Title: Automated early yield prediction in vineyards from on-the-go image acquisition

Authors: Arturo Aquino, Borja Millan, Maria-Paz Diago, Javier Tardaguila

Instituto de Ciencias de la Vid y del Vino (University of La Rioja, CSIC, La Rioja Regional Government), 26006, Logroño, La Rioja, Spain

Corresponding author: Javier Tardaguila (javier.tardaguila@unirioja.es).

Abstract

Early grapevine yield assessment provides information to viticulturists to help taking management decisions to achieve the desired grape quality and yield amount. In previous works, image analysis has been explored to this effect, but with systems performing either manually, on a single variety or close to harvest-time, when there are few rectifiable agronomic aspects. This study presents a solution based on image analysis for the non-invasive and in-field yield prediction in vines of several varieties, at phenological stages previous to veraison, around 100 days from harvest. To this end, an all-terrain vehicle (ATV) was modified with equipment to autonomously capture images of 30 vine segments of five different varieties at night-time. The images were analysed with a new image analysis algorithm based on mathematical morphology and pixel classification, which yielded overall average Recall and Precision values of 0.8764 and 0.9582, respectively. Finally, a model was calibrated to produce yield predictions from the number of detected berries in images with a Root-Mean-Square-Error per vine of 0.16 kg. This accuracy makes the proposed methodology ideal for early yield prediction as a very helpful tool for the grape and wine industry.

Keywords: Grapevine early yield prediction; image analysis; non-invasive sensing; precision viticulture.

1 Introduction

Among all collectable data from a vineyard, grapevine yield estimation outstands for its economical relevance (Wolpert and Vilas, 1992; Martin *et al.*, 2002; Dunn, 2010), and also for being key to help optimizing plant growth and to improve fruit quality (Dunn and Martin, 2003). Yield variability within a vineyard has been proved to be high (Bramley and Hamilton, 2004). Classical yield estimation methods, which consist on manual collection and weighting of the crop yield in a given and limited number of plants previous to harvest is tedious and insufficient to obtain representative yield data. Consequently, non-invasive imaging-based methods are being investigated to make possible the efficient and continuous capture of detailed information from vines throughout their life cycle (Spalding and Miller, 2013; Li *et al.*, 2014).

Grapevine yield is determined by the yield components, defined as the number of clusters, the number of berries per cluster and the berry weight (Tardáguila *et al.*, 2012). Alternatively, yield can also be estimated from the total number of berries and the berry weight (Nuske *et al.*, 2014). Whatever the case, the number of berries is the most labile variable determining yield (Anderson *et al.*, 2008). It is highly influenced by the weather conditions during inflorescence development and berry-set, when it gets fully established and remains mostly invariable until harvest (May, 2004).

Imaging-based developments aimed at yield estimation can be found in the literature under two differentiated approaches. A first set of proposals is framed within the manual acquisition of images, thereby focusing on the discrete analysis of samples from the vineyard. Contrary, the second set studies yield estimation by means of on-the-go image acquisition using modified human-driven, or autonomous vehicles, with the ambition of making possible the continuous and massive analysis. Within the first approach, Chamelat *et al.* (2006) presented a method based on pixel classification to detect grape-pixel aggregations in vine RGB images manually taken in the field under daylight. Reis *et al.* (2011) also proposed the detection of grape-pixel aggregations in outdoor RGB images that were taken using the camera flash. They employed colour analysis and non-linear image processing to detect aggregations of white-grape pixels. The method was later

improved and extended to work with red grapes also (Reis *et al.*, 2012). On the other hand, Rahman and Hellicar (2014) studied the detection and classification of undeveloped and mature clusters of white grapes. Their development consisted on the extraction of the clusters from the background as a first step, to finally perform pixel classification by means of texture analysis. Other approaches involved colour histogram classification, RGB thresholding and fuzzy clustering to estimate cluster weight in vine images (Liu *et al.*, 2013). A more complex capturing device was developed by Fernández *et al.* (2013), which consisted on a sensor rig composed of a CCD camera and a servo-controlled filter wheel. This device was used to collect multispectral imagery, which was analysed employing a sequential masking algorithm based on k-means clustering. Aquino *et al.* (2017) proposed a methodology based on mathematical morphology and false positives filtering by means of supervised learning to count the number of berries in RGB cluster images taken in the vineyard with a smartphone under natural sunlight. Very recently, Mack *et al.* (2017) described a method based on 3D reconstruction of high-resolution laser range data to model sampled grape clusters.

Regarding the methods for on-the-go vineyard monitoring, the literature is rather limited. Some works involved image acquisition after veraison (Font *et al.* 2015; Rose *et al.* 2016), very close to harvest, when the time window for cluster thinning, to either regulate the yield or to promote increased allocation of carbohydrates to specific clusters (to improve fruit ripening) is very limited. Font *et al.* (2015) equipped a ground vehicle with a reflex camera and artificial lighting to capture vine images of a single table grape variety (Flame seedless) at night-time right before harvest and reported yield prediction errors of 16%. More recently, a multi-view-stereo system coupled to a LED external illumination system was mounted in an autonomous moving platform (0.3 km/h), to acquire night-time images, and the estimation of the number of berries in the cluster images was attempted (Rose *et al.*, 2016). The choice of earlier yield estimation, before veraison, was taken by several authors. Likewise, Nuske *et al.* (2014) used a tractor, which moved around 5 km/h, to mount RGB cameras and artificial illumination for image capture in the vineyard at night time. They collected data from six grapevine varieties in several growing seasons and vineyards, at various timings before harvest, and developed a supervised classifier for berry

detection fed with a wide set of descriptors (above 30) of texture, colour and shape. Average yield estimation accuracy of 11% was only provided for Flame seedless, using images acquired at harvest time (Nuske *et al.* 2014). A different approach was followed by Liu *et al.* (2017), who developed a computer vision system for early grape yield estimation based on shoot detection using videos acquired with a low-cost camera at daytime using a white background which moved alongside the row.

In view of the missing factors and lessons learned from previous works, the goal of the present study was the development of a comprehensive technological solution for automated early grapevine yield prediction based on images acquired on-the-go at a speed similar or faster than most agricultural vehicles. In contrast to other available tools focused mostly on yield estimation of a single variety at pre-harvest, this proposal analyses vine images of tens of clusters of five different grapevine varieties at earlier stages; concretely at phenological stages between berry-set and cluster closure (around 100 and 120 days before harvest) This earlier stage at which yield information is acquired could significantly improve its impact, since grape-growers could perform viticultural practices to rectify certain key parameters more effectively. Additionally, a secondary goal of the present work was to simplify the classifier for berry detection and counting and to evaluate the potential development of a unique solution involving multiple grapevine varieties.

2 Materials and methods

2.1 Experimental design

The trials were carried out during season 2015 in a commercial vineyard located in Falces (lat. 42º27'46.0''N; long. 1º48'12.9''W; Navarra, Spain). Five grapevine (*Vitis vinifera* L.) varieties (red and white) were considered for this study: Albariño, Cabernet Sauvignon, Syrah, Tempranillo and Viognier. The vines were planted in year 2009 on rootstock Richter 110 in N-S orientation, trained to vertical shoot positioning (VSP) trellis system with 2 and 1 m inter-row and intra-row distances, respectively. Vines were defoliated from node one to six only on their east side after berry set. Depending on the variety, at the time of image acquisition in the field (23rd June, 2015), the clusters were at phenological stages between K and L, according to the

scale proposed by Baggiolini (1952). Following this scale, phenological stage K refers to that at which berries have the 50% of their final size; it is also denoted as pea-size stage. With regard to phenological stage L, also called cluster closure, it is reached when berries have about 70% of their final size, and they start to touch each other within the cluster.

2.2 Acquisition of reference data

For every cultivar, ten consecutive sampling segments composed of three adjacent vines each, were labelled and delimited ($5\ varieties \times 10\ segments \times 3\ vines$); thirteen additional segments of non-defoliated vines were also selected. Each three-vine segment constituted a unique sampling point in which the produced yield was individually weighted and registered at the end of the season using a hanging scale (Kern CH15K20, Kern & Sohn GmbH, Balingen, Germany).

