



# Infrared Spectroscopy and Chemometric Applications for the Qualitative and Quantitative Investigation of Grapevine Organs

Elizma van Wyngaard<sup>1</sup>, Erna Blancquaert<sup>1</sup>, H el ene Nieuwoudt<sup>1</sup> and Jose Luis Aleixandre-Tudo<sup>1,2\*</sup>

<sup>1</sup> South African Grape and Wine Research Institute (SAGWRI), Department of Viticulture and Oenology, Stellenbosch University, Stellenbosch, South Africa, <sup>2</sup> Instituto de Ingenier a de Alimentos para el Desarrollo (IIAD), Departamento de Tecnolog a de Alimentos, Universidad Polit cnica de Valencia, Valencia, Spain

## OPEN ACCESS

### Edited by:

Marcos Egea-Cortines,  
Universidad Polit cnica de Cartagena,  
Spain

### Reviewed by:

Federico Marini,  
Sapienza University of Rome, Italy  
Omar Vergara-D az,  
Universidade Nova de Lisboa,  
Portugal

### \*Correspondence:

Jose Luis Aleixandre-Tudo  
joaltu@sun.ac.za

### Specialty section:

This article was submitted to  
Technical Advances in Plant Science,  
a section of the journal  
Frontiers in Plant Science

**Received:** 10 June 2021

**Accepted:** 09 August 2021

**Published:** 03 September 2021

### Citation:

van Wyngaard E, Blancquaert E, Nieuwoudt H and Aleixandre-Tudo JL (2021) Infrared Spectroscopy and Chemometric Applications for the Qualitative and Quantitative Investigation of Grapevine Organs. *Front. Plant Sci.* 12:723247. doi: 10.3389/fpls.2021.723247

The fourth agricultural revolution is leading us into a time of using data science as a tool to implement precision viticulture. Infrared spectroscopy provides the means for rapid and large-scale data collection to achieve this goal. The non-invasive applications of infrared spectroscopy in grapevines are still in its infancy, but recent studies have reported its feasibility. This review examines near infrared and mid infrared spectroscopy for the qualitative and quantitative investigation of intact grapevine organs. Qualitative applications, with the focus on using spectral data for categorization purposes, is discussed. The quantitative applications discussed in this review focuses on the methods associated with carbohydrates, nitrogen, and amino acids, using both invasive and non-invasive means of sample measurement. Few studies have investigated the use of infrared spectroscopy for the direct measurement of intact, fresh, and unfrozen grapevine organs such as berries or leaves, and these studies are examined in depth. The chemometric procedures associated with qualitative and quantitative infrared techniques are discussed, followed by the critical evaluation of the future prospects that could be expected in the field.

**Keywords:** infrared spectroscopy, chemometrics, grapevine organs, non-invasive, carbohydrates, nitrogen, amino acids

## INTRODUCTION

The agricultural sector is entering its fourth revolution, thus moving toward more sustainable farming practices. Improved sustainability is achieved through the implementation of digital technologies and precision farming or precision viticulture (Lopo et al., 2015). The practice of precision viticulture can be used to successfully manage the challenges of global competition, decreasing natural resources, and increasing environmental pressures (De Ordu a, 2010; Fraga et al., 2012). This practice uses innovative technologies and data science, of which infrared spectroscopy is an example.

The application of infrared technologies in the agricultural sector can supply the means to measure intact, fresh, and unfrozen samples directly, and implement precision viticulture successfully. Infrared spectroscopy is becoming increasingly popular in the agricultural

and the agri-food industries for its capability to supply rapid results while remaining cost-effective (Lopo et al., 2015; Diago et al., 2018; Cuq et al., 2020). Infrared spectroscopy provides an invaluable tool for qualitative and quantitative applications in plant and food production. This review article will present an overview of infrared spectroscopy applications, focusing on near infrared (NIR) and mid infrared (MIR) spectroscopy, (i) qualitative investigation of grapevine organs will be examined. Then, (ii) discussing the quantification of specifically carbohydrates, nitrogen, and amino acids using infrared technologies, (iii) assessing the chemometric applications used for qualitative and quantitative analysis and associated performance evaluation indices, and (iv) concluding with future prospects anticipated in the field of infrared spectroscopy applications for viticultural investigations.

## QUALITATIVE SPECTRAL INVESTIGATION OF GRAPEVINE ORGANS

Infrared spectroscopy provides information on multiple properties of a sample simultaneously and therefore is considered a fingerprinting technique (Cozzolino, 2014; Dos Santos et al., 2017). When looking at complex organic samples, this spectral fingerprint subjected to chemometric methods can be used to investigate and elucidate compositional characteristics in the sample, as well as the relationship between several metabolites in the plant. Additionally, important information about the similarities or dissimilarities of groups can emerge (Cozzolino et al., 2011; Dos Santos et al., 2017). Thus, infrared spectroscopy technologies are a valuable tool for data acquisition.

A study conducted by Lopo et al. (2015) used a portable NIR instrument to scan fresh soil and leaf samples to identify soil type. Using principal component analysis (PCA) and supervised partial least squares discriminant analysis (PLS-DA) the authors grouped the soil samples according to soil type. The leaf samples showed similar separation which correlated to soil type and indicated the feasibility of scanning fresh leaves for soil type determination. PCA was used to extract groupings while PLS-DA was used for model development to determine soil type (Lopo et al., 2015). This study indicated the benefit of using infrared spectroscopy for the direct measurement of fresh grapevine material.

Musingarabwi et al. (2016) studied Sauvignon blanc berry samples at five distinct phenological stages namely green, pré-véraison, véraison, post-véraison, and ripe using NIR and MIR spectroscopy. Berry samples were analyzed as fresh and frozen, and homogenized to a pulp. The best separation between phenological stages was seen using PCA for fresh homogenized samples measured with MIR. Additionally, separation was also seen using orthogonal PLS-DA for NIR and MIR data. This study linked the absorption bands associated with sugars and organic acids to the separation seen between phenological stages. However, there was some overlapping of certain phenological stages, specifically post-véraison and

ripe, that showed some degree of similarity between these stages. Although homogenized samples were used in this study, the fresh samples showed the best separation (Musingarabwi et al., 2016). The results showed that using fresh samples could be beneficial and valuable. This study also demonstrated the capability of infrared spectroscopy to detect variability throughout the growing season.

Furthermore, Dos Santos Costa et al. (2019) investigated the development of classification models for three maturation stages for Shiraz and Cabernet Sauvignon berries. Whole berries were scanned throughout the growing season using visible/near (Vis-NIR) infrared spectroscopy. PCA was used to identify the clustering of the samples based on maturation stages. Although some overlapping was seen, three stages were identified namely green, véraison, and ripe. Once these clusters were identified, supervised PLS-DA models were compiled for classification of grapes according to maturation stage. Maturation stages could be successfully predicted with 93.15% accuracy (Dos Santos Costa et al., 2019). These results not only emphasized that infrared spectroscopy can detect changes occurring during the growing season but also that these changes could be predicted and monitored.

Additionally, Cuq et al. (2020) sampled fresh berries at two phenological stages, pea-size and véraison. Fresh leaves were also sampled at véraison and separated into leaf blades and leaf petioles. PCA was performed on the data and clear separation was seen between the four sample groups. The PCA loadings showed spectral regions that discriminated between the grapevine organs. The same regions were identified when the spectra of each organ was averaged and compared directly (Cuq et al., 2020). Although the results showed that spectral regions could be used to distinguish between organs, the wavenumbers or regions associated with each organ type were not specified.

Two studies measuring dried and ground grapevine samples used PCA and found separation based on grapevine organs (Schmidtke et al., 2012; De Bei et al., 2017). Schmidtke et al. (2012) used MIR and identified separation between trunk and root samples. Separation along PC1 positively correlated to regions from 1,650 to 1,550  $\text{cm}^{-1}$  with smaller contributions from 1,500 to 1,300  $\text{cm}^{-1}$ , while PC2 corresponded with the region between 1,000 and 875  $\text{cm}^{-1}$ . De Bei et al. (2017) used NIR spectroscopy and noted that the spectra looked similar for trunk and leaf samples with prominent peaks at 4,300, 5,200, and 7,000 but that leaf samples had an additional peak at 5,900  $\text{cm}^{-1}$ . Although certain wavenumbers associated with grapevine organ groupings were identified in these studies, multiple wavenumbers and regions need to be examined in future research.

A study using Vis-NIR investigated the spectra and response of basal, young, and apical leaves on iron deficiency in young rootstock vines (Rustioni et al., 2017). The spectra were transformed, normalized, and compared directly. The leaves had responded differently to the iron deficiency and these differences were evident in the spectra. Not only were differences seen between leaves from the same shoot, but differences between the veins and interveinal area of a specific leaf were also observed. These differences were mostly based on the chlorophyll synthesis

and pigmentation distribution in the leaves (Rustioni et al., 2017). The results indicated that spectroscopy is able to detect subtle differences between leaves of the same shoot and even variation within one leaf.

Recently, non-linear methods were also used for classification purposes (Fuentes et al., 2018; Murru et al., 2019). Grape samples have been classified according to variety and ripeness using Fourier transform infrared spectroscopy (FTIR) and artificial neural networks (ANN). The specific compounds influencing the classification were identified (Murru et al., 2019). Machine learning algorithms together with NIR spectra of grapevine leaves were utilized to compile ANN models for cultivar classification. Classification with 92% accuracy was achieved leading to enhanced capabilities for ampelography (Fuentes et al., 2018). Both linear and non-linear methods need to be considered in future spectral investigations of grapevine organs for increased knowledge.

The studies discussed in this section have reported separation between grapevine organs or phenological stages (Schmidtke et al., 2012; Musingarabwi et al., 2016; De Bei et al., 2017; Dos Santos Costa et al., 2019; Cuq et al., 2020). However, the reasons for the differences in spectral properties observed for the grapevine organs or phenological stages were not fully investigated and could have provided interesting insights. Numerous changes occur in grapevine organs throughout the growing season and lead to large heterogeneity between organs and phenological stages (Hunter et al., 1995; Zapata et al., 2004; Holzapfel et al., 2010; Rossouw et al., 2017). Infrared spectroscopy could be used to monitor changes throughout the growing season and link the changes to specific spectral regions or wavenumbers of interest. Infrared spectroscopy could also be used to identify the characteristic spectral properties of grapevine organs at different phenological stages. Currently, available literature shows a lack of interpretation of the spectral differences perceived for various grapevine organs or phenological stages. More research is required to investigate the reason for the changes in spectral properties throughout the growing season.

## SPECTROSCOPY TECHNIQUES FOR QUANTIFICATION OF CARBOHYDRATES, NITROGEN, AND AMINO ACIDS IN GRAPEVINES

Although the studies discussed above reported qualitative methods to investigate grapevine organs, most of the same studies focused on quantification methods (Schmidtke et al., 2012; De Bei et al., 2017; Cuq et al., 2020). The following sections focus on the quantification of carbohydrates, nitrogen, and amino acids, henceforth referred to as key metabolites, using infrared spectroscopy methods. Firstly, the importance and role of key metabolites will be discussed. Next, the non-direct approaches will be evaluated, followed by the proposed methods using direct, non-invasive measurement of fresh grapevine material.

