Article

Prediction of Soluble Solids Content by Means of NIR Spectroscopy and Relation with *Botrytis cinerea* Tolerance in Strawberry Cultivars

Manuela Mancini [ID], Luca Mazzoni [ID], Rohullah Qaderi [ID], Elena Leoni, Virginia Tonanni [ID], Francesco Gagliardi, Franco Capocasa [ID], Giuseppe Toscano [ID] and Bruno Mezzetti *[ID]

Department of Agricultural, Food and Environmental Sciences, Università Politecnica delle Marche, Via Brecce Bianche 10, 60131 Ancona, Italy
* Correspondence: b.mezzetti@staff.univpm.it; Tel.: +39-071-2204640

Abstract: Strawberry fruits are particularly appreciated by consumers for their sweet taste related to their soluble solids content (SSC). However, strawberries are characterized by a short shelf-life and high susceptibility to tissue infection, mainly by *Botrytis cinerea.* The SSC determination of strawberry fruit through traditional destructive techniques has some limitations related to the applicability, timing, and number of samples. The aims of this study are (i) to verify if any relation between SSC and *B. cinerea* susceptibility in the fruits of five strawberry cultivars occurs and (ii) to determine the SSC of strawberry fruits through near infrared spectroscopy (NIR). Principal component analysis was used to search for spectral differences among the strawberry genotypes. The partial least squares regression technique was computed in order to predict the SSC of the fruits collected during two harvesting seasons. Moreover, variable selection methods were tested in order to improve the models and get better predictions. The results demonstrated that there was a high correlation between SSC and *B. cinerea* susceptibility (R^2 up to 0.87). The SSC was predicted with a standard error of 0.84 °Brix and R^2_p 0.75 (for the best model), which indicated the possibility to use the models for screening applications. NIR spectroscopy represents an important non-destructive alternative and finds remarkable applications in the agro-food market.

Keywords: post-harvest; grey mold infection; PLS regression; non-destructive technique; sugar analysis

1. Introduction

Strawberry demand is increasing in the fruit and vegetable market around the world, thanks to consumers who are looking for more high-quality fruits in terms of taste, appearance, and nutritional compounds. Strawberries are non-climacteric fruits, so they must be harvested at near-perfect ripeness to ensure the best quality in terms of flavor and firmness. At present, strawberries for sale are graded according to external qualities, such as uniform distribution, bright red color, shape, absence of surface defects, and size, but internal quality parameters, such as taste, sweetness, and sourness, are also fundamental for determining the quality of fruits, and they strongly influence the evaluation by end consumers [1]. However, due to their delicate structure and high metabolic rate during storage, strawberries are characterized by a very short shelf life and high susceptibility to tissue damage and infection, mainly by various plant pathogens [2]. The main pathogen affecting strawberries in post-harvest is *Botrytis cinerea*, which is the causative agent of gray mold, followed by *Penicillium* spp., *Colletotrichum* spp., *Mucor* spp., and *Rhizopus stolonifer* [3]. Post-harvest diseases caused by these pathogens are the result of latent infections initiated in the field during the growing season and infections caused by injuries during harvesting and handling [4]. Early detection of fruit affected by the fungus is of great importance both for producers and consumers. Fungal contamination poses a particular danger to packers during storage, transportation, and distribution. This is because even a very small number

of infected fruits can spread the infection to neighboring healthy strawberries to cause food loss and less safe products. Some fungal genera and species can produce mycotoxins that cause allergies and infections in susceptible individuals [5].

The traditional methods to determine the qualitative characteristics of fruits and fungal infection are mainly destructive. On the one hand, clear and reliable information is guaranteed, while on the other hand, there are various limitations regarding the amplitude of the products that can be analyzed, due to the time and cost required for analysis. In fact, destructive methods are not applicable to the entire production batch, but only to a certain representative number of fruits. Destructive methods for measuring these parameters are described in many studies. Sweetness is usually measured with a manual or automatic refractometer and is generally determined by the content of soluble solids (SSC), which is expressed in units of °Brix [6,7]. Mycotoxin contamination and fungal infection are commonly measured through several physicochemical and/or microbiological techniques (mass spectrometry, gas and liquid chromatography) [8]. In the last few years, researchers have been increasingly looking for non-destructive and fast analytical technologies, especially those based on optical properties [9,10]. This allows real-time evaluation and analysis of a wider range of samples and requires minimal sample preparation [11].

In recent years, various optical methodologies have been successfully applied to evaluate quality parameters, such as firmness, acidity, and soluble solids [12–15]. Among these methods, near infrared spectroscopy (NIR) is a valuable alternative. This technique uses the infrared region of the electromagnetic spectrum to non-destructively study the chemical and physical properties of the samples. Indeed, chemical bonds (NH-, CH-, and OH- groups) in molecules vibrate differently and at different energy levels depending on the molecular structure of the analyzed sample [16]. This technology has already been applied in various fields, such as industry [17], pharmaceuticals [18], energy [19,20], and the food industry [21,22]. Spectroscopic techniques produce datasets containing a high amount of data and information, which can be highly exploited using multivariate data analysis techniques or chemometrics. Furthermore, chemometrics can be used to handle effects common in spectroscopic techniques, such as peak shifts, interfering signals, and baseline shifts [23].

To the best of our knowledge, a number of studies have recently been carried out to evaluate the organoleptic and nutritional quality of strawberries using near infrared spectroscopy. The main disadvantage of these studies is that the number of selected fruits is usually limited, which can reduce the accuracy of the developed predictive models [24]. In addition, only a few cultivars are usually tested (only one in some cases) they mostly employ different types of sample treatments (e.g., different pesticide treatments, different culture systems, different post-harvest storage conditions), and several quality parameters are taken into account at the same time [25–27].

In this context, the first aim of this work was to characterize a key qualitative characteristic, such as soluble solids content, in five strawberry genotypes for two consecutive harvest seasons using traditional destructive methods and NIR spectroscopy in order to develop a regression model for the prediction of the considered parameter. In particular, partial least squares regression (PLS) models have been developed using data from the two harvesting seasons (2020 and 2021), and different variable selection techniques have been applied to the pretreated dataset in order to improve the prediction models. The introduction of non-destructive phenotyping methods makes it possible to monitor quality parameters, ensure fruit quality before the sale, and develop predictive models to speed up the selection process. A second aim of this study was to evaluate the relationship between strawberry sugar composition and susceptibility/tolerance to *Botrytis cinerea*. For this purpose, fruits of the same genotypes were also subjected to controlled in vitro infection, and their tolerance/susceptibility to this pathogen was monitored and measured for 5 consecutive days. Taken together, the two aims of this study will reveal if there is any relationship between the sugar content and the tolerance/susceptibility of strawberry fruits, and if it will be possible to build a prediction model for the SSC of strawberry fruits through

non-destructive NIRS technique. The final goal was the prediction of strawberry SSC and, consequently, its tolerance or susceptibility to *B. cinerea* infection, without destroying the fruits.

2. Materials and Methods

2.1. Plant Material

The qualitative parameters assessment was performed on five strawberry cultivars: Cristina, Romina, Sibilla, Silvia (in the 2020 harvest season), and Sveva (only in the 2021 harvest season). The analyzed strawberries were cultivated at the Didactic and Experimental Farm Center for Agricultural Research "P. Rosati", Agugliano (AN, Italy, 43_320N–13_220E) during the 2019–2020 season (4 cultivars) and during the 2020–2021 season (5 cultivars). The cultivation method used cold-stored "A" type plants planted in an open field at the end of July, with 8 plots of each genotype, for a total of 32 plants of each genotype. This experimental design allows us to standardize the natural variation in fruit parameters depending on the location of the plot in the field. The precocity index (PI), defined as the average of the weighted days needed to collect the whole production of a cultivar from January 1 for each genotype in 2020 and 2021, is reported in Table 1. From each plot, 15 fruits were selected, for a total of 60 fruits of each genotype, which were sent to the laboratory for analysis. The fruits were collected on two harvest dates, for a total of 120 fruits per genotype. The fruits were carefully processed, individually numbered, and divided into two groups: Control (not infected with *B. cinerea* spores) and Treated (infected with *B. cinerea* spores). Then, fruits were subjected to an external check of the *B. cinerea* infection level through the McKinney index for five consecutive days (starting from harvest day, which corresponded to day 0 and coincided with Mondays) and analyzed one by one by means of NIR spectroscopy (the non-destructive method). Then, the fruits were frozen separately and analyzed one by one after a few days for soluble solids content, given that the extraction of more juice from thawed fruits is easier.

Table 1. Precocity index (PI) of the strawberry cultivars for the years 2020 and 2021.

CULTIVAR	2020	2021
Cristina	155	157
Romina	147	151
Sibilla	140	146
Silvia	146	151
Sveva	-	154

2.2. Soluble Solids Content Analysis

The soluble solids content (SSC) of each fruit was measured using an automatic temperature-compensated digital refractometer (Palette PR101, Atago, Tokyo, Japan). For this analysis, frozen fruits, which were previously analyzed by NIR, were thawed one by one and pressed to extract the juice. A few drops of this juice were placed on the surface of a refractometer slide, and the °Brix value for each fruit was measured twice [12].

2.3. Botrytis Cinerea Infection

The *B. cinerea* level of infection of the fruit of the different cultivars was assessed by using the method described by Siedliska et al. [28]. Before inoculation, *B. cinerea* was cultured on agar for two weeks at room temperature. The fungal conidia were scraped off the surface of the agar with a sterile glass rod in a small volume of sterile water and diluted to a concentration of approximately 10^6 mL^{-1} spores.

During inoculation, strawberry fruits of tested cultivars were immersed in *B. cinerea* spore solutions (concentration 10^6 spores mL^{-1}) or sterile distilled water, for the Control, for 30 s and dried at room temperature in an air chamber. Throughout the experiment

(5 days), from day 0 (D0) to day 4 (D4), the Treated and Control fruits were stored in a dark room at a relative humidity of 90% and a temperature of 20 ± 1 °C [29,30].

The severity of damage was recorded on a 6-point empirical scale according to the surface of fruit that was infected: 0, healthy fruit; 1, 1% to 20%; 2, 21% to 40%; 3, 41% to 60%; 4, 61% to 80%; 5, more than 81% of the fruit surface was infected and sporulated. The decay index, or McKinney index (MKI) was calculated from the following equation [31]

$$I = \left[\sum \left(\frac{(d \times f)}{(N \times D)} \right) \right] * 100$$

After daily evaluation, fruits were stored at -80 °C for analysis of the SSC.