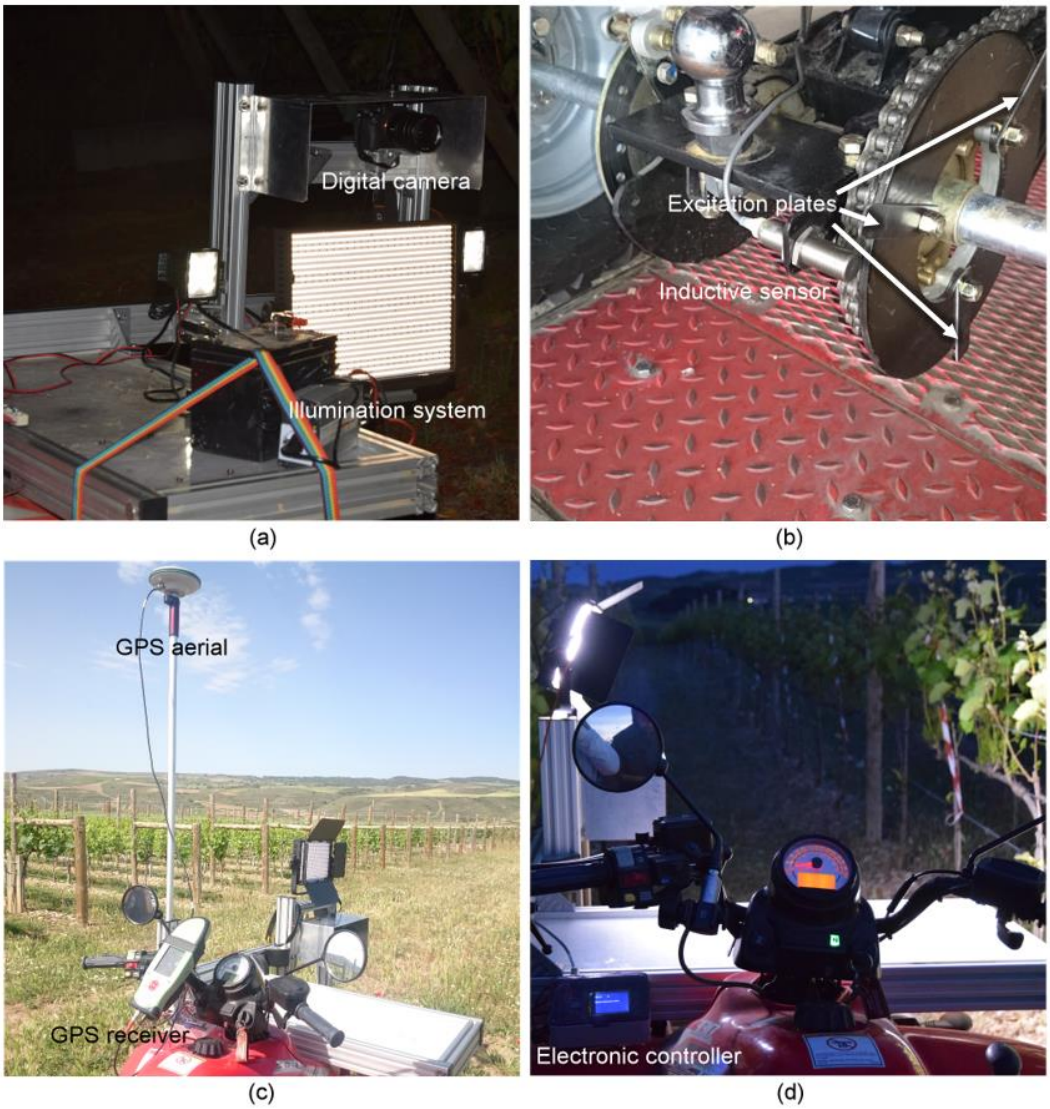
2.3 Image acquisition

The vine segments were photographed 'on-the-go' at night time without user intervention. To this effect, a sensor-equipped all-terrain vehicle (ATV) was driven through the vineyard at 7 km/h. Image capture automation was achieved by adapting the vehicle to incorporate the following elements:

- A mirror-less RGB camera (Sony α7II, Sony, Tokyo, Japan) equipped with a Zeiss 24/70mm lens with optical stabilisation. This camera mounts a 24 Mpx CCD sensor, incorporates a 5-axis image stabilisation system and provides high shutter speed, quick image storage and low noise generation (Figure 1-(a)). These features allowed to capture and store on-the-go three images per wheel-spin at 7 km/h, producing high quality images despite the vibrations caused by the ATV's engine and the irregular ground's surface. For the experiments, the camera was set in manual mode, configuring the aperture in f/4, shutter speed in 1/2500 seconds, ISO sensitivity in 5000 and focus in manual mode.
- A white-light LED panel (Figure 1-(a)) to provide controlled artificial illumination for vineyard monitoring at night time. By means of illumination and camera parametrization, it

was possible to isolate in the image the vines under evaluation from those in the adjacent row. Should image acquisition be conducted at day time, the sunlight would equally illuminate the whole scene, producing images in which the vines under study were hardly distinguishable from those at their back (this fact is illustrated in Figure 2).

- A modular and flexible structure built with commercial aluminium profiles for sensor attachment. The structure consisted on an upfront and a rear tray, plus an adjustable arm to be installed in the front tray (Figure 1-(a)). The arm was designed to mount the camera and illumination system, making the combination adjustable to different vineyard heights and widths. The arm was adjusted for the camera to be at around 1.5 meters from the canopy.
- An inductive sensor installed in the rear axle for camera triggering; the sensor produced three activation pulses per wheel-spin (Figure 1-(b)).
- A GPS receiver (Leica Zeno 10 Global Positioning System; Heerbrug, St. Gallen, Switzerland) for image georeferencing (Figure 1-(c)).
- A custom-built electronic control system for managing the signals generated by the installed devices (GPS and inductive sensor) and triggering the camera using an isolated signal (Figure 1-(d)). The system also allowed for data storage in an SD-card and for showing capture-status information in a 4.9” TFT screen.



154

155 **Figure 1.-** Modified all-terrain vehicle (ATV) for vineyard monitoring: (a) view of the front
156 showing the installed attachment structures, the illumination system and the capturing device;
157 (b) inductive sensor which generates triggering signals for the camera as the excitation plates
158 are detected (thee signals per wheel turn); (c) detail of the GPS installed in the vehicle; (d)
159 electronic controller for handling the triggering signals and GPS information.

Figure 2.- Images of the same vine taken on-the-go at day time (a) and with artificial illumination at night time (b). As it can be observed, the vegetation at the back plane is hardly distinguishable from that at the front plane in image (a). This difficulty is solved with artificial illumination and camera parametrization (b).

2.4 Image analysis algorithm for berry detection in grapevine images

The methodology developed for berry segmentation in digital images acquired on-the-go can be divided into three main steps. The first phase consisted on pre-processing the images to discard redundant data; this is, those canopy segments captured in more than one image. Then, in a second phase, a set of berry candidates was detected by means of morphological filtering. Finally, those candidates not corresponding to berries, i.e. false positives, were discarded using a supervised-learning approach. The berry-segmentation algorithm was developed and tested using Matlab (R2015a, Mathworks, Natick, MA, USA). Figure 3 includes a flow-chart diagram illustrating the image analysis algorithm described below.

2.4.1 Image pre-processing

Three images per wheel-spin were acquired on-the-go with the sensing vehicle, which ensured a full coverage of the canopy under evaluation. However, this approach also produced duplication of information in images, as two consecutive images always contained a repeated vine segment (Figure 4-(a)-(c)). To avoid the analysis of redundant data, images were matched using the 'Auto Blend' tool provided by Adobe Photoshop CC 2015 (Adobe Systems Incorporated; San José,

California, USA) to this effect. By this way, matched images were created for being piecewise analysed afterwards (Figure 4-(d)).

