Computers and Electronics in Agriculture 230 (2025) 109856

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Quantitative prediction of grape ripening parameters combining an autonomous IoT spectral sensing system and chemometrics

Alessio Tugnolo [a], Hugo M. Oliveira [c], Valentina Giovenzana [a,*], Natacha Fontes [b], Sara Silva [b], Cristina Fernandes [b], António Graça [b], Alessia Pampuri [a], Andrea Casson [a], João Piteira [c], Paulo Freitas [c], Riccardo Guidetti [a], Roberto Beghi [a]

[a] Department of Agricultural and Environmental Sciences (DiSAA), Università degli Studi di Milano, via G. Celoria 2, 20133 Milan, Italy
[b] Sogrape Vinhos, S.A., Aldeia Nova, Avintes, Portugal
[c] INL, International Iberian Nanotechnology Laboratory, Avenida Mestre José Veiga s/n, 4715-330, Braga, Portugal

A R T I C L E I N F O

Keywords:
Proximal sensing
Sensors network
Viticulture
Precision agriculture
Chemometrics

A B S T R A C T

The research presented in this study offers a contribution to the field of viticulture by testing at lab scale an innovative approach for monitoring grape ripening using an autonomous proximal sensing technology.

By leveraging an IoT spectral sensing system, termed i-Grape, the research aims to remotely monitor vineyards and provide real-time data on grape ripening status. This system, consisting of tailored optical, host, and controller modules, offers a novel solution for continuous monitoring throughout the crop season, overcoming limitations associated with traditional sampling methods.

The study conducted comprehensive sampling in the viticulture area of the Douro Valley, collecting data from cv. Touriga Nacional and Touriga Franca. Both optical and wet-chemistry analyses were performed on the grape samples to develop predictive models for ripening parameters, including Total Soluble Solids (TSS), Potential Alcohol (PA), pH, Titratable Acidity (TA), Total Polyphenols (TP), and Extractable Anthocyanins (EA).

Exploratory analysis of the optical data revealed insights into the behaviour of the spectral readouts over time, highlighting the evolution of grape ripening and the potential interference factors that need to be addressed for accurate modelling. Pre-processing techniques, including background subtraction and Log10 transformation, were employed to enhance the quality of the optical data and improve model performance.

Overall, predictive PLS models with good performance were obtained for the estimation of the technological ripening parameters (RPD = 2.76 and $R^2 = 0.86$ for TSS; RPD = 2.58 and $R^2 = 0.85$ for PA; RPD = 3.65 and $R^2 = 0.92$ for TA; RPD = 2.27 and $R^2 = 0.79$ for pH), establishing a solid ground for the application of this sensing strategy in the field. For the phenolic parameters (TP and EA), the performance of the models is still insufficient (RPD = 1.28 and $R^2 = 0.51$ for TP; RPD = 1.55 and $R^2 = 0.58$ for EA). A comparison with existing literature highlighting the advancements achieved in terms of predictive performance and operational capabilities has been reported. The potential of the i-Grape system to revolutionize grape ripening monitoring by offering a cost-effective, non-destructive, and scalable solution for vineyard management has been demonstrated at lab scale.

In conclusion, the research laid the groundwork for further advancements in optical sensing technology for viticulture, opening up avenues for future research in optimizing hardware design, data processing algorithms, and field implementation strategies to realize the full potential of IoT-based solutions in precision agriculture.

1. Introduction

The wine sector is evolving in an increasingly competitive international scenario characterised by new producing countries with innovative strategies in production and trade. In this highly competitive market, it is now well-accepted that the quality of a wine depends mainly on the qualitative features in terms of the chemical characteristics of the grapes used to produce it (Giovenzana et al., 2018). Moreover, the wine sector is adapting to changing lifestyles which accelerated due to the Covid-19 pandemic (Cavallo et al., 2020). For instance,

* Corresponding author.
E-mail address: valentina.giovenzana@unimi.it (V. Giovenzana).

https://doi.org/10.1016/j.compag.2024.109856
Received 8 March 2024; Received in revised form 12 July 2024; Accepted 17 December 2024
Available online 8 January 2025

sustainability concerns have increased but also led to new behaviours such as consuming more wine at home (Eurostat., 2019; Vallone et al., 2019). All these aspects are contributing to a growing interest in more sustainable production management, as well as in the quality of the raw material – the grapes.

The quality of the grapes relies on their chemical composition, which makes the information about grapes' ripening status the key element to producing high-quality wines. Indeed, this heterogeneity of grape's chemical composition can be affected by multiple factors, and it has a significant effect on the wine composition and quality (Kontoudakis, et al., 2011). The conventional method for monitoring these parameters relies on the random sampling of berries in a pre-defined area, followed by different destructive wet-chemistry assays for the different target parameters (OIV, 2022). This approach is limited by the reduced number of samples that can be collected due to the logistic constraints (e.g. the distance from the lab to the field) and also the time gap between sample collection and results. Furthermore, the use of a destructive assay leads to the production of a significant amount of chemical waste, which is a critical factor in seeking for a more environmentally sustainable production (Casson et al., 2019).

Alternatives to this current standard methodology have been proposed over the last decade, with a particular emphasis on non-destructive optical methods that commonly include prediction models. One example is the use of a portable Vis/NIR spectrophotometer for the quantification of the major technological parameters (Total Soluble Solids, pH, Total Acidity) of the maturation process (Giovenzana et al., 2014, Giovenzana et al., 2015). More recently, portable devices based on discrete optical bands have been developed to the same end (Pampuri et al., 2021a). Additionally, fluorescence-based instruments (Agati et al., 2007). have been successfully used to predict phenolic parameters. More recently, methods based on computer vision have become an area of growing interest due to the possibility of acquiring multiple images and automating them (Vrochidou et al., 2021). Although these examples represent a significant improvement over the current state of the art, they are still unable to be a continuous source of analytical information. Most of these methods still require manual data collection (the instruments need an operator or are mounted in a vehicle that needs to be driven across the vineyard) (Pampuri et al., 2021a; Fernández-Novales et al. 2019; Urraca et al., 2016; Giovenzana et al., 2014, 2015; Matese and Filippo Di Gennaro, 2015). Moreover, computer vision methods also require significant computing power to develop the prediction models and then predict the target parameters. The combination of these two factors (in addition to the cost of spectrophotometers and cameras) makes it difficult to use these analytical tools on a large scale in the field, which impacts the number of samples that can be collected during the complete period of a crop season (~3 months). In this context, an Internet of Things (IoT) spectral sensing system (i-Grape) has been proposed in Oliveira et al. (2024) describing the concept of application and the technical specification. Such a device can be installed in proximity to the target (bunch) to remotely monitor the vineyard during the crop season. However, to make this device practical for winemakers, the development of models that can predict grape maturation parameters with accuracy and reliability comparable to conventional methods (including measurement scales and associated error margins) is essential.