2.4. NIR Analysis

In this study, the analysis was performed with a Fourier transform (FT) NIR spectrophotometer (FT-NIR mod. Nicolet iS10, Thermo, Waltham, MA, USA), equipped with an integrative sphere. The spectra were acquired in the near infrared range, namely from 10,000 to 4000 cm^{-1}. The analysis was performed using a continuous flow of nitrogen, which was aimed at minimizing the moisture content in the instrument during the assessment and reducing the variability of the spectral analysis. Each spectrum was recorded at 8 cm^{-1} resolution, and averaged 32 scans, which resulted in 1557 absorbance data. The background spectrum was scanned hourly to reflect the entire electromagnetic signal and reduce variations due to the environmental conditions rather than the sample.

For the 2020 season, each fruit was analyzed twice. The first measurement was performed on a point along the fruit equator. The fruit was then turned 180° for the successive measurement. The resulting dataset (DB-2020) consists of 2784 observations x 1557 wavenumbers (1392 samples). For the 2021 season, an additional NIR measurement was added, so each fruit was analyzed in triplicate to better describe the fruit variability. The resulting dataset (DB-2021) consists of 3096 observations x 1557 wavenumbers (1032 samples).

2.5. Multivariate Data Analysis

First, one-way ANOVA was carried out to determine statistically significant differences between the cultivars and the SSC content, as well as between cultivars and *Botrytis cinerea* infection at different storage days. To investigate which specific groups differed from each other, the Tukey test was performed. This post-hoc test is based on the studentized range distribution and compares all possible pairs of means to figure out which groups differ from each other.

Chemometrics was used to extract useful information for the spectral data. Before model computation, in order to minimize the effects of baseline shifts, interfering signals, and peak shifts, the spectral data were pretreated. Different pretreatments, including standard normal variate (SNV), multiplicative scatter correction (MSC), first and second derivative spectra (from Savitzky–Golay [32] with 13 or 21 smoothing points and a second order polynomial), and a combination of the previous ones were applied in order to remove undesired physical phenomena [23]. As last, each sample was averaged across the replicates.

Principal component analysis (PCA) was used as an explorative method to show the statistical variance of all the samples on a chemical basis and cluster them according to spectral similarity. PCA computation was performed on the pretreated dataset and on mean-centered data. A confidence ellipse was calculated for each cultivar group using the mean score values as the center and the standard deviation of each variability direction as the radius of the ellipse. The loading plot of the two first PCs was investigated to identify the compounds associated with the separation of samples in the PCA space.

Partial least squares regression (PLS) was used to predict the SSC content (°Brix) of both DB-2020 and DB-2021 datasets. The technique works by evaluating the relationship between non-spectral dependent values (Y) and independent spectral values (X) to then

maximize the covariance of their scores [33]. PLS models were validated using venetian blind-cross validation (five segments) and an external test set. The dataset was split into training and test sets using a duplex algorithm [34].

To evaluate the performance of the developed model and its ability to estimate unknown samples, different parameters were taken into account. In detail, the parameters were R^2_{cv}, or the coefficient of determination of cross-validation, R^2_p, or the coefficient of determination of prediction, the root mean square error of cross validation (RMSECV) and the root mean square error of prediction (RMSEP). The ratio of performance to deviation (RPD) and the range error ratio (RER) were also computed. An RPD with values of 2–3 indicated a discrete model, 3–4 indicated a good model for screening applications, and values > 4 indicated an excellent model [35]. By contrast, RER values < 6, between 7 and 20, and >20 indicated a poor, suitable, and excellent model, respectively [36]. For each model, the number of latent variables (LVs) was chosen considering the RMSECV plot against the number of LVs searching for a local minimum. In addition, the U vs. T (scores of Y and scores of X) plot and the loading plot were also analyzed.

It is well known that the performance of the prediction models can be enhanced by using variable selection methods in comparison to that of the full-spectrum PLS model [37]. Therefore, after the selection of the optimal number of LVs, the PLS models were calculated again on the wavenumbers, which were selected on the basis of the following different variable selection methods: recursive weighted partial least squares (rPLS) [38] and interval partial least squares (iPLS) [39]. Both backward and forward iPLS were performed with a number of intervals varying from 50 to 100.

All the data analyses were performed using Matlab (ver. R2022a, The MathWorks, Natick, MA, USA) and in-house functions based on existing algorithms.

3. Results

3.1. Analysis of Variance on Soluble Solid Content

The results of the one-way ANOVA are represented with a box and whisker plot (Figure 1). The boxes are used to compare groups; if the boxes of two cultivars do not overlap, this is strong evidence that the means differ. Cultivars with the same letters are considered not significantly different ($p \leq 0.05$). The horizontal black line is the mean value, the box represents a group of numerical data through their quartiles, and the whiskers represent the max and min values for each cultivar.

Figure 1a reports the results of harvest season 2020 and clearly shows that Silvia was the cultivar with the lowest value of soluble solid content (b) with respect to the other cultivars (a). For season 2021 (Figure 1b), the results clearly show that Sveva and Silvia had the lowest values of soluble solid content (b), Cristina and Romina had similar characteristics (ab), and Sibilla had the highest SSC content value (a). The differences in fruit SSC were not surprising and were, in fact, expected. In fact, it is well known that the genotype is one of the main factors affecting the SSC of strawberry fruits. Furthermore, the harvest season (harvest year but also different dates across the same year) is also a key factor influencing the SSC content, as well as the environmental conditions (light exposure, temperature, radiation, and water management) [1]. Other key factors, such as the soil conditions, the harvest method, and the cultivation conditions, were not considered responsible for these differences, since they were the same for all the cultivars.
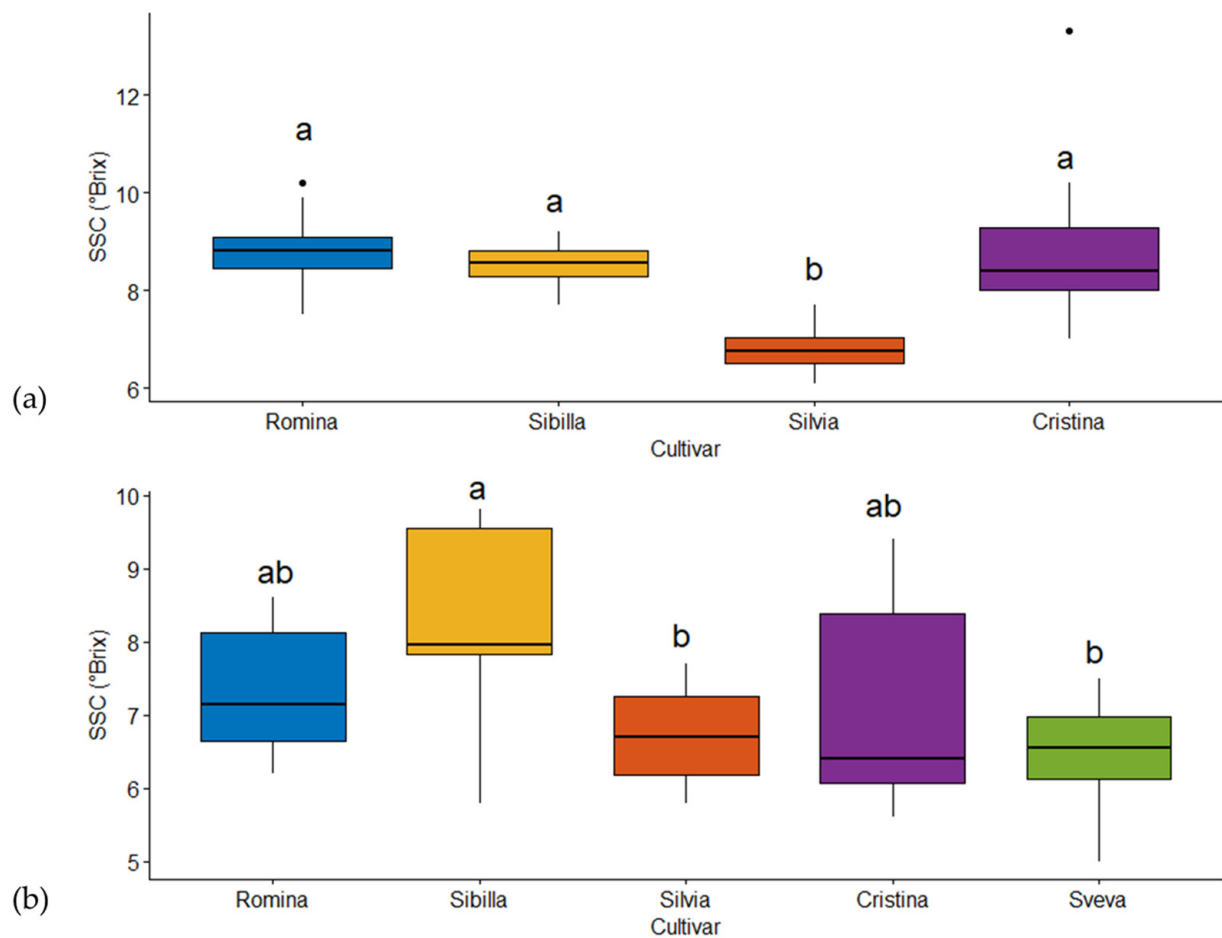
Figure 1. Results of one-way ANOVA for harvest seasons 2020 (**a**) and 2021 (**b**). The mean value of each cultivar is represented with the horizontal black line, the boxes cover the interquartile interval where 50% of the data is found, and they indicate the upper and lower quartiles. The whiskers indicate the maximum and minimum values of SSC content (°Brix) within each cultivar. Different lowercase letters indicate a significant difference ($p \leq 0.05$, Tukey test) between the strawberry cultivars.

3.2. Analysis of Variance on Botrytis Cinerea Susceptibility/Tolerance Trial

Data on *B. cinerea* infection tolerance/susceptibility of the 2020 harvest year revealed that Cristina was the most sensitive to *B. cinerea* in control conditions (without manual infection) with a more than 70% MKI result. At the same conditions, Romina (about 30% MKI at day 4) and Silvia (about 35% MKI at day 4) were the most tolerant to this pathogen (Figure 2). Contrarily, Romina was found to be the most sensitive cultivar when the fruits were inoculated with *B. cinerea* after harvest (treated trial), with an almost 100% MKI result at day 4, followed by Cristina (about 90% MKI), Sibilla (about 85% MKI), and, finally, Silvia (about 70% MKI), which was found to be the most tolerant cultivar in both control and treated conditions (Figure 2).

