Lund University GEM thesis series nr 22

Comparison of multi-temporal and multispectral Sentinel-2 and Unmanned Aerial Vehicle imagery for crop type mapping

Stephania Zabala

2017

Department of Physical Geography and Ecosystem Science
Lund University
Sölvegatan 12
S-223 62 Lund
Sweden

LUND UNIVERSITY

UNIVERSITY OF TWENTE.

FACULTY OF GEO-INFORMATION SCIENCE AND EARTH OBSERVATION

Comparison of multi-temporal and multispectral Sentinel-2 and Unmanned Aerial Vehicle imagery for crop type mapping

by

Stephania Zabala

Thesis submitted to the department of Physical Geography and Ecosystem Science, Lund University, in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation for Environmental Modelling and Management

Thesis assessment Board

First Supervisor: *David E. Tenenbaum* (Lund University)

Exam committee:
Jonas Ardö
Hongxiao Jin

Disclaimer

This document describes work undertaken as part of a program of study at the University of Lund. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the institute.

Course title: Geo-information Science and Earth Observation for Environmental Modelling and Management (GEM)

Level: Master of Science (MSc)

Course duration: September 2016 until June 2017

Consortium partners:

The GEM master program is a cooperation of departments at 5 different universities:
University of Twente, ITC (The Netherlands)
University of Lund (Sweden)
University of Southampton (UK)
University of Warsaw (Poland)
University of Iceland (Iceland)

ABSTRACT

Precision Agriculture aims to maximize crop production and the efficiency of land use to meet the increased demand for food while minimizing environmental impact and economic cost of food production. Crop type maps are needed for Precision Agriculture applications and remote sensing techniques are an efficient way to produce this information. However, at present, there is no single sensor that can provide sufficient data to assess the various growth stages and temporal changes of crops over the growing cycle. Therefore, this study investigated the possibility of combining Sentinel-2A and Unmanned Aerial Vehicle (UAV) data for crop monitoring. We evaluated the potential of the spectral, spatial and temporal information of Sentinel-2A imagery for crop type mapping at the plot level in southern Sweden. We explored the compatibility of spectral bands between MicaSense RedEdge and Sentinel-2A (S2A) to assess the utility of UAV observations in complementing and replacing satellite imagery with cloud-cover and noise. Moreover, we examined the seasonal variation of crops based on an annual time-series of S2A NDVI together with available UAV NDVI data.

A supervised Random Forest classifier (RF) was calibrated and validated with ground truth data. We tested the performance of the Variable Selection using Random Forest (VSURF) algorithm to reduce the number of covariates in the classification and to eliminate redundancy in the dataset. We used S2A imagery from 12 dates (from April through July). The number of variables used in the classification was reduced from 145 to 8, with an accuracy of 93% and Kappa of 0.92. Regarding key spectral information, we found that red-edge and shortwave infrared (SWIR) were of high value for crop mapping. Also, the blue band appeared to be important for differentiating crops, together with the maximum NDVI for the growing season. Conversely, bands in the near infrared were amongst the least important for the classification of crops in the study area.

Three atmospherically corrected S2A images were compared to three orthomosaics consisting of raw image values and reflectances. The comparison was based on averaged UAV pixels falling into S2A pixel sized cells. The band-by-band analysis evaluated the correlation and mean differences of reflectances and three vegetation indices (NDVI, EVI, and GCC). The results showed that the correlation of reflectances between sensors improved after the radiometric calibration performed by ATLAS. The most correlated bands were red and NIR, closely followed by the green band. However, statistically, significant differences were found in the actual physical units. Similarly, vegetation indices (VI) reduced the variability in the data and showed stronger correlations, although significant differences were also found mostly with EVI. VIs values from S2A imagery were higher in bare soil and lower in green areas compared to those from UAV orthomosaics.

The S2A NDVI time-series for crop pixels showed potential to provide seasonality information that can be of high value for various agriculture applications, including crop monitoring. NDVI derived from UAV orthomosaics were used to complement the time-series and to evaluate how well they represent the temporal variation. TIMESAT was used to improve data quality and produce smooth seasonal curves. Results showed that despite absolute differences between the indices obtained from both sensors, UAV observations could provide continuity to the S2A time-series and improve up-scaling of vegetation phenology.

Keywords: UAV, remote sensing, precision agriculture, Random Forest, classification, Micasense

Acknowledgments

I would like to thank my supervisor at Lund University, Prof. David E. Tenenbaum, and Prof. Lars Eklundh for suggesting the topic of this research and providing me invaluable support throughout the investigation.

Special thanks to Ryan Davidson and Erik Rasmusson, who provided me with field data from SITES. Anders Christensson and Bertil Hansson who kindly provided me with data from their agriculture fields.

Thank you to Prof. Anton Vrieling, Prof. Gabriel Parodi, and Caroline Gevaert from the University of Twente for their guidance in the first stage of this research.

To Ximena Tagle who helped me with the UAV image processing and Zhanzhang Cai for his help running TIMESAT with Sentinel-2 images.

Many thanks to Per-Ola Olsson who collected the images during all the UAV missions, without his help this research would not have been possible.

This work is dedicated to my parents for their constant support. To the amazing humans who developed computers, the internet, satellites and all the current technology that, among others, enable us to monitor the Earth to try to come up with solutions to live together in harmony. To the online community who share their knowledge and help others advance in their careers. Last but not least, to the European Commission and Erasmus Mundus + for the opportunity of my lifetime.

Abbreviations

AC	Atmospheric Correction
AGL	Above Ground Level
AOT	Aerosol Optical Thickness
APS	Advanced Photo System
BOA	Bottom of the atmosphere
BRDF	Bidirectional Reflectance Distribution Function
CMOS	Complementary Metal-Oxide-Semiconductor
COFS	Commercial off-the-shelf
DEM	Digital Elevation Model
DN	Digital number
DN	Digital number
DNG	Digital Negative
DSLR	Digital Single-Lens Reflex
ESA	European Spatial Agency
EVI	Enhanced Vegetation Index
ExG	Excess Green
GCC	Green Chromatic Coordinate
GCP	Ground Control Points
GLCM	Gray Level Co-occurrence Matrix
GPS	Global Positioning System
GRC	Ground Resolution Cell
GSD	Ground Sampling Distance
IMU	Inertial Measurement Unit
LAI	Leaf Area Index
MDA	Mean Decrease in Accuracy
MDG	Mean Decrease in Gini
MLC	Maximum Likelihood Classification
MLC	Maximum likelihood classification
NDVI	Normalized Difference Vegetation Index
NIR	Near infrared
OBIA	Object-based image analysis
OOB	Out-of-the-bag
PA	Precision Agriculture
PDGS	Payload Data Ground Segment
QI	Quality Indicators
RF	Random Forest
RMSE	Root Mean Square Error
RTK	Real Time Kinematic
RTM	Radiative Transfer Model
S2A	Sentinel-2A
SCM	Scene Classification Map
SfM	Structure-from-motion
SIFT	Scale-Invariant Feature Transform
SITES	Swedish Infrastructure for Ecosystem Science
SNAP	Sentinel Application Platform
TIFF	Tagged Image File Format
TOA	Top of the atmosphere
TOW	Take-Off Weight
UAV	Unmanned Aerial Vehicles
UgCS	Universal Ground Control System
VI	Variable Importance
VTOL	Vertical take-off and landing
WV	Water Vapor

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. Introduction

1.1. Background

Increasing food demand due to increasing population is a problem that requires special attention. Intensifying agricultural production can adversely impact the environment, especially through land cover change and its resulting impacts on water quality (Zhang & Kovacs, 2012). Therefore, effort has to be put into the development of methodologies that help cope with this demand efficiently and minimize environmental deterioration. Numerous initiatives in agriculture aim to reduce the yield gap between realized productivity and the best that can be achieved by using current technologies and management systems (Talebpour, Türker, & Yegül, 2015), to increase crop production and increase the efficiency of land use. Food security policies rely on up-to-date spatial information such as that provided by Precision Agriculture (PA).

PA has been proposed as a way to achieve food security by monitoring crops closely to prevent damage resulting from undesired events through early intervention. PA is the application of a set of techniques and methods to retrieve useful information for monitoring and managing crops while considering landscape heterogeneity and variability within and between fields (Gago et al., 2015). The aim of PA is to optimize production efficiency and quality, improve land management, minimize environmental impact and risk, and reduce the uncertainties associated with the decisions required for local and site-specific management of crops (Rokhmana, 2015; Schellberg, Hill, Gerhards, Rothmund, & Braun, 2008).

Precision information is required periodically to detect variation in the fields, in order to determine the most effective management strategy and adjust agricultural practices accordingly. Remote sensing techniques have been used for agricultural applications for over three decades to collect data in crop fields at different spatial and temporal scales. Such techniques generally rely on the relationship between plant optical properties and bio-physiological parameters (Bolton & Friedl, 2013; Stroppiana et al., 2015). Monitoring crop development is an important decision support tool in remote sensing applications for Precision Agriculture since it allows assessment of the most critical stages of growth. Phenological monitoring also improves the understanding of crop development and growth process (Ballesteros, Ortega, Hernández, & Moreno, 2014b).

Satellite images and aerial photos collected during the growing season have been used to monitor crops. Several remote sensing techniques have been developed for data derived from sensors like MODIS (Bolton & Friedl, 2013; M. Wang, Tao, & Shi, 2014), SPOT-Vegetation (Kowalik, Dabrowska-Zielinska, Meroni, Raczka, & de Wit, 2014), NOAA-AVHRR (Balaghi, Tychon, Eerens, & Jlibene, 2008; Moriondo, Maselli, & Bindi, 2007), Landsat (Inglada et al., 2015) and others. However, it has been acknowledged that previously operational satellite sensors do not meet the requirements to simultaneously satisfy the needed temporal frequency and spatial resolution for such applications (Vega, Ramírez, Saiz, & Rosúa, 2015; Stafford, 2000).

Monitoring the development of crops is challenging in heterogeneous and cloud-prone landscapes. Crop area extent and crop type maps provide crucial information for this purpose, and to produce these we must know what crop is grown where, and when and how it is grown since only then we can rely on them for proper monitoring and management. Different crops exhibit similar spectral responses and appearance characteristics, thus imposing additional difficulties for identification. When data availability from a single sensor is insufficient, complementary use of different sensors can be useful to obtain a continuous time-series of high temporal and spatial resolution imagery that highlights phenological differences to improve the distinction between land cover types. The success of crop identification depends on various factors, including the choice of the spectral, spatial and temporal resolutions of the sensor, and the choice of a suitable classification procedure (Belgiu & Drăgu, 2016).

Image classification is one of the most often used methods of information extraction in remote sensing and is used to produce thematic maps from imagery (Choodarathnakara, Kumar, Koliwad, & Patil, 2012). Mapping vegetation in crop fields via image classification, i.e. crop mapping, is an important step for PA, since it provides important baseline information for numerous agricultural decision support and monitoring applications, such as crop area estimation, crop yield prediction or water demand calculation (Torres-Sánchez, Peña, de Castro, & López-Granados, 2014; Löw, Michel, Dech, & Conrad, 2013).

Recently, a non-parametric machine learning algorithm known as the Random Forest (RF) classifier has been used to effectively perform image classification (Belgiu & Drăgu, 2016). Non-parametric classifiers are not constrained to assumptions like the need for a parametric distribution of the spectral characteristics of input data, which is the case with parametric classifiers like Maximum Likelihood Classification (MLC). Inglada, Vincent, Arias, & Marais-Sicre (2016) highlighted the adequacy of the RF classifier for land cover mapping and justified the choice of this approach by noting that RF can yield high-quality mapping for a variety of crop types. In another study, Immitzer, Vuolo, & Atzberger (2016) used the supervised RF classifier for land cover classification in agriculture and forestry. Löw et al. (2013) reported that RF has successfully been applied to crop classification.

The continuous improvement of remote sensing technologies in terms of spectral, spatial and temporal resolutions now allows for monitoring crop growth and, the accurate classification of crop types at the field level. Commercial off-the-shelf Unmanned Aerial Vehicles (UAVs) have emerged as an attractive platform for acquiring high-resolution images at low altitude. Among the advantages are that UAVs provide a low-cost image acquisition platform with high spatial resolution and flexible flight scheduling. UAVs can carry miniature narrowband, hyperspectral radiometers or thermal cameras to capture patterns in biophysical variables. Revisit times can be optimized to match the phenological cycle of specific species, thus maximizing the availability of information needed for crop management. In addition, due to the flight altitude images are free of cloud contamination (Anderson & Gaston, 2013). UAVs have the potential to match with small-scale measurements and in-field ground truth measurements can better relate to remotely sensed data. Useful pioneering studies are available on vegetation mapping projects based on UAV imagery (Michez, Piégay, Lisein, Claessens, & Lejeune, 2016; Diaz-Varela, Zarco-Tejada, Angileri, & Loudjani, 2014).

At the same time, the recently launched Sentinel-2A and Sentinel-2B satellites carry a unique combination of global coverage: systematic acquisition, high revisit frequency, wide field of view, high spatial resolution and a large number of spectral bands (Gascon, 2014). At the time of this writing, the Sentinel-2B system has not reached its full acquisition capabilities. Another major advantage is that Sentinel-2 data is freely available online. Nevertheless, as with any other optical satellite sensor, cloud contamination limits applications such as crop growth monitoring, where detecting rapid surface changes is crucial.

One solution to overcome the discrepancy in requirements is to combine both sensors. However, prior to this, it is important to recognize the degree of agreements between sensors, as it is critical for the consistent retrieval of biophysical parameters and for detecting land-cover changes. Many studies have focused on data fusion approaches to generate high-resolution multispectral imagery for image sharpness and better visualization (Tu, Su, Shyu, & Huang, 2001). Only a few have examined those techniques in terms of outputs calibrated to spectral reflectance which is critical for many quantitative remote sensing applications (Gao et al., 2015).

This study takes advantage of considerable recent improvements in sensor technology and builds upon the extensive body of research that has been conducted over the past decades on crop monitoring to examine the quantitative changes in radiometry caused by crop phenology and the compatibility of radiometrically calibrated surface reflectances from a UAV and atmospherically corrected Sentinel-2A imagery. To accomplish this, physical parameters such as surface reflectance and vegetation indices were compared on a band-by-band basis. Biophysical parameters provide a

comparable measure from different sensors; reflectance acquired from the UAV can be aggregated and linearly compared to Sentinel-2A data when they are acquired from the same date and location. Another goal of this research is to explore the performance of the Random Forest classifier using Sentinel-2A multi-temporal data to differentiate crops at the plot level. To the best of the author's knowledge, there is no literature or publicly available product comparing the spectral response of UAV sensors and Sentinel-2A for crop mapping and monitoring at field level, nor studies exploiting the temporal information of Sentinel-2A. Within this context, this study contributes to the current state-of-the-art by using Sentinel-2A data to map crop types using actual Sentinel-2A data.

1.2. Research Problem

While UAV platforms have been exploited increasingly according to the state-of-the-art review by Colomina & Molina (2014), Sentinel-2A has been recently launched in June 2015 and only a few useful pioneering studies are available on Sentinel-2A data (Immitzer et al., 2016; Inglada et al., 2016), besides studies using simulations of Sentinel-2A time-series based on SPOT4 (Take5) (Battude et al. 2016; Inglada et al. 2015). However, at present, there is no single sensor that can provide sufficient data to assess the various growth stages and temporal changes of crops over the growing cycle. UAVs and Sentinel-2A offer a high potential to provide data to produce maps with enough detail to monitor individual parcels and in this way fill the gap between the availability of timely and accurate crop type maps and the user's needs.

Combining imagery from different sensors via data fusion is a commonly used practice to obtain higher quality results than an individual sensor can provide. These practices have been more common with moderate and medium spatial resolution sensors such as MODIS and Landsat (Knauer, Gessner, Fensholt, & Kuenzer, 2016). Few studies have combined data from high-resolution sensors like Tewes et al. (2015), that used RapidEye and MODIS to monitor vegetation. Similarly, UAV observations can be used to replace satellite imagery with cloud-cover, enlarge satellite time-series, or up-scale and generalize data on a larger scale.

Nevertheless, in applications where biophysical parameters matter and data consistency is required, spectral responses from both sensors should be examined to acknowledge the discrepancies that may arise due to differences in position, spectral bands, spatial resolutions, acquisition time and others, and to be aware of the suitability of the products for various applications. This study focuses on the comparison of two cost-efficient technologies for the acquisition of very high-resolution imagery, UAVs, and Sentinel-2A, to provide insight about the quality of the data products used for monitoring crops. We also examine how well Sentinel-2A can reproduce what it is captured by a UAV sensor and if UAV observations can add relevant information to time-series of vegetation indices to monitor the spatial and temporal variation of crop conditions.

1.3. Research Objectives

The main objectives of this study are to explore the utility of multi-temporal Sentinel-2A (S2A) data for identifying different crops and to examine the degree to which UAV sensors, specifically the MicaSense RedEdge, can complement or replace S2A data. At the same time, it is also of interest to evaluate the performance of the Random Forest classifier, implemented with a supervised approach using time-series imagery from Sentinel-2A, for crop type mapping and crop monitoring.

To achieve these objectives, the following specific objectives are defined:

1. To compare atmospherically corrected S2A imagery and UAV orthomosaics generated from raw image values and radiometrically corrected reflectance data.
2. To compare vegetation indices at the plot level generated from UAV orthomosaics and S2A imagery.

3. To conduct image classification with multi-temporal S2A imagery using the Random Forest classifier and a supervised approach. And to evaluate the accuracy and parsimony of the resulting classification.
4. To explore the potential of S2A time-series to detect phenological cycles of crops in the study area and to determine to what extent can UAV observations provide data when satellite imagery is not available.

The following research questions will be addressed:

1. What, if any, are the differences in reflectance values between imagery collected by UAVs and S2A?
   Hypothesis: Reflectance from UAV orthomosaics is linearly related to S2A imagery and there is no statistically significant difference between the spectral responses in the data recorded by these sensors.
2. What is the degree of correlation of vegetation indices calculated from UAVs and satellite sensors' imagery?
   Hypothesis: The correlation of vegetation indices is strong.
3. What is the accuracy of the resulting land cover classification using the described methods? Is it a function of the temporal range of the data included in the classification?
   Hypothesis: The accuracy of the land cover classification increases as more temporal data is included in the model.
4. Can S2A time-series data capture crop phenology in the study area?
   Hypothesis: Crop phenology can be seen consistently from S2A time-series.

2. Review of UAV and Sentinel-2A concepts

This chapter presents the main concepts used in the methodological framework to compare the compatibility of Unmanned Aerial Vehicles and Sentinel-2A for crop type mapping and monitoring.

2.1. Remote Sensing of Vegetation

Remote sensing allows us to collect information about objects or areas from a distance. The basis of many remote-sensing technologies is the detection of electromagnetic radiation by a sensor. Electromagnetic radiation interacts with the atmosphere and with objects at the surface of the Earth. Every object on Earth absorbs, reflects and transmits a portion of the incident radiation received at different wavelengths. The reflected radiation, together with the radiation emitted by the Earth, can be captured by remote sensors. Such remote sensors measure radiances that are recorded as digital numbers (DN). DNs are the quantized samples value of the electric signal that is generated by the detector. "The DNs correspond to photon energy incident upon the detector and radiances at the detector, but have not a meaningful physical unit" (ITC, 2013), since the degree to which DN correspond to radiances on the ground depends on many factors such as scene illumination, atmospheric scattering and absorption, and detector-response characteristics.

Based on the surface properties of an object, the reflectance response varies along the electromagnetic spectrum. Reflectance is the fraction of irradiance that is reflected as a function of wavelength (Eq. 1). Differences in the reflectance make it possible to identify different Earth surface objects by analyzing their reflectance response or spectral signature as a function of wavelength in a spectral response curve. Spectral reference characteristics allow a clear distinction between three basic types of Earth features: green vegetation, dry bare soil, and clear water.

$$\rho(\lambda) = \frac{radiance\ (\lambda)}{irradiance\ (\lambda)} \times \pi \quad (1)$$

Where:

ρ is reflectance as a function of wavelength
Irradiance is the energy reaching the surface as a function of wavelength [W m^{-2}]
Radiance is the energy reflected by the surface as a function of wavelength [W m^{-2} sr^{-1} μm^{-1}]

In optical remote sensing of vegetation, chlorophyll absorption is a fundamental biophysical variable that allows differentiation between vegetated and non-vegetated areas, and between different types of vegetation. The spectral signature of vegetation clearly shows that due to the absorption of red light by leaf pigments like chlorophyll for photosynthesis, the reflectance in the red region of the electromagnetic spectrum is very low (Fig.1). The two optimum wavelength intervals for sensing chlorophyll absorption are believed to be 0.45 – 0.52 μm and 0.63 – 0.69 μm (Jensen, 2000). The spongy mesophyll cells and the amount of water in the plant cause scattering of the NIR light, thus a high reflectance in this region of the spectrum. Light in wavelengths at 0.97, 1.19, 1.45, 1.94, and 2.7 μm is absorbed by the content of water in the leaf (Jensen, 2000).

Chlorophyll *a* and *b* pigments tend to dominate the characteristics of the resulting spectral signature over other pigments, but pigments like carotenes and xanthophyll absorb significant portions of the blue incident energy, and phycoerythrin absorbs predominantly in the green region. When a plant undergoes senescence or encounters stress, chlorophyll may disappear and allow other pigments to become dominant and have a much higher reflectance, particularly in the green and red portions of the spectrum (Jensen, 2000). In this way, increased reflectance in portions of the visible spectrum is considered to be the most consistent response to plant stress detectable by optical remote sensing, rather than NIR reflectance, since only once the stress has developed sufficiently can a change in the NIR reflectance be noticed.

The accuracy and precision with which a remote sensor can distinguish different types of vegetation depend on the spectral bands, the spatial and temporal resolution, and the choice of a suitable classification procedure. Information derived from remote sensing imagery is useful for decision-making in crop management, yield forecasting, and environmental protection. The continuous improvement of remote sensing technologies, combined with more powerful computers and new geomatic procedures for extracting information make remote sensing approaches more useful for monitoring vegetation (Michez et al., 2016a).

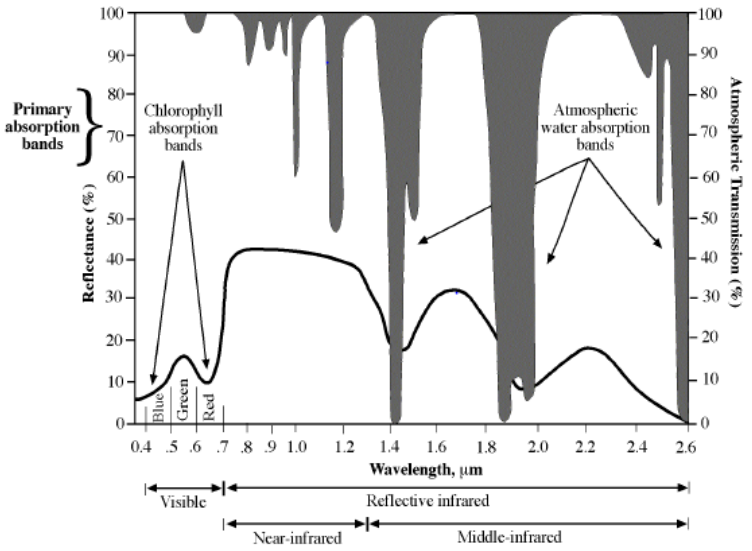


Figure 1. Spectral reflectance characteristics of healthy green vegetation over the wavelength interval 0.4 – 2.6 μm. Figure also shows atmospheric water absorption bands (modified from Jensen, 2000)

2.2. Vegetation Phenology

Spectral signatures of vegetation appear to be very similar throughout the visible and NIR, especially in broadband multispectral sensors. Besides considering the spectral response of vegetation, different temporal characteristics, such as information about the phenological (growth) cycles can provide valuable information for identifying different vegetation types or for extracting useful vegetation biophysical information (e.g., biomass). Similarly, hyperspectral sensors may identify unique absorption features (Jensen, 2000).

Phenology is the study of periodical phenomena of organisms and how these are influenced by seasonal and inter-annual variation in climate (Jin & Eklundh, 2014). Many plant species leaf-out, grow to maturity and reach senescence at about the same time each year, but unusually cold or warm temperatures in spring or fall can shift their annual phenological cycles by several weeks. Therefore, it is important to check if the remotely sensed data were collected in a typical or atypical year in order to properly interpret plants' phenology in that year.

