Article

Miniaturized NIR Spectrometers in a Nutshell: Shining Light over Sources of Variance

Giulia Gorla [1] , Paolo Taborelli [1] , Hawbeer Jamal Ahmed [2] , Cristina Alamprese [3] , Silvia Grassi [3] ,
Ricard Boqué [2] , Jordi Riu [2] and Barbara Giussani [1,*]

[1] Department of Science and High Technology, University of Insubria, Via Valleggio 9, 22100 Como, Italy
[2] Chemometrics, Qualimetrics and Nanosensors Group, Department of Analytical Chemistry and Organic
    Chemistry, Universitat Rovira i Virgili, Campus Sescelades, 43007 Tarragona, Spain
[3] Department of Food, Environmental and Nutritional Sciences, Università degli Studi di Milano,
    Via Giovanni Celoria 2, 20133 Milan, Italy
*   Correspondence: barbara.giussani@uninsubria.it

Abstract: The increasing portability and accessibility of miniaturized NIR spectrometers are promoting the spread of in-field and online applications. Alongside the successful outcomes, there are also several problems related to the acquisition strategies for each instrument and to experimental factors that can influence the collected signals. An insightful investigation of such factors is necessary and could lead to advancements in experimental set-up and data modelling. This work aimed to identify variation sources when using miniaturized NIR sensors and to propose a methodology to investigate such sources based on a multivariate method (ANOVA—Simultaneous Component Analysis) that considers the effects and interactions between them. Five different spectrometers were chosen for their different spectroscopic range and technical characteristics, and samples of worldwide interest were chosen as the case study. Comparing various portable sensors is interesting since results could significantly vary in the same application, justifying the idea that this kind of spectrometer is not to be treated as a general class of instruments.

Keywords: Miniaturized NIR spectrometers; sources of variance; food applications; ASCA; instrument stability

1. Introduction

According to the state of the art, miniaturized near infrared (NIR) spectrometers have proved useful for a wide range of analytical challenges [1–5]. However, the potential applications and relative performance may differ from instrument to instrument [6]. Coupling NIR measurements with chemometrics techniques has become extremely established during the years [7] because of the complex and self-correlated information contained in the spectra. Moreover, frequently the analytical information about the analyte of interest is not evident in the spectra but included in the data due to correlation and interactions with other sample features (i.e., physical form and components of samples).

In general, field spectrometers can be transportable (mounted to a vehicle for field use), portable in a suitcase (total weight >4 kg), and handheld (<1 kg). A miniaturized instrument is one that is no larger than the size of a book [1]. Scientists working with NIR miniaturized instruments usually compare the outcomes from benchtop and handheld devices, mainly based on the typical performance parameters of prediction or classification models. Commonly, results show better performances for the benchtop equipment than for the different portable devices [8–11]. This is totally reasonable given that the robustness of the benchtop spectrometers has been achieved through years of analyses and technological development. Apart from that, standard benchtop NIR devices are specialized equipment, often used in laboratories (controlled environments) by experienced staff due to their size and cost. The portability and accessibility (including a lower cost, sometimes with a

decrease of one order of magnitude from benchtop devices) of miniaturized sensors offer on one hand the opportunity of in-field and online analysis, but on the other hand there are several issues related to their use by non-expert personnel [6]. Additionally, what is nowadays evident in the use of benchtop systems, as for instance the choice of the suitable sample cell and the analytical procedure to collect spectra, it is not so straightforward for miniaturized sensors, being necessary in most cases to develop a specific analytical procedure tailored to the sample and to the miniaturized NIR instrument. While the use of several chemometric strategies has been reported for model development [12], experiences with different acquisition strategies for miniaturized sensors [13,14] are rarely encountered in the literature and their influence on the resulting spectral data has only recently been investigated [15].

Other interesting aspects arise when comparing different portable sensors since their behaviour in the same application may significantly vary [16–18]. The main reasons for these variations could be related to the different factors that affect the measurements during an NIR analysis, especially in the case of diffuse external reflectance. Above all, the spectrometers cover different spectroscopic ranges, thus the radiation penetrates differently in the samples. Moreover, the sample granulometry, shape, and colour could influence at several levels of the analysis output. On top of that, technological differences [19,20] can play a determinant role during the analysis. For instance, the sensors scan sample areas of different size, and this is a key point when dealing with inhomogeneous samples.

In this context, an examination of the potential sources of variance that may arise when using handheld instruments is interesting for several reasons, including potential advancements in experimental set-up and data modelling, as well as identification of the limits of experimental measurements. Methods to highlight the specific characteristics of spectrometers and to perform careful use of them can be of interest at many levels considering that the time and effort required to train and validate models may be higher than the cost of a single new sensor. First, information about the experimental setup can benefit the application development. Then, the data analysis can take advantage of a prudent development of models resulting in better performances for the final application, and, finally, the spectrometer itself could be technologically optimized.

ANOVA—Simultaneous Component Analysis (ASCA) has already been proved efficient in combining the advantages of design of experiments and multivariate exploratory analysis. In preliminary analysis for method set-up, joining experimental design and a powerful tool for an easy interpretation could really be a game changer. ASCA allows the user to study whether a set of data is significantly affected by certain experimental parameters or factors and to evaluate the magnitude of the influences. It has been successfully applied to produce process understanding through spectroscopic measurement [21–24], study influences due to sample positioning [25], or to common measurement procedures during the use of miniaturized NIR [15].

The aim of this work is to propose a methodology to investigate these sources of variance using a multivariate method that takes into account effects and interactions between them. In this research, an experimental design data structure was planned by considering, as factors, the type of sample, the power supply system, the timing of background acquisition, the session of analysis, and the replicates with the focus of studying the influence of these factors on the spectroscopical signal. The significant main and interaction effects were investigated by using ASCA. In addition, highlights of the spectrometers' characteristics related to the different samples are reported.

In this study, reflectance spectra of different samples (i.e., granulated sugar, sugar lumps, brown sugar, milled rice, and red rice) were acquired under different conditions with five miniaturized spectrometers. Reflectance spectroscopy is well known for carrying chemical and physical information about samples [26]. The five selected sensors were chosen for the different covered spectral ranges and optical configurations. The spectrometers did not represent all the miniaturized NIR available on the market. Still, they could offer an interesting overview of the possible sources of variance and examples to show methodologi-

cal strategies. Sugar and rice samples were chosen to investigate the peculiarity and sources of variance of the spectrometers under study. The selection of the samples was mainly based on their known stability over time and the worldwide interest in these matrices. Moreover, rice and sugar present physical characteristics similar to many other powdered and granulated samples. The main differences among the samples were granulometry, colour, and physical state. Rice samples of medium grain size permit a different light path and interaction different from those of the smaller granules of sugar. Milled rice and red rice were considered for the different influences on the external reflection spectra due to colour and chemistry. In the same way, brown and white sugar are interesting for the different particle size and colour. Moreover, the analysis of granulated and lump sugar samples could unlock considerations on the packing form.

2. Materials and Methods

2.1. Samples

Granulated sugar (i.e., sucrose—Eridania semolato classico), sugar lumps (Eridania zollette classico), granulated brown sugar (Esselunga zucchero di canna grezzo), milled rice (Esselunga carnaroli), and red rice (Riso Mittino–Rosso Ermes) were purchased from a local supermarket in Como (Italy) and maintained at room temperature in a protected environment (sealed containers) for the duration of the experiments. The same samples were packaged under nitrogen and sent to the different laboratories where the experiments took place. No pretreatments were conducted on the samples. All spectra were acquired with the following fixed order: brown sugar (A), granulated sugar, sugar lumps, milled rice, red rice, and brown sugar (B). A and B are used to intend the same sample, analyzed at the beginning and end of an analytical session.

2.2. Spectrometers

Since miniaturized spectrometers available on the market present different technological characteristics and could be used with different acquisition configurations, the main features and the details of the analysis carried out with the five sensors used in this study are reported. The miniaturized NIR spectrometers under investigation have different technological and spectroscopic features and prices. Examples of configurations used to carry out the analyses are reported in Supplementary Figure S1. The variability in the technologies is also reflected in the availability of software, applications, cloud-solutions, and other connecting solutions. Moreover, the parameters to be set-up through the interfaces are mainly different. There is no homogeneity in the instrumental parameters and for each spectrometer an ad hoc optimization is required. The main information on the spectrometers is summarized in Table 1.

2.2.1. SCiO

The 1080 spectra were recorded in Tarragona (ES) with SCiO NIR spectrometer hardware version 1.2 from Consumer Physics (Herzliya, Israel). The sensor was operated by the SCiO "The Lab" mobile application on an Android phone using Bluetooth connection. The spectrometer allowed the acquisition under charge and standing on its own battery. Spectra were acquired after calibration with the built-in calibration device in the sensor cover. After the first background, needed to start the analyses, all the following acquisitions can be set by the operator. The sensor was positioned with the window and detector facing downward. The granulated sugar samples and rice grains were placed in a 40 mm diameter petri dish without applying any pressure apart from that of the sensor during the analysis (contact measurements). A single sugar lump was placed directly under the spectrophotometer window paying attention to the obscuration of both the light source and the detector. No additional acquisition parameters (i.e., exposure time and spectrum averaging) can be optimized. Data were stored in The Lab cloud environment and then the raw data were exported to a local Windows portable computer and used for further data

analysis. The exported NIR spectra consisted in 331 individual variables. Supplementary Figure S2 shows the dataset structure.

