

An algorithm to automate the filtering and classifying of 2D LiDAR data for site-specific estimations of canopy height and width in vineyards

Anice Cheraiet, Olivier Naud, Mathilde Carra, Sébastien Codis, Frédéric

Lebeau, James Taylor

▶ To cite this version:

Anice Cheraiet, Olivier Naud, Mathilde Carra, Sébastien Codis, Frédéric Lebeau, et al.. An algorithm to automate the filtering and classifying of 2D LiDAR data for site-specific estimations of canopy height and width in vineyards. Biosystems Engineering, 2020, 200, pp.450-465. 10.1016/j.biosystemseng.2020.10.016 . hal-03110685

HAL Id: hal-03110685 https://hal.inrae.fr/hal-03110685

Submitted on 15 Dec 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

An algorithm to automate the filtering and classifying of 2D LiDAR data for site-specific estimations of canopy height and width in vineyards

- 3 A. Cheraiet^{a, b, *}, O. Naud^a, M. Carra^a, S. Codis^b, F. Lebeau^{a,c} and J.A. Taylor^a.
- 4 aITAP, INRAE, Montpellier SupAgro, University of Montpellier, 361 rue Jean-François Breton,
- 5 34000 Montpellier, France
- 6 ^bIFV French Vine and Wine Institute, 3430 Route de l'Espiguette, 30240 Le Grau-du-Roi, France
- 7 ^cBiosystems Dynamics and Exchanges (BioDynE), TERRA Teaching and Research Center, Gembloux
- 8 Agro-Bio Tech, University of Liege, Gembloux, Belgium
- 9 *Correspondence: <u>anice.cheraiet@vignevin.com</u>
- 10

11 Abstract

12 The 3D characterisation of individual vine canopies with a LiDAR sensor requires point cloud 13 classification. A Bayesian point cloud classification algorithm (BPCC) is proposed that combines an automatic filtering method (AFM) and a classification method based on clustering to process LiDAR 14 data. Data were collected on several grape varieties with two different modes of training. To evaluate 15 16 the quality of the BPCC algorithm and its influence on the estimation of canopy parameters (height and width), it was compared to an expert manual method and to an established semi-automatic 17 research method requiring interactive pre-treatment (PROTOLIDAR). The results showed that the 18 19 AFM filtering was similar to the expert manual method and retained on average 9% more points than 20 the PROTOLIDAR method over the whole growing season. Estimates of vegetation height and width 21 that were obtained from classification of the AFM-filtered LiDAR data were strongly correlated with estimates made by the PROTOLIDAR method ($R^2 = 0.94$ and 0.89, respectively). The classification 22 23 algorithm was most effective if its parameters were permitted to be variable through the season. 24 Optimal values for classification parameters were established for both height and width at different 25 phenological stages. On the whole, the results demonstrated that although the BPCC algorithm 26 operates at a higher level of automation than PROTOLIDAR, the estimates of canopy dimensions in 27 the vineyards were equivalent. BPCC enables the possibility to adjust the spray rate according to local 28 vegetative characteristics in an automated way.

- 29
- 30 Key words

31 Ground-based 3D LiDAR, Point clouds partition, Automatic filtration, Canopy dimensions, Variable

32 rate, Crop protection.

33 Nomenclature

AFM	automatic filtering method
BBCH	describes the phenological development of grapes using the BBCH-scale
BPCC	Bayesian point cloud classification
CCC	correlation coefficient of concordance
CMM	conventional manual measurements
CV-RMSE	coefficient of variation of root mean square error, %
D	distance between the LiDAR travel line and the Line of Trunk, m
D_{ϵ}	distance indicative of the angle range (ϵ) of LiDAR beams intercepted by
	the ground in the grassed zone, m
D_{δ}	distance indicative of the angle range (δ) of LiDAR beams intercepted by
	the ground in the inter-row, m
HEF	human expert filtration
HG	height of grassed zone above ground, m
HS	height of the LiDAR above ground, m

Hsc	height of the start of canopy growth above the ground, m
H_{ϵ}	distance defined along the y-axis between the LiDAR emission point and the first (closest) 5% beams intercented in the grassed zone, m
LiDAR	light detection and ranging
LoT	line of trunks
PPP	plant protection products
vine unit	the area of foliage corresponding to 0.5 m before and after the vine trunk; signifies a standardised individual vine
VH	vegetation height, m
VW	vegetation width, m
$\beta_{\rm H}$	adjustable threshold during the season defined to estimate VH
$\beta_{\rm w}$	adjustable threshold during the season defined to estimate VW
ΔW	distance interval between two consecutive vertical scans, m
$\Delta \theta$	$\Delta \theta$ angular resolution of the scans, degree
θ	angular resolution of the scans, degree
μ	population mean of the points along the y or z axis, m
ρ	radial distance, m
σ	standard deviations of the points along the y or z axis, m
δ	angle range where the LiDAR beams are removed as intercepted by the ground in the inter-row, in degrees
3	angle range where the LiDAR beams are removed as intercepted in the grassed zone, in degrees

34

35 1. Introduction

Over the past two decades, various advances towards more precise and efficient spray systems have 36 37 been proposed for different crops, including vineyards (Siegfried et al., 2007; Walklate & Cross., 38 2013). Although these advances differ in their assumptions and calculations, most of them are based on a characterisation of the canopy. The important factors to consider to ensure an efficient spray 39 40 application process are the geometric characteristics of the canopy (Solanelles et al., 2006; Llorens et 41 al., 2011a) and the relationship between the quantity of plant protection products (PPP) sprayed and 42 the deposits obtained on the foliage, expressed as a quantity per surface area of organs to protect (Gil 43 et al., 2014). As stated by Gil et al. (2013), the risk levels to harm sensitive non-target areas during the 44 spray application process are related to dose rates and will depend on both the total amount of PPP 45 sprayed and the spraying efficiency over the entire canopy. It has been stated that correctly targeting and adjusting deposition to canopy dimensions/structure will lead to a considerable increase in the 46 efficiency of applications (Vercruysse et al., 1999; Gil et al., 2007), thereby reducing the total amount 47 48 of PPP required in accordance with EU objectives (Llorens et al., 2010). This has led to the 49 development of variable rate spraying technologies and methodologies (Gil et al., 2013). These 50 techniques hypothesise that foliar application should target similar deposits per quantity of vegetation 51 to be protected, regardless of the canopy shape or density. In this context, the development of 52 precision spraying technologies that take into account the dimensional characteristics of the canopy to 53 regulate nozzle flow is one of the levers that has been identified to reduce PPPs in perennial crops 54 (Berk et al., 2016).

55

Canopy dimensions can be retrieved manually (Viret et al., 2005; Rosell Polo et al., 2009) or obtained
from sensor measurements (Rosell et al., 2012). Manual measurements are time-consuming and have
limited suitability under production conditions. Using them requires an extrapolation of measurements

59 from a few locations across the entire field, which generally implies some assumptions about the

homogeneity of crop characteristics within a production system. This disregards the knowledge that 60 61 canopy size exhibits spatial variation in vineyard blocks (Tisseyre et al., 2008; Taylor et al., 2013). In 62 order to increase spatial resolution to account for known variability in canopy size, vineyard canopy 63 structure can be indirectly estimated using various types of sensors. The literature includes numerous studies that have characterised vine dimensions from the scale of the estate to the individual vine 64 65 (Rosell et al., 2012; Arnó et al., 2017). Sensors used to date include ultrasonic sensors (Gil et al., 2007; Llorens et al., 2011a), stereo vision imagery (Andersen et al., 2005) including unmanned aerial 66 vehicle (UAV) mounted photogrammetry (Mathews et al., 2013; Miranda et al., 2017; de Castro et al., 67 2018) and 2D terrestrial Laser imaging Detection And Ranging (LiDAR) sensors (Poni et al., 1996; 68 69 Rosell Polo et al., 2009; Siebers et al., 2018).

70

71 The use of laser sensors to digitise the 3D features (or characteristics) of crops (particularly in 72 viticulture) has been established for some decades but is still mainly limited to the research domain. 73 An early attempt to use laser scanning in viticulture was the study by Poni et al. (1996), who used a 2D LiDAR mounted on an arc-shaped structure to simply calculate the light interception of each vine 74 organ (leaves, trunk, cordon etc...). Since this initial work, interest and development in the use of 75 76 LiDAR in vineyards has increased and it is becoming more frequently used to non-destructively characterise vegetation structure, shape and biomass (Colaço et al., 2018; Jaakkola et al., 2010). Using 77 78 LiDAR sensing to measures distances from the sensor to a target over a plane, has a particular interest 79 for the real-time determination of canopy structure during spray operations. 3D scanning is possible when a 2D LiDAR is deployed on a moving platform (Rovira-Más et al., 2006) with a well-80 81 determined method of geo-referencing the LiDAR data. Canopy characterisation using 2D LiDAR has 82 been proposed in vineyard studies (Palacin et al., 2007; Sanz et al., 2018) and 3D point clouds have 83 been used to digitally reconstruct and describe the geometric characteristics of vegetation cover with a 84 high level of accuracy (Moorthy et al., 2011). A system developed by Rosell et al. (2012) made it 85 possible to obtain 3D digitised point clouds of crops, from which a large amount of information, such as height, width, volume, leaf area index and leaf area density, could be obtained for a plant or an area 86 87 of the crop. Arnó et al. (2013) concluded that LiDAR systems were able to measure the geometric characteristics of plants with sufficient precision for most site-specific agriculture applications. 88

89

For high-resolution canopy characterisation, LiDAR systems have an advantage over ultrasonic and 90 stereoscopic imagery approaches because of their ability to provide information on both canopy 91 92 dimensions and density. Ultrasonic sensors were used before LiDAR systems became affordable and 93 available (Schumann et al., 2005), but did not gain widespread popularity. This was due to issues regarding the large angle of divergence of the wave beams (which limits the resolution and accuracy 94 95 of the measurements) (Stajnko et al., 2012), the need for multiple sensors to cover vine and tree crops 96 (Lee et al., 2009) and limitations with the proximity to the crop at which the sensor can be effectively deployed (Llorens et al., 2011a). Recent advances in UAV-based photogrammetry have indicated a 97 98 high potential for their use in mapping canopy shape (de Castro et al., 2018). However, mapping 99 canopy density with stereoscopy is still an issue (Torres-Sánchez et al., 2018), and this is critical for 100 modelling spray deposition and adjusting sprayer operation (Campos et al., 2019). Moreover, UAVbased sensors are also not suitable for real-time applications and require a pre-application survey 101 combined with a prescription mapping approach. While LiDAR systems could equally be used pre-102 spraying to develop prescription spray maps, they also have the potential to be used in front of a 103 sprayer to generate on-the-go, real-time 3D information for variable-rate spraying (Llorens et al., 104 105 2010). In the latter real-time use-case for LiDAR, robust and rapid data processing methods will be required to ensure that correct information is transferred to the spray control system. 106 107

The literature presents different types of vegetative indicators, such as the tree row volume (TRV) 108 (Byers et al., 1971; Sanz et al., 2013) and the leaf wall area (LWA), which can be used to characterise 109 vegetation structure from canopy dimensions. These are measured either manually or with sensors. 110 111 There are high resolution variants of the LWA, such as the pixelated leaf wall area (PLWA) (del-Moral-Martínez et al., 2015) and the leaf wall area by points (LWApts) proposed by Bastianelli et al. 112 113 (2017). When LWA is constant, PLWA and LWApts may exhibit variations due to changes in canopy density. The tree area index (TAI) (Walklate et al., 2002) is based on the notion of light interception 114 and integrates both canopy density and variations of geometry surface area density (SAD) (Schultz, 115 1995). All of these vegetative indicators aim to simplify the complex structure of vegetation by 116 117 describing it as a simple geometrical form, sometimes with a feature representing density. However, 118 before these indicators can be calculated from sensor-based data, different processes are required to 119 obtain the primary canopy dimensions from these data.

