Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/issn/15375110

Research Paper

A novel approach to upgrade infrared spectroscopy calibrations for nutritional contents in fresh grapevine organs

Elizma van Wyngaard [a], *Erna Blancquaert* [a], *Hélène Nieuwoudt* [a], *Jose L. Aleixandre-Tudo* [a,b,*]

[a] South African Grape and Wine Research Institute (SAGWRI), Department of Viticulture and Oenology, Stellenbosch University, South Africa

[b] Institute of Food Engineering (Food-UPV), Food Technology Department, Polytechnic University of Valencia, Spain

ARTICLE INFO

Article history:
Received 21 December 2022
Received in revised form
13 July 2023
Accepted 14 July 2023

Keywords:
viticulture
berries
shoots
leaves
nitrogen
carbon

Infrared spectroscopy is widely used in viticulture. Spectroscopy correlates spectral properties with reference data to obtain calibrations later used to predict the analyte content in new samples with a single spectral measurement. However, the main limitation lies in generating the reference data required to build robust prediction calibrations. This study proposes a data generation strategy to obtain reference data for larger spectral datasets. A reduced sample set was used to develop initial calibrations. These initial calibrations were subsequently applied to predict the reference data in larger spectral datasets. Calibrations for nitrogen, carbon, and hydrogen content were then attempted using the larger generated datasets. The initial nitrogen calibrations per organ showed coefficients of determination in validation (R^2val) between 80.08 and 89.93%. The root mean square errors of prediction (RMSEP) ranged from 0.10 to 0.18% dry matter, and the residual predictive deviations in validation (RPD) were between 2.27 and 3.19. The larger predicted datasets showed improved prediction accuracy with coefficients of determination in validation values above 91.79%, root mean square errors of prediction below 0.14% dry matter, and residual predictive deviations in validation above 3.49. The carbon calibrations showed, on average, a 20% increase in the coefficient of determination in validation decreased root mean square errors of prediction and increased residual predictive deviations in validation. The hydrogen calibrations showed a similar increase in prediction accuracy. The results showed the suitability of using reduced sample sets to generate the reference data of larger datasets capable of yielding more accurate prediction calibrations.

* *Corresponding author*. South African Grape and Wine Research Institute (SAGWRI), Department of Viticulture and Oenology, Stellenbosch University, South Africa.
E-mail address: joaltu@sun.ac.za (J.L. Aleixandre-Tudo).

https://doi.org/10.1016/j.biosystemseng.2023.07.008
1537-5110/

Nomenclature

ANOVA	analysis of variance
C	carbon
CHNS	carbon, hydrogen, nitrogen, sulphur
CO_2	carbon dioxide
CV	coefficients of variation
DM	dry matter
EL	Eichhorn-Lorenz
FT-NIR	Fourier transform near infrared
H	hydrogen
H_2O	water
HSD	honestly significant difference
LOQ	limit of quantification
Max	maximum
Min	minimum
MIR	mid infrared
MPA	multi-purpose analyser
MSC	multiplicative scatter correction
N	nitrogen
N_2	nitrogen gas
NIR	near infrared
NO	nitric oxide
O	oxygen
PC	principal component
PCA	principal component analysis
R^2	coefficient of determination/R squared
R^2cal	coefficient of determination for calibration
R^2val	coefficient of determination for validation
RMSE	root mean square error
RMSEE	root mean square error of estimation
RMSEP	root mean square error of prediction
RPD	residual predictive deviation/ratio of prediction to deviation
RPDcal	residual predictive deviation/ratio of prediction to deviation for calibration
RPDval	residual predictive deviation/ratio of prediction to deviation for validation
S	sulphur
SD	standard deviation
SNV	standard normal variate/vector normalisation
SO_2	sulphur dioxide
SO_3	sulphur trioxide
WO_3	tungsten (vi) oxide

1. Introduction

Infrared spectroscopy techniques are widely applied in the agriculture and are proposed as a powerful tool to obtain relevant information from fresh plant material. However, the infrared investigations of fresh grapevine organs are still in their infancy and yield large complex datasets requiring novel data exploration strategies. The applications of infrared technologies in the viticultural field with direct measurements are becoming increasingly popular and new approaches are constantly emerging (Baca-Bocanegra et al., 2019; Diago et al., 2018; Tardaguila et al., 2017). Infrared spectroscopy supplies the means for rapid and reliable measurements for quantitative analysis.

Few studies have investigated spectroscopy applications with measurements taken directly in fresh grapevine organs. Cuq et al. (2020) investigated the prediction of nitrogen, carbon, and hydrogen content in berries at two phenological stages and leaves at one phenological stage using infrared spectroscopy. However, despite the ability of the models to predict these compounds, their translocation within different organs, as well as the source-sink relationship between them is still not completely understood.

Furthermore, studies on fresh grapevine organs often focused on a single organ. Fuentes et al. (2018) investigated leaves for classification. Another study assessed berries at three maturation stages and found separation between green, véraison, and ripe berries, although some overlapping occurred (Dos Santos Costa et al., 2019). Moreover, a study performed on homogenised berries using mid infrared (MIR) showed the ability to distinguish between phenological stages such as pre-véraison and véraison (Musingarabwi et al., 2016).

Although infrared technologies supply the means to rapidly measure a large number of samples, obtaining reference data is often time consuming and reduces the number of samples that can be analysed and included during model optimisation. In addition, the analyses used to obtain reference data are often expensive, again limiting the sample's set size. Therefore, the cost and time-efficient generation or reference data to produce large sample sets could be beneficial. Larger datasets could yield more accurate predictions as increased variability can be included. In most infrared applications, spectral data from various cultivars, regions, and organs are often combined during quantitative analysis (Cuq et al., 2020; Fernandez-Novales et al., 2012; Schmidtke et al., 2012). However, if larger datasets are obtained, specialised calibrations per cultivar or organ could be explored, potentially leading to improved calibrations.

As stated, infrared studies are often limited by the amount of reference data that can be analysed. Limited reference data leads to studies that include limited cultivars, regions, vintages, or phenological stages (Cuq et al., 2020; Diago et al., 2018; Dos Santos Costa et al., 2019; Fernández-Novales et al., 2019). On the contrary, if a successful data generation strategy is developed, a wider variety of spectral properties, from different cultivars, regions or vintages, could be explored simultaneously. More comprehensive studies could also be performed throughout the growing season at different phenological stages and climatic conditions. A limited amount of reference samples containing both spectral properties and reference data could be used to construct calibrations capable of predicting the reference data in larger datasets. However, the selected samples must represent the larger dataset and include sufficient population variability. This study used the content of carbon, hydrogen, nitrogen, and sulphur (CHNS) to explore a novel reference data generation approach.

The CHNS concentration in a vineyard indicates its nutritional status. The nutritional status of grapevines is crucial for viticulturists and can influence various factors, including growth, fruit set, and grape yield (Leibar et al., 2017; Moghimi et al., 2020; Rodriguez-Lovelle & Gaudillère, 2002). These

factors, in turn, affect grape composition and impacting must and wine quality. Fertilisation is an essential tool for viticulturists to manage grapevine nutrition; however, unbalanced fertilisation adversely affects grape quality (Cuq et al., 2020; Leibar et al., 2017). Monitoring the CHNS composition of grapevine organs throughout the growing season could provide valuable information to aid viticultural decisions regarding fertilisation.

Nitrogen content in grapevines affects growth, berry development, and grape quality. Berry quality is influenced by amino acid and aromatic precursor content. Additionally, nitrogen availability influences grapevine carbon balance (Rodriguez-Lovelle & Gaudillère, 2002). Carbon composition also plays a vital role in vine growth, berry sugar accumulation, and anthocyanin biosynthesis (Cuq et al., 2020; Pastenes et al., 2014; Rossouw et al., 2018). The hydrogen content in grapevines is relevant for various chemical reactions related to respiration, carbohydrate accumulation, and the formation of multiple compounds, including sugars, starch, amino acids, and organic acids (Hellman, 2003, pp. 5–19). Sulphur content in grapes are linked to the production of amino acids and sulphur containing aromatic compounds that impact grape must quality (Bruwer et al., 2019; Lacroux et al., 2008). Monitoring nitrogen, carbon, hydrogen and sulphur concentrations throughout the growing season could provide important information on the source-sink relationship between these compounds and the grapevine organs (Cuq et al., 2020; Pastenes et al., 2014; Rodriguez-Lovelle & Gaudillère, 2002).

Therefore, the study's main aim was to investigate the use of a limited set of samples to optimise calibrations with larger sets of data. This study proposes a reference data generation strategy using a limited number of samples as starting point. NIR spectral properties were obtained from direct shoots, leaves, and berries measurements. The content of carbon, nitrogen, hydrogen, and sulphur as reference data was obtained. The suitability of the initial sample set was assessed regarding the representativeness and variability of the more extensive data set. The model prediction accuracy of the large data set was finally assessed after model optimisation.