Table 1. Main technical specifications and characteristics of selected miniaturized NIR spectrometers. Parameters are extracted from the technical sheet of the vendors. LED: light-emitting diodes, LVF: linear variable filter.

Spectrometer (Vendor)	Source	Wavelenght Selector	Detector	Wavelenght Region	Spectral Resolution (at Wavelength)	Weight (Dimension)	Connectivity	Approximate Order of Unit Cost ($)
SCiO (Consumer Physics)	LED	Bandpass filter	Si photodiode (array, 12 elements)	740–1070 nm	N/A	35 g (5.4 × 3.6 × 3.5 cm)	Bluetooth (Cloud service)	<1000
MicroNIR 1700 (VIAVI)	Tungsten halogen	LVF	InGaAs (array; 128 elements)	908–1676 nm	12.5 (1000) 25 (2000)	64 g (4.5 × 5.0 cm)	USB cable	<30,000
AvaSpec-Mini-NIR (AVANTES)	Tungsten halogen		InGaAs (array; 256 pixels)	900–1750 nm	12(1000) 24 (2000)	185 g (9.5 × 6.8 × 2.0 cm)	USB cable	<10,000
NeoSpectra Scanner (Si-Ware Systems)	Tungsten halogen	MEMS Michelson interferometer	InGaAs (single element)	1350–2550 nm	16 (1550)	1 kg (17.8 × 9.1 × 6.2 cm)	Bluetooth	<5000
NeoSpectra development kit (Si-Ware Systems)	Tungsten halogen	MEMS Michelson interferometer	InGaAs (single element)	1350–2558 nm	16 (1550)	17 g (32. × 3.2 × 2.2 cm)	USB cable	<5000

2.2.2. MicroNir OnSite-W

The 810 spectra were recorded in Milan (IT) by using the handheld MicroNIR[TM] OnSite spectrometer (VIAVI Solutions Inc., Santa Rosa, CA, USA). The sensor was operated by MicroNIR Pro software suite and a Windows laptop connected through a cable. The spectrometer allowed the acquisition only when connected to a computer. External white and black references were scanned manually before spectra acquisition. After the first background, needed to start the analyses, two possible solutions for the following background acquisitions are possible: performing the background acquisition when suggested by the integrated application and collecting the background when considered necessary by the operator. The sensor was positioned with the window facing downward and blocked on a holder system. The samples of granulated sugar and rice were placed in a 40 mm diameter petri dish without applying any pressure apart from that of the sensor that was placed in contact. In the case of sugar lumps, a single cube was placed directly under the spectrophotometer window paying attention to cover the whole acquisition window. The spectra acquisitions were performed with 12.5 μs integration time and 200 scans at 80 Hz. Data were stored in .CSV files and then imported into MATLAB for data analysis. The exported NIR spectra consisted of 125 individual variables. Supplementary Figure S3 shows the dataset structure.

2.2.3. AvaSpec-Mini-NIR

The 1080 spectra were recorded in Como (IT) with AvaSpec-Mini-NIR from Avantes coupled with AvaLight-HAL-S-Mini2 source. The 540 spectra were acquired with a reflection fiber probe (7 × 400 μm fibers, 2 m length, SMA term) while 540 spectra were acquired using AvaSphere-50-LS-HAL. The spectroscopic range covered was 972–1701 nm. The sensor was operated by the AvaSoft 8 application on a Windows laptop connected through a cable. The spectrometer needs to be connected to a computer and the light source needs to be connected to a power supply with a power cable. Spectra were acquired after calibration with a black reference with the source turned off and then the total reflectance reference with a white standard (WS-2). After the first background, all the following background acquisition can be set by the operator. The analysis with the optical fiber were conducted by

placing the reflectance probe in a holder (RPH-1) and by placing granulated sugars and rice samples in a 10 mL beaker without applying any pressure, whereas a single sugar lump was placed directly under the holder. An integration time of 15 ms and 10 average scans were set as acquisition parameters. Data were stored directly in .CSV and then imported into MATLAB for elaboration. The exported NIR spectra consisted of 236 individual variables. Supplementary Figure S4 shows the dataset structure.

2.2.4. NeoSpectra Scanner

The 1080 spectra were acquired in Como (IT) using the handheld NIR spectrometer from Si-Ware Systems (18.5 × 4.5 × 8 mm with a weight of approximately 730 g). The spectroscopic range covered was 1351–2559 nm. The sensor was operated by the proprietary mobile application on an Android phone using Bluetooth connection. The spectrometer allowed the acquisition under charge and standing on its own battery. Spectra were acquired after calibration with a 100% reflectance reference with a Spectralon® standard, approximately 15 min after the spectrometer was turned on. After the first background, all the others can be set by the operator. The sensor was positioned with the window facing downward in the case of granulated sugars and rice samples, presented to the spectrometer in a 40 mm diameter petri dish without applying any pressure. A single sugar lump was placed directly over the spectrophotometer window while the sensor was standing with the window facing upward. A time scan of 5 s without data interpolation were used. Data were autosaved in the default format ".Spectrum" and then converted to .txt files and processed with MATLAB. The exported NIR spectra consisted of 74 individual variables. Supplementary Figure S5 shows the dataset structure.

2.2.5. NeoSpectra Micro Development Kit

The 540 spectra were collected in Tarragona (ES) using the handheld NIR spectrometer from Si-Ware Systems (2 × 32 × 22 mm with a weight of 17 g). The spectroscopic range covered was 1351–2558 nm. The device was connected to a Windows personal computer via a universal serial bus (USB) and operated by the software package SpectroMOST. Spectra were acquired after calibration with a Spectralon (99% reflectance) reflectance standard. Apart from the first background, all the subsequent background acquisitions can be set by the operator. Granulated sugars and rice samples were presented to the spectrometer in an in-house constructed cell [14,24] without applying any pressure. The sugar lump was placed directly over the spectrophotometer window by covering it. A time scan of 5 s without data interpolation was used. The exported NIR spectra consisted of 134 individual variables. Supplementary Figure S6 shows the dataset structure.

2.3. Experimental Set-Up: Sources of Variability

The experimental set-up considered as the main influencing factors: (1) the type of sample, (2) the order of replicate, (3) the session of analysis, (4) the power supply during spectra acquisition, and (5) the timing of background.

Type of sample identifies different samples, i.e., granulated sugar (i.e., sucrose), sugar lumps, brown sugar, milled rice, and red rice, which were analyzed in fifteen experimental replicates for each sample during each session from the first replicate to the fifteenth. Each experimental replicate is intended as the displacement and the repositioning of the samples on the sample holder. The analytical session identifies three independent sessions of analysis. At the end of the acquisition of a sequence of samples, the spectrometers were switched off; a new session started only after the cool-down and warm-up time suggested by the manufacturers. Concerning the charge condition, SCiO and NeoSpectra Scanner sensors allow the recording of spectra with the instruments under charge or operating on their own battery. MicroNIR[TM] OnSite, AvaSpec-Mini-NIR, and NeoSpectra development kit could work only when connected to a computer and power supply. Timing of background is used to intend background before each sample, background before each session, or for the MicroNir[TM] Onsite, background when suggested by the

instrument software. Indeed, MicroNIR^{TM} OnSite has an automatic alarm indicating the need for a background reference (dark and white references are available). The timing of background factor was intended to evaluate the elapsed time between a background acquisition and another.

Moreover, existing factors related to instrumental acquisition range and technological features were considered (Table 1). AvaSpec-Mini-NIR allows the acquisition with different technological solutions: integrating sphere and fiber optic cable.

2.4. Chemometrics Analysis

The datasets obtained for each instrument were considered firstly as a whole ("all samples"), and then divided for rice and sugar samples. Afterwards, models for the spectra of each type of sample were considered. Data mean centering was set as the default for preprocessing in the data analysis.

Exploratory analysis was carried out with Principal Component Analysis (PCA) [27]. It was performed on each of the mean-centered datasets to identify gross errors using the Q residuals and Hotelling's T2 statistics. The spectra identified as outliers were removed from further analyses. Common preprocessing methods, such as multiplicative signal correction (MSC) and standard normal variate (SNV), were also applied [28,29].

ANOVA–Simultaneous Component Analysis (ASCA) [30,31] was used to evaluate the significance of the main considered factors and their interactions. ASCA was proposed for the analysis of multivariate datasets obtained through designed experiments [32]. The model consists in the partitioning of the original data matrix into a set of matrices corresponding to design factors (e.g., *a* and *b*) and their interactions (*ab*) and decomposing them through simultaneous component analysis (SCA):

$$X_{original} = X_{mean} + X_a + X_b + X_{ab} + X_{res}$$

where X_{res} represents the residuals and X_{mean} is equal to zero since the original matrix is centered. The resulting matrices consist in the mean profiles calculated by averaging all the replicates at a specific level of each factor or interaction. For instance, in the case of spectra acquired with the spectrometer under charge or on battery, half of the rows in the $X_{power\ supply}$ will contain the mean profile of the spectra in which the spectrometer was under charge and the other half will contain the average of the spectra acquired with the sensors operating on their own battery.

