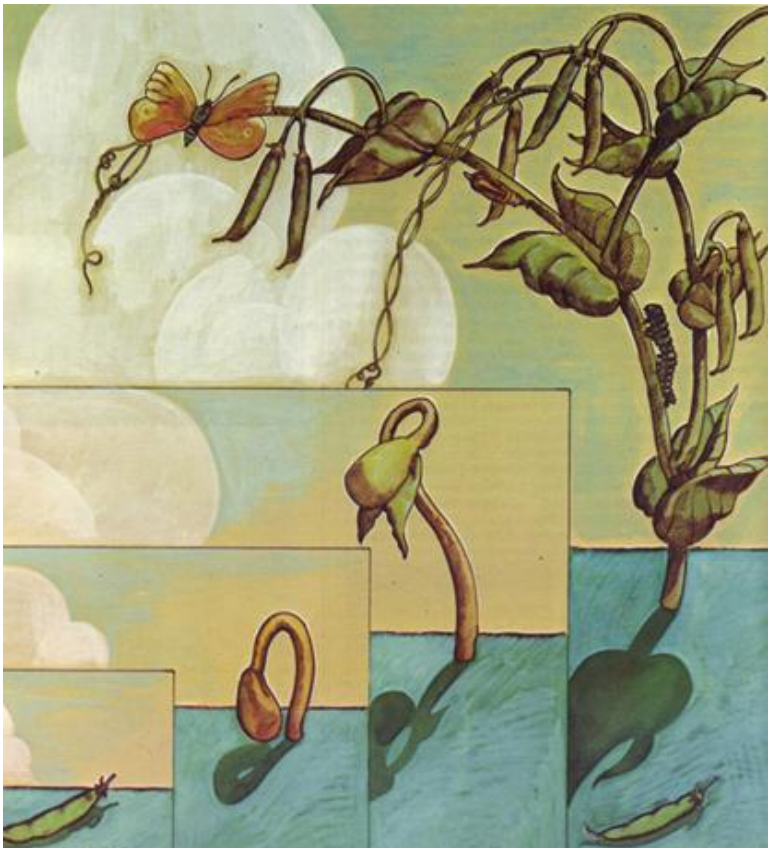
Centre for Geo-Information

Thesis Report GIRS-2013-09

Opportunities of hyperspectral vegetation indices to assess nitrogen and chlorophyll content in potato, maize and grassland

Ilina B. Kamenova

04.04.2013



WAGENINGEN UNIVERSITY
WAGENINGEN UR

Opportunities of hyperspectral vegetation indices to assess nitrogen and chlorophyll content in potato, maize and grassland

Ilina B. Kamenova

Registration number: 851113 – 417 - 050

Supervisors:

Lammert Kooistra
Petra van Vliet

A thesis submitted in partial fulfilment of the degree of Master of Science
at Wageningen University and Research Centre,
The Netherlands

Thesis code number: GRS-80436 Date: 04. 04. 2013
Thesis Report: GIRS-2013-09 Wageninegen, The Netherlands
Wageningen University and Research Centre
Laboratory of Geo-Information Science and Remote Sensing

Foreword

The subject of my thesis report is the nitrogen estimation based on vegetation indices and hyperspectral data analysis in precision agriculture. It focuses on a study area in the Netherlands but the methods can be easily transferred and used elsewhere around the globe. I have been always fascinated about the nature and the use of remote sensing gave me a good opportunity to 'concatenate' the knowledge about geography, biology and mapping I had in prior. This report is a result of over half a year work. During this period I was reading, writing, programming, analysing, getting lost and finding the way back with the priceless help of my supervisors Lammert Kooistra and Petra van Vliet. Without them it has not been possible to conduct this research and write my thesis. I would like to express my special thanks to all the people who helped me to go through this challenging experience with a smile. I will always remember the time spent in Gaia with my fellow classmates, the coffee brakes, and the ironic sense of humour after the long days. I want to thank to: Kim and Jose for introducing me to the wonderful world of R; Sebastian for the inspiration which kicked me in to become an MGI student; Christine and Lingtong for the nice company, dinners, proof reading and everything else. Furthermore I'd like to thank my family and friends for supporting me while working on my research and keeping me motivated during the times when I needed them most.

Abstract

To reduce environmental pollution from agricultural activities, a reliable indicator of crop nitrogen status is needed for site specific fertilization management in agricultural fields. To characterize the spatial variability of nitrogen over large fields, using traditional methods such as soil testing, plant nutrient analysis and SPAD meter many point samples are required. Because of the strong correlation between leaf chlorophyll and leaf nitrogen content in green vegetation, remote sensing techniques have the potential to evaluate spatial variability over large agricultural fields with lower costs than the traditional plant destructive methods. The main objectives of this research were to: (a) test the ability of hyperspectral vegetation indices to estimate nitrogen and chlorophyll content and to compare it with indices calculated from broad band sensors (Landsat TM and Sentinel-2); (b) to calibrate and validate simple regression models relating vegetation indices and in situ measurements. The hyperspectral data were gathered for grassland, potato and maize crops for 10 fields in Noord- Brabant, the Netherlands using the Airborne Prism Experiment (APEX) sensor. The data was subsequently simulated to the spectral bandwidths of Sentinel-2 and Landsat TM. Field measurements of nitrogen and chlorophyll concentrations were sampled in the investigated agricultural fields.

Various vegetation indices were calculated based on the original APEX bands and also based on the simulated spectra from Sentinel-2 and a Landsat TM. Subsequently the hyperspectral and broadband indices were tested for estimating nitrogen and chlorophyll content using regression analysis. In the regression model linear and exponential relationships were investigated. On the subject of nitrogen estimation in grassland the best performance was achieved by REP model (R^2 =0.61 linear and R^2 =0.71 exponential); in potato by MTCI (R^2 = 0.65 linear and R^2 =0.75 exponential); in maize there were no significant correlation for any of the tested indices. The best chlorophyll correlation was with MTCI (R^2 = 0.9 linear and R^2 =0.93 exponential). Coefficients obtained by the best performing indices per crop were used for calculation of nitrogen maps, showing the spatial variability of nitrogen content over agricultural fields.

There was no general index able to predict the nitrogen variability for all crops, hence the choice of explanatory variable should be crop specific. However the best results for both potato and grassland were obtained from indices estimating the red-edge position (REP and MTCI). Despite the non- linear correlation, nitrogen maps can be calculated using exponential relationships as long as they are in the range of the observed values used for model calibration. The resulting nitrogen maps could be a valuable input for precision farming applications, taking into account the accuracy of the estimated nitrogen and the spatial accuracy. The correlations between vegetation indices based on Sentinel-2 in terms of R^2 and RMSEP values were similar to APEX indices and proved the utility of the upcoming sensor for estimation of canopy properties. Results also confirmed the importance of red-edge bands, which absence in Landsat TM makes the estimation of chlorophyll and nitrogen insufficient.

Table of Contents

1. Introduction

The first chapter of the thesis will provide brief information about the research topic. The relevance of the research will be discussed and what has already been done in the scientific community. These topics will be further elucidated in chapter 2. The main research objectives and the connected research questions will be outlined, finally in the last section of this chapter we will discuss how the rest of thesis will lead to an answer of the posed research questions.

1.1 Context and Background

Nitrogen is of particular interest in ecological and agricultural studies, because nitrogen availability can affect the rate of key ecosystem processes, including primary production (Vitousek and Howarth 1991). Plants need at least seventeen elements to grow. Three of these elements- carbon, oxygen and hydrogen are referred to as 'building blocks'. The other fourteen elements, such as nitrogen (N), potassium (K) and phosphorus (P) are referred to as 'nutrients'. Nitrogen has traditionally been considered one of the most important nutrients. It is an essential component of the proteins that build cell material and plant tissues (Henry, Sullivan et al. 1999). In addition, it is necessary for the function of other essential biochemical agents, including chlorophylls A and B; chloroplast enzymes of the Calvin cycle which are dominated by ribulose-1,5-bisphosphate carboxylase oxygenase (RuBisCO); high energetic compounds such as ATP and NADPH (Evans 1989); and the nucleic acids DNA and RNA.

The absorbed solar radiation by a leaf is a function of the photosynthetic pigment content. Thus chlorophyll content can directly determine the photosynthetic potential and primary production (Filella, Serrano et al. 1995). Chlorophylls A and B are essential pigments in the light-dependent reactions of photosynthesis for the conversion of light energy to chemical energy. One molecule of chlorophyll absorbs one photon and loses one electron, which leads to reduction of NADP to NADPH and synthesis of ATP. These compounds take part next within the Calvin cycle where the enzyme RuBisCO captures CO_2 in bounds of sugars. It has been estimated that RuBisCO accounts for a quarter of leaf nitrogen and it is probably the most abundant protein of the world (Portis Jr and Parry 2007). Chlorophylls can give an indirect estimation of the nutrient status, because part of the leaf nitrogen is incorporated in chlorophyll. Despite the relatively low nitrogen content of chlorophyll (4 mol /mol $^{-1}$), strong correlations are found between chlorophylls and nitrogen in green leaves, because of the large amount of protein that complexes the photosynthetic pigment (Evans 1989). Furthermore, leaf chlorophyll content is indicative for health status evaluation and is closely related to plant stress and senescence (Merzlyak, Gitelson et al. 1999; Carter and Knapp 2001).

Of all the major plant nutrients, N is often the most important determinant of plant growths and crop yield (Henry, Sullivan et al. 1999). The productivity and dynamics of many unmanaged terrestrial and marine ecosystems, and most agricultural and managed-forestry ecosystems, are limited by the supply of biologically available nitrogen (Vitousek and Howarth 1991). Although an artificial supply of nitrogen to crops is fundamental to optimize crop yields, mismanagement of N and its excessive application, causes many negative effects which has dramatically altered the global nitrogen cycle. Combined N in the atmosphere and precipitation fertilizes in natural ecosystems result in eutrophication, lowered biodiversity, N leakage, while acidity from nitric oxide and ammonia oxidation results in acid lakes and streams, and declining health of forests (Keeney and Hatfield 2008). The effect of anthropogenic activities on the N cycle has been addressed to some extent. Europe has had some success using rules and fines to modify the fertilizer and animal farm waste. Educational programs need to be further developed to modify human behaviour including the way farmers manage N fertilizers in their farms (Tilman, Cassman et al. 2002).

One of the largest challenges facing the agricultural sector is to produce enough food for the growing population while at the same time protecting the environment and human health from excess supply of fertilizers. This challenge requires knowledge about the crop status and good understanding for the responsible biological processes (Matson, Naylor et al. 1998). Precision agriculture or precision farming is a management strategy aiming to improve the decision making process related to crop and pasture management by farmers using integrated information technologies (e.g. ,GIS, GPS, remote sensing etc.) next to traditional practices. The variable rate technology adopted by this management practice aims to maximize the productivity applying specific inputs, such as fertilizers, for specific conditions at specific location and specific time (Moran, Inoue et al. 1997). Among the fertilizers, nitrogen is essentially the most important one.

Different methods have been developed and implemented for assessment of crop nitrogen status. Traditionally, nitrogen content has been estimated through soil testing and plant tissue analysis, which involves destructive field sampling and laboratory chemical analysis (Daughtry, Walthall et al. 2000). Improved non-destructive technology, such as SPAD meter (Minolta Osaka Co., Ltd. Japan), is based on measuring leaf transmittance in two wave bands centred at 650 and 940 nm, which well correlate to leaf chlorophyll content (Uddling, Gelang-Alfredsson et al. 2007). These methods require laborious field measurements and are costly in terms of time and money. Even though SPAD is a great improvement in the non-destructive analysis it can be hardly extended to cover large areas. Ideally a method is required that is accurate, non-destructive, simple to use, covering large areas (Curran, Windham et al. 1995).

Remote sensing techniques satisfy these requirements by measuring the optical properties of leaves and canopy that are driven by the biochemical and biophysical properties of the plant. Field experiments on leaf and canopy level with multispectral sensors demonstrated the sensitivity of visible (VIS), near infrared (NIR) and short wave infrared (SWIR) part of electromagnetic spectrum to pigments and nitrogen content. The development of hyperspectral instruments boosted the use of remote sensing data for biochemical quantification (Hansen and Schjoerring 2003). Hyperspectral remote sensing has large amount of spectral information measured by using narrow contiguous bands (Figure 1A), which can provide a new means for non-destructive, fast, and real-time monitoring of plant biochemical parameters. Remote sensing images are capable to provide spatial information of vegetation features and have the potential to sense various biophysical and biochemical parameters of plants. This is possible due to specific interactions of solar light and vegetation along the electromagnetic spectrum, this interaction is due to the relationships between the optical, morphological and physiological properties of vegetation (Knipling 1970).

In optical remote sensing for vegetation monitoring (400-2500 nm), three parts of the spectrum are of particular interest: VIS-NIR 400-700 nm; NIR 700-1100 nm; and SWIR 1100- 2500 nm. Every region is characterized by specific radiation- vegetation interactions influenced by leaf biochemical and structural canopy characteristics. In the visible region of electromagnetic spectrum the reflection of light is mainly regulated by the pigment absorption within the leaves. Thus the reflection in VIS provides information about the chlorophyll content in plants. Chlorophylls absorb radiation mainly in the blue (≈450 nm) and red (≈680 nm) wavelengths. In the NIR wavelengths the reflectance is caused by the internal cellular structure (Knipling 1970) and provides information for biomass and leaf area index (LAI).

Between the red and NIR wavelengths, leaf reflectance is associated with the transition from chlorophyll absorption to leaf scattering. This position of the maximum slope of reflectance is called red-edge position (REP) (Horler, Dockray et al. 1983). In the third region of interest SWIR reflection of solar light is influenced by the major absorption features of water and minor absorption features related to foliar biochemical including nitrogen. Determination of N content in this SWIR is very complex, due to the strong effect of water absorption and the impossibility to associate a specific nitrogen absorption feature among the abundance of biochemicals (Kokaly and Clark 1999).

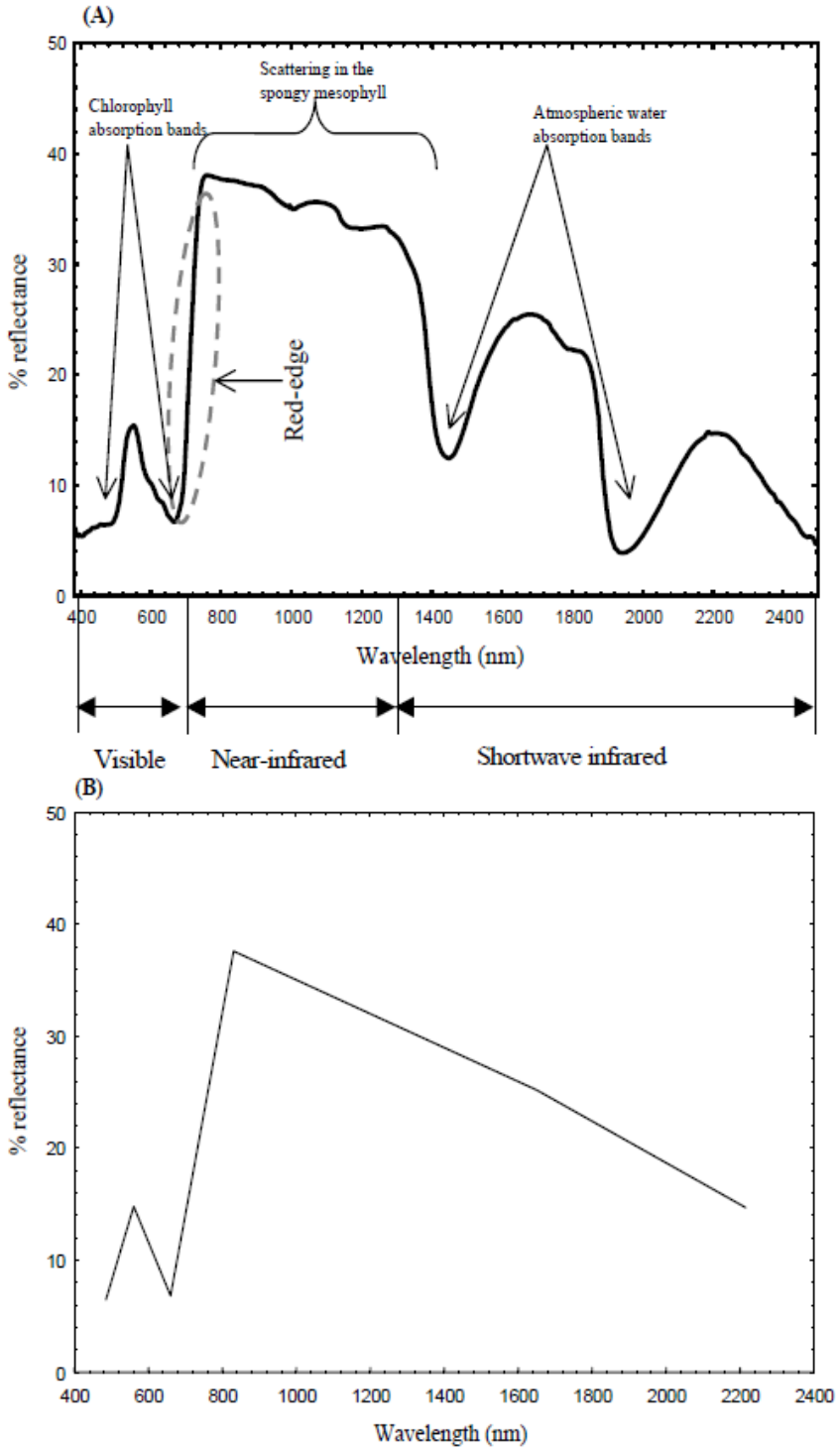


Figure 1. (A) Contiguous spectrum of healthy green vegetation using handheld GER 3700 spectrometer and (B) the same spectrum re-sampled to 6 bands of Landsat TM (Cho and Skidmore 2006)

1.2 Problem definition

Estimates of biochemical and biophysical parameters over large areas may be obtained using remote sensing data acquired from air and space platforms. Most of the remote sensing products for quantifying those parameters are derived from the conventional broadband (multispectral) sensors such as SPOT, Landsat TM, MODIS, NOAA- AVHRR, DMC and others. The multispectral scanners can collect data in multiple spectral bands in the range from 300 to approximately 1400 nm (Lillesand, Kiefer et al. 2004p. 325). Broadband sensors collect data in discrete channels with fairly wide bandwidths over which the reflectance data is averaged. A major limitation of the broadband remote sensing products is that due to the averaging over certain wave widths of interest, information is lost which is available in specific narrow bands (Figure 1). Even though remote sensing has proven to be reliable source for estimating biochemical and biophysical properties of green vegetation variables, broadband sensors rather deliver inadequate or supplied information for this purpose (Thenkabail, Enclona et al. 2004). The limitations of broadband data analysis can be illustrated by vegetation indices (VIs), which saturate beyond certain level of biomass and leaf area index (LAI) (Thenkabail, Smith et al. 2000).

Normalized difference vegetation index (NDVI) is one of the most commonly used vegetation indices for crop growth monitoring calculated by broadband sensors (Huang et all 2004). It combines NIR spectral band, representing scattering of solar radiation by the canopy, with red band, representing absorption by chlorophyll. The problem using NDVI for chlorophyll estimation is that NDVI saturates at higher chlorophyll levels, because the absorption in the red bands is too prominent and this results in loss of sensitivity. Due to lower absorption by chlorophyll in the red-edge region (referred earlier to as REP), use of such a band reduces the saturation effect, and reflectance still remains sensitive to chlorophyll absorption at it moderate to high values (Gitelson and Merzlyak, 1996). Red-edge bands are not available at the current operational multispectral devises and the VIs calculated based on this data suffer from the 'saturation effect'.

Recent advances in hyperspectral remote sensing or (imaging spectroscopy) demonstrate great utility for variety of vegetation monitoring applications. Hyperspectral data analyses are superior to traditional broadband analyses in spectral information. Many studies explore hyperspectral remote sensing of vegetation and agricultural croplands and its utility for monitoring purposes. Some examples include (a) detecting plant stress (Thenkabail, Enclona et al. 2004), (b) measuring chlorophyll content of plants (Haboudane, Miller et al. 2002), (c) extracting biochemical variables such as nitrogen (Chen, Haboudane et al. 2010; Clevers and Kooistra 2012), (d) modelling biophysical and yield characteristics of crops (Thenkabail, Smith et al. 2000), etc.

Hyperspectral sensors (sometimes referred to as imaging spectrometers) are instruments that acquire images in many, very narrow, contiguous spectral bands throughout the visible near-IR, mid-IR and thermal IR portions of the spectrum. These systems typically collect 200 and more bands of data (Lillesand, Kiefer et al. 2004). Examples of these sensors are: the Airborne Visible-infrared Imaging Spectrometer (AVRIS); Compact Airborne Spectrographic Imager (CASI); Airborne Prism Experiment (APEX). Although most of the sensors are airborne, the potential of hyperspectral sensors has motivated also the launch of spaceborne sensors for example Hyperion on board of Earth observing-1 (EO-1).The APEX imaging spectrometer is a development of Swiss-Belgian consortium on behalf of ESA. It is intended as a validation and calibration device for future spaceborne hyperspectral imagers.

It should be noted that the use of hyperspectral data is much more complex and extensive than the multispectral data. As mentioned before imaging spectrometers usually gather data in hundreds of near- contiguous narrow bands and the data volume is vast. The increase in data volume poses challenges in data storing and handling. This

issue makes it essential to develop methods to handle the high-dimensional data or build specialized optimal sensors to gather data for specific applications in optimal band selection, excluding the redundant bands. (Thenkabail, Enclona et al. 2004) Therefore prior knowledge of specific, optimal bands is crucial to reduce costs in data storing and handling. Methods and approaches of hyperspectral data analysis need to be established in order to achieve highest accuracies, indices and wavebands should be developed that best model biochemical quantities and identify and eliminate redundant bands. In hyper spectral data analysis (Thenkabail 2012) advocate quick elimination of redundant bands and identification of most useful bands on a first place. This process should establish categorization to achieve higher accuracy and to develop indices that best model biophysical and biochemical quantities. The process involves review of performance of VIs that will help to establish the wavebands associated with these indices in the study of vegetation and agricultural crops.

Future generation of sensors will address the issues posed by conventional broadband sensors and the limited spectral information they provide next to the vast and mostly redundant data offered by hyperspectral acquisition methods. These sensors will have more than 10 and less than 50 bands, these bands are narrower than broadband, but they are not contiguous. In this sense they will have an intermediate position between broad and hyperspectral remote sensing. They aim at high quality precision agricultural red-edge applications and should introduce a unique combination of spectral and spatial resolutions (Herrmann, Pimstein et al. 2011). The upcoming satellites Sentinel-2 and Venμs are about to be launched in 2014. They are intended for environmental and agricultural applications. Both sensors have four red-edge bands which are particularly sensitive to biochemical content and vegetation stress, they will be particular suitable for precision agriculture tasks, such as site specific management.

1.3 Research objectives

The aim of this study was to investigate the possibilities to derive nitrogen and chlorophyll content of potato, maize and grassland using data derived from APEX imaging spectrometer. For this purpose a statistical approach using regressive models were adopted, relating in situ measurements and different hyperspectral vegetation indices (HVIs). The study was targeted to the following objectives:

A. To test the ability of hyperspectral data analysis to estimate nitrogen and chlorophyll content and compare with the potenential of Landsat TM and Sentinel-2.

B. To calibrate and validate empirical model relating hyperspectral data and in situ measurements.

- Perform literature review and report narrow bands and HVIs which are sensitive to chlorophyll and N.
- Calculate HVIs which are either chlorophyll or nitrogen related as reported in the literature.
- Establish quantitative model for monitoring canopy nitrogen and chlorophyll content for potato, maize and grassland.
- To validate the model and test ability of the model to predict nitrogen or chlorophyll in other locations.

1.4 Research questions

The following research questions will be addressed in order to meet the objectives of this study:

1. Are broadband VIs able to assess nitrogen and chlorophyll content for the specific crops and locations?
2. Which narrow bands have highest correlation with nitrogen and chlorophyll in potato, maize and grassland according to the literature?
3. Which HVIs best characterize nitrogen and chlorophyll with the specific crops?
4. Are the results crops specific?
5. Is the model valid to predict biochemical contents in other locations?

1.5 Thesis outline

The thesis report consists of 6 chapters. So far in Chapter 1, context and background of the research were discussed and the problem definition was introduced. In Chapter 2 the physical properties of light which determine its behaviour in interaction with vegetation and the mechanics of photosynthesis will be explained. Consequently, how these processes can be quantified with remote sensing techniques. Chapter 3 takes a closer look at the methodology using the background research from Chapter 2, while chapter 4 will discuss the achieved results with the proposed methods. The gathered results are discussed in broader context and linked to scientific literature in chapter 5. Finally, in chapter 6 the conclusive remarks and the recommendations for further study are synthesised.

2. Theoretical framework

In this chapter, we will take a closer look at the theoretical framework of this study. Section 2.1 presents the physical and physiological basis of remote sensing. Section 2.2 discusses the plants characteristics and studies dealing with remote sensing of vegetation. Section 2.3 will discuss the bands found to be sensitive to chlorophyll and nitrogen in other studies. Section 2.4 discusses the core of the current study: the retrieval approaches for nitrogen and chlorophyll.

2.1 Physical and physiological basis of remote sensing

The domain of optical remote sensing used for vegetation monitoring is usually between 400 and 2500 nm range in the electromagnetic spectrum. It can be divided into three spectral regions of interest : Visible region (VIS) in the range between 400- 700 nm; near infrared region (NIR) between 100 – 1100nm; and short wave infrared (SWIR), between 1100 and 2500 nm. These regions are characterized by specific light-vegetation interactions, which are influenced by leaf biochemical and structural canopy characteristics. The amount of radiation reflected by vegetation in the photosinthetically active radiation region (PAR) of the electromagnetic spectrum (400- 700 nm) is regulated by pigment absorption within leaves. Chlorophyll-a and chlorophyll- b absorb the greatest proportion of radiation and provide energy for the reactions of photosynthesis, while the carotenoids protect the radiation centres from excessing light back to the environment (Blackburn 1999).

The absorbed solar radiation by a leaf is a function of the photosynthetic pigment content. Thus chlorophyll content can directly determine the photosynthetic potential and primary production (Filella, Serrano et al. 1995). Chlorophylls A and B are essential pigments in the light-dependent reactions of photosynthesis for the conversion of light energy to chemical energy. One molecule of chlorophyll absorbs one photon and loses one electron, which leads to reduction of NADP to NADPH and synthesis of ATP. These compounds take part next within the Calvin cycle where the enzyme RuBisCO captures CO_2 in bounds of sugars. It has been estimated that RuBisCO accounts for a quarter of leaf nitrogen and it is probably the most abundant protein of the world (Portis Jr and Parry 2007). Chlorophylls can give an indirect estimation of the nutrient status, because part of the leaf N is incorporated in chlorophyll. Despite the relatively low nitrogen content of chlorophyll (4 mol /mol $^{-1)}$, strong correlations are found between chlorophyll and nitrogen in green leaves, because of the large amount of protein that complexes the photosynthetic pigment in vivo (Evans 1989). Furthermore, leaf chlorophyll content is indicative for health status evaluation and is closely related to plant stress and senescence (Merzlyak, Gitelson et al. 1999; Carter and Knapp 2001).

Carotenoids usually are represented by two type of carotens (α and β) and xantophyles (lutein, zeaxantin, violaxantin, antheraxantin and neoxantin), which have strong absorption in the blue region of the spectrum. Several specific and important physiological functions have been attributed to carotenoids because of their unique physiochemical and photophysical properties: structural role in the organisation of photosynthetic membranes, participation in light harvesting, energy transfer (Lichtenhaler, 1989). Chlorophylls absorb the solar radiation mainly in the blue (~450nm) and the red (~680nm) wavelengths, whereas carotenoids have an absorption feature in the blue region overlapping with the chlorophyll (Hatfield, 2008). The red absorption peak is solely due to the presence of chlorophylls but low concentrations might saturate at the red region, thus making it insensitive to high chlorophyll contents. Longer wavelengths red edge (~ 700 nm), or shorter (~550) wavelengths in the green

part of the spectra are therefore preferred because the reflectance id more sensitive to moderate to high chlorophyll content (Hatfield, 2008).

In the NIR wavelengths (700-1100nm), the high reflectance is caused by the internal leaf structure. Between the red and near infrared wavelengths, the leaf reflectance is associated with the transition from chlorophyll absorption to leaf scattering (Knipling 1970). The position referred to this point is the so called red edge. The red edge is the maximum slope of the reflectance spectrum between 650 and 800 nm. An increase of amount of chlorophyll in the canopy, results in broadening of the red absorption feature and shifting the red edge position towards linger wavelengths (Curran 1989). The SWIR (1100- 2500nm) is dominated by water absorption and minor absorption features related to other biochemicals, including nitrogen(Curran 1989) .However, determining the nitrogen content from content from reflectance measurements on fresh leaves or canopies is extremely complex due to the strong effect of water and due to impossibility to associate specific absorption features with the chemical abundance which influences the SWIR part of the spectra (Kokaly and Clark 1999).

The relationship between nitrogen supply and chlorophyll has long been observed. Since part of the leaf nitrogen is contained in chlorophyll molecules, the amount of available nitrogen determines the amount of chlorophyll formed in green vegetation, if the other requirements for chlorophyll formation such as light, iron supply, magnesium etc. are present in sufficient quantities. The nitrogen/chlorophyll relation can be influenced also by environmental conditions (nutrients and water stress), leaf position of the canopy, temperature, and growth stage and most important it is crop specific (Hatfield, Gitelson et al. 2008) .Since nitrogen availability influence chlorophyll production, which actually produces changes in leaf and canopy spectra, reflectance can be used to assess both nitrogen and chlorophyll concentrations.

2.2 Investigated plants

The investigated plants in this study are potato (*Solanum sp*.), grass (*Loium sp*.) and maize (*Zea sp*.) (Figure 2). Most of the recent literature for crops focuses on nitrogen status assessment of rice (*Oriza sativa*), maize (*Zea mays*) and wheat (*Triticum sp*.) (Hansen and Schjoerring 2003; Hatfield, Gitelson et al. 2008; Chen, Haboudane et al. 2010) as they are the most common produced cereals worldwide and are staple food for the global population . Other crops that have been relatively widely studied are cotton, sorghum, sunflowers and sugarcane. Potato has been also mentioned in recent studies (Thenkabail, Smith et al. 2000; Jain, Ray et al. 2007; Herrmann, Pimstein et al. 2011; Clevers and Kooistra 2012), but the number of studies focusing on potato is rather limited. Leaf scale measurements have been used to understand the physiological processes governing plant growth and to estimate leaf nitrogen content as an indicator of crop nutritional status (Curran 1989). The use of canopy reflectance has shown good results for deriving crop nitrogen status for precision agriculture and yield estimation.

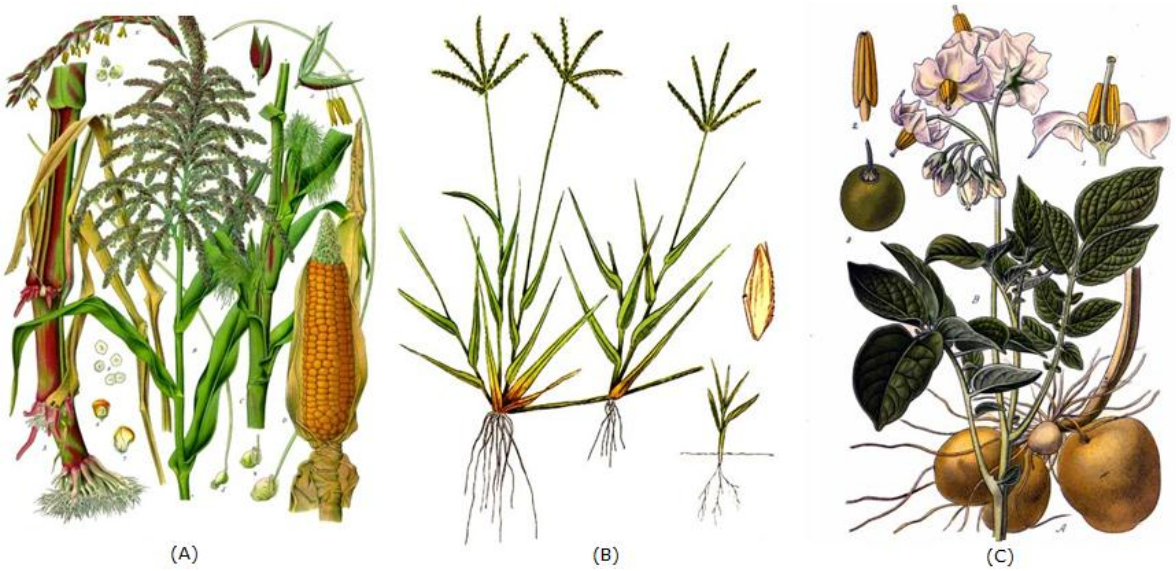


Figure 2. Investigated Plants (A) *Zea sp.* (B) *Lollium sp*. (C) *Solanum sp*.

Some authors have more specifically addressed the issue of estimating photosynthetic content concentrations as indicators of physiological vegetation conditions due to their direct relationship with reflectance (Haboudane, Miller et al. 2002). Principally statistical approaches such a simple linear and non- linear regression and multiple linear regression have been used with calculation of vegetation indices. Partial least square regression has also been shown to be promising (Hansen and Schjoerring 2003). Finally, Artificial Neural Networks were found suitable for crop nitrogen assessment (Yi, Huang et al. 2007). Concerning pastures, the majority of studies have been performed in monoculture pastures, mainly bermudagras (*Cynodon dactylon*) (Starks, Zhao et al. 2008) and ryegrass (*Lollium sp.) (Mutanga, Skidmore et al. 2003)* besides their economic relevance, monoculture pastures are particularly suited for reflectance spectroscopy experiments, since their canopy structure is vertically and horizontally homogeneous compared to mixed species and canopies.

2.3 Review of bands sensitive to nitrogen and chlorophyll content

Thenkabail 2012 performed exhaustive literature review regarding hyperspectral narrowbands in the study of vegetation and agricultural studies. This general term includes monitoring of biochemical and biophysical properties; water content and sensitivity; physiological properties. The authors identified 28 optimal bands which are based on frequency of occurrences in numerous studies discussed later. For the purpose of this study only the bands which are sensitive to chlorophyll and nitrogen in table 3 are listed. The optimal hyperspectral narrowbands determined by (Thenkabail 2012)are based on (a) identifying redundant bands, (b)modelling by linking crop biophysical and biochemical variables, (c) identifying wavebands, through statistical and other approaches, that best separate vegetation characteristics.

These prominent wavebands include (i) two blue bands that are specially sensitive to leaf chlorophyll and senescing conditions (Thenkabail, Smith et al. 2000; Thenkabail, Smith et al. 2002; Thenkabail, Enclona et al. 2004; Thenkabail, Enclona et al. 2004); (ii) four green bands: cantered at 515, 520 nm (Thenkabail, Enclona et al. 2004; Thenkabail, Enclona et al. 2004) 525, 550 and 575 nm(Thenkabail, Smith et al. 2002; Chan and Paelinckx 2008). These bands are overwhelmingly sensitive to biochemical properties 550 nm is strongly correlated with total chlorophyll. The green band cantered at 520 nm

provides the most rapid positive change in reflectance per unit change in wavelength anywhere in the visible portion of the spectrum. The green band centred at 575 nm provides the most negative change in reflectance per unit change in wavelength anywhere in the visible portion of the spectrum. Overall, the green bands are very sensitive to plant nitrogen and pigment; (iii) two red bands:cantered at 675 nm (chlorophyll absorption maxima) and 682 nm (most sensitive to biophysical quantities and yield) (Thenkabail, Enclona et al. 2004; Thenkabail, Enclona et al. 2004; Chan and Paelinckx 2008).

The NIR part of the spectrum is highly sensitive to changes in biophysical quantities and plant structure for biochemical assessment the (iv), however red-edge bands offer more information in biochemical analysis, cantered at 700 and 720 nm (sensitive to vegetation stress) and 740 nm (sensitive to nitrogen content). The red-edge bands are especially sensitive to crop stress and changes in total chlorophyll and have the potential to form useful indices. They are also sensitive to senescing rates, chlorophyll changes, browning, ripening, carotenoids (Thenkabail, Smith et al. 2002; Thenkabail, Enclona et al. 2004; Chan and Paelinckx 2008; le Maire, François et al. 2008). The SWIR bands (v) are overwhelmingly sensitive to moisture and biochemical properties (Thenkabail, Smith et al. 2002; Chan and Paelinckx 2008), although determining the nitrogen content from reflectance measurements on fresh leaf or canopies is extremely complex, due to water absorbing features (Kokaly and Clark 1999).