Therefore, this study aims to utilize and test the i-Grape prototype by developing new prediction models for estimating ripening parameters based on optical data collected in a lab environment for two grape varieties, cv. Touriga Nacional and cv. Touriga Franca. The research thoroughly examines various aspects of multivariate model development, such as the evolution of optical signals throughout the crop season, potential external interferences affecting measurements, and data preprocessing. These new features will introduce new capabilities for assessing the grape's ripening status in a fast, non-destructive fashion, which would positively impact the harvesting operations (namely operating procedures, scheduling, and classification).

2. Materials and methods

2.1. Spectral sensing system basic specs

The spectral sensing system (Fig. 1) used to measure the reflectance of the grapes in a lab environment is a tailor-made IoT end node which was introduced (for the first time) in detail by Oliveira et al. (2024) Briefly, the hardware system is composed of three modules: (i) an optical module, which includes the optical components (LEDs and photodetectors), (ii) a host module, which includes the LED driver and the analog front-end for signal condition, and (iii) a controller module, which is a platform that controls the operation of the system, namely the spectral sensor protocol sequence, the data storage, and connectivity of the system. The optical module, which contacts with the grape berry, is composed of four photodiodes (PDs) with an active area of 700 x 700 μm surrounded by four chip-level LEDs (~300 x 300 μm) with dominant wavelengths of 530, 630, 690, and 730 nm. The optical module was produced in RIGID.flex PCB fabrication process from Würth Elektronik (Waldenburg, Germany), where the flexible polyimide section connects the optical head to the host module.

In this work, the system used for data acquisition included a single spectral sensor considering its use in a lab environment. The optical module was placed onto a 3D-printed holder, where the grape was placed and covered before the measurement cycle began. The optical data acquired was stored locally on a micro-SD card (the radio transmission feature of the controller was disabled).

The measurement cycle included the consecutive illumination of a single grape berry by an individual LED followed by the signal acquisition in each of the four photodetectors (also labelled as channels for the sake of simplicity). In the end, an acquisition with all LEDs off was also performed to evaluate any background effects. Therefore, a single measurement cycle was composed of 20 individual readouts (sixteen for the four combinations of LED-PDs plus four corresponding to the acquisition in the absence of illumination).

2.2. Sampling

The experimental activity took place in the viticulture area of the Douro Valley at Quinta do Seixo (Valença do Douro, Tabuaço, Portugal, latitude: N 41°10′4.44″ and longitude: W 7°33′18.36″) owned by Sogrape Vinhos since 1987. The vineyard covers 71 ha of century-old vineyards on a steep-slope landscape on the south bank of the Douro River. In more than 30 years of commercial activity, the company collected data by building a historical dataset that was used to verify the ripening trends consistency during the experimentation. Activities began from late July to mid-September 2020 by collecting a total of 106 grape samples (1 sample is equivalent to 200 grape berries randomly collected) from six dates (Table 1). The sampling was performed on cv. Touriga Nacional (TN) and Touriga Franca (TF). TN samples were collected from a pilot-vineyard of around 1 ha total area at Quinta do Seixo (Tabuaço, Portugal) whereas TF samples came from different vineyards across the Douro Valley region. A total of 106 samples (75 from TN and 31 from TF) were analysed using both the spectral sensing system and the standard procedures for the reference wet-chemical analyses.'

2.2.1. Optical analysis

Fig. 2 shows the experimental protocol used for analyzing each sample with the spectral sensing system and conventional wet-chemistry assays. Each sample, consisting of 200 berries collected from the pilot vineyard at Quinta do Seixo in Tabuaço, Portugal, underwent a two-step analysis process. First, 30 berries from each sample were optically analyzed using the spectral sensor. The optical signatures of these 30 berries were averaged to obtain a representative signal for the entire sample of 200 berries. Subsequently, the same 200 berries were analyzed using the traditional wet-chemistry assays employed by the winery, Sogrape Vinhos. During the optical analysis, each berry was

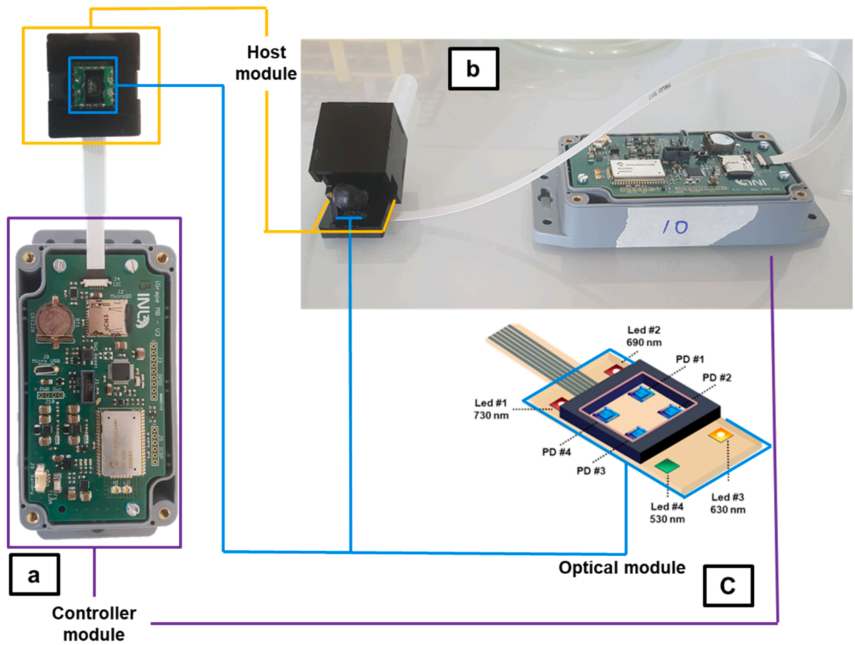

Fig. 1. Spectral sensing system with optical sensing head (a), operative representation of the system for data acquisition (b), and schematic representation of the optical head structure (c).

Table 1
Full calendar and number of samples collected in 2020 campaign for *cv.* Touriga Nacional (TN) and Touriga Franca (TF).

	Varieties	
Time	TN	TF
t0	July 23rd	July 23rd
Samples	14	10
t1	August 5th	August 10th
Samples	11	3
t2	August 20th	August 17th
Samples	12	5
t3	September 3rd	August 24th
Samples	12	6
t4	September 10th	August 31st
Samples	12	3
t5	September 17th	September 14th
Samples	14	4
Tot samples	**75**	**31**

placed on the optical module and covered with a 3D-printed cap before the optical measurement routine was manually initiated by the operator.

2.2.2. *Reference analysis*

The reference analyses were performed in the Sogrape's laboratory (Quinta do Seixo, Tabuaço, Portugal) immediately after the sampling in the field and the optical analysis. A total of 200 random berries (2 to 3 berries from each grape bunch) per sample were collected and analysed from each plot. The reference analyses performed were: (i) Total Soluble Solids (TSS, °Brix) using a digital refractometer (PAL-1 ATAGO, Japan), (ii) Potential Alcohol (PA, % vol), (iii) Titratable Acidity (TA, g of tartaric acid L^{-1}) using an automatic titrator (TitroMatic KF 1S, Crison Instruments, Italy), (iv) Total Polyphenols (TP, mg/dm^{-3}) calculated by direct measurement of the optical density at 280 nm, under a 1-cm quartz cell (Ribéreau-Gayon et al., 1998), (v) Extractable Anthocyanins (EA, mg/dm^{-3}) calculated through Glories' method (Ribéreau-Gayon et al., 2006) based on the Optical Density measurements at 280 and 520 nm using a UV/Vis spectrophotometer and (vi) pH (pH meter, PCE Inst. GmbH, Germany).