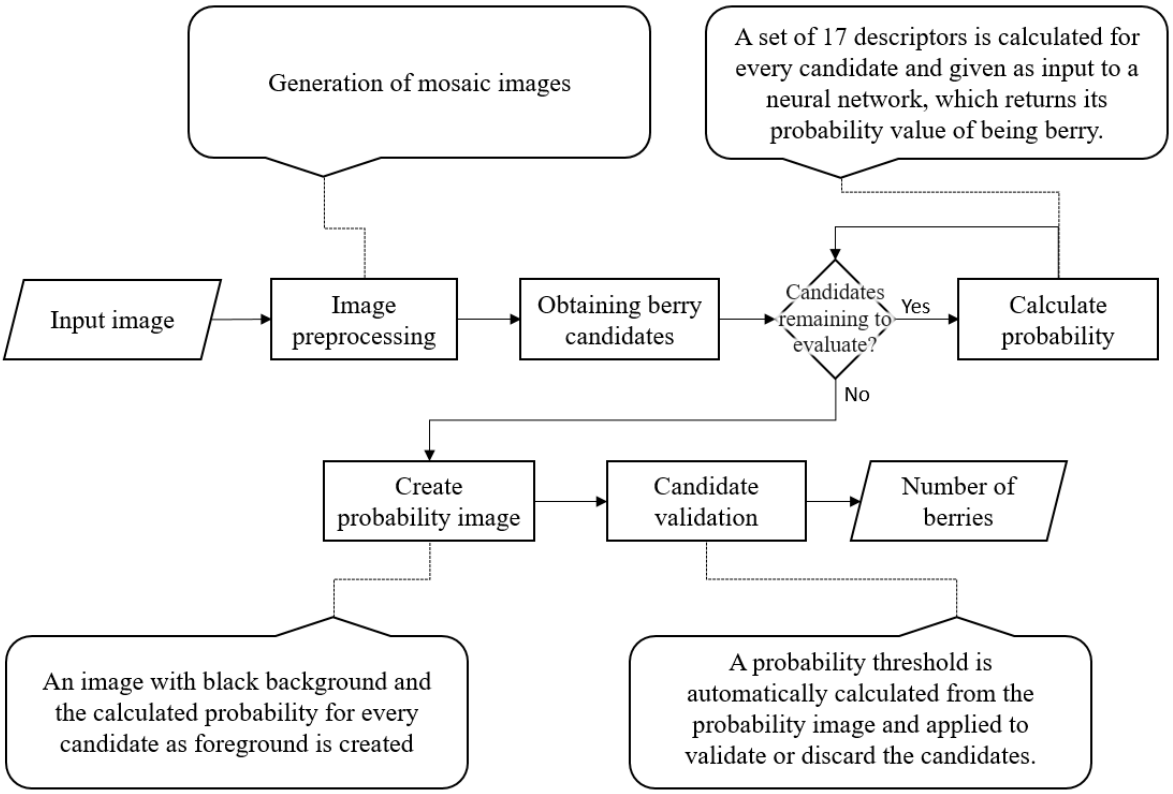


Figure 3.- Flow-chart diagram of image analysis algorithm for detecting berries.

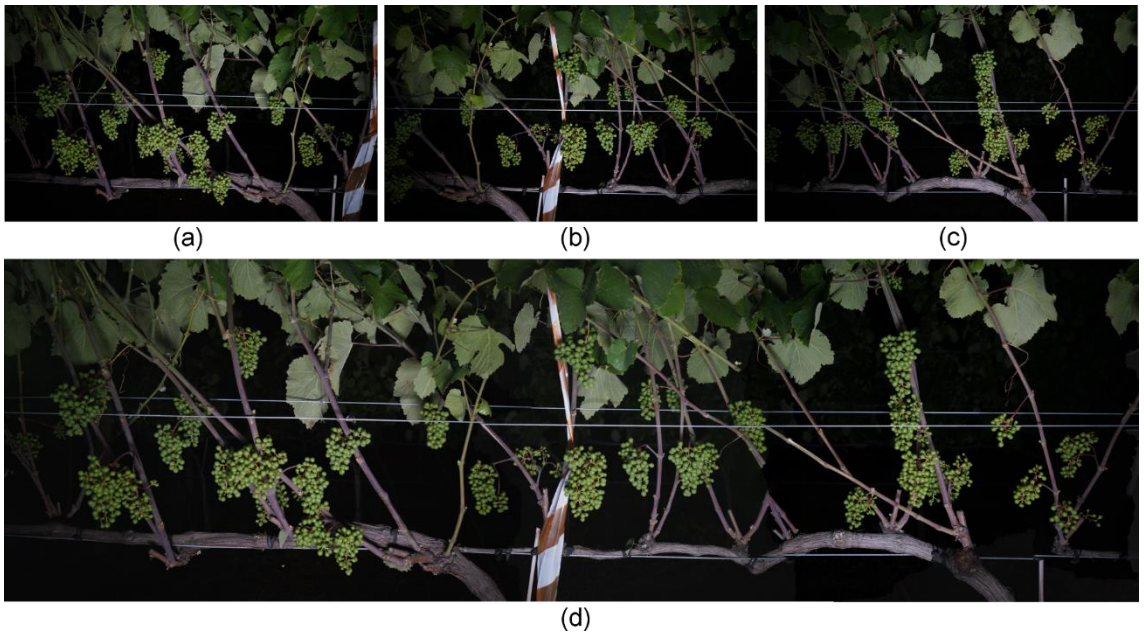


Figure 4.- Result of image matching: (a), (b) and (c) are images taken on-the-go with the

mobile sensing platform; (d) result of matching the overlapping areas in images (a), (b) and (c).

2.4.2 Extraction of berry candidates

All images were taken in the camera-native RGB colour scheme, but they were converted to the CIE 1976 *L*a*b** colour space (CIELAB) at this point (Connolly and Fleiss, 1997). This scheme favoured the following processing approach, which explodes features related to colour and luminosity independently. Contrary to the RGB scheme, the CIELAB space decouples the illumination and colour information. Indeed, it provides a luminosity layer '*L**', and the chromaticity layers '*a**' and '*b**' describing colour information along the red-green and blue-yellow axis, respectively. Thus, the colour information is represented in the '*a*b**' sub-space, keeping the luminosity information isolated in the '*L**' plane.

As a first step, a 5×5 Gaussian filter was applied to the grey-level image of the *L** channel, I_L, for salt-and-pepper noise filtering. Then, a set of berry candidates was generated by finding the connected components in the image corresponding to the circular light reflection pattern taking place on the curved surface of berries; this effect was enunciated as the Lambert's cosine law (Smith, 2007). These bright regions were computed using the extended *h*-maxima transform, which finds in the image those regional maxima with a height less than or equal to *h*. The transform firstly calculates the morphological reconstruction of I_L from marker $I_L - h$ (consult details on morphological reconstruction at Soille (2004))

$$I_{filt} = R_{I_L}(I_L - h), h = 3 \quad (1)$$

to finally extract the surviving regional maxima in image I_{filt} fulfilling the *h* constraint by computation of

$$I_{max} = I_{filt} - R_{I_{filt}}(I_{filt} - 1) \quad (2)$$

and thresholding this resulting image

$$I_{maxBin}(x, y) = \begin{cases} 255 & if\ I_{max}(x, y) > 0 \\ 0 & otherwise \end{cases} \quad (3)$$

The effect of the parameter h in the algorithm's results is studied in the results section. Finally, the set of berry candidates, $Berry_{cand}$, was obtained by calculating the connected components, CC_i, remaining in the binary image I_{maxBin}:

$$Berry_{cand} = \{CC_i \subseteq I_{maxBin}\} \quad (4)$$

2.4.3 Validation of berry candidates

This post-processing step is aimed at discarding false positives by analysis of the set of connected components previously obtained. To this effect, a set of 17 descriptors was computed for every candidate, and provided as input to a feedforward multilayer perceptron neural network (NN) trained with supervised learning. This classifier produced a probability for the candidate to truly represent a berry. Finally, a probability threshold to discard wrong candidates was automatically established.

Let *L*, *A* and *B* be the images from the *L**, *a** and *b** channels of an image in the CIE 1976 *L*a*b** colour space. With these definitions, the set of berry descriptors is formulated as:

- Circularity (one descriptor): it is applied as defined in Aquino *et al.* (2017), and evaluates the circularity of a connected component representing a berry candidate by calculation of the quotient between its minor and major axis length:

$$f_1 = \frac{min\ (R_{E_i}^1, R_{E_i}^2)}{max\ (R_{E_i}^1, R_{E_i}^2)} \quad (5)$$

  where $R_{E_i}^1$ and $R_{E_i}^2$ are the two radii of the ellipse having the same normalized second central moments as the connected component.

- Normality (one descriptor): this descriptor strengthens the detection of berries under the assumption that the light reflection produced on their surface describes a 2-dimensional Gaussian distribution in an image. It is computed on image *L* and formulated as:

$$f_2 = \frac{d_{S_i^{Gauss}}^{0º} + d_{S_i^{Gauss}}^{45º} + d_{S_i^{Gauss}}^{90º} + d_{S_i^{Gauss}}^{135º}}{\left(\hat{S}_i^{0º} + d_{S_i}^{0º}\right) + \left(\hat{S}_i^{45º} + d_{S_i}^{45º}\right) + \left(\hat{S}_i^{90º} + d_{S_i}^{90º}\right) + \left(\hat{S}_i^{135º} + d_{S_i}^{135º}\right)} \quad (6)$$

  A full description and study on this descriptor is developed in Aquino *et al.* (2017).

- Size (one descriptor): assuming that all berries within a cluster have a similar size, and that the distance between a given berry and its closest neighbour is also quite stable, this descriptor penalizes those candidates having an incongruent ratio size/distance.

