GIMA

Geographical Information Management and Applications

Determining yield-limiting factors using optical satellite data

Author: Angelina Savchuk
Supervisor: Dr. Jan Clevers
Responsible professor: Dr. Ron van Lammeren

August 15, 2019

Universiteit Utrecht

TU Delft

UNIVERSITY OF TWENTE.

WAGENINGEN UR

Abstract

The constantly growing population and food demand requires us to use agricultural land as efficiently as possible. The high costs of fertilisers and plant protection products together with their impact on the environment motivate farmers to seek and implement new technologies for maximising the yield while reducing the costs of production. In order to achieve that, farmers need to address Yield-Limiting Factors (YLFs). This is where remote sensing comes in handy - instead of using ground-based techniques for the detection and monitoring of yield-limiting factors that are costly and time-consuming, satellite imagery is offering a cheap and fast alternative.

In this research, the focus is put on yield-limiting factors for corn crops. The test fields are located in the central part of Ukraine, namely in the Cherkasy province. Two cornfields were selected for analysis. Whereas field 6 was used as the main research field, field 10 was used for validation purposes to compare the results obtained from both fields. The size of the fields is 227 ha and 181 ha respectively, making this research important for large-scale farming. Eight satellite images were selected, one for every month in the growing season (March to October 2017). Those images were visually and statistically analysed for correlation with the final yield and several yield-limiting factors.

Near Infrared (NIR), Normalized Difference Vegetation Index (NDVI) and Atmospherically Resistant Vegetation Index (ARVI) were used to test the possibility of the detection of yield variability using remote sensing data. NIR was determined to be the most correlated with the final yield (with the maximum R^2 of 0.72). NDVI was the second-best tool for the detection of yield variability with the maximum R^2 reaching 0.698. ARVI did not show any particularly outstanding results. The average amount of explained variation in NIR is around 28%; it is then followed by NDVI of 19% and ARVI of 14%.

The results show that September is the best moment for the detection of yield variability, the highest correlation can be found at that time (R^2 = 0.72). The beginning of the growing season can also produce a high correlation considering that it is only bare soil ($R^2 \approx 22\%$ on average).

Organic Matter (OM) content was deemed as the YLF that has the greatest potential for detection using satellite imagery in field 6 (R^2 of 0.368). The reason for that is the big variations in OM content in field 6. Soil Potential of Hydrogen (pH) and Electrical Conductivity (EC) showed smaller correlation (R^2 = 0.038 and R^2 = 0.051, respectively), while elevation needed a different approach before the correlation could be found (R^2 = 0.312).

During the validation of the results, it was found that a lot of them are field dependent and cannot be reproduced on other fields/farms. Especially when it comes to yield-limiting factors, no consistent results were found; hence there is no specific method which is the most optimal for yield-limiting factors detection. The same applies to the best moment for the detection of yield-limiting factors. For different yield-limiting factors different VI/spectral bands showed the highest results, and it also varied per field. The same goes for the temporal aspects of detection. However, both fields showed that September was found to be the month of the highest correlation between yield variability and optical satellite imagery.

Acknowledgements

Throughout the writing of this thesis, I have received a great deal of support and assistance. First of all, I would like to express my deepest gratitude to Jan Clevers for supervising and guiding me in the writing of this thesis, his always-to-the-point comments and suggestions, expertise, quick reactions, patience and support. I honestly believe that this thesis would not have been written/finished if not for Jan's guidance.

The data for the research was kindly provided by Neo B.V. and the Kischenzi farm, for which I would like to thank them.

I would like to thank Corné van der Sande for allowing me to have a brand new look on the use of remote sensing in agriculture, for all his support and expertise and for being my contact point between NEO B.V. and the Kischenzi farm.

A big thank you to NEO B.V. and Cornelis Valk for giving me the opportunity to work on this topic.

My warmest regards to Kees Huizinga and his family for allowing me to work on such an interesting topic, for letting me stay at his farm during some field research and for making this stay very comfortable and enjoyable. To Henrik Laursen, for making time in his busy agenda to answer my questions and for providing insights only he can know. To Sasha, for showing me around the farm and for helping me figure out the new world of agriculture.

A special thank you to Maurits, for his unconditional support, spelling checks and motivational speeches. I would like to also thank my fellow students for answering my questions. A special thank you to Aaron, who was always asking me about the progress of this thesis and his explanations of terms or topics that were new to me. Also, to Vivian, who volunteered to review my thesis.

Last but not least, I would like to thank my parents for making my work on this thesis their priority and for creating the most stimulating environment and conditions for that.

This topic was relatively new to me, however, I honestly enjoyed researching it and learned a lot of new things on the way. I hope that its results will be useful and innovative and that they will provide some new insights that can be used by the Kischenzi farm in their future managerial practices.

Sincerely,
Angelina Savchuk
August 13, 2019

Contents

List of Figures

List of Tables

Glossary

ARVI Atmospherically Resistant Vegetation Index.

CEC Cation-Exchange Capacity.

EC Electrical Conductivity.

GNDVI Green Normalized Difference Vegetation Index.

IR Infrared.

LAI Leaf Area Index.

NDVI Normalized Difference Vegetation Index.

NIR Near Infrared.

OM Organic Matter.

pH Potential of Hydrogen.

SRTM Shuttle Radar Topography Mission.

Vis-NIR Visible and Near Infrared.

YLF Yield-Limiting Factor.

1 Introduction

1.1 Context

Identifying yield-limiting factors Yield-Limiting Factors (YLFs) that are influencing the harvest is what farmers, together with scientists, have been doing for a long time. Knowing in time which factors reduce the yield, would give the farmer an opportunity to adapt his managerial behaviour and take those factors into consideration for a better yield. Many yield-limiting factors are well known and investigated, for example, soil properties, weather, pests, diseases, field management (Paz et al., 2001), weed (Subedi and Ma, 2009), topography (Changere and Lal, 1997), row spacing and management practices (Mueller et al., 2012).

While multiple research projects were done to determine yield-limiting factors, not many of them made use of remote sensing data. However, this source of data can give a whole new angle of crop and yield observation, since it *"can potentially provide observations for every single field in a region for every single growing season"* (Lobell, 2013). Today's developments make it possible to use remote sensing techniques to monitor even within-field variations.

The monitoring of small areas dedicated to agriculture using remote sensing techniques is called precision agriculture. Precision agriculture (sometimes also called precision farming), as defined by Seelan et al. (2003), is *"the use of modern technologies to improve crop yield, provide information to enable better in-field management decisions, reduce chemical and fertilizer costs through more efficient application, permit more accurate farm records, increase profit margin and reduce pollution."*

Remote sensing techniques have great advantages over traditional (ground) measurements and survey techniques when it comes to the surveillance of the crops (Clevers, 1997). First of all, it can collect information from large areas much faster compared to traditional methods. Second of all, the revisiting time could be quite short, especially if the data is collected from multiple satellites. To increase the list of advantages even more, it is important to mention that the costs for obtaining and processing the satellite data are constantly decreasing, while the quality of data is increasing (Chuvieco, 2016).

Satellite data has been named as a great potential for precision agriculture and yield-gaps evaluation according to Lobell (2013). The same author argues that the high-resolution satellite data is becoming more and more available and affordable, which offers farmers great opportunities for using it in the determination and supervision of yield-limiting factors instead of traditional techniques.

Given the aforementioned advantages, it is wise to investigate how remote sensing techniques can be applied more often in the field of precision agriculture. This master thesis will focus on yield-limiting factors that can be determined and supervised using satellite data.

1.2 Problem statement

It is relatively easy to prove the interconnection between the yield-limiting factors and the yield, but it is time-consuming due to the necessity of collecting the information from the ground, its processing and analysis. Once the limiting factors are known, they can be quantified and monitored. Monitoring the yield using ground survey is a time consuming task, especially if it concerns large areas. Here is where the remote sensing techniques might come in handy. By using remote sensing techniques, a farmer can have constant supervision on the influence of yield-limiting factors on his crops without a need to collect ground data, which is especially useful for large-scale farming.

The research was initiated by the need of one large-scale farm to obtain more information and therefore knowledge from their satellite data. The farm, named ”Kischenzi”, is located in Ukraine and possesses over 16 000 ha of fields of different crops. The management of the farm concluded that it could save large amounts of money by properly addressing yield-limiting factors - limited yield on such a large scale makes the losses remarkable. The company decided to make use of satellite data to improve their management of yield-limiting factors. Satellite data is probably the only possible way of doing that due to the enormous size of the farm. The farm has been provided with the satellite images by a Dutch company called NEO B.V. While the data is provided, there was a lack of knowledge within the farm management on how yield-limiting factors can be monitored with satellite data. This research aims at covering this lack of knowledge and investigating which factors can be monitored with satellite data and providing the technique on how this could be achieved.

1.3 Research objectives

The ultimate goal of this research is to investigate how remote sensing data can contribute to better crop productivity. Better crop productivity can be ensured when factors that limit its capacity are defined. Once the limiting factors are defined, it is possible to investigate which of them can be observed with remote sensing data. This research will focus on topography and soil properties as possible yield-limiting factors. The anticipated results of the research will contain information about which yield-limiting factors can be monitored with remote sensing and how and to what extent remote sensing can replace traditional techniques of yield monitoring.

1.4 Research questions

The main research question is as follows :

To what extent can we relate yield-limiting factors to within-field variation as expressed by vegetation indices derived from optical data?

To help answer this research question, several subquestions will be used:

1. To what extent can remote sensing be used for detection of spatial yield variability?
2. Which vegetation indices and/or spectral bands are the most optimal for detecting yield variability?
3. Which time frame is most suitable for identifying yield variability?
4. Which factors can be identified as yield-limiting factors and show potential for detection in remote sensing data?

1.5 Societal relevance

The anticipated results of this research are expected to help farmers match the actual yield as close as possible with the potential yield. This will increase the crop productivity and maximum economic return, from which both farmer and society will benefit (Loomis and Williams, 1963).

Also, if the conclusions from the results of this research are derived correctly, it might help the farmer in reducing the amounts of inputs of plant protection products, such as herbicides (Kennedy et al., 2012), insecticides and fungicides (Marty et al., 1993) and the input of fertilisers.

This all will result in maximising the benefits for the farmer and at the same time reducing the negative effect on the environment.

1.6 Scientific relevance

A study conducted by Prasad et al. (2006) suggests the use of high-resolution satellite data to improve the detection of yield-limiting factors. While multiple types of research have been done to determine the yield-limiting factors with ground survey techniques, only a few of them made use of remote sensing techniques for determining if those factors can be seen or detected with optical data (Oliver et al. (2013), Labus et al. (2002)). This research will attempt to fill the gap of scientific knowledge in the field of remote sensing and precision agriculture.

1.7 Constraints and limits

For this research, a few testing fields will be selected to perform the analysis on. For the reason of consistency of the analysis, those fields will be of a specific crop type that will be defined in a later stage of the research and described in section 3.5. The reason for that is that different crops have different growing peculiarities and it would obstruct the research if those have to be taken into account.

Also, even if the data exists for several growing seasons, only the data from the growing season of 2017 will be used for the analysis.

This research will only focus on the following factors: soil properties and topography. In case new data sets will become available, new factors can be included for analysis.

2 Theoretical framework

What is a yield variation, what does a yield-limiting factor mean, what can vegetation indices tell us - those are questions that can be answered with a literature review. This chapter aims at providing background for subquestions 1, 2 and 3. The chapter commences with more general theoretical concepts like precision farming, yield-limiting factors, yield variation and it is followed by an explanation of multiple vegetation indices and how they can be used in remote sensing.

2.1 Precision farming and remote sensing

"Remote sensing, with the high variety of spectral ranges and the fine spatial and temporal resolution currently available, is a tool of great value for various applications in agriculture" (Baghdadi and Zribi, 2016).

Courault et al. (2017) have described a few angles of opportunities that remote sensing gives to crop monitoring:

- it is able to monitor large areas of land and give information on numerous practices and land uses that are engendered by agriculture. This information is valuable, it can be used for generating agricultural statistics which in their turn can be used in agri-environmental monitoring and modelling.
- it offers the continuous monitoring of crops, their production potential, their irrigation needs by remotely monitoring soil and vegetation properties. Courault et al. (2017) mention that this is important for forecasting purposes, yield security and efficient resource management.
- its ability to assess the influence of agriculture to net emissions of CO_2 and other gases that have an impact on the environment.
- due to constantly improving spatial and temporal resolution of remote sensing data, it is used more and more often for decision support for farming activities on intra-field variability - what is best known as precision farming. This is done to ensure better productivity and overall efficiency by an intelligent use of inputs, which also results in a better impact on the environment.

Remote sensing gave, without a doubt, a whole new view to precision farming. Where only a few decades ago it was not yet possible to even imagine the use of remote sensing techniques in the daily managerial field practices, today it is used not only on large scales, but for applications that are meant to monitor the within-field variations.

2.2 Remote sensing and soil mapping

Another example of the use of remote sensing in agriculture is mapping soil types and properties. Many researchers, such as for example Forkuor et al. (2017) and Ge et al. (2011), claim remote sensing as the most precise and efficient technique for mapping soil types and its properties. It has been proved that a great amount of soil properties that are important for agriculture can be successfully mapped and monitored with remote sensing, namely *"including textures,* *organic and inorganic carbon content, macro- and micro-nutrients, moisture content, cation exchange capacity, Electrical Conductivity (EC), Potential of Hydrogen (pH), and iron"*(Ge et al., 2011) .

The mapping of soil surface properties is usually done with Visible and Near Infrared (Vis-NIR) (400–2500 nm) spectral bands (Gomez and Lagacherie, 2017).That can be done with high spatial resolution even on large areas.

However, there are also some limitations in extracting soil properties from remotely sensed data.

- Due to the low surface penetration of electromagnetic waves in the Vis-NIR only upper surface properties can be mapped (Gomez and Lagacherie, 2017).
- Different components affect the soil spectrum what results in wrong interpretations. The most important constituents that influence the spectral signature of the soil are minerals (clay, ferric and carbonate), Organic Matter (OM) and water (Baghdadi and Zribi, 2016).
- *"Some primary soil properties, such as pH and Cation-Exchange Capacity (CEC), have no particular spectral characteristics and, therefore, cannot be determined by direct spectral analysis "*(Baghdadi and Zribi, 2016).

Despite the popularity of remote sensing techniques for mapping the soil, there is not one particularly good approach that will work for all types of soil, on all properties and in all study areas. While choosing a method, one should evaluate carefully its advantages and disadvantages. However, since consistently more research and innovation are aiming at replacing traditional ground techniques with remote sensing, in the near future it is very likely that remote sensing will become the most used tool for mapping soil types and properties (López-Granados et al., 2005).

2.3 Yield variability

In the past, a field was wrongly assumed to be homogeneous and all cultivation practices, levels of input and other managerial techniques were applied at fixed rate or proportion over the whole field (Joernsgaard and Halmoe, 2003). The yield variability is not a new discovery, however, until recently there weren't any techniques that would allow to take the yield variability into account while planning managerial practices for the future season. Precision farming is this new technique that was built around the knowledge that yield can vary on the inter or intra level (Zarco-Tejada et al., 2014).

Yield variability is a consequence of influence of yield-limiting factors on the crop (Carter and Dean, 1960). As mentioned before, a field cannot be homogeneous, and yield-limiting factors interact with field properties differently what results in differences in yield. Analysis of yield variability is an important issue in agricultural studies (Kravchenko and Bullock, 2000) and is of main concern to farmers, who want to maximise the profit, and society, which is interested in making the use of soil as efficient as possible.

Due to its importance, the question on how we can measure the yield variability arises. Stafford et al. (1996) suggest that measuring the yield variability using mapping techniques and spatial sciences will become an important tool for achieving this. Since even small changes in location can produce big differences in yield, it is important to map the yield with a suitable spatial resolution to be able to extract useful information from those maps and convert it to strategic decisions based on the extracted information afterwards. Stafford et al. (1996) state that *"a number of systems have been developed that determine the grain flow into the combine by indirect methods such as measurement of engine speed or the torque on the grain tank filling auger"*. Simply said, the combine is measuring how many grains per unit area is collected at a known position in the field. That is one of the ways to measure yield and this technique was used to collect the yield data for this research. Appendix A is an example of what a yield-variability map can look like.

Usually, yield variability is not explained by a single limiting factor (Melakeberhan, 2002). It is more likely a combination of multiple factors that do or do not interact with each other. In the next section an overview of yield-limiting factors is made and their way(s) of influencing the yield variability are explained.

2.4 Yield-limiting factors

"Yield potential is the yield of a crop cultivar when grown in an environment to which it is adapted, with non-limiting water and nutrient supplies, and with pests, weeds, and diseases effectively controlled" (Edreira et al., 2017). Simply said, it is the maximum yield a farmer can expect if he succeeds in providing the crop with the most convenient conditions for its growth. As derived from the definition, yield-limiting factors play an important role in matching the real yield as close as possible to the potential yield. This chapter will provide the reader with a theoretical underpinning of yield-limiting factors and will go into details about every factor that is taken into account in this research.

Crop yields are influenced by a multitude of factors (Nielsen, 1998) (see Figure 1). Some of them influence the yield directly, some only influence while interacting with other factors. What makes studying yield-limiting factors even more difficult is that some factors appear every year, while others only occasionally.

Different theories and opinions exist about yield-limiting factors and their way and extent to which they influence the yield. Also, many research projects have been done to determine what are the most important yield-limiting factors and again here opinions of the scientists differ. While some claim water to be the most responsible for yield factor (Clark et al., 1999), others name climate (Downing, 2013) or topography (Kravchenko and Bullock, 2000) as the most important yield-limiting factors. Nevertheless, there are two statements on which most of the authors agree unanimously:

- Yield-limiting factors differ per crop type. (f.e. where one soil property can be a limiting factor for crop A, it can be a stimulating factor for crop B.) - *"Different crops prefer different soils* (Aires, 2016)".
- Weather interacts with most of them (Nielsen, 1998).

A study made by Tittonell and Giller (2013) points out the difference between yield defining, yield-limiting and yield-reducing factors. Often those are misused and even if they are interrelated, there is a significant difference between them. Below, an overview can be found of main categories and factors that belong to these categories (Figure 1). For this study, there will be no distinction made between yield-defining, limiting or reducing factors and all of them will be considered as limiting.

2.4.1 Soil properties

It has been proven that yield strongly depends on soil properties (Dampney and Moore, 1999). Many of these influence the yield, but for this research only a few soil properties will be taken

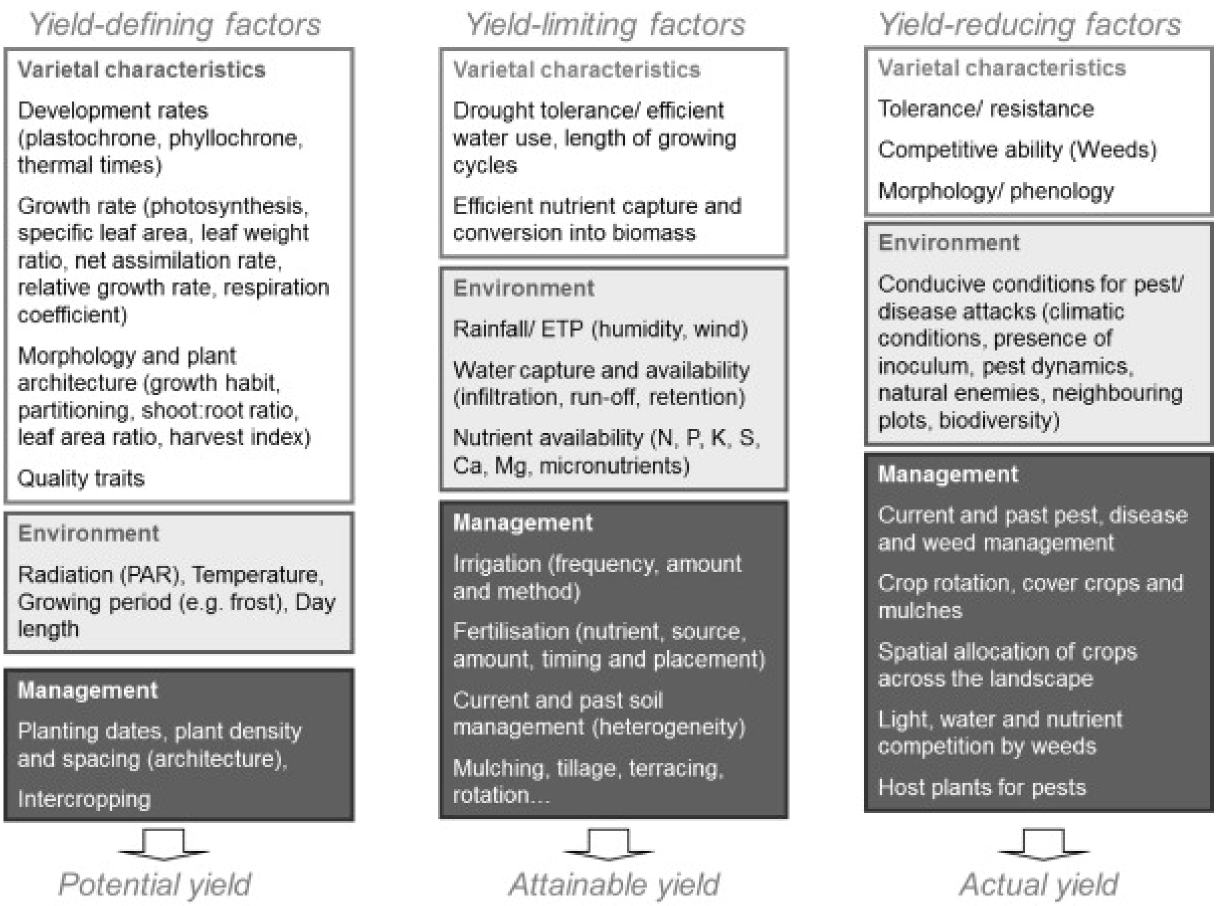


Figure 1: A representation of yield-defining, yield-limiting and yield-reducing factors according to Tittonell and Giller (2013).

into account. It should be noted that soil properties might not always be a limiting factor on their own (Joernsgaard and Halmoe, 2003). By interacting with weather or farming practices they can influence the yield differently from year to year, which makes the analysis even more complex.

In the following sections soil properties such as texture pH, electrical conductivity and organic matter content will be studied in more detail.

2.4.1.1 Soil pH

Soil pH is an indication of the acidity or alkalinity of a soil. It is measured in pH units. Soil pH is considered as a yield-limiting factor when the level of pH is too low or too high for what the specific crop can tolerate. It is believed that the most convenient pH for plants is when the pH is neutral or close to it - from 6.6 to 7.3 according to Table 1 (Stites, 2011).

However, it is wrong to assume that a pH of 7.0 would be ideal for any plant. Some crops can only grow in acid soils and a pH of 7.0 would be a limiting factor for its growth. For example, blueberries and potatoes have a pH tolerance of 4.5-5.0 and 4.5-6.0 accordingly,

Table 1: Classification of soil pH according to NRCS (1998)

Category	Range
Extremely acid	3.5 – 4.4
Very strongly acid	4.5 – 5.0
Strongly acid	5.1 – 5.5
Moderately acid	5.6 – 6.0
Slightly acid	6.1 – 6.5
Neutral	6.6 – 7.3
Slightly alkaline	7.4 – 7.8
Moderately alkaline	7.9 – 8.4
Strongly alkaline	8.5 – 9.0

which makes them crops that require soils that vary between extremely acid to moderately acid according to Table 1 (Stites, 2011).

Therefore, it can be concluded that an optimal soil pH is unique per crop and it can be both a limiting and a stimulation factor depending on a crop type. The pH tolerance of corn crops is discussed in section 3.5.

2.4.1.2 Soil EC

Electrical conductivity is *"the ability of a material to transmit (conduct) an electrical current and is commonly expressed in units of milliSiemens per meter (mS/m)"* (Grisso et al., 2005a).

Soil EC does not influence plant growth directly, but it is often used as an indirect measurement of the available nutrients for plants (Grisso et al., 2005b). According to Grisso et al. (2005a), it is possible to determine other soil properties using EC such as:

- **Water-holding capacity/drainage.** There is a clear textural difference between dry and wet soil and this can be seen with EC.
- **Cation exchange capacity** is influenced by the amount of clay and organic matter and CEC can be related to EC through its clay percentage.
- **Porosity** is directly related to EC - the bigger it is, more freely it conducts electricity.
- **Salinity** can be identified with EC through the amount of salt concentration. Usually EC is used as a measure of soil salinity (USDA, 2011).
- **Temperature** variation provokes variations in soil EC.

2.4.1.3 OM content Soil organic matter serves as a reservoir of nutrients and water in the soil (Landsberg and Gower, 1997). Soil OM consists of plant and animal residues at various stages of decomposition and usually only 5% of it mineralizes per year. The percentage can increase if temperature, oxygen and water availability are favourable for decomposition.

Soil OM is an important yield-limiting factor, because it contains nutrient supply for the plant and if this supply is broken, plant growth will be limited (Hijbeek et al., 2017).

2.4.2 Topography

Topography is another important and well known yield-limiting factor. Kravchenko and Bullock (2000) identified two of most common ways of influence of topography on the yield: redistribution and availability. Water has an impact on how the organic matter, soil particles and nutrition are redistributed over a sloping surface, what results in changes in physical and chemical attributes of the soil. Also, topography can be an extra limiting factor in combination with other factors, as for example water. The amount of the water that is present for the plant is decisive for its growth and topography can play an important role in water distribution, therefore being a factor that contributes to the yield. There are several topography properties that can act as yield-limiting factors, namely slope, curvature, landscape position and elevation (Kravchenko and Bullock, 2000).

2.5 Vegetation indices

Vegetation indices are used to highlight the differences in green vegetation using mathematical equations and transformations. It is a mathematical combination or transformation of two or more spectral bands that is used to accentuate the spectral properties of green plants and make them look distinct (Dutta, 2017).

Green plants have distinct spectral reflectance patterns in the Vis-NIR spectral bands, with a very low reflectance in blue and red and a slightly higher in green. This is the reason why leaves appear green. However, the spectral reflectance of green leaves in Near Infrared (NIR) is very high, what allows us to distinguish plants from soil and water (see Figure 2)(Parida, 2006). The big contrast of absorbtion and reflectance in the Vis-NIR spectral bands can be used for different quantitative indices of vegetation conditions. These mathematical quantitative combimations and transformations are called vegetations indices (Panda et al., 2010).