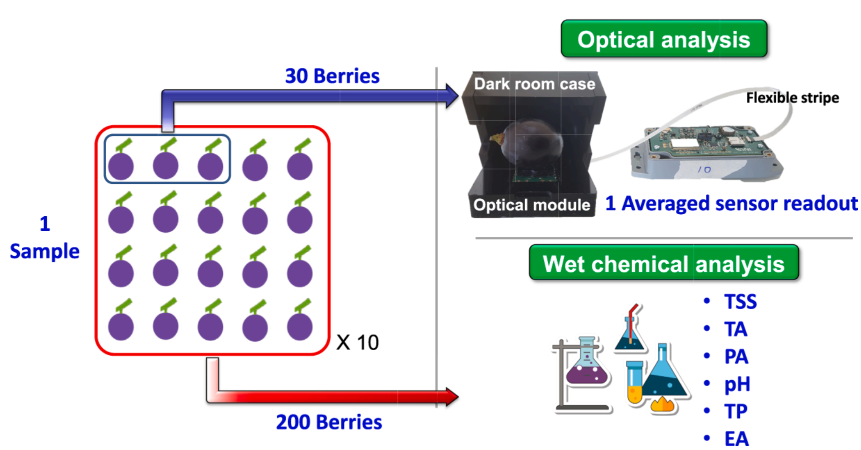

Fig. 2. Sampling procedure for data collection (optical and reference).

2.3. Modelling

Multivariate data analysis was carried out using Matlab® version 2020b (The MathWorks, Inc., Natick, MA, USA). This involved both the PLS-Toolbox package (Eigenvector Research, Inc., Manson, Washington), built-in, and custom functions.

Different spectral pre-processing techniques were considered to remove any irrelevant information (offsets and scattering) that could negatively affect the regression models. By the end, the readouts were treated row-wise with Log10 (to improve correlation with the reference data) and autoscaled column-wise (before applying scale-dependent multivariate analysis methods such as Principal Components Analysis (PCA) or Partial Least square Regression (PLS) to minimize location differences and to give the same importance to all the optical outputs.

PCA was applied to the data matrix of readouts to investigate the variability in the optical results. Subsequently, latent variable modelling was conducted using the PLS method, which linearly maximizes the covariance between the optical readouts and the reference analyses (TSS, PA, pH, TA, TP, and EA) (Oliveri et al., 2020). To verify the robustness of the developed PLS models, a uniform and representative subdivision of the samples was performed using the Kennard–Stone duplex algorithm (Kennard & Stone, 1969). Such method was used to partition the entire dataset (106 sample units) into the calibration set (75 sample units, thus the 70 % of 106) and prediction set (the rest of 31 sample units used as external validation set).

The model's accuracy was then assessed using several metrics: RMSE (root mean square error), bias, and R2 (coefficient of determination). High model performances are associated with low values of error and bias, and high values of R^2 (as maximum equal to 1). Besides, the RPD (residual prediction deviation, i.e. the ratio between the standard deviation related to the response variable of the set of data and the root mean squared error associated) was calculated. As a reference, a RPD between 1.5 and 2 indicates the capability of the model to distinguish low from high values of the response variable. When the RPD falls within 2 and 2.5, the model is capable of providing approximate quantitative predictions. Instead, an RPD higher than 2.5 indicates a high level of prediction accuracy (Nicolai et al., 2007).

These evaluations were conducted for the calibration set (using RMSE, Calibration Bias, and R^2), and the prediction set (external validation, evaluated with RMSEP, Prediction Bias, and R^2Pred). Moreover, the cross-validation strategy (internal validation) was used to choose the optimal number of latent variables (model complexity) to maximize the model reliability, balancing good predictions, and preventing overfitting (Tugnolo et al., 2021). Thus, an internal cross-validation set obtained using the venetian blinds method with 5 data splits was evaluated with RMSECV, CV Bias, and R^2CV.

Finally, to assess the contribution of each combination of LED-PD to build the six independent predictive models, the VIP (Variable Importance in the Projection) scores were calculated. The VIP score is the squared function of the PLS weights taking into account the amount of explained "y" (response variable) variance in each dimension. The VIP value is calculated for each variable (combination of LED-PD). Therefore, the VIP scores provide information about the significance of each variable on the latent variables (LV). A higher VIP score indicate increased importance of the corresponding variable. A threshold criterion for variable selection is set at VIP score values greater than 1.0, as these variables are deemed to have the most significant influence on the model (Pampuri et al., 2021b).

3. Results and discussions

3.1. Reference data analysis

This study covered the grape ripening evolution across the whole timeframe of the process, from veraison to harvest. Figs. 3 and 4 summarize the descriptive statistics (for each sampling time) related to the grape's wet-chemistry reference analysis (TSS, PA, pH, TA, TP, and EA) carried out on the two varieties considered (TN and TF). The mean, median, interquartile range (IQR), and data range were represented in the graphs. Moreover, potential outliers (observations beyond the data range whisker length) were also reported. By default, a potential outlier is a value that is more than 1.5 times the IQR (away from the bottom or top of the box). However, no reference sample was considered a true outlier to be discarded from the data set.