From the interpretative point of view, the procedure can be thought as a combination of analysis of variance (ANOVA) and PCA under the scheme constraint. The effect of factors and interactions are evaluated by the sum-of-squares of the corresponding *j* submatrix (according to the previous equation, *j* could be *a*, *b*, *ab* or *res*).

$$SSQ_j = X_j^2$$

The magnitude of the calculated effect is indicative of the influence magnitude of the specific factor or interaction on the data. The significance is then assessed by counterposing the effect to a null-distribution, nonparametrically estimated by permutation test. If the sum of squares is systematically larger than the values of the null-distribution, a *p*-value < 0.05 is obtained and the tested effect is assumed to be significant.

Models for each instrument were calculated and 200 permutations were used to evaluate the significance of the factors. The factors taken into consideration during the study were type of sample, order of replicates, session, power supply condition within the session, and timing of background. Two-way interactions were also calculated. Firstly, the whole samples were considered, while considering the factor sample. In this case, the brown sugar was included only once even if acquired twice. Subsequently, the ASCA model for each of the samples was carried out. In the brown sugar model, the factor "At the begging or end of the session" was added and all the spectra acquired were considered.

Data analysis was performed using routines and toolboxes developed in the MATLAB R2021a environment (the Mathworks Inc., Natick, MA, USA). Principal component analysis and ASCA were carried out by PLS-Toolbox v. 9.0 (Eigenvector Inc., Manson, WA, USA).

3. Results and Discussion

3.1. Spectra and Features

Figure 1 shows the average spectra of granulated sugar collected with the five spectrometers and the two configurations for the AvaSpec-Mini. The different wavelength ranges are also represented. While benchtop instruments typically cover the entire range represented (800–2500 nm) [33–35], miniaturized spectrometers typically cover narrower ranges. When dealing with miniaturized sensors, the specific range of acquisition may influence analysis in different ways and careful evaluations need to be pursued for the application of interest. In addition to this, the technological development of different manufacturers should be intended for different scopes, for instance, use in the field or online implementation. With respect to the color [36,37] SCiO works in a spectral range that reaches the visible part of the electromagnetic spectrum, adding information that may not be of interest.

A priori knowledge about the applicability of a certain sensor is not straightforward to be evaluated only based on the spectral ranges [38]. Indeed, when dealing with reflectance NIR spectroscopy the information correlated to the properties of interest is not always strictly related to evident chemical spectral information of the analyte or bulk. On the contrary, it is quite probable that the information is distributed over wider spectral ranges or reflected in the physical information that non-exclusively depends on the spectral range [7,39]. Physical parameters, such as compactness and particle sizes, also influence the acquisition of external reflectance NIR spectra because the resulting spectra are strictly related to light path, energy, and power.

The miniaturization and the different engineering designs clearly influenced the working and performance characteristics (e.g., spectral region, sensitivity and precision, spectral resolution, and recordable intensity). Spectra might also be affected by the practical procedures, i.e., sample located near the light source or at a given distance. Figure 1 shows that even in the regions shared by different devices the signal shapes are slightly different, and the definition of the bands can vary. The overall intensity captured by the spectrometers is also quite different. For the two configurations of Ava-Spec-Mini-NIR it was found that there was, as expected, a higher intensity for the use of the integrating sphere than for the optical fiber. As for the direct comparison of NeoSpectra devices, lower values of reflectance (up to 17%) were registered with the NeoSpectra development kit compared to the Neospectra Scanner. These results could be ascribed to the differences in the spot and detector sizes, the number of light sources (seven tungsten halogen lamps for the Neospectra Scanner and three for the Neospectra development kit), and the analytical measurement procedures.

In Supplementary Figure S7, the average spectrum of each sample is represented for each analytical configuration and spectrometer. As expected, the red rice reflects the radiation less than the milled rice with every instrument and configuration. The noticeable differences are related to both the physical and chemical differences of the two samples. The comparison of the spectra of granulated and lump sugar highlights the influence of the compactness and granulometry in the investigated spectrometers. Generally, less intense signals were obtained for the sugar lumps even if the difference decreased at higher wavelengths. Peculiar spectra were obtained with SCiO whose spectra showed a consistent change of shape for the different rice samples. The same appeared to be true when comparing the spectra of granulated sugar and brown sugar samples, where their physical and chemical characteristics were identifiable at first sight.
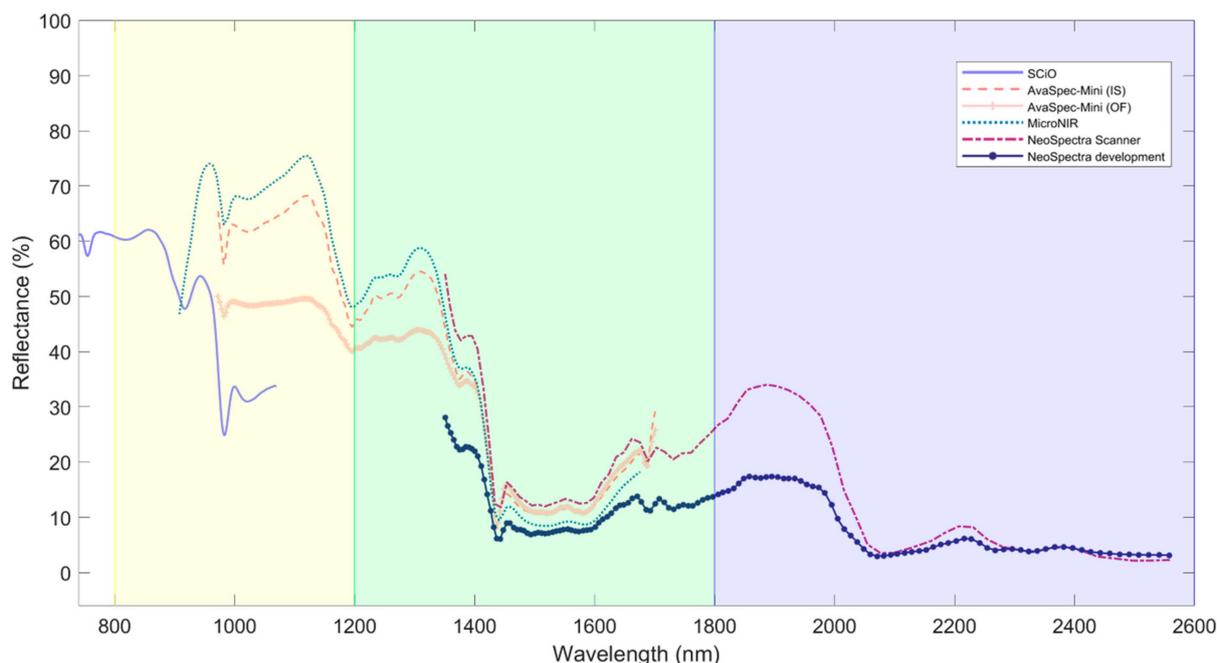
Figure 1. Averaged spectra of granulated sugar acquired with the different spectrometers and optical configurations. The background colors indicate the different NIR regions [38]: 800–1200 nm, Region 1 in yellow; 1200–1800 nm, Region 2 in green; 1800–2600 nm, Region 3 in blue.

3.2. Exploratory Analysis

The replicates obtained for each sample were used to build an **X** matrix, one for each instrument configuration for a total of six datasets. Raw data were centered and then analyzed using Principal Component Analysis (PCA) to detect the presence of gross outliers, which were not detected. In the case of multiple spectra emerging as possible outliers on Hotelling's T2 and Q residuals statistics, possible trends related to the factors under investigation were studied to understand if they represent the intrinsic variability contained in the experimentations. Figure 2 shows the score plot obtained for the whole dataset under investigation, whereas in Figure 3 the loadings are shown. The way in which the replicates and samples are distributed over PC1 and PC2 for the spectrometers is noteworthy. The arrangement of the sample and their dispersion over each sample centroid is characteristic of each sensor. The study of the distribution of samples and replicates may provide interesting insights. The dispersion of the clouds of points related to each sample can be used as semi-quantitative information related to the precision of the analysis. Similar replicas led to closer points in principal component space, while more significant variability in the replicas is reflected in broader disposition.

For what concerns the spectra acquired with different spectrometers and configuration for the same samples, the variability could be related to the interaction between the compactness and sample granulometry and the intrinsic characteristics of the sensors. The influence of different compactness has been partially addressed in [40], while consideration on different granulometry due to sample pretreatment can be found in [41]. These topics emerge here as a possible future research direction.