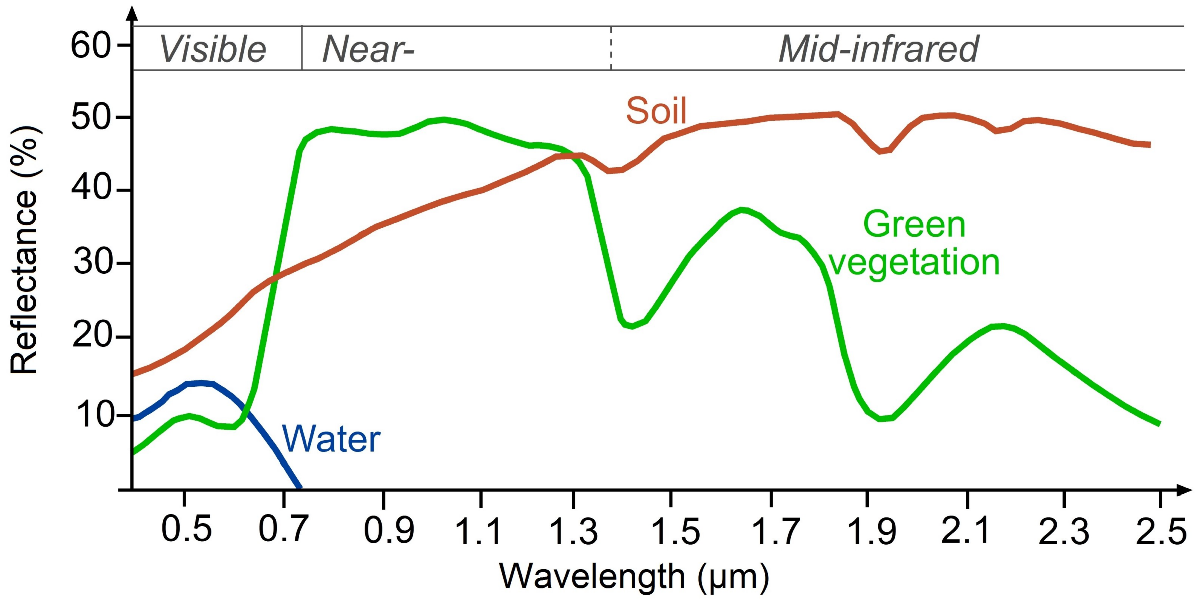


Figure 2: Reflectance of water, soil and vegetation at different wavelengths.

Source : http://www.seos-project.eu/modules/remotesensing/remotesensing-c01-p05.html

There are different types of vegetation indices, all of them using different bands to give an estimate on Leaf Area Index (LAI), percentage of green cover, chlorophyll content, biomass and plant moisture. However, for this research only the following indices will be used - Normalized Difference Vegetation Index and Atmospherically Resistant Vegetation Index (ARVI), as well as Near Infrared band. They will be discussed in more detail in the following sections.

2.5.1 NDVI

Normalized Difference Vegetation Index is, without doubt, the most widely used vegetation index according to Gurgel and Ferreira (2003). It is used to calibrate crop yield indicators, to supervise crop development, to estimate yield and detect disease and weak spots. NDVI is calculated using the following formula:

$$NDVI = (NIR - Red)/(NIR + Red)$$

According to Prasad et al. (2006), by analysing NDVI values over time it is possible to supervise vegetation growth and health status, which is crucial for this research. Since NDVI represents vegetation greenness, it shows us how healthy the vegetation is at given time and space. With this information we can derive weak spots that might later transform into low yield. NDVI is also a helpful index for analysing plant's phenology according to Tzima (2017).

There are great advantages of using NDVI for agricultural purposes - it is reliable, robust and widely available (spectral bands used for the computations can be found on all optical satellites) (Baghdadi and Zribi, 2016). In the same time, it should be noted that it also has its drawbacks - such as weak correction of atmospheric effects, ratio-based formula and sensibility to canopy background brightness (Huete et al., 2002), which can greatly impact the results.

2.5.2 ARVI

Atmospherically Resistant Vegetation Index was developed to improve the performance of NDVI, since the latest is very sensitive to atmospheric influences. Thus, the main characteristic of ARVI index (and its difference from NDVI index) is its ability to limit the influence of atmospheric effects (Liu et al., 2004). This is done by using the blue band in conducting atmospheric corrections on the red band - in comparison to the red band, the blue one is more easily scattered by the atmosphere particles. ARVI is four times less sensitive to atmospheric scattering in comparison to NDVI (Huete et al., 1997). ARVI is calculated with the following equation:

$$ARVI = (NIR – (2 * Red) + Blue)/(NIR + (2 * Red) + Blue)$$

2.6 Vegetation growth stages

Remote sensing is used to supervise crops during different periods of their phenological stages. A bunch of factors define the plant's growth stage and it allows us to monitor its development from above. Current techniques make it possible to monitor the tiniest development in the vegetation growth stages - from the bud-break moment to its senescence (see Figure 3). This is achievable with supervision of various levels of vegetation's "greenness" using various vegetation indices, that describe the health status of the crop (Baghdadi and Zribi, 2016).

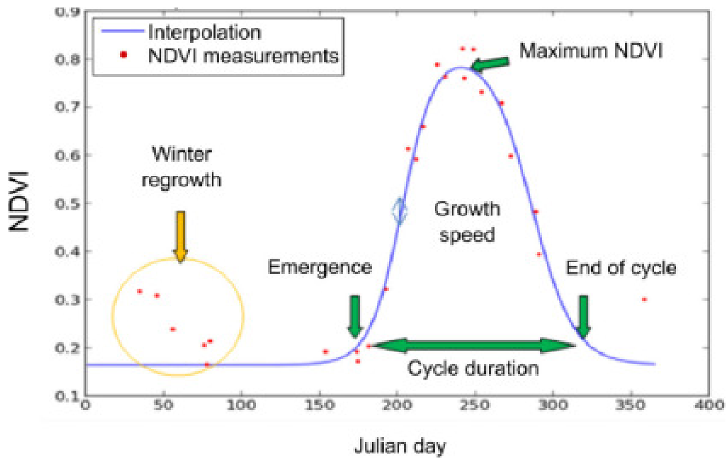


Figure 3: Example of a NDVI temporal profile for a summer crop

Naturally, different crops have different vegetation growth stages. Additionally, the same crop can have great variation in vegetation growth stages in time and space even within the same field. This is explained by differences in primary conditions, crop management and interaction with climate.

3 Materials and methods

3.1 Study area

The study area of the research is located in central Ukraine, namely in the Cherkasy Oblast (province) (see Figure 4). The region lies within the Ukrainian forest steppe ecotone. This geographical position is more than favourable for agriculture, which is the main economic activity of this region.

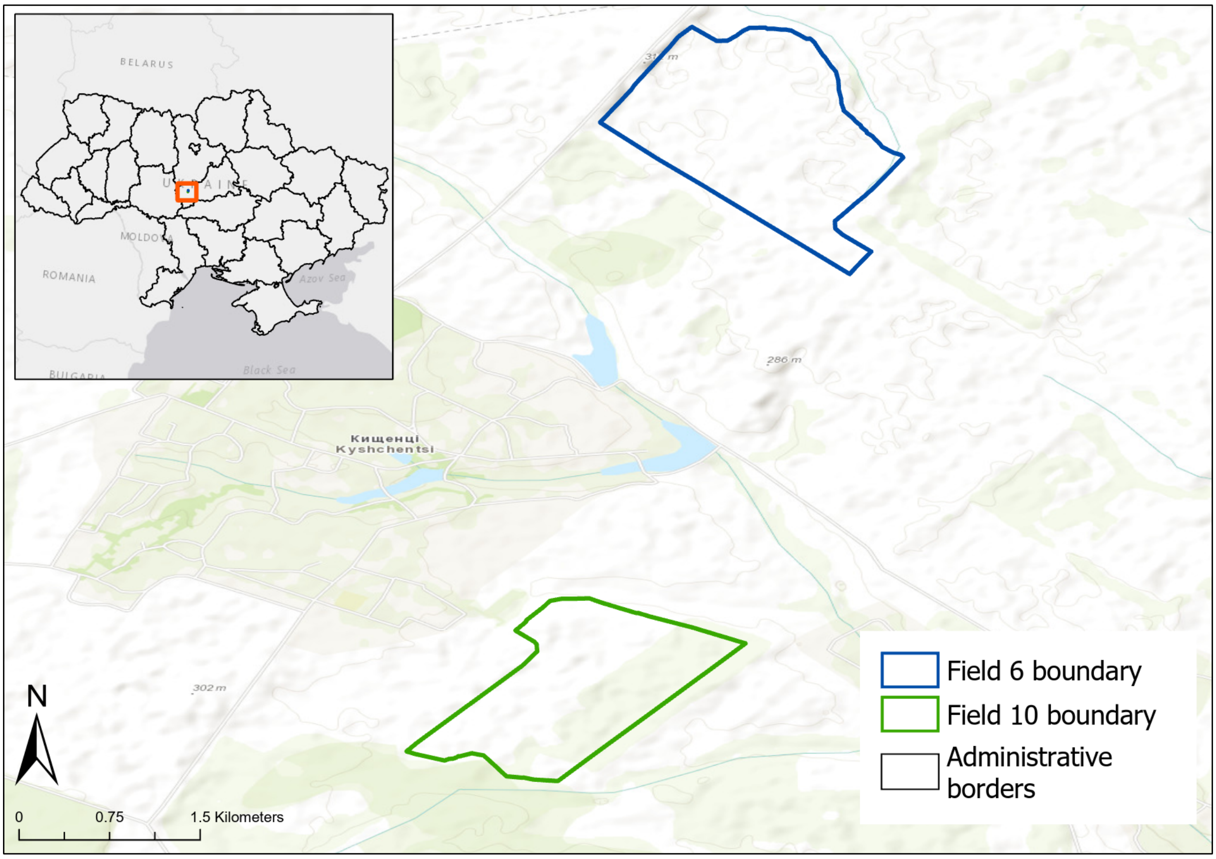


Figure 4: Study area

The choice of the area was due to the presence of an agricultural company - a farm called Kischenzi - in this region that is oriented towards precision agriculture, hence the availability of large amounts of data for the research. The company has provided the indispensable validation data for the research that was not available via any other sources. Also, the region is known for its high agriculture productivity due to the presence of chernozems - an extremely fertile soil that can produce high agricultural yields (Demydenko and Velychko, 2015).

The overall study area is around 16 000 hectares. However, the research will not concentrate on all fields but will instead focus on only a few, namely field 6 and field 10. The most accurate and complete data sets exists for those fields. It is important to note that the size of those fields is impressive - field 6 is around 227 hectares large, field 10 is 181 hectares. Together, they provide around 4 km^2 area for the research.

3.2 Data

The research is focusing on the use of optical satellite data for the determination of yield-limiting factors. The executive methodology is entirely based on the data analysis, therefore data is the most important component of the research. Below, an overview of the available data is provided.

3.2.1 Satellite data

Satellite data is the core of this project. Neo B.V. - a Dutch company specialising in earth observation - is supplying the research with the processed optical data. Optical data is obtained from satellites Landsat 8 and Sentinel-2A and 2B. The spatial resolution of Landsat 8 is 15m, whereas both Sentinel satellites have a spatial resolution of 10m. The data contains Red, Green, Blue, NIR spectral bands and vegetation indices such as NDVI, NDI7 and ARVI (see Table 2).

Table 2: Vegetation indices

Index	Abbreviation	Formula	Source
NDVI	Normalized Difference Vegetation Index	NDVI = $\frac{NIR - VIS}{NIR + VIS}$	Wang et al. (2007)
NDI7	Normalized Difference Index 7	NDI7 = $\frac{B5 - B7}{B5 + B7*}$ * B - the band indication in Landsat TM	Jin et al. (2015)
ARVI	Atmospherically Resistant Vegetation Index	ARVI = $\frac{IR - rb}{IR + rb}$ **rb* is a combination of the reflectances in the Blue (b) and Red (r) bands	Kaufman and Tanre (1992)

Neo B.V. has developed an online portal, called Cropmonitor, where the satellite data can be seen. This portal is multifunctional - it is possible to view data from different indices, dates and levels (field or segments). However, the online portal is created for viewing data only - it is not possible to extract data, so the company has provided original files for the analysis.

3.2.2 Harvest data

Harvest data is collected by the farm during harvesting with a John Deere S690 combine and then processed with the software from The Climate Corporation. An example of harvest data can be seen in the annex A. This data is stored in the portal with other information obtained from the John Deere tractor. This data set contains other variables that are used throughout the research such as planting date, harvest date, population and hybrid.

3.2.3 Soil properties data

Soil properties data were collected using the software and hardware from Veris Technologies and not by using the technique described in section 2.2. The soil properties database contains information about pH, electrical conductivity, organic matter, cation exchange capacity, slope and curve. An example of soil properties data can be found in annex B.

3.3 Data quality

As was already mentioned before, data constitutes the base of this research. However, the quality of the data must be investigated prior to its use. Since the data is provided by different parties, it must be first evaluated to see if it is appropriate and suitable for the research. Issues such as calibration, spatial and temporal scaling must be evoked.

Mostly, the harvest and soil properties data have to be evaluated since they were collected with machines by people who might not have the necessary geographical education and experience.

Data quality assessment will be performed according to six main data criteria of evaluation according to Batini et al. (2009), such as :

- **Completeness**. It will evaluate if all data sets and data items are recorded. F.e. is there data for the whole field or there are parts of the field that are not covered (missing)?
- **Uniqueness** ensures that there are no double entries.
- **Timeliness** evaluates to which extent data represents reality from the required point in time and at the same time it evaluates how big the impact of date and time is on data accuracy.
- **Validity** of the data represents standards, syntax and rules that were defined for the dataset.
- **Accuracy** measures to what extent the data reflects the real-world object or event that is identified by it.
- **Consistency** verifies if the data sets can be compared to other data sets containing similar information.

Also, since spatial data is the type of data that is used for this research, positional accuracy has to be evaluated as well (Chrisman et al., 2006).

3.4 Data availability

Data availability analysis is not part of the data quality assessment but it is closely related to it. There are a few aspects that can be related to data availability, such as costs, openness, time etc, but this research will only focus on the temporal aspect of data availability.

3.4.1 Satellite data

Since satellite data is the core of this research, it is wise to start the analysis with the availability of satellite imagery. Optical remote sensing is widely used in agriculture and has a great list of advantages. However, its biggest disadvantage is its weather conditions intolerance - which means that f.e. clouds can impede optical imagery. In contrast, radar sensing, is not affected by cloud cover, but the use of radar data is not the scope of this research.

Below an overview is made of satellite images that are available for the research, the date and the satellite with which the image was taken are provided as well (see Figure 5). In total, 46 images from three different satellites are available. However, not all of them are usable.

Out of 46 images only 33 are not covered by clouds on the research fields (see Figure 6). Also, due to climate characteristics of the study area, the soil is usually covered with snow during the whole winter. In 2017, the fields were totally covered by snow until the 4th of March (judging from satellite images) which makes all the images until this day irrelevant for the research. The snow partially stayed until the end of April, but it was mostly on the roads and field borders, which is not obstructing the research but has to be taken into account nevertheless.

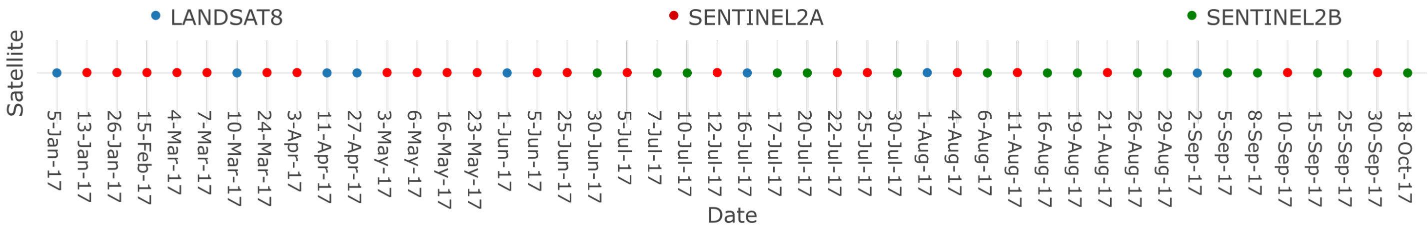


Figure 5: Overview of satellite data availability

Out of 33 images that are not covered by cloud, four still have snow (see Figure 7). Therefore a total of 29 images is used in this research.

Figure 6: Optical image obstructed by cloud cover

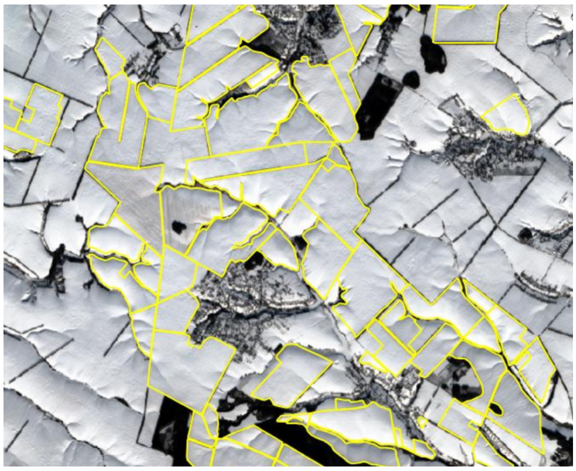


Figure 7: Snow cover over the fields as seen by optical imagery

3.4.2 Harvest data

The availability of harvest data differs per field and per crop, but it is important to mention that due to large sizes of fields in the study area, usually the harvest dates are spread over multiple days (it can go up to six different dates). This is also true for planting dates - even if they tend to be more homogeneous, there are some fields where planting dates differ. For field 6, the planting day is unique - the 23rd of April 2017, while harvest days are multiple - from 20 to 23 of November 2017. For field 10, there are two planting days - the 25th and 26th of April 2017, harvest days are varying between the 15th to the 18th of November 2017.

3.5 Crop

In this section the properties of the selected research crop are discussed. It is also discussed what are the most favourable soils of this crop, its growing rhythm and its known yield-defining factors.

3.5.1 Soil requirements for corn planting

For its optimal growth, corn requires soil which is rich in organic matter and moisture and is well-drained (Delp, 2015). Those conditions have to be complemented with warm temperatures and plenty of sunlight. There are also requirements for the soil temperatures prior to seed planting - it has to be around 15°and 18°, otherwise seeds will not develop correctly. The most optimal soil pH to grown corn is 5.8 and 6.8.

3.5.2 Corn growth stages

Knowing the growth stage of corn is important not only for the farmer to match his managerial practices according to the stage, but also for this research. It is important to know what the growth stage of the crop is when we look at a satellite image. This will help us derive the most optimal time slot for the detection of yield-limiting factors using remote sensing data.

There are multiple corn growth stages as can be seen in Figure 8. Those can be grouped in four big categories: emergence, vegetative, reproductive and maturity. Below, the leaf collar method (as suggested by Ransom and Enders (2013)) is used to briefly describe all growth stages of corn:

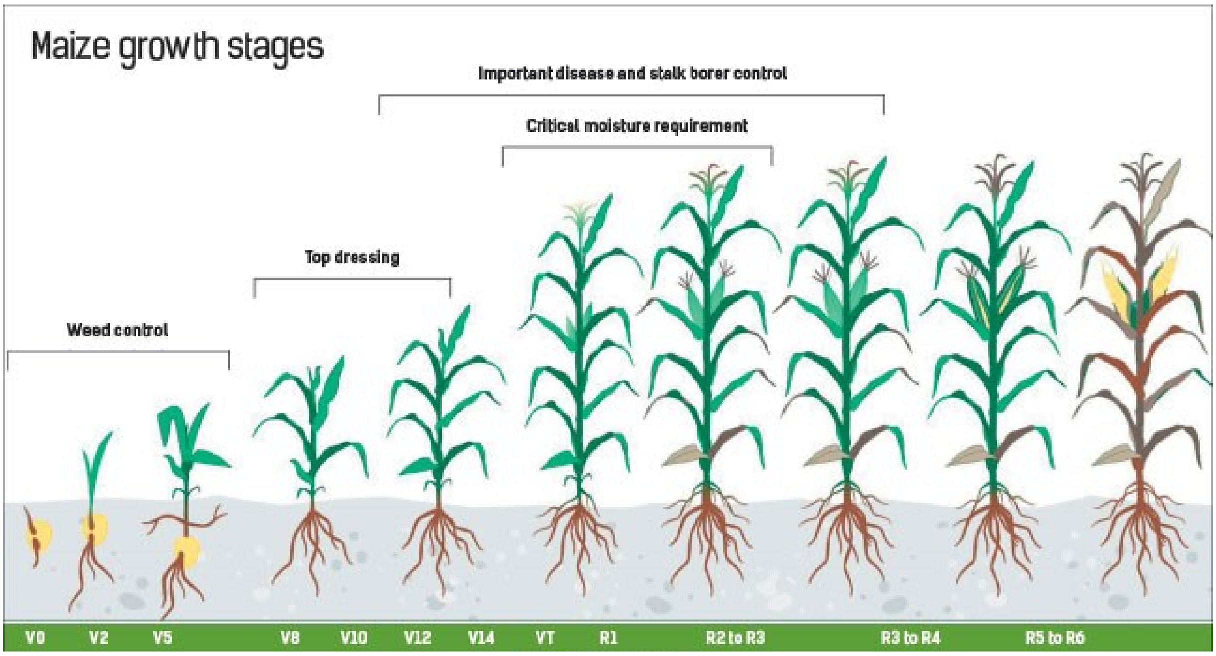


Figure 8: Corn vegetation growth stages

Source: www.farmersweekly.co.za/crops/field-crops/maize-production-managing-critical-plant-growth-stages/

- **VE. Germination and Emergence.** Once planted, corn seeds absorb water from the soil and begin to emerge. This process can take from five days in case of ideal climate conditions up to three weeks or more, if planting occurred during cool temperatures. This stage is crucial for the plant to sustain fungal infections - the longer it takes for the plant to break out the more chances it has to get infected in the later stage.

  In this stage, the collar is appearing on the first leaf. The next stages will all be determined by counting the visible collars (thus named V1, V2, V3 etc) until the tassel is emerged.

- **V3.** The collar of the third leaf is visible. This occurs usually two weeks after emergence. Photosynthesis is very active, but the main growing point of the plant remains under the ground.

- **V6.** The collar of the sixth leaf is visible. Tassel formation has commenced. The growing point rises above the surface and the stalk gets longer.
- **V9.** The collar of the the ninth leaf is visible. In this stage, tassel is rapidly developing. This occurs approximately 45 days after the emergence. If the plant is lacking macro- and/or micro-nutrients, it will be apparent. Therefore, soil nutrient and water needs are decisive for the plant growth at this stage due to the growth rate.
- **V12.** The collar of the 12th leaf is visible. It usually occurs six weeks after emergence. Adequate moisture and nutrients are critical during this time. Failure to provide it will diminish harvested yield. At this stage, the potential of kernels per row is determined.
- **V18.** The plant is about a week away from silk emergence, it develops rapidly. Silks will begin to grow as tassel is emerging. Brace roots are now growing.
- **VT. Tasseling.** During this stage, tassel, as well as all leaves, are fully emerged two to three days prior to silking, the plant has reached its full height. Hail is a danger, as it can result in loss of tassel, which means 100% yield lost.
- **R1. Silking.** At this stage, the plant is about 55 to 66 days after emergence. This stage commences when silks are appearing and pollination occurs. The plant is exceptionally sensitive to stress at this time - the largest yield reduction occurs with stress at this stage.
- **R2 - R3.** Those stages occur around 10 to 20 days after silking. Kernels are small, white and contain 80% moisture. Silks are getting darker and dryer.
- **R4 - R5.** Those stages occur about 24 to 42 days after silking. Almost all kernels are dented or denting and they have about 55% moisture.
- **R6. Physiological Maturity** This stage occurs about 55 days after midsilk. All kernels have attained maximum dry weight and their moisture content is varying between 30 to 35%. The biggest part of leaves are turned yellow or brown, but the stalk may be still green.

3.5.3 Known limiting factors

Before discussing already known limiting factors for corn, it is important to explain the difference between continuous corn (CC) production and corn rotated with soybean (CS) (other crops are rarely used). Changing crop type is a common practice among farmers and for corn crops it is even more important as production type might be one of the limiting factors for corn yield (Heggenstaller, 2004).

Several studies agree on the fact that continuous corn production might act as a limiting factor on its own. A recent study conducted by Gentry et al. (2013), has proven that it is easier to maximise yield of corn rotated with soybean than continuous corn. The authors have also determined three factors that explain 99% of difference between CC and CS yield : soil N (nitrogen) supply, field production type history and weather. Even though the crop production type can not be monitored with satellite data, it is important to take this knowledge into account as it might impact the research outcomes.

Another study conducted by Below (2008) has determined seven main corn yield-limiting factors (see Table 3). The author decided not to include several important factors that are considered as yield-limiting due to them being *"onetime improvements (e.g., tile drainage or waterways), they protect rather than increase yield (e.g., weed or pest control), or they involve decisions that do not need to be made every year (e.g., soil pH and nutrient levels)"* (Below, 2008). Also, the author is highlighting the interaction between the factors, which can be both negative and positive. Beneath, each factor is discussed more in depth.

Table 3: Seven most important yield-limiting factors according to Below (2008)

Factor	**Value(%)**
Weather	27
Nitrogen	26
Hybrid	19
Previous crop	10
Plant population	8
Tillage	6
Growth regulators	4

- **Weather.** Weather is by far the most important and most unpredictable yield-limiting factor for corn crops. While there are multiple weather factors that can impact the yield, the most important ones are rainfall and temperature. Even if the most optimal managerial practices are in place, weather will interact with all of them.
- **Nitrogen.** The presence of fertiliser N is heavily impacted by weather (rainfall), therefore it makes it highly dependent on the latest. Together, weather and the availability of fertiliser N explain more than 50% of yield variation.
- **Hybrid.** Hybrids are new breeds of corn developed to optimise corn production. Currently, commercial hybrids are being developed to help the plant overcome drought, increase kernel size etc.
- **Previous crop.** As described previously, previous crop can play an important role in defining yield.
- **Plant population.** Even though it has a smaller influence on yield as a stand-alone factor, it interacts intensively with all other factors, especially weather. Unfavourable weather conditions will negatively impact yield if plant population is high.
- **Tillage.** This factor is also known mostly for interaction with other yield-limiting factors. Weather, availability of fertiliser N and hybrid are the factors it interacts the most with.
- **Growth regulators.** Those are chemical components that lead to a positive change (growth) of the plant and therefore also yield.

3.6 Analysis

In this section, an overview on analysis techniques that will be used in the research is provided.

1. The first step prior to direct data analysis would be data quality assessment. The quality of the data will be reviewed and questioned according to the methodology described in section 3.3.
2. The next step would be to analyse the yield data on yield variability. This will allow us to identify weak spots on the yield map that are useful for the study of the effect of yield-limiting factors on the yield, as mentioned in section 2.3. This will be done by mapping the yield data.

A visual study of yield spatial distribution can already indicate some present spatial patterns that might be explained in the later stage of the research.

3. A selection of satellite images to be used in the research will be made. As 29 available images exist and three layers (NDVI, ARVI and NIR) are used, it is becoming a large task to analyse all of them. Therefore, one image per month will be selected for the further analysis.

4. Satellite images will be visually analysed for the spatial correlation between them and yield variability. Also bare soil images from the beginning of the season (before crop development) will be analysed.

5. The yield data will be compared with selected spectral bands and VIs (see section 2.5), they will be tested statistically. This will be done by visualising their scatter plots and analysing corresponding statistics, which will indicate if there is a correlation between two datasets.

6. A few objectives of this research are aimed at discovering which VIs and dates are the most optimal for yield-limiting factors detection by using satellite data. In order to answer this question with the data analysis and not with the literature review, several repetitions of a similar type are needed. For example, it could be possible that for a VI_a, the most optimal day of detection is day X, but for a VI_b, the most optimal day is day Y. To be able to verify this, 24 spatial analysis are needed (3 x 8 selected satellite images).

   In order to simplify this work, a ModelBuilder tool from ESRI will be used to allow an automatic processing of a large amount of data. In the end, a Table containing multiple statistic indication about correlation between variables VI and date will be created, which will help answering subquestions two and three.