First, a set of minimum distances among candidates, D_{min}, is calculated by computing, for all of them, the minimum Euclidean distance between themselves and their closest neighbour:

$$D_{min} = \left\{ d_j^{min} \middle| \, j = 1, \dots, n \right\}$$

where

$$d_j^{min} = min \left\{ d_{j,k} \middle| \sqrt{\left(x_j^c - x_k^c\right)^2 + \left(y_j^c - y_k^c\right)^2}, k = 1, \dots, n \right\} \quad (7)$$

denoting *n* the number of connected components in $Berry_{cand}$, and x_l^c and y_l^c the coordinates of the central pixel of the *l*th connected component. Outliers, considered as those distances d_j^{min} out of 2 times plus/minus the standard deviation of D_{min}, where excluded ultimately from D_{min}. Finally, for a given connected component representing a berry candidate, CC_i, the normalized descriptor was defined as:

$$f_3 = \begin{cases} 1 & if \ \ max\left(R_{E_i}^1, R_{E_i}^2\right) \Big/ d > 1 \\ max\left(R_{E_i}^1, R_{E_i}^2\right) \Big/ d & otherwise \end{cases} \quad (8)$$

where

$$d = \overline{D_{min}}$$

- Colour discrimination (six descriptors): the algorithm described in this work was designed to analyse clusters at phenological stages before veraison, when berries are green regardless the grapevine variety is red or white. This colour feature was used to define four statistical descriptors.

Let CC_i be a connected component belonging to the set of berry candidates, $Berry_{cand}$; this is, CC_i represents a berry candidate. Thus, two $n \times n$ sub-images centred in the centroid of

CC_i are extracted from the images *A* and *B*; these sub-images are called $A_i^{n\times n}$ and $B_i^{n\times n}$. The n value was defined to the sub-images mainly contained the berry under evaluation, but not its surroundings. To make it automatically adaptive to different berry sizes, n was formulated exploiting the set of minimum distances among candidates, D_{min}, previously defined. First, the quartiles of the set, $q_1, \dots, q_4$, were calculated. Thus, for a given candidate *i*, its corresponding *n* value was obtained by computing:

$$n = \begin{cases} q_1 & if\ d_i^{min} < q_1 \\ q_2 & if\ q_1 \le d_i^{min} \le q_2 \\ q_3 & if\ d_i^{min} > q_2 \end{cases} \quad (9)$$

Next, two sets made up with the pixel values in the two sub-images, but excluding those pixels belonging to CC_i, were extracted:

$$\begin{aligned} p_A &= \{p_A^x | p_A^x \in A_i^{n\times n} \wedge p_A^x \notin CC_i\} \\ p_B &= \{p_B^x | p_B^x \in B_i^{n\times n} \wedge p_B^x \notin CC_i\} \end{aligned} \quad (10)$$

thus being p_A and p_B the sets from the sub-images $A_i^{n\times n}$ and $B_i^{n\times n}$, respectively. Note that the connected component CC_i was the segmented bright region produced by light reflection on the berry's surface, so it does not provide colour information. This is why its pixels were excluded from the sets. Then, the average and standard deviation calculated on the sets p_A and p_B, were finally selected as descriptors:

$$\begin{aligned} f_4 &= \overline{p_A}, \quad & f_5 &= \sigma(p_A) \\ f_6 &= \overline{p_B}, \quad & f_7 &= \sigma(p_B) \end{aligned} \quad (11)$$

Additionally, the normalized grey-level value of the central pixel in the sub-images, p_A^c and p_A^c, were also considered as descriptors:

$$f_8 = p_A^c/255; \quad f_9 = p_A^c/255 \quad (12)$$

As illumination is not completely homogeneous in the image (check Figure 4), note that f_4, f_5, f_6 and f_7, might not be flexible enough to give the correct response for two berries with substantial different illumination. To achieve this flexibility, f_8 and f_9 provides a local normalized intensity reference.

- Aggregation (one descriptor): Since berries develop in vines configuring clusters, they appear on images as forming constellations. This is why a relatively isolated candidate can be considered as having a very low probability of being an actual berry. This descriptor penalizes those candidates that do not belong to an aggregation of at least five candidates separated by a meaningful distance. This is, for a given candidate CC_i separated from its fifth closest neighbour by a distance $d_{i,5}$, the aggregation descriptor is calculated as:

$$f_{10} = \begin{cases} 1 & if \ {}^{d_{i,5}}/_{5d} > 1 \\ {}^{d_{i,5}}/_{5d} & otherwise \end{cases} \quad (13)$$

where *d* is the distance measurement defined in the size descriptor (eq. 8).

- Hu moments (seven descriptors): the Hu moment invariants (Hu, 1962) are a set of seven moments that are invariant under size, translation and rotation, and are used to recognize known patterns in images. The seven Hu moment invariants, $\phi_1, \dots, \phi_7$, can be derived from combinations of regular moments, as it was formulated by Gonzalez *et al.* (2004).

For most grapevine varieties, berries exhibit a circular shape in the images. This feature is distinctive from other objects in the scene and exploited for discrimination by means of modelling using the defined set of moments.

Given a connected component CC_i, that belongs to the set of berry candidates $Berry_{cand}$, a $n \times n$ sub-image from the image *L*, $L^{n\times n}$, centred in its centroid is extracted. Then, and image containing the external morphological gradient is computed as:

$$L^{n\times n}_{Grad} = \delta_B(L^{n\times n}) - \varepsilon_B(L^{n\times n}) \quad (14)$$

where δ and ε stands for the morphological dilation and erosion operators using the circular

structuring element *B* (Soille ,2004). Thus, a set of seven descriptors based on the Hu moments and added to the original classifier is defined to be calculated on $L_{Grad}^{n\times n}$ as:

$$f_{1i} = |log\ (\phi_i)|, i = 1, 2, \dots, 7 \quad (15)$$

With the set of the defined descriptors $\{f_{11}, \dots, f_{17}\}$, the NN was implemented with 17 input neurons, nine neurons in one hidden layer, and one neuron in the output layer. The NN was configured to apply the transfer function logarithmic-sigmoid, thus producing real outputs in the interval [0, 1]. These real values required additional processing to obtain a binary classification result which consisted on assigning the values 0 and 1 to false and true positives, respectively. To this end, the real NN's outputs were considered as probability values of the candidates to represent actual berries, and were used to create a probability map in form of image (check Figure 5-(a) and (b)). Finally, this image was binarized using the threshold automatically provided by the Otsu's method (Nobuyuki, 1979), and the connected components with a resulting value equal to 0 were discarded (check Figure 5-(c) and (d)).

2.5 Evaluation of the image analysis algorithm

The NN was trained to enable the recognition of berry candidates. To this end, the images used in the training process were firstly analysed following the methodology described in section 2.3.2. to obtain berry candidates. Then, in the output images, false berries were manually labelled, thus generating two sets of cases: positive cases (true positives) given as input to the NN to produce a 1 output, and negative cases given to generate a 0 output. The number of images required for an optimum training was empirically assessed using an iterative approach. First, four images were selected, and the cases contained in them were divided into three sets, namely training (75% of cases), validation (15% of cases) and testing (15% of cases). Thus, the training set was used to adjust the weights of the NN, the validation set prevented from overfitting issues and the testing set served to evaluate the performance of the NN. For each iteration, an additional image was incorporated, being its cases distributed into the sets respecting the previously defined weights. Comparing the different versions, the performance of the NN trained with seven and eight images

was similar, being the latter saved. Six out of the eight images (every image contains a three-vine segment hereafter) ultimately used for training were from the set of defoliated vines and two from the non-defoliated one.

