University of Southern Queensland

Faculty of Health, Engineering and Sciences

Sugarcane Yield Estimation by UAV Photogrammetry Survey

A dissertation submitted by

Clancy Sharp

in fulfilment of the requirements of

ENG4111 and ENG4112 Research Project

towards the degree of

Bachelor of Spatial Science (Honours) (Surveying)

Submitted: October 2019

ABSTRACT

Sugarcane is a giant tropical grass in the botanical genus, *Saccharum*, the stalks of which are the world's primary source of sugar (sucrose). After wheat, sugarcane is the second largest export crop in Australia with a total annual revenue of about $2.5 billion AUD.

There is a need for accurate and efficient yield estimation models for sugarcane crops, primarily because most of the cane is forward sold in the months leading up to harvest, and for logistical reasons including equipment allocation and harvest scheduling. Existing methods rely on hyperspectral satellite imagery and grower's estimates, both of which have some limitations.

Unmanned aerial vehicles (UAVs), mounted with a visual spectrum (red-green-blue, i.e. RGB) camera may present an efficient and cost-effective method for capturing spatial and spectral data about sugarcane crops and, if processed and analysed properly, this data could be used to estimate the quantity of usable cane stalks in a canefield. Such a technique would be valuable for the sugarcane industry.

In this research project, the existing literature relating to crop height determination by UAV photogrammetry survey, visible-band spectral analysis of vegetation, and sugarcane yield estimation has been reviewed. A methodology was developed and a field study carried out to survey sugarcane crops using a consumer-grade UAV at approximately monthly intervals for three months leading up to harvest, to process the data into 3D digital models using photogrammetry software, to analyse the spatial and spectral properties of the data to find correlations with empirical yield data as recorded during a monitoring survey of the harvest, and to develop yield prediction models using linear regression and multiple linear regression techniques.

The results demonstrate that UAV-based photogrammetry is a suitable method to create digital models of the crop's surface, and that the height of this surface model correlates strongly with empirical yield at all survey epochs. Such a technique is useful for assessing crop variability within fields. Unfortunately, however, mature cane is vulnerable to damage by wind and rain, which can affect its height and subsequently thwart observations about growth rate and yield predictions that are based on height. Visible-band vegetation indices exhibited low or erratic correlations with yield and were subject to influence from many factors including changing ambient light conditions and yellowing of the cane due to frost, thus rendering them an unreliable predictor of yield.

The conclusions of this project indicate promising potential for UAV photogrammetry survey in the sugarcane industry, with recommendations for future research to improve the yield prediction models by input of additional independent variables to overcome the obstacles discovered in this project.

University of Southern Queensland

Faculty of Health, Engineering and Sciences

ENG4111 and ENG4112 Research Project

Limitations of Use

Certification

I certify that the ideas, designs and experimental work, results, analyses and conclusions set out in this dissertation are entirely my own effort, except where otherwise indicated and acknowledged.

I further certify that the work is original and has not been previously submitted for assessment in any other course or institution, except where specifically stated.
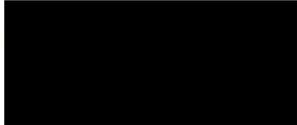
Clancy Sharp

Student Number: █████████

Date: 11 September 2019

Acknowledgements

Studying does hurt. It's a bit like going to the gym. – Dr. Glenn Campbell (2019)

Studying requires one to exercise the mental muscles of self-discipline and critical, rational thinking and, just like going to the gym, the benefits only become apparent after one has persevered through discomfort.

This project was conducted under the supervision of Dr. Glenn Campbell, who I thought might study for fun before the gym analogy. I would like to thank Glenn, not for helping me lift the barbell, but for giving me tips on technique and gently putting on a few extra weights if I wasn't straining. I took inspiration knowing that Glenn was also a surveyor that embarked on a remote sensing topic for his PHD thesis, compared to which my study is just a molehill, but which I felt had some elements of overlap with my topic.

Sugarcane farmers also abide by the reap-what-you-sow axiom in the most literal sense. Malcolm Warren and Johan Lambrechts from Sunshine Sugar at the Condong Sugar Mill, and Paul O'Keeffe, the canegrower and landowner of the subject site, and the harvesting crew, all shared their time and expertise on sugarcane and allowed me to ask naïve questions and run around with my drone and survey pole during one of the busiest times of the year. May you reap the benefits of improved yield estimations in the future.

I would like to say a special *grazie amore* to my partner, Carolina, for being so supportive and cooking so many more dinners than could ever be considered fair this semester. I think I owe you a few dinners now. Believe me, I cannot wait to do just that.

LIST OF FIGURES

LIST OF TABLES

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AHD   Australian height datum

CSM   Crop surface model (elevation relative to terrain)

DEM   Digital elevation model (elevation model relative to datum)

DTM   Digital terrain model (elevation model relative to datum)

GCP   Ground control point

GLI   Green leaf index

GSD   Ground sampling distance

NGRDI   Normalised green red difference index

NIR   Near infrared

PH   Average plant height

PH_{CSM} Average plant height from crop surface model (within a harvest bin polygon)

RGB   Red green blue (refers to visible band of electromagnetic spectrum)

RMSE   Root mean square error

SfM   Structure from motion

TCH   Tonnes of cane per hectare

TGI   Triangular greenness index

VARI   Visible atmospherically resistant index

VI   Vegetation index

VI_{RGB} Vegetation Index using red green blue (visible band) values

VI_{HS} Vegetation index using hyperspectral (eg. Infrared) band values

CHAPTER 1: INTRODUCTION

1.1 Introduction

Each year in Australia, about 35 million tonnes of sugarcane stalks are grown, harvested and crushed to produce about 4.5 million tonnes of sugar, which is used to produce a variety of foodstuffs as well as biofuel ethanol. There is a need for accurate and efficient yield estimation models for sugarcane crops, primarily because most of the cane is forward-sold in the months leading up to harvest, and also for logistical reasons including equipment allocation and harvest scheduling calculations, which aim to allow the cane to grow as long as possible, but still be harvested before the onset of the wet season in November.

Existing yield estimation methods rely primarily on hyperspectral satellite imagery, which is limited because it only observes spectral properties of the upper storey of cane and lacks information about the cane's height, and grower's estimates, which are also limited because the grower cannot easily observe the interior of the field.

In the past decade, UAVs have been used in an ever-widening variety of applications including environmental resource monitoring, media production, hobbyist racing, military surveillance and even commercial deliveries. In many survey practices, they are also fast becoming a standard tool for capturing aerial photographs to be used in conjunction with photogrammetry software to create digital 3D models of a site.

Consumer-grade UAVs are lowering in cost whilst at the same time increasing in capability and are often equipped not only with a high-resolution camera, but also an array of other sensors for manoeuvrability, collision-avoidance, and flight automation (Perritt 2017).

UAV hardware has evolved hand in hand with the complex software used to drive them and process the data. Aerial photogrammetry is no longer the domain of only the cartographer or the surveyor, but much of the innovation is coming from software engineers and programmers, robotics engineers and even computer game developers.

As UAV-software systems become increasingly automated, perhaps less skill is required to capture high quality aerial imagery. For spatial scientists though, unique opportunities can be found through their ability to analyse and interpret the imagery, as well as integrate it with spatial data from other platforms. There is potential for the surveyor's expertise to be exported to other industries in innovative ways.

In precision agriculture, high accuracy spatial data has potential to be used to predict yield of crops. Sugarcane is one crop where this is needed and will be the focus of this research.

1.2 Aim

The aim of this project is to estimate sugarcane yield using imagery captured by visible spectrum, i.e. red green blue (RGB), camera mounted to unmanned aerial vehicle (UAV), by applying 3D photogrammetry techniques integrated with spectral analysis techniques.

1.3 Objectives

To achieve this aim, the following objectives have been set for this project:

1. Review existing literature relating to crop height measurement and yield estimation by UAV photogrammetry survey and/or RGB spectral analysis.
2. Design a model to estimate sugarcane yield using data captured by UAV survey.
3. Gather field data at the subject canefields to parameterise the model, including UAV photogrammetry survey at several epochs, topographic survey of the terrain, and empirical yield data at time of harvest.
4. Parameterise the model and evaluate the effectiveness of predictions.

CHAPTER 2: Literature review

The accuracy and suitability of UAV-based photogrammetry for performing topographic and other types of surveys is an area of ongoing research. Studies that compare and validate the results against 3D scanner and total station-acquired data have been valuable in determining the strengths and limitations of photogrammetry in a variety of conditions.

In agriculture, the results of UAV surveys are often validated by assessing statistical correlations with other remotely sensed data, like hyperspectral aerial or satellite imagery, or with empirical observations about the subject crops, like plant height or biomass.

In order to make a useful contribution to the area of study, it is first necessary to understand what has been done previously in these areas by others, and then to build upon that. This research project draws on, integrates, and builds upon knowledge from 6 main areas:

- Vegetation heighting by UAV photogrammetry
- Using crop height to estimate yield
- Vegetation health assessment by spectral analysis
- Sugarcane yield estimation
- Precision agriculture tools for the sugarcane industry
- Combining spatial and spectral data using multiple linear regression

2.1 Introduction

Photogrammetry means measurement from photographs, and usually refers to deducing 3-dimensional information about an object using overlapping photos of the object taken from different perspectives. The simplest form of this is stereoscopy, a technique by which two overlapping images are used to infer depth information by examining the apparent displacement of an object between the images. Stereoscopy is as old as photography itself, and, indeed, the eyes of humans (and other animals) perceive depth in this same way, a principle that was understood before the invention of cameras.

It is also possible to measure 3-dimensional information using a single camera, by taking a photo then moving the camera to another position and taking another photo. A minimum of two photos are needed, provided that the camera position and orientation (known as exterior orientation), and information about the camera such as principal distance, principal point of autocollimation, lens distortions (known as internal orientation) are known (USQ 2018).

However, if there are many overlapping photos of the object, it is not necessary for the external or internal orientation to be known. These can be computed simultaneously along with the geometry of the object by using the bundle adjustment technique (USQ 2018). Structure from Motion (SfM) is a technique for performing this automatically with computers that utilise machine vision algorithms to identify and track common points in different photos, regardless of the orientation or scale of the photos (Malambo et al. 2018).

The development of SfM computing technology, with its automated image matching capabilities and highly redundant, iterative bundle adjustment, has enabled consumer-grade, non-metric cameras mounted to UAVs to be used for affordable and efficient 3D modelling in a wide range of applications (Malambo et al. 2018).

By 2019, many practitioners are now aware of the usefulness of this technology in applications such topographic surveys on construction sites or mine sites and stockpile volume surveys, where vegetation cover is minimal, and the UAV-mounted camera has line of sight access to the terrain. In such applications, vegetation is often considered a nuisance because it obstructs the camera view and can cause errors in volume results if the model is not processed properly to exclude vegetation.

However, there are also many other applications for this technology, in which the vegetation itself is the object of interest, including precision agriculture, forestry, and natural resource management.

2.2   Vegetation heighting by UAV photogrammetry

In 2003, researchers (Okuda et al. 2004) used aerial photographs and stereoscopy to calculate the height of trees in tropical rainforest in Malaysia by aerial triangulation. They also measured the trunk diameter of the trees during terrestrial surveys and used this data to establish, by regression analysis, an allometric relationship between tree height and trunk diameter. From this, they were able to estimate the total above-ground tree biomass (AGTB) within the rainforest by observing the mean canopy height (MCH) of trees and multiplying by the expected mass, which was predicted using regression analysis of the scatter plot as shown in Figure 1 below. In the same study, Okuda et al. (2004) also used aerial photogrammetry to measure canopy surface height in a 2.5 m grid over the study area and found that the mean canopy surface height for each 20 m x 20 m quadrat was significantly correlated with TAGB estimated using the abovementioned method. These results suggest that there is some uniformity of density within the Malaysian rainforest, which enabled the above-ground biomass to be estimated once the topography of the canopy is known.

Okuda et al. (2004) successfully demonstrated that photogrammetry can be used not only to create a 3D digital surface of the canopy but can be used to further infer qualitative properties of the forest using

known correlations between variables. The aerial photography was captured from a manned aircraft, as this was before UAVs had become popular and widely available.
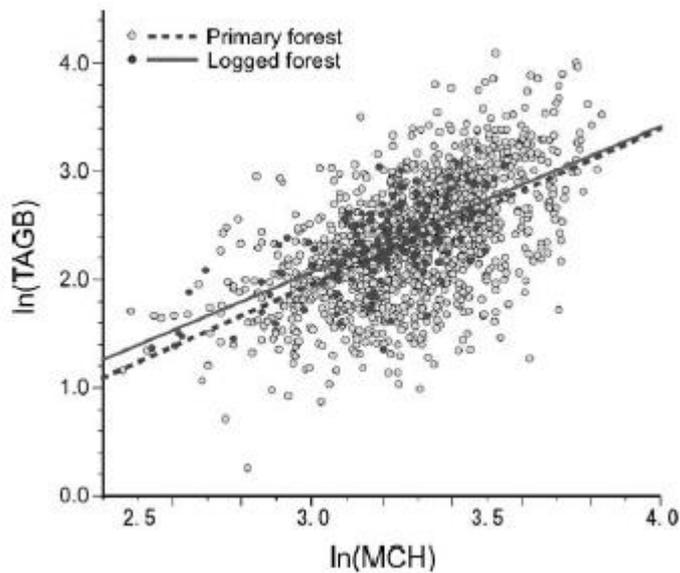
Figure 1: Scatter plot and regression analysis to establish relationship between MCH and TAGB in Malaysian forests (Okuda et al. 2004)

Almost a decade later, UAVs were becoming popular and widely available for surveying and agriculture. In 2013, a group of researchers (Bendig et al. 2013) used an UAV mounted with a camera to conduct photogrammetric surveys of barley crops in west Germany. Surveys were conducted at several epochs throughout the growing of the barley to observe change over time, i.e. a *multi-temporal* approach as depicted in Figure 2 below. They conducted these surveys on 16 plots of barley, comprising 4 replications of each of 4 different cultivars, with the aim of assessing growth variability and its dependency on cultivar, fungicide treatment and stress.
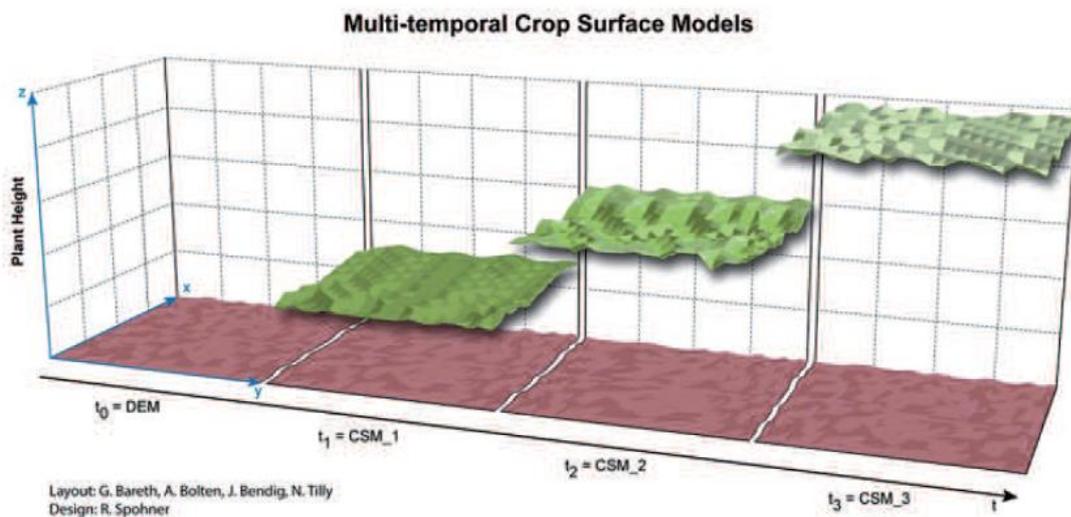
Figure 2: Change of crop height over time depicted by multi-temporal digital crop surface models (Bendig et al. 2013)

Bendig et al. (2013) used specialized photogrammetry SfM software, Agisoft PhotoScan, to generate digital crop surface models (CSMs) of the canopy of the barley for each of the epochs and compared the CSMs to monitor growth rates for each of the plots.

In 2014, researchers in Germany posited that there exists a correlation between crop height and crop yield, or biomass, for many crops including maize and winter wheat (Grenzdörffer 2014). In their study, they did not actually attempt to predict yield of these crops, but rather they compared 3D SfM point clouds computed with several different software programs, Figure 3 below shows a cross section for comparison.
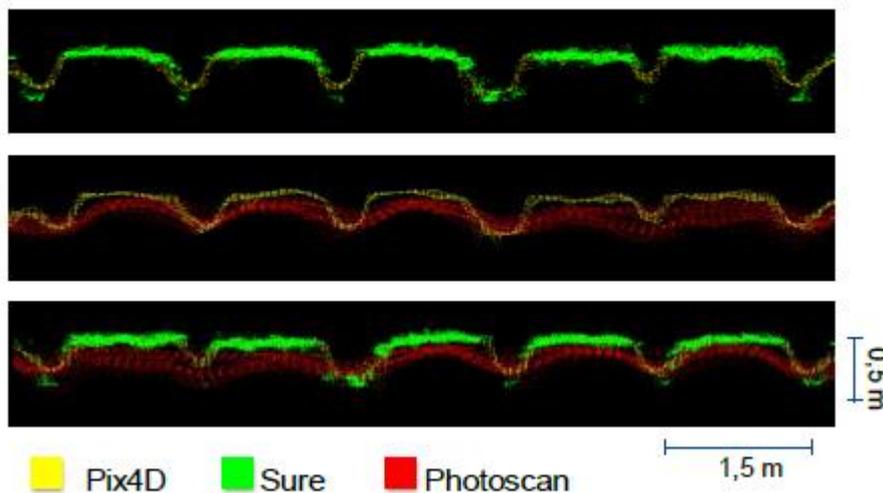
Figure 3: Cross section of 3D point clouds of crop surface computed with software packages Pix4D, Sure, Photoscan (Grenzdörffer 2014)

The suitability of SfM for measuring crops' surfaces has been further tested by comparison with Terrestrial Lidar Survey (TLS) acquired data. Malambo et al. (2018), used aerial imagery acquired from a consumer-grade DJI Phantom 3 UAV, a predecessor to the Phantom 4 used in this study, to create SfM crop surface models of maize and sorghum and used linear regression to compare the results with those obtained by (TLS) for validation. Lidar is considered a suitable benchmark against which results can be measured because it uses actively emitted light pulses to measure direct radiations to 3D point positions, in contrast to the passive technique of SfM photogrammetry which relies on intersection of angles after many variables have been approximated and is hence more error prone (Malambo et al. 2018).

Visual comparison of a longitudinal section of the SfM and TLS point clouds (Figure 4) shows that the SfM cloud is very sparse at the tops of the plants. One possible reason for this is that the tops of plants tend to move in the wind, which would result in discrepancies between the many solutions derived from many matching image pairs for a leaf that is moving, and SfM bundle adjustment algorithms reject

outliers that do not meet a prescribed level of redundancy between matching pairs (Luhmann et al. 2013).
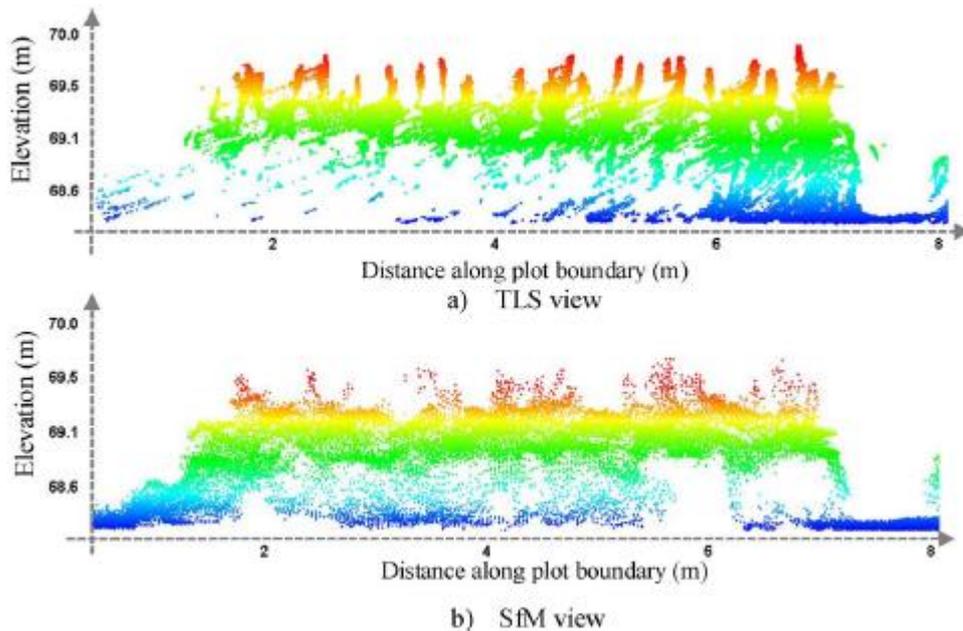
Figure 4: Longitudinal section view of crop point clouds created from TLS and SfM acquired data (Malambo et al. 2018)

The SfM and Lidar point cloud data of 360 crop plots at the 90th, 95th, 99th percentiles and maximum height were compared by linear regression. The R square and Root mean square error (RMSE) correlation results are summarised in Table 1 below, which shows reasonably strong correlation and confirms suitability for SfM as a viable and cost effective alternative to TLS for measuring crop height.

Table 1: Correlation between Lidar and SfM point clouds for sorghum and maize (Malambo et al. 2018)

Regression model	Maize			Sorghum		
	No. plots	R^2	RMSE(m)	No. plots	R^2	RMSE(m)
$90th_{TLS}$ vs $90th_{SfM}$	110	0.60	0.14	250	0.74	0.12
$95th_{TLS}$ vs $95th_{SfM}$	110	0.65	0.12	250	0.77	0.12
$99th_{TLS}$ vs $99th_{SfM}$	110	0.71	0.14	250	0.73	0.13
Max_{TLS} vs Max_{SfM}	110	0.66	0.12	250	0.75	0.13

The crops discussed so far of barley, wheat, maize and sorghum are all from the botanical family, *Poaceae*, which contains all the grasses including sugarcane. Whilst these grasses all have similar structural characteristics and the conclusions may be transferable to sugarcane, there is fortunately a body of work that has applied UAV-based photogrammetry specifically to sugarcane.

Work by De Souza et al. (2017) focused on comparing height maps of sugarcane generated using UAV-acquired images with empirical, ground-based measurements of the cane's height. The researchers used a survey levelling staff to painstakingly measure the average height of mature cane in live sample plots.

They did not measure to the top of the uppermost leaves, but rather the useable stalk length from the soil to the insertion of the top visible dewlap (TVD) leaf as shown in Figure 5.
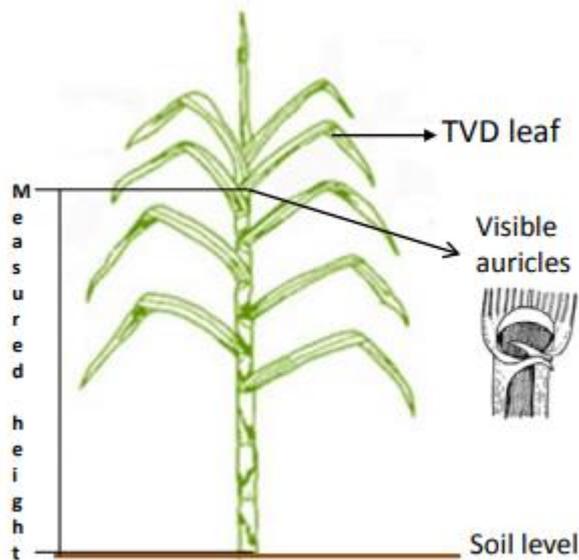
Figure 5: Ground truth stalk measurement technique to top visible dewlap (TVD) leaf, used by De Souza et al. (2017) and Portz et al. (2012)

The UAV imagery was acquired from an altitude of 200 m, with a 75% forward lap and 70% side lap, providing a ground sampling distance (GSD) of 106 mm / pixel. The study also compared results of using imagery acquired from the UAV flying in north-south lines with that from east-west lines. The authors conclude that "different flight directions influenced the generation of the crop surface model" (De Souza et al. 2017), and further conclude that when both sets of imagery are processed into a single surface model the results more closely match ground-based measurements of the cane, in line with the general principle that larger sample sizes give results closer to the true value (provided the data is free from systematic or gross errors).