7. In case correlation between yield-variability and VI is significant, other data will be used to help answering the question about what might cause this variability. Therefore soil and elevation data will be tested as possible yield-limiting factors and tested for statistical correlation with satellite images.

3.7 Validation

The validation of the results will be accomplished on the basis of matching spatial patterns in satellite and yield data. Harvest data will be used as ground truth data and likely for validation purposes.

The main analysis will be carried out on field 6 and field 10 will be used to validate the outcomes of this analysis. It will be tested whether the spectral bands or VI that identified the yield-variability on field 6 are showing the same result on field 10. The same will be done on time - is the most optimal period derived from field 6 the same for field 10? Finally, it will be tested if the yield-limiting factors identified for field 6 are the same for field 10.

The outcomes of the validation process will give us an estimate on how much yield-limiting factors are field-dependent. It will also give an insight about the degree of possibilities on re-using these research results on other fields and/or farms.

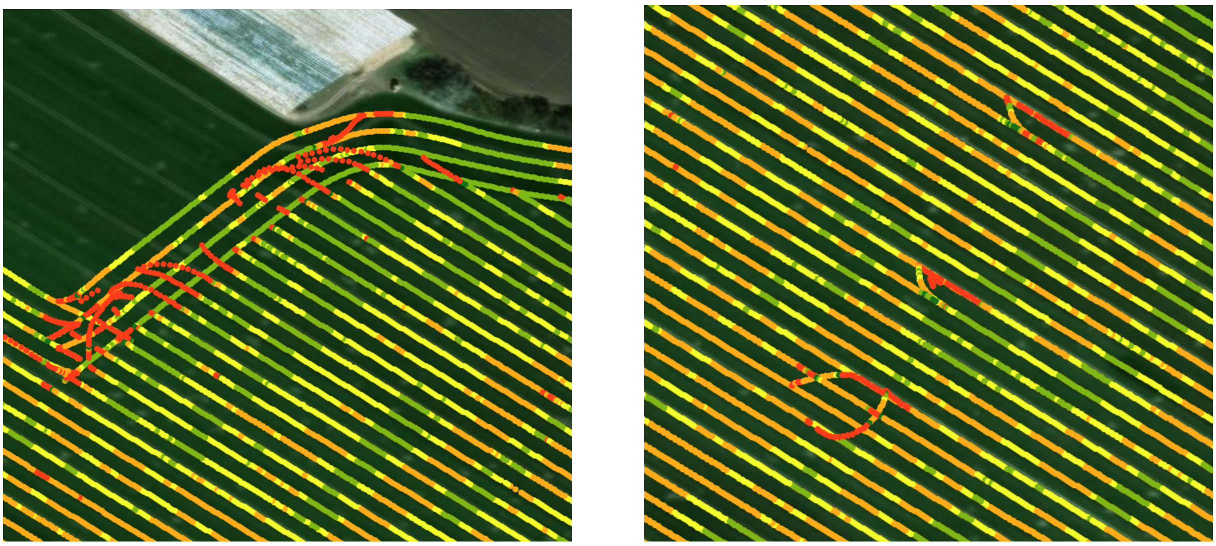
4 Results

As was mentioned before, the majority of data sets used in this research are collected by people who are not specialists in geo-information and are not trained for this. This data collection was never the ultimate goal, but rather a side bonus from an operation that was carried out at the same time as other tasks. For example, yield data was collected during harvesting which means that the priority during this time was harvesting the yield and not data collection. The same applies for soil data. Even though here data collection was the primary purpose, it was not collected by specialists. The only type of data used in this research that was collected and processed by the company specialised in geo-information field and image processing is satellite imagery. In the next chapters we will look more in detail into used data sets and assess their quality.

4.1 Yield data

Yield data is a relatively well collected, precise, accurate, consistent, continuous and complete data set, despite the fact that is was collected by a person who does not have a background in geographical information science. There are some small issues that needed to be taken care of before this data set could be used for the research, they will be described below.

First of all, data is collected differently alongside the borders of the field than on the rest of the field (see Figure 9a). This is due to the turning trajectory of the tractor. Since the data collecting device does not stop working during those turns, they are also appearing on the dataset. In order to prevent those from impacting the results of the research, the decision was made to perform an 85 meters inside buffer. This will assure that all tractor rotation points are not included in the data set and will not influence the analysis. The amount was based on a visual assessment of the data extent impacted by the turning trajectory.



(a) Border errors *(b) Row errors*

Figure 9: 2 types of error in yield data

Secondly, as can be seen in Figure 9b, sometimes the tractor has jumped lines that result in wrong paths and values. Unfortunately, there is not a uniform tool that can correct those issues (as they are quite often reoccurring throughout the data set in both fields). A manual correction is used to correct the path and delete wrong values.

Also, due to the nature of the data (points) and its prompt to double values, it was decided to first convert the vector points to raster data set in order to calculate neighbouring values that are missing. This will be done using the ArcGis Pro ”Point to raster tool”, that will transform the vector data into a uniform grid raster data set with a cell resolution of 10 meters (as it is the resolution of satellite images). The advantage of using this tool is that if there are multiple points that fell into the same cell, the mean was calculated and only one value was added to the grid. It helped illuminate the issue of double values.

Once yield data was in a raster grid format, it was again converted to vector using the ArcGis Pro ”Raster to point” tool. This was necessary to make a vector grid data set that can be used for statistical tests and to overcome the problem of connecting different data sets - from now on values from other data sets will be added to this one. This will ensure that all values are attributed to the exact same point.

Data quality assessment for yield data from both fields was performed accordingly to the methodology described in section 3.3. Below the results of this assessment can be found:

- The data set is **complete**, as the whole field is covered.
- The **uniqueness** of the data is not maintained throughout the field. As was shown in Figure 9, the double entries exist and they do impact the research. Therefore, as described earlier, a decision was made to manually correct those problems by a) deleting them manually and b) using the ”Point to raster” tool.
- As yield data can be collected only once per season, the **timeliness** is correct.
- **Validity** of the data set is hard to evaluate as it might have been defined by the manufacturer of the collection hard- and software and this is not known. It is known that the direct collector of the data has not installed any standards, syntaxes or rules while collecting the data set.
- The **accuracy** of the data set was tested with the main agronomist of the farm who confirmed that the actual yield matches the one represented by this data.
- The **consistency** of this data set was compared on the ”Elevation” attribute. The elevation values were compared to elevation values from two other data sets - namely soil data set and Shuttle Radar Topography Mission (SRTM) data set and the values matched with a difference of around two meters in some places (this can be explained by the resolution of the comparing data set, f.e. SRTM data has 30 meters resolution and is less detailed).

To conclude, the yield data can be named as suitable for the research. It is important to mention that it has some less strong sides (like uniqueness and validity) and those might impact the research outcomes. However, no other sources of similar data exist and this data set is in the core of this research, so it will be nevertheless used for the analysis with extra precautions.

4.2 Soil data

The soil data used in this study was collected by the main agronomist using Veris Technologies scanners. It was then send to the same company for processing and calibrating the results. In general, the data set is continuous and representative, however, it has some drawbacks that will be discussed later in this chapter.

The soil data contains information about organic matter content, Electrical Conductivity (EC) and pH. The data also contains a report about data quality and issues that occurred with

Veris Techologies. For example, for field 6, a conclusion is made that *"There is an unnatural data break within the field."* (see Annex C) This data break is however totally natural and will be discussed in more detail in section 4.4.1.

Another drawback of OM data is that it is not stored as absolute values, but as ratio between the Red and Infrared (IR). Veris Techologies has developed a method that allows to quantify the amount of organic matter content in a soil by using the difference in reflectance of the soil in Red and IR band. In Figure 10, this ratio can be seen. Smaller ratio represents area with a high organic matter content.

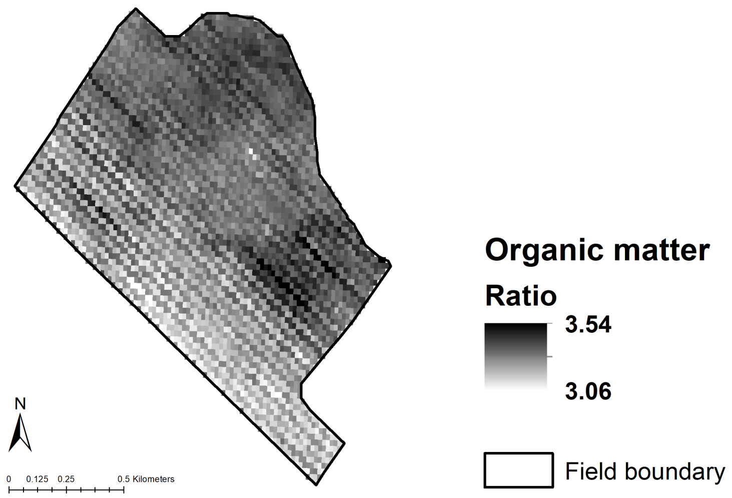


Figure 10: Map of organic matter content for field 6

From the map of soil pH (Figure 11), it can be observed that the spatial resolution is much worse in comparison to the organic matter map. Such coarse resolution will negatively impact the results and will have an effect on the statistical analysis, since a lot of details will be lost.

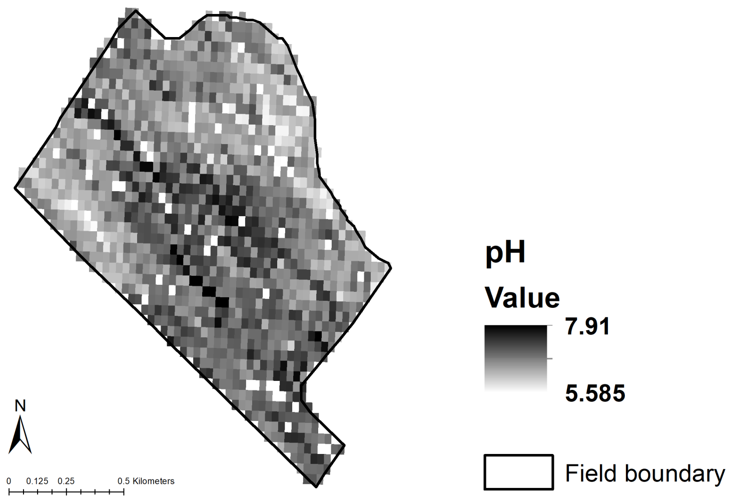


Figure 11: Map of pH content for field 6

Electrical conductivity data is the most complete feature of this data set (see Figure 12). Its spatial resolution is very detailed and the data is provided not only in the form of ratio but also as absolute values for both shallow and deep measurements.

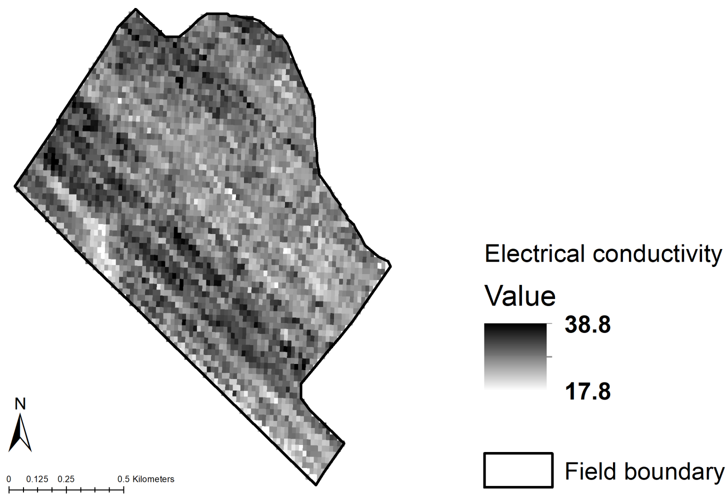


Figure 12: Map of average electrical conductivity for field 6

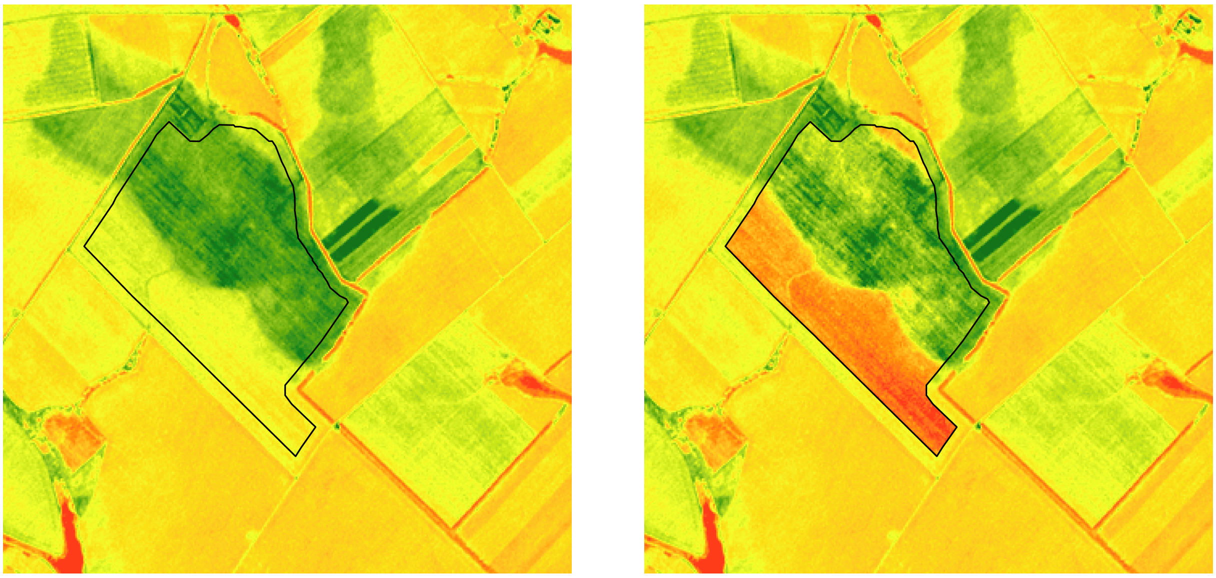
It is relatively hard to assess the quality of soil data, since not a lot of data sets of needed spatial resolution exist to make a comparison with. More comments about soil data quality will be made after this data set will be used in the analysis. For the primary quality assessment it can be stated that:

- The **completeness** of this data set can be questioned. It does cover the entire field, however not on the same resolution level for all attributes. Therefore, some parts are less covered than others.
- The data is **unique**, there are no double entries.
- Since soil properties do not change drastically over time, the **timeliness** of the data set is judged correct.
- It is now known which standards, syntax and rules were set during the collection of this data. However, the company has processed the collected data and **validated** the dataset.
- The data is **accurate**.
- It is hard to compare the **consistency** of this dataset. The spatial resolution is very detailed and no other data set exist that can provide such detailed information for comparison. Furthemore, the main agronomist who is very familiar with the field has confirmed that overall the soil data represent the real-world situation.

4.3 Satellite data

Satellite data is collected by satellites Landsat 8 and Sentinel 2A and 2B. However, only Sentinel 2 images are used in the analysis. As they were processed by a company specialised in image processing and delivered as final images, it is hard to evaluate the quality of processing. Therefore, the quality assessment will only focus on processed images and leave all the other features out of discussion.

- All of the used images are **complete**. The images that have a partial cloud cover were not selected for the analysis.



(a) Entire area *(b) Field boundary*

Figure 13: Differences in Blue band visualisation for a non-cropped and cropped image

- As the data is stored in raster format, all the values are **unique** - one cell has one value.
- The data represents the state of the fields at a given date, therefore its **timeliness** is absolute.
- As being simply a user of the data set, it is hard to assess the **validity** of the data. However, knowing that is was collected and processed by professionals, it can be assumed that is was collected using predefined standards, syntaxes and rules.
- As many agricultural projects and tools heavily rely on the satellite data collected by Sentinel and Landsat satellites, its **accuracy** can be graded as very good.
- The **consistency** of the data set was tested by calculating NDVI from original bands and comparing it to the NDVI provided by NEO. The values matched 100%, meaning that the data is accurate.

To conclude, the satellite images are of a high quality and precision. The added value they provide to the research is unquestionable. However, due to their high spatial resolution (10 meters), they become quite voluminous to work with. Also, the provided images were covering a large surface area. Therefore, all the calculations for VI were based on that total surface, meaning that every pixel plays a role in the final visualisation. In order to enhance the visualisation of a particular field and reduce the file size, the raster images were cropped to the research fields. This has shifted minimum and maximum values and made the differences on the intra-field level much clearer (see Figure 13). This adds a totally new level of details and offers new angles for analysis. It also speeds up the processing and calculation time.

In order to make the statistical comparison between the yield and satellite data possible, the latter needed to be converted into a vector format. The yield data file (already converted to a vector grid) was used as a base to append values from satellite images. This was done using ArcGis tool ”Extract Multi Values to Points”. All the values of spectral bands were added to a vector file that made it possible to perform statistical analysis on the correlation between yield and satellite images, as for every yield point a value representing a spectral band or VI is now present.

4.3.1 Image selection

Once all the operations on satellite images described in section 4.3 were carried out, a problem of too many images and too many spectral bands that needs to be analysed occurred. A decision was made to select one image per month which would make analysis less time-consuming. In the end, only eight images were used for statistical analysis, all images obtained with Sentinel 2 satellite at 10 meters resolution.

Below an overview of dates of selected images for statistical analysis is made. In the annex D, an overview of all images in true colour for field 6 can be seen.

1. 4 March 2017
2. 3 April 2017
3. 6 May 2017
4. 5 June 2017
5. 17 July 2017
6. 11 August 2017
7. 15 September 2017
8. 18 October 2017

4.4 Yield variability

A visual study of the yield variability map revealed two interesting points for further discussion - zonal division of the field and a clear impact of geographical features on yield. In this chapter, we will first discuss the zonal division and its possible causes and then look into what kind of geographical features is impacting the yield.

4.4.1 Zonal division

By only looking at a yield variability map it is already possible to divide the field into two zones: one where crop is performing much better (lower zone) and another - where crop is clearly performing less well (upper zone) (see Figure 14).

One of the reasons why the upper part of the field is less productive might be weeds infection. According to the main agronomist, field 6 had weeds issues, which, in his opinion, severely affected the yield. However, not the entire field was infected. He indicated that the location of the weeds problem was at the same spots where the lower yield can be observed.

On the other hand, lesser productivity can also be explained by other factors. If we take a closer look in Figure 15, which shows us bare soil satellite image of field 6, we see a clear difference in colour between the lower and upper part. Those are differences in soil type - the lower part represents black soil which is rich in organic matter, whereas the upper part is more clay soil with a smaller amount of organic matter (information provided by main agronomist and Figure 10). As was already discussed in section 3.5.1, corn prefers a soil rich in organic matter. There is clearly a visual correlation between the bare soil image and the yield variability map, as the lower part (rich in OM soil) produced higher yield than the upper part (clay soil).

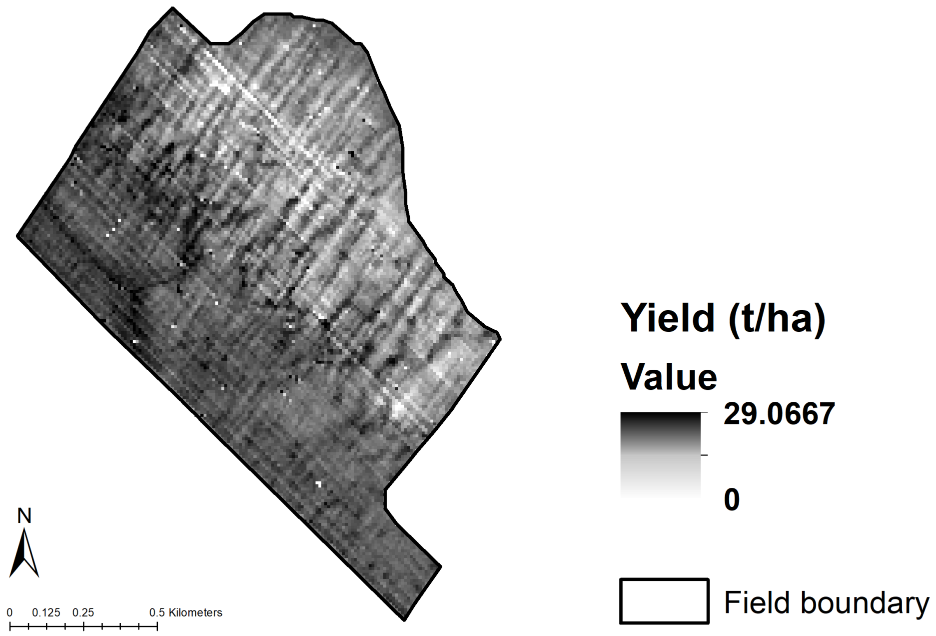


Figure 14: Yield variability map for field 6

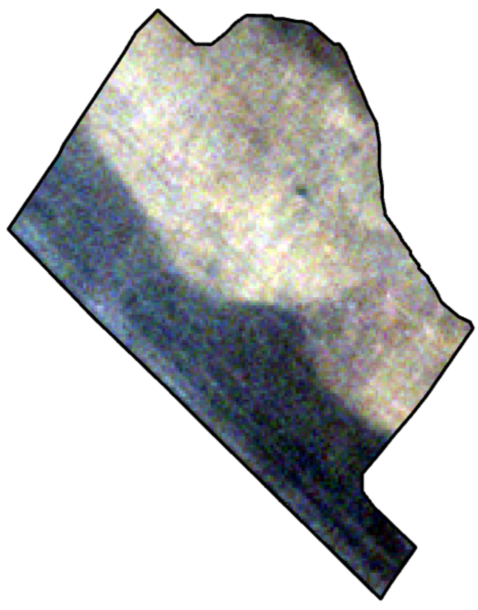


Figure 15: Satellite image of field 6 taken on the 4th of March 2017 shown in true colour

4.4.2 Topography

In section 2.4.2, it was mentioned that topography is a known limiting factor that is influencing the yield variability. If comparing the yield variability map (Figure 14) to the elevation map (see Figure 16) some similarities are clearly visible.

The impact of some geographical features can be clearly seen (see Figure 17). From analysing satellite and elevation data, as well as from features mentioned by the main agronomist, it is known that this geographical feature is the centre of the slope - a place where water flows

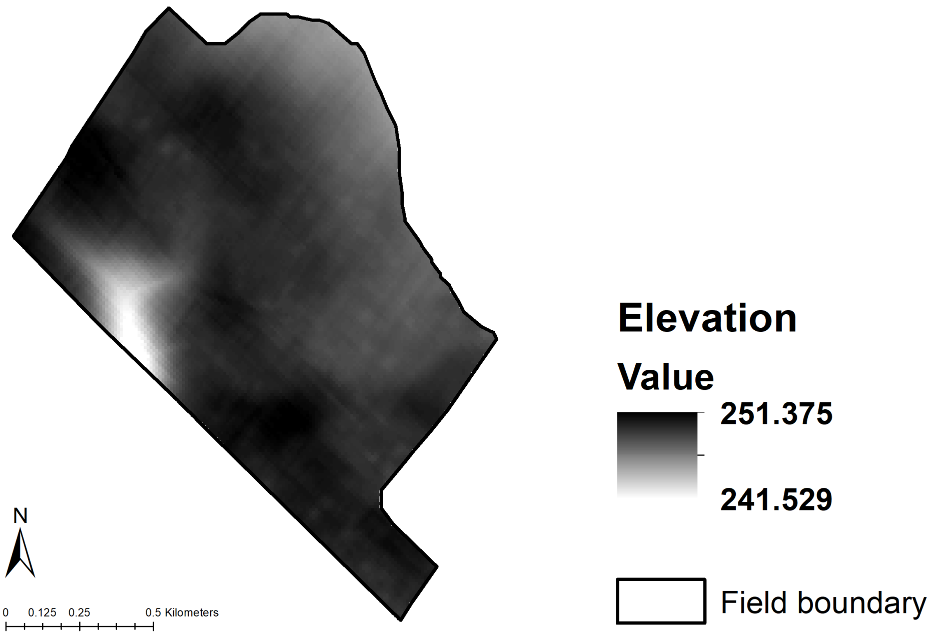


Figure 16: Elevation map of field 6

and accumulates. This feature is visible on almost all satellite images (see annex D).

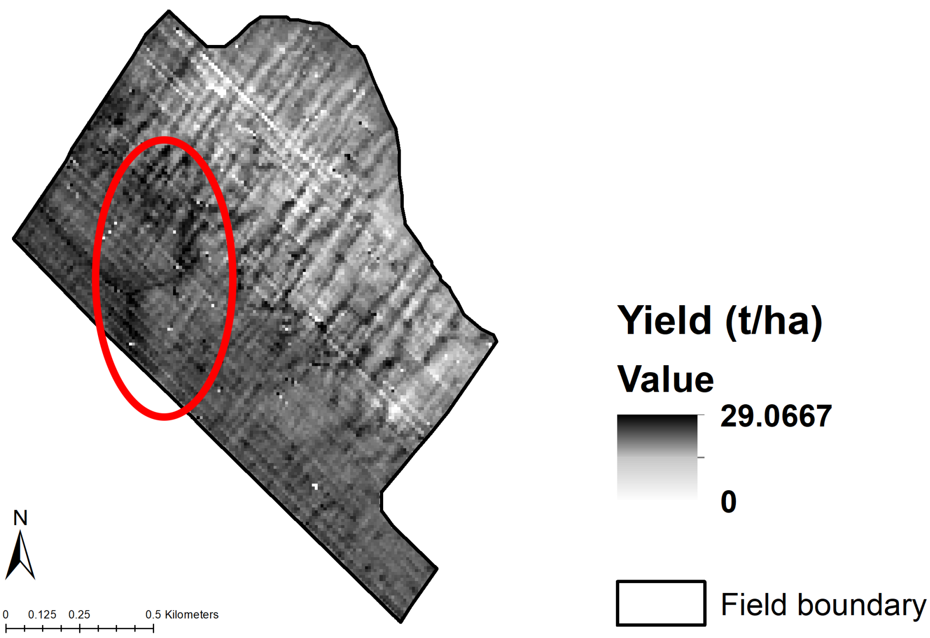


Figure 17: Impact of geographical features

The elevation within field 6 varies between 241 meters for the lowest part and 251 for the highest part. 10 meters difference is not radical, therefore small differences in altitude are not clearly visible on the elevation map. In the upper part of the field, a series of small hills is present, however they are not visible on this map. Those hills are the best to be spotted in Figures 23 and 24. They have a big impact on yield - a small change in altitude provokes big differences in the final yield.

4.5 Visual analysis of optical satellite images

In this chapter the selected satellite images are analysed in more detail and they are visually compared to the yield variability map.

4.5.1 4 March 2017

The earliest cloud and snow free image of field 6 was made on the 4th of March (Figure 15). This is an image of a bare soil a few days after the snow has vanished, since only two weeks before the field was entirely covered by snow. A clear difference in soil types between the lower and upper part of the field can be seen on this picture. As mentioned earlier, the upper, lighter in colour, part of the field represents the clay soil with a lesser amount of organic matter, while the lower part is the Chernozem soil with a high percentage of organic matter. The centre of the slope can be seen, but not very clearly.

While it is hard to correlate this image with the yield map visually, one thing is sure - the difference in soil types can be clearly seen on the yield variability, as there too the upper part has produced less yield than the lower one.

