

Vineyard yield estimation using 2-D proximal sensing: a multitemporal approach

Chris Hacking¹, Nitesh Poona¹ and Carlos Poblete-Echeverría^{2,*}

¹Department of Geography and Environmental Studies, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa

²South African Grape And Wine Research Institute (SAGWRI), Department of Viticulture and Oenology, Faculty of AgriSciences, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa

*corresponding author: cpe@sun.ac.za

ABSTRACT

Vineyard yield estimation is a fundamental aspect in precision viticulture that enables a better understanding of the inherent variability within a vineyard. Yield estimation conducted early in the growing season provides insightful information to ensure the best fruit quality for the maximum desired yield. Proximal sensing techniques provide 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-dimensional (2-D) proximal sensing and computer vision techniques in a vertical shoot positioned (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. Preceding the multitemporal analysis for yield estimation, an unsupervised *k*-means clustering (KMC) algorithm was evaluated for image segmentation on the final dataset captured before harvest, yielding bunch-level segmentation accuracies as high as 0.942, with a corresponding F1-score of 0.948. The segmentation yielded a pixel area (cm²), which served as input to a cross-validation model for calculating bunch mass (g). The ‘calculated mass’ was linearly regressed against the ‘actual mass’, indicating the capability for estimating vineyard yield. Results of the multitemporal analysis showed that the final stage of berry ripening was the ideal phenological stage for yield estimation, achieving a global *r*² of 0.790.

KEYWORDS

proximal remote sensing, computer vision, yield estimation, precision viticulture, image segmentation, unsupervised classification, colour thresholding

INTRODUCTION

Yield estimation is fundamental for 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 (e.g. 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). Due to the limitations of traditional methods, the combination of proximal sensing and computer vision techniques has been investigated as an alternative for yield estimation. Studies have been conducted early in the season (Liu *et al.*, 2017), around bunch closure (Aquino *et al.*, 2018) and prior to harvest (Millan *et al.*, 2018). Proximal sensing incorporates modern terrestrial sensors to capture 2-dimensional (2-D) and 3-dimensional (3-D) datasets in a non-destructive, cost-effective manner (Font *et al.*, 2015; Marinello *et al.*, 2016). Researchers have applied computer vision techniques for fruit characterisation and yield estimation (Mochida *et al.*, 2018). Computer vision incorporates multiple techniques of image processing to interpret and extract accurate information from digital datasets (Gonzalez *et al.*, 2009). Computer vision techniques – including segmentation, feature extraction and classification – are ideal for extracting information from raw datasets in agriculture (Mochida *et al.*, 2018; Tian *et al.*, 2020). Tian *et al.* (2020) have presented a detailed review of computer vision in agriculture.

In recent years, several authors (e.g. Aquino *et al.*, 2018; Diago *et al.*, 2015; Liu *et al.*, 2013; Millan *et al.*, 2018) have highlighted the potential of 2-D red, green, and blue (RGB) proximal sensing and related computer vision techniques for yield estimation. A limited number of studies have investigated 3-D RGB-Depth (RGB-D) proximal sensing and computer vision 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,

significantly 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 potential or in situ data collection. By contrast, 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). The authors did not note significant natural illumination interference from the RGB methodology for in-situ data collection, concluding that the RGB methodology was better suited for yield estimation – specifically in-situ acquisition (Hacking *et al.*, 2019). Changes in natural illumination generate some inconsistencies in the imagery analyses when in-situ RGB data collection is implemented (Font *et al.*, 2015). In this regard some authors suggest the use of images acquired at night with external light sources (Nuske *et al.*, 2014; Reis *et al.*, 2012) or the standardisation of natural illumination using a constant methodology for image acquisition, i.e. time during the day, solar angle, distance, and climatic conditions (Hacking *et al.*, 2019). Yield estimation using 2-D RGB data uses one of two common computer vision approaches: bunch detection (image segmentation) or berry detection. Image segmentation (e.g. Font *et al.*, 2015; Liu & Whitty, 2015), applied at pixel-level, differentiates bunches from the background. A relevant bunch metric, e.g. a pixel count metric (Liu *et al.*, 2013), is then employed to estimate the final yield. Berry detection techniques (e.g. Grossetête *et al.*, 2012; Nuske *et al.*, 2014; Zabawa *et al.*, 2020) generally bypass the pixel-level segmentation (bunch detection) and bunch metric approach. In this instance, individual berries are detected and counted for estimating the yield, commonly incorporating a historical berry mass during the estimation process (Nuske *et al.*, 2014). A limitation of berry detection algorithms is the requirement for berry mass data, which guides the estimation process during computation (Aquino *et al.*, 2018). Both computer vision techniques require bunch or berry detection prior to yield estimation. However, bunch detection via image segmentation is more widely applied (Font *et al.*, 2015; Hacking *et al.*, 2019; Millan *et al.*, 2018).

Image segmentation is one of the biggest challenges in yield estimation using proximal sensing and related computer vision techniques (Millan *et al.*, 2018). Segmentation methods generally use some form of colour thresholding

(CT) or image classification to classify pixel values according to relevant classes (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; S Liu *et al.*, 2013; Millan *et al.*, 2018) have evaluated 2-D proximal sensing and image segmentation for yield estimation pre- and post-harvest. Traditionally, bunch detection has relied on user-selected colour thresholds (known as colour thresholding – CT) for image segmentation by selecting the relevant pixels within the defined thresholds (Dunn & Martin, 2004; Hacking *et al.*, 2019; Reis *et al.*, 2012). CT has been successfully implemented for yield estimation (Diago *et al.*, 2015; Hacking *et al.*, 2019) with in-situ bunch-level segmentation results ranging between $r^2 = 0.625$ (full canopy treatment) and $r^2 = 0.742$ (leaf removal treatment) (Hacking *et al.*, 2019). 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.

More automated approaches employ classification methods for 2-D image segmentation, to reduce the human intervention during bunch detection. Most 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). While supervised segmentation methods still require human interaction to train the classifiers, there is more automated computation applied in the background, enabling higher segmentation accuracies. Examples include the use of an adaptive boosting (AdaBoost) classification framework, accuracy = 0.966 (Luo *et al.*, 2016), and the implementation of the Mahalanobis distance clustering algorithm, accuracy = 0.980 (Diago *et al.*, 2012). An alternative classification approach is an unsupervised classification, which completely forgoes manual training (Correa *et al.*, 2012). For example, k-means clustering (KMC) (Arthur & Vassilvitskii, 2007) is a popular unsupervised technique that computes the average squared distance between pixels to determine suitable clusters. 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) employed unsupervised feature

selection and shoot classification algorithms for estimating vineyard yield from early-stage shoot detection at the start of the growing season. The authors presented an average shoot-detection accuracy of 0.868, with an F1-score of 0.900. However, the study did not employ unsupervised image classification techniques to segment the image. To date, the authors are unaware of any yield estimation research which 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 veraison (the onset of ripening, whereby the bunch undergoes a colour transition), using artificial lighting to detect individual berries. Grossetête *et al.* (2012) achieved an r^2 of 0.920 for berry detection, whereas 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 veraison and prior to harvest (maximum ten-day window preceding harvest). Multitemporal data acquisition enables a more detailed understanding of phenological stages within the vines (Padua *et al.*, 2020). A better understanding of the phenological development during the season is likely to influence proximal sensing capabilities for improved yield estimation results, warranting the investigation of multitemporal yield estimation. This study investigated the use of 2-D proximal sensing and related computer vision techniques for yield estimation in a vertical shoot position (VSP) 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 implement a novel experiment to determine the optimal timeframe (phenological stage) in the growing season for yield estimation in a vineyard using a 2-D methodology.

MATERIALS AND METHODS

1. Study site

The study was conducted at Stellenbosch University’s Welgevallen Experimental Farm, Stellenbosch, South Africa (33° 56’ 26’’ S; 18° 51’

56° E). Data was collected in situ from a Shiraz vineyard (Figure 1), trained to a VSP system. The vineyard was planted in 2000, approximately 157 m above sea level, in a North-South row direction with 2.7×1.5 m vine spacing. Datasets were collected for 16 individual vines 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, typical of a Mediterranean climate (Conradie *et al.*, 2002).

2. Data acquisition

The 2-D image data was captured in situ at bunch-level. Multitemporal data was acquired weekly for 50 individual bunches (bunch-level), over 12 weeks (8 December 2018 to 25 February 2019). 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 (i.e. start of veraison). The final bunch-level dataset was captured on 25 February 2019, yielding a total of 12 datasets captured at bunch-level. Table 1 indicates the E-L phenological stage (Dry & Coombe, 2004) for the 12 data acquisition dates. A total of

16 vines, from which the 50 bunches were sampled, were harvested on 27 February 2019, when final reference measurements were recorded for both bunch- and plant-level under laboratory conditions.

2.1. Reference measurements

Bunch volume (cm³) was recorded per individual bunch (×50 bunches), collected with the weekly RGB data acquisition. Figure 2a 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 script in MATLAB® (v.2018b, The MathWorks Inc., Natick, United States). The difference in the two levels yielded a bunch displacement value, which was used as a proxy for bunch volume (Ferreira & Marais, 1987).