The most productive growth period of vegetation, when scientists may want to collect the maximum amount of remote sensing data, is often associated with intense periods of precipitation and cloud cover. Therefore, the phenological calendar of natural vegetation systems and managed agricultural systems may provide useful information about an optimum time of the year to collect data when there still might be a shift in the growth cycle of plant species or crop types that will allow a distinction among them.

Jensen (2000) mentions that for purposes of classification of vegetation, it is wise to collect remote sensing data early in the growing season when the vegetation is developing at different rates when hopefully there is a detectable difference in their spectral signals. However, it is more useful to collect remote sensing data at the peak of the growing season if the purpose is to monitor the biomass of vegetation.

2.3. Vegetation Indices

Usually, vegetation indices are dimensionless radiometric measures extracted from remotely sensed data. They are used to model various vegetation biophysical variables. At the same time, vegetation indices exploit the characteristic response of vegetation over the different wavelengths of the electromagnetic spectrum to maximize the contrast of their spectral signatures to differentiate vegetation types. In general, healthy or dense vegetation will absorb most of the red light that it receives and will reflect back a large portion of the NIR light; conversely when vegetation is sparse or not so healthy, there is a decrease in the NIR reflectance and an increase in the red as there is less chlorophyll to absorb the light.

Table 1. Vegetation indices

Name	Formula	Source
NDVI	$NDVI = \frac{NIR - Red}{NIR + Red}$	Tucker 1979
EVI	$EVI = 2.5 * \frac{NIR - Red}{1 + NIR + 6 * R - 7.5 * B}$	Kerr and Ostrovsky 2003
GCC	$GCC = \frac{G}{R + G + B}$	Richardson et al. 2009

The Normalized Difference Vegetation Index (NDVI) is a common technique used to help identify the presence of vegetation and to provide a measure of its health and vitality (Jensen, 2005). NDVI combines the information of the red and NIR wavelengths into a single representative value. NDVI is also useful in identifying other features in an image; those that show a distinct behavior in the red and NIR spectrum. It has to be noted, however, that NDVI is sensitive to canopy background

variations and saturates at very dense vegetation levels while others indices, such as Enhanced Vegetation Index (EVI), still respond. EVI is a modification of the NDVI with a soil adjustment factor and two correction coefficients for atmospheric aerosol scattering. EVI has improved vegetation monitoring through a separation of the canopy background signal and a reduction in atmospheric influences (Jensen, 2005). Similarly, the Green Chromatic Coordinate (GCC) is a color index that attempts to suppress the effect of changes in scene illumination and was found to outperform NDVI when using near-surface remote sensing images (Richardson, Braswell, Hollinger, Jenkins, & Ollinger, 2009).

2.4. Image Classification

Image classification is often described as the process of assigning pixels to nominal classes. The input for this process is an image with multiple bands, and the output is an image in which each cell has a thematic code. Classification is generally conducted to extract land information from remote sensing imagery for two purposes: The first is to generate maps of categorical variables (thematic classes) such as crop mapping, while the second is to generate maps of continuous variables for quantification of biophysical variables for modelling such as LAI, biomass, tree volume, yield, etc.

The process of image classification typically involves five steps (ITC, 2013):

1. Selection and preparation of remote sensing images considering the spatial resolution, spectral bands, sensor, acquisition date(s).
2. Definition of clusters in the feature space depends on the data available and the knowledge about the study area to select a supervised or unsupervised classification approaches.
3. Algorithm selection whether parametric or non-parametric.
4. Classification at pixel or object level
5. Validation of the classification.

Supervised classifiers learn the characteristics of target classes from training samples and identify these learned characteristics in the unclassified data. When the knowledge of the characteristics of targets is limited or classes have not yet been defined, an unsupervised classification can be conducted where clustering algorithms are used to partition the feature space into a number of clusters (ITC, 2013).

Parametric classifiers like Maximum Likelihood Classification (MLC) give good results when dealing with unimodal data, but when handling multi-modal datasets they have limitations because of the assumption of normally distributed data and remotely sensed data rarely have normal distributions (Belgiu & Drăgu, 2016). Non-parametric machine learning classifiers like Random Forest (RF), Support Vector Machine (SVM) and Artificial Neural Networks (ANN) can overcome these limitations since they are not constrained by assumptions regarding frequency distribution (Löw et al., 2013).

Pixel-based classification is often used to derive thematic classes from multi-band images based only on their spectral characteristics. This technique assigns each pixel to only one class, which is not a problem when dealing with relative small ground resolution cells where different land cover classes are less likely to occur. However, even with higher spatial resolution imagery, in some applications, an approach purely based on the spectral characteristics of pixels is unable to distinguish desired classes. One solution is to use a different approach, for example, instead of classifying every pixel separately based only on spectral properties, object-based classification considers segments with homogeneous features such as shape, texture, and context, besides the spectral information.

An important part of using classification is the validation of the results. The most commonly used method for accuracy assessment may be derived from a confusion or error matrix. A confusion

matrix is a cross-tabulation of the classes in the classification result against the reference data derived preferably from field observations, also referred to as validation or testing sample set, although sources of higher accuracy are sometimes used. Another way of measuring the overall accuracy of the map is the Kappa coefficient, which can be derived from the confusion matrix. Kappa coefficient takes into consideration the fact that even assigning labels at random will result in a certain degree of accuracy. When the labels are assigned randomly a Kappa of 0 is expected. Or to put this another way, a result of Kappa ~ 0 indicates the classification is no better than one constructed by random assignment of classes. If Kappa is greater than 0.80, it means that there is a strong agreement between the classification map and the reference data (and the performance is considerably better than could be achieved by chance), if Kappa values fall between 0.40 and 0.80 there is a moderate agreement and if less than 0.40 there is a poor agreement (Jensen, 2005).

2.4.1. Random Forest Classifier

Random Forest (RF) is a classifier that has become increasingly popular in remote sensing due to its classification accuracy and speed of processing. This approach is not used only for classification, but also to assess variable importance. It has been implemented in several software packages, such as eCognition, R software and Matlab (Belgiu & Drăgu, 2016).

RF is a "black-box" machine learning ensemble classifier that uses a set of classifications and regression trees (CART) to make a prediction. The trees are created by drawing a random subset of training samples through replacement, following a bagging approach, to produce multiple decision trees. The RF classifier is sensitive to the sampling procedure, to the size and representativeness of the training sample. The samples are divided into two groups: In-bag samples and out-of-the-bag (OOB) samples. About two-thirds of the samples are used for training the tree (in-bag samples), and the remaining one-third of the samples are used for validating and estimating the performance of the RF model (OOB samples). The estimated error of the model is called OOB error; this error can be used to as a reliable measure of classification accuracy (Belgiu & Drăgu, 2016).

Two parameters need to be set in order to produce the trees: (1) The number of decision trees to be generated (*Ntree*), and (2) the number of variables to be selected and tested for the best split when expanding the trees (*Mtry*). Based on a review on RF in remote sensing, Belgiu & Drăgu (2016) suggest a value of 500 for *Ntree* and mention that *Mtry* is usually set to the square root of the number of input variables. Before the final classification, the class assignment probabilities calculated by all produced trees are averaged and a new unlabeled data input is evaluated against all decision trees created in the ensemble. Then, each tree votes for a class membership, and finally the membership class with the maximum votes will be the one that is finally selected (Belgiu & Drăgu, 2016).

In addition, RF functions provide information about variables and datasets that have greatest ability to discriminate between classes; one of these is variable importance (VI). VI can be used to identify the optimal variables computed from a single sensor, from several different sensors, or from a combination of remote sensing and ancillary geographic data. VI measurement could be used to identify the most suitable seasons for identifying desired classes. To increase the classification accuracy, input datasets identified as important during the first run of the classification can be used to rerun the classification until the OOB error no longer varies. This procedure is known as iterative backward feature elimination (Belgiu & Drăgu, 2016).

VI can be calculated using the Mean Decrease in Accuracy (MDA) or the Mean Decrease in Gini (MDG). MDA considers the difference between OOB error from the random permutations of the values of different variables and the OBB error from the original dataset. MDG measures how much a variable reduces the Gini Impurity metric in a particular class (Belgiu & Drăgu, 2016). Gini Impurity refers to how often an element selected randomly from the set would be incorrectly labeled if it were randomly labeled according to the distribution of labels in the subset. The majority of the studies reported by Belgiu & Drăgu (2016) used the MDA to determine the VI.

The RF classifier has been used successfully to extract information from imagery acquired from various remote sensors, including Landsat (Frazier, Coops, Wulder, & Kennedy, 2014), MODIS (Deng & Wu, 2013), IKONOS (H. Wang, Zhao, Pu, & Zhang, 2015), RapidEye (Löw et al., 2013), WorldView-2 (Räsänen, Rusanen, Kuitunen, & Lensu, 2013), SAR (Inglada et al., 2016), LiDAR (Li, Cheng, Chen, Chen, & Liu, 2015) and UAV (Michez, Piégay, Jonathan, Claessens, & Lejeune, 2016). Moreover, it allows the user to combine multi-source and multi-temporal data to improve classification accuracy. However, only relevant datasets should be included to reduce the computational burden (Belgiu & Drăgu, 2016).

2.5. Unmanned Aerial Vehicles (UAVs)

Unmanned aerial vehicles carrying imaging payloads are frequently used for remote sensing applications to acquire high-resolution images for the monitoring of various aspects of agriculture and the environment. Progress made in the miniaturization and reduction in the cost of sensors, Global Position System (GPS) devices, and embedded computers have increased the possibility of remote sensing using commercial off-the-shelf (COTS) UAVs (Diaz-Varela et al., 2014).

The two most important characteristics of UAV platforms are said to be the very high spatial and temporal resolutions they support, which allow the recognition of events occurring at a local scale in a specific time window (Michez, Piégay, Lisein, et al., 2016). These advantages make UAVs feasible for monitoring crops in greater detail, capturing important phenological stages for Precision Agriculture, which is a classic limitation of traditional remote sensing platforms. The flexibility in acquisition times and the lower cost of operation may exceed the demand from traditional manned aircraft and other platforms like WorldView-2, WorldView-3, GeoEye-1, IKONOS, Quickbird, and RapidEye.

Along with these advantages, there are well-known technical problems for UAVs, such as engine power, short flight duration limited by the weight of the payload and battery life, stability in winds and turbulence, and variation in sun angle (Zhang & Kovacs, 2012). Another key issue that restricts sensor selection is UAV payload capacity. Several consumer cameras have been tested due to their low weights and low costs in various studies; for instance, by Lelong et al. (2008). However, many problems arise with these cameras: These include limited optical quality, zoom lenses, fully automatic focusing and a lack of a NIR band for vegetation surveys.

Different types of vehicles have different capabilities, with advantages and disadvantages depending on the purpose. Compromises must be made between ease of flying, distance covered, payload capacity and takeoff and landing requirements (Ballesteros, Ortega, Hernández, & Moreno, 2014a). UAVs can be classified according to the method of takeoff and landing; whether vertical or horizontal, they can be either fixed-wing aircraft or rotary-wing in design (helicopter and quadcopter) with various remotely sensed sensors used for data collection, such as optical cameras, NIR and multispectral, hyperspectral, thermal, and laser scanners. For an overview of the current status of the main UAV technologies, the reader is referred to Colomina & Molina (2014) or Anderson & Gaston (2013).

The entire unmanned aerial system consists of (1) the vehicle, (2) the radio control transmitter, (3) a ground station with the software for mission planning and flight control, and (4) a telemetry system. The main tasks of the handheld radio control transmitter are to start the engines, manage takeoff and landing, control the complete flight in the manual mode, and activate the autonomous navigation system. The ground station works as an interface between the operator and the vehicle and includes support software that allows the UAV settings to be configured, implements the flight route plan with the waypoint module, and monitors the flight. The telemetry system collects relevant flight data that includes vehicle's position, flight altitude, flight speed, flight time, battery level, and radio control signal quality, among many others. Detailed information about the configuration of the UAV flights and specification of the vehicle and the camera used can be found in Torres-Sánchez, López-Granados, De Castro, & Peña-Barragán (2013).

The main components of the navigation system in a UAV are a GPS receiver, an Inertial Measurement Unit (IMU), a magnetic compass, and an altimeter to ensure flight stabilization. The GPS receiver reports and controls the location of the UAV. The IMU uses an accelerometer and a gyroscope to measure acceleration and rotation of the UAV. This electronic device allows basic stabilization of the UAV and facilitates the manual control task of the pilot. The magnetic compass indicates direction relative to the Earth's magnetic field, and the altimeter measures the distance of the vehicle from the ground to support accurate flights and landings.

The recent and rapid growth of the use of UAVs in environmental and agricultural applications has prompted the development of innovative methodologies. Several studies were conducted using UAVs to assess the potential of images to support Precision Agriculture. Lelong et al. (2008) assessed UAV imagery for monitoring a wheat crop in small plots. Ballesteros et al. (2014b) examined the relationship between green canopy cover and leaf area index in order to characterize crop growth. They concluded that high-resolution images obtained with UAVs, together with proper treatment, might be a useful tool for precision monitoring of crop growth and development, and has the potential to advise farmers on water requirements, yield prediction, and weed and insect infestation, among others. Stroppiana et al. (2015) tested the correlation between reflectance in the spectral channels and vegetation indices derived from imagery acquired with a multispectral sensor on board a UAV in an effort to estimate rice yield. Their results suggested that UAVs are a potential platform for data acquisition for crop monitoring.

Few studies have taken advantage of the temporal resolution of UAV imagery. Michez et al. (2016) proposed a methodological framework to explore multi-temporal imagery to identify variables and most relevant image scale in order to characterize riparian forest species and health condition. In Precision Agriculture, Bendig et al. (2014) computed a UAV derived digital surface model time-series to measure the growth rate of experimental field crops, while Torres-Sánchez et al. (2014) used multi-temporal imagery to differentiate weeds in wheat fields in the early season and map vegetation fraction using visible spectral indices.

During a UAV mission, several images are captured which have different geometric and radiometric features due to the low altitude of the flight, various viewing angles, the stability of the camera, the motion of the camera, etc. Usually, these images need to be combined to obtain a scene image, often referred as orthomosaic or orthophoto. To overcome geometric differences traditional airborne photogrammetry or remote sensing techniques are difficult to apply. However, state-of-the-art computer vision techniques such as Structure-from-motion (SfM) enable the generation of 2-D imagery, i.e. orthomosaic, from the collection of large sets of multi-angle images (Diaz-Varela et al., 2014). More about this technique is discussed in Section 2.5.2.

2.5.1. UAV Data Acquisition

The flight can be controlled manually or can be done under autonomous control between GPS waypoints using pre-programmed flight plans. Manual control is generally more useful for inspections and autonomous for gathering information, in order to create a map or a 3D model after the flight is completed.

Before starting the UAV mission, the design of the flight path covering the area of interest needs to be defined, along with the altitude of the flight, and the sensors and UAV configurations. The flight path is typically done using a flight planning software, many of which are offered by the UAV manufacturers. The path usually follows a pattern of parallel lines, commonly referred as transects, which are connected to a series of waypoints. The flight path should ensure that enough images are captured and that they overlap to the degree required for the reconstruction of the scene (Greenwood, 2015).

Forward and side overlap between images allows the identification of common points between images during the processing stage in order to successfully reconstruct the image of the entire study area. Forward overlap, or along-track, refers to the percentage of image overlap between its

previous or next image with respect to the flight direction. Forward overlap ensures that the distortion introduced by the relief does not cause gaps between the images. Side overlap, or cross-track, refers to the percentage of image overlap between images from one track and the next with respect to the orthogonal direction to the UAV. To achieve a certain image overlap, the flight speed, the interval at which the camera is triggered, the altitude of the fight, the distance between the transects, and the internal geometry of the camera being used should be considered. However, today's flight planning software will automatically calculate all these parameters (Greenwood, 2015).

The altitude of the flight, the focal length of the camera, and the size of the sensor determine the ground sampling distance (GSD), also referred to as pixel size on the ground. Resolution in aerial photography is measured as GSD, which is equivalent to the ground resolution cell (GRC) (Greenwood, 2015). The GSD is the distance between the centers of two adjacent resolution cells measured on the ground (ITC, 2013), or the length on the ground corresponding to the side of one pixel in the image (GRC). Defining the appropriate flight altitude is a key aspect in PA applications, and it depends on the scale of what needs to be captured by the imagery. J. Torres-Sánchez et al. (2014), acquired images with a UAV over a wheat field in the early-season period and studied the influence of flight altitude. The results suggest that a flight altitude of 30 m provides better accuracy for vegetation fraction mapping using visible spectral indices, than an altitude of 60 m. However, it is important to acknowledge that flight altitude is a trade-off between greater detail, and area covered.

Based on the area to be mapped, the specifications of the sensors and the flight altitude, the ground control software, like MDC, Mission Planer or UgCS, generates a series of transects with waypoints and displays the estimated ground sampling distance to ensure adequate overlap and area coverage. Using the software, the flight route can be uploaded to the UAV computer. Connectivity and proper functioning of the cameras and the whole system should be checked. Usually, image triggering is activated by the UAV according to the programmed flight path. At the moment of each shoot, the on-board computer system records a timestamp, the GPS location, the flight altitude, and vehicle principal axes (pitch, roll, and yaw) (Greenwood, 2015).

It is also important to take into consideration the flying mode of the UAV. There are three modes; (1) manual; (2) stop mode; and (3) cruising mode. The first mode is used when no flight planning is available. The other two modes require planning. When configured in stop mode, the UAV stops in each waypoint position and only then the image is taken. When in cruising mode, the UAV continues flying at the same speed while images are being taken. Flying in cruising mode reduce the flight time by 75% (Mesas-Carrascosa et al., 2015). The turning mode can be of three types, namely (1) bank turn; (2) adaptive bank turn; and (3) stop and turn. In the first one, the vehicle keeps its speed when turning, so it will deviate from its route. In the second mode, the vehicle slows down so as to follow the route. When in stop and turn mode, the vehicle stops and then turns (DJI, 2014).

Another aspect to consider when acquiring UAV images is radiometric calibration. As mentioned earlier, sun angle and clouds may affect the radiometric quality of images. During pre-processing routines, images should be corrected for changing illumination. For this effect, *in situ* radiometric measurements should be conducted. There are several ways of doing this; the most common being placing reflectance targets or panels (that have a high albedo and are spectrally homogeneous) on the ground with known reflectance which are later found in the images, as is the case with GCPs. Digital images captured by each camera band/channel can be spectrally corrected by applying an empirical linear regression to convert digital numbers to reflectance values in all the bands (Torres-Sánchez et al., 2014). A similar result can be achieved using a spectrometer, where DN can be calibrated to surface reflectance values using an empirical model and *in situ* spectral measurements (Stroppiana et al., 2015). This type of correction is called relative correction since the user makes a scene-dependent relative atmospheric correction. It is also possible to use radiative transfer models (RTM) for atmospheric correction, in which case the correction would be "absolute correction" since the atmospheric influence is modeled and removed (ITC, 2013).

The weather has become the most important operational consideration for flights. Special care is needed to operate UAVs. Winds, turbulence and cold temperatures at designated altitudes are some of the meteorological conditions that affect UAVs and must be accounted for (Teets, Donohue, Underwood, & Bauer, 1998). In terms of radiometric quality, although low altitude flights limit atmospheric effects, weather conditions still pose some limitations during image acquisition. Sun angle and cloud shadows may reduce the quality of the orthophoto. Acquiring images at high solar angles around solar noon is recommended to reduce the effect of shadow. Another factor that influences image quality is blurriness. According to Wijesingha (2016), images can be made blurry by the motion of the UAV, cloud shadow and wind. Greenwood (2015) adds turbulence to the list of causes of blurriness and argues that reducing the speed of the UAV's flight and stabilizing the camera with a gimbal system can ameliorate the problem.

2.5.2. UAV Image Processing

Raw images acquired from a UAV contain geometric distortions due to variations in camera orientation and rotation. Usually, all of these images need to be combined to obtain a single scene image for further analysis. Image mosaicking is the process of combining all the individual and overlapped images. Accurate orthomosaic generation relies on rigorous photogrammetric methods. Recent developments in the computer vision field, have developed methods such as Structure-from-motion that when integrated with photogrammetry techniques like relative orientation, bundle adjustment, feature detection, and Multi-View Stereo (MVS) they are ideal processes for use with UAV acquired imagery.

Structure-from-Motion (SfM) has been frequently applied (Westoby, Brasington, Glasser, Hambrey, & Reynolds, 2012) to process UAV images and create orthomosaics: in this case motion refers to the camera motion, and structure refers to the 3-D point coordinates. SfM is a technique that is well suited for a set of multiple overlapping images derived from a moving sensor, such as UAVs. It allows the structure reconstruction of a scene from a number of images with corresponding points. SfM algorithms can be used to generate orthophotos and 3-D surface models from a collection of 2-D digital images by calculating the 3-D structure of a scene. This technique differs from conventional photogrammetry as the geometry of the scene, camera position and orientation are simultaneously and automatically solved using a highly redundant, iterative bundle adjustment procedure based on matching features automatically extracted from the multiple views overlapping images (Ma, Soatto, Košecká, & Sastry, 2004; Westoby et al., 2012).

One key development in SfM is the Scale-Invariant Feature Transform (SIFT) algorithm for feature detection and matching. The first stage of a SfM process is to extract features in single images that can be matched to their corresponding feature in other images, followed by a rough alignment of images with enough number of features. These matched points establish the relative location of the sensors during the flight and are used to calculate simultaneously the sensor parameters of each image. The SIFT method uses an incremental approach in which the bundle adjustment of an initial image pair is repeated sequentially incorporating more images at each iteration. The final product of the SIFT process is a sparse 3-D point cloud that represents the matching features and camera location for each photo (Mesas-Carrascosa et al., 2015; Nesbit, 2014).

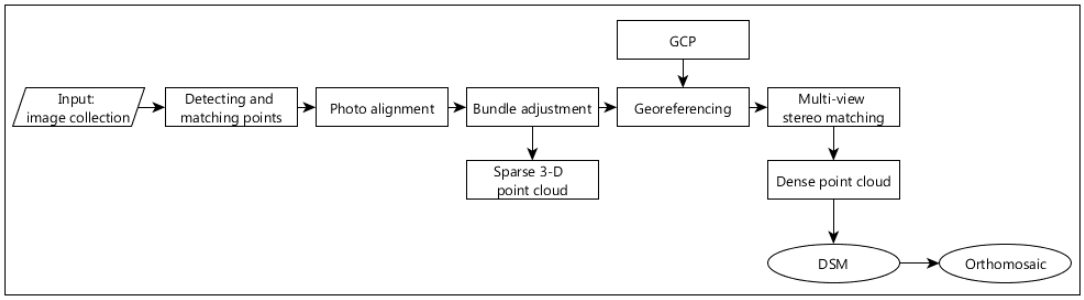


Figure 2. Simplified SfM-MVS workflow to produce an orthomosaic from UAV images

Moreover, an additional step that comprises Multiple-View Stereo techniques follows the SfM processing, since SfM alone has proven insufficient for studies with high accuracy and fine detail (Nesbit, 2014). SfM-MVS produce a reconstruction of the surface by a dense point cloud with up to three orders of magnitude more points than the sparse point cloud generated only by SfM (Mesas-Carrascosa et al., 2015). The final products of the image processing are a digital surface model (DSM) and an orthomosaic image or orthophoto. Several open source software and commercial packages that have implemented SfM-MVS processes are available; among the most popular are Agisoft PhotoScan, Pix4Dmapper, Visual SFM and 123D Catch.

2.5.3. Geometric Calibration

In order to assign spatial coordinates to the orthophoto, it is required to georeference the photos during image pre-processing. Georeferencing is an essential step in image processing since it allows the orthomosaic to be associated with a real-world scale and integrated with other spatial data. Geometric correction allows users to accurately measure distances in the resulting orthomosaic. Ground control points (GCPs) or targets placed on the ground, with surveyed coordinates before the UAV flight, are used to improve the accuracy of georeferencing images. GCPs enable the image processing software to know the real world coordinates of a number of visibly identifiable locations in the imagery. GCPs also are used to improve the SfM process during photo alignment and calculations of the camera calibration parameters by minimizing the error between the modeled locations of the points and measured locations (Cook, 2016). However, some software packages no longer require GCPs to create reasonably geographically accurate images, they instead use UAV acquired images roughly geotagged based on GPS position data embedded in the camera. However, depending on the type of GPS receiver, the coordinates might not be accurate enough for detailed studies.

