ALMA MATER STUDIORUM

UNIVERSITA' DI BOLOGNA

SCUOLA DI SCIENZE

SECOND CYCLE DEGREE in

ENVIRONMENTAL ASSESSMENT AND MANAGEMENT,

Curriculum Climate KIC

Automatic methods for crop classification by merging radar (Sentinel1) and optical (Sentinel2) satellite images and Artificial Intelligence analysis

Thesis in Environmental, Political and Economic Management Systems

Relatore

Prof.re Andrea Contin

Correlatore   Presents

Dott.re Nicolas Greggio   Beatrice Gottardi

Controrelatore

Dott.re Alessandro Buscaroli

Academic year 2018-2019

24th of March 2020

"Start with why…"

Index

ABBREVIATIONS

AAN	Agricultural Area of the Netherlands
ACC	Accuracy
ANN	Artificial Neural Network
API	Application Programming Interface
AQ60	Cloud Bitmask for Sentinel2
BSI	Bare Soil Index
BRP	Basic Registration of crop Plots
CAP	Common Agricultural Policy
CCRS	Canada Center for Remote Sensing
CNN	Convolution Neural Network
DL	Deep Learning
EM	ElectroMagnetic
EO	Earth Observation
ESA	European Space Agency
ETM+	Enhanced Thematic Mapper Plus
EUMETSAT	Exploitation of Meteorological Satellites
GEE	Google Earth Engine
GRD	Ground Range Detected
IDE	Integrated Development Environment
IW	Interferometric Wide
KNN	K-Nearest Neighbours
LPIS	Land Parcel identification System
ML	Machine Learning
MSI	Multi-Spectral Instrument (MSI)
MV	Microwave electromagnetic spectrum
NASA	National AeroSpace Administration
NDVI	Normalized Difference Vegetation Index
NIR	Near Thermal Infrared electromagnetic spectrum
OA	Overall Accuracy
OLI	Operational Land Imager
OOB	Out-Of-Bag

RA	Random Accuracy
RF	Random Forest
RFC	Random Forest Classifier
RS	Remote Sensing
SAR	Synthetic-Aperture Radar
SR	Surface Reflectance
SSD	Spatial Sampling Distance
STOWA	Foundation for Applied Water Management Research in the Netherlands
SVM	Support Vector Machine
SWIR	ShortWave InfraRed electromagnetic spectrum
TIR	Thermal Infrared electromagnetic spectrum
TIRS	Thermal Infrared Sensor
TOA	Top-of-Atmosphere
VI	Vegetation Index
VIS	Visible electromagnetic spectrum

1 Introduction

Human primary needs for food, water, fuel, clothing and shelter must be met from the land, which is in limited supply. As population and aspirations increase, land becomes an increasingly scarce resource. Studying land use and land cover is important to make the best use of this limited resource.

Land use must change to meet new demands. However, land use changes bring up conflicts between competing uses, i.e. between the interests of individuals and the common good. Land covered by towns and industry is no longer available for farming. Likewise, the development of new farmland competes with forestry, water supplies and wildlife.

Land-use planning is the systematic assessment of land and water resource, the development of options for land use and the study of the socio-economic conditions in order to select and adopt the best land-use choices (Sicre, et al., 2020). Its purpose is to select and put into practice those land uses that will best weight the needs of people and the protection of resources for future generations. The driving forces in planning are the need for improved management and the need for a different land uses due to changing circumstances.

Planning to make the best use of land is not a new idea. Over the years, farmers have made plans, season after season, deciding what to grow and where to grow it. Their decisions have been made according to their own needs, their knowledge of the land and the technology, labour and capital available. As the size of the area, the number of people involved and the complexity of the problems increased, so did the need for information and rigorous methods of analysis and planning (FAO, 1993).

1.1 Remote sensing for Earth Observation, Land Use and Land Cover

Earth Observation (EO) is the gathering of information about planet Earth's physical, chemical and biological systems via remote sensing technologies, usually involving satellites carrying imaging devices. EO is used to monitor and assess the status of, and changes in, the natural and manmade environment. Space-based technologies deliver

reliable and repeat-coverage datasets, which, combined with research and the development of appropriate methods, provide a unique mean for gathering information concerning the planet. Examples include the monitoring of the state and evolution of our environment, be it land, sea or air, and the ability to rapidly assess situations during crises such as extreme weather events or during conflicts (European Commission, 2016).

Remote sensing by imaging systems consists in a process where the interaction between incident radiation and the target of interest is involved. The following seven points, well explained in the "Fundamental of Remote Sensing" by CCRS (Canada Center for Remote Sensing, 2015), shortly describe the process. In Figure 1, the whole procedure is represented.
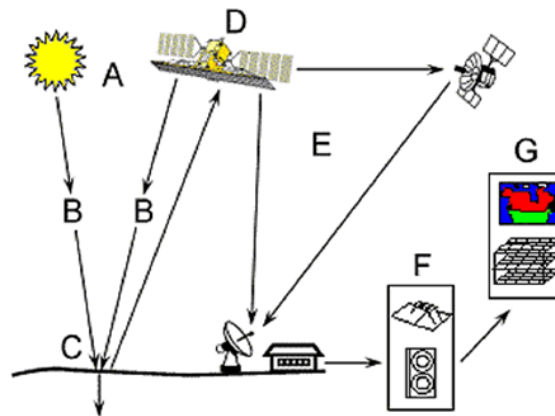
Figure 1: Picture representing the elements of a remote sensing system. (Canada Center for Remote Sensing, 2015)

1. **Energy Source or Illumination (A)**. The requirement number one for remote sensing is to have an energy source which provides electromagnetic energy to the target of interest. The energy source could be the Sun (passive remote sensing, e.g. optical satellite) or a wave emitted by the satellite (active remote sensing, e.g. radar satellites). Also, the direction of the electromagnetic radiation is used to gather information.

2. **Radiation and the Atmosphere (B)**. On the way from source to target, the wave will interact with the medium it passes through. This interaction may take place a second time as the energy comes back to the source and is detected. Part of the electromagnetic wave could be absorbed so that it does not reach the ground, or the reflected wave

could be absorbed so that it does not reach the detector. It is important to identify the so-called "atmospheric windows" to define the characteristics of the sensors on board, and to create specific filters. In the analysis of the data, these interferences between the atmosphere and the wave have to be removed, e.g. with atmospheric correction and cloud filters (more details in section 3.1).

3. **Interaction with the Target (C)**. The radiation interacts with the target depending on the properties of both the target and the radiation. Part of it can be absorbed, reflected, scattered or its wavelength modified.

4. **Detection and Recording by the Sensor (D)**. After the energy has been scattered by, or emitted from, the target, a sensor is required to collect and record the radiation.

5. **Transmission, Reception, and Processing (E)**. The detected signal has to be transmitted, often in electronic form, to a receiving and processing station where data are translated into an "image", i.e., a collection of information which includes images but also other data or meta-data about the collected radiation (polarisation, direction, etc.).

6. **Interpretation and Analysis (F)**. The processed image is interpreted, visually and/or digitally, to extract information about the target. This step allows converting an image into physical measurements. The conversion is performed using calibration and cleaning from background noise.

7. **Application (G)**. Finally, the information extracted from the collection of images, called "imagery", describes some target properties, assist in solving a particular environmental problem or describe processes.

The study described in the thesis mainly falls into the last point of the process.

Recently, the improvement of remote sensing for EO allowed the development of better monitoring and a better description of the changes in the Earth surface. A great amount of data has been stored and data analysts are working together with physical earth systems scientists in order to analyse this information. Traditional physical models are no more the only prediction methods which can be applied: artificial intelligence is now supporting big data analysis and extrapolations. For example, satellite images have been largely exploited for mapping continental areas (Hościło & Lewandowska, 2019), for

analysing crops (Kristof Van Tricht, 2018; Sicre, et al., 2020), for estimating biophysical parameters (Betbeder, et al., 2016), etc.

The exploitation of multispectral optical sensors to identify differences among various vegetation types (see section 2.2.4) has various proven methods, e.g. using vegetation indexes (Nathalie Pettorelli, 2005; Xingwang Fan, 2016). However, cloud-free or weather independent data are necessary to map cloud-prone regions. Very recently, given the complementary nature of optical and radar signals, notably their different penetration capacities (see section 2.2.3), they have been used in synergy to improve performances in agricultural land monitoring (Steinhausena, et al., 2018). Moreover, vegetative cover is characterized by strong variations within relatively short time intervals. These dynamics are challenging for land cover classifications but deliver crucial information that can be used to improve the performances of a machine learning algorithm that detects earth's surface differences. Studying the time variability of the collected data may provide useful information both for understanding differences among crops and variability over the phenology of the plant (Bargiel, 2017; Kristof Van Tricht, 2018).

Satellite images are heavy datasets that can be difficult to handle, particularly on local workstations, when artificial intelligence algorithms are developed and applied. This has been a hurdle that had to be overcame in order to allow open and free access to these types of datasets and permit a knock-on disrupting added value to worldwide users, from policy makers to commercial and private users. A new tool, launched by Google in 2014, called Google Earth Engine (GEE), allows "computing in the cloud" and fast visualisation of the results. Since 2017, it has started to receive more and more interest, as a promising interface to analyse satellite images for environmental assessment and for crop mapping (Fuyou, et al., 2019), decreasing the need for computing time on local machines.

Satellite data are a valuable source for land use and land cover mapping (European Commission, 2016). Land use and land cover maps can support the understanding of coupled human-environment systems and can provide important information for environmental modelling, water resources, food waste management and new bioenergy resources implementation.

Many different campaigns have been fostered from Countries, private and public companies all over the World, for both space and ground sectors. In the last decades, NASA from USA and ESA from EU have acquired a crucial role in aerospace innovation and research, producing a relevant amount of data and making them available openly to scientists. In particular, the Copernicus Progamme launched by ESA, provides a unique combination of free access, high spatial resolution, wide field of view and spectral coverage. It represents a major step forward compared to previous missions. Therefore, the use of these new data sources may increase the quality of final crop mapping results.

1.2   The Copernicus Programme

Copernicus is the European Union's Earth Observation and Monitoring programme, looking at our planet and its environment for the ultimate benefit of all European citizens. Thanks to a variety of technologies, from satellites in space to measurement systems on the ground, in the sea and in the air, Copernicus delivers operational data and information services openly and freely available for a wide range of application areas.
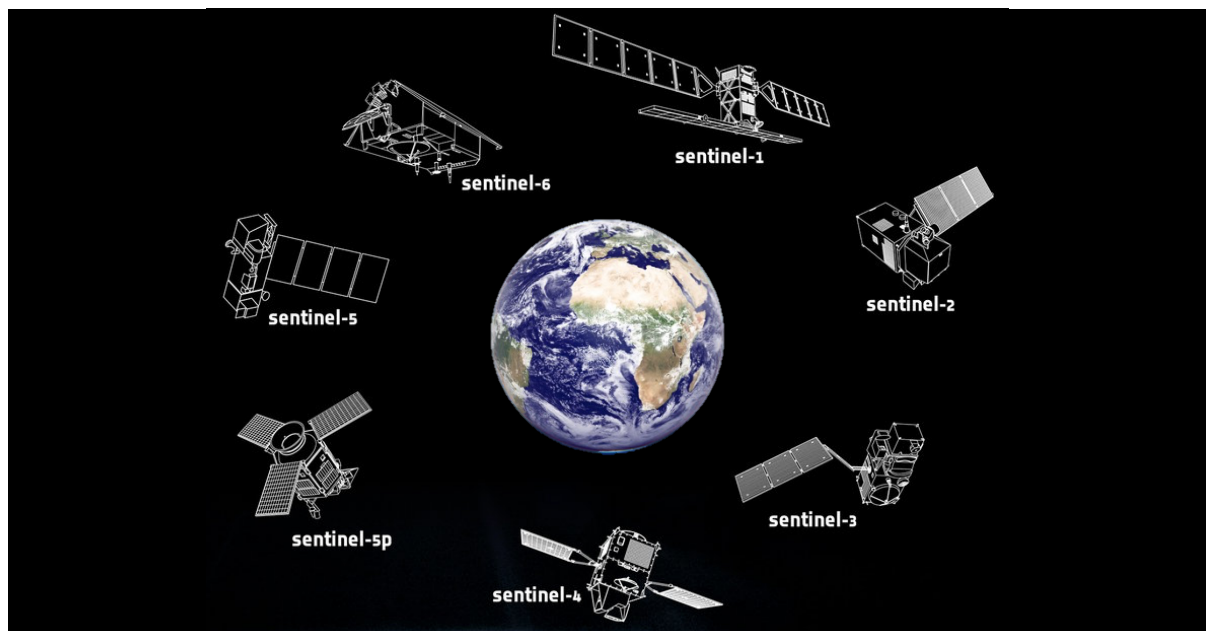

Figure 2: The Sentinels, the six missions launched till 2019 from Copernicus Programmed (source: ESA)

In addition to supporting environmental protection and the security of citizens, Copernicus stimulates European companies to explore new business opportunities and

foster job creation. Due to its boost to many real applications that can support decision makers and new start-ups, this Programme has a high potential for the general public.

As described in the European Commission Report "Europe's Eyes on Earth" (2015), the ''Copernicus programme is supported by a family of dedicated, EU-owned satellites (the Sentinels) specifically designed to meet the needs of the Copernicus services and their users". Since the launch of the first satellite – Sentinel-1A – in 2014, the European Union set in motion a process to place a constellation of more than a dozen satellites in orbit during the next ten years (see **Error! Reference source not found.**).

The Sentinels "fulfil the need for a consistent and independent source of high-quality data for the Copernicus services" (European Commission , 2015). These satellites are equipped with different sensors on board, able to detect different features of air, water and ground matrixes. An in-depth description of Sentinel-1 and Sentinel 2 missions is given in sections 2.2.3 and 2.2.4. The global coverage of Sentinel1 and Sentinel2 is shown in Figure 3.

✓ Two satellites in a 12 day orbit
✓ Repeat frequency: 6 days (important for coherence)
✓ Revisit frequency: (asc/desc & overlap): 3 days at the equator, <1 day at high latitudes (Europe ~ 2 days)
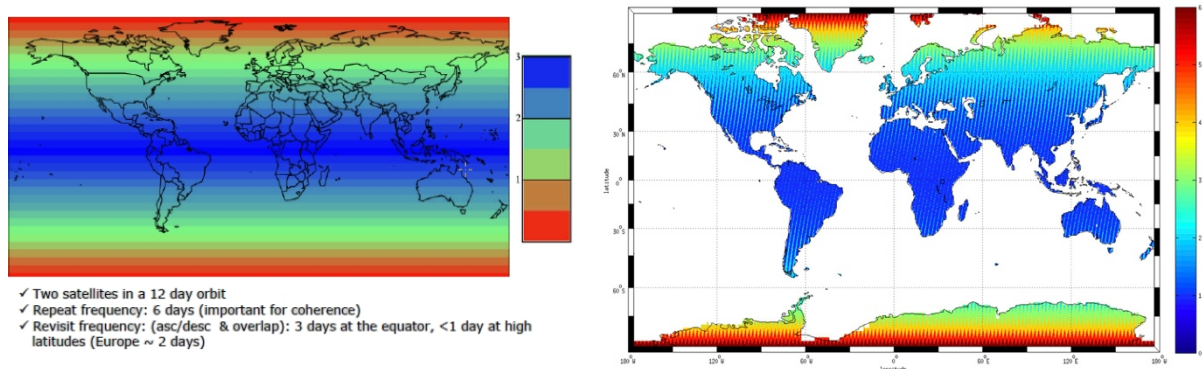
Figure 3: Representation of number of passages on global scale. a) Sentinel1 (including the number of revisit frequency, ascending and descending, within a 12 day orbit around the globe – every three days in blue – less than 1 day in red); b) Sentinel2 missions, including the geometric revisit frequency of a single satellite (for every track at the equator – blue – there are up to 6 tracks at poles – red). Source: ESA.

Copernicus also builds on existing space infrastructure: satellites operated by the European Space Agency (ESA), the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT), the EU Member States and other third countries and commercial providers, and draws on a large number of in situ (meaning on-site or local) measurement systems put at the disposal of the programme by the EU Member States. These include sensors placed on the banks of rivers, carried through the air by

weather balloons, pulled through the sea by ships, or floating in the ocean. In situ data are used to calibrate, verify and supplement the information provided by satellites, which is essential in order to deliver reliable and consistent data over time (European Commission , 2015).

The Copernicus Services transform this wealth of satellite and in situ data into value-added information, by processing and analysing the data, integrating them with other sources and validating the results. Datasets stretching back for years and decades are made comparable and searchable, thus ensuring the monitoring of changes; patterns are examined and used to create better forecasts, for example, of the ocean and the atmosphere. Maps are created from imagery, features and anomalies are identified and statistical information is extracted.

1.3   Classification of the electromagnetic spectrum

The frequency bands that are used in Remote Sensing (RS) can be divided into 5 fundamental ranges of the electromagnetic (EM) spectrum:

- Visible (VIS – 0.4μm$<\lambda<0.7\mu$m),
- Near Thermal Infrared (NIR – 0.7μm$<\lambda<1.1\mu$m),
- Shortwave Infrared (SWIR – 1.1μm$<\lambda<3\mu$m) distinguished into SWIR$_1$ (lower part of the interval) and SWIR$_2$ (higher part of the interval),
- Thermal Infrared (TIR – 4μm$<\lambda<20\mu$m) and
- Microwave (MV – 1mm$<\lambda<$1m).

In turn, the microwave spectrum can be divided into 6 bands according to the frequency: P (from 0.3 GHz), L, S, C (3.9 GHz to 5.8 GHz), X, K (up to 36 GHz). Longer wavelength microwave radiation can penetrate through cloud cover and dust. Therefore, different in frequencies allow different penetration of the wave to the target. Microwave range is used by radar satellite, while from VIS to SWIR (TIR) waves are detected by optical sensors.

The signal is not detectable through all the electromagnetic spectrum due to atmospheric interferences, especially for λ <1 mm. The Atmospheric windows are spectral regions in which there is a particular transparency of the atmosphere, so that the transmission through the atmosphere is maximum. As shown in Figure 4, optical

sensors cannot acquire all the spectrum, hence the signal is divided into different bands. Observing the earth surface, in particular vegetation cover, every band can highlight different aspects of the canopy, e.g. characteristic colours, photosynthetic activity, response to irrigation or drought periods, etc.