120

The first challenge is to obtain a complete 3D point cloud of the entire canopy. Typically, this has 121 required the merging of data collected from the left and right sides of the vineyard (or orchard) row at 122 potentially different times, i.e. during different transects (Sanz et al., 2004). Various tedious and 123 124 difficult methodologies have been proposed, such as placing reference elements at specific points in the row that can be identified within the canopy point cloud. This complicates data management 125 (Rosell et al., 2009; Sanz et al., 2013). Subsequently, other developments have improved this process 126 with the coupling of global navigation satellite system (GNSS) positioning (Llorens et al., 2011b; 127 Escolà et al., 2017) and inertial measurement units (IMUs) (del-Moral-Martínez et al., 2016). 128 129 However, GNSS and IMUs both require high quality, expensive specialised equipment. This increases the cost and the processing required and affects the transferability of the research methods into 130 commercial applications. Furthermore, obtaining scans of both sides of the canopy requires sensors to 131 be deployed in every vineyard row. While this has been possible to date in research-based studies, the 132 reality of agricultural practices is that vineyard traffic is usually only every second or third row 133 depending on equipment configuration. It is more likely that only one side of the canopy (a half-134 canopy scan) will be sensed during any single vineyard operation. This remains problematic, as 135 approaches to estimate canopy dimensions from 'half-canopy' (one-side) LiDAR scans, and their 136 137 accuracy, have not yet been well-developed.

138

139 The second challenge is the filtering procedure of the 3D point cloud. Given the large number of beams emitted by a 2D LiDAR, the selection and classification of "points of interest" becomes an 140 important pre-processing task before canopy dimensions and vegetative indicators can be calculated. A 141 significant number of points are intercepted in regions that are not relevant for the calculation of 142 vegetative indicators, such as the ground, grassed areas, the vine trunk, adjacent rows or the trellis 143 wires (Bastianelli et al., 2017). However, in the available literature on applications of mobile 2D 144 LiDAR in vineyards and orchards, there are very limited explanations and details on the procedures 145 146 for filtering 3D point clouds. In many studies, the goal was to establish the proof of concept and the 147 data filtering was performed with intensive human intervention (Palacin et al., 2007; Rinaldi et al., 148 2013). This laborious human intervention at the pre-processing step is not practical if LiDAR is to be deployed in production contexts. Rapid, repeatable, robust filtering methods are needed to ensure the 149 correct estimation of simple vegetative parameters, such as vegetation height or width. These methods 150 151 need to be effective at all stages of canopy development, from small open canopies during early shoot development to large, potentially dense canopies late in the season. Research methods developed and 152 used to date have tended to focus on filtering and pre-processing data obtained at specific growth 153 stages, not collectively across all growth stages. If LiDAR, or any other sensing technology, is to be 154 155 successfully incorporated into variable-rate PPP spraying regimes, the technology must be effective

- across a wide range of canopy sizes and adaptable to changing canopy conditions. Arguably, the most
- 157 important period for applying PPP is when the canopy size and shape is rapidly developing during 158 early to mid-season shoot growth. The need for rapid and robust filtering of these large 3D datasets
- early to mid-season shoot growth. The need for rapid and robust filtering of these large 3D datasets will become even more critical when real-time processing is required for on-the-go applications in
- 160 spatially variable canopy systems.
- 161
- 162 The research presented here aims to address these issues of half-canopy scans and an evolving canopy 163 structure by proposing and testing a novel method for the automated pre-processing and filtering of 164 LiDAR data. The method was designed to remain effective as canopy size and shape change quickly 165 through the first half of the growing season and to be applicable in commercial agricultural situations.
- 166 The specific objectives of this work were to:
- (1) propose an adaptable algorithm that applies an automatic filtering method to remove artefacts and
 non-vine data from 2D LiDAR data collected from only one side of the vine canopy, and then
 classifies and separates the canopy zone from other vine components (trunk, vegetation, trellis wires)
 without any operator intervention,
- 171 (2) use the proposed algorithm to estimate canopy height and width from LiDAR surveys in several
- 172 vineyard blocks in southern France and,
- (3) assess the quality of these estimations of canopy dimensions by comparing them to canopydimensions derived from an existing standard LiDAR data filtering method, which is not automated
- and requires human intervention, and to conventional manual canopy measurements.
- 176

2. Materials and methods

178 2.1. Fields trials

179 A vineyard with four different blocks ("Les pins", "Aglae", "Terre blanche" and "Franquet") of four 180 different varieties of Vitis vinifera L. cv (Marselan, Cabernet Sauvignon, Chardonnay and Petit 181 Verdot), with contrasting vigour, was chosen for the study in 2019. Located in Grabels, close to Montpellier (Hérault, France), the study vineyard is characteristic of a vineyard from the south of 182 France, both in terms of grape varieties and training systems. The rows were north-south oriented for 183 184 "Les pins", and northeast-southwest for "Aglae", "Terre blanche" and "Franquet". Two different 185 training systems were used: Royat cordon for "Les pins", "Aglae" and "Franquet" and Guyot for 186 "Terre Blanche". Vines were trained (one carrying wire and one trellising wire) in all blocks. Rows were separated by distance D_{ir}, with D_{ir} equal to 2.5 m in all blocks and vine spacing in the row was 1 187 188 m. For each block, 20 vines were selected and their trunks geolocated with a LEICA Viva GS10 dual-189 frequency GNSS receiver equipped with a Siemens MC75 GSM/GPRS individual module, triple-190 frequency antennas (GPS/GLONASS/Galileo) LEICA AS10 and CS10 radio controls. The same vines were followed throughout the season. 2D LiDAR and manual characterisation of vegetation were 191 192 carried out on seven dates during the season (T1: 2019/04/29, T2: 2019/05/13, T3: 2019/05/21, T4: 193 2019/05/28, T5: 2019/06/20, T6: 2019/07/18, T7: 2019/07/31). These dates correspond respectively to the following BBCH scale growth stages (Lorenz et al., 1994): three leaves spread out (14), four to six 194 leaves spread out (53), separate flower buds (57), beginning of flowering (61), flowering (70), berry 195 196 development (76), bunch closure (81).

197 198

2.2. Measurement system

199 2.2.1. Conventional manual measurement (CMM)

Two different canopy parameters were manually measured at each vine: canopy height (m) and canopy width (m). Manual observations were performed according to the protocol of Manktelow and Praat (1997). Briefly, canopy height was defined from the first leaf above the trunk to the highest leaf in the canopy in the area above the vine trunk. For canopy width, the canopy zone was divided into three equal vertical sections and a measurement made horizontally in each section between the external canopy leaves with a 2 m ruler. The three measurements were averaged to give the mean canopy width. Each measurement aimed to include > 99 % of the canopy (i.e. some protruding branches were ignored).

208

209 2.2.2. LiDAR sensor specifications

210 Data acquisition unit

211 A Sick LMS100 (SICK AG, Düsseldorf, Germany) 2D LiDAR sensor was used in the study. The 212 LMS100 LiDAR is a fully-automatic divergent laser scanner based on time-of-flight (TOF) measurement with a systematic error of \pm 30 mm, a selectable angular resolution ($\Delta\theta$) set to 0.5° and a 213 214 range of 270°. With these settings, there were 541 distances (ρ , from the sensor to the interception point) that corresponded to one complete laser mirror rotation. This set of 541 distances is called a 215 "scan" throughout the article and scans were repeated at 50 Hz. The Sick LMS100 laser emission 216 wavelength is 905 nm (near infrared) and it is Class one eye-safe. This sensor was coupled to a Real 217 218 Time Kinematic (RTK) GNSS receiver (Teria GSM correction, Vitry-sur-Seine, France) and an 219 Effibox data acquisition unit (Effidence society, Romagnat, France) that was used as a data-logger. 220 After surveys, the data were transferred to a laptop over a Wi-Fi network. The sensors were mounted 221 on a dedicated stainless-steel mast placed behind a tractor according to a previously described procedure (Bastianelli et al., 2017) at a height ranging from 1.0 m to 1.40 m above the ground level 222 223 (HS). Height was adjusted up during the season to account for increasing canopy height (Figure 1A). Collectively the sensors and mobile equipment provided a 3D measurement system. 224

225

The tractor was driven along the vineyard rows at a constant forward travel speed (FTS) (Figure 1B) of 5 km h⁻¹, with a systematic error of \pm 0.21 km h⁻¹ (IFV, internal report, October, 2018). The 20 target vines were located in various locations along the vineyard rows. The RTK-GNSS was used to identify the starting point of these 20 target vines, after which the scans were aggregated, using a fixed forward distance based on the constant tractor speed, to generate a 3D point cloud reconstruction of the vine environment (Figure 1B). During the trials, only one side of the canopy was scanned for each vineyard row.

