Centre for Geo-Information

Thesis Report GIRS-2017-25 August 2017

Predicting sugarcane yields in Rwanda using remotely sensed imagery

J.P.A. Goris

WAGENINGEN
UNIVERSITY & RESEARCH

PREDICTING SUGARCANE YIELDS IN RWANDA USING REMOTELY SENSED IMAGES

J.P.A. Goris

Registration number: 910122270070

Supervisor:

Dr. ir. Sytze de Bruin

A thesis submitted in partial fulfilment of the degree of Master of Science

at Wageningen University and Research Centre,

The Netherlands.

August 2017

Wageningen, The Netherlands

Thesis code number: GRS-80436
Thesis Report: GIRS- 2017-25
Wageningen University and Research Centre
Laboratory of Geo-Information Science and Remote Sensing

ABSTRACT

Sugarcane is one of the most important crops in Rwanda. Demand on the national market is claimed to rise, while domestic production only covers 30% of the demand. Precision agriculture can help increase sugarcane yields. Monitoring of crop status could lead to a more realistic prediction of the yield than current yield predictions. In this research, remotely sensed imagery aimed to provide information about crop status. This information was used in a yield prediction model. Multiple models were evaluated and the most suitable model was chosen based on the requirements and available data. Green chlorophyll index data were calculated from the green and near-infrared light reflectance band and used as an indicator of crop status. The bulk harvest of sugarcane stalks was deemed linearly related to the green chlorophyll index times the incident photosynthetically active radiation. Historic data about yield per hectare per zone were used to calibrate the regression model. As a result of poorly calibrated remotely sensed imagery and lack of data covering the full growth cycle, the predicted yields are not accurate enough for Kabuye Sugar Work to be useful. In future endeavours, crop status over the whole growth cycle and correctly calibrated remotely sensed imagery are necessary to predict sugarcane yield more accurately.

KEYWORDS

Yield prediction, Rwanda, sugarcane, remote sensing, vegetation index, green chlorophyll index, , Kabuye sugar works

TABLE OF CONTENTS

1. INTRODUCTION

1.1. CONTEXT AND BACKGROUND

In Rwanda, agriculture accounted for a third of the GDP in 2012 (RDB, 2016). For 84% of the population, agricultural activities are the main source of income (Ansoms, 2008). One of the crops that is important in Rwanda is sugarcane. From sugarcane, multiple products are made, of which sugar is the most important. The Rwanda development board predicted a rise in demand of sugar in Rwanda, while production currently seems to cover only 30% of the national demand (RDB, 2016). Rwanda has one factory that processes sugarcane into sugar. Sugarcane cultivation provides both employment in the factory and on agricultural lands. Currently 10000 jobs exist as the result of the sugarcane industry (Veldman and Lankhorst, 2011). Opportunities lie within increasing production of sugarcane. Increased production could lead to the creation of jobs, more food security. Furthermore, with an increased production of sugarcane, Rwanda has to import less sugar from other countries. This means less money will leave the country (Ansoms, 2008).

Rwanda has a suitable climate for the cultivation of sugarcane. For sugarcane to reach optimum yield, it needs between 2000 mm and 2300 mm of precipitation, together with a well-drained, fertile soil (Yamane, 2016). The cultivation of sugarcane in Rwanda is done with ratoon cycles. A ratoon cycle is when the sugarcane is harvested, leaving the lowest part of the stem in the field. From the old stem, a new sugarcane crop will grow, that will give a slightly lower yield next harvest (Henry and Ellis, 1995). The maximum sucrose level of sugarcane is reached after 18 months in Rwanda (Veldman and Lankhorst, 2011). During these 18 months, improved capacity to monitor crops can help farmers respond to diseases such as smut, cope with floods and identify potential other plant stress (de Bruin, 2015). Furthermore, monitoring crop maturity could improve the estimation of the right harvesting time. Together with optimizing water use and fertiliser, this will eventually lead to an improved yield (N.L. Ministry of Foreign Affairs, 2012).

Crop monitoring techniques can be a valuable data source for precision agriculture. Yield prediction can benefit from monitoring negative environmental impacts and crop status which can be acted upon. The sugarcane crop can be monitored using remotely sensed imagery to know when maturity is reached. Additionally, monitoring of crop status could lead to a more realistic prediction of the yield than traditional yield predictions. In that sense, crop monitoring is a useful tool to help improve sugarcane yield prediction.

In the eight months the factory is crushing cane, it requires a minimum input of sugarcane to keep it running, but also has a limit to the amount of sugarcane that can be processed. Harvesting of sugarcane is done by hand and thus requires a lot of labourers. Therefore, distribution of harvesting the sugarcane plots is an important facet of the sugarcane processing, The prediction of the yield plays an important role in harvesting the sugarcane and keeping the only sugar factory in Rwanda, Kabuye Sugar Works (KSW), running. After sugarcane has been harvested, deterioration of sugarcane stalks happens. In many sugar mills, time-lag between harvest and milling of cane ranges between three to seven days, resulting in deterioration of sugarcane stalks and consequently in a loss of sugar production (Solomon et al.,

2001). Knowing when and how much sugarcane per zone is harvested, could potentially lead to an optimal input-flow of sugarcane, improving the sugarcane value chain.

The “Sugar: make it work” project in Rwanda aims to improve the sugarcane value chain. By investing in water management infrastructure, KSW hopes floods will be reduced and the available marshland will be optimally used. With the introduction of remote sensing technology and a yield prediction model, farmers are supported in making decisions about crop management and harvest timing. This will eventually result in higher yields for farmers.

1.2. Problem definition

1.2.1. Crop models

Yield losses due to negative environmental factors (e.g. flooding, diseases) and poor management skills (such as early harvest) are a limiting factor in optimizing yield. Crop monitoring can be a viable tool to counter this constraint. With yield prediction models, it would be possible to predict at which moment in time a certain amount of sugarcane can be harvested. Crop growth monitoring systems (CGMS) have been developed to monitor crop state and make yield forecasts (van Ittersum et al., 2003).

The development of relations between crop characteristics and remote sensing observations has played an important role in many agricultural studies (Baret and Guyot, 1991). By making measurements of the reflectance of sugarcane canopy, information about the plant can be obtained. Indicators of the biophysical and biochemical parameters of a crop derived from spectral data are also known as vegetation indices (VIs). Numerous VIs have been formed from multispectral reflectance data and have been linked to vegetation variables such as above ground biomass.

Using remote sensing, crop production has been estimated with a simple or multiple regression (Delécolle et al., 1992). Moulin et al. (1998) identified the use of empirical methods, semi-empirical methods and mechanistic methods to vegetation indices with crop characteristics. In chapter 2.3. these methods will be further elaborated.

Many models using VIs as an input have been used in applied- agricultural studies. Many relationship were established between VIs and the leaf area index (LAI) (Baret and Guyot, 1991). VIs have been considered as a measure of crop chlorophyll content, plant height, a measure of crop density and photosynthetically active biomass of a crop (Wiegand et al., 1991). Peng and Gitelson (2011) related VIs to the gross primary production of crops .

Models that are not necessarily dependent on VIs also exist. Since the 1960s crop models have been developed for multiple applications and a variety of vegetation types. For example, SUCROS was developed to get insight into the potential production situation (Van Keulen, 1982). The SUCROS model formed the basis for other models such as WOFOST (Diepen et al., 1989), MACROS (Penning de Vries, 1989) and ORYZA (Bouman, 2001). Spitters and Schapendonk (1990) developed a model called LINTUL, which is based on the light use efficiency (LUE) of crops.

For the prediction of sugarcane yield, currently two main models are in use: the APSIM-sugarcane model and the Canegro model (Lisson et al., 2005). These models can describe the entire crop cycle, but require

much data regarding soil, fertilization, weather, management and parameters of plant growth (Dorigo et al., 2007). Many model components can be added to make the models deal with different conditions, such as water and nitrogen limitation, weeds, pests, diseases and air pollution (van Ittersum et al., 2003).

The crop models are also known as soil-vegetation-atmosphere models (SVAT). Within these models, vegetation variables are linked to atmospheric variables (e.g. weather condition), soil variables (e.g. nutrient availability) and management variables. The output SVAT models commonly is the accumulated biomass(Delécolle et al., 1992).

1.2.2. LINKING REMOTE SENSING TO CROP MODELS

Remotely sensed data can provide information about crop state variables at times when observations are made, making the model more independent on the times of these observations (Delécolle et al., 1992). Remotely sensed data can be obtained by satellite sensors, aerial sensors or terrestrial sensors.

The spectral response of sugarcane canopies can provide insight into biophysical properties of sugarcane such as canopy chlorophyll content, nutrient deficiency, water stress, leaf area index (LAI) and absorbed Photosynthetically Active Radiation (APAR) (Abdel-Rahman and Ahmed, 2008). Although VIs are not a direct measurement of the biophysical properties of a plant, they are strongly related. This suggests that biophysical properties of a plant – such as chlorophyll content – might be replaced by indices.

Gitelson et al. (2006) found that the gross primary product (GPP) of a crop can be a function of LAI- and chlorophyll related VIs and incident photosynthetically active radiation. Chlorophylls are essential pigments that turn light energy into stored chemical energy. The amount of solar radiation that is absorbed by a leaf is related to the amount of pigment that leaf contains, hence the photosynthetic potential and primary production of a crop are determined by the chlorophyll content of that crop (Curran et al., 1990). Gitelson et al. (2005) found a linear relation between the total chlorophyll content and the red-edge chlorophyll index with a R^2 of 0.95.

Dorigo et al. (2007) explored several ways to use remote sensing in crop models.

- The first method is by forcing a remotely sensed observation into a crop model, replacing the state variable of the crop model. At each model time step (week or day) in the simulation, an observation derived from remote sensing is needed and implemented into the crop growth model. To derive values at intermediate time-steps, interpolation techniques such as linear interpolation can be used. With the forcing method, the model forgets its own data and follows the observed state variables. The duration of a time-steps can be chosen according to convenience but has to be consistent. A disadvantage of this method is that the observation errors are also included. These errors are propagated into the model when forcing is used to assimilate the remotely sensed data.
- The second method is by recalibrating. With recalibration of a crop model, selected parameters are adjusted to attain an optimal fit between the remotely sensed observation and the simulated variable. By running the model with various combinations of parameters, the

sensitive and uncertain model parameter values are obtained within realistic ranges. This method is more flexible with assimilating remotely sensed data into the model than the forcing method. A disadvantage is that the model requires a lot of computation time.

- The last method is updating the state variable continuously whenever an remote sensed observation is available. It assumes that an observed state variable at a certain time will improve the accuracy of a state variable on succeeding days. This method is also more flexible in assimilating remotely sensed data into the model and is more flexible in terms of data availability. However, errors in both the state variables and the observed variables can propagate to the final output of a model.

When the methods of either forcing or updating (which is a form of forcing) are applied, substitution of a simulated state variable (e.g. LAI) with an observed one suggests the simulation is flawed in a way that biophysical processes are not described correctly by the model (Moulin et al., 1998). From this point of view, calibration is more suitable when dealing with a mechanistic crop growth model.

In all cases, the usability of remote sensing in crop models is conditioned by spatial resolution, frequency of observation and the observed wavelengths. Many methods have multiple components to simulate the time profile of crop state variables. These components all require input data. For example, SVAT-models require input on soil, vegetation and atmosphere conditions combined with information about management, irrigation, fertilization and plant date (van Ittersum et al., 2003). These data are not available in Rwanda, which imposes a challenge on finding a method that is able to solve the problem of when to harvest and predict the sugarcane yield. Depending on the VIs that have been or will be obtained in Rwanda through the use of remote sensing, a method must be chosen which can implement these VIs to provide a prediction of the optimum harvest time and sugarcane yield.

1.3. Research objectives and questions

The main objective is to develop an approach for predicting sugarcane yields by adapting a model that assimilates remote sensing data.

The research questions are:

1) What are existing methods for predicting sugarcane harvest yield using remotely sensed data?
2) What inputs and parameters are needed for implementing these methods?
3) Given data availability within the “Sugar: make it work” project, which is a viable model for predicting sugarcane yields?
4) Are the available data suitable for obtaining accurate yield predictions using the selected model?

1.4. REPORT OUTLINE

In the following chapter, the different methods for predicting yield using remote sensing are examined. First, the relationship between remote sensing and the crop status is explained. An explanation will be given about how crop status is derived from remotely sensed images. Then, the different methods are disclosed and explained. In chapter 3, the study area and the chosen method are elaborated on. Data availability and the underlying methods to predict yield are described. Furthermore, how to examine the accuracy of the data and the prediction are described in this chapter. In chapter 4, the results of the prediction are presented, as well as the accuracy that is achieved with the chosen model. The results of the data influencing the model and data quality are presented. In chapter 5, the results are discussed and the constraints of the chosen model are evaluated, followed by the discussion of the prediction results. In chapter 6, conclusions are drawn from the results and the discussion. In chapter 7, some recommendations are given for future endeavours to predict sugarcane yields in Rwanda.

2. PREDICTION OF CROP PRODUCTION USING REMOTELY SENSED DATA

2.1. CANOPY STATE VARIABLES

The modelling of agroecosystems started in the 1960s when computational power no longer was an obstacle (Passioura, 1996). The main aim of these models was to acquire information about the underlying processes of crop development and functioning (van Ittersum et al., 2003). Since then, agroecosystem models have been developed with a wide variety of applications, for numerous types of vegetation and at different scale levels. The main applications of crop models are research, as a crop management tool or to make a policy analysis (Boote et al., 1996).

At the same time that crop models were developing, remote sensing techniques were being developed. In 1972, the first civil satellite that made earth observation possible was launched. It demonstrated that remote sensing is a suitable tool to monitor bio- and geophysical processes at global and regional scales (Goward and Williams, 1997). Since then, scientists have retrieved canopy state variables and assimilated remote sensing data into crop models.

To derive canopy state variables, spectral bands are measured in the wavelength range between 400 nanometres (nm) and 2500 nanometres. The spectral reflectance of canopy is governed by three principal factors (Dorigo et al., 2007):

- The optical properties of the vegetation elements.
- The arrangement of the canopy elements (canopy structure).
- The optical properties of the soil beneath the canopy.

Furthermore, atmospheric conditions affect the quantity of the solar radiance incident on the canopy and the radiance reaching the sensor (Richter and Schläpfer, 2002). Another factor that influences the measured spectral reflectance, is the view and illumination geometrical configuration (Baret et al., 2007).