In the 2021 harvest season, Sveva and Cristina were the most sensitive cultivars to *B. cinerea* in control conditions, and each reached almost 100% MKIs, even when they were not infected manually with *B. cinerea* spores. In addition to the high susceptibility of these two cultivars, this high data could be also explained by the bad weather conditions (heavy rains before fruit collection) that favored infection by gray mold and other fungi and bacteria, which spoil the fruit even without artificial contamination [40]. As expected, Sveva and Cristina were also found to be the most susceptible cultivars in the treated trial, together with Silvia (100% MKI) and Sibilla (around 90% MKI). The most tolerant cultivar in the treated trial was Romina with about 70% MKI (Figure 3).
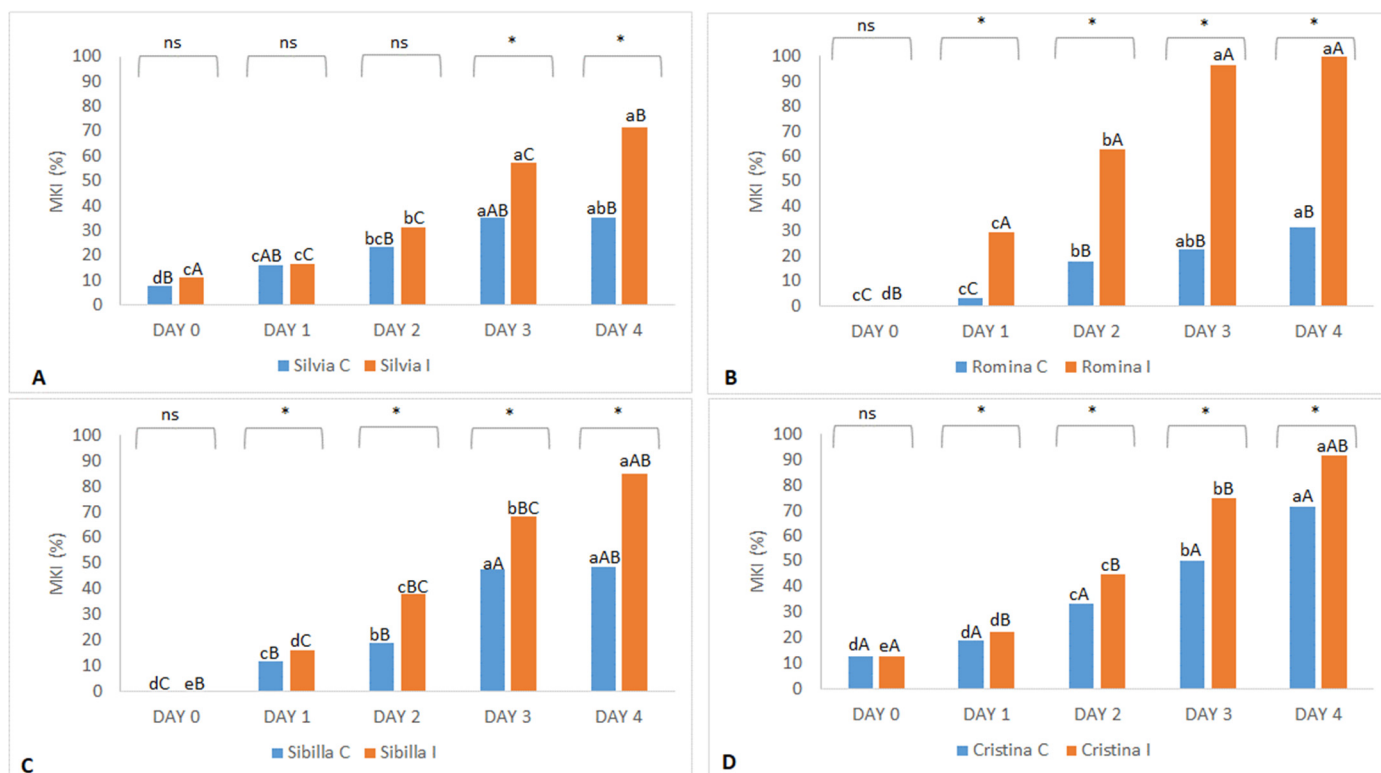
Figure 2. McKinney index (MKI) of postharvest decay of the (**A**) Cristina, (**B**) Romina, (**C**) Sibilla, and (**D**) Cristina strawberry fruits in the 2020 harvest season. Blue bars represent the control samples (not infected), and orange bars represent the infected samples (with *B. cinerea* spores). Strawberries were stored for 4 days at $20 \pm 1\ °C$, and 90% relative humidity. Different lowercase letters indicate a significant difference ($p \leq 0.05$, Tukey test) between fruits of the same treatment of the same cultivar at different days. Asterisks indicate a significant difference ($p \leq 0.05$, Tukey test) between control and infected fruits of the same cultivar at the same days (ns: not significant). Different uppercase letters indicate a significant difference ($p \leq 0.05$, Tukey test) between fruits of the same treatment of different cultivars at the same days.

In summarizing both harvest years of study, Silvia, Romina, and Sibilla confirmed a good tolerance to *B. cinerea* in control conditions, while Cristina and Sveva were the most susceptible. When manually inoculated (treated trial), all genotypes developed higher levels of fruit infection than the control (except Sveva and Cristina in 2021, which showed similar results). Sibilla was the cultivar that showed better average values of tolerance to *B. cinerea* in both years, while Cristina and Sveva were the most susceptible. Silvia and Romina showed contrasting results, being among the most (Silvia) and least (Romina) tolerant in 2020 and presenting the opposite behavior in 2021. Thus, these results evidenced the importance of also considering the environmental interaction during the ripening period in determining the fruit response to the disease.

When looking at the SSC and *B. cinerea* susceptibility/tolerance data, it is possible to identify a connection between the content of fruit sugars content and the *B. cinerea* infection, depending on the cultivar. The correlation between MKI at day 4 for strawberries and SSC was highly positive ($R^2 = 0.87$ for control and 0.61 for treated). These data refer to the 2020 harvest season, which performed better than the 2021 season for the climatic conditions in pre-harvest. These results were in line with our findings, with the fruits of the Silvia cultivar being less sugary (Figure 1) and having the lowest susceptibility to *B. cinerea* (Figure 2), especially in the treated trial. On the contrary, Romina and Cristina, which presented higher sugar content in their fruits, had the most susceptibility to *B. cinerea* in treated and control

conditions, respectively. Therefore, from these results, it can be highlighted how the higher sugar content facilitates a greater development of *B. cinerea* infection [41].
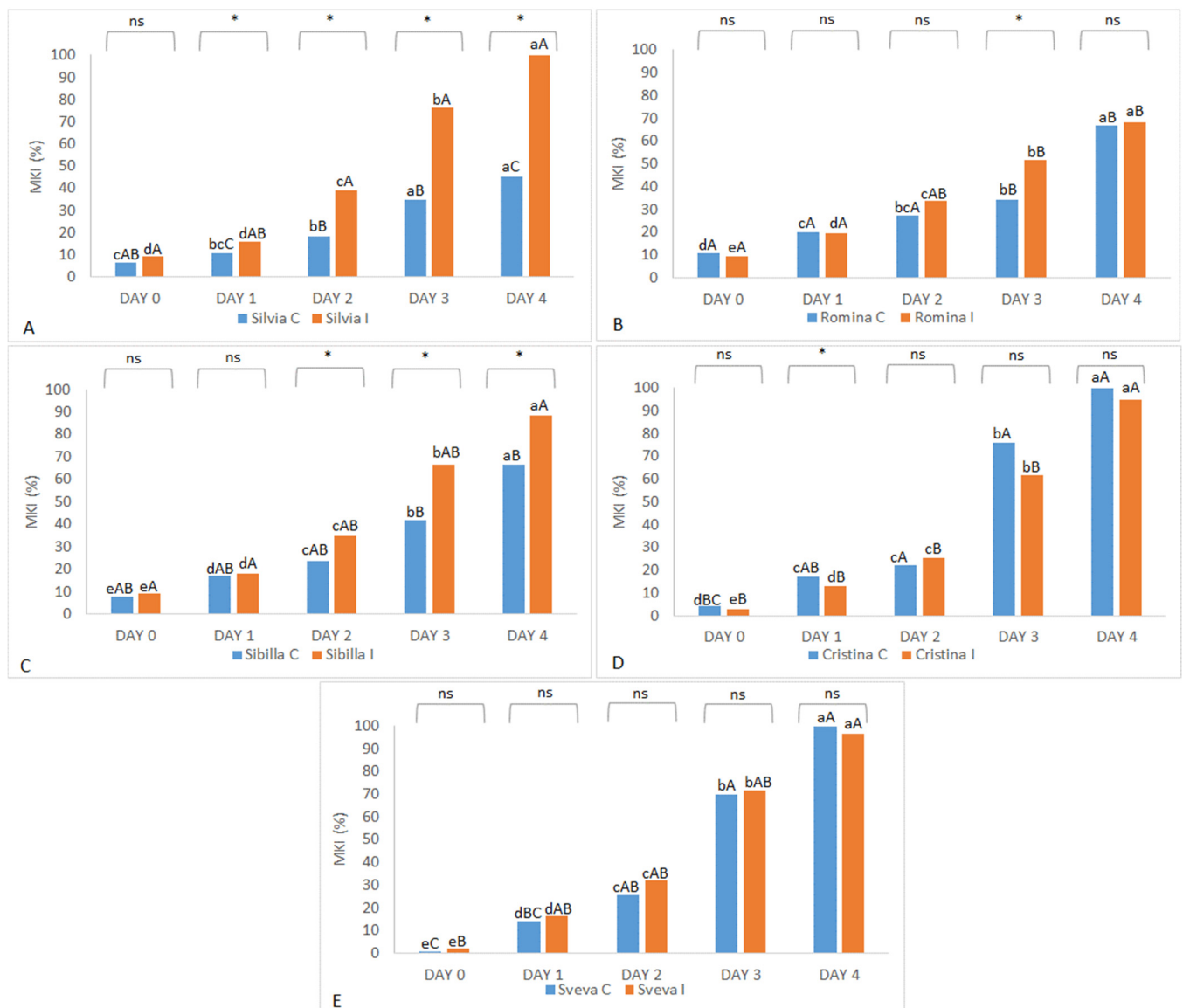

Figure 3. McKinneyindex (MKI) of postharvest decay of the (**A**) Cristina, (**B**) Romina, (**C**) Sibilla, (**D**) Cristina, and (**E**) Sveva strawberry fruits in the 2021 harvest season. Blue bars represent the control samples (not infected), and orange bars represent the infected samples (with *B. cinerea* spores). Strawberries were stored for 4 days at $20 \pm 1\ °C$, and 90% relative humidity. Different lowercase letters indicate a significant difference ($p \leq 0.05$, Tukey test) between fruits of the same treatment of the same cultivar at different days. Asterisks indicate a significant difference ($p \leq 0.05$, Tukey test) between control and infected fruits of the same cultivar at the same days (ns: not significant). Different uppercase letters indicate a significant difference ($p \leq 0.05$, Tukey test) between fruits of the same treatment of different cultivars at the same days.