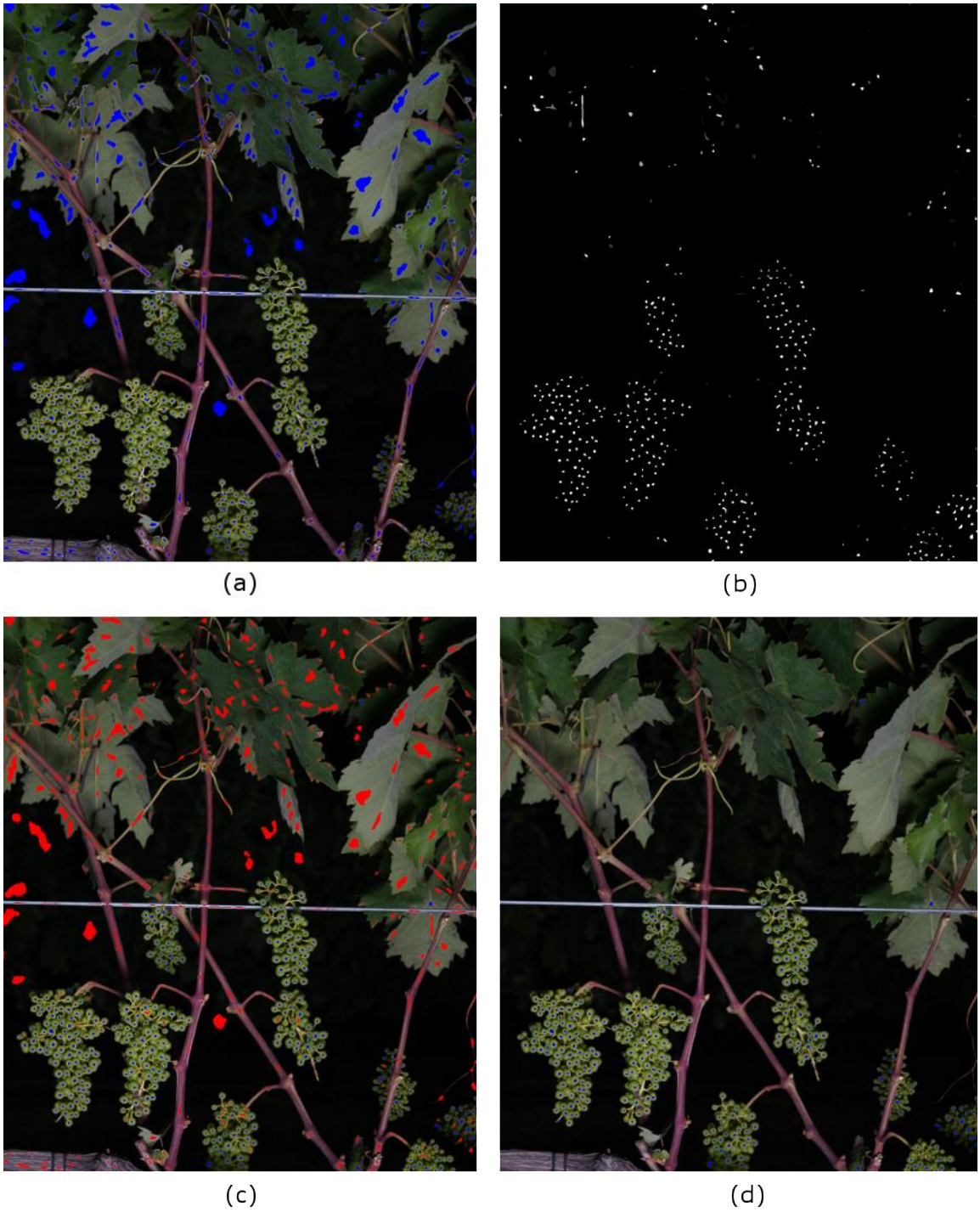


Figure 5.- Illustration of the process for discarding false positives: (a) image resulting from obtaining berry candidates (candidates represented in blue colour); (b) Probability image created

by representing the candidates with the probability computed by the neural network (the probability is scaled to the interval [0, 255] for image representation); the Otsu's method applied to this image provided a probability threshold of 114; (c) image showing the discarded and validated candidates in red and blue colour, respectively, decided by applying the calculated Otsu's threshold; (d) image illustrating the final results after false positives discarding.

Once the NN was optimised, the remaining images not used for training (44 defoliated and 11 non-defoliated) were analysed with the algorithm, and results evaluated by confronting output images with reference data. This reference set was created by manually labelling grapes in the images, making use of a tool developed for this purpose. Hence, for every image, and by comparison between the manual and automated outcomes, true positives (*TP*), false positives (*FP*) and false negatives (*FN*) were calculated per image according to the following definitions:

- *TP*: berries automatically detected corresponding to actual berries labelled in the reference data.
- *FP*: berries automatically detected not corresponding to actual berries in the reference data. When more than one berry candidate represents the same actual berry, only one was counted as *TP* and the rest were also considered as *FP*.
- *FN*: actual berries labelled in the reference data not found by the segmentation algorithm.

This approach allowed to measure (per image) the following two metrics based on contingency tables for binary classification:

$$RC = \frac{TP}{TP + FN}; PC = \frac{TP}{TP + FP} \quad (16)$$

where *RC* stands for Recall, which is the percentage of actual berries detected by the algorithm, whereas *PR* denotes Precision, and calculates the percentage of correctly detected berries. Additionally, the Pearson's coefficient of determination (R^2) was applied to assess the relationships between the actual berry number per image manually measured and the number of berries automatically detected by the algorithm.

The influence and determination of the *h* parameter value (eq. 1) was studied by analysing a subset

of images varying the *h* value, and assessing *h*-associated outputs using *RC* and *PR*. These pairs of values were combined with the harmonic mean of *RC* and *PC*, called *F*-measure, to decide the most optimum *h* value:

$$F = 2\frac{PC \times RC}{PC + RC} \quad (17)$$

This analysis is developed in section 3.1.

2.6 Evaluation of yield prediction

The number of detected berries was used for yield estimation. First, an average berry weight per variety was calculated from samples taken at the end of the previous season (measurements were taken from the same vineyard in October 2014). These average berry weights were then combined by multiplication with the number of visible berries detected per segment by the algorithm. Thus, the 'visible' yield per segment was predicted. These 'visible' weights per segment were correlated to the corresponding actual manually weighted yields to obtain a linear estimation model. Finally, this linear estimation model was applied to estimate the actual yield per segment and variety.

Yield predictions were assessed using the Root-Mean-Square-Error (RMSE), defined as:

$$RMSE_V = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i^v - y_i^v)^2}{n}} \quad (18)$$

where $\hat{y}_i^v$ and y_i^v are the predicted and actual yield for the *i*-th segment of the *v* variety, respectively, and *n* is the number of image segments. To avoid differences in yield between different varieties, a relative square error expressed as percentage was calculated as:

$$SE_v = \frac{RMSE_v}{\frac{\sum_{i=1}^{n} y_i^v}{n}} \times 100 \quad (19)$$

where *n* is the number of images for the variety *v* under evaluation. Finally, the relative error expressed as percentage was also computed:

$$|E_v| = \frac{|\sum_{i=1}^{n}(\hat{y}_i^v - y_i^v)|}{\sum_{i=1}^{n} y_i^v} \times 100 \quad (20)$$

3 Results and discussion

3.1 Evaluation results of the image analysis algorithm

The *h* parameter of the *h*-maxima transform (eq. 1) sets a threshold to decide whether a regional maximum is significant enough to be considered as a berry candidate. Therefore, the *h* value influenced the number, shape and size of the found berry candidates to be filtered afterwards by the NN, so it was firstly analytically established.

A subset of eight images was analysed with the algorithm varying the *h* value from 2 to 6, and Recall (*RC*) and Precision (*PR*) outputs were calculated for each value by means of comparison to the corresponding reference images. Table 1 shows the obtained average values of Recall ($\overline{RC}$) and Precision ($\overline{PR}$) for the different *h* values, and the *F*-measure calculations for every pair. The *F*-measure calculated for all $\overline{RC}$ and $\overline{PR}$ pairs reached the maximum at $h = 3$. At that point, the maximum $\overline{RC}$ was also obtained (maximum berry detection capability), coupled with a competitive $\overline{PC}$ value (low level of berry misclassification). Therefore, according to this analysis, although other values could also be valid (as claimed by other *F*-measure values), the *h* parameter was set to 3 hereafter.

Table 1.- Results of the segmentation methodology on a subset of images for several *h* values (see eq. 1).

***h* value**	$\overline{RC}$	$\overline{PR}$	$F - measure$
2	0.7976	0.9479	0.8663
3	0.8145	0.9532	0.8784
4	0.8077	0.9568	0.8759
5	0.7836	0.9602	0.8629
6	0.7599	0.9663	0.8508

With this parameter of the algorithm optimized, and the adaptive size of *n* used in eq. 10 defined, its performance was evaluated without interferences. Figure 6 shows a correlation study between the number of visible berries manually counted per image and automatically detected by the algorithm. The obtained R^2 value for the training (Figure 6-(a)) and external validation sets (Figure 6-(b)) showed a strong relationship between both magnitudes. Moreover, the determination coefficient, R^2, was higher and the slope closer to 1 for the external validation set, which discarded NN overfitting and guaranteed its generalization. Additionally, the measured *p*-values indicated no statistical differences between the manual and automated berry count for both cases ($p > 0.05$). On the other hand, Table 2 includes the algorithm's evaluation results in terms of *RC* and *PR* for the training and validation sets, detailed per variety and considering all images together.