Comparison of Landsat 7 and 8 bands with Sentinel-2
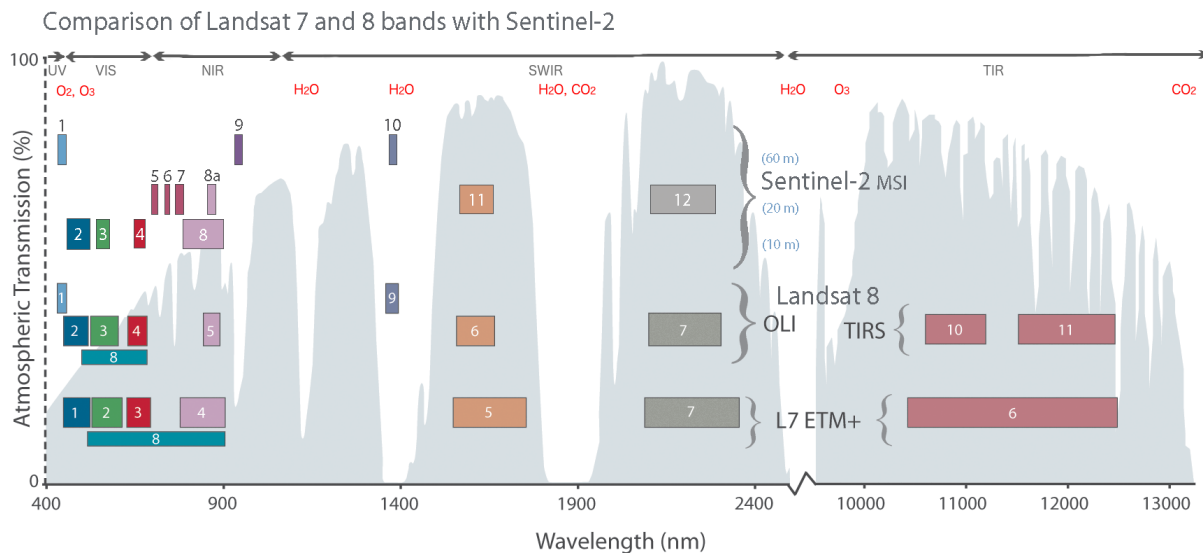
Figure 4: Atmospheric windows. The atmospheric transmission (in grey) of the electromagnetic spectrum depends on atmospheric components absorbance (red). Boxes represent bands of detection signals from different optical remote sensors. Sentinel-2 Multi-Spectral Instrument (MSI) collects 13 bands with three different spatial resolutions from UV to SWIR wavelengths. Landsat 8 with Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) and Landsat 7 with Enhanced Thematic Mapper Plus (ETM+) sensor are part of NASA earth observation missions. Source: NASA (Rocchio, 2020; S.K. Alavipanah, 2010).

Concerning radar images, not only the intensity of the signal is relevant but also its orientation in space. A factor that influences the interaction mechanisms on the target is polarization. In fact, a polarized electromagnetic wave interacts preferentially with the structures oriented along the polarization plane. Changes in polarization are caused by different modes of interaction with the surface. With Sentinel1 radar sensors it is possible to detect backscatter intensity from 4 types of polarization.

- VV and HH dual polarization (vertical emitted end vertical received or horizontal emitted and received)
- VH and HV dual cross-polarisation (vertical emitted and horizontal received or horizontal emitted and vertical received)

In particular, three different mechanisms can happen on a vegetation cover (see Figure 5): ground, stems and leaves responds with different surface and volume backscatter

mechanisms, where backscattering from vegetation canopy is called "volume" backscattering. Stem-ground double-bounce typically occurs due to a mixed effect among different surfaces, so that the radiation is reflected towards the radar sensor, resulting in high backscatter values and a tendency of change in polarization. On the other hand, on a simple surface the reflection and the volume scattering (as on plants' leaves) backscatter may be attenuated due to a reflection away from the radar sensor. Wave polarimetry intensity depends on how and how much the radiation is scattered, which in turn depends on how the canopy is structured. Hence, Synthetic-Aperture Radar (SAR) images are representative of the structure of the canopy cover.
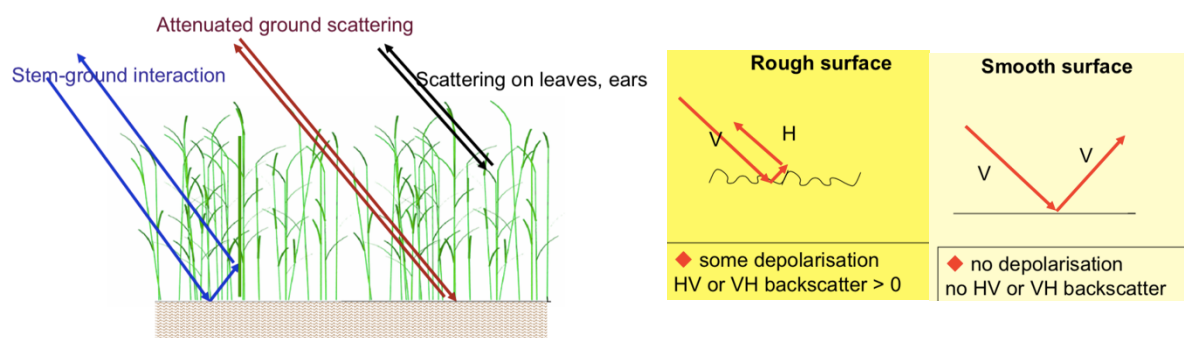
Figure 5: Backscatter mechanisms on a crop canopy. Bouncing rays are subject to a change in polarisation. When a vegetation cover is developing, the interaction with the ground surface is attenuated. An increasing in roughness causes a difference in VV/VH ratio.

1.4   Use of with Artificial Intelligence

Artificial intelligence (AI) can be used for many reasons and many algorithms can be built to solve the same problem (Géron, 2017). In principle this technique is based on programming computers to learn from data, or, better, giving them the ability to learn without being explicitly programmed for. Many complex decisions (like differentiate between a dog and a cat, a maize and a wheat field, automatic drive etc.) can be broken down into small steps, so that each step is simple and it can be translated into a computer program. Being able to rely on experience, AI helps a machine to recognize objects that it was never confronted with before, especially when the amount of data is increasing in time, like constantly updating datasets.

Machine Learning (ML) is a field of AI and Deep Learning (DL) is a subset of ML. ML, in turn, can be broadly divided into three subcategories depending on the training method: supervised (where input data are labelled), unsupervised (where data do not

need labels), and reinforcement learning (which works on a system of reward and punishment feedbacks).

There are many algorithms for each of these subsets. In particular, examples of Supervised Learning are Regression, Decision Tree, Random Forest (RF), K-Nearest Neighbours (KNN), Support Vector Machine (SVM) etc.
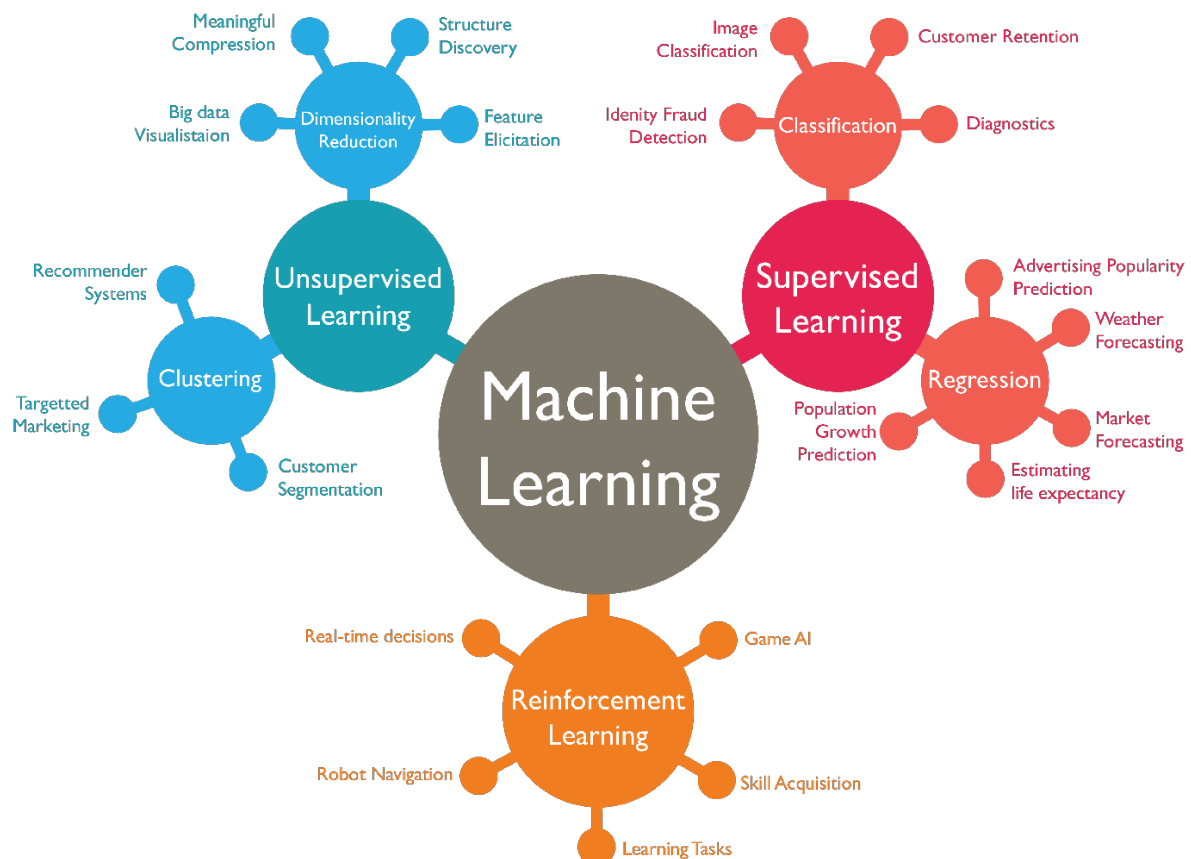
Figure 6: Machine Learning categories with few applications examples.

DL includes all ML algorithms in which the tasks are broken down and distributed into consecutive layers. The more layers there are, the more automatized the entire process is. In DL the algorithms get closer to a network of neurons. Algorithms like artificial neural networks (ANN), convolution neural network (CNN) etc. are used, and sometimes it is even difficult to understand how the machine is effectively learning.

The understanding of which method to use, and the level of complexity it needs, allows to be more efficient during the workflow and towards the achievement of the results. Figure 7 shows how, by increasing the "depth" of the method, an increasing amount of data inputs is needed to reach full performance.
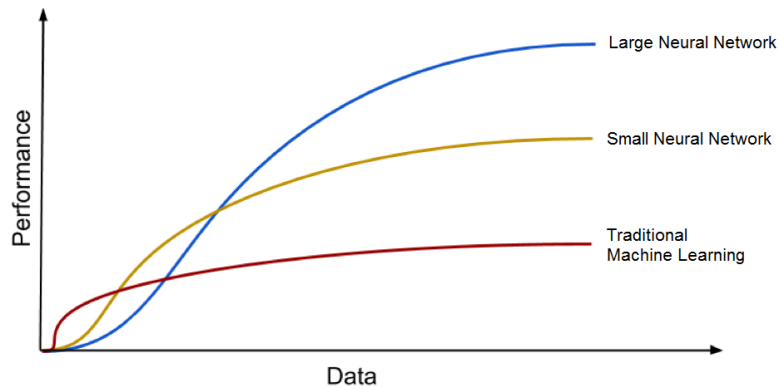
Figure 7: Graph showing how data availability can be related to the method's choice, in order to have higher performances (in terms of computing time, speed, results). Source: deeplearning-academy.com NOTE: graph is representational only and does not depict actual data.

In the remote sensing community, ML is largely used, especially for classification purposes and due to the lack of the big amount of labelled data needed to train the machine. For classification purposes, like identifying differences and collect similar patterns into some given classes, there are few methods that, for their straightforward application, fast computational time, and high accuracy, are more attractive.

Some examples of supervised classification methods used primarily by experts are Support Vector Machine (Mountrakis, et al., 2011) and Random Forest (Breiman, 2001). The second one is much more used for crop mapping due to the higher accuracy it gets (Belgiu & Drăguţ, 2016; Griffiths, et al., 2019; Noi & Kappas, 2017; Sitokonstantinou, et al., 2018). The Random Forest Classifier (see Figure 8) represents an approach that has proven to perform fast and accurate for large features space and variegated training datasets. RF generates multiple decision trees (the number is chosen by the programmer) by randomly drawing samples with replacement from the training data and determining the best split at each decision tree node by considering a given value of randomly selected features (Breiman & Cutler, 2003).
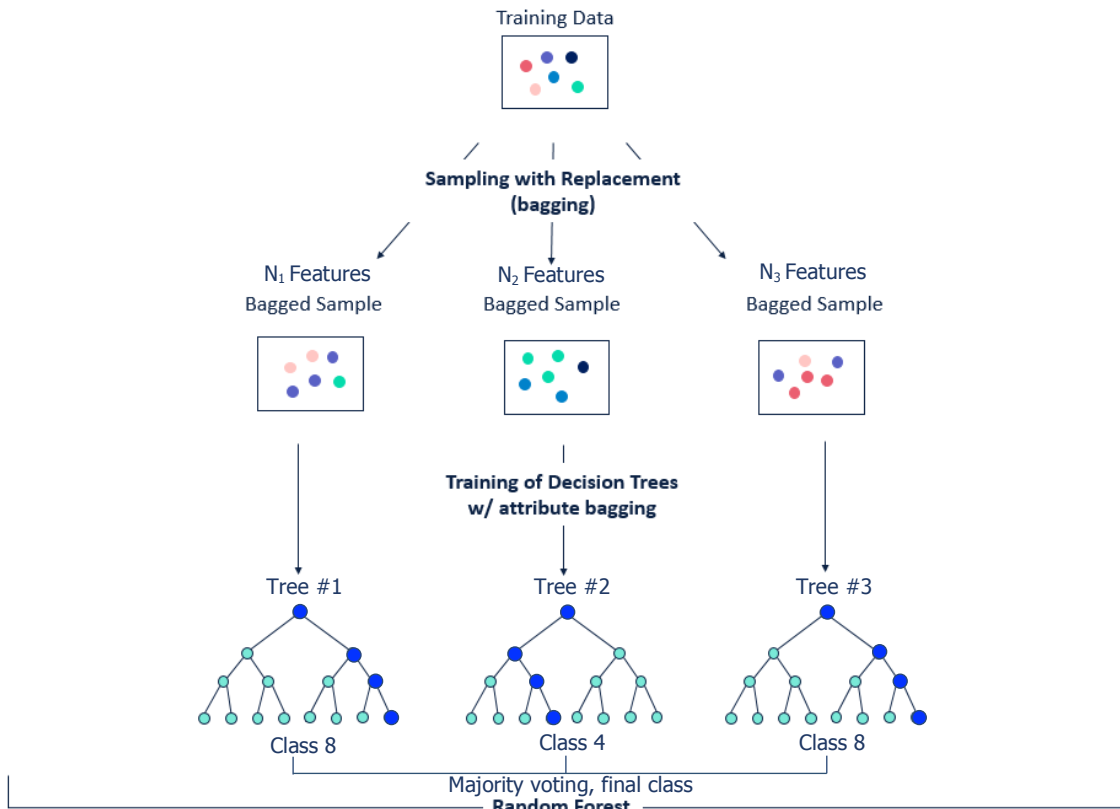
Figure 8: Random Forest Classifier diagram

In order to teach to a machine how and what to learn, a certain amount of data is needed for training and validating the method. A number of examples need to be provided to the machine. This dataset of examples (named labelled data) has to be built by humans (or it is automatized if the algorithm is sufficiently complex for unsupervised learning or DL) and is considered the truth compared to what the machine tries to predict. In Remote Sensing, this dataset is also called groundthruth set and it is divided into a trainset and a testset.

In ML, the process is generally divided into two main steps:

1) Training and validation of the model.

   The training session consists in teaching to the computer to understand how to associate a picture (or an email, or any object) to a certain label that characterizes the class. For example, the information "this is a dog" is linked to a picture of a dog.

The process lies on the ability to analyse which features all the dog's pictures have in common and how they differ from other classes.

This step includes the change of some hyperparameters of the classifier, or a change in the algorithm, in order to reduce the amount of errors. A process called cross validation is commonly used for this purpose: the trainset is divided into complementary subsets, on which to separately run the model with different hyperparameters, testing it on the remaining parts. The model that performs the best is trained over the whole trainset and the generalized errors are measured on the test set. In RF classifier the cross validation is somehow included already in the algorithm itself. Compared with other ML methods, the function Out-Of-Bag (OOB) uses approximately two-thirds of the training samples in the training model. The remaining samples (out-of-bag) are used to validate the model. Hence RF will not overfit the data and the OBB accuracy is an overall unbiased estimation.

2) Test and accuracy measurement.

Once the machine has understood common patterns of a class, its performances are tested. The model is fed with the images of the testset (with no information on the true classification) and it tries to classify them. This process allows to understand if the model generalizes well over data that were never seen before. The programmer knows the actual class and can understand the accuracy of the model, normally expressed as confusion matrixes, as shown in Figure 9. The accuracy is higher if more objects are correctly predicted. Accuracy (ACC) is calculated as the number of all correct predictions divided by the total samples. The best accuracy is 1.0 (100%), whereas the worst is 0.0. Since ACC is not a useful metric of a classification system when there are some classes that are not equally represented, other measurements are used in addition. Kappa statistic, for example, compares the accuracy of the system to the accuracy of a random system.

		Actual class	
		Cat	Non-Cat
Predicted class	Cat	True positives (TP)	False positives (FP) (Type error I)
	Non-cat	False negatives (FN) (Type error II)	True negatives (TN)

True Positive Rate, sensibility

$$TPR = \frac{TP}{TP+FN}$$

True Negative Rate, selectivity

$$TNR = \frac{TN}{FP+TN}$$

$$\text{Accuracy (ACC)} = \frac{TP+TN}{TP+FN+FP+TN}$$

$$\text{Random Accuracy (RA)} = \frac{(TN+FP)\cdot(TN+FN)+(FN+TP)\cdot(FP+TP)}{(TP+FN+FP+TN)^2}$$

$$\text{Kappa statistic (K)} = \frac{ACC-RA}{1-RA}$$

$$\text{Error rate (ERR)} = \frac{FP+FN}{TP+FN+FP+TN}$$

Figure 9: Example of confusion matrix, over an example of identification of cat images. True Positive Rate, True Negative Rate Accuracy, K coefficient, Error rate are some commonly used metrics to identify the performance of a classification model.

1.5 Pixel based vs Parcel Based classification

Two types of classifications can be used: pixel-based and parcel-based. Pixel-based classification consists in using information from single pixels of the image to detect common patterns. This type of analysis often lead to misclassifications due to the land cover's spectral variability, bare soil background reflectance, atmospheric effects and mixed pixels present at the boundaries between parcels. With the increasing amount of easily accessible high resolution satellite images (with pixel size smaller than the parcel area), the attention on how to pre-process images is growing. Post-classification processing is also an important step in improving the quality of classifications when using information based on single pixels (i.e. reducing "salt and pepper" effect) (Lu & Weng, 2007; Sitokonstantinou, et al., 2018; Sicre, et al., 2020). Grouping pixels into delimited objects before classification can overcome these problems. The remote sensed image is coupled with vector parcel boundaries layer. The classification is made by unifying multiple pixel information into one single list of features statistically depending on the delimited area. This process can perform a more reliable and accurate classification (Peña-Barragán, et al., 2011).

