# *Grapevine yield estimation using image analysis for the variety Arinto*

***Ruben Bonaria***

Dissertation to obtain a Master's Degree in

**Viticulture and Oenology Engineering**

**Jury:**

PRESIDENT

PhD Joaquim Miguel Rangel da Cunha Costa, Assistant Professor at Instituto Superior de Agronomia, Universidade de Lisboa.

MENBERS

PhD Carlos Manuel Antunes Lopes, Associate Professor with Habilitation at Instituto Superior de Agronomia, Universidade de Lisboa;

PhD Ricardo Nuno da Fonseca Garcia Pereira Braga, Assistant Professor at Instituto Superior de Agronomia, Universidade de Lisboa.

2019

**Acknowledgement**

I'm thankful to the University of Turin and the University of Lisbon for giving me the opportunity to live part of my studying experience abroad.

I'm grateful to Professor Carlos Lopes and Professor Vittorino Novello for their support in the writing of the thesis.

I'm grateful to Gonçalo Victorino, for the continue support during all the research work and for the patience in answering my innumerable questions.

I'm grateful to Gonçalo, Luigi Mauro, Giuseppe Samà and Beatrice Carmignani, who made sure that the research work was a wonderful experience.

I'm thankful to the *Instituto Português do Mar e da Atmosfera* for the meteorological data used in this thesis.

I'm grateful to my parents, my grandma and my uncle Gilberto for supporting me in every way for this year in Lisbon and I thank my brother for his moral support.

I'm grateful to Beatrice Foscoli for being a point of reference, a safe haven on stormy days, a source of inspiration, for being a true friend. Furthermore, to had make me know his housemate, Federico Barbieri, a true force of the nature.

I'm grateful to Catarina Mattos who gave me every time her support and make me feel light and free as a feather hovering in the wind.

I'm thankful to all my classmates and all my new friends for their fantastic energy and for made this experience unforgettable.

I'm thankful to all my old friends that supported me during all the years before arriving at this point.

**Abstract**

Yield estimation can lead to difficulties in the vineyard and winery, if it is done inaccurately following wrong procedures, doing a non-representative sampling or for the human error. Moreover, the traditional yield estimation methods are time consuming and destructive because they need someone that goes into the vineyard to count the yield components and that take out from the vineyard inflorescence or bunches to count and weight the flowers and the berries. To avoid these problems and the errors that can occur on this way, the development and application of new and innovative techniques to estimate the yield through the analysis of RGB images taken under field conditions are under study from different groups of research.

In our research work we've studied the application of counting the yield components in the images throughout all the growing season. Furthermore, we've studied two different algorithms that starting from the survey of canopy porosity and/or visible bunches area, can help to do an estimation of the yield.

The most promising yield estimation, based on the counting of the yield components done through image analysis, was found to be at the phenological stage of four leaves out, which shown a mean absolute percent error (MA%E) of 32 ± 2% and an correlaion coefficient ($r_{Obs,Est}$) between observed and estimated shoots of 0.62.

The two algorithms used different models: for estimating the area of the bunches covered by leaves and to estimate the weight of the bunches per linear canopy meter. When the area of the bunches without leaf occlusion was estimated, an average percentage of occlusion generated by the bunches on the other bunches of 8%, 6% and 12% respectively at pea size, veraison and maturation, was used to estimate the total area of the bunches. When the total area of the bunches per linear canopy meter was estimated the two models to estimate the grape weight were used. Finally, to estimate the weight at harvest, the growth factors of 6.6 and 1.7 respectively, at pea size and veraison were used. The first algorithm shown a MA%E, between the estimated and observed values of yield, of -33.59%, -9.24% and -11.25%, instead the second algorithm shown a MA%E of -6.81%, -1.35% and 0.01% respectively at pea-size, veraison and maturation.

Key words: Arinto, Image Analysis, Precision Viticulture, Robotics, Yield Estimation, Vinbot.

**Resumo**

A estimativa de rendimento pode levar a dificuldades na vinha e na adega, se for realizada de maneira incorreta, fazendo uma amostragem não representativa ou por erro humano. Além disso, os métodos tradicionais de estimativa de produção são demorados e destrutivos, porque precisam de alguém que se desloque à vinha para contar componentes de rendimento e que recolha inflorescências ou cachos da vinha, destrutivamente, para contar e pesar as flores e os frutos. Para evitar estes problemas, o desenvolvimento e a aplicação de técnicas novas e inovadoras para estimar o rendimento através da análise de imagens RGB tiradas em condições de campo tem sido estudado em diferentes grupos de investigação.

A mais promissora estimativa de rendimento, baseada na contagem dos componentes de produção feitos pela análise de imagens, acabou revelando-se na fase fenológica de folhas livres, que obteve um erro percentual absoluto médio (EPAM) de 32 ± 2% e um coeficiente correlativo ($r_{Obs,Est}$) entre brotos observados e estimados de 0.62.

Os dois algoritmos usaram diferentes modelos: para estimarem a área dos cachos coberta por folhas e para estimar o peso dos cachos por metro linear do dossel. Quando a área dos cachos sem oclusão foliar foi estimada, uma média percentual de oclusão gerada pelos cachos em outros cachos de 8%, 6% e 12%, respectivamente em bago de ervilha, pintor e maturação, foi utilizada para estimar a área total dos chachos. Quando a área total de cachos por metros linear do dossel foi estimada, dois modelos para estimar o peso da uva foram utilizados. Finalmente, os fatores de crescimento de 6.6 e 1.7 respectivamente de bago de ervilha pintor foram utilizados para estimar o peso na colheita. O primero algoritmo mostrou um EPAM, entre os valores estimados e observados do campo, de -33.59%, -9.24% and -11.25% respectivamente em bago de ervilha, pintor e maturação. O segundo algoritmo mostrou um EPAM, entre os valores estimados e observados do campo, de -6.81%, -1.35% and 0.01%, respectivamente em bago de ervilha, pintor e maturação.

Palavras-chave: Arinto, Análise de Imagem, Viticultura de Precisão Robótica, Estimativa de Rendimento, Vinbot.

**Resumo alargado**

A estimativa de rendimento pode levar a dificuldades na vinha e na adega, se for realizada de maneira incorreta, fazendo uma amostragem não representativa ou por erro humano. Além disso, os métodos tradicionais de estimativa de produção são demorados e destrutivos, porque precisam de alguém que se desloque à vinha para contar componentes de rendimento e que recolha inflorescências ou cachos da vinha, destrutivamente, para contar e pesar as flores e os frutos. Para evitar estes problemas, o desenvolvimento e a aplicação de técnicas novas e inovadoras para estimar o rendimento através da análise de imagens RGB tiradas em condições de campo tem sido estudado em diferentes grupos de investigação.

Em nosso trabalho de pesquisa nós estudamos a aplicação da contagem dos componentes do rendimento nas imagens por toda a temporada de crescimento. Ademais, estudamos dois diferentes algoritmos (A e B) que, começando pela pesquisa da porosidade de dossel e/ou áreas de cachos visíveis, pode ajudar a fazer uma estimativa do rendimento.

A mais promissora estimativa de rendimento, baseada na contagem dos componentes de produção feitos pela análise de imagens, acabou revelando-se na fase fenológica de folhas livres, que obteve um erro percentual absoluto médio (EPAM) de 32 ± 2% e um coeficiente correlativo ($r_{Obs,Est}$) entre brotos observados e estimados de 0.62.

Os dados para construir os modelos estimativos utilizados em nossos algoritmos para a previsão de rendimentos foram obtidos pela análise de imagens coletadas em 20 metros linear do dossel para cada fase fenológica (bago de ervilha, pintor e maturação) dentro do campo. A coleta de imagens foi realizada em dossel não perturbado, parcialmente e completamente desfolhadas e para cada remoção de camadas dos cachos (a camada do cacho é representada por cachos que na imagem não estão cobertas e estão cobrindo outros cachos) da zona de frutificação. Ademais, todos os cachos foram pesados e parte deles foi fotografada no laboratório e suas imagens foram também analisadas para a coleta de dados para a construção do modelo.

Os dois algoritmos A e B usaram diferentes modelos: dois diferentes modelos para estimarem a área dos cachos coberta por folhas e dois diferentes modelos para estimar o peso dos cachos por metro linear do dossel.

Das imagens tiradas em dossels não perturbado, parcialmente e completamente desfolhados, foram coletados os dados de porosidade e de área de cachos projetada.

Quando a área dos cachos sem oclusão foliar foi estimada, uma média percentual de oclusão gerada pelos cachos em outros cachos de 8%, 6% e 12%, respectivamente em bago de ervilha, pintor e maturação, foi utilizada para estimar a área total de cachos sem oclusão de cacho pelo

cacho. Esses dados foram obtidos pela análise de imagens das imagens coletadas nas fotos tiradas no metros linear do dossel sujeita a remoção do cacho.

Quando a área total de cachos por metros linear do dossel foi estimada, dois modelos para estimar o peso da uva foram utilizados, obtendo a estimativa do peso. Os dados utilizados para construir os dois modelos foram coletados pesando os cachos pela análise de imagens das fotos dos cachos tiradas dentro do laboratório e pelas fotos tiradas no campo, obtendo a área do cacho projetada e a área total de cachos projetados por metro linear do dossel.

Finalmente, os fatores de crescimento de 6.6 e 1.7 respectivamente de bago de ervilha pintor são utilizados para estimar o peso na colheita. Os fatores de crescimento foram calculados dividindo a média de peso dos cachos por metro linear do dossel na maturação por isso em bago de ervilha e pintor.

O algoritmo A envolve o uso de dois modelos estimativos: o primeiro usou a porosidade de dossel para estimar a área dos cachos sem oclusão foliar; o segundo usou a área projetada do cacho para estimar o peso dos cachos por metro linear do dossel. Os resultados desse algoritmo mostrou um EPAM, entre os valores estimados e observados do campo, de -33.59%, -9.24% and -11.25% respectivamente em bago de ervilha, pintor e maturação.

O algoritmo B envolve o uso de dois modelos estimativos: o primeiro usou a porcentagem de área visível de cachos para estimar a área dos cachos sem oclusão foliar; o segundo usou o total de área de cachos projetada sem oclusão foliar e sem oclusão de cacho pelo cacho para estimar o peso dos cachos por metro linear do dossel. Os resultados desse algoritmo mostrou um EPAM, entre os valores estimados e observados do campo, de -6.81%, -1.35% and 0.01%, respectivamente em bago de ervilha, pintor e maturação

# Table of contents

## List of figures

## List of tables

## Acronyms

**a*** greenness-redness coordinate

**AC** accuracy

**APCFM** Agropalynoclimatological forecast models

**ASI** Autonomous Solution

**b*** blueness-yellowness coordinate

**BR** Blue, Red

**BxB** bunch by bunch

**CCR** correct classification rate

**CIE** Commission Internationale d'Eclairage

**CNNs** convolutional neural networks

***Csa* C**: warm temperature; s:summer dry; a:hot summer

**ELA** Exposed Leaf area

**Est** estimated

**F1** F1score

**Faster RCN** architecture of the CNNs

**FCNs** Fully Convolutional Networks

**GB** Green, Blue

**GNSS** Global Navigation Satellite System

**GPS** Global Positioning System

**HSB** Hue, Saturation and Brightness

**HVS** human visual system

**ICT** Information and Communication Technology

**IOU** Intersection Over Union

**IPMA** Instituto Português do Mar e da Atmosfera

**ISA** Instituto superior de Agronomia

**L*** Luminance coordinate

**LAI** leaf area index

**LCM** Linear canopy meter/s

**LiDAR** Laser Rangefinder

**MA%E** mean absolute percent error

**MAE** Mean Absolute Error

**MTLS** mobile terrestrial laser scanner

**NN** neural network

**Obs** observed

**PQA** Point Quadrat Analysis

**PR** Precision

$r_{Obs,Est}$ correlation coefficient

$R^2$ determination coefficient

**RC** Recall

**RG** Red, Green

**RGB** Red, Green, Blue

**RMSE** Root Mean Square Error

**ROS** Robot Operating System

**SLAM** Simultaneous Localisation and Mapping

**SPs** smart-points

**SVM** Support Vector Machine

**TTMs** Trellis tension monitors

**U-Net** architecture of FCNs

**VB** Visible Bunches

**VRC** Vision Robotics Corporation

## 1 Introduction

There are different problems with yield forecasting: it is a prediction and not a punctual measurement, there is spatial and temporal variation, therefore growers must keep their records accurate. There are several reasons linked to temporal variation of the yield which are the effect of yield during the previous season, fruitfulness, fruit-set and berry development (Kultural and O'Daniel, 2007). The spatial variation is caused by the type of soil, exposition, age of the vine and altitude. Moreover, yield forecast can be easily done inaccurately, as it is common to forecast by a subjective "eye balling" of a block. Even when objective methods are used unreliable numbers can be in their calculation. Another problem can be that an inappropriate sampling procedure is done (Clingeleffer, 2001). For these reasons, there is a strong demand to develop better yield forecast methods (Dunn, 2010).

The development and application of new and innovative techniques is a key issue in viticulture research to improve grape-growing sustainability and quality of the production. The goal is to develop a simple, cheap, fast, accurate and robust image analysis methodology to be applied to RGB (Red, Green, Blue) images taken automatically under field conditions (Diago *et al*., 2014). To this regard the EU VINBOT (Autonomous cloud-computing vineyard robot to optimize yield management and wine quality) project focused on vineyard yield estimation was funded under the FP7 SME program (Lopes *et al*. 2017; Guzman *et al*., 2016 a).

### Aim of the work

The grapevine varieties under study by our research group are Syrah, Encruzado and Arinto; the last one variety is the object of this thesis.

Lopes *et al*. (2017) obtained an underestimation of the actual yield which however presented a similar trend to the actual values. While getting a significant coefficient of determination ($R^2$) of 0.31* regarding the relationship between actual and estimated values, the equation of the fitted line indicates that the Vinbot algorithms underestimated the yield by an additive factor. This underestimation may be explained by different kind of occlusions, as noted by Nuske *et al*. (2014 b) who proposed some solutions based on modelling and calibration of occlusion ratio by the bunch and leaf induced occlusions.

Therefore, our aims were: analysing the existing relationship between several yield components throughout the season and the final yield; developing models to estimate the non-visible grape bunches at harvest and at earlier stages of grapevine development to estimate the final yield; contributing to the development of the Vinbot platform by improving its yield estimation algorithms and methodology.

## 2 Literature review

Since vineyards, like other open field crops, haven't got a growing substrate and uniform microclimatic conditions (Boselli *et al.*, 2016) and this variability causes different vine physiological response with direct consequences on grape quality and yield (Matese and Di Gennaro, 2015), it's necessary to know the pedological characteristics (structure, texture, depth, geomorphology, exposure) and the hydrogeological conditions (meteorological trends, surface and underground water flows) to balance the production factors and differentiate production quality (Boselli *et al.*, 2016). The introduction of new technologies for supporting vineyard management can increase the efficiency and quality of production and, at the same time, improve the economic and environment sustainability (Boselli *et al*., 2016; Matese and Di Gennaro, 2015).

To practice the precision viticulture, it's necessary keeping in mind that there are two types of geodiversity: permanent and temporary diversity. Permanent geodiversity is determined by factors that don't change easily over time, such as the physical structure of the soil and subsoil. Temporary geodiversity is determined by a constant modification of environmental parameters and the availability of nutrients for the crop (Boselli *et al.*, 2016).

To detect the variability in the vineyard and to manage different operations needed to obtain the best quality of the grape and improve the economic and environmental sustainability there are different technologies for precision viticulture as reviewed by Boselli *et al.*, (2016) and by Matese and Di Gennaro (2015): optoelectronic, that is a branch of electronics, it studies electronic devices that interact with light and their applications, it represents the interface between the electrical and the optical domain; remote and proximal sensing to study the vineyard status like, vigour, water availability, plant and pathogen attacks and soil conditions and other parameters such as fruit composition and yield can be estimated using principles of precision viticulture (Gatti *et al*., 2017; Lopes *et al.*, 2017); ICT (Information and Communication Technology) technology that represent all the methods and technics used in transmission, reception and elaboration of the information, web and digital technologies (Boselli *et al*., 2016); systems of satellite navigation, like GNSS, Global Navigation Satellite System (Boselli *et al*., 2016; Matese and Di Gennaro, 2015).

Although the precision viticulture is based on remotely sensed multispectral images acquired by satellites and manned and unmanned aircraft (Matese and Di Gennaro, 2015), various new and more flexible devices for proximal sensing based on vision systems, laser scanning, ultrasonic and spectral acquisition are developed (Gatti *et al*., 2017).

### 2.1 Robots in viticulture

As reviewed by Matese and Di Gennaro (2015) the use of the robots in viticulture is in a stage of prototype, but some projects are already on the market.

In Spain it was developed, in VineRobot project, a robot provided with sensors of fluorescence, multispectral and RGB cameras for machine vision, thermal infrared and GPS (Global Positioning System). It was designed to perform a proximal monitoring of various parameters like yield, vigour, water stress and quality of the grapes (Diago *et al.,* 2015a).

Another robot called Wall-Ye, developed by Christophe Millot for vineyard monitoring, can move independently along the rows, acquiring data for each vine and produce a very highly detailed vineyard map. Using a monitoring system based on many optical sensors, it can perform correct displacements within the vineyard and carry out precision pruning, respecting the structure of each vine. It can prune about 600 plants per day (it has an autonomy of 12 hours) (Diago *et al.,* 2015b).

In Israel it was developed a prototype for vineyard spraying, called VineGuard. It can move within the vineyard using a complex set of sensors, and an arm for harvest, using an artificial intelligence for localisation assessment of the maturation state, selection and detachment of the grapes from vine (VineGuard, 2017).

In France it was developed a robot called Vitirover that can cut the grass up to 2-3 cm from the vine with a cutting height between 4 and 10 cm and can work with a maximum slope of 15%. The power system is completely independent thanks to a solar panel. This robot is slow needing 100 hours to cover 1 ha. It can work using GPS coordinates or can be controlled by remote using an application on mobile device (Vitirover, 2018).

The VRC (Vision Robotics Corporation) has developed a prototype to perform a precision pruning using sensors to perform a 3D reconstruction of the vine. The robot identifies the point where to cut according to the specification provided by the user, using two hydraulic shears (Matese and Di Gennaro 2015).

The ASI (Autonomous Solution) has developed a robot tractor prototype that can be driven by remote or can be autonomous but can also endowed with a cab for the presence of one operator on board. It is a real tractor capable to do agricultural tools like a normal tractor (Matese and Di Gennaro 2015).

## 2.2 Importance of the yield forecasting

It is known that competition on wine market is increasing with the increase of globalization. The increasing in competition forces the producers to professionalize their product regarding aspects like product differentiation (Dunn, 2010), regional pricing negotiations, crusher intake scheduling, tank space allocation, investment in the winery and development of marketing strategies (logistic problems of production and distribution) (Clingeleffer, 2001).

Although there are various other reasons that lead producers to increase their competitiveness, the ones mentioned have all something in common: they are dependent on yield forecasting. Inaccurate yield forecasting can lead to difficulties at the vineyard and winery (MacMillan and Fisher, 2005; Clingeleffer, 2001). In the vineyard, accurate yield forecasting can help to know how much crop expect or assist decision making on how maintain the yield for the future crop and to meet a target value of fruit composition (Kultural and O'Daniel, 2007). Indeed, concerning the last point, the growers can expect penalties when their grapes do not achieve the minimum standard quality to sell to winery, ranging from price reduction to the total rejection of the production (MacMillan and Fisher, 2005).

## 2.3 Strategies for yield forecast

Grape yield is defined by the yield components, involving the number of grape clusters, the number of berries per cluster and the berry size (Tardaguila *et al*., 2012). The number of berries per cluster is a very labile variable, more than other yield components, even within a given genotype (Anderson *et al*., 2008; Diago *et al*., 2015). It is influenced by the number of flowers per inflorescence and by the fruit-set rate, which are highly dependent on the weather conditions during inflorescence development at bud dormancy and berry set, respectively (May, 2004). Flower number per inflorescence shows a strong variability among vines and within inflorescences of the same vine, therefore, a count of the flower number per inflorescence is essential for accurate assessment of fruit set (Diago *et al*., 2014) while the number of berries per cluster is fully established at berry set and remains mostly invariable until harvest (Aquino *et al*., 2017).

The knowledge of the rate of fruit set, at very early stages, is of great value for grape growers, as this variable can be used to estimate the final yield at harvest, provided a historical value of average berry weight, and the average cluster number per vine for each vineyard is available for each vineyard (Aquino *et al*., 2015). Flowering and fruit set, together with berry size, have also a great impact on grape and wine quality, since they define the number of berries per cluster and contribute to determine the cluster architecture and compactness, which are a recognized key indicators of grape and wine quality (Matthews *et al*., 2007). Despite its importance, limited flower counting, and fruit-set estimation are currently carried out in commercial vineyards, as manual flower counting is very laborious and destructive (Aquino *et al*., 2015).

Traditional yield estimation methods are based on yield components, that must be collected every year: number of bunches per unit length of vine row, number of berries per bunch, bunch weight and berry weight. The accuracy of the forecast, based on measurement, depends on the inputs to the formulae that are used: use of the correct formula, accurate knowledge of patch dimensions, adequate sampling to estimate means of yield components and accurate prediction of components not yet determined at forecasting time (Martin and Dunn, 2003).

Yield predictions can be attempted at any time during the growing cycle, although they become more accurate later in the cycle (Folwell *et al*., 1994).

