Laboratory of Geo-Information Science and Remote Sensing

Thesis Report GIRS-2015-17

DERIVING ALERTING SERVICES FROM CROP MONITORING TIME-SERIES IN PRECISION AGRICULTURE

Ratih Nurhayati

May 2015

WAGENINGEN UNIVERSITY
WAGENINGEN UR

Deriving Alerting Services from Crop Monitoring Time-Series in Precision Agriculture

Ratih Nurhayati

Registration number 88 01 21 680 100

Supervisor:

Dr. ir. Lammert Kooistra

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

May, 2015
Wageningen, The Netherlands

Thesis code number: GRS-80436
Thesis Report: GIRS-2015-17
Wageningen University and Research Centre
Laboratory of Geo-Information Science and Remote Sensing

This thesis is dedicated to my mother and the memory of my father,
Yeyet Hidayat (1953-2014).

Acknowledgement

I could never thank enough to the people around me during this thesis research. First of all, I would like to praise my God for giving me strength to pass all the challenges I had during my thesis research and for writing the best scenario for my life. I also want to thank the Netherlands' government for giving me the opportunity to study in Wageningen University through the StuNed Scholarship Programme. A special thanks to my great supervisor Lammert Kooistra for his guidance and never ending support during the whole thesis work process. He always tracked the report progress through the whole process and encouraged me to think creatively. Lammert was always available for discussions, both direct assistance and even via video chat and voice calls. He also gave valuable remarks and constructive inputs for my thesis report. Overall, it was a great experience for me to be supervised by Lammert Kooistra which also gave me new knowledge of precision agriculture and control chart theory. I would also like to thank Willy ten Haaf for keeping track of my study progress and reminding me to finish my study on time. The thanks also goes to Ms. Antoinette Stoffers for taking care of the presentation schedule and the organizational part of my thesis work.

Moreover, I would like to say thanks for all the support, love and prayers that my father gave to me until he passed away on September 3rd, 2014. My special thanks go to my mom Sri Astuti for her love and support. Both my parents have always been so supportive to my educational and professional careers. I would never be in the Netherlands without their support.

The most special thanks goes to my beloved boyfriend Bendik Svendsen who always supported me when I felt down and always gave me all the love and time he had. He also helped me with the Python programming and was never bored to proofread my report. I would never finish this research on time without his help and all the support from his parents Gro Andersen and Petter Svendsen.

Finally, I also want to thank my MGI friends, especially Marthalina Indhawati, Sukmo Pinuji, Siti Hamim Latifah and Nikolaos Tziolas, for their support during the completion of this thesis work.

Abstract

Tracking down the crop nutrient status over the growing season is the most principle step for monitoring crop health status in precision agriculture. As a measure of the crop response to nitrogen application, the nutrient status is also related to chlorophyll content and an indicator of photosynthetic activity. To detect the growth problem within the crop fields, the crop N status monitoring can be used as a solution. The main objective of this study was to derive an alerting service for potato growth development from crop monitoring time-series in the precision agriculture. A potato field located in the South of the Netherlands was used as a case study to answer the research objectives. The first step in this research was comparing the three different N status measurement methods over the temporal development. After that, the relationships between crop biophysical and biochemical parameters over the growing season were analysed. The crop biophysical indicator that is highly related to the potato N status was identified from the regression analysis between the eight well-known VIs (NDVI, EVI, WDVI, REP, MCARI/OSAVI, TCARI/OSAVI, CI_{green} and $CI_{red\text{-}edge}$) and the chlorophyll content (on the leaf and canopy level). The results showed that the chlorophyll ratio index TCARI/OSAVI has the strongest relationship with the leaf chlorophyll among the eight VIs, even though the coefficient of determination value was relatively low (R^2=0.517). With a high coefficient of determination value (R^2=0.858), $CI_{red\text{-}edge}$ has the strongest relationship with the chlorophyll canopy. Next, the TCARI/OSAVI and $CI_{red\text{-}edge}$ time series data over the growing season were used in the time series similarity measures. The time series similarity measures based on the distance measures (Manhattan distance, Euclidean distance and Root Mean Square distance) and the correlation measures were used to calculate the crop growth deviation from each experimental plot or subplot towards two selected reference plots: maximum yield and mean curve approach. The results showed that both Euclidean distance and RMSD were the best similarity measures in characterizing the growth status of each experimental plot and subplot over the growing season. However, the Euclidean distance was used in this research to derive the alerting service using the Control Chart theory. The alerting services were available in the 30x30 m experimental plot level and 13x30 m experimental subplot level. In the plot level, the alerting services were acquired using the two different reference plot approach. The results showed that the alerting services were able to give alerts to specific plot or subplot with considering the changes in the plot or subplot condition): changes from “in control” state (green) to alert state (yellow) or “out-of-control” state (red). The alerting services were validated using fused satellite and UAV imagery dataset (STRS dataset) and twelve subplots validation set from the Cropscan dataset. The validation results showed that the alerting services were working properly for both dataset.

***Keywords*: nutrient status, nitrogen, maximum yield approach, mean curve approach, time series similarity measures, remote sensing, control chart, alerting service.**

Table of Contents

List of Figures

AMSS	:	Angular Metric for Shape Similarity
APAR	:	Absorbed Photosynthetically Active Radiation
C	:	Carbon
CCA	:	Canonical Correlation Analysis
CI_{green}	:	Green Chlorophyll Index
$CI_{red\text{-}edge}$	:	Red Edge Chlorophyll Index
CIRAD	:	Centre de coopération Internationale en Recherche Agronomique pour le Développement
CL	:	Control Limit
CO_2	:	Carbon Dioxide
CTVI	:	Corrected Transformed Vegetation Index
D_{CC}	:	Coefficient Correlation Distance
D_E	:	Euclidean Distance
D_{Man}	:	Manhattan Distance
D_{Mink}	:	Minkowski Distance
DAP	:	Days after Planting
DSM	:	Digital Surface Model
DTW	:	Dynamic Time Wrapping
DVI	:	Difference Vegetation Index
EDR	:	Edit Distance on Real sequence
ERP	:	Edit distance with Real Penalty
EVI	:	Enhanced Vegetation Index
FOV	:	Field of View
GIS	:	Geographic Information Systems
GPS	:	Global Positioning System
GRS	:	Geo-information Science and Remote Sensing
IAASTD	:	International Assessment of Agricultural Knowledge, Science and Technology for Development
INRA	:	French National Institute for Agricultural Research
KBDI	:	Keetch-Byram Drought Index
LAI	:	Leaf Area Index
LCL	:	Lower Control Limit
LCSS	:	Longest Common Sub Sequences
LPM	:	Local Pattern Match
LWL	:	Lower Warning Limit
MCARI	:	Modified Chlorophyll Absorption Ratio Index
MODIS	:	Moderate Resolution Imaging Spectroradiometer

MIR : Mid Infrared
MSAVI : Modified Soil-adjusted Vegetation Index
N : Nitrogen
NDVI : Normalized Difference Vegetation Index
NDWI : Normalized Difference Water Index
NIR : Near Infrared
NO_x : Nitrogen Oxides
NO_3-N : Nitrogen Nitrate
NPCI : Normalized Pigments Chlorophyll Ratio Index
NRVI : Normalized Ratio Vegetation Index
NSE : Nitrate Specific Electrode
OSAVI : Optimized Soil-Adjusted Vegetation Index
OWG : Onder Water Gewicht (Under Water Weight)
PAR : Photosynthetically Active Radiation
PSNC : Petiole Sap Nitrate Concentration
PSRI : Plant Senescence Reflectance Index
PVI : Perpendicular Vegetation Index
R^2 : Coefficient of determination
REIP : Red-Edge Inflection Point
REP : Red-Edge Position
RS : Remote Sensing
RMSD : Root Mean Square Distance
RVI : Ratio Vegetation Index
SAVI : Soil-Adjusted Vegetation Index
SPAD : Soil Plant Analysis Development
SpADe : Spatial Assembling Distance
SSCM : Site-Specific Crop Management
STARFM : Spatial and Temporal Adaptive Reflectance Fusion Model
SVD : Singular Vector Decomposition
TCARI : Transformed Chlorophyll Absorption in Reflectance Index
TQuEST : Threshold Query Based Similarity
TSAVI : Transformed Soil Adjusted Vegetation Index
TTVI : Thiam's Transformed Vegetation Index
TVI : Transformed Vegetation Index
UAV : Unmanned Aerial Vehicles
UCL : Upper Control Limit
UWL : Upper Warning Limit
VI : Vegetation Index
VRT : Variable Rate Technology
WDVI : Weighted Difference Vegetation Index

1.1 Context and background

In the coming decades, global agriculture must simultaneously produce more food to feed a growing population whilst adapting to climate change, an increasing threat to agricultural yields (IAASTD, 2009; INRA/CIRAD, 2011; Lobell et al., 2011; The Hague Conference, 2010; Foresight, 2011). This scenario calls for the introduction of modern technologies to improve crop yield, provide information to enable better in-field management decisions, reduce chemical and fertilizer costs through more efficient application, permit more accurate farm records, increase profit margin and reduce pollution. In other words, there is a need for farming with precision to optimize inputs and outputs (Seelan et al., 2003). Precision agriculture, with its ability as a key of sustainable intensification method, aims to maximize the agricultural production in a sustainable manner (The Royal Society, 2009).

Remote sensing technology that allows non-destructive acquisition of information about the Earth's surface can facilitate the implementation of precision agriculture (Liaghat & Balasundram, 2010). During the last two decades, development in remote sensing data acquisition capabilities, data processing and interpretation of ground based, airborne and satellite observations have made it possible to couple remote sensing technologies and precision crop management systems (Waheed et al., 2006). With the ability to identify variation in biophysical parameters, such as canopy nitrogen content and plant biomass (Clevers & Kooistra, 2012), remote sensing has a key role in agricultural monitoring (Jones & Vaughan, 2010).

Farmers throughout the world are constantly searching for ways to maximize their returns and precision agriculture has potential to support this (Seelan et al., 2003). Along with the development of precision agriculture, the use of global navigation satellite systems, remote sensing and near-sensing instruments on tractors, as well as in situ wireless sensor networks, provide the modern farmer with a wealth of data (Kooistra et al., 2012). Through these methods, the density of the time-series data has increased. Even though the time-series data from remote sensing are highly available, the knowledge of making benefit of the data to optimize the management activities and natural resources in the precision agriculture has not been used.

The current issue faced by most farmers is to detect which areas of their fields that have a growing problem and need immediate action. Combining the field knowledge of the farmers and the available remote sensing data will, hopefully, be able to solve the problem.

1.2 Problem definition

Remote sensing has been widely used in the agricultural sector for many years. The ability of remote sensing to produce repeatable measurements from a field, without destructive

sampling of the crop, has made remote sensing methods valuable for precision agriculture applications (Hatfield & Prueger, 2010). The availability of consistent time-series sensor data, with a high spatial and temporal resolution to detect anomalies in crop development, is a critical user requirement for the application of remote sensing in precision agriculture (Gebbers & Adamchuk, 2010; Hatfield & Prueger, 2010). The prodigious innovation in the spatial, spectral and temporal resolution has contrived tremendous data availability that can be used to monitor crop growth and real-time detection.

Of particular significance to precision agriculture is the nutrient status, a measure of the crop response to nitrogen application, which is related to chlorophyll content and which can be as indicator of photosynthetic activity (Haboudane et al., 2002). Currently, there are different spectral vegetation indices to measure the nitrogen status in crops. Spectral Vegetation Indices (VIs) calculated as linear combinations of near infrared (NIR) and VIS red reflectance have been found to be well correlated with canopy cover, leaf area index (LAI) and absorbed photosynthetically active radiation (APAR) (Broge & Mortensen, 2012; Elvidge & Chen, 1995; Myneni & Williams, 1994). Indices such as the ratio vegetation index and normalized difference vegetation index (NDVI) perform exceptionally well for tracking green biomass or leaf area index through the season, or for detecting uneven patterns of growth within a field (Jackson & Huete, 1991; Wiegand et al., 1991). The most representative VI for nitrogen status or chlorophyll content of crops will be able to provide the best reflection of the crop growth condition.

VIs can be derived from satellite-mounted sensors or sensors mounted on tractors (Yara N-sensor, Crop Circle, Cropscan (by hand), and Greenseeker) (Duisterwinkel, 2013). Since vegetation indices have the ability to identify variation in crop properties, the availability of VI time-series data is increasing. Several studies (Sakamoto et al., 2005; Xin et al., 2002; Zhang et al., 2003) have used VI time series data from satellites to monitor global and local crops phenology. However, there is still a lack of scientific knowledge to use time-series analysis techniques to optimize the management activities and use of natural resources in precision agriculture.

Recently, a new method has been developed to investigate abrupt changes in the satellite image time series data for forested areas (Verbesselt et al., 2010). The method iteratively estimates the time and number of abrupt changes within time series, and characterizes change by its magnitude and direction. Time series similarity offers its ability to derive statistical inferences about the relationship between time series of different data sets (Bretherton et al., 1992 and Tippett et al., 2008). There are several time-series similarity methods for change detection explained by Lhermitte et al. (2011). Current research (Kooistra et al., 2012) has adopted the Euclidean distance to calculate deviating nitrogen conditions between two different points in time and detect a potential need for management action. This research gives an idea of how to use time series techniques to detect crop status changes for a local scale application in precision agriculture.

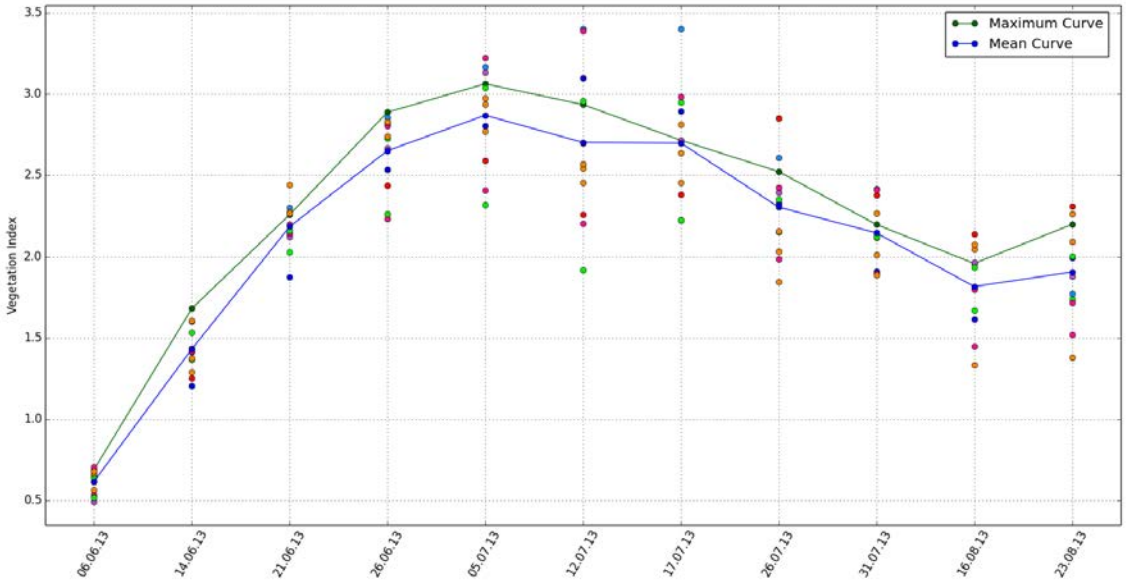


Figure 1: Potato Growth Mean and Maximum Curve Simulation (Simulated Values)

As for detecting crop growth anomalies using the time series technique, one should be able to compare the time-series spectrum with the reference spectrum of the healthy crop. A representative VI can be used as a time-series dataset to produce a crop growth reference curve. There are different ways to determine the time-series reference spectrum. First, deciding the reference curve by selecting the area that produces the maximum yield at the end of the growing season and using its VI time-series data for the whole growing season as a reference (Figure 1). Second, calculating the VI mean value for each observation date and deriving a mean curve crop growth as reference curve. The second method can be used for near real time alerting services. On the other hand, the first method is still hard to use since the curve is only available at the end of the growing season, but can be used as a long term solution.

In summary, after deciding the reference curve, the deviation of the crop growth can be detected. Whether the crop growth needs management attention or not will depend on the threshold values. This study will derive an alerting service to detect changes in the crop fields. The changes detection will be derived from the established reference curve deviation using time series analysis. Finally, an alerting service from crop monitoring time-series in precision agriculture will be created.

1.3 Research objectives and research questions

The main objective of this research is to derive an alerting service from crop monitoring time-series in the precision agriculture. This involves determining the VI that is most sensitive to the crop health status as well as the time series technique that can be used to detect changes in the crop compared to the reference curve. Finally, to be able to detect the changes, the thresholds of the alerting service from the reference curve have to be

established. The potato field, which is located in the South of the Netherlands, will be used as a case study to answer the research objectives.

To achieve these objectives, the following research questions will be answered:

Q1: Which VIs provide a good representation of the N status in the potato crop?

Q2: How can the optimal growth curve for the potatoes - the reference curve - which is represented by VI time-series, be determined?

Q3: Which time-series analysis method can be adopted to evaluate deviation from the established reference curve to characterize the growth status?

Q4: Which thresholds are relevant to detect deviating plant growth and which can be used to derive the alerting service?

Q5: How can the accuracy of the potato growth alerting service be validated?

1.4 Outline of the report

This thesis report consists of six chapters. In the second chapter a complete literature review about main topics of the research are conferred: an introduction of precision agriculture including remote sensing application in precision agriculture and comprehensive explanations about: nitrogen status assessment methods, similarity measures for time series analysis, and statistical control chart. The third chapter mainly explains about the materials and methods involved in the research which starts with the detail of the study area; the data used; and data analysis methods to answer each research question in Section 1.3. The achieved results from each proposed method are gathered in the fourth chapter. While in the fifth chapter, the results are mainly discussed in broader context and linked to the scientific literature. Finally, the last chapter elucidates overall recommendations and remarks for future research.

This chapter presents a complete literature review on the main topics of the research. The main topics covered are: the introduction of precision agriculture (Section 2.1); crop nitrogen status assessment methods in the precision agriculture (Section 2.2); remote sensing based VIs adopted for precision agriculture (Section 2.3); time series similarity measures (Section 2.4); and, the control chart theory (Section 2.5).

2.1 Introduction of Precision Agriculture

Sustainability was originally used to refer the agricultural and industrial technologies that are able to prevent or reduce environmental degradation and are often associated with economic activity (Bongiovanni & Lowenberg-DeBoer, 2004). Adding the sustainability term to the concept of agricultural and food systems can be traced to environmental concerns that started to appear in the 1950s – 1960s (Pretty, 2008). The concerns mainly related to the fact that world population continues to increase which, as a result, leads to an increase in the absolute demand for food. On the other hand, the increase of agricultural production leads to significant amounts of pollutants in the environment brought by fertilizers and pesticides during the agricultural process. This created a growing pressure for farmers to keep producing more foods whilst keeping the right amounts of fertilizer and pesticides in their crop fields. Site-specific crop management (SSCM) with its ability to manage the farm fields' variability can be a smart solution. SSCM is the idea of doing the right thing, at the right place and at the right time. This developing system for agricultural management is incorporating variable-rate technology, which is also known as one of the choices within Precision Agriculture. Precision agriculture uses intensive data and information collection, processing in time and space to make more efficient use of farm inputs, leading to improved crop production and environmental quality (Harmon et al., 2005). While farmers tend to treat their fields uniformly in conventional agriculture, precision agriculture introduces the benefits of field micromanaging. With precision agriculture, the crop fields can be divided into management zones that each receives customized management inputs based on varying soil types, landscape position, and management history (Mulla, 2013).

From the mid-1970s to the early 1980s, better field investigation methods (including soil survey, soil sampling, aerial photography, and crop scouting) resulted in a better awareness of soil and crop condition variability within fields (Robert, 2012). The field investigation methods lead to an understanding of crop management by zones to increase the profit and protect the environmental condition. Precision agriculture introduced new technologies, such as global positioning system (GPS), geographic information systems (GIS) and microcomputers in the late 1990s. These technologies provide farmers the information collection about their fields. At that time, farmers increasingly realized how much farming data were spatially correlated (Blackmore, 2003). GPS is used to locate an exact point location within a field. The GPS working process is started with the radio signal

transmittance from twenty-four satellites. A GPS receiver requires at least four satellites to determine a point of location on earth, of which the fourth satellite provides error correction. An exact point is found later by determining the distance from three satellites.

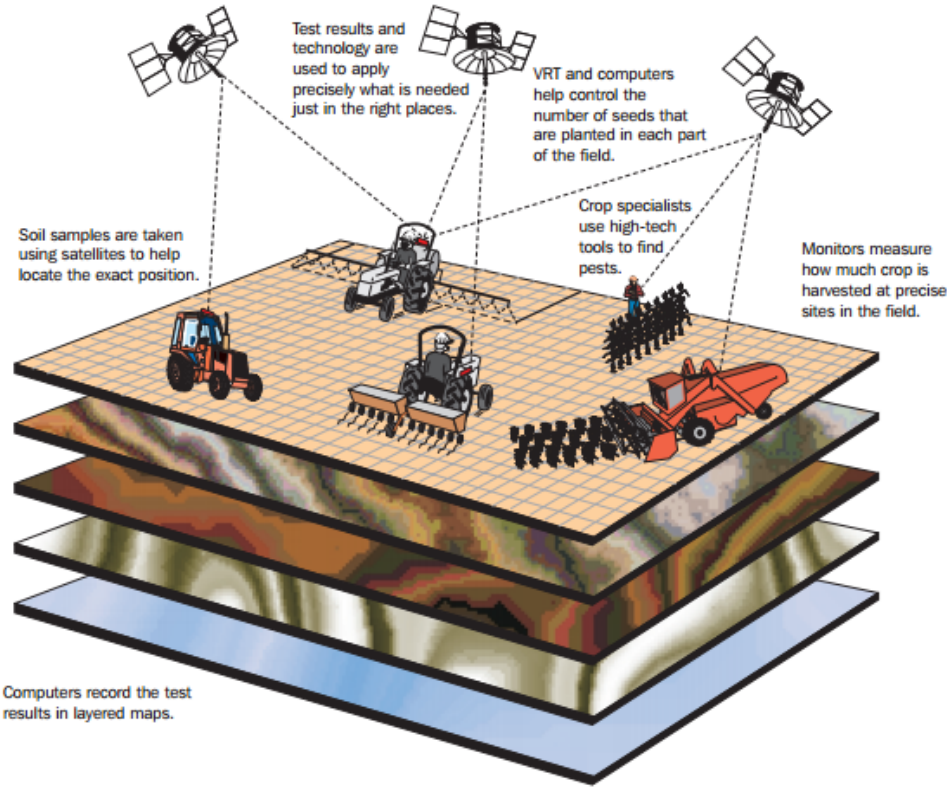


Figure 2: Precision Agriculture system (Picture copyright by CAERT, Inc.)

The other technology, GIS, refers to computer software that provides data storage, retrieval, and transformation of spatial data (Rains & Thomas, 2009). Field-attributes data, such as nutrient level, soil type, and fertilization level, will be stored as layers in the GIS software for precision agriculture (**Error! Reference source not found.**. The field location is normally stored by its latitude and longitude as determined using GPS. Using the combination a GPS and GIS, farmers are able to record the positions and variability of a field, store the data in the software, and analyse the characteristics between layers to develop application maps or apply variable rate technology (VRT) to vary the output of fertilizer and pesticides. Variable rate fertilizer application is using GPS to tell the operator the exact location of the tractor within a field. Then the GPS links with GIS to tell the controller about the field characteristics in that location. After getting information about the precise amount of fertilizer to apply at the location, the controllers then manipulates the machinery to apply the optimal level of fertilizer (Rains & Thomas, 2009).

Precision agriculture quickly generated strong interest, mainly because of associated new technologies, the concept made good sense, and it offered new routes for agro-industries and agri-businesses. Within the last decade, precision agriculture has been applied to a

variety of practices, crops and countries. From its first idea to manage variable rate applications of fertilizers, now the use of precision agriculture encompasses all management practices on a spatial and temporal basis (Robert, 2012). This is clearly evident by the development of several new technologies and manuscripts published in the area of precision agriculture in international journals and also by the variety of papers presented at the major international conferences on precision agriculture from different countries around the world (Bakhtiari & Hematian, 2013). New technologies involved in precision agriculture involve data collection and information management, as well as technological advances in field positioning, remote sensing, computer processing, yield monitoring and sensor design (Mulla & Schepers, 1997).

2.1.1 Remote Sensing in Precision Agriculture

The beneficial use of spatial imagery in agriculture for crop management has been known as early as 1929 when aerial photography was used to map soil resources (Seelan et al., 2003). The ability of remote sensing to derive information on the earth surface without making direct contact makes it beneficial for agricultural use. The introduction of satellites as a new technology for remote sensing acquisition methods offered farmers a greater result than the aerial photography method. In the beginning of its development, remote sensing had not made significant inroads to precision agriculture due to its long data acquiring time. Precision agriculture needs a frequent, or even near real time, information of crop condition throughout the growing season and often requires high spatial resolution. In the past, satellites were unable to fulfil the farmers' needs due to the long pre-processing time of satellite images, low availability of satellite images within a growing season, and an expensive cost of the high spatial satellite images. To solve the problems, efforts subsequently started the design of satellite imaging systems that had the higher spatial resolution and quicker revisit cycles required for precision agriculture (Mulla, 2013). IKONOS, a high spatial resolution satellite, was launched in 1999 to collect 4 m resolution imagery in 4 multispectral bands and 1 panchromatic band with a 5 days revisit cycle. In 2001, another high resolution satellite known as Quickbird was launched to produce 0.6 – 2.4 m image resolution in the blue, green, red, and near infrared with a revisit frequency of 1 – 3 days. Wu et al. (2007) made a comparison of Quickbird satellite imagery with petiole nitrate concentrations and SPAD chlorophyll readings to detect nitrogen status of potato canopies. These two satellites have steadily gained a substantial base of commercial subscribers interested in precision agriculture applications (Mulla, 2013). The development of satellite remote sensing for precision agriculture has been followed by the launch of three other satellites: RapidEye, GeoEye, and WorldView. With the improvement of spatial and spectral resolution of satellite images, farmers are now able to use the data for different kind of precision agriculture applications.

In addition to satellite-mounted sensors, low altitude sensors and sensors mounted on tractors also provide remote sensing data collection. Low-altitude remote sensing is

controlled from the ground to capture images at lower heights. Helicopters, airplanes and unmanned aerial vehicles (UAVs) can produce images with higher resolution and better details of soil and crop status, which are requirements for precision agriculture method. One advantage of the low-altitude remote sensing is their ability to provide real time basis of image and location information. This information stored in a mounted-microprocessor or mounted-storing device can be downloaded later and also can be carried out through internet (Swain et al., 2007). Apart from all described remote sensing methods above, there are also sensors that can be carried by a scout to the field and are able to spot check the health of plants and soil properties. These sensors use light reflectance on the leaf to determine chlorophyll levels, which is known to be related nitrogen levels in the plants (Rains & Thomas, 2009).

Remote sensing can be used for precision agriculture in a number of ways including by providing input on soil and plant condition and variability to the overall management and decision support system (Brisco et al., 1998). The integration of data coming from remote sensing platforms such as satellites, aircrafts, UAV, sensors mounted on the tractor and hand-held sensors, are able to provide farmers with the field-scale analysis. This benefit means remote sensing has an important role in the precision agriculture method. Although RS cannot capture all types of agricultural information, it can reliably provide accurate and timely information to guide agronomic and economic decision-making (Liaghat & Balasundram, 2010).

2.1.2 Remote Sensing for Vegetation

Keeping up with the information about nutrient status, water-stress condition and the possibility of insect attacks are the main concerns of agricultural operations. The conventional method using visual examination has several restrictions; human eye and aerial photography are unable to discriminate between healthy foliage and foliage in a stress condition. Modern remote sensing methods are able to overcome this problem by recording beneficial information related to the interaction of electromagnetic radiation with soil or plant material (Mulla, 2013). Using the knowledge of plant structure and how energy can be recorded throughout sensors are the essence of remote sensing methods.

Vegetation interacts with solar radiation in a different way compared to other materials. The absorbance, transmittance and reflectance of electromagnetic radiation depend on the plant foliage materials. The plant foliage materials, such as water content, carbon content, nitrogen content and, most importantly, photosynthetic pigments, of each plant are influencing the spectral characteristics as shown in Figure 3. The amount of radiation reflected from plants is inversely related to the radiation absorbed by plant pigments, and varies within the wavelength of incident radiation (Mulla, 2013). In the red and blue parts of the visible wavelengths, the reflectance is mainly influenced by the photosynthetic pigments. The chlorophyll pigment in green-leaf chloroplasts strongly absorbs radiation in

red and blue visible wavelengths, removes these colours from white lights, and leaves the predominant but diminished reflectance for visible wavelengths concentrated in the green (Lillesand et al., 2007). Other plant pigments, such as anthocyanin and carotenoids, are also important (Blackburn, 2007). In the near-infrared (NIR) plateau, reflectance is affected by multiple scattering of photons within the leaf, related to the internal structure, fraction of air spaces and air-water interfaces that refract light within leaves (Jacquemoud and Ustin, 2008). The reflectance in the mid-infrared (MIR) is primarily influenced by water content. The primary and secondary absorptions of water in leaf reflectance are greatest in spectral bands centred at 1450, 1940 and 2500 nm, with important secondary absorptions at 980 nm and 1240 nm (Carter, 1991).

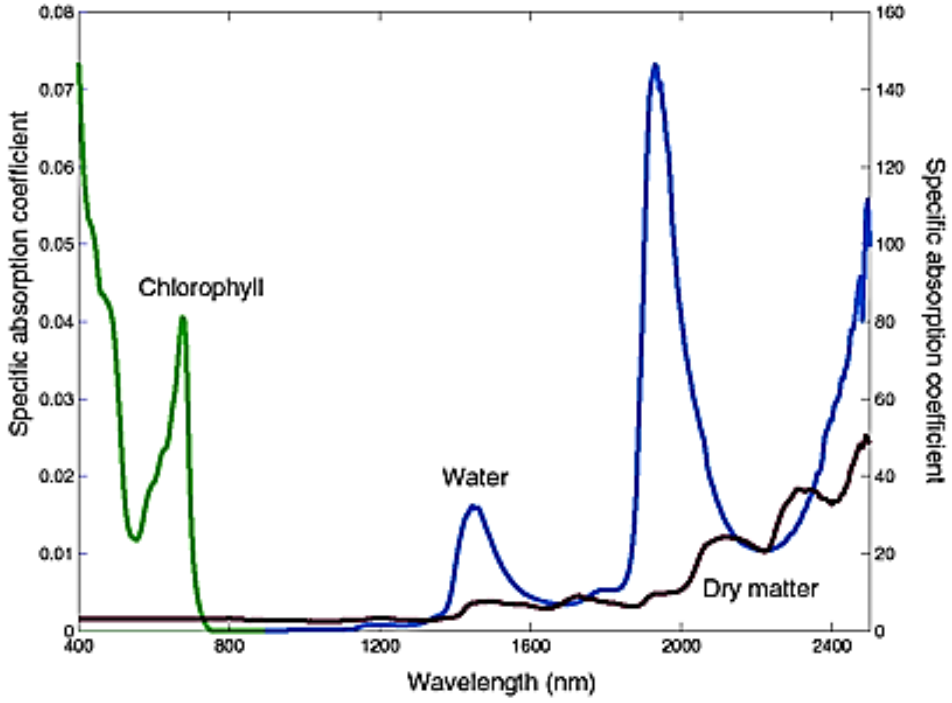


Figure 3: Specific absorption coefficient of (left scale) chlorophyll a+b ($cm^2 \mu g^{-1}$) and (right scale) water ($cm^2 g^{-1}$) and dry matter ($cm^2 g^{-1}$). Picture taken from Jacquemoud et al. (2000)

The high contrast in the reflectance behaviour between the red and NIR area (Figure 4: Typical spectral signature for vegetation (Picture adapted from Gaussman, 1977) can form vegetation indices that are able to monitor plant-health issues such as salt excessive, drought, nitrogen deficiencies and fungal pathogens. A high level of water related stress results in the noticeable changes of the photosynthetic pigments. The changes lead to yellow coloured leaves which provides an indication of chlorosis (reflectance of red wavelengths are equal to green). The phenomenon can be detected earlier using remote sensing, especially with the use of hyperspectral imaging. Besides the ability to detect the drought problems within the plants, hyperspectral and multispectral imaging are also able to detect the fungal pathogens which can cause serious yields losses. The changes of leaf pigments and moisture content within the plants can be used as an indication of pathogens detection needs.

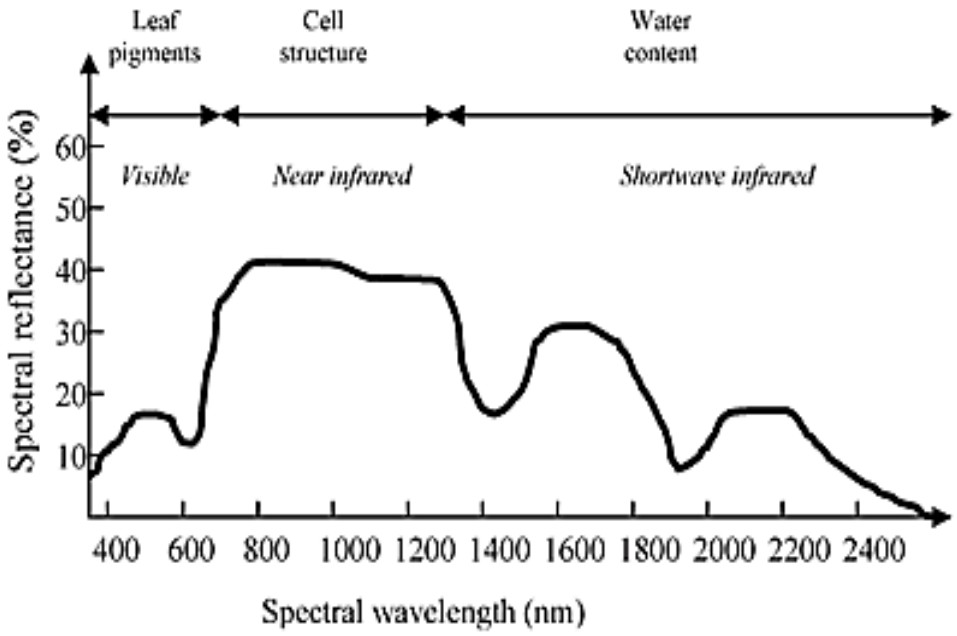


Figure 4: Typical spectral signature for vegetation (Picture adapted from Gaussman, 1977)

The successful nitrogen SSCM application is relying on the ability to identify areas in the field that have a differential response to nitrogen. The identification can be done using stem juices analysis or chlorophyll content estimation. The chlorophyll pigments in the leaf tissues are related to the concentration of leaf nitrogen (Thomas & Gausmann, 1977). Approximately 75% of the plant's total nitrogen is contained in the chloroplasts, mostly in Rubisco and chlorophyll binding proteins (Lawlor, 1993). Therefore, remote sensing offers spectral indices derived from the reflectance in the visible wavelengths and NIR to detect the nitrogen status and nitrogen deficiencies detection within the plants.

2.2 Nitrogen status assessment methods in precision agriculture

Besides hydrogen, carbon and oxygen, nitrogen (N) is required by plants in the largest quantity and most frequently becomes the limiting factor in crop productivity. For potato crops, proper rate and timing of N application are critical factors in optimizing potato tuber yield and quality, and minimizing environmental pollution (Wu et al., 2007). N deficiency can substantially reduce yield, whereas excessive N application can delay tuber maturity, lower tuber quality and increase the chance of nitrate contamination of surface and groundwater (Errebhi et al., 1998 and Zvomuya et al., 2003). To solve this dilemma, an adequate assessment of nitrogen status and its variability in the crop field have to be determined. At this time, there are various methods available to obtain timely assessment of the crop N status. Crop N status assessment can be carried out at the tissue, leaf and canopy level. Methods based on tissue analysis, such as Kjeldahl-digestion and Dumas-combustion (Muñoz-Huerta et al., 2013), have been widely applied to plants due to their reliability in organic nitrogen determination, but they are time-consuming and destructive. To minimize the drawbacks from Kjeldahl-digestion and Dumas-combustion methods, another method

such as Petiole Sap Nitrate Concentration (PSNC) test is now mainly used to assess the crop N status in the plant tissue measurement (Goffart et al, 2008). N assessment at the leaf or plant level involves the tools that are normally used to measure leaf chlorophyll content, which is highly related to plant N status (*i.e.*, Minolta SPAD-502, Dualex, and Chlorophyll fluorescence). In this report, the most common commercial handheld chlorophyll meters - Minolta SPAD - will be used to assess the crop N status for leaf level measurement. At the canopy scale, N status assessment relies on remote sensing methodology (based on spectral canopy characteristics) which aims to estimate canopy structure parameters such as LAI (Goffart et al, 2008). The most frequent tools for ground-based remote sensing are hand-held radiometer Cropscan, Yara N-Sensor and GreenSeeker, and QuickBird for satellite-mounted sensors. An N status assessment at the canopy scale will not be carried out in this report.

2.2.1 Kjeldahl Digestion and Dumas Combustion

One of the methods which have been widely used to determine nitrogen in plant tissue is the Kjeldahl digestion. This method was proposed by Johan Kjeldahl in 1883 (Kjeldahl, 1883). The procedure can be divided into three major steps. The first step is called wet digestion; the collected sample is mixed with sulphuric acid in a Kjeldahl flask then heated until it clarifies as CO_2 evolves. The second process's (distillation) main purpose is turning ammonium sulphate solution from the first process into ammonia. Before further processing, ammonia solution is heated to release the gas, which passes through a condenser and is then trapped in a receiving solution (boric acid, standard acid (HCl) or sulphuric acid) contained in a flask. T last step is ammonium estimation (Muñoz-Huerta et al, 2013). This method has some drawbacks, one of which is the inability to measure nitrate and nitrite. Kjeldahl Digestion is only able to measure organic nitrogen bound in the tri-negative state (Pontes et al., 2009).

To overcome the Kjeldahl method deficiencies, Jean-Baptiste Dumas proposed a combustion method for total nitrogen determination (Dumas, 1831). The Dumas method is being revived due to the availability of automated analytical instruments which determine C, H, N, and S on the same sample and O with a simple modification. The Dumas method normally involves an initial oxidation step, followed by the passage of the gases through a reduction furnace to reduce NO_x to N_2 (Carter, 1993). After that, carbon dioxide and gaseous nitrogen are separated and CO_2 is removed to measure only the gas nitrogen concentration. The final product of the combustion procedure is N_2 instead of ammonia, but an incomplete combustion causes nitrogen loss in the sample (Unkovich et al., 2008). In this report, the combination of wet-chemistry analysis and the Dumas combustion method are applied to get the N concentration percentage within the study area.

2.2.2 Petiole Sap Nitrate Concentration (PSNC)

Due to its efficiency in the measurement process, the petiole sap nitrate concentration (PSNC) test has become the most commonly used method to assess the N status in crops. The PSNC test is able to perform a quick diagnostic in monitoring nitrate level changes of a crop. The procedure starts with collecting the youngest mature leaves and petioles (the transition between the stem and the leaf blade) from the field. After the handling and storing process, the petioles are cut into half inch pieces, and then the sap is extracted to determine the nitrate content of the samples (Hochmuth, 2012). Two different types of instruments are reported in several studies - Nitrate Specific Electrode (NSE), and the combination of nitrate test strips and a hand-held reflectometer (Goffart et al., 2008 and Muñoz-Huerta et al, 2013).

In general, the nitrate test strips method turns the two-reactive zones into a red-violet combination when exposed to nitrate contained in the sample. A handheld-reflectometer Nitrachek (Montemurro, 2010) is produced to measure nitrate test colour changes. The reflectometer measures the amount of light reflected from the test strip pad, converts this reflectance to concentration, and displays solution concentrations in $mgL^{-1}NO3^{-}$ (Jemison and Fox, 1988). Due to the high value of tissue nitrate-N concentrations of many crops during early growing season, a sample dilution is required for the direct measurement using reflectometer (Williams & Maier, 1990). One study showed an excellent relation between the Nitrachek readings and standard concentrations of nitrate solutions (MacKerron et al., 1995 and Goffart et al., 2008). The PSNC values from the NSE instrument are significantly correlated to the dry petiole nitrate concentration values which are obtained from the same sampled material, although appreciable scatter of the data appears around the best-fit regression (Westcott et al., 1993 and Zhang et al., 1996). The scatter of the data are caused by a meter limitation and the use of dry petioles.

Goffart et al. (2008) reported that PSNC has been shown to be very responsive to various rates of nitrogen supply, either in soil or from fertilizer application. However, nitrogen can be stored in the other parts of the plant, as reduced N in leaf lamina or as nitrate in stems and lower leaves, which makes it immeasurable by the PSNC test (MacKerron et al., 1995). This shows one of the limitations using PSNC measurement to assess plant N status. PNSC changes over the growing season does not only depend on the N factor within the crop field. Weather conditions (rain periods) and soil nitrogen supply availability are responsible for temporary peaks in PSNC values, probably reflecting an increase in the supply of nitrate to the roots (MacKerron et al., 1995). Based on the same N-uptake values observed for different cultivars during the growing season, significantly different PSNC values can be observed (Laurent & Lancelot, 1999). Cultural practices such as mode of application, irrigation and types of N fertilizer can significantly influence the PSNC values. Therefore, crop management applications have to be considered while giving an analysis from the PSNC test.

Even though the PSNC test are considered as the quickest method among the other invasive methods to assess the crop N status, it still remains time consuming compared to methods using chlorophyll meters or a radiometer (Goffart et al., 2008). Several drawbacks such as the inaccurate results compared to an analytical laboratory method and inconsistent readings due to extreme environments should be considered while using the PSNC test.

2.2.3 Chlorophyll Meters

Chlorophyll is the most important biomolecule for photosynthesis process, responsible for allowing plants to absorb energy from the light. Leaf chlorophyll content can be used as a nitrogen status indicator because this is an essential element in photosynthetic protein synthesis (Taiz & Zeiger, 2010 and Demotes-Mainard et al., 2008). Furthermore, the deficiency of nitrogen usually leads to green colour loss in the leaves and a decrease in the leaf area and photosynthesis process (Bojović & Marković, 2009). Based on this fact, leaf chlorophyll content measurement tools (i.e. Minolta SPAD-502, Dualex, and Chlorophyll fluorescence) have been used to assess N status in the crops. These tools are part of the N status assessment at the leaf level, which are normally known as chlorophyll meters. There are two similar devices working with the same wavelengths: Soil Plant Analysis Development (SPAD)-502 chlorophyll meter (Minolta, Japan) and Hydro N-tester or HNt (Yara, Oslo, Norway). Dualex is a newer device that focuses on polyphenolic compound content measurement in the leaves using two wavelengths (375 nm and 650 nm), whereas the SPAD-502 and N-tester use two light sources on 650 nm and 940 nm. Both SPAD-502 and N-tester (Figure 5) provide a leaf greenness measurement, which is highly correlated to leaf chlorophyll and nitrogen content. The two instruments have a different sampling method: SPAD has individual measurement, while the N-tester gives a mean value of 30 readings (Gianquinto et al., 2011). Since the leaf analysis in this study has been done using SPAD-502, emphasis is specified to the SPAD-502 chlorophyll meter.

The SPAD-502 determines the relative chlorophyll content by measuring the absorbance of the leaf between two wavelength regions. By taking advantage of the chlorophyll characteristic, the SPAD-502 utilizes two light–emitting diodes at 650 nm (peak chlorophyll absorption area) and 940 nm (non-chlorophyll absorption area), and a photodiode detector to sequentially measure light transmission through leaves from red and infrared light (Markwell et al., 1995). The light transmitted by a leaf is then converted into electrical signals, amplified, and converted into digital signals (Gianquinto, 2004). The SPAD-502 chlorophyll index digital display value ranges from 0 to 50, while it ranges from 0 to 800 for N-tester measurement (Goffart et al., 2008). The values measured by SPAD-502 are proven to give a good estimation of the amount of chlorophyll present in the potato plant leaf (Vos & Bom, 1993). Uddling et al. (2007) have determined the non-linear relationship between chlorophyll concentration and SPAD values for potato crops with a relatively-good correlation value (r^2=0.46).

Figure 5: SPAD-502 chlorophyll meter and N-tester
(Photos taken from The Garden Professors™ blog and Agri Con website respectively)

The PSNC test determines the concentration of nitrate as the inorganic N reserve (Westcott et al., 1993), whereas SPAD meter estimates the N status based on its correlation with the chlorophyll content (Martin, 1995 and Gerendás & Pieper, 2001). Compared with the PSNC response to nitrogen, the time-course of values from the SPAD and N-tester reacts poorly to increasing N fertilizer rates applied at planting time (Gianquinto et al., 2004). This is related to the deficiencies of the optical sensing method which does not allow over fertilized plant detection. Due to its low sensitivity in the potato crop case, the SPAD or N-tester can only be used to compare low N-uptake crops (non-fertilized plot) and high N-uptake crops (fertilized plot). The method can then only detect deficiency situations for N (Goffart et al., 2008). In potato crops, the SPAD-502 chlorophyll meter does not respond as rapid as PSNC. Based on the study conducted by Wu et al. (2007), the SPAD-502 was able to detect severe N deficiencies about one month after emergence, while the PSNC test responded within two weeks. Apart from the proneness, chlorophyll meters are relatively easy to use and provide rapid on-site indications of the nutritional status of the crops (Gianquinto et al., 2011). However, Martínez and Guiamet (2004) showed that time of measurement, irradiance and plant water status must be considered when using the SPAD chlorophyll meter.

