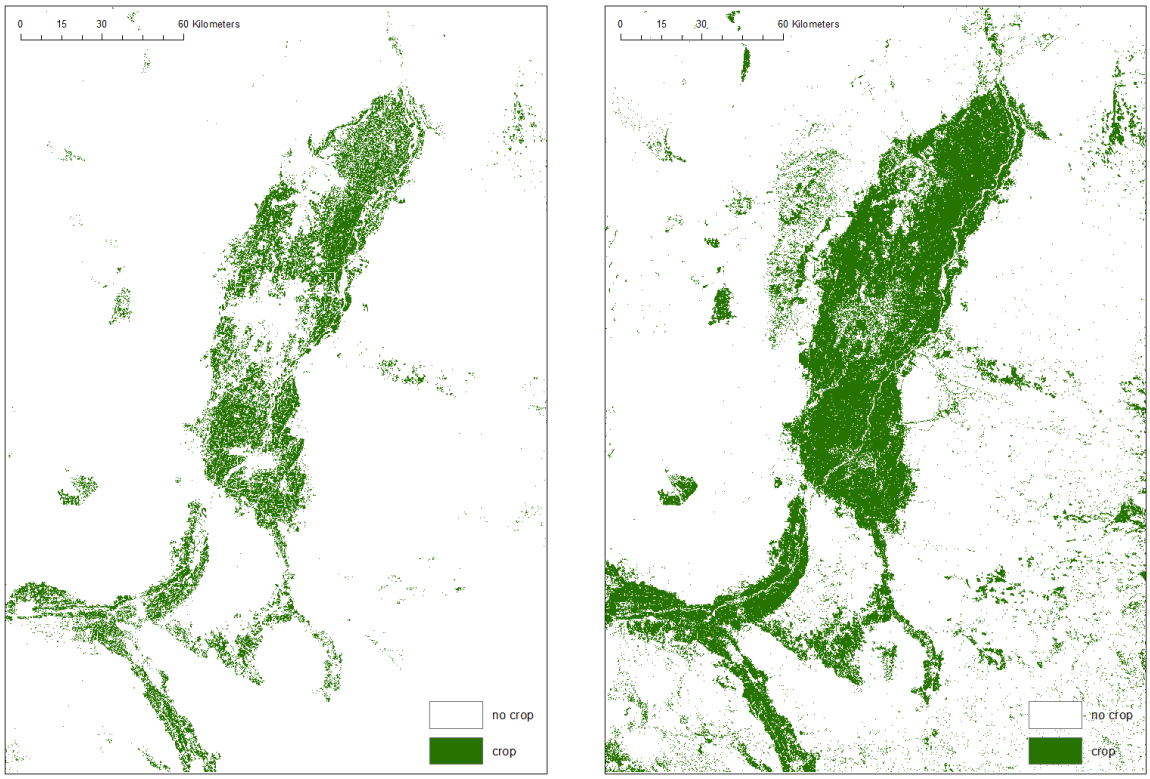

Figure 18: Cropland masks generated with (left) and without (right) in situ reference data.

The performance metrics, computed with the same *in situ* validation dataset for both masks, show mask B is indeed less accurate than mask A (Table 7). The overall accuracy calculated with the *in situ* validation dataset is 64%, which is much lower than the 93% accuracy of mask A. The low overall accuracy is mainly due to the high amount of "no crop" misclassified as "crop", as reflected by the low precision of "no crop" (0,40) and the low recall of "crop" (0,55). On the other hand, very little "crop" pixels were misclassified as "no crop" in mask B (less than in mask A) and in the remote cropland areas, mask B seems to correctly identify more cropland than mask A (Figure 20).

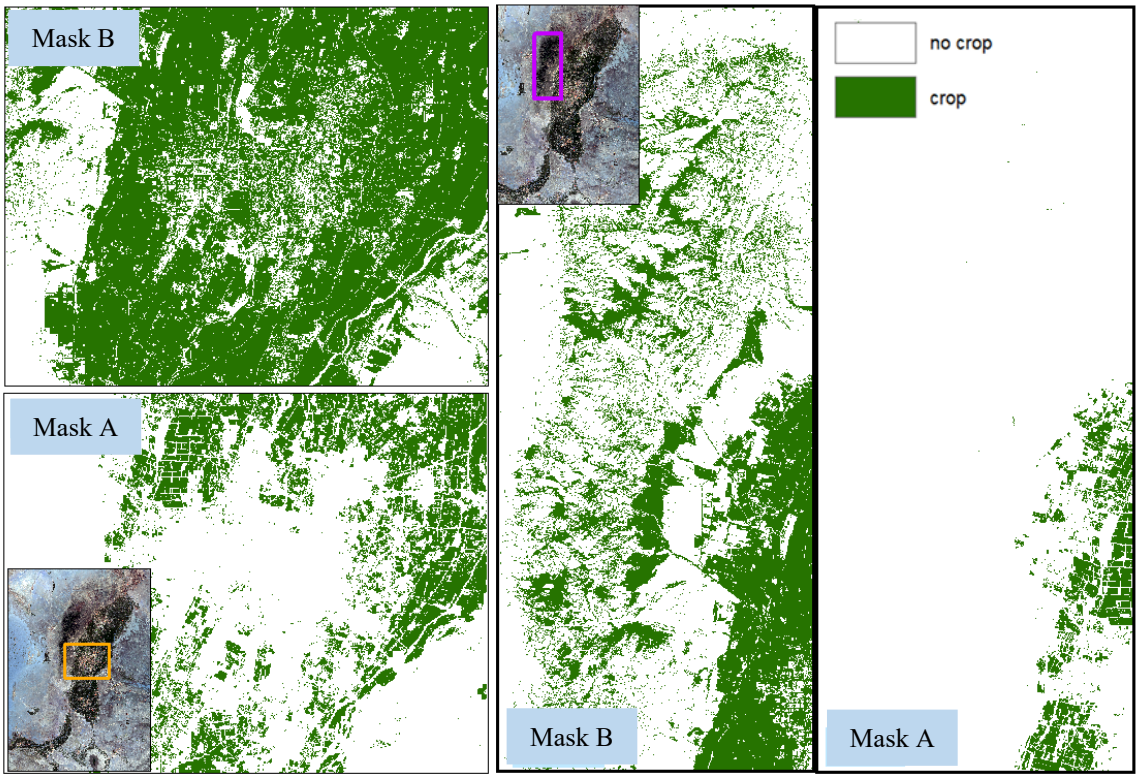

Figure 19: Close-up on the main areas of misclassification: the urban area of Yingchuan (left) and the mountainous area in the north-west of the region (right).

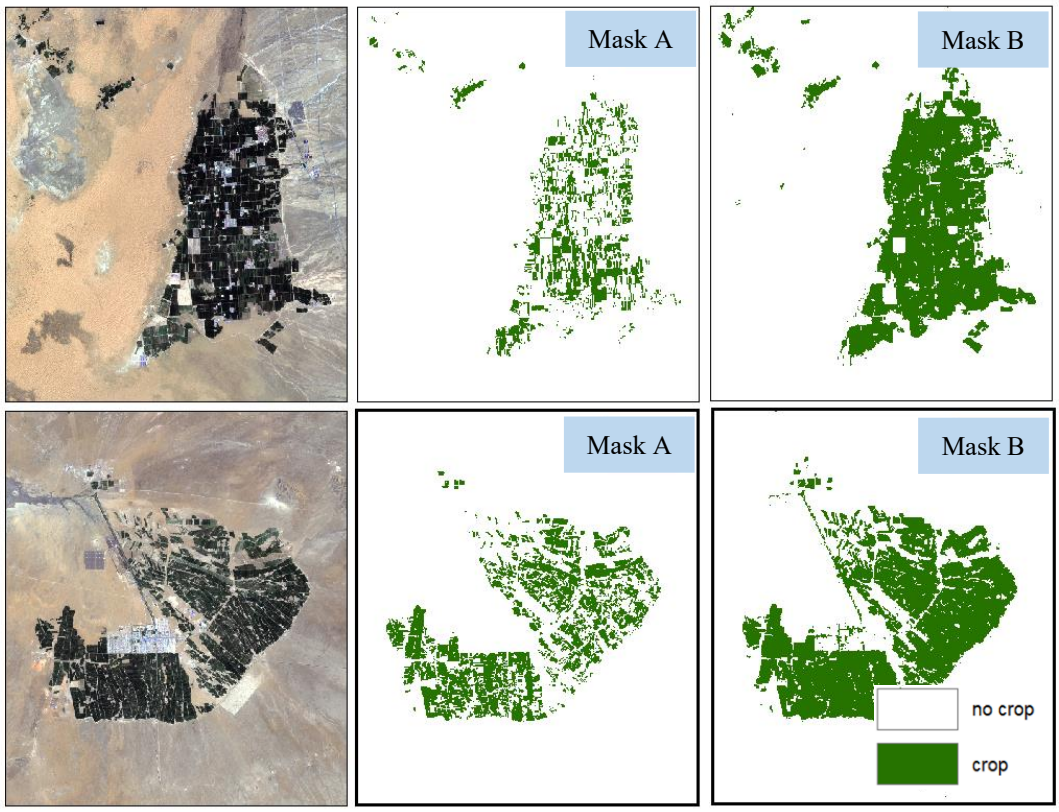

Figure 20: Close-up on remote cropland areas where mask B seems to correctly identify more cropland than mask A.

Table 7: Confusion matrix and performance metrics for cropland mask A and cropland mask B, computed with the in situ validation dataset.

Cropland mask A		Classified		Cropland mask B		Classified	
		no crop	**crop**			**no crop**	**crop**
Ref	**no crop**	13485	952	**Ref**	**no crop**	7809	11807
	crop	1323	18664		**crop**	502	14306
Precision		0,93	0,93	**Precision**		0,40	0,97
Recall		0,91	0,95	**Recall**		0,94	0,55
F1-score		**0,92**	**0,94**	**F1-score**		**0,56**	**0,70**
OA		**0,93**				**0,64**	

Statistically, the accuracy of mask B seems excessively low, relative to the visual aspect of the mask. This is due to the fact that the validation dataset is spatially correlated to the *in situ* training dataset and lacks a good spatial distribution over the study area. Therefore, the accuracy of the classification in remote areas is not represented in the performance metrics. This also puts into perspective the very high overall accuracy obtained for mask A.

In order to better evaluate mask B, a second validation dataset was generated, based on a 5 km grid of points covering most of the study area. The points were manually labelled as "crop" or "no crop", using Sentinel-2 imagery and Google Earth as visual reference. Evaluating and comparing both masks, based on this spatially better distributed dataset should be more adequate. New performance metrics were computed based on the validation grid (Table 8).

Table 8: Confusion matrix and performance metrics for cropland mask A and cropland mask B, computed with the 5 km validation grid.

Cropland mask A		Classified		Cropland mask B		Classified	
		no crop	**crop**			**no crop**	**crop**
Ref	**no crop**	1318	18	**Ref**	**no crop**	1188	148
	crop	94	140		**crop**	20	214
Precision		0,99	0,60	**Precision**		0,89	0,91
Recall		0,93	0,89	**Recall**		0,98	0,59
F1-score		**0,96**	**0,71**	**F1-score**		**0,93**	**0,72**
OA		**0,93**				**0,89**	

The metrics for mask A mostly change in terms of F1-scores, especially for "crop", which reflects the remote cropland misclassified as "no crop". The overall accuracy stays as high, because it is mostly influenced by the very high amount of "no crop" points in the validation grid. For mask B, the F1-scores and the overall accuracy increase significantly with the new validation method. The F1-score of the "crop" is even slightly higher than the one in mask A (72% and 71% respectively). The overall accuracy of mask B increased to 89%. But again the "no crop" validation points weigh much more than "crop" points in this overall accuracy. Indeed, the validation grid contains 4 times more "no crop" points than "crop" points.

