2-D AND 3-D PROXIMAL REMOTE SENSING FOR YIELD ESTIMATION IN A SHIRAZ VINEYARD

CHRISTOPHER JAMES HACKING

Thesis presented in partial fulfilment of the requirements for the degree of Master of Science in the Faculty of Science at Stellenbosch University.

Supervisors: Dr Carlos Poblete-Echeverria

Mr Nitesh Poona

April 2020

DECLARATION

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

This thesis includes two original manuscripts, constituting Chapter 3 and Chapter 4:

Nature of Contribution
Chapter 3 Poblete-Echeverría was responsible for the conceptualisation of the work. Manzan acquired the data and assisted with data analysis. I (Hacking) was responsible for the literature review and main analysis, and I wrote the manuscript. Poblete-Echeverría and Poona provided content supervision and helped with editing the manuscript. Hacking C, Poona N, Manzan N & Poblete-Echeverría C 2019. Investigating 2-D and 3-D Proximal Remote Sensing Techniques for Vineyard Yield Estimation. *Sensors* 19, 17: 3652.
Chapter 4 I (Hacking) conceptualised this work, with the help of my supervisors. I completed the literature review, acquired and analysed the data, and wrote the manuscript. My supervisors were responsible for content supervision and the editing of the manuscript. Hacking C, Poona N & Poblete-Echeverría C 2020. Vineyard yield estimation using 2-D proximal sensing: A multitemporal approach. *OENO One* 54, 4: 793-812.

Date: March 2020

SUMMARY

Precision viticulture aims to minimise production input expenses through the efficient management of vineyards, yielding the desired quantity and quality, while reducing the environmental footprint associated with modern farming. Precision viticulture practices aim to manage the inherent spatial variability in vineyards. Estimating vineyard yield provides insight into this process, enabling informed managerial decisions regarding production inputs. At the same time, yield information is important to the winery, as it facilitates logistical planning for harvest.

Traditional yield estimation methods are destructive by nature and require in-situ sampling, which is labour-intensive and time-consuming. Proximal remote sensing (PRS) presents a suitable alternative for estimating yield in a non-destructive manner. PRS employs terrestrial proximal sensors for data acquisition that can be combined with computer vision (CV) techniques to process and analyse the data, generating the estimated yield for the vineyard. This research intends to investigate 2-dimensional (2-D) and 3-dimensional (3-D) PRS and related CV techniques for estimating yield in a vertically shoot position (VSP) trellised Shiraz vineyard.

This research is presented as two components. The first component evaluates 2-D and 3-D methodologies for estimating yield in a vineyard. Three experiments are presented at bunch- and plant-level, incorporating both laboratory and in-situ experimental conditions. Under laboratory conditions (bunch-level only), the 2-D methodology achieved an r^2 of 0.889, while the 3-D methodology achieved a higher r^2 of 0.950. Both methodologies demonstrate the potential of PRS and associated CV techniques for estimating yield. The in-situ plant-level results favoured the 2-D methodology (full canopy (FC): $r^2 = 0.779$; leaf removal (LR): $r^2 = 0.877$) over the 3-D methodology (FC: $r^2 = 0.487$; LR: $r^2 = 0.623$). The general performance of the 2-D methodology was superior, and thus implemented in the subsequent component.

Component two set out to determine the ideal phenological stage for estimating yield. The 2-D methodology was employed with slight improvements and multitemporal digital imagery were acquired on a weekly basis for 12 weeks; culminating in a final acquisition two days prior to harvest. This component also successfully implemented image segmentation using an unsupervised k-means clustering (KMC) technique, an improvement to the colour thresholding (CT) technique implemented in component one. The ideal phenological stage was approximately two weeks prior to harvest (final stages of berry ripening), which achieved a global (bunch-level: 50 bunches) r^2 of 0.790 for estimating yield.

This research successfully implements 2-D and 3-D PRS and CV techniques for estimating yield in a Shiraz vineyard, and thereby accomplishes the aim of this research. The research demonstrates the suitability of the methodologies – specifically the 2-D methodology, which demonstrated superior performance (simple data acquisition and analysis with competitive results). Future research could refine the presented methodologies for operational use.

KEY WORDS

Proximal remote sensing; yield estimation; RGB; RGB-D; image segmentation; surface reconstruction; precision viticulture; multitemporal; Kinect; computer vision

OPSOMMING

Presisie-wingerdbou het ten doel om die produksie insetkoste te verminder deur doeltreffende bestuur van wingerde, die gewenste opbrengs en kwaliteit te lewer, en terselfdertyd die omgewingsvoetspoor van moderne boerdery te verminder. Presisie wingerdboukundige praktyke is daarop gemik om die inherente ruimtelike variasie in wingerde te bestuur. Opbrengsberaming in die wingerd gee insig in hierdie proses, wat ingeligte bestuursbesluite rakende produksie-insette moontlik maak. Terselfdertyd is opbrengs inligting belangrik vir die kelder, aangesien dit die logistieke beplanning tydens oestyd vergemaklik.

Tradisionele opbrengsberamingsmetodes is destruktief van aard en benodig in-situ monsterneming, wat arbeidsintensief en tydrowend is. Kort-afstand waarneming (KAW) is 'n geskikte alternatief om die opbrengs op nie-destruktiewe wyse te skat. KAW gebruik land-gebasseerde kort-afstand sensors vir die insamel van data wat gekombineer kan word met rekenaarvisie-tegnieke om die data te verwerk en te ontleed, wat die geskatte opbrengs vir die wingerd lewer. Hierdie navorsing het ten doel om tweedimensionele (2-D) en driedimensionele (3-D) KAW en verwante rekenaarvisietegnieke te ondersoek om die opbrengs in 'n vertikale loot geposisioneerde (VLP) opgeleide Shiraz-wingerd te skat.

Hierdie navorsing word as twee komponente aangebied. Die eerste komponent evalueer 2-D en 3-D metodologieë vir die beraming van die opbrengs in 'n wingerd. Drie eksperimente word op tros- asook plantvlak aangebied, sowel as laboratorium- en in-situ. Onder laboratoriumtoestande (slegs op trosvlak) het die 2-D-metodologie 'n r^2 van 0,889 behaal, terwyl die 3-D-metodologie 'n hoër r^2 van 0,950 behaal het. Albei metodologieë demonstreer die moontlikheid van KAW en gepaardgaande rekenaarvisie-tegnieke om die opbrengs te skat. Die plant-vlak in-situ resultate het die 2-D-metodologie (vol-lower $r^2 = 0.779$; blaarverwydering $r^2 = 0.877$) bevoordeel bo die 3-D-metodologie (vol-lower $r^2 = 0.487$; blaarverwydering $r^2 = 0.623$). Die algehele prestasie van die 2-D-metodologie was beter en is gevolglik in die daaropvolgende komponent gebruik.

Komponent twee het ten doel gehad om die ideale fenologiese stadium vir die beraming van opbrengs te bepaal. Die 2-D-metodologie is met geringe verbeterings gebruik en multitemporale digitale beelde is weekliks ingesamel oor 12 weke, met die laaste beelde verkry twee dae voor oes. Hierdie komponent het ook beeldsegmentering suksesvol geïmplementeer met behulp van 'n onbewaakte k-gemiddeld groeperingstegniek, 'n verbetering in die kleurdrempelwaarde-tegniek wat in komponent een geïmplementeer is. Die ideale fenologiese stadium was ongeveer twee weke voor oes (finale

stadiums van korrelrypwording), wat 'n algehele (trosvlak: 50 trosse) r^2 van 0,790 behaal het om die opbrengs te skat.

Hierdie navorsing implementeer suksesvol 2-D en 3-D KAW en rekenaarvisietegnieke om opbrengs in 'n Shiraz-wingerd te skat, en hierdeur is die doel van die navorsing wel bereik. Die navorsing toon die geskiktheid van die metodologieë, spesifiek die 2-D-metodologie wat uitstekende prestasie getoon het (eenvoudige data-verkryging en -ontleding met mededingende resultate). Toekomstige navorsing kan die voorgestelde metodes vir operasionele gebruik verder verfyn.

TREFWOORDE

Kort-afstand waarneming; opbrengsskatting; RGB; RGB-D; beeldsegmentering; oppervlak rekonstruksie; presisie wingerdbou; multitemporeel; Kinect; rekenaarvisie

ACKNOWLEDGEMENTS

I sincerely thank:

- Mr Nitesh Poona for his mentorship and supervision during my Masters, and the numerous discussions – no matter the topic;
- Dr Carlos Poblete-Echeverria for his supervision and affording me the opportunity to work on this project;
- Winetech for their financial support;
- Dr Albert Strever for the many discussions over a cup of coffee;
- Ms Kelly McDowall for her professional language editing of this thesis and manuscripts;
- the staff and fellow students of the Department of Geography and Environmental Studies for all the constructive feedback;
- the Department of Viticulture and Oenology and the staff of Welgevallen Experimental Farm for all the assistance;
- Mr Kyle Loggenberg for his friendship and the many hours of procrastination over the past several years;
- Mr Kobus Bredell and Mr Aloïs Houeto for their assistance during the field season, especially with all the early morning data acquisition sessions. Their comradeship made the journey considerably more bearable;
- Ms Jayde Bromwich for her continued support and friendship, as well as her assistance with data acquisition during the field season; and
- my parents for their continued love and support.

Maintain course and speed

– J.G. Geddes (1925–2016)

CONTENTS

TABLES

FIGURES

ACRONYMS AND ABBREVIATIONS

2-D	2-Dimensional
3-D	3-Dimensional
AdaBoost	Adaptive Boosting
ATV	All-Terrain Vehicle
CT	Colour Thresholding
CV	Computer Vision
EM	Electromagnetic
EOS	Electro-Optical System
FC	Full Canopy
FN	False Negative
FP	False Positive
GDP	Gross Domestic Product
HSV	Hue, Saturation, Value
IR	Infrared
KMC	K-Means Clustering
LAI	Leaf Area Index
LiDAR	Light Detection and Ranging
LR	Leaf Removal
Mgt	Millardet et de Grasset
NDVI	Normalised Difference Vegetation Index
NIR	Near Infrared
NN	Neural Network
PRS	Proximal Remote Sensing

RGB	Red, Green, Blue
RGB-D	RGB-Depth
RMSE	Root Mean Square Error
ROI	Region of Interest
RTAB-Map	Real-Time Appearance-Based Mapping
SfM	Structure from Motion
SVM	Support Vector Machine
SWIR	Short Wave Infrared
TBAV	Total Bunch Area of Vine
TBVV	Total Bunch Volume of Vine
TN	True Negative
TP	True Positive
UAS	Unmanned Aerial System
VIS	Visible
VSP	Vertical Shoot Positioned

CHAPTER 1: INTRODUCTION

This chapter introduces the reader to the research. A brief background to the research is provided, followed by the problem statement, aim and objectives, description of the study area, and the research methodology and design. Finally, the thesis structure is outlined.

1.1 RESEARCH BACKGROUND

Advances in technology over the past three decades have revolutionised the agricultural industry, giving rise to the term 'precision agriculture' (Mulla 2013). Precision agriculture can be broadly defined as the efficient management of agricultural crops with inherent spatial variability for economic profit, while reducing the environmental impacts of farming (Blackmore 2003). Precision viticulture applies the same key aspects as precision agriculture in the daily management of the vineyard (Arnó et al. 2009). The inherent spatial variability within vineyards (Taylor et al. 2005) requires the use of precision viticulture techniques to delineate parcels[1] of varying quality within the vineyard (Matese & Di Gennaro 2015). The interpretation of spatial variability is considered the main advantage of precision viticulture (Arnó et al. 2009).

Like many countries, South Africa has seen an increase in precision viticulture practices (Arnó et al. 2009). Efficient management ensures appropriate irrigation and chemical application while specific agricultural tasks are performed at the correct phenological stage (Font et al. 2015). Various aspects of the vineyard, such as vineyard vigour, yield and quality, can be monitored, providing parcel-specific information (Mathews & Jensen 2013; Millan et al. 2019; Rose et al. 2016; Tang et al. 2016). Additionally, vine shape and size can be monitored per individual vine (Matese & Di Gennaro 2015). The field of remote sensing is now widely used for advanced vineyard monitoring (Cunha, Marçal & Silva 2010).

Remote sensing has enabled the remote monitoring of vineyards, acquiring information at various resolutions (per vineyard, farm, or region) and thereby resulting in better-informed decision-making (Matese & Di Gennaro 2015). Remotely sensed data acquired via spaceborne (satellite) or airborne (manned aircraft, and unmanned aerial system – UAS) platforms can be used to monitor the soil and

[1] *Agricultural parcel* can be defined as a continuous area of land representing a single crop group – i.e. crop type or cultivar. Alternatively, a parcel can define a specific area within a crop group, such as an area within a vineyard with homogeneous properties (European Parliament and the Council of the European Union 2013).

vines in the vineyard (Hall et al. 2002; Matese et al. 2015). Vegetation indices, such as the Normalised Difference Vegetation Index (NDVI), can be computed from remotely sensed data and used to estimate pruning weight (Dobrowski, Ustin & Wolpert 2003), canopy area (Hall, Louis & Lamb 2008), leaf area index (LAI) (Towers, Strever & Poblete-Echeverría 2019), and crop yield (Cunha, Marçal & Silva 2010). In addition, remote sensing can be utilised to map grape quality from hyperspectral data (Martín et al. 2007) or quantify LAI from digital images acquired with a UAS platform (Mathews & Jensen 2013). The application of remote sensing in precision viticulture is vast, while the advances in terrestrial sensor technology have resulted in the progressive field of proximal remote sensing (PRS) (Mulla 2013).

The use of terrestrial proximal sensors (PRS) in precision viticulture has gained traction since the turn of the century, with sensors frequently being mounted on commercial farming equipment (Arnó et al. 2009; Mulla 2013). Numerous proximal sensors have been attached to mechanical harvesters for monitoring yield during harvest (Matese & Di Gennaro 2015). Yield information provides insight regarding a vineyard's inherent spatial variability, a vital aspect of precision viticulture (Bramley & Hamilton 2004). Ideally, yield information is collected prior to harvest (i.e. yield estimation). Yield estimates determined during the growing season provide useful information to the vineyard manager and winemaker (Diago et al. 2015). Early season yield estimation, for example prior to vèraison (the onset of ripening indicated by the change of colour in the berries), facilitates the implementation of managerial decisions to safeguard production quantity and quality (Grossetête et al. 2012; Nuske et al. 2014). Additionally, yield estimates aid the winery with logistical planning for harvest (De la Fuente et al. 2015; Dunn & Martin 2004). The commercial benefit of yield knowledge prior to harvest has seen the expansion of research using PRS in precision viticulture (Aquino et al. 2018; Liu, Marden & Whitty 2013; Marinello et al. 2016; Millan et al. 2018; Nuske et al. 2014; Rose et al. 2016).

PRS techniques can acquire 2-dimensional (2-D) and 3-dimensional (3-D) datasets which can be analysed with specialised computer vision (CV) techniques for estimating vineyard yield (Aquino et al. 2018; Marinello et al. 2016). Recent studies (Diago et al. 2015; Font et al. 2015; Liu, Marden & Whitty 2013; Millan et al. 2018) favour the use of 2-D methodologies (incorporating PRS and CV techniques), such as capturing red, green, and blue (RGB) images with a digital camera, over 3-D methodologies, such as RGB-Depth (RGB-D) imagery, for yield estimation. Pixel-level processing (i.e. image classification) (Diago et al. 2012) is generally favoured over fruit-level processing (i.e. berry detection) (Nuske et al. 2011). Several forms of pixel-level processing use image segmentation techniques for bunch detection, such as rudimentary colour thresholding (CT) (Reis et al. 2012) or semi-supervised classification (Liu & Whitty 2015); effectively separating the bunch from the

background. From the segmented bunch, a yield metric, such as the number of pixels or a pixel-based bunch perimeter, can be calculated and used for yield estimation purposes by means of statistical regression (Liu, Marden & Whitty 2013).

The lesser favoured 3-D approach uses PRS techniques to acquire 3-D data, processing the data with CV techniques for vineyard yield estimation. The added spatial dimension of depth to standard RGB imagery has resulted in RGB-D sensors that are capable of capturing 3-D datasets, with pixels represented in the RGB colour space (Bengochea-Guevara et al. 2018; Quan et al. 2017). RGB-D sensors, for instance the Kinect™ (Microsoft, Redmond, United States), represent ideal low-cost sensors for 3-D data acquisition, enabling 3-D yield estimation via volume quantification (Andújar et al. 2016; Marinello et al. 2016; Wang & Li 2014). There is limited research regarding the use of RGB images processed with advanced CV techniques to reconstruct the captured scenes as 3-D models; point clouds, for example (Herrero-Huerta et al. 2015; Rose et al. 2016). However, this technique is costly, as it requires an advanced understanding of the process and a computer capable of processing large amounts of data (Rose et al. 2016). CV techniques play an important role in the processing of 2-D and 3-D PRS datasets, subsequently enabling accurate vineyard yield estimation – a vital metric in understanding the spatial variability of a vineyard.

1.2 PROBLEM STATEMENT

A 2015 report by Conningarth Economists estimated that the South African wine industry employed more than 300 000 people, both directly and indirectly (Conningarth Economists 2015). The report stated that the wine industry contributed over R36.1 billion (1.2%) to the country's Gross Domestic Product (GDP) in 2013 (Conningarth Economists 2015). South Africa maintains an influential stake in the global wine industry. It ranked ninth in world production in 2018, producing over 824 million litres of wine (SAWIS 2018). At the end of 2018, there were approximately 93 000 ha of cultivated wine-producing vineyards in South Africa, with Stellenbosch holding the highest hectarage (16%) among the ten established wine regions in South Africa (Floris-Samuels 2018). The South African wine industry, evidently important to the country's economy, requires careful management during production, and is guided by precision viticulture practices (Arnó et al. 2009). Yield information plays a vital role in precision viticulture, helping both the wine producers and the wine industry to obtain statistically accurate information (SAWIS 2019). Capturing accurate yield estimates is therefore crucial.

Traditional yield estimation methods are destructive, inefficient, and labour-intensive (Herrero-Huerta et al. 2015). Sampling requires the removal of healthy bunches during the growing season to

measure the bunch mass, before extrapolating the measurements to the entire vineyard for yield estimation, for example (Wolpert & Vilas 1992). Supplementing these measurements with historical data can improve the capabilities of manual yield estimation models (De la Fuente et al. 2015). However, the advancement of PRS and CV techniques, which enable a spatial understanding of the vineyard, have overcome the limitations of traditional methods for yield estimation (Nuske et al. 2011).

Accurate yield monitoring of entire vineyards is possible when integrating proximal sensors with harvesters, thereby removing the subjective sampling practices common in traditional methods (Matese & Di Gennaro 2015). Capturing yield information during harvest provides a better indication of the inherent variability within vineyards (Taylor et al. 2005). However, yield information is frequently desired prior to the actual harvest (Nuske et al. 2014). The last two decades have produced several approaches using PRS and related CV techniques, either 2-D (Dunn & Martin 2004; Millan et al. 2018) or 3-D (Herrero-Huerta et al. 2015; Marinello et al. 2016), to estimate a vineyard's yield before the vines are harvested. The majority of these studies have conducted yield estimation prior to harvest, i.e. the fruit is still on the vines, with data acquisition generally occurring a few days before harvest (Font et al. 2015; Millan et al. 2018). Ideally, yield estimations need to be determined early in the season (Aquino et al. 2018; Liu et al. 2017), maximising the amount of information that can be extracted from the data.

Two noticeable shortfalls were identified during a review of the literature. Firstly, various studies present a solid progression for yield estimation in the precision viticulture domain, especially in recent years. However, to date no study has undertaken a direct comparison between 2-D and 3-D methodologies for yield estimation prior to harvest. Secondly, despite limited yield estimation research portraying the use of PRS techniques for data acquisition early in the season, no studies have conducted yield estimation research on a multitemporal scale with weekly data acquisition.

Considering the gaps identified in the existing research, specifically regarding PRS and related CV techniques for yield estimation, the following two research questions were formulated:

1. Which PRS methodology, 2-D or 3-D, is better suited for yield estimation before harvest?
2. What is the optimal phenological stage during the growing season for estimating vineyard yield using 2-D PRS and related CV techniques?

1.3 AIM AND OBJECTIVES

The overarching aim of this research is to investigate 2-D and 3-D PRS and related CV techniques for yield estimation in a Shiraz vineyard.

To fulfil the above-mentioned aim, the following primary objectives were conceptualised:

1. Evaluate the use of 2-D (RGB) and 3-D (RGB-D) methodologies for yield estimation prior to harvest.
2. Investigate 2-D PRS and related CV techniques for yield estimation using multitemporal RGB data.

1.4 STUDY AREA

The study was undertaken at Welgevallen, a commercially operated experimental farm owned by Stellenbosch University (Stellenbosch University 2013). Figure 1.1B illustrates the approximate extent of Welgevallen (33°56’33.66” S; 18°51’59.20” E), situated in Stellenbosch, Western Cape, South Africa (Figure 1.1A). Stellenbosch forms part of the Western Cape’s coastal wine-grape region, boasting the highest number of private wine cellars in South Africa – more than 165 (SAWIS 2018). The coastal wine-grape region is ideal for vineyards and is characterised by a Mediterranean climate (Conradie et al. 2002). On average, Stellenbosch receives 700 mm of rainfall per year, with most rain occurring during June, July and August, the cooler winter months, where temperatures average a daily high of 15°C (Bekker et al. 2016). In contrast, the summer months are hot and dry, with average temperatures reaching a daily high of 27°C (Bekker et al. 2016).

Data was captured in a Shiraz vineyard (101–14 Mgt rootstock) situated on the Welgevallen experimental farm, as illustrated in Figure 1.1C. Data acquisition focused on 31 vines across three rows. The vineyard lies approximately 157 m above sea level and was planted in the year 2000. A North-South row direction and vine spacing of 2.7×1.5 m was implemented. A seven-wire vertical shoot position (VSP) trellis system is used, with three sets of adjustable canopy guidewires. A drip irrigation system is used to maintain a regular irrigation schedule throughout the long, dry summers.

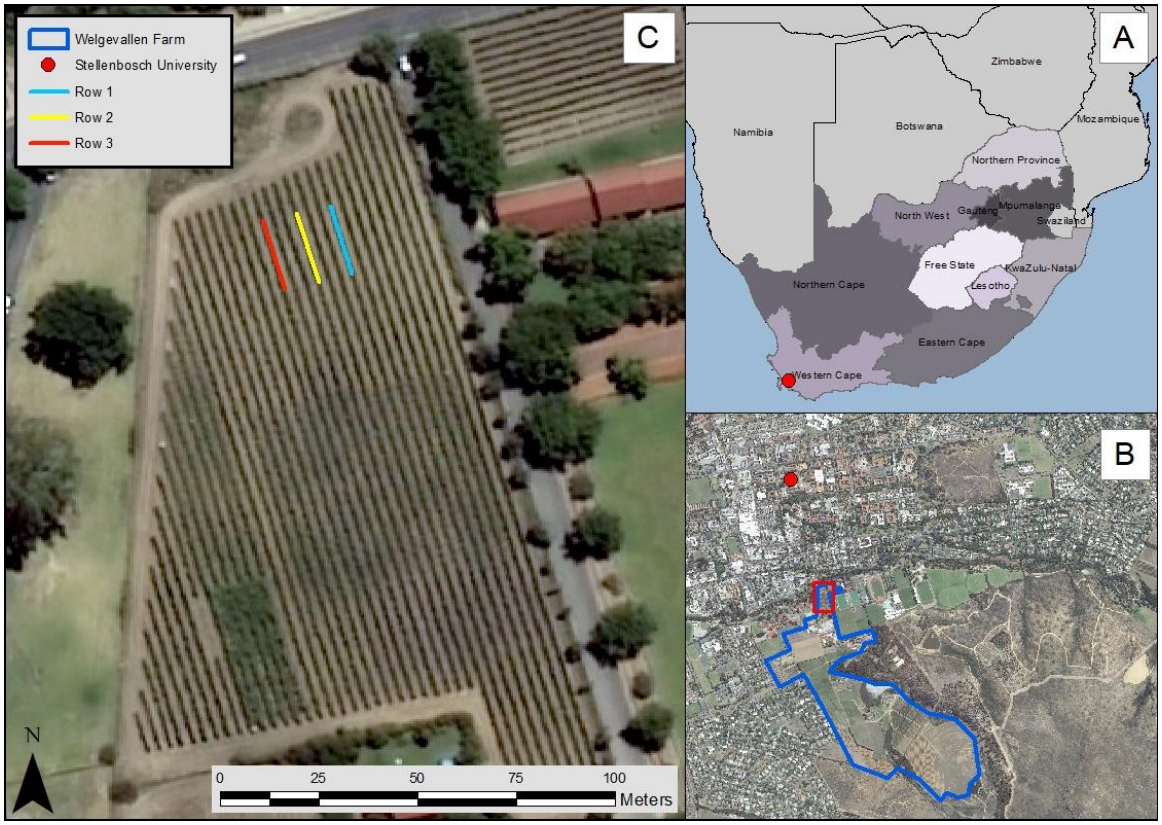


Figure 1.1 Stellenbosch University is located in Stellenbosch, Western Cape, South Africa (A). The university owns Welgevallen farm (B), where the Shiraz vineyard (C) is situated.

1.5 METHODOLOGY AND RESEARCH DESIGN

The research employed empirical methods to investigate the suitability of PRS and related CV techniques for vineyard yield estimation. Quantitative data was utilised to achieve the research objectives outlined in Section 1.3. The objectives were divided into two components, as portrayed in the research design outlined in Figure 1.2. The research investigated data acquired by a 2-D RGB camera and a 3-D RGB-D sensor, accompanied by manually collected reference measurements. Applied CV techniques combined qualitative (e.g. visual interpretation regarding suitable threshold selection) and quantitative (e.g. calculating bunch area) methods during image pre-processing.

The first component of the research used primary data collected during the 2016/2017 growing season to evaluate 2-D and 3-D PRS methodologies for yield estimation. The study was set up as a series of three experiments. The first experiment captured 2-D and 3-D data of individual bunches (subsequently referred to as ‘bunch-level’) under laboratory conditions, and experiment two captured in-situ data for the same bunches. Experiment three captured in-situ data of individual vines (termed ‘plant-level’). The in-situ datasets (captured in experiment two and three) were collected under two canopy treatments: full canopy (FC), standard canopy conditions in the bunch zone; and leaf removal

(LR), 100% leaf removal applied in the bunch zone. The refined methodologies are detailed in Chapter 3.

The second component of the research employed primary data collected throughout the 2018/2019 growing season. In-situ RGB imagery was collected at bunch-level on a weekly basis, from 8 December 2018 to 25 February 2019, totalling 12 temporal datasets. An additional, once-off plant-level dataset was captured on 25 February 2019 to complement the final bunch-level dataset. Weekly volume-reference measurements of the individual bunches were captured, with a final reference dataset captured after harvest. This component aimed to determine the ideal phenological stage for yield estimation by evaluating the multitemporal data captured. Chapter 4 specifies the methodologies of this component in more detail.

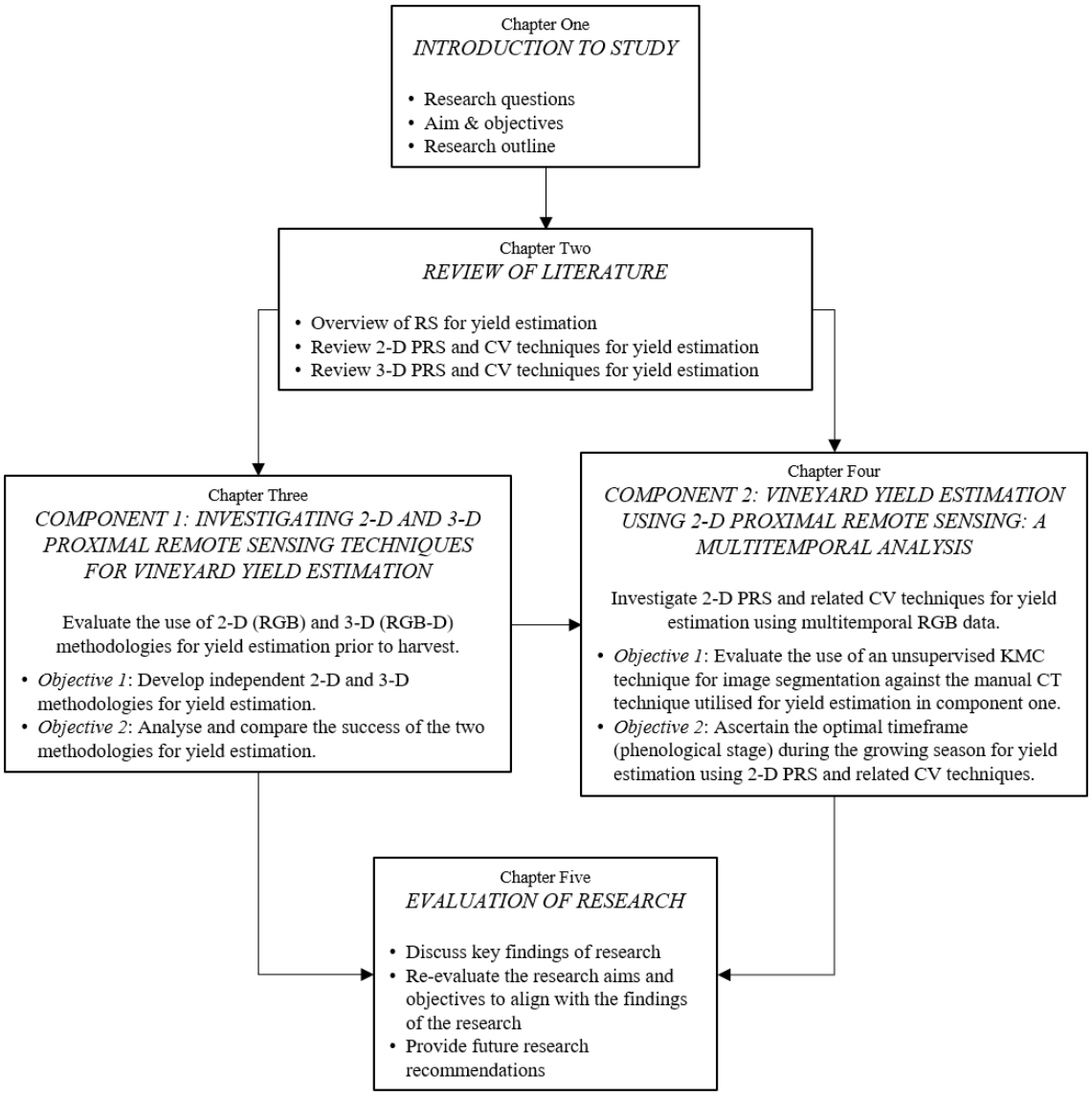


Figure 1.2 Research design and thesis structure.