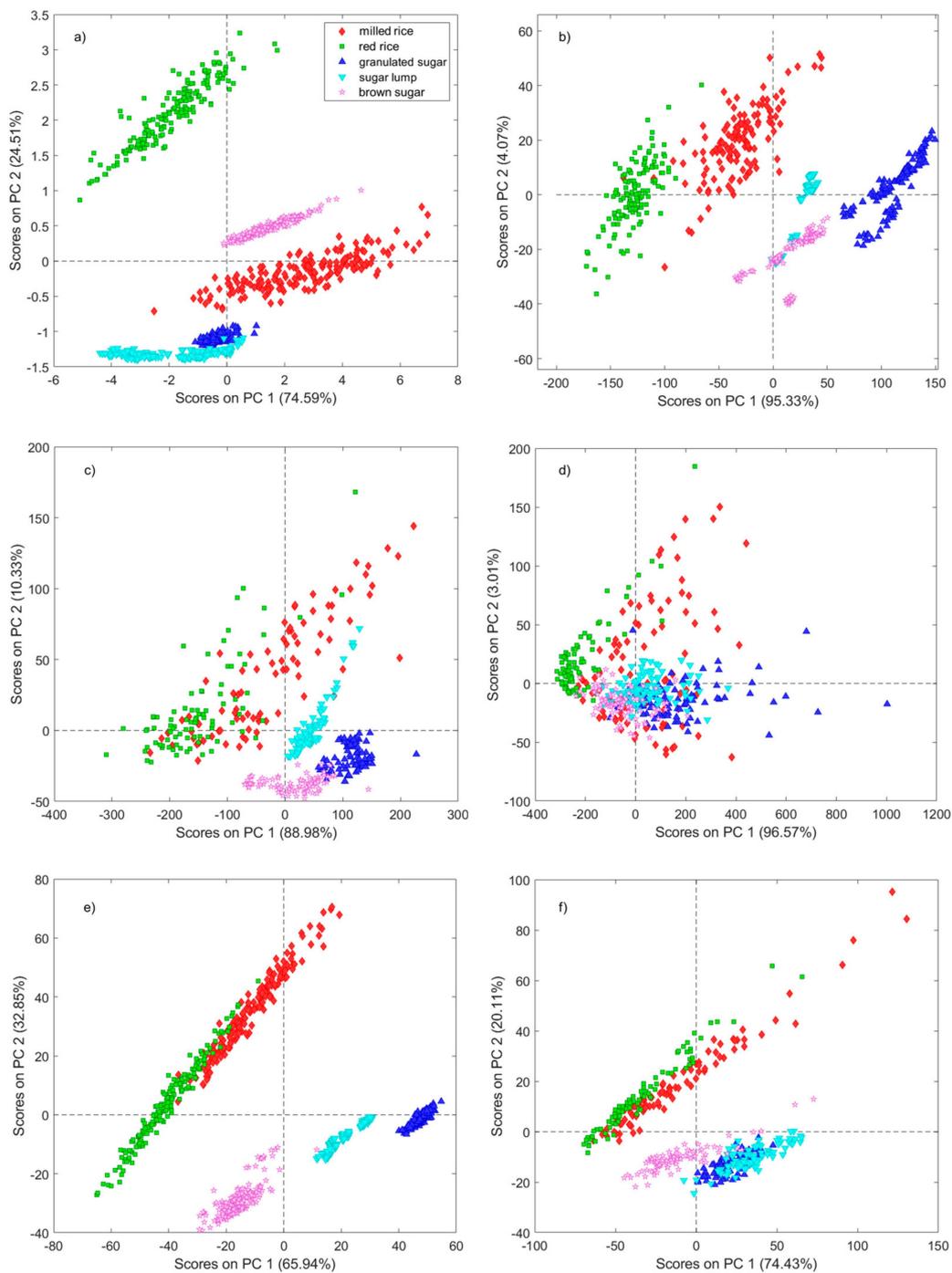**Figure 2.** PCA results on mean-centered data: score plot for (**a**) SCiO, (**b**) Viavi Micro-NIR, (**c**) AvaSpec with integrating sphere, (**d**) AvaSpec with optical fiber, (**e**) NeoSpectra Scanner, and (**f**) NeoSpectra Development kit.

Figure 3. PCA results on mean−centered data: loading plot for (**a**) SCiO, (**b**) Viavi Micro−NIR, (**c**) AvaSpec with integrating sphere, (**d**) AvaSpec with optical fiber, (**e**) NeoSpectra Scanner, and (**f**) NeoSpectra Development kit.
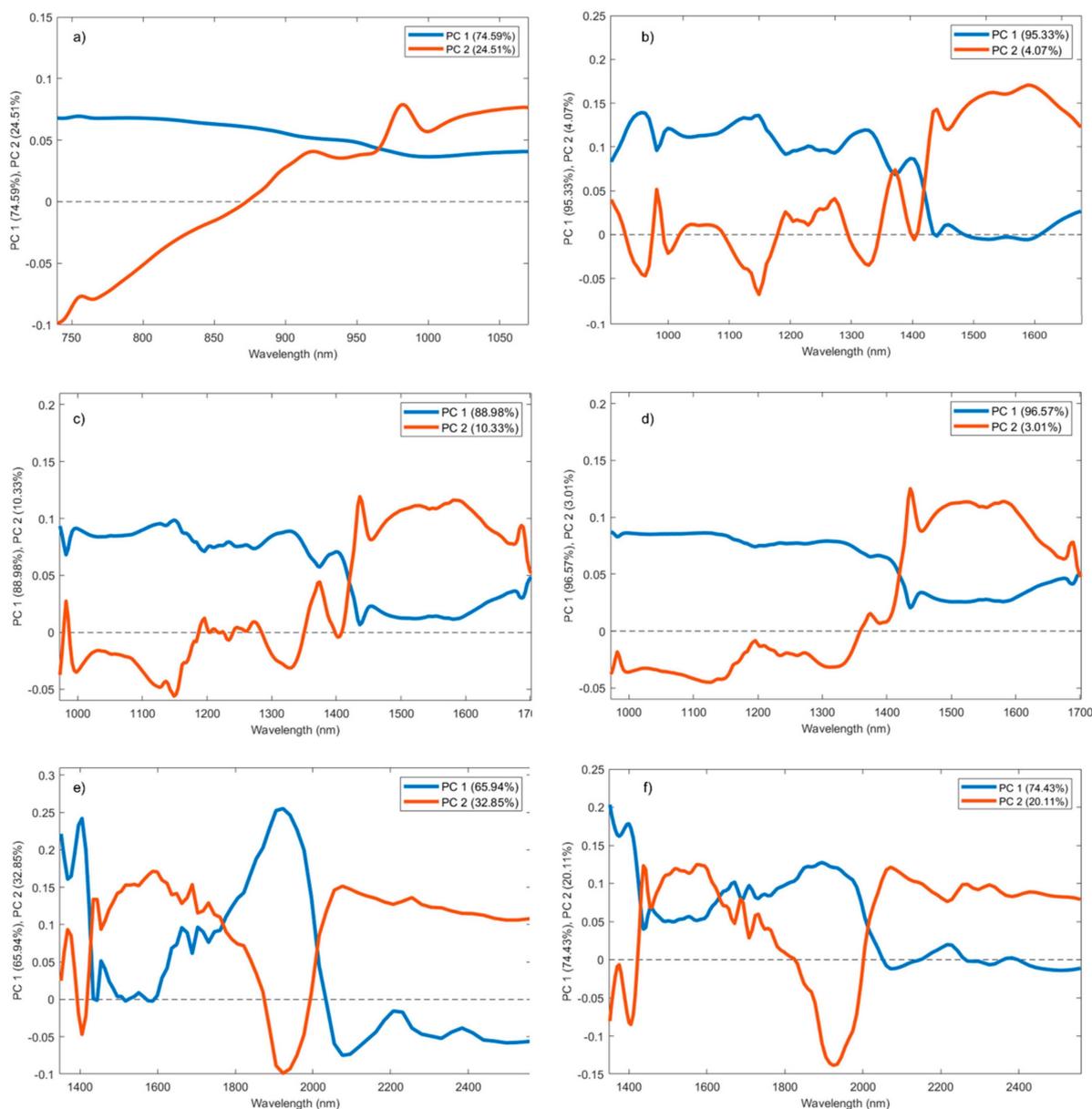
The separation between the five samples is not always accomplished in the scores plot, and it strictly depends on the ability of each spectrometer in capturing the substantial physical and chemical differences among the samples. Overall, sugar lump and granulated sugar spectra seem to have the highest precision for each sensor since the replicates appear closer in the PCA space than those of the other samples. These could be related to the analytical practices along with the sample characteristics; indeed, the presentation of the sample to the sensor is intuitively more constant for the packed form than for the loose one. Other interesting insights came from the fact that when the same sensor is used but different acquisition configurations are selected (Figure 2c,d), the scores are found to be grouped in different ways, and PCA models explained different percentages of variance. According to the comparison of the loading plots for the AvaSpec−Mini−NIR data (Figure 3c,d), the distribution of the samples in the score plot could be ascribed to the same spectral regions independently from the acquisition configuration. Indeed, the shape of loadings

looked remarkably similar. The differences encountered in the scores could be related to the different detail of the features registered. Neither the acquisition in the same spectral range nor with similar engineering technologies showed the same distribution of the samples (Figure 2e,f).

All these results lead to the conclusion that it is not straightforward to build a priori theories about instruments' performances in a specific application. The shapes found in the loadings suggested that both the physical and chemical characteristics of the samples were responsible for the score grouping or dispersion.