When any sort of forecast is done, to avoid any bias in sampling, it's important include weak and missing plants. It is recommended a "stratified random" sampling, the block is split into equal sized segments and vines randomly selected within these. We must use the samples number calculated as showed in equation 1, to obtain a representative sample (Dunn 2010):

$$n = \frac{t^2 * CV^2}{PE^2} \qquad \text{(Eq. 1)}$$

where *n* is the number of samples required, *t* is the t value from Student *t* tables, *CV* is the coefficient of variation and *PE* is the percentage of error from the mean for the desired estimate. The use of the correct formula is important and must also considered the harvest efficiency that reflects of the ratio of the actual delivery of grape to the winery to the amount of grape in the block in the harvest time. The typical harvest efficiency factors individuated by Dunn (2010) are: meticulous hand harvesting very close to the winery (1); hand harvesting with transfer to a distant winery (0.95); very efficient machine harvesting with small transport losses (0.90); inefficient machine harvesting with transport losses (0.85).

"More experience and more data would increase the accuracy of estimates in upcoming seasons. Perennial record keeping for cluster weights from year to year improves crop estimation" (Sabbatini *et al*., 2012).

### 2.3.1Traditional yield estimation methods

The decision of the number of buds to retain at pruning is the first method to establish the yield, but it isn't always accurate due to seasonal variation of bud fruitfulness and bunch weight. Fruitfulness varies somewhat due to many factors: genetic characteristics, vine management and weather conditions during the previous year. So, there is a need for bud fruitfulness testing to obtain a more accurate yield prediction using the number of buds retained at pruning. To generate a target bud number to retain at winter pruning we need a target yield, vine density, bud fruitfulness (which can be obtained by bud dissection) and bunch weight data plugged into the "pruning equation". Bud dissection can only determine fruitfulness, the potential bunch size is formed around the budburst period of the new season which is influenced by the temperature and therefore it will influence yield

potential. The number of flowers that set determines the number of berries per cluster, that therefore impacts on the potential bunch weight. The number of berries that set is also heavily dependent on weather conditions, mostly wind and rain, when flowering take place (MacMiller and Fisher, 2005).

After budburst, to assess the potential cluster size, one way is counting the flowers, although the process is labour-intensive and impractical for most growers. Good correlations exist between cluster length, dry weight and flower number, although these can be valid for a population of clusters in any given year, is not true in another season, for another variety or other site. The most promising of the early predictors of harvest cluster weight is the number of primary branches on the cluster (Sabbatini *et al*., 2012).

The number of primary branches is a structural character that is determined soon after budburst and as it remains relatively constant until flowering the timing of sampling is flexible; another advantage is that the measure is non-destructive. Large seasonal deviation of cluster weight can be detected and a strong functional relationship between the number of primary branches and the number of flowers per cluster was observed and it remains relatively stable season to season. However, some places can present difficulties when conditions during fruit-set are adverse. The relationships differ between the varieties, so the growers must establish their own predictive relationships for other varieties (Dunn, 2010).

Another method consists of using the number of clusters and the cluster weight at harvest time of the previous year, using the equation 2 (Sabbatini *et al*., 2012):

$$PY = \frac{ANV * NC * CW}{1000} \qquad \text{(Eq. 2)}$$

- *PY*: predicted yield, (ton/ha)
- *ANV*: actual number of producing vines/ha
- *NC*: number of cluster/vine
- *CW*: cluster weight (kg)

The number of cluster varies with the type and intensity of canopy management practices (level of pruning, cluster thinning) and it can be counted (with sampling) in different period but the counting is easier at bloom, because they are more visible. The weight varies from year to year depending on weather conditions, level of canopy management, irrigation, fertilisation, fungal disease, insect feeding damage and bird depredation. Cluster weight must be obtained from the same vines where cluster numbers were counted (Sabbatini *et al*., 2012).

The lag-phase method is like the method discussed before but is based on collecting cluster weight during the lag-phase. At this stage the cluster weight has reached half of the final weight although the multiplier can change among cultivars and seasons, therefore each grower must calculate their own. The formula used in this method is equation 3 (Sabbatini *et al.*, 2012):

$$PY = \frac{ANV * NC * LagCV * HM}{1000} \quad \text{(Eq. 3)}$$

- *PY*: predicted yield (ton/ha)
- *ANV*: actual number of producing vines/ha
- *NC*: number of clusters/vine
- *LagCW*: cluster weight at lag-phase (kg)
- *HM*: harvest multiplier

### 2.3.2 Airborne pollen samples for early-season estimates

This method is based on the determination of potential production by measuring airborne pollen concentration in the air at flowering with Cour traps followed by an assessment of possible impact of post-flowering conditions. Agropalynoclimatological forecast models (APCFM) can explain from 97% to 99% of the annual variability in regional wine production, considering agronomic and weather conditions after flowering (Cunha *et al*., 2003). Considering the result that are been obtained in Portugal, the APCFM is a valuable tool to do a good prevision at flowering, in regions characterized by late flowering and consistent post-flowering water stress. Anyway, additional parameters are needed, in regions characterized by early flowering or rainy post-flowering period and great interannual variation in water stress, such as disease occurrence, agronomic and weather condition after flowering (Cunha *et al*., 2003).

### 2.3.3 Trellis force estimation

Trellis tension monitors (TTMs) is a system that measure the changes in the tension of the trellis support wire on which grapevines are trained. Before the bloom the estimates of yield are irrelevant because the oscillation of the tension is ascribable to the canopy growth, only after bloom the differences in the tension is ascribable to fruit mass (Blom and Tarara, 2009).

This technique requires a permanent infrastructure to be installed and the variation in the yield is not the only cause of the variations in trellis tension, the temperature and vine size that can differ for each plant can give rise to loss of accuracy when the trellis tension is related to the yield (Nuske *et al*., 2014).

## 2.4 Image analysis

The development and application of new and innovative techniques, with the objective of monitoring the vineyard is a key issue in viticulture research to improve grape-growing sustainability and quality of the production, through the analysis of RGB images taken under field conditions. However, the performance of a computer vision system only based on colour information is dependent on many factors that must be studied and understood as illumination conditions, acquisition angle and object composition (Diago *et al*., 2014).

### 2.4.1 Image related information

#### *2.4.1.1 Colour spaces – HVS, CIE*

Colour is the way the HVS (human visual system) measures the visible part of the electromagnetic spectrum, approximately between 300 and 830 nm. The HVS based colour spaces include: the RGB colour space and the phenomenal colour spaces. These colour spaces are motivated by the properties of the HVS (Tkalčič and Tasič, 2003). The main disadvantage of the RGB colour space is the high correlation between its components: about 0.78, 0.98 and 0.94 between BR, RG and GB, respectively (Palus, 1998; Tkalčič and Tasič, 2003).

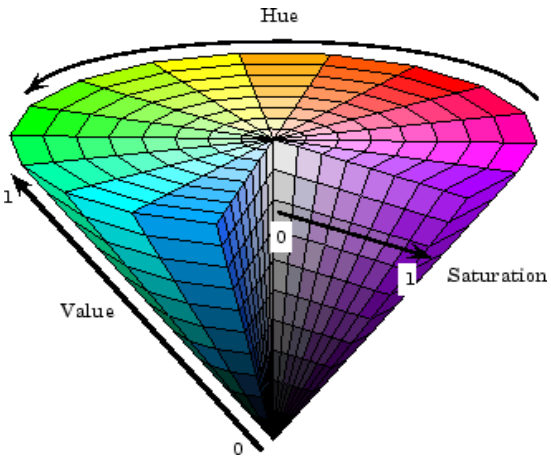



*Figure 1) A phenomenal colour space (source: Mathworks (2019)).*

The phenomenal colour spaces (Fig. 1) are the mind's representation of colours; these colour spaces use 3 attributes to describe a colour: Hue, Saturation and Brightness, (HSB). Hue is the attribute which tells if the colour is red, green, yellow, blue… Saturation is the level of non-whiteness. An extremely saturated colour has only one spectral component while an unsaturated colour has lots of white added. Brightness is a measure of the light intensity. These colour spaces are deformations of the RGB colour space (Cotton, 1996; Tkalčič and Tasič, 2003). Although the phenomenal colour spaces are very intuitive to use, they have a few shortcomings which limit their use in practical applications. Since they are mostly linear transformations from RGB they do not include any information about chromaticity and white point. (Poynton, 1997; Tkalčič and Tasič, 2003).

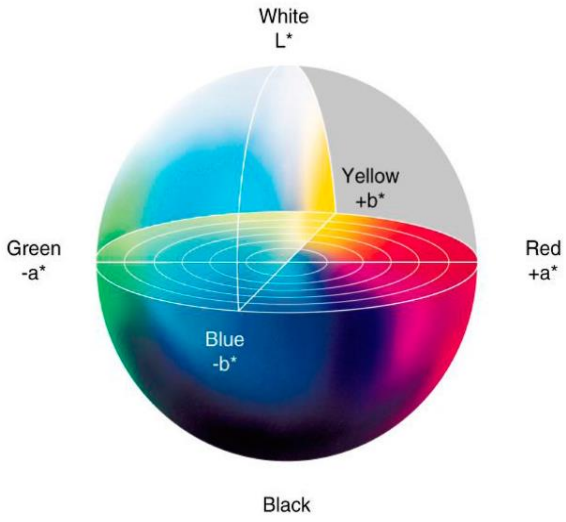



*Figure 2) CIE Lab colour space (source: Cortez et al., 2017).*

CIE Lab colour space (Fig. 2), is an international standard for colour measurement, developed from the *Commission Internationale d'Eclairage* (CIE) in 1976, this colour space is a transformation from the CIE XYZ colour space that describes any colour thanks to tristimulus positive values. A very important attribute of the CIE XYZ colour space is that it is not device dependent, and therefore also its colour spaces-transformations. CIE Lab colour space normalizes its values by the division with the point, it is represented by three coordinates called L*, a*, b*. The term L* is the luminance ranging from 0 to 100, which is combined with the other two coordinates: a* represents the greenness-redness, whereas b* represents the blueness-yellowness (CIE Colorimetry, 1986; Tkalčič and Tasič, 2003; Wyszecki and Stiles, 1982).

#### *2.4.1.2 Traditional and machine learning approaches to image analysis*

While traditional approaches to image segmentation (process of partitioning an image into multiple segments) it is typically used to locate objects and boundaries in images (Tan, 2016)) employ handcrafted heuristic criteria (e.g., intensity and colour distributions) to identify appropriate image, deep learning convolutional neural networks (CNNs), which is an evolution of the standard neural network (NN), allow learning descriptive criteria of the desired image regions just from the image data itself (Rudolph *et al*., 2018).

NN consists of many simple, connected processors called neurons, which produce a sequence of real-valued activations. Input neurons get activated through sensors perceiving the environment, while others get activated through weighted connections from previously active neurons (Schmidhuber, 2015).

CNNs used for image classification classify complete images and generally follow a common structure that shows two phases. The first is the feature extraction phase, where multiple convolution layers and pooling layers generate successively more complex class characteristic image features (in the convolution layers) thereby down sampling the image size (in the pooling layers). The other phase regards the classification, multiple fully connected layers derive class labels based on the derive image feature (Rudolph *et al*., 2018). CNNs have established themselves as a state-of-the-art method for many tasks of image processing, including image classification (Krizhevsky *et al.,* 2012; Simonvan and Darrel, 2015) as well as, more recently, image segmentation (Long *et al.,* 2015; Ronneberger *et al.,* 2015).

Two approaches to CNN-based image segmentations was used by Long *et al*. (2015). Firstly, they used the Fully Convolutional Networks (FCNs) for image segmentation which shows two phases: the feature extraction phase (as in the classification networks) followed by a decoder phase that results in a classification on the original image resolution (i.e. assigns a class label to each pixel of the image); then they used U-Net (Ronneberger *et al.,* 2015) that is an architecture of FCNs and it can be trained end-to-end, which was used for the segmentation (Long *et al*., 2015). Faster RCN is another kind of architecture of the CNN (Ren *et al*., 2015).

### 2.4.2 Image analysis for early stage forecasting

#### 2.4.2.1 Shoots counting

Counting grapevine shoots early in the growing season is critical for adjusting management practices, but it is expensive and time consuming and moreover, these measures are often biased prone to error. An automated system has the potential to eliminate human error and reduce labour costs (Liu *et al*., 2017). Liu *et al*. (2017) investigated the ability to forecast yield using a low-cost vision system along with a novel shoot detection framework. It was investigated without a manual labelling to build a classifier, in certain range of illumination changes and different scenarios with various noise. It was done to evaluate the ability of the unsupervised feature selection and the optimum phenological stage for imaging shoots.

The common challenges for shoots detection in vineyards are varying lighting conditions, undesired objects in the field of view (e.g. posts, cordon, grass, animals, wire, reflections), change of shoot position in the field of view, shadows and barren cordon. To overcome these challenges Liu *et al*. (2017), used some approaches that divide images into sub-windows and process them all as object candidates (Chamelat *et al*., 2006). The common method for segmenting targets by image processing is to form a binary image using a threshold value. The problem with this image processing is that when collecting the images in movement the threshold value cannot be fixed since the illumination condition changes, but this can be overcome using the dynamic threshold method proposed by Otsu (1979), that improves the results. The model proposed by Otsu, however cannot guarantee good results for shoot segmentation when the image contains complicated objects (Liu *et al*., 2017).

The method, proposed by Liu *et al*. (2017) for shoots detection segments, consists in the use of potential shoots patches, by Gaussian fitting based on colour histograms for automatically locating the threshold value accurately. It then combines different scalar features into a feature vector to use as a descriptor of potential shoots.

Being the supervised learning approach very time-consuming Liu *et al*. (2017) used an unsupervised feature selection based on three correlation filters (Kendall *et al*., 1946; Higgins, 2004; He *et al*., 2006).

The shoots detection framework accuracy was calculated to be 86.83% with an F1 score (annex 1) of 0.90 across the four experimental blocks.

Liu *et al*. (2017) then introduced a procedure for converting shoots counts from videos to yield estimates using the equation 4 that shows the proposed relationship between shoot counts and estimated yield:

$$PY = NS \times PRV \times R_{BS} \times BW \times (1 - PR) \times H_e \times (1 - SP) \quad \text{(Eq. 4)}$$

Where $PY$ is the total predicted yield (mass of fruit without rachis assuming machine harvested); $NS$ is the number of shoots detected from videos; $PRV$ is the proportion of recorded VID (should videos of rows be missing or incomplete); $R_{BS}$ is the ratio of bunches to shoots from historical data; $BW$ is the average bunch weight at harvest in previous seasons; $PR$ is the proportion of rachis weight to bunch weight; $H_e$ is the harvester efficiency factor; $SP$ is the percentage of any destructively sampled fruit before harvest (Liu *et al*., 2017).

The results of the counting obtained by Liu *et al*. (2017) resulted most promising when videos were captured around E-L stage 9 (Coombe, 2004). At this stage, the absolute predicted yield estimation error of the system ranged from 1.18% to 36.02%, which can be considered very good results considering that the forecast regards a very early stage of the growing season (Liu *et al*., 2017).

#### 2.4.2.2 Flowers and inflorescence counting

Flower development, flowering and fruit set rate vary between cultivars, locations and seasons, therefore accurately assessing of these yield components is a key opportunity to determine the potential yield early in the season (Dry *et al*., 2010). The number of flowers per inflorescence show a strong variability among vines and within inflorescences of the same vine, a count of the flower number per inflorescence is essential for accurate assessment of fruit set (May, 2004 b). To do an evaluation of the number of flowers, an automated flower counting systems have recently been developed to process images of grapevine inflorescences and to asset the fruit-set ratios without arduous manual counting (Liu *et al*., 2018).

Diago *et al*., (2014) developed a new image analysis methodology to be applied to RGB images taken under field conditions to estimate the number of flowers per inflorescence automatically, which however require a uniform background colour and cannot be applied to the entire vineyard but only for localized samples.

The image processing method proposed and tested by Diago *et al*. (2014) involves three stages. The first step (image pre-processing) involves conversion of the image from RGB to CIE Lab colour space and a segmentation by means of a threshold, separating the inflorescence from the background. The second step (flower counting) consists in the detection and counting of the points

of higher light reflection: the flowers usually present a higher light reflection than other image area components. To find and identify the brighter points of the light of the image (L* coordinate) Diago *et al*. (2014) did a computation of the extended maxima transform (suppression of all local maxima lower than a threshold). Finally, the last step (image post-processing) is necessary to remove material other than flowers from the bright areas selected, this filtering operation is based on three criteria, in three consecutive steps: region size filtering, distance between brighter areas and shape of brighter areas (Diago *et al*.*,* 2014).

To validate the image analysis method Diago *et al*. (2014) determined the actual flower number per inflorescence manually by individually detaching the flowers from the rachis and moreover the estimation of the flower number on imaged inflorescences was also done manually on printed images. The determination coefficient ($R^2$) of the relationships between the number of flowers estimated either manual or automatically with the developed method, to actual flower number per inflorescence, was always higher than 0.80 ($\rho \leq 0.001$) (Diago *et al*., 2014).

The methodology proposed and validated by Diago *et al*. (2014) was used by Aquino *et al*. (2015) to implement artificial vision algorithms aimed at counting the number of flowers per inflorescence in the image implemented into the application vitisFlower® was tested on two devices. It gave results which indicated that more than 84% of flowers in the images were identified, producing less than 6% of detection errors in terms of average of Recall (RC) and Precision (PR) (annex 1) (Aquino *et al*., 2015).

A lower RC that was obtained by one of the devices used on Chenin Blanc may be explained by non-optimum acquisition conditions and potentially by the huge degree of compactness of the flower buttons of this cultivar at phenological stage of acquisition of the images. This fact led them to delineating the image acquisition settings that yielded the best application behaviour, these include: analysing inflorescences facing the sun; casting a shadow on the inflorescence to create a homogeneous scene; moreover, they recommended the use of the camera flash, if the illumination is poor due to low natural-light conditions (Aquino *et al*., 2015).

To evaluate the behaviour of variety-independent predictive models and yield prediction capabilities on a wide set of varieties Millan *et al*. (2017) worked using an evolved version, for the stability improvement, of the algorithm described in Diago *et al*. (2014) testing different models: single-variable linear model, multivariable linear model and non-linear model. They founded that the existent relationship between the flower number visible on images and the actual flower number per inflorescence is predominantly linear and with little influence of other variables associated with variability between cultivars (Millan *et al*, 2017).

The obtained results, about the image analysis algorithm for flower detection, were considered satisfactory and consistent among varieties, attending to the measured averages and standard

deviations. The RC values were all over 0.80 (except for Airen with RC=0.79) and in conjunction with PR values higher that 0.75 this validates the general applicability of the image analysis algorithm in flower detection (Millan *et al*., 2017).

As described before, different sensor-based methods were developed for flower quantification based on images of individual captured grapevine inflorescences (Diago *et al*., 2014; Aquino *et al*., 2015 a, b; Millan *et al*., 2017). These approaches require images of a single inflorescence in front of well distinguishable backgrounds which makes screenings of large plant sites more difficult and laborious (Rudolph *et al*., 2018). Therefore, another approach to detect inflorescences and flowers was studied and presented by Rudolph *et al*. (2018).

The work, of Rudolph *et al*. (2018) was focused on quantifying inflorescences and single flowers in unprepared field images of grapevines, where no artificial background or light was applied. They identified and localized inflorescence areas by partitioning the image into classes: 'inflorescence' and 'non-inflorescence' by assigning a class label to each individual pixel. The trained FCN used achieved a mean Intersection Over Union (IOU) (annex 1) of 87.6% on the test data set. Individual flowers were extracted from the areas representing the inflorescence using Circular Hough Transform. The flower extraction achieved a recall of 80.3% and a precision of 70.7% using the segmentation derived by the trained FCN model (Rudolph *et al*., 2018).

The study presented by Rudolph *et al*., (2018) was done to do an efficient screening of large sets of grapevines. This is important for studies regarding the development of early yield prediction models, for objective characterization and multi-year monitoring of breeding material.

### *2.4.3 Image analysis for yield estimation from fruit set to harvest*

Being that clusters per vine and berries per cluster account for 60% and 30% of variation in yield per vine respectively, an image analysis approach to estimate the 90% of the variation in yield was proposed by Nuske *et al.* (2011) doing an accurate count of the berries into the vineyard. Their approach consists in estimating the total number of berries, essentially combining clusters per-vine and berries per cluster in one measurement. The challenges in visually detecting grape berries is gave by their varying appearance under different lighting, the lack of colour contrast to the background, and occlusions causing not all grapes to be visible. Lack of colour contrast is an important issue that occurs in the white-grape varieties and all the grape varieties prior to veraison. They tried to solve these challenges, using shape and texture cues for detection (Nuske *et al*., 2011).

The issue of occlusion means it is not possible to detect and count all berries on a vine, but the berry count can be a reliable estimator of yield, even though the algorithm used only accounts for a percentage of all grape berries on a vine, given that the percentage of berries not detected is relatively constant from vine to vine (Nuske *et al*., 2011).

The berry detection was done using a sideways-facing camera and lighting on a small vineyard utility vehicle to capture the pictures which were processed with an algorithm that works following three distinct stages (Nuske *et al*., 2011): detection of the potential berry locations; identification of the potential locations that have similar appearance to grape berries, to classify the detected points which appear most like grapes; grouping of the neighbouring berries into cluster eliminating the false positive (Nuske *et al*., 2011).

Using the automatically berry counts with the actual harvest crop weights they obtained a linear relationship with a $R^2$ of 0.74. The average error obtained with their method was 9.8% of the actual harvest weight that already exceed what is possible obtain with the traditional practises used to estimate the yield (Nuske *et al*., 2011).

Nuske *et al*. (2014 a) presented an extension of their method reported by Nuske *et al*. (2011), demonstrating two different ways to calibrate the image berry measurements to harvest yield and collecting image data at various stages during the growing season.

In the work carried out by Nuske *et al.* (2011) it was obtained an estimate of the actual berry count using the number of berries founded by the visual detection algorithm with a yield forecasting function. After that, using the expected berry weight at harvest it was obtained the yield prediction.