2.3 Vegetation Indices

Applying variable rate nitrogen fertilization is seen as one of the objectives in precision agriculture applications. Due to the high influence of nitrogen supply to the yield production, farmers and agricultural managers are interested in crop status monitoring within specific critical points in time: first in earlier growth stages, in order to supply adequate fertilizers quantities for a normal growth of the crop; and second, during an advanced development stage for health monitoring and the prediction of yield (Haboudane et al., 2002). This purpose can mostly be fulfilled by remote sensing methods which are able

to provide time-specific information about the conditions of crop parameters. Baret et al. (2007) has proven the ability of remote sensing observations to quantify crop stress level using a variable derived from the leaf area index (LAI) and $Chlorophyll_{ab}$ multiplication as the representative of the nitrogen content in the canopy level. Nitrogen concentration in green vegetation is related to chlorophyll content, and therefore indirectly to one of the basic plant physiological processes: photosynthesis (Haboudane et al., 2002). Nitrogen shortage within the crops will lead to lower chlorophyll content, which will increase the reflectance in visible wavelengths. Based on this condition, one way to assess nitrogen variability is by estimating crop canopy chlorophyll concentration from crop leaves and canopy reflectance (Daughtry et al., 2000).

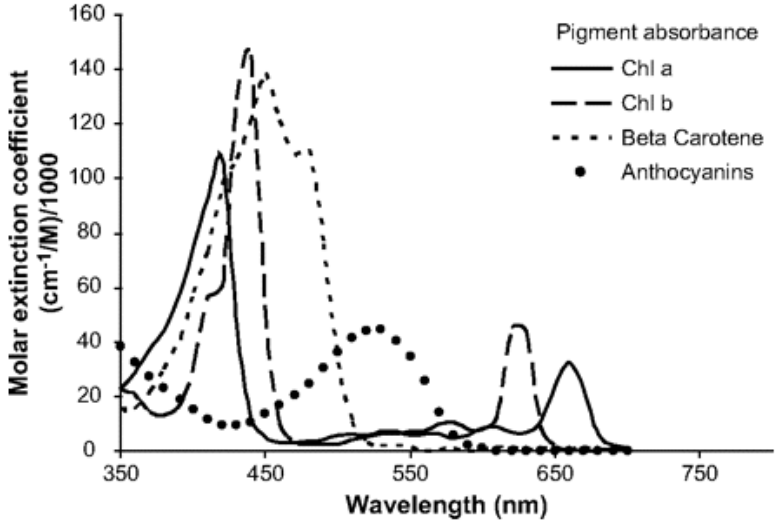


Figure 6: Absorption spectra of the major plant pigments (Blackburn, 2007).

Extracting spectral information related to N uptake using various VIs has been developed to enhance the capability of detecting canopy N variations (Blackmer et al., 1996). A wide variety of chlorophyll indices have been established to maximize the relationship between the measured spectral response and plant N status. Many of these indices employ narrow band (hyperspectral) data that include bands within the red edge region (Daughtry et al. 2000, Haboudane et al. 2002). Clevers and Kooistra (2012) investigated the use of cropscan datasets to find the best representative VI for estimating canopy chlorophyll content in a potato crop. The research involved six different VIs to investigate the best VI for monitoring chlorophyll and nitrogen content within the crop field.

Increasing efforts have focused on understanding the relationships between vegetation optical properties and photosynthetic pigments concentrations within green leave tissues such as chlorophylls and carotenoids (Wu et al., 2008). From the optical point of view, photosynthesis pigments have different spectral behaviour, with specific absorption features at different wavelengths (Figure 6Figure 6: Absorption spectra of the major plant pigments (Blackburn, 2007).), which allow remote sensing techniques to discriminate their respective effects on vegetation reflectance spectra (Chappelle et al., 1992; Blackburn,

1998; and Haboudane et al., 2002). These characteristics have supported the development of various approaches based on model inversion or the use of semi-empirical and empirical methods to estimate the chlorophyll content at the leaf and canopy level (Daughtry et al., 2000).

Table 1: Vegetation Indices for this thesis report

Index	Formulation	Source
NDVI	$\frac{NIR - R}{NIR + R}$	(Tucker, 1979)
EVI	$2.5 * \frac{NIR - R}{1 + NIR + 6R - 7.5B}$	(Kerr and Ostrovsky, 2003)
WDVI[1]	$NIR - C * R$	(Clevers, 1989)
REP	$700 + 40\ \frac{(R_{670} + R_{780})/2 - R_{700}}{R_{740} - R_{700}}$	(Guyot and Baret, 1988)
MCARI/OSAVI[2]	$\frac{[(R_{700} - R_{670}) - 0.2(R_{700} - R_{550})](R_{700}/R_{670})}{(1 + 0.16)(R_{800} - R_{670})/(R_{800} + R_{670} + 0.16)}$	(Daughtry et al., 2000)
TCARI/OSAVI[2]	$\frac{3[(R_{700} - R_{670}) - 0.2(R_{700} - R_{550})(R_{700}/R_{670})]}{(1 + 0.16)(R_{800} - R_{670})/(R_{800} + R_{670} + 0.16)}$	(Haboudane et al., 2002)
$CI_{red\text{-}edge}$	$(R_{780}/R_{710}) - 1$	(Gitelson and Merzlyak, 2003) (Gitelson et al., 2006)
CI_{green}	$(R_{780}/R_{550}) - 1$	(Gitelson and Merzlyak, 2003) (Gitelson et al., 2006)

[1] C = 2 for the current study area (Kooistra et al. 2013). [2] Wavelength at 800 will be replaced by 780 nm

VIs are combinations of surface reflectance at two or more wavelengths built to emphasize a specific vegetation property. The development of different spectral band combinations, known as VIs, has a main purpose of decreasing the spectral effects caused by external factors such as atmosphere and soil background. As one of the empirical methods, VIs can be classified into two groups: *slope-based* and *distance-based* VIs (Jackson & Huete, 1991). Slope-based VIs use simple linear combinations that emphasize the difference in vegetation spectral reflectance in the red and NIR wavelengths. The slope-based VIs include the RATIO, NDVI, SAVI, RVI, NRVI, TVI, CTVI, TTVI and EVI (Silleos, et al., 2006). Different from slope-based VIs, the distance-based VIs require the establishment of a soil line. Pixels that are located near the soil line are assumed to represent the soil, and those far away are assumed to represent vegetation (Mróz & Sobieraj, 2004). To get the VIs' value, the calculations are made by measuring the perpendicular distance of each point compared to the soil line. The parent index from this group is the Perpendicular Vegetation Index, which was suggested by Richardson and Wiegand in 1977. The other indices that are part of distance-based VIs are

PVI1, PVI2, DVI, TSAVI, MSAVI and WDVI. In addition to those two VI groups, there are orthogonal VIs and Red Edge Inflection Point (REIP).

VIs have been highly developed for agricultural remote sensing studies and management purposes in the last 30 years. However, the effectiveness of these indices is dependent on crop phenological development and management practices (Hatfield & Prueger, 2010). After conducting research on four crops (corn, soybean, wheat and canola) over eight years, Hatfield and Prueger (2010) found six different VIs that are most useful depending on crop phenology and management practices: (a) simple ratios for biomass; (b) NDVI for intercepted photosynthetically active radiation (PAR); (c) SAVI for LAI at early growth stages; (d) EVI for LAI at later stages; (e) CI_{green} for leaf chlorophyll; (f) NPCI for chlorophyll during later stages; and, (g) PSRI to quantify plant senescence.

To answer the research questions in this report, the calculations of the selected VIs (Table 1) will be made. These VIs are known as the most commonly used indices - both in terms of multispectral and hyperspectral indices - to generate the nitrogen content and chlorophyll information within the crops. The reflectance values for VI calculation purposes were derived from Cropscan measurements. To calculate the TCARI/OSAVI and MCARI/OSAVI based on the given formula in Table 1, the wavelengths of 550, 670, 700 and 800 nm are needed. However, the CropScan MSR16R is not able to measure at 800 nm. Therefore, the defined wavelength on 800 nm will be replaced by the wavelength at 780 nm.

2.4 Similarity Measures in Time-Series Analysis

The number of satellite sensors and close sensing sensors that are able to provide data with high temporal resolution has increased in recent years. Most of these sensors are able to acquire daily information of terrestrial processes, which makes it possible to monitor the crop condition continuously and dynamically (Ji-hua & Bing-fang, 2008). VIs allow reliable spatial and temporal comparison of terrestrial photosynthetic activity and canopy structural variations (Huete et al., 2002). Therefore, VI time-series data are able to represent the true surface measurements which are beneficial for crop monitoring purposes both in large scale and small scale areas. The analysis of the VI time-series through time permits the extraction of appropriate metrics to describe vegetation dynamics, allowing better monitoring and understanding of the biophysical changes in the vegetation cover and phenology in different ecosystems (Bradley & Mustard, 2008). In terms of providing precision agriculture with the crop temporal information for management purposes, the analysis of VI time-series could be applied. Zarco-Tejada et al. (2005) have used time-series data from several VIs, calculated from airborne hyperspectral sensors, to understand within-field yield variability in cotton (*Gossypium hirsutum L.*) over an entire growing season. NDVI time-series have also been used for estimating the bimodal agriculture areas using Fourier analysis (Canisius et al., 2007).

To analyse the temporal changes within the crop development phase, there are different type of time-series analysis methods that can be applied to detect the changes within the time-series data. Lhermitte et al. (2011) explained that the possible methods range from a point time scale (e.g., post-classification analysis of single date images and interpretation of the results over time) to a bi-temporal (e.g., bi-temporal change detection) and continuous timescale (e.g., classification and change detection based on temporal trajectory analysis). Temporal similarity of trajectories is performed the same as similarity analysis on time series data sequences (Dodge et al., 2009). Based on Serrà and Arcos (2014), similarity measure is the most essential ingredient of time series clustering and classification systems. However, different studies have shown the various purposes of time series similarity measures. Tippet et al. (2008) have used time series similarity in the form of regression-based methods as an absolute criterion to derive statistical inferences between different time series datasets. For this first purpose, various methods have been applied in remote sensing studies to resolve the relation between remote sensing time series data with bio-physical and geophysical variables (Lhermitte et al., 2011). The methods can be reflected in different approaches which range from singular vector decomposition (SVD) and canonical correlation analysis (CCA), to more commonly employed methods based on regression and correlation analysis (Buermann et al., 2003; Verbesselt et al., 2006; Lhermitte et al., 2011). The cross-correlation analysis were applied to study the connection between a meteorological drought index KBDI and remote sensing index (NDWI) which are both known to be related to vegetation moisture dynamics (Verbesselt et al., 2006).

Second, time series similarity has been employed as a relative criterion to numerically characterize the relationship between time series, not to derive statistical inferences, but to provide a decision criterion to cluster/discriminate time series (Lhermitte et al., 2011). As an example of this time series similarity context, Huang and Siegert (2006) have used NDVI time series data to monitor the desertification processes in North China. The classification methods started with detecting the temporal behaviour pattern of each vegetation cover (NDVI peak locations in growing season or non-growing season), the similarity measures then supply the premise to cluster the satellite pixel time series data in homogeneous groups based on minimization of within group temporal similarity and maximization of between group temporal similarity (Lhermitte et al., 2011). As a comparison to this approach, the similarity measures can be used for change detection purposes. A study conducted by Linderman et al. (2005) showed the changes detected in overall vegetation activity and its timing across Sub-Saharan Africa between different years using MODIS data includes the examination of magnitude, extent and nature of changes in photosynthetic activity. This study showed the possibility of similarity measures discriminating changes in time-series.

Similarity measures can be generally categorized into four classes: lock-step measures; elastic measures; threshold-based measures; and, pattern-based measures (Ding et al., 2008). The most widely known measure for lock-step measures is Euclidean distance, which

is defined as the square root of the sum of the squared differences between corresponding data points in two time series data sets (Nakamura et al., 2013). Besides being parameter-free and easy to implement, the complexity of evaluating Euclidean distance and its variants from L_p-norms is linear (Ding et al., 2008). To solve different sampling rates problems in the lock-step measures, DISSIM distance was introduced by Frentzos et al. (2007). Dynamic time wrapping (DTW) is one of the examples from the elastic measures group (Figure 7). As a classic approach to compute similarity between two time series, DTW works by warping the time series in the temporal domain so that the accumulated cost of this alignment is minimal (Serrà & Arcos, 2014). Based on Nakamura et al. (2013), other measures that fall into the elastic measures category are longest common sub sequences (LCSS), edit distance with real penalty (ERP), edit distance on real sequence (EDR) and angular metric for shape similarity (AMSS). The Threshold query based similarity (TQuEST) distance was presented by Aßfalg et al. (2006). This measure starts with accepting a user-provided threshold indicated by τ and converts the sequence data to *threshold crossing*. The *threshold crossing* will be treated as points in two-dimensional space which is composed only of data points above the τ (Nakamura et al., 2013). The similarity output is then defined as the Minkowski sum (Ding et al., 2008). The last class is a pattern-based measure which is known as Spatial Assembling Distance (SpADe). Based on Chen et al. (2007), SpADE is able to measure distances between shape-based time series of which the calculation is based on the detection of the best combination of local pattern match (LPMs) by calculating the shortest path in matching matrix.

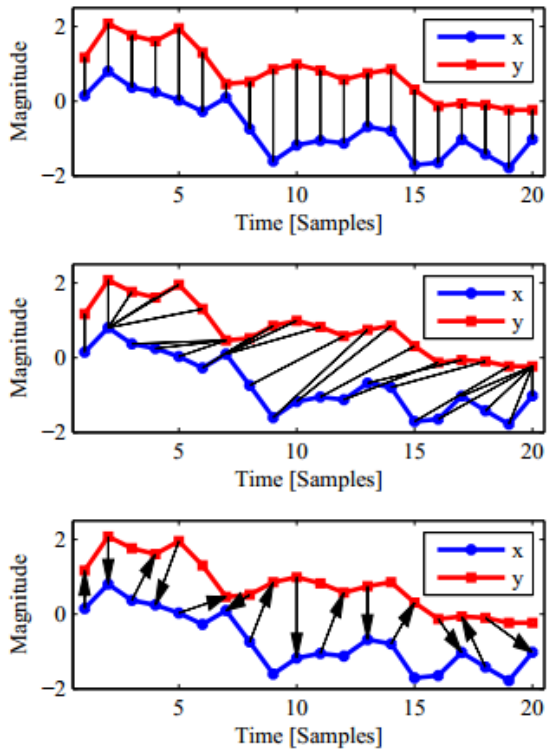


Figure 7: The illustration of Euclidean distance (top), DTW alignment (centre), and MJC (bottom). X axis indicates each time point or sample, while y axis shows magnitude (in this context, magnitude is analogy of the value from each sample). Picture is taken from Serrà and Arcos (2014).

The time series similarity measures can also be categorized based on the data characteristics. The measures can be divided into three major categories dependent on whether they work (i) directly with original time series data, (ii) indirectly with transformations taken out from the original time series data, or (iii) indirectly with metrics derived from the original time series data (Lhermitte et al., 2011). In this thesis report, the data came from the original time series data. Therefore, the original time series data approaches will be used to detect the area within the crop field that needs a management action. The most widely known approaches can be distinguished into: (i) distance measures and (ii) correlation measures.

Known as the most commonly used distance measures, Minkowski distance (D_{Mink}) is a generalization of both the Euclidean distance (D_E) and the Manhattan distance (D_{Man}). Based on Lhermitte et al. (2011), the Minkowski distance between two individual time series $f^p(t)$ and $f^q(t)$ collected in t time, for pixels p and q respectively, is given by:

$$D_{Mink} = \left(\sum_{t=1}^{N} \left| f_t^p - f_t^q \right|^r \right)^{\frac{1}{r}} \quad (1)$$

, where f_t^p is the $f^p(t)$ time series value at moment t and number of samples N in the time series while r indicates a user-defined integer. If the r value in Equation (1) is defined as 1 then this equation indicates the Manhattan distance D_{Man}, while the Euclidean distance D_E is indicated by r=2. The infinite form of D_{Mink}, which indicates by $r = \infty$ is also known as Chebyshev distance. For the relatively large training set size, Euclidean distance is known to be strongly competitive with more complex similarity measures (Ding et al., 2008). However, these distance measures are very sensitive to noise and requires the time series to have equal length.

The other methods that are part of the time series similarity approach are correlation measure and root mean square distance (RMSD). The original time series data can use these two approaches without making assumptions on interpolation or curve fitting (Lhermitte et al., 2010). The most commonly used correlation measure, described by Liao (2005), is Pearson's cross-correlation coefficient (D_{CC}). This cross-correlation coefficient is defined as the degree of linear relationship between time series (Lhermitte et al., 2011):

$$D_{CC} = \frac{\sum_{t=0}^{N-1} \left[\left(f_t^p - \bar{f}^p \right) * \left(f_{t-s}^q - \bar{f}^q \right) \right]}{\sqrt{\sum_{t=0}^{N-1} \left(f_t^p - \bar{f}^p \right)^2} * \sqrt{\sum_{t=0}^{N-1} \left(f_{t-s}^q - \bar{f}^q \right)^2}} \quad (2)$$

Where f_t^p and f_t^q are time series values at moment t and the mean values of the corresponding series are indicated by $\bar{f}^p$ and $\bar{f}^q$. *N* is the length of the time series, while s is specifying the delay (lag) between both time series. If s = 0, then the D_{CC} is computing the similarity between two time series without time shift (Lhermitte et al., 2010). The value of D_{CC} ranges from -1 to 1. When the D_{CC} shows a value that is close to 1, it indicates the increase in linear relationship and shows the opposite if the value is close to -1 (Lhermitte et al., 2011).

The RMSD values quantify the straight-line inter-point distance in a multi-temporal space (Lhermitte et al., 2010). The low values give representation of high temporal similarity. The RMSD measure defined as:

$$RMSD = \sqrt{\frac{\sum_{t=0}^{N-1}\left(f_t^p - f_t^q\right)^2}{N^2}} \quad (3)$$

Where f_t^p and f_t^q are time series values at moment t and N is the length of the time series. Both D_{CC} and RMSD are often used as similarity measures for remote sensing time series. Lhermitte et al. (2010) has applied these two similarity measures in the control pixel selection.

2.5 Control Chart

The basic idea of control charts was first developed by Shewhart (1931) for industrial product control purposes. A control chart is a plot which contains a time sequence with additional decision lines. The decision lines are used to determine whether a process is in control or not (Ryan, 2011). The control chart construction is based on the statistical principle. The statistical principle allows the balance between fast "out of control" occurrence detection with a minimum amount of *false alarms* (Ryan, 2011).

The concepts underlying the control chart are the natural variability in any manufacturing process that can be quantified with a set of control limits and the variation exceeding these limits signals a change in the process (SAS Institute Inc., 1999). This idea applies to more than just manufacturing process. Control charts were also used by Podur et al. (2002) to detect the significant changes in the mean and variance of the annual fire occurrences and burned areas in Canada. While Moameni and Zinck (1997) have also used statistical quality control charts to investigate variability in soil properties and control the mean of soil variables.

The principle of the control chart is shown in Figure 8. The central line (CL) on the control chart indicates the average (expected value) of the summary statistic when the process is in statistical control. The upper and lower control limit, shown as UCL and LCL, indicate the range of variation to be expected in the summary statistic when the process is in statistical control. The control limits are commonly computed as 3.09σ limits, representing three

standard errors of variation in the summary statistic above and below the central line. The orange colour dash-lines at ± 1,96σ are called warning lines or, relating to their position, the lower warning limit (LWL) and upper warning limit (UWL) (Massart et al., 1997).

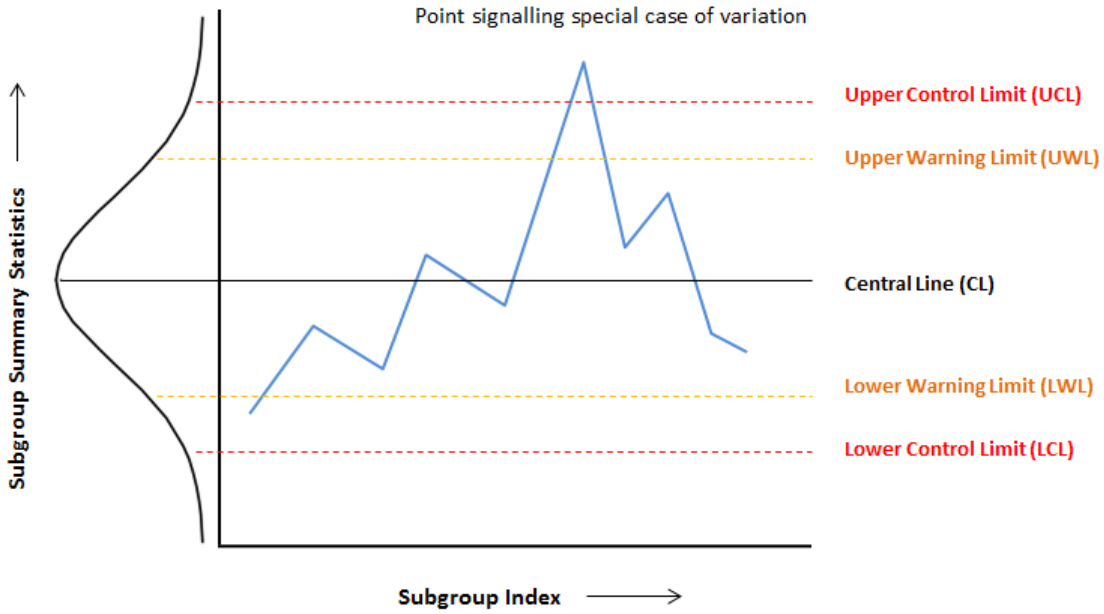


Figure 8: The control chart basic form

The bell-shaped curve at the left side of the vertical line in Figure 8 shows an ideal condition of normal distribution in the data (monitored process). As explained before, the CL supposedly signifies the value of mean, standard deviation or other statistics (Moameni & Zinck, 1997). The two small areas in the curve, which are located below the LCL and above UCL, indicate that the process is out of control. Therefore, the probabilities of these areas are required to be as low as possible. If the purpose of the chart is to control the process mean, μ, while the control limits were rounded to ±3σ, then the probability would be 0.00135 for each area (or 0.0027 for both) if X has a normal distribution (Moameni & Zinck, 1997). If the normal distribution of the process is unknown, then 3σ or 3.09σ limits are more appropriate to use (Moameni & Zinck, 1997).

In this report, control charts will be used to establish the alerting service threshold. The similarity measures values will be used as the input process of the control chart. Both control and warning limits will be decided based on the management needs, which later control whether the potato crop conditions are under control or not.

3.1 Study Area

To address the research questions, a case study of an agricultural field situated in the southern part of Netherlands will be included. The study area selected for this purpose is a potato field near the Dutch village of Reusel in the province of Noord-Brabant, located at 51°19' N and 5°10'14" E (Figure 9).

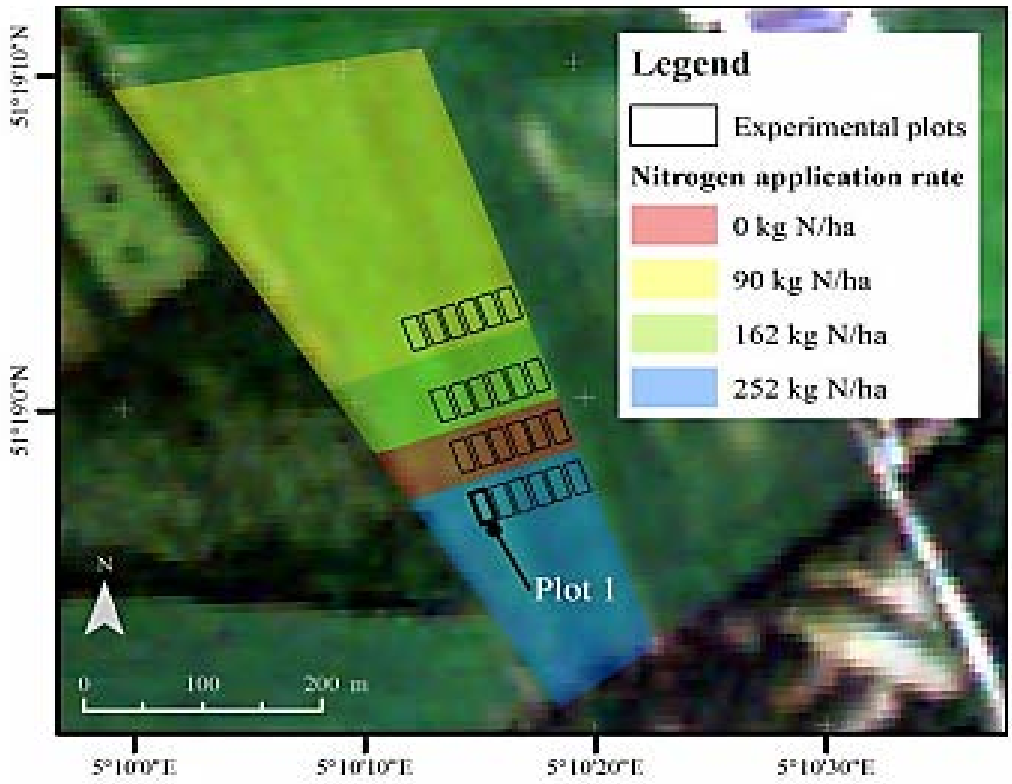


Figure 9: Location of the study area, setup of experimental subplots and Nitrogen application rates on the potato fields (Gevaert et al., 2015)

This field was the subject of a research project executed by the Wageningen University (WU) Laboratory of Geo-information Science and Remote Sensing (GRS) under the Smart Inspectors project. For the purpose of this study, four distinct nitrogen fertilization rates (0, 90, 162 and 252 kg N/ha) were applied to the field in the beginning of the 2013 growing season (Table 2). Additional fertilizer was also applied during the growing season based on the recommendations from sensors attached to the tractor. The variable rate technology was used to control the fertilizer level in the experimental fields. As the sensor attached on the tractor detected the nitrogen deficiencies, the fertilizer was applied automatically to the field. Another fertilization (N-vlb) applied to the plant leaves occurred as part of the management application to each experimental plot which summed up to 32 kg N ha^{-1} throughout the growing season.

Table 2: The initial, additional and Total N fertilization in kg per hectare (ha) applied to the experimental plots over the growing season

Plot	Organic Manure (N) + Liquid Fertilizer (NLV)	Fertilization over the growing season			Total N
		N2 (5 July 2013)	N3 (18 July 2013)	N-vlb	
A	252	-	42.7	32	326.7
B	0	29	47.5	32	108.5
C	162	-	50.2	32	244.2
D	90	-	44.4	32	166.4
E	252	-	-	32	284.0
F	0	-	-	32	32.0
G	162	-	-	32	194.0
H	90	-	-	32	122.0
I	252	-	45.3	32	329.3
J	0	22	42.4	32	96.4
K	162	-	60.2	32	254.2
L	90	-	53.2	32	175.2

Twelve 30 m x 30 m experimental plots (three per fertilization rates) were defined in the fields. For the purpose of this research, each experimental plot was divided in half to analyse the left and right parts of the tractor driving plots separately. The tractor driving path was not included in the experimental plots to avoid the disturbance in the surface reflectance. Therefore, the 24 13×30 m experimental subplots will be used in this study (Figure 10).

3.2 Available Data

For this study area, three datasets will be used: one obtained from cropscan, one from fused imagery and crop data which contain the N status (N%, plant sap and SPAD Chlorophyll meter data), LAI and chlorophyll data from the potato field.

3.2.1 Cropscan Dataset

The average spectral reflectance for the 24 experimental subplots were obtained on a weekly basis between June 6, 2013 and August 23, 2013 (Figure 10). Eleven spectral measurements were taken per experimental subplot using a Cropscan Multispectral Radiometer (Table 4). The cropscan dataset for the 24 subplots with 11 observations over the growing season, including a broad range of plant health parameters, will be used to define the most representative VI for potato crop health status. The spectral measurements from the 24 experimental subplots will be used to calculate the eight VIs: Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), Weighted Difference Vegetation Index (WDVI), Red-Edge Position (REP), the ration between MCARI and OSAVI (MCARI/OSAVI), the ration between TCARI and OSAVI (TCARI/OSAVI), Chlorophyll Index (CI) green and Chlorophyll Index (CI) red-edge. The VIs will be calculated based on the formulas presented in Table 1.

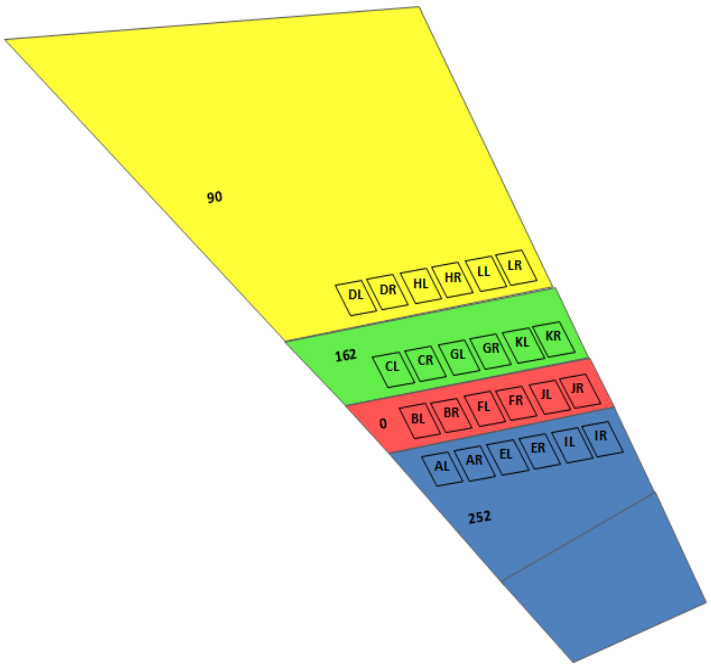


Figure 10: Twenty-four experimental subplots location

Table 3: Specifications of CS MSR16R

Spectral band position (nm)	Band width (nm)
490	7.3
530	8.5
550	9.2
570	9.7
670	11
700	12
710	12
740	13
750	13
780	11
870	12
940	13
950	13
1000	15
1050	15
1650	200

A Cropscan Multispectral Radiometer (MSR16R, Cropscan Inc.) has sixteen spectral bands, which measures simultaneously the reflected and incoming radiation in narrow spectral bands (Table 3). Reflectance is measured through a 28 degree field-of-view (FOV) aperture and incoming radiation is measured through a cosine-corrected sphere. Calibration is performed by pointing the 28 degree FOV aperture towards the sun using an opal glass. With this calibration, the spectral reflectance is derived (Clevers & Kooistra, 2012).

3.2.2 Fused satellite and UAV Imagery data

The UAV hyperspectral images were obtained with the Specim ImSpector V10 2/3" spectrograph. A GPS inertia navigation system (INS, XSens, and MTi-G-700) and a Panasonic GXI +14 mm camera obtained the geographical location and provided data for a Digital Surface Model (DSM) which the hyperspectral images were orthorectified. The system was mounted on an Aerialtronics Altura AT8 octocopter. This platform has a maximum payload of 2 kg and a flight-time of 5-8 minutes (Kooistra et al., 2013). Since the operational costs of the UAV are high, available imagery from the growing season are limited. UAV imagery are available from four dates: June 6th, June 14th, July 5th and July 17th.

The satellite data was generated from Formosat-2 images, with eight cloud-free scenes over the study area from March 1st to September 25th. However, the spatial and spectral resolutions of the Formosat-2 imagery are inadequate for precision agriculture applications. Therefore, the STR dataset, which is a simulated dataset from Formosat-2 imagery and UAV imagery, will be used in this research. This dataset is a result of previous research performed by Gevaert et al. (2015). The imagery from both platforms were combined in two ways.

Firstly, data fusion methods brought the spatial resolution of the Formosat-2 imagery (8 m) down to the spatial resolution of the UAV imagery (1 m). Two data fusion methods were applied: an unmixing-based algorithm and the Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM).

3.2.3 Crop Data

3.2.3.1 Crop N Status over growing season

The crop N status was measured over the growing season using three methods: the Dumas-combustion method, PSNC test and chlorophyll meter (Minolta SPAD-502). Both the Dumas-combustion method and the PSNC test are crop N status assessments at the plant tissue level, while measurement using SPAD-502 is carried out at the leaf level. Due to its destructive sampling and time consuming drawbacks, the nutrient status measurement using the Dumas-combustion method were completed on the plots on a monthly basis. As explained by Clevers and Kooistra (2012), the measurement started with sampling from the experimental plots and was followed by wet-chemistry analysis. The destructive sampling was taken by harvesting one row of potatoes over one linear meter each time (0.75 m^2). The next step was determining the vegetation dry weight and N concentration after going through a drying process for 24 hours at 70^{o}C. The latter process was based on the Dumas method (Hansen, 1989; Clevers & Kooistra, 2012). The end results of this process were fresh and dry aboveground weights, dry matter content, total N concentration in the aboveground parts and total N content. Within this report, the N concentration will be used to compare the N status in the crop with other assessment methods.

Table 4: Dumas-combustion, PSNC test, SPAD-502, Cropscan and LAI acquired dates in the experimental plots over 2013 growing season

Dumas-combustion	PSNC test	Chlorophyll meter (SPAD)	Cropscan	LAI
-	06 June 2013	06 June 2013	06 June 2013	06 June 2013
14 June 2013	14 June 2013	14 June 2013	14 June 2013	14 June 2013
-	21 June 2013	21 June 2013	21 June 2013	21 June 2013
-	26 June 2013	26 June 2013	26 June 2013	26 June 2013
-	05 July 2013	05 July 2013	05 July 2013	05 July 2013
-	12 July 2013	12 July 2013	12 July 2013	12 July 2013
-	17 July 2013	17 July 2013	17 July 2013	17 July 2013
26 July 2013	26 July 2013	26 July 2013	26 July 2013	26 July 2013
31 July 2013	31 July 2013	31 July 2013	31 July 2013	31 July 2013
-	-	16 August 2013	16 August 2013	16 August 2013
-	23 August 2013	23 August 2013	23 August 2013	23 August 2013

The PSNC test was easier to apply and less destructive compared to the Dumas-combustion method. In this report, the PSNC test was applied to the experimental plots ten times during the growing season (Table 4). The measurement values were averaged per plot. The PSNC

test gave the NO3 blade ppm value which will be used to compare the N status condition. The Dumas-combustion and PSNC test were only applied to eight plots (A, B, C, D, E, F, G and H). However, the chlorophyll meter measurement (SPAD-502) included all experimental plots (the measurement was acquired for both the left and right side). Besides measuring the N status in the potato crops, SPAD-502 was also used for calculating the chlorophyll (g/m^2) which will be explained later in this report.

3.2.3.2 Crop biophysical and biochemical variables

Crop biophysical and biochemical variables calculated in this report will be the Leaf Area Index (LAI) and leaf chlorophyll. The measurements were applied to all experimental plots using a handheld chlorophyll meter (Minolta SPAD-502) and Plant Canopy Analyser (LAI-2000) to calculate leaf chlorophyll and LAI, respectively. Between June 6th 2013 and August 23rd 2013, six SPAD leaf chlorophyll readings were produced per row (12 readings per plot) on a weekly basis. To get the SPAD value for each plot (total of 24 experimental subplots), the average value was generated from 12 readings. One of the purposes of this study is to find the best VI that is most representative for the N status in the crops. Since a direct relation between the VI and the N status cannot be acquired due to the limited N status data during the growing season, the leaf chlorophyll and chlorophyll canopy value will be used to find the best VI.

The amount of chlorophyll in the leaf level cannot be directly translated from the SPAD measurement values. The SPAD measurement values, which range from 0 to 50, are the relative amount of chlorophyll present in the leaves, and are shown as digital numbers without the unit of measurement. In order to translate SPAD readings to leaf chlorophyll concentration values, an exponential regression proposed by Uddling et al. (2007) can be used. In this report, the leaf chlorophyll concentration values will be attained from:

$$y = 0.913e^{0.0415\,x} \qquad (4)$$

Where y = chlorophyll concentration and x = SPAD-502 value. The chlorophyll canopy values in this report will be multiplication results from chlorophyll leaf concentration and LAI. The LAI measurement was acquired on the same day as the chlorophyll meter measurement. The LAI readings were taken from the third and tenth row of each plot and six LAI values were taken per row. Each reading per plot was the average value from 12 LAI readings. The total LAI readings over the growing season were 11 weekly measurements starting from 6 June until 23 August 2013 (Table 4).

Basically, the field measurements were taken for each 13 x 30 m experimental subplots (Figure 10). These 13 x 30 m subplots are results from the large plot (30 x 30 m) division into two part, left side and right side. For comparing the N status which was derived from three different methods, the large plots (i.e. A through L) will be used. The same large plots will also be used in identifying relationships between leaf chlorophyll with plant sap and

calculating the regression analysis to find the best representative VI for crop N status. Conversely, in the similarity measures with mean curve approach, the 13 x 30 m subplots will be used (i.e. plot AL, AR, BL, BR, CL…, and LR).

3.3 Methods

The outline of the proposed methodology in this thesis research is presented in the flowchart below (Figure 11). The flowchart explains the *step-by*-step methods that will be applied in order to achieve the research objectives in Section 1.3. The research is divided into five phases, of which each phase is intended to answer each research question. The first phase focuses mainly on analysing the best VI that gives a good representation in estimating the potato N status over the growing season. Before analysing the best VI, several analyses related to crop biophysical and biochemical parameters will be conducted beforehand. The analysis will start by comparing the N status over the growing season, which is derived from three different methods: the Dumas-combustion test, PSNC test and chlorophyll meter. Secondly, the relation between the N status in crops with chlorophyll concentration both in the leaf and canopy level will be identified, where the chlorophyll concentration in the leaf level will be calculated using Equation (3). Next, the VIs' pattern over the growing season will be analysed. The purpose of this analysis is to find out the variability in each experimental plot over the growing season. The last part of this phase is to derive the VIs' value from the Cropscan data and identify the best VI that is able to give a good representation of the N status in the potato crop over the growing season. This phase is explained in section 3.3.1.

After deciding the best representative VI for the N status, the time series profile of the VI will be set for each experimental plot. The next phase of the research is to determine the optimal growth curve for the potatoes - the reference curve - represented by the time-series of the VI identified in phase one. In this report, two approaches are used to construct the reference curve: the maximum yield approach and the mean curve approach. For the maximum yield approach, the yield analysis will be used. The methods that will be used in the second phase are explained in section 3.3.2.

Phase three is a reflection of the third research question. The main purpose of phase three is to confer the time-series analysis method that can be adopted to evaluate deviation from the established reference curve. Performing the time series similarity measures on the VI time series data will make it possible to characterize the growth status within the experimental plots. Therefore, the growth deviation, which is indicated by the VI values, can be tracked along the growing season. This phase is extensively explained in section 3.3.3. Using the characteristic of the data, the proper time series similarity measures will be chosen to solve the problem.

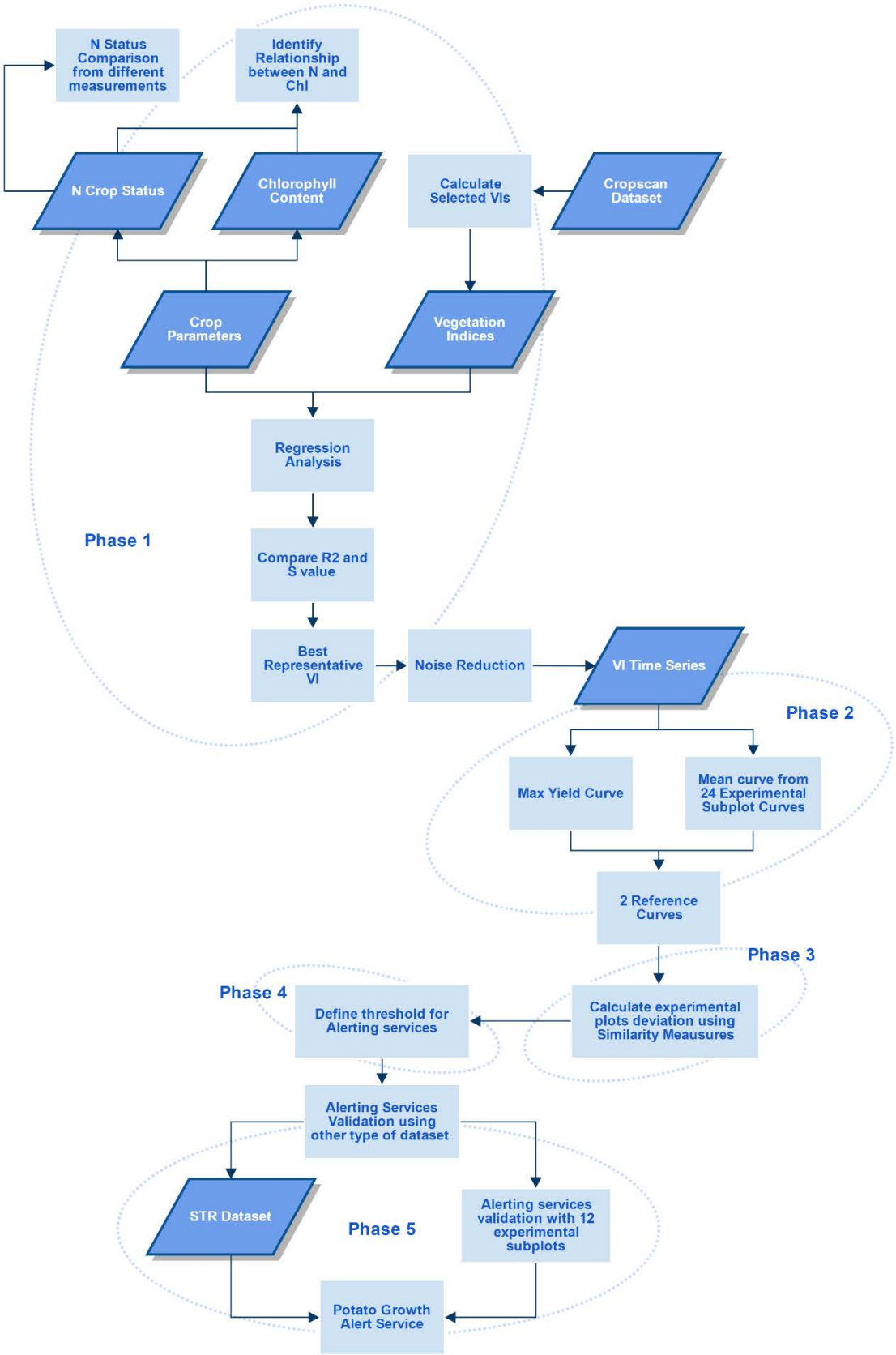


Figure 11: Methodology Flowchart

The fourth phase of the methodology flowchart is to determine the threshold value for the maximum acceptable difference, or deviation, between two VI time series. This phase has a relevancy to detect deviating plant growth which can be used to derive an alerting service. This phase is described in section 3.3.4.

After accomplishing phase four, the last phase is to validate the alerting service. Two approaches will be adopted to validate the accuracy of the potato growth alerting services. Firstly, the image based dataset of Geveart et al. (2015) will be used to check the alerting service's threshold sensitivity to plant growth deviation. Secondly, the 12 validation subplots will be used to evaluate the alerting service for the subplots level. Section 3.3.5 gives a complete explanation of this last phase.

3.3.1. Vegetation Indices for potato N status (Phase 1)

3.3.1.1 N Status comparison with different measurement methods

The N status measurements over the growing season in this report used three different methods: wet chemistry analysis + the Dumas-combustion method; PSNC test, also known as the Plant Sap test; and chlorophyll meter analysis (Minolta SPAD-502). To give an overview of N status changes within the potato crop over the growing season, all N status values from the different methods were plotted over time. Although the number of observations varied between each method (Table 4), the overall conditions in the beginning and at the end of the growing season were still comparable. The N status development plots were made using the *Python* programming language. Besides comparing the N status in the potato crop with the three different measurement methods, the N status development plots will also give information related to the N status for each 30 x 30 m experimental plot over the growing season.

To examine the relation between the N statuses acquired from the three different methods, a regression model will be created using the *Minitab 14* software. Since the data from the Dumas-combustion method only are available on three dates (14 June, 26 July and 31 July), any regression analysis related to this method will be calculated using 24 samples for each variable. The relationships among these three methods will be assessed by the coefficient of determination (R^2).

3.3.1.2 Relationships between crop biophysical and biochemical parameters

The relationships between the N status in the potato crop with the chlorophyll concentration values (both at the leaf level and canopy level) are essential information. As explained by Taiz and Zeiger (2010), leaf chlorophyll content can be used as an N status indicator. Therefore, it is necessary to check the relationship amongst these crop parameters as a validation phase before making use of chlorophyll content in further calculations. The leaf chlorophyll concentration values will be calculated using Equation (4).

This parameter will then be paired with the N status value to find their relation using regression analysis.

The other crop biophysical parameter - chlorophyll canopy - will also be paired with the crop N status to find their relationship. The chlorophyll canopy values will be calculated using:

$$Chlorophyll\ Canopy = leaf\ chlorophyll * LAI \quad (5)$$

Besides investigating the overall relationship between the N crop status and the chlorophyll content (at leaf and canopy level), the influences of temporal and spatial scales and different fertilizer levels will also be analysed. The analysis will use a stepwise regression that is provided by Minitab 14. The most influencing factor for the relationship between the N crop status and the chlorophyll content values will later be identified using this analysis.

3.3.1.3 Relationships between VIs and chlorophyll content

The relationships between the eight different VIs with chlorophyll content will be investigated to examine the most representative VI that is able to detect the N status in the potato crop over the growing season. The previous study conducted by Clevers and Kooistra (2012) used six VIs (REP, MTCI, MCARI/OSAVI, TCARI/OSAVI, CI_{green}, and $CI_{red\text{-}edge}$) to find the best VI for crop chlorophyll content. In this research, well-known satellite-derived VIs are also considered, including NDVI, WDVI and EVI. VIs used in this report will be calculated using the formulas in Table 1 with reflection values from Cropscan measurements. Regression analyses will then be applied to find the best VI.