No explanation for the difference in results between flight directions is offered, and it is possible that the differences the authors observed are a result of random errors or from other changing conditions such as ambient light or wind on the cane. As this conclusion was based on only a single flight in each direction, further testing to compare surface models from different flights in *the same* direction could confirm or challenge this conclusion.

Regardless of the questionable conclusion, the research by De Souza et al. (2017) is valuable because it compares empirical height measurements of sugarcane with UAV SfM data captured from a considerably high altitude, and the results, importantly, as depicted in Figure 6 below, showed CSM cane heights from SfM were consistently lower than actual cane heights. The blue column in Figure 6 shows that CSM height converges with actual height with incorporation of more overlapping photos. It appears, then, that the CSM represents a 'pseudo-surface' that is not only lower than the tops of the

8

uppermost leaves, but can be lower than the TVD surface that is the tops of the useable stalks. This may have implications for regression line between CSM height and true height; the effect would be a y-intercept of a value of less than zero.
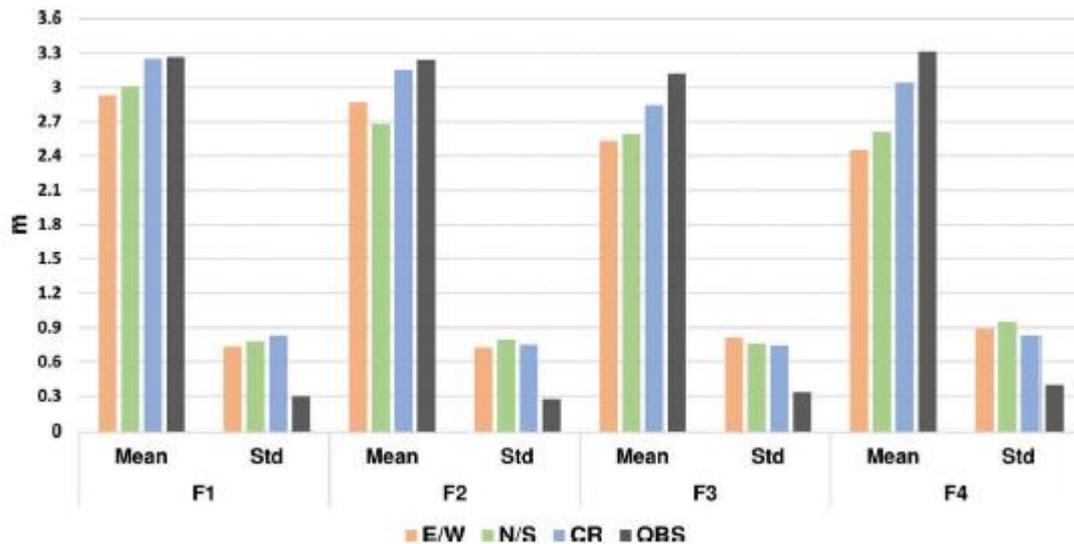
Figure 6: Histogram showing mean and standard deviation of each approach for measuring height (E/W, N/S, combined E/W & N/S (CR) and Observed) for four test fields (De Souza et al. 2017)

This body of work by others discussed so far has focused mainly on *if* an accurate crop surface model can be captured by UAV photogrammetry survey. Their results shown that it can indeed be done, but this outcome was not so obvious to begin with, because crop surfaces, especially crops in the grass family, *Poaceae*, have several characteristics that might seem fundamentally incompatible with the principles of photogrammetry. Firstly, stalks and leaves of these crops move and sway back and forth in even a gentle breeze, and this movement would produce large errors when computing elevations by intersecting narrow angles from the perspective of a UAV if outliers were not sufficiently eliminated. Upper leaves, being more exposed to breezes, are likely to be more to move and thus be eliminated, which might account for the lower CSM height obtained by De Souza et al. (2017). Secondly, these crops often span large areas and appear very monotonous and repetitive to the human eye (Figure 7); one square meter looks indistinguishable from the next and it seems almost impossible to match up leaves and stalks between photo pairs. Finally, such crops do not actually have a well-defined canopy surface but rather, in the case of sugarcane, leaves that point upward and taper to a point, creating a fuzzy, spiky field as seen in Figure 8.

Figure 7: Top view close-up (taken from 120 m high) of the subject sugarcane crops of this study, depicting monotonous and repetitive pattern

Figure 8: Side view of the subject sugarcane crops of this study, depicting the spiky, fuzzy, poorly defined surface

The fact that the SfM software can resolve a point cloud as well as it does for these crop surfaces is testament to the algorithm's ability to use the high redundancy of overlapping photos to eliminate pixels of leaves that are swaying in the wind. The monotonous texture of crop surfaces does not pose the same problems for machine vision as for human vision, because algorithms do not 'see' a field of grass, they systematically scan the images and match up similar pixel value sequences between images, first aligning images using the most stand-out and inimitable patterns, then performing iterations to match increasingly subtle patterns (Luhmann et al. 2013).

2.3   Using crop height to estimate yield

The literature contains a body of work by researchers who used CSMs to predict yield or biomass of crops with regression models that establish correlation between crop height and empirical, ground-based, measurements of yield.

Portz et al. (2012) did not use a UAV, photogrammetry, or digital CSMs, but rather compared empirical measurements of sugarcane stalk height (measured to TVD as in Figure 5) with empirical measurements of dry biomass, using linear regression to establish correlation for both early and late season varieties of cane in their earlier stages of growth (stalk length <1.4 m). The authors note in their conclusions, that the linear form of the regression function, depicted in Figure 9, distinguishes sugarcane from other Poaceae crops, which tend to exhibit upturned curves as will be discussed below. Perhaps this is because this is because most of the biomass of sugarcane lies in the thick cylindrical stalk, which, as it grows, may increase in length without increasing proportionally in diameter.
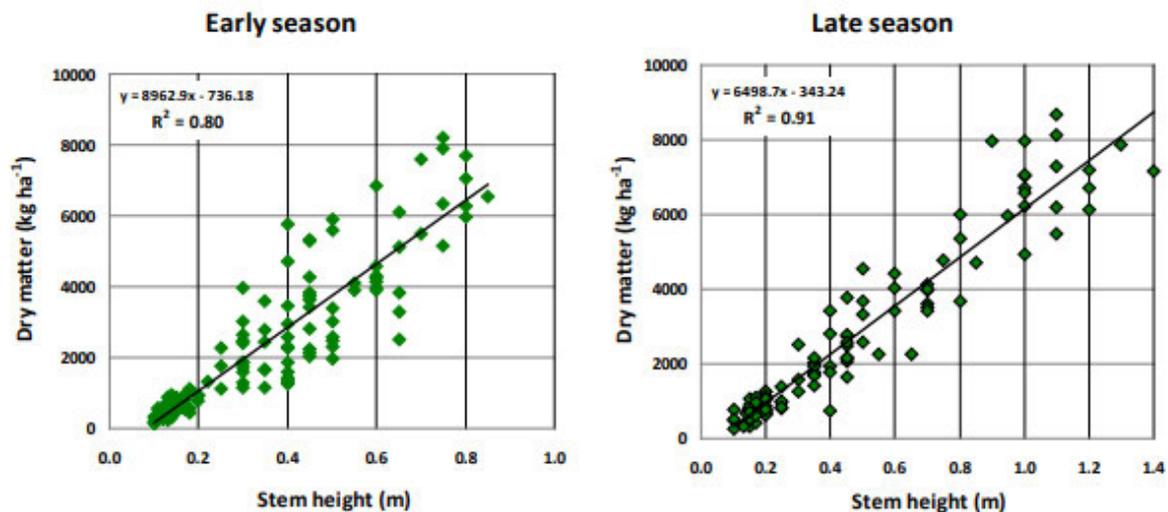
Figure 9: Correlations between sugarcane height and biomass (Portz et al. 2012)

Pioneers in the use of UAVs combined with SfM photogrammetry for this crop yield estimations were seeking to improve on other recently-devised methods that used UAVs combined with hyperspectral sensors that used Vegetation Indices (VIs) but not CSMs (Hunt et al. 2011), or terrestrial laser scanners, which used CSMs but could not measure large areas of crops efficiently (Tilly et al. 2014).

In 2013, researchers in Germany set up 36 small (3 x 7 m) plots of barley, with each plot divided into a 3 x 5 m area for CSM measurement and a 2 x 5 m area reserved for destructive sampling (Bendig et al. 2014). GCPs were placed at approximately 15 m spacings and a UAV fitted with a 16-megapixel RGB camera was flown at an altitude of 50 m above ground level (a GSD of 0.009 m), capturing around 400 images on each of 6 flights throughout the growing season plus one of the bare terrain before planting. This resulted in an image overlap of >9 images for any part of the test area. On the same day as the flights, small destructive samples were taken, and biomass was weighed both fresh and after drying.

Point clouds were generated using Agisoft PhotoScan Professional, and sample area polygons were defined 0.3m in from the edges of the barley plots to eliminate anomalies in the outer edges of the crop, average plant height (PH) was computed within these polygons and this was plotted firstly against plant

height reference measurements taken from the ground, then against destructive biomass sampling results to develop yield prediction regression models.
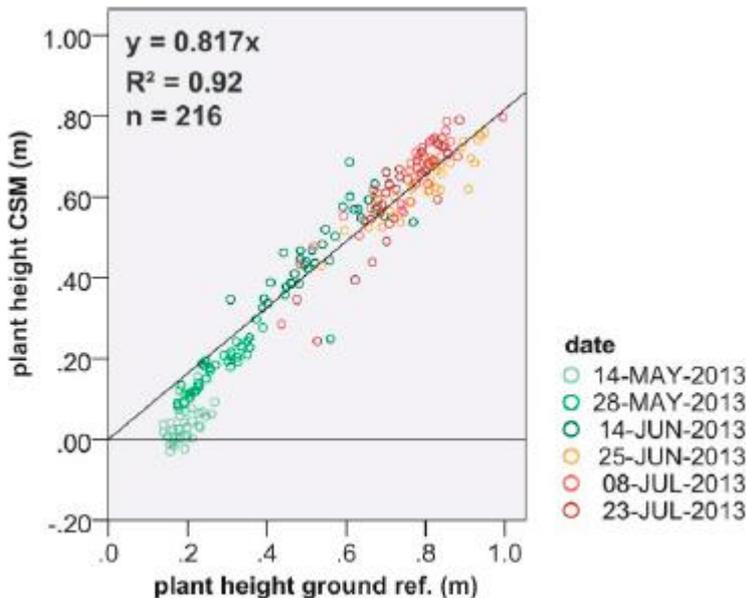
Figure 10: A scatter plot showing correlation between barley plant height as measured from the ground (x axis) against as measured from photogrammetry-derived CSM (y axis) (Bendig et al. 2014)

Figure 10 above shows a strong correlation between plant height by CSM (PH_{CSM}) and by ground measurements, though PH_{CSM} is only about 0.817 of the ground-measured height. Bendig et al. (2014) posit that this is because the PH_{CSM} value is an 'average' within the sample polygon, whereas ground measurements were taken to the highest points of the plants. The R^2 value of 0.92 means that 92% of the variance on plant PH_{CSM} is explained by the regression line. N is the total number of samples, at 216, which is the 36 barley plots for each of 6 UAV flights.

The regression line in Figure 10 has the y-intercept fixed at zero, which seems like a reasonable thing to do when comparing different measurements of the same variable on both axes. However, in light of other studies (De Souza et al. 2017) that demonstrate SfM does not perform well at detecting the upper extremities of grass limbs, it is arguable that the PH_{CSM} is not an 'average' in the sense that it depicts the mean of the full heights of all plants, tall and short, within the sample area, but rather it is a 'pseudo-surface', somewhat below the upper plant extremities, where the SfM algorithms find sufficiently redundant pixel solutions to meet their pre-programmed tolerances. By this argument, it would be better to let the regression calculation do its work and accept whatever coefficient and constant is returned, *especially* if one has the luxury of plotting against ground truthed values of the same variable. Certainly, a visual assessment of the scatter plot in Figure 10 would support this argument, as a negative constant and greater coefficient would clearly improve the fit.

To develop yield prediction models, the 216 samples were randomly split into 70% calibration datasets and 30% validation datasets, and exponential regression curves with PH_{CSM} as the independent variable and fresh and dry biomass as dependent variables.
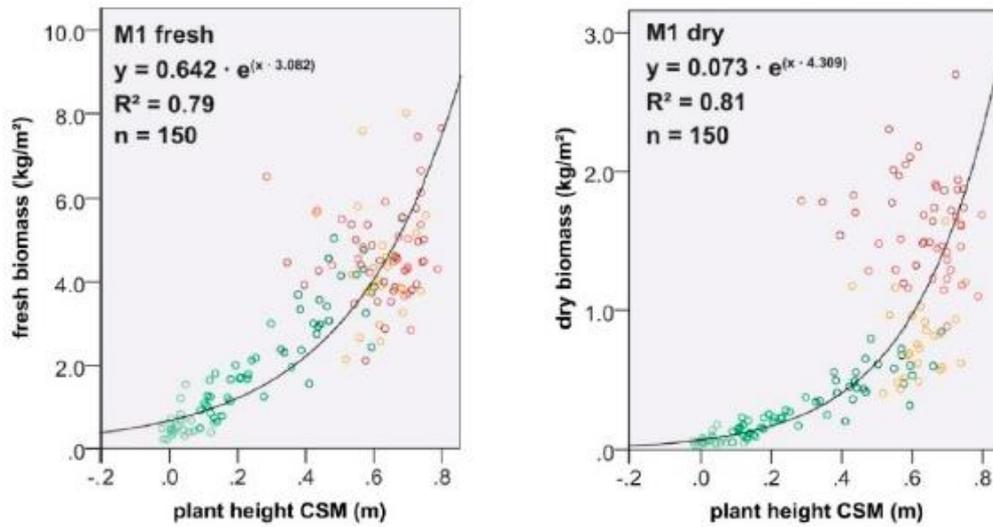
Figure 11: Yield prediction calibration models that exponential regression curves to describe relationship between CSM-derived plant height and fresh biomass (left) and dried biomass (right) by Bendig et al. (2014)

Bendig et al. (2014) have found that, with barley, unlike sugarcane, the relationship between plant height and biomass is best described by an upturned curvilinear function. This makes sense considering the development of the heavy grain-bearing head after around the 40th day of growth as seen in Figure 12 below, which coincides with the yellow dots in Figure 11.
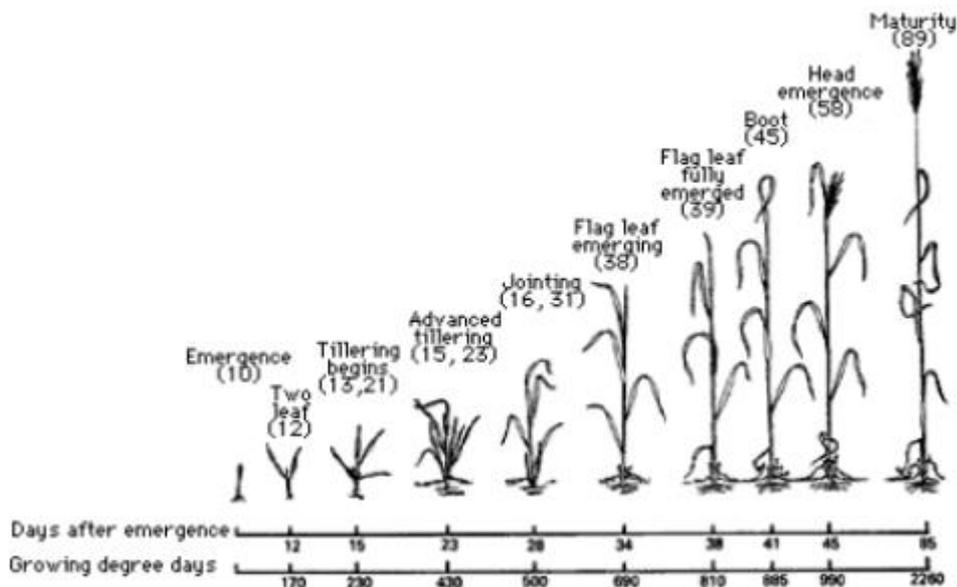
Figure 12: Timeline of barley plant development (https://extension.umn.edu)

2.4 Vegetation health assessment by spectral analysis

Vegetation indices (VIs) are expressions of the amounts of light from different wavelengths reflected by plants. There are many different indices, whose formulas exaggerate various bands of the electromagnetic spectrum, or ratios between bands, with the aim of inferring properties of the vegetation, such as the amount of chlorophyll, which are indicative of plant health.

Some of the most commonly used, and most successful, indices for determining plant health examine the ratio of red to near-infrared reflected light (Grenzdörffer 2014). This is because chlorophyll, which is a photosynthetic pigment used by plants to convert light into chemical energy, strongly absorbs red light (and other light within the 400 to 700 nanometer range) whilst the cell wall within the leaves strongly reflects near-infrared light (within the 700 to 1100 nanometer range) (Gates 1980). This causes healthy leaves to reflect a ratio of these two spectral bands that is quite unique, enabling it to be distinguished from vegetation that is low in chlorophyll or from other reflective surfaces such as soil or water.

Consumer grade UAVs, however, are typically not equipped with cameras that can detect near-infrared light. Rather, they utilise RGB cameras and can only detect wavelengths within the visible region of the electromagnetic spectrum, between about 380 to 740 nanometers (Star 2005). Many VIs that use only RGB light (VI_{RGB}) have been developed and tested to assess health of vegetation, with varying degrees of success. Research by others on this topic will be reviewed below, but first it is important to understand that RGB pixel colours are only a proxy for reflectance, and to consider some of the the factors that can cause them to be a poor proxy.

2.4.1 Orthomosaics from RGB Images

Due to the phenomena of reflectance and scattering, the appearance of vegetated surfaces (i.e. their apparent reflectance, defined by the observable wavelengths of light and their intensity) varies depending on the direction from which the vegetation is viewed and the angle of the light source, collectively called the Bidirectional reflectance distribution function (BRDF), or simply 'angular effects' (Aasen 2016). Smooth and regular mirror-like surfaces, such as the surface of a lake, tend to exhibit *specular reflectance* in which the angle of incidence is equal to the angle of reflectance for all incoming light rays. Rough surfaces that have facets in all directions, such as freshly fallen snow, tend to exhibit *Lambertian scattering* in which incoming rays are reflected equally in all directions of the hemisphere. Plant canopy surfaces typically exhibit a type of *anisotropic asymmetric scattering* in which light is most strongly reflected back in the direction it came from, and as a result will appear different depending on the viewing angle. The zone that appears brightest in aerial images is called a 'hotspot' (Aasen 2016).
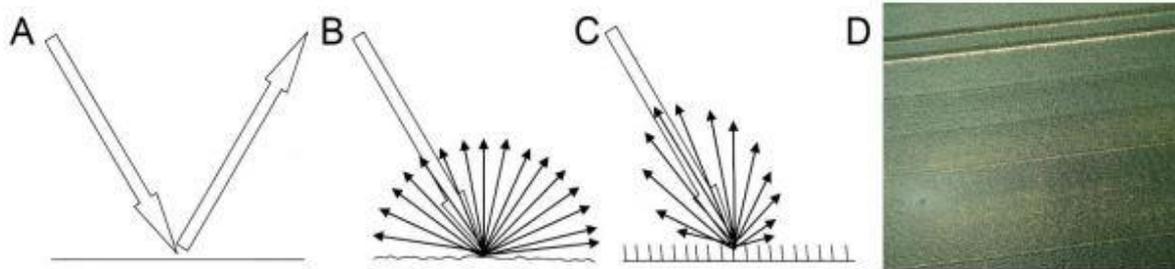
Figure 13: Specular reflectance (A), Lambertian scattering (B), anisotropic asymmetric scattering (C) and an aerial RGB image of a wheat field showing a 'hotspot' resulting from anisotropic reflectance, with the shadow of a UAV visible in the centre (Aasen 2016)

This property of plant canopy surfaces presents a problem when applying vegetation indices to single camera images. Bendig et al. (2015) addressed the issue by using Agisoft PhotoScan Professional software to generate an orthophoto mosaic, where a composite image is constructed from many overlapping images taken from a range of angles. Each pixel's colour value is taken from the most appropriate photo, as determined by the Agisoft software's 'mosaicing blending mode' (Bendig et al. 2015). Unfortunately, details on how, i.e. the algorithms by which, 'the most appropriate photo' is selected are not available due to the proprietary nature of the software.

Nevertheless, this technique has been used to eliminates hotspots, correcting somewhat for anisotropic asymmetric scattering in in predicting sugarcane yield by Sanches et al. (2018), who used a DJI Phantom 3 with RGB camera to generate orthomosaics, and employed vegetation indices to assess the degree of canopy closure. Admittedly, however, detecting spectral differences between bare earth and living plants is likely to be far easier than detecting the more subtle spectral differences between similar living plants with slightly varying yields.

The concept of using orthomosaics to reduce angular effects was not adopted Possoch et al. (2016) in assessing suitability of VI_{RGB} to predict biomass of grassland in Germany. Instead, researchers used a single image taken from an altitude sufficient to capture the entire field experiment. Lens distortions and effects including vignetting and chromatic aberration were then supposedly corrected for using *PTLens* software, and image was georeferenced using GCPs. This study did not find strong correlations between biomass and VI_{RGB} in the subject grassland, and noted difficulties in correcting the distorted images (Possoch et al. 2016). It should be noted, however, that in the same study, ground-based hyperspectral measurements taken with a handheld spectrometer also did not show strong correlations with biomass, indicating that the lack of correlation may be a characteristic of the grassland rather than a flaw in the method.

Whilst use of orthophoto mosaics may help to overcome some obvious angular effects like hotspots, there are other factors that contribute to RGB pixel values not being a true representation of the

vegetation's colour. Importantly, the camera's sensor is only capable of sensing a few narrow bands of the electromagnetic spectrum, the exact values of these bands varies between different sensor models, and such details about the hardware are not usually provided to the user in the product specifications for consumer-grade cameras.

Furthermore, the camera settings such as ISO, aperture, and shutter speed all affect the recorded value of the pixels within an image. In some cameras these can be manually set to fixed values, but given that the ambient light is also changing as the sun moves and from atmospheric effects including cloud cover, it may be just as valid to use the camera in 'automatic' mode as has been done by Sanches et al. (2018). Vegetation indices, as opposed to absolute colour values, fortunately reduce the impact of such brightening or darkening effects because they express only the ratio between colour bands.

The RGB pixel values are therefore only a proxy for reflectance, and information gathered from vegetation indices assumes that the difference between the observed proxy and the true reflectance is the same for the subjects being compared.

2.4.2   RGB Vegetation Indices

The use of RGB vegetation indices (VI_{RGB}) by Possoch et al. (2016) and Sanches et al. (2018) has already been described above, which was unsuccessful in the case of the former (but neither was hyperspectral index) to predict grassland biomass, and was successful in the case of the latter, but only for observing obvious differences between bare earth and living plants.

VI_{GRB} has also been used successfully to estimate flower number in oilseed canola (Wan et al. 2018), who found that visible indices performed slightly better than hyperspectral indices to detect the bright yellow flowers of the plant.

In sugarcane, however, it is not brightly coloured flowers, but chlorophyll content that has been demonstrated to correlate with yield (Robson et al. (2013), reviewed below). A study by Hunt et al. (2013) compares the performance of several hyperspectral indices (VI_{HS}), using Airborne Visible InfraRed Imaging Spectrometer (AVIRIS) and Thematic Mapper (TM) remotely sensed imagery, with four VI_{GRB}s for correlations with leaf chlorophyll content at the canopy scale in maize crops. Results are shown in Table 2 below.

Table 2: Table of correlations between leaf chlorophyll content in corn and several VIHS and VIRBG (Hunt et al. 2013)

Index	AVIRIS	TM bands	Camera bands
RVI	0.82	0.82	–[a]
NDVI	0.82	0.82	–
SAVI	0.74	0.64	–
MSAVI	0.76	0.66	–
OSAVI	0.79	0.73	–
EVI	0.73	0.61	–
TVI	0.64	0.50	–
MTVI2	0.72	0.64	–
CVI	0.92	0.91	–
gNDVI	0.89	0.88	–
CI-G	0.90	0.89	–
NGRDI	−0.92	−0.89	−0.87
GLI	−0.91	−0.90	−0.89
VARI	−0.91	−0.91	−0.84
NDREI	0.76	–	–
CI-RE	0.76	–	–
MTCI	0.89	–	–
MCARI	−0.89	–	–
TCARI	−0.88	–	–
TCI	−0.89	–	–
TCARI/OSAVI	−0.89	–	–
MCARI/MTVI2	−0.89	–	–
TGI	−0.91	−0.91	−0.92

Results obtained by Hunt et al. (2013) suggest that the VI_{RGBS}, NGRDI, VARI, GLI, and TGI are suitable indicators of chlorophyll. It is these VI_{RGBS} that will be used in this study, their formulas are given later in the methodology.