The vegetation elements such as leaves, stems and possibly fruits determine the canopy reflectance (Jacquemoud and Baret, 1990). The chemical constituents such as chlorophyll absorb the solar radiation while structural elements such as cell walls scatter the radiation. Other absorbers of solar radiation are other pigments, water, proteins and cellulose (Faurtyot and Baret, 1997, Jacquemoud and Baret, 1990). The amount of organic matter, minerals, surface roughness and soil water content are factors that determine the soil reflectance (Baumgardner et al., 1986). However, soil reflectance can be neglected when the LAI of a canopy is greater than three (Atzberger et al., 2003).

The arrangement of canopy elements accounts for the directional variation and the magnitude of the reflectance from the canopy. The distribution of the vegetated and non-vegetated areas, the LAI and the leaf angle distribution all contribute to the magnitude and directional distribution of the reflectance of solar radiation by the land surface (Kuusk, 1995).

These are the main factors that determine what the magnitude of these wavelengths is. The reflective remote sensing data can then be used to retrieve biophysical and biochemical variables from the

measured crop. Establishing a relation between reflected radiation and crop state variables is crucial to a quantitative interpretation of the data (Verhoef, 1984).

2.2. Relating remote sensing data to canopy state variables

Methods for retrieving biophysical and biochemical variables of a crop from remotely sensed measurements can be divided in two main categories: statistical methods and physical methods.

2.2.1. Statistical methods

With a statistical method, a relationship is described between the spectral signature of a crop canopy reflectance and a certain biophysical or biochemical variable of that crop.

The measurements made through the means of remote sensing are related to the biophysical or biochemical data that are acquired by measurements in the field or in a laboratory. The relationship can differ per crop and the phenological development stages of a crop. Simple or multiple regressions are the most used – but not exclusive – techniques to link spectral signatures to variables of interest (Clevers, 1989). To filter out background effects and undesired variation effects of reflectance caused by the factors that were discussed in chapter 2.1., VIs are used. Huete et al. (2002) defined VIs as a combination of two or more spectral bands to enhance the contribution of vegetation properties that allow for spatial and temporal comparisons of photosynthetic activity and canopy structure of vegetation.

2.2.2. Physical methods

To estimate the canopy properties through a physical approach, a canopy reflectance model is inverted. A canopy reflectance model calculates the directional reflection as a function of canopy properties. These properties describe measurement and soil conditions (Verhoef, 1984). An example of a canopy reflectance model is the SAIL model. When inverting a canopy reflectance model, a set of properties is sought that leads to the closest match between the bi-directional reflectance simulated by the canopy reflectance model and the measured reflectance at the sensor (Dorigo et al., 2007).

By establishing a relation between the measured reflection of a crop and the biophysical and biochemical variables of a crop, the measured reflection can indirectly be used in crop models to substitute for plant properties.

2.3. YIELD PREDICTION MODELS

Different methods exist for making a yield prediction. Remotely sensed data can be related to canopy state variables measured for that moment in time. The next three subchapters will cover the empirical, semi-empirical and mechanistic methods that were used to predict crop yield in previous researches.

2.3.1. EMPIRICAL MODELS

Empirical methods are based on direct observations, measurements and historic data. Muchow et al. (1998) used historical block productivity data of sugarcane plots, collected over six years. These blocks were classified by district, farm, block, paddock, variety, crop class (planted or ratoon), harvest date and yield to optimize the harvest date and yield of sugarcane. They made a prediction of total sugar yield by averaging the sugar yield per hectare based on the area of farm paddocks within a geographical zone by crop class by harvest month by harvest age. They found the ideal age when the sugarcane should be harvested in that region and the ideal number of ratoons. The authors argue that by applying their method to other countries and regions offers potential in optimizing productivity and profitability in sugarcane production. A drawback of empirical methods to predict yield is that they require extensive historic records and cannot deal with a non-stationary process..

2.3.2. SEMI-EMPIRICAL MODELS

Semi-empirical methods are relying to some extend on observations. Statistical techniques are used to obtain a correlation between the target and the measured spectral signature. Peng and Gitelson (2012) related gross primary production of a crop to VIs especially focusing on chlorophyll content and photosynthetically active radiation (PAR). Crop gross primary production (GPP) is defined as the rate at which a crop captures and stores carbon as biomass (Peng and Gitelson, 2012). The model they used successfully estimated the GPP of maize and wheat. They tested different VIs to estimate the GPP of soybean and found that the red-edge chlorophyll index and green chlorophyll index were the best indices to do this. Gitelson et al. (2006) found a non-species-specific relation between the crop GPP and the product of total crop chlorophyll content (Chl) and the incident photosynthetically active radiation (PAR_{in}). The total crop chlorophyll content can be used to estimate the GPP at that moment. The GPP is proportional to the chlorophyll content multiplied with the incident PAR, therefor the proportional sign $\propto$ is used. They suggested a simple model to remotely assess the instantaneous GPP (Gitelson et al., 2006):

$$GPP \propto Chl \cdot PAR_{in}$$

Furthermore, they found that several VIs were related to the chlorophyll content of a crop. NDVI, TVI, MTVI and WDRVI (Table 1) were used to estimate the LAI which relates to the chlorophyll content (Ciganda et al., 2008). Also, Gitelson et al. (2005) found a significant relationship between the VI green chlorophyll index (CI_{green}) or Red-edge chlorophyll index ($CI_{red\ edge}$) and the crop chlorophyll content of maize and soybean. Both VIs performed well on both crops. In Table 1, the formulation of the VIs are shown.

TABLE 1. VEGETATION INDICES STRONGLY RELATED TO CROP CHLOROPHYLL CONTENT

Index	Formulation	Source
Normalized Difference Vegetation Index (NDVI)	$(\rho_{NIR} - \rho_{red}) / (\rho_{NIR} + \rho_{red})$	(Rouse Jr et al., 1974)
Triangular Vegetation Index (TVI)	$0.5 \times [120 \times (\rho750 - \rho550) - 200 \times (\rho670 - \rho550)]$	(Broge and Leblanc, 2001)
Modified TVI (MTVI)	$1.2 \times [1.2 \times (\rho800 - \rho550) - 2.5 \times (\rho670 - \rho550)]$	(Haboudane et al., 2004)
Wide Dynamic Range Vegetation Index (WDRVI)	$(\alpha \times \rho_{NIR} - \rho_{red}) / (\alpha \times \rho_{NIR} + \rho_{red})$, $0 < \alpha < 1$ $(\alpha \times \rho_{NIR} - \rho_{red}) / (\alpha \times \rho_{NIR} + \rho_{red}) + (1 - \alpha) / (1 + \alpha)$, $\alpha = 0.2$	(Gitelson, 2004)
Green Chlorophyll Index (CI green)	$\rho_{NIR} / \rho_{green} - 1$	(Gitelson et al., 2005)
Red edge chlorophyll index (CI red edge)	$\rho_{NIR} / \rho_{red\ edge} - 1$	(Gitelson et al., 2005)

This indicates that chlorophyll- and LAI-related VIs can be used as a proxy for the chlorophyll content in a crop. The model for estimating the GPP at one instant can then be defined as (Gitelson et al., 2006):

$$GPP \propto VI \cdot PAR_{in}$$

Another research by Bastiaanssen and Ali (2003) combined the photosynthetically active radiation model of Monteith and Moss (1977) with the light use efficiency model of Field et al. (1995)) and the model of surface energy balance by Bastiaanssen et al. (1998) to estimate crop growth and forecast crop yield. They suggested a model to calculate the accumulated above ground biomass for a whole growth cycle of a crop:

$$B_{act}^{tot} \approx \varepsilon \sum (APAR(t) \cdot t)\,(kg\ m^{-2})$$

Where B_{act}^{tot} (kg m^{-2}) is the accumulated biomass above ground in a period t and ε (g MJ^{-1}) is the light use efficiency of the crop and t is the period over which the accumulation takes place. They used the NDVI to derive the fraction of PAR that was absorbed by the crop. This fraction of the PAR is called APAR (W m^{-2}). To get the light use efficiency (ε) they used two environmental conditions, namely soil moisture and heat. They were calculated using multiple functions with a number of variables. These variables include daily actual hours of sunshine, monthly mean air temperature, surface albedo, surface temperature and NDVI.

The model of the accumulated biomass above ground by Bastiaanssen and Ali (2003) assumes the light use efficiency is constant during the whole growth cycle. van Heerden et al. (2010) found that the LUE is not constant during the growth period. Hence, the accumulation of biomass is not a linear function. The phenomena that reduced growth in sugarcane are defined as reduced growth phenomena (van Heerden et al., 2010). When only a part of the growth cycle is measured, one has to keep in mind that

the measured crop is in a certain phenological stage. Crops in different phenological stages have different light use efficiency, resulting in a different accumulated biomass for that measured period.

2.3.3. MECHANISTIC MODELS

Mechanistic methods are based on understanding the behaviour of system components. Mechanistic models are made to simulate the time profile of the most important crop state variables such as LAI, biomass and development stage, together with water and nutrient fluxes, energy and carbon (Moulin et al., 1998). To simulate the time-lapse of these variables, information on soil and climate conditions is needed. The SUCROS model designed by Spitters et al. (1989) is an example of a mechanistic model. This model calculates the daily rate of CO_2 assimilation from the incoming radiation, temperature and LAI. Other kinds of models are the LINTUL (light interception and utilisation) models. These kind of models use a linear relation between the biomass produced and the radiation that is intercepted by a crop (Monteith and Moss, 1977). Spitters and Schapendonk (1990) developed a module for the LINTUL model that calculates the crop growth based on the light use efficiency (LUE) concept.

The development of a crop and the portion of assimilates to the leaves are described using empirical relationships. Using inaccurate coefficients to calculate daily crop development may lead to errors in estimating the total biomass (Porter, 1984). Because of the mechanistic nature, these models require data on many of variables that influence the crop growth.

Clevers (1997) used optical remote sensing in combination with the SUCROS model to estimate the yield of sugar beets. Measuring the WDVI (Weighted Difference Vegetation Index) of the crop is done by finding the weighted difference between the NIR and red reflectance. The assumption is made that the ratio of NIR and red reflectance of bare soil is constant. This gives the formula to calculate the WDVI:

$$WDVI = NIR - (c \cdot red)$$

Where *NIR* is the total measured NIR reflectance, *red* is the measured red reflectance and *c* is the slope of soil line, i.e. the ratio between NIR and red reflectance of the soil. Then, the WDVI was used to estimate the LAI. The following formula was applied:

$$LAI = -\frac{1}{\alpha} \cdot \ln(1 - \frac{WDVI}{WDVI_{\infty}})$$

Where α is a parameter that describes at which rate the function runs to its asymptotic value. $WDVI_{\infty}$ is the asymptotic value of the WDVI. The simulated LAI was used as an input in the SUCROS model to initialize and calibrate the model within plausible ranges of parameters. Crop reflectance was measured approximately every 10 days with a ground-based crop scan radiometer

Different methods to predict yields of various crops exist. Choosing the most suitable model to predict sugarcane yield depends on the required model input data and the available data. In the next chapter, the project requirements and the available data are discussed to select the model that fits the given data availability and model requirements.

3. METHODS AND DATA

To develop an approach to predict the sugarcane yield in Rwanda, an inquiry was made into the available data and the requirements of the model.

3.1. THE PROJECT AND DATA

3.1.1. THE PROJECT

The project "Sugar: make it work" has been focusing on improving the sugar sector in Rwanda (N.L. Ministry of Foreign Affairs, 2012). The goal is to make Rwanda self-proficient in covering domestic sugar need through a combination of water management and precision agriculture, which will help advance the sugarcane agriculture significantly. KSW has access to 3100 hectares of land situated around Rwanda's capital Kigali. Out of these 3100 hectares, only 1750 ha are being cultivated. Outgrowers cultivate 60% of that land. Outgrowers are independent farmers that convey their produced sugarcane to the sugar factory (N.L. Ministry of Foreign Affairs, 2012). In Figure 1, the locations of the sugarcane plots are shown in grey.

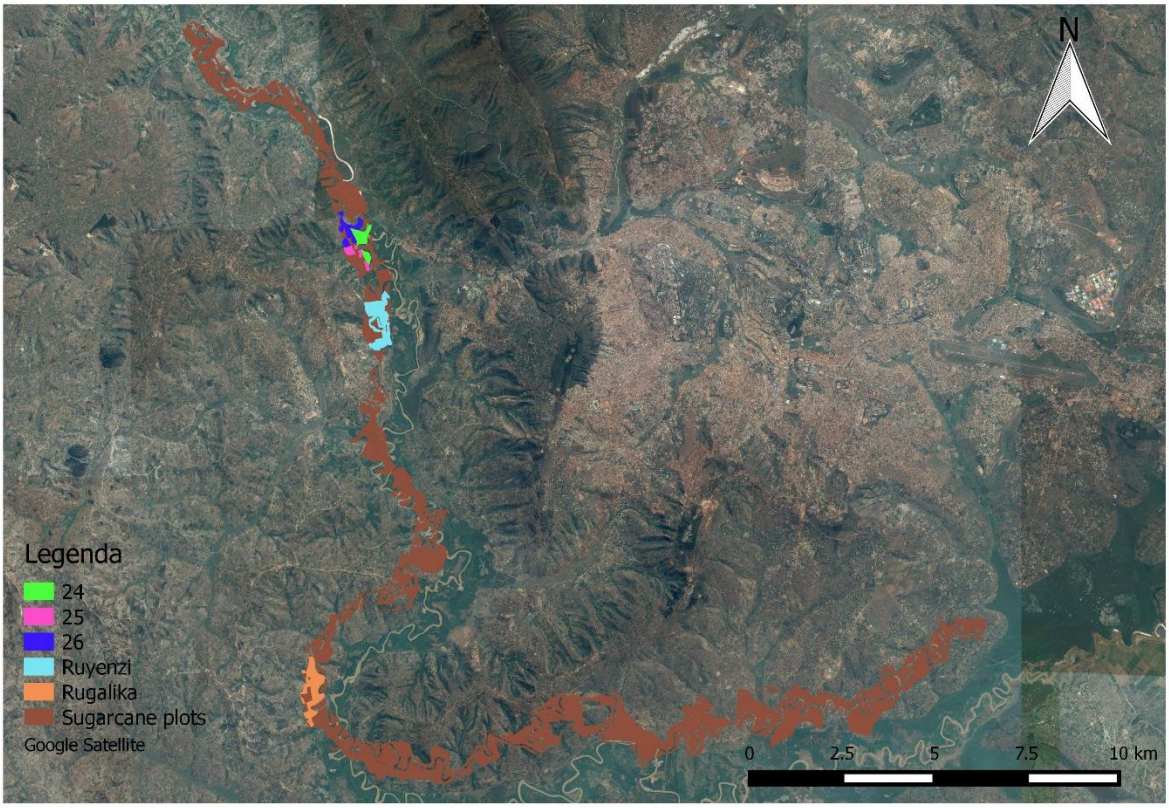

FIGURE 1. LOCATION OF SUGARCANE FIELDS OWNED BY KSW.