To tackle this balance issue, ¾ of the "no crop" points in the validation grid were randomly selected and deleted from the dataset. This way, a balanced validation grid was obtained with

the same amount of "crop" and "no crop" reference points. The performance metrics computed with this balanced grid are shown in Table 9.

Table 9: Confusion matrix and performance metrics for cropland mask A and cropland mask B, computed with the balanced validation grid.

Cropland mask A		Classified		Cropland mask B		Classified	
		no crop	**crop**			**no crop**	**crop**
Ref	**no crop**	248	4	**Ref**	**no crop**	225	27
	crop	93	140		**crop**	19	214
Precision		0,98	0,60	**Precision**		0,89	0,92
Recall		0,73	0,97	**Recall**		0,92	0,89
F1-score		**0,84**	**0,74**	**F1-score**		**0,91**	**0,90**
OA		**0,80**				**0,91**	

The results of this third validation confirm mask A underestimates the cropland area (0,60 precision) and illustrate the main advantages of mask B, namely a spatially homogenous precision and accuracy. This balanced validation grid however overlooks the very high accuracy of mask A in the main agricultural area, as the number of reference pixels is very low.

Overall, the cropland mask generated without *in situ* reference data (mask B) remains less accurate than the one based on the *in situ* training dataset (mask A). In general, mask B tends to overestimate the cropland area. In terms of details, like plot boundaries, mask A performs much better than mask B. This is majorly due to the coarse 300m resolution of the CCI LC product mask B is based on. With the development of higher resolution global landcover products, this Sen2-Agri product could become very promising.

On the other hand, mask A underestimates the cropland area, especially in remote areas. The better performance of mask B in the remote areas could be due to the different algorithm used for the mask without *in situ* data, more precisely to the number of features used as input for the classification. As algorithm A works with a high number of features, the random forest classifier tends to overfit the reference training data. The crops in the remote areas probably have slightly different spectral signatures or temporal profiles due to divergent cultural practices or environments. Therefore, algorithm A tends to overlook those remote plots in the classification. On the other hand, algorithm B being more generic classifies them as cropland. This could also be due to the training dataset of mask B, covering those remote areas as well, while the *in situ* training data used for mask A is limited to the main agricultural area. Further experiments need to be carried out to draw conclusions on the matter.

6.1.2. Impact of the algorithm

Following the last experiment and in order to draw conclusions on the comparison of both cropland mask algorithms, a second experiment on the cropland mask production was carried out. The idea is to use the same *in situ* training dataset as input for algorithms A and B. This way, it is possible to assess the performance of algorithm B and actually compare it to the more complex one with *in situ* data (algorithm A). The new cropland mask generated with the *in situ* training dataset in algorithm B will be referred to as mask C.

From a first visual analysis (Appendix 2), it seems clear mask C is much more accurate than mask B, especially in the main central agricultural area and the mountainous area. Moreover, mask C looks very similar to mask A, suggesting it as accurate as mask A. The combined map of both mask A and mask C confirm the similarity (Appendix 3). When zooming in on different areas however, some differences appear more clearly (Figure 21). In the main central agricultural area, mask C seems to have identified less cropland and the salt and pepper effect appears to be stronger within the plots. In the more remote agricultural areas however, mask C appears to be much more accurate than mask A.

Statistically, mask C performs much more like mask A than mask B, as expected from the visual analysis (Table 10). The overall accuracy of 90% (computed with in situ data) and 78% (computed with the balanced validation grid) is indeed much closer to mask A's accuracies (93% and 80%).

Table 10: Confusion matrices and performance metrics of the cropland mask generated with in situ data and the second algorithm, computed with the *in situ* validation dataset (left) and the balanced validation grid (right).

Cropland mask C (in situ validation)		**Classified**	
		no crop	**crop**
Ref	**no crop**	21348	541
	crop	2989	9546
Precision		0,98	0,76
Recall		0,88	0,95
F1-score		**0,92**	**0,84**
OA		**0,90**	

Cropland mask C (balanced validation grid)		**Classified**	
		no crop	**crop**
Ref	**no crop**	248	4
	crop	102	131
Precision		0,98	0,56
Recall		0,71	0,97
F1-score		**0,82**	**0,71**
OA		**0,78**	

With both validation methods, better performance metrics are obtained for mask A than for mask C, meaning algorithm A delivers a more accurate result than algorithm B with the same training data. The only advantage in algorithm B is that, thanks to its more generic aspect, it does not fit the training data as close as mask A and is thereby more accurate in the remote areas, presenting slightly different cropland spectro-temporal profiles.

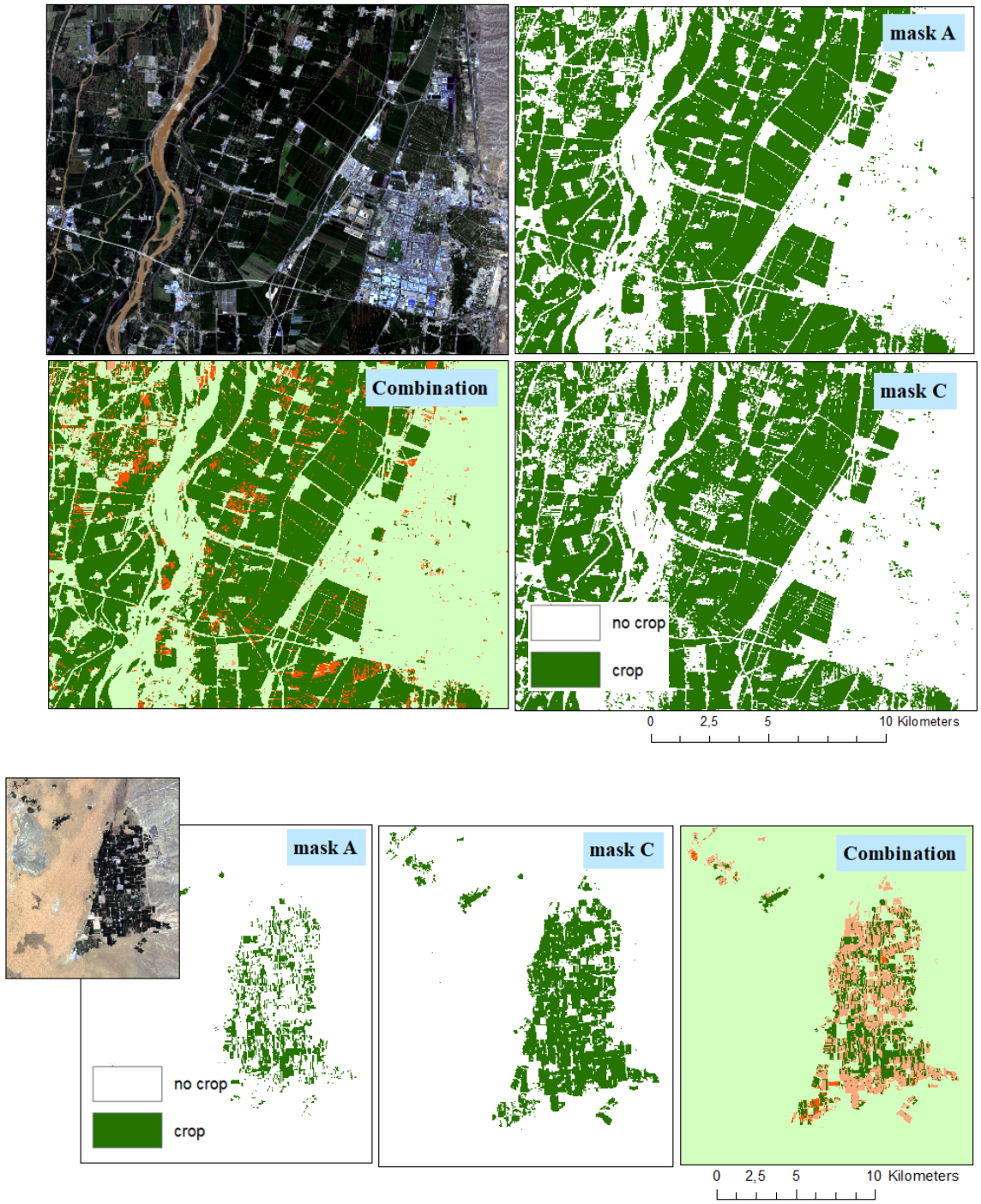

Figure 21: Close-up on the main agricultural area (above) and on the remote agricultural area (below). From left to right: the cropland mask generated with *in situ* data and algorithm A (mask A), the cropland mask generated with *in situ* data and algorithm B (mask C) and the combination of both cropland masks.

6.1.3. Impact of the training data on algorithm B

In this third experiment, the aim is to evaluate the performances of algorithm B with different reference rasters as training data source. As demonstrated in the previous experiments, the low accuracy of mask B was mainly due to the coarse spatial resolution and the low accuracy of the CCI LC map. The objective of these last experiments on the Sen2-Agri crop masks is to verify if algorithm B can deliver a cropland mask with a higher accuracy than the basemap used as training data source. In the first step of the experiment cropland mask B is used as reference data for a new cropland mask, instead of the CCI LC product. The thereby produced cropland mask is then in turn used as reference data for another one and so on. In a second iterative experiment, the reference data for the first mask is cropland mask A. The process is then iterated in the same manner as for the previous experiment.

Iterations based on cropland mask B

From a first visual analysis it seems already clear the cropland mask does not change much with each iteration (Figure 22). The urban area still contains a lot of misclassified cropland pixels, the mountainous area is still misclassified as crop as well. The statistical analysis confirms the very small change. The overall accuracy gets slightly worse with the iterations (Table 11), no matter which validation method was used. The confusion matrices and the performance metrics are shown in Appendix 4.

Table 11: Overall accuracies of the cropland masks along the iterations, computed with the in situ validation dataset (left) and the validation grid (right).

Overall Accuracy	In situ validation	Validation grid	Balanced grid
Mask B	0,64	0,89	0,91
CM B1	0,62	0,88	0,90
CM B2	0,62	0,88	0,90

The results of this experiment do not allow for any conclusion to be made on the impact of the training dataset. It is unclear whether the negative result is due to the too poor accuracy of the first cropland mask compared to the CCI LC map it is based on, or to the algorithm and the input features. In order to shed light on the issue, a second iterative experiment is carried out, using mask A as training data source.

Iterations based on cropland mask A

Using mask A as reference, a new cropland mask was generated through algorithm B. Visually, the evolution along the iterations seems more significant than in the previous experiment (Figure 23). In general, more cropland is identified with each iteration. In some cases the more generic algorithm B correctly classifies new plots as cropland in the new iterations, improving the masks quality. In remote areas, the iterations contribute greatly to the quality of the cropland mask (Figure 24). While algorithm A tends to overlook a lot of cropland in those areas, due to overfitting, the iterations with the more generic algorithm B allow to better classify the remote areas. This improvement is probably due both to the nature of the algorithm and to the presence of a few plot samples in those areas in the mask used as reference data.

On the other hand, because of the more generic aspect, other green landcovers in the urban area and the mountainous area are increasingly misclassified as cropland along the iterations (Figure 25).

Overall, the statistics confirm the visual analysis (Table 12). When evaluating the mask with the spatially correlated *in situ* validation dataset, the overall accuracy decreases strongly with the iterations. In opposition, the overall accuracy, computed using the validation grid covering the whole study area in a systematic manner, slightly increases from 93% for the reference crop mask to 94% for the first iteration. The improvement appears even stronger with the balanced validation grid, showing how the iterations correct the errors of mask A. Indeed, the number of "crop" pixels misclassified as "no crop" decreases with the iterations. The complete confusion matrices and performance metrics of this experiment are shown in Appendix 5.