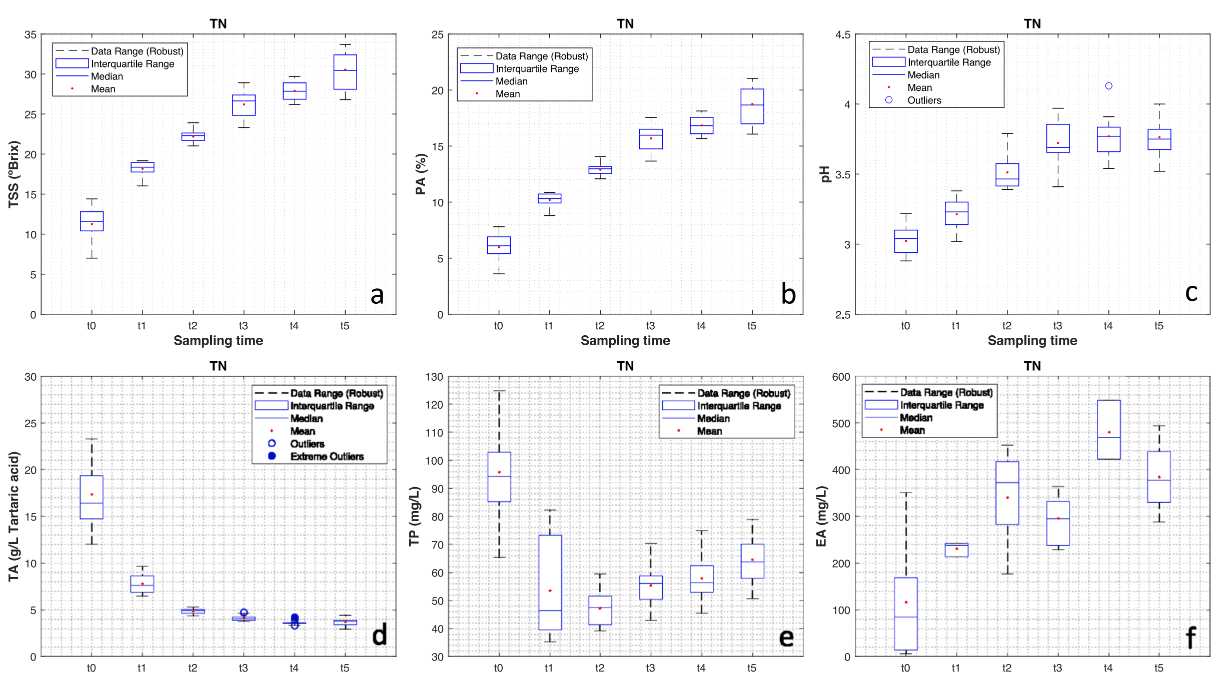

Fig. 3. Descriptive statistics of the reference analysis for TSS (a), PA (b), pH (c), TA (d), TP (e), and EA (f) obtained from *cv.* Touriga Nacional (TN) samples at each sampling time.

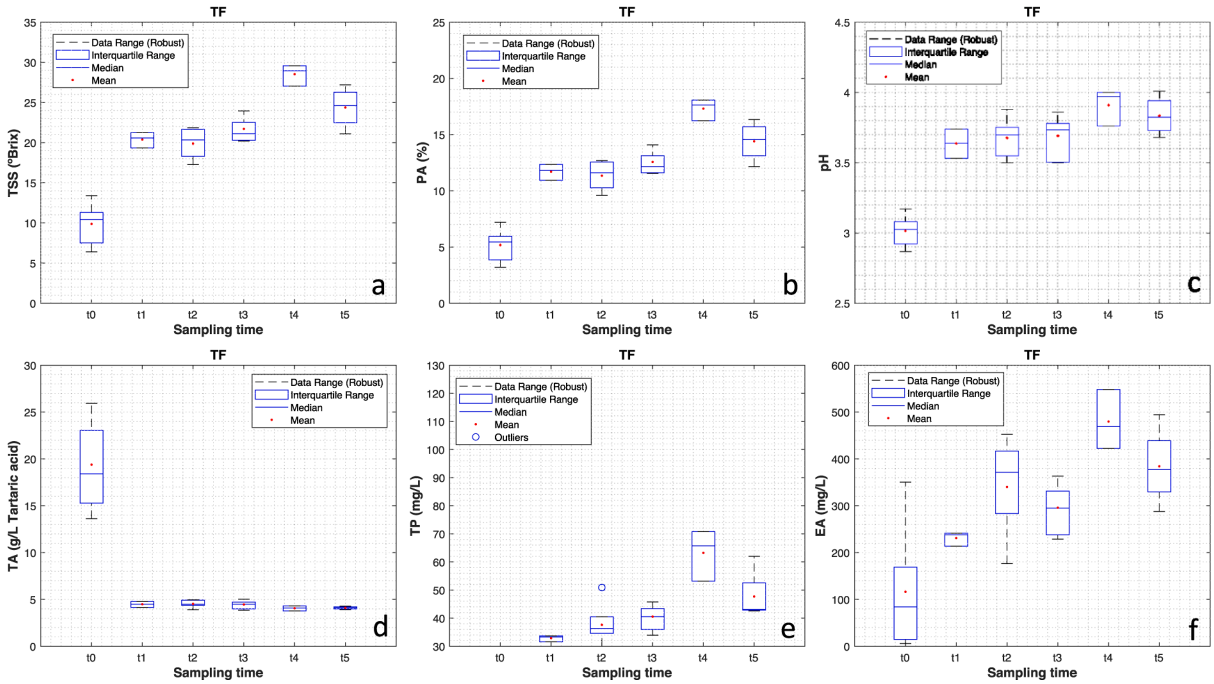

Fig. 4. Descriptive statistics of the reference analysis for TSS (a), PA (b), pH (c), TA (d), TP € and EA (f) obtained from *cv.* Touriga Franca samples at each sampling time.

Overall, for variety TN (Fig. 3), a clear trend over time for both technological and phenolic maturation parameters was observed. This fact can be associated with the sampling strategy design that consisted of multiple sample collections within the same parcels of the vineyards. TSS (Fig. 3a) ranged from 6.4 °Brix to 33.7 °Brix and it follows that PA (Fig. 3b) ranged from about 3.2 % to 21.04 %. Especially, for these two parameters, very high values were obtained at the end of the sampling campaign due to the use of this grape to produce fortified wine (e.g. Port wine).

Concerning TF variety (Fig. 4), the sampling campaign was performed in different vineyards, altitudes, and soil. Due to this additional geographic variability of the samples, a wider standard deviation in the maturation values was observed when compared to TN. Such geographical variability is also found at t4, where higher maturation values were obtained compared to t5 suggesting a sampling of grape fully ripe and ready to be harvested. Regarding TP (Fig. 4e), no samples were analysed at t0.

3.2. Optical data exploratory analysis

The lab optical acquisitions were managed considering the potential conditions that the sensor has to face once placed into the grape bunch in the field: (i) the noise produced by environmental light, (ii) the physical features of the grape berries, (iii) the increase in the size of the bunch, (iv) the variable optical gap and (v) the sensor position inside the grape bunch which may change during the ripening process. To minimize these issues and maximize the information collected, the measurements were performed using a case to reproduce a dark room to have similar field conditions to a sampling performed overnight in future application.

Firstly, a visual interpretation of the optical readouts acquired was performed. Fig. 5 shows the mean sensors' readouts (20 variables) at each sampling time coming from the signal recorded by each PD (labelled as ch1, ch2, ch3, and ch4 for the sake of simplicity). A more focused description of the optical signals at each sampling time has been reported in Figure S1. The measurement routine turns on one LED at a time starting following the order: 730, 690, 630, and 530 nm. Finally, the signal with all the switched LEDs off has been collected with the four PDs to detect background signals. Overall, the final output provides a complete profile that is comparable with the typical Vis/NIR spectrum of dark grape berries as reported by Pampuri et al. (2021a) and Giovenzana et al. (2018). Higher reflectance values associated with complete green samples (very unripe berries) were observed at the beginning of the

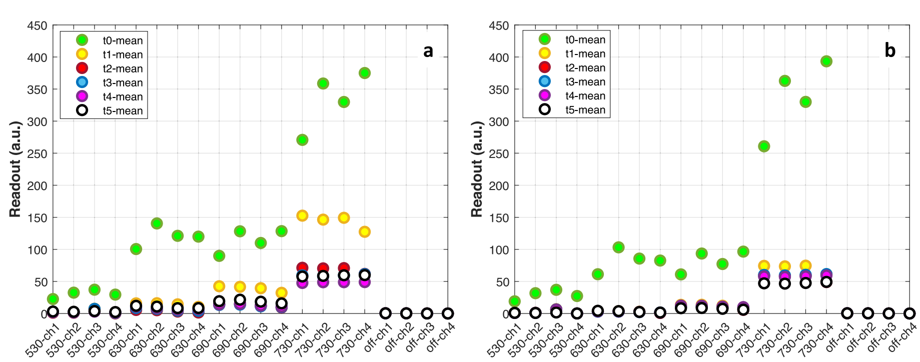

Fig. 5. Sensor readouts of each LED read by each photodetector (–ch) including the background condition with LED switched off for cv. Touriga Nacional mean optical outputs (a) and cv. Touriga Franca mean optical outputs (b).

ripening season (t0), which is due to the low concentration of anthocyanins. Then, as the ripening process progresses, the color evolution due to the accumulation of phenolic compounds leads to lower reflectance values (Bigard et al., 2019). Focusing on each variety (TN and TF), the mean reflectance impact for TN (Fig. 6a) remains qualitatively more distributed along the time than TF (Fig. 6b) where the mean signal after t0 remains almost stable, especially for the readouts at 530 and 630 nm.