2.5 Sugarcane yield estimation

The main variables within a sugarcane crop that contribute to final yield are number of stalks, thickness of stalks, and length of stalks. Since it is impracticable to measure these directly in large crops, efforts are instead focused on estimating overall plant height, plant diameter, number of stalks per plant, and number of plants in the crop (De Souza et al. 2017).

Also in the sugarcane industry, a team of researchers in Brazil (Zhao et al. 2016) experimented with using satellite imagery to predict sugarcane yield. More specifically, they were using time-series satellite images to monitor spectral properties of the cane and attempting to correlate this with sucrose content.

Some of the more recent efforts to predict sugarcane yield focus on a multi-model approach (Dias & Sentelhas 2017), whereby the results from 3 different simulation models for predicting yield are compared to the mean of the 3 models, then adjusted by the application of a correction factor which

brings all estimates closer to the mean. The 3 models are FAO-AZM, DSSAT/CANEGRO and APSIM-Sugarcane, which are complex mathematical algorithms that require the input sugarcane cultivar and growing conditions parameters. Dias & Sentelhaus (2017) identify that there are significant discrepancies in the yield volumes predicted using each of these methods. This study highlights the need for more accurate prediction methods.

Analysis of, and comparison with these simulations is outside the scope of the proposed research, as they do not use aerial photogrammetry or any other remote sensing platform to monitor the sugarcane. These models and the multi-model approach have been included in this literature review because they represent the current state of the technology for sugarcane yield forecasting.

2.6 Precision Agriculture Tools for the Sugar Industry

Precision agriculture applications for remote sensing of sugarcane were investigated and developed in a collaborative project by the Department of Agriculture, Fisheries and Forestry, QLD, the UNE Precision Agriculture Research Group and CSIRO, which ran from 2011 to 2013 (Robson et al. 2013).

A range of active and passive remote sensing systems were tested and compared, including passive satellites ALOS, SPOT5, SPOT4, RapidEYE, IKONOS, GeoEYE and Raptor aerial active sensor; an instrument that attaches to a low-flying aircraft records reflectance of light emitted from its LEDs. The authors found SPOT5 satellite imagery (now decommissioned) to be very useful because of the spectral bands imaged, which included green, red, near infrared and mid infrared, that allowed for testing of several accepted vegetation indices, of which GNDVI (green normalised difference vegetation index) proved to have the highest correlation with sugarcane yield in terms of tonnes of cane per hectare (TCH).

Table 3 is a correlation matrix comparing the more widely used NDVI value, derived from remotely sensed imagery, with TCH and commercial cane sugar (CCS) for a site in Bundaberg, with strong TCH correlations of >0.7.

Table 3: Correlation matrix of R values comparing NDVI values derived from imagery captured by various remote sensing platforms with yield and commercial cane sugar (CCS) for a site in Bundaberg.

Block 1	Raptor 23.02.11	Raptor 23.03.11	IKONOS 23.03.11	SPOT 27.03.11	IKONOS 30.04.11	Raptor 02.05.11	TCH	CCS
Raptor 23.02.11	1	0.8738	0.5380	0.5498	0.4024	0.3436	0.6291	-0.5211
Raptor 23.03.11		1	0.7588	0.8292	0.6264	0.5913	0.7845	-0.3755
IKONOS 23.03.11			1	0.9242	0.9620	0.8501	0.8408	-0.4565
SPOT 27.03.11				1	0.8875	0.8645	0.8460	-0.2630
IKONOS 30.04.11					1	0.9082	0.7510	-0.4206
Raptor 02.05.11						1	0.7474	-0.3440
TCH							1	-0.5614
CCS								1

The Raptor aerial sensor not only performs marginally worse but is prohibitively expensive to transport the aircraft on demand. Satellite platforms provide a reasonably cost effective and accurate method of gauging NDVI values that have a strong correlation with yield. Unfortunately, however, the resolution of the images of the best performer, SPOT5, is 10 m, which means they are not suitable for identifying small areas of damaged or poorly performing cane within paddocks, for example due to weed infestation or damage by feral animals (Robson et al. 2013).

One of the other major limitations of satellite imagery is that it is obstructed by cloud cover. Unfortunately, the sugarcane growing regions of New South Wales and Queensland tend to have a high degree of cloud cover in the period between January and May, which coincides with the transition from vegetative development to maturation of the sugarcane (Robson et al. 2013). It is within this crucial period that farmers, if they had access to high resolution yield maps, might be able to take final measures to rectify any problems.

Another tool of precision agriculture used in the sugarcane industry is the mechanical yield monitor, which can be fitted to harvesters to monitor the spatial variability of cane within fields, presented as a yield map in Figure 14. The yield map clearly shows that the cane is not distributed homogeneously but can vary by more than 200 t/ha. These maps are excellent for showing stronger and weaker performing areas of fields, albeit at a time when it's too late to do anything about it until the next season. It should be noted that the harvesters in the Tweed Valley, where this research is being conducted, do not have yield monitors fitted to them, which, if they did, would be useful for ground truthing the UAV photogrammetric data.
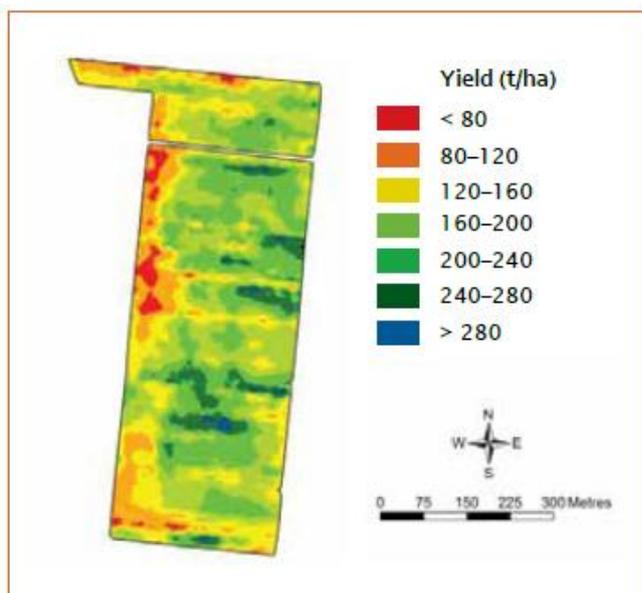
Figure 14: A yield map produced with data collected from a yield monitor fitted to a harvester (Olsen & Hussey 2015), showing variability of more than 200 t/ha

2.7 Combining spatial and spectral data with multiple linear regression

Multiple linear regression (MLR) is a statistical technique that uses multiple independent variables to predict the outcome of a dependent variable. In this study the independent, i.e. explanatory, variables are sugarcane height and spectral properties, and the dependent, i.e. response, variable is sugarcane yield.

A similar approach was adopted by Geipel et al. (2014) to predict corn yield. Building upon previous research that found simple linear regression to be the most suitable method for establishing relationships between corn between height and yield, the researchers in Germany conducted experiments to integrate an vegetation index variable into the prediction model using MLR (Geipel et al. 2014).

Geipel et al. (2014) parameterised the MLR model with empirical harvested corn grain yields of 64 sample plots and evaluated the effectiveness of predictions using RMSE statistical indicator to assess residuals. The empirical yields, predicted yields, and residuals are depicted in Figure 15 below. The RMSE is 0.68 t/ha, equivalent to 8.8% of the total yield.
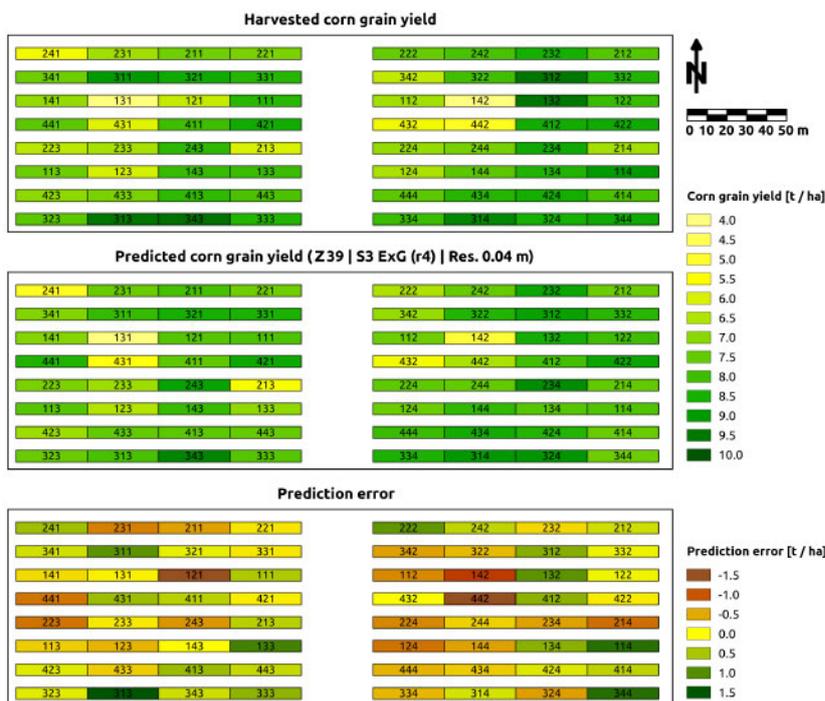
Figure 15: Spatial illustration of plot-wise distribution of harvested corn grain yield (top), predicted corn grain using MLR (middle), and prediction residual (bottom) (Geipel et al. 2014).

In another study (Bendig et al. 2015) researchers compared estimations of barley biomass by linear regression using PH_{CSM} and several VI_{RGB} and VI_{HS} alone and in combination by MLR. Results are shown in Figure 16 below using scatter plots and R square values of predicted against observed biomass.
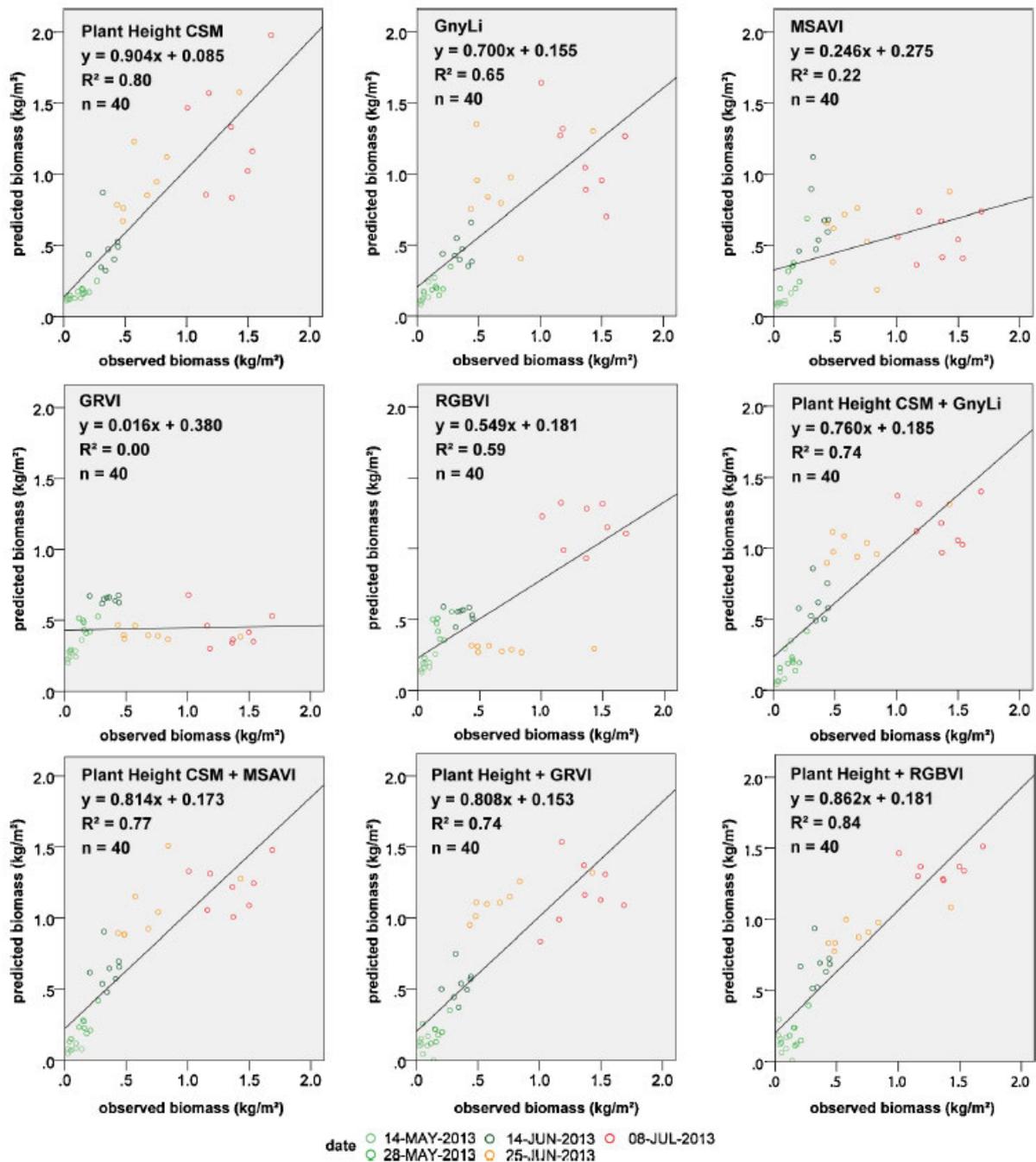
Figure 16: Barley biomass scatter plots and regression funtions for predictions made using various linear regression and MLR models (Bendig et al. 2015)

The methodology by Bendig et al. (2015) entailed gathering empirical biomass samples from all the sample plots at each survey epoch, which enabled detailed analysis of how biomass changed over time. In this research project, only empirical sugarcane yields will only be measured at final harvest, and survey data from all survey epochs will be analysed by regression with final yield. The methodology is described in detail in the next chapter.

CHAPTER 3: Methodology

3.1 Study site and field design

The study site is a 380 m x 100 m area at Tumbulgum, New South Wales (Latitude: 38.28 S, Longitude: 153.49). The site contains 2 individual sugarcane fields (numbered 205 & 206), each with an area of about 1.7 ha.

This site has been selected because these fields are planted with one of the most commonly grown varieties of sugarcane in the Tweed region, Q240, which is an early season variety of sugarcane and is typically harvested in July. These fields are the most suitable because they are growing in side by side in similar conditions.

Figure 17 below shows the study site and the layout of the fields.

Figure 17: Study site. Google Earth aerial imagery showing location of the site and layout of the fields

3.2 Ground Control Survey

Prior to UAV survey, 8 control marks were placed in the ground around the perimeter of the site, as shown in Figure 18 below. The marks are a nail in timber dumpy peg driven into the gravel roads. In principle it is favourable to have control points evenly spaced around the outer extents of the subject site so that the model is being interpolated between these 'fixed' points rather than extrapolated outside them. Accuracy would be further improved by having additional control points toward the centre of this site, but in this is not possible due to the density and height of the sugarcane.
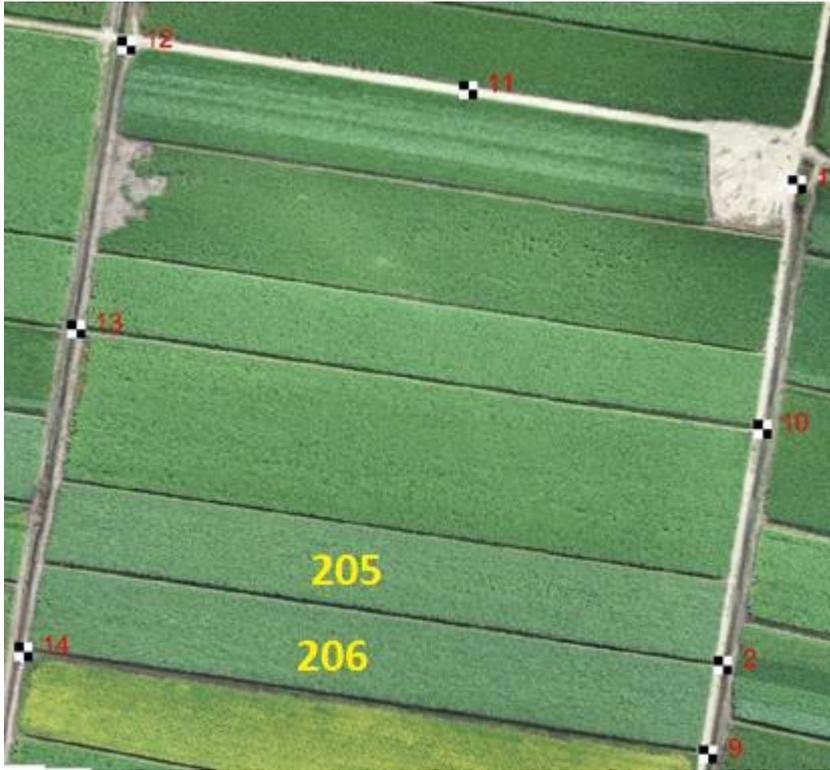
Figure 18: Aerial orthomosaic image of the site showing ground control points

After each mark was placed in the ground, its position was measured using a Leica GS03 GNSS sensor with Real Time Kinematic (RTK) corrections. The easting and northing coordinates are computed on the Map Grid Australia (MGA) Zone 56 UTM projection, based on the Geodetic Reference System 1980 (GRS80) ellipsoid, and the height is computed using AUSGeoid09 geoid model giving elevation relative to the Australian Height Datum (AHD). Points were observed for 2 minutes (120 epochs) at the beginning of the project in April, and again for 2 minutes at the end of the project in July, to obtain a 3D positional uncertainty of ≤10 mm. Figure 19 below shows the GS03 mounted to survey pole being placed on the control peg ready for GNSS observations.

Figure 19: Placement and GNSS observation of control points

During the aerial surveys, a large and highly visible target was temporarily placed on each ground control point, so they are visible in the UAV-acquired images. Figure 20 below shows one of the targets, which are made from squares of plywood painted black and white with a hole drilled in the centre that sits directly on the control point nail.
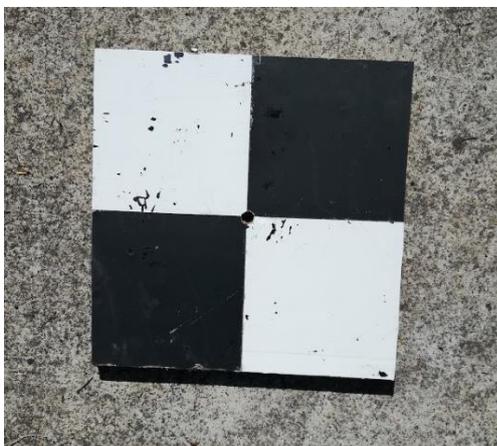
Figure 20: A target made from painted plywood

3.3 UAV Photogrammetry Survey of Sugarcane

3.3.1 Platform

The UAV is a Phantom 4 Pro V2.0 quadcopter by DJI. It has a diagonal size of 350 mm (without propellers) and weighs 1375 g (including battery and propellers) (DJI 2019). It has a maximum speed of 50 km/hr in P-mode (positioning mode) i.e. when all sensors are active, which is the standard mode for this type of use (DJI 2019).

In addition to the camera, the UAV has the following built-in sensors to assist in positioning and collision avoidance:

- Internal measurement unit (IMU) with 3-axis accelerometer and gyroscope for stabilisation and acceleration and spin tracking.

- GNSS receiver that uses GPS/GLONASS satellites for point positioning.

- Vision System –Stereoscopic camera pairs (facing forward, backward, and downward) and integrated processor for real time depth sensing of surfaces with a clear pattern and adequate lighting. Obstacle sensory range of 0.7 to 30 m.

- Infrared Sensing System - Active infrared emitter and sensor on each side for detecting proximity to surfaces with sufficient reflectivity. Obstacle sensory range of 0.2 to 7m.

- Sonar sensing system to assist with safe landing.

The battery is a rechargeable 5870 mAh 15.2 V lithium polymer (LiPo 4S) 'Intelligent Flight Battery' that weights 468 g. When fully charged it gives a maximum flight time of approximately 30 minutes (DJI 2019).

The photogrammetry survey flights follow pre-determined flight path at prescribed altitude and and image capture intervals, being controlled remotely by a smart phone running the DroneDeploy flight plan app, connected by usb cable to the controller. The controller communicates with the UAV by radio transmission at up to 7 km range (DJI 2019).

Figure 21: DJI Phantom 4 Pro V2.0 quadcopter UAV and controller with smartphone connected by usb cable

3.3.2 Sensor

The RGB (red-green-blue band) camera mounted to the UAV has the following specifications (DJI 2019):

- Make: DJI
- Model: FC6310s
- Sensor type: 1" CMOS (complementary metal oxide semiconductor)
- Effective megapixels: 20
- FOV 84° 8.8 mm/24 mm (35 mm format equivalent) f/2.8 - f/11 auto focus at 1 m - ∞
- Iso range: 100-3200
- Mechanical shutter speed: 8 – 1/2000 s
- Electronic shutter speed: 8 – 1/8000 s

In addition to storing the red, green and blue values for each pixel, the camera stores 132 entries of metadata about each photo, including:

- Date Time Original
- Gps Altitude
- Gps Longitude
- Gps Latitude
- Speed X
- Speed Y

- Focal Length In35Mm Format
- Fov
- Shutter Speed
- Aperture
- Scale Factor35Efl
- Calibrated Optical Centre X

- Speed Z
- Roll
- Pitch
- Yaw

- Calibrated Optical Centre Y
- Calibrated Focal Length
- Circle Of Confusion

3.3.3 Data Acquisition

Prior to UAV take off, the timber targets were placed on the ground control points. The thickness of the targets is 20mm, which was noted and added to the RL of the mark observed previously by GNSS.

The UAV and the controller are powered on, the smartphone is connected to the controller and the DroneDeploy app is activated. The app is used to select the flight area on the map screen, and altitude, front and side overlap are nominated at 75% front overlap and 65% side overlap. The map area (in hectares), the duration, the number of images and number of batteries required are automatically calculated. Figure 22 below shows a series of screenshots from the DroneDeploy app during flight planning.
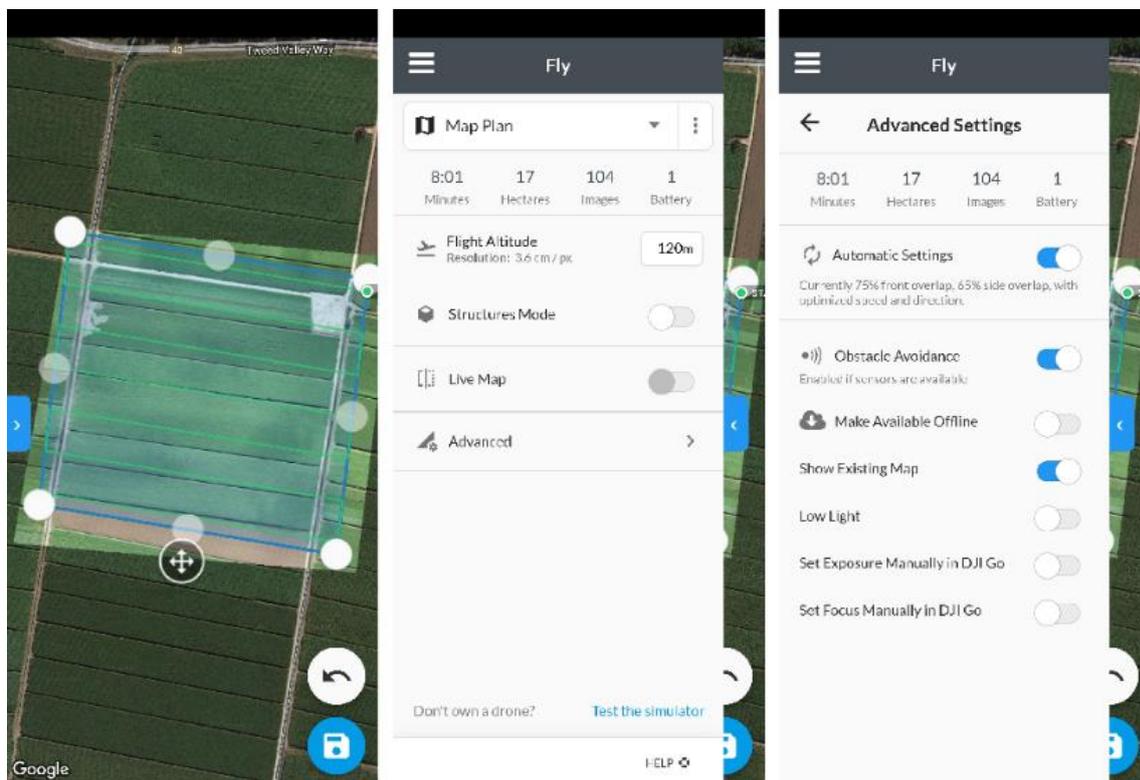
Figure 22: A series of screenshots from the DroneDeploy app

After the flight planning parameters are confirmed, the UAV and app/controller system conducts automated pre-flight checks. Finally, the UAV takes off, ascends to the correct height, and completes

the flight path, taking photos at the prescribed intervals before returning and landing in the same location.