233 234

2.3. Vine unit local 3D point cloud construction

235 A vine unit, corresponding to an individual vine, was defined according to the direction of travel (x), 236 considering 0.5 m before and 0.5 m after the vine trunk centre (Figure 1B). Vertical scans of the vine canopy were obtained from the 2D LiDAR. Each scan was composed of distances between the LiDAR 237 and objects in the path of the laser beam. The coordinate system origin (O) was defined as the first 238 position of the 2D LiDAR during the measurement on the studied vine unit. The time stamp t (in 239 240 seconds) was given by the Effibox acquisition unit. The distance interval between two consecutive 241 vertical scans (ΔW) was 0.028 m along the direction of travel of the tractor (Figure 1B). For each 242 point of the cloud, the x coordinate was calculated by multiplying t by the travel speed. The y and z 243 coordinates (informing on canopy height and width respectively) were obtained by a polar (ρ , θ) 244 (Figure 1A) to Cartesian (y, z) coordinates transformation. Therefore, the 3D point cloud of the vine unit was generated within a Cartesian coordinate system. 245

Figure 1 near here

246

247

248

249

250 251

2.4. Filter algorithms and LiDAR data analysis

252 In this section, the methodologies of the two approaches to be compared are presented. The first, 253 PROTOLIDAR, is considered here as a standard approach. It requires human intervention and is based

- on work by Rinaldi et al. (2013). The second is the novel algorithm BPCC. 254
- 255

2.4.1. PROTOLIDAR methodology 256

257 The data files were analysed using the open source statistical software R (Version 1.2.5001) (R Development Core Team, 2019) and the PROTOLIDAR package (PROcess TO LIDAR Data) 258 (Rinaldi et al., 2013). PROTOLIDAR contains three functions to characterise the vine canopy (height, 259 width and front view) from the LiDAR point cloud. The tool performs statistical analysis on the 260 outputs and estimates the leaf area index (LAI), LWA and TRV. For the pre-processing (filtering), the 261 methodology described in Rinaldi et al. (2013) was used. The 3D point cloud was trimmed using the 262 Extract plant grapevine function with manually defined thresholds, leaving only the area of interest 263 (i.e. the canopy). This function removed areas of the 3D point cloud that were not associated with the 264 canopy, including LiDAR returns from ground and under vine weeds as well as vines in neighbouring 265 266 rows.

- Once the data had been filtered to a canopy-only response, the PROTOLIDAR package allowed user-267 defined parameters to be set to characterise the vegetation. The functions Width_canopy and 268 Height_canopy permit the characterisation of vegetation height (VH) and vegetation width (VW) 269 respectively. The minimum possible height was defined manually as the cordon height. VH was 270 estimated from the lowest registered point of canopy LiDAR returns above the defined cordon height 271 (denoted as Height start canopy (Hsc)) to the highest registered point of canopy returns along the y-272 273 axis. VW was estimated using the same methodology as for VH, but by considering points along the zaxis.
- 274 275

2.4.2. LiDAR Bayesian point cloud classification algorithm (BPCC) 276

The BPCC is a 2-stage algorithm. It comprises an automatic filter to remove points of non-interest and 277 a hierarchical cluster-based method to derive canopy dimensions. The two stages are presented in their 278 279 respective subsections.

280

2.4.2.1. Automatic filter method (AFM) 281

282 As the 2D LiDAR sensor scans the entire vineyard, not just the vine canopy, points that belong to the 283 canopy must be automatically identified and distinguished from points associated with other elements 284 (ground, non-vine vegetation, etc...). This pre-processing is critical to estimate canopy dimensions (height and width) as accurately as possible. The filtering of the raw data was carried out using 4 285 functions that eliminate LiDAR returns from areas of non-interest associated with: (1) the ground in 286 the inter-row, (2) adjacent rows, (3) undervine and inter-row vegetation (weeds) and (4) obstacles too 287 288 close to the sensor to be canopy.

289

290 (1) Inter-row ground filtering: beams intercepted by the ground in the inter-row must be removed from 291 the raw data. It is assumed that this zone corresponds to half of the distance from the sensor to the line 292 of trunks and equates to a distance D_{δ} of 0.625 m in these vineyards (Figure 2). Depending on the 293 height of the LiDAR (HS), the beams in the interval $[0; \delta]$ are removed. D_{δ} is not fixed and should be adjusted for changes in row width and canopy vigour and shape if transferred to other production 294 295 systems. The value of the angle δ is calculated as follows:

296
$$\delta = atan\left(\frac{D_{\delta}}{HS}\right)$$

298 (2) Filtering of adjacent rows: in a first pass, points intercepted more than two rows away (> 8 m) from 299 the LiDAR sensor were removed from the raw data. Then, assuming that the tractor has a straight 300 trajectory centred in the inter-row, with a systematic error of ± 0.035 m, the distance from the centre of 301 the inter-row to the trunk line (LoT) can be used to identify and delete points associated with the 302 opposite side of the canopy or adjacent rows. The filter value (D) is therefore half the row width (D_{ir}).

- 303
- $D = \left(\frac{D_{ir}}{2}\right)$

305 (3) Grassed zone filtering: vegetation present under the vine or in the inter-row must be removed from 306 the raw data to avoid its inclusion in the calculation of the vegetative parameters. For this purpose, the 307 height of the grassed zone (HG) could be set as a constant threshold, which would need to be adjusted 308 between systems, or alternatively derived from the LiDAR data, so that it is automated. In the latter 309 case, HG can be derived under the assumption that there is only grass below the LiDAR sensor and 310 that HS is known. In this case, a distance for filtering the grassed zone (D_{ϵ}) can be defined as:

312 $D_{\varepsilon} = D - D_{\delta}$

313 The beams are removed at the angles ε in the interval that is considered as the grassed zone (Figure 2). 314 The value of the ε angle is defined as:

315

311

316
$$\varepsilon = atan\left(\frac{D_{\varepsilon}}{HS}\right)$$

Subsequently, a distance H_{ϵ} can be calculated as the average distance of the nearest 5% of intercepted points to the LiDAR emission point in the grassed zone, as defined by the angle ϵ . The 5% threshold was based on previous unpublished research using this setup. HG is therefore defined as:

320 321

 $HG = HS - H_s$

This filtration threshold (HG) will evolve during the season according to the acquisition date and the characteristics of the ground cover in the blocks (Figure 2). The angle ε may need to be altered in vineyards with differing canopy and ground cover conditions to those in southern France. The relative importance of this filter will vary depending on how precisely ground cover in the vineyard is managed.

327

(4) Near point filtering: beams intercepted at a distance too close to the 2D LiDAR to be canopy need 328 329 to be removed from the raw data. Most of these are likely to be associated with large insects or 330 random, untrained or broken shoots. The filter value was set at a constant 0.5 m, which was based on 331 the operating range of the 2D LiDAR sensor and the expectation that the canopy is vertically trained 332 (Figure 2). This fixed threshold will again need to be adapted when transferred into vineyards with 333 different training and trellising modes; however, once determined, it should be a fixed value to automate this filtering process. In practice, this filtering represents a tiny fraction of point removal by 334 335 AFM.

336

Before the application of these four filters, no pre-processing or filtering was applied to the raw point
clouds. Generic parameters based on vineyard characteristics were set and all the above defined filters
were applied automatically. At the end of this step, it was possible to separate the intercepted points

340 into two categories (Figure 2): (1) points intercepted outside the zone of interest (in blue) that have

been eliminated, and (2) points intercepted in the zone of interest (trunk, vegetation and trellis wire)(in red).

Figure 2 near here

- 343
- 344
- 345
- 346
- 347

348 2.4.2.2. Clustering methodology

The determination of canopy dimensions from the pre-processed LiDAR data consisted of two parts; (a) a 1D cluster analysis based on the vegetation height from the LiDAR point clouds to identify different components of the vine and trellis, followed by a Bayes classification, and (b) a statistical test (thresholding) to delimit two dimensional parameters (vegetation height and vegetation width), associated with the canopy area defined from the classification process.

354

355 (a) 1D hierarchical cluster and Bayes classification

The LiDAR point cloud expressed the canopy information in a 3D space. Points therefore 356 357 corresponded to heterogeneous distributions, like multivariate clusters, of discrete objects within the 358 sample space according to their positioning. Field observations suggested the presence of at least three different 'groups' within the general area of interest for the canopy. A "low" group associated with 359 LiDAR returns from the trunk and low-hanging or poorly placed shoots; a "high" group, particularly 360 early in the season, associated with LiDAR returns from trellis wires and infrastructure; and a 361 362 "transition" or central group associated with LiDAR returns from the canopy (Figures. 3A, 3B and 3C). The spatial location (on a vertical axis) of the high and low groups is static, as the trellis wires 363 and vine trunks are fixed. It is predominantly the transition group associated with the canopy that is 364 dynamic and changing as the season progresses. As the vine grows, the transition group will merge 365 with the high group and obstruct the trellising wires. 366

367 Hierarchical cluster analysis was performed to determine if there were two or three unique 368 combinations of Gaussian distributions along the height axis. Given the expected overlapping Gaussian distributions of the 2D LiDAR groups, a hierarchical clustering algorithm based on a 369 370 Gaussian mixing model (Fraley et al., 2007) was used. Hierarchical clustering defines classes by 371 grouping the most similar observations in a hierarchical fashion and is based on functions that 372 combine model-based hierarchical clustering (expectation-maximisation) and the Bayesian 373 Information Criterion (BIC). The clustering was conducted using the mclust package (Fraley et al., 374 2012) in R.

375 Once the points in the point cloud had been clustered and points associated with (or likely to be 376 associated with) the canopy had been identified, the canopy dimensions were calculated. The canopy 377 point cloud will follow a Gaussian distribution (Figures. 3A, 3B and 3C). Therefore, a choice must be made on which values of this distribution should be used to determine canopy dimensions. In the 378 standard approach of Rinaldi et al. (2013), extreme values were used for width and for the maximum 379 380 height, while the minimum height (Hsc) was defined manually as the cordon height. However, in the 381 case of an automated system, as proposed here, this may not be sensible as some outlying values may 382 be retained and will unduly influence the dimension calculations. A sensitivity analysis on the choice of a statistical threshold for defining the vegetation height and width was carried out. The distribution 383 of the points along the y axis were filtered based on standard deviations (σ_H) from the population mean 384 (μ_H). The thresholds were established as follows: $\mu_H +/- (\beta_H * \sigma_H)$ with β_H a parameter. Candidate 385 386 values for $\beta_{\rm H}$ were selected as follows: 0.5; 1; 1.5; 2; 2.5; 3. For each set of data corresponding to the 387 same phenological stage (from T1 to T7) and for each $\beta_{\rm H}$ values (6 in total), the absolute error

(expressed in m) between the dimensions of the vegetation canopy measured manually by an operator 388 389 in the field and estimated by the clustering method was calculated for the vegetation. Any y-values that were not in the respective interval were excluded from the analysis. The evolution of the absolute 390 391 error according to the selected candidate $\beta_{\rm H}$ values allowed the identification of the optimum $\beta_{\rm H}$ value that minimised the absolute error (vs. manual measurement). Thus, a set of seven phenology 392 393 dependent $\beta_{\rm H}$ thresholds were defined that covered the whole season.