Statistical Analysis

Both simple linear regression and non-linear regression will be used to model the relationship between the chlorophyll content (both in the leaf and canopy level) and predictor variables (eight VIs). Goodness of fit, or how well a model fits the data, will be evaluated based on the coefficient of determination (R^2). The selection of the best VI will be based on the combination of the R^2 value and standard error of the regression (S). The S value provides important information that R^2 cannot provide.

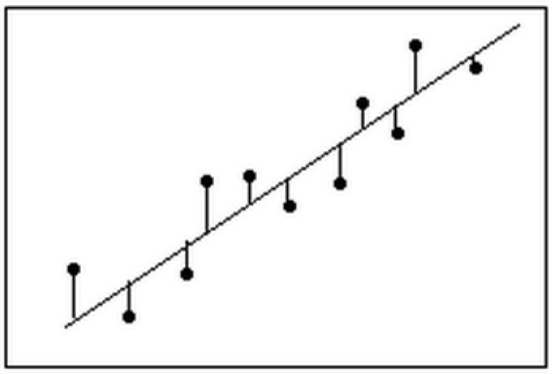


Figure 12: Illustration of S calculation

The S value, as shown in Figure 12, represents the average distance that is calculated from the observed values to the regression line (Frost, 2014). Therefore, the S value explains how wrong the regression model is. A low S value specifies that the observations are closer to the fitted line. The best VI will be decided by the highest R^2 value and the lowest S value.

3.3.2 Optimal growth curve for potato crop (Phase 2)

To answer the second research question, a reference curve needs to be constructed in Phase 2. In this report, there are two approaches to acquire the reference curve: maximum yield and mean curve approach. These two approaches have their own emergence - the maximum yield approach is most likely to achieve the long term goal, while the mean curve can be used in near real time purposes. The reason that the maximum yield approach should be used for a long term goal alerting service is that maximum yield value can only be generated at the end of the growing season. This means that this approach cannot be used for near real time acquisition.

The mean curve is generated from calculating the mean value from 24 experimental subplots for each observation date over the growing season. The mean curve end result will be considered as the optimal growth curve for potatoes (reference curve). Deviations of the 12 experimental plots from the reference curve can then be calculated. The method presented by Bala and Islam (2009) gives an insight of mean curve approach in constructing a regression model to predict the potato yield in Bangladesh.

Maximum yield approach

Constructing a reference curve based on the maximum yield approach is performed by selecting the experimental plot which produces the highest amount of yield. This information can be gathered from the yield analysis which is normally done by the farm management. When the field that produces the maximum yield has been found, the VI time series value from this experimental plot will be used as the reference curve.

Mean curve approach

After selecting the best representative VI for potato health status in Phase 1, the time-series of the selected VI will be calculated for all 24 experimental subplots over the growing season. To generate a reference curve using the mean curve approach, the 24 subplots will be used to get the mean value for each observation date and will be assigned as the reference value (Figure 13). After establishing the reference curve, the 12 plots (A, B, C, D, E, F, G, H, I, J, K, and L) will then be compared to the reference curve using the similarity measures which will be explained later in Section 3.3.3. The other data from the STR dataset will be used for the validation process (Figure 11).

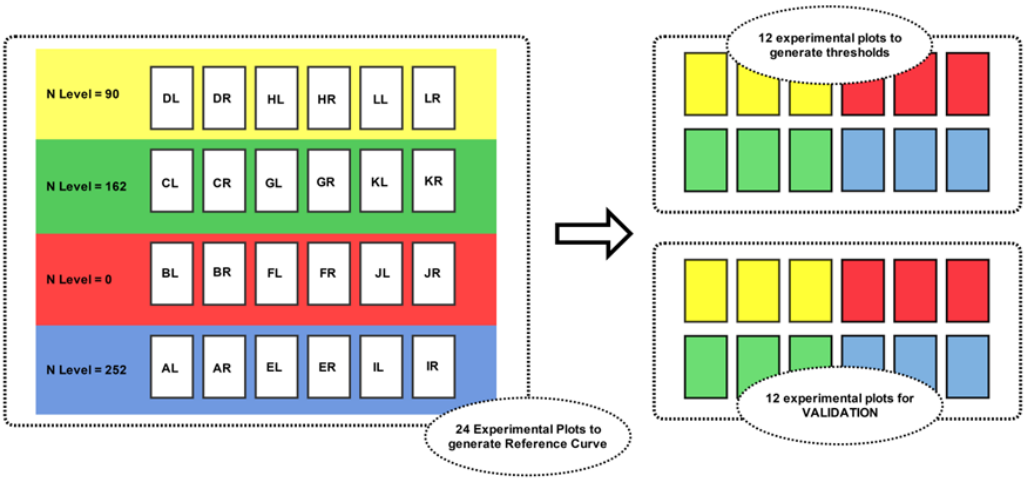


Figure 13: Sampling design for mean curve approach

As seen in Figure 13, the 24 experimental subplots will be divided into two groups: the calibration (thresholds) dataset and the validation dataset. The calibration dataset will be used in Section 4.3 and 4.4 for performing the similarity measures and control chart in the subplots level using the mean curve approach; whilst the validation dataset will be used in Section 4.5 for alerting service validation purposes in the subplots level.

3.3.3 Similarity measures in time series analysis for deviation detection (Phase 3)

The purpose of this research question is establishing proper similarity measures to detect the difference between temporal profiles. Therefore, the deviation between the experimental plots' time series profiles with the reference curve can be calculated. The deviation or difference occurring in the similarity measures will be able to emphasize the variety of crop development patterns due to the different fertilizer treatments applied over the growing season. There are several time-series analysis methods for change detection explained by Lhermitte et al. (2011). Similarity measures such as Euclidean Distance, Manhattan Distance, Principal Component Analysis (D_{PCA}) and Fourier based were revealed sensitive to quantify the difference in time-series values (Lhermitte et al., 2011). A previous research (Kooistra et al., 2012) adopted the Euclidean distance between two individual points in time to calculate deviating nitrogen conditions and a potential need for a management action in the field.

Similarity is considered as the measure that establishes an absolute value of resemblance between two vectors, in principle isolated from the rest of the vectors and without assessing the location inside the solution space. The time series distance measures are normally divided into four categories: shape based, edit based, features based and structure based. The commonly used similarity measures for original time-series data will be performed in this report. Therefore, the similarity measures that will be used in this report comes from the shape based category and correlation measures as an additional method. The measures from the shape based measures category that will be used in this report are:

Manhattan distance (D_{Man}) and Euclidean distance (D_E). As it is explained in Section 2.4, these two measures are derived from Minkowski distance and explained by:

$$D_{Man} = \sum_{t=1}^{N} \left| f_t^p - f_t^q \right| \quad (6)$$

$$D_E = \sqrt{\sum_{t=1}^{N} \left(f_t^p - f_t^q \right)^2} \quad (7)$$

, where f_t^p is the $f^p(t)$ time series value at moment t and number of samples N in the time series. Using these two similarity measures, the difference between the two time-series (reference plot with other experimental plot) will be calculated. The distance will be calculated for each observation date over the growing season. The distance between two points at moment t is calculated using all measurements at moment *t* and before *t*, except for the first point measurement. To illustrate this calculation, a first distance measurement only calculates the distance at the first date of measurement. While the second distance measurement calculates the accumulation of the values from first date and second date to measure the difference.

The other similarity measures that will be used in this report are Pearson's correlation coefficient (D_{CC}) and root mean square distance (RMSD). Both these measures follow the same rules as Manhattan and Euclidean distance. For D_{CC}, the correlation of two time series at moment *t* is translated as the accumulation of all measurements at *t* and before *t*. However, this does not apply to the measurement on the first observation date, as this is the initial time series value. As it explained in Section 2.4, Lhermitte et al. (2011) defined the Pearson's correlation coefficient with the equation:

$$D_{CC} = \frac{\sum_{t=0}^{N-1} \left[\left(f_t^p - \bar{f}^p \right) * \left(f_{t-s}^q - \bar{f}^q \right) \right]}{\sqrt{\sum_{t=0}^{N-1} \left(f_t^p - \bar{f}^p \right)^2} * \sqrt{\sum_{t=0}^{N-1} \left(f_{t-s}^q - \bar{f}^q \right)^2}} \quad (2)$$

Where f_t^p and f_t^q are time series values at moment t and the mean values of the corresponding series are indicated by $\bar{f}^p$ and $\bar{f}^q$. *N* is the length of the time series, while s is specifying the delay (lag) between both time series. A D_{CC} value that is close to 1 gives the indication of an increasing linear relationship, whereas a value of -1 indicates a decrease in the linear relationship (Lhermitte et al., 2010). In this report, the s was specified as 0 which means the D_{CC} estimates time series similarity without the time shift.

Using the RMSD, the straight-line inter-point distance was calculated for each corresponding observation date over the growing season. The RMSD is explained as:

$$RMSD = \sqrt{\frac{\sum_{t=0}^{N-1}\left(f_t^p - f_t^q\right)^2}{N^2}} \quad (3)$$

Where f_t^p and f_t^q are time series values at moment t and N is the length of the time series. The same rules that are applied to D_{Man}, D_E, and D_{CC} also applied to RMSD. A low RMSD value indicates a high temporal similarity between two time series in moment *t*, while a high value indicates the opposite. To calculate the RMSD between each assigned reference plot in section 3.3.2 with the other experimental plots, equation (3) will be applied.

3.3.4 Plant growth alerting service (Phase 4)

As shown in the methodology flowchart (Figure 11), the purpose of Phase 4 is to define the thresholds for the alerting services. The thresholds in the alerting services indicates whether the growth development of each experimental plot, when compared to the reference curve, is considered as within the control condition or not. To specify these alerting service thresholds, the control chart method will be used in this report. The results of the distance measures and correlations between two comparable time series in Phase 3 will be checked using the control chart theory to see whether these values lay in the acceptable area of the control chart or not. If the distance measures or the correlation values from Phase 3 are located in the area above the upper control limit or below the lower control limit, then the control chart will trigger alerts for the N status condition in the potato crop. This alert will inform the farmer of that at a specific point in time in a specific experimental plot, the growth pattern of the potato crop did not follow the reference curve (healthy potato) and needs management attention. In this report, the mean of the data will be used to construct the control charts. The control limits will be computed as μ+3σ (UCL) and μ-3σ (LCL), whilst the warning limits will be calculated as μ ± 2σ.

Three different results will be derived from Section 4.3: a plant growth alerting service using the maximum yield approach for the plot level, a plant growth alerting service using the mean curve approach for the plot level, and a plant growth alerting service using the mean curve approach for the subplots level. In the subplots level, only the alerting service using the mean curve approach will be tested. The reason for this is that the mean curve approach is assumed to be the best approach for within season monitoring.

3.3.5 Alerting services validation (Phase 5)

The alerting services in Phase 4 will be constructed using two different reference curve approaches in the plot level: maximum yield and mean curve approach. Therefore, to validate the alerting services, both approaches will be evaluated using image based dataset of Geveart et al. (2015). The STR dataset is a simulated dataset from the Formosat-2 imagery and the UAV imagery which will be used to test the accuracy of the alerting services for different type of spectral and spatial resolution datasets.

The first step in the validation phase is to generate the VI time series from the 12 experimental plots from the STR dataset. Then the distance differences between the experimental plots and the potato growth reference curve (both approaches) will be calculated using time series similarity measures. After calculating the deviation from the 11 experimental plots (maximum yield approach) or 12 plots (mean curve approach) to the reference curve, the plots that have an alert during the growing season will be investigated. To validate the alerting services, the yield result for the plots that have an alert will be compared with others to see whether they have a low yield result at the end of the growing season or not. In this validation, the management application (e.g., nutrient application, irrigation) for the experimental plots will be considered, since it might contribute to the potato yield end-result. Finally, to validate the alerting service using the mean curve approach in the subplots level, the validation subplot dataset (Figure 13) will be used in this phase.

4.1 Vegetation Indices for potato N status (Phase 1)

To answer the Phase 1 main question which is defining the best VI that gives a good representation in estimating the potato N status over the growing season, two other analyses were conducted. The first analysis compared the N crop status over the growing season using three different measurement methods: the Dumas-combustion method, PSNC test or plant sap, and chlorophyll meter using Minolta SPAD-502. The second analysis was conducted to see the relation between plant biophysical and biochemical parameters. After that, regression analysis was used to find the best VI that best represents the N potato status.

4.1.1 N Status comparison with different measurement methods

N status plots comparison

The N potato status was acquired using three measurement types: the Dumas-combustion method, PSNC tests and the chlorophyll meter method. The Dumas-combustion analysis determines the information related to fresh and dry above ground weight, dry matter content and total N concentration in the above ground parts. In this report, N concentration was selected as the N potato status that will be compared to the other two methods in the next step of this section. The PSNC tests were acquired for ten observation dates over the growing season for the 30x30 m experimental plots. However, only eight plots had the plant sap NO3 information from the PSNC test (A, B, C, D, E, F, G and H). The last measurement was conducted at the leaf level using the chlorophyll meter method. Minolta SPAD-502 was used to measure the N status in 24 13x30 m experimental plots for 11 observation dates over the growing season. Since the N status values over the growing season from all three measurements should be compared, the uniformity of the data was set as such: eight 30x30 m plots (A, B, C, D, E, F, G and H) were involved and 10 observation dates were used (except for N concentration which only has 3 observation dates).

All the N status values, which were generated from three different measurements, showed an overall decreasing pattern over the growing season. The highest N status readings were found in approximately 46 until 56 days after planting (DAP) for all the three methods. These values then smoothly decreased along the growing season in the SPAD measurement readings, while a more rapid decline occurred in the PSNC test values, also known as petiole sap NO_3 values. The decreasing pattern found in the petiole sap values from the PSNC test over the growing season was most likely due to the depletion of soil nitrogen supply and the increase of the potato canopy size.

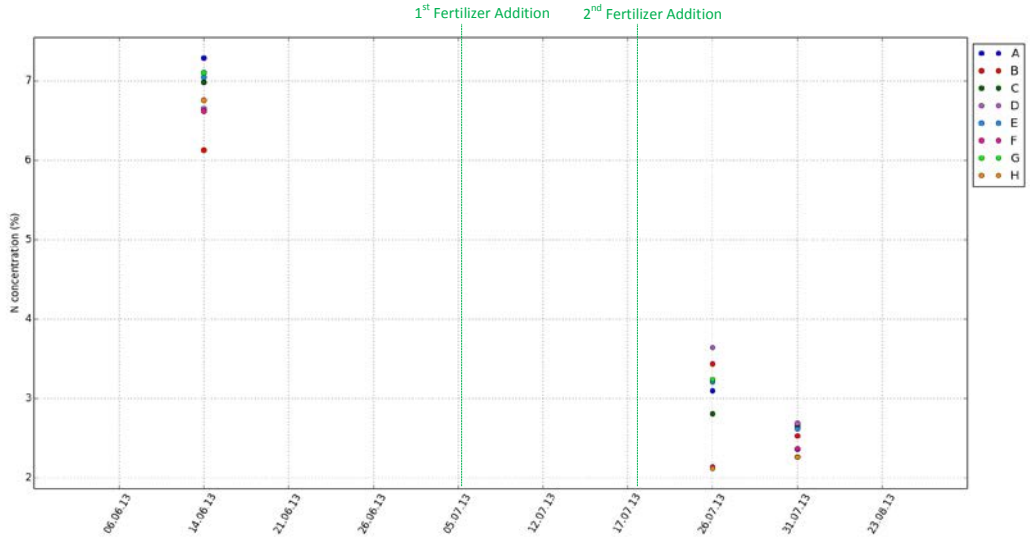


Figure 14: N status values over the growing season acquired from Dumas-combustion method

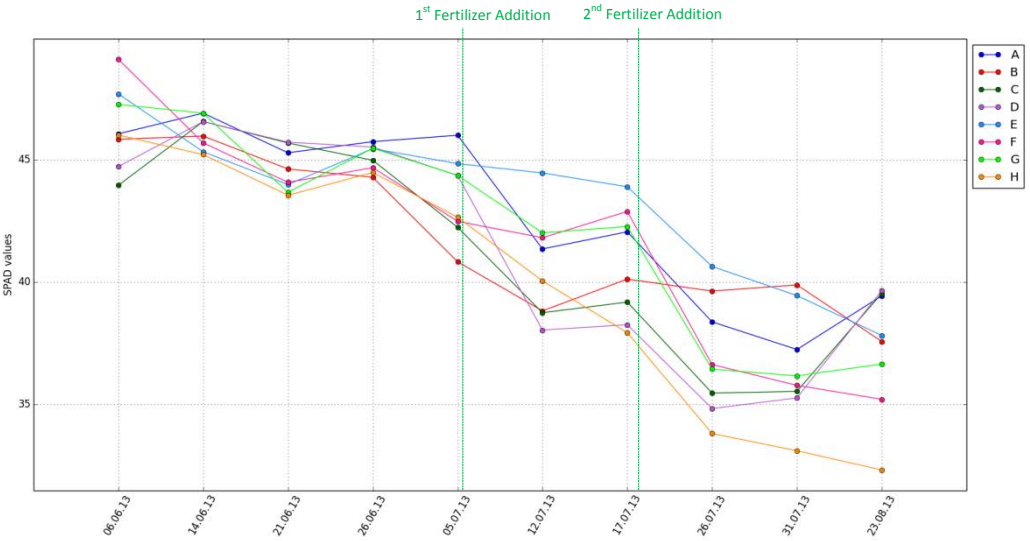


Figure 15: N status values over the growing season acquired from chlorophyll meter (SPAD) measurement

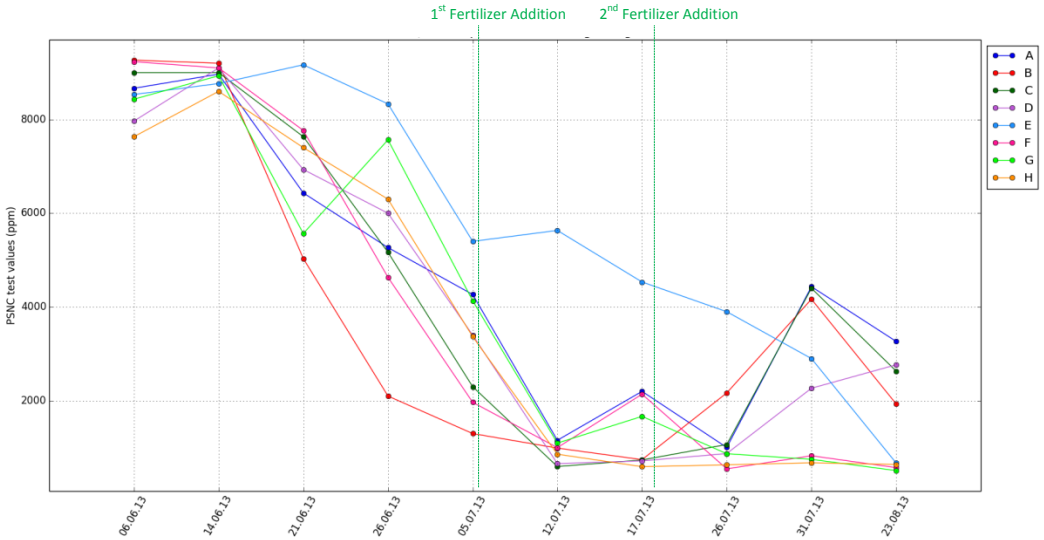


Figure 16: N status values over the growing season acquired from PSNC (Plant sap) test

The PSNC test values at the end of the growing season was relatively low: 1934 – 3267 ppm NO_3 for plot A – D and 510 – 674 ppm NO_3 for plot E – H, but these values were acceptable. These results show that the potato crops had sufficient nitrogen supply to meet the tuber yield and size distribution targets.

The plot with the highest N fertilizer among these eight experimental plots (plot E) shows the delayed decline over the growing season. This shows in both the PSNC test plot (Figure 16) and the SPAD measurement plot (Figure 15). This condition cannot be tracked down by the N concentration value from the Dumas-combustion method since there were only three observation dates available over the growing season.

Table 5: Management decision in the experimental plots over the growing season

	April	May	June						July								August	
	21	24	06	08	14	15	21	26	05	07	12	16	17	18	26	31	16	23
Planting																		
Emergence																		
Additional Fertilizer (specific plots)																		
Irrigation																		
Field Observation																		

Gradual steep decreases were found in almost all nitrate concentration readings (PSNC tests) from 26 June until 12 July 2013, with only plot E showing different behaviour (increasing value on 12 July). Similar, but less intense, decreasing patterns were also found in the SPAD readings for the same dates. The decrease appearing in both the SPAD and PNSC readings were most likely due to the potato development, while the "dry event" influenced the N uptake from the soil.

As shown in Table 5, additional fertilizer was applied twice to the specific plots. The first additional fertilizer was applied to plot B and plot J on 5 July 2013 with the amount of 29 and 22 kg per hectare, respectively. This fertilizer application influenced both the SPAD and PSNC tests on multiple readings in following measurement periods. The effects of the fertilizer application was shown in the nitrate concentration reading with an increasing value for plot B on 26 July 2013. A week after (31 July 2013), an increasing value was detected in the SPAD readings for experimental plot B, while the values for the eight other plots decreased (Figure 15). The second fertilizer application were also observed in the SPAD readings on 16 August 2013 (this date was not included in the plot, due to the data uniformity). Even though the main increase appeared a month after the application, the increasing values of plot A, B, C and D still appeared on 23 August 2013. Differing from the SPAD readings, the second fertilizer event which was applied to plot A, B, C, D, I, J, K and L, with the amount of fertilizer as shown in Table 2, was affecting the nitrate concentration readings on 31 July. Figure 16 shows an increase in the plot A, B, C and D readings for the measurement on 31 July 2013 which is most likely influenced by the fertilizer addition in the crop field. The effect of irrigation on the experimental plots still could not be found in the three N status plots.

Another difference was shown at the end of the growing season. The N concentration plots show that the values were relatively low with no specific difference between each of the plots (the values were uniformly low), while in the SPAD and PSNC test plots the values for each plot were more diverse.

Relationship of different N status measurements

The relationship between N concentrations with the other two N status measurement methods were done using the regression analysis of 24 samples over the growing season. As seen in Figure 17, the N concentration values in the beginning and at the end of the growing season (the only available measurement) show significant relationships to both SPAD and PSNC test values. These were shown by relatively high R^2 values: 0.868 for the relationship between N concentration with SPAD values; and 0.848 for the relationship between N concentration and nitrate nitrogen from the PSNC test.

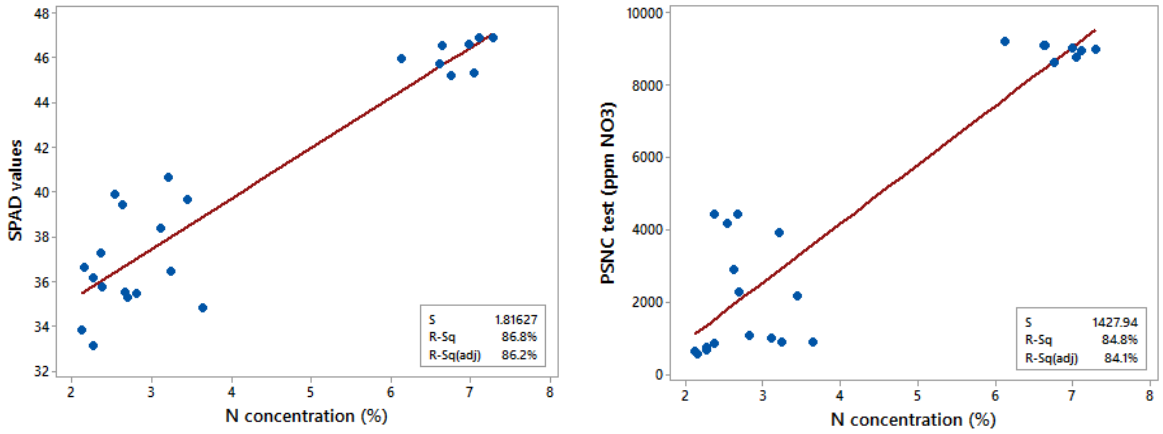


Figure 17: Relationship between N concentration (Dumas-combustion method) with SPAD values (left) and with PSNC test method (right) from the measurements taken in the beginning and at the end of growing season.

However, these relationships only reflected the condition of observations in the beginning and at the end of the growing season. As shown in Table 4, the observation dates that were available for the N concentration values were 14 June, 26 July and 31 July 2013. Since the data was not available after additional fertilizers were applied to the experimental plots, the effect of the fertilizer treatment could not be compared through the relationship plots in Figure 17.

Figure 18 shows the relationship between the PNSC test values (nitrate nitrogen concentration) and the SPAD measurement. From 80 samples gathered from ten observation dates for eight 30x30 m experimental plots (A through H), the PSNC test values range from 510 to 9266 ppm NO_3. The values range from 32.31 to 49.11 for the SPAD measurements. The R^2 value of 0.75 shows that the relationship between these two measurement methods is moderately high. Both measurements have high N status values at the beginning of the growing season, but decreases over time.

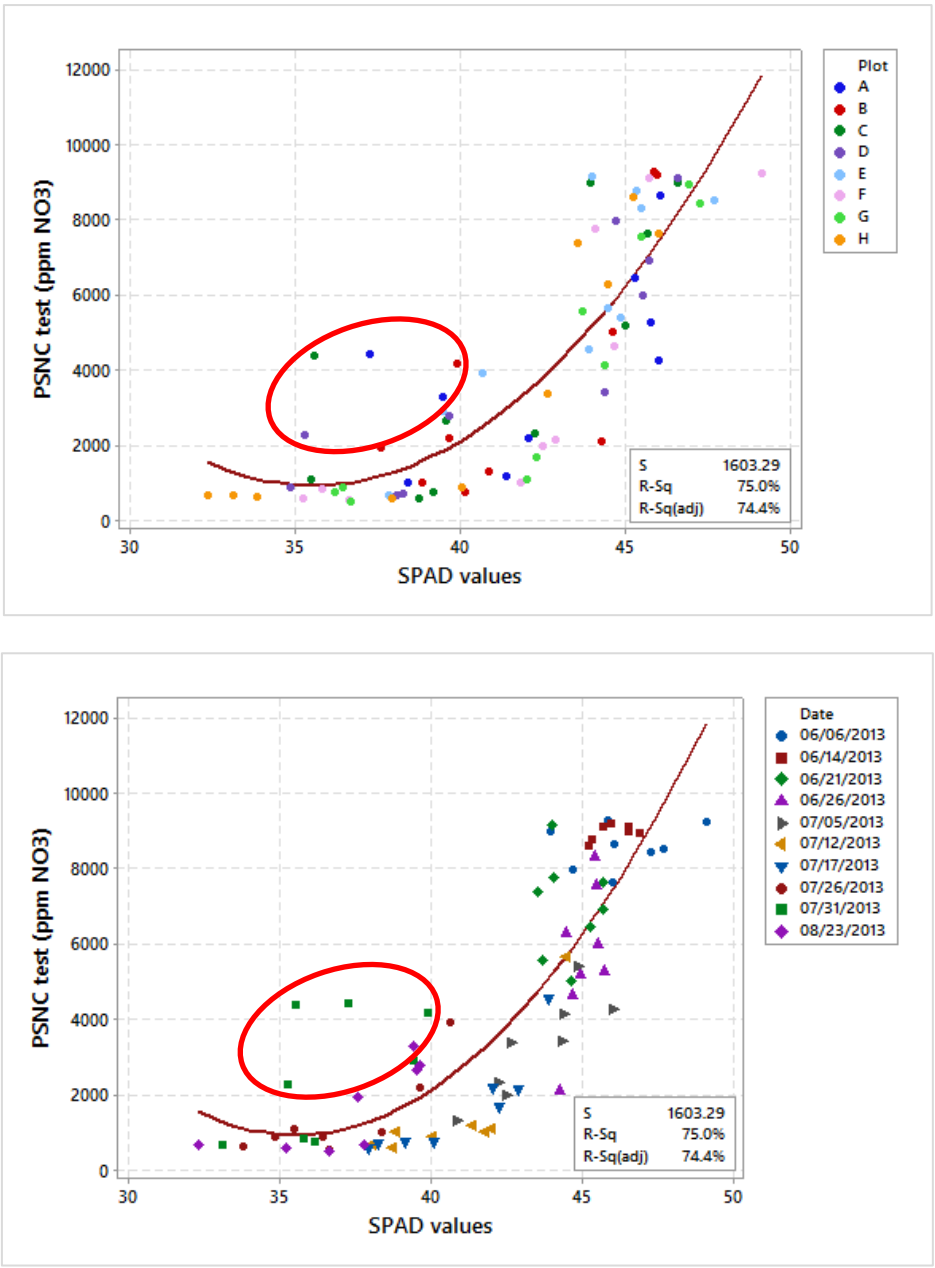


Figure 18: Relationship between PNSC test and SPAD values with additional explanatory variables: based on plots (above) and based on observation date (below)

The SPAD values (N leaf level) were shown to be sensitive to the high nitrate concentration values acquired from the PSNC test, but deviated more when the nitrate concentration values were low. Figure 18 shows that the SPAD values were varied around 32 – 40 for the nitrate concentration value around 800 ppm NO_3.

There were some values that showed different behaviour from the rest of the observation values (the red circle). In Figure 18, the values from plot A, B, C and D were acquired on 31 July 2013. These N status values were located away from the regression line. The values were high for the PSNC test and relatively low for the SPAD measurement. This behaviour indicates that a management application was applied to the experimental plots. Based on the given information in Table 5, a fertilizer application was applied to the eight

experimental plots (A, B, C, D, I, J, K and L) on 26 July 2013. Based on this information it can be concluded that the values for plot A, B C and D from the PSNC tests increased on 31 July due to the response from the fields initiated by the applied fertilizer application. Since the measurement data from the SPAD and PSNC tests were gathered on the same date, it can be concluded that the PSNC test is more sensitive to the N application.

4.1.2. Relationships between crop biophysical and biochemical parameters

Figure 19 shows the relationship between the leaf chlorophyll and the nitrate concentration derived from the PSNC tests with all data included in the calculation. The overall relationship between these two variables is relatively significant, with the coefficient of determination R^2 equal to 0.745. Since the leaf chlorophyll values were calculated from the SPAD values using Equation (4), then the uncommon behaviour that occurred in Figure 18 was expected in Figure 19. As shown in Figure 18, there was an uncommon behaviour taking place in the observations collected from July 31 for plot A, B, C and D. These outlier values show that the influence of additional fertilizer application in the specific experimental plots can be detected by the PSNC test as soon as one week after the application was applied. These values were also shown in the relationship between PSNC test values with leaf chlorophyll (red circle).

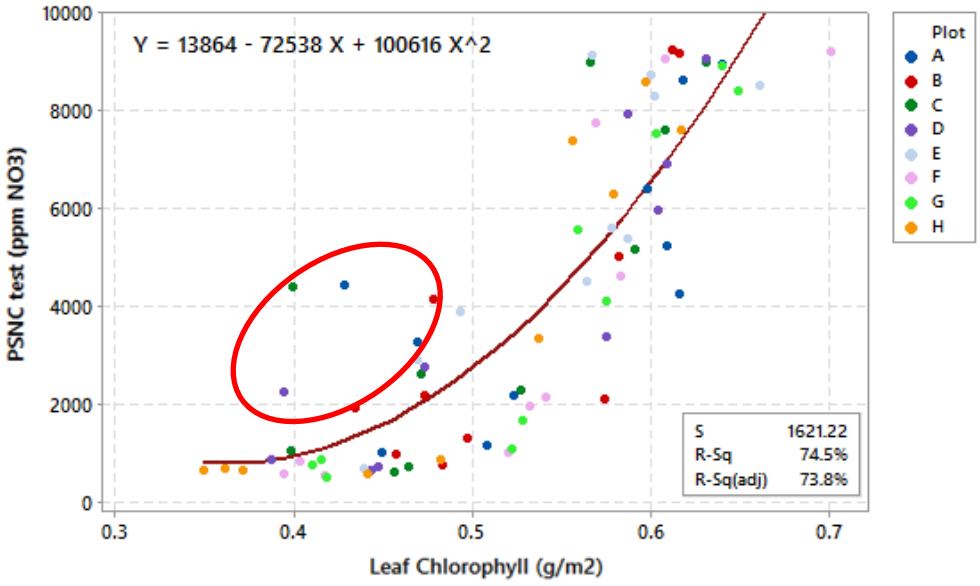


Figure 19: Scatter plot of PSNC test (Y) and leaf chlorophyll (X) using all data from 10 observation dates

Apart from that, the relationship between the PSNC test values with leaf chlorophyll concentration shows a better result than the relationship between the PSNC test and chlorophyll canopy. The coefficient of determination (R^2= 0.075), as shown in Figure 20, clearly indicates that the relationship between the PSNC test and the chlorophyll canopy is significantly low. This condition might be explained by the fact that both parameters represent the values at different scales. The PSNC test was conducted at the leaf level, while

the chlorophyll canopy represents the chlorophyll content at the canopy level. Therefore, the relation between these two variables is insignificant.

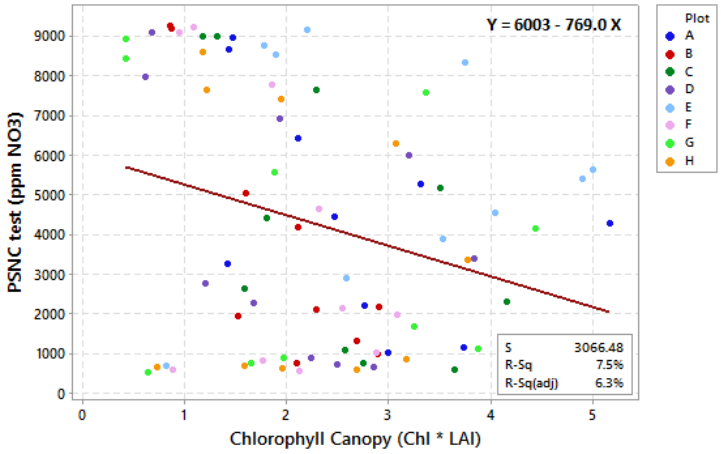


Figure 20: Scatter plot of PSNC test (Y) and canopy chlorophyll (X) using all data from 10 observation dates

Figure 21 shows the relationship between the nitrate nitrogen values from the PSNC test and the leaf chlorophyll (g/m^2) based on the date of the observations. The relationship between these two variables showed insignificant results at the beginning of the growing season (6 June – 26 June). The R^2 values on 6 June and 21 June were almost close to 0. This was another indication that there was almost no relation between the PSNC test values with leaf chlorophyll at the beginning of the growing season. Conversely, the rest of the calculations show relatively significant results for the two variables. The most significant relationship between the two variables was shown on the seventh observation (17 July). On 31 July the PSNC test was able to detect the second additional fertilizer application, which resulted in an insignificant relationship (p-value = 0.157) between the two variables. The overall relationship between the PSNC test and leaf chlorophyll can be seen from Table 6.

The previous result shows that the relationship of the PSNC test and the canopy chlorophyll for all observations was insignificant (Figure 20). One of the purposes of this section is to find out the relationship between PSNC test values with chlorophyll canopy based on the date of observations. As it had been carried out (Table 6), the relationship between these two parameters shows insignificant results at the beginning of the growing season between the first and third observation. Besides those particular dates, the relationship between these two values over the growing season were relatively significant. Therefore, a recalculation was made to see the overall relationship between the PSNC test values with chlorophyll canopy by excluding the observations between 6 until 21 June. As shown in Appendix 2, the coefficient of determination still shows a relatively low value (R^2 = 0.231). This emphasizes that the insignificant result between these two variables was caused by different measurement levels at the crop, not due to outlier values on specific observation dates.

Table 6: R^2, S value and p-value for the relationship between PSNC test – leaf chlorophyll and PSNC test – Chlorophyll Canopy at each date of observations

Date	PSNC test - Leaf Chlorophyll			PSNC test - Chlorophyll Canopy		
	R-Sq	S	p-value	R-Sq	S	p-value
06/06/2013	0.055	611.769	0.575	0.005	627.866	0.867
14/06/2013	0.218	186.161	0.243	0.192	189.236	0.277
21/06/2013	0.006	1415.470	0.859	0.476	1027.270	0.058
26/06/2013	0.390	1607.510	0.098	0.631	1249.850	0.018
05/07/2013	0.752	728.545	0.005	0.768	704.387	0.004
12/07/2013	0.632	1102.560	0.018	0.700	994.930	0.010
17/07/2013	0.768	694.650	0.004	0.682	813.406	0.012
26/07/2013	0.686	685.065	0.011	0.726	640.349	0.007
31/07/2013	0.303	1505.100	0.157	0.383	1415.870	0.102
23/08/2013	0.636	752.871	0.018	0.737	639.669	0.006

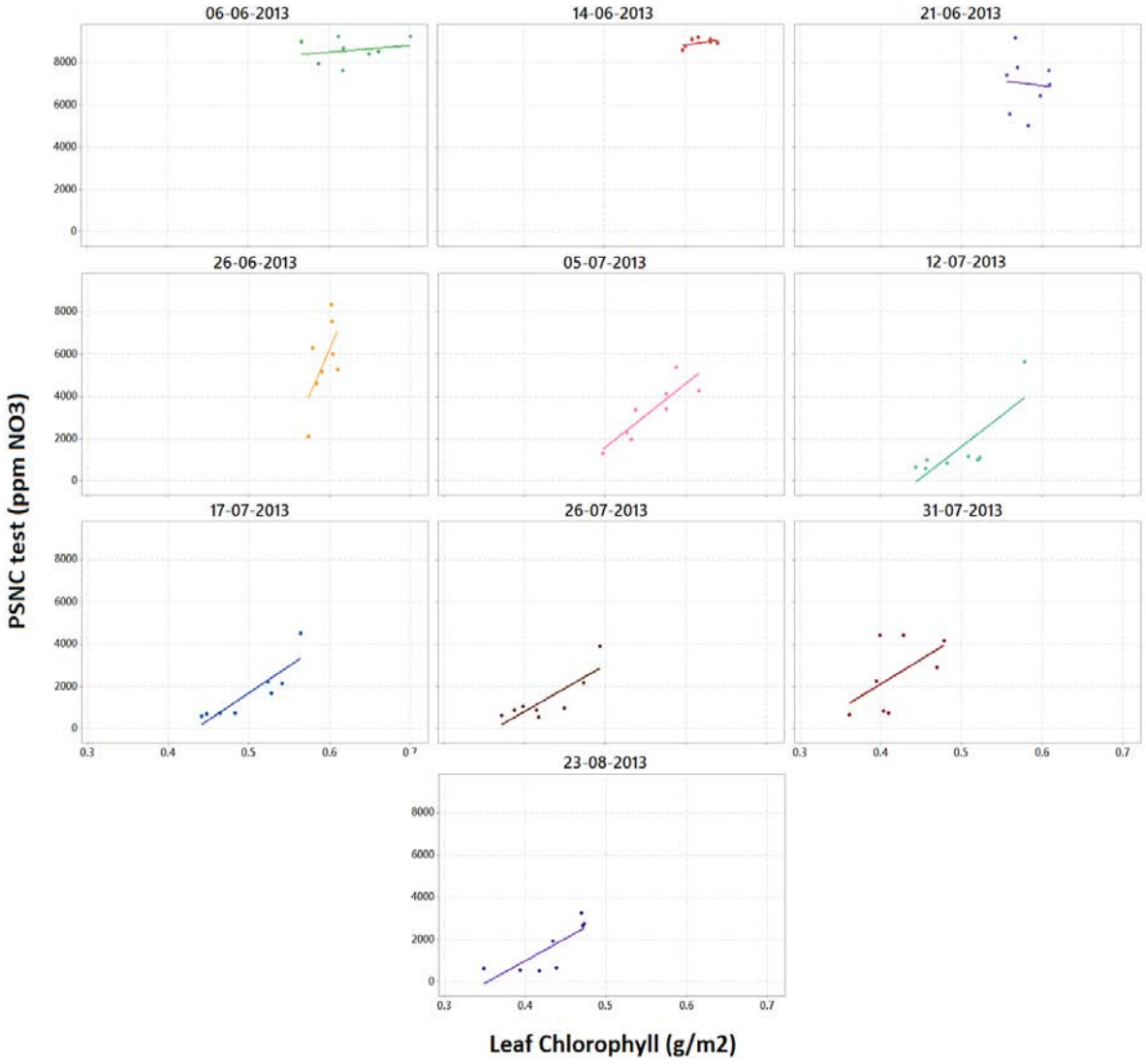


Figure 21: Relationship between PSNC test (Y) and leaf chlorophyll (X) based on the specific observation time.

The relationship between the PSNC test and the leaf chlorophyll based on the initial fertilizer applied to the experimental plots before planting is shown in Figure 22. The relationships between the PSNC test values and the leaf chlorophyll for all the fertilization levels (0, 90, 162, 252 kg N h^{-1}) show significant results (p-value < 0.05) with moderately high R^2 values. The highest relationship between the PSNC test and the leaf chlorophyll was found in the plots with an initial fertilizer level of 90 kg N h^{-1}, while the lowest one was the plots with an initial fertilizer level of 252 kg N h^{-1}.

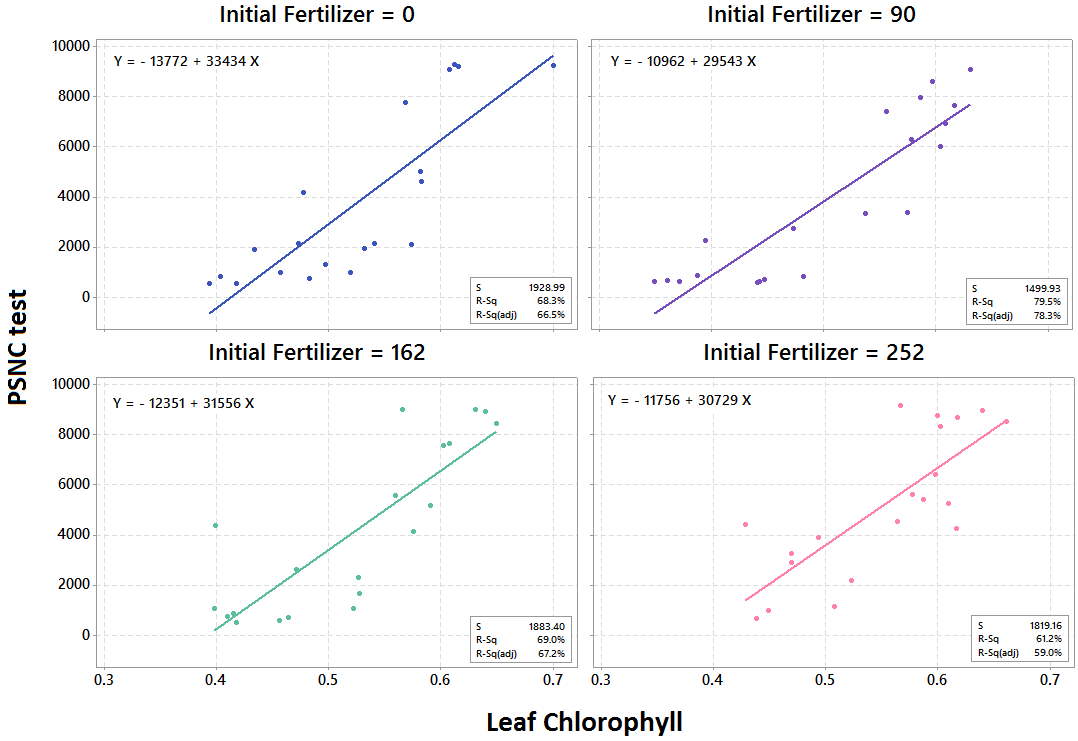


Figure 22: Scatter plot of PSNC test (Y) and leaf chlorophyll (X) using all data from 10 observation dates

To investigate the effects of the temporal aspects and level of initial fertilizer on the relationship between the PSNC test values and the leaf chlorophyll, a regression analysis was applied to the observed data over the growing season with the observation date and fertilizer level as the explanatory variables. Using Minitab 12, the coefficient of determination showed an increasing value (R^2=0.9118). This explains that the relationship between the PSNC test values and the leaf chlorophyll is considered to be significant with the influence of two explanatory variables (fertilization level and date of observation).

4.1.3 Relationships between VIs and chlorophyll content

The last step of phase 1 was to identify the best VI that is able to give a good representation of the N status in potato crops over the growing season. To find this VI, eight VIs derived from Cropscan data were regressed to leaf chlorophyll content and chlorophyll canopy which was calculated from the SPAD measurement data.