The above procedure for UAV photogrammetry survey was executed at approximate monthly intervals, on clear sunny days with minimal wind, until the harvest of the sugarcane. Surveys were conducted on the following dates:

1. Friday, April 26, 2019 (223 days after planting, 86 days before harvest)
2. Monday, May 16, 2019 (243 days after planting, 66 days before harvest)
3. Friday, June 14, 2019 (272 days after planting, 37 days before harvest)
4. Friday, July 12, 2019 (300 days after planting, 9 days before harvest)
5. Sunday, July 21, 2019 (Day of harvest. Survey conducted immediately prior to harvest)

3.3.4 UAV Data Processing and Digital Surface Model

After the UAV survey data acquisition, back at home using PC, the images were uploaded to SfM photogrammetry software package Agisoft Metashape Professional Version 1.5.2 for processing and generation into 3D digital model.

After uploading the images, an 'align photos' function of the software is used to generate a preliminary 'sparse' point cloud, comprising several thousand points whose positions have been resolved simultaneously, along with internal and external camera calibration parameters, via the SfM bundle adjustment algorithms. However, as a whole, this point cloud's absolute position, orientation and scale are based on best-fit to the UAV's GPS coordinates as recorded in the EXIF data of the images, and are thus approximate only, within several metres.

It is necessary to manually identify the GCPs in the images for accurate co-registration of the point cloud with the survey control system. A csv (comma-separated value) file containing the coordinates of the GCPs was uploaded, and the coordinate system nominated by its EPSG code (European Petroleum Survey Group code) i.e. 28356 for GDA94/MGA Zone 56.

Each GCP is then manually marked in every image in which they are visible. The software prompts to the most likely region of the image, based on the cloud's current fix, to assist the user with finding the GCP. Initially, there is a displacement of several metres, but iterative adjustments improve the accuracy as GCPs are marked, and eventually the software's estimates are within a pixel or two. Figure 23 below shows a screenshot of this step, with GCP target visible in the aerial image.
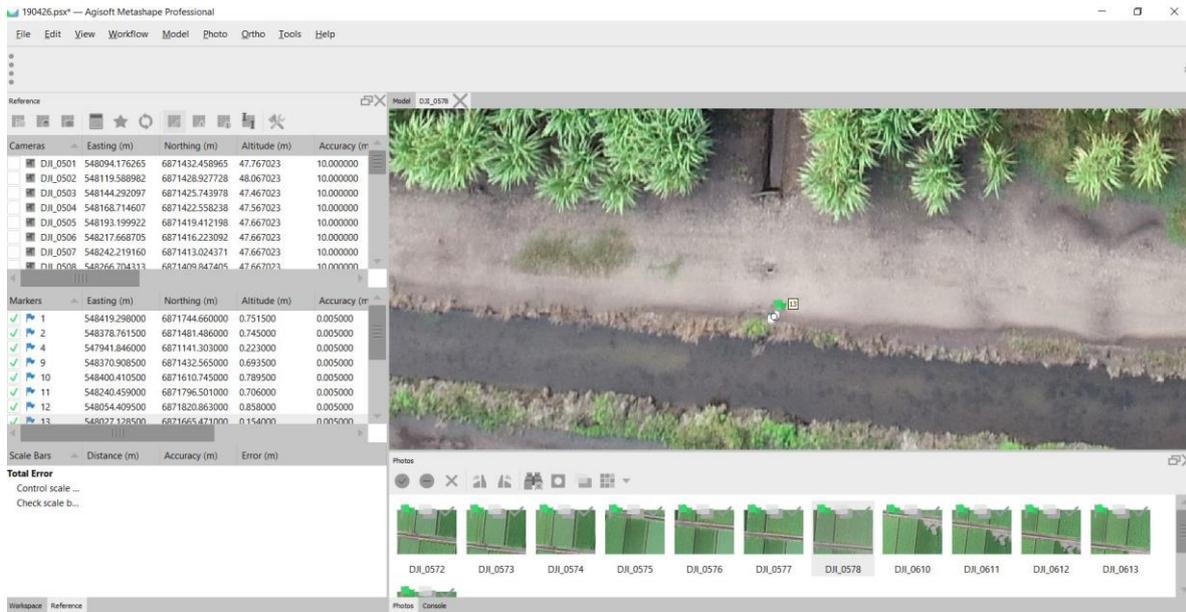
Figure 23: A screenshot from Agisoft Metashape during the procedure for manual identification of GCPs

Figure 24 below shows a typical image of a GCP target in a survey image, taken from a height of 120 m. This image demonstrates the degree of pixellation due to the camera's resolution and the digital zoom during GCP identification. The black and white squares of the target also appear blurred due to glare from the white paint. Nevertheless, it is possible to determine the centre of the target as illustrated by the red dot.

Figure 24: A close-up of a GCP target from one of the survey images, taken from a height of 120 m. Red dot illustrates user's judgement of the centre of the target.

Figure 25 below shows the completed dense point cloud and UAV camera locations. Figure 26 shows a low oblique close-up view of the point cloud surface of the sugarcane.
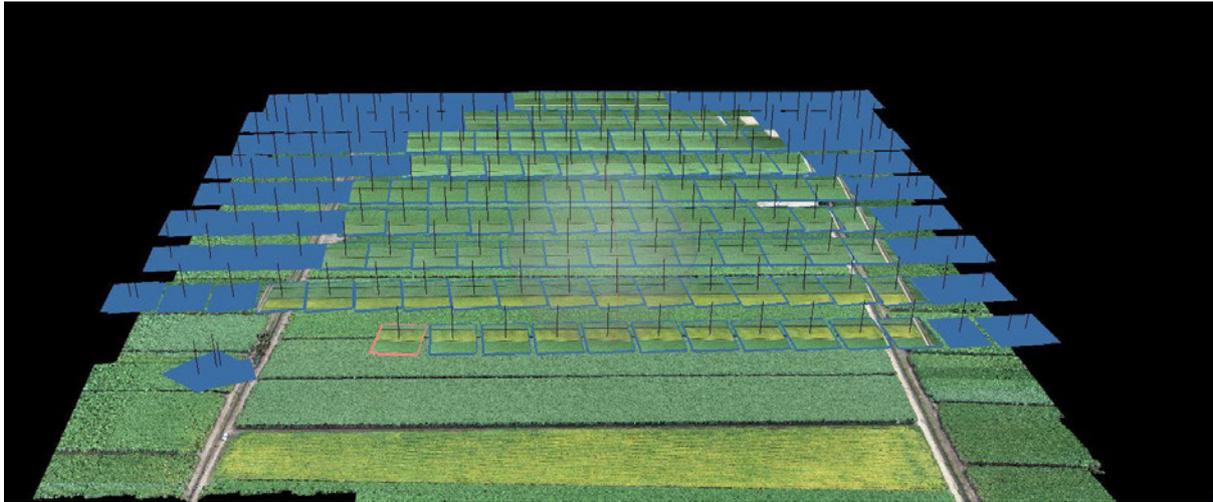
Figure 25: 3D point cloud of the study site with UAV camera locations displayed.
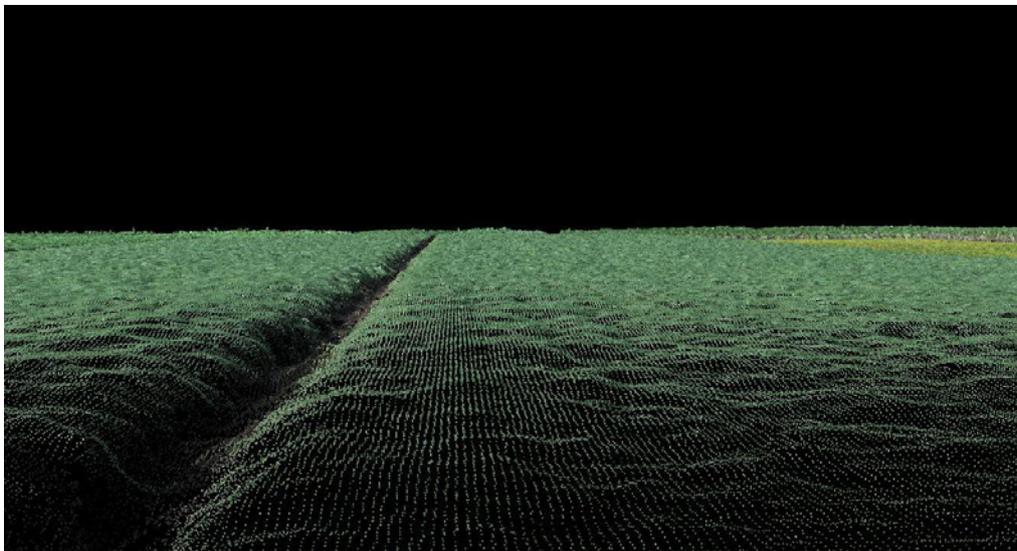
Figure 26: A low oblique close-up view of 3D Point cloud of surface of sugarcane. A drainage channel between the two subject fields is visible.

After generation of the point cloud and orthophoto in Agisoft, the following raster file types were exported for each survey epoch at a nominal Ground Sampling Distance (GSD) of 100 mm, i.e. each pixel represents a 100 mm x 100 mm square of ground:

- GeoTIFF orthomosaic - Raster image containing RGB (3 band) data of pixels at a nominated resolution, georeferenced to the project coordinate system
- GeoTIFF elevation map – Raster image containing elevation (single band) data at a nominated reolution, georeferenced to the project coordinate system.

3.4   Digital Terrain Model

The digital terrain model (DTM) is the topography of the soil surface beneath the sugarcane. It was surveyed by UAV photogrammetry survey after the cane was harvested, since, prior to that, the terrain is not visible from the air due to the height and density of the sugarcane vegetation.

For this step, the data capture and processing method was almost identical to as described above, with the only difference being that some additional ground control points were be placed toward the centre of the canefields, now that they were accessible, to improve accuracy.

3D point cloud and GeoTIFF images were generated as described in the previous step.

Figure 27: A recently harvested canefield in the Tweed Valley

3.5   Crop Surface Model

Previous studies by others (De Souza et al. 2017) computed the crop surface model (CSM) raster by subtracting the DTM raster from the DSM raster. For this study, this was performed using the inbuilt raster calculator in QGIS software. The workflow and concept are depicted in the figures below:
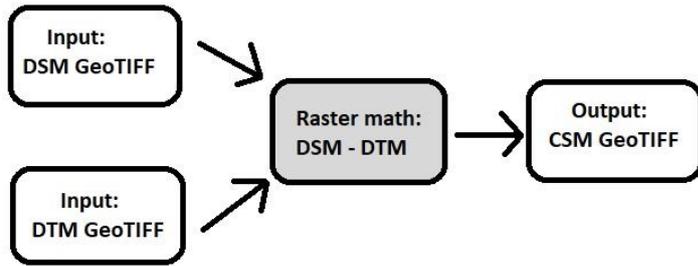
Figure 28: Workflow to compute CSM from DSM and DTM
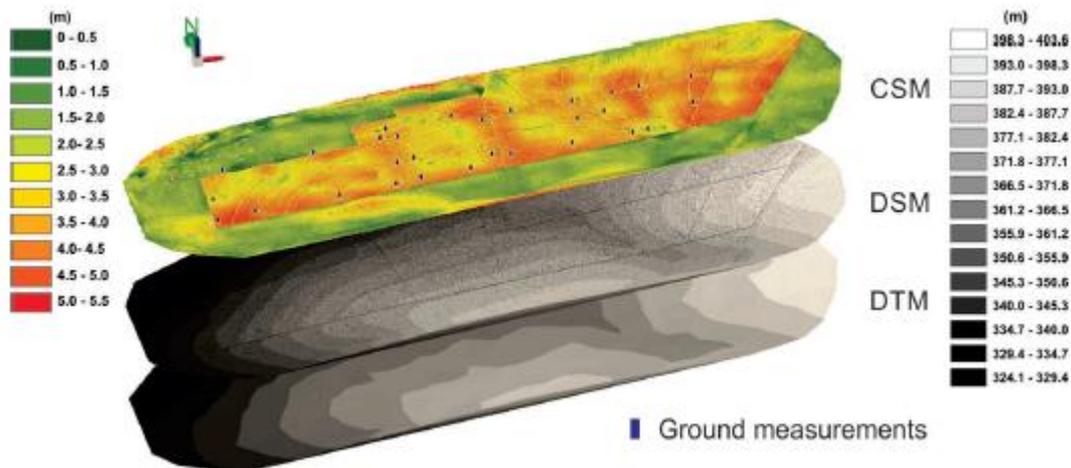
Figure 29: The CSM (shown in colour) represents crop height relative to terrain, whilst the DSM and DTM (shown in greyscale) represent elevations relative to a common datum (De Souza et al. 2017)

The output is a single band numerical raster (that can be rendered in shades of grey or false colour) in which the pixel value represents the height of the cane's surface above the ground.

3.6 Vegetation Indices

The visible band vegetation indices (VI$_{RGB}$) in Table 4 have been used by others previously (Hunt et al. 2013) to assess chlorophyll content in vegetation, which is generally indicative of plant health. The hypothesis is that chlorophyll content will correlate with yield independently from the height of the crop surface model. The software QGIS has an inbuilt Raster Calculator that was used to produce the vegetation index rasters by manually inputting the formula in terms of red, green and blue band values. These formulas were applied to the orthomosaic GeoTIFFs for each survey epoch. Output is a single-band numerical raster for each of the four VI$_{RGB}$S in Table 4 at of the five survey epochs, to give 20 VI$_{RGB}$ rasters.

Table 4: Vegetation indices from RGB images (Hunt et al. 2013)

Index Name	Abbreviation	Equation
Normalized green red difference index	NGRDI	$(R_g - R_r)/(R_g + R_r)$
Green leaf index	GLI	$(2 \cdot R_g - R_r - R_b)/(2 \cdot R_g + R_r + R_b)$
Visible atmospherically resistant index	VARI	$(R_g - R_r)/(R_g + R_r - R_b)$
Triangular greenness index	TGI	$-0.5[(\lambda_r - \lambda_b)(R_r - R_g) - (\lambda_r - \lambda_g)(R_r - R_b)]$

3.7 Multi-Temporal Series Graphs

Graphs were produced in Excel to depict the change over time of PH_{CSM} and each of the four VI_{RGBS} for all 27 harvest pin polygons. These graphs were visually assessed to determine if the change of these independent variables over time conforms with the expectations for them to correlate positively with yield.

Since yield of sugarcane should increase over time, the hypothesis is that the values for PH_{CSM} and each of the four VI_{RGBS} will also increase over time.

3.8 Empirical Yield Data

Actual yield data was gathered during harvesting, by monitoring the path of the harvester and logging its position, using a Leica GS03 GNSS RTK sensor, each time it filled up a bin.

The harvest procedure involves the harvester moving along cutting one row of cane at a time. After being cut at ground level by spinning blades beneath the harvester, the stalks enter the feed train where a series of rollers and winnowing mechanisms remove excess non-useable vegetation, before the useable stalks are ejected into a bin that is being towed alongside the harvester by a tractor. Once the bin is full, the harvester slows down while the tractor tows away the full bin and a second tractor towing a new empty bin immediately moves in to resume the process. Each bin holds about 11.8 tonnes, and an accurate weight is measured by a calibrated load cell after the bin is filled.

On the day of harvest, using the GNSS RTK sensor mounted on a range pole, a point for the starting position of the harvester was logged, and each time a bin was filled a position was logged for that full bin, and the accurate weight of the bin was advised by the operator over UHF radio. This data was used to compute polygons representing the area of field covered by each harvest bin.

It is important to note here that the canegrower (P. O'Keeffe, 2019) mentioned that, in his experience, the edge rows always contain more cane than the interior rows because they receive more sunlight and have more space to spread out.

Figure 30: The harvester that will be used to harvest the sugarcane

In this way, 27 sample measurements, i.e. full bins, were obtained for the study site (i.e. 14 bins for field 205 and 13 bins for field 206).

3.9   Separation of Samples into Datasets for Parameterisation and Validation

Of the 27 sample bins, the four bins containing the southern and northern edge rows from fields 205 and 206 were rejected as outliers because these rows contained anomalous amounts of cane and could not be considered part of the same population as the interior rows.

The remaining 23 bins were randomly separated into a parameterisation dataset (containing 15 bins, i.e. approximately two thirds of the sample) for developing the prediction models, and a validation dataset

(containing 8 bins, i.e. approximately one third of the sample) for testing and evaluating the effectiveness of the prediction models.

3.10 Simple Linear Regression Statistical Model

Linear regression analysis was computed using Microsoft Excel's data analysis tool pack, with plant height from the crop surface model (PH_{CSM}) as the independent variable and yield as the dependent variable.

The PH_{CSM} value, expressed in metres, was computed for each sample bin polygon at each epoch using the zonal statistics tool in QGIS. This value is the average of the values of all CSM raster pixels within a harvest bin polygon, which represents the average height of the cane for that region of the field.

The yield value, expressed in tonnes of cane per hectare (TCH), was computed for each sample bin polygon by dividing the tonnage of the sample bin by the area of the polygon, then scaling up to hectares.

The parameterisation dataset was used to compute a separate regression line for each survey epoch. These linear functions were then applied to the validation dataset for their epoch, and the effectiveness of these prediction models was evaluated using the statistical indicator Root Mean Square Error (RMSE) to compare residuals between computed yields and empirical yields.

A yield map raster, in which pixel value denotes TCH, of the study site (using survey data from day of harvest only) was produced by applying the linear regression formula to the CSM raster.

3.11 Multiple Linear Regression Statistical Model

The statistical technique of Multiple Linear Regression (MLR) was implemented by using Microsoft Excel's data analysis tool pack to compute yield prediction models with two independent variables; PH_{CSM} and one of the four VI_{RGB}s, and yield as the dependent variable.

The VI_{RGB} value, expressed as a number, was computed for each sample bin polygon at each survey epoch using the zonal statistics tool in QGIS. This value is the average of the values of all VI_{RGB} raster pixels within a harvest bin polygon, which represents the average VI_{RGB} value for that region of the field.

The values for PH_{CSM} and yield are as computed in the previous step.

The relationship between the independent variables themselves and with the dependent variable was checked using correlation matrices containing Pearson's correlation coefficient, *r*. Importantly, there

should not be strong correlation between the independent variables. Correlation between independent variables in MLR is known as *multicollinearity*. This means the variables are predictive of each other, and thus redundant and invalid for input into an MLR model. The effect of multicollinearity on MLR prediction is that a small change in independent variable value will produce a large change in the dependent variable, often resulting in erroneous predictions on samples other than the sample used to parameterise the model.

Yield as a function of the independent variables is expressed by the MLR equation:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip} + \epsilon$$

where, for $i = n$ observations:

y_i = dependent variable

x_i = expanatory variables

β_0 = y-intercept (constant term)

β_p = slope coefficients for each explanatory variable

ϵ = the model's error term (also known as the residuals)

The parameterisation dataset for each epoch was used to compute y-intercept and independent variable coefficients for that epoch. This was repeated for each of the four VI_{RGB}s to obtain 20 MLR functions.

These MLR functions were applied to the validation dataset for their epoch, and the effectiveness of these prediction models was evaluated using the statistical indicator Root Mean Square Error (RMSE) to compare residuals between computed yields and empirical yields, and by further examining the contribution, as a percentage, of each independent variable to the computed yield value.

Four yield map rasters of the study site (one for each VI_{RGB}, using survey data from day of harvest only) were produced by applying the MLR formulas to the CSM raster in combination with each VI_{RGB} raster.

CHAPTER 4: Results

4.1 Introduction

This chapter presents the results obtained through the methodology. The results of the harvest and the UAV survey of the bare terrain will be given first, even though they were the final field surveys, because they are necessary to compute the results of the other field surveys, which will be given in chronological order thereafter. Sections of this chapter are as follows:

4.2 Monitoring of harvest – 21 July 2019

4.3 Bare terrain UAV survey – 3 August 2019

4.4 Sugarcane UAV survey – 26 April 2019

4.5 Sugarcane UAV survey – 16 May 2019

4.6 Sugarcane UAV survey – 14 June 2019

4.7 Sugarcane UAV survey – 12 July 2019

4.8 Sugarcane UAV survey – 21 July 2019

Then, the results will be presented as a multi-temporal Series to observe change of the CSM and RGB spectral properties over time.

4.2 Monitoring of harvest – 21 July 2019

As the cane was harvested, the path of the harvester was monitored, and its position logged each time a bin was filled.

Figure 31 shows the run of the harvester to fill each bin, denoted by coloured polygons. The bins are coloured randomly for visual discernment of individual polygons.

Fortunately, the planting rows were originally set out using RTK GNSS-guided machinery and are therefore straight and of even widths. The average row width of 1.806 m was adopted by dividing the number of rows by total width, and check measurements taken during harvest confirmed that actual planting centrelines were within, at worst, about 0.15 m of best-fit alignment.
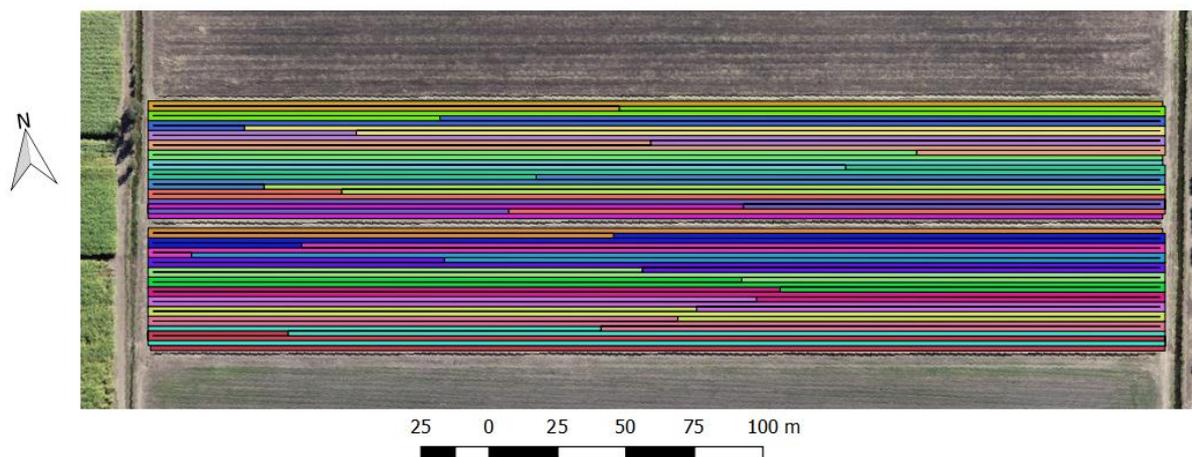
Figure 31: Harvest bin polygons

27 bins were filled in total. The harvester started at the south and worked towards the north, bins are numbered in sequential order of filling. Table 5 shows the results of the harvest.