Table 1. Optimal hyperspectral narrowbands recommended in the study of nitrogen and chlorophyll in agricultural crops. Taken from (Thenkabail 2011, p.28)

Wave band	Waveband centre (nm)	Importance
Blue band	466	Chlorophyll: chlorophyll a and b
	490	Senescing and loss of chlorophyll, ripening, crop yield
Green bands	515	Nitrogen: leaf nitrogen, wetland vegetation studies
	520	Pigment, biomass changes
	525	Vegetation vigour, pigment, nitrogen
	550	Chlorophyll and biomass: total chlorophyll; chlorophyll/carotenoid ratio, vegetation and nutritional and fertility level
	575	Vegetation vigour, pigment, nitrogen
Red bands	675	Chlorophyll absorption maxima: greatest crop-soil contrast is around this band for most crops in growing conditions.
	682	Biophysical quantities and yield and chlorophyll absorption.
Red edge bands	700	Stress and chlorophyll: nitrogen stress, crop stress, crop growth and stage studies
	720	Stress and chlorophyll: nitrogen stress, crop stress, crop growth and stage studies. Red shift for healthy vegetation, blue shift for stressed vegetation.
	740	Nitrogen accumulation: leaf nitrogen an accumulation. Red shift for healthy vegetation, blue shift for stressed vegetation.
SWIR bands	1316	Nitrogen: leaf nitrogen content of crops
	2173	Protein, nitrogen
	2359	Cellulose, protein, nitrogen: sensitive to crop stress, lignin and starch

2.4 Retrieval approaches

Hyper spectral data has the potential to measure the reflected radiation from many plants, thus making N assessment feasible on canopy level. The most widely used approaches for nitrogen and chlorophyll estimation are based on regressive models. They relate in situ measurements and Vegetation Indices (VIs). Vegetation indices are mathematical transformations of the original spectral reflectance that are designed to reduce the additive and multiplicative errors associated with atmospheric effects, solar illumination, soil background effects, and sensor viewing geometry (Huete 1988) . Spectral VIs can be computed by using broadband data as well narrowband data. Even though both VIs are computed using the same algebraic equations, the Hyperspectral Vegetation Indices (HVIs) have greater dynamic range and offer greater opportunity in finding the right index to predict certain biophysical or biochemical variable.

There are two significant limitations of broadband derived VIs. As mentioned before broadband vegetation indices tend to saturate beyond certain level of LAI or canopy cover. Secondly most of the indices for monitoring biochemical properties of plants are constructed with red and NIR spectral measurements and offer only a few VIs. The broadband indices are significantly correlated with crop agronomic variables, such as chlorophyll content and LAI, however a large proportion of variability in modelling biochemical quantities is not explained by broadband indices. The HVIs overcome these limitations to a certain degree. They have a larger dynamic range and the larger number of HVIs offers greater opportunity to find the right index for studying certain vegetation variables. For example the derivative analysis used for calculation of the red edge position (REP) is developed for continuous spectra. It identifies the steep transition between the reflectance in red wavelengths and the NIR reflectance. It has been recognised as good estimator of plant chlorophyll and nitrogen content (Curran, Windham et al. 1995; Richardson, Duigan et al. 2002).

The relationship between nitrogen and chlorophyll amount in vegetation needs to be further explored, because the remote sensing observations are gathering information mainly on chlorophyll content, not directly on nitrogen content. It is assumed that both are significantly correlated. Also it should be further investigate to which extend the correlations are crop and place specific and if the results from one location could be transferred to another location (Clevers and Kooistra 2012). Simple ratio (SR) and normalized difference (ND) indices have been used and are still widely used for many applications mainly due to their simplicity. These indices are usually based upon the contrast between two spectral bands. The challenge is to discover combination of two narrow bands providing essential information to the target parameters. Common approach is to calculate all possible combinations and to identify the combination that has highest coefficient of determination (R^2) with the target variable at this approach the index is valid only for the specific dataset (Thenkabail, Smith et al. 2002).

Chlorophyll and soil sensitive indices have been developed to enhance sensitivity to photosynthetic pigments and to reduce the influence of canopy architecture and the background. The Chlorophyll Absorption Reflectance Index (CARI) was proposed by (Kim et al.1994) To reduce the effect of the non-photosynthetic parts of plants. Modified Chlorophyll Absorption Reflectance Index (MCARI) was introduced by (Daughtry, Walthall et al. 2000) It was intended to reduce the combined effect of non-photosynthetic parts and soil background. (Haboudane, Miller et al. 2002) found it is still influenced by non-photosynthetic elements at low chlorophyll concentrations and proposed the Transformed Chlorophyll Absorption Reflectance Index (TCARI). The family of Soil Adjusted Vegetation Indices (SAVI) introduced by (Huete 1988) is intended to further reduce the contribution of background soil reflectance and includes Optimized Soil Adjusted Vegetation Index (OSAVI) and the Transformed Soil Vegetation Adjusted Index (TSAVI). Since neither sensitivity to chlorophyll, nor intensity of the background was

achieved with the above mentioned indices (Haboudane, Miller et al. 2002) proposed MCARI/OSAVI and TCARI/OSAVI ratios. The red edge identifies the steep transition between the absorption feature in red wavelengths and the high NIR reflectance. The Red Edge Position (REP) has been used to estimate plant N content in crops and pastures and several techniques have been proposed to locate REP. The main drawback of this model is the saturation effect at higher chlorophyll contents. Range of hyperspectral vegetation indices was tested in this study for nitrogen and chlorophyll estimation, they are listed in Table 2.

Physically based reflectance models are increasingly being applied to describe the interactions between solar radiation and the biochemical vegetation parameters. They can be numerically inverted to retrieve canopy parameters from radiometric measurements. These models can be applied only to parameters directly involved in the radiative transfer and in this case for chlorophyll rather than N. The models are very complex and computational demanding. In general statistical approaches are the simplest way to predict N and are suitable when the goal is prediction of in situ quantities then understanding the radiative process. For the purpose of this study regressive models based on HVIs are preferable to physically based models that are complex to design.

Table 2. Vegetation indices evaluated in this study.

Index	Formulation	Reference
REP	$(((R_{670} + R_{780})/2 - R_{700})/(R_{740} - R_{700}))*40 + R_{700}$	(Guyot and Baret 1988)
MTCI	$(R_{754} - R_{709})/(R_{709} - R_{681})$	(Dash and Curran 2004)
MCARI/OSAVI	$(((R_{700} - R_{670})- 0.2*(R_{700} -R_{550}))*(R_{700}/R_{670}))/(1.16*(R_{800} - R_{670})/(R_{800} + R_{670} +0.16))$	(Daughtry, Walthall et al. 2000)
MCARI/OSAVI RE	$(((R_{750} - R_{705})- 0.2*(R_{750} -R_{550}))*(R_{750}/R_{705}))/(1.16*(R_{750} - R_{705})/(R_{750} + R_{705} +0.16))$	(Wu, Niu et al. 2008)
TCARI/OSAVI	$((R_{700} - R_{670}) - 0.2*(R_{700} - R_{550})*(R_{700}/R_{670}))*3/(1.16*(R_{800} - R_{670})/(R_{800} + R_{670} + 1.16))$	(Haboudane, Miller et al. 2002)
TCARI/OSAVI RE	$((R_{750} - R_{705}) - 0.2*(R_{750} - R_{550})*(R_{750}/R_{705}))*3/(1.16*(R_{750} - R_{705})/(R_{750} + R_{705} + 1.16))$	(Wu, Niu et al. 2008)
CI red edge	$(R_{780}/R_{709}) - 1$	(Gitelson, Keydan et al. 2006)
CI green	$(R_{780}/R_{550}) - 1$	(Gitelson, Keydan et al. 2006)
NDNI	$(\log(1/ R_{1510}) - \log(1/R_{1680}))/(\log(1/ R_{1510}) + \log(1/R_{1680}))$	(Serrano, Penuelas et al. 2002)
SIPI	$(R_{800} - R_{445})/(R_{800} - R_{681})$	(Penuelas, Filella et al. 1995)
DCNI	$(R_{720} - R_{700})/(R_{700} - R_{670})/(R_{720} - R_{670} + 0.03)$	(Chen, Haboudane et al. 2010)
NDRE	$(R_{790} - R_{720})/(R_{790} + R_{720})$	(Tilling, O'Leary et al. 2007)
NDRE1	$(R_{740} - R_{705})/(R_{740} + R_{705})$	(Gitelson and Merzlyak 1994)
NDRE2	$(R_{780} - R_{705})/(R_{780} + R_{705})$	(Barnes, Clarke et al. 2000)
NDVI	$(R_{800} - R_{670})/(R_{800} + R_{670})$	(Rouse, Haas et al. 1974)
CCCI	$((R_{790} - R_{720})/(R_{790} + R_{720}))/((R_{800} - R_{670})/(R_{800} + R_{670}))$	(Barnes, Clarke et al. 2000)
WDRVI	$(0.2*R_{800} - R_{670})/(0.2*R_{800} + R_{670})$	(Gitelson 2004)

R_{XXX} refers to the reflectance factor wavelength XXX nm

3. Materials and methods

In the previous chapter the theoretical background of the thesis research has been discussed, aiming to investigate and test the opportunities of vegetation spectral indices to assess crop nitrogen and chlorophyll content in potato, maize and grassland. In this chapter the analysed data will be introduced and the pre- processing steps will be explained. The following chapter contains also detailed description of the methodology and the analysis steps, which address the research objectives and questions.

3.1 Study area

The study area is located in the South of the Netherlands in the province of Noord Braband, close to the village of Reusel (Figure 3). The APEX image is covering area of 11.4 km^2 which extends over the Dutch state border. Parts of the image cover Belgium territory. However the study area is situated only in the borders of Netherlands since the field sampling doesn't cover any areas in Belgium. Three types of agricultural crops are sampled and tested in the study area: grassland, maize and potato.

In one of the potato fields was prepared experimental design with 12 plots (30*30 m). They were supplied with different levels of fertilization. Before the planting four levels of fertilization were applied and after planting three types of treatments were performed during the growing season (Figure 4). CL treatment was the first type, there was no additional fertilization applied. In plots with TTW fertilization treatment, additional nitrogen was applied few times during the season and at MB only ones at the growing season, at a specific point when the crop vegetation closes its canopy.

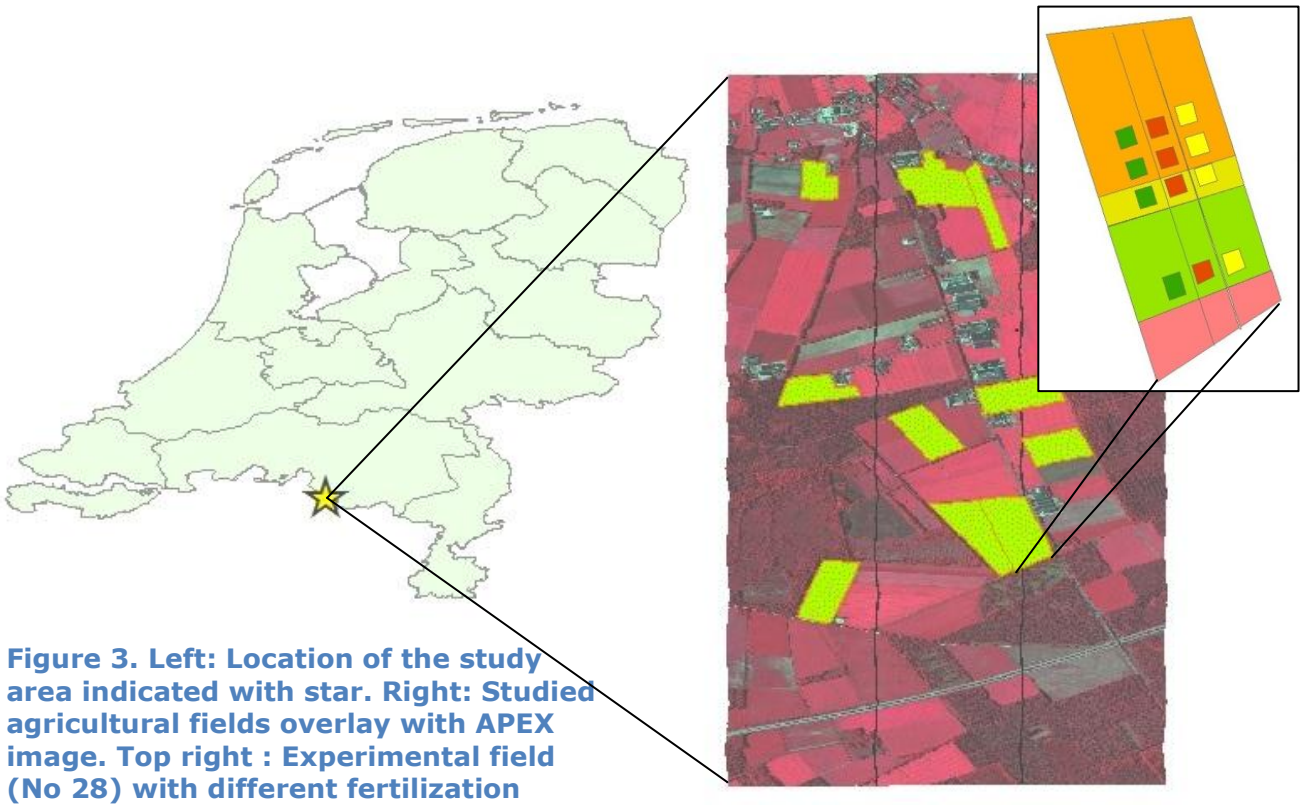


Figure 3. Left: Location of the study area indicated with star. Right: Studied agricultural fields overlay with APEX image. Top right : Experimental field (No 28) with different fertilization levels

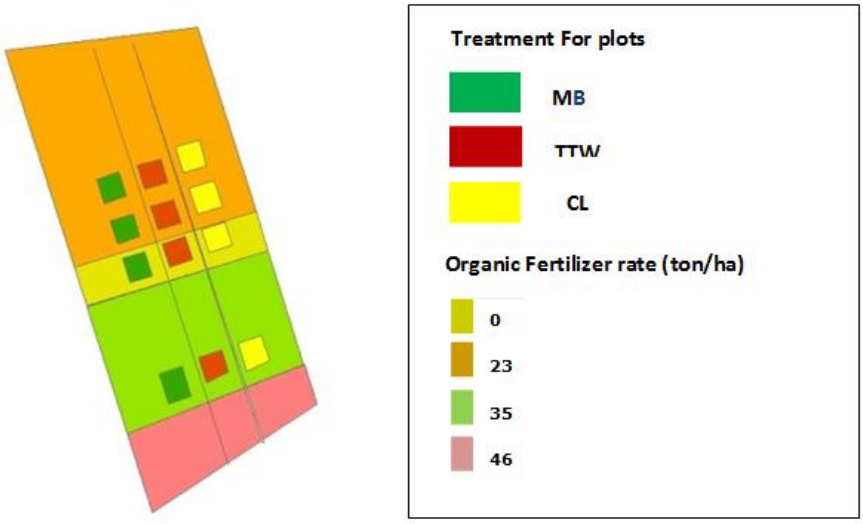


Figure 4 Experimental field No 28. On the left hand side is depicted a map representing the initial fertilization levels and plots with different treatments. On the right hand side is shown the legend and the corresponding amounts of fertilizer

3.2 Available data

3.2.1 APEX dataset

APEX is an airborne (dispersive push broom) imaging spectrometer. It is developed and constructed by a Swiss-Belgian consortium on behalf of the European Space Agency (ESA). It is intended as a future simulator and also as calibration and validation device for future spaceborne (hyperspectral) imagers of ESA. APEX sensor is operational as an advanced scientific instrument on board of an aircraft for the European remote sensing community. It records hyperspectral information in over 300 bands between ~400 nm and 2500 nm. The spectral resolution varies for different bands: over VIS-NIR range bands vary between 0.55 and 8 nm, over SWIR between 6.2 - 11 nm. The APEX imaging flight above the study area was performed by VITO on Monday 27- 06- 2011 around 17: 45 local time. VITO is an independent research and advisory organisation based in Flanders, Belgium. VITO acts as APEX co- investigator and is providing the geometrically and atmospherically corrected hyperspectral data in ENVI format. For the current study, ground measurements were performed both by WUR and BLGG research. They were used for higher level of image processing.

The geometric correction was performed by VITO's own developed C++ module and the data was projected to geographic coordinate system WGS84 with pixel size of about 7 m^2. The atmospheric correction of the acquired APEX data was done by the radiative transfer code following algorithms given in (de Haan, Hovenier et al. 1991; de Haan and Kokke 1996). Wavelength depended spectral smoothing of the data was performed to remove noise and spikes remaining after atmospheric correction. There were two bold black lines along the image. They were due to presence of wires placed on the entry slit of the airplane during the acquisition flight. Unfortunately some data were lost as the lines are going through the studied fields (Sindy Sterckx 2011).

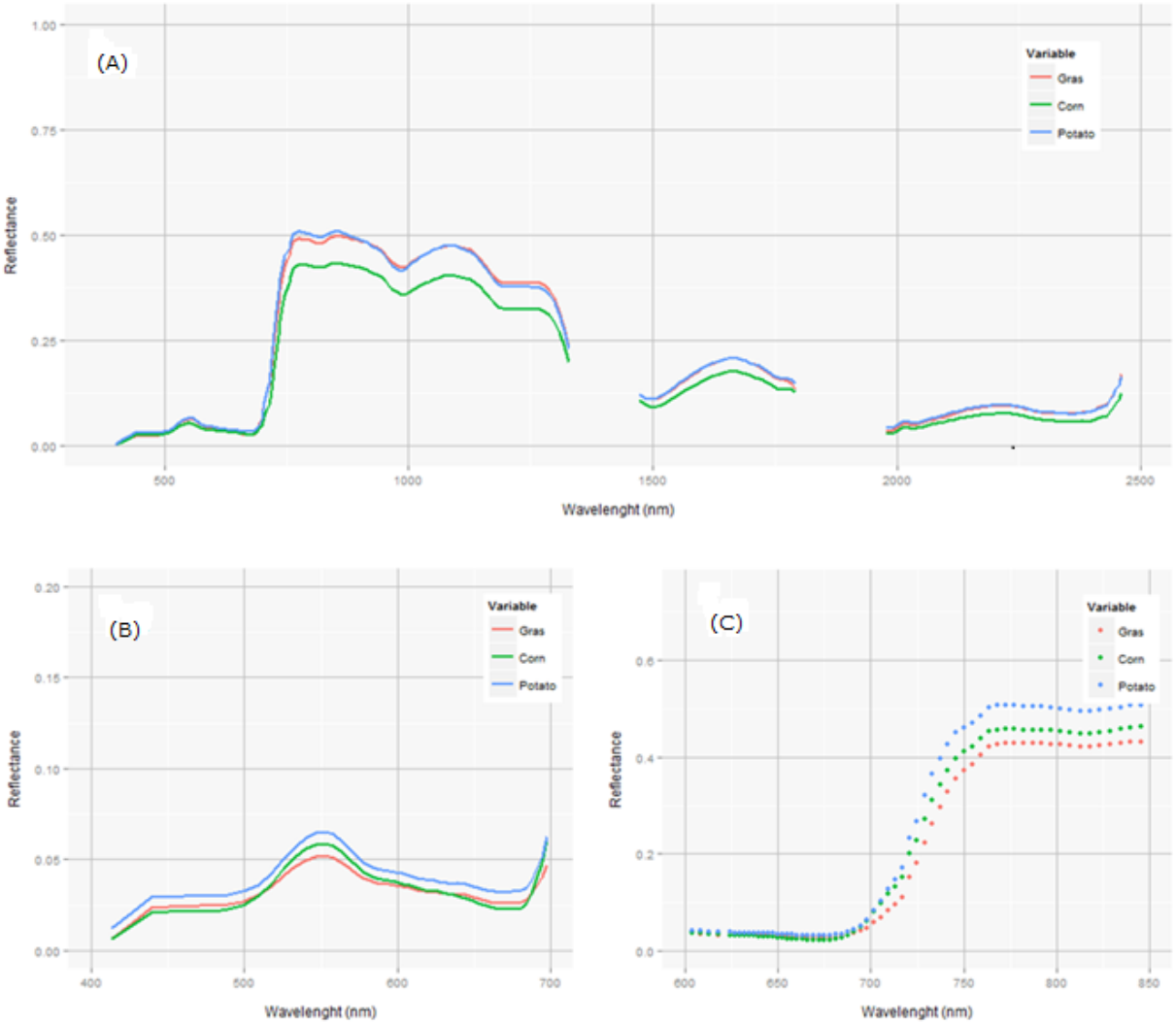


Figure 5 Averaged potato, maize and grass spectral reflectance obtained by APEX sensor illustrated for (A) along the full spectra without water absorbing bands (B) along VIS spectra (C) along red edge region

The available hyperspectral data derived from APEX consists of 288 bands in the range between 399 nm and 2461nm. These data in hundreds of bands were composed into a single hyperspectral 'data cube'. Click on any pixel will provide continuous spectrum of vegetation categories (Figure 5). The band setting and the many narrow contiguous bands along VIS, NIR and SWIR makes APEX dataset particularly suitable for vegetation and crops monitoring. At Figure 5 are depicted the averaged reflectance signatures from the studied crops. As discussed before (Section 1.2) the hyperspectral narrowband data has many benefits but also drawbacks due to the large storage volumes and long computation time required for analysis, due to the high interband correlation, which results in multiple measurements of the same quantity. To coop with these challenges, data high dimensionality can be reduced. Thereby one of the key steps in the hyperspectral data pre-processing is to indicate and to remove redundant bands from future analysis. In this study bands removal wasn't required due to calculation of HVIs, where only few sensitive bands from specific already known wavelengths were used.

With visualisation purposes Lambda (λ_1 = 399 – 2460 nm) by Lambda (λ_1 = 399 – 2460 nm) plots of R^2 values are calculated and plotted. Figure 6 shows an interband correlation of the reflection of all potato samples analysed in this study gathered by APEX. It represents a matrix where all 288 APEX bands are plotted against the same 288 bands. In each cell of the matrix was depicted the resulting R^2 between each band pair responses. In the figure least redundant bands (R^2 values < 0.01) are shown in magenta to green. Bands with higher than R^2 > 0.01 correlation are shown in white. Since the correlations above and below the diagonal mirror each other it will suffice to consider the correlations that are either below or above the diagonal of a matrix (Thenkabail, Smith et al. 2000).

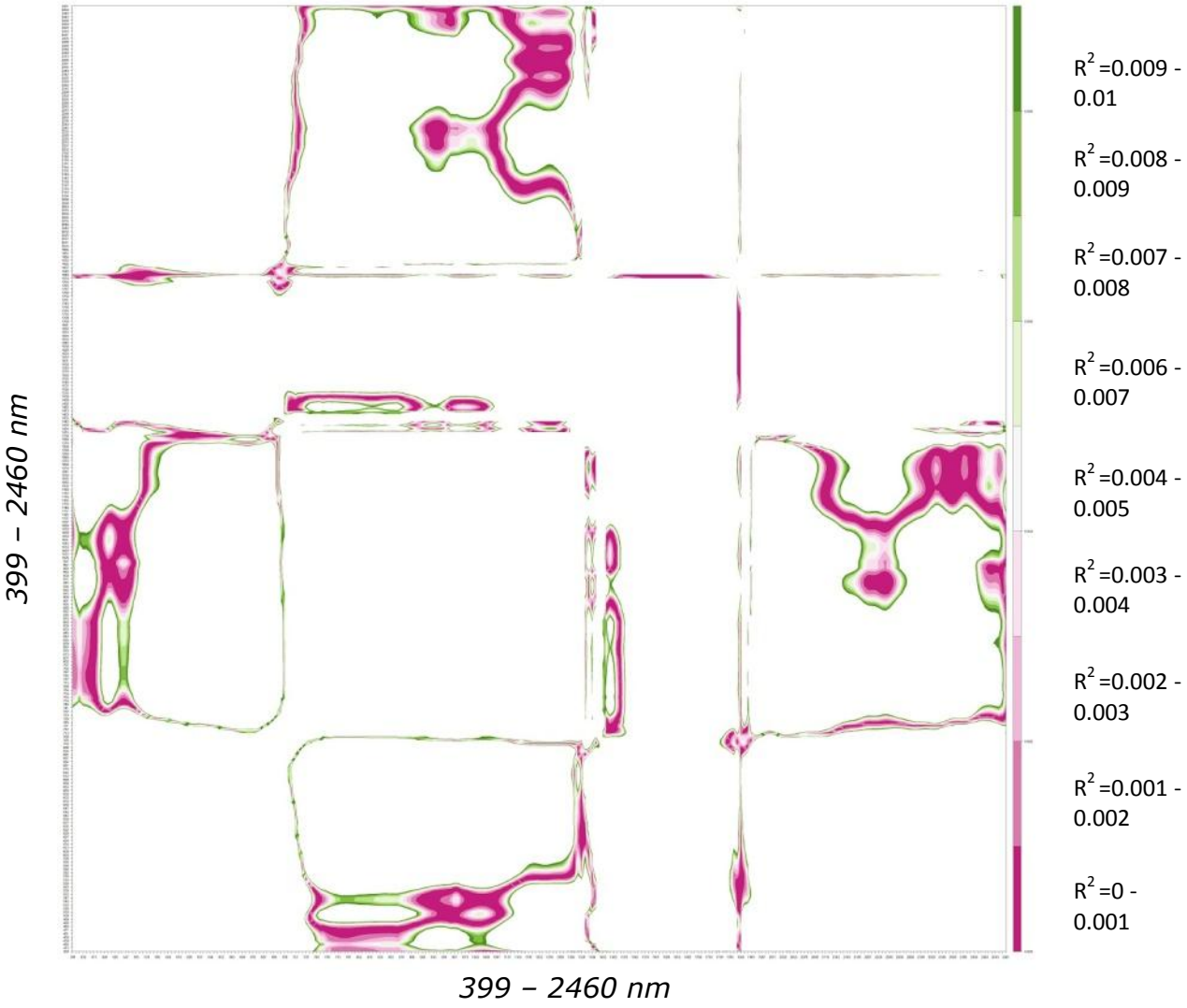


Figure 6 Redundant bands and distinctly unique bands. This matrix represents all APEX bands from potato plotted against the same bands on *x* and *y* axis. In the cells are depicted the R^2 values between bands. This Lambda (λ_1) by Lambda (λ_2) plot of APEX bands shows redundant bands (higher correlation higher redundancy) and distinctly unique bands (lower redundancy and greater the uniqueness). The R^2 values on the right side are corresponding to the colours in the plot.

3.2.2 Field data collection

For the purpose of this study selective field sampling was performed over the study area in 7 agricultural fields and in 3 pastures. The field campaign was carried out in June 2011 in two consecutive days (27 and 28 June). Samples were taken in 4 potato, 3 maize fields and 3 pastures. In each field five different locations were examined, exception was the field with the experimental fertilization levels, where 12 locations were sampled. The locations were chosen taking into consideration the plant condition in different sites of each agricultural field, estimation was done by inspecting Google Earth images and choosing suitable locations for measurements. As much diverse plant condition was targeted, thus differences in the nutrient status. The range of nutrient values is assumed to lead to better correlation with the reflection data derived from APEX.

In each of the five locations per field, two samples (A and B) were taken, close to each other. These paired samples were selected with respect of the plant condition as well, but they were aimed to be similar in contrast to the locations choice. Both A and B samples were weighted in fresh condition for above ground mass. Each A sample was analysed in BLGG laboratory, estimating dry mass (g/kg product) and nutrient quantities including total nitrogen (g/kg dry mass). The nutrient readings from A samples were be used for calculations, correlating nitrogen status and the reflectance data. B samples were weighted, but only the B samples from the experimental field with different fertilization levels (No 28) were analysed for nutrients. These B samples from field No 28 were later used for independent validation of prediction model for potato fields (Section 3.6), rest of the B samples weren't further analysed.

Every sample position was located using Differential Global Positioning System (DGPS) and projected in Rijks Driehoekstelsel (RD). The field data was organized in two separate datasets: one including biomass and nutrient characteristics and second including the positions of all samples. The chlorophyll readings are estimated using SPAD meter (SPAD models 501 Minolta corporation, Ltd., Osaka, Japan) in potato field No 28 only. The chlorophyll is measured at the same locations, where the measurements for nutrient analysis are taken, for both A and B samples which results in total 24 measurements.

3.3 Data pre- processing

In the following section the pre-processing steps and the resulting dataset will be described. The data was obtained by different sources and most of it didn't follow the same format, structure and projections. The first step in the methodology was to extract, structure, transform and match all the data together to derive the final data set required for the analysis.

3.3.1 APEX

APEX data as described (section 3.2.1) was atmospherically corrected and geometrically rectified by the data provider VITO. The assigned geographic coordinate system WGS 84 wasn't changed to the metric RD projection, because any kind of raster image transformation causes data lost, due to pixel resampling. However vector transformations, especially point dataset can be transformed to another projection with less inaccuracy. The vector dataset with all the locations of the field samples was transformed to the source APEX image projection. After the transformations of the field samples, which will be discussed in the next section, all the data were fitted together in ArcGIS and visually inspected for misplacements and distortions.

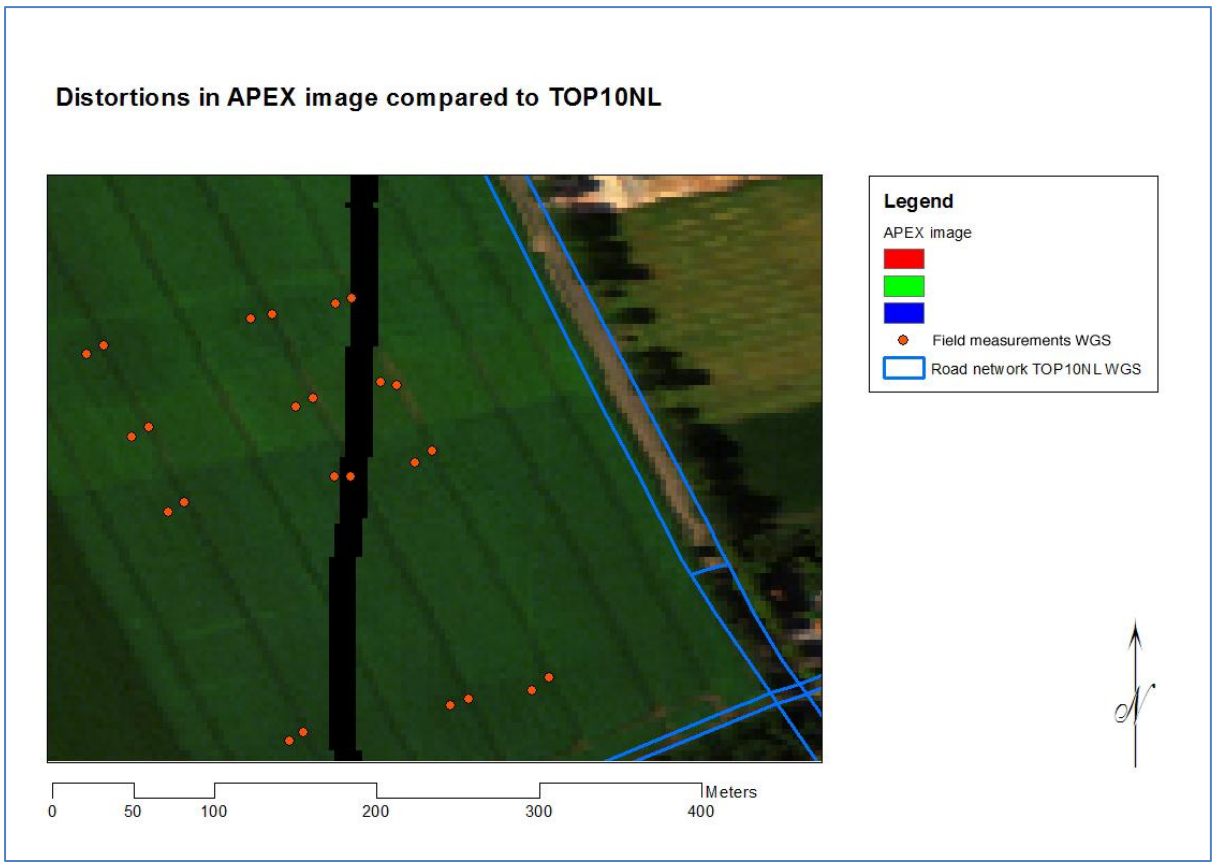


Figure 7 Distorted area in APEX image in compared to TOP10NL road network

Severe distortions were noticed in the geometry of the image, namely a rotation in the lower parts of the image. The distortion was increasing rapidly in south-east direction. The distortion is proved by comparing with other independent sources: aerial photography and NL Top 10 road network (Figure 7). Further geometrical correction of the image was not feasible with respect of time schedule and available resources. The alternative was to adapt the ground measurements to the image distorted areas, which will be described in the next section.

3.3.2 Field data

DGPS locations

During the field campaign all the sampled locations were located using DGPS, with spatial accuracy of about 50 cm. The data was projected to RD from the data provider. The dataset available was a dbf file consisting of all sample locations with an appropriate naming per sample. The dbf file was converted to shp file in ArcGIS and transformed to WGS84. For the purpose the 'project' tool from ESRI was used with geographic transformation 'Amersfort to WGS 2'. The inaccuracy of the DGPS was taken into account buffering the point locations with 1 m radius.

As already mentioned, it was noticed that some of the points didn't correspond with the expected locations overlaying the APEX image. Some of the points in the lower part of the study area were manually corrected. The main distorted field was No 28 – a potato field, which was essential for the analysis as there were 12 pairs of samples tested for nutrients and also for chlorophyll. The sampling plan in this particular field included

samples along the tractor driving path on the left and right hand side at distance of 3 m. This expert knowledge was the leading guide to match the points with the right position with respect to the image, and later on to extract the information at the right location where actually the sampling occurred. The remaining points were not manually corrected as no information was available for their locations. Small shifts were assumed not to cause large effects, unless they don't fall into trenches between rows, where soil reflectance is disturbing the results.

Biomass and nutrients

Information about the fresh biomass (FB) of each sample (A and B) in each location was available from BLGG. All the samples were weighted and then with respect to the area coverage of the field sampling plots, calculations were made to estimate the fresh biomass in kg per hectare. The nitrogen (g/kg dry matter) and dry matter (DS) (g/kg product) were derived from laboratory analysis. To make the link between nitrogen concentrations per volume and remote sensing measurements, first the nitrogen content per area should be calculated. Because the nitrogen measurements were estimated per unit dry matter, not per fresh biomass and some transformations were required first. Equations 1 and 2 were used to make the necessary calculations. After estimating the nitrogen concentrations in the corresponding units, they were merged with the buffered DGPS locations together in one dataset. This dataset was used in the further steps of data pre-processing.

$$\mathbf{DS}\ (\boldsymbol{kg/ha}) = \left(\frac{\mathbf{DS(g/kg}\ \boldsymbol{product})*\mathbf{FB}\ (\boldsymbol{kg/ha})}{\mathbf{1000}}\right) \quad \textbf{Equation 1}$$

$$\mathbf{N}(\boldsymbol{kg/ha}) = \left(\frac{\mathbf{N}(\boldsymbol{g/kg\ DS})}{\mathbf{1000}}\right) * \mathbf{DS}\ (\boldsymbol{kg/ha}) \quad \textbf{Equation 2}$$

Extracting remote sensing information

Next logical step was to join the field data via the spatial coordinates of the buffered locations to the reflection measurements gathered by the APEX sensor. The locations of field samples were with radius of 1 m^2, that means much less than the APEX pixel which was ~ 7 m^2.

Each buffered location spatially matched with 1 to 4 pixels from the image. When more than 1 pixel corresponds to a location then the average of all the matching pixels was extracted to a table, using 'raster' package in R. The merge of all the tables resulted in one table for each crop. Each table consists from:

1. Name of the Field- Sample location – indication of A or B sample
2. Crop
3. Nitrogen concentration
4. Chlorophyll concentration in field No 28 only
5. 288 bands measurements derived from APEX

With this step the data volume was decreased considerably and only the few target pixels were extracted from the massive hyperspectral image.

3.3.4 Sentinel -2 and Landsat data simulations

One of the main research objectives of this study was to test the ability of hyperspectral data analysis in estimating nitrogen and chlorophyll content and to compare with the potenential of the new generation of multispectral sensors such as Sentinel-2 and the broadly used Landsat TM. This sensor is expected to be lounched in 2014 and the ambitions are to provide band setting suiting vegetation monitoring. The main improvement over sensors as for instance Landsat TM, SPOT, DMC etc are the bands centered around the Red Edge position, which are cantered at 705 nm and 740 nm and band width of 15 nm. These wavelengths are essential for estimating biochemical quanities (Dash and Curran 2004)

Sentinel-2 bands were simulated calculating simple averaged reflectance of APEX bands over the band width of the respective Sentinel- 2 bands (Clevers and Kooistra 2012). The 6 bands from Landsat TM, without the thermal one were used for the simulation, using the same principle as for Sentinel- 2 simulations. The resulting datasets were used for further calculations including vegetation broadband indices.

Table 3 Specification of Multi Spectral Instrument (MSI) on the Sentinel-satellite system with the corresponding APEX bands used for simulations

Sentinel-2	Sentinel-2 Band center (nm)	Band width (nm)	Spatial resolution (m)	Spectral range of Sentinel-2 (nm)	APEX range used (nm) for simulation
Band 1	443	20	60	433:453	438:450
Band 2	490	65	10	457,5:522,5	461:517
Band 3	560	35	10	542,5:577,5	546:574
Band 4	665	30	10	650:680	652:677
Band 5	705	15	20	697,5:712,5	701:713
Band 6	740	15	20	732,5:747,5	733:745
Band 7	783	20	20	773:793	777:792
Band 8	842	115	10	784,5:899,5	787:894
Band 8a	865	20	20	855:875	857:869
Band 9	945	20	60	935:955	936:950
Band 10	1375	30	60	1360:1390	1366:1386
Band 11	1610	90	20	1565:1655	1573:1663
Band 12	2190	180	20	2100:2280	2103:2275

3.4 Vegetation indices

The data analysis was divided in two main parts. The first step (Section 3.4.1) was to calculate nitrogen and chlorophyll suitable HVIs based on literature review. In addition (Section 3.4.2) the same broadband VIs were calculated based on Sentinel-2 and Landsat TM simulations, derived from the hyperspectral dataset. In step two (Section 3.5) the results from the calculations were further analysed in regression models aiming to determine the best predictors (HVIs and BVIs) for nitrogen and chlorophyll content.

