University of Tartu
Faculty of Science and Technology
Institute of Ecology and Earth Sciences
Department of Geography

Master's thesis in Geoinformatics for Urbanised Society (30 ECTS)

Clustering analysis of spatiotemporal Sentinel-2 data of agricultural parcels in Estonia for damaged crop delineation

Anton Kostiukhin

Supervisors:
Ph.D. Alexander Kmoch
Ph.D. Tanel Tamm
Ph.D. Indrek Sünter

Tartu 2022

Abstract

Clustering analysis of spatiotemporal Sentinel-2 data of agricultural parcels in Estonia for damaged crop delineation

The objective of the master's thesis was to perform an unsupervised classification of Sentinel-2 time-series data of agricultural parcels into zones with damaged and healthy vegetation. To achieve the objective, data from field surveys, reference information about the fields, parcel borders, and time-series of various optical vegetation indices which are based on Sentinel-2 imagery from the 2020-2021 agricultural season were used. Temporal variations were analyzed on a parcel-wide scale first, then division into zones was performed using a clustering algorithm, and lastly, pixel-based time-series classification was performed using decision trees.

The results show that on different time scales, NDVI, NDWI, TCA, and NDYI vegetation indices exhibit temporal variations between areas of healthy and damaged green vegetation, while PSRI provides comparable results for drought-induced crop damage and healthy vegetation, making its usage difficult for damage delineation. Clustering of time-series of vegetation indices revealed spatial patterns that generally coincide with observed damaged areas. The classification exploited variability in descriptive statistics between damaged and healthy zones, providing up to 73% overall accuracy on early damage classification if no distortions were introduced.

Keywords: clustering, agriculture, crop damage, time-series
CERCS code: T181 – Remote sensing

Abstrakt

Ajalis-ruumiliste Sentinel-2 andmete klasteranalüüs kahjustatud põllukultuuride piiritlemiseks Eesti põllumajandusmaal

Antud magistritöö eesmärgiks oli teostada põllumajanduslike maa-alade juhendamata klassifitseerimine Sentinel-2 aegridade abil kahjustatud ja terve taimestikuga vööndidesse. Eesmärgi saavutamiseks kasutati väliuuringute andmeid ning põldude, maatükkide piiride ja erinevate optiliste taimestikuindeksite aegridu, mis põhinevad 2020-2021 põllumajandushooaja Sentinel-2 kujutistel. Ajalisi variatsioone analüüsiti esmalt kogu maatüki skaalal, seejärel jagati maatükk rühmitusalgoritmi abil tsoonideks ja lõpuks viidi läbi pikslipõhine aegridade klassifitseerimine otsustuspuude abil.

Tulemustest selgus, et erinevatel ajaskaaladel näitavad NDVI, NDWI, TCA ja NDYI taimestikuindeksid tervete ja kahjustatud vegetatsiooniga alade vahel ajalisi erinevusi. Samas võib PSRI indeksi kasutamine taimekahjustuste piiritlemiseks olla raskendatud, sest see näitas sarnaseid tulemusi põua poolt kahjustatud ja terve vegetatsiooni vahel. Taimestikuindeksite aegridade klasterdamine tõi esile ruumilised mustrid, mis üldiselt langesid kokku vaadeldud kahjustatud aladega. Kirjeldav statistika näitas varieeruvust kahjustatud ja tervete tsoonide vahel, näidates, et klassifikatsiooni tulemused võivad kuni 73% täpsusega määrata varajase taimekahju kui puuduvad deformatsioonid põldudel.

Märksõnad: klasteranalüüs, põllumajandus, saagi kahjustumine, aegread
CERCS kood: T181 – Kaugseire

Table of Contents

1. Introduction

Currently, around 11% of the global land surface is used for agriculture purposes (Garcia Millan et al., 2020). As the world population grows and global warming changes the climate in arable lands (Carter et al., 2018), the pressure on the farmers increases since they are expected to fulfill the yield requirements and provide enough material for food production. Consequently, the agricultural sector is at great risk as unforeseen events might reduce the yield. To mitigate possible consequences and stay in business, farmers invest in crop insurance policies that reimburse monetary losses in case of low yield due to damaged crops. As of 2017, several European countries such as Spain, France, and Germany have more than 50% of agricultural parcels insured, while other EU members are in the process of establishing such policies (Baimisheva et al., 2019). Traditional on-the-ground methods of crop loss estimation are labor-intensive, biased (Gobbo et al., 2021), and imply substantial difficulties in case of parcels with an area of thousands of square meters. As a result, insurance companies search for ways to improve the process of loss estimation (Garcia Millan et al., 2020) and claims reporting. In this case, satellite data has the potential to be an efficient alternative to on-the-ground inspections, since optical instruments provide excellent data for existing highly developed methods for land monitoring (Tamm et al., 2016).

Sentinel-2 A and B are some of the most advanced non-commercial optical satellites monitoring the Earth's surface (Segarra et al., 2020). The pair of satellites provide high temporal resolution of European countries where revisit time may be as often as every 3 days with a spatial resolution of the images acquired though on board Multi-Spectral Instrument (MSI) reach 10 meters. The images consist of different bands that cover various parts of the spectrum. Utilizing the bands in the combination of spectral vegetation indices (VIs), that are sensitive to various crop parameters, and classification methods, it is possible to obtain up to 95% accuracy in crop classification tasks (Segarra et al., 2020). High accuracy is reached in many studies that apply supervised classification techniques to vegetation indices from Sentinel-2 with additional information in form of microwave vegetation indices from Sentinel-1 data (Phiri et al., 2020). However, this kind of approach requires an additional amount of data for training, and careful model development (Abdi, 2020), yet it does not guarantee a better overall performance compared with fast and less data demanding unsupervised approaches (Phiri et al., 2020). Moreover, long-term data unavailability due to power failure of the main instrument on board Sentinel-1B (Copernicus Open Access Hub, News, 2022) implies that existing models that rely on both Sentinel-1 and -2 might not be able to obtain necessary data anymore. To avoid the described problem and to reduce data dependency, it is possible to utilize only optical data.

There are numerous studies available that explore the usefulness of Sentinel-2 data in agriculture (Phiri et al., 2020), but only a limited number of papers focus on crop damage detection using Sentinel-2 data. In addition, almost no research is done in the scope of unsupervised classification techniques for crop damage detection on agricultural parcels. It is particularly true in case of Estonia, where only few studies have been carried out. Despite that drought conditions in Europe between 1949 and 2018 were mostly stable (Jaagus et al., 2022), detrimental effects of droughts (Páscoa et al., 2018) and a general increase in drought risk across mid-latitudes (Cook et al., 2018) create a negative picture for agricultural business.

Thus, the work aims to analyze Sentinel-2 time series data to perform an unsupervised classification into damaged and healthy crop zones within agricultural parcels. The classification is to be performed at the end of the agricultural season. Eventually, the classification is to be used as a part of an automatic damage loss compensation process. To reach the objective, the following research questions are defined:

1. Can an unsupervised algorithm detect variations in crop development within agricultural parcels?
2. Which optical vegetation indices provide the most information about crop damage?
3. What statistical parameters from time series of which vegetation indices can be used to perform the classification into healthy and damaged areas?

2. Theoretical overview

2.1. Satellite time series for crop monitoring

The popularity of Sentinel-2 grows as studies that utilize the data for different purposes are published each year. The interest in the data is conditioned by its quality: high spatial and temporal resolution with the availability of 13 different spectral bands open a wide range of opportunities for the analysis.

Generally, spectral bands are capable of estimating crop parameters, however, not all crop features can be directly sensed. In order to obtain specific biophysical properties of the crops, vegetation indices (VIs) in form of spectral transformations can be used (Segarra et al., 2020). This is one of the most popular techniques for vegetation monitoring and numerous studies have employed the indices themselves or other data derived from them on a temporal scale. Some studies, for example by Ghosh et al. (2018), focused on a small scale, using vegetation indices for crop monitoring within the agricultural parcels. It has been proved that the indices can be successfully applied for crop growth monitoring (Phiri et al. et al., 2020), as, for example, Normalized Difference Vegetation Index (NDVI) has a positive correlation with crop biophysical parameters (Ghosh et al., 2018). On a spatial inter-field scale, variations of the VIs can be utilized for crop damage detection (Ghosh et al., 2018) using, for example, classification techniques (Sosa et al., 2021) or yield estimation models (Gobbo et al., 2021).

Among the published studies of interest, that use classification techniques, the classification tasks are mainly performed with the help of supervised classification techniques that focuses on objects present in the images (Phiri et al., 2020). In this case, however, there are several major disadvantages, that include big data consumption along with high computational requirements (Enderle et al., 2005). A less popular alternative to supervised classification are pixel-based unsupervised classification methods, as they are fast, easy to apply (Sosa et al., 2021), and do not require big training datasets. K-means is one of the most widely used algorithms of the kind (Phiri et al., 2020; Sosa et al., 2021). Focusing on known cases of hail damage in Argentina, Sosa et al. (2021) performed a field classification into homogeneous damage zones (HDZs) applying k-means to standard deviation from a set of spectral vegetation indices before and after the date of the storm. The main drawback of the proposed method is that it focuses on cases with existing damage, and it does not perform classification into predefined classes, but rather into synthetic zones. It means that these clusters do not necessarily represent homogeneous damage zones within the fields, because in case of uniform conditions over the area, the clustering algorithm performs artificial division based on the assumption of the existence of three different damage zones. As a result of such classification, a supervised decision should be made to determine whether such division reflects a spatial pattern or not.

Vegetation indices and clustering algorithms are widely used for classification techniques, particularly, for crop damage identification. Nevertheless, few available studies did perform clustering into homogeneous zones with the aim to ease the work of insurance claims adjusters, but none of them focused on the classification into predefined user classes.

2.2. Optical satellite data

The Sentinel missions are a part of the Copernicus programme run by the European Space Agency under a Global Monitoring for Environment and Security (GMES) European Union initiative (Drusch et al., 2012). The Sentinel programme includes 7 missions with various radar and spectral instruments that are used for land, ocean, and atmospheric monitoring. A passive Multispectral Instrument (MSI) is one of them, onboard Sentinel-2 A and B satellites, launched in 2015 and 2017 respectively. The satellites have been designed to complement each other and provide frequent imagery with a revisit time of 5 days at the equator with identical illumination conditions (Drusch et al., 2012; Nico, 2019). The satellites operate at 786 km altitude on the Sun-synchronous orbit and acquire images over land and coastal areas from −56° to 84° latitude. The data acquired through MSI is divided into 13 bands in the visible, the near infra-red, and a short wave infra-red parts of the spectrum. Bands #2 (490nm – blue), #3 (560nm – green), #4 (665nm – red), and #8 (842nm – near-infrared) have 10m spatial resolution with coverage of ~100km x 100km, which makes them compatible with some other satellites and allows to perform precise land cover classification (Drusch et al., 2012). Spatial resolution of other bands varies between 20m and 60m (see Table 1) as they are utilized for analysis purposes other than vegetation:

Table 1. Sentinel-2 bands

Band number	Central wavelength, nm	Spatial resolution, m	Primary usage
1	443	60	Cloud screening and atmospheric correction
2	490	10	Vegetation monitoring
3	560	10	
4	665	10	
5	705	20	Snow/ice/cloud detection
6	740	20	
7	783	20	
8	842	10	Vegetation monitoring
8a	865	20	Snow/ice/cloud detection
9	945	60	Cloud screening and atmospheric correction
10	1380	60	
11	1610	20	Snow/ice/cloud detection
12	2190	20	

Sentinel-2 data is available in cartographic UTM projections, depending on the location, and is primarily provided at two different processing levels: Level-1C (L1C) or Top-Of-Atmosphere (TOA) and Level-2A (L2A) or Bottom-Of-Atmosphere (BOA) (Drusch et al., 2012; Domnich et al., 2021). The main difference between these two processing levels is that L2A products have removed atmospheric effects, and they have been corrected using a Digital Elevation Model (DEM) (see Figure 1).

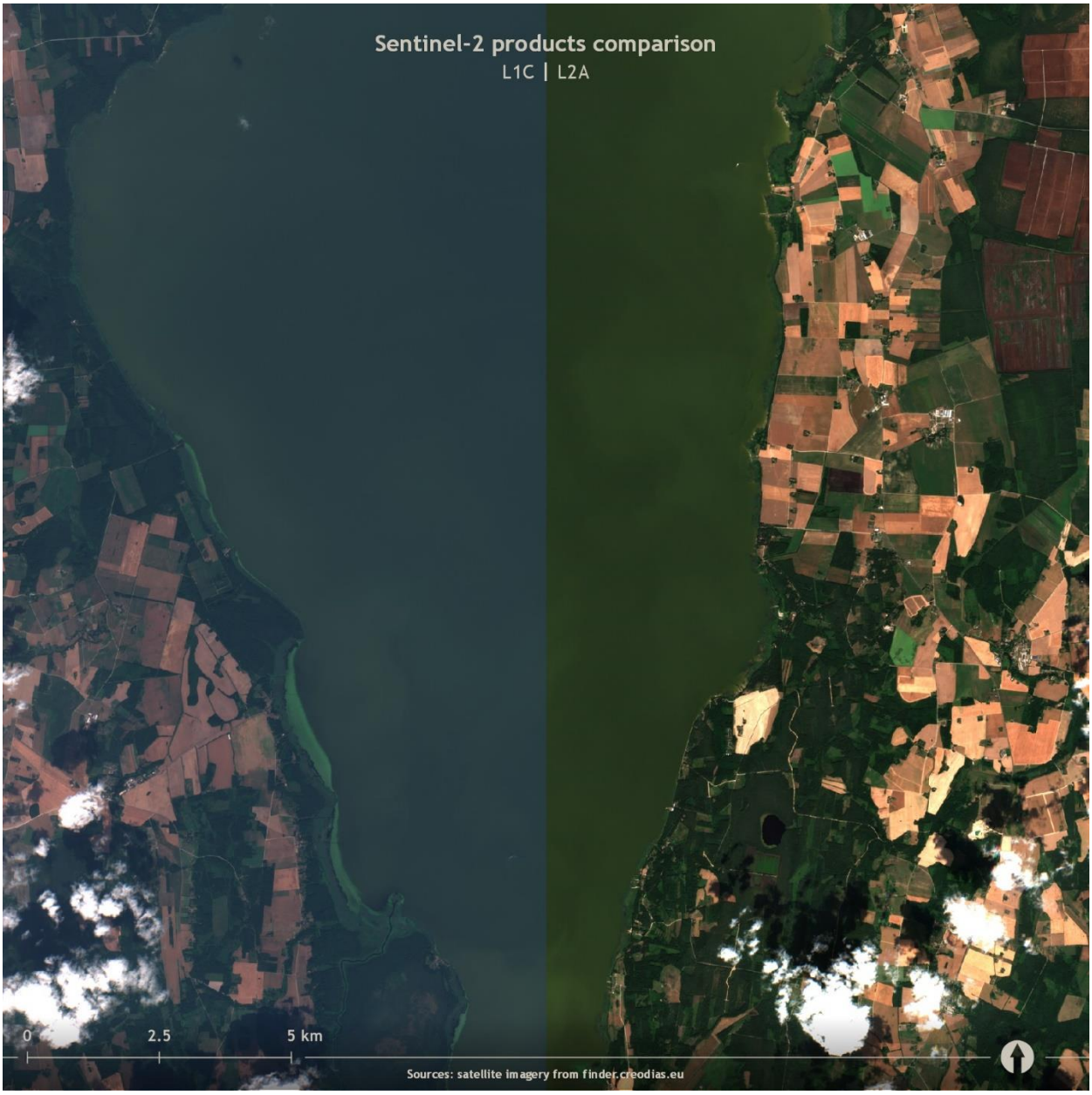


Figure 1. Sentinel-2 L1C (left) and L2A (right) product types comparison

2.3. Vegetation indices

Since optical remote sensing provides a range of bands that cover various parts of the spectrum, it makes it a convenient instrument for vegetation monitoring. Unique characteristics of green plants such as moisture and concentration of chlorophyll A and B allow exploiting these, as in this case, visible radiation in the red channel (665nm) and the main part of the blue channel (490nm) is absorbed by leaves, while radiation in the green (560nm) part of the spectrum is reflected (Bannari et al., 1995; Sharma et al., 2015). The reflectance intensity in optical sensors is primarily determined by the properties of the observed features on the ground (Sharma et al., 2015) (see Figure 2):

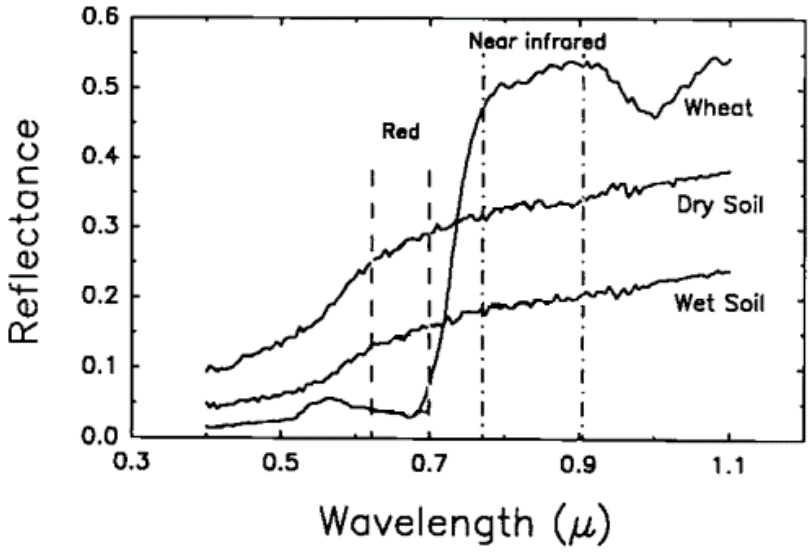


Figure 2. Reflectance intensity of red and NIR bands of LANDSAT TM for different features
Source: Adapted from Jackson et al., 1991

As single spectral measurement does not always provide enough information about the observed vegetation (Segarra et al., 2020) as well as the values might be affected by adjacent features (Bannari et al., 1995), the bands are utilized in form of vegetation indices (VIs). The indices use data from several spectral bands in form of rationing, differencing, or sums in a way, so the result enhances the visual difference between plants and their surroundings (Jackson et al., 1991), as well as it reduces the influence of soil brightness, shadows, and other environmental effects (Bannari et al., 1995). As of now, there are more than 40 various indices available, that allow the assessment of various parameters and general condition of greenery (Bannari et al., 1995). Within this study, five widespread indices that proved to provide the highest amount of information about vegetation conditions in the fields were selected.

Normalized Difference Vegetation Index or NDVI, is one of the most utilized coefficients for monitoring land cover using satellite imagery (Sulik et al., 2016) because it is related to the amount of chlorophyll in the plants, and it is linearly correlated with vegetation density (Zhang et al., 2013). The index assesses the quality of vegetation by measuring the difference between near-infrared and red parts of the spectrum (Sosa et al., 2021):

$$NDVI = \frac{B_{842} - B_{665}}{B_{842} + B_{665}} \quad \text{(Eq. 1)}$$

where B_{842} and B_{665} represent reflectance in near-infrared (842nm) and red (665nm) channels respectively (Tang et al., 2015). The final value ranges from -1 to 1, where high values represent areas with dense and healthy vegetation. More precise interpretation, however, depends on many factors, among which are geographical location, soil, and vegetation type (Carlson et al., 1997; Gutman et al., 1998; Zhang et al., 2013). Table 2 reflects NDVI values for bare soil ($NDVI_{soil}$) and vegetation ($NDVI_{veg}$) obtained from various studies from different parts of the world:

Table 2. Study areas and used NDVI values

Reference	Coverage	$NDVI_{soil}$	$NDVI_{veg}$	Data
Carlson et al. (1997)	State College, PA (approximately 40" N and 76" W)	-0.1…0.2	0.5...0.8	Not specified: satellite with zenith angles from 0 to 20 degrees from nadir
Gutman et al. (1998)	Global	0.04	0.52	NOAA/AVHRR data
Sobrino et al. (2000)	Morocco	0.2	0.5	NOAA/AVHRR multi-temporal imagery
Zeng (2000)	Global	0.04	0.52, varies depending on the land cover	AVHRR; International Geosphere-Biosphere Program (IGBP) land cover classification
Jiapaer et al. (2011)	China, Taklamakan desert outskirts	0.1	0.6	Cloud-free Landsat 5 TM images; MODIS L1B; FVC in situ data
Zhang et al. (2013)	Western China	0.071	0.641	HJ-1/HIS hyperspectral imager; FVC in situ data

NDVI, however, is not reliable in case of flowering vegetation that changes its color within a growing season. Presence of any other color apart from green will result in low NDVI values which might be misinterpreted. To perform a more accurate estimation of yellow color, and thus estimate crops health at the flowering stage, it is possible to use Normalized Difference Yellowness Index or NDYI:

$$NDYI = \frac{B_{560} - B_{490}}{B_{560} - B_{490}} \quad \text{(Eq. 2)}$$

where B_{560} and B_{490} represent reflectance in green and blue parts of the visible spectrum respectively (D'Andrimont et al., 2020). Since NDYI is sensitive to the amount of yellow, it overcomes drawbacks of NDVI when applied to flowering rapeseed. During flowering, NDYI is expected to increase, while the relative level should be higher after the flowering stage, which confirms the growth of a reproductive canopy. The index is crucial for rapeseed health estimation as the crop is the most vulnerable to external factors during flowering, while the amount of flowers linearly affects the yield (Sulik et al., 2016; D'Andrimont et al., 2020). Among different studies, NDYI level of healthy vegetation varies from 0.25 to 0.59 (Zhang et al., 2011; D'Andrimont et al., 2020).

Normalized Difference Water Index or NDWI stands for two different indicators: the first one was proposed by Gao (1996) to observe changes in the water content of greenery; the second one was proposed by McFeeters (1996) to record changes in waterbodies. Since waterbodies are not a part of this work, the index defined by Gao (1996) is used:

$$NDWI = \frac{B_{842} - B_{1610}}{B_{842} + B_{1610}} \quad \text{(Eq. 3)}$$

where B_{842} represent reflectance in near-infrared wavelengths, B_{1610} in short-wave infrared. According to Gao's (1996) work, while NDVI values over 0.63 become insensitive to the amount

of greenery, NDWI values vary, which indicates the sensitivity to the amount of liquid in vegetation. Gao (1996) defines NDWI values for the High Plains area in Colorado (USA) as -0.022, 0.084, and 0.215 for bare soil, grass, and crops respectively. Other studies, for example, Gu et al. (2007), who focused on the Flint Hills area (USA), state that NDWI values below 0.3 represent drought conditions.

To detect the maturation phase of plant development, a Plant Senescence Reflectance Index was introduced by Merzlyak et al. (1999). The index utilizes red, blue, and red-edge bands to detect the process of senescence:

$$PSRI = \frac{B_{665} - B_{490}}{B_{740}} \quad \text{(Eq. 4)}$$

where B_{665} and B_{490} represent reflectance in visible red and blue parts of the spectrum correspondingly, B_{740} represent reflectance in the red-edge band. The result belongs to the -1…1 interval, where values from -0.1 to 0.2 represent green vegetation (Zhang et al., 2018), above 0.2 – plants at the senescence stage. Das et al. (2015), however, defines thresholds differently, stating that for green vegetation the value generally ranges from -0.2 to 0.4. The indicator can be especially useful during later stages of plant development, where NDVI would be difficult to interpret due to similar values for soil and vegetation.

A group of useful spectral indices for the classification of different landscape features as well as vegetation monitoring (Xue et al., 2017) can be obtained from a linear combination of existing bands (Mostafiz et al., 2018) using Tasseled Cap Transformation (TCT). TCT creates a new set of indices - TC Brightness (TCB), TC Vegetation (TCV), and TC Wetness (TCW). The indices are calculated based on the sensor-specific derived coefficients. The coefficients that provide the highest accuracy in case of Sentinel-2 MSI data are presented in Table 3 (Shi et al., 2019).

Table 3. Sentinel-2 Tasselled Cap Coefficients
Source: Adapted from Shi et al., 2019

TC Feature	TC coefficients for corresponding Sentinel-2 Bands					
	Band 2	Band 3	Band 4	Band 8	Band 11	Band 12
TCB	0.3510	0.3813	0.3437	0.7196	0.2396	0.1949
TCV	-0.3599	-0.3533	-0.4734	0.6633	0.0087	-0.5856
TCW	0.2578	0.2305	0.0883	0.1071	-0.7611	-0.5308

TC Brightness (TCB) and TC Vegetation (TCV) are of particular interest: TC Brightness is a measure of the ground where soil corresponds to higher values of the index (Kauth et al., 1976). Differences in soil characteristics have the biggest influence on the index, while vegetation cover affects the least:

$$TCB = X_2 * B_2 + X_3 * B_3 + X_4 * B_4 + X_8 * B_8 + X_{11} * B_{11} + X_{12} * B_{12} \quad \text{(Eq. 5)}$$

where X_n stands for the coefficient corresponding to the TCB index and Sentinel-2 band number B_n in Table 3.

Tasseled Cap Vegetation (TCV) captures a unique reflectance pattern of green vegetation and thus is defined as a contrast between the aggregation of visible bands and near-infrared bands (Crist et al., 1984):

$$TCV = X_2 * B_2 + X_3 * B_3 + X_4 * B_4 + X_8 * B_8 + X_{11} * B_{11} + X_{12} * B_{12} \quad \text{(Eq. 6)}$$

where X_n stands for the coefficient corresponding to the TCV index and Sentinel-2 band number B_n in Table 3.

Combined in pairs, TCT-derived indices form various panes (Mostafiz et al., 2018) where the vegetation pane (see Figure 3) is a combination of TC Vegetation and TC Brightness. On the pane, the indices form Tasseled Cap Angle (TCA) which reflects the proportion between vegetation and non-vegetation per unit area (Gómez et al., 2011):

$$TCA = \arctan\left(\frac{TCV}{TCB}\right) \quad \text{(Eq. 7)}$$

where *TCV* is TC Vegetation, *TCB* is TC Brightness.

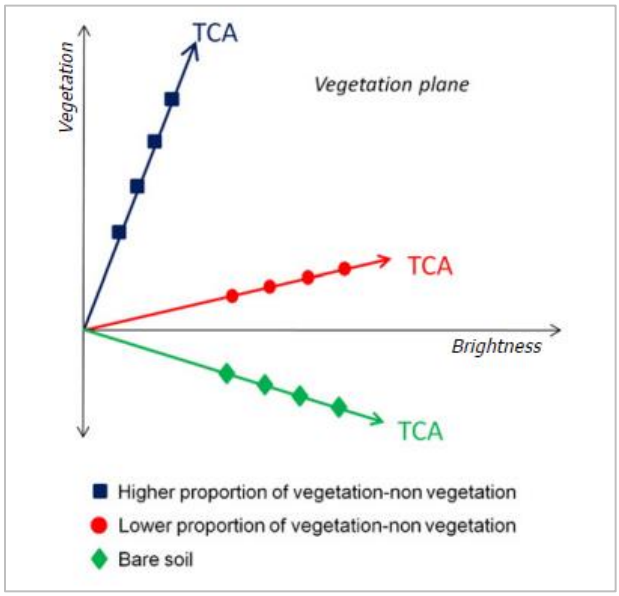


Figure 3. Vegetation plane from TC Vegetation and TC Brightness
Source: Adapted from Gómez et al., 2011

More concrete interpretation of the values depends on a scene (Crist et al., 1984), but generally negative TCA value stands for bare soil (Gómez et al., 2011), while the global distribution of land cover types in the vegetation plane can be seen in Figure 4 (Zhang et al., 2002):

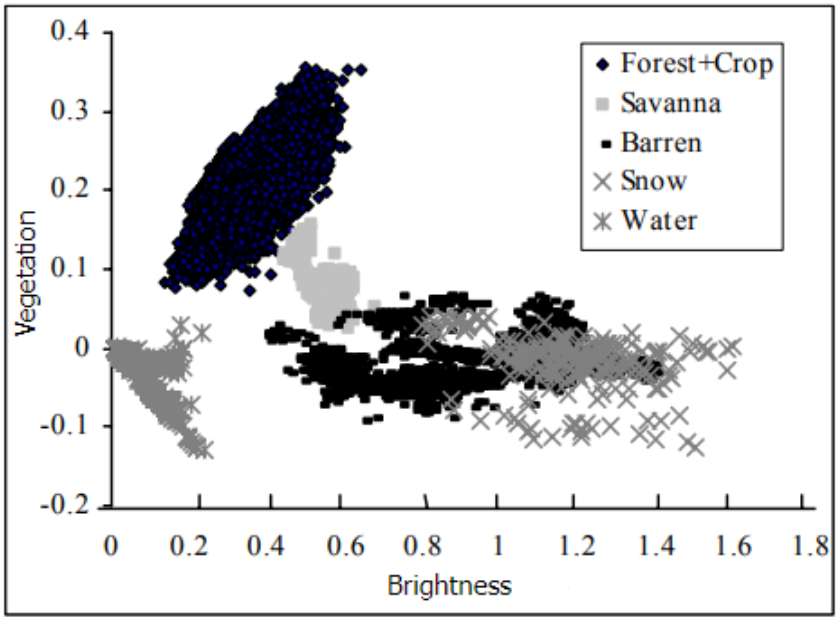


Figure 4. Global distribution of land cover types in the vegetation plane
Source: Adapted from Zhang et al., 2002

2.4. Crops ontogenesis

After being planted, crops start their lifecycle and during the process of growing, change their morphological structure at different phases of development (Komarov et al., 2003). Generally, each period is characterized by a unique set of features. Since optical remote sensing acquires images at various wavelengths, color or structural changes directly affect the values. Awareness of these natural phases plays a crucial role in differentiating between natural developments in plants' structure and deviations in the fields.