Table 5: Results of harvest

Bin No.	Mass of cane (tonnes)	Polygon Area m²	Yield (TCH)	Canefield	Comments
1	11.97	1430.1	83.703	206	Contains southern edge row + others
2	11.97	1545.4	77.456	206	Interior rows
3	11.86	1389.9	85.332	206	"
4	11.86	1351.5	87.753	206	"
5	11.92	1378.8	86.449	206	"
6	11.92	1354.7	87.993	206	"
7	11.89	1313.9	90.495	206	"
8	11.89	1273.8	93.342	206	"
9	11.80	1208.6	97.636	206	"
10	11.80	1115.3	105.805	206	"
11	11.85	1266.6	93.561	206	"
12	11.85	1133.7	104.527	206	"
13	11.97	974.7	122.803	206	Contains northern edge row + others
14	11.87	1060.4	111.938	205	Contains southern edge row + others
15	11.79	1184.5	99.537	205	Interior rows
16	11.79	1229.2	95.916	205	"
17	11.82	1135.2	104.126	205	"
18	11.82	1159.8	101.917	205	"
19	11.80	1136.1	103.866	205	"
20	11.50	1126.9	102.050	205	"

21	11.86	1173.9	101.031	205	"
22	11.86	1164.6	101.835	205	"
23	11.88	1145.1	103.745	205	"
24	11.88	1138.5	104.344	205	"
25	11.88	1210.3	98.158	205	"
26	11.88	1221.0	97.297	205	"
27	11.50	978.5	117.529	205	Contains northern edge row + others

4.3   Bare terrain UAV survey – 3 August 2019

The final field survey of the bare terrain could not be conducted until almost 2 weeks after the harvest of the cane because immediately after the harvest there was a thick layer of 'cane trash' i.e. leaves and stalk debris covering the ground, which needed to be dried and baled up before UAV survey of the terrain was possible.

Details about the survey and the flight conditions are given in Appendix B.

Figure 32 shows an RGB orthophoto mosaic of the site. The subject fields appear lighter in colour than the adjacent fields because there is still a very fine layer of dry cane trash on the surface. Planting rows are visible as east-west striations. Drainage channels are visible as darker lines between fields.
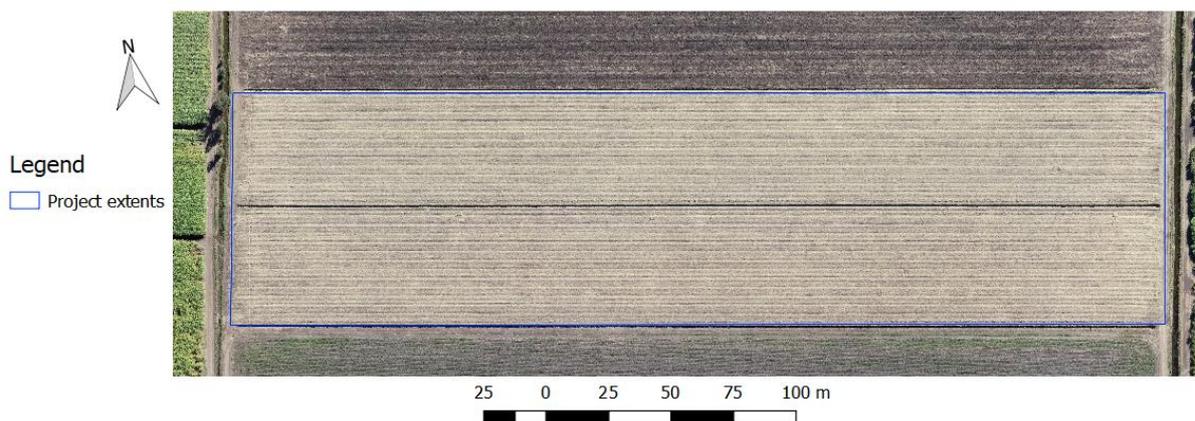
Figure 32: RGB orthophoto mosaic of site after harvest – 3 August

Figure 33 shows the DTM in raster format, where AHD elevation is denoted by pixel colour. The topography is gently sloping downwards from east to west. Drainage channels between fields are prominent. Planting rows are also still visible because of the slightly raised ridges of the rows. The site has a minimum elevation of about 0.2 AHD to a maximum of about 0.8 AHD. This range of

approximately 0.6 m highlights the necessity of the DTM, which is subtracted from the DEM of the cane surface to obtain accurate CSM.
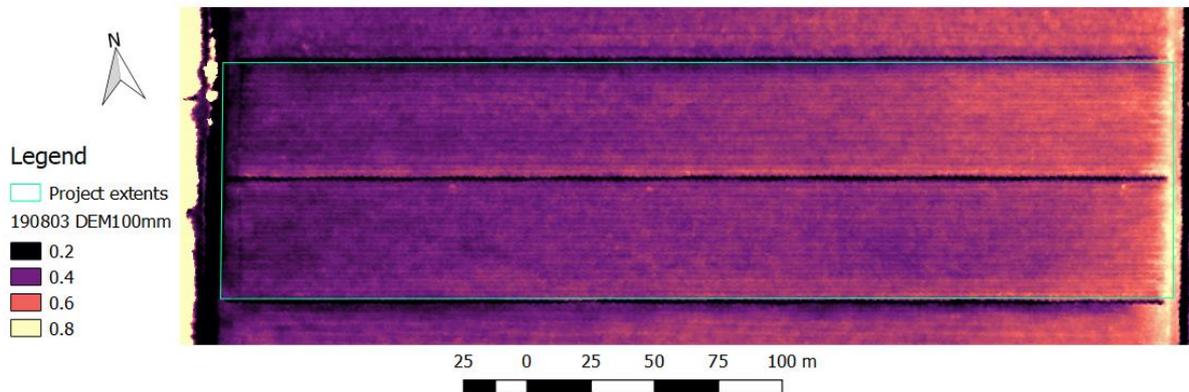
Figure 33: DTM raster of terrain after harvest – 3 August

4.4 Sugarcane UAV survey – Multi-temporal series

As described in the methodology, UAV surveys of the crop's surface were conducted at approximately monthly intervals. The images were processed into RGB and CSM GeoTIFF rasters, then, for each of the 27 harvest bin polygons, the average value for PH_{CSM} and each of the VI_{RGB}s was computed. These results are shown below.

4.4.1 RGB orthomosaic multi-temporal series

Figure 34 shows the RGB orthomosaics of the site from each survey epoch in chronological order as a multi-temporal series. In the second frame (16/05/19), cloud shadows are visible as large dark areas. In general, there is a perceptible 'yellowing' of the cane over time and, during the last two frames, numerous small dark spots appear. It should be noted that between the 3rd and 4th frame, in early June, there was period of heavy wind and rain. Details about the survey and the flight conditions are given in Appendix B.
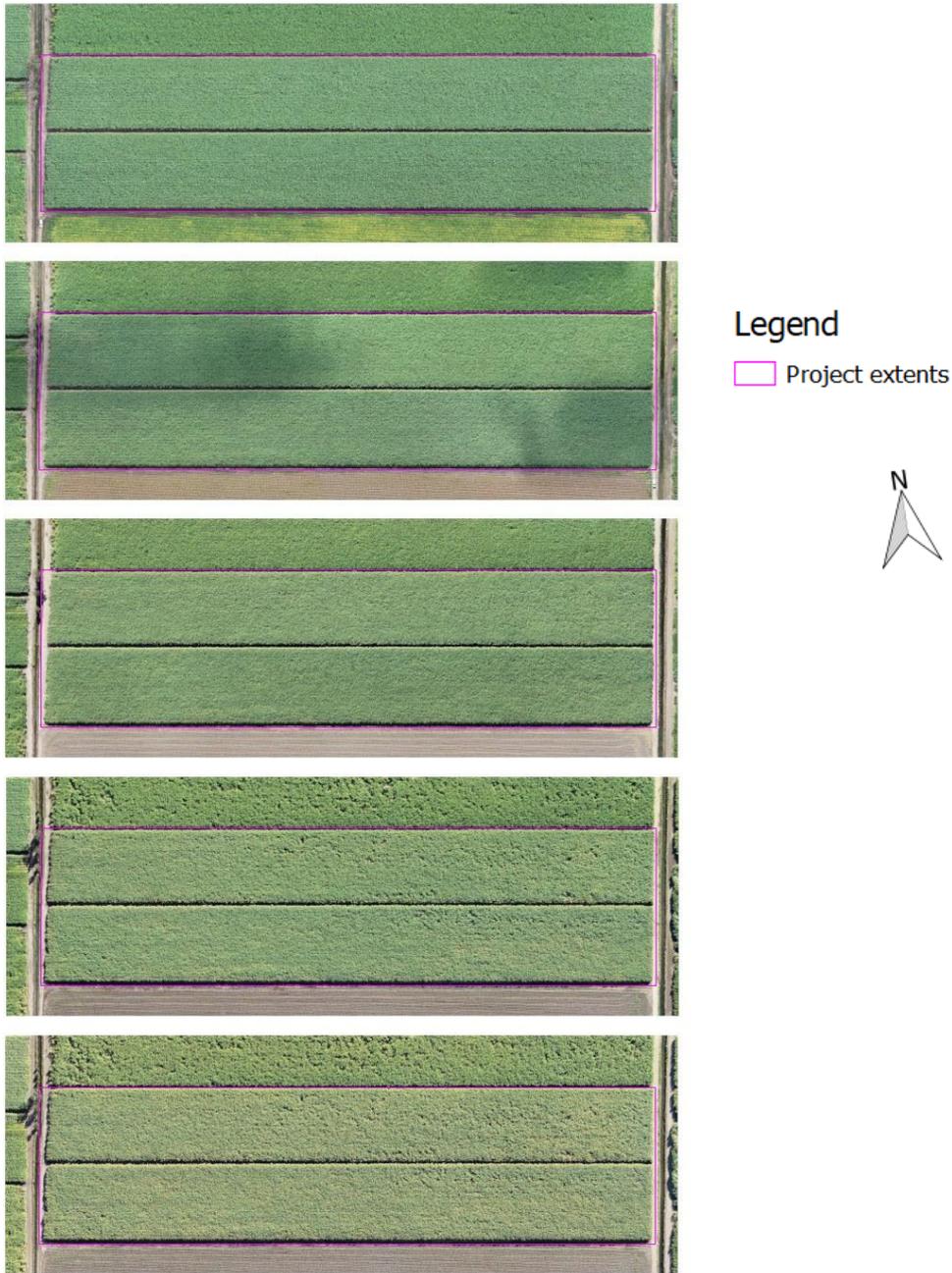
Figure 34: Multi-temporal series of the RGB orthomosaics from each survey epoch in chronological order.

4.4.2 VI$_{RGB}$ change over time

Figure 35 shows change over time for each of the four VI$_{RGBS}$. Each chart in the figure shows the 27 sample bins depicted by randomly coloured lines. There is a general downward trend, most notably for NGRDI and VARI, which corresponds with the yellowing of the cane apparent in Figure 34. The formulas for the VI$_{RGBS}$, given in Table 4 previously, express greenness, and hence chlorophyll, exaggerated in various way. This decrease in chlorophyll over time was not expected, as it was intended to be a predictor of yield, which presumably should increase over time. It should be noted that the yellowing of the cane also corresponds the onset of cold winter weather, in addition to the maturation

of the cane as it approaches harvest time. It is unclear which of these factors caused the was responsible for the yellowing, or if it was a combination of both.
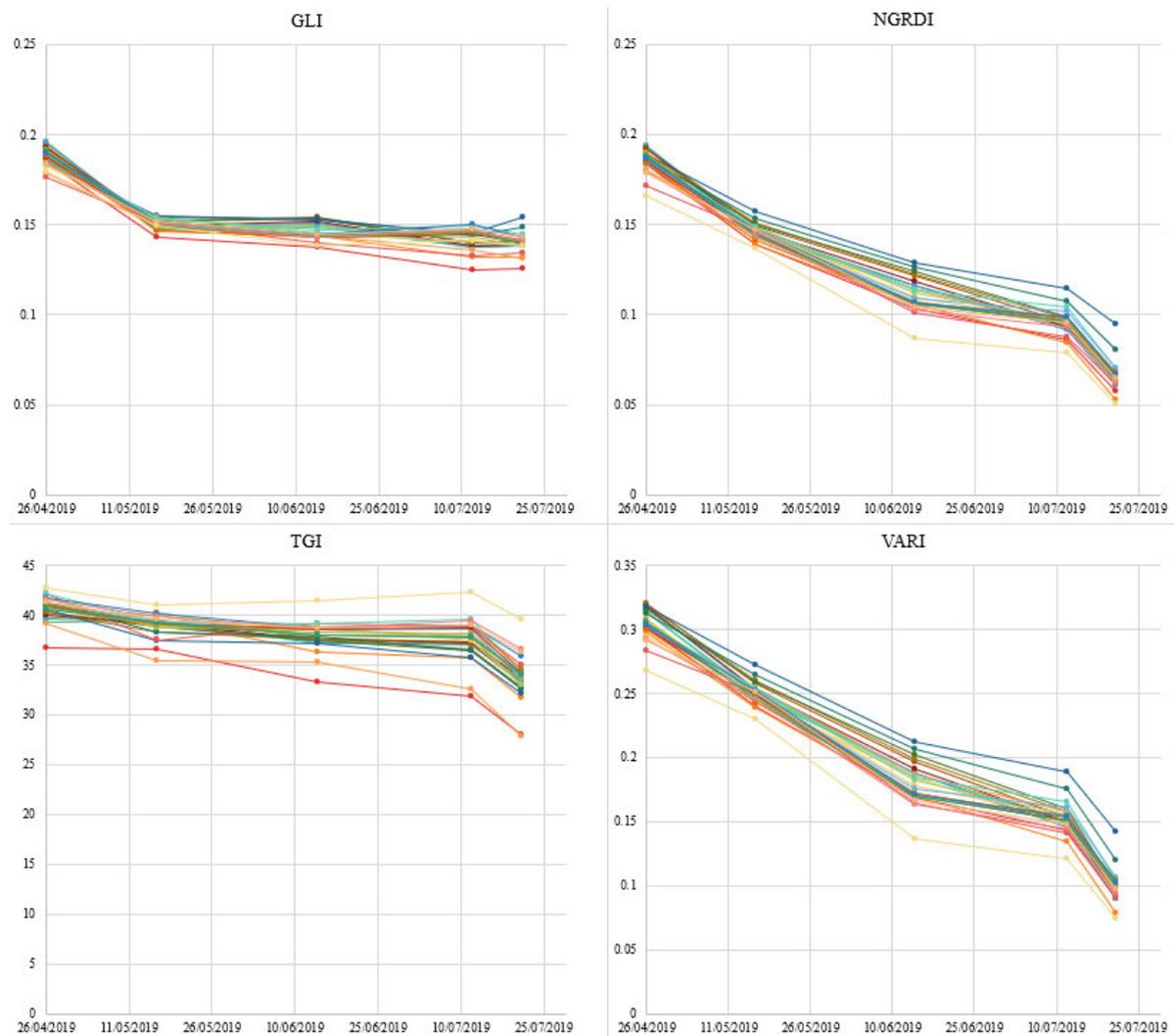
Figure 35: Change over time for each of the four VI_RGBS. Each chart in the figure shows the 27 sample bins depicted by randomly coloured lines

4.4.3   Crop surface model multi-temporal series

Figure 36 shows the CSM elevation raster of the site from each survey epoch in chronological order as a multi-temporal series. The crop is growing vertically for the first three frames, then shows a decrease in height at the fourth frame, and more vertical growth for the final frame. This decrease corresponds with the period of wind and rain in early June, and with the appearance of the dark spots in Figure 34, which are also visible in this series and are therefore sunken patches of cane.
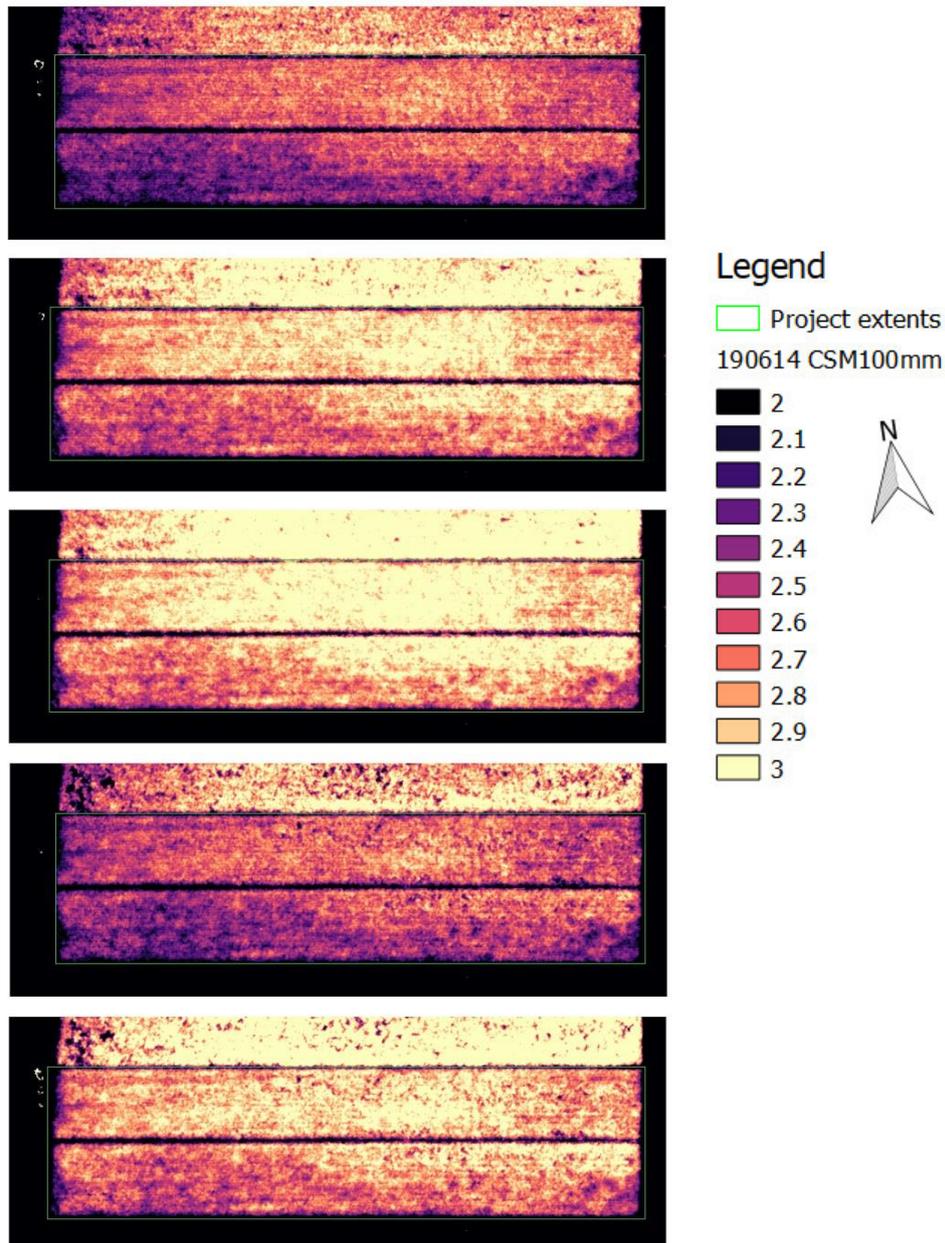
Figure 36: Multi-temporal series of the CSM elevation raster from each survey epoch in chronological order.

4.4.4   CSM plant height change over time

The fluctuation in height over time is clearly presented in Figure 37 below. Average CSM plant height for each of the 27 harvest bin polygon samples (PH$_{CSM}$) are depicted by randomly coloured lines. The height values increase, then decrease, then increase again in almost perfect lockstep. This represents their growth, their damage and 'slumping' due to wind and rain, and their partial recovery before harvest. Yield, however, should presumably grow over time, so this fluctuation is problematic for inferring growth rate as a predictor of yield.
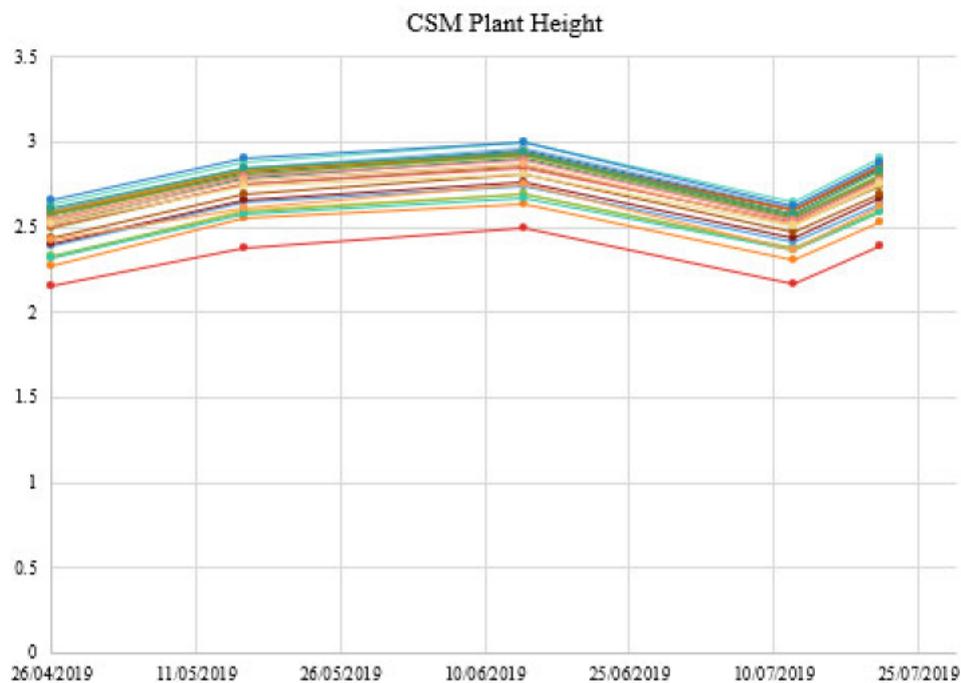
Figure 37: PH$_{CSM}$ change over time, with each of the 27 sample bins depicted by randomly coloured lines

4.5   Rejection of Outliers and Testing of Regression Function Types

Using the yield data from the harvest and the PH$_{CSM}$ from final crop survey on 21st of July, values for all 27 bins were compared on a scatter plot in Figure 38 below to observe correlation.
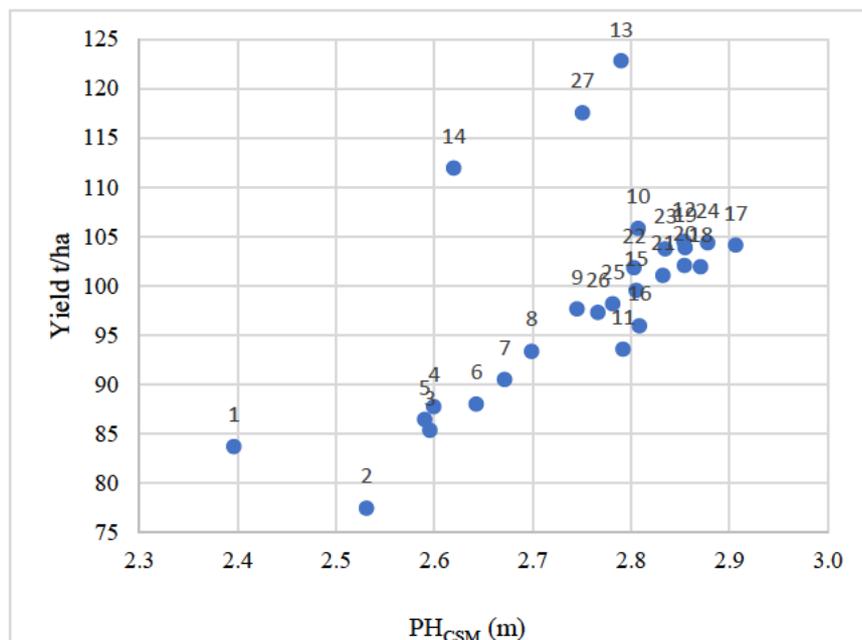
Figure 38: Scatterplot of all bins showing CSM height from 21 July and sugarcane yield from harvest results

In Figure 38 there is an obvious linear pattern between most of the sample bins except for bins 1, 13, 14 and 27, however these outlier bins fall in a linear pattern within themselves. These outliers are the bins that contain the edge rows of the sugarcane fields, as noted previously in Table 5. This supports the canegrower's anecdote that edge rows always contain more cane. These bins are hereafter excluded from the study because they are not part of the same population as the other bins.

Several different regression function types were then tested on the remaining 23 bins, shown graphically in Figure 39 below. The regression equations and the R square value for each function type are given in Table 6.