Final reference measurements were recorded for the 16 harvested vines under laboratory conditions. Volume (cm³) (Figure 2b) and mass (g) (Figure 2c) was recorded per bunch, for each of the 50 individual bunches sampled from the 16 vines. Individual bunch mass (g) was recorded using

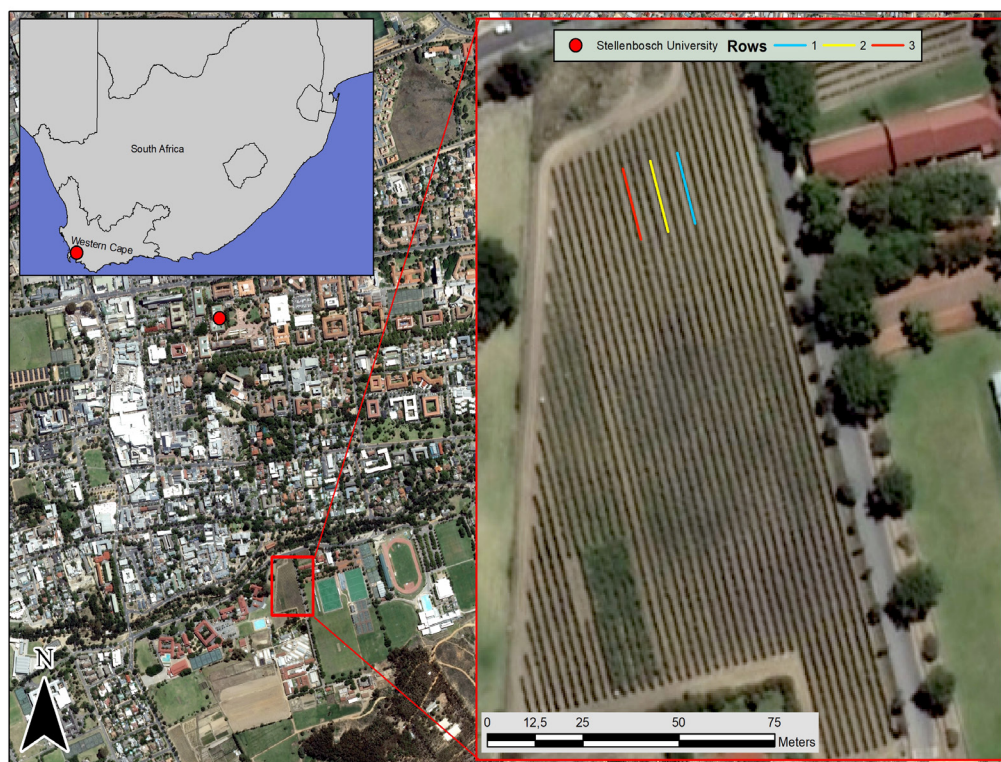


FIGURE 1. Location of the Shiraz vineyard situated on the Welgevallen Experimental Farm in Stellenbosch, South Africa. The red inset map illustrates the location of the three rows within the vineyard.

TABLE 1. Phenological stages for the 12 experimental weeks.

Date	12/8/2018	12/13/2018	12/20/2018	12/27/2018	1/3/2019	1/10/2019
E-L Stage	29	31	32	33	34	35
Date	1/17/2019	1/24/2019	1/31/2019	2/7/2019	2/14/2019	2/25/2019
E-L Stage	35	35	36	37	38	39

Phenological stages are indicated using the E-L scale (Dry & Coombe, 2004).

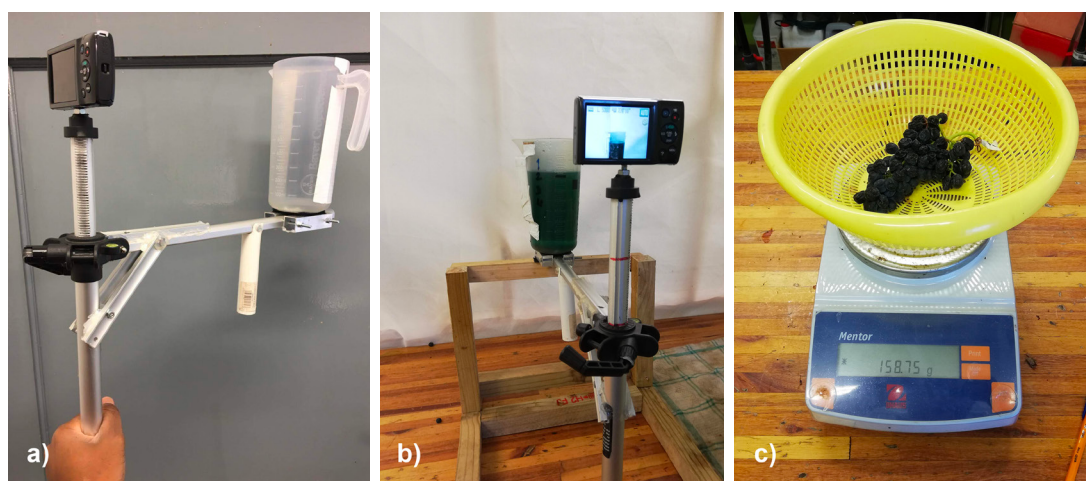


FIGURE 2. Reference measurement systems measuring bunch displacement in the field (a). Laboratory measurements captured bunch displacement (b) and bunch mass (c).



FIGURE 3. Bunch-level data acquisition system (a), with an example captured on 8 December 2018 (b), and the same bunch captured on 25 February 2019 (c). Refer to Table 1 for phenological stage.

a Mentor scale (Ohaus, Parsippany, United States). For consistency, the same volume measurement system (Figure 2a) was used in the laboratory, with the addition of a custom-built base that held the system level and steady (Figure 2b). Individual bunch measurements per vine were aggregated, yielding the respective vine's volume (cm^3) and mass (g).

2.2. Bunch-level image acquisition

Bunch-level data was acquired weekly for 50 individual bunches, statistically sampled from the 16 vines. The number of bunches

per vine was determined shortly after flowering on 21 November 2018. Subsequently, the number of bunches per vine was plotted using a box and whisker diagram, and four vines randomly selected from each quartile, yielding a total of 16 vines. For ease of logistics, 20 % of bunches were selected per individual vine, resulting in either two, three, or four bunches being selected per vine. A total of 52 bunches were selected. However, two bunches broke off during the season. Consequently, a total of 50 bunches were monitored throughout the growing season. Bunch-level RGB imagery was acquired using

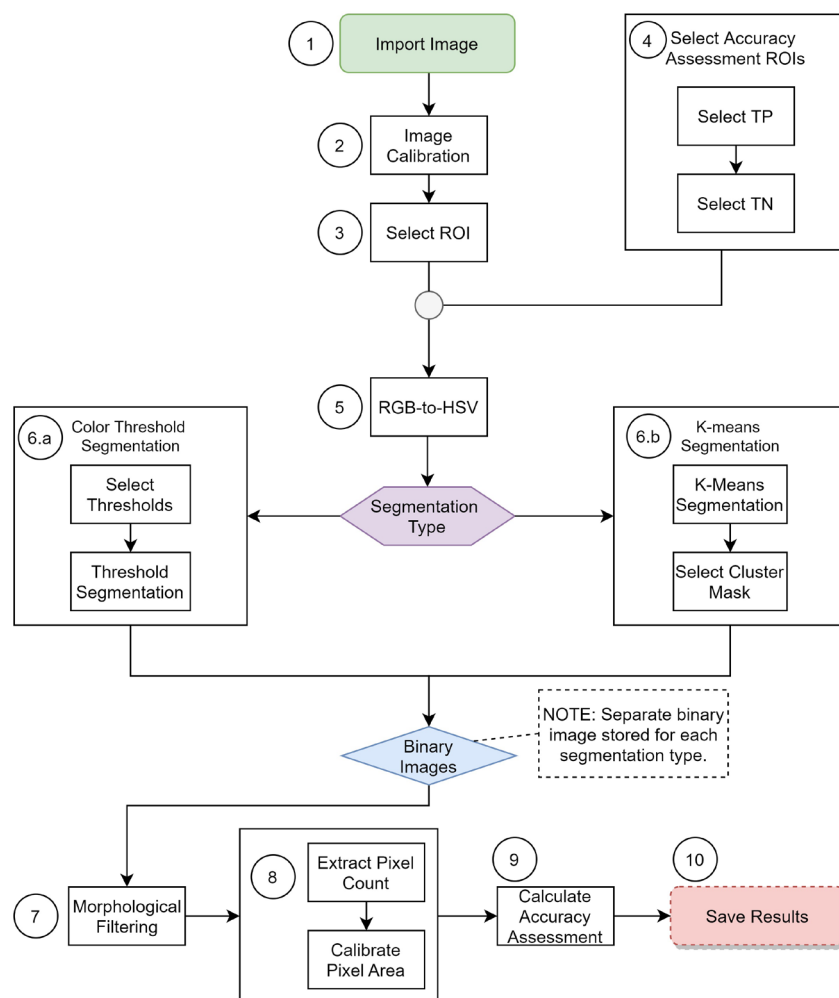


FIGURE 4. Image analysis workflow.

the same system used for collecting the reference volume measurements (Figure 2a), with the inclusion of a white background (Figure 3a). The white background produced a distinct image contrast; an improvement in the methodology presented by 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 (between 08:00 and 10:00) under natural illumination. An umbrella provided shade over the bunch being photographed.