Nuske *et al*. (2014 a) to better understand the individual causes of the bias, the single first order linear factor was divided, analysing three physical bias and two bias in the visual detection process: self-occlusion (berries hidden behind berries within the same grape cluster); cluster-occlusion (berries hidden behind other grape cluster); vine-occlusion (berries hidden behind the leaves and shoots of the vine); detection (false positive and false negative in the detection of the berries); mis-registration (errors, that can occur after several overlapping of the images, of double-counted or mistakenly not counted berries). Then all these bias terms were combined as linear factors.

The final image measurement model regarding the estimation of self-occlusion (assuming the visible berry count is proportional to the total berry count), the convex hull, shows a $R^2$ of 0.92 which is the best of the three image measurements studied by Nuske *et al*. (2014 a).

Moreover, Nuske *et al.* (2014 b) worked on automated visual yield estimation working during the night, using a camera mounted on a ground-vehicle, setting the camera and the illumination to optimize for low motion blur. Their approach is different from the other work that take in consideration only one cues as either colour (Diago *et al*., 2012; Dunn and Martin, 2004), shape (Rabatel and Guizard, 2007) or texture (Grossette *et al*., 2012) taking into consideration all these main visual cues that grape berries have. This because the colour on its own is not suitable for distinguishing green grapes on a background of leaves, the grape shape is difficult to identify in cluttered images with leaves and other spurious contours and the grape texture can be less distinguishing under certain illumination conditions.

Nuske *et al*. (2014 b) introduced an important innovation in their method presenting a modality to eliminate the double-counting problem from overlapping imagery and the challenge of geometrically referencing the measurements by estimating camera position along the row. Their approach obtained a $R^2$ from 0.6 to 0.73, depending on the dataset used, for the linear relationship between their automatically generated berry counts and actual harvest crop weights.

Aquino *et al*. (2017), similarly to Aquino *et al.* (2015), proposed a solution to predict the number of the berries present in the grape cluster by analysing a 2D image. The algorithm used for segmentation and counting of the berries is composed by two main stages: in a first step, the berry candidates are extracted by the detection of the bright spots generated by the light reflection finding the regional maxima of illumination. In the second step, false positives (FP) (annex 1) are eliminated from the counting by the algorithm, that uses a set of six morphological and statistical descriptors to train, test and compare a three-layer NN (Neural Network) and an optimised SVM (Support Vector Machine)

The results obtained by Aquino *et al*. (2017) show a $R^2$ of 0.83 between the berry number counted by manually labelling images and the actual berry number per cluster and a $R^2$ of 0.75 between the berry number automatically counted by the algorithm developed in their study and the actual berry number per cluster.

Aquino *et al*. (2018) presented a smartphone application, vitisBerry® that permit the assessment of berry number in clusters at phenological stages between fruit-set and cluster-closure, using an implementation of the image analysis algorithm presented by Aquino *et al.* (2017) that provide 1.63% and 7.67% of Recall and Precision improvement, respectively. Using vitisBerry and vitisFlower it is possible to monitor the selected clusters from pre-flowering to veraison (Aquino *et al.,* 2018).

In a recent work, carried out by Santos *et al*. (2019) on grape detection, segmentation and tracking using neural networks and three-dimensional association, demonstrated that grape cluster can be successfully detected. They obtained a F1-score up to 0.91 for instance segmentation, a fine separation of each cluster from other structures in the image that allows a more accurate assessment of fruit size and shape. They have also showed that 3-D models produced by structure-from-motion or SLAM (simultaneous localisation and mapping) can be employed to track fruits, avoiding double counts and increasing tolerance to errors in detection. Moreover, they said that is necessary looking for more sophisticated scene understanding systems able to robustly cope with occlusions and other sources of errors.

Proximal sensing using affordable cameras combined with computer vision have seen a promising alternative, strengthened after the advent of convolutional neural networks (CNNs) as an alternative for challenging pattern recognition problems in natural images. Off-the-shelf RGB cameras and computer vision can provide affordable and versatile solutions for fruit detection. Wine grapes

present large variations in shape, size, colour and structure, even for the same grape variety. In computer vision, classic machine learning and pattern recognition have been replaced by modern deep learning techniques, which are able to address the enormous variability in object appearance (Santos *et al*., 2019).

State-of-the-art computer vision systems based on deep convolutional neural networks (LeCun *et al*., 2015) can deal with variations in pose, shape, illumination and large inter-class variability (He *et al*., 2016; Krizhevsky *et al*., 2012; Simonyan and Zisserman, 2014) essential features needed for robust recognition of complex objects in outdoor environments. CNNs invariant to local translation give vision systems robustness in situations where a feature's presence is more important than its exact location (Goodfellow *et al*., 2016): Nuske *et al*. (2014 b) reported that variations in the berry candidate location by detection affect berry classification. While the CNNs can encode variance regarding pose, colour and illumination, if the training data presents enough examples of such variation, which relieves the need for controlled imaging, illumination and camera settings (Santos *et al*., 2019).

Looking at the presented studies, we know that we can detect and recognize, with good approximation, what we can see with our eyes in the pictures through computer vision, but now the problem that must be resolved is the estimation of what we can't see, which is what we are looking for in our research work.

### 2.4.4 Secondary traits – canopy features

Other parameters exist that are easily detected using computer vision and that might help to build models to estimate the yield.

#### *2.4.4.1 Canopy Porosity*

Canopy porosity is a relevant canopy feature because it indicates the level of fruit exposure and air circulation, which are important factors for fruit quality and health. The ideal grapevine canopy may have between 10 to 20 % (Palliotti and Silvestroni, 2004) or 20 to 40% gaps (Smart, 1987) which guarantees an adequate sunlight capture and reduces shading. Point Quadrat Analysis (PQA) is the standard method to evaluate canopy porosity, but it is laborious and time consuming. Therefore, a new, objective, non-invasive, image-based method was developed and compared with PQA, in different field conditions (Diago *et al.,* 2016). The technique of PQA involves the insertion of a probe through the canopy of grapevines and counting the number and parts of the vine that contacts the probe. The proportion of gaps is quantified by dividing the number of gaps by the total number of insertions. Minimum 50 passes through the canopy is recommended to accurately quantify the gaps (Smart 1987).

The image-based method to evaluate the canopy porosity was implemented using a clustering algorithm based on the Mahalanobis distance described by Diago *et al*. (2012) to identify the pixels

corresponding to the canopy porosity. The Mahalanobis distance measures the similarity between an unknown and a known sample group considering the correlations the different variances of a data set for each direction (Diago *et al.,* 2016). The $R^2$ of the regressions between the percent of gaps, using both methods, was 0.90, and $R^2$ of the global regression was 0.93. Considering the good results obtained and the easy-to-use implementation of this image-based method, new technologies as machine vision, can be used as decision support tools (Diago *et al.,* 2016).

#### 2.4.4.2 Bunches occlusion

As suggested by Nuske *et al*. (2014 a) and Lopes *et al*. (2017) an important parameter, that can improve the yield estimation, is the study and the prevision of the bunch by bunch occlusion (berries hidden behind other grape cluster) to avoid the bias in the estimation of the total area of the bunches present on the vine.

#### 2.4.4.3 Leaf area

Surface area assessment is normally carried out when the canopy is fully grown. This is usually done between veraison and harvest. It can be calculated for continuous canopies whit equation 5 (Smart and Robinson,1991):

$$ELA\left(\frac{m^2}{ha}\right) = (2 \times H + D) \times n^{\circ}\ of\ canopy\ meters/ha \qquad \text{(Eq. 5)}$$

del-Moral-Martinez *et al*. (2016) studied the application of a mobile terrestrial laser scanner (MTLS) to map the vineyard by estimating its leaf area. The leaf area index (LAI) is defined as the one-side leaf area per unit ground area and is probably the most widely used index to characterize grapevine vigour.

They used the equation 6 to calculate LAI:

$$LAI = \frac{(S_L + S_R + S_T) \times C}{d_r \times L} \qquad \text{(Eq. 6)}$$

where $S_L(m^2)$ and $S_R$ ($m^2$) were the left and right leaf wall areas, respectively, $S_T$ ($m^2$) was the top area, C (dimensionless) was the ratio of the foliar surface to the enveloping area, $d_r$ was the row spacing (m), and L (m) was the section length.

So del-Moral-Martinez *et al.* (2016) were also able, using the MTLS, to calculating the enveloping vegetative area of the canopy (also known as exposed leaf area), which is the sum of the leaf wall areas for both sides of the row (excluding gaps) and the projected upper area (that was computed considering the distance between the points where the right and left leaves wall area were georeferenced).

### 2.5 Vinbot robot platform

Vinbot is an autonomous cloud-computing vineyard robot to optimize yield management and wine quality. It is an EU project (Powerful precision viticulture tool to break traditional yield estimation in vineyards) funded under the FP7 SME program. Vinbot is formed by a consortium of developers and end-users (Guzman *et al*., 2016 a).

Vinbot is an all-terrain autonomous mobile robot with a set of sensors (is possible set all the sensors that we need) capable of capturing and analysing vineyard images and 3D data by means of cloud computing applications. Vinbot responds to a need to boost the quality of European wines by implementing precision viticulture, to estimate the yield (amount of fruit per square meter of the vine area). In addition, Vinbot estimates the amount of leaves, grapes and other data in the vine throughout the entire vineyard via computer vision and other sensors and generates online yield and vigour maps to help wine growers to optimize their management strategies (Guzman *et al.*, 2016 a).

An important part of the research and development effort was related with the selection of the solutions able to do the task without exceeding the budget, in summary to find a compromise between price, reliability and simplicity (easy deployment) of the solution. The implementation of the robotized solution is done thanks to The Robot Operating System (ROS) which is a set of software libraries and tools that help you build robot applications (Guzman *et al.*, 2016 a).

Regarding machine vision procedures, the approach “Convolutional Neural Networks inside of Deep Learning Field” was used for image segmentation and grape recognition in the previous studies carried out by Lopes *et al*. (2017). Now the new objective of our research group is to develop and train the FCN (Fully Convolutional Networks) to be used to segment the vineyard images (Duarte, 2019) and then use new algorithms to estimate the leaves and bunch by bunch (BxB) occlusion.

## 3 Material and methods

### 3.1 Localization and characterisation of the vineyard

The experiment was carried out during the vintage 2019 in the experimental vineyard called “vinha da Meia Encosta” located in the Instituto superior de Agronomia (ISA) (Portugal, Lisbon, Tapada da Ajuda: 38°42’24” N 9°11’05” W) which is shown in the figure 3.


*Figure 3) Picture captured from Google maps, vineyard of white-vines of ISA, located in Lisbon, Portugal. The blue part highlights the location of Arinto spot.*

This vineyard was planted in 2006 with the following varieties: Alvarinho, Arinto, Moscatel de Setubal and Viosinho grafted on rootstock 1103 Paulsen (*Vitis berlandieri* X *Vitis rupestris*) and Encruzado, Macabeu and Moscatel Galego grafted on rootstock 110 Richter (*Vitis berlandieri* X *Vitis rupestris*). It has an area of 1.46 ha, and the distance between the plants is 1 m and the distance between rows is 2.5 m determining a plant density of 4000 plants/ha. The rows are oriented Azimut 21° and exposed to Est caused by a slight slope (maximum of slope is 9%). This research thesis is focus on the *Vitis vinifera* L., variety Arinto which is planted from the 50$^{th}$ row to the 68$^{th}$ rows in an area of 0.38 ha. The vines are trained to a vertical shoot positioning and spur pruned on a unilateral Royat cordon.

The soil of the vineyard is a clay loam with 1.6% organic matter and a pH of 7.8.-7 (Teixeira *et al.,* 2018). It is described as a reddish-brown clay, not basaltic limestone. It presents a profile of type Ap (B) C characterized by a high content in colloids of montmorillonite, which gives it high plasticity when wet and hard when it’s dry; it can be cracked when the moisture content is very low. The expandability and the field capacity values are high with a high usable capacity in the first 50 cm. Its permeability is rapid to moderate (Sarmento, 1969). The vineyard is drip irrigated and standard cultural practices in the ISA vineyards were applied to all the Arinto plot.

### 3.2 Climatic characterisation of Tapada da Ajuda

The climate in Lisbon is classified *Csa* (C: warm temperature; s:summer dry; a:hot summer) as established by the classification of Köppen and Geiger (2006), with higher precipitation in the winter than in the summer, as shown in figure 4. The annual mean temperature is 15.4 °C and the annual mean precipitation (1973 to 2000) is 725.8 mm (IPMA, 2019).

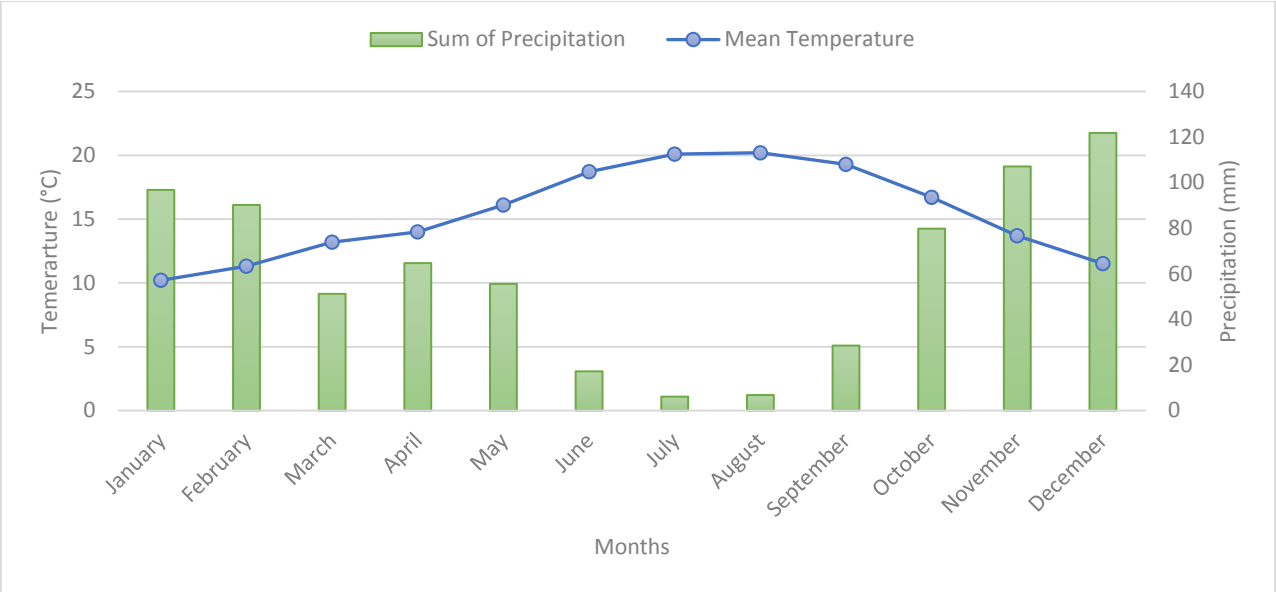



*Figure 4). Climatic data from years 1973 to 2000 for each month; the mean temperature and the sum of precipitations registered in Lisbon.*

It is possible to observe an increment of the mean temperature from 15.4 °C (considering the years from 1973 to 2000) to 16.5 °C (considering the years from 1973 to 2019) (NOAA, 2019).

From the figure 5 it is possible observe an increase of the mean temperatures in August and in the other months until December. On the other hand, during the other months until july the mean temperatures are lower or equal to the general mean. Moreover, during 2019 is possible observe how the mean temperature is oscillatory and dissimilar from the regular growing of the temperature as it happens in the 2018.

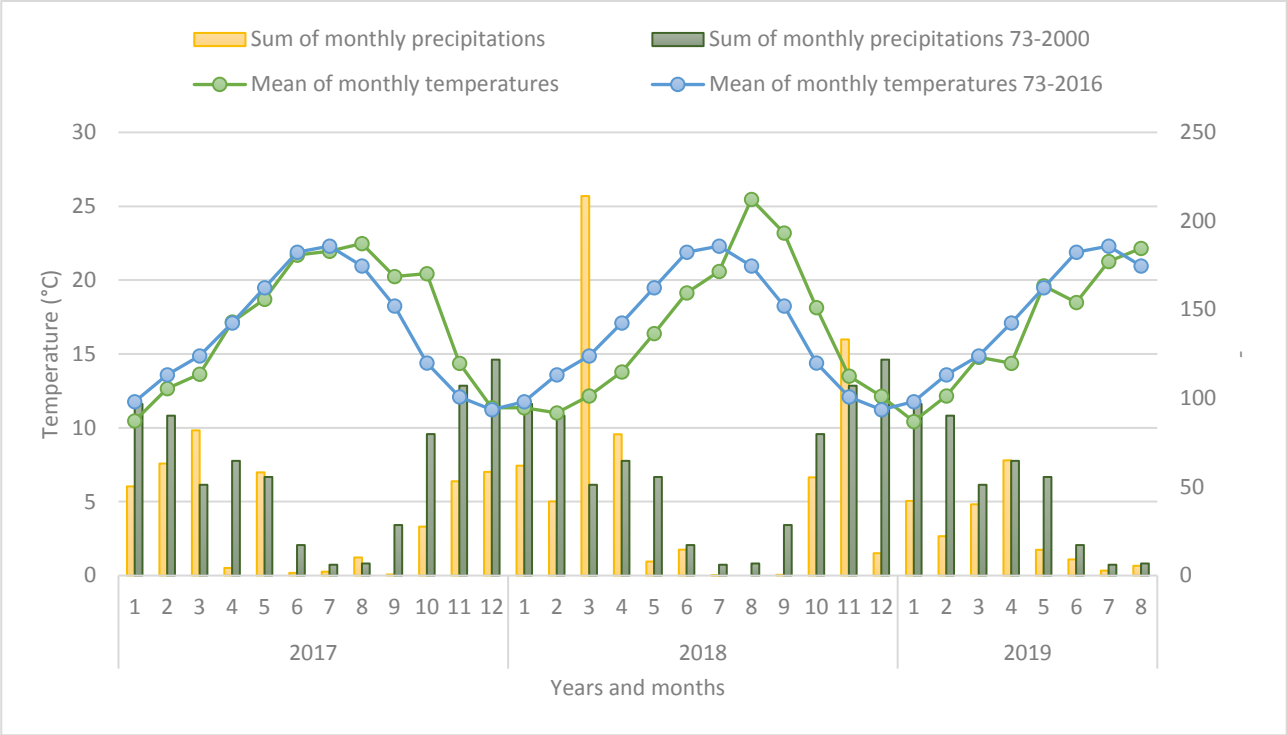



*Figure 5). Mean monthly temperatures and of sum monthly of precipitations of the last 3 growing circles (source:* IPMA*) compared to the mean of the monthly temperatures from 1973 to 2016 (NOAA, 2019) and with the sum of monthly precipitations from 1973 to 2000 (IPMA).*

### 3.3 Studied variety

*Vitis vinifera* L., variety Arinto has the synonymy of Pedernã in the delimited region of Vinhos Verdes. Arinto is a variety with low fruitfulness but very large bunches, features that allowed compensate for the lower bunch number. Ampelographic characteristics (Eiras-Dias *et al*., 2011):

- The budbreak is late
- The shoot, shown in figure 6A, has its extremity with a form completely open and a high density of creeping hairs. It presents: an erect habit, a medium vigour and the fertility of basal buds and a weak intensity of anthocyanin coloration of the buds.
- The cluster shown in figure 6B is long, with conical shape and presents 1 or 2 wings; it has a medium level of compactness and a high weight. The berry has a rounded shape and medium size. The pulp is colourless and little hard.
- The young leaf presents the upper edge in bronze colour. The adult leaf shown in figure 6C presents a big and irregular profile; the margin is characterized by convex teeth of medium length and a medium level of swelling. The petiolar sinus presents overlapping lobes in brace form. The underside edge is characterized by a medium-high density of hair.

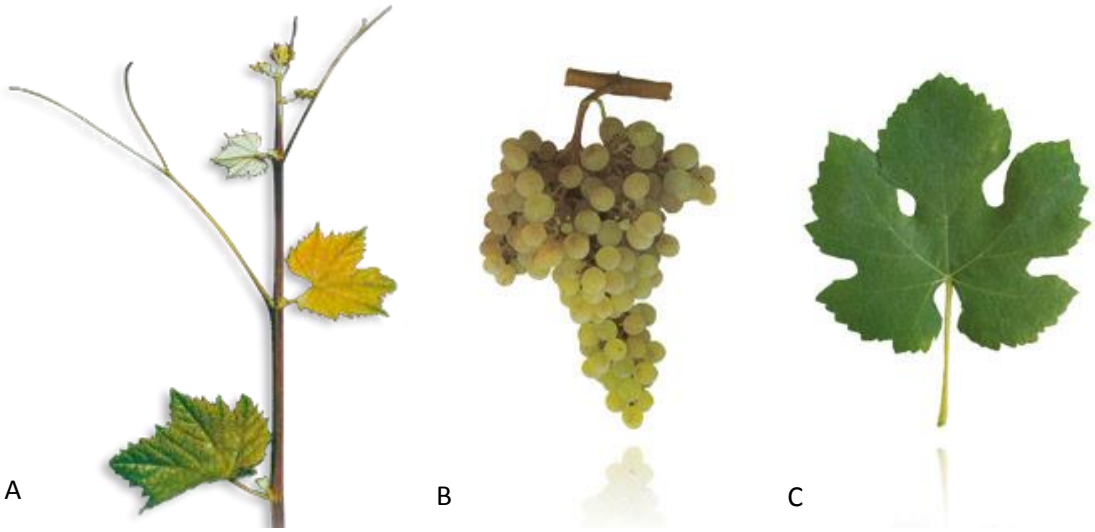



*Figure 6) Picture of sprout (A) (Antunes et al., 2011), bunch (B) and leaf (C) of Arinto (Wines of Portugal).*

### 3.4 Vinbot

The Vinbot robot platform is based on a Summit XL HL mobile base able to carry up to 65 kg payload and consists of (Fig. 7) (Lopes *et al*., 2017):

- A robotic platform: durable, mobile, with ROS Indigo and Ubuntu 14.04;
- RGBD Kinect v2 camera to take images of the vine;
- 2D range finders to navigate the field and to obtain the shape of the canopies;
- A small computer for basic computational functions, connected to a communication module that use Bluetooth and Wi-fi;
- An optional RTK-DGPS high accuracy rover, optional base and associated communication devices;
- A cloud-based web application to process images and create 3D maps;
- User friendly HMI to define navigation and data acquisition missions.