394

395 With regards to the estimation of canopy width, a similar sensitivity analysis was carried out in the z 396 axis using the methodology described above for canopy height (y axis). However, as only one side of 397 the vine was scanned, a symmetry hypothesis was used based on the observations of Arnó et al. (2015) to estimate full canopy width from half canopy width. The half width of canopy was estimated using 398 the line of trunks (LoT) as the upper limit. The z-values of the point cloud followed an exponential 399 distribution over the interval [$\mu_w - (\beta_w * \sigma_w)$; LoT]. The lower thresholds were established as follows: 400 401 $\mu_w - (\beta_w * \sigma_w)$ (with μ_w and σ_w respectively the mean and standard deviation of the z values of the 402 points defined in the foliar zone). For each phenological stage and for each β_w value, the absolute error 403 was calculated (expressed in m) between the manually measured full canopy width and the canopy width estimated by the clustering method. Calculating this over a range values (0.5; 1; 1.5; 2; 2.5; 3) 404 405 allowed an adjustable threshold (β_w) to be defined at each observed phenological stage along the 406 season that minimised the absolute error against the manual measurements.

407

408 The parameterisation of the adjustable (temporal) threshold for defining the canopy zone was carried 409 out for different phenological stages. This is needed because (1) the number of groups defined by the algorithm changes with vine development, decreasing from three at the start of the season to two 410 groups by mid/late-season once the trellising wire is covered by foliage and (2) the geometry of the 411 412 canopy evolves with vine management operations, such as lifting, trimming and topping, that are 413 linked to phenological development, and they have a potential impact on canopy dimensions.

414

415 (b) Estimation of canopy height and width based on adaptive thresholding in the canopy zone Given the preferred β_H and β_w values for the vine phenological stage, canopy dimensions in the y and z 416 417 axes can be derived from the 2D LiDAR points classed as the canopy zone. Height and width were calculated for each vine unit. 418

- 419
- 420 Vegetation height (VH) in m was defined as:
- 421

422 $VH = (\mu_H + (\beta_H * \sigma_H)) - (\mu_H - (\beta_H * \sigma_H))$

423 where $\beta_{\rm H}$ is the threshold identified from the sensitivity analysis for a particular phenological stage; $\mu_{\rm H}$ 424 and $\sigma_{\rm H}$ are respectively the mean and standard deviation of the y values defined in the canopy zone 425

- 426 Vegetation width (VW) in metre (m) was derived from a LiDAR scan of only one side of the vineyard row. Therefore, VW was calculated as double the width of one side and defined as: 427
- 428 429

 $VW = (D - (\mu_w - (\beta_w * \sigma_w)) * 2$

where D is the distance between the LiDAR travel line and the LoT (in m), βw is the threshold 430 431 identified from the sensitivity analysis for a particular phenological stage and μ_w and σ_w are 432 respectively the mean and standard deviation of the z values defined in the canopy zone.

- 433
- 434 2.5. Error assessment of result

435 2.5.1. Quantitative comparison of filter methods

436 In order to compare the effect of the AFM (first step in the BPCC) and the PROTOLIDAR package on the raw point cloud data, both approaches were compared with an intensive expert classification of the 437 438 entire point cloud, which is termed a "Human Expert Filtration" (HEF) approach. The HEF consisted of manually tagging all the intercepted points and using the expert's knowledge to classify each 439 440 LiDAR return into a group (inter-row ground, adjacent rows, grassed zone, near point or canopy). This 441 was a very laborious process and was only performed on a few vine units at different phenological stages to illustrate and compare how the three different filtering methods were performing. To describe 442 the differences between the HEF, the PROTOLIDAR and AFM methodologies, a distribution of the 443 444 intercepted points in the four groups defined by the applied filters was studied and a comparison of the 445 percentage of the points retained to calculate vegetative parameters after the filtration steps was 446 performed on two 3D LiDAR point clouds from an acquisition made on three vines (Vitis vinifera L. 447 cv. Marselan) at three different stages - BBCH 14, 57 and 76 (Table 1). It should be noted that the PROTOLIDAR method is a global and non-specific method for filtering intercepted LiDAR points. It 448 449 did not offer the possibility to class the filtered points according to groups (inter-row ground, grassed zone, etc). Consequently, only total data removed, and not associated groupings, are reported for the 450 451 PROTOLIDAR method.

452

453 2.5.2. Sensitivity study on thresholding in the clustering methodology

454 A sensitivity analysis was used to select $\beta_{\rm H}$ and $\beta_{\rm w}$ thresholds that minimised the absolute difference with manual measurement of canopy dimensions. To evaluate the accuracy and precision of the 455 456 automatic clustering method within the sensitivity analysis, several statistical tests were performed on 457 the absolute errors measured on the 560 vine units. To test the accuracy between the different methods, an ANOVA test was performed on the absolute error values per vine unit by aggregating the 458 data for each phenological stage (n = 80 vine units) and significant differences between the groups 459 460 determined by a Tukey Honest Significance Difference (Tukey-HSD) post-hoc test. The variance of the absolute error is a measure of the precision of the method, with a low variance indicating a high 461 462 precision. To test for differences in the variance between groups, a pairwise test was done using Bartlett's test ($\alpha = 0.05$) and p values were adjusted using the Bonferroni method (Westfall et al., 463 464 1997).

465

466 2.5.3. Comparison of derived canopy dimensions between the PROTOLIDAR - BPCC - CMM 467 methodologies

The coefficient of variation of root mean square error (CV-RMSE) and the correlation coefficient of concordance (CCC; Lin et al. 1989) were calculated to evaluate the quality and concordance of the estimations of canopy height and width between the established methods (PROTOLIDAR and manual observations) and the new method (BPCC) for all vine units over the entire season. The R² was used to evaluate the fit of these regressions.

473

474 **3. Results and discussions**

475 3.1. Quantitative comparison of filtering methods

Table 1 shows the number of points removed and the percentage of points retained using the three different filtration methods for three vine units in one vineyard (*Vitis vinifera* L. cv. Marselan), representing three of the phenological stages measured (14, 57 and 76). The results in Table 1 are presented to illustrate the behaviour of the three filters, not to present a complete analysis over all 560 vines. The HEF approach is slow and laborious and could not be performed on all vines. It can be observed that the total number of points intercepted increased throughout the growing season, as the canopy size increased. The percentage of points retained differed between the PROTOLIDAR and 483 AFM (Table 1), with more points preserved with the AFM regardless of the phenological stage. There 484 were respectively 2%, 9% and 12% more points preserved with the AFM at BBCH 14, 57 and 76. Although the HEF method retained the highest percentage of points for all three phenological stages, it 485 retained on average only 1.3 % more points than the AFM. Overall the removal rate was 486 approximately 50 - 60% of the data for the three filters and three stages (Table 1). This is expected 487 488 considering the wide scan angle relative to the canopy area that results in a large amount of data being collected from areas of non-interest. The similarity between HEF and AFM in these three vine units 489 indicated that AFM retained a sensible level of information for subsequent analysis. Overall, the AFM 490 method mimicked the expert approach more closely and retained a larger percentage of data to carry 491 492 through to the next stage than the PROTOLIDAR method.

493

At BBCH 14, there was little difference between the AFM and PROTOLIDAR methodologies (Table 1). This can be explained by the almost non-existent grassed zone that limited errors when classifying the canopy zone. Additionally, at this growth stage, vegetation was sparse permitting the LiDAR laser beams to penetrate the inner surface of the canopy. There was no shadowing effect, allowing the PROTOLIDAR method to estimate the canopy area with high precision. However, at more advanced growth stages (BBCH 57 and 76) there was a significant difference between the AFM and PROTOLIDAR filtration methods with the PROTOLIDAR filter removing more points.

501

The PROTOLIDAR methodology only relies on this filtering step to eliminate erroneous data and to 502 503 define the canopy zone. Thus, the accuracy of the filtering directly affects the accuracy of the 504 PROTOLIDAR estimations of canopy dimensions. In addition, it should be noted that the 505 PROTOLIDAR method is based on a hypothesis of total propagation of LiDAR laser beams through the vegetation to define the canopy width. However, with high-density canopies, the LiDAR beams 506 507 cannot penetrate deeply in or through the canopy. Instead the majority of the LiDAR returns are from the outer surface of the canopy. This is a "shadowing effect" and reduces the amount of information 508 related to the inner surface of the canopy. Returns from the far-side of dense canopies are very limited, 509 510 reducing the precision of canopy width estimations. This shadowing effect is one of the main drawbacks of the laser measurement system (Van der Zande et al., 2006). 511

512

513 While Table 1 provides an example comparison including the HEF approach, the filtered points from 514 AFM and PROTOLIDAR methods were calculated and compared over the entire vegetation season from the 560 vine units. From this global analysis, the AFM method retained on average 9% more 515 points than the PROTOLIDAR method for calculating canopy dimensions (data not shown). However, 516 it should be noted that the performance of PROTOLIDAR filtering may be affected by imperfect 517 adjustment of filtering parameters. Although a considerable amount of time was spent manually 518 519 optimising the filter parameter settings, there may be a better set of parameters that could have been used for the 560 vine units used for comparison. 520

Table 1. Points deleted from three different filtration methods, for three vine units in one vineyard (*Vitis vinifera* L. cv. Marselan) at BBCH stages 14, 57 and 76 to illustrate the differences between the automatic filtration method (AFM), the human expert filtration (HEF) and the PROTOLIDAR methodologies. Deleted points are classed according to the four groups defined in the AFM. The percentage of points retained by each filter at each stage indicates the data that are available to be used in the derivation of canopy dimensions post-filtration.