The remote sensing data do not cover all the areas shown in the map above, but several zones were flown with a drone. Zones consist of several individual plots where sugarcane is grown. The zones 24, 25, 26, Rugalika and Ruyenzi were flown at least one time during the data collection period. These zones belong to the estate fields.

3.1.2. DATA

In table 1, the data that was available for this research is presented.

TABLE 2. AVAILABLE DATA

Data	Details
Yield at zone level	Yields at zone level for the seasons 2014-2015 and 2015-2016 provided by KSW. For the season 2015-2016 the yields for zone Ruyenzi were missing.
Sugarcane varieties at plot level	The sugarcane variety that each plot had for a specific growing season.
Planted or ratoon at plot level	Information of the crop at a plot is a newly planted crop or a crop that is a ratoon.
Sugarcane planting data at plot level	The date, month and year at which the crop was planted or at which the ratoon started
Indirect measurements of crop state variables through remote sensing	Aerial photographs of plots within zones 24, 25, 26, Rugalika and Ruyenzi. These aerial photographs were acquired at June 2016, august 2016 and September 2016. The company Milan Innovincy provided the Aerial photographs. Information about which reflectance bands were measured is given in table 2.
Meteorological data on solar radiation	Solar radiation data from starting date 1 October 2014 till 31 December 2016. The data was provided by awhere and NASA power. Historical meteorological data on solar energy fluxes were provided by aWhere (https://developer.awhere.com/api/about-our-data/weather-data) and NASA power (https://power.larc.nasa.gov/).
Flood data	The dates at which a flood started and ended. In between these dates, some areas in the study area were fully flooded are partly flooded.

In June, August and September a drone was flown above individual plots in the zones 24, 25, 26, Rugalika and Ruyenzi. These zones are located in the area shown in figure 1. In the months the drone was flown, a couple of days were spent flying a drone above the plots to measure the spectral reflectance of the sugarcane crops on a plot. These days in which the remote sensing data was acquired are called a measurement campaign. In table 2, the measured reflectance bands are shown per measurement campaign.

TABLE 3. MEASURED REFLECTANCE BANDS PER MEASUREMENT CAMPAIGN

Measurement campaign	Measured reflectance bands
June	442 nm, 560 nm, 670 nm, 700 nm, 710 nm, 740 nm, 750 nm, 780 nm and 830 nm
August	560 nm, 670 nm, 700 nm, 710 nm, 740 nm, 750 nm, 780 nm and 830 nm
September	560 nm, 670 nm, 700 nm, 710 nm, 740 nm, 750 nm, 780 nm and 830 nm

3.2. PREDICTING YIELD

3.2.1. MODEL CHOICE

Various methods that implement remote sensing techniques exist that give an estimation of yield. From these methods, the most suitable had to be chosen for this research. This was done considering the data availability and the requirements of the model.

To choose the model a two main limiting factors are taken into account:

1. The number of time steps for remote sensing data was very limited. Imagery was only available for 3 moments in time covering a period of 18 months (one growth cycle). What's more is that these 3 moments in time were 3 consecutive months. This meant the model had to deal with a lack of time-consistent input of data and still had to be able to perform. Between the 3 months, simple linear interpolation has been carried out. The measured VI values are representing the whole growth cycle.
2. The historic data of the yield per zone was only available for the seasons 2014-2015 and 2015-2016.

The extent to which an empirical model is representative for the system to represent is depended on the recorded data. The empirical model used by Muchow et al. (1998) requires extensive records of historic data. They used block data acquired over six years. In this research, the recorded data has a limited time span. Because of this limited time span of the historic yield and crop data for the sugarcane production in Rwanda, the empirical model is deemed unsuitable for the prediction of sugarcane yield.

No day to day remote sensing imagery was available for the calibration of a mechanistic model. Furthermore, no in-field measurements were made of the sugarcane crop during the different phenological stages. Due to the unavailability of the a large time-series of remote sensing imagery and in-field measurements, parameter calibration could not be carried. Parameter calibration is needed for a mechanistic model to properly make a prediction for sugarcane yield in Rwanda. Therefore, it is not possible to utilize a mechanistic model in this research

The remaining option is to use a semi-empirical model. The semi-empirical model was chosen because of the minimum requirements of data input. The lack of available historic data can be mitigated with addition of the remote sensing imagery.

3.2.2. THE MODEL

With the input of a VI and the PAR, the model $GPP \propto VI \times PAR_{in}$ has been adopted by this project. The median VI over a plot was extracted out of the remote sensing imagery. The median was calculated per plot because the zones did not have remote sensing imagery available for all plots in that zone. This way, the mean production per hectare could be calculated per zone with the available plots. The median was chosen because of its robustness to outliers. The plot name and date of planting was added to the datasets of the plots for later visualisation. A temporal interpolation was made between the VI values to get the daily values of the median VI. These were then multiplied with the PAR to get the daily GPP values per plot. A summation was made of the daily GPP to get the total GPP per plot over the measured period (in this case 3 months). Next, a summation was made of the GPP in the plots and then divided by the number of plots to get the mean GPP in a zone.

It is assumed the fields in a zone are all of a similar size, which in reality is not the case. This decision was made because the plots do not differ in size much. Weighing the plots according to size complicates the model, while it has minimal effect on the outcome of the model. This is because the plot sizes balance each other out. Using the historic yield data as an input, a linear model was fitted with linear regression.

A distinction is made between sugarcane crops of certain age categories in a zone. Crops that have approximately the same age and are in the same phenological stage are put together into the same age category. This is done, because sugarcane growth is not constant during the different growth stages (chapter 2.3.2.). The first age category only has three months starting at 1 October 2014 and ending at 31 December 2014. The second to the sixth age categories all have four months starting at 1 January 2015.

The flowchart in figure 2 shows the steps that were taken to come to the prediction of the production per age category per zone. All of these steps were carried out in the integrated development environment Rstudio for the programming language R.

Yield prediction per age category per zone

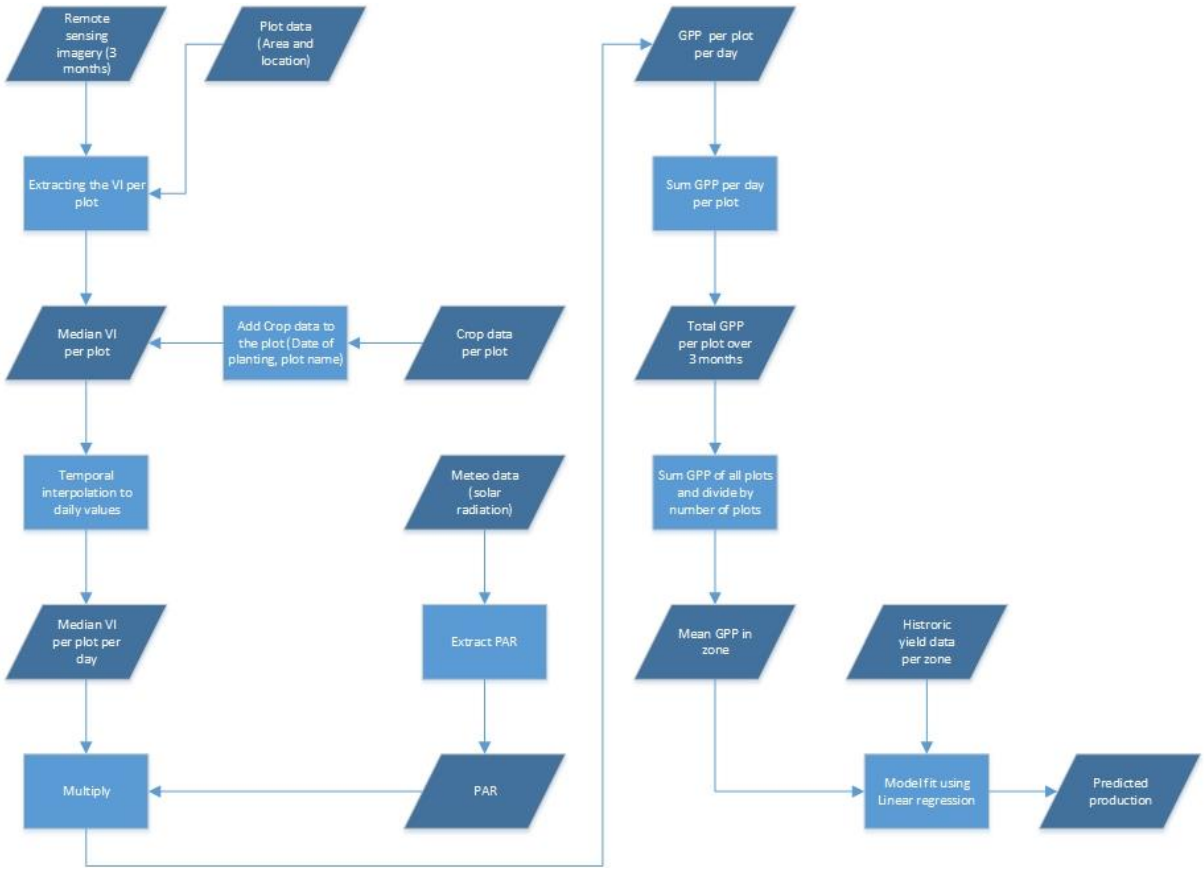

FIGURE 2. FLOWCHART OF THE INDIVIDUAL STEPS TO PREDICT PRODUCTION PER AGE CATEGORY PER ZONE. DATA IS SHOWN IN DARK BLUE, WHILE PROCESSES ARE SHOWN IN LIGHT BLUE.

Hence, the model to make a prediction of the production per age category per zone is as follows:

$$Production_{zone} = a + b \sum_{j=1}^{m} \sum_{i=1}^{n} PAR_{ij} \cdot VI_{ij} + \varepsilon$$

$Production_{zone}$ represents the production in a zone in tonnes per hectare
a represents the intercept
b is the slope coefficient
m represents the number of plots in a zone
n represents the number of days between first and last measurement
PAR_{ij} represents the incident photosynthetically active radiation for field j, on day i
VI_{ij} represents the vegetation index for field j, on day i
ε represents the stochastic remainder, which is deemed normally distributed with mean zero and standard deviation σ

3.2.3. VEGETATION INDEX

This research used the green chlorophyll index (CI_{green}) as a vegetation index. This VI has been chosen because CI green has a linear relation with total chlorophyll content and remains sensitive within a wide range of this chlorophyll content as can been seen in Figure 3 (Peng et al., 2011). The green chlorophyll index is calculated using the green band (560 nm) and the NIR band (830 nm) with the equation: $CI_{green} = \frac{\rho NIR}{\rho green} - 1$.

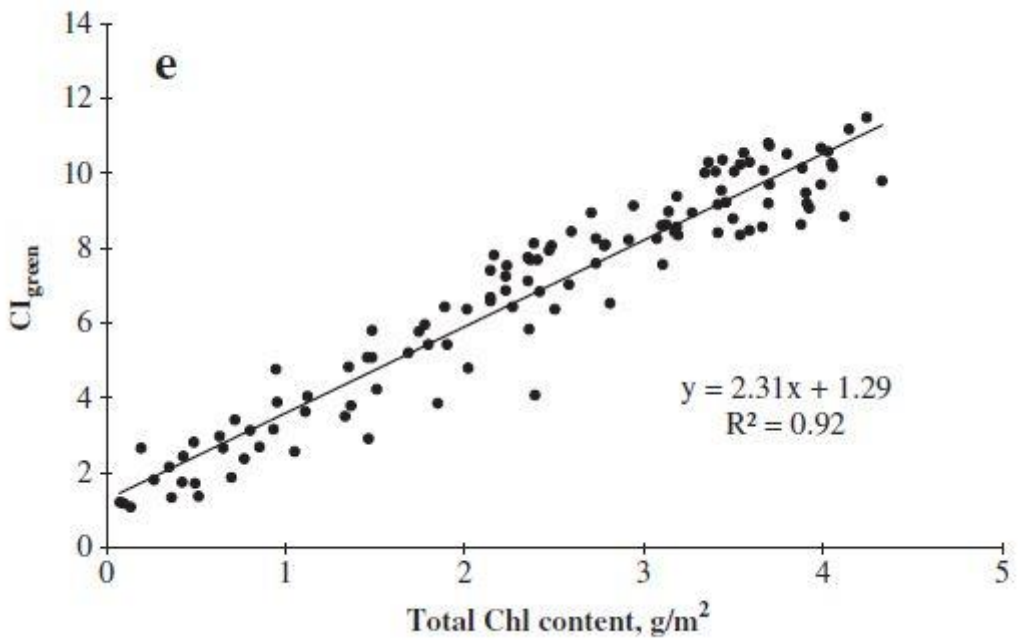

FIGURE 3. RELATION CI GREEN WITH TOTAL CHLOROPHYLL CONTENT (PENG ET AL., 2011).

3.2.4. METEOROLOGICAL DATA

Photosynthetically active radiation (PAR) is the radiation between 0.4 and 0.7 µm that can be used by plants for photosynthesis. PAR is a fraction of all the incoming sunlight *K↓*. The PAR to *K↓* ration can depend on factors such as visibility, optical depth and ozone amount (Frouin and Pinker, 1995). However, a value between 45% and 50% is commonly accepted to represent the amount of PAR the earth receives in 24 hours (Moran et al., 1995). For this research, a value of 48% was chosen similar to the research as Moran et al. (1995). This value has no influence on the predictive power of the linear model. The following function was used:

$$PAR = 0.48 \cdot K_{\downarrow 24}$$

PAR represents the photosynthetically active radiation in Watt per square meter (Wm^{-2})
$K_{\downarrow 24}$ represents the amount of incoming sunlight in 24 hours in Watt per square meter (Wm^{-2})

First, to derive the PAR, we need data on incoming solar radiation. Second, the formula above was applied to the solar radiation to get the PAR on a specific day. Data used on incoming solar radiation was provided by aWhere and NASA power. To determine if the amount of solar radiation had influence on the outcome of the model both sources were used as an input.