Fig. 6 shows the PCA outcomes (scores and loadings) performed on the optical readouts after autoscaling (Log_{10} transformation has not been used during the exploratory analysis of the optical data). The score plot (Fig. 6a) shows the behaviour of the samples (labelled according to the sampling time) in the orthogonal space defined by PC1 and PC2. About 80 % of the total variability is described by PC1 which highlights the pigmentation process that occurred from t0 to t1/t2. This behaviour has been confirmed by the constant positive loadings of PC1 (Fig. 6b). Given the variables with the LED switched on, the primary factor describing this component is likely the decreasing interactance trend in the readout. This behavior is attributed to pigmentation, reflecting the increasing concentration of anthocyanins, which is closely related to the passage of time and the maturation process. Moreover, looking into the loadings of PC2, the readouts without illumination (the background) are the main descriptor of such principal component (almost 9 % of the total variability). Thus, since the data variance of the background is ascribable as a source of noise, this source of interference has been removed from the calibration dataset.

Therefore, the raw optical data matrix has been mathematically treated to reduce the background effects. In detail, for each readout with LED switched on (the reflectance readout of each LED at each channel), the corresponding background value was subtracted. Then, the new data matrix was re-explored to study the effects of pre-treatment, as well as to verify that the information contained is related to the chemical characteristics of the sample.

Fig. 7 shows the PCA outcomes performed on the pre-processed data matrix after readout transformation using Log_{10} and autoscaling (column-wise normalization). Fig. 7a shows the combined contribution of the PC1 and PC2 scores related to the two main information contained in the new dataset. Firstly, at positive values of PC1, a clear separation (from negative to positive values of PC3) of the samples (very unripe berries with green skin) at t0 according to the two varieties considered (TN and TF). Instead, at negative values of PC1, the maturation process after veraison has been highlighted from positive to negative values of PC3. No background contribution appeared on the three major components (Fig. 7c).

3.3. Models calculation

Following the promising outcomes from this initial phase of data processing exploration using PCA, six independent regression models were developed following the PLS method for predicting technological and phenolic grape ripening parameters using the optical readouts after Log10 transformation and autoscaling normalization. The PLS method has been chosen (between the various regression methods) considering the technical features of the spectral sensing system. PLS can analyse data with strongly collinear, noisy, and numerous X-variables (readouts) (Wold et al., 2001). At this stage, such an assumption fits properly with the proposed system as it can acquire optical readouts at 4 different wavelengths with 4 PDs providing at the end 16 outputs highly correlated where the PDs physical position (on the optical spot) and the grape geometry influence the optical information. PLS has proven to be a capable simple linear method to be able to minimize such sources of noise maximizing the covariance between the readouts and the six reference parameters.

To build robust models capable of predicting unknown samples, the original dataset of 106 samples was split (using the Kennard–Stone duplex algorithm) into two independent datasets: a calibration set (comprising 70 % of the samples) and a prediction set (with the remaining 30 % of the samples). The descriptive statistics of the total and split datasets are summarized in Table 2. Instead, the estimation performance of the prediction set is summarised in Table 3 (supplementary information about the figure of merits in calibration, cross-validation "CV" and prediction has been reported in Figures S2 and S3). Since the PLS method, like many other multivariate calibration methods (e.g., support vector machines and artificial neural networks), is prone to over-fitting or under-fitting (which may limit their predictive ability), the cross-validation step is essential to determine the correct number of latent variables to address these issues (Gowen et al., 2011). Indeed, the permutation of different sets of cross-validation provided RMSECV values that were comparable with the RMSEP values, suggesting a low risk of incorrect model fitting. Focusing on the model's performance in terms of RPD, promising results were obtained for the technological maturation parameters regression models (RPD > 2.2 for TSS, PA, TA, and pH). However, fairly lower performances (RPD of 1.28 and 1.55 for TA and EA, respectively) have been obtained for the phenolic maturation predictive models (compared with the technological ones) suggesting the capability of the models to only discriminate between low and high values of TP and EA. This poor predictive performance is remarkably noticeable in the moments close to the harvest where the high anthocyanin concentrations generate very low reflectance readouts (especially with the LED at 530 and 630 nm) that do not allow an accurate estimate of phenolic parameters in the crucial moment of the maturation process.

VIP scores were calculated (Fig. 8) from each independent PLS model. The VIP highlighted the crucial combination of LED-PD for predicting TSS, PA, TA, pH, TP, and EA. As expected, the LED and PD position influenced the model building. Indeed, although the LED illumination is the same for the quartets of PDs, the amount of photons received by each PD changes according to the proximity to the LED. Such an effect is well highlighted in almost all the models for the third

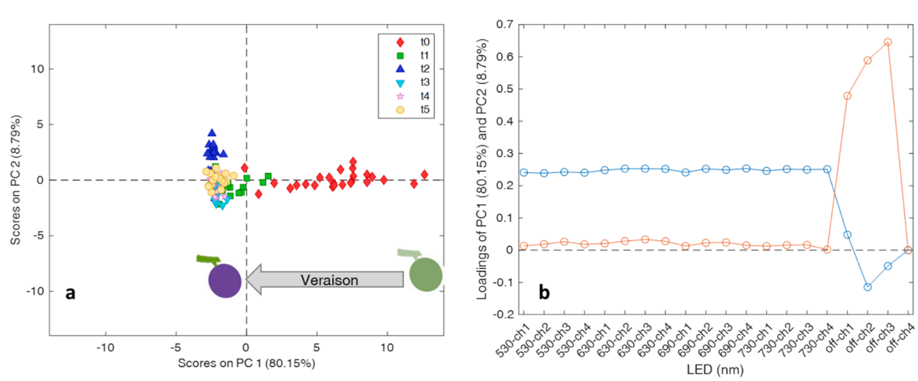

Fig. 6. PCA outcomes (scores and loadings) on spectral sensors readouts (raw data). (a) scores plot, labeled according to the sampling time, and (b) loadings plot of the first two principal components.