According to Skellern et al. (2017), rapeseed is planted in early July and is expected to be harvested in mid-July the following year (see Figure 5):

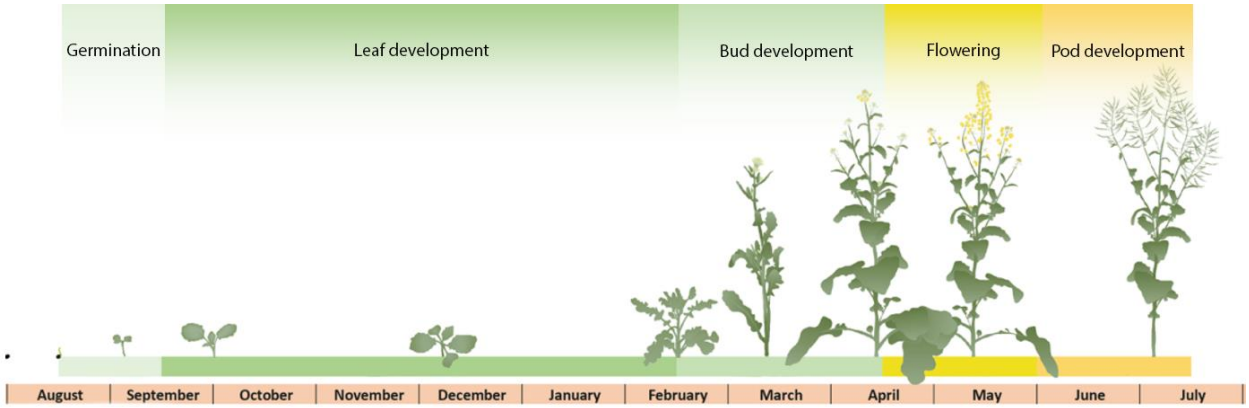


Figure 5. Rapeseed development stages
Source: Adapted from Skellern et al., 2017

The actual time of planting and harvesting varies depending on the weather conditions and geographical location, however, the most important morphological developments occur in the same order. As plants become greener and taller, NDVI value increases. At the beginning of May rapeseed enters the phase of flowering, which can be characterized by a rapid and significant decrease in NDVI level (D'Andrimont et al., 2020), because yellow flowers contribute to red light (Sulik et al., 2016). This stage is crucial in rapeseed yield estimation because low solar irradiance or deficiencies in water supply result in a reduced number of flowers, which causes a smaller number of pods to form (D'Andrimont et al., 2020). As a result, a low amount of flowers indicate unhealthy developing plants. Flowering can be detected by using vegetation indices that are sensitive to changes in the amount of yellow, for example, Normalized Difference Yellow Index (see Eq. 2) (D'Andrimont et al., 2020). After flowering (see Figure 6), the crops develop a reproductive canopy and enter the last stage of pod development becoming brown over time.

Figure 6. Rapeseed growing stages: (a) pre-flowering, (b) flowering, (c) post-flowering
Source: adapted from D'Andrimont et al., 2020

Winter wheat and spring barley can be characterized as plants that "green-up" and then "green-down" (Sulik et al., 2016). Since they do not have a flowering stage, the NDVI curve is more stable across the lifespan (see Figure 7), which makes the index very useful for estimating the health of the crop, as deviations might indicate non-homogeneous developments (Wang et al., 2021):

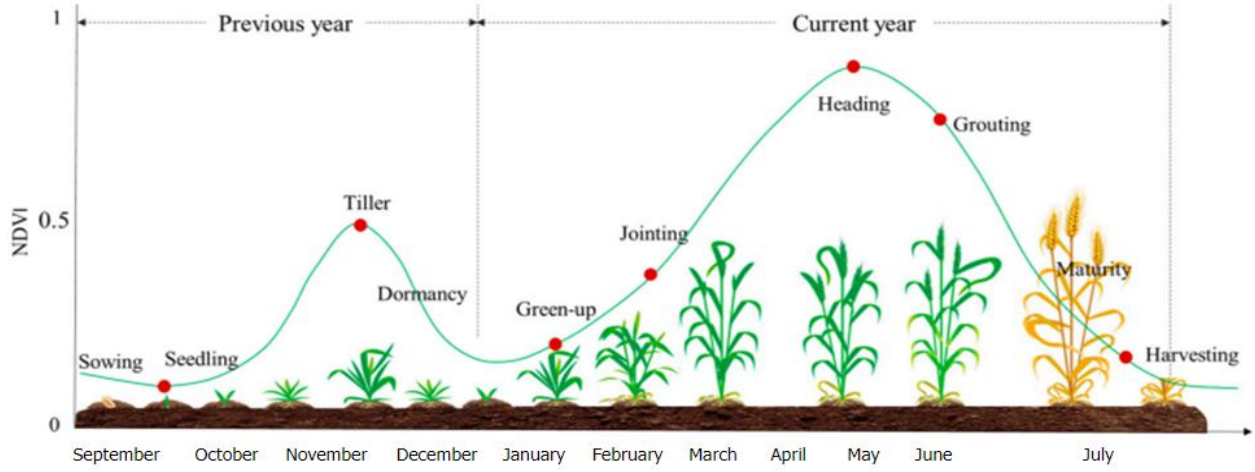


Figure 7. Winter wheat growth cycle and relative NDVI values
Source: adapted from Wang et al., 2021

The main growing stages of Winter Wheat after dormancy are presented in Figure 8:

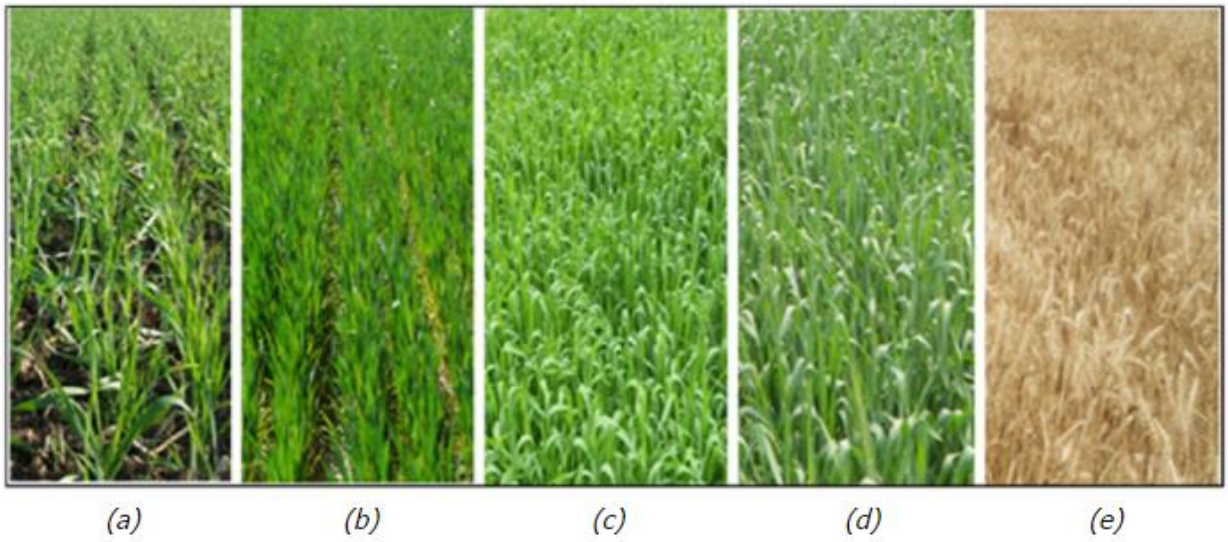


Figure 8. Major stages of winter wheat growing: (a) tillering, (b) jointing, (c) booting, (d) heading, (e) ripening
Source: Adapted from Kumar et al., 2017

2.5. Self-Organizing Maps (SOM) clustering

The analysis approaches of previous studies that utilized k-means clustering used statistical parameters of the time series, such as standard deviation (Sosa et al., 2021), to perform classification into damage zones. The approach of this work focuses on the temporal changes in the crop growth, and therefore, self-organizing maps (SOM) algorithm was chosen. The method and its modifications were successfully applied for classifying time series data in previous studies (Dresp-Langley et al., 2018; Fortuin et al., 2018), including the Earth observation sector, as SOM provides an effective approach for dust sources identification on satellite images (Miljković, 2017).

A self-organizing map is a two-layers (input and computational layers) artificial neural network which can produce an easily visualizable low dimensional representation of high-dimensional data (see Figure 9) (Miljković, 2017):

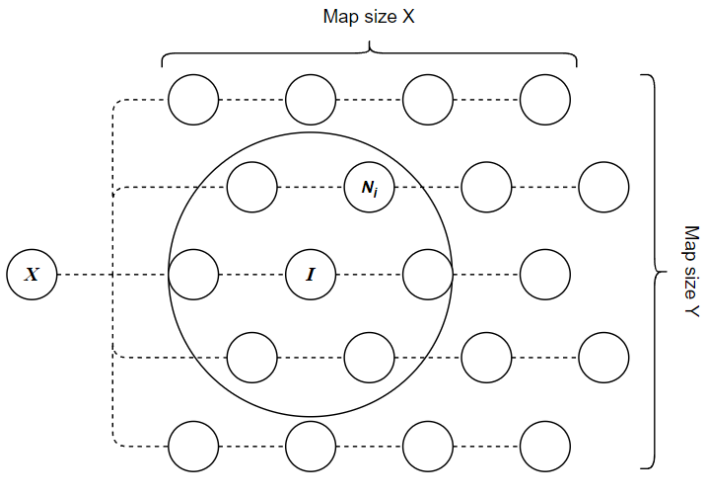


Figure 9. SOM example network: (*X*) input data element, (*I*) winning neuron, (N_i) index number of a neuron
Source: Adapted from Kohonen, 2013

For SOM to provide a result, a model should be trained first. The general working principle includes several steps: 1) based on the input data, weights for each neuron (N_i) are initialized; 2) step (t) is set to 0; 3) for a random input data element (*X*) the winning neuron (*I*) (also called best matching unit 1 (BMU1)) is determined using its weights (w_I):

$$\underset{I}{\operatorname{argmin}} ||X - w_I|| \qquad \text{(Eg. 8)}$$

4) weights of each neuron (N_i) are adapted using predefined $\eta(t)$ *learning rate* and *neighborhood function* $h(I)$:

$$w_n = w_n + \eta(t)h(I)(X - w_n) \qquad \text{(Eq. 9)}$$

5) t increased by 1 and if $t < t_{max}$ go to step 3, where t_{max} is the total number of data entries.

As a result of one iteration, a data element is placed on the map. Neurons that lie within the neighborhood of *I* neuron (circle in Figure 9) match better with the input *X* than ones beyond the neighborhood distance. Thus, more similar data entries will be associated close to each other, while less similar will be placed further away on the map.

The optimal settings depend on the different parameters, such as the size of the dataset and the length of the time series. The number of clusters is generally determined by "a rule of thumb", where the size of the grid for a dimensionality reduction task would be determined as:

$$S_x * S_y = 5 * \sqrt{N} \qquad \text{(Eq. 10)}$$

where S_x and S_y are X and Y sizes of the grid, *N* is the number of samples in the dataset (Shalaginov et al., 2015). In case the goal is to get the most optimal results, the learning rate, neighborhood distance, and the number of iterations can be adjusted with the help of metrics that can be provided by various SOM implementations (Vettigli, 2018).

3. Data and methods

3.1. Study area

Study areas are located in Southern Estonia. The climate of the country is generally moderate and varies between inland and coast parts of the country. Spring is mild with an average temperature of around 5°C, with rains in late April. Summer is generally warm, with an average temperature of around 20°C, but can be up to 30°C, while southern parts of the country are warmer than the seaside and islands. In autumn average temperature decreases to ~7°C, with strong winds and heavy precipitation. The average temperature in the wintertime is just below 0°C, however -10°C to -20°C frostbites are quite common (Estonian Environment Agency, 2022). Most agricultural activities are performed within a season between April and September.

For the research, 23 agricultural fields with various crop types in different conditions around Tartu city were chosen (see Figure 10). Eight of the fields represent winter wheat, eight rapeseed, and seven spring barley in various conditions:

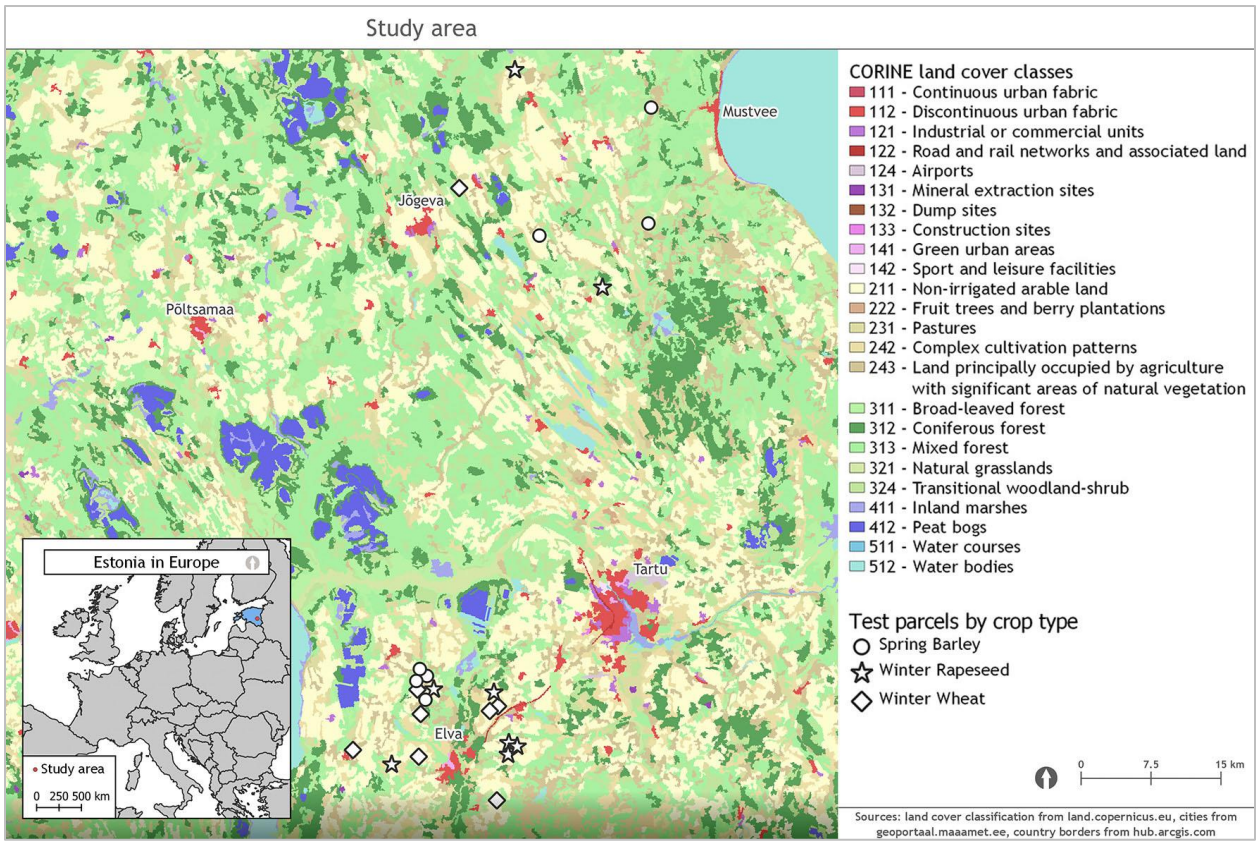


Figure 10. Study area CORINE (2018) land cover map

3.2. Data

3.2.1. Satellite images

The primary data utilized in the research is Sentinel-2 time series over Tartu city (Estonia) vicinities. Originally the satellite images (products) are provided by ESA and can be obtained

through various services, e.g. Creodias (Creodias, 2022) or Sentinel Hub (Sinergise Ltd., 2022). The dataset consists of 73 L2A products, where the date of the earliest product is 2021-04-01, the latest is 2021-09-28. These dates cover the agricultural season and provide some time buffers before and after the growth period, that help to understand time patterns better. The spatial resolution of images varies, as is shown in Table 1. Temporal resolution varies from 2 to 3 days as it is shown in the figure below (see Figure 11), but depending on the exact location up to ~60% of acquisitions can be obscured by the clouds.

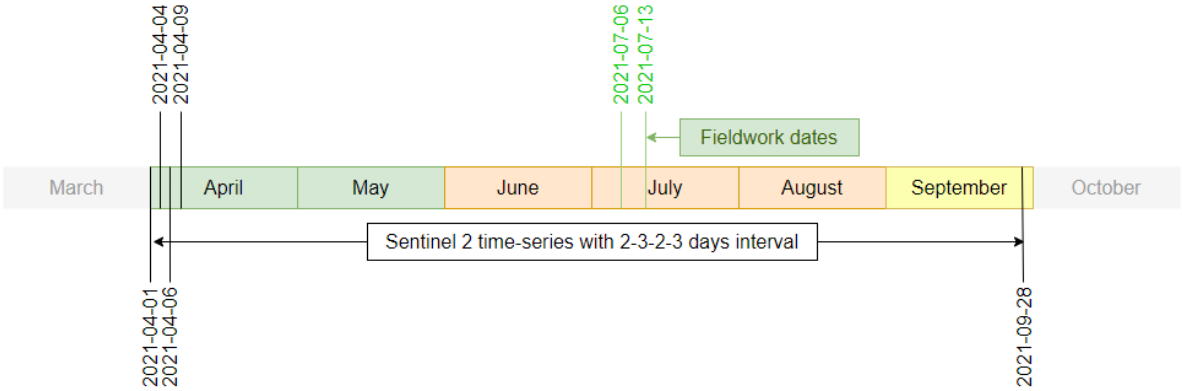


Figure 11. Dataset coverage and fieldwork dates

In addition to the satellite imagery, information about plowing and crop harvesting dates for each field is present and utilized for the data analysis.

3.2.2. Field surveys

Since the resolution of the satellite images imposes some limitations on the accuracy of the estimations that are based purely on the images, in-person trips to one of the test fields were organized. Agricultural parcel #55 has a 50.33 ha area of winter wheat, with visible damage in the central part of the field and along its west and north borders (see Figure 12).

Figure 12. Sentinel-2 true-color image of agricultural parcels #57 and #55, July 5th, 2021

To better understand the nature of the damage as well as its size, on-the-ground images were taken. Below, example images from the 6th of July 2021 at about 17:30 using a smartphone camera are provided. All images taken are georeferenced to ease data analysis and description. From an on-the-ground perspective, it was possible to estimate the density of healthy vegetation (see Figure 13) and damaged areas (see Figure 14 and Figure 15). Along the edge of the parcel (see Figure 14), the difference between healthy vegetation and dry damage areas can be seen. There is no clear border between the zones, as the number of healthy crops gradually decreases over the distance of 3 to 4 meters:

Figure 13. The density of healthy winter wheat, Parcel #55, Northern side, July 6th, 2021

Figure 14. The damaged western side of the parcel #55, July 6th, 2021

Figure 15. Broken and dry stems of winter wheat, Parcel #55, Western side, July 6th, 2021

It was not possible to see the whole field from the ground due to slight elevation in the central part and high-density healthy vegetation, which obstructs the movements and visibility. Thus, to obtain a general picture of the area, airborne images were taken using a drone (DJI Mavic 2 Pro, UT Geography department). Examples from the 13th of July 2021 at 13:00 are provided below (see Figure 16, Figure 17).

Figure 16. Drone image of the field #55, Northern direction, July 13th, 2021

Figure 17. Parts of the central damaged area of the field #55, Southern direction, July 13th, 2021

In the figures provided, the shape of the damaged traces on the ground resembles a tractor's axle track, which indicates a possible presence of agricultural vehicles along the lines. The pattern matches among on-the-ground observations, aerial imagery, and satellite images used for the analysis.

Overall, 9 different aerial scenes for parcels #55 and #57, and 30 various on-the-ground compositions for multiple agricultural parcels were taken.

3.2.3. Estonian Land Board aerial imagery

As it is seen in section 3.2.2. Field surveys, for a drone it is problematic to cover the whole agricultural field with an area of 50 ha. In addition, inspecting each field in person takes a sufficient amount of resources. To facilitate the process of data analysis and damage estimation, oblique aerial photographs provided by the Estonian Land Board (Estonian Land Board Orthophoto, 2021) were used. The images cover most agricultural parcels of interest from the end of June until mid-July. An example of a complete image of parcels #55 and #57 can be seen in Figure 18, parcel geometries in Figure 12.

Figure 18. Oblique aerial photograph of parcels #57 (upper-left) and #55 (center), June 9th, 2021
Source: Estonian Land Board Orthophoto, 2021

Because images are in high resolution and were taken from different angles, it is possible to see a detailed picture of damaged areas within agricultural parcels.

3.2.4. Reference data

Additional textual information was obtained from KappaZeta Ltd., which includes the following information for some parcels and crop types: field area, crop type, and sort, date of sowing, and date of harvesting.

3.3. Methodology

The methodology of the current study consists of several steps that include data pre-processing and analysis techniques applied to the data. A general overview of the methodology is provided in Figure 19, while a more detailed description of the steps performed is provided in the paragraphs of the current section.

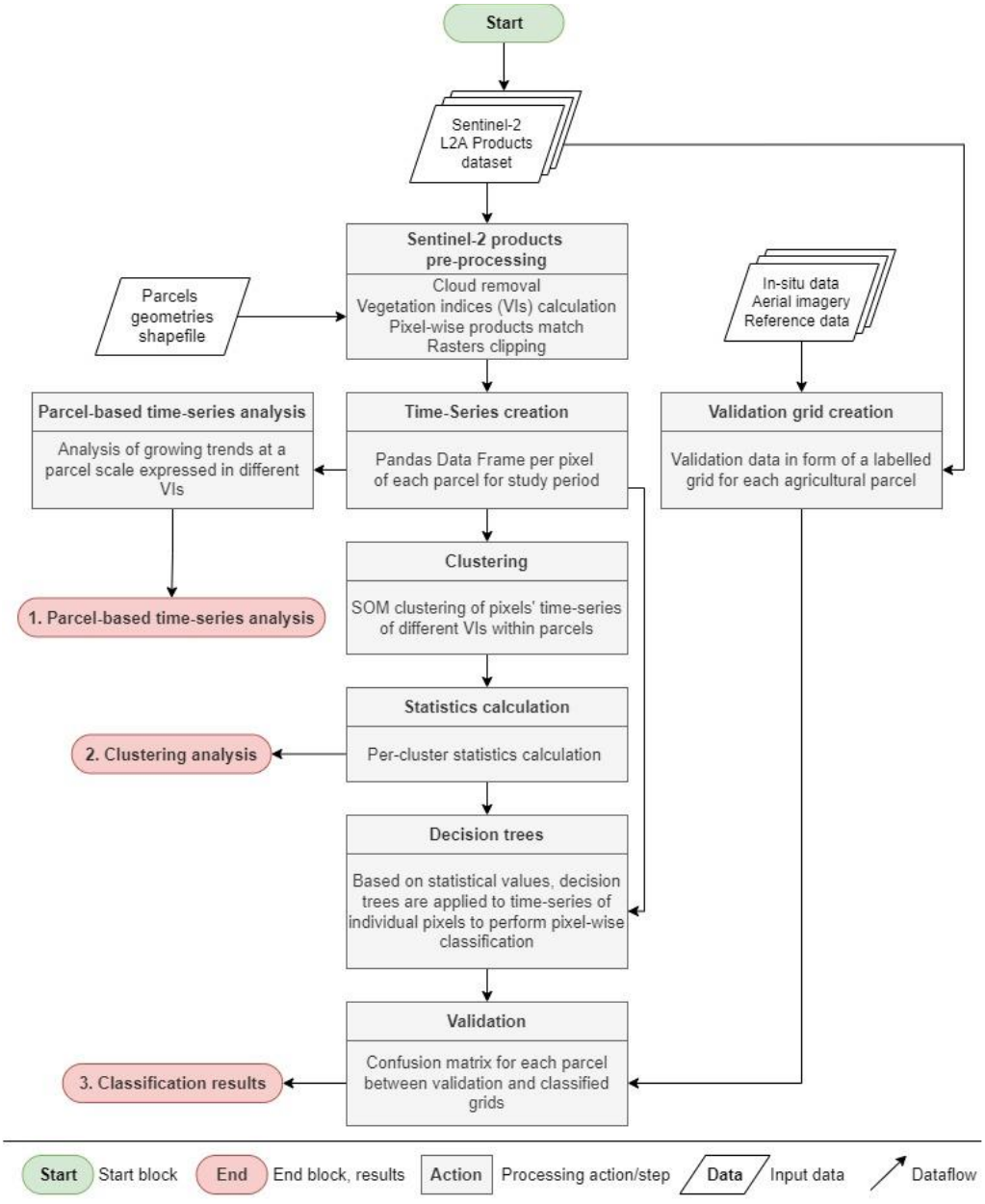


Figure 19. Methodology workflow

3.3.1. Sentinel-2 products pre-processing

The data was obtained from ESA in form of .SAFE (Sentinel Online, 2022) L2A products, which were partly processed in KappaZeta Ltd. Bands with lower than 10m resolution were resampled to 10m spatial resolution. To remove unwanted distortions on the images, KappaMask (Domnich et al., 2021) cloud mask algorithm that masked out cloud and cloud shadow pixels was applied. In addition to the default set of bands, vegetation indices were calculated:

1) NDVI (Eq. 1)
2) NDWI (Eq. 3)
3) PSRI (Eq. 4)

The calculated indices were added as additional bands to multiband raster files in .tiff format. Other vegetation indices – NDYI (Eq. 2), TCB (Eq. 5), TCV (Eq. 6), TCA (Eq. 7) were calculated using GDAL library under QGIS 3.22.3 Python environment.

To make the rasters ready for analysis, several extra processing steps were performed. Firstly, it was made sure that rasters obtained from different dates match each other pixel-wise. For that, a comparison of rasters extents and pixel sizes was performed. This step is necessary to ensure that during the time series creation pixels will not be shifted and, consequently, replaced by values from another pixel. Since all rasters matched each other, no extra actions were performed. Secondly, the rasters were clipped by the fields' geometry to keep the areas of interest. Visual inspection was performed to make sure the area was cut correctly, and no unrelated zones were present within the field area. In cases when parcel geometries covered unwanted areas, such as roads or forests, the polygons were manually corrected.