1.6 STRUCTURE OF THESIS

This chapter has established the research problem, aim and objectives, as well as the study area, research design, and research methodology. The remaining chapters in this thesis are structured as follows:

Chapter 2 provides an in-depth review of the literature. An overview of remote sensing in precision viticulture, specifically with regard to yield estimation, is provided. Subsequently, PRS and related CV techniques for yield estimation are reviewed in alignment with the research aim.

Chapter 3 presents the first component of this research, as outlined in the research design (Figure 1.2). This chapter investigates 2-D and 3-D PRS methodologies for vineyard yield estimation. Both methodologies and findings associated with yield estimation are presented.

Chapter 4 presents the second component of this research. Chapter 4 compares supervised and unsupervised CV techniques for the processing of the RGB datasets. Additionally, the chapter reports the findings of temporal yield prediction during the growing season.

Chapter 5 evaluates the research in its entirety, drawing relevant conclusions from the results and re-evaluating the research aim and objectives. Future research recommendations are outlined.

Chapter 3 and Chapter 4 are presented as journal articles. Chapter 3 has been published in a special issue (Emerging Sensor Technology in Agriculture) of *Sensors*. Chapter 4 has been submitted for peer-review.

CHAPTER 2: REMOTE SENSING IN PRECISION VITICULTURE

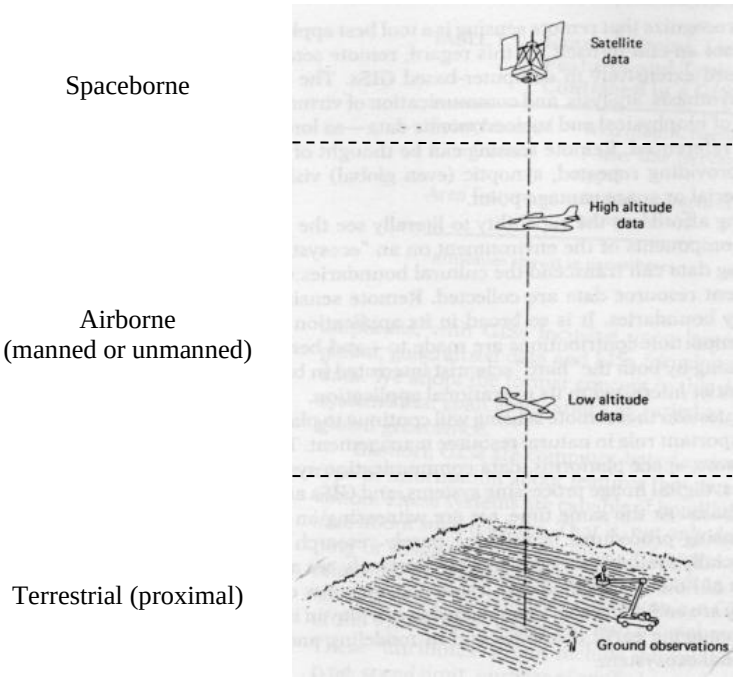
This chapter provides a broad review of remote sensing in precision viticulture, focusing on PRS for yield estimation in a vineyard. Additionally, the chapter reviews the current 2-D and 3-D PRS and related CV techniques employed for yield estimation in precision viticulture. Thereafter, a final summary of the literature is presented.

2.1 INTRODUCTION

Remote sensing is the science of monitoring features on earth (for example, vineyards) without directly interacting with said features (Hall et al. 2002). It has various applications (Cunha, Marçal & Silva 2010; Millan et al. 2018) that facilitate sustainable farming practices, such as precision viticulture (Blackmore 2003). Precision viticulture embraces the use of modern technology to overcome the inherent spatial variability within a vineyard by economically maximising the quality and quantity of the harvest while reducing the environmental impact of farming (Blackmore 2003). Remote sensing has therefore become an important part of precision viticulture, empowering informed decision-making from vineyard observations – specifically regarding the vines and soil (Arnó et al. 2009).

2.1.1 Remote sensing

Remote sensing can capture vast amounts of data with different spatial and temporal resolutions, providing information to the end user (Usha & Singh 2013). Conventional remote sensing employs optical sensors to capture surface features from a distance (Matese & Di Gennaro 2015). To this end, optical sensors capture electromagnetic (EM) radiation, which is reflected off the earth's surface, and store it as images (Ferrer et al. 2019). EM radiation is emitted from the sun and commonly transferred in three ranges of the EM spectrum: i) visible (VIS: 400–700 nm), ii) near infrared (NIR: 700–1300 nm), and iii) shortwave infrared (SWIR: 1300–2500 nm) (Mulla 2013). Optical sensors operate within these three ranges and can store multiple individual wavelengths, thereby preserving the EM radiation as images (Hall et al. 2002). Optical sensors are commonly housed by spaceborne, airborne, and terrestrial platforms for data acquisition (Matese et al. 2015; Matese & Di Gennaro 2015). Figure 2.1 indicates the platforms' positions relative to the ground.



Source: El-Khoury (2019)

Figure 2.1 Different remote sensing platforms: spaceborne, airborne, and terrestrial.

Spaceborne platforms (i.e. satellites) are placed into orbit for acquiring data of the earth's surface (Hall et al. 2002). Satellites are a common source of data, capable of remotely monitoring vast extents from a single image (Wójtowicz, Wójtowicz & Piekarczyk 2016). Historically, the level of information that could be extracted from satellite data was limited by the spatial resolution (pixel size; i.e. minimum definable object) of the sensor and the temporal resolution (revisit time) of the satellite (Usha & Singh 2013). For example, Landsat 1 was launched in 1972 with a spatial resolution of 80 m and a temporal resolution of 18 days (Nasa 2019). Advances in satellite technology over the last five decades has culminated in sub-metre spatial resolutions and revisit times of one to three days (Matese & Di Gennaro 2015). However, the high spatial and temporal resolution of modern satellite imagery has a substantial price tag attached to it (Hall et al. 2002).

Airborne remote sensing platforms are either manned (e.g. an aeroplane) or unmanned (e.g. UAS) (Matese et al. 2015), and can be equipped with various sensors, such as a multispectral (Ferrer et al. 2019) or hyperspectral (Martín et al. 2007) sensor. Airborne platforms can also facilitate specialised sensors uncommon to satellite platforms, such as LiDAR (light detection and ranging), which captures 3-D data (Mathews & Jensen 2012). In comparison to spaceborne platforms, airborne platforms facilitate ad hoc data acquisition with high spatial resolution at reduced costs, especially when a UAS is employed (Rey-Caramés et al. 2015). The sensor's proximity to the ground directly influences the spatial resolution, in turn controlling the acquisition cost.

Terrestrial remote sensing (i.e. PRS) incorporates an assortment of proximal sensors for in-situ data acquisition (Mulla 2013). PRS enables site-specific data to be collected in real time, monitoring both the vines (Diago et al. 2012) and the soil (Priori, Martini & Costantini 2010). For this reason, proximal sensors are commonly attached to farming equipment, acquiring data while traversing the vineyard (Matese & Di Gennaro 2015). PRS is becoming increasingly common in precision viticulture (Mulla 2013).

2.1.2 Remote sensing in precision viticulture

Precision viticulture utilises remote sensing as a tool to monitor various aspects of a vineyard, such as production inputs (e.g. water, fertilisers, and pesticides), vine phenology (e.g. canopy area and vigour) and crop health (e.g. canopy health and grape quality) (Arnó et al. 2009). Satellite data with a coarse spatial resolution limits the amount of information that can be extracted, and is typically monitored at vineyard-level (Matese & Di Gennaro 2015). This level of information provides the farm manager with a broader overview (of the farm) that is less precise at the plant-level (Wójtowicz, Wójtowicz & Piekarczyk 2016). Examples include the delineation of crop boundaries (Rydberg & Borgefors 2001), mapping different crop types (Schultz et al. 2015) and assessing general crop conditions (Doraiswamy et al. 2004).

Developments in spaceborne and airborne technology have enabled parcel-level (and, in some cases, plant-level) observations with high spatial resolutions (Johnson et al. 2001). The last decade has seen a significant rise in the deployment of UAS platforms for data acquisition, and they have become increasingly popular in precision viticulture (Di Gennaro et al. 2019). For example, Mathews and Jensen (2013) operated a UAS for quantifying LAI within a vineyard, while Weiss and Baret (2017) described the 3-D macro-structure of a vineyard using UAS-captured RGB imagery. Monitoring the vine's water stress is another example of airborne remote sensing in precision viticulture (Bellvert et al. 2014; Matese et al. 2018). Examples of additional precision viticulture applications that use remote sensing (both spaceborne and airborne) include indicating vine vigour and biomass (Hall et al. 2002); calculating LAI (Towers, Strever & Poblete-Echeverría 2019); the determination of pruning mass (Dobrowski, Ustin & Wolpert 2003; Smit, Sithole & Strever 2010) and vineyard variability (Hall, Louis & Lamb 2008); and estimating canopy area (Tang et al. 2016) and expected yield (Cunha, Marçal & Silva 2010; Ferrer et al. 2019; Hall et al. 2011; Sun et al. 2017).

In the last two decades, PRS techniques for data acquisition have increased in popularity within precision viticulture (Dunn & Martin 2004; Marinello et al. 2017; Nuske et al. 2011). They present flexible and cost-effective alternatives to spaceborne and airborne remote sensing platforms, and are

capable of high spatial and temporal resolutions (Mulla 2013). PRS uses various proximal sensors to facilitate in-situ data acquisition (Nuske et al. 2014). These include optical cameras (Aquino et al. 2018), spectral sensors (Loggenberg et al. 2018), depth sensors (Marinello et al. 2016), thermal sensors (Fuentes et al. 2012), and EM soil sensors (Priori et al. 2018). PRS applications vary from mapping soil variability (Priori et al. 2018) to monitoring the vine's canopy and health (Mulla 2013). For instance, calculating the LAI of the vineyard using either LiDAR (Arnó et al. 2013), 2-D digital imagery (Fuentes et al. 2014), or 3-D RGB-D data (Marinello et al. 2017); using PRS for early disease detection (Gallo et al. 2017); or modelling water stress using hyperspectral (Loggenberg et al. 2018) and thermal (Fuentes et al. 2012) sensors. PRS has also been implemented at bunch- (Font et al. 2015) and berry-level (Nuske et al. 2011) for yield estimation purposes. Evidently, PRS is an important tool in precision viticulture.

2.1.3 Yield estimation

Yield information provides insight into the inherent spatial variability of vineyards, a vital component of precision viticulture (Arnó et al. 2009). Advanced yield estimates determined early in the growing season enable managerial decisions to be made in order to achieve an optimal fruit quality out of a desired yield quantity (Arnó et al. 2009; Nuske et al. 2014). Remote sensing has facilitated yield estimation in vineyards (Sun et al. 2017), while overcoming limitations associated with traditional methods – such as destructive sampling (Wolpert & Vilas 1992).

Yield estimates can be computed from spaceborne, airborne, or terrestrial (PRS) sensors (Wójtowicz, Wójtowicz & Piekarczyk 2016). Typically, remotely sensed data from spaceborne platforms (satellites) utilises vegetation indices – a product of multispectral data – for estimating the yield (Cunha, Marçal & Silva 2010; Sun et al. 2017). NDVI, a popular choice, signifies photosynthetically active biomass (Hall et al. 2011). This can be used for estimating canopy area (Tang et al. 2016) and LAI (Towers, Strever & Poblete-Echeverría 2019); both techniques demonstrate promise for estimating yield (Sun et al. 2017). In 2010, Cunha, Marcal and Silva (2010) extracted NDVI values from satellite data and correlated the values with historical yield values to estimate the respective season's yield. The authors were able to achieve an r^2 (coefficient of determination) value between 0.593 and 0.774 for yield estimation. More recently, Sun et al. (2017) extracted NDVI and LAI values from satellite images to compute separate spatial variability estimates within a vineyard. The authors achieved similar experimental results, with the highest r^2 values of 0.689 and 0.672 from the respective NDVI and LAI values.

While satellite data demonstrates promise for yield estimation, there are several limitations to consider. For instance, cloud cover prevents optical sensors from acquiring data on overcast days (Hall et al. 2002), while vegetation indices can be influenced by the soil or other photosynthetic properties of the canopy (Mulla 2013). Moreover, the acquisition of high-resolution satellite data (commercial satellites) that is capable of producing yield estimates at parcel- or plant-level can be extremely expensive (Matese & Di Gennaro 2015). Unfortunately, this is not always feasible for smaller farms, particularly in third-world countries (Mulla 2013). The alternative is acquiring free satellite data, at the cost of a coarser spatial resolution (Matese & Di Gennaro 2015). It is therefore important to consider the desired spatial and temporal resolution of satellite data, and the associated costs. Airborne platforms, both manned (such as an aeroplane (Ferrer et al. 2019; Hall et al. 2011)) and unmanned (such as a UAS (Rey-Caramés et al. 2015)), present an alternative solution to satellite data. Airborne sensors can operate under elevated cloud cover (Usha & Singh 2013) and are capable of higher spatial and temporal resolutions at a reduced cost (Mathews 2015). However, airborne (and spaceborne) sensors are unable to operate at bunch- and berry-level, which would require terrestrial sensors for yield estimation at this scale (Matese & Di Gennaro 2015).

The last decade has seen a significant increase in precision viticulture research that uses PRS and related CV techniques for vineyard yield estimation (Diago et al. 2012; Di Gennaro et al. 2019; Rose et al. 2016). PRS employs digital (optical) sensors for data acquisition (Millan et al. 2018), directly monitoring the vine's fruit (bunches) and not the canopy (Liu, Marden & Whitty 2013). Research has successfully investigated various 2-D (Aquino et al. 2018; Diago et al. 2012; Nuske et al. 2014) and 3-D (Herrero-Huerta et al. 2015; Marinello et al. 2016; Rose et al. 2016) methodologies for vineyard yield estimation. However, research conducted in recent years has favoured 2-D methodologies (Aquino et al. 2018; Di Gennaro et al. 2019; Millan et al. 2018). The subsequent sections of this chapter will review relevant literature that used 2-D and 3-D PRS and related CV techniques for yield estimation. This is summarised in Table 2.1.

Table 2.1 Relevant 2-D and 3-D PRS and CV techniques for yield estimation

Reference (Chronological)	Cultivar	Phenological Stage	Scale	Trellis system	Sensor (2-D or 3-D)	CV Technique	Yield Metric	Yield Results
(Dunn & Martin 2004)	Cabernet Sauvignon	Pre-harvest	Bunch	VSP	2-D: RGB	2-D: CT	Pixel count	$r^2 = 0.720$
(Nuske et al. 2011)	Traminette & Riesling	Pre-harvest	Berry	VSP	2-D: RGB	2-D: Radial symmetry transform	Berry count	$r^2 = 0.740$
(Diago et al. 2012)	Tempranillo	Pre-harvest	Bunch	VSP	2-D: RGB	2-D: Mahalanobis distance	Pixel count	$r^2 = 0.730$
(Grossetête et al. 2012)	–	Pre-vèraison	Berry	–	2-D: Smartphone + Flash	2-D: Spectral peak detection	–	–
(Reis et al. 2012)	Red & White	Pre-harvest	Bunch	–	2-D: RGB + Flash	2-D: CT	–	–
(Liu, Marden & Whitty 2013)	Shiraz	Post-harvest	Bunch	Lab	2-D: RGB	2-D: CT	Volume, pixel count, perimeter, berry count, berry size	$r^2 = 0.770$
(Nuske et al. 2014)	×4	Pre-vèraison & pre-harvest	Berry	Multiple	2-D: RGB + Flash	2-D: Radial symmetry transform + spectral peak detection [similar]	Berry count	$r^2 = 0.600$ – 0.730
(Font et al. 2015)	Flame Seedless	Pre-harvest	Bunch	Table grape	2-D: RGB + Flash	2-D: CT	Pixel count	Yield error = 16%
(Diago et al. 2015)	×7	Post-harvest	Berry	Lab	2-D: RGB + Flash	Canny algorithm	Berry count	$r^2 = 0.840$
(Liu & Whitty 2015)	Shiraz & Cabernet Sauvignon	Pre-harvest	Bunch	VSP	2-D: RGB	2-D: SVM	–	–

Continued overleaf

Table 2.1 Continued.

Reference (Chronological)	Cultivar	Phenological Stage	Scale	Trellis system	Sensor (2-D or 3-D)	CV Technique	Yield Metric	Yield Results
(Herrero-Huerta et al. 2015)	Tempranillo	Pre-harvest	Bunch	VSP	2-D: RGB	3-D: Surface reconstruction	Volume	$r^2 = 0.780$
(Ivorra et al. 2015)	×10	Post-harvest	Bunch	Lab	2-D: Stereo	3-D: Surface reconstruction	Volume	$r^2 = 0.820$ (Volume)
(Liu, Whitty & Cossell 2015a)	Shiraz & Cabernet Sauvignon	Post-harvest	Berry	Lab	2-D: RGB + Flash	3-D: Sphere reconstruction	Berry count with sparsity factor	$r^2 = 0.850$
(Luo et al. 2016)	Summer Black	Pre-harvest	Bunch	Y & T trellis	2-D: RGB	AdaBoost	–	–
(Marinello et al. 2016)	Dan Ben Hannah & Dauphine	–	Bunch	Table grape	3-D: RGB-D	–	Volume	10–15% mass error deviations
(Rose et al. 2016)	Riesling	Pre-harvest	Bunch	VSP	2-D: Stereo + Flash	3-D: SfM	–	–
(Aquino et al. 2018)	×5	Pre-vèraison	Berry	VSP	2-D: RGB	NN	Berry count	RMSE = 0.480 kg
(Millan et al. 2018)	×5	Pre-harvest	Berry	VSP	2-D: RGB	Mahalanobis Distance	Berry count	$r^2 = 0.810$

2.2 2-D PRS AND CV FOR YIELD ESTIMATION

The literature presents multiple studies that incorporate 2-D techniques, whereby RGB images are acquired at bunch- and plant-level (Millan et al. 2018). The proximal sensor employed for capturing high quality (i.e. very high spatial resolution) images is a digital camera (Nuske et al. 2011). Although digital cameras capture images in the RGB colour space, the images are typically transformed into an alternative colour space during image processing (Liu, Marden & Whitty 2013). The hue, saturation, and value (HSV) colour space, among others, is a common substitute for processing yield estimates (Font et al. 2015). The transformed colour spaces are better suited for image segmentation, a CV technique employed for bunch detection (Luo et al. 2016).

Segmentation is commonplace in image processing, yielding a binary output – bunch and background (Liu & Whitty 2015). Dunn and Martin (2004) presented one of the first 2-D PRS studies for yield estimation, where colour thresholds were manually selected for image segmentation – effectively 'detecting' the bunches in the image. From the binary image, the number of pixels representing the 'bunch' class were counted, producing a yield metric. The authors regressed this metric with the known yield from harvest, achieving an overall r^2 of 0.720 (Dunn & Martin 2004). This was the first implementation of PRS and CV techniques in precision viticulture for yield estimation. Almost a decade later, advancing techniques saw studies incorporating supervised image classification techniques (Diago et al. 2012) and alternative yield metrics (Liu, Marden & Whitty 2013), achieving comparable yield estimation r^2 values of 0.730 and 0.770, respectively.

Around the same period, Nuske et al. (2011) presented an algorithm for berry detection, as opposed to bunch detection (via image segmentation). The authors used radial symmetry to detect berries and determine a berry count for yield estimation, achieving an r^2 value of 0.740 – on par with the segmentation results. In more recent years, studies have investigated the use of semi-supervised classifiers for bunch detection before vèraison (Aquino et al. 2018), and before harvest (Millan et al. 2018). Further details are provided in the following sub-sections.

2.2.1 Bunch detection

Using 2-D PRS and CV techniques for yield estimation relies heavily on being able to successfully detect the bunches in the image and separate them from the rest of the image (background). Bunch detection can be facilitated by CV techniques, such as image segmentation, which divides the image into different parts, generally yielding a binary image (bunch and background) (Font et al. 2015). Image segmentation depends on the colour properties of the image to differentiate the different segments (Reis et al. 2012). CT was one of the first techniques where thresholds were manually

selected for the respective colour channels, red, green, and blue, in the case of Dunn and Martin (2004). The thresholds were applied to the image, and the pixels that fell within the specified thresholds were selected, representing bunch. Additionally, Reis et al. (2012) and Liu, Marden and Whitty (2013) successfully implemented CT using the RGB colour space. The principle is rudimentary, yet effective if correctly implemented.

To remove human dependency from the processing activity, more automated techniques have been developed for bunch detection (Diago et al. 2012; Liu & Whitty 2015; Luo et al. 2016). For example, Diago et al. (2012) incorporated a Mahalanobis distance classifier, which required supervised training for the seven classes defined. The classifier computes the distance between the image's colour properties and the colour properties of the training dataset, statistically classifying the image according to the similarities present between colour properties (Diago et al. 2012; Font et al. 2015). After the classifier was trained, independent images were classified into seven clusters (segments/classes), including a cluster representing bunches. The authors achieved a correct classification accuracy of 98% for the grape (bunch) class (Diago et al. 2012). In 2015, Font et al. assessed several different pixel-based segmentation techniques for in-situ detection of red table grape bunches under artificial light at night. CT and Mahalanobis distance were among the techniques tested. The authors concluded that thresholding in the hue component of the HSV colour space achieved the lowest segmentation error (13.550%).

Liu and Whitty (2015) and Luo et al. (2016) utilised supervised classifiers for bunch detection in vineyards. However, these studies did not conduct yield estimations. Liu and Whitty (2015) presented a bunch detection approach that consisted of three major steps. First, potential bunch areas were determined via thresholding in the H and V channels of the HSV colour scheme. Second, a supervised feature selection was implemented for bunch detection from the potential bunch areas that were previously determined. Finally, a support vector machine (SVM) determined the appropriate training size for this process. An SVM determines the appropriate hyperplane to separate the data within the training dataset, enabling values to be classified according to the side of the hyperplane they occur on. Liu, Whitty and Cossell (2015b) were able to achieve a bunch detection accuracy of 0.880, with a total recall of 0.916. In comparison, Luo et al. (2016) achieved a higher bunch detection accuracy of 0.937. The authors created four linear classification models from different colour components, an improvement to the two components (channels) used by Liu and Whitty (2015). Subsequently, Luo et al. (2016) implemented the adaptive boosting (AdaBoost) framework to weight the linear classification models and create a single strong classifier. Interestingly, Luo et al. (2016) reported less over-segmentation than Liu and Whitty (2015) when directly comparing the two methodologies.

Morphological operators are commonly implemented post-segmentation to further refine the binary image at pixel-level. A popular combination of morphological operators includes erosion and dilation filters (Millan et al. 2018). Simply put, dilation closes gaps in the binary image, while erosion removes outlying pixels (Chudasama et al. 2015). Several studies (Diago et al. 2012; Font et al. 2015; Millan et al. 2018) included morphological operators post-segmentation, presenting improved yield estimation results. For instance, Font et al. (2015) implemented a combination of operators that reduced the segmentation error (13.550%) even further (10.010%).

Image segmentation techniques have evolved over the last two decades, enabling more accurate bunch detection (Luo et al. 2016). While CT was once the standard (Dunn & Martin 2004), it has since been replaced by supervised techniques (Diago et al. 2012; Liu & Whitty 2015). The main limitation of CT is the reliance on a specialist to select the thresholds, a biased step in the process. Nonetheless, CT is still a good benchmark for assessing new techniques on small datasets (Font et al. 2015). Supervised techniques have gained traction, as they limit the human involvement in training the classifier, making the final segmentation process more automated and suitable for larger datasets (Luo et al. 2016). The training stage is still susceptible to human-induced errors and dependent on the colour properties of the image (Diago et al. 2012).

With advancing techniques, such as unsupervised classification, the limitations of supervised techniques could be bypassed. Examples of unsupervised classification in precision viticulture are extremely limited, as the classifiers are susceptible to noise, hindering the classification of a busy environment like a vineyard (Diago et al. 2012). Nonetheless, Correa et al. (2012) investigated the use of unsupervised classification for feature extraction in a vineyard. The authors achieved an overall image segmentation accuracy of 0.950, which aligns with the classification accuracy presented by Diago et al. (2012). However, to date, no study has incorporated unsupervised image segmentation techniques, such us k-means clustering (KMC) (Arthur & Vassilvitskii 2007), for bunch (or berry) detection when estimating vineyard yield. Future yield estimation research should investigate the potential of unsupervised classification techniques for image segmentation.

2.2.2 Berry detection

Berry detection techniques have been presented in several studies (Aquino et al. 2018; Diago et al. 2015; Grossetête et al. 2012; Millan et al. 2018), whereby CV is used to count individual berries, working at a finer scale (compared to bunch- or even vineyard-level) for estimating yield. Grossetête et al. (2012) presented an early yield estimation technique where images were captured prior to vèraison, at night, using a smartphone. The smartphone's integrated flash was the source of artificial

illumination, causing a differentiable specular reflection on the centre of the berry. Effectively, finding these specular reflections enabled the authors to determine the centre of each berry, and thus determine a berry count per bunch, achieving an r^2 of 0.920 (based on a polynomial model, not linear) between estimated and actual berry count (Grossetête et al. 2012).

More recently, Aquino et al. (2018) and Millan et al. (2018) – the same research group – investigated the capability of on-the-go captured images for yield estimation prior to vèraison and prior to harvest, respectively. Both studies conducted data acquisition at night, using an all-terrain vehicle (ATV) with a customised digital camera and artificial lighting setup. Aquino et al. (2018) conducted the research prior to vèraison. The authors detected individual berry candidates using three main steps: i) mosaic overlapping images together, ii) determining berry candidates through morphological filtering in the L*a*b colour space (converted from RGB colour space), and iii) employing supervised neural network (NN) classifiers to remove false berry candidates. The NN classification was based on 17 descriptors – including colour properties – to determine the probability of the selected candidate being a berry, before a threshold was applied to remove non-berry candidates. This study was able to detect berries from an external validation dataset (defoliated vines), with a recall of 0.876 and a precision of 0.958.

Berry detection has demonstrated promise as an alternative option to bunch detection for estimating yield (Nuske et al. 2011). Berry detection has been successfully implemented prior to vèraison and after vèraison on white cultivars (Nuske et al. 2014). This can be achieved because the CV techniques implemented in these cases do not rely on the colour properties of the image, and thus green grapes and green leaves (the background) do not interfere with each another. However, the acquisition of data at night suits the controlled illumination, which can be logistically challenging to implement and is therefore a potential hindrance to commercialisation (Aquino et al. 2018).

2.2.3 Yield estimation metric

An estimation metric is a quantifiable product extracted from the 2-D data and employed for estimating bunch or berry mass, depending on the scale of analysis (Aquino et al. 2018). At bunch-level, Liu, Marden and Whitty (2013) investigated five yield metrics on Shiraz bunches under laboratory conditions: i) volume, ii) pixel area (number of pixels), iii) perimeter, iv) berry number, and v) berry size. The authors concluded that the best metric for estimating individual bunch mass was the pixel area, yielding an r^2 of 0.770 (Liu, Marden & Whitty 2013). This is one of the simplest metrics, as it requires a pixel count from the segmented binary image to indicate ‘pixel area’ (i.e. bunch area) (Liu, Marden & Whitty 2013). This has been used as a yield metric on several occasions

(Diago et al. 2012; Dunn & Martin 2004; Font et al. 2015). Employing the pixel count as a metric is only suitable at bunch-level, as it is extracted from the segmented image during bunch detection.

Berry detection techniques yield a berry count per bunch, which is used for estimating yield (Nuske et al. 2011). It is common to include the historical berry masses for yield estimation, supposedly improving the estimation capabilities (Aquino et al. 2018). For instance, Nuske et al. (2014) (continuation of Nuske et al. (2011)) counted the number of visible berries and compared this to the harvest yield. The authors investigated the proposed technique on several cultivars over four seasons, with data acquisition occurring before vèraison and before harvest (ranging between one to ten days prior to harvest). The following r^2 values were achieved: i) Traminette: $r^2 = 0.730$, ii) Riesling: $r^2 = 0.670$, iii) Flame seedless: 0.600, and iv) Chardonnay: 0.650 (Nuske et al. 2014). Similarly, Aquino et al. (2018) used the number of berries for estimating yield. However, the authors included historical berry masses. An average root mean square error (RMSE) of 0.480 kg was attained per image segment (consisting of three vines). Several studies (Diago et al. 2015; Millan et al. 2018; Nuske et al. 2011) have presented comparable results using a berry count.