4.5.2 3 April 2017

The image that was made on the beginning of April (Figure 18) still mostly shows the differences in bare soil, with the exception of some green plants emerging at the upper left part of the field. These are, most likely - weeds, since nothing has been planted yet as the planting date has not come yet (planting date is the 22nd of April). The contrast between the soil types is even more apparent than on image 15 and the centre of the slope is clearly visible.

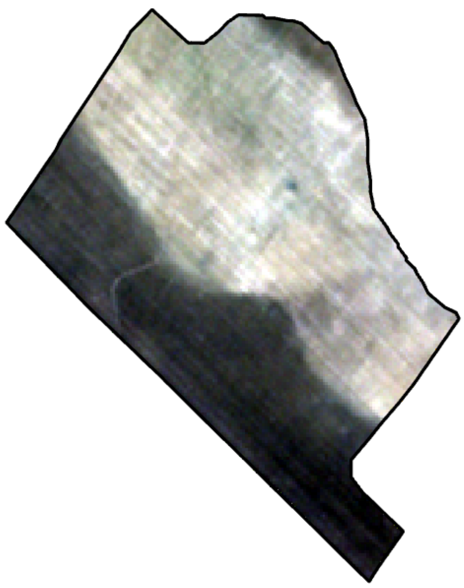


Figure 18: Satellite image of field 6 taken on the 3rd of April 2017 shown in true colour

If we divide this images into zones as it was done with the yield variability map, it can be observed that they are visually similar - same lower/upper division with a clear difference in the centre of a slope area.

4.5.3 6 May 2017

The satellite image (Figure 19) taken at the beginning of May, approximately two weeks after the planting date does not differ much from the previous two. The plant is in the phase of emergence according to the corn growth stage described in section 2.6. It is not possible to see the plant yet, as it takes up to three weeks for the plant to emerge.

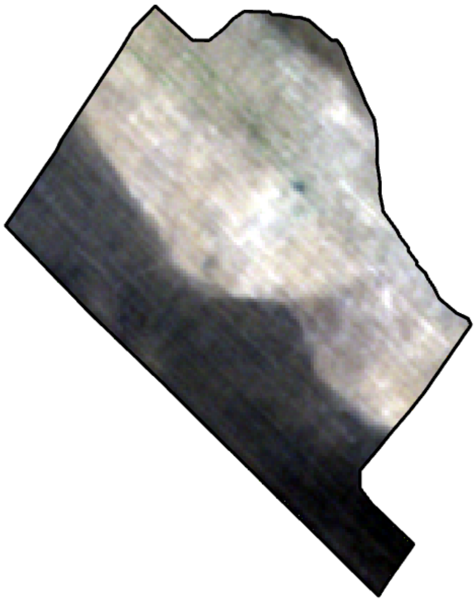


Figure 19: Satellite image of field 6 taken on the 6th of May 2017 shown in true colour

4.5.4 5 June 2017

The image taken in the beginning of June (Figure 20) represents the plant in its V3 - V6 growing stages. This means that the plant has emerged but the main growing point of the plant still remains under the ground, that is why the image does not look much different from previous ones, even if one month and two weeks have passed since the planting day.

4.5.5 17 July 2017

A lot of changes have happened during the month of June and the first half of July, that is why the image taken on the 17th of July (see Figure 21) looks very different from all the previous ones. This image is very difficult to visually interpret as it is very dark. The lower part of the field looks very dark and as if there is no vegetation there. However, it is only an illusion, since the vegetation is developing there too, only the spatial resolution of the satellite image (which is 10 meters) is not enough to show it. In majority, it is still bare soil, with some parts covered with vegetation. The plant is entering its V9 - V18 stage - the most rapid growth stage.

An interesting point to mark is that the green vegetation at the upper left corner is probably not the corn plant itself. As it was already appearing on the images since the month of April, therefore before the seeding, it cannot possibly be corn. Taking into account the weeds problems mentioned earlier (see section 4.4), it is very likely that this green vegetation are weeds.

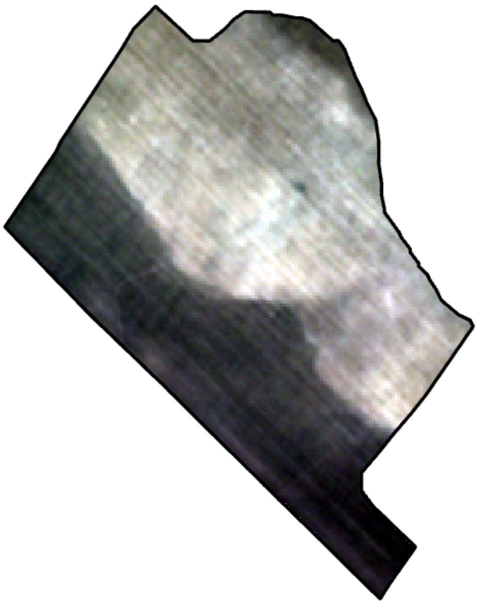


Figure 20: Satellite image of field 6 taken on the 5th of June 2017 shown in true colour

Figure 21: Satellite image of field 6 taken on the 17th of July 2017 shown in true colour

4.5.6 11 August 2017

Around half August, the satellite image (Figure 22) looks more homogeneous, even though the soil differences are still visible. The plant is entering the tasseling stage, but as can be seen, the plant development is not equal throughout the field - some parts are still in their previous growth stages.

This satellite image begins to have clearly visible similarities with the yield map - on the upper part, horizontal lines can be seen, they can also be seen on the yield map. The horizontal

Figure 22: Satellite image of field 6 taken on the 11th of August 2017 shown in true colour

lines are differences in elevation, they can also be seen at the elevation map (see Figure 16). The centre of the slope can also be seen on the image.

4.5.7 15 September 2017

The image of the 15th of September (Figure 23) is now totally covered by plants, no soil is visible. The lower part appears more green than the upper part and a lot of differences can be seen in elevation. What should be noted is that while elevation differences are very clearly visible on this image, they are not that visible on the elevation map itself. That means that the smallest difference in elevation has an impact on how the plant grows and its future yield. Since the biggest difference between the lowest and the highest point registered at this field is 10 meters, it means that even centimetre change in elevation has an impact on the plant development.

The centre of the slope appears very dark as well. This can be explained by the moisture. As this is the lowest point, it is where all the water flows and where the water stays. The plant has passed the tasseling stage and is entering the silking phase. This phase is very important for the future yield - the stress can cause the largest yield reduction at this time.

The image of mid-September looks almost identical to the yield map. All the elements that differ in the yield variability map are visible here too. By only looking at this image it is possible to get a very good and precise idea on how the future harvest will look like.

4.5.8 18 October 2017

The final available image of field 6 was made on the 18th of October (see Figure 24). It is approximately one month before the harvest day. The image looks very dark, with a lot of contrast between the dark and light zones. The light zones represent lower areas, where the yield is higher. This image visually correlates very well with the yield map, they are almost identical.

Figure 23: Satellite image of field 6 taken on the 15th of September 2017 shown in true colour

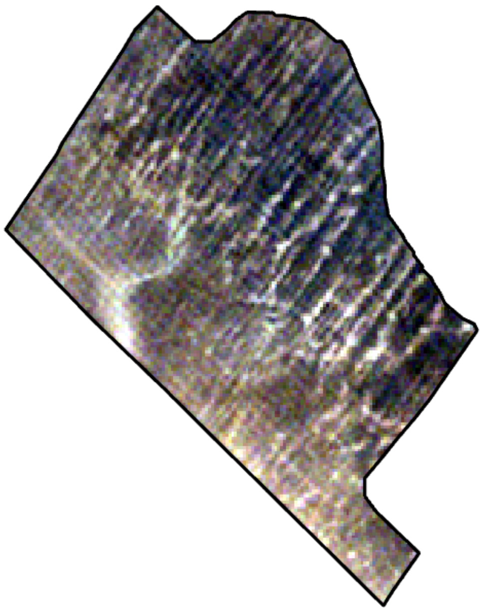


Figure 24: Satellite image of field 6 taken on the 18th of October 2017 shown in true colour

Since the last available image was taken approximately one month before the harvest, it is impossible to see what transformations occurred in between the 18th of October and the end of November. Most likely, no major changes occurred as the plant development was already predetermined in the earlier stages.

To conclude, during some dates satellite images visually correlate well with the yield variability map, however on some dates this correlation is very low. What is interesting to note is that soil plays an important role in the visual analysis - bare soil images can give a good indication for the future yield differences. Late images have shown a strong connection between them and the yield variability map, which makes remote sensing a good technique for the detection of yield variability.

4.6 Yield variability detection using NIR, NDVI and ARVI

Vegetation indices are mostly used to supervise the state of the crop and its performance, but theoretically, they can also be used for detecting yield variability. Vegetation indices are capable of detecting plants that are developing slower than their neighbours or are affected by unfavourable conditions. The same reasons will influence the final yield, therefore the correlation between remote sensing vegetation indices and yield is direct. In this chapter, this connection will be studied and it will be described in more detail which vegetation indices or bands are the most optimal for the detection of this correlation.

For field 6, series of statistical tests were performed to test the coefficient of determination (R^2) between the values obtained from satellite images and yield. Below, an overview can be found of the R^2 test per date (see Table 4). The p-value for all those tests is smaller than 0.05, which means that the results are statistically significant and can be taken into account.

Table 4: R^2 values of NIR band and vegetation indices with yield

Date	**NIR**	**NDVI**	**ARVI**
4 March 2017	0.172	**0.191**	0.016
3 April 2017	**0.241**	0.046	0.024
6 May 2017	**0.261**	0.013	0.156
5 June 2017	**0.216**	0.012	0.071
17 July 2017	0	0.078	**0.106**
11 August 2017	0.11	**0.211**	0.05
15 September 2017	**0.722**	0.698	0.697
18 October	**0.5**	0.282	0.009

As can be observed in the Table 4, the R^2 varies between 0% to 72.2%. Considering the fact that satellite images are taken from the space and the yield map was created on the ground using a combine - 72% correlation between them is a very good result. The average correlation in NIR is around 28%, it is then followed by NDVI - with a mean of 19% of correlation and ARVI - 14%. What is surprising is that the lowest and the highest percentages of correlation are both found in Near Infrared band.

Out of eight days of observations, the highest degree of correlation between satellite images and yield can be found in the NIR band. NIR has the highest resemblance to the yield in five images, which means in more than half of the observations. That means that Near Infrared band can be used in 62.5% of cases for the comparison of remote sensing images to final yield and it will produce the best results. NDVI has the highest correlation in two occasions and ARVI has only one. Scatter plots for the NIR band against yield will be discussed with more details in section 4.8 (see appendix E for all scatter plots).

4.7 Temporal differences in yield variability detection

In this section a focus will be made on the temporal aspect of the yield variability detection using remote sensing techniques.

An interesting point to notice is that all the indices/bands performance lines (see Figure 25) are different. While the three lines have slightly different beginnings, all three lines have the same ending - they all start to grow from mid-August and attain 70% of correlation by month of September. Unfortunately, there are no observations available after the 18th of October, therefore it is impossible to study the correlation right before the harvesting.

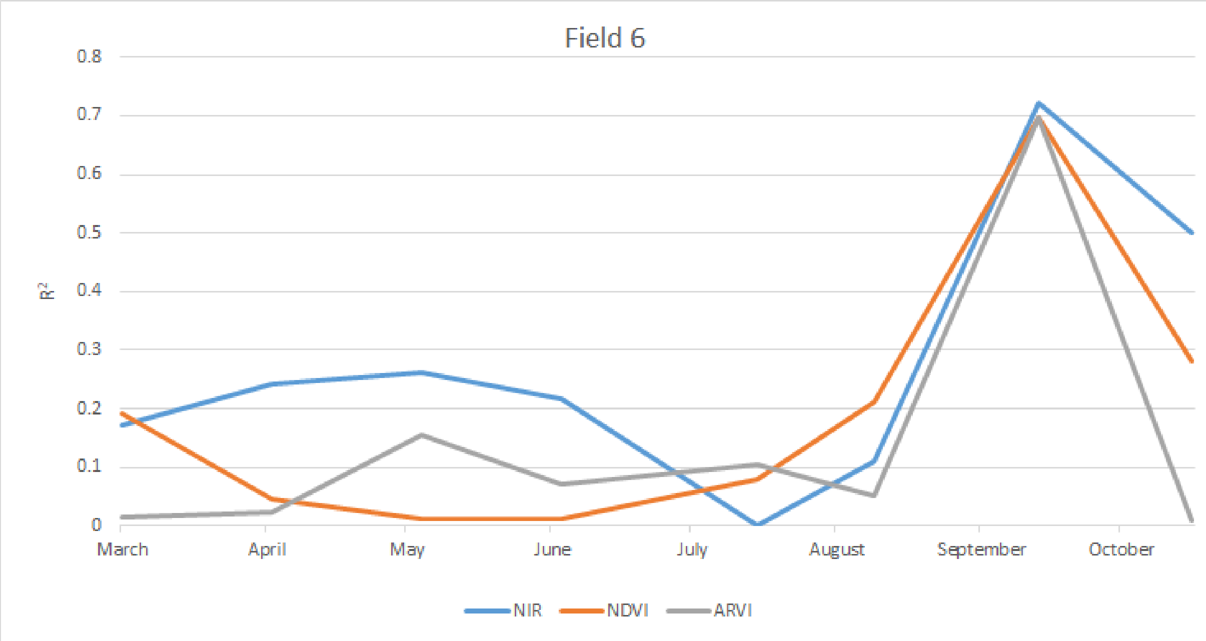


Figure 25: R^2 of NIR, NDVI, ARVI during the growing season for field 6

In the beginning of the growing season, NIR has the highest correlation with the final yield. This is due to the fact that soil has a lot of differences in organic matter content. NIR correlation with yield varies between 17% to 26% for bare soil images. This is a high result, considering that it is only bare soil and no vegetation yet. This indicates that soil has a big impact on the final yield and this impact can be detected by NIR.

In July, a lot of changes happen to the crop. The crop enters the stage of rapid development and by half July the field starts to be covered with vegetation. It is therefore very hard to distinguish healthy from weak vegetation. Bare soil is still visible and it makes it hard to detect yield variability, since it affects the general image. While the correlation in NIR is 0, other indices show a much better result. ARVI, especially, can detect 10% correlation, NDVI is showing 8%. According to those results, July is the worst moment for detection of yield variability.

From the month of August, the situation is rapidly changing. The field looks very saturated from space, since the vegetation is growing very densely. The highest correlation is again back to around 20% as in the beginning of the season and it is found in the NDVI band. This can also be seen in Figure 22, the same features as on the final yield map (Figure 14) start to appear, while they were not visible yet on the image from July (Figure 21).

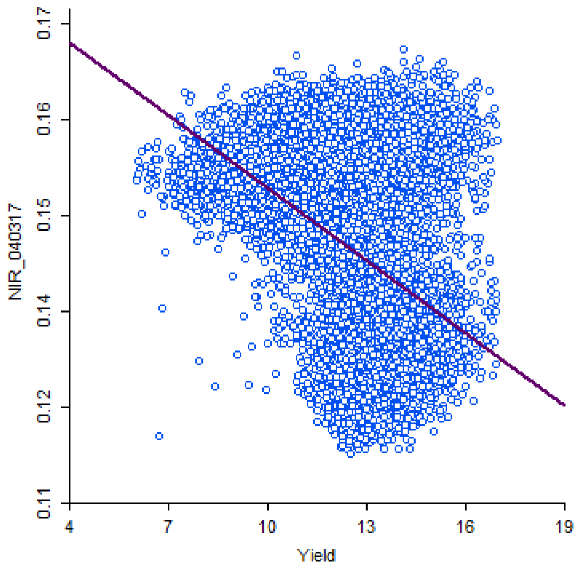
The most important change occurred in September. There is an increase of 50% in correlation comparing to the previous image - now all the bands and vegetation indices have a R^2 of minimum 0.697 and maximum 0.722. Mid-September results reach the highest percentage of correlation between images taken from the satellite and yield, therefore it is considered as the best moment for yield variability detection in this study.

A decrease of correlation can be observed in October, however the correlation stays relatively high with a R^2 of 0.5. In the end of the growing season, NIR outperforms again NDVI and ARVI.

4.8 Yield variability through NIR

The correlation between yield and satellite images was already discussed in sections 4.4 and 4.5. As determined in section 4.6, the most optimal band for evaluating this correlation is NIR. This section describes the statistical relationship between yield and NIR band in more detail.

As previously discovered, a bare soil image can be visually correlated with the final yield (see section 4.4.1). This correlation can also be seen in the scatter plot (see Figure 26). With a R^2 of 0.172, there is a relatively high correlation between the final yield and NIR values from the 4th of March and with a p-value smaller than 0.05 this correlation is statistically significant. However, as can be seen in the scatter plot, the values are not distributed perfectly along the fitting line. The values are grouped in two big clusters, with an upper cluster being wider than the lower one.



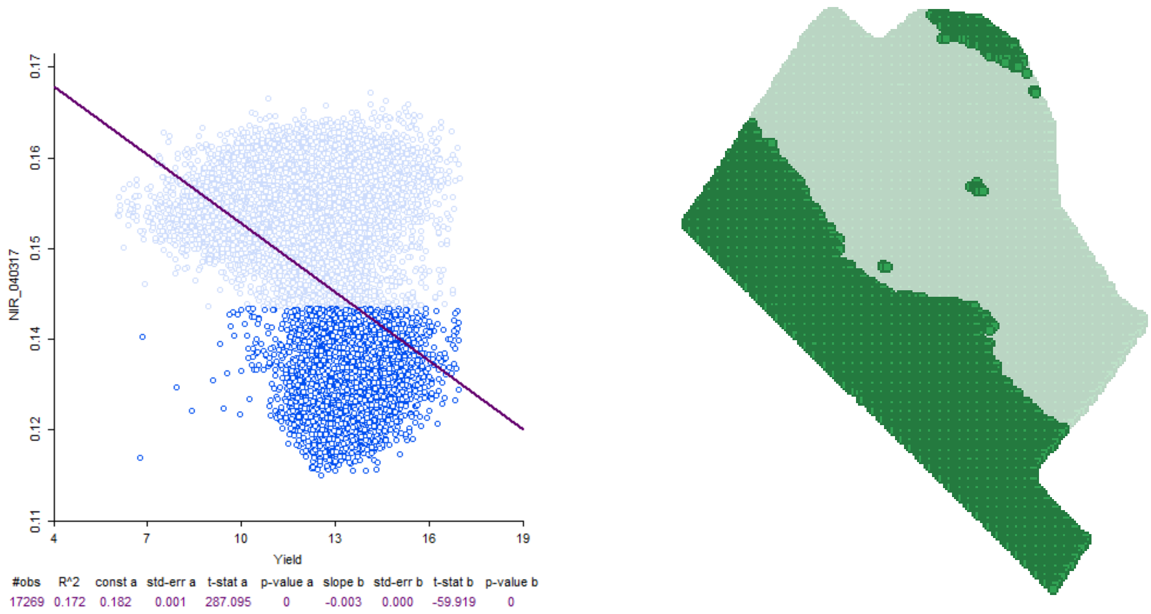
#obs	R^2	const a	std-err a	t-stat a	p-value a	slope b	std-err b	t-stat b	p-value b
17269	0.172	0.182	0.001	287.095	0	-0.003	0.000	-59.919	0

Figure 26: Scatter plot of yield (in t/ha) versus NIR values on March, 4th 2017

By selecting the lower cluster values from the scatter plot (Figure 27(a)), it can be determined why the values are grouped into two clusters. If we take a closer look at where the clustered values are located (see Figure 27(b)), we can observe that the lower cluster represents the darker part of the satellite image (Figure 15) and higher yield area (Figure 14).

This clustering is explained by the high reflectance of low organic matter soil in NIR and lower reflectance for a soil that is richer in organic matter, which appears darker on the image. That is why on the scatter plot two clusters are present - therefore one that represents high reflected part and a lower reflected one. What is important to notice from Figure 26, is that the low reflectance cluster has higher values along the x-axis, which represent yield. While the upper cluster also has a number of high yield values, it is clearly visible that the lower reflectance cluster has higher yield values.

The differences in reflectance of different soil types explain the negative slope of fitted line in Figure 26. Higher NIR values represent low organic matter soil which is less suitable for



(a) Scatter plot selection

(b) Selection location on a map

Figure 27: Scatter plot selection

corn growth, therefore it produces lower yield. Lower NIR values represent darker soil which is rich in organic matter that is more fertile, and therefore the yield is higher, which is visible on the scatter plot.

Similar patterns are observed in other scatter plots of yield and NIR up until the month of July (they can be found in annex E). Until July, the correlation is always negative, R^2 varies between 0.172 and 0.216 and values are still grouped in two clusters in the same way - upper and lower. However, a change occurs in July - because of the high plant density and a little variation in vegetation, NIR values do not correlate at all with yield (see Figure 28) - R^2 is 0 and it can be seen that the points are randomly distributed.

Starting from the month of August, the correlation pattern changes. The correlation is now positive and will remain like this until the end of the growing season. In August, the correlation between yield and NIR values is still relatively low (see Figure 29). Even though the R^2 is low (0.110), the values are now distributed more homogeneously and along the fitting line. The positive correlation now indicates that the field is now at least partly covered with vegetation. Since healthy vegetation is highly reflective in NIR, the correlation is now positive, with high NIR values representing healthy vegetation and not anymore bare soil.

However, the most drastic change occurs in September (see Figure 30). Two months before the harvesting date, the coefficient of determination is 0.722, meaning that NIR variation of that day can explain 72% of the final yield. The scatter plot of this day looks very homogeneous, with points distributed very close to the fitting line, showing an almost perfect example of a linear regression. The positive correlation is explained by a high reflectance of healthy plants in NIR. As can be observed from the scatter plot (Figure 30), higher NIR value equals higher yield. Since the entire field is now covered with vegetation (see Figure 23), it is possible to determine weak spots that will more likely produce less yield. These spots will have a smaller reflectance in NIR.

The trend of positive high correlation between NIR band and yield remains strong until the end of the growing season. In October, the correlation is still strong, however, less strong than

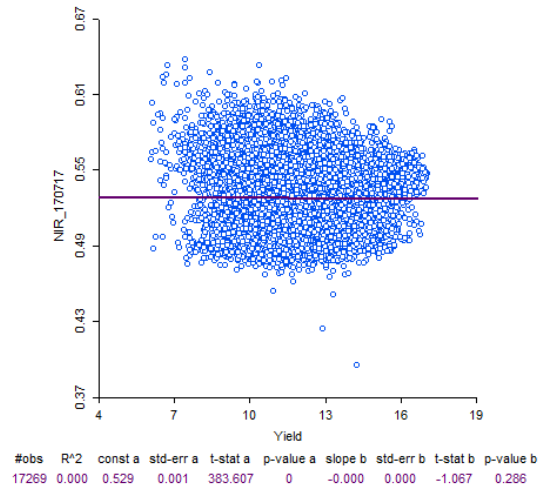


Figure 28: Scatter plot of yield (in t/ha) versus NIR values on July, 17th 2017

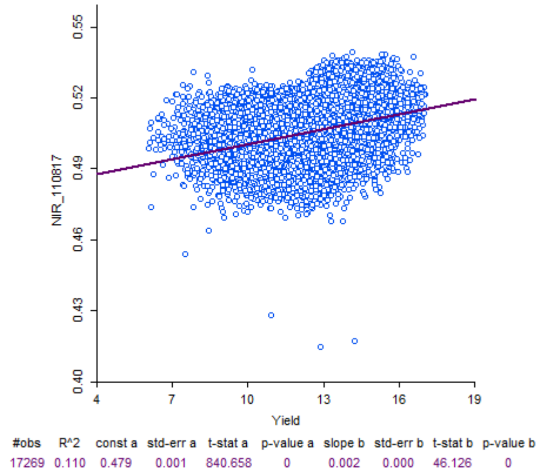


Figure 29: Scatter plot of yield (in t/ha) versus NIR values on August, 11th 2017

in the previous month - with the R^2 of 0.5 in October compared to 0.722 in September. When comparing the scatter plot of October (Figure 31) to the one from September (Figure 30), not a lot has changed. However, the points on the scatter plot from October, especially the ones located in the end of the graph, are distributed less close to the fitting line.

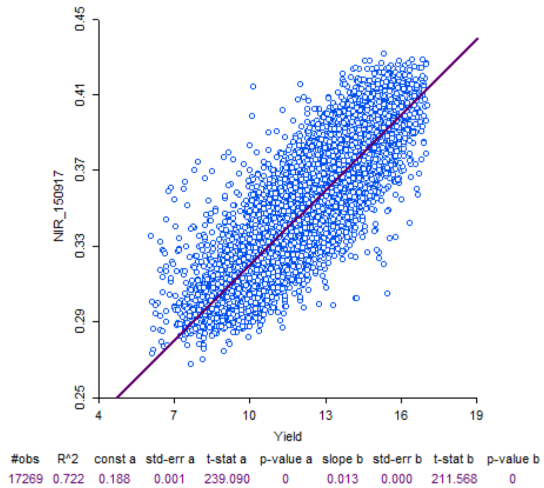


Figure 30: Scatter plot of yield (in t/ha) versus NIR values on September, 15th 2017

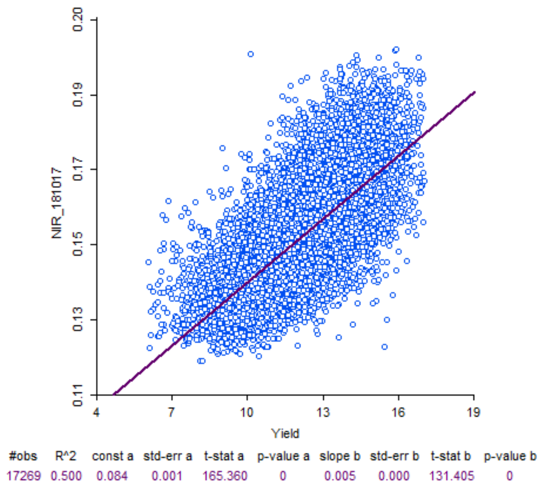
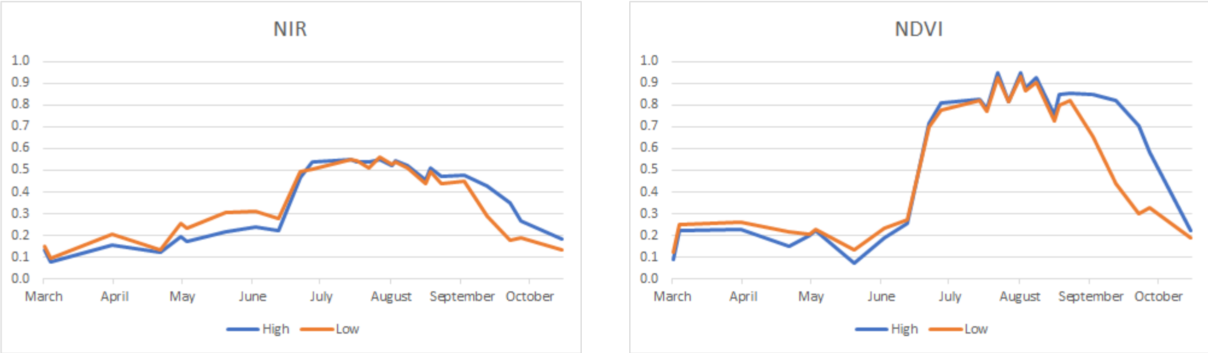


Figure 31: Scatter plot of yield (t/ha) versus NIR values on October, 18th 2017

4.9 Temporal profiles of high and low areas

From the section 2.6, it is known that it is possible to supervise the development of the plant through its phenological phases. This is usually done with Normalized Difference Vegetation Index, as it is the most commonly used vegetation index that is believed to be the most efficient in the detection of differences in crop development. However, from the results obtained in previous chapters (see section 4.6), we learned that NDVI is not always the most optimal index to use for the detection of vegetation variability. In fact, Near Infrared band has been proven to



(a) NIR *(b) NDVI*

Figure 32: NIR and NDVI temporal profiles of high and low yield areas

be the most efficient in detecting the differences within field from space. In order to see how both methods perform over a growing season, two points from the field 6 were selected - one that is located in the low yield area, another one - in the high yield. For those points, a NIR and a NDVI temporal profile (Figure 32a and Figure 32b, respectively) were made to compare their performances.