Generally, a minimum number of GCPs is required by the software for the referencing process to function, Greenwood (2015) recommends around five. Mesas-Carrascosa et al. (2015) tested two GCP configurations (4 and 5) over a wheat field of 1.12 ha and found that placing 5 GCPs provides the best setting to maximize spatial resolution. Nevertheless, it is necessary to consider the area to be covered; more ground control points will yield more accurate results. With respect to distribution, GCPs should be scattered around the area to be mapped for better results. The distribution and number of GCP influences the geometric accuracy of the model. The size of the GCP, as a rule of thumb, should be a minimum of 2 or 3 times larger than the image spatial resolution in order to avoid mixed pixels (mixed land cover) and recognize it in the images during pre-processing routines (ITC, 2013).

The geometric accuracy can be estimated using the root mean square error (RMSE) statistic. The RMSE measures the error between two datasets by comparing a predicted value and an observed one. It is important to note that the RMSE is a convenient measure of overall accuracy, however, it does not say anything about which parts of the image are accurately georeferenced and which parts are not. Note also that the RMSE is only valid for the area bounded by the GCPs. The RMSE in the x-direction, m_x, is calculated with Eq. 2, for the y-direction with Eq. 3 and the overall error, m_p is calculated by Eq. 4 (ITC, 2013).

$$m_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{x}_i - x_i)^2} \tag{2}$$

$$m_y = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{3}$$

$$m_p = \sqrt{m_x^2 + m_y^2} \tag{4}$$

Where:
n is the number of GCP

$\hat{x}_i, \hat{y}_i$ are the geographic coordinates in the image of point i after georeferencing
x_i, y_i are the geographic coordinates location of the GCP in the point i

2.5.4. Radiometric Calibration

The low altitude flight of UAV typically provides images that have a higher radiometric homogeneity when compared to higher altitude aerial or satellite imagery (Lelong et al., 2008). However, the variation of the solar illumination angle during the day, together with bidirectional reflectance distribution function (BRDF) effects, and cloud shadows, may affect the radiance values measured by the sensor (Zhang & Kovacs, 2012). Radiometric calibration provides transfer functions to transform DN values to at-sensor radiance, which in turn have to be corrected by the user for atmospheric effects to finally obtain at-sensor or surface reflectance (ITC, 2013).

Applications for which the actual radiance at ground level is needed to infer chemical properties of surface material must include rigorous atmospheric correction (AC) procedures. The correction can be relative or absolute. Relative correction is based on ground reflectance properties, while absolute correction is based on atmospheric process information (ITC, 2013). There are some techniques that can be applied to improve the radiometric quality of UAV images, whether in visible or thermal ranges. Atmospheric correction for the visible range will be described since it is the range of the spectrum captured by the sensor in this research.

Sun elevation correction

When analyzing a sequence of images taken of a particular area on different dates, or when using UAVs, for a set of images of the same date taken at different times with which we would like to make a mosaic, any changes in illumination will cause difficulties. To correct for sun elevation, each image can be normalized as if it was taken with the sun at its zenith. To achieve this, every pixel value of an image is divided by the sine of the sun elevation angle at the time of data acquisition. The sun elevation angle usually can be found in the metadata file supplied with an image and some websites provide estimations of angle. Considering that the atmosphere is dynamic, mainly at low altitudes, this is an approximate correction.

Relative AC methods based on ground reflectance

These methods rely on the assumption that, for one sensor band, a linear trend establishes the relation between the radiances at top of the atmosphere (TOA) and at ground level. According to ITC (2013), there are two methods of AC, namely two reflectance measurements and two reflectance surfaces.

Two reflectance measurements: For the application of this method the use of a portable radiometer is required. The radiometer should be capable of measuring in the same wavelength range as the image band to be corrected. The image output is an absolute atmospherically corrected image, that is to say, that it has physical units and the calculated ground radiances are compatible with actual atmospheric constituents and it can be used for multi-temporal comparison, parameter evolution assessment, and flux quantification.

Two reflectance surfaces: This method works by normalizing radiance values on individual bands and is valid for establishing a basis for uniform comparison, e.g. to derive land properties, or to study the evolution of non-flux related parameters such as indexes. The method relies on the existence of at least one dark and one bright reflective invariant area that are considered to retain their reflective properties over time. These areas can be artificial like reflectance/calibration targets or natural like deep reservoir lakes, sandy beaches or desserts, and large asphalted areas. The supposition is that the reflectance should always be the same for these pixels. If a difference occurs, it is attributed to a different state of the atmosphere. In this case, the atmospheric composition is unknown, but its influence is measurable by detecting the change in radiance for the reflective invariant areas. The image output allows comparative results since it matches reflectance that is

compatible with the atmosphere of a similar image taken on a different date, but absolute values of radiance are not obtained.

Relative AC methods, such as the one just described above, often use empirically derived models to convert radiance data to reflectance data based on reflective invariant areas. Techniques known as flat-field correction and empirical-line correction are examples. In the first one, the conversion is achieved by dividing pixel-by-pixel the whole dataset by the mean value of the reflectance target within the area. The second one requires a spectrometer to be used in the field to characterize the spectra of a dark and a bright target, to later force with an empirical correction model the best fit between sets of field and image spectra that represent the same ground areas. This method removes "atmospheric effects, residual instrument artifacts, and viewing geometry effects" (ITC, 2013).

Absolute AC methods based on atmospheric processes

Although these methods require a thorough description of the components in the atmospheric profile that is rarely available, they allow us to evaluate ground reflectance for any atmospheric condition. The output is an image that corresponds with the reflectance of the ground pixels with a maximum estimated error of 10%.

Radiative transfer models: Atmospheric constituents attenuate processes at different wavelengths and RTM can be used for calculating radiances for a variety of atmospheric and surface conditions. These models require a detailed description of the atmospheric component at specific altitudes throughout the atmosphere. Because of the many components to consider, RTMs are based on atmospheric common profiles, called standard atmospheres, which correspond to average atmospheric conditions for different parts of the Earth and are used as input for the models. Different models have different requirements and yield different accuracies.

2.6. Sentinel-2A

Sentinel-2A is a satellite sensor launched on 23 June 2015 by the European Copernicus program that provides geographical information in areas of land monitoring, emergency management, and security. Data can be modified and adapted by users interested in thematic areas such as mapping of land cover, classification and change detection, as well as forest management, water management, monitoring of vegetation, agriculture and food security. Sentinel-2A images provide complementary data to missions such as LANDSAT, SPOT-Vegetation and ENVISAT/MERIS sensors (ESA, 2017).

Sentinel-2A is a polar-orbiting, sun-synchronous mission that offers systematic global coverage of land surfaces, between latitudes 56° South and 83° North, as well as observation of inland waterways and coastal areas. The system has the following characteristics; high revisit time (every 5 days at equator under the same viewing conditions with two satellites); high spatial resolutions: 10m, 20m and 60m at-ground; multispectral information from 13 bands covering the visible, near infrared and short wave infrared part of the spectrum; and a swath width of 290 km, larger than previous multispectral optical missions such as SPOT and LANDSAT (ESA, 2017).

Table 2. Spectral bands and spatial resolution of Sentinel-2A

Band number	Band name	Central wavelength (nm)	Bandwidth (nm)
10m spatial resolution bands			
2	Blue	490	65
3	Green	560	35
4	Red	665	30
8	NIR	842	115
20m spatial resolution bands			
5	RedEdge	705	15

6	Vegetation RedEdge	740	15
7	Vegetation RedEdge	783	20
8a	Vegetation RedEdge	865	20
11	SWIR-Snow/ice/clouds	1610	90
12	SWIR-Snow/ice/clouds	2190	180
60m spatial resolution bands			
1	Coastal aerosol	443	20
9	Water vapor	945	20
10	SWIR-Cirrus	1375	30

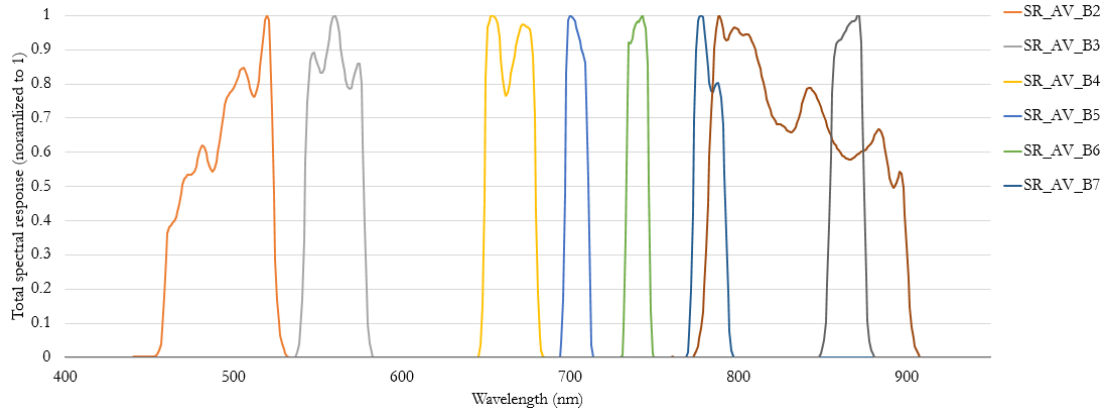


Figure 3. Sentinel-2A spectral response average – VNIR

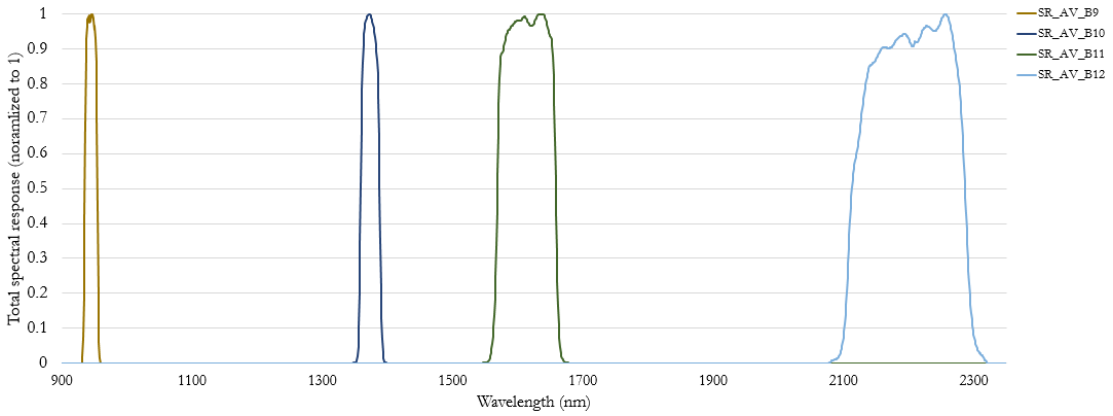


Figure 4. Sentinel-2A spectral response average - SWIR

Radiometric resolution

The capacity of an instrument to distinguish differences in light intensity or reflectance is called radiometric resolution and it is expressed as a bit number. The greater the radiometric resolution, the more capable the device on detecting differences in intensity or reflectance, and more accurate the sensed image will be. Sentinel-2A has a radiometric resolution of 12 bits, meaning that the acquired image can contain 2^{12} potential light intensity values (ESA, 2017).

Product types and data formats

Two types of Sentinel-2A products are available to users: Level-1C and Level-2A. The products are in Sentinel-SAFE format that includes image data in JPEG2000 format, quality indicators, auxiliary data, and metadata. The SAFE format wraps a folder containing image data in a binary data format and product metadata in XML. Level-1C images are processed by the Payload Data Ground Segment (PDGS) and distributed online. Level-1C product results from using a Digital Elevation Model (DEM) to project the image in cartographic geometry; it contains per-pixel TOA reflectances along with the parameters to transform them into radiances. In Level-1C products, pixel coordinates refer to the upper left corner of the pixel. Level-2A products are generated on the user side from Level-1C products through Sentinel-2A Toolbox to obtain bottom of the atmosphere (BOA) reflectance in cartographic geometry. The granules or tiles of both products are 100 x 100 Km^2 orthoimages in UTM/WGS84 projection and have a size of approximately 500 MB (ESA, 2017).

2.6.1. Sentinel-2A Data Acquisition

The Copernicus Open Access Data Hub in the official website of European Spatial Agency (ESA) provides complete, free and open access to Sentinel-2A products. Data is available to download at https://scihub.copernicus.eu/ from an interactive graphical user interface or in the form of scripting interface. Data can also be retrieved from other websites like the PEPS platform that redistributes products of Sentinel-2A and others satellites from the family (https://peps.cnes.fr/rocket/#/home).

2.6.2. Image Processing

Data acquired from Sentinel-2A is processed from Level-0 to Level-1C by the PDGS. The Sentinel-2A ground segment is in charge of the acquisition, processing, archiving and dissemination to the final users. Level-0, Level-1A, and Level-1C are not available to users.

To obtain Level-1C products from Level-1B products, two successive operations are performed: The first is a resampling of the Level-1B image to produce an orthoimage in TOA reflectance, and calculations of cloud and land/water masks for a tile in the Level-1C geometry. The resampling is conducted in five steps: (1) Selection of tiles intersecting the image footprint, (2) Projection (geographic coding by bi-linear interpolation), (3) Computation of resampling grids linking the image in native geometry to the target geometry (orthoimage), (4) Radiometric interpolation to estimate radiance values of each spectral band in the geometry of the orthoimage using the resampling grids, and finally (5) TOA reflectance calculation in the target geometry. A Level-2A image is produced through the Sentinel Application Platform (SNAP) toolbox using a plugin called Sen2Cor. Sen2Cor performs atmospheric, terrain and cirrus correction of TOA Level-1C input data. The Sen2cor output is an orthoimage BOA corrected reflectance product. Additional outputs are an Aerosol Optical Thickness (AOT) map, Water Vapor (WV) map, a Scene Classification Map (SCM) and Quality Indicators (QI) for cloud and snow probabilities (ESA, 2017).

3. Data and Methodology

3.1. Study Area

The Swedish Infrastructure for Ecosystem Science (SITES) is a national coordinated infrastructure for terrestrial and limnological field research. SITES has several research stations throughout Sweden. Lönnstorp Research Station is located in the Municipality of Lomma, Skåne County, Sweden, and is an experimental area dedicated to "studies of cropping systems ecology, with a focus on the design, sustainable development and assessment of arable cropping systems, in conventional and organic farming" (SLU, 2017). The Lomma region is characterized by an oceanic climate with an average annual precipitation of 666 mm. The soil type is loam with about 15 % clay and 3 % organic material.

Lönnstorp station covers an area of about 60 ha, where the main crop rotations are winter wheat, sugar beets, spring barley and winter rapeseed. The study was conducted on experimental plots that were part of an ongoing SITES project since 2015 called Agroecological Field Experiment (SAFE). In the growing season of 2016-2017, a total of 110 micro-plots of different sizes covering an area of 17 ha were delineated and sown with winter crops (wheat, rye, and rapeseed) in mid-August and early September. In addition to the experimental fields of Lönnstorp, 4 adjacent fields of Bomhög farm covering an area of 30 ha were included to increase the size of the study area (Fig.5).

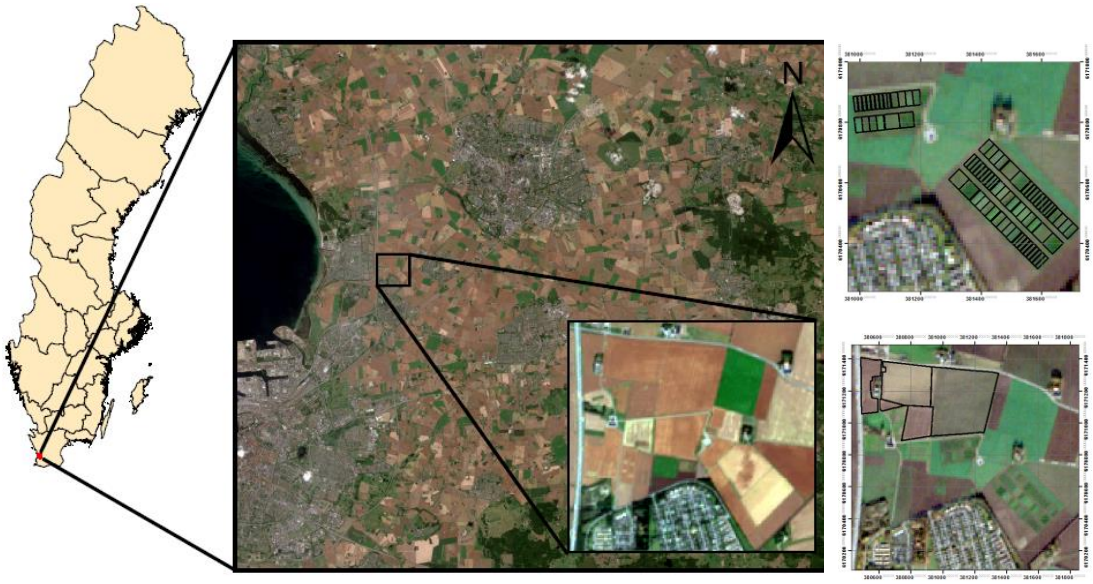


Figure 5. Location of the study area in the Municipality of Lomma, Skåne County, Sweden (left). SAFE plots distribution (upper-right), Bomhög plots (bottom-right).

3.2. Data

3.2.1. UAV imagery

A rotary-wing quadcopter platform with vertical takeoff and landing (VTOL), model Explorian 4, manufactured by Pitchup AB in Sweden, was used to collect aerial images at a constant flight altitude over the experimental crop fields, with a nadir view of direction and around noon. This UAV is equipped with four brushless rotor motors powered by a battery and can fly by remote control or autonomously with the aid of its GPS receiver and its waypoint navigation system. The technical specifications provided by the manufacturer and the operational configurations of the UAV are shown in Table 3.

Table 3. Technical specifications and operational configurations of UAV Explorian 4

Technical specifications	
Propulsion system	Four 22.2 V brushless electric motors.
Propellers	Four carbon fiber propellers of 45 cm diameter each.
Energy provider	One 24.0 V (22000 mA) battery.
Height	85 cm
Width	85 cm, 100 cm when propellers expanded.
Landing gear height	40 cm
Maximum TOW	7 Kg
Maximum payload weight	3.5 Kg
Max speed	13.88 m/s
Controller	FrSky Remote control of 500 m radius
Average flying time	Up to 37 minutes (depends on payload, wind, and weather)
Operational configurations	
Turning method	Adaptive Bank Turn
Flying altitude	75 m
Flying speed	5 m/s

Source: Pitchup AB (2015)

The sensors mounted on the UAV to acquire the imagery include a visible spectrum camera, a multispectral and a thermal camera. The visible and the thermal cameras were not used in this study. The MicaSense RedEdge is a lightweight multispectral camera designed especially for UAV mapping with a resolution of 1980 x 960 pixels and five spectral narrow bands described in Table 4. The spectral response of each band is shown in Fig. 6. The camera weighs 150 g and is powered by the UAV battery. The lens achieves 8.2 cm/pixel at 120 m above ground level (AGL). The images are 12-bit resolution that can be stored in either 12-bit DNG RAW format or 16-bit TIFF RAW format. Metadata tags are embedded for each image in standard format since the sensor has its own positioning system. The camera was set to operate in automatic mode, which adjusts the shutter speed (exposure time) and the aperture (F-stop) optimally. Image triggering was activated by the UAV according to the programmed flight route. At each shoot, the UAV computer system records a timestamp, the GPS location in WGS84 coordinate system, the altitude, and vehicle principal axes (pitch, roll and heading).

Table 4. Spectral bands, center wavelength and bandwidth of MicaSense RedEdge as compared to Sentinel-2A

	UAV		**Sentinel-2A**	
Band Name	**Bandwidth (nm)**	**Center Wavelength (nm)**	**Center Wavelength (nm)**	**Bandwidth (nm)**
Blue	20	475	490	65
Green	20	560	560	35
Red	10	668	665	30
Red Edge	10	717	705	15
NIR	40	840	842	115

Based on the distribution of the crop fields, the path was designed using the ground control software called UgCS (Universal Ground Control System) and then uploaded to the UAV computer system. The flight path considered forward and side overlaps of 80%. Six GCPs evenly distributed were placed on the ground and their position was measured accurately using a GPS real-time kinematic (RTK) satellite navigation receiver that recorded the coordinates in the reference system SWEREF99. For the radiometric correction of the images in the pre-processing stage, photos of the reflectance panels designed for use with the MicaSense RedEdge were captured before the flight ensuring no shadows were covering the panel.

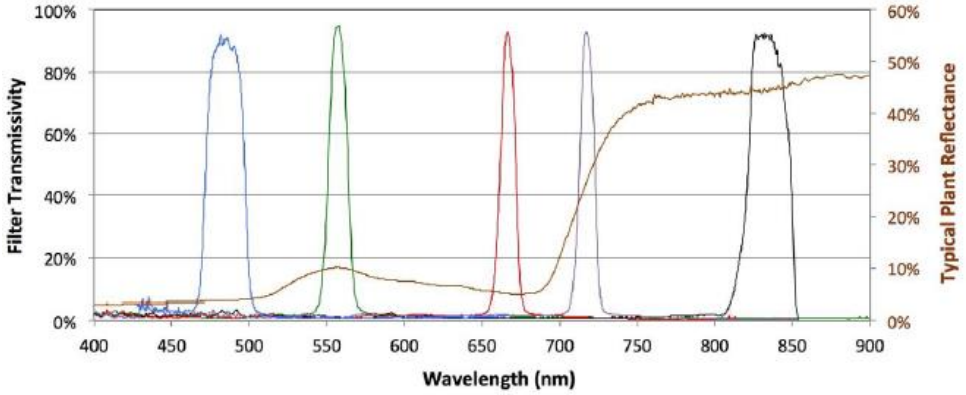


Figure 6. MicaSense RedEdge spectral response average (Source: MicaSense 2015)

Before the flight, the connectivity and proper functioning of the camera were checked. Then, using the radio control transmitter, the pilot launched the UAV and later activated the automatic mode. While the UAV was flying to the waypoints along the flight lines, another person was controlling the information provided by the telemetry system (UAV position, altitude, speed and battery level). Once the route was completed, or earlier if the battery was low or the conditions were unstable, the pilot manually landed the UAV. Finally, right after landing, another set of photos of the reflectance panels were taken. Images were acquired on five different dates during early-season crop development (emergence, tillering and growth after dormancy) and covered between 15 and 30 plots (Table 5).

Table 5. UAV missions

Date	Flight time	Images	Time	Illumination	Coverage (ha)
03-Nov-2016	8 min	158	11-12 am	Clear sky	5.44
02-Dec-2016	8 min	117	15-16 pm	Clear sky	3.64
07-Mar-2017	8 min	139	11-12 am	Overcast	4.15
13-Mar-20117	8 min	136	11-12 am	Overcast	4.36
24-Mar-2017	11 min	172	11-12 am	Clear sky	5.34

3.2.2. Sentinel-2A imagery

For the present study, all available Sentinel-2A Level-1C (TOA) products were examined for cloud coverage in the study area for the year 2016 and 2017. Only cloud-free images were downloaded from the official website of ESA and from the French website PEPS. A total of 17 images of tile 33UUB were downloaded, 15 from 2016 and 2 from 2017.

3.2.3. Field data

Inspections of the crops were conducted from November 2016 to March 2017 during the UAV missions. Phenological observations were performed at the same plots covered by the UAV flights and the height of the crops was recorded in each visit. In addition, on March 13th, 2017, spectral reflectances of three crops (winter wheat, winter rye, winter rapeseed) and bare soil were obtained using an ASD FieldSpec FR spectroradiometer with FOV 7.5° (350-2500 nm).

In addition, the Swedish Infrastructure for Ecosystem Science (SITES) provided information of the crop calendar for the years 2016-2017, and a land cover polygon layer of experimental plots. The phenological information of the crops is summarized in Table 6.

Table 6. Crop calendar 2016 in Lönnstorp farm showing the key phenological stages starting from the seeding, rising, blooming and harvest. The months where crops can be found on the ground are shown in grey.