3.3.2. Time series creation

At this step, time series were created within individual agricultural parcels by taking a pixel number and extracting vegetation indices values from the available rasters from different dates. The result was a pandas data frame with "pixel_id", "date", and "value" columns. For pixels where data is not available due to the clouds, a "no data" value was used.

3.3.3. Time series gaps filling algorithm

Due to the fact that some of the areas were masked out, each parcel has its uneven distribution of the raster values over the period, which might lead to wrong data interpretation. This step does not have a separate entity in the methodological workflow provided in Figure 19, but it is performed at the beginning of "Parcel based time series analysis", "Clustering", and "Decision trees" stages to ensure that time series have identical length and uniform temporal sampling. For that, the "no data" value gaps of time series were filled by linear interpolation, while values at the beginning and end of the period by nearest-neighbor extrapolation (Figure 20). For interpolation and extrapolation no spatially neighboring pixels were considered.

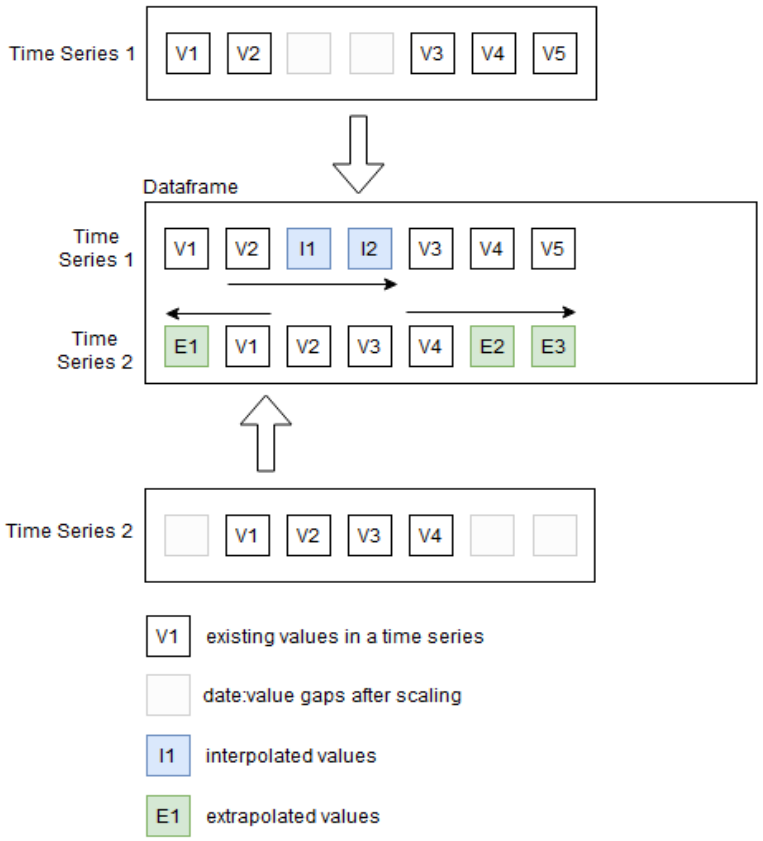


Figure 20. Time series gaps filling algorithm (Time Series 1-2 are examples of time series for individual pixels within a parcel area)

3.3.4. Parcel-based time series analysis

To see general patterns across the croplands, parcel-wide time series were created for selected vegetation indices using an average over all pixel values within individual agricultural parcels. To ensure correct data visualization and interpretation, the gaps in the data were preliminarily filled using time series gaps filling algorithm described in section 3.3.3. Visual inspection and interpretation of the results are described in corresponding results sections.

3.3.5. Clustering

At this step, clustering of previously created pixels' time series was performed using a self-organizing maps (SOM) clustering approach. Particularly, MiniSOM (Vettigli, 2018) implementation of the SOM algorithm was applied to time series of individual pixels within parcels for each vegetation index separately. The settings of the model were set as follows: the size of the map is 3x1 – 3 clusters, as it is the number of zones visually defined by insurance claims adjusters (Sosa et al., 2021); a maximum number of iterations equals the number of pixels in a grid for each field to ensure that every time series would be observed by the model at least once; learning rate – 0.15; sigma – 0.5; time-window interval used for clustering varies depending on the crop type and vegetation index:

1) For winter wheat fields time series of NDVI, NDWI, and TCA indices are clustered from 2021-04-01 to 2021-07-25, so they focus on crops development tendency over the whole growing period; PSRI is clustered from 2021-06-03 to 2021-08-07, so it covers the ripening stage; NDYI is not used for the clustering of this crop type because winter wheat is not a flowering plant.
2) For winter rapeseed time series of NDVI, NDWI, and TCA indices are clustered similarly to winter wheat, from 2021-04-01 to 2021-07-25; PSRI is clustered from 2021-06-20 to 2021-08-14, during senescence; NDYI is clustered during flowering, from 2021-05-16 to 2021-06-18.
3) For spring barley time series of NDVI, NDWI, and TCA indices are clustered from 2021-05-01 to 2021-07-30; PSRI is clustered from 2021-06-23 to 2021-08-14; NDYI is not used for the clustering of this crop type because spring barley is not a flowering plant.

To ensure correct data visualization and interpretation, the gaps in the time series data were preliminarily filled using time series gaps filling algorithm described in section 3.3.3.

3.3.6. Statistics calculation

For each generated cluster a set of statistical parameters was calculated using the *NumPy* python package. Statistics included standard deviation, mean, median, min, max, and accumulation, which were averaged over clusters' entries. In addition, the slope-intercept trend line (Eq. 11) per each cluster was calculated using the original Mann-Kendall test from the *pyMannKendall* python package using statistical significance value $p = 0.05$:

$$y = mx + b \quad \text{(Eq. 11)}$$

where y is y and x is x coordinate in a cartesian coordinate system, m – slope, b – y-intercept.

For Mann-Kendall test the time series data was supplied as a series of values without corresponding dates.

3.3.7. Decision trees

To perform scene classification, simple decision trees based on if-else conditions for vegetation indices were created. Threshold values used in the decision trees were defined based on the statistics obtained from a previous step considering the theoretical values outlined in other sources. The decision trees were applied to pixels' time series of various lengths, depending on the clustering intervals. To ensure correct data visualization and interpretation, the gaps in the time series data were preliminarily filled using time series gaps filling algorithm described in section 3.3.3. The classification was performed into three classes: 1 – "bare soil", 2 – "medium damage", 3 – "no damage".

3.3.8. Validation grid creation

Based on the satellite images, in-situ observations, drone images, and oblique aerial photographs, validation data was created manually. Validation data has been created for each agricultural parcel in form of a grid, that spatially corresponds to Sentinel-2 imagery pixels at 10m resolution, while, in addition, each cell has a number that corresponds to the image pixel id underneath the grid. The cells were manually labeled into 3 classes that represent different health conditions at the end of the growing season (25th of July 2021) on the fields (see Figure 21): 1 – "bare soil", 2 – "medium damage", 3 – "no damage".

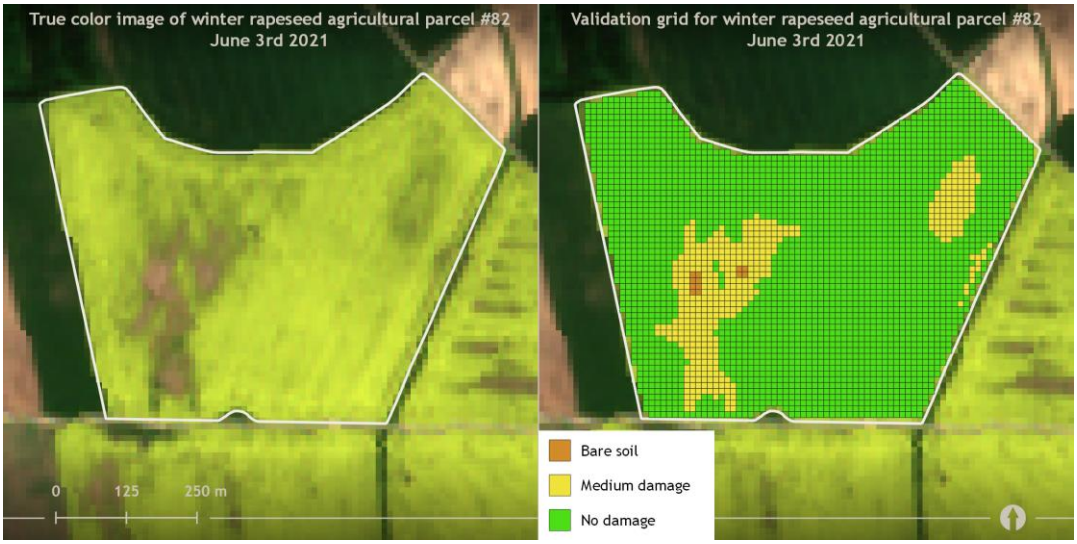


Figure 21. Validation grid example

3.3.9. Validation of the results

Within this study, a pair of vegetation index and decision tree is considered the most accurate if it is sensitive to all the damage classes present in the scene, therefore, equally high performance across the damage classes represents the most favorable classification results. Those pairs that misclassify 100% of one or more damage classes within a parcel are considered poor-performing since they are not sensitive to all damage present in the area.

The quality of the classifications was assessed using the *sklearn* (Pedregosa et al., 2011) python package. For each scene and crop type, a normalized confusion matrix was created using *confusion_matrix* function. Precision, recall, and f1-score were calculated using precision_score, recall_score, and f1_score functions respectively. The described metrics were calculated with *macro* averaging parameter that does not account for the label imbalance, which makes it possible to eliminate the bias induced by the dominant dense healthy vegetation group. Accuracy was not calculated since it provides misleading results in case of imbalanced datasets (Kulkarni et al., 2020). To illustrate the imbalance, support values that indicate an actual number of occurrences of the class in the dataset are provided as well.

For the validation, 7 parcels from winter wheat and 7 parcels from winter rapeseed, and 6 parcels from spring barley were used.

4. Results

As it was mentioned before, the whole time period is covered by 73 Sentinel-2 images, but due to weather conditions only around 30 products are available, mostly before the harvesting dates during spring and early summer (see Figure 23, Figure 32). However, the exact number of clear images varies as some fields are located as far as 50 kilometers away from each other.

4.1. Winter wheat

Eight winter wheat agricultural parcels in various conditions are used in this study. True color images that reflect damaged and healthy areas within the parcels before the ripening are presented in Figure 22:

Figure 22. Sentinel-2 true-color image, winter wheat parcels at the end of the growing season, July 5th, 2021

Croplands vary in a way, so they reflect different spatial and temporal patterns of damage: #6, #54, #57, #93 have visibly damaged areas along the borders and within, that occurred at different stages of growth, while #10, #59, #68, #75 have more homogeneous conditions with relatively smaller damaged areas.

4.1.1. Parcel-based time series analysis

Parcel #55 was taken for the parcel-based time series analysis of winter wheat, as it was the main place of the fieldwork. Temporal changes in vegetation indices at a parcel level are provided in Figure 23.

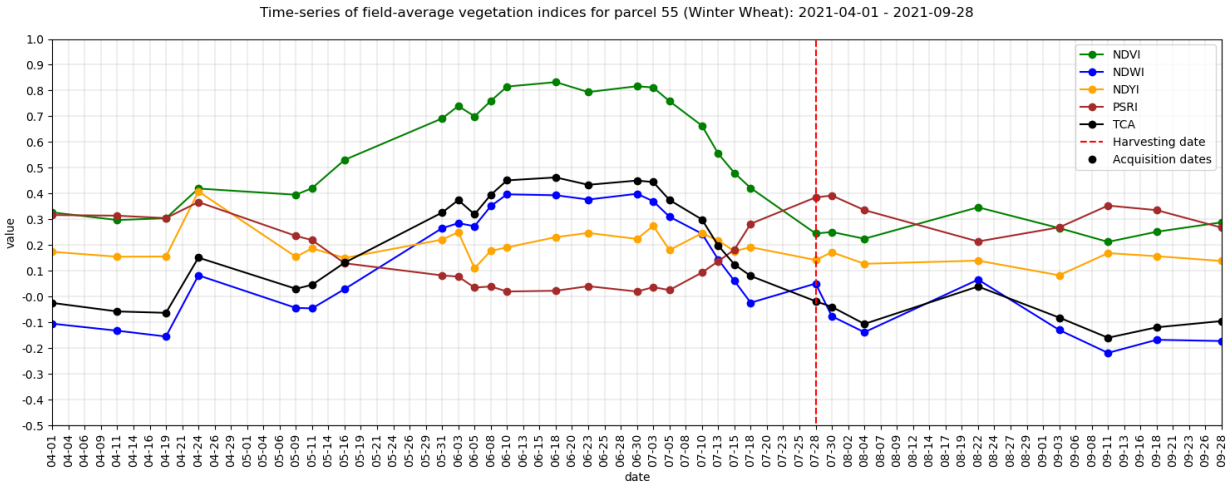


Figure 23. Time series of vegetation indices for parcel #55

From the figure, NDVI level of the parcel started to grow between the 19th and the 24th of April, which indicates the first stages of greenery development. Biomass intensity was growing up until the 18th of June when it reached its peak at around 0.83. The index remained at about the same level until the 3rd of July when it started to rapidly decrease because crops were entering the ripening stage.

Having a similar pattern to NDVI, NDWI provides additional information about the area. The index was negative at a level of -0.1 until the 24th of April, when it spiked to ~0.1 as NDVI increased. Afterward, the index was slowly declining and on the 11th of May started to grow reaching a plateau of ~0.4 on the 10th of June. It was stable until the ripening stage on the 3rd of July when it started to decrease. Paired with NDVI, NDWI values below 0 until the 11th of May indicate bare soil, while a rise to 0.4 on the 10th of June shows healthy vegetation.

In case of TCA, which is scene-specific, values below 0 indicate bare ground, as a level of greenness in this case below zero. This supports the assumption made from NDVI and NDWI as it correlates with low values of both indices. Across the growing period, the index has identical to NDVI and NDWI tendency. The first rise happened on the 24th of April from ~-0.05 to 0.15. The index then was declining to 0 until the 9th of May and started to increase afterward, reaching a plateau at a level of 0.45, with a decline after 3rd of July down to 0.

Opposite to NDVI is PSRI index. Despite being sensitive to aging, it had high values at the early stages of crops development during the low NDVI period. Later, when NDVI values were indicating healthy vegetation, PSRI was expectedly low and started to increase from the 5th of July, about the day when crops started to ripen. As the plants were turning yellow, the index was growing up to 0.39, pointing out the process of senescence. Other combinations of NDVI, NDWI,

TCA, together with a high PSRI index, for example, 18th of July, show ready-to-be-harvested crops at the end of the agricultural season.

NDYI was fluctuating mostly between 0.1 and 0.25 with expected no signs of flowering.

After the harvesting on the 28th of July, all the indices were mostly stable with slight elevations that occur due to preparation activities for the next farming season.

4.1.2. Clustering analysis

Because parcel-wide time series statistics average the data across the field, it is not possible to see the difference in development between non-homogenous areas within individual plots, which are presented in Figure 22. Clustering, in turn, differentiates between non-homogeneous areas and allows for gathering more precise statistics (see Figure 24). The parcel represents mostly healthy areas with damage that occurred at the early stages of plants development:

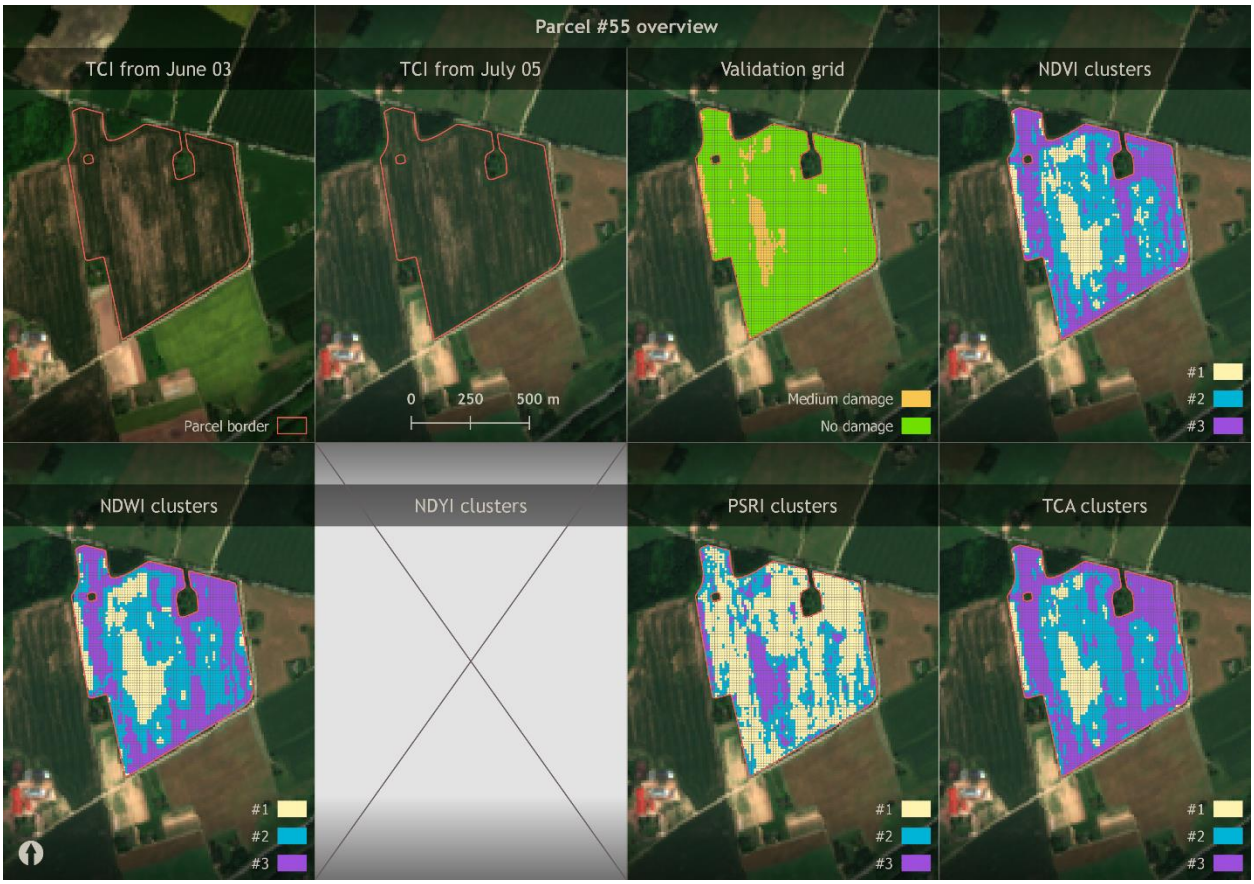


Figure 24. Overview and clustering results visualization for winter wheat parcel #55

TCI images show the condition of the parcel one month before, and the day of entering the ripening stage on the 5th of July, so the development dynamics can be seen visually. A validation grid is the grid of ground truth data used to compare the classified results against.

It is noticeable that NDVI, NDWI, PSRI, and TCA clusters resemble the pattern depicted in the validation grid, and, therefore, can be used for data generalization.

To start with NDVI, cluster visualization is provided in Figure 25. Spatially cluster 1 overlaps with the damaged area, while clusters #2 and #3 represent more healthy parts of the cropland. The average line for cluster #1 (in red) as well as all entities lay lower than the field average (in blue). Clusters #2 and #3 have different patterns, as temporal variations mostly lay above the field average. In addition, the damaged cluster has a lower trendline as its NDVI peak was not as high as in densely vegetated clusters. Based on the statistical description of the groups, #1 can be characterized as the weakest class among all, with the smallest values for every parameter (see Table 4).

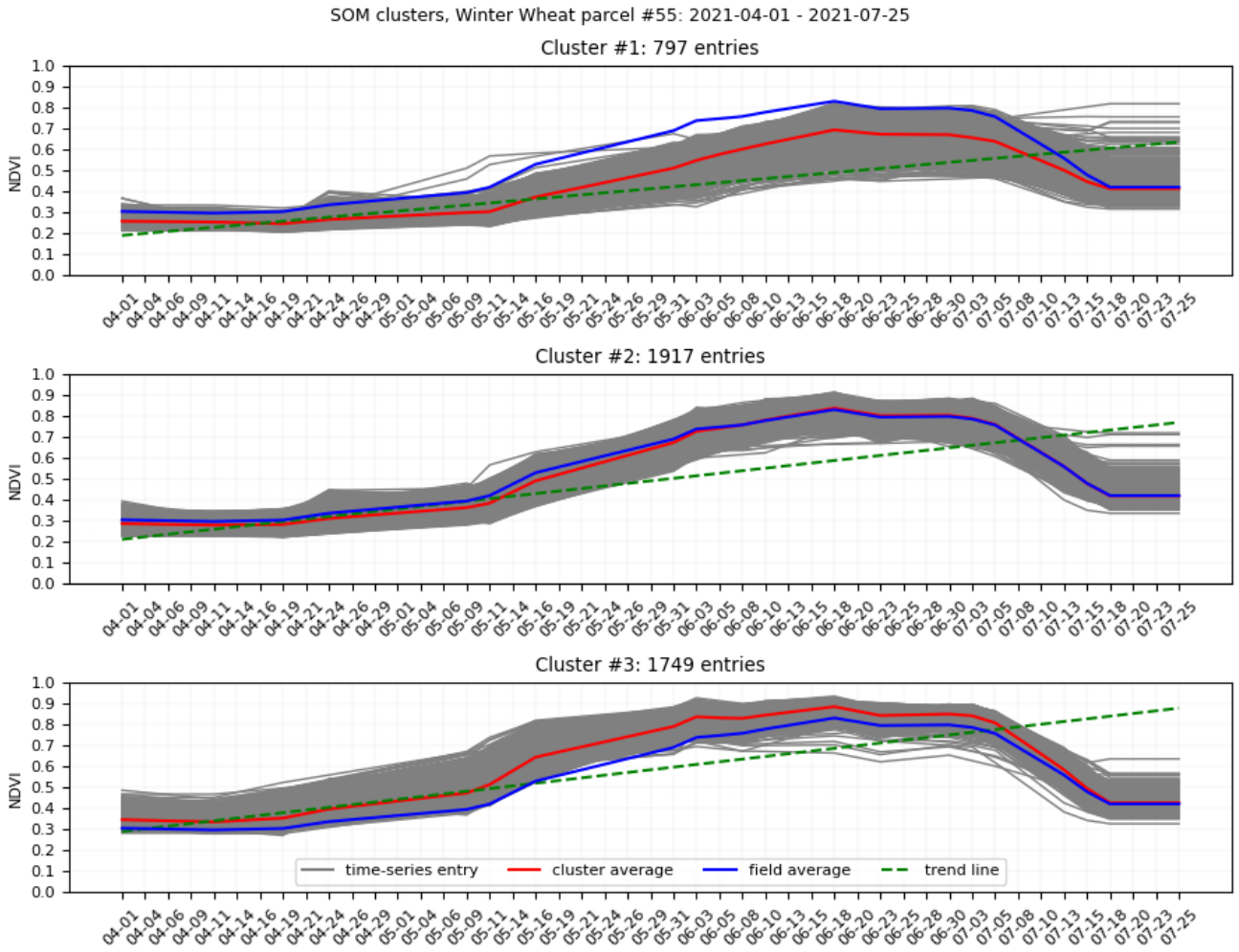


Figure 25. NDVI SOM clusters, winter wheat parcel #55

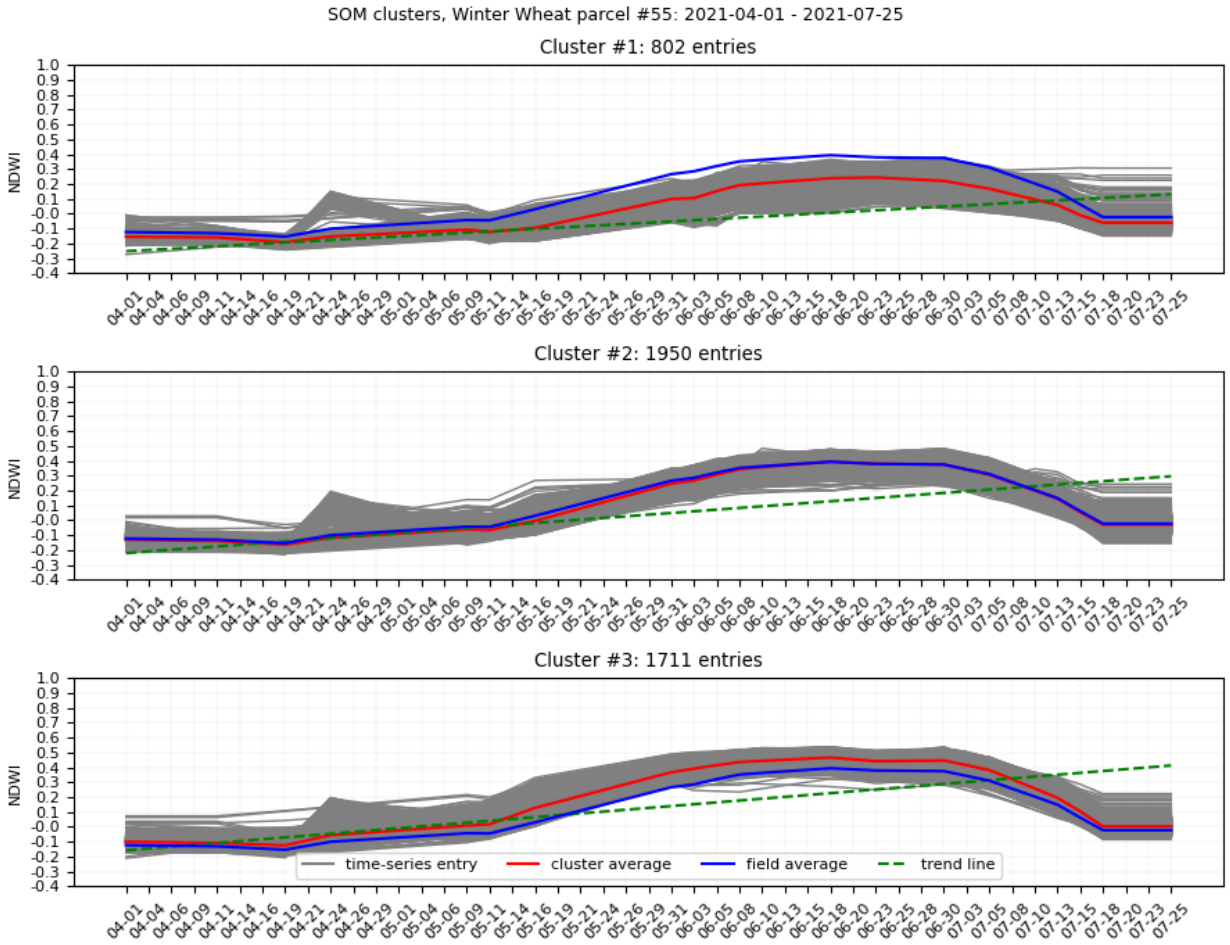


Figure 26. NDWI SOM clusters, winter wheat parcel #55

Similar patterns are observed for NDWI clusters (see Figure 26), where group #1 is the weakest growing, #3 is a healthy area, and #2 is an intermediate zone. Despite differences in value growth, the classes have similarities. Their average lines, as well as most of the entries, have values under 0 during the first month, which indicates homogeneous conditions across the field, whilst the growth dynamic can be described as problematic for category #1, as particular entries are around 0 during the peak of the season. From the detailed statistics group #1 has the lowest coefficients for all the parameters with medians standing out due to their negative value (see Table 4).