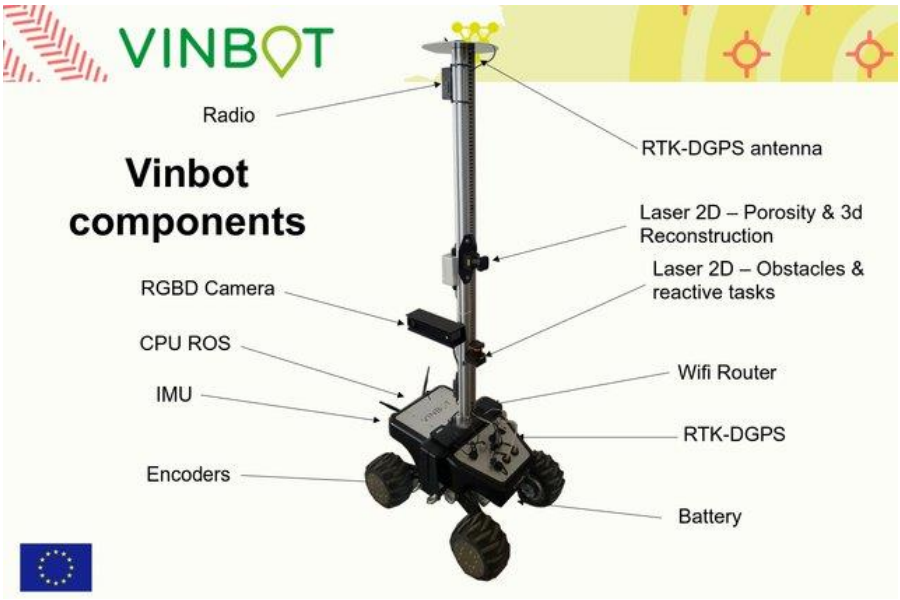



*Figure 7) VinBot component sensor and component distribution (Guzman* et al*., 2016 b).*

The RGBD Kinect v2 has two cameras, the RGB camera collects images with a resolution of 1920x1080 pixels, whereas the IR camera is used for the real-time acquisition of depth-maps and IR data with a 512x424 pixels resolution. The whole acquisitions can be carried out with a framerate up to 30 Hz. A last feature to be mentioned is the field of view for depth sensing of 70 degrees horizontally and 60 degrees vertically.

This unmanned platform is capable of autonomously navigating over rough terrain and can climb up to 45º with a battery capacity of up to 8 hours to photograph and digitize the rows of vines and although on this date collect only data with a LiDAR (Laser Rangefinder) system and an RGB-D camera, the mobile platform can incorporate numerous other sensors and technologies. (Vinbot).

### 3.5 Experimental Design

In the field we selected 4 smart-points (SPs) in different rows to obtain the most possible variability (rows n° 52-56-60-65). For each smart-point we took in consideration 25 linear canopy meters (LCM), divided as shown in figure 8: 10 meters (highlighted in red colour) were dedicated to the image collection with Vinbot throughout all the growing season; 15 LCM where dedicate to the image collection with Nikon RGB camera for three phenological stages: 5 LCM (highlighted in blue) for pea size, 5 LCM (highlighted in green) for veraison and 5 LCM (highlighted in yellow) for full maturation. The LCM dedicated to Vinbot sessions were set with a meter scale (divided in step of 10 cm) fixed on the vines, for all the season to have every time the same vine reference in the images. Regarding the others 15 LCM we used a transportable scale that we fixed on the vines, because it was necessary only one time for each 5 LCM, in the corresponding smart-point and phenological stage.

*Figure 8) Rows (n° 52-56-60-65) chosen for the location of the 4 smart-points, and the relative 10 and 5 meters where we collected the data: (red) 10m for the Vinbot sessions; (blue) 5m for pea size; (green) 5m for veraison; (yellow) 5m for full maturation.*

### 3.6. Vinbot sessions for algorithms testing

Since the Vinbot's navigation system is still under development (it will use odometry system for the self-navigation), we used a wireless joystick system to pilot the Vinbot and using a smartphone, to communicate with Vinbot through Wi-fi, we managed the collection of the images (one image per LCM). The image collection was done at the phenological stage of 4 leaves out, flowering, pre-bloom, pea-size, veraison and full maturation; it was done in two steps: with and without a blue background on the same linear canopy meters and only on one side of the row. Therefore, we collected 20 images for smart-point (4 SPs), during 6 phenological stages with a total of 480 images (divided in images with and without background). Beside the image collection was also done the ground truth, observed (Obs) yield components (shoots, inflorescences and clusters) for each meter.

We calculated also the exposed leaf area (using the method presented in the section 2.4.4.3) for each meter at the stage of veraison and full maturation. We measured the amplitude of the fruit zone for every meter and checking the visibility of the Vinbot camera, using the smart-phone application, we tried to find the best distance to conciliate the closer distance from the canopy with the visibility of scale and all full fruit zone. The data of exposed leaf area are not reported in this thesis because they resulted not useful to develop the algorithms for yield estimation. The data of fruit zone

amplitude were registered because it is an important parameter that permits to better manage the position of the camera along the vertical support of the Vinbot.

### 3.7 Nikon sessions for algorithms building

We collected images with a commercial camera Nikon D5200 (Manualslib, 2019a) during three phenological stages (pea size, veraison and maturation) in the 4 SPs using blue background. The sessions of image collection were set in two steps: the first one was done to evaluate the canopy porosity and the occlusion in the visual detection of the bunches area generated by leaf occlusion at different levels of defoliation; the second one was done to evaluate the percentage of BxB occlusion. During the first step 15 images were collected for each SP and for each phenological stage, obtaining a total number of 180 images. The pictures were collected on three levels of defoliation:

- on non-disturbed canopy vines (Fig. 9A)
- on canopy vines partially defoliated at different levels of defoliation (Fig. 9B):
    - SP1 low level of defoliation;
    - SP2 medium level of defoliation;
    - SP3 high level of defoliation;
    - SP4 different levels of defoliation for each of the 5 meter: low, medium, high, medium, low.
- on canopy vines completely defoliated (Fig. 9C).


*Figure 9). Image (A) plant without defoliation; image (B) plant with high level of defoliation; image (C) plant totally defoliated with all the bunches.*

During the second step the number of images collected was variable between every linear canopy meter, because it is dependent on the number of layers generated by the superposition of bunches. After the images collection done in the first step, to be able to evaluate the BxB occlusion, it was necessary to take out from the canopy vine the first layer of bunches (taking the decision looking to the image with all the layers on the vine). This was done for every layer removal as shown into the figure 10 A-C.


*Figure 10) (A) Canopy vine without the first bunches layer; (B) canopy vine without the first and second bunches layer; (C) canopy vine without bunches.*

The bunches took out from the vine were collected in plastic bags and labelled with a code (# of SP; # of meter; # of layer). We chose one or two meters per SP at every phenological stage (depending on the number of bunches present), from where the bunches took out from the canopy vine were labelled singularly with a code (# of SP; # of meter; # of layer; # of bunch) to then use them for the detail measurements into the lab.

As for the lienear meters of canopy vines dedicated to he algorithms testing the exposed leaf area was calculated.

### 3.8 Detailed measurements

The bunches, clustered per smart point, meter and layer, were counted and weighted together. The other bunches coming from the chosen LCM and individually labelled were inserted with their own label in aluminium containers and using a computer for data input they were assessed in detail by the following proceedings:

1. two images per bunch were collected with blue background using the same Nikon camera used into the field. To maintain the same distance from the bunch, the camera was mounted on a tripod and moved, to take two pictures, between position referment present on the floor. The bunches were hung on a bar with the use of a metal spring paper clip tied to the bar in a precise point designated with two black marks made with a permanent marker.
2. cluster weight was measured using a digital table scale (KERN FCB version 1.4).
3. cluster volume was assessed using the water displacement method: a volumetric cylinder (NORMAX 1000:10 ±10) was filled until a certain level with water and after the cluster insertion the final water level was recorded. The subtraction of the initial volume of water to the final volume obtained after the bunch insertion gives the corresponding volume of the bunch.

4. rachis and wings (if present) length were assessed after detachment of the berries, using a ruler meter, scaled in millimetres;
5. berries weight was measured;
6. calculation of berries number: the berries for every bunch were distributed in casual order on the table making sure that they were spaced from each other, to avoid contacts that would have given problems in the subsequent phase of image analysis. The berries, well distributed on the table with the presence of the corresponding bunch label, were photographed using automatic exposure control and flash with the SONY digital camera model DSC-H90 (manualslib, 2019b). The camera was fixed on a support (goose neck) to maintain the same distance from the table to guaranty the possibility to use of the same scale in the image analysis.

### 3.9 Image analysis

The image analysis was carried out using “ImageJ”, a public domain Java image processing program available for Windows, Mac OS, Mac OS X and Linux. It can: display, edit, analyse, process, save and print 8-bit, 16-bit and 32-bit images; read many image formats including TIFF, GIF, JPEG, BMP, DICOM, FITS and "raw"; calculate area and pixel value statistics of user-defined selections; can measure distances and angles; create density histograms and line profile plots. It supports standard image processing functions such as contrast manipulation, sharpening, smoothing, edge detection and median filtering. ImageJ was designed with an open architecture that provides extensibility via Java plugins (Rasband).

#### 3.9.1 Field image analysis

As explained above we took images in the field with Nikon camera and with Vinbot. From the images collected with Nikon camera different parameters were extracted: image area, background area and area of visible bunches. The first step that must be done for every image is the setting of the scale to convert the image pixels in centimetres using the meter scale fixed on the photographed plants, (Fig.11A). To avoid the prospective problems the scale must be set using the full meter present in the picture. Then the original image was then cut to select only the canopy, (Fig. 11B).

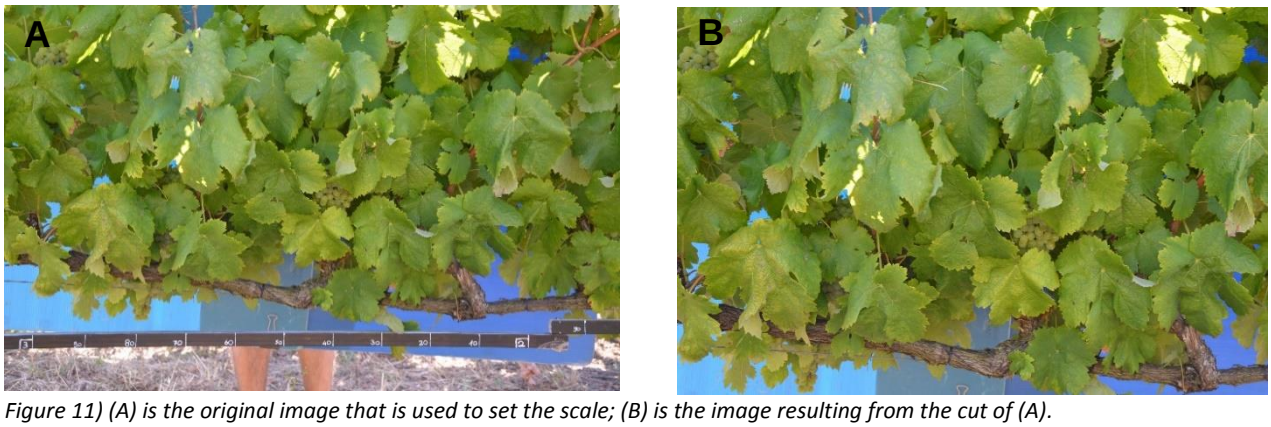



*Figure 11) (A) is the original image that is used to set the scale; (B) is the image resulting from the cut of (A).*

On this new image it was firstly calculated the total area and then the area of the porosity that corresponds at the blue background. To obtain the selection of the background was necessary converting the image from RGB colour space to CIE Lab colour space with the function of ImageJ "Lab stack" that gives back the converted image and divided in three images which represent the 3 channels L*, a, and b, as shown in the figure 12A-C, respectively.

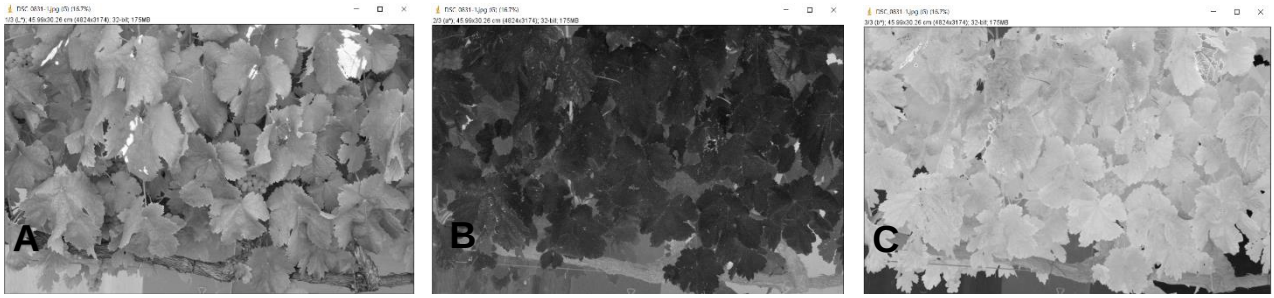
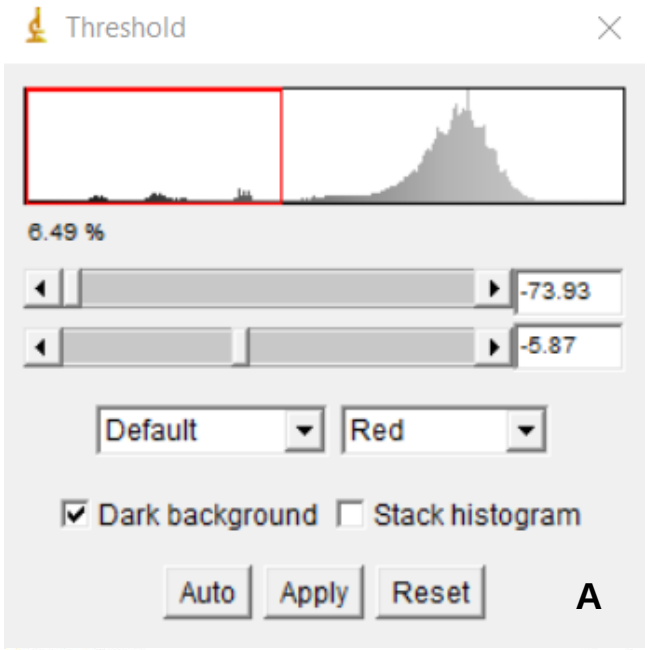



*Figure 12) (A) is the image representing the channel L*; (B) is the image representing the channel a; (C) is the image representing the channel b.*




Then, by adjusting the threshold (Fig. 13A) of the channel b image was possible to highlight the ROI, (Fig. 15B) to then create its selection (Fig. 15C) and measuring it.

The projected bunches area was calculated measuring the handmade selection of the visible bunches on all taken images.

As presented above in the section (3.7) every linear canopy meter was photographed at variable levels of defoliation obtaining different levels of canopy porosity and bunches visibility.

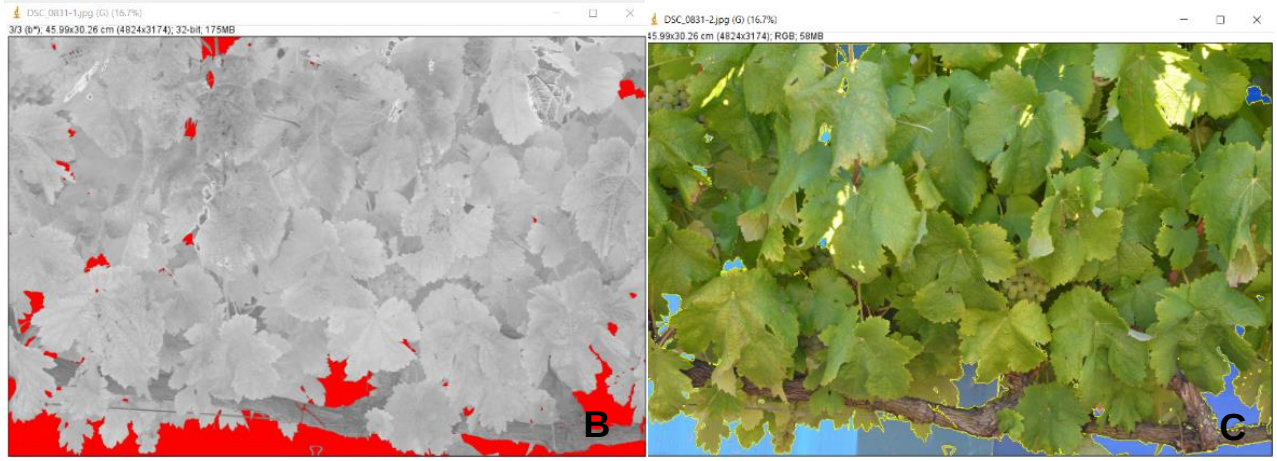



*Figure 13) (A) Controller window to manage the threshold to highlight the background pixels; (B) channel b image with highlighted background; (C) selection of the highlighted background.*

The percentage of canopy porosity was calculated for all levels of defoliation with the equation 7:

$$\%\ porosity\ of\ the\ canopy = \frac{background\ area}{total\ image\ area} \times 100 \quad \text{(Eq. 7)}$$

The projected bunches area was measured on non-disturbed, partially defoliated and full defoliated (without leaf occlusion) canopy for every LCM. It was used to calculate its corresponding percentage of visible bunches (VB) on the corresponding area of bunches without leaf occlusion, at every level of defoliation, using the equation 8:

$$\%\ VB = \frac{projected\ area\ of\ visible\ bunches}{projected\ area\ of\ bunches\ without\ leaves\ occlusion} \times 100 \quad \text{(Eq. 8)}$$

On the images, collected to evaluate the bunch by bunch occlusion, the fruiting areas (all visible bunches in the picture) composed by all layers together and the areas of the bunches layers in first plane were projected. These areas were used to calculate the bunch by bunch occlusion ($B \times B$), applying the *Set Theory (Al-Amin 2017)*. From the image with all the layers we can get the union area ($A \cup B$) (Fig. 14A) that represents the union between the bunches area of the first layer ($A$) (Fig. 14B) and the area of the bunches in background ($B$) (Fig. 14C); ($A$) was get from the same image from where was get ($A \cup B$). Instead ($B$) was get from the image took after the removal on the first layer, so ($B$) represents also ($A \cup B$) but of the second set. Then to calculate the percentage of bunch occlusion, it was necessary the calculation of the intersection between the two sets ($A \cap B$) that can be calculated with equation 9:

$$A \cap B = A + B - A \cup B \quad \text{(Eq. 9)}$$

The intersection ($A \cap B$) represents the area of ($B$) covered by ($A$), to obtain the percentage of bunch by bunch occlusion, we have just to use the equation 10:

$$\%\ B \times B = \frac{A \cap B}{A \cup B} \times 100 \quad \text{(Eq. 10)}$$

To obtain the total percentage of bunch by bunch occlusion we must sum the percentages of occlusion caused by each layer.


*Figure 14) (A) Selected area of bunches representing the $A \cup B$; (B) area of the layer that was took out from the plant representing the set $A$; (C) area of bunches without the occlusion caused by the layer took out representing $B$.*

The images collected with Vinbot were analysed doing a visual detection (counting) of the yield components simulating the automatic visual detection of the future algorithms that will be used, assuming that the algorithms can have the same level of accuracy as the human eye. Moreover, for the phenological stages of pea-size, veraison and full-maturation the images were also analysed in the same way of the images took with Nikon camera for the models building, for calculating the area of visible bunches and the percentage of canopy porosity on the non-disturbed canopies.

### 3.9.2 Image analysis of bunches and berries

The image analysis carried out on the images collected in the lab, has the objective of measure the area of the bunches and the number of the berries. The area of bunches was measured on the images took with Nikon camera. The images were scaled using the bunch label and the metal spring paper clip in the two images of the same bunch (Fig. 15A). The images were then cut by maintaining only the bunch and the blue background in the new image (Fig. 15B). It was done in order to convert the images from RGB colour space to CIE Lab colour space to be able to manage the colour threshold to highlight the bunch area (Fig. 15E) dividing the bunch pixels from the background (Fig. 15C). From the highlighted area of the bunch was created its selection (Fig. 15D) and it was measured. This procedure was done for both side of the same bunch to then be able to calculate the mean area of the bunch.