2. Materials and methods

2.1. Sample collection

Fresh grapevine material was collected from five growing sites in the Stellenbosch district (South Africa). Twenty vineyard blocks were sampled over two vintages (2019–2020; 2020–2021), with seven cultivars included (Genus: *Vitis;* Species: *Vitis vinifera;* Cultivar: Chardonnay, Sauvignon blanc, Shiraz, Cabernet Sauvignon, Merlot, Malbec, and Pinotage). The growing season ranged from November to March for both vintages. One additional shoot sampling was conducted at dormancy in July 2021. Five intact shoots containing leaves and berries were collected from each block per month in the early mornings (06:00–10:00). The intact grapevine canes were carefully transported to the laboratory. The samples were then separated into shoots, leaves, and berries, and analyses were performed within 36 h of sampling. Samples were stored at 4 °C overnight as required (Cuq et al., 2020). Phenological stages were assigned at sampling based on the modified Eichhorn-Lorenz system of Coombe (1995). A wide range of phenological stages was incorporated per grapevine organ, with shoots ranging from EL15 (8 Leaves present) to the end of leaf fall (EL47). Berries were sampled between EL29 (Peppercorn-size) and EL39 (Over-ripe) and leaves from EL15 (8 Leaves present) to EL41 (Cane maturity). In total, 3530 samples were collected for shoots (n = 1514), leaves (n = 1540), and berries (n = 476).

2.2. Infrared spectroscopy analysis

NIR measurements were obtained using the multi-purpose analyser (MPA) Fourier transform near infrared (FT-NIR) instrument (Bruker Optics, Ettlingen, Germany) fitted with a fiber-optic solid probe. The spectral range of 12,000 to 4000 cm^{-1} was used with a resolution of 4 cm^{-1} at 10 KHz. Shoot and leaf measurements averaged 16 sample scans, and the berries averaged 64 scans. Shoots were divided into top, middle, and bottom, and each shoot section was scanned two to four times, depending on the length. Shoot sections above 35 cm were measured or scanned at four height points on alternating sides of the shoot section starting from the bottom. Sections below 35 cm were scanned twice. Five replicates were measured per leaf surface in an anti-clockwise direction corresponding to the leaf lobes. The first measurement was at the top left basal lobe, the second at the left lateral lobe, the third at the apical lobe, the fourth at the right lateral, and finally, at the top right basal lobe. Five berries were randomly selected for measurement after all berries were removed from a bunch and homogenised. Berries were measured from the side in an upright position. The total number of spectral scans was 13,828 for shoots (n = 3747), leaves (n = 7701), and berries (n = 2380). Further information about the sample distribution according to categorical variables can be found in van Wyngaard et al. (2022).

2.3. Reference analysis (CHNS)

2.3.1. Sample selection

Sample selection had to be carefully considered to represent the categorical variables and retain the variability seen in the larger datasets. The available resources determined the initial sample set size while considering a minimum number of samples required to attempt calibrations with a 50:50 split between calibration and validation. First, a representative number of samples were selected for the categorical variables vintage (120 × 2 vintages) and organ type (80 × 3 organ types). Secondly, randomised sample selection was performed on the remaining categorical variables to maintain the original proportions and variability in the dataset. Representative samples per growing site, cultivar, and phenological stage were thus selected after randomising the data. The total number of samples to be analysed for the reference data was 240.

2.3.2. Sample preparation

Shoot, leaf, and berry samples were dried and ground to a homogenous powder for analysis. Shoots and leaves were freeze-dried using a BETA 1–8 LDplus/2–8 LDplus freeze-dryer (Martin Christ, Germany) until a constant weight was

achieved (60 h for shoot samples and 12 h for leaf samples). Two strategies adapted from Cuq et al. (2020) were developed for the berries. Green berries were oven dried for 48 h at 80 °C. Véraison to ripe berries were first freeze-dried for 24 h, followed by oven drying for 48 h at 80 °C. After drying, the samples were milled using an MM400 mixer mill (Retsch, Germany) to obtain a homogenous powder and stored in airtight falcon tubes awaiting analysis.

2.3.3. CHNS analysis

The total nitrogen, carbon, hydrogen, and sulphur analysis was performed using the Vario EL Cube Elemental Analyzer (Elementar, Frankfurt, Germany). An analytical balance was used to weigh approximately 30 mg of dried, homogenous sample in an aluminium foil boat along with 5 mg of tungsten (VI) oxide (WO_3) as a catalyst for the combustion reaction. The sample was introduced at a temperature of 1050 °C into the combustion column and then dosed with oxygen at 2 bars. The C, H, N, and S elements bound in the sample combusted into their gaseous forms producing carbon dioxide (CO_2), water (H_2O), nitrogen (N_2), nitric oxide (NO), sulphur dioxide (SO_2), and sulphur trioxide (SO_3). Argon 5.0 carrier gas at 1 bar was then used to move the gaseous products through the reduction column, housing copper wire heated to 850 °C to reduce the NO to N_2 and the SO_3 to SO_2. Additionally, silver wool in the reduction column bound any volatile halogen-bound compounds produced during combustion. Finally, the pure gasses were transported to the absorption column. The N_2 gas not absorbed by the column reaches the thermal conductivity detector first while the absorption column absorbs the other gasses (CO_2, H_2O, SO_2) at room temperature. The absorption column was then sequentially heated to 60 °C to desorb the products before transferring them to the thermal conductivity detector with 220 ml/min airflow. All results were reported in percentage dry matter (%DM) with the limits of quantification (LOQ) for N, C, H, and S at 0.000, 0.238, 3.257, and 0.28% DM, respectively.

2.4. Chemometric analysis

PLS regression was performed using OPUS software (OPUS v. 7.2 for Microsoft, Bruker Optics, Ettlingen, Germany). As mentioned in section 2.2, each leaf was scanned five times, shoot sections were measured two to four times, and berries were measured individually. Therefore, several leaves, shoots or berries were used to obtain a single reference data point. This led to having multiple spectra that corresponded to a single reference data point, giving rise to two possible model optimisation strategies. Under this scenario, the spectral data contained in the single reference data point could be averaged. On the contrary, the reference data values can be replicated for each of the spectra included in the sample. Several modelling strategies were attempted to optimise the accuracy of the calibrations. First, calibrations for averaged spectra per reference value, and calibrations where reference data were repeated based on spectral repeats per sample, were explored. Moreover, a dataset split 50:50 was used to include sufficient samples in the validation and calibration datasets. Different calibration/validation split methods were explored, including a random split based on the Kennard-Stone algorithm. Additionally, various manual dataset splits were explored corresponding to 1-1-1, 2-2-2, and 2-1-1, with the first number representing the first test sample allocated, the second number the block length of test samples, and the third number the gap before the calibration samples. Optimal calibrations based on the coefficient of determination and root mean square error were selected based on all the above considerations. Spectral region selection and multiple pre-processing algorithms, including no spectral data pre-processing, constant offset elimination, straight-line subtraction, standard normal variate (SNV), also known as vector normalisation, min–max normalisation, multiplicative scattering correction (MSC), first derivative, and second derivative were also used. Combinations of pre-processing methods were also investigated, such as first derivative with straight line subtraction, first derivative with SNV, and first derivative with MSC. The optimised initial calibrations were then used to predict the nutrient content of each spectrum in the larger datasets. Calibrations with spectral pre-processing, wavenumber selection, and several calibration/validation strategies were attempted with the complete set of spectral properties and reference data.

Calibrations were evaluated using various performance evaluation indices. The coefficient of determination (R^2) was used to explain the variance in the calibration (R^2cal) and validation (R^2val) datasets. Values close to 100% indicate that a high percentage of the variance is explained by the response variable (Bureau et al., 2019; Cozzolino et al., 2011; van Wyngaard et al., 2021). Furthermore, the root mean square error of estimation (RMSEE) and prediction (RMSEP) show the model's ability to predict future samples accurately. RMSEE and RMSEP can be expressed in percentages or units of the measured compound. Residual predictive deviation (RPD) is calculated using the ratio of the standard deviation of the response variable to the RMSEE (RPDcal) and RMSEP (RPDval). Higher RPD values indicate higher model accuracy and increased capability of predicting new samples. Rank is also reported and represents the optimal number of latent variables or principal components. High rank can indicate overfitting, while low rank can lead to underfitting or underrepresenting the variability in the dataset (Aleixandre-Tudo et al., 2019; Cozzolino et al., 2011).

Agricultural applications with RMSE values below 20% for prediction models are reported as acceptable in the literature (Aleixandre-Tudo et al., 2019; Torchio et al., 2013). The interpretation of RPD values differs slightly based on the application. In general, RPD below 3 shows that discrimination between low and high predictions is possible, while RPD above 5 indicates very good to excellent predictability for quality and process control applications (Cozzolino et al., 2011). However, RPD values above 2.5 are considered acceptable for wine and grape applications, while values below 2 are mostly suited for low, medium, and high screening purposes (Aleixandre-Tudo et al., 2019; Cozzolino et al., 2008; Cuq et al., 2020).