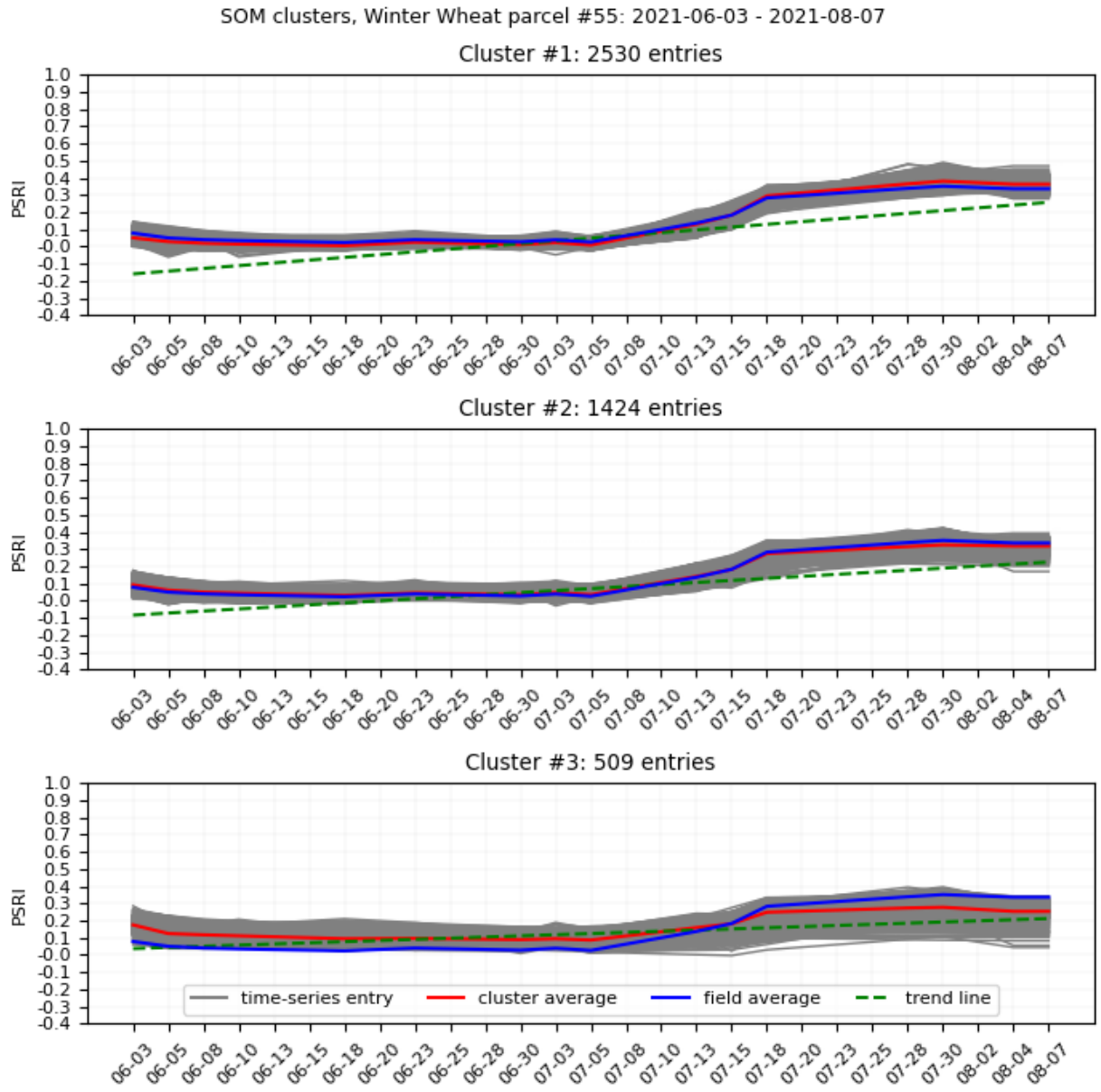


Figure 27. PSRI SOM clusters, winter wheat parcel #55

PSRI is sensitive to senescence and since drying plants in the problematic areas undergo a progressive decline of their living characteristics, the index tends to be higher in those areas at the beginning of the period, compared with healthy vegetation. Cluster #3 in Figure 27 is such a case, as it overlaps with a damaged zone on the field. Within this cluster, vegetation that wasn't turning brown has more horizontal time variations, which result in a smaller standard deviation. The cluster has higher values at the beginning of the time periods, and smaller values at the end, compared with healthy developing areas. This indicates that the damaged vegetation didn't change its state or color. At the same time, average, median, and min values are higher than in other clusters, which indicates that vegetation wasn't green enough during its expected peak. Healthy clusters can be characterized by close to 0 or negative min values and a high peak (see Table 4).

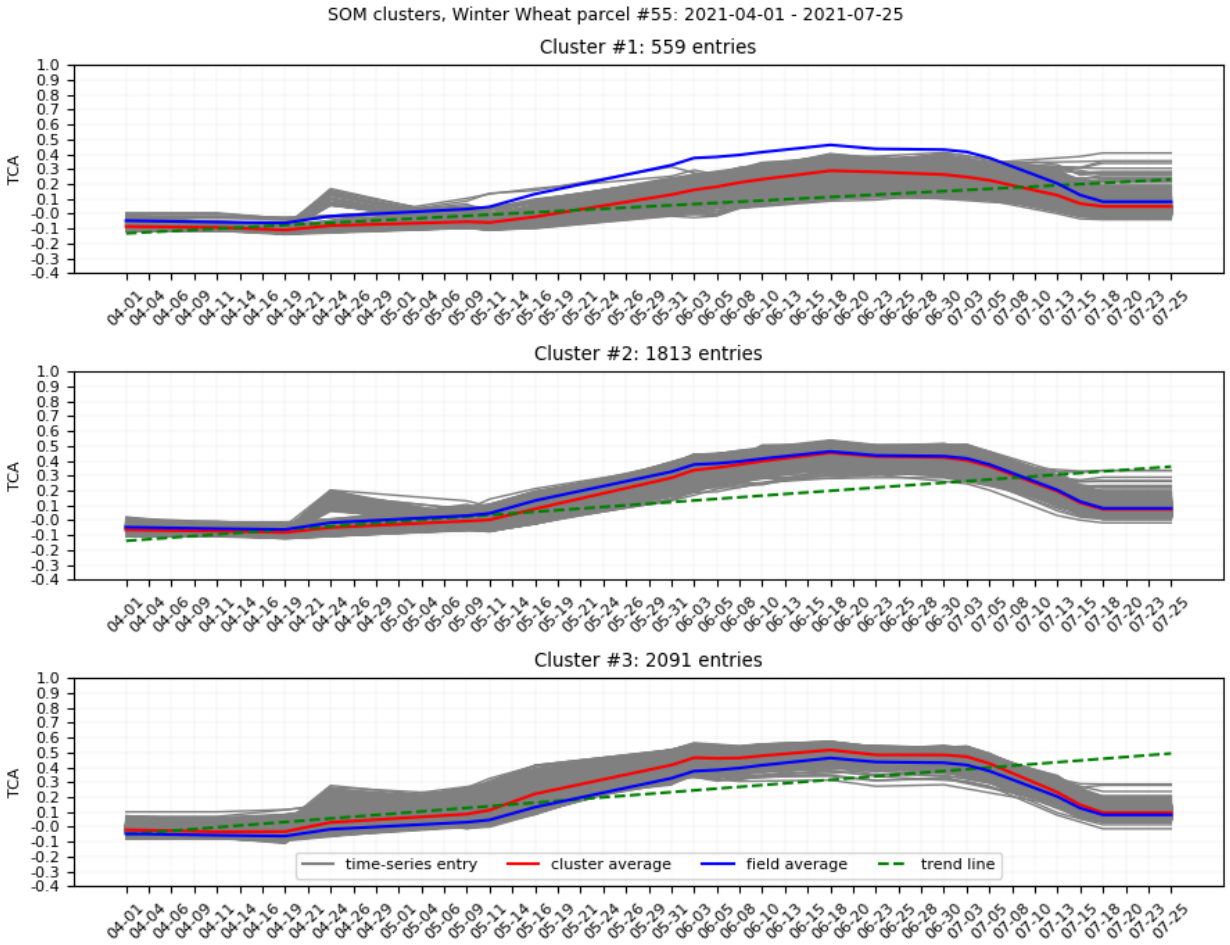


Figure 28. TCA SOM clusters, winter wheat parcel #55

Similar to NDVI, TCA cluster #1 represents a damaged zone on parcel #55 (see Figure 28). As values below 0 indicate bare soil, manual comparison of various pixels suggests that 0.4 can be defined as a threshold for healthy vegetation. In this case, values in class #1 mostly belong one of the damaged vegetation classes, while groups #2 and #3 have entries that experience healthy growth.

Table 4. Averaged statistics per generated clusters for winter wheat parcel #55

VI	#	std.	average	median	min	max	sum	trend	slope	intercept
NDVI	1	0.159	0.443	0.414	0.247	0.696	20.829	increasing	0.0097	0.191
	2	0.203	0.531	0.492	0.280	0.840	24.951	increasing	0.0122	0.212
	3	0.202	0.604	0.585	0.336	0.887	28.368	increasing	0.0128	0.290
NDWI	1	0.152	0.005	-0.061	-0.192	0.244	0.254	increasing	0.0083	-0.253
	2	0.202	0.092	0.037	-0.165	0.394	4.323	increasing	0.0113	-0.221
	3	0.217	0.163	0.127	-0.125	0.466	7.645	increasing	0.0124	-0.158
PSRI	1	0.151	0.142	0.049	0.005	0.381	3.831	increasing	0.016	-0.160
	2	0.120	0.144	0.070	0.030	0.325	3.890	increasing	0.0119	-0.084
	3	0.073	0.163	0.124	0.086	0.277	4.411	increasing	0.0067	0.037
TCA	1	0.139	0.066	0.049	-0.110	0.289	3.117	increasing	0.0079	-0.133
	2	0.192	0.155	0.111	-0.083	0.454	7.306	increasing	0.0109	-0.139
	3	0.20	0.230	0.222	-0.033	0.517	10.806	increasing	0.0118	-0.051

4.1.3. Classification and validation

Clustering of vegetation indices for winter wheat croplands showed the differences in temporal and spatial variations between healthy and unhealthy areas, making it possible to differentiate these zones based on the calculated statistics. Thus, decision trees that perform parcel area classification based on time series data are presented in Figure 29.

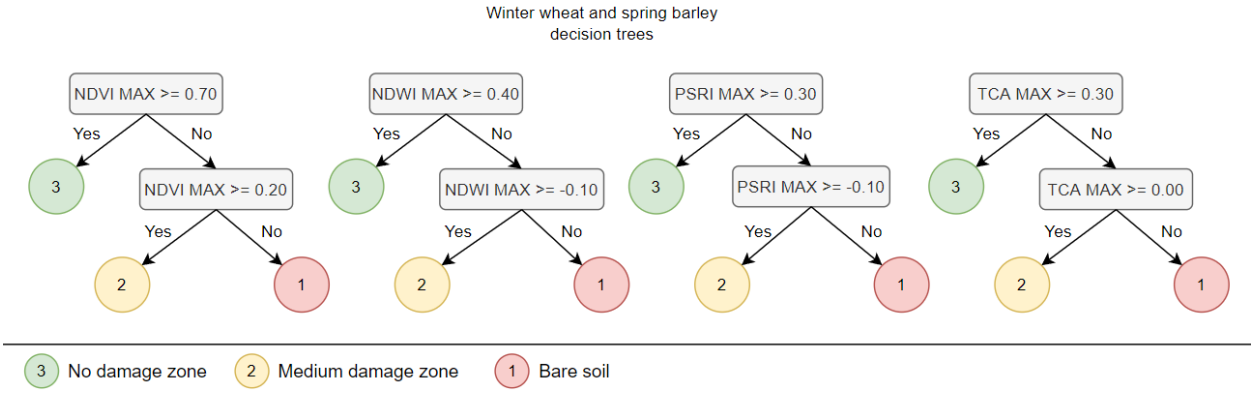


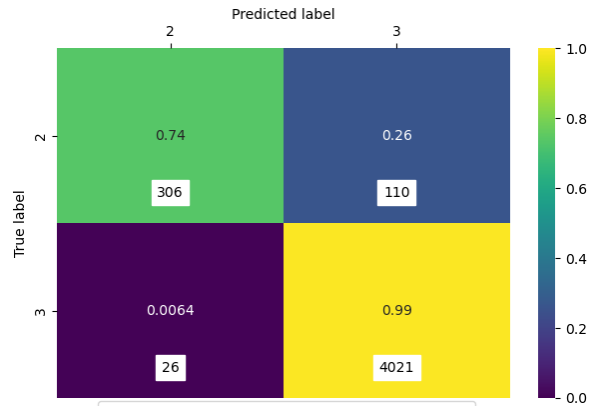
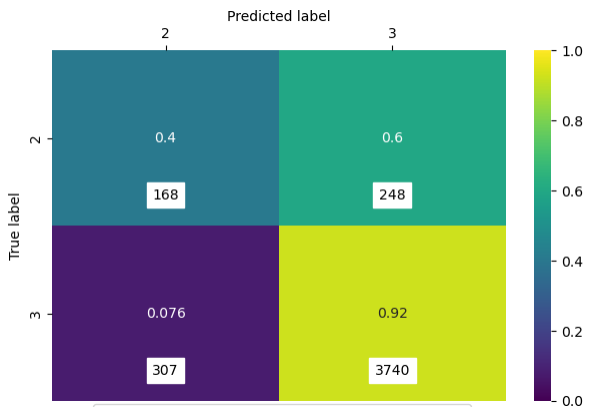
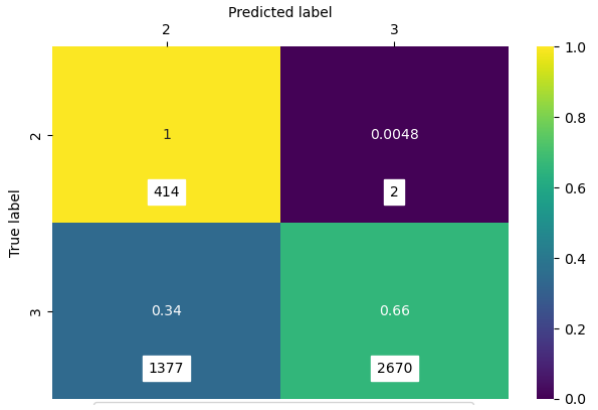
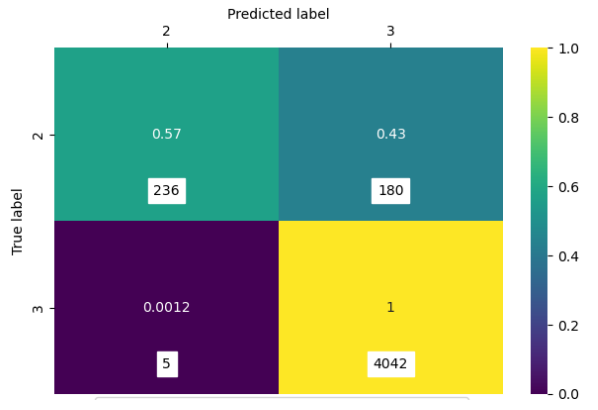
Figure 29. Decision trees for winter wheat and spring barley agricultural parcels

Normalized confusion matrices of the classification results of agricultural parcel #55 are provided in Table 5, while results across test winter wheat parcels are provided in Table 6. Non-normalized class counts are presented in white rectangles within confusion matrices' classes.

From the Table 5 it can be seen that NDVI and TCA indices with the combination of the abovementioned decision trees are the most accurate, as they have the highest F1-scores: 0.901 and 0.848 respectively. Both indices classify 99% of "no damage" zone correctly, however, NDVI has a better score in the "medium damage" zone and, therefore, would be a preferred choice over all pairs. The other vegetation indices show less accurate scores. In case of NDWI, despite detecting 100% of "medium damage" area, it has a low precision, as 34% of "no damage" class was misclassified, resulting in the lowest precision score among all confusion matrices. A combination of PSRI and the corresponding decision tree has the poorest performance, with the

lowest F1-score among presented matrices. Due to the fact that the “bare soil” group was not presented in the field, it is not displayed in the confusion matrices. The spatial pattern of the results is provided in Figure 30.

Table 5. Confusion matrices for the classifications performed on winter wheat agricultural parcel #55: (2) “medium damage”, (3) “no damage” zones

<table>
<tr><th colspan="2">NDVI confusion matrix</th><th colspan="2">PSRI confusion matrix</th></tr>
<tr><td colspan="2"> 
</td><td colspan="2"> 
</td></tr>
<tr><td>Precision: 0.948</td><td>Recall: 0.865</td><td>Precision: 0.948</td><td>Recall: 0.865</td></tr>
<tr><td>F1 score: 0.901</td><td></td><td>F1 score: 0.901</td><td></td></tr>
<tr><th colspan="2">NDWI confusion matrix</th><th colspan="2">TCA confusion matrix</th></tr>
<tr><td colspan="2"> 
</td><td colspan="2"> 
</td></tr>
<tr><td>Precision: 0.615</td><td>Recall: 0.827</td><td>Precision: 0.615</td><td>Recall: 0.827</td></tr>
<tr><td>F1 score: 0.585</td><td></td><td>F1 score: 0.585</td><td></td></tr>
</table>

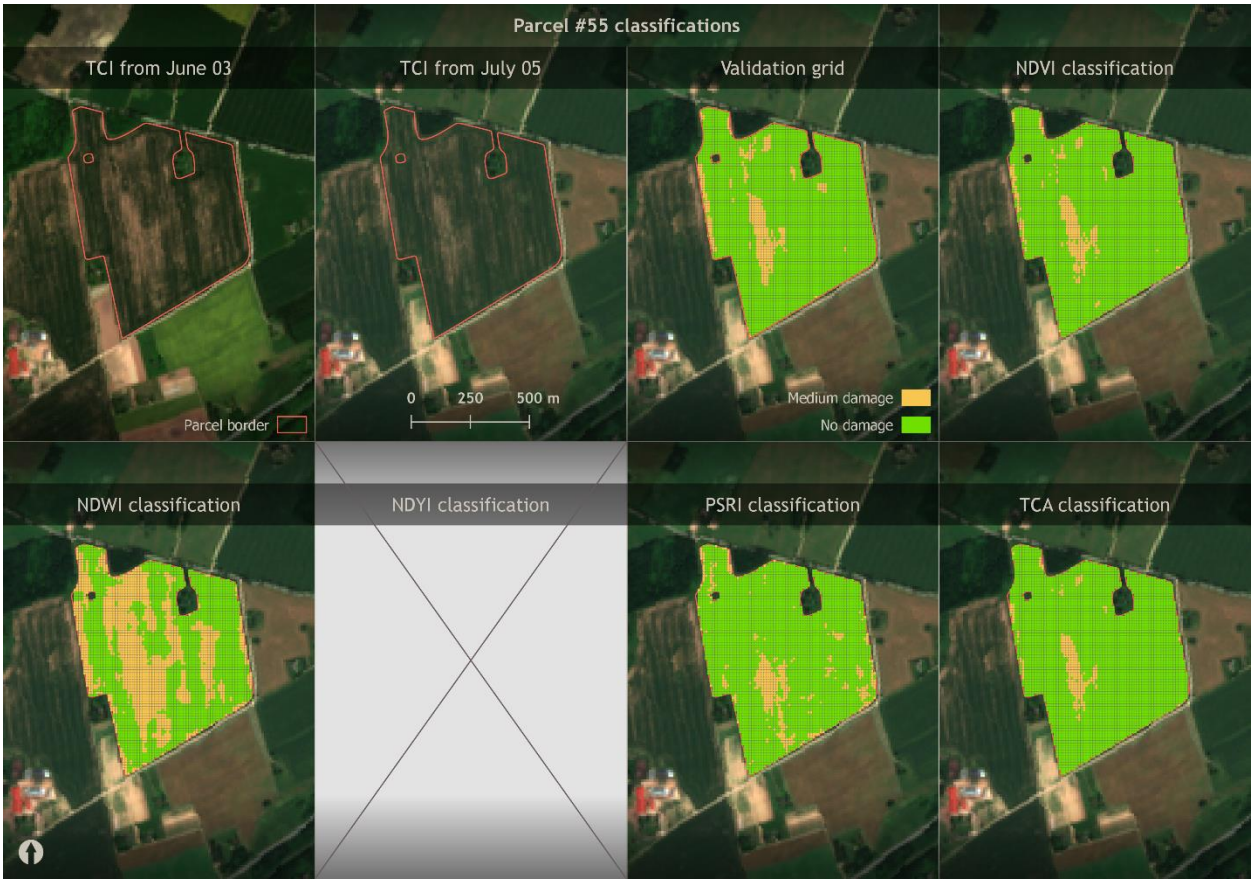
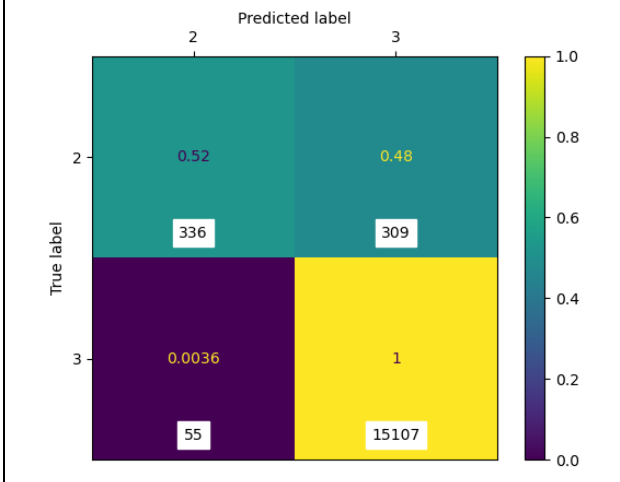
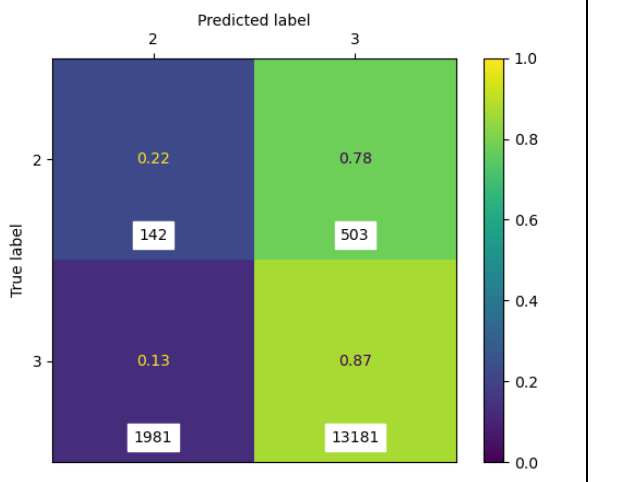
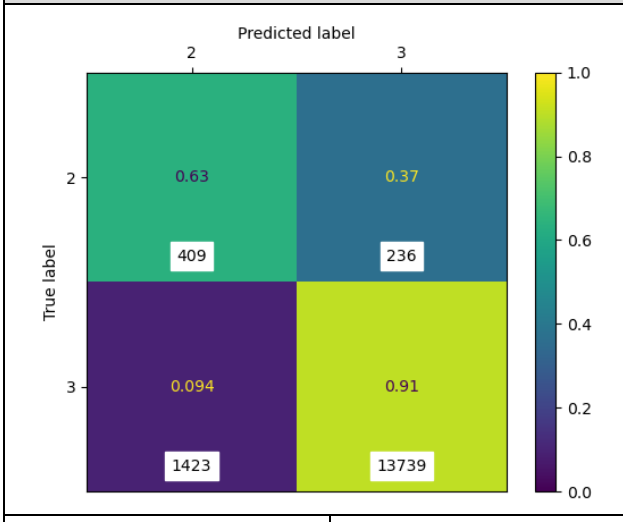
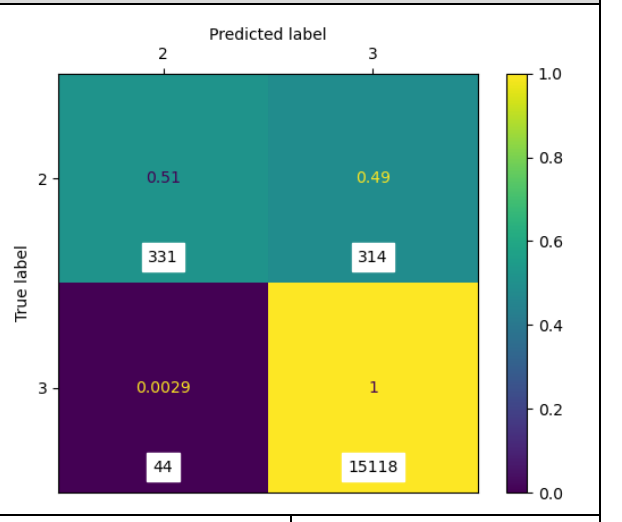


Figure 30. Classification results for winter wheat agricultural parcel #55

The results across test winter wheat parcels are similar to the ones obtained for agricultural unit #55. In Table 6 NDVI and TCA show the best performance with the highest F1-score of 0.818. NDWI provides a balance between the relative amount of correctly classified areas in groups #2 and #3, as well as it has the highest score in the “medium damage” group across all confusion matrices. This is, however, at the expense of the low precision. A combination of PSRI and the corresponding decision tree has the poorest performance, with the lowest F1-score among presented matrices. Due to the fact that the “bare soil” group was not presented on the winter wheat agricultural parcels, it is not displayed in the confusion matrices.

Table 6. Confusion matrices for the classifications performed on test winter wheat agricultural parcels: (2) “medium damage”, (3) “no damage” zones

NDVI confusion matrix		PSRI confusion matrix	
			
Precision: 0.920	Recall: 0.759	Precision: 0.920	Recall: 0.759
F1 score: 0.818		F1 score: 0.818	
NDWI confusion matrix		**TCA confusion matrix**	
			
Precision: 0.603	Recall: 0.770	Precision: 0.603	Recall: 0.770
F1 score: 0.637		F1 score: 0.637	

4.2. Winter rapeseed

Eight winter rapeseed agricultural parcels in various conditions are used in this study. Damaged areas are the most prominent during flowering stage in case of winter rapeseed, therefore, true-color images that reflect damaged and non-damaged areas of the selected parcels at the peak of flowering are presented in Figure 31:

Figure 31. Sentinel-2 true-color image, winter rapeseed parcels at the peak of flowering, June 3rd 2021

For winter rapeseed, parcels #21, #31, #46, and #48 have a uniformly high level of flowering and the whole area of the fields represents healthy developing crops. Agricultural parcels #78, #82, #83, and #85 represent mixed conditions since they have a considerable amount of non-flowering damaged areas that can be visually distinguished on the images.

4.2.1. Parcel-based time series analysis

Parcel #82 was taken for parcel-based time series analysis of winter rapeseed as it represents visible damaged and healthy areas (see Figure 32).

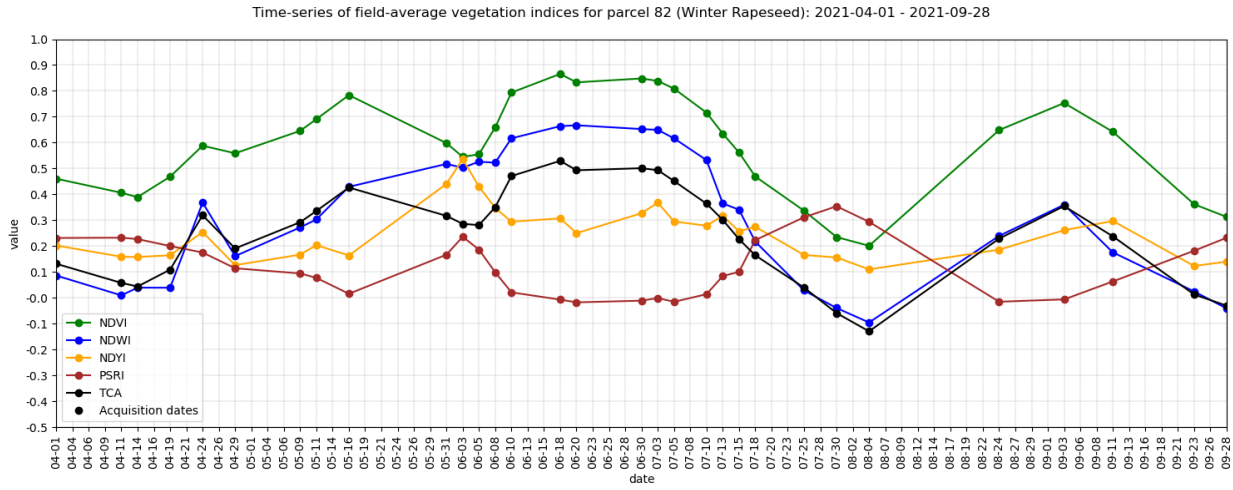


Figure 32. Time series of vegetation indices for parcel #82

Within the time interval of interest, the time pattern of winter rapeseed differs from wheat, as rapeseed develops flowers during its growth. In this case, the NDVI starts with a small decrease between the 1st and the 14th of April, which might indicate hilling. Later, NDVI level of the parcel grows from 0.39 on the 14th of April to 0.78 on the 16th of May, followed by a decrease down to 0.55 on the 3rd of June during flowering. After that, NDVI started to rise to 0.87 in mid-June entering a short plateau, after which NDVI decreased rapidly as crops started to ripen.