Table 12: Overall accuracies of the cropland masks along the iterations, computed with the in situ validation dataset (left) and the validation grid (right).

Overall accuracy	In situ validation	Validation grid	Balanced grid
Mask A	0,93	0,93	0,80
CM A1	0,80	0,94	0,88
CM A2	0,70	0,94	0,89

Considering these results, it seems the iterations with algorithm B contribute to the crop mask quality outside of the field campaign area, but not particularly inside the field campaign area. The statistics however indicate one iteration is beneficial for the crop mask accuracy. This means algorithm B allows to generate a cropland mask with a higher accuracy than the reference basemap, as long as the basemap's accuracy is above a certain threshold to start with.

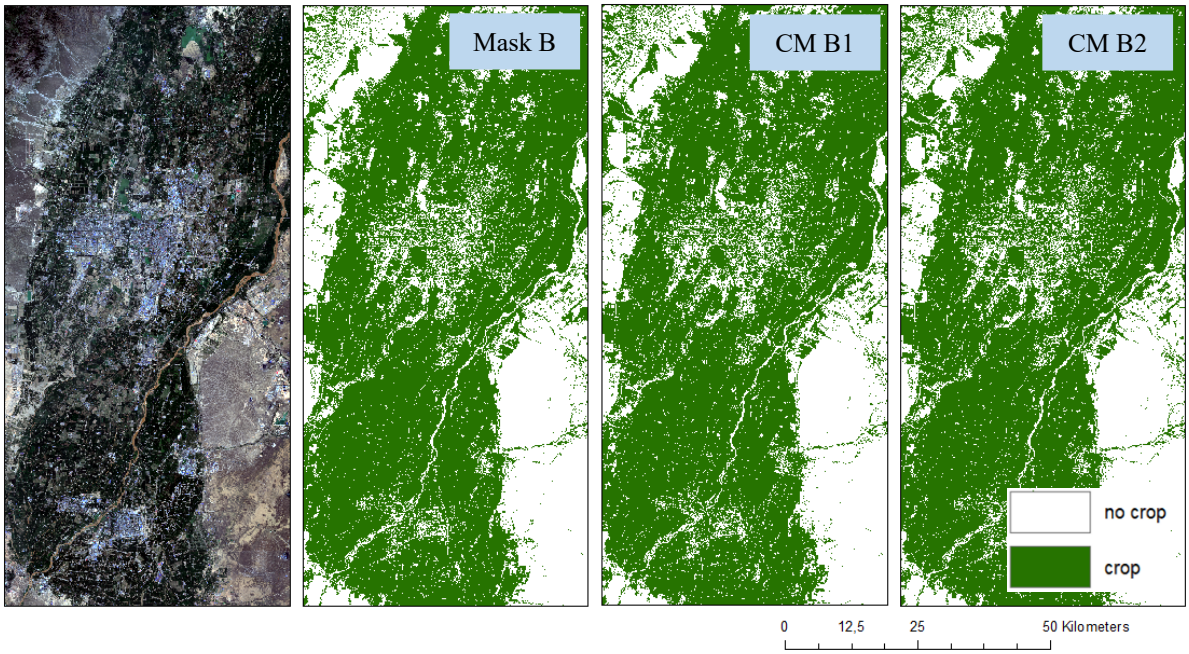

Figure 22: Evolution of cropland mask along the iterations. Cropland mask B generated with the CCI LC product as reference, cropland mask 1 (CM B1) generated with mask B as reference and cropland mask 2 (CM B2) generated with CM B1 as reference.

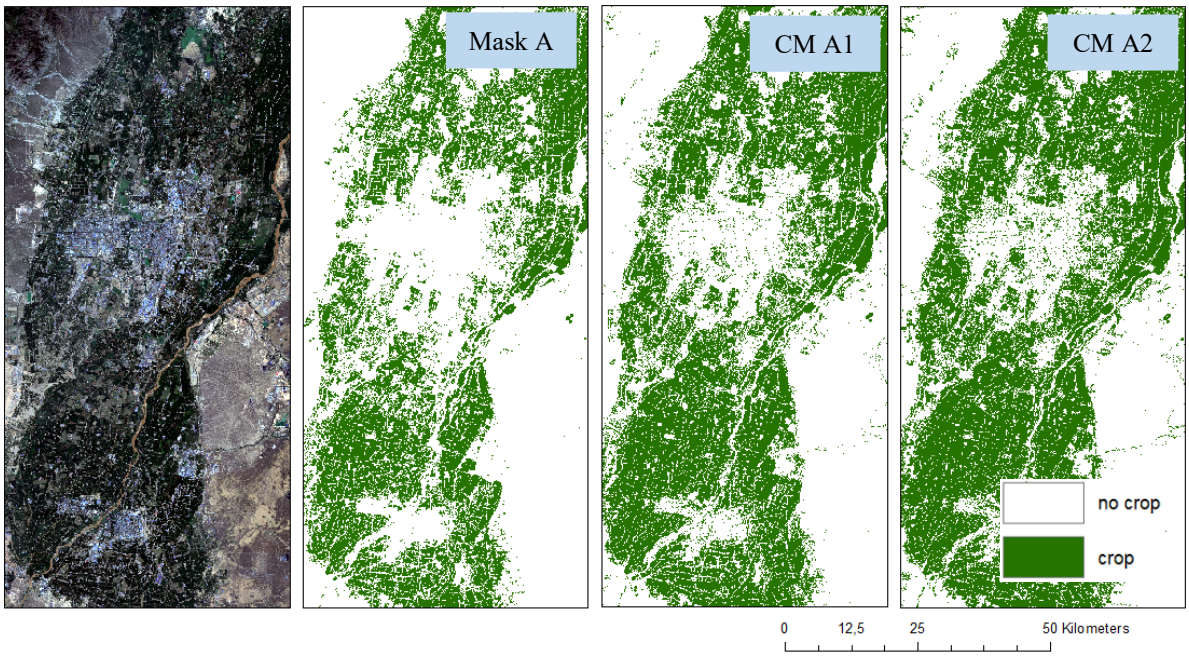

Figure 23: Evolution of cropland mask along the iterations. Cropland mask A (generated with *in situ* data as reference), cropland mask 1 (CM A1) generated with mask A as reference and cropland mask 2 (CM A2) generated with CM A1 as reference.

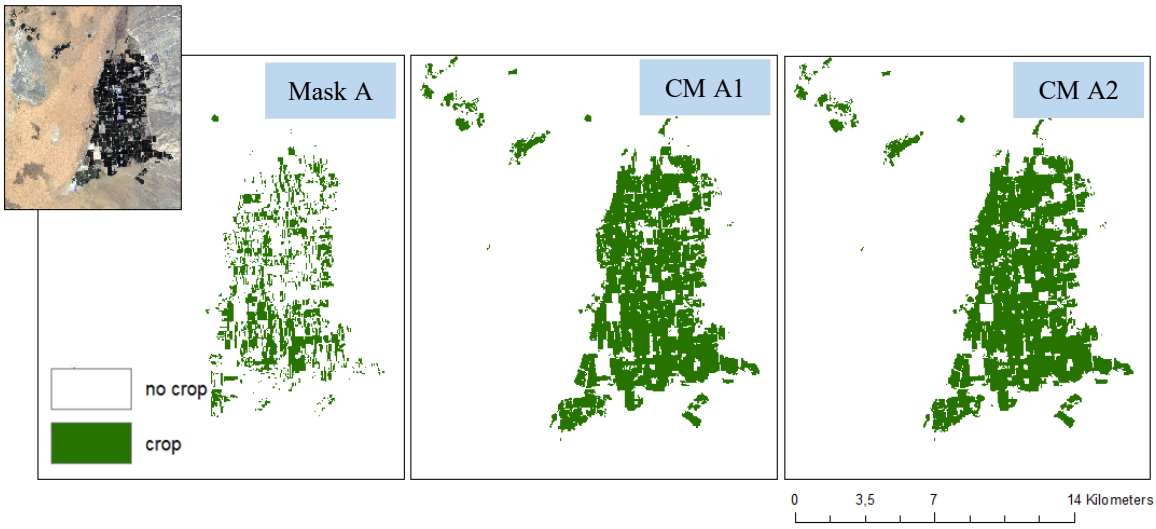

Figure 24: Close-up on a remote agricultural area and its evolution along the cropland mask iterations.

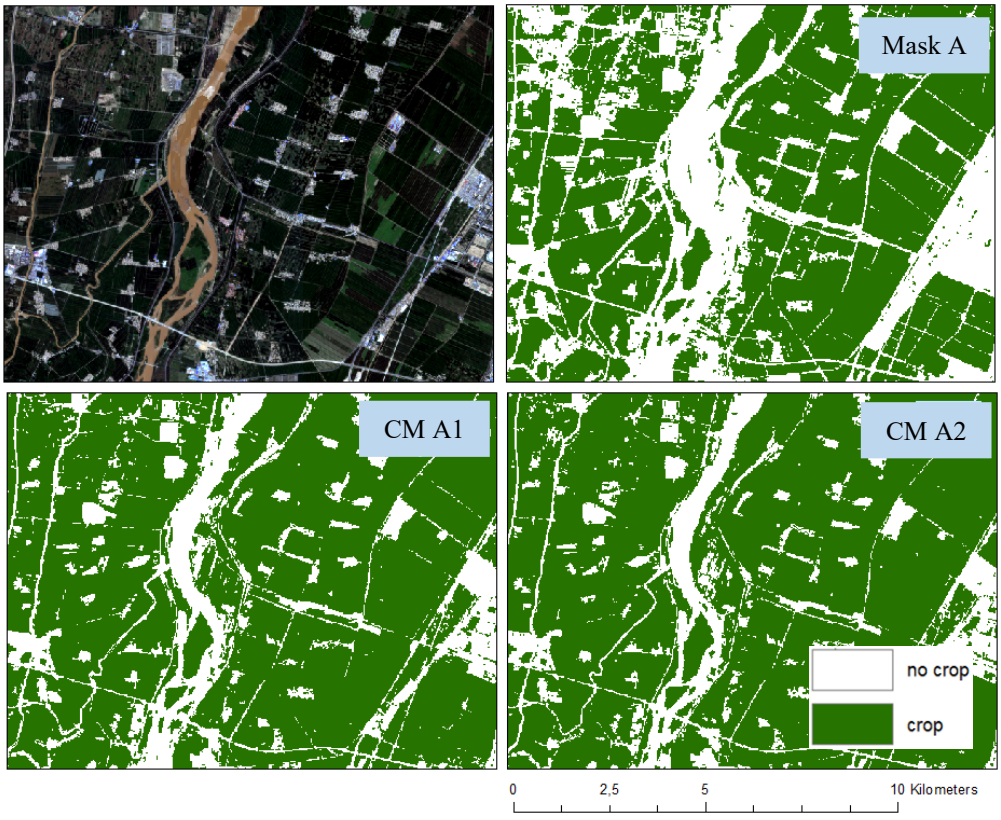

Figure 25: Close-up in the main agricultural area and its evolution along the cropland mask iterations.

6.1.4. Conclusions on Sen2-Agri's cropland mask production

Based on all three experiments, several conclusions can be drawn on the cropland mask processors of Sen2-Agri.

First of all, it is clear that the poor quality of mask B is entirely due to the coarse resolution and the low accuracy of the CCI LC map it is based on, as the same algorithm performed much better with mask A or *in situ* data as input. This is promising for the generation of cropland masks without *in situ* data, considering the increasing quality and resolution of available basemaps.