528

Methodology	Number of deleted points after filtration step			Percentage of points				
			retained after filtration step					
	Inter-row	Grassed	Adjacent	Near				
	ground	zone	rows	points				
BBCH 14 - total number of intercepted points in the vine unit = 2732								
HEF	450	154	966	0	43			
PROTOLIDAR	$\mathbf{N}\mathbf{A}^\dagger$	$\mathbf{N}\mathbf{A}^\dagger$	$\mathbf{N}\mathbf{A}^\dagger$	$\mathbf{N}\mathbf{A}^{\dagger}$	39			
AFM	470	149	986	0	41			
BBCH 57 - total number of intercepted points in the vine unit = 4136								
HEF	469	34	48					
PROTOLIDAR	$\mathbf{N}\mathbf{A}^\dagger$	$\mathbf{N}\mathbf{A}^\dagger$	$\mathbf{N}\mathbf{A}^{\dagger}$	$\mathbf{N}\mathbf{A}^{\dagger}$	37			
AFM	479	377	1335	34	46			
BBCH 76 - total number of intercepted points in the vine unit = 5014								
HEF	542	489	1345	0	52			
PROTOLIDAR	$\mathbf{N}\mathbf{A}^\dagger$	$\mathbf{N}\mathbf{A}^\dagger$	$\mathbf{N}\mathbf{A}^\dagger$	$\mathbf{N}\mathbf{A}^{\dagger}$	39			
AFM	562	501	1369	0	51			

529 [†] The PROTOLIDAR filter did not permit classification of deleted points into groups

530

531 3.2. Sensitivity study on thresholding in the clustering methodology

This analysis concerned the influence of the statistical thresholds, $\beta_{\rm H}$ and $\beta_{\rm w}$, on the calculation of 532 vegetation height (VH) and width (VW) in the clustering method. The evolution of the absolute errors 533 534 (in m) between canopy height and width from CMM and VH and VW are shown in Tables 2 and 3 respectively. For both VH and VW, the values of β_H and β_w that minimised the absolute error when 535 compared with CMM were not constant during the growing season. Table 2 shows that for the 536 phenological stages BBCH 14, 53 and 57, a value of 2 for $\beta_{\rm H}$ (equivalent to $\pm 2 \sigma$ or the retention of 95 537 538 % of the data) gave the lowest absolute error to define the canopy height from the BPCC filtered LiDAR data. However, for phenological stages BBCH 61 and 81, a value of 3 ($\pm 3 \sigma$ or retention of 539 540 99.7 % of the data) minimised the absolute error relative to CMM. This is explained by the elongation and the lifting of shoots between BBCH 57 and 61 that moves the canopy into and above the upper 541 542 trellis wire. Before this, there is a clear third (high) group associated with LiDAR returns from the 543 upper trellis wire that needs to be considered in the determination of canopy dimensions.

544

The sensitivity analysis associated with β_w indicated a preferred value of 2 for the earliest and latest observed phenological stages (BBCH 14 and 81), and a value of 3 for all other stages in order to reduce the absolute error. Thus a more severe trimming (lower β_w value) is needed early and late in the season. No clear reason was found for this empirical result. It may be associated with a less dense canopy (more open foliage) at both these phenological stages that is associated respectively with earlyseason leaf/shoot expansion and late-season leaf senescence.

551

The best performed thresholds ($\beta_{\rm H}$ and $\beta_{\rm w}$) at each phenological stage are shown in bold in Tables 2 and 3. The absolute errors between canopy height and width from the PROTOLIDAR and CMM at 554 each phenological stage are also shown. This permits an indirect comparison of absolute error for the BPCC and PROTOLIDAR methods (both relative to CMM). For canopy height, the optimised BPCC 555 method outperformed (lower absolute error) the PROTOLIDAR at all phenological stages, with the 556 difference in absolute error rising from 0.08 m early in the season to 0.16 m late in the season (Table 557 2). For canopy width, the response was different. Earlier in the season (BBCH 14), when the 558 559 vegetation was not very dense, the absolute error associated with the PROTOLIDAR method was less than the BPCC method (0.11 m vs 0.17 m respectively) (Table 3). However, as the canopy developed 560 and the vegetation became denser, the optimised BPCC estimated canopy width with less absolute 561 error than PROTOLIDAR, with an average difference ≥ 0.16 m from BBCH 61 onwards (Table 3). 562 563 This can be explained by the increasing influence of shadowing effects in the LiDAR data as the canopy develops. The PROTOLIDAR depends on LiDAR returns from the distal part of the canopy to 564 estimate canopy width. With larger, denser canopies, these returns are greatly reduced, generating less 565 certainty in the shape of the distal part of the canopy and therefore more error in canopy width 566 567 estimation. Under these conditions, it appears that an estimation of canopy width based on half-row LiDAR scans and the assumption of a symmetrical canopy structure is more accurate. The 568 PROTOLIDAR method was developed under the assumption that a good quality 3D point cloud of the 569 570 canopy is available, i.e. scanned from both sides in the case of larger canopies. It is not surprising that the absolute error with the PROTOLIDAR canopy width estimations increases overtime with canopy 571 development. However, as noted in the introduction, a clear need for the industry is to have LiDAR 572 processing systems that can operate with half-row scans. The BPCC method permitted estimations of 573 the height and width of individual vine canopies (i.e. a site-specific estimation) with absolute errors < 574 575 0.2 m in both height and width at any point throughout the season and < 0.15 m at growth stages up to 576 and including flowering (BBCH 61). The exception to this was canopy width estimations very early in the season (BBCHH 14) with BPCC (absolute error compared to CMM was 0.17 m). However, at this 577 578 stage, the canopy is still small, shoots can still be randomly organised thereby generating measurement 579 or scanning anomalies, and issues with PPP coverage are unlikely in small open canopies. Therefore, 580 this result was not considered detrimental to the potential adoption of the BPCC.

581

582 The sensitivity analysis of both $\beta_{\rm H}$ and $\beta_{\rm w}$ indicated that a dynamic threshold value is preferable for 583 calculating vegetation height and width with the clustering method. The optimum threshold can be 584 associated with the management and the architecture of the vine, that itself can be modelled or sensed, 585 enabling the threshold to be programmed in the clustering method based on vine management and phenology. This makes the BPCC less subject than PROTOLIDAR to operator interpretation for 586 calculating VW, particularly at the beginning and end of the season (Table 3). It should be noted that 587 the absolute error was highest towards the end of the season, a period when, typically, PPP are applied 588 less frequently in vinevards. In this study, the parameterisation of the number of clusters defined by 589 the algorithm was performed in a supervised mode. In future developments, a statistical test could be 590 used to support the automatic determination of the number of clusters to be defined. This would be 591 592 important in vineyards at mid-season when there is potentially a clear difference between high and low 593 vigour areas in a vineyard in regards to the location of shoots relative to the upper trellis wire.

Table 2. Pairwise differences of the means of absolute errors (in m) for the vegetation height parameter of the conventional manual measurement (CMM) and PROTOLIDAR and clustering methodologies, grouped by BBCH stage class with *p* value from ANOVA. Variances between groups that differed significantly using Bartlett's test with the Bonferroni adjustment are in italics. Considering the clustering method, for each BBCH stage, the $\beta_{\rm H}$ threshold that minimised the absolute mean error rate with CMM is in bold.

601

Stage BBCH	clustering $\beta_{\rm H} = 1$	clustering $\beta_{\rm H}$ =2	clustering $\beta_{\rm H} = 3$	PROTOLIDAR	p-value
14	0.14	0.08	0.21	0.12	<i>p</i> ≺ 0.001
53	0.14	0.09	0.24	0.14	<i>p</i> ≺ 0.001
57	0.18	0.11	0.33	0.17	<i>p</i> = 0.017
61	0.24	0.21	0.12	0.19	<i>p</i> = 0.038
70	0.25	0.23	0.13	0.21	<i>p</i> = 0.027
76	0.33	0.29	0.14	0.24	<i>p</i> = 0.039
81	0.36	0.32	0.16	0.26	<i>p</i> = 0.025

602

Table 3. Pairwise differences of the means of absolute errors (in m) for the vegetation width parameter of the conventional manual measurement (CMM) and PROTOLIDAR and clustering methodologies, grouped by BBCH stage and with p value from ANOVA. Variances between groups that differed significantly using Bartlett's test with the Bonferroni adjustment are in italics. Considering the clustering method, for each BBCH stage, the β_w threshold that minimised the absolute mean error rate with CMM is in bold.

 Stage BBCH	clustering $\beta_{\rm w} = 1$	clustering $\beta_w = 2$	clustering $\beta_w = 3$	PROTOLIDAR	p-value
 14	0.37	0.28	0.17	0.11	p = 0.019
 53	0.38	0.11	0.21	0.18	p = 0.027
 57	0.22	0.13	0.26	0.21	p = 0.032
 61	0.36	0.14	0.22	0.24	p = 0.041
 70	0.20	0.17	0.29	0.28	p = 0.038
 76	0.32	0.19	0.24	0.31	p = 0.026
81	0.39	0.28	0.19	0.32	p = 0.039

609

610 3.3. Classification of the intercepted points of a 3D LiDAR point cloud

611 In order to illustrate the classification method used by the BPCC, a detailed analysis was performed on 612 a single vine unit at three phenological stages for height (Figures 3 and 4) and width (Figure 5). Figure 3 presents histograms of the points intercepted in the different compartments of a vine (trunk, leaf area 613 and trellis wires) along the height axis (y) at BBCH 14, 57 and 76. It visualises the change in the 614 number of defined clusters during the season, with a decrease from three to two clusters after BBCH 615 61, when the upper trellis wire is covered by the canopy (Figure 3C). The adaptive threshold for the 616 grassed zone filter (HG) also changed during the season, increasing from 0.25 m at BBCH 14 (Figure 617 3A) to 0.35 m at BBCH 76 (Figure 3C). Across all 560 vines, HG varied between 0 and 0.4 m on 618 different days and in different blocks (data not shown). The distribution of intercepted points in the 619 620 canopy zone followed a Gaussian distribution throughout the growing season (solid black lines in

Figure 3). Independently of phenological stage, a higher number of intercepted points, associated with 621 622 a higher density of vegetation, was found in the centre of the canopy zone (in green) (Figures 3A-C). This can be explained by the Royat cordon training system used in this block. This is explained by the 623 624 presence of primary shoots and the first three leaves in the central zone at the beginning of vegetation (BBCH 14), by the appearance of flower buds which are transformed into bunches mid-season (BBCH 625 626 57), and by the mechanical action of pruning, which induces the development of secondary shoots in 627 the central zone of the canopy later in the season (BBCH 76). 628 629 Figure 4 presents similar information to the statistical distributions in Figure 3, but in the form of a 2D plot along the row. It clearly illustrates changes in LiDAR returns associated with the under-vine 630 631 grassed zone as well as canopy height over the course of the season. The issue with the proximity of the upper canopy to the upper trellis wire mid-season (Figure 4B) and its potential effect on the 632 PROTOLIDAR method for height estimation is clear. While the PROTOLIDAR overestimated height 633 at BBCH 57 (and similar stages), the use of a three-class hierarchical clustering with a moderate level 634 635 of trimming ($\beta_{\rm H} = 2$) provided more accurate canopy height (VH) estimations (Table 2). 636 637 Figure 3 near here Figure 4 near here 638 639 640 An alternative view of the LiDAR returns from the same vine, at the same three phenological stages, to illustrate changes in canopy width is shown in Figure 5. This is a cross-section through the canopy 641 642 of a scan taken from the left-hand side of the image. Early in the season, the small vine size allows a 643 good characterisation of the full canopy from the half-row scan (Figure 5A). However, as noted previously, the characterisation of the distal side of the canopy is problematic with half-row scans as 644 the canopy develops, which leads to issues with underestimating full canopy width with the 645 646 PROTOLIDAR method. The decrease in the density of LiDAR returns from the far side of the canopy is obvious from midseason onwards (Figures 5B-C). 647 648 649 Figure 5 near here 650 651 3.4. Comparison of canopy dimensions derived from the PROTOLIDAR - BPCC - CMM 652 methodologies Table 4 presents the CCC, R² and CV-RMSE statistics from comparisons of estimations of canopy 653 height and width from the BPCC, PROTOLIDAR and CMM on all 560 vine units. The CMM was 654 considered as a reference observation. There was a strong statistical relationship between all pairwise 655 comparisons (Table 4). The BPCC generated similar results to the PROTOLIDAR, with height 656 657 estimations slightly more similar than width estimations between the two methods. This indicated that the proposed automated approach was similar to the more manually demanding PROTOLIDAR 658 659 method for vegetation height and width estimation. It is acknowledged that the relationship between 660 the width estimations is likely to change if the PROTOLIDAR method is applied to a full canopy scan 661 as it was initially intended to be.