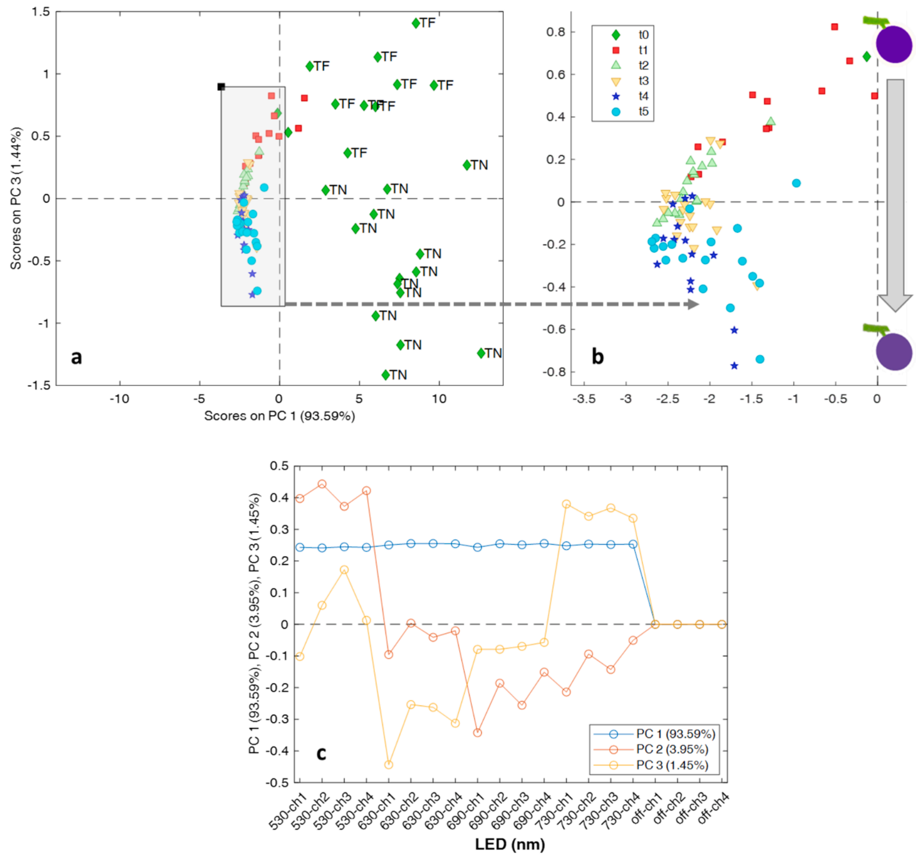

Fig. 7. PCA outcomes (scores and loadings) after data pre-processing: 8a) scores plot, colored according to the sampling time; (8b) scores plot inset; 8c) loadings plot of the three lower order components.

Table 2
Descriptive statistics of grape composition parameters after Kennard–Stone duplex algorithm dataset division*.

Parameter	Dataset				Calibration				Prediction			
	mean	std	n °	range	mean	std	n °	range	mean	std	n °	range
TSS (°Brix)	21.30	7.19	106	6.40–33.70	20.83	7.39	75	7.00–33.70	22.45	6.64	31	6.40–33.10
PA (%)	12.47	4.73	106	3.20–21.04	12.16	4.87	75	3.60–21.04	13.22	4.34	31	3.20–20.61
TA (g/L $_{\text{Tartaric acid}}$)	7.84	6.18	106	2.94–25.93	8.35	6.27	75	2.94–23.26	6.60	5.87	31	3.38–25.93
pH	3.50	0.34	106	2.87–4.13	3.46	0.35	75	2.87–4.13	3.58	0.30	31	2.92–4.01
TP (mg/L)	59.52	20.40	96	25.80–124.70	61.37	21.61	68	25.80–124.70	55.03	16.58	28	31.60–102.80
EA (mg/L)	333.43	166.55	106	3.48–708.06	317.92	167.96	75	3.48–656.99	370.95	159.52	31	5.80–708.06

*TSS = total solids soluble; PA = potential alcohol; TA = titratable acidity; TP = total polyphenols; EA = extractible anthocyanins; std = standard deviation; n° = number of samples.

Table 3
Figures of merit of the PLS models. *.

Qualitative parameter	LVs	RMSE	R^2	RMSEP	R^2 Pred	Bias Pred	RPD
TSS (°Brix)	4	1.95	0.92	2.41	0.86	−0.07	2.76
PA (%)	4	1.33	0.92	1.68	0.85	0.03	2.58
TA (g/L $_{\text{Tartaric acid}}$)	4	1.35	0.95	1.61	0.92	−0.26	3.65
pH	4	0.12	0.89	0.13	0.79	0.005	2.27
TP (mg/L)	2	11.03	0.73	12.98	0.51	−5.82	1.28
EA (mg/L)	4	72.2	0.81	103	0.58	−15.13	1.55

* LVs = number of latent variables; RMSE = root mean square error of calibration; R^2 = coefficient of determination; RMSEP = root mean square error of cross-validation; R^2 Pred = coefficient of determination in Prediction; Bias Pred = Prediction Bias; RPD = residual prediction deviation.

channel (ch3) at 530 nm.

Overall, the wavelength at 730 nm holds major importance in constructing quite all the models. Instead, at 530 nm, while it frequently surpasses the threshold value of 1, its significance is attributed solely to channel number three (ch3), which receives more light than the others. This implies that enhancing the power of the green LED (530 nm) could potentially be beneficial for enhancing chemical information collection at this wavelength and, consequently, improving the predictive performance of the models. Only one exception is made for the TP model (the worst one) where 530-ch1, 530-ch2 and 530-ch4 seems to be more informative than the third one (530-ch3). Such a result combined with the poor performance of the model suggests a significant presence of noise in the optical data mainly driven by the LED at 530 nm. Its inefficiency in capturing chemical information related to phenolic maturation can be ascribable to potentially low power and diminished effectiveness caused by the purplish colour development.

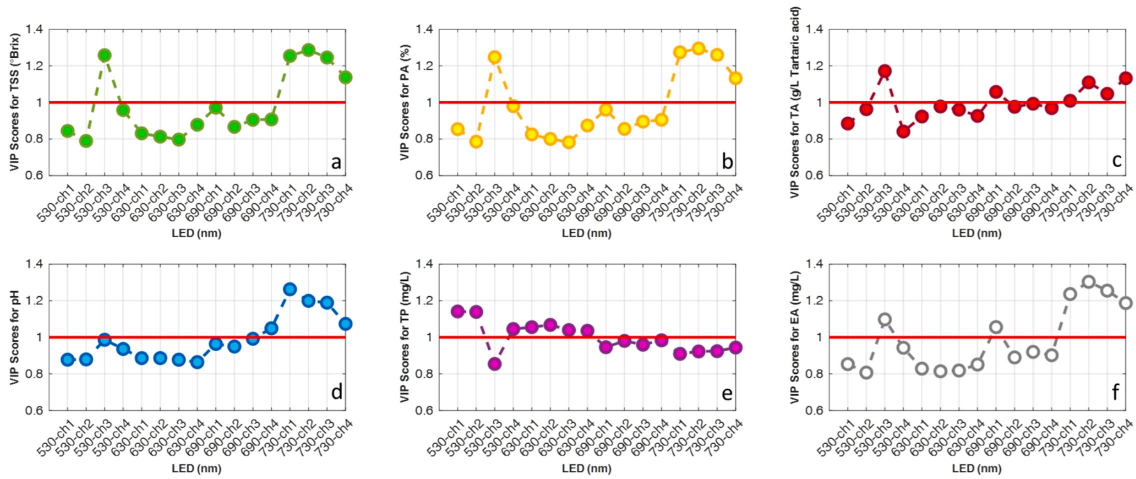

Fig. 8. Variable importance in projection scores of the PLS models for TSS (a), PA (b), TA (c), pH (d), TP and EA (f).