3. Results

The chemical results of the CHNS reference analysis are first discussed. After that, the initial calibrations with a reduced

dataset are examined. The initial calibrations are then used to predict the levels of the reference compounds in the larger dataset. Furthermore, the reference data of the reduced data set obtained with the established method described in section 2.3.3 and the predicted reference data of the large data set are compared using various statistical methods. The last section compares the initial calibrations with those optimised with the larger dataset.

3.1. *Reference analysis results*

Using the randomisation approach described (section 2.3.1), 240 samples were selected. The initial reduced sample subsets' spectral data and the larger datasets were investigated using principal component analysis (PCA) (Fig. 1). The distribution of the sample set per organ type showed the reduced sample set (80 samples per organ type) to represent the larger spectral datasets, with similar distribution and variability included. The spectral distribution of the smaller datasets (in red) for berries (Fig. 1A), shoots (Fig. 1B), and leaves (Fig. 1C) compared to the larger spectral datasets (green) are shown in Fig. 1.

The descriptive statistics of the initial datasets for nitrogen, carbon, and hydrogen content are shown in Table 1. The minimum (Min) and maximum (Max) values, mean, standard deviation (SD), and coefficients of variation (CV) are included. The sulphur analysis proved to be below the limit of quantification and is therefore not shown. Similar results for sulphur levels below the limit of detection or quantification have been reported for fresh grapevine berries and leaves (Cuq et al., 2020). The nitrogen values in the leaves were the highest with a mean of 2.04% DM, the berries were second highest at 1.06% DM and the shoots showed the lowest mean percentage at 0.61% DM. The nitrogen concentrations ranged from 1.32 to 3.71% DM, 0.14–2.55% DM, and 0.34–1.61% DM for the leaves, berries, and shoots, respectively. These results agree with the values reported in the literature (Cuq et al., 2020; Leibar et al., 2017). Cuq et al. (2020) reported nitrogen values ranging from 0.39 to 2.51% DM, with a mean of 1.20% DM, and a standard deviation (SD) of 0.57. However, these values were for both berries and leaves. Another study found mean nitrogen values of 2.77% DM in leaves (Leibar et al., 2017). The CV values for nitrogen were high, possibly indicating more variation during the growing season.

The carbon concentrations were comparable between the grapevine organs, with the berries showing slightly higher values between 42.19 and 48.96% DM, the leaves ranging from 40.35 to 46.71% DM, and the shoots from 39.10 to 46.20% DM. Similar mean and SD values were seen for the three organs, and comparable carbon values ranging from 34.82 to 49.09% DM with a mean of 43.44% DM and SD of 2.98 were reported in the literature (Cuq et al., 2020). The hydrogen values between the grapevine organs were alike, with the leaves displaying a slightly higher mean value of 7.23% DM, while the berries and the shoots were 6.56 and 6.82% DM, respectively. The leaves values ranged from 6.00 to 8.03% DM, the berries from 5.56 to 7.95% DM, and the shoots from 5.65 to 7.58% DM. However, the hydrogen values Cuq et al. (2020) reported were lower, ranging from 3.17 to 6.58% DM with a mean concentration of 5.61% DM. The CV values for carbon and hydrogen were low, possibly indicating less variation in their concentrations over the growing season.

3.2. *Initial calibrations for reference data*

Calibrations for each grapevine organ and reference compound were compiled using the initial spectral and reference data. As mentioned in section 2.4, the initial calibrations were explored with average spectra per reference value and repeating reference value per spectral repeat. The optimal calibrations were selected based on the coefficient of determination and root mean square error. When spectra were averaged, 80 samples per organ were used, 40 for each calibration and validation, while with repeating reference values, the spectra per model increased. Overall, the averaged calibrations showed greater performance accuracy except for three calibrations. The shoots nitrogen ($n = 228$), leaves carbon ($n = 595$), and the leaves hydrogen ($n = 595$) calibrations with repeating reference values showed better performance. The split between calibration and validation datasets and the performance evaluation indices are reported in Table 2.

The nitrogen calibration for berries showed acceptable accuracy with the coefficient of determination values (R^2) of 89.17% for calibration (R^2cal) and 89.93% for validation (R^2val). The root mean square error of the calibration (RMSEE) was 19.30% and 17.18% for validation (RMSEP). The residual predictive deviation (RPD) for the calibration (RPDcal) and validation (RPDval) were 3.04 and 3.19, respectively. The leaves nitrogen calibration showed similar results, with the R^2cal at 89.10% and the R^2val at 89.93%. However, the RMSEE and RMSEP values were lower at 8.52 and 7.11%, respectively. Similar RPD values were seen, with RPDcal of 3.03 and RPDval of 3.18. Although the nitrogen calibration for shoots showed a slightly higher R^2cal of 91.65%, the R^2val was lower at 80.08%. The RMSEE was 11.33%, the RMSEP 16.46%, the RPDcal at 3.46, and the RPDval at 2.27. Comparable results were reported by a similar study conducted on fresh grapevine berries and leaves with R^2val of 91%, RMSEP of 0.17% DM, and RPDval of 3.32 for nitrogen (Cuq et al., 2020), while the RMSEP reported in this study as a percentage of dry matter was also between 0.10 and 0.18% DM.

The carbon calibrations overall performed poorer than the nitrogen calibrations but still yielded moderately accurate predictions. The berries' carbon calibration showed R^2cal of 70.93%, R^2val of 67.47%, RMSEE of 1.67%, RMSEP of 2.03%, RPDcal of 1.85, and RPDval of 1.79. The leaves' carbon calibration reported a higher R^2cal of 80.37%. However, the R^2val was slightly lower at 63.98%. The RPDcal was 2.26, and RPDval was 1.67. The carbon calibrations for the shoots showed similar results for the R^2cal (80.41%) and the highest R^2val (68.80%). Low RMSE were reported (below 1.6%), with similar RPDcal and RPDval of 2.26 and 1.86, respectively. Literature reported a lower R^2val of 49% for carbon calibrations of fresh berries and leaves (Cuq et al., 2020). In this study, the RMSEP per % DM was below 1, and RPDval between 1.47 and 2.61 were seen, while Cuq et al. (2020) found a RMSEP of 2.24% DM and RPDval of 1.33.

The hydrogen calibrations showed more variation in the predictability of the organs. Fairly accurate results were yielded for the hydrogen calibration of the berries, with R^2cal of