In both graphs, it can be seen that the low yield areas have higher reflectance until approximately mid-summer. This is due to the high reflectance of soil with a low content of organic matter. After mid-summer, for both NIR and NDVI, high yield areas begin to show higher reflectance. The reason for this change is that the vegetation start to appear and part with a high OM is performing slightly better. This lasts until September, even though higher yield areas tend to show higher values in both graphs, the difference between them is minimum. The differences begin to be very distinguishable in late August for NDVI and beginning of September for NIR.

NDVI detects the differences in high and low yield areas better in comparison to NIR. Especially from mid-August until beginning of October the differences between NDVI values for high and low yield are quite considerable. By only looking at NDVI graphs it is possible to predict which areas will have lower or higher yield from mid-August.

Figure 32b (NDVI temporal profile for field 6) can be compared to the Figure 3 (an example of a NDVI temporal profile for a summer crop). On both Figures the same patterns can be observed – small reflectance in the beginning (because of bare soil and no vegetation), emergence, rapid growth, then it attains its maximum NDVI (around 25th of July in our case) and after that NDVI values start to decrease – the crop is getting mature and it approaches the end of a cycle.

It can be concluded that the differences in crop development within the same field are the best to be spotted with NDVI, even though NIR is capable of spotting them too. However, the differences in NIR values for high and low yield areas are less substantial. The beginning and mid growing season are not the most optimal moments for difference detection, as in the beginning lower yield areas have a higher reflectance and in the middle of the growing seasons there are almost no differences between high and low yield areas. Observations made during mid-August until beginning of October are the most likely to detect future yield variation.

4.10 Yield-limiting factors and satellite images

In the previous sections, some preliminary conclusions were made on what is influencing the yield variability map of field 6. From the visual analysis (section 4.5), it is known that yield is correlated with organic matter and that topography has an impact on yield as well (sections

4.4.1 and 4.4.2). In this section, the statistical relationship between possible yield-limiting factors such as soil pH, electrical conductivity, organic matter content and elevation and satellite imagery will be studied.

4.10.1 pH

Several opinions exist on the possibility of predicting pH values using remote sensing techniques - according to Baghdadi and Zribi (2016), it is not possible since pH does not show any particular spectral characteristics, while Ge et al. (2011) argue that it is possible (see chapter 2.2). In this section this possibility will be tested.

Statistical tests on the correlation between pH and satellite data did not produce any significant results with one exception (see Figure 33). On June the 5th, the R^2 attains 0.105 - which means that 10% of variation within pH of the soil can be detected with NDVI.

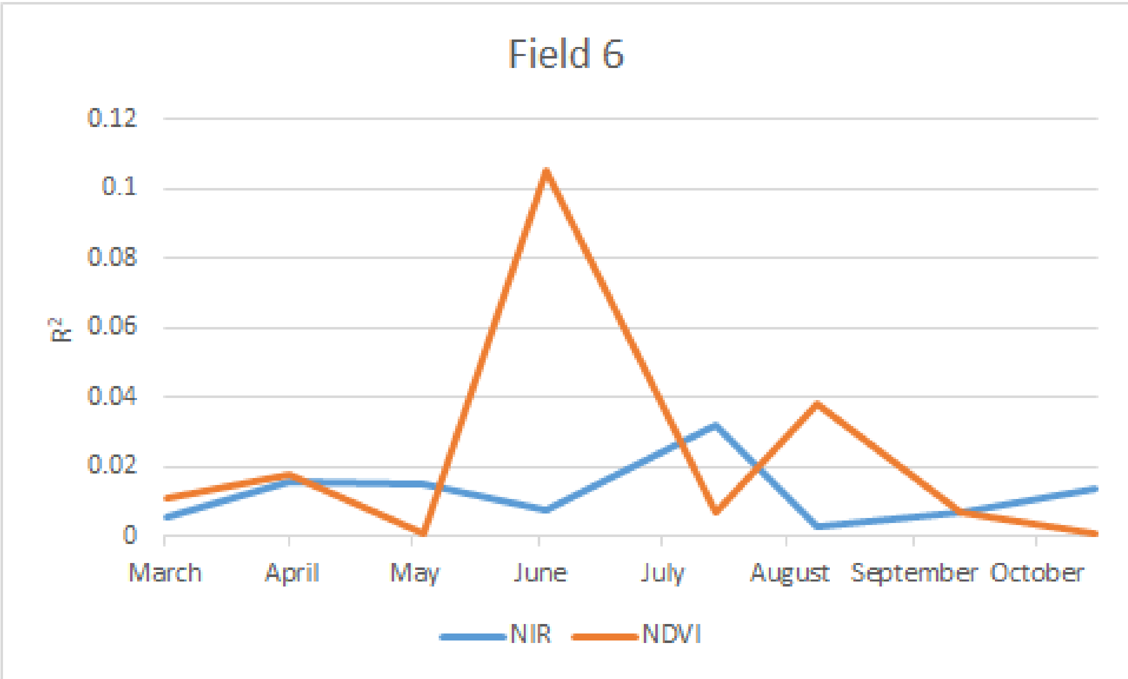


Figure 33: R^2 values of NIR and NDVI against soil pH

The correlation is low for the other days. Only in the month of July it slightly rises to the R^2 of 0.032 with NIR and in the month of August with NDVI of 0.038 (see Table 5). It is interesting to notice that here the situation is totally the opposite than for detection of yield variability through NIR and NDVI - the months of July and August showed the worst results. The results for ARVI are not shown and discussed here because of their low R^2 - smaller than in NDVI.

There are several possible explanations from the low correlation.

1. The very low spatial resolution of pH data does not allow to determine the relationship between pH and satellite imagery.

2. NIR, NDVI and ARVI are not the most optimal ways for this task.

3. As mentioned in chapter 2.2, remote sensing might not be suitable for pH detection.

Table 5: R^2 values of NIR band and NDVI with pH

Date	NIR	NDVI
4 March 2017	0.006	0.011
3 April 2017	0.016	0.018
6 May 2017	0.015	0.001
5 June 2017	0.008	0.105
17 July 2017	0.032	0.007
11 August 2017	0.003	0.038
15 September 2017	0.007	0.007
18 October	0.014	0.001

To conclude, it can be said that remote sensing is capable of detecting the differences in soil pH that in their turn will act as yield-limiting factors. However, the level of this detection is not enough - a 10% detection of total pH variation that was achieved in June with NDVI is relatively a small percentage.

4.10.2 Electrical conductivity

The tests were performed on both shallow and deep measurements of EC. However, in this section, only the results for shallow measurements are discussed, as the correlation between satellite imagery and deep EC (for the results on deep EC, see annex F).

Statistical tests performed on electrical conductivity and NIR and NDVI have shown on average slightly better results than on pH, but they are not groundbreaking (Annex F). It is important to mention that ARVI is overcoming NIR and NDVI during some months, which was not the case for other tests (see Figure 34).

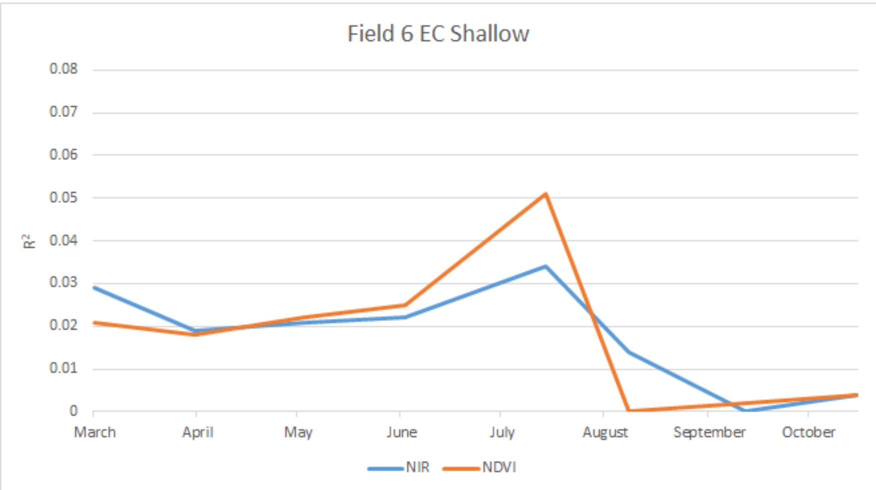


Figure 34: R^2 values of NIR and NDVI against EC

From the beginning until the middle of the season, R^2 varies between 0.018 and 0.034 for both NIR and NDVI. The difference with ARVI is that it is able to detect the variation one month earlier than NIR and NDVI. The maximum R^2 - 0.051 is achieved in June with ARVI and a month later with NDVI and is also 0.051. After the month of August, there is almost no correlation at all amongst vegetation indices and NIR with electrical conductivity.

To conclude, mid-summer is the most optimal moment to depict variations in EC with satellite derived data. It is hard to say which of the vegetation indices or spectral band should be used, as they all have different peaks of best performances. One of the reasons why overall EC is correlated better with satellite imagery than pH is the much higher spatial resolution of EC data. A much higher correlation was expected for EC since the spatial resolution of this data set is twice as high in comparison to pH, however, it was not the case.

4.10.3 Organic matter

The correlation between organic matter and satellite imagery is significant - during the first month of observations, the correlation reaches 34.2% in NDVI and 29.6% in NIR (Table 6). What can be observed on satellite images from section 4.5 is translated statistically in the Table 6: from March until June the differences in soil types are clearly visible on satellite images and R^2 values in NIR are high for those months. From July onwards, the correlation drops significantly for NIR, however the situation is different for NDVI.

Table 6: R^2 values of NIR band and NDVI with OM

Date	NIR	NDVI
4 March 2017	0.296	0.342
3 April 2017	0.368	0.112
6 May 2017	0.361	0.053
5 June 2017	0.346	0.156
17 July 2017	0.005	0.046
11 August 2017	0.037	0.231
15 September 2017	0.066	0.115
18 October	0.099	0.018

As can be seen in Figure 35, NDVI detects better the variation on organic matter on the first month of observations, in March. After March, its performance is lower when compared to NIR. However, the situation changes from July until September. The difference is especially substantial during the month of August - while NDVI can detect 23% of variation, NIR is barely detecting 4%. From October onwards NIR is leading again.

Out of three soil properties used in this study as possible yield-limiting factors, organic matter is showing the best correlation with satellite imagery. It also shows the best correlation with the final yield - 11% of yield variability can be explained with variations in OM (see Figure 36). What can also be observed in Figure 36 is that the relationship between OM and yield is negative. The reason for that is that organic matter data is not stored in absolute values, but as ratio of Red and Infrared, in which, areas with the high organic matter content have a smaller ratio and vice versa. This is the reason why the relationship between yield and organic matter is negative - because a smaller OM ratio actually means higher OM absolute values.

The same reasoning can be applied while studying the scatter plots of organic matter versus NIR and NDVI. In the beginning of the growing season, the correlation between NIR and OM is positive - mostly because dark soil has lower reflectance and this dark soil has smaller OM ratio values (see Figure 37a). In August, it changes and now the relationship is negative, when it should have been positive if the values for OM were stored as absolute. In Figure 37b, it can be seen that small OM values have a higher NDVI. Since in August the field is covered with vegetation, high NDVI represents healthier vegetation.

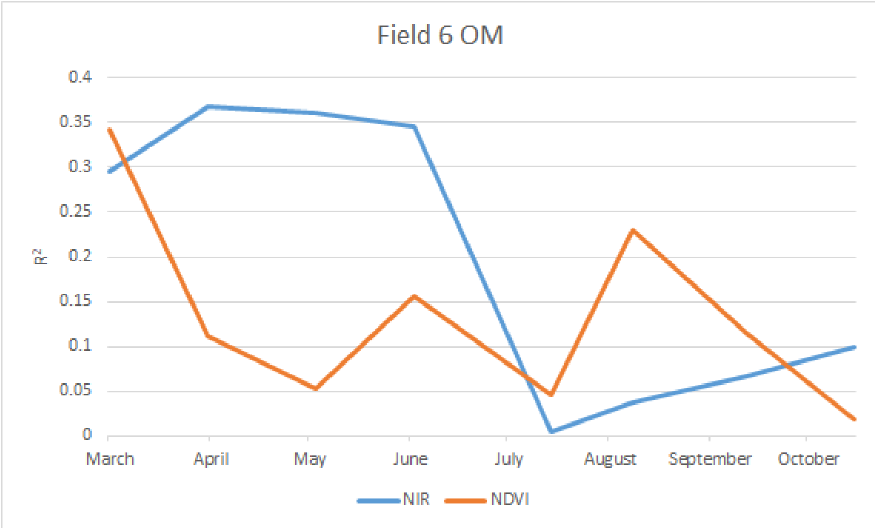
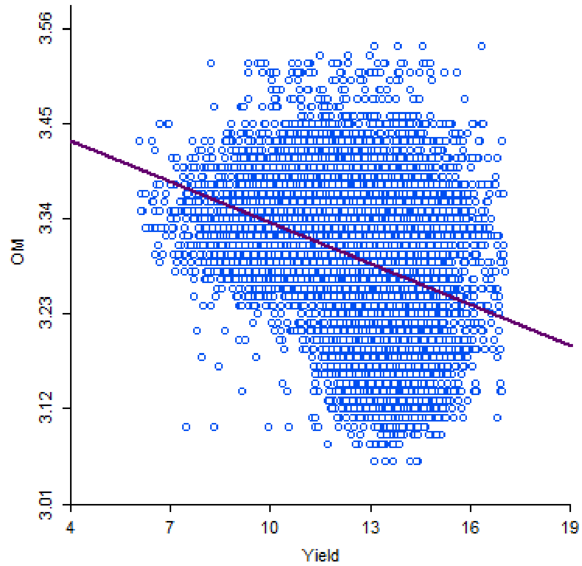


Figure 35: R^2 values of NIR and NDVI against OM



#obs	R^2	const a	std-err a	t-stat a	p-value a	slope b	std-err b	t-stat b	p-value b
17173	0.114	3.494	0.004	826.717	0	-0.016	0.000	-47.092	0

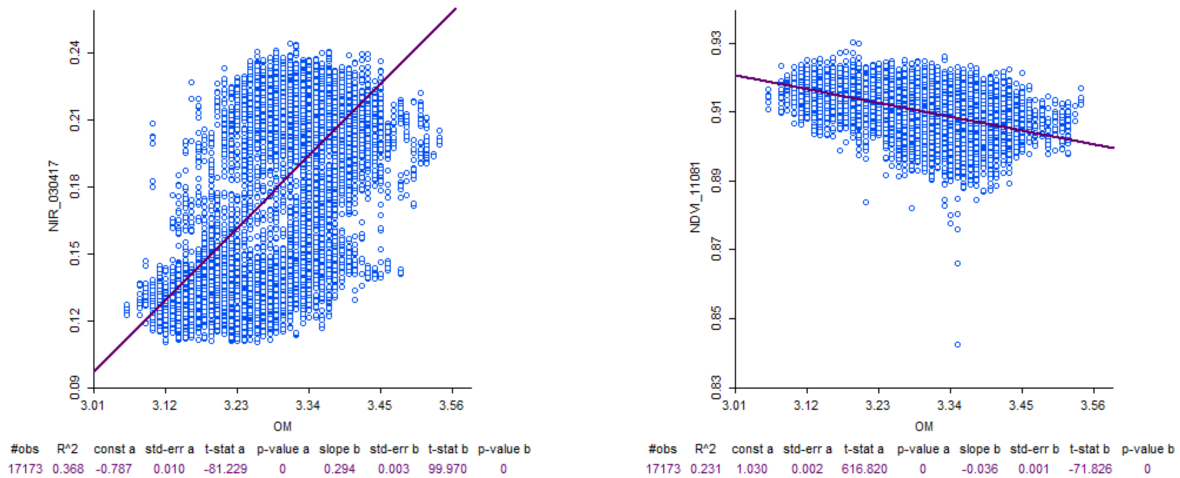
Figure 36: Scatteplot of yield against organic matter

Having such a high coefficient of determination and considering the big amount of data observations, it can be concluded that OM can be successfully monitored from space. Organic matter does influence the yield and this influence can be predicted. The most optimal way of predicting is to use NIR band and to use satellite images that represent bare soil.

4.10.4 Elevation

At first, it can seem that there is no immediate relation between elevation and yield. However, as many researches have proved, elevation does have an impact on yield (2.4.2). Yet, proving this relationship for field 6 is not that straightforward. As can be seen in Figure 38, R^2 is equal to 0, which means that there is no correlation between elevation and yield in field 6.

However, one of the possible explanations of why there is no correlation might be in the type



(a) OM against NIR in April *(b) OM against NDVI in August*

Figure 37: Comparison of scatter plots OM against NIR and NDVI

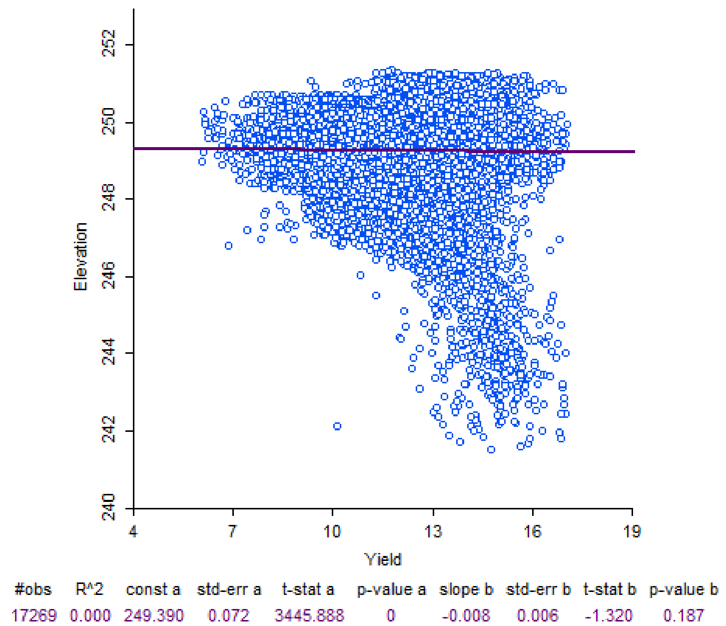


Figure 38: Scatter plot of yield against elevation

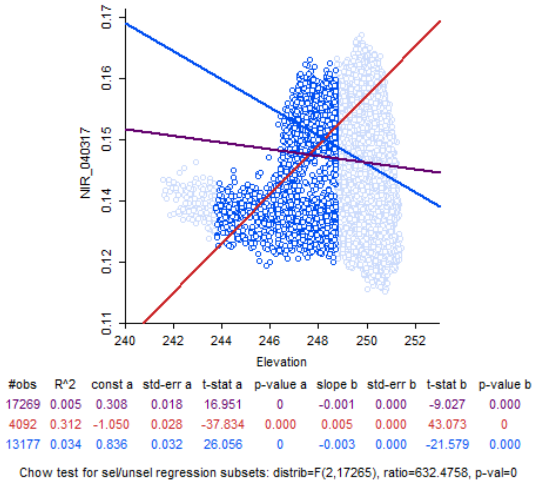
of the regression analysis used. As can be seen in the Figure 38, the points are not distributed linearly. The way the points are distributed suggests that there is no linear relationship between yield and elevation. Therefore, other type(s) of regression analysis should be used to detect this relationship.

The low correlation is also found between elevation and satellite imagery. As can be seen from Table 7, the highest R^2 value reaches only 0.049, which means that only 4.9% of variation within elevation can be explained with satellite imagery. In those statistical tests linear regression analysis is used, as other regression types are not the scope of this study.

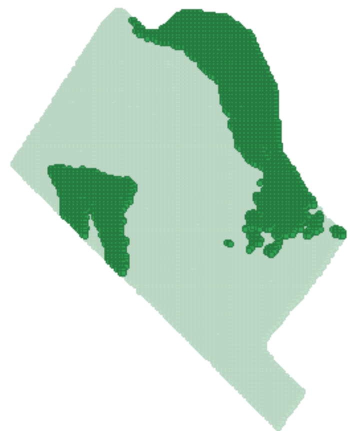
Another proof of a wrong choice of linear regression can be seen in Figure 39. The points located in the middle of a scatter plot were selected and this selection revealed an interesting

Table 7: R^2 values of NIR band and NDVI against elevation

Date	NIR	NDVI
4 March 2017	0.005	0.014
3 April 2017	0.011	0.023
6 May 2017	0.01	0.003
5 June 2017	0.004	0.048
17 July 2017	0.001	0.007
11 August 2017	0.028	0.049
15 September 2017	0.023	0.011
18 October	0.004	0.013



(a) Scatter plot points selection



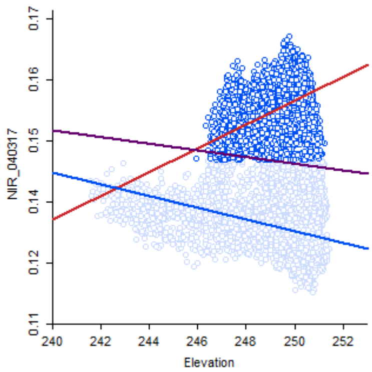
(b) Location of points on the map

Figure 39: Differences in R^2 values in low areas detected through NIR in March

fact - R^2 raised from 0.005 to 0.312 for the selected points. As can be seen from the Figure 39b, the selected points represent the low areas of the field 6.

A similar pattern can be observed for higher areas of field 6 (see Figure 40b). The R^2 for those points is lower - only 0.202 (see Figure 40a). However, in comparison to a total R^2 of all points, the increase is impressive - it is 40 times higher.

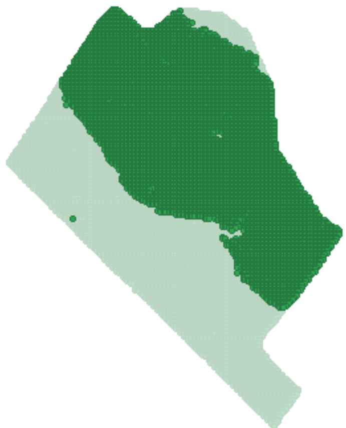
To conclude, elevation can be successfully detected using remote sensing techniques, however it should be done with precaution. Linear regression analysis should not be used, as the nature of relationship between elevation and yield and satellite imagery is not linear. In some cases, linear fit can used. For example, in low areas of field 6, NIR is able to detect more than 30% of variation in elevation from bare soil.



#obs	R^2	const a	std-err a	t-stat a	p-value a	slope b	std-err b	t-stat b	p-value b
17269	0.005	0.308	0.018	16.951	0	-0.001	0.000	-9.027	0.000
9750	0.202	-0.434	0.012	-36.685	0.000	0.002	0.000	49.712	0
7519	0.148	0.423	0.008	52.348	0	-0.001	0.000	-36.154	0.000

Chow test for sel/unsel regression subsets: distrib=F(2,17265), ratio=61853.9350, p-val=0

(a) Scatter plot points selection



(b) Location of points on the map

Figure 40: Differences in R^2 values in high areas detected through NIR in March

5 Validation

In section 4, the results of this study are described in detail. However, as mentioned in the beginning, all the tests were performed on one single field. The fundamental base of precision agriculture is the knowledge that no fields are alike, and even within the same field, there are a lot of variations. The chapter aims to test whether the knowledge obtained in section 4 can be reproduced in other fields and possibly even farms. A similar analysis will be performed on another field (field 10), and the results for both field 6 and 10 will be compared.

5.1 Yield variability in field 10

The first major difference between field 6 and field 10 is their differences in the final yield. While yield for field 6 varies a lot - between 5 to 20 t/ha on average, field 10 has a variation of 9 to 14 t/ha (excluding highest values, see Annex A). Yield in field 6 is less evenly distributed, with several high-yield zones and a few low-yield ones. Field 10 does not have so much variety and is more homogeneously distributed.

From Figure 41, it is possible to make some comparisons with field 6. Field 10 also has a clear impact of geographical features on its final yield, and it can also be divided into zones with higher and lower yield (lower and upper part, respectively). In the next sections, those impacts will be discussed in more detail.

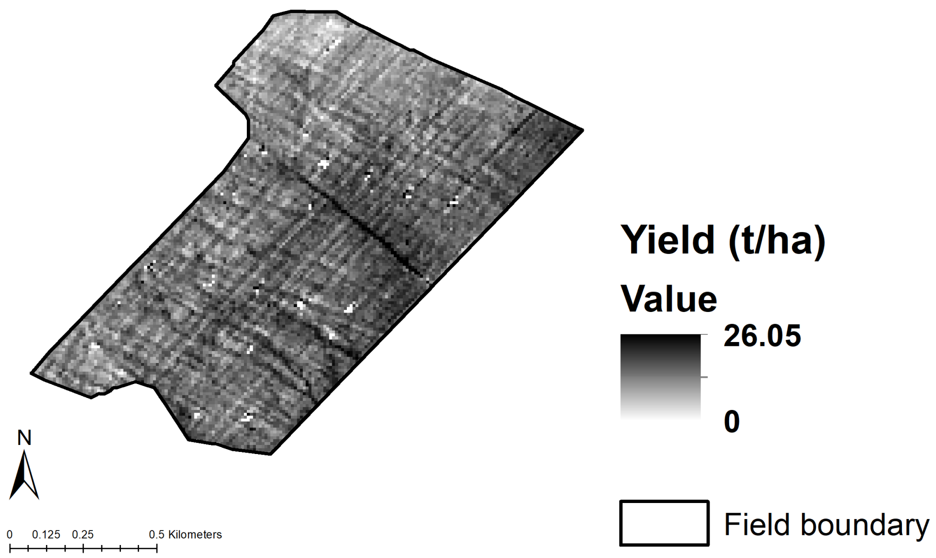


Figure 41: Map of yield variability for field 10

5.1.1 Soil variability

Another big difference between field 6 and 10 is soil. For field 6, it was determined that there are big differences in soil types and properties, especially in organic matter content. For field 10, it was difficult to determine soil types - the bare soil image of field 10 looks homogeneous (see Figure 42). From the results described in section 4.10.3, it is known that variations in organic matter can be detected with remote sensing. It is hard to spot soil variations in field 10 (Figure 42) and the most probable reason why is the smaller variation of organic matter content in field 10.