3.4. Comparison with the state of the art

Several analytical approaches have been sought for effective and non-destructive methods that could explore a large number of samples and provide a rapid and comprehensive overview of the ripening process. Indeed, the scientific literature reports different applications where the advantages of optics, initially using benchtop instruments and then handheld and/or autonomous devices are widely explored (Tardaguila et al., 2021). Technological advancement and cost reduction (due to the massification of miniaturized optical components) have driven many fields of application, including viticulture (Krause et al., 2021). Certainly, the performance of these devices varies according to their technical features but also to the method of sampling and model development.

The outcomes attained in this study are comparable with those found in the literature using full-range Vis/NIR spectrometers. In 2015, Giovenzana et al. obtained comparable results using a portable Vis/NIR device (400–1000 nm Jaz, Ocean Optics, Dunedin, FL) for the development of PLS models (RPD of 2.26, 2.66, 1.61, 1.51, and 1.98 for the estimation of TSS, TA, PA, EA, and TP, respectively).

One step forward has been given by Fernández-Novales, et al. (2019), who reported the quantification of TSS, anthocyanins, and TP in grape berries under field conditions using an on-the-go Vis/NIR spectroscopic system (570 and 990 nm, Polytec GmbH, Waldbronn, Germany) acquiring from a moving platform at 0.30 m of distance. The authors provided a real alternative application with a significant predictive performance for two of the three parameters under study (RPD of 3.52, 2, and 1.31, for the estimation of TSS, anthocyanins, and TP, respectively). This approach was capable of non-destructively appraising and mapping the vineyard grape composition variability with a high spatial and temporal resolution. However, the topography is an important factor that (in some extreme hilly and mountainous conditions) could reduce the mapping effectiveness of such an approach. More recently, Pampuri et al. (2021a) proposed a compact optical device (covering 12 wavelengths, from 450 nm to 860 nm) based on LEDs and multi-spectral sensors that integrate Gaussian filters into standard CMOS silicon via nano-optic deposited interference filter technology (AMS, models AS7262 visible and AS7263 NIR, Premstaetten, Austria). Although the device presented by Pampuri et al. (2021a) and the ones presented in this work share some similar acquisition features, the results showed a lower predictive performance (RPD of 1.45, 1.44, 1.33 for the estimation of TSS, TA, and pH, respectively) when compared to the results presented in Table 3. This suggests improvements in both technological aspects (customization of the device's architecture and definition of the optical bands) and methodological aspects (sampling and modelling). It is important to note that none of these instruments were capable of autonomous operation, potentially delivering useful information with acceptable uncertainty across both spatial and temporal resolutions. Moreover, the storing of such additional information over the years contributes to the development of an historical dataset that helps monitor climatic impacts, optimize quality, plan resources, support research, enhance marketing, and ensure regulatory compliance, ultimately improving wine production and sustainability.

4. Conclusions and future perspectives

In this work, an optical prototype based on spectral sensing technology (monitoring of four optical bands in Vis/NIR range), for grape maturation control was used for the prediction of the technological parameters of grape's ripening at a lab-scale. The readouts acquired with the miniaturized sensor provided reliable data to develop prediction models for the target parameters.

Overall, predictive models with good performance were obtained for the estimation of the technological ripening parameters (RPD = 2.76 and $R^2 = 0.86$ for TSS; RPD = 2.58 and $R^2 = 0.85$ for PA; RPD = 3.65 and $R^2 = 0.92$ for TA; RPD = 2.27 and $R^2 = 0.79$ for pH), establishing a solid ground for the application of this sensing strategy in the field. For the phenolic parameters (TP and EA), the performance of the models is still insufficient (RPD = 1.28 and $R^2 = 0.51$ for TP; RPD = 1.55 and $R^2 = 0.58$ for EA), which is consistent with previous works reported in the literature (Kemps et al., 2010; Guidetti et al., 2010; Aleixandre-Tudo et al., 2019). An approach based on fluorescence signals (Agati et al., 2013) instead of the diffuse reflectance based on Vis/NIR spectral bands used here can be the key to improving this lower performance.

This new generation of spectral sensing devices can be viewed as a starting point of a new concept of cost-effective IoT end-nodes which could be distributed in the vineyard at fruit-set (open bunch) and that will then be enveloped by the grape bunch as the grape grows on it. Its stand-alone operation makes it possible to acquire optical data and predict the most important ripening parameters directly from the field measurements taken inside the bunch.

Further developments of the work described here will include the expansion of the prediction models to other grape varieties (including white grapes), as well as the definition of the data pipeline for processing field data. Moreover, several potential challenges in the system's field operation can be expected. The first challenge pertains to the sensor's positioning, which may change throughout the season due to alterations in the grape berry, such as dehydration or varying weather conditions. Another issue is determining the number of sensors needed to monitor a specific area, which remains unresolved. Finally, applying prediction models developed in one crop season to the following year might be problematic if there are substantial differences in the grape's ripening progression between years.

CRediT authorship contribution statement

Alessio Tugnolo: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Data curation. **Hugo M. Oliveira:** Writing – review & editing, Visualization, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Valentina Giovenzana:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Formal analysis. **Natacha Fontes:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization. **Sara Silva:** Writing – review & editing, Data curation. **Cristina Fernandes:** Validation, Supervision, Resources. **António Graça:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization. **Alessia Pampuri:** Writing – review & editing, Methodology, Investigation, Formal analysis. **Andrea Casson:** Writing – review & editing, Investigation, Formal analysis. **João Piteira:** Writing – review & editing, Supervision, Software, Project administration, Methodology, Funding acquisition, Conceptualization. **Paulo Freitas:** Writing – review & editing, Validation, Supervision, Project administration, Funding acquisition, Conceptualization. **Riccardo Guidetti:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization. **Roberto Beghi:** Writing – review & editing, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Funding

This research has been founded by European Horizon 2020, project i-GRAPE (grant N° 825521).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors wish to thank Prof. Paulo Freitas for the supervision of the research projects. The authors also acknowledge the i-Grape consortium for the fruitful discussions and inputs.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compag.2024.109856.

Data availability

Data will be made available on request.