20

Three assumptions are made when using a parcel-based classification on multi-temporal series:

- the boundaries will not change in time;
- during the period of study only one crop type is grown in the parcel;
- in the same parcel there will be no subdivision of the field for multiple crops.

For this study, these points are considered satisfied. First, it is very unlikely that field boundaries would change in such a short period of time as two successive years. Second, crops with sufficiently long phenology have been chosen for the classification. Third, the parcels boundary layer, with which the model is built, corresponds exactly to the specified crop cultivated inside.

1.6 Study goal

The goal of this study is to create a model that identifies mayor regional crop types from satellite images, and apply it to The Netherlands.

The model combines time series of radar and multispectral images from Sentinel 1 and Sentinel 2 satellites, respectively. The combination of multi-sensors and multi-temporal images enhances the classification over areas with relevant cloud coverage through the year and gives more information about growth peculiarities.

The groundtruth dataset comes from the Dutch Basic Registration of crop Plots (BRP). This crop-parcel-layer is an element of the Land Parcel identification System (LPIS) that the Country uses for the Common Agricultural Policy (CAP), the EU support for agricultural income. The study relies on a parcel-based classification.

The scripts are entirely implemented in the Google Earth Engine, combining one year of satellite data (from autumn 2017 to autumn 2018) in a multiband image to feed in the classifier. Standard deviation and several vegetation indexes were added in order to have more variables for each 12-day-median image composite. The processing pays particular attention to the time variability of the mean values of each field. This provides useful information both for understanding differences among crops and the variability over the phenology of the plant.

The machine-learning algorithm used in the thesis is a Random Forest Classifier. If sufficient accuracy is achieved, further application of the model on the crop in the following years will be examined.

The specificity of the study is to apply state of the art in remote sensing processes using only open source datasets and tools, over a cloud prone region. To enhance the performance, this work has also created a time series profile, able to identify similar field spectra for the same crop class throughout the agricultural year.

Providing a free crop mapping method is a valuable step not only for control and monitoring, but also for quantifying residual biomass, organize better logistic arrangements, and many other applications.

2 Study area and Data

2.1 Study area

The Netherlands has a surface of 41,873 km². It lies between 50° and 54° N in latitude, and between 3° and 8° E in longitude. The area can be considered flat, with about 26% of the country located below the sea level and 50% one meter above the sea level. The maximum altitude is 322m a.s.l. in the southeast (see Figure 10.c) (Blom-Zandstra, et al., 2009).

The climate is classified as Cfb according to the Koeppen-Geiger system, indicating a mild, marine climate with warm summers and no dry season. Average yearly temperature between 1981 and 2010 span from 9.6°C in the northeast to 11.1°C in the southwest. The amount of dry days during the year is on average higher in the south than in the north, with opposite trends for the relative humidity, that is higher in winter in the north (90%) and lower in summer in the south (76%) (Koninklijk Nederlands Meteorologish Instituut, 2011). Average yearly rainfalls are mostly concentrated in the central and northern part of the country (975 mm/y). Research shows that the intensity of precipitations has increased in the last decades due to global warming (KNML, 2020). The increasing sea level (by about 2 mm/year) brings salinity stress to the soil. Together with the increase in precipitation extremes, this may cause relevant effects to the agricultural and natural systems in the medium-long period.

The country was formed by delta deposits from the Rhine and Meuse rivers. Soil moisture, shown in Figure 10.a, is correlated to the soil types (Figure 10.b) and the elevation (Figure 10.c). Clay soils will be more affected by floods while the more elevated, south-eastern, sandy soils will suffer more in the dry seasons. In particular, the annual average temperature was 10.9 °C in 2017, 11.3 °C in 2018 and 11.2 °C in 2019. These years are all within the last 6 consecutive warmest years. Precipitations were 862 mm/y, 607 mm/y and 783 mm/y, respectively. All the three last years were warmer and dryer than the climatic mean but relatively similar between them.
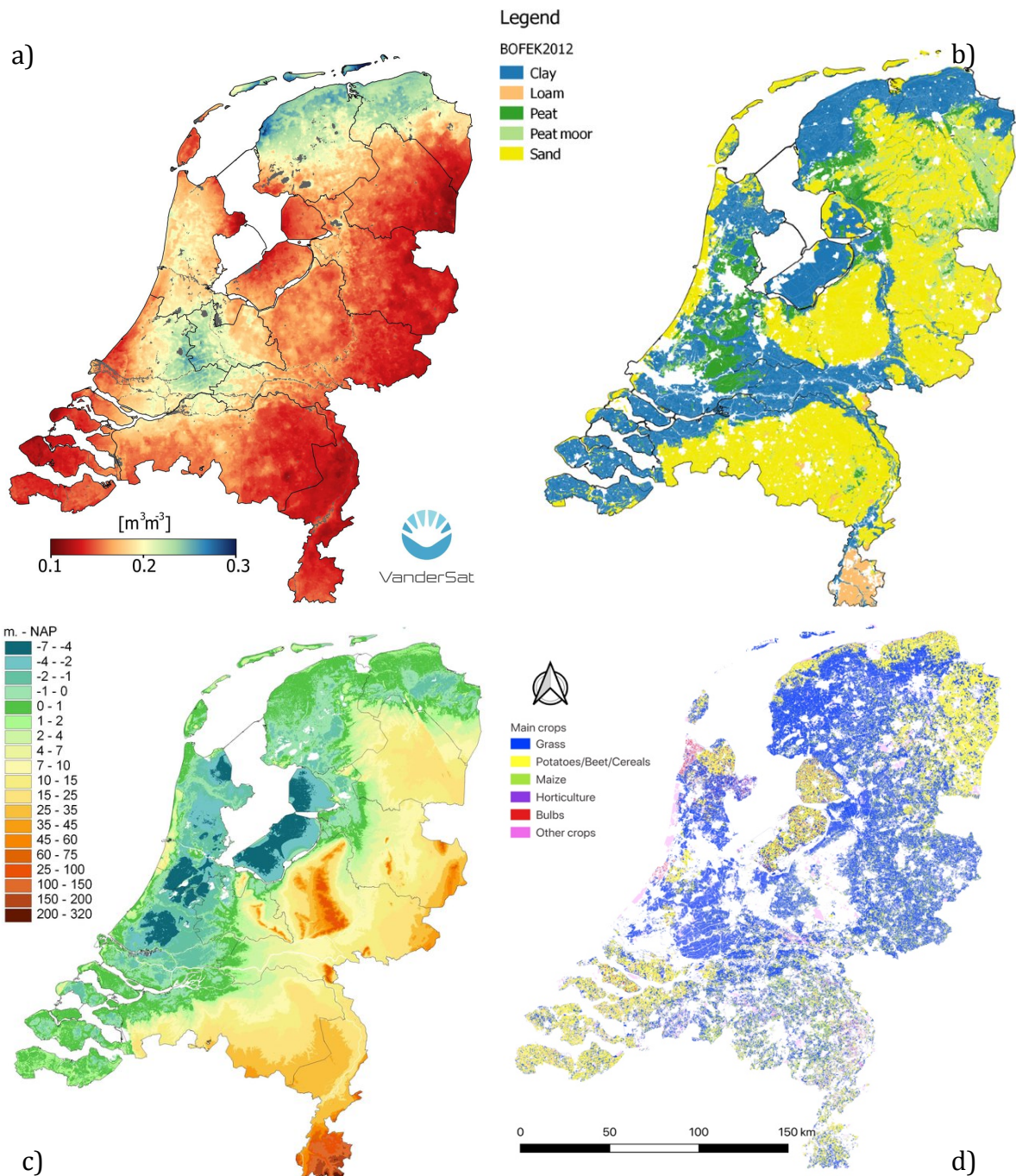
Figure 10: Overview of study area. a) Soil moisture (Source: ESA's Soil Moisture and Ocean Salinity mission). b) Soil types, based on the Soil Physical Map of the Netherlands (Kroes, et al., 2018; Wosten, et al., 2013). c) Altitude in m a.s.l. (Blom-Zandstra, et al., 2009). d) Simplified basic registration of crop plots. The boundaries of the agricultural plots are based on the Agricultural Area of the Netherlands (AAN) (Source: PDOK, geoportal of the Netherlands).

Even if the agricultural sector of the Netherlands has developed vertical greenhouses, more than the 50% of the total surface is still used for open cultures.

Excluding grasslands (50% of the total), the main open-air cultivated crops are maize (25%), potatoes (19%), wheat (13%), sugar beet (10%), barley (4%) and vegetables. It is interesting to insert in the list bulbs and flower fields: blooming crops have a strong fingerprint signals for remote sensors. Phenology development of plants in the region has been taken into consideration for choosing the more representative and relevant crop classes for this study. More details about major crop calendar in this territory are shown in Table 1, extrapolated from the 2018 "Water Guide Agriculture Report" by the Foundation for Applied Water Management Research in the Netherlands (STOWA).

Table 1: Crop calendar for open cultures in which the growth season takes place. Crop development (green) begins after preparatory work has taken place (plowing and seeding) indicated with orange. For winter wheat and tulip, an extra long period is taken into account before crop development actually takes place, as a result of vernalization (influence of growth processes by cold). Source: (STOWA, 2018)

	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct
Grass	growth	growth	growth	growth	growth	growth	growth	growth	growth	growth	growth	growth
Maize						sowing	growth	growth	growth	growth	growth	harvest
Sugar beet	harvest				sowing	growth	growth	growth	growth	growth	growth	harvest
Barley, summer					sowing	growth	growth	harvest				
Potatoes, consumption					sowing	growth	growth	growth	growth	harvest		
Potatoes, seeds					sowing	growth	growth	growth	harvest			
Wheat, winter	sowing	sowing	growth	growth	growth	growth	growth	growth	harvest			sowing
Onion, sowing					sowing	sowing	growth	growth	growth	growth		
Tulip	sowing	sowing	growth	growth	growth	growth	harvest					sowing
Lily								sowing	growth	growth	growth	harvest
Leeks	growth	harvest						sowing	growth	growth	growth	growth

Legend:
sowing — orange
growth — teal
harvest — red

2.2 Materials and Data

Three different sources of data have been used for this study: SAR images from Sentinel-1, MSI images from Sentinel-2, both provided by the Google Earth Engine Data Catalog, and a vectorised parcel boundary layer provided by the LPIS National system.

2.2.1 Field Data

BRP is a dataset freely provided by the national geoportal. The boundaries of the agricultural plots are based on the Agricultural Area of the Netherlands (AAN). The

owner of this information is the "Ministry of Economic Affairs and Climate - Netherlands Enterprise Agency", which uses it for planning the CAP.

This dataset is updated annually at the 15th of May. The user of the parcel must indicate his crop plots annually and indicate which crop is grown on the relevant plot. Every polygon is geo-localized and associated with the label of the crop type currently cultivated in the area. The total amount of parcels corresponds to about 774000 plots, divided into 312 classes, where the majority are classified like grassland (see Figure 10.d and Figure 11).

Figure 11: BRP (Basic Registration of crop plots) dataset overview.

From the BRP dataset a certain amount of crop classes was selected to build the classifier. Following the STOWA Report, it has been decided to map only the more common crop types in the country due to their higher relevance for the final crop mapping of the region. 22 classes were used: Potatoes, consumption; Potatoes, seeds; Potatoes, starch; Corn, corncob mix; Corn, energy; Corn, grain; Corn, cut; Corn, sugar; Wheat, winter; Wheat, summer; Sugar beet; Winter barley; Summer barley; Tulip, flower bulbs and tubers; Lily, bulbs and tubers; Onions, sowing; Onions, silver; Leek, winter, production; Leeks, winter, seeds and propagating material; Leek, summer, production; Leeks, summer, seeds and propagating material. All classes not interesting for this study are classified in a final class named "Others".

The BRP dataset is used to construct the groundtruth dataset for the analysis.

2.2.2   The Google Earth Engine

Cloud-computing based calculations and computer capabilities have been improving in recent years. Big companies like Google, Amazon and Microsoft have introduced cloud-computing tools. Google, for example, provides a constantly updated catalogue of satellite imagery for cloud-computing on a global scale on the GEE (Fuyou, et al., 2019; Gorelicka, et al., 2017)

GEE is a platform for scientific analysis and visualization of geospatial datasets, for academic, non-profit, business and government users. Based on user-developed algorithms, it combines an expanding catalogue of open source satellite imagery with global-scale analysis capabilities and Google "makes it available to detect changes, map trends, and quantify differences on the Earth's surface". (Google, 2020)

On this platform, many different dataset sources are made available. Not only pre-processed and constantly updated satellite imagery (including, but not only, the Copernicus Programme) but also products from large scale mapping validated from acknowledged institutions (e.g., global precipitation measurements, USGS National Land Cover Database, etc.). Images fed into the GEE are pre-processed to facilitate fast and efficient access.

Furthermore, in order to allow fast visualization during algorithm elaboration, pyramids of reduced-resolution tiles are created for each image and stored in the tile database. This power-of-two downscaling enables having data ready at multiple scales without significant storage overhead, and aligns with the common usage patterns in web-based mapping (Gorelicka, et al., 2017).

Very recently, increasing research studies have been built using GEE for different purposes (Carrasco, et al., 2019; Fuyou, et al., 2019). It assembles a conventional ML method and provides a visual user interface. The Earth Engine Code Editor (see **Error! Reference source not found.**) is a web-based Integrated Development Environment (IDE) for the Earth Engine JavaScript Application Programming Interface (API). Functions ready to use, and fast to understand scripts are shown and described. It allows importing and exporting of materials in the main geospatial file formats. There is an intuitive visualization board where to observe mapped results.
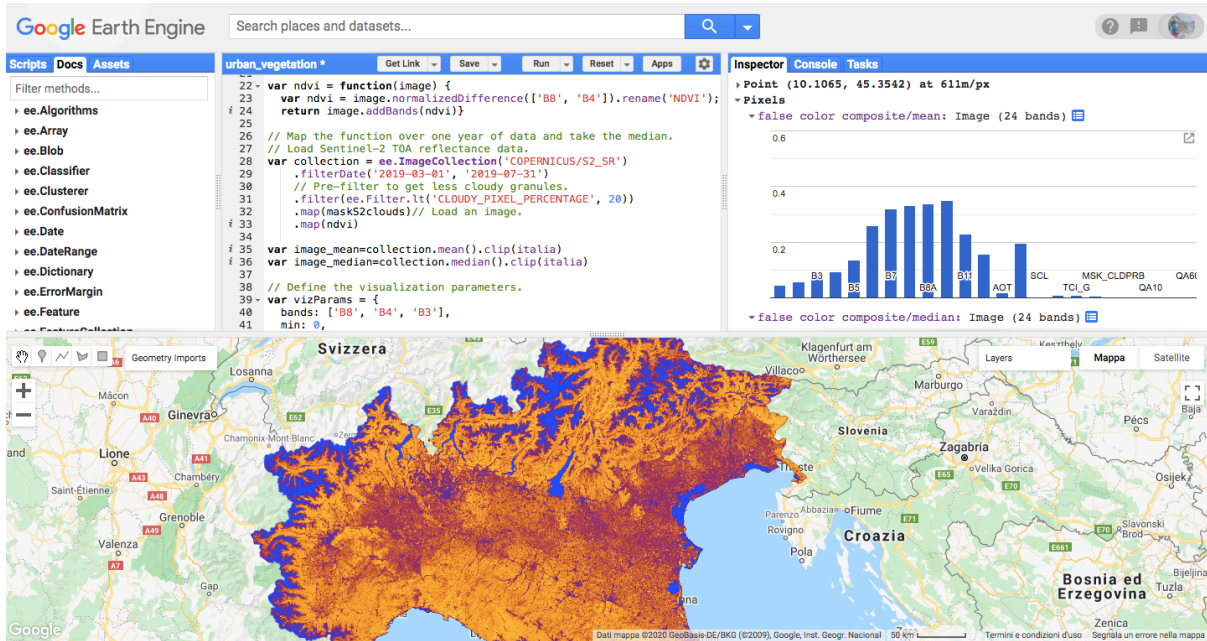
Figure 12: Google Earth Engine Code Editor, the Earth Engine interactive development environment.

Despite of the great capability to rapidly compute tasks and to visualise results, one GEE feature has to be highlighted: scaling errors.

A given amount of computing power and of memory is assigned to each user account in order to introduce some control to the system. Although a script may be a valid JavaScript, without logical errors, and it may represent a valid set of instructions for the server in parallelizing and executing the computations, the resulting objects may be too big or take too long to compute. In this case, an error indicating that the algorithm can't be scaled is given. These errors are generally the most difficult to diagnose and to resolve. Examples of this type of error include (Google, 2019):

- "Computation timed out"
- "Too many concurrent aggregations"
- "User memory limit exceeded"
- "Internal server error"

A full community of users and Earth Engine developers are very active and provide a prompt response on forums and online networks of professionals, offering support with errors, as well as many advices for reducing the development effort.

2.2.3   Sentinel1: SAR sensor

The Sentinel1 mission (see Figure 13) involves a constellation of two polar-orbiting satellites (A and B), operating day and night performing C-band (centre frequency 5.405 GHz) synthetic aperture radar imaging. This enables to acquire imagery regardless of the weather.
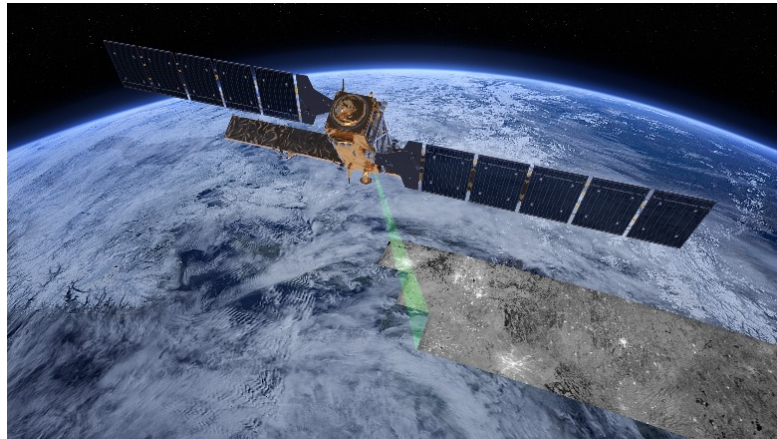
Figure 13: Sentinel1 satellite from Copernicus Programme. Source: ESA

The satellites flight at 693km of altitude, with a near-polar sun-synchronous orbit. Each satellite is potentially able to map the global landmasses in the Interferometric Wide (IW) swath mode once every 12 days, in a single pass (ascending or descending). The two-satellite constellation offers a 6-days exact repeat cycle at the equator. Since the orbit track spacing varies with latitude, the revisit rate is significantly greater at higher latitudes than at the equator. Over the Netherland territory it overlaps every day in some areas.