Secondly, using the same *in situ* reference data as input, algorithm B provides a cropland mask with a spatially more homogenous accuracy than the first algorithm. Nevertheless, the overall accuracy of mask A is higher than the overall accuracy of mask C. Algorithm A can thereby be considered very accurate in a homogenous agricultural landscape, and with a spatially distributed training dataset, covering the whole variety of cropland and other land covers. On the other hand, algorithm B, as it doesn't tend to overfit the training data, is better adapted to map larger and more heterogenous landscapes. The input data for algorithm B doesn't need be as diverse in terms of spatial distribution and crop types as for algorithm A, but the size of the training dataset is more important for algorithm B. The best result was obtained by using mask A as input in the algorithm B, as the reference dataset was large and the second algorithm allowed to correctly classify even the more remote areas.

Furthermore, the results of these experiments put the conventional validation method and performance metrics into perspective. F1-scores and the overall accuracy cannot be entirely relied on to evaluate a classification but need to be complemented by a visual analysis. The considerable difference between the two overall accuracies computed for mask B emphasizes the impact of spatial distribution of the validation dataset on statistical evaluation.

6.2. Crop type map

The crop type map generated with the Sen2-Agri system, based on the previously obtained cropland mask (mask A), using images of the whole 2017 growing season and the *in situ* training dataset, is shown in Figure 26 and Appendix 6. From a visual analysis of the full extent, the main crops, namely maize and rice clearly stand out followed by wheat. Those three main crops are indeed the major crop types grown in this region (cfr. Chapter 3). Based on a closer look on the plots, rice and maize seem to be well distinguished from one another, while maize and wheat appear to be mixed inside some plots. Generally, there is an undeniable salt and pepper effect, due to the pixel-based approach.

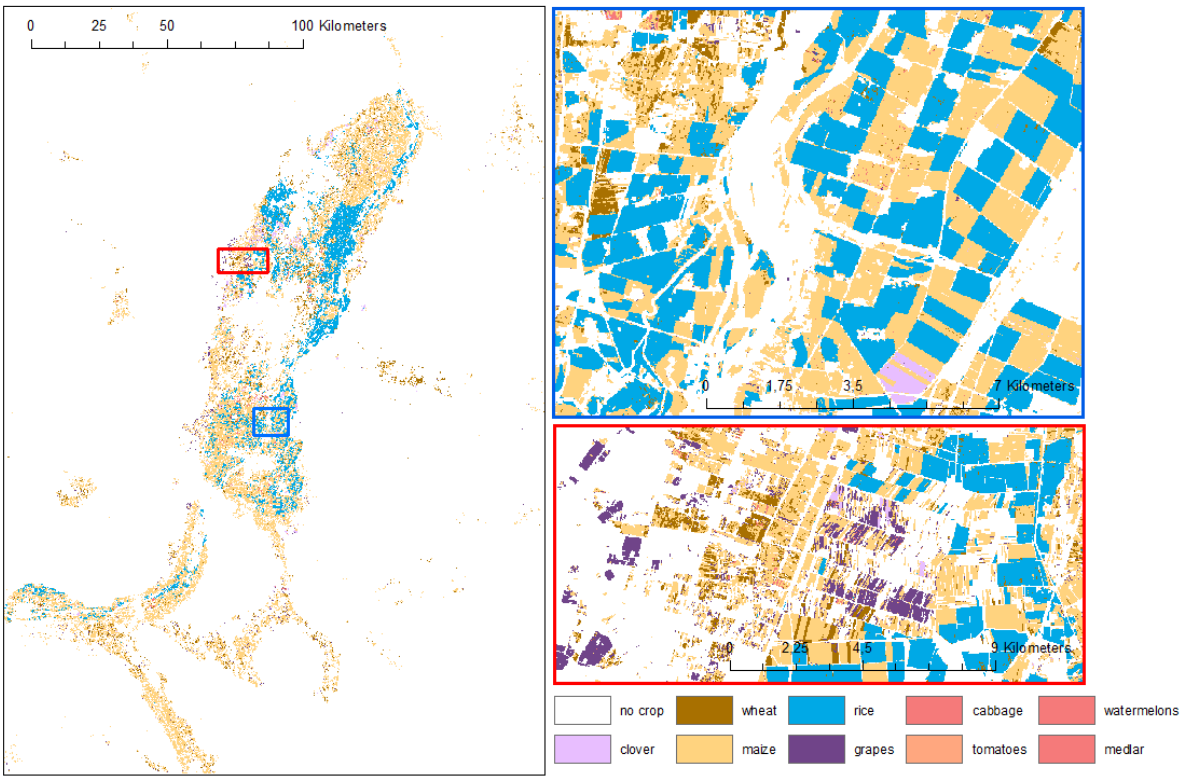

Figure 26: 2017 crop type map generated in Sen2-Agri with *in situ* reference data and the cropland mask.

6.2.1. Statistical evaluation

To assess the quality of the classification, a confusion matrix was generated between the crop type map and the *in situ* validation dataset containing each crop type, and quality metrics were computed (Table 13). A rough crop area estimation was carried out as well, in order to rank the crop types in terms of importance.

Based on a simple pixel counting, Figure 27 gives an estimation of the area covered by each crop type. According to the confusion matrix, this area estimation is biased, but it gives a general idea of the relative importance of each crop in the study area. The three main crops mentioned before are followed by three minor crops, namely grapes, cabbage and clover. The three least represented crops (tomatoes, watermelons and medlar) cover a negligible area compared to the others and will therefore be further referred to as marginal crops.

Table 13: Confusion matrix of the crop type map generated using imagery from the whole season. Precision, recall, F1-scores and overall accuracy based on the matrix. The overall accuracy was also computed for the crop classes only, without including the no crop.

Refernece \ Classified	Clover	Wheat	Maize	Rice	Grapes	Cabbage	Tomatoes	Watermelons	Medlar	No Crop	Precision
Clover	**815**	3	17	22	0	2	0	1	0	60	0,89
Wheat	16	**678**	22	114	29	6	0	0	6	298	0,58
Maize	45	264	**3170**	184	294	84	56	10	29	259	0,72
Rice	89	12	196	**3334**	0	13	4	0	0	44	0,90
Grapes	2	17	21	0	**942**	0	0	1	3	755	0,54
Cabbage	2	2	1	0	0	**214**	0	48	0	17	0,75
Tomatoes	25	0	1	0	0	0	**0**	0	0	10	0,00
Watermelons	0	0	9	0	0	0	0	**0**	0	76	0,00
Medlar	0	0	1	0	8	0	0	0	**0**	204	0,00
No Crop	6	84	387	114	58	35	0	0	4	**21201**	0,97
Recall	0,82	0,64	0,83	0,88	0,71	0,60	0,00	0,00	0,00	0,92	
F1-score	**0,85**	**0,61**	**0,77**	**0,89**	**0,61**	**0,67**	**0,00**	**0,00**	**0,00**	**0,95**	

Overall accuracy	**0,88**
Overall accuracy (crop type classification)	**0,73**

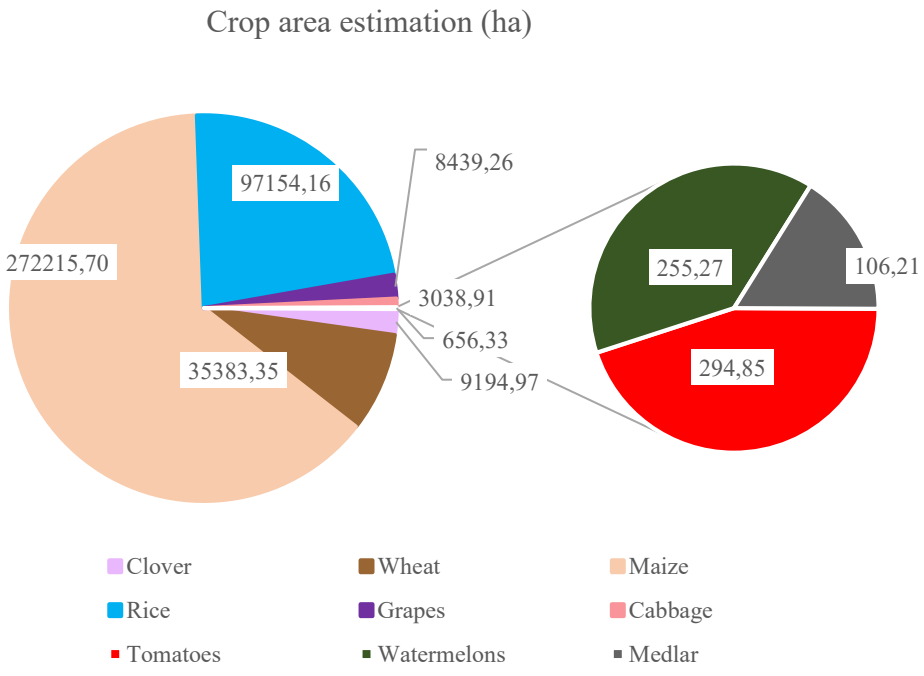

Figure 27: Rough crop area estimation, based on pixel counting.

Several observations can be made, based on the confusion matrix and the performance metrics.

First of all, the overall accuracy computed from the confusion matrix is 88%, which seems high. However, as the "no crop" pixels are numerous, they weigh a lot in the overall accuracy. It is therefore interesting to calculate the overall accuracy of the crop type classification only, excluding the "no crop" pixels. The result (73%) is significantly lower, meaning the cropland mask contributes largely to the overall accuracy of the crop type map. The real accuracy of the crop type classification is however 73%.

Secondly, when analysing the individual performances for each class, certain crops clearly stand out in terms of classification accuracy. Rice was classified with high accuracy (89% F1-score). It was barely confused with other crops, meaning its spectral signature and its temporal features are very specific and different from other crops. Clover was also classified quite accurately (85% F1-score). Other crops, like maize, wheat, grapes or cabbage have lower F1-scores, ranging from 61% (wheat and grapes) to 77% (maize). Those crops seem to be less distinguishable. Wheat and grapes were confused with maize quite often. Other landcovers ("no crop") were misclassified as maize as well. Another notable error is the confusion between grapes and other land covers, probably bare soil.

Finally, it is clear the marginal classes have been completely misclassified or overlooked by the classification model. They have not once been classified correctly. While the main and minor crops show very high F1-scores, the three marginal crop types score 0 in both precision and recall. This is majorly due to their poor representation in the training dataset, compared to other classes. In a random forest classifier, each tree uses a subset of the total set of features to define decision rules for each node (cfr. 1.1.1) As a result, classes that are poorly represented in the complete training dataset (cfr. Table 5) have a smaller chance to be represented in the feature subset for each tree. At the end, when performing the majority voting, those classes tend to be overlooked.

6.2.2. Solution for the marginal classes

A potential solution to tackle the issue of the marginal classes was to gather the three marginal classes in one "other" class for the main classification, to perform a separate classification with only those three classes and then substitute the "other" class with the separate marginal classification. The resulting crop type map's confusion matrix and performance metrics are shown in Appendix 7. Figure 28 shows the F1-scores of the initial classification and of the separate classification of the marginal classes. There is no significant improvement, except for the watermelons scoring 30%, which is still very low. The F1-scores of the other crop types remain nearly the same and the overall accuracy of the classification is the same as the initial one (80% and 73% with and without no crop pixels respectively).