## Importance of Carbohydrates, Nitrogen, and Amino Acid Analysis in Grapevines

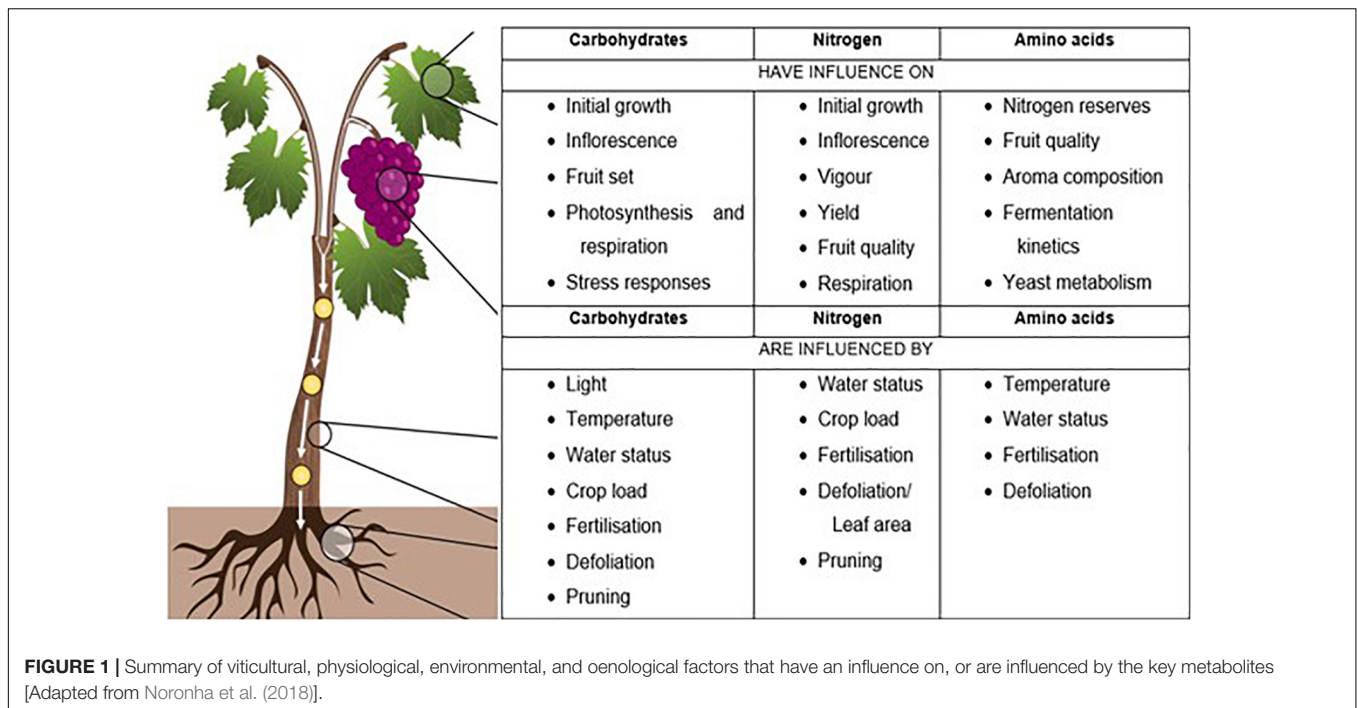
Carbohydrates, nitrogen, and amino acids form a key part of grapevine physiology affecting growth, yield, and grape quality (Holzapfel et al., 2010; Rossouw et al., 2017). The grapevine source-sink balance causes the concentration of these metabolites to continuously change throughout the growing season (Hunter et al., 1995; Zapata et al., 2004; Rossouw et al., 2017). Carbohydrate and nitrogen reserves play an integral role in vegetative growth and fruiting responses, and are influenced by various factors (Schmidtke et al., 2012; Li-Mallet et al., 2016; Rossouw et al., 2017). The amino acid composition of grape must affects fermentation kinetics, yeast metabolism, and aroma composition. Thus, amino acid content contributes directly to wine quality (Fernández-Novales et al., 2019). **Figure 1** summarizes the numerous factors that contribute to carbohydrate, nitrogen, and amino acid concentrations in grapevines.

Despite the importance of these key metabolites, the current analysis methods are costly and time-consuming (Edwards et al., 2011; López et al., 2012; Schmidtke et al., 2012). Using wet chemistry methods to monitor compounds continuously throughout the growing season is not yet feasible. By the time the results are obtained with these methods, concentrations in the grapevine would have changed. The measurement of the key metabolites during the growing season in berries, shoots and leaves could provide valuable information to aid fertilization, irrigation, canopy management, and winemaking decisions (Fernández-Novales et al., 2019; Cuq et al., 2020). New methods for quantifying these metabolites need to be investigated to increase our knowledge, facilitate management decisions, and implement precision viticultural practices.

## Non-direct Quantification Using Infrared Spectroscopy

The quantification of carbohydrate and nitrogen reserves using infrared spectroscopy with non-direct methods was investigated (Schmidtke et al., 2012; Smith et al., 2014; De Bei et al., 2017; Jones et al., 2020). Non-direct methods still use destructive sampling where samples need to be removed in the field and transported to laboratories. The methods also use extensive sample preparation, where the sample material needs to be freeze-dried to remove moisture, and then ground to a powder to obtain a completely homogenized sample. Favorable results were shown for numerous infrared spectroscopy applications and multivariate regression techniques to quantify chemical components in grapevine samples (Schmidtke et al., 2012; Smith et al., 2014; De Bei et al., 2017; Jones et al., 2020).

The use of attenuated total reflectance Fourier transform infrared spectroscopy (ATR-FT-IR) was investigated to assess nitrogen and starch reserves in grapevine wood and root samples (Schmidtke et al., 2012). Over a thousand samples were selected spanning over four vintages, two countries, five locations, and three cultivars, leading to a suitably representative dataset. Cross validation was used with a data split of ten for the calibration subsets. Accurate models were obtained with



PLS regression and support vector machine (SVM) regression using dried and powdered samples (Schmidtke et al., 2012). Although sample preparation was still used, the improvement from the existing wet chemistry methods was substantial. This research demonstrated that infrared spectroscopy could be used to develop accurate calibrations for complex compounds across various vintages and locations. Both linear and non-linear regression techniques performed equally well with SVM regression leading to marginally improved models. This study also showed the diverse applications of infrared spectroscopy when sufficient variability is included during the model development stage (Schmidtke et al., 2012).

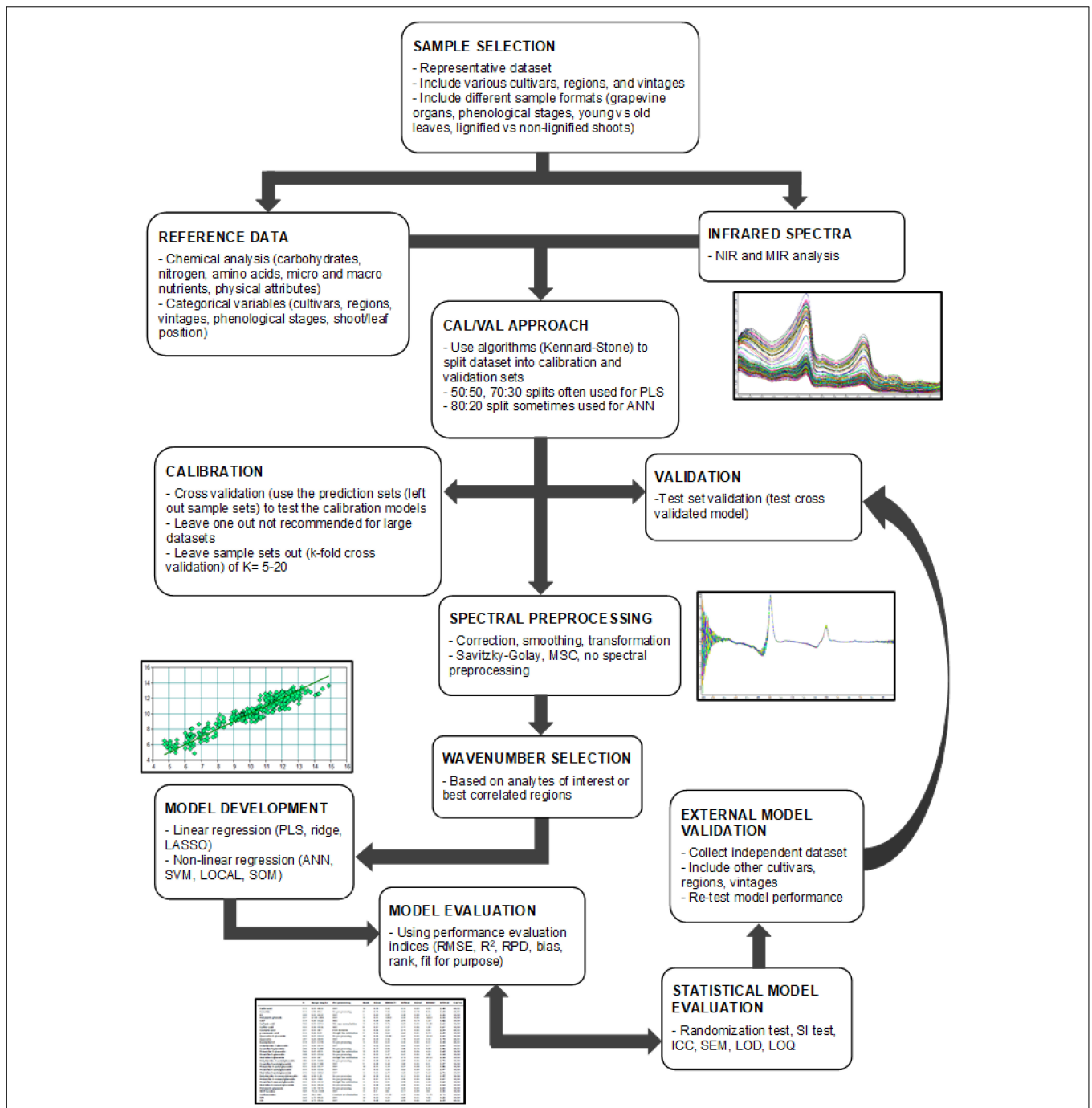
A related study used NIR spectroscopy to assess non-structural carbohydrates in grapevine trunks and leaves (De Bei et al., 2017). Models were obtained for trunk and leaf samples separately for starch, sugar, and total non-structural carbohydrates using PLS regression. Test set validation was used with the calibration set corresponding to two thirds of the dataset, and validation consisting of one third. Sampling was performed at three phenological stages and clear separation was seen between the phenological stages in the trunk samples' PCA plot. However, during model development the data for all the stages were combined, although separate models were developed for trunk and leaf samples. The study reported that the models could not provide quantifiable data but that the practical viticultural applications were still significant. The sample set for this study was limited with only 261 trunk samples and 222 leaf samples included. The samples were all collected from one cultivar and vineyard on which irrigation treatments were tested (De Bei et al., 2017). In future research several cultivars and locations could be investigated to increase dataset representativeness as well as model variability

and accuracy. Although a limited number of samples could be used to compile sufficient models, special care needs to be taken to ensure that the samples account for the largest possible variability and are representative. Additionally, the irrigation treatment could have negatively influenced the model performance.

A recent study investigated the feasibility of using NIR reflectance spectroscopy to quantify starch in ground and intact grapevine cane samples (Jones et al., 2020). Samples were collected during dormancy and first measured directly with and without bark, and then freeze-dried and ground. Direct measurement led to poor calibration models with no substantial improvement after bark removal (Jones et al., 2020). The calibrations using the ground samples yielded accurate results in agreement with previous studies (Schmidtke et al., 2012; De Bei et al., 2017; Jones et al., 2020).

As discussed, various authors have investigated infrared spectroscopy for the quantification of carbohydrates and nitrogen in grapevine plant material (Schmidtke et al., 2012; De Bei et al., 2017; Jones et al., 2020). However, extensive sample preparation methods were still employed. Most sample preparations change the anatomical and physical properties of the sample. Even fixing the samples in resin could lead to structural and chemical changes that could alter the infrared spectra. Thus, it is suggested that using fresh samples could be less invasive, preserving the sample integrity and structure (Türker-Kaya and Huck, 2017). Current research suggests that methods consisting of direct measurement of intact, fresh samples using infrared spectroscopy with no sample preparation could lead to accurate calibrations and will be investigated in the next section (Lopo et al., 2015; Musingarabwi et al., 2016; Diago et al., 2018; Fernández-Novales et al., 2019; Cuq et al., 2020; Jones et al., 2020).