3.3. Spectra

Figure 4 reports the average spectra of the five cultivars considered in this study for both DB-2020 and DB-2021. Only the fruits analyzed at D0 (and without any infection of *B. cinerea*) were considered and were pretreated using a first derivative (Savitzky–Golay with 21 smoothing points and a second order polynomial)—21der1—and SNV. It should be noted that the interpretation of first derivative spectra is more complex than the raw spectra,

because a peak of maximum absorbance on the original spectra corresponds to zero in the 1st derivative [42]. For this reason, zero points corresponding to peak in the raw spectra were selected. It is important to take into account that differences between the location of bands in the raw spectrum and in the derivative of the spectrum can occur, by even more than 20 cm^{-1} (5 nm) [43]. This depends not only on the data processing algorithms used, but also on the shape of the spectrum. In fact, the particle size influences the radiation scattering properties of the samples, and spectral changes, such as baseline offset or slope and band shifts to higher or lower wavenumber regions, can be observed [43].
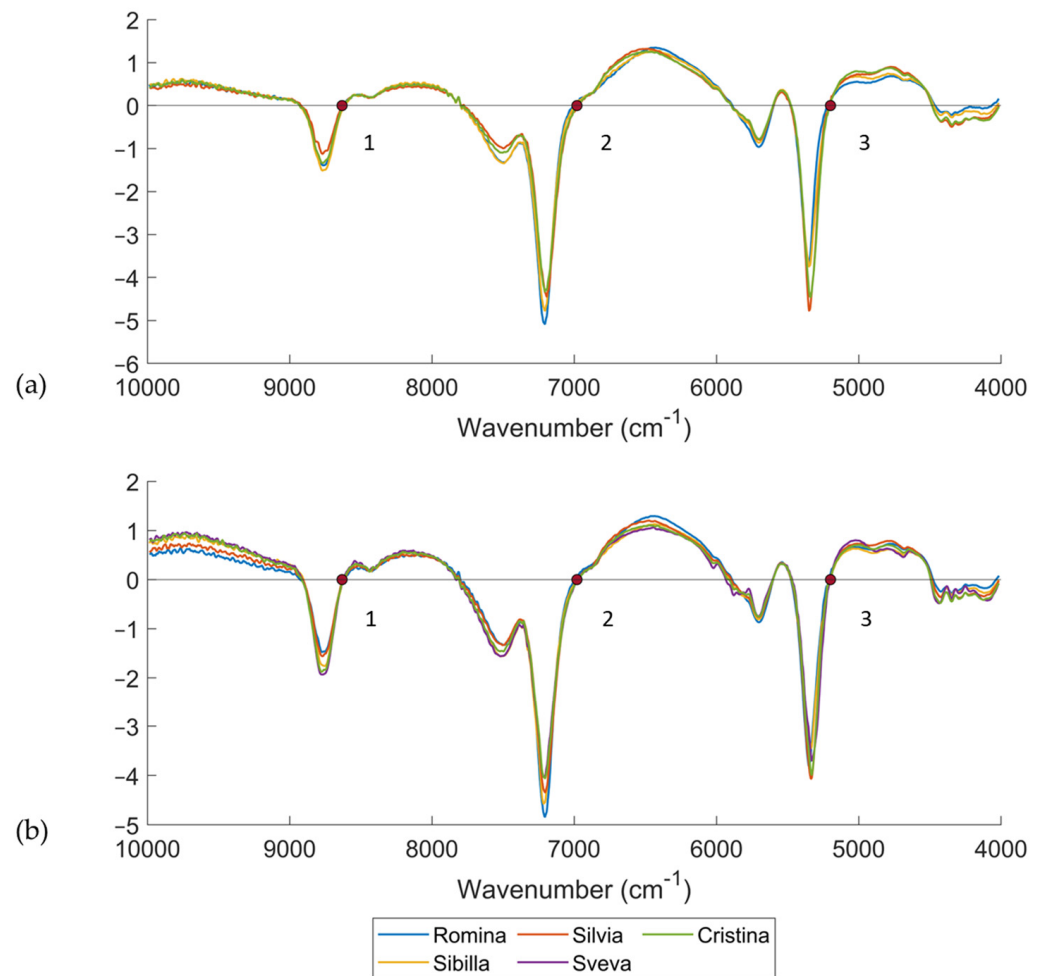

Figure 4. Average spectra of DB-2020 (**a**) and DB-2021 (**b**) cultivars pretreated using first derivative (Savitzky–Golay filter, 21 smoothing points, 2nd polynomial order) and SNV and considering only the first day of analysis i.e., the non-infected material/samples. The most important wavenumbers are marked with red dots.

Three significant bands dominated the spectra. The band at 5199 cm^{-1} (3) was related to the OH bonds and was much higher for the Cristina and Silvia cultivars than for the others [36,44]. The band at 7004 cm^{-1} (2) was assigned to O-H and C-H combinations, and it involved qualitative parameters, such as the soluble solids content or the titratable acidity [44]. As seen from the plot of spectra, it can be observed that the highest band among the cultivars was reached by Sibilla and Romina, which were the cultivars with the highest soluble solids contents. Lastly, the band at 8632 cm^{-1} (1) was in the region of the second overtone of CH. These three major bands were seen in all strawberry cultivars, though the spectral position differed due to the typical characteristics of each cultivar. In fact, wavelength drift or shift in informative peaks may be due to differences in chemical composition [45].

It is important to take into consideration that water constitutes about 80–90% of fruit and vegetables; consequently, the water absorption bands dominate the NIR spectrum [46], with the second major component being carbohydrates. In addition, the effective NIR absorption bands (of large molecules such as fruits) are relatively broad and overlapping, and their assignment results are complex. This is due to the different chemical environments of each O–H and C–H bond in water and sugar molecules [47].

3.4. Principal Component Analysis (PCA)

PCA was used to determine the differences between the cultivars considered in this study and, in general, for looking at groupings among the analyzed samples. The score plot showed the differences/groupings among the samples, while the loading plot was investigated to identify the compounds associated with the differences among the cultivars.

Figure 5a reports the score plot of the PCA computed on the DB-2020 considering only the D0, i.e., the non-infected material/samples. A total of 432 samples belonging to 4 cultivars were used for PCA computation. The spectra were pretreated using a first derivative (Savitzky–Golay with 21 smoothing points and second order)—21der1—and SNV. All the NIR measurements were averaged across the replicates. The first principal component (PC1) separated Romina and Sibilla, which were the cultivars having fruits with the highest SSC content, from Silvia and Cristina. These findings were confirmed by the loading plot (Figure 5b). To make an easier interpretation of the loadings, a comparison with the pretreated average spectra of the different cultivars (Figure 4a) was also carried out. The most relevant wavenumbers were 8531, 6862, and 5172 cm^{-1} for the first loading, and 7247 and 5172 cm^{-1} for the second loading. In addition to bands at 7247 cm^{-1}, all of the most relevant NIR regions were confirmed by the spectral differences previously reported in Figure 4a. The band at 6862 cm^{-1} was related to first overtone of water [48] and the band at 5172 cm^{-1} was related to an OH combination of water [46,49], while the bands at 8531 and 7247 cm^{-1} were related to the second overtone of CH [49] and first overtone of CH, respectively.

Figure 5c reports the score plot of the PCA computed on the DB-2021 considering only the first day of analysis, i.e., the non-infected material/samples. A total of 240 fruit samples belonging to 5 cultivars were used for PCA computation. As was done with DB-2020, the spectra were pretreated using a first derivative (Savitzky–Golay with 21 smoothing points and second order) and SNV, and all the NIR measurements were averaged across replicates. The second principal component (PC2) separated Romina and Sibilla, which were the cultivars with the highest SSC content, from the others. Cristina and Sveva were located in the bottom right part of the score plot, while Silvia was the most different cultivar from the others. PCA nicely confirmed the ANOVA results.

The loading plot was used for investigating the most influencing wavenumbers responsible for the separation of the samples in the PCA space (Figure 5d). The most relevant wavenumbers were 8940, 6962, and 5296 cm^{-1} for the first loading and 6862 and 5184 cm^{-1} for the second loading. In addition to bands at 5296 cm^{-1}, all of the most relevant NIR regions were confirmed by the spectral differences previously reported in Figure 4b. As last, it is interesting to note that the same wavenumbers for the PCA for DB-2020 were detected for DB-2021, which indicated the same chemical information. The bands at 5296 and 5184 cm^{-1} were related to the OH combination bonds, while bands at 6962 and 6862 cm^{-1} were assigned to the first overtone of CH and OH combinations [48,49] and first overtone of OH bonds [46,48], respectively. The last band at 8940 cm^{-1} was assigned to third overtone of CH stretching [49].

To conclude, it can be stated that the difference between the two harvesting seasons was mainly related to the number of cultivars and samples analyzed. The different location of the cultivars in the PCA score plot was only related to the number of samples and cultivars considered in the PCA computation. In DB-2020, PC1 divided Sibilla and Romina from the other cultivars, while, in DB-2021, it was PC2 dividing Sibilla and Romina from the other cultivars. The two aforementioned loadings detected the same wavenumbers,