Interferometric-Wide swath mode is one of the four different operational modes of acquisition of the signal. It is the main acquisition mode over land and satisfies the majority of service requirements. It acquires data with a 250 km swath at 5 m by 20 m spatial resolution (single look, scene). The incidence angle is 20° to 45° and the polarization modes are VV, VH, HH, HV.

Radar images are built from the intensity of the backscattered signal detected by the sensor. In an active system, the sensor, which is a transceiver, radiates electromagnetic power. It exploits the phenomenon of diffusion (scattering) that occurs when an object is hit by an electromagnetic wave. Then, the energy is re-radiated with a spatial

distribution that depends on the shape and material of the target (dielectric constant of the medium and roughness), as well as on the system parameters (e.g., polarisation).

Since clouds do not affect radar acquisition, it does not need to filter data. From November 2017 to November 2018 from 113 up to 300 passages per pixel have been counted (see the histogram in Figure 14(left)).
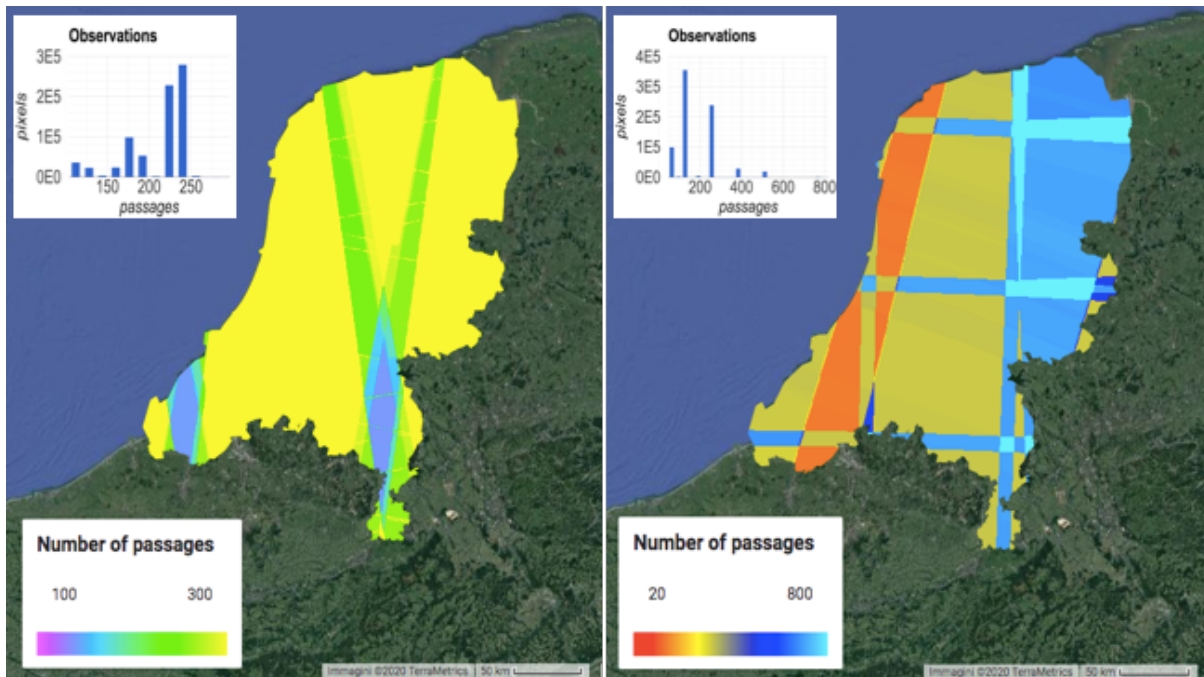
Figure 14. Number of observations from Sentinel1 (left) and Sentinel2 (right) in the year of acquisition (1st of Nov 2017-1st of Nov 2018) on the Netherlands. Each imagery absolute frequency per pixel is shown on the occurrence histogram (this representation has 300m resolution per pixel.).

The GEE collection of Sentinel-1 repository includes the processed S1 Ground Range Detected (GRD) scenes, to generate a calibrated, ortho-corrected product. The imagery is daily updated and new assets are provided by ESA (the data provider) within two days after they become available. GRD scenes have up to 10 meters resolution, 4 bands combinations (corresponding to single or dual polarization) and three instrument modes.

GEE uses the following pre-processing steps (as implemented by the Sentinel-1 Toolbox) to derive the backscattering coefficient in each pixel: thermal noise removal, radiometric calibration and terrain correction. The backscattering coefficient is converted to decibels by log scaling ($\sigma°_{dB}=10*\log_{10}(\text{intensity})$) because it can vary by several orders of magnitudes. The intensity of backscattering measures whether the

radiated terrain scatters the incident microwave radiation away from (dB > 0) or towards (dB < 0) the SAR sensor.

2.2.4 Sentinel2: MSI, multispectral instrument

The Copernicus Sentinel-2 mission (see Figure 15) includes a constellation of two polar-orbiting satellites placed in the same sun-synchronous orbit, phased at 180° to each other. It aims at monitoring variability in land surface conditions, and its wide swath width (290 km) and high revisit time (10 days at the equator with one satellite, and 5 days with 2 satellites under cloud-free conditions which results in 2-3 days at mid-latitudes) will support monitoring of Earth's surface changes. It systematically acquires data over land and coastal areas from latitudes between 56° S and 84° N.
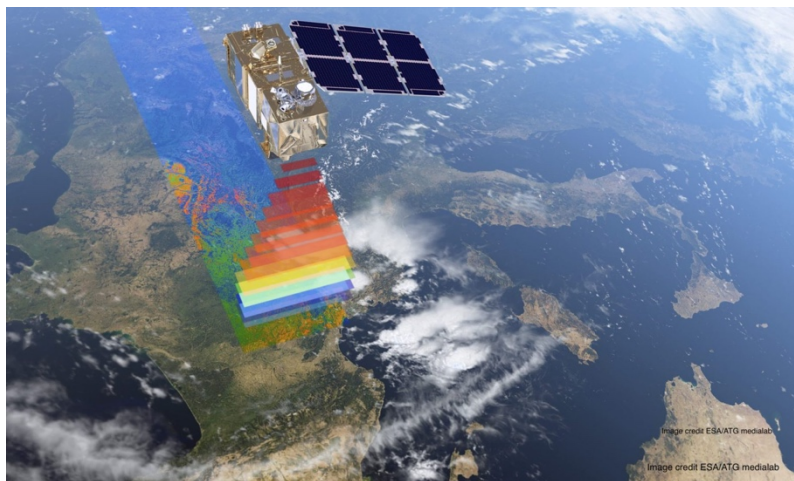
Figure 15. Sentinel-2 satellite from Copernicus Programme. Source: ESA

Sentinel2 is the sharpest free satellite imagery available today. The MSI is an innovative high-resolution multispectral imager. MSI is a passive instrument: the sensor is a receiver and measures the spontaneous emission of electromagnetic energy by the observed object. It features 13 spectral bands (as shown in Figure 3) including 3 bands for atmospheric corrections, spanning from the VNIR (Visible and Near Infrared), to the SWIR featuring 4 spectral bands at 10 m, 6 bands at 20 m and 3 bands at 60 m Spatial Sampling Distance (SSD). The signal is not detectable through all the electromagnetic spectrum due to atmospheric interferences. Within the Sentinel-2 acquisition range there are two windows in the SWIR and one that includes the NIR and VIS ranges. Each of these bands have their specific proposes due to their different interaction with the medium (the spectral band specification is shortly resumed in annex 1).

The GEE Sentinel-2 imagery is provided at a pre-processing level 2A[1], computed by running *sen2cor.* This is a processor generated by the ESA Payload Data Ground Segment and performs the conversion from the level 1C Top-of-Atmosphere (TOA) product to the orthorectified, atmospherically corrected, Surface Reflectance (level 2A-SR). Level 1C TOA comes from the precursor levels 1B, 1A and 0 (compressed raw data). These coarse previous levels are not disseminated to users but are necessary phases to elaborate the signal transmitted directly by the satellite into usable information for the final applications.

At the 2A-RS level, in addition to the bands described above, ESA provides some supplementary information as, for example, a bitmask band with cloud mask data (AQ60) (see more details in section 3.1.3).

Figure 14 (right) shows how the images are distributed on the territory per pixel (300m of resolution). In some areas there are about 60 passages along the year of the study, concentrated in the southwest, while in other parts of the country the number rises up to almost 770. These observations are affected by cloud coverage hence a filtering process has to be done before further analysis.

---

[1] From the mid-March 2018, the Level-2A became an operational product, beginning with coverage of the Euro-Mediterranean region. Global coverage started in December 2018.

3   Methods

The imagery analysis has been divided into five parts:

a) Filtering of the reference field data, choosing a certain amount of polygons per class, and of the satellite data (anomalous values, clouds) with the addition of further vegetation indexes.

b) Filling the gaps in the optical images and creation of an image composite, piling radar and optical images at 24-days interval along the year.

c) Creation of training and test sets by by intersecting subsets of groundtruth dataset with all pixels' information of the images throughout the year.

d) Training of the RF classifier.

e) Evaluation of the accuracy and validation of the model.

3.1   Preliminary data processing

3.1.1   BRP dataset

In order to create a certain amount of groundtruth data, representative and not too heavy for computational work, the BRP datasets were filtered. On the GEE, a script was developed able to select from the shapefile a collection of features ("ee.FeatureCollection") of up to 500 polygons in each crop class, out of a total 774822 polygons. This process is commonly named "stratified sampling". Those polygons must represent an area of more than 100m^2 each, because a field size sufficiently big for pixel statistical processing is needed. The polygons are chosen in randomly selected positions and the seed changes for every class. Random sampling for spatial dataset is rather critical when classification approaches are involved. The closer the observations are located to each other, the more similar they are. This can produce overoptimistic results (overfitting) when the test set contains observations which are somewhat similar to observations in the training set. However, this is more relevant when a pixel-based classification is conduced (Schratz, 2018).

In order to reduce the field edges interference on the signal response, every polygon was buffered, i.e. areas within 10 m from the perimeter were discarded. The list of

training and test polygons was than saved to be used further on in the code (Export.table.toAsset exports a FeatureCollection as an Earth Engine asset).

Due to GEE limitations of computing power and of memory (see section 2.2.2), there are several limitations on the size and shape of GEE table assets, which are computable list of features (points or polygons) containing given variables (named properties) and georeferenced geometries. These tables can be exported or imported as well as manipulate on GEE freely once they respect certain constraints:

- maximum of 100 million features,
- maximum of 1000 properties (columns),
- maximum of 100,000 vertices for each row's geometry,
- maximum of 100,000 characters per string value.

Annex 2 shows the number of polygons per class fed into the classifier.

3.1.2 Sentinel1 data

VV and VH dual-band cross-polarization (vertical transmitted and horizontal received) were chosen on the Interferometric Wide (IW) swath mode by Sentinel1. Another band was successively added, called the Vegetation Index (VI), (Hosseini & McNairn, 2017) in order to enhance the amount of information per pixel. Before any calculation, backscattering values were converted into the natural scale.

VI is computed as:

$$VI = VV - VH \tag{Eq. 1}$$

where VV is the backscatter intensity ($\sigma°$) of vertically transmitted and received beams, and VH is $\sigma°$ of vertically transmitted and horizontally received beam.

Before feeding these data to the algorithm, a filter for cleaning backscattering intensity is implemented. A mask that deletes values lower than -30dB is applied, reducing some anomalies in the signal (see an example in Figure 16).
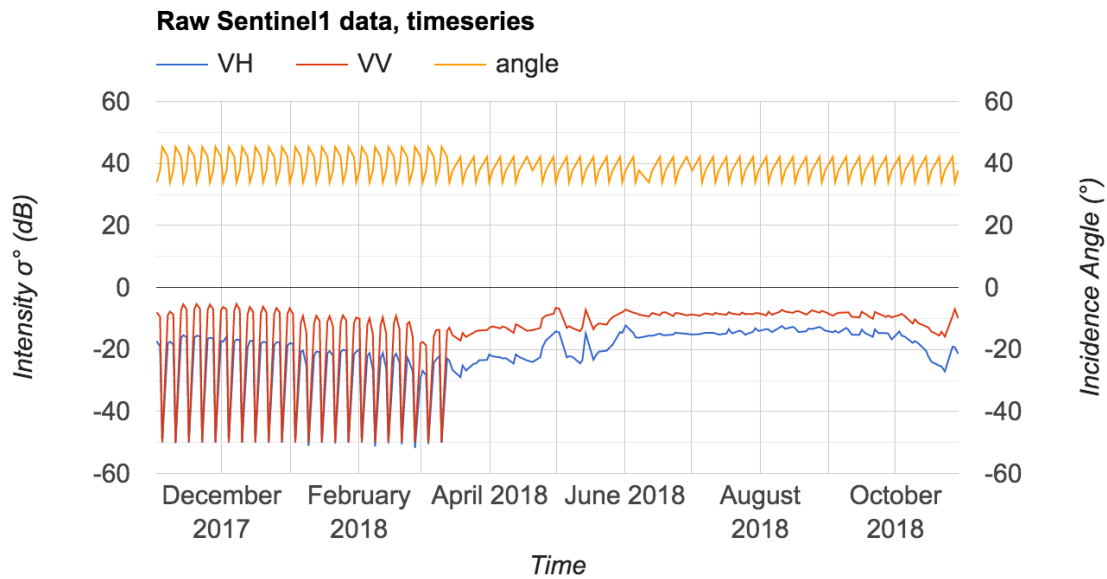
Figure 16. Example of the time variability of the backscattering signal of one pixel of a Sentinel-1 image. The yellow curve shows the approximate viewing incident angle (from 20° to 45°). The red and blue curves show VV and VH dual-band cross-polarized dB intensity.

Another cause of non-valid data is the overlaying zones among different Sentinel-1 images. While for optical images an overlapping means higher occurrence of observation, the difference of acquisition angle of radar swats can compromise the overall backscattering value. An increase of noise is expected due to non-aligned data (Fuyou, et al., 2019). A function that computes the "geometry erosion" of Sentinel1 GRD scenes, reduces this effect eliminating images' edges. In order to prepare the SAR images for the classifier, all σ° values were converted into log scale: the ML algorithm can understand better differences if values are in this format.

3.1.3 Sentinel2 data

For MSI images, the following bands were included in the analysis: B2, B3, B4, B5, B6, B7, B8, B9, B11, B12. Moreover, two other indexes, Bare Soil Index (BSI) (Wanhui, et al., 2004) and Normalized Difference Vegetation Index (NDVI) (Tucker, 1979; Hosseini & Saradjian, 2011) were added to the information of each pixel of the optical image. These indexes are largely used in remote sensing to detect and enhance anomalies in vegetation changes. NDVI (Eq. 2) is used due to its capability to determine the photosynthetic performance of plants, because healthy canopy absorbs more red light (VIS) than NIR light. BI, derived by SWIR band

BSI is defined as:

$$\text{BSI} = \frac{(\text{SWIR}_2+\text{RED})-(\text{NIR}+\text{BLUE})}{(\text{SWIR}_2+\text{RED})+(\text{NIR}+\text{BLUE})}$$), is sensitive to the water content in the soil.

$$\text{NDVI} = \frac{\text{NIR}-\text{RED}}{\text{NIR}+\text{RED}} \hspace{3cm} \textbf{(Eq. 2)}$$

where RED and NIR parameters correspond to intensity in B4 and B8 respectively.

BSI is defined as:

$$\text{BSI} = \frac{(\text{SWIR}_2+\text{RED})-(\text{NIR}+\text{BLUE})}{(\text{SWIR}_2+\text{RED})+(\text{NIR}+\text{BLUE})} \hspace{2cm} \textbf{(Eq. 3)}$$

where SWIR$_2$, RED, NIR and BLUE parameters coincide with B12, B4, B8 and B2 intensities.

As optical images are sensitive to cloud coverage, a filtering process based on thresholds on the percentage of pixel with clouds (maximum 80% per scene) was applied. Then, a cloud mask was applied to eliminate bad observations. To elaborate this mask, an artificial bitmask band (QA60, 60m of resolution) provided by GEE was used.

Bitmask for QA60 (Google, 2020):

- Bit 10: Opaque clouds
    - 0: No opaque clouds
    - 1: Opaque clouds present
- Bit 11: Cirrus clouds
    - 0: No cirrus clouds
    - 1: Cirrus clouds present

This masking method is not perfect (not all the clouds are identified and flagged on the QA60 band) and many developers are focussing now on elaborating models able to perform more precise cloud detection. For the purposes of this work it has been decided to not investigate further on this topic, using instead the solution provided directly by GEE at a 2A-SR level of pre-processing.

3.2 Multitemporal analysis: gap filling (spatial and temporal filling)

A multitemporal series of images was created in order to build the complete list of information to associate to every polygon. The 12-day composite is a stack of images

where each of them is created by statistically associating values in a 24-days acquisition period (12 days before and after the measuring day). Every pixel should have at least 2 radar and 1 optical acquisition per sample in clear conditions.

In this study, the median values of every band of every pixel, in addition to their standard deviation, were employed. Moving median values are more recommended than mean values for remote sensing. Statistically, the moving average is optimal for recovering the underlying trend of the time series when the fluctuations about the trend are normally distributed. However, if the fluctuations are assumed to be Laplace distributed[2], the moving median is statistically better. Simple statistics of time series are commonly used, because representative values can be calculated regardless of the length of the series or the lack of complete data (Fuyou, et al., 2019).

The operation on GEE for the median applied over the image collection of every moving 24 days sampling period is ee.imageCollection.median(). It reduces an image collection by calculating the median of all values at each pixel across the stack of all matching bands. Bands are matched by name.

In order to have full coverage for every pixel of all multi-sensor images some gap filling was applied. For all the Sentinel-2 data, which were masked due to cloud presence, a mosaicking procedure was implemented. As explained in Figure 17, when an image is masked by clouds, it is substituted with a patch, recovering the lack of data for that image. The patch is created by using the median values of a longer period, 80 days. Empirical tests suggested that 80 days are sufficient to have a good coverage over the territory: shorter periods would have caused lack of information in cloudy conditions. Obviously, this method allows gap filling (spatial and temporal) when periods of cloud coverage (detected with the frequency of satellite passages) are not longer than 80 days.

---

[2] Laplace distribution represents the maximum entropy probability distribution. According to the principle of maximum entropy, if nothing is known about a distribution except that it belongs to a certain class (usually defined in terms of specified properties or measures), then the distribution with the largest entropy should be chosen as the least-informative default. The motivation is twofold: first, maximizing entropy minimizes the amount of prior information built into the distribution; second, many physical systems tend to move towards maximal entropy configurations over time. Source: (Lawrence, 2013).