**FIGURE 2 |** Regression workflow for model development using infrared spectra and grapevine reference data. NIR, near infrared; MIR, mid infrared; PLS, partial least squares; ANN, artificial neural networks; MSC, multiplicative scatter correction; LASSO, least absolute shrinkage and selection; SVM, support vector machines; LOCAL, locally weighted; SOM, Self-organizing maps; RMSE, root mean square error; RPD, residual predictive deviation; SI, slope and intercept; ICC, inter-class correlation coefficient; SEM, standard error of measurement; LOD, limit of detection; LOQ, limit of quantification.

## Direct Quantification Techniques Using Infrared Spectroscopy

The investigation of using infrared spectroscopy for the non-invasive measurement of fresh grapevine samples could lead to direct, in-field applications allowing for the implementation

of precision viticulture. Although limited studies have been reported on fresh samples, the feasibility of using infrared spectroscopy as a direct, non-invasive quantification method for nitrogen, amino acids, and other metabolites, are explored in this section (Fernández-Navales et al., 2019; Cuq et al., 2020).

A method using on-the-go, contactless NIR explored the possibility of monitoring grapevine water status (Diago et al., 2018). Predictive models were compiled with  $R^2$  values ranging from 0.68 to 0.85. The models were used to spatially map the water status variability of a vineyard on different dates, providing useful information and facilitating decisions regarding irrigation schedules (Diago et al., 2018).

Vis-NIR reflectance spectroscopy was also used to explore quality attributes in intact grape berries (Dos Santos Costa et al., 2019; Fernández-Novales et al., 2019). Fernández-Novales et al. (2019) investigated amino acid and total soluble solid content using PLS with fivefold cross validation. Samples were collected from one cultivar and one vineyard with 128 grape clusters sampled at five phenological stages. Accurate models were developed for total soluble solids, but the amino acid calibration models were only sufficient for screening purposes (Fernández-Novales et al., 2019). The quantification of amino acid content of berries, even just for screening purposes, could greatly aid harvesting and oenological decisions. Dos Santos Costa et al. (2019) achieved robust prediction models for total soluble solids and total anthocyanins using 432 Shiraz and 576 Cabernet Sauvignon berries. Separate models for Shiraz and Cabernet Sauvignon berries were developed, as well as a combined model. Similar results were reported for the separate and combined models (Dos Santos Costa et al., 2019). These studies were conducted on limited cultivars and vineyards, and although a large number of samples were included the representativeness of those samples were restricted. Future research should include various cultivars, regions, grapevine organs, and phenological stages to include sufficient variability and representativeness in the dataset. The similar results found for separate models per cultivar and for combined models suggest that individualized models should be explored and could lead to more accurate prediction calibrations.

Lastly, a study conducted in southern France used NIR spectroscopy to investigate and assess macro-elements in fresh grapevine leaves and berries (Cuq et al., 2020). The macro elements that were considered in this study were carbon (C), hydrogen (H), nitrogen (N), and sulfur (S). Four cultivars were sampled across 63 plots, and leaf petioles, leaf blades, pea-size berries and véraison berries were collected for each plot leading to 252 total samples. Each sample was made up of a representative number of leaves (50) or berries (200). The blades and petioles of the same leaves, and berries at two phenological stages (pea-size and véraison) were measured as fresh and dried (homogenized) samples. PLS regression using test set validation was employed with a 75:25 split. During model development all data for leaves and berries were combined, but separate models for fresh and dried samples were compiled. Model performance showed only the models for nitrogen and carbon:nitrogen ratio was usable according to the residual predictive deviation (RPD) for fresh and dried samples. The models for the dried samples still performed somewhat better than the fresh samples (Cuq et al., 2020).

However, the predictive difference observed for fresh and dried samples could be offset by the time needed for sample preparation. Using fresh samples could also lead to direct, in-field applications. As previously discussed, this study found separation

based on grapevine organs and phenological stage (Cuq et al., 2020). However, during model development, all the data was combined into a single database. Models for each organ were considered but the data was not shown, and the models were said to be inconclusive and less accurate than the combined data. The inconclusive results for the individual models could be due to limited samples, organs, or phenological stages in this study leading to poor representativeness in the dataset.

The studies discussed in this section investigated infrared spectroscopy for quantification of amino acids, quality parameters, and macro elements in fresh grapevine organs (Dos Santos Costa et al., 2019; Fernández-Novales et al., 2019; Cuq et al., 2020). Although the quantification of some key metabolites was reported in fresh grapevine organs, carbohydrate determination has not yet been attempted. The studies mostly focused on berries and although some reported sampling throughout the growing season, the data from all phenological stages were combined for model development. Similarly, when more than one grapevine organ was sampled, the data for all organs were combined into one dataset. The fact that grapevine organs are extremely heterogeneous, based on their morphological, anatomical, and chemical structure could lead to large variations between grapevine organs, and even within one organ at different phenological stages. This variation could explain why models that combine all the data from different organs, and phenological stages, into a single database for model development are not leading to very accurate prediction models.

By monitoring fresh grapevine organs at several phenological stages across the growing season, specific models for each organ, and possibly phenological stage, could be investigated. These individualized models could lead to more accurate predictions of complex chemical compounds such as carbohydrates, nitrogen, and amino acids. The successful monitoring of these key metabolites during the growing season could greatly aid the implementation of precision viticulture.

## CHEMOMETRIC APPLICATIONS

The investigation and interpretation of infrared spectra is a complex process. The MIR and NIR spectra show peaks and vibrations for all the major molecular bonds in the measured sample. Additionally, NIR also includes the combinations and overtones present (Massart, 1973; Massart et al., 1988; Varmuza and Filzmoser, 2016). Infrared spectroscopy generates spectral data that contains immense amounts of information. Therefore, as previously stated, infrared spectroscopy is seen as a fingerprinting technique because it provides information about several properties of a sample simultaneously (Cozzolino, 2014; Dos Santos et al., 2017).

Chemometrics is needed to decipher this information and is defined as the procedure of extracting relevant information from chemical data using mathematical and statistical tools (Massart, 1973; Massart et al., 1988; Varmuza and Filzmoser, 2016). Chemometrics, and more specifically multivariate data analysis (MDVA), has been extensively used for qualitative and quantitative applications in the agricultural industry

**TABLE 1** | Chemometric techniques and calibration parameters reported in discussed literature.

References	Chemometric/MDVS technique used	Calibration parameters reported
De Bei et al., 2017	PLS	$R^2_{cal}$ , $R^2_{val}$ , SEC <sub>V</sub> , SEP, Slope, Bias, Rank <sup>a</sup> , RPD
Schmidtke et al., 2012	PLS and SVM	$R^2_{val}$ , RMSEP, Bias, Rank
Jones et al., 2020	PLS	$R^2_{cal}$ , $R^2_{val}$ , RMSEC, RMSEV, RPD, Bias
Fernández-Navales et al., 2019	PLS	SEC, $R^2_{cal}$ , SEC <sub>V</sub> , $R^2_{cv}$ , RPD, SEP, $R^2_{pred}$ , Bias, Slope, Rank
Cuq et al., 2020	PLS, PCA	$R^2_{cal}$ , RMSEC, RPD <sub>cal</sub> , $R^2_{val}$ , RMSEV, RPD <sub>val</sub> , $R^2_{pred}$ , RMSEP, RPD <sub>pred</sub> , percentage (reported for PCA)
Diago et al., 2018	PLS	$R^2_{cal}$ , RMSEC, $R^2_{cv}$ , RMSECV, $R^2_{pred}$ , RMSEP
Dos Santos Costa et al., 2019	PCR, PLS, MLS for calibration models. PCA-LDA, PCA-QDA, PCA-LDA Mahalanobis, PLS-DA <sup>b</sup> for maturation classification.	$R^2_{cal}$ , RMSEC, SEC, $R^2_{cv}$ , RMSECV, SEC <sub>V</sub> , $R^2_{pred}$ , RMSEP, SEP, Bias, Rank Percentage (reported for classification), confusion matrix
Petrovic et al., 2020	PLS	$R^2_{cal}$ , RMSEC, RPD <sub>cal</sub> , $R^2_{val}$ , RMSEV, RPD <sub>val</sub> , Slope, Bias, Rank
Aleixandre-Tudo et al., 2018a	PLS	$R^2_{cal}$ , RMSECV, RPD <sub>cal</sub> , $R^2_{val}$ , RMSEV, RPD <sub>val</sub> , Rank, ICC, SEM
Aleixandre-Tudo et al., 2018b	PLS	$R^2_{cal}$ , RMSECV, $R^2_{val}$ , RMSEV, RPD, Bias, Rank, Slope, ICC, SEM, LOD, LOQ
Wang et al., 2017	PCR, PLS, SMLR, BPNN <sup>c</sup>	$R^2_{cal}$ , $R^2_{val}$ , RMSEC, RMSEV
Quentin et al., 2017	PLS	$R^2_{cal}$ , RMSEC, SEC, RPD <sub>cal</sub> , $R^2_{pred}$ , RMSEP, RPD <sub>pred</sub> , Slope, Bias
Ramirez et al., 2015	PLS	$R^2_{cal}$ , $R^2_{cv}$ , $R^2_{val}$ , RMSEC, RMSECV, RMSEP, RPD <sub>cv</sub> , RPD <sub>val</sub> , Rank

<sup>a</sup>Rank defined as the number of latent variables or principal components/partial least square terms in calibration.

<sup>b</sup>PCA using linear discriminant analysis (PCA-LDA), quadratic discriminant analysis (PCA-QDA), linear discriminant analysis using Mahalanobis distance (PCA-LDA Mahalanobis) and PLS-DA.

<sup>c</sup>SMLR, stepwise multiple linear regression; BPNN, back propagation neural network.

(Damberg et al., 2015; Varmuza and Filzmoser, 2016; Williams, 2019). Qualitative applications will be mentioned in this section, while quantitative chemometric applications will be discussed in more detail.

The chemometric techniques most often used in the literature discussed in this review for infrared spectroscopy data are PCA and PLS regression, as shown in **Table 1**. PCA is mostly utilized as a qualitative method for screening, grouping, extraction, and compression of multivariate data. Using mathematical procedures, the correlated response variables in spectral data are transformed into non-correlated variables known as principal components (PC's) (Cozzolino et al., 2011; Dos Santos Costa et al., 2019; Cuq et al., 2020).

Although most literature in the agricultural and viticultural field report the use of PLS, it is worth mentioning that other regression techniques have recently been proposed that could outperform PLS. Ridge regression and lasso (least absolute shrinkage and selection operator) regression are variable selection methods that can be applied to spectroscopy data to reduce variables used during model development leading to more interpretable models (Piaskowski et al., 2016; Frizzarin et al., 2021; Gao et al., 2021). Although limited agricultural applications for ridge and lasso regression have been reported, it has been successfully employed to predict quality traits in cow's milk (Frizzarin et al., 2021) as well as lignin content in poplar trees (Gao et al., 2021). However, in a study conducted on NIR data seven different regression methods were investigated including PLS, ridge, and lasso regression, to assess carbohydrates in wheat samples with PLS yielding the most accurate results (Piaskowski et al., 2016).