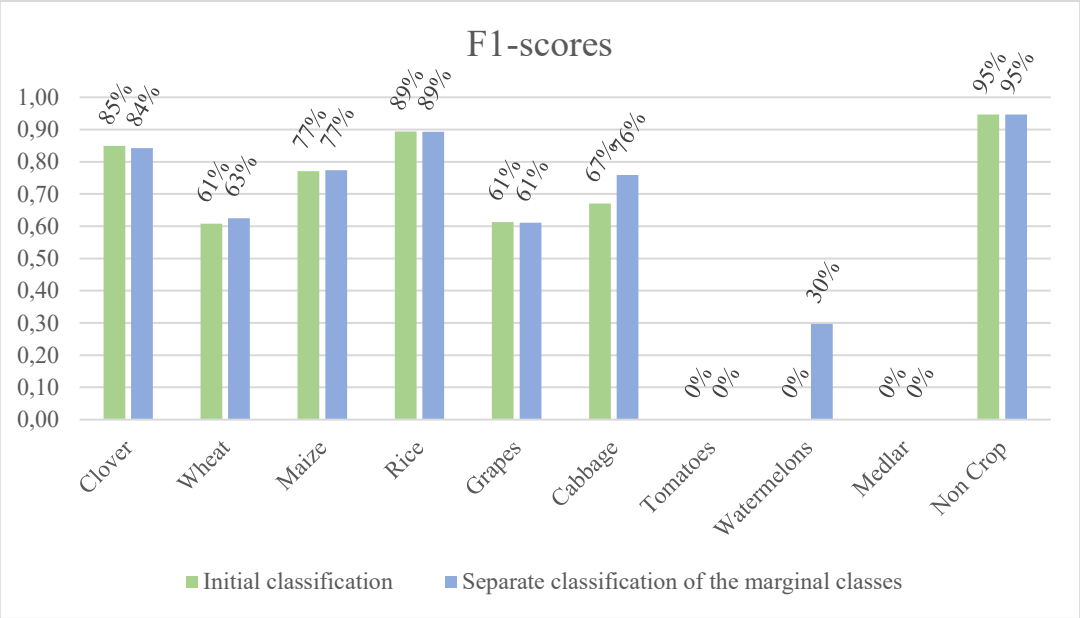

Figure 28: F1-scores of the different classes in the initial classification and in the attempted separate classification of the three marginal crops (tomatoes, watermelons and medlar).

Mapping marginal crop types over such a large area in Sen2-Agri remains an issue. But the main purpose of the system is to map the main crops in a region and in this case, main crops (rice especially) have been classified with a high accuracy. The distinction of maize and wheat could still be improved, by using different sections of the time series for example. Another way to improve the crop type classification would be to collect *in situ* training data with a better spatial distribution over the study area and a more systematic representation of the different crop types.

6.3. Crop mapping along the growing season

Four additional crop type maps were generated using images of different dates, like illustrated in Figure 15. The maps were evaluated the same way as the previous ones, delivering F1-scores and overall accuracies as shown on Figure 29. The confusion matrices and performance metrics of each crop type map are shown in Appendix 8 to 11. Based only on the overall accuracy, it seems an accurate crop type map can be generated from the end of May on (1-2 months after sowing), which is unexpectedly early. When considering the F1-scores of the different crops along the season however, it appears only rice and clover are classified accurately from the end of May on. Other classes F1-scores, like wheat or maize increase until the end of July. Rice plots can be identified early on, because they are characterised by surface water, which largely influences their spectral signature. Clover fields are sawn earlier in the season compared to wheat or maize for example and have a very characteristic spectral signature.

Considering these results, it seems possible to generate an overall accurate crop type map from the end of June on (3 months after sowing). If applied in near-real-time, this could allow to make interesting predictions and estimations on crop production, long before harvest.

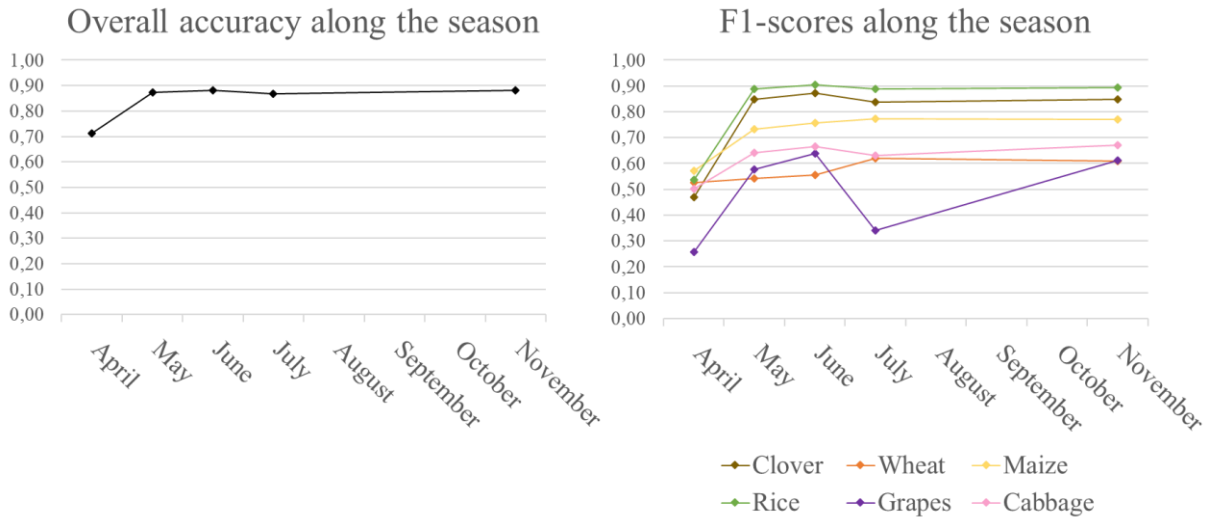

Figure 29: Evolution of the F1-scores and the overall accuracy along the growing season.

7. Near-real-time crop mapping (2018)

Following the experiments carried out on the 2017 growing season and considering their results, a cropland mask and a crop type map were generated in near-real-time for the 2018 growing season. A field campaign was carried out collecting new training and validation data for this season, giving particular attention to the sampling strategy, in order to improve the quality of the crop mapping products generated with the Sen2-Agri system.

7.1. Data sources

7.1.1. Field campaign sampling design

The new field campaign was carried out at the end of June 2018, using the same GPS picture method as in 2017 (cfr. 5.1.1). Considering the results of 2017, additional attention was given to the distribution of the sample points, aiming to provide a more accurate crop map for the 2018 growing season.

For the 2018 sampling strategy, a 10 km grid was drawn over the study area. 14 single 100 km² grids were then selected across the irrigated area, responding to two main criteria, namely spatial distribution and crop variety coverage. The sampling itinerary went from grid to grid, gathering data along all the passable roads of each grid and along the roads connecting the grids. About 8000 pictures were taken and interpreted through the Quick Photo Data Processor (Figure 30). The grids and gathered reference points are shown in Figure 31. In the same way as for previous season, polygons were created, based on the reference points. Again, complementary polygons were added *a posteriori* to cover the "no crop" land cover diversity. As opposed to the 2017 dataset, the training and validation subsets for 2018 were not defined completely randomly. All samples outside the grids and a random 25% subset of the samples inside the grids formed the validation subset. The remaining 75% of the samples inside the grids were used for training. The idea is to reduce the spatial correlation between training and validation data and thereby increase the relevance of the validation and the computed performance metrics. Table 14 shows the number of polygons and estimated number of pixels per class in the training and validation datasets.

Another issue in the 2017 crop mapping results were the marginal classes (tomatoes, watermelons and medlar), which were represented very poorly in the *in situ* dataset collected in 2017. As a result, those crops were completely overlooked by the classifier. Considering this, particular attention was given to those crops during the 2018 field campaign. The sampling itinerary crossed areas where watermelon or medlar are grown more intensively, allowing to gather more reference points for those crops. Tomatoes and cabbage were grouped, together with other vegetables, to form one single "vegetable" class. Additionally, in order to better balance the final training dataset between the different crop types, a random 25% subset of maize polygons was removed from the dataset.

Figure 30: Examples of pictures taken during the 2018 field campaign for later crop type retrieval in the Quick Photo Data Processor. From left to right: rice, medlar, maize, watermelon, wheat and grapes.

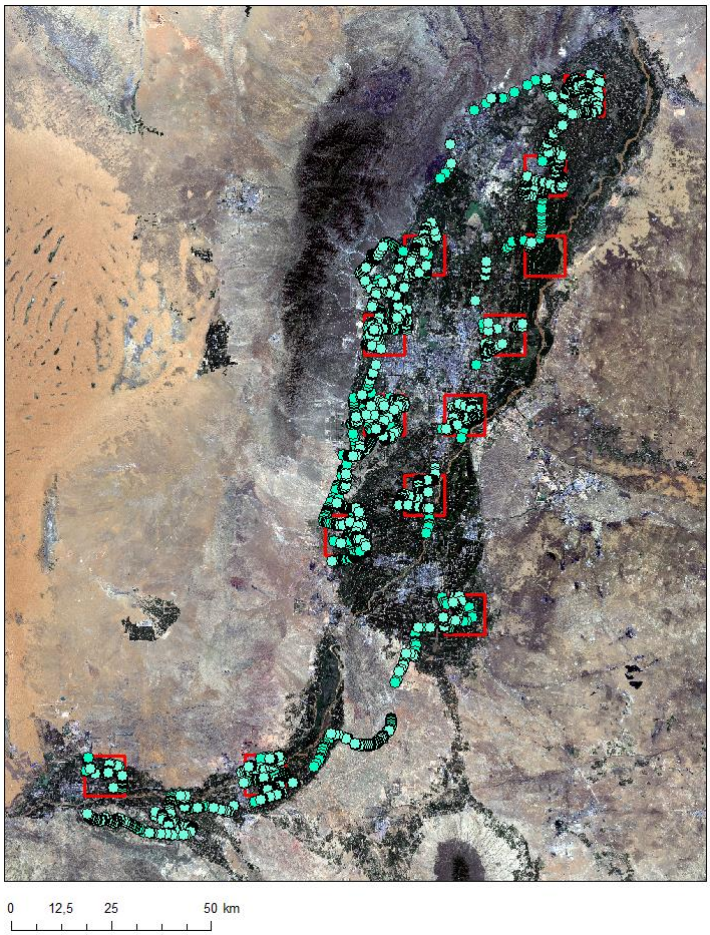

Figure 31: June 2018 field campaign grid strategy and sample point distribution.

Table 14: Training and validation datasets for the 2018 growing season.

		Samples		Pixels (10m)	
	Land Cover	**Training**	**Validation**	**Training**	**Validation**
Crops	Clover	44	41	12703	19585
	Wheat	187	158	19238	14411
	Rice	111	60	70641	63648
	Corn	299	232	51599	51967
	Grapes	52	51	17366	12616
	Vegetables	131	126	31436	29580
	Watermelons	22	8	6976	1789
	Medlar	39	55	7745	7328
No crops	Plantations	136	152	12172	19506
	Grassland	6	5	7068	4700
	Reed	13	7	848	334
	Forests	11	9	4054	1883
	Bare Soils	49	81	11732	20184
	Build-up	90	74	5205	6050
	Greenhouse	15	7	1018	1420
	Water bodies	43	35	3367	1904
	Total	**1248**	**1101**	**263168**	**256905**

7.1.2. Satellite imagery

For the 2018 growing season 34 Sentinel-2 A and 36 Sentinel-2 B bottom of atmosphere images were successfully processed by Sen2-Agri. The resulting time series is shown on Figure 32. Those images were used, together with the *in situ* training dataset to generate a cropland mask and a crop type map with Sen2-Agri.