It is well known that one of the main sources of repeatability error in spectroscopy is related to scattering effects [42]. Since the final user is likely to use one of the most common preprocessing techniques for NIR spectra to remove scattering influences, investigations were carried out on how this affects the dataset under study. As for raw data, using MSC (Supplementary Figures S8 and S9) or SNV (Supplementary Figures S10 and S11) resulted in different patterns depending on the preprocessing and the spectrometer used for data acquisition.

In the case of SCiO, the use of MSC (which is dependent on the mean spectra of the dataset and corrects for scattering correlated to the wavelength), highly enhances the variability of the red rice while the scores for the spectra of other samples appeared all in the center of the model. The first two components were not sufficient for the discrimination of sample groups. The preprocessing correction was found to be substantially incompatible with the data under investigation, which varied based on chemical, physical, and also color features, the latter highly influencing the SCiO signal. A solution could reside in the use of different reference spectra during the MSC procedure according to the type of sample. On the contrary, when SNV is applied to the data, red rice and brown sugar emerged as distinctive groups, although very separated from the other samples.

Application of the scattering removal preprocessing techniques to MicroNir, AvaSpec, and NeoSpectra Scanner data resulted in very similar outputs. The application of such preprocessing techniques enhances the variability related to the discrimination between samples according to the macroscopic differences between rice and sugar spectra. The use of MSC or SNV did not differently affect the modelling phase. Spectra acquired with the optical fiber were consistently more variable than the ones obtained with the integrating sphere. The variability related to the chemistry of the samples predominantly emerged as the main variance in data, while white granulated and lump sugar samples turned out to be superimposed in the space of the first two PCs.

NeoSpectra development kit spectra showed an interesting behavior for MSC preprocessed data. The main variance (first PC) includes the intrinsic variability of the rice groups while the second component highlights spectroscopic differences between rice and sugar samples. The opposite behavior was found when applying SNV: the first PC was mainly related to the differences between rice and sugar spectra, whereas the second was necessary to explain the variability of the rice sample spectra.

The results obtained after applying spectra preprocessing methods depend on the spectrometer and the samples under investigation, which showed that generalized guidance could not be the most suitable. Consequently, a careful evaluation is needed for the choice of the instrument and the evaluation of the performances for a specific application, even in the case of preprocessing methods.

3.3. Studying the Factors Influencing the Analysis

3.3.1. Preliminary Models

Initially, ASCA models including all the spectra and the spectra divided according to the macro−chemical differences (rice or sugar samples) were calculated and investigated. The multivariate method permitted the separation of the influences of the investigated factors and their interactions. The interpretation of the results is somehow immediate. Some factors resulted significant in different models, whereas others were not, thus allowing the identification of their connection with a specific spectrometer and/or the sample

characteristics. The type of sample is the main significant factor in each model that included this source of variability even if with different magnitudes depending on the model and instrument or acquisition configuration considered. Other factors showed a behavior more related to the specific instrument. Results are reported in Supplementary Tables S1–S5.

3.3.2. ASCA Models on Different Samples

The outcomes of the ASCA models for each sample type were easily evaluated by visualizing and interpreting scores and loadings as in a common PCA. Some selected results for the sub−models obtained are represented in Figure 4, where some factors significant for the sensors are shown. The obtained scores for the sub−models are represented according to the factor levels, and spider graphs are used to enhance the scores centroid. Tendencies and grouping behaviors reflected the significance of the factors. The results are discussed below according to the spectrometer.

SCiO

ASCA models were carried out on each sample type to enhance the recurrence of significant factors. The influence of spectra preprocessing for scattering removal was also investigated. The results are summarized in Table 2, which shows the percentage each effect contributes to the sum of squares. The significant effects are highlighted in bold.

From the models of raw spectra, the timing of background emerged as significant for the majority of the samples, whereas other factors and interactions are sparse. For instance, the order of replicates was found significant only for sugar lump spectra, for which the interaction between the order of replicates and the power supply was also significant. When preprocessing was applied to data, interesting results arose. First, the magnitudes of the effects were identical, clearly highlighting that the multiplicative scattering is not present in the data. Secondly, some factors significant for specific samples became negligible when the scattering was removed, whereas others emerged. In general, it is relevant to notice that preprocessing increased the number and extent of significant factors and interactions. A reason could be that when removing the artifacts created by the scattering, the sources of variance typical for the instrument gain importance relative to the variation of the analysis. The sub−model obtained for the factor timing of background of red rice spectra is reported in Figure 4a,b.

The experimental procedure included the acquisition of brown sugar spectra at the beginning and at the end of each session in order to verify the factor and interaction related to the sample position in the order of analyses. The factor *begin or end of the session* was considered as well as the interactions with the other factors. No significant effects were found.
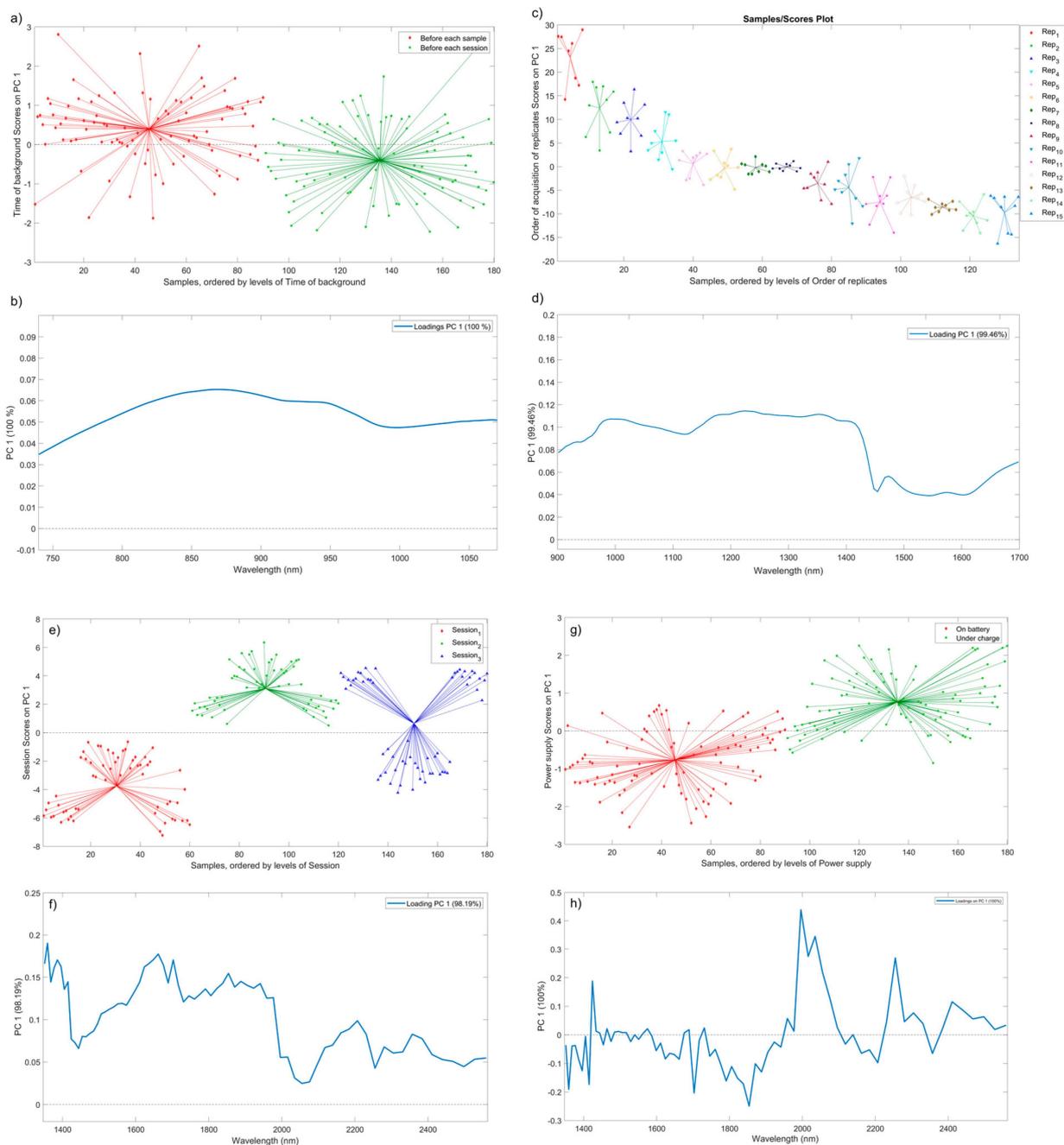
Figure 4. ASCA sub−models of raw data. Scores (**a**) and loadings (**b**) of the ASCA sub−model for the factor timing of background. Instrument: SCiO. Sample: red rice. Scores (**c**) and loadings (**d**) of the ASCA sub−model for the factor order of replicates. Instrument: MicroNir. Sample: brown sugar. Scores (**e**) and loadings (**f**) of the ASCA sub−model for the factor session. Instrument: NeoSpectra Scanner. Sample: sugar lumps. Scores (**g**) and loadings (**h**) of the ASCA sub−model for the factor power supply. Instrument: NeoSpectra Scanner. Sample: granulated sugar.

MicroNir OnSite−W

The results of ASCA models on each sample group with and without signal preprocessing are summarized in Table 3, which shows the percentage each effect contributes to the sum of squares. The significative effects are highlighted in bold.