Regression methods, such as PLS regression, are employed for classification purposes (PLS-DA) as well as quantification

using spectral and reference data to compile prediction models. Chemometrics uses statistical techniques and reference data to extract and correlate the spectral data with a property of interest. This chemometric approach uses model development, including model calibration and validation, for the prediction and quantification of chemical and physical properties (Cozzolino et al., 2011; Schmidtke et al., 2012; De Bei et al., 2017; Bureau et al., 2019). **Figure 2** shows the main steps during regression model development implemented on infrared spectra and grapevine reference data. Regression techniques employed for quantification purposes will be discussed extensively. Preprocessing techniques, model development steps, and performance evaluation indices and other methods used to compare model performance will be evaluated.

## Preprocessing Techniques

Before multivariate regression techniques are applied to spectral data, preprocessing is often used to enhance spectral features and remove undesirable sources of variation or irrelevant information (Dos Santos et al., 2017; Bureau et al., 2019). Various spectral preprocessing techniques exist, such as scattering or baseline correction, smoothing or transforming to first or second derivatives, standardization, and normalization. During model development some of the preprocessing methods are often investigated and the optimal technique or combination of techniques are chosen based on model performance (Nicolai et al., 2007; Cozzolino et al., 2011; Varmuza and Filzmoser, 2016). As discussed in previous sections the study done by Cuq et al. (2020) was one of few that used fresh grapevine material. They employed preprocessing to eliminate light scattering effects and compensate for baseline bias and offset. They used

multiple methods including the Savitzky-Golay algorithm and multiplicative scatter correction (MSC).

The Savitzky-Golay algorithm uses smoothing and transformation to remove random noise, baseline shifts and superimposed peaks to the first or second derivative. These derivatives help to highlight small peaks and bands and to elucidate overlapping peaks. The application of the MSC method to spectral data reduces the effect of light scattering and linearizes the spectra. MSC compensates for the non-uniform light scattering throughout the sample caused by particle size, refractive index, and radiation wavelengths (Nicolai et al., 2007; Cozzolino et al., 2011; Varmuza and Filzmoser, 2016). Both these methods were proven to be useful when dealing with whole, fresh grapevine organs (Cuq et al., 2020).

Various other studies also used the Savitzky-Golay algorithm and MSC for preprocessing and these methods were often used in conjunction (Schmidtke et al., 2012; Dos Santos Costa et al., 2019; Fernández-Navales et al., 2019). Schmidtke et al. (2012) applied these algorithms to spectra from grapevine samples that were dried and ground to a powder, while Dos Santos Costa et al. (2019) and Fernández-Navales et al. (2019) used fresh grape berries. The wide application of these algorithms to dried and fresh samples show their applicability in the field of viticulture (Schmidtke et al., 2012; Dos Santos Costa et al., 2019; Fernández-Navales et al., 2019; Cuq et al., 2020).

The sample type, instrument, and purpose of the analysis all contribute to choosing an appropriate preprocessing technique (Stuart, 2004; Varmuza and Filzmoser, 2016; Türker-Kaya and Huck, 2017; Aleixandre-Tudo et al., 2019; Bureau et al., 2019). Often numerous preprocessing techniques and combinations need to be investigated to find the correct approach. The correct choice could also be not to implement any preprocessing techniques. The scattering components in some NIR spectra could include physical information of the sample such as density and removing scattering effects could reduce a sample's physical information. Using smoothing techniques often only marginally improves the calibration model or even causes unfavorable effects. The use of smoothing could lead to decreased prediction accuracy because of correlations introduced in the noise structure (Olivieri, 2015). The option of using preprocessing or the raw spectra should be carefully considered before model development is initiated.

## Model Development

During linear regression model development, the dataset is often subdivided into a calibration set that is used to construct the model and a validation set used to test the model. This is done using various methods including test set validation, cross validation with leave-one-out or k-fold cross validation. During test set validation the dataset is split using algorithms, such as Kennard-Stone, based on a percentage (Porep et al., 2015; Petrovic et al., 2020). The split, between calibration and validation datasets, generally applied is 70:30 or 50:50 for PLS calibrations or even 80:20 for machine learning models. These dataset splits differ based on sample type and application (Fuentes et al., 2018; Bureau et al., 2019; Murru et al., 2019; Petrovic et al., 2020).

Alternatively, cross validation can be used. This involves leaving out one sample at a time for the construction of the calibration model, and then using the left-out sample for validation. However, with larger datasets it is often difficult to leave out one sample at a time because of the numerous calibrations generated when using this method and the minimal effect that leaving one sample out has on model performance. In the instance of large datasets, cross validation uses sample sets of up to ten or twenty samples that are left out at a time and used for validation (Cozzolino et al., 2011; Varmuza and Filzmoser, 2016). Similarly, when using k-fold cross validation the dataset is split into K subsets. One subset is applied for validation while the other K-1 subsets are used for calibration. This is repeated K times so that after K iterations all data is used for calibration as well as validation. K values between five and ten are commonly used (Paiva et al., 2021; Santos et al., 2021).

Although test set and cross validation datasets are considered as independent from a statistical viewpoint, model accuracy and robustness should preferably be tested with an independent, completely external sample test set. The samples for the external test set should be collected under the same conditions as the calibration set. Samples collected from different vintages, cultivars, and regions should be included in the external test set to assess the model's robustness and ability to handle spectral variability in samples beyond those used in the calibration set (Cozzolino et al., 2011; Dos Santos et al., 2017; Bureau et al., 2019). Including sufficient variability could be difficult in agriculture and viticulture where ample variations exist between years because of vintage and climatic effects. However, if sufficient sample variability was included in previous vintages the model should have a better ability to predict future samples.

## Performance Evaluation Indices

The application of various preprocessing strategies, multivariate regression techniques and model optimizations lead to multiple calibration and validation models that need to be assessed based on their performance. Multiple performance evaluation indices or calibration parameters are used to report the results of calibration and validation models. The indices test the model's accuracy and reliability (Aleixandre-Tudo et al., 2018b; Bureau et al., 2019; Dos Santos Costa et al., 2019; Fernández-Navales et al., 2019; Williams, 2019). The most commonly used parameters in the discussed literature are summarized in **Table 1** and the values associated with them in **Tables 2, 3**.

One of the most widely used parameters is the root mean square error (RMSE) that is used for calibration (RMSEC), cross validation (RMSECV) and prediction (RMSEP). RMSECV indicates the possible error for future predictions and RMSEP estimates the model's ability to accurately predict new samples. Alternatively, the standard error of calibration (SEC), cross-validation (SECV) and standard error of prediction (SEP) could be used. RMSEP and SEP values are related but unlike RMSEP, SEP is independent of bias, while RMSEP include bias error. Some authors prefer using SEP together with bias while others favor RMSEP and some authors report all of the above as shown in **Tables 1–3** (Aleixandre-Tudo et al., 2018b; Bureau et al., 2019;



**TABLE 2** | Direct and non-direct applications of infrared spectroscopy for quantification tasks in grapevines.

References	Application	Sample type	Sample collection	IR technique	Chemometric/MDVS technique used	Calibration accuracy
De Bei et al., 2017	Total non-structural carbohydrates (TNC), starch and sugar	Non-direct Trunk and leaf samples—Freeze dried, and ground to powder	One cultivar, two vintages, four phenological stages	NIR	PLS	<p><b>Trunk models:</b>  TNC: <math>R^2_{cal} = 0.84</math>, <math>R^2_{val} = 0.82</math>, SECV = 10.92 mg/g, SEP = 12.65 mg/g, PC's = 11, RPD = 2.17  Starch: <math>R^2_{cal} = 0.84</math>, <math>R^2_{val} = 0.80</math>, SECV = 10.60 mg/g, SEP = 12.51 mg/g, PC's = 11, RPD = 2.51</p> <p><b>Leaf models:</b>  TNC: <math>R^2_{cal} = 0.92</math>, <math>R^2_{val} = 0.86</math>, SECV = 17.83 mg/g, SEP = 22.86 mg/g, PC's = 5, RPD = 3.48  Starch: <math>R^2_{cal} = 0.93</math>, <math>R^2_{val} = 0.88</math>, SECV = 16.18 mg/g, SEP = 21.13 mg/g, PC's = 7, RPD = 3.73</p>
Schmidtke et al., 2012	Nitrogen and starch reserves	Non-direct Root and wood samples—Freeze dried, and ground to powder	Three cultivars, four vintages, five regions (across two countries), 35 vineyards	ATR-FT-IR	PLS and SVM	<p><b>Nitrogen</b>  SVM: <math>R^2 = 0.98</math>, RMSEP = 0,07% DW  PLS: <math>R^2 = 0.97</math>, RMSEP = 0,08% DW</p> <p><b>Starch</b>  SVM: <math>R^2 = 0.95</math>, RMSEP = 1,56% DW  PLS: <math>R^2 = 0.95</math>, RMSEP = 1,43% DW</p>
Jones et al., 2020	Starch	Direct and non-direct Cane wood	Two cultivars, five vineyards, three vintages	NIR	PLS	<p><b>Starch -ground samples</b>  <math>R^2 = 0.88</math>, RMSEV = 0.73 mg/g, RPD = 2.85</p> <p><b>Starch—intact samples with bark</b>  <math>R^2 = 0.20</math>, RMSEV = 0.86 mg/g, RPD = 1.08</p> <p><b>Starch—intact samples without bark</b>  <math>R^2 = 0.36</math>, RMSEV = 0.75 mg/g, RPD = 1.24</p>
Fernández-Novales et al., 2019	Amino acids	Direct Berries—fresh	One cultivar, one vineyard, five sampling dates	Vis + SW-NIR (Visible + Short wave NIR) and NIR	PLS	<p><b>Asparagine</b>  SEC = 0.38 mg N/l, <math>R^2_{cal} = 0.71</math>, SECV = 0.44 mg N/l, <math>R^2_{cv} = 0.64</math>, RPD = 1.63, SEP = 0.45 mg N/l, <math>R^2_{pred} = 0.66</math></p>
Cuq et al., 2020	C, H, N, S	Direct Leaves and berries—fresh	Four cultivars, 63 vineyards, two phenological stages	NIR	PLS, PCA	<p><b>Nitrogen</b>  <math>R^2_{cal} = 0.90</math>, RMSEC = 0.181% DW, <math>RPD_{cal} = 3.14</math>, <math>R^2_{val} = 0.84</math>, RMSECV = 0.237% DW, <math>RPD_{val} = 2.41</math>, <math>R^2_{pred} = 0.91</math>, RMSEP = 0.172% DW, <math>RPD_{pred} = 3.32</math></p>
Dos Santos Costa et al., 2019	Quality (total soluble solids, total anthocyanins, yellow flavonoids) and maturation	Direct Berries—fresh	Two cultivars, one region, one vintage	Vis/NIR	PCR, PLS, MLS for calibration models.  PCA-LDA, PCA-QDA, PLS-DA for maturation classification.	<p><b>Shiraz model</b>  TSS: <math>R^2_{cal} = 0.97</math>, RMSEC = 0.89%, SEC = 0.89%, <math>R^2_{cv} = 0.94</math>, RMSECV = 1.33%, SECV = 1.33%, <math>R^2_{pred} = 0.95</math>, RMSEP = 1.15%, SEP = 1.13%</p> <p><b>Cabernet Sauvignon model</b>  TSS: <math>R^2_{cal} = 0.97</math>, RMSEC = 1.08%, SEC = 1.09%, <math>R^2_{cv} = 0.97</math>, RMSECV = 1.13%, SECV = 1.13%, <math>R^2_{pred} = 0.96</math>, RMSEP = 1.39%, SEP = 1.40%</p> <p><b>Combined model</b>  TSS: <math>R^2_{cal} = 0.96</math>, RMSEC = 1.21%, SEC = 1.21%, <math>R^2_{cv} = 0.95</math>, RMSECV = 1.37%, SECV = 1.37%, <math>R^2_{pred} = 0.95</math>, RMSEP = 1.38%, SEP = 1.39%</p>