3.4.1. HVIs

A number of spectral indices have been developed to measure chlorophyll content and nitrogen content. They were considered as a good estimator of these properties. There are indices which better correlate with nitrogen and as well indices that better correlate with nitrogen concentrations. All the listed indices in Table 2 were calculated, using the wavebands as reported in the literature. The indices evaluated in the study can be grouped based on their characteristics. There are chlorophyll sensitive indices, nitrogen

sensitive indices and structural indices as for example the commonly used NDVI and its modification WDRVI. The last two indices were tested because they are commonly used and can be calculated for most of the existing optical sensors, not because they had proved significant correlation with nitrogen or chlorophyll. The chlorophyll indices are utilizing the bands in the red- edge position which was proved by previous research (Horler, Dockray et al. 1983; Curran, Windham et al. 1995; Dash and Curran 2004; Clevers and Kooistra 2012) to be particularly suitable for chlorophyll estimation. Because of the strong correlation between chlorophyll and nitrogen these indises are suitable for nitrogen as well. There is one more peculiar index NDNI, index claimed to be suitable for nitrogen and utilizing bands in SWIR region. SWIR bands are often mentioned in literature as sensitive directly to nitrogen concentration. The sensitivity of those indices will be compared to the sensitivity of broadband spectral indices, comparing R squared values, RMSE and RMSEP. Broad band vegetation indices were calculated based on the simulated data from previous step.

3.4.2. Broad band VIs

All the discussed indices were calculated based on data simulated from APEX spectral responses for the future Sentinel- 2 and Landsat TM. Often the indices cannot be calculated with the exact wavelengths as mentioned in the literature, because they are hyperspectral indices and require very specific bands, most of the cases not available by the multispectral sensors. For the multispectral sensors approximately close wavebands were chosen to calculate each index. As for instance NDNI was impossible to calculate with the bands available from Sentinel- 2; other indices are calculated with small shifts in the wavelengths of bands. The shifts vary from 3 nm to 14 nm.

For example the original formula of MTCI (Dash and Curran 2004)and NDVI (Rouse, Haas et al. 1974) was transformed to match with the spectral bands from Sentinel- 2 and Landsat TM. Most of the hyper spectral indices were not suitable for Landsat TM bands, only CI green, SIPI, NDVI and WDRVI, were calculated based on Landsat bands. Based on Sentinel bands all the indices except NDNI were calculated. In Equations 3 to 8 R stands for reflectance in a certain wavelength of the spectra, expressed in nanometres.

Formulations for index calculations on APEX data:

$$\mathbf{MTCI = (R_{754} - R_{709})/(R_{709} - R_{681})} \quad \textbf{Equation 3}$$

$$\mathbf{NDVI = (R_{800} - R_{670})/(R_{800} + R_{670})} \quad \textbf{Equation 4}$$

Indices calculated on Sentinel-2 bands, indicating the band centre, they look like:

$$\mathbf{MTCI = (R_{740} - R_{705})/(R_{705} - R_{665})} \quad \textbf{Equation 5}$$

$$\mathbf{NDVI = (R_{783} - R_{665})/(R_{783} + R_{665})} \quad \textbf{Equation 6}$$

Indices calculated on Landsat TM bands, MTCI cannot be calculated:

$$\mathbf{MTCI = NA} \quad \textbf{Equation 7}$$

$$\mathbf{NDVI = (R_{NIR} - R_{RED})/(R_{NIR} + R_{RED})} \quad \textbf{Equation 8}$$

The index values for each sensor were scatter-plotted against nitrogen contents (kg/ha) to provide general sensitivity examination and to obtain the R^2 (Tables 6 and 9, Section 4) values and equations for simple linear regression between each index and nitrogen content.

3.5 Regression analysis

Comparison of the results derived from previous steps namely VIs calculations were performed fitting linear and exponential regression models. Regression analysis was performed on the same way for broadband vegetation indices and also hyperspectral vegetation indices. The comparison between them was essential to test the hypothesis, that HVIs can predict better biochemical properties of crops and explain a significantly higher proportion of their variability, overcoming the limitations posed by broadband VIs. The accuracy of all indices was assessed for each crop individually. The judgment of all results was performed by comparing coefficients of determination (R^2) and calculation of root mean square error Root mean square of prediction (RMSE). In addition normalize root mean square of prediction was be calculated (Appendix 1 and 2). Scatter plots derived from the models are depicted in Appendix 4 and 5.

Root mean square error (RMSE) or also called root mean square of deviation (RMSD) is a frequently used measure of the difference between values predicted by a model and the values actually observed from the environment that has been modelled. These individual differences are also called residuals, and the RMSE serves to aggregate them into a single measure of predictive power. The RMSE of a model prediction with respect to the estimated variable *y* is defined as the square root of the mean squared error (Equation 9) where y are observed values and ŷ is modelled value at place i. The calculated RMSE values will have the same units as the estimated variable. Respectively RMSE for nitrogen/chlorophyll concentrations can be directly compared to the average values and standard deviation of observed values. The RMSE values can be used to distinguish model performance in a calibration period with that of a validation period as well as to compare the individual model performance to that of other predictive models. Non dimensional form of RMSE was be also calculated, because often RMSE isn't easy to compare with other units or crops in our case. The approach is to calculate a normalized root mean square error (NRMSE) (Equation 10). It normalizes RMSE to the range of the observed data, it can be also normalized to the mean of the observed data.

$$\mathbf{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad \textbf{Equation 9}$$

$$\mathbf{NRMSE} = \left(\frac{\mathbf{RMSE}}{\mathbf{x_{max} - x_{min}}}\right) * \mathbf{100} \qquad \textbf{Equation 10}$$

3.6 Model Validation

After the calibration of the models we had to validate the predictive ability of the models. In order to do the validation two approaches were used. For potato enough data was available to perform external validation. In potato crop, independent dataset from the available B samples from field No 28 can be spared for validation. For grass and maize samples are available only 15 samples and external validation isn't feasible, there cross-validation can be carried out. The independent validation denoted as root mean square error of prediction (RMSEP). Is defined from Equation 11 by taking a validation sample y(i) and the prediction for the validation sample based on calibration ŷ(i).

$$\mathbf{RMSEP} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_{test\,(i)} - \hat{y}_{training(i)}\right)^2} \qquad \textbf{Equation 11}$$

It should be noted that calculation of RMSE based on the calibration sample only gives too optimistic results for prediction error, because a model may fit the calibration data well without being valid for new data. Hence it is crucial to have a realistic validation sample in order to judge the true performance of the predictors and calculate RMSEP. For grass and maize samples proper validation sample weren't available and leave- one-out cross- validation were used. The formula for calculation root mean square error of cross validation (RMSECV) is the same as for external validation (RMSEP). With the difference that ŷ is a prediction of y (i) based on a n-1 remaining samples in such a way that each sample is left out exactly once, if n is the number of all samples. Error is estimated for each point left out of the model. Subsequently the resulting error is summed and averaged to a single RMSECV value.

3.7 Nitrogen maps

The regression analysis was performed in order to give an indication of vegetation indices able to give a good of prediction of nitrogen content for the investigated crops. The judgement of the best fit can be done based on R^2 and RMSE values. However more important is not how model fits the available data but to which extent it is able to predict nitrogen content for external samples. The validation of the models and the resulting RMSECV and RMSEP values were very important because based on them can be decided which estimator to use for upscaling the model and to transfer the regression coefficients to other locations. The best models were assessed using both R^2 and RMSEP values.

The functions derived from the best performing regression models with relatively low error of the prediction were utilized for calculation of nitrogen maps, to show the spatial variability of nitrogen over the fields. The maps were calculated in two steps. The first step was to calculate vegetation index map using the spectral band values of the original APEX image. In the second step nitrogen map was calculated using the coefficients derived from the regression analysis. The calculation of the nitrogen maps was executed for potato fields, mainly because the knowledge about the fertilization levels in the experimental field, which provided the opportunity for visual comparison and inspection of the emerging patterns.

4. Results

The results obtained from the regression analysis will be discussed in this chapter. Firstly the results obtained for the grassland sites will be presented and secondly the results from the potato sites. The results for maize site will not be discussed in this chapter, as all the vegetation indices regressed against nitrogen content resulted in statistically insignificant models. However, the summary statistics for maize is presented in Appendix 3.4; results for maize regression models are presented in Appendices 1 and 2; the scatter plots are depicted in Appendix 4. Results for the relations between vegetation indices and nitrogen at grassland (Section 4.1) and potato (Section 4.2) will be presented and compared between the three types of sensors investigated in the study. In addition, the results of the relationships between VIs and chlorophyll will be presented for the potato fields (Section 4.3). The comparison will be followed by the results from model validation and the corresponding RMSEP values. Finally nitrogen maps derived from the APEX image are presented to show the spatial variability of nitrogen for the potato fields (Section 4.4).

4.1 Performance of vegetation indices in grassland for nitrogen estimation

For grassland, three fields were examined. Each field was sampled in 5 locations, resulting in a total of 15 plots. During the field campaign, fresh biomass and nutrient content were analysed for each plot. From all the measured nutrients the most prominent for plant development is nitrogen. Following Equations 1 and 2 (Section 3.3.2) the total nitrogen content was calculated in kilogram per hectare. This unit is often used in agricultural studies and by farmers to calculate amounts for field fertilisation practices. In the case of grassland plots, the range of nitrogen content varies between 23.77 kg/ha and 100.64 kg/ha. Other summary statistics regarding the grass samples are listed in Table 5. For grassland samples no measurements of chlorophyll are available and the results are shown only for nitrogen content.

An overview of the R^2 values of the relation between nitrogen and the selected 17 VIs for both linear and exponential models can be found in Table 6. Indicated with an asterisk are the corresponding significance levels from the F statistics. More detailed information about other model diagnostics can be found in Appendix 1 and 2, including R^2 adjusted, RMSE, F statistics, and the exact probability values. The model diagnostics are calculated for all the investigated indices and their correlation with nitrogen content based on the spectral information delivered from the APEX image and the simulated reflectance for Sentinel-2 and Landsat TM (Section 3.3.4). The band setting of Sentinel-2 and Landsat TM didn't allow calculation of some of the hyperspectral vegetation indices. The number of vegetation indices calculated re respectively: 17, 16 and 4 (Table 6).

4.1.1 Hyper spectral vegetation indices based on APEX data

To examine the relationship between nitrogen content as a quantitative outcome and a vegetation index as a single quantitative explanatory variable, simple linear regression was considered as the most simple and prominent approach. To investigate non- linear relationships also exponential models were calculated, using log transformation of the independent variable and linear fitting. The index providing the strongest linear relationship with nitrogen content was REP ($R^2 = 0.612$, significant at $p < 0.001$).

Table 4 Summary statistics for the 15 plots of the grassland sites (harvested in June 2011)

Grass	Minimum	Maximum	Mean	Standard deviation	Coefficient of variation (%)
Fresh weight (kg/ha)	2840,00	23640,00	11026,67	7896,44	71,61
Dry weight (kg/ha)	445,88	3049,56	1436,34	992,05	69,07
N content (kg/ha)	23,77	100,64	54,06	27,08	50,10

Figure 8 shows that the three grassland fields were in a different range of development but that the combined dataset shows a significant correlation. The exponential model performed with a higher value of the coefficient of determination (R^2 = 0.713, $p < 0.001$). The strong non- linear relationship was expected and reported before in literature. Clevers and Kooistra 2012 found a strong non-linear relationship based on PROSAIL simulations (Figure 10) in broad range of REP values between 670 and 740 units of REP (R^2 =0.92). The strong linear relationship gained in the current study could be explained by the relatively short range of REP values between 720 and 722 of measured grassland canopy. In this narrower range a linear relation could be more prominent. The MTCI performed comparably well to REP for both linear and the exponential model (Table 6). The two indices are designed to locate the red-edge-position. MTCI was proposed as a better index then REP, because of its sensitivity to a wider range of chlorophyll and respectively nitrogen content. MTCI is an index based on the specifications of MERIS sensor (Dash and Curran 2004).

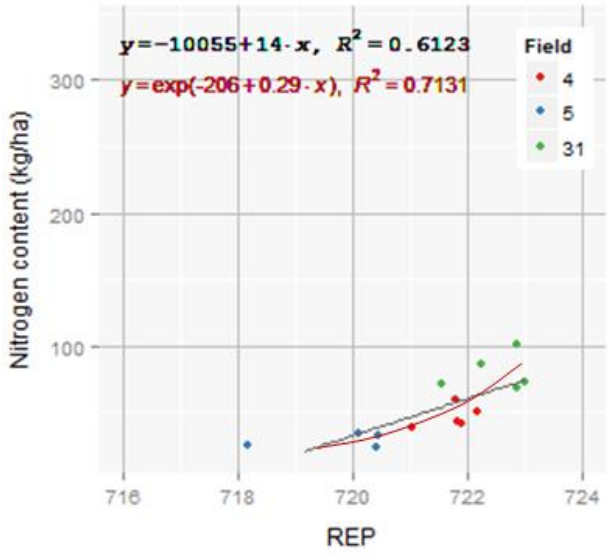


Figure 8 Relationship between REP and nitrogen content for the grassland sites.

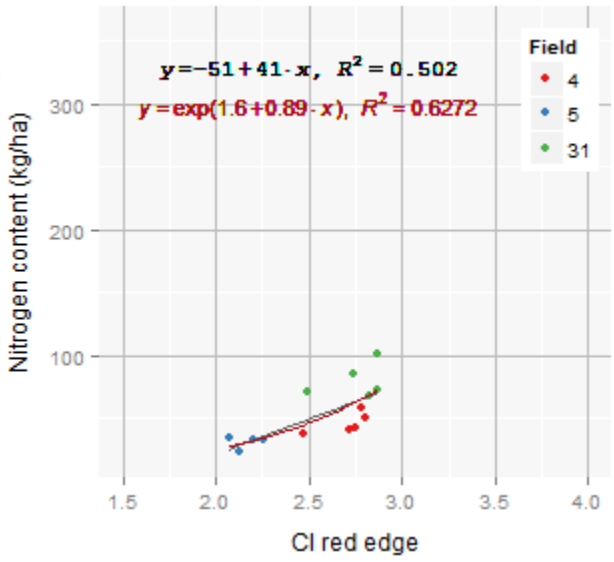


Figure 9 Relationship between CI red edge and nitrogen content and for the grassland sites.

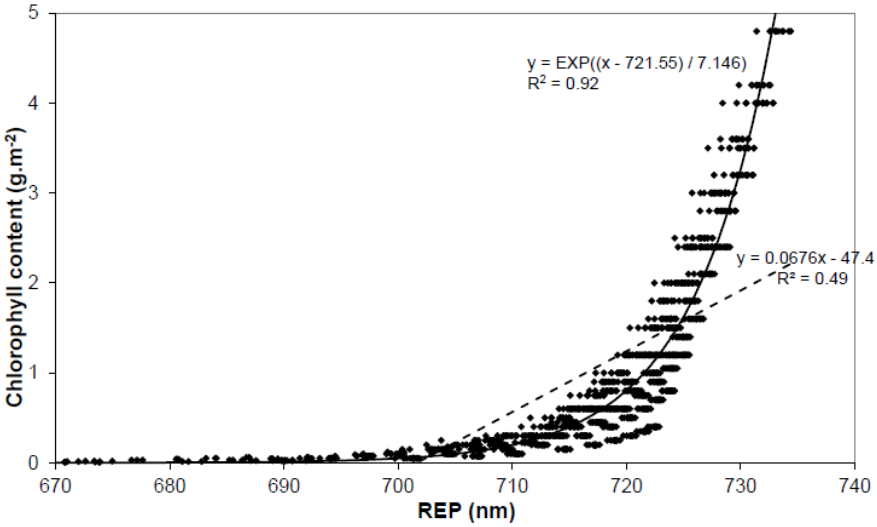


Figure 10 Relationship between REP and chlorophyll content (PROSAIL simulations) The solid line shows the exponential fit and the dashed line shows a linear fit. (taken from Clevers, Kooistra 2012)

The model using MCARI/OSAVI ratio index suggested by (Daughtry, Walthall et al. 2000) was insignificant despite it is a chlorophyll model and insensitive to confounding factors as a soil background. Reason could be the state of the grasslands, where no soil fraction was exposed due to the canopy structure. Similar results are obtained for TCARI/OSAVI. The modified ratios of MCARI/OSAVI and TCARI/OSAVI proposed by (Wu, Niu et al. 2008) including bands at 705 and 750 nm were stronger correlated with nitrogen, confirming results from previous studies (Clevers and Kooistra 2012) Furthermore, the MCARI/OSAVI red edge index explains in the exponential model the largest portion of variation in the response variable, with R^2= 0.736, significant at p <0.001. The chlorophyll sensitive indices is CI green and CI red edge, showed strong correlations with nitrogen, stronger for the index including the red edge (Figure 9), which also showed stronger exponential behaviour (R^2=0.627, p <0.001), than linear (R^2=0.501, p <0.01)

Another group of indices including red edge bands was the group of 'normalized difference red-edge' (NDRE or red-edge NDVI). Three versions were evaluated in this study as reported in the literature: NDRE 1 using 740 and 705 nm (Gitelson and Merzlyak 1994) NDRE 2 is utilizing 780 and 705 nm (Barnes, Clarke et al. 2000) and NDRE using 790 and 705 nm (Tilling, O'Leary et al. 2007). The best prediction in terms of R^2 was achieved by the last version mentioned, both for linear and for the exponential fit of the regression model. Two structural indices were also tested for nitrogen prediction despite the fact they are not reported to be sensitive to plant pigments. NDVI is one of them and is broadly used in broad range of remote sensing applications. However, the regression model for estimation of nitrogen content showed insignificant results (Figure 12). The ratio of NDRE and NDVI index (CCCI) was proposed by (Barnes, Clarke et al. 2000). This ratio (Figure 11) showed a strong linear correlation with nitrogen (R^2 = 0.604, p <0.001) and stronger exponential (R^2 = 0.717, p <0.001). Three more indices were tested, which are more peculiar in respect to the indices mentioned before. The Structure Intensive Pigment Index (SIPI) is an index developed by (Penuelas, Filella et al. 1995). It was intended as a biochemical index, sensitive to chlorophylls and carotenoids, exploiting bands in NIR, blue and red part of spectra. The regression resulted in insignificant linear model and exponential model with relatively low R^2 value. The normalized difference nitrogen index (NDNI) is the only index proposed which utilises bands in SWIR wavelengths. They are claimed to be sensitive to

nitrogen contents and more specifically to the first and second overtone to N-H band vibration. NDNI includes two bands in this region, namely 1510nm and 1680nm. Unfortunately none of the models showed any statistical significance (Table 6).

Table 5 Overview of R2 values of the linear and exponential relationships between indices and nitrogen content for grassland. For Sentinel-2 and Landsat TM not all values are given as they could not be calculated due the limited number of bands of the sensors.

Grass	APEX		Sentinel_2		Landsat TM	
Index name	R^2 linear	R^2 exponential	R^2 linear	R^2 exponential	R^2 linear	R^2 exponential
REP	0,612***	0,713***	0,606***	0,717***		
MTCI	0,599***	0,711***	0,525**	0,642***		
MCARI/OSAVI	ns	ns	ns	ns		
MCARI/OSAVI $_{RE}$	0,607***	0,736***	0,597***	0,726***		
TCARI/OSAVI	ns	0,339*	ns	ns		
TCARI/OSAVI $_{RE}$	0,416**	0,546**	0,331*	0,451**		
CI red edge	0,501**	0,627***	0,487**	0,612***		
CI green	0,348*	0,472**	0,355*	0,478**	0,29*	0,407*
NDNI	ns	ns				
SIPI	ns	0,348*	ns	0,291*	0,276*	0,386*
DCNI	ns	ns	ns	ns		
NDRE	0,523**	0,646***	0,65***	0,501**		
NDRE1	0,363*	0,493**	0,377*	0,564**		
NDRE2	0,422**	0,551**	0,439**	0,755***		
NDVI	ns	0,316*	ns	0,316*	ns	0,313*
CCCI	0,604***	0,717***	0,712***	0,801***		
WDRVI	ns	0,318*	ns	0,319*	ns	0,315*

Notes: *p < 0.05; **p < 0.01; ***p < 0.001

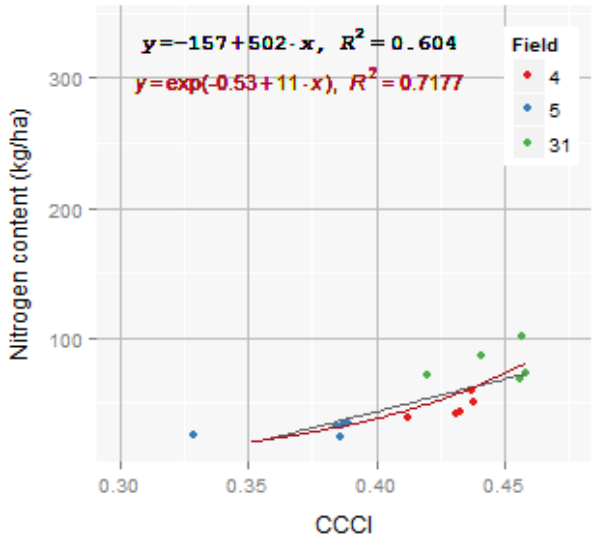


Figure 11 Relationship between CCCI and nitrogen content for the grassland sites.

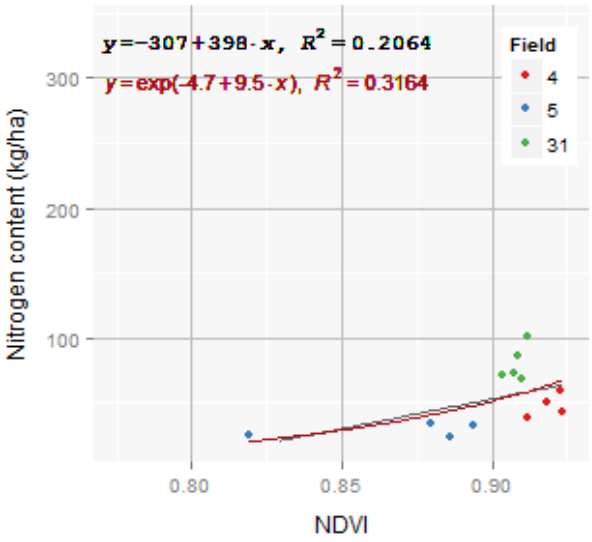


Figure 12 Relationship between NDVI and nitrogen content for the grassland sites.

4.1.2 Performance of VIs based on Sentinel- 2 and Landsat TM for grassland

The simulated Sentinel- 2 and Landsat TM band reflectance values don't suffice calculation of all the mentioned indices investigated (Table 2, Section 2.4). However most of the vegetation indices can be calculated using Sentinel-2 simulations, with only one exception, namely the NDNI, because the required bands in the SWIR are not available from Sentinel-2. Landsat did not provide a sufficient number of bands so only 4 indices were calculated (SIPI, CI green, NDVI, WDRVI). Results from the regression analysis of the available indices are also summarized in Table 6 in terms of R^2 values with the corresponding probability value derived from F test marked with an asterisk. They are compared with the values of the coefficients of determination derived from regressions based on APEX based hyperspectral VIs.

The results for the grassland indices, based on Sentinel-2 band positions, all had similar performance in terms of R^2 values to the indices based on APEX. Some of them (REP, MTCI, MCARI/OSAVI red edge, TCARI/OSAVI red edge, CI indices) had slightly less accurate predictions of the linear models. Surprisingly the group of 'Normalized Difference Red-Edge' NDRE indices performed with even higher accuracy judged on R^2 values, where the CCCI achieved the impressive R- squared for linear model of 0.712; for the exponential $R^2 = 0.801$, both significant at $p < 0.001$.

The results achieved from Landsat regressions performed with relatively lower accuracy than APEX and Sentinel-2 (Table 6). This could be expected taking into account the lower spectral resolution of the broad band sensor. Exception in this respect makes SIPI which attain statistically significant results for linear ($R^2 = 0.276$, $p < 0.05$) and exponential model ($R^2 = 0.386$, $p < 0.05$). The relationship isn't strong but is better than the relationship gained by APEX and Sentinel-2.

4.1.3 Validation of grassland models

For validation of nitrogen prediction models in grassland a leave one out cross validation approach (LOOCV) was adopted. The results in terms of RMSECV values are summarized in Table 7. The results from the validation of maize fields are also available in Appendix 1 and 2, however they all weren't significant. The LOOCV is a method suitable for datasets, which are relatively small (15 grass samples). RMSECV was calculated for the 15 iterations and then averaged for all the cases. The results from LOOCV in terms of RMSECV showed slightly larger error then the RMSE from the calibration model using all the samples for its prediction (Appendix 1 and 2).

The lowest error for a linear model was achieved by MTCI with 17.3 kg/ha. This value can be compared to the average value of standard deviation value measured for nitrogen content of 27.08 kg/ha and average value of 54.06 kg/ha (Table 5). The corresponding value of the exponential model was 14.9 kg/ha. The REP achieved very similar results as for linear model RMSECV = 17.4 kg/ha and RMSECV =14.6 kg/ha for the exponential model. The indices from the NDRE group resulted in larger errors, which is in accordance with the estimated goodness of fit diagnosed by R^2 values. The prediction error estimated by the indices based on simulated data, were slightly larger than the original indices based on APEX. Structural indices as NDVI and WDRVI had the largest error values and were both not insignificant for linear model ($p > 0.5$) but significant for the exponential model.

Table 7 Overview of RMSECV values of the linear and exponential model validation in grassland. Leave one out cross validation.

Grass	APEX		Sentinel-2		Landsat TM	
Index name	RMSECV linear	RMSECV	RMSECV	RMSECV	RMSECV	RMSECV
REP	17.437	14.608	17.844	14.583		
MTCI	17.322	14.904	18.788	16.472		
MCARI/OSAVI	na	na	na	na		
MCARI/OSAVI RE	16.616	14.586	16.990	14.726		
TCARI/OSAVI	na	ns	na	na		
TCARI/OSAVI RE	19.753	18.905	20.925	20.458		
CI red edge	19.035	17.046	19.088	17.365		
CI green	20.682	20.356	20.589	20.351	21.336	21.588
NDNI	na	na				
SIPI	na	21.268	24.571	23.926	21.829	20.720
DCNI	na	na	na	na		
NDRE	19.343	16.446	15.955	13.847		
NDRE1	22.124	19.501	22.193	19.274		
NDRE2	21.620	18.409	21.559	18.082		
NDVI	na	22.116	na	22.082	na	22.123
CCCI	17.483	14.697	14.252	12.894		
WDRVI	na	22.228	na	22.179	na	22.240

4.2 Performance of vegetation indices in potato site for nitrogen estimation

Distributed over two consecutive days in June 2011, four potato fields (27 locations) were sampled twice (A and B sample) on a relatively small distance. The corresponding A samples were analysed for total nitrogen concentration, dry matter and fresh weight. B samples are only weighted for fresh biomass. Field 28 was an exception in the sampling plan, where all 12 locations were sampled, both A and B samples. In this field experimental design was conducted with four different fertilization levels, which resulted in higher variation in nitrogen content of the potato plants sampled in the field, than other fields. The summary statistics of the potato field measurements are presented in Table 8. The range of potato samples is larger compared to the grass samples, with mean value of 202.5 kg/ha and standard deviation of 77.48 kg/ha. At potato field No 28 also chlorophyll measurements from SPAD are available. The chlorophyll readings are summarized in Table 8, it should be noticed that these measurements are from one field only (both A and B samples) and are in different units than nitrogen statistics (g/m^2).

4.2.1 Hyper spectral vegetation indices based on APEX data for nitrogen estimation

At almost all potato plots spectral data was available, but because there were two wires in front the slit of the APEX sensor during the acquisition flight, regions with no data occurred in the image. These no-data regions coincided with three field measurements in the potato fields, these locations were not further analysed. Following the same methodology as for the grassland sites, hyperspectral and broadband vegetation indices were calculated. The regression analysis explored linear and also non- linear relationships. Table 9 provides a summary of R^2 values for linear and exponential estimators. The significance levels are marked with an asterisk. More detailed information is listed in Appendix 1 and 2; all the calculated scatter plots are depicted in Appendix 4. The strongest linear correlation between nitrogen and vegetation indices was found for MTCI (R^2= 0.655, significant at $p < 0.001$). However the exponential relationship yielded stronger correlation (R^2 = 0.746). Scatter plot is depicted in Figure 13. Again the REP showed a significant linear relationship for the observed values between 720 and 727 nm for nitrogen content by potatoes, but as expected even stronger exponential relationship. These two indices are exploiting bands along the red-edge point and had also a significant correlated with grassland samples.

Table 8 Summary statistics for the 27 plots (only A samples) of the potato sites (harvested in June 2011), chlorophyll readings are available in field No 28 (A and B samples)

Potato	Minimum	Maximum	Mean	Standard deviation	Coefficient of variation (%)
Fresh weight (kg/ha)	18400,00	80264,66	47950,18	17143,55	35,75
Dry weight (kg/ha)	2208,00	6020,80	4220,68	1128,35	26,73
N content (kg/ha)	81,66	356,43	202,52	77,48	38,26
Chlorophyll content (g/m2)	0.41	0.73	0.59	0.09	15,69

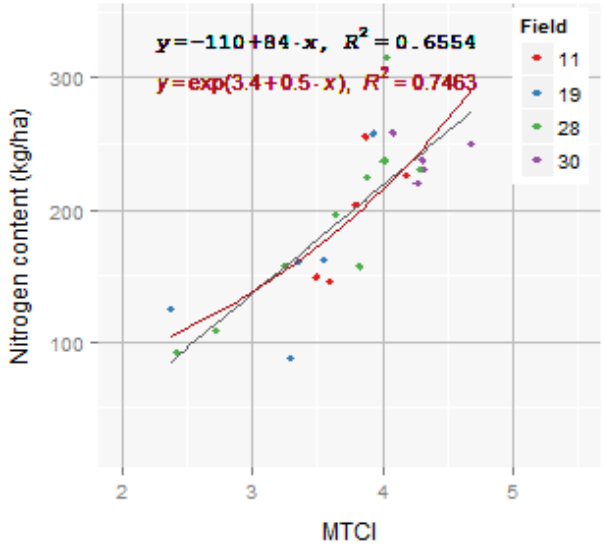


Figure 13 Relationship between REP and nitrogen content for the grassland sites.

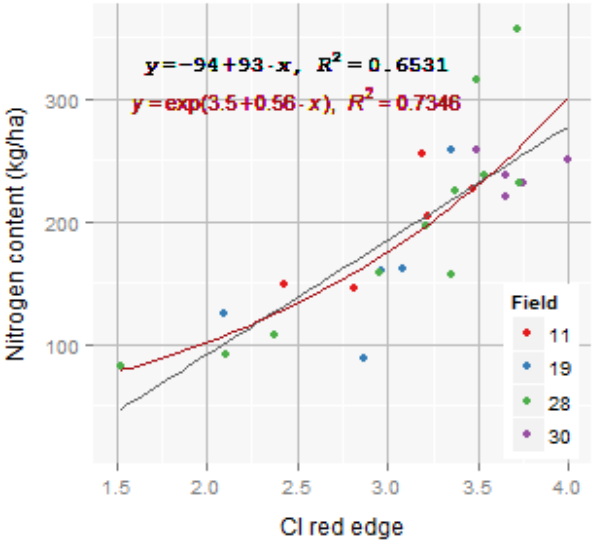


Figure 14 Relationship between CI red edge and nitrogen content for the grassland sites.

The MCARI/OSAVI ratio index resulted in a significant linear model for potato (R^2 = 0.276, p <0.01). Compared to the grassland model, which wasn't significant for both linear and exponential model. The reason could be that more data is available for potato and the larger variation due to the fertilization levels in the potato field No 28 (Appendix 3.4). However the explanatory power of this index is low. The modified MCARI/OSAVI RE version using 705 and 750 nm wavelengths gave again much stronger correlation also at higher significance level compared to the original formulation of the index. TCARI/OSAVI also performed better in the modified version proposed by (Wu, Niu et al. 2008). It resulted in a linear relationship of R^2 = 0.653, very close to the best prediction of MTCI and exponential model of R^2 = 0.733.

The chlorophyll indices CI red edge and CI green were strongly linear correlated with the nitrogen content. CI red edge (R^2 = 0.653, p <0.001) (Figure 14) and CI green (R^2 = 0.638, p <0.001), however the exponential models were better correlated. The results again prove the tendency discovered by grassland for better performance of exponential models. From the group of normalized difference red-edge indices, the best prediction was obtained from NDRE, with R^2 = 0.626 and R^2 = 0.731 for linear and exponential relation respectively. The ratio between NDRE and NDVI was also very successful in this respect (Figure 15), especially when compared to the results obtained, using NDVI and WDRVI. The last two indices resulted in significant models but the saturation effect of the red band is still present (Figure 16). NDNI and SIPI were also used in linear nitrogen content regression models, but similar to the grassland models they resulted in insignificant models or with very low prediction power.

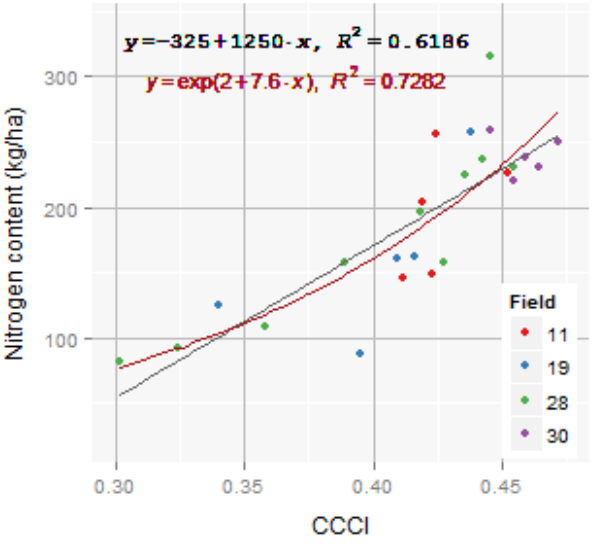


Figure 15 Relationship between CCCI and nitrogen content for the grassland sites.

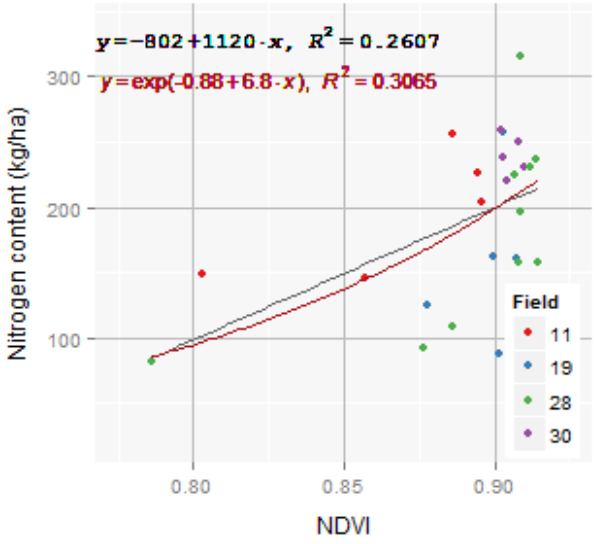


Figure 16 Relationship between NDVI and nitrogen content for the grassland sites.

Table 9 Overview of R^2 values of the linear and exponential relationships between indices and nitrogen content. For Sentinel-2 and Landsat TM not all values are given as they could not be calculated due the limited number of bands of the sensors.

Potato	APEX		Sentinel-2		Landsat TM	
Index name	R^2 linear	R^2 exponential	R^2 linear	R^2 exponential	R^2 linear	R^2 exponential
REP	0,619***	0,724***	0,621***	0,729***		
MTCI	0,655***	0,746***	0,654***	0,748***		
MCARI/OSAVI	0,276**	0,31**	ns	ns		
MCARI/OSAVI RE	0,553***	0,627***	0,505***	0,578***		
TCARI/OSAVI	0,549***	0,666***	0,437***	0,547***		
TCARI/OSAVI RE	0,653***	0,733***	0,651***	0,734***		
CI red edge	0,653***	0,735***	0,647***	0,729***		
CI green	0,638***	0,726***	0,632***	0,719***	0,609***	0,693***
NDNI	ns	0,155*				
SIPI	ns	ns	ns	ns	0,199*	0,203*
DCNI	0,312**	0,364**	0,238*	0,276**		
NDRE	0,626***	0,731***	0,645***	0,671***		
NDRE1	0,559***	0,657***	0,574***	0,69***		
NDRE2	0,577***	0,678***	0,59***	0,738***		
NDVI	0,261**	0,307**	0,277**	0,33**	0,326**	0,388***
CCCI	0,619***	0,728***	0,633***	0,727***		
WDRVI	0,273**	0,318**	0,291**	0,342**	0,344**	0,404***

Notes: *p < 0.05; **p < 0.01; ***p < 0.001

4.2.2 Performance of VIs based on Sentinel- 2 data and Landsat TM at potato site

The results from the nitrogen regression models based on indices calculated from simulated spectra are presented in Table 9 next to the results from the original data acquired by the APEX sensor. Similar to the grassland examples, all the indices were related in a linear and exponential way. In general, the exponential relationships showed higher R^2 values. Most of the indices based on Sentinel- 2 had similar performance in terms of R^2 values compared to APEX indices. MTCI, MCARI/OSAVI RE, TCARI/OSAVI, TCARI/OSAVI RE, CI had similar results and only slightly less accurate than the calculations based on APEX. Surprisingly all the NDRE indices, NDVI, REP and MTCI, showed better predictions than the models based on the original data. There are only few indices based on Landsat but the results for CI green turned out to be promising (R^2 = 609, p <0.001) for the linear model. Also the regression models of SIPI are significant at $p < 0.05$ with a higher R^2 value than the original APEX and the simulated Sentinel-2 indices.