Recent years have seen more research conducted with the ‘berry count’ metric (Aquino et al. 2018; Millan et al. 2018). However, these results are still comparable to pixel metrics (Font et al. 2015; Liu, Marden & Whitty 2013). The limitations to consider occur during the bunch or berry detection, as detailed in Section 2.2.1 and Section 2.2.2, respectively. The logistical challenges associated with berry-level techniques could be a potential hindrance to future research, as was the case in this research. Nevertheless, 2-D PRS and related CV techniques are evidently suitable for conducting yield estimation in a non-destructive manner within a vineyard.

2.3 3-D PRS AND CV FOR YIELD ESTIMATION

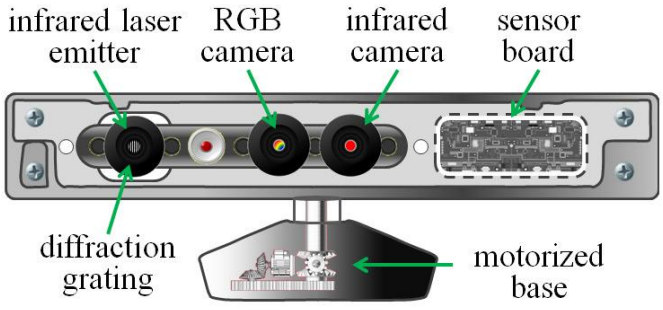
A limited amount of research has investigated 3-D PRS and related CV techniques for vineyard yield estimation (Marinello et al. 2016; Rose et al. 2016). 3-D methodologies employ a proximal 3-D sensor, such as an RGB-D sensor, for data acquisition and make use of 3-D CV techniques to process the data for computing yield estimates (Marinello et al. 2016). Several studies (Herrero-Huerta et al. 2015; Ivorra et al. 2015; Liu, Whitty & Cossell 2015a; Rose et al. 2016) have employed 2-D PRS techniques for data acquisition but implemented 3-D CV techniques for data processing, yielding 3-D volumetric models. For the purposes of this research, these studies have been incorporated as ‘3-D techniques’.

This research incorporated a Microsoft Kinect™ RGB-D sensor for data acquisition. However, research employing an RGB-D sensor for yield estimation in precision viticulture is extremely

limited, with only a single study (Marinello et al. 2016) employing a Kinect sensor. For this reason, a wider review of the literature was conducted regarding yield estimations with Kinect (RGB-D) sensors. In addition, a review of the 3-D CV techniques employed for processing the 2-D PRS acquired data, producing yield estimates, is outlined in this section.

2.3.1 RGB-D sensors for yield estimation

An RGB-D sensor is effectively an RGB camera that can capture depth. This enables the acquisition of 3-D datasets representing objects, while retaining the object's colour properties (Bengochea-Guevara et al. 2018). The Kinect V1 (also known as the Kinect 360) sensor was developed by Microsoft Corporation (Redmond, Washington, USA) as a gaming device that connected with the popular Xbox console. However, with slight alterations to the power supply, the Kinect becomes a suitable low-cost RGB-D sensor (Bengochea-Guevara et al. 2018). The Kinect V1 measures an image's depth by projecting an infrared (IR) pattern out of the IR sensor (Figure 2.2) and recording the distortion returned to the sensor, enabling the depth to be calculated (Wasenmüller & Stricker 2016). The sensor is extremely sensitive to the environment's lighting conditions, as solar irradiance interferes with the IR signal, thus distorting and limiting the sensor's depth capabilities (Marinello et al. 2017). Various software platforms enable the Kinect V1 to capture 3-D models of the object. For example, the Kinect Fusion software (included in Microsoft's Software Development Kit 1.8 for Windows (Microsoft 2013)) captures a mesh of the object, while the Real-Time Appearance-Based Mapping (RTAB-Map) (Labbe 2018) software can capture a point cloud of the object. Both software suites result in a 3-D model of the object, capable of determining the object's volume when deploying CV techniques.



Source: Marinello et al. (2016: 877)

Figure 2.2 Schematic of a Kinect V1 sensor.

To date, Marinello et al. (2016) are the only authors to operate the Kinect V1 as a 3-D (RGB-D) PRS sensor for data acquisition in a table grape vineyard, specifically for yield estimation. The authors investigated the optimal distance and angle for an RGB-D sensor to be mounted for estimating bunch mass in a table grape vineyard. They concluded that a side-on perpendicular acquisition angle with a

distance of approximately 1 m provided the best results, attaining error deviations between 10–15% for yield estimates. The study was presented with extremely limited detail regarding the experiment's methodology, prohibiting a solid interpretation of the research and how it was conducted, and thus questioning the validity of the results achieved.

Alternatively, horticulture studies utilising a Kinect sensor have been conducted by Andujar et al. (2016) and Wang and Li (2014), estimating cauliflower and sweet onion yields, respectively. Andujar et al. (2016) acquired 3-D point clouds of cauliflowers under controlled laboratory conditions. Following data acquisition, the point clouds were filtered and cleaned before applying a *Screened Poisson Surface Reconstruction* (Kazhdan & Hoppe 2013) to enclose the point cloud. The *Screened Poisson Surface Reconstruction* expands on the *Poisson Surface Reconstruction* presented by Kazhdan, Bolitho and Hoppe (2006), where watertight surfaces are created from all points within a spatially oriented point set, cast as a spatial Poisson problem. The *Screened Poisson Surface Reconstruction* (Kazhdan & Hoppe 2013) can also be applied to mesh datasets (refer to Kazhdan, Bolitho and Hoppe (2006) and Kazhdan and Hoppe (2013) for further detail). Once the 3-D model (point cloud or mesh) is 'watertight', the volume can be extracted from the model and used to estimate the yield (Andújar et al. 2016). The authors were able to achieve an r^2 of 0.868 for yield estimation – computed from linear regression between estimated volume (cm^3) and actual mass (g) (Andújar et al. 2016). Similarly, Wang and Li (2014) captured 3-D point clouds of sweet onions and successfully regressed the estimated volume (cm^3) against the known volume (cm^3), achieving an estimation accuracy of 0.953.

Evidently, the use of the Kinect as an RGB-D sensor presents a cost-effective option to accurately estimate the volume of a fruit or vegetable. These high results were obtained under laboratory conditions, with controlled lighting (Wang & Li 2014). The interference of solar irradiance with the Kinect is a potential limitation to in-situ data acquisition (Andújar et al. 2016), but this is yet to be investigated with respect to estimating vineyard yield.

2.3.2 2-D PRS with 3-D CV techniques for yield estimation

The advancement of CV techniques has enabled the 3-D processing of yield estimates from standard RGB (2-D) digital images (Ivorra et al. 2015). 2-D PRS data acquisition can employ a stereo camera, whereby two RGB images are captured from two separate lenses within the camera (Ivorra et al. 2015). The slight variation in viewing angle from the two lenses combined with the internal camera calibrations enables the 3-D reconstruction of the object during data processing (Klodt et al. 2015). Ivorra et al. (2015) successfully estimated total berry mass when using 3-D CV techniques to

reconstruct the bunches imaged with a stereo camera under laboratory conditions. The authors achieved an r^2 value of 0.820 for bunch volume estimation and an r^2 of 0.830 for berry mass estimation. The following year saw Rose et al. (2016) employ a five-camera system to capture in-situ stereo images of grape vines. The authors modelled various aspects of the vines, yielding metrics capable of yield estimation. However, no yield estimates were computed.

A similar approach to yield estimation was presented by Herrero-Huerta et al. (2015). In this case, the authors used a single-lensed digital camera to acquire multiple images of the object from varying angles. Owing to the various angles, the 3-D geometry of the scene was reconstructed with specialised CV techniques (Herrero-Huerta et al. 2015), resulting in a 3-D point cloud of the object. This 3-D reconstruction technique is similar to that of Structure-from-Motion (SfM), exemplified by different applications (Díaz et al. 2018; Jay et al. 2015). Utilising this 3-D CV technique, Herrero-Huerta et al. (2015) achieved a bunch mass r^2 value of 0.778 during yield estimation calculations. This result was slightly lower than the r^2 obtained by Ivorra et al. (2015), but data acquisition occurred in situ (Herrero-Huerta et al. 2015).

While limited research has used 3-D CV techniques for estimating yield, several studies (Ivorra et al. 2015; Wang & Li 2014) have demonstrated the potential. Advances in 3-D CV techniques in the near future could yield more research in this domain, while improved PRS technology is required to overcome said limitations of in-situ data acquisition (Andújar et al. 2016). 3-D techniques have displayed the potential to challenge 2-D techniques for estimating vineyard yield, where a direct comparison could prove fruitful in this regard.

2.4 LITERATURE SUMMARY

Remote sensing has been successfully implemented in precision viticulture for several decades (Matese & Di Gennaro 2015; Mulla 2013). Remote sensing applications and sensor platforms have varied over the years, with this review focusing on yield estimation – enabling a better understanding of the inherent spatial variability within vineyards (Hall et al. 2011). Traditional methods are limited by their destructive, labour-intensive and time-consuming nature (De la Fuente et al. 2015). Remote sensing overcomes these limitations, with research investigating spaceborne (Cunha, Marçal & Silva 2010), airborne (Hall et al. 2011), and terrestrial (Millan et al. 2018) platforms for estimating vineyard yield at various scales, from vineyard-level down to plant-, bunch- and even berry-level. Unfortunately, spaceborne and airborne platforms can be limited by adverse weather conditions, especially cloud cover, and acquiring high spatial resolution data can be an expensive feat (Hall et al. 2002).

In recent years, the literature has focused on PRS and CV techniques for estimating vineyard yield and is considered a 'hot topic' in the precision viticulture domain (Liu, Whitty & Cossell 2015b). PRS techniques employ terrestrial sensors for data acquisition, capturing either 2-D or 3-D datasets of the vine's fruit – i.e. the bunches (Herrero-Huerta et al. 2015; Millan et al. 2018). Digital cameras that acquire 2-D RGB imagery (Aquino et al. 2018) are the most popular sensor choice, with an RGB-D sensor (Marinello et al. 2016) being an alternative for 3-D data acquisition. Subsequently, CV techniques are used for data processing, yielding a suitable metric (such as 'pixel count' or 'berry count') from the 2-D data for estimating the yield (Liu, Marden & Whitty 2013). On the contrary, 3-D data is volumetric, and the extracted bunch volume is directly used for estimating yield (Marinello et al. 2016). CV techniques are manual and semi-automated processes that make use of supervised learning techniques, striving towards automation.

This review of the literature has identified the lack of research utilising 3-D PRS techniques for data acquisition; for instance, acquiring data with an RGB-D sensor. Furthermore, the literature lacks a direct operational comparison between 2-D and 3-D methodologies. Regarding CV techniques for yield estimation, potential research could investigate unsupervised techniques for bunch and berry detection, removing the human aspect associated with existing techniques. Nonetheless, the future of PRS and CV techniques for vineyard yield estimation is a prosperous one. Commercial implementation of such techniques provides the opportunity to revolutionise the viticulture industry, and entire agriculture industry, maximising crop outputs (quantity and quality) in an economical and environmentally sustainable manner.

CHAPTER 3: INVESTIGATING 2-D AND 3-D PROXIMAL REMOTE SENSING TECHNIQUES FOR VINEYARD YIELD ESTIMATION

Hacking C, Poona N, Manzan N & Poblete-Echeverría C 2019. Investigating 2-D and 3-D Proximal Remote Sensing Techniques for Vineyard Yield Estimation. *Sensors* 19, 17: 3652.

3.1 ABSTRACT

Vineyard yield estimation provides the winegrower with insightful information regarding the expected yield. This facilitates managerial decisions to achieve maximum quantity and quality, whilst assisting the winery with planification. The use of PRS technology and CV techniques for yield estimation has produced limited success within viticulture. In this study, 2-D RGB and 3-D RGB-D (Kinect) imagery was investigated for yield estimation in a VSP trellised vineyard. The research presents three experiments; including two measurement levels and two canopy treatments. RGB imagery (bunch- and plant-level) underwent image segmentation before estimating fruit area using a calibrated pixel area. RGB-D imagery captured at bunch-level (mesh) and plant-level (point cloud) was reconstructed for fruit volume estimation. The RGB and RGB-D measurements utilised cross-validation to determine fruit mass, which was subsequently used for yield estimation. Experiment one's (laboratory conditions) bunch-level results achieved a high yield estimation agreement with RGB-D imagery (r^2 = 0.950), which outperformed RGB imagery (r^2 = 0.889). Both RGB and RGB-D performed similarly in experiment two (bunch-level), whilst RGB outperformed RGB-D in experiment three (plant-level). The RGB-D sensor (Kinect) is suited to ideal laboratory conditions, whilst the robust RGB methodology is suitable for both laboratory and in-situ yield estimation.

3.2 INTRODUCTION

Modern day viticulture has seen an increase in the use of robust scientific methods combined with new technologies to improve overall production (Arnó et al. 2009). Precision farming – a direct result of the modernisation in farming – can be discipline-specific, i.e. specific in horticulture or viticulture. Precision viticulture aims to effectively manage production inputs to improve yield and grape quality, while reducing the environmental impact of farming (Blackmore 2003). The use of remote sensing technology in precision viticulture allows variability to be monitored at vineyard level, per individual block, or on a vine basis. Aspects such as vine shape, size and vigour can be observed, providing more accurate yield and fruit quality information (Matese & Di Gennaro 2015).

Yield estimation provides information to the winegrower that can be used to manage the vineyard, optimising quality and yield (Nuske et al. 2014). Awareness of the estimated yield allows the vineyard manager to manipulate the vines to obtain the desired grape characteristics and have a better planification for the winery to use during the wine-making process (Cunha, Marçal & Silva 2010). Accurate yield forecasting assists with logistical planning both during and post-harvest; for example what volume will be harvested, where the grapes will be stored, as well as an expected market price (Cunha, Marçal & Silva 2010).

Wolpert and Vilas (1992) outlined a two-step method for vineyard yield estimation. The start of the process determines the number of clusters situated on individual vines early in the season. Subsequent determination of cluster weights occurs at vèraison. Unfortunately, the two-step method is labour-intensive, error prone, and destructive in the estimation process. To overcome the limitations of the two-step method, modern techniques have employed sensors attached to automatic harvesters to monitor yield during the harvesting process (Matese & Di Gennaro 2015). Yield estimation prior to harvest is becoming possible with increasing accuracies when making use of non-invasive PRS and CV techniques (Dunn & Martin 2004; Liu, Marden & Whitty 2013; Marinello et al. 2016; Nuske et al. 2011).

A common PRS approach employs 2-D RGB imagery captured with a digital camera for yield estimation (for example, Diago et al. (2012); Dunn and Martin (2004); and Liu, Marden and Whitty (2013)). The 2-D approach can be generalised into two steps, namely image segmentation – to differentiate the bunch from the background, and yield estimation – using a suitable bunch metric, i.e. the pixel count of bunch area in an image. For example, Diago et al. (2012) used image classification (for segmentation) and a bunch metric to classify ‘background noise’ and ‘grape’ classes. The authors achieved a testing r^2 of 0.730. An alternative segmentation approach to image classification uses RGB CT to differentiate grapes from the background. Dunn and Martin (2004) presented an RGB CT technique that applied specific thresholds to the colour properties of an RGB image, generating a binary image; background and grapes. The authors achieved an r^2 value of 0.720. A similar thresholding approach was adopted by Liu, Marden and Whitty (2013) and Font et al. (2015). Specifically, Liu, Marden and Whitty (2013) introduced a bunch-level experiment under laboratory conditions with manual CT for image segmentation, which resulted in a yield estimation r^2 of 0.770. Additionally, the authors presented a more complex automatic process for image segmentation on the same dataset presented by Dunn and Martin (2004), resulting in an improved r^2 of 0.860 (Liu, Marden & Whitty 2013).

The second step, yield estimation, depends on the selected bunch metric. A simple pixel count of the segmented bunches is a favoured metric (Diago et al. 2012; Dunn & Martin 2004), with adaptions to the pixel count presented by Font et al. (2015) and Liu, Marden and Whitty (2013). Liu, Marden and Whitty (2013) tested five metrics, namely i) volume, ii) pixel count, iii) perimeter, iv) berry number, and v) berry size. Yield estimation using the pixel count produced superior results over the remaining metrics (Liu, Marden & Whitty 2013). Contrary to Liu, Marden and Whitty (2013), Nuske et al. (2011) avoided image segmentation and used a berry detection algorithm to determine a berry count. The use of the berry count as a yield estimation metric provided an r^2 of 0.740. Subsequent work on multiple multi-temporal datasets produced r^2 values between 0.600–0.730 (Nuske et al. 2014).

A less common PRS approach utilises an RGB-D camera to capture a 3-D model in either mesh or point cloud format, representing the bunch or vine in a 3-D coordinate system (Bengochea-Guevara et al. 2018). The use of the Microsoft Kinect™ represents an ideal low cost RGB-D sensor for in-situ imaging of vines (Bengochea-Guevara et al. 2018). The resulting 3-D models can be used to extract volumetric measurements for yield estimation. A limited number of studies have investigated the utility of the Kinect sensor for volume estimation. For example, Wang and Li (2014) and Andújar et al. (2016) employed the Kinect sensor under laboratory conditions for volume estimation of sweet onions and cauliflowers, respectively. To date, the only use of an RGB-D Kinect sensor for yield estimation within viticulture was presented by Marinello et al. (2016). The authors assessed sensor position for volume estimation of table grape clusters by testing two viewing angles, side-on and bottom-up. Multiple sensor-target distances were tested for the side-on viewing angle. Marinello et al. (2016) concluded that a side-on viewing angle with a sensor-target distance of 0.8–1.0 m generated the best results.

PRS (RGB and RGB-D) and related CV techniques incorporated into a suitable methodology present a viable solution for vineyard yield estimation. The established use of RGB imagery is evident, while the novel use of RGB-D imagery shows promise for future yield estimation. However, to date, no study has investigated 2-D and 3-D methodologies side-by-side for vineyard yield estimation. Investigating these two methodologies in a commercial vineyard could provide insight to their capabilities and operational potential.

A key aspect for consideration when implementing these methodologies within a vineyard is the canopy coverage. The combination of essential canopy management practices and the vineyard's trellis system – particularly a VSP system – directly influences canopy coverage, and inevitably the success of PRS and related CV techniques for yield estimation. High canopy coverage in the bunch

zone results in bunch occlusion from the sensor. The incorporation of a canopy treatment was therefore proposed for this study.

The aim of this study was to investigate 2-D (RGB) and 3-D (RGB-D) methodologies for yield estimation in a VSP vineyard by means of bunch area/volume estimation. The study was undertaken as three experiments, occurring under laboratory and field conditions. Field conditions were conducted at both bunch- and plant-level. Two canopy treatments occurred from direct canopy manipulation and were defined as FC and LR. Hypothetically, the LR treatment will produce better estimation results. Furthermore, to achieve the aim of the study, two objectives were determined: to develop independent 2-D and 3-D yield estimation methodologies, and to analyse and compare the success of the two methodologies for yield estimation.

3.3 MATERIALS AND METHODS

3.3.1 Study site

The study was carried out at the end of the 2016/17 growing season in a drip-irrigated Shiraz vineyard at the Welgevallen Experimental farm located in Stellenbosch, South Africa (33 56’ 26” S; 18 51’ 56” E). The vineyard was planted in the year 2000 with a grapevine spacing of 2.7×1.5 m in a North-South orientation and lies approximately 157 m above sea level. A seven-wire hedge VSP trellis system with three sets of moveable canopy wires is used in the vineyard. The Stellenbosch area falls within the coastal wine grape region of the Western Cape, which is characterised by a Mediterranean climate with long, dry summers (Conradie et al. 2002).

A total of 31 individual vines were selected across three rows and sampled for this study (Figure 3.1).

Figure 3.1 Location of the Shiraz vineyard in Stellenbosch, South Africa. Inset map (red rectangle) shows the three rows used for data collection.

3.3.2 Data acquisition

Data was acquired between 28 February and 3 March 2017 (harvest), where data collected in situ conformed with the two canopy treatments. The purpose behind the canopy treatments was gaining direct line of sight to the bunches, which were generally hidden by the vine's canopy. No manipulation of the canopy, otherwise normal canopy conditions was classified as the FC treatment. The alternative LR treatment occurs after manual manipulation of the canopy, resulting in full leaf removal in the bunch zone, effectively displaying the bunches.

Figure 3.2 illustrates the data acquisition process that resulted in a total of ten datasets, as outlined by each step:

1. RGB and RGB-D imagery acquired for the FC treatment taken at bunch-level (n = 21; randomly selected and labelled bunches from the 31 vines) and plant-level (n = 31; individual vines). The resulting datasets (n = 4) included RGB: bunch; RGB-D: bunch; RGB: vine; and RGB-D: vine.
2. Manual manipulation of the canopy resulted in the environment for the LR treatment datasets (n = 4) to be captured. These datasets were identical to the FC datasets and concluded the in-situ datasets (n = 8). At this point the 31 vines were harvested.

3. RGB and RGB-D imagery of the harvested bunches (n = 21) captured under laboratory conditions resulted in the final two datasets. At this point, all datasets (n = 10) have been captured.
4. Reference measurements, captured under laboratory conditions, included mass (g) and displacement (ml) for the individual bunches (n = 21) and individual vines (n = 31).

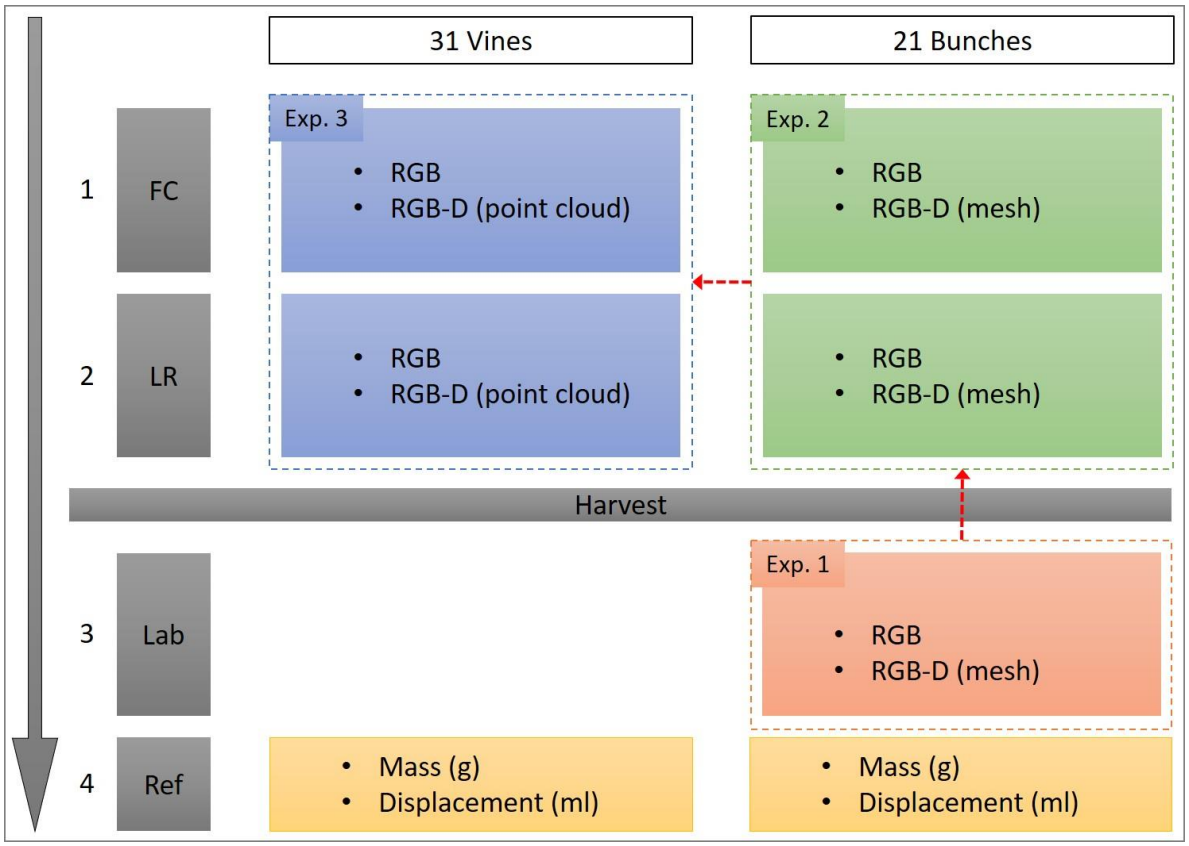


Figure 3.2 Data acquisition protocol used in this study. Order of acquisition indicated by the grey arrow. *{Key: FC = full canopy; LR = leaf removal; Lab = laboratory; Ref = reference measurements; Exp = experiment.}*

RGB and RGB-D images were captured by two PRS sensors. A Nikon D3200 digital single-lens reflex camera was used for capturing 24.2-megapixel RGB images. The camera captured images in *auto* mode with the flash disabled. The second sensor, a Microsoft Kinect™ V1 was used to capture RGB-D imagery as either a mesh (bunch-level) or a point cloud (plant-level). The following data acquisition subsections provide experiment-specific details.

3.3.2.1 Reference measurements

Reference measurements were collected under laboratory conditions for the 21 individual bunches and the 31 individual vines. Individual bunch mass (g) was recorded with a Mentor scale (*Ohaus*, Parsippany, NJ, USA), and individual vine mass (g) was recorded with a Viper SW scale (*Mettler Toledo*, Columbus, OH, USA). Bunch/vine volume measurements were recorded as the displacement

(ml) of water when bunches were submerged in a container of water (Ferreira & Marais 1987). The mass and volume measurements were used as reference measurements for the estimated measurements derived from the two methodologies.

3.3.2.2 Experiment one: Individual bunches under laboratory conditions

The 21 individual bunches were imaged using the RGB camera and Kinect sensor under laboratory conditions, i.e. laboratory illuminated with white fluorescent lights and natural light entering through the windows. Each bunch was suspended from a tripod against a green background to maximise image contrast (Figure 3.3A).

a) *RGB imagery:* The camera was placed parallel to the suspended bunch at a distance of 60 cm. A single image per bunch was captured for image processing. A ruler was included in each image for length reference (Figure 3.3B).

b) *RGB-D (Kinect) imagery:* The Kinect sensor was placed 60 cm proximal to the target and captured individual meshes per bunch; resulting in a solid 3-D model of the bunch. The Kinect sensor coupled with the Kinect Fusion software (part of Microsoft's Software Development Kit 1.8 for Windows (Microsoft 2013)) running on a laptop can capture individual meshes. Meshes were captured in *.stl* format with a resolution of 640 voxels/m, and a voxel resolution of 256x256x256 voxels. A white surface was positioned directly behind the bunch to improve contrast and depth determination; as illustrated by a captured mesh seen in Figure 3.3C. During mesh capture, the entire bunch system was rotated, providing different angles of view.

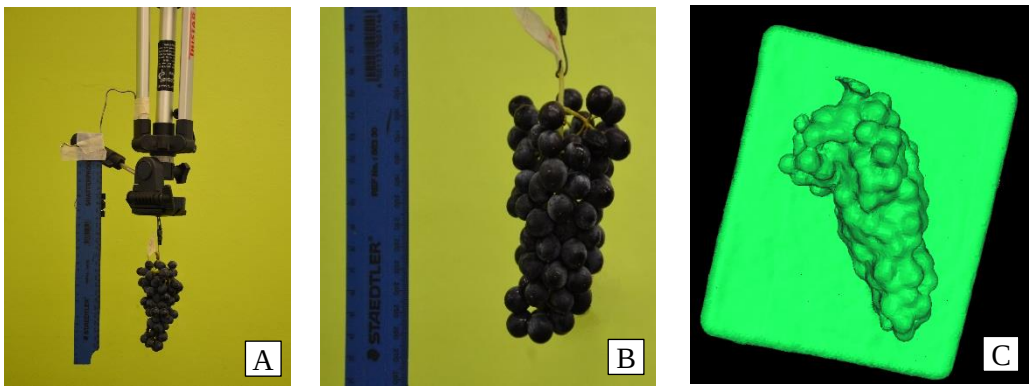


Figure 3.3 Data acquisition under laboratory conditions. (A) Experimental setup for image capture; (B) RGB image of an individual bunch with a ruler for reference length; and (C) RGB-D (Kinect mesh) of an individual bunch.

3.3.2.3 Experiment two: Individual bunches in field conditions

Images of the same 21 individual bunches were captured in in-situ conditions under both canopy treatments with the same RGB and RGB-D proximal sensors as used in experiment one. Here individual bunches were still attached to the respective vines.

a) *RGB imagery:* The camera captured images of the individual bunches for both FC (Figure 3.4A) and LR (Figure 3.4B) treatments. The camera was positioned approximately 40 cm from the bunch being imaged, maintaining the reference length (ruler) in each image. Image acquisition occurred between 12:00 and 13:00, under natural solar illumination.

b) *RGB-D (Kinect) imagery:* The same software setup as in experiment one was used, which allowed the Kinect to capture an individual mesh per bunch for FC (Figure 3.4C) and LR (Figure 3.4D) treatments. Imagery was captured after sunset (approximately 20:00) with artificial illumination. The Kinect was held approximately 60 cm from the bunch, moving the Kinect around the bunch axis by hand. For the LR treatment a board was placed behind the bunches, thereby creating an artificial background to improve volume extraction (Figure 3.4D).