**TABLE 3 |** Direct and non-direct application of infrared spectroscopy for quantification tasks in other plants.

References	Application	Sample type	Sample collection	IR technique	Model accuracy
Veselá et al., 2007	Fat, nitrogen, and moisture	Cocoa powder	100 different brands	NIR and FTIR	<b>Nitrogen</b> NIR: $R^2 = 0.98$ , RMSECV = 1.7%, RMSEP = 0.09% FTIR: $R^2 = 0.95$ , RMSECV = 3.9%, RMSEP = 0.14% OPA with NIR and MIR: $R^2 = 0.96$ , RMSECV = 2.2%, RMSEP = 0.10%
Ramirez et al., 2015	Non-structural carbohydrates, sugar, and starch	Leaves and wood samples (including branch, stem, and root) – Dried and ground to powder	82 native tree species, four sites, two countries	NIR	<b>Total NSC (all tissues)</b> $R^2_{cal} = 0.93$ , $R^2_{cv} = 0.88$ , $R^2_{val} = 0.91$ , RMSEC = 1.12%, RMSECV = 1.43%, RMSEP = 1.34%, RPD <sub>cv</sub> = 2.88, RPD <sub>val</sub> = 3.26 <b>Total NSC (leaves)</b> $R^2_{cal} = 0.98$ , $R^2_{cv} = 0.94$ , $R^2_{val} = 0.68$ , RMSEC = 0.71%, RMSECV = 1.12%, RMSEP = 2.63%, RPD <sub>cv</sub> = 4.22, RPD <sub>val</sub> = 1.45 <b>Total NSC (stem and branches)</b> $R^2_{cal} = 0.97$ , $R^2_{cv} = 0.93$ , $R^2_{val} = 0.87$ , RMSEC = 0.67%, RMSECV = 1.00%, RMSEP = 1.22%, RPD <sub>cv</sub> = 3.86, RPD <sub>val</sub> = 2.58 <b>Total NSC (roots)</b> $R^2_{cal} = 0.99$ , $R^2_{cv} = 0.94$ , $R^2_{val} = 0.91$ , RMSEC = 0.55%, RMSECV = 1.11%, RMSEP = 1.18%, RPD <sub>cv</sub> = 4.22, RPD <sub>val</sub> = 3.41
Wang et al., 2017	Nitrogen	Leaves - fresh	One pear orchard, two vintages	Portable Vis/NIR	<b>Nitrogen</b> $R^2_{cal} = 0.86$ , $R^2_{val} = 0.81$ , RMSEC = 0.12, RMSEV = 0.13
Quentin et al., 2017	Non-structural carbohydrate, total soluble sugar, and starch	Leaves - Dried and ground to powder	24 Eucalyptus trees, over 8 months	NIR	<b>Total NSC</b> $R^2_{cal} = 0.88$ , RMSEC = 2.90%, SEC = 2.91%, RPD <sub>cal</sub> = 2.79, $R^2_{pred} = 0.91$ , RMSEP = 2.35%, RPD <sub>pred</sub> = 2.73

Dos Santos Costa et al., 2019; Fernández-Navales et al., 2019; Williams, 2019).

These parameters' values are given in the same units as the measured compounds and should be as small as possible (Tables 2, 3). The threshold or accuracy will depend on the unit of measure for the compounds and the sensitivity of the existing analysis. The parameters can also be expressed as a percentage calculated using the population mean of the calibration or validation set used and a percentage below 20% is regarded as acceptable for analytical methods (Cozzolino et al., 2008; Torchio et al., 2013; Alexandre-Tudo et al., 2019). These calculations do not include possible errors associated with the reference methods, but despite these limitations, they are still the most commonly used (Dos Santos et al., 2017).

Another performance index that is often used is the coefficient of determination or R squared ( $R^2$ ).  $R^2$  is used to explain the variance of the response variable in the calibration ( $R^2_{CAL}$ ) and validation ( $R^2_{VAL}$ ) sets. The value should be close to 1 so that as much variance as possible is explained for the response variable in the dataset (Cozzolino et al., 2011; Bureau et al., 2019).

The ratio of prediction to deviation or residual predictive deviation (RPD) is also often used to evaluate the predictive ability of a model (Ramirez et al., 2015; De Bei et al., 2017; Quentin et al., 2017; Cuq et al., 2020; Jones et al., 2020). RPD is calculated as the ratio of the standard deviation of the response variable to the RMSEP or RMSECV (Alexandre-Tudo et al., 2019; Bureau et al., 2019). Values between 2 and 3 have been reported as acceptable for wine and grape applications (Alexandre-Tudo et al., 2019; Cuq et al., 2020). In addition, other authors have interpreted values below 3 as adequate for screening and values above 5 capable of classification, quality, and process control tasks (Tables 2, 3) (Ramirez et al., 2015; De Bei et al., 2017; Quentin et al., 2017; Cuq et al., 2020; Jones et al., 2020). However, the RPD values' interpretation is somewhat controversial since these thresholds were not determined using statistical basis (Cozzolino et al., 2011). Furthermore, other authors have argued that reporting both  $R^2$  and RPD is redundant since the calculation of RPD is inversely related to R squared (Minasny and McBratney, 2013).

Other statistics such as bias, rank and fit for purpose criterion are also used to assess model performance. Bias is measured as the difference between expected values (predicted) and true values (reference data) of a distribution. The selection of an optimal number of latent variable or principal components (often called rank) when using multivariate regression techniques such as PLS is also extremely important. If a model's rank is too high there might be over-fitting and too low rank could lead to a model incapable of capturing all the variability present in the dataset (Cozzolino et al., 2011; Alexandre-Tudo et al., 2019).

The fit for purpose criterion also needs to be considered after model development and judges the applicability of the model for routine use. The models need to be interpreted based on application as well as statistics (Cozzolino et al., 2011; Bureau et al., 2019). The successful implementation of a method also depends on model robustness. A calibration model is robust when the prediction accuracy is independent of external

factors. Including a large representative sample set consisting of various cultivars, regions, vintages, and climatic conditions in the calibration model will help to ensure robustness (Cozzolino et al., 2011; Dos Santos et al., 2017; Petrovic et al., 2020).

In agricultural applications even less accurate prediction models can supply the ability to screen samples for low and high values. Rapid screening can be invaluable compared to the existing time-consuming and destructive methods. Although each of the performance evaluation indices do not supply all the answers, by investigating, and evaluating, all of them a clear picture of the regression techniques' performance could emerge.

**Tables 2, 3** list the studies using direct and indirect measurement of grapevines and other plants, respectively. The sample types, analytical application, and sampling procedures are reported together with the infrared and chemometric techniques used. Most of the studies reported numerous calibrations. However, only the most accurate calibration results were included in the tables. Most of the information shown in these tables were discussed in previous sections.

## Statistical Tests for Model Comparison

During model development with agricultural samples, a large number of models are often created to investigate the dataset. The values of the performance evaluation indices such as RMSE are often directly compared to evaluate the models' predictive ability. Direct comparison of these absolute values can show differences between the models but not if the differences are significant in terms of model performance. Comparing the values directly might not be the best strategy to evaluate prediction performance. Statistical tests to compare the difference in observed and predicted values could be applied to evaluate significant differences (Olivieri, 2015; Alexandre-Tudo et al., 2018a, 2019; Petrovic et al., 2020).

A randomization test has been investigated for the comparison of RMSE values. Hypothesis testing is employed to determine if calibrations differ significantly with the null hypothesis stating that the two compared RMSE values are equal ( $RMSE_1 = RMSE_2$ ) and the alternative hypothesis proposing that one value is larger or different than the other ( $RMSE_1 > RMSE_2$ ) (Van der Voet, 1994; Olivieri, 2015; Alexandre-Tudo et al., 2019; Petrovic et al., 2020). Randomization testing could determine if significant differences between sample types, infrared methods or multivariate statistical techniques are present (Petrovic et al., 2020). Some literature has also suggested randomization testing for classification problems. Several classification methods could be compared, testing the hypothesis that two classification methods have similar classification ability (Van der Voet, 1994).

Other methods for model comparison, such as the slope and intercept test (SI test), have also been proposed in the literature (Linnet, 1993; Alexandre-Tudo et al., 2018a,b). The SI test investigated the systematic error between the predicted values and the reference data and used a combined analysis of the regression line's slope and intercept. SI testing can show if differences observed between predicted and reference values are due to random noise or not. The SI test can be used to evaluate and compare model performance and measurement methods

such as infrared methods compared to reference methods (Linnet, 1993; Alexandre-Tudo et al., 2018a,b).

Another proposed approach is the use of the inter-class correlation coefficient (ICC) and the standard "typical" error of measurement (SEM) calculated from the ICC. ICC is sensitive to detecting systematic error and both ICC and SEM are often used in reliability studies (Alexandre-Tudo et al., 2018b). Reliability is defined as the consistency of measurements and ICC can be used to test the reliability of an instrument, person, or prediction value. The ICC values are unitless and are reported as a relative measure of reliability with values between 0 and 1. Values closer to 0 show no reliability and values closer to 1 indicate higher reliability. The magnitude of the ICC values depends on both the between sample variability as well as variability within the dataset (Yen and Lo, 2002; Weir, 2005; Alexandre-Tudo et al., 2018a,b).

Therefore, the heterogeneity of the sample set should be considered when using ICC. Large ICC values could mask systematic errors when between-sample variability is high and low ICC values could still be found with a low systematic error and little between-sample variability. These values could lead to the conclusion that if samples are homogenous, it could be difficult to differentiate between them although the systematic error is small (Yen and Lo, 2002; Weir, 2005; Alexandre-Tudo et al., 2018a,b).

The misinterpretation of the ICC values can be avoided by investigating the SEM in conjunction with ICC values. SEM values have the same unit as the measurement of interest and provide an absolute measure of reliability while ICC is a relative measure. SEM quantifies the precision of separate measurements and shows measurement error. SEM could also be used to construct confidence intervals for separate measurements and to determine the minimum difference needed to show true variance between separate measurements (Yen and Lo, 2002; Weir, 2005; Alexandre-Tudo et al., 2018a,b).

Lastly, PLS calibrations could be assessed using the limit of detection (LOD) and limit of quantification (LOQ) adapted for multivariate purposes. LOD and LOQ values could still indicate the minimum concentration accurately predicted by the PLS calibration. Although, some studies have suggested that these values should be calculated for each sample. The multivariate LOD and LOQ will then be represented by a concentration range rather than a single value. LOD is seen as a good indicator of the quality of a calibration model because it incorporates the sensitivity and precision of analytical measurements (Allegrini and Olivieri, 2014; Olivieri, 2015; Alexandre-Tudo et al., 2018b). Calibrations and methods could be evaluated and assessed using the statistical model comparison methods discussed. During model development, calibrations for each grapevine organ or developmental stage could be compiled and compared using the comparison methods.

## Non-linear Regression Techniques

The multivariate regression techniques discussed up to this point have all been linear regression methods, but sometimes agricultural models behave in a non-linear way. Linearity, or the lack thereof, can be assessed using statistical tests. Only when a statistically significant improvement of the prediction model

is found, non-linear methods can be explored (Olivieri, 2015). Recently, non-linear regression techniques such as artificial neural networks (ANN) and Kernel-based techniques such as least squares support vector machines (LS-SVM) have been used more frequently (Fuentes et al., 2018; Murru et al., 2019).

Although ANN may perform better than linear techniques in some cases, the results are often difficult to understand, visualize, and interpret. Kernel-based techniques are often more favorable because they allow interpretation of the calibration model. LS-SVM has the benefit of including an added regularization parameter. The regularization parameter penalizes the use of large regression coefficient values leading to improved robustness of calibration models. Kernel versions of PCR and PLS as a logical extension of ordinary PCR and PLS have also been described. These statistical methods could be easier to work with for someone already familiar with PCR and PLS (Nicolai et al., 2007; Cozzolino et al., 2011).