In case of TCA, the pattern is similar to NDVI, as one of the components – TCV, represents the amount of greenness within a unit area. TCA rises and falls on the same dates as NDVI: the first peak is on the 16th of May, which is followed by a drop on the 3rd/5th of July, with its absolute peak of 0.53 on the 18th of June. The second TCA peak is higher (0.42 on 05-16 vs. 0.53 on 06-18), which might be caused by more dense vegetation. At the end of July TCA falls below 0 at the same time as NDVI declines down to 0.2.

NDWI being similar to NDVI and TCA in case of winter wheat parcels, shows a peculiar behavior when applied to winter rapeseed. The index still resembles the tendency of the biomass indices, but during flowering, NDWI remains sensitive to the amount of water in the plants regardless of their color. Steady index increase from the 29th of April to its peak at 0.66 on the 20th of June shows desirable developments in the field. Values under 0 at the end of July overlap with the same of TCA and low NDVI, ultimately indicating the bare ground at that time.

A useful for the winter rapeseed crop type NDYI index expectedly reacted to the amount of yellow during flowering in a form of a peak at 0.54 on the 3rd of June, while the average NDYI value before flowering was lower than after: ~0.17 vs. ~0.25. The index reacts exactly as it is described in D'Andrimont et al. (2020), making it an important indicator for rapeseed yield estimation.

Lastly, PSRI, which in addition to being sensitive to senescence, had increased values during the flowering stage as well.

There is no extra data available for the agricultural parcel, however, the harvesting was expected to take place in early August. For a short period after that, the grass was growing on the field. That

can be seen as NDVI, NDWI, TCA increase, and PSRI decrease. Late August and September is the time when preparations for the next farming season are made, so vegetation indices that estimate greenness and water content of the vegetation decrease, PSRI increases, which ultimately indicate low vegetation with bare soil at the end of September.

4.2.2. Clustering analysis

A short parcel overview and spatial patterns of vegetation indices clustering are provided in Figure 33. Field #82 experienced early damage and during flowering had some visible areas of damaged vegetation that didn't produce any flowers. After a month, the area has partially regrown, however, no yield was expected from that zone. Visually, NDVI, NDWI, NDYI, and TCA clusters provide similar to the validation grid spatial patterns, while PSRI has partial overlap with the damaged vegetation validation classes.

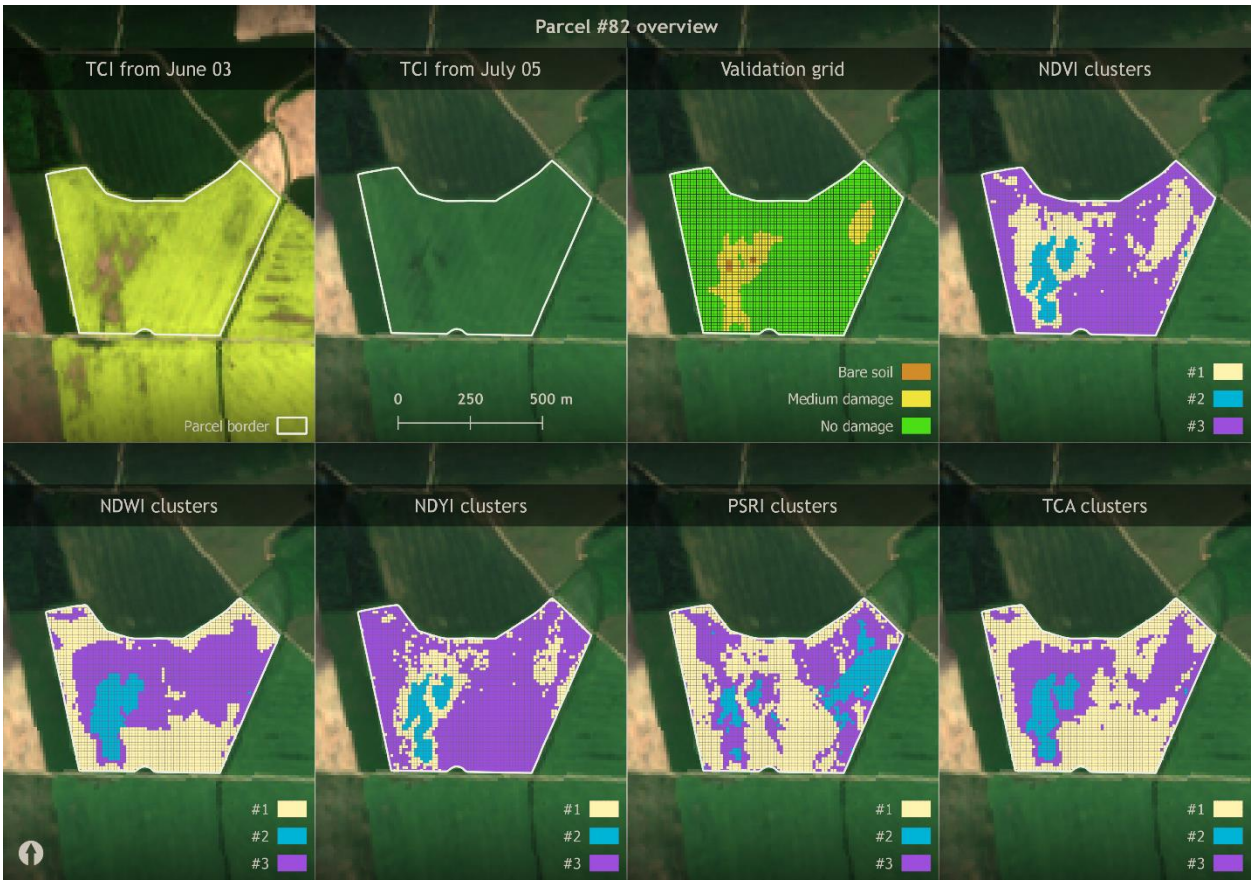


Figure 33. Overview and clustering results visualization for winter rapeseed parcel #82

From Figure 34, NDVI cluster #2 represent a damaged area within the field. Compared with the two other groups, the cluster can be described as the weakest growing with entries that lay below a healthy vegetation threshold during both growing peaks on the 16th of May and 18th of June. At the end of the growing stage, however, the vegetation intensity of cluster #2 reaches its peak at a level of around 0.8 and evens out with neighboring areas. The peak is followed by a comparatively slow decrease during the ripening stage. Late growth, lack of flowers, and slow decline indicate that condition of winter rapeseed is not healthy within this area. From Table 7 the cluster has the lowest total accumulation of NDVI, while other parameters, for example, cluster maximum, are at the same level as other clusters, which makes it difficult to distinguish groups from each other.

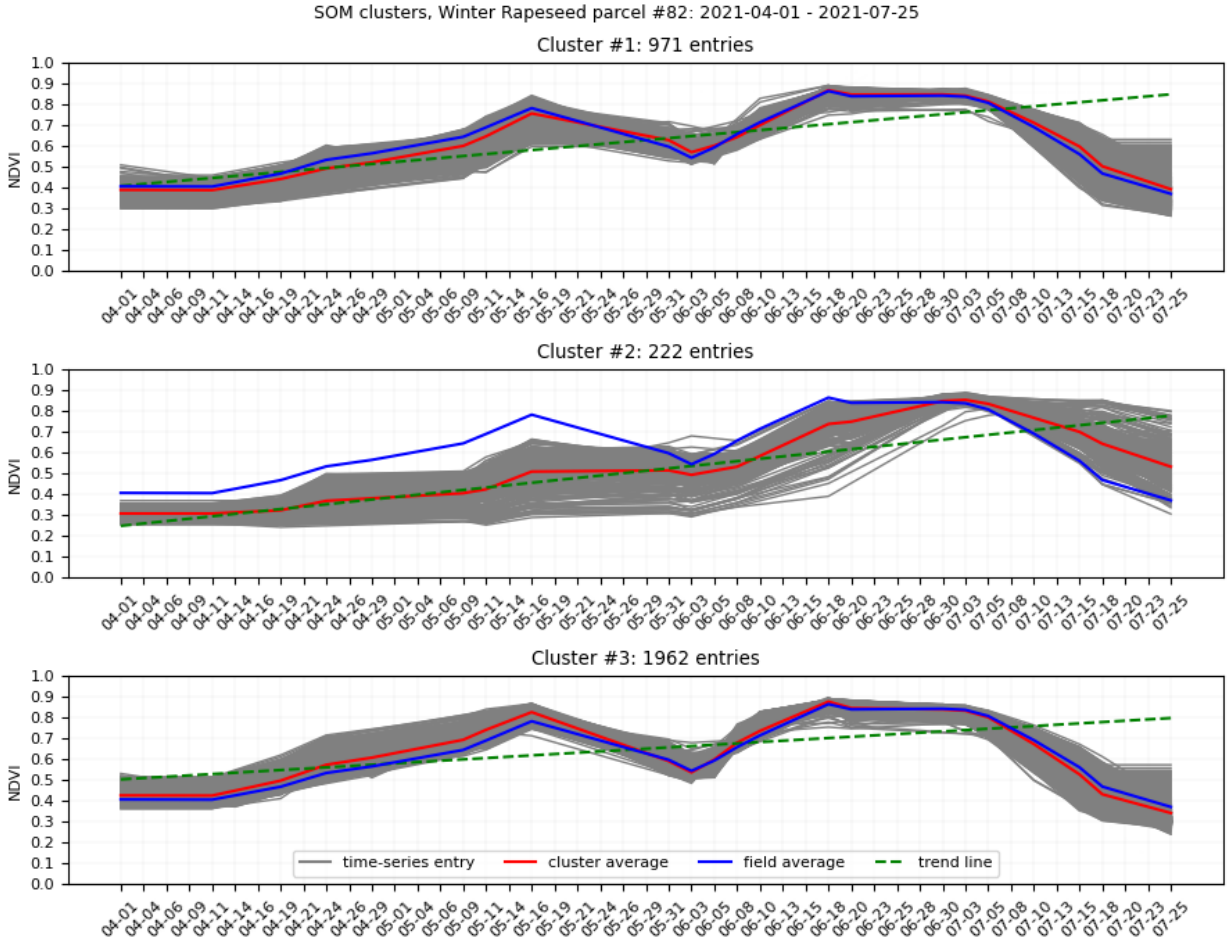


Figure 34. NDVI SOM clusters, winter rapeseed parcel #82

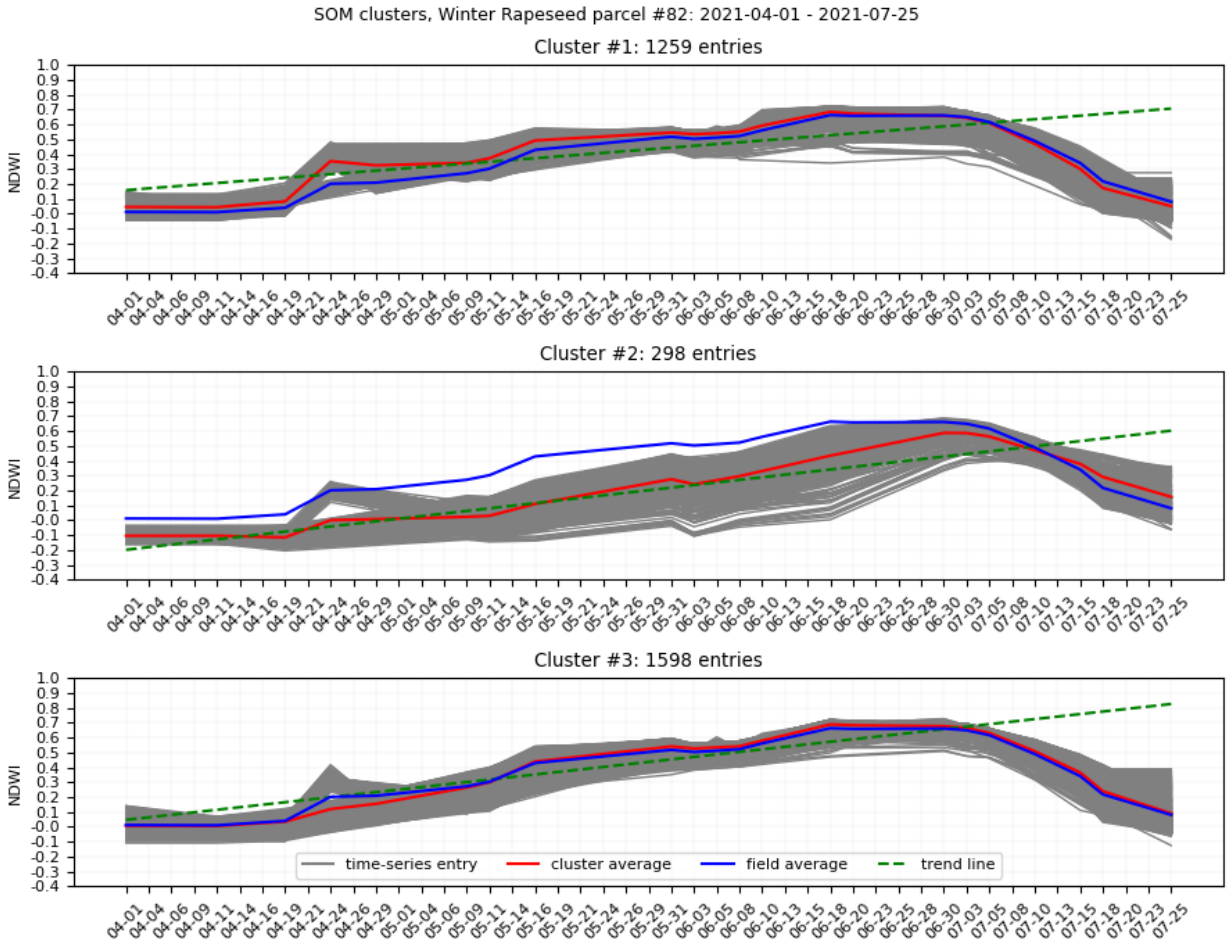


Figure 35. NDWI SOM clusters, winter rapeseed parcel #82

Weak growth of biomass within the damaged zones can be seen at NDWI clusters as well (see Figure 35). Group #2 overlaps with the area of interest and visually has values that indicate low water content in the vegetation. Similar to NDVI, vegetation growth before the ripening stage evens out the maximum statistical parameter used for winter wheat classification (see Table 7).

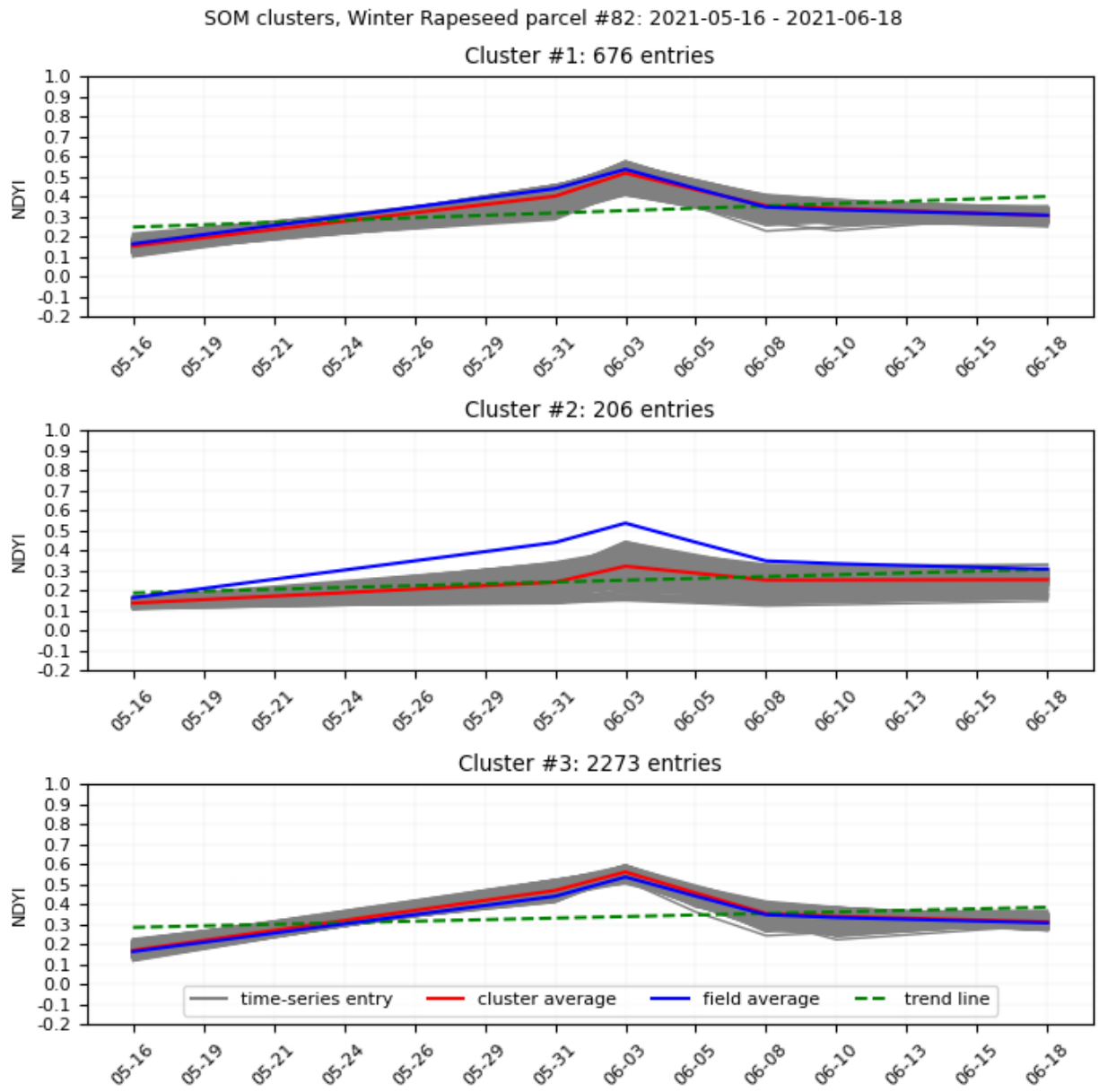


Figure 36. NDYI SOM clusters, winter rapeseed parcel #82

In case of NDYI clusters (see Figure 36), group #2 overlaps with the damaged zones. Unlike other clusters, the group has entries that lack any temporal variation of the index, ultimately lowering the flowering intensity within the group (see Table 7). Clusters #1 and #3 show a higher variation as they represent more healthy areas.

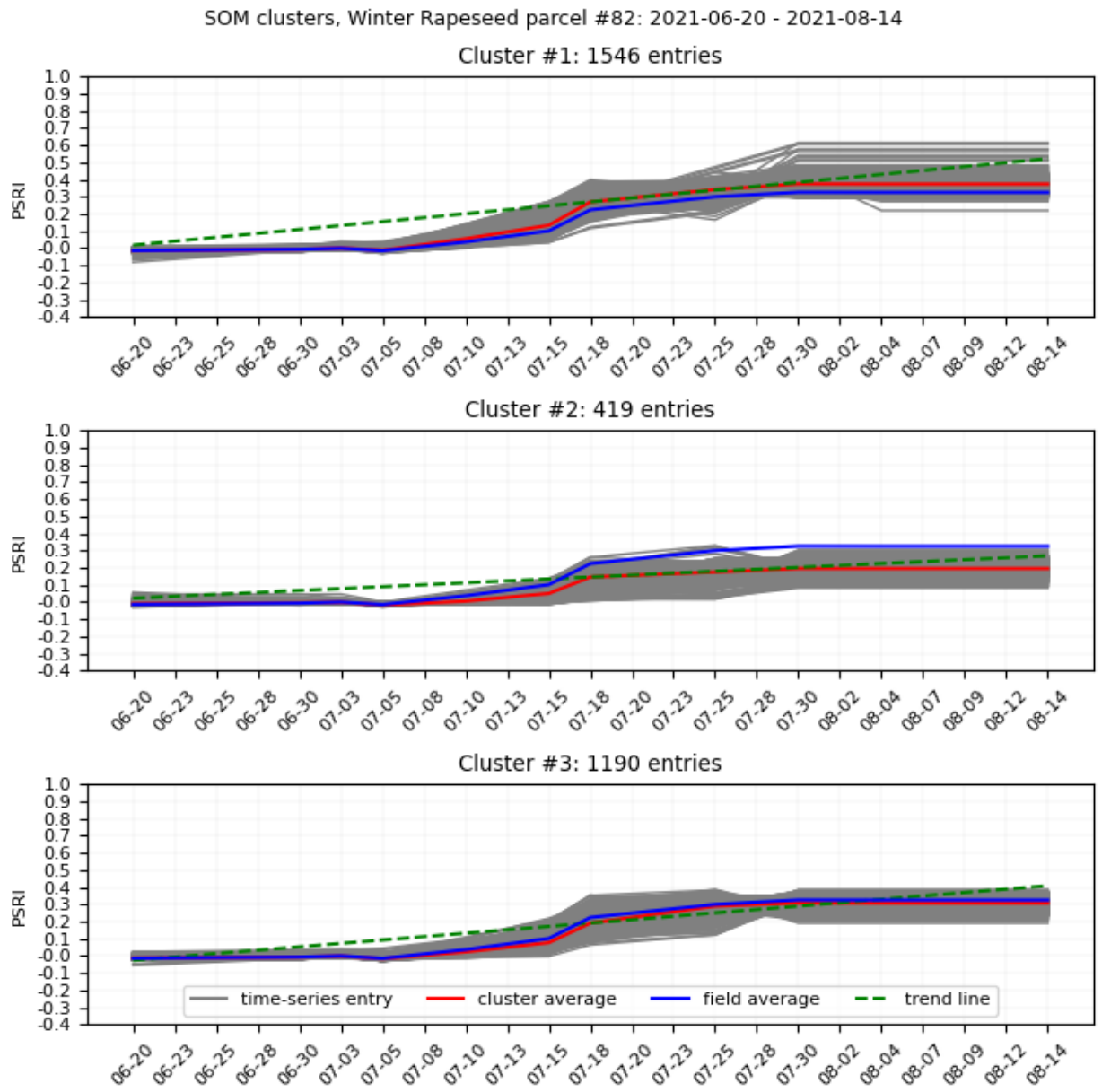


Figure 37. PSRI SOM clusters, winter rapeseed parcel #82

Spatial pattern of PSRI poorly highlights the damaged zone as it also includes healthy areas. Nevertheless, group #2 represents most of the damaged zone, group #3 partly damaged zones. Over the time-period group #2 has the lowest PSRI values that do not exceed the healthy vegetation threshold (see Figure 37), while cluster 3 has entries above 0.2, which is seen in the statistics as well (see Table 7).

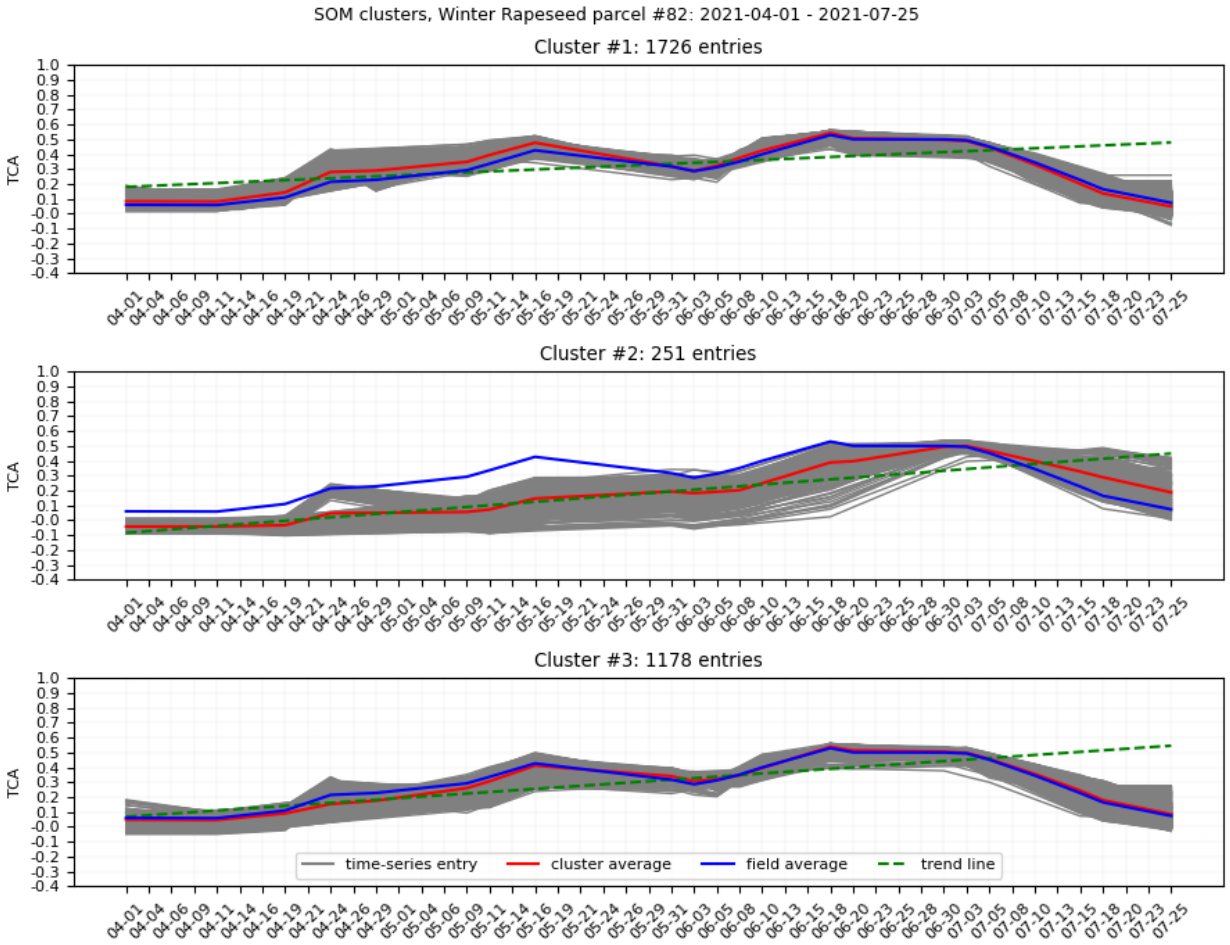


Figure 38. TCA SOM Clusters, winter rapeseed parcel #82

Similar to NDVI, TCA cluster #2 represents the damaged zones on parcel #82. Together with negative NDWI, TCA at the beginning of the time interval shows that the cluster area was mostly bare ground (see Figure 38), while the rest of the agricultural parcel had sprouts. During biomass intensity peaks, TCA didn't exceed a 0.4 threshold used for winter wheat classification. At the end of the season, however, cluster #2 has risen to the field average values due to weed growth.