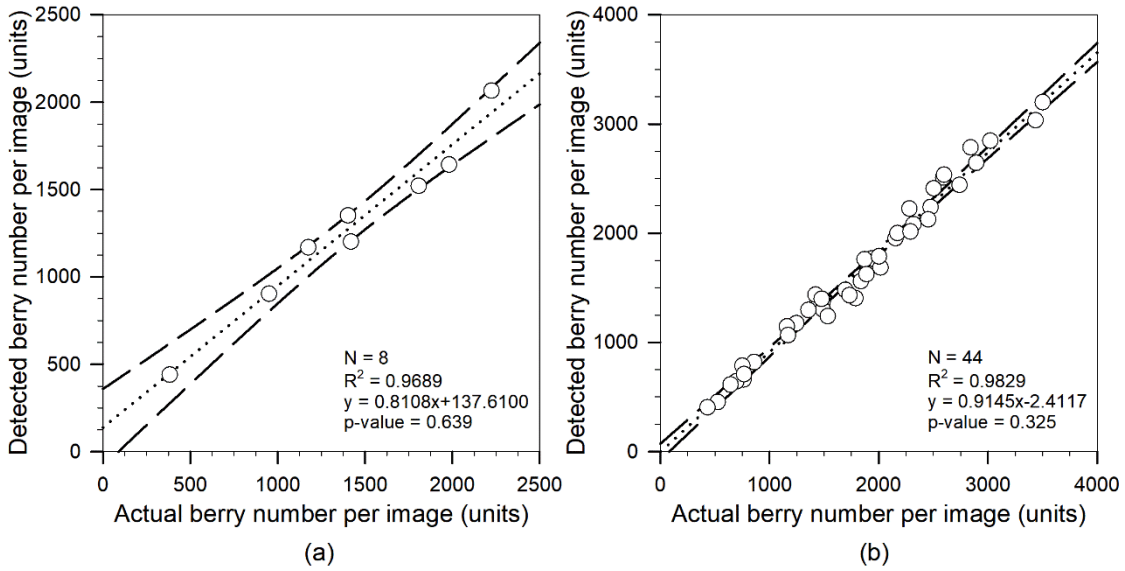


Figure 6.- Correlation study between the actual berry number per image and the berry number per image detected by the algorithm in the training set (a), and in the set for external validation (b).

RC and *PR* values were more consistent for the external validation set than for the training set. Indeed, *RC* and *PR* values ranged in the intervals [0.8041, 0.9200] and [0.8682, 0.9908] for the training set, and in [0.8327, 0.9067] and [0.9363, 0.9836] for the validation set. These figures

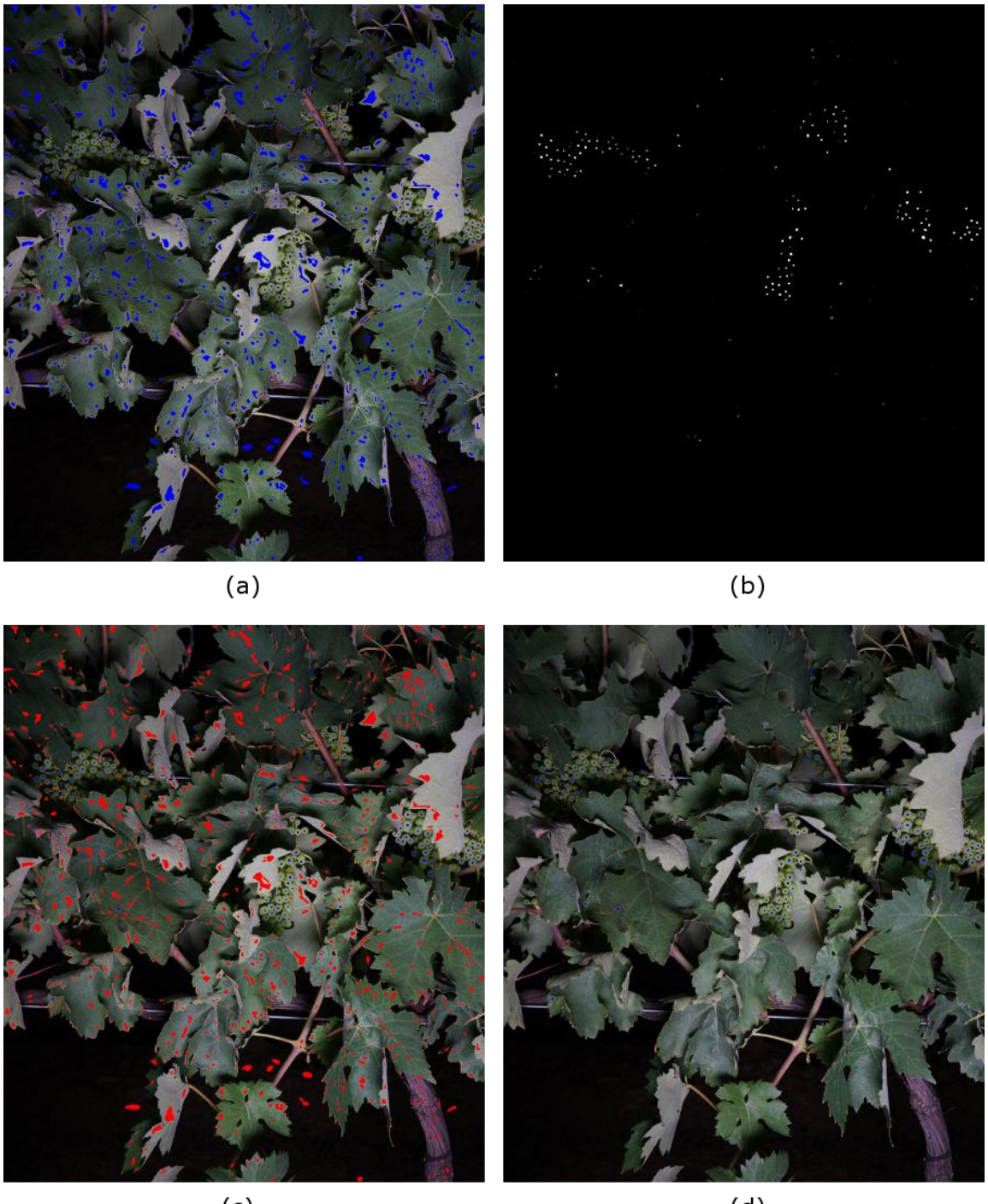
corroborated the achieved accuracy and generalization of the implemented NN.

Table 2.- Results of the image analysis algorithm for detecting berries measured on the training (N = 8) and validation (N = 44) set containing images from five grapevine varieties. Results are given in terms of Recall and Precision, detailed per variety and considering all images as a whole.

	Training set		**External validation**	
Grapevine variety	***RC***	***PR***	***RC***	***PR***
Albariño	0.9140	0.9515	0.8838	0.9836
Cabernet Sauvignon	0.8567	0.8682	0.8891	0.9363
Syrah	0.8041	0.9696	0.8327	0.9545
Tempranillo	0.8340	0.9908	0.8589	0.9674
Viognier	0.9200	0.9679	0.9067	0.9455
Overall	0.8746	0.9373	0.8764	0.9582

As it was mentioned before, the images evaluated above referred to defoliated vines. Defoliation is a common viticultural practice to promote air ventilation within the grapevine canopy to decrease fungal disease incidence, and to favour a more homogeneous and improved fruit ripening and composition (Diago *et al.*, 2010; Diago *et al.*, 2012). For the aims of the experiments of this work, it also provided better fruit exposure by limiting occlusions from leaves. Furthermore, non-defoliated vine images provide random fruit exposure, which make them unsuitable for the accurate yield predictions aimed by this work. Nevertheless, strictly focusing on the image analysis algorithm, the evaluation of non-defoliated images was considered an interesting stress test. Indeed, the set used for training the NN was configured to maximize its performance on defoliated vine images, only containing examples from two non-defoliated vine images. Therefore, challenging the trained NN to the non-defoliated vine images may provide a better idea of the berry descriptors' behaviour under non-optimum conditions. Figure 7 shows an

410 analysed sub-image of a non-defoliated vine. Table 3 contains the results produced by the
411 algorithm analysing the set of eleven segments of non-defoliated vines; results are given per
412 variety and considering all images as a whole in terms of *RC* and *PR*. As it can be confirmed, the
413 calculated average *RC* was similar to the one measured for images of defoliated vines (0.8483 vs
414 0.8746). Contrary, the average *PR* was lower for the non-defoliated images, but still remaining at
415 good performance rates (0.8544 vs 0.9582). Consequently, it can be concluded that these results
416 increase confidence in the robustness of the design set of berry descriptors.



417

418 **Figure 7.-** Illustration of the process for discarding false positives on a sub-image from a non-
419 defoliated vine: (a) detected berry candidates highlighted in blue dots; (b) probability image
420 generated by the neural network and to be thresholded applying the Otsu's method; (c) the
421 discarded and validated candidates after thresholding represented in red and blue colour,
422 respectively; (d) image resulting after false positives discarding.

Table 3.- Results of the image analysis algorithm for berry detection measured on the external validation set of images from non-defoliated vines (N = 11) of five different grapevine varieties.