In recent viticultural research, non-linear methods have been employed for various classification purposes (Fuentes et al., 2018; Murru et al., 2019). Machine learning techniques such as ANN have been implemented together with NIR to develop models capable of cultivar classification based on leaf samples (Fuentes et al., 2018). FTIR and ANN have been utilized for the classification of grape samples according to cultivar and ripeness levels (Murru et al., 2019). LOCAL (locally weighted) regression has also been applied for the prediction of red grape quality parameters, and was found to perform better than PLS (Damberg et al., 2006).

Other non-linear regression techniques, not widely used in the field of viticulture, should also be considered. Self-organizing maps (SOM) is a type of neural network specifically suited to large and multi-dimensional datasets making it ideal for spectral data (Tan et al., 2013; Wehrens and Kruisselbrink, 2018; Xu et al., 2019). SOM is an unsupervised clustering algorithm unique in that it transforms complex data into visually interpretable clusters and still preserve the topological properties of the input space (Milovanovic et al., 2019; Xu et al., 2019). This is achieved using a neighborhood function to plot objects into a two-dimensional space with similar objects close together and dissimilar objects further apart. SOMs can be implemented to assess clustering in data as well as investigate the structure within the clusters (Wehrens and Kruisselbrink, 2018; Milovanovic et al., 2019).

In a recent study PCA and SOM were used to investigate the classification of wine based on origin. Both methods yielded acceptable clustering results, although it was stated that SOM treatments provided better resolution. Although this study was not conducted on spectral data, the feasibility of using SOM as an alternative to PCA was shown. The added advantage of the unsupervised nature of SOMs is that no assumptions are made during clustering, visualization, and construction of the data (Milovanovic et al., 2019). The use of SOM in conjunction with PLS has also been proposed to improve spectral prediction models. SOM is initially used to cluster the data based on spectral variables or samples, after which PLS is performed on the clusters, leading to a consensus model. These models fully incorporate the sample information while highlighting the role of applicable

variables and samples, and reducing the effect of useless variables or samples (Tan et al., 2013; Xu et al., 2019). Often a combination of linear and non-linear methods needs to be considered together with the technique's purpose (prediction or classification) to achieve the desired outcome.

## Final Remarks

The main advantage of chemometric methods is that they look beyond the one-dimensional approach and investigate the sample in its entirety, making it well suited for use with spectroscopy. Both multivariate and infrared spectroscopy techniques do not assess just a single component but the interactions, interferences, and combined effects of the whole sample matrix. Many modern applications of spectroscopy techniques in the food industry are based on indirectly measuring chemical and physical properties. Instrumental techniques established using infrared spectroscopy are often correlated methods meaning that the measurement variable does not directly correlate with the compound of interest or the concentration of the compound (Gishen et al., 2005). The spectral regions identified in the prediction model might not correlate directly to the compound of interest. The complexity of spectra makes it difficult to correlate prediction models with the specific or relevant functional groups of a reference compound (Cozzolino, 2014). The correlation of the prediction models with other spectral regions can be explained by the fact that infrared spectroscopy measures all components in the sample, as well as the interaction between compounds, and the interference or combined effects they have on each other. Therefore, these interactions or interferences could contribute to the data used in the prediction model. Multivariate data analysis techniques can incorporate and investigate all the aspects of spectroscopy data making it the favored approach (Gishen et al., 2005).

Furthermore, new chemometric techniques and new ways of reporting results are continuously being developed. Recent studies have proposed using statistical tests such as the randomized test, SI test, ICC, SEM, LOD, and LOQ values to compare model performance (Allegrini and Olivieri, 2014; Olivieri, 2015; Alexandre-Tudo et al., 2018a,b). Most studies still use direct comparison of the absolute values of the model performance indicators for the comparison of calibrations. Direct comparison is not the most suitable approach since it does not indicate if these values differ significantly, and if there truly are significant differences between the calibrations. The comparison methods could also be used for various purposes including testing the reliability of measurements, comparing instrumentation, detecting systematic error, and reporting the sensitivity of the calibration (Yen and Lo, 2002; Weir, 2005; Allegrini and Olivieri, 2014; Olivieri, 2015; Alexandre-Tudo et al., 2018a,b). The future application of infrared spectroscopy will be elucidated in the following section.

## DISCUSSION AND FUTURE PROSPECTS

Despite numerous articles on infrared spectroscopy techniques published in scientific literature, the majority described feasibility studies (Dos Santos Costa et al., 2019;



Fernández-Novales et al., 2019; Cuq et al., 2020; Jones et al., 2020). The heterogeneous nature of grapevine samples further complicates the matter, and a more representative sample set might be needed to accurately capture the variability, including various cultivars, regions, vintages, organs, and phenological stages. Thus, two contrasting approaches to optimize prediction models could be to construct a universal calibration or to build individualized calibrations per grapevine organ or phenological stage. Both these approaches will be discussed in this section.

However, the number of samples is not the most important factor, but rather how representative the samples are of future datasets. Studies conducted on just one cultivar or region could still measure thousands of samples without capturing the variability present in a vineyard (Schmidtke et al., 2012; De Bei et al., 2017; Rossouw et al., 2017; Fernández-Novales et al., 2019). During sample selection special care needs to be taken to incorporate representative samples. This could include sampling from various cultivars, regions, and vintages. Calibration development has also been mostly based on experimental data using cross-validation techniques where no independent sample set was used to validate the calibration model (Damberg et al., 2015). Using a dataset of external and new samples to validate the calibration model could lead to more robust models capable of dealing with samples from different vintages, cultivars, or regions.

The frequent use of less representative sample sets of only one cultivar, organ, or phenological stage could be caused by expensive and time-consuming reference methods that are needed during calibration development. The reference methods require intense analytical, and human resources. Often cultivars, regions, vintages, organs, and phenological stages need to be analyzed to accurately capture the variability required for robust prediction models (Schmidtke et al., 2012; De Bei et al., 2017; Rossouw et al., 2017; Fernández-Novales et al., 2019).

In viticultural studies it is often difficult to include cultivars, regions, and vintages in a single study. Still, large variability of the samples is commonly needed to achieve sufficient robustness in the calibration (Schmidtke et al., 2012; De Bei et al., 2017; Dos Santos Costa et al., 2019; Fernández-Novales et al., 2019; Cuq et al., 2020). Similarly, conducting experiments on various climatic conditions, viticultural practices, or multiple growing regions could be challenging. It is not yet feasible to conduct studies including all these factors simultaneously. Therefore, collaborations among universities, research groups, and countries could be considered to develop a universal calibration capable of predicting multiple key metabolites in grapevines across the world. However, this will require global cooperation, immense datasets, and years to achieve.

Hyperspectral imaging (HSI) has also been proposed as a solution for the non-invasive quality assessment of fruits and vegetables. The main advantage of HSI is that it simultaneously provides the spatial and spectral information of the whole sample, whereas NIR provides the spectra of a given spot (Lorente et al., 2012; Chandrasekaran et al., 2019). The use of HSI for various quantification applications in grape berries has been reported, such as pH, anthocyanin, and soluble solid content (Chandrasekaran et al., 2019). A recent study done on grapevine bunches found that HSI was capable of distinguishing

between healthy bunches and bunches infected with powdery mildew (Pérez-Roncal et al., 2020). Hyperspectral imaging has also been successfully applied to classification tasks such as the identification of grape varieties based on leaf spectra (Diago et al., 2013). The monitoring of vegetative indices throughout the growing season using HSI on grapevine leaves was also proposed (Yang et al., 2021). The use of NIR and MIR spectroscopy has been widely applied and researched for various agricultural and viticultural applications, unlike HSI. Although NIR and MIR spectroscopy generate extensive datasets, these datasets are still easier to analyze than the even more extensive datasets generated with HSI. While most research on HSI has been done in the last 2 years, these studies showed the significant potential of using HSI in future for non-invasive applications on grapevines.

Furthermore, studies have investigated portable infrared devices for in-field applications. Portable devices are mostly used for qualitative purposes such as classification (Gutiérrez et al., 2015; Lopo et al., 2015) while the samples for direct quantification are still taken to the laboratory (Fernández-Novales et al., 2019; Cuq et al., 2020). The few studies involving portable devices for viticultural applications have shown very contradicting results. Some recent studies have reported promising results for the prediction of water status in vineyards. The quantitative studies confirmed that NIR spectroscopy was robust and capable of reliably evaluating water status across diverse environmental conditions (Tardaguila et al., 2017; Diago et al., 2018). Although bench-top and portable devices represent different measurement technologies, similar measurement parameters were reported in a comparison study (Rodgers et al., 2017). However, portable devices often do not cover the entire spectral range and use fiber optics which could potentially introduce spectral noise (Reeves, 2010; Pasquini, 2018; Cuq et al., 2020). It has also been suggested that portable devices could lack the same quality components (detectors) found in bench-top instruments (Zumba et al., 2018).

Some studies have also raised the concern that extreme environmental conditions such as high temperatures or relative humidity could affect spectra (Zumba et al., 2018; Baca-Bocanegra et al., 2019). Temperature and humidity could affect samples' moisture content and thus influence spectra (Zumba et al., 2018). A recent study investigated different detectors and sample temperatures under simulated field conditions. It was found that the temperature of the detector could influence spectral reproducibility, although sample temperature did not have the same effects. However, very few samples were included in the study (Phuphaphud et al., 2020).

The monitoring of grapevine organs throughout the growing season using infrared spectroscopy could yield valuable information. Extensive knowledge on the morphological and anatomical changes occurring in grapevine organs during the growing season are available, but the changes influencing spectral properties are not well understood. Changes have been observed in infrared spectra between different grapevine organs or phenological stages (Schmidtke et al., 2012; De Bei et al., 2017; Dos Santos Costa et al., 2019; Fernández-Novales et al., 2019; Cuq et al., 2020), but the reason for the differences have not been examined or explained.

Further investigation is needed to understand the relationship between morphological and anatomical changes, and spectral properties. These investigations could lead to the linking of certain morphological or anatomical changes to specific spectral properties. Spectral regions could be correlated to changes occurring throughout the growing season such as lignification or leaf aging. Once identified, the regions could be used for future monitoring of changes or used for classification purposes. The classifications could in turn lead to more individualized calibrations.

The heterogeneity found between grapevine organs and phenological stages has not often been considered during calibration development. Most calibrations have been compiled by combining all data from various organs and stages to yield one prediction model (Schmidtke et al., 2012; De Bei et al., 2017; Dos Santos Costa et al., 2019; Fernández-Navales et al., 2019; Cuq et al., 2020). Calibrations could be developed not just for each grapevine organ (shoots, leaves and berries), but possibly also for different phenological stages of the organs. Specific calibrations for young leaves vs. old leaves, or green shoots vs. lignified shoots could be established. This could possibly lead to more accurate and robust individualized calibration models, as an alternative to compiling a universal calibration.

The future of infrared spectroscopy lies in the rapid prediction of key metabolites in vineyards (Schmidtke et al., 2012; Rossouw et al., 2017; Cuq et al., 2020). Infrared technologies could lead to the continuous measurement and monitoring of the metabolites throughout the growing season under diverse conditions, which has not been feasible until now. There is still a lack of knowledge regarding the mobilization, accumulation, and storage of carbohydrates, nitrogen, and amino acids under different viticultural and climatic conditions. Infrared spectroscopy might supply the solution to monitor the metabolites on a continuous basis per vineyard block or even per grapevine. Continuous measurements could greatly increase our knowledge of the metabolites' movement throughout the growing season under divergent conditions. This knowledge in turn could aid the implementation of precision viticulture and assist with viticultural decisions regarding fertilization, irrigation, and harvesting.