The data aWhere provides is collected all over the world from multiple meteorological data sources. Sources include satellite, Doppler radar, ground stations, private weather stations and global forecast models. The meteorological data was delivered without gaps in the time period. For more information about aWhere, the reader is referred to http://www.awhere.com/

The NASA power project is part of the applied sciences program set up by NASA to provide society the benefits of earth science, information and technology. A satellite measures solar energy fluxes which are used to study the climate and its processes. For more information about the NASA power project, the reader is referred to https://power.larc.nasa.gov/

3.2.5. CALIBRATION OF THE MODEL

The historic yields per hectare at zone level are used to fit the model for different age categories. Fitting was done by performing linear regression using the ordinary least squares method. This was done using the function "lm"in R. The function takes a data frame as an input.

With the gathered historic data about yield in tonnes per hectare at zone level and the date of planting for sugarcane at field level, the yield could be calculated for sugarcane that belonged one of the 6 age categories. There was no distinction made between ratooning crops and planted crops. Due to the small number of remote sensing observations of the plots, planted and ratooning crops had to be merged together to make the fitting of the model possible. For the same reason, a distinction between crop varieties was also neglected.

The model was calibrated by first making the data frame in which the mean GPP for the plots in a zone and the historic yields were stored. The historic yields that were used were of the 2014 – 2015 season as a consequence of lacking historic yields in the season 2015 – 2016. The structure of the data frame for age category x resemble the structure specified below:

Age category x:

Zone	Historic $yield_{2014\text{-}2015}$	$\frac{GPP\ (VI\ \cdot\ PAR)}{m_{plots}}$

3.3. DATA QUALITY INDICATORS

3.3.1. MEASURED CI VALUES COMPARED TO SIMULATED PROFILE

Viña et al. (2011) found a strong linear relation between green LAI and canopy chlorophyll content. They found the linear relation of $CI_{gr} = 1.677 \cdot LAI + 0.994$ between LAI and the green chlorophyll index. This relation was used to transform the LAI to the green chlorophyll index. This temporal profile of the green chlorophyll index of sugarcane was then calibrated to actual observations and used to be compared with the measured plots.

In Figure 4, a LAI profile for the sugarcane in Rwanda computed with the WOFOST crop growth model (de Bruin, 2014) is shown. The LAI curve represents the potential growth, meaning the growth was not limited by draught, flooding, nutrient deficiency and diseases (de Bruin, 2014). The growth profile was not calibrated to the sugarcane varieties used in Rwanda; however, it was calibrated to achieve a growing period of approximately 18 months. This is when the maturity of the sugarcane crop is reached. The maximum LAI was reached between the 5th and 6th month after planting.

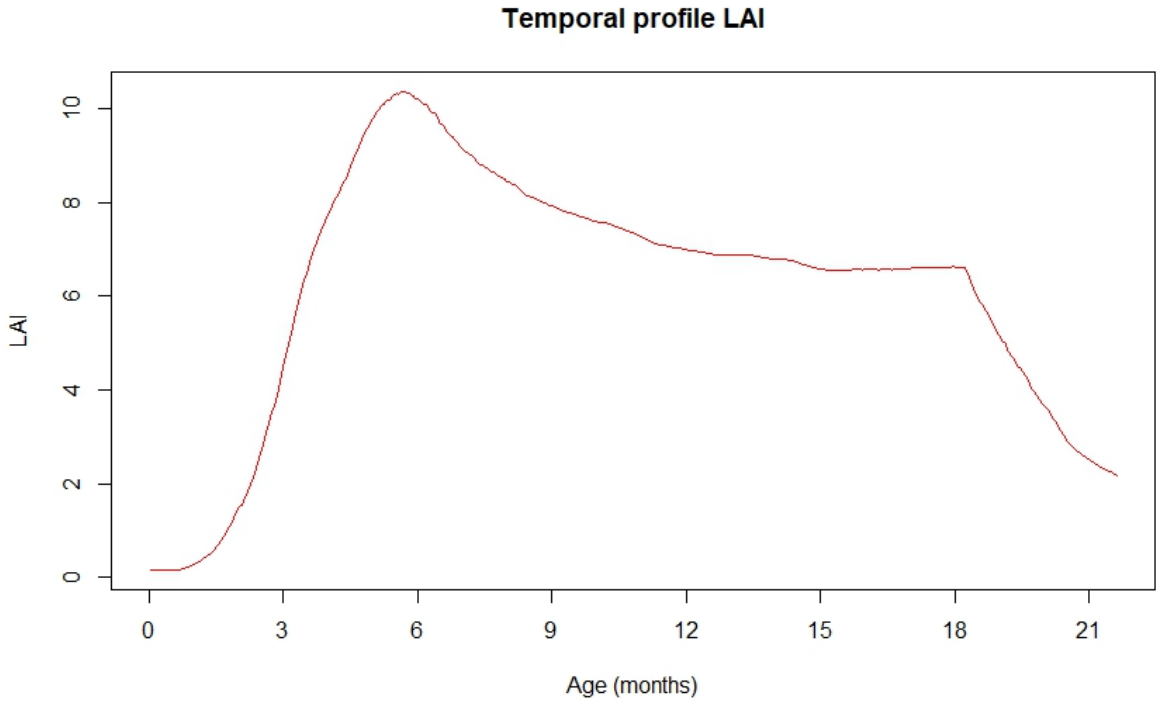

FIGURE 4. TEMPORAL LAI PROFILE OF SUGARCANE SIMULATED BY THE WOFOST MODEL. THE PROFILE REPRESENTS POTENTIAL GROWTH.

Using the relation found by Viña et al. (2011), the potential green chlorophyll index curve for sugarcane in Rwanda was constructed, which can be seen in Figure 5. The potential green chlorophyll index is the green chlorophyll index that can be reached under optimal growing conditions. The CI green was calibrated to observations of sugarcane plots in Rwanda acquired with a hyperspectral mapping system (HYMSY) that had been done by de Bruin (2015). The CI green was calibrated to 75% of the optimum LAI profile. This simulated 75% profile was used as a means to compare measured CI green values of the plots in the remaining part of this research and was labelled as the simulated growth profile. The

transformation from the LAI from Figure 4 to the temporal green chlorophyll index produced the following graph presented in Figure 5

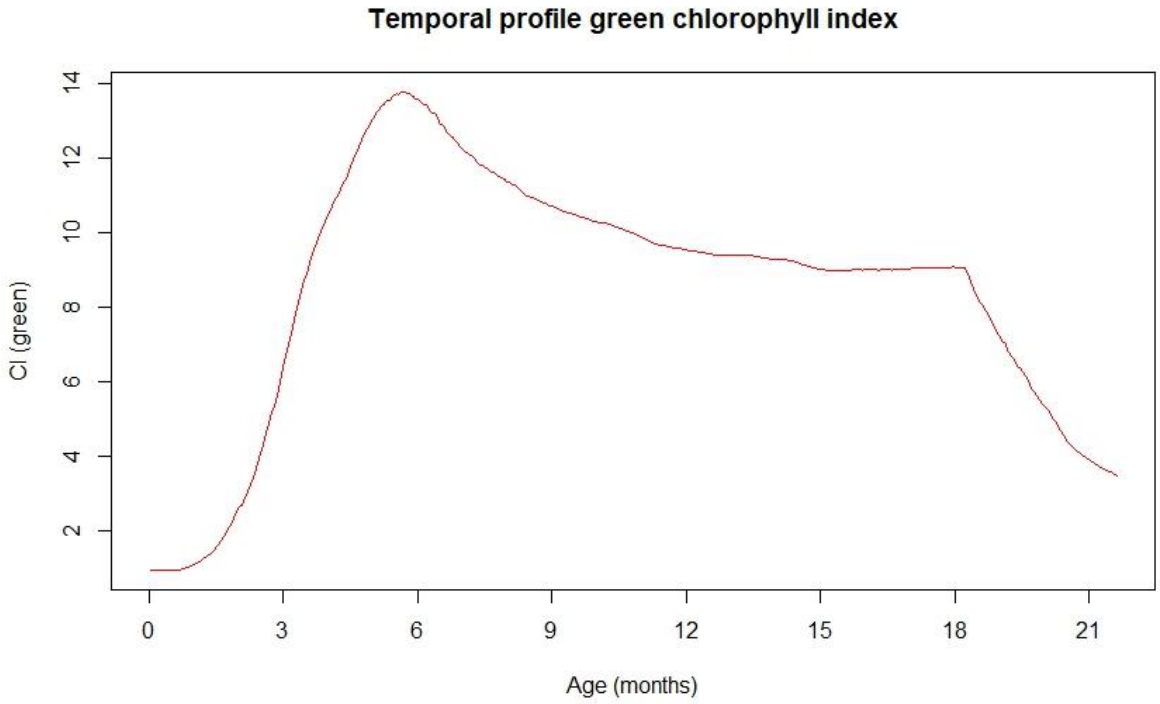

FIGURE 5. TEMPORAL GREEN CHLOROPHYLL INDEX PROFILE DERIVED FROM THE SIMULATED LAI PROFILE. THE PROFILE REPRESENTS THE POTENTIAL GROWTH.

The simulated growth profile of sugarcane was then compared with the measured CI green values of the plots of which the data was available for three consecutive months. A visual assessment was made between the measured growth profiles and the simulated growth profiles. The shape was evaluated of the individual growth curves and compared with the simulated growth curve. For example, a young crop can have exponential growth in the first five months. After that the VI is expected to decline gradually until 18 months. After that, fast deterioration was expected.

The measured CI green values were also compared with flood data shown in Figure 6, to see if the floods affected the CI green values. Data about the floods were recorded for the season 2015-2016. The figure shows when the river overflowed and affected plots in the vicinity of the river. Floods occurred in May of 2015 after which no floods occur for a couple of months. Regular floods occurred at the end of 2015 and the beginning of 2016 until June.

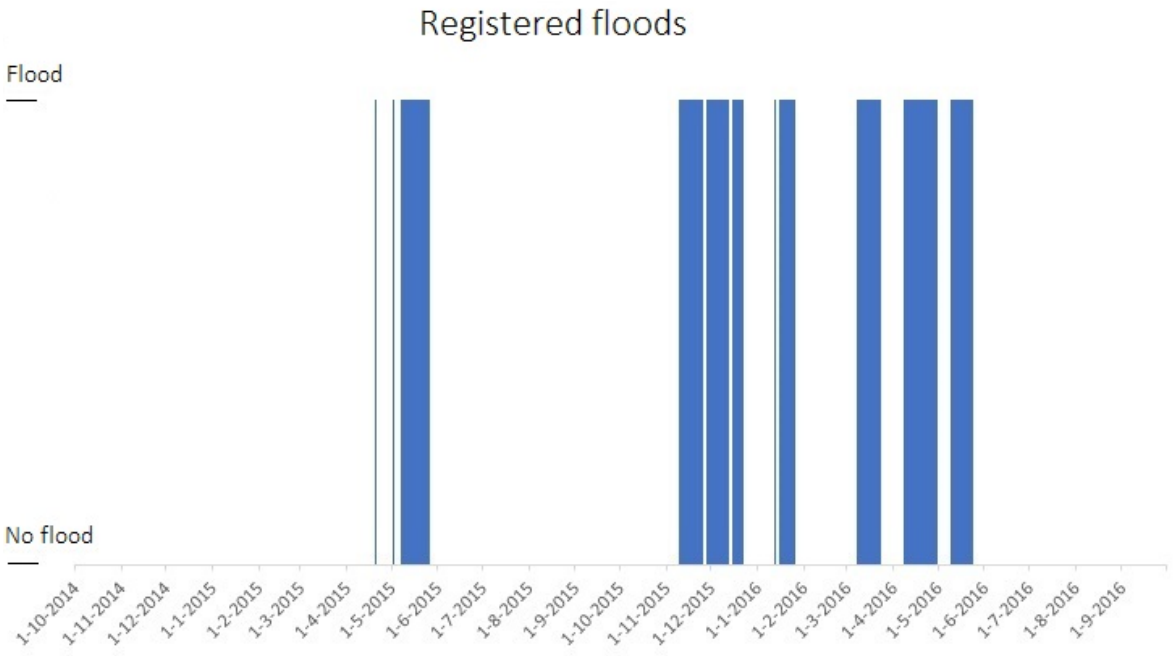

FIGURE 6. REGISTERED FLOODS IN THE PERIOD BETWEEN 1 OCTOBER 2014 AND 31 DECEMBER 2016.

3.3.2. SPACE-TIME SUBSTITUTION

In a space-time substitution, spatial phenomena are used to comprehend and model temporal processes (Blois et al., 2013) . In a space-time substitution it is assumed that temporal and spatial variation are equal. In this research, the lacking VI measurements over time for a plot, are compensated with the VI values of another plot in with sugarcane in another growth stage.

This space-time substitution was done with all the plots were remote sensing imagery had been available for (also the plots that were not measured multiple times). The space-time substitution incorporates more plots and in that way differs from the combined plots (fig 6). The space-time substitution represents all individual plots measured. Not only the ones that had multiple measuring moments in time. The space-time substitution used more samples and gives a better overview of how the plots performed compared to the ideal growth situation.

A growth curve based on the space-time substitution was made by calculating the conditional mean using the LOESS method. This technique fits a smooth curve through points in a scatter plot using local weighted regression. To make a fit at point x, points are taken from the neighbourhood of x and weighted by their distance from x. The size of the neighbourhood was set at 0.75 (the default setting). This means that a proportion of 0.75 of the points are used. Weighting is done using the tricubic weighting: $1 - ((\frac{dist}{maxdist})^3)^3$.

Through visual assessment, the simulated growth profile (Fig. 5) was compared with the CIgr profiles of the space-time substitutions. The overall shape of the conditional means was evaluated, together with the extent of the VI values.

3.3.3. Model accuracy

The fitting function "lm" in R gives a prediction of the yield in tonnes per hectare. At the same time, the fitting function gives the R^2. The R^2 expresses how much variation in the yield can be predicted by the independent variables (GPP and solar radiation). The R^2 has been used as in indicator for the model accuracy in this research.