Grapevine variety	*RC*	*PR*
Albariño	0.8494	0.8361
Cabernet Sauvignon	0.8552	0.8358
Syrah	0.7741	0.8746
Tempranillo	0.8750	0.8635
Viognier	0.9103	0.8262
Overall	0.8483	0.8544

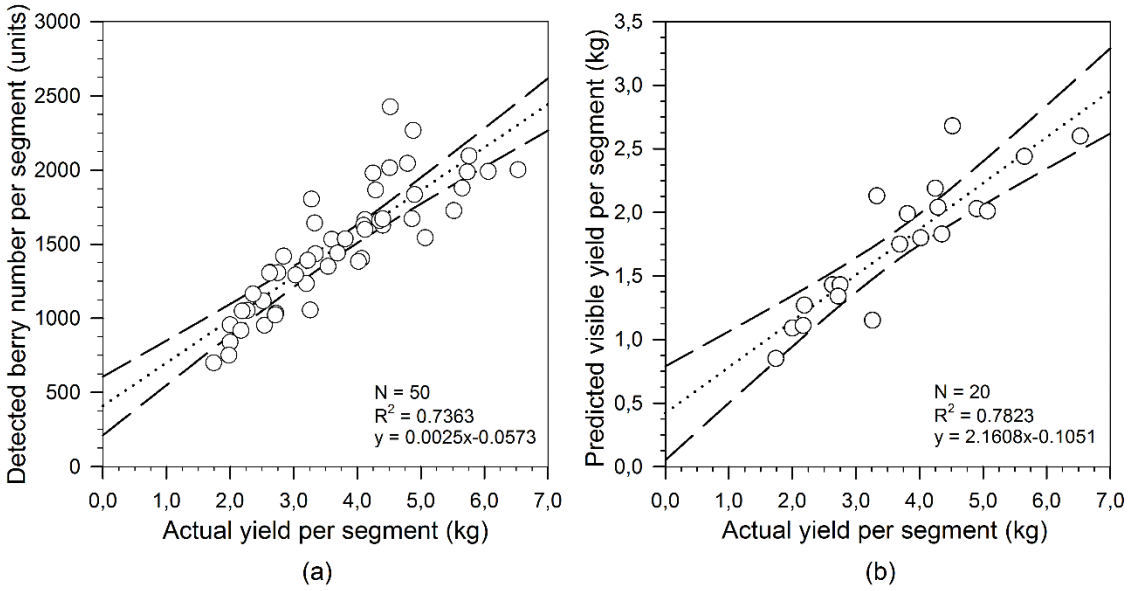
3.2 Evaluation results of yield prediction

From the regressions shown in Figure 8-(a), the number of detected visible berries was a good indicator of the actual yield. Based on this finding, a linear model for yield prediction was built. First, 20 segments, four per variety, which accounted for the whole dataset variability were selected for training. Then a predicted 'visible' yield was calculated for these segments by multiplying the available average berry weight per variety (computed from historical data) by the number of automatically detected berries. Thus, the weights corresponding to the berries visible in the segments were estimated. Finally, these predicted 'visible' yield values were linearly correlated to the actual yield per segment (reference data described in section 2.2) manually harvested to build the model (Figure 8-(b)). Thus, the linear model obtained in Figure 8-(b) was used to generate the yield predictions per variety detailed in Table 4 for the 30 segments kept for external validation. According to the results considering all images together, an average square error of 0.16 kg was made per vine, since the measured $RMSE_v$ of 0.48 kg referred to image segments composed of three vines. On the other hand, taking into account the relative square error (SE_v) per variety, the predictions ranged in the interval [8.74, 16.86] (expressed as %). This variability can be explained by the different phenological stages at which images were taken

442 depending on the variety (ranging between K and L (Baggiolini, 1952)). Indeed, there was a great
443 variability in cluster compactness between varieties at both stages, which influenced the
444 percentage of visible berries. Therefore, according to the results and this factor, it can be inferred
445 that specific predictive models for the stages K and L may surely produce even better predictions.
446 Curiously, the obtained overall relative error ($|E_v|$) was very close to 0, concretely 1.30. Contrary
447 to the SE_v, errors are accumulated with their signs (positive or negative) in the $|E_v|$ calculation,
448 and errors by excess and defect neutralized each other in this case.

449 Compared to previous studies, Font *et al.* (2015) achieved a relative square error of 16.0 % when
450 all the 25 cluster images (acquired at harvest time) were analysed together. In comparison, the
451 presented methodology yielded 12.83 % of error when considering the 30 segments containing
452 90 vines. Therefore, besides being more accurate, the presented methodology provides a more
453 efficient tool for vineyard management, as it performs at earlier stages on images containing tens
454 of clusters. Nuske *et al.* (2014) linearly correlated the berry counts produced by their algorithm
455 with the actual harvest crop weights collected from five *Vitis vinifera* L. and one table-grape
456 varieties. Images were taken at different phenological stages, ranging from berry-set to pre-
457 harvest. The measured correlation scores varied from R^2 = 0.60 to 0.73 depending on the variety,
458 although results were not specified per phenological stage nor in terms of weight error. The
459 present work casted a R^2 value of 0.7823 in yield prediction for the training set with five different
460 varieties, being this result specifically assessed at early phenological stage. In the study of Nuske
461 *et al.* (2014), yield prediction accuracy was evaluated along two seasons for a single table grape
462 cultivar (Flame Seedless) and yield prediction errors ranging from 6.5 to 11.7 % were reported.
463 These yield prediction accuracy values are similar to those obtained in the present work for five
464 grapevine varieties (Table 4). However, differently from this study in which yield prediction was
465 accomplished around 100 days before harvest, yield estimation was conducted at harvest time in
466 the work of Nuske *et al.* (2014). There are some studies in the literature facing yield prediction at
467 early stages in different crops, such as apple (Zhou *et al.*, 2012) or mango (Payne *et al.*, 2013). In
468 grapevine, early yield prediction (pre-veraison) provide outstanding advantages with respect to

469 pre-harvest yield estimation, when only limited corrective actions addressed to comply with
470 controlled yield regulations can be performed. In terms of the classifier's parsimony, a simple
471 image classifier fed with 17 descriptors was used in the present study, while Nuske *et al.* (2014)
472 compared three sets of texture descriptors giving as result classifier configurations of more than
473 30 descriptors after applying PCA analysis for dimensionality reduction.



474

475 **Figure 8.-** (a) Correlation study between the actual yield weighted per segment (three vines) and
476 the number of visible berries detected by the image analysis algorithm. (b) Calculation of the
477 linear model for yield prediction using a training subset of 20 image segments. The model was
478 computed by correlating the predicted 'visible' weights in the segments to the actual yield per
479 segment manually harvested. The predicted 'visible' weights per segment were obtained by
480 multiplying the number of detected berries by an average berry weigh measured per variety during
481 seasons previous to these experiments.

482

483 **Table 4.-** Results on yield prediction obtained for the 30 segments (three vines per segment) in
484 the external validation set. Outcomes are detailed per each of the five grapevine varieties (six
485 segments per variety).

Grapevine variety	Berry weight (g)*	Predicted visible yield (kg)	Predicted yield (kg)**	Actual yield (kg)	$RMSE_v$ (kg)	SE_v (%)	$\lvert E_v \rvert$ (%)
Albariño	1.10	12.05	25.41	27.18	0.58	12.80	6.51
Cabernet Sauvignon	1.09	9.99	20.95	20.23	0.42	12.46	3.56
Syrah	1.30	12.97	27.29	24.11	0.58	16.86	13.19
Tempranillo	1.30	11.81	25.00	26.32	0.46	8.74	5.01
Viognier	1.21	7.31	15.17	14.53	0.27	11.16	4.40
Overall	-	54.14	113.83	112.37	0.48	12.83	1.30

* Average berry weight per variety measured on a set of samples acquired from the same vines during seasons previous to this experiment.

** It was calculated by applying the linear model in Figure 8-(b).

The presented methodology paves the way towards its implementation in all sort of agricultural moving vehicles during a time window of two to three weeks, prior to veraison, to get an accurate estimation of final yield. In addition to the competitive advantage of having such early yield estimation, for logistic, economical, and vineyard management purposes, the developed image-based method can be used to generate maps representing the spatial yield variability in vineyards expected at harvest time. To maximize the exploitation of these valuable data, these maps could be available to viticulturists through smartphone or tablet applications, as well as through web services, directly in the vineyard, and could also be used by the new generation of variable rate machinery. Overall it constitutes a key tool for the grape and wine industry as it may help to define in a proper way, viticultural operations for yield adjustments to either meet legal regulations (as in many wine appellations), as well as to support decision making regarding harvest planning and arrangements.