## REFERENCES

- Alexandre-Tudo, J. L., Nieuwoudt, H., Alexandre, J. L., and du Toit, W. (2018a). Chemometric compositional analysis of phenolic compounds in fermenting samples and wines using different infrared spectroscopy techniques. *Talanta* 176, 526–536. doi: 10.1016/j.talanta.2017.08.065
- Alexandre-Tudo, J. L., Nieuwoudt, H., and du Toit, W. (2019). Towards on-line monitoring of phenolic content in red wine grapes: A feasibility study. *Food Chem.* 270, 322–331. doi: 10.1016/j.foodchem.2018.07.118
- Alexandre-Tudo, J. L., Nieuwoudt, H., Olivieri, A., Alexandre, J. L., and du Toit, W. (2018b). Phenolic profiling of grapes, fermenting samples and wines using UV-Visible spectroscopy with chemometrics. *Food Control* 85, 11–22. doi: 10.1016/j.foodcont.2017.09.014
- Allegrini, F., and Olivieri, A. C. (2014). IUPAC-consistent approach to the limit of detection in partial least-squares calibration. *Anal. Chem.* 86, 7858–7866. doi: 10.1021/ac501786u
- Baca-Bocanegra, B., Hernández-Hierro, J. M., Nogales-Bueno, J., and Heredia, F. J. (2019). Feasibility study on the use of a portable micro near infrared

## CONCLUSION

Infrared spectroscopy has emerged as a rapid and reliable quantification method for agricultural crops. With the added benefit that infrared technology can be used for the direct measurement of intact plant material, it could lead to future agricultural applications. Infrared technologies could be implemented for the continuous monitoring of key metabolites in grapevine organs throughout the growing season. Continuous monitoring could greatly increase the knowledge of the movement and effects of these compounds under varying conditions.

The investigation of the grapevine organs' spectral properties at different phenological stages could elucidate the changes occurring in the organs throughout the growing season. The observed spectral changes could lead to the classification of grapevine organs to develop individualized calibrations, possibly leading to improved quantification. Individualized calibrations based on grapevine organs or phenological stages could be considered to compensate for the heterogeneity in grapevines and develop more robust prediction models. The field of infrared spectroscopy is complex and comprehensive and could lead to specialized solutions for the broader agricultural sector, and more specifically the viticultural industry.

## AUTHOR CONTRIBUTIONS

EW wrote the manuscript with support of JA-T, EB, and HN. EB and JA-T conceived the study, oversaw the overall direction, and planning and execution of the manuscript. EB, JA-T, and HN provided critical feedback on the manuscript. All authors helped shape the review manuscript.

## ACKNOWLEDGMENTS

The authors gratefully acknowledge the Oppenheimer Memorial Trust. A local scholarship (OMT Award Ref. 21579/01) was awarded to EW.

- spectroscopy device for the “in vineyard” screening of extractable polyphenols in red grape skins. *Talanta* 192, 353–359. doi: 10.1016/j.talanta.2018.09.057
- Bureau, S., Cozzolino, D., and Clark, C. J. (2019). Contributions of fourier-transform mid infrared (FT-MIR) spectroscopy to the study of fruit and vegetables: A review. *Postharvest Biol. Technol.* 148, 1–14. doi: 10.1016/j.postharvbio.2018.10.003
- Chandrasekaran, I., Panigrahi, S. S., Ravikanth, L., and Singh, C. B. (2019). Potential of near-infrared (NIR) spectroscopy and hyperspectral imaging for quality and safety assessment of fruits: an Overview. *Food Anal. Methods* 12, 2438–2458. doi: 10.1007/s12161-019-01609-1
- Cozzolino, D. (2014). Use of infrared spectroscopy for in-field measurement and phenotyping of plant properties: Instrumentation, data analysis and examples. *Appl. Spectrosc. Rev.* 49, 564–584. doi: 10.1080/05704928.2013.878720
- Cozzolino, D., Cynkar, W. U., Damberg, R. G., Mercurio, M. D., and Smith, P. A. (2008). Measurement of condensed tannins and dry matter in red grape homogenates using near infrared spectroscopy and partial least squares. *J. Agric. Food Chem.* 56, 7631–7636. doi: 10.1021/jf801563z

- Cozzolino, D., Cynkar, W. U., Shah, N., and Smith, P. (2011). Multivariate data analysis applied to spectroscopy: Potential application to juice and fruit quality. *Food Res. Int.* 44, 1888–1896. doi: 10.1016/j.foodres.2011.01.041
- Cuq, S., Lemetter, V., Kleiber, D., and Levasseur-Garcia, C. (2020). Assessing macro-element content in vine leaves and grape berries of *Vitis vinifera* by using near-infrared spectroscopy and chemometrics. *Int. J. Environ. Anal. Chem.* 100, 1179–1195. doi: 10.1080/03067319.2019.1648644
- Damberg, R. G., Cozzolino, D., Cynkar, W. U., Janik, L., and Gishen, M. (2006). The determination of red grape quality parameters using the LOCAL algorithm. *J. Near Infrared Spectrosc.* 14, 71–79. doi: 10.1255/jnirs.593
- Damberg, R., Gishen, M., and Cozzolino, D. (2015). A review of the state of the art, limitations, and perspectives of infrared spectroscopy for the analysis of wine grapes, must, and grapevine tissue. *Appl. Spectrosc. Rev.* 50, 261–278. doi: 10.1080/05704928.2014.966380
- De Bei, R., Fuentes, S., Sullivan, W., Edwards, E. J., Tyerman, S., and Cozzolino, D. (2017). Rapid measurement of total non-structural carbohydrate concentration in grapevine trunk and leaf tissues using near infrared spectroscopy. *Comput. Electron. Agric.* 136, 176–183. doi: 10.1016/j.compag.2017.03.007
- De Orduña, R. M. (2010). Climate change associated effects on grape and wine quality and production. *Food Res. Int.* 43, 1844–1855. doi: 10.1016/j.foodres.2010.05.001
- Diago, M. P., Fernandes, A. M., Millan, B., Tardaguila, J., and Melo-Pinto, P. (2013). Identification of grapevine varieties using leaf spectroscopy and partial least squares. *Comput. Electron. Agric.* 99, 7–13. doi: 10.1016/j.compag.2013.08.021
- Diago, M. P., Fernández-Navales, J., Gutiérrez, S., Marañón, M., and Tardaguila, J. (2018). Development and validation of a new methodology to assess the vineyard water status by on-the-go near infrared spectroscopy. *Front. Plant Sci.* 9:1–13. doi: 10.3389/fpls.2018.00059
- Dos Santos, C. A. T., Páscoa, R. N. M. J., and Lopes, J. A. (2017). A review on the application of vibrational spectroscopy in the wine industry: From soil to bottle. *TrAC Trends Anal. Chem.* 88, 100–118. doi: 10.1016/j.trac.2016.12.012
- Dos Santos Costa, D., Oliveros Mesa, N. F., Santos Freire, M., Pereira Ramos, R., and Teruel Mederos, B. J. (2019). Development of predictive models for quality and maturation stage attributes of wine grapes using vis-nir reflectance spectroscopy. *Postharvest Biol. Technol.* 150, 166–178. doi: 10.1016/j.postharvbio.2018.12.010
- Edwards, E. J., Downie, A. F., and Clingeleffer, P. R. (2011). A simple microplate assay to quantify nonstructural carbohydrates of grapevine tissues. *Am. J. Enol. Vitic.* 62, 133–137. doi: 10.5344/ajev.2010.10051
- Fernández-Navales, J., Garde-Cerdán, T., Tardaguila, J., Gutiérrez-Gamboa, G., Pérez-Álvarez, E. P., and Diago, M. P. (2019). Assessment of amino acids and total soluble solids in intact grape berries using contactless Vis and NIR spectroscopy during ripening. *Talanta* 199, 244–253. doi: 10.1016/j.talanta.2019.02.037
- Fraga, H., Malheiro, A. C., Moutinho-Pereira, J., and Santos, J. A. (2012). An overview of climate change impacts on European viticulture. *Food Energy Secur.* 1, 94–110. doi: 10.1002/fes3.14
- Frizzarin, M., Gormley, I. C., Berry, D. P., Murphy, T. B., Casa, A., Lynch, A., et al. (2021). Predicting cow milk quality traits from routinely available milk spectra using statistical machine learning methods. *J. Dairy Sci.* 104, 7438–7447. doi: 10.3168/JDS.2020-19576
- Fuentes, S., Hernández-Montes, E., Escalona, J. M., Bota, J., Gonzalez Viejo, C., Poblete-Echeverría, C., et al. (2018). Automated grapevine cultivar classification based on machine learning using leaf morpho-colorimetry, fractal dimension and near-infrared spectroscopy parameters. *Comput. Electron. Agric.* 151, 311–318. doi: 10.1016/j.compag.2018.06.035
- Gao, W., Shu, T., Liu, Q., Ling, S., Guan, Y., Liu, S., et al. (2021). Predictive modeling of lignin content for the screening of suitable poplar genotypes based on fourier transform-raman spectrometry. *ACS Omega* 6, 8578–8587. doi: 10.1021/ACSOMEGA.1C00400
- Gishen, M., Damberg, R. G., and Cozzolino, D. (2005). Grape and wine analysis - Enhancing the power of spectroscopy with chemometrics. A review of some applications in the Australian wine industry. *Aust. J. Grape Wine Res.* 11, 296–305. doi: 10.1111/j.1755-0238.2005.tb00029.x
- Gutiérrez, S., Tardaguila, J., Fernández-Navales, J., and Diago, M. P. (2015). Support vector machine and artificial neural network models for the classification of grapevine varieties using a portable NIR spectrophotometer. *PLoS One* 10:1–16. doi: 10.1371/journal.pone.0143197
- Holzappel, B. P., Smith, J. P., Field, S. K., and James Hardie, W. (2010). “Dynamics of Carbohydrate Reserves in Cultivated Grapevines,” in *Horticultural Reviews*, ed. J. Janick (New Jersey, NJ: Wiley), 143–211. doi: 10.1002/9780470543672.ch3
- Hunter, J. J., Ruffner, H. P., and Volschenk, C. G. (1995). Starch concentration in grapevines leaves, berries and roots and the effect of canopy management. *South Afr. J. Enol. Vitic.* 16, 35–40. doi: 10.21548/16-2-2270
- Jones, J., Eyles, A., Claye, C., Rodemann, T., Damberg, B., and Close, D. (2020). Prediction of starch reserves in intact and ground grapevine cane wood tissues using near-infrared reflectance spectroscopy. *J. Sci. Food Agric.* 100, 2418–2424. doi: 10.1002/jsfa.10253
- Li-Mallet, A., Rabot, A., and Geny, L. (2016). Factors controlling inflorescence primordia formation of grapevines: their role in latent bud fruitfulness? A review. *Botany* 94, 147–163. doi: 10.1139/cjb-2015-0108
- Linnet, K. (1993). Evaluation of regression procedures for methods comparison studies. *Clin. Chem.* 39, 424–432. doi: 10.1093/clinchem/39.3.424
- López, R., Tenorio, C., Rosa Gutiérrez, A., Garde-Cerdán, T., Garijo, P., González-Arenzana, L., et al. (2012). Elaboration of Tempranillo wines at two different pHs. Influence on biogenic amine contents. *Food Control* 25, 583–590. doi: 10.1016/j.foodcont.2011.11.029
- Lopo, M., Dos Santos, C. A. T., Páscoa, R. N. M. J., Graça, A. R., and Lopes, J. A. (2015). Non-invasive real-time monitoring of vineyard soils, berries and leaves with FT-NIR spectroscopy. *BIO Web Conf.* 5:01003. doi: 10.1051/bioconf/20150501003
- Lorente, D., Aleixos, N., Gómez-Sanchis, J., Cubero, S., García-Navarrete, O. L., and Blasco, J. (2012). Recent advances and applications of hyperspectral imaging for fruit and vegetable quality assessment. *Food Bioproc. Technol.* 5, 1121–1142. doi: 10.1007/s11947-011-0725-1
- Massart, D. L. (1973). The use of information theory for evaluating the quality of thin-layer chromatographic separations. *J. Chromatogr. A* 79, 157–163. doi: 10.1016/S0021-9673(01)85284-5
- Massart, D. L., Deming, S. N., Michotte, Y., Kaufman, L., and Vandeginste, B. G. M. (1988). *Chemometrics: A Textbook*. Amsterdam: Elsevier.
- Milovanovic, M., Žeravik, J., Obořil, M., Pelcová, M., Lacina, K., Cakar, U., et al. (2019). A novel method for classification of wine based on organic acids. *Food Chem.* 284, 296–302. doi: 10.1016/j.foodchem.2019.01.113
- Minasny, B., and McBratney, A. B. (2013). Why you don't need to use RPD. *Pedometron* 33, 14–15.
- Murru, C., Chimento-Trinchet, C., Díaz-García, M. E., Badía-Laiño, R., and Fernández-González, A. (2019). Artificial neural network and attenuated total reflectance-fourier transform infrared spectroscopy to identify the chemical variables related to ripeness and variety classification of grapes for Protected Designation of Origin wine production. *Comput. Electron. Agric.* 164:104922. doi: 10.1016/j.compag.2019.104922
- Musingarabwi, D. M., Nieuwoudt, H. H., Young, P. R., Eyéghè-Bickong, H. A., and Vivier, M. A. (2016). A rapid qualitative and quantitative evaluation of grape berries at various stages of development using Fourier-transform infrared spectroscopy and multivariate data analysis. *Food Chem.* 190, 253–262. doi: 10.1016/j.foodchem.2015.05.080
- Nicolai, B. M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K. I., et al. (2007). Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. *Postharvest Biol. Technol.* 46, 99–118. doi: 10.1016/j.postharvbio.2007.06.024
- Noronha, H., Silva, A., Dai, Z., Gallusci, P., Rombolà, A. D., Delrot, S., et al. (2018). A molecular perspective on starch metabolism in woody tissues. *Planta* 248, 559–568. doi: 10.1007/s00425-018-2954-2
- Olivieri, A. C. (2015). Practical guidelines for reporting results in single- and multi-component analytical calibration: A tutorial. *Anal. Chim. Acta* 868, 10–22. doi: 10.1016/j.aca.2015.01.017
- Paiva, D. N. A., Perdiz, R., de, O., and Almeida, T. E. (2021). Using near-infrared spectroscopy to discriminate closely related species: a case study of neotropical ferns. *J. Plant Res.* 134, 509–520. doi: 10.1007/S10265-021-01265-9
- Pasquini, C. (2018). Near infrared spectroscopy: A mature analytical technique with new perspectives - A review. *Anal. Chim. Acta* 1026, 8–36. doi: 10.1016/j.aca.2018.04.004
- Pérez-Roncal, C., López-Maestresalas, A., Lopez-Molina, C., Jarén, C., Urrestarazu, J., Santesteban, L. G., et al. (2020). Hyperspectral imaging to assess the presence