Figure 17: Gap filling method over long-lasting clouds for producing the time-series image composite. The GEE function *.mosaic()* allows merging patches of different overlapping images over the masked area.

Once all the gaps are filled, standard deviation bands were added to the list. Overall, the image composite is made up with 31 images (one every 12 days) in 30 bands, derived by the optical and the radar sensors. In

Figure 18, the name of all the bands are listed.

The classification is done on a matrix where every "image" contains all the time series information of each band (median values of the 12-day composite) spatially averaged over all pixel contained in each crop field (parcel-based classification) and the type of cultivation information (the label on which the classifier is trained and tested).

n	Band name	n	Band name	n	Band name
0	B2	10	NDVI	20	B11_stdDev
1	B3	11	BSI	21	B12_stdDev
2	B4	12	B2_stdDev	22	NDVI_stdDev
3	B5	13	B3_stdDev	23	BSI_stdDev
4	B6	14	B4_stdDev	24	VV
5	B7	15	B5_stdDev	25	VH
6	B8	16	B6_stdDev	26	VI
7	B9	17	B7_stdDev	27	VV_stdDev
8	B11	18	B8_stdDev	28	VH_stdDev
9	B12	19	B9_stdDev	29	VI_stdDev
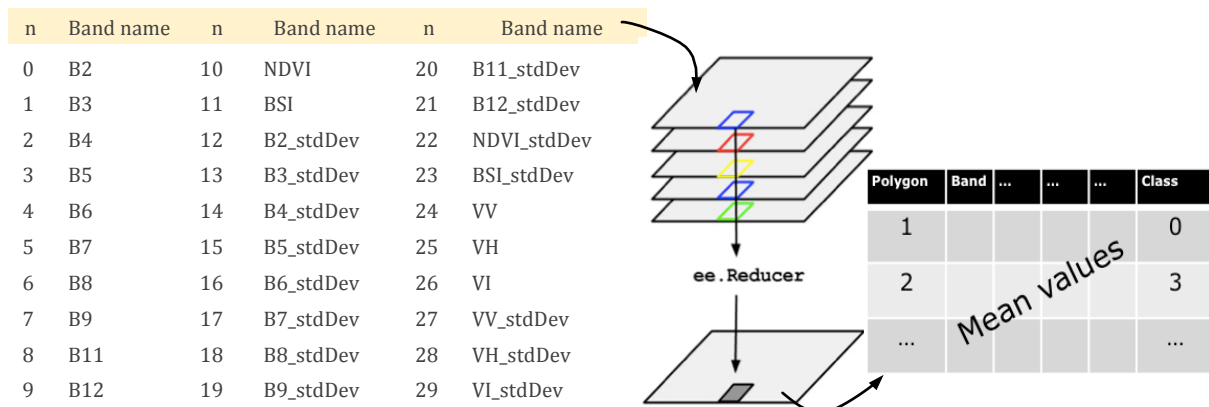
Figure 18: Method of creating the dataset for the classification process. The stack of 12 days interval images composed by the 30 bands is merged into a single image. The ee.Reducer function is applied on this image. The function computes the average over pixels included in each field boundaries of the BRP layer.

The values of each pixel of all the 12-day composite are merged together in a single object with 930 bands (30 bands times 31 images), representing the variability in time and space of each parcel of the Netherland territory. Therefore, with the function

ee.Image.ReduceRegions() to every selected BRP polygon, all mean values of the pixels included in every parcel were computed for each band.

This method converts a stack of images into a collection of objects (or "features" in GEE) each with a given crop class and more than 900 properties associated, for exporting into a local cloud repository. Finally, polygons boundaries were rebuffered with an offset of 10m externally in order to recreate the real fields areas. This amount of information is at the limit of the computing power for a single GEE user. The largest computing time is taken by this step.

3.3 Random Forest Classifier and accuracy assessment

At this stage of the analysis, despite the gap filling procedure, about 5% of polygons still appears with null properties. Too persistent cloud coverage interfered with the calculation of all bands for those areas. These were discarded, because the RFC needs a dataset where all objects have the same amount of properties.

Figure 19: View of polygons subsets: training and test set.

In order to train and test the classifier, the dataset (8335 objects, see Annex 2 for class grouping) was divided into a trainset and a testset (80% / 20% respectively), as shown in Figure 19. The drawing is random and proportioned to the class occurrence in order to maintain the representativeness of the territory.

The RFC can be tuned through several parameters. The most relevant for this study are the number of decision trees (set as 100), the minimum size of a terminal node (set as 10) and the choice of whether the classifier should run in out-of-bag mode (selected as true). Other parameters were left at their default value. The choice of the value of the main parameters, number of trees and number of variables in each node, which determine the running time, was done empirically monitoring the accuracy of the measurements. The value was chosen at a level above which the classification accuracy does not increase significantly.

The function employed is ee.Classifier.randomForest(). This function creates a classifier, which uses the Random Forest algorithm. During training, the accuracy increases. The classifier is considered as trained when the accuracy converges. The training accuracy is computed with the confusion matrix (see section 4.4 and Figure 25). Once the classifier is ready, it is possible to validate the model with the test set, the remaining 1667 objects. Hence, another confusion matrix for the test accuracy is computed. The difference among the train and test accuracy allows understanding if there are over-fitting bias. Moreover, the K coefficient is computed to compare the accuracy of the model to the accuracy of a random system by weighting good predicted features with class abundance.

The classifier performances have been independently assessed for each class.

3.4   Summary of the workflow

The workflow of the study is summarized in Figure 20. At first, the analysis was based on data collecting and preparation, with the merging of all the images and satellite information with groundthruth data. Second, more vegetation indexes were added to the bands already available on GEE. The process of segmentation of those values to fields' areas and analysing the spatial average of medians through time are the core part of this phase. Third, the datasets are fed into an RFC in order to build a model able to detect differences among many plant phenology. The accuracy in classifying the crop classes was then examined in order to evaluate the possibility to apply the model to year 2019. Annex 3 shows links references of the function to create the image composite and the RFC algorithm used for this study.
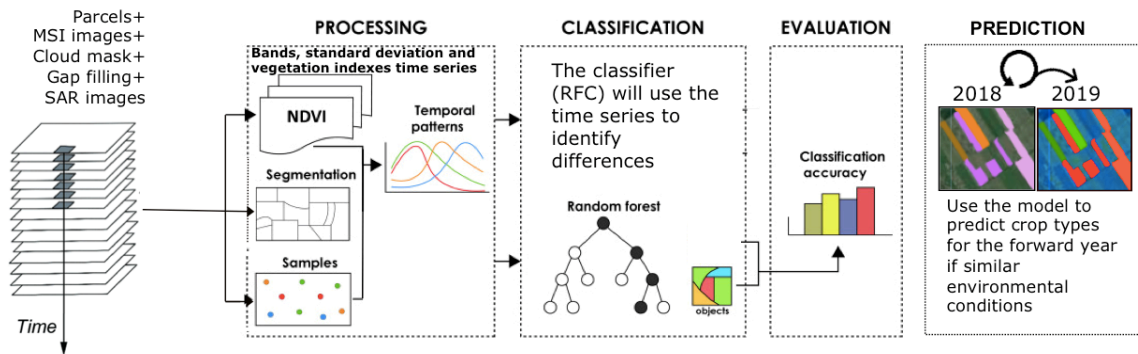
Figure 20: Workflow of the study.

3.5 Predictions

As a further study, it has decided to test how an RFC built on the 2018 agricultural year can perform in the following year, with only radar imagery.

In order to build the model, the same approach as described in previous sections has been used. 2019 images of the same period of the year, were collected. After submitting all scenes to the filtering and cleaning processes, they were intersected with all the selected groundthruth fields' polygons, as described in the section 3.2. A new dataset of 8812 polygons, with their mean features derived from Sentinel1 images, makes up the test set. Every polygon includes 6 variables (VV, VH and VI with their standard deviations) referred in each 12-days time step from November 2018 to November 2019, for a total of 187 properties.

This new testset was fed into the RFC built, trained and tested on the previous year. Results of predicted agricultural classes were compared to BRP from 2019 in order to check the confusion matrix outcome and assess the accuracy.

Moreover, in order to investigate how much the accuracy of 2019 prediction could be affected by an earlier crop development due to temperature and precipitation anomalies, a third test was computed with a dataset built selecting radar images of the same period shifted in time two weeks back.

4 Results

4.1 Valid observations

As described previously, both SAR and MSI images needed some data filtering. Figure 21 reports the geographical distribution of all valid observations derived from the raw dataset. Figure 21 has to be compared with Figure 14, where all observations are shown. A reduced number of acquisitions per pixel is the result of the procedure described in section 3.
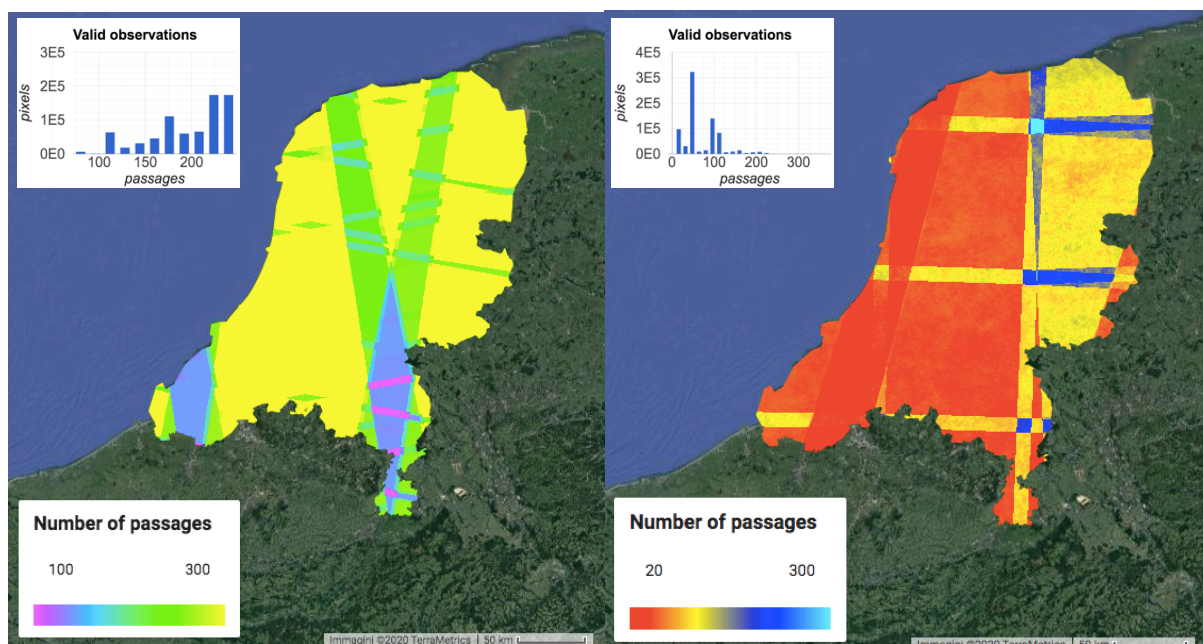
Figure 21: Number of valid observations from Sentinel1 (left) and Sentinel2 (right) in the year of acquisition (1st of Nov 2017-1st of Nov 2018). This representation has a 300 m resolution per pixel.

As expected, SAR images are in larger number than MSI ones (on average 216 and 202, before and after data filtering). The average number of optical valid occurrence is reduced from 189 to 76 after the selection process. Moreover, the coverage is not homogeneous over the full territory. The visible large diagonal stripes over both maps come from the overlapping scenes due to high latitude orbits (ascending and descending).

4.2   Time series trends, cleaned data and indexes (BSI, NDVI, VI)

After filtering anomalous observations and after creating the composite images, the trend through the year can be reconstructed. Watching a single field, for example, it is possible to see how these two satellites recognise crop development.

The NDVI, BSI, VV, VH and VI time series for a potato and a wheat field are shown in Figure 22. As a comparison, the same time series are shown in Figure 23 for urban areas and water surfaces.

The characteristic increase in NDVI during plant development, following by decrease during ripening and harvest, are well correlated with the decrease in BSI. While the crop is growing, the difference between VV and VH, represented by VI, is decreasing, while, overall, backscatter intensities have correlated trends with NDVI and BSI.
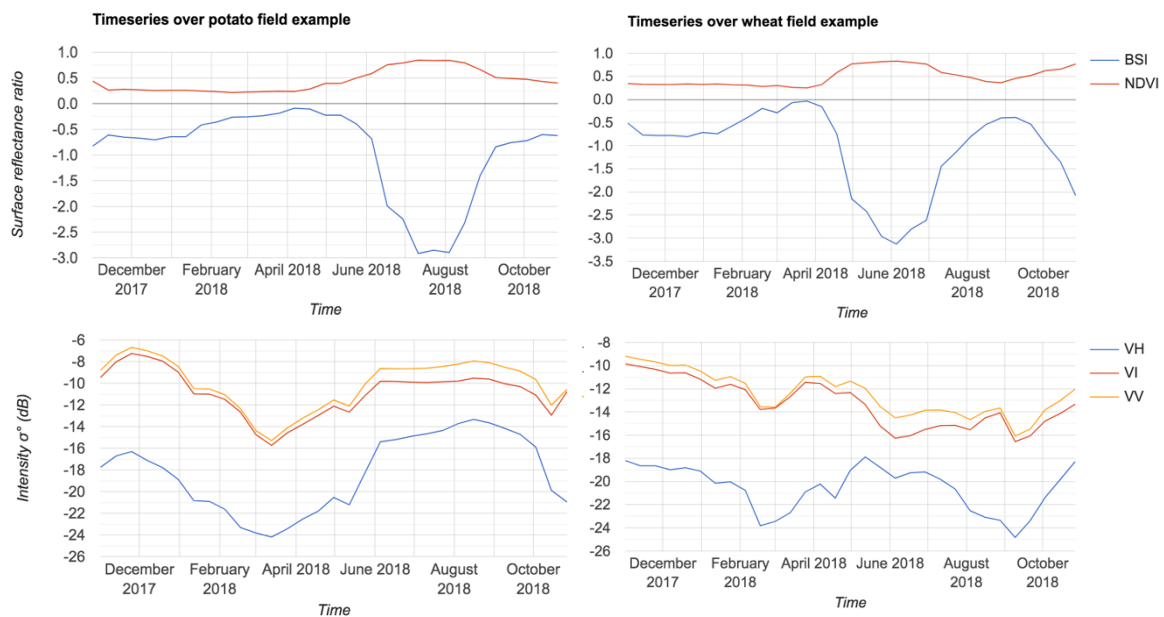
Figure 22: Upper panel: NDVI and BSI from Sentinel 2; lower panel: VV, VH and VI $\sigma°$ intensities from Sentinel1. Left graphs: potato field; right graphs: summer wheat field. Curves report the mean values of the field area throughout the 12-days moving median of the year from 2017/11 to 2018/11.
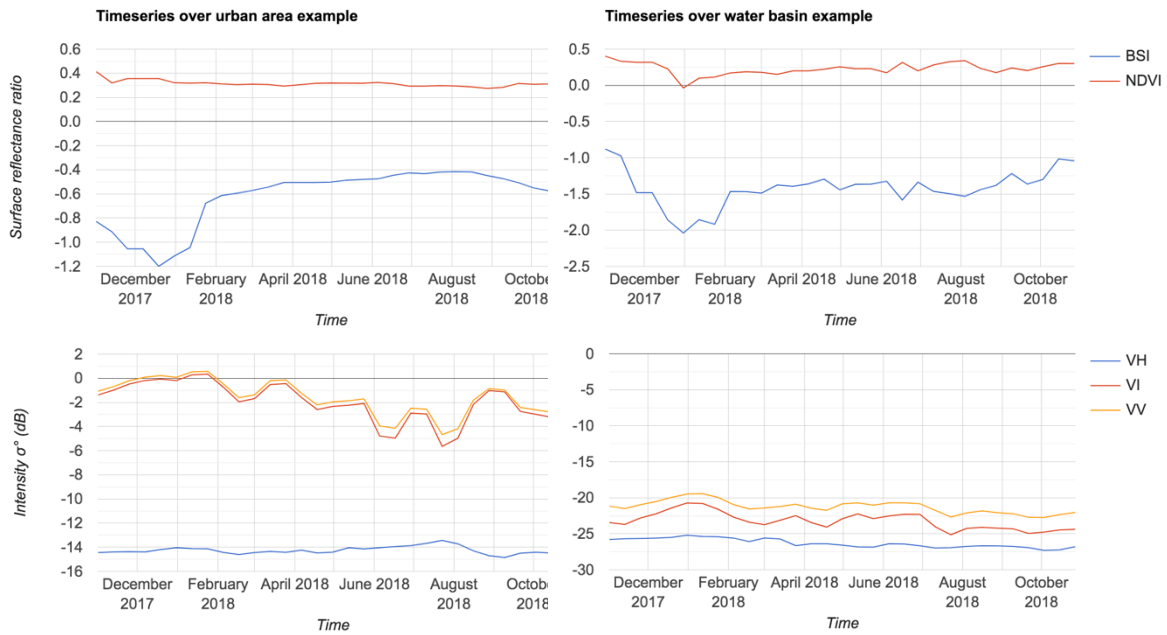
Figure 23: Upper panel: NDVI and BSI from Sentinel 2; lower panel: VV, VH and VI $\sigma°$ intensities from Sentinel1. Left graphs: urban areas; right graphs: water basins. Curves report the mean values of the field area throughout the 12-days moving median of the year from 2017/11 to 2018/11.

It is interesting to notice the differences between crops due to seasonality and to the development cycle, and the correlated behaviour of all parameters especially compared with signal response of urban areas and water basins. While the formers have a harmonic trend during time, the latters tend to stay stable and with either high or low intensity responses, especially with $\sigma°$ values. VH intensities are more representative for surface roughness: it is visible how urban areas and water basins do not have any variation throughout the period of study compare to the two crop fields.