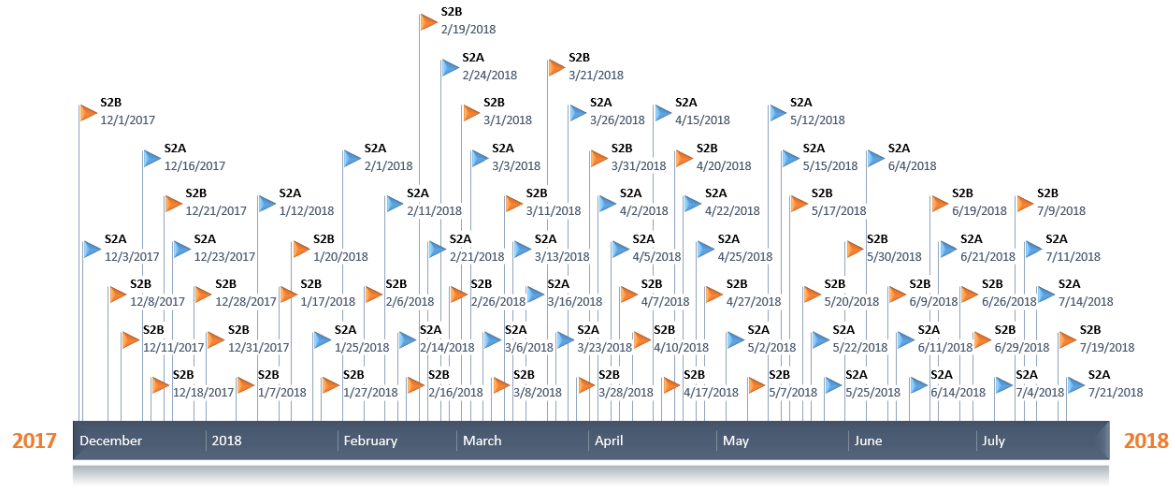

Figure 32: Sentinel-2 A and B time series of successfully processed images by atmospheric correction for the 2018 growing season.

7.2. Near-real-time crop mapping results

Using the sampled training data and the available image time series, near-real-time crop mapping was performed through the Sen2-Agris system, with the same configurations as for last season. The only difference lies in the distribution and size of the in situ dataset and in the density of the Sentinel-2 time series. A cropland mask was generated and then used to produce a crop type map. Both products were evaluated visually and statistically, with the predefined validation dataset. The aim is to assess if the better distributed *in situ* dataset improved the crop mapping performances.

7.2.1. Cropland mask

The cropland mask generated in near-real-time with Sen2-Agri is shown Figure 33. Visually, it seems to have a similar quality to mask A of the 2017 growing season. In the main central agricultural area, the urban areas are correctly identified as "no crop" and are not infested with too many misclassified "crop" pixels. The mountainous area in the north is not confused with crop. The field boundaries are more or less clear.

In order to assess the contributions of the better distributed training data, it is interesting to have a closer look at some remote areas. First of all, Figure 34 shows an area in the south of the region, that was not crossed by the field campaign itinerary in 2017 but was in 2018. The 2018 cropland mask identifies much more cropland than the 2017 cropland mask. The additional cropland identified in 2018 is mostly actual cropland, meaning the 2018 mask performed better in this area, thanks to the training data that was sampled there. Secondly, other remote agricultural areas, that were not sampled in 2017 nor in 2018, were better classified in 2018 as well. Figure 35 shows such an area in the west of the region. The 2018 mask identifies much more actual cropland than the 2017 mask. The better distribution of the *in situ* training data enhanced the accuracy of the crop mask, even in remote areas that did not contain any training samples.

The 2018 cropland mask was evaluated statistically by computing a confusion matrix between the mask and the *in situ* validation dataset. The resulting matrix and computed performance metrics are shown in Table 15. As the validation dataset was less spatially correlated to the training dataset than in 2017, the validation can be considered a bit more relevant. The high F1-scores (90% for "no crop" and "97%" for "crop") and the overall accuracy of 95% confirm the very high quality of this near-real-time 2018 cropland mask.

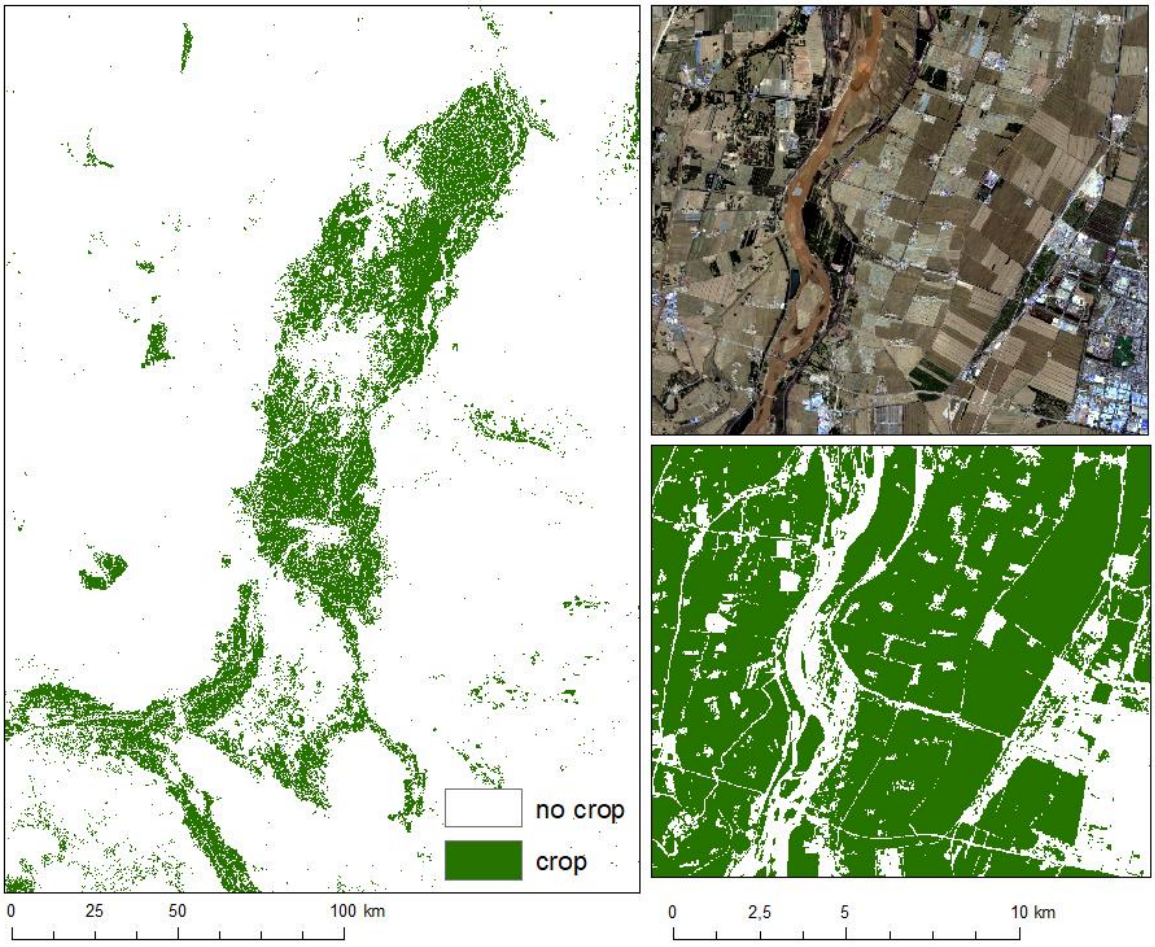

Figure 33: Near-real-time cropland mask for 2018.

Table 15: Confusion matrix and performance metrics of the 2018 cropland mask.

2018 Cropland Mask		**Classified**	
		no crop	**crop**
Ref	**no crop**	51446	5486
	crop	6404	192644
Precision		0,90	0,97
Recall		0,89	0,97
F1-score		**0,90**	**0,97**
OA		**0,95**	

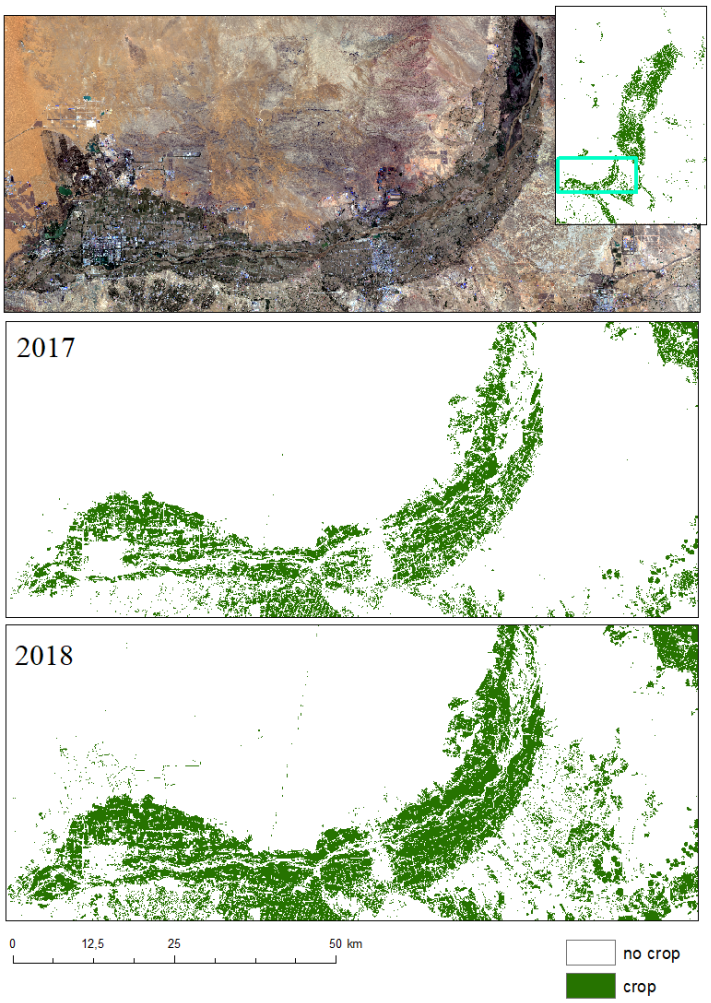

Figure 34: Close-up on the southern agricultural area; cropland mask of 2017 (above) and 2018 (below).

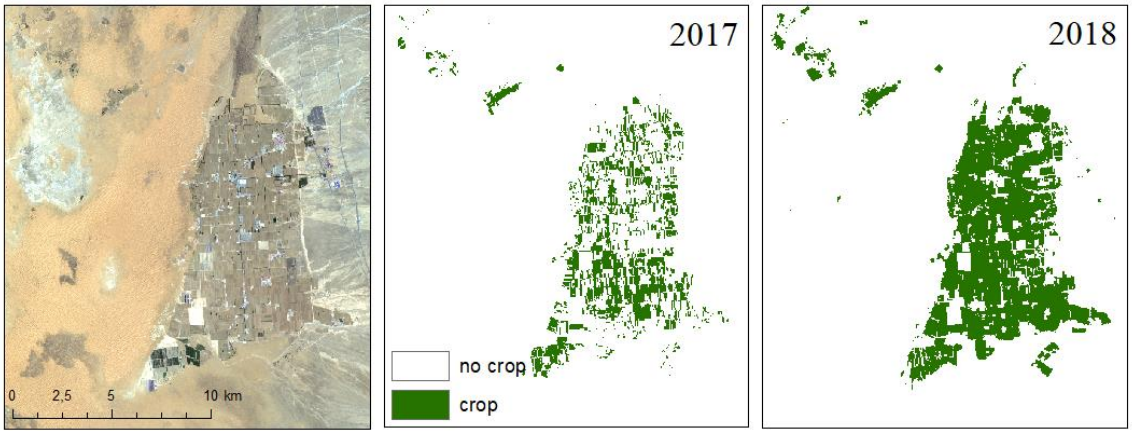

Figure 35: Close-up on a remote agricultural area; cropland mask of 2017 (left) and 2018 (right).