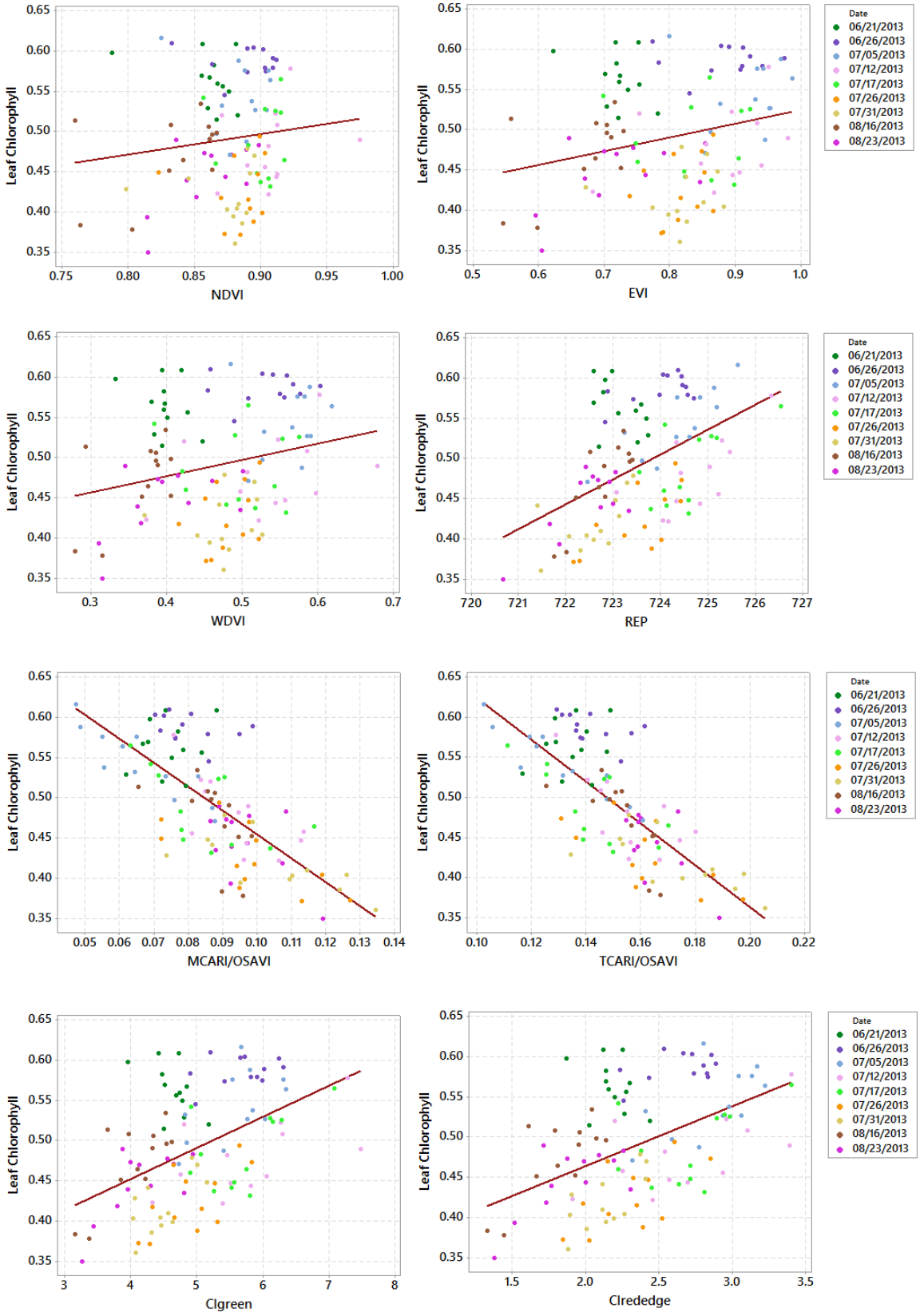


Figure 23: Relationships between leaf chlorophyll (g/m^2) and eight different vegetation indices (observations from 6 – 14 June were excluded).

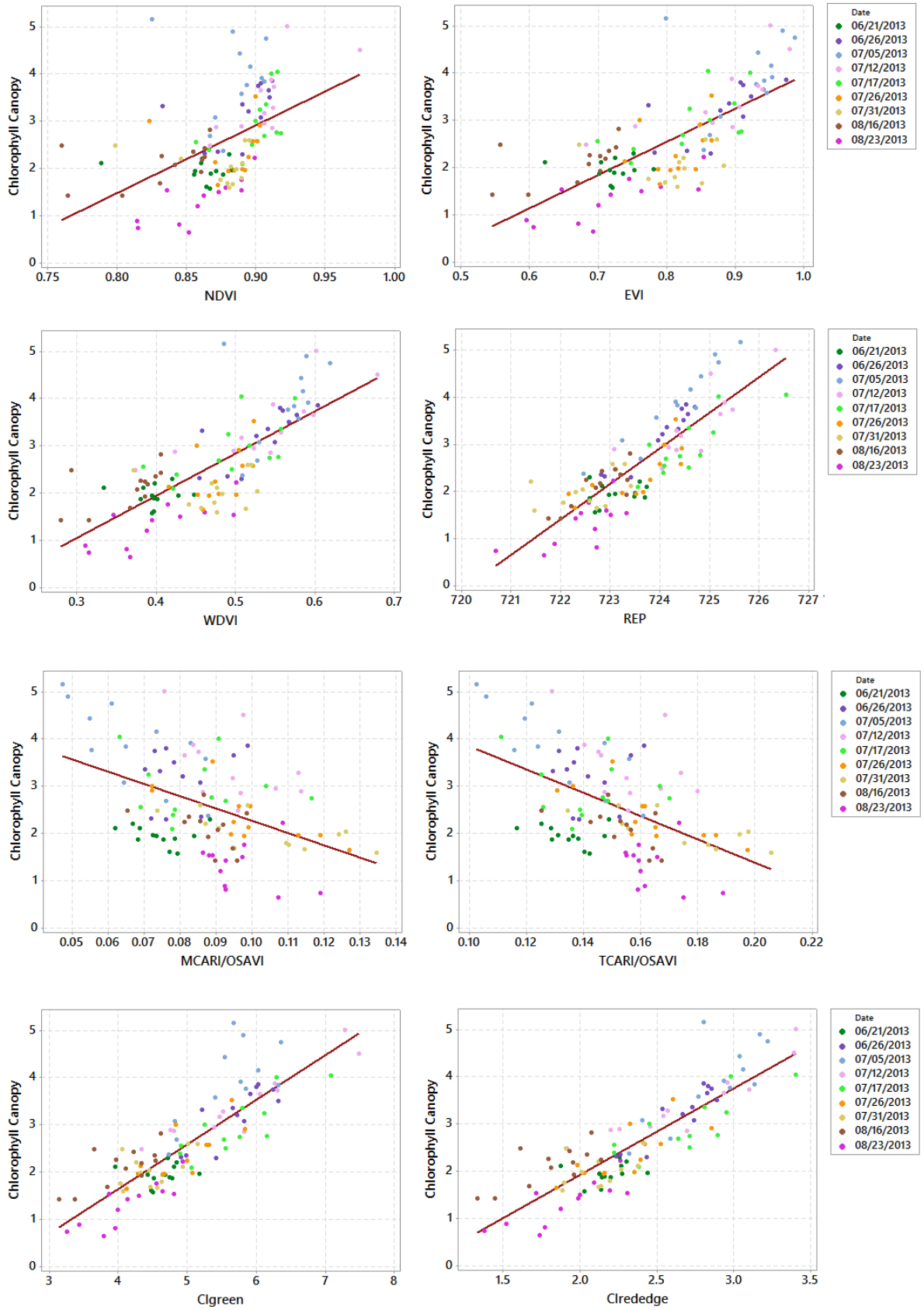


Figure 24: Relationships between chlorophyll canopy and eight different vegetation indices (observations from 6 – 14 June were excluded).

Table 7: The summary of R^2, S-Values and p-values between leaf chlorophyll or chlorophyll canopy with eight different vegetation indices for all observations (above) and with data exclusion (below).

Variable Y	Variable X	R-Sq	S-value	p-value	R-Sq	S-value	p-value
			Linear			Exponential	
Leaf Chlorophyll	NDVI	0.213	0.0719270	0.0000	0.185	0.0726900	0.0000
	EVI	0.139	0.0752517	0.0000	0.114	0.0759821	0.0000
	WDVI	0.127	0.0757778	0.0000	0.105	0.0768093	0.0000
	REP	0.017	0.0803944	0.1330	0.026	0.0797850	0.0810
	TCARI/OSAVI	**0.514**	**0.0565482**	**0.0000**	**0.517**	**0.0551240**	**0.0000**
	MCARI/OSAVI	0.461	0.0595242	0.0000	0.465	0.0572310	0.0000
	CI-green	0.041	0.0794002	0.0190	0.027	0.0797850	0.0008
	CI-rededge	0.052	0.0789601	0.0090	0.036	0.0810029	0.0000
			Linear			Quadratic	
Chlorophyll Canopy	NDVI	0.308	0.8960160	0.0000	0.562	0.7153960	0.0000
	EVI	0.551	0.8960160	0.0000	0.758	0.5322330	0.0000
	WDVI	0.607	0.6751370	0.0000	0.776	0.5114330	0.0000
	REP	0.390	0.8413750	0.0000	0.390	0.8440140	0.0000
	TCARI/OSAVI	0.002	1.0756600	0.5840	0.095	1.0282600	0.0000
	MCARI/OSAVI	0.000	1.0768500	0.9080	0.172	0.9838710	0.0000
	CI-green	0.736	0.5531390	0.0000	0.814	0.4660870	0.0000
	CI-rededge	**0.751**	**0.5375120**	**0.0000**	**0.858**	**0.4080120**	**0.0000**

Variable Y	Variable X	R-Sq	S-value	p-value	R-Sq	S-value	p-value
			Linear			Exponential	
Leaf Chlorophyll	NDVI	0.057	0.0673971	0.0130	0.063	0.0664005	0.1800
	EVI	0.091	0.0661600	0.0010	0.094	0.0657810	0.0090
	WDVI	0.082	0.0665147	0.0030	0.084	0.0662321	0.0060
	REP	0.237	0.0606071	0.0000	0.253	0.0600102	0.0000
	TCARI/OSAVI	**0.579**	**0.0450599**	**0.0000**	**0.595**	**0.0398210**	**0.0000**
	MCARI/OSAVI	0.524	0.0479100	0.0000	0.538	0.0534560	0.0000
	CI-green	0.270	0.0593016	0.0000	0.278	0.0592158	0.0000
	CI-rededge	0.281	0.0588529	0.0000	0.288	0.0586235	0.0000
			Linear			Quadratic	
Chlorophyll Canopy	NDVI	0.428	0.7287620	0.0000	0.470	0.7045450	0.0000
	EVI	0.628	0.5878550	0.0000	0.700	0.5300250	0.0000
	WDVI	0.659	0.5630450	0.0000	0.718	0.5139270	0.0000
	REP	0.726	0.5044690	0.0000	0.739	0.4944780	0.0000
	TCARI/OSAVI	0.291	0.8111600	0.0000	0.319	0.7992010	0.0430
	MCARI/OSAVI	0.214	0.8540420	0.0000	0.251	0.8379280	0.0260
	CI-green	0.769	0.4628310	0.0000	0.771	0.4632210	0.3670
	CI-rededge	**0.789**	**0.4421380**	**0.0000**	**0.814**	**0.4171770**	**0.0000**

Leaf chlorophyll was selected due to its good relationship with nitrate concentration for all observations over the growing season, whilst the chlorophyll canopy was expected to give a good representative of the nitrogen status for the canopy level. A relationship analysis between the eight VIs and leaf chlorophyll content was done twice. The first analysis included all observation dates, the results of which can be seen in Appendix 3. In the second

analysis, the relationships were calculated without including the observations on 6 June and 14 June 2013. The exclusion of these observations was due to the values in the beginning of the growing season being considered to have a different pattern than the remaining observations. The same regression analyses were also applied twice in calculating the relationships between the chlorophyll canopy and the eight vegetation indices.

The scatter plot in Appendix 3 shows that the NDVI, EVI and WDVI were bad indices to determine the chlorophyll at the leaf level. These indices were easily saturated in the high values which made them insensitive to the crop development. The REP, CI_{green}, and $CI_{red\text{-}edge}$ were also found insignificant to the leaf chlorophyll content, which can be seen from their low R^2 values and unknown pattern in the scatter plot (Appendix 3). The low relationship between leaf chlorophyll and the three indices (REP, CI_{green}, and $CI_{red\text{-}edge}$) were suspected to be due to the influence of the observation on 6 June and 14 June 2013. The values from these dates were not following the overall pattern. The scattering points observed from these dates were most likely the results of tissue age and irradiance during plant growth. Therefore, the second regression analyses were made to analyse the relationships of the leaf chlorophyll with all indices without the influence of the first two observation dates. However, the R^2 values from the second analysis show an increasing result for REP, CI_{green}, and $CI_{red\text{-}edge}$. Conversely, R^2 values from the other three vegetation indices (NDVI, EVI and WDVI) were decreasing. Figure 23 shows that the REP, CI_{green}, and $CI_{red\text{-}edge}$ have a linear relationship with the leaf chlorophyll in the second regression analysis. On the contrary, both the MCARI/OSAVI and TCARI/OSAVI have non-linear relationships with the leaf chlorophyll content, either in the first or the second regression analysis. Based on the summary from Table 7, the TCARI/OSAVI has the strongest relationship with leaf chlorophyll both in the first and the second regression analyses. The R^2 and S-value for the relationship between the TCARI/OSAVI and leaf chlorophyll are 0.517 and 0.055, respectively (using exponential regression with all observations included). The R^2 value and S-value for the relationship between the TCARI/OSAVI and leaf chlorophyll excluding the observation on 6 June and 14 June were 0.595 and 0.0398. The MCARI/OSAVI has the second strongest relationship with the leaf chlorophyll.

The same procedures were applied to find the best relationship between the chlorophyll canopy and the eight vegetation indices. Among the eight vegetation indices, five indices (EVI, WDVI, REP, CI_{green} and $CI_{red\text{-}edge}$) were shown to have a significant linear relationship with the chlorophyll canopy (Appendix 4). In the second regression analysis, most of the R^2 values increased and the S-values were smaller than in the first analysis. $CI_{red\text{-}edge}$ had the strongest relationship with the chlorophyll canopy which is shown by its relatively high R^2 (0.858) and the lowest S-value (0.4080120) compared to the other seven VIs (Table 7). The second best VI, both from the first and second regression analysis results, was CI_{green}.

The exclusion of the observation on 6 June and 14 June 2013 changed the R^2 values, but it did not change the results of the best VI for both the leaf chlorophyll and the chlorophyll

canopy. The VI that had the strongest relationship with the leaf chlorophyll or chlorophyll canopy will be used in the next phase to generate the reference curve for potato growth. In this report, the TCARI/OSAVI showed the strongest relationship with leaf chlorophyll, whilst $CI_{red\text{-}edge}$ showed the strongest relationship towards chlorophyll canopy. Therefore, these two VIs will be used in further phases.

4.2 Optimal growth curve for potato crop (Phase 2)

Based on the result from Phase 1, the TCARI/OSAVI and $CI_{red\text{-}edge}$ were proven to give a good representation of the N potato status in the leaf and canopy level respectively. Therefore, the values from these two indices will be used to construct the optimal growth curve, also known as a reference curve, over the growing season. In this phase, the reference curve will be calculated using two different approaches: the maximum yield and the mean curve approach. In the beginning of the research, an optimum quality approach was considered as one of the methods to generate the reference curve. The optimum quality of the potato is the equilibrium point which meets the best value of the yield harvester and the potato's under water weight (OWG). However, the available OWG data are only available for 8 experimental plots (Table 8). Therefore, the optimum quality approach was taken out from this research report.

The two reference curves from this phase (based on the maximum yield and the mean curve approach) will be used in Phase 3 as reference curves to calculate the difference (deviation) between each experimental plot for the TCARI/OSAVI and $CI_{red\text{-}edge}$ using similarity measures. However, only the mean curve approach will be used in Phase 3 to perform the similarity measures in the subplots level. The difference of each experimental plot curve or subplot curve compared to the reference curve will give appropriate information to the farmer regarding potato growth development.

4.2.1 Maximum yield approach

The maximum yield approach was selected as one of the reference curve construction approaches due to that its results are normally expected at the end of the growing season. Based on the yield harvester map in Appendix 5, the average yield harvester values for each experimental plot were calculated using R and ArcMap 10.2 (the tractor path was excluded from the calculation). As shown in Table 8, the maximum yield was produced by Plot C (73,887.89 kg/ha). Therefore, Plot C will be used as the reference curve based on the maximum yield approach.

In section 4.3.1, the TCARI/OSAVI and $CI_{red\text{-}edge}$ values over the growing season from each experimental plot will be compared to the values from Plot C. The deviation or the difference between each plot from its reference plot (Plot C) will be calculated using similarity measures for time series analysis as explained in section 3.3.3. Experimental plots

plot that have a growth problem are expected to deviate far from the reference plot over the growing season.

Table 8: Yield harvester for 12 experimental plots

Experimental Plot	Initial Fertilizer	Total Fertilizer	OWG TTW 5/9	Yield harvester	
	kg/ha	kg/ha		ton/ha	kg/ha
A	252	326.7	401.82	72.57	65,833.98
B	0	108.5	434.32	70.33	63,800.08
C	162	244.2	391.38	81.45	73,887.89
D	90	166.4	418.47	73.20	66,403.55
E	252	284.0	403.17	71.73	65,068.33
F	0	32.0	431.32	58.71	53,262.33
G	162	194.0	394.74	71.95	65,269.66
H	90	122.0	411.38	65.38	59,309.26
I	252	329.3	n.a.	78.36	71,087.53
J	0	96.4	n.a.	66.30	60,146.35
K	162	254.2	n.a.	81.31	73,765.82
L	90	175.2	n.a.	74.44	67,534.92

Note: 1 ton/ha = 907.18474 kg/ha (Source: Google). OWG= Potato's Under Water Weight. OWG represents the quality of the potato.

4.2.2 Mean curve approach

As seen in Figure 13, the reference curve based on the mean curve approach was established by calculating the average value from 24 experimental subplots on each observation date.

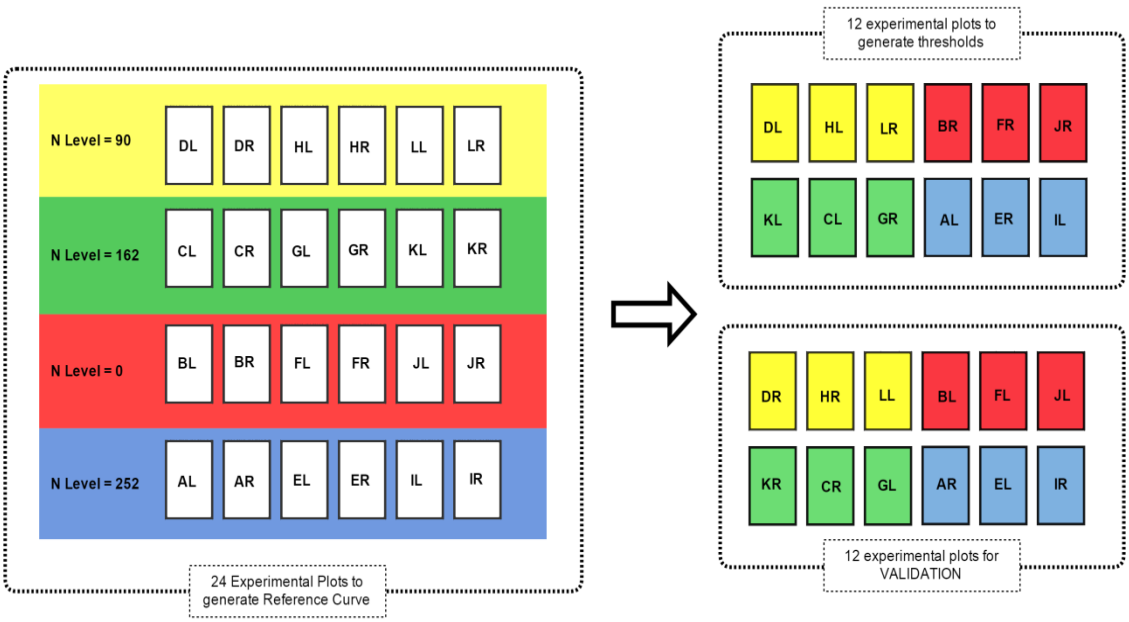


Figure 25: Reference curve based on the mean curve approach design (left). Calibration and validation subplot dataset design (right).

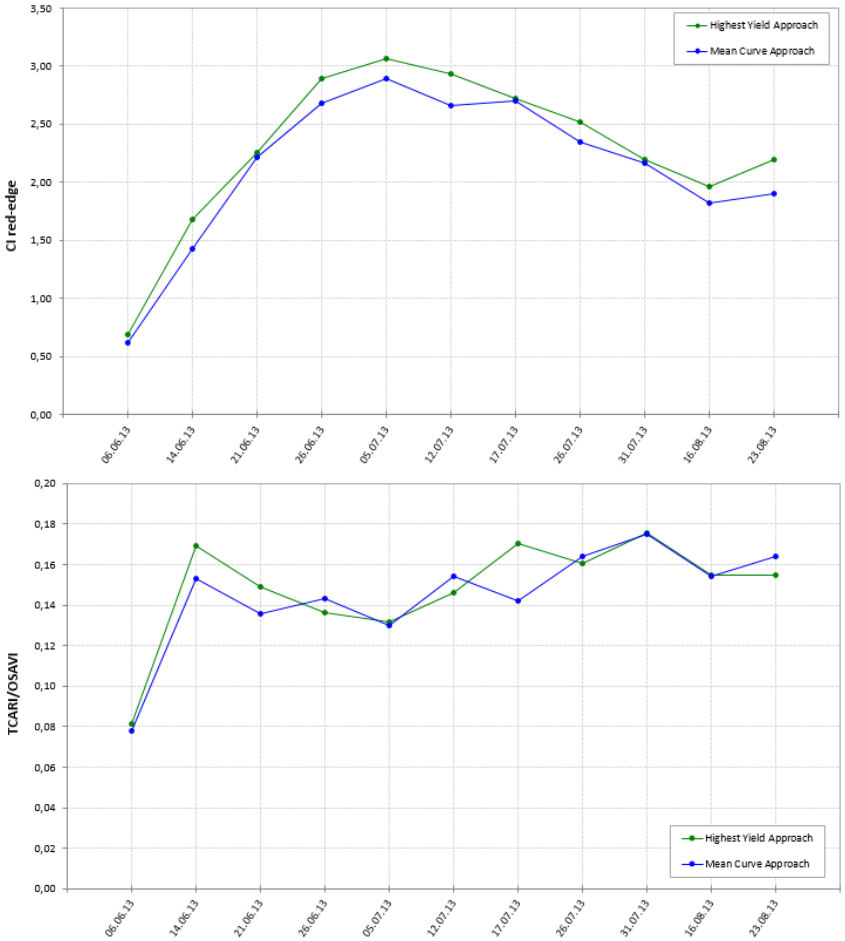


Figure 26: Optimum growth curve for potato using $CI_{red\text{-}edge}$ (top). Optimum growth curve for potato using TCARI/OSAVI (down). Optimum growth curve with highest yield approach (green plot) and mean curve approach (blue plot).

In section 4.3.2, the TCARI/OSAVI and $CI_{red\text{-}edge}$ values over the growing season from 12 experimental plots (A through L) will be compared to the reference curve from the mean curve approach (Figure 26). The mean curve approach will also be used in section 4.3.3 to perform similarity measures on the subplots level using 12 experimental subplots from the calibration dataset (Figure 25). The deviation or the distance difference between each plot and subplot from the reference plot will be calculated using similarity measures for time series analysis as explained in section 3.3.3.

4.3 Time series similarity measures for deviation detection (Phase 3)

Performing similarity measures to the temporal series of VIs from the experimental plots is an important phase when deriving the alerting service. The two different reference curves which was established during Phase 2 were simulated as the optimum potato growth curve. For Phase 3 four different time series similarity measures were conducted to calculate the distance difference between each experimental plot and subplot relative to the reference curve. Since there are two reference curve approaches, the analysis at the experimental plots level (i.e.: plot A, B, C) will be explained for each approach (Section 4.3.1 and Section 4.3.2), while the analysis at the subplots level (i.e.: plot AR, BL or CR) will be made using the mean curve approach (Section 4.3.3). Plots or subplots that have severe differences or

deviations from the reference plot are assumed to experience growth related issues due to the improper amount of N in the crops. To derive the alerting service, the distance values established during this phase will be plotted in the next phase by using the control chart theory.

4.3.1 Similarity measures using maximum yield approach in the plot level

There were four similarity measures (D_{Man}, D_E, D_{CC}, and RMSD) evaluated in this phase to investigate whether the four distinct initial fertilization levels applied to the experimental plots resulted in a significant temporal profile difference. Two VIs ($CI_{red\text{-}edge}$ and TCARI/OSAVI) were used as the time series input data. The $CI_{red\text{-}edge}$ values were used in the similarity measures as the representation of the N status in the canopy level whilst the TCARI/OSAVI was assumed to give a good representation of the N status at the leaf level. Therefore, the analysis was done separately.

$CI_{red\text{-}edge}$ (N status indicator in the canopy level)

Appendix 6 and Figure 27 show the results of the time series similarity measures using the Manhattan distance (D_{Man}) and the Euclidean distance (D_E), where both measures were derived from the Minkowski distance. There were relatively small differences detected between the $CI_{red\text{-}edge}$ values from the reference plot (Plot C) compared with the other 11 experimental plots in the beginning of the growing season (Appendix 6). However, each plot started to deviate away from the reference curve along with the crop development over the growing season. This was indicated by the increase of the D_{Man} values starting from the fourth observation (26 June 2013). The highest deviation differences were clearly seen from the plots with an initial fertilization level of 0 kg N h^{-1}. Starting from the fourth observation (26 June 2013), plots with an initial fertilizer level of 0 kg N h^{-1} (plot F, J, and B) were deviating away from the reference plot. The sudden increase in the distance difference from these plots occurred on the sixth observation (12 July 2013). The distance values for plot F, J, and B were relatively similar at the 5^{th} and 6^{th} observation. However, at the end of the growing season, plot F had the highest deviation distance to the reference plot, followed by plot J and H. As shown in the case by D_{Man}, plot F had the highest distance value. This was due to the two additional fertilizer applications that was applied to plot J and B on 5^{th} and 18^{th} of July. Furthermore, the additional fertilizer applications caused difference reductions for plot J and B. As a consequence, plot B and J show a less deviating curve compared with plot F at the end of the growing season.

It was expected that the distances between the plots compared to their reference plot (Plot C) would depend on their fertilization levels at the beginning of the growing season. Based on the same initial fertilizer level (252 kg N h-1) plot A, E, and I were expected to have closely related D_{Man} values over the growing season. However, plot E showed close distances to reference plot on the first five observations (6 June – 5 July), whilst plot A and I deviated slightly from the reference plot.

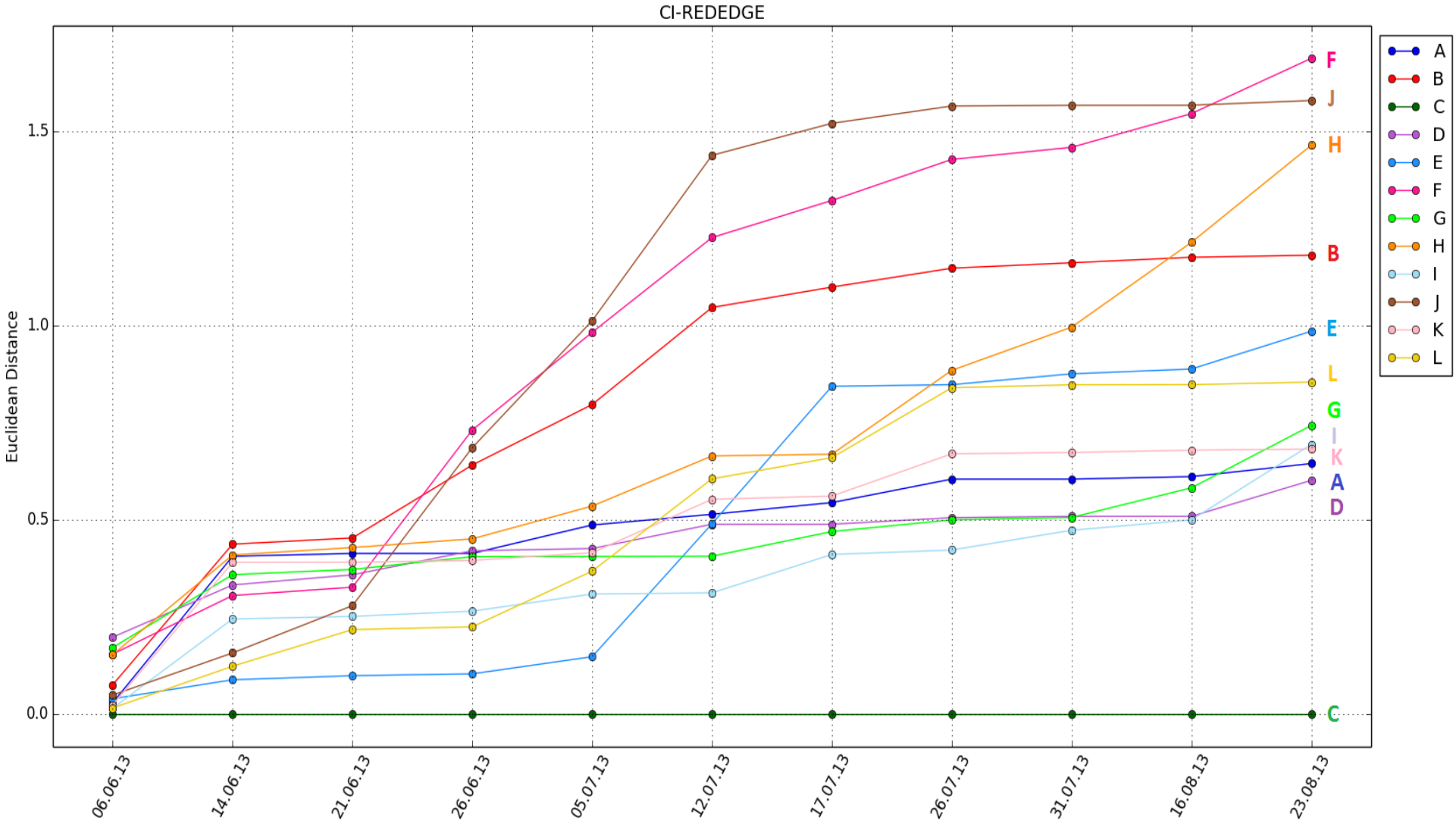


Figure 27: $CI_{red\text{-}edge}$ time series similarity measures using Euclidean Distance method. Plot C was assigned as the reference plot (maximum yield approach). The X-axis shows weekly observations over the growing season. The Y-axis shows the Euclidean distance between the experimental plots and the reference plot (Plot C) for each observation date.

At the end of the growing season, the distance order was changed due to the additional fertilizer applied to plot A and I on 18 July 2013. Plot A and I deviated less from the reference curve compared to plot E at the end of the growing season. As for plot G and K, these deviated moderately from plot C in the beginning of the growing season. Even though they had a similar initial fertilization level (162 kg N ha-1) with reference plot C, the two plots showed a different growth pattern with this plot from 6 June until 5 July 2013. However, starting from the sixth observation, plot G and K had the least distance (a more similar growth pattern) from plot C as expected.

The time series similarity measures using D_E shows a better illustration of the crop condition over the growing season (Figure 27). Differing from D_{Man}, D_E was able to capture the distance variations from each plot starting from the first observation. Even though the order of the distance was not based on the initial fertilizer level, the D_E values clearly show that each plot performs slightly different from the reference plot on the first observation on 6 June 2013. The distances for plot B, F and J from the reference plot C increased significantly on the fourth observation (26 June 2013). It is clearly seen that plots with an initial fertilizer level of 0 kg N h^{-1} did not perform as well with plots that received initial fertilizers. To pursue the ideal crop development, which is indicated by a small distance difference to the reference plot, the first additional fertilizer were applied to plot B and J with amounts of 29 kg N h^{-1} and 22 kg N h^{-1} respectively on 5 July 2013. The second additional fertilizer was applied on 18 July 2013 to plot B and J. However, these two plots were unable to keep up with an optimal crop development. This resulted in the high distance differences between reference plot C with plot B and J at the end of the growing season. The significant distance difference of plot B, F and J to its reference curve were positively related with the yield production data in Table 8. Based on Table 8, these three plots produced the lowest yield at the end of the growing season.

The order of the distance difference for all the plots at the end of the growing season were well explained by D_E measure. Based on the information of management activities in Table 2, there were additional fertilizer applications applied to eight plots (A, B, C, D, I, J, K and L) on 18 July 2013. The effect of the fertilizer applications can be clearly seen in Figure 27. At the end of the growing season, the plots that did not get an additional fertilizer application (E, F, G and H) on 18 July 2013 show higher deviation from the reference plot compared with the other plots in the same initial fertilization group. As an example, plot H had a higher distance difference compared with the other two plots in its group (plot D and plot L) which had the same initial fertilization level of 90 kg N h^{-1}.

Appendix 7 shows the result of the time series similarity measures using the correlation coefficient distance (D_{CC}). Differing from the other three similarity measures used in this report (D_{Man}, D_E and RMSD), D_{CC} values ranged from 1 to -1. A D_{CC} value that is close to 1 gives an indication of an increasing linear relationship, whereas a value of -1 indicates a decrease in the linear relationship (Lhermitte et al., 2010). In the beginning of the growing

season (06 June 2013), all plots showed linear relationships to the reference plot C. Starting from the third observation (21 June 2013), plot F had a decreasing relationship with the reference plot C, but showed improvements starting from the fifth observation. At the end of the growing season, plot F had a relatively good relation with plot C. This result was quite opposite with the result from the other three similarity measures, where plot F had a relatively high difference with plot C at the end of the growing season. Plot B and J started to decrease on the sixth observation (12 July 2013) and kept decreasing until the end of the growing season. On 26 July 2013 (the 8th observation), the plots were mainly formed into three groups. The first group consisted of plot D, G, A and I (all of which have the closest positive relationship with reference plot C). The second group consisted of plot F, K, L and E; and the last group consisted of plot B, J and H. The order of the plots in terms of the distance difference from the reference plot was slightly different from the other three similarity measure methods. The most distinct difference between the D_{CC} approach and the other three distance measures was that D_{CC} was not able to detect plot F as one of the most deviating plots from reference plot C. The result of the time series similarity measure using the Root Mean Square Distance (RMSD) method can be seen in Appendix 8. RMSD was able to detect crop development changes for each plot based on the different initial fertilizer levels and management activities (additional fertilizer application and irrigation). Similar with D_E, the RMSD method was able to detect the distance variability of all plots starting from the first observation. The order of the RMSD distance difference between the plots and reference plot C throughout the growing season were similar to the Euclidean distance. The effects of the additional fertilizer application applied to the eight experimental plots on 18 July 2013 were also visible in the end of growing season.

TCARI/OSAVI (N status indicator in the leaf level)

As described in the beginning of section 4.3, the time series similarity measures would be carried out two times by using the $CI_{red\text{-}edge}$ and TCARI/OSAVI time series data. Appendix 9 shows the time series similarity measure performed using D_{Man}. The D_{Man} differences between the 11 experimental plots to the reference plot C were relatively small in the beginning of the growing season. In the case of D_E and RMSD, the distance differences were more significant in the beginning of the growing season (Figure 28 and Appendix 9). Since the four distinct initial fertilizers were applied to the soil during the planting time, the clear distance differences from each plot were expected to start from the first observation (6 June 2013). Based on the distance difference clear illustration given in the beginning of the growing season, D_E and RMSD were more suitable to characterize the growth status over the growing season. Differing from the results of the similarity measures using the $CI_{red\text{-}edge}$ time series, Plot F and J did not appear to have the highest distance differences with the reference plot C at the fourth observation using the TCARI/OSAVI time series data. Figure 28 shows that starting from the third observation (21 June 2013); plot A, plot B, plot E and plot K had the highest distance differences compared to the reference plot C.

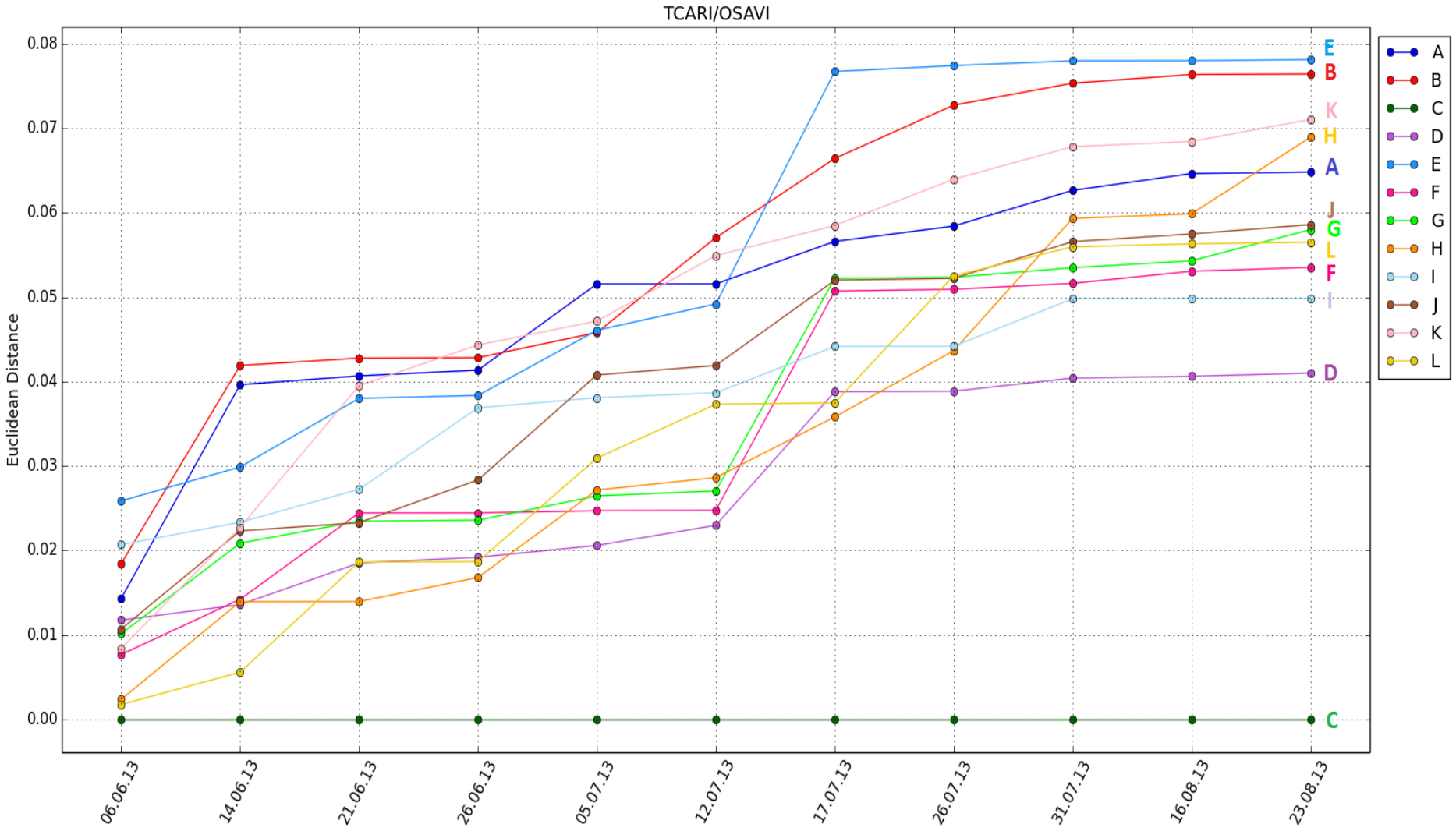


Figure 28: TCARI/OSAVI time series similarity measures using Euclidean Distance method. Plot C was assigned as the reference plot (maximum yield approach). The X-axis shows weekly observations over the growing season. The Y-axis shows the Euclidean distance difference between the experimental plots and the reference plot (Plot C) for each observation date.

The high deviation values between plot A, B, E and K compared to the reference plot C continued until the end of the growing season. The same pattern was found using the D_{CC} method (Appendix 9). As seen in Appendix 9, the correlation between plot F, B and K to the reference plot C started to decrease as early as on the second observation date (14 June 2013). However, the relation between plot K and reference plot C started to improve on the ninth observation date. The correlation between plot E and the reference plot C decreased steeply on the seventh observation date (17 July 2013). On the sixth observation date a steep decrease also occurred for plot B.

The plots that had similar TCARI/OSAVI values with the reference plot C in the beginning of the growing season were plot H, K and F (Figure 28). However, plot L started to deviate away from the reference plot on the fifth observation (5 July 2013), whilst plot H started to deviate at the end of the growing season. The effects of the additional fertilizer application on 5 July 2013 applied to plot B and J was not seen in the similarity measures using D_E. However, on the eight observation date (26 July 2013) they could be slightly seen using the D_{CC} method. On the eight observation and onwards, Plot E, F, G, and H clearly showed sharp deviations from the other plots as an effect from the additional fertilizer application applied on 18 July 2013. Even though TCARI/OSAVI was assigned as the best representative VI for leaf chlorophyll (R^2=0517), this VI did not perform as well as $CI_{red\text{-}edge}$. This fact was proven by the inability of the TCARI/OSAVI to detect the high deviation of plot F and J as the $CI_{red\text{-}edge}$ did. The TCARI/OSAVI also detected high distance values from plots that had a relatively high initial fertilizer level, such as plot K. This information may indicate that the TCARI/OSAVI could possibly lead to false alerts in Section 4.4.

4.3.2 Similarity measures using mean curve approach in the plot level

The second similarity measures were applied using the mean curve approach. Instead of choosing one of the experimental plots as the reference plot, this method calculated the average (mean value) from all 24 experimental subplots (Figure 13) for each observation date to construct it. Based on the explanation in section 4.2.2, the 12 experimental plots were compared to the mean curve (reference plot) using four similarity measure methods (D_{Man}, D_E, D_{CC}, and RMSD). In section 4.4, the distance values from these 12 experimental plots will be used to set up the control limits in the control chart. At the end, the approach to generate the control limits (the thresholds of alerting service) will be tested using the same 12 experimental plots from the STR datasets.

Mean curve approach with $CI_{red\text{-}edge}$ time series data

Figure 29 shows the time series similarity measures performed with D_E for the 12 experimental plots (A, B, C, D, E, F, G, H, I, J, K, and L). Similar with the previous analysis using the maximum yield approach in section 4.3.1, D_E and RMSD were able to detect the distance differences variability in the beginning of the growing season (6 June 2013), while the results from the D_{Man} approach showed relatively similar distance values from all plots

on 6 June (Appendix 10). Plot J and F (initial fertilizer of 0 kg N ha^{-1}) deviated significantly from the reference plot starting from the fourth observation (26 June 2013), whilst with D_{Man}, these two plots started to deviate significantly on the fifth observation (5 July 2013). However, even though D_{Man} performed slightly worse than RMSD and D_E in the beginning of the growing season, the range of the distance difference values were bigger (0 – 3.5) than D_E (0 – 1.2) or RMSD (0 – 0.20). Differing from the result using the maximum yield approach (Figure 27), plot B (initial fertilizer level of 0 kg N ha^{-1}) did not deviate significantly from the reference plot on the second observation (14 June 2013). On the contrary, Plot C (initial fertilizer of 162 kg N ha^{-1}) deviated the most from the reference plot on the second observation. However, Plot C started to perform better from the fifth observation (5 July 2013) and onwards. This development was most likely due to the additional fertilizer of 50.2 kg N ha^{-1} applied to plot C on 18 July 2013. As the crops in plot C reacted to the fertilizer application the distance differences to the reference plot started to decrease from 26 July 2013 and onwards.

In the previous result (Figure 27), plot A performed generally well against the reference plot C with the maximum yield approach. Even though the crop development in plot A was not completely similar to the reference plot C, the distance differences between plot A and the reference plot C were relatively low over the growing season. The time series similarity measure performed by D_E showed a different result: Plot A was significantly deviating from the reference plot starting from the fifth observation (Figure 29). On that observation date (5 July 2013), the distance difference of plot A from the reference plot was even higher than plot B (initial fertilizer of 0 kg N ha^{-1}). However, plot A performed slightly better after the fertilizer application applied on 18 July 2013. Plots with an initial fertilizer of 162 kg N ha^{-1} (plot C, G, and K) showed relatively similar crop developments with the reference curve over the growing season; their D_E values or RMSD were relatively low over the temporal development (Figure 29 and Appendix 10). The effects of the second fertilizer application (18 July 2013) were clearly seen in the plots with an initial fertilizer of 90 kg N ha^{-1}. On the ninth observation and onwards, plot H deviated away from the reference plot. This is shown by the high distance difference from plot H to the reference plot on 31 July 2013 and onwards; whilst the distance difference values of plot D and L were relatively much lower than plot H.

The time series similarity measures performed by D_{CC} can be seen in Appendix 10. D_{CC} was able to emphasize low relationships between two plots with an initial fertilizer of 0 kgNha^{-1} (Plot F and B) to the reference plot. As we can see in Appendix 10, the relationship of plot F and B to the reference plot started to significantly decrease on the second observation date (14 June 2013) until the end of the growing season, while the other plot with an initial fertilizer of 0 kg N ha^{-1}, plot J, had a good relationship with the reference plot until the fourth observation. However, this similarity measure was not clearly able to detect the expected decrease in the relationship between plot H and the reference plot.