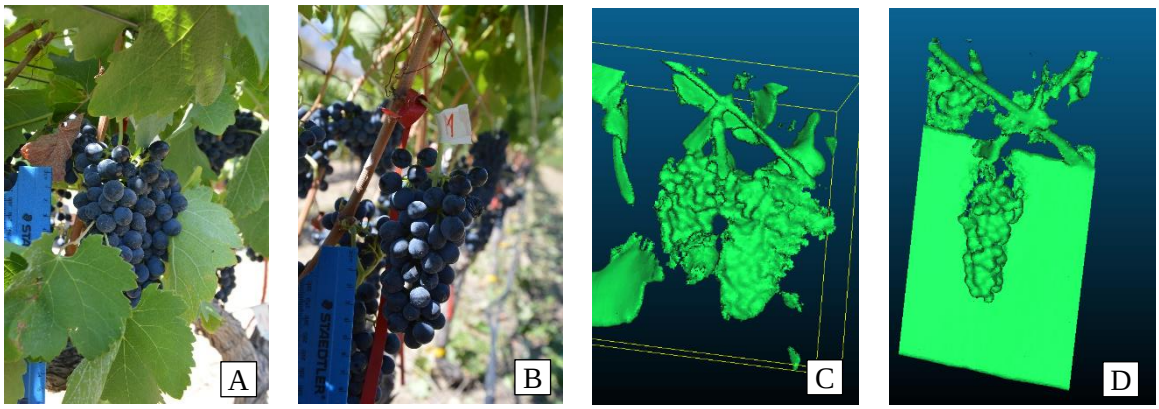


Figure 3.4 Data acquisition of individual bunches in field. (A) RGB image with FC; (B) RGB image with LR; (C) RGB-D (Kinect mesh) with FC; and (D) RGB-D (Kinect mesh) with LR.

3.3.2.4 Experiment three: Individual vines in field conditions

The same RGB and RGB-D sensors were used to capture in-situ images at plant-level, as presented in this experiment. This saw 31 individual vines imaged for both FC and LR treatments.

a) *RGB imagery:* Imagery was captured at 12:00, under natural solar illumination. The camera was positioned 2 m from the vine capturing an image for each side of the canopy, and repeated for both canopy treatments (FC: Figure 3.5A; LR: Figure 3.5B).

b) *RGB-D (Kinect) imagery:* The Kinect sensor captured point clouds – instead of meshes –of the 31 vines, due to the scale difference. Point clouds consisted of thousands of individual points to create 3-D models; for both the FC and LR treatments (Figure 3.5C,D). The RTAB-Map software (Labbe 2018) was used for point cloud modelling and regeneration. In RTAB-Map (Labbe 2018), the default filtering parameter and a 3-D cloud decimation ('thinning' of the point cloud) value of 2 were used during exportation of the point clouds in *.ply* format. The point cloud was captured per individual row, repeated for both sides of the canopy. The Kinect sensor was hand-held approximately 2 m from the vines at a perpendicular angle and was moved in a north to south direction. Imagery was collected at 19:00, immediately prior to sunset, using the last natural illumination of the day.

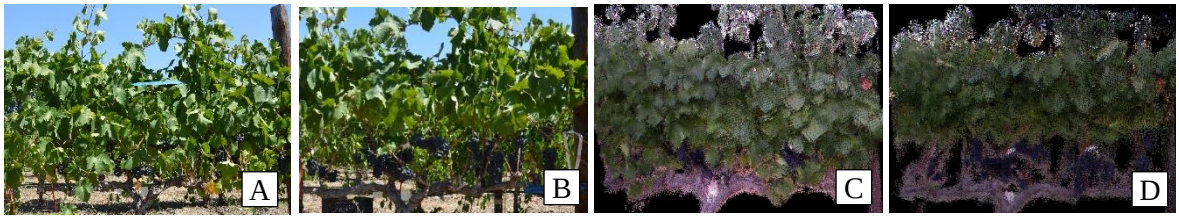


Figure 3.5 Experiment three data examples at plant-level. RGB imagery of FC (A) and LR (B) treatments. RGB-D (Kinect point cloud) of FC (C) and LR (D) treatments.

3.3.3 Data analysis

Data pre-processing and analysis occurred sequentially from experiment one (indicated via red arrows in Figure 3.2). The canopy treatments existing in experiment two and three had no effect on how the data was analysed for yield estimation. The proposed RGB and RGB-D methodologies were created on the LR datasets, before being directly applied to the FC datasets.

3.3.3.1 RGB imagery

RGB images were processed using a custom script in MATLAB® (The MathWorks Inc. 2018) as follows:

1. The reference length (obtained from the ruler) in each image was used to scale the image pixels, creating a calibration value in cm^2.
2. Manual selection of the region of interest (ROI) containing the relevant bunch/bunches. ROIs were strategically digitised so as to capture minimum background.
3. The masked RGB images were then converted to the HSV colour space and segmented using MATLAB's (The MathWorks Inc. 2018) *Colour Thresholder* app, part of the *Image Processing Toolbox*™. It was visually evident when selecting the threshold values that the lighting

conditions influenced the values. Therefore, separate threshold values were determined for the respective experiments. Threshold values were computed using a random sample from the specific dataset; equivalent to 25% of the experiment's dataset.

4. After image segmentation, the number of segmented pixels were determined (adaption of the pixel count metric (Liu, Marden & Whitty 2013)) and converted into pixel area (cm^2) using the calibration value.

Figure 3.6 illustrates the image segmentation process for experiment one at bunch-level (Figure 3.6A,B), and experiment three at plant-level (Figure 3.6C,D). For experiment three, segmentation produced a segmented bunch area image (Figure 3.6D) for a single side of the vine, with a similar image for the reverse side of the vine. To obtain a single area value per vine, the Total Bunch Area of Vine (*TBAV*) was calculated using Equation 3.1:

$$TBAV = (Ae + Aw)/2 \qquad \text{Equation 3.1}$$

where *Ae* was the area of the east-facing side of the vine, and *Aw* represents the vine's west-facing area. The *TBAV* was in cm^2.

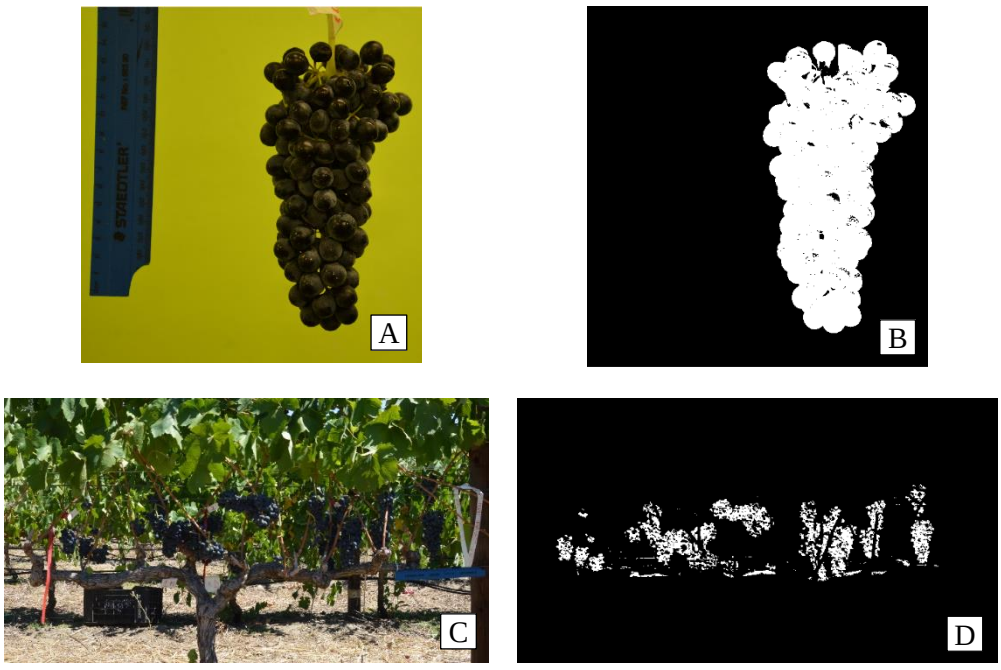


Figure 3.6 (A) Represents the original RGB image, with (B) illustrating the segmented binary image at bunch-level. (C) An RGB image of an east-facing vine, with (D) the segmented binary image at plant-level.

3.3.3.2 RGB-D (Kinect) imagery

Due to the two image data types captured – mesh vs. point cloud – RGB-D imagery was processed differently for experiment one and two (mesh: bunch-level), and experiment three (point cloud: plant-level).

Data processing for experiment one and two progressed as follows:

1. Each mesh was manually cleaned and sectioned to remove any background, using the *Cross-Section* tool in CloudCompare (CloudCompare 2018).
2. The cleaned mesh was reconstructed in MeshLab (Cignoni et al. 2008) using the *Screened Poisson Surface Reconstruction* (Kazhdan & Hoppe 2013) method – see Figure 3.7. Additional cleaning to the mesh was completed ad hoc in MeshLab (Cignoni et al. 2008).
3. Volume (cm^3) of the mesh was calculated in CloudCompare (CloudCompare 2018).

The *Screened Poisson Surface Reconstruction* (Kazhdan & Hoppe 2013) allows the back of the mesh, which was 'open' (Figure 3.7A), to be reconstructed, becoming 'watertight' (as seen in Figure 3.7B).

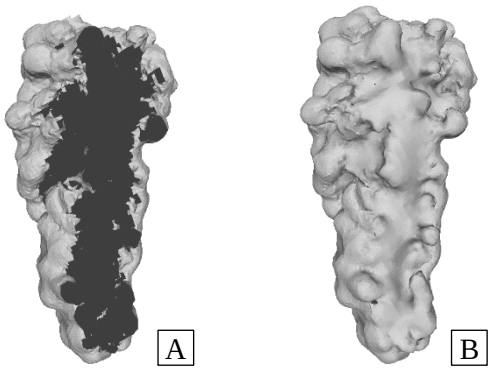


Figure 3.7 (A) Example of a mesh prior to reconstruction, and (B) the same mesh after Poisson reconstruction.

The nature of the point cloud data requires different processing steps to that of the mesh data. This was largely attributed to the necessity of closing the points of the point cloud, thereby producing a 'watertight mesh' for volume extraction.

The point cloud data of experiment three was processed as follows:

1. Point clouds were imported into CloudCompare (CloudCompare 2018), and subsequently cleaned and sectioned to individual vines, focusing on the bunch zone.
2. Bunches were segmented from the point cloud using their colour properties. CloudCompare's (CloudCompare 2018) *Filter Points by Value* tool incorporates user-defined thresholds,

manipulating the RGB colour space of the point cloud. Threshold values were determined on a random sample (25%) of the LR dataset.

3. A custom-built script in R statistical software (R Core Team 2019) was used for calculating the segmented point cloud's volume, representing the vine's bunches.

Figure 3.8 illustrates the raw point cloud (Figure 3.8A) captured by the Kinect sensor, with the segmented point cloud (Figure 3.8B) displaying the bunches.

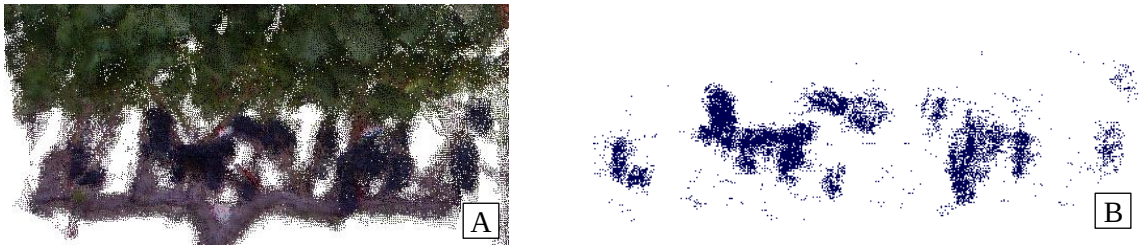


Figure 3.8 (A) The Kinect point cloud for a LR treatment vine (east side) and (B) the segmented point cloud of the same vine.

The custom script in R statistical software (R Core Team 2019) incorporated the *alphashape3d* package v1.3 (Lafarge & Pateiro-Lopez 2017) to compute the 3-D shape for volume calculation purposes. The *alphashape3d* (Lafarge & Pateiro-Lopez 2017) package incorporates the *α-shape* algorithm (Lafarge et al. 2014) to recover the geometric structure of the 3-D point cloud for volume calculation. The *α-shape* algorithm (Lafarge et al. 2014) requires a specific alpha value for computation; hence, the alpha value directly influences the total volume calculated. To adjust an alpha value for the experimental conditions, three levels of alpha were tested on 25% of the dataset and linearly regressed against the reference values. Similar investigative experiments were conducted by Rueda-Ayala et al. (2019) and Ribeiro et al. (2017) to determine experiment-specific alpha values.

The process described above for experiment three was repeated for both sides of the vine, resulting in two volume measurements per vine. A single volume value per vine was obtained via the Total Bunch Volume of Vine (*TBVV*) calculation, Equation 3.2:

$$TBVV = (Ve + Vw)/2 \qquad \text{Equation 3.2}$$

where *Ve* was the volume of the vine's east-facing side and *Vw* was the volume of the west-facing side. The resultant *TBVV* was in cm^3 per vine.

3.3.4 Cross-validation

Five-fold cross-validation was used to develop the yield estimation model for each dataset. Cross-validation was implemented using the *Caret* package (Kuhn 2008) in R statistical software (R Core Team 2019), repeated ten times for model robustness. The model produced 'fitted values', which represented the estimated yield (in grams). Following this, the estimated yield (g) values were linearly regressed against the actual mass (g) to produce a final r^2 value (coefficient of determination), indicating the potential for yield estimation. The RMSE was computed from the linear regression, indicating the yield estimation error (in grams).

3.4 RESULTS

3.4.1 Reference measurements

Results of the reference measurements indicated a strong relationship between mass and volume at bunch-level ($r^2 = 0.971$) and plant-level ($r^2 = 0.996$). The established relationships between mass and volume served as the basis for the subsequent experiments, which were used to evaluate the 2-D and 3-D methodologies.

3.4.2 Pre-processing

The complexity of the RGB-D (Kinect) datasets required two additional pre-processing steps. The first step was the determination of an alpha value required for volume calculation using the *alphashape3d* package (Lafarge & Pateiro-Lopez 2017). The second step was volume correction for all Kinect datasets due to volume estimation errors in the datasets.

3.4.2.1 *alphashape3d's* adjusted alpha value

Table 3.1 summarises the three alpha values tested, with Figure 3.9 illustrating the output based on the alpha values. An alpha value of 0.01 (Figure 3.9C) produced the highest coefficient of determination ($r^2 = 0.605$). An alpha value lower than 0.01 underestimated volume (Figure 3.9B), whereas alpha values higher than 0.01 over-estimated volume (Figure 3.9D). An alpha value of 0.01 was subsequently used for all further analyses.

Table 3.1 Relevant alpha values tested for the *alphashape3d* package in the custom R script.

Alpha	r^2	Estimated volume (cm^3)	Actual volume (cm^3)
0.005	0.520	1 017	15 695
0.010	0.605	13 412	15 695
0.050	0.506	120 077	15 695

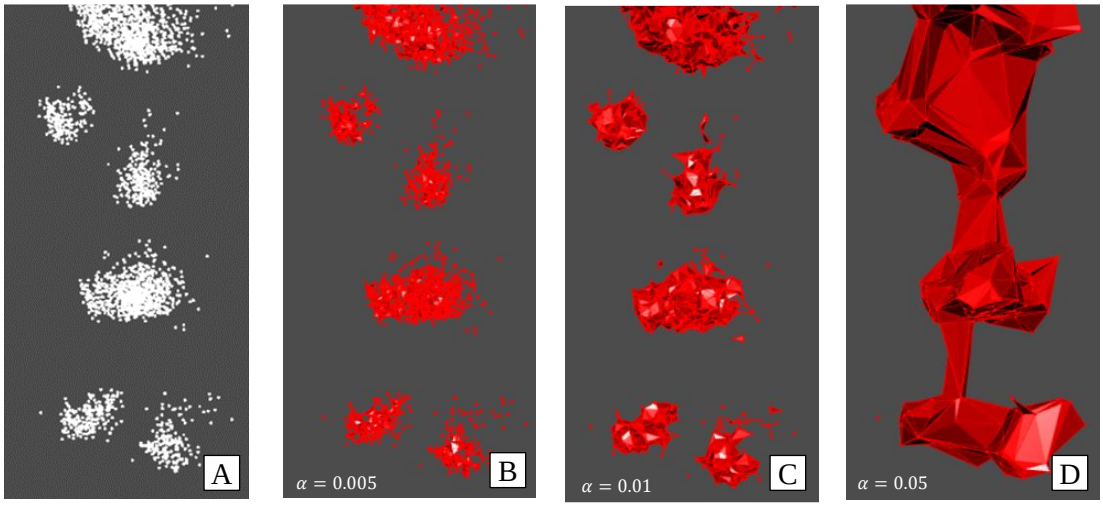


Figure 3.9 Point cloud reconstruction testing the alpha value for the *alphashade3d* package. The original point cloud before reconstruction (A), and after reconstruction (B), (C), and (D).

3.4.2.2 Kinect volume correction

Figure 3.10 shows a volume estimation error present in experiment one's data, which aligns with a subsequent review of the literature (Andújar et al. 2016; Marinello et al. 2016; Wang & Li 2014). The 21 bunches have a mean actual volume (the reference volume) of 144.952 cm^3 and a mean estimated Kinect volume of 175.672 cm^3. Overestimation by the Kinect V1 sensor was evident from the results. Volume correction via cross-validation was therefore subsequently incorporated into the methodology; where a mean corrected volume of 144.952 cm^3 was achieved. Thereafter, the correction was applied to the remaining Kinect datasets.

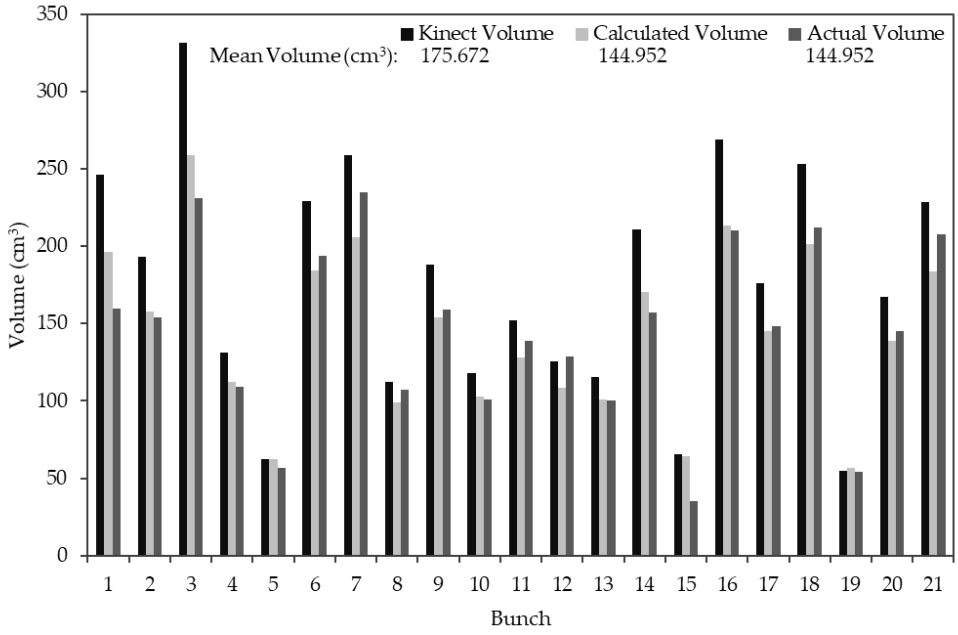


Figure 3.10 Experiment one's results, for the 21 individual bunches, illustrating the volume estimation error by the Kinect sensor.

3.4.3 RGB results

Figure 3.11 illustrates the results for the three experiments that used 2-D RGB digital imagery. The best results were obtained in experiment one (Figure 3.11A), which produced an r^2 of 0.889 and an RMSE of 17.987 g. The level of accuracy achieved in experiment one can be contributed to the controlled laboratory conditions and supports the proposed methodology for yield estimation. Applying this methodology to in-situ bunches produced less accurate results, as seen in experiment two. Experiment two's FC treatment (Figure 3.11B) produced the lowest yield estimation results for 2-D RGB imagery, with an r^2 of 0.625 and an RMSE of 27.738 g. The LR treatment's (Figure 3.11C) r^2 and RMSE values were 0.742 and 25.066 g, respectively. The lesser FC values were directly attributed to the canopy coverage present, as the bunches were partially occluded from the sensor's view. The effect of the canopy treatment was evident when comparing the results. At plant-level (experiment three), the same pattern was present between the canopy treatments. The FC (Figure 3.11D) treatment of experiment three produced an r^2 of 0.779, while the LR (Figure 3.11E) treatment produced an even higher r^2 of 0.877. The respective RMSE values were 559.357 g and 443.235 g. The success of yield estimation in experiment three, specifically the LR treatment, supports the methodology's capability for 2-D RGB yield estimation. The in-situ yield estimation was more preferable at plant-level than that of bunch-level, which may be attributed to the lighting conditions, and the success of the CT technique for bunch segmentation at plant-level.

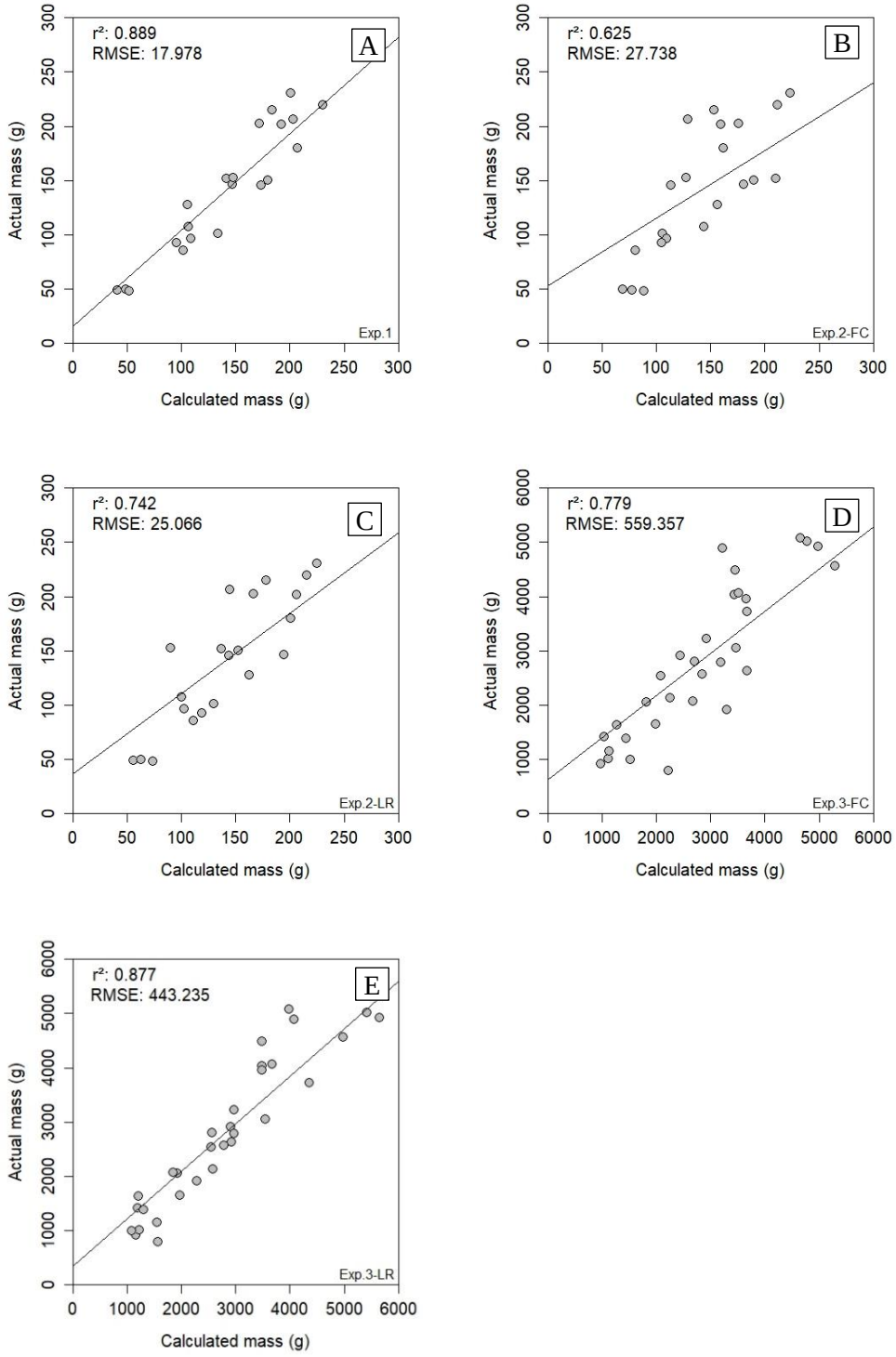


Figure 3.11 RGB results presented for the three experiments; experiment one (A), experiment two FC (B) and LR (C), and experiment three FC (D) and LR (E). *{Key: Exp. – experiment; FC – full canopy; LR – leaf removal.}*

3.4.4 RGB-D results

Figure 3.12 illustrates the RGB-D results obtained for the three experiments. The unrivalled results obtained in experiment one (Figure 3.12A) produced an r^2 of 0.95 and an RMSE of 12.458 g – the best performing results presented in this study. The Kinect sensor favoured the controlled conditions of the laboratory, specifically the artificial illumination as a source of light. Applying the same methodology to experiment two (in-situ bunches) resulted in a lower yield estimation performance for both canopy treatments. The FC treatment (Figure 3.12B) produced an abnormally low r^2 of 0.0203, with an RMSE of 8.081 g. Upon analysing the results, a statistical outlier in the data was noticed. With the removal of this outlier, the modified FC results (n = 20) improved drastically, with a new r^2 of 0.609 and an RMSE of 26.79 g (Figure 3.12C). In contrast, the LR treatment (Figure 3.12D) resulted in an r^2 of 0.756 and an RMSE of 24.601 g, which aligned with the LR results for bunch-level obtained with RGB imagery (Figure 3.11C). At plant-level, the unfavourable results of the FC treatment (Figure 3.12E) generated an r^2 of 0.487 and an RMSE of 673.535 g. However, the LR treatment (Figure 3.12F) provided some promise for the Kinect sensor at plant-level, achieving an r^2 of 0.623 and an RMSE of 652.959 g. The same effect of the canopy treatment was evident in the RGB-D results as in the RGB results, with the LR treatment producing a better yield estimation agreement. The results of experiment three indicated a limitation within the proposed methodology for RGB-D yield estimation at plant-level. Such limitation could be from the data acquisition process, or the image segmentation within the data analysis.

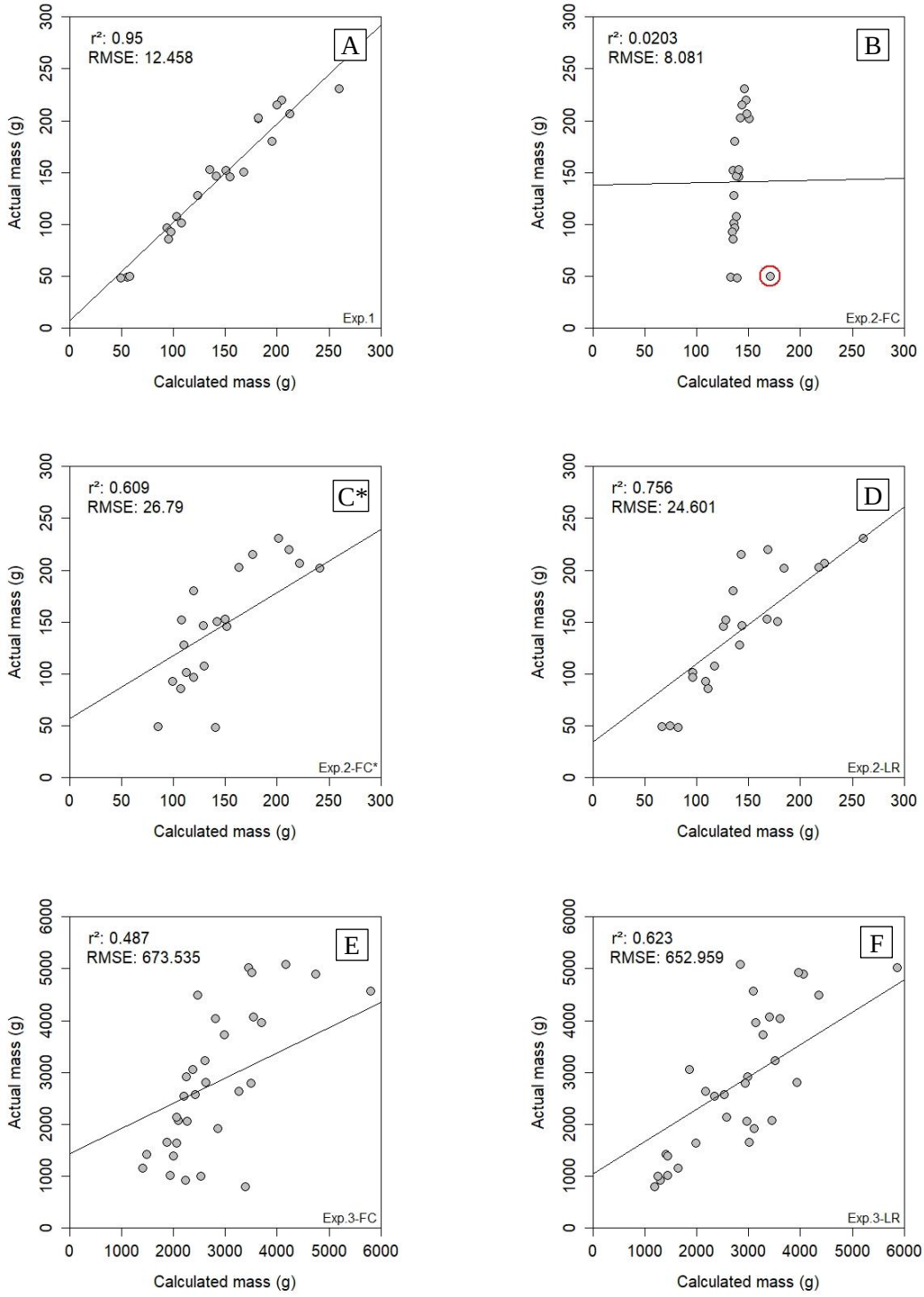


Figure 3.12 Presented RGB-D results of the three experiments; experiment one (A), experiment two FC (n = 21) (B), experiment two FC with statistical outlier removed (n = 20) (C), experiment two LR (D), and experiment three FC (E) and LR (F). *{Key: Exp. – experiment; FC – full canopy; LR – leaf removal; *statistical outlier removed, resulting in 20 bunches.}*

The poor results obtained in experiment two's FC dataset can be attributed to the mesh reconstruction step in the data analysis process. Exaggerated reconstruction of bunches presents a potential limitation of the *Screened Poisson Surface Reconstruction* algorithm (Kazhdan & Hoppe 2013), as depicted in Figure 3.13. The statistical outlying bunch circled in red (Figure 3.12B) produced a Kinect volume of 856 cm3 as illustrated in Figure 3.13A, when the actual volume was only 35 cm^3. An example of an accurately reconstructed mesh (Figure 3.13B) was included for visual comparison. This anomaly was unavoidable during data processing.