4.3   Classification

Figure 24 represents the 8335 randomly selected polygons of the 22 classes distributed on the Dutch territory as classified by RFC. The visualisation is done with the GEE Code Editor visualization panel. The classification outcome shows the geographical allocation of crop types: different classes are mixed together, while in some areas there are several groups of single crop type. The accuracy assessment would show if this feature could be related to field proximity or not.
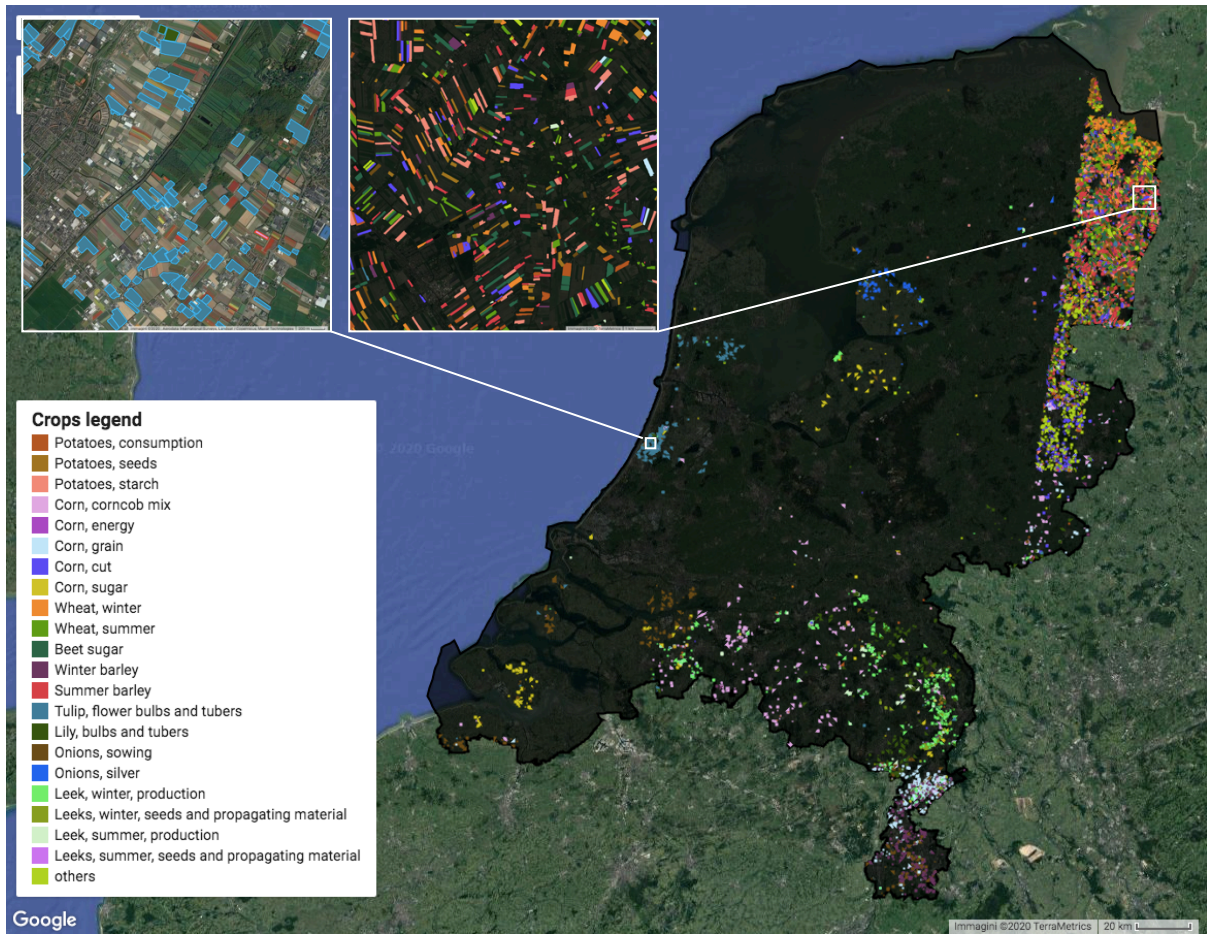
Figure 24: Crop-based classification output, Sentinel1 and Sentinel2 input images

4.4 Accuracy

As described in section 3.3, confusion matrices have been computed in order to measure the accuracy of the classifier.

4.4.1 Training

After the injection of the 6617 features selected for the training part, RFC created an optimal forest of decision trees, with a performance described by the confusion matrix shown in Figure 25. The overall goodness of the classification is shown by the fact that the confusion matrix has a diagonal structure. The overall accuracy is of 76,77%.

		Actual Classes																					
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Predicted Classes	0	211	58	78	2	1	3	0	2	0	0	13	0	0	0	9	0	1	1	1	5	0	2
	1	79	295	56	3	0	0	1	10	0	0	4	0	0	1	0	2	1	1	0	6	0	1
	2	71	27	245	0	0	0	0	2	0	1	8	0	0	0	8	1	0	0	0	0	0	0
	3	1	5	1	269	7	45	36	23	0	8	3	0	0	0	3	2	1	7	3	11	0	1
	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	0	0	0	56	3	234	0	7	0	0	0	1	2	0	3	3	0	1	2	7	1	0
	6	10	5	2	49	24	85	342	5	0	12	1	1	5	0	3	5	0	2	0	2	0	6
	7	1	1	0	0	2	0	0	103	0	0	0	0	0	0	0	2	0	1	1	5	0	0
	8	0	0	0	1	0	0	0	0	358	14	0	4	3	0	0	0	0	0	0	0	0	0
	9	0	1	1	2	0	1	1	0	33	308	1	4	46	1	2	0	0	4	0	3	0	6
	10	9	2	13	0	0	1	0	1	0	3	350	0	0	0	10	0	1	0	0	0	0	0
	11	0	0	0	0	0	5	0	1	8	1	0	399	0	0	0	0	0	5	0	0	0	0
	12	2	0	0	1	0	0	2	1	2	40	0	8	335	1	0	3	0	1	0	0	0	1
	13	1	0	0	0	0	1	0	1	0	0	0	1	0	280	4	0	0	1	0	2	0	0
	14	1	0	2	2	0	3	0	5	0	0	6	0	0	0	300	3	1	3	2	2	2	1
	15	1	2	0	2	0	4	1	15	0	0	2	0	0	1	13	246	13	6	7	7	0	0
	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	48	0	0	0	0	0
	17	1	0	0	2	0	0	0	6	0	0	0	0	0	3	10	2	0	249	19	72	12	0
	18	0	0	0	0	0	1	0	0	0	0	0	0	0	0	3	1	0	5	38	1	17	0
	19	1	1	0	1	0	1	0	5	0	1	0	0	0	1	1	2	0	21	3	80	2	0
	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0
	21	1	2	0	0	0	6	1	3	0	5	0	6	2	0	1	0	2	6	0	4	0	389
Total		390	399	398	390	37	390	384	190	401	393	388	424	393	288	370	273	68	314	77	208	35	407
Overall Accuracy												76,77%												

Figure 25: **Training Confusion Matrix of the RFC using Sentinel1 and Sentinel2 images, with the following numbering of classes: 0-Potatoes, consumption; 1-Potatoes, seeds; 2-Potatoes, starch; 3-Corn, corncob mix; 4-Corn, energy; 5-Corn, grain; 6-Corn, cut; 7-Corn, sugar; 8-Wheat, winter; 9-Wheat, summer; 10-Sugar beet; 11-Winter barley; 12-Summer barley; 13-Tulip, flower bulbs and tubers; 14-Lily, bulbs and tubers; 15-Onions, sowing; 16-Onions, silver; 17-Leek, winter, production; 18-Leeks, winter, seeds and propagating material; 19-Leek, summer, production; 20 Leeks, summer, seeds and propagating material; 21-Others.**

4.4.2 Testing

After the training, the model was tested to understand how well it could generalize on data that it never saw before. The remaining 30% of the dataset was fed to the RFC and compared to the known classes. The yellow colour intensity in Figure 26 is proportional to the percentage of entries with respect to the total entries in the class (the actual groundthruth).

	Actual Classes																					
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
0	60	14	18	0	0	0	0	0	0	0	2	0	0	0	1	0	0	1	0	0	0	0
1	16	80	17	0	0	0	0	2	0	0	0	0	1	0	0	0	0	0	0	1	0	0
2	27	5	64	0	0	1	0	0	0	0	5	0	0	0	3	0	0	0	0	0	0	0
3	0	1	2	76	4	20	13	4	0	2	1	0	0	1	0	0	0	1	1	4	1	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	11	0	64	0	4	0	0	0	1	0	0	0	1	0	0	1	1	0	0
6	1	0	0	16	6	19	97	1	0	3	2	0	0	0	0	3	1	0	0	0	0	2
7	0	0	0	0	1	0	0	29	0	0	0	0	0	0	1	1	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	87	5	0	1	1	0	0	0	0	0	0	1	0	0
9	0	0	0	0	0	0	0	0	12	87	0	0	20	0	1	0	0	0	0	0	0	2
10	2	0	0	0	0	0	0	0	0	0	98	0	1	0	4	0	0	0	0	1	0	1
11	0	0	0	0	0	1	0	0	0	0	0	70	0	1	0	0	0	1	0	0	0	1
12	0	0	0	0	0	0	0	0	0	8	0	0	82	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	1	0	0	0	2	0	66	0	0	0	1	0	2	0	0
14	1	0	1	1	0	0	1	2	0	1	3	0	0	0	79	0	0	2	1	3	0	0
15	1	0	0	0	0	0	0	6	0	0	1	0	1	0	3	68	2	2	0	2	1	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	17	0	0	0	0	0
17	0	0	0	0	0	1	0	1	0	0	0	0	0	1	1	2	0	61	7	29	2	0
18	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	13	1	2	0
19	2	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	3	0	18	1	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	0	1	0	0	0	2	1	1	0	1	0	1	0	0	0	0	1	0	0	1	0	87
Total	110	101	102	104	11	109	112	51	99	107	112	75	107	69	93	75	21	73	23	64	7	93
TPR	0,55	0,79	0,63	0,73	0,00	0,59	0,87	0,57	0,88	0,81	0,88	0,93	0,77	0,96	0,85	0,91	0,81	0,84	0,57	0,28	0,00	0,94
TNR	0,77	0,76	0,77	0,76	0,76	0,77	0,75	0,76	0,75	0,75	0,75	0,75	0,76	0,75	0,75	0,75	0,76	0,76	0,76	0,78	0,76	0,75

Overall Accuracy	75,84%
Error	24,16%
Kappa statistic	74,43%

Figure 26. **Test Confusion Matrix** of the RFC using Sentinel1 and Sentinel2 images, **the numbering of classes is the same as in Figure 23. Overall accuracy, Sensitivity (TPR), Sensibility (TNR), Kappa coefficient and Error metrics of the classification system are reported.**

In addition, other metrics as classifier performance estimators were measured. The overall accuracy is 75,84%, similar to the training set, which shows that there is no evident indication of overfitting. The Kappa coefficient is 74,43% and the error is 24,16%. In Figure 26 TPR and TNR cells are coloured with a scale from red (lowest value) to green (highest value). The sensibility (TPR) shows how good is the model to find the real class, while the sensitivity (TNR) is the capability to discriminate well all cases that are not included in a certain class. TPR is also called "consumer's accuracy" for its estimate of how much in percentage all areas identified as one class are actually correct. TPR and TNR are, as expected, complementary to their selves.

From the total count of polygons from both the confusion matrices it is possible to see that some classes are under-represented. Also, as shown in Figure 27, some crops are more easily detected.
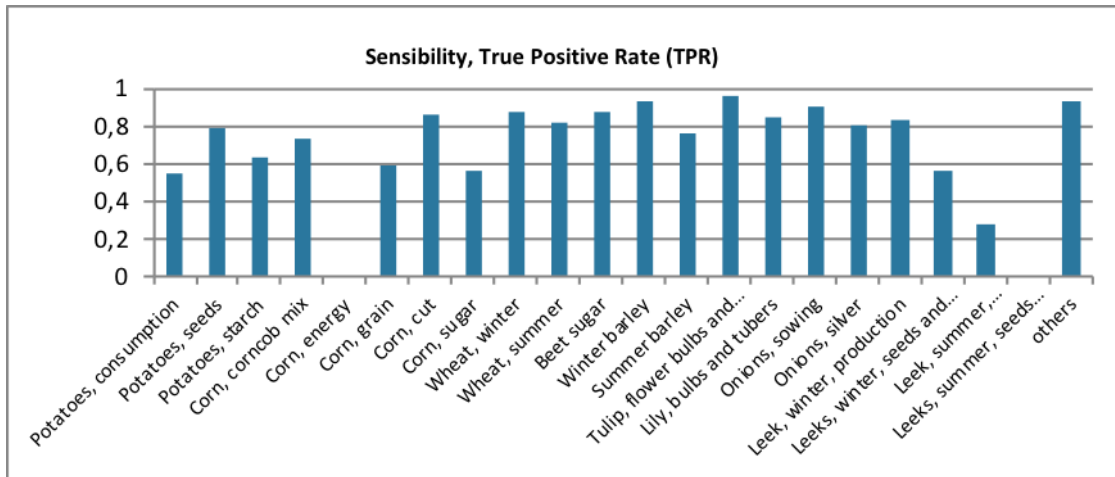
Figure 27. TPR of the classification model per class

It is interesting to represent the accuracy of all classification on a map where blue polygons are truly classified and red ones the not well classified. Figure 28 shows the geographical localisation of classified polygons with respect to a correct or incorrect outcome. From this kind of figures it is possible to estimate if there is an inhomogeneous distribution of wrong results that could depend on specific factors.

Correctly classified   Wrongly classified
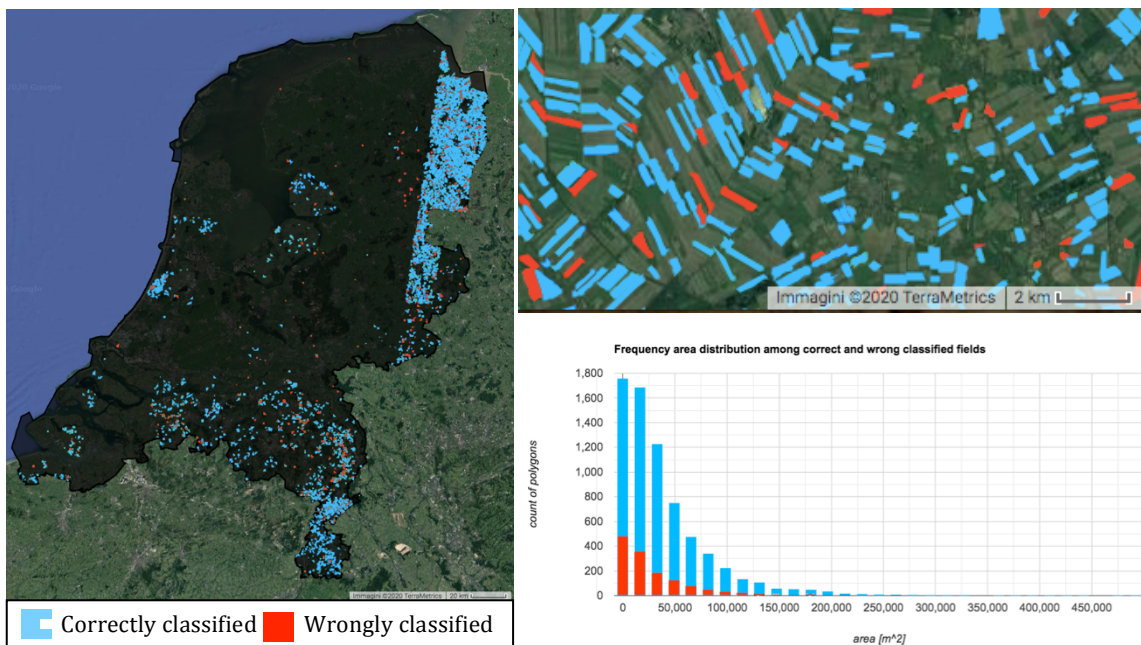
Figure 28: Geographical representation of true (in blue) and false (in red) classification. On the right the national overview, on the left a zoom in and the histogram of correct and incorrect classifications frequency by field area. Satellite images on the background come from GEE visualization panel.

Polygons proximity, topography, specific crop type misclassification, lack of good observations and field size are some of possible causes of model performance interference. Comparing this map with the one in Figure 24 can help to understand

better the classification model. Red polygons are generally equally distributed over the territory; no more dense errors areas are occurred. The number of misclassified crop fields as a function of surface magnitude has a monotonous decreasing trend similar to the one of correctly classified samples.

4.5   Results using only the Sentinel1 imagery

To understand more deeply the sources of variability in the results as well as to investigate which method can perform better results from different satellite imagery as input of the machine learning algorithm have been used. This assessment improves the comprehension of the system and has the aim to observe if some crops are better recognized from radar sensors only.

4.5.1   RFC model for 2018

As described in section 3.5, a further study on only Sentinel1 classification model has been computed. Here to follow are reported confusion matrices of training and test processes concerning the period from November 2017 and November 2018. In both Figure 29 and Figure 30 is visible the expected diagonal structure, however from class 0 to 6 and classes 18, 19 and 20 a higher confusion in the predicted values was observed.

Parcels used for this part of the study are the same as the one used for the combined Sentinel1 and Sentinel2 analysis. The overall accuracy of the training set is 73,70% and for the test set is 73,60%. This model is apparently not affected by overfitting like the one described in the previous section. K coefficient is 72,05%, while for the model with Sentinel2 is 74,43%. A small decrease is observed in both OA and K. However, as shown in Figure 31, true positive rates report a similar variability among crop classes as in the combined classification model. Interesting results are noticed in some classes like for example winter barley and sowing onion with a TPR value of 94% and 93%, respectively. In addiction, some other classes like corn energy and leek are particularly not well classified by the model, mainly because of their low frequency.

		Actual Classes																					
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Predicted Classes	0	213	46	75	6	0	4	0	5	0	0	8	0	0	0	6	2	0	1	1	6	0	2
	1	91	306	65	0	0	1	0	19	0	0	3	0	0	0	0	7	2	2	0	5	0	0
	2	75	26	252	1	0	0	0	0	0	1	3	0	0	0	17	0	0	0	0	0	1	0
	3	6	5	4	234	12	74	58	13	0	12	2	4	1	0	3	2	2	8	1	11	0	4
	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	9	4	3	66	6	256	40	7	0	17	2	3	3	3	10	5	1	7	1	8	0	10
	6	4	1	0	64	26	62	297	3	0	3	0	0	1	0	1	2	0	2	0	7	0	0
	7	0	4	1	8	2	3	1	154	0	0	1	0	1	0	3	5	0	14	1	11	1	2
	8	0	0	0	1	0	2	0	0	361	15	0	4	2	0	5	2	0	3	10	2	6	1
	9	0	0	0	4	0	6	3	2	37	281	0	3	60	7	2	1	1	4	14	5	0	17
	10	6	2	8	1	0	1	0	4	0	1	367	0	1	0	24	1	0	0	0	1	0	1
	11	0	0	0	1	0	1	0	2	0	0	0	361	3	3	1	0	0	5	2	5	3	3
	12	0	0	0	1	0	1	0	1	4	30	0	6	335	0	2	1	1	0	1	6	0	2
	13	0	0	0	2	0	3	0	1	1	3	0	1	0	375	2	1	0	6	1	17	0	5
	14	1	0	1	2	0	0	0	1	0	0	10	0	0	0	310	0	0	1	1	0	0	0
	15	0	0	0	4	1	0	5	19	0	3	4	0	0	0	8	367	17	9	5	17	2	0
	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	52	0	0	0	0	0
	17	1	0	0	1	0	1	0	8	0	3	2	0	0	4	3	2	0	256	15	66	9	0
	18	0	0	0	0	0	0	0	0	0	0	0	0	1	0	6	0	0	7	33	1	10	0
	19	2	1	0	0	0	0	0	1	0	0	0	0	0	1	1	1	0	14	1	54	0	0
	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	21	0	0	0	0	0	4	1	3	0	10	0	6	2	6	9	0	1	2	1	2	2	355
Total		408	395	409	396	47	419	405	243	403	379	402	388	410	399	413	399	77	341	88	224	34	402
Overall Accuracy												73,70%											

Figure 29: Training confusion matrix, using only Sentinel1 images, from Nov 2017 to Nov 2018. Enumeration of classes is the same as in Figure 25.