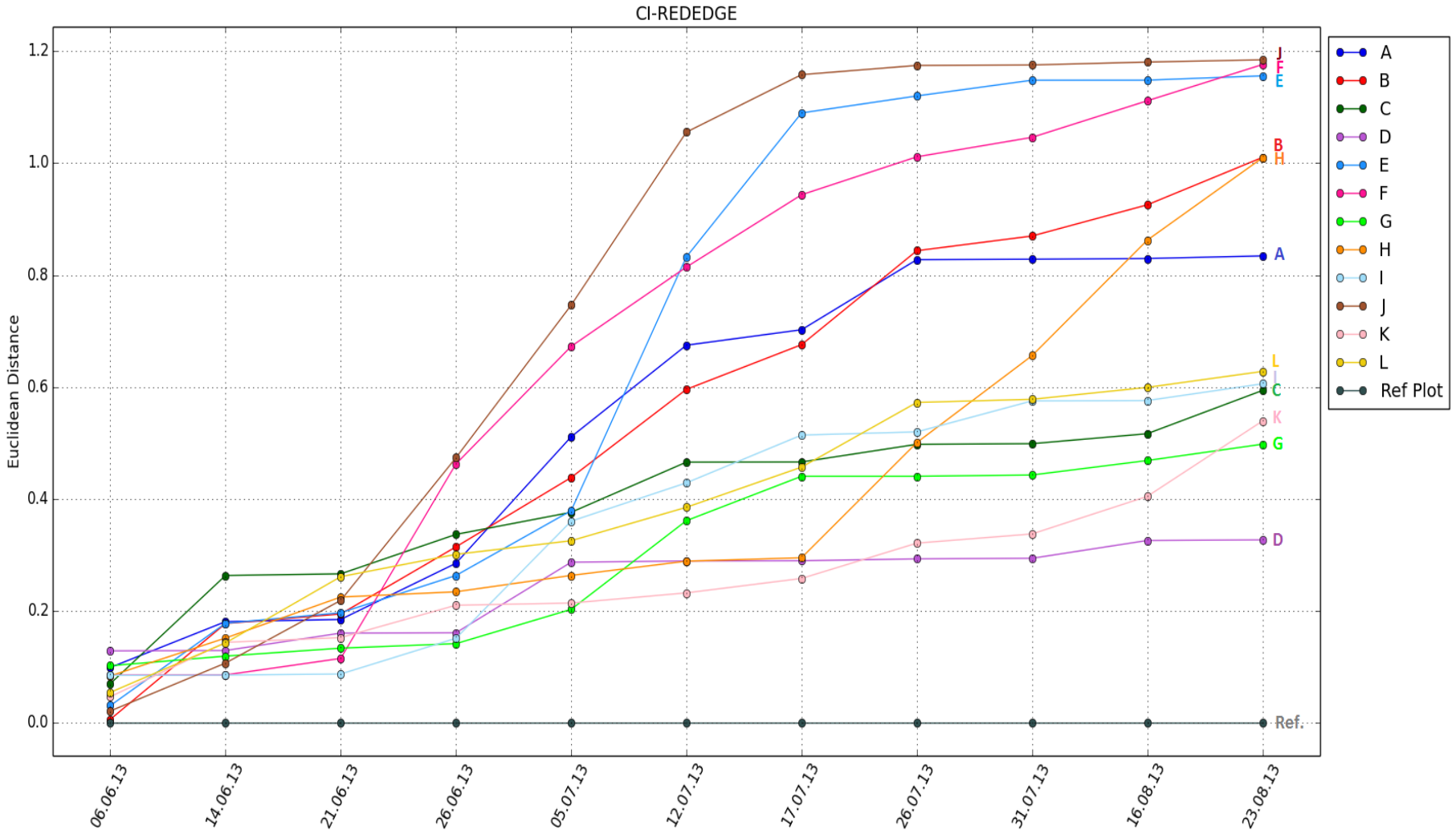


Figure 29: $CI_{red\text{-}edge}$ time series similarity measures using Euclidean Distance method for the plot level. Mean curve was assigned as the reference plot. The X-axis shows weekly observations over the growing season. The Y-axis shows the Euclidean distance between the experimental plots and the reference plot for each observation date.

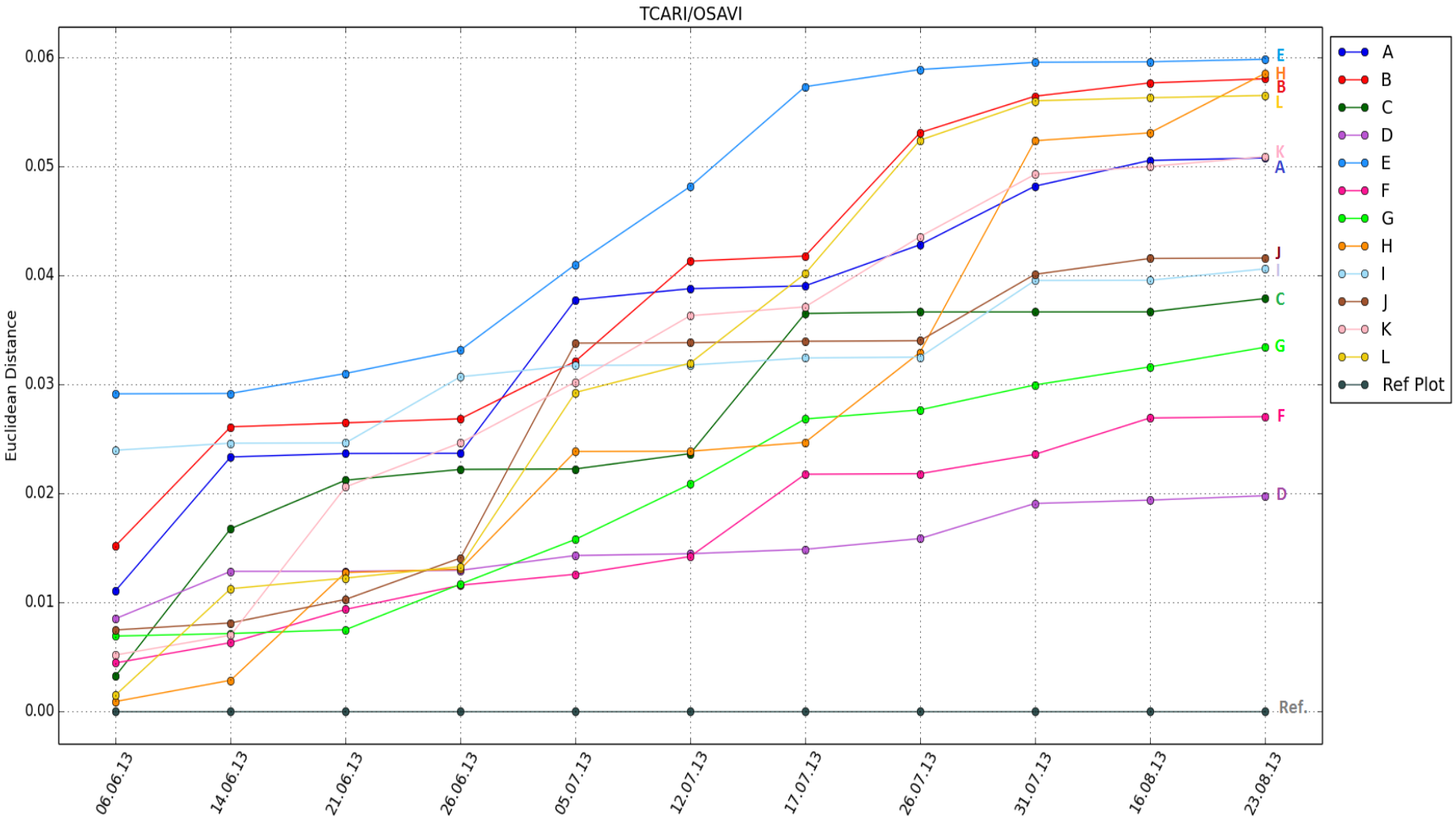


Figure 30: TCARI/OSAVI time series similarity measures using Euclidean Distance method for the plot level. Mean curve was assigned as the reference plot. The X-axis shows weekly observations over the growing season. The Y-axis shows the Euclidean distance between the experimental plots and the reference plot for each observation date.

Mean curve approach with TCARI/OSAVI time series data

Figure 30 shows the time series similarity measure performed with D_E using the mean curve approach as the reference curve. This approach gave essentially the same portrait as the one given by the maximum yield approach (Figure 28). The TCARI/OSAVI was not precisely able to differentiate the growth development of the experimental plots based on the initial fertilizer level. This can be seen from the time series similarity measures in Figure 30 and Appendix 11. From Figure 30, we can see that the distance difference values from the plots with an initial fertilizer of 0 $kgNha^{-1}$ (plot B, plot F, and plot J) were not located close to each other; there were no specific deviation pattern. These three plots were expected to have approximately similar distances from the reference plot. However, the TCARI/OSAVI was only able to detect the growth problem, which was indicated by the high distance difference values, from plot B; whilst plot F's distance difference values were relatively low over the growing season. This pointed out one of the evidences that TCARI/OSAVI was not able to perform a good potato growth detection.

4.3.3 Similarity measures using mean curve approach in the subplots level

For the subplots level, the similarity measures were applied to both the $CI_{red\text{-}edge}$ and TCARI/OSAVI time series using the mean curve approach as the reference plot. Based on the explanation in section 4.2.2, the 12 experimental subplots from the calibration dataset (plot AL, BR, CL, DL, ER, FR, GR, HL, IL, JR, KL, and LR) were compared to the mean curve (reference plot) using the four similarity measure methods (D_{Man}, D_E, D_{CC}, and RMSD). Later, in section 4.4.3, the distance values from these 12 experimental subplots will be used to set up the control limits in the control chart. At the end, the approach to generate the control limits (the thresholds of the alerting service) will be tested in Section 4.5.3 with another 12 experimental subplots from the validation dataset (plot AR, BL, CR, DR, EL, FL, GL, HR, IR, JL, KR, and LL).

Mean curve approach with $CI_{red\text{-}edge}$ time series data

Figure 31 shows the time series similarity measures performed by D_E for the 12 chosen subplots (AL, BR, CL, DL, ER, FR, GR, HL, IL, JR, KL, and LR). Similar with the previous analysis using the maximum yield approach in section 4.3.1, D_E and RMSD were able to detect the distance differences variability in the beginning of the growing season (6 June 2013), while the results from the D_{Man} approach showed relatively similar distance values from all plots on 6 June (Appendix 12).

Plot FR (initial fertilizer of 0 kg N ha^{-1}) started to deviate away from the reference curve on the fourth observation until the end of the growing season. The other two subplots with similar initial fertilizers, plot BR and plot JR, deviated significantly from the reference plot starting from the sixth observation (Figure 29). Differing from the results shown in Section 4.3.1, plot AL, the left side of plot A, showed high distance difference values from the

reference plot starting from 5 July until the end of the growing season. This result was not a form of false detection from the Euclidean measure approach. Yet, the result was related to the underdevelopment of plot AL which was proven by the low yield production at the end of the growing season (Appendix 5). On the ninth observation (31 July 2013), there were two distinct groups seen in Figure 29: group one consisted of plot LR, IL, GR, CL, KL, and DL; while the second group consisted of plot ER, BR, JR, HL, AL and FR. The effects of the additional fertilizer application were slightly visible with the D_E and RMSD approach. As an example: plot DL and LR showed smaller deviations to reference plot compared with plot HL starting from 31 July 2013 (2 weeks after the additional fertilizer application was applied).

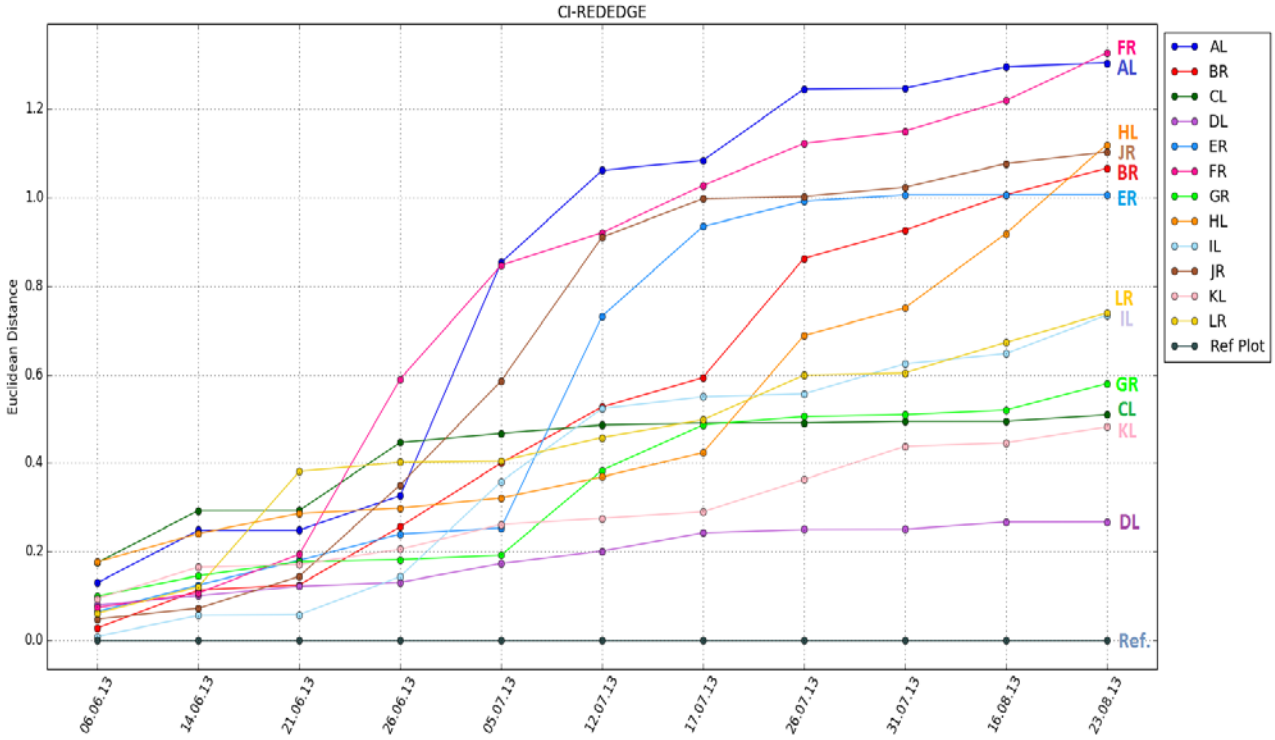


Figure 31: $CI_{red-edge}$ time series similarity measures using Euclidean Distance method for the subplot level. Mean curve was assigned as the reference plot. The X-axis shows weekly observations over the growing season. The Y-axis shows the Euclidean distance between the experimental plots and the reference plot for each observation date.

The time series similarity measure performed with D_{CC} can be seen in Appendix 12. In the beginning of the growing season, all twelve subplots show linear relationships to the reference plot (the mean curve approach). On the sixth observation date (12 July 2013), the relationship between plot JR and the reference plot decreased and continued over the temporal development. On the same observation date a major decrease was also found in the relationship between plot BR and the reference plot. However, the D_{CC} method could not detect the expected decrease in the relationship between plot FR and the reference plot as D_E and RMSD did. This draws the conclusion that D_E and RMSD are the best similarity measures to detect the difference between each plot to the reference plot (both approaches).

Mean curve approach with TCARI/OSAVI time series data

In Section 4.3.1, we saw that performing time series similarity measures on the TCARI/OSAVI time series data gave undesired results. Differing from the results using the $CI_{red\text{-}edge}$ data, some plots that had a low initial fertilizer level did not significantly deviate from the reference plot C (the maximum yield approach) as expected. In this section, four similarity measures were performed on the TCARI/OSAVI data from 12 subplots (AL, BR, CL, DL, ER, FR, GR, HL, IL, JR, KL, and LR).

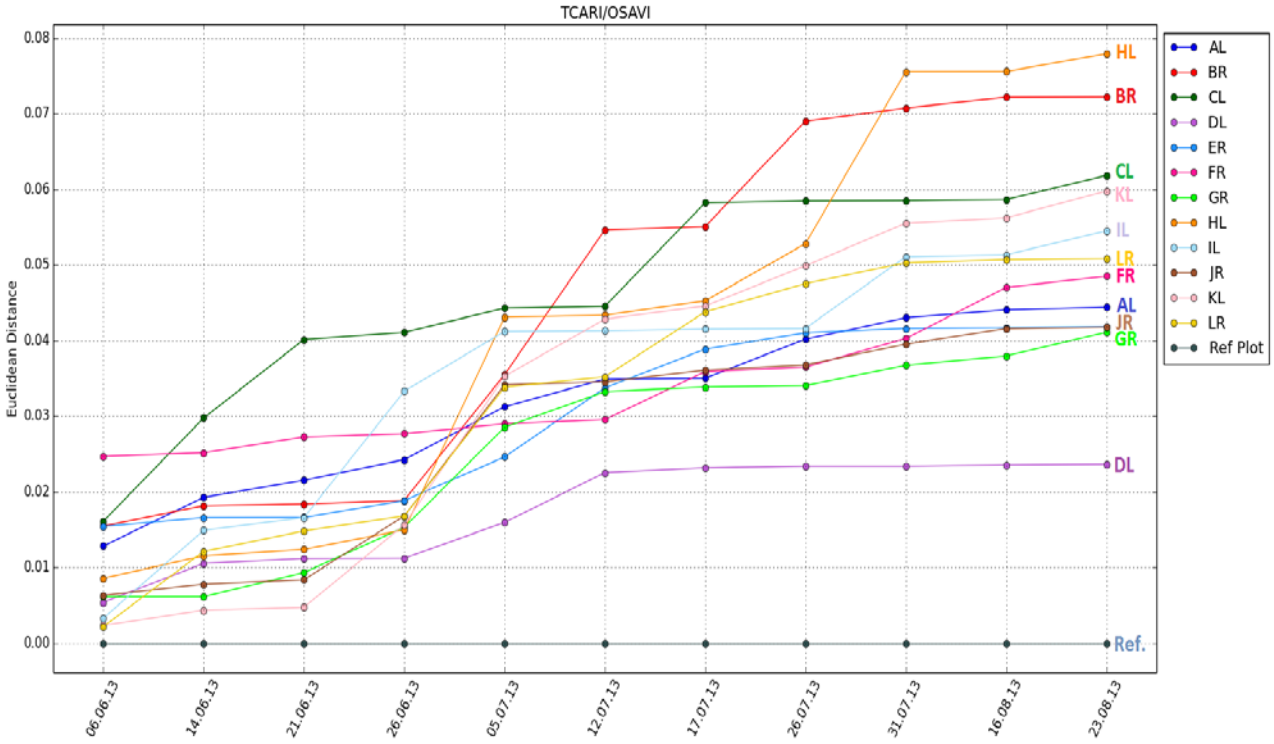


Figure 32: TCARI/OSAVI time series similarity measures using Euclidean Distance method for the subplot level. Mean curve was assigned as the reference plot. The X-axis shows weekly observations over the growing season. The Y-axis shows the Euclidean distance between the experimental plots and the reference plot for each observation date.

Figure 32 shows the time series similarity measures performed with D_E. In the beginning of the growing season, the distance difference from plot FR (initial fertilizer of 0 kg N ha^{-1}) to the reference plot was relatively high. Over the temporal development, the distance difference values showed significant improvements: plot FR did not deviate much from the reference plot, starting from the sixth observation until the end of the growing season. Similar with plot FR, the distance differences of plot JR (initial fertilizer of 0 kg N ha^{-1}) to the reference plot were relatively medium over the growing season and considered as low at the end of the growing season. These results, once again, show the inability of TCARI/OSAVI to represent the potato growth condition. The plots that should have deviated away from the reference plot (i.e.: FR and JR) did not perform as they were supposed to, whilst the plot that had a high initial fertilizer level (i.e.: plot KL) tended to have highly deviating values from the reference plot starting from the fifth observation (Appendix 13).

4.4 Plant growth alerting service using control chart theory (Phase 4)

In this phase, the statistical control charts theory was used to determine the plant growth alerting service. Based on the theory of control charts, a set of thresholds was constructed to detect a so called "out of control process". In this report, the control limits of the control chart were computed as $\mu + 3\sigma$ (UCL) and $\mu - 3\sigma$ (LCL), whilst the warning limits were calculated as $\mu \pm 2\sigma$ (UWL and LWL). Several dataset combinations were tested to find the best observation date for constructing the control chart thresholds (limits). There were three main different alerting services constructed in this section: an alerting service based on the maximum yield approach for the plot level, an alerting service based on the mean curve approach for the plot level and an alerting service based on the mean curve approach for the subplot level. Therefore, the results from each alerting service are shown separately.

4.4.1 Plant growth alerting service for maximum yield approach (Plot level)

Time series data from two VIs were used in the previous section to calculate the distance difference between each plot to the reference plot. Because of this, two control charts were established in this section: a control chart that was based on the distance values from the $CI_{red\text{-}edge}$ time series data and one that was based on the distance values from the TCARI/OSAVI time series data. Based on the analysis in Section 4.3, there were two similarity measures that represented the growth deviation detection of each plot compared to the reference plot the best over the growing season: D_E and RMSD. However, D_E was not used to generate the alerting service in this section.

4.4.1.1 Alerting service for Euclidean distance values from $CI_{red\text{-}edge}$ time series

Figure 33 shows the alerting service for the first six observation dates using D_E values with control chart theory. The D_E values from the first four observation dates (6 June, 14 June, 21 June and 26 June) were used to construct the thresholds (limits) of the control charts. The yellow lines in Figure 33 and Figure 34 indicate the warning limits which were calculated using $\mu \pm 2\sigma$; whilst the red lines (UCL and LCL) were calculated using $\mu \pm 3\sigma$. Distance values from the plots that were located above the upper warning limit or below the lower warning limit would receive an alert. The alert would then be considered as growth problem indication within the experimental plot. If the distance value from a plot was located above the upper the UCL or below the LCL, then the plot would be classified as "out of control".

With the thresholds constructed from the first four observations, control charts from 6 June until 21 June 2013 show that all plots were in control (Figure 33). Starting from 26 June, the control charts show that plot B, F and J (yellow dots) were located above the Upper Warning Limit (UWL). This shows that these plots started to deviate away from the reference plot at this date, thus triggering the alert state. On the next observation (7 July 2103), Plot F and J entered the out of control state, while plot B was still at the warning state.

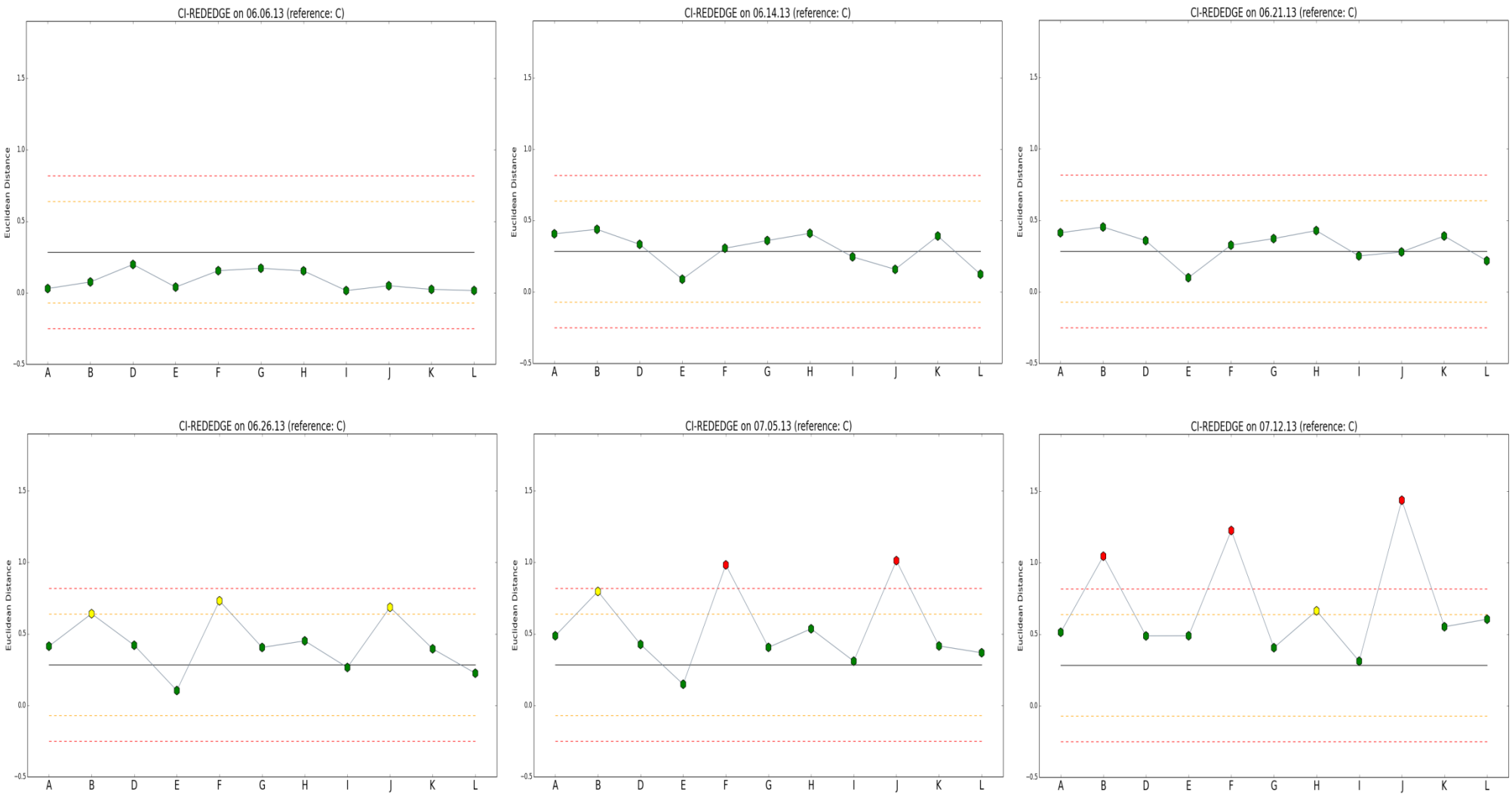


Figure 33: Control charts of the first six observations date which show the status of each plot (Constructed using Python Programming Language). The black line indicates control line (CL), yellow lines indicate warning limits (UWL&LWL) and red lines indicate control limits (UCL&LCL). First 4 observations (6 June, 14 June, 21 June and 26 June) are used as calibration dataset.

Table 9: Alert service for 11 experimental plots at each specific observation date using Euclidean distance values (from CIred-edge data). Green indicates plot is under control, yellow means plot is in "alert" state (above 2σ from the mean) and red indicates that the plot is out of control (above 3σ from the mean).

Date	A	B	D	E	F	G	H	I	J	K	L
06 - Jun	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢
14 - Jun	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢
21 - Jun	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢
26 - Jun	🟢	🟡	🟢	🟢	🟡	🟢	🟢	🟢	🟡	🟢	🟢
05 - Jul	🟢	🟡	🟢	🟢	🔴	🟢	🟢	🟢	🔴	🟢	🟢
12 - Jul	🟢	🔴	🟢	🟢	🔴	🟢	🟡	🟢	🔴	🟢	🟢
17 - Jul	🟢	🔴	🟢	🔴	🔴	🟢	🟡	🟢	🔴	🟢	🟡
26 - Jul	🟢	🔴	🟢	🔴	🔴	🟢	🔴	🟢	🔴	🟡	🔴
31 - Jul	🟢	🔴	🟢	🔴	🔴	🟢	🔴	🟢	🔴	🟡	🔴
16 - Aug	🟢	🔴	🟢	🔴	🔴	🟢	🔴	🟢	🔴	🟡	🔴
23 - Aug	🟡	🔴	🟢	🔴	🔴	🟡	🔴	🟡	🔴	🟡	🔴

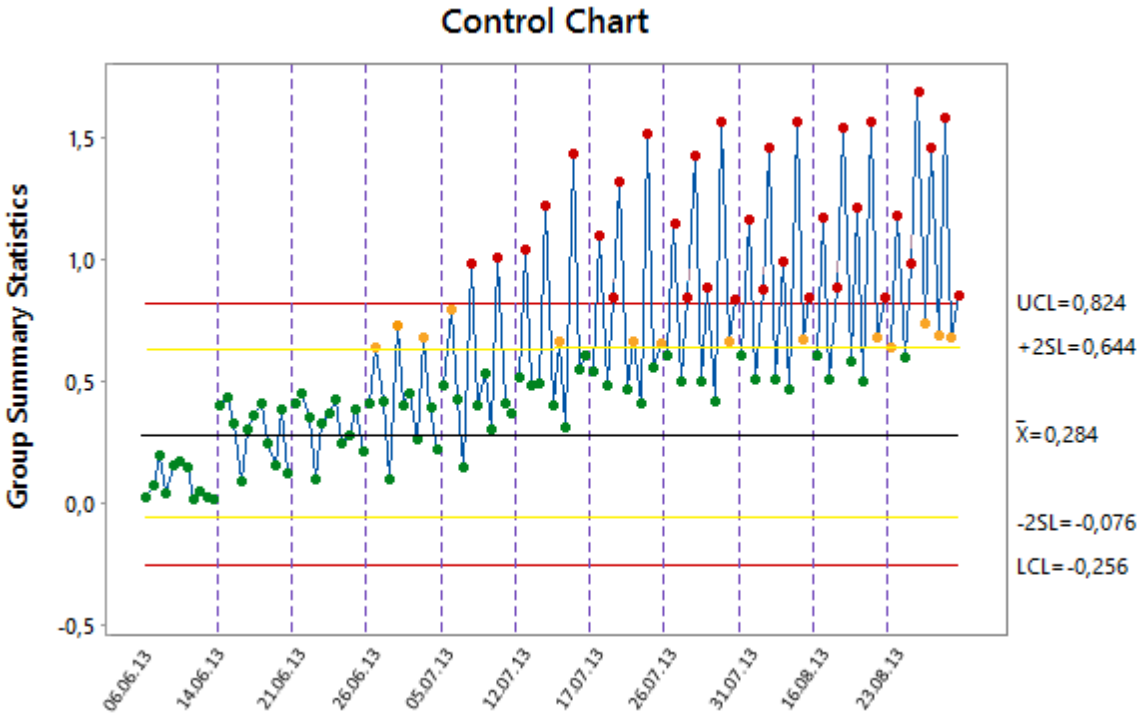


Figure 34: Control chart for all plots together over the growing season (constructed using Minitab 17). The black line indicates control line (CL), yellow lines indicate warning limits (UWL&LWL) and red lines indicate control limits (UCL&LCL). The first 4 observations (6 June, 14 June, 21 June and 26 June) are used as calibration dataset.

As shown in Figure 33, the thresholds from the control chart concept were able to determine the status for every experimental plot on each specific observation date. The control chart approach enables the ability to determine which plots are close to or far from

the reference plot. In that sense, the plots that are far from the reference plot at some point (over the thresholds) will trigger an alert and an appropriate management action should be initiated thereafter. The alerting service summary (from both Figure 33 and Figure 34) for all experimental plots over the growing season can be seen in Table 9. Green circles represent plots that are within the acceptable areas (in control condition). Yellow circles represent plots in the alert state (above 2σ from the mean), while red circles represent an out of control state (above 3σ from the mean).

Table 10: The order of yield harvester (from highest to lowest)

Experimental Plot	Initial Fertilizer	Total Fertilizer	Yield harvester	
	kg/ha	kg/ha	ton/ha	kg/ha
C	162	244.2	81.45	73,887.89
K	162	254.2	81.31	73,765.82
I	252	329.3	78.36	71,087.53
L	90	175.2	74.44	67,534.92
D	90	166.4	73.20	66,403.55
A	252	326.7	72.57	65,833.98
G	162	194.0	71.95	65,269.66
E	252	284.0	71.73	65,068.33
B	0	108.5	70.33	63,800.08
J	0	96.4	66.30	60,146.35
H	90	122.0	65.38	59,309.26
F	0	32.0	58.71	53,262.33

Table 9 shows that plots B, F and J were in the alerting state starting from the fourth observation (26 June). Plot F and J were assigned as out of control states starting from the fifth observation until the end of the growing season; whilst plot B started to be in out of control state at 12 July 2013 (sixth observation). The purpose of the alerting service was to detect plots with growth problems indicated by high distance differences to the reference plot. These plots were then expected to produce low yields at the end of the growing season if no management application was applied. To prove whether the alerting service using the control chart theory worked properly, the status of three plots (plot B, F and J) were cross-checked with the yield harvester data in Table 10. Table 10 shows that these three plots (plot B, F and J) were producing low amounts of yield at the end of the growing season. Plot H, which was in an alert state starting from the sixth observation, also produced low yields at the end of the growing season (Table 10). This proves that the alerting service with threshold values (UWL, LWL, UCL and LCL) constructed from the first four observation dates worked properly.

However, this alerting service (Table 9) could not detect the changes in the plots with management applications interference, for example: the additional fertilizer on 5 July and 18 July did not change the state of the experimental plots. This was due to the use of D_E (Equation 7): a cumulative calculation was applied to all iterations following the first one. As

a consequence, an increasing trend appeared in the D_E data series (Figure 35) which lead to the inability of the alerting service to detect management applications.

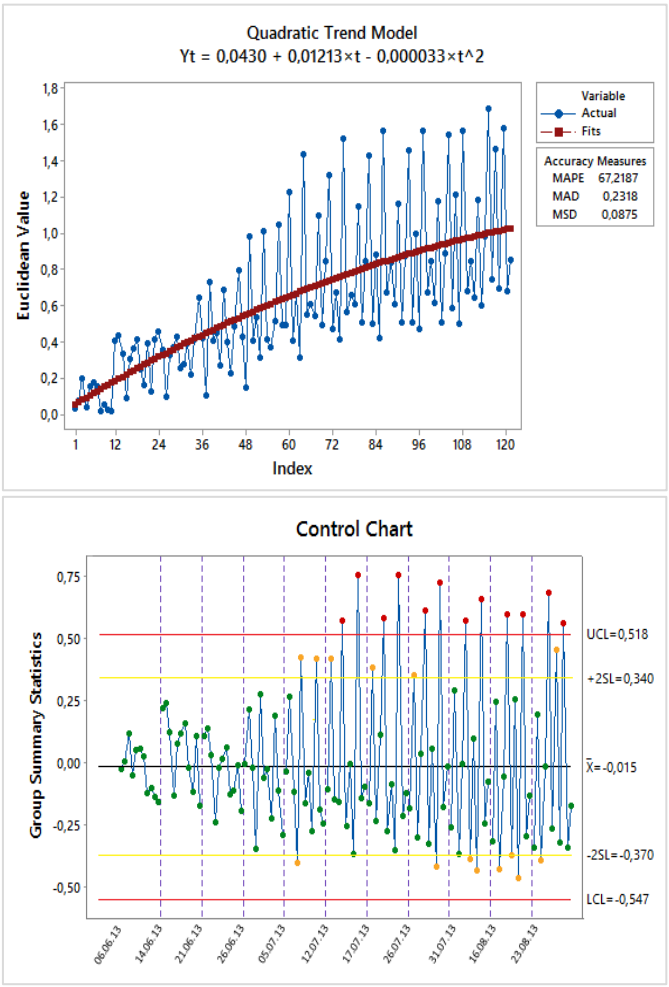


Figure 35: Trend analysis for Euclidean distance value over the growing season (top) and control chart over the growing season using residual data from de-trended process (down)

Two actions were initiated to solve the problem: (1) the de-trended D_E values were used to construct the control charts and (2) the D_E values were used to construct the control charts with the cumulative thresholds method applied to each observation date. To initiate the first action, the data was de-trended and the trend influence were taken out from the D_E data series before constructing the control charts. Next, the residual data series (the de-trended data) were used to perform the control chart approach. In this report, the control chart thresholds were constructed from the first four observation dates (Figure 35). Table 11 (top) shows the alerting service summary for all plots over the growing season using the de-trended approach. Differing from the alerting service without de-trending process, this alerting service was able to track the influence of the fertilizer application. The effects of the first and second additional fertilizer application on specific experimental plots can be seen on the alerting service: the alert state of plot B was changed to normal state on 31 July 2013 after the additional fertilizer was applied. However, this method was unreliable to get within season detection. The trend pattern was only possible to calculate after all the observations

were collected. In other words, the detection could only be made at the end of the growing season. Therefore, this method was more suitable for the evaluation process than the detection phase.

The second solution used the cumulative thresholds method for each observation date. For example, the data from the first observation date were used to construct the thresholds for the first control chart (6 June 2013). For the second control chart (14 June 2013), data from the first and second observation date were used to construct the control chart threshold. This method made the thresholds shift for each observation date, which followed the trend pattern in the data.

Table 11: Alert service for 11 experimental plots at each specific observation date using de-trended data from Euclidean distance values (top); and using Euclidean distance values with cumulative thresholds method (bottom).

Date	A	B	D	E	F	G	H	I	J	K	L
06 - Jun	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢
14 - Jun	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢
21 - Jun	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢
26 - Jun	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢
05 - Jul	🟢	🟢	🟢	🟡	🟡	🟢	🟢	🟢	🟡	🟢	🟢
12 - Jul	🟢	🟡	🟢	🟢	🔴	🟢	🟢	🟢	🔴	🟢	🟢
17 - Jul	🟢	🟡	🟢	🟢	🔴	🟢	🟢	🟢	🔴	🟢	🟢
26 - Jul	🟢	🟡	🟢	🟢	🔴	🟢	🟢	🟡	🔴	🟢	🟢
31 - Jul	🟢	🟢	🟢	🟢	🔴	🟡	🟢	🟡	🔴	🟢	🟢
16 - Aug	🟢	🟢	🟡	🟢	🔴	🟡	🟢	🟡	🔴	🟢	🟢
23 - Aug	🟢	🟢	🟡	🟢	🔴	🟢	🟡	🟢	🔴	🟢	🟢

Date	A	B	D	E	F	G	H	I	J	K	L
06 - Jun	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢
14 - Jun	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢
21 - Jun	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢
26 - Jun	🟢	🟢	🟢	🟢	🟡	🟢	🟢	🟢	🟡	🟢	🟢
05 - Jul	🟢	🟡	🟢	🟢	🟡	🟢	🟢	🟢	🔴	🟢	🟢
12 - Jul	🟢	🟡	🟢	🟢	🟡	🟢	🟢	🟢	🔴	🟢	🟢
17 - Jul	🟢	🟢	🟢	🟢	🟡	🟢	🟢	🟢	🔴	🟢	🟢
26 - Jul	🟢	🟢	🟢	🟢	🟡	🟢	🟢	🟢	🟡	🟢	🟢
31 - Jul	🟢	🟢	🟢	🟢	🟡	🟢	🟢	🟢	🟡	🟢	🟢
16 - Aug	🟢	🟢	🟢	🟢	🟡	🟢	🟢	🟢	🟡	🟢	🟢
23 - Aug	🟢	🟢	🟢	🟢	🟡	🟢	🟡	🟢	🟡	🟢	🟢

Appendix 14 shows the control charts for the first eight observation dates with thresholds that changed over the temporal development. The summary of the alerting service which performed the cumulative threshold method can be seen in Table 11 (bottom). Similar with the de-trended method, this method was unable to give fast alert detections and insensitive to detect growth problems in plots with a medium amount of initial fertilizer (for example:

90 kg N ha^{-1}). As seen in Table 10, plot H (initial fertilizer of 90 $kgNha^{-1}$) was producing the second lowest yield at the end of the growing season. However, this plot was categorized as an alert at the end of the growing season, which made it impossible to initiate any management actions and fix the growth problem (Table 11). Based on these results, the best way to construct the alerting service is using the D_E data series where the control chart thresholds are calculated from the first four observations.

4.4.1.2 Alerting service for Euclidean distance values from TCARI/OSAVI time series

As it was emphasized in Section 4.3.1, the TCARI/OSAVI was not able to represent the potato growth condition. In the time series similarity measures performed in that section, the TCARI/OSAVI data did not give expected results. For example: the plots that should have deviated away from the reference plot (i.e.: F and J) did not perform as expected, whilst the plots that had a high initial fertilizer level (i.e.: plot K) tended to deviate highly from the reference plot starting from the fifth observation (5 July 2013). These unexpected findings influenced the alerting service results. Appendix 15 shows the control charts for the first six observations using TCARI/OSAVI Euclidean distance data series with thresholds generated from the first four observation dates.

Table 12: Alerting service for 11 experimental plots at each specific observation date using TCARI/OSAVI Euclidean distance data series

Date	A	B	D	E	F	G	H	I	J	K	L
06 - Jun	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢
14 - Jun	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢
21 - Jun	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢
26 - Jun	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢
05 - Jul	🟡	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟡	🟢
12 - Jul	🟡	🟡	🟢	🟡	🟢	🟢	🟢	🟢	🟢	🟡	🟢
17 - Jul	🟡	🔴	🟢	🔴	🟡	🟡	🟢	🟢	🟡	🟡	🟢
26 - Jul	🟡	🔴	🟢	🔴	🟡	🟡	🟢	🟢	🟡	🟢	🟡
31 - Jul	🔴	🔴	🟢	🔴	🟡	🟡	🔴	🟡	🟡	🔴	🟡
16 - Aug	🔴	🔴	🟢	🔴	🟡	🟡	🔴	🟡	🟡	🔴	🟡
23 - Aug	🔴	🔴	🟢	🔴	🟡	🟡	🔴	🟡	🟡	🔴	🟡

As we can see from the alerting service summary (Table 12), alerts started to appear for plot A and K from the fifth observation date (5 July 2013). On the following observation, another two plots (plot B and E) triggered alerts. Since plot A and K were supposed to have good growth conditions, the alerts for these were considered as “false alarms”. Plots that should have gotten alerts early on, such as plot F and plot J, did not start to trigger alerts before on the seventh observation date. The alerting service using the TCARI/OSAVI data was not completely ineffective, but half of the results ended up triggering false alarms. Therefore, the alerting service in this report performed better using the results generated from D_E or RMSD using the $CI_{red\text{-}edge}$ data series.

4.4.2 Plant growth alerting service for mean curve approach (Plot level)

Based on the D_E measure executed by the within season alerting service using the mean curve approach in Section 4.3.2, the alerting service for the mean curve approach at the plot level was carried out in this section.

Table 13: Alerting service for 12 experimental plots using $CI_{red\text{-}edge}$ Euclidean distance data series with mean curve approach.

DATE	A	B	C	D	E	F	G	H	I	J	K	L
06 - Jun	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢
14 - Jun	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢
21 - Jun	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢
26 - Jun	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢
05 - Jul	🟢	🟢	🟢	🟢	🟢	🔴	🟢	🟢	🟢	🔴	🟢	🟢
12 - Jul	🔴	🟡	🟢	🟢	🔴	🔴	🟢	🟢	🟢	🔴	🟢	🟢
17 - Jul	🔴	🔴	🟢	🟢	🔴	🔴	🟢	🟢	🟢	🔴	🟢	🟢
26 - Jul	🔴	🔴	🟢	🟢	🔴	🔴	🟢	🟢	🟡	🔴	🟢	🟡
31 - Jul	🔴	🔴	🟢	🟢	🔴	🔴	🟢	🟡	🟡	🔴	🟢	🟡
16 - Aug	🔴	🔴	🟢	🟢	🔴	🔴	🟢	🔴	🟡	🔴	🟢	🟡
23 - Aug	🔴	🔴	🟡	🟢	🔴	🔴	🟢	🔴	🟡	🔴	🟡	🟡

The alerting service was initiated to check the plant growth development for the 12 experimental plots over the growing season. The observation dates from 6 June until 5 July 2013) were used to calculate the control chart thresholds (UCL, UWL, LWL and LCL). From Table 13 we can see that plot F and plot J were in the out of control state starting from the fifth observation (5 July 2013) until the end of the growing season. While plot B triggered a warning alert on the sixth observation (12 July 2013), plot A and E started to be in the out of control state at the same time. On 26 July 2013, both plot I and L triggered alerts which continued until the end of the growing season.

The major difference between the results from Table 12 and Table 13 is that plot A was in an out of control state starting from 12 July 2013 until the end of the growing season using the mean curve approach, while the results using the maximum yield approach did not show this behaviour. However, the alerting service using the mean curve approach was able to detect the warning state from the potato growth over the growing season even though the accumulation effects were inevitable.

4.4.3 Plant growth alerting service for mean curve approach (Subplot level)

The time series similarity measures that were performed in Section 4.3.3 used the $CI_{red\text{-}edge}$ time series data with the reference plot taken from the mean curve approach. In this section, the alerting service was initiated to check the plant growth development for the subplots from the calibration dataset (Plot AL, BR, CL, DL, ER, FR, GR, HL, IL, JR, KL, and LR) using the best time series similarity measure (D_E) data series.