7.2.2. Crop type map

Using the cropland mask, the available Sentinel-2 time series and the *in situ* training dataset representing the 8 crop types, a near-real-time crop type map was generated. The result is shown in Figure 36 and in Appendix 12. The crop type map was validated with the *in situ* validation dataset. The confusion matrix and the computed performance metrics are shown in Table 16.

Visually, the complete crop type map seems satisfying. Apart from an unavoidable salt and pepper effect due to the pixel-based approach, the plots are clearly identified with different crop types. The overall accuracy of the crop type map is 86%, which is lower than the 88% accuracy computed for the 2017 crop type map. This comparison needs to be put into perspective. First of all, the validation dataset of 2017 was strongly correlated with the training dataset of the same year, while the spatial correlation between training and validation subsets was reduced in the 2018 *in situ* dataset (7.1.1). Secondly, the overall accuracy of the crop type classification only (excluding the "no crop" pixels) is 85%, which is much higher than the one computed in 2017 (73%), meaning the actual crop type classification, without considering the binary "crop"/"no crop" mask, was more accurate in 2018.

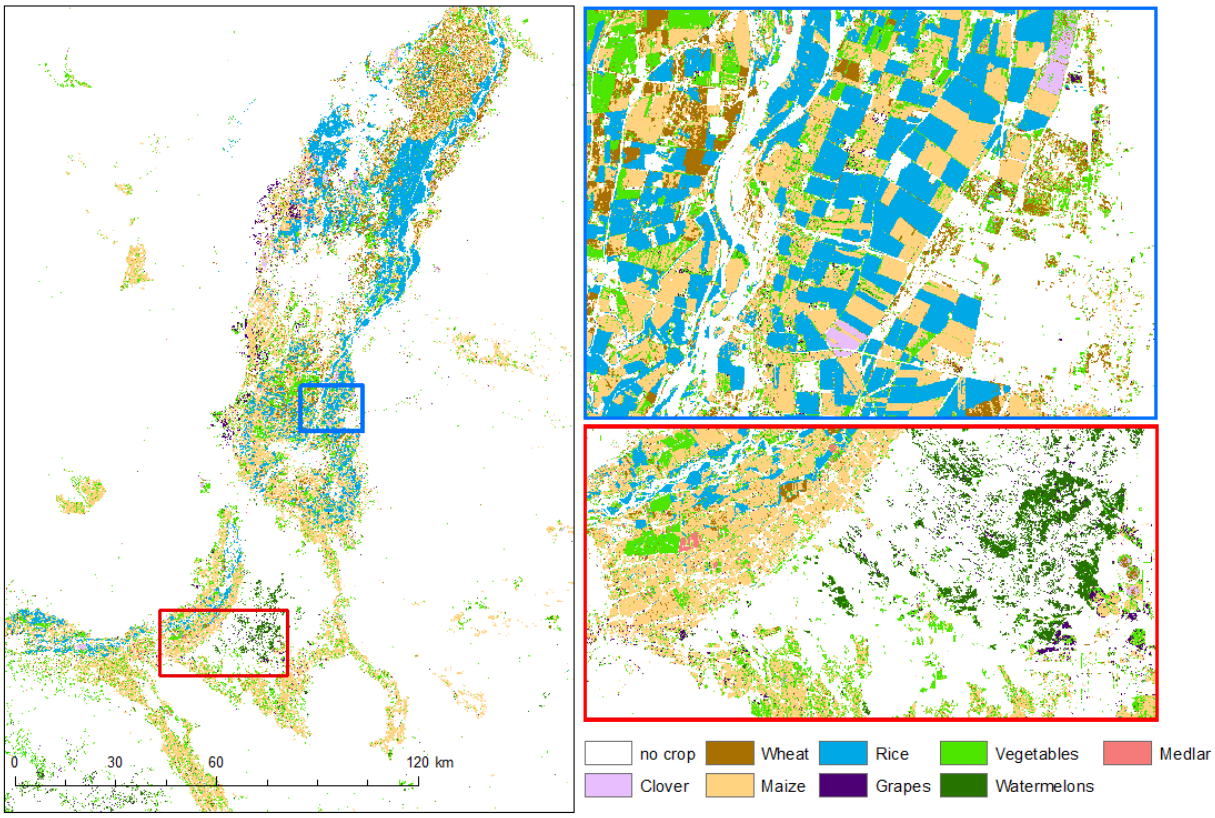

Figure 36: Near-real-time crop type map obtained in July 2018.

The main crops (rice, maize and wheat) are well distinguished, even in the north of the agricultural area, where many small wheat and maize fields form a very scattered landscape. The F1-score computed for rice is 97%. The F1-scores of maize and wheat are lower (86% and 70%), reflecting the frequent confusion between both crops and with other classes as well.

Grapes have an even lower 66% F1-score. They were mainly confused with the "vegetable" class and with "no crop", probably bare soil.

The new "vegetable" class covers a significant area compared to the separate vegetable classes in the 2017 classification and most vegetable plots were classified correctly (83% precision). Many other classes' plots are however contaminated with "vegetable" pixels. This is probably due to the great variety of crops and thus of spectral signatures inside the "vegetable" class, causing the classification model to confuse other crop types with this more generic class. Despite this, the mixed "vegetable" class scored a 79% F1-score, which is very high, considering the tomatoes class was not classified correctly once in 2017.

The two remaining marginal classes, medlar and watermelon were classified much better than in 2017. Medlar still has a low 33% F1-score. It was confused a lot with "vegetables" and with "no crop" (probably bare soil, for the same reasons as grapes), causing a low recall of 53%, but on the other hand, most medlar crops were correctly identified, which is shown by the high 90% precision. Watermelons were identified with a very high accuracy this season (93% F1-score). The field campaign itinerary crossed an area in the south, growing the majority of the watermelons of the whole region (Figure 36, bottom right). This area had completely been overlooked by the 2017 classification.

Overall, this crop type map is more accurate than the one generated last season. As the training dataset was better distributed between the different crop types, none of them was overlooked by the classification model. Another reason for the better quality of the cropland mask and the crop type map of 2018 is probably the higher density of the Sentinel-2 time series. Sentinel-2 B images were available from the start of season on, while the first Sentinel-2 B images of 2017 were from July.

Table 16: Confusion matrix of the near-real-time crop type map of 2018. Precision, recall, F1-scores and overall accuracy based on the matrix. The overall accuracy was also computed for the crop classes only, without including the no crop.

		Classified									
		Clover	**Wheat**	**Maize**	**Rice**	**Grapes**	**Vege-tables**	**Water-melons**	**Medlar**	**No Crop**	**Precision**
Refernece	**Clover**	**13479**	4510	692	45	0	416	0	4	53	0,70
	Wheat	21	**11109**	1356	204	255	411	0	25	277	0,81
	Maize	661	847	**43404**	2198	361	966	31	183	836	0,88
	Rice	38	41	353	**63550**	5	403	0	21	128	0,98
	Grapes	0	0	669	0	**7138**	2372	41	2	2353	0,57
	Vegetables	200	327	1979	154	427	**23318**	60	105	1417	0,83
	Watermelons	0	0	64	0	0	37	**1615**	0	8	0,94
	Medlar	5	150	585	25	338	1159	0	**3946**	1230	0,53
	No Crop	37	1010	1857	222	478	1744	18	112	**51420**	0,90
	Recall	0,93	0,62	0,85	0,96	0,79	0,76	0,92	0,90	0,89	
	F1-score	**0,80**	**0,70**	**0,86**	**0,97**	**0,66**	**0,79**	**0,93**	**0,33**	**0,90**	

Overall accuracy	**0,86**
Overall accuracy (crop only)	**0,85**

8. Conclusions and prospects

In the framework of the Dragon 4 cooperation and the recent development of the Sentinel-2 for Agriculture system, this master thesis aimed at evaluating the current capacities for near-real-time crop mapping with Sentinel-2 and Gaofen-1 satellite data. The first objective was therefore to analyse and compare the data quality of both satellites and assess their compatibility for crop mapping. Then the focus was set on the evaluation of the Sentinel-2 for Agriculture system, by using it to map an agricultural area in the Ningxia Hui Autonomous region in China.

The first part of this work was dedicated to evaluating Gaofen-1 imagery and comparing it to Sentinel-2. Theoretically, both satellites have very similar features and in practice, according to their analysis, their spatial and spectral resolutions are compatible. There is however an issue in terms of coregistration and geometrical compatibility. The inter-band coregistration of Gaofen-1 is right, but there is a shift between images of different dates. Moreover, there is an even more significant geometrical shift between Gaofen-1 and Sentinel-2 images. In order to consider the joint use of both satellites time series for crop mapping, it is crucial to first tackle this geometrical issue. As the shifts are not linear, further investigation is needed on the complex geometrical correction of Gaofen-1 data. In this work, the focus was further set on crop mapping with Sentinel-2 data alone.

In the second part of this thesis, the performances of the Sentinel-2 for Agriculture system were evaluated. Two specific products were focussed on, namely the cropland mask and the crop type map. Results showed a generally satisfying accuracy for both products (maximum overall accuracy of 95% and 85% respectively), confirming the high performance of the Sen2-Agri system for crop mapping.

Through several experiments on the cropland mask, a main observation was made. The most impacting factor is the training data. Both cropland mask algorithms are strongly influenced by the quality of the training data. The first algorithm tends to fit the training data very closely and therefore the *in situ* training dataset needs to be well distributed - spatially and in terms of land cover variety. The second algorithm, on the other hand, is more generic, so the distribution of the training data is less important, but the size and resolution of the dataset matter. Results showed that the second algorithm can deliver a very accurate cropland mask if the reference basemap it is trained on exceeds a certain accuracy threshold. These results are promising for the generation of a cropland mask without *in situ* data. Furthermore, through these experiments, the conventional validation methods were put into perspective, showing the sensitivity of performance metrics to the quality and the size of the validation dataset.

The crop type map generated through Sen2-Agri very accurately identified the main crops of the region, as they were dominant in the training dataset. The classification of the marginal crops was however problematic. The results showed the importance of the balance between different classes in the training dataset. The crop mapping experiment along the season showed it is possible to generate an accurate crop type map from the end of June on, three months after

the first crops are sowed and long before the harvest. This allows to use near-real-time crop mapping as a tool to make estimations and predictions on the coming crop productions of a region.

Finally, the near-real-time cropland mask and crop type map generated in 2018 confirmed the importance of the training and validation data. With a specific sampling strategy, a better distributed *in situ* training and validation dataset was obtained., Thanks to the better spatial distribution, the accuracy of the cropland mask was more homogeneous over the whole study area. Thanks to the better balance between the different crops, the crop type map was much more accurate for the marginal classes of the region.

To sum up, in the future Gaofen-1 could contribute to crop mapping in combination with Sentinel-2 if a solution is found for the geometrical correction. Sentinel-2 for Agriculture is a very efficient system for crop mapping and the most crucial factor for obtaining accurate products, if enough imagery is available, is the quality of the training and validation data.

Over the past decades, agricultural monitoring and crop mapping have developed tremendously to reach the current level of precision and accuracy, but further research is still needed to enhance performances over the globe, contribute to a better food production and management system and face the challenges ahead.