which indicated the same chemical information related to the SSC between the two datasets and, consequently, the two harvesting seasons.
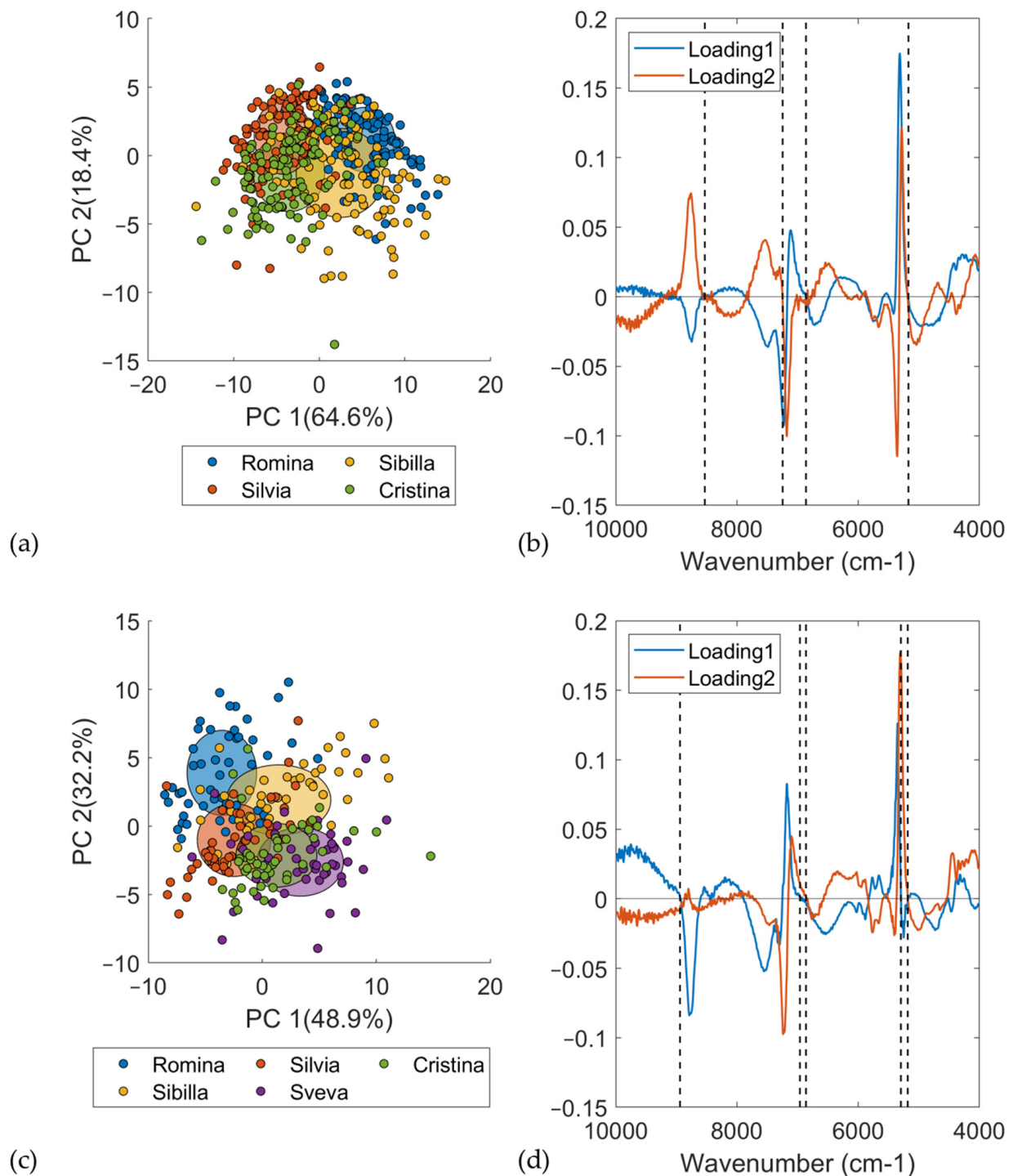

Figure 5. PCA score plots of the two first principal components (**a**) and the respective loadings based on the pretreated spectra of the DB-2020 samples (**b**). PCA score plots of the two first principal components (**c**) and the respective loadings based on the pretreated spectra of the DB-2021 samples (**d**). The most important wavenumbers are marked with vertical dotted lines.

3.5. PLS on DB-2020

PLS regression models were developed on DB-2020 using both the whole dataset (WDB-2020) and a reduced dataset (RDB-2020), i.e., considering only the strawberry fruits

analyzed from D0 to D2 after *B. cinerea* infection and excluding the fruits with a high score of damage.

For WDB-2020, 300 samples were used as a training set and 130 samples as a test set; while for RDB-2020, 240 samples were used as a training set and 100 samples as a test set. The best prediction models were obtained from RDB-2020. In the interest of providing complete information, the best results for WDB-2020 were obtained using a first derivative (Savitzky–Golay with 21 smoothing points and second order)—21der1—or 21der1 and SNV as pretreatments. The models used seven LVs and returned an $R^2_p = 0.65$, and a RMSEP of 0.99 and 1.03 °Brix, respectively. Table 2 reports the results of the best PLS models developed from the RDB-2020 dataset. Second derivative and MSC/SNV pretreatments were also tested without obtaining acceptable results. The best model for the prediction of the SSC parameter was obtained using a first derivative (Savitzky–Golay with 21 smoothing points and second order) as a pretreatment. The model used seven LVs and had an $R^2_p = 0.69$, a RMSEP = 0.91 °Brix, a RPD = 1.80, and a RER = 7.58. The values indicate the possibility to use the model for screening applications only. Considering that the current dataset, consisting of strawberry fruits from different cultivars, is relatively complex, seven latent variables are an acceptable number.

Table 2. Summary of partial least squares regression (PLS) results and summary of PLS results applying variable selection techniques to RDB-2020 for SSC (°Brix) prediction models. The best PLS models are marked in bold (xder1 = first derivative with x number of smoothing points; SNV = standard normal variate; LVs = latent variables; RMSECV = root mean square error of cross validation; RMSEP = root mean square error of prediction; RER = range error ratio; RPD = ratio of performance to deviation).

Model	Pretreatment	LVs	RMSECV	R^2_{cv}	RMSEP	R^2_p	RER	RPD
PLS	13der1 + SNV	5	1.13	0.55	1.02	0.50	6.76	1.61
PLS	13der1	7	1.02	0.59	0.92	0.69	7.50	1.78
PLS	21der1 + SNV	7	0.84	0.75	0.93	0.63	7.42	1.76
PLS	**21der1**	**7**	**0.86**	**0.71**	**0.91**	**0.69**	**7.58**	**1.80**
PLS	Cut_13der1	7	0.94	0.65	0.88	0.73	7.84	1.86
PLS	Cut_21der1	7	0.84	0.72	0.88	0.72	7.84	1.86
rPLS	Cut 21der1	6	0.78	0.76	0.91	0.7	7.58	1.80
iPLS (for 50)	Cut 21der1	6	0.89	0.69	1.05	0.59	6.57	1.56
iPLS (for 100)	Cut 21der1	5	0.99	0.61	1.01	0.62	6.83	1.62
rPLS	**Cut 13der1**	**5**	**0.84**	**0.75**	**0.84**	**0.75**	**8.21**	**1.95**
iPLS (for 50)	Cut 13der1	9	0.84	0.72	0.91	0.7	7.58	1.80
iPLS (for 100)	Cut 13der1	7	0.84	0.72	0.88	0.71	7.84	1.86

In order to improve the model performance and get a better prediction, PLS models were computed again by only considering the wavenumber range between 9000 cm^{-1} up to 4000 cm^{-1}, since it contains the most significant information. The results of the best performing PLS models are reported in Table 2. First derivative (Savitzky–Golay with 13 or 21 smoothing points and second order) was used as a pretreatment and returned models with seven LVs, an $R^2_p = 0.73 - 0.72$, a RMSEP = 0.88 °Brix, a RPD = 1.86, and a RER = 7.84.

Considering the slight improvement of the prediction results, the iPLS and rPLS have were as variable selection techniques on the reduced wavenumber range (9000 cm^{-1} up to 4000 cm^{-1}). The iPLS was tested using both forward and backward methods and at different intervals (50 and 100). Table 2 reports the results of the best prediction models. It should be noted that the iPLS using the backward method did not return any good results. When comparing the different figures of merit, the rPLS was the superior variable selection method performed by estimating the first derivative with Savitzky–Golay, 13 or 21 smoothing points, and a second order polynomial fit. There was a significant improvement using 13 smoothing points. In the model obtained using the rPLS method (five LVs), the RMSEP

value decreased from 0.88 °Brix to 0.84 °Brix, the R^2_p increased from 0.73 to 0.75, the RPD increased from 1.86 to 1.95, and the RER increased from 7.84 to 8.21.

3.6. PLS on DB-2021

As with the DB-2020 dataset, PLS regression models were developed for DB-2021 using both the whole dataset (WDB-2021) and a reduced dataset (RDB-2021), i.e., considering only the strawberry analyzed from D0 to D2 after *B. cinerea* infection and excluding the fruits with a high score of damage.

For the WDB-2020, 400 samples were used as a training set and 162 samples as a test set, while for the RDB-2020, 312 samples were used as a training set and 133 samples as a test set. In this case, the best prediction models were obtained from WDB-2021. Before PLS model computation, one outlier was removed from the training set. The most promising results from the RDB-2021 were obtained using a first derivative (Savitzky–Golay with 21 smoothing points and second order)—21der1—or 21der1 and SNV as pretreatments. The former model (21der1) used nine LVs and returned an R^2_p = 0.65 and a RMSEP of 0.75 °Brix. Table 3 reports the results of the best models developed from the WDB-2021 dataset. Second derivative and MSC/SNV pretreatments were also tested, but without obtaining acceptable results. The best model for the prediction of the SSC parameter was obtained using first derivative (Savitzky–Golay with 21 smoothing points and 2nd order) and SNV as pretreatment. The model used nine LVs and had an R^2_p = 0.70, a RMSEP = 0.87 °Brix, a RPD = 1.93, and a RER = 8.51. The values indicate the possibility to use the model for screening applications only.

Table 3. Summary of partial least squares regression (PLS) results and summary of PLS results applying variables selection techniques to WDB-2020 for SSC (°Brix) prediction models. The best PLS models are marked in bold (xder1 = first derivative with x number of smoothing points; SNV = standard normal variate; LVs = latent variables; RMSECV = root mean square error of cross validation; RMSEP = root mean square error of prediction; RER = range error ratio; RPD = ratio of performance to deviation).

Model	Pretreatment	LVs	RMSECV	R^2_{cv}	RMSEP	R^2_p	RER	RPD
PLS	13der1 + SNV	7	0.94	0.61	1.03	0.49	7.18	1.63
PLS	13der1	7	0.84	0.64	0.92	0.71	8.04	1.83
PLS	**21der1 + SNV**	**9**	**0.84**	**0.67**	**0.87**	**0.70**	**8.51**	**1.93**
PLS	21der1	7	0.88	0.64	0.93	0.65	7.96	1.81
PLS	Cut_13der1	7	0.86	0.63	0.92	0.71	8.04	1.83
PLS	Cut_21der1	7	0.87	0.65	0.93	0.64	7.96	1.81
PLS	Cut_21der1 + SNV	7	0.91	0.6	0.97	0.63	7.63	1.73
rPLS (8 LVs)	**21der1 + SNV**	**6**	**0.82**	**0.68**	**0.91**	**0.67**	**8.13**	**1.85**
iPLS (for 50)	21der1 + SNV	8	0.9	0.62	0.95	0.65	7.79	1.77
iPLS (for 100)	21der1 + SNV	6	0.93	0.59	0.98	0.61	7.55	1.71
rPLS (7 LVs)	**13der1**	**6**	**0.81**	**0.66**	**0.91**	**0.71**	**8.13**	**1.85**
iPLS (for 50)	13der1	7	0.78	0.69	1	0.97	0.70	1.68
iPLS (for 100)	13der1	7	0.77	0.7	0.99	0.68	1.70	1.70

As for DB-2020, the PLS models were computed again to only consider the wavenumber range between 9000 cm^{-1} to 4000 cm^{-1}. The results of the best performing PLS models are reported in Table 3. First derivative (Savitzky–Golay with 13 smoothing points and second order) was used as a pretreatment and returned a model with seven LVs, an R^2_p = 0.71, a RMSEP = 0.92 °Brix, a RPD = 1.83, and a RER = 8.04.