Table 7. Averaged values per generated clusters for winter rapeseed parcel #82

VI	#	std.	average	median	min	max	sum	trend	slope	intercept
NDVI	1	0.156	0.619	0.629	0.389	0.870	29.110	increasing	0.0096	0.409
	2	0.175	0.538	0.514	0.308	0.855	25.279	increasing	0.0115	0.249
	3	0.155	0.638	0.652	0.342	0.878	30.002	increasing	0.0064	0.504
NDWI	1	0.219	0.391	0.433	0.043	0.685	18.368	increasing	0.0119	0.158
	2	0.225	0.200	0.202	-0.115	0.588	9.422	increasing	0.0174	-0.198
	3	0.241	0.361	0.437	0.005	0.689	16.960	increasing	0.0169	0.046
NDYI	1	0.098	0.295	0.309	0.165	0.546	7.383	increasing	0.0069	0.225
	2	0.103	0.291	0.256	0.167	0.558	7.266	no trend	0.0041	0.208
	3	0.061	0.222	0.240	0.138	0.339	5.554	increasing	0.0070	0.155
PSRI	1	0.169	0.193	0.270	-0.015	0.374	4.431	increasing	0.0229	0.017
	2	0.092	0.096	0.145	-0.018	0.195	2.208	increasing	0.0112	0.022
	3	0.143	0.152	0.191	-0.017	0.309	3.500	increasing	0.0198	-0.026
TCA	1	0.150	0.313	0.330	0.049	0.544	14.697	increasing	0.0065	0.180
	2	0.170	0.19	0.182	-0.042	0.500	8.937	increasing	0.0116	-0.085
	3	0.158	0.287	0.307	0.045	0.539	13.469	increasing	0.0104	0.068

4.2.3. Classification and validation

With the existing amount of information about winter rapeseed agricultural parcels, similar to winter wheat decision trees were made Figure 39. Unlike winter wheat or spring barley, here NDYI is used to estimate the damage at the flowering stage:

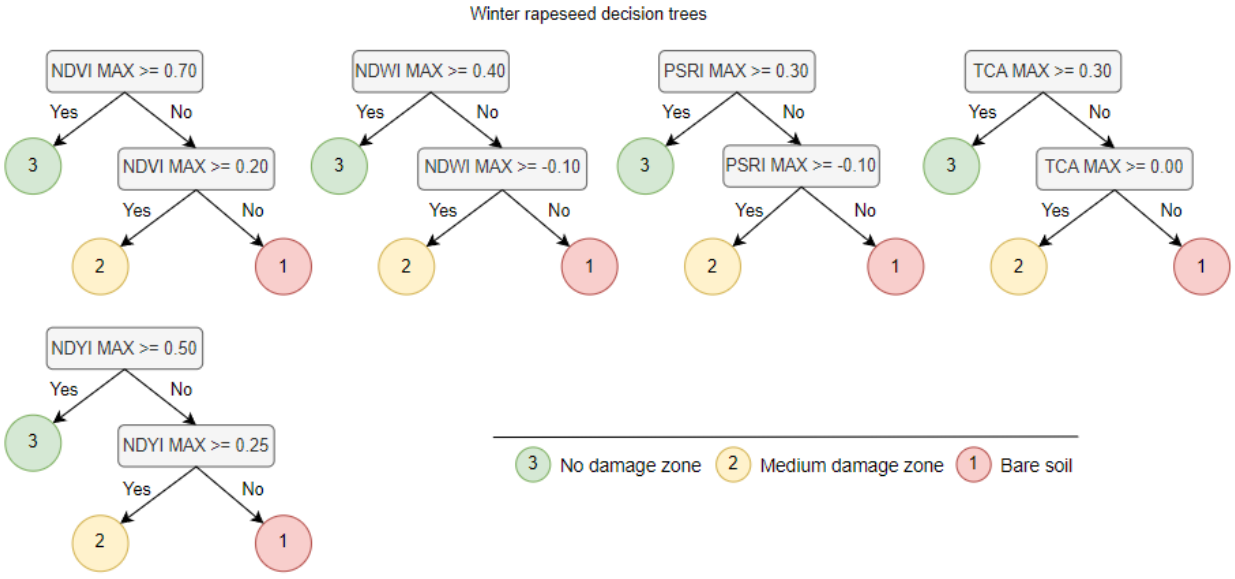
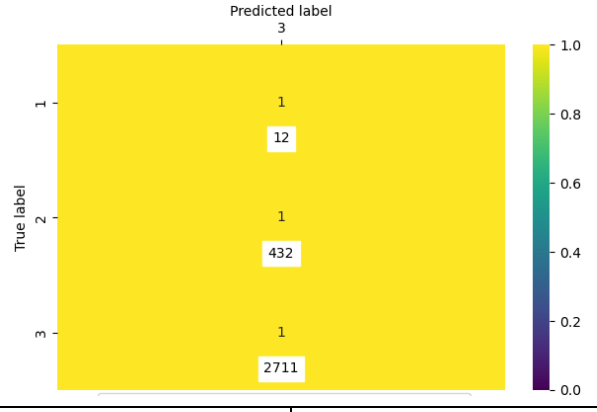
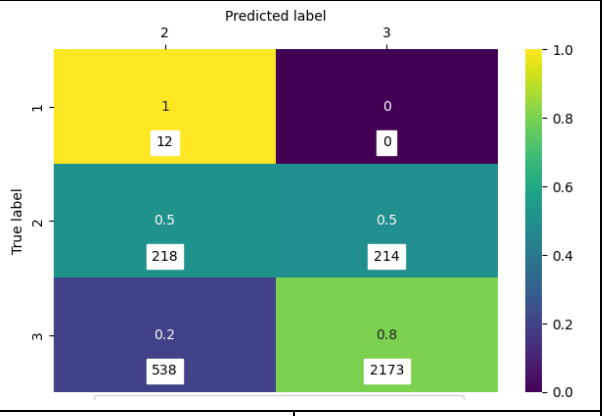
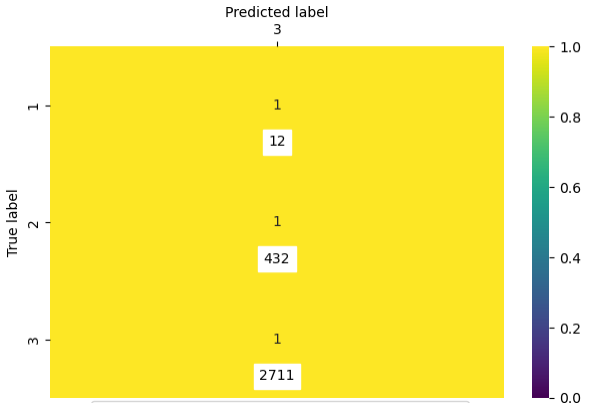
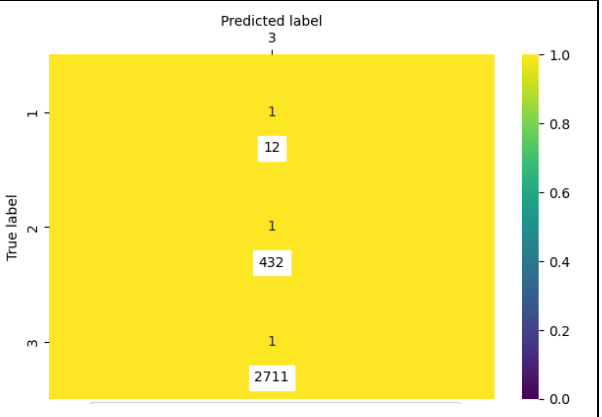
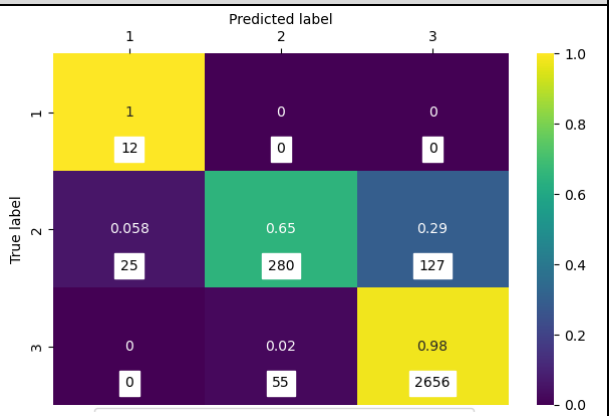


Figure 39. Decision trees for winter rapeseed

Normalized confusion matrices of the classification results of agricultural parcel #82 are provided in Table 8, while results across all winter rapeseed parcels are provided in Table 9. Non-normalized class counts are presented in white rectangles within confusion matrices' classes.

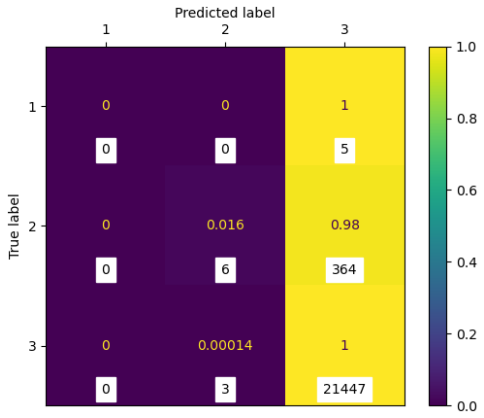
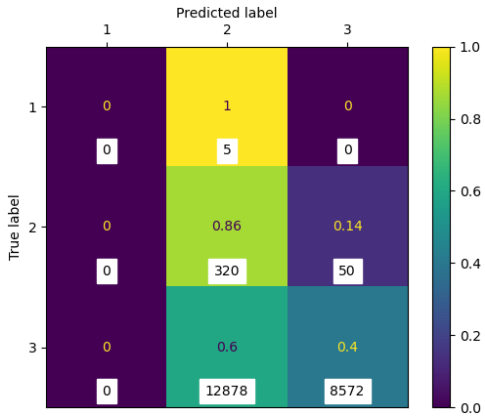
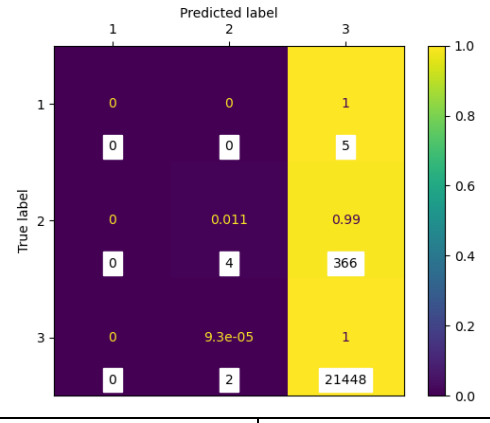
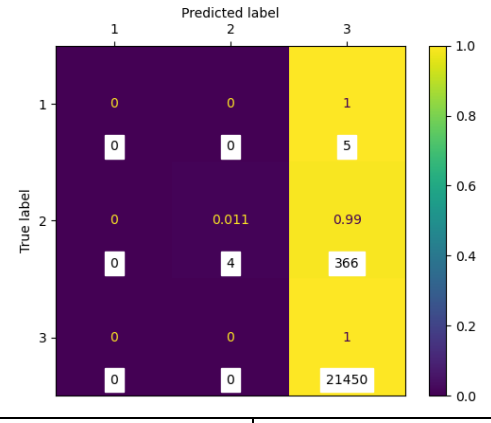
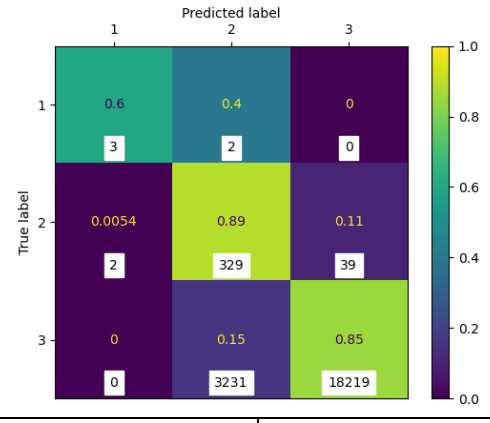
Table 8. Confusion matrices for the classifications performed on winter rapeseed agricultural parcel #82: (1) “bare soil”, (2) “medium damage”, (3) “no damage” zones

NDVI confusion matrix		PSRI confusion matrix	
			
Precision: 0.286	Recall: 0.333	Precision: 0.398	Recall: 0.435
F1 score: 0.308		F1 score: 0.405	
NDWI confusion matrix		**TCA confusion matrix**	
			
Precision: 0.286	Recall: 0.333	Precision: 0.286	Recall: 0.333
F1 score: 0.308		F1 score: 0.308	
NDYI confusion matrix			
			
Precision: 0.705	Recall: 0.876		
F1 score: 0.729			

As a result of the winter rapeseed agricultural parcel #82 classification and overall results from test parcels in Table 9, vegetation indices that estimate greenness or water content with the combination of the proposed decision trees performed poorly. None of the NDVI, NDWI, PSRI, or TCA decision trees classified “bare soil” on winter rapeseed parcels on a defined time interval, while “medium damage” and “no damage” groups are largely misclassified. The best overall performance was obtained from a pair of NDYI and its decision tree, which was highly sensitive to the “bare soil” and “no damage” classes, and with more than a half correctly classified “medium damage” class. The spatial pattern of the results for agricultural parcel #82 is provided in Figure 40.

Figure 40. Classification results for winter rapeseed agricultural parcel #82

Table 9. Confusion matrices for the classifications performed on test winter rapeseed agricultural parcels: (1) “bare soil”, (2) “medium damage”, (3) “no damage” zones

<table>
<tr><th colspan="2">NDVI confusion matrix</th><th colspan="2">PSRI confusion matrix</th></tr>
<tr><td colspan="2"> 
</td><td colspan="2"> 
</td></tr>
<tr><td>Precision: 0.550</td><td>Recall: 0.339</td><td>Precision: 0.339</td><td>Recall: 0.421</td></tr>
<tr><td>F1 score: 0.341</td><td></td><td>F1 score: 0.206</td><td></td></tr>
<tr><th colspan="2">NDWI confusion matrix</th><th colspan="2">TCA confusion matrix</th></tr>
<tr><td colspan="2"> 
</td><td colspan="2"> 
</td></tr>
<tr><td>Precision: 0.550</td><td>Recall: 0.337</td><td>Precision: 0.661</td><td>Recall: 0.337</td></tr>
<tr><td>F1 score: 0.338</td><td></td><td>F1 score: 0.338</td><td></td></tr>
<tr><th colspan="2">NDYI confusion matrix</th><th colspan="2"></th></tr>
<tr><td colspan="2"> 
</td><td colspan="2"></td></tr>
<tr><td>Precision: 0.563</td><td>Recall: 0.780</td><td></td><td></td></tr>
<tr><td>F1 score: 0.562</td><td></td><td></td><td></td></tr>
</table>

4.3. Spring barley

Overall, spring barley agricultural parcels are in better condition than winter crop types. Healthy conditions prevail across agricultural parcels of this type around Tartu city, with only a small number of fields that experience growth problems. Seven spring barley agricultural parcels were selected that represent various conditions: #109, and #113 are uniformly healthy while other parcels have different sizes of damaged areas (see Figure 41).

Figure 41. Spring barley parcels at the end of the growing season, July 5th, 2021

Across the selected agricultural parcels, sowing of spring barley took place in late April, while dates of harvesting vary from mid-July in case of parcel #125, which showed a rapid development, to early August in case of other croplands.

4.3.1. Parcel-based time series analysis

Examples of previously described winter crops reflect the damage that occurs at the early stages of development. This is however not always the case, and to depict a growth pattern with mid-season crop damage, spring barley cropland #111 was taken for parcel-based time series analysis.

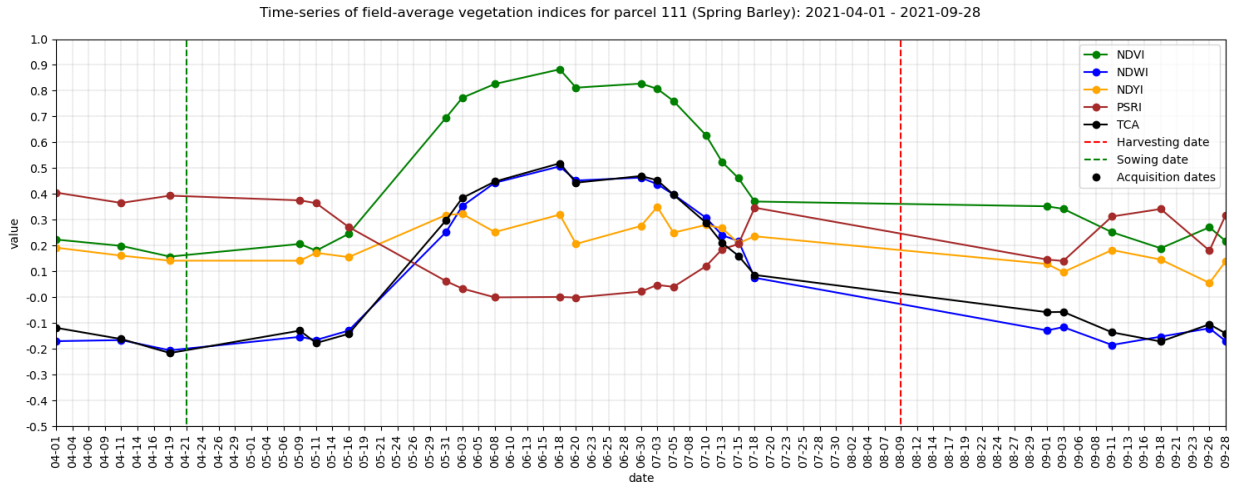


Figure 42. Time series of vegetation indices for parcel #111

From Figure 42 the sowing took place on the 21st of April while crops were harvested on the 9th of August. Conveniently, most of the acquisitions are available during growth, and only the second part of July and August don't have cloud-free images. Nevertheless, the general growth pattern is visible and resembles the one of winter wheat as both plants "green-up" and "green-down".

Immediately after sowing the level of biomass remains steady at a level of bare soil of ~0.2 until the 11th of May when it starts to grow. The growth continues until the 18th of June when it peaks at a level of 0.89. The peak is followed by a sudden drop on the next acquisition date by ~0.1 which is followed by stabilization at a level of 0.8. At the beginning of July, the crops start to ripen resulting in NDVI level decrease until the 18th of July which is the last available cloud-free acquisition before the harvesting.

The levels of NDWI and TCA are identical, while general time variations are similar to the biomass level. Both indices start at a negative plateau of -0.2 with steady growth from the 11th of May to the 18th of June when the indices peak at a level of 0.5. NDWI and TCA then experience the same drop by 0.1 on the 20th of Jule with a gradual decrease during the ripening from the 3rd of July to the 18th of July.

PSRI index was at a level of 0.4 before and immediately after sowing, with a gradual decrease to the level of 0 during the time when crops were standing green, and a gradual increase to the level of 0.34 as crops were ripening.

NDYI was fluctuating mostly between 0.1 and 0.30 with expected no signs of flowering.

4.3.2. Clustering analysis

A short parcel overview and spatial patterns of vegetation indices clustering are provided in Figure 43. Field #111 experienced a delay in the crop development mostly in the central part of the parcel. As of July 5th, the size of the problematic area has increased towards the North-Western side of the parcels with a clear indication of the damaged crops. The negative tendency within the damaged area continued until the harvesting (see Figure 44).

Visually, TCA provides the best overlap with damaged and no damaged areas, while NDVI, NDWI, and PSRI partly resemble validation grid spatial patterns.

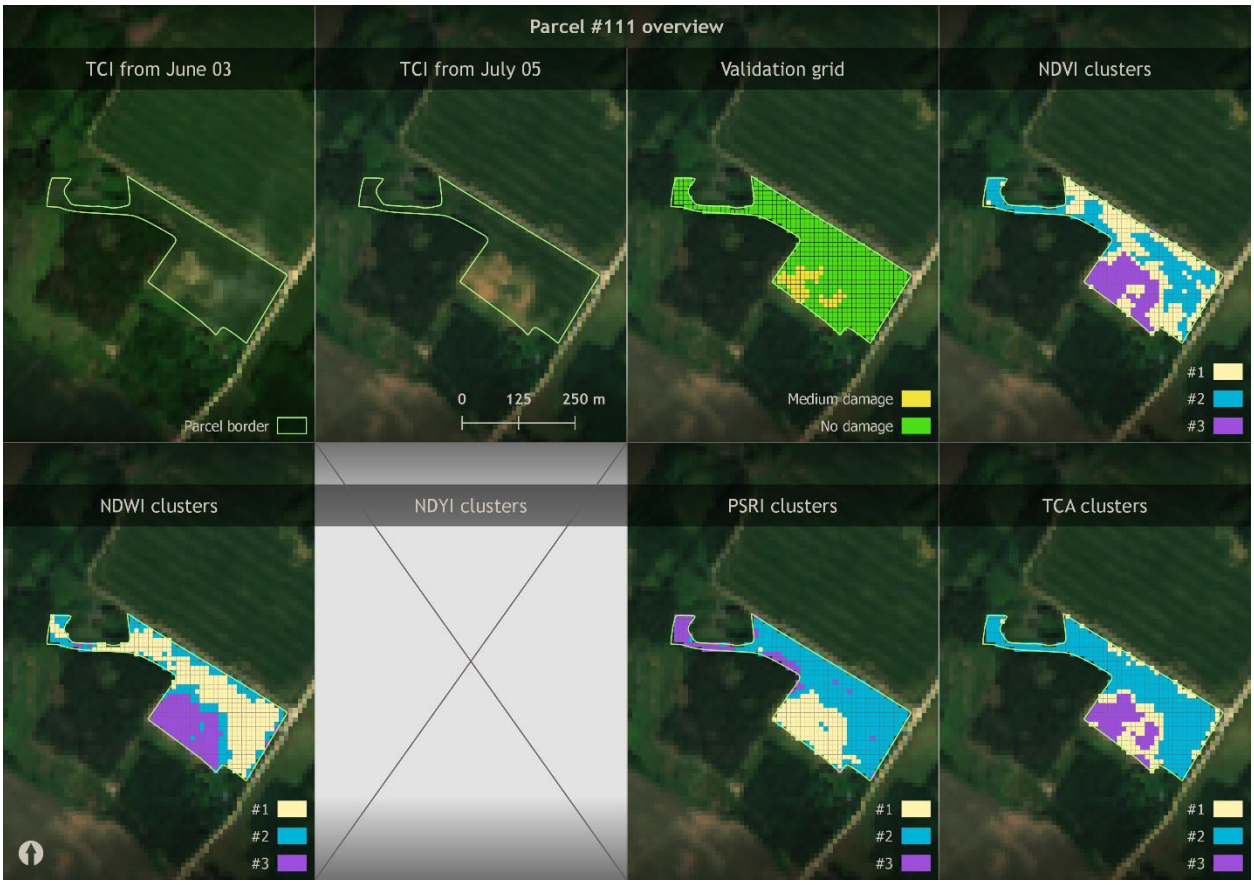


Figure 43. Overview and clustering results visualization for spring barley parcel #111

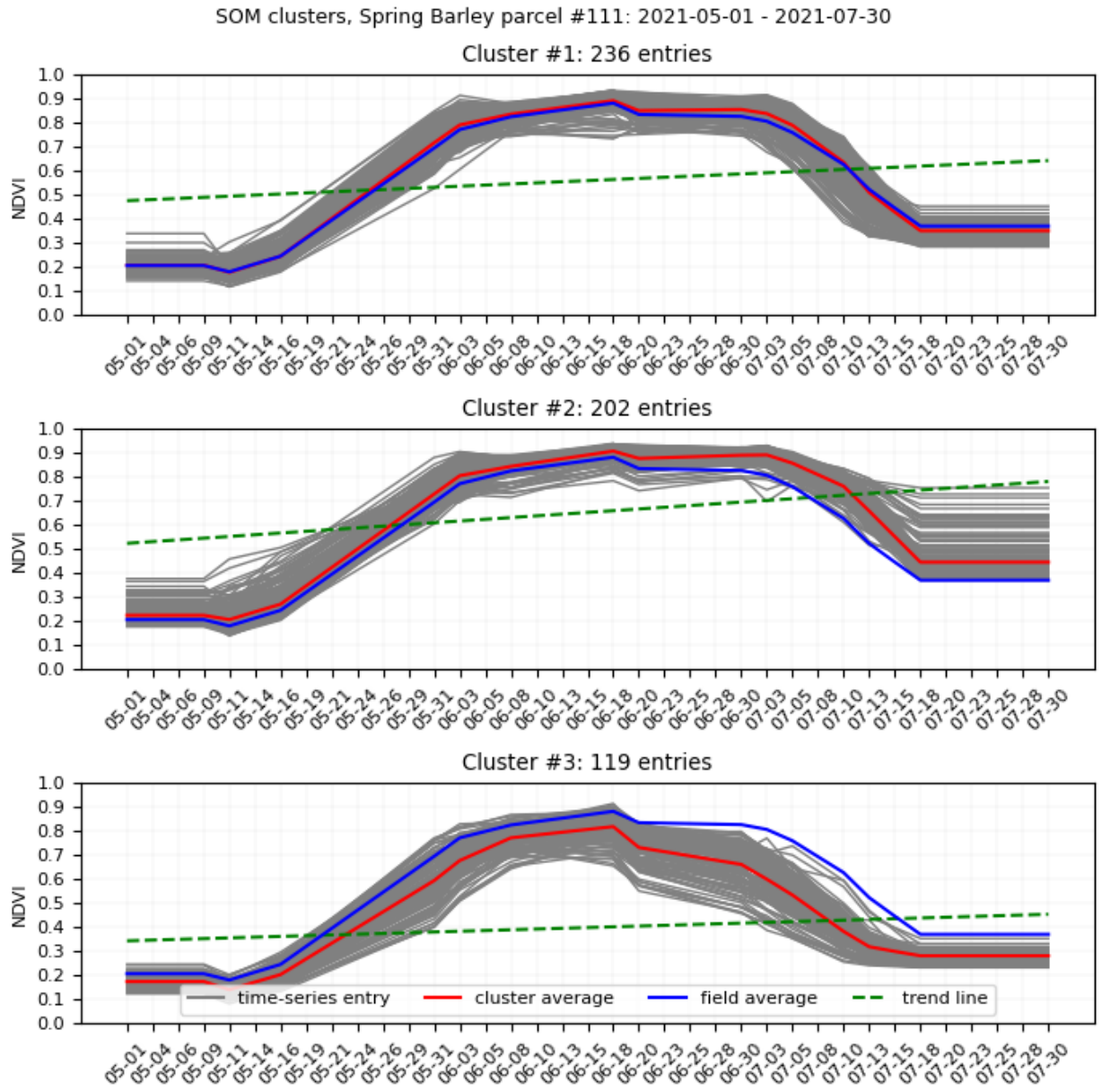


Figure 44. NDVI SOM clusters, spring barley parcel #111

NDVI cluster #3 overlaps with the "medium damage" zone depicted in the validation grid, while clusters #1 and #2 reflect healthy conditions on the field. It is noticeable that the overall growing tendency is weaker in cluster #3, as, for example, the lowest biomass value of cluster #3 on the 13th of June is about 0.68 versus 0.73 for clusters #1 and #2 (see Figure 44). In addition, the sudden biomass drop on the 18th of June is more significant among entries of cluster #3 along with a subsequent rapid decline in the biomass index immediately after the drop. The result of a rapid decline in the damage zone can be seen as almost twice lower NDVI values compared with clusters that overlap with the "no damage" zone and have biomass level of about 0.8-0.9 on the 30th of July before the ripening. Based on the cluster statistics (see Table 10), cluster #3 is characterized by the smallest values across all three groups.