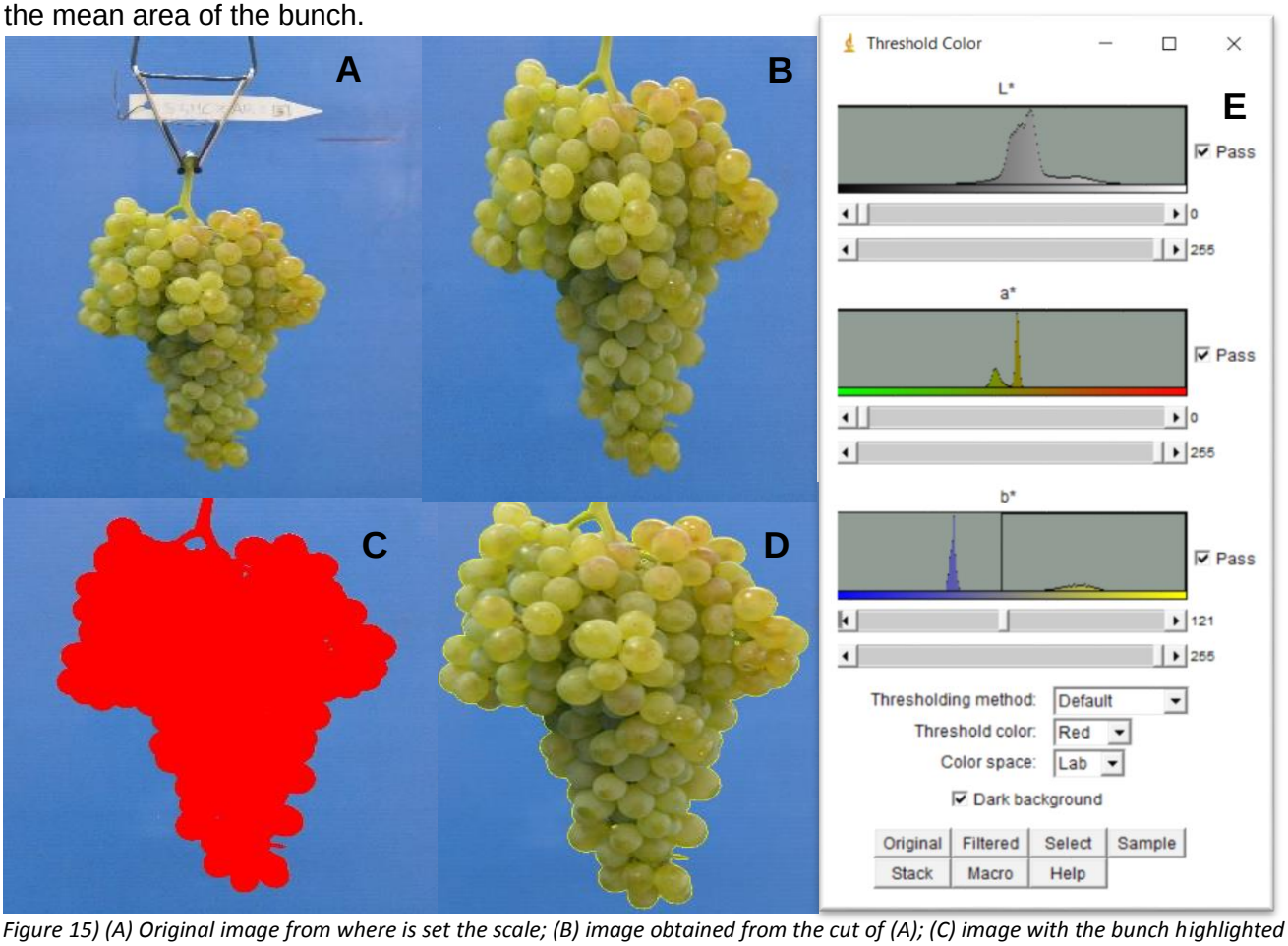



*Figure 15) (A) Original image from where is set the scale; (B) image obtained from the cut of (A); (C) image with the bunch highlighted using colour threshold showed in (E); (C) image with the bunch area selected.*

The berries images were scaled using the label of the corresponding bunch (Fig. 16A). The images were cut to eliminate noise sources for image analysis maintaining only the berries (Fig 16B).

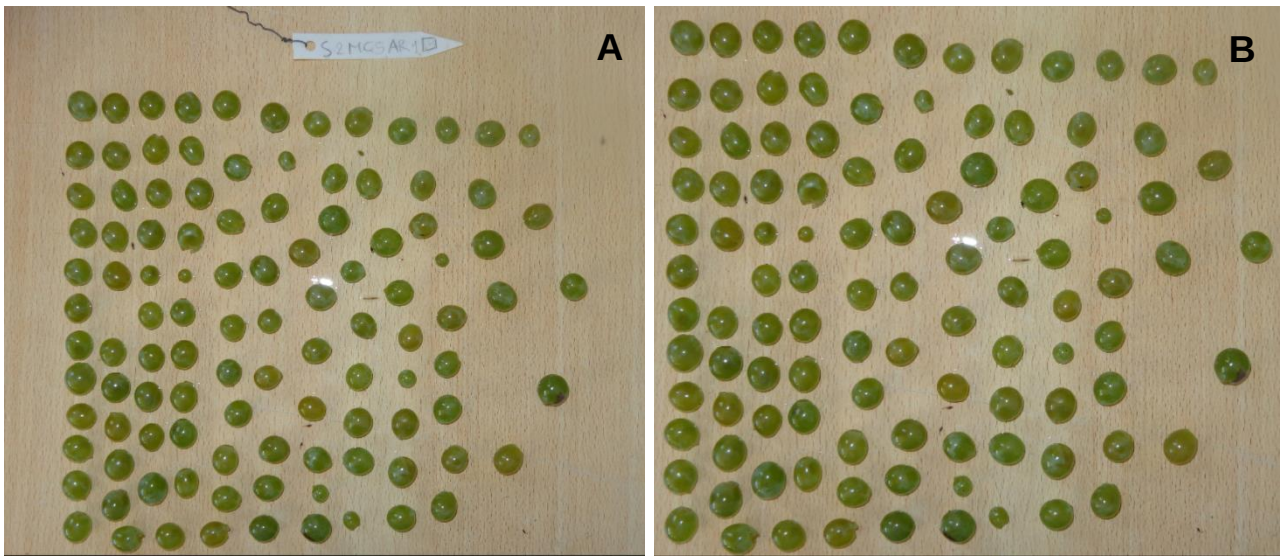



*Figure 16) (A) Original image; (B) is the image that come from the cut of the previous image.*

The cut images were converted from RGB colour space to HSB colour space that gives back three images which represent the channels Hue (Fig 17A), Saturation (Fig. 17B) and Brightness, (Fig. 17C).

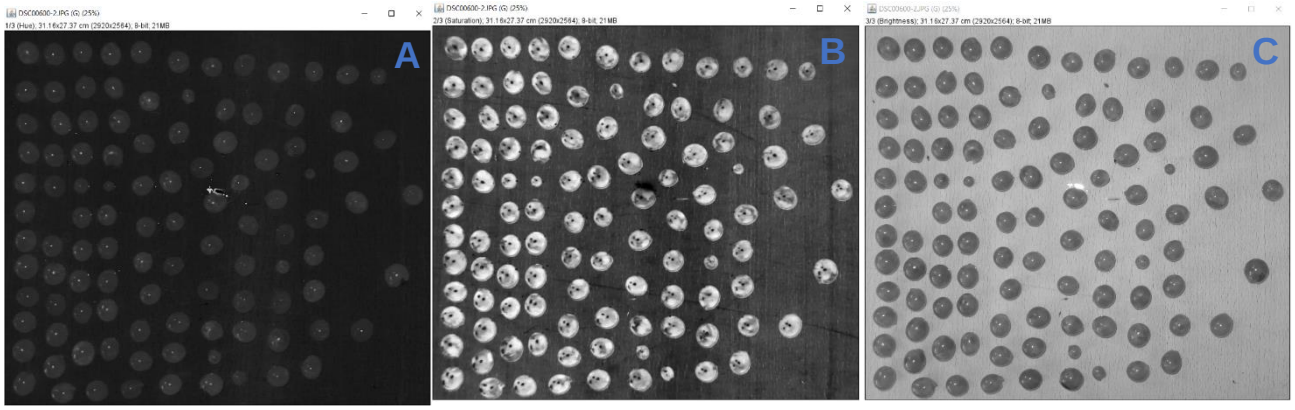



*Figure 17). (A) Image of Hue channel; (B) image of Saturation channel; (C)image of Brightness channel.*

It allows a better selection of the berries which is done by managing the threshold (Fig. 18A) of the Brightness channel. In the background it was present some noise represented by drawn plot of the table where the berries were positioned (Fig. 18B). This noise was eliminated by applying the "smooth" function (with this filter, the data points of a signal are modified so that individual points that are "higher" than the immediately adjacent points, presumably because of noise, are reduced and points that are "lower" than the adjacent points are increased (O'Haver T., 2019)) (Fig. 19A). After having highlighted the berries it was necessary set the measures wanted and after it, using the ImageJ function "Analyse particles" (Fig. 19B), the berries are counted, and their area and perimeter assessed. To avoid problems in the measures is was necessary set the general size (giving to the program the minimum and the maximum of the berries size range) of the particles that must be

analysed. Finally, from the output results of the program, the mean area and perimeter of the berries for every bunch was calculated.

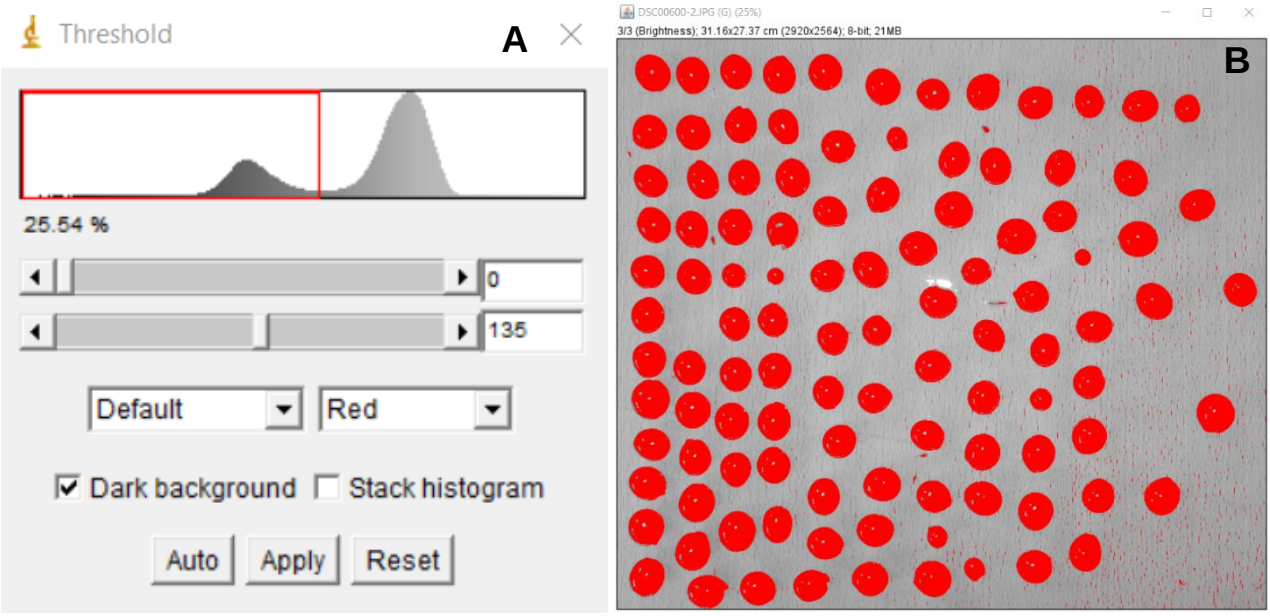



*Figure 18) (A) window of control of threshold used to select the berries; (B) image with highlighted berries with noise present.*

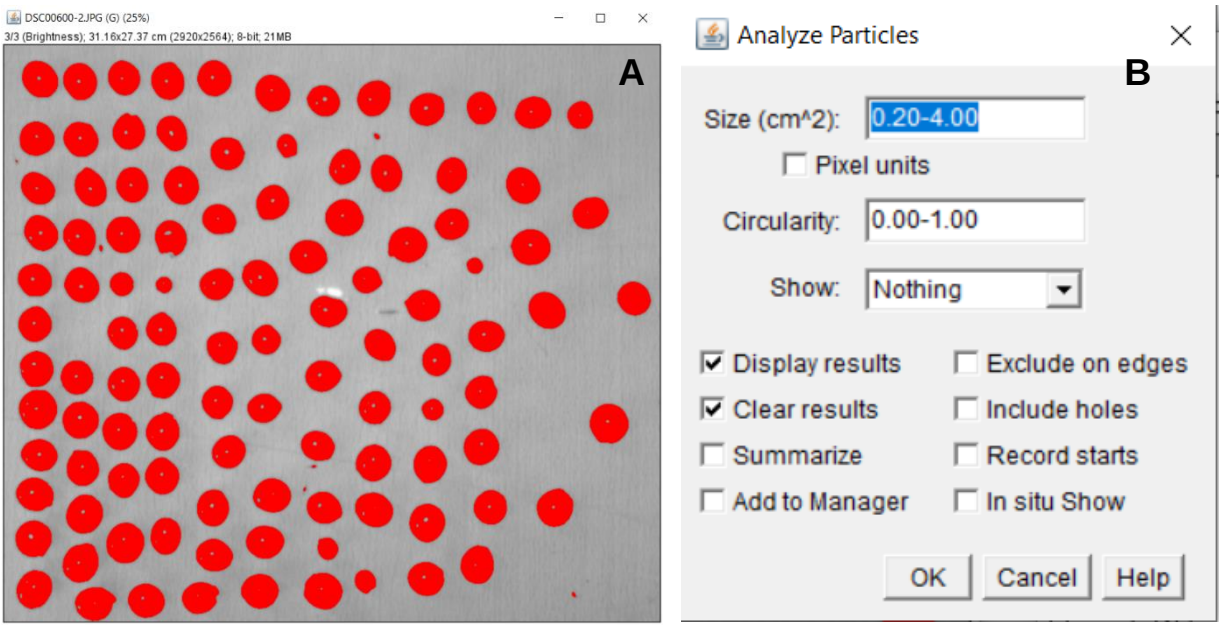



*Figure 19) (A) Image of the highlighted berries after the elimination of the noise caused from the brightness of the table; (B) window used to set the characteristics of the particles to analyse.*

### 3.10 Models and structure of the algorithms used for yield estimation

The models used to build the algorithms to estimate the yield are different. The algorithm previously used by Lopes *et al*. (2017), here denominated algorithm A, concerns the use of two models. The first one, which has the aim to estimate the bunches area without leaf occlusion, is built using the relationship between porosity and the percentage of visible bunches. The second one was built using the relationship between the weight and the area of bunches (measurements done into the lab for every single bunch collected) with the aim to convert the estimated area of the bunches into the corresponding weight. To replace the models used by Lopes *et al.* (2017), other models were developed and used in the algorithm B. The first model, that has the aim to estimate the bunches area without leaf occlusion, was built using the projected visible bunches area (from non-disturbed canopy) and the percentage of projected visible bunches area calculated on the total projected bunches area per linear canopy meter without leaf occlusion. The second model, that has the aim to estimate the weight of the bunches, was built using the total bunches area per linear canopy meter, which was obtained by adding to the total projected bunches area without leaf occlusion, the calculated % of BxB occlusion per linear canopy meter, and the total bunches weight per linear canopy meter.

#### 3.10.1 Models to estimate the occluded bunches area by leaves

The first model implemented into the algorithm A, to estimate the leaves' occlusion takes into consideration the percentage of porosity present into the fruiting zone of the canopy (parameter that can be easily calculated by the Vinbot using Lidar technology) to estimate the percentage of the visible bunches area to then estimate the bunches area without leaves' occlusion. This because, when there is more porosity on the vine could be possible to see a higher percentage of bunches, the fraction of bunches occluded by leaves is dependent of canopy porosity (fraction of gaps in the fruiting zone; Smart and Robinson, 1991). In order to use the model of the porosity also on vines that can be submit to defoliation, (which is a current cultural practice) it was built using all the levels of porosity present on non-disturbed, partially defoliated and full defoliated canopies.

The second model, implemented into the algorithm B, to estimate the leaves' occlusion, is based on the idea that the amount of visible bunches' area we can see is dependent on the factors that determine the final yield but also on the varietal characteristics and on the training system which determine the percentage of visible bunches' area respect to their total visible area without leaves' occlusion. The probable most important varietal characteristics that could determine the amount of visibility of the bunches is represented by the morphology and the average size of the leaves; another varietal parameter could be the distribution of the bunches along the shoots. The training system, determining the canopy architecture, can change the ratio between the visible bunches area of non-disturbed canopy and the visible bunches on full defoliated canopy.

### 3.10.2 Estimation of bunch by bunch occlusion

Bunch by bunch occlusion is an important parameter which was investigated with the aim to estimate the occluded part of the total bunches' area by other bunches, to obtain a better estimation of the total area of the bunches per linear canopy meter and therefore a better estimation of the final yield. For the reasons explained in the section 4.2.1 we used the average percentage of bunch by bunch occlusion calculated for every phenological stage as presented into the section 3.9.1.

### 3.10.3 Models to estimate the bunches weight

The first model implemented into the algorithm A, to estimate the grape weight was built using the data collected into the lab: bunch weight and the mean projected area of the bunch (from two images of the same bunch).

The second model implemented into the algorithm B has the same objective of the first model, but it was built to be closer to the real conditions of the field (the data collection was done into the field, with a bigger size of the projected area). To build the first model, it was calculated the mean area of the projected areas by the two images taken for each bunch (used for the detail measurements into the lab). Instead, into the field the projected area of the bunches has the characteristic to be more variable because the position and the side of the bunches visible into the images is casual. Therefore, by way of example, if we were to record an area of bunches in the field showing all, their side of minor area in the image, the result of the application of the first model would bring us to underestimate their weight; as otherwise the contrary, seeing all the sides of greater area of the bunches, the same model would bring us to over-estimate their weight. Therefore, not being able to know the orientation of the bunches on the plant, which determines the area which is recorded in the images, the second model records the variability of the total area of the bunches of each single linear canopy meter and its relationship with their relative total weight. The total area of the bunches per linear canopy meter was calculated adding to the projected area of bunches not covered by leaves the percentage of the bunch by bunch occlusion.

### 3.10.4 Estimation of yield at harvest

In order to estimate the final yield at harvest, the estimated weights of the bunches per linear canopy meter at pea size and at veraison were multiplied for their correspondent growth factors. The growth factors were calculated dividing the mean bunches' weight per linear canopy meter at maturation by that at pea-size and at veraison.

### 3.10.5 Structure of the algorithms used for yield estimation

As it is possible see in the figure 20, the structure of the algorithms A and B is the same, what it changes are the models used for the estimation of the bunches area without leaves' occlusion and for the estimation of bunches weight per linear canopy meter.

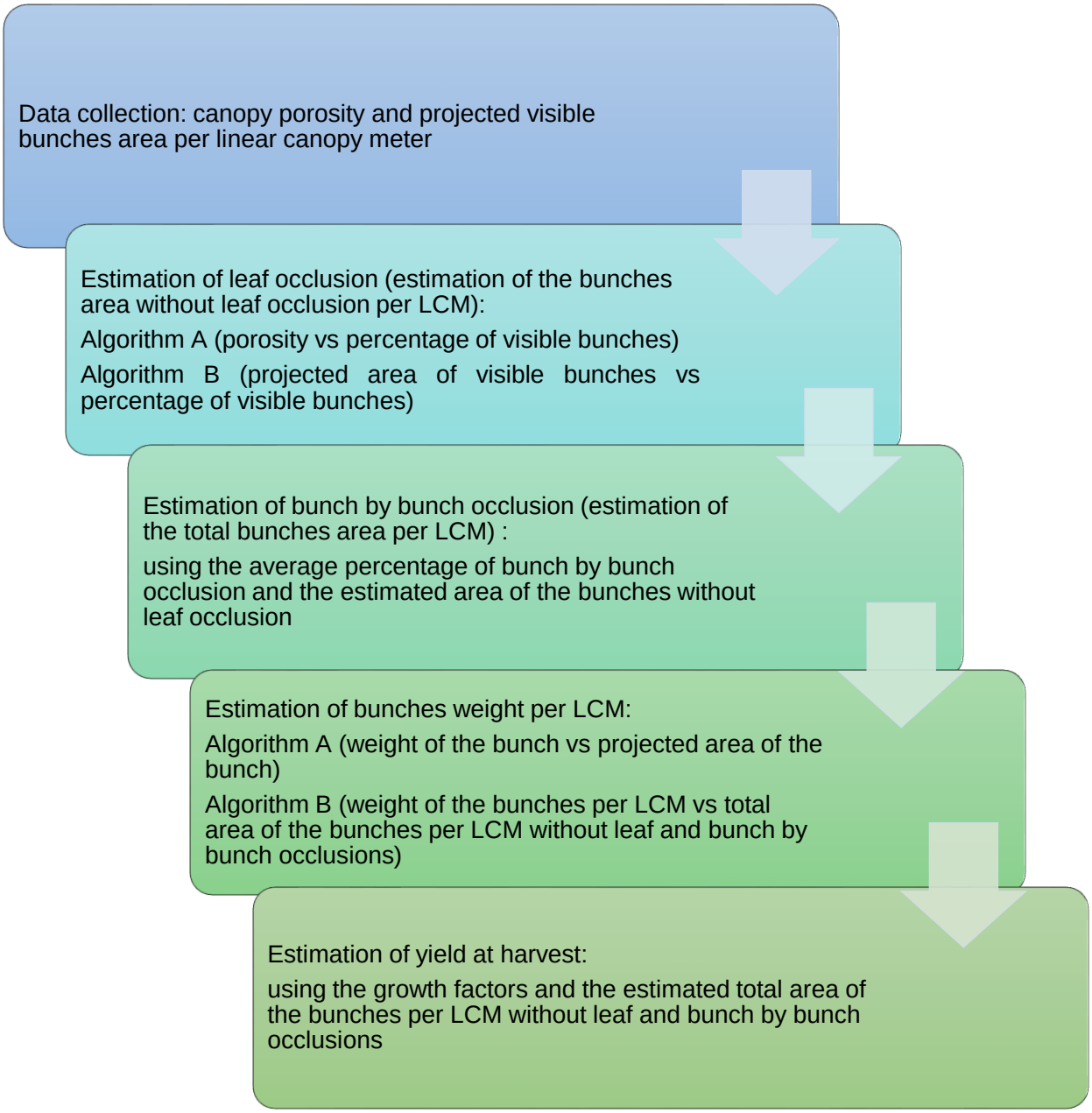



*Figure 20) Process diagram of the algorithms A and B, after the first box, in all boxes is presented the estimation's kind and the models used.*

## 4 Results and discussion

### 4.1 Yield components analysis

In order to verify the capability of the pictures taken with Vinbot, to be a good instrument to detect the yield components, them were counted directly into the vineyard (Observed) and on the pictures (Estimated) assuming that the future recognition algorithms can work as the human eyes. The counting on the images was done for all phenological stages, as presented in table 1.