662

663 When compared with CMM, the BPCC slightly outperformed PROTOLIDAR (higher CCC and R² 664 and lower CV-RMSE) for the estimation of canopy height and width. However, for both approaches 665 the relationship with CMM was strong over the entire season. For height, the improved fit with BPCC 666 resulted from the two stage filtering and classification approach that adapted to canopy development 667 and provided more accurate mid- and late-season estimations. For width, the PROTOLIDAR was 668 more accurate at early stages, when the canopy was small and open, allowing impacts to be made throughout the canopy (Table 3). As canopy size and density increased, impacts were less likely to
occur in distal parts of the canopy and width estimations with the PROTOLIDAR from a half-row
scan became less accurate than the BPCC method, which assumed symmetry and a fixed distance from
the LoT (Table 3).

673

Table 4. Results of the comparison between PROTOLIDAR methodology, a LiDAR Bayesian point

675 cloud classification algorithm (BPCC) and conventional manual measurement (CMM) in the

Pairwise comparisons		CCC		R ²		CV-RMSE (%)	
		VH	VW	VH	VW	VH	VW
PROTOLIDAR	BPCC	0.97	0.92	0.94	0.89	5	7
PROTOLIDAR	СММ	0.92	0.87	0.91	0.83	12	15
BPCC	СММ	0.94	0.90	0.92	0.85	10	13

676 estimation of vegetation height (VH) and width (VW) for all vines over the entire season.

677

678 3.5. Future applications for dose management of PPP and precision viticulture

The method for expressing PPP doses currently used in French viticulture is based on a fixed dose, 679 defined per ha ground surface area (Codis et al., 2016). In this context, a system for expressing PPP 680 681 doses that explicitly takes into account the evolution of the structure of the plant to be protected, as well as spatial variability in this evolution, would be an important step toward more efficient 682 683 agricultural practices. The development of precision spraying technologies has been identified as a key area for more efficient viticulture (and agriculture) practices (Berk et al., 2016). As such, the 684 685 automated method for analysis of LiDAR point clouds proposed here is an important step forward. In this work, the method and validation were focussed on the determination of canopy dimensions from 686 sensor data. By themselves, canopy dimensions are limited. Once obtained they are typically used to 687 calculate vegetative indicators, such as the TRV indicator that has been used to adjust PPP dose rates 688 in Switzerland (Viret and Höhn, 2008). More recently, the LWA indicator has been proposed by the 689 chemical industry as a new method to report dose expression at the European level (Wohlhauser, 690 2009). Ideally these indicators, which have both been used for dose adjustment and dose expression 691 692 (Llorens et al., 2010; Walklate et al., 2011), should also incorporate information related to canopy 693 density/porosity for still more accurate dose management (Pergher and Petris, 2008).

694

However, when these vegetative indicators are manually determined, there is still a possibility that the canopy structure metric is over- or under-estimated (Rüegg et al., 2001). Predictive modelling and real-time observation of spray deposition patterns are capable of providing a feed-back mechanism to correct misapplications (either over or under applications) (Saddem et al., 2017). These are not yet well developed or commercialised but they could also form an important part of any future PPP application system.

701

Vineyards and orchards present a wide variety of different canopy characteristics. Although the results presented here are only from one vineyard (over four different blocks) scenario, the BPCC has been designed to be flexible so that it can be adapted to a variety of training systems and production situations. The choice of thresholds used here may need to be altered for other types of production systems, and the relative importance of the four filtering algorithms in the AFM may change. However, once the parameters and thresholds have been set, the algorithm should run in a fully automated manner, permitting it to be used in on-the-go applications.

710 The concept of Line of Trunks (LoT) (del-Moral-Martínez et al. 2015) and the use of vine symmetry 711 (Arnó et al., 2015) to model canopy width have been used here. This permitted a more automated 712 approach to modelling canopy width and generated better estimates of canopy width from one-sided canopy scans from BBCH 53 (mid- and late-season) (Table 3). Early season width estimations were 713 better with PROTOLIDAR but the mean absolute error at BBCH 14 was low for BPCC (0.17 m). The 714 715 accuracy achieved with BPCC negated the need for scanning both sides of the canopy. This is important as scanning both sides requires the fusion of point clouds from both sides, with issues such 716 as rectification and harmonisation. Considering multi-row spraying, scanning only one canopy side 717 means fewer LiDAR sensing systems to be installed on a sprayer with real-time dose control. 718 719

720 Although focussed on canopy dimensions here, the automatic classification of different components of vines potentially provides additional, automatically collected information that could be used for vine 721 722 and vineyard management, e.g. spraying for variable grass height/growth or estimating vine trunk 723 diameter.

724

725 4. Conclusions

726 In this study, a LiDAR BPCC was proposed that combined an AFM and a clustering method to automate the 3D digital characterisation of the dimensions of vineyard canopies from LiDAR data. 727 The BPCC only required basic configuration related to vineyard set up to operate autonomously. To 728 evaluate the efficacy of the BPCC filtering of LiDAR point clouds, it was compared to a manual 729 human expert filter (HEF) and to a semi-automatic method requiring manual pre-processing 730 731 (PROTOLIDAR). The results obtained from data collected on several grape varieties in a two different 732 training modes demonstrated that the BPCC filtered the LiDAR point clouds in an equivalent way to HEF and to the well-accepted PROTOLIDAR research method. Hierarchical classification and 733 trimming of the AFM filtered LiDAR data yielded estimations of canopy height and width that were 734 strongly correlated with equivalent PROTOLIDAR estimations. The classification was most effective 735 when the threshold in the trimming process was permitted to be variable along the season. Empirical 736 737 results provided clear indications of the preferred threshold value for both height and width at different phenological stages. When a dynamic threshold was used, the canopy dimensions from the BPCC 738 739 process were closer to manual canopy observations than the equivalent PROTOLIDAR estimations. 740 These results demonstrated that although operating at a higher level of automation, which is more 741 suited to on-the-go processing, the proposed BPCC was more effective than the PROTOLIDAR to 742 filter point cloud data and to estimate canopy dimensions in vineyards from half-row scans. This is a first iteration of a potential automated LiDAR processing algorithm. Further improvements are needed 743 before commercial deployment, in particular to provide a more robust temporal estimator for the 744 determination of the trimming threshold and for determining the preferred number of classes for a 745 given phenological stage or canopy size. The BPCC method also needs to be validated over other 746 training systems.

- 747
- 748

749 **Acknowledgments**

750 The authors are indebted to the numerous individuals that assisted in collecting the field data presented 751 in this work: Hudebine Y., Riberolles X., Trinquier E., Delpuech X., Rico A., de Jesus P., Delpuech

X., Verges A., Lienard A., Kazakos A., Mariette A., de Runicki V., Bastianelli M. We also want to 752

753 thank the staff members of the vine estate Domaine Mas Piquet for making their plots available for our

754 measurements. The first author also addresses his personal thanks to especially Abdelghani Cheraiet and Louisa Moussaoui. 755

This work was supported by the French National Research Agency under the Investments for the 756