Table 6: Regression equations and R square values

Regression function type	Equation	R^2
Linear (intercept fixed at 0)	$35.061x$	0.6918
Linear	$67.601x - 90.001$	0.9009
Exponential	$12.965e^{0.7264x}$	0.9003
Logarithmic	$184.2\ln(x) - 90.285$	0.9038
Quadratic	$-74.276x^2 + 472.17x - 639.96$	0.9102
Cubic	$192.54x^3 - 1648x^2 + 4759.5x - 4527.8$	0.911
Power	$12.91x^{1.9805}$	0.9042

In Table 6, linear regression with the y-intercept fixed at zero clearly performs worse than all other functions and is not suitable for yield prediction. The other functions all have a very strong R square value of >9, with quadratic and cubic performing marginally better than the others. As noticeable by visual inspection of Figure 39, all the curved functions are almost straight. All these functions are suitable for yield prediction models within the PH_{CSM} ranges observed, but Linear has been selected for this study for its simplicity and ease of incorporation into MLR models. The form of the regression line was critically reviewed in Chapter 2 and will be further discussed in Chapter 5.
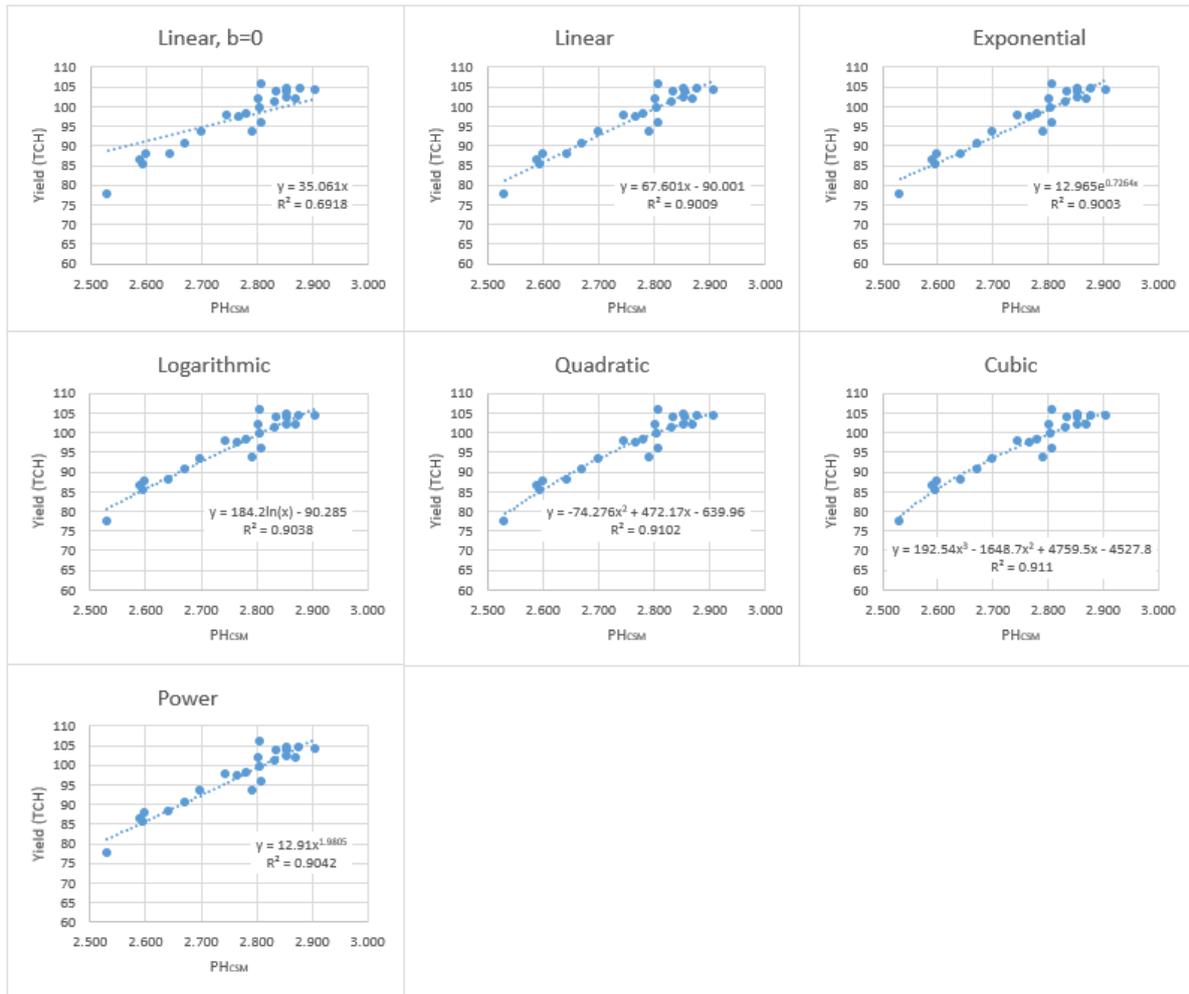
Figure 39: Testing of different regression function types using PH_{CSM} and yield from survey on 21st July (harvest day)

4.6 Selection of datasets for parameterisation and validation

The 23 sample bins were separated into two datasets; approximately two thirds of the bins (i.e. 15 samples) for parameterisation and approximately one third (i.e. 8 samples) for testing of prediction models.

The values for independent variables of all harvest bins at all survey epochs are given in Appendix C.

4.7 Regression analysis and yield prediction models

4.7.1 Preliminary analysis

Using the parameterisation dataset for each survey epoch, data was analysed for correlations between variables.

Table 7 below contains correlation matrices, produced with Microsoft Excel's data analysis tool, with input of the parameterisation dataset for each survey epoch. The tables contain Pearson's correlation coefficient, *r*, for each variable pair.

Table 7: Correlation matrices for each survey epoch, showing correlations between final yield, PH$_{CSM}$ and each of the VI$_{RGBS}$

Epoch :	26/04/19					
	Yield	*PH*	*GLI*	*NGRDI*	*TGI*	*VARI*
Yield	1					
PH	0.94	1				
GLI	-0.65	-0.74	1			
NGRDI	-0.29	-0.48	0.83	1		
TGI	0.74	0.89	-0.87	-0.76	1	
VARI	-0.04	-0.28	0.56	0.91	-0.56	1

Epoch:	16/05/19					
	Yield	*PH*	*GLI*	*NGRDI*	*TGI*	*VARI*
Yield	1					
PH	0.94	1				
GLI	0.56	0.59	1			
NGRDI	0.47	0.35	0.86	1		
TGI	-0.49	-0.47	-0.66	-0.76	1	
VARI	0.40	0.25	0.77	0.99	-0.74	1

Epoch:	14/06/19					
	Yield	*PH*	*GLI*	*NGRDI*	*TGI*	*VARI*
Yield	1					
PH	0.96	1				
GLI	-0.05	-0.19	1			
NGRDI	0.08	-0.04	0.95	1		
TGI	0.58	0.64	-0.54	-0.53	1	
VARI	0.10	-0.01	0.93	1.00	-0.52	1

Epoch:	12/07/19					
	Yield	*PH*	*GLI*	*NGRDI*	*TGI*	*VARI*
Yield	1					
PH	0.97	1				
GLI	0.78	0.83	1			
NGRDI	0.69	0.66	0.50	1		
TGI	0.35	0.42	0.70	-0.16	1	
VARI	0.59	0.55	0.33	0.98	-0.33	1

Epoch:	21/07/19					
	Yield	*PH*	*GLI*	*NGRDI*	*TGI*	*VARI*
Yield	1					

PH	0.96	1			
GLI	0.62	0.56	1		
NGRDI	0.57	0.52	0.96	1	
TGI	0.26	0.27	0.08	-0.16	1
VARI	0.62	0.58	0.95	0.99	-0.11	1

The table shows that PH$_{CSM}$ at all epochs has a very strong positive correlation with final yield. The correlation of vegetation indices GLI, NGRDI and VARI with final yield changes erratically between epochs, with each exhibiting medium correlation (R value of <-0.5 or >0.5) in at least one epoch. Importantly, none show strong consistent correlation, and hence don't exhibit multicollinearity, with PH$_{CSM}$, which suggests they may be suitable for input into an MLR model. GLI, NGRDI and VARI often show strong multicollinearity with each other, indicating they should not be put into an MLR model together. The vegetation index TGI correlates poorly or erratically with all other variables, which indicates purely coincidental, poor, or non-existent correlation.

For the sake of thoroughness, experimentation will proceed to develop MLR models using PH$_{CSM}$ with each of the VI$_{RGB}$s as described in the methodology.

4.7.2   Simple linear regression - PH$_{CSM}$ to predict yield

Table 8 shows linear regression statistics and performance on validation datasets for final yield predictions using PH$_{CSM}$ alone at each survey epoch. Validation scatter plots are also given below in Figure 41.

Table 8: Linear regression yield prediction modelling using PH$_{CSM}$

Survey Epoch	26/04/19	16/05/19	14/06/19	12/07/19	21/07/19
Days after planting	223	243	272	300	309
Days before harvest	86	66	37	9	0
Regression Statistics (Performance on Parameterisation Dataset)					
Observations	15	15	15	15	15
R	0.9377	0.9427	0.9572	0.9680	0.9555
R Square	0.8793	0.8886	0.9162	0.9370	0.9130
Standard Error	2.8646	2.7521	2.3877	2.0697	2.4324
RMSE (TCH)	2.6668	2.5621	2.2228	1.9268	2.2645
Coefficients					
Intercept	-70.2137	-101.1445	-100.1325	-109.6572	-100.0760
PH$_{CSM}$	66.7113	71.7752	68.8480	82.1021	71.4203
Performance on Validation Dataset					
Observations	8	8	8	8	8

R^2	0.9442	0.9282	0.9014	0.9052	0.9029
RMSE (TCH)	2.1841	2.5437	2.6995	2.9551	2.7382

Table 8 shows that PH_{CSM} is a very good predictor of yield, with R square values of >0.9 for all validation datasets. The RMSE values are very similar for the parameterisation and validation datasets, with the models for 26/04/19 and 16/05/19 coincidentally performing even better on the validation datasets than parameterisation datasets. These models from these earliest epochs also perform better than the models from the other epochs on the validation datasets, which may be coincidence, but is interesting because they are the most temporally distant from the final harvest, and yet they predict it better. Perhaps this is because, prior to the damage to the cane from the wind and rain in early June, which may have been spatially heterogenous, the cane height correlates more coherently with final yield.

The model for 12/07/19 performs marginally worse than the other models, despite having the strongest performance with the parameterisation dataset, however this difference is so small it is probably just a coincidence.

The intercept values range from -70.213 to -109.657, which indicates that these models are not suitable for PH_{CSM} values below the ones obtained in this study, i.e. for cane in earlier stages of growth. There are several logical reasons for this that will be discussed in the Chapter 5.

4.7.3   Multiple linear regression - PH$_{CSM}$ + VI$_{RGB}$ to predict yield

A detailed table is given in Appendix D that shows multiple linear regression statistics and performance on validation datasets for final yield predictions using PH$_{CSM}$ in combination with each VI$_{RGB}$ at each survey epoch.

To summarise these results, RMSE values for each prediction model at each survey epoch are given in Table 9 below for the parameterisation and validation datasets, and a table of residuals (RMSE validation – RMSE parameterisation) is given.

Table 9: RMSE values for all prediction models at all epochs

Table of RMSE values for parameterisation dataset

| | | *Survey Epoch* | | | | |
		26/04/19	16/05/19	14/06/19	12/07/19	21/07/19
Model	PH$_{CSM}$	2.6668	2.5621	2.2228	1.9268	2.2645
	PH$_{CSM}$ + GLI	2.6049	2.5606	1.9861	1.9057	2.1195
	PH$_{CSM}$ + NGRDI	2.2304	2.3117	2.0245	1.8652	2.1535
	PH$_{CSM}$ + TGI	2.1484	2.5252	2.1885	1.8662	2.2641
	PH$_{CSM}$ + VARI	2.0310	2.2107	2.0421	1.8637	2.1642

Table of RMSE values for validation dataset

| | | *Survey Epoch* | | | | |
		26/04/19	16/05/19	14/06/19	12/07/19	21/07/19
Model	PH$_{CSM}$	2.1841	2.5437	2.6995	2.9551	2.7382
	PH$_{CSM}$ + GLI	2.4708	2.5722	3.2470	2.8874	2.7382
	PH$_{CSM}$ + NGRDI	2.7720	3.0671	3.2078	3.2875	3.0579
	PH$_{CSM}$ + TGI	2.3024	2.4826	2.7092	2.9825	2.7469
	PH$_{CSM}$ + VARI	2.8187	3.1735	3.1856	3.2692	3.0119

Table of residuals (RMSE validation – RMSE parameterisation)

| | | *Survey Epoch* | | | | |
		26/04/19	16/05/19	14/06/19	12/07/19	21/07/19
Model	PH$_{CSM}$	-0.4827	-0.0184	0.4766	1.0283	0.4737
	PH$_{CSM}$ + GLI	-0.1342	0.0115	1.2608	0.9817	0.6187
	PH$_{CSM}$ + NGRDI	0.5416	0.7554	1.1833	1.4223	0.9044
	PH$_{CSM}$ + TGI	0.1541	-0.0426	0.5207	1.1162	0.4828
	PH$_{CSM}$ + VARI	0.7877	0.9628	1.1435	1.4056	0.8477

In Table 9, RMSE validation – RMSE parameterisation values are useful for assessing the performance of the model for making predictions; the lower the value the better. Higher residuals may indicate a degree of overfitting in the parameterisation of these models, because they perform much worse on the validation dataset.

In general, PH_{CSM} alone performs better than the MLR models, with two notable anomalies: PH_{CSM} + TGI is the best performer on 16/05/19 and PH_{CSM} + GLI is the best performer on 12/07/19.

These RMSE residuals and how they change over time is depicted graphically in Figure 40 below. Interestingly, the models tend to perform better the more temporally distant from the final harvest. This should not be interpreted to mean distant yield forecasts are more accurate than close ones, but rather, because, in a kind of 'temporal overfitting', the models have been parameterised using known final yields, and there is a more coherent correlation between dependant and independent variables in the earlier growth before, one suspects, the cane is damaged by wind and rain.
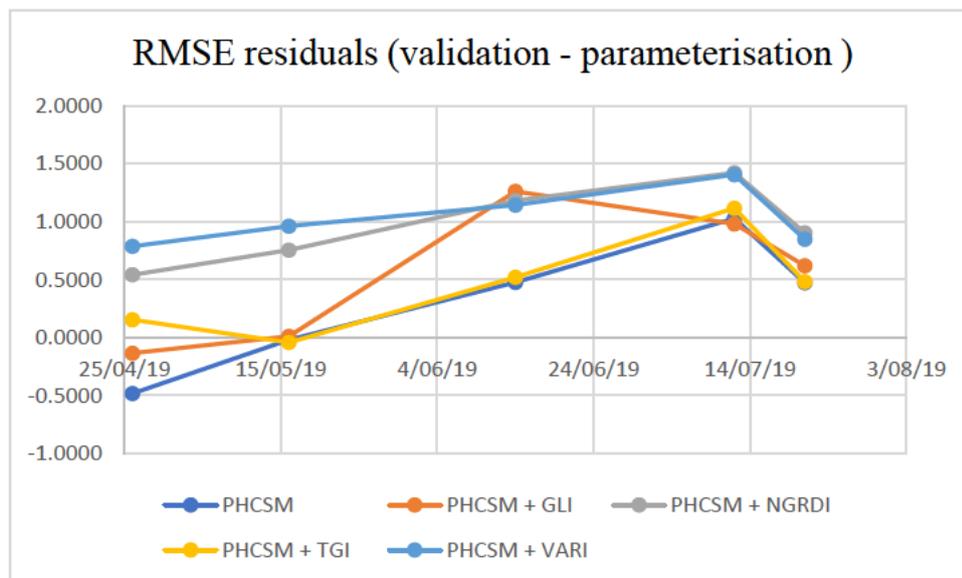
Figure 40: RMSE residual change over time

4.7.4   Validation scatter plot summary

Figure 41 below contains validation scatterplots for all models at all epochs, with predicted yield on the x-axis and final yield on the y-axis
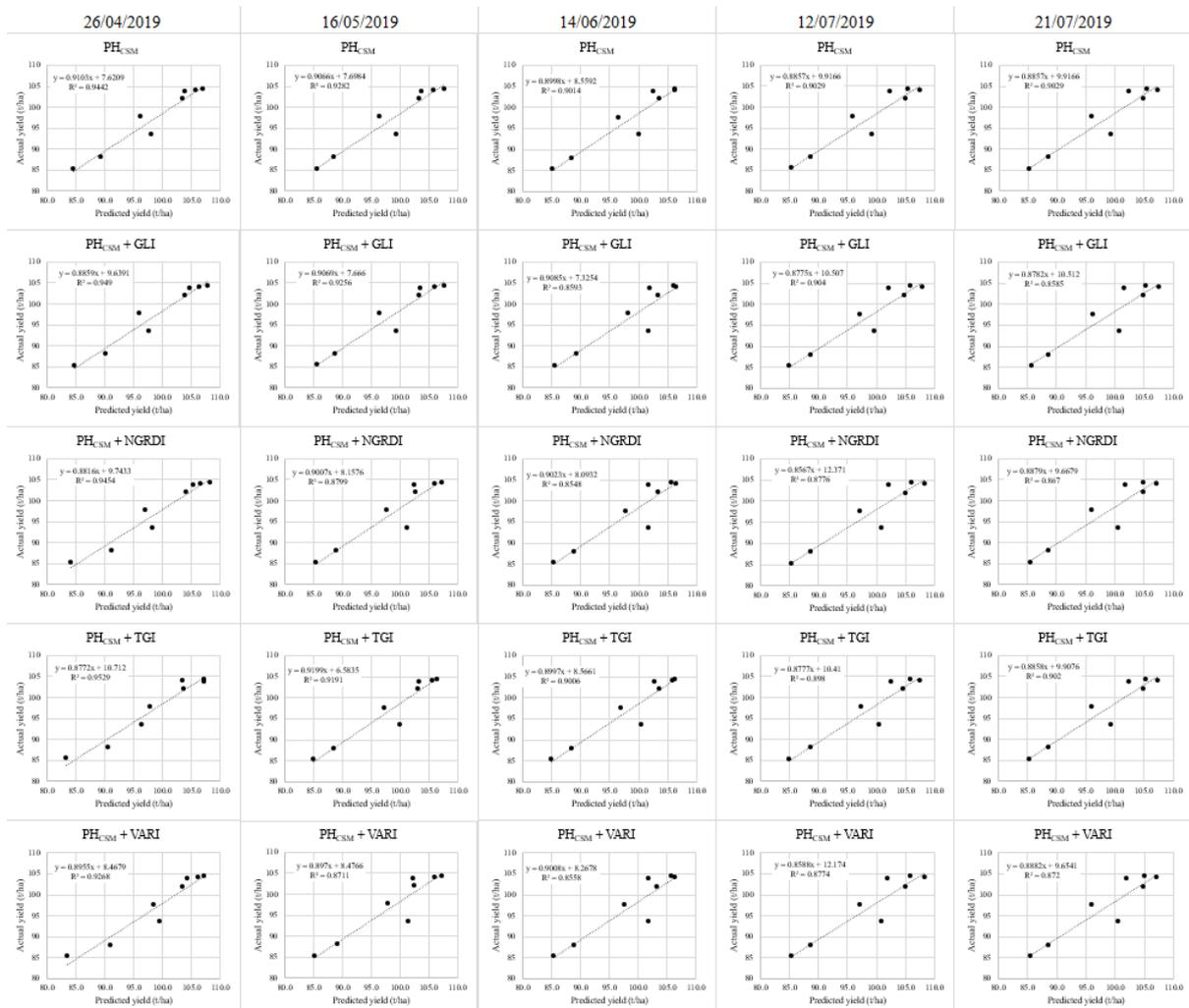
Figure 41: Validation scatterplots for all models at all epochs, with predicted yield on the x-axis and final yield on the y-axis

4.8   3D MLR surfaces and Yield maps

In the figures of the subsections below, as a demonstration using only the data from the final survey epoch on 21/07/19, the linear regression functions are depicted as 3D surface graphs with the two independent variables along the x and y axes, and the dependent variable (yield) on the z axis. Raster yield maps have been computed by applying the regression formulas to the CSM and VI_{RGB} raster GeoTIFFs.

4.8.1   Linear regression PH_{CSM} only

Figure 42 has been computed using linear regression function developed previously in Section 4.7.2
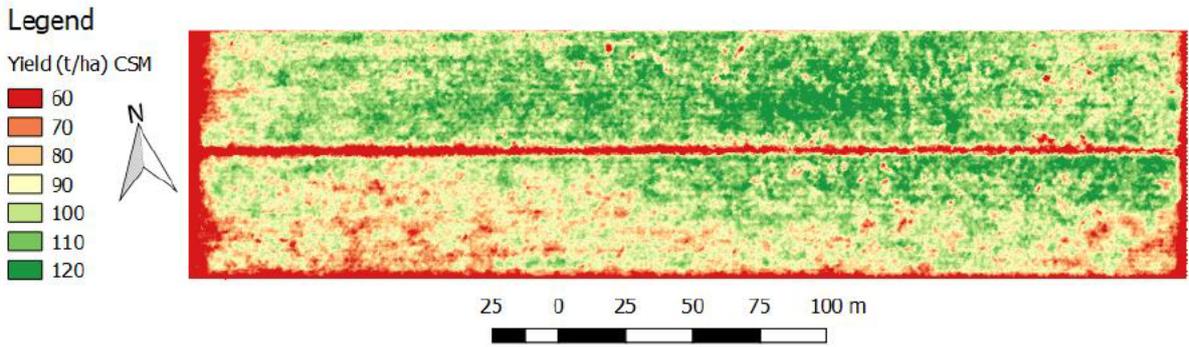
Figure 42: Yield map using PH$_{CSM}$ linear regression model

4.8.2 MLR PH$_{CSM}$ + GLI

Figure 43 below shows MLR 3D surface for this prediction model and a yield map computed using this function.
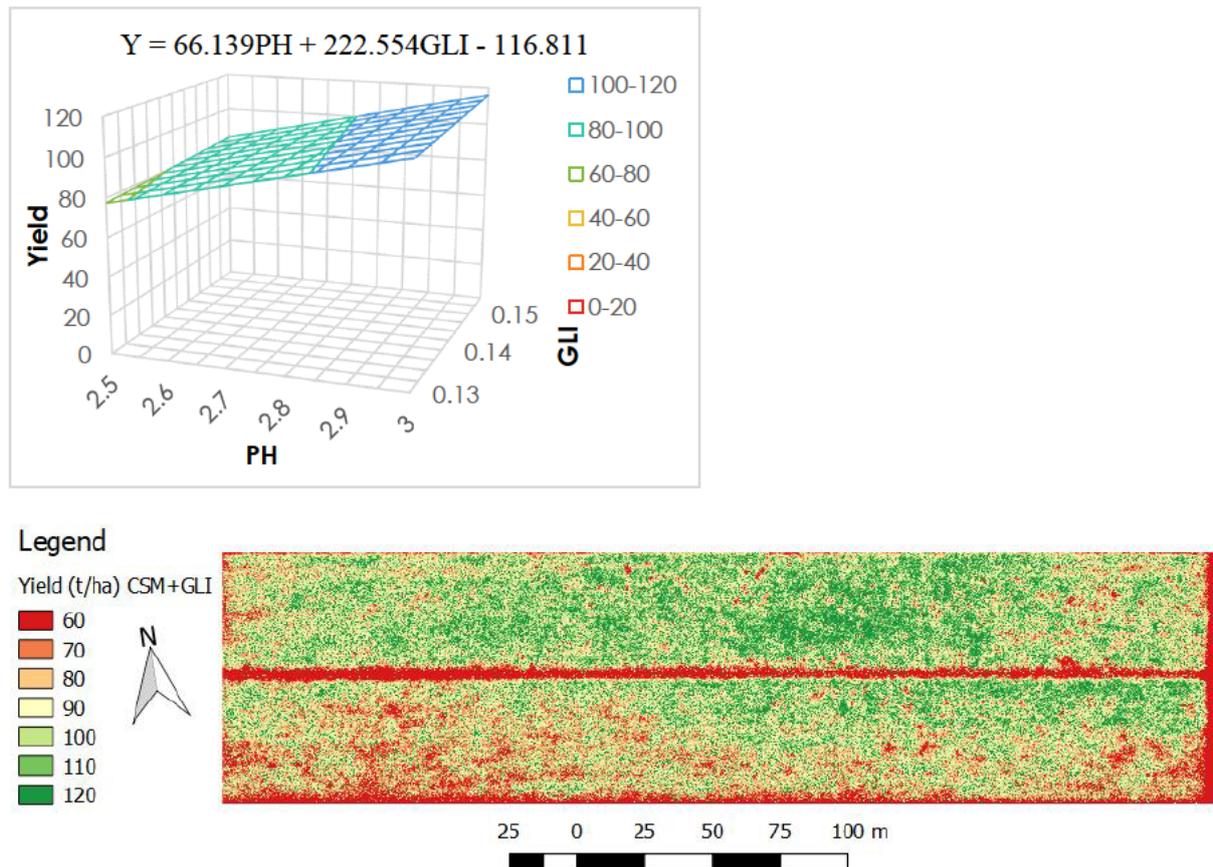
Figure 43: MLR surface of PH + GLI and yield map using this function