A prediction with the historic yield data and the mean GPP per zone per age category as an input, was made using the predict function in R. The prediction interval was set 95% and 75%. The output of the predict function is a mean prediction, a lower limit and an upper limit in which the yield can be predicted with 95% certainty or 75% certainty.

3.3.4. Remote sensing imagery

The outcome of the model is dependent on the quality of the input imagery. The imagery was delivered by the company Milan Innovincy. The remotely sensed imagery was visually inspected. Irregularities were looked for within plots and between plots. Irregularities could be large differences between plots, sharp lines between two plots and spots with outliers. Furthermore, the median green chlorophyll index values at plot level were compared with the index values of a green chlorophyll index growth curve at the specific age of that plot. This way, the state of the sugarcane crop can be examined.

Remote sensing imagery acquired on the campaign of August 2016 was checked on the calibration of the remote sensing imagery. This already had been done in research by the de Bruin (2016).

4. RESULTS

In this chapter, first the VI values of fields that have been measured three times are given and compared with the simulated VI values of sugarcane during the growth cycle. Additionally, a space-time substitution is shown of the VI values of all measured plots. These are also compared with the optimum growth cycle. Then a prediction is made of the sugarcane yield for several age categories and zones. Theses prediction are shown along with the accuracy of the prediction. Lastly, the remote sensing and meteorological data are assessed.

4.1. GROWTH PROFILES

The combination of all the VI profiles of the individual plots are shown in Figure 7. It shows temporal profiles of the median CI_{green} at plot level for plots having three measurement moments in time. The green line represents the simulated CI_{green} profile.

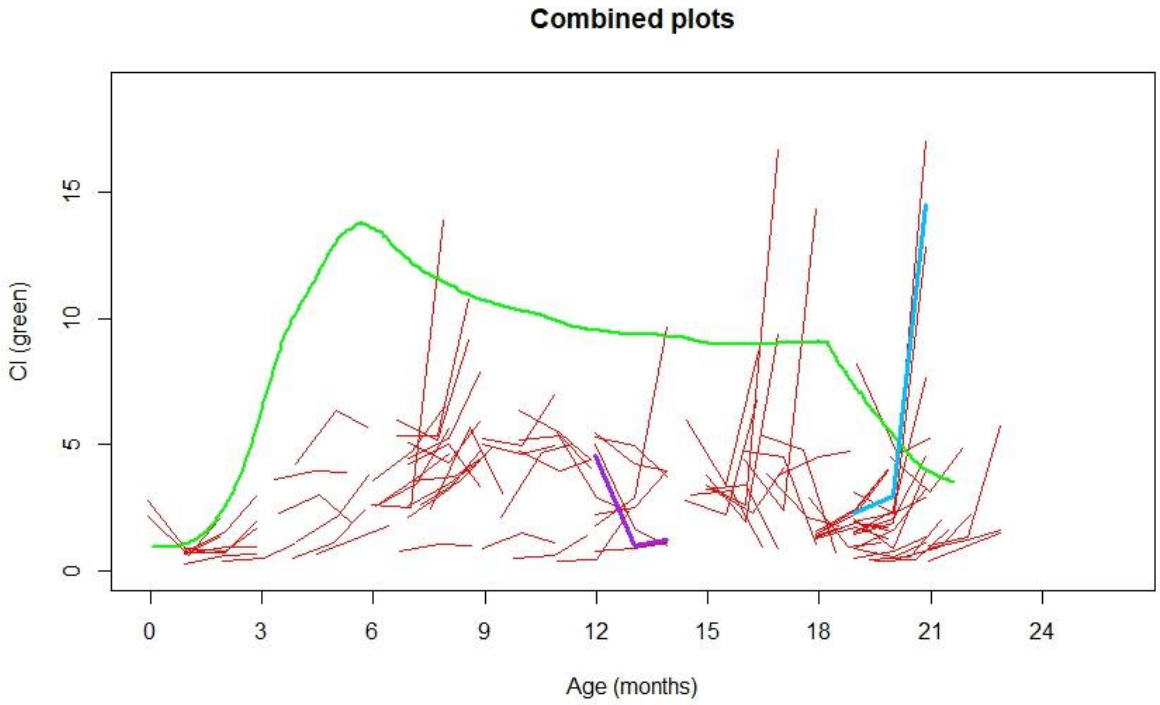

FIGURE 7. TEMPORAL GREEN CHLOROPHYLL INDEX PROFILES OF ALL THE INCLUDED SINGLE PLOTS (INDICATED WITH THE RED COLOUR) MEDIANS AT PLOT LEVEL. ALSO SHOWN IS THE SIMULATED GROWTH PROFILE (INDICATED WITH THE GREEN COLOUR).

In Figure 7, the combined plots departed greatly from the simulated CI green growth curve. In the first two campaigns measured CIgr medians at plot level were typically far below the simulated curve while the third campaign produced several unrealistically high values. Furthermore, from the age of 6 months, a 39 out of the 83 plots showed an increased green chlorophyll index compared with a decrease in the simulated growth profile.

In Figure 7, plot 6 in zone Ruyenzi-2 is indicated in the colour cyan. In this plot, the green chlorophyll index rose from 3 to 15. Transforming the CI green back to the LAI using the relation Viña et al. (2011) found, this would imply an increase in the LAI from 0.8 to 8.0 in a period of two months at the end of the growing cycle. In contrast, the simulated growth profile declined from a green chlorophyll index of 8 to 3 (LAI from 3.8 to 0.8). In figure 7, plot 6 is not the only plot that showed an increase where the simulated growth profile showed a decline. Plots in other zones show similar patterns, meaning it is not bound to only one zone.

The purple line in Figure 6 shows plot 23 in zone 24, where the green chlorophyll index declined after which it increased again. Comparing it with the simulated growth profile, the decline of the single plot was very steep. The increase from month 13 to month 14 does not coincide with the simulated growth profile.

4.2. SPACE-TIME SUBSTITUTION

In the following figures the green chlorophyll index is shown over time for all the plots that were measured. The plots are categorized by their zones.

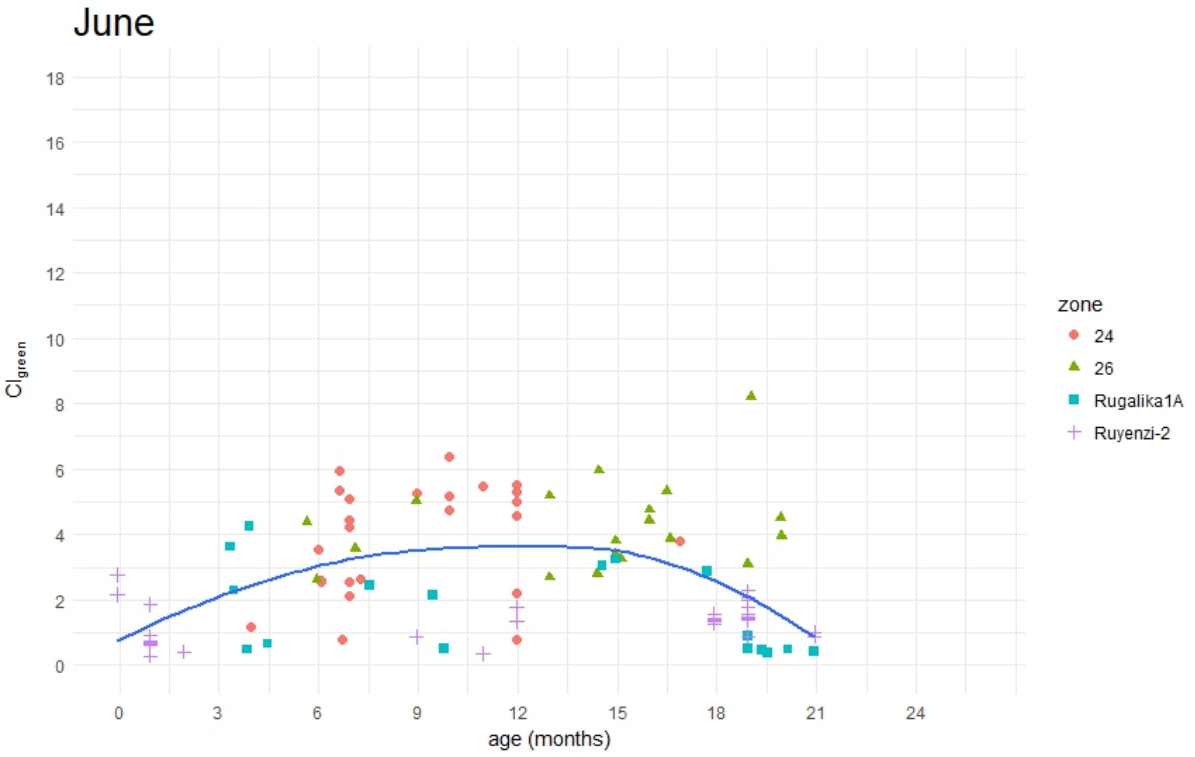

FIGURE 8. CI GREEN VALUES OF ALL THE PLOTS MEASURED IN JUNE. THE BLUE LINE RESPRESENTS THE CONDITIONAL MEAN.

Figure 8 shows the measurements made of the plots in June. He blue line represents the growth curve. The line starts at a green chlorophyll index of 1, which is the same as the simulated growth curve. However, after the first month there is no rapid incline of the growth curve. Also, the green chlorophyll index values only go as high as four, which is low when compared with the simulated growth curve. The oldest age of an individual plot is around 21 months.

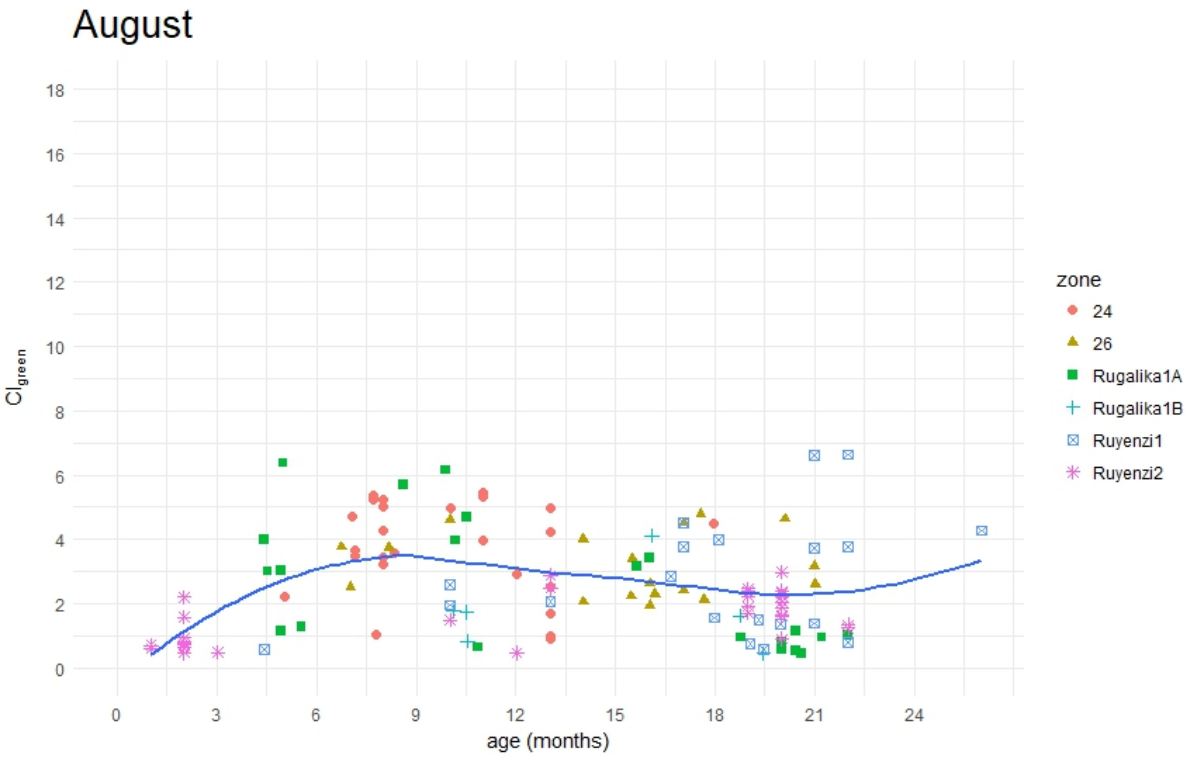

FIGURE 9. CI GREEN VALUES OF ALL THE PLOTS MEASURED IN AUGUST. THE BLUE LINE REPRESENTS THE CONIDTIONAL MEAN

Figure 9 shows the measurements of the plots made in August. The plots still show a low green chlorophyll index. The growth curve shows an upwards movement from week 90 caused by an outlying field in zone Ruyenzi-1 around month 25.