	Crop calendar 2016											
Time	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Winter-rapeseed												
Sugar-beat												
Winter-wheat												
Spring barley/Lupin												
Grass-legume ley												
Spring wheat/Fava bean												
Spring wheat												
Spring barley												
Spring rapeseed												

	Seeding		Rising		Blooming		Harvest		On the ground

3.2.4. Sampling Strategy

The sample set for the calibration and validation of the supervised classification was generated following a stratified random sampling strategy including 10 strata (Built up area, rapeseed, spring barley, sugar beet, wheat, fava bean/wheat, grass-legume ley, kernza, kernza/lucerne, and spring barley/lupin). The sampling unit was one Sentinel-2A pixel (10 x 10 m). Based on the land cover polygon layer provided by SITES a set of 10 pixels per number of bands (10x11) for each class was considered. The sample size used for testing the performance of the Random Forest model (25% of the total sample set) was enough to achieve 85% accuracy according to the formula (Eq. 4) for the binomial probability theory suggested by Fitzpatrick-Lins (1981) for a land-use classification map.

$$N = \frac{z^2(p)(q)}{E^2} \quad (4)$$

Where:
z is the z-score for the standard normal deviate for the 95% two-sided confidence interval (1.96)
p is the estimated error of success (85%)
E is the level of error allowed (5%)
q is the percentage of error (q=1 – p)

3.3. Methods

3.3.1. UAV imagery pre-processing

Agisoft PhotoScan Professional Edition (Agisoft LLC, St. Petersburg. Russia) was used to generate orthomosaics from images captured by MicaSense RedEdge. Before starting the pre-processing, based on the image quality value estimated by the software, images were inspected and discarded if they were blurry. According to the Agisoft LLC (2016) user guide, images with a quality value of less than 0.5 units are recommended to be disabled and excluded from the photogrammetric processing. After Wijesingha (2016), the threshold considered in this study was 0.65.

The process followed to generate the orthomosaics is summarized in Fig. 7. The first step in the process was image alignment. The software looks for common points in the images and matches them, also finds the position of the camera for each image and refines camera calibration

parameters. The output of this process was a sparse point cloud. The next step was to georeference the images using the coordinates of the GCP. To optimize cameras, a routine was followed using the gradual selection tool to filter out points from the sparse point cloud based on reconstruction uncertainty, projection accuracy, and re-projection error. Once again the images were aligned, this time with the highest accuracy. Based on the estimated camera positions and images themselves a dense point cloud was built by the software. The last step was to build the DSM and based on this, the orthomosaic.

The orthomosaics were then projected to WGS 1984 UTM Zone 33N and the cell size resampled to 0.05 x 0.05 m with nearest neighbor interpolation since the original map was produced under geographic coordinates WGS 1984.

Another set of orthomosaics was generated by the ATLAS cloud-based data platform, a service provided by MicaSense for the RedEdge camera (https://www.micasense.com/atlas/), to obtain reflectance values from the radiometrically corrected orthomosaics. The reflectance values were scaled by a factor of 32768 for storage as 16-bit data.

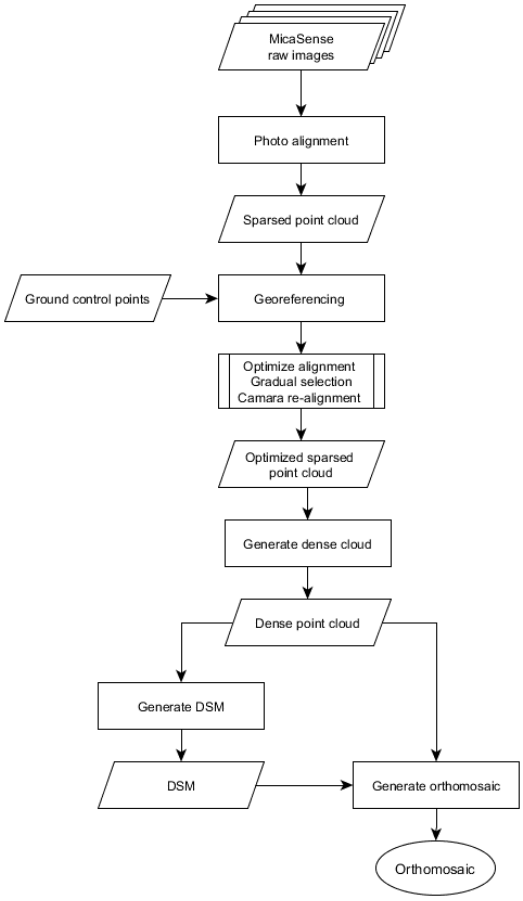


Figure 7. Workflow for orthomosaic generation using Agisoft Photoscan

3.3.2. Sentinel-2A imagery pre-processing

Images were processed to Level-2A BOA reflectance using SNAP 5.0.0 with the plugin Sen2cor 2.3.1 in the Sentinel-2A toolbox and then were clipped to the study area extent. Bands acquired at 20 m resolution (5, 6, 7, 8a, 11, 12) and at 60 m (9) were resampled using the nearest neighbor method to obtain a layer stack of 11 spectral bands at 10 m and to minimize changes to pixel values. Bands 1 and band 10 were excluded since the first one is used for aerosol detection and the second one is omitted from the output after atmospheric correction. The reflectance values were kept scaled by a factor of 10000 for file storage and precision considerations.

3.3.3. Image Classification and Accuracy Assessment

Image classification was performed using RF supervised classification implemented within the R package "randomForest". Regarding the parameters *ntree* (number of trees to be grown in the run) and *Mtry* (number of features used in each split) we used the default values and set *ntree* to 500 trees and *Mtry* to the square root of the total number of input features.

A total of 12 images from 2016 were used to run the classification, one from April, five from May, three from June, and three from July (Table 7). Ten land cover classes were considered: Built up area, rapeseed, spring barley, sugar beet, wheat, fava bean/wheat, grass-legume ley, kernza, kernza/lucerne, and spring barley/lupin. The sample set for training and testing the model was composed of 1097 observations derived from the land cover polygons provided by SITES and the UAV orthomosaic of August. Of this sample set, 75% was used to train the model and 25 % to test.

The classification was performed five times with different explanatory variables each time in an attempt to find a parsimonious relationship between S2A time-series and land cover classes. The first model was constructed using 11 Sentinel-2A spectral bands, i.e. 2, 3, 4, 5, 6, 7, 8, 8A, 9, 11, 12, and NDVI for a single date image of June 11th (12 explanatory variables). The second model included the same 11 spectral bands but from 12 different images from April to July. In addition, NDVI for each image date and an extra covariate with the maximum NDVI of the growing season making a total of 145 explanatory variables.

Table 7. Sentinel-2A imagery used in the study for the image classification, the comparison between sensors with raw and reflectance values, and imagery used for the annual NDVI-time series extraction

Nº	Date	Image classification	UAV image pair	NDVI time-series
1	22/4/2016	X		X
2	2/5/2016	X		X
3	5/5/2016	X		X
4	12/5/2016	X		X
5	22/5/2016	X		X
6	25/5/2016	X		X
7	4/6/2016	X		X
8	11/6/2016	X		X
9	24/6/2016	X		X
10	11/7/2016	X		X
11	21/7/2016	X		X
12	24/7/2016	X		X
13	31/07/2016			X
14	12/9/2016			X
15	8/11/2016		X	X
16	28/11/2016		X	X
17	27/1/2017			
18	11/3/2017		X	

The results of the two first models were evaluated to identify variables that were most relevant to differentiate land cover classes in the study area. Once the variables were identified, the third classification was run using a model that considered only these variables, thus, a total of 30 variables were included in the classification. Furthermore, the third model was evaluated in the same way to build a fourth model with 15 variables. Finally, the fifth model consisted of the 8 most important variables. To identify these variables, we used the Variable Importance function (VI) and the Variable Selection Using Random Forest (VSURF) implemented within the R packages "randomForest" and "VSURF" respectively.

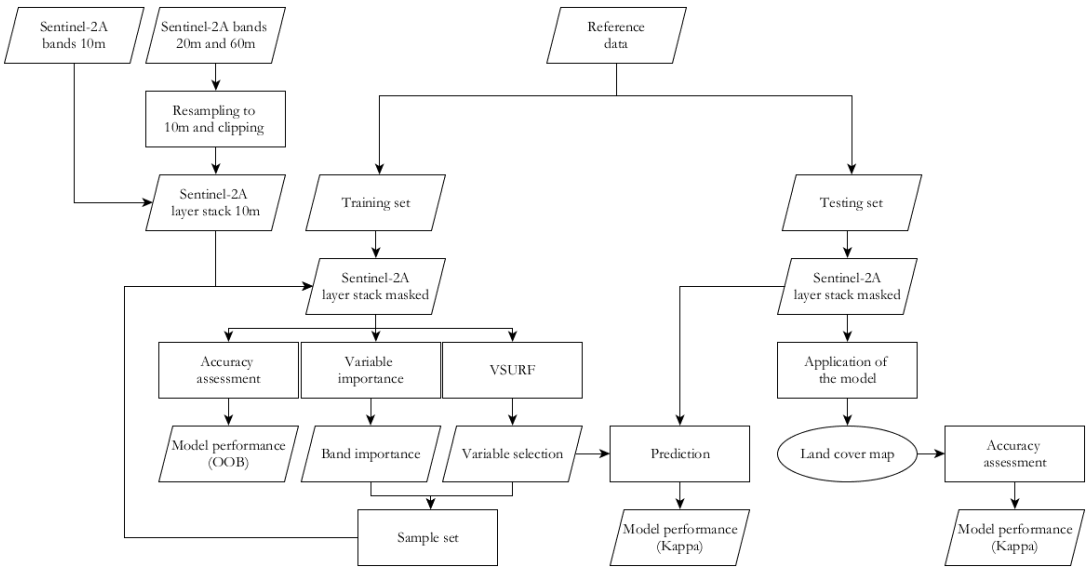


Figure 8. Workflow for image classification and validation using S2A

Variable importance considers the following measures of importance: Mean decrease in accuracy and mean decrease in Gini, which measures how much better than random a predictor variable is in successfully classifying data. Mean decrease in accuracy identifies variables that decrease the accuracy of the Random Forest due to exclusion or permutation. Therefore, variables with large mean decreases in accuracy are more important for the classification of the data. The mean decrease in Gini measures how each variable contributes to the homogeneity of the node in the resulting Random Forest (0 homogeneous and 1 heterogeneous). Every time a split of a node is made on a variable the Gini impurity criterion for the two descendent nodes is less than the parent node A low Gini (i.e. higher decrease in Gini) signifies that a predictor variable plays a greater role in partitioning the data into the defined classes (Breiman & Cutler, 2007).

VSURF is a package in the R software based on the Random Forest algorithm that handles high-dimensional data for variable selection. It returns two subsets of variables: 1) a subset including some redundancy which can be relevant for interpretation and 2) a smaller subset corresponding to a set of variables that are selected to avoid redundancy in the model and focus closely on the prediction objective. The process of variable selection follows three steps. The first step, thresholding, runs *ntree* Random Forest trees (2000 default) and eliminates irrelevant variables based on the variable importance (Gini index) computed during the Random Forest processes, the second step is the interpretation where the objective is to find important variables highly related to the response variable, even if there is redundancy. The third is the prediction, to find a smaller number of variables with very low redundancy and a set that are sufficient for a good enough (essentially, parsimonious) prediction of the response variable (Genuer, Poggi, & Tuleau-Malot, 2015). The initial model is based only on the most important variable, and other variables are added to the model in a stepwise manner. A variable must have a bigger impact on the model accuracy than a noise variable (variables left out by interpretation) to remain in the final model (Michez, Piégay, Lisein, et al., 2016).

The accuracy of the model was assessed with the confusion matrix, the out-of-bag error (OOB), Kappa coefficient and by applying the model to the test set. The workflow for image classification and validation using S2A imagery is shown in Fig. 8.

3.3.4. Spectral bands comparison

In this phase, UAV intensity and reflectance values were extracted to find the correlation with Sentinel-2A reflectance values on a band-by-band basis. An average of the UAV pixel values

covered by a Sentinel-2A pixel (10 x 10 m) was used for the comparison. The average was based on presumably pure pixels, meaning that pixels with only one land cover were selected.

Two band-by-band comparisons were explored. First, the correlation between the intensity values of the orthomosaics generated using Agisoft Photoscan and the atmospherically corrected reflectance values of Sentinel-2A imagery and second, the correlation between these Sentinel-2A images with radiometrically corrected reflectances of the orthomosaics produced by the ATLAS platform. In this analysis, three pairs of images were compared; November 3rd, 2016 with November 8th, 2016, December 2nd, 2016 with November 28th, 2016, and March 13th with March 11th, 2017. Then, the compatibility of the spectral data from the UAV bands and the Sentinel-2A bands was analyzed through statistical methods. A t-test compared two samples means to evaluate if there is a significant difference between the spectral responses recorded by the two sensors. All the available UAV images were paired with the most contemporaneous available Sentinel-2A image (Table 8) as this is a realistic situation in which there is no exact date match due to cloudy periods or no available imagery from the provider.

Table 8. Dates of the UAV and Sentinel-2A imagery used in the research

UAV imagery	**Sentinel-2A imagery**	**±Δt**
November 3rd, 2016	November 8th, 2016	5
December 2nd, 2016	November 28th, 2016	4
March 13th, 2017	March 11th, 2017	2

In addition, statistical analysis was conducted to addresses two questions: (1) Does the radiometric correction performed by ATLAS to the UAV orthomosaics improve the correlation with Sentinel-2A imagery? (2) Are the spectral bands of the MicaSense sensor on board of the UAV compatible with the coinciding Sentinel-2A bands? The answers to these questions provide insights into the compatibility of sensors and the extent to which one can replace the other when observations for a time-series are missing.

The first question requires us to explore the correlation of reflectance values of S2A images and raw image values from UAV orthomosaics on a band-by-band basis. A linear model was fitted to UAV-S2A image pairs of November 3rd with 8th, December 2nd with November 28th (2016) and March 13th with 11th (2017). The coefficients of determination were calculated and compared to assess the strength of the relationship between both sensors.

The second question requires us to compare the mean difference of reflectance values and raw values from UAV orthomosaics with the most coinciding S2A imagery on a band-by-band basis. The null hypothesis (H_0) assumes that the true mean difference between the paired samples is zero. Under this model, all observable differences are explained by random variation. Conversely, the alternative hypothesis (H_A) assumes that the true mean difference between the paired samples is not equal to zero.

H_0: There is no significant difference of reflectances between sensors ($\mu_1 = \mu_2$).

H_A: There is a statistically significant difference of reflectances between sensors ($\mu_1 \neq \mu_2$).

The chosen probability of making a Type I error (rejecting a true H_0) was $\alpha=0.05$. The statistical test selected to compare the spectral bands of the UAV and Sentinel-2A was the paired two samples for mean t-test. Paired Student's t-test is usually used to compare means of two sets of paired samples, taken from two populations with unknown variances (Quinn & Keough, 2002).

The application of a t-test depends upon several assumptions; it is important to know whether these assumptions are satisfied. Three main assumptions should be met to appropriately apply t-test: Both samples should: 1) be random, 2) be independent, meaning that one observation is not affected by the value of another observation, and 3) come from a normally distributed population with unknown but equal variances, such that when the *t* procedures are applied to the differences,

we assume that the differences are normally distributed. This last assumption rules out outliers and skewed distributions.

Spatial data are often non-independent. The effect of ignoring dependence when it is present in the data is to reject null hypotheses more frequently than we should. Over rejection of the null hypothesis is more pronounced as the degree of dependence increases. In case of violating the assumption of homoscedasticity, heterogeneity causes slightly more null hypotheses to be rejected than desired, but the effect is not strong. Generally, outliers may be due to recording errors, to the sample not being entirely from the same population, or to the values being from the same, but no-normal population. A risk of violating the outlier's assumption is that because neither the sample mean nor the sample variance is resistant to outliers; a large outlier can inflate the sample variance, decreasing the *t* statistic and thus lowering the chance of rejecting the null hypothesis.

The *t* procedures are relatively robust to deviations from assumptions of normality of the population and homoscedasticity except in the case of outliers or strong skewness (Quinn & Keough, 2002). A normal quantile plot and a boxplot were used to explore whether the observed distribution of the sample conforms to the normal population distribution. In addition, data was tested for normality using two tests: Shapiro-Wilk and Anderson-Darling (null hypothesis against the assumption of normality) in RStudio.

Because of the risk of the t-test incorrectly rejecting the null hypothesis that the mean of the paired differences is 0 due to skewness in the data, we employed a logarithmic transformation to promote normality and at the same time solve for possible outliers, skewness and asymmetry. Furthermore, we proceeded with the t-test analysis on the transformed data and the inspection of p-values. However, as the transformation and the previous tests were not very reassuring concerning the assumption that the data are normally distributed, we applied the Wilcoxon signed rank test to the data to corroborate the results obtained with the t-test even though the resulting p-values were not close to the significance level. Wilcoxon signed rank test is a distribution-free inference procedure, i.e. a nonparametric procedure, that does not make assumptions regarding how the underlying data are distributed but still assumes symmetry.

Moreover, we compared crop reflectances from both sensors to *in situ* measurements collected using a hyperspectral radiometer for one single day (March 13th).

3.3.5. Vegetation indices

Vegetation indices were calculated for UAV and Sentinel-2A image pairs (Table 8) to evaluate the compatibility of the sensors. The multispectral vegetation indices calculated in this study were Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI) and Green Chromatic Coordinate (GCC). The comparison was based on the methodology described in Section 3.3.4.

The statistical analysis conducted to compare vegetation indices calculated from imagery from both sensors addressed the following question: Can vegetation indices produced from UAV orthomosaics be used for upscaling data to Sentinel-2A imagery scale? Three vegetation indices (NDVI, EVI, and GCC) were calculated from orthomosaics and Sentinel-2A imagery. Again, both the t-test: paired two samples for mean and Wilcoxon signed rank statistical test were applied in RStudio to determine whether the vegetation index mean is significantly different between sensors.

H_0: The difference in vegetation index mean between sensors is zero ($\mu_1 = \mu_2$).	H_A: The difference in vegetation index mean between sensors is not zero ($\mu_1 \neq \mu_2$).

In addition, to exploring upscaling opportunities, NDVI time-series data from Sentinel-2A from 2016 were used to assess how well the satellite captured the temporal variation of the crops and to evaluate the potential of UAV observations in replacing noisy or missing S2A observation. TIMESAT (Jönsson et al. 2017, unpublished) was used to fit a double logistic function to define better the curve of crop development. Orthomosaics generated from acquired imagery in

November and December were used to extract NDVI values of different crops and compare with NDVI predictions constructed based on S2A imagery.

4. Results

4.1. Image classification and accuracy assessment

The first Random Forest classification based on a Sentinel-2A image from June 11th, 2016 gave an accuracy of 91.6% with an OOB error of 8.38 (Fig. 9). The performance of the classifier using the test set to predict the land cover map gave an accuracy of 92.3% with a Kappa coefficient of 0.91. The misclassification rate for predicting the land cover class of the training set can be seen in the confusion matrix obtained by the model (see Appendix 1). The land cover classes with higher errors of omission were fava bean/wheat and grass-legume ley accounting for 26% of the OOB error. Fava bean/wheat was the class that scored highest on both error of omission (57%) and error of commission (50%) in the validation of the classifier. On the other hand, the classes where the model was more accurate were built up area, spring barley, rapeseed, sugar beet and wheat (see Appendix 1).

Figure 9. Land cover classification map single date image of June 11th, 2016 with 12 explanatory variables

According to the Variable Importance function of the Random Forest model (Fig. 10 a), the variables that had the highest impact on accuracy are bands 11 (1610 nm), 9 (945 nm), 2 (490 nm), 6 (740 nm) and 3 (560 nm). Those that had less impact were NDVI, band 7 and band 4 probably due to the similar spectral response that crops exhibit on these bands as shown in Fig. 11. These findings are also confirmed by Immitzer et al. (2016), who found that Sentinel-2A blue and red bands were more important than red and NIR for crop classification in Austria.

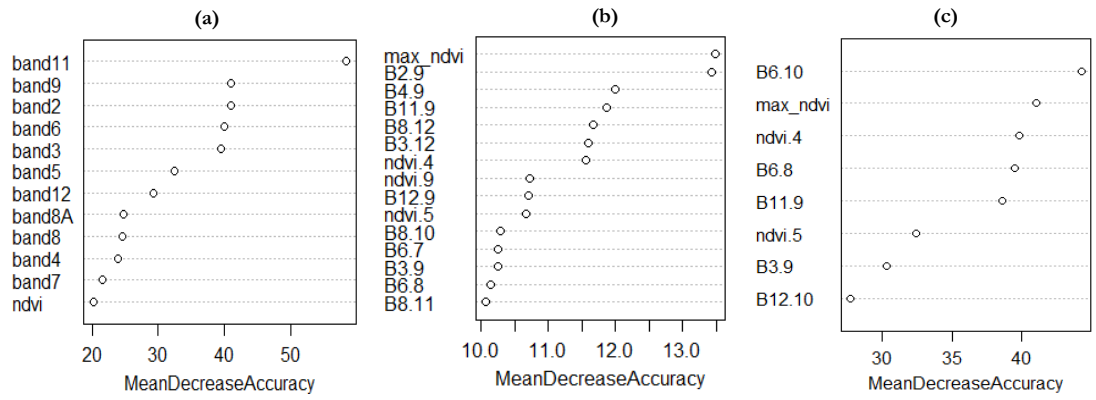


Figure 10. Variable importance plot using a) single date imagery with all 12 explanatory variables for the classification, b) multi-temporal imagery with 30 explanatory variables for the classification and showing only the 15 most important variables after the VSURF prediction step, c) multi-temporal imagery with 8 explanatory variables for the classification and showing all 8 variables in order of importance from top to bottom.

The second classification based on the 12 Sentinel-2A images, NDVI layers for each date and maximum NDVI of the season resulted in the highest accuracy 97% and the lowest OOB error of 3.1%. However, it included 145 explanatory variables. Grass-legume ley was the class with the lowest producer accuracy (85.7%). The model predicted successfully built up areas with 100% user accuracy. The 145 variables were evaluated using VSURF of which 87 were found to be relevant and were used in the interpretation step. After eliminating redundancy in the set, the prediction step selected 19 variables that were combined with those that were important to the result according to the Variable Importance function making a total of 30 explanatory variables that were feed to the next classifier. The first three dates of the dataset (April 22th, May 2nd and May 5th) were not selected. This might be due to the fact that growth of crops resumes after dormancy and starts to rise in mid-May according to the crop calendar in Table 6.

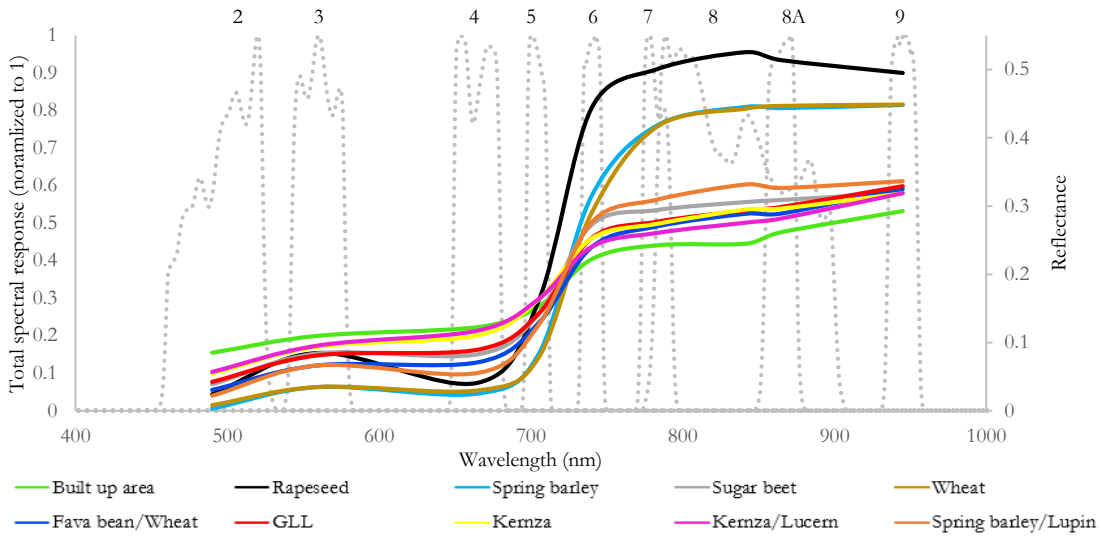


Figure 11. Spectral signature of crops in Lönnstorp based on 9 bands of Sentinel-2A imagery of June 11th, 2016

An interesting finding was that red-edge band 5 (705 nm) was also absent from the important variables. Figure 11 shows that many spectral signatures of crops are overlapping at 705 nm, this could be the reason why band 5 does not seems to be of value for differentiating these crops. On the other hand, 63% of the variable set was comprised of bands 3, 6, 8, 12 and NDVI layers, this is possible since at these wavelengths the spectral response of crops varies (see Appendix 2 for bands 11 and 12). The high frequency of NDVI in the set of important variables can be explained due to the different response of crops to the Red and NIR at different phenological stage.