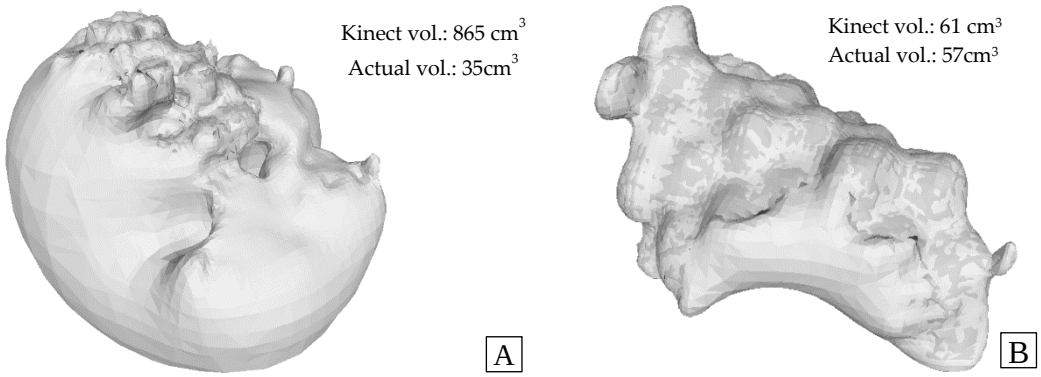


Figure 3.13 Illustration of *Screened Poisson Surface Reconstruction*. The reconstructed bunch (circled in red – Figure 3.12B) with the incorrect volume (A), and an example of a reconstructed bunch of the correct volume (B).

3.5 DISCUSSION

To date, most studies have employed 2-D RGB imagery for vineyard yield estimation at both bunch- and plant-level. However, only Marinello et al. (2016) have investigated the use of 3-D RGB-D imaging for vineyard yield estimation. The research presented here investigated 2-D and 3-D methodologies, incorporating PRS and CV techniques for vineyard yield estimation. The study was undertaken as three experiments, consisting of bunch-level and plant-level datasets, with in-situ measurements captured for the two canopy treatments (FC and LR). The following subsections discuss the results for the two methodologies in further detail.

3.5.1 Using 2-D RGB imagery for yield estimation

The presented results using RGB imagery for yield estimation are robust in nature and support the implemented 2-D methodology. Experiment one (r^2 = 0.889; RMSE = 17.978 g) illustrated the success of RGB imagery in a controlled environment for yield estimation at bunch-level. At plant-level, similar results were presented for the LR treatment in experiment three (r^2 = 0.877; RMSE =

443.235 g). The success of the methodology under both laboratory and field conditions supports the use of CT for image segmentation; as well as the adapted pixel area metric for yield estimation.

CT for image segmentation was favoured by several studies (Dunn & Martin 2004; Font et al. 2015; Liu, Marden & Whitty 2013). In this study, the use of the HSV colour space for thresholding has proven fruitful. The HSV colour space is supported by Font et al. (2015), who achieved favourable results (estimation error of 13.55%) when working with the H layer for segmentation purposes. At bunch-level, experiment one's result (r^2 = 0.889) shows an improvement to the result (r^2 = 0.770) presented by Liu, Marden and Whitty (2013); similarly conducted under laboratory conditions. This was a noteworthy improvement, specifically considering the manual nature of the CT technique. At plant-level, the manually produced results of experiment three (FC: r^2 = 0.779; LR: r^2 = 0.877) align with the automated classification results (r^2 = 0.865) of Liu, Marden and Whitty (2013). Here, Liu, Marden and Whitty (2013) use the same 1×1 m image dataset of Dunn and Martin (2004), as opposed to the plant-level imagery used in this study. Additionally, the CT approach of in this study outperformed the image classification approach (test r^2 = 0.730) (Diago et al. 2012) to segmentation.

This study presented an adaption of the pixel count metric for yield estimation, which expands on current literature (Diago et al. 2012; Dunn & Martin 2004; Font et al. 2015; Liu, Marden & Whitty 2013). Out of the five metrics tested, Liu, Marden and Whitty (2013) concluded that the pixel count produced the best results.

This study found the incorporation of a calibration length (the ruler) for pixel count resulted in an improved quantitative pixel area (cm^2) for yield estimation. Again, the results at bunch-level (r^2 = 0.889) improved the bunch-level results (r^2 = 0.770) conducted in laboratory conditions as presented by Liu, Marden and Whitty (2013), who were specifically testing the various yield estimation metrics. At plant-level, the LR treatment in this study (r^2 = 0.877) outperformed all current literature to date, with the FC treatment (r^2 = 0.779) representing a slight improvement for in-situ measurements. The presented pixel area (cm^2) metric also improved on the berry count results (Nuske et al. 2011; Nuske et al. 2014), with the highest berry count r^2 (0.740) presented still being lower than the plant-level pixel area r^2 (0.779) for the FC treatment in this study. However, the potential of the berry count is applicable across all cultivars, as it does not depend on the colour of the berry (Nuske et al. 2011).

The limitation of slight distance variations between the camera and bunches within each image was resolved by the incorporation of the reference length (the ruler). The necessity of determining the calibration length for each image outweighs the added processing requirement of this step. Overall, this allows improved yield estimation success, as represented by the results in this study. Future work

could attempt to automate this process. However, a more restricting limitation was the human involvement in determining the appropriate threshold values. This could explain the lowered in-situ estimation performance at bunch-level. Future work could investigate a more automated methodology, thus alleviating this limitation.

3.5.2 Using 3-D RGB-D imagery for yield estimation

To date, the basic work presented by Marinello et al. (2016) is the only literature supporting the use of the Kinect V1 sensor for vineyard yield estimation. Marinello et al. (2016) used table grape clusters to determine the optimal viewing angle and distance for the sensor. The authors concluded a side-on view of the cluster with a distance between 0.8–1.0 m produced the least variability in mass estimations. Ergo, the current findings of this study, which incorporated RGB-D imagery for yield estimation across the three experiments, are the most comprehensive findings to date. The presented study exemplifies the potential of 3-D PRS and related CV techniques for yield estimation, specifically a cost-effect sensor like the Kinect V1.

The nature of this study's datasets resulted in separate data analysis between the bunch- and plant-level datasets. Astounding results were obtained in experiment one (r^2 = 0.950; RMSE = 12.458 g). It was evident that the Kinect sensor favours ideal laboratory conditions, allowing accurate yield estimation at bunch-level. The ramification of less-favourable conditions, such as in-situ monitoring, became apparent in experiment two (bunch-level) and three (plant-level).

Although no bunch-level studies have used a Kinect sensor for vineyard yield estimation, a similar approach under laboratory conditions estimating the volume of cauliflowers was presented by Andújar et al. (2016). The authors were able to achieve an r^2 of 0.868 when regressing the estimated volume against the known fruit mass. On the contrary, the methodology used in this study actually enables the relationship between fruit volume and mass to be modelled, allowing the use of the adjusted mass (calculated from the model) for subsequent yield estimation. A fundamental difference between this study's methodology and that of Andújar et al. (2016) is the method in which the 3-D model was captured. The methodology in this study captured a 3-D mesh of the fruit, while (Andújar et al. 2016) captured a 3-D point cloud of the vegetable, which required model reconstruction; constructively producing a 'watertight mesh'. The improved results at this level (r^2 = 0.950) could be accredited to this fundamental difference. Interestingly, both methodologies used the same 3-D reconstruction method; *Screened Poisson Surface Reconstruction* (Kazhdan & Hoppe 2013) found in MeshLab (Cignoni et al. 2008).

The primary limitation of the proposed methodology at bunch-level was a combination of dataset quality and the *Screened Poisson Surface Reconstruction* method (Kazhdan & Hoppe 2013). The consequence of a poor-quality mesh became apparent when bunch reconstruction resulted in large defects, as seen in Figure 3.13A. Such imperfections directly affect the potential for accurate yield estimation, exemplified by the results of experiment two's FC dataset (r^2 = 0.0203; in Figure 3.12B). Future research into this matter is advisable, so as to gain a better understanding of this shortfall within the methodology.

This study presented the novel use of a Kinect RGB-D sensor for in-situ vineyard yield estimation at plant-level (experiment three). The presented plant-level methodology produced promising results for the LR treatment (r^2 = 0.623; RMSE = 652.959 g), while the effect of the canopy coverage was evident in the FC treatment (r^2 = 0.487; RMSE = 673.535 g). Future work could improve the presented methodology, potentially overcoming several limitations. Additionally, the nature of the Kinect V1 sensor results in the fundamental constraint of being suited to indoor use only, as solar irradiance produces excessive interference in the captured imagery (Andújar et al. 2016). Future work should make use of the Kinect V2 sensor, as there are several improvements within the sensor that allow improved outdoor imagery to be captured (Wasenmüller & Stricker 2016). RGB-D sensors, specifically the Kinect V2, are being incorporated in unmanned vehicles as cheap sensor alternatives for vineyard modelling and yield estimation, as demonstrated by Lopes et al. (2017).

3.5.3 Operational potential of developed methodologies

Both the presented 2-D RGB and 3-D RGB-D methodologies achieved acceptable accuracies across the three experiments. The results in this study (Figure 3.11 and Figure 3.12) support the use of PRS and related CV techniques for vineyard yield estimation, especially for VSP-trained Shiraz vineyards. The nature of the presented work was conceptualised to assess 2-D and 3-D PRS sensors side-by-side; a novelty in the vineyard yield estimation domain.

Experiment one illustrated the capability of these two methodologies for successful yield estimation of individual bunches, where the Kinect RGB-D sensor (r^2 = 0.950) outperformed the digital RGB sensor (r^2 = 0.889). The suitability of the lighting under laboratory conditions coupled with the Kinect's ability to capture a 3-D model of the bunch both contribute to the success of the Kinect sensor over the RGB sensor. Nonetheless, robust methodologies were established in the controlled environment.

Experiment two tested the established methodologies in situ, under both FC and LR canopy treatments. The produced results of experiment two – as well as experiment three – confirmed the

hypothesis of a superior yield estimation agreement under the LR canopy treatment. If good canopy management practices are established early in the season, the potential of improved yield estimation results to those obtained under the FC treatment could be achieved. Both sensors produced similar results for FC (approximate $r^2 = 0.610$; using Kinect's modified results (n = 20)) and LR (approximate $r^2 = 0.750$) treatments in experiment two.

The success of the two PRS methodologies can be differentiated at plant-level by the results of experiment three. Conversely to experiment one, the RGB methodology outperformed the RGB-D methodology for yield estimation. The RGB results ($r^2 = 0.877$) outclassed the RGB-D results ($r^2 = 0.487$) under the FC treatment, whereas a smaller margin between the RGB ($r^2 = 0.779$) and RGB-D ($r^2 = 0.623$) sensors occurred under the LR treatment. Inference behind the differing results can lead to the following observations. The different lighting conditions influenced the results; as the RGB imagery was collected at midday, while the RGB-D imagery was collected immediately prior to sunset. Additionally, the continuous movement of the Kinect sensor (RGB imagery captured in a stationary position) could be a reason for the lowered RGB-D results. Further research is encouraged, using a standardised experimental setup, where feasible, to create more favourable conditions for both sensors to operate under.

Overall, the use of the Kinect as a cost-effective RGB-D sensor for vineyard yield estimation, specifically at bunch-level is supported by the results obtained for experiment one. However, the robustness of the RGB methodology is evident across all three experiments, with substantial plant-level results obtained in situ. To this end, the results obtained in this study support the potential for operationalisation of both PRS sensors.

3.6 CONCLUSION

A novel approach to a side-by-side investigation of 2-D and 3-D PRS and related CV techniques for successful vineyard yield estimation has been presented. This study assessed RGB imagery captured by a digital camera, and RGB-D imagery captured by a Microsoft Kinect V1 sensor across three experiments; with in-situ measurements captured under two canopy treatments. The results of this study show that the Kinect RGB-D sensor produced the highest yield estimation agreement under laboratory conditions (bunch-level). At bunch-level, the RGB and RGB-D methodologies performed equally under both canopy treatments for in-situ yield estimation. At plant-level, the best in-situ results were obtained using the RGB imagery, which outperformed the RGB-D results. Both sensors support the use of PRS data acquisition and related CV techniques for vineyard yield estimation, with improved the accuracies presented. The results of this study confirm the operational potential of 2-D

RGB imagery for accurate yield estimation. It is recommended that future work investigate a more automated RGB methodology, suitable for operational environments. Regarding the presented RGB-D methodology, the Kinect demonstrates the potential for vineyard yield estimation using 3-D RGB-D imagery. Future work should investigate the use of the Kinect V2 sensor coupled with suitable lighting conditions for in- situ yield estimation.

CHAPTER 4: VINEYARD YIELD ESTIMATION USING 2-D PROXIMAL REMOTE SENSING: A MULTITEMPORAL ANALYSIS

Hacking C, Poona N & Poblete-Echeverría C 2020. Vineyard yield estimation using 2-D proximal sensing: A multitemporal approach. *OENO One* 54, 4: 793-812.

4.1 ABSTRACT

Vineyard yield estimation is a fundamental aspect of precision viticulture that enables a better understanding of the inherent variability within a vineyard. Yield estimations conducted early in the growing season provide insightful information, ensuring the best fruit quality for the maximum desired yield. PRS techniques enable non-destructive, in-situ data acquisition for yield estimation during the growing season. This study aimed to determine the ideal phenological stage for yield estimation using 2-D PRS and CV techniques (2-D methodology) in a VSP vineyard. To achieve this aim, multitemporal digital imagery was acquired weekly over a 12-week period, with a final acquisition two days prior to harvest. Prior to the multitemporal analysis for yield estimation, an unsupervised KMC algorithm was evaluated for image segmentation on the final dataset prior to harvest, yielding bunch-level segmentation accuracies as high as 0.942, with respective F1-scores of 0.948. The segmentation yielded a pixel area (cm^2), a subsequent input to a cross-validation model utilised to calculate bunch mass (g). The 'calculated mass' was linearly regressed against the 'actual mass', indicating the capability for estimating vineyard yield. From the multitemporal analysis, the final stage of berry ripening was determined as the ideal phenological stage for yield estimation, achieving a global (50 bunches) r^2 of 0.790.

4.2 INTRODUCTION

Yield information is fundamental to precision viticulture practices, providing important information to both the vineyard manager and the winemaker (Nuske et al. 2014). Yield estimates determined early in the growing season facilitate managerial decisions to regulate vine growth, optimising the desired balance between grape quality and quantity at harvest (Aquino et al. 2018). Leading up to harvest, accurate information on the expected yield provides the winery with relevant estimates to guide logistical planning for the harvest period (De la Fuente et al. 2015). Traditional yield estimation methods (De la Fuente et al. 2015; Wolpert & Vilas 1992) are notoriously destructive, labour-intensive and time-consuming (Diago et al. 2015). The combination of PRS and CV techniques has been investigated as an alternative for yield estimation, due to the limitations of traditional methods. Studies have been undertaken both early in the season (Liu et al. 2017) and before harvest (Millan et

al. 2018). PRS incorporates modern terrestrial sensors to capture 2-D and 3-D datasets in a non-destructive, economic manner (Font et al. 2015; Marinello et al. 2016). Researchers employ CV techniques, which emulate human vision to interpret and extract accurate information from digital datasets (Gonzalez, Woods & Eddins 2009), for image processing and yield estimation.

In recent years, multiple studies (Diago et al. 2015; Liu, Marden & Whitty 2013; Millan et al. 2018) have investigated 2-D RGB PRS and related CV techniques for yield estimation, highlighting the potential of RGB imagery for yield estimation. On the contrary, limited research has investigated 3-D RGB-D PRS and CV techniques for yield estimation (Marinello et al. 2016). Hacking et al. (2019) evaluated the use of 2-D (RGB) and 3-D (RGB-D) methodologies for yield estimation before harvest. Although the RGB-D methodology performed well under laboratory conditions (r^2 = 0.950), the in-situ data acquisition suffered interference caused by solar irradiance, lowering the yield estimation capabilities. Andújar et al. (2016) noted that an RGB-D sensor, such as the Microsoft Kinect™ V1 (Microsoft, Redmond, United States), experiences interference caused by solar irradiance, thereby limiting the sensor's acquisition process for in-situ data collection. On the contrary, the 2-D RGB results for yield estimation were more robust for in-situ and laboratory conditions, yielding r^2 values between 0.625 and 0.889. Hacking et al. (2019) concluded that the RGB methodology was better suited for yield estimation, and suggested a more automated approach be investigated.

Yield estimation using 2-D RGB data requires specific image processing, using one of two common CV techniques; bunch detection (image segmentation) and berry detection. Numerous studies (Font et al. 2015; Liu and Whitty 2015) incorporate image segmentation, applied at pixel-level, to differentiate the bunches from the background. A relevant bunch metric, e.g. a pixel count metric (Liu, Marden & Whitty 2013), is then employed to estimate the final yield. Berry detection techniques (Grossetête et al. 2012; Nuske et al. 2014) bypass the pixel-level segmentation (bunch detection) and bunch metric approach. The limitation of the berry detection algorithms was the requirement of berry mass data, which is entered into the algorithm, guiding the estimation process during computation (Aquino et al. 2018). Both techniques require bunch or berry detection prior to yield estimation. However, bunch detection via image segmentation has been more widely applied (Font et al. 2015; Hacking et al. 2019; Millan et al. 2018).

Image segmentation is pertinent to bunch detection – one of the biggest challenges in yield estimation using PRS and related CV techniques (Millan et al. 2018). Various segmentation methods generally use some form of CT or image classification to classify pixel values according to relevant classes defined (i.e. bunch and background) (Diago et al. 2012; Reis et al. 2012). Post-segmentation, a

morphological operation is commonly applied for ‘filtering’ and ‘cleaning’ the segmented image (Millan et al. 2018).

Several studies (Diago et al. 2015; Font et al. 2015; Hacking et al. 2019; Liu, Marden & Whitty 2013; Millan et al. 2018) have evaluated 2-D PRS and related CV techniques that incorporate image segmentation for yield estimation pre- and post-harvest. When evaluating bunch-yield post-harvest under laboratory conditions, Diago et al. (2015) designed a methodology for berry detection using CT, achieving a global (seven red varietals) r^2 of 0.840 for yield estimation. Hacking et al. (2019) achieved comparable accuracies at bunch-level using CT. Under laboratory conditions (post-harvest), the authors presented an r^2 value of 0.889 for yield estimation, with in-situ (pre-harvest) r^2 values of 0.625 (FC treatment) and 0.742 (LR treatment) obtained at bunch-level. Hacking et al. (2019) also presented an r^2 value of 0.877 for plant-level yield estimation under an LR canopy treatment (pre-harvest). Millan et al. (2018) employed a Boolean model for berry estimation before harvest, tested at bunch- and plant-level. The study achieved a yield estimation r^2 value of 0.810 for individual vines under a similar LR canopy treatment, as presented in Hacking et al. (2019).

Originally, bunch detection relied on human-selected colour thresholds to perform image segmentation by selecting the relevant pixels within the defined thresholds (Dunn & Martin 2004; Hacking et al. 2019; Reis et al. 2012). Although CT has been successfully applied (Hacking et al. 2019), the methodology is inherently biased and dependent on accurate threshold selection by a trained specialist. In an attempt to remove the human element during bunch detection, more automated approaches have seen the incorporation of classification methods for 2-D image segmentation.

The majority of these classification methods are supervised and rely on training the segmentation model prior to computation (Diago et al. 2012; Font et al. 2015; Liu & Whitty 2015; Luo et al. 2016; Millan et al. 2018). For example, Diago et al. (2012) implemented a supervised classification methodology which employed Mahalanobis distance (clustering algorithm) to segment the image into several classes. The authors attained an accuracy of 0.980 for bunch detection. More recently, Luo et al. (2016) presented a binary bunch detection methodology that utilised colour components from the images to build linear classification models. These were incorporated into a strong classifier using the AdaBoost framework, achieving a classification accuracy of 0.966. Both these methods required the classifiers to be manually trained at one point or another.

An alternative classification approach is unsupervised classification which completely forgoes manual training (Correa et al. 2012). For instance, KMC (Arthur & Vassilvitskii 2007) is a popular

unsupervised technique that computes the average squared distance between pixels, thereby determining a suitable cluster for the pixels. According to Diago et al. (2012), an unsupervised classification model can be intrinsically unreliable when tasked with classifying an environment with varying structural components, such as a vineyard. Nonetheless, Liu et al. (2017) incorporated unsupervised feature selection and shoot classification algorithms when estimating vineyard yield from early-stage shoot detection at the start of the season. The authors presented an average shoot-detection accuracy of 0.868, with an F1-score of 0.900. However, this study did not employ unsupervised image classification techniques to segment the image. To date, no yield estimation research has adopted unsupervised image segmentation for bunch detection.

A limited number of studies have attempted yield estimation early in the growing season. For example, Grossetête et al. (2012) and Aquino et al. (2018) undertook bunch detection at night between flowering and vèraison, using artificial lighting to detect individual berries. Grossetête et al. (2012) achieved an r^2 of 0.920 for berry detection, while Aquino et al. (2018) used the number of berries to estimate yield and achieved a training r^2 of 0.782. Yield estimations have also been conducted during flowering (Liu et al. 2018; Millan et al. 2016) and early shoot detection stages (Liu et al. 2017). Nuske et al. (2014) investigated various cultivars over four seasons, with data acquisition both prior to vèraison and prior to harvest (maximum ten-day window preceding harvest). Varying phenological stages could influence the final results for yield estimation. To date, no study has conducted yield estimation on a multitemporal scale with weekly data acquisition, ranging from before vèraison all the way to harvest.

This study investigated the use of 2-D PRS and related CV techniques for yield estimation in a Shiraz vineyard, using multitemporal RGB data acquired weekly over a 12-week period. The specific objectives of the study were to evaluate the use of an unsupervised KMC technique for bunch detection during the image segmentation process; and to ascertain the optimal timeframe (phenological stage) in the growing season for yield estimation in a vineyard when utilising a 2-D methodology.

4.3 MATERIALS AND METHODS

4.3.1 Study site

The study was conducted at Stellenbosch University's Welgevallen Experimental Farm, situated in Stellenbosch, South Africa (33° 56' 26" S; 18° 51' 56" E). Data was collected in situ from a Shiraz vineyard (Figure 4.1), using a VSP trellis system. The vineyard was planted approximately 157 m

above sea level in the year 2000 in a North-South row direction and with 2.7×1.5 m vine spacing. Datasets were collected from 16 individual vines spread across three rows during the 2018–2019 growing season. The vineyard, which lies in the coastal wine region of the Western Cape, experiences long and dry summers, exhibiting conditions similar to those of the Mediterranean climate (Conradie et al. 2002).

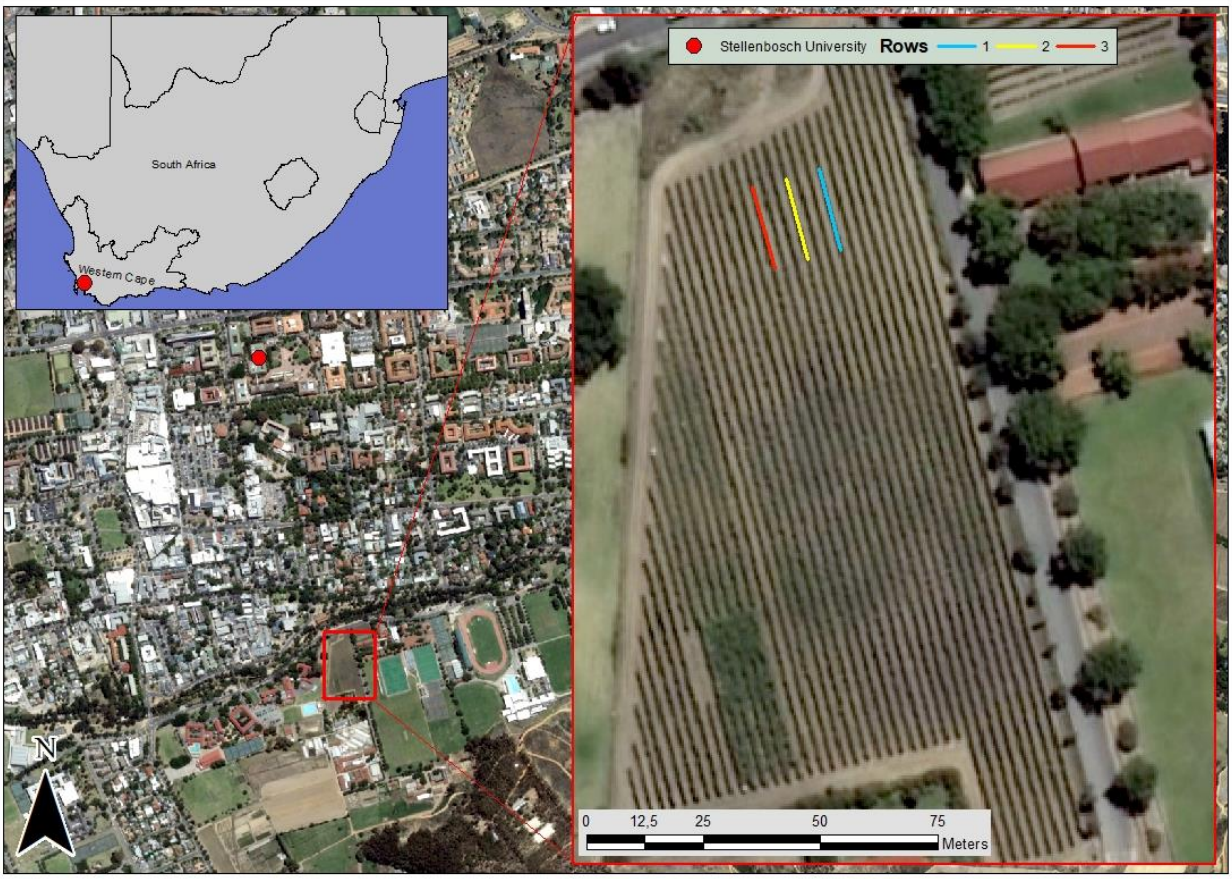


Figure 4.1 Location of the Shiraz vineyard, situated on the Welgevallen Experimental Farm in Stellenbosch, South Africa. The red inset map illustrates the location of the three rows within the vineyard.

4.3.2 Data acquisition

The 2-D image data was captured in situ, at bunch- and plant-level, as illustrated in Figure 4.2. Multitemporal data acquisition monitored 50 individual bunches (bunch-level) weekly, over 12 weeks (08 December 2018 to 25 February 2019) during the 2018–2019 growing season. Weekly displacement measurements were collected for the respective bunches, which served as reference data. The initial bunch-level dataset was acquired on 8 December 2018, approximately four weeks before the onset of ripeness (start of vèraison). The final bunch-level dataset was captured on 25 February 2019, totalling 12 datasets captured at bunch-level. Additionally, a single plant-level dataset, consisting of 16 individual vines, was captured on 25 February 2019. The vines were

harvested on 27 February 2019, where final reference measurements were recorded for both bunch- and plant-level under laboratory conditions.