4.2.3 Validation of potato models

Independent validation was used to assess the predictive power of models for the potato sites, because the data size, allowed to spare measurements as an independent test dataset. The samples in field 28 were all nutrient analysed, both A and B samples. The B samples were used as test data set and all A samples including the A samples from fields No 11, 19, 30, 28 were used as training data. In total 27 observations are used for model training and 10 for model testing. The results from the independent validation are the most unbiased test for the model prediction power, because the predictions of the model are compared with real values not used by the model for calibration (Arboretti Giancristofaro and Salmaso 2007). The RMSEP values for the validation of nitrogen in potato are summarized in Table 10.

The results from the independent validation resulted in much larger values of RMSEP then the calculated RMSE values. The difference between RMSE and RMSEP value in some cases was almost double. For example, CI green had a RMSEP of 77.1 kg/ha compared to a RMSE of 41.0 kg/ha (Appendix 1.1). For REP and MTCI the estimated RMSEP was respectively 82.2 kg/ha and 81.6 kg/ha from a linear model. The corresponding RMSE values for REP and MTCI were: 42 kg/ha and 40 kg/ha (Appendix 1.1). The errors estimated for the models based on simulated data had similar values as the errors gained by the original indices (Appendix 1 and 2).

After further analysis of the validation result the testing data set (B samples from the field with fertilization experiment), one sample stood out with an extremely high nitrogen content (474.4 kg/ha). Compared to the maximum nitrogen value in the calibration samples for the potato samples (356.4 kg/ha; Table 8) this was considerably higher. The very large value was caused by the large biomass and the relatively low dry matter of the sample. When replacing the values in equations 1 and 2 (Section 3.3.2) the resulting nitrogen content was extremely high. The high nitrogen sample was located in a plot with high fertilization and the high content can be explained to certain degree. As the nitrogen concentration for this test sample was outside the nitrogen range of the calibration set it was decided to also calculate the RMSEP for the corrected test set without the sample with the high nitrogen concentration. If we exclude the sample, the RMSEP values from the validation model the RMSEP value decreased rapidly (Table 10), comparable with the RMSE derived from the calibration model. The outlying observation was very influential because only 10 samples were available for the validation of the nitrogen models of the potato samples.

Table 10 Overview of RMSEP values for the independent validation of the linear and exponential model in potato. Both results for original test set (n=10) and after correction for outliers (n=9) are presented.

Potato	APEX (n=10)		APEX Corrected (n=9)	
Index name	RMSEP linear model	RMSEP exponential	RMSEP linear model	RMSEP exponential
REP	82,293	82,871	44,409	47.991
MTCI	81,514	83,210	43,139	48.554
MCARI/OSAVI	106,027	112,639	68,763	75.298
MCARI/OSAVI RE	77,961	78,571	32,850	35.851
TCARI/OSAVI	88,862	91,728	49,804	54.578
TCARI/OSAVI RE	77,498	77,576	38,185	42.272
CI red edge	78,096	78,408	39,656	44.175
CI green	77,117	76,605	38,361	42.09
NDNI	105,987	109,378	65,971	66.839
SIPI	106,877	111,503	71,837	74.6
DCNI	105,391	113,044	67,711	76.13
NDRE	80,558	80,309	41,688	44.218
NDRE1	83,476	85,200	41,120	44.718
NDRE2	82,640	83,766	41,142	44.275
NDVI	94,562	98,505	53,165	56.816
CCCI	82,723	83,519	44,607	48.396
WDRVI	93,003	96,810	51,439	55.197

4.3 Performance of vegetation indices in potato for chlorophyll estimation

The estimation of the chlorophyll content was a main objective of the study next to the retrieval of nitrogen content, because the two variables are strongly correlated and furthermore the chlorophyll is the driving force of the photosynthesis, which provides the primary productivity in all ecosystems (Vitousek, Cassman et al. 2002). The same comparison as for potato and grassland results was prepared for the diagnostic statistics form chlorophyll predictive models based on potato canopy reflectance (Table 11). All the results are summarized in Appendix 1.4 and 2.4 and all the scatter plots are depicted in Appendix 5. As expected the correlation between chlorophyll and remote sensing data is stronger than the correlation between spectral data and nitrogen. The reason is that the vegetation- light relationships are influenced mainly of the photosynthetic active pigments, such as chlorophylls and carotenoids. They absorb light in the visible spectra, namely in blue and red region. The best fit of a model correlating chlorophyll content and vegetation index was achieved by MTCI, REP, TCARI/OSAVI RE and CI indices in exponential models. The linear and exponential models gave also high R^2 values, especially TCARI/OSAVI and MTCI. The NDRE group of vegetation indices performed also with good results, for instance CCCI, with R^2 = 0.883 in linear model and R^2 = 9.14 for exponential model. The structural models as NDVI and WDRVI had a poorer fit. The lowest fit was observed between NDNI and chlorophyll, which was expected as only bands in SWIR range are used for the index calculation and they are not sensitive to chlorophyll content.

Table 11 Overview of R^2 values of the linear and exponential relationships between indices and chlorophyll content in potato samples (n = 21). For Sentinel-2 and Landsat TM not all values are given as they are not tested due the limited number of bands of the sensors.

Potato Chlorophyll	APEX		Sentinel 2		Landsat TM	
Index name	R^2 linear	R^2 exponential	R^2 linear	R^2 exponential	R^2 linear	R^2 exponential
REP	0,872***	0,906***	0,872***	0,907***		
MTCI	0,905***	0,927***	0,909***	0,934***		
MCARI/OSAVI	0,64***	0,596***	0,248*	0,199*		
MCARI/OSAVI RE	0,79***	0,839***	0,742***	0,797***		
TCARI/OSAVI	0,9***	0,929***	0,76***	0,814***		
TCARI/OSAVI RE	0,893***	0,923***	0,892***	0,921***		
CI red edge	0,881***	0,913***	0,885***	0,917***		
CI green	0,892***	0,92***	0,888***	0,916***	0,874***	0,905***
NDNI	0,261*	0,323**				
SIPI	0,362**	0,413**	0,388**	0,439**	0,46***	0,525***
DCNI	0,881***	0,853***	0,766***	0,723***		
NDRE	0,864***	0,902***	0,874***	0,899***		
NDRE1	0,817***	0,868***	0,824***	0,873***		
NDRE2	0,826***	0,875***	0,833***	0,879***		
NDVI	0,602***	0,668***	0,614***	0,68***	0,65***	0,714***
CCCI	0,883***	0,914***	0,878***	0,896***		
WDRVI	0,623***	0,688***	0,638***	0,703***	0,674***	0,737***

Notes: *p < 0.05; **p < 0.01; ***p < 0.001

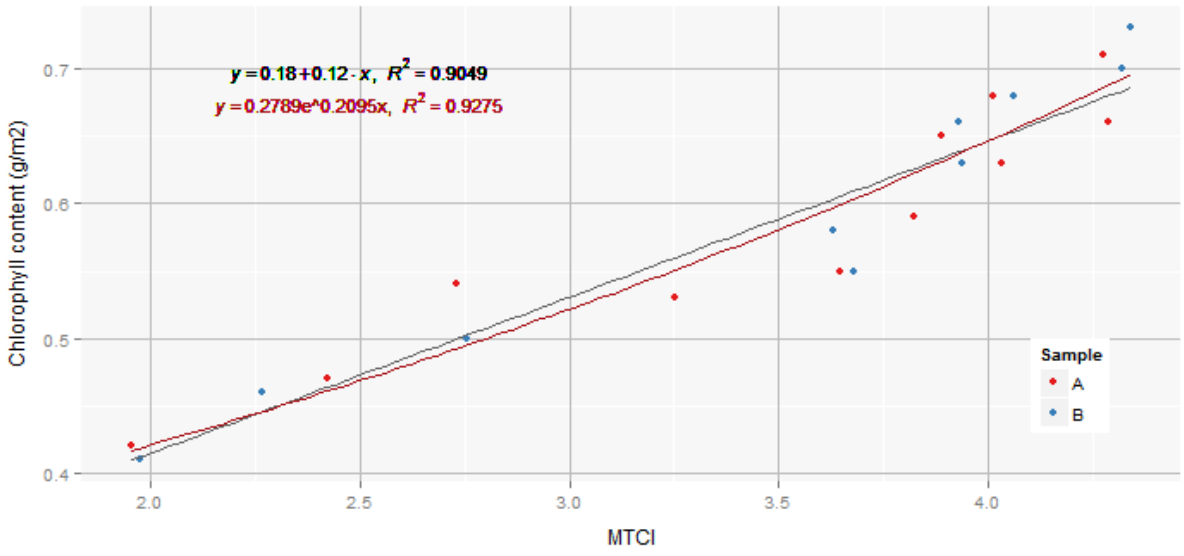


Figure 17 Relationship between MTCI and chlorophyll content for potato samples in field 28

4.4 Nitrogen maps based on the best model predictions

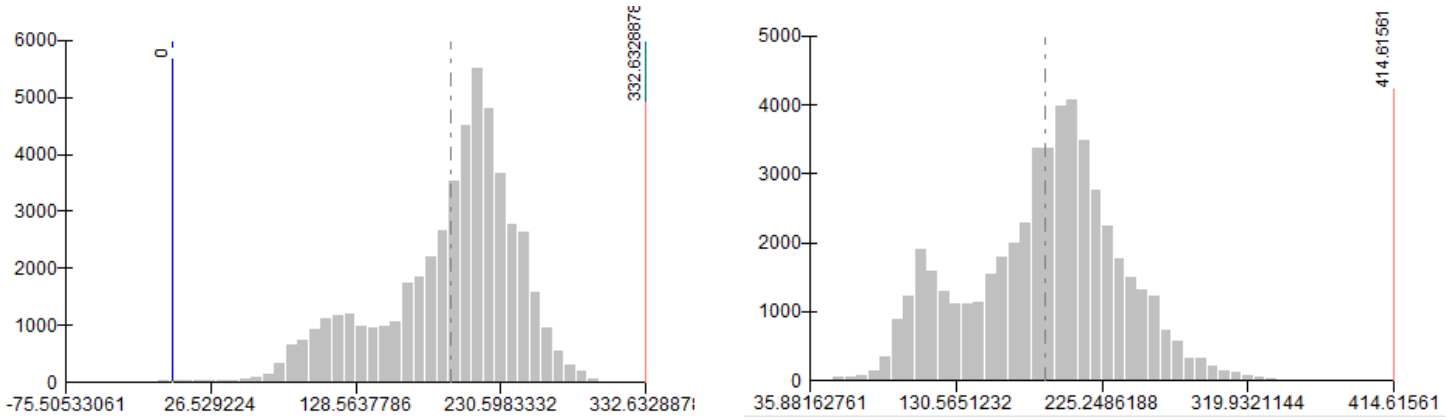
The regression models were assessed using the R^2 values, RMSEP. The best models were used for calculation of nitrogen maps for the investigated fields. MTCI was the index with highest R^2 value among other indices for potato and the RMSEP showed relatively low error of 43.1 nitrogen kg/ha (as a relative to an average measured N content of 202.5 kg/ha). The maps were calculated in two steps. First a vegetation index map using the spectral band values of the original APEX image was calculated and second a nitrogen content map was calculated, using the coefficients derived from the regression analysis. The resulting in MTCI index values were in the range between 0.4 and 5.3. In the second step the coefficients from the linear and exponential model and the MTCI maps were used for calculation of the nitrogen content map (Figure 19). The resulting map consisted of one layer, whereby each pixel in the map had a value for the predicted nitrogen content.

The values for the map based on the linear coefficients of the MTCI relation (Figure 19 (A, B, C)) ranged between -75.5 and 332.6 kg/ha as it can be seen in the associated map histogram for the map (Figure 18 (A)). Negative nitrogen content is of course not possible in the reality, the values were result of a simple linear mathematical function (Figure 13). However a small fraction of the pixels had negative nitrogen value which can be attributed to pixels with exposed soil cover. The values computed using exponential coefficients of the MTCI relation were between 35 kg/ha and 414.6 kg/ha. (Figure 18 (B)). It should be pointed out that the RMSEP for the exponential model is larger than the RMSEP from the linear model. If we compare the mean values of the predicted nitrogen from linear, exponential model and actual observations (Table 8 and Table 12), the linear model had better prediction, also with respect to the maximum value.

Nitrogen maps of the different fields (Figure 19) showed clear nitrogen patterns. Field 28 (Figure 19 A) was an experimental field with 3 initial fertilization levels (Figure 4, Section 3.1) and 4 different treatments during the growing season. There is a clear pattern showing the stripes of plots with different levels of initial fertilization. The purple stripe shows the plot with lowest nitrogen content which was expected because this plot didn’t get any initial fertilization. There was not much knowledge about the fertilization in other fields but in field 19 (Figure 19 B and E) is also a clear pattern dividing the field in two parts with lower nitrogen content on the west side. Field 11 (Figure 19 C and F) was a homogeneous in terms of nitrogen content and only the tractor driving paths are shown with low nitrogen concentration. The driving paths should have value 0 of nitrogen content but because of the scattered reflection from neighbouring pixels they have higher index values and respectively higher nitrogen concentration.

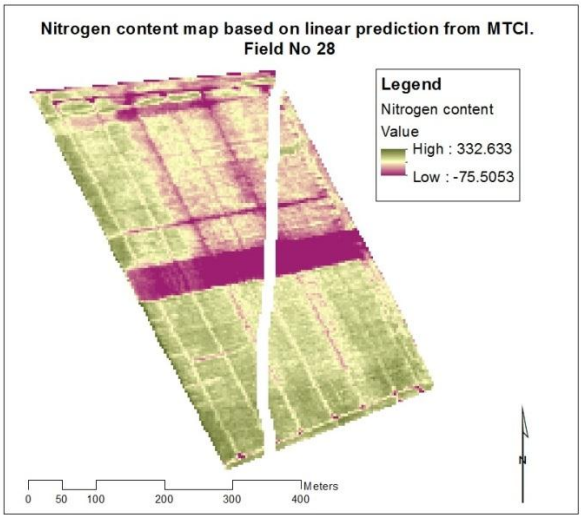
Table 12 Summary of the predicted nitrogen content maps based on MTCI

	Linear prediction	Exponential prediction
Minimum	-75.51	35.88
Median	207.7	196
Mean	195.6	189.4
Maximum	332.6	414.6

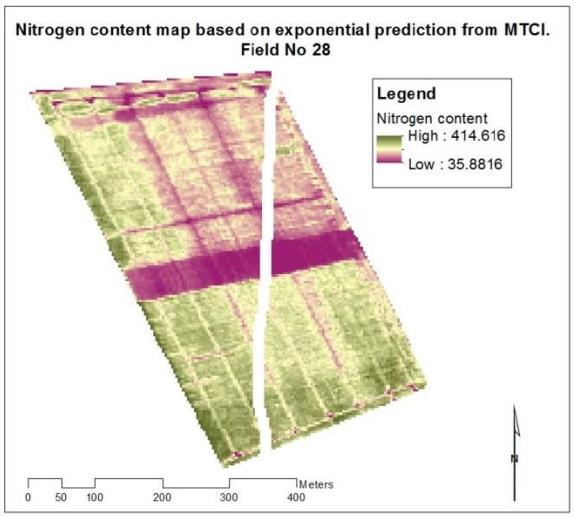


(A) **(B)**

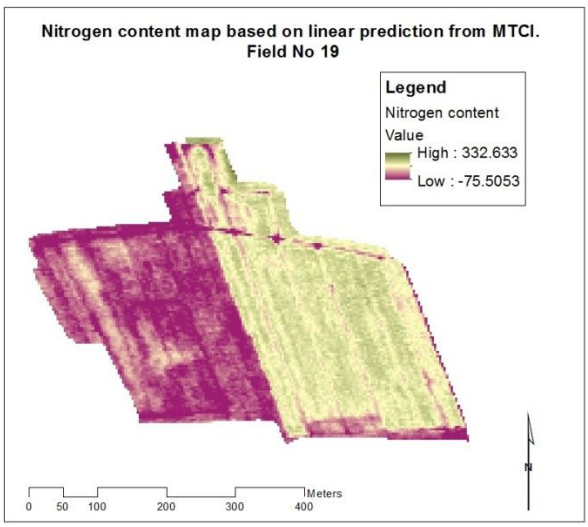
Figure 18 Histograms of predicted nitrogen content maps distribution from MTCI. On the left hand side is depicted the histogram based on linear coefficients and on the right hand side is the histogram based on exponential coefficients. The mean value is marked with a dashed line, the zero value with blue line and the maximum value with red line.



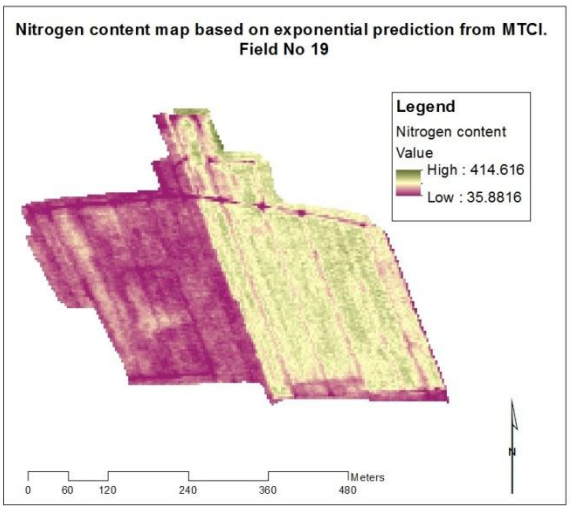
(A)



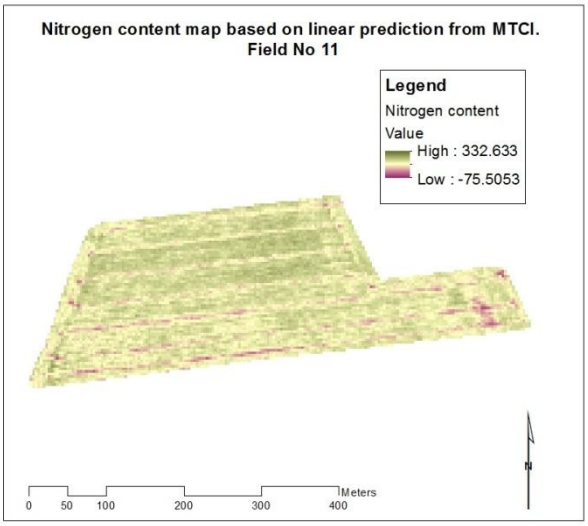
(D)



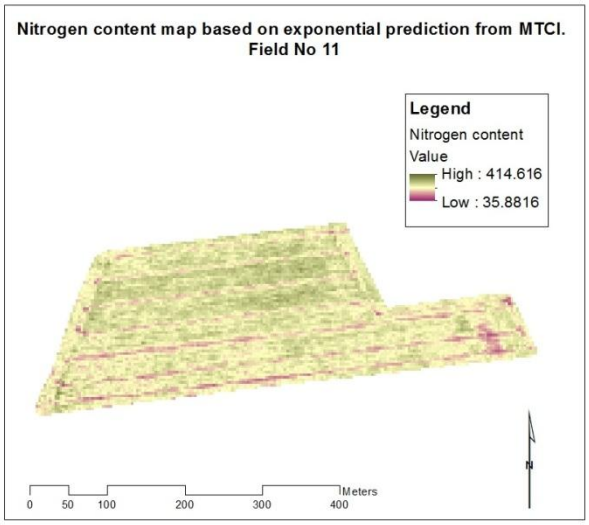
(B)



(E)



(C)



(F)

Figure 19 Predicted nitrogen content (kg/ha) maps for potato. On the left hand side are the prediction maps based on linear coefficients and on the right hand side are the exponential.

5. Discussion

In this research a range of different types of vegetation indices was tested to explore their potential in estimating the content of either chlorophyll or nitrogen for three different crops. The potential of the indices was tested using remote sensing data acquired by the image spectrometer APEX on board of an aircraft at 27 June 2011 and field observations gathered in two consecutive days coinciding with the date of the acquisition flight. The remote sensing data and the field measurements were correlated and regression models were built to calculate linear and nonlinear functions aiming to predict nitrogen and chlorophyll from the vegetation indices. There was little agreement in literature over which vegetation indices have the strongest relationship with chlorophyll and nitrogen (Chen, Haboudane et al. 2010). This study confirmed that there was no general index, showing best results for all crops. However, the group of indices calculating the red- edge position (REP and MTCI) had the most robust predictions. Other indices utilizing red- edge bands also showed good potential for nitrogen and chlorophyll estimation. The vegetation indices were tested for three types of crops: maize, grassland and potato. Unfortunately all the regression models for maize samples showed insignificant results (Appendix 1 and 2). The reason for the poor performance in the maize site was the lack of spread of observations as visible in the scatter plots in Appendix 4. The samples from maize had very low variation in terms of nitrogen content and in terms of reflectivity and vegetation index values. The standard deviation of nitrogen in maize was 13.6 kg/ha and for grassland and potato it was respectively: 27 kg/ha and 77.5 kg/ha (Appendix 3.3). The small variation caused lack of correlation with the vegetation indices.

Regressions for grass resulted in significant models for REP, MTCI, MCARI/OSAVI RE, TCARI/OSAVI RE, CI indices, NDRE indices. In potato the performance was slightly better and there almost all indices except NDNI and SIPI were significant. The grassland parcels were in different management stages as apparent from the scatter plots presented at Appendix 4. The samples in each field were clustered in groups as they had similar properties regarding spectral reflectance and nitrogen content values. The best performing index at grasslands was REP, next to it also MCARI/OSAVI RE and CI red edge performed with high values of R^2. For potato the best performing index was MTCI and the second best with the same value of R^2 were CI red edge and TCARI/OSAVI RE. The fertilization levels at potato fields were the major reason of the much higher variation at physiological properties of potato crops (Table 8, Section 4.2.1; Appendix 3.4). The higher variation was a reason for larger spread and better fit of the regression models for the potato fields. The results were crop specific and the main reason is the structure of the canopy, because it can significantly influence changes in canopy spectra (Stroppiana, 2009). Another reason was that nitrogen content range varies among crops and coefficients from regression model derived from potato model cannot be used to predict nitrogen at grassland.

The visible and red- edge position were reported as important spectral regions for nitrogen assessment. The red-edge region could be better suited than the red bands, which are strongly influenced by structural canopy parameters. The de-correlation of vegetation indices to factors other than nitrogen content had proved to be rather difficult to achieve (Stroppiana 2009). Canopy architecture can significantly influence changes in canopy spectra. Since NIR wavelengths are mainly influenced by fresh biomass, vegetation indices that use visible and red-edge position should be preferred (Stroppiana 2009). The red edge position was a good estimator of chlorophyll and respectively

nitrogen content, but being less sensitive at higher contents of chlorophyll (Figure 10, Section 4.1.1). This saturation effect is still a problem for predictions at higher concentration levels (Clevers and Kooistra 2012). Results from the current study confirmed that the indices utilising bands in the red- edge position have greater correlation to vegetation indices than indices using red and NIR positioned bands. The modified versions of TCARI/OSAVI and MCARI/OSAVI proposed by (Wu, Niu et al. 2008), composed of bands positioned at 705 nm and 750 nm showed better linearity than the original versions utilizing band lengths at 670 nm and 800 nm. Major reason was that the reflectance at 670 nm was getting quickly saturated at relatively low chlorophyll content. Despite the observed linear relationships between the modified vegetation indices by Wu et al 2008, they performed with higher correlation at exponential models for grassland and potato sites. Results were summarized in table (Table 6 and Table 9, Section 4). The non- linearity of the regression models was a general trend regarding all indices incorporating red-edge position bands.

Strong non-linear relationship between REP and chlorophyll was reported by (Clevers and Kooistra 2012), using PROSAIL simulations, with large range of REP values. In the current study REP had the strongest linear correlation among all indices with respect to nitrogen in grassland site, also strong linear prediction at potato fields in terms of nitrogen and also chlorophyll. The reason for the good linear prediction could be that narrower range of REP values was used in this study for model calibration. Also the results obtained by PROSAIL simulations are only describing chlorophyll and REP relationship, they cannot be directly compared to the nitrogen models. MTCI is a vegetation index designed to be sensitive to wide range of chlorophyll content (Dash and Curran 2004) and in this study it performed better than REP in respect to linearity, for chlorophyll and also nitrogen estimation. The chlorophyll indices presented by (Gitelson, Keydan et al. 2006)are using simple equation incorporating NIR band and red- edge band (710 nm) or green band (550 nm). These two indices were expected to have linear relation to chlorophyll and their major advantage was be the absence of the saturation effect obtained with REP indices (Clevers and Kooistra 2012). However in, grassland and potato, again stronger exponential relationship was observed than linear. In the potato crops regarding nitrogen content CI green and CI red edge performed with similar accuracy (Table 9, Section 4.2.1). But with respect to chlorophyll fit of CI green was better, both at linear and exponential model.

The NDRE indices using red- edge bands, also called 'red-edge NDVI' showed promising results in this study. There were different bands used in the calculation of this type of index. In this study three versions of NDRE were tested and their predictive power was compared (Table 2, Section 2.4) The best prediction was achieved for all the tests by the version proposed by (Tilling, O'Leary et al. 2007). The same formulation of NDRE was incorporated in the NDRE/NDVI ratio, called canopy chlorophyll content index, which turned out to be a robust index with performance similar to the performance achieved by REP and MTCI. NDVI, WDRVI saturated at very low nitrogen levels and could not be used for estimation of chlorophylls as they are related to the physical properties of the canopy. NDNI is the only one index investigated for direct nitrogen estimation from SWIR wavelengths, but it didn't showed satisfactory results in this aspect. The water content in the SWIR range is obscuring any relation with nitrogen (Curran 1989). Another point for discussion is the spatial resolution and the accuracy of geometry of the APEX image. The image was found to be distorted, especially in the south- east part. There the DGPS locations and the hyperspectral image had a shift of few pixels (Figure 7). Despite the correction there is still an uncertainty in the exact positions of field measurements with respect to the image. Furthermore the pixel size of APEX is about 7 m^2 and the sample size of the field measurements was 0.25 m^2 for the grassland and about 1m^2 for potato and maize. This point makes the relation between nitrogen readings and spectral reflectance per pixel also uncertain, because the effect of mixed pixels.

The results from the current study showed that REP, MCARI/OSAVI and CCCI performed best at grassland sites. For the potato the results were slightly different and there the best performance was achieved by MTCI. These results were in accordance with previous research (Clevers and Kooistra 2012). Based on the prediction of MTCI, nitrogen maps were created of the potato field (Figure 11) with the fertilization experiment (Field 28) and two others (Field 11 and Field 19) as the fertilization levels would be helpful for visual comparison of the predicted nitrogen content in the field. Both linear and exponential functions were used to upscale the prediction of the regression models and to calculate the nitrogen maps. The calculations based on linear models resulted in negative values of predicted nitrogen content at lower values of the vegetation index (Figure 19 (A, B, C)). After investigation of the pixels with negative values it was concluded that they coincide with the tractor diving paths where bare soil is exposed. These locations were not part of the regression model and respectively the predictor (the vegetation index) is outside of the range, which makes the prediction uncertain. The negative values derived from index predictions are another clear indication of the strong non-linear relationship between vegetation indices and chlorophyll. The choice of index and model in prediction of nitrogen content should be based on careful consideration of model diagnostics, taking into account validation results on a first place, because in the calibration period the model diagnostics may be over positive. That means that the model will very likely work better for the data used to fit it than for any other data. Analysts usually refer to this fact as the "principle of optimism" (Picard and Cook 1984).

Although we can make a prediction of the true mean for any pixel value of the vegetation index value it will be unwise to use extrapolation to make predictions outside the range of the index value that we have available for study. On the other hand it is reasonable to interpolate, to make predictions for unobserved values in between the range of the index values. That means that the linear model can still be used if the low index values are removed from the map, which can be easily done by excluding the bare soil pixels using a NDVI threshold. The predicted values from the exponential model were all positive also for low values of the explanatory variable, but they were overestimating the prediction at high index values. MTCI was shown to have strong linear relationship with chlorophyll in PROSAIL simulations and obtained the strongest linear relationship with chlorophyll and nitrogen in the current study in terms of R^2 values (Table 9 and Table 11). Despite that, the prediction maps had negative results for low index values. However, the linear model was predicting better according to the independent validation, it also better models the mean and maxim values (Table 12), if we don't use the vegetation index values falling outside the range from observed locations, we can consider linear predictions of MTCI as better then exponential . It is hard to say which prediction model works better for every index, since the results from the independent validations are very similar between the linear and exponential models (Appendix 3). The results from independent validation are claimed to be the best way to evaluate performance of regression models (Arboretti Giancristofaro and Salmaso 2007). However the choice and the proper use of validation samples can be tricky and very influential as the validation at potato crops proved. In this case only one outlier changed the RMSEP dramatically.

Earlier studies have shown that red-edge bands are highly significant for chlorophyll retrieval (Horler, Dockray et al. 1983; Daughtry, Walthall et al. 2000; Dash and Curran 2004; Clevers and Kooistra 2012). This point was also confirmed by the current study, other sources of data were explored to further investigate the red– edge bands. Recently the significance of the red edge bands on Sentinell-2 for estimating chlorophyll content has been shown through simulations studies (Herrmann, Pimstein et al. 2011). In this study we further elaborated the utility of Sentinel-2 bands for calculation of vegetation indices. When calculating the indices mentioned (Table 2, Section 2.4) using the simulated Sentinel-2 data, all these indices performed on similar way to the indices calculated on APEX bands. See table (Table 6 and Table 9, Section 4).

When the index values, calculated by continuous spectra, were related to those calculated by Sentinel- 2, the correlation was almost perfect (r >0.99). This result was also confirmed by (Herrmann, Pimstein et al. 2011). It reassures the high similarity of R^2 values for the tested indices for the two formations of data and that there is no advantage for the continuous over the superspectral data. The correlation with indices calculated on Landsat were also with very high correlation coefficient, for example NDVI (r >0.98), but this has no practical merit since NDVI was not a good predictor of chlorophyll and nitrogen. The absence of red-edge indices on board of Landsat TM makes the sensor with very limited value for nitrogen and chlorophyll retrieval. However CI green index can be calculated based on NIR and Green Landsat TM bands and to be used for chlorophyll prediction with the impressive result with $R^2 = 0.9$, significant at $p < 0.001$ (Table 11, Section 4.3). Apparently the exact position of the bands and the band width is not critical in estimating the chlorophyll content or nitrogen content.

6. Conclusions and recommendations

Results from this study indicate that the hyperspectral vegetation indices are a promising tool to derive biochemical characteristics from arable crops, such as chlorophyll and nitrogen. The best performing vegetation indices regarding all investigated crops in this study were REP, MTCI, CI red edge and also NDRE. In the formulation of those indices mainly bands from the red-edge region are used. This proves again the relevance of this spectral region for agricultural applications in accordance with the literature review. The index with highest correlation with nitrogen for grassland was REP (R^2 =0.612 linear and R^2 =0.713 exponential), for potato MTCI (R^2 = 0.655 linear and R^2 =0.746 exponential), for maize none of the indices showed statistically significant results. The best fit for chlorophyll model was achieved by MTCI (R^2 =0.905 linear and R^2 =0.927 exponential). The results were crop specific and main reason could be the influence by the canopy architecture as shown in previous studies using PROSAIL modelling.

Compared to multispectral sensors like Landsat TM the availability of narrow and contiguous bands is fundamental for retrieving nitrogen content. There is vast literature available, most of which has focussed on the use of field spectrometers and few studies have used airborne hyperspectral sensors, whereas space-borne instruments still suffer from problems for this type of applications: atmospheric correction, spatial resolution and signal to noise ratio. The upcoming Sentinel- 2 system has 12 bands, two of which are centred at the red-edge region (705 nm and 740 nm). They have good potential of retrieving canopy nitrogen and chlorophyll with high spatial resolution (20 m) and short revisit time. This study confirmed in accordance with previous studies the strength of the relation between vegetation indices based on Sentinel-2 band setting with chlorophyll and nitrogen.

Statistical regression methods, based on the use of vegetation indices, have been the most widely exploited approaches for retrieving nitrogen content from leaf and canopy spectra. However several studies shown that the performance of vegetation indices could be a function of site and vegetation characteristics and none of the indices appears to be robust enough the sites. Empirical models remain therefore applicable mainly at local and regional scale. The high correlation observed between chlorophyll and nitrogen is the basis for using spectra for nitrogen assessment and the relation between spectral changes and physiological processes should be further investigated.

In summary there are also some other priorities to be addressed by further research. There is a need for further understanding of the physiological processes involved in the relation between plant nitrogen content and canopy spectra by exploiting simulations from radiative transfer modelling. Also more knowledge is needed to evaluate the influence of vegetation indices of canopy architecture, other nutrients and other crop growing conditions, such as moisture, soils, irrigation. In this study one nitrogen index utilizing wavelengths in SWIR was tested (NDNI), which resulted in insignificant results, however the potential of SWIR for nitrogen motoring should be further explored. Nitrogen content prediction models should be further investigated for maize, experimental design with fertilization levels could considerably improve the prediction for this crop.

7. References:

Arboretti Giancristofaro, R. and L. Salmaso (2007). "Model performance analysis and model validation in logistic regression." Statistica **63**(2): 375-396.

Barnes, E., T. Clarke, et al. (2000). Coincident detection of crop water stress, nitrogen status and canopy density using ground based multispectral data. Proceeding International Conference on Precision Agriculture, 5th, Bloomington, MN.

Blackburn, G. A. (1999). "Relationships between spectral reflectance and pigment concentrations in stacks of deciduous broadleaves." Remote Sensing of Environment **70**(2): 224-237.

Carter, G. A. and A. K. Knapp (2001). "Leaf optical properties in higher plants: Linking spectral characteristics to stress and chlorophyll concentration." American Journal of Botany **88**(4): 677-684.

Chan, J. C. W. and D. Paelinckx (2008). "Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery." Remote Sensing of Environment **112**(6): 2999-3011.

Chen, P., D. Haboudane, et al. (2010). "New spectral indicator assessing the efficiency of crop nitrogen treatment in maize and wheat." Remote Sensing of Environment **114**(9): 1987-1997.

Cho, M. A. and A. K. Skidmore (2006). "A new technique for extracting the red edge position from hyperspectral data: The linear extrapolation method." Remote Sensing of Environment **101**(2): 181-193.

Clevers, J. G. P. W. and L. Kooistra (2012). "Using hyperspectral remote sensing data for retrieving canopy chlorophyll and nitrogen content." IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **5**(2): 574-583.

Cohen, Y., V. Alchanatis, et al. (2010). "Leaf nitrogen estimation in potato based on spectral data and on simulated bands of the VENμS satellite." Precision Agriculture **11**(5): 520-537.

Curran, P. J. (1989). "Remote sensing of foliar chemistry." Remote Sensing of Environment **30**(3): 271-278.

Curran, P. J., W. R. Windham, et al. (1995). "Exploring the relationship between reflectance red edge and chlorophyll concentration in slash pine leaves." Tree Physiology **15**(3): 203-206.

Dash, J. and P. J. Curran (2004). "The MERIS terrestrial chlorophyll index." International Journal of Remote Sensing **25**(23): 5403-5413.

Daughtry, C. S. T., C. L. Walthall, et al. (2000). "Estimating maize leaf chlorophyll concentration from leaf and canopy reflectance." Remote Sensing of Environment **74**(2): 229-239.

de Haan, J. F., J. W. Hovenier, et al. (1991). "Removal of atmospheric influences on satellite-borne imagery: A radiative transfer approach." Remote Sensing of Environment **37**(1): 1-21.

de Haan, J. F. and J. M. M. Kokke (1996). "Remote sensing algorithm development toolkit I Operationalization of atmospheric correction methods for tidal and inland waters." pp. 91.

Evans, J. R. (1989). "Photosynthesis and nitrogen relationships in leaves of C3 plants." Oecologia **78**(1): 9-19.

Filella, I., L. Serrano, et al. (1995). "Evaluating wheat nitrogen status with canopy reflectance indices and discriminant analysis." Crop Science **35**(5): 1400-1405.

Gitelson, A. and M. N. Merzlyak (1994). "Spectral reflectance changes associated with autumn senescence of Aesculus hippocastanum L. and Acer platanoides L. leaves. Spectral features and relation to chlorophyll estimation." Journal of Plant Physiology **143**: 286-286.

Gitelson, A. A. (2004). "Wide Dynamic Range Vegetation Index for Remote Quantification of Biophysical Characteristics of Vegetation." Journal of Plant Physiology **161**(2): 165-173.

Gitelson, A. A., G. P. Keydan, et al. (2006). "Three-band model for noninvasive estimation of chlorophyll, carotenoids, and anthocyanin contents in higher plant leaves." Geophysical Research Letters **33**(11).

Gitelson, A. A. and M. N. Merzlyak (1996). "Signature analysis of leaf reflectance spectra: algorithm development for remote sensing of chlorophyll." Journal of Plant Physiology **148**(3): 494-500.

Guyot, G. and F. Baret (1988). Utilisation de la haute resolution spectrale pour suivre l'etat des couverts vegetaux. Spectral Signatures of Objects in Remote Sensing.

Haboudane, D., J. R. Miller, et al. (2002). "Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture." Remote Sensing of Environment **81**(2-3): 416-426.

Hansen, P. M. and J. K. Schjoerring (2003). "Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression." Remote Sensing of Environment **86**(4): 542-553.

Hatfield, J. L., A. A. Gitelson, et al. (2008). "Application of spectral remote sensing for agronomic decisions." Agronomy Journal **100**(3 SUPPL.): S117-S131.

Henry, C., D. Sullivan, et al. (1999). "Managing Nitrogen from Biosolids." Biosolids Management Guidelines for Washington State.