The third classification used 30 variables as inputs to predict the land cover classes of the study area. Despite reducing the number of explanatory variables more than 80%, the accuracy of the classification did not decrease (96.2%). The model predicted with 100% accuracy the classes built up areas and kernza and had the same poor performance as the previous model in predicting fava bean/wheat, with an error of commission of 21%.

The result of the VSURF was 24 variables in the interpretation step and 15 in the prediction step (Fig. 10 b). More than 66% of coinciding important variables were found between VI and VSURF. This time only the 15 variables predicted by VSURF were used as input for the fourth classification. Apart from band 5, the second VSURF eliminated bands 7, 8A and 9. The images of June and July resulted in the most frequently selected among the variables.

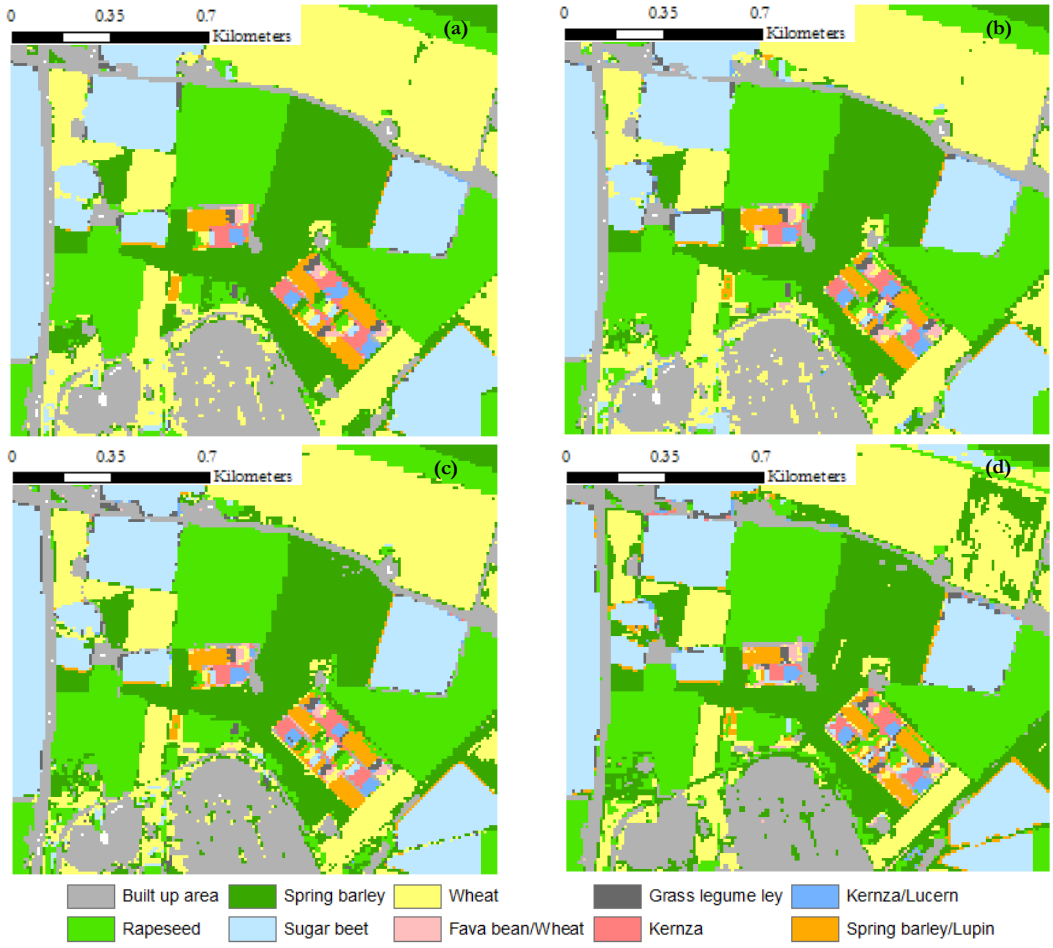


Figure 12. Land cover classification maps. a) Multi-temporal imagery from April until June with 145 explanatory variables. b) Multi-temporal imagery with 30 explanatory variables. c) Multi-temporal imagery with 15 explanatory variables. d) Multi-temporal imagery with 8 explanatory variables.

In the fourth classification, the number of explanatory variables was reduced by 90%. The 15 remaining variables from the algorithm for variable selection predicted 10 land cover classes with the same accuracy (96.2%). The model performed poorly again for the classes fava bean/wheat and grass-legume ley, but very well with built up areas. However, with the testing set, the model predicted five classes with 100% of accuracy: built up area, spring barley, fava bean/wheat, kernza, and kernza/lucerne. The overall classification accuracy was 97.8% and the Kappa coefficient was 0.9746.

In addition, the test sample set was used with the predict method of VSURF, which permits to predict the outcomes for new data with RF using only the variables selected by VSURF. The

accuracy using 8 variables was 94.8% compared to the accuracy of 96.2% using 145 (see Appendix 1). The 8 important variables selected were: two single date image NDVI from May 12th and 22nd; band 6 from June 11th; band 3 and 11 from June 24th; band 6 and 12 from July 11th; and maximum NDVI of the growing season. It can be noted that two of the important bands for vegetation, Red (band 2) and NIR (band 8), are absent from the last set of variables, however, their importance might be accounted by the presence of three NDVI variables. With respect to the timing, it seems that images from June and July provide enough information for a good distinction of crops in the study area, and even with one single date image, the accuracy of the classification it is still acceptable.

The last classification attempt was using these 8 variables (Fig. 10 c). The overall accuracy was 95% with the training set and 93.3% with the testing set and a Kappa coefficient of 0.92. Figure 12 shows the four land cover maps using the validation set as input for the model. The previous results show that VSURF was a good method for finding a smaller number of important variables highly related to the response variable (i.e. crop class) and eliminating redundancy from the set. In the case where the study area would be bigger, a smaller set of variables sufficient to achieve a good parsimonious prediction of the response variable could decrease the computational load and the time required to process the data.

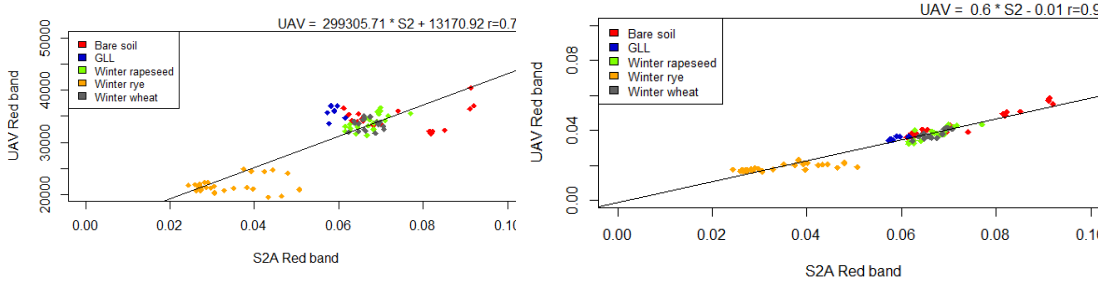
4.2. Spectral bands comparison

The coefficients of determination obtained from comparing radiometrically corrected orthomosaics and Sentinel-2A imagery were higher than those correlations obtained with raw orthomosaics. Coefficients of determination for each band of the three most coinciding dates are given in Table 9, and in Fig. 13 the plots for the last pair of images for bands red, NIR and red-edge are shown. According to the results, the radiometric correction performed by ATLAS in the first and last orthomosaics improved the correlation more than for the orthomosaic of December 2nd.

There is a visible difference in the quality of the orthomosaic of December (see Appendix 3) compared with the other orthomosaics which can perhaps explain the lower coefficients obtained for this pair. Nevertheless, the coefficients of determination of the other two pairs were higher than 0.60 with *p-value*<0.05 which means that there is a relationship between the two sensors that it is unlikely that it is due to chance, however, the relationship is not very strong. A special situation occurs in the red-edge band since it is the least coinciding between the two sensors and this is reflected in the coefficient of determination as it is the lowest of all bands.

Table 9. Coefficients of determination based on a band-by-band comparison

	November 3rd with 8th, 2016		**December 2nd with November 28th, 2016**		**March 13th with 11th, 2017**	
	r^2 (raw)	r^2 (atlas)	r^2 (raw)	r^2 (atlas)	r^2 (raw)	r^2 (atlas)
Blue	0.05	0.63	0.33	0.24	0.2	0.8
Green	0.02	0.68	0.48	0.53	$5.6x10^{-3}$	0.69
Red	0.57	0.71	0.17	0.13	0.79	0.94
RedEdge	0.11	0.58	0.37	0.39	0.28	0.42
NIR	0.44	0.74	0.46	0.61	0.83	0.96



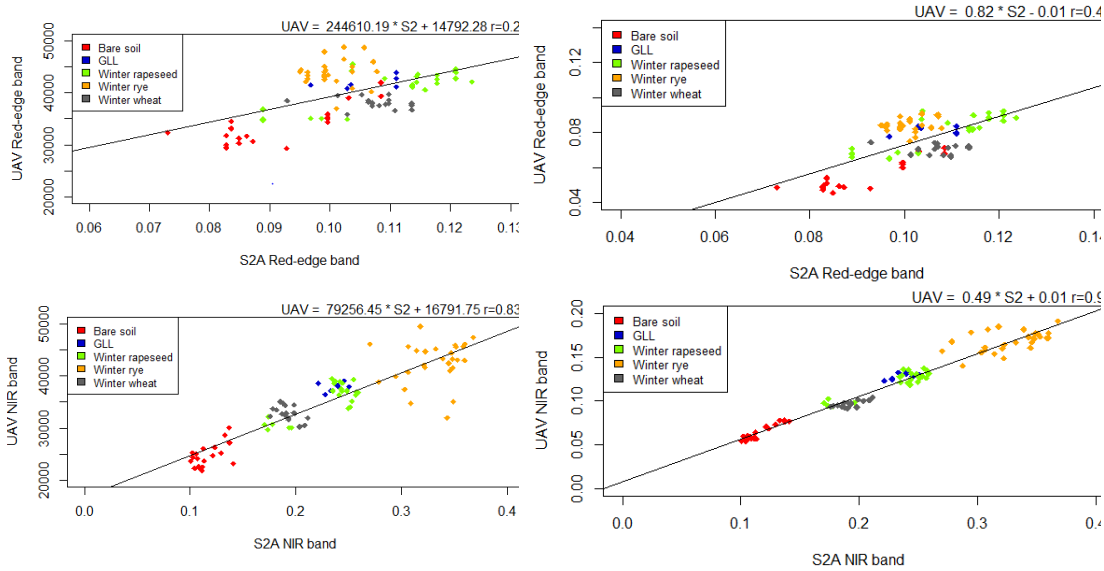


Figure 13. Band-by-band comparison between UAV and S2A. Raw UAV orthomosaic (left) of March 13th and radiometrically corrected (right) with S2A image of March 11th, 2017.

The results suggest that while more than 60% of the variation in the data can be explained by the linear model, the remaining variation that cannot be accounted by the model might be due to the difference in acquisition dates, differences in the radiometric correction and differences in the center wavelength and bandwidth of the sensors (Table 4).

Further analyses of the data showed that no anomalous outliers were present, but also showed signs of skewness and asymmetry (Fig. 14). However, the paired difference was more symmetric, meaning that the sample sets had similar skewness. The logarithmic transformation employed solved for skewness and promoted normality in most of the band-by-band pairs. Since the skewness was not severe, and the sample size was large, which tends to improve the accuracy of p-values and critical values from the *t* distribution, we considered that the t-test performed adequately.

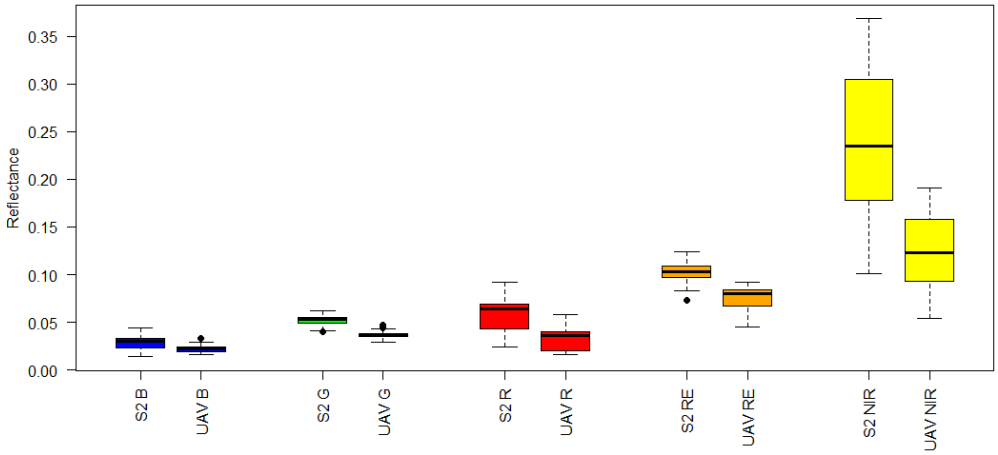


Figure 14. Boxplot of reflectances per band for both sensors. UAV image of March 13th and S2A image of March 11th.

Regarding the assumption of independence, if we would have incorrectly assumed independence within the samples and the t-critical values should have to be larger, the t-statistic obtained is much larger and at the same time the *p-value* obtained is very small, hence, we would have still detected a significant departure from 0 of the mean of the paired difference.

Table 10. t-test and Wilcoxon signed rank test of reflectance data for the first image pair. UAV orthomosaic of November 3rd and Sentinel-2A image of November 8th, 2016

	November 3rd with 8th, 2016				
	B	G	R	RedEdge	NIR
Observations	192				
df	191				
Pearson correlation (r)	0.79	0.82	0.84	0.76	0.86
Wilcoxon signed rank	*p<0.001*	*p<0.001*	*p<0.001*	*p<0.001*	*p<0.001*
t-stat	-6.45	22.52	-12.53	19.21	-4.46
P(T≤t) two-tail	*p<0.001*	*p<0.001*	*p<0.001*	*p<0.001*	*p<0.001*
t-critical two-tail	1.97				

Table 11. t-test and Wilcoxon signed rank test of reflectance data for the second image pair. UAV orthomosaic of December 2nd and Sentinel-2A image of November 28th, 2016

	December 2nd with November 28th, 2016				
	B	G	R	RedEdge	NIR
Observations	74				
df	73				
Pearson correlation (r)	0.49	0.73	0.35	0.63	0.78
Wilcoxon signed rank	0.25*	*p<0.001*	*p<0.001*	*p<0.001*	*p<0.001*
t-Stat	1.19	10.52	19.19	13.17	10.47
P(T≤t) two-tail	0.24*	*p<0.001*	*p<0.001*	*p<0.001*	*p<0.001*
t-critical two-tail	1.99				

The t-scores from the paired two samples for mean t-test showed that there is a statistically significant difference between mean reflectances of UAV and mean reflectances of Sentinel-2A based on a band-by-band comparison. The test statistic obtained for all image pairs is beyond t-critical. Based on this data it appears that there is a significant difference between both sensors (*p-value<0.001*). With 95% confidence interval, we reject the null hypothesis since there is evidence that the mean of the reflectance of UAV and S2A is different than zero. The results of the Wilcoxon signed rank test agreed with the results obtained with the t-test as shown in Tables 10-12.

The only non-significant difference from the whole analysis was found in band blue on the comparison between the orthomosaic of December 2nd and the S2A imagery of November 28th that on the previous band-by-band analysis of reflectance showed the lowest correlation.

Table 12. t-test and Wilcoxon signed rank test of reflectance data for the third image pair. UAV orthomosaic of March 13th and Sentinel-2A image of March 11th, 2017

	March 13th with 11th, 2017				
	B	G	R	RedEdge	NIR
Observations	95				
df	94				
Pearson correlation (r)	0.89	0.83	0.97	0.65	0.98
Wilcoxon signed rank	*p<0.001*	*p<0.001*	*p<0.001*	*p<0.001*	*p<0.001*
t-stat	-15.50	-60.35	-31.69	-27.02	-25.85
P(T≤t) two-tail	*p<0.001*	*p<0.001*	*p<0.001*	*p<0.001*	*p<0.001*
t-critical two-tail	1.98				

The differences between *in situ* and measurements are pronounced at the green and NIR bands in both winter rye and winter rapeseed crops. Contrarily, reflectances are more consistent in bands blue, red and red-edge (Fig. 15). It should be noted that in general, data from S2A imagery on this date had higher values than the UAV data as shown in the boxplot. However, for the other dates UAV derived reflectances were higher than those from S2A, this explains why the t-stat results are positive for most of the bands in November and December.

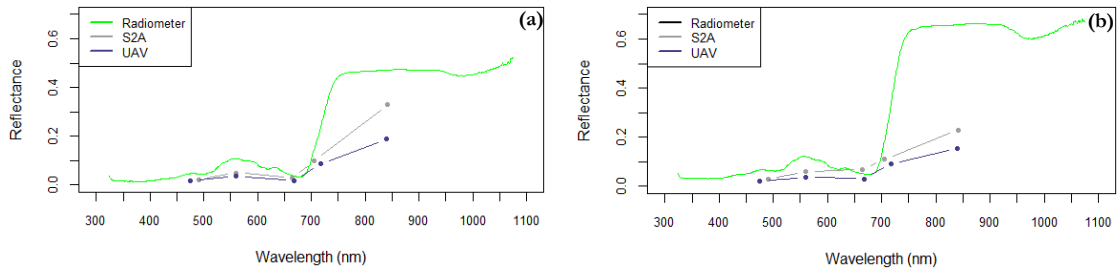


Figure 15. Spectral signature of crops from *in situ* and remotely derived measurements. a) Winter rye b) Winter rapeseed

4.3. Vegetation indices

Unlike the comparison band-by-band with reflectance data, the coefficients of determination of vegetation indices calculated from comparison of the two sources of imagery are stronger. There is no consistent pattern between indices on different observations dates. Figure 16 shows that the correlation varies from one image pair to the other. Based on this data, EVI and NDVI correlate the best, closely followed by GCC, but with lower coefficients in all the analyzed pairs. GCC and NDVI mean values are similar between sensors, whereas with EVI it seems that there is a disagreement. EVI and GCC values are higher when calculated from UAV imagery and NDVI appears to capture a larger variability using data from both sensors (Fig. 17).

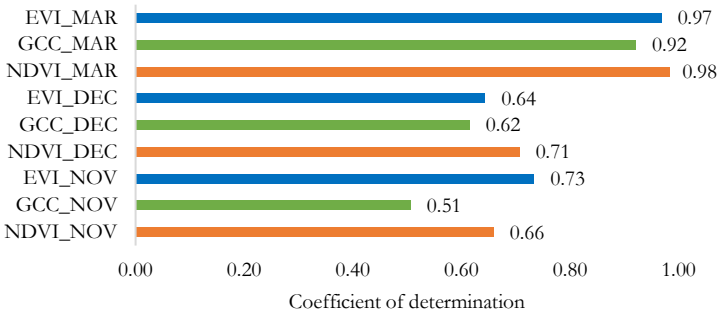


Figure 16. Coefficient of determination of vegetation indices

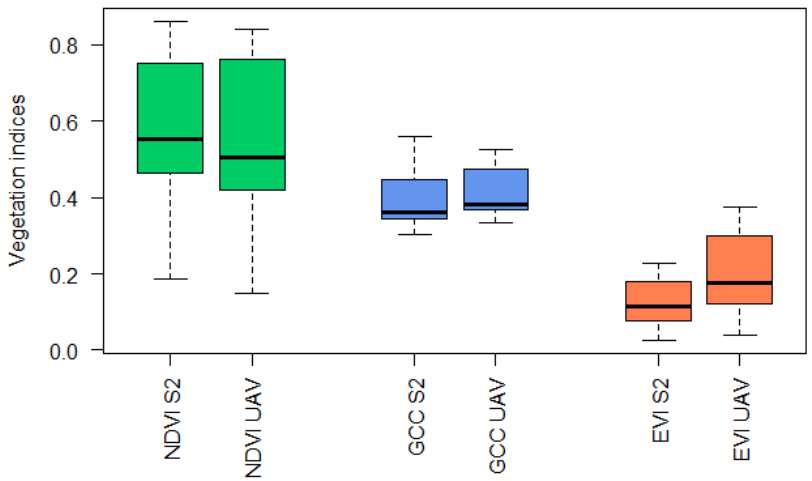


Figure 17. Boxplot of vegetation indices per sensor. UAV image March 13th with S2A March 11th.

Although the correlation seems stronger, the paired t-test shows that there are still significant differences between sensors. The statistical test conducted with vegetation indices produces results that are not consistent in the three days of observation. The first image pair compared (see image

pairs in Table 8) shows that there is no statistical evidence that the mean difference between NDVI calculated from UAV and S2A imagery is significantly different from zero (α=0.05). However, the difference in means of GCC and EVI is statistically significant (*p-value*<0.001 and t-statistic>t-critical), therefore we reject the null hypothesis in favor of the alternative (Table 13).

Table 13. t-test and Wilcoxon signed rank test of vegetation indices for the first image pair. UAV orthomosaic of November 3rd and Sentinel-2A image of November 8th, 2016

	November 3rd with 8th, 2016		
	NDVI	GCC	EVI
Observations		192	
df		191	
Pearson correlation (r)	0.81	0.71	0.86
Wilcoxon signed rank	0.72*	*p<0.001*	*p<0.001*
t-stat	-0.66*	31.47	33.27
P(T≤t) two-tail	0.51	*p<0.001*	*p<0.001*
t-critical two-tail		1.97	

If we analyze where most of these differences occur, we can find that when subtracting the S2A imagery from the UAV orthomosaic, S2A gives lower values for bare soil areas shown in brown and conversely the UAV yields higher NDVI values for green areas like spring wheat and sugar beet (Fig. 18). Areas with sparsely cover, like the case of crops in the green-up stage like winter rye and grass-legume ley, have almost the same NDVI values.

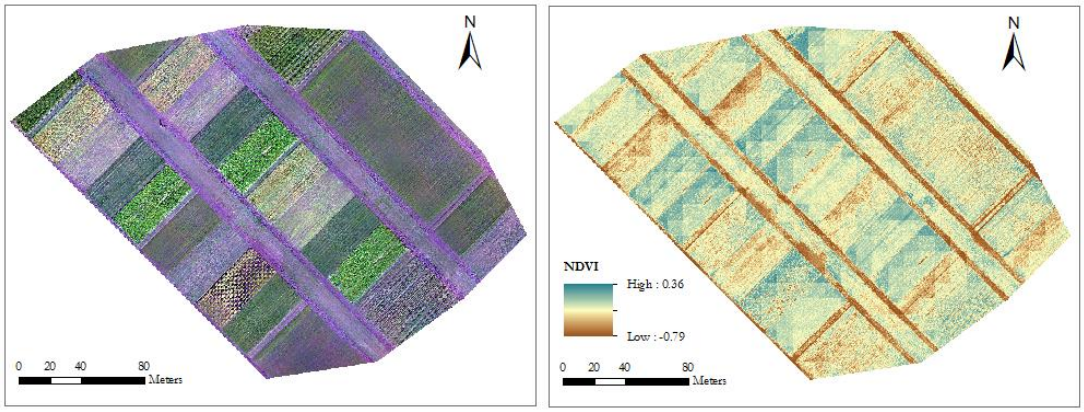


Figure 18. UAV orthomosaic of November 3rd and NDVI difference between UAV imagery and S2A imagery of November 8th, 2016

The analysis of the second image pair reveals that there are statistically significant differences only with EVI and not with NDVI and GCC (Table 14). This is interesting given that the previous analysis showed that the orthomosaic of December 2nd had the lowest correlation on a band-by-band analysis, but it was also the only image pair where one band (blue) had no significant difference between sensors.