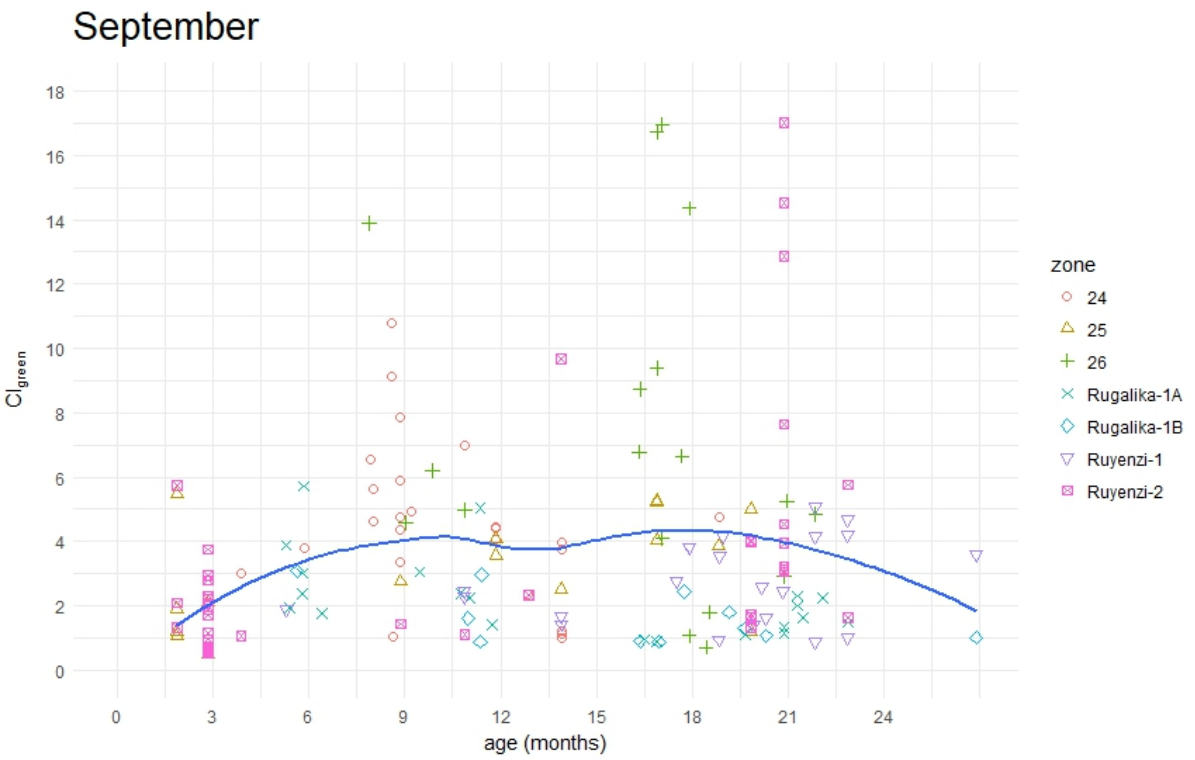

FIGURE 10. CI GREEN VALUES OF ALL THE PLOTS MEASURED IN SEPTEMBER. THE BLUE LINE REPRESENTS THE CONDITIONAL MEAN.

Figure 10 shows the measured plots in September. Around week 40, week 72 and week 90 a lot of high green chlorophyll index values are visible. When comparing the plots of zone 24, zone 26 and zone Ruyenzi-2 with the same plots measured in August, a sharp increase in the CI green values are visible.

The conditional mean of the green chlorophyll index in September still does not reach much higher than four. Even with the high values of zone 24, 26 and Ruyenzi-2.

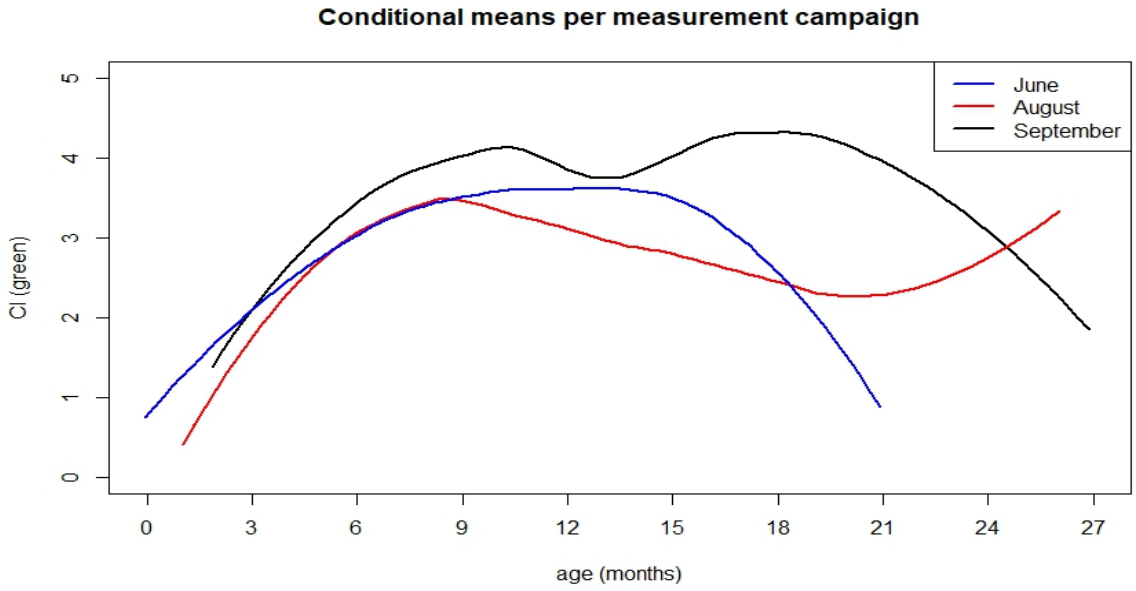

FIGURE 11. CONDITIONAL MEAN OF THE CI GREEN OF EACH MEASUREMENT CAMPAIGN.

In Figure 11 the conditional means of all measurement campaigns are plotted against each other. What is noticeable is that all the profiles have a different shape. The June profile grows till its peak is reached at 14 months, after which it declines. The august profile grows till the age of nine months after which it declines, till it starts growing again at the age of 21 months. The September profile grows till ten months. After that is a short dip at 13 months. Then at the age of 18 months it reached its peak again and declines till the age of 27 months.

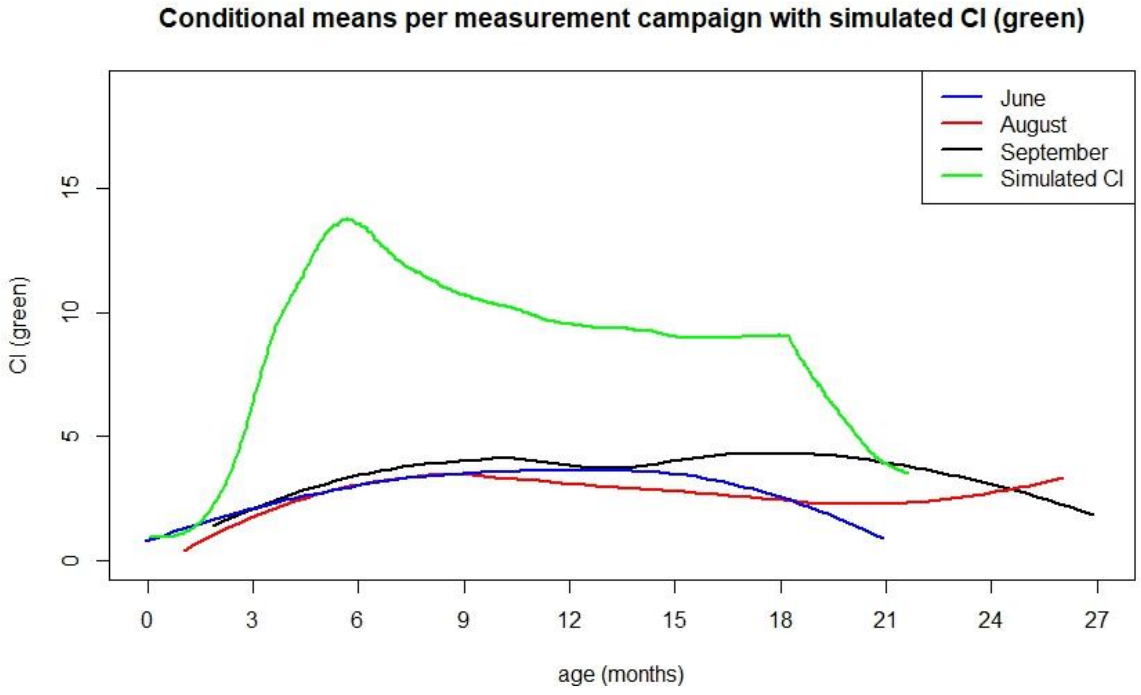

FIGURE 12.CONDITIONAL MEANSS OF THE CI GREEN OF EACH MEASUREMENT CAMPAIGN TOGETHER WITH THE SIMULATED GROWTH PROFILE OF SUGARCANE.

When the growth profiles are compared to the simulated green chlorophyll index in Figure 12, it can be seen that they do not resemble each other. The simulated growth profile reaches its peak at six months with a value of 14, while the highest value of one of the growth profiles reaches 4.5.

4.3. MODEL ACCURACY

In Figure 13, the incoming solar radiation in megajoule per square meter per day is displayed for both aWhere and NASA power. The black line represents the solar radiation measured by aWhere. The red line represents the solar radiation measured by NASA power.

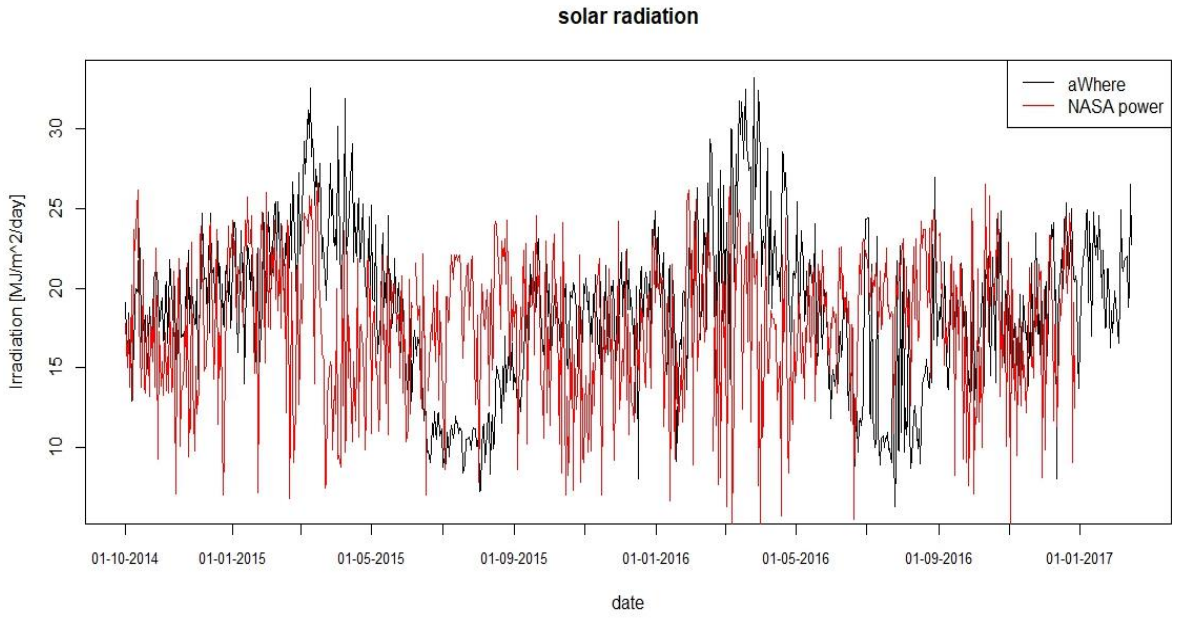

FIGURE 13. SOLAR RADIATION MEASURED BY AWHERE AND NASA POWER FOR THE COORDINATES LONGITUDE 30.1355 AND LATITUDE -1.9641 WHICH REPRESENTS THE AREA IN WHICH THE STUDY AREA LIES. THE DATA WAS COLLECTED BETWEEN 1 OCTOBER 2014 AND 1 JANUARY 2017.

The black line of the data collected by aWhere differs from the data collected by NASA power in a way that there is a more seasonal pattern in the data collected by aWhere. The aWhere data shows a peak in measured solar radiation around march. The solar radiation is lowest around July. The data collected by NASA power does not show this seasonal pattern.

The tables below represent the predicted yield for the mean GPP ($\sum VI \cdot PAR$) per age category per zone. The results were obtained using the meteorological data acquired of both aWhere and NASA power.

Fit represents the mean predicted yield. The lower (LWR) and upper (UPR) represent the lower and upper limit in which the yield can be predicted with a pre-determined certainty. In this case a prediction is made of the yield with 95% certainty and 75% certainty. For age category 1 (date of planting between 1 October 2014 and 31 December 2014); R^2 amounted to 0.5.

TABLE 4. PREDICTIONS AGE CATEGORY 1. FIT REPRESENTS THE MEAN PREDICTED VALUE OF THE SUGARCANE YIELD IN TONNES PER HECTARE. LWR AND UPR REPRESENT RESPECTIVELY THE LOWER AND UPPER BOUNDERY IN WHICH THE YIELD CAN BE PREDICTED WITH A PRE-DETERMINED CERTAINTY (IN THIS CASE 95% AND 75%).

ZONE	METEO	FIT	LWR 95 %	UPR 95%	LWR 75%	UPR 75%
26	aWhere	62.03	-575.50	699.56	-59.10	183.16
RUGALIKA	aWhere	115.11	-575.04	805.26	-16.02	246.24
RUYENZI	aWhere	76.35	-50.98	655.68	-33.72	186.42
26	NASA	62.89	-675.69	801.47	-77.44	203.23
RUGALIKA	NASA	109.38	-657.84	876.61	-36.39	255.16
RUYENZI	NASA	81.22	-564.20	726.63	-41.42	203.85

R^2 age category 1 aWhere: 0.50. *R^2 age category 1 NASA power: 0.36.*

TABLE 5. PREDICTIONS AGE CATEGORY 2.

ZONE	METEO	FIT	LWR 95%	UPR 95 %	LWR 75%	UPR 75%
24	aWhere	97.57	-112.77	307.91	19.18	175.96
26	aWhere	93.95	-100.80	288.70	21.37	166.53
RUGALIKA	aWhere	81.98	-114.79	278.75	8.65	155.32
RUYENZI	aWhere	79.50	-127.97	286.98	2.18	156.83
24	NASA	87.14	-120.87	295.14	9.62	164.66
26	NASA	87.02	-124.32	298.37	8.26	165.79
RUGALIKA	NASA	89.71	-129.05	308.47	8.18	171.24
RUYENZI	NASA	89.14	-113.08	291.35	13.77	164.50

R^2 age category 2 aWhere: 0.07. *R^2 age category 2 NASA power: 0.002.*

TABLE 6. PREDICTIONS AGE CATEGORY 4.

ZONE	METEO	FIT	LWR 95%	UPR 95 %	LWR 75%	UPR 75%
24	aWhere	95.80	-88.08	279.68	27.27	164.33
26	aWhere	101.62	-97.84	301.08	27.28	175.96
RUGALIKA	aWhere	82.87	-97.23	262.98	15.75	150.00
RUYENZI	aWhere	72.71	-134.30	279.74	-4.44	149.87
24	NASA	104.92	-29.46	239.30	54.84	155.00
26	NASA	103.21	-29.72	236.14	53.67	152.75
RUGALIKA	NASA	92.39	-34.86	219.64	44.97	139.82
RUYENZI	NASA	52.49	-106.32	211.29	-6.70	111.67

R^2 age category 4 aWhere: 0.16. *R^2 age category 4 NASA power: 0.56.*

For all the zones in the age categories, the lower boundary 95% (LWR 95%) and upper boundary 95% (UPR 95%) are far apart, meaning there is a wide range in which the yield can be predicted with 95% certainty. The lower boundary suggests the yields would be negative, with the upper boundary the yields

could go up to 805.262 tonnes per hectare. Even for the lower and upper boundary at 75% certainty, the values are far apart. Zone Rugalika in age category 4 has the smallest width, ranging from 15.750 to 149.997.