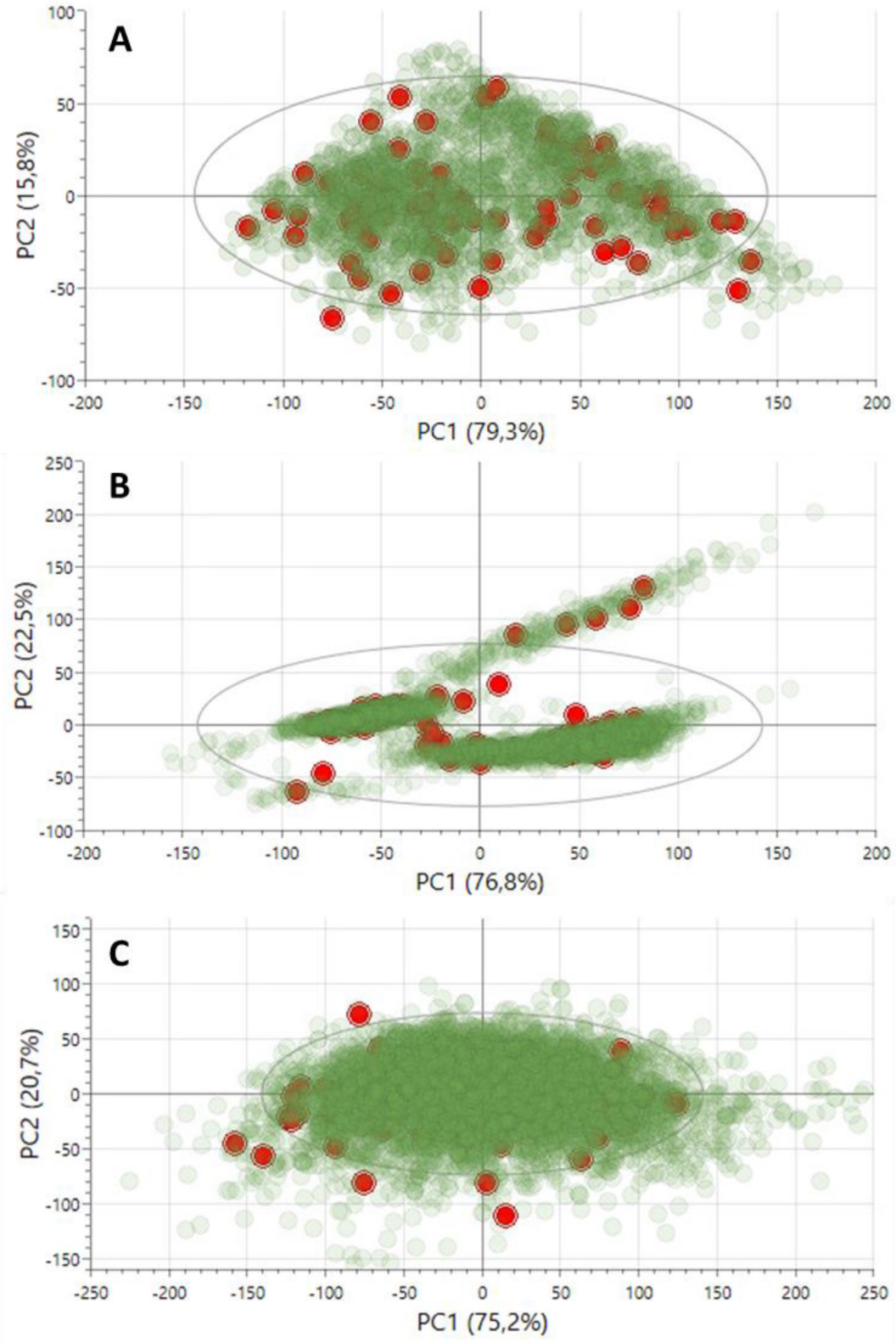

Fig. 1 – Principal component analysis (PCA) of the spectral distribution of the initial dataset (red – solid dot) compared to the larger dataset (green – transparent dot) for berries (A), shoots (B), and leaves (C).
***95% confidence *ellipse* using *Hotelling* T-squared.**

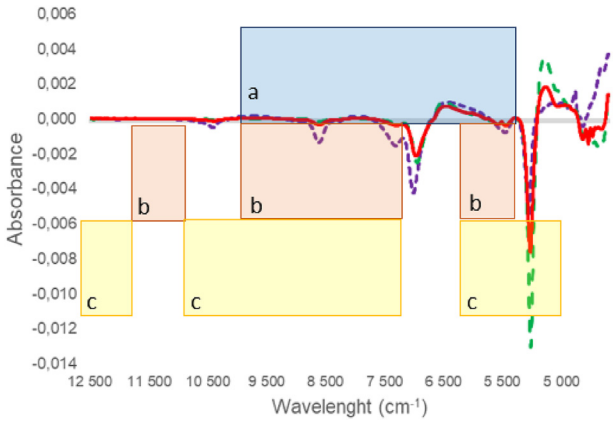

Fig. 2 – Average pre-processed spectra for berries (---), leaves (---), and shoots (—) for the initial reference data with spectral regions used for the calibrations of nitrogen (a), carbon (b), and hydrogen (c).

89.82% and R^2val of 85.27%. RMSEE of 2.67% was seen with RMSEP at 2.83%, RPDcal at 3.13, and RPDval at 2.61. The leaves hydrogen calibrations showed comparable R^2cal (85.38%) but lower R^2val (74.36%). The RMSEE and RMSEP were comparable at 2.59 and 3.33%, respectively. The RPDcal was at 2.62 and the RPDval below 2. The shoots calibration showed the lowest accuracy overall for the reference compounds and specifically for hydrogen. R^2cal of 68.45% and R^2val of 52.35% were obtained, with RMSE above 3% and RPD values below 2. A similar study reported R^2val of 56%, RMSEP of 0.27% DM, and RPDval of 1.45 for hydrogen calibrations constructed from fresh berries and leaves (Cuq et al., 2020). However, the prediction results for berries and leaves performed much better, and the shoots' calibrations were comparable. These results indicated that specialised calibrations could perform better than global calibrations combining all possible data.

The spectral pre-processing methods reported in this study included SNV, MSC, straight line subtraction, and no spectral pre-processing in specific calibrations was preferred. Some of the same pre-processing methods were reported in the literature, including no spectral pre-processing for fresh grapevine material (Cuq et al., 2020) as well as dried material (De Bei et al., 2017; Leibar et al., 2017; Schmidtke et al., 2012).

The spectral regions used during the development of each calibration are summarised in Table 2. The average raw spectra per grapevine organ were calculated, and although some spectral differences could be seen, pre-processing was used to highlight relevant bands in the spectra. First-order Savitzky–Golay spectral derivatives were applied, which include an initial smoothing step and lead to spectra where small bands are emphasised and overlapping peaks are resolved (Cuq et al., 2020; Rinnan, van den Berg, & Engelsen, 2009). A similar approach has been proposed by Cuq et al. (2020) to highlight the spectral regions that differ between grapevine organs. Figure 2 combines the pre-processed spectra per grapevine organ with the spectral regions used for the nitrogen, carbon, and hydrogen calibrations. Figure 2 visually shows the spectral regions for the reference compounds and the overlapping areas between them.

The spectral region for the nitrogen calibrations ranged from 5700 to 9950 cm^{-1}. The carbon calibrations showed some overlapping regions between 5700 and 6550 cm^{-1} and 7400 to 9950 cm^{-1}, with an additional region between 10,800 and 11,650 cm^{-1}. Similarly, the hydrogen calibrations overlapped between 5700 and 6550 cm^{-1} and 7400 to 9950 cm^{-1}. However, three unique spectral regions were seen at 4850 to 5700 cm^{-1}, 9950 to 10,800 cm^{-1}, and 11,650 to 12,500 cm^{-1}. The pre-processed spectra for the shoots, leaves, and berries showed spectral regions of interest between 4100 and 5700 cm^{-1}, and 6800 to 7400 cm^{-1}, with minor differences seen at 8500 to 9000 cm^{-1}, and 10,200 to 10,500 cm^{-1}. The area showing the most prominent differences (4100–5500 cm^{-1}) was only included in the hydrogen calibration. The region with the second largest differences (6800–7400 cm^{-1}) was only included in the nitrogen calibrations.

Similar spectral regions were reported in the literature and previous studies (Cuq et al., 2020; van Wyngaard et al., 2022). The investigation of raw spectra showed that the regions between 4500 and 5300 cm^{-1} and 6000 to 7300 cm^{-1} are associated with shoots, leaves, and berries, while 8000 to 8700 cm^{-1} and 9500 to 10,500 cm^{-1} relate only to berries. This corresponds to the regions seen in Fig. 2, with the additional bands shown for berries on the left side of the graph. A similar study reported comparable regions (6900–7400, 8300–9100, and 10,250–10,950 cm^{-1}) for fresh berries and leaves and associated them with the first vibrational overtones of OH, CH, and NH (Cuq et al., 2020). The investigation of dried trunk and leaf

Table 1 – Descriptive statistics of the nitrogen, carbon, and hydrogen reference analysis for the initial dataset.

	Min*	Max*	Mean*	SD*	CV (%)**
Berries Nitrogen (% DM)	0.14	2.55	1.06	0.57	53.69
Shoots Nitrogen (% DM)	0.34	1.61	0.61	0.23	37.44
Leaves Nitrogen (% DM)	1.32	3.71	2.04	0.47	23.26
Berries Carbon (% DM)	42.19	48.96	45.33	1.49	3.28
Shoots Carbon (% DM)	39.10	46.20	44.78	1.27	2.85
Leaves Carbon (% DM)	40.35	46.71	43.92	1.21	2.76
Berries Hydrogen (% DM)	5.56	7.95	6.56	0.49	7.47
Shoots Hydrogen (% DM)	5.65	7.58	6.82	0.35	5.13
Leaves Hydrogen (% DM)	6.00	8.03	7.23	0.47	6.46

*Minimum (Min), maximum (Max), mean, and standard deviation (SD) are shown in % dry matter (DM).
**Coefficient of variation (CV) is shown in percentage.

Table 2 – Initial calibrations per reference compound (nitrogen, carbon, hydrogen) and per grapevine organ (berries, leaves, shoots).