Table 14. t-test and Wilcoxon signed rank test of vegetation indices for the second image pair. UAV orthomosaic of December 2nd and Sentinel-2A image of November 28th, 2016

	December 2nd with November 28th, 2016		
	NDVI	GCC	EVI
Observations		74	
df		73	
Pearson correlation (r)	0.84	0.78	0.80
Wilcoxon signed rank	0.85*	0.33*	*p<0.001*
t-stat	-0.29*	1.49*	13.73
P(T≤t) two-tail	0.77	0.14	*p<0.001*
t-critical two-tail		1.99	

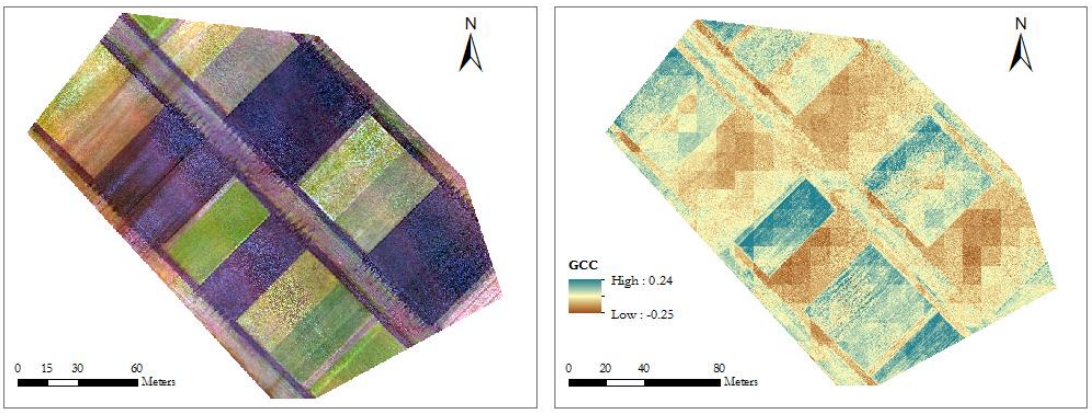


Figure 19. UAV orthomosaic of December 2nd and GCC difference between UAV imagery and S2A imagery of November 28th, 2016

Similarly, the image difference of GCC for the next pair (Fig. 19) shows that values of the vegetation index using the UAV imagery give approximately 20% higher values in crop areas while areas with bare soil have less error.

Moreover, the statistical test performed on the third image pair resulted in significant differences in the three indices (Table 15). Results suggest that whether there is a relationship between sensors, it is difficult to relate reflectance values between sensors.

Table 15. t-test and Wilcoxon signed rank test of vegetation indices for the third image pair. UAV orthomosaic of March 13th and Sentinel-2A image of March 11th, 2017

	March 13th with 11th, 2017		
	NDVI	GCC	EVI
Observations		95	
df		94	
Pearson correlation (r)	0.99	0.96	0.98
Wilcoxon signed rank	*p<0.001*	*p<0.001*	*p<0.001*
t-stat	-11.44	7.98	15.42
P(T≤t) two-tail	*p<0.001*	*p<0.001*	*p<0.001*
t-critical two-tail		1.98	

The results of the image differencing of EVI agrees with the previous findings. The green areas shown in the orthomosaic in Fig. 20 correspond to winter rye and winter rapeseed that give higher EVI values when using the UAV imagery than when using S2A.

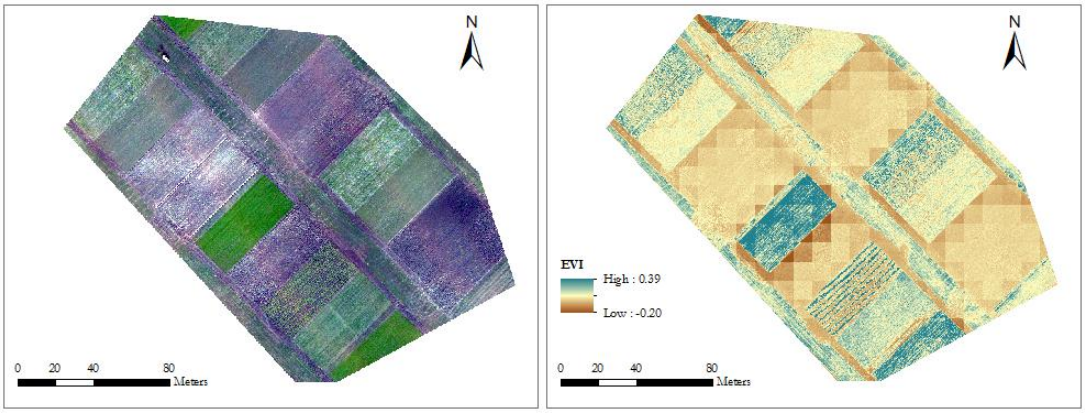


Figure 20. UAV orthomosaic of March 13th and EVI difference between UAV imagery and S2A imagery of March 11th, 2017

Senitnel-2A NDVI time-series allows us to observe the temporal development and phenology of the crops in the study area. In Fig. 21 we can see that time-series of crop pixels allow us to roughly estimate the start, end, and length of the season, as well as the estimations of the maximum NDVI and the time of the middle of the season (Jönsson & Eklundh, 2004). Sentinel-2A observations of the growing season in 2016 seem to represent well the temporal variation of the crops in the study area. However, as expected, some of the observations contain noise that affects the estimation of the seasonality parameters and no obvious peak of photosynthetic activity can be recognized in crops like Fava bean/Wheat apart from finding sudden drops of NDVI.

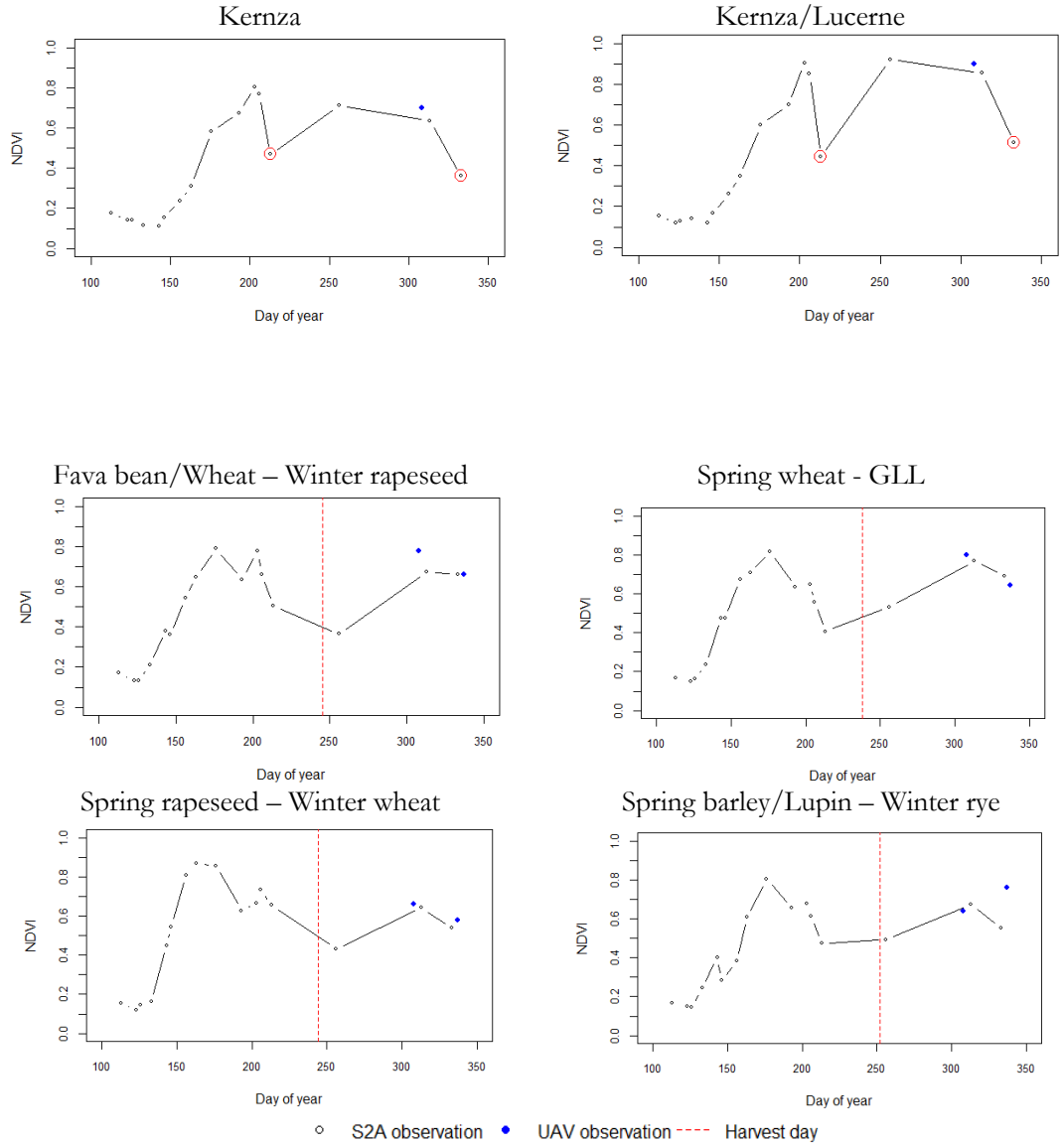


Figure 21. Time-series of S2A NDVI data for crop pixels

Many studies use crop phenology metrics derived from vegetation indices to estimate yield (Mkhabela, Mkhabela, & Mashinini, 2005; M. Wang et al., 2014). Yield is most influenced by crop conditions during the heading or peak phase of the growth (Wall, Larocque, & Léger, 2008). Identifying key stages of crop phenology is important for the development of satellite-based crop yield estimation models. Even though there are significant differences in the actual units of reflectances and vegetation indices, UAV observations appear to have the potential to replace noisy

pixels and/or complement observations to represent better the phenology of crops and thus contribute to an improvement in the accuracy of yield predictions (Fig. 21).

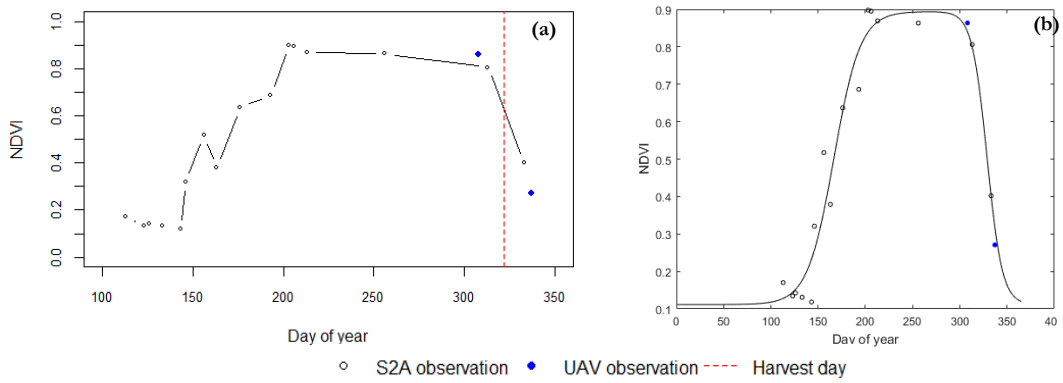


Figure 22. Time-series of S2A NDVI data for sugar beet. a) Sugar beet pixel extraction, b) Smoothed curve after fitting the function.

The two dips that appear on the plots of kernza and kernza/lucerne were due to events of clipping that occurred in July (DOY=207) and November (DOY=330). The blue circles represent the UAV observations of November and December, which correspond to the end of the season and beginning of winter crops after the harvest (red dashed line). The harvest time of the crops differs very little, however earlier in the phenological cycle, i.e. rising and blooming, crops can be distinguishable due to the spectral differences when they are developing at different rates.

In Fig. 22 a) we show the result of the extraction of NDVI time-series of one sugar beet pixel and in b) the same extraction after fitting a double logistic function. The algorithm of TIMESAT seems to reduce the noisy signal, improve the data quality and transform the data into a smoothed seasonal curve that highlights phenology parameters. Moreover, the UAV observations complement the time-series and coincided with the estimation of the temporal variation despite the significant differences found in the previous analyses.

5. Discussion

5.1. Evaluation of multi-temporal S2A data for crop identification

We found high accuracies in the outputs of the digital image classification using a supervised Random Forest approach. Based on the results, a single date Sentinel-2A image from mid-June has the potential to distinguish the different crops in the study area through this classification approach with a Kappa coefficient of 0.91. The spectral signatures of crops derived from the same image corroborate this finding, as crops exhibit different responses in mid-June, especially in SWIR bands (Appendix 2). According to the variable importance plot, differences in NDVI between crops did not seem to impact the accuracy of the classification more than red and NIR bands alone did, this probably due to the vegetation index not adding additional information. The single date image from June 11th corresponds to the day of the year 163, about two months ahead of harvest, meaning that it is possible to have a clear distinction between crops early in the season. It is important to mention that this result is supported by a strong sampling set, comprised of 808 pixels for training and 269 more for testing, that allowed identification of intercropping plots like spring barley/lupin, fava bean/wheat, and kernza/lucerne, although with slightly more error than the other classes.

Including multi-temporal data increased the accuracy of the classification by 4%, but considered a very large number of covariates. However, the variable importance function of the randomForest package and VSURF algorithm were successful in reducing the number to 8 variables with the highest impact in the classification while maintaining an overall accuracy above 90%. This poses an opportunity to reduce the computational burden when increasing the size of the study area. Moreover, the multi-temporal evaluation shows that June and July are the most optimal months

for differentiating crops in the study area. The results are already comparable to the outcome of the study presented in Immitzer et al. (2016) where they investigated the classification performance using a single image to classify crops in Austria and achieved accuracies around 77-83%.

Amongst the most important spectral bands, we found one in the red-edge (B6), two in the SWIR (B11 and B12) and one in the visible spectrum (B3). Immitzer et al. (2016) observed the same behavior in their study regarding the importance of SWIR and red-edge bands of S2A for crop classification.

We consider that the proportion of background material present within the field of view of the sensors allowed the discrimination between vegetation types since the background material is closely related to the stage of the plant in its phenological cycle. In this case, imagery from early in the growing season when crops are developing at different rates yields different percent canopy closure, and that could create differences in their spectral signatures.

Song et al. (2017) used maximum NDVI to separate different crops as well as to separate these from non-crop land covers. In this study, maximum NDVI of the growing season turned out to be very important as well for the accuracy of the classification. Since neither band red nor NIR were included among the 8 important variables, we can suggest that their information was retained in the NDVI from single date imagery instead. In general, the results suggest that the classifier performed well, considering the size of the experimental plots and that the harvest date of crops was variable since July 11th.

5.2. Band differences

MicaSense RedEdge has very well corresponding central wavelengths to Sentinel-2A, although their bandwidths are narrower than those of S2A. There was a clear improvement in the correlation of Sentinel-2A and MicaSense RedEdge spectral bands after the radiometric calibration performed by ATLAS, except with data from December 2nd. This orthomosaic was greatly affected by blurriness caused by wind and varying sun illumination conditions due to the time of the flight, this could explain the difference in the results. In general, red and NIR bands showed stronger correlation. Tewes et al. (2015) suggested that achieving better results with these bands may be because for vegetated land surfaces the reflected radiation in the NIR is higher than in visible wavelengths and therefore less affected by radiation variations. Conversely, the least correlated band was the red-edge band. Discrepancies between sensors regarding the center wavelength and bandwidth of red-edge may have caused a decrease in the correlation. The large variability found in the data from the NIR band of both sensors could be due to the broader bandwidth, 40 nm for MicaSense and 115 nm for S2A, compared to the other channels.

Despite the strong correlation of bands, a significant difference in reflectance, between MicaSense RedEdge and Sentinel-2A, was found with the statistical testing. This means that one should be aware of these differences when combining data from both sensors since these differences could affect the performance of models based on biophysical variables and they could lead to biases in the estimation of the timing of phenologic events. Certain seasonal transitions could be effectively observed at nearly the same time when observed from both platforms, while others may not.

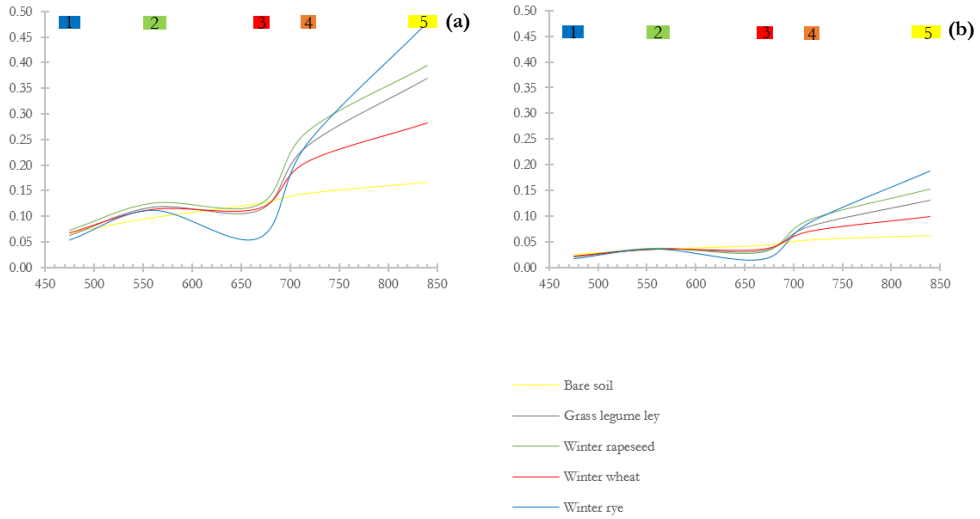


Figure 23. Spectral signature of crops from UAV imagery. a) March 7th, 2017, b) March 13th, 2017.

Besides differences in spatial resolution, the poor agreement can be explained by factors such as observation quality, varying atmospheric conditions, and viewing geometry (Vrieling et al., 2017). Probably the most important source of dissimilarities in reflectance between the sensors is due to the observation dates not coinciding. The most coincident pair had a difference of 2 days between observations, followed by 4 and 5 days for the remaining image pairs. The sensitivity of the sensors may be another important source of discrepancy. However, an interesting finding was that the spectral response of crops between two consecutive UAV observations differs greatly, even when flying with the same configurations and at the same time of the day (Fig. 23). This means that the differences are not only between Sentinel-2A and UAV data but are also between different observations with the same sensor.

Compared to field reflectances, UAV and S2A values were generally lower; this could be explained by how the field measurements were taken. The radiometer was pointing directly to the vegetation, thus reducing the influence of the soil background to the spectra.

The influence of different land covers contributing to the signal of a S2A pixel cannot be ruled out, even if we tried to include only pure S2A pixels in the analyses. The measurements would most likely contain a variety of surface materials due to the growing stage of the crops, i.e. winter wheat and winter rye, as at the time this study was conducted the canopy did not cover the ground completely, thus the measurement of the sensor consist of mixed signals from multiple reflectance signatures. Moreover, it can be possible that a simple spatial average to degrade high-resolution UAV observations to a S2A footprint might not be representing well the signal captured by the satellite.

It should be noted that we do not know what is behind the radiometric calibration performed by ATLAS and in which way this could be improved. In this sense, there is a potential to increase the compatibility between UAVs and satellite sensors by improving the radiometric calibration according to local conditions. At the end of this study, three MosaicMill reflectance panels were available but the attempt of calibrating the images was not successful given that the 44% and 23% panels appeared saturated in the images with the current configuration of the UAV sensor.

5.3. Vegetation indices

The comparison of vegetation indices between sensors showed stronger correlation. NDVI and GCC showed to normalize the data better compared to EVI as shown in the boxplot in Fig. 17. Similar results were found with the other two image pairs. Although, as was the case with the reflectances, there were absolute differences between vegetation indices derived from the UAV and S2A, in this case, less statistically significant differences were found.

In general, UAV vegetation indices values were higher than those from S2A; this was found with two out of three image pairs (except March) where the UAV measurements were higher in most of the bands. This could be explained since there is less interaction of the reflected radiation with the atmosphere before it reaches the sensor.

From the image differencing approach, we can see that the S2A values were higher in bare soil and lower in crop areas. NDVI derived from S2A imagery was likely more influenced by the soil background as in the image differencing bare soil areas showed very low values (-0.79) compared to the differences found with GCC and EVI. The difference between sensors is smaller when crops are starting to develop.

The NDVI time-series using cloud-free and atmospherically corrected images with Sen2Cor highlighted the seasonal development of the crops in the study area. Most of the NDVI observations derived from the UAV showed the potential to complement the time-series. Nevertheless, as expected, we found noisy data in both S2A and UAV images, probably arising from varying atmospheric conditions, radiometric calibration, blurriness effects (in the case of UAV imagery) and others. Improved radiometric calibration (as discussed above) may make spectral reflectances and derived products, such as vegetation indices, more comparable. The fitted function could reduce the impact of these issues and smooth the data to produce and improve the quality of data for a better estimation of phenological parameters.

For this study, data availability restricted us from evaluating the agreement of S2A and UAV observations at the beginning and peak of the season. However, scaling between UAV derived vegetation indices and S2A showed strong agreement, UAV observations may thus provide more confidence in the accuracy of transition dates at key development stages. Nevertheless, when combining UAV imagery with other sensors, data should be validated before making a detailed analysis, especially in landscapes with fine-scale land cover variability as this could lead us to conclude that UAVs and satellite could indicate different phenology (Vrieling et al., 2017).

6. Conclusions and Recommendations

In the context of crop monitoring for Precision Agriculture practices, the aim of this study was to explore a multi-temporal Sentinel-2A dataset for crop type mapping at the plot level and to assess the performance of the methodology used. Another major ambition was to examine the compatibility of S2A data with that derived from the MicaSense RedEdge sensor, operated on-board a UAV.

This study demonstrated that cloud-free S2A atmospherically corrected images applied to a pixel-based supervised Random Forest classification yielded satisfactory results for crop identification in the study area. The accuracy achieved with a single-date S2A image from mid-June (two months before harvest) was comparable to the accuracy of a multi-temporal stack of imagery from the whole growing season.

The radiometric calibration of the orthomosaics performed by ATLAS improved the correlation of spectral bands between UAV and S2A data in terms of reflectance information. Moreover, we found a stronger agreement of vegetation indices (NDVI, EVI, GCC) between both sensors using UAV calibrated images.

S2A annual time-series data effectively represented the characteristic phenological development of the different crops in the study area and displayed agreement with the crop calendar. Although, in cloud-prone landscapes like in Sweden, frequent good-quality observations for phenology assessments are rare and difficult to obtain. UAV data showed the potential to complement satellite data, as well as to be useful for validating phenological estimates derived from satellite data, to provide more confidence in the accurate determination of transition dates for key development stages, and to improve classification of crops with this information.

Despite the many advantages of UAVs, among these that they are able to capture the small-scale spatial variability of crops, their usefulness is hindered by their relatively low area coverage per flight. For this reason, UAV data should be validated and used as a representative high-resolution input for various models, because while S2A has a lower spatial resolution, it represents an important tool for global monitoring due to its frequent and extensive coverage.

Finally, we consider that MicaSense RedEdge offers an interesting alternative for small-scale farmers to take advantage of multispectral sensors to monitor their crops, while at the same time promoting Precision Agriculture practices.