*Table 1) Observed (Obs) and estimated (Est) yield components with their relative root mean square error (RMSE); mean absolute percent error (MA%E) between the ground truth (GT) and image counting (IMG); regression coefficient between GT and IMG ($r_{Obs,Est}$). Significance: ns: not significant; *: significant with ρ≤0.05; **: very significant with ρ≤0.01; ***: highly significant with ρ≤0.001.*

| | 4 leaves out (shoots) | Flowering (inflorescences) | Pre-bloom (inflorescences) | Pea-size (bunches) | Veraison (bunches) | Maturation (bunches) |
|---|---|---|---|---|---|---|
| **Obs** | 16.08±3.6 | 7.93±1.9 | 7.93±1.9 | 9.45±2.6 | 9.45±2.6 | 9.45±2.6 |
| **Est** | 10.82±3.28 | 2.50±1.18 | 3.63±1.63 | 2.95±1.58 | 2.48±1.40 | 3.43±1.92 |
| **MA%E** | 32±2% | 68±3% | 52±3% | 68±3% | 72±3% | 63±3% |
| $r_{Obs,Est}$ | 0.62*** | 0.40* | $0.12^{ns}$ | 0.47** | $0.26^{ns}$ | 0.40* |

The estimated yield components are, for every stage low as compared to the observed yield components, which is due to the occlusion generated by different organs of the vines into the canopy. Data shows that the average number of bunches is higher than the number of inflorescences, indicating that there might have been an underestimation of the number of inflorescences obtained by manual counting because it was extremely difficult due to the presence of a dense canopy.

The lower mean absolute percent error (MA%E) was obtained at phenological stage of four leaves out, this because the shoots are the organs easier to count, this error is higher than the error of 11.82% founded by Liu *et al*. (2017). The calculated correlation coefficient ($r_{Obs,Est}$) is highly significative.

Regarding the inflorescences we observed a higher MA%E at flowering, probably because the inflorescences was littler and therefore more difficult to identify compared to their counting at pre-bloom, when they were bigger and easier to count. The $r_{Obs,Est}$ was significant at flowering meanings that the error is, more or less constant, indeed at pre-bloom the $r_{Obs,Est}$ is not significant, showing that even if the MA%E is lower the influence of more and bigger leaves, some times, limit the visibility of the inflorescences at different levels.

Relatively at the bunches detection we can see that the MA%E is higher at veraison, probably because, the canopy development is a its maximum during this phenological stage (Lamb *et al.*, 2014) and the bunches are not at their full size. Instead at pea size we find a lower error, probably due to the not full developing of the canopy, that permit to observe more bunches even if they are smaller compared to veraison. At maturation we can observe a reduction of the MA%E compared to

pea size and veraison, probably due to the increase of bunch size and the contemporary stop of canopy development. We can observe a very significant $r_{Obs,Est}$ at pea size for the same reason presented to explain the MA%E, but it can be explained also by facts presented after in the section 4.2.1, regarding the number of layers present at this phenological stage. At veraison the $r_{Obs,Est}$ is not significant, it is probably due to the same reason presented for phenological stage of pre-bloom: presence of dense canopy that cover the bunches (which are also not full developed) at different levels. At maturation the $r_{Obs,Est}$ becames newly significant probably because the bunches are bigger and the leaf occlusion has a more regular influence on bunch visibility.

As presented in the section 3.5, in the 40 meters dedicated to test the algorithms, we used a meter left in the same position on the plants over all the time as suggest by the experience of the previous research team of ISA, that worked on the same topic. This was done to avoid issues with the correct assignation of yield components at the same meter during all the phenological stages, but even if it can be considered a good approach than to use a mobile meter, it is not a final and complete solution if the images are collected in a non-automatic way. The non-automatic navigation brings to have different acquisition angles of the images due at the irregularity of the soil and plants position. The problems that we had can be avoided through the future Vinbot's self-navigation because it will probably stabilise the problem taking images with the same angle. Even if the problem of acquisition angle of the images will remain a problem also with the self-navigation, it could be resolved during the image analysis using the solution proposed by Nuske *et al.* (2014 b) or it can be eliminated with the use of CNNs as affirmed by Santos *et al*. (2019). Another important thing that can be useful is the use of a better camera with higher definition and a system that can modify the height of the camera to manage the differences of the canopy within the varieties.

### 4.2 Data obtained from the vines dedicated to model building

To build the models for yield estimation we analysed in detail 20 LCM selected for each phenological stage. In the table 2, are presented the weight of the berries and of the bunches (obtained from the detail measurement done into the lab for the bunches selected, as explained into chapter material and methods), they show a bigger increase between pea size and veraison and a lower increment between veraison and maturation, as it's also shown by the mean bunches weight per LCM and by the growth factor. The growth factor shows a high increase of the bunches' weight from pea size to maturation and an increase around the double from veraison to maturation. The percentage of BxB occlusion decrease from pea-size to veraison and increase from veraison to maturation, its lower value was found at veraison and the higher at maturation.

*Table 2) Mean berry and bunch weight (g), mean bunches weight per linear LCM of vine (g/m), and mean percentage of bunch on bunch occlusion (mean % of BxB occlusion) for three phenological stages. The growth factor from pea size and from veraison to maturation. The Results from Smart-points dedicated to model building.*

| | **Phenological stage** | | |
|---|---|---|---|
| | ***Pea-size*** | ***Veraison*** | ***Maturation*** |
| **berry weight (g)** | 0.209±0.014 | 0.971±0.032 | 1.370±0.042 |
| **bunch weight (g)** | 65.7±5.1 | 244.8±20.5 | 427.0±26.5 |
| **bunches weight per m (g/m)** | 546.9±49.4 | 2122.6±20.2 | 3641.1±253.4 |
| **% of BxB occlusion** | 8% | 6% | 12% |
| **Growth factor** | 6.6 | 1.7 | 1 |

### 4.2.1 Bunch by bunch occlusion

Table 3 shows the data relative to the clustered LCM according to the same number of bunches' layers. We can observe that the number of layers grows during the season and therefore there is change also in the number of LCM with the same number of layers. This can be caused by the growth of the berries. This changes can also be due to the different number of bunches per vine, to the position of the shoots and to the position of the bunches along the shoot.

*Table 3) Meters clustered by the same number of layers, number of meters with the same number of layers for each phenological stage.*

| ***Pea size*** | **1 layer** | **2 layers** | **3 layers** | **4 layers** |
|---|---|---|---|---|
| **Canopy segment (m) with the same number of layers** | 5 | 11 | 4 | - |
| ***Veraison*** | **1 layer** | **2 layers** | **3 layers** | **4 layers** |
| **Canopy segment (m) with the same number of layers** | 3 | 11 | 5 | 1 |
| **Maturation** | **1 layer** | **2 layers** | **3 layers** | **4 layers** |
| **Canopy segment (m) with the same number of layers** | 2 | 6 | 9 | 3 |

Looking at the Fig. 21 A-C we can observe that the increase of layers' number (in addition to increasing the BxB occlusion) could bring the bunches at being closer to the external part of the canopy, having higher probability to be less covered by the leaves and therefore more visible.

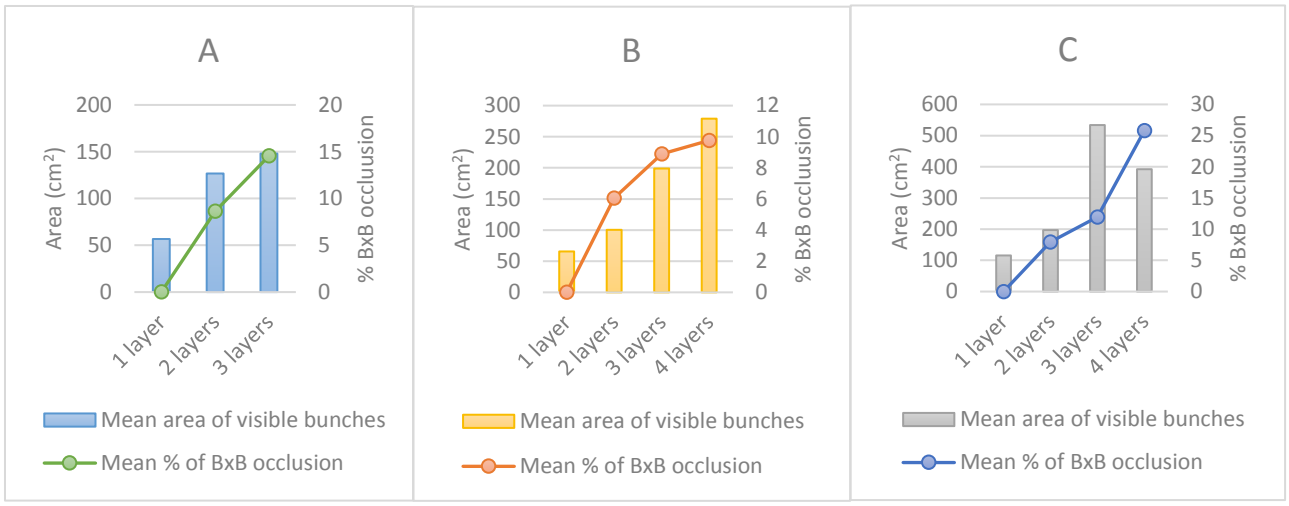



*Figure 21) Mean projected area of visible bunches on non-disturbed canopies and mean percentage of bunch on bunch (BxB) occlusion of the LCM clustered by the same number of layers at pea-size (A), veraison (B) and maturation (C).*

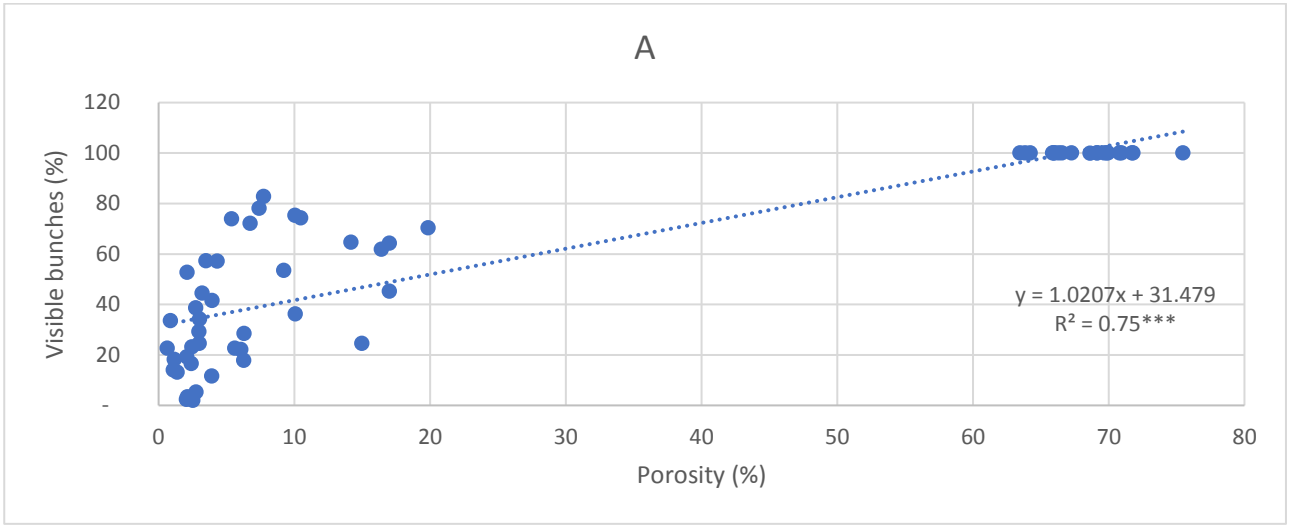
In some cases, the tendency of the increase of the visible area with the increase of the number of layers changes after pea-size when the layers are higher than three. This is probably due to the smaller number of bunches into each layer and thus a lesser number of bunches in the layer closer to the external part of the canopy, limiting therefore the general visibility of the bunches; this situation is easily to observe at maturation where a higher number of vines with four layers are present. Anyway, we can observe it only if we look at the average data, because there are other factors which determine the bunches visibility and the % of BxB occlusion. Probably the area of visible bunches is a strong predictor, but not the only one, that could be used to estimate the % of BxB occlusion, using multiple linear regression. So, not founding a good way to estimate the % of BxB occlusion to be used into the yield estimation algorithms we used the mean % of BxB occlusion founded for each phenological stage.

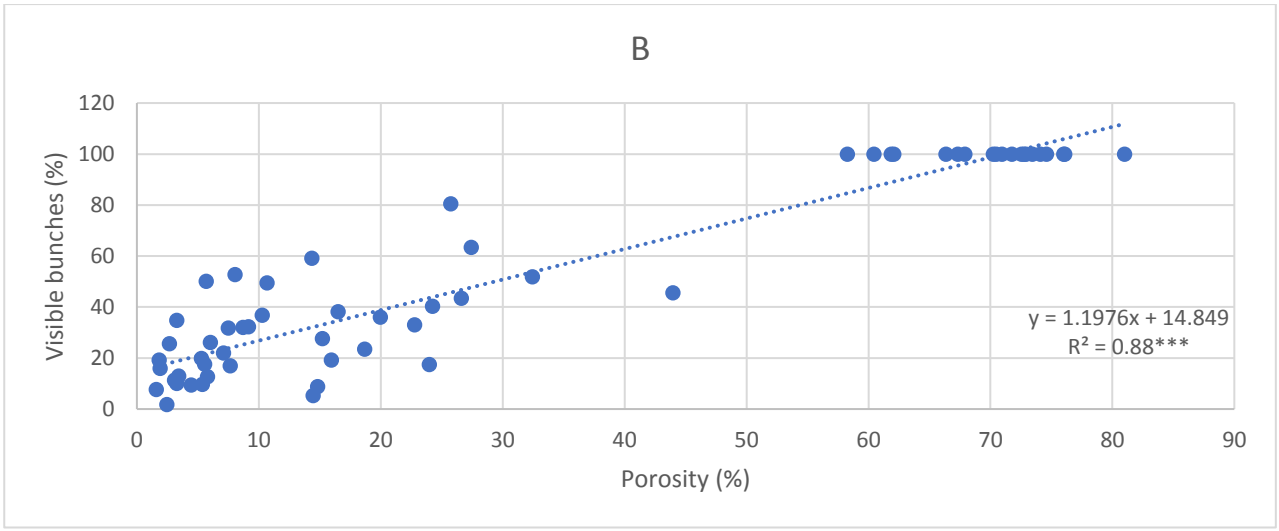
The parameters shown in the table 3 and in the figure 20 A-C could be used for future research, but they need to be studied deeper on a higher number of samples.

### 4.3 Estimation models

#### 4.3.1 Models to estimate the occluded bunches area by leaves

The first model, implemented in the algorithm A and described in the section 3.10.1, shown a $R^2$ of the linear regression between the fraction of visible bunches (dependent variable) and porosity (independent variable) of 0.75, 0.88 and 0.88 respectively at pea, veraison and at maturation (Fig. 22 A-C). These results show that the canopy porosity explains a high percentage of the ratio between the visible bunches area of non-disturbed canopy and the visible bunches area on full defoliated canopy. This allows to consider this variable as a good explanatory variable to predict the percentage of visible bunches area, and therefore to estimate the total area of bunches without leaf occlusion in any of the phenological phases considered.



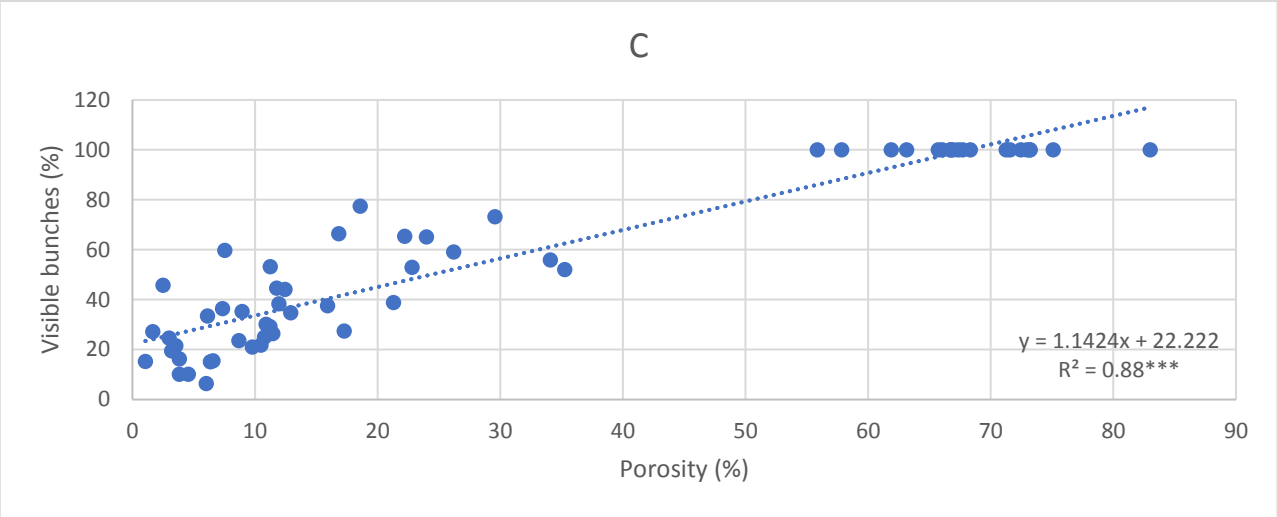
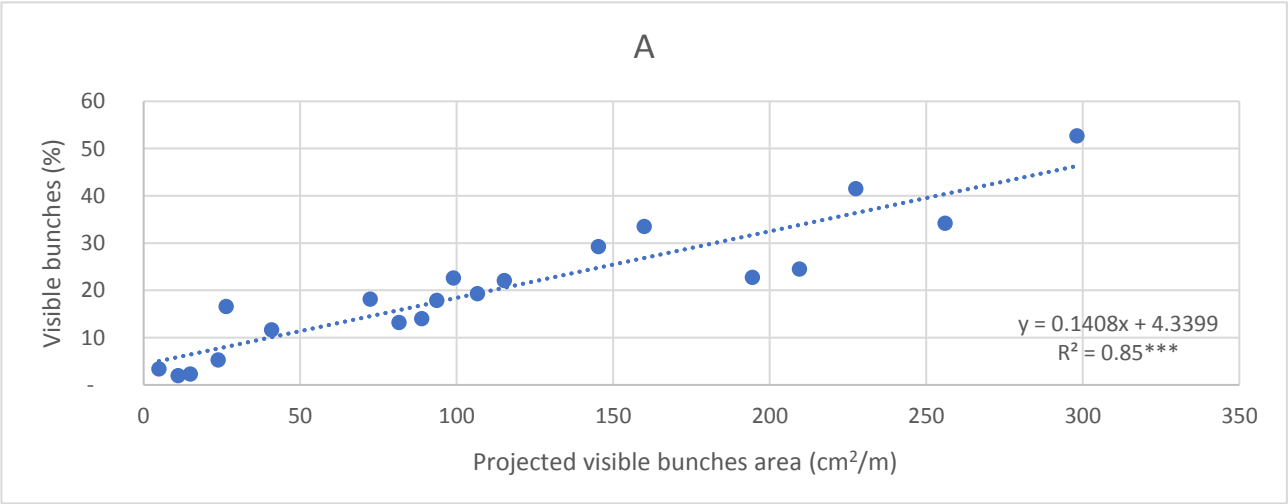



*Figure 22) Relationship between canopy porosity (independent variable) and percentage of bunches not covered by leaves (dependent variable), with respective regression equation and coefficient of determination ($R^2$), pea size phenological stage (A), n = 60. Veraison phenological stage (B), n = 60. Maturation phenological stage (C), n = 60. The *** indicates a significant $R^2$ ($p \leq 0.001$).*

The second model, implemented in the algorithm B and described in the section 3.10.1, shown a $R^2$ of the linear simple regression between the projected visible bunches area (independent variable) and the percentage of visible bunches (dependent variable) of 0.85, 0.57 and 0.69 respectively at pea-size, veraison and maturation (Fig. 23 A-C). These results show that the projected visible bunches area on non-disturbed canopy explains a high percentage of the ratio between the visible bunches area of non-disturbed canopy and the visible bunches area on full defoliated canopy. This allows to consider this variable as a good explanatory variable to predict the percentage of visible bunches area, and therefore to estimate the total area of bunches without leaf occlusion in any of the phenological phases considered.