Nitrogen	Spectral Pre-processing	Spectral Regions (cm^{-1})	Cal/Val**	Rank	R^2cal	R^2val	RMSEE(% DM)*	% RMSEE*	RMSEP (% DM)*	% RMSEP*	RPDcal	RPDval	Bias
Berries	Vector normalisation (SNV)	8250-7400; 6550-5700	40/40	7	89,17	89,93	0,21	19,30	0,18	17,18	3,04	3,19	−0,0270
Leaves	Multiplicative scattering correction (MSC)	9100-8250; 7400-5700	40/40	7	89,10	89,93	0,17	8,52	0,15	7,11	3,03	3,18	−0,0204
Shoots	No spectral data pre-processing	9950-8250; 6550-5700	114/114	12	91,65	80,08	0,07	11,33	0,10	16,46	3,46	2,27	−0,0148
Carbon													
Berries	Straight line subtraction	11,650-10800; 9950-9100; 8250-7400	40/40	4	70,93	67,47	0,75	1,67	0,92	2,03	1,85	1,79	0,1880
Leaves	No spectral data pre-processing	9100-7400; 6550-5700	297/298	13	80,37	63,98	0,53	1,20	0,72	1,64	2,26	1,67	0,0043
Shoots	Straight line subtraction	9950−8250	40/40	4	80,41	68,80	0,69	1,54	0,58	1,29	2,26	1,86	0,1590
Hydrogen													
Berries	Multiplicative scattering correction (MSC)	12,500-11650; 10,800-9100; 8250-7400; 5700-4850	40/40	7	89,82	85,27	0,18	2,67	0,19	2,83	3,13	2,61	−0,0025
Leaves	Straight line subtraction	9100-7400; 6550-5700	298/297	11	85,38	74,36	0,19	2,59	0,24	3,33	2,62	1,97	0,0032
Shoots	Straight line subtraction	9100−7400	40/40	6	68,45	52,35	0,21	3,06	0,25	3,60	1,78	1,47	−0,0459

*RMSE as percentage dry matter (% DM) and as percentage calculated with the population mean.
**Cal/Val – spectra included for each calibration and validation dataset split.

samples identified three mutual absorption bands (4300, 5200, and 7000 cm^{-1}) and various additional bands for leaves (4900, 5900, 6600, and 8300 cm^{-1}) (De Bei et al., 2017). The mutual bands were associated with starch (CH, OH related bands), while the additional leaf bands corresponded to protein and nitrogen.

3.3. *Predictions of the large datasets using the initial calibrations*

The initial calibrations in the previous section were used to predict the larger dataset' nitrogen, carbon, and hydrogen content. The predicted results were compared to the initial reference analysis and investigated further for accuracy and validation. The descriptive statistics of the initial reference analysis are shown in Table 1, and the predicted values are summarised in Table 3. However, prior to calculating the descriptive statistics for the predicted values, outlier removal was implemented. Outlier removal for the predicted values was based on the minimum (Min) and maximum (Max) range of values seen for the reference analysis. They were removed when the predicted values for the nitrogen, carbon, and hydrogen were more than 0.1–0.2% DM below or above the minimum and maximum reference values. However, the outliers removed per dataset were very limited, and the percentage of outliers removed (%OR) was never higher than 4.1%. As expected, outlier removal led to comparable minimum and maximum results, as shown in Tables 1 and 3 Furthermore, the mean, SD, and CV were similar for the reference and predicted values. The CV values for the nitrogen were once again higher, as was seen for the reference values, while the carbon and hydrogen CV values were low. The CV values could indicate variation over the growing season, with nitrogen showing more variability than carbon and hydrogen.

The nitrogen, carbon, and hydrogen concentrations over the growing season were also explored for the initial reference and predicted data. This was done to assess if the predicted data behaved similarly to the reference data and to validate the predicted results from the initial calibrations. The nitrogen reference data in the berries showed the most noticeable change over the growing season (Fig. 3). However, the carbon and hydrogen reference data overall remained constant across progressing phenological stages. The predicted values showed the same trend for nitrogen in Fig. 4, with the most prominent decrease observed in berries. Literature has reported a similar reduction in nitrogen concentration in leaves and berries over the growing season (Cancela et al., 2018; Ferrara et al., 2018; Romero et al., 2010; Rossouw et al., 2017).

The reference data of the initial dataset and the predicted values showed the most prominent trend for the berries' nitrogen. The nitrogen content in berries per phenological stage was further investigated and compared to determine whether significant differences occurred during the growing season. One-way analysis of variance (ANOVA) and post-hoc analysis using Tukey's honestly significant difference (HSD) test was employed. Significant p-values were reported using one-way ANOVA for both the reference and predicted datasets below the significance level of 0.1% ($p < 0.001$), showing that

Table 3 – Descriptive statistics for the nitrogen, carbon, and hydrogen reference analysis for the predicted dataset.

	Min*	Max*	Mean*	SD*	CV (%)**
Berries Nitrogen (% DM)	0.04	2.32	1.13	0.53	46.74
Shoots Nitrogen (% DM)	0.17	1.54	0.63	0.21	32.82
Leaves Nitrogen (% DM)	1.10	3.89	2.08	0.43	20.67
Berries Carbon (% DM)	42.11	48.73	45.36	1.02	2.25
Shoots Carbon (% DM)	40.44	46.41	44.62	1.15	2.57
Leaves Carbon (% DM)	40.08	46.83	43.98	0.87	1.98
Berries Hydrogen (% DM)	5.30	8.12	6.56	0.49	7.52
Shoots Hydrogen (% DM)	5.81	7.61	6.83	0.25	3.67
Leaves Hydrogen (% DM)	5.92	8.13	7.18	0.37	5.17

*Minimum (Min), maximum (Max), mean, and standard deviation (SD) are shown in % dry matter (DM).
**Coefficient of variation (CV) is shown in percentage.

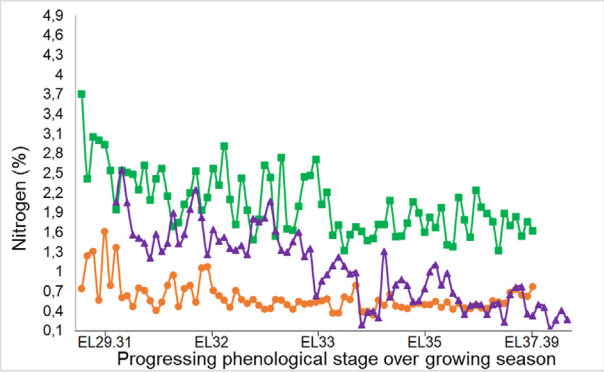

Fig. 3 – Nitrogen trends for the initial reference data at progressing phenological stages over the growing season for shoots (●), leaves (■), and berries (▲) (240 samples). Phenological stages are EL29.31 – Peppercorn-size to Pea-size, EL32 – Bunch closure, EL33 – Hard-green, EL35 – Véraison, EL37.39 – Almost-ripe, Harvest, and Over-ripe (Coombe, 1995).

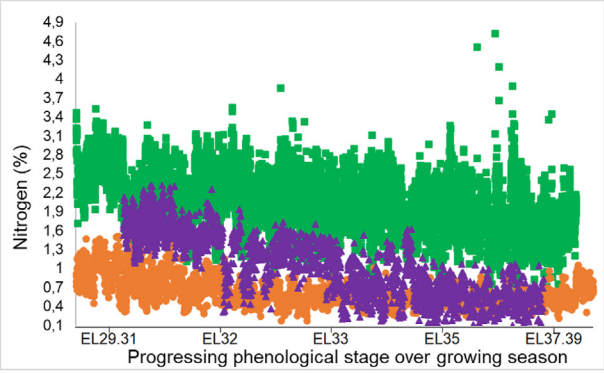

Fig. 4 – Nitrogen trends for the predicted data at progressing phenological stages over the growing season for shoots (●), leaves (■), and berries (▲) (13,828 samples). Phenological stages are EL29.31 – Peppercorn-size to Pea-size, EL32 – Bunch closure, EL33 – Hard-green, EL35 – Véraison, EL37.39 – Almost-ripe, Harvest, and Over-ripe (Coombe, 1995).

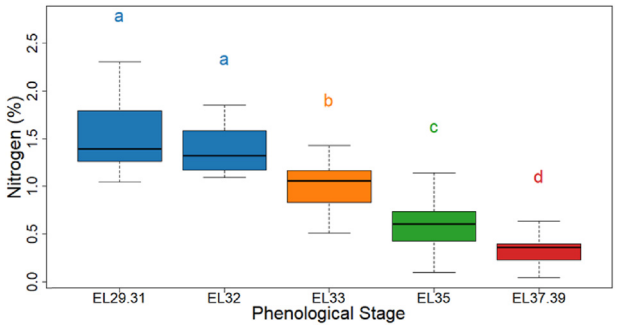

Fig. 5 – Box plot representing nitrogen distribution per phenological stage for the berries' initial reference data. *Different letters indicate significant differences at 5% significance level ($p < 0.05$). **EL29.31 – Peppercorn-size to Pea-size (a), EL32 – Bunch closure (a), EL33 – Hard-green (b), EL35 – Véraison (c), EL37.39 – Almost-ripe, Harvest, and Over-ripe (d) (Coombe, 1995).

differences between the phenological stages could be detected. Post-hoc analyses were further used to determine which phenological stages differed, and box plots of the data distribution were used with significant letters corresponding to a 5% significance level ($p < 0.05$).