		Actual Classes																					
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Predicted Classes	0	56	12	23	0	0	0	0	2	0	0	1	0	0	0	2	0	0	1	0	1	0	0
	1	21	76	11	0	0	0	0	5	0	0	0	0	0	0	0	3	0	0	0	6	0	0
	2	10	11	55	0	0	0	1	0	0	0	1	0	0	0	2	0	0	0	0	1	0	0
	3	1	2	0	67	2	19	19	7	0	3	0	1	0	0	0	0	0	2	1	4	0	0
	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	0	2	0	20	0	41	6	2	0	9	2	1	0	0	4	0	0	2	0	4	0	5
	6	2	0	0	14	2	17	67	1	0	1	0	0	0	0	3	0	0	0	0	0	0	0
	7	0	0	0	3	2	0	0	41	0	0	0	0	0	0	0	1	0	0	1	5	0	0
	8	0	0	0	0	0	0	0	0	88	14	0	1	2	1	1	0	0	1	1	1	2	0
	9	0	0	0	0	0	3	1	0	6	83	1	1	12	5	2	0	0	4	0	0	0	2
	10	1	0	2	0	0	1	0	0	0	1	89	0	0	0	3	0	0	3	0	0	0	0
	11	0	0	0	0	0	0	0	0	2	0	0	105	0	0	0	0	0	1	1	0	0	0
	12	0	0	0	0	0	0	0	0	1	6	0	1	75	1	0	1	1	0	0	2	1	1
	13	0	0	0	0	0	0	0	0	0	0	0	0	0	88	0	0	1	2	0	1	0	1
	14	1	0	0	0	0	0	0	0	0	0	4	0	0	0	64	0	0	1	0	0	0	0
	15	0	0	0	0	0	0	1	8	0	0	0	0	0	0	1	94	0	1	0	0	1	0
	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0
	17	0	0	0	0	0	0	0	2	0	1	0	0	0	0	1	1	0	69	4	20	1	0
	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	1	3	0	3	0
	19	0	2	0	0	0	0	0	0	0	0	0	1	0	3	0	0	0	3	1	14	0	0
	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	21	0	0	0	0	0	0	0	0	0	3	0	1	1	3	2	1	0	0	1	0	1	89
Total		92	105	91	104	6	81	95	68	97	121	98	112	90	101	87	101	12	91	13	59	9	98
TPR		0,61	0,72	0,60	0,64	0,00	0,51	0,71	0,60	0,91	0,69	0,91	0,94	0,83	0,87	0,74	0,93	0,83	0,76	0,23	0,24	0,00	0,91
TNR		0,74	0,74	0,74	0,74	0,74	0,75	0,74	0,74	0,73	0,74	0,73	0,72	0,73	0,73	0,74	0,72	0,74	0,73	0,74	0,75	0,74	0,73
Overall Accuracy												73,60%											
Error												26,40%											
K coefficient												72,05%											

Figure 30: Test confusion matrix using only Sentinel1 images, from Nov 2017 to Nov 2018. This RFC algorithm was used to classify parcels for the following year. TPR, TNR, OA, ERR and K describe the performance of the model.
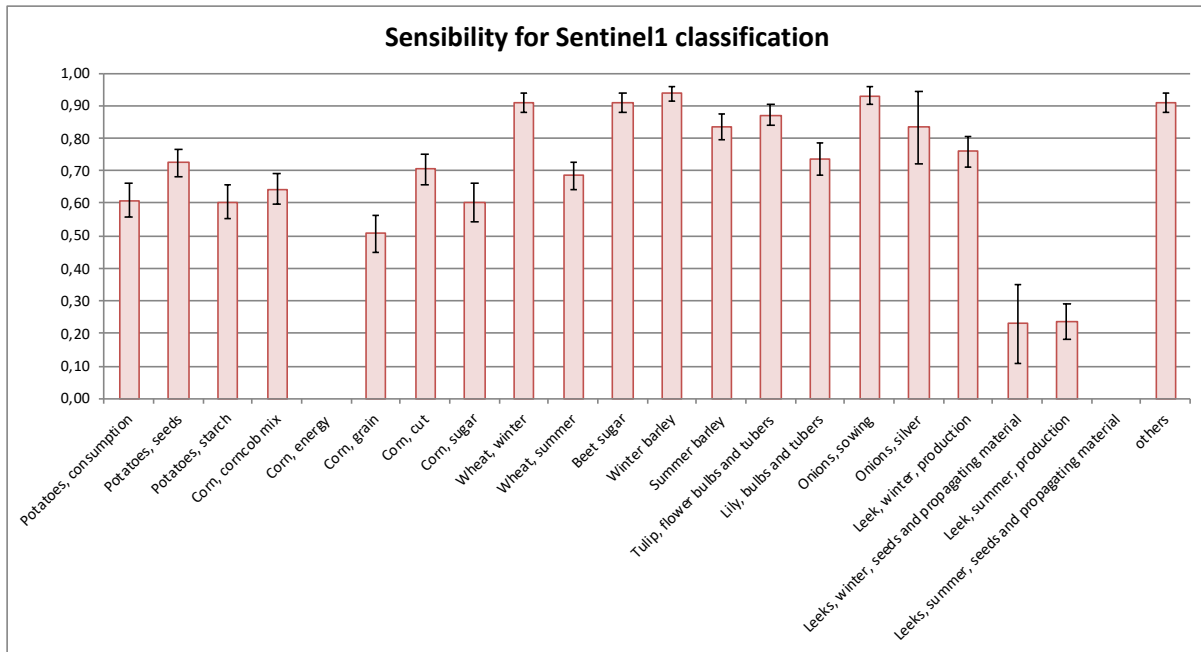
Figure 31: Sensibility of the RFC model with only radar images.

4.5.2 Prediction for 2019

Once the RFC model has been trained and tested on the 2018 agricultural year, it was tested over satellite's observations in 2019 in the same period of the year (from 1st of November 2018 to 1st of November 2019). The same polygons were used as input. Only after performing the classification, results were compared with the 2019 version of BRP dataset. Figure 32 represents the resulting confusion matrix.

Some crop classes were not present among the parcels, i.e. potato for consumption and silver onion. A new class had to be defined due to some parcels that were classified in BPR but were not possible to evaluate due to no groundthruth declaration concerning them. Overall accuracy and K coefficient are clearly lower with respect to the others test set confusion matrices. Cell colours help to understand the magnitude of misclassification.

Due to the low measured performance, a further analysis was performed.

When using a classifier built on a different year than the one it is tested on, a problem can rise if the environment conditions are not equivalent. Even if the weather (temperatures and precipitations) were similar among years 2017, 2018 and 2019, some differences in other parameters as well as anticipation of plants growth

51

development may occur. To test this last hypothesis, the same classifier was used with as, input, a shift of two weeks back of the radar observation. In Figure 33 the resulting test confusion matrix is shown. The structure is similar to the previous one. All metrics show a worse classification performance.

Predicted \ Actual	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	23
0	0	74	260	14	0	3	12	7	0	0	20	0	2	2	8	0	0	2	0	0	0	40	169
1	0	90	38	0	0	0	1	9	0	0	3	0	0	1	1	0	0	1	0	8	0	44	79
2	0	9	401	1	0	0	0	0	0	1	19	0	0	0	1	0	0	0	0	0	0	8	21
3	0	0	0	75	7	53	195	4	2	3	4	0	0	0	1	1	0	2	1	1	0	24	2
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	3	132	3	197	418	5	5	29	10	2	7	4	3	2	0	3	0	12	1	174	7
6	0	1	0	8	0	8	44	0	1	0	1	0	0	0	0	0	0	1	0	1	0	12	0
7	0	0	3	15	1	5	44	26	2	0	3	0	0	0	1	4	0	3	1	7	1	63	2
8	0	0	2	0	0	0	1	0	685	19	0	2	26	0	1	0	0	1	1	0	2	87	1
9	0	0	0	2	0	4	4	2	10	72	1	0	38	0	0	0	0	2	6	3	0	62	0
10	0	13	384	25	0	15	6	2	1	4	739	2	1	1	10	4	0	4	1	2	0	112	95
11	0	0	0	1	0	1	4	1	6	2	1	85	3	1	0	0	0	1	2	2	0	112	0
12	0	0	0	2	0	0	2	0	4	2	0	2	251	1	0	0	0	0	1	1	0	29	1
13	0	0	0	0	0	0	1	0	0	0	0	0	0	24	0	0	0	0	0	9	0	140	1
14	0	25	70	1	0	0	1	1	0	1	147	0	0	0	119	0	0	1	0	1	0	66	228
15	0	1	1	9	1	11	75	2	40	15	19	0	25	1	4	78	0	5	5	14	0	142	4
16	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
17	0	0	1	2	0	7	9	16	4	10	3	0	11	4	0	6	0	172	22	57	14	167	1
18	0	0	0	0	0	0	1	0	0	0	2	0	1	0	1	0	0	3	16	1	4	17	0
19	0	0	0	0	0	0	0	1	0	2	1	0	1	2	0	0	0	0	1	33	0	39	2
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	0	0	2	1	0	4	4	0	72	19	1	7	13	4	3	0	0	7	6	2	9	755	8
																							0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Total	0	213	1165	288	12	308	822	76	832	179	974	100	380	45	153	95	0	208	63	154	31	2093	
TPR	-	0,42	0,34	0,26	0,00	0,64	0,05	0,34	0,82	0,40	0,76	0,85	0,66	0,53	0,78	0,82	-	0,83	0,25	0,21	0,00	0,36	
TNR	0,47	0,47	0,49	0,48	0,47	0,46	0,52	0,47	0,43	0,47	0,43	0,47	0,46	0,47	0,47	0,47	0,47	0,46	0,47	0,48	0,47	0,51	

Overall Accuracy	43,83%
Error	52,85%
K coefficient	39,52%

Figure 32: Test confusion matrix out came from prediction of 2019 crop fields. Only radar images as input were used (period of observation ingested: 1st November 2018-1st November 2019). Class 23 represents parcels that could not be possible to evaluate due to the absence of groundthruth data of the respective year. TPR, TNR, OA, ERR and K describe the performance of the model.

4.6   Comparison of the different RFCs

A comparison of true positive rates in all the study cases reported, helps to evaluate which method is better for classifying certain classes. Figure 34 reports the histogram representing the sensibility of the methods for all 22 classes and the four cases: combined radar and optical, radar only 2018, radar only 2019 and radar only 2019 shifted by two weeks.

Even if the classification with combined optical and radar images have a higher overall accuracy compared to the use of only radar data, the results for most of the classes are within the estimated statistical error.

														Actual Classes										
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	23
Predicted Classes	0	0	92	214	23	1	9	10	2	0	0	28	0	2	2	5	0	0	1	2	3	0	36	167
	1	0	53	6	1	0	1	1	7	0	0	1	0	0	3	0	0	0	1	0	8	0	91	60
	2	0	3	88	1	0	0	0	0	0	1	3	0	0	0	2	0	0	0	0	0	0	9	6
	3	0	0	0	72	5	71	174	3	3	5	3	1	2	2	0	0	0	3	0	3	0	43	1
	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	0	0	2	76	1	112	294	7	2	24	5	1	9	1	2	2	0	2	1	8	1	129	3
	6	0	0	0	6	0	9	25	0	0	1	1	0	0	0	1	0	0	0	0	3	0	12	0
	7	0	19	16	59	2	39	99	28	9	0	6	2	6	0	1	6	0	2	0	6	1	101	11
	8	0	0	1	0	0	0	0	0	573	6	0	23	3	1	0	0	0	2	4	0	0	66	1
	9	0	0	0	2	0	5	5	0	72	52	1	1	55	0	0	0	0	2	6	3	0	42	0
	10	0	20	675	2	0	7	4	1	1	1	747	1	1	1	18	4	0	1	0	3	0	81	175
	11	0	0	0	0	0	0	1	0	1	1	0	23	15	1	0	0	0	0	0	2	2	17	1
	12	0	0	0	3	0	7	5	2	11	3	0	18	101	4	0	0	0	1	1	9	0	77	0
	13	0	0	0	0	0	0	0	0	0	0	0	0	0	19	0	0	0	0	0	3	0	77	1
	14	0	19	138	1	0	0	1	1	0	0	128	1	0	0	112	0	0	0	0	0	0	64	172
	15	0	7	8	15	1	24	144	0	16	31	24	5	91	0	5	67	0	2	2	8	0	146	7
	16	0	0	0	1	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	1	0	5	0
	17	0	0	14	23	1	15	42	24	5	10	22	1	56	2	2	16	0	179	16	64	7	202	4
	18	0	0	0	0	0	0	2	1	1	3	1	0	4	0	1	0	0	2	13	0	2	25	2
	19	0	0	0	0	1	0	2	0	0	1	1	0	0	3	1	0	0	2	0	22	0	33	0
	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	21	0	0	3	3	0	9	13	0	138	40	2	23	34	6	3	0	0	8	18	8	18	837	10
																									0
	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Total		0	213	1165	288	12	308	822	76	832	179	974	100	380	45	153	95	0	208	63	154	31	2093	621
TPR		-	0,25	0,08	0,25	0,00	0,36	0,03	0,37	0,69	0,29	0,77	0,23	0,27	0,42	0,73	0,71	-	0,86	0,21	0,14	0,00	0,40	
TNR		0,35	0,36	0,40	0,36	0,35	0,35	0,39	0,35	0,32	0,36	0,30	0,36	0,36	0,35	0,35	0,35	0,35	0,34	0,36	0,36	0,36	0,34	

Overall Accuracy	35,44%
Error	61,87%
K coefficient	30,26%

Figure 33: Test confusion matrix out came from prediction of 2019 crop fields, anticipating of two weeks the satellite observations data ingestion (from 15th of October 2018 to 15th of October 2019). Class 23 represents parcels that could not be possible to evaluate due to the absence of groundthruth data of the respective year. TPR, TNR, OA, ERR and K describe the performance of the model.
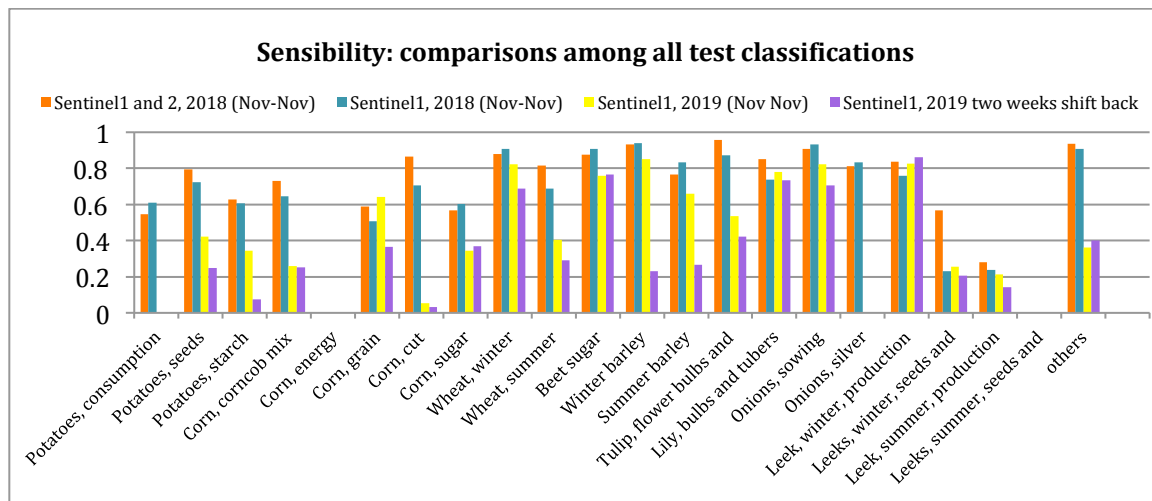

Figure 34: Comparisons of TPRs of all RFC tests computed in the study.

5 Discussion

The data preparation for the classifier is the most time-consuming step in the entire study. Because it deals with a considerable amount of data from different sources, many different processes are taken in consideration. The unbalanced number of images between Sentinel1 and Sentinel2 is substantial. The cloud mask applied allows to detect a certain type of clouds but some, not sharp-edged, shapes are difficult to flag and discard. On the other hand, using a cloud mask instead of just dropping scenes with cloud presence helps to recover pixels that can actually be used thus increasing the number of valid observations in some areas.

About Sentinel1 images, $\sigma°$ intensity is conditioned by soil moisture, and in general water surfaces. Therefore, backscatter values may be biased due to weather conditions (i.e. floods, precipitations, drought as well as irrigation). The use of a moving median statistics should reduce this issue.

In general, time series and multisensor composition enhance the robustness of the model. A comparison between a combined SAR and MSI model with the one using only SAR shows a gain of 2.2% in the overall accuracy. Merging information allows distinguishing better over crops with similar response signals but with different harvesting periodicity like, for example, lilies and tulips or winter and summer wheat. Monitoring over time demonstrates the harvest frequency throughout the year. On the other hand, it can classify differently crops with similar phenology like onion for sowing and silver-onion. Additional information derived from merging SAR and MSI improves the classification accuracy, as in (Fuyou, et al., 2019) (Kristof Van Tricht, 2018).

For the reliability of the study, it is important to remember the assumptions made about parcel-based analysis, described in section 1.5. The possibility that for future implementation some boundaries will change is a relevant factor. In this case further information about updated parcel distribution have to be acquired. From the analysis conducted over the 2019 agricultural year with Sentinel1 imagery there were no evident case of wrong classification due to change in parcels boundaries. However, this problem is relevant, especially in view of the low performance of the same model applied for two following years.

Observing the trends in Figure 22 it is interesting to see how two fields of potatoes and wheat describe exactly what it has been reported in the 2018 STOWA report (STOWA, 2018). Potatoes are sowed between March and April. The crop cycle proceeds till September when the tuberous is harvested. Wheat is seeded in winter (November-December) and grows until the harvesting season in August. The signals are correlated as expected from the literature and this correlation is clearly seen from the analysis. BSI and NDVI have opposite trends, with NDVI describing crop photosynthetic activity. Concerning Sentinel1 bands, the change in polarization represented with the VH backscatter intensity which depends on the increase in the roughness of the surface and the development of leaves (volume backscatter). In general VV band (vertical polarisation emitted and received) have a higher intensity than VH (vertical emitted and horizontal received) and for vegetation cover they vary between -30dB and -5dB (contrary to other land uses as shown in Figure 23).