4.8.3   MLR PH$_{CSM}$ + NGRDI

Figure 44 below shows MLR 3D surface for this prediction model and a yield map computed using this function.
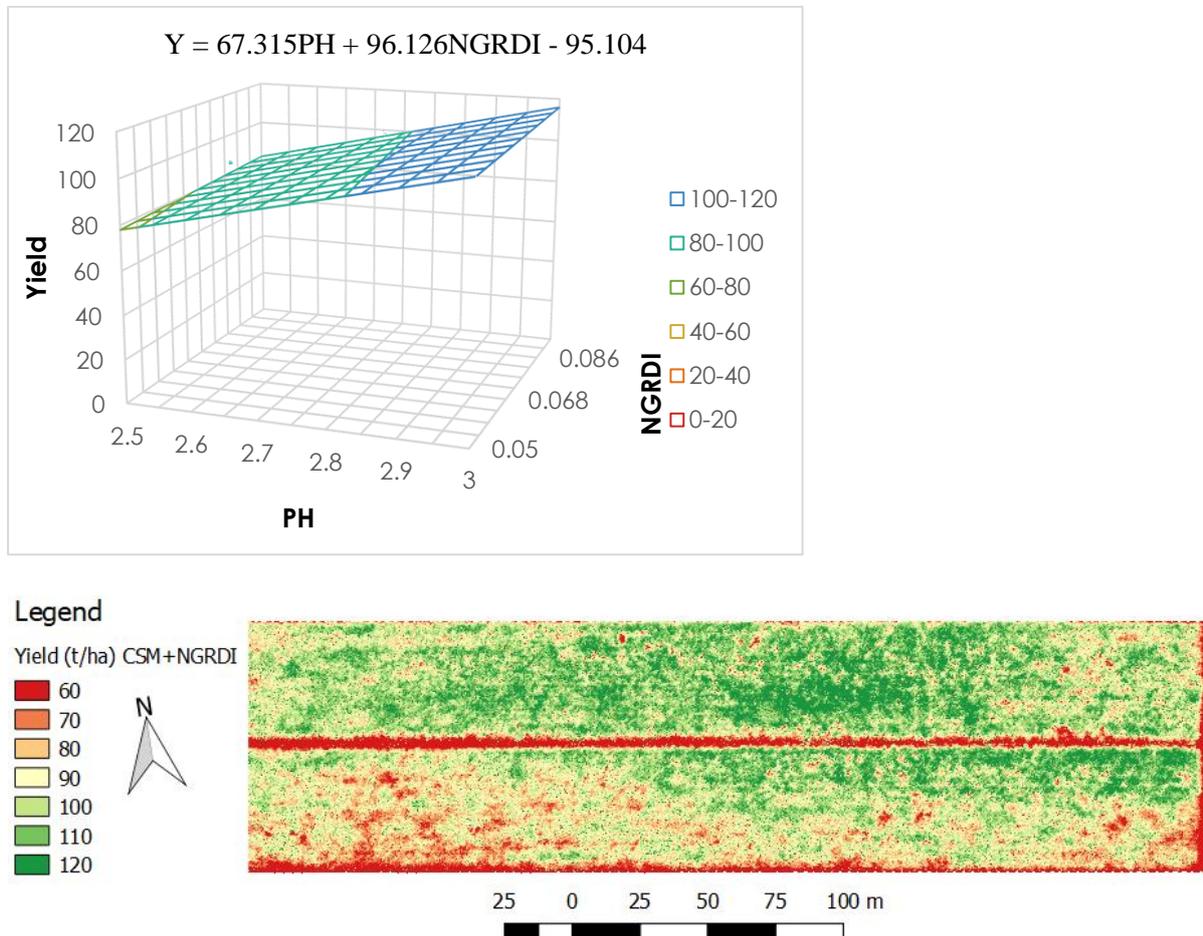
Figure 44: MLR surface of PH + NGRDI and yield map using this function

4.8.4   MLR PH$_{CSM}$ + TGI

Figure 45 below shows MLR 3D surface for this prediction model and a yield map computed using this function.

$$Y = 71.540PH - 0.035TGI - 99.242$$
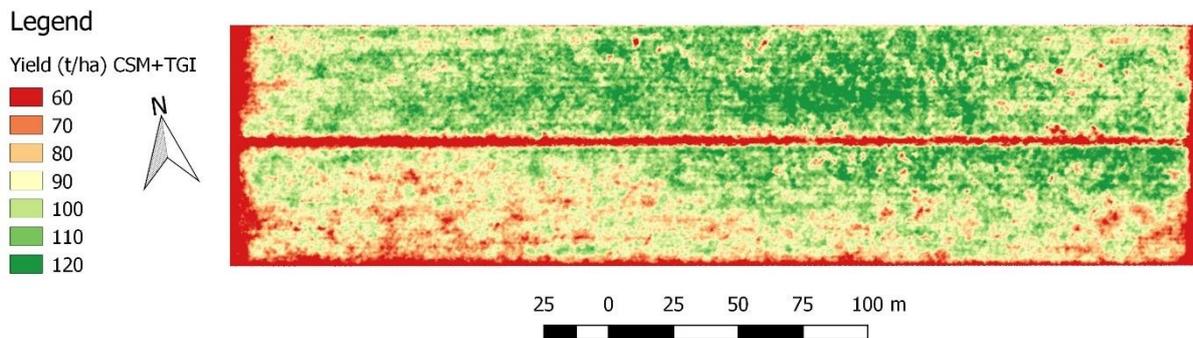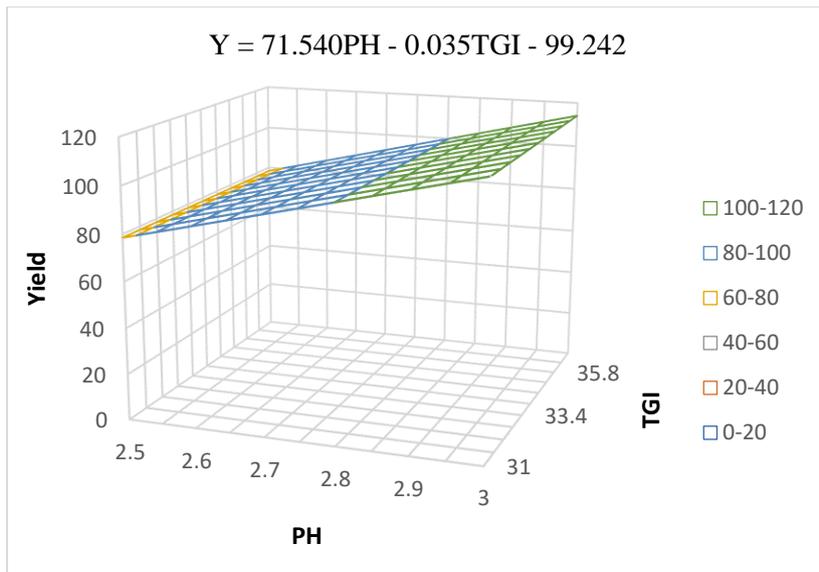
Figure 45: MLR surface of PH + NGRDI and yield map using this function

4.8.5   MLR PH$_{CSM}$ + VARI

Figure 46 below shows MLR 3D surface for this prediction model and a yield map computed using this function.

$$Y = 66.828PH + 63.505VARI - 93.729$$
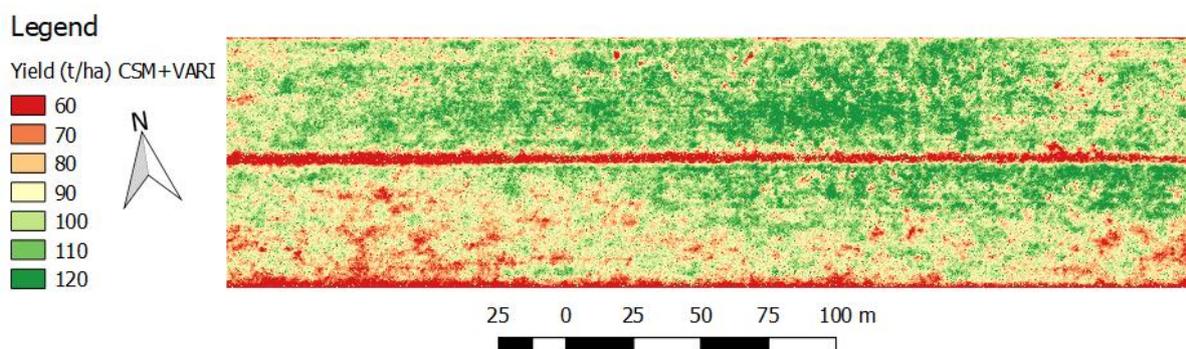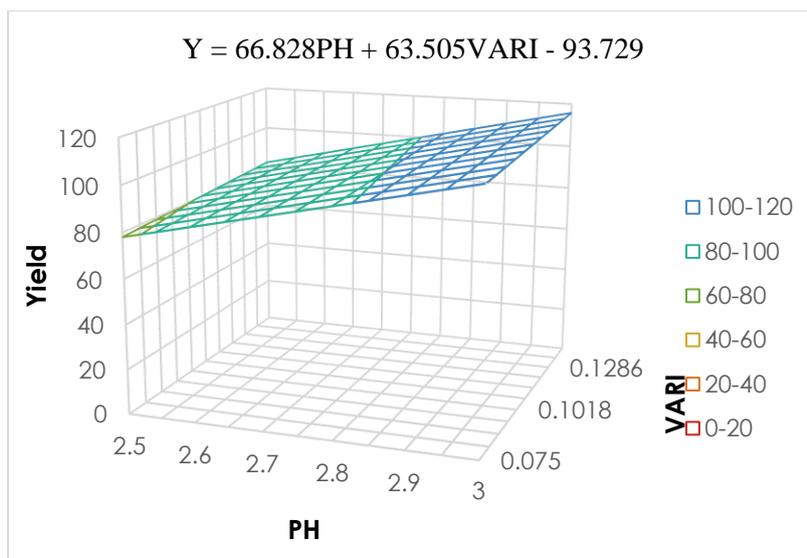
Figure 46: MLR surface of PH + VARI and yield map using this function

4.9   Contribution to solution by vegetation indices

All yield maps in Section 4.8 above show, in general, a similar spatial distribution of yield to the first map that uses PH_{CSM} alone. The MLR maps seem only different in that they are grainier, or 'noisier'; the variations in the spatial distribution of yield that they introduce into the model appear as tiny 'flecks', rather than over plant or row-sized areas that one would expect yield variations to occur.

Figure 47 below shows a close-up view of the PH_{CSM}+GLI yield map to demonstrate this. There are tiny visible green (high-yielding) flecks on the red (low-yielding) background, and vice versa. This will be discussed further in the next chapter.

Figure 47 also shows an indicative percentage (based on 21/07/19 data) of how much the VI_{RGB} contributes to the solution for the MLR models. This is to illustrate that most of the result is due to PH_{CSM}, an effect of its higher statistical correlation with yield.
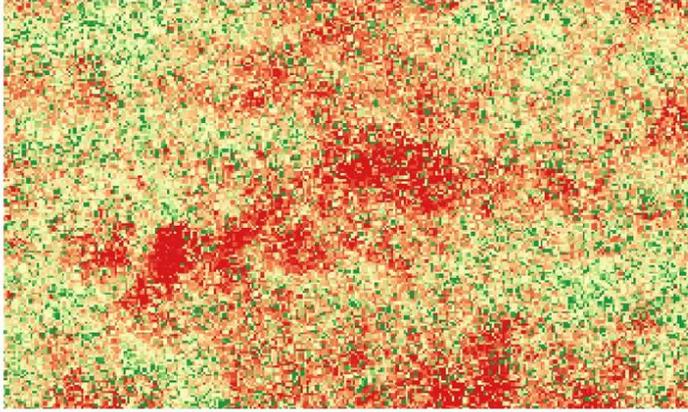
MLR Model	How much the Vegetation Index contributes to the result
PH+TGI	0.57%
PH+VARI	3.10%
PH+NGRDI	3.11%
PH+GLI	14.25%

Figure 47: A close-up of an MLR model showing noise introduced by vegetation index (example shown is GLI), and percent of the solution that vegetation indices contribute.

CHAPTER 5: Discussion

5.1 Introduction

The prediction models developed in this project are mathematical functions that express sugarcane yield (tonnes of cane per hectare), in terms of crop surface model height and vegetation indices at various points throughout the duration of its growth. Essentially, properties of the sugarcane that are difficult or impracticable to measure directly are being inferred from other properties that have been measured with relative ease by UAV photogrammetry survey.

5.2 What really constitutes 'yield'?

The quantity that is really being estimated here is the mass of useable sugarcane stalks (Y) within an area. To break it down further, Y is the product of the average thickness (radius) of the stalks (r_S), the average length of the useable portion of the stalks (l_S), the total number of stalks (n_S), and the average density (mass/volume) of the stalks (d_S), at any given point in time.

Figure 48: Depiction of the physical variables of cane comprising yield

Mathematically, this could be expressed by the equation:

$$Y = \pi r_S{}^2 \times l_S \times n_S \times d_S \tag{1}$$

To forecast Y for a future point in time though would require the application of some growth function for each of these variables. Such growth functions are difficult to establish, however, because, even if every stalk in an entire canefield was painstakingly measured throughout the duration of its growth, these observed rates of change are not likely to be replicated exactly in any other canefield, or indeed that same canefield in the next season, because they are influenced by a myriad of biotic (cane variety, soil microbes, weeds, insects, etc.) and abiotic (nutrients, rainfall, temperature, etc.) factors.
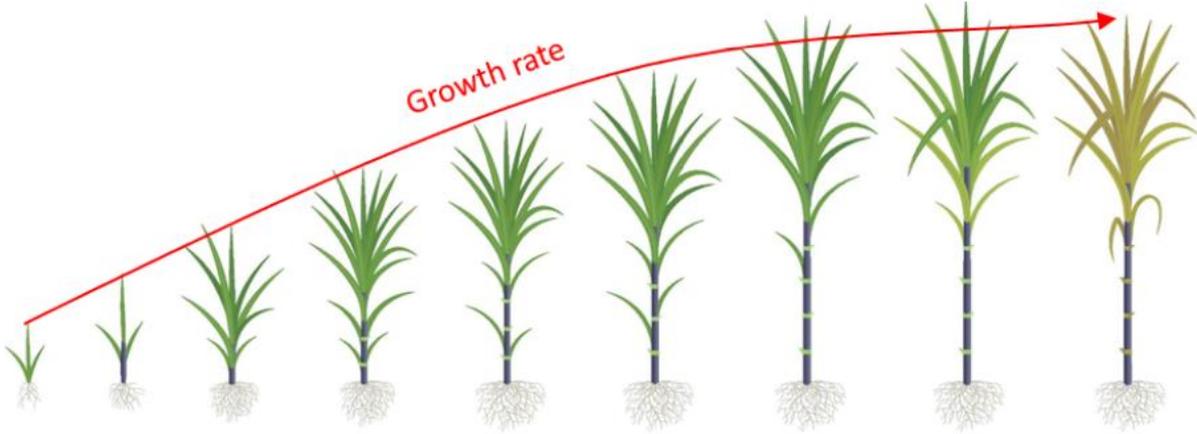
Figure 49: Depiction of growth of cane over time

Suffice it to say, Y tends to increase over time. Therefore, for forecasting of a future Y, the above equation (1) requires a factor for growth due to time (g_T). Thus, generically:

$$Y_T = \pi r_S^2 \times l_S \times n_S \times d_S \times g_T \qquad (2)$$

In this study, none of the properties of cane comprising the above equation are measured directly. However, the plant height from the crop surface model (PH_{CSM}) is a good proxy for the length of the stalks, especially if the cane is standing fairly upright as it tends to do, and requires only corrections by a coefficient, f, to account for the degree to which it is leaning, bent or bowed, or any other such errors proportional to its length, and a constant, k, to account for the unusable portion of the top of the plant and any consistent difference between the digital surface and the 'true' surface of the top of the useable portion of the stalks, which will be discussed further below). Thus, equation (2) becomes:

$$Y_T = \pi r_S^2 \times n_S \times d_S \times g_T \times f PH_{CSM} + k \qquad (3)$$

Equation (3) permits all the variables that are not being directly observed in this study to be multiplied together into a single unknown coefficient, m, and all errors requiring correction by a constant to be added together into a single unknown, b, then equation (3) takes the linear form:

$$Y = mPH_{CSM} + b \qquad (4)$$

5.3 Linear regression

In this section, linear regression of PH_{CSM} against yield is discussed. Multiple linear regression is omitted and will be discussed in the next section.

In this study, m and b in equation (4) are deduced by linear regression of PH_{CSM} against empirical Y values of harvest bins. 15 of the 27 harvest bin samples are used to parameterise the prediction models and the remaining 8 of 27 bins is used to evaluate their effectiveness.

Linear regression, as distinct from other curvilinear regressions that use exponential, logarithmic, polynomial, etc. curves, was used in this study not only for its ease or simplicity, but for its suitability for the dataset as depicted in Figure 39. Tests were conducted with the other curve types and they did not significantly improve the fit i.e. they did not reduce the r-square value, and the polynomials that slightly improved the fit were downturned curves.

This is a point of curiosity that requires some discussion. For instance, as an analogy, if one were to use regression analysis to predict the volume of solids that were roughly cubes by measuring their height only, and aimed to achieve this by establishing correlation between the height and volume of known samples on a scatter plot, then the function of best fit would not be a line, it would be a 3rd order polynomial function that is upturned. The same should also apply for cylinders of an approximately fixed radius to length ratio, as one might expect for sugarcane stalks.

Figure 50: A sugarcane stalk collected after harvest, with thickness of 40 mm (banana for scale)

However, that fact that yield increases approximately linearly with height suggests that, in the subject fields at any given epoch, whilst the average lengths of stalks varies between regions of the field (for example, PH$_{CSM}$ ranges from 2.4 to 2.9 m in the final epoch), the average thickness does not vary. Either that, or there are some compensating variations in the other variables between tall and short cane, for example if short cane is indeed thinner but is denser. This seems less likely, but is not impossible if, for instance, the outer skin had a constant thickness for all lengths of stalk and was denser than the inner pith; the skin to pith ratio would be greater for shorter, thinner stalks.

If it is true that at any point in time the average lengths of stalks varies between regions of the field but the average thickness does not, then it must also be true that over time they increase in both length and thickness, since young and immature cane stalks are not always 40 mm thick as seen in Figure 50. This is an obvious statement, but the implication is that the multi-temporal series depicting change in PH$_{CSM}$ over time cannot be used to infer change in yield over time, because the relationship between height

and yield must vary over time, as average thickness increases, and no yield samples were obtained for any time other than at the final harvest.

To reiterate and emphasise, the linear regression in this study describes only the relationship between PH_{CSM} *at any given epoch* (all epochs exhibit a similar linear pattern), and final yield *on the day of harvest*. It does not describe the relationship between height and yield at any other time. It is tempting, but wrong, to infer, for example, that when an area of the crop that was 3 m high at harvest was only 2.5 m high at some earlier point in time, it had the same yield as the final yield for another area of the crop that was 2.5 m high at harvest. There is insufficient data, and the shape of the regression line implies that cane thickness at any time other than harvest cannot be inferred from PH_{CSM} in this study.

5.3.1 The y-intercept

The y-intercept of the linear and multilinear functions derived in this study fall within the range of around -70 to -100. This means the prediction models are not suitable for application to cane at earlier stages of growth than those used to develop the models. Indeed, for crop heights of less than around 1.4 m, these models would compute a negative yield value, which is obviously nonsensical. However, there are valid reasons for computed yield to be zero for a non-zero value of a predictor variable.

Firstly, it does indeed make sense that there is some point in the cane's growth before which the yield would be zero or virtually zero. For instance, when the cane is very young, as with all members of the Poaceae family, which are monocotyledons, it just a single leaf with no segmented stalk whatsoever. It is only as the plant grows, with newer leaves unfurling from the top as lower, older leaves wither and fall off, that this segmented stalk comes into being. Furthermore, it is not until the plant reaches a certain level of maturity that its stalk synthesises and accumulates sucrose in concentrations worth extracting commercially. The exact height that this occurs at is unknown and no doubt differs between varieties and with variations in other biotic and abiotic factors, but it not inconceivable that it would be somewhere close to the 1.4 metres suggested by these prediction models.

Secondly, the digital model of the crop's surface is not the true top of the useable portion of the sugarcane stalk. It is a pseudo-surface, whether it be above or below this nominal line, that is a suitable representative by virtue of the fact that it shows strong correlation with empirical yield.

5.4 Multiple linear regression

In multiple linear regression, additional independent variables are sought that account for residuals in single linear regression resulting in an improved fit. All independent variables should also be independent from the each other, and thus produce a continuum of values along an axis perpendicular to both the first independent variable and the dependent variable. Hence a multiple linear regression prediction model can be best visualised as a 3-dimensional surface, as in the figures in Section 4.8.

The first problem with introducing the VI_{RGB} variables in this study is that they showed, over time, that chlorophyll was decreasing. This does not support the hypothesis that chlorophyll correlates with yield, at least in the mature cane of this study. Logically, this alone is enough to reject these VI_{RGB} variables, but they were persevered with for the sake of thoroughness and in case they showed negative correlation with yield, or in some other way correlated with the residuals of yield in linear regression with PH_{CSM}.

The second problem with the VI_{RGB} variables is that they introduced only noisy flecks to yield maps; erratic peaks and troughs of yield that were most likely a result of reflections and light glinting off the leaves of the cane, darks shadows, saturated pixels or other optical effects that caused them to be a poor proxy of reflectance. The yield maps are a good tool for visually assessing the effect that added variables contribute.

The third problem is that the index that performed the best in making predictions in Table 9, i.e. GLI, was still worse than PH alone and also produced the noisiest-looking yield map, purely because, as seen in Figure 47, the contribution of the other VI_{RGBS} was so small it had negligible effect on the appearance of the map.

5.5   Growth rate and damaged cane

As seen in Figure 37, in the first three epochs of April-May-June, the height of the cane is increasing at a decreasing rate, i.e. its growth is decelerating. This is perhaps to be expected for a crop approaching maturity, especially as the seasons transition from the warmer, wetter autumn months to the cooler winter months. Then, suddenly, in a manner uncharacteristic of plant growth, by the fourth epoch (12th July), the height of the cane apparently decreases. In fact, the cane stalks did not shrink or decrease in length, but rather they were caused to lean over or slump by two weeks of very windy and rainy weather conditions, before partially recovering by the final epoch on 21st of July.

Indeed, many other canefields in the region sustained even more severe damage than the subject fields during this weather event, some to the point where the cane was lying flat. This is clearly visible from Google Earth aerial imagery (Figure 51 below), which coincidentally was recently updated to show imagery captured on the 11th of July i.e. within one day of the survey epoch of this study (12th July) in which the slump in cane height was observed.

Figure 51: Examples of canefields severely affected by wind. Google Earth (2019) – Imagery date 11/07/2019

Upon enquiry about this phenomenon of sugarcane falling over, a representative from Sunshine Sugar (M. Warren, 2019) advised that it is very common, especially for sugarcane in the advanced stages of growth, i.e. when it is taller, and happens every year to a more or lesser degree.

It is fortunate that the subject fields sustained only comparatively minor damage. It is also fortunate that the subject fields did sustain some damage, because it revealed a major weakness in the in the PH_{CSM} regression model that would otherwise seem to be such a strong predictor of yield. It revealed, which is not apparent when looking at any of the survey epochs in isolation, but only when the multi-temporal series is viewed, that correlation, however strong, cannot necessarily be used to make absolute predictions if there is some systematic error that the measurement method fails to detect.