The boxplot representing the reference data (Fig. 5) showed significant differences between all the phenological stages except EL29-31 (Peppercorn-size to Pea-size) and EL32 (Bunch-closure). These phenological stages are very close together, and the smaller number of samples per stage for the reference data could also have led to these results. Furthermore, the box plot for the predicted values (Fig. 6) showed significant differences between all the stages. Once again, the decreasing trend of nitrogen was seen over the growing season, and the post-hoc test additionally showed

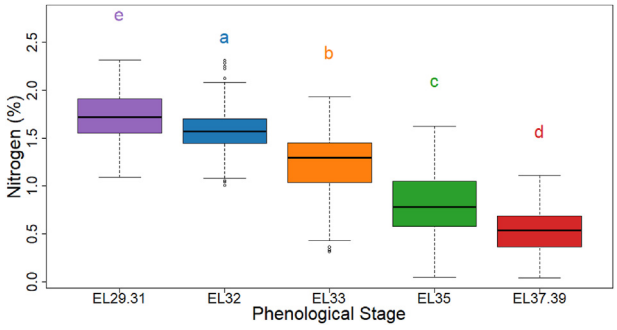

Fig. 6 – Box plot representing nitrogen distribution per phenological stage for the berries' predicted data. *Different letters indicate significant differences at 5% significance level ($p < 0.05$). **EL29.31 – Peppercorn-size to Pea-size (e), EL32 – Bunch closure (a), EL33 – Hard-green (b), EL35 – Véraison (c), EL37.39 – Almost-ripe, Harvest, and Over-ripe (d) (Coombe, 1995).

the decreases to be significant between each phenological stage.

The descriptive statistics (Tables 1 and 3), trend graphs (Figs. 3 and 4), and box plots (Figs. 5 and 6) all showed that the predicted data behaved similarly to the reference data and could therefore be used for further calibration development.

3.4. *Comparison of initial calibrations to larger dataset calibrations*

The calibrations per organ were investigated using the larger dataset with the predicted reference data. The calibration statistics are summarised in Table 4. The number of spectra included in the calibration and validation split, with the percentage outliers removed (%OR) for each model is shown in Table 4. The total spectra were 2380, 7701, 3747 for berries, leaves and shoots, respectively. The berries nitrogen calibration reported R^2cal of 93.64% and R^2val of 92.66%, with RMSEE of 11.60% and RMSEP of 12.86%. The RPD values were 3.97 and 3.70 for calibration and validation, respectively. The leaves nitrogen calibration showed similarly accurate results with R^2cal of 91.92%, R^2val of 91.79%, RMSEE of 5.95%, RMSEP of 5.85%, RPDcal of 3.52, and RPDval of 3.49%. The shoots nitrogen calibration showed slightly better results, with R^2cal and R^2val 93.98% and 92.85%, respectively, while the RMSEE was at 8.06% and the RMSEP at 8.77%. The RPD values for the calibration were above 4, with the RPDval reported at 3.74. Overall, extremely accurate prediction results were observed for nitrogen for all three organs.

The berries' carbon calibration showed the highest prediction accuracy with R^2cal of 94.60%, R^2val of 90.65%, RMSEE of 0.50%, and RMSEP of 0.71%. The RPDcal value was 4.30, and the RPDval was 3.27. The carbon calibration for the leaves showed a slight decrease in predictability compared to the berries' calibration. R^2cal of 86.39% and R^2val of 84.99% were observed for the leaves' carbon calibration with RMSE below 0.8% and RPD values above 2.5. The shoots' carbon calibration showed R^2cal and R^2val of 91.26% and 90.06%, respectively. RMSE values were also close to 0.8%, RPDcal at 3.38, and RPDval at 3.17. Overall, all three organs showed very good to excellent accuracy for carbon prediction.

The hydrogen calibration for the berries showed improved results with R^2cal of 92.72%, R^2val of 89.01%, RMSEE of 1.97%, RMSEP of 2.54%, RPDcal of 3.71, and RPDval of 3.02. Further improvements were seen for the leaves' hydrogen calibration with R^2cal and R^2val at 95.22% and 94.83%, respectively, RMSE below 1.2%, and RPD values above 4.4. The hydrogen calibration for the shoots displayed the slightest improvement. The shoots hydrogen calibration showed R^2cal of 80.53% and R^2val of 75.22%, with RMSEE of 1.60%, RMSEP of 1.84%, RPDcal at 2.27, and RPDval at 2.01.

The calibrations per grapevine organ for nitrogen, carbon, and hydrogen showed more accurate predictability than reported in the literature for fresh grapevine material (Cuq et al., 2020). Cuq et al. (2020) found R^2cal, R^2val, RPDcal, and RPDval values of 90%, 84%, 3.14, and 2.41 for nitrogen, respectively. The carbon prediction results were R^2cal of 67%, R^2val of 57%, RPDcal at 1.74, and RPDval at 1.49. Hydrogen was poorly predicted at R^2cal of 54%, R^2val of 50%, RPDcal at 1.46, and RPDval at 1.38 (Cuq et al., 2020). The results found in this study

Table 4 – Calibrations using larger dataset per reference compound (nitrogen, carbon, hydrogen) and per grapevine organ (berries, leaves, shoots).

Nitrogen	Spectral pre-processing	Spectral regions (cm^{-1})	Cal/Val (%OR)**	Rank	R^2cal	R^2val	RMSEE (% DM)*	% RMSEE*	RMSEP (% DM)*	% RMSEP*	RPDcal	RPDval	Bias
Berries	Straight line subtraction	12,500-11650; 6550-5700	1180/1179 (0.9%)	5	93,64	92,66	0,13	11,60	0,14	12,86	3,97	3,70	0,0067
Leaves	Straight line subtraction	10,800-9950; 6550-5700	3803/3786 (1.5%)	7	91,92	91,79	0,12	5,95	0,12	5,85	3,52	3,49	−0,0005
Shoots	First derivative	9950-7400; 6550-5700	1873/1874 (0.0%)	8	93,98	92,85	0,05	8,06	0,06	8,77	4,08	3,74	0,0013
Carbon													
Berries	Straight line subtraction	11,650-10800; 8250-7400	1180/1179 (0.9%)	7	94,60	90,65	0,23	0,50	0,32	0,71	4,30	3,27	−0,0419
Leaves	First derivative	9100-7400; 6550-5700	3850/3849 (0.0%)	8	86,39	84,99	0,32	0,73	0,34	0,77	2,71	2,58	0,0037
Shoots	Straight line subtraction	10,800-9950; 9100-8250	1797/1796 (4.1%)	7	91,26	90,06	0,34	0,75	0,37	0,82	3,38	3,17	−0,0090
Hydrogen													
Berries	First derivative + Vector normalisation (SNV)	9100-8250; 5700-4850	1172/1167 (1.7%)	7	92,72	89,01	0,13	1,97	0,17	2,54	3,71	3,02	0,0092
Leaves	First derivative + MSC	9100-7400; 6550-5700	3850/3851 (0.0%)	8	95,22	94,83	0,08	1,12	0,08	1,18	4,58	4,40	−0,0034
Shoots	Straight line subtraction	9950-9100; 8250-7400	1873/1874 (0.0%)	7	80,53	75,22	0,11	1,60	0,13	1,84	2,27	2,01	0,0032

*RMSE as percentage dry matter (% DM) and as percentage calculated with the population mean.
**Cal/Val – spectra included for each calibration and validation dataset split and the percentage outliers removed (%OR).

showed significant improvement from the reported results. The pre-processing methods reported per calibration were similar for nitrogen and carbon including straight line subtraction and first derivative. The hydrogen calibration reported using the first derivative with SNV and MSC, and straight-line subtraction. Similar studies have also reported using the first derivative and MSC (Cuq et al., 2020; Dos Santos Costa et al., 2019).

The calibrations developed per grapevine organ from the larger dataset with predicted values showed substantially improved accuracy compared to the initial calibrations using a reduced sample set (Tables 2 and 4). The initial nitrogen calibrations reported R^2 values ranging from 80.08 to 91.65%, RMSE from 7.11 to 19.30%, and RPDs between 2.27 and 3.46 across all organs. The predicted datasets for nitrogen calibrations of all organs reported increased R^2 values (91.79–93.98%), lower RMSE (5.85–12.86%), and higher RPDs (3.49–4.08). Similar improved results were seen for the carbon calibrations with the initial R^2 values between 63.98 and 80.41%, and those with the larger predicted datasets from 84.99 to 94.60%. The RMSE values decreased from the initial calibrations (1.20–2.03%) to the predicted dataset calibrations (0.50–0.82%). Increases in the RPD were also seen from the initial (1.67–2.26) to the predicted (2.58–4.30) calibrations. The predicted reference data calibrations for hydrogen showed considerable improvement in R^2 values between 75.22 and 95.22% from the initial values (52.35–89.82%). Decreased RMSE values were seen for the predicted calibrations (1.12–2.54%) compared to the initial calibrations (2.59–3.60%). The RPD values also increased from the initial calibrations (1.47–3.13) to the predicted (2.01–4.58).

The raw spectra per grapevine organs were averaged and pre-processed using the Savitzky–Golay algorithm with first-order derivative, as discussed in section 3.2 (Fig. 7). The spectral graph obtained was very similar to Fig. 2, and the same absorption bands of interest were seen (4100–5700, 6800–7400, 8500–9000, and 10,200–10,500 cm^{-1}). In our previous studies, three spectral regions of interest were identified (5115–5240, 8830–9800, and 10,600–11,300 cm^{-1}) (van Wyngaard et al., 2022) and similar regions have been reported in the literature for fresh grapevine leaves and berries (Cuq et al., 2020). These regions can be observed in Fig. 7, with more distinct differences between 5115 and 5240 cm^{-1} and subtle differences in the other regions (8830–9800, 10,600–11,300 cm^{-1}).