Both values are highly unrealistic, with registered yields of approximately 70 tonnes per hectare for planted crops and approximately 80 tonnes per hectare for ratooning crops. This wide range is an indicator that the model is not accurate. This could also be deduced from the low R^2.

While the R^2 values of the age categories with the NASA power data differed from the age categories with the aWhere data, the range of the yield prediction is still very wide for each age category. The R^2 of 0.5642 for age category 4 was the highest R^2 in the results. However, the lower boundary for 95% is negative for all zones and the upper boundary lies above 200. For 75% certainty the lower boundary comes close to the registered yields of 70 and 80 tonnes per hectare for the zones 24, 26 and Rugalika. The upper boundary still lies way above these registered yields as well as realistic yields in the region.

4.4. REMOTE SENSING IMAGERY

When doing a first reconnaissance of the remote sensing data a few things stand out. First of all, when visually assessing the green chlorophyll index for, in this case Rugalika-1a in June, there sharp differences between the individually measured plots. In Figure 14 it can be seen that there were large differences between different plots with a sharp line dividing the fields, with the plots on the left that are more bright than the plots on the right. This sharp line suggest differences in the primary acquired data, in this case the different wavelengths. This indicates that the measuring circumstances varied.

Furthermore, in the remote sensing data there were some far outliers. In Figure 15, the green chlorophyll index went as high as 663.

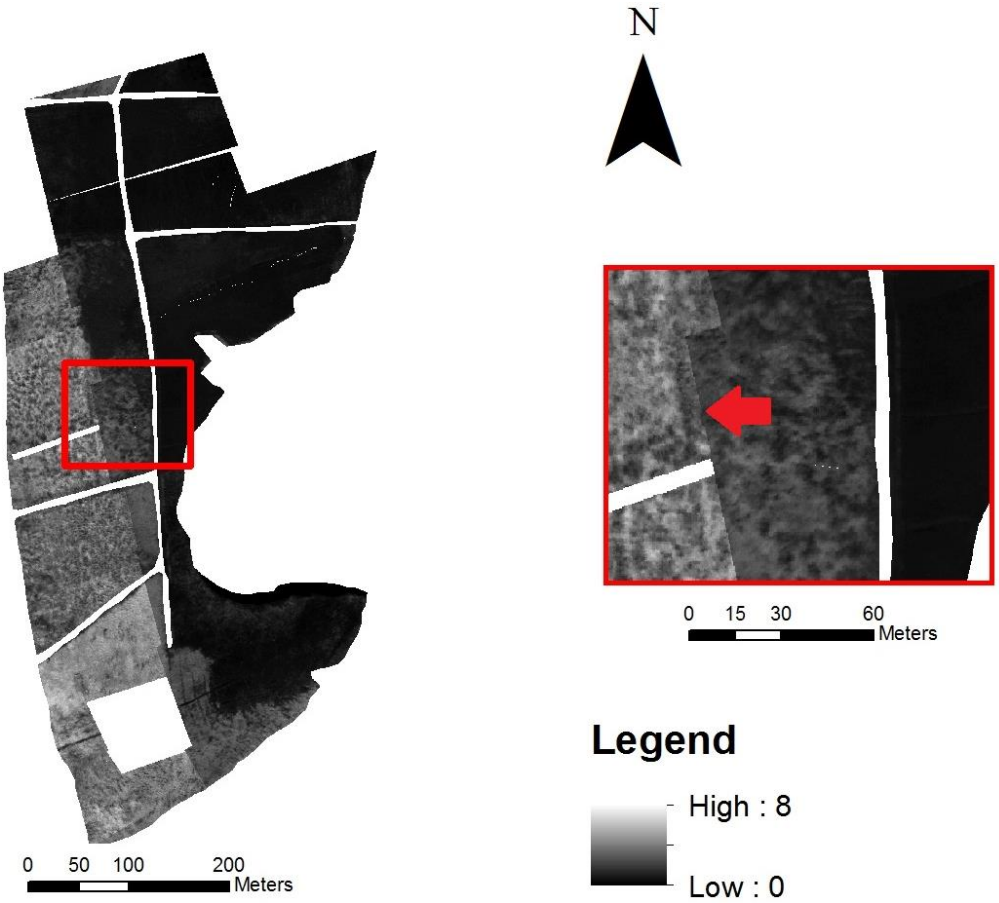

FIGURE 14. CI GREEN IN RUGALIKA. SHARP DIFFERENCES CAN BE OBSERVED BETWEEN AREAS THAT WERE MEASURED DURING DIFFERENT FLIGHTS AT THE SAME MEASUREMENT CAMPAIGN.

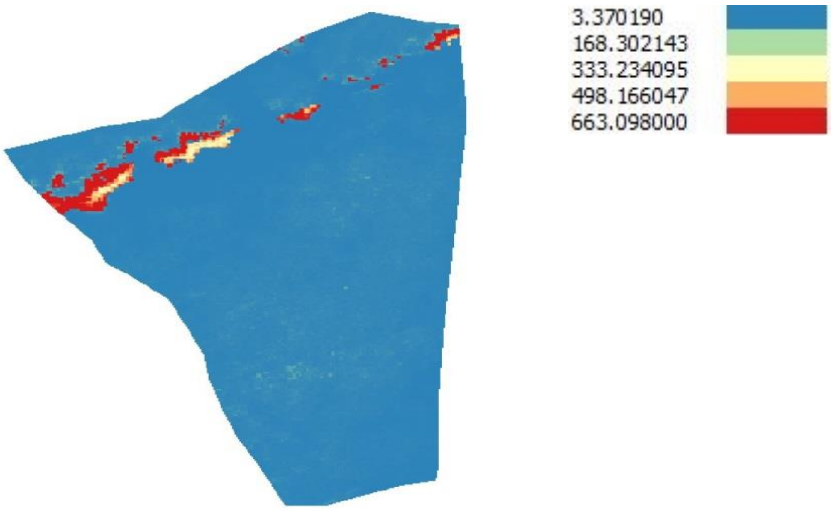

FIGURE 15. THE GREEN CHLOROPHYLL INDEX IS DISPLAYED FOR A PLOT. THE YELLOW AND RED AREAS SHOW EXTRAORDINARY VALUES FOR THE VI.

4.4.1. CALIBRATION OF REMOTELY SENSED IMAGERY

To calibrate the measuring equipment, targets with a pre-set reflectance across all wavelengths were used. Those targets have a flat response at 6% and 33% reflectance and are placed in the field. In previous research by de Bruin (2016), indicators of image quality from point pairs were contrived. The point pairs were acquired over the calibration targets.

de Bruin (2016) found that the responses fluctuated from the lines of 6% and 33% reflectance. The reflectance was not a linear line, but varied throughout the spectrum. For the 6% reflectance, the green band usually had a small overestimation (0.5% – 1% overestimation) while the NIR had a greater overestimation (2% – 6% overestimation). At the 33% reflectance, the green band had a greater underestimation (2% – 33% underestimation), while the NIR had a smaller underestimation (1% – 3% underestimation). On the whole, the reflectance of 6% tended to be overestimated, while the reflectance of 33% were underestimated. These dispositions affect VIs near absorbance and reflectance bands producing lower VI values.

5. DISCUSSION

In this chapter the adopted model will be evaluated. Its constraints and how these could have an effect on the outcome of the model are discussed. Furthermore, the results of the model are assessed together with the data quality.

5.1. EVALUATING THE MODEL

Regarding the model (section 3.2.2.), one could argue the light use efficiency (LUE) model $GPP \propto VI \times PAR_{in}$ is a suitable model to accurately predict sugarcane yield. It is arguable that the yield of a crop can be predicted using a linear model with two variables: the vegetation index of a crop and the incoming photosynthetically active radiation. Peng and Gitelson (2012) found an instantaneous relation between GPP and $VI \times PAR_{in}$ in soybean and maize. The relation for sugarcane GPP and vegetation index times photosynthetically active radiation may vary from the relation soybean and maize have. The model used in this research does not take into account this factor which leaves the question if this model represents sugarcane growth accurately.

5.2. DATA LIMITATIONS

Firstly, the calibration of the model (section 3.2.5.) depended on historic data. Historic data was available for this research, but not for every zone. Furthermore, the timescale of this data was very limited, including only two seasons. Historic yields over a longer time scale are required to properly calibrate the model, because the model currently is heavily dependent on the yields of the two seasons. The measured yields could deviate from yields due to negative or positive contemporary factors. The time scale of the remote sensing measurements should be assessed in further research.

A major constraint in the model is the limiting number of VI observations. The measurements made in the months July, August and September were deemed to represent the whole growth cycle. The VI is an indicator about crop status, but over a small time scale the crop status cannot represent the whole growth cycle. When predicting the sugarcane yield, a time-series of the whole growth cycle is a prerequisite. Any influences that happened before or after the measurements were taken were not accounted for in the yield prediction. An assessment on how many remote sensing observations over the growth period are required for the model to properly represent crop growth should be made.

The data of NASA power and aWhere do not follow the same trends. It is hard to tell which one lies closest to the reality because there were no measurements with the required time scale available that were made at the location itself. Both sources use different methods to acquire the incoming solar radiation.

Despite the large differences in solar radiation, the impact of the solar radiation is minimal on the outcome of the model. This is due to GPP per day per plot being summed to a total GPP over 3 months per plot (see Fig. 2). This makes the incoming photosynthetically active radiation less significant if the source provides the PAR in a consistent fashion.

Some other inconsistencies in the data prevented the model from giving more reliable results:

- Not all the plots in the zones were incorporated in the model for that specific zone, because no remote sensing imagery was available for all 3 measuring campaigns for these plots. Depending on the data availability for a plot, the plot was incorporated into the model. To increase the accuracy of the prediction of the yield of a zone, incorporation of all the fields would benefit the model.
- No distinction was made between planted crops and ratoon crops. As a consequence of no data being available for the plots in a zone for a certain age category. Dividing the model to represent planted and ratoon crops could increase the accuracy of the model.
- The spatial scale of the model used in this research is at zone level. Delécolle et al. (1992) argued that assessment of large agricultural regions should be performed at field level, since other factors like crop variety and soil type could vary. Calibrating at field level requires historic yield data for each field which was not available for this research. This would increase accuracy in future attempts to predict yield.

5.3. EVALUATING THE RESULTS

In the space-time substitution, the plots of the conditional mean do not resemble the plot of the simulated growth profile (Figure 10). The mean value of the combined plots was constantly lower than that of the simulated growth profile. The sharp increase in the first few months is missing. Furthermore, the plots of the individual fields do not follow the shape of the simulated growth profile at any time. The conditional means of all the plots are not higher than a green chlorophyll index of five. Also, the shape does not match the shape of the simulated growth profile.

When inspecting the individual plot 6 in the zone Ruyenzi-2 (the cyan line in Figure 7), the growth curve does not look anything like the simulated growth curve. It is expected for the sugarcane to have a declining green chlorophyll index from around the 15^{th} month (∼450 days). However, the plot in Figure 5 has a sharp increase in the green chlorophyll index. This would imply a LAI increase from 0.795 to 7.951 in a period of two months at the end of the growing cycle. Compared to the simulated growth profile, this is not credible.

The individual plot 23 in zone 24 (the purple line in Figure 7) shows a decline in CI green after which it increases again. This is unusual for the age of the crop.

Plot 23 in zone 24 was planted at 1 July 2015. The moments of measuring were at June, August and September 2016. During this period no floods were recorded (Fig. 6). The decline of the green chlorophyll index is not related to floods according to flood data, which means other factors account for unrealistic values of the green chlorophyll index.

Two reasons could contribute to the individual plots not resembling the simulated growth profile: the age of the crop is communicated wrong or something went wrong with the calibration of the measuring equipment.

In chapter 4.4.1. it is shown that the calibration of the measuring equipment hadn't been properly. The dispositions lowered the CI green values for that measuring campaign. This explanation seems plausible in the case of the remotely sensed imagery that was obtained in Rwanda. The mean and median CI green values are consistently lower than the simulated growth profile. Also, when inspecting Figure 11, difference between lines of flights can be observed which indicates flawed calibration of the measuring equipment.

6. CONCLUSIONS AND RECOMMENDATIONS

In this chapter, following the points that are made in the discussion, the research question are answered. Succeeding the conclusions, a couple of recommendations are made to improve the prediction of sugarcane yield in future research.

1. What are existing methods for predicting sugarcane harvest yield using remotely sensed data?

Implementation of remotely sensed data is possible in empirical, semi-empirical and mechanistic models. However, remotely sensed imagery is not a necessity for an empirical method since such a method is based on historic data about yield at a specific location and specific growing conditions. In mechanistic methods, it is possible for remotely sensed imagery to be an input. For the prediction of yield, remotely sensed imagery is a crucial component in the semi-empirical model.

2. What inputs and parameters are needed for implementing these methods?

Depending on the model, inputs range from historic yield and conditions to input of the mechanistic nature. Inputs for the empirical model are based on historic measurements of the crop. Crop class, area, yield, harvest age and harvest month are all variables used to predict yield in the empirical method. For the mechanistic method, all variables that explain yield can be added to make a model. Soil data, weather data, crop data, and management data are all potential inputs. For the method that was used in this research, crop data - acquired out of a remote sensing image - in the form of the green chlorophyll index was used. The incoming solar radiation was multiplied with the median green chlorophyll index in a plot to get the GPP for a day in that plot. Then, this was multiplied with the number of plots in a zone and the number of days in the growth period to predict the gross in a zone over a certain period of time.

3. Given data availability within the "Sugar: make it work" project, which is a viable model for predicting sugar cane yields?