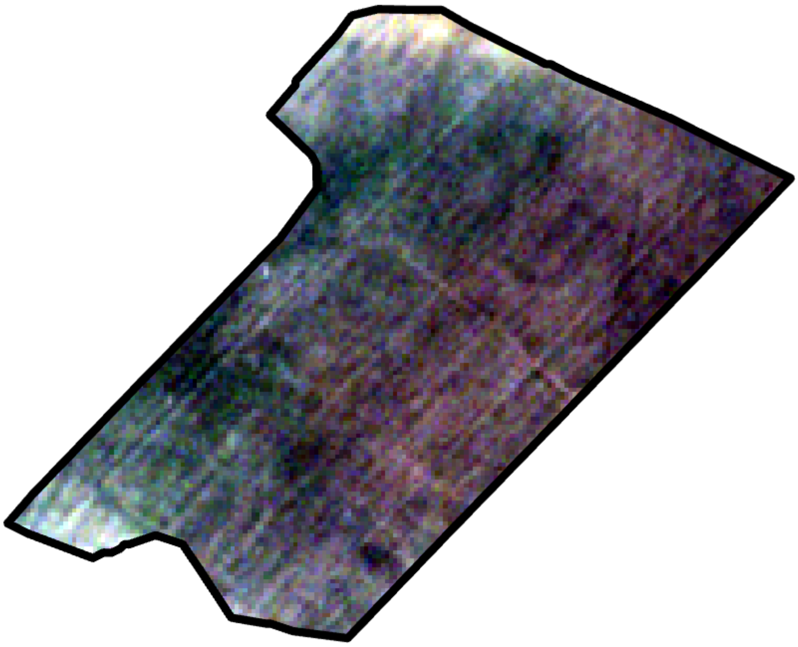


Figure 42: Satellite image of field 10 taken on the 4th of March 2017 shown in true colour

This hypothesis is partly proven by the OM soil map for field 10 (Figure 43) and by the main agronomist - he indicated that there are no considerable variations in OM content in field 10. The reason for partial proof is that the OM map is made using relative values instead of absolute.

5.1.2 Topography

A third, major difference between field 6 and 10 is elevation. While field 6 has only 10 meters of differences in altitude, field 10 has 18 - almost twice as much (Figure 44). As mentioned in section 4.4.2, a small change in attitude can provoke big changes in yield. Similarly to field 6, the impact of topography on yield can be clearly seen in field 10. A long white line that can be seen on the elevation map of field 10 is the centre of a slope, a place where water flows. It is also the area with a higher yield. This is another proof that yield is (at least visually) highly correlated with elevation. In section 5.4, this correlation will be studied using statistical methods.

5.2 VI and time

In comparison to field 6, the correlation between yield and satellite imagery in field 10 is low (see Figure 45). On average, in field 10, NIR has detected 13% of variation (versus 27.8% in field 6), NDVI 9.2% (versus 19.1%) and ARVI - 3.9% (versus 14.1%). This is a relatively big difference, considering the fact that the maximum correlation between yield and satellite images in field 6 was 72%, in field 10 it reaches only 36.6% (see Table 8).

From the Figure 45, it is clear that there is no mid-summer drop in correlation in field 10 as it is in field 6 (except ARVI). The correlation between bare soil images and yield is low, but

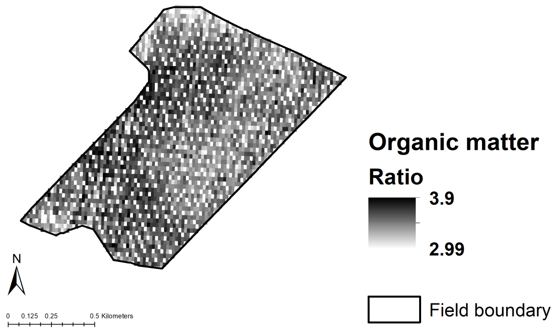


Figure 43: Map of organic matter content for field 10

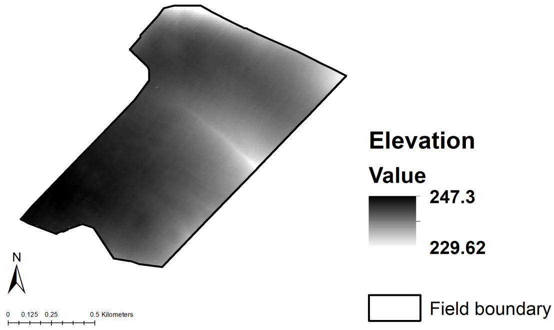


Figure 44: Elevation map of field 10

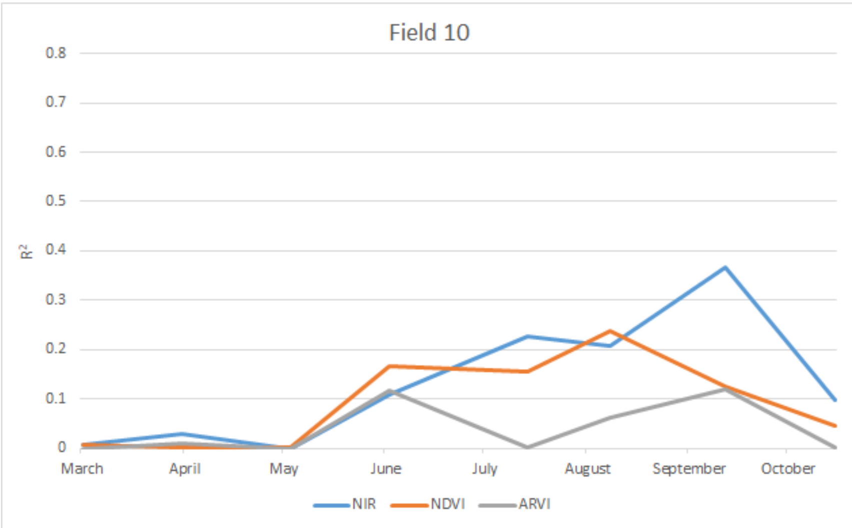


Figure 45: R^2 of NIR, NDVI, ARVI during the growing season for field 10

it starts growing when vegetation starts to appear, unlike in field 6, where correlation is high when there is no vegetation yet.

Table 8: R^2 values of NIR band and vegetation indices with yield for field 10

Date	**NIR**	**NDVI**	**ARVI**
4 March 2017	**0.007**	0.006	0.à
3 April 2017	**0.03**	0.002	0.009
6 May 2017	0	**0.001**	0
5 June 2017	0.108	**0.167**	0.117
17 July 2017	**0.227**	**0.156**	0.003
11 August 2017	0.207	**0.238**	0.063
15 September 2017	**0.366**	0.125	0.121
18 October	**0.097**	0.046	0.001

Despite the differences in the amount of correlation and peak performances, the results for field 6 and 10 are quite comparable regarding the following aspects:.

- Both fields show the best results in NIR.
- ARVI is not able to detect the correlation in those fields.

- The correlation reaches its maximum in September.

5.3 NIR/NDVI temporal profiles

As described in section 5.1, the differences in yield in field 10 are not as big as they are in field 6. Therefore, a decision was made to not make NIR or NDVI profiles for field 10 as it is believed not to show a lot of variation.

5.4 Yield-limiting factors

There are a lot of differences between field 6 and 10 when it comes to yield-limiting factors and their detection with satellite imagery. The differences can be seen in the amount of correlation, methods and timing. It was assumed that the same yield-limiting factors would have the highest correlation with satellite imagery, but this assumption was wrong. Organic matter was the YLF that showed the best results for detection in field 6. The highest R^2 for field 6 is 0.368; however, in field 10, the maximum R^2 attained is 0.172 - twice as small (see Annex G). The reason for this was discussed in section 5.1.1 - it is because of the smaller variation of organic matter content in field 10.

On the other hand, elevation showed better possibilities for detection in field 10. In April, the correlation between elevation and NDVI reached almost 32% (see Annex J). The linear regression issue is still relevant for field 10, and it is assumed that it is possible to attain a higher coefficient of determination using other regression analysis.

While EC of soil produced more or less the same results for both fields, soil pH showed a higher correlation in field 10. With the maximum R^2 of 0.127, it overpowered the best result of field 6 by four times (R^2 = 0.038) (Annex H).

The most optimal vegetation indices or spectral bands that should be used for YLF detection differs per yield-limiting factor. Elevation is the only exception - NDVI has the highest correlation in both fields. The most optimal timing also differs: the best dates for field 6 in some cases are the least optimal for field 10. For more detailed description of R^2 values of yield-limiting factors versus satellite imagery refer to Annexes G, H, I and J.

6 Discussion

This study has produced a variety of results; some of them were expected as they were mentioned in the reviewed literature, some of them were not. In this chapter, the results of this research will be discussed and compared to other similar studies. Also, difficulties and issues that occurred during the analysis will be mentioned.

6.1 NIR, NDVI, ARVI

6.1.1 NIR vs NDVI

While testing the detection of yield variability through optical satellite imagery, it was concluded that NIR is the most optimal method for that. However, in practice and theory, NIR is not as widely used as, for example, NDVI is. One of the reasons why might be in the easy definition and application of Normalized Difference Vegetation Index - healthier vegetation has a higher NDVI value. This is easy to understand and to interpret and gives farmers a tool to supervise their crops without immersing themselves into the difficulties of remote sensing world. NDVI can be still used for detection and prediction of yield variability, but considering the outcomes of this research, it is highly recommended to use NIR instead as it produces better results. In the next paragraphs, some assumptions will be made on why NIR performs better.

There are several explanations why NIR has the highest correlation with the yield. Campbell and Wynne (2011) have mentioned that NIR radiation can transmit the light through up to eight leaf layers. This alone gives a lot more information about vegetation differences below the top of the canopy, which might be the explanation why NIR correlates the most with the yield. The other possible reason for high correlation in NIR band is that Near Infrared light is reflecting off the soil. The soil has a lower reflectance in the NIR band, therefore it is easier to identify differences in soil properties, and as was concluded earlier, soil properties have a great influence on yield.

The soil reflectance in NIR can also explain the 20% correlation with yield during the beginning of the growing season when the field does not have any vegetation yet, only bare soil. Other indices did not demonstrate such a high correlation.

Campbell and Wynne (2011) also argue that NIR is able to detect plants exposed to stress (f.e. diseases, drought) better than other vegetation indices or spectral bands. On the other hand, they also mention the disadvantage of NIR. According to the authors, Near Infrared is not the most optimal in the detection of the differences in corn once tassels has formed, since tassels conceal green vegetation too much. However, this is not found in our study - from three bands/VI available, NIR outperformed all others even after tasseling stage.

Since a lot of researches are focusing on the relationship between yield and NDVI data, it was already proven that NDVI could have a high correlation with yield. In this study, only two images out of eight have shown the best correlation with NDVI. In comparison to NIR, NDVI has around 9% less correlation on average on all observations. NDVI has the potential to detect 70% of final yield. This is, however, the only high correlation number for this index found in this study.

6.1.2 Underperformance of ARVI

In this study, ARVI did not show remarkable results when it comes to yield detection. On only one occasion it correlated better than the rest, and it is during mid-summer. From this, it can be concluded that ARVI has not achieved its primary goal - improving the results of NDVI by correcting the aerosol effects. However, other research suggest that ARVI is indeed capable of showing better results than NDVI (Liu et al., 2004). Therefore, more research is needed to determine whether ARVI can be used for detection of yield variability.

6.2 Yield prediction and timing

Yield prediction using remote sensing techniques was one of the core questions of this research. NIR showed exceptional results - a bare soil image made on the 4th of March before any seeding took place correlates with the final yield on 19%. This means that around 20% in yield variation can be foreseen one and a half month before planting by just looking at the bare soil satellite image. The highest R^2 is achieved in September, when both NIR and NDVI produce the R^2 of around 0.7. Armed with this information, the farmer has two whole months before the harvesting to adapt his managerial practices in order to estimate and optimise the future yield.

Previous researches have confirmed that it is indeed possible to obtain a strong relationship between crop yield and satellite imagery. Wiegand et al. (1986) confirmed that the best moment for detection of this relationship is during the grain filling stage (R2-R6 stages, see section 2.6). Similar research realised by Tucker et al. (1980) conducted on wheat found a R^2 of 0.64 correlation between yield and NDVI 5 weeks before the harvesting. Another research on wheat and NDVI, claims that it is possible to predict yield 30 days before harvest (Lopresti et al., 2015).

Returning to research on corn crops, Teal et al. (2006) found that the strongest relationship between NDVI and yield can be found in V8-V9 stages ($R^2 = 0.77$) and that all the measurements after the V9 stages failed to distinguish the variation because of the canopy closure. Gao et al. (2018) claim that *"the best period for yield prediction for corn was during the middle of the growing season from day 192 to 236 (early July to late August, 1–3 months before harvest)"*. Other similar researches found aR^2 of 0.5 between NIR and yield (Campbell and Wynne, 2011). As can be concluded, there is no unique answer to when is the most optimal moment for yield prediction. This proves what was also discovered in this study. In addition, it is also a crop dependent.

This research only focused on NIR, NDVI and ARVI. Therefore, it would be of interest to conduct similar research on other vegetation indices or spectral bands. For example, Shanahan et al. (2001) suggest using Green Normalized Difference Vegetation Index (GNDVI) as it can be highly correlated with grain yield - up to 0.92.

Another small drawback of this study is the absence of satellite data close to the harvesting date. The last image available was made one month before the harvest, and it excluded the possibility to study the correlation of yield and satellite data right before the harvesting.

6.3 Yield-limiting factors

A variety of yield-limiting factors were tested for correlation with optical satellite imagery. Some of them showed potential for detection, while others did not. Organic matter content was found to be the most detectable with NIR. As described previously, soil reflects NIR and since there was a lot of variations in OM in field 6, up to 36% of variaton was detected with NIR.

From this, it can be concluded that fields with relatively substantial variation of organic matter will demonstrate a higher correlation with NIR, especially in the beginning of the growing season. The same can not be said about other soil properties.

pH and electrical conductivity showed a relatively small ability of detection with remote sensing. Here, the quality of soil data should be discussed. It is very likely that the quality of soil data impedes the study of the soil properties with remote sensing. It is advisable to not draw conclusions about the possibilities of detection of pH and EC with remote sensing from this study. It is recommended to repeat the study with a data set of soil properties of higher quality, especially because other authors have found a relationship between spectral bands and EC. Shanahan et al. (2001) found that Mid-Infrared, together with Near Infrared are the most correlated with shallow electrical conductivity values (which was not the case in this study).

This study has used simple NIR and NDVI values to test the possibility of determination of pH. The maximum found R^2 is relatively small - 0.105. However, it is likely that NIR can still be used for that purpose but in a slightly different manner. Tekin et al. (2013) suggest to use On-Line Visible and Near Infrared Spectroscopy for predicting pH values. With this method, they managed to obtain high accuracy detection (R^2 = 0.85). Their study was made using ground machinery, which is not the scope of our research. It should be further investigated if their findings can be reproduced with optical satellites.

Elevation at first did not produce any consistent results. It was determined that the possible explanation was in the type of regression used. Several studies (Whetton et al. (2017), Cheng et al. (2017)) have already addressed this issue, but since non-linear regression was not the scope of this study, their results were not implemented in this work. It can be assumed that some YLF would show a higher correlation if they are tested for a non-linear correlation.

It should be noted that it was not possible to obtain weather data, while multiple authors (Below (2008), Gentry et al. (2013)) named it as the most important yield-limiting factor. For further research, it would be interesting to investigate the possibilities that weather data can offer.

6.4 Re-productivity of the results

The comparison of the results in two fields revealed a few points for discussion. While overall, the results were quite different, they had some common points.

The most important reason influencing the results is the big variation in soil OM in field 6. Field 10 is more homogeneous, that is why NIR failed to detect this variation. Soil pH and EC were better detected in field 10, which brings us to the conclusion that all the results are field-dependent.

The same is confirmed by other studies. McConkey et al. (2004) have determined in their research that date of highest correlation between NDVI and yield was not consistent. The same principle applies for yield-limiting factors - in some cases NIR has performed better, in some cases NDVI showed the best results. In two test fields, the moments when the highest correlation can be found are different. It can be suggested to supervise the situation on the field level and using several years of observations. This research has only used the data from 2017 growing season, and it is not possible to say whether the same patterns were detected in previous or will be repeated in the next growing seasons. Therefore, a multiyear research based on one single field is suggested to produce more accurate results.

In order to help the farmer in making the most out of remote sensing observations, it is advised to focus on one single field and one single YLF. Knowing beforehand which soil properties have the most variations in this precise field, will help him determine the right method to use. It will also help to select the right tool (VI or spectral band) and the most optimal moment.

7 Conclusions

In this concluding chapter, the main research questions are answered on the basis of the results obtained from the analysis chapters.

7.1 To what extent can remote sensing be used for detection of spatial yield variability?

Remote sensing has produced some outstanding results when it comes to detection of yield variability. Both visual and statistical analysis allow quite accurate predictions of the final yield.

During some dates, satellite images visually correlate well with the yield variability map, however on some dates, this correlation is very low. Early season images showing the bare soil (from March until June) can give a good indication for the future yield differences and will help identify areas with possible lower yield because of organic matter content. Mid-summer images are not able to express the potential differences in the final yield. The late-season images (September and October) are looking almost identical to the final yield and can be safely used for the final yield estimation.

7.2 Which vegetation indices and/or spectral bands are the most optimal for detecting yield variability?

Near Infrared showed the highest correlation between satellite imagery and yield. It is able to detect up to 72% of variation. Out of eight days of observation throughout the growing season, NIR has the highest correlation in five. It is then followed by NDVI, which performed slightly better in field 10 than in field 6. ARVI did not show any remarkable results and is considered to be not suitable for detection of yield variability.

7.3 Which time frame is most suitable for identifying yield variability?

Statistically, the best moment to predict corn yield is September. It is able to detect up to 72% variation in yield variability in all three tested vegetation indices and spectral band - NIR, NDVI and ARVI. Observations from October can also be used as they are able to predict up to 50%.

Another moment that can provide useful insights into the future yield variability is the beginning of the growing season. NIR is especially capable of detecting around 22% of yield variability from the bare soil images.

7.4 Which factors can be identified as yield-limiting factors and show potential for detection in remote sensing data?

The yield-limiting factors that can be identified with remote sensing data differ per field. For field 6, the most detected YLF was organic matter. The most optimal way of detecting the organic matter variation is considered to be NIR in the beginning of the growing season ($R^2 \approx 0.34$). Elevation also could be detected; however, a special approach is required.

Field 10 showed a great potential for detection of elevation and of soil pH. 30% of variation in elevation can be explained with NDVI measurements from April. Variation in soil pH in field 10 was the highest in August (R^2 = 0.127 for NIR and R^2 = 0.108 for NDVI). Remote sensing showed only a slight ability for detection of electrical conductivity (maximum R^2 = 0.051 (NDVI) for field 6 and R^2 = 0.068 (NDVI) for field 10).

To conclude, remote sensing shows strong potential to detect yield variability. NIR has been proven to show the greatest ability to detect this variation (R^2 = 0.72), and its peak performance is two months before the harvesting date. Several yield-limiting factors can be successfully monitored with optical imagery. However, the methods and timing for this detection are not consistent. Organic matter content of the soil is the most correlated with satellite imagery YLF (R^2 = 0.368), it is then followed by elevation. pH and electrical conductivity can also be monitored to a certain extent, but require a different approach.

References

Aires, A. Conventional and organic farming—does organic farming benefit plant composition, phenolic diversity and antioxidant properties? In *Organic Farming-A Promising Way of Food Production*. InTech, 2016.

Baghdadi, N. and Zribi, M. *Land Surface Remote Sensing in Agriculture and Forest*. Elsevier, 2016.

Batini, C., Cappiello, C., Francalanci, C., and Maurino, A. Methodologies for data quality assessment and improvement. *ACM computing surveys (CSUR)*, 41(3):16, 2009.

Below, F. E. The seven wonders of the corn yield world. In *2008 Illinois Crop Protection Technology Conference*, page 86. Citeseer, 2008.

Campbell, J. B. and Wynne, R. H. *Introduction to remote sensing*. Guilford Press, 2011.

Carter, H. and Dean, G. Income, price, and yield variability for principal california crops and cropping systems. *California Agriculture*, 30(6):175–218, 1960.

Changere, A. and Lal, R. Slope position and erosional effects on soil properties and corn production on a miamian soil in central ohio. *Journal of Sustainable Agriculture*, 11(1): 5–21, 1997.

Cheng, L., SUN, B.-c., TANG, H.-j., WANG, T.-y., Yu, L., ZHANG, D.-f., XIE, X.-q., SHI, Y.-s., SONG, Y.-c., YANG, X.-h., et al. Simple nonlinear model for the relationship between maize yield and cumulative water amount. *Journal of integrative agriculture*, 16(4):858–866, 2017.

Chrisman, N., Devillers, R., Jeansoulin, R., Fisher, P., Comber, A., Wadsworth, R., Riazanoff, S., Santer, R., Lowell, K., van Oort, P., et al. *Fundamentals of spatial data quality*. Wiley, 2006.

Chuvieco, E. *Fundamentals of satellite remote sensing: An environmental approach*. CRC press, 2016.

Clark, M. S., Horwath, W. R., Shennan, C., Scow, K. M., Lantni, W. T., and Ferris, H. Nitrogen, weeds and water as yield-limiting factors in conventional, low-input, and organic tomato systems. *Agriculture, Ecosystems & Environment*, 73(3):257–270, 1999.

Clevers, J. A simplified approach for yield prediction of sugar beet based on optical remote sensing data. *Remote Sensing of Environment*, 61(2):221 – 228, 1997.

Courault, D., Demarez, V., Guérif, M., Le Page, M., Simonneaux, V., Ferrant, S., and Veloso, A. Contribution of remote sensing for crop and water monitoring. In *Land Surface Remote Sensing in Agriculture and Forest*, pages 113–177. Elsevier, 2017.

Dampney, P. and Moore, M. Precision agriculture in england: current practice and research-based advice to farmers. *Precision Agriculture*, pages 661–673, 1999.

Delp, R. Ideal climate soil for corn growth. *Home Guides*, 2015. URL `https://homeguides.sfgate.com/ideal-climate-soil-corn-growth-37426.html`.

Demydenko, O. and Velychko, V. Humus state of chernozem at different ways of tillage in the agrosystems of the left-bank forest steppe of ukraine. *Agricultural science and practice*, 2 (3):61–76, 2015.

Downing, T. E. *Climate change and world food security*, volume 37. Springer Science & Business Media, 2013.

Dutta, R. Introduction to satellite derived vegetation indices. United Nations Economic and Social Commission for Asia and the Pacific, 2017.

Edreira, J. I. R., Mourtzinis, S., Conley, S. P., Roth, A. C., Ciampitti, I. A., Licht, M. A., Kandel, H., Kyveryga, P. M., Lindsey, L. E., Mueller, D. S., et al. Assessing causes of yield gaps in agricultural areas with diversity in climate and soils. *Agricultural and Forest Meteorology*, 247:170–180, 2017.

Forkuor, G., Hounkpatin, O. K., Welp, G., and Thiel, M. High resolution mapping of soil properties using remote sensing variables in south-western burkina faso: A comparison of machine learning and multiple linear regression models. *PloS one*, 12(1):e0170478, 2017.

Gao, F., Anderson, M., Daughtry, C., and Johnson, D. Assessing the variability of corn and soybean yields in central iowa using high spatiotemporal resolution multi-satellite imagery. *Remote Sensing*, 10(9):1489, 2018.

Ge, Y., Thomasson, J. A., and Sui, R. Remote sensing of soil properties in precision agriculture: A review. *Frontiers of Earth Science*, 5(3):229–238, 2011.

Gentry, L. F., Ruffo, M. L., and Below, F. E. Identifying factors controlling the continuous corn yield penalty. *Agronomy Journal*, 105(2):295–303, 2013.

Gomez, C. and Lagacherie, P. Mapping of primary soil properties using optical visible and near infrared (vis-nir) remote sensing. In *Land Surface Remote Sensing in Agriculture and Forest*, pages 1–35. Elsevier, 2017.

Grisso, R. D., Alley, M. M., Holshouser, D. L., and Thomason, W. E. Precision farming tools. soil electrical conductivity. *Virginia cooperative extension*, 2005a.

Grisso, R. D., Alley, M. M., Phillips, S. B., and McClellan, P. Interpreting yield maps:" i gotta yield map, now what?". *Virginia Cooperative Extension*, 2005b.

Gurgel, H. and Ferreira, N. Annual and interannual variability of ndvi in brazil and its connections with climate. *International Journal of Remote Sensing*, 24(18):3595–3609, 2003.

Heggenstaller, A. Yield-limiting factors in continuous corn production. *Crop Insights*, 2004. URL `https://www.pioneer.com/home/site/us/agronomy/library/yield-limiting-factors-continuous-corn/`.

Hijbeek, R., van Ittersum, M. K., ten Berge, H. F., Gort, G., Spiegel, H., and Whitmore, A. P. Do organic inputs matter–a meta-analysis of additional yield effects for arable crops in europe. *Plant and Soil*, 411(1-2):293–303, 2017.

Huete, A., Didan, K., Miura, T., Rodriguez, E., Gao, X., and Ferreira, L. Overview of the radiometric and biophysical performance of the modis vegetation indices. *Remote Sensing of Environment*, 83(1):195 – 213, 2002. ISSN 0034-4257. doi: https://doi.org/10.1016/

S0034-4257(02)00096-2. URL http://www.sciencedirect.com/science/article/pii/S0034425702000962. The Moderate Resolution Imaging Spectroradiometer (MODIS): a new generation of Land Surface Monitoring.

Huete, A., Liu, H., Batchily, K., and van Leeuwen, W. A comparison of vegetation indices over a global set of tm images for eos-modis. *Remote Sensing of Environment*, 59(3):440 – 451, 1997. ISSN 0034-4257. doi: https://doi.org/10.1016/S0034-4257(96)00112-5. URL http://www.sciencedirect.com/science/article/pii/S0034425796001125.

Jin, X., Ma, J., Wen, Z., and Song, K. Estimation of maize residue cover using landsat-8 oli image spectral information and textural features. *Remote Sensing*, 7(11):14559–14575, 2015.

Joernsgaard, B. and Halmoe, S. Intra-field yield variation over crops and years. *European Journal of Agronomy*, 19(1):23–33, 2003.

Kaufman, Y. J. and Tanre, D. Atmospherically resistant vegetation index (arvi) for eos-modis. *IEEE transactions on Geoscience and Remote Sensing*, 30(2):261–270, 1992.

Kennedy, K., Schroeder, T., Shaw, M., Haynes, D., Lewis, S., Bentley, C., Paxman, C., Carter, S., Brando, V. E., Bartkow, M., et al. Long term monitoring of photosystem ii herbicides–correlation with remotely sensed freshwater extent to monitor changes in the quality of water entering the great barrier reef, australia. *Marine Pollution Bulletin*, 65(4):292–305, 2012.

Kravchenko, A. N. and Bullock, D. G. Correlation of corn and soybean grain yield with topography and soil properties. *Agronomy Journal*, 92(1):75–83, 2000.

Labus, M., Nielsen, G., Lawrence, R., Engel, R., and Long, D. Wheat yield estimates using multi-temporal ndvi satellite imagery. *International Journal of Remote Sensing*, 23(20): 4169–4180, 2002.

Landsberg, J. J. and Gower, S. T. *Applications of physiological ecology to forest management*. Academic Press, 1997.

Liu, G.-R., Liang, C.-K., Kuo, T.-H., Lin, T.-H., et al. Comparison of the ndvi, arvi and afri vegetation index, along with their relations with the aod using spot 4 vegetation data. *Terrestrial, Atmospheric and Oceanic Sciences*, 15(1):15–31, 2004.