As well as for DB-2020, variable selection techniques on the best models, reported in Table 3, were also used. iPLS was tested using both forward and backward methods and at different intervals (50 and 100), but the iPLS using the backward method did not return any good results. Table 3 reports the results of the best prediction models using variable selection techniques. When comparing the different figures of merit, the rPLS

was the superior variable selection method when using a first derivative (Savitzky–Golay with 13 smoothing points and second order). In any case, the model did not show any significant improvement. Both the RMSEP value (from 0.92 °Brix to 0.93 °Brix) and the R^2_p (0.71) remained almost constant. In contrast, the RPD decreased from 1.83 to 1.81, and the RER decreased from 8.04 to 7.96. This demonstrates that some of the deleted variables were somehow important for the prediction of the SSC.

3.7. Model Interpretability

Variable selection has different benefits, including improving the performance of models in terms of prediction, reducing computation time, and facilitating data visualization/interpretability of the models. This last aspect is particularly important for an overall better understanding of the process/results. For this reason, Figure 6 reports the average spectrum with variables selected by the rPLS for both the DB-2020 and DB-2021 datasets. It is interesting to note that both models selected almost the same spectral areas. In detail, the first (8950–8500 cm^{-1}) and second (7320–7100 cm^{-1}) areas matched and were assigned to the second overtone of CH and the first overtone of CH combination bonds, respectively. The rPLS model for DB-2020 also selected variables in the region between 6800 and 6290 cm^{-1}, which were assigned to the first overtone of OH bonds, while the rPLS model for DB-2021 selected only some wavenumbers in that region. On the other hand, the rPLS for DB-2021 selected variables in a broader third area (5600–5270 cm^{-1}), which was assigned to OH combination bands, while the rPLS for DB-2020 had a more narrow range (5120–5460 cm^{-1}).
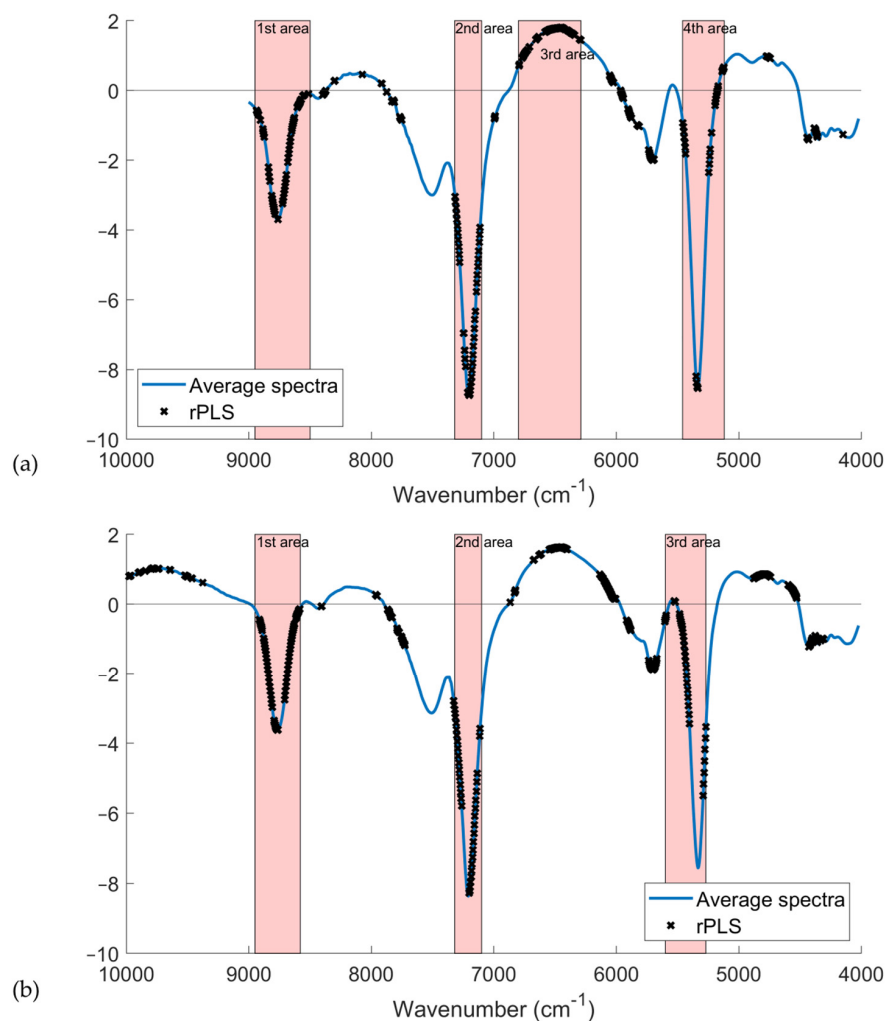

Figure 6. (**a**) Average spectrum of DB-2020 dataset pretreated using first derivative (Savitzky–Golay filter, 13 smoothing points, second polynomial order) on the spectral range from 9000 to 4000 cm^{-1};

(**b**) average spectrum of DB-2021 dataset pretreated using first derivative (Savitzky–Golay filter, 13 smoothing points, second polynomial order) on all the spectral range. The variables selected by rPLS are marked with black crosses. The red areas mark the most important wavenumber range for the prediction of SSC content (°Brix).

The fact that the models did not differ significantly is an indication that the system is robust, and that the information on the SSC content wass spread throughout three/four regions of the spectra. This is in line with the findings reported in Sections 3.3 and 3.4.

3.8. Global PLS

The DB-2020 and DB-2021 datasets were merged to investigate if it was possible to develop a global model able to predict the SSC of strawberry fruits from different seasons. First of all, DB-2020 was used as a training set and DB-2021 as a test set, and vice versa. The results demonstrated that it is not possible to develop a model on the data of one harvest season and predict the samples of another season. The prediction model should also contain a percentage of the strawberry fruits of the season to be predicted. This demonstrated the season variability of the fruit, whose characteristics are strongly influenced by environmental and pedoclimatic conditions changing year to year. Even fruits of the same cultivar can present quite different qualitative characteristics in different harvest seasons [1]. Of course, the model maintenance is expected to last at least a few years until fruit variability is fully described, which would then ensure having strong models and more reliable results.

For this reason, calibration models were cross-validated by leaving out samples from both seasons at the same time; the same applied to the test set. A total of 692 fruit samples were used as a training set and 300 samples as a test set using a duplex algorithm. In detail, around 30% of DB-2020 (132 samples) and 30% of DB-2021 (168 samples) were used as a test set, and the remaining samples were used as a training set. One outlier sample was removed before PLS computation (as also reported in Section 3.6). Table 4 reports the results of the best models developed from the merged DB-2020 and DB-2021 datasets. The best prediction model was obtained using a first derivative (Savitzky–Golay with 21 smoothing points and second order) as pretreatment. The model used eight LVs and returned an R^2_p = 0.66, a RMSEP of 0.98 °Brix, a RPD = 1.70, and a RER = 10.20. PLS models were also computed to only considering the wavenumber range between 9000 cm^{-1} to 4000 cm^{-1}. The results are also reported in Table 4. The best performing models were developed by considering a first derivative (Savitzky–Golay with 13 or 21 smoothing points and second order) and SNV as pretreatments. In any case, the models did not improve the performance with respect to the full-range model.

Variable selection techniques were applied to the best performing model reported in Table 4. Table 4 reports the results of the best prediction models using variable selection techniques. rPLS led to variable reduction without any improvement in the prediction performance. iPLS (forward, 50 intervals) led to a significant improvement. The RMSEP value decreased from 0.98 °Brix to 0.90 °Brix, the R^2_p increased from 0.66 to 0.71, the RPD increased from 1.70 to 1.86, and the RER increased from 10.20 to 11.11. The values indicate the possibility to use the model for screening applications.

As for the other prediction models, the variables selected by the forward iPLS (50 size interval) method were investigated to figure out the most relevant wavenumbers for the prediction of the SSC and to facilitate the data interpretability of the model. It can be noted in Figure 7 that the iPLS method applied to the merged DB-2020 and DB-2021 selected the same spectral areas reported in Figure 6, which confirmed the robustness of the model and that the information of the SSC content was spread throughout the same region of the spectrum.

Table 4. Summary of partial least squares regression (PLS) results and summary of PLS results applying variables selection techniques to the merged DB-2020 and DB-2021 datasets for SSC (°Brix) prediction models. The best PLS models are marked in bold (xder1 = first derivative with x number of smoothing points; SNV = standard normal variate; LVs = latent variables; RMSECV = root mean square error of cross validation; RMSEP = root mean square error of prediction; RER = range error ratio; RPD = ratio of performance to deviation).