7. References

Agisoft LLC. (2016). Agisoft PhotoScan User Manual. Retrieved from http://www.agisoft.com/pdf/photoscan-pro_1_2_en.pdf

Anderson, K., & Gaston, K. J. (2013, April). Lightweight unmanned aerial vehicles will revolutionize spatial ecology. *Frontiers in Ecology and the Environment*. Ecological Society of America. http://doi.org/10.1890/120150

Balaghi, R., Tychon, B., Eerens, H., & Jlibene, M. (2008). Empirical regression models using NDVI, rainfall and temperature data for the early prediction of wheat grain yields in Morocco. *International Journal of Applied Earth Observation and Geoinformation*, *10*(4), 438–452. http://doi.org/10.1016/j.jag.2006.12.001

Ballesteros, R., Ortega, J. F., Hernández, D., & Moreno, M. A. (2014a). Applications of georeferenced high-resolution images obtained with unmanned aerial vehicles. Part I: Description of image acquisition and processing. *Precision Agriculture*, *15*(6), 579–592. http://doi.org/10.1007/s11119-014-9355-8

Ballesteros, R., Ortega, J. F., Hernández, D., & Moreno, M. A. (2014b). Applications of georeferenced high-resolution images obtained with unmanned aerial vehicles. Part II: application to maize and onion crops of a semi-arid region in Spain. *Precision Agriculture*, *15*(6), 593–614. http://doi.org/10.1007/s11119-014-9357-6

Battude, M., Al Bitar, A., Morin, D., Cros, J., Huc, M., Marais Sicre, C., … Demarez, V. (2016). Estimating maize biomass and yield over large areas using high spatial and temporal resolution Sentinel-2 like remote sensing data. *Remote Sensing of Environment*, *184*, 668–681. http://doi.org/10.1016/j.rse.2016.07.030

Belgiu, M., & Drăgu, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, *114*, 24–31. http://doi.org/10.1016/j.isprsjprs.2016.01.011

Bendig, J., Bolten, A., Bennertz, S., Broscheit, J., Eichfuss, S., & Bareth, G. (2014). Estimating Biomass of Barley Using Crop Surface Models (CSMs) Derived from UAV-Based RGB Imaging. *Remote Sensing*, *6*(11), 10395–10412. http://doi.org/10.3390/rs61110395

Bolton, D. K., & Friedl, M. A. (2013). Forecasting crop yield using remotely sensed vegetation

indices and crop phenology metrics. *Agricultural and Forest Meteorology*, *173*, 74–84. http://doi.org/10.1016/j.agrformet.2013.01.007

Breiman, L., & Cutler, A. (2007). Random forests — Classification description: Random forests. Retrieved April 18, 2017, from https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm#varimp

Choodarathnakara, A. L., Kumar, D. T. A., Koliwad, D. S., & Patil, D. C. G. (2012). Mixed Pixels : A Challenge in Remote Sensing Data Classification for Improving Performance. *International Journal of Advanced Research in Computer Engineering & Technology*, *1*(9), 261–271.

Colomina, I., & Molina, P. (2014). Unmanned aerial systems for photogrammetry and remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *92*, 79–97. http://doi.org/10.1016/j.isprsjprs.2014.02.013

Cook, K. L. (2016). An evaluation of the effectiveness of low-cost UAVs and structure from motion for geomorphic change detection. *Geomorphology (in Revisions)*, *278*, 195–208. http://doi.org/10.1016/j.geomorph.2016.11.009

Deng, C., & Wu, C. (2013). The use of single-date MODIS imagery for estimating large-scale urban impervious surface fraction with spectral mixture analysis and machine learning techniques. *ISPRS Journal of Photogrammetry and Remote Sensing*, *86*, 100–110. http://doi.org/10.1016/j.isprsjprs.2013.09.010

Diaz-Varela, R. A., Zarco-Tejada, P. J., Angileri, V., & Loudjani, P. (2014). Automatic identification of agricultural terraces through object-oriented analysis of very high resolution DSMs and multispectral imagery obtained from an unmanned aerial vehicle. *Journal of Environmental Management*, *134*, 117–126. http://doi.org/10.1016/j.jenvman.2014.01.006

DJI. (2014). Ground Station Wireless Data-link User Manual. Retrieved from http://dl.djicdn.com/downloads/groundstation/en/Ground_Station_User_Manual_en_v3.04.pdf

ESA. (2017). Sentinel family. Retrieved February 25, 2017, from https://earth.esa.int/web/sentinel/home

Fitzpatrick-Lins, K. (1981). Comparison of sampling procedures and data-analysis for a land-use and land-cover map. *Photogrammetric Engineering and Remote Sensing*, *47*(3), 343–351. Retrieved from http://ezproxy.utwente.nl:2062/full_record.do?product=WOS&search_mode=GeneralSearch&qid=1&SID=T2RGxcXTzaIcCUtgcPQ&page=1&doc=1

Frazier, R. J., Coops, N. C., Wulder, M. A., & Kennedy, R. (2014). Characterization of aboveground biomass in an unmanaged boreal forest using Landsat temporal segmentation metrics. *ISPRS Journal of Photogrammetry and Remote Sensing*, *92*, 137–146. http://doi.org/10.1016/j.isprsjprs.2014.03.003

Gago, J., Douthe, C., Coopman, R. E., Gallego, P. P., Ribas-Carbo, M., Flexas, J., … Medrano, H. (2015). UAVs challenge to assess water stress for sustainable agriculture. *Agricultural Water Management*. http://doi.org/10.1016/j.agwat.2015.01.020

Gao, F., Hilker, T., Zhu, X., Anderson, M., Masek, J., Wang, P., & Yang, Y. (2015). Fusing Landsat and MODIS Data for Vegetation Monitoring. *IEEE Geoscience and Remote Sensing Magazine*, *3*(3), 47–60. http://doi.org/10.1109/MGRS.2015.2434351

Gascon, F. (2014). Sentinel-2 Products and Algorithms, Workshop "Preparations for Sentinel 2 in Europe." Oslo: EuroSDR, Norway Digital and Norwegian Space Centre. Retrieved from http://www.eurosdr.net/workshops/preparations-sentinel-2-europe

Genuer, R., Poggi, J.-M., & Tuleau-Malot, C. (2015). VSURF: An R Package for Variable Selection Using Random Forests. *The R Journal*, *7*(2), 19–33. Retrieved from https://journal.r-project.org/archive/2015-2/genuer-poggi-tuleaumalot.pdf

Greenwood, F. (2015). Drones and aerial observation: New Technologies for Property Rights, Human Rights, and Global Development. In *Drones and aerial observation* (pp. 35–47). New America. Retrieved from http://drones.newamerica.org/primer/Chapter 4.pdf

Immitzer, M., Vuolo, F., & Atzberger, C. (2016). First experience with Sentinel-2 data for crop and tree species classifications in central Europe. *Remote Sensing*, *8*(3). http://doi.org/10.3390/rs8030166

Inglada, J., Arias, M., Tardy, B., Hagolle, O., Valero, S., Morin, D., … Koetz, B. (2015). Assessment of an operational system for crop type map production using high temporal and spatial resolution satellite optical imagery. *Remote Sensing*, *7*(9), 12356–12379. http://doi.org/10.3390/rs70912356

Inglada, J., Vincent, A., Arias, M., & Marais-Sicre, C. (2016). Improved Early Crop Type Identification By Joint Use of High Temporal Resolution SAR And Optical Image Time Series. *Remote Sensing*, *8*(5), 362. http://doi.org/10.3390/rs8050362

ITC. (2013). *The core of GIScience: a process-based approach*. Enschede: Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente. Retrieved from https://blackboard.utwente.nl/bbcswebdav/pid-879149-dt-content-rid-1751335_2/courses/P15-EDU-109/CoreBook2012_2013.pdf

Jensen, J. R. (2005). *Introductory digital image processing : a remote sensing perspective* (3rd ed.). California: Prentice Hall. Retrieved from https://books.google.se/books/about/Introductory_Digital_Image_Processing.html?id=F84PAQAAIAAJ&redir_esc=y

Jin, H., & Eklundh, L. (2014). A physically based vegetation index for improved monitoring of plant phenology. *Remote Sensing of Environment*, *152*, 512–525. http://doi.org/10.1016/j.rse.2014.07.010

Jönsson, P., Cai, Z., Melaas, E., Friedl, M. A., & Eklundh, L. (2017). A Method for Robust Estimation of Seasonality for Landsat and Sentinel-2 Time Series Data.

Jönsson, P., & Eklundh, L. (2004). TIMESAT - A program for analyzing time-series of satellite sensor data. *Computers and Geosciences*, *30*(8), 833–845. http://doi.org/10.1016/j.cageo.2004.05.006

Knauer, K., Gessner, U., Fensholt, R., & Kuenzer, C. (2016). An ESTARFM fusion framework for the generation of large-scale time series in cloud-prone and heterogeneous landscapes. *Remote Sensing*, *8*(5). http://doi.org/10.3390/rs8050425

Kowalik, W., Dabrowska-Zielinska, K., Meroni, M., Raczka, T. U., & de Wit, A. (2014). Yield estimation using SPOT-VEGETATION products: A case study of wheat in European countries. *International Journal of Applied Earth Observation and Geoinformation*, *32*(1), 228–239. http://doi.org/10.1016/j.jag.2014.03.011

Lelong, C. C. D., Burger, P., Jubelin, G., Roux, B., Labbé, S., & Baret, F. (2008). Assessment of

unmanned aerial vehicles imagery for quantitative monitoring of wheat crop in small plots. *Sensors*, *8*(5), 3557–3585. http://doi.org/10.3390/s8053557

Li, X., Cheng, X., Chen, W., Chen, G., & Liu, S. (2015). Identification of Forested Landslides Using LiDar Data, Object-based Image Analysis, and Machine Learning Algorithms. *Remote Sensing*, *7*(8), 9705–9726. http://doi.org/10.3390/rs70809705

Löw, F., Michel, U., Dech, S., & Conrad, C. (2013). Impact of feature selection on the accuracy and spatial uncertainty of per-field crop classification using Support Vector Machines. *ISPRS Journal of Photogrammetry and Remote Sensing*, *85*, 102–119. http://doi.org/10.1016/j.isprsjprs.2013.08.007

Ma, Y., Soatto, S., Košecká, J., & Sastry, S. S. (2004). Estimation of Multiple Motions from Two Views. In *An Invitation to 3-D Vision: From Images to Geometric Models* (pp. 228–260). New York, NY: Springer New York. http://doi.org/10.1007/978-0-387-21779-6_7

Mesas-Carrascosa, F. J., Torres-Sánchez, J., Clavero-Rumbao, I., García-Ferrer, A., Peña, J. M., Borra-Serrano, I., & López-Granados, F. (2015). Assessing optimal flight parameters for generating accurate multispectral orthomosaics by UAV to support site-specific crop management. *Remote Sensing*, *7*(10), 12793–12814. http://doi.org/10.3390/rs71012793

Michez, A., Piégay, H., Jonathan, L., Claessens, H., & Lejeune, P. (2016). Mapping of riparian invasive species with supervised classification of Unmanned Aerial System (UAS) imagery. *International Journal of Applied Earth Observation and Geoinformation*, *44*(August), 88–94. http://doi.org/10.1016/j.jag.2015.06.014

Michez, A., Piégay, H., Lisein, J., Claessens, H., & Lejeune, P. (2016). Classification of riparian forest species and health condition using multi-temporal and hyperspatial imagery from unmanned aerial system. *Environmental Monitoring and Assessment*, *188*(3), 146. http://doi.org/10.1007/s10661-015-4996-2

Mkhabela, M. S., Mkhabela, M. S., & Mashinini, N. N. (2005). Early maize yield forecasting in the four agro-ecological regions of Swaziland using NDVI data derived from NOAA's-AVHRR. *Agricultural and Forest Meteorology*, *129*(1–2), 1–9. http://doi.org/10.1016/j.agrformet.2004.12.006

Moriondo, M., Maselli, F., & Bindi, M. (2007). A simple model of regional wheat yield based on NDVI data. *European Journal of Agronomy*, *26*(3), 266–274. http://doi.org/10.1016/j.eja.2006.10.007

Nesbit, P. R. (2014). *Uninhabited Aerial Vehicles and Structure from Motion: A Fresh Approach to Photogrammetry*. California State University. Retrieved from http://pqdtopen.proquest.com/doc/1626726998.html?FMT=AI

Quinn, G. P., & Keough, M. J. (2002). *Experimental Design and Data Analysis for Biologists. Experimental design and data analysis for biologists* (Vol. 277). Cambridge University Press. http://doi.org/10.1016/S0022-0981(02)00278-2

Räsänen, A., Rusanen, A., Kuitunen, M., & Lensu, A. (2013). What makes segmentation good? A case study in boreal forest habitat mapping. *International Journal of Remote Sensing*, *34*(23), 8603–8627. http://doi.org/10.1080/01431161.2013.845318

Richardson, A. D., Braswell, B. H., Hollinger, D. Y., Jenkins, J. P., & Ollinger, S. V. (2009). Near-surface remote sensing of spatial and temporal variation in canopy phenology. *Ecological Applications*, *19*(6), 1417–1428. http://doi.org/10.1890/08-2022.1

Rokhmana, C. A. (2015). The Potential of UAV-based Remote Sensing for Supporting Precision Agriculture in Indonesia. *Procedia Environmental Sciences*, *24*, 245–253. http://doi.org/10.1016/j.proenv.2015.03.032

Schellberg, J., Hill, M. J., Gerhards, R., Rothmund, M., & Braun, M. (2008). Precision agriculture on grassland: Applications, perspectives and constraints. *European Journal of Agronomy*. http://doi.org/10.1016/j.eja.2008.05.005

SLU. (2017). Swedish Infrastructure for Ecosystem Science. Retrieved March 5, 2017, from http://www.slu.se/en/Collaborative-Centres-and-Projects/sites/?si=53A2DE4F0AC29D651569BB2780B9DD2C&rid=73070847&sn=sluEPi6-prodSearchIndex

Song, X.-P., Potapov, P. V, Krylov, A., King, L., Di Bella, C. M., Hudson, A., … Hansen, M. C. (2017). National-scale soybean mapping and area estimation in the United States using medium resolution satellite imagery and field survey. *Remote Sensing of Environment*, *190*, 383–395. http://doi.org/10.1016/j.rse.2017.01.008

Stroppiana, D., Migliazzi, M., Chiarabini, V., Crema, A., Musanti, M., Villa, P., … Saffi, C. A. (2015). Rice yield estimation using multispectral data from uav: a preliminary experiment in northern italy. *Igarss 2015*, 4664–4667. http://doi.org/10.1109/IGARSS.2015.7326869

Talebpour, B., Türker, U., & Yegül, U. (2015). The Role of Precision Agriculture in the Promotion of Food Security. *International Journal of Agricultural and Food Research*, *4*(1), 1–23. Retrieved from https://www.researchgate.net/profile/Babak_Talebpour/publication/277306789_The_Role_of_Precision_Agriculture_in_the_Promotion_of_Food_Security/links/5566b2e508aeccd77735a775.pdf

Teets, E. H., Donohue, C. J., Underwood, K., & Bauer, J. E. (1998). Atmospheric Considerations for Uninhabited Aerial Vehicle (UAV) Flight Test Planning. http://doi.org/10.2514/6.1998-748

Tewes, A., Thonfeld, F., Schmidt, M., Oomen, R., Zhu, X., Dubovyk, O., … Schellberg, J. (2015). Using RapidEye and MODIS Data Fusion to Monitor Vegetation Dynamics in Semi-Arid Rangelands in South Africa. *Remote Sensing*, *7*(6), 6510–6534. http://doi.org/10.3390/rs70606510

Torres-Sánchez, J., López-Granados, F., De Castro, A. I., & Peña-Barragán, J. M. (2013). Configuration and Specifications of an Unmanned Aerial Vehicle (UAV) for Early Site-Specific Weed Management. *PLoS ONE*, *8*(3), e58210. http://doi.org/10.1371/journal.pone.0058210

Torres-Sánchez, J., Peña, J. M., De Castro, A. I., & López-Granados, F. (2014). Multi-temporal mapping of the vegetation fraction in early-season wheat fields using images from UAV. *Computers and Electronics in Agriculture*, *103*, 104–113. http://doi.org/10.1016/j.compag.2014.02.009

Tu, T. M., Su, S. C., Shyu, H. C., & Huang, P. S. (2001). A new look at IHS-like image fusion methods. *Information Fusion*, *2*(3), 177–186. http://doi.org/10.1016/S1566-2535(01)00036-7

Vega, F. A., Ramírez, F. C., Saiz, M. P., & Rosúa, F. O. (2015). Multi-temporal imaging using an unmanned aerial vehicle for monitoring a sunflower crop. *Biosystems Engineering*, *132*, 19–27. http://doi.org/10.1016/j.biosystemseng.2015.01.008

Vrieling, A., Skidmore, A. K., Wang, T., Meroni, M., Ens, B. J., Oosterbeek, K., … Paganini, M.

(2017). Spatially detailed retrievals of spring phenology from single-season high-resolution image time series. *International Journal of Applied Earth Observation and Geoinformation*, *59*, 19–30. http://doi.org/10.1016/j.jag.2017.02.021

Wall, L., Larocque, D., & Léger, P. (2008). The early explanatory power of NDVI in crop yield modelling. *International Journal of Remote Sensing*, *29*(8), 2211–2225. http://doi.org/10.1080/01431160701395252

Wang, H., Zhao, Y., Pu, R., & Zhang, Z. (2015). Mapping Robinia pseudoacacia forest health conditions by using combined spectral, spatial, and textural information extracted from IKONOS imagery and random forest classifier. *Remote Sensing*, *7*(7), 9020–9044. http://doi.org/10.3390/rs70709020

Wang, M., Tao, F. L., & Shi, W. J. (2014). Corn yield forecasting in northeast china using remotely sensed spectral indices and crop phenology metrics. *Journal of Integrative Agriculture*, *13*(7), 1538–1545. http://doi.org/10.1016/S2095-3119(14)60817-0

Westoby, M. J., Brasington, J., Glasser, N. F., Hambrey, M. J., & Reynolds, J. M. (2012). "Structure-from-Motion" photogrammetry: A low-cost, effective tool for geoscience applications. *Geomorphology*, *179*, 300–314. http://doi.org/10.1016/j.geomorph.2012.08.021

Wijesingha, J. (2016). *Geometric quality assessment of UAV-borne Remote Sensing (RS) products for Precision Agriculture (PA)*. Lund University.

Zhang, C., & Kovacs, J. M. (2012). The application of small unmanned aerial systems for precision agriculture: A review. *Precision Agriculture*, *13*(6), 693–712. http://doi.org/10.1007/s11119-012-9274-5

Appendix 1. Confusion matrices

Confusion matrix model 1 with single-date image and 12 variables

	Built up area	Rapeseed	Spring barley	Sugar beet	Wheat	Fava bean/ Wheat	Grass-legume ley	Kernza	Kernza/ Lucerne	Spring barley/ Lupin	Total	Error of commission	User accuracy
Built up area	116	0	0	2	0	2	0	0	0	1	121	4.1	95.9
Rapeseed	0	114	1	1	0	0	0	0	1	1	118	3.4	96.6
Spring barley	0	0	77	0	2	0	1	0	0	2	82	6.1	93.9
Sugar beet	2	0	0	123	0	0	1	2	1	5	134	8.2	91.8
Wheat	1	0	2	0	97	5	0	1	0	4	110	11.8	88.2
Fava bean/Wheat	1	0	0	0	0	25	2	0	1	1	30	16.7	83.3
Grass legume ley	1	0	0	1	0	2	27	0	1	1	33	18.2	81.8
Kernza	1	0	0	0	0	0	0	62	1	0	64	3.1	96.9
Kernza/Lucerne	0	0	0	1	0	0	1	4	38	0	44	13.6	86.4
Spring barley/Lupin	0	1	0	6	0	2	2	1	0	75	87	13.8	86.2
Total	122	115	80	134	99	36	34	70	43	90	823		
Error of omission	4.9	0.9	3.8	8.2	2.0	30.6	20.6	11.4	11.6	10.0			
Producer accuracy	95.1	99.1	96.3	91.8	98.0	69.4	79.4	88.6	88.4	90.0			

Overall classification accuracy	91.62
OOB estimate of error rate	8.38

Confusion matrix prediction based on model 1

	Built up area	Rapeseed	Spring barley	Sugar beet	Wheat	Fava bean/ Wheat	Grass-legume ley	Kernza	Kernza/ Lucerne	Spring barley/ Lupin	Total	Error of commission	User accuracy
Built up area	30	1	0	1	0	0	0	0	0	0	32	6.3	93.8
Rapeseed	0	37	0	0	1	0	0	0	0	0	38	2.6	97.4
Spring barley	0	1	33	0	1	0	0	0	0	0	35	5.7	94.3
Sugar beet	0	0	0	50	1	0	0	0	0	0	51	2.0	98.0
Wheat	0	0	0	0	36	1	0	0	0	0	37	2.7	97.3
Fava bean/Wheat	0	0	0	0	2	3	0	0	0	1	6	50.0	50.0
Grass legume ley	0	0	0	0	0	1	8	0	0	1	10	20.0	80.0
Kernza	0	0	0	0	0	1	0	10	1	0	12	16.7	83.3
Kernza/Lucerne	0	0	0	0	0	0	0	2	15	0	17	11.8	88.2
Spring barley/Lupin	0	0	1	1	1	1	1	0	0	31	36	13.9	86.1
Total	30	39	34	52	42	7	9	12	16	33	274		
Error of omission	0.0	5.1	2.9	3.8	14.3	57.1	11.1	16.7	6.3	6.1			
Producer accuracy	100.0	94.9	97.1	96.2	85.7	42.9	88.9	83.3	93.8	93.9			

Overall classification accuracy	92.34
Kappa coefficient	0.9122

Confusion matrix model 2 with 145 variables

	Built up area	Rapeseed	Spring barley	Sugar beet	Wheat	Fava bean/ Wheat	Grass-legume ley	Kernza	Kernza/ Lucerne	Spring barley/ Lupin	Total	Error of commission	User accuracy
Built up area	113	0	0	0	0	0	0	0	0	0	113	0.0	100.0
Rapeseed	1	108	1	0	1	0	0	0	0	0	111	2.7	97.3
Spring barley	0	0	85	0	0	0	0	0	0	3	88	3.4	96.6
Sugar beet	0	0	0	137	0	0	1	1	0	2	141	2.8	97.2
Wheat	0	0	0	0	110	1	0	0	0	1	112	1.8	98.2
Fava bean/Wheat	0	0	0	0	3	19	1	1	0	0	24	20.8	79.2
Grass legume ley	0	0	0	0	0	0	24	0	0	3	27	11.1	88.9
Kernza	0	0	0	0	0	0	0	54	1	0	55	1.8	98.2
Kernza/Lucerne	0	0	0	0	0	0	0	1	45	0	46	2.2	97.8
Spring barley/Lupin	0	1	0	0	0	0	2	0	0	88	91	3.3	96.7
Total	114	109	86	137	114	20	28	57	46	97	808		
Error of omission	0.9	0.9	1.2	0.0	3.5	5.0	14.3	5.3	2.2	9.3			
Producer accuracy	99.1	99.1	98.8	100.0	96.5	95.0	85.7	94.7	97.8	90.7			

Overall classification accuracy	96.91
OOB estimate of error rate	3.09

Confusion matrix prediction based on model 2

	Built up area	Rapeseed	Spring barley	Sugar beet	Wheat	Fava bean/ Wheat	Grass-legume ley	Kernza	Kernza/ Lucerne	Spring barley/ Lupin	Total	Error of commission	User accuracy
Built up area	36	0	0	0	0	0	0	0	0	0	36	0.0	100.0
Rapeseed	0	44	0	0	1	0	0	0	0	0	45	2.2	97.8
Spring barley	0	0	27	0	0	0	0	0	0	0	27	0.0	100.0
Sugar beet	0	0	0	43	0	0	0	0	0	0	43	0.0	100.0
Wheat	0	0	0	0	36	1	0	0	0	0	37	2.7	97.3
Fava bean/Wheat	0	0	0	0	1	8	0	0	0	0	9	11.1	88.9
Grass legume ley	0	0	0	1	0	1	11	0	0	0	13	15.4	84.6
Kernza	0	0	0	0	0	0	0	19	0	0	19	0.0	100.0
Kernza/Lucerne	0	0	0	0	0	0	0	1	13	0	14	7.1	92.9
Spring barley/Lupin	0	1	0	0	0	1	0	0	0	24	26	7.7	92.3
Total	36	45	27	44	38	11	11	20	13	24	269		
Error of omission	0.0	2.2	0.0	2.3	5.3	27.3	0.0	5.0	0.0	0.0			
Producer accuracy	100.0	97.8	100.0	97.7	94.7	72.7	100.0	95.0	100.0	100.0			