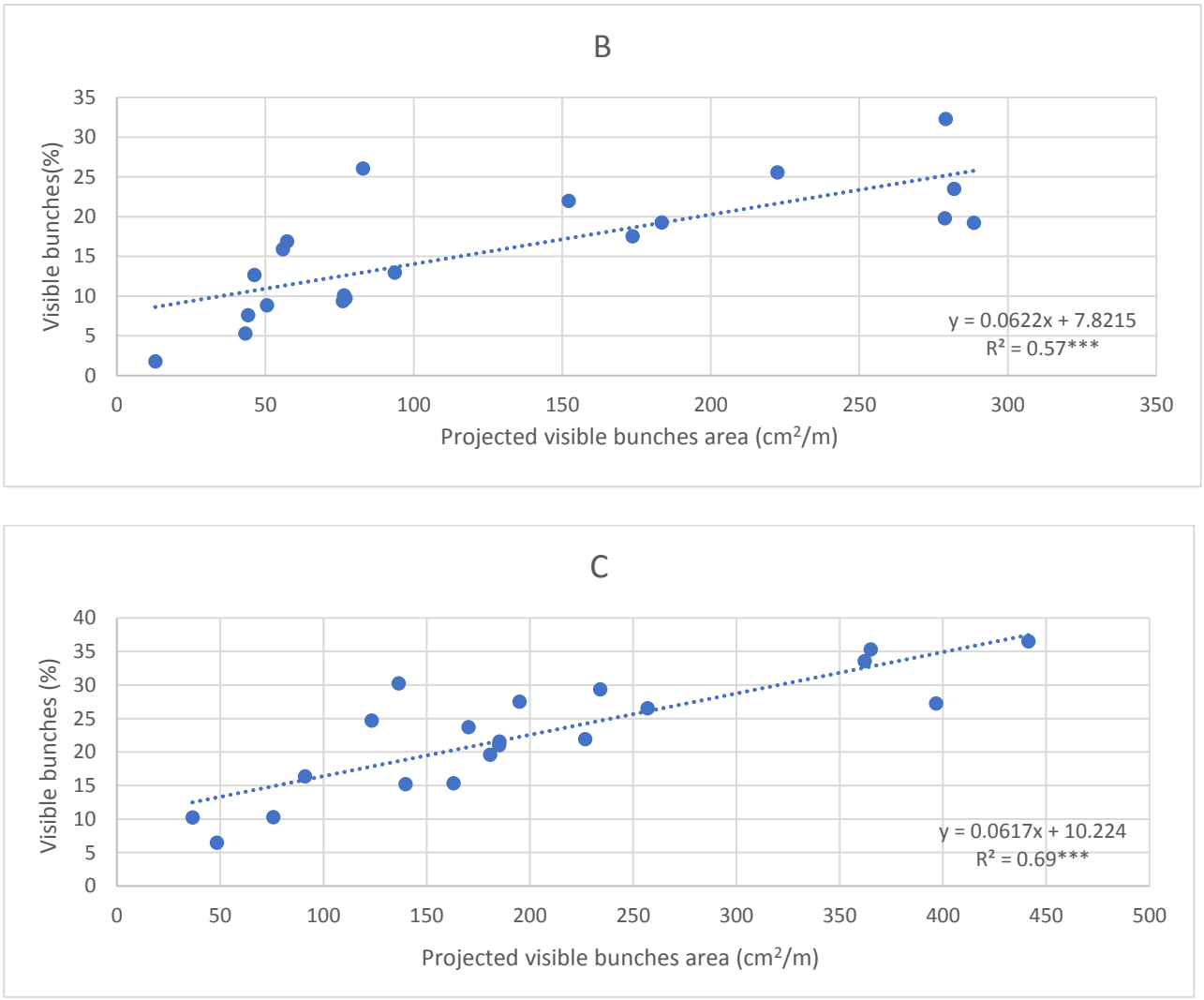



*Figure 23) Relationship between projected area of visible bunches per meter (independent variable) and percentage of bunches not covered by leaves (dependent variable), with respective regression equation and coefficient of determination ($R^2$). Pea size phenological stage, n = 20. Veraison phenological stage (B), n = 20. Maturation phenological stage (C), n = 20. The *** indicates a significant R2 (p≤ 0.001).*

### 4.3.2 Models to estimate the bunches weight

The first model, implemented in the algorithm A and described in the section 3.10.3, shown a $R^2$ of the linear simple regression between the projected bunch area in the image (independent variable) and the real weight of the bunch (dependent variable) of 0.83, 0.79 and 0.87 respectively at pea-size, veraison and maturation (Fig. 24 A-C). These results show that the projected area explains a high percentage of bunch weight variability, allowing to consider this variable as a good explanatory variable to predict bunch weight, in any of the phenological stages considered.

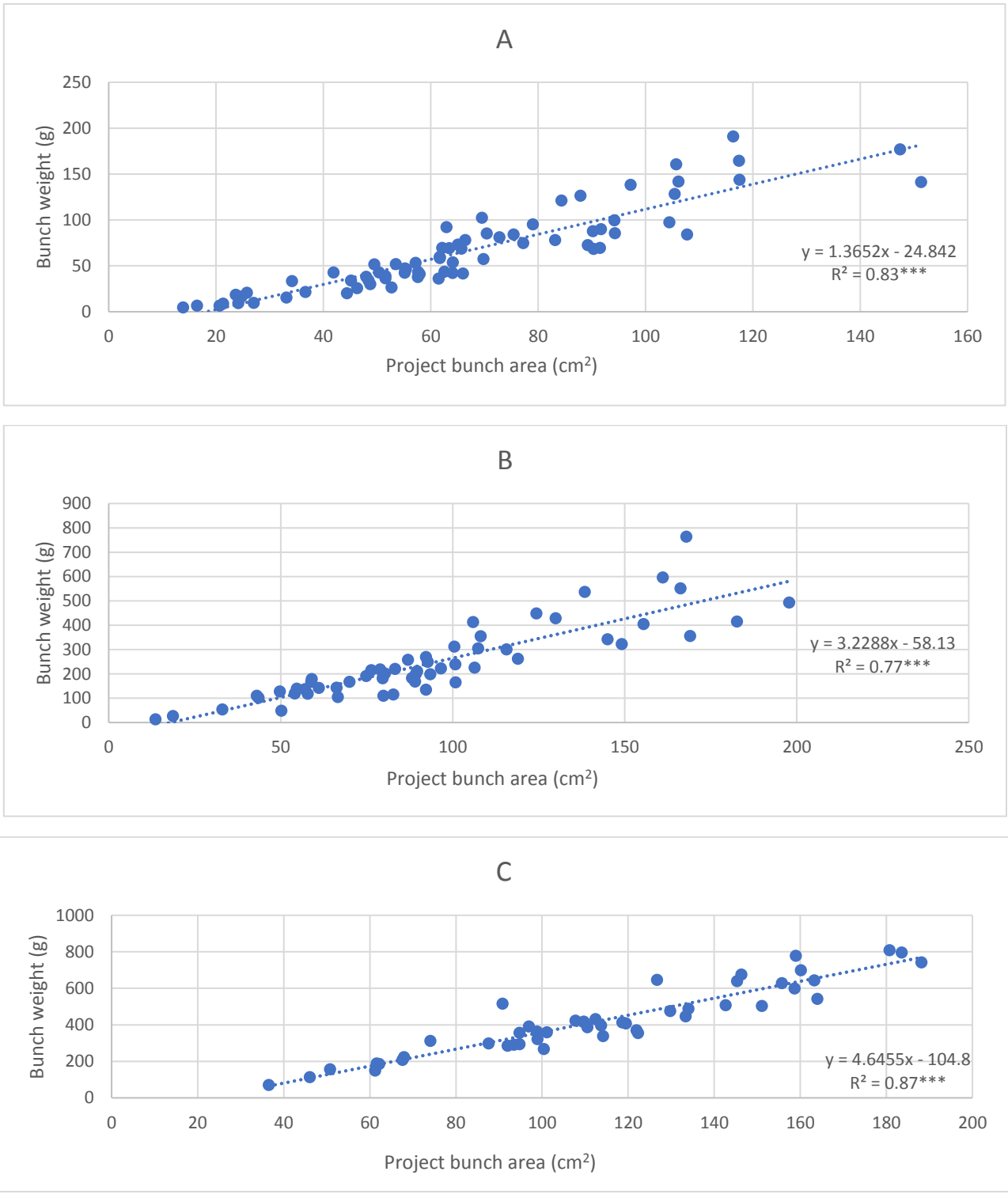



*Figure 24) Relationship between the projected area of the bunch (independent variable) and the weight of the bunch (dependent variable), with respective regression equation and coefficient of determination ($R^2$). Pea size phenological stage (A), n = 73. Veraison phenological stage, n = 56 (B). Maturation phenological stage (C), n = 49. The *** indicates a significant R2 ($p \leq 0.001$).*

The second model, implemented in the algorithm B and described in the section 3.10.3, shown a $R^2$ of the linear simple regression between the total area of bunches per linear canopy meter (independent variable) and the real weight of the bunches per linear canopy meter (dependent variable) of 0.79, 0.86 and 0.87 respectively at pea-size, veraison and maturation (Fig. 25 A-C).

These results show that the total area of bunches per meter explains a high percentage of bunch weight per meter variability, allowing to consider this variable as a good explanatory variable to predict bunches weight per meter, in any of the phenological phases considered.

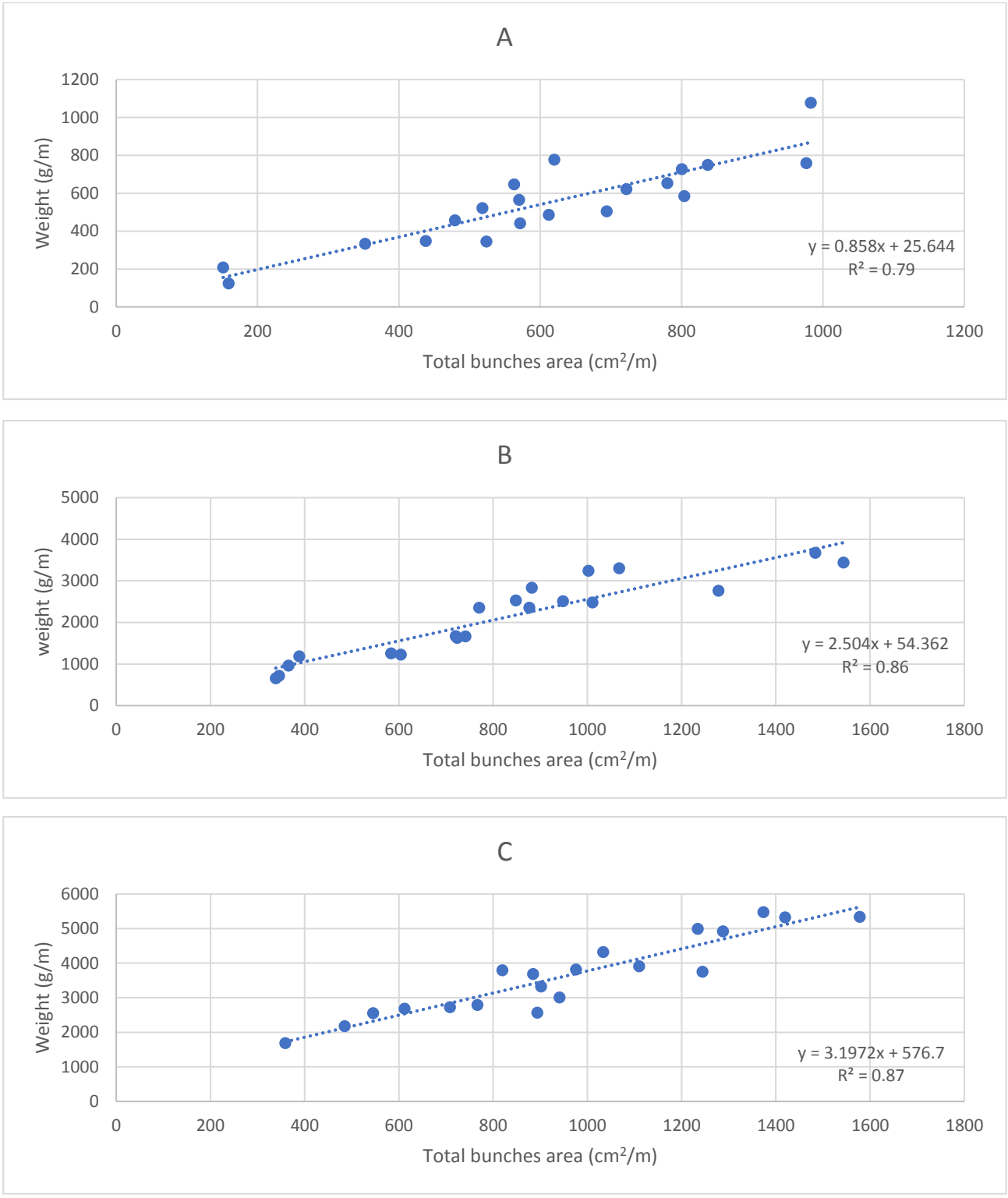



*Figure 25) Relationship between the total area of the bunches per meter (independent variable) and the weight of the bunches per meter (dependent variable), with respective regression equation and coefficient of determination ($R^2$). Pea size phenological stage, n = 20; veraison phenological stage, n = 20; maturation phenological stage, n = 20. The *** indicates a significant $R^2$ ($p \leq 0.001$).*

### 4.4 Application results of yield estimation algorithms

The two algorithms were tested on the same 40 LCM analysed at pea size, veraison and maturation. The average weight per meter calculated at harvest was of 3084.0±178.3 g/m. Looking figure 34 is possible see that the algorithm A gives a MA%E between the estimated and the real yield bigger than the MA%E obtained with the algorithm B for all phenological stages.

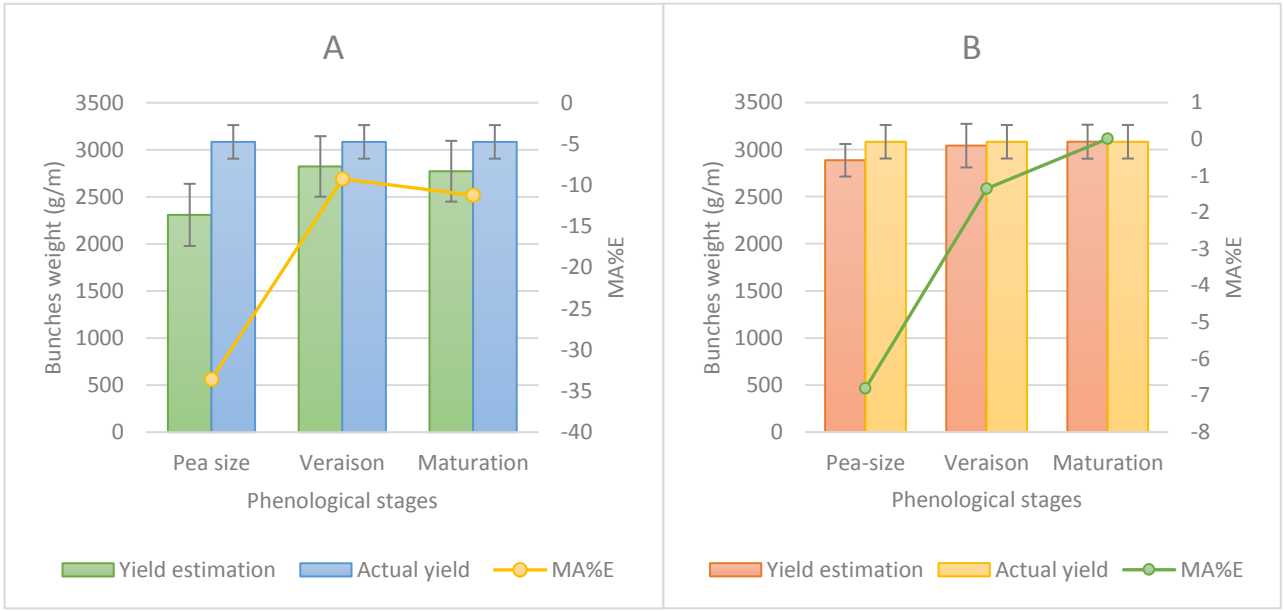



*Figure 26) Comparison between estimated and observed yield per meter (g/m) and (mean absolute percent error) MA%E at pea size, veraison and maturation for algorithm (A) and (B).*

Looking at the table 4 is possible see that the Root Mean Square Error (RMSE) of the mean estimated weight per linear canopy meter is similar at veraison between the two algorithms, but for the other two phenological stages the values are close to be the double for the algorithm A. For both algorithms the higher Mean Absolute Error (MAE) is at pea size, while we founded the lower MAE at maturation and at veraison for the algorithms B and A, respectively. Looking at the values of MA%E we can observe that the algorithm B gives better results against the algorithm A for all phenological stages. If we extend the prevision at one hectare we observe easier the entity of the error. Yield forecast is very important for several reasons, one of them is the organisation of the winery: with the result of the algorithm A we will have an underestimation of about 9 hL/ha of must, if we consider a 70% the ratio between weight and must, pressing the grape; with the estimation obtained with the algorithm B we will have an overestimation of about 1 L/ha. In line with what Folwell *et al*. (1994) stated: yield predictions can be attempted at any time during the growing cycle, becoming more accurate later in the cycle, with the algorithm B we had a decrease of the error from pea size to maturation.

*Table 4)* Root mean square error (RMSE) per meter (g/m) of the yield estimation, mean absolute error (MAE) calculated per meter (g/m) and for hectare (kg/ha) and mean absolute percent error (MA%E) of the mean estimated yield of the algorithms A and B at pea size, veraison and maturation.

| | **Algorithm A** | | | **Algorithm B** | | |
|---|---|---|---|---|---|---|
| **Phenological stage** | Pea size | Veraison | Maturation | Pea size | Veraison | Maturation |
| **RMSE (g/m)** | 330.5 | 321.6 | 322.9 | 172.9 | 231.8 | 180.6 |
| **MAE (Kg/ha)** | -3102 | -1044 | -1327 | -786 | -164 | 1.6 |
| **MAE (g/m)** | -775.6 | -261.0 | -311.8 | -196.6 | -41.0 | 0.4 |
| **MA%E** | -33.59% | -9.24% | -11.25% | -6.81% | -1.35% | 0.01% |

The results, of the algorithm B obtained for each phenological stage, are better than the results of the algorithm A and also better than the MA%E of 9.8% obtained by Nuske *et al.* (2011). Being a new algorithm, it will need to be tested on different cultivars, sites, and seasons. If its excellent results will be confirmed, they may justify the collection of the data necessary to adapt the algorithm for other vineyards. Moreover, more experience and data would increase the accuracy (Sabbatini *et al*., 2012) of the estimation algorithm and a different experimental design will give the possibility to build and test the algorithm for different levels of defoliation.

## 5 Conclusions

With the exception of the number of shoots, the estimated number of yield components, obtained by visual counting on the images, were lower as compared to the observed yield components in the field, which is due to the occlusion generated by different organs of the vines into the canopy. The lower mean absolute percent error (MA%E), was obtained at the phenological stage of four leaves out, this because the shoots are the organs easier to count and the calculated correlation coefficient ($r_{Obs,Est}$) between observed and estimated shoots was 0.62 . So, the most promising yield estimation, based on the counting of the yield components done through image analysis, was at the phenological stage of four leaves out.

Two algorithms were developed to estimate the yield. They use two different models to estimate the area of the bunches occluded by leaves and two different models to estimate the weight of the grape. In order to estimate the area of the bunches occluded by leaves, in the first algorithm was used a model based on the relationship between the fraction of visible bunches (dependent variable) and porosity (independent variable). The linear regression between these two variables shown a $R^2$ of 0.75, 0.88 and 0.88 respectively at pea size veraison and at maturation (with $\rho \leq 0.001$ for all phenological stages). In the second algorithm it was used a model based on the relationship between the projected visible bunches area in the image (independent variable) and the percentage of visible bunches (dependent variable). The linear regression between these two variables shown a $R^2$ = 0.85, 0.57 and 0.69 respectively at pea-size, veraison and maturation (with $\rho \leq 0.001$ for all phenological stages).

In order to estimate the weight of the bunch, in the first algorithm it was used a model based on the relationship between the projected bunch area in the image (independent variable) and the real weight of the bunch (dependent variable). The linear regression between these two variables shown a $R^2$ = 0.83, 0.79 and 0.87 respectively at pea-size, veraison and maturation (with $\rho \leq 0.001$ for all phenological stages).In the second algorithm it was used a model based on the relationship between the total area of bunches per linear canopy meter (independent variable) and the real weight of the bunches per linear canopy meter (dependent variable). The linear regression between these two variables shown a $R^2$ = 0.79, 0.86 and 0.87 respectively at pea-size, veraison and maturation (with $\rho \leq 0.001$ for all phenological stages).

The algorithms estimated the bunch by bunch occlusion using the average percentage of bunch by bunch occlusion obtained for every phenological stage in the LCM used for the models building. The values of the average percentage of bunch by bunch occlusion were 8%, 6% and 12% respectively at pea-size, veraison and maturation. Then to estimate the yield at harvest from the estimated weight for the different phenological stages it were used the growth factors of 6.6 and 1.7 obtained at pea-size and veraison, respectively.

All these linear regressions showed a high $R^2$ with high level of significance, but when the algorithms were applied to estimate the yield on the 40 LCM used to test the algorithms, we obtained the best results with the second algorithm. The first algorithm shown a MA%E, between the estimated and observed values of yield, of -33.59%, -9.24% and -11.25% respectively at pea-size, veraison and maturation. Instead the second algorithm shown a MA%E between the estimated and observed values of yield of -6.81%, -1.35% and 0.01% respectively at pea-size, veraison and maturation.

## 6 References


Al-Amin, M. (2017). Set theory is the ultimate branch of Mathematics. *International Journal of Innovative Knowledge* Concepts, 5(1).

Altman, D. G., & Bland, J. M. (1994). Diagnostic tests. 1: Sensitivity and specificity. *BMJ: British Medical Journal*,*308*(6943), 1552.

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. *American Journal of Enology and Viticulture*,*59*(2), 188-193.

Antunes M., Lehmann J., Eiras-Dias J. E. and Böhm J. (2011). *Arinto, Atlas das Castas da Península Ibérica*, viewed in 08/01/2020, <https://www.vinetowinecircle.com/castas_post/arinto/>

Aquino, A., Barrio, I., Diago, M. P., Millan, B., & Tardaguila, J. (2018). VitisBerry: An Android-smartphone application to early evaluate the number of grapevine berries by means of image analysis. *Computers and Electronics in Agriculture*,*148*, 19-28.

Aquino, A., Diago, M. P., Millán, B., & Tardáguila, J. (2017). A new methodology for estimating the grapevine-berry number per cluster using image analysis. *Biosystems engineering*,*156*, 80-95.