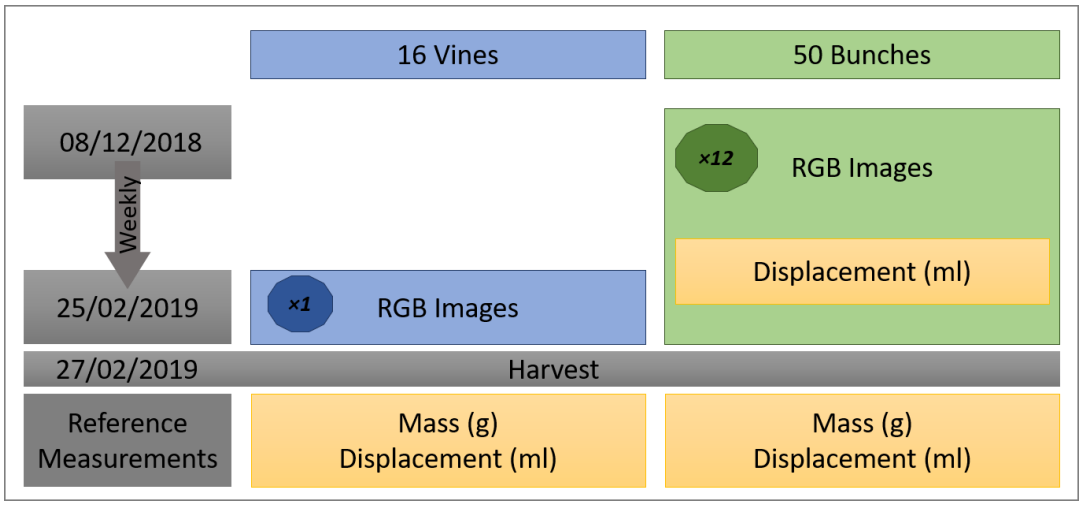


Figure 4.2 Image and reference data acquisition for the 2018–2019 growing season.

4.3.2.1 Reference measurements

Bunch volume (cm^3) was recorded per individual bunch (×50 bunches), collected with the weekly RGB data acquisition (Figure 4.2). Figure 4.3A illustrates the volume acquisition system designed and built for this specific purpose. The system captures a single image of the water level before and after the bunch was submerged using a digital *PowerShot-ELPH160* camera (Canon, Tokyo, Japan). The displaced volume was extracted using a custom-built script in MATLAB® (The MathWorks Inc. 2018). The difference in the two levels yielded a bunch displacement value, which was used as a proxy for bunch volume (Ferreira & Marais 1987).

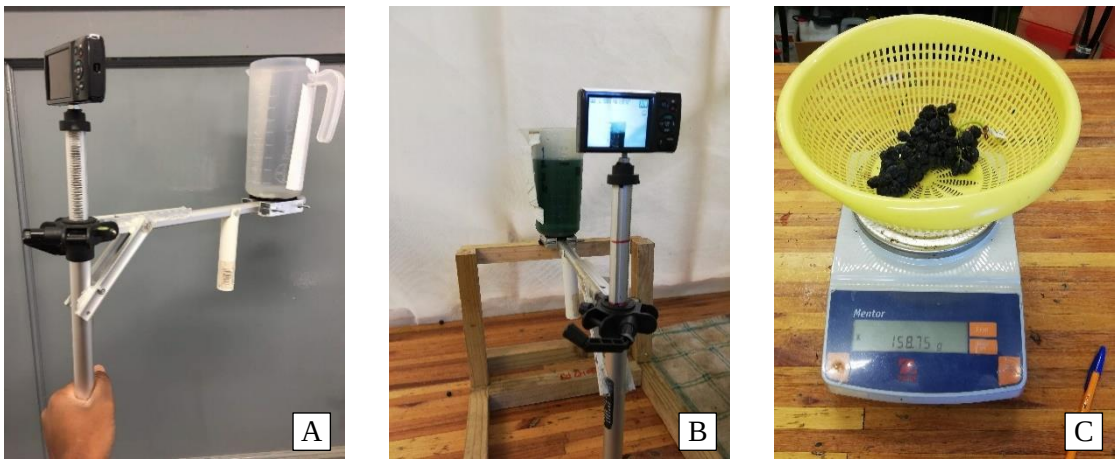


Figure 4.3 Reference measurement systems measuring bunch displacement in the field (A). Laboratory measurements captured bunch displacement (B) and bunch mass (C).

Final reference measurements were recorded for the harvested vines under laboratory conditions. This process saw volume (cm^3) (Figure 4.3B) and mass (g) (Figure 4.3C) measurements individually recorded per bunch harvested from the 16 vines. The reference measurements of the 50 individual bunches (sampled from the 16 vines) were noted separately during this process. Individual bunch mass (g) was recorded under laboratory conditions using a Mentor scale (Ohaus, Parsippany, United States). For consistency, the same volume measurement system was used in the laboratory, with the addition of a custom-built base that held the system level and steady (Figure 4.3B). Individual bunch measurements per vine were aggregated, yielding the vine's individual volume (cm^3) and mass (g).

4.3.2.2 Bunch-level image acquisition

Bunch-level data was acquired weekly, monitoring 50 individual bunches that were statistically sampled. Shortly after flowering (21 November 2018), the number of bunches per vine was determined. Subsequently, the number of bunches per vine was plotted using a box and whisker diagram, where four vines were randomly selected from each quartile – totalling 16 vines. Owing to logistics, it was decided to monitor 20% of the bunches per individual vine selected, yielding either two, three, or four bunches per vine. This resulted in a total of 52 bunches selected. However, two bunches broke off during the season. A total of 50 bunches were therefore monitored throughout the season.

Bunch-level RGB imagery was acquired by the same system that was utilised for collecting the reference volume measurements (Figure 4.3A), incorporating a white background (Figure 4.4A). The white background produced a distinct image contrast, an improvement on the methodology presented in Hacking et al. (2019). The acquisition system maintained a parallel and perpendicular distance of 40 cm between the camera and each bunch. Where necessary, occluding leaves were manually concealed during image acquisition. Images were captured during the early morning (varied between 08:00 and 10:00) under natural illumination. An umbrella provided shade over the bunch being photographed.

Figure 4.4B is an example of a bunch captured on 8 December 2018, forming part of the first dataset. This clearly depicts the 1×1 cm black calibration squares (Figure 4.4B) that were placed on the white background for scale reference. The calibration squares were incorporated to improve the ruler-calibration technique presented in Hacking et al. (2019). The ruler was limited to a single dimension for image calibration, whereas the calibration square considers the 2-D image plane when calibrating the pixel area. The same bunch, as in Figure 4.4B, was captured on 25 February 2019 (Figure 4.4C), two days prior to harvest, illustrating the bunch development from the first image to the final image.

Figure 4.4 Bunch-level data acquisition system (A), with an example captured on 08 December 2018 (B), and the same bunch captured on 25 February 2019 (C).

4.3.2.3 Plant-level image acquisition

RGB images were acquired for 16 vines on 25 February 2019, producing the plant-level dataset. Two images were captured per vine, one for the east-facing canopy and one for the west-facing canopy. An EOS 650D (Canon, Tokyo, Japan) digital single-lens reflex camera was used in *Auto* mode for image acquisition. The camera was positioned perpendicular to the vine at a distance of 2 m, capturing the entire vine's canopy (Figure 4.5A). The addition of the white background served two purposes: to remove background noise from the image (Figure 4.5A), and for image calibration (Figure 4.5B). The white background was a vital improvement to the methodology presented in Hacking et al. (2019), and was held directly behind the vine during image acquisition. Imagery was captured in the morning during 'golden-hour', the natural illumination directly after sunrise (and just before sunset), which is dependent on location and solar inclination, resulting in an indirect warm lighting environment (timeanddate.com 2019).

Figure 4.5 Plant-level data acquisition with the white background behind the vine (A) and the calibration square on the background (B).

4.3.3 Data analysis

The analysis of the data, based on the RGB methodology for yield estimation presented in Hacking et al. (2019), was twofold, aligning with the objectives of this study. The first component analysed CV techniques for image segmentation. Section 4.3.3.1 details the custom script used for image segmentation, while Section 4.3.3.2 details the process employed to assess the segmentation accuracy. The second component (Section 4.3.3.3) conducted statistical analysis using cross-validation to determine the estimated yield, which was then regressed with the actual yield measured at harvest. Additionally, the estimated mass per bunch was averaged per vine and used to infer the vine's yield at harvest, detailed in Section 4.3.3.4.

4.3.3.1 Image segmentation

Image segmentation for both bunch- and plant-level was completed with a custom script in MATLAB® (The MathWorks Inc. 2018). Figure 4.6 illustrates the flow diagram of the custom script, outlined in three stages:

Stage 1: Calibration and pre-processing

1. The raw RGB image was imported for processing.
2. Semi-automated image calibration was computed using the black calibration squares (Figure 4.4B and Figure 4.5B), yielding a calibrated pixel-area (cm^2) coefficient.
3. The ROI was manually digitised; whereby minimal background was included.
4. Segmentation accuracy assessment ROIs (selection of five 'bunch' and five 'background' ROIs were manually digitised). Details are provided in Section 4.3.3.2.
5. Image converted from RGB to HSV colour space.

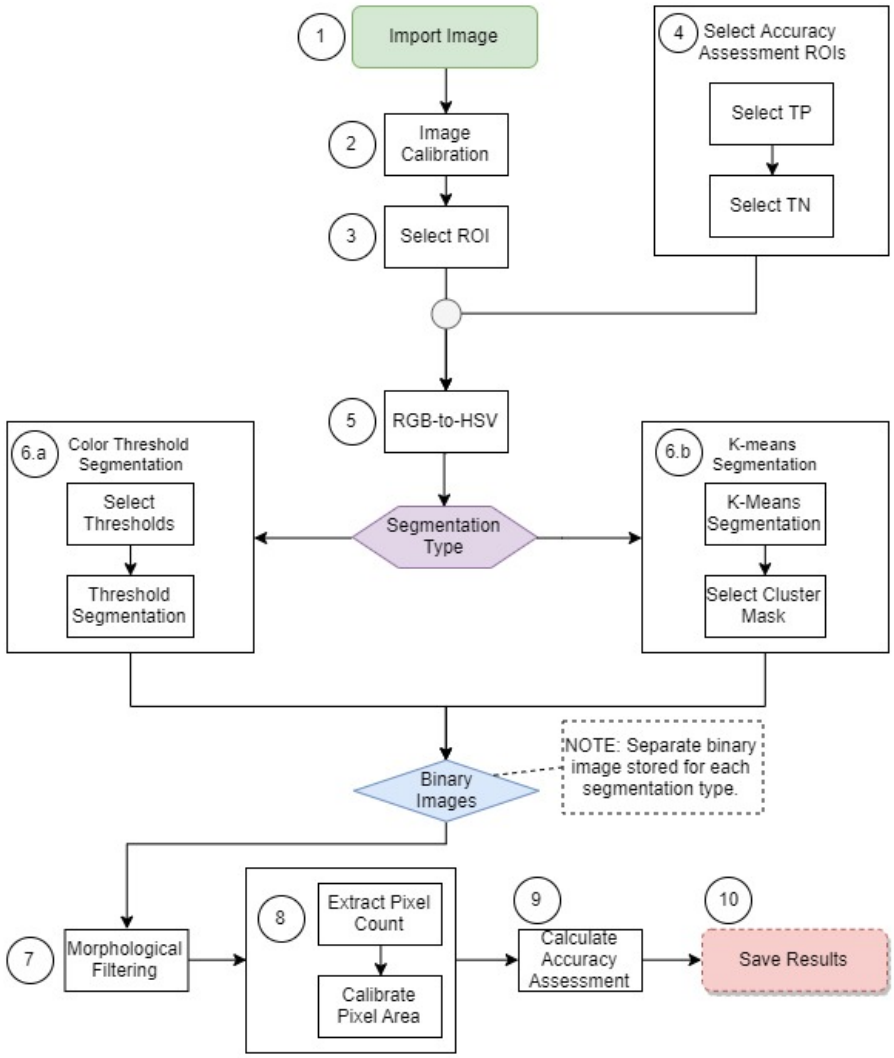


Figure 4.6 Flow diagram of the image analysis script executed in MATLAB® (The MathWorks Inc. 2018).

Stage 2: Image segmentation

Two segmentation techniques were assessed: a manual CT technique (Hacking et al. 2019), and an unsupervised KMC technique (Arthur & Vassilvitskii 2007). CT was implemented as the bench-mark segmentation technique in this study, where colour thresholds are manually selected for segmentation. The KMC technique requires a *k*-value to define the number of clusters to segment the image. KMC classifies the image into the relevant clusters by minimising the average squared distance between points within the same cluster, thereby classifying the pixels into the respective cluster (Arthur & Vassilvitskii 2007).

The script proceeded from the previous step as follows:

6. Image segmentation:
    a. CT required the manual determination of suitable threshold limits based on the HSV colour space, using MATLAB's (The MathWorks Inc. 2018) *Colour Thresholder app* – encompassed in the *Image Processing Toolbox*™. For each dataset processed, the first image per row was used for selecting the threshold limits, which was then applied to the entire row. The output was a segmented binary image: bunch and background.
    b. The *'imsegkmeans'* (i.e. KMC) algorithm (Arthur & Vassilvitskii 2007), part of the *Image Segmentation* section in MATLAB's (The MathWorks Inc. 2018) *Image Processing Toolbox*™, applied a value of '*k=3'* and '*k=4*' at bunch- and plant-level, respectively. The algorithm repeated the clustering process three times, with the remaining parameters set to default. See the documentation for detail (The MathWorks Inc. 2019). The algorithm utilises the HSV properties of the image for clustering. Post-classification, the appropriate cluster mask was manually selected, yielding a segmented binary image – bunch and background.

Stage 3: Post-segmentation processing

7. A sequence of morphological operations, first dilation and then erosion, were applied to the segmented binary image. Dilation 'grows' or 'thickens' the binary image – effectively filling any holes, while erosion 'shrinks' or 'thins' the binary image – removing any outliers; both within certain thresholds (Gonzalez, Woods & Eddins 2009). The morphological operators used a 2-D 'disk' structure with a radius of '10' pixels for image refinement.
8. The number of pixels representing 'bunch' in the binary image were counted. Bunch area (cm^2) was determined using the number of pixels and the calibrated pixel-area coefficient (Step 2).
9. Segmentation accuracy assessment was calculated for both segmentation techniques (Section 4.3.3.2).
10. Results were exported and saved.

4.3.3.2 Segmentation accuracy assessment

Segmentation accuracy was calculated per image, part of the custom script detailed in Section 4.3.3.1. The ROIs defined for assessing segmentation accuracy (Step 4, Section 4.3.3.1) represented actual (ground-truth) pixel values for bunch and background. These values were utilised when assessing the segmentation techniques (Step 9, Section 4.3.3.1).

Table 4.1 presents a confusion matrix, where the following accuracy assessment metrics (Liu & Whitty 2015) were evaluated for individual pixels post-segmentation, inclusive of the morphological operators:

- *True positive (TP)*: pixel manually labelled as bunch and automatically detected as bunch.
- *True negative (TN)*: pixel manually labelled as background and automatically detected as background.
- *False negative (FN)*: pixel manually labelled as bunch, but automatically detected as background.
- *False positive (FP)*: pixel manually labelled as background, but automatically detected as bunch.

Table 4.1 Confusion matrix for binary classification

		Actual	
		Bunch	Background
Predicted	Bunch	TP	FP
	Background	FN	TN

Adapted from Luque et al. (2019: 218).

Using the metrics defined above, *accuracy*, *recall*, *precision* and *F1-score* were calculated. *Accuracy* indicates the percentage of correctly classified pixels, both as bunch and background, from the total metric population – defined by Equation 4.1 (Liu & Whitty 2015):

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FN} + \text{FP}} \qquad \text{Equation 4.1}$$

Recall calculates the percentage of pixels correctly classified as bunch from the manually labelled bunch-pixels, defined by Equation 4.2 (Aquino et al. 2018):

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \qquad \text{Equation 4.2}$$

Precision indicates the percentage of pixels correctly classified as bunch out of the population of pixels classified as bunch, irrespective of the manual label – defined by Equation 4.3 (Aquino et al. 2018):

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \qquad \text{Equation 4.3}$$

Finally, the *F1-score* combines *recall* and *precision* into a single metric, indicating the success of the binary classification; bunch and background in this instance. Equation 4.4 defines the *F1-score*, a

value between '0' and '1', where '1' indicates a perfect harmonic mean between *recall* and *precision* (Aquino et al. 2018):

$$\text{F1-score} = 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad \text{Equation 4.4}$$

4.3.3.3 Yield estimation

A custom yield estimation script was compiled using R statistical software (R Core Team 2019), which integrated the *Caret* package (Kuhn 2008) for cross-validation. Five-fold cross-validation, repeated ten times for robustness, was incorporated for developing the yield estimation model, which was executed and evaluated across three steps:

1. Bunch area (cm^2) was linearly regressed against the respective volume measurement (cm^3) in the first cross-validation model. The yield estimation model began with this step to include the multitemporal reference measurements (bunch volume). The model produced 'fitted' volume (cm^3) values.
2. The 'fitted' volume (cm^3) values were then linearly regressed against the actual mass (g) in a subsequent cross-validation model, yielding estimated mass (g) from the 'fitted values'. This step was justified by the established relationship between bunch mass and volume presented in Hacking et al. (2019).
3. The estimated mass (g) was then linearly regressed against the actual mass (g), yielding an r^2 value (coefficient of determination) that indicated the potential for yield estimation. Additionally, the RMSE was calculated from the linear regression, which specified the yield estimation error in grams.

4.3.3.4 Inferred plant-level yield estimation

Plant-level yield was inferred from the estimated bunch-level yield, determined for the relevant date from the multitemporal datasets. Equation 4.5 was formulated to serve this purpose:

$$V_m = \frac{\sum_{m=1}^{n} B_m}{n} \times N \quad \text{Equation 4.5}$$

where V_m represented the inferred vine mass; calculated by the sum of the individual bunch masses (B_m), divided by the number of bunches monitored (n) on the respective vine – yielding a mean bunch mass per vine. The mean bunch mass (g) was then multiplied by the vine's bunch population (N –

manually counted at harvest), yielding the inferred vine mass (g). The coefficient of determination (r^2) was calculated between V_m and the actual vine mass (g) determined at harvest.

4.4 RESULTS

The results presented in this section are founded on the linear relationship between bunch mass (g) and volume (cm^3), presented in Hacking et al. (2019). Upon evaluating the reference measurements for the data, the authors established strong relationships at bunch-level (r^2 = 0.971; 21 bunches) and at plant-level (r^2 = 0.996; 31 vines). The reference measurements collected for this study portrayed near-identical relationships at bunch- (r^2 = 0.981; 50 bunches) and plant-level (r^2 = 0.995; 16 vines).

4.4.1 Segmentation results

4.4.1.1 Harvest: Bunch-level

Table 4.2 presents the segmentation results for the two techniques evaluated at bunch-level (Figure 4.7A–C). The KMC technique achieved a segmentation accuracy of 0.942 and F1-score of 0.948, outperforming the CT technique (accuracy = 0.939; and F1-score = 0.943). Owing to the KMC technique being unsupervised, the segmentation process becomes more robust than the CT technique, which is dependent on human-selected thresholds. It is visually evident that the thresholds manually selected during the CT technique were not completely inclusive of the entire bunch in certain situations, resulting in the incorrect segmentation of the bunch as background (Figure 4.7B).

Table 4.2 Segmentation accuracy results for bunch detection on 25 February 2019, computed at bunch-level (50 bunches) and plant-level (16 vines).

	Accuracy		**Precision**		**Recall**		**F1-Score**	
	CT	**KMC**	**CT**	**KMC**	**CT**	**KMC**	**CT**	**KMC**
Bunch-level	0.939	0.942	0.976	0.954	0.933	0.959	0.943	0.948
Plant-level	0.916	0.900	0.862	0.789	0.904	0.966	0.865	0.864

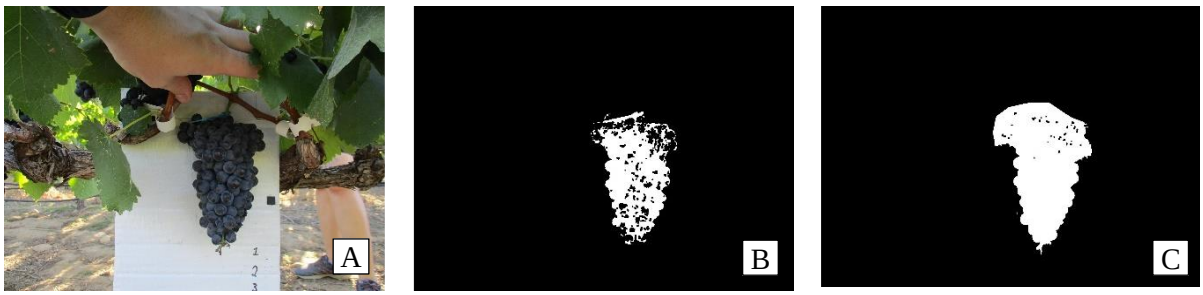


Figure 4.7 Visual representation of segmentation results at bunch-level: raw image (A), CT (B), and KMC (C).

The following observations can be deduced from the precision and recall results, distinguishing the two segmentation techniques. The CT technique produced fewer FPs when segmenting the image than the KMC technique, as the CT technique achieved a higher precision – 0.976 (Table 4.2). FPs indicate that the background has been incorrectly classified as bunch, illustrated by the area circled in red in Figure 4.8C. On the contrary, the KMC technique was more effective for correctly segmenting the bunch, yielding more TPs and fewer FNs than the CT technique; encapsulated by the higher recall (KMC = 0.959; CT = 0.933). A reduced recall was visually evident by the increased presence of FNs in the segmented image (Figure 4.8). Effectively, portions of the bunch were incorrectly classified as background; i.e. the black pixels within the bunch (Figure 4.8C). Segmentation errors can be attributed to inherent variability in the HSV within the image (Figure 4.8B).

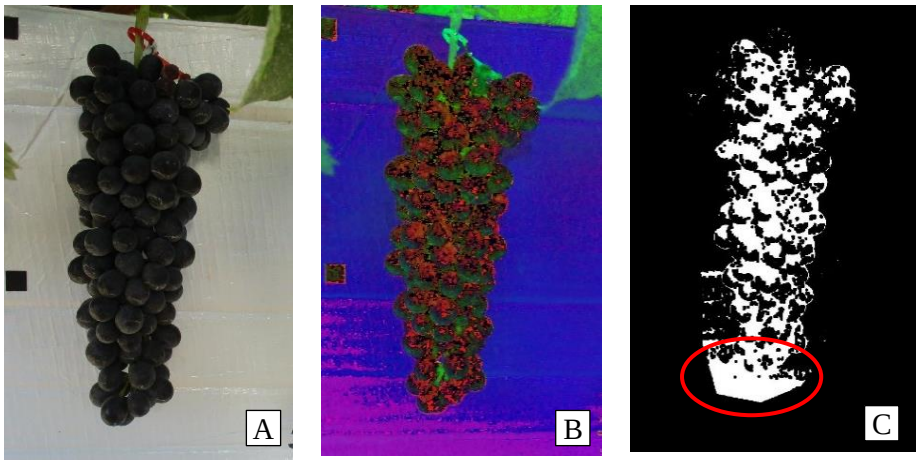


Figure 4.8 Example of precision and recall errors; original RGB image (A), converted HSV image (B), and segmented image (C).

4.4.1.2 Harvest: Plant-level

Table 4.2 presents the segmentation results achieved at plant-level on 25 February 2019, accompanying the bunch-level results presented in Section 4.4.1.1. The results indicated the CT technique performed better than the KMC technique at plant-level, evident when analysing the respective accuracy (CT = 0.916; and KMC = 0.900) and F1-scores (CT = 0.865; and KMC = 0.864). However, the margin differentiating the two techniques was again negligible, where accuracy varied by 0.016 and F1-scores varied by 0.001. Figure 4.9A–C illustrates the segmentation outputs for the CT (Figure 4.9B) and KMC (Figure 4.9C) techniques. At this scale, it is extremely difficult to visually differentiate between the techniques. Similarly to the bunch-level results (Section 4.4.1.1), CT produced a higher precision value than KMC, and, conversely, KMC yielded a higher recall than CT. This supports the observation that the CT technique yields more FNs and less FPs than the KMC

technique – in other words, the KMC technique is better at correctly classifying pixels as bunch, and the CT technique is better at correctly classifying pixels as background.

Figure 4.9 Image segmentation results for bunch detection at plant-level; raw image (A), CT (B), and KMC (C).

4.4.1.3 Multitemporal: Bunch-level

Upon evaluating the segmentation results for both bunch- and plant-level, the decision was made to solely utilise the KMC technique for the multitemporal bunch-level analysis. Figure 4.10 presents the multitemporal segmentation results, accuracy and F1-score, for the 50 bunches. Both indicate the segmentation performance, with indistinguishable values following the same pattern during the data acquisition timeframe (12 weeks). The KMC technique achieved an average accuracy of 0.955 and F1-score of 0.956 during the 12-week period. This was reduced by the large trough in the graph during the middle of the acquisition period (27 December 2018 to 24 January 2019), which aligned with vèraison. During this period, the bunches underwent a colour change, from green to purple/black berries (Figure 4.11). This process restricted the KMC's ability to accurately detect the entire bunch, often resulting in portions of the bunch being omitted from the final binary image (as evidenced in Figure 4.11C). The lowest segmentation results occurred on 17 January 2019 (Figure 4.10), where both accuracy and F1-score achieved a value of 0.859. On the contrary, the best results occurred between 7 February 2019 (accuracy = 0.992 and F1-score = 0.993) and 14 February 2019 (accuracy = 0.991 and F1-score = 0.992), where there was a difference of 0.001 between each date.

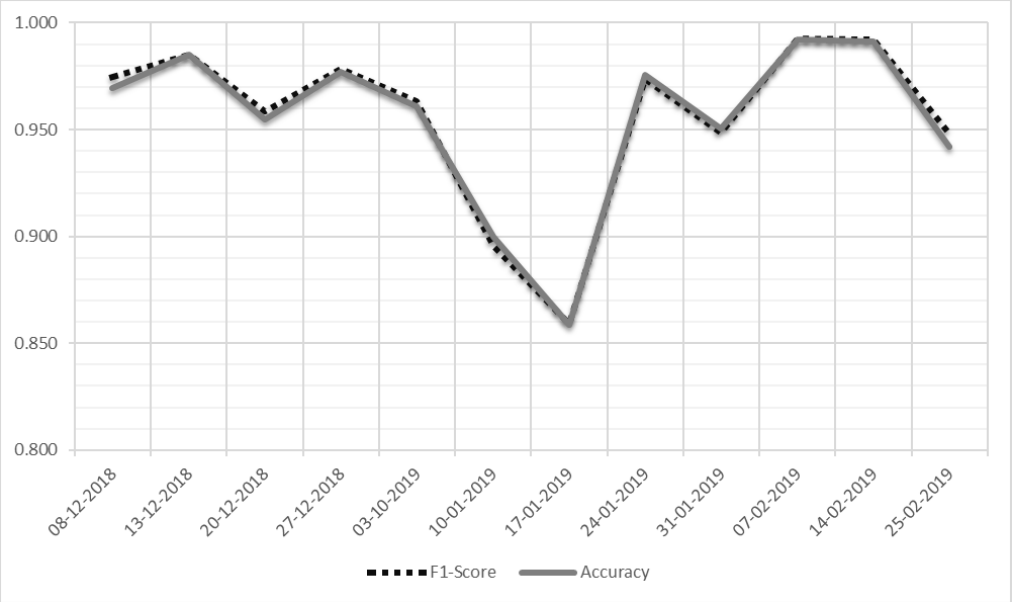


Figure 4.10 Achieved segmentation results, accuracy and F1-score, using the KMC technique for the multitemporal RGB data.

Figure 4.11 The process of vèraison, where the bunch experiences a colour change during ripening; before vèraison (A), during vèraison (B), binary segmentation during vèraison (C), and after vèraison (D).

Outside the vèraison timeframe, three other dates experienced a reduction in both accuracy and F1-score. On 20 December 2018 and 31 January 2019, a maximum reduction of approximately 0.030 was experienced. This had no visible effect on the subsequent yield estimation presented in Section 4.4.2.1. The third date, 25 February 2019, had an approximate reduction of 0.050 from the previous date. Upon visual inspection of the processed results, these reductions in accuracy and F1-score were attributed to either a single bunch with a very low recall – increased FNs – or several bunches with slightly lowered precision (in comparison to the remainder of the sample population) – increased FPs. The increased FNs and FPs were likely caused by environmental conditions such as image shadow, something that naturally fluctuates and is generally uncontrollable under solar illumination.

4.4.2 Yield estimation: Multitemporal results

During the data acquisition period, field observations noted the three rows were at different phenological stages. When the yield estimation results were evaluated, the varying phenological stages became evident, aligning with the field observations. The decision was therefore made to present the results as a consolidated, global dataset, as well as individual rows, yielding an additional three sub-datasets. At bunch-level (Section 4.4.2.1), the global dataset represented 50 bunches and 16 vines at plant-level (Section 4.4.2.2).

4.4.2.1 Bunch-level

The multitemporal yield estimation results (r^2 values) achieved at bunch-level are presented in Figure 4.12. The graph presents a general trend where the values increased at the start of the multitemporal datasets, which was followed by a decrease in the r^2 values as the bunches experienced vèraison. Subsequent to vèraison, the r^2 values increased again, achieving the best results roughly two weeks prior to harvest. Each individual row obtained the lowest r^2 value within a two-week window (03 January to 17 January 2019), supporting the in-situ observations regarding the varying phenological stages between the rows. The global bunch r^2 reached its lowest value (0.349) on 10 January 2019, and its highest value (0.790) on 14 February 2019.