The spectral regions reported in Table 4 for the per organ calibrations developed with the predicted datasets (Fig. 7) showed similarities to those seen for the calibrations using reference data (Fig. 2). However, it seems that more extensive regions of the spectra were incorporated for each reference compound with overlapping regions between 5700 and 6550 cm^{-1} and 7400 to 11,650 cm^{-1}. The larger regions could indicate that more comprehensive spectral regions are needed during the calibration development of larger datasets with more variability. An additional region (4000–4850 cm^{-1}) was seen for the global nitrogen calibration and was previously reported in the literature to correspond to nitrogen in a wide range of plant leaves (Johnson, 2001). Interestingly, once again, the spectral region showing the most prominent differences (4100–5500 cm^{-1}) between the organs was only used

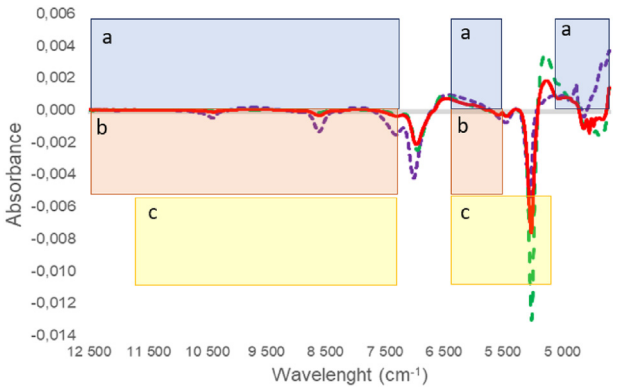

Fig. 7 – Average pre-processed spectra for berries (---), leaves (---), and shoots (——) for the predicted data with spectral regions used for the calibrations of nitrogen (a), carbon (b), and hydrogen (c).

in the hydrogen calibration for berries. Literature has suggested that the 5200 cm^{-1} absorption band could be linked to the OH first overtones and O–H stretching/HOH deformation of water (De Bei et al., 2017; Fernández-Novales et al., 2019; González-Caballero et al., 2012). The possible association with water could explain why the large peak shows differences between the organs but was not included in the calibrations, except for hydrogen.

4. Discussion

The nitrogen, carbon, and hydrogen concentrations reported in fresh grapevine shoots, leaves, and berries were comparable with the literature (Cuq et al., 2020; Leibar et al., 2017). However, these studies did not report the levels of these compounds throughout the growing season. This study showed that carbon and hydrogen concentrations remained constant, but nitrogen concentrations decreased in leaves and berries (Figs. 3 and 4). This behaviour can be attributed to the source-sink relationship and how nitrogen translocation changes between organs at different phenological stages. The decrease in nitrogen concentration in the leaves could indicate remobilisation towards the berries that acts as a stronger sink during early ripening before véraison. Later in the season, the leaves could also translocate nitrogen to the permanent grapevine organs for storage (Ferrara et al., 2018). These changes are in line with the results observed in Fig. 3, which shows the nitrogen concentration in the berries remaining constant during early ripening and then decreasing towards the final stages of ripening.

Infrared spectroscopy can efficiently analyse numerous heterogeneous viticultural samples leading to large datasets containing thousands of samples. However, there will always be a restricted capacity when performing reference analyses on large datasets due to resource constraints. Although the initial calibrations per organ and reference compound demonstrated good to moderate prediction ability, they were used to predict the reference data in large datasets and led to

accurate calibrations being developed. The increased accuracy of the calibrations validates the data generation strategy and demonstrates that a reduced sample set could be used to upgrade the calibrations. In addition, most viticultural and oenological infrared applications report small sample sets (below 100) for calibration development (Dambergs et al., 2015). The considerably large sample numbers of up to 7701 in our predicted datasets could have influenced the calibration's accuracy (Porep et al., 2015). However, the literature reports that sample variability has a more significant impact than sample numbers (Bureau et al., 2019; Porep et al., 2015). Selecting samples that are representative and include sufficient variability could improve calibrations. Thus, especially for agricultural and viticulture samples, various cultivars, growing sites, vintages, grapevine organs, and phenological stages should be included (Bureau et al., 2019). Although initial reference samples were selected represent all the categorical variables, the variability in the predicted datasets, with thousands of samples, could have still been increased. The updated calibrations for nitrogen, carbon, and hydrogen were also more accurate than those reported in the literature for fresh grapevine organs (Cuq et al., 2020).

The application of these calibrations could supply information on nutritional content per organ during the growing season (Cuq et al., 2020; Pastenes et al., 2014; Rodriguez-Lovelle & Gaudillère, 2002; Rossouw et al., 2017). The accumulation and mobilisation of nitrogen and carbon constantly change during the growing cycle. Furthermore, respiration reactions linked to hydrogen can change rapidly (Hellman, 2003, pp. 5–19). The ability to measure and monitor these changes could supply relevant knowledge on the growth dynamics and nutritional status of a vineyard and facilitate fertilisation decisions (Rodriguez-Lovelle & Gaudillère, 2002; Schmidtke et al., 2012). For example, lower nitrogen concentrations before véraison could aid berry development, while enlargement could be inhibited. The nitrogen content could also influence the amount of amino acid and aromatic precursors in grapes and thus affect grape quality (Rodriguez-Lovelle & Gaudillère, 2002). Fertilisation applications per organ, such as foliar or bunch treatments, could be considered if nutritional content is decreased.

The measurement of nutritional content could also lead to a better understanding of the source (leaves and shoots) sink (berries) relationship (Cuq et al., 2020; Leibar et al., 2017; Pastenes et al., 2014; Rodriguez-Lovelle & Gaudillère, 2002) leading to specialised decisions regarding fertilisation. Viticulturists are under increasing pressure to use fewer resources while achieving optimal grape quality (Leibar et al., 2017). Improved resource management, in turn, could optimise berry development and lead to better grape and must quality (Moghimi et al., 2020; Rodriguez-Lovelle & Gaudillère, 2002).

5. Conclusion

The possibility to use a reduced set of samples to generate reference data in larger datasets was the main finding of this study. This might lead to studies being able to explore larger sample sets with expanded variability containing data from multiple cultivars, regions, and vintages simultaneously. Previous infrared and viticultural studies have been limited in the number of samples and reference compounds analysed. This strategy could supply the means to do more holistic viticultural monitoring throughout the growing season. For example, studies exploring various climatic conditions on a larger scale could be implemented. Various growth parameters could be explored with the prediction of chemical components to better understand the plant and grape development during the season. In addition, the data generation strategy presented here is not only applicable to grapevines and could therefore be implemented for broader agricultural uses. Furthermore, the ability to generate larger datasets could lead to novel model optimisation strategies. In previous spectroscopy studies, data from various cultivars, regions, organs, and phenological stages are often combined to obtain sufficient samples for calibration optimisation. However, if larger datasets were generated, specialised calibrations per, for example, cultivar, organ or phenological stage could be explored. These calibrations could lead to improved prediction accuracy and calibrations capable of more specific and targeted predictions.

Declaration of competing interest

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

Acknowledgments

The authors gratefully acknowledge the Oppenheimer Memorial Trust, for the local scholarship (OMT Award Ref. 21579/02) awarded to Ms Elizma van Wyngaard.

REFERENCES

Aleixandre-Tudo, J. L., Nieuwoudt, H., & du Toit, W. (2019). Towards on-line monitoring of phenolic content in red wine grapes: A feasibility study. *Food Chemistry, 270*, 322–331. https://doi.org/10.1016/j.foodchem.2018.07.118

Baca-Bocanegra, B., Hernández-Hierro, J. M., Nogales-Bueno, J., & Heredia, F. J. (2019). Feasibility study on the use of a portable micro near infrared spectroscopy device for the "in vineyard" screening of extractable polyphenols in red grape skins. *Talanta, 192*, 353–359. https://doi.org/10.1016/j.talanta.2018.09.057

Bruwer, F. A., du Toit, W., & Buica, A. (2019). Nitrogen and sulphur foliar fertilisation. *South African Journal for Enology & Viticulture, 40*, 1–16. https://doi.org/10.21548/40-2-3281

Bureau, S., Cozzolino, D., & Clark, C. J. (2019). Contributions of Fourier-transform mid infrared (FT-MIR) spectroscopy to the study of fruit and vegetables: A review. *Postharvest Biology and Technology, 148*, 1–14. https://doi.org/10.1016/j.postharvbio.2018.10.003

Cancela, J. J., Fandiño, M., González, X. P., Rey, B. J., & Mirás-Avalos, J. M. (2018). Seasonal variation of macro and micronutrients in blades and petioles of Vitis vinifera L. cv.