- of powdery mildew (*Erysiphe necator*) in cv. Carignan noir grapevine bunches. *Agron* 10, 88–103. doi: 10.3390/AGRONOMY10010088
- Petrovic, G., Aleixandre-Tudo, J.-L. L., and Buica, A. (2020). Viability of IR spectroscopy for the accurate measurement of yeast assimilable nitrogen content of grape juice. *Talanta* 206:120241. doi: 10.1016/j.talanta.2019.120241
- Phuphaphud, A., Saengprachatanarug, K., Posom, J., Maraphum, K., and Taira, E. (2020). Non-destructive and rapid measurement of sugar content in growing cane stalks for breeding programmes using visible-near infrared spectroscopy. *Biosyst. Eng.* 197, 76–90. doi: 10.1016/j.biosystemseng.2020.06.012
- Piaskowski, J. L., Brown, D., and Campbell, K. G. (2016). Near-infrared calibration of soluble stem carbohydrates for predicting drought tolerance in spring wheat. *Agron. J.* 108, 285–293. doi: 10.2134/AGRONJ2015.0173
- Porep, J. U., Kammerer, D. R., and Carle, R. (2015). On-line application of near infrared (NIR) spectroscopy in food production. *Trends Food Sci. Technol.* 46, 211–230. doi: 10.1016/j.tifs.2015.10.002
- Quentin, A. G., Rodemann, T., Doutreleau, M. F., Moreau, M., and Davies, N. W. (2017). Application of near-infrared spectroscopy for estimation of non-structural carbohydrates in foliar samples of *Eucalyptus globulus* Labillardière. *Tree Physiol.* 37, 131–141. doi: 10.1093/treephys/tpw083
- Ramirez, J. A., Posada, J. M., Handa, I. T., Hoch, G., Vohland, M., Messier, C., et al. (2015). Near-infrared spectroscopy (NIRS) predicts non-structural carbohydrate concentrations in different tissue types of a broad range of tree species. *Methods Ecol. Evol.* 6, 1018–1025. doi: 10.1111/2041-210X.12391
- Reeves, J. B. (2010). Near- versus mid-infrared diffuse reflectance spectroscopy for soil analysis emphasizing carbon and laboratory versus on-site analysis: Where are we and what needs to be done? *Geoderma* 158, 3–14. doi: 10.1016/j.geoderma.2009.04.005
- Rodgers, J., Zumba, J., and Fortier, C. (2017). Measurement comparison of cotton fiber micronaire and its components by portable near infrared spectroscopy instruments. *Text. Res. J.* 87, 57–69. doi: 10.1177/0040517515622153
- Rossouw, G. C., Smith, J. P., Barril, C., Deloire, A., and Holzapfel, B. P. (2017). Implications of the presence of maturing fruit on carbohydrate and nitrogen distribution in grapevines under postveraison water constraints. *J. Am. Soc. Hortic. Sci.* 142, 71–84. doi: 10.21273/jashs03982-16
- Rustioni, L., Grossi, D., Brancadoro, L., and Failla, O. (2017). Characterization of iron deficiency symptoms in grapevine (*Vitis* spp.) leaves by reflectance spectroscopy. *Plant Physiol. Biochem.* 118, 342–347. doi: 10.1016/j.plaphy.2017.06.031
- Santos, J. X., dos, Vieira, H. C., Souza, D. V., Menezes, M. C., de, et al. (2021). Discrimination of “Louros” wood from the Brazilian Amazon by near-infrared spectroscopy and machine learning techniques. *Eur. J. Wood Wood Prod.* 79, 989–998. doi: 10.1007/S00107-021-01685-3
- Schmidtke, L. M., Smith, J. P., Müller, M. C., and Holzapfel, B. P. (2012). Rapid monitoring of grapevine reserves using ATR-FT-IR and chemometrics. *Anal. Chim. Acta* 732, 16–25. doi: 10.1016/j.aca.2011.10.055
- Smith, J. P., Schmidtke, L. M., Müller, M. C., and Holzapfel, B. P. (2014). Measurement of the concentration of nutrients in grapevine petioles by attenuated total reflectance Fourier transform infrared spectroscopy and chemometrics. *Aust. J. Grape Wine Res.* 20, 299–309. doi: 10.1111/ajgw.12072
- Stuart, B. (2004). *Infrared spectroscopy: Fundamentals and applications*. New Jersey, NJ: John Wiley and Sons, Ltd.
- Tan, C., Chen, H., Wang, C., Zhu, W., Wu, T., and Diao, Y. (2013). A multi-model fusion strategy for multivariate calibration using near and mid-infrared spectra of samples from brewing industry. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* 105, 1–7. doi: 10.1016/J.SAA.2012.12.023
- Tardaguila, J., Fernández-Novales, J., Gutiérrez, S., and Diago, M. P. (2017). Non-destructive assessment of grapevine water status in the field using a portable NIR spectrophotometer. *J. Sci. Food Agric.* 97, 3772–3780. doi: 10.1002/jsfa.8241
- Torchio, F., Río Segade, S., Giacosa, S., Gerbi, V., and Rolle, L. (2013). Effect of growing zone and vintage on the prediction of extractable flavanols in winegrape seeds by a FT-NIR method. *J. Agric. Food Chem.* 61, 9076–9088. doi: 10.1021/jf401955m
- Türker-Kaya, S., and Huck, C. W. (2017). A review of mid-infrared and near-infrared imaging: Principles, concepts and applications in plant tissue analysis. *Molecules* 22, 168–187. doi: 10.3390/molecules22010168
- Van der Voet, H. (1994). Comparing the predictive accuracy of models using a simple randomization test. *Chemom. Intell. Lab. Syst.* 25, 313–323. doi: 10.1016/0169-7439(94)85050-X
- Varmuza, K., and Filzmoser, P. (2016). *Introduction to multivariate statistical analysis in chemometrics*. Florida, FL: CRC press.
- Veselá, A., Barros, A. S., Synytsya, A., Delgadillo, I., Čopíková, J., and Coimbra, M. A. (2007). Infrared spectroscopy and outer product analysis for quantification of fat, nitrogen, and moisture of cocoa powder. *Anal. Chim. Acta* 601, 77–86. doi: 10.1016/j.aca.2007.08.039
- Wang, J., Shen, C., Liu, N., Jin, X., Fan, X., Dong, C., et al. (2017). Non-destructive evaluation of the leaf nitrogen concentration by in-field visible/near-infrared spectroscopy in pear orchards. *Sensors* 17, 538–552. doi: 10.3390/s17030538
- Wehrens, R., and Kruisselbrink, J. (2018). Flexible self-organizing maps in kohonen 3.0. *J. Stat. Softw.* 87, 1–18. doi: 10.18637/jss.v087.i07
- Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J. Strength Cond. Res.* 19, 231–240. doi: 10.1007/978-3-642-27872-3\_5
- Williams, P. (2019). *Near Infrared Technology: Getting the best out of light*. Stellenbosch: African Sun Media.
- Xu, Z., Chen, X., Meng, L., Yu, M., Li, L., and Shi, W. (2019). Sample consensus model and unsupervised variable consensus model for improving the accuracy of a calibration model. *Appl. Spectrosc.* 73, 747–758. doi: 10.1177/0003702819852174
- Yang, Z., Tian, J., Feng, K., Gong, X., and Liu, J. (2021). Application of a hyperspectral imaging system to quantify leaf-scale chlorophyll, nitrogen and chlorophyll fluorescence parameters in grapevine. *Plant Physiol. Biochem.* 166, 723–737. doi: 10.1016/J.PLAPHY.2021.06.015
- Yen, M., and Lo, L. H. (2002). Examining test-retest reliability: An intra-class correlation approach. *Nurs. Res.* 51, 59–62. doi: 10.1097/00006199-200201000-00009
- Zapata, C., Deléens, E., Chaillou, S., and Magné, C. (2004). Mobilisation and distribution of starch and total N in two grapevine cultivars differing in their susceptibility to shedding. *Funct. Plant Biol.* 31, 1127–1135. doi: 10.1071/fp04028
- Zumba, J., Rodgers, J., and Indest, M. (2018). Impact of temperature and relative humidity on the near infrared spectroscopy measurements of cotton fiber micronaire. *Text. Res. J.* 88, 2279–2291. doi: 10.1177/0040517517720499

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 van Wyngaard, Blancauert, Nieuwoudt and Aleixandre-Tudo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.