Herrmann, I., A. Pimstein, et al. (2011). "LAI assessment of wheat and potato crops by VENμS and Sentinel-2 bands." Remote Sensing of Environment **115**(8): 2141-2151.

Herrmann, I., A. Pimstein, et al. (2011). "LAI assessment of wheat and potato crops by VENμS and Sentinel-2 bands." Remote Sensing of Environment **115**(8): 2141-2151.

Horler, D. N. H., M. Dockray, et al. (1983). "The red edge of plant leaf reflectance." International Journal of Remote Sensing **4**(2): 273-288.

Huete, A. R. (1988). "A soil-adjusted vegetation index (SAVI)." Remote Sensing of Environment **25**(3): 295-309.

Jain, N., S. Ray, et al. (2007). "Use of hyperspectral data to assess the effects of different nitrogen applications on a potato crop." Precision Agriculture **8**(4): 225-239.

Keeney, D. R. and J. L. Hatfield (2008). "The nitrogen cycle, historical perspective, and current and potential future concerns." New York, NY: Academic Press: p. 1-18.

Knipling, E. B. (1970). "Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation." Remote Sensing of Environment **1**(3): 155-159.

Kokaly, R. F. and R. N. Clark (1999). "Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression." Remote Sensing of Environment **67**(3): 267-287.
le Maire, G., C. François, et al. (2008). "Calibration and validation of hyperspectral indices for the estimation of broadleaved forest leaf chlorophyll content, leaf mass per area, leaf area index and leaf canopy biomass." Remote Sensing of Environment **112**(10): 3846-3864.
Lillesand, T. M., R. W. Kiefer, et al. (2004). Remote sensing and image interpretation, John Wiley & Sons Ltd.
Matson, P. A., R. Naylor, et al. (1998). "Integration of environmental, agronomic, and economic aspects of fertilizer management." Science **280**(5360): 112-115.
Merzlyak, M. N., A. A. Gitelson, et al. (1999). "Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening." Physiologia Plantarum **106**(1): 135-141.
Moran, M. S., Y. Inoue, et al. (1997). "Opportunities and limitations for image-based remote sensing in precision crop management." Remote Sensing of Environment **61**(3): 319-346.
Mutanga, O., A. K. Skidmore, et al. (2003). "Discriminating tropical grass (Cenchrus ciliaris) canopies grown under different nitrogen treatments using spectroradiometry." ISPRS Journal of Photogrammetry and Remote Sensing **57**(4): 263-272.
Penuelas, J., I. Filella, et al. (1995). "Assessment of photosynthetic radiation-use efficiency with spectral reflectance." New Phytologist **131**(3): 291-296.
Picard, R. R. and R. D. Cook (1984). "Cross-validation of regression models." Journal of the American Statistical Association **79**(387): 575-583.
Portis Jr, A. R. and M. A. J. Parry (2007). "Discoveries in Rubisco (Ribulose 1,5-bisphosphate carboxylase/oxygenase): A historical perspective." Photosynthesis Research **94**(1): 121-143.
Richardson, A. D., S. P. Duigan, et al. (2002). "An evaluation of noninvasive methods to estimate foliar chlorophyll content." New Phytologist **153**(1): 185-194.
Rouse, J. W., R. H. Haas, et al. (1974). "Monitoring the vernal advancements and retro gradation of natural." Vegetation, Greenbelt: 371.
Serrano, L., J. Penuelas, et al. (2002). "Remote sensing of nitrogen and lignin in Mediterranean vegetation from AVIRIS data: Decomposing biochemical from structural signals." Remote Sensing of Environment **81**(2): 355-364.
Sindy Sterckx, K. M. (2011). "APEX 2011 –June campaign Reusel delivery report."
Starks, P. J., D. Zhao, et al. (2008). "Estimation of nitrogen concentration and in vitro dry matter digestibility of herbage of warm-season grass pastures from canopy hyperspectral reflectance measurements." Grass and Forage Science **63**(2): 168-178.
Thenkabail, P. S. (2012). "Hyperspectral remote sensing of vegetation."
Thenkabail, P. S., E. A. Enclona, et al. (2004). "Hyperion, IKONOS, ALI, and ETM+ sensors in the study of African rainforests." Remote Sensing of Environment **90**(1): 23-43.
Thenkabail, P. S., E. A. Enclona, et al. (2004). "Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications." Remote Sensing of Environment **91**(3-4): 354-376.
Thenkabail, P. S., R. B. Smith, et al. (2000). "Hyperspectral vegetation indices and their relationships with agricultural crop characteristics." Remote Sensing of Environment **71**(2): 158-182.
Thenkabail, P. S., R. B. Smith, et al. (2002). "Evaluation of narrowband and broadband vegetation indices for determining optimal hyperspectral wavebands for agricultural crop characterization." Photogrammetric Engineering and Remote Sensing **68**(6): 607-621.
Tilling, A. K., G. J. O'Leary, et al. (2007). "Remote sensing of nitrogen and water stress in wheat." Field Crops Research **104**(1): 77-85.
Tilman, D., K. G. Cassman, et al. (2002). "Agricultural sustainability and intensive production practices." Nature **418**(6898): 671-677.
Uddling, J., J. Gelang-Alfredsson, et al. (2007). "Evaluating the relationship between leaf chlorophyll concentration and SPAD-502 chlorophyll meter readings." Photosynthesis Research **91**(1): 37-46.
Vitousek, P. M., K. Cassman, et al. (2002). "Towards an ecological understanding of biological nitrogen fixation." Biogeochemistry **57-58**: 1-45.
Vitousek, P. M. and R. W. Howarth (1991). "Nitrogen limitation on land and in the sea: How can it occur?" Biogeochemistry **13**(2): 87-115.
Wu, C., Z. Niu, et al. (2008). "Estimating chlorophyll content from hyperspectral vegetation indices: Modeling and validation." Agricultural and Forest Meteorology **148**(8-9): 1230-1241.
Yi, Q. X., J. F. Huang, et al. (2007). "Monitoring rice nitrogen status using hyperspectral reflectance and artificial neural network." Environmental Science and Technology **41**(19): 6770-6775.

Appendix 1: Linear Model Summary

1.1 Nitrogen Linear Model Summary APEX

Potato								
Index name	R2	AdjR2	RMSE	RMSEP	NRMSE (%)	F - statistic	DF	p-value:
REP	0.619	0.603	42.084	82.293	60.500	38.982	(1, 24)	0.000001878
MTCI	0.655	0.641	40.021	81.514	57.600	45.640	(1, 24)	0.000000548
MCARI.OSAVI	0.276	0.246	58.002	106.027	83.400	9.156	(1, 24)	0.005835000
MCARI.OSAVI.RE	0.553	0.535	45.569	77.961	65.500	29.717	(1, 24)	0.000013320
TCARI.OSAVI	0.549	0.530	45.770	88.862	65.800	29.245	(1, 24)	0.000014850
TCARI.OSAVI.RE	0.653	0.638	40.164	77.498	57.800	45.147	(1, 24)	0.000000598
CI.red.edge	0.653	0.639	40.153	78.096	57.800	45.186	(1, 24)	0.000000721
CI.green	0.638	0.623	41.001	77.117	59.000	42.351	(1, 24)	0.000001189
NDNI	0.114	0.077	64.157	105.987	92.300	3.099	(1, 24)	0.091080000
SIPI	0.137	0.101	63.329	106.877	91.100	3.812	(1, 24)	0.062650000
NDRE	0.626	0.610	41.688	80.558	60.000	40.184	(1, 24)	0.000001489
NDRE1	0.559	0.541	45.273	83.476	65.100	30.422	(1, 24)	0.000011337
NDRE2	0.577	0.560	44.316	82.640	63.700	32.798	(1, 24)	0.000006693
NDVI	0.261	0.230	58.617	94.562	84.300	8.463	(1, 24)	0.007693000
CCCI	0.619	0.603	42.102	82.723	60.600	38.928	(1, 24)	0.000001898
WDRVI	0.273	0.242	58.140	93.003	83.600	8.998	(1, 24)	0.006210000
Grass								
Index name	R2	AdjR2	RMSE	RMSEP(LOOCV)	NRMSE (%)	F - statistic	DF	p-value:
REP	0.612	0.582	13.999	17.437	60.200	20.529	(1,13)	0.000564500
MTCI	0.599	0.568	14.234	17.322	61.200	19.431	(1,13)	0.000707100
MCARI.OSAVI	0.003	-0.074	22.453	25.302	96.500	0.033	(1,13)	0.858300000
MCARI.OSAVI.RE	0.607	0.577	14.092	16.616	60.600	20.088	(1,13)	0.000617300
TCARI.OSAVI	0.235	0.176	19.662	22.226	84.500	3.996	(1,13)	0.066960000
TCARI.OSAVI.RE	0.417	0.372	17.171	19.753	73.800	9.286	(1,13)	0.009349000
CI.red.edge	0.502	0.464	15.866	19.035	68.200	13.103	(1,13)	0.003112000
CI.green	0.348	0.298	18.147	20.682	78.000	6.952	(1,13)	0.020530000
NDNI	0.008	-0.068	22.392	26.003	96.200	0.105	(1,13)	0.751600000
SIPI	0.245	0.187	19.536	21.235	83.900	4.216	(1,13)	0.060720000
NDRE	0.523	0.486	15.525	19.343	66.700	14.260	(1,13)	0.002309000
NDRE1	0.364	0.315	17.931	22.124	77.100	7.435	(1,13)	0.017288490
NDRE2	0.422	0.378	17.087	21.620	73.400	9.505	(1,13)	0.008728697
NDVI	0.206	0.145	20.027	22.543	86.100	3.382	(1,13)	0.088880000
CCCI	0.604	0.574	14.147	17.483	60.800	19.830	(1,13)	0.000650900
WDRVI	0.206	0.145	20.028	22.208	86.100	3.380	(1,13)	0.088950000
Corn								
Index name	R2	AdjR2	RMSE	RMSEP(LOOCV)	NRMSE (%)	F - statistic	DF	p-value:
REP	0.067	-0.005	13.043	15.134	93.300	0.934	(1,13)	0.351300000
MTCI	0.111	0.043	12.731	14.684	91.100	1.628	(1,13)	0.224300000
MCARI.OSAVI	0.002	-0.075	13.489	15.615	96.500	0.028	(1,13)	0.868600000
MCARI.OSAVI.RE	0.178	0.115	12.242	13.745	87.600	2.818	(1,13)	0.117100000
TCARI.OSAVI	0.059	-0.014	13.101	15.184	93.700	0.812	(1,13)	0.383800000
TCARI.OSAVI.RE	0.121	0.054	12.659	14.391	90.600	1.793	(1,13)	0.203500000
CI.red.edge	0.117	0.049	12.687	14.502	90.800	1.728	(1,13)	0.211400000
CI.green	0.101	0.032	12.806	14.591	91.600	1.456	(1,13)	0.249100000
NDNI	0.144	0.078	12.496	14.132	89.400	2.182	(1,13)	0.163500000
SIPI	0.171	0.108	12.293	13.927	87.900	2.687	(1,13)	0.125100000
NDRE	0.102	0.033	12.795	14.669	91.500	1.481	(1,13)	0.245200000
NDRE1	0.136	0.070	12.549	14.166	89.800	2.053	(1,13)	0.175493800
NDRE2	0.120	0.052	12.669	14.389	90.600	1.771	(1,13)	0.206090800
NDVI	0.113	0.045	12.718	14.291	91.000	1.657	(1,13)	0.220400000
CCCI	0.092	0.022	12.866	14.792	92.000	1.322	(1,13)	0.270900000
WDRVI	0.110	0.042	12.738	14.324	91.100	1.610	(1,13)	0.226700000

1.2 Nitrogen Linear Model Summary Sentinel-2

Potato								
Index name	R2	AdjR2	RMSE	RMSEP	NRMSE (%)	F - statistic	DF	p-value:
REP	0.621	0.606	41.953	82.179	60.300	39.375	(1, 24)	0.000001740
MTCI	0.654	0.639	40.124	83.612	57.700	45.283	(1, 24)	0.000000584
MCARI.OSAVI	0.140	0.104	63.215	114.392	90.900	3.913	(1, 24)	0.059505610
MCARI.OSAVI.RE	0.505	0.484	47.961	80.130	69.000	24.493	(1, 24)	0.000047307
TCARI.OSAVI	0.437	0.414	51.142	87.015	73.600	18.647	(1, 24)	0.000235143
TCARI.OSAVI.RE	0.651	0.636	40.303	79.074	58.000	44.671	(1, 24)	0.000000651
CI.red.edge	0.647	0.632	40.514	78.482	58.300	43.957	(1, 24)	0.000000740
CI.green	0.632	0.616	41.373	77.082	59.500	41.163	(1, 24)	0.000001237
SIPI	0.147	0.112	62.949	105.101	90.500	4.149	(1, 24)	0.052833860
NDRE	0.645	0.630	40.632	83.277	58.400	43.561	(1, 24)	0.000000794
NDRE1	0.574	0.557	44.483	82.547	64.000	32.372	(1, 24)	0.000007343
NDRE2	0.590	0.573	43.666	77.978	62.800	34.501	(1, 24)	0.000004652
NDVI	0.277	0.247	57.958	93.488	83.400	9.206	(1, 24)	0.005722528
CCCI	0.633	0.618	41.304	79.674	59.400	41.381	(1, 24)	0.000001187
WDRVI	0.291	0.261	57.420	91.867	82.600	9.831	(1, 24)	0.004487910
Grass								
Index name	R2	AdjR2	RMSE	RMSEP(LOOCV)	NRMSE (%)	F - statistic	DF	p-value:
REP	0.606	0.576	14.103	17.844	60.600	20.034	(1, 13)	0.000624150
MTCI	0.525	0.489	15.493	18.788	66.600	14.374	(1, 13)	0.002243950
MCARI.OSAVI	0.000	-0.077	22.481	25.447	96.600	0.000	(1, 13)	0.983209000
MCARI.OSAVI.RE	0.597	0.566	14.267	16.990	61.300	19.282	(1, 13)	0.000729490
TCARI.OSAVI	0.101	0.032	21.313	23.713	91.600	1.465	(1, 13)	0.247755900
TCARI.OSAVI.RE	0.331	0.279	18.393	20.925	79.000	6.421	(1, 13)	0.024933430
CI.red.edge	0.487	0.448	16.097	19.088	69.200	12.359	(1, 13)	0.003798874
CI.green	0.355	0.305	18.060	20.589	77.600	7.145	(1, 13)	0.019155450
SIPI	0.198	0.136	20.133	24.571	86.500	3.210	(1, 13)	0.096468290
NDRE	0.650	0.623	13.305	15.955	57.200	24.116	(1, 13)	0.000284555
NDRE1	0.377	0.329	17.747	22.193	76.300	7.861	(1, 13)	0.014923410
NDRE2	0.439	0.395	16.844	21.559	72.400	10.159	(1, 13)	0.007139028
NDVI	0.208	0.147	20.012	22.582	86.000	3.406	(1, 13)	0.087837020
CCCI	0.712	0.690	12.066	14.252	51.800	32.134	(1, 13)	0.000076850
WDRVI	0.208	0.147	20.007	22.215	86.000	3.416	(1, 13)	0.087450060
Corn								
Index name	R2	AdjR2	RMSE	RMSEP(LOOCV)	NRMSE (%)	F - statistic	DF	p-value:
REP	0.081	0.010	12.945	14.974	92.600	1.146	(1, 13)	0.303805700
MTCI	0.107	0.038	12.764	14.617	91.300	1.551	(1, 13)	0.234932800
MCARI.OSAVI	0.003	-0.074	13.487	15.696	96.500	0.033	(1, 13)	0.857603600
MCARI.OSAVI.RE	0.186	0.124	12.181	13.613	87.100	2.978	(1, 13)	0.108050800
TCARI.OSAVI	0.073	0.002	12.998	14.878	93.000	1.031	(1, 13)	0.328426900
TCARI.OSAVI.RE	0.119	0.051	12.678	14.350	90.700	1.749	(1, 13)	0.208775100
CI.red.edge	0.110	0.042	12.738	14.564	91.100	1.612	(1, 13)	0.226536300
CI.green	0.094	0.025	12.852	14.614	91.900	1.354	(1, 13)	0.265521700
SIPI	0.180	0.117	12.227	13.820	87.500	2.857	(1, 13)	0.114800300
NDRE	0.081	0.010	12.949	14.982	92.600	1.138	(1, 13)	0.305391700
NDRE1	0.133	0.067	12.571	14.203	89.900	2.001	(1, 13)	0.180714900
NDRE2	0.120	0.052	12.667	14.414	90.600	1.774	(1, 13)	0.205774000
NDVI	0.137	0.071	12.545	14.049	89.700	2.065	(1, 13)	0.174362400
CCCI	0.068	-0.003	13.035	15.128	93.200	0.954	(1, 13)	0.346617900
WDRVI	0.134	0.067	12.567	14.088	89.900	2.010	(1, 13)	0.179738200

1.3 Nitrogen Linear Model Summary Landsat TM

Potato								
Index name	R2	AdjR2	RMSE	RMSEP	NRMSE (%)	F - statistic	DF	p-value:
CI.green	0.609	0.593	42.621	76.527	61.300	37.403	(1, 24)	0.000002565
SIPI	0.199	0.165	61.027	101.686	87.800	5.950	(1, 24)	0.022480080
NDVI	0.326	0.298	55.978	90.667	80.500	11.597	(1, 24)	0.002326145
WDRVI	0.344	0.316	55.237	88.879	79.500	12.558	(1, 24)	0.001653534
Grass								
Index name	R2	AdjR2	RMSE	RMSEP(LOOCV)	NRMSE (%)	F - statistic	DF	p-value:
CI.green	0.290	0.235	18.944	21.336	81.400	5.308	(1,13)	0.038374230
SIPI	0.276	0.221	19.125	21.829	82.200	4.964	(1,13)	0.044168370
NDVI	0.206	0.144	20.038	22.483	86.100	3.363	(1,13)	0.089645720
WDRVI*	0.205	0.144	20.041	22.191	86.100	3.359	(1,13)	0.089816000
Corn								
Index name	R2	AdjR2	RMSE	RMSEP(LOOCV)	NRMSE (%)	F - statistic	DF	p-value:
CI.green	0.110	0.041	12.743	14.933	91.200	1.600	(1,13)	0.228097800
SIPI	0.275	0.219	11.497	14.176	82.300	4.934	(1,13)	0.044715560
NDVI	0.144	0.078	12.492	13.678	89.400	2.191	(1,13)	0.162629200
WDRVI	0.141	0.075	12.515	13.912	89.500	2.136	(1,13)	0.167638000

1.4 Chlorophyll Linear Model Summary APEX

Potato								
Index name	R2	AdjR2	RMSE	RMSEP(LOOCV)	NRMSE (%)	F - statistic	DF	p-value:
REP	0.872	0.865	0.034	0.037	34.900	129.147	(1, 19)	0.000000001
MTCI	0.905	0.900	0.029	0.031	30.100	180.853	(1, 19)	0.000000000
MCARI.OSAVI	0.640	0.621	0.057	0.063	58.500	33.820	(1, 19)	0.000013312
MCARI.OSAVI.RE	0.790	0.779	0.043	0.047	44.700	71.624	(1, 19)	0.000000072
TCARI.OSAVI	0.900	0.895	0.030	0.032	30.800	171.549	(1, 19)	0.000000000
TCARI.OSAVI.RE	0.893	0.887	0.031	0.033	31.900	158.468	(1, 19)	0.000000000
CI.red.edge	0.881	0.875	0.033	0.035	33.600	140.918	(1, 19)	0.000000000
CI.green	0.892	0.886	0.031	0.033	32.100	156.274	(1, 19)	0.000000000
NDNI	0.261	0.222	0.082	0.088	83.900	6.703	(1, 19)	0.018005220
SIPI	0.362	0.328	0.076	0.081	78.000	10.772	(1, 19)	0.003920312
NDRE	0.864	0.856	0.035	0.038	36.000	120.296	(1, 19)	0.000000001
NDRE1	0.817	0.808	0.041	0.046	41.700	85.105	(1, 19)	0.000000019
NDRE2	0.826	0.817	0.040	0.044	40.700	90.095	(1, 19)	0.000000012
NDVI	0.602	0.581	0.060	0.068	61.500	28.765	(1, 19)	0.000035547
CCCI	0.883	0.877	0.032	0.035	33.400	143.653	(1, 19)	0.000000000
WDRVI	0.623	0.603	0.058	0.065	59.900	31.439	(1, 19)	0.000020875

1.5 Chlorophyll Linear Model Summary Sentinel-2

Potato								
Index name	R2	AdjR2	RMSE	RMSEP(LOOCV)	NRMSE (%)	F - statistic	DF	p-value:
REP	0.872	0.865	0.034	0.037	35.000	128.954	(1, 19)	0.000000001
MTCI	0.909	0.904	0.029	0.031	29.500	189.095	(1, 19)	0.000000000
MCARI.OSAVI	0.248	0.208	0.082	0.092	84.600	6.265	(1, 19)	0.021600220
MCARI.OSAVI.RE	0.742	0.729	0.048	0.052	49.600	54.681	(1, 19)	0.000000528
TCARI.OSAVI	0.760	0.748	0.046	0.051	47.800	60.213	(1, 19)	0.000000263
TCARI.OSAVI.RE	0.892	0.886	0.031	0.034	32.100	156.262	(1, 19)	0.000000000
CI.red.edge	0.885	0.879	0.032	0.035	33.100	146.650	(1, 19)	0.000000000
CI.green	0.888	0.882	0.032	0.034	32.700	150.388	(1, 19)	0.000000000
SIPI	0.388	0.356	0.074	0.080	76.300	12.045	(1, 19)	0.002560711
NDRE	0.874	0.867	0.034	1.909	34.700	131.593	(1, 19)	0.000000001
NDRE1	0.824	0.815	0.040	0.103	40.900	89.180	(1, 19)	0.000000013
NDRE2	0.833	0.824	0.039	0.683	39.900	94.434	(1, 19)	0.000000008
NDVI	0.614	0.594	0.059	0.067	60.600	30.282	(1, 19)	0.000026182
CCCI	0.878	0.871	0.033	0.036	34.100	136.606	(1, 19)	0.000000000
WDRVI	0.638	0.619	0.057	0.064	58.700	33.465	(1, 19)	0.000014215

1.6 Chlorophyll Linear Model Summary Landsat TM

Potato								
Index name	R2	AdjR2	RMSE	RMSEP(LOOCV)	NRMSE (%)	F - statistic	DF	p-value:
CI.green	0.874	0.867	0.034	0.036	34.700	131.611	(1, 19)	0.000000001
SIPI	0.460	0.432	0.070	0.089	71.700	16.190	(1, 19)	0.000725940
NDVI	0.650	0.631	0.056	0.075	57.800	35.242	(1, 19)	0.000010277
WDRVI	0.674	0.657	0.054	0.071	55.700	39.303	(1, 19)	0.000005094

Appendix 2: Exponential Model Summary

2.1 Nitrogen Exponential Model Summary APEX

Potato								
Index name	R2	AdjR2	RMSE	RMSEP	NRMSE (%)	F - statistic	DF	p-value:
REP	0.724	0.713	41.434	82.871	59.600	62.983	(1, 24)	0.000000036
MTCI	0.746	0.736	39.789	83.210	57.200	70.612	(1, 24)	0.000000013
MCARI.OSAVI	0.310	0.281	60.653	112.639	87.200	10.780	(1, 24)	0.003137207
MCARI.OSAVI.RE	0.627	0.611	44.100	78.571	63.400	40.274	(1, 24)	0.000001464
TCARI.OSAVI	0.666	0.652	45.200	91.728	65.000	47.922	(1, 24)	0.000000370
TCARI.OSAVI.RE	0.733	0.722	39.096	77.576	56.200	65.997	(1, 24)	0.000000024
CI.red.edge	0.735	0.724	39.688	78.408	57.100	64.952	(1, 24)	0.000000028
CI.green	0.726	0.715	40.763	76.605	58.600	62.493	(1, 24)	0.000000039
NDNI	0.155	0.120	65.834	109.378	94.700	4.401	(1, 24)	0.046626400
SIPI	0.127	0.091	64.354	111.503	92.600	3.506	(1, 24)	0.073362050
NDRE	0.731	0.719	40.156	80.309	57.800	65.113	(1, 24)	0.000000027
NDRE1	0.657	0.643	42.056	85.200	60.500	45.980	(1, 24)	0.000000516
NDRE2	0.678	0.665	41.258	83.766	59.300	50.604	(1, 24)	0.000000236
NDVI	0.307	0.278	58.595	98.505	84.300	10.607	(1, 24)	0.003345319
CCCI	0.728	0.717	41.162	83.519	59.200	64.290	(1, 24)	0.000000030
WDRVI	0.318	0.289	58.136	96.810	83.600	11.181	(1, 24)	0.002705650
Grass								
Index name	R2	AdjR2	RMSE	RMSEP(LOOCV)	NRMSE (%)	F - statistic	DF	p-value:
REP	0.713	0.691	13.011	14.608	55.900	32.308	(1, 13)	0.000074911
MTCI	0.712	0.690	13.329	14.904	57.300	32.126	(1, 13)	0.000076947
MCARI.OSAVI	0.002	-0.075	23.007	25.648	98.900	0.024	(1, 13)	0.878750800
MCARI.OSAVI.RE	0.737	0.717	13.046	14.586	56.100	36.387	(1, 13)	0.000042191
TCARI.OSAVI	0.339	0.289	19.979	21.906	85.900	6.679	(1, 13)	0.022664480
TCARI.OSAVI.RE	0.547	0.512	17.179	18.905	73.800	15.671	(1, 13)	0.001634517
CI.red.edge	0.627	0.598	15.332	17.046	65.900	21.868	(1, 13)	0.000433407
CI.green	0.472	0.432	18.423	20.356	79.200	11.634	(1, 13)	0.004644953
NDNI	0.033	-0.042	22.976	26.300	98.700	0.438	(1, 13)	0.519693200
SIPI	0.349	0.299	19.873	21.268	85.400	6.966	(1, 13)	0.020420400
NDRE	0.647	0.620	14.745	16.446	63.400	23.812	(1, 13)	0.000300728
NDRE1	0.493	0.454	17.733	19.501	76.200	12.656	(1, 13)	0.003505138
NDRE2	0.551	0.517	16.632	18.409	71.500	15.984	(1, 13)	0.001517622
NDVI	0.316	0.264	20.403	22.116	87.700	6.018	(1, 13)	0.029036080
CCCI	0.718	0.696	13.127	14.697	56.400	33.056	(1, 13)	0.000067162
WDRVI	0.318	0.266	20.506	22.228	88.100	6.073	(1, 13)	0.028436320
Corn								
Index name	R2	AdjR2	RMSE	RMSEP(LOOCV)	NRMSE (%)	F - statistic	DF	p-value:
REP	0.066	-0.005	13.073	15.146	93.500	0.925	(1, 13)	0.353621000
MTCI	0.112	0.043	12.776	14.743	91.400	1.634	(1, 13)	0.223532500
MCARI.OSAVI	0.001	-0.076	13.519	15.650	96.700	0.007	(1, 13)	0.933636500
MCARI.OSAVI.RE	0.189	0.127	12.279	13.749	87.800	3.032	(1, 13)	0.105256000
TCARI.OSAVI	0.057	-0.016	13.141	15.230	94.000	0.786	(1, 13)	0.391550300
TCARI.OSAVI.RE	0.125	0.057	12.704	14.426	90.900	1.852	(1, 13)	0.196722100
CI.red.edge	0.121	0.053	12.731	14.544	91.100	1.782	(1, 13)	0.204868800
CI.green	0.107	0.039	12.846	14.610	91.900	1.565	(1, 13)	0.232915400
NDNI	0.161	0.096	12.549	14.151	89.800	2.495	(1, 13)	0.138242600
SIPI	0.186	0.124	12.370	14.035	88.500	2.976	(1, 13)	0.108187000
NDRE	0.101	0.032	12.828	14.699	91.800	1.461	(1, 13)	0.248233600
NDRE1	0.139	0.073	12.600	14.213	90.100	2.107	(1, 13)	0.170336300
NDRE2	0.122	0.055	12.713	14.426	90.900	1.809	(1, 13)	0.201663500
NDVI	0.121	0.053	12.758	14.298	91.300	1.782	(1, 13)	0.204837000
CCCI	0.089	0.019	12.898	14.829	92.300	1.267	(1, 13)	0.280756300
WDRVI	0.118	0.050	12.779	14.330	91.400	1.734	(1, 13)	0.210600300

2.2 Nitrogen Exponential Model Summary Sentinel-2

Potato								
Index name	R2	AdjR2	RMSE	RMSEP	NRMSE (%)	F - statistic	DF	p-value:
REP	0.729	0.717	40.459	82.809	58.200	64.455	(1, 24)	0.000000030
MTCI	0.748	0.737	38.752	85.988	55.700	71.188	(1, 24)	0.000000012
MCARI.OSAVI	0.150	0.114	65.548	120.205	94.300	4.231	(1, 24)	0.050713340
MCARI.OSAVI.RE	0.578	0.560	46.744	81.391	67.200	32.833	(1, 24)	0.000006642
TCARI.OSAVI	0.547	0.528	49.368	88.250	71.000	28.980	(1, 24)	0.000015796
TCARI.OSAVI.RE	0.734	0.723	39.076	79.484	56.200	66.288	(1, 24)	0.000000023
CI.red.edge	0.729	0.718	39.582	79.030	56.900	64.657	(1, 24)	0.000000029
CI.green	0.719	0.707	40.906	76.591	58.800	61.424	(1, 24)	0.000000045
SIPI	0.140	0.104	63.974	109.693	92.000	3.901	(1, 24)	0.059866580
NDRE	0.671	0.657	41.516	84.770	59.700	48.971	(1, 24)	0.000000310
NDRE1	0.690	0.677	40.869	83.519	58.800	53.518	(1, 24)	0.000000148
NDRE2	0.738	0.727	41.583	77.511	59.800	67.567	(1, 24)	0.000000019
NDVI	0.330	0.302	57.781	97.411	83.100	11.813	(1, 24)	0.002152031
CCCI	0.727	0.716	42.661	80.044	61.400	63.888	(1, 24)	0.000000032
WDRVI	0.342	0.315	57.257	95.629	82.400	12.494	(1, 24)	0.001690687
Grass								
Index name	R2	AdjR2	RMSE	RMSEP(LOOCV)	NRMSE (%)	F - statistic	DF	p-value:
REP	0.717	0.696	12.976	14.583	55.800	32.978	(1, 13)	0.000067923
MTCI	0.642	0.615	14.871	16.472	63.900	23.339	(1, 13)	0.000328091
MCARI.OSAVI	0.013	-0.063	23.051	25.653	99.100	0.165	(1, 13)	0.690930600
MCARI.OSAVI.RE	0.726	0.705	13.134	14.726	56.400	34.437	(1, 13)	0.000055160
TCARI.OSAVI	0.187	0.125	22.032	23.979	94.700	2.999	(1, 13)	0.106980800
TCARI.OSAVI.RE	0.451	0.409	18.622	20.458	80.000	10.672	(1, 13)	0.006127523
CI.red.edge	0.612	0.582	15.673	17.365	67.400	20.489	(1, 13)	0.000569030
CI.green	0.478	0.438	18.354	20.351	78.900	11.890	(1, 13)	0.004323310
SIPI	0.291	0.237	20.572	23.926	88.400	5.341	(1, 13)	0.037869170
NDRE	0.501	0.463	17.464	13.847	75.000	13.066	(1, 13)	0.003142294
NDRE1	0.564	0.530	16.278	19.274	69.900	16.787	(1, 13)	0.001259893
NDRE2	0.755	0.736	12.333	18.082	53.000	40.050	(1, 13)	0.000026232
NDVI	0.316	0.264	20.372	22.082	87.500	6.013	(1, 13)	0.029095790
CCCI	0.801	0.786	11.366	12.894	48.800	52.369	(1, 13)	0.000006585
WDRVI	0.319	0.266	20.465	22.179	87.900	6.083	(1, 13)	0.028326610
Corn								
Index name	R2	AdjR2	RMSE	RMSEP(LOOCV)	NRMSE (%)	F - statistic	DF	p-value:
REP	0.081	0.010	12.976	14.985	92.800	1.148	(1, 13)	0.303449100
MTCI	0.107	0.038	12.804	14.654	91.600	1.550	(1, 13)	0.235146000
MCARI.OSAVI	0.006	-0.071	13.518	15.720	96.700	0.076	(1, 13)	0.786483200
MCARI.OSAVI.RE	0.203	0.141	12.224	13.606	87.500	3.303	(1, 13)	0.092277750
TCARI.OSAVI	0.075	0.003	13.037	14.900	93.300	1.047	(1, 13)	0.324769000
TCARI.OSAVI.RE	0.123	0.056	12.721	14.375	91.000	1.831	(1, 13)	0.199080400
CI.red.edge	0.113	0.045	12.776	14.589	91.400	1.661	(1, 13)	0.219986400
CI.green	0.101	0.032	12.888	14.620	92.200	1.459	(1, 13)	0.248639700
SIPI	0.196	0.134	12.310	18.911	88.100	3.174	(1, 13)	0.098171470
NDRE	0.138	0.071	12.616	14.987	90.300	2.076	(1, 13)	0.173277400
NDRE1	0.123	0.056	12.708	14.236	90.900	1.828	(1, 13)	0.199463900
NDRE2	0.081	0.011	12.979	14.442	92.800	1.150	(1, 13)	0.303028000
NDVI	0.145	0.079	12.588	14.066	90.100	2.207	(1, 13)	0.161237500
CCCI	0.068	-0.004	13.062	15.128	93.400	0.949	(1, 13)	0.347859100
WDRVI	0.142	0.076	12.611	14.103	90.200	2.152	(1, 13)	0.166168200

2.3 Nitrogen Exponential Model Summary Landsat TM

Potato								
Index name	R2	AdjR2	RMSE	RMSEP	NRMSE (%)	F - statistic	DF	p-value:
CI.green	0.693	0.680	41.738	75.991	60.000	54.178	(1, 24)	0.000000133
SIPI	0.203	0.170	61.603	106.356	88.600	6.122	(1, 24)	0.020813480
NDVI	0.388	0.362	55.288	94.337	79.500	15.184	(1, 24)	0.000683782
WDRVI	0.404	0.379	54.567	92.294	78.500	16.266	(1, 24)	0.000484622
Grass								
Index name	R2	AdjR2	RMSE	RMSEP(LOOCV)	NRMSE (%)	F - statistic	DF	p-value:
CI.green	0.407	0.361	19.461	21.588	83.600	8.913	(1, 13)	0.010529100
SIPI	0.386	0.339	19.210	20.720	82.600	8.184	(1, 13)	0.013377450
NDVI	0.313	0.260	20.418	22.123	87.700	5.930	(1, 13)	0.030041270
WDRVI	0.315	0.262	20.519	22.240	88.200	5.974	(1, 13)	0.029534770
Corn								
Index name	R2	AdjR2	RMSE	RMSEP(LOOCV)	NRMSE (%)	F - statistic	DF	p-value:
CI.green	0.117	0.049	12.784	14.563	91.500	1.719	(1, 13)	0.2125005
SIPI	0.292	0.238	11.529	13.862	82.500	5.368	(1, 13)	0.03745984
NDVI	0.152	0.087	12.537	14.561	89.700	2.338	(1, 13)	0.1502158
WDRVI	0.149	0.084	12.560	14.387	89.900	2.282	(1, 13)	0.1548209

2.4 Chlorophyll Exponential Model Summary APEX

Potato								
Index name	R2	AdjR2	RMSE	RMSEP(LOOCV)	NRMSE (%)	F - statistic	DF	p-value:
REP	0.906	0.901	0.032	0.034	32.600	182.449	(1, 19)	0.000000000
MTCI	0.927	0.924	0.027	0.029	27.900	242.890	(1, 19)	0.000000000
MCARI.OSAVI	0.596	0.575	0.055	0.061	56.700	28.063	(1, 19)	0.000041086
MCARI.OSAVI.RE	0.839	0.831	0.042	0.045	43.100	99.334	(1, 19)	0.000000006
TCARI.OSAVI	0.929	0.926	0.027	0.029	27.600	249.875	(1, 19)	0.000000000
TCARI.OSAVI.RE	0.923	0.919	0.028	0.030	29.100	227.198	(1, 19)	0.000000000
CI.red.edge	0.913	0.909	0.030	0.032	30.900	200.224	(1, 19)	0.000000000
CI.green	0.920	0.916	0.028	0.030	29.100	219.822	(1, 19)	0.000000000
NDNI	0.323	0.287	0.083	0.089	85.200	9.065	(1, 19)	0.007187041
SIPI	0.413	0.382	0.077	0.082	79.100	13.387	(1, 19)	0.001669156
NDRE	0.902	0.897	0.033	0.035	33.500	175.196	(1, 19)	0.000000000
NDRE1	0.868	0.861	0.037	0.040	37.900	125.312	(1, 19)	0.000000001
NDRE2	0.875	0.868	0.036	0.039	37.200	132.649	(1, 19)	0.000000001
NDVI	0.668	0.650	0.058	0.075	59.700	38.198	(1, 19)	0.000006134
CCCI	0.914	0.909	0.030	0.595	31.200	201.453	(1, 19)	0.000000000
WDRVI	0.688	0.672	0.057	0.458	58.300	41.891	(1, 19)	0.000003342