Mencía and Sousón. *Journal of Plant Nutrition and Soil Science, 181*, 498–515. https://doi.org/10.1002/jpln.201700446

Coombe, B. G. (1995). Growth stages of the grapevine: Adoption of a system for identifying grapevine growth stages. *Australian Journal of Grape and Wine Research, 1*, 104–110. https://doi.org/10.1111/j.1755-0238.1995.tb00086.x

Cozzolino, D., Cynkar, W. U., Dambergs, R. G., Mercurio, M. D., & Smith, P. A. (2008). Measurement of condensed tannins and dry matter in red grape homogenates using near infrared spectroscopy and partial least squares. *Journal of Agricultural and Food Chemistry, 56*, 7631–7636. https://doi.org/10.1021/jf801563z

Cozzolino, D., Cynkar, W. U., Shah, N., & Smith, P. (2011). Multivariate data analysis applied to spectroscopy: Potential application to juice and fruit quality. *Food Research International, 44*, 1888–1896. https://doi.org/10.1016/j.foodres.2011.01.041

Cuq, S., Lemetter, V., Kleiber, D., & Levasseur-Garcia, C. (2020). Assessing macro-element content in vine leaves and grape berries of vitis vinifera by using near-infrared spectroscopy and chemometrics. *International Journal of Environmental Analytical Chemistry, 100*, 1179–1195. https://doi.org/10.1080/03067319.2019.1648644

Dambergs, R., Gishen, M., & 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. *Applied Spectroscopy Reviews, 50*, 261–278. https://doi.org/10.1080/05704928.2014.966380

De Bei, R., Fuentes, S., Sullivan, W., Edwards, E. J., Tyerman, S., & Cozzolino, D. (2017). Rapid measurement of total non-structural carbohydrate concentration in grapevine trunk and leaf tissues using near infrared spectroscopy. *Computers and Electronics in Agriculture, 136*, 176–183. https://doi.org/10.1016/j.compag.2017.03.007

Diago, M. P., Fernández-Novales, J., Gutiérrez, S., Marañón, M., & Tardaguila, J. (2018). Development and validation of a new methodology to assess the vineyard water status by on-the-go near infrared spectroscopy. *Frontiers in Plant Science, 9*, 1–13. https://doi.org/10.3389/fpls.2018.00059

Dos Santos Costa, D., Oliveros Mesa, N. F., Santos Freire, M., Pereira Ramos, R., & Teruel Mederos, B. J. (2019). Development of predictive models for quality and maturation stage attributes of wine grapes using vis-nir reflectance spectroscopy. *Postharvest Biology and Technology, 150*, 166–178. https://doi.org/10.1016/j.postharvbio.2018.12.010

Fernández-Novales, J., Garde-Cerdán, T., Tardáguila, J., Gutiérrez-Gamboa, G., Pérez-Álvarez, E. P., & 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. https://doi.org/10.1016/j.talanta.2019.02.037

Ferrara, G., Malerba, A. D., Matarrese, A. M. S., Mondelli, D., & Mazzeo, A. (2018). Nitrogen distribution in annual growth of 'Italia' table grape vines. *Frontiers in Plant Science, 9*, 1374. https://doi.org/10.3389/fpls.2018.01374

Fuentes, S., Hernández-Montes, E., Escalona, J. M., Bota, J., Gonzalez Viejo, C., Poblete-Echeverría, C., Tongson, E., & Medrano, H. (2018). Automated grapevine cultivar classification based on machine learning using leaf morpho-colorimetry, fractal dimension and near-infrared spectroscopy parameters. *Computers and Electronics in Agriculture, 151*, 311–318. https://doi.org/10.1016/j.compag.2018.06.035

González-Caballero, V., Sánchez, M. T., Fernández-Novales, J., López, M. I., & Pérez-Marín, D. (2012). On-vine monitoring of grape ripening using near-infrared spectroscopy. *Food Analytical Methods, 5*, 1377–1385. https://doi.org/10.1007/s12161-012-9389-3

Hellman, E. W. (2003). *"Grapevine Structure and Function," in* Oregon Viticulture. Corvallis, OR, USA: Oregon State University Press. https://doi.org/10.1016/b978-012379062-0/50004-4

Johnson, L. F. (2001). Nitrogen influence on fresh-leaf NIR spectra. *Remote Sensing of Environment, 78*, 314–320. https://doi.org/10.1016/S0034-4257(01)00226-7

Lacroux, F., Tregoat, O., Van Leeuwen, C., Pons, A., Tominaga, T., Lavigne-Cruège, V., et al. (2008). Effect of foliar nitrogen and sulphur application on aromatic expression of Vitis vinifera L. cv. Sauvignon blanc. *Journal International des Sciences de la Vigne et du Vin, 42*, 125–132. https://doi.org/10.20870/oeno-one.2008.42.3.816

Leibar, U., Pascual, I., Aizpurua, A., Morales, F., & Unamunzaga, O. (2017). Grapevine nutritional status and K concentration of must under future expected climatic conditions texturally different soils. *Journal of Soil Science and Plant Nutrition, 17*, 385–397. https://doi.org/10.4067/S0718-95162017005000028

Moghimi, A., Pourreza, A., Zuniga-ramirez, G., Williams, L. E., & Fidelibus, M. W. (2020). A novel machine learning approach to estimate grapevine leaf nitrogen concentration using aerial multispectral imagery. *Remote Sensing, 12*, 3515. https://doi.org/10.3390/rs12213515

Musingarabwi, D. M., Nieuwoudt, H. H., Young, P. R., Eyéghè-Bickong, H. A., & 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 Chemistry, 190*, 253–262. https://doi.org/10.1016/j.foodchem.2015.05.080

Pastenes, C., Villalobos, L., Ríos, N., Reyes, F., Turgeon, R., & Franck, N. (2014). Carbon partitioning to berries in water stressed grapevines: The role of active transport in leaves and fruits. *Environmental and Experimental Botany, 107*, 154–166. https://doi.org/10.1016/j.envexpbot.2014.06.009

Porep, J. U., Kammerer, D. R., & Carle, R. (2015). On-line application of near infrared (NIR) spectroscopy in food production. *Trends in Food Science and Technology, 46*, 211–230. https://doi.org/10.1016/j.tifs.2015.10.002

Rinnan, Å., van den Berg, F., & Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *TrAC, Trends in Analytical Chemistry, 28*, 1201–1222. https://doi.org/10.1016/j.trac.2009.07.007

Rodriguez-Lovelle, B., & Gaudillère, J. P. (2002). Carbon and nitrogen partitioning in either fruiting or non-fruiting grapevines: Effects of nitrogen limitation before and after veraison. *Australian Journal of Grape and Wine Research, 8*, 86–94. https://doi.org/10.1111/j.1755-0238.2002.tb00216.x

Romero, I., García-escudero, E., & Martín, I. (2010). Effects of leaf position on blade and petiole mineral nutrient concentration of Tempranillo grapevine (Vitis vinifera L.). *American Journal of Enology and Viticulture, 61*, 544–550. https://doi.org/10.5344/ajev.2010.09091

Rossouw, G. C., Orchard, B. A., Šuklje, K., Smith, J. P., Barril, C., Deloire, A., & Holzapfel, B. P. (2017). Vitis vinifera root and leaf metabolic composition during fruit maturation: Implications of defoliation. *Physiologia Plantarum, 161*, 434–450. https://doi.org/10.1111/ppl.12604

Rossouw, G. C., Šuklje, K., Smith, J. P., Barril, C., Deloire, A., & Holzapfel, B. P. (2018). Vitis vinifera berry metabolic composition during maturation: Implications of defoliation. *Physiologia Plantarum, 164*, 120–133. https://doi.org/10.1111/ppl.12715

Schmidtke, L. M., Smith, J. P., Müller, M. C., & Holzapfel, B. P. (2012). Rapid monitoring of grapevine reserves using ATR-FT-IR and chemometrics. *Analytica Chimica Acta, 732*, 16–25. https://doi.org/10.1016/j.aca.2011.10.055

Tardaguila, J., Fernández-Novales, J., Gutiérrez, S., & Diago, M. P. (2017). Non-destructive assessment of grapevine water status in the field using a portable NIR spectrophotometer. *Journal of the Science of Food and Agriculture, 97*, 3772–3780. https://doi.org/10.1002/jsfa.8241

Torchio, F., Río Segade, S., Giacosa, S., Gerbi, V., & Rolle, L. (2013). Effect of growing zone and vintage on the prediction of extractable flavanols in winegrape seeds by a FT-NIR method. *Journal of Agricultural and Food Chemistry, 61*, 9076–9088. https://doi.org/10.1021/jf401955m

Van Wyngaard, E., Blancquaert, E., Nieuwoudt, H., & Aleixandre-tudo, J. L. (2022). Infrared spectroscopy investigation of fresh grapevine (Vitis vinifera) shoots , leaves , and berries using novel chemometric applications for viticultural data. *Computers and Electronics in Agriculture, 203*, Article 107481. https://doi.org/10.1016/j.compag.2022.107481