| Model	Pretreatment	LVs	RMSECV	R^2_{cv}	RMSEP	R^2_p	RER	RPD
PLS	13der1 + SNV	8	0.95	0.61	0.99	0.65	10.10	1.69
PLS	13der1	7	0.98	0.59	1.08	0.58	9.26	1.55
PLS	21der1 + SNV	7	0.97	0.59	1.07	0.59	9.35	1.56
PLS	**21der1**	**8**	**0.91**	**0.64**	**0.98**	**0.66**	**10.20**	**1.70**
PLS	**Cut_13der1 + SNV**	**8**	**0.92**	**0.64**	**0.98**	**0.66**	**10.20**	**1.70**
PLS	Cut_21der1	7	0.97	0.6	1.06	0.60	9.43	1.58
PLS	**Cut_21der1 + SNV**	**9**	**0.89**	**0.66**	**0.98**	**0.66**	**10.20**	**1.70**
rPLS (8 LVs)	21der1	7	0.82	0.71	0.98	0.66	10.20	1.70
iPLS (for,50)	**21der1**	**8**	**0.79**	**0.73**	**0.90**	**0.71**	**11.11**	**1.86**
iPLS (for,100)	21der1	10	0.83	0.7	1.05	0.6	9.52	1.59
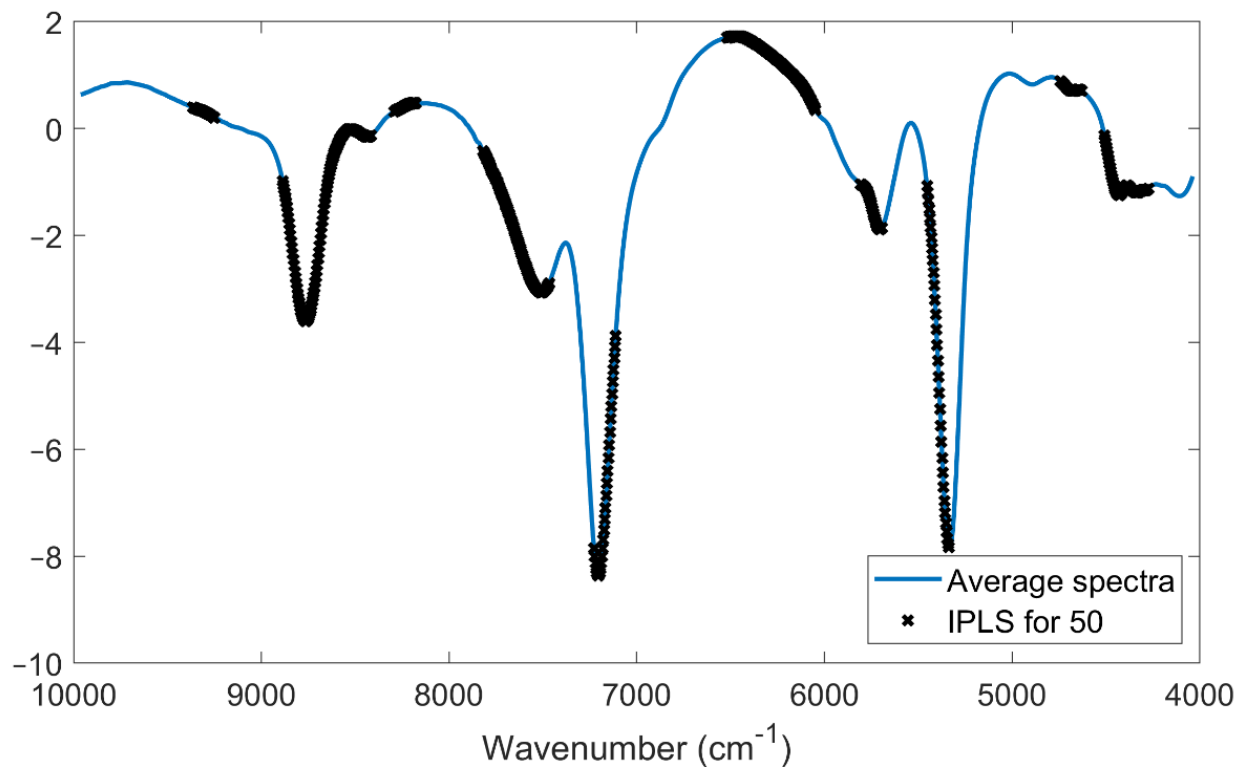
Figure 7. The average spectrum of the merged DB-2020 and DB-2021 datasets pretreated using first derivative (Savitzky–Golay filter, 21 smoothing points, second polynomial order) on the full spectral range. The variables selected by forward iPLS (50 intervals) are marked with black crosses. The red areas mark the most important wavenumber range for the prediction of SSC content (°Brix).

4. Discussion

The results of this present study are in agreement with those obtained by other authors; Włodarska et al. [24] found in the NIR range an external test set for predicting strawberry juice (R^2_p = 0.98, RMSEP = 0.25%) and fruit (R^2_p = 0.93, RMSEP = 0.46%) SSC. Amodio et al. [26] also used an external test set to predict SSC in strawberry fruits with an R^2_p = 0.85 and a RMSEP of 0.58%. Shen et al. [25] reported that postharvest strawberry quality (SSC) could be predicted online with an R^2_p = 0.733, a RMSEP = 0.699 °Brix, and a

RPD of 1.96. Also, the study of Sánchez et al. [15] used the LOCAL-PLS algorithm to predict the SSC in nine strawberry cultivars; they reported an R2p = 0.69 and a SEP = 0.88%. In a recent study, Xie et al. [50] investigated the relationship between optical properties (i.e., the absorption coefficient and scattering coefficient) and the internal quality of the strawberry (i.e., SSC and moisture content) by using partial least squares regression and support vector machine methods. The authors demonstrated the correlation between optical properties and strawberry quality, which confirmed that the development of non-destructive techniques will open up interesting perspectives for fruit internal quality control.

As previously mentioned, the strawberry chemical composition is strongly influenced by many factors, including pedoclimatic and environmental conditions. The different characteristics of the 2020 and 2021 harvest seasons, especially regarding weather conditions, affected the sensorial characteristics of the strawberry fruits in terms of SSC, which also caused differences in the results for the same cultivars. Furthermore, the presence of heavy rains close to the harvest moment affected the fruit integrity of the external surface; this led to a higher fruit susceptibility to *B. cinerea*, which, in turn, can alter neutral sugar and sugar acid levels in the infected host tissues [41], which affects the good results and the reproducibility of the prediction model.

The NIR method does not replace laboratory methods, but can be used in screening and evaluation processes that give an indication of the quality of fruit products. This method has the potential to become an important decision-making tool for the various stakeholders involved in the strawberry fruit chain. Moreover, this technology can be used both at line and offline. In this case, the main benefits gained from implementing this technology are simplified analytical execution, lower cost, and faster execution.

5. Conclusions

The results of this study led to interesting conclusions that can be summarized as follows:

- Genotype is an important factor in the SSC variability, with some cultivars (Sibilla and Romina) presenting higher amounts than others (Cristina, Silvia, and Sveva);
- The cultivation conditions affect the SSC content, which modify the amount of sugars in the same cultivar if examined during different harvest seasons;
- Cristina and Sveva were the most susceptible cultivars to *B. cinerea* infection, while Romina, Silvia, and Sibilla showed different behaviors according to the harvest year considered;
- The fruit SSC, serving as ripening initiation signal, resulted in quite a good correlation with the fruit susceptibility to *B. cinerea* infection;
- NIR spectroscopy, coupled with multivariate analysis, is a promising tool for the prediction of SSC in strawberry samples;
- PLS regression models for the prediction of SSC gave good results for DB-2020, for DB-2021, and for the merged DB-2020 and DB-2021 datasets;
- Variable selection techniques took the same spectral areas in all the datasets investigated, which confirmed the robustness of the models and that the information for the SSC prediction was related to the same region of the spectrum, which corresponded to OH bonds of water and CH bonds of carbohydrates.

To comply with the initial aims of the study, the results demonstrated how NIR spectroscopy, coupled with chemometrics, is a promising tool for the prediction of the SSC content of strawberries in screening applications. In addition, the relationship has been demonstrated between strawberry sugar composition and susceptibility/tolerance to *Botrytis cinerea* infection.

Author Contributions: Conceptualization, M.M., L.M., G.T., F.C. and B.M.; methodology, M.M., L.M., E.L. and R.Q.; software, M.M., L.M., F.G. and V.T.; validation, M.M. and L.M.; formal analysis, E.L., R.Q., V.T. and F.G.; investigation, M.M. and L.M.; resources, B.M., F.C. and G.T.; data curation, M.M., F.G., V.T. and L.M.; writing—original draft preparation, M.M. and L.M.; writing—review and

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Di Vittori, L.; Mazzoni, L.; Battino, M.; Mezzetti, B. Pre-Harvest factors influencing the quality of berries. *Sci. Hortic.* **2018**, *233*, 310–322. [CrossRef]
2. Pan, L.; Zhang, W.; Zhu, N.; Mao, S.; Tu, K. Early detection and classification of pathogenic fungal disease in post-harvest strawberry fruit by electronic nose and gas chromatography–mass spectrometry. *Food Res. Int.* **2014**, *62*, 162–168. [CrossRef]
3. Feliziani, E.; Romanazzi, G. Postharvest decay of strawberry fruit: Etiology, epidemiology, and disease management. *J. Berry Res.* **2016**, *6*, 47–63. [CrossRef]
4. Michailides, T.J.; Morgan, D.P.; Luo, Y. Epidemiological Assessments and Postharvest Disease Incidence. In *Postharvest Pathology. Plant Pathology in the 21st Century (Contributions to the 9th International Congress)*; Prusky, D., Gullino, M., Eds.; Springer: Dordrecht, The Netherlands, 2009; Volume 2.
5. Gallo, A.; Giuberti, G.; Frisvad, J.C.; Bertuzzi, T.; Nielsen, K.F. Review on mycotoxin issues in ruminants: Occurrence in forages, effects of mycotoxin ingestion on health status and animal performance and practical strategies to counteract their negative effects. *Toxins* **2015**, *7*, 3057–3111. [CrossRef] [PubMed]
6. Mezzetti, B.; Balducci, F.; Capocasa, F.; Zhong, C.-F.; Cappelletti, R.; Di Vittori, L.; Mazzoni, L.; Giampieri, F.; Battino, M. Breeding Strawberry for Higher Phytochemicals Content and Claim It: Is It Possible? *Int. J. Fruit Sci.* **2016**, *16*, 194–206. [CrossRef]
7. Capocasa, F.; Balducci, F.; Di Vittori, L.; Mazzoni, L.; Stewart, D.; Williams, S.; Hargreaves, R.; Bernardini, D.; Danesi, L.; Zhong, C.-F.; et al. Romina and Cristina: Two New Strawberry Cultivars with High Sensorial and Nutritional Values. *Int. J. Fruit Sci.* **2016**, *16*, 207–219. [CrossRef]
8. Sanzani, S.M.; Reverberi, M.; Geisen, R. Mycotoxins in harvested fruit and vegetables: Insights in producing fungi biological role, conducive conditions, and tools to manage postharvest contamination. *Postharvest Biol. Technol.* **2016**, *122*, 95–105. [CrossRef]
9. Butz, P.; Hofmann, C.; Tauscher, B. Recent Developments in Noninvasive Techniques for Fresh Fruit and Vegetable Internal Quality Analysis. *J. Food Sci.* **2006**, *70*, R131–R141. [CrossRef]
10. Giovenzana, V.; Beghi, R.; Civelli, R.; Guidetti, R. Optical techniques for rapid quality monitoring along minimally processed fruit and vegetable chain. *Trends Food Sci. Technol.* **2015**, *46*, 331–338. [CrossRef]
11. Flores, K.; Sanchez, M.-T.; Perez-Marin, D.; Guerrero, J.-E.; Garrido-Varo, A. Feasibility in NIRS instruments for predicting internal quality in intact tomato. *J. Food Eng.* **2009**, *91*, 311–318. [CrossRef]
12. Mancini, M.; Mazzoni, L.; Gagliardi, F.; Balducci, F.; Duca, D.; Toscano, G.; Mezzetti, B.; Capocasa, F. Application of the non-destructive NIR technique for the evaluation of strawberry fruits quality parameters. *Foods* **2020**, *9*, 441. [CrossRef] [PubMed]
13. Clément, A.; Dorais, M.; Vernon, M. Nondestructive Measurement of Fresh Tomato Lycopene Content and Other Physicochemical Characteristics Using Visible-NIR Spectroscopy. *J. Agric. Food Chem.* **2008**, *56*, 9813–9818. [CrossRef] [PubMed]
14. Nicolaï, B.M.; Defraeye, T.; De Ketelaere, B.; Herremans, E.; Hertog, M.L.; Saeys, W.; Torricelli, A.; VandenDriessche, T.; Verboven, P. Nondestructive Measurement of Fruit and Vegetable Quality. *Annu. Rev. Food Sci. Technol.* **2014**, *5*, 285–312. [CrossRef] [PubMed]
15. Sánchez, M.-T.; De La Haba, M.J.; Benítez-López, M.; Novales, J.F.; Garrido-Varo, A.; Perez-Marin, D. Non-destructive characterization and quality control of intact strawberries based on NIR spectral data. *J. Food Eng.* **2012**, *110*, 102–108. [CrossRef]
16. Pissard, A.; Pierna, J.A.F.; Baeten, V.; Sinnaeve, G.; Lognay, G.; Mouteau, A.; Dupont, P.; Rondia, A.; Lateur, M. Non-Destructive measurement of vitamin C, total polyphenol and sugar content in apples using near-infrared spectroscopy. *J. Sci. Food Agric.* **2012**, *93*, 238–244. [CrossRef]
17. Iyakwari, S.; Glass, H.; Rollinson, G.K.; Kowalczuk, P.B. Application of near infrared sensors to preconcentration of hydrothermally-formed copper ore. *Miner. Eng.* **2016**, *85*, 148–167. [CrossRef]
18. Pedersen, T.; Rantanen, J.; Naelapää, K.; Skibsted, E. Near infrared analysis of pharmaceutical powders with empirical target distribution optimization (ETDO). *J. Pharm. Biomed. Anal.* **2020**, *181*, 113059. [CrossRef]
19. Mancini, M.; Rinnan, Å.; Pizzi, A.; Toscano, G. Prediction of gross calorific value and ash content of woodchip samples by means of FT-NIR spectroscopy. *Fuel Process. Technol.* **2018**, *169*, 77–83. [CrossRef]
20. Toscano, G.; Rinnan, Å.; Pizzi, A.; Mancini, M. The Use of Near-Infrared (NIR) Spectroscopy and Principal Component Analysis (PCA) To Discriminate Bark and Wood of the Most Common Species of the Pellet Sector. *Energy Fuels* **2017**, *31*, 2814–2821. [CrossRef]