Figure 4.12 Bunch-level r^2 values evaluating yield estimation from multitemporal RGB data. The global dataset (all rows) consisted of 50 bunches, with row one (14 bunches), two (20 bunches) and three (16 bunches) presented as sub-datasets.

On 25 February 2019 – two days prior to harvest – a reduced r^2 value for yield estimation was presented. This drop in r^2 value may be explained by overripe bunches observed at harvest. An example of this is illustrated in Figure 4.13, representing the same bunch on 14 February and 25 February 2019. The extracted area on 14 February (Figure 4.13A) was 112.453 cm^2, which shrank down to 88.893 cm^2 on 25 February (Figure 4.13B). This aligns with the reference volume measurements recorded on 14 February (244.350 cm^3) and 25 February (111.821 cm^3).

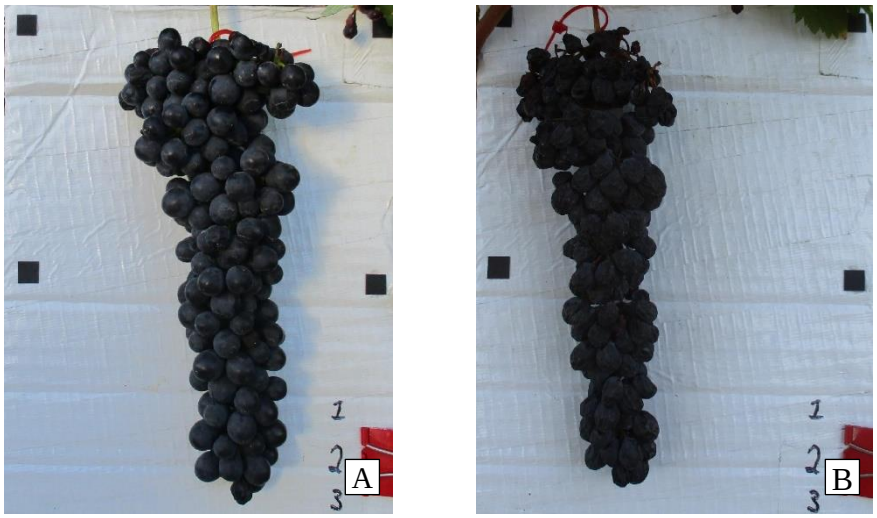


Figure 4.13 A bunch imaged on 14 (A) and 25 (B) February 2019, illustrating an overripe bunch observed on 25 February.

The best yield estimation for the global dataset was observed on 14 February 2019, achieving an r^2 value of 0.790 and an RMSE of 26.918 g (see Figure 4.14A for more detail). The highest r^2 results for both row two (r^2 = 0.697; Figure 4.14C) and three (r^2 = 0.913; Figure 4.14D) occurred on the same day, with RMSE values of 30.710 g and 17.833 g, respectively. Although the highest r^2 (0.932) for row one was achieved on 7 February 2019, the r^2 (0.914) achieved on 14 February 2019 (Figure 4.14B) was still the highest of the three rows. The points circled by the dotted line in Figure 4.14C exemplify over-estimated yields, likely caused by overripe bunches.

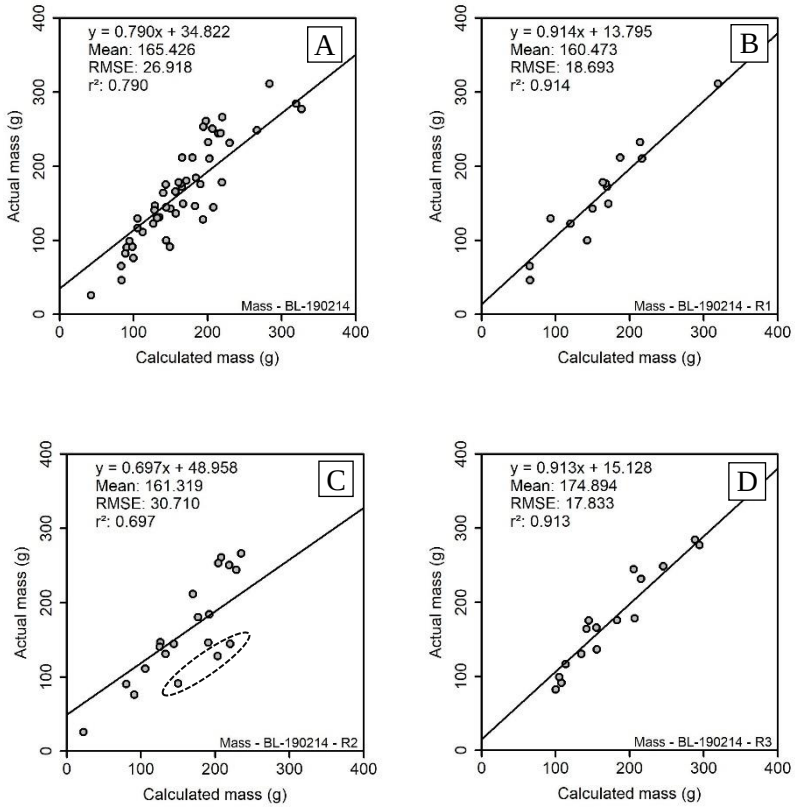


Figure 4.14 Relationships between estimated and actual yield on 14 February 2019; global dataset – 50 bunches (A), row one – 14 bunches (B), row two – 20 bunches (C), and row three – 16 bunches (D).

4.4.2.2 Plant-level

Figure 4.15 presents the vine-inferred yield estimation results achieved using Equation 4.5. The average harvest yield per vine (from 16 vines) was 2.831 kg. From a quick visual inspection, it was evident that both row one and three performed relatively well throughout the growing season, while row two performed poorly during the season, and only improved towards harvest. The reduced results of row two prompted further investigation.

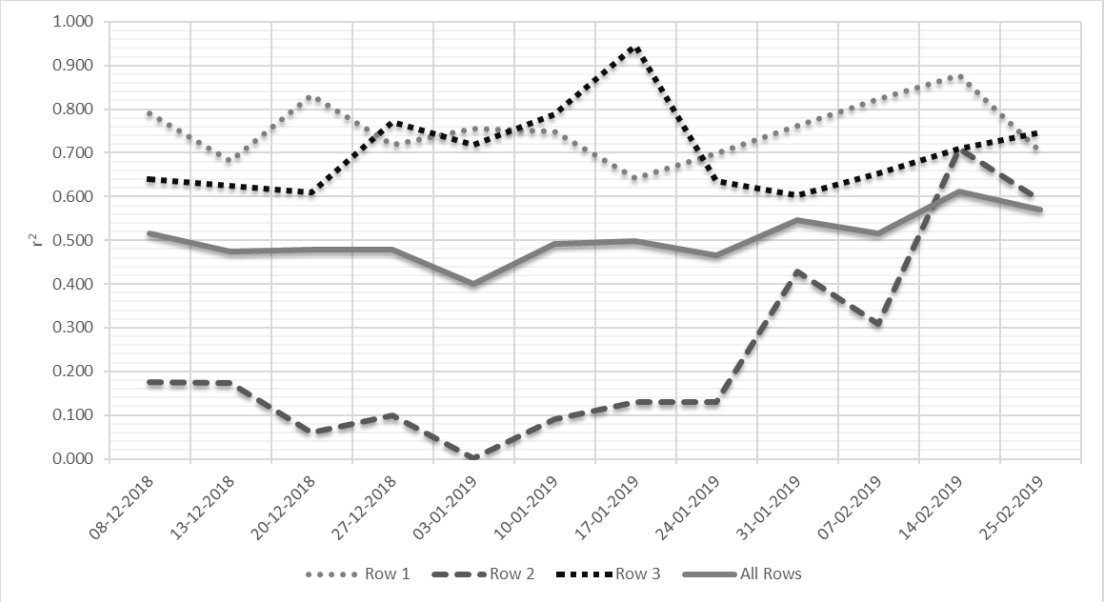


Figure 4.15 Vine-inferred r^2 values, indicating the potential of this technique for yield estimation. Evaluated for the global dataset (vines = 16), and individual rows: one (vines = 5), two (vines = 6), and three (vines = 5).

Figure 4.16 was incorporated to further scrutinise the vine-inferred data, specifically row two – represented by vines 6 to 11. Upon investigation, it was noted that the inferred mass of vine 10 (3421.247 g) was significantly higher than the actual mass (1478.420 g) from harvest. Visual investigation of the vine and the vine's bunches from images captured at harvest indicated severe damage, both to the canopy (Figure 4.17A) and the bunches (Figure 4.17B). The canopy damage left the bunches exposed, resulting in overripe bunches with bird damage; reducing the final yield at harvest, and negatively influencing the results presented.

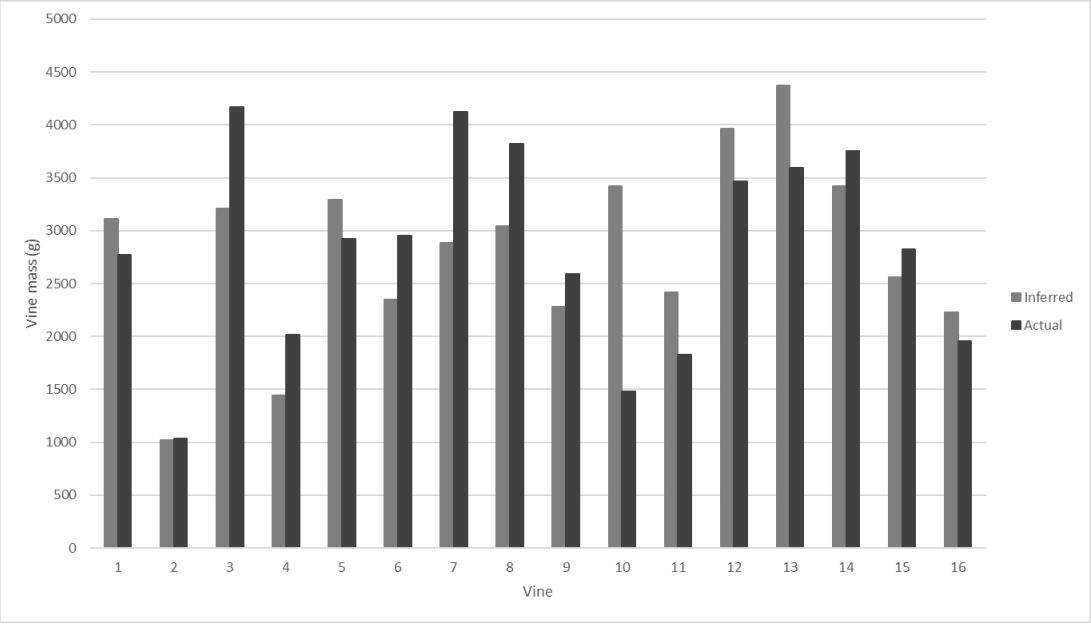


Figure 4.16 Inferred (03 January 2019) vs. actual (harvest – 27 February 2019) mass per individual vine.

Figure 4.17 Indication of canopy damage (A) to vine 10 (Figure 4.16), and subsequent bunch damage (B) from same vine.

Figure 4.18 illustrates a improvement to the vine-inferred yield estimation results, attained by omitting vine 10 from the results. The adjusted results were more uniform, with r^2 values fluctuating between 0.564 and 0.945 (inclusive of individual rows). The highest global r^2 value improved from 0.612 (Figure 4.15) to 0.702 (Figure 4.18), with the lowest global r^2 improving from 0.399 to 0.593. The fluctuation of the results may be attributed to the bunch count per vine. In this case, the number of bunches per vine was counted at harvest and not captured on each respective date.

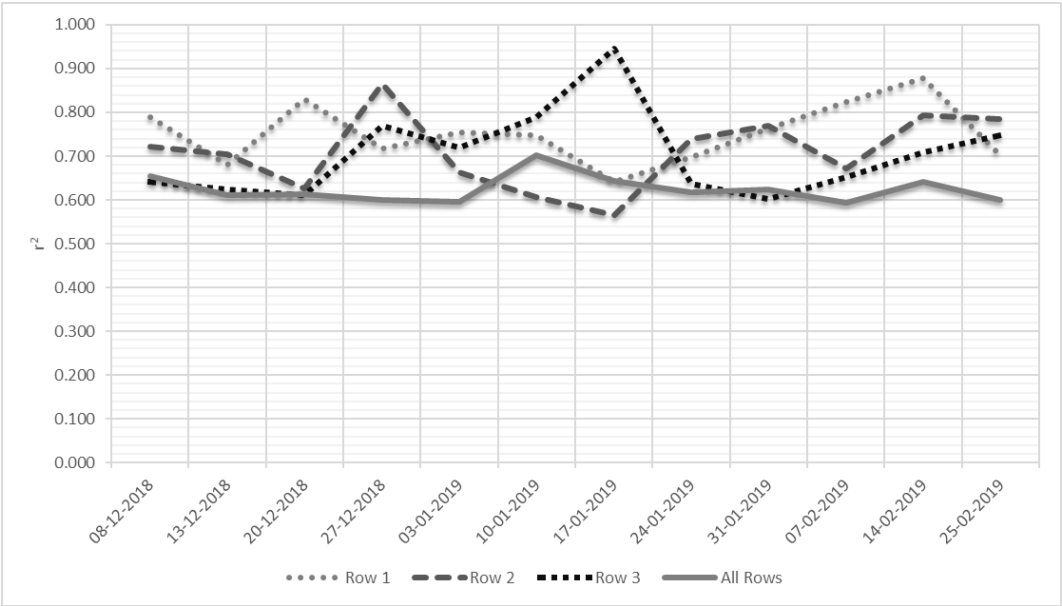


Figure 4.18 Adjusted (vine 10 removed) r^2, indicating vine-inferred yield estimation results for the global dataset (vines = 15) and individual rows (vines = 5) – one, two, and three.

4.5 DISCUSSION

Researchers have successfully utilised 2-D PRS and related CV techniques for yield estimation before vèraison (Aquino et al. 2018; Liu et al. 2017), at harvest (both pre- and post-harvest) (Diago et al. 2015; Millan et al. 2018), or both (Nuske et al. 2014). This study successfully implemented unsupervised KMC for image segmentation, replacing the CT technique presented in Hacking et al. (2019). This study set out to determine the ideal phenological stage for yield estimation in a vineyard. To this end, a multitemporal analysis using RGB data was performed over a 12-week period (8 December 2018 to 25 February 2019), concluding with harvest on 27 February 2019. The following subsections outline the presented results.

4.5.1 Image segmentation

The segmentation results provided insight regarding the success of the two techniques evaluated. Comparing the bunch-level segmentation results (Table 4.2) indicated that the KMC technique (accuracy = 0.942 and F1-score = 0.948), although marginally, outperformed the CT technique (accuracy = 0.939 and F1-score = 0.943). Comparing these results to the in-situ bunch-level segmentation results (accuracy = 0.781 and F1-score = 0.842) of Hacking et al. (2019), both the CT technique and the KMC technique presented in this study performed better. The improved segmentation results may be attributed to the white background that was included behind the bunches for image contrast. By comparison, the segmentation results of this study compare favourably with several studies conducted before harvest (for example, Font et al. (2015) and Luo et al. (2016)). Font et al. (2015) evaluated various segmentation techniques for bunch detection of ‘Flame Seedless’ table-grapes preceding harvest, where CT in the H layer from the HSV colour space achieved the least error (0.136) prior to morphological filtering, which further reduced the error by 0.036. Luo et al. (2016) investigated the automatic bunch detection of ‘Summer Black’ grapes, utilising multiple colour components and the AdaBoost framework for classification purposes. The authors achieved a bunch classification accuracy of 96.56% when performed at bunch-level under various greenhouse and outdoor conditions.

At plant-level (Table 4.2), the CT technique (accuracy = 0.916 and F1-score = 0.865) outperformed the KMC technique (accuracy = 0.900 and F1-score = 0.864). The results of this study aligned with those of Aquino et al. (2018), Hacking et al. (2019), and Millan et al. (2018). For example, Aquino et al. (2018) presented bunch-detection results (recall = 0.876 and precision = 0.958) prior to vèraison, whereas Millan et al. (2018) presented bunch-detection results (recall = 0.560 and precision = 0.790) prior to harvest. There are several experimental similarities between these two studies as the research

was conducted by the same study group during the same season (Aquino et al. 2018; Millan et al. 2018). However, the canopy in both these studies had undergone a leaf removal treatment, thus exposing the bunches. Similarly, Hacking et al. (2019) presented comparable results (accuracy = 0.932 and F1-score = 0.833) at plant-level, where the canopy had also undergone leaf removal.

Evaluating the KMC and CT techniques at bunch- and plant-level yielded extremely similar results. The unsupervised clustering of the KMC technique ultimately outperforms the manual CT technique, due to the CT technique being dependent on human-selected thresholds. Unsupervised classifications tend to struggle in 'busy images'; for example, a vineyard with varying structural properties (Diago et al. 2012). The incorporation of the white background, specifically at plant-level, to improve the image contrast might have potentially improved the KMC's ability for successful segmentation. Inevitably, the bunch-level segmentation performed better than the plant-level segmentation, as the bunch-level imagery was 'clean' (i.e. simple contrast due to white background).

The multitemporal image segmentation subsequently utilised the KMC technique, attributable to the better performance of the technique. The novelty of this research visually indicated the effect of vèraison on bunch detection when employing pixel-based segmentation techniques that rely on the image's colour properties (Liu, Marden & Whitty 2013). The segmentation results (accuracy = 0.977) achieved in this study prior to vèraison aligned with the results of Aquino et al. (2018), while the segmentation results (accuracy = 0.942) obtained before harvest were an improvement on the results presented by Millan et al. (2018).

4.5.2 Yield estimation: Bunch-level

This study employed a 2-D methodology for yield estimation at bunch-level, with the novelty of a multitemporal analysis spanning 12 weeks from before vèraison right up to harvest. The best yield estimation result (r^2 = 0.790 and RMSE = 26.918 g) for the global dataset (all 50 bunches) was obtained approximately two weeks prior to harvest, on 14 February 2019.

The lowest yield estimation r^2 (0.349 – Figure 4.12) occurred on 10 January 2019, coinciding with the middle of vèraison. Maximum colour variation was therefore present in the bunches. Evidently, vèraison negatively influenced the 2-D methodology used for yield estimation, as the colour properties were incorporated into image segmentation. Although the lowest segmentation results (17 January 2019 – Figure 4.10) were not obtained on the same day, they coincided with the same noticeable trough, attributed to vèraison. The effect of vèraison reduced the area detected per bunch, ultimately reducing the estimated yield during this window. The presented methodology, reliant on pixel-level segmentation for bunch detection, was therefore deemed ineffective during this window,

spanning roughly four weeks: from 27 December 2018 to 24 January 2019. Future research should avoid this period of phenological growth.

The highest yield estimation result (r^2 = 0.790; 14 February 2019) achieved at bunch-level was a slight improvement on the respective result (r^2 = 0.742) of Hacking et al. (2019). The current study implemented several improvements, yielding the higher r^2 value. These included the incorporation of a white background for image contrast during data acquisition, an improved calibration technique, and utilisation of the KMC technique for segmentation – although this presented comparable segmentation results, as outlined in Section 4.4.1. Diago et al. (2015) conducted a similar bunch-level study under laboratory conditions post-harvest. The authors estimated the number of berries, employed as a yield metric for estimating the final yield. The authors achieved a global (various varietals tested) r^2 of 0.840, slightly higher than the r^2 (0.790) obtained in this study. Similarly, Liu, Marden and Whitty (2013) conducted bunch-level yield estimation under laboratory conditions. The authors were assessing various pixel-level yield metrics for yield estimation. They were able to achieve an r^2 of 0.770 when utilising a pixel count metric, the basis behind the pixel area (cm^2) metric employed in this study. While the results of this study fall between those of Liu, Marden and Whitty (2013) and Diago et al. (2015), it is important to remember that this study was conducted in-situ and incorporated an unsupervised KMC technique for image segmentation.

When investigating the global yield estimation results (Figure 4.12), the best r^2 (0.604) prior to vèraison occurred at the onset of ripeness, on 27 December 2018. An alternative approach to yield estimation prior to vèraison was conducted by Millan et al. (2016). The authors estimated the number of flowers per inflorescence, and subsequently used this as a proxy for yield estimation, achieving an r^2 of 0.490. However, the r^2 value increased when the authors incorporated historical data, such as fruit set rate (r^2 = 0.790) and average berry weight (r^2 = 0.910), bypassing the r^2 presented in this study. Evidently, the incorporation of historical data can improve yield estimation results, as in this case. Aquino et al. (2018) investigated berry detection techniques for yield estimation during the bunch development phase, the final stage before the ripening process begins. Here, the authors incorporated historical berry weights during the yield estimation process, yielding an r^2 value of 0.782. Although this result outperformed the global result of this study (r^2 = 0.604), the results in this study per individual row, both row one (r^2 = 0.732) and row three (r^2 = 0.852) performed on par, if not better.

With respect to row two, the general performance was lower than rows one and three. This lowered performance is not necessarily due to the presented methodology, but more likely the reference measurements captured at harvest. Overripe bunches, some with bird damage, negatively influenced

the reference measurements, as indicated in Section 4.4.2.2. The variation that occurred in the results of this study may be attributed to the different phenological stages (evidenced in Figure 4.12) of the rows. Similarly, Aquino et al. (2018) discussed the variation present in grape compactness during the early stages, and how this negatively influenced their results. This is an important aspect for future consideration. Nonetheless, the 2-D methodology implemented in this study present a strong alternative for estimating a vineyard's anticipated yield from data captured prior to vèraison.

4.5.3 Yield estimation: Plant-level

Traditional yield estimation methods incorporate destructive sampling to measure bunch mass, and subsequently infer the mass per vine to estimate vineyard yield (De la Fuente et al. 2015). For example, De la Fuente et al. (2015) presented various yield estimation models that required traditional destructive sampling, achieving a highest r^2 of 0.749. By comparison, the vine-inferred yield estimation results from this study (Section 4.4.2.2), specifically the results presented in Figure 4.18 (vines = 15), achieved a highest global r^2 of 0.702. The acquisition technique omitted the destructive sampling, resolving a fundamental limitation of traditional methods. The global r^2 results (Figure 4.18) ranged between 0.593 and 0.702 throughout the entire multitemporal analysis, aligning with the results (0.600 – 0.730) presented by Nuske et al. (2014), where data was collected prior to vèraison and prior to harvest. Nuske et al. (2014) utilised a berry count algorithm for yield estimation, and captured data at night using artificial illumination. While berry count has been successfully implemented by various authors (Aquino et al. 2018; Diago et al. 2015; Nuske et al. 2014), the logistical requirements for data acquisition at night can be complex.

Hacking et al. (2019) presented plant-level yield estimation results under two canopy treatments: FC (r^2 = 0.779) and LR (r^2 = 0.877). While these both outperformed the global vine-inferred results in this study, the novelty of the inference technique is evident. The influence of the damaged bunches, either overripe or bird-damage, is unaccounted for in these results, and needs to be taken into consideration when evaluating them. Further refinement of the inference technique presents great potential, as illustrated by several r^2 peaks per individual rows, ranging between 0.831 and 0.945. This non-destructive sampling technique for inferring vine yield could hold future operational capabilities.

4.5.4 Operational limitations

Several limitations were noted during the evaluation of the research and the presented 2-D methodology for yield estimation. Limitations were observed during the segmentation process, as well as the yield estimation process – relevant to both bunch- and plant-level estimates. Refining the

below-mentioned limitations may lead to operationalisation of the presented techniques, especially the vine-inferred technique for yield estimation.

Image segmentation was conducted using the unsupervised KMC technique. However, the implementation still required the manual selection of the appropriate cluster post-segmentation, yielding a binary image of bunch and background. Additionally, ROI selection prior to this was a manual process. Future research should build on the presented methodology and aim to completely automate image processing. Alternatively, future work could compute bunch detection at object-level, and not pixel-level. For example, adapting geographic object-based image analysis techniques, thus segmenting the images at an object-level for bunch detection (Blaschke et al. 2014).

Evaluating the yield estimation aspect of this study yielded two noteworthy limitations. The yield estimation results were presented as a global dataset, as well as individual rows. Although this gave insight regarding varying phenological stages between the three rows, it yielded conservative (between 14 and 20 bunches, and 5 to 6 vines) sample sizes, and could be considered statistically insufficient for linear regression. On account of logistical limitations, the global dataset used in this study was restricted, as described in Section 4.3.2.2. This limited the sample size per row. While the results of this study presented phenological variance between the three rows, this is also an inherent factor within vines, an aspect worthy of consideration. Ensuring phenological uniformity among the bunches at the time of sampling is therefore vital for future research.

Damage sustained to bunches by the time of harvest significantly reduced the final reference mass recorded, likely influencing the results. The main damage was overripe bunches, which was attributed to climatic conditions, thus delaying harvest. Bird damage was another cause of bunch damage, an intrinsic aspect of commercial farming. The bunch damage was an unfortunate, uncontrollable circumstance.

4.6 CONCLUSION

This study set out to determine the optimal phenological stage for yield estimation. A novel approach was presented, using 2-D PRS and related CV techniques to estimate bunch mass from multitemporal RGB imagery. Initially, the study evaluated KMC, an unsupervised technique to undertake image segmentation for bunch detection. This was evaluated against CT, a manual technique. Comparatively, the two techniques performed similarly, but the unsupervised clustering capability of the KMC technique removed the human limitation to selecting the appropriate colour thresholds for CT. KMC was utilised for the subsequent multitemporal dataset's segmentation. The results of this study achieved the best yield estimates during the final stages of berry ripening (during sugar

development), which occurred in the final weeks prior to harvest. An alternative stage for yield estimation would be at the end of berry formation, immediately prior to the first signs of vèraison.

For the duration of vèraison, the bunches' colour development reduced yield estimates. Future colour-based research should avoid this phenological window. Yield estimation should be conducted prior to vèraison or harvest – depending on what the estimated yield data will be used for.

CHAPTER 5: RESEARCH DISCUSSION AND CONCLUSION

This chapter serves as the final chapter, concluding the research presented. The key findings of the research are re-examined and contextualised prior to revisiting the aim and objectives of the research. Subsequently, notable limitations are discussed and complemented with future research recommendations.

5.1 KEY FINDINGS

Yield information is an important aspect within precision viticulture. It enables a better understanding of the spatial variability within a vineyard, while providing the vineyard manager and winemaker with strategic, quantifiable data (Aquino et al. 2018). Yield information can be directly extracted from the physical bunch characteristics (Liu, Marden & Whitty 2013). The reference measurements obtained during this research indicated a strong linear relationship between bunch mass (g) and volume (cm^3), which was used to establish the base of this research. At bunch-level, the relationship achieved a minimum (between the two components) r^2 value of 0.971, while the minimum plant-level relationship yielded an r^2 of 0.995 – a near perfect linear relationship (1:1).

The first component (Chapter 3) evaluated the potential of 2-D and 3-D PRS techniques for data acquisition, and the ability for estimating vineyard yield before harvest. Standard CV techniques were employed for data analysis, allowing the research to focus on the two PRS methodologies. A depth sensor (Kinect V1) was employed to acquire RGB-D (3-D) data under laboratory conditions (experiment one), where it performed extremely well for yield estimation (r^2 = 0.950) at bunch-level. By comparison, RGB (2-D) data acquired using a digital camera produced a lower r^2 of 0.889 for yield estimation. Nonetheless, these yield estimation results illustrated the capability of 2-D and 3-D PRS and related CV techniques, outperforming similar studies (Diago et al. 2015; Liu, Marden & Whitty 2013) conducted under laboratory conditions.

Experiment two (bunch-level) and experiment three (plant-level) were conducted in situ, with two canopy treatments: FC and LR. Comparable yield estimation results (r^2) between the 2-D (FC = 0.625 and LR = 0.742) and 3-D (FC = 0.609 and LR = 0.756) PRS methodologies were achieved in experiment two. However, it was observed that the 2-D (RGB) results (FC = 0.779 and LR = 0.877) had outperformed the 3-D (RGB-D) results (FC = 0.487 and LR = 0.594) in experiment three. As experiment three's data acquisition occurred under natural illumination (experiment two used artificial illumination at night), the effect of solar irradiance (Andújar et al. 2016) on the Kinect sensor became apparent through the poor results. The simplicity and success of the RGB methodology for

yield estimation illustrated commercial viability. This methodology was therefore exclusively used in the subsequent component (component two).

Building on from the first component, the second component (Chapter 4) utilised the RGB methodology to determine the optimal phenological stage for yield estimation. 2-D PRS techniques acquired multitemporal RGB data weekly, resulting in a total of 12 datasets over a three-month period. In an attempt to reduce human error and improve the segmentation results obtained in component one (using CT), the KMC technique was applied and compared with the CT technique. Both segmentation approaches yielded comparable accuracies at bunch-level (CT = 0.939; KMC = 0.942) and plant-level (CT = 0.916 and KMC = 0.900). While the results are similar, the unsupervised nature of the KMC technique is favoured, as it overcomes the limitation of manual threshold selection (Dunn & Martin 2004).

The multitemporal analysis produced the best global yield estimation (r^2 = 0.790) approximately two weeks prior to harvest. If the estimated yield of a vineyard is desired early in the season, then an alternative phenological stage could be immediately prior to vèraison (global yield estimation: r^2 = 0.604). However, this result is considered low for operational use. Further refinement of the methodology is necessary for pre-vèraison yield estimation. Additionally, the multitemporal analysis clearly illustrated the effect of vèraison on the RGB methodology for estimating yield, evidenced by the drop in r^2 values (0.349 – being the lowest) during the four-week period from 27 December 2018 (r^2 = 0.604) to 24 January 2019 (r^2 = 0.557). The bunch-level yield estimation results achieved in the second component were an improvement on the in-situ bunch-level results of component one.