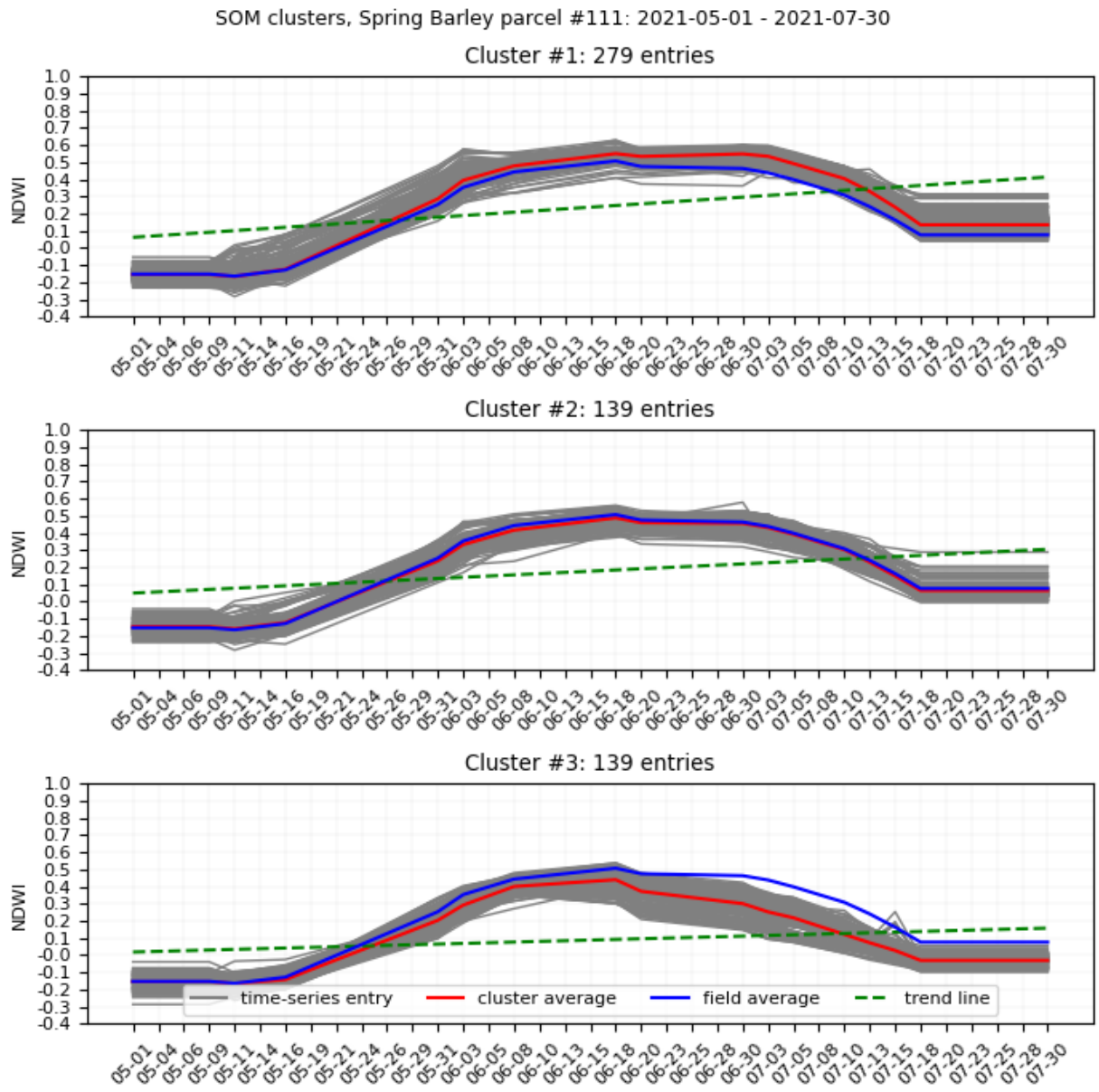


Figure 45. NDWI SOM clusters, spring barley parcel #111

NDWI cluster #3 overlaps with the "medium damage" class presented on the validation grid and the field and can be characterized as the weakest (Table 10). Based on the time series from Figure 45, noticeable differences between clusters #3 and clusters #2 and #1 start with the decline after the 18th of June, which is the most prominent among entries of cluster #3. Later, on the 10th of July, the damaged areas have 0 level of water content, which indicates almost dry vegetation, compared with "no damage" zones with moisture content at a level of 0.4 (see Figure 45).

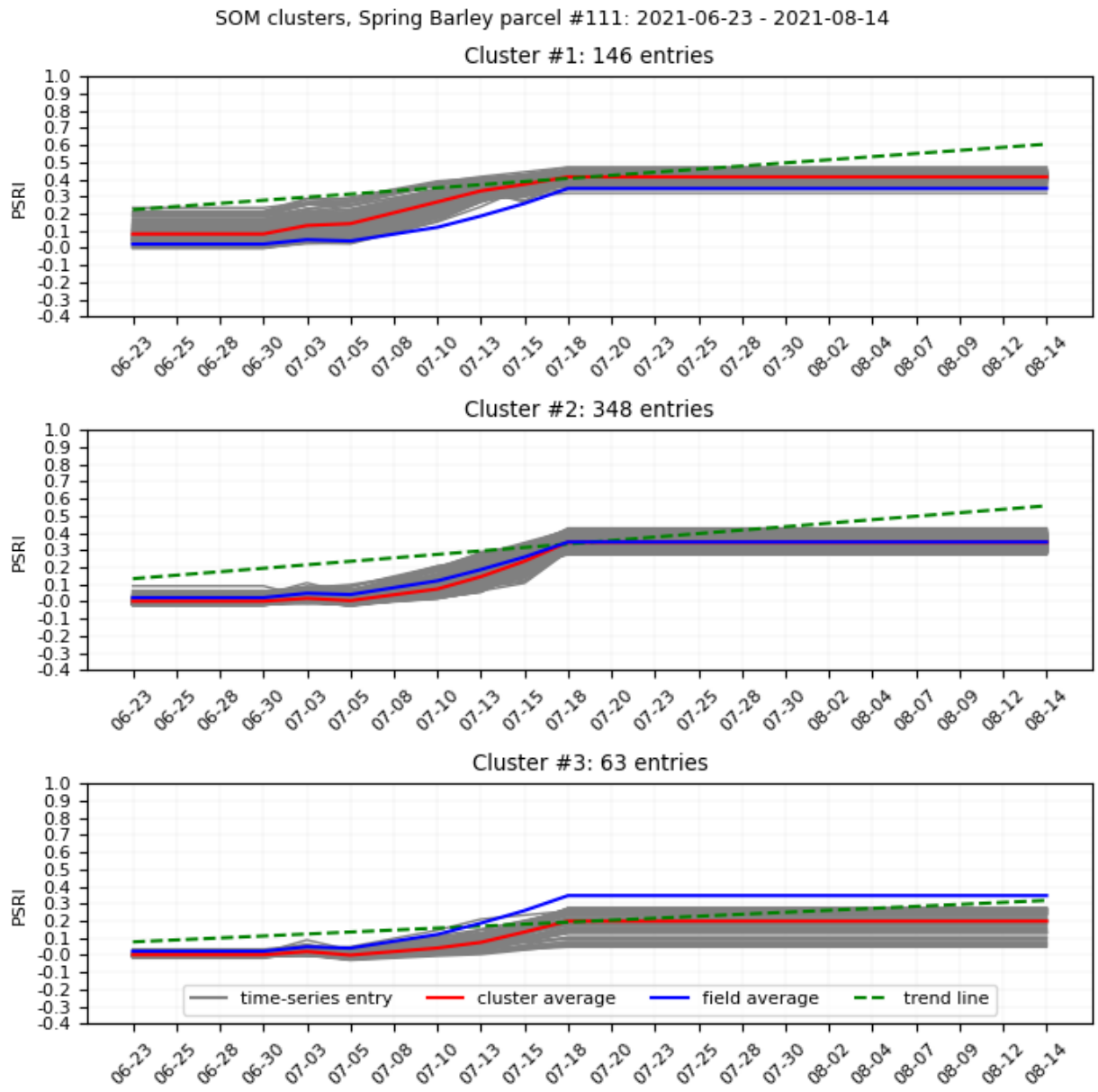


Figure 46. PSRI SOM clusters, spring barley parcel #111

PSRI cluster #1 overlaps with the "medium damage" zone, and #2 overlaps with the "no damage" zone. As the damage started to occur mid-season, cluster #1 has comparatively high values of PSRI that indicate a low level of greenness among entries (see Figure 46). Such early growth doesn't occur in clusters #2, and #3, which mostly represent areas on the outskirts of the parcel. Cluster #2 shows a late expected PSRI increase from ~0 to ~0.4 during the process of ripening. However, based on the statistical values (see Table 10) and time series variations (Figure 46) across the generated clusters, it is difficult to distinguish areas which groups correspond to damaged or non-damaged areas, as minimum and maximum are similar.

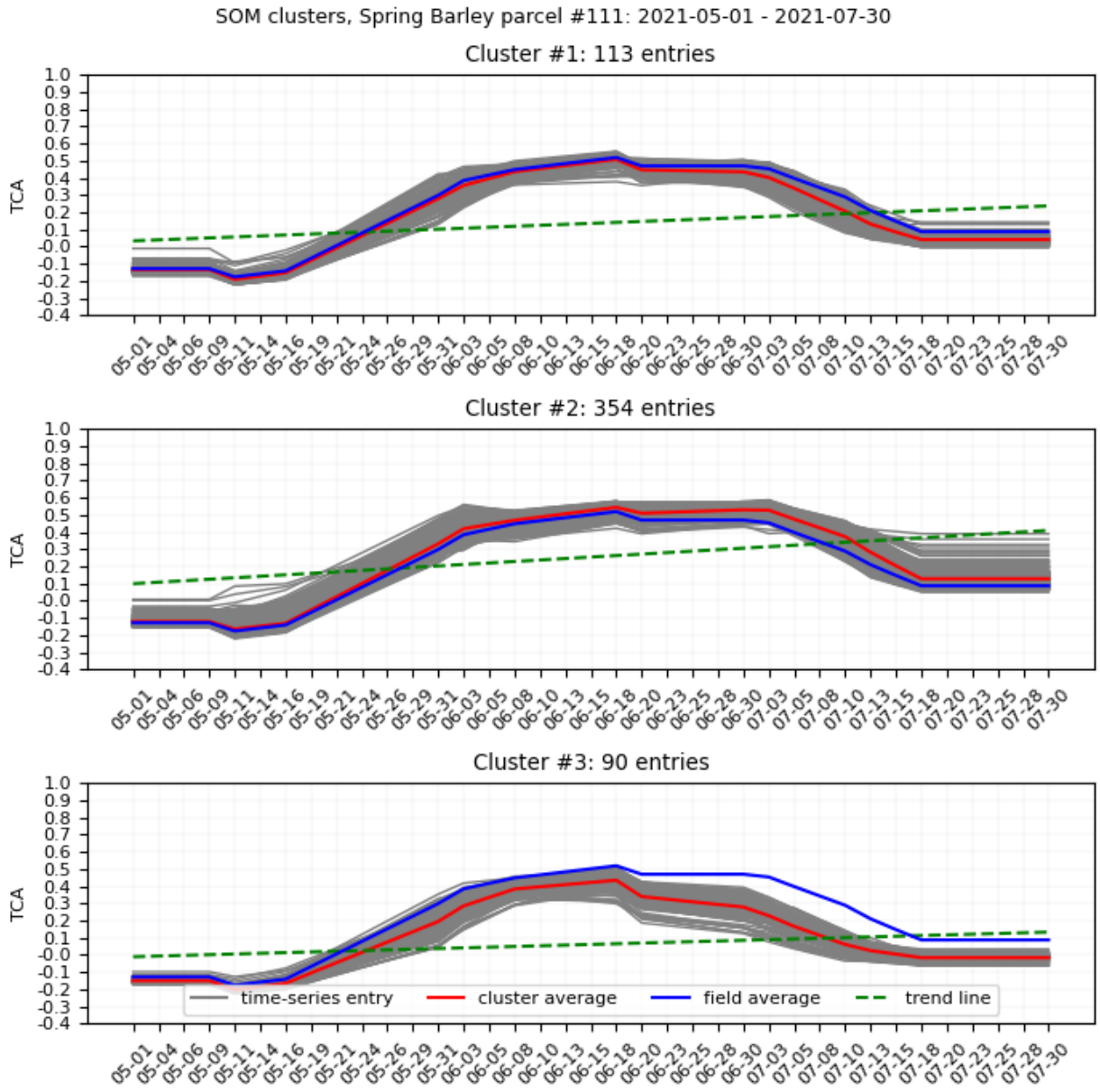


Figure 47. TCA SOM clusters, spring barley parcel #111

Lastly, in case of TCA, cluster #3 overlaps with the "medium damage" class presented on the validation grid. The cluster statistics show the lowest values across generated groups (see Figure 47, Table 10). The index behaves similarly to NDVI with the most prominent differences between no damage and damaged zones from the 18th of June (see Figure 47).

Table 10. Averaged values per generated clusters for spring barley parcel #111

VI	#	std.	average	median	min	max	sum	trend	slope	intercept
NDVI	1	0.257	0.563	0.560	0.179	0.893	20.826	no trend	0.0046	0.477
	2	0.251	0.608	0.653	0.207	0.908	22.494	increasing	0.0071	0.524
	3	0.226	0.458	0.399	0.141	0.819	16.957	no trend	0.0031	0.344
NDWI	1	0.257	0.243	0.237	-0.167	0.550	8.992	increasing	0.0098	0.062
	2	0.227	0.189	0.177	-0.160	0.488	7.001	increasing	0.0071	0.049
	3	0.205	0.118	0.088	-0.166	0.440	4.367	no trend	0.0039	0.018
PSRI	1	0.138	0.307	0.415	0.081	0.415	6.751	increasing	0.0182	0.224
	2	0.155	0.212	0.346	0.001	0.346	4.672	increasing	0.0203	0.133
	3	0.088	0.121	0.198	0.000	0.198	2.671	increasing	0.0115	0.077
TCA	1	0.230	0.177	0.135	-0.191	0.508	6.537	no trend	0.0057	0.033
	2	0.246	0.242	0.254	-0.166	0.544	8.942	increasing	0.0086	0.099
	3	0.200	0.103	0.061	-0.203	0.435	3.818	no trend	0.004	-0.011

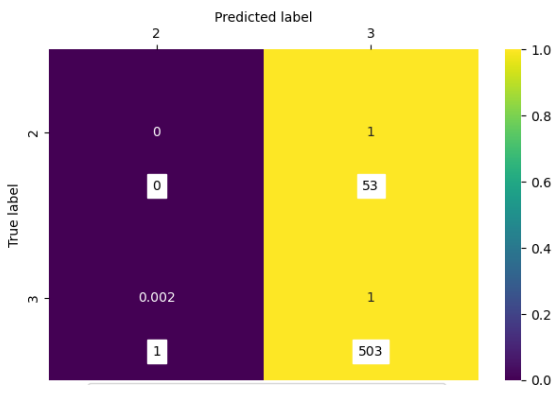
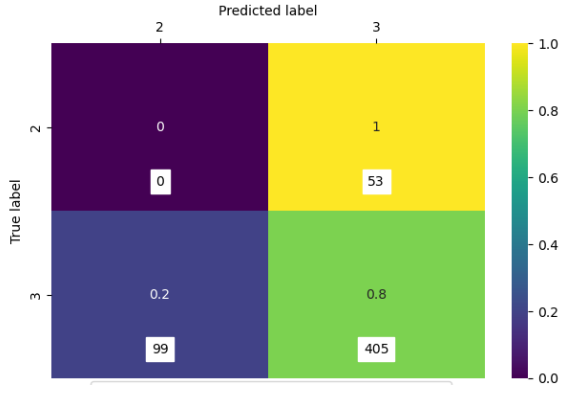
4.3.3. Classification and validation

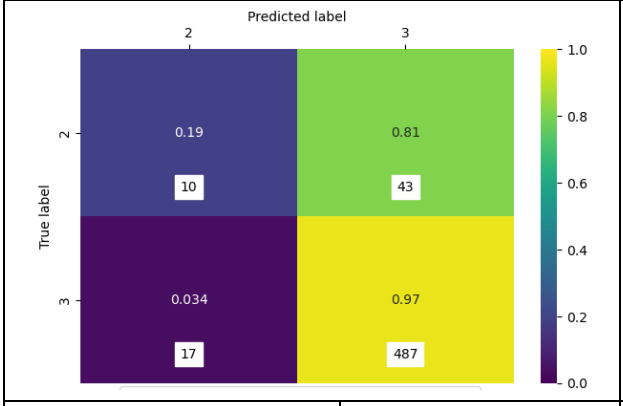
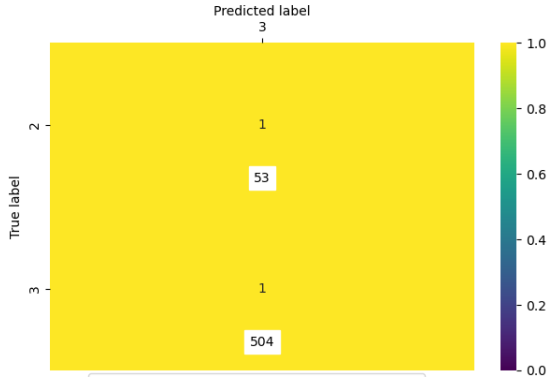
Due to similarities in the development of winter wheat and spring barley, analogous decision trees provided in Figure 29 were used for spring barley classification.

Normalized confusion matrices of the classification results of agricultural parcel #111 are provided in Table 11, while results across spring barley test parcels are provided in Table 12. Non-normalized class counts are presented in white rectangles within confusion matrices' classes.

From the Table 11 time series classification of all vegetation indices performed poorly because of the temporal pattern of the damage, including NDWI, where only 19% of the "medium damage" group was classified correctly. Due to the fact that the "bare soil" group was not presented in the field, it is not displayed in the confusion matrices. The spatial pattern of the results is provided in Figure 48.

Table 11. Confusion matrices for the classifications performed on spring barley agricultural parcel #111: (2) "medium damage", (3) "no damage" zones

NDVI confusion matrix		PSRI confusion matrix	
			
Precision: 0.452	Recall: 0.499	Precision: 0.442	Recall: 0.402
F1 score: 0.475		F1 score: 0.421	

NDWI confusion matrix		TCA confusion matrix	
			
Precision: 0.645	Recall: 0.577	Precision: 0.452	Recall: 0.5
F1 score: 0.596		F1 score: 0.475	

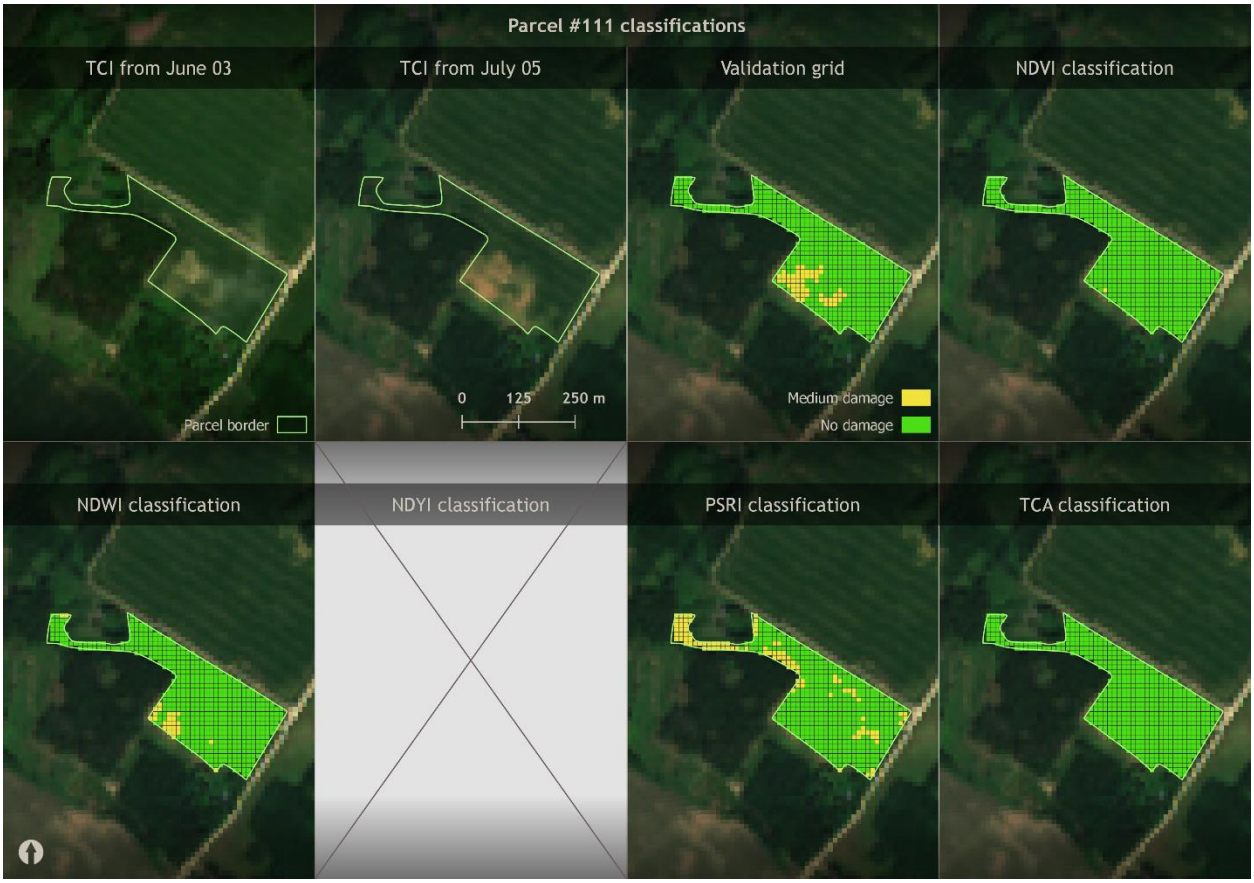
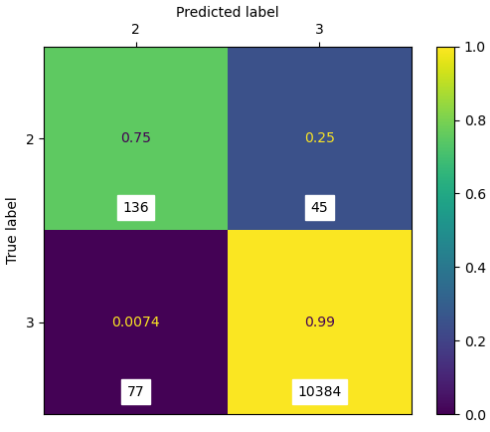
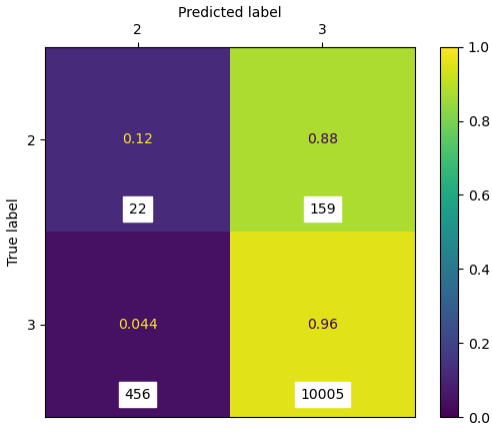
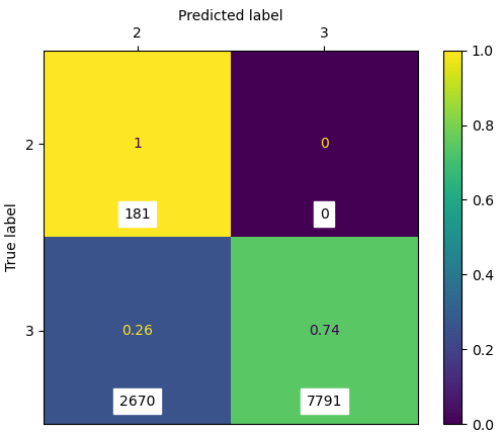
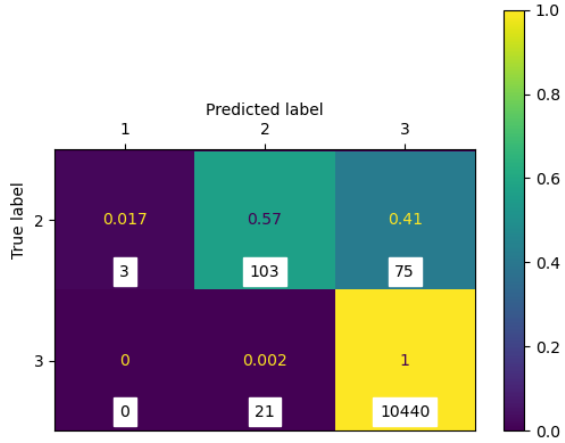


Figure 48. Classification results for spring barley agricultural parcel #111

The classification results across test parcels of spring barley are drastically different, as NDVI scores the first in classification with 99% correct classification in the “no damage” and 75% in “medium damage” zones. The biomass index is followed by TCA which has a smaller amount of the “medium damage” area classified correctly – only 57% with consequent smaller precision and recall. NDWI and PSRI have the poorest performance with the lowest F1-scores among presented matrices.

Table 12. Confusion matrices for the classifications performed on test spring barley agricultural parcels: (2) “medium damage”, (3) “no damage” zones

<table>
<tr><th colspan="2">NDVI confusion matrix</th><th colspan="2">PSRI confusion matrix</th></tr>
<tr><td colspan="2"> 
</td><td colspan="2"> 
</td></tr>
<tr><td>Precision: 0.817</td><td>Recall: 0.872</td><td>Precision: 0.515</td><td>Recall: 0.539</td></tr>
<tr><td>F1 score: 0.842</td><td></td><td>F1 score: 0.518</td><td></td></tr>
<tr><th colspan="2">NDWI confusion matrix</th><th colspan="2">TCA confusion matrix</th></tr>
<tr><td colspan="2"> 
</td><td colspan="2"> 
</td></tr>
<tr><td>Precision: 0.532</td><td>Recall: 0.872</td><td>Precision: 0.608</td><td>Recall: 0.522</td></tr>
<tr><td>F1 score: 0.487</td><td></td><td>F1 score: 0.557</td><td></td></tr>
</table>

5. Discussion

Since remote sensing provides a wide range of opportunities for crop monitoring, EO data is utilized for a variety of agricultural monitoring tasks, while damage assessment is one of the most useful use cases of EO data for farmers and insurance companies. The results of this study will be compared with other studies that focused on temporal and spatial patterns of crops and clustering.

5.1. Clustering

In this study, a self-organizing maps clustering algorithm was applied to the time series of vegetation indices per pixel. It was found that clustering allows to find similarities in crop development over time and group them together. On a temporal scale from sowing to harvesting, vegetation indices such as NDVI, NDWI, and TCA are of great value, while more specific indices, for example, NDYI and PSRI, that focus on flowering and senescence, can be used to estimate temporal variations during important phenological stages of the plants growth. Deviations of the vegetation indices which are expressed on a temporal scale will result in a separate group defined by a clustering algorithm. On a spatial scale, the groups delineate areas that reflect homogenous growing patterns regardless of overall field condition.

Such a temporal perspective on the development of the crops, extends a spatial theory and confirms an assumption made by Ghosh et al. (2018), who concluded that inter-field variation of vegetation indices can be used for crop monitoring (Ghosh et al., 2018). Similarly, the main results of Sosa et al. (2021), who employed k-means clustering of standard deviation values of microwave and spectral vegetation indices before and after hailstorm damage, show that k-means allows dividing the area into homogeneous damage zones. The clustering can be applied to croplands in any conditions, but in case of uniform development across the field, the algorithm is forced to artificially divide the area into zones that statistically have no difference from each other. Similar behavior is observed in the current study (see clusters #2 and #3, Figure 24), but as it was concluded by Sosa et al. (2021), it is not a mistake of the classification algorithm, but rather that all generated groups, in fact, belong to one.

Sole clustering of damaged fields into three spatial zones was enough in the scope of work by Sosa et al. (2021), for the reason that the study aimed to support decisions made by insurance claim adjusters. However, in the scope of this work, which is focused on parcels with both healthy and unhealthy vegetation, bare clustering results are not enough, since clustering itself cannot provide enough information about the state of vegetation in the generated zones. To provide zones that estimate damage degree, classification based on temporal statistics was performed.

5.2. Temporal variations of vegetation indices

General temporal patterns and clustering of selected spectral vegetation indices provided an extensive overview of the agricultural parcels of studied winter and spring crop types.

Despite being different kinds of crops that are sowed at different times, the main findings in case of winter wheat and spring barley are similar, as the growth pattern consists of a gradual increase

followed by a decrease in the amount of biomass. An analogous pattern is observed by Ghosh et al. (2018), who studied potato crop using Sentinel-2 imagery. Across winter wheat and spring barley, NDVI index allows differentiating between damaged areas with dry plants and healthy crops, where damage will stand out with low values of the index. The values of non-damaged vegetation observed in this study are alike to ones obtained for potato crop by Ghosh et al. (2018) or arctic plants by Zagajewski et al. (2017): biomass level lower than 0.2 indicates bare soli, values higher than 0.7 is a clear proof of no damage per unit area. These extreme values match the winter rapeseed growing pattern as well, but the mid-season flowering stage should be taken into account.