2.5 Chlorophyll Exponential Model Summary Sentinel-2

Potato								
Index name	R2	AdjR2	RMSE	RMSEP(LOOCV)	NRMSE (%)	F - statistic	DF	p-value:
REP	0,907	0,902	0,032	0,034	32,400	184,218	(1, 19)	0,000000000
MTCI	0,934	0,930	0,026	0,028	27,000	267,764	(1, 19)	0,000000000
MCARI.OSAVI	0,199	0,157	0,083	0,094	85,100	4,721	(1, 19)	0,042656850
MCARI.OSAVI.RE	0,797	0,787	0,047	0,051	48,500	74,731	(1, 19)	0,000000052
TCARI.OSAVI	0,814	0,805	0,044	0,047	45,200	83,391	(1, 19)	0,000000022
TCARI.OSAVI.RE	0,921	0,917	0,029	0,031	29,400	220,841	(1, 19)	0,000000000
CI.red.edge	0,917	0,912	0,029	0,032	30,300	208,891	(1, 19)	0,000000000
CI.green	0,916	0,912	0,029	0,031	29,900	208,366	(1, 19)	0,000000000
SIPI	0,439	0,410	0,075	0,081	77,600	14,880	(1, 19)	0,001061020
NDRE	0,899	0,894	0,032	0,034	32,900	169,747	(1, 19)	0,000000000
NDRE1	0,873	0,867	0,036	0,039	37,500	130,858	(1, 19)	0,000000001
NDRE2	0,879	0,873	0,036	0,039	36,700	138,287	(1, 19)	0,000000000
NDVI	0,680	0,663	0,057	0,074	58,600	40,416	(1, 19)	0,000004240
CCCI	0,896	0,891	0,032	0,034	32,500	164,065	(1, 19)	0,000000000
WDRVI	0,703	0,687	0,055	0,060	56,900	44,878	(1, 19)	0,000002101

2.6 Chlorophyll Exponential Model Summary Landsat TM

Potato								
Index name	R2	AdjR2	RMSE	RMSEP(LOOCV)	NRMSE (%)	F - statistic	DF	p-value:
CI.green	0.905	0.900	0.031	0.033	32.000	181.105	(1, 19)	0.000000000
SIPI	0.525	0.500	0.070	0.089	72.100	21.033	(1, 19)	0.000201666
NDVI	0.714	0.699	0.054	0.070	55.500	47.429	(1, 19)	0.000001438
WDRVI	0.737	0.723	0.052	0.067	53.600	53.138	(1, 19)	0.000000648

Appendix 3: Independent validation correction

3.1 Independent validation result before correction

Potato	APEX		Sentinel_2		Landsat TM	
Index name	RMSEP linear model	RMSEP exponential	RMSEP linear model	RMSEP exponential	RMSEP linear model	RMSEP exponential
REP	82.293	82.871	82.179	82.809		
MTCI	81.514	83.210	83.612	85.988		
MCARI/OSAVI	106.027	112.639	114.392	120.205		
MCARI/OSAVI RE	77.961	78.571	80.130	81.391		
TCARI/OSAVI	88.862	91.728	87.015	88.250		
TCARI/OSAVI RE	77.498	77.576	79.074	79.484		
CI red edge	78.096	78.408	78.482	79.030		
CI green	77.117	76.605	77.082	76.591	76.527	75.991
NDNI	105.987	109.378				
SIPI	106.877	111.503	105.101	109.693	101.686	106.356
DCNI	105.391	113.044	107.390	114.537		
NDRE	80.558	80.309	83.277	84.770		
NDRE1	83.476	85.200	82.547	83.519		
NDRE2	82.640	83.766	77.978	77.511		
NDVI	94.562	98.505	93.488	97.411	90.667	94.337
CCCI	82.723	83.519	79.674	80.044		
WDRVI	93.003	96.810	91.867	95.629	88.879	92.294

3.2 Independent validation result for potato after correction

Potato	APEX		Sentinel_2		Landsat TM	
Index name	RMSEP linear model	RMSEP exponential	RMSEP linear model	RMSEP exponential	RMSEP linear model	RMSEP exponential
REP	44,409	47.99165	43,711	47.27601		
MTCI	43,139	48.55494	44,997	50.57939		
MCARI/OSAVI	68,763	75.29856	79,713	84.28920		
MCARI/OSAVI RE	32,850	35.85171	33,824	36.59070		
TCARI/OSAVI	49,804	54.57876	45,795	47.93020		
TCARI/OSAVI RE	38,185	42.27279	40,322	44.36829		
CI red edge	39,656	44.17593	39,672	44.32670		
CI green	38,361	42.09532	38,198	41.94839	37,231	41.11673
NDNI	65,971	66.83930				
SIPI	71,837	74.60016	69,636	72.49675	64,592	68.07821
DCNI	67,711	76.13777	71,425	78.90766		
NDRE	41,688	44.21893	41,328	44.58928		
NDRE1	41,120	44.71873	41,410	44.34732		
NDRE2	41,142	44.27583	41,344	45.58874		
NDVI	53,165	56.81686	51,631	55.56296	48,128	52.39030
CCCI	44,607	48.39629	43,599	48.64386		
WDRVI	51,439	55.19730	49,886	53.93214	46,335	50.68385

3.3 Summary statistic for the 15 plots of the maize sites

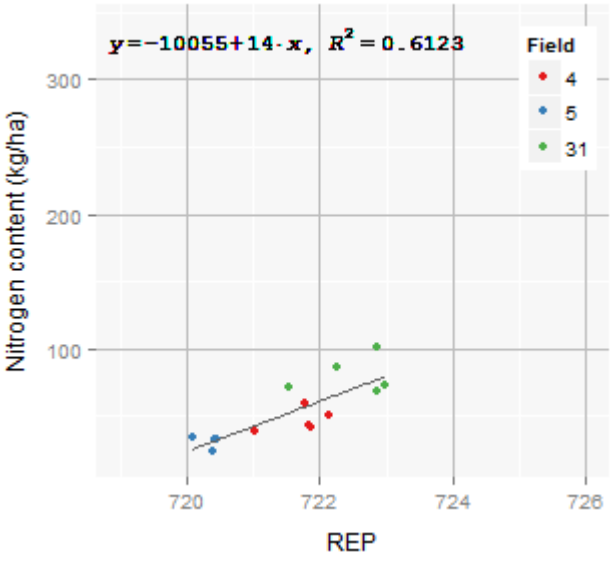
Maize	Minimum	Maximum	Mean	Standart deviation	Coeficient of variation (%)
Fresh weight (kg/ha)	29182.22	56435.56	41823.59	8149.79	19.49
Dry weight (kg/ha)	2655.58	5453.32	3661.07	673.31	18.39
N content (kg/ha)	90.82	132.48	109.92	13.68	12.44

3.4 Summary statistic of nitrogen content in potato sites per field

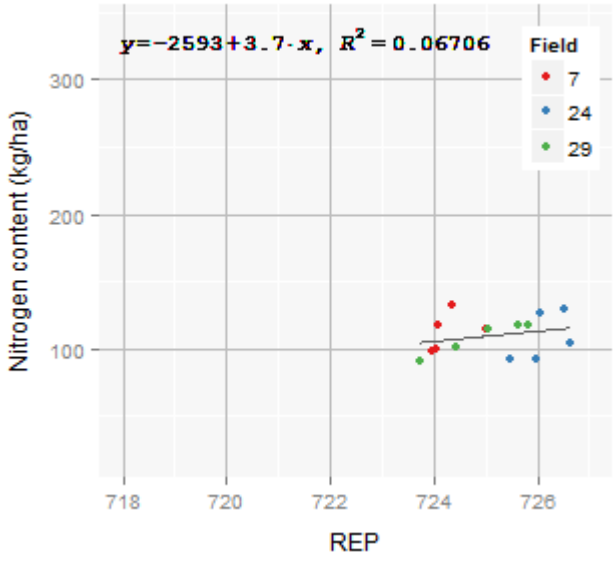
Potato per field	Minimum	Maximum	Mean	Standard deviation	Coefficient of variation (%)
N content field 11 (kg/ha)	145.12	255.27	195.77	48.02	0.25
N content field 19 (kg/ha)	87.83	258.08	158.66	63.32	0.40
N content field 30 (kg/ha)	220.40	258.32	239.44	15.10	0.06
N content field 28 (kg/ha)	81.66	356.43	196.12	88.48	0.45