5.2 REVISITING THE AIM AND OBJECTIVES

Utilising PRS and related CV techniques for yield estimation has gained traction in recent years, overcoming common limitations of both traditional yield estimation methods and traditional remote sensing techniques (Matese & Di Gennaro 2015). This research aimed to investigate 2-D and 3-D PRS and related CV techniques for yield estimation in a Shiraz vineyard. Two primary objectives were formulated to accomplish this aim, and they became the core aim of component one and two.

Component one (objective one) evaluated the use of 2-D (RGB) and 3-D (RGB-D) PRS and related CV techniques for yield estimation prior to harvest. Both the 2-D and 3-D methodologies (incorporating PRS and CV techniques) were implemented successfully, respectively acquiring RGB and RGB-D datasets across three experiments. Simple CV techniques (such as CT segmentation for bunch detection) were utilised for estimating bunch and vine yield from these datasets. Evaluating the 2-D and 3-D results illustrated the potentiality of PRS and CV techniques for yield estimation.

However, the 2-D methodology was arguably the better of the two and better suited for future research.

The second component (objective two) investigated the use of a 2-D methodology to determine the ideal phenological stage for yield estimation using multitemporal RGB data. The first aspect of this component investigated the unsupervised KMC technique for image segmentation, which produced comparable segmentation results to the manual CT technique. However, due to the KMC technique being unsupervised, it was deemed a better segmentation technique than CT, a manual thresholding technique. The KMC technique was implemented as the segmentation technique for the subsequent multitemporal data. The second aspect of the component, determining the optimal phenological stage for yield estimation, yielded two options: the final stages of berry ripening (in this case, two weeks prior to harvest), and immediately prior to the onset of ripening (vèraison).

The combined results of component one and two – Chapter 3 and Chapter 4 – supported the RGB methodology for yield estimation. This methodology incorporated 2-D PRS techniques for in-situ data acquisition and relevant CV techniques for image processing. Subsequently, linear regression indicated the correlation between the estimated and actual yields, whereby the highest r^2 value of 0.889 was achieved.

The research objectives conceptualised in Section 1.3 were both successfully achieved, thereby fulfilling the aim of this research to investigate 2-D and 3-D PRS and related CV techniques for yield estimation in a Shiraz vineyard.

5.3 LIMITATIONS AND FUTURE RESEARCH RECOMMENDATIONS

The presented 2-D (RGB) and 3-D (RGB-D) results demonstrated the potential of PRS and CV techniques for vineyard yield estimation. However, this research was only conducted on a single cultivar (Shiraz) vineyard with a VSP trellis system. Further investigation is required to assess the viability of the methodology for different cultivars and trellis systems. While conducting this research, several other limitations were noted and require future research.

The limited performance of the Kinect V1 (RGB-D) highlighted the effect of solar irradiance (Andújar et al. 2016) on in-situ data acquisition during daylight. Although the sensor performed well under laboratory conditions, the poor in-situ performance resulted in the decision to not use the RGB-D sensor for the subsequent component of this research. Nevertheless, future research should investigate the use of an improved sensor, such as the Kinect V2 (Wasenmüller & Stricker 2016), for night-time data acquisition utilising artificial illumination. Additionally, analysing the RGB-D data

required significant manual pre-processing; another limitation of the 3-D methodology. It is recommended that future research investigate ways to automate the pre-processing of 3-D data.

In recent years, the use of 2-D PRS and related CV techniques has been a popular choice for yield estimation in the precision viticulture domain (Aquino et al. 2018; Font et al. 2015; Liu et al. 2017; Millan et al. 2018). Typically, RGB imagery is captured with a digital camera and used for estimating yield (Dunn & Martin 2004). The 2-D methodology presented in this research conformed with these trends, yielding improved results with fewer limitations, especially when compared with the 3-D methodology. Component two saw the implementation of various refinements to the 2-D methodology that alleviated the limitations of component one. These refinements included the following: i) the image calibration was changed from a ruler (single dimension calibration) to a square (enabling 2-D calibration); ii) a white background was incorporated for image contrast (reduced background noise in the images); and iii) the unsupervised KMC technique was implemented for image segmentation (overcoming the manual limitation of the CT technique). Some limitations still persist, such as uneven lighting conditions during data acquisition, the manual selection of calibration squares and ROIs, as well as the manual selection of the ‘bunch’ cluster post-segmentation (Section 4.3.3.1). Fine-tuning the methodology could resolve these limitations in the future.

It was noted in component two that the phenological progression between the three rows was approximately three weeks apart, resulting in a lowered global r^2. Future research should be mindful of varying phenological stages, an inherent attribute of vines. Moreover, it was observed in component two that a large portion of the bunches were overripe at harvest (effect of climatic conditions), resulting in a reduced mass (i.e. yield). This inevitably reduced the final results.

The 2-D and 3-D methodologies that were implemented both demonstrated great potential for estimating yield in a vineyard. Be that as it may, operationalising the 2-D methodology is certainly feasible if the above-mentioned limitations are addressed. With refinement, the presented methodology could be a robust, non-destructive, user-friendly and cost-effective solution for yield estimation. Further research could see the 2-D methodology utilised for estimating yield before vèraison and preceding harvest, providing maximum insight at vital stages to the vineyard manager and winemaker.

5.4 CONCLUSION

Yield estimation is an important factor in precision viticulture, enabling a better understanding of the inherent variability within a vineyard, while providing actionable information to the vineyard manager and winemaker. The advancement of PRS techniques has facilitated in-situ data acquisition, overcoming the destructive nature of traditional sampling methods. Improved CV techniques further enable a more automated estimation process. The combination of these techniques is becoming a commercially viable option for yield estimation.

This research investigated both 2-D and 3-D PRS and CV techniques for yield estimation in a Shiraz vineyard. Two key outcomes of the research summarise the core findings. Firstly, the 2-D methodology was better suited for yield estimation. Secondly, the ideal phenological stage for yield estimation was during the final stages of berry ripening, after vèraison. However, an alternative phenological stage, immediately prior to vèraison (onset of ripening), could be suitable for early yield estimation. Further investigation with the presented 2-D methodology could result in operational solutions for estimating vineyard yield. Advanced knowledge of the estimated yield could enable managerial decisions that achieve maximum quantity and quality, while reducing the environmental footprint of agriculture.

[Word count: 25 288]

REFERENCES

Andújar D, Ribeiro A, Fernández-Quintanilla C & Dorado J 2016. Using depth cameras to extract structural parameters to assess the growth state and yield of cauliflower crops. *Computers and Electronics in Agriculture* 122: 67–73.

Aquino A, Millan B, Diago MP & Tardaguila J 2018. Automated early yield prediction in vineyards from on-the-go image acquisition. *Computers and Electronics in Agriculture* 144: 26–36.

Arnó J, Martínez-Casasnovas JA, Ribes-Dasi M & Rosell JR 2009. Review. Precision viticulture. Research topics, challenges and opportunities in site-specific vineyard management. *Spanish Journal of Agricultural Research* 7, 4: 779–790.

Arnó J, Escolà A, Vallès JM, Llorens J, Sanz R, Masip J, Palacín J & Rosell-Polo JR 2013. Leaf area index estimation in vineyards using a ground-based LiDAR scanner. *Precision Agriculture* 14, 3: 290–306.

Arthur D & Vassilvitskii S 2007. k-means++: *The advantages of careful seeding*. SODA '07: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms held 7–9 January 2007, New Orleans, Louisiana.

Bekker SJ, Hoffman JE, Jacobs SM, Strever AE & van Zyl JL 2016. Ecophysiology, vigour, berry and wine characteristics of grape-vines growing on and off heuweltjies. *South African Journal of Enology and Viticulture* 37, 2: 176–192.

Bellvert J, Zarco-Tejada PJ, Girona J & Fereres E 2014. Mapping crop water stress index in a 'Pinot-noir' vineyard: Comparing ground measurements with thermal remote sensing imagery from an unmanned aerial vehicle. *Precision Agriculture* 15, 4: 361–376.

Bengochea-Guevara JM, Andújar D, Sanchez-Sardana FL, Cantuña K & Ribeiro A 2018. A low-cost approach to automatically obtain accurate 3D models of woody crops. *Sensors* 18, 1: 30.

Blackmore S 2003. The role of yield maps in Precision Farming. PhD thesis. Cranfield: Cranfield University.

Blaschke T, Hay GJ, Kelly M, Lang S, Hofmann P, Addink E, Feitosa R, van der Meer F, van der Werff H, van Coillie F & Tiede D 2014. Geographic object-based image analysis - Towards a new paradigm. *ISPRS Journal of Photogrammetry and Remote Sensing* 87: 180–191.

Bramley RGV & Hamilton RP 2004. Understanding variability in winegrape production systems. *Australian Journal of Grape and Wine Research* 10, 1: 32–45.

Chudasama D, Patel T, Joshi S & Prajapati G 2015. Image segmentation using morphological operations. *International Journal of Computer Applications* 117, 18: 16–19.

Cignoni P, Callieri M, Corsini M, Dellepiane M, Ganovelli F & Ranzuglia G 2008. *MeshLab : An open-source mesh processing tool*, 129–136. Scarano V De Chiara R & Erra U (eds). Sixth Eurographics Italian Chapter Conference held 2–4 July 2008, Pisa, Italy. The Eurographics Association.

CloudCompare. 2018. CloudCompare Version 2.10.alpha [GPL Software] [online]. Available from: http://www.cloudcompare.org/. [Accessed 29 November 2018].

Conningarth Economists 2015. *Final Report: Macro-economic impact of the wine industry on the South African economy (also with reference to the impacts on the Western Cape)*. [online]. Available from: http://www.sawis.co.za/info/download/Macro-economic_impact_study_-_Final_Report_Version_4_30Jan2015.pdf. [Accessed 23 September 2019].

Conradie W, Carey V, Bonnardot V, Saayman D & van Schoor L 2002. Effect of eifferent environmental factors on the performance of sauvignon blanc grapevines in the Stellenbosch/Durbanville districts of South Africa. I. Geology, soil, climate, phenology and grape composition. *South African Journal for Enology and Viticulture* 23, 2: 78–91.

Correa C, Valero C, Barreiro P, Diago M & Tardáguila J 2012. *Feature extraction on vineyard by Gustafson Kessel FCM and K-means*. 16th IEEE Mediterranean Electrotechnical Conference. Yasmine Hammamet, Tunisia. IEEE.

Cunha M, Marçal ARS & Silva L 2010. Very early prediction of wine yield based on satellite data from vegetation. *International Journal of Remote Sensing* 31, 12: 3125–3142.

De la Fuente M, Linares R, Baeza P, Miranda C & Lissarrague J 2015. Comparison of different methods of grapevine yield prediction in the time window between fruitset and veraison. *Journal International Sciencedes Sciences de la Vigne et du Vin* 49: 27–35.

Diago MP, Correa C, Millán B, Barreiro P, Valero C & Tardaguila J 2012. Grapevine yield and leaf area estimation using supervised classification methodology on RGB images taken under field conditions. *Sensors* 12, 12: 16988–17006.

Diago MP, Tardaguila J, Aleixos N, Millan B, Prats-Montalban JM, Cubero S & Blasco J 2015. Assessment of cluster yield components by image analysis. *Journal of the Science of Food and Agriculture* 95: 1274–1282.

Díaz CA, Pérez DS, Miatello H & Bromberg F 2018. Grapevine buds detection and localization in 3D space based on Structure from Motion and 2D image classification. *Computers in Industry* 99: 303–312.

Di Gennaro SF, Toscano P, Cinat P, Berton A & Matese A 2019. A low-cost and unsupervised image recognition methodology for yield estimation in a vineyard. *Frontiers in Plant Science* 10: 559.

Dobrowski SZ, Ustin SL & Wolpert JA 2003. Grapevine dormant pruning weight prediction using remotely sensed data. *Australian Journal of Grape and Wine Research* 9, 3: 177–182.

Doraiswamy PC, Hatfield JL, Jackson TJ, Akhmedov B, Prueger J & Stern A 2004. Crop condition and yield simulations using Landsat and MODIS. *Remote Sensing of Environment* 92: 548–559.

Dunn GM & Martin SR 2004. Yield prediction from digital image analysis: A technique with potential for vineyard assessments prior to harvest. *Australian Journal of Grape and Wine Research* 10, 3: 196–198.

El-Khoury D. 2019. Use of Remote Sensing in Natural Resource Management [online]. Available from: https://staff.aub.edu.lb/~webeco/rs lectures.htm. [Accessed 22 November 2019].

European Parliament and the Council of the European Union 2013. Regulation (EU) 1306/2013 of the European Parliament and of the Council of 17 December 2013 on the financing, management and monitoring of the common agricultural policy and repealing Council Regulations (EEC) No 352/78, (EC) No 165/94, (EC) No 2799/98, (EC) No 814/2000, (EC) No 1290/2005 and (EC) No 485/2008. *Official Journal of the European Union* 56, L 347: 549–607.

Ferreira JHS & Marais PG 1987. Effect of rootstock cultivar, pruning method and crop load on Botrytis cinerea rot of Vitis vinifera cv. Chenin blanc grapes. *South African Journal of Enology & Viticulture* 8, 2: 41–44.

Ferrer M, Echeverría G, Pereyra G, Gonzalez-Neves G, Pan D & Mirás-Avalos JM 2019. Mapping vineyard vigor using airborne remote sensing: relations with yield, berry composition and sanitary status under humid climate conditions. *Precision Agriculture*: 1–20.

Floris-Samuels B 2018. *Status of wine-grape vines as on 31 December 2018*. [online]. Available from: http://www.sawis.co.za/info/download/Vineyards_2018_final.pdf. [Accessed 20 September 2019].

Font D, Tresanchez M, Martínez D, Moreno J, Clotet E & Palacín J 2015. Vineyard yield estimation based on the analysis of high resolution images obtained with artificial illumination at night. *Sensors* 15, 4: 8284–8301.

Fuentes S, De Bei R, Pech J & Tyerman S 2012. Computational water stress indices obtained from thermal image analysis of grapevine canopies. *Irrigation Science* 30, 6: 523–536.

Fuentes S, Poblete-Echeverria C, S O-F, Tyerman S & De Bei R 2014. Automated estimation of leaf area index from grapevine canopies using cover photography, video and computational analysis methods. *Australian Journal of Grape and Wine Research* 20, 3: 465–473.

Gallo R, Ristorto G, Daglio G, Massa N, Berta G, Lazzari M & Mazzetto F 2017. New solutions for the automatic early detection of diseases in vineyards through ground sensing approaches integrating LiDAR and optical sensors. *Chemical Engineering Transactions* 58: 673–678.

Gonzalez R, Woods R & Eddins S 2009. *Digital image processing using MATLAB*. s.l.: Gatesmark.

Grossetête M, Berthoumieu Y, Da Costa J-P, Germain C, Lavialle O & Grenier G 2012. *Early estimation of vineyard yield: Site specific counting of berries by using a Smartphone*. International Conference on Agriculture Engineering held 8–12 July 2012, Valencia, Spain.

Hacking C, Poona N, Manzan N & Poblete-Echeverría C 2019. Investigating 2-D and 3-D proximal remote sensing techniques for vineyard yield estimation. *Sensors* 19, 17: 3652.

Hall A, Louis JP & Lamb DW 2008. Low-resolution remotely sensed images of winegrape vineyards map spatial variability in planimetric canopy area instead of leaf area index. *Australian Journal of Grape and Wine Research* 14, 1: 9–17.

Hall A, Lamb DW, Holzapfel B & Louis J 2002. Optical remote sensing applications in viticulture: A review. *Australian Journal of Grape and Wine Research* 8: 36–47.

Hall A, Lamb DW, Holzapfel BP & Louis JP 2011. Within-season temporal variation in correlations between vineyard canopy and winegrape composition and yield. *Precision Agriculture* 12, 1: 103–117.

Herrero-Huerta M, González-Aguilera D, Rodriguez-Gonzalvez P & Hernández-López D 2015. Vineyard yield estimation by automatic 3D bunch modelling in field conditions. *Computers and Electronics in Agriculture* 110: 17–26.

Ivorra E, Sánchez AJ, Camarasa JG, Diago MP & Tardaguila J 2015. Assessment of grape cluster yield components based on 3D descriptors using stereo vision. *Food Control* 50: 273–282.

Jay S, Rabatel G, Hadoux X, Moura D & Gorretta N 2015. In-field crop row phenotyping from 3D modeling performed using Structure from Motion. *Computers and Electronics in Agriculture* 110: 70–77.

Johnson LF, Bosch DF, Williams DC & Lobitz BM 2001. Remote sensing of vineyard management zones: Implications for wine quality. *Applied Engineering in Agriculture* 17, 4: 557–560.

Kazhdan M, Bolitho M & Hoppe H 2006. *Poisson Surface Reconstruction*, 61–70. SGP '06 Proceedings of the fourth Eurographics symposium on Geometry processing held 26–28 June 2006, Cagliari, Italy. The Eurographics Association.

Kazhdan M & Hoppe H 2013. Screened Poisson surface reconstruction. *ACM Transactions on Graphics* 32, 3: 29.

Klodt M, Herzog K, Töpfer R & Cremers D 2015. Field phenotyping of grapevine growth using dense stereo reconstruction. *BMC Bioinformatics* 16, 1: 143.

Kuhn M 2008. Building predictive models in R using the caret package. *Journal of Statistical Software* 28, 5: 1–26.

Labbe M 2018. RTAB-Map. [online]. Available from: http://introlab.github.io/rtabmap/. [Accessed 1 December 2018].

Lafarge T & Pateiro-Lopez B 2017. alphashape3d: Implementation of the 3D Alpha-shape for the reconstruction of 3D sets from a point cloud. *R package version 1.3*. [online]. Available from: http://cran.r-project.org/package=alphashape3d. [Accessed 6 May 2019].

Lafarge T, Pateiro-López B, Possolo A & Dunkers JP 2014. R implementation of a polyhedral approximation to a 3D set of points using the α -shape. *Journal of Statistical Software* 56, 4: 1–19.

Liu S, Cossell S, Tang J, Dunn G & Whitty M 2017. A computer vision system for early stage grape yield estimation based on shoot detection. *Computers and Electronics in Agriculture* 137: 88–101.

Liu S, Marden S & Whitty M 2013. *Towards automated yield estimation in viticulture*. Australasian Conference on Robotics and Automation held 2–4 December 2013, Sydney, Australia.

Liu S & Whitty M 2015. Automatic grape bunch detection in vineyards with an SVM classifier. *Journal of Applied Logic* 13, 4: 643–653.

Liu S, Whitty M & Cossell S 2015a. *A lightweight method for grape berry counting based on automated 3D bunch reconstruction from a single image*. International Conference on Robotics and Automation (IEEE) (Workshop on Robotics in Agriculture: 4) held 2–6 February 2015, Sydney, Australia.

Liu S, Whitty M & Cossell S 2015b. *Automatic grape bunch detection in vineyards for precise yield estimation*, 238–241. 14th IAPR International Conference on Machine Vision Applications (MVA) held 18–22 May 2015, Tokyo, Japan. MVA organization.

Liu S, Li X, Wu H, Xin B, Petrie PR & Whitty M 2018. A robust automated flower estimation system for grape vines. *Biosystems Engineering* 172: 110–123.

Loggenberg K, Strever A, Greyling B & Poona N 2018. Modelling water stress in a Shiraz vineyard using hyperspectral imaging and machine learning. *Remote Sensing* 10, 2: 202.

Lopes C, Torres A, Guzman R, Graca J, Reyes M, Vitorino G, Braga R, Monteiro A & Barriguinha A 2017. *Using an unmanned ground vehicle to scout vineyards for non-intrusive estimation of canopy features and grape yield*, 16–21. 20th GiESCO International Meeting held 5–9 November 2017, Mendoza, Argentina.

Luo L, Tang Y, Zou X, Wang C, Zhang P & Feng W 2016. Robust grape cluster detection in a vineyard by combining the AdaBoost framework and multiple color components. *Sensors* 16, 12: 2098.

Luque A, Carrasco A, Martín A & de las Heras A 2019. The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition* 91: 216–231.

Marinello F, Pezzuolo A, Meggio F, Martinez-Casasnovas J, Yezekyan T & Sartori L 2017. Application of the Kinect sensor for three dimensional characterization of vine canopy. *Advances in Animal Biosciences* 8, 2: 525–529.

Marinello F, Pezzuolo A, Cillis D & Sartori L 2016. *Kinect 3D reconstruction for quantification of grape bunches volume and mass*. Engineering for Rural Development Conference held 20–22 May 2016, Jelgava, Latvia.

Martín P, Zarco-Tejada PJ, González MR & Berjón A 2007. Using hyperspectral remote sensing to map grape quality in "Tempranillo" vineyards affected by iron deficiency chlorosis. *Vitis - Journal of Grapevine Research* 46, 1: 7–14.

Matese A, Baraldi R, Berton A, Cesaraccio C, Di Gennaro S, Duce P, Facini O, Mameli MG, Piga A & Zaldei A 2018. Estimation of water stress in grapevines using proximal and remote sensing methods. *Remote Sensing* 10: 114.

Matese A & Di Gennaro SF 2015. Technology in precision viticulture: A state of the art review. *International Journal of Wine Research* 7: 69–81.

Matese A, Toscano P, Di Gennaro SF, Genesio L, Vaccari FP, Primicerio J, Belli C, Zaldei A, Bianconi R & Gioli B 2015. Intercomparison of UAV, aircraft and satellite remote sensing platforms for precision viticulture. *Remote Sensing* 7, 3: 2971–2990.

Mathews AJ 2015. A practical UAV remote sensing methodology to generate multispectral orthophotos for vineyards : estimation of spectral reflectance using compact digital cameras. *International Journal of Aplied Geospatial Research* 6, 4: 65–87.

Mathews AJ & Jensen JLR 2012. An airborne LiDAR-based methodology for vineyard parcel detection and delineation. *International Journal of Remote Sensing* 33, 16: 5251–5267.

Mathews AJ & Jensen JLR 2013. Visualizing and quantifying vineyard canopy LAI using an unmanned aerial vehicle (UAV) collected high density structure from motion point cloud. *Remote Sensing* 5: 2164–2183.

Microsoft. 2013. Kinect for Windows SDK 1.8. [online]. Available from: https://www.microsoft.com/en-us/download/details.aspx?id=40276. [Accessed 16 May 2019].

Millan B, Aquino A, Diago MP & Tardaguila J 2016. Image analysis-based modelling for flower number estimation in grapevine. *Journal of the Science of Food and Agriculture* 97, 3: 784–792.

Millan B, Diago MP, Aquino A, Palacios F & Tardaguila J 2019. Vineyard pruning weight assessment by machine vision: towards an on-the-go measurement system. *OENO One* 53, 2: 307–319.

Millan B, Velasco-Forero S, Aquino A & Tardaguila J 2018. On-the-go grapevine yield estimation using image analysis and boolean model. *Journal of Sensors* 2018: 1–14.

Mulla DJ 2013. Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems Engineering* 114, 4: 358–371.

Nasa 2019. A Landsat Timeline [online]. Available from: https://landsat.gsfc.nasa.gov/. [Accessed 30 October 2019].

Nuske S, Achar S, Bates T, Narasimhan S & Singh S 2011. *Yield estimation in vineyards by visual grape detection*, 2352–2358. IEEE International Conference on Intelligent Robots and Systems held 25–30 September 2011, San Francisco, USA. IEEE.

Nuske S, Wilshusen K, Achar S, Yoder L, Narasimhan S & Singh S 2014. Automated visual yield estimation in vineyards. *Journal of Field Robotics* 31, 5: 837–860.

Priori S, Martini E & Costantini E 2010. *Three proximal sensors for mapping skeletal soils in vineyards*, 121–124. 19th World Congress of Soil Science, Soil Solutions for a Changing World held 1–6 August 2010 Brisbane, Australia.

Priori S, L'Abate G, Fantappie M & Costantini E 2018. Mapping soil spatial variability at high detail by proximal sensors for new vineyard planning. *International Journal of Environmental Quality* 30: 9–15.

Quan Q, Lanlan T, Xiaojun Q, Kai J & Qingchun F 2017. *Selecting candidate regions of clustered tomato fruits under complex greenhouse scenes using RGB-D data*, 389–393. 3rd International Conference on Control, Automation and Robotics, held 22–24 April 2017, Nagoya, Japan. IEEE.

R Core Team 2019. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. [online]. Available from: https://www.r-project.org/. [Accessed 5 March 2019].

Reis MJCS, Morais R, Peres E, Pereira C, Contente O, Soares S, Valente A, Baptista J, Ferreira PJSG & Bulas Cruz J 2012. Automatic detection of bunches of grapes in natural environment from color images. *Journal of Applied Logic* 10, 4: 285–290.

Rey-Caramés C, Diago MP, Martín MP, Lobo A & Tardaguila J 2015. Using RPAS multi-spectral imagery to characterise vigour, leaf development, yield components and berry composition variability within a vineyard. *Remote Sensing* 7: 14458–14481.

Ribeiro A, Bengochea-Guevara JM, Conesa-Muñoz J, Nuñez N, Cantuña K & Andújar D 2017. RGB-D monitoring of woody crops using an unmanned ground vehicle. *Advances in Animal Biosciences* 8, 2: 210–215.

Rose JC, Kicherer A, Wieland M, Klingbeil L, Töpfer R & Kuhlmann H 2016. Towards automated large-scale 3D phenotyping of vineyards under field conditions. *Sensors* 16, 12: 2136.

Rueda-Ayala V, Peña J, Höglind M, Bengochea-Guevara J & Andújar D 2019. Comparing UAV-based technologies and RGB-D reconstruction methods for plant height and biomass monitoring on grass ley. *Sensors* 19, 3: 535.

Rydberg A & Borgefors G 2001. Integrated method for boundary delineation of agricultural fields in multispectral satellite images. *IEEE Transactions on Geoscience and Remote Sensing* 39, 11: 2514–2520.

SAWIS 2018. South African Wine Industry 2018 Statistics Nr. 43. 43. [online]. Available from: http://www.sawis.co.za/. [Accessed 25 September 2019].

SAWIS 2019. SA Wine Industry Information and Systems [online]. Available from: http://www.sawis.co.za/. [Accessed 25 September 2019].

Schultz B, Immitzer M, Formaggio AR, Del' Acro Sanches I, Luiz A & Atzberger C 2015. Self-guided segmentation and classification of multi-temporal Landsat 8 images for crop type mapping in Southeastern Brazil. *Remote Sensing* 7: 14482–14508.

Smit JL, Sithole G & Strever AE 2010. Vine signal extraction - an application of remote sensing in precision viticulture. *South African Journal of Enology and Viticulture* 31, 2: 65–74.

Stellenbosch University 2013. Experimental Farms [online]. Available from: http://www.sun.ac.za/english/faculty/agri/departments1/experimental-farms. [Accessed 28 September 2019].

Sun L, Gao F, Anderson MC, Kustas WP, Alsina MM, Sanchez L, Sams B, Mckee L, Dulaney W, White WA, Alfieri JG, Prueger JH, Melton F & Post K 2017. Daily mapping of 30 m LAI and NDVI for grape yield prediction in California vineyards. *Remote Sensing* 9: 317.

Tang J, Woods M, Cossell S, Liu S & Whitty M 2016. Non-productive vine canopy estimation through proximal and remote sensing. *IFAC-PapersOnLine* 49, 16: 398–403.

Taylor J, Tisseyre B, Bramley R & Reid A 2005. A comparison of the spatial variability of vineyard yield in European and Australian production systems. *Precision Agriculture*: 907–914.

The MathWorks Inc. 2018. Matlab R2018b v9.5.0.944444. [online]. Available from: https://ww2.mathworks.cn/en/. [Accessed 8 January 2019].

The MathWorks Inc. 2019. imsegkmeans: K-means clustering based image segmentation [online]. Available from: https://www.mathworks.com/help/images/ref/imsegkmeans.html?searchHighlight=imsegkmeans&s_tid=doc_srchtitle. [Accessed 11 November 2019].

timeanddate.com 2019. The Golden Hour - When Sunlight Turns Magical [online]. Available from: https://www.timeanddate.com/astronomy/golden-hour.html. [Accessed 15 September 2019].

Towers PC, Strever A & Poblete-Echeverría C 2019. Comparison of vegetation indices for leaf area index estimation in vertical shoot positioned vine canopies with and without grenbiule hail-protection netting. *Remote Sensing* 11, 9: 1073.

Usha K & Singh B 2013. Potential applications of remote sensing in horticulture - A review. *Scientia Horticulturae* 153: 71–83.

Wang W & Li C 2014. Size estimation of sweet onions using consumer-grade RGB-depth sensor. *Journal of Food Engineering* 142: 153–162.

Wasenmüller O & Stricker D 2016. *Comparison of Kinect V1 and V2 depth images in terms of accuracy and precision*. ACCV Workshops. Taipei, Taiwan: 34–45.

Weiss M & Baret F 2017. Using 3D point clouds derived from UAV RGB imagery to describe vineyard 3D macro-structure. *Remote Sensing* 9, 2: 111.

Wójtowicz M, Wójtowicz A & Piekarczyk J 2016. Application of remote sensing methods in agriculture. *Communications in Biometry and Crop Science* 11, 1: 31–50.

Wolpert JA & Vilas EP 1992. Estimating vineyard yields: Introduction to a simple, two-step method. *American Journal of Enology and Viticulture* 43, 4: 384–388.