Overall classification accuracy	97.03
Kappa coefficient	0.9662

Confusion matrix model 3 with 30 variables

	Built up area	Rapeseed	Spring barley	Sugar beet	Wheat	Fava bean/ Wheat	Grass-legume ley	Kernza	Kernza/ Lucerne	Spring barley/ Lupin	Total	Error of commission	User accuracy
Built up area	113	0	0	0	0	0	0	0	0	0	113	0.0	100.0
Rapeseed	1	107	1	0	0	1	0	0	1	0	111	3.6	96.4
Spring barley	0	0	85	0	0	0	0	0	0	3	88	3.4	96.6
Sugar beet	0	0	0	136	0	0	2	1	0	2	141	3.5	96.5
Wheat	0	0	0	0	109	2	0	0	0	1	112	2.7	97.3
Fava bean/Wheat	0	0	0	1	2	19	0	1	0	1	24	20.8	79.2
Grass legume ley	0	0	0	2	0	0	22	0	0	3	27	18.5	81.5
Kernza	0	0	0	0	0	0	0	55	0	0	55	0.0	100.0
Kernza/Lucerne	0	0	0	1	0	0	0	2	43	0	46	6.5	93.5
Spring barley/Lupin	0	1	0	0	1	0	1	0	0	88	91	3.3	96.7
Total	114	108	86	140	112	22	25	59	44	98	808		
Error of omission	0.9	0.9	1.2	2.9	2.7	13.6	12.0	6.8	2.3	10.2			
Producer accuracy	99.1	99.1	98.8	97.1	97.3	86.4	88.0	93.2	97.7	89.8			

Overall classification accuracy	96.16
OOB estimate of error rate	3.84

Confusion matrix prediction based on model 3

	Built up area	Rapeseed	Spring barley	Sugar beet	Wheat	Fava bean/ Wheat	Grass-legume ley	Kernza	Kernza/ Lucerne	Spring barley/ Lupin	Total	Error of commission	User accuracy
Built up area	36	0	0	0	0	0	0	0	0	0	36	0.0	100.0
Rapeseed	0	44	0	0	1	0	0	0	0	0	45	2.2	97.8
Spring barley	0	0	27	0	0	0	0	0	0	0	27	0.0	100.0
Sugar beet	0	0	0	43	1	0	0	0	0	0	44	2.3	97.7
Wheat	0	0	0	0	36	1	0	0	0	0	37	2.7	97.3
Fava bean/Wheat	0	0	0	0	0	9	0	0	0	0	9	0.0	100.0
Grass legume ley	0	0	0	1	0	1	11	0	0	0	13	15.4	84.6
Kernza	0	0	0	0	0	0	0	20	1	0	21	4.8	95.2
Kernza/Lucerne	0	0	0	0	0	0	0	0	12	0	12	0.0	100.0
Spring barley/Lupin	0	1	0	0	0	0	0	0	0	24	25	4.0	96.0
Total	36	45	27	44	38	11	11	20	13	24	269		
Error of omission	0.0	2.2	0.0	2.3	5.3	18.2	0.0	0.0	7.7	0.0			
Producer accuracy	100.0	97.8	100.0	97.7	94.7	81.8	100.0	100.0	92.3	100.0			

Overall classification accuracy	97.40
Kappa coefficient	0.9662

Confusion matrix model 4 with 15 variables

	Built up area	Rapeseed	Spring barley	Sugar beet	Wheat	Fava bean/ Wheat	Grass-legume ley	Kernza	Kernza/ Lucerne	Spring barley/ Lupin	Total	Error of commission	User accuracy
Built up area	113	0	0	0	0	0	0	0	0	0	113	0.0	100.0
Rapeseed	1	108	1	0	0	1	0	0	0	0	111	2.7	97.3
Spring barley	0	0	85	0	0	0	0	0	0	3	88	3.4	96.6
Sugar beet	0	0	0	138	0	0	1	1	0	1	141	2.1	97.9
Wheat	0	0	0	0	106	5	0	0	0	1	112	5.4	94.6
Fava bean/Wheat	0	0	0	1	1	21	0	1	0	0	24	12.5	87.5
Grass legume ley	0	0	0	2	0	0	21	0	0	4	27	22.2	77.8
Kernza	0	0	0	0	0	0	0	54	1	0	55	1.8	98.2
Kernza/Lucerne	0	0	0	1	0	0	0	2	43	0	46	6.5	93.5
Spring barley/Lupin	0	1	0	0	1	0	1	0	0	88	91	3.3	96.7
Total	114	109	86	142	108	27	23	58	44	97	808		
Error of omission	0.9	0.9	1.2	2.8	1.9	22.2	8.7	5.2	2.3	19.6			
Producer accuracy	99.1	99.1	98.8	97.2	98.1	77.8	91.3	94.8	97.7	80.4			

Overall classification accuracy	96.16
OOB estimate of error rate	3.84

Confusion matrix prediction based on model 4

	Built up area	Rapeseed	Spring barley	Sugar beet	Wheat	Fava bean/ Wheat	Grass-legume ley	Kernza	Kernza / Lucerne	Spring barley/ Lupin	Total	Error of commission	User accuracy
Built up area	36	0	0	0	0	0	0	0	0	0	36	0.0	100.0
Rapeseed	0	44	0	0	1	0	0	0	0	0	45	2.2	97.8
Spring barley	0	0	27	0	0	0	0	0	0	0	27	0.0	100.0
Sugar beet	0	0	0	43	1	1	0	0	0	0	45	4.5	95.5
Wheat	0	0	0	0	36	1	0	0	0	0	37	2.7	97.3
Fava bean/Wheat	0	0	0	0	0	9	0	0	0	0	9	0.0	100.0
Grass legume ley	0	0	0	1	0	0	11	0	0	0	12	8.3	91.7
Kernza	0	0	0	0	0	0	0	20	0	0	20	0.0	100.0
Kernza/Lucerne	0	0	0	0	0	0	0	0	13	0	13	0.0	100.0
Spring barley/Lupin	0	1	0	0	0	0	0	0	0	24	25	4.0	96.0
Total	36	45	27	44	38	11	11	20	13	24	269		
Error of omission	0.0	2.2	0.0	2.3	5.3	18.2	0.0	0.0	0.0	0.0			
Producer accuracy	100.0	97.8	100.0	97.7	94.7	81.8	100.0	100.0	100.0	100.0			

Overall classification accuracy	97.77
Kappa coefficient	0.9746

Confusion matrix model 5 with 8 variables

	Built up area	Rapeseed	Spring barley	Sugar beet	Wheat	Fava bean/ Wheat	Grass-legume ley	Kernza	Kernza / Lucerne	Spring barley/ Lupin	Total	Error of commission	User accuracy
Built up area	111	0	0	0	0	0	0	0	0	0	111	0.0	100.0
Rapeseed	0	118	0	0	0	0	0	0	0	1	119	0.8	99.2
Spring barley	0	0	83	0	0	0	0	0	0	2	85	2.4	97.6
Sugar beet	0	0	0	124	0	0	6	1	0	3	134	7.5	92.5
Wheat	0	2	0	0	111	4	1	0	0	2	120	7.5	92.5
Fava bean/Wheat	1	1	0	1	1	18	0	0	0	2	24	25.0	75.0
Grass legume ley	0	0	0	1	0	0	27	0	0	1	29	6.9	93.1
Kernza	0	0	0	0	0	0	0	60	1	0	61	1.6	98.4
Kernza/Lucerne	0	0	0	0	1	0	1	5	38	0	45	15.6	84.4
Spring barley/Lupin	0	0	0	0	0	2	0	0	0	78	80	2.5	97.5
Total	112	121	83	126	113	24	35	66	39	89	808		
Error of omission	0.9	2.5	0.0	1.6	1.8	25.0	22.9	9.1	2.6	12.4			
Producer accuracy	99.1	97.5	100.0	98.4	98.2	75.0	77.1	90.9	97.4	87.6			

Overall classification accuracy	95.05
OOB estimate of error rate	4.95

Confusion matrix prediction based on model 5

	Built up area	Rapeseed	Spring barley	Sugar beet	Wheat	Fava bean/ Wheat	Grass-legume ley	Kernza	Kernza / Lucerne	Spring barley/ Lupin	Total	Error of commission	User accuracy
Built up area	38	0	0	0	0	0	0	0	0	0	38	0.0	100.0
Rapeseed	0	36	1	0	0	0	0	0	0	0	37	2.7	97.3
Spring barley	0	0	26	0	1	0	0	0	0	0	27	3.7	96.3
Sugar beet	0	0	0	51	0	0	0	1	0	0	52	1.9	98.1
Wheat	0	0	2	0	26	0	0	0	0	0	28	7.1	92.9
Fava bean/Wheat	0	0	0	0	2	6	0	0	0	0	8	25.0	75.0
Grass legume ley	0	0	0	0	0	0	8	0	0	1	9	11.1	88.9
Kernza	0	0	0	0	0	1	0	13	1	0	15	13.3	86.7
Kernza/Lucerne	0	0	0	0	0	0	0	0	13	0	13	0.0	100.0
Spring barley/Lupin	0	1	1	0	1	4	1	0	0	34	42	19.0	81.0
Total	38	37	30	51	30	11	9	14	14	35	269		
Error of omission	0.0	2.7	13.3	0.0	13.3	45.5	11.1	7.1	7.1	2.9			
Producer accuracy	100.0	97.3	86.7	100.0	86.7	54.5	88.9	92.9	92.9	97.1			

Overall classification accuracy	93.31
Kappa coefficient	0.9234

Appendix 2. Spectral signature of crops on 11 spectral bands of Sentinel-2A

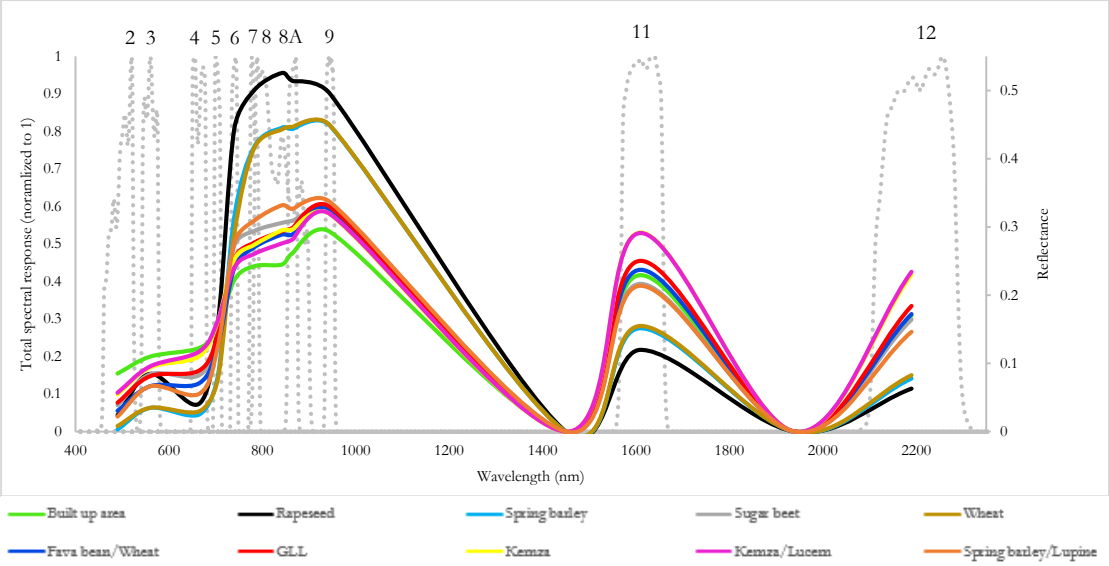
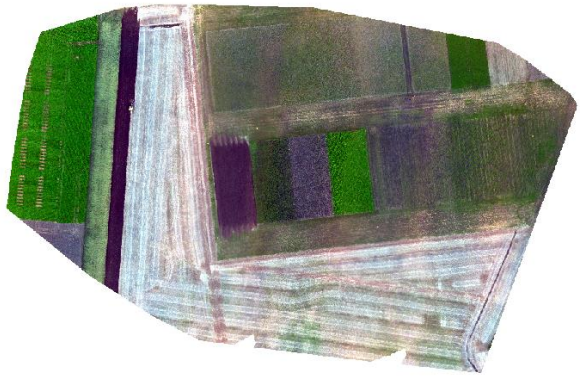


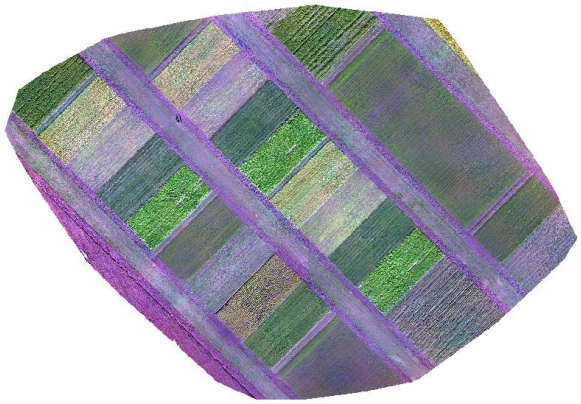
Figure 25. Spectral signature of Lönnstorp crops on 11 bands of Sentinel-2images of June 11th, 2016

Appendix 3. Orthomosaics with ATLAS radiometric correction

August 12th, 2016



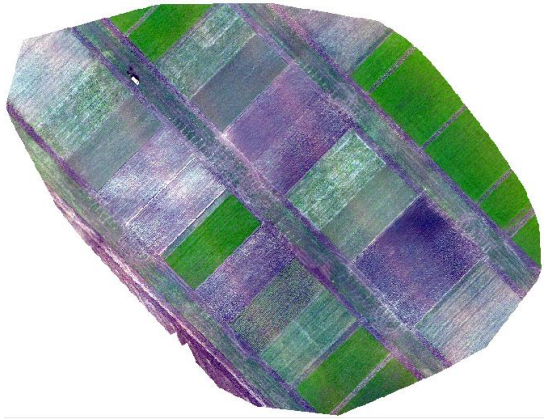
November 3rd, 2016



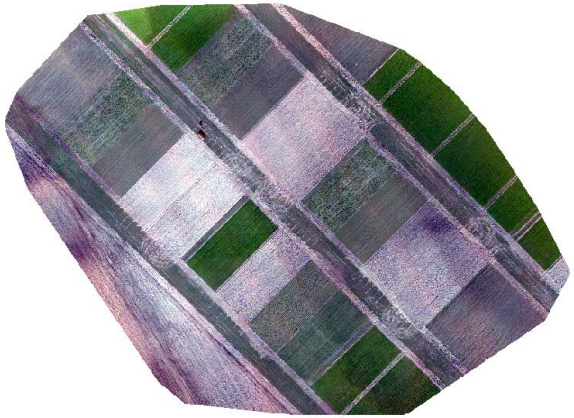
December 2nd, 2016

March 7th, 2017

March 13th, 2017



March 24th, 2017



Appendix 4. t-test and Wilcoxon signed rank test of reflectance data for other image pairs

	August 12th with July 31st, 2016				
	B	G	R	RedEdge	NIR
Observations	90				
df	89				
Pearson correlation (r)	0.33	0.09	0.17	-0.03	0.9
Wilcoxon signed rank	*p<0.001*	*p<0.001*	*p<0.001*	*p<0.001*	*p<0.001*
t-stat	11.22	21.26	5.631	21.22	18.13
P(T≤t) two-tail	*p<0.001*	*p<0.001*	*p<0.001*	*p<0.001*	*p<0.001*
t-critical two-tail	1.98				

	March 7th with 11th, 2017				
	B	G	R	RedEdge	NIR
Observations	80				
df	79				
Pearson correlation (r)	0.83	0.77	0.94	0.72	0.97
Wilcoxon signed rank	*p<0.001*	*p<0.001*	*p<0.001*	*p<0.001*	*p<0.001*
t-Stat	78.28	59.63	35.92	26.23	21.93
P(T≤t) two-tail	*p<0.001*	*p<0.001*	*p<0.001*	*p<0.001*	*p<0.001*
t-critical two-tail	1.99				

	March 24rd with 11th, 2017				
	B	G	R	RedEdge	NIR
Observations	101				
df	100				
Pearson correlation (r)	0.70	0.11	0.84	0.32	0.92
Wilcoxon signed rank	*p<0.001*	*p<0.001*	*p<0.001*	*p<0.001*	*p<0.001*
t-stat	17.36	18.29	9.30	38.99	18.43
P(T≤t) two-tail	*p<0.001*	*p<0.001*	*p<0.001*	*p<0.001*	*p<0.001*
t-critical two-tail	1.98				

Appendix 5. Boxplot of reflectances per band for both sensors

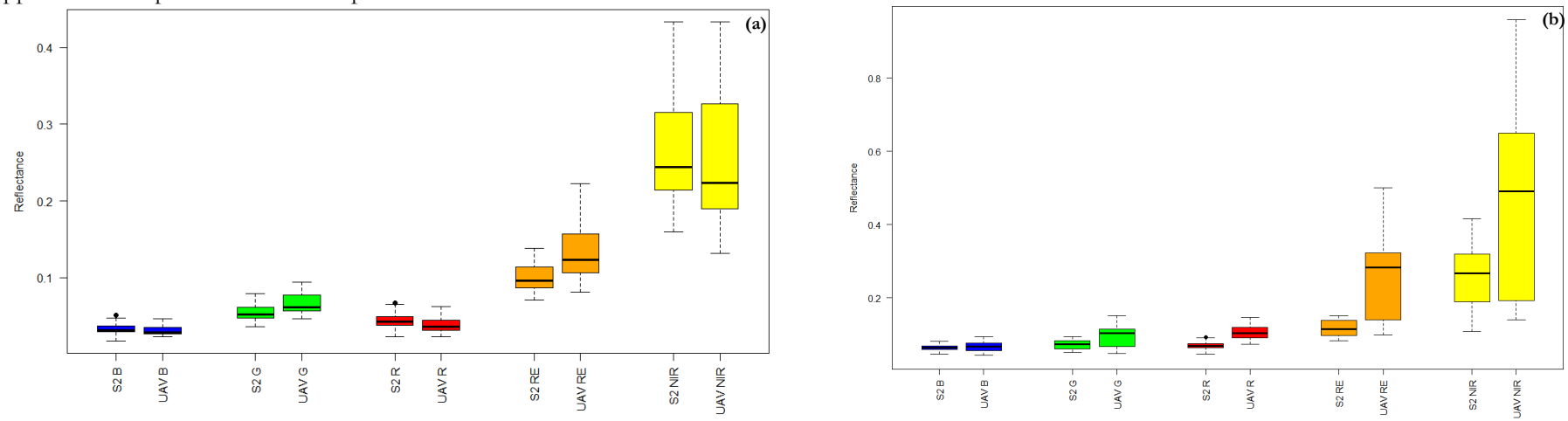


Figure 26. a) November 3rd with 8th pair, b) December 2nd with November 28th pair

Appendix 6. Boxplot of vegetation indices per sensor

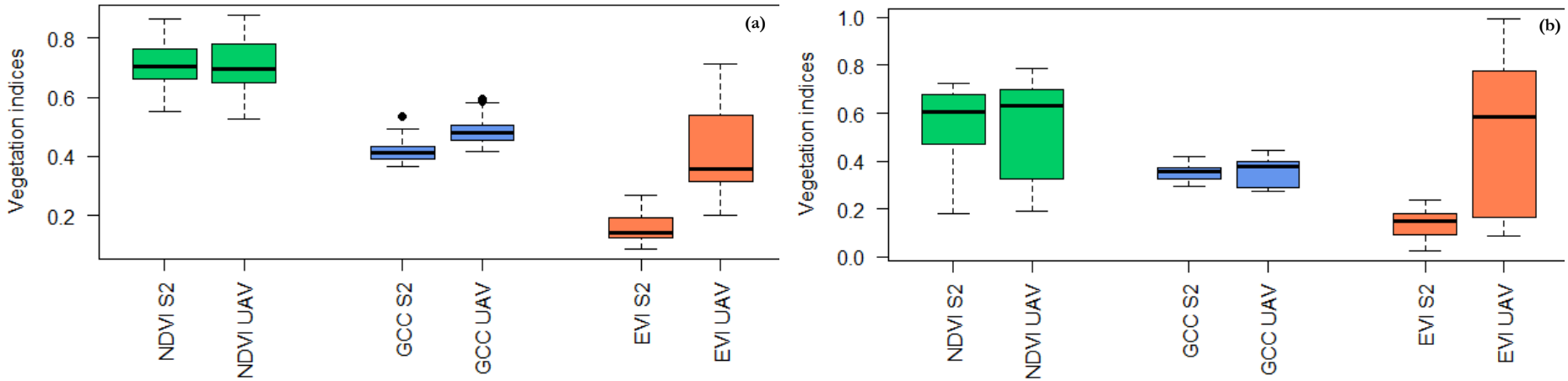


Figure 27. a) November 3rd with 8th pair, b) December 2nd with November 28th pair

Department of Physical Geography and Ecosystem Science, Lund University

Lund University GEM thesis series are master theses written by students of the international master program on Geo-information Science and Earth Observation for Environmental Modelling and Management (GEM). The program is a cooperation of EU universities in Iceland, the Netherlands, Poland, Sweden and UK, as well a partner university in Australia. In this series only master thesis are included of students that performed their project at Lund University. Other theses of this program are available from the ITC, the Netherlands (www.gem-msc.org or www.itc.nl).

The student thesis reports are available at the Geo-Library, Department of Physical Geography and Ecosystem Science, University of Lund, Sölvegatan 12, S-223 62 Lund, Sweden. Report series started 2013. The complete list and electronic versions are also electronic available at the LUP student papers (https://lup.lub.lu.se/student-papers/search/) and through the Geo-library (www.geobib.lu.se).

1 Soheila Youneszadeh Jalili (2013) The effect of land use on land surface temperature in the Netherlands

2 Oskar Löfgren (2013) Using Worldview-2 satellite imagery to detect indicators of high species diversity in grasslands

3 Yang Zhou (2013) Inter-annual memory effects between Soil Moisture and NDVI in the Sahel

4 Efren Lopez Blanco (2014) Assessing the potential of embedding vegetation dynamics into a fire behaviour model: LPJ-GUESS-FARSITE

5 Anna Movsisyan (2014) Climate change impact on water and temperature conditions of forest soils: A case study related to the Swedish forestry sector

6 Liliana Carolina Castillo Villamor (2015) Technical assessment of GeoSUR and comparison with INSPIRE experience in the context of an environmental vulnerability analysis using GeoSUR data

7 Hossein Maazallahi (2015) Switching to the "Golden Age of Natural Gas" with a Focus on Shale Gas Exploitation: A Possible Bridge to Mitigate Climate Change?

8 Mohan Dev Joshi (2015) Impacts of Climate Change on *Abies spectabilis*: An approach integrating Maxent Model (MAXent) and Dynamic Vegetation Model (LPJ-GUESS)

9 Altaaf Mechiche-Alami (2015) Modelling future wheat yields in Spain with LPJ-GUESS and assessing the impacts of earlier planting dates

10 Koffi Unwana Saturday (2015) Petroleum activities, wetland utilization and livelihood changes in Southern Akwa Ibom State, Nigeria: 2003-2015

11 José Ignacio Díaz González (2016) Multi-objective optimisation algorithms for GIS-based multi-criteria decision analysis: an application for evacuation planning

12 Gunjan Sharma (2016) Land surface phenology as an indicator of performance of conservation policies like Natura2000

13 Chao Yang (2016) A Comparison of Four Methods of Diseases Mapping

14 Xinyi Dai (2016) Dam site selection using an integrated method of AHP and GIS for decision making support in Bortala, Northwest China

15 Jialong Duanmu (2016) A multi-scale based method for estimating coniferous forest aboveground biomass using low density airborne LiDAR data

16 Tanyaradzwa J. N. Muswera (2016) Modelling maize (Zea Mays L.) phenology using seasonal climate forecasts

17 Maria Angela Dissegna (2016) Improvement of the GPP estimations for Sudan using the evaporative fraction as water stress factor

18 Miguel G. Castro Gómez (2017) Joint use of Sentinel-1 and Sentinel-2 for land cover classification: A machine learning approach

19 Krishna Lamsal (2017) Identifying potential critical transitions in a forest ecosystem using satellite data

20 Maimoona Zehra Jawaid (2017) Glacial lake flood hazard assessment and modelling: a GIS perspective

21 Tracy Zaarour (2017) Application of GALDIT index in the Mediterranean region to assess vulnerability to sea water intrusion

22 Stephania Zabala (2017) Comparison of multi-temporal and multispectral Sentinel-2 and Unmanned Aerial Vehicle imagery for crop type mapping