Table 14: Alerting service for 12 experimental subplots using $CI_{red\text{-}edge}$ Euclidean distance data series with mean curve approach

DATE	AL	BR	CL	DL	ER	FR	GR	HL	IL	JR	KL	LR
06 - Jun	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢
14 - Jun	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢
21 - Jun	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢
26 - Jun	🟢	🟢	🟢	🟢	🟢	🟡	🟢	🟢	🟢	🟢	🟢	🟢
05 - Jul	🔴	🟢	🟢	🟢	🟢	🔴	🟢	🟢	🟢	🟡	🟢	🟢
12 - Jul	🔴	🟢	🟢	🟢	🟡	🔴	🟢	🟢	🟢	🔴	🟢	🟢
17 - Jul	🔴	🟡	🟢	🟢	🔴	🔴	🟢	🟢	🟢	🔴	🟢	🟢
26 - Jul	🔴	🔴	🟢	🟢	🔴	🔴	🟢	🟡	🟢	🔴	🟢	🟡
31 - Jul	🔴	🔴	🟢	🟢	🔴	🔴	🟢	🟡	🟡	🔴	🟢	🟡
16 - Aug	🔴	🔴	🟢	🟢	🔴	🔴	🟢	🔴	🟡	🔴	🟢	🟡
23 - Aug	🔴	🔴	🟢	🟢	🔴	🔴	🟡	🔴	🟡	🔴	🟢	🟡

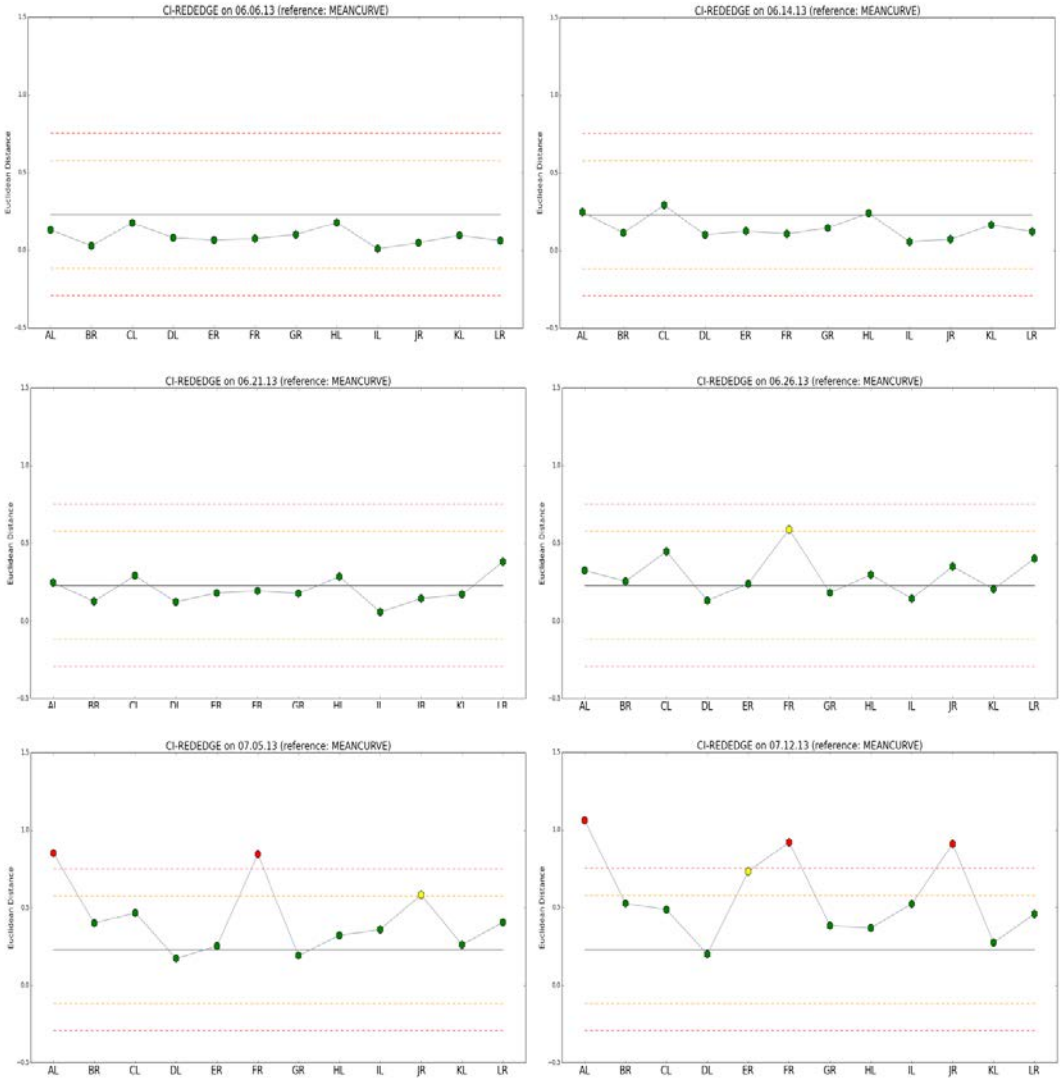


Figure 36: Control charts of the first six observations date which show the status of each subplot (Constructed using Python Programming Language). The black line indicates control line (CL), yellow lines indicate warning limits (UWL&LWL) and red lines indicate control limits (UCL&LCL). First 5 observations (6 June, 14 June, 21 June, 26 June and 5 July) were used as calibration dataset.

Figure 36 shows the control charts for the 12 experimental subplots from the first six observation dates (6 June – 12 July) based on the D_E values. The observation dates from 6 June until 5 July 2013) were used to calculate the control chart thresholds (UCL, UWL, LWL and LCL). The summary of the control chart results can be seen in Table 14. From Table 14, we can see that plot FR was in the alert state starting from the fourth observation (26 June 2013). On the next observation (5 July 2013), plot AL and FR were in the out of control state; whilst plot JR triggered an alert. Plot HL was in the alert state from the eight until the ninth observation (26 July – 31 July 2013) and the out of control state starting from the tenth observation date until the end of the growing season. At the end of the growing season, plot CL, DL and KL were still in the control state. Any plots in an alert state would be an indication to the farmers that necessary measures and immediate actions should be taken.

4.5 Alerting services validation (Phase 5)

In this section, two approaches were carried out to validate the alerting services in section 4.4. The first approach used the STR dataset from the Geveart et al. (2015) study to validate the alerting service using the maximum yield approach and the mean curve approach in the plot level. While in the second approach, the 12 experimental subplots from the validation dataset (AR, BL, CR, DR, EL, FL, GL, HR, IR, JL, KR, and LL) were used to validate the alerting service with the mean curve approach in the subplot level. Before performing the control chart theory, each approach would calculate the D_E values from each experimental plot or subplot to the selected reference plot (plot C for maximum yield approach and "the mean value" plot for the mean curve approach).

4.5.1 Validation of the alerting service with maximum yield approach for the plot level

To generate the alerting service, the $CI_{red\text{-}edge}$ values were calculated from the daily STR dataset. For validation purposes, the $CI_{red\text{-}edge}$ time series were used to calculate the similarity measure (D_E) on a weekly basis from each experimental plot (A, B, D, E, F, G, H, I, J, K and L) compared to the reference plot (Appendix 17). Since plot C produced the highest yield at the end of the growing season, it was assigned as the reference plot. The distance values were then plotted in the control chart and analysed using the threshold values which were calculated using the data from the first four observation dates (6 June, 14 June, 21 June and 26 June 2013).

Table 15 shows the alerting service summary for the 11 experimental plots from the first observation date (6 June) until the ninth observation (31 July). The alerting service was only able to analyse the crop's state until the ninth observation due to that the STR dataset were only available from 6 June until 2 August 2013 (Appendix 17).

Table 15: Alerting service for 11 experimental plots at each specific observation date using Euclidean distance from STR dataset

DATE	A	B	D	E	F	G	H	I	J	K	L
06 - Jun	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢
14 - Jun	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢
21 - Jun	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢
26 - Jun	🟢	🟡	🟢	🟢	🟢	🟢	🟢	🟢	🔴	🟢	🟢
05 - Jul	🟢	🔴	🟢	🟢	🟢	🟢	🟢	🟢	🔴	🟢	🟢
12 - Jul	🟢	🔴	🟡	🟢	🟡	🟢	🟡	🟢	🔴	🟢	🟢
17 - Jul	🟢	🔴	🟡	🟢	🟡	🟢	🔴	🟢	🔴	🟢	🟢
26 - Jul	🟢	🔴	🔴	🟢	🔴	🟢	🔴	🟢	🔴	🟢	🟢
31 - Jul	🟢	🔴	🔴	🟢	🔴	🟢	🔴	🟢	🔴	🟢	🟢

In Table 15 we can see that plot B started to receive an alert on 26 June 2013; whilst plot J entered the out of control state at the same day. On 12 July 2013, three plots (D, F and H) were in the alert state and started to be in the out of control state on 17 July (plot H) and 25 July (plot D and F). By cross checking the results with the yield harvester data in Table 10, we can see that the alerting service successfully detected the plots with the lowest yield production using the information from the STR dataset.

4.5.2 Validation of the alerting service with mean curve approach for the plot level

Before constructing the alerting service based on the mean curve approach, the D_E values were calculated from each experimental plot to the reference plot over the growing season. To be able to compare the alerting service that used the Cropscan data (Section 4.4.2) with the alerting service established in this section, the STR dataset were recalculated, similar with Section 4.5.1, from daily to weekly observations.

Table 16: Alerting service for 12 experimental plots using Euclidean distance values from STR dataset

DATE	A	B	C	D	E	F	G	H	I	J	K	L
06 - Jun	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢
14 - Jun	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢
21 - Jun	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢
26 - Jun	🟢	🟡	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🔴	🟢	🟢
05 - Jul	🟢	🔴	🟢	🟢	🟡	🟢	🟢	🟢	🟢	🔴	🟢	🟢
12 - Jul	🟢	🔴	🟢	🟢	🔴	🟢	🟢	🟢	🟢	🔴	🟢	🟢
17 - Jul	🟢	🔴	🟡	🟢	🔴	🟢	🟢	🟢	🟡	🔴	🟢	🟢
26 - Jul	🟡	🔴	🟡	🟢	🔴	🟢	🟢	🔴	🟡	🔴	🟢	🟢
31 - Jul	🟡	🔴	🔴	🟡	🔴	🟢	🟢	🔴	🔴	🔴	🟢	🟢

Table 16 shows the alerting service for the 12 experimental plots using the mean curve approach to generate the reference plot. In this validation phase, the STR dataset were used to evaluate the effectivity of the alerting service using different sources of data. From Table 16 we can see that Plot B got an alert on the fourth observation date, while plot J started to receive out of control states from here on until the end of the growing season. Plot E was in the alert state on the fifth observation date (5 July 2013), but changed to the out of control phase on the next observation date (12 July 2013). However, plot F did not receive any alerts over the growing season when the mean curve approach was applied. As seen in Table 15, plot F started to get alerts on 12 July 2013 from the alerting service that used the maximum yield approach (plot C as reference plot). Even though this alerting service was not able to detect the growth problem in plot F, it still detected the growth problems from the other plots that known for their low yield productions at the end of growing season (i.e.: plot B, J, H and E).

4.5.3 Validation of alerting service with mean curve approach for the subplot level

As depicted in Figure 25, the validation of this alerting service (the mean curve approach) was done using the 12 experimental subplots from the validation dataset. The Euclidean distance between these 12 experimental subplots (AR, BL, CR, DR, EL, FL, GL, HR, IR, JL, KR, and LL) and the reference curve ("mean value plot") were calculated over the growing season. These values were then plotted using the control chart theory to detect the plots with growth problems. Based on the analysis in Section 4.4.3, the control chart thresholds were calculated using the dataset from the first five observation dates. The same threshold calculation approach was applied in this section to check the accuracy of the alerting service.

Table 17: Alerting service for 12 experimental subplots at each specific observation date using Euclidean distance values from $CI_{red\text{-}edge}$ data series

DATE	AR	BL	CR	DR	EL	FL	GL	HR	IR	JL	KR	LL
06 - Jun	●	●	●	●	●	●	●	●	●	●	●	●
14 - Jun	●	●	●	●	●	●	●	●	●	●	●	●
21 - Jun	●	●	●	●	●	●	●	●	●	●	●	●
26 - Jun	●	●	●	●	●	●	●	●	●	●	●	●
05 - Jul	●	●	●	●	●	●	●	●	●	●	●	●
12 - Jul	●	●	●	●	●	●	●	●	●	●	●	●
17 - Jul	●	●	●	●	●	●	●	●	●	●	●	●
26 - Jul	●	●	●	●	●	●	●	●	●	●	●	●
31 - Jul	●	●	●	●	●	●	●	●	●	●	●	●
16 - Aug	●	●	●	●	●	●	●	●	●	●	●	●
23 - Aug	●	●	●	●	●	●	●	●	●	●	●	●

Table 17 shows that on the fifth observation date (5 July) plot JL started to get an alert; whilst the other 10 plots were still in the in control state. On the sixth observation date, plot

EL entered the out of control state and continued in this state until the end of the growing season. For plot FL, alerts were seen from 12 July until 26 July 2013, but started showing out of control states starting from 31 July 2013. Plot BL entered the alert state on the sixth observation date and plot HR started to be in the alert state from the ninth observation date.

Table 18: Yield Harvester for 12 experimental subplots (highest to lowest)

Experimental Plot	Initial Fertilizer kg/ha	Total Fertilizer kg/ha	Yield harvester	
			ton/ha	kg/ha
CR	162	244.2	83.96	76,169.88
IR	252	329.3	80.71	73,220.77
KR	162	254.2	80.12	72,679.82
AR	252	326.7	78.10	70,852.18
DR	90	166.4	75.09	68,115.97
GL	162	194.0	72.10	65,406.84
LL	90	175.2	71.57	64,923.84
EL	252	284.0	69.26	62,828.50
HR	90	122.0	68.68	62,308.17
BL	0	108.5	66.46	60,291.50
JL	0	96.4	64.68	58,681.08
FL	0	32.0	60.32	54,718.66

Even though the reference plot used for this alerting service was based on the mean curve approach, the service was able to detect plots that produced low yields (Table 18) at the end of the growing season. This proves that the control chart thresholds effectively detected the growth issues over the growing season.

This chapter reflects on the results considering the research questions proposed in Section 1.3. The results are mainly discussed in broader context and linked to the scientific literatures.

5.1 Vegetation Indices for potato N status (Phase 1)

There were three analyses enclosed in Phase 1: the N crop status comparison using three different measurements over the growing season; analysing the relations between the plant's biophysical and biochemical parameters; and finding the best VI that best represented the N potato status. Therefore, the discussion will be based on each step explained in Section 4.1.

5.1.1 N Status comparison with different measurement methods

Three N status measurements were plotted over time (Figure 14-16) to analyse possible trend patterns over the growing season. Haboudane et al. (2004) stated in their study that plant N variations may cause biochemical changes at the leaf and canopy level (i.e. LAI). The N status from all three measurements showed relatively high values in the beginning of the growing season and started to decline over the development of the potato growth. The declining trend was related to the development of the potato canopy size which was indicated by the increase of the plant biomass and LAI. This condition can be seen in Appendix 18 where the LAI from all plots show relatively high values while the N status values from the PSNC test are on their lowest (5 July - 12 July).

Based on the information of management decision in Table 5, additional fertilizer applications were applied to selected plots on two observation dates over the growing season, i.e.: on 5 July and 18 July. These fertilizer applications were supposed to influence the N status readings around three weeks after the application was applied. Figure 15 and Figure 16 show that the PSNC test were more sensitive to the fertilizer application than the SPAD chlorophyll meter. This behaviour was visible during the observation on 31 July 2013, where the PSNC test values from plot A, B, C, D, I, J, K and L increased (higher) while the SPAD values were still low. This shows that PSNC test were able to respond faster (approximately 2 weeks after the application was applied) than the SPAD readings. The results were similar to a study conducted by Wu et al. (2007). One of the points discussed in their study was that SPAD readings did not respond as rapidly as petiole NO_3-N concentrations to N fertilization throughout the growing season (Wu et al., 2007). The difference between the two measurements could possibly have occurred due to the different level of measurements: the PSNC test was acquired in the plant tissue level, while the SPAD chlorophyll meter was acquired in the leaf level. Figure 18 shows the relationship between the PNSC test values (nitrate nitrogen concentration) with the SPAD measurement.

The overall relationship between the nitrate nitrogen concentrations (PSNC test) and the SPAD measurement were 0.75 with p-value < 0.05 (significant). However, the SPAD values deviated more in the low nitrate concentration values. This condition is shown in Figure 18: the SPAD values varied around 32 – 40 for nitrate concentration values around 800 ppm NO_3.

5.1.2. Relationships between crop biophysical and biochemical parameters

The overall relationship between the leaf chlorophyll and the nitrate concentration derived from the PSNC test was moderately high (R^2=0.745) as seen in Figure 19. However, an uncommon pattern was seen in the PSNC test values on 31 July for plot A, B, C and D. The PSNC values for these plots had increased while the leaf chlorophyll values showed relatively low values. This was related to the ability of the PSNC test to respond to the additional fertilizer application on 18 July 2013. Apart from that, the relationship between PSNC test and leaf chlorophyll was better than the overall relationship (all observations were bulked) between the PSNC test and the canopy chlorophyll (R^2= 0.075).

The overall relationship between the PSNC test and the canopy chlorophyll showed insignificant results when all data were bulked into the calculation. However, better results were seen for the relationships based on the specific time of the growing season. From Table 6, we can see that the relationship between the PSNC test and the chlorophyll canopy were significant starting from the fourth observation (26 June 2013) until the end of the growing season. This proved that the chlorophyll canopy values were able to give good representations of the N status in the canopy level. Therefore, in section 4.1.3 the best VI was not only identified for the leaf chlorophyll, but also for the chlorophyll canopy.

5.1.3. Relationships between VIs and chlorophyll content

The first objective of this thesis research was to find the best VI that would provide a good representation of the N status in the potato crop over the growing season. Eight VIs that are normally used in precision agriculture fields were calculated to analyse their ability to detect changes in the crop N status over the growing season. Since the crop N status were well related to both the leaf chlorophyll and the chlorophyll canopy, the regression analysis were applied to find the best relation between the eight VIs with chlorophyll content in the leaf and canopy level. The results in Table 7 show that the chlorophyll ratio index TCARI/OSAVI represented the best VI for leaf chlorophyll, even though the coefficient of determination value was relatively low (R^2=0.517). Similar results were also found in a research conducted by Wu et al. (2008). In their study, the correlation coefficient between the original TCARI/OSAVI [670,800] and the chlorophyll content was 0.4984. However, a strong correlation was achieved if the wavelengths that were used in TCARI/OSAVI (670 and 800 nm) were replaced by 705 and 750 nm. In the case of this study, the relation between the TCARI/OSAVI and the leaf chlorophyll became weaker in line with the temporal development. This fact was proven in the time series similarity measures and alerting

service derived from this VI. The TCARI/OSAVI was not able to perfectly detect the crop N development over the growing season (Figure 28) and tended to give 50% of false alarms in the alerting service. In other words, even though this VI showed the highest relation with leaf chlorophyll compared with other VIs, the values of the correlation itself should be considered.

The best VI for chlorophyll canopy based on the results in Table 7 was $CI_{red\text{-}edge}$. The relationship between the $CI_{red\text{-}edge}$ and the chlorophyll canopy was relatively high, which was shown by the coefficient of determination value = 0.858. This result was related to the significance of wavelengths in the red-edge region to estimate the chlorophyll content (Clevers & Kooistra, 2012). The result was also confirmed by the study of Clevers and Gitelson (2013) on another potato site. In their study, the results of R^2 value of the linear relationship between the $CI_{red\text{-}edge}$ and nitrogen content on the potato site was relatively high (0.89). Even though the result was not shown in this report, an additional calculation was made for the TCARI/OSAVI. TCARI/OSAVI [750,710] was calculated from the Cropscan data and the relationship between this VI and the chlorophyll canopy showed a relatively high value (R^2=0.784); whilst the relationship between TCARI/OSAVI [750,710] and the leaf chlorophyll showed an insignificant result (R^2=0.08).

5.2 Optimal growth curve for potato crop (Phase 2)

As explained in Section 4.2, the optimal growth curve for potato growth in this report was acquired using two different approaches: (1) the maximum yield approach and (2) the mean curve approach. In the beginning of the study, there was another approach that was considered important to use: the optimum quality approach. However, the available data related to the potato quality were not available for all experimental plots and subplots level. Therefore, this approach was eliminated from this study.

The importance of the maximum yield approach is that the potato development curve is expected to follow the curve from the plot that produces the highest yield at the end of the growing season. Therefore, any experimental plot that has a VI curve apart from the reference curve (maximum yield plot) are considered to have a growth problem. Using control chart theory these growth problems will result in alerts, similar to those triggered by the alerting services in Section 4.4. However, a reference curve that is the result of the maximum yield approach can only be used in on-coming growing seasons. This is due to the yield harvester information only being available at the end of the growing season. To solve the problem with real time detection or within season detection, assuming that there are no history of previous yield harvester data, the mean curve approach can be used. In the study conducted by Bala and Islam (2009), the mean curve approach was generated by calculating the average value of the VI on a specific date from all available potato plots consisting of different levels of initial fertilizer. In Section 4.3.2 and 4.3.3, the mean curve was assumed to be the optimum growth curve for the potato crop. Therefore, in those two sections, the

plots that deviated away from the mean curve (at certain thresholds) was considered as plots with a growth problem and in need of immediate action. Since the mean curve approach was assumed as the best approach for within season detection or real time detection, the analysis in the subplots level was only made using this approach (Section 4.3.3).

5.3 Time series similarity measures for deviation detection (Phase 3)

The third objective in this research was to analyse the time-series similarity measures that could be adopted to evaluate the deviation from the established reference curve to characterize the growth status. The idea of using time series similarity measures to detect changes in the potato crop N status was first applied in the study conducted by Kooistra et al. (2012). In their study there were three aspects which needed to be done before the N status changes could be detected. First, a crop biophysical indicator highly related to the potato N status over the growing season had to be defined. After that, the growth reference curve of the potato crop over the growing season had to be determined to compare the status of experimental plots. Then, the method to detect the change in the crop N status had to be chosen. The steps in their study were adapted in this thesis research. The crop biophysical indicator highly related to the potato N status was identified from the regression analysis between the VIs and chlorophyll content (leaf and canopy level). Based on the result from section 4.1.3, the best biophysical indicators were TCARI/OSAVI and $CI_{red\text{-}edge}$.

In the second step, obtaining the optimum growth curve or reference curve, was also done in this research by using two different approaches: the maximum yield curve and the mean curve approach. For the maximum yield approach, plot C was selected as the reference curve based on the yield harvester information in Table 8; whilst the reference curve using the curve approach was generated from the average values of all experimental plots on each observation date (Figure 25). As explained in Section 5.2, the mean curve approach was considered as a better approach to carry out within season alerting service. Therefore, for the analysis in the subplot level in Section 4.3.3, only the reference plot based on the mean curve approach was applied to the data series.

The literature study was carried out in Section 2.4 to check the capability of time series similarity measures in detecting changes within VI time series data. A previous study conducted by Lhermitte et al. (2011) explained that similarity measures based on distance measures or correlation measures were able to detect the changes in time series data. Based on this study, four similarity measures (D_{Man}, D_E, D_{CC}, and RMSD) were performed to two VIs ($CI_{red\text{-}edge}$ and TCARI/OSAVI) with two different reference curve approaches (maximum yield and mean curve approach) in two experimental area levels (the plot level and the subplot level). The results of the similarity measures were given in Section 4.3.1 – 4.3.3. In general, three analysis were made: similarity measures using the maximum yield

approach for the plot level, similarity measures using the mean curve approach for the plot level, and similarity measures using the mean curve approach for the subplot level.

Similarity measures using maximum yield approach for the plot level

In Section 4.3.1, four similarity measures (D_{Man}, D_E, D_{CC}, and RMSD) had been carried out on the two best biophysical indicators (TCARI/OSAVI and $CI_{red\text{-}edge}$) time series data. Plot C was assigned as the reference curve based on the maximum yield approach, as explained in section 4.2.1. In this plot level analysis, each experimental plot (A, B, D, E, F, G, H, I, J, K and L) was compared to the reference plot C and the distance differences were calculated. Based on the results on Figure 27, Appendix 6, and Appendix 8; the time series similarity measures based on distance measures were generally able to detect the changes of the potato crop growth development. This can be seen by the ability of the three similarity measures to respond to the changes in the N status of the potato crops based on their initial fertilizer level and additional fertilizer over the growing season. However, the sensitivity of each of them was divergent. In the beginning of the growing season, D_E and RMSD were able to give a better portrait of the growth condition in the crops (Figure 27 and Appendix 8). Differing from the D_{Man} values that were relatively similar for all experimental plots, D_E and RMSD values were showing more variations since the first observation. The variations were expected as early as the first observation (6 June 2013) since the initial fertilizers were applied to the soil before planting.

From Appendix 6, we can see that the D_{Man} values were ranging from 0 to 5 over the growing season, whilst the D_E values were ranging from 0 – 1.5. The difference between these two values were related to the definition of the D_E itself. Based on the definition in Cha (2007), D_E is known as the shortest distance between two points, while D_{Man} is explained as the sum of the absolute differences of their Cartesian coordinates (Xu, 2104). From the visual illustration given by the spatial study of Shadid et al. (2009) and Xu (2014), we can see that D_E tends to underestimate the distance while D_{Man} tends to overestimate the distance (Appendix 19). However, in this report, D_E and RMSD were proven to show a better growth detection than D_{Man} over the growing season. As seen in Figure 27 and Appendix 8, the difference between one plots to another were clearer over the growing season. This means that these two similarity measures were able to differentiate plots based on their initial and additional fertilizer level. The effects of the additional fertilizer on 5 and 18 July 2013 were also clearer in the time series similarity measures performed by D_E and RMSD than from D_{Man} (Figure 27, Appendix 6, and Appendix 8).

In the case of the similarity measures based on the correlation coefficient (Appendix 7), the result was quite different from the distance measures. The D_{CC} values from all experimental plots in the beginning of the growing season showed a strong relationship to reference plot C (value=1). Over the temporal development, the relationships of the experimental plots to the reference plot C varied based on the initial fertilizer level. The relationships of the plots

with 0 $kgNha^{-1}$ initial fertilizer (plot B and plot J) decreased over the time. However, D_{CC} was not able to clearly distinguish the real growth pattern from the experimental plots; for example: plot F that should have had a low relationship with reference plot C, showed the opposite result.

The results from the time series similarity measures based on distance measures can be used by the farmer or the farm management to check the growth condition (N status) of the crops on a specific time and location. However, the selection of the VI is very important. In Section 4.3.1, the similarity measures were carried out twice: (1) to $CI_{red\text{-}edge}$ time series and (2) to the TCARI/OSAVI time series. These two VIs were selected based on their relationship with chlorophyll content in Section 4.1.3. The relationship between $CI_{red\text{-}edge}$ with the chlorophyll canopy was relatively high (R^2 = 0.858); whilst the relationship between the TCARI/OSAVI with the leaf chlorophyll was 0.517 (all data bulked). Based on the D_E result in Figure 28, the TCARI/OSAVI was not able to differentiate between the plots with high initial fertilizer levels and the ones with low initial fertilizer levels. The TCARI/OSAVI tended to give false growth conditions over the growing season. Similar conditions were also found in the results from the other three similarity measures (Appendix 9). These conditions were most likely related to the low coefficient correlation values between TCARI/OSAVI and leaf chlorophyll (R^2 = 0.517). Generally, leaf chlorophyll is influenced by the development stage of the potato crop. After the full crop coverage, the potato structure will influence the vegetation reflectance and the TCARI/OSAVI values. Therefore, the TCARI/OSAVI was not able to give a good representation of the N status condition for the crops in the leaf level.

Similarity measures using mean curve approach for both plot and subplot level

In Section 4.3.2 and 4.3.3, the time series similarity measures were conducted with the reference plot based on the mean curve approach for both plot and subplot level. The reference plot based on the mean curve approach was acquired from the average value of the 24 experimental subplots (Figure 25) for each observation date over the temporal development. Selecting the right reference plot is the most important step when constructing an alerting service, because different reference plots gives different results. The maximum yield approach proved to give a good representation of the crop growth and N status condition over the growing season (Section 4.3.1). However, the mean curve approach was initiated to tackle the problem that the maximum yield approach had. As explained in Section 4.2, to obtain the optimum growth curve based on the maximum yield approach, the yield harvester information needed to be gathered at the end of the growing season. This made the maximum yield approach unusable for within season detection.

The similarity measures based on the distance measures (D_{Man}, D_E, and RMSD) and the correlation measure (D_{CC}) were performed on both the $CI_{red\text{-}edge}$ and TCARI/OSAVI data series. The results from the similarity measures on the plot level using $CI_{red\text{-}edge}$ (Figure 29

and Appendix 10) showed slightly different results than the one from maximum yield approach (Figure 27 and Appendix 6 - Appendix 8). With this reference curve approach, plot C (the plot that produced the highest yield) deviated away from the reference curve at the beginning of the growing season and starting to be in a good condition over the temporal development. However, this approach was able to perfectly detect the high deviation from plots with an initial fertilizer level of 0 $kgNha^{-1}$ (plot B, F and J). The same conditions were also applied in the subplot level. In the case of the TCARI/OSAVI data series, the results of the time series similarity measures for both levels (plot and subplot) showed that this VI tended to give a false condition of the potato growth development. Similar with the results from the maximum yield approach, in Figure 30 and Figure 32 we can see that two plots with an initial fertilizer level 0 $kgNha^{-1}$ (plot J and plot F) did not deviate much from the reference plot. This made a final conclusion that the similarity measures performed with the mean curve approach gave comparable results with the ones from the maximum curve approach. However, choosing the best VI series had to be done carefully.

5.4 Plant growth alerting service using control chart theory (Phase 4)

The main objective of this research was to derive an alerting service from the crop monitoring time series which in this report the time series data were taken from the Cropscan dataset. To derive an alerting service for the crop development, the range of acceptable growth conditions needed to be set. Later, four different thresholds (LWL, UWL, LCL and UCL) were calculated in Section 4.4 to decide which experimental plots or subplots that should get an alert over the crop developments. These thresholds were used in the statistical control charts in Section 4.4.1 – 4.4.3 as part of the alerting services. Following the similarity measures analyses in Section 4.3, the alerting services were mainly carried out in three different parts: an alerting service based on the maximum yield approach for the plot level, an alerting service based on the mean curve approach for the plot level and an alerting service based on the mean curve approach for the subplot level.

The alerting service based on the maximum yield approach for the plot level

The alerting service based on the maximum yield approach in Table 9 used the D_E values from section 4.3.1. The control chart thresholds were calculated from D_E values from the first (6 June 2013) until the fourth observation date (26 June 2013). As seen from the summary in Table 9, the alerting service using the maximum yield approach (Plot C as the reference curve) was able to detect the plots with the alert state as early as the fourth date of observation. The plots that mainly produced low yields at the end of the growing season (Table 10) were detected perfectly by the alerting service. Plot F (the lowest yield) started to get an alert on 26 June 2013 and changed to the out of control state on 5 July 2013 until the end of the growing season. Plot B and J were in the alert state on 26 June 2013 (fourth observation) and changed to the out of control state on the fifth (Plot J) and sixth observation (Plot B). As seen in Figure 35, the D_E values data series that were plotted in the

control chart had an increasing trend. Even though the increasing trend appeared in the data series, the control chart approach in Figure 33 still fulfilled the Nelson rules (Nelson, 1984). However, this control chart was not able to detect the management applications interference due to the cumulative calculation in the distance measure (Equation 7). To improve the control chart result, two actions were initiated: (1) the de-trended D_E values were used to construct the control charts and (2) the D_E values were used to construct the control charts with the cumulative thresholds method applied to each observation date. The alerting service from the first solution (Table 11), control chart with de-trended data, successfully detected the crop growth condition by emphasizing the influence of the additional fertilizer application. However, the trend pattern in the data series was only possible to obtain while all the observations over the growing season had been collected. In other words, the detection could only be made at the end of the growing season. Therefore, this method is more suitable for the evaluation process than the detection phase.

The alerting service derived from the second solution is shown in Table 11 (down). Similar with the de-trended method, this method was also unable to give fast alert detection and insensitive to detect growth problem in the plots with medium amount of initial fertilizer (i.e. 90 $kgNha^{-1}$). During the research, an additional control chart method (I-MR chart) was applied. However, this method was only giving an alert to plot F over the growing season. Therefore, the best way to construct the alerting service based on the maximum yield approach was by using the D_E data series where the control chart's thresholds were calculated from the first four observations.

In section 4.4.1, there were two D_E data series used to generate the alerting services in Table 9 and Table 12: the data series from $CI_{red\text{-}edge}$ (Figure 27) and the data series from TCARI/OSAVI (Figure 28). As it has been explained, the alerting service that was based on the maximum yield approach using the $CI_{red\text{-}edge}$ Euclidean distance data series (Table 9) was able to detect the plots with growth problems. Differing from this result, the alerting service that was based on the TCARI/OSAVI tended to give false alarms. As explained in Section 4.4.1, plot K, with an initial fertilizer of 162 $kgNha^{-1}$, got alerts starting from the fifth observation. This condition did not fit the yield harvester information in Table 8. Therefore, the alerts that were given to plot K, one of the examples, were considered as false alarms. The TCARI/OSAVI was proven to give false alarms and was deemed unfit to be the input data for the alerting services.

The alerting services based on the mean curve approach for both plot and subplot level

The alerting services based on the mean curve approach for both plot and subplot level can be seen in Table 13 and Table 14. The different colours that appeared in the alerting service summary in both Table 13 and Table 14 indicated different states of the crop development. The green circle is a sign of the under control state from the crop development. If the distance difference from the plot or subplot to the reference plot was located above UWL (2 standard deviations from the mean) or below the LWL, then the experimental plot or

subplot would be assigned with the yellow circle (warning state). If the condition worsened and the distance difference values was larger than three standard deviations from the mean, the plot or subplot would be shown as red circle.

From Table 13, we can see that plot F and plot J were in the out of control state starting from the fifth observation until the end of the growing season. Plot A and E were in the out of control state on the sixth observation (12 July 2013); while plot B got an alert on the same date and changed to the out of control state on the seventh observation. The alerting service in the plot level using the mean curve approach resulted in the condition that plot A was in the out of control state starting from the sixth observation. This condition was different from the other approach (Table 9) where plot A got an alert only at the end of the growing season (23 August 2013). As it can be seen from the D_E measure on Figure 29, the mean curve approach was able to better emphasize the real condition of plot A. In the beginning of the growing season, the left side of plot A had a growth problem (Table 14); whilst the right side of plot A tended to grow normally. However, the crop condition from the left part of plot A got better during the development of the crop. Since the $CI_{red\text{-}edge}$ time series values for plot A was calculated from the average value of both the right and the left side of plot A, the general condition for plot A became less good. This condition was captured by the alerting service using the mean curve approach.

Table 19: Yield harvester for 12 experimental subplots in the calibration dataset

Experimental	Initial Fertilizer	Total Fertilizer	Yield harvester	
Plot	kg/ha	kg/ha	ton/ha	kg/ha
KL	162	254,2	82,56	74.899,56
CL	162	244,2	78,90	71.579,97
LR	90	175,2	76,80	69.673,52
IL	252	329,3	76,45	69.354,27
BR	0	108,5	75,27	68.287,80
ER	252	284,0	74,74	67.805,90
GR	162	194,0	71,83	65.166,10
DL	90	166,4	71,12	64.521,77
AL	252	326,7	68,28	61.946,01
JR	0	96,4	68,25	61.910,97
HL	90	122,0	61,03	55.363,34
FR	0	32,0	56,71	51.441,91

In the subplot level, the alerting service also gave good growth detections. As we can see in Table 14, subplots that produced low yields at the end of the growing season (Appendix 5), such as FR, JR, ER, AL and BR, clearly got alerts as early as the fourth observation date. Subplot FR was in an alert state on the fourth observation, while subplot JR started to get an alert on the fifth observation date (5 July 2013). These two subplots produced low yields at the end of the growing season: 56.71 ton/ha and 68.25 ton/ha respectively (Table 19). The other subplot that produced low yields at the end of growing season, subplot AL, was in the out of control state starting from the fifth observation date. With its ability to detect the growth problem within the experimental subplots over the growing season, the alerting

service using the mean curve approach was proven to give accurate alerts. Alerting services that detect when plots changes their state from green to yellow circle or from green to red circle, can be used by farmers to take additional measurements or management actions to solve the problems within the crop field.

5.5 Alerting services validation (Phase 5)

To validate the alerting services based on the maximum yield approach and mean curve approach, the STR dataset from a previous study conducted by Gevaert et al. (2015) were used in this report. The purpose of this validation was to evaluate the effectivity of the alerting service, including the calculation to construct the thresholds, from different sources of data. There were two approaches involved in the validation process: (1) using the STR dataset from the Geveart et al. (2015) to test the alerting service in the plot level; and (2) using the 12 experimental subplots from the validation dataset (AR, BL, CR, DR, EL, FL, GL, HR, IR, JL, KR, and LL) to validate the alerting service with the mean curve approach in the subplot level.

The result of the alerting service validation with the maximum yield approach for the plot level can be seen in Table 15. Similar with the alerting service result from the Cropscan dataset in Section 4.4.1, both plot B and J started to get an alert from the fourth observation date. Plot F, another plot from the plot group with an initial fertilization level of 0 kg N ha^{-1}, started to be in the alert state on the sixth observation date. The main difference between the two alerting services, the one using the Cropscan dataset and the one using the STR dataset, was that plot D, which got an alert on 12 July 2013 (Table 15), did not get any alerts in the alerting service using the Cropscan data. The difference occurred due to the different equipment that were used to collect the data within the field. However, the alerting service still successfully detected the plots with the lowest yield production (Table 10).

The STR dataset was originally available on a daily basis (Appendix 17). However, for validation purposes, the alerting service from the STR dataset was calculated based on a weekly basis. This made the alerting services from the two different datasets easier to compare. The alerting service that used the mean curve approach for the STR dataset can be seen in Table 16. This alerting service was not able to detect the growth problem in plot F; while the growth problems from the other plots that were known for their low yield productions at the end of growing season (i.e.: plot B, J, H and E) were still detected. Therefore, the alerting service still counted as an accurate tool to derive crop growth alerts from different types of data sources.

The second validation process using the 12 experimental subplots from the validation dataset was done to check the alerting service with the mean curve approach for the subplot level. This alerting service revealed new information that within a 30 x 30 m experimental plot the crops would sometimes react differently towards the initial or additional fertilizer. The alerting services based on the experimental subplot gave a better

alert of the crop growth condition over the growing season. The alerting service in Table 17 was able to detect plots that produced low yields (Table 18) at the end of the growing season. This proves that the control chart thresholds effectively detected the growth issues in the subplot level over the growing season.

CHAPTER 6: Conclusions and Recommendations

Tracking down the crop nutrient status over the growing season is the most principle step for monitoring crop health status in precision agriculture. As a measure of the crop response to nitrogen application, nutrient status is also related to chlorophyll content and an indicator of photosynthetic activity. Ground based measurements of crop parameters data; close sensing data; fused satellite and UAV Imagery data were used in this study to identify the relationship between crop N status and chlorophyll content over the growing season. Based on the result from this study: the relationship between the PSNC tests (nitrogen nitrate) and the leaf chlorophyll was relatively high; whilst the relationship between the nitrogen nitrate and chlorophyll canopy was insignificant when all the data were bulked. As explained in this study, the relationship between the chlorophyll canopy and the nitrogen nitrate improved when the analyses were based on specific dates of observations in the productive part of the season. This study also analysed three different N status measurement methods: the Dumas-combustion method, the PSNC test or plant sap, and chlorophyll meter using Minolta SPAD-502; and compared the relationships between chlorophyll content and the crop biophysical indicators.

The data series of the best representative crop biophysical indicator were used in this study to perform the time series similarity measures over the growing season. Then the distance differences from the similarity measures were plotted using control chart theory to derive the alerting service. The following **conclusions** were made based on all the results and discussions from this study:

- A similar pattern was found when all three N status measurements were plotted over time: the N status from all three measurements showed relatively high values in the beginning of the growing season and started to decline over the development of the potato growth.

- The PSNC test was clearly able to respond faster (approximately 2 weeks) than the SPAD readings to the additional fertilizer application during the growing season.

- The overall relationship between the leaf chlorophyll and the nitrate concentration derived from the PSNC test was moderately high (R^2=0.745). However, the PSNC test responded faster to the additional fertilizer application on 18 July 2013.

- The overall relationship between the PSNC test and the canopy chlorophyll showed insignificant results when all the data were bulked into the calculation. However, better results were seen for the relationships based on specific points in time within the growing season.

- Considering the results from the coefficient of determination (R^2), S values and p-values; the following conclusions were made:

(a) The TCARI/OSAVI index was found to be the best VI for representing the N status at the leaf level over the growing season. However, its low R^2 value (0.517) affected the performance of this index in the time series similarity measures analyses and the results from the alerting services.

(b) $CI_{red\text{-}edge}$ seemed to be the best VI to represent the N status condition at the canopy level over the growing season.

- The study proved that the time series similarity measures based on the distance measures were able to detect the changes within the experimental plots or subplots with different level of accuracy. However, the selection of VI data series was very important. The time series similarity measures using the two VIs gave different results:

  (a) The similarity measures performed using the $CI_{red\text{-}edge}$ time series data were able detect the growth problems from the experimental plots or subplots and also able to distinguish the plots or subplots based on their initial fertilizer level.

  (b) The results from the similarity measures that were performed using the TCARI/OSAVI data series showed that: the TCARI/OSAVI tended to give a false condition of the potato growth development and could not distinguish the plots based on their initial fertilizer level.

- The time series similarity measures based on the mean curve approach were able to give good results; as good as the ones from the maximum yield approach. Therefore, if there is no history of previous yield harvester data, reference plots based on the mean curve approach can be used as an alternative for within season detection.

- The control chart thresholds can be determined using control chart theory. However, choosing the best control chart and determining the thresholds should be done carefully by paying attention to the Nelson's rule.

- Since the alerting services in this study were unable to detect the effects of the management activities due to the effect of the D_E cumulative calculation, then changes from green circles to yellow or red circles could be an indication for the farmers to take immediate actions (i.e.: additional measurement or other management actions).

- The alerting services also proved that they properly detected plots with growth problems or plots that produced low yields at the end of the growing season using the STRS dataset (Fused satellite and UAV Imagery data).

Recommendations:

- As explained before, the alerting services in this study were unable to detect the effects of the management activities due to the effect of the D_E cumulative calculation. Two actions were initiated to solve this problem. However, the strategy could not be used due

to its low sensitivity and late detection. It would be very beneficial if subsequent studies can find solutions to this problem.

- To be able to compare the results from the alerting service that used the Cropscan dataset with the one that used the STR dataset, the STR dataset was changed from daily to weekly observations. In subsequent studies, the daily observations from the high temporal resolution data, as in the STR dataset, should be used to derive the alerting services.

- Subsequent studies using fused satellite and UAV imagery data, such as the STRS dataset in this study, should review the data interpolation carefully. As seen in this study, the data after the interpolation took place (Appendix 20) showed a different pattern than what it was supposed have.

Aßfalg, J., Kriegel, H. P., Kröger, P., Kunath, P., Pryakhin, A., & Renz, M. (2006). Similarity search on time series based on threshold queries. In Advances in Database Technology-EDBT 2006 (pp. 276-294). Springer Berlin Heidelberg.

Bakhtiari, A. A., & Hematian, A. (2013). Precision Farming Technology, Opportunities and Difficulty. International Journal for Science and Emerging Technologies with Latest Trends, 5(1), 1-14.

Bala, S. K., & Islam, A. S. (2009). Correlation between potato yield and MODIS-derived vegetation indices. *International Journal of Remote Sensing*, *30*(10), 2491-2507.

Baret, F., Houles, V., & Guérif, M. (2007). Quantification of plant stress using remote sensing observations and crop models: the case of nitrogen management. Journal of Experimental Botany, 58(4), 869-880.

Blackburn, G. A. (1998). Spectral indices for estimating photosynthetic pigment concentrations: a test using senescent tree leaves. International Journal of Remote Sensing, 19(4), 657-675.

Blackburn, G. A. (2007). Hyperspectral remote sensing of plant pigments. Journal of Experimental Botany, 58(4), 855-867.

Blackmer, T. M., Schepers, J. S., & Varvel, G. E. (1994). Light reflectance compared with other nitrogen stress measurements in corn leaves. *Agronomy Journal*, *86*(6), 934-938.

Blackmer, T. M., Schepers, J. S., Varvel, G. E., & Walter-Shea, E. A. (1996). Nitrogen deficiency detection using reflected shortwave radiation from irrigated corn canopies. *Agronomy journal*, *88*(1), 1-5.

Blackmore, B.S. 2003. The role of yield maps in precision farming. Doctoral Thesis, Silsoe College, Cranfield University

Bojović, B., & Marković, A. (2009). Correlation between nitrogen and chlorophyll content in wheat (Triticum aestivum L.). Kragujevac Journal of Science, 31, 69-74.

Bongiovanni, R., & Lowenberg-DeBoer, J. (2004). Precision agriculture and sustainability. Precision Agriculture, 5(4), 359-387.

Bradley, B. A., & Mustard, J. F. (2008). Comparison of phenology trends by land cover class: a case study in the Great Basin, USA. Global Change Biology, 14(2), 334-346.

Brisco, B., Brown, R. J., Hirose, T., McNairn, H., & Staenz, K. (1998). Precision agriculture and the role of remote sensing: a review. Can J Remote Sens, 24(3), 315-327.

Broge, N. H., & Mortensen, J. V. (2002). Deriving green crop area index and canopy chlorophyll density of winter wheat from spectral reflectance data.Remote sensing of environment, 81(1), 45-57.

Bretherton, C. S., Smith, C., & Wallace, J. M. (1992). An intercomparison of methods for finding coupled patterns in climate data. Journal of climate, 5(6), 541-560.

Buermann, W., Anderson, B., Tucker, C. J., Dickinson, R. E., Lucht, W., Potter, C. S., & Myneni, R. B. (2003). Interannual covariability in Northern Hemisphere air

temperatures and greenness associated with El Niño-Southern Oscillation and the Arctic Oscillation. Journal of Geophysical Research: Atmospheres (1984–2012), 108(D13).

Canisius, F., Turral, H., & Molden, D. (2007). Fourier analysis of historical NOAA time series data to estimate bimodal agriculture. International Journal of Remote Sensing, 28(24), 5503-5522.

Carter, G. A. (1991). Primary and secondary effects of water content on the spectral reflectance of leaves. American Journal of Botany, 916-924.

Carter, M. R. (Ed.). (1993). Soil sampling and methods of analysis. CRC Press.

Cha, S. H. (2007). Comprehensive survey on distance/similarity measures between probability density functions. City, 1(2), 1.

Chappelle, E. W., Kim, M. S., & McMurtrey III, J. E. (1992). Ratio analysis of reflectance spectra (RARS): an algorithm for the remote estimation of the concentrations of chlorophyll a, chlorophyll b, and carotenoids in soybean leaves. Remote Sensing of Environment, 39(3), 239-247.

Chen, Y., Nascimento, M. A., Ooi, B. C., & Tung, A. (2007, April). Spade: On shape-based pattern detection in streaming time series. In Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on (pp. 786-795). IEEE.

Clevers, J. G., & Gitelson, A. A. (2013). Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and-3. International Journal of Applied Earth Observation and Geoinformation, 23, 344-351.