Lobell, D. B. The use of satellite data for crop yield gap analysis. *Field Crops Research*, 143: 56–64, 2013.

Loomis, R. S. and Williams, W. A. Maximum crop productivity: an extimate. *Crop Science*, 3 (1):67–72, 1963.

López-Granados, F., Jurado-Expósito, M., Pena-Barragan, J., and García-Torres, L. Using geostatistical and remote sensing approaches for mapping soil properties. *European Journal of Agronomy*, 23(3):279–289, 2005.

Lopresti, M. F., Di Bella, C. M., and Degioanni, A. J. Relationship between modis-ndvi data and wheat yield: A case study in northern buenos aires province, argentina. *Information processing in Agriculture*, 2(2):73–84, 2015.

Marty, J.-L., Mionetto, N., Noguer, T., Ortega, F., and Roux, C. Enzyme sensors for the detection of pesticides. *Biosensors and Bioelectronics*, 8(6):273–280, 1993.

McConkey, B., Basnyat, P., Lafond, G., Moulin, A., and Pelcat, Y. Optimal time for remote sensing to relate to crop grain yield on the canadian prairies. *Canadian Journal of Plant Science*, 84(1):97–103, 2004.

Melakeberhan, H. Embracing the emerging precision agriculture technologies for site-specific management of yield-limiting factors. *Journal of nematology*, 34(3):185, 2002.

Mueller, N. D., Gerber, J. S., Johnston, M., Ray, D. K., Ramankutty, N., and Foley, J. A. Closing yield gaps through nutrient and water management. *Nature*, 490(7419):254–257, 2012.

Nielsen, B. Understanding factors that limit yield in corn. University Lecture, 1998.

NRCS. Soil quality indicators: ph, 1998. URL `https://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/nrcs142p2_052208.pdf`.

Oliver, M., Bishop, T., and Marchant, B. *Precision agriculture for sustainability and environmental protection*. Routledge, 2013.

Panda, S. S., Ames, D. P., and Panigrahi, S. Application of vegetation indices for agricultural crop yield prediction using neural network techniques. *Remote Sensing*, 2(3):673–696, 2010.

Parida, B. R. Analysing the effect of severity and duration of agricultural drought on crop performance using terra-modis satellite data and meteorological data. *ITC*, 2006.

Paz, J. O., Batchelor, W. D., Tylka, G., and Hartzler, R. A modeling approach to quantify the effects of spatial soybean yield limiting factors. *Transactions of the ASAE*, 44(5):1329, 2001.

Prasad, A. K., Chai, L., Singh, R. P., and Kafatos, M. Crop yield estimation model for iowa using remote sensing and surface parameters. *International Journal of Applied Earth Observation and Geoinformation*, 8(1):26–33, 2006.

Ransom, J. and Enders, G. Corn growth and management quick guide. *North Dakota State University, Fargo, North Dakota*, 2013.

Seelan, S. K., Laguette, S., Casady, G. M., and Seielstad, G. A. Remote sensing applications for precision agriculture: A learning community approach. *Remote Sensing of Environment*, 88(1):157 – 169, 2003.

Shanahan, J. F., Schepers, J. S., Francis, D. D., Varvel, G. E., Wilhelm, W. W., Tringe, J. M., Schlemmer, M. R., and Major, D. J. Use of remote-sensing imagery to estimate corn grain yield. *Agronomy Journal*, 93(3):583–589, 2001.

Stafford, J., Ambler, B., Lark, R., and Catt, J. Mapping and interpreting the yield variation in cereal crops. *Computers and Electronics in Agriculture*, 14(2-3):101–119, 1996.

Stites, D. The effect of soil ph on crop yield. *The Morning Sun*, 2011. URL `http://www.morningsun.net/x286173897/The-effect-of-soil-pH-on-crop-yield`.

Subedi, K. and Ma, B. Assessment of some major yield-limiting factors on maize production in a humid temperate environment. *Field Crops Research*, 110(1):21 – 26, 2009.

Teal, R., Tubana, B., Girma, K., Freeman, K., Arnall, D., Walsh, O., and Raun, W. In-season prediction of corn grain yield potential using normalized difference vegetation index. *Agronomy Journal*, 98(6):1488–1494, 2006.

Tekin, Y., Kuang, B., and Mouazen, A. Potential of on-line visible and near infrared spectroscopy for measurement of ph for deriving variable rate lime recommendations. *Sensors*, 13(8):10177–10190, 2013.

Tittonell, P. and Giller, K. E. When yield gaps are poverty traps: The paradigm of ecological intensification in african smallholder agriculture. *Field Crops Research*, 143:76–90, 2013.

Tucker, C., Holben, B., Elgin Jr, J., McMurtrey III, J., et al. Relationship of spectral data to grain yield variation. *Photogrammetric Engineering and Remote Sensing*, 46(5):657–666, 1980.

Tzima, K. Adding the precision in farming using remote sensing, 2017. URL `https://geoawesomeness.com/adding-precision-farming-using-remote-sensing/`.

USDA. Soil quality indicators, 2011.

Wang, F.-M., Huang, J.-F., Tang, Y.-L., and Wang, X.-Z. New vegetation index and its application in estimating leaf area index of rice. *Rice Science*, 14(3):195–203, 2007.

Whetton, R., Zhao, Y., Shaddad, S., and Mouazen, A. M. Nonlinear parametric modelling to study how soil properties affect crop yields and ndvi. *Computers and electronics in agriculture*, 138:127–136, 2017.

Wiegand, C. L., Richardson, A. J., Jackson, R. D., Pinter, P. J., Aase, J. K., Smika, D. E., Lautenschlager, L. F., and McMurtrey, J. Development of agrometeorological crop model inputs from remotely sensed information. *IEEE Transactions on Geoscience and Remote Sensing*, GE-24(1):90–98, 1986.

Zarco-Tejada, P., Hubbard, N., and Loudjani, P. Precision agriculture: An opportunity for eu farmers—potential support with the cap 2014-2020. *Joint Research Centre (JRC) of the European Commission*, 2014.

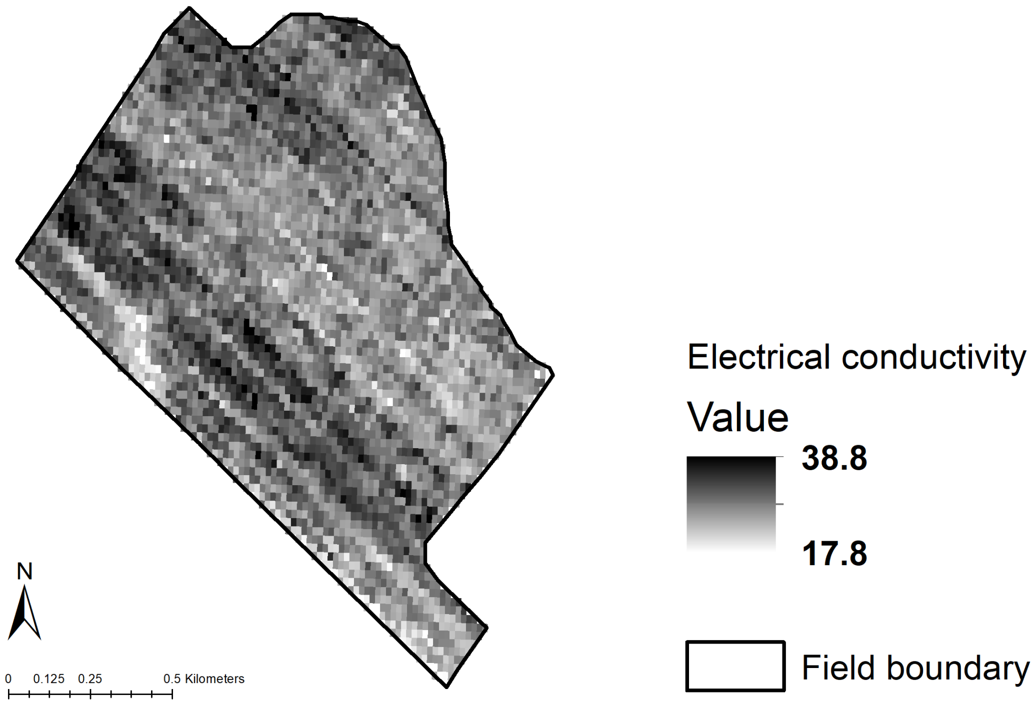
A Examples of harvest data

Figure 46: Yield map of field 6

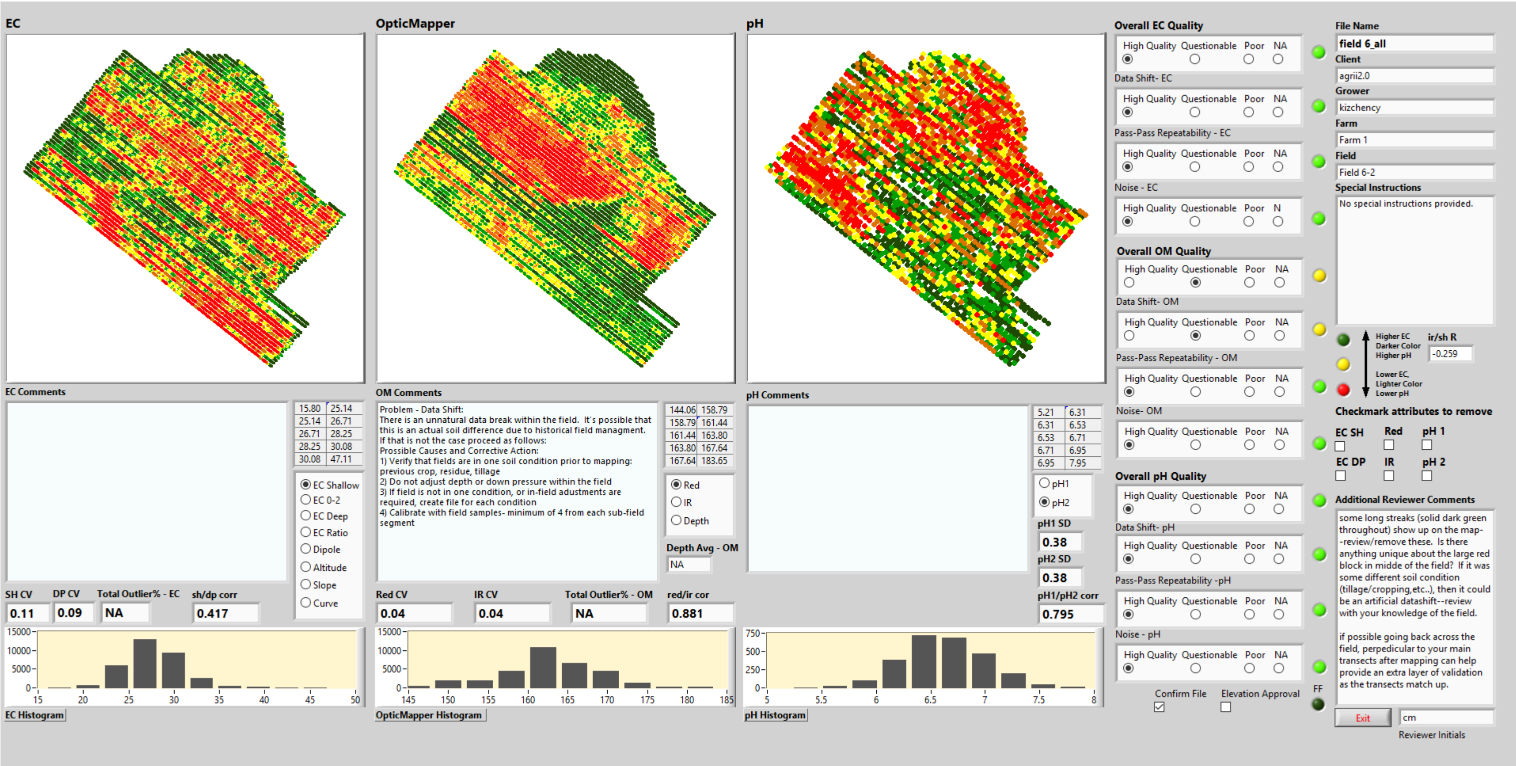
Figure 47: Yield map of field 10

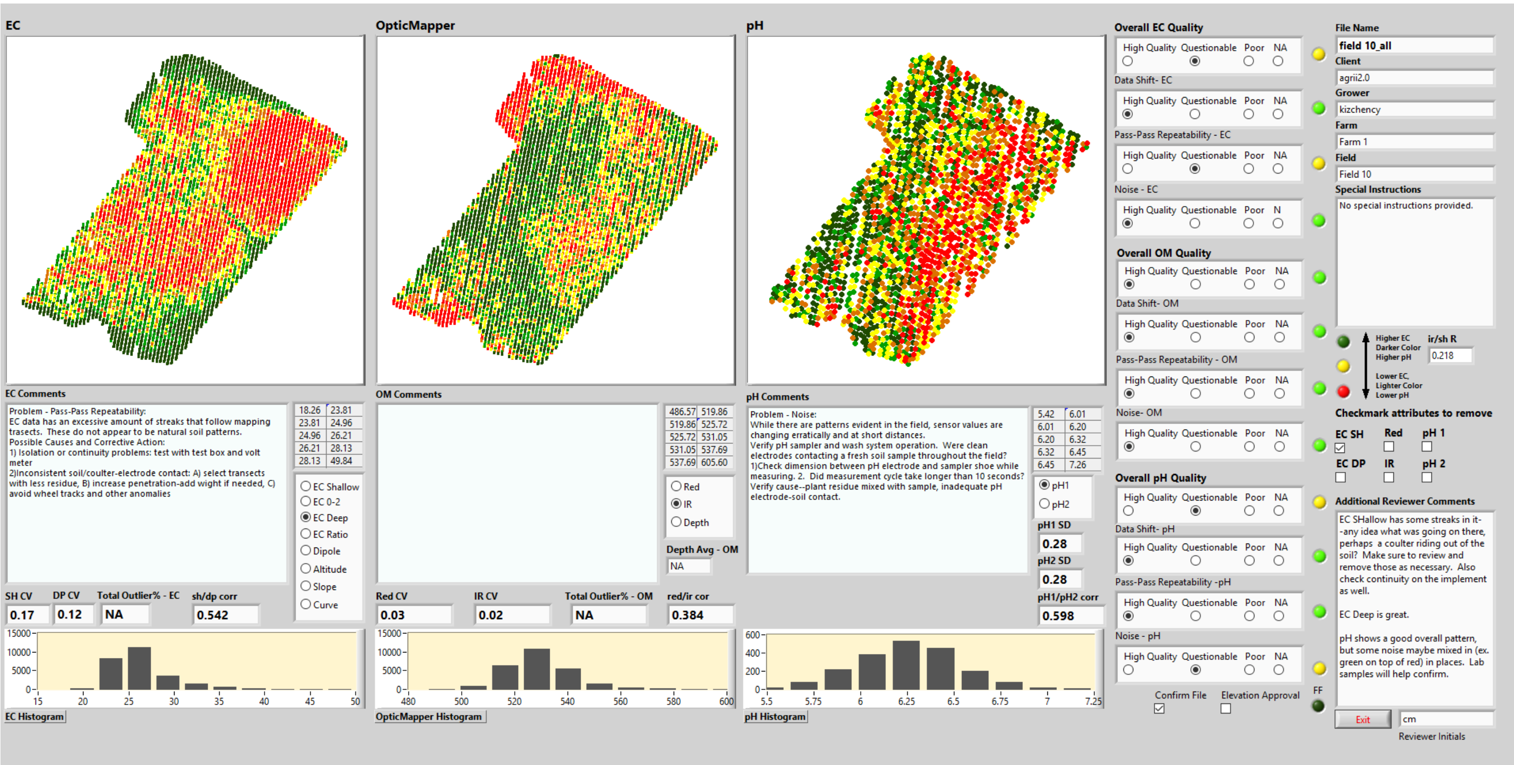
B Example of soil properties data

The picture below represents the shallow electrical conductivity. The unit of measurement is mSm^{-1} .

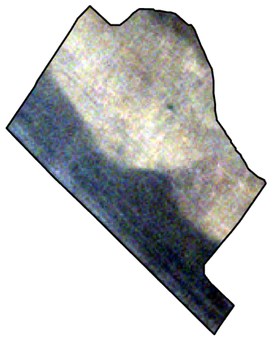


C Soil processing reports by Veris Techologies

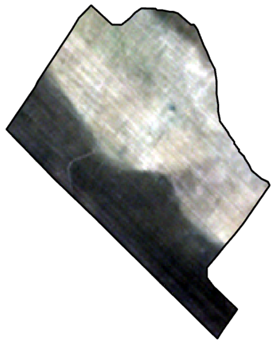



EC
OpticMapper
pH
Overall EC Quality
High Quality Questionable Poor NA
Data Shift- EC
Pass-Pass Repeatability - EC
Noise - EC
Overall OM Quality
Data Shift- OM
Pass-Pass Repeatability - OM
Noise- OM
Overall pH Quality
Data Shift- pH
Pass-Pass Repeatability -pH
Noise - pH
Confirm File
Elevation Approval
FF
File Name
field 10_all
Client
agrii2.0
Grower
kizchency
Farm
Farm 1
Field
Field 10
Special Instructions
No special instructions provided.
Higher EC
Darker Color
Higher pH
Lower EC,
Lighter Color
Lower pH
ir/sh R
0.218
Checkmark attributes to remove
EC SH
Red
pH 1
EC DP
IR
pH 2
Additional Reviewer Comments
EC SHallow has some streaks in it--any idea what was going on there, perhaps a coulter riding out of the soil? Make sure to review and remove those as necessary. Also check continuity on the implement as well.
EC Deep is great.
pH shows a good overall pattern, but some noise maybe mixed in (ex. green on top of red) in places. Lab samples will help confirm.
Exit
cm
Reviewer Initials
EC Comments
Problem - Pass-Pass Repeatability:
EC data has an excessive amount of streaks that follow mapping trasects. These do not appear to be natural soil patterns.
Possible Causes and Corrective Action:
1) Isolation or continuity problems: test with test box and volt meter
2)Inconsistent soil/coulter-electrode contact: A) select transects with less residue, B) increase penetration-add wight if needed, C) avoid wheel tracks and other anomalies
18.26 23.81
23.81 24.96
24.96 26.21
26.21 28.13
28.13 49.84
EC Shallow
EC 0-2
EC Deep
EC Ratio
Dipole
Altitude
Slope
Curve
SH CV
0.17
DP CV
0.12
Total Outlier% - EC
NA
sh/dp corr
0.542
15000
10000
5000
0
15 20 25 30 35 40 45 50
EC Histogram
OM Comments
486.57 519.86
519.86 525.72
525.72 531.05
531.05 537.69
537.69 605.60
Red
IR
Depth
Depth Avg - OM
NA
Red CV
0.03
IR CV
0.02
Total Outlier% - OM
NA
red/ir cor
0.384
15000
10000
5000
0
480 500 520 540 560 580 600
OpticMapper Histogram
pH Comments
Problem - Noise:
While there are patterns evident in the field, sensor values are changing erratically and at short distances.
Verify pH sampler and wash system operation. Were clean electrodes contacting a fresh soil sample throughout the field?
1)Check dimension between pH electrode and sampler shoe while measuring. 2. Did measurement cycle take longer than 10 seconds?
Verify cause--plant residue mixed with sample, inadequate pH electrode-soil contact.
5.42 6.01
6.01 6.20
6.20 6.32
6.32 6.45
6.45 7.26
pH1
pH2
pH1 SD
0.28
pH2 SD
0.28
pH1/pH2 corr
0.598
600
400
200
0
5.5 5.75 6 6.25 6.5 6.75 7 7.25
pH Histogram

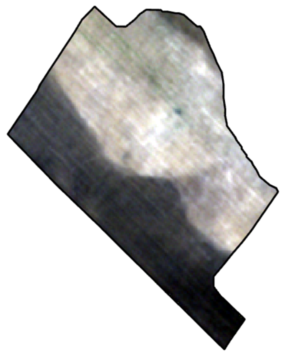
D Overview of satellite images in true colour for field 6



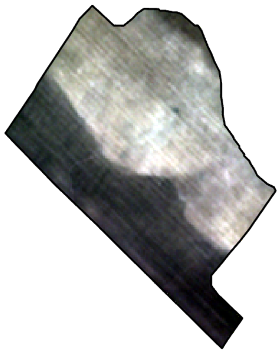
(a) 4 March 2017



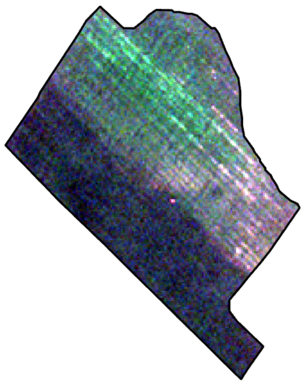
(b) 3 April 2017



(c) 6 May 2017



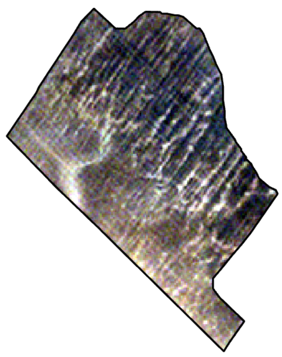
(d) 5 June 2017



(e) 17 July 2017

(f) 11 August 2017

(g) 15 September 2017

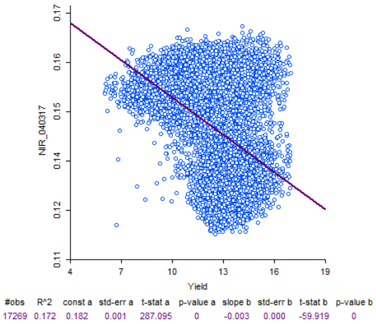


(h) 18 October 2017

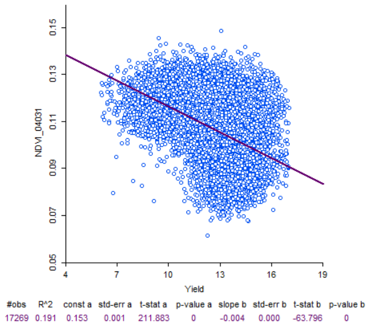
Figure 48: Overview of satellite images in true colour for field 6

E Scatter plots of yield versus band or vegetation indices

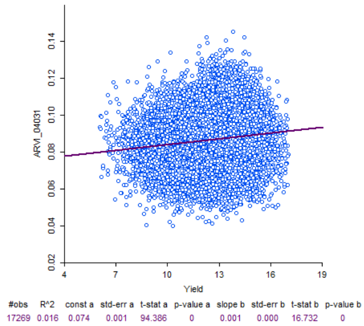
E.1 4th of March 2017



(a) NIR



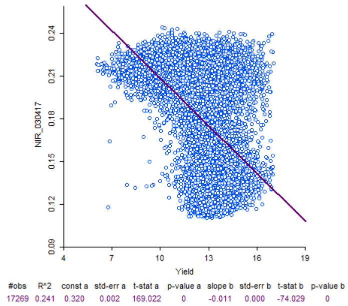
(b) NDVI



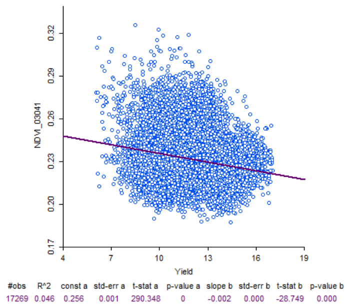
(c) ARVI

Figure 49: Scatter plots of satellite data taken on 4th of March 2017 versus final yield

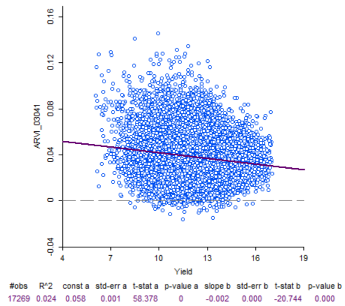
E.2 3rd of April 2017



(a) NIR



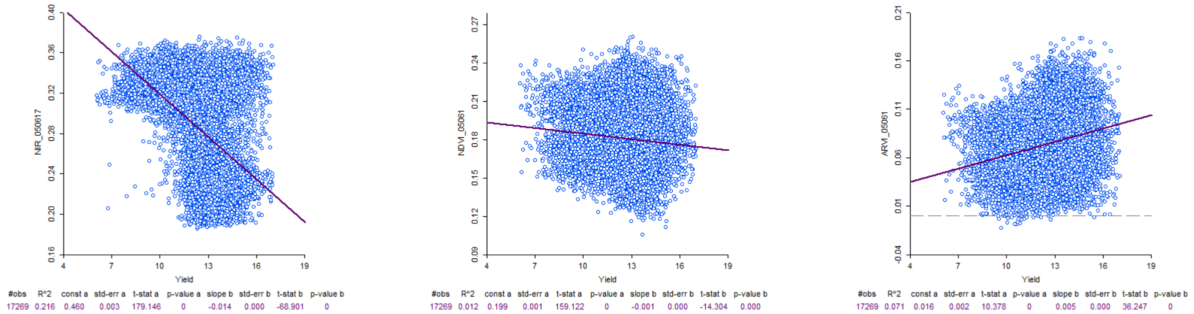
(b) NDVI



(c) ARVI

Figure 50: Scatter plots of satellite data taken on 3rd of April 2017 versus final yield

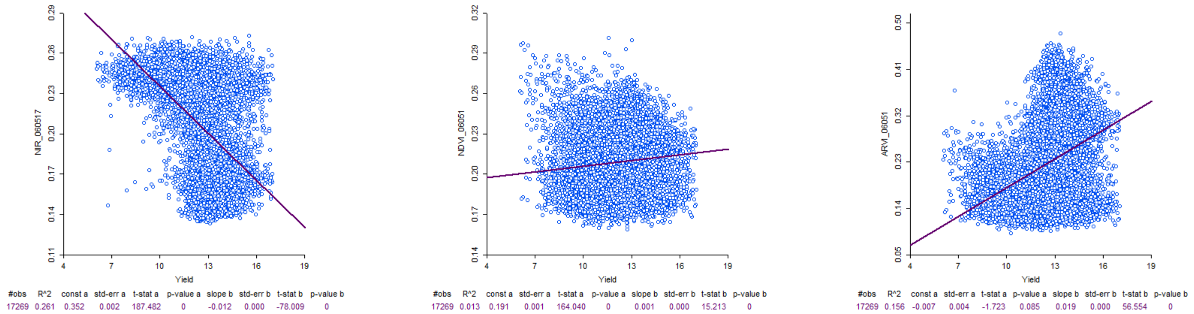
E.3 6th of May 2017



(a) NIR *(b) NDVI* *(c) ARVI*

Figure 51: Scatter plots of satellite data taken on 6th of May 2017 versus final yield