757 Future Program, referred to as ANR-16-CONV-0004. Anice Cheraiet's PhD is cofunded by #DigitAg 758 and IFV Acta.

759

760 *References*

- Andersen, H. J., Reng, L., & Kirk, K. (2005). Geometric plant properties by relaxed stereo vision
 using simulated annealing. Computers and Electronics in Agriculture, 49(2), 219-232.
 https://doi.org/10.1016/j.compag.2005.02.015
- 764 Arnó, J., Escolà, A., Vallès, J. M., Llorens, J., Sanz, R., Masip, J., Palacín, J., & Rosell-Polo, J. R.
- (2013). Leaf area index estimation in vineyards using a ground-based LiDAR scanner. Precision
 Agriculture, 14(3), 290-306. https://doi.org/10.1007/s11119-012-9295-0
- Arnó, J., Escolà, A., Masip, J., & Rosell-Polo, J. R. (2015). Influence of the scanned side of the row in
 terrestrial laser sensor applications in vineyards: Practical consequences. Precision Agriculture, 16(2),
 110, 128, https://doi.org/10.1007/s11110.014.0264.7
- 769 119-128. <u>https://doi.org/10.1007/s11119-014-9364-7</u>
- Arnó, J., Escolà, A., & Rosell-Polo, J. R. (2017). Setting the optimal length to be scanned in rows of
 vines by using mobile terrestrial laser scanners. Precision Agriculture, 18(2), 145-151.
 https://doi.org/10.1007/s11119-016-9451-z
- 773 Bastianelli, M., Rudnicki, V. D., Codis, S., Ribeyrolles, X., & Naud, O. (2017). Two vegetation
- indicators from 2D ground Lidar scanner compared for predicting spraying deposits on grapevine.
- 775 Proceedings of the 2017 EFITA WCCA conference, Montpellier, France, 153-154.
- 776 Berk, P., Hocevar, M., Stajnko, D., & Belsak, A. (2016). Development of alternative plant protection
- product application techniques in orchards, based on measurement sensing systems: A review.
 Computers and Electronics in Agriculture, 124, 273-288.
- 779 <u>https://doi.org/10.1016/j.compag.2016.04.018</u>
- Byers, R. E., Hickey, K. D., & Hill, C. H. (1971). Base gallonage per acre. Virginia Fruit, 60(8), 1923.
- 782 Campos, J., Llop, J., Gallart, M., García-Ruiz, F., Gras, A., Salcedo, R., & Gil, E. (2019).
- Development of canopy vigour maps using UAV for site-specific management during vineyard
 spraying process. Precision Agriculture, 20(6), 1136-1156. <u>https://doi.org/10.1007/s11119-019-09643-</u>
- 785 <u>z</u>
- 786 Codis, S. (2016). Stakes for a new model of dose expression in viticulture: advantages and points to
- 787 be taken into consideration. In Proceedings of the EPPO Workshop on harmonized dose expression for
- the zonal evaluation of plant protection products in high growing crops, Austrian Agency for Healthand Food Safety Vienna, 12-13.
- Colaço, A. F., Molin, J. P., Rosell-Polo, J. R., & Escolà, A. (2018). Application of light detection and
 ranging and ultrasonic sensors to high-throughput phenotyping and precision horticulture: Current
- status and challenges. Horticulture research, 5(1), 1-11. https://doi.org/10.1038/s41438-018-0043-0
- 793 de Castro, A., Jiménez-Brenes, F., Torres-Sánchez, J., Peña, J., Borra-Serrano, I., & López-Granados,
- F. (2018). 3-D Characterization of Vineyards Using a Novel UAV Imagery-Based OBIA Procedure
 for Precision Viticulture Applications. Remote Sensing, 10(4), 584.
- 796 <u>https://doi.org/10.3390/rs10040584</u>
- 797 del-Moral-Martínez, I., Arnó, J., Escolà, A., Sanz, R., Masip-Vilalta, J., Company-Messa, J., &
- Rosell-Polo, J. (2015). Georeferenced Scanning System to Estimate the Leaf Wall Area in Tree Crops.
 Sensors, 15(4), 8382-8405. https://doi.org/10.3390/s150408382
- 800 del-Moral-Martinez, I., Rosell-Polo, J. R., Company, J., Sanz, R., Masip, J., Martínez-Casasnovas, J.
- 801 A., & Arnó, J. (2016). Mapping Vineyard Leaf Area Using Mobile Terrestrial Laser Scanners: Should
- 802 Rows be Scanned On-the-Go or Discontinuously Sampled? Sensors, 16(1), 119. 803 https://doi.org/10.3390/s16010119
- 804 Escolà, A., Martínez-Casasnovas, J. A., Rufat, J., Arnó, J., Arbonés, A., Sebé, F., Pascual, M.,
- 805 Gregorio, E., Rosell-Polo, J. R. (2017). Mobile terrestrial laser scanner applications in precision

- fruticulture/horticulture and tools to extract information from canopy point clouds. Precision
 Agriculture, 18(1), 111–132. <u>https://doi.org/10.1007/s11119-016-9474-5</u>
- Fraley, C., & Raftery, A. (2007). Model-based Methods of Classification : Using the mclust Software
 in Chemometrics. Journal of Statistical Software, 18(6). https://doi.org/10.18637/jss.v018.i06
- 810 Fraley, C., & Raftery, A. E. (2012). Mclust Version 4 for R: Normal Mixture Modeling for Model-
- 811 Based Clustering, Classification, and Density Estimation. Technical Report No. 597. Department of
- 812 Statistics, University of Washington.
- Gil, E., Escolà, A., Rosell, J. R., Planas, S., & Val, L. (2007). Variable rate application of plant
 protection products in vineyard using ultrasonic sensors. Crop Protection, 26(8), 1287-1297.
 https://doi.org/10.1016/j.cropro.2006.11.003
- Gil, E., Llorens, J., Llop, J., Fàbregas, X., Escolà, A., & Rosell-Polo, J. R. (2013). Variable rate
 sprayer. Part 2 Vineyard prototype: Design, implementation, and validation. Computers and
 Electronics in Agriculture, 95, 136-150. https://doi.org/10.1016/j.compag.2013.02.010
- 819 Gil, E., Arnó, J., Llorens, J., Sanz, R., Llop, J., Rosell-Polo, J., Gallart, M., & Escolà, A. (2014).
- Advanced Technologies for the Improvement of Spray Application Techniques in Spanish Viticulture:
 An Overview. Sensors, 14(1), 691-708. <u>https://doi.org/10.3390/s140100691</u>
- Jaakkola, A., Hyyppä, J., Kukko, A., Yu, X., Kaartinen, H., Lehtomäki, M., & Lin, Y. (2010). A low-
- 823 cost multi-sensoral mobile mapping system and its feasibility for tree measurements. ISPRS Journal of
- Photogrammetry and Remote Sensing, 65(6), 514-522. <u>https://doi.org/10.1016/j.isprsjprs.2010.08.002</u>
- 825 Lee, K. H., & Ehsani, R. (2009). A Laser Scanner Based Measurement System for Quantification of
- 826 Citrus Tree Geometric Characteristics. Applied Engineering in Agriculture, 25(5), 777-788.
 827 <u>https://doi.org/10.13031/2013.28846</u>
- Lin, L. I.-K. (1989). A Concordance Correlation Coefficient to Evaluate Reproducibility. Biometrics,
 45(1), 255. <u>https://doi.org/10.2307/2532051</u>
- Llorens, J., Gil, E., Llop, J., & Escolà, A. (2010). Variable rate dosing in precision viticulture: Use of
 electronic devices to improve application efficiency. Crop Protection, 29(3), 239-248.
 https://doi.org/10.1016/j.cropro.2009.12.022
- 833 Llorens, J., Gil, E., Llop, J., & Escolà, A. (2011a). Ultrasonic and LiDAR Sensors for Electronic
- 834 Canopy Characterization in Vineyards: Advances to Improve Pesticide Application Methods. Sensors,
- 835 11(2), 2177-2194. https://doi.org/10.3390/s110202177
- Llorens, J., Gil, E., Llop, J., & Queraltó, M. (2011b). Georeferenced LiDAR 3D Vine Plantation Map
 Generation. Sensors, 11(6), 6237-6256. <u>https://doi.org/10.3390/s110606237</u>
- 838 Lorenz, D. H., Eichorn, K. W., Bleiholder, H., Klose, U., Meier, U & Weber, E. (1994).
- 839 Phänologische Entwicklungsstadien der Weinrebe (Vitis vinifera L. spp. Vinifera). (Phenological
- 840 stages of grapevine (Vitis vinifera L. spp. Vinifera)). Viticultural and Enological Science 49, 66-70
- 841 Manktelow, D. W. L., & Praat, J. P. (1997). The tree-row-volume spraying system and its potential
- use in New Zealand. Proceedings of the New Zealand Plant Protection Conference, 50, 119-124.
 https://doi.org/10.30843/nzpp.1997.50.11360
- 844 Mathews, A., & Jensen, J. (2013). Visualizing and Quantifying Vineyard Canopy LAI Using an
- 845 Unmanned Aerial Vehicle (UAV) Collected High Density Structure from Motion Point Cloud.
 846 Remote Sensing, 5(5), 2164-2183. <u>https://doi.org/10.3390/rs5052164</u>
- 847 Miranda, C., Santesteban, L. G., Escalona, J. M., De Herralde, F., Aranda, X., Nadal, M., Intrigliolo,
- 848 D., Castel, J., Royo, J., & Medrano, H. (2017). Allometric relationships for estimating vegetative and
- 849 reproductive biomass in grapevine (Vitis vinifera L.): Allometric relations for grapevine biomass.
- Australian Journal of Grape and Wine Research, 23(3), 441-451. https://doi.org/10.1111/ajgw.12285
- 851 Moorthy, I., Miller, J. R., Berni, J. A. J., Zarco-Tejada, P., Hu, B., & Chen, J. (2011). Field
- 852 characterization of olive (Olea europaea L.) tree crown architecture using terrestrial laser scanning

- 853
 data.
 Agricultural
 and
 Forest
 Meteorology,
 151(2),
 204-214.

 854
 https://doi.org/10.1016/j.agrformet.2010.10.005

 204-214.
- 855 Viret, O., Siegfried, W., Wohlhauser, R. (2005). Crop adapted spraying in viticulture. Leaf volume
- dependant pesticide dosage for a precise and ecological application. 8thworkshop on spray application
 techniques in fruit growing, Barcelona, Spain, s. 23-24.
- Viret, O. & Höhn, H. (2008). Application de la dose selon la méthode du TRV. Revue Suisse de
 Viticulture, Arboriculture et Horticulture 40, 50-51.
- 860 Palacin, J., Palleja, T., Tresanchez, M., Sanz, R., Llorens, J., Ribes-Dasi, M., Masip, J., Arnó, J.,
- 861 Escolà, A., & Rosell, J. R. (2007). Real-Time Tree-Foliage Surface Estimation Using a Ground Laser
- 862 Scanner. IEEE Transactions on Instrumentation and Measurement, 56(4), 1377-1383.
 863 <u>https://doi.org/10.1109/TIM.2007.900126</u>
- 864 Pergher, G., & Petris, R. (2008). Pesticide dose adjustment in vineyard spraying and potential for dose
- reduction. Manuscript ALNARP 08 011. Agricultural Engineering International. CIGR Ejournal X(May).
- 867 Poni, S., Lakso, A., Intrieri, C., Rebucci, B., & Filippetti, I. (1996). Laser scanning estimation of
- relative light interception by canopy components in different grapevine training systems. VITISGEILWEILERHOF, 35, 177–182.
- Rinaldi, M, Llorens, J., & Gil, E. (2013). Electronic characterization of the phenological stages of
 grapevine using a LiDAR sensor. Precision Agriculture Wageningen: Academic Publishers), 13, 603609.
- 873 Rosell, J. R., Llorens, J., Sanz, R., Arnó, J., Ribes-Dasi, M., Masip, J., Escolà, A., Camp, F.,
 874 Solanelles, F., Gràcia, F., Gil, E., Val, L., Planas, S., & Palacín, J. (2009). Obtaining the three-
- dimensional structure of tree orchards from remote 2D terrestrial LiDAR scanning. Agricultural and
- 876 Forest Meteorology, 149(9), 1505-1515. https://doi.org/10.1016/j.agrformet.2009.04.008
- Rosell, J.R., & Sanz, R. (2012). A review of methods and applications of the geometric
 characterization of tree crops in agricultural activities. Computers and Electronics in Agriculture, 81,
 124-141. <u>https://doi.org/10.1016/j.compag.2011.09.007</u>
- 880 Rosell Polo, J. R., Sanz, R., Llorens, J., Arnó, J., Escolà, A., Ribes-Dasi, M., Masip, J., Camp, F.,
- 881 Gràcia, F., Solanelles, F., Pallejà, T., Planas, S., Gil, E., Val, L., & Palacín, J. (2009). A tractor-
- 882 mounted scanning LiDAR for the non-destructive measurement of vegetative volume and surface area
- of tree-row plantations: A comparison with conventional destructive measurements. Biosystems
 Engineering, 102(2), 128-134. <u>https://doi.org/10.1016/j.biosystemseng.2008.10.009</u>
- Rovira-Más, F., Reid, J. F., & Zhang, Q. (2006). Stereovision data processing with 3d density maps
 for agricultural vehicles. Transactions of the ASABE, 49,1213-1222.
- 887 Rüegg, J., Siegfried, W., Raisigl, U., Viret, O., Steffek, R., Reisenzein, H., & Persen, U. (2001).
- 888 Registration of plant protection products in EPPO countries: Current status and possible approaches to
- harmonization. EPPO Bulletin, 31(2), 143-152. https://doi.org/10.1111/j.1365-2338.2001.tb00983.x
- Saddem, R., Naud, O., Cazenave, P., Godary-Dejean, K., & Crestani, D. (2017). Precision spraying:
 from map to sprayer control using model-checking. Journal of Agricultural Informatics, 8(3), 1–10.
- 892 Sanz, R., Palacin, J., Sisó, J., Ribes-Dasi, M., Masip, J., Arnó, J., Llorens, J., Valles, J.M., & Rosell, J.
- 893 (2004). Advances in the measurement of structural characteristics of plants with a LiDAR scanner.
- 894 International Conference on Agricultural Engineering, Leuven (Belgium). Book of Abstracts of the
- 895 AgEng 2004 Conference, Paper No. 27, 400–401.
- 896 Sanz, R., Rosell, J. R., Llorens, J., Gil, E., & Planas, S. (2013). Relationship between tree row LiDAR-
- 897 volume and leaf area density for fruit orchards and vineyards obtained with a LiDAR 3D Dynamic
- 898 Measurement System. Agricultural and Forest Meteorology, 171-172, 153-162. 899 https://doi.org/10.1016/j.agrformet.2012.11.013