Figure 3b shows an individual bunch captured on 8 December 2018, forming part of the first dataset. The image clearly shows the 1×1 cm black calibration squares (Figure 3b) placed on the white background for scale reference. The calibration squares served 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 illustrated in Figure 3b is illustrated in Figure 3c; captured 25 February 2019, two days prior to harvest. These two images illustrate the bunch development from 8 December 2018 to 25 February 2019.

3. Data analysis

Data analysis was based on the RGB methodology for yield estimation presented by Hacking *et al.* (2019). The first component employed computer vision techniques for image segmentation (Section 3.1. Image segmentation) details the custom script used for image segmentation, with Section 3.2. (Segmentation accuracy assessment) detailing the assessment of segmentation accuracy. The second component (Section 3.3. Yield estimation) conducted a 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 3.4. (Inferred plant-level yield).

3.1. Image segmentation

Image segmentation was undertaken using a custom script in MATLAB®. Figure 4 illustrates the workflow using a custom script, executed 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 3b), yielding a calibrated pixel-area (cm²) coefficient.
3. The region of interest (ROI) was manually digitised to select the area of the image representing bunch, whereby minimal background was included.
4. Segmentation accuracy assessment ROIs (Regions of interest); based on a selection of five 'bunch' and five 'background' ROIs that were manually digitised. Details are provided in Section 3.2. (Segmentation accuracy assessment).
5. Image converted from RGB to hue, saturation, and value (HSV) colour space.

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, where colour thresholds are manually selected for segmentation. The KMC technique requires a *k*-value that defines the number of clusters in the segmented 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 Step 5 as follows:

6. Image segmentation:
 - a. CT required the manual determination of suitable threshold limits based on the HSV colour space, using MATLAB's *Colour*

Threshold app, included in the *Image Processing Toolbox*™ (The MathWorks Inc., 2018). For each dataset processed, the first image per row was used for selecting the threshold limits (i.e. lower and upper thresholds in H, S, and V), which was then applied to the entire row. The thresholds were empirically selected using the *Colour Threshold* app, which automatically illustrates the segmentation output in real-time according to the selected thresholds. The output was a segmented binary image, i.e. bunch and background.

- b. The '*imsegkmeans*' (i.e. KMC) algorithm (Arthur & Vassilvitskii, 2007), part of the Image Segmentation section in MATLAB's *Image Processing Toolbox*™ (The MathWorks Inc., 2018), applied an empirically selected value of '*k* = 3' (three clusters: bunch, background, and ROI excluded background) for the bunch-level segmentation. Visual interpretation confirmed the suitability of the chosen value for *k*, yielding the respective three clusters per image. The remaining parameters of the algorithm were set to their default values (The MathWorks Inc., 2019). Using default values was deemed sufficient for the purpose of this study. The KMC algorithm utilises the HSV properties of the image for clustering. Post-classification, the appropriate cluster mask was manually selected, yielding a segmented binary image, i.e. bunch and background.

Stage 3: Post-segmentation processing

7. A sequence of morphological operations; dilation followed by erosion, were applied to the segmented binary images. Dilation 'grows' or 'thickens' the binary image effectively filling any holes, whereas erosion 'shrinks' or 'thins' the binary image removing any outliers. The morphological operators used a 2-D 'disk' structure with a radius of '10' pixels for image refinement (Gonzalez *et al.*, 2009).
8. All pixels representing 'bunch' in the binary image were counted. Bunch area (cm²) was determined using the number of pixels, and the calibrated pixel-area coefficient obtained from Step 2 in the Stage 1.
9. Segmentation accuracy was calculated for both segmentation techniques (Section 3.2. Segmentation accuracy assessment).
10. Results were exported and saved.

Table 2 summarises the datasets utilised for image segmentation. Segmentation accuracy was calculated for each individual image as detailed in Section 3.2. (Segmentation assessment).

3.2. Segmentation accuracy assessment

Segmentation accuracy was calculated per image, using the custom script detailed in Section 3.1. (Image segmentation). The ROIs defined for assessing segmentation accuracy (Step 4, Section 3.1. Image segmentation) represented actual (reference) pixel values for bunch and background. These values were utilised for assessing the segmentation techniques (Step 9, Section 3.1. Image segmentation).

Several accuracy assessment metrics (Liu & Whitty, 2015) were evaluated using a confusion matrix (Table 3), 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.

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 1 (Liu & Whitty, 2015):

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP}$$

Recall calculates the percentage of pixels correctly classified as bunch from the manually labelled bunch-pixels; defined by Equation 2 (Aquino *et al.*, 2018):

$$Recall = \frac{TP}{TP + FN}$$

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 3 (Aquino *et al.*, 2018):

$$Precision = \frac{TP}{TP + FP}$$

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 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):

$$F1 - score = 2 \times \frac{Recall \times Precision}{Recall + Precision}$$

3.3. Yield estimation

A custom yield estimation script that integrated the *Caret* package (Kuhn, 2008) for cross-validation, was compiled using R statistical software (v3.5.2, R Core Team, 2019, Vienna, Austria). Five-fold cross-validation, repeated ten times for robustness, was employed for developing the yield estimation

TABLE 2. Summary of datasets utilised during segmentation.

Dataset	Number of images	Resolution
Harvest: Bunch-level	50 (single date)	5152 × 3864 pixels
Multitemporal: Bunch-level	50 (12 dates – inclusive of the harvest dataset)	5152 × 3864 pixels

TABLE 3. Confusion matrix for binary classification.

		Actual	
		Bunch	Background
Predicted	Bunch	TP	FP
	Background	FN	TN

Adapted from Luque *et al.*, 2019 (p. 218)

model. This was completed for each individual data acquisition date. Cross-validation was executed and evaluated across three steps:

1. Bunch area (cm²) was linearly regressed against the respective volume measurement (cm³) in the first cross-validation model. The yield estimation model started with this step to include the multitemporal reference measurements (bunch volume). The model produced ‘fitted’ volume (cm³) values.
2. The ‘fitted’ volume (cm³) 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, quantifying the yield estimation error in grams.

3.4. Inferred plant-level yield

The plant-level yield was inferred from the estimated bunch-level yield, determined for the relevant date from the multitemporal datasets. Equation 5 was formulated from the traditional equation (Komm & Moyer, 2015) employed for yield estimation to serve this purpose:

$$V_m = \frac{\sum_{m=1}^n B_m}{n} \times N$$

where V_m represents the inferred vine-mass calculated from the sum of the individual bunch masses (B_m), divided by the number of bunches monitored (n) on the respective vine, thus 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 using V_m and the actual vine mass (g) determined at harvest.

RESULTS

The results presented in this section are based on the linear relationship between mass (g) and volume (cm³) established by Hacking *et al.* (2019). This relationship sets the context for the results presented in this study.

1. Segmentation results

1.1. Harvest: Bunch-level

Table 4 presents the segmentation results for the two techniques evaluated at bunch-level. The KMC technique achieved a segmentation accuracy of 0.942 and F1-score of 0.948 (Figure 5c), outperforming the CT technique, which achieved an accuracy of 0.939; and F1-score of 0.943 (Figure 5b). The higher accuracy achieved using the KMC technique may be attributed to the KMC technique being unsupervised, and therefore more robust than the CT technique, which is dependent on user-selected thresholds. It is evident from Figure 5b that the thresholds manually selected during the CT technique were not completely inclusive of the entire bunch in certain situations, resulting in misclassification of bunch as background.

TABLE 4. Segmentation accuracy results for bunch detection (50 bunches) on 25 February 2019.

	Accuracy	Precision	Recall	F1-Score
CT	0.939	0.976	0.933	0.943
KMC	0.942	0.954	0.959	0.948



FIGURE 5. Visual representation of segmentation results at bunch-level: raw image (a), CT (b), and KMC (c).

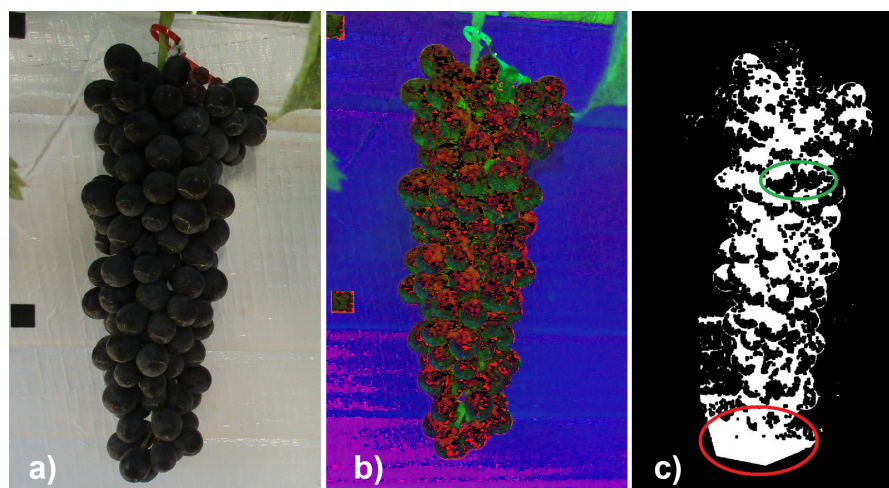


FIGURE 6. Example of precision and recall errors; original RGB image (a), converted HSV image (b), and segmented image (c).