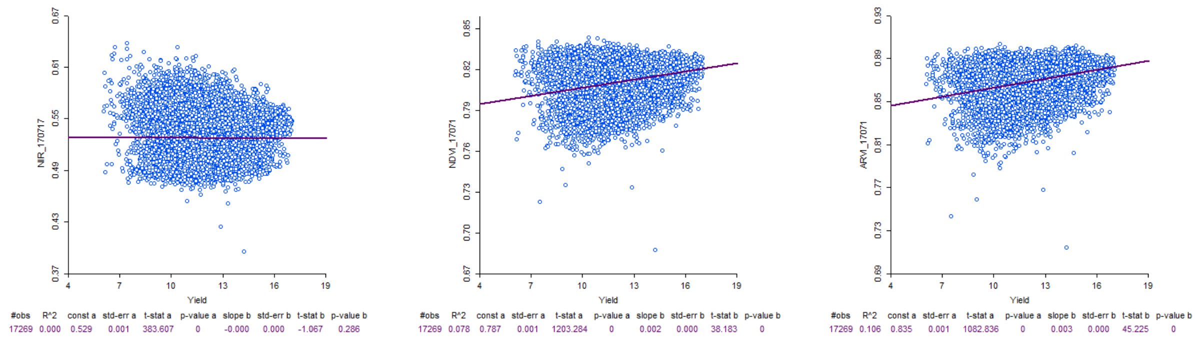
E.4 5th of June 2017



(a) NIR *(b) NDVI* *(c) ARVI*

Figure 52: Scatter plots of satellite data taken on 5th of June 2017 versus final yield

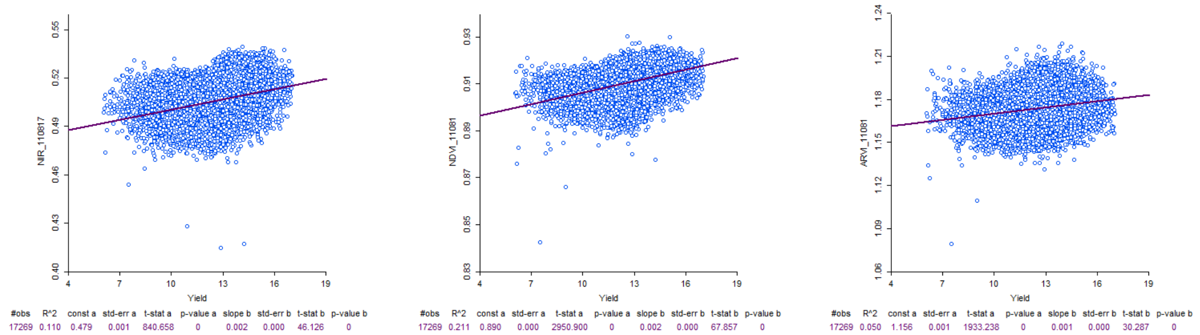
E.5 17th of July 2017



(a) NIR *(b) NDVI* *(c) ARVI*

Figure 53: Scatter plots of satellite data taken on 17th of July 2017 versus final yield

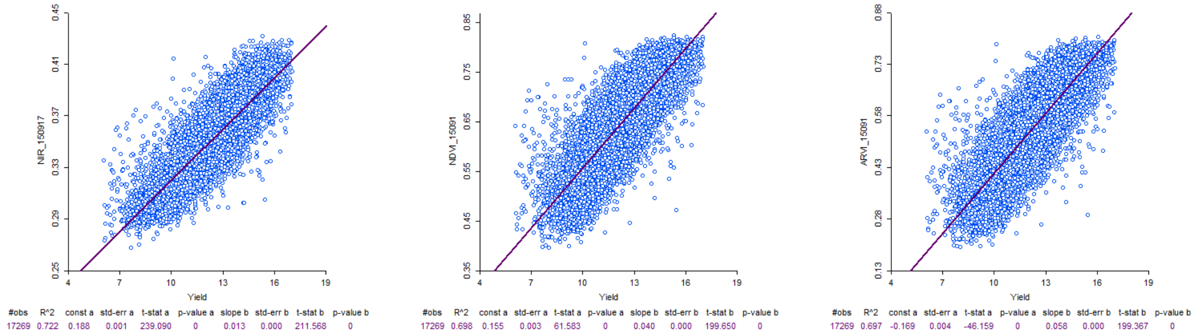
E.6 11th of August 2017



(a) NIR *(b) NDVI* *(c) ARVI*

Figure 54: Scatter plots of satellite data taken on 11th of August 2017 versus final yield

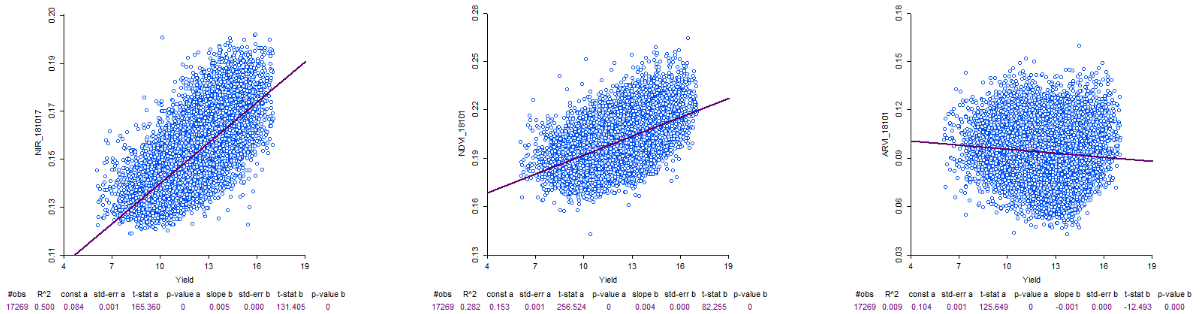
E.7 15th of September 2017



(a) NIR *(b) NDVI* *(c) ARVI*

Figure 55: Scatter plots of satellite data taken on 15th of September 2017 versus final yield

E.8 18th of October 2017



(a) NIR *(b) NDVI* *(c) ARVI*

Figure 56: Scatter plots of satellite data taken on 18th of October 2017 versus final yield

F R^2 values of NIR band and NDVI with EC Deep

Table 9: R^2 values of NIR band and NDVI with shallow measurement of electrical conductivity

Date	**NIR**	**NDVI**	**ARVI**
4 March 2017	0.029	0.0.021	0.004
3 April 2017	0.019	0.018	0.018
6 May 2017	0.021	0.022	0.014
5 June 2017	0.022	0.025	0.051
17 July 2017	0.034	0.051	0.047
11 August 2017	0.014	0	0.001
15 September 2017	0	0.002	0.001
18 October	0.004	0.004	0.006

Table 10: R^2 values of NIR band and NDVI with deep measurement of electrical conductivity

Date	**NIR**	**NDVI**
4 March 2017	0.003	0.002
3 April 2017	0	0.006
6 May 2017	0	0.001
5 June 2017	0.001	0.011
17 July 2017	0.018	0.018
11 August 2017	0	0.006
15 September 2017	0	0.001
18 October	0.005	0.001

G R^2 of Organic Matter content against NIR and NDVI for field 6 and 10

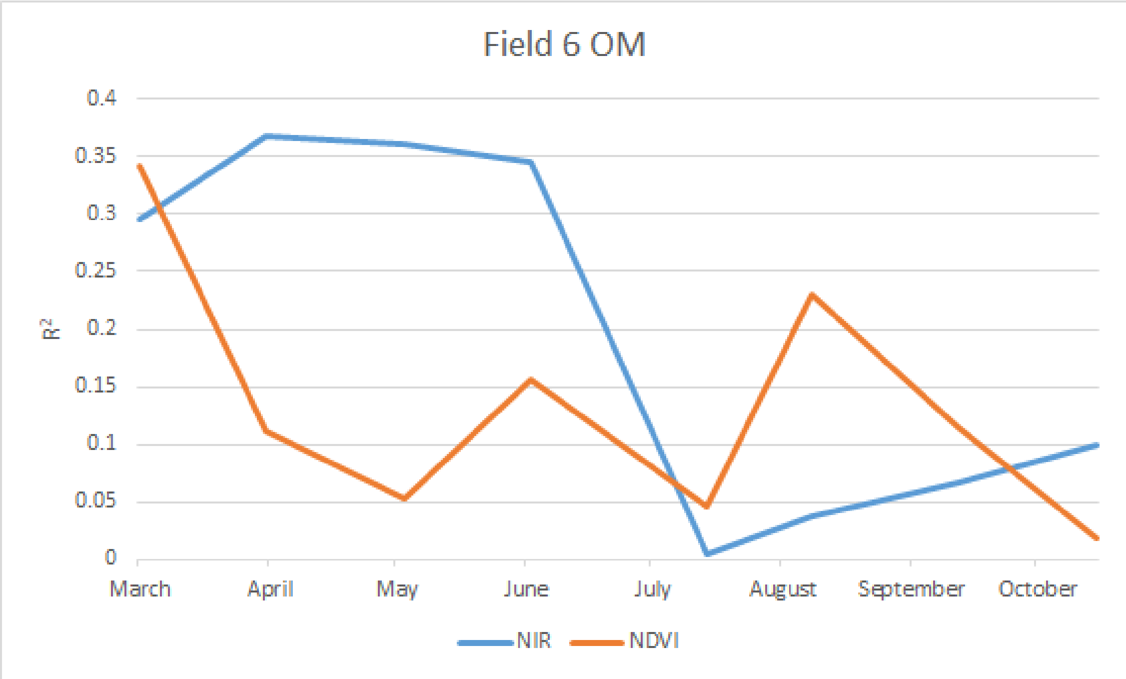


Figure 57: R^2 of Organic Matter content against NIR and NDVI in field 6

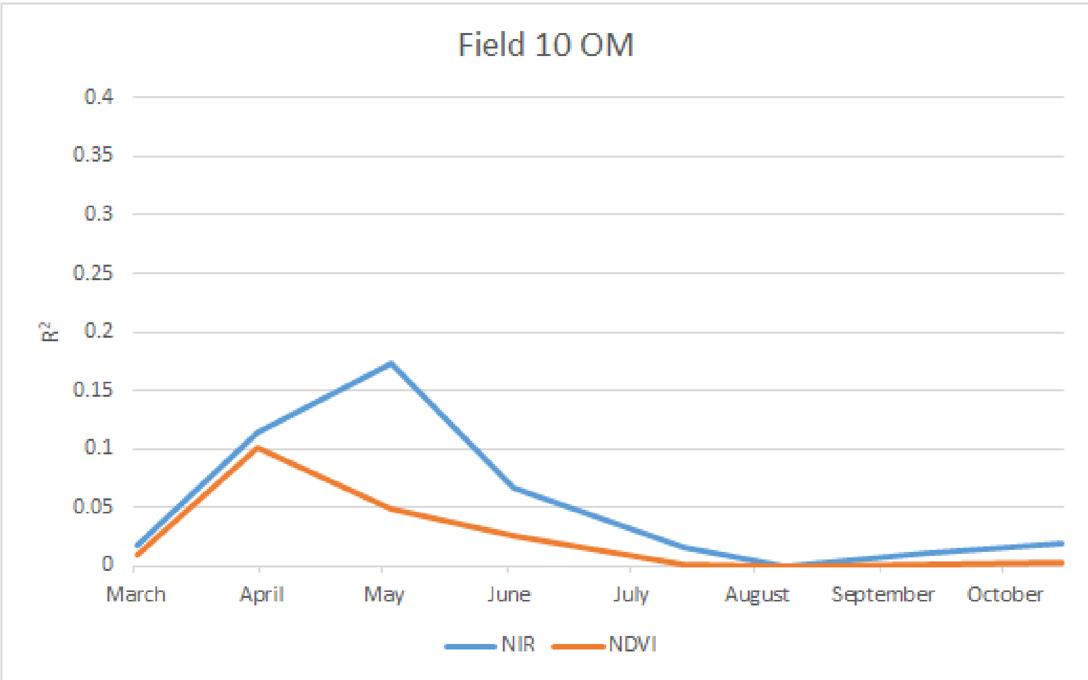


Figure 58: R^2 of Organic Matter content against NIR and NDVI for field 10

H R^2 of pH content against NIR and NDVI for field 6 and 10

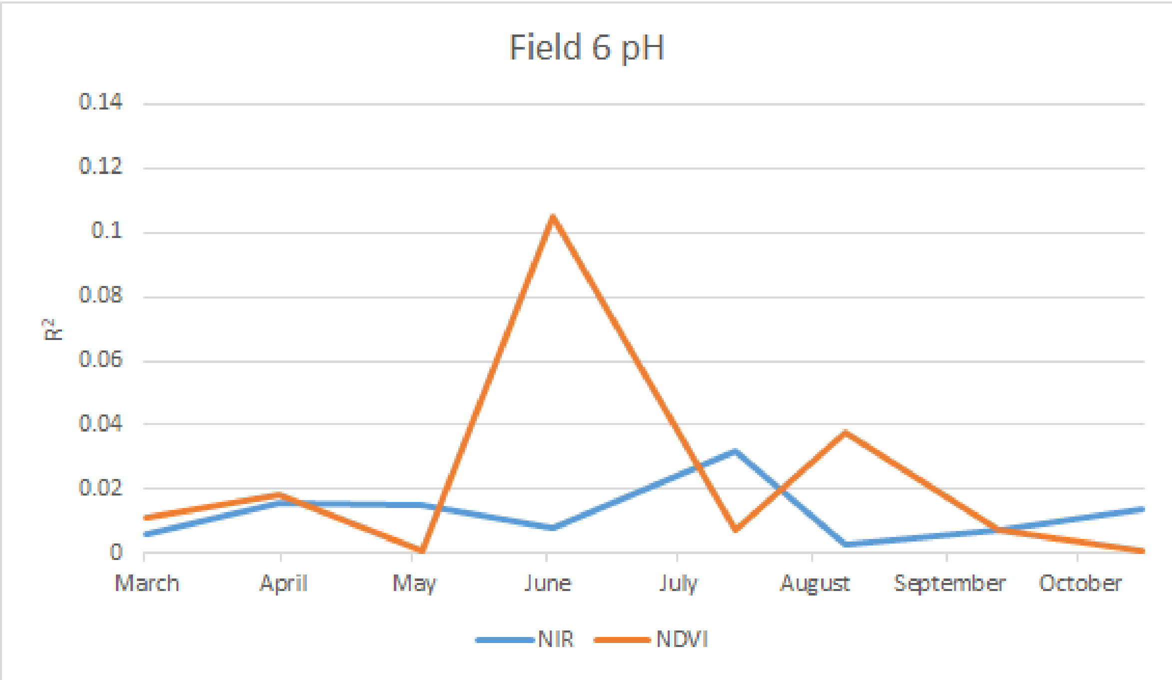


Figure 59: R^2 of pH content against NIR and NDVI in field 6

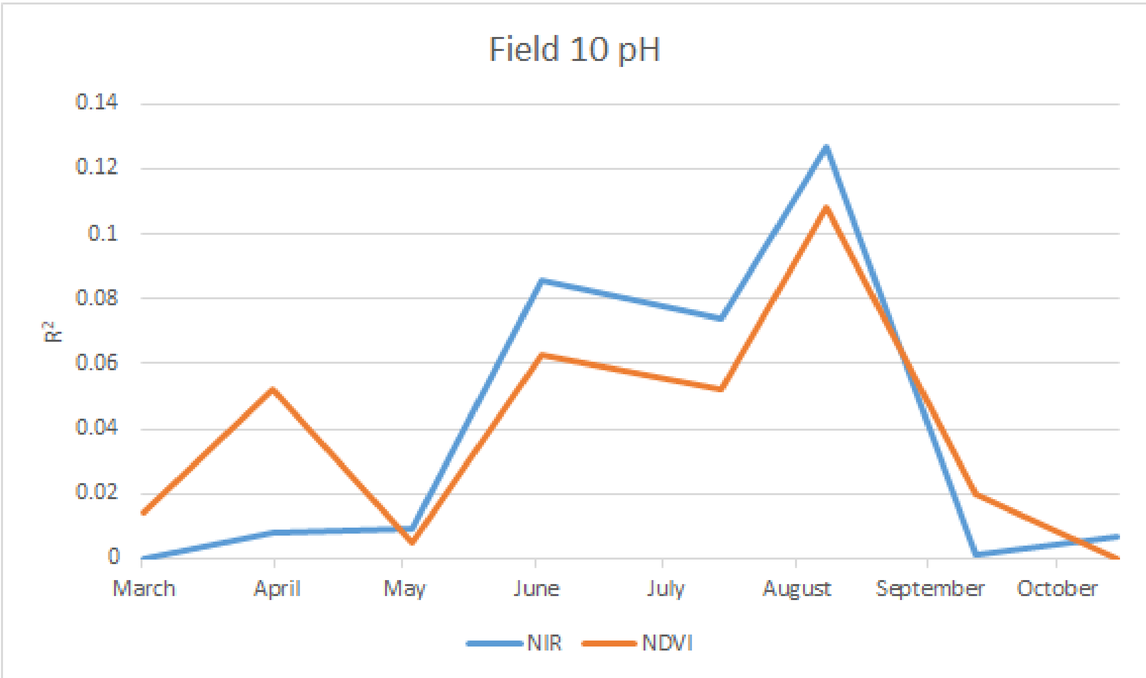


Figure 60: R^2 of pH content against NIR and NDVI for field 10

I R^2 of EC content against NIR and NDVI for field 6 and 10

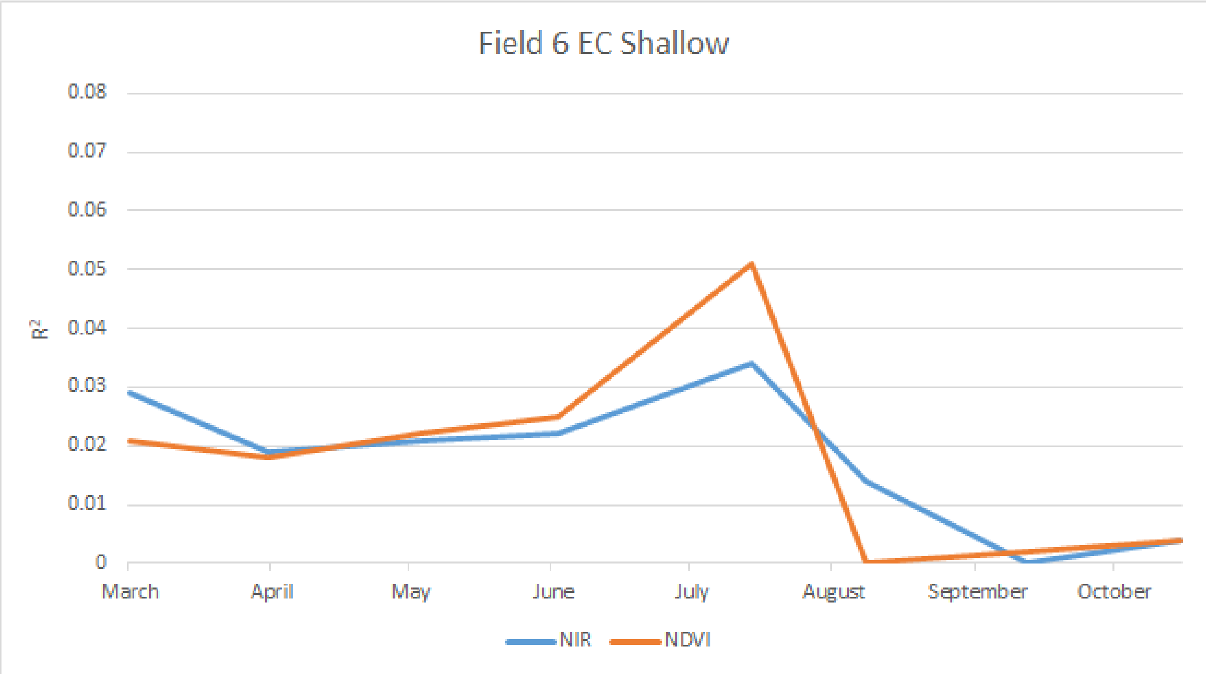


Figure 61: R^2 of electrical conductivity content against NIR and NDVI in field 6

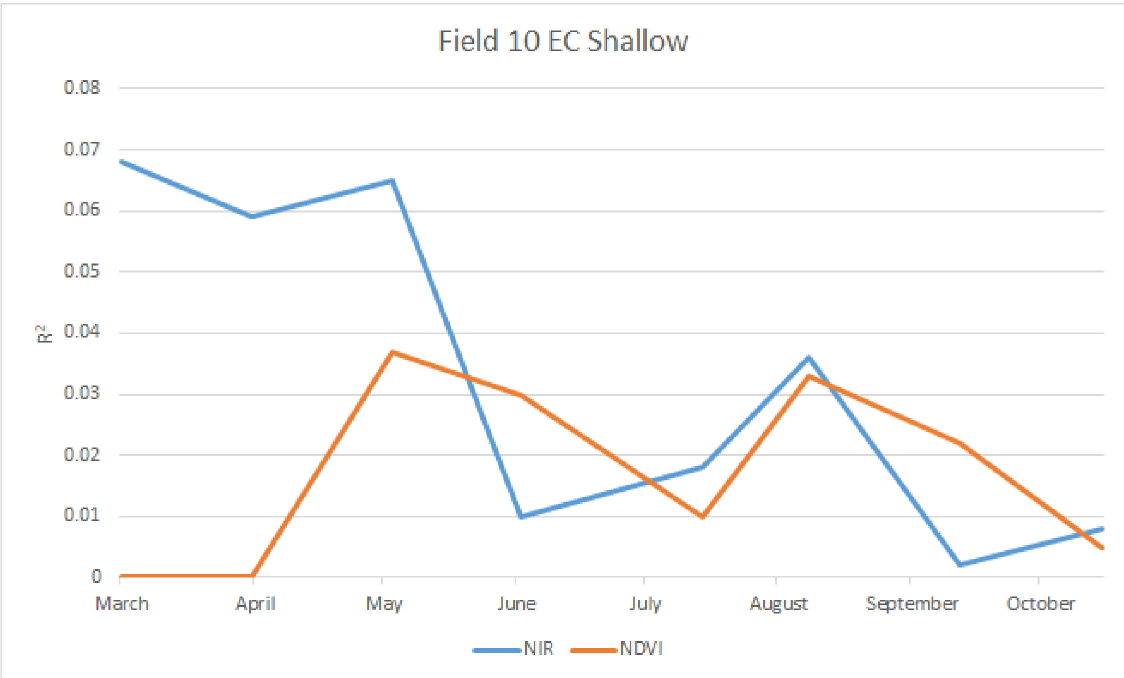


Figure 62: R^2 of electrical conductivity content against NIR and NDVI for field 10

J R^2 of elevation against NIR and NDVI for field 6 and 10

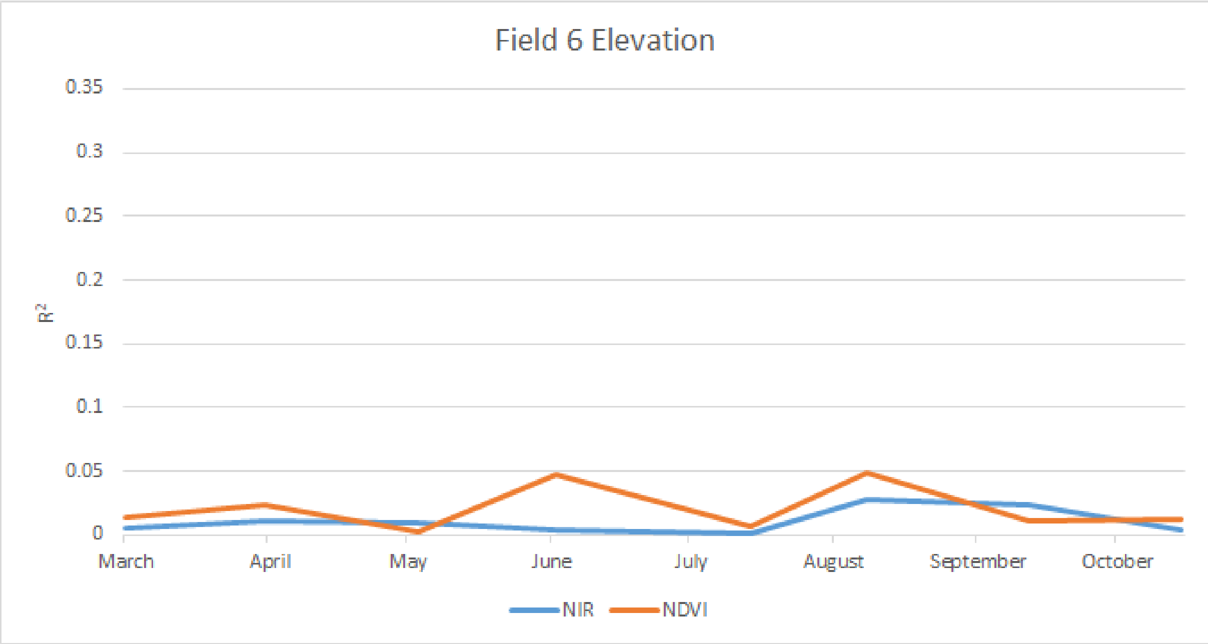


Figure 63: R^2 of elevation content against NIR and NDVI in field 6

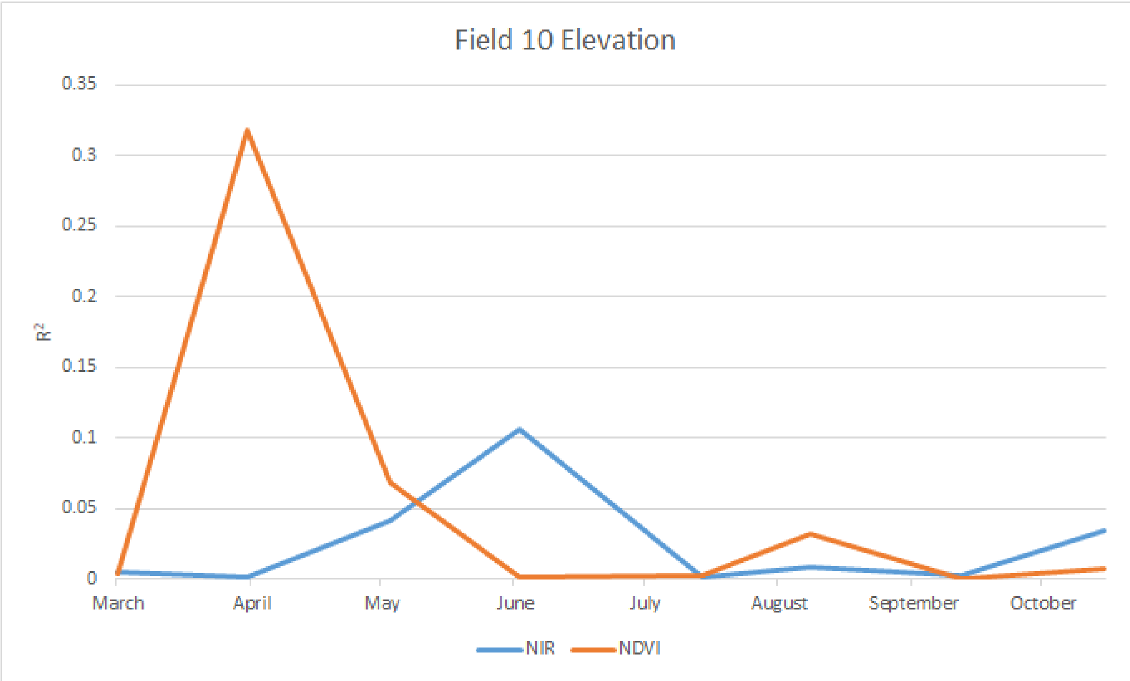


Figure 64: R^2 of elevation conductivity content against NIR and NDVI for field 10