Aquino, A., Millan, B., Gaston, D., Diago, M. P., & Tardaguila, J. (2015) a. vitisFlower®: Development and testing of a novel android-smartphone application for assessing the number of grapevine flowers per inflorescence using artificial vision techniques.*Sensors*,15(9), 21204-21218.

Aquino, A., Millan, B., Gutiérrez, S., & Tardáguila, J. (2015). Grapevine flower estimation by applying artificial vision techniques on images with uncontrolled scene and multi-model analysis. *Computers and Electronics in Agriculture*, 119, 92-104.

Behroozi-Khazaei, N., & Maleki, M. R. (2017). A robust algorithm based on color features for grape cluster segmentation. *Computers and electronics in agriculture*,*142*, 41-49.

Blom, P. E., & Tarara, J. M. (2009). *"Trellis tension monitoring improves yield estimation in vineyards". HortScience*, *4*(3), 678-685.

Borgogno-Mondino E., A. Lessio, L. Tarricone, V. Novello, L. de Palma (2018). A comparison between multispectral aerial and satellite imagery in precision viticulture*. Precision Agric 19:195–217*.

Boselli M. *et al*., (2016). Progress in viticulture, chapter 12 (*Precision viticulture*) *Ermes Frazzi, Massimo Vincini, Edises*, ISBN 978-887-959-906-1.

Chamelat, R., Rosso, E., Choksuriwong, A., Rosenberger, C., Laurent, H., & Bro, P. (2006, November). Grape detection by image processing. In *IECON 2006-32nd Annual Conference on IEEE Industrial Electronics* (pp. 3697-3702). *IEEE*.

CIE Colorimetry (2nd edn). *Publication 15.2, CIE,* Vienna (1986).

Clingeleffer P. (2001). Crop Development, Crop Estimation and Crop Control to Secure Quality and Production of Major Wine Grape Varieties: A National Approach*. Department of N Folwell R.J., Santos D.E., Spayd S.E., Porter L.H*.

Coombe, B. G. (2004). Grapevine growth stages-The modified EL system.*Viticulture*, *1*, 150-154.

Correa, C., Valero, C., Barreiro, P., Diago, M. P., & Tardáguila, J. (2011, November). A comparison of fuzzy clustering algorithms applied to feature extraction on vineyard. In *Proceedings of the XIV Conference of the Spanish Association for Artificial Intelligence*.

Cortez, R., Luna-Vital, D. A., Margulis, D., & Gonzalez de Mejia, E. (2017). Natural pigments: stabilization methods of anthocyanins for food applications. *Comprehensive Reviews in Food Science and Food Safety*, *16*(1), 180-198.

Cotton, S. D. O. (1996). Colour, colour spaces and the human visual system. *School of Computer Science, University of Birmingham*. Technical Report.

Cubero S., Diago M. P., Blasco J., Tardáguila J., Prats-Montalbán J. M., Ibáñez J., Tello J. & Aleixos N. (2015). A new method for assessment of bunch compactness using automated image analysis. *Australian journal of grape and wine research*, *21*(1), 101-109.

Cunha, M., Abreu, I., Pinto, P., & de Castro, R. (2003). Airborne pollen samples for early-season estimates of wine production in a Mediterranean climate area of northern Portugal*. American Journal of Enology and Viticulture*, *54*(3), 189-194.

de Carvalho Cardoso, J. V. (1965). Os solos de portugal: sua classificação caracterização e génese. *Secretaria de estado da agricultura. Direcção-geral dos serviços agrícolas.*

del-Moral-Martínez, I., Rosell-Polo, J., Sanz, R., Masip, J., Martínez-Casasnovas, J., & Arnó, J. (2016). Mapping vineyard leaf area using mobile terrestrial laser scanners: should rows be scanned on-the-go or discontinuously sampled?. *Sensors, 16(1), 119.*

Diago, M. P., & Tardaguila, J. (2015a). Vinerobot: on-the-go vineyard monitoring with non-invasive sensors. *In 19th Meeting of the Group of international Experts of vitivinicultural Systems for CoOperation (GiESCO)*.

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

Diago, M. P., Krasnow, M., Bubola, M., Millan, B., & Tardaguila, J. (2016). Assessment of vineyard canopy porosity using machine vision. *American Journal of Enology and Viticulture*, *67*(2), 229-238.

Diago, M. P., Sanz-Garcia, A., Millan, B., Blasco, J., & Tardaguila, J. (2014). Assessment of flower number per inflorescence in grapevine by image analysis under field conditions. *Journal of the Science of Food and Agriculture*, *94*(10), 1981-1987.

Diago, Maria P and Tardaguila, Javier (2015b). A new robot for vineyard monitoring [online]. Wine & Viticulture Journal, Vol. 30, No. 3,: 38, 41-42 ISSN: 1838-6547. Viewed in 02/05/2019. <https://search.informit.com.au/documentSummary;dn=256729208411211;res=IELHSS>

Dry, P. R., Longbottom, M. L., McLoughlin, S., Johnson, T. E., & Collins, C. (2010). Classification of reproductive performance of ten winegrape varieties. *Australian Journal of Grape and Wine Research*, *16*, 47-55.

Duarte M. J. (2019). VINBOT: robot vision and learning from vineyards. Introduction to the research in Electrical and computer engineering. Supervisors: Prof. José Santos-Victor and Prof. Jorge Salvador Marques. Instituto Superior Técnico, Lisboa.

Dunn G. (2003). Predicting bunch weight early in the season*. The Australian & New Zealand Grapegrower & Winemaker.* November, pp. 29-30.

Dunn G. (2010). Yield forecasting*. Innovators network, fact sheet, University of Melbourne*.

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.

Eiras-Dias, J. E., Faustino, R., Clímaco, P., Fernandes, P., Cruz, A., Cunha, J., ... & Castro, R. D. (2011). Catálogo das castas para vinho cultivadas em Portugal. *Vol. I, Instituto da Vinha e do Vinho, Lisboa*.

Folwell RJ, Santos DE, Spayd SE, Porter LH and Wells DS, (1994). Statistical technique for forecasting Concord grape production. *Am J Enol Vitic* 45:63–70.

Gatti M., Garavani A., Vercesi A., and Poni S. (2017), Ground-truthing of remotely sensed within-field variability in cv. Barbera plot for improving vineyard management*. Australian Society of Viticulture and Oenology Inc*. pp. 399-408, doi: 10.1111/ajgw.12286.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. Book in preparation for MIT Press.*URL¡ http://www. deeplearningbook. org*.

Grossetete, M., Berthoumieu, Y., Da Costa, J. P., Germain, C., Lavialle, O., & Grenier, G. (2012). Early estimation of vineyard yield: site specific counting of berries by using a smartphone. In *International Conference of Agricultural Engineering—CIGR-AgEng*.

Guzmán, R., Ariño, J., Navarro, R., Lopes, C. M., Graça, J., Reyes, M., ... & Braga, R. (2016 b, November). Autonomous hybrid GPS/reactive navigation of an unmanned ground vehicle for precision viticulture-VINBOT. In *62nd German Winegrowers Conference At: Stuttgart*.

Guzman, R., Navarro, R., Beneto, M., & Carbonell, D. (2016 a). Robotnik—Professional service robotics applications with ROS. *In Robot Operating System (ROS) (pp. 253-288). Springer, Cham.*

Hastie, T., Tibshirani, R., Friedman, J., & Franklin, J. (2005). The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer*, *27*(2), 83-85.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

He, X., Cai, D., & Niyogi, P. (2006). Laplacian score for feature selection. In *Advances in neural information processing systems* (pp. 507-514).

Higgins, J. J. (2004). An introduction to modern nonparametric statistics. *Brooks/Cole Cengage Learning.*

IPMA, Instituto Português do Mar e da Atmosfera. Viewed in 20/04/2019. <http://www.ipma.pt/pt/oclima/normais.clima/1971-2000/001/>

Ivorra, E., Sánchez, A. J., Camarasa, J. G., Diago, M. P., & Tardáguila, J. (2015). Assessment of grape cluster yield components based on 3D descriptors using stereo vision. *Food Control*, *50*, 273-282.

Kendall, M. G. (1946). The advanced theory of statistics. *The advanced theory of statistics.*, (2nd Ed).

Kicherer, A., Klodt, M., Sharifzadeh, S., Cremers, D., Töpfer, R., & Herzog, K. (2017). Automatic image-based determination of pruning mass as a determinant for yield potential in grapevine management and breeding. *Australian journal of grape and wine research*, *23*(1), 120-124.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., & Rubel, F. (2006). World map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift*, *15*(3), 259-263.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Image net classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).

Lachat, E., Macher, H., Mittet, M. A., Landes, T., & Grussenmeyer, P. (2015). First experiences with Kinect v2 sensor for close range 3D modelling. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *40*(5), 93.

Lamb, D. W., Weedon, M. M., & Bramley, R. G. V. (2004). Using remote sensing to predict grape phenolics and colour at harvest in a Cabernet Sauvignon vineyard: Timing observations against vine phenology and optimising image resolution. *Australian Journal of Grape and Wine Research*, *10*(1), 46-54.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning*. Nature*, *521*(7553), 436.

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

Liu, S., 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.

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3431-3440).

Lopes, C. M., Torres A., Guzman R., Graca J., Reyes M., Vitorino G., Braga R., Monteiro A., Barriguinha A. (2017), Using an unmanned vehicle to scout vineyards for non-intrusive estimation of canopy features and grape yield . *GiESCO Mendoza conference paper*, pp 16-21.

MacMiller I. and Fisher K. (2005). Yield estimation – can we do it better? The Tinlins McLaren Vale Shiraz experience*. The Australian & New Zealand Grapegrower & Winemaker, n° 500 pp. 26-28.*

MacQueen, J. (1967, June). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (Vol. 1, No. 14, pp. 281-297).

Manualslib (2019a), *the ultimate manuals library*, *Nikon D5200 User Manual, single-lens reflex digital camera*, viewed in 20/05/2019. <https://www.manualslib.com/manual/745198/Nikon-D5200.html>

Manualslib (2019b) *the ultimate manuals library*, *Sony DSC-H90 Specifications, cyber-shot digital still camera*, viewed in 02/05/2019. <https://www.manualslib.com/manual/262880/Sony-Dsc-H90.html#product-DSC-H90B>

Martin, R. D. S., & Dunn, G. (2003). How to forecast wine grape deliveries*. Technique report, Department of Primary Industries*.

Matese A. and Di Gennaro S. F. (2018), Practical Applications of a Multisensor UAV Platform Based on Multispectral, Thermal and RGB High Resolution Images in Precision Viticulture*, Agriculture 8, 116; doi:10.3390/agriculture8070116.*

Matese, A., & Di Gennaro, S. F. (2015). Technology in precision viticulture: A state of the art review*. Int. J. Wine Res*, *7*, 69-81.

Mathworks (2019) *Convert from HSV to RGB Color Space*. Viewed in 02/05/2019. <https://www.mathworks.com/help/images/convert-from-hsv-to-rgb-color-space.html>

May, P. (2004). Development after fertilization. In *P. May (Ed.), Flowering and fruit set in grapevines. Adelaide (Australia): Lythrum Press.*

May, P. (2004, b). Flowering and Fruitset in Grapevines. *Lythrum Press*.

Millan, B., Aquino, A., Diago, M. P., & Tardaguila, J. (2017). Image analysis-based modelling for flower number estimation in grapevine. *Journal of the Science of Food and Agriculture*, *97*(3), 784-792.

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*.

Nisbet R., Miner G., Yale K. (2018). Chapter 20 - Significance versus Luck in the Age of Mining: The Issues of P-Value “Significance” and “Ways to Test Significance of Our Predictive Analytic Models”, Editor(s): Robert Nisbet, Gary Miner, Ken Yale, *Handbook of Statistical Analysis and Data Mining Applications (Second Edition), Academic Press, Pages 753-765.*

NOAA, 2019. *National Environmental Satellite, Data and Information Service*. viewed in 29/08/2019. <https://www7.ncdc.noaa.gov/CDO/cdoselect.cmd?datasetabbv=GSOD&countryabbv=&georegionabbv=&resolution=40>

Nuske, S., Achar, S., Bates, T., Narasimhan, S., & Singh, S. (2011, September). Yield estimation in vineyards by visual grape detection. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 2352-2358). *IEEE.*

Nuske, S., Gupta, K., Narasimhan, S., & Singh, S. (2014 a). Modeling and calibrating visual yield estimates in vineyards. In *Field and Service Robotics* (pp. 343-356). Springer, Berlin, Heidelberg.

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

Nuzzo, V., & Matthews, M. A. (2005, September). Berry size and yield paradigms on grapes and wines quality. In *International Workshop on Advances in Grapevine and Wine Research 754* (pp. 423-436).

O'Haver T. (2019). *A Pragmatic Introduction to Signal Processing, Smoothing*. Viewed in 03/06/2019. <https://terpconnect.umd.edu/~toh/spectrum/Smoothing.html>

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics, 9(1), 62-66.*

Palliotti, A., and O. Silvestroni. 2004. Ecofisiologia applicate alla vite. In: Viticoltura ed enologia biologica. *Eugenio Cozzolino (ed), pp. 41-88. Edagricole. Bologna, Italia*.

Palus, H. (1998). Colour spaces, chapter 4. *Chapmann and Hall,* 67.

Pinelli P., Romani A., Fierini E., Agati G. (2018). Prediction models for assessing anthocyanins in grape berries by fluorescence sensors: Dependence on cultivar, site and growing season*. Food Chemistry* 244, 213-223.

Poynton, C. (1997). Frequently asked questions about color. *Retrieved June 19*, 2004.

Poynton, C. A. (1995, February). A Guided Tour of Colour Space. In *New Foundation for Video Technology: The SMPTE Advanced Television and Electronic Imaging Conference* (pp. 167-180). SMPTE.

Rabatel, G., & Guizard, C. (2007). Grape berry calibration by computer vision using elliptical model fitting. In *European Conference on Precision Agriculture* (Vol. 6, pp. 581-587).

Rafael. C. González, Woods, R. E., & Eddins, S. L. (2004). Digital Image Processing Using MARLAB. *Pearson.*

Rasband W. *ImageJ, introduction*. Viewed in 20/4/2019. <https://imagej.nih.gov/ij/docs/intro.html>

Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Springer, Cham.

Rosebrock A. (November 7, 2016) *Machine Learning, Object Detection, Tutorials*, viewed in 03/05/2019 <https://www.pyimagesearch.com/2016/11/07/intersection-over-union-iou-for-object-detection/>

Rudolph, R., Herzog, K., Töpfer, R., & Steinhage, V. (2018). Efficient identification, localization and quantification of grapevine inflorescences in unprepared field images using Fully Convolutional Networks. *arXiv preprint arXiv:1807.03770*.

S. CS. CUBERO, M.P. DIAGO, J. BLASCO, J. TARDAGUILA, J.M. PRATS-MONTALBÁN, J. IBÁÑEZ, J. TELLO & ALEIXOS (2015). A new method for assessment of bunch compactness using automated image analysis*. Australian journal of grape and wine research*,*21*(1), 101-109.

Sabbatini P., Dami I. and Howell G.S., (2012). Predicting Harvest Yield in Juice and Wine Grape Vineyards. *Extension Bulletin 3186, Michigan State University*.

Santos, T. T., de Souza, L. L., Santos, A. A. D., & Avila, S. (2019). Grape detection, segmentation and tracking using deep neural networks and three-dimensional association. *arXiv preprint arXiv:1907.11819*.

Sarmento, A. M. (1969). Elementos para a Elaboração do Plano de Trabalho na Exploração Agrícola da Tapada da Ajuda. *Relatório Final do Curso de Engenheiro Agrónomo, Instituto Superior de Agronomia – Universidade Técnica de Lisboa (atual Universidade de Lisboa).*

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, *61*, 85-117.

Silva, C., & Ribeiro, B. (2009, September). Improving text classification performance with incremental background knowledge. In *International Conference on Artificial Neural Networks* (pp. 923-931). Springer, Berlin, Heidelberg.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Škrabánek, P., & Majerík, F. (2017, July). Detection of grapes in natural environment using HOG features in low resolution images. In *Journal of Physics: Conference Series* (Vol. 870, No. 1, p. 012004). IOP Publishing.

Smart, R. E. (1986, August). Influence of light on composition and quality of grapes. In *Symposium on Grapevine Canopy and Vigor Management, XXII IHC 206* (pp. 37-48).

Smart, R., & Robinson, M. (1991). Sunlight into wine: a handbook for winegrape canopy management. Winetitles.

Son, J., Inoue, N., & Yamashtia, Y. (2010). Geometrically local isotropic independence and numerical analysis of the Mahalanobis metric in vector space. *Pattern Recognition Letters*, *31*(8), 709-716.

Stat Trek (2019) *teach yourself statistics Statistics Dictionary, Accurancy*. Viewed in 03/07/2019. <https://stattrek.com/statistics/dictionary.aspx?definition=accuracy>

Tan, Y. (2016). GPU-based parallel implementation of swarm intelligence algorithms. *Morgan Kaufmann. Chapter 11, pp. 167-177.*

Tardaguila, J., Blanco, J. A., Poni, S., & Diago, M. P. (2012). Mechanical yield regulation in winegrapes: Comparison of early defoliation and crop thinning. *Australian Journal of Grape and Wine Research*, *18*(3), 344-352.

Teixeira, G., Monteiro, A., Santos, C., & Lopes, C. M. (2018). Leaf morphoanatomy traits in white grapevine cultivars with distinct geographical origin. *Ciência e Técnica Vitivinícola*.

Tkalcic, M., & Tasic, J. F. (2003). Colour spaces: perceptual, historical and applicational background (Vol. 1, pp. 304-308). *IEEE.*

Vinbot, *powerful precision viticulture tool to break traditional yield estimation in vineyards, technical features.* Viewed in 02/05/2019. <http://vinbot.eu/robot-2/robot/>

VineGuard: Futuristic Concept Robot for Agriculture (2017) The VineGuard project. Viewed in 20/04/2019.<http://robotics.bgu.ac.il/index.php/Development_of_an_Autonomous_vineyard_sprayer>.

Vitirover, micro winery robotics (2018). Saint-Émilion: Vitirover. Viewed in 20/04/2019. <http://www.vitirover.com/>.

Wells D.S., (1994). Statistical technique for forecasting Concord grape production. *Am. J. Enol. Vitic., 45, 63-70. Natural Resources and Environment, Project number CSH 96/1, pp. 1-159.*

Wines of Portugal, a world of difference. *Grape varieties, Arinto*. Viewed in 02/03/2019. <http://www.winesofportugal.info/pagina.php?codNode=1092>

Wyszecki Gunter, & Stiles, W. S. (1982). Color science: concepts and methods, quantitative data and formulae. *New York, NY: John Wiley & Sons.*

# 7 Annexes

## 7.1 Annex 1 (validation metrics)

The validation metrics are used to assed the performance of algorithms and models. They are based on contingency table for binary classification (Tab. A1):

*Table A1) Contingency table from binary classification (source: rebuilding from Silva and Ribeiro, 2009)*

| | Class Positive | Class Negative |
|---|---|---|
| Assigned Positive | TP<br>(True Positive) | FP<br>(False Positive) |
| Assigned Negative | FN<br>(False Negative) | TN<br>(True Negative) |

Where TP positive is something that is assigned as positive and it results in a true assignation; FP is something that is assigned as positive, but it doesn't result in a true assignation; TN is something that is assigned as negative and it results as a true negative assignation; FN is something that is assigned as negative, but it doesn't result as a true negative assignation (Aquino et al., 2017).

Several validation metrics based on the matrix present above, include recall (RC) or sensitivity (SE), precision (PR) or specificity (SP), accuracy (AC), correct classification rate (CCR) and F1 score (F1) which can be calculated using the equations A1-5, respectively (Behroozi-Khazaei and Maleki, 2017; Liu et al., 2017):

$$SE = \frac{TP}{TP + FN} \times 100 \quad \text{(Eq. A1)}$$

$$SP = \frac{TN}{TN + FP} \times 100 \quad \text{(Eq. A2)}$$

$$AC = \frac{TP}{TP + FP} \times 100 \quad \text{(Eq. A3)}$$

$$CCR = \frac{TP+TN}{TP+FP+TN+FN} \times 100 \quad \text{(Eq. A4)}$$

$$F1 = 2 * \frac{SP * SE}{SP + SE} \quad \text{(Eq. A5)}$$

Sensitivity is the proportion of true positives that are correctly identified by the test. Specificity is the proportion of true negatives that are correctly identified by the test. Sensitivity and specificity are one approach to quantifying the diagnostic ability of the test (Altman and Bland, 1994). Accuracy refers to how close a statistic sample is to a population parameter (Stat trek, 2019). Correct classification rate (CCR) is overall accuracy without regard to what type of errors are present (Nisbet et al., 2018). F1 Score, which refers to a measure of good binary classification by considering sensitivity and

specificity together, statistically it is a weighted average of SE and SP and it reaches its best value at 1 and worst at 0 (Liu et al., 2017).

Another important evaluation metric, used to measure the accuracy of an object detector on a dataset, is the Intersection over Union (IoU). Typically, Intersection over Union is used to evaluate the performance of Histogram of oriented gradients (HOG) object detectors and Convolutional Neural Network detectors. Any algorithm that provides predicted bounding boxes as output can be evaluated using IoU (Rosebrock, 2016).

In order to apply IoU to evaluate an (arbitrary) object detector is need the ground-truth bounding boxes (the hand labelled bounding boxes from the testing set that specify where in the image our object is) and the predicted bounding boxes from the model. IoU (Eq. A6) is simply a ratio where in the numerator is computed the area of overlap between the predicted bounding box and the ground-truth bounding box ($A \cap B$) and the denominator is the area of union ($A \cup B$) (in other words, the area encompassed by both the predicted bounding box and the ground-truth bounding box) (Rosebrock, 2016).

$$IoU = \frac{A \cap B}{A \cup B} \qquad \text{(Eq. A6)}$$