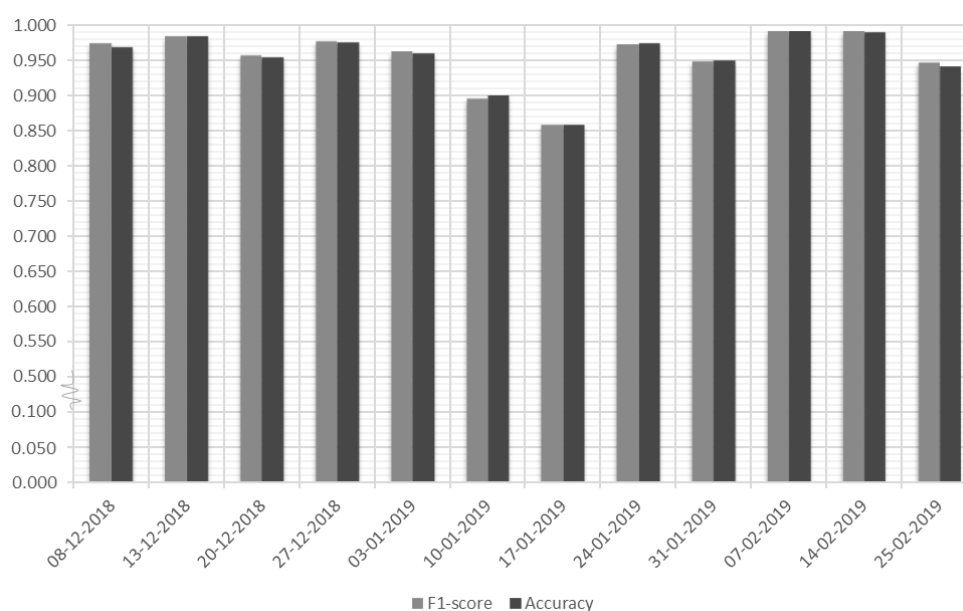


FIGURE 7. Segmentation results for KMC. Take note of the truncated y-axis for visualisation. See Table 1 for phenological information.

An evaluation of the precision and recall results illustrate that the CT technique produced fewer FPs compared with the KMC technique. The CT technique achieved a higher precision of 0.976 (Table 4). FPs indicate that the background has been incorrectly classified as bunch, illustrated by the area circled in red in Figure 6c. On the contrary, the KMC technique was more effective at correctly segmenting the bunch, yielding more TPs and fewer FNs compared with the CT technique; illustrated 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 6c – red circle). Effectively,

portions of the bunch were incorrectly classified as background, i.e. the black pixels within the bunch (Figure 6c – green circle). Segmentation errors may be attributed to inherent variability in the HSV within the image (Figure 6b).

1.2. Multitemporal: Bunch-level

Upon evaluating the segmentation results, the decision was made to solely utilise the KMC technique for the multitemporal bunch-level analysis, applied to all subsequent results. Figure 7 presents the multitemporal segmentation results (accuracy and F1-score)

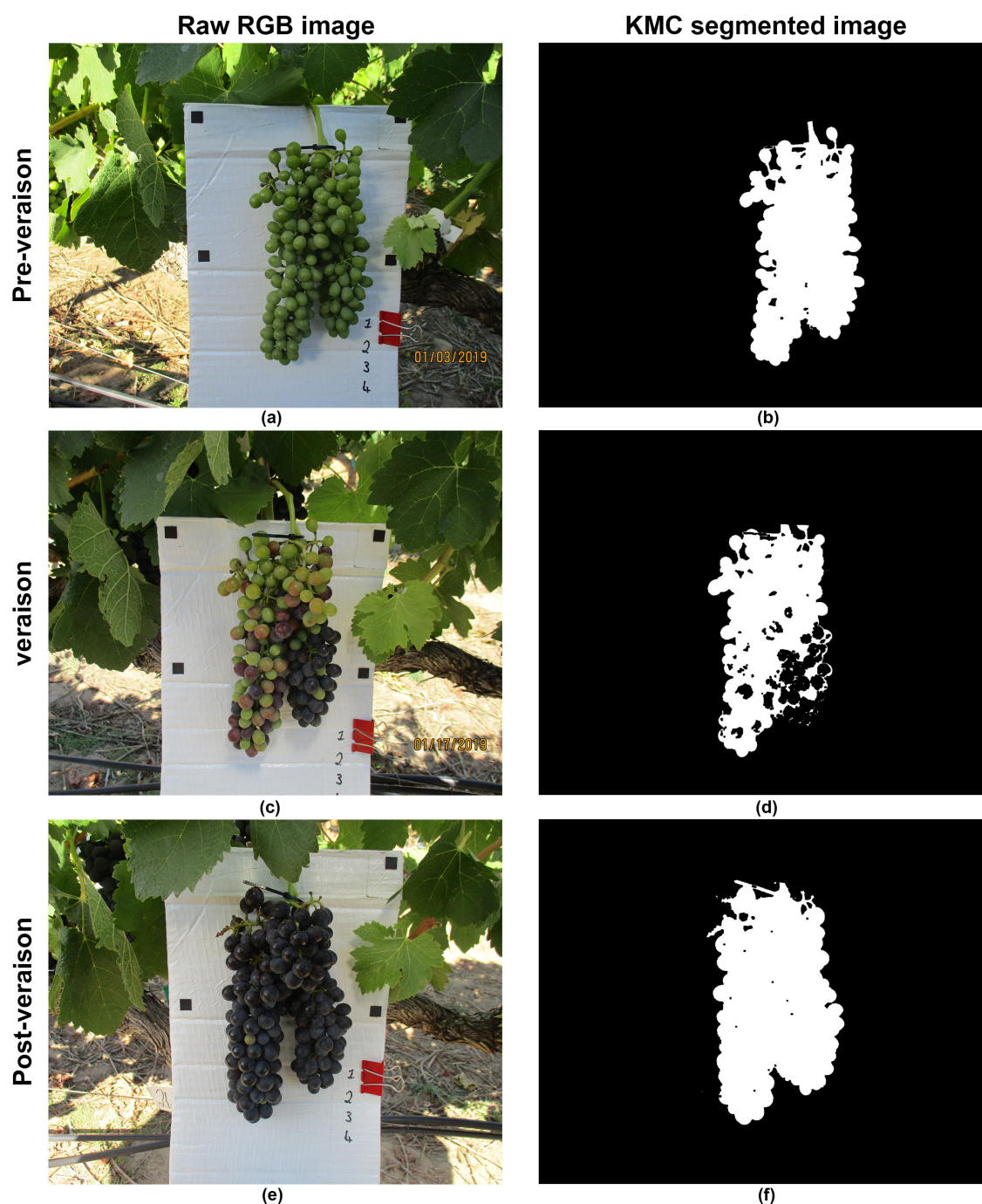


FIGURE 8. Difference in berry colour and segmentation results pre-veraison (a-b), during veraison (c-d), and post-veraison (e-f). Refer to Table 1 for phenological stages.

for the 50 bunches. Both metrics indicate the segmentation performance, with indistinguishable values following the same pattern during the data acquisition timeframe (i.e. 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 influenced by the reduced values during the middle of the acquisition period (27 December 2018 to 24 January 2019), which aligned with

veraison. During this period, the berries underwent a colour change from green to purple/black (Figure 8). 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 8d). The lowest segmentation results were for 17 January 2019 (phenological state: E-L 35 (Dry & Coombe, 2004), where both accuracy

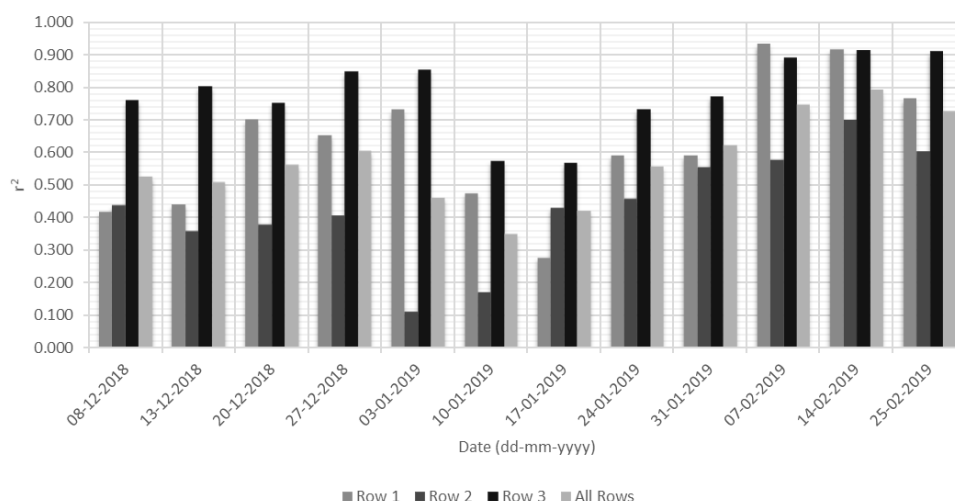


FIGURE 9. Bunch-level r^2 values evaluating yield estimation from multitemporal RGB data. The global dataset (all rows) consisted of 50 bunches, with row one (14 bunches), two (20 bunches) and three (16 bunches) presented as data subsets.