Appendix 4: Scatter plots of nitrogen models

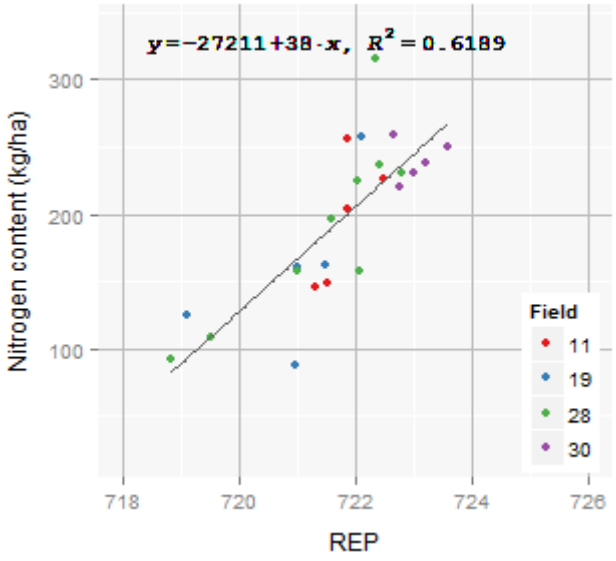
Grassland



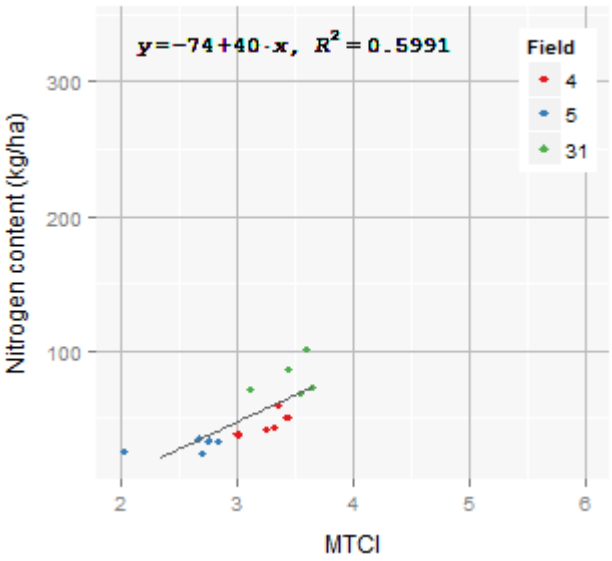
Maize



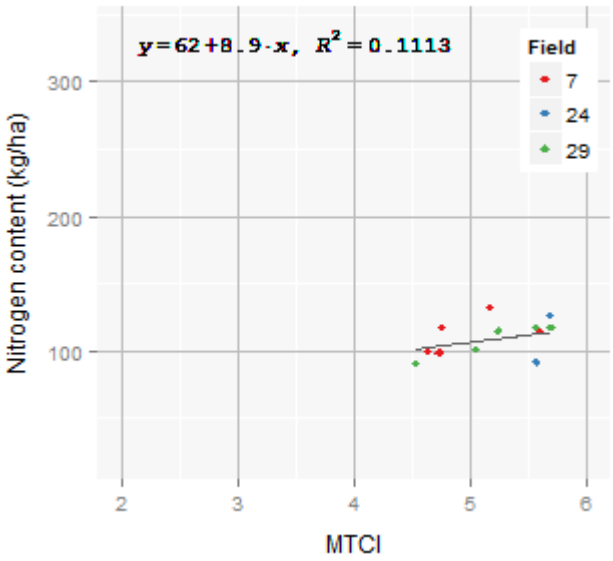
Potato



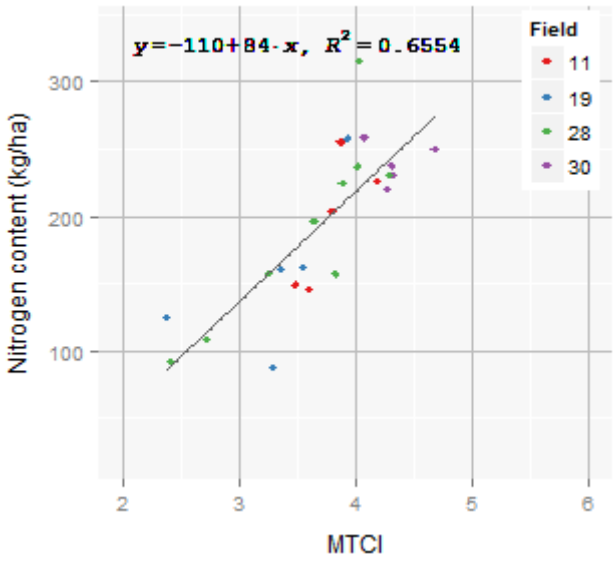
Grassland



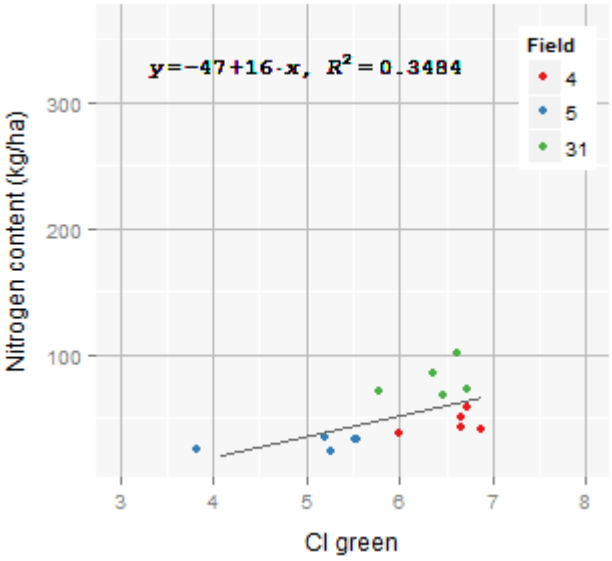
Maize



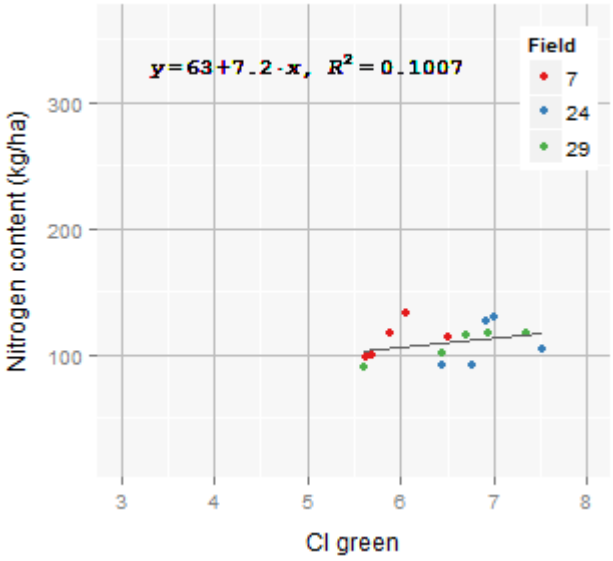
Potato



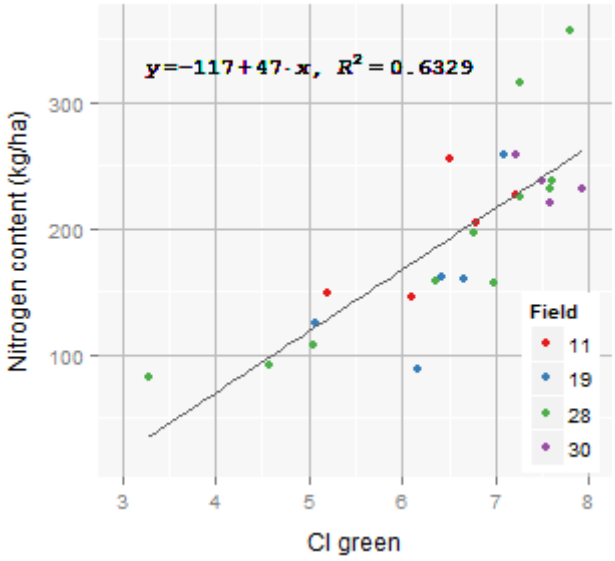
Grassland



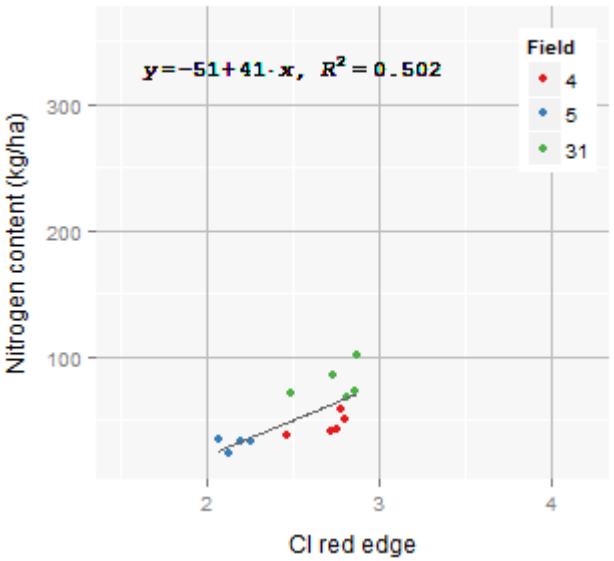
Maize



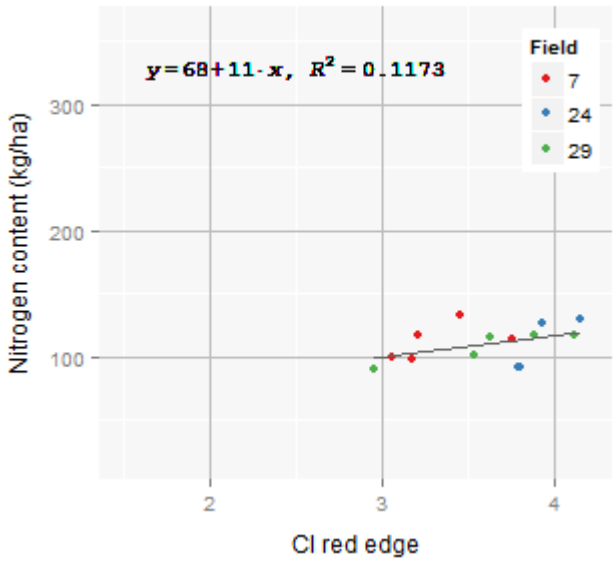
Potato



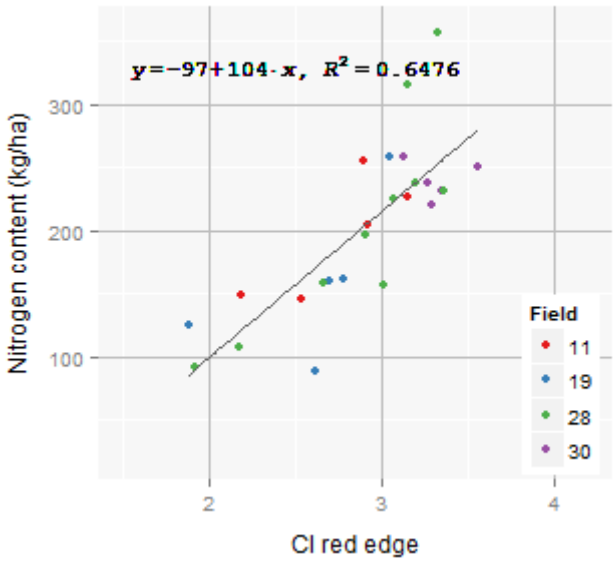
Grassland



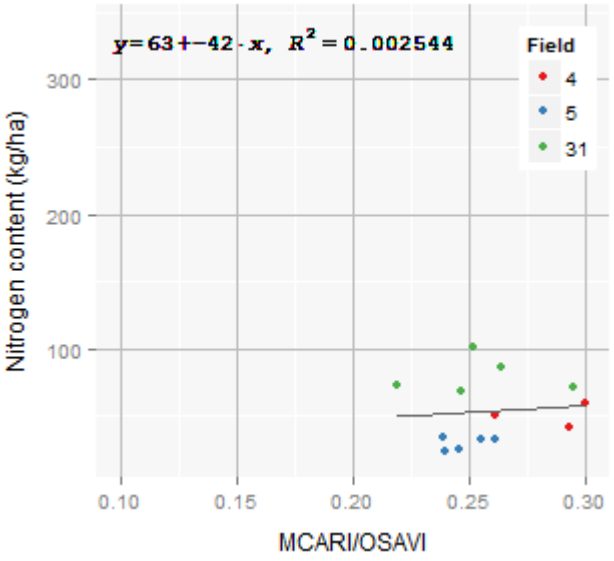
Maize



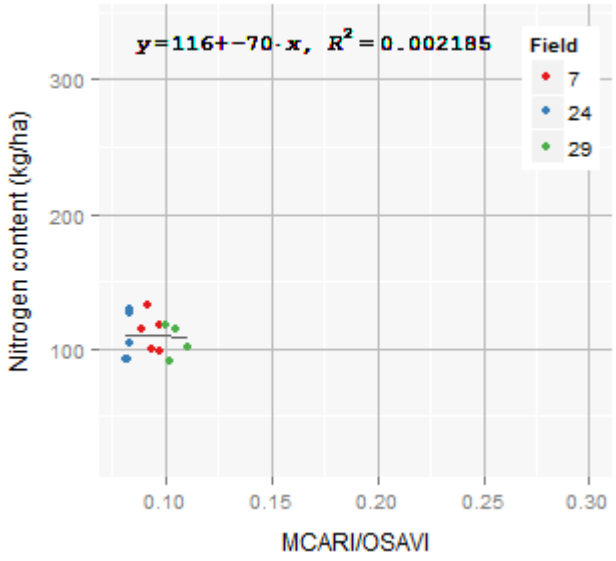
Potato



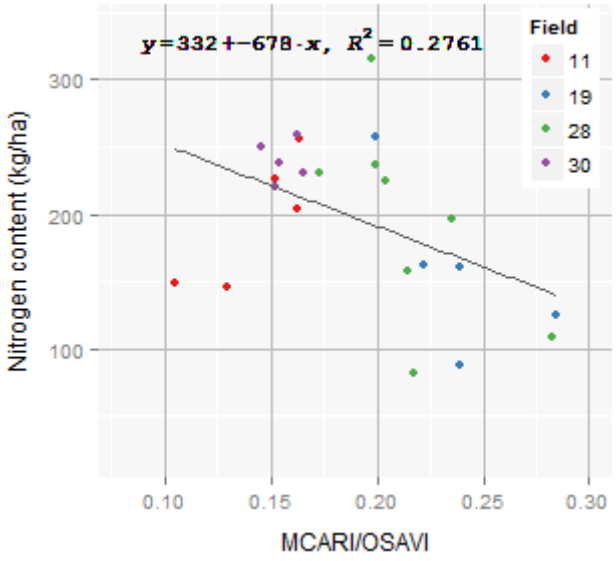
Grassland



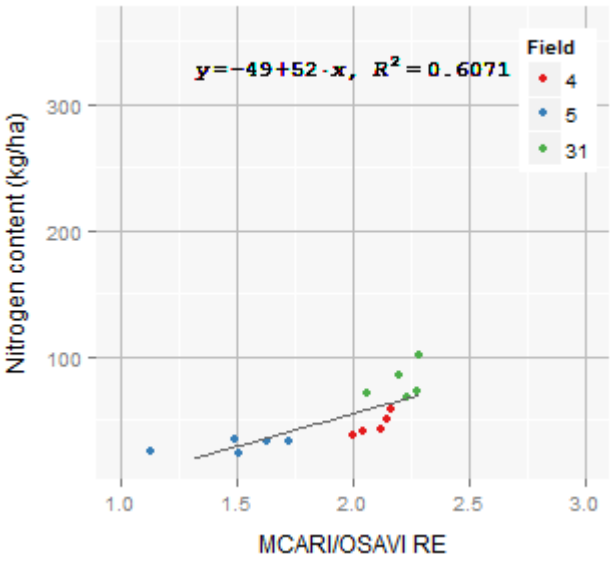
Maize



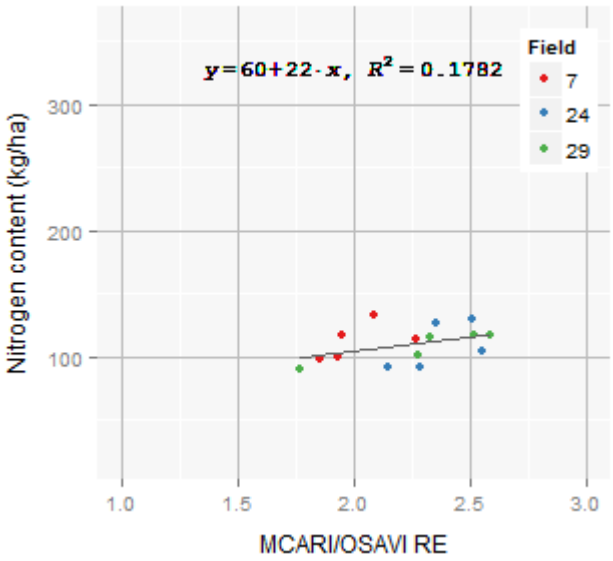
Potato



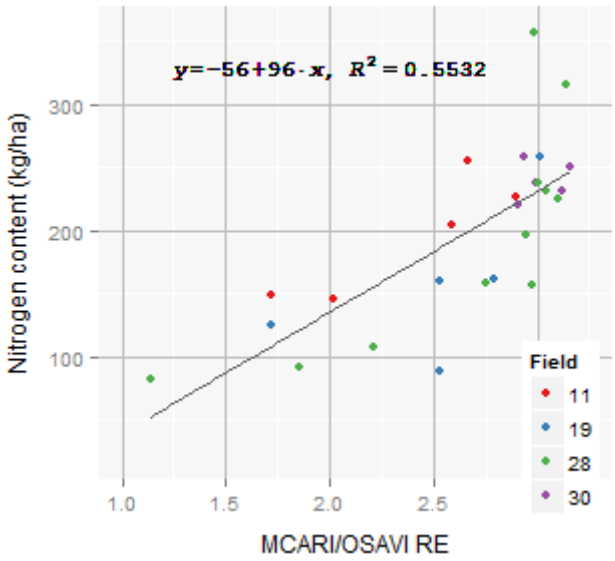
Grassland



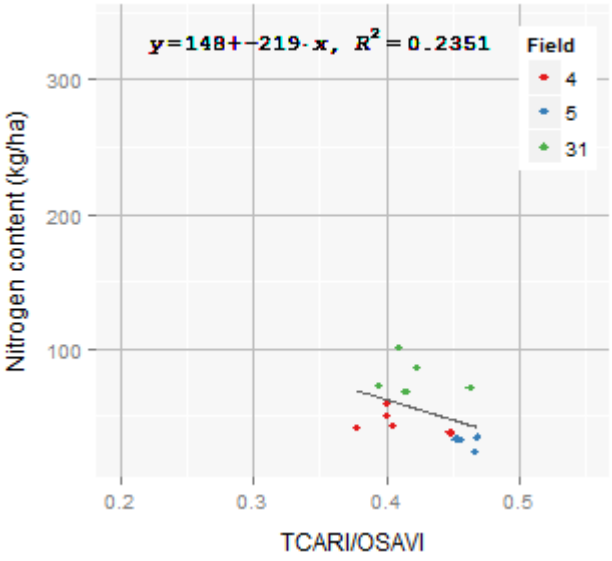
Maize



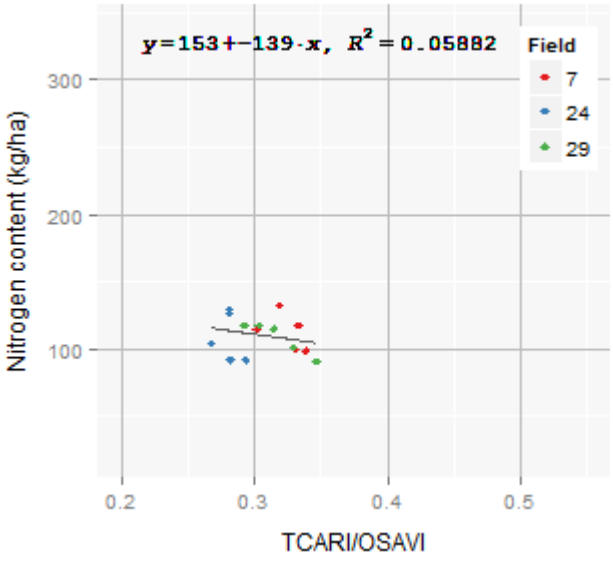
Potato



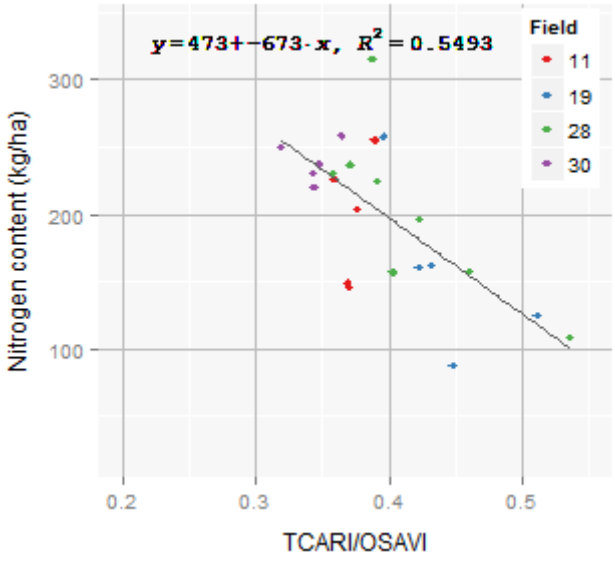
Grassland



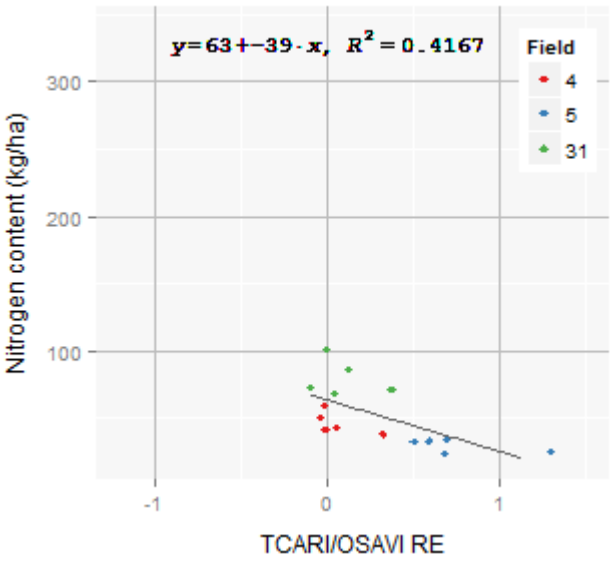
Maize



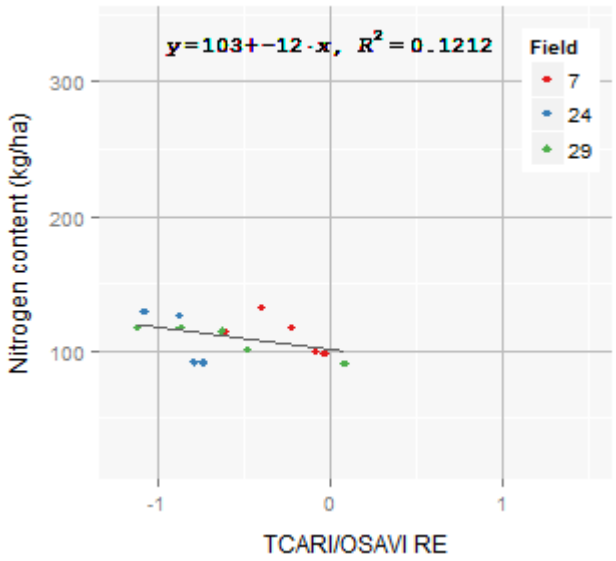
Potato



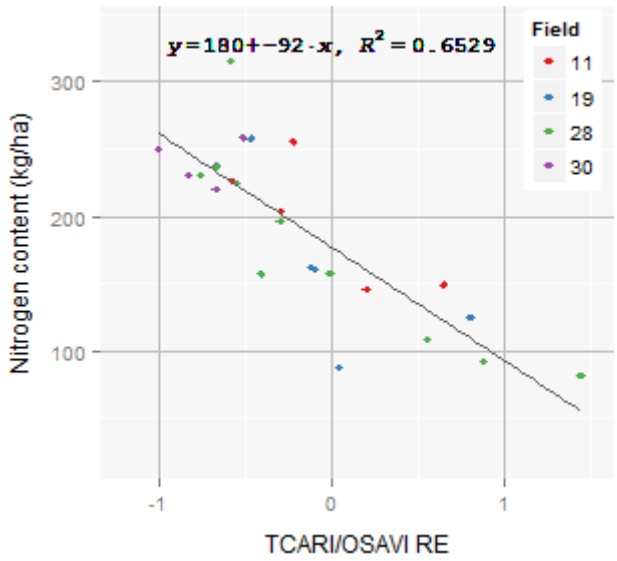
Grassland



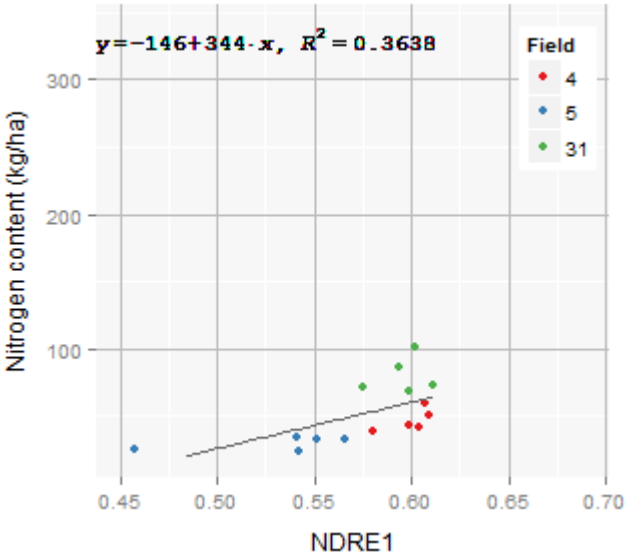
Maize



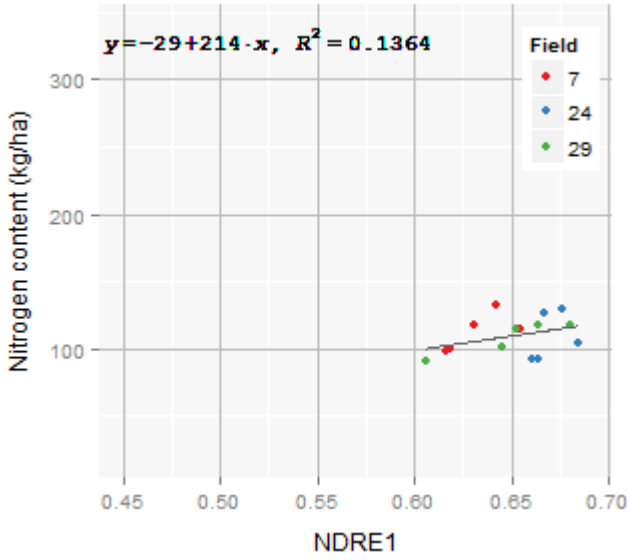
Potato



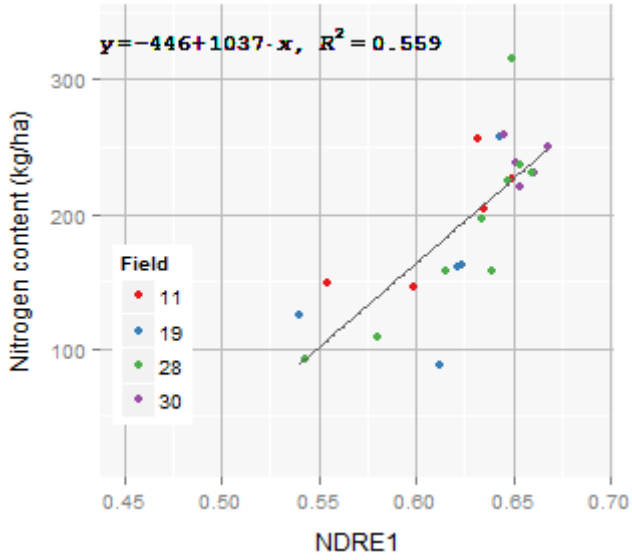
Grassland



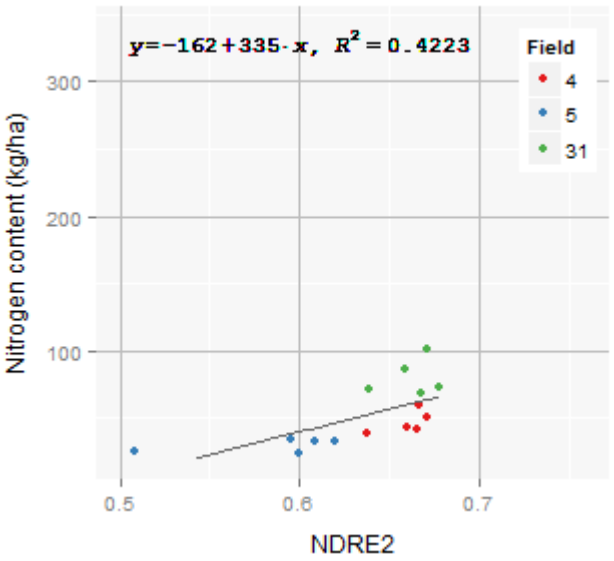
Maize



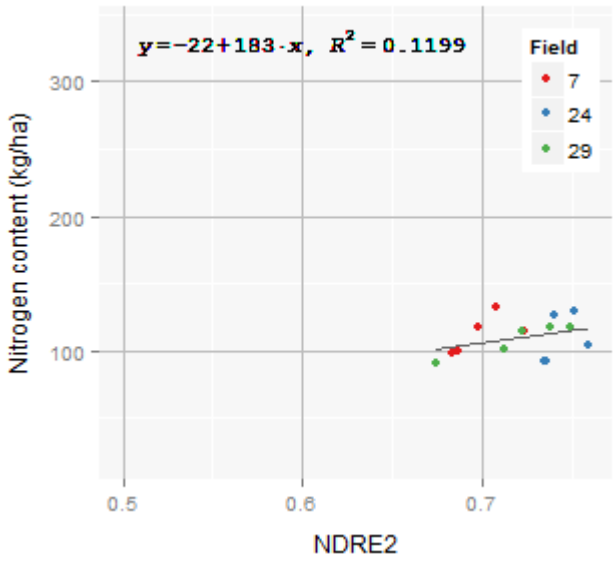
Potato



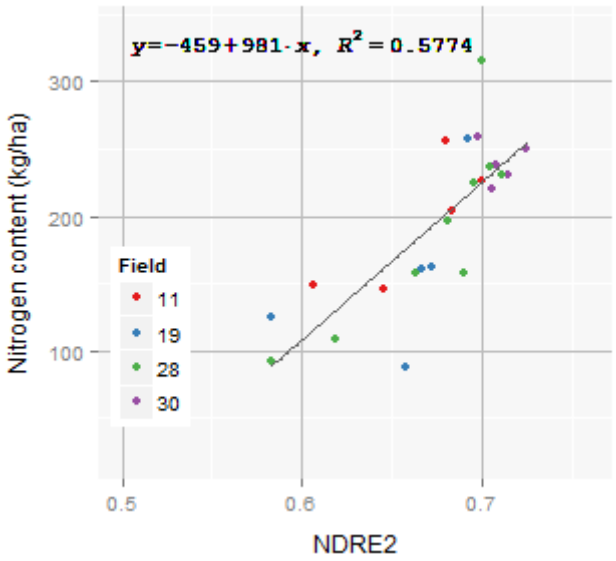
Grassland



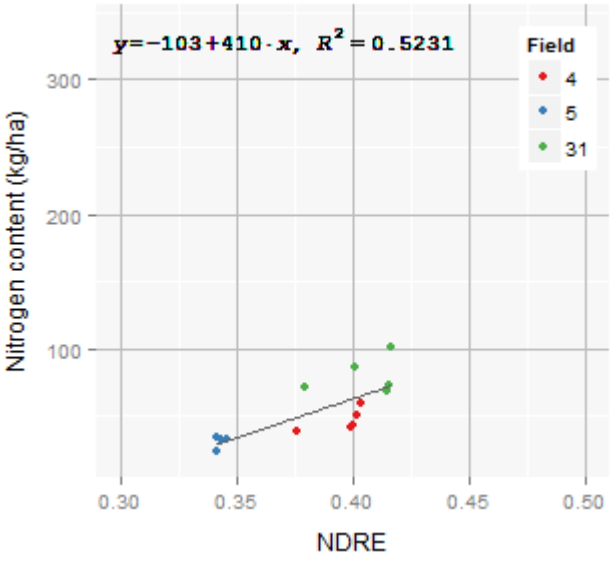
Maize



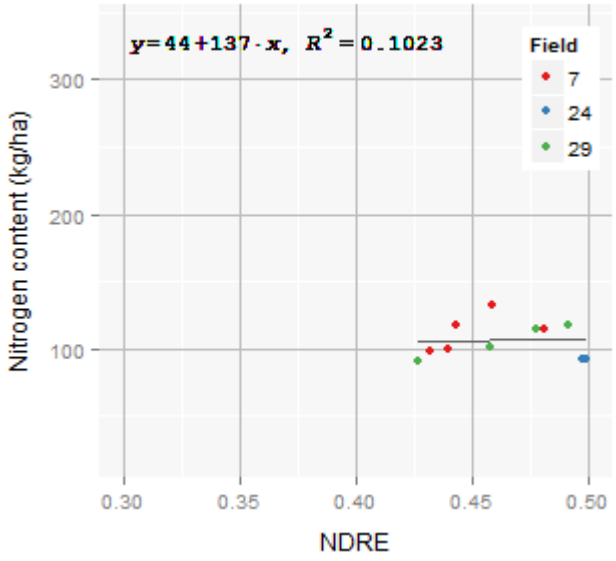
Potato



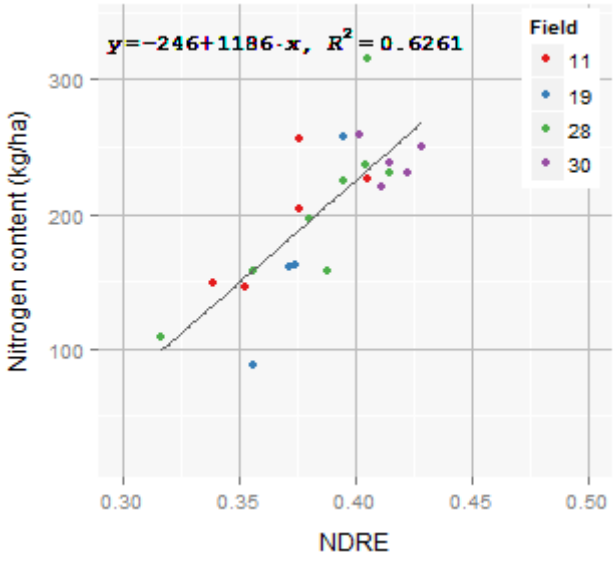
Grassland



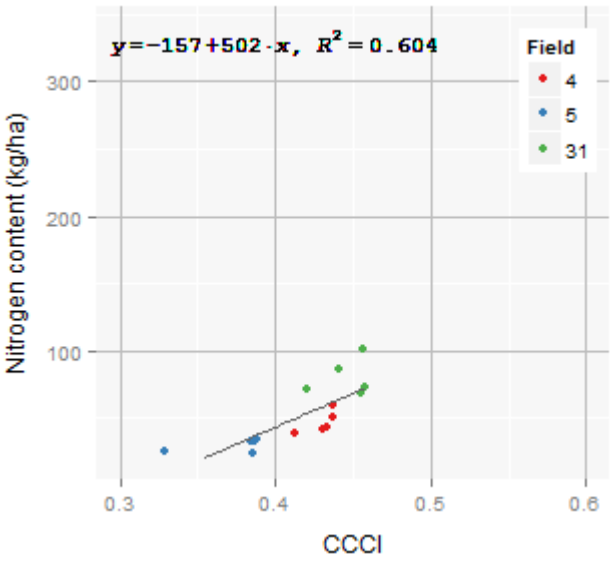
Maize



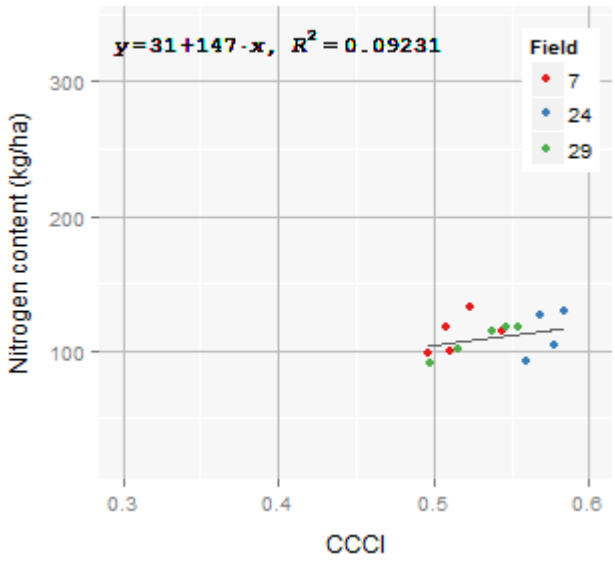
Potato



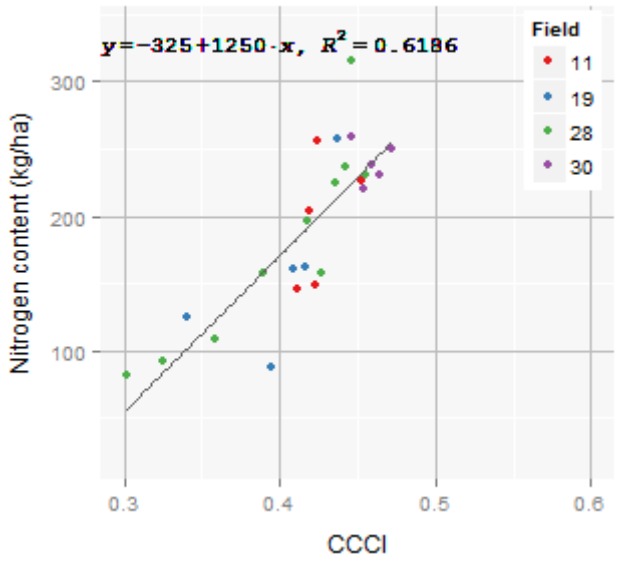
Grassland



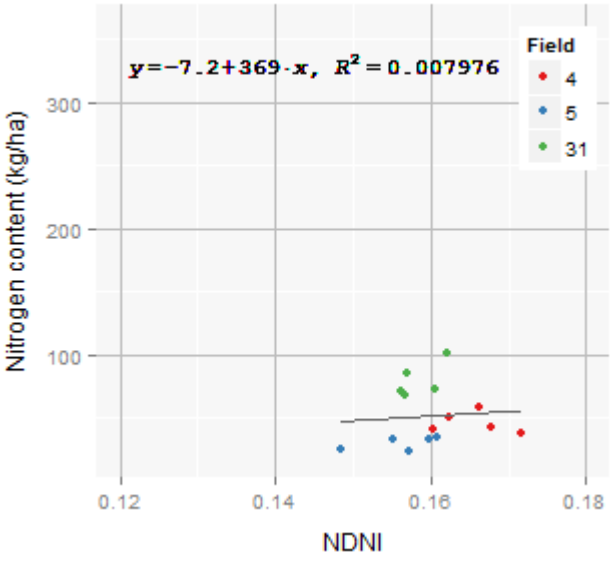
Maize



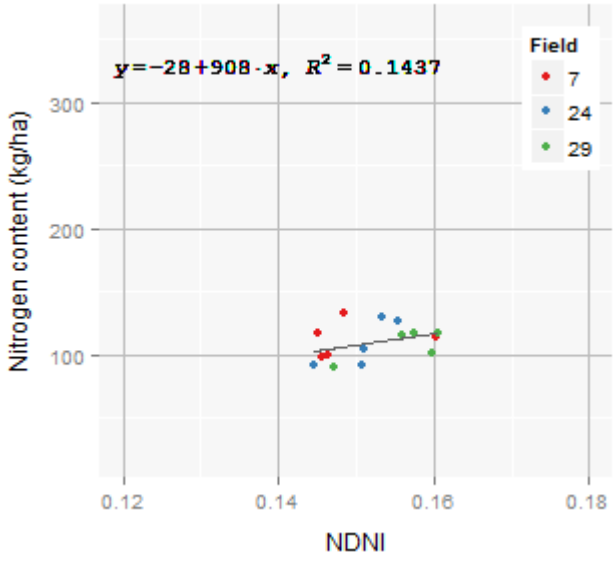
Potato



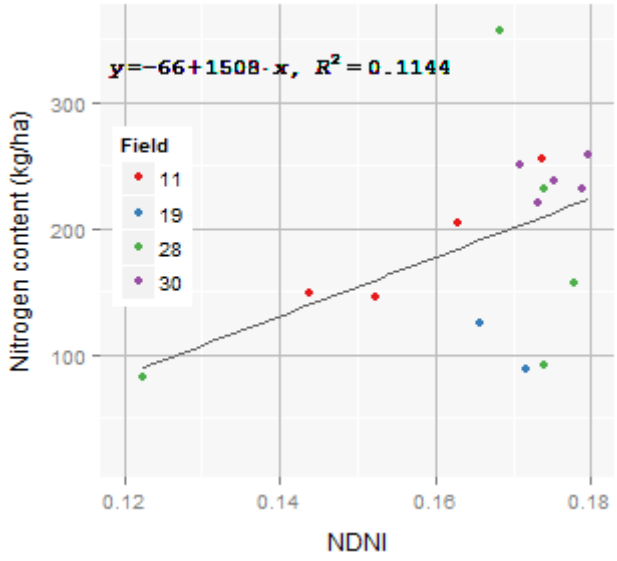
Grassland



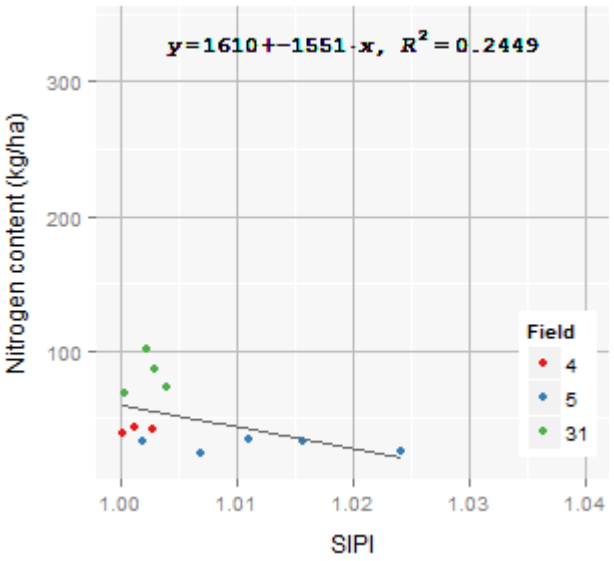
Maize



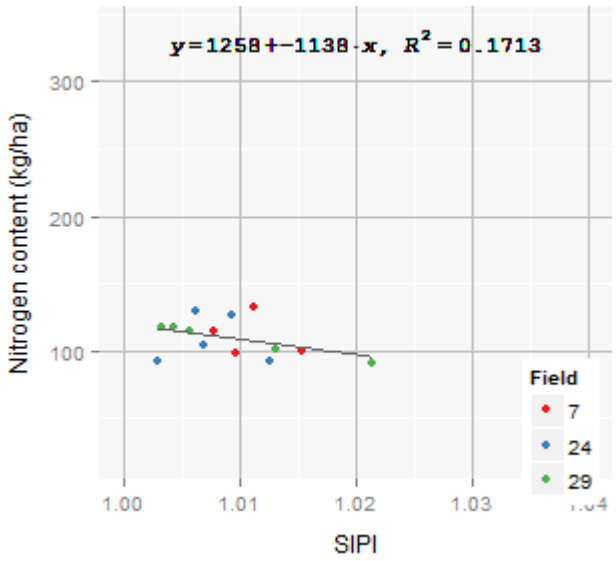
Potato



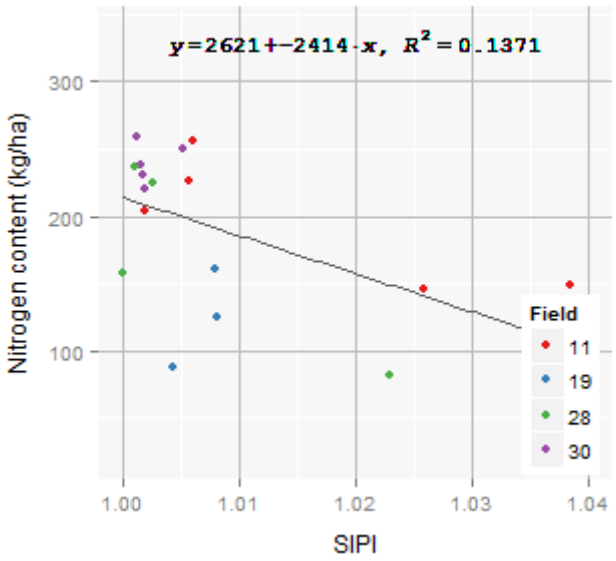
Grassland



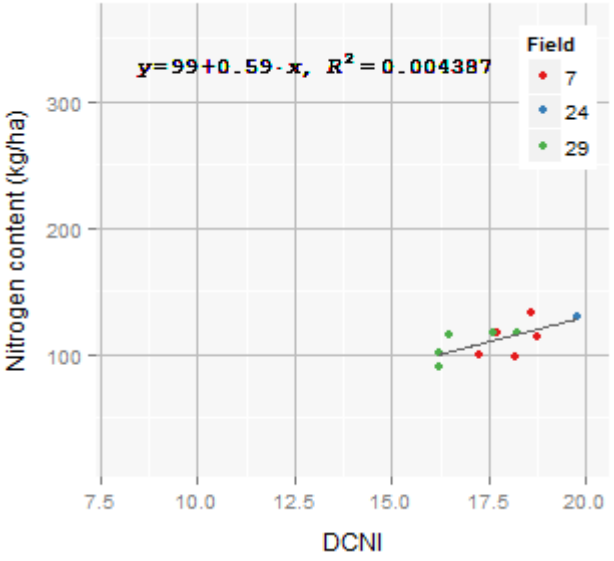
Maize



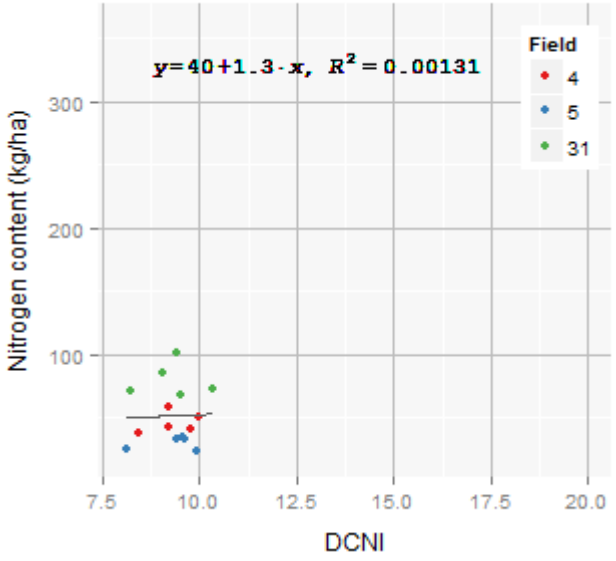
Potato



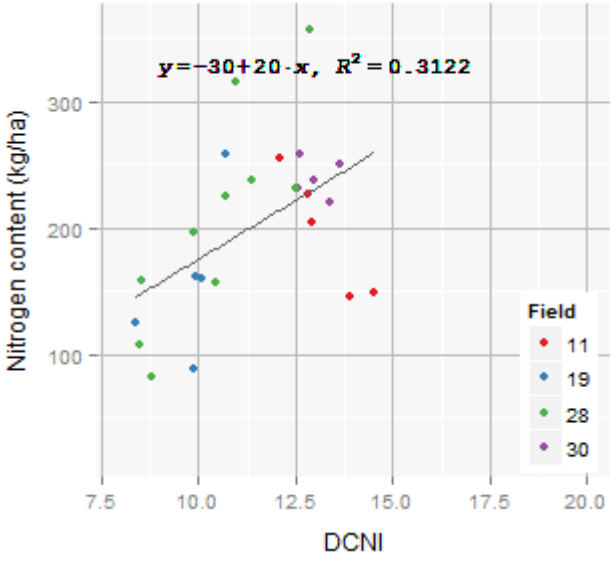
Grassland



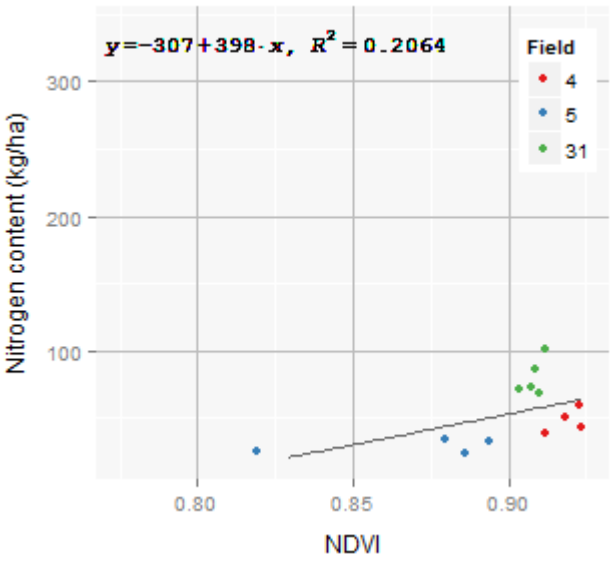
Maize



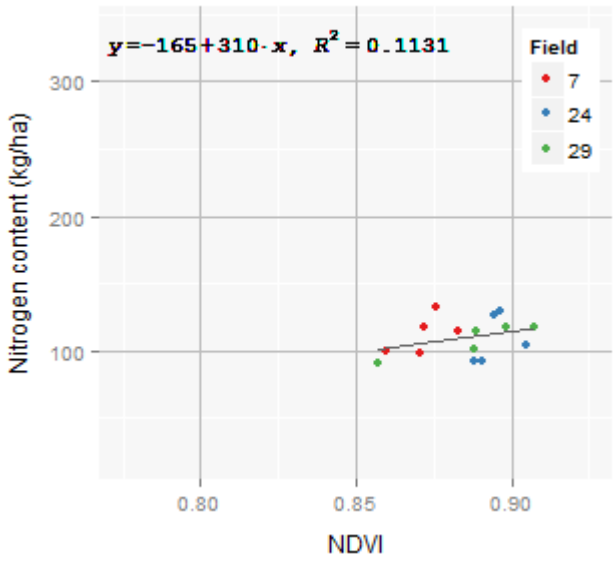
Potato



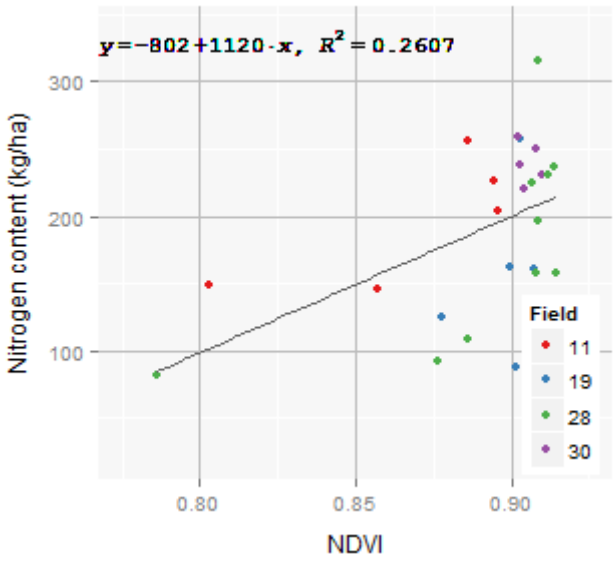
Grassland



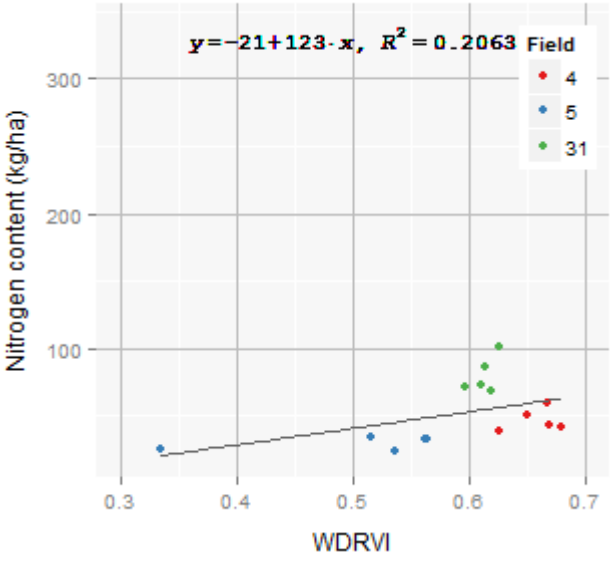
Maize



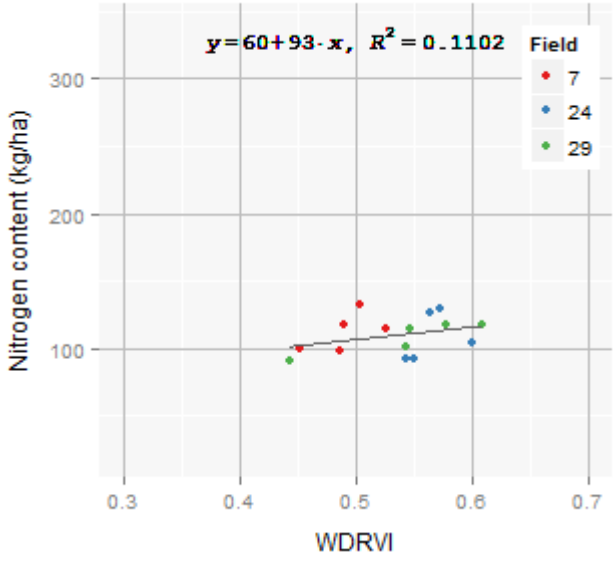
Potato



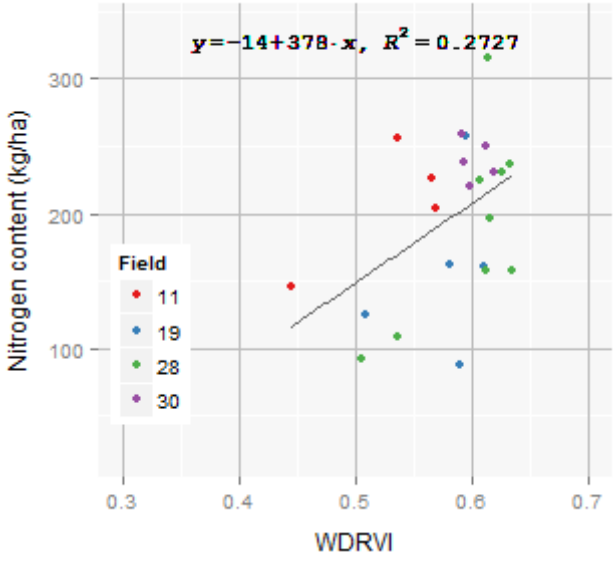
Grassland



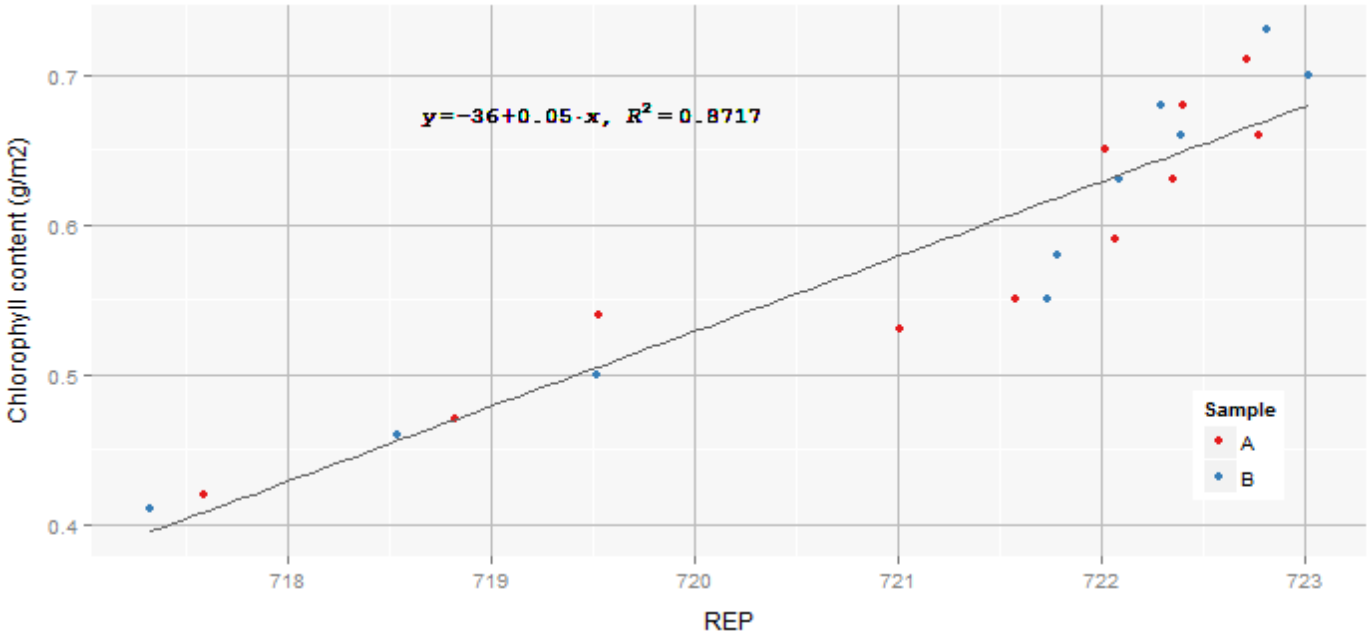
Maize

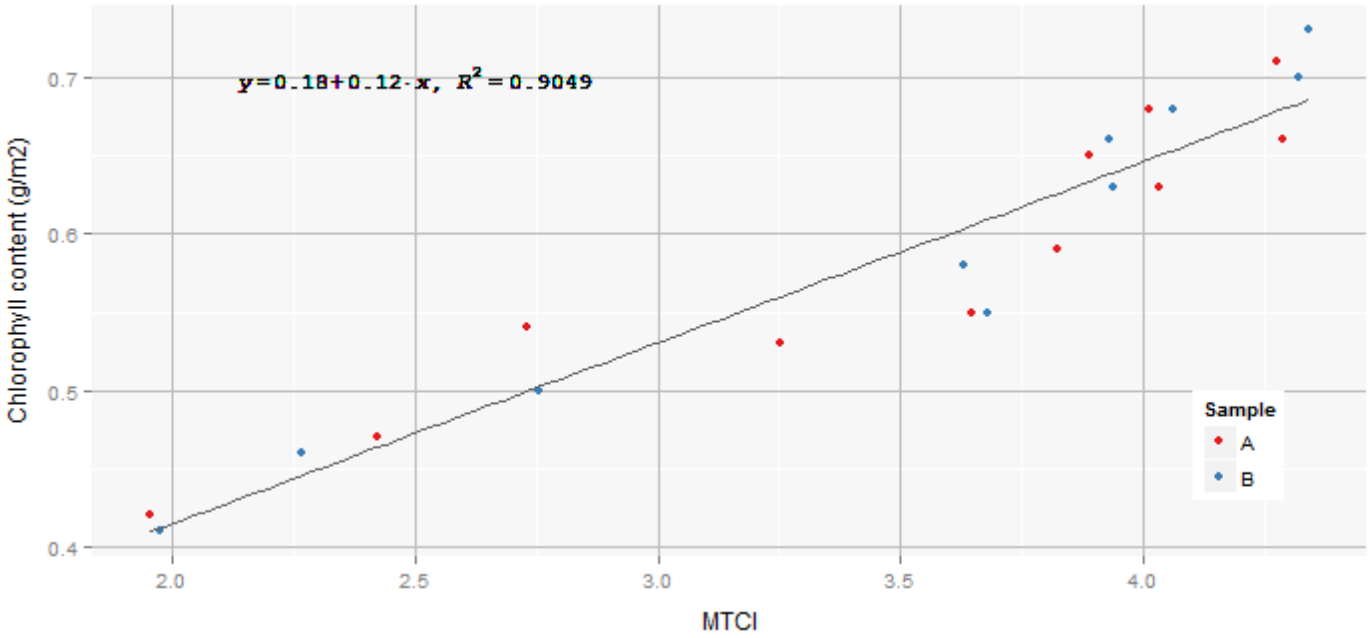


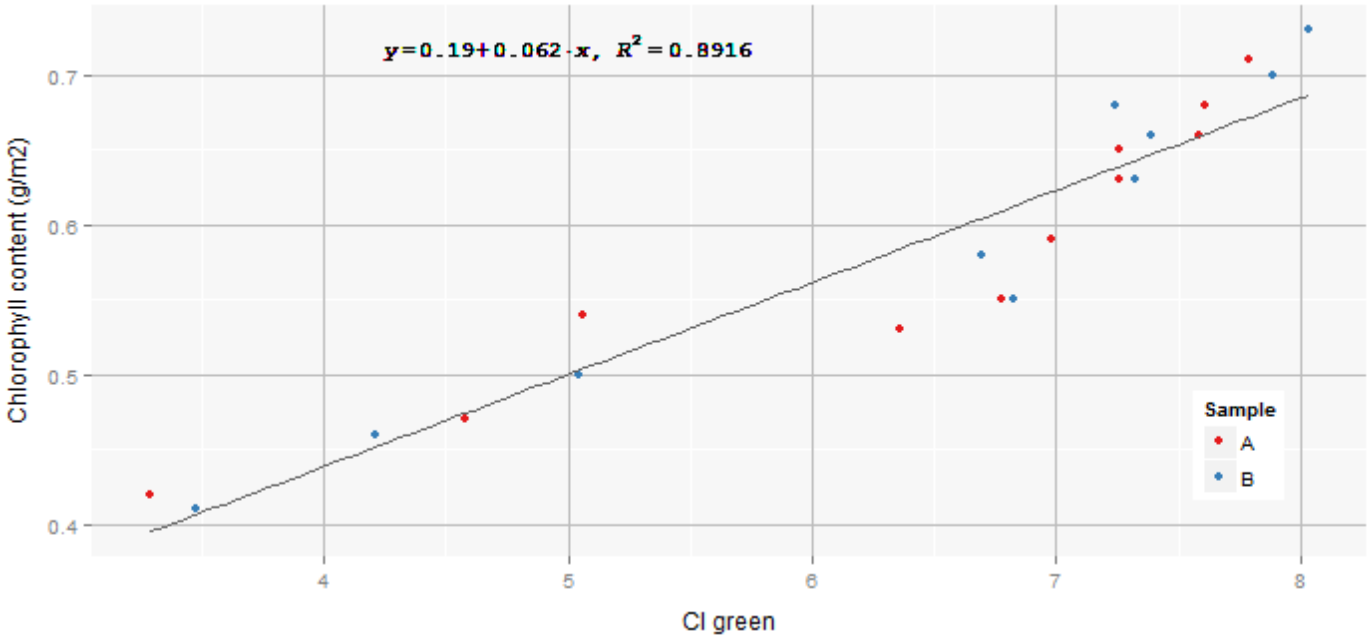
Potato

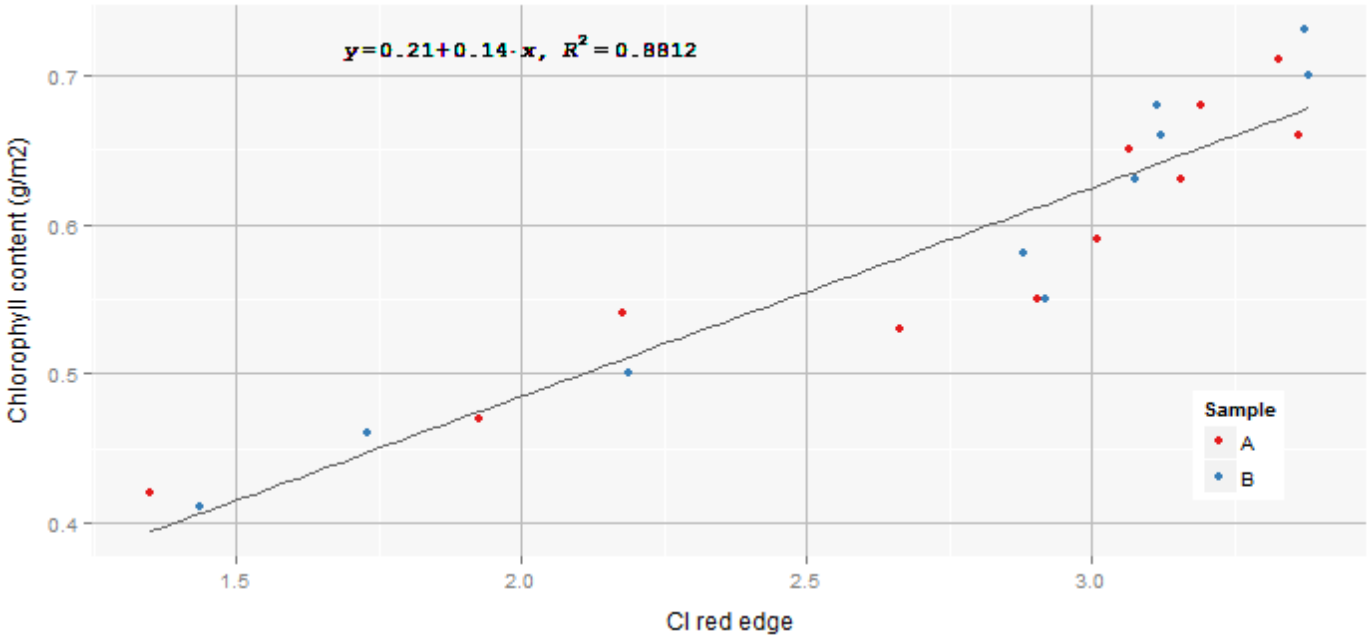


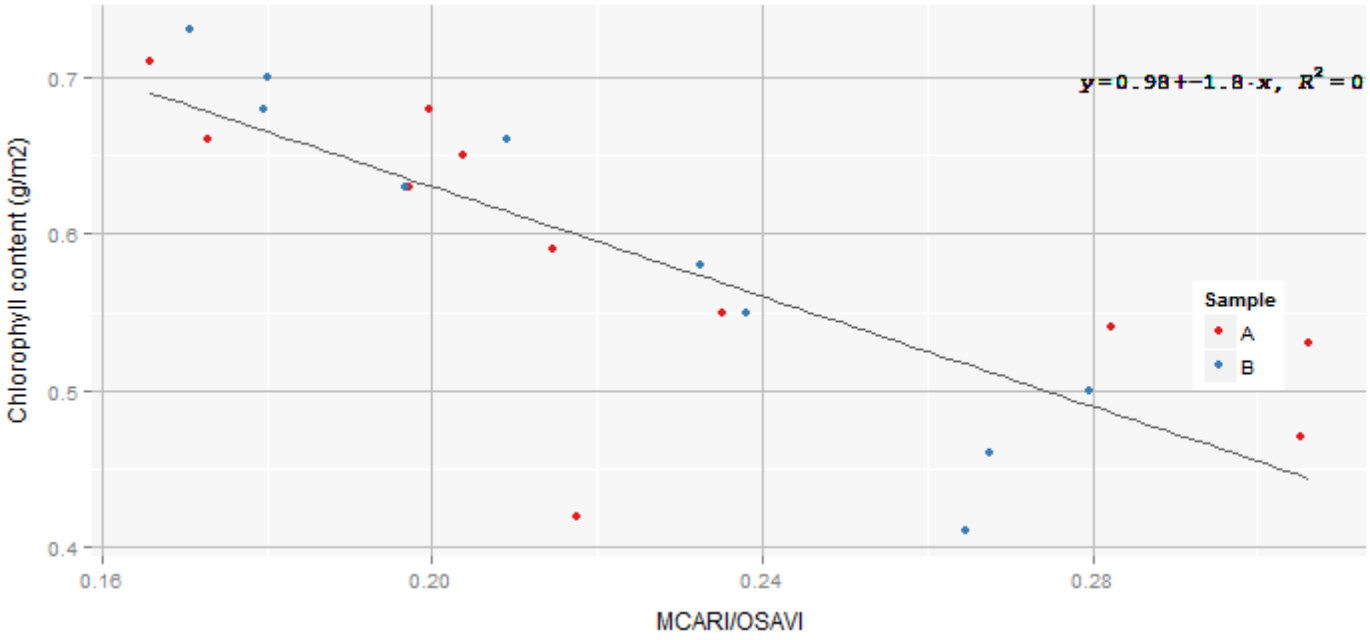
Appendix 5: Scatter plots of chlorophyll models