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

Daughtry, C. S. T., Walthall, C. L., Kim, M. S., De Colstoun, E. B., & McMurtrey Iii, J. E. (2000). Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote sensing of Environment*, *74*(2), 229-239.

Darvishzadeh, R. (2008, May). Hyperspectral remote sensing of vegetation parameters using statistical and physical models. ITC.

Demotes-Mainard, S., Boumaza, R., Meyer, S., & Cerovic, Z. G. (2008). Indicators of nitrogen status for ornamental woody plants based on optical measurements of leaf epidermal polyphenol and chlorophyll contents. Scientia horticulturae, 115(4), 377-385.

Ding, H., Trajcevski, G., Scheuermann, P., Wang, X., & Keogh, E. (2008). Querying and mining of time series data: experimental comparison of representations and distance measures. Proceedings of the VLDB Endowment, 1(2), 1542-1552.

Dodge, S., Weibel, R., & Laube, P. (2009). Exploring movement-similarity analysis of moving objects. SIGSPATIAL Special, 1(3), 11-16.

Dumas, J.B.A. 1831. Procedes de L'analyse Organique. Ann. Chim. Phys., 47: 198–205.

Efron, B., & Gong, G. (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician*, *37*(1), 36-48.

Elvidge, C. D., & Chen, Z. (1995). Comparison of broad-band and narrow-band red and near-infrared vegetation indices. Remote sensing of environment, 54(1), 38-48.

Foresight. (2011). Migration and global environmental change: future challenges and opportunities. Final project report. Futures. London: Government Office for Science.

Frentzos, E., Gratsias, K., & Theodoridis, Y. (2007, April). Index-based most similar trajectory search. In Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on (pp. 816-825). IEEE.

Frost, J. (2014, January 23). Regression Analysis: How to Interpret S, the Standard Error of the Regression. Retrieved February 2, 2015.

Gebbers, R., & Adamchuk, V. I. (2010). Precision agriculture and food security.Science, 327(5967), 828-831.

Gerendás, J., & Pieper, I. (2001). Suitability of the SPAD meter and the petiole nitrate test for nitrogen management in nursery potatoes. In Plant Nutrition (pp. 716-717). Springer Netherlands.

Gevaert, C. M., Suomalainen, J., Tang, J., & Kooistra, L. (2015). Generation of Spectral–Temporal Response Surfaces by Combining Multispectral Satellite and Hyperspectral UAV Imagery for Precision Agriculture Applications. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.

Gianquinto, G., Goffart, J. P., Olivier, M., Guarda, G., Colauzzi, M., Dalla Costa, L., ... & Mackerron, D. K. L. (2004). The use of hand-held chlorophyll meters as a tool to assess the nitrogen status and to guide nitrogen fertilization of potato crop. Potato Research, 47(1-2), 35-80.

Gianquinto, G., Orsini, F., Sambo, P., & D'Urzo, M. P. (2011). The use of diagnostic optical tools to assess nitrogen status and to guide fertilization of vegetables. HortTechnology, 21(3), 287-292.

Gitelson, A. A., & Merzlyak, M. N. (2003). Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of plant physiology*, *160*(3), 271-282.

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

Goffart, J. P., Olivier, M., & Frankinet, M. (2008). Potato crop nitrogen status assessment to improve N fertilization management and efficiency: past–present–future. Potato Research, 51(3-4), 355-383.

Guyot, G., & Baret, F. (1988, April). Utilisation de la haute resolution spectrale pour suivre l'etat des couverts vegetaux. In *Spectral Signatures of Objects in Remote Sensing* (Vol. 287, p. 279).

Hatfield, J. L., & Prueger, J. H. (2010). Value of using different vegetative indices to quantify agricultural crop characteristics at different growth stages under varying management practices. Remote Sensing, 2(2), 562-578.

Haboudane, D., Miller, J. R., Tremblay, N., Zarco-Tejada, P. J., & Dextraze, L. (2002). Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. Remote sensing of environment, 81(2), 416-426.

Haboudane, D., Miller, J. R., Pattey, E., Zarco-Tejada, P. J., & Strachan, I. B. (2004). Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. Remote sensing of environment, 90(3), 337-352.

Hansen, B. (1989). Determination of nitrogen as elementary N, an alternative to Kjeldahl. Acta Agriculturae Scandinavica, 39(2), 113-118.

Harmon, T., Kvien, C., Mulla, D., Hoggenboom, G., Judy, J., Hook, J., et al. (2005). Precision agriculture scenario. In P. Arzberger (Ed.), NSF workshop on sensors for environmental observatories. Baltimore, MD, USA: World Tech. Evaluation Center.

Hayes, D. J., & Sader, S. A. (2001). Comparison of change-detection techniques for monitoring tropical forest clearing and vegetation regrowth in a time series. *Photogrammetric engineering and remote sensing*, *67*(9), 1067-1075.

Hochmuth, G. (1994). Plant petiole sap-testing: Guide for vegetable crops. University of Florida Coop. Ext. Serv. Circular 1144.

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote sensing of environment, 83(1), 195-213.

International Assessment of Agricultural Knowledge, Science and Technology for Development (IAASTD). McIntyre BD, Herren HR, Wakhungu J, Watson RT, eds. (2009). Agriculture at a crossroads: a synthesis of the global and sub-global IAASTD reports. Washington, DC: Island Press.

INRA/CIRAD. (2011). Agrimonde: scenarios and challenges for feeding the world in 2050. Versailles: Editions Quae.

Jacquemoud, S., Bacour, C., Poilve, H., & Frangi, J. P. (2000). Comparison of four radiative transfer models to simulate plant canopies reflectance: Direct and inverse mode. Remote Sensing of Environment, 74(3), 471-481.

Jacquemoud, S., & Ustin, S. L. (2008). Modelling leaf optical properties. Photobiological Sciences Online.

Jemison Jr, J. M., & Fox, R. H. (1988). A quick-test procedure for soil and plant tissue nitrates using test strips and a hand-held reflectometer 1. Communications in Soil Science & Plant Analysis, 19(14), 1569-1582.

Ji-hua, M., & Bing-fang, W. (2008). Study on the crop condition monitoring methods with remote sensing. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 37(B8), 945-950.

Jones, H. G., & Vaughan, R. A. (2010). Remote sensing of vegetation: principles, techniques, and applications. Oxford university press.

Kjeldahl, JGCT (1883). Neue method zur Bestimmung des stickstoffs in organized chen körpern. Fresenius' Journal of Analytical Chemistry, 22 (1), 366-382.

Kooistra, L., Beza, E., Verbesselt, J., & van den Borne, J. (2012). Integrating remote-, close range-and in-situ sensing for high-frequency observation of crop status to support precision agriculture. In Sensing a Changing World II, 1–7. Wageningen, the Netherlands

Laurent, F., & Lancelot, F. (1999). Petiole sap NO3 test: calibration and validation, or threshold values to diagnose crop N status and decide a second N application. In EAPC Abstracts of Conference Papers and Posters. 14th Triennial Conference of the European Association for Potato Research, Sorrento, Italy (pp. 462-463).

Lawlor, D. W. (1993). Photosynthesis: molecular, physiological and environmental processes (No. Ed. 2). Longman scientific & technical.

Lhermitte, S., Verbesselt, J., Verstraeten, WW, & Coppin, P. (2010). A pixel-based regeneration index using time series similarity and spatial context. Photogram Metric Engineering & Remote Sensing, 76 (6), 673-682.

Lhermitte, S., Verbesselt, J., Verstraeten, W. W., & Coppin, P. (2011). A comparison of time series similarity measures for classification and change detection of ecosystem dynamics. Remote Sensing of Environment, 115(12), 3129-3152.

Liao, T. W. (2005). Clustering of time series data—a survey. Pattern recognition, 38(11), 1857-1874.

Liaghat, S., & Balasundram, S. K. (2010). A review: The role of remote sensing in precision agriculture. American Journal of Agricultural and Biological Sciences, 5(1), 50.

Lillesand, T. M., Kiefer, R. W., & Chipman, J. W. (2004). Remote sensing and image interpretation (No. Ed. 5). John Wiley & Sons Ltd.

Lobell, D. B., Schlenker, W., & Costa-Roberts, J. (2011). Climate trends and global crop production since 1980. Science, 333(6042), 616-620.

MacKerron, D. K. L., Young, M. W., & Davies, H. V. (1995). A critical assessment of the value of petiole sap analysis in optimizing the nitrogen nutrition of the potato crop. Plant and soil, 172(2), 247-260.

Markwell, J., Osterman, J. C., & Mitchell, J. L. (1995). Calibration of the Minolta SPAD-502 leaf chlorophyll meter. Photosynthesis Research, 46(3), 467-472.

Martin, R.J. (1995). Evaluation of rapid field methods for determining the nitrogen status of potato crops. Proceedings Agronomy Society of New Zealand 25, 91-95.

Massart, DL, Vandeginste, BGM, Buydens, LMC, De Jong, GJ, Lewi, PJ, and Smeyers-Verbeke, J. (1997). Handbook of Chemometrics and Qualimetrics (Data Handling in Science and Technology). *Vol. 20Elsevier, Amsterdam*.

Montemurro, F. (2010). Are organic N fertilizing strategies able to improve lettuce yield, use of nitrogen and N status? Journal of plant nutrition, 33(13), 1980-1997.

Mróz, M., & Sobieraj, A. (2004). Comparison of several vegetation indices calculated on the basis of a seasonal SPOT XS time series, and their suitability for land cover and agricultural crop identification. Technical Sciences, 7, 39-66.

Mulla, D. J., & Schepers, J. S. (1997). Key processes and properties for site-specific soil and crop management. In F. J. Pierce, & E. J. Sadler (Eds.), the state of site specific management for agriculture (pp. 1e18). Madison, WI, USA: ASA/CSSA/SSSA

Mulla, D. J. (2013). Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. Biosystems Engineering, 114(4), 358-371.

Muñoz-Huerta, R. F., Guevara-Gonzalez, R. G., Contreras-Medina, L. M., Torres-Pacheco, I., Prado-Olivarez, J., & Ocampo-Velazquez, R. V. (2013). A Review of Methods for Sensing the Nitrogen Status in Plants: Advantages, Disadvantages and Recent Advances. Sensors, 13(8), 10823-10843.

Myneni, R. B., & Williams, D. L. (1994). On the relationship between FAPAR and NDVI. Remote Sensing of Environment, 49(3), 200-211.

Nakamura, T., Taki, K., Nomiya, H., Seki, K., & Uehara, K. (2013). A shape-based similarity measure for time series data with ensemble learning. Pattern Analysis and Applications, 16(4), 535-548.

Nelson, L. S. (1984). Column: Technical Aids: The Shewhart Control Chart--Tests for Special Causes. Journal of quality technology, 16(4).

Podur, J., Martell, D. L., & Knight, K. (2002). Statistical quality control analysis of forest fire activity in Canada. Canadian Journal of Forest Research, 32(2), 195-205.

Pontes, F. V., Carneiro, M. C., Vaitsman, D. S., da Rocha, G. P., da Silva, L. I., Neto, A. A., & Monteiro, M. I. C. (2009). A simplified version of the total kjeldahl nitrogen method using an ammonia extraction ultrasound-assisted purge-and-trap system and ion chromatography for analyses of geological samples. Analytica chimica acta, 632(2), 284-288.

Pretty, J. (2008). Agricultural sustainability: concepts, principles and evidence. Philosophical Transactions of the Royal Society B: Biological Sciences, 363(1491), 447-465.

Rains, G. C., & Thomas, D. L. (2009). Precision farming: an introduction.

Richardson, A. J., & Wiegand, C. L. (1977). Distinguishing vegetation from soil background information. PE & RS, 43 (1977), 1541 – 1552.

Robert, P. C. (2002). Precision agriculture: a challenge for crop nutrition management. Plant and soil, 247(1), 143-149.

Ryan, T. P. (2011). Statistical methods for quality improvement. John Wiley & Sons.

Sakamoto, T., Yokozawa, M., Toritani, H., Shibayama, M., Ishitsuka, N., & Ohno, H. (2005). A crop phenology detection method using time-series MODIS data. Remote sensing of environment, 96(3), 366-374.

SAS Institute Inc. (1999), SAS/ETS User's Guide, Version 8, Cary, NC: SAS Institute Inc.

Schlerf, M., Atzberger, C., & Hill, J. (2005). Remote sensing of forest biophysical variables using HyMap imaging spectrometer data. *Remote Sensing of Environment*, *95*(2), 177-194.

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

Serrà, J., & Arcos, J. L. (2014). An empirical evaluation of similarity measures for time series classification. Knowledge-Based Systems.

Shahid, R., Bertazzon, S., Knudtson, M. L., & Ghali, W. A. (2009). Comparison of distance measures in spatial analytical modeling for health service planning. BMC health services research, 9(1), 200.

Silleos, N. G., Alexandridis, T. K., Gitas, I. Z., & Perakis, K. (2006). Vegetation indices: advances made in biomass estimation and vegetation monitoring in the last 30 years. Geocarto International, 21(4), 21-28.

Swain, K. C., Jayasuriya, H. P. W., & Salokhe, V. M. (2007). Low-altitude remote sensing with unmanned radio-controlled helicopter platforms: A potential substitution to satellite-based systems for precision agriculture adoption under farming conditions in developing countries.

Taiz, L., & Zeiger, E. (2010). Plant physiology. Sunderland, MA: Sinauer Associates.

The Hague Conference on Agriculture, Food Security and Climate Change. (2010). Chair's summary.

The Royal Society. 2009. Reaping the benefits: Science and the sustainable intensification of global agriculture. London: The Royal Society.

Thomas, J. R., & Gausman, H. W. (1977). Leaf reflectance vs. leaf chlorophyll and carotenoid concentrations for eight crops. Agronomy journal, 69(5), 799-802.

Tippett, M. K., DelSole, T., Mason, S. J., & Barnston, A. G. (2008). Regression-based methods for finding coupled patterns. Journal of Climate, 21(17), 4384-4398.

Uddling, J., Gelang-Alfredsson, J., Piikki, K., & Pleijel, H. (2007). Evaluating the relationship between leaf chlorophyll concentration and SPAD-502 chlorophyll meter readings. Photosynthesis Research, 91(1), 37-46.

Unkovich, M., Herridge, D., Peoples, M. A. R. K., Cadisch, G., Boddey, B., Giller, K., ... & Chalk, P. (2008). Measuring plant-associated nitrogen fixation in agricultural systems. Australian Centre for International Agricultural Research (ACIAR).

Verbesselt, J., Jönsson, P., Lhermitte, S., Jonckheere, I., van Aardt, J., & Coppin, P. (2006). Relating time-series of meteorological and remote sensing indices to monitor vegetation moisture dynamics. Signal and image processing for remote sensing, 153-173.

Verbesselt, J., Hyndman, R., Newnham, G., & Culvenor, D. (2010). Detecting trend and seasonal changes in satellite image time series. Remote sensing of Environment, 114(1), 106-115.

Vos, J., & Bom, M. (1993). Hand-held chlorophyll meter: a promising tool to assess the nitrogen status of potato foliage. Potato Research, 36(4), 301-308.

Waheed, T., Bonnell, R. B., Prasher, S. O., & Paulet, E. (2006). Measuring performance in precision agriculture: CART—A decision tree approach.Agricultural water management, 84(1), 173-185.

Westcott, M. P., Rosen, C. J., & Inskeep, W. P. (1993). Direct measurement of petiole sap nitrate in potato to determine crop nitrogen status. Journal of plant nutrition, 16(3), 515-521.

Wiegand, C. L., Richardson, A. J., Escobar, D. E., & Gerbermann, A. H. (1991). Vegetation indices in crop assessments. Remote Sensing of Environment, 35(2), 105-119.

Wu, J., Wang, D., Rosen, C. J., & Bauer, M. E. (2007). Comparison of petiole nitrate concentrations, SPAD chlorophyll readings, and QuickBird satellite imagery in detecting nitrogen status of potato canopies. Field Crops Research, 101(1), 96-103.

Wu, C., Niu, Z., Tang, Q., & Huang, W. (2008). Estimating chlorophyll content from hyperspectral vegetation indices: Modeling and validation. Agricultural and forest meteorology, 148(8), 1230-1241.

Xin, J., Yu, Z., van Leeuwen, L., & Driessen, P. M. (2002). Mapping crop key phenological stages in the North China Plain using NOAA time series images.International Journal of Applied Earth Observation and Geoinformation, 4(2), 109-117.

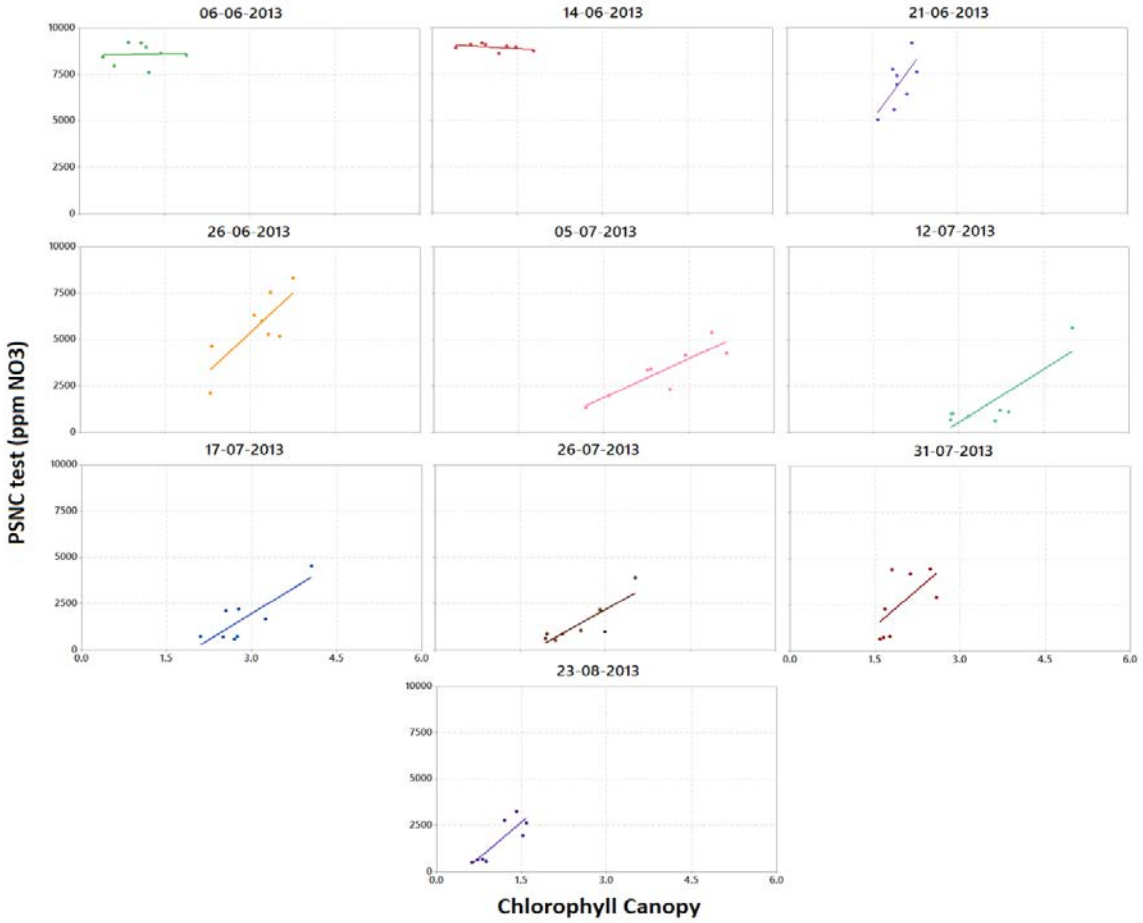
Xu, W. (2014). Spatial Model-aided indoor tracking (Doctoral dissertation, Delft University of Technology, Delft University of Technology).

Zarco-Tejada, P. J., Ustin, S. L., & Whiting, M. L. (2005). Temporal and spatial relationships between within-field yield variability in cotton and high-spatial hyperspectral remote sensing imagery. Agronomy Journal, 97(3), 641-653.

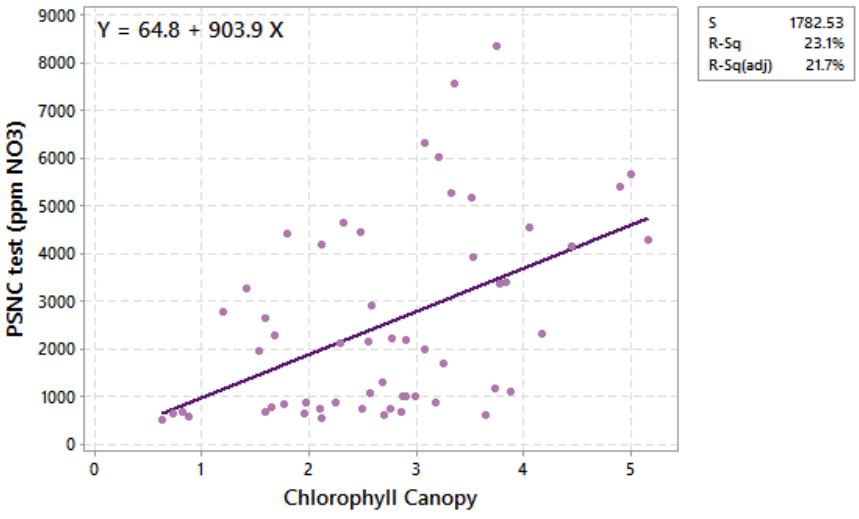
Zhang, H., Smeal, D., Arnold, R. N., & Gregory, E. J. (1996). Potato nitrogen management by monitoring petiole nitrate level. Journal of plant nutrition, 19(10-11), 1405-1412.

Zhang, X., Friedl, M. A., Schaaf, C. B., Strahler, A. H., Hodges, J. C., Gao, F., ... & Huete, A. (2003). Monitoring vegetation phenology using MODIS. Remote sensing of environment, 84(3), 471-475.

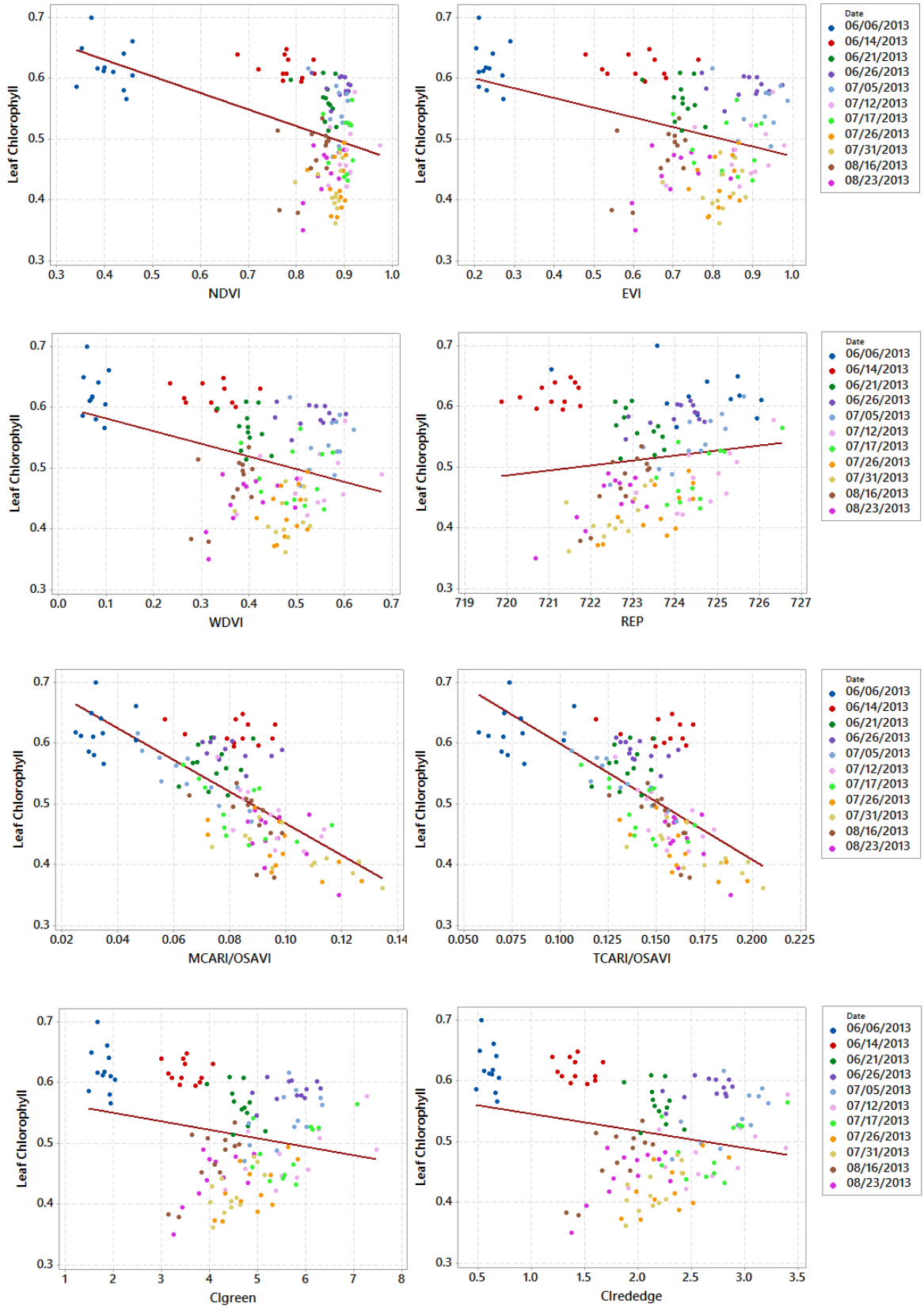
Appendix 1: Relationship between Chlorophyll Canopy and Nitrogen Nitrate (PSNC test) based on the specific observation time.



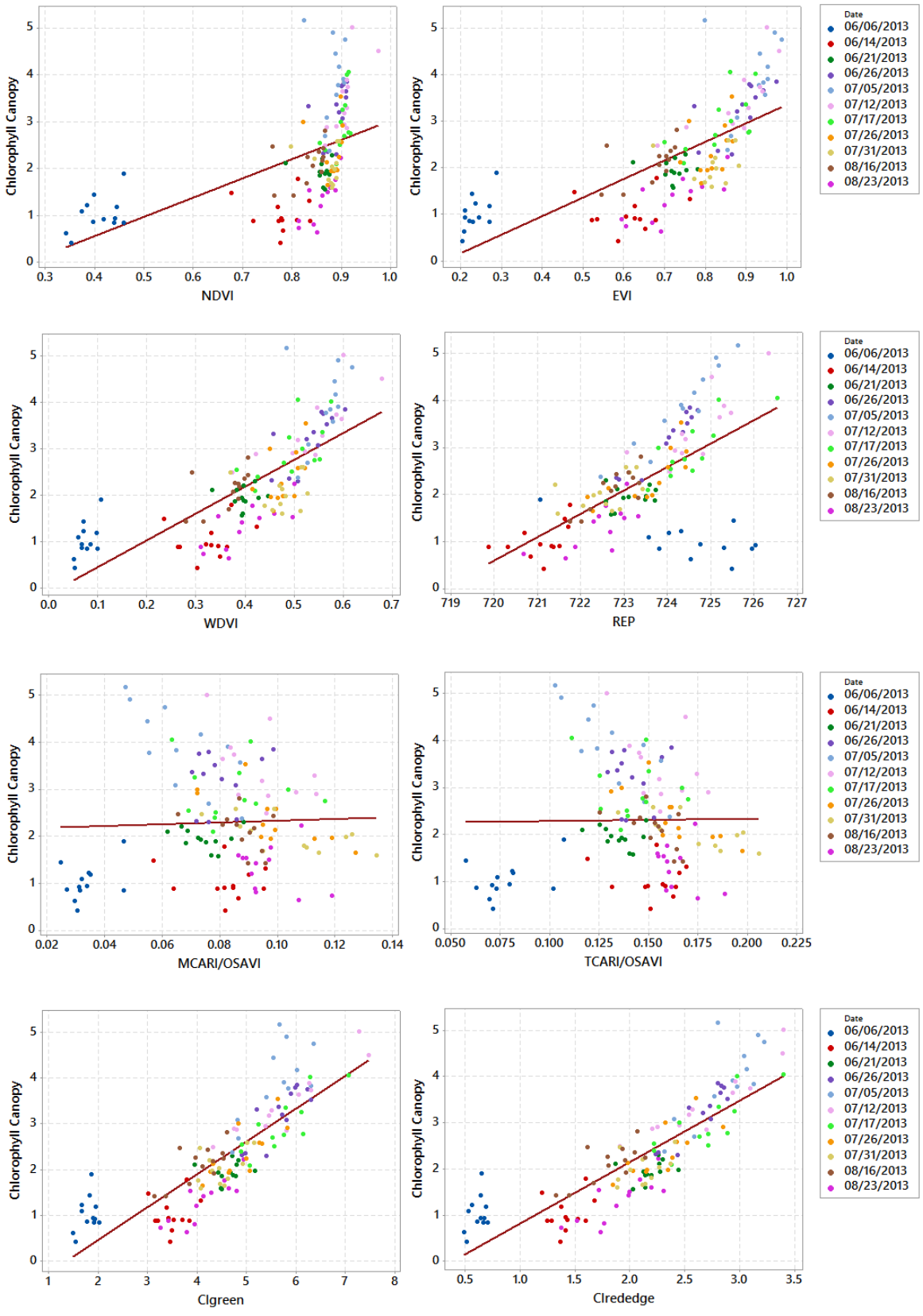
Appendix 2: Recalculation of the relationship between PSNC test and Chlorophyll Canopy (exclude observation on 6 – 21 June 2013)



Appendix 3: Relationships between leaf chlorophyll (g/m^2) and eight different vegetation indices.

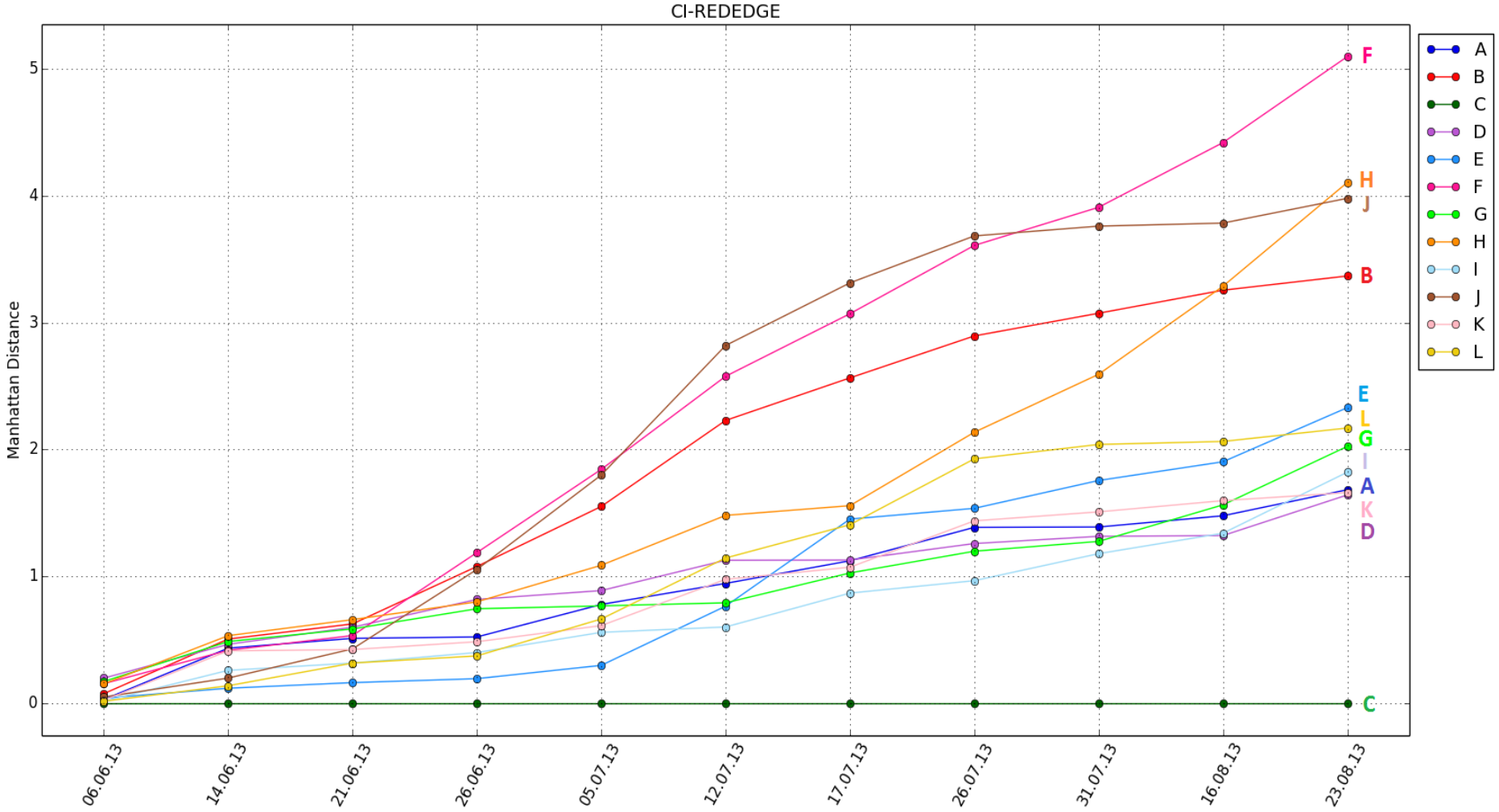


Appendix 4: Relationships between chlorophyll canopy and eight different vegetation indices.

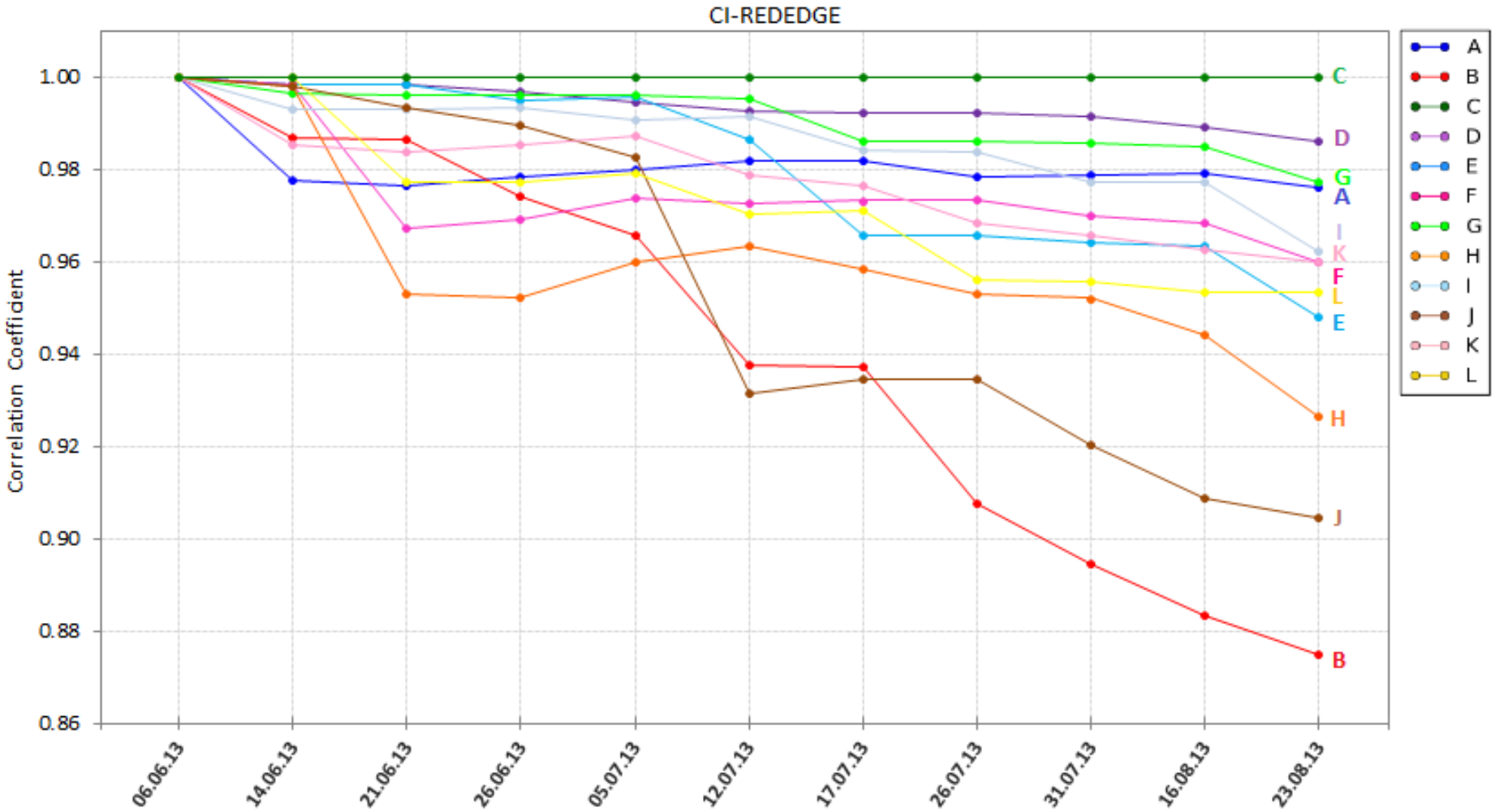


Appendix 5: Yield harvester map from the potato field

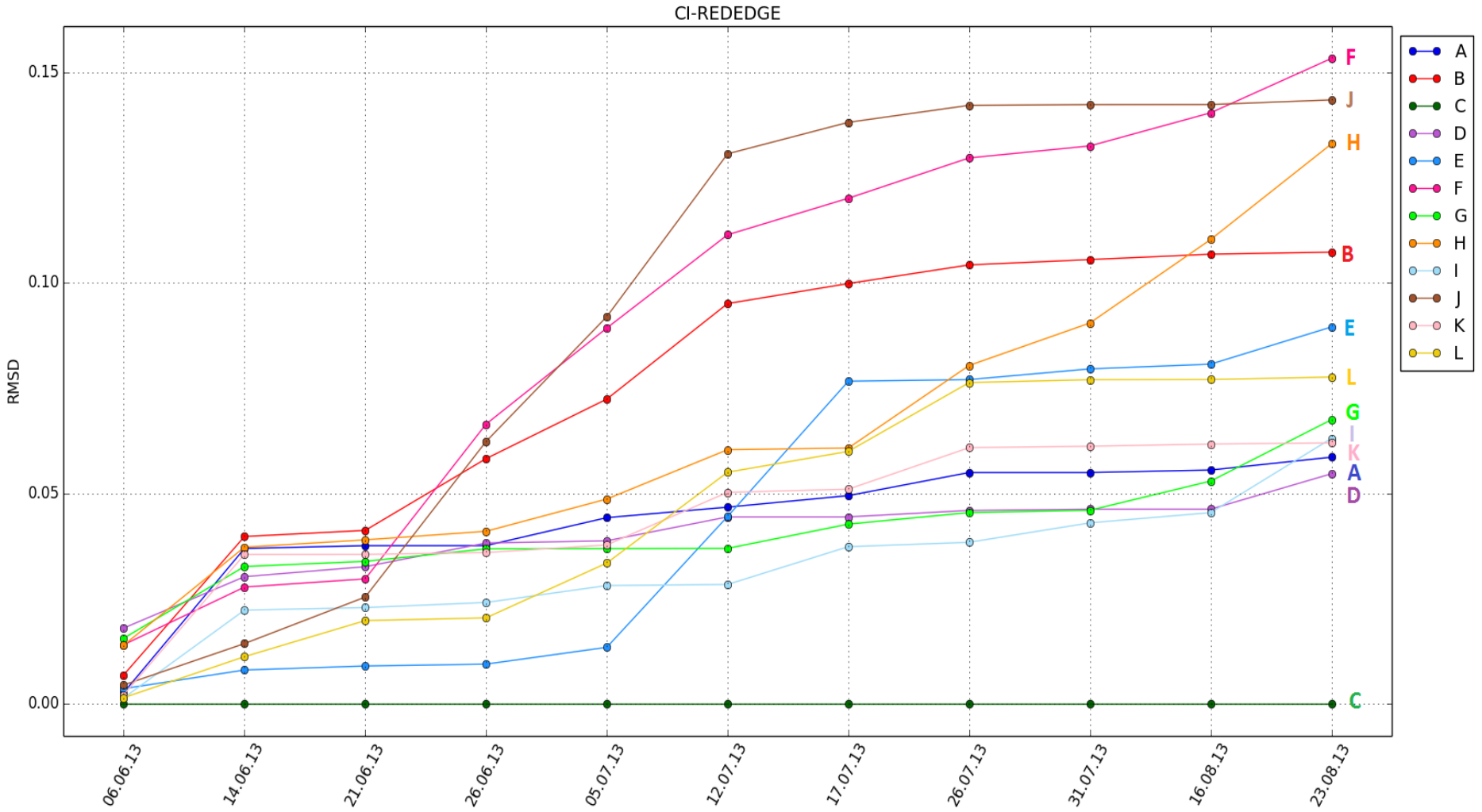
Appendix 6: $CI_{red\text{-}edge}$ time series similarity measures using Manhattan Distance method. Plot C was assigned as the reference plot (maximum yield approach). The X-axis shows weekly observations over the growing season. The Y-axis shows the Manhattan distance between the experimental plots and the reference plot (Plot C) for each observation date.



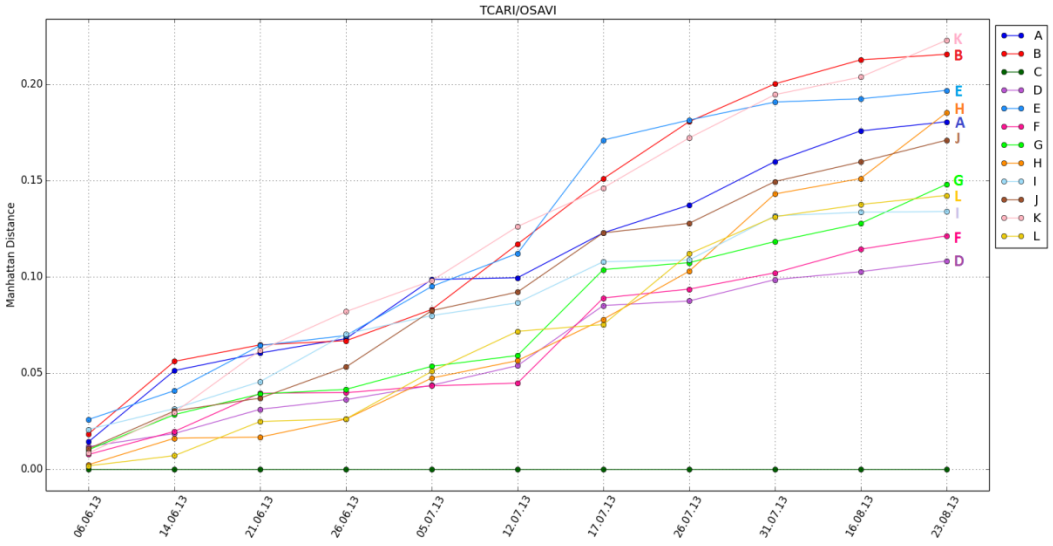
Appendix 7: $CI_{red\text{-}edge}$ time series similarity measures using Correlation Coefficient method. Plot C was assigned as the reference plot (maximum yield approach). The X-axis shows weekly observations over the growing season. Y-axis shows the correlation coefficient between the experimental plots and the reference plot (Plot C) for each observation date.

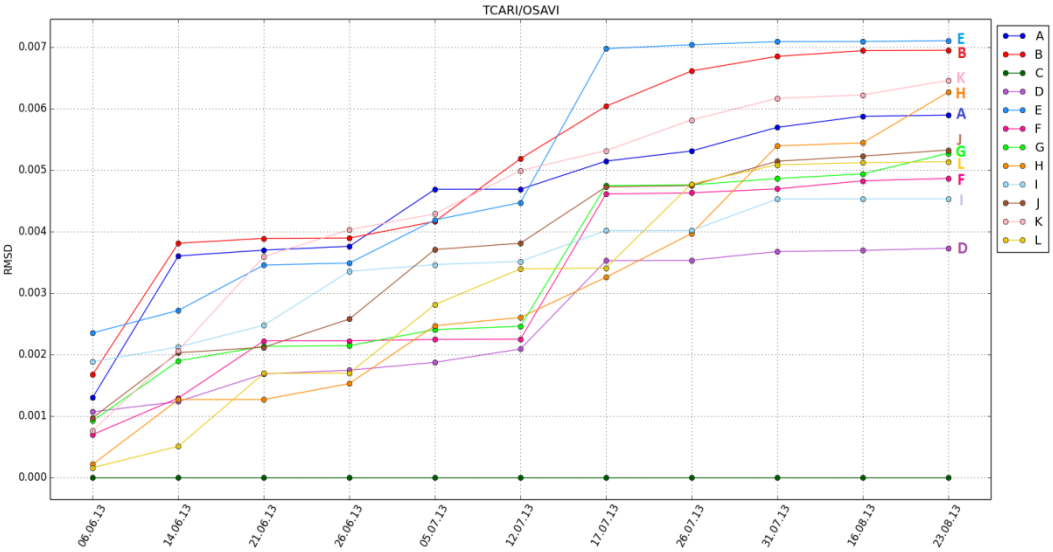


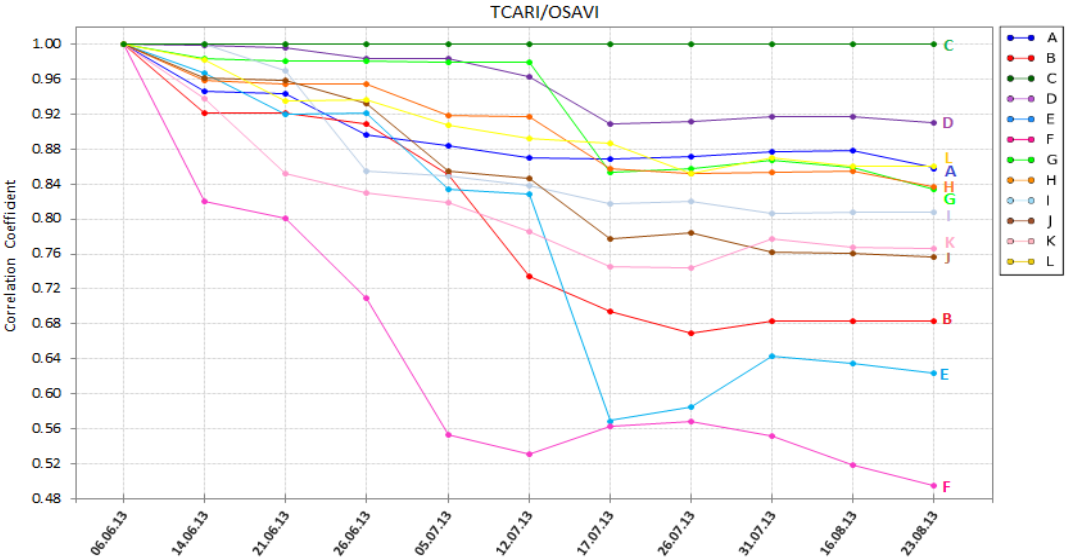
Appendix 8: $CI_{red\text{-}edge}$ time series similarity measures using RMSD method. Plot C was assigned as the reference plot (maximum yield approach). The X-axis shows weekly observations over the growing season. Y-axis shows the RMSD difference between the experimental plots and the reference plot (Plot C) for each observation date.