References

Bagan, H., Wang, Q., Watanabe, M., Yang, Y., Ma, J., 2005, Land cover classification from MODIS EVI times-series data using SOM neural network. *Int. J. Remote Sens. 26 (22), 4999–5012.*

Bargiel D., 2017, A new method for crop classification combining time series of radar images and crop phenology information, *Remote Sensing of Environment 198, 369-383.*

Becker-Reshef I., Justice C.O., Sullivan M., Vermote E.F., Tucker C., Anyamba A. Small J., Pak E., Masuoka E., Schmaltz J., *et al.* 2010, Monitoring global croplands with coarse resolution Earth observation: The Global Agriculture Monitoring (GLAM) project. *Remote Sens., 2, 1589–1609.*

Belgiu M., Drăguţ L., 2016, Random forest in remote sensing: A review of applications and future directions, *ISPRS Journal of Photogrammetry and Remote Sensing, 114, 24–31.*

Belgiu M., Csillik O., 2018, Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis, *Remote Sensing of Environment, 204, 509-523.*

Breiman L., 1999, Random forests—random features. *Technical Report 567, Statistics Department, University of California, Berkeley.*

Carrão, H., Gonçalves, P., Caetano, M., 2008. Contribution of multispectral and multitemporal information from MODIS images to land cover classification. *Remote Sens. Environ. 112, 986–997.*

Castillejo-González, I.L., López-Granados, F., García-Ferrer, A., Pena-Barragán, J.M., Jurado-Expósito, M., Sánchez de la Orden, M., González-Audicana, M., 2009. Object- and pixel-based analysis for mapping crops and their agroenvironmental associated measures using QuickBird imagery. *Comput. Electron. Agric. 68, 207–215.*

Cohen J., 1960, A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20, pp. 37–46.*

Conradsen K., Nielsen A.A., Skriver H., 2016. Determining the points of change in time series of polarimetric SAR data. *IEEE Trans. Geosci. Remote Sens. 54, 3007–3024.*

Dash, J., Mathur, A., Foody, G.M., Curran, P.J., Chipman, J.W., Lillesand, T.M., 2007, Land cover classification using multi-temporal MERIS vegetation indices. *Int. J.Remote Sens. 28 (6), 1137–1159.*

Davidson A.M., Fisette T., Mcnairn H. & Daneshfar B. 2017. Detailed crop mapping using remote sensing data (Crop Data Layers). In: J. Delincé (ed.), *Handbook on Remote Sensing for Agricultural Statistics* (Chapter 4). Handbook of the Global Strategy to improve Agricultural and Rural Statistics (GSARS): Rome.

Foerster S., Kaden K., Foerster M., Itzerott S., 2012, Crop type mapping using spectral–temporal profiles and phenological information, *Computers and Electronics in Agriculture 89, 30-40.*

Furby, S., Caccetta, P., Wu, X., Chia, J., 2008. Continental scale land cover change monitoring in Australia using Landsat imagery. In: *International Earth Conference: Studying, Modeling and Sense Making of Planet Earth, Mytilene, Lesvos, Greece, June 2008, pp. 1–8.*

Galton F., 1892. Finger Prints (London: Macmillan).

Gómez C., White J.C., Wulder M. A., 2016, Optical remotely sensed time series data for land cover classification: A review, *ISPRS Journal of Photogrammetry and Remote Sensing, 116, 55-72.*

Homer, C., Huang, C., Yang, L., Wylie, B., Coan, M., 2004, Development of a 2001 national land cover database for the United States. *Photogramm. Eng. Remote Sens. 70 (7), 829–840.*

Inglada J., Arias M., Tardy B., Valero S., Morin D., Sepulcre G., Dedieu G., Bontemps S., Defourny P., Koetz B., 2015, Assessment of an Operational System for Crop Type Map Production Using High Spatial and Temporal Resolution Satellite Optical Imagery, *Remote Sens., 7, 13208-13232.*

Jakubauskas M. E., Legates D. R., Kastens J. H., 2002, Crop identification using harmonic analysis of time-series AVHRR NDVI data, *Computers and Electronics in Agriculture 37, 127-139.*

Jia, K., Liang, S., Zhang, N., Wei, X., Gu, X., Zhao, X., Yao, Y., Xie, X., 2014, Land cover classification of finer remote sensing data integrating temporal features from time series coarser resolution data. ISPRS J. Photogramm. *Remote Sens. 93, 49– 55*.

Julea A., Méger N., Rigotti C., Trouvé E., Bolon P., Lăzărescu V., 2011. Mining pixel evolutions in satellite image time series for agricultural monitoring. In: *Perner, P. (Ed.), Advances in Data Mining. Applications and Theoretical Aspects chapter Mining Pixel Evolutions in Satellite Image Time Series for Agricultural Monitoring. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 189–203.*

Li J., Zheng G., Liu H., Wang L., Tang Z., Shi H., Guo W., Wang H. Situation Analysis of Ningxia Province. *China Climate Change Partnership Framework - Enhanced strategies for climate-proofed and environmentally sound agricultural production in the Yellow River Basin (C-PESAP)*

Löw F. & Duveiller G. 2014, Defining the Spatial Resolution Requirements for Crop Identification Using Optical Remote Sensing, *Remote Sens., 6, 9034-9063.*

Matton N., Sepulcre Canto G., Waldner F., Valero S., Morin D., Inglada J., Arias M., Bontemps S., Koetz B., Defourny P., 2015, An Automated Method for Annual Cropland Mapping along the Season for Various Globally-Distributed Agrosystems Using High Spatial and Temporal Resolution Time Series, *Remote Sens., 7, 13208-13232.*

Mascolo L., Lopez-Sanchez J.M., Vicente-Guijalba F., Nunziata F., Migliaccio M., Mazzarella G., 2016. A complete procedure for crop phenology estimation with PolSAR data based on the complex Wishart classifier. IEEE Trans. Geosci. Remote Sens. 54, 6505–6515.

Maus V., Camara G., Cartaxo R., Sanchez A., Ramos F.M., Queiroz G.R.D., 2016, A time-weighted dynamic time warping method for land-use and land-cover mapping, *IEEE J. Select. Top. Appl. Earth Observ. Remote Sens. 1–11.*

Mayaux, P., Bartholomé, E., Fritz, S., Belward, A., 2004. A new land cover map of Africa for the year 2000. J. *Biogeogr. 31, 861–877.*

Meier U., Bleiholder H., 2006. Growth stages of mono-and dicotyledonous plants. *BBCH Monograph, Federal Biological Research Centre for Agriculture and Forestry.*

Odenweller, J.B., Johnson, K., 1984. Crop identification using Landsat temporalspectral profiles. *Rem. Sens. Environ. 14, 39–54.*

Ok A. O., Akar O., Gungor O., 2012, Evaluation of random forest method for agricultural crop classification, *European Journal of Remote Sensing, 45:1, 421-432.*

Osman J., Inglada J., Dejoux J.-F., 2015, Assessment of a Markov logic model of crop rotations for early crop mapping, *Computers and Electronics in Agriculture 113, 234-243.*

Pal M., 2007, Random forest classifier for remote sensing classification, *Int. Jour. Remote Sens. 26 (1), 217-222.*

Pontius Jr. R. G., Mollines M., 2011. Death to Kappa : birth of quantity disagreement and allocation disagreement for accuracy assessment. *Int. Journal of Remote Sensing, Vol. 32, No. 15, 4407-4429*

Rabiner L.R., Juang B.H., 1993, Fundamentals of Speech Recognition*, PTR Prentice Hall, Englewood Cliffs, N.J.*

Radoux, J., Lamarche, C., Van Bogaert, E., Bontemps, S., Brockmann, C., Defourny, P., 2014, Automated training sample extraction for global land cover mapping. *Remote Sens. 6, 3965–3987.*

Udroiu, C., Grosu, A., Popescu, R., Nicola, L., Manda, R., Savinaud, M., Bellemans, N., Bontemps, S., 2017, Seninel-2 Agriculture Software User Manual, *UCL-Geomatics 2016*

Sakoe H., Chiba S., 1978, Dynamic programming algorithm optimization for spoken word recognition, *IEEE Trans. Acoust. Speech Signal Process. 26, 43–49.*

Valero S., Morin D., Inglada J., Sepulcre G., Arias M., Hagolle O., Dedieu G., Bontemps S., Defourny P. & Koetz B. 2015, Production of a Dynamic Cropland Mask by Processing Remote Sensing Image Series at High Temporal and Spatial Resolutions, *Remote Sens., 8, 55.*

Vicente-Guijalba F., Martinez-Marin T., Lopez-Sanchez J.M., 2015. Dynamical approach for real-time monitoring of agricultural crops. *IEEE Trans. Geosci. Remote Sens. 53, 3278–3293.*

Waldner F., Lambert M.-J., Li W., Weiss M., Demarez V., Morin M., Marais-Sicre C., Hagolle O., Baret F. and Defourny P., 2015, Land Cover and Crop Type Classification along the Season Based on Biophysical Variables Retrieved from Multi-Sensor High-Resolution Time Series, *Remote Sensing 2015, 7, 10400-10424.*

Xiong J., Thenkabail P.S., Gumma M.K., Teluguntla P., Poehnelt J., Congalton R.G., Yadav K., Thau D., 2017, Automated cropland mapping of continental Africa using Google Earth Engine cloud computing, *ISPRS Journal of Photogrammetry and Remote Sensing, 126, 225–244.*

Yang, C., Everitt, J.H., Murden, D., 2011. Evaluating high resolution SPOT 5 satellite imagery for crop identification. *Comput. Electron. Agric. 75, 347–354.*

Near-real-time crop mapping in Ningxia (China) using Sen2-Agri system: Contributions of Sentinel-2 and Gaofen-1 satellite data.

Présenté par Mathilde De Vroey

Abstract

In the framework of the Dragon 4 cooperation (between ESA and the Chinese MOST) and the recent development of the Sentinel-2 for Agriculture system, the aim of this master thesis is to evaluate the current capacities for near-real-time crop mapping with Sentinel-2 and Gaofen-1 satellite data.

The first objective was to analyse and compare the data quality of both satellites and assess their compatibility for crop mapping. The results of this joint imagery analysis showed a potential compatibility of both satellites but revealed an issue in the coregistration of Gaofen-1 and a significant geometrical shift between both satellites' images. Concluding further research is needed before considering a combination of Sentinel-2 and Gaofen-1 for crop mapping, this study further focussed on crop mapping with Sentinel-2 only.

The focus was then set on the evaluation of the Sentinel-2 for Agriculture system performances, by using it to map an agricultural area in the Ningxia Hui Autonomous region in China.

As a first step, several experiments were carried out on the cropland mask and the crop type map products generated for the 2017 growing season. The general accuracy of the cropland mask was assessed as well as its sensitivity to the training data and to the classification algorithm. The experiments led to the conclusion that the training dataset has a great impact on the resulting cropland mask and that its spatial distribution and land cover diversity coverage are crucial. The accuracy of the crop type classification was assessed for the different crops of the region. The system delivered a very accurate classification for the main crops but showed difficulties for identifying the more marginal crop types. The capacity of crop mapping along the season was also evaluated. Results revealed an accurate crop map could be generated from the end of June on.

The second stage was to perform near-real-time crop mapping for the 2018 growing season, considering the previous observations. Thanks to a better distributed training dataset, better crop mapping performances were achieved, confirming the importance of training data in the Sen2-Agri system.

To sum up, this study showed Gaofen-1 could contribute to crop mapping in combination with Sentinel-2 in the future, if a solution is found for the geometrical correction. Secondly, it confirmed Sentinel-2 for Agriculture is a very efficient system for crop mapping. Finally, experiments showed the most crucial factor for obtaining accurate products, if enough imagery is available, is the quality of the training and validation data.

Année académique 2017-2018