The Light use efficiency (LUE) model – where the gross primary product of the sugarcane crop is proportional to the green chlorophyll index multiplied with the photosynthetically active radiation ($GPP \propto VI \times PAR_{in}$) – was the most suiting model. Comparing the available data with the requirements of the model, the semi-empirical approach was the most convenient and applicable method. However, given the poor accuracy of the model, it is arguable if even this method is viable for predicting the yield. Given that there are doubts about the consistency and accuracy of the data, no method will currently be able to predict the yield accurate enough for KSW to be useful.

4. Are the available data suitable for obtaining accurate yield predictions using the selected model?

Very low accuracy was achieved with the chosen model. The R-squared of the models for the different age categories did not exceed 0.5. For this research, this low value of the R^2 is not sufficient. The 95% prediction intervals for yield extended from -88 to 280 in the best case. The 75% prediction intervals extended from 27 to 164 Both prediction intervals indicate large uncertainty about yields using currently

available data in combination with the LUE approach. This range is useless when estimating how much sugarcane at one moment will be delivered to the sugarcane factory for processing. KSW aims to have the model predict the yield within a margin of 10 tonnes per hectare. With the results of the chosen model, this is not viable. In future endeavours to predict sugarcane yield, the data has to be more consistent, accurate and precise. Otherwise, no model will be able to make an accurate estimation of the yield.

For the prediction of sugarcane yield, quality of data plays an important role. For the data to be useful it has to be consistent and accurate. Flying regularly (having more time-steps) will aid the calibration of the model to get a more accurate result of the predicted yield. Assessment of how many time-steps are required should be assessed in further research. In order to avoid inconsistency in the image data, calibrating the measuring equipment using a reflectance panel each mission is a prerequisite.

Additionally, a database with general information about the historically obtained yields, floods, diseases, crop variety, date of planting and if the crop is planted or ratooned will aid the model to give more accurate yield predictions (per crop class). Storing information about the historic yield per hectare per zone over a longer time will aid in the calibration of the model. Monitoring general information and keeping track of the yields that have been obtained in the past per hectare per zone is happening already. This is a step in the right direction and has to be further developed.

To measure incoming solar radiation, equipment is needed at a location adjacent to the sugarcane plots. A measuring device at each zone would be preferable to make accurate predictions for each zone.

With the requisition of more data, the opportunity arises to implement other models. With the collection of soil data, management data and additional environmental data, mechanistic models like SUCROS and LINTUL will be able to make a yield prediction. Adding remote sensing and weather data could provide real-time data on crop status. Real-time information about crop status provides possibilities to counter negative influences on crops.

A reliable model is conceivable using accurate remote sensing data, weather data, the date of planting of the sugarcane and area of a plot. This data is already collected. A lot of progress can be made, when data is more carefully acquired.

7. REFERENCES

ABDEL-RAHMAN, E. M. & AHMED, F. B. 2008. The application of remote sensing techniques to sugarcane (Saccharumspp. hybrid) production: a review of the literature. *International Journal of Remote Sensing,* 29**,** 3753-3767.

ANSOMS, A. 2008. Privatisation's bitter fruit: the case of Kabuye Sugar Works in Rwanda. *L'Afrique des grands lacs: annuaire,* 2009**,** 55-70.

ATZBERGER, C., JARMER, T., SCHLERF, M., KÖTZ, B. & WERNER, W. Retrieval of wheat bio-physical attributes from hyperspectral data and SAILH+ PROSPECT radiative transfer model. Proceedings of the 3rd EARSeL Workshop on imaging spectroscopy, 2003. Citeseer, 473-482.

BARET, F. & GUYOT, G. 1991. Potentials and limits of vegetation indices for LAI and APAR assessment. *Remote sensing of environment,* 35**,** 161-173.

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**,** 869-880.

BASTIAANSSEN, W., MENENTI, M., FEDDES, R. & HOLTSLAG, A. 1998. A remote sensing surface energy balance algorithm for land (SEBAL). 1. Formulation. *Journal of hydrology,* 212**,** 198-212.

BASTIAANSSEN, W. G. M. & ALI, S. 2003. A new crop yield forecasting model based on satellite measurements applied across the Indus Basin, Pakistan. *Agriculture Ecosystems & Environment,* 94**,** 321-340.

BAUMGARDNER, M. F., SILVA, L. F., BIEHL, L. L. & STONER, E. R. 1986. Reflectance properties of soils. *Advances in agronomy,* 38**,** 1-44.

BLOIS, J. L., WILLIAMS, J. W., FITZPATRICK, M. C., JACKSON, S. T. & FERRIER, S. 2013. Space can substitute for time in predicting climate-change effects on biodiversity. *Proceedings of the National Academy of Sciences,* 110**,** 9374-9379.

BOOTE, K. J., JONES, J. W. & PICKERING, N. B. 1996. Potential uses and limitations of crop models. *Agronomy Journal,* 88**,** 704-716.

BOUMAN, B. 2001. *ORYZA2000: modeling lowland rice*, IRRI.

BROGE, N. H. & LEBLANC, E. 2001. Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote sensing of environment,* 76**,** 156-172.

CIGANDA, V., GITELSON, A. & SCHEPERS, J. 2008. Vertical profile and temporal variation of chlorophyll in maize canopy: Quantitative "crop vigor" indicator by means of reflectance-based techniques. *Agronomy Journal,* 100**,** 1409-1417.

CLEVERS, J. 1989. Application of a weighted infrared-red vegetation index for estimating leaf area index by correcting for soil moisture. *Remote Sensing of Environment,* 29**,** 25-37.

CLEVERS, J. G. P. W. 1997. A simplified approach for yield prediction of sugar beet based on optical remote sensing data. *Remote Sensing of Environment,* 61**,** 221-228.

CURRAN, P. J., DUNGAN, J. & GHOLZ, H. 1990. Exploring the relationship between reflectance red edge and chlorophyll content in slash pine. *Tree physiology,* 7**,** 33-48.

DE BRUIN, S. 2014. Sugar make it work: Methodology development remote sensing for sugarcane - final version. Wageningen University.

DE BRUIN, S. 2015. Sugar make it work: Contributions to result 3. Wageningen University.

DE BRUIN, S. 2016. Precision agriculture results of the "Sugar make it work" project: current status and prospects. Wageningen University.

DELÉCOLLE, R., MAAS, S. J., GUÉRIF, M. & BARET, F. 1992. Remote sensing and crop production models: present trends. *ISPRS Journal of Photogrammetry and Remote Sensing,* 47**,** 145-161.

DIEPEN, C. V., WOLF, J., KEULEN, H. V. & RAPPOLDT, C. 1989. WOFOST: a simulation model of crop production. *Soil use and management,* 5**,** 16-24.

DORIGO, W. A., ZURITA-MILLA, R., DE WIT, A. J. W., BRAZILE, J., SINGH, R. & SCHAEPMAN, M. E. 2007. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. *International Journal of Applied Earth Observation and Geoinformation,* 9**,** 165-193.

FAURTYOT, T. & BARET, F. 1997. Vegetation water and dry matter contents estimated from top-of-the-atmosphere reflectance data: a simulation study. *Remote Sensing of Environment,* 61**,** 34-45.

FIELD, C. B., RANDERSON, J. T. & MALMSTRÖM, C. M. 1995. Global net primary production: combining ecology and remote sensing. *Remote sensing of Environment,* 51**,** 74-88.

FROUIN, R. & PINKER, R. T. 1995. Estimating photosynthetically active radiation (PAR) at the earth's surface from satellite observations. *Remote Sensing of Environment,* 51**,** 98-107.

GITELSON, A. A. 2004. Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation. *Journal of plant physiology,* 161**,** 165-173.

GITELSON, A. A., VINA, A., CIGANDA, V., RUNDQUIST, D. C. & ARKEBAUER, T. J. 2005. Remote estimation of canopy chlorophyll content in crops. *Geophysical Research Letters,* 32.

GITELSON, A. A., VIÑA, A., VERMA, S. B., RUNDQUIST, D. C., ARKEBAUER, T. J., KEYDAN, G., LEAVITT, B., CIGANDA, V., BURBA, G. G. & SUYKER, A. E. 2006. Relationship between gross primary production and chlorophyll content in crops: Implications for the synoptic monitoring of vegetation productivity. *Journal of Geophysical Research: Atmospheres,* 111.

GOWARD, S. N. & WILLIAMS, D. L. 1997. Landsat and earth systems science: development of terrestrial monitoring. *Photogrammetric engineering and remote sensing,* 63**,** 887-900.

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**,** 337-352.

HENRY, P. & ELLIS, R. Soil as a factor in sugarcane ratoon yield decline on an irrigated estate in Swaziland. Proc. Int. Soc. Sugarcane Technol, 1995. 236-245.

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**,** 195-213.

JACQUEMOUD, S. & BARET, F. 1990. PROSPECT: A model of leaf optical properties spectra. *Remote sensing of environment,* 34**,** 75-91.

KUUSK, A. 1995. A fast, invertible canopy reflectance model. *Remote Sensing of Environment,* 51**,** 342-350.

LISSON, S. N., INMAN-BAMBER, N. G., ROBERTSON, M. J. & KEATING, B. A. 2005. The historical and future contribution of crop physiology and modelling research to sugarcane production systems. *Field Crops Research,* 92**,** 321-335.

MONTEITH, J. L. & MOSS, C. 1977. Climate and the efficiency of crop production in Britain [and discussion]. *Philosophical Transactions of the Royal Society of London B: Biological Sciences,* 281**,** 277-294.

MORAN, M. S., MAAS, S. J. & PINTER JR, P. J. 1995. Combining remote sensing and modeling for estimating surface evaporation and biomass production. *Remote Sensing Reviews,* 12**,** 335-353.

MOULIN, S., BONDEAU, A. & DELECOLLE, R. 1998. Combining agricultural crop models and satellite observations: from field to regional scales. *International Journal of Remote Sensing,* 19**,** 1021-1036.

MUCHOW, R. C., HIGGINS, A. J., RUDD, A. V. & FORD, A. W. 1998. Optimising harvest date in sugar production: a case study for the Mossman mill region in Australia: II. Sensitivity to crop age and crop class distribution. *Field Crops Research,* 57**,** 243-251.

N.L. MINISTRY OF FOREIGN AFFAIRS 2012. Sugar: make it work. A Dutch-Rwanda PPP initiative.

PASSIOURA, J. B. 1996. Simulation models: science, snake oil, education, or engineering? *Agronomy Journal,* 88**,** 690-694.

PENG, Y. & GITELSON, A. A. 2011. Application of chlorophyll-related vegetation indices for remote estimation of maize productivity. *Agricultural and Forest Meteorology,* 151**,** 1267-1276.

PENG, Y. & GITELSON, A. A. 2012. Remote estimation of gross primary productivity in soybean and maize based on total crop chlorophyll content. *Remote Sensing of Environment,* 117**,** 440-448.

PENG, Y., GITELSON, A. A., KEYDAN, G., RUNDQUIST, D. C. & MOSES, W. 2011. Remote estimation of gross primary production in maize and support for a new paradigm based on total crop chlorophyll content. *Remote Sensing of Environment,* 115**,** 978-989.

PENNING DE VRIES, F. W. T. 1989. *Simulation of ecophysiological processes of growth in several annual crops*, Int. Rice Res. Inst.

PORTER, J. 1984. A model of canopy development in winter wheat. *The Journal of Agricultural Science,* 102**,** 383-392.

RDB. 2016. *Ideal conditions for high value products* [Online]. Rwanda development board. Available: http://www.rdb.rw/fileadmin/user_upload/Documents/Agriculture/Sugar%20cane%20production%20and%20processing.pdf [Accessed].

RICHTER, R. & SCHLÄPFER, D. 2002. Geo-atmospheric processing of airborne imaging spectrometry data. Part 2: atmospheric/topographic correction. *International Journal of Remote Sensing,* 23**,** 2631-2649.

ROUSE JR, J. W., HAAS, R., SCHELL, J. & DEERING, D. 1974. Monitoring vegetation systems in the Great Plains with ERTS.

SOLOMON, S., SHAHI, H., SUMAN, A., GAUR, A., DEB, S. & SINGH, I. A SURVEY OF POST-HARVEST BIOLOGICAL LOSSES OF SUCROSE IN INDIAN SUGAR FACTORIES: AN EMERGING CHALLENGE. Proc. Int. Soc. Sugar Cane Technol, 2001. 380-381.

SPITTERS, C., KEULEN, H. V. & VAN KRAALINGEN, D. 1989. A simple and universal crop growth simulator: SUCROS87.

SPITTERS, C. & SCHAPENDONK, A. 1990. Evaluation of breeding strategies for drought tolerance in potato by means of crop growth simulation. *Genetic Aspects of Plant Mineral Nutrition.* Springer.

VAN HEERDEN, P. D. R., DONALDSON, R. A., WATT, D. A. & SINGELS, A. 2010. Biomass accumulation in sugarcane: unravelling the factors underpinning reduced growth phenomena. *Journal of Experimental Botany,* 61**,** 2877-2887.

VAN ITTERSUM, M. K., LEFFELAAR, P. A., VAN KEULEN, H., KROPFF, M. J., BASTIAANS, L. & GOUDRIAAN, J. 2003. On approaches and applications of the Wageningen crop models. *European journal of agronomy,* 18**,** 201-234.

VAN KEULEN, H. 1982. Penning de Vries, FWT, Drees, EM, 1982. A summary model for crop growth. *Simulation of Plant Growth and Crop Production. Simulation Monographs. Pudoc, Wageningen, The Netherlands***,** 87-98.

VELDMAN, M. & LANKHORST, M. 2011. Socio-Economic Impact of Commercial Exploitation of Marshes in Rwanda: Cane sugar production in rural Kigali. RCN contribution to ILC Collaborative Research Project on Commercial Pressures on Land, Rome.

VERHOEF, W. 1984. Light scattering by leaf layers with application to canopy reflectance modeling: The SAIL model. *Remote Sensing of Environment,* 16**,** 125-141.

VIÑA, A., GITELSON, A. A., NGUY-ROBERTSON, A. L. & PENG, Y. 2011. Comparison of different vegetation indices for the remote assessment of green leaf area index of crops. *Remote Sensing of Environment,* 115**,** 3468-3478.

WIEGAND, C., RICHARDSON, A., ESCOBAR, D. & GERBERMANN, A. 1991. Vegetation indices in crop assessments. *Remote Sensing of Environment,* 35**,** 105-119.

YAMANE, T. 2016. *Sugarcane* [Online]. Encyclopedia Brittanica. Available: https://www.britannica.com/plant/sugarcane [Accessed].