- 900 Sanz, R., Llorens, J., Escolà, A., Arnó, J., Planas, S., Román, C., & Rosell-Polo, J. R. (2018). LiDAR
- 901 and non-LiDAR-based canopy parameters to estimate the leaf area in fruit trees and vineyard.
- 902AgriculturalandForestMeteorology,260-261,229-239.
- 903 <u>https://doi.org/10.1016/j.agrformet.2018.06.017</u>
- 904 Schultz, H. (1995). Grape canopy structure, light microclimate and photosynthesis. I. A two-
- 905 dimensional model of the spatial distribution of surface area densities and leaf ages in two canopy 906 systems. Vitis, 34, 211–215.
- Schumann, A. W., & Zaman, Q. U. (2005). Software development for real-time ultrasonic mapping of
 tree canopy size. Computers and Electronics in Agriculture, 47(1), 25-40.
 https://doi.org/10.1016/j.compag.2004.10.002
- 910 Siebers, M., Edwards, E., Jimenez-Berni, J., Thomas, M., Salim, M., & Walker, R. (2018). Fast
- 911 Phenomics in Vineyards: Development of GRover, the Grapevine Rover, and LiDAR for Assessing
- 912 Grapevine Traits in the Field. Sensors, 18(9), 2924. <u>https://doi.org/10.3390/s18092924</u>
- 913 Siegfried, W., Viret, O., Huber, B., & Wohlhauser, R. (2007). Dosage of plant protection products 914 adapted to leaf area index in viticulture. Crop Protection, 26(2),73-82. https://doi.org/10.1016/j.cropro.2006.04.002 915
- Solanelles, F., Escolà, A., Planas, S., Rosell, J. R., Camp, F., & Gràcia, F. (2006). An Electronic
 Control System for Pesticide Application Proportional to the Canopy Width of Tree Crops.
- 918 Biosystems Engineering, 95(4), 473-481. https://doi.org/10.1016/j.biosystemseng.2006.08.004
- 919 Stajnko, D., Berk, P., Lešnik, M., Jejčič, V., Lakota, M., Štrancar, A., Hočevar, M., & Rakun, J.
- 920 (2012). Programmable Ultrasonic Sensing System for Targeted Spraying in Orchards. Sensors, 12(11),
 921 15500-15519. <u>https://doi.org/10.3390/s121115500</u>
- Taylor, J., Tisseyre, B., Bramley, R., & Reid, A. (2005). A comparison of the spatial variability of
 vineyard yield in European and Australian production systems. Precision Agriculture, 5, 907-914.
- 924 Tisseyre, B., Mazzoni, C., & Fonta, H. (2008). Within-field temporal stability of some parameters in
- 925 viticulture: Potential toward a site specific management. OENO One, 42(1), 27-39.
- 926 Torres-Sánchez, J., López-Granados, F., Borra-Serrano, I., & Peña, J. M. (2018). Assessing UAV-
- collected image overlap influence on computation time and digital surface model accuracy in olive
 orchards. Precision Agriculture, 19(1), 115-133. https://doi.org/10.1007/s11119-017-9502-0
- 929 Van der Zande, D., Hoet, W., Jonckheere, I., van Aardt, J., & Coppin, P. (2006). Influence of
 930 measurement set-up of ground-based LiDAR for derivation of tree structure. Agricultural and Forest
 931 Meteorology. 141, 147–160
- 932 Vercruysse, F., Steurbaut, W., Drieghe, S., & Dejonckheere, W. (1999). Off target ground deposits
- 933 from spraying a semi-dwarf orchard. Crop Protection, 18, 565-570.
- 934 Walklate, P. J., Cross, J. V., Richardson, G. M., Murray, R. A., & Baker, D. E. (2002). IT-
- 935 Information Technology and the Human Interface. Biosystems Engineering, 82(3), 253-267.
- 936 <u>https://doi.org/10.1006/bioe.2002.0082</u>
- Walklate, P. J., Cross, J.V., & Pergher, G. (2011). Support system for efficient dosage of orchard and
 vineyard spraying products. Computers and Electronics in Agriculture 75, 355–362
- Walklate, P. J., & Cross, J. V. (2013). Regulated dose adjustment of commercial orchard spraying
 products. Crop Protection, 54, 65-73. https://doi.org/10.1016/j.cropro.2013.07.019
- Westfall, P. H., Johnson, W. O., & Utts, J. M. (1997). A Bayesian perspective on the Bonferroni
 adjustment. Biometrika, 84(2), 419-427.
- 943 Wohlhauser, R. (2009). Dose rate expression in tree fruits the need for harmonization approach from
- 944 a chemical producer industry perspective. In Proceedings of the Tree Fruit Dose Adjustment
- 945 Discussion Group Meeting, Wageningen, The Netherlands, September 2009.
- 946

947

948 Figures caption

Figure 1. A: Representation of the scanning procedure showing polar (distance, ρ , and angle, θ) and 949 Cartesian (x, y, z plane) coordinate reference systems. B: Overhead view of two simulated scans along 950 the row (x, z plane) showing projected LiDAR returns for a 1 m vine unit (0.5 m either side of the 951 952 trunk). The shaded area indicates the progressive reduction in LiDAR returns across the crosssectional area of the canopy. Legend: O - origin, LoT - Line of Trunks, D - distance between the 953 LiDAR travel line and the LoT, $\Delta\theta$ - angular resolution of the scans, HS - height of the LiDAR above 954 ground and HG - height of the grassed zone above the ground, ΔW - distance interval between two 955 956 consecutive vertical scans and FTS - forward travel speed of tractor used to mount the LiDAR.

957

Figure 2. View (y, z plane) of a LiDAR point cloud corresponding to a vine unit at BBCH 14 to illustrate the points deleted in the automated filtration step (blue) and points retained (red) to calculate canopy dimensions. The four filter functions applied to this LiDAR point cloud (in dotted lines) are (1) inter-row ground filter based on angle (δ), (2) filtering of adjacent rows based on the distance (D),

962 (3) grassed zone filtering, based on the height of the grassed zone (HG) and (4) near point filtering.

963

Figure 3. Examples of the vertical distribution of the intercepted points in the 3D LiDAR point clouds 964 on one vine unit at three phenological stages illustrating how the BPCC algorithm filters and classifies 965 966 the point cloud into different zones (non-vine ground vegetation in white, trunk zone in red, canopy 967 zone in green and trellis wire in blue). A = BBCH 14 (early season); B = BBCH 57 (mid-season) and 968 C = BBCH 76 (mid-late season). A and B have three distinct zones (trunk – canopy – trellis wire), while C exhibits only two zones as the trellis wire is covered by the canopy. The horizontal lines (in 969 yellow) represent the σ -based thresholds ($\beta_{\rm H}$: with $\sigma = 2$ for A and B, and 3 for C) that were used in 970 971 the BPCC to define the canopy zone. The dotted line indicates the threshold used for the grassed zone filter (HG) that changes as the under-vine vegetation grows. The black line represents the distribution 972 973 of points in the canopy zone only.

974 Figure 4. Examples of the LiDAR point clouds on one vine unit seen from the inter-row to illustrate how the proposed BPCC filtering and classification algorithm defines the canopy height (VH) and the 975 976 undervine grass height (HG) at different phenological stages of the season. (A, B and C: same stages 977 as Fig. 3). The horizontal lines (in yellow) represent the σ -based thresholds ($\beta_{\rm H}$: with $\sigma = 2$ for A and B, and 3 for C) that were used in the BPCC to define the canopy zone at each stage. The dashed line 978 979 indicates the derived threshold for the grassed zone filter (HG) that changes as the under vine 980 vegetation grows. For comparison, the canopy height derived from the PROTOLIDAR method is shown in B illustrating the effect of the trellis wire on VH estimates with PROTOLIDAR with larger 981 982 canopies.

Figure 5. Examples of cross-sections ('scans') of LiDAR points for one vine unit at three different phenological stages (same vine and stages as shown in Fig. 4), illustrating how the proposed BPCC filtering and classification algorithm and the PROTOLIDAR method define canopy width (VW). The vertical lines represent the extremes of canopy width from both approaches. The BPCC has a 5% threshold to estimate the half-row width (solid line) and the distal extreme is estimated assuming symmetry (dashed lines). The PROTOLIDAR derives width directly from the LiDAR returns and underestimates canopy width in larger canopies (B and C) relative to the BPCC.