One of the weaknesses of this study is not a sufficient amount of ground truth information, which introduces difficulties with NDVI interpretation. An additional index in form of TCA helps with this process. Liu et al. (2018) who studied desertification processes using tasseled cap transformation indices, obtained values of –0.05 for TCA in desert lands over several years. Similar results were obtained by Gómez et al. (2011), who interpreted TCA values below zero as non-vegetated areas. Such interpretation corresponds with the temporal pattern of winter wheat and winter rapeseed, for example, during flowering (see Figure 38) or after harvesting (see Figure 23).

In case of NDWI, 5 years-long analysis of drought using MODIS data, Gu et al. (2007) concluded, that NDWI for non-drought classified areas exceeds 0.4, while zones with moderate and severe drought have values under 0.3, with bare soil below -0.1– 0 (Gao, 1996). Similar values with light deviations up to 10% were obtained for winter wheat and winter rapeseed at their peak of growth, while the clear difference between damaged and not damaged spring barley areas can be seen in Figure 45. The index shows a remarkable behavior during winter rapeseed flowering, as it remains sensitive to the amount of water in the crops despite color change (see Figure 32).

Since NDYI was purposefully chosen to estimate flowering vegetation, the index was not useful for winter wheat and rapeseed. For winter rapeseed, in turn, temporal variations of the index from damaged areas clearly show low flowering intensity (see Figure 36) compared with non-damaged areas.

PSRI index is not as widespread as the other indices presented in the study. This makes it difficult to find its application to various crop types. Within this study, the index showed response not only to the process of senescence, but also to the amount of yellow reflected by yellow dry stems on the fields (see Figure 15), for example, high values in clusters #1 and #2 (see Figure 36). It makes the usage of the index complicated for damage estimation during the ripening stage.

5.3. Time series classification

Time series classification was performed using simple decision trees with thresholds for different vegetation densities, which are sensitive to various problems and reflect homogeneous damage zones within agricultural parcels.

Since the main damage on the selected agricultural parcels was drought, which causes crops to dry and change their color to brown/yellow, vegetation indices such as NDVI, NDWI, and TCA were

sensitive to the damage. Using defined thresholds for the maximum value of time series, the classification performed well on those agricultural parcels, where the damage occurred at the early stages of crop development. In case of winter crop types, the damage that happened during the winter and spring seasons was seen starting from the very beginning of the time series in April. As damage moves towards mid-season and end-season, it becomes more difficult to be spotted using a simple maximum parameter of the time series, as it will reflect healthy plants at the beginning of growth. Similarly, weed growth in the areas with early damage introduces distortions (see Figure 34) making NDVI, NDWI, and TCA unable to differentiate between grass and actual crops. These are major issues that don't allow to obtain accurate classification results for some temporal damage patterns.

The issue can be partly solved by using additional vegetation indices that focus on particular phenological stages of the crops. As NDYI can be easily detected (D'Andrimont et al., 2020) using the time series of Sentinel-2, simple thresholds can accurately classify the parcels into homogeneous damage zones (see Table 8). During ripening, the results obtained from PRSI of senescence vegetation turned out to be similar to damaged dry areas, which caused uncertainties in the classifications that lead to poor results.

To summarize, a set of advantages can be highlighted, as the analysis provided an extensive overview of temporal changes of vegetation indices that reflect various temporal patterns, taking into account phenological features of the crops around Tartu city in Estonia. Time series of vegetation indices are sensitive to changes and clustering can detect those. The unsupervised approach, which is computationally inexpensive, doesn't require big datasets or historical data, and can accurately estimate areas that experienced damage at the early stages of crop development or during important developing stages, such as flowering. At the same time, there are several drawbacks, connected both to the data and classification approach. For example, due to an insufficient amount of ground truth data, it was difficult to avoid ambiguous interpretations of vegetation indices. In addition, calculated statistics is greatly dependent on the sensed time intervals and spatial resolution (Gobbo et al., 2021), as well as it varies even across parcels of the same type. That implies substantial obstacles for unsupervised classification, as universal threshold values cannot be used and relative differences within the agricultural parcel should be prioritized.

6. Conclusion

As a result of this study, it becomes clear that it is possible to detect damaged areas of vegetation on agricultural parcels using solely Sentinel-2 spatio-temporal data. Spectral transformations in form of vegetation indices allow to increase the difference among various vegetation conditions on the ground, while simple data extraction techniques, such as clustering, can generalize temporal variations of vegetation indices to produce a spatio-temporal pattern for the area. Thus, answering the question if an unsupervised algorithm can detect variations in crop development within agricultural parcels, it is true to say that an unsupervised self-organizing maps algorithm can detect variations in crop development within agricultural parcels.

Although the data analyzing approach provided in the current study is simple, the data should be utilized carefully. The answer to the question which vegetation indices provide the most information about anomalies is not straightforward. Generally, in case of plants that "green-up and down", NDVI, NDWI, and TCA are the most useful, as they provide the most accurate representation of cropland conditions in case of drought-induced damage. The indices can be supported by NDYI for winter rapeseed, as there are existing methods for flowering detection and this work shows its efficiency for damage delineation. Visually, PSRI provides the lowest sensitivity if measured shortly before and during the senescence as it shows similar results for dry and ready-to-be harvested crops.

As some studies suggest, basic descriptive statistics can be used to characterize damage occurring in the field. A successfully applied in another work standard deviation did not provide any useful information in the current study, so answering the question what statistical parameters of the time series can be used for classification, the "maximum" parameter of a time series can be used, as it was exploited to perform a scene classification in the current work. The feature demonstrated a possibility to differentiate between healthy areas and regular damage on the agricultural parcels if used as a part of a decision tree. The parameter, however, is affected by emerging weed and should be applied carefully. At the same time, there is not enough scientific evidence to utilize other characteristics of the time series, such as minimum value, trend, or median, as they greatly depend on time intervals used for calculation.

To achieve a better performance in unsupervised classification, a novel approach that focuses on particular stages of crop development is required, while more advanced supervised techniques, such as random forest classifier, can be of great value as well.

Summary

The thesis aimed to analyze Sentinel-2 time series data to perform an unsupervised classification into damaged and healthy crop zones within agricultural parcels. The classification was to be performed at the end of the agricultural season so afterward it can be used as a part of an automatic damage claim report.

The aim of the work was achieved by first analyzing Sentinel-2 time series data in form of vegetation indices on a parcel-wide scale; second, applying self-organizing maps clustering algorithm to the time series of vegetation indices; third, classifying pixels' time series into several zones that represent damage zones within individual agricultural parcels of winter wheat, winter rapeseed, and spring barley during 2020-2021 agricultural season.

The main findings of the work show that on a time scale from early April to late July, NDVI, NDWI, and TCA vegetation indices are sensitive to various growth patterns and ultimately have different signatures for healthy green and damaged crops. On a shorter sensing interval of one month, NDYI exhibits peaks for healthy flowering rapeseed and remains flat for damaged areas. During the senescence stage, PSRI provides similar results between drought-induced damaged areas and healthy vegetation, making it difficult to differentiate between the zones.

Clustering of the abovementioned vegetation indices grouped similar temporal variations and revealed spatial patterns, that generally outline damaged and healthy zones on croplands. Descriptive statistics obtained from the generalized data was primarily affected by the sensed period, and it showed variability not only between damaged and healthy zones within individual fields, but also among the same zone type across all agricultural parcels of a type.

Due to inconsistency among values from various parcels, the classification was performed using only the "maximum" parameter of the time series as an indicator of damaged or healthy vegetation. The accuracy of classification of NDVI, NDWI, and TCA reaches 80% in case of croplands that experience early crop damage, the accuracy of NDYI classification during winter rapeseed flowering provides up to 73% correctly classified areas if no distortions caused by weed were introduced.

Ajalis-ruumiliste Sentinel-2 andmete klasteranalüüs kahjustatud põllukultuuride piiritlemiseks Eesti põllumajandusmaal

Anton Kostiukhin

Kokkuvõte

Tänapäeval põhjustab põllumajandussektorile suurimat ohtu maailma pidevalt kasvava rahvaarvuga kaasnev tootmisvajaduse tõus ning globaalse soojenemise ja ettenägematute ilmastikunähtuste tõttu suurenenud saagi kahjustumine (Carter et al., 2018). Saagikuse vähenemise rahaliste kahjude leevendamiseks lasevad põllumehed oma saagi kindlustada. Kindlustusfirmad kasutavad kahjude hindamiseks tõõjõumahukaid lokaalseid meetodeid, muutes kindlustuskahjude väljamaksmise keeruliseks ja kallutatuks (Gobbo et al., 2021). Kaugseire pakub potentsiaalset alternatiivi kohapealsetele meetoditele, hõlbustades kahjude hindamise protsessi nii põllumeestele kui ka kindlustusfirmadele (Garcia Millan et al., 2020). Sentinel-2 A ja B satelliidid on ühed kõrgetasemelisemad maaseire instrumendid, pakkudes kõrgekvaliteedilisi sisendandmeid olemasolevatele maapinna monitoorimise meetoditele (Tamm et al., 2016).

Antud magistritöö eesmärgiks on analüüsida Sentinel-2 aegridade andmeid ja teostada juhendamata klassifikatsioon, mis klassifitseerib alad tervete ja kahjustatud põllukultuuride tsoonidesse põllumajandushooaja lõpus. Antud klassifikatsioon võimaldab automatiseerida kahjude tuvastamise ja hindamise protsesse. Magistritöö eesmärgi täitmiseks püstitati järgmised uurimisküsimused:

1. Kas juhendamata algoritm suudab tuvastada variatsioonid taimede arengus põllumajanduslike maa-alade raames?
2. Millised optilised taimestikuindeksid annavad enim teavet saagikahjustuste kohta?
3. Milliseid aegridade analüüsi statistilisi parameetreid saab milliste taimestikuindeksite puhul kasutada, et klassifitseerida taimekasvualad kahjustatud ja terveteks tsoonideks?

Üleilmselt on vaid ükskikud uurimused vaadelnud saagikahjude piiritlemist. Olemasolevad uurimused kasutavad peamiselt taimestikuindekseid, mis on põllukultuuride omaduste suhtes tundlikumad kui üksikud spektriribad (Segarra et al., 2020) ning näitavad viljatüüpide ruumilist varieerumist erinevates seisundites aladel (Ghosh et al., 2018). Taimestikuindeksid ühildatakse kas arvutuslikult kulukate juhendatud klassifitseerimistehnikatega, kiirete ja lihtsalt rakendatavate juhendamata klassifitseerimistehnikatega (Phiri et al., 2020; Sosa et al., 2021) või saagikuse hindamise mudelitega (Gobbo et al., 2021). Juhendamata meetodeid kasutavad uuringud keskenduvad otsustustoetussüsteemidele, mis klassifitseerivad maa-alad sünteetilistesse tsoonidesse. Selline saagikahjude tuvastamine nõuab täiendavat järelevalvet ega kajasta alati põllumaade ruumilist varieeruvust, kuna klastermeetodid rühmitavad maa-ala olenemata põllukultuuride seisundist(Sosa et al., 2021). Seetõttu ei ole üheski uuringus põllukultuuride kahjustusi manuaalselt määratletud klassidesse klassifitseeritud. Eestis on taolisi uuringuid tehtud väga vähe, mistõttu on juhendamata meetodite rakendamine puudulik.

Käesolevas uuringus kasutatud andmed koosnevad 73 Sentinel-2 L2A sisendist, väliuuringutelt saadud teabest põllukultuuride seisundi kohta, aeropiltidest (Maa-ameti Ortofoto, 2021), maatükkide piiridest ja põldude taustainfost, mis saadi KappaZeta OÜ-st.

Antud töö metoodika jaguneb mitmeks etapiks. Esiteks eeltöödeldi andmed, eemaldades piltidelt pilved KappaMask (Domnich et al., 2021) pilvemaski abil, arvutati taimestikuindeksid GDAL raamatukogu abil, ning lõigati rastrid kasutades põllumaade piire. Teiseks loodi üksikute maatükkide sees iga piksli jaoks taimestikuindeksite aegrida. Kolmandaks rakendati üksikute maatükkide taimestikuindeksite aegridadele iseorganiseeruvate kaartide (SOM) (Vettigli, 2018) klastrite algoritm. Klastrite põhjal arvutati kirjeldav statistika. Neljandaks loodi statistiliste väärtuste põhjal aegridade klassifitseerimise otsustuspuud. Lõpuks valideeriti klassifikatsioon valideerimisruudustiku suhtes kasutades segadusmaatrikseid.

Tulemustest selgub, et põllumajandushooajal on taimestikuindeksid tundlikud põllukultuuride seisundi suhtes ja neil on erinevad tunnusmärgid tervena arenevatele ja kahjustatud põllukultuuridele. Üldisi kasvumustreid saab jälgida NDVI, NDWI ja TCA ajalistest variatsioonidest, mis näitavad vastavalt rohelust, niiskusesisaldust ning taimestikuga kaetud ja katmata alade suhet pindalaühiku kohta. Kuna mõnel taimel on spetsiifilised arengufaasid, näiteks rapsi õitsemise faas, saab NDYI-d kasutada kahjustatud ja terve rapsiseemne erinevuse tuvastamiseks. Vananemisstaadiumis annab PSRI sarnaseid tulemusi põua poolt kahjustatud piirkondade ja terve taimestiku vahel, mistõttu on tsoonide eristamine keeruline.

Taimestikuindeksite aegridade klasterdamine näitab kasvumustrite ajalis-ruumilisi variatsioone. Ruumilised mustrid vastavad üldiselt kahjustatud ja terve taimestikuga aladele. Klastrite kirjeldav statistika näitas eeldatavaid erinevusi kahjustatud ja tervete tsoonide vahel üksikute maatükkide raames. Erinevusi leidus aga ka sama tsoonitüübi puhul mitmel identse põllukultuuritüübiga maatükil. See näitab, et kahjustatud ja tervete põllukultuuride tunnusmärgid on põldudel erinevad, mistõttu on raske määratleda väga täpseid mustreid. Lisaks mõjutavad tulemusi vaadeldav periood ja põllukultuuride arengu ajalised muutused.

Kirjeldava statistika varieeruvuse tõttu viidi klassifikatsioon läbi ainult „maksimaalse“ aegrea parameetri abil. See parameeter toimib indikaatorina, mis eristab kahjustatud taimestikuga alasid terve taimestikuga aladest. NDVI, NDWI ja TCA klassifikatsiooni täpsus ulatub varajase saagikahjustusega põllumaadel 80%-ni. NDYI klassifikatsioon prognoosib talirapsi õitsemise ajal õigesti kuni 73% aladest. Mida vähem on põldudel deformatsioone, seda täpsem on klassifikatsioon, mistõttu võib klassifitseerimise täpsus näiteks umbrohu korral langeda alla 50%.

Acknowledgements

I would like to thank my supervisors, Alexander Kmoch, Tanel Tamm, and Indrek Sünter, for guidance and technical support throughout the work. I am also grateful to the whole KappaZeta team for their support and feedback about the work. I am exceptionally thankful to professor Tõnu Oja from University of Tartu for joining filed surveys and drone flying.

References

Abdi, A.M., 2020. Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data. *GIScience & Remote Sensing*, *57*(1), pp.1-20.

Baimisheva, T.A., Kurmaeva, I.S., Gazizyanova, Y.Y., Baimeshev, R.H., Aiesheva, G.A., 2019. State regulation systems of agricultural insurance. In *IOP Conference Series: Earth and Environmental Science* (Vol. 315, No. 2, p. 022090). IOP Publishing.

Bannari, A., Morin, D., Bonn, F., Huete, A., 1995. A review of vegetation indices. *Remote sensing reviews*, *13*(1-2), pp.95-120.

Carlson, T.N., Ripley, D.A., 1997. On the relation between NDVI, fractional vegetation cover, and leaf area index. *Remote sensing of Environment*, *62*(3), pp.241-252.

Carter, C., Cui, X., Ghanem, D., Mérel, P., 2018. Identifying the economic impacts of climate change on agriculture. *Annual Review of Resource Economics*, *10*, pp.361-380.

Cook, B.I., Mankin, J.S., Anchukaitis, K.J., 2018. Climate change and drought: From past to future. *Current Climate Change Reports*, *4*(2), pp.164-179.

Crist, E.P., Cicone, R.C., 1984. A physically-based transformation of Thematic Mapper data---The TM Tasseled Cap. *IEEE Transactions on Geoscience and Remote sensing*, (3), pp.256-263.

D'Andrimont, R., Taymans, M., Lemoine, G., Ceglar, A., Yordanov, M., van der Velde, M., 2020. Detecting flowering phenology in oil seed rape parcels with Sentinel-1 and-2 time series. *Remote sensing of environment*, *239*, p.111660.

Das, P.K., Seshasai, M.V.R., 2015. Multispectral sensor spectral resolution simulations for generation of hyperspectral vegetation indices from Hyperion data. *Geocarto International*, *30*(6), pp.686-700.

Domnich, M., Sünter, I., Trofimov, H., Wold, O., Harun, F., Kostiukhin, A., Järveoja, M., Veske, M., Tamm, T., Voormansik, K., Olesk, A., 2021. KappaMask: AI-Based Cloudmask Processor for Sentinel-2. *Remote Sensing*, *13*(20), p.4100.

Dresp-Langley, B., Wandeto, J. M. & Nyongesa, H. O., 2018. Using the quantization error from Self-Organizing Map (SOM) output for fast detection of critical variations in image *ISTE OpenScience.*

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote sensing of Environment*, *120*, pp.25-36.

Enderle, D.I., Weih Jr, R.C., 2005. Integrating supervised and unsupervised classification methods to develop a more accurate land cover classification. *Journal of the Arkansas Academy of Science*, *59*(1), pp.65-73.

Fortuin, V., Hüser, M., Locatello, F., Strathmann, H., Rätsch, G., 2018. Som-vae: Interpretable discrete representation learning on time series. *arXiv preprint arXiv:1806.02199*.

Gao, B.C., 1996. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote sensing of environment*, *58*(3), pp.257-266.

Garcia Millan, V.E., Rankine, C., Sanchez-Azofeifa, G.A., 2020. Crop loss evaluation using digital surface models from unmanned aerial vehicles data. *Remote Sensing*, *12*(6), p.981.

Ghosh, P., Mandal, D., Bhattacharya, A., Nanda, M.K., Bera, S., 2018. Assessing crop monitoring potential of Sentinel-2 in a spatio-temporal scale. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *42*, pp.227-231.

Gobbo, S., Ghiraldini, A., Dramis, A., Dal Ferro, N., Morari, F., 2021. Estimation of Hail Damage Using Crop Models and Remote Sensing. *Remote Sensing*, *13*(14), p.2655.

Gómez, C., White, J.C., Wulder, M.A., 2011. Characterizing the state and processes of change in a dynamic forest environment using hierarchical spatio-temporal segmentation. *Remote Sensing of Environment*, *115*(7), pp.1665-1679.

Gu, Y., Brown, J.F., Verdin, J.P., Wardlow, B., 2007. A five-year analysis of MODIS NDVI and NDWI for grassland drought assessment over the central Great Plains of the United States. *Geophysical research letters*, *34*(6).

Gutman, G., Ignatov, A., 1998. The derivation of the green vegetation fraction from NOAA/AVHRR data for use in numerical weather prediction models. *International Journal of remote sensing*, *19*(8), pp.1533-1543.

Jaagus, J., Aasa, A., Aniskevich, S., Boincean, B., Bojariu, R., Briede, A., Danilovich, I., Castro, F.D., Dumitrescu, A., Labuda, M., Labudová, L., 2022. Long-term changes in drought indices in eastern and central Europe. *International Journal of Climatology*, *42*(1), pp.225-249.

Jackson, R.D., Huete, A.R., 1991. Interpreting vegetation indices. *Preventive veterinary medicine*, *11*(3-4), pp.185-200.

Jiapaer, G., Chen, X., Bao, A., 2011. A comparison of methods for estimating fractional vegetation cover in arid regions. *Agricultural and Forest Meteorology*, *151*(12), pp.1698-1710.

Kauth, R.J., Thomas, G.S., 1976. The tasselled cap--a graphic description of the spectral-temporal development of agricultural crops as seen by Landsat. In *LARS symposia* (p. 159).

Kohonen, T., 2013. Essentials of the self-organizing map. *Neural networks*, *37*, pp.52-65.

Komarov, A.S., Palenova, M.M., Smirnova, O.V., 2003. The concept of discrete description of plant ontogenesis and cellular automata models of plant populations. *Ecological Modelling*, *170*(2-3), pp.427-439.

Kulkarni, A., Chong, D., Batarseh, F.A., 2020. Foundations of data imbalance and solutions for a data democracy. In *data democracy* (pp. 83-106). Academic Press.

Kumar, P., Prasad, R., Gupta, D.K., Mishra, V.N., Vishwakarma, A.K., Yadav, V.P., Bala, R., Choudhary, A., Avtar, R., 2018. Estimation of winter wheat crop growth parameters using time series Sentinel-1A SAR data. *Geocarto international*, *33*(9), pp.942-956.

Liu, Q., Liu, G., Huang, C., 2018. Monitoring desertification processes in Mongolian Plateau using MODIS tasseled cap transformation and TGSI time series. *Journal of arid land*, *10*(1), pp.12-26.

McFeeters, S.K., 1996. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International journal of remote sensing*, *17*(7), pp.1425-1432.

Merzlyak, M.N., Gitelson, A.A., Chivkunova, O.B., Rakitin, V.Y.U., 1999. Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. *Physiol. Plant. 106,* 135–141.

Mostafiz, C., Chang, N.B., 2018. Tasseled cap transformation for assessing hurricane landfall impact on a coastal watershed. *International journal of applied earth observation and geoinformation*, *73*, pp.736-745.

Miljković, D., 2017. Brief review of self-organizing maps. In *2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)* (pp. 1061-1066). IEEE.

Páscoa, P., Gouveia, C.M., Trigo, R.M., 2018. The influence of drought on vegetation activity assessed by NDVI-MODIS on Estonia. In *EGU General Assembly Conference Abstracts* (p. 18068).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., 2011. "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, p.

Phiri, D., Simwanda, M., Salekin, S., Nyirenda, V.R., Murayama, Y., Ranagalage, M., 2020. Sentinel-2 data for land cover/use mapping: A review. *Remote Sensing*, *12*(14), p.2291.

Segarra, J., Buchaillot, M.L., Araus, J.L., Kefauver, S.C., 2020. Remote sensing for precision agriculture: Sentinel-2 improved features and applications. *Agronomy*, *10*(5), p.641.

Shalaginov, A., Franke, K., 2015. A New Method for an Optimal SOM Size Determination in Neuro-Fuzzy for the Digital Forensics Applications. *Advances in Computational Intelligence. IWANN 2015. Lecture Notes in Computer Science,* Volume 9095, p. 549–563.

Sharma, L.K., Bu, H., Denton, A., Franzen, D.W., 2015. Active-optical sensors using red NDVI compared to red edge NDVI for prediction of corn grain yield in North Dakota, USA. *Sensors*, *15*(11), pp.27832-27853.

Shi, T., Xu, H., 2019. Derivation of tasseled cap transformation coefficients for Sentinel-2 MSI at-sensor reflectance data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *12*(10), pp.4038-4048.

Skellern, M.P., Cook, S.M., 2018. The potential of crop management practices to reduce pollen beetle damage in oilseed rape. *Arthropod-Plant Interactions*, *12*(6), pp.867-879.

Sobrino, J.A., Raissouni, N., 2000. Toward remote sensing methods for land cover dynamic monitoring: Application to Morocco. *International journal of remote sensing*, *21*(2), pp.353-366.

Sosa, L., Justel, A., Molina, Í., 2021. Detection of Crop Hail Damage with a Machine Learning Algorithm Using Time Series of Remote Sensing Data. *Agronomy*, *11*(10), p.2078.

Sulik, J.J., Long, D.S., 2016. Spectral considerations for modeling yield of canola. *Remote Sensing of Environment*, *184*, pp.161-174.

Tamm, T., Zalite, K., Voormansik, K., Talgre, L., 2016. Relating Sentinel-1 interferometric coherence to mowing events on grasslands. *Remote Sensing*, *8*(10), p.802

Tang, B.H., Shao, K., Li, Z.L., Wu, H., Tang, R., 2015. An improved NDVI-based threshold method for estimating land surface emissivity using MODIS satellite data. *International Journal of Remote Sensing*, *36*(19-20), pp.4864-4878.

Vettigli, G., n.d. *Minisom: Minimalistic and Numpy-Based Implementation of the Self Organizing Map.* Available at: https://github.com/JustGlowing/minisom [Accessed 08 January 2022].

Wang, S., Rao, Y., Chen, J., Liu, L., Wang, W., 2021. Adopting "Difference-in-Differences" Method to Monitor Crop Response to Agrometeorological Hazards with Satellite Data: A Case Study of Dry-Hot Wind. *Remote Sensing*, *13*(3), p.482.

Xue, J., Su, B., 2017. Significant remote sensing vegetation indices: a review of developments and applications. J Sens 2017: 1–17.

Zagajewski, B., Tømmervik, H., Bjerke, J.W., Raczko, E., Bochenek, Z., Kłos, A., Jarocińska, A., Lavender, S., Ziółkowski, D., 2017. Intraspecific differences in spectral reflectance curves as indicators of reduced vitality in high-arctic plants. *Remote Sensing*, *9*(12), p.1289.

Zhang, X., Schaaf, C.B., Friedl, M.A., Strahler, A.H., Gao, F., Hodges, J.C., 2002. MODIS tasseled cap transformation and its utility. In *IEEE International Geoscience and Remote Sensing Symposium* (Vol. 2, pp. 1063-1065). IEEE.

Zhang, X., Liao, C., Li, J., Sun, Q., 2013. Fractional vegetation cover estimation in arid and semi-arid environments using HJ-1 satellite hyperspectral data. *International Journal of Applied Earth Observation and Geoinformation*, *21*, pp.506-512.

Zhang, T., Vail, S., Duddu, H.S., Parkin, I.A., Guo, X., Johnson, E.N., Shirtliffe, S.J., 2021. Phenotyping Flowering in Canola (Brassica napus L.) and Estimating Seed Yield Using an Unmanned Aerial Vehicle-Based Imagery. *Frontiers in Plant Science*, *12*, p.1178.

Zhang, Z., Liu, M., Liu, X., Zhou, G., 2018. A new vegetation index based on multitemporal Sentinel-2 images for discriminating heavy metal stress levels in rice. *Sensors*, *18*(7), p.2172.

Copernicus Open Access Hub, News., 2022. Copernicus Sentinel-1B anomaly (6th update), https://scihub.copernicus.eu/news/News01026 (last viewed 22.04.2022)

Creodias, 2022. Knowledgebase, EO Finder, https://finder.creodias.eu/ (last viewed 08.04.2022)

Estonian Environment Agency, 2022. Weather database, http://www.ilmateenistus.ee/kliima/kliimanormid/ohutemperatuur/?lang=en (last viewed 08.04.2022)

Estonian Land Board Orthophoto, 2021. Fotoladu - aerofotode digitaalarhiiv, https://fotoladu.maaamet.ee/ (last viewed 08.04.2022)

Sentinel Online, 2022. Sentinel-2 data formats, https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi/data-formats (last viewed 08.04.2022)

Sinergise Ltd., 2022. Sentinel Hub, EO Browser, https://apps.sentinel-hub.com/eo-browser (last viewed 08.04.2022)

Non-exclusive licence to reproduce thesis and make thesis public

I, Anton Kostiukhin,

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright,

   "Clustering analysis of spatiotemporal Sentinel-2 data of agricultural parcels in Estonia for damaged crop delineation",

   supervised by Ph.D. Alexander Kmoch, Ph.D. Tanel Tamm, Ph.D. Indrek Sünter.

2. I grant the University of Tartu a permit to make the work specified in p. 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 3.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.

3. I am aware of the fact that the author retains the rights specified in p. 1 and 2.

4. I certify that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Anton Kostiukhin
29/05/2022