4 Conclusions

This study presents a new, comprehensive technological solution for non-invasive and automated early grapevine yield prediction based on RGB images acquired on-the-go from a vehicle moving at a speed similar or higher than most agricultural vehicles. The developed tool was able to accurately predict yield by monitoring vines at phenological stages between berry-set and cluster-closure, that is around 100 days prior to grape harvest. Yield prediction models can be developed for individual varieties, but a global, multi-varietal solution has also proved effective, although larger training datasets acquired at similar working conditions should be added to maintain or increase the performance of the global solution. According to the detailed results per variety, it was found that even better outcomes could be achieved by developing prediction models adaptable to the specific phenological stage of the vineyard under analysis.

This advanced yield prediction would be of great utility for viticulturists to better adopt grapevine management operations, such as cluster thinning to regulate the yield in an efficient way, not only to comply with yield regulations but also to promote improved ripening and quality of the remaining fruit. Additionally, for the wine industry, early yield estimation can be also very beneficial, as the supply chain from grapegrowing to wine marketing strongly depends on anticipated yield estimation well before harvest. The presented solution enables the acquisition of a large amount of georeferenced images in a short time, hence of the spatial variability of yield prediction within a vineyard plot. This fact will favour the application of vineyard precision management techniques as well as the adoption of the new generation of variable rate machinery.

Acknowledgments

The authors want to acknowledge the European Union for funding the VineRobot project within the Seventh Programme for research, technological development and demonstration (Grant Agreement No 610953). Thanks also to the Agencia de Desarrollo Económico de La Rioja (ADER) for funding the VINETICS project (2012-I-IDD-0009). Borja Millán would especially like to acknowledge the research founding FPI grant 536/2014 by the University of La Rioja. Dr. Maria P. Diago is funded by the Spanish Ministry of Economy and Competitiveness (MINECO) with a Ramon y Cajal grant RYC-2015-18429. The authors are grateful to Vitis Navarra (Larraga,

Navarra, Spain) for the use of their vineyards to carry out this study.

References

Anderson, M.M., Smith, R.J., Williams, M.A., Wolpert, J.A., 2008. Viticultural Evaluation of French and California Pinot Noir Clones Grown for Production of Sparkling Wine. Am. J. Enol. Viticult. 59, 188-193.

Aquino, A., Diago, M.P., Millan, B., Tardaguila, J., 2017. A new methodology for estimating the grapevine-berry number per cluster using image analysis. Biosyst. Eng. 156, 80-95.

Baggiolini, M., 1952. Stades repères de l'abricotier. Revue Romande d'Agriculture, de Viticulture et d'Arboriculture 8, 28-29.

Bramley, R.G.V., Hamilton, R.P., 2004. Understanding variability in winegrape production systems. 1 - Within vineyard variation in yield over several vintages. Aust. J. Grape Wine R. 10, 32-45.

Chamelat, R., Rosso, E., Choksuriwong, A., Rosenberger, C., Laurent, H., Bro, P., 2006. Grape Detection by Image Processing. In: Proc. of the 32nd Annual Conference on IEEE Industrial Electronics, Paris, France, pp. 3697–3702.

Connolly, C., Fliess, T.A., 1997. Study of efficiency and accuracy in the transformation from RGB to CIELAB color space. IEEE Trans. Image Process. 6, 1046-1047.

Diago, M.P., Vilanova, M., Tardaguila, J., 2010. Effects of timing of manual and mechanical early defoliation on the aroma of Vitis vinifera L. Tempranillo wine. Am. J. Enol. Vitic. 61, 382–391.

Diago, M.P., Ayestarán, B., Guadalupe, Z., Garrido, Á., Tardaguila, J., 2012. Phenolic composition of Tempranillo wines following early defoliation of the vines. J. Sci. Food Agric. 92, 925–34.

Dunn, G.M. Martin, S.R., 2003. The current status of crop forecasting in the australian wine industry. In: ASVO Seminar Series, Grapegrowing at the Edge, Tanunda, Barossa Valley, South

554 Australia, pp. 4-8.

555 Dunn, G, 2010. Yield Forecasting. Grape and Wine Research and Development Corporation –
556 Australian Government, Australia.

557 Fernández, R., Montes, H., Salinas, C., Sarria, J., Armada, M., 2013. Combination of RGB and
558 Multispectral Imagery for Discrimination of Cabernet Sauvignon Grapevine Elements. Sensors-
559 Basel 13, 7838-7859.

560 Font, D., Tresanchez, M., Martínez, D., Moreno, J., Clotet, E., Palacín, J., 2015. Vineyard Yield
561 Estimation Based on the Analysis of High Resolution Images Obtained with Artificial
562 Illumination at Night. Sensors-Basel 15, 8284-8301.

563 Gonzalez, R.C., Woods, R.E., Eddins, S.L. (Eds.), 2004. Digital Image Processing Using
564 MATLAB. Pearson - Prentice Hall, Upper Saddle River, New Jersey, USA.

565 Hu, M.K., 1962. Visual Pattern Recognition by Moment Invariants. IRE Trans. Info. Theory 8,
566 179–187.

567 Li, L., Zhang, Q., Huang, D., 2014. A review of imaging techniques for plant phenotyping.
568 Sensors-Basel 14, 20078-20111.

569 Liu, S., Marden, S., Whitty, M., 2013. Towards Automated Yield Estimation in Viticulture. In:
570 Proceedings of the Australasian Conference on Robotics and Automation, Sydney, Australia.

571 Liu, S., Cossell, S., Tang, J., Dunn, G., Whitty, M. 2017. A computer vision system for early stage
572 grape yield estimation based on shoot detection. Comput. Electron. Agric. 137, 88-101.

573 Mack, J., Lenz, C., Teutrine, J., Steinhage, V., 2017. High-precision 3D detection and
574 reconstruction of grapes from laser range data for efficient phenotyping based on supervised
575 learning. Comput. Electron. Agric. 135, 300–311.

576 Martin, S.R., Dunn, G.M., Hoogenraad, T., Krstic, M.P., Clingeleffer, P.R., Ashcroft, W.J., 2002
577 Crop forecasting in cool climate vineyards. In: Proceedings of the 5th International Symposium
578 on Cool Climate Viticulture and Enology, Melbourne, Australia.

May, P., 2004. Development after fertilization, 2004. In: May, P. (Eds.), Flowering and Fruitset in Grapevines. Lythrum Press, Adelaide, Australia.

Nobuyuki, O., 1979. A Threshold Selection Method from Gray-Level Histograms. IEEE T. Syst. Man Cyb. 9, 62-66.

Nuske, S., Wilshusen, K., Achar, S., Yoder, L., Narasimhan, S., Singh, S., 2014. Automated visual yield estimation in vineyards. J. Field Robot. 31, 837-860.

Payne, A.B., Walsh, K.B., Subedi, P.P., Jarvis, D., 2013. Estimation of mango crop yield using image analysis–segmentation method. Comput. Electron. Agric. 91, 57-64.

Rahman, A., Hellicar, A., 2014. Identification of mature grape bunches using image processing and computational intelligence methods. In: Proceeding of the IEEE Symposium on Computational Intelligence for Multimedia, Signal and Vision Processing, Orlando, USA, pp. 1-6.

Reis, M.C., Morais, R., Pereira, C., Soares, S., Valente, A., Baptista, J., Ferreira, P.J.S.G., Cruz, J.B., 2011. Automatic Detection of White Grapes in Natural Environment Using Image Processing. In: Proceedings of the International Conference on Soft Computing Models in Industrial and Environmental Applications, Salamanca, Spain, pp. 19-26.

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. J. Appl. Logic 10(4), 285-290.

Rose, J.C., Kicherer, A., Wieland, M., Klingbeil, L., Töpfer, R., Kuhlmann, 2016. Towards automated large-scale 3D phenotyping of vineyards under field conditions. Sensors-Basel 16, 2136.

Smith, W.J. (Eds.), 2007. Modern Optical Engineering 4th ed. The Design of Optical Systems. McGraw-Hill Education – Europe, USA.

Soille, P. (Eds.), 2004. Morphological Image Analysis. Principles and Applications 2nd ed.

604 Springer – Verlag, Berlin, Germany.

605 Spalding, E.P., Miller, N.D., 2013. Image analysis is driving a renaissance in growth
606 measurement. Curr. Opin. Plant Biol. 16, 100-104.

607 Tardaguila, J., Blanco, J.A., Poni, S., Diago, M.P., 2012. Mechanical yield regulation in
608 winegrapes: comparison of early defoliation and crop thinning. Aust. J. Grape Wine R. 18, 344-
609 352.

610 Wolpert, J.A., Vilas, E.P., 1992 Estimating vineyard yields: Introduction to a simple, two-step
611 method. Am. J. Enol. Viticult. 43(4), 384–388.

612 Zhou, R., Damerow, L., Sun, Y., Blanke, M.M., 2012. Using colour features of cv. 'Gala' apple
613 fruits in an orchard in image processing to predict yield. Precis. Agric. 13(5), 568-580.