21. Sánchez, N.N.; Martínez-Marín, A.; Polvillo, O.; Fernández-Cabanás, V.; Carrizosa, J.; Urrutia, B.; Serradilla, J. Near Infrared Spectroscopy (NIRS) for the determination of the milk fat fatty acid profile of goats. *Food Chem.* **2016**, *190*, 244–252. [CrossRef]

22. Porep, J.; Kammerer, D.R.; Carle, R. On-Line application of near infrared (NIR) spectroscopy in food production. *Trends Food Sci. Technol.* **2015**, *46*, 211–230. [CrossRef]

23. Rinnan, Å.; van den Berg, F.; Engelsen, S.B. Review of the Most Common Pre-Processing Techniques for near-Infrared Spectra. *TrAC Trends Anal. Chem.* **2009**, *28*, 1201–1222. [CrossRef]

24. Włodarska, K.; Szulc, J.; Khmelinskii, I.; Sikorska, E. Non-Destructive determination of strawberry fruit and juice quality parameters using ultraviolet, visible, and near-infrared spectroscopy. *J. Sci. Food Agric.* **2019**, *99*, 5953–5961. [CrossRef]

25. Shen, F.; Zhang, B.; Cao, C.; Jiang, X. On-Line discrimination of storage shelf-life and prediction of post-harvest quality for strawberry fruit by visible and near infrared spectroscopy. *J. Food Process. Eng.* **2018**, *41*, e12866. [CrossRef]

26. Amodio, M.L.; Ceglie, F.G.; Chaudhry, M.M.A.; Piazzolla, F.; Colelli, G. Potential of NIR spectroscopy for predicting internal quality and discriminating among strawberry fruits from different production systems. *Postharvest Boil. Technol.* **2017**, *125*, 112–121. [CrossRef]

27. Yazici, A.; Tiryaki, G.Y.; Ayvaz, H. Determination of pesticide residual levels in strawberry (Fragaria) by near-infrared spectroscopy. *J. Sci. Food Agric.* **2020**, *100*, 1980–1989. [CrossRef] [PubMed]

28. Siedliska, A.; Baranowski, P.; Zubik, M.; Mazurek, W.; Sosnowska, B. Detection of fungal infections in strawberry fruit by VNIR/SWIR hyperspectral imaging. *Postharvest Biol. Technol.* **2018**, *139*, 115–126. [CrossRef]

29. Wang, Q.; Tao, S.; Dubé, C.; Tury, E.; Hao, Y.J.; Zhang, S.; Khanizadeh, S. Postharvest changes in the total phenolic content, antioxidant capacity and LPhenylalanine ammonia-Lyase activity of strawberries inoculated with botrytis cinerea. *J. Plant Stud.* **2012**, *1*, 11. [CrossRef]

30. Guidarelli, M.; Carbone, F.; Mourgues, F.; Perrotta, G.; Rosati, C.; Bertolini, P.; Baraldi, E. Colletotrichum acutatum interactions with unripe and ripe strawberry fruit and differential responses at histological and transcriptional levels. *Plant Pathol.* **2011**, *60*, 685–697. [CrossRef]

31. McKinney, H.H. Influence of soil temperature and moisture on infection of wheat seedlings by Helmintosporium sativum. *J. Agric. Res.* **1923**, *26*, 195–218.

32. Savitzky, A.; Golay, M.J.E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* **1964**, *36*, 1627–1639. [CrossRef]

33. Wold, S.; Martens, H.; Wold, H. The Multivariate Calibration Problem. In *Chemistry Solved by the PLS Method BT—Matrix Pencils*; Kågström, B., Ruhe, A., Eds.; Springer: Berlin/Heidelberg, Germany, 1983; pp. 286–293.

34. Snee, R.D. Validation of Regression Models: Methods and Examples. *Technometrics* **1977**, *19*, 415–428. [CrossRef]

35. Sørensen, L.K. Application of Reflectance near Infrared Spectroscopy for Bread Analyses. *Food Chem.* **2009**, *113*, 1318–1322. [CrossRef]

36. Williams, P.; Norris, K. *Near-Infrared Technology in the Agricultural and Food Industries*; American Association of Cereal Chemists, Inc.: St. Paul, MN, USA, 1987; ISBN 091325049X.

37. Andersen, C.M.; Bro, R. Variable Selection in Regression—A Tutorial. *J. Chemom.* **2010**, *24*, 728–737. [CrossRef]

38. Rinnan, Å.; Andersson, M.; Ridder, C.; Engelsen, S.B. Recursive Weighted Partial Least Squares (RPLS): An Efficient Variable Selection Method Using PLS. *J. Chemom.* **2014**, *28*, 439–447. [CrossRef]

39. Nørgaard, L.; Saudland, A.; Wagner, J.; Nielsen, J.P.; Munck, L.; Engelsen, S.B. Interval Partial Least-Squares Regression (IPLS): A Comparative Chemometric Study with an Example from Near-Infrared Spectroscopy. *Appl. Spectrosc.* **2000**, *54*, 413–419. [CrossRef]

40. Parikka, P. The effect of irrigation method on the quality and shelf-life of strawberry fruit in organic production. *Acta Hort.* **2006**, *708*, 319–322. [CrossRef]

41. Petrasch, S.; Knapp, S.J.; Van Kan, J.A.; Blanco-Ulate, B. Grey mould of strawberry, a devastating disease caused by the ubiquitous necrotrophic fungal pathogen Botrytis cinerea. *Mol. Plant Pathol.* **2019**, *20*, 877–892. [CrossRef]

42. Fagan, C.C.; Everard, C.D.; McDonnell, K. Prediction of Moisture, Calorific Value, Ash and Carbon Content of Two Dedicated Bioenergy Crops Using near-Infrared Spectroscopy. *Bioresour. Technol.* **2011**, *102*, 5200–5206. [CrossRef]

43. Schwanninger, M.; Rodrigues, J.C.; Fackler, K. A Review of Band Assignments in near Infrared Spectra of Wood and Wood Components. *J. Near Infrared Spectrosc.* **2011**, *19*, 287–308. [CrossRef]

44. Magwaza, L.S.; Opara, U.L.; Nieuwoudt, H.; Cronje, P.J.R.; Saeys, W.; Nicolaï, B. NIR Spectroscopy Applications for Internal and External Quality Analysis of Citrus Fruit—A Review. *Food Bioprocess Technol.* **2012**, *5*, 425–444. [CrossRef]

45. Zude, M. *Optical Monitoring of Fresh and Processed Agricultural Crops*, 1st ed.; CRC Press: Boca Raton, FL, USA, 2008.

46. Gómez, A.H.; He, Y.; Pereira, A.G. Non-Destructive Measurement of Acidity, Soluble Solids and Firmness of Satsuma Mandarin Using Vis/NIR-Spectroscopy Techniques. *J. Food Eng.* **2006**, *77*, 313–319. [CrossRef]

47. Guthrie, J.A.; Walsh, K.; Reid, D.; Liebenberg, C.J. Assessment of Internal Quality Attributes of Mandarin Fruit. 1. NIR Calibration Model Development. *Aust. J. Agric. Res.* **2005**, *56*, 405–416. [CrossRef]

48. Wang, H.; Peng, J.; Xie, C.; Bao, Y.; He, Y. Fruit Quality Evaluation Using Spectroscopy Technology: A Review. *Sensors* **2015**, *15*, 11889–11927. [CrossRef] [PubMed]

49. Golic, M.; Walsh, K.; Lawson, P. Short-Wavelength Near-Infrared Spectra of Sucrose, Glucose, and Fructose with Respect to Sugar Concentration and Temperature. *Appl. Spectrosc.* **2003**, *57*, 139–145. [CrossRef] [PubMed]
50. Xie, D.; Liu, D.; Guo, W. Relationship of the Optical Properties with Soluble Solids Content and Moisture Content of Strawberry during Ripening. *Postharvest Biol. Technol.* **2021**, *179*, 111569. [CrossRef]