Table 2. Summary of results for ASCA models of SCiO data. The effects are expressed as percentage of contribution to the sum of squares of the data. Models calculated on each sample. Factors that show a *p*-value < 0.05 after 200 permutations are reported in bold. Brown sugar A indicates the spectra acquired at the beginning of each analytical session. Brown sugar B indicates the spectra acquired at the end of each analytical session. *None* is used to identify the default preprocessing.

	Preprocessing	Sample	Factors				Interactions						Residuals
			Order of Replicates	Session	Power Supply	Timing of Background	Order of Replicates × Session	Order of Replicates × Power Supply	Order of Replicates × Timing of Background	Session × Power Supply	Session × Timing of Background	Power Supply × Timing of Background	
Effect (percentage contribution to the sum of squares)	None	Milled rice	10.56	0.04	0.34	0.21	15.53	8.48	5.18	0.72	2.60	**4.10**	52.23
		Red rice	7.49	**3.51**	1.19	**8.73**	13.90	10.80	4.30	1.26	0.02	**1.82**	46.99
		Granulated sugar	9.23	1.90	0.37	**4.40**	7.81	5.76	3.20	2.18	0.39	**7.87**	56.90
		Sugar lump	**20.51**	0.87	0.31	0.43	8.22	**17.94**	5.70	1.17	0.33	0.55	43.94
		Brown sugar A	8.9	0.98	0.08	**4.85**	11.5	5.5	10.07	3.32	1.32	0.35	53.11
		Brown sugar B	6.71	1.71	0.01	**2.53**	14.96	7.67	9.1	0.33	1.27	0.97	54.74
	MSC	Milled rice	7.64	**4.35**	**2.01**	**18.82**	10.30	4.96	4.75	0.89	1.40	0.31	44.58
		Red rice	8.87	**2.92**	0.98	**13.77**	12.99	4.13	5.29	1.07	0.84	0.30	48.83
		Granulated sugar	5.59	**7.04**	**5.60**	**33.75**	5.89	4.55	3.03	**1.78**	1.05	**3.06**	28.65
		Sugar lump	**18.25**	**3.14**	**1.54**	**16.50**	6.15	**13.94**	3.81	**3.34**	0.57	**1.02**	31.75
		Brown sugar A	8.62	**7.06**	**6.16**	**6.47**	8.79	5.19	6.93	**7.23**	**4.61**	0.81	38.15
		Brown sugar B	2.80	0.63	**0.84**	**59.74**	4.47	2.36	2.15	**1.40**	1.23	**1.61**	22.78
	SNV	Milled rice	7.64	**4.35**	**2.01**	**18.82**	10.30	4.96	4.75	0.89	1.40	0.31	44.58
		Red rice	8.87	**2.92**	0.98	**13.77**	12.99	4.13	5.29	1.07	0.84	0.30	48.83
		Granulated sugar	5.59	**7.04**	**5.60**	**33.75**	5.89	4.55	3.03	**1.78**	1.05	**3.06**	28.65
		Sugar lump	**18.25**	**3.14**	**1.54**	**16.50**	6.15	**13.94**	3.81	**3.34**	0.57	**1.02**	31.75
		Brown sugar A	8.62	**7.06**	**6.16**	**6.47**	8.79	5.19	6.93	**7.23**	**4.61**	0.81	38.15
		Brown sugar B	2.80	0.63	**0.84**	**59.74**	4.47	2.36	2.15	**1.40**	1.23	**1.61**	22.78

For the raw data, the factor *timing of background*, and the interaction between *session × timing of background* were significant for all the samples under investigation. The importance of the session of analysis is manifested for almost all the samples. The order of replicates was significant for milled rice and brown sugar (Figure 4c,d). In addition, the values for the residuals were found to be mainly different depending on the sample considered. This could be ascribed to the concurrence of other factors not taken under control and interactions of higher levels.

Preprocessing was applied to the data and, as for SCiO, the calculated effects were identical showing that the multiplicative scattering is not present in the data. Preprocessing in this case did not change the significant factors and interactions apart from two specific samples with similar particle sizes: *session* for red rice and *session × timing of background* for milled rice. Generally, with the higher quality of the preprocessed data, the value of the considered effects increased. An exception was found for the influence of the timing of background, which increased for rice samples and decreased for sugars.

The study conducted to identify if the acquisition of a sample at the beginning or at the end of the session may affect the resulting spectra showed that the factor *begin and end of the session* had a *p*−value of 0.075, whereas its interaction with the timing of background had an effect of 15.04 and a *p*−value of 0.005.

Table 3. Summary of results for ASCA models of MicroNIR^{TM} OnSite data. The effects are expressed as percentage of contribution to the sum of squares of the data. Models calculated on each sample. Factors that shown a p−value < 0.05 after 200 permutations are reported in bold. Brown sugar A indicates the spectra acquired at the beginning of each analytical session. Brown sugar B indicates the spectra acquired at the end of each analytical session. *None* is used to identify the default preprocessing.

| | | | Factors | | | Interactions | | | |
	Preprocessing	Sample	Order of Replicates	Session	Timing of Background	Order of Replicates × Session	Order of Replicates ×Timing of Background	Session × Timing of Background	Residuals
Effect (percentage contribution to the sum of squares)	None	Milled rice	**16.78**	**11.17**	**5.50**	7.77	**20.25**	**12.36**	26.18
		Red rice	10.72	2.32	**7.28**	16.82	19.81	**10.12**	32.93
		Granulated sugar	8.19	**8.50**	**35.87**	0.93	1.78	**39.01**	5.73
		Sugar lump	0.09	**13.62**	**69.73**	0.23	0.14	**15.68**	0.51
		Brown sugar A	**12.64**	**11.33**	**60.87**	0.85	1.14	**11.24**	1.94
		Brown sugar B	**12.21**	**11.66**	**60.63**	1.15	1.05	**11.03**	2.27
	MSC	Milled rice	**12.97**	**12.20**	**11.98**	10.42	11.50	**16.54**	24.40
		Red rice	8.03	**9.89**	**21.91**	9.88	12.21	**18.60**	19.49
		Granulated sugar	9.10	**6.90**	**32.35**	1.86	2.40	**40.83**	6.57
		Sugar lump	0.79	**17.39**	**44.93**	0.20	0.22	**35.95**	0.52
		Brown sugar A	**23.65**	**13.79**	**27.77**	1.49	2.02	**25.83**	5.45
		Brown sugar B	**22.07**	**14.11**	**28.64**	1.84	1.89	**24.78**	6.67
	SNV	Milled rice	**12.97**	**12.20**	**11.98**	10.42	11.50	**16.54**	24.40
		Red rice	8.03	**9.89**	**21.91**	9.88	12.21	**18.60**	19.49
		Granulated sugar	9.10	**6.90**	**32.35**	1.86	2.40	**40.83**	6.57
		Sugar lump	0.79	**17.39**	**44.93**	0.20	0.22	**35.95**	0.52
		Brown sugar A	**23.65**	**13.79**	**27.77**	1.49	2.02	**25.83**	5.45
		Brown sugar B	**22.07**	**14.11**	**28.64**	1.84	1.89	**24.78**	6.67

AvaSpec−Mini−NIR

Results for the AvaSpec−Mini−NIR data are summarized for raw and preprocessed data in Table 4. The significant effects are highlighted in bold.

At a first glance, the significant factors and interactions are the same regardless of acquisition configuration and preprocessing method. The slight differences in the significant effects for the acquisition configuration in raw data could be attributed to the higher intrinsic variability of data acquired with the fiber optic cable compared to the integrating sphere. Indeed, the spectrometer and detector were the same, but the light source and light path through the fiber or the sphere were substantially different, together with the portion of the sample exposed to the acquisition windows. As a general rule, the effect values obtained after the application of MSC or SNV were higher than for raw data and very similar but not identical within them. Interestingly, factors significant only for some samples when raw spectra were considered were found significant for all the samples under investigation when a scattering correction method was applied.

Drawing attention to the factor related to different sessions of analysis and the interaction with the timing of background, interesting aspects arise depending on the samples and configuration. The fact that milled rice spectra resulted in peculiar behavior (only few

factors influence the analysis) with integrating sphere configuration could reside in the intrinsic characteristics of the samples, which typically influence NIR acquisition: the grain size and the color.

The factor dealing with the acquisition of brown sugar spectra at the beginning or the end of a session and its interactions with the other factors were investigated. Results allowed for identification of a significant influence of the interaction *session $\times$ begin or end of the session* (effect 3.33 and p−value 0.02) in the spectra obtained with the optical fiber. Data registered with the integrating sphere did not show any significant effect according to the p−values related to the order of samples' acquisition.