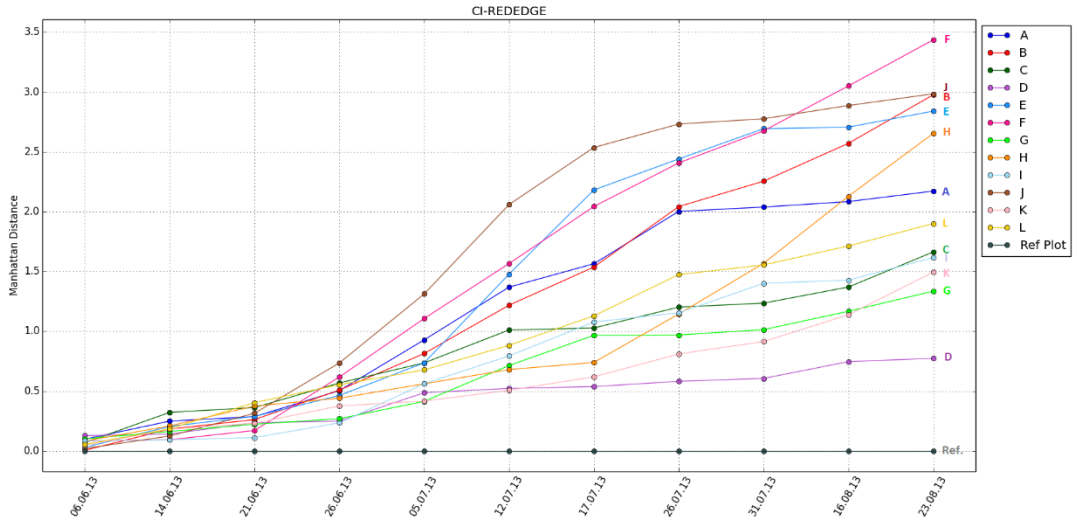
Appendix 9: TCARI/OSAVI time series similarity measures using Manhattan Distance, RMSD methods and Coefficient Correlation. The Y-axis shows the distance measures between the experimental plots and the reference plot (Plot C) for each observation date.

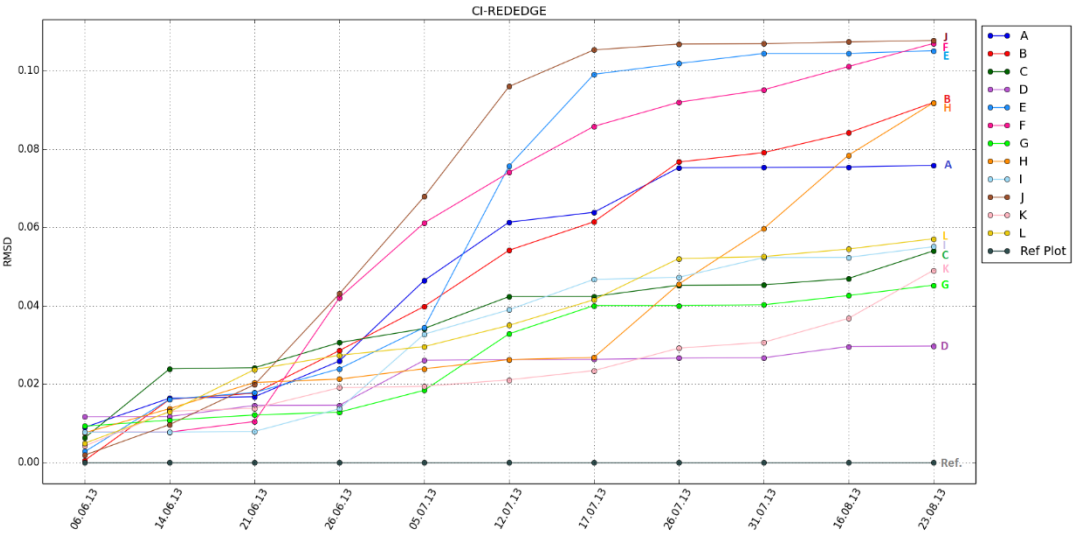


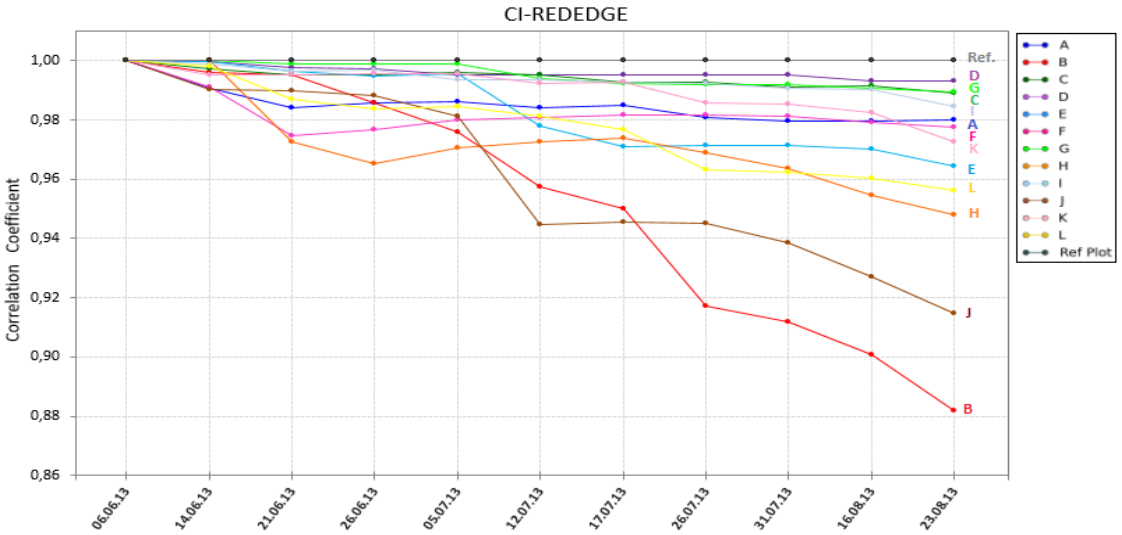




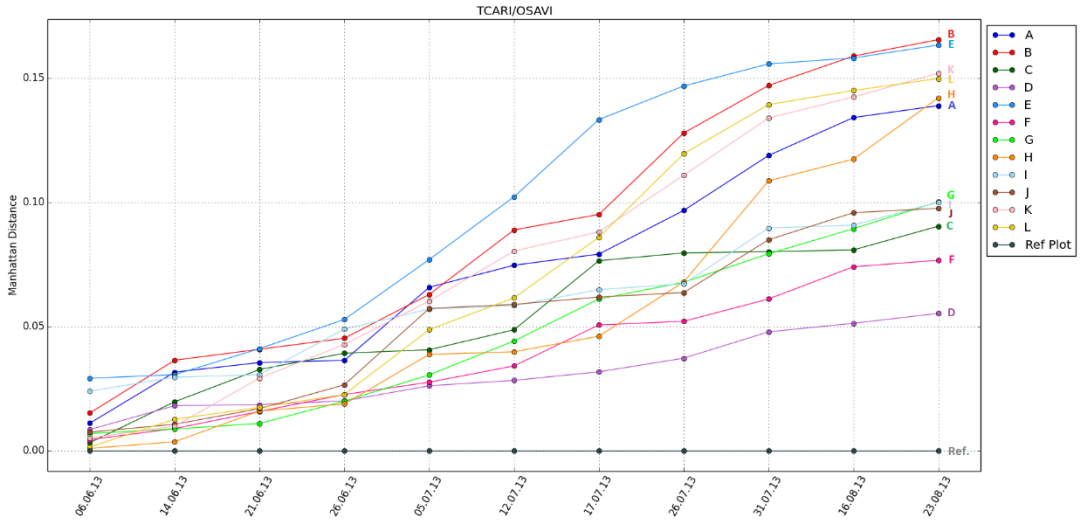
Appendix 10: $CI_{red-edge}$ time series similarity measures using Manhattan Distance, RMSD methods and Coefficient Correlation. The Y-axis shows the distance measures between 12 experimental plots and the reference plot (mean curve approach) for each observation date.

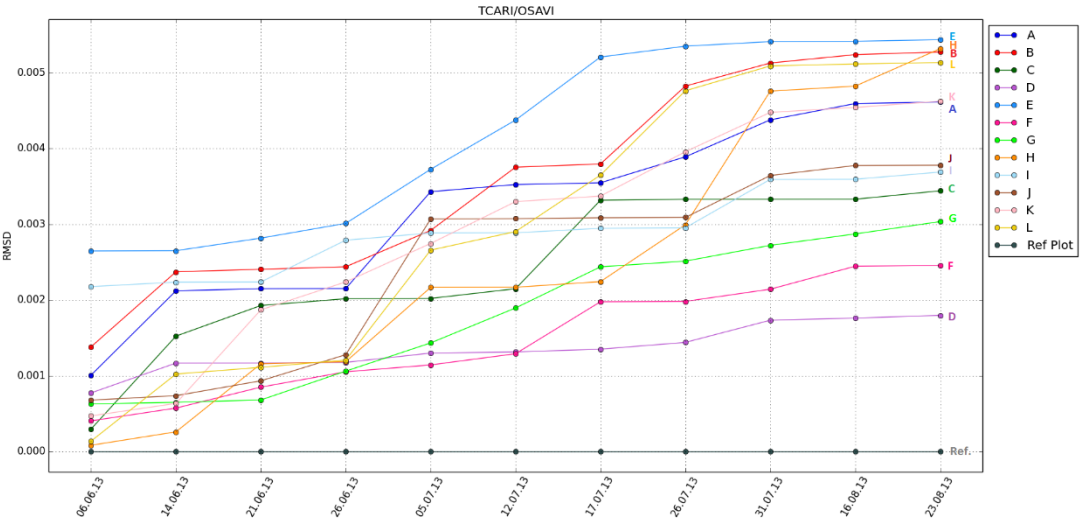


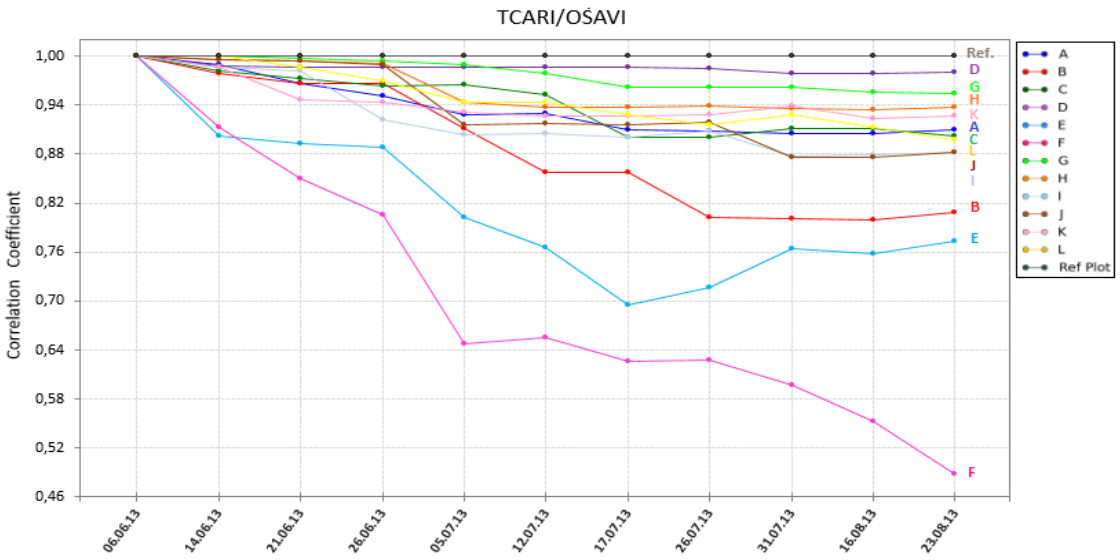




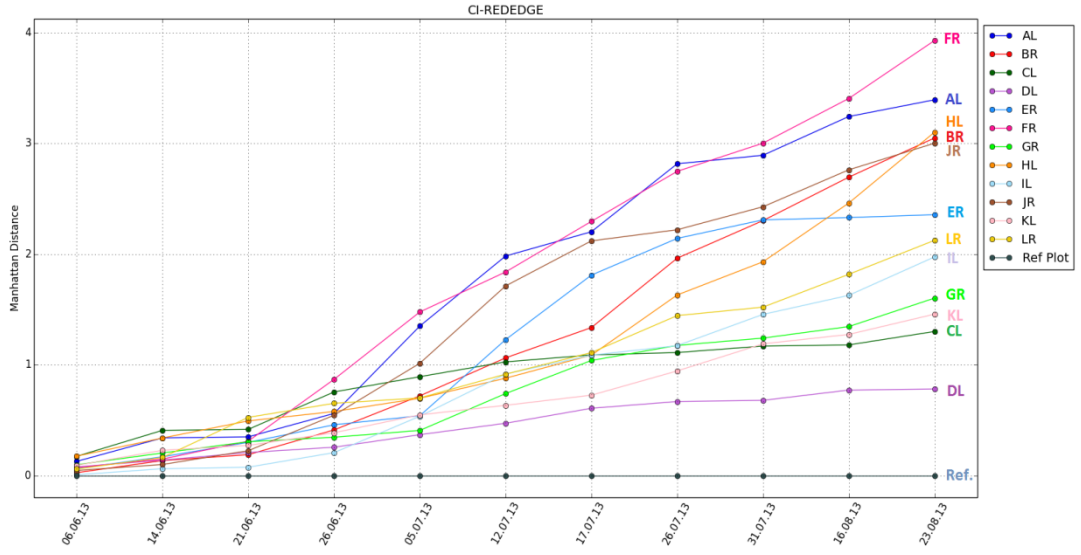
Appendix 11: TCARI/OSAVI time series similarity measures using Manhattan Distance, RMSD methods and Coefficient Correlation. The Y-axis shows the distance measures between 12 experimental plots and the reference plot (mean curve approach) for each observation date.

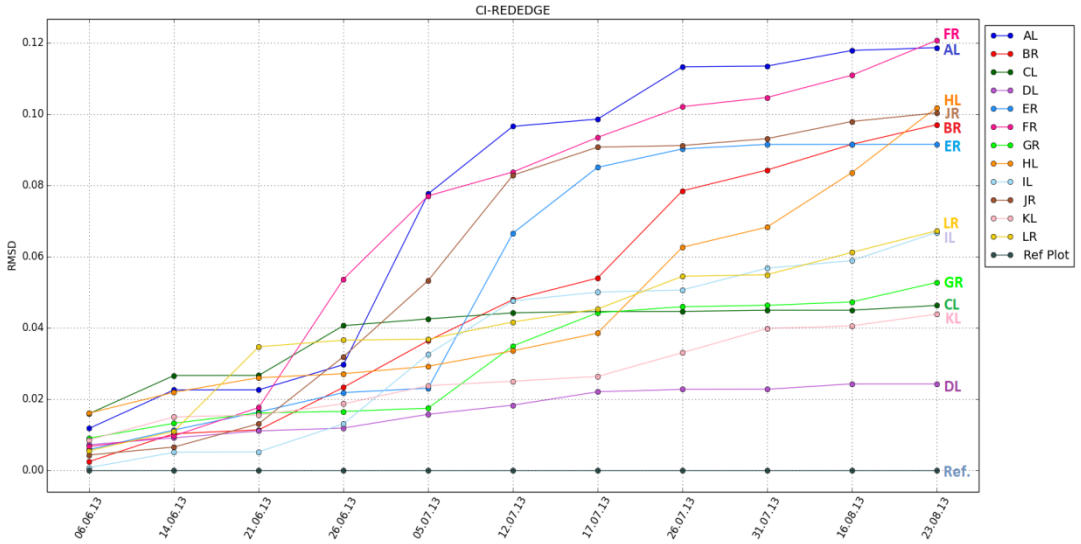


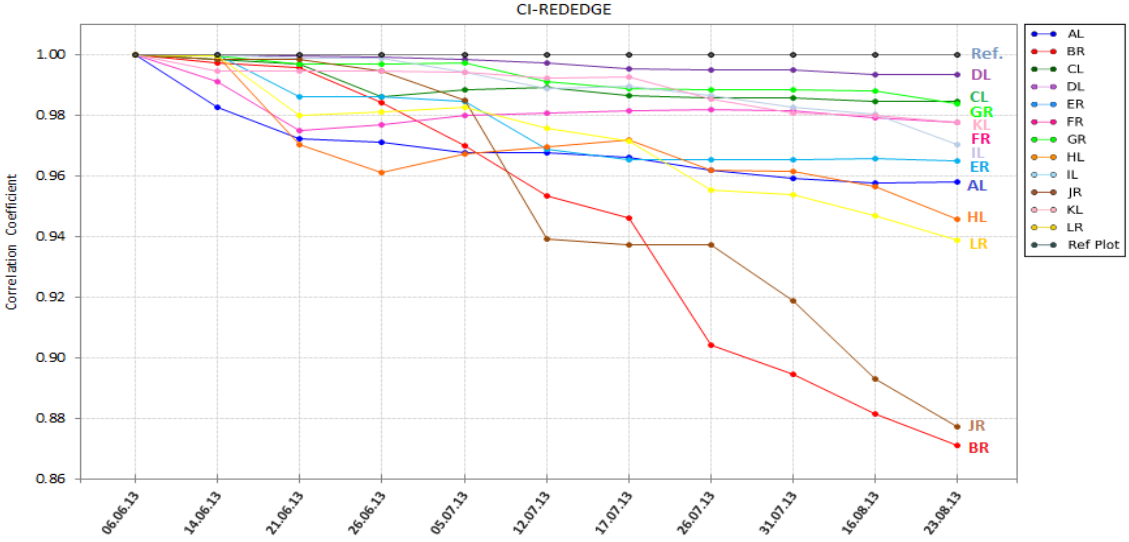




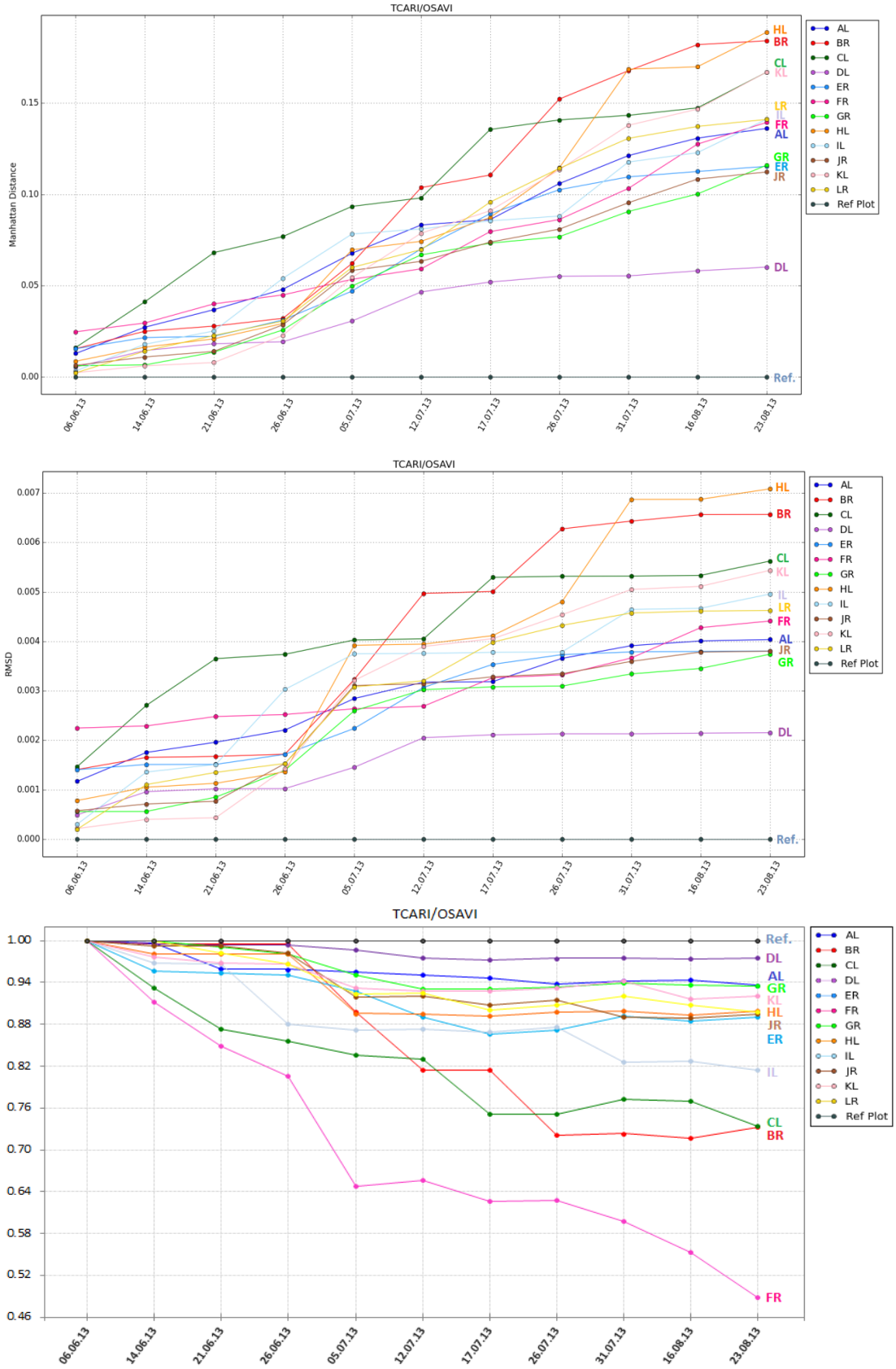
Appendix 12: $CI_{red\text{-}edge}$ time series similarity measures using Manhattan Distance, RMSD methods and Coefficient Correlation for 12 experimental subplots. The Y-axis shows the distance measures between the subplots and the reference plot (mean curve approach) for each observation date.



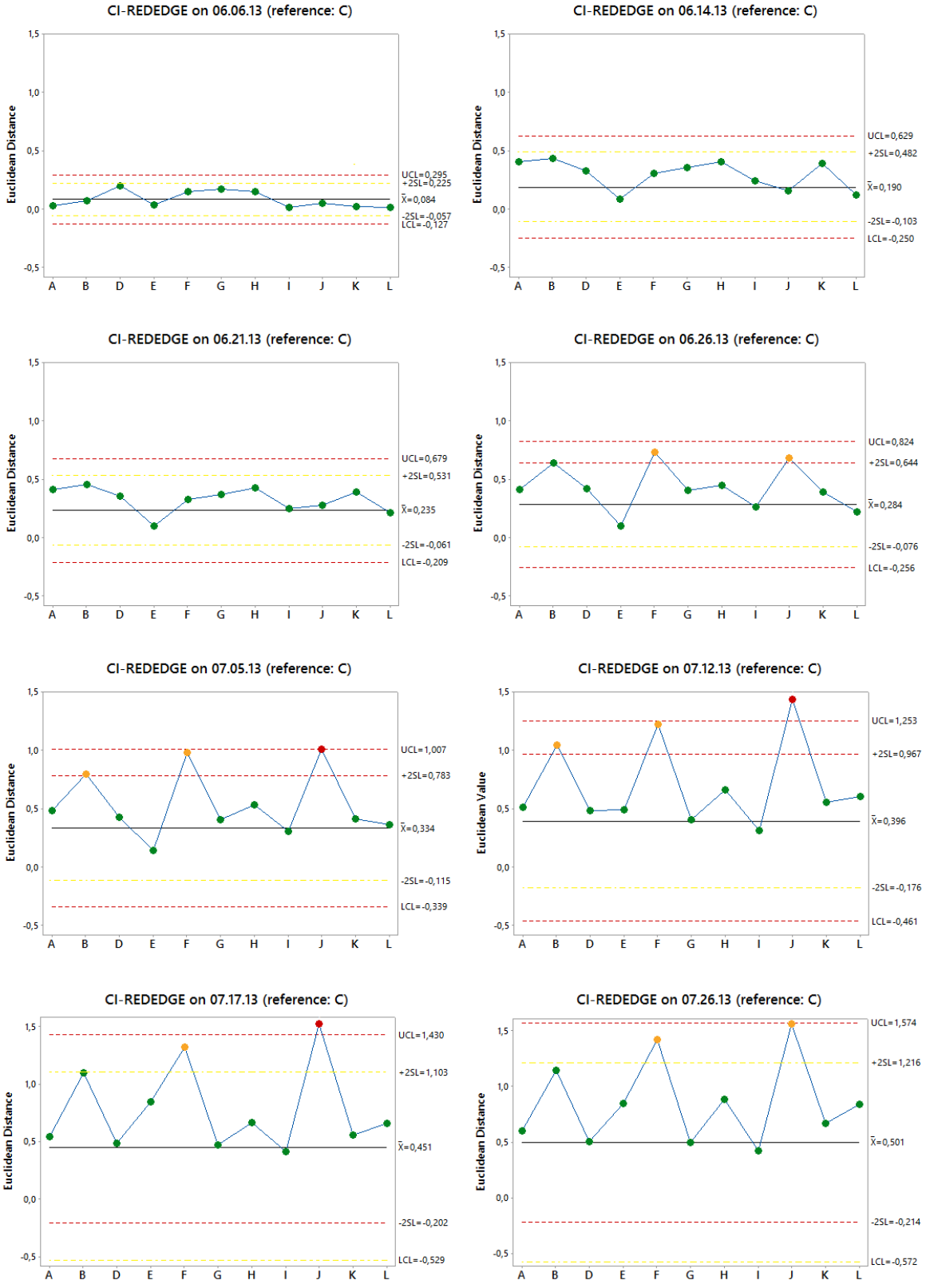




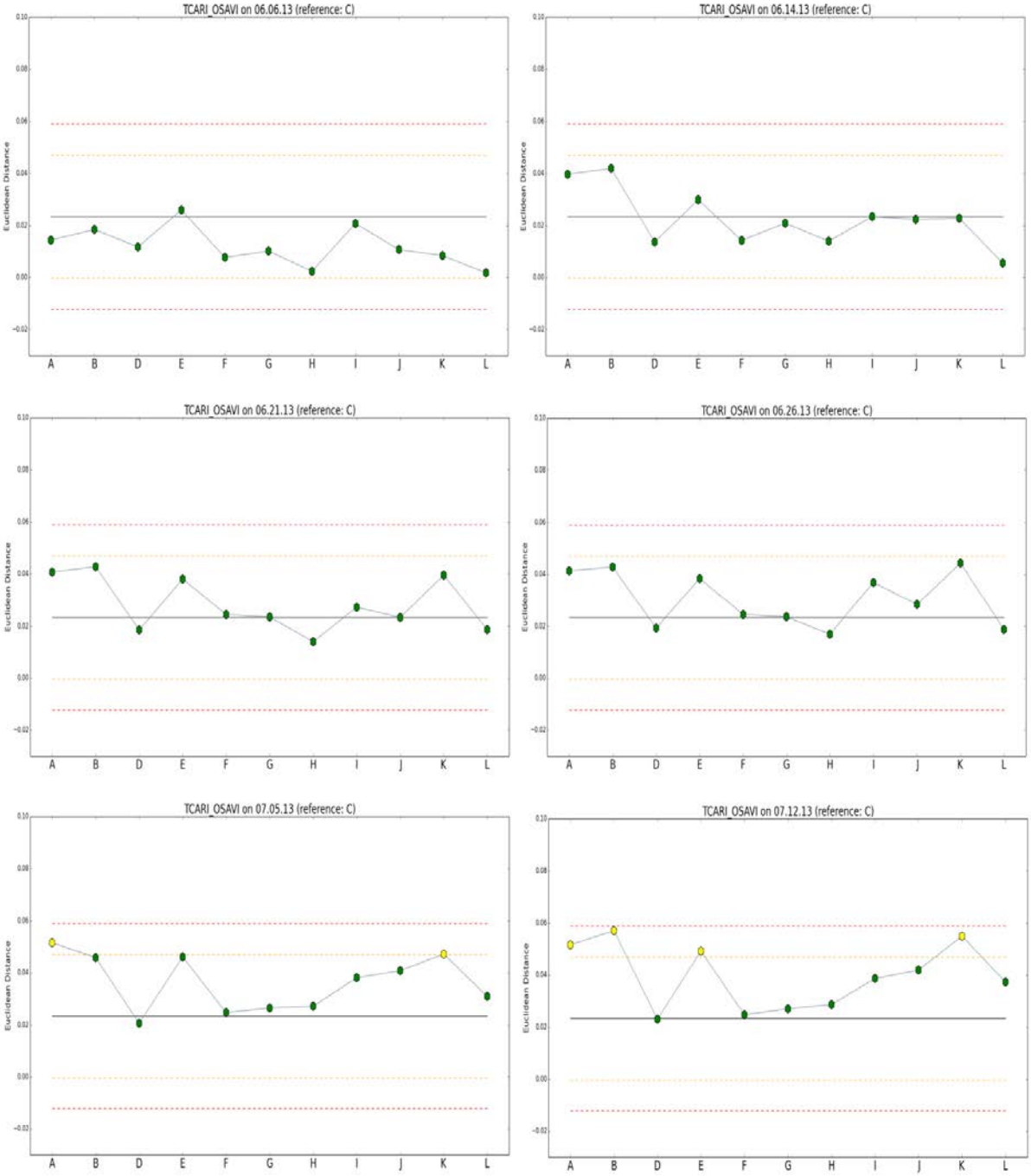
Appendix 13: TCARI/OSAVI time series similarity measures using Manhattan Distance, RMSD methods and Coefficient Correlation for 12 experimental subplots. The Y-axis shows the distance measures between the subplots and the reference plot (mean curve approach) for each observation date.



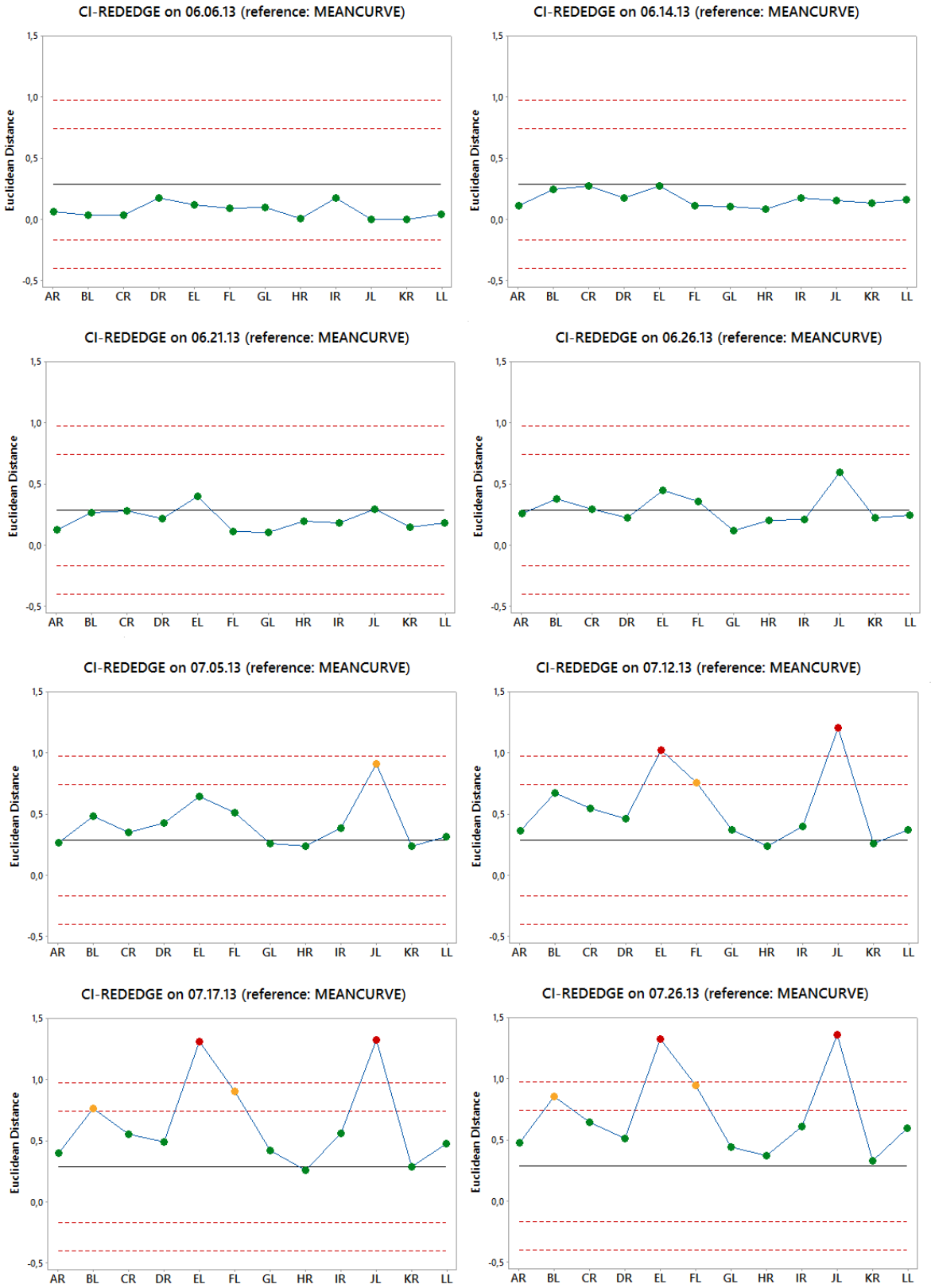
Appendix 14: $CI_{red\text{-}edge}$ control charts of the first eight observations date using cumulative threshold method (the thresholds were recalculated for each observation using cumulative calculation). Control charts were constructed using Python.



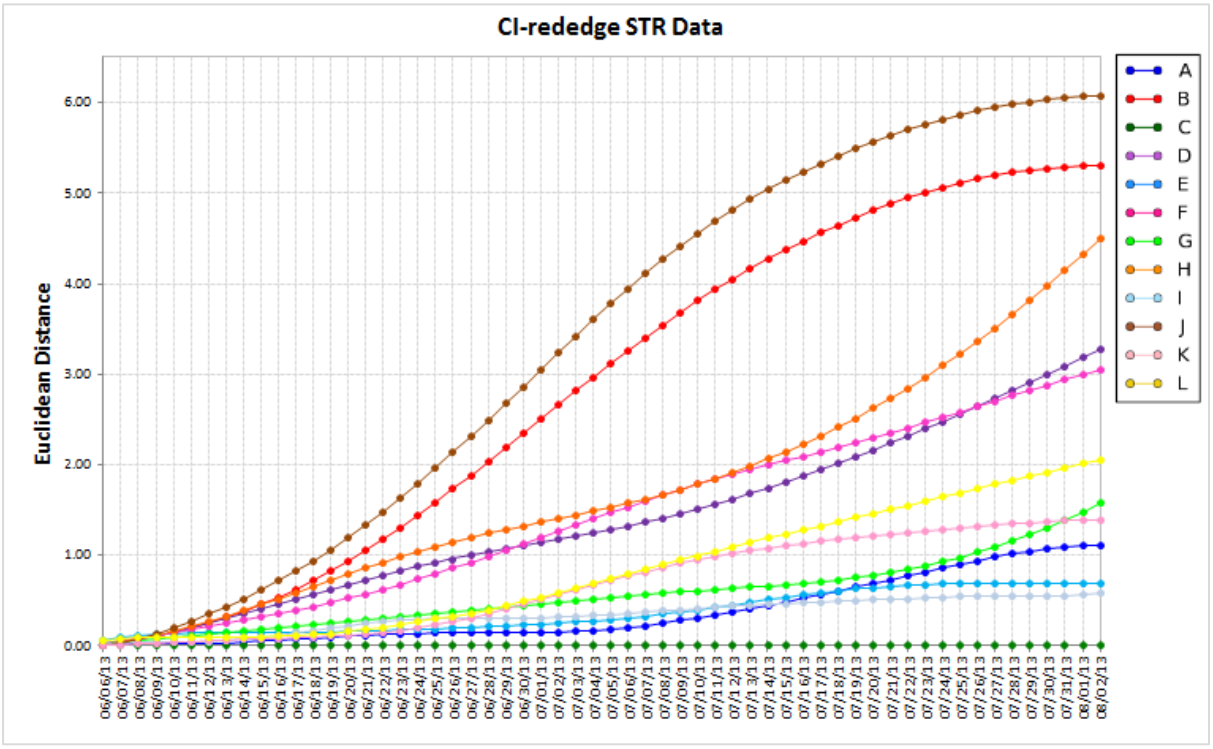
Appendix 15: TCARI/OSAVI Control charts of the first six observations date which show the status of each plot (Constructed using Python Programming Language). The black line indicates control line (CL), yellow lines indicate warning limits (UWL&LWL) and red lines indicate control limits (UCL&LCL). First 4 observations (6 June, 14 June, 21 June and 26 June) are used as calibration dataset.

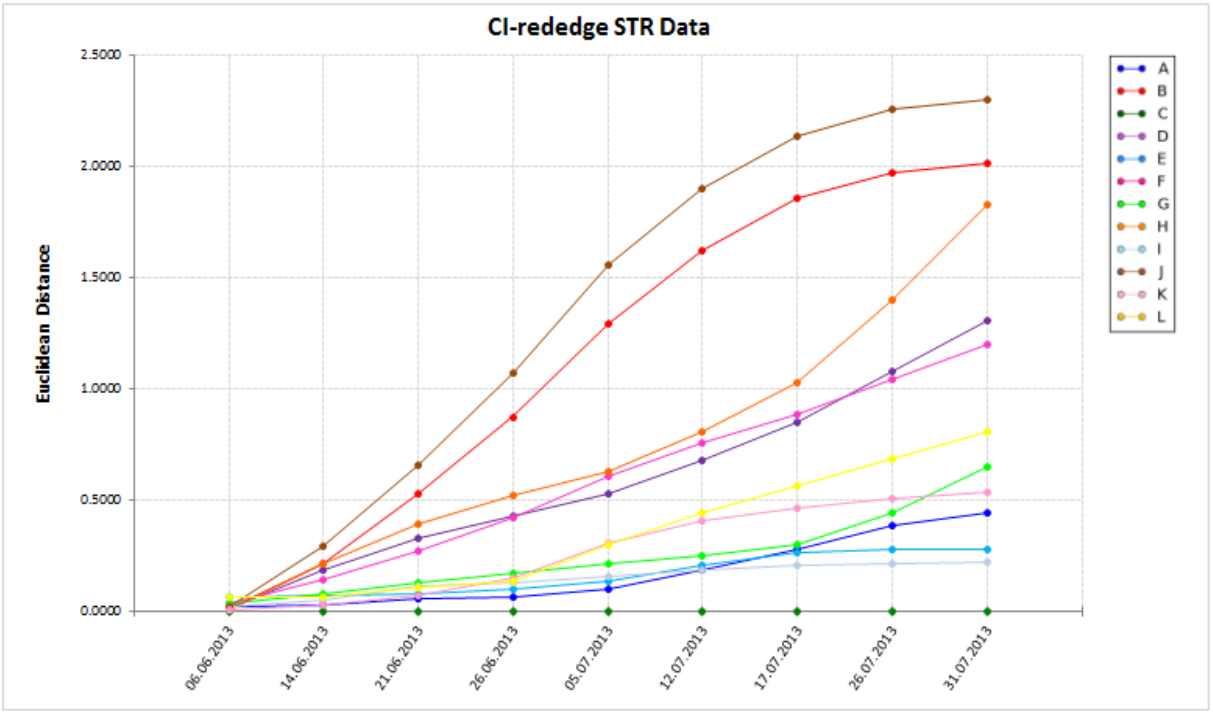


Appendix 16: $CI_{red\text{-}edge}$ control charts of the first eight observations date for 12 experimental subplots. The thresholds were calculated using the first five observation date data.

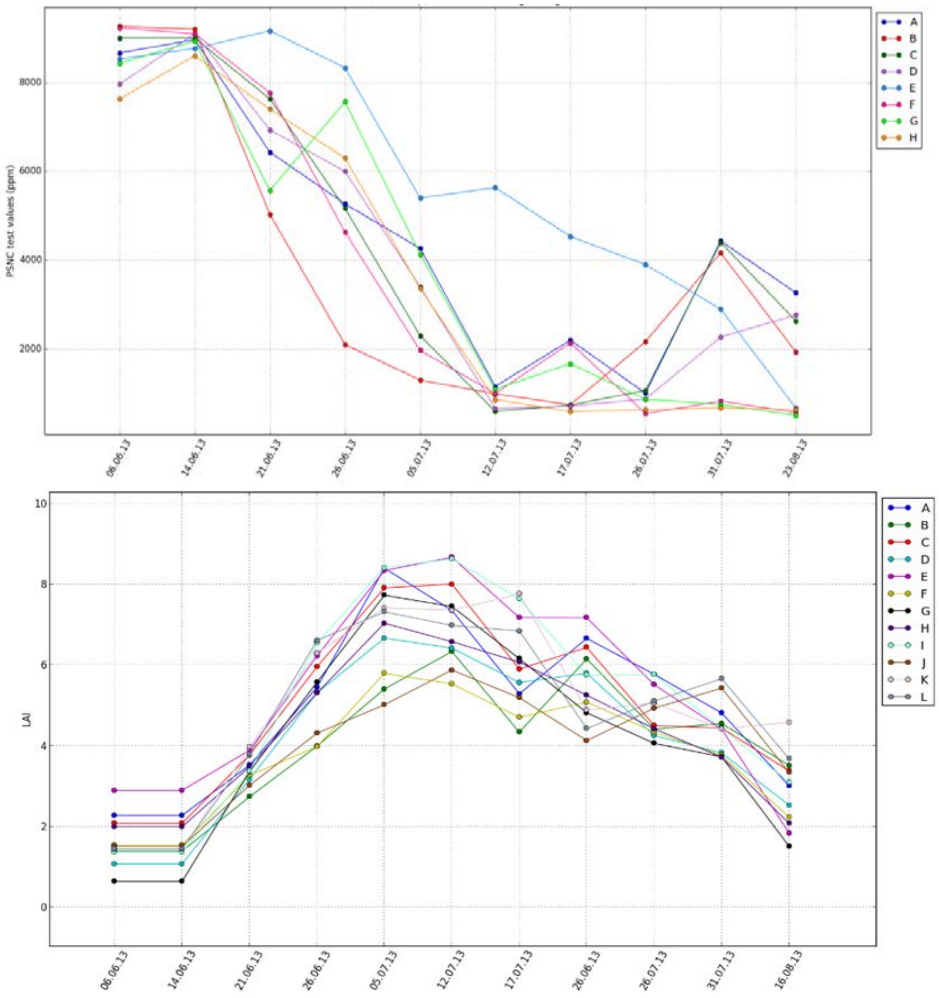


Appendix 17: The Euclidean distance for $CI_{red\text{-}edge}$ time series from STR dataset over the growing season: daily values (top) and weekly values (down).

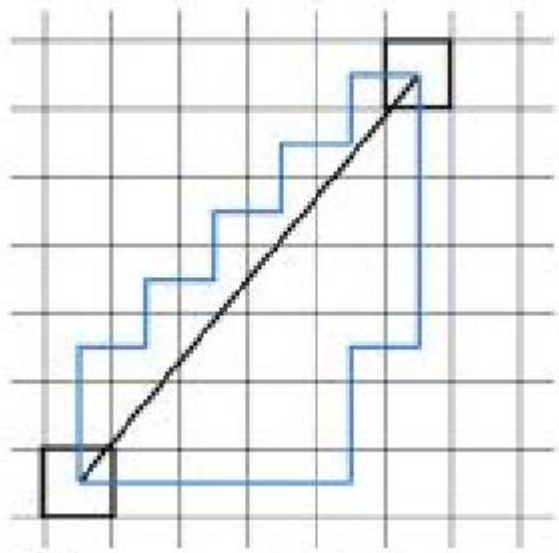




Appendix 18: The PSNC test values and LAI over the growing season.



Appendix 19: The illustration of Euclidean distance and Manhattan distance. The black line is Euclidean distance and the blue lines are the possible Manhattan distance.



Appendix 20: The $CI_{red\text{-}edge}$ time series from the STR data and Cropscan data. The $CI_{red\text{-}edge}$ interpolated values after 17 July 2013 from the STR dataset were overestimated.