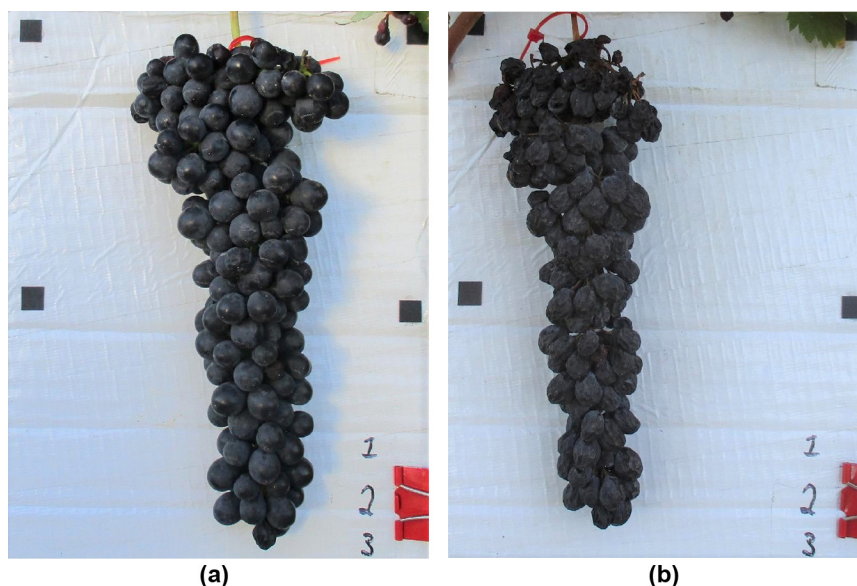


FIGURE 10. The same bunch imaged on 14 (a) and 25 (b) February 2019, illustrating the over-ripeness observed on 25 February.

and F1-score were 0.859. The best results were obtained for 7 February 2019 (accuracy = 0.992 and F1-score = 0.993) and 14 February 2019 (accuracy = 0.991 and F1-score = 0.992).

Outside the veraison timeframe, three other dates experienced a reduction in both accuracy and F1-score. On 20 December 2018 and 31 January 2019, a maximum accuracy and F1-score reduction of approximately 0.030 occurred. This had no visible effect on the yield estimation presented in Section 2.1 (Bunch-level). The third date, 25 February 2019, had an approximate accuracy and F1-score reduction of 0.050 from the previous date (14 February 2019). The reduction in segmentation

accuracy and F1-score for these three dates was likely due to the classes ($k = 3$) implemented for the KMC segmentation technique. These results clearly indicate the influence of veraison when using proximal sensing techniques for bunch detection and subsequent yield estimation.

2. Yield estimation: Multitemporal results

The three selected rows had a different phenological progression during the growing season, due to soil conditions and canopy practices implemented in previous years. The varying phenological stages became evident when evaluating the results for yield estimation, which were according to the field

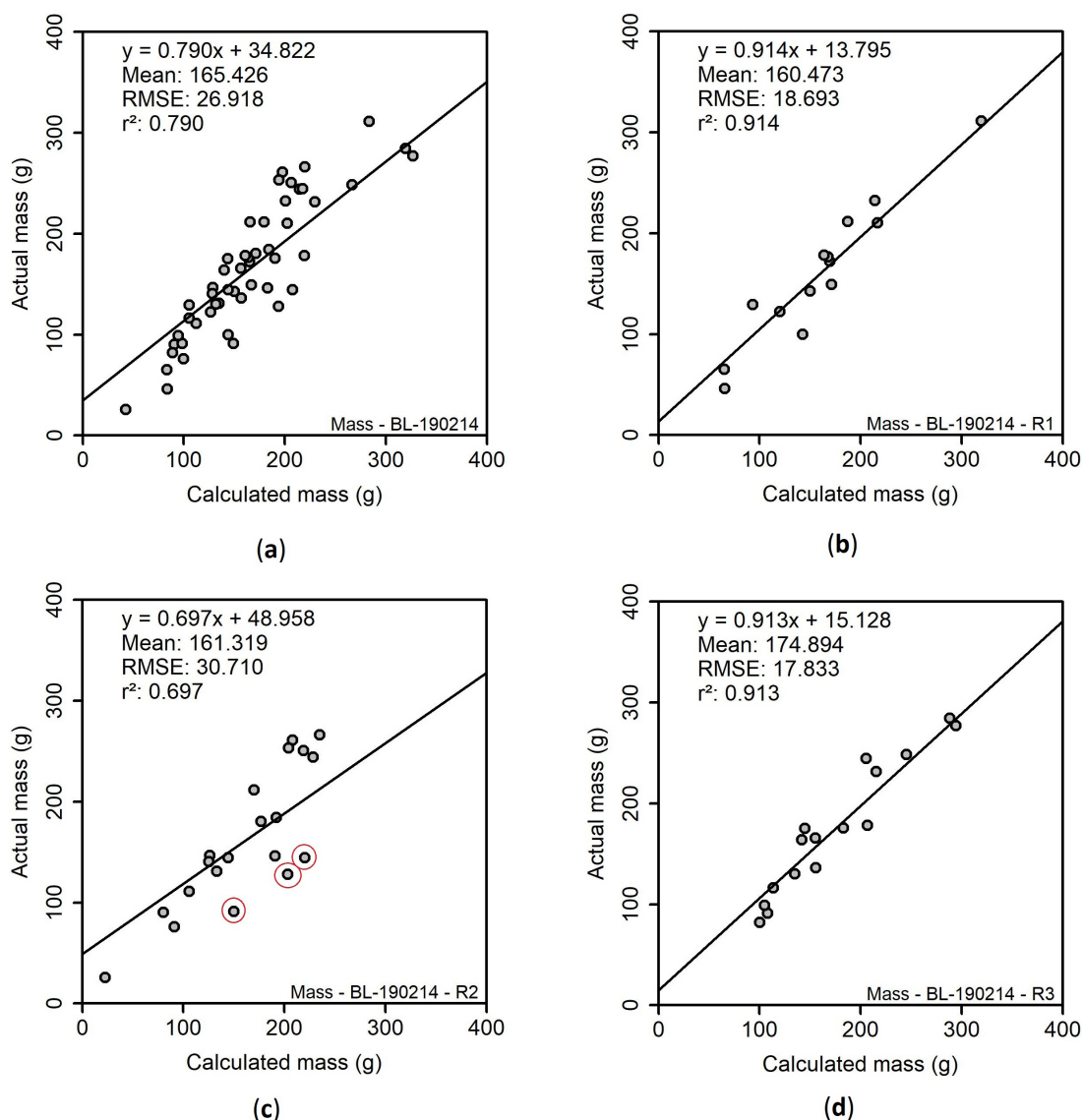


FIGURE 11. Relationship between the estimated and actual yield on 14 February 2019 for the global dataset (50 bunches) (a), row one (14 bunches) (b), row two (20 bunches) (c), and row three (16 bunches) (d).

observations. The decision was therefore made to present the results as a consolidated, global dataset, as well as individual rows (i.e. three data subsets). The global datasets represented 50 bunches (Section 2.1. Bunch-level) and 16 vines (Section 2.2 Plant-level).

2.1. Bunch-level

The multitemporal yield estimation results (estimated vs. actual mass yielding r^2 values) obtained at bunch-level are presented in Figure 9. The graph shows a general trend with r^2 values increasing at the start of the growing season, followed by a decrease in r^2 values toward veraison. Post-veraison, r^2 values increase, with best results approximately two weeks prior to harvest. Each individual row obtained the lowest r^2

value within a two-week window (3 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.

On 25 February 2019, two days prior to harvest, a reduced r^2 value was obtained. This decline r^2 value may be explained by the overripe bunches observed at harvest (Figure 10). The extracted area on 14 February (Figure 10a) was 112.453 cm², which reduced to 88.893 cm² on 25 February (Figure 10b) for the same bunch. This area reduction correlates with the reference volume measurements recorded on 14 February (244.350 veraison cm³) and 25 February (111.821 cm³).