Table 4. Summary of results for ASCA models of AvaSpec−Mini−NIR data. The effects are expressed as percentage of contribution to the sum of squares of the data. Models calculated on each sample. Factors that shown a p-value < 0.05 after 200 permutations are reported in bold. Brown sugar A indicates the spectra acquired at the beginning of each analytical session. Brown sugar B indicates the spectra acquired at the end of each analytical session. *None* is used to identify the default preprocessing.

| | | Optical Fiber | | | | | | | Integrating Sphere | | | | | | |
| | | Factors | | | Interactions | | | | Factors | | | Interactions | | | |
Preprocessing	Sample	Order of Replicates	Session	Timing of Background	Order of Replicates × Session	Order of Replicates × Timing of Background	Session × Timing of Background	Residuals	Order of Replicates	Session	Timing of Background	Order of Replicates × Session	Order of Replicates × Timing of Background	Session × Timing of Background	Residuals
None	Milled rice	14.92	4.41	3.10	30.50	13.22	3.67	30.18	14.00	1.52	0.43	30.34	16.67	1.43	35.61
	Red rice	17.14	0.64	**4.91**	23.04	21.88	2.78	29.61	13.61	1.87	**3.91**	34.07	11.97	4.85	29.72
	Granulated sugar	8.86	**4.39**	0.84	38.18	18.28	0.89	28.54	5.70	**7.61**	**21.64**	17.29	4.81	**25.93**	17.02
	Sugar lump	9.76	3.50	**7.86**	24.35	**22.14**	**18.17**	14.22	12.19	**14.50**	0.52	30.36	10.75	**10.38**	21.30
	Brown sugar A	19.08	**12.32**	**6.55**	17.97	13.25	3.37	27.45	5.27	**15.72**	**24.29**	8.77	4.07	**27.23**	14.65
	Brown sugar B	12.59	0.48	2.38	30.63	11.63	0.74	41.56	4.55	5.97	**10.77**	8.64	5.02	**55.13**	9.92
MSC	Milled rice	0.03	**42.79**	**14.27**	0.05	0.02	**42.80**	0.05	13.01	4.28	3.78	27.90	13.24	4.86	32.93
	Red rice	18.34	**6.45**	**3.52**	24.71	12.71	**4.31**	29.95	14.95	**9.26**	6.94	33.62	11.55	**4.54**	19.15
	Granulated sugar	9.84	**13.79**	3.39	25.61	13.41	**5.63**	28.32	4.87	**14.27**	27.11	13.51	4.38	**20.27**	15.58
	Sugar lump	11.21	**5.03**	**6.30**	28.56	10.67	**6.10**	32.13	8.49	**9.56**	10.94	13.45	6.35	**37.11**	14.10
	Brown sugar A	0.03	**42.80**	**14.27**	0.03	0.02	**42.82**	0.03	12.22	**7.06**	**5.89**	16.14	6.71	**24.35**	27.62
	Brown sugar B	11.21	**5.03**	**6.30**	28.56	10.67	**6.10**	32.13	8.49	**9.56**	10.94	13.45	6.35	**37.11**	14.10
SNV	Milled rice	0.17	**42.38**	**14.09**	0.37	0.12	**42.52**	0.35	13.00	4.29	3.79	27.89	13.23	4.87	32.94
	Red rice	18.31	**6.41**	**3.50**	24.78	12.74	**4.32**	29.95	14.93	**9.27**	6.95	33.61	11.55	**4.55**	19.15
	Granulated sugar	9.84	**13.80**	3.40	25.60	13.41	**5.63**	28.32	4.87	**14.27**	27.11	13.51	4.38	**20.27**	15.58
	Sugar lump	11.21	**5.05**	**6.31**	28.54	10.64	**6.12**	32.12	8.49	**9.56**	10.94	13.45	6.35	**37.11**	14.10
	Brown sugar A	0.15	**42.55**	**14.28**	0.13	0.10	**42.64**	0.15	12.22	**7.06**	**5.89**	16.14	6.71	**24.36**	27.62
	Brown sugar B	11.21	**5.05**	**6.31**	28.54	10.64	**6.12**	32.12	8.49	**9.56**	10.94	13.45	6.35	**37.11**	14.10

Effect (percentage contribution to the sum of squares)

NeoSpectra Scanner

Remarkable aspects were highlighted from the modelling of every single type of sample and the application of signal preprocessing (Table 5).

Table 5. Summary of results for ASCA models of Neospectra Scanner data. The effects are expressed as percentage of contribution to the sum of squares of the data. Models calculated on each sample. Factors that shown a *p*-value < 0.05 after 200 permutations are reported in bold. Brown sugar A indicates the spectra acquired at the beginning of each analytical session. Brown sugar B indicates the spectra acquired at the end of each analytical session. *None* is used to identify the default preprocessing.

| | Preprocessing | Sample | Factors | | | | Interactions | | | | | | Residuals |
			Order of Replicates	Session	Power Supply	Timing of Background	Order of Replicates × Session	Order of Replicates × Power Supply	Order of Replicates × Timing of Background	Session × Power Supply	Session × Timing of Background	Power Supply × Timing of Background	
Effect (percentage contribution to the sum of squares)	None	Milled rice	6.72	0.86	1.40	0.36	14.18	9.57	8.91	1.80	0.94	1.26	54.01
		Red rice	7.57	0.15	1.14	1.10	11.77	**14.68**	**12.49**	1.17	0.64	0.04	49.25
		Granulated sugar	4.98	**3.72**	**3.94**	**6.19**	9.10	3.06	7.09	2.44	1.91	**5.50**	52.06
		Sugar lump	0.51	**15.60**	**6.41**	**15.98**	0.64	0.21	0.22	**8.99**	**5.11**	**33.42**	12.92
		Brown sugar A	5.09	**6.38**	**5.59**	1.19	8.06	3.34	5.59	**5.66**	**9.91**	**9.83**	39.36
		Brown sugar B	5.38	**4.25**	**5.92**	**5.78**	10.09	3.57	6.20	**3.53**	**8.15**	2.51	44.62
	MSC	Milled rice	7.35	1.30	0.66	0.54	15.63	7.79	7.81	1.25	0.86	1.06	55.75
		Red rice	8.16	0.62	1.19	0.60	15.02	**15.38**	**11.42**	1.15	0.45	0.35	45.66
		Granulated sugar	2.45	**9.15**	**13.86**	**12.60**	4.01	1.63	2.43	**7.12**	**5.75**	**16.46**	24.54
		Sugar lump	2.91	**9.31**	**9.12**	**12.05**	1.38	0.55	0.49	**9.31**	**7.78**	**36.60**	10.51
		Brown sugar A	7.02	**4.20**	**4.55**	**5.51**	11.32	5.11	5.52	**2.93**	**5.57**	4.41	43.85
		Brown sugar B	5.91	**5.93**	**4.66**	**3.82**	9.22	4.70	5.57	**5.26**	**5.80**	4.33	44.80
	SNV	Milled rice	5.98	1.92	0.87	0.64	15.74	7.47	7.98	1.43	0.86	**2.13**	54.97
		Red rice	8.16	0.62	1.18	0.60	15.06	**15.26**	**11.36**	1.16	0.46	0.36	45.77
		Granulated sugar	2.45	**9.15**	**13.86**	**12.60**	4.01	1.63	2.43	**7.12**	**5.75**	**16.46**	24.54
		Sugar lump	2.91	**9.31**	**9.12**	**12.05**	1.38	0.55	0.49	**9.31**	**7.78**	**36.60**	10.51
		Brown sugar A	7.02	**4.21**	**4.55**	**5.51**	11.31	5.12	5.52	**2.94**	**5.57**	4.41	43.85
		Brown sugar B	5.91	**5.93**	**4.66**	**3.82**	9.23	4.70	5.57	**5.27**	**5.80**	4.33	44.79

Based on the analysis of raw data, session (Figure 4e,f), timing of background, and power supply are significant factors, as well as their interaction and the interaction with the timing of background. The discrepancy among residual values calculated in these models is consistent depending on the sample, indicating that other factors or higher−level interactions could be involved. The influencing factors in the spectra of this instrument were found substantially independent of scattering removal. The main difference when applying MSC (or SNV) resided in the magnitude of the effects for the significant factors and interactions.

The analysis carried out with the brown sugar spectra acquired at the beginning and the end of the session proved that there is not such a significant influence of the factor itself. In contrast, the interactions *session × timing of background*, *session × begin or end of the session*, and *timing of background × begin or end of the session* were significant (*p*−values—0.005, 0.025, and 0.005, respectively) with a small contribution (% sum of squares—8.80, 1.77, and 2.01, respectively).

NeoSpectra Development Kit

Looking into the results obtained for each sample (Table 6), it appears that disparate factors and interactions for different samples could influence the spectra.

Table 6. Summary of results for ASCA models of Neospectra development kit data. The effects are expressed as percentage of contribution to the sum of squares of the data. Models calculated on each sample. Factors that shown a *p*-value < 0.05 after 200 permutations are reported in bold. Brown sugar A indicates the spectra acquired at the beginning of each analytical session. Brown sugar B indicates the spectra acquired at the end of each analytical session. *None* is used to identify the default preprocessing.