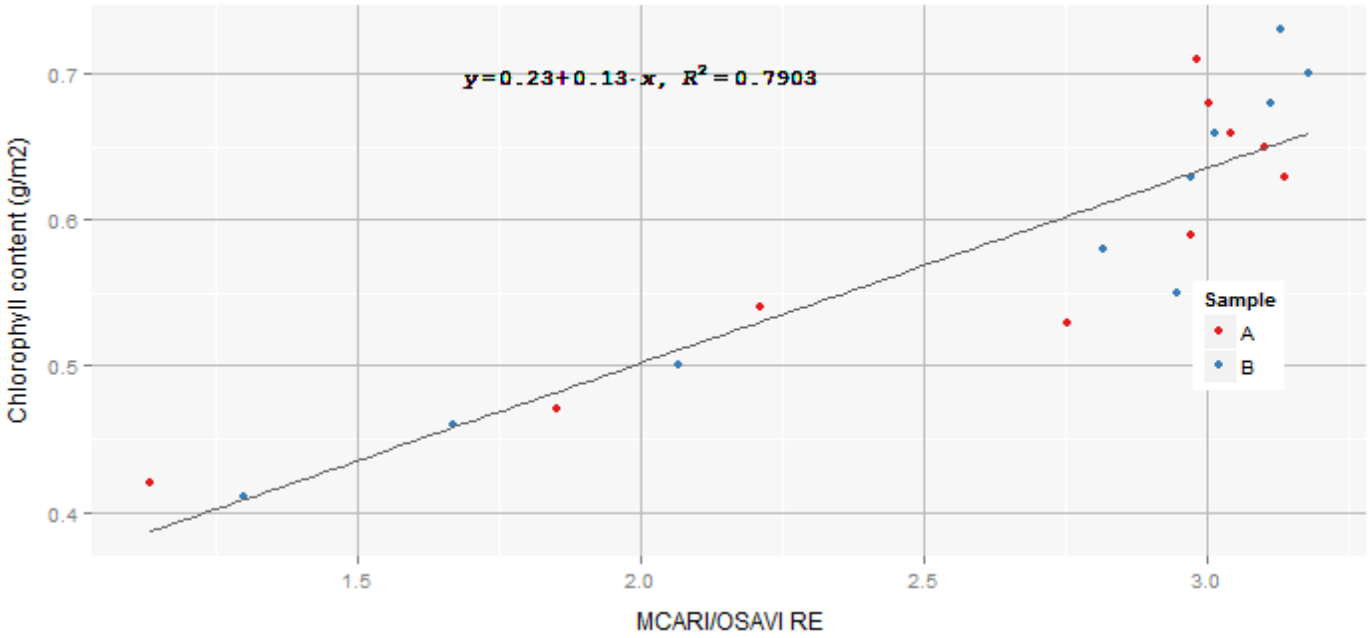


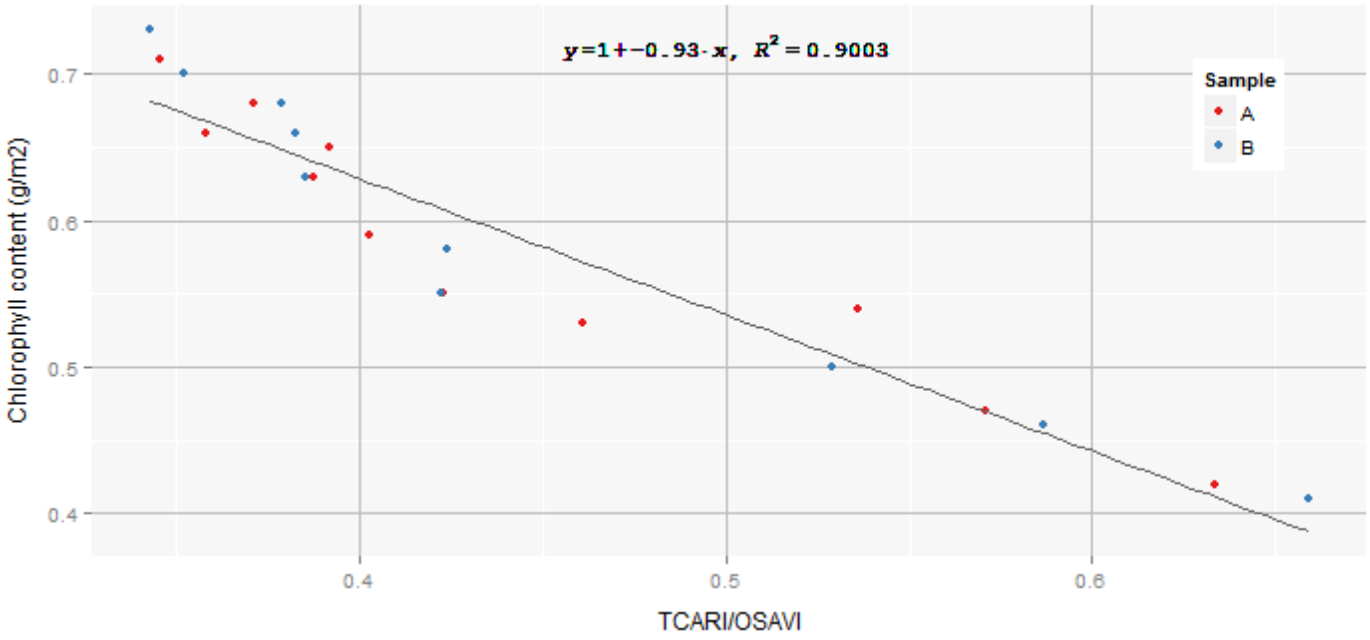


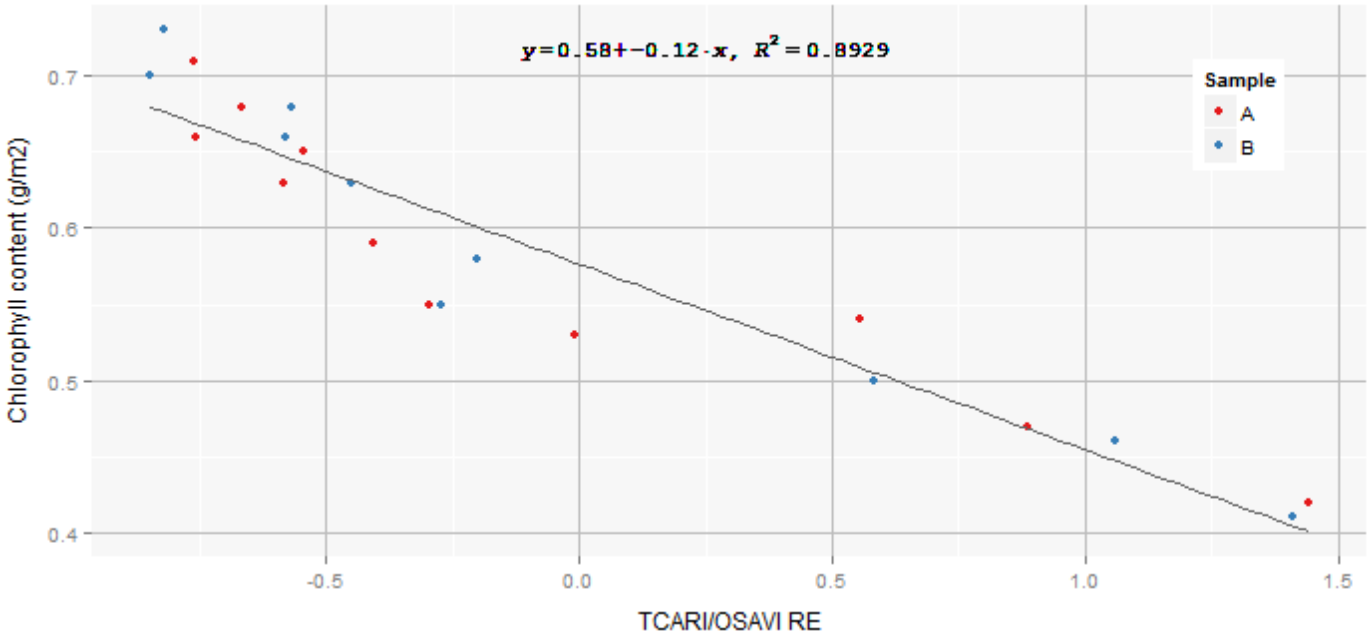

y=0.19+0.062·x, R² = 0.8916
Chlorophyll content (g/m2)
0.7
0.6
0.5
0.4
4
5
6
7
8
Cl green
Sample
A
B

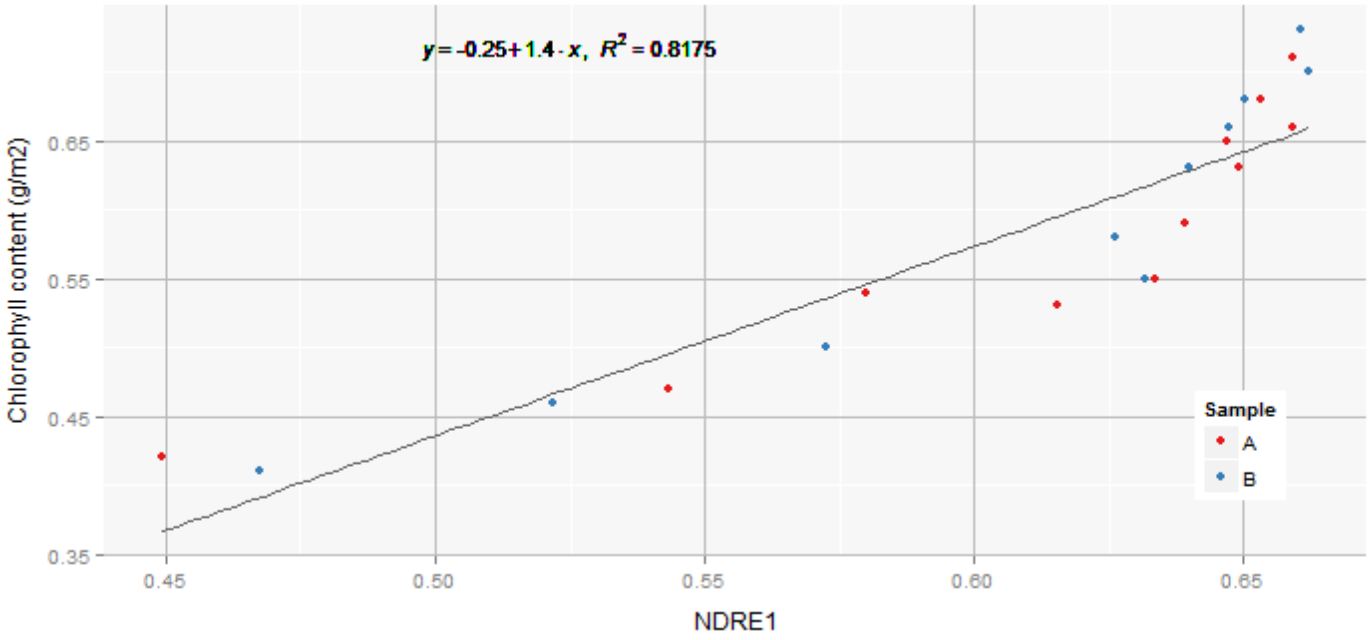

y=0.21+0.14·x, R² = 0.8812
Chlorophyll content (g/m2)
0.7
0.6
0.5
0.4
1.5
2.0
2.5
3.0
Cl red edge
Sample
A
B

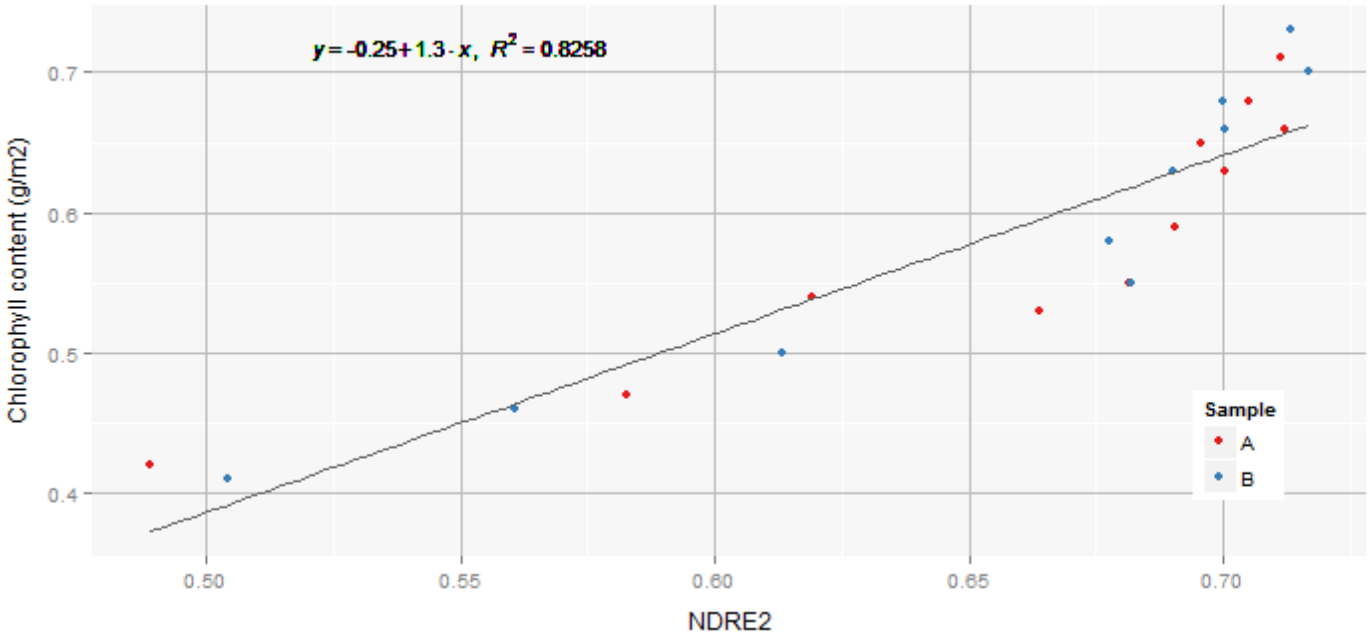

y=0.98+−1.8·x, R² = 0
Chlorophyll content (g/m2)
0.7
0.6
0.5
0.4
0.16
0.20
0.24
0.28
MCARI/OSAVI
Sample
A
B

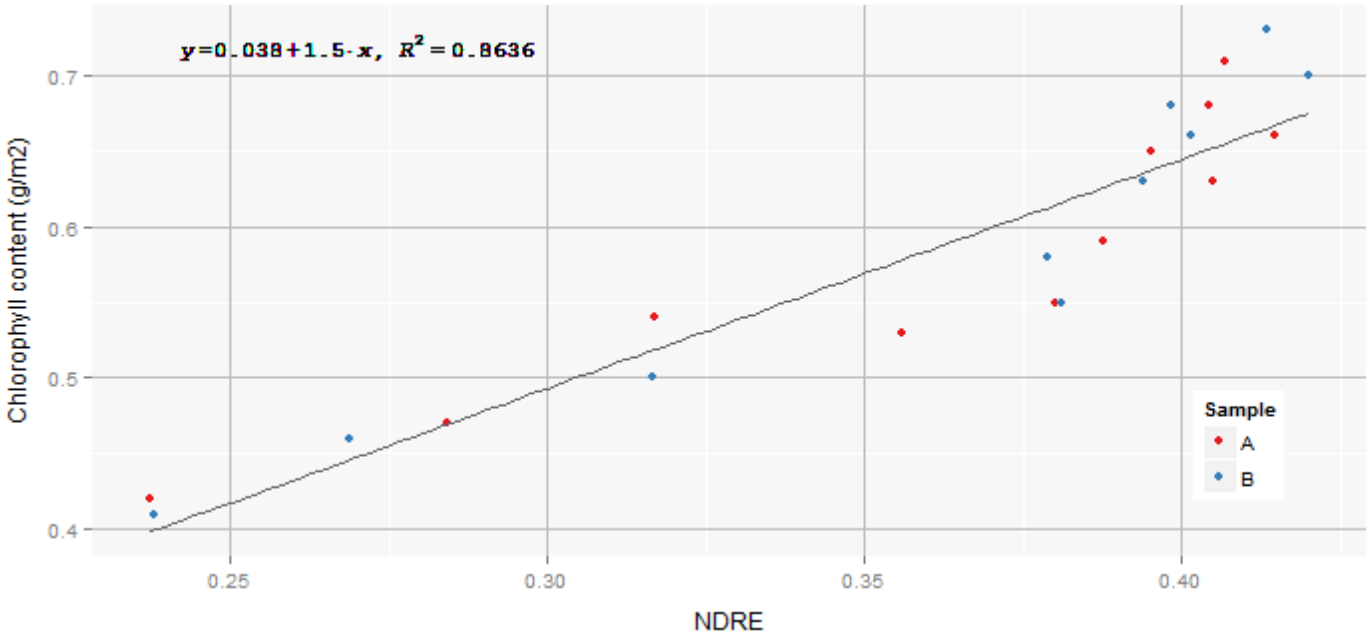

y=0.23+0.13·x, R²=0.7903
Chlorophyll content (g/m2)
MCARI/OSAVI RE
Sample
A
B

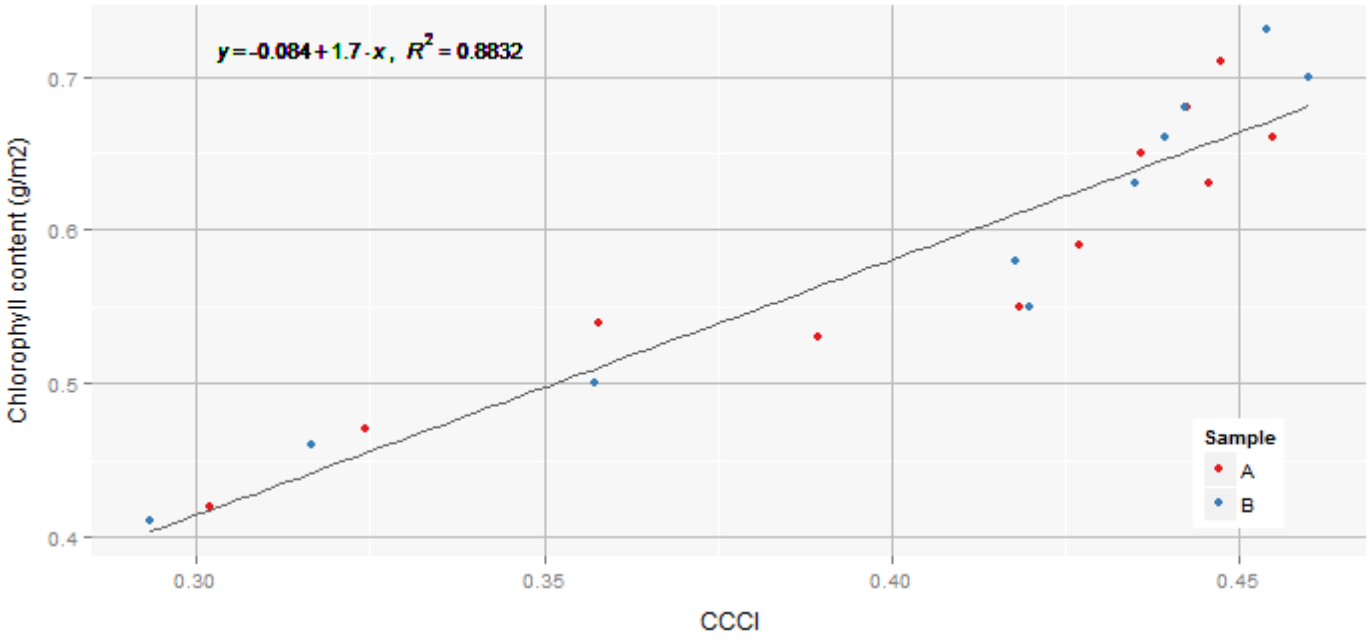

y=1+−0.93·x, R²=0.9003
Chlorophyll content (g/m2)
TCARI/OSAVI
Sample
A
B

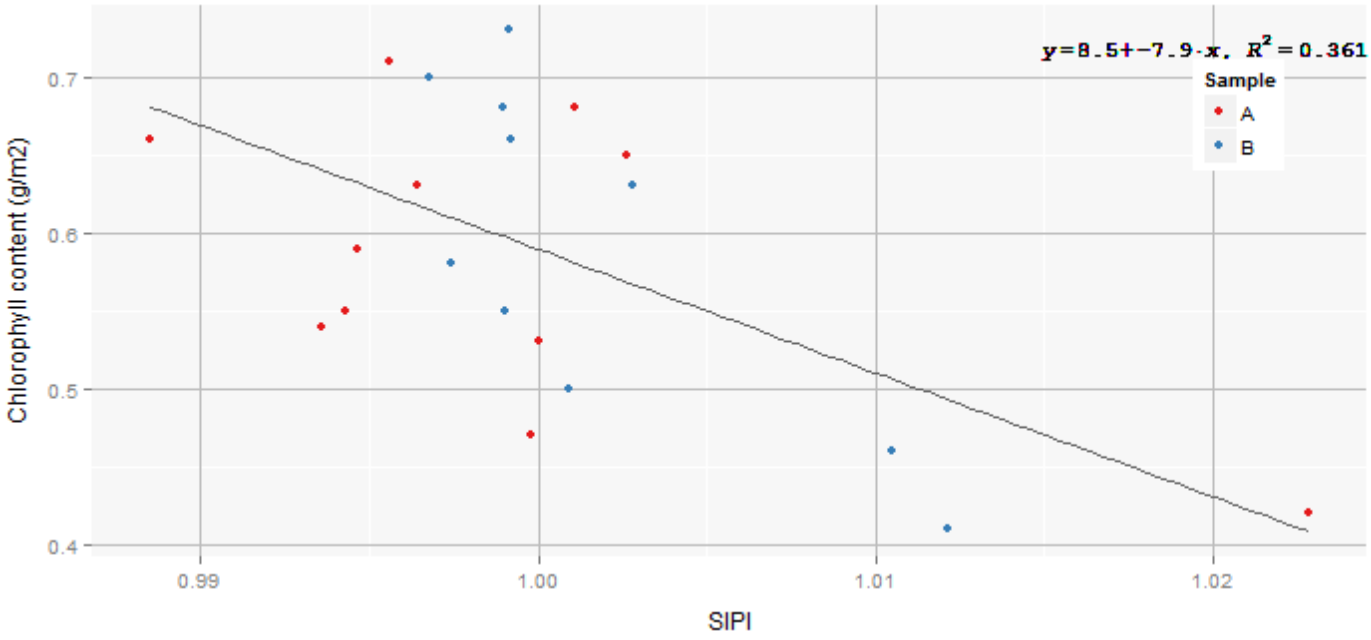

y=0.58+−0.12·x, R²=0.8929
Chlorophyll content (g/m2)
TCARI/OSAVI RE
Sample
A
B

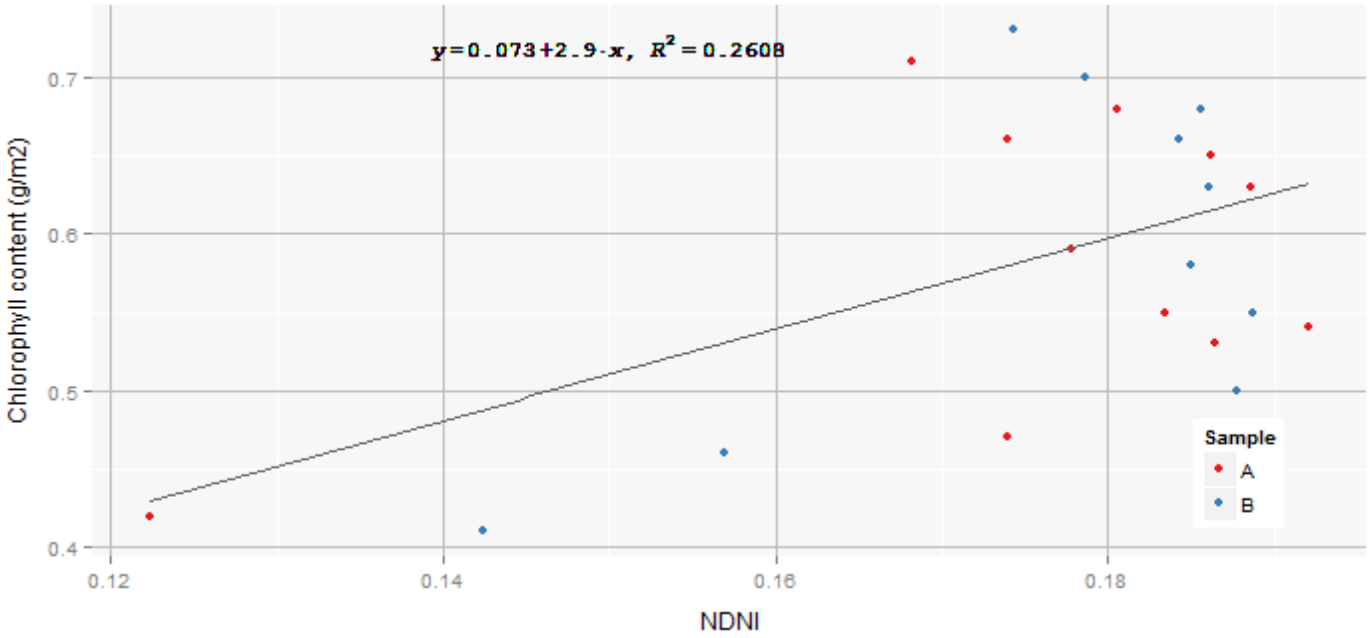

y = -0.25+1.4 · x, R² = 0.8175
Chlorophyll content (g/m2)
0.65
0.55
0.45
0.35
0.45
0.50
0.55
0.60
0.65
NDRE1
Sample
A
B

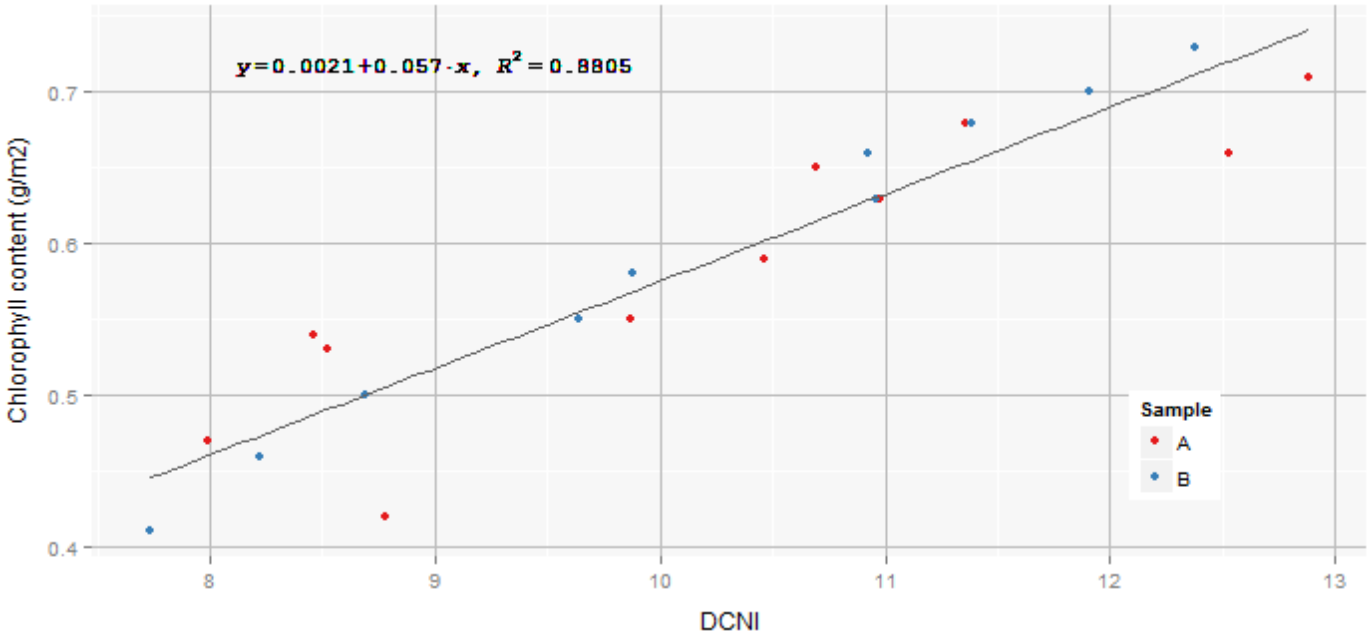

y = -0.25+1.3 · x, R² = 0.8258
Chlorophyll content (g/m2)
0.7
0.6
0.5
0.4
0.50
0.55
0.60
0.65
0.70
NDRE2
Sample
A
B

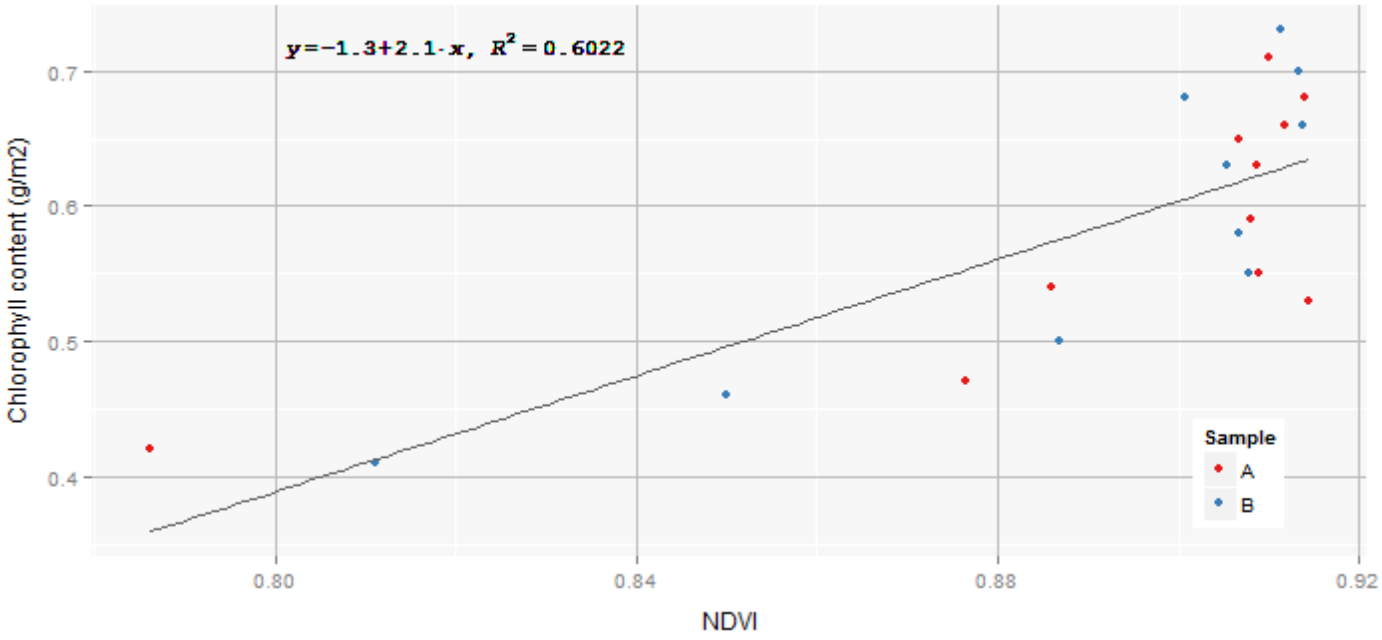

y = 0.038+1.5 · x, R² = 0.8636
Chlorophyll content (g/m2)
0.7
0.6
0.5
0.4
0.25
0.30
0.35
0.40
NDRE
Sample
A
B

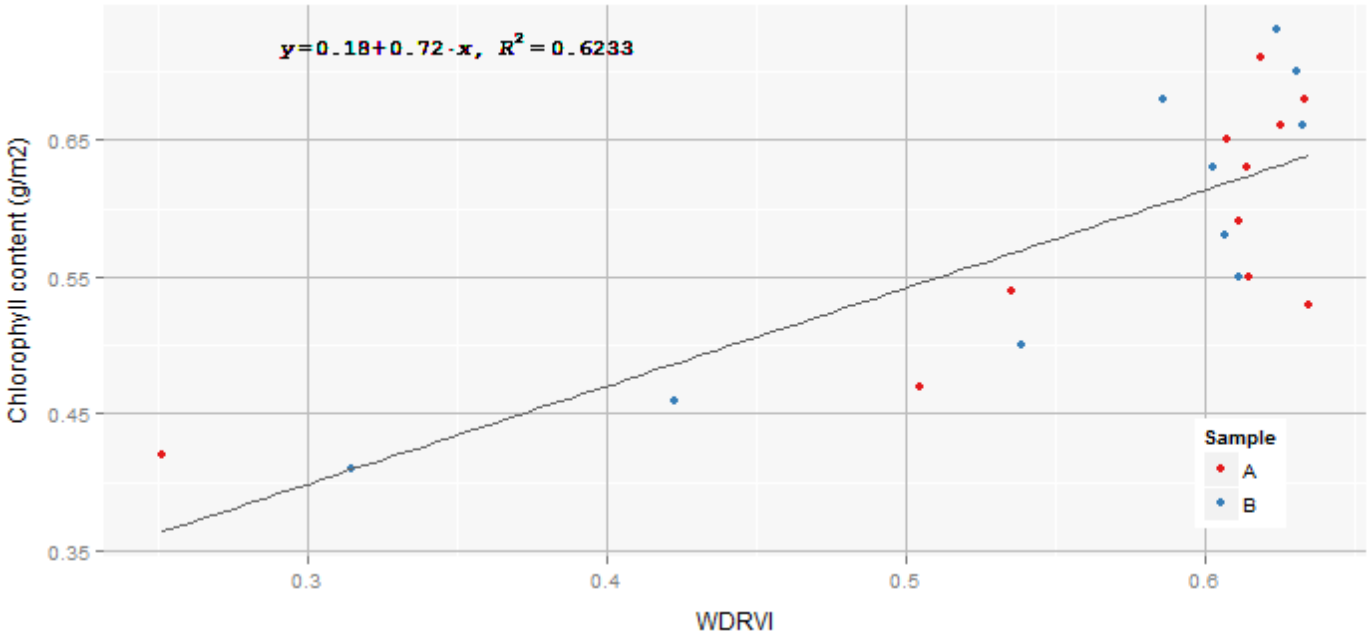

y=-0.084+1.7·x, R² = 0.8832
Chlorophyll content (g/m2)
0.7
0.6
0.5
0.4
0.30
0.35
0.40
0.45
CCCI
Sample
A
B


y=8.5+−7.9·x, R² = 0.361
Chlorophyll content (g/m2)
0.7
0.6
0.5
0.4
0.99
1.00
1.01
1.02
SIPI
Sample
A
B


y=0.073+2.9·x, R² = 0.2608
Chlorophyll content (g/m2)
0.7
0.6
0.5
0.4
0.12
0.14
0.16
0.18
NDNI
Sample
A
B


y=0.0021+0.057·x, R² = 0.8805
Chlorophyll content (g/m2)
0.7
0.6
0.5
0.4
8
9
10
11
12
13
DCNI
Sample
A
B


y=−1.3+2.1·x, R² = 0.6022
Chlorophyll content (g/m2)
0.7
0.6
0.5
0.4
0.80
0.84
0.88
0.92
NDVI
Sample
A
B


y=0.18+0.72·x, R² = 0.6233
Chlorophyll content (g/m2)
0.65
0.55
0.45
0.35
0.3
0.4
0.5
0.6
WDRVI
Sample
A
B