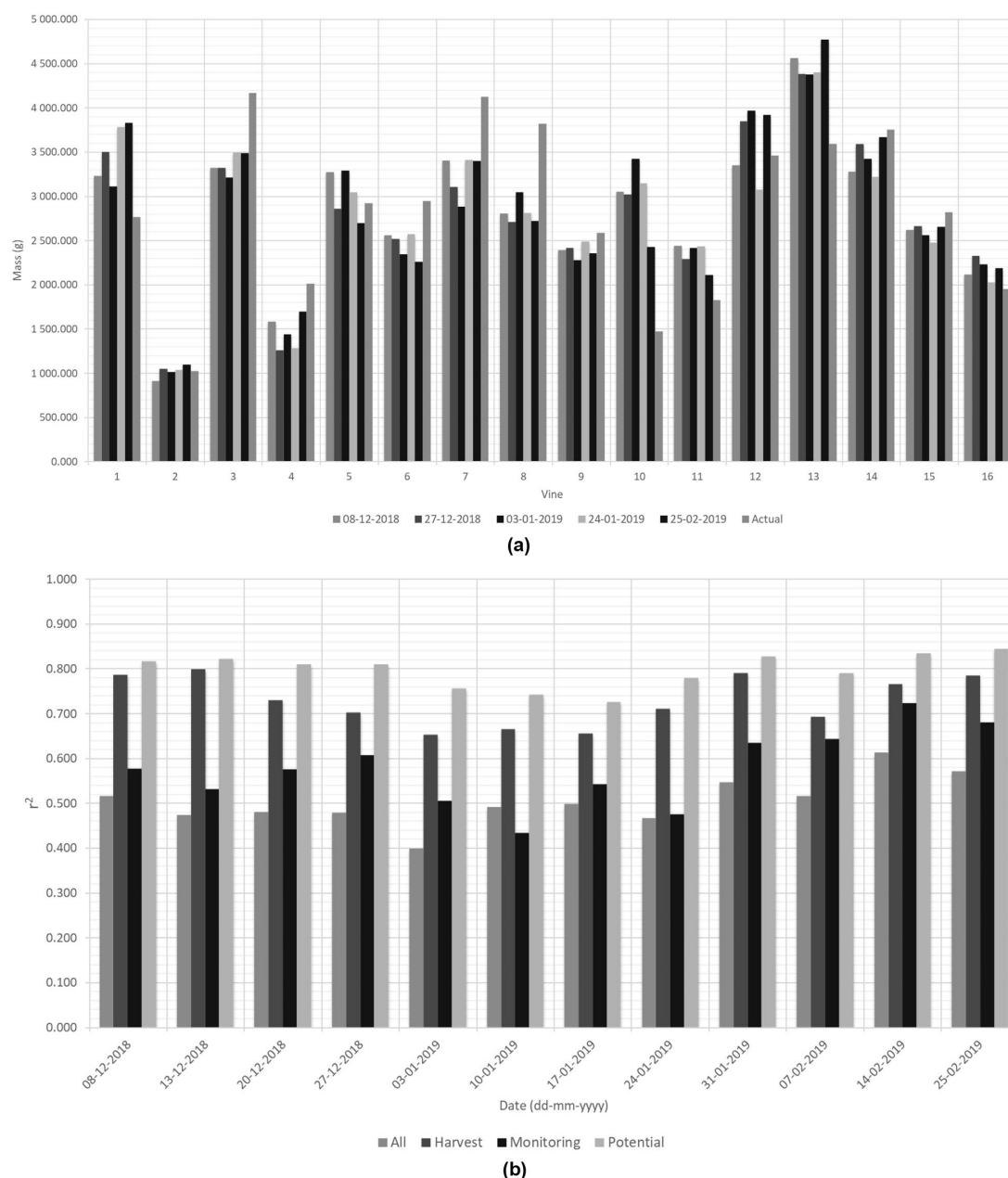


FIGURE 12. Vine-inferred mass for plant-level yield estimation across five dates (a). Vine-inferred r^2 values for the 12 dates, with four scenarios: i) 'All'; ii) 'Harvest'; iii) 'Monitoring'; and iv) 'Potential' (b). 'All' indicates the global dataset with 16 vines. 'Harvest' represents 13 vines where vines 1, 10, and 13 have been excluded from the analysis due to problems encountered at harvest. The 'Monitoring' scenario presents 13 vines with vines 3, 7, and 8 removed from the analysis owing to problems during the monitoring stage. 'Potential' presents the best-case scenario where six vines (3× 'Harvest' and 3× 'Monitoring') have been omitted, resulting in 10 vines. Refer to Table 1 for phenological stages.

The best yield estimation for the global dataset (50 veraison bunches) was observed on 14 February 2019, with an r^2 value of 0.790 and an RMSE of 26.918 g (see Figure 11a for more details). The highest r^2 results for both row two ($r^2 = 0.697$; Figure 11c) and row three ($r^2 = 0.913$; Figure 11d) were obtained on the same day, with RMSE values of 30.710 g and 17.833 g,

respectively. Although the highest r^2 value (0.932) for row one was achieved on 7 February 2019, the r^2 value (0.914) achieved on 14 February 2019 (Figure 11b) was still the highest of the three rows. The data points circled in Figure 11c represent over-estimated bunch yields, likely caused by overripe bunches.

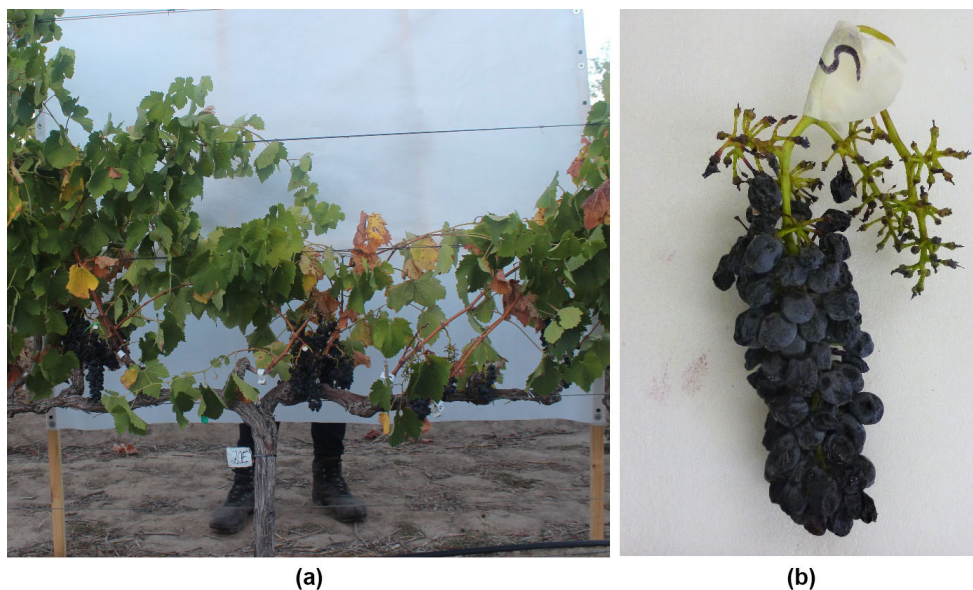


FIGURE 13. Example of canopy damage (a), and subsequent bunch damage (b) from the same vine.

2.2. Plant-level

Figure 12 presents the vine-inferred results obtained from the multitemporal data. It was evident from Figure 12a, which presents the estimated mass (g) per vine, that the presented technique performed better for certain plants. An in-depth analysis was done to investigate the problems behind the lower performing vines with two main issues detected. The first issue (Figure 12b; ‘Scenario – Harvest’), caused by over-ripeness and bird-damage, was apparent in vines 1, 10, and 13 (Figure 12a). A clear effect of this issue was illustrated by Figure 13 (vine 10), where damage to the canopy (Figure 13a) resulted in excessive bunch exposure, leading to overripe bunches with bird damage (Figure 13b). In this case, the actual yield measured at harvest was lower than the expected yield due to the physical damage of the bunches, thereby negatively influencing the results. The second issue (Figure 12b; ‘Scenario – Monitoring’) detected was related to the selection of the target bunches. When the target bunches were judged to be a poor representative sample of the vine, i.e. the target bunches were significantly smaller than the average size of the remaining bunch population, then the vine’s mass was underestimated as experienced in the vines 3, 7, and 8 (Figure 12a). These two problems yielded a decrease in the general performance of the vine-inferred methodology during the multitemporal analysis (Figure 12b).

Figure 12b presents the coefficient of determination (r^2) for the vine-inferred yield estimation across the 12 weeks. The global dataset (all 16 vines)

obtained a maximum r^2 value of 0.612 on 14 February 2019. To quantify the effect of the above-mentioned issues, the global dataset was analysed under three scenarios: i) ‘Harvest’; ii) ‘Monitoring’; and iii) ‘Potential’. By omitting three vines for the ‘Harvest’ and ‘Monitoring’ scenarios, the highest respective r^2 values were achieved on 13 December 2018 ($r^2 = 0.799$) and 14 February 2019 ($r^2 = 0.724$). The third scenario (‘Potential’) considers the ten ideal vines by omitting six vines (3× ‘Harvest’ and 3× ‘monitoring’) in total. The ‘Potential’ scenario achieved a maximum $r^2 = 0.844$ on 25 February 2019. A general trough (Figure 12b) was noticed across the multitemporal results which was influenced by the change in colour of the fruit during veraison. During this period, the model tends to underestimate the expected yield per vine. These results show that similar yield estimation values can be obtained during the early period, from fruit-set until pre-veraison, as are obtained post-veraison (late in the season).