References

Agati, G., Meyer, S., Matteini, P., Cerovic, Z.G., 2007. Assessment of anthocyanins in grape (Vitis vinifera L.) berries using a noninvasive chlorophyll fluorescence method. J. Agric. Food Chem. 55 (4), 1053–1061.
Agati, G., D'Onofrio, C., Ducci, E., Cuzzola, A., Remorini, D., Tuccio, L., Lazzini, F., Mattii, G., 2013. Potential of a multiparametric optical sensor for determining in situ the maturity components of red and white Vitis vinifera wine grapes. J. Agric. Food Chem. 61 (50), 12211–12218.
Aleixandre-Tudo, J.L., Nieuwoudt, H., du Toit, W., 2019. Towards on-line monitoring of phenolic content in red wine grapes: A feasibility study. Food Chem. 270, 322–331.
Bigard, A., Romieu, C., Sire, Y., Veyret, M., Ojeda, H., Torregrosa, L., 2019. The kinetics of grape ripening revisited through berry density sorting. OENO One 53 (4).
Casson, A., Beghi, R., Giovenzana, V., Fiorindo, I., Tugnolo, A., Guidetti, R., 2019. Visible Near Infrared Spectroscopy as a Green Technology: An Environmental Impact Comparative Study on Olive Oil Analyses. Sustainability 11 (9), 2611.
Cavallo, C., Sacchi, G., Carfora, V., 2020. Resilience effects in food consumption behaviour at the time of Covid-19: Perspectives from Italy. Heliyon 6 (12).
Eurostat. (2019). https://ec.europa.eu/eurostat/web/products-eurostat-news/-/edn-20201119-2.
Fernández-Novales, J., Tardáguila, J., Gutiérrez, S., Diago, M.P., 2019. On-the-go vis+sw− nir spectroscopy as a reliable monitoring tool for grape composition within the vineyard. Molecules 24 (15), 2795.
Giovenzana, V., Beghi, R., Malegori, C., Civelli, R., Guidetti, R., 2014. Wavelength selection with a view to a simplified handheld optical system to estimate grape ripeness. Am. J. Enol. Vitic. 65 (1), 117–123.
Giovenzana, V., Civelli, R., Beghi, R., Oberti, R., Guidetti, R., 2015. Testing of a simplified LED based VNIR system for rapid ripeness evaluation of white grape (Vitis vinifera L.) for Franciacorta wine. Talanta 144, 584–591.
Giovenzana, V., Beghi, R., Tugnolo, A., Brancadoro, L., Guidetti, R., 2018. Comparison of two immersion probes coupled with visible/near infrared spectroscopy to assess the must infection at the grape receiving area. Comput. Electron. Agric. 146, 86–92.
Gowen, A.A., Downey, G., Esquerre, C., O'donnell, C.P., 2011. Preventing over-fitting in PLS calibration models of near-infrared (NIR) spectroscopy data using regression coefficients. J. Chemom. 25 (7), 375–381.
Guidetti, R., Beghi, R., Bodria, L., 2010. Evaluation of grape quality parameters by a simple Vis/NIR system. Trans. ASABE 53 (2), 477–484.
Kemps, B., Leon, L., Best, S., De Baerdemaeker, J., De Ketelaere, B., 2010. Assessment of the quality parameters in grapes using VIS/NIR spectroscopy. Biosyst. Eng. 105 (4), 507–513.
Kennard, R.W., Stone, L.A., 1969. Computer aided design of experiments. Technometrics 11, 137–148.
Kontoudakis, N., Esteruelas, M., Fort, F., Canals, J.M., De Freitas, V., Zamora, F., 2011. Influence of the heterogeneity of grape phenolic maturity on wine composition and quality. Food Chem. 124 (3), 767–774.
Krause, J., Grüger, H., Gebauer, L., Zheng, X., Knobbe, J., Pügner, T., Kicherer, A., Gruna, R., Längle, T., Beyerer, J., 2021. SmartSpectrometer—embedded optical spectroscopy for applications in agriculture and industry. Sensors 21 (13), 4476.
Matese, A., Filippo Di Gennaro, S., 2015. Technology in precision viticulture: A state of the art review. International journal of wine research 69–81.
Nicolai, B.M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K.I., Lammertyn, J., 2007. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. Postharvest Biol. Technol. 46 (2), 99–118.
OIV, 2022. Compendium of International Methods of Wine and Must Analysis, Paris.
Oliveira, H.M., Tugnolo, A., Fontes, N., Marques, C., Geraldes, Á., Jenne, S., Zappe, H., Graça, A., Giovenzana, V., Beghi, R., Guidetti, R., Piteira, J., Freitas, P., 2024. An autonomous Internet of Things spectral sensing system for in-situ optical monitoring of grape ripening: design, characterization, and operation. Comput. Electron. Agric. 217. https://doi.org/10.1016/j.compag.2023.108599.
Oliveri, P., Malegori, C., Mustorgi, E., & Casale, M. (2020). Application of chemometrics in the food sciences.
Pampuri, A., Tugnolo, A., Giovenzana, V., Casson, A., Guidetti, R., Beghi, R., 2021a. Design of cost-effective LED based prototypes for the evaluation of grape (Vitis vinifera L.) ripeness. Comput. Electron. Agric. 189, 106381.
Pampuri, A., Tugnolo, A., Bianchi, D., Giovenzana, V., Beghi, R., Fontes, N., Hugo, M.O., Casson, A., Lucio, B., Guidetti, R., 2021b. Optical specifications for a proximal sensing approach to monitor the vine water status in a distributed and autonomous fashion. Biosyst. Eng. 212, 388–398.
Ribéreau-Gayon P., Dubourdieu D., Donèche B., Lonvaud A., 1998. Traité d'Œnologie Tome 1: Microbiologie du Vin, Vinifications. Dunod, Paris.
Ribéreau-Gayon, P., Glories, Y., Maujean, A., & Dubourdieu, D. (Eds.). (2006). Handbook of enology, volume 2: the chemistry of wine-stabilization and treatments (Vol. 2). John Wiley & Sons.
Tardaguila, J., Stoll, M., Gutiérrez, S., Proffitt, T., Diago, M.P., 2021. Smart applications and digital technologies in viticulture: A review. Smart Agric. Technol. 1, 100005.
Tugnolo, A., Giovenzana, V., Beghi, R., Grassi, S., Alamprese, C., Casson, A., Casiraghi, E., Guidetti, R., 2021. A diagnostic visible/near infrared tool for a fully automated olive ripeness evaluation in a view of a simplified optical system. Comput. Electron. Agric. 180, 105887.
Urraca, R., Sanz-Garcia, A., Tardaguila, J., Diago, M.P., 2016. Estimation of total soluble solids in grape berries using a hand-held NIR spectrometer under field conditions. J. Sci. Food Agric. 96 (9), 3007–3016.
Vallone, M., Alleri, M., Bono, F., Catania, P., 2019. Quality evaluation of grapes for mechanical harvest using vis NIR spectroscopy. Agric. Eng. Int. CIGR J. 21 (1), 140–149.
Vrochidou, E., Bazinas, C., Manios, M., Papakostas, G.A., Pachidis, T.P., Kaburlasos, V.G., 2021. Machine vision for ripeness estimation in viticulture automation. Horticulturae 7 (9), 282.
Wold, S., Sjöström, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. Chemom. Intel. Lab. Syst. 58 (2), 109–130.