Knowing the time series of several parameters, like vegetation indexes, for a large amount of fields and for many different crop types has a great potential for many applications. It is possible to study trends of phenology correlated with ground measures of the physical environment parameters. For example soil moisture compared with NDVI can help detecting localized needs of irrigation. Furthermore, creating average standard spectra of certain crops throughout the plant development might be used for evaluate the health of a specific farmer's field at a short-term notice.

Although many studies use mainly NDVI, VV and VH (Kristof Van Tricht, 2018), the choice of using many bands from Sentine1 and Sentinel2 increases the amount of information on which the classifier can work and find common trends for the same crop which result in an increased accuracy. Generally, using more satellite information increase, indeed, performances of the model.

Thanks to the OOB mode of RFC it is possible to avoid overfitting problems. RFC draws randomly a number equal to the square root of the features list with replacement. The OOB features of the training set are used for cross validating the model. The measured values of accuracy (75.84%) indicate that the model generalizes well over new data. The high K coefficient (74,43%) also indicate that the classification model is very far from a random association. These results, are similar to the observation with only radar data, and differ by only 2.2% for OA and by 2,4% for K, from the combined model. These

results shows that to create a machine learning classifier using the Random Forest algorithm from a national subset of crop parcels is possible and merging different sensors helps to increase accuracy.

Comparison with other articles results (even if they use different methodology) (Kristof Van Tricht, 2018) a maximum overall accuracy of 82% (kappa 0.77) was reached. As OA strictly depends on true positives and true negatives over the total amount of inputs, when there are classes that are not well represented, the metric is affected. OA can be increased by

- removing the misclassified classes (i.e. classes 4 and 20),
- undersampling the most frequent classes (reducing the number of samples),
- creating two different models for similar class abundance
- oversampling less frequent classes with other sources of dataset.

Observing the histograms in Figure 27, Figure 31 and Figure 34 and the test confusion matrices (Figure 26, Figure 30, Figure 32, Figure 33) it is clear that some classes are better detected than others. While some, like corn energy and leeks, have very low or even zero true positive rate, some others like tulips, winter barely, sowing onions are recognized best. Tulips, in particular, are actually the easiest to recognize due to their characteristic strong blooming colours.

It is interesting to notice that for winter crops of the same species the sensibility is higher. This could depend on more data acquired during the developing season, compared with the crops planted in summer and developing through and after the last part of the acquisition period (ending in November).

The selection of classes was driven by most common crop types in the Netherlands, as described in the STOWA Report, together with the high variability of the labels reported to in BRP dataset. Since there was the possibility to discriminate among different cultural types of the same species (i.e. potatoes for consumption, starch and seed or onion for sowing or consumption), it would have been interesting to know if a machine learning classifier had perceived the difference. For some crops, specifications about different use destinations result in a different response on the field that can be possibly detected by the satellite. For example, corn classes' subsets are divided in:

- corncob mix, which is partly harvested maize (corn cob plus some part of crop residues)
- energy, that is a particular maize with focus on production yield; the nutritional value is of no concern, and the whole plant is generally harvested.
- grain, it is a maize harvested for grains, where all crop residues are left behind
- cut; it is a silage maize: the main maize type in NL.; the whole plant is harvested; it is used to store for feeding in the winter, and fermented.
- sugar, this corn type is harvested for the corn cob, mainly for human consumption, generally before full ripening.

Corn cut was the sub class that got the higher accuracy. However, it is one of the most represented (see in Annex 2 the number of parcels per class). In general, confusion matrices clearly report that crop classes of the same crop family are subject to more confusion. Results reported by Fuyou et al., 2019 identify that corn class, grouped in one unique class, can be detected with high accuracy, however no literature exists on a comparison among different yeld use.

Classification accuracy differs not only among crop classes but, for the same class, among different observation sensors. Figure 34 shows that corn classes are better detectable using MSI sensor in addition to SAR. This happens also to tulip, lily, onion and leek classes. Wheat, sugar beet and barley have, on the contrary, higher TPRs using only SAR, though within the statistical errors. Potato's classes do not show any relevant preference. Plants better identified with optical images have a higher intensity in colour changes and plants, better identified by radar images, have specific structural shape of surfaces covered by the crop. It is important to notice that sentinels sensors cannot identify row distribution on the ground since their resolution is worst than 10m.

A decrease in accuracy can be related to other reasons. Hence, when applying the classifier to new crops or new geographical regions, it is crucial that training data are representative of these new cases. For example, climatic conditions should be similar in order to not change substantially the time evolution of the plant growth.

Polygons proximity in the dataset, topography, specific crop type misclassification, lack of good observations and field size are some of possible causes of degraded model

performances. From Figure 28 it is possible to evaluate if there is a spatial distribution of misclassified polygons that could be due to environmentally adverse conditions i.e. high soil moisture or slope/topography shadowing effects especially for radar images. None of those causes seems to have relevant effects on the correctness of the classification. A comparison with Figure 27 shows that the particular grouped distribution of same crops fields it is not due to geographical biases, because the wrong classified plots are homogeneously distributed on the territory. Moreover, parcel's area among correct and incorrect subsets has the same trend of occurrence, hence neither surface magnitude of the crop seems interfere with results. Anyhow, fields smaller than $100m^2$ were filtered in advance. Increasing in sensor's ground resolution might improve classification of even smaller parcels.

Concerning the prediction of crop types in the year 2019 with Sentinel1 images, results shows interesting performances for several classes. However, the value of OA and K coefficient values indicate that the method needs improvement. For classes like winter wheat, barley and leek and sowing onion, TPRs are comparable with the results of 2018. Separating these crop classes in a different model and performing it with different increasing time delays of the satellite images could help to understand better the classification workflow.

GEE is a valid instrument for fast processing and visualization projects. It is easy to learn and with increasing number of developers that create ad hoc tools to import. The biggest hurdle is to find a trade-off between computing time length and memory usage in order to fall into memory errors. Dividing the workflow in several separate steps instead of running all the script in one shot might help. When many complex features (e.g. polygons) are involved, it is advisable to simplify edges and reduce vertices. When performing long runs it is suggested to export the results instead of printing them on the console output. The visualisation panel instead gives a fast representation of geolocalised features and satellite scenes that improve a lot the data pre-processing.

The algorithm, written for this study, is scalable. Once the classifier is trained, it is possible to run the model over all Dutch territory segmented into several patches.

6 Conclusions

Agriculture is a fast-growing sector when it comes to the utilization of EO-based products. It plays a crucial role in global economy and is rapidly evolving especially due to climate change and increasing production demand. As population and aspirations increase, land, water and energy become fundamental resources. Hence, smart and more efficient land cover monitoring can help to tackle several significant challenges. This study aims to exploit open- and free-accessible cloud-computing tools and datasets for crop mapping, which have the potential of a knock-on disrupting added value in EO on worldwide users, from policy makers to commercial and private users.

At present, large scale crop mapping is mostly derived from coarse resolution satellite images. The recent availability of free and open EO datasets improves radically the applications in this field. High resolution SAR and optical imagery at 10 m resolution enable the user to elaborate agricultural land use monitoring in detail and can be applied in large scale with the support of cloud computing technology.

Sentinel1 and Sentinel2 from the Copernicus Programme can give a relevant improvement for land use monitoring. Google Earth Engine demonstrates to be a powerful tool in executing a parcel-based crop classification.

The thesis project consisted in developing a machine-learning RF classifier that identifies mayor regional crop types from radar and multispectral images. It combines 12-days moving median composites of Sentinel1 and Sentinel2 satellites trough one agricultural year (2017-2018). Some commonly used vegetation indexes as NDVI have been used in addiction to images bands intensities. The workflow has been applied to the territory of the Netherlands. Groundtruth dataset comes from national Basic Registration of crop Plots.

The combination of multi-sensors and multi-temporal images enhances the classification over areas with relevant cloud coverage through the year and gives more information about growth phenology. Accuracy assessment gives interesting results especially for some crop classes' detection (corn, wheat, sugar beat, onions and tulip),

reliable at the 85% - 96% level. Among these, higher accuracy has been observed for winter crops of the same species.

A comparison with only Sentinel1 imagery shows a small reduction of 2% of the overall accuracy of the classification. The model has been applied as such for the following year, with mixed results, so new approaches have to be explored.

A further in the research would be to more deeply learn the model mechanism. A sensitivity analysis allows investigating if some variables have higher importance in the classification performances than others. Moreover, monitoring the receiver operating characteristic curve (ROC curve) might help to understand which thresholds of classifier hyperparameters settings achieve higher accuracy with respect to each crop types. Also, the application of simple deep learning methods specialised in time series analysis might give an improvement to the results.

The quantification of specific crop types with a very high accuracy and early in the crop development can greatly improve actual rural monitoring. Knowing in advance specific information extremely useful for harvesting quantification gives commercial values to yields and food waste as well as identifies possible bio-sources for energy production.

7   Bibliography

Alavipanah, S.K. et al., 2010. Criteria of selecting satellite data for studying land resources. *Desert*, pp.83-102.

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

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

Betbeder, J., Fieuzal, R. & Baup, F., 2016. Assimilation of LAI and Dry Biomass Data From Optical and SAR Images Into an Agro-Meteorological Model to Estimate Soybean Yield. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.

Blom-Zandstra, M., Paulissen, M., Agricola, H. & Schaap, B., 2009. How will climate change affect spatial planning in agricultural and natural environments? Examples from three Dutch case study regions. *IOP Conference Series Earth and Environmental Science* .

Breiman, L., 2001. Random Forest. *Machine Learning*, 45, pp.5-32.

Breiman, L. & Cutler, A., 2003. *Classification/clustering*. [Online] Available at: https://www.stat.berkeley.edu/~breiman/forests/cc_home.htm [Accessed 2020].

Canada Center for Remote Sensing, 2015. *Fundamentals of Remote Sensing*. Natural Resources Canada.

Carrasco, L., O'Neil, A.W., Morton, R.D. & Rowland, C.S., 2019. Evaluating Combinations of Temporally Aggregated Sentinel-1, Sentinel-2 and Landsat 8 for Land Cover Mapping with Google Earth Engine. *Remote Sensing*.

ESA, 2017. *Soil moisture in the Netherlands.* [Online] Available at: https://www.esa.int/ESA_Multimedia/Images/2017/08/Soil_moisture_in_the_Netherlands [Accessed 2020].

European Commission , 2015. *Copernicus Europe's eyes on Earth*. Luxembourg : Publications Office of the European Union.

European Commission, 2016. *EU SCIENCE HUB*. [Online] Available at: https://ec.europa.eu/jrc/en/research-topic/earth-observation [Accessed 2020].

FAO, 1993. *Guidelines for Land-Use Planning.* Rome: FAO.

Fuyou, T. et al., 2019. Efficient Identification of Corn Cultivation Area with Multitemporal Synthetic Aperture Radar and Optical images in the Google Earth Engine Cloud Platform. *Remote Sensing*.

Géron, A., 2017. *Hands-On Machine Learning with Scikit-Learn and TensorFlow*. O'REILLY.

Google, 2019. *Debugging guide*. [Online] Available at: https://developers.google.com/earth-engine/debugging#scaling-errors [Accessed 2020].

Google, 2020. *A planetary-scale platform for Earth science data & analysis*. [Online] Available at: https://earthengine.google.com/ [Accessed 2020].

Google, 2020. *Sentinel-2 MSI: MultiSpectral Instrument, Level-2A*. [Online] Available at: https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR [Accessed 2020].

Gorelicka, N. et al., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 201, pp.18-27.

Griffiths, P., Nendel, C. & Hostert, P., 2019. Intra-annual reflectance composites from Sentinel-2 and Landsat for national-scale crop and land cover mapping. *Remote Sensing of Environment*, 220, pp.135-51.

Hościło, A. & Lewandowska, A., 2019. Mapping Forest Type and Tree Species on a Regional Scale Using Multi-Temporal Sentinel-2 Data. *Remote Snsing*.

Hosseini, M. & McNairn, H., 2017. Using multi-polarization C- and L-band synthetic aperture radar to estimate biomass and soil moisture of wheat fields. *International Journal of Applied Earth Observation and Geoinformation*, 58, pp.50-64.

Hosseini, M. & Saradjian, M.R., 2011. Multi-index-based soil moisture estimation using MODIS images. *International Journal of Remote Sensing*, 32, pp.6799-809.

KNML, 2020. *Climate of the Netherlands*. [Online] Available at: https://www.knmi.nl/kennis-en-datacentrum/dossier/klimaat-van-nederland [Accessed 2020].

Koninklijk Nederlands Meteorologish Instituut, 2011. *De Bosatlas, van het klimaat*. Groningen: Noordhoff publishers/KNMi.

Kristof Van Tricht, A.G.S.G.I.P., 2018. Synergistic use of radar Sentinel-1 and Optical Sentinel-2 Imagery for crop mapping: a case studty of Belgium. *Remote Sensing*.

Kroes, J. et al., 2018. Impact of capillary rise and recirculation on simulated crop yields. *Hydrology and Earth System Science*, 22, pp.2937–52.

Lawrence, J., 2013. Distribution of the Median in Samples from the Laplace Distribution. *Open Journal of Statistics*, 3, pp.422-26.

Lu, D. & Weng, Q., 2007. A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, 28, pp.823-70.

Mountrakis, G., Im, J. & Ogole, C., 2011. Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66, pp.247-59.

Nathalie Pettorelli, J.O.V.A.M.J.-M.G.C.J.T.N.C.S., 2005. Using the satellite-derived NDVI to assess ecological responses to environmental change. *TRENDS in Ecology and Evolution*, 20, pp.503-10.

Noi, T.P. & Kappas, M., 2017. Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. *Sensors*, 18, p.18.

OECD, 2014. *The Space Economy at a Glance 2014*. OECD Publishing.

Oppermann, A., 2019. *Artificial Intelligence vs. Machine Learning vs. Deep Learning*. [Online] Available at: https://www.deeplearning-academy.com/p/ai-wiki-machine-learning-vs-deep-learning [Accessed 2020].

Peña-Barragán, J.M., Ngugi, M.K., Plant, R.E. & Six, J., 2011. Object-based crop identification using multiple vegetation indices, textural features and crop phenology. *Remote Sensing of Environment*, 115, pp.1301-16.

Rocchio, L., 2020. *Sentinel-2A Launches—Our Compliments & Our Complements.* [Online] Available at: https://landsat.gsfc.nasa.gov/sentinel-2a-launches-our-compliments-our-complements/ [Accessed 2020].

Schratz, P., 2018. *mlr.* [Online] Available at: https://mlr.mlr-org.com/articles/tutorial/handling_of_spatial_data.html [Accessed 2020].

Sicre, C.M., Fieuzal, R. & F.Baup, 2020. Contribution of multispectral (optical and radar) satellite images to the classification of agricultural surfaces. *International Journal of Applied Earth Observation and Geoinformation*, 84.

Sitokonstantinou, V. et al., 2018. Scalable Parcel-Based Crop Identification Scheme Using Sentinel-2 Data Time-Series for the Monitoring of the Common Agricultural Policy. *Remote Sensing*, 10, p.911.

Steinhausena, M.J., Wagnerc, P.D., Narasimhand, B. & Waskea, B., 2018. Combining Sentinel-1 and Sentinel-2 data for improved land use and land cover mapping of monsoon regions. *International Journal Applied Earth Observation and Geoinformation*, 73, pp.595-604.

STOWA, 2018. *Water Guide Agriculture: Instruments for Quantifying Effects of Water Management and Climate On Agricultural Production (in Dutch: Waterwijzer Landbouw: Instrumentarium voor Kwantificeren van Effecten van Waterbeheer en Klimaat op Landbouwproductie).* Amersfoort : STOWA.

Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote sensing of environment*, 8, pp.127-50.

van den Berg, F. et al., 2017. *An improved soil organic matter map for GeoPEARL_NL.* Wageningen: Wageningen university and Research.

Wanhui, C. et al., 2004. Monitoring the seasonal bare soil areas in Beijing using multitemporal TM images. *IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium*.

Wosten, J.H.M. et al., 2013. *BOFEK2012, the new soil physical schematisation of The Netherlands, (in Dutch: BOFEK2012, de nieuwe, bodemfysische schematisatie van Nederland)*. Wageningen: Alterra-rapport.

Xingwang Fan, Y.L., 2016. A global study of NDVI difference among moderate-resolution satellite sensors. *ISPRS Journal of Photogrammetry and Remote Sensing* , 121, pp.177-91.

8   Annexes

1.

Spectral bands (centre wavelength in nm/ SSD in m)	Mission objective	Measurement or calibration
B1 (443/20/60) B2 (490/65/10) B12 (2190/180/20)	Aerosols correction	Calibration bands
B8 (842/115/10) B8a (865/20/20) B9 (940/20/60)	Water vapour correction	
B10 (1375/20/60)	Cirrus detection	
B2 (490/65/10) B3 (560/35/10) B4 (665/30/10) B5 (705/15/20) B6 (740/15/20) B7 (775/20/20) B8 (842/115/10) B8a (865/20/20) B11 (1610/90/20) B12 (2190/180/20)	Land cover classification, Leaf chlorophyll content, Leaf water content, LAI (Leaf Area Index), FAPAR (Fraction of Absorbed Photosynthetically Active Radiation), Snow/ice/cloud, Mineral detection	Land measurement bands

Figure a: MSI spectral band specification. Source ESA

2.

Table 2: Counting of polygons per class from the BRP dataset before and after filtering

Class name	Class number	Polygons tot	Polygons after filter
Potatoes, consumption	0	16532	500
Potatoes, seed	1	9811	500
Potatoes, starch	2	7682	500
Corn, corncob mix	3	1615	494
Corn, energy	4	54	48
Corn, grain	5	4190	499
Corn, cut	6	79299	496
Corn, sugar	7	338	241
Wheat, winter	8	18228	500
Wheat, summer	9	4819	500
Beet sugar	10	16812	500
Winter barley	11	2158	499

Summer barley	12	7847	500
Tulip, flower bulbs and tubers	13	3262	357
Lily, bulbs and tubers	14	1504	463
Onions, sowing	15	4671	348
Onions, silver	16	89	89
Leek, winter, production	17	448	387
Leeks, winter, seeds and propagating material	18	102	100
Leek, summer, production	19	305	272
Leeks, summer, seeds and propagating material	20	46	42
others	21	489215	500

3. Code references

Function that creates the 12-days image composite from Sentinel1 and Sentinel2 imagery:

https://code.earthengine.google.com/3c3543f2668d65e95d1e6cd39d8dfb18

Random Forest Classifier algorithm:

https://code.earthengine.google.com/960baddc550596bed5c022475cb97a39

9 Acknowledgments