DISCUSSION

Researchers have successfully utilised 2-D proximal sensing and related computer vision techniques for yield estimation before veraison (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). The present study successfully implemented unsupervised KMC for image segmentation, replacing the CT technique presented by Hacking *et al.* (2019). Additionally, this study set out to determine the optimal phenological stage for

yield estimation in a vineyard. To this end, a multitemporal RGB data was acquired over a 12-week period (8 December 2018 to 25 February 2019), concluding with the harvest on 27 February 2019.

1. Image segmentation

The segmentation results presented in Section 1. (Segmentation results) provided insight regarding the success of the two techniques evaluated in this study. An evaluation of the bunch-level segmentation results in Table 4 indicated that the KMC technique (accuracy = 0.942 and F1-score = 0.948) outperformed the CT technique (accuracy = 0.939 and F1-score = 0.943). Comparing these results with the in situ bunch-level segmentation results (accuracy = 0.781 and F1-score = 0.842) of Hacking *et al.* (2019), illustrates that 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 used during image acquisition, to improve image contrast.

Our results compare favourably with several studies conducted prior to harvest. For example, 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. The authors achieved a bunch classification accuracy of 96.56 % at bunch-level under various greenhouse and outdoor conditions.

While the KMC and CT techniques yielded similar results as represented in Table 4, it was the unsupervised clustering of the KMC technique which ultimately outperformed the manual CT technique which is dependent on user-selected thresholds. Generally, unsupervised classification tends to struggle in ‘busy images’ such as a vineyard with varying structural properties (Diago *et al.*, 2012), but the inclusion of the white background enabled this limitation to be overcome. This combination of the background and KMC technique is simple and practical for bunch-level data acquisition, with adequate performance for bunch segmentation as demonstrated by the results. Therefore, the KMC technique was employed for the subsequent multitemporal image

segmentation. The novelty of this research (i.e. multitemporal analysis) highlighted the effect of veraison on bunch detection when employing pixel-based segmentation techniques that rely on the image’s colour properties (Liu *et al.*, 2013). The segmentation results (accuracy = 0.977) achieved in this study prior to veraison aligned with the results of Aquino *et al.* (2018), whereas the segmentation results (accuracy = 0.942) obtained before harvest were an improvement on the results presented by Millan *et al.* (2018).

2. Yield estimation: Bunch-level

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

The lowest yield estimation r^2 (0.349 – Figure 9) occurred on 10 January 2019, coinciding with veraison. Maximum colour variation was therefore present in the bunches. Evidently, veraison negatively influenced the 2-D methodology used for yield estimation, as the colour properties are used for image segmentation. Although the lowest segmentation results (17 January 2019; Figure 7) were not obtained on the same day, they coincided with the same noticeable trough, attributed to veraison. The effect of veraison reduced the area detected per bunch, ultimately reducing the estimated yield during this phenological window. The presented methodology, which is reliant on pixel-level segmentation for bunch detection, was therefore deemed ineffective during this phenological window, spanning roughly four weeks from 27 December 2018 to 24 January 2019. Future research should investigate this period of phenological growth to determine its suitability for yield estimation.

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, resulting in a high r^2 value. These improvements included the use of a white background for image contrast during data acquisition, an improved calibration technique (using the black calibration squares), and the use of the KMC technique for segmentation, although KMC yielded comparable segmentation results (Section 1. Segmentation results). 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 varieties tested) r^2 of 0.840, slightly higher than the r^2 (0.790) obtained in this study. Similarly, Liu *et al.* (2013) conducted bunch-level yield estimation under laboratory conditions. The authors assessed various pixel-level metrics for yield estimation, achieving an r^2 value of 0.770 using a pixel count metric; the basis for the pixel area (cm^2) metric employed in this study. When comparing the results of this study to those of Liu *et al.* (2013) and Diago *et al.* (2015), it is important to note that this study was conducted in situ and employed an unsupervised KMC technique for image segmentation.

The global yield estimation results (Figure 9) show that the best r^2 (0.604) prior to veraison occurred at the onset of ripeness, on 27 December 2018. An alternative approach to yield estimation prior to veraison was conducted by Millan *et al.* (2016). The authors estimated the number of flowers per inflorescence, using this estimate as a proxy for yield estimation, and achieved an r^2 of 0.490. However, the r^2 value increased when the authors included historical data, such as fruit set rate ($r^2 = 0.790$) and average berry weight ($r^2 = 0.910$), surpassing the r^2 values obtained in this study. Evidently, historical data can improve yield estimation results. Aquino *et al.* (2018) investigated berry detection techniques for yield estimation during the bunch development phase, i.e. the final stage before the ripening process begins. The authors included 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 of this study per individual row; row one ($r^2 = 0.732$) and row three ($r^2 = 0.852$) performed on par, if not better.

The general performance for row two was lower than rows one and three. This lowered performance is not necessarily due to the presented methodology, but more likely due to the reference measurements captured at harvest. Overripe bunches, some with bird damage, negatively influenced the reference measurements, as noted in Section 2.2. (Plant-level). The variation in the results of this study may be attributed to the different phenological stages (evidenced in Figure 9) 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. Nevertheless, the 2-D methodology implemented in this study represents a strong alternative for estimating a vineyard's yield from data captured prior to veraison.

3. Yield estimation: Plant-level

Section 2.2. (Plant-level) presented the plant-level results with four different scenarios as seen in Figure 12b: 'All', 'Harvest', 'Monitoring', and 'Potential'. While the 'All' scenario represented the entire global dataset, the maximum achieved r^2 was only 0.612. Through the removal of problematic vines (under 'Harvest' and 'Monitoring' scenarios), the 'Potential' scenario was able to achieve a higher $r^2 = 0.844$. The improved results outperformed the various traditional yield estimation models presented by De la Fuente *et al.* (2015) which achieved a maximum $r^2 = 0.749$. Although both methodologies employ similar inference techniques, our data acquisition process was non-destructive, thereby resolving a fundamental limitation of traditional methodologies – i.e. destructive sampling.

Comparing these results to the work of Nuske *et al.* (2014) who achieved r^2 values between 0.600 – 0.730, one can see that the 'All' scenario yielded similar results, while the 'Potential' scenario yielded significantly higher results. While Nuske *et al.* (2014) obtained data both prior to veraison and prior to harvest, their overall methodology was fundamentally different. The authors utilised a berry count algorithm for yield estimation and captured data on-the-go at night using artificial illumination. Although 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. Interestingly, Zabawa *et al.*, (2020) have presented a modern methodology which utilises a grape harvester refitted with sensors for counting grapevine berries, overcoming certain lighting limitations of previous studies (Font *et al.*, 2015; Nuske *et al.*, 2011). However, the modifications applied to the harvester should be considered an expensive alternative which will not always be practical, especially in developing countries.

When evaluating the results on a multitemporal scale, the reduced r^2 values during veraison was evident, and aligned with the multitemporal results presented in Section 2.1. (Bunch-level). This provides clear evidence for the effect veraison has on proximal sensors combined with computer

vision for yield estimation. With that said, the results obtained for the ‘Potential’ scenario were of similar value both prior to veraison and prior to harvest. In practical terms, this is a very positive result demonstrating the potential for early yield estimation. Future research should investigate possible operational capabilities of this non-destructive sampling technique for inferring vine yield.

4. Operational potential and future research

The multitemporal nature of this research has provided insight to determining suitable phenological time frames for yield estimation – with promising results. The best results were obtained immediately prior to harvest (E-L stages 37-38), although similarly promising results (Figure 12b) were obtained between the completion of berry formation, i.e. E-L stages 33-34, and the onset of veraison, i.e. E-L stage 35. These results illustrate the potential for early yield estimation. From an operational perspective, being able to estimate the final yield six to eight weeks prior to harvest would be beneficial for managing the logistical aspects involved at harvest.

This research did not focus specifically on commercial implementation. However, there are several points worth noting in this regard. The yield estimation potential at plant-level via inference is of significant value as this methodology is simple to implement, requiring limited resources and expertise. The semi-automated image processing is straightforward, and the costs involved in data acquisition (Figure 3a) are comparatively low compared with more automated methods. Additionally, the simple use of an umbrella alleviates illumination variance that is associated with daytime data acquisition. While this follows inference techniques similar to traditional methods (De la Fuente *et al.*, 2015), a key aspect beneficial to the operational implementation was the conservative sampling and acquisition process which enabled the sampled bunches to remain on the vines for harvest, i.e. there was no destructive sampling.

Although the presented 2-D methodology for yield estimation demonstrated operational potential, several limitations were noted during the evaluation of the research. Regarding image processing, manual intervention is still required for ROI and cluster selection. The methodology incorporating unsupervised KMC segmentation is thus semi-supervised. Further research is required to develop a fully automated image processing

chain. Suitable alternatives worth investigating include deep learning (Badrinarayanan *et al.*, 2017; Razavian *et al.*, 2014) and object-based image analysis (Blaschke *et al.*, 2014). Both these techniques have strengths and limitations that are potentially capable of outperforming traditional image segmentation techniques in complex environments like vineyards. Additionally, logistical limitations restricted the sample sizes due to time constraints. These reduced sample sizes are statistically small for linear regression and the reader needs to be cognitive of this fact. Future research should have a more statistically sound sample size to substantiate the findings of this research.