CHAPTER 6: Conclusions

This project was conducted with the aim of estimating sugarcane yield using imagery captured by visible spectrum, i.e. red green blue (RGB), camera mounted to unmanned aerial vehicle (UAV), by applying 3D photogrammetry techniques integrated with spectral analysis techniques.

Existing literature relating to crop height measurement and yield estimation by UAV photogrammetry survey and/or RGB spectral analysis was reviewed. The body of work completed previously by others revealed useful techniques and insights, and details about setbacks they experienced, enabling this study to build on this and focus primarily on the knowledge gap in applying digital surface model parameters and visible spectrum indices to yield prediction in sugarcane crops using multiple linear regression.

Field data was gathered at the subject canefields, including UAV photogrammetry survey at several epochs, topographic survey of the terrain, and empirical yield data at time of harvest. This data was processed and used to parameterise several models and evaluate the effectiveness of predictions.

Results of this study show that average plant height from digital crop surface models (PH_{CSM}) has a very strong correlation with sugarcane yield, even stronger than remotely sensed hyperspectral indices that are currently used in the industry. Unfortunately, though, the coefficient of the PH_{CSM} variable in linear regression is sensitive to changes in the height of the cane resulting from damage due to wind and rain. It is not possible to ascertain an accurate coefficient unless the model is parameterised with simultaneous empirical yield values, such as those obtained at harvest, thus limiting the usefulness of the model for forecasting yields. Nevertheless, PH_{CSM} is excellent for assessing variability within fields several months in advance of harvest, even if absolute values are not known. This is useful because results show there are indeed considerable variations within fields, which growers, given access to this information, may have a chance to ameliorate before final maturation of the cane.

Visible spectrum vegetation indices (VI_{RGB}s) are not reliable in predicting yield in mature cane, primarily because the cane yellows as it approaches maturity, which is paradoxical to the intention of the VI_{RGB}s. In any case, however, the PH_{CSM} correlation with yield was so strong that it left little room for another independent variable to improve the model without introducing multicollinearity.

Finally, this study demonstrated that photogrammetry using a consumer-grade UAV can be used to develop reasonably precise digital models of vegetation, and explored a possible application for this technology in precision agriculture

CHAPTER 7: Recommendations for further work

- Any additional independent variables to be input into MLR models with PH_{CSM} should focus on accounting for the degree of damage/slumping of the cane due to wind and rain, i.e. correction factors.
- Potential correction factors could be hyperspectral indices or ground-based observations about the degree of damage.
- Parameters for different varieties of cane should also be established.
- VI_{RGB}s may be useful for predicting yield and variability earlier in the growing season and should be investigated.

References

Aasen, H 2016, 'The acquisition of Hyperspectral Digital Surface Models of crops from UAV snapshot cameras', Universität zu Köln.

Bendig, J, Bolten, A & Bareth, G 2013, 'UAV-based imaging for multi-temporal, very high Resolution Crop Surface Models to monitor Crop Growth VariabilityMonitoring des Pflanzenwachstums mit Hilfe multitemporaler und hoch auflösender Oberflächenmodelle von Getreidebeständen auf Basis von Bildern aus UAV-Befliegungen', *Photogrammetrie-Fernerkundung-Geoinformation*, vol. 2013, no. 6, pp. 551-62.

Bendig, J, Bolten, A, Bennertz, S, Broscheit, J, Eichfuss, S & Bareth, G 2014, 'Estimating biomass of barley using crop surface models (CSMs) derived from UAV-based RGB imaging', *Remote Sensing*, vol. 6, no. 11, pp. 10395-412.

Bendig, J, Yu, K, Aasen, H, Bolten, A, Bennertz, S, Broscheit, J, Gnyp, M & Bareth, G 2015, *Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley*, vol. 39.

De Souza, CHW, Lamparelli, RAC, Rocha, JV & Magalhães, PSG 2017, 'Height estimation of sugarcane using an unmanned aerial system (UAS) based on structure from motion (SfM) point clouds', *International Journal of Remote Sensing*, vol. 38, no. 8-10, pp. 2218-30.

Dias, HB & Sentelhas, PC 2017, 'Evaluation of three sugarcane simulation models and their ensemble for yield estimation in commercially managed fields', *Field Crops Research*, vol. 213, pp. 174-85.

DJI 2019, *DJI Phantom 4 Pro V2.0 - Specs, Tutorials & Guides - DJI*, https://www.dji.com/au/phantom-4-pro-v2/info>.

Gates, DM 1980, *Biophysical Ecology*, Springer-Verlag, New York.

Geipel, J, Link, J & Claupein, W 2014, 'Combined Spectral and Spatial Modeling of Corn Yield Based on Aerial Images and Crop Surface Models Acquired with an Unmanned Aircraft System', *Remote Sensing*, vol. 6, no. 11.

Grenzdörffer, GJ 2014, 'Crop height determination with UAS point clouds', *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*.

Hunt, ER, Hively, WD, McCarty, GW, Daughtry, CST, Forrestal, PJ, Kratochvil, RJ, Carr, JL, Allen, NF, Fox-Rabinovitz, JR & Miller, CD 2011, 'NIR-green-blue high-resolution digital images for assessment of winter cover crop biomass', *GIScience & remote sensing*, vol. 48, no. 1, pp. 86-98.

Hunt, ERJ, Doraiswamy, PC, McMurtrey, JE, Daughtry, CST & Perry, EM 2013, 'A visible band index for remote sensing leaf chlorophyll content at the canopy scale', *Publications from USDA-ARS / UNL Faculty*.

Luhmann, T, Robson, S, Kyle, S & Boehm, J 2013, *Close-range photogrammetry and 3D imaging*, Walter de Gruyter.

Malambo, L, Popescu, SC, Murray, SC, Putman, E, Pugh, NA, Horne, DW, Richardson, G, Sheridan, R, Rooney, WL, Avant, R, Vidrine, M, McCutchen, B, Baltensperger, D & Bishop, M 2018, 'Multitemporal field-based plant height estimation using 3D point clouds generated from

small unmanned aerial systems high-resolution imagery', *International Journal of Applied Earth Observation and Geoinformation*, vol. 64, pp. 31-42.

Okuda, T, Suzuki, M, Numata, S, Yoshida, K, Nishimura, S, Adachi, N, Niiyama, K, Manokaran, N & Hashim, M 2004, 'Estimation of aboveground biomass in logged and primary lowland rainforests using 3-D photogrammetric analysis', *Forest Ecology and Management*, vol. 203, no. 1, pp. 63-75.

Olsen, S & Hussey, B 2015, 'Precision Agriculture for the sugarcane industry'.

Perritt, HH 2017, *Domesticating drones : the technology, law, and economics of unmanned aircraft*, Routledge, Abingdon, Oxon ;.

Portz, G, Amaral, Ld & Molin, JP 2012, 'Measuring sugarcane height in complement to biomass sensor for nitrogen management', in *Proceedings of 10th International Conference on Precision Agriculture (ISPA), IN, USA (CD-ROM)*.

Possoch, M, Bieker, S, Hoffmeister, D, Bolten, AA, Schellberg, J & Bareth, G 2016, 'Multi-Temporal crop surface models combined with the rgb vegetation index from UAV-based images for forage monitoring in grassland', *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, vol. 2016-, pp. 991-8.

Robson, A, Abbott, C, Bramley, R & Lamb, D 2013, 'Final report SRDC project DPI021 Remote sensing-based precision agriculture tools for the sugar industry'.

Star, C 2005, *Biology: Concepts and Applications*, Thompson Brooks/Cole.

Tilly, N, Hoffmeister, D, Cao, Q, Huang, S, Lenz-Wiedemann, V, Miao, Y & Bareth, G 2014, 'Multitemporal crop surface models: accurate plant height measurement and biomass estimation with terrestrial laser scanning in paddy rice', *Journal of Applied Remote Sensing*, vol. 8, no. 1, p. 083671.

USQ 2018, 'Module 6 - Close Range Photogrammetry', *SVY3400 Advanced Surveying - Study Book*.

Wan, L, Li, Y, Cen, H, Zhu, J, Yin, W, Wu, W, Zhu, H, Sun, D, Zhou, W & He, L 2018, *Combining UAV-Based Vegetation Indices and Image Classification to Estimate Flower Number in Oilseed Rape*, vol. 10.

Zhao, Y, Justina, DD, Kazama, Y, Rocha, JV, Graziano, PS & Lamparelli, RAC 2016, 'Dynamics modeling for sugar cane sucrose estimation using time series satellite imagery', in *SPIE Remote Sensing*, SPIE, p. 11.

Appendix A – Project Specification

ENG4111/4112 Research Project
Project Specification

For:	Clancy Sharp
Title:	*Estimation of Sugarcane Yield by UAV Photogrammetry Survey*
Major:	Surveying
Supervisors:	Glenn Campbell
Enrolment:	ENG4111 - EXT S1, 2019 ENG4112 - EXT S2, 2019
Project Aim:	To use spectral and spatial data captured by RGB camera mounted to unmanned aerial vehicle (UAV) to produce raster maps of cane height and cane health, to use this data to predict sugarcane yield.
Programme:	Version 2, 26th March 2019 1. Literature review and background research relating to current efforts of yield estimation for sugarcane and other crops, particularly if methods are like those proposed for the project. 2. Design a field measurement programme in consultation with cane growers, with focus on the two main varieties of cane in the Tweed region, and which will be harvested within the project timeframe. 3. Perform control surveys and topographic surveys of terrain (once only) and UAV photogrammetry surveys at three epochs. 4. Process data to create DTM (of terrain), DSM (of sugarcane), height map rasters, vegetation index images, vegetation quality raster. 5. Produce yield maps, validating against preliminary yield estimates from canegrowers and satellite imagery methods. 6. Evaluate predictions using final tonnage reports from sugar mill after harvest and propose model improvements. 7. Write dissertation.
If time and resources permit:	8. Compare yield map with soil test report / map for correlation with nutrient levels (nitrogen, phosphorous, potassium). 9. Prepare variable fertiliser application map for yield optimisation.

Appendix B – Survey and Flight Conditions

Table 10: Survey and flight conditions - 26 April

Date:	Friday, 26 April 2019
Time:	09:00 AM
Cloud cover:	Clear
Wind direction / speed:	W / 7 km/h
Sun azimuth / altitude	49.08° / 32.50°
UAV altitude:	120 m
Photo front overlap:	75%
Photo side overlap:	65%

Table 11: Survey and flight conditions - 16 May

Date:	Thursday, 16 May 2019
Time:	11:30 AM
Cloud cover:	20%
Wind direction / speed:	S / 18 km/h
Sun azimuth / altitude	3.92° / 42.61°
UAV altitude:	120 m
Photo front overlap:	75%
Photo side overlap:	65%

Table 12: Survey and flight conditions - 14 June

Date:	Friday, 14 June 2019
Time:	11:50 AM
Cloud cover:	Clear
Wind direction / speed:	NE / 10 km/h
Sun azimuth / altitude	-1.12° / 38.48°
UAV altitude:	120 m
Photo front overlap:	75%
Photo side overlap:	65%

Table 13: Survey and flight conditions - 12 July

Date:	Friday, 12 July 2019
Time:	09:30 AM
Cloud cover:	Clear
Wind direction / speed:	Calm
Sun azimuth / altitude	38.03° / 29.21°
UAV altitude:	120 m

Photo front overlap:	75%
Photo side overlap:	65%

Table 14: Survey and flight conditions - 21 July (immediately before harvest)

Date:	Sunday, 21 July 2019
Time:	09:00 AM
Cloud cover:	Clear
Wind direction / speed:	W / 6 km/h
Sun azimuth / altitude	45.39° / 25.84°
UAV altitude:	120 m
Photo front overlap:	75%
Photo side overlap:	65%

Table 15: Survey and flight conditions - 3 August (13 days after harvest)

Date:	Saturday, 3 August 2019
Time:	09:30 AM
Cloud cover:	Clear
Wind direction / speed:	Calm
Sun azimuth / altitude	41.27° / 32.66°
UAV altitude:	120 m
Photo front overlap:	75%
Photo side overlap:	65%

Appendix C – Sample polygon values for independent variables

BIN No.	PH$_{CSM}$	RGB Vegetation Indices			
		GLI	NGRDI	TGI	VARI
1	2.152	0.185	0.184	36.771	0.303
2	2.275	0.194	0.186	39.341	0.298
3	2.320	0.195	0.189	39.799	0.306
4	2.327	0.196	0.193	39.591	0.314
5	2.322	0.195	0.193	39.257	0.314
6	2.393	0.196	0.194	39.733	0.318
7	2.406	0.193	0.192	40.006	0.318
8	2.442	0.192	0.192	40.325	0.320
9	2.494	0.189	0.190	40.298	0.319
10	2.575	0.188	0.188	40.947	0.316
11	2.524	0.186	0.187	41.312	0.313
12	2.555	0.185	0.189	40.553	0.318
13	2.511	0.177	0.171	42.121	0.284
14	2.427	0.179	0.180	39.160	0.303
15	2.559	0.183	0.180	41.716	0.295
16	2.575	0.185	0.183	41.770	0.301
17	2.637	0.189	0.186	42.119	0.305
18	2.605	0.188	0.186	41.368	0.303
19	2.589	0.187	0.185	41.092	0.302
20	2.582	0.188	0.184	41.151	0.300
21	2.594	0.192	0.190	40.804	0.308
22	2.578	0.191	0.188	40.983	0.305
23	2.610	0.190	0.188	40.646	0.306
24	2.656	0.190	0.187	41.714	0.304
25	2.549	0.184	0.179	41.566	0.292
26	2.527	0.184	0.178	41.399	0.291
27	2.519	0.179	0.166	42.826	0.268

Survey epoch: 26/04/219

BIN No.	PH$_{CSM}$	GLI	NGRDI	TGI	VARI
1	2.380	0.143	0.139	36.653	0.241
2	2.559	0.147	0.139	39.542	0.239
3	2.604	0.147	0.143	40.328	0.246
4	2.594	0.148	0.145	40.070	0.250

Survey epoch: 16/05/219

5	2.575	0.147	0.144	39.474	0.249
6	2.644	0.149	0.146	39.483	0.252
7	2.665	0.149	0.147	39.169	0.254
8	2.697	0.152	0.149	38.841	0.258
9	2.754	0.152	0.151	38.274	0.260
10	2.828	0.153	0.150	38.911	0.259
11	2.793	0.154	0.154	38.275	0.265
12	2.832	0.155	0.157	37.494	0.273
13	2.750	0.151	0.146	37.535	0.250
14	2.616	0.149	0.147	35.400	0.255
15	2.809	0.153	0.147	38.856	0.251
16	2.816	0.154	0.149	38.818	0.254
17	2.886	0.153	0.149	39.334	0.255
18	2.849	0.150	0.145	39.269	0.249
19	2.835	0.149	0.145	39.132	0.249
20	2.835	0.147	0.142	39.071	0.243
21	2.839	0.149	0.143	39.050	0.246
22	2.819	0.150	0.145	39.172	0.248
23	2.854	0.150	0.144	39.322	0.246
24	2.910	0.152	0.147	40.209	0.252
25	2.799	0.150	0.144	39.957	0.246
26	2.776	0.152	0.147	39.675	0.251
27	2.745	0.150	0.137	40.983	0.231

Survey epoch: 14/06/219

1	2.498	0.138	0.103	33.270	0.169
2	2.639	0.144	0.106	36.387	0.170
3	2.691	0.150	0.115	37.684	0.185
4	2.693	0.150	0.116	37.594	0.188
5	2.668	0.149	0.116	37.150	0.187
6	2.740	0.151	0.116	37.347	0.188
7	2.770	0.152	0.118	37.613	0.191
8	2.808	0.154	0.122	37.581	0.197
9	2.858	0.153	0.122	37.555	0.199
10	2.946	0.153	0.124	37.528	0.202
11	2.909	0.153	0.126	37.831	0.207
12	2.957	0.152	0.129	37.109	0.213
13	2.848	0.140	0.101	39.117	0.164
14	2.752	0.144	0.113	35.340	0.187
15	2.916	0.145	0.110	38.840	0.178
16	2.918	0.146	0.112	38.677	0.182
17	2.998	0.147	0.114	39.207	0.185
18	2.960	0.145	0.109	38.600	0.176
19	2.933	0.143	0.107	38.575	0.172
20	2.935	0.143	0.107	38.709	0.172
21	2.944	0.144	0.107	38.284	0.172

22	2.926	0.143	0.106	38.075	0.171
23	2.943	0.144	0.106	37.967	0.171
24	2.999	0.145	0.106	38.752	0.171
25	2.890	0.143	0.103	38.768	0.164
26	2.869	0.144	0.105	38.756	0.167
27	2.816	0.138	0.087	41.541	0.136

Survey epoch: 12/07/219

1	2.173	0.125	0.086	31.838	0.143
2	2.308	0.132	0.084	35.755	0.134
3	2.373	0.138	0.092	37.295	0.147
4	2.381	0.138	0.094	37.322	0.149
5	2.369	0.138	0.093	37.215	0.149
6	2.416	0.138	0.091	36.988	0.146
7	2.442	0.138	0.093	37.389	0.149
8	2.469	0.140	0.095	37.121	0.153
9	2.516	0.140	0.098	36.613	0.158
10	2.571	0.140	0.099	36.499	0.160
11	2.551	0.144	0.108	36.448	0.175
12	2.604	0.145	0.115	35.792	0.189
13	2.534	0.133	0.087	38.857	0.141
14	2.373	0.136	0.096	32.583	0.157
15	2.550	0.141	0.096	37.493	0.154
16	2.555	0.143	0.097	37.837	0.156
17	2.652	0.149	0.104	39.562	0.166
18	2.613	0.148	0.102	38.672	0.162
19	2.599	0.145	0.096	38.798	0.151
20	2.597	0.146	0.098	38.986	0.154
21	2.579	0.146	0.096	38.217	0.150
22	2.555	0.146	0.096	37.727	0.151
23	2.582	0.147	0.096	37.825	0.151
24	2.631	0.151	0.099	39.049	0.154
25	2.538	0.147	0.093	39.519	0.145
26	2.525	0.148	0.095	38.969	0.148
27	2.509	0.141	0.079	42.389	0.121

Survey epoch: 21/07/219

1	2.396	0.126	0.058	27.945	0.090
2	2.531	0.132	0.053	31.667	0.078
3	2.595	0.139	0.062	33.495	0.092
4	2.600	0.139	0.063	33.402	0.095
5	2.590	0.138	0.063	33.481	0.094
6	2.643	0.138	0.061	33.263	0.090
7	2.671	0.139	0.063	33.762	0.094
8	2.699	0.141	0.064	33.225	0.094
9	2.745	0.141	0.067	32.797	0.100
10	2.807	0.141	0.067	32.680	0.099

11	2.792	0.148	0.081	32.658	0.120
12	2.854	0.154	0.095	32.103	0.142
13	2.790	0.134	0.061	34.978	0.091
14	2.620	0.131	0.062	27.862	0.097
15	2.805	0.138	0.063	32.966	0.095
16	2.808	0.139	0.063	33.020	0.094
17	2.906	0.144	0.070	34.284	0.105
18	2.871	0.143	0.070	33.632	0.107
19	2.855	0.140	0.065	34.217	0.099
20	2.854	0.141	0.068	34.565	0.104
21	2.832	0.139	0.065	33.901	0.101
22	2.803	0.140	0.067	33.978	0.104
23	2.835	0.139	0.063	34.212	0.098
24	2.878	0.143	0.067	35.865	0.102
25	2.781	0.141	0.062	36.588	0.094
26	2.766	0.143	0.065	36.325	0.098
27	2.750	0.138	0.050	39.561	0.074

Appendix D – Multiple linear regression statistics and performance

MLR (PH$_{CSM}$+GLI)

Survey Epoch	26/04/19	16/05/19	14/06/19	12/07/19	21/07/19
Days after planting	223	243	272	300	309
Days before harvest	86	66	37	9	0
Regression Statistics (Performance on Parameterisation Dataset)					
Multiple R	0.9407	0.9427	0.9660	0.9687	0.9611
R Square	0.8849	0.8887	0.9331	0.9384	0.9238
Adjusted R Square	0.8657	0.8702	0.9219	0.9281	0.9111
Standard Error	2.9124	2.8629	2.2206	2.1307	2.3697
Observations	15	15	15	15	15
RMSE (TCH)	2.6049	2.5606	1.9861	1.9057	2.1195
Coefficients					
Intercept	-122.6665	-105.5742	-142.5324	-104.8434	-116.8105
PH	72.5857	71.1535	70.6241	86.7321	66.1392
GLI	199.8021	40.8651	253.8220	-115.5751	222.5544
Performance on Validation Dataset					
Observations	8	8	8	8	8
R Square	0.9490	0.9256	0.8593	0.9040	0.9029
RMSE (TCH)	2.4708	2.5722	3.2470	2.8874	2.7382

MLR (PH$_{CSM}$+NGRDI)

Survey Epoch	26/04/19	16/05/19	14/06/19	12/07/19	21/07/19
Days after planting	223	243	272	300	309
Days before harvest	86	66	37	9	0
Regression Statistics (Performance on Parameterisation Dataset)					
Multiple R	0.9569	0.9536	0.9646	0.9700	0.9598
R Square	0.9156	0.9093	0.9305	0.9410	0.9213
Adjusted R Square	0.9015	0.8942	0.9189	0.9311	0.9082
Standard Error	2.4937	2.5845	2.2635	2.0853	2.4077
Observations	15	15	15	15	15
RMSE (TCH)	2.2304	2.3117	2.0245	1.8652	2.1535
Coefficients					

	26/04/19	16/05/19	14/06/19	12/07/19	21/07/19
Intercept	153.2719	-131.7097	-114.5108	-108.1661	-95.1040
PH	74.1831	67.6718	69.1748	77.3832	67.3159
NGRDI	344.9442	286.4500	119.6172	107.8355	96.1262

Performance on Validation Dataset

Observations	8	8	8	8	8
R Square	0.9454	0.8799	0.8548	0.8776	0.8670
RMSE (TCH)	2.7720	3.0671	3.2078	3.2875	3.0579

MLR (PH$_{CSM}$+TGI)

Survey Epoch	26/04/19	16/05/19	14/06/19	12/07/19	21/07/19
Days after planting	223	243	272	300	309
Days before harvest	86	66	37	9	0

Regression Statistics (Performance on Parameterisation Dataset)

Multiple R	0.9600	0.9444	0.9585	0.9700	0.9555
R Square	0.9217	0.8918	0.9187	0.9409	0.9130
Adjusted R Square	0.9086	0.8738	0.9052	0.9311	0.8985
Standard Error	2.4020	2.8233	2.4468	2.0865	2.5313
Observations	15	15	15	15	15
RMSE (TCH)	2.1484	2.5252	2.1885	1.8662	2.2641

Coefficients

Intercept	32.6675	-62.0076	-82.5652	-97.4915	-99.2422
PH	94.9131	69.4826	71.9133	84.5275	71.5401
TGI	-4.2580	-0.8385	-0.6930	-0.4849	-0.0345

Performance on Validation Dataset

Observations	8	8	8	8	8
R Square	0.9529	0.9191	0.9006	0.8980	0.9020
RMSE (TCH)	2.3024	2.4826	2.7092	2.9825	2.7469

MLR (PH$_{CSM}$+VARI)

Survey Epoch	26/04/19	16/05/19	14/06/19	12/07/19	21/07/19
Days after planting	223	243	272	300	309
Days before harvest	86	66	37	9	0

Regression Statistics (Performance on Paramaterisation Dataset)

Multiple R	0.9644	0.9576	0.9640	0.9701	0.9594
R Square	0.9300	0.9171	0.9292	0.9411	0.9205
Adjusted R Square	0.9183	0.9032	0.9174	0.9312	0.9073
Standard Error	2.2707	2.4716	2.2831	2.0836	2.4197
Observations	15	15	15	15	15

RMSE (TCH)	2.0310	2.2107	2.0421	1.8637	2.1642
Coefficients					
Intercept	-138.7420	-135.9221	-111.9954	-108.6889	-93.7295
PH	71.3356	68.4300	68.9576	78.5067	66.8279
VARI	186.0797	174.9485	63.5526	52.7273	63.5054
Performance on Validation Dataset					
Observations	8	8	8	8	8
R Square	0.9268	0.8711	0.8558	0.8774	0.8720
RMSE (TCH)	2.8187	3.1735	3.1856	3.2692	3.0119