Intra-vine bunch sampling yielded various bunch sizes as highlighted by the ‘Monitoring’ scenario in Section 2.2. (Plant-level) Ensuring phenological uniformity within bunches at the time of sampling is recommended for future research, as well as the definition of a protocol to select representative bunches. Damage sustained to bunches prior to harvest reduced the final reference mass recorded, likely influencing the results. Bunch damage is inherent in commercial farming, therefore this aspect must be considered as a factor in the estimation methods.

Refining these limitations may lead to the practical use of the presented techniques, especially the vine-inferred technique for yield estimation. Future research should investigate the scalability of the presented methodology by incorporating a more automated, on-the-go approach to data acquisition (Aquino *et al.*, 2018; Millan *et al.*, 2018). It is our recommendation to further commercial yield estimation research during the final stages of bunch formation, prior to veraison (E-L stage 35).

CONCLUSIONS

This study set out to determine the optimal phenological stage for yield estimation using 2-D proximal sensing and related computer vision techniques at both bunch- and plant-level. Two segmentation techniques were compared at bunch-level prior to harvest, namely KMC and CT. Based on the results of this study, the following conclusions are drawn:

1. Both segmentation techniques performed well. However, the unsupervised KMC technique removed the human limitation of having to select appropriate colour thresholds, as required by CT.

2. The best yield estimates were obtained during the final stages of berry ripening (during sugar development), in the final weeks prior to harvest (phenological E-L stages 36-38). An alternative window is at the end of berry formation (E-L stage 33/34), immediately prior to the first signs of veraison (E-L stage 35).
3. For the duration of veraison, the bunches' colour development significantly reduced yield estimates. Future colour-based research should avoid this phenological window. Yield estimation should be conducted prior to veraison or prior to harvest, depending on the intended use of the yield estimation data.

Acknowledgements: This research was funded by Winetech, project DVO 07: "Near-real-time characterization of vines for more efficient vineyard management". The authors would like to thank Kobus Bredell, Jayde Bromwich, Aloïs Houeto, Aladino Jasse, Albert Strever and Tatiana Pizzol for their technical insight and research support during the season.

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. <https://doi.org/10.1016/j.compag.2016.01.018>
- Aquino, A., Millan, B., Diago, M. P., & Tardaguila, J. (2018). Automated early yield prediction in vineyards from on-the-go image acquisition. *Computers and Electronics in Agriculture*, 144, 26–36. <https://doi.org/10.1016/j.compag.2017.11.026>
- 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*, 1027–1035. <http://www.stanford.edu/~dathur/kMeansPlusPlus.pdf>
- Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12), 2481–2495. <https://doi.org/10.1109/TPAMI.2016.2644615>
- Blaschke, T., Hay, G. J., 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. <https://doi.org/10.1016/j.isprsjprs.2013.09.014>
- Conradie, W., Carey, V., Bonnardot, V., Saayman, D., & van Schoor, L. (2002). Effect of different 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. <http://www.sawislibrary.co.za/dbtextimages/ConradieWJ.pdf>
- Correa, C., Valero, C., Barreiro, P., Diago, M., & Tardaguila, J. (2012). Feature extraction on vineyard by Gustafson Kessel FCM and K-means. 16th IEEE Mediterranean *Electrotechnical Conference*. <https://doi.org/10.1109/MELCON.2012.6196477>
- 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 Science de La Vigne et Du Vin*, 49, 27–35.
- Diago, Maria P, Tardaguila, J., Aleixos, N., Millan, B., Prats-Montalban, J. M., Cubero, S., & Blasco, J. (2015). Assessment of cluster yield components by image analysis. *Journal of the Science of Food and Agriculture*, 95, 1274–1282. <https://doi.org/10.1002/jsfa.6819>
- Diago, Maria Paz, 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. <https://doi.org/10.3390/s121216988>
- Dry, P., & Coombe, B., (2004). *Viticulture 1 – Resources*, 2nd Ed. "Revised version of grapevine growth stages – The modified E-L system". Winetitles, Adelaide, Australia.
- Dunn, G. M., & Martin, S. R. (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. <https://doi.org/10.1111/j.1755-0238.2004.tb00022.x>
- Ferreira, J. H. S., & Marais, P. G. (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. <https://doi.org/10.21548/8-2-2313>
- 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. <https://doi.org/10.3390/s150408284>
- Gonzalez, R., Woods, R., & Eddins, S. (2009). Digital image processing using MATLAB (Vol. 2011, Issue 2). *Gatesmark Publishing*. <https://doi.org/10.1109/IEMBS.2011.6091204>

- Grossetête, M., Berthoumieu, Y., Da Costa, J.-P., Germain, C., Laviolle, O., & Grenier, G. (2012). Early estimation of vineyard yield: Site specific counting of berries by using a Smartphone. *International Conference on Agriculture Engineering*. <https://hal.archives-ouvertes.fr/hal-00669502>
- 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. <https://doi.org/10.3390/s19173652>
- Komm, B., & Moyer, M. (2015). Vineyard Yield Estimation. In WSU Extension (pp. 1–11).
- Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, 28(5), 1–26. <https://doi.org/10.18637/jss.v028.i05>
- 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. <https://doi.org/10.1016/j.compag.2017.03.013>
- Liu, S, Marden, S., & Whitty, M. (2013). Towards automated yield estimation in viticulture. Australasian Conference on Robotics and Automation.
- Liu, S, & Whitty, M. (2015). Automatic grape bunch detection in vineyards with an SVM classifier. *Journal of Applied Logic*, 13(4), 643–653. <https://doi.org/10.1016/j.jal.2015.06.001>
- Liu, Scarlett, Li, X., Wu, H., Xin, B., Petrie, P. R., & Whitty, M. (2018). A robust automated flower estimation system for grape vines. *Biosystems Engineering*, 172, 110–123. <https://doi.org/10.1016/j.biosystemseng.2018.05.009>
- 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. <https://doi.org/10.3390/s16122098>
- 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. <https://doi.org/10.1016/j.patcog.2019.02.023>
- Marinello, F., Pezzuolo, A., Cillis, D., & Sartori, L. (2016). Kinect 3D reconstruction for quantification of grape bunches volume and mass. *Engineering for Rural Development*, 876–881.
- Millan, B., Aquino, A., Diago, M. P., & 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. <https://doi.org/10.1002/jsfa.7797>
- 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. <https://doi.org/10.1155/2018/9634752>
- Mochida, K., Koda, S., Inoue, K., Hirayama, T., Tanaka, S., Nishii, R., & Melgani, F. (2018). Computer vision-based phenotyping for improvement of plant productivity: a machine learning perspective. *Journal of GigaScience*, 8(1), 1–12. doi: 10.1093/gigascience/giy153
- 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. <https://doi.org/10.1002/rob.21541>
- R Core Team (2019). R : A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. <https://www.r-project.org/>
- Padua, L., Adao, T., Sousa, A., Peres, E. & Sousa, J. (2020). Individual grapevine analysis in a multi-temporal context using UAV-based multi-sensor imagery. *Journal of Remote Sensing* 12(1), 139. <https://doi.org/10.3390/rs12010139>
- Razavian, A. S., Azizpour, H., Sullivan, J., & Carlsson, S. (2014). CNN features off-the-shelf: An astounding baseline for recognition. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 512–519. <https://doi.org/10.1109/CVPRW.2014.131>
- Reis, M. J. C. S., Morais, R., Peres, E., Pereira, C., Contente, O., Soares, S., Valente, A., Baptista, J., Ferreira, P. J. S. G., & Bulas Cruz, J. (2012). Automatic detection of bunches of grapes in natural environment from color images. *Journal of Applied Logic*, 10(4), 285–290. <https://doi.org/10.1016/j.jal.2012.07.004>
- The MathWorks Inc. (2018). Matlab R2018b v9.5.0.944444. (9.5.0.944444). *The MathWorks Inc.* <https://www.mathworks.cn/en/>
- The MathWorks Inc. (2019). *imsegkmeans*: K-means clustering based image segmentation. *Matlab Documentation*. https://www.mathworks.com/help/images/ref/imsegkmeans.html?searchHighlight=imsegkmeans&s_tid=doc_srchttitle
- Tian, H., Wang, T., Liu, Y., Qiao, X., & Li, Y. (2020). Computer vision technology in agricultural automation – A review. *Journal of Information Processing in Agriculture*, 7(1), 1–19.
- Wolpert, J. A., & Vilas, E. P. (1992). Estimating vineyard yields: Introduction to a simple, two-step method. *American Journal of Enology and Viticulture*, 43(4), 384–388.
- Zabawa, L., Kicherer, A., Klingbeil, L., Topfer, R., Kuhlmann, H., & Roscher, R. (2020). Counting of grapevine berries in images view semantic segmentation using convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 164, 73–83.