			Factors			Interactions			
Preprocessing		Sample	Order of Replicates	Session	Timing of Background	Order of Replicates × Session	Order of Replicates × Timing of Background	Session × Timing of Background	Residuals
Effect (percentage contribution to the sum of squares)	None	Milled rice	22.06	3.11	0.81	22.38	23.23	0.37	28.04
		Red rice	15.97	2.45	0.99	32.20	12.90	4.30	31.17
		Granulated sugar	13.49	**12.12**	**16.76**	23.34	9.61	**4.01**	20.66
		Sugar lump	**18.09**	2.50	**21.56**	16.17	**19.74**	**5.87**	16.07
		Brown sugar A	8.64	**19.96**	**5.33**	21.40	10.44	**14.68**	19.56
		Brown sugar B	6.85	2.52	**31.25**	23.98	10.32	**3.94**	21.14
	MSC	Milled rice	13.83	0.44	1.58	33.10	20.02	1.05	29.96
		Red rice	11.73	2.35	0.12	37.79	15.61	3.73	28.68
		Granulated sugar	14.47	1.16	0.99	32.60	14.70	2.57	33.50
		Sugar lump	18.68	2.19	**2.66**	24.90	19.07	1.71	30.79
		Brown sugar A	10.90	**11.61**	1.45	29.32	11.97	**4.11**	30.62
		Brown sugar B	16.55	2.97	**4.79**	28.09	14.19	2.70	30.71
	SNV	Milled rice	15.53	0.64	0.91	31.18	21.26	1.02	29.46
		Red rice	16.19	3.01	0.97	**37.37**	15.72	2.90	23.84
		Granulated sugar	14.46	1.19	1.01	32.59	14.68	2.59	33.47
		Sugar lump	18.67	2.19	**2.70**	24.87	19.04	1.72	30.81
		Brown sugar A	10.88	**11.66**	1.49	29.34	11.99	**4.12**	30.52
		Brown sugar B	16.58	3.01	**4.80**	28.02	14.22	2.71	30.65

None of the factors under investigation were identified as significant from the permutations test for the spectra of rice samples. Despite that, the residuals are compatible with the ones of the previous instruments and with the other samples. The timing of background was found significant with a relatively different magnitude for the spectra of the sugar samples. The same is true for the interaction *session × timing of background*. Conversely to what was obtained for the other spectrometers, the application of a preprocessing method in this case generally reduced the effect of significant factors, strictly depending on the sample. The order of the sample in the session, investigated through the replicates of brown sugar, demonstrated that the interaction *session × begin or end of the session* was the only one significant for the spectra acquired by the NeoSpectra development kit.

These results enhanced the importance of a comprehensive study of the signals. In this case, the results of the ASCA models suggested that this spectrometer could be the best option among the tested ones. However, from a thoughtful evaluation of the characteristics

of the signal, the circumstances appeared different. First of all, the intensities of the spectra registered with the NeoSpectra development kit were considerably lower than the ones of the spectra acquired with other sensors (Supplementary Figure S7). The inspection of the S/N ratio and the relative standard deviation was carried out, and results were directly compared with those achieved for the NeoSpectra Scanner spectra. The reason for such comparison resides in the similarity of the sensors' conceptualization. The S/N was calculated as the ratio between the mean spectra for each sample and the standard deviation of the spectra of the same sample. The relative standard deviation was estimated as $100 \times$ standard deviation of spectra for each sample divided by the mean spectrum for the same sample. In general, the S/N was smaller with the relative standard deviation higher for the Neospectra development kit (Figure 5). Therefore, fewer factors and interactions were found to be significant in the ASCA models due to the low quality and lack of sensitivity of the sensor and not because of reduced influences in the spectra.
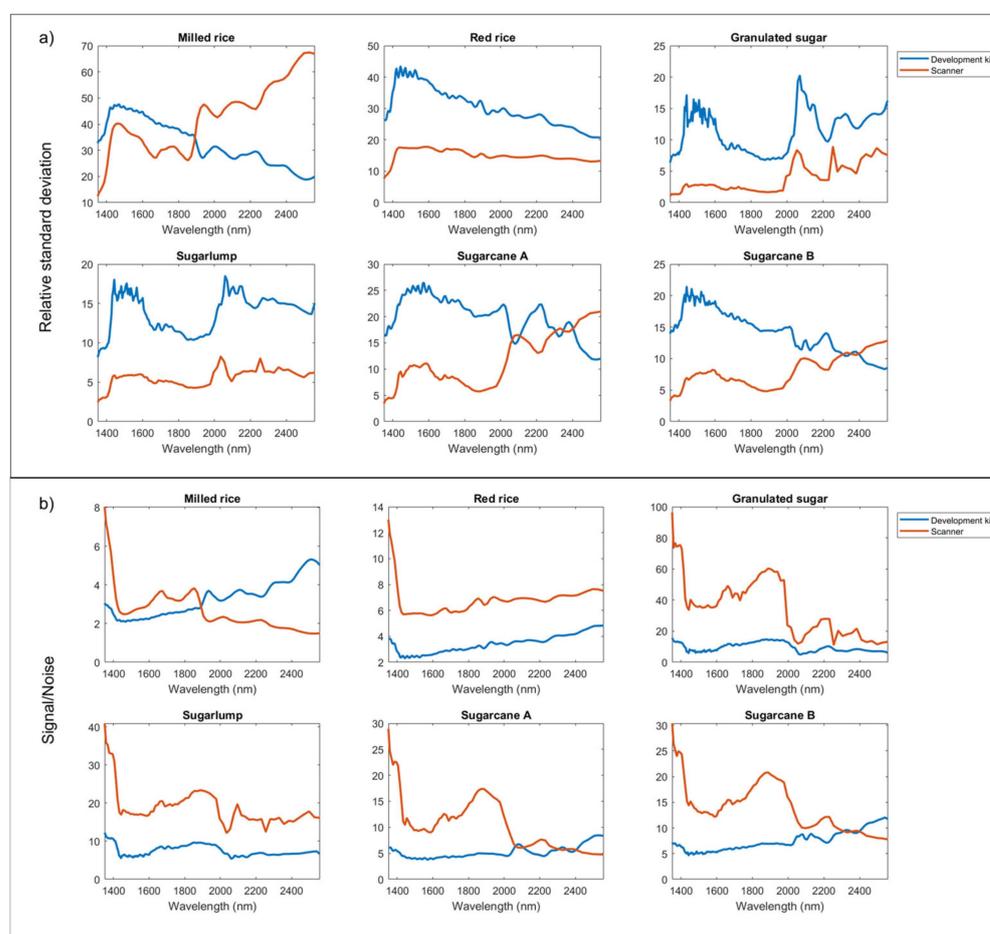
Figure 5. Relative standard deviation (**a**) and signal to noise ratio (**b**) for the spectra of all samples: results for spectra acquired with NeoSpectra scanner are in orange, while results for the spectra of NeoSpectra development kit are represented in blue.

3.4. Emerging Rules of Thumb

The feasibility and utility of miniaturized NIR devices has been proved in the literature in several applications [2], and the low cost of the sensors mainly led to their wide use in industries of several fields. In addition, the potentialities in gaining chemical and physical information have been widely proved. Despite this, it is quite common to underestimate the importance of the method set−up procedure in the analytical method and the effects to consider during the development of the acquisition strategy [1]. Under these circumstances, the results obtained through this study are meaningful for several

reasons. ASCA allowed the separation of the influences of different factors and interactions on the spectra. Moreover, an esteem of the magnitude of the effects was made available and interpretable through the loadings of each sub−model. Typical behaviors about the main effects studied could be stated for each spectrometer. The class of miniaturized spectrometers is proved to be plenty of differences between technologies and applicability. The outcome for spectrometers measuring the same spectral range could not be expected to be consistent without examining the specific application. Indeed, the intrinsic characteristics of each instrument and the features reflected in different experimental procedures make the influences in the spectra arduous to generalize. Even when the same spectrometer is used but with a different acquisition configuration, as in the case of AvaSpec−Mini−NIR, the variabilities in the data may change. In the same way, the factors to consider can vary depending on how a specific technology is assembled in the final spectrometer, as was proved for NeoSpectra sensors.

Some good practices may pop up from the insights obtained in this work. First, when considering a new application of a miniaturized spectrometer, a previous feasibility analysis has to be carried out for several factors. The method proposed could be used to identify the main factors for the specific case study. Within this frame, the optimization of the background timing was proved essential. Indeed, the acquisition of the background before each sample could be the optimal condition, but being time consuming, the identification of an optimal timing for the parameter is fundamental. Nonetheless, the evaluation of power supply conditions could be of interest when the spectrometer allows analysis standing in battery mode. In some cases, systematic error related to the order of acquisition was identified. Consequently, when more than one experimental replicated spectrum is needed, the influence of subsequent acquisitions is to be taken into consideration, as well as the order of samples under analysis. The general guidelines could reside in the random spectra acquisition for samples with similar characteristics.

With a view to developing predictive or classification models where the ordinary user will likely consider preprocessing methods, a deep investigation must be carried out [43] according to the emphasized influences found on the factors under study. The calibration set for the modelling phase should be constructed to include the sources of variability related to the instrument and its application [44,45] and should consider the reference errors. Similarly, it will most likely be necessary to include the same variabilities for reliable external validation. For instance, if the analysis session is a significant factor, a representative calibration set should include data from different sessions of analyses. Similarly, an actual external validation set could be composed of spectra acquired in different sessions. Other relevant implications could be highlighted for monitoring measurement systems and model maintenance (i.e., the calibration transfer strategies between different instruments and within the same instrument) [46–49].

4. Conclusions

Replicated spectra of samples with different physical and chemical characteristics and of worldwide interest were analyzed with five miniaturized NIR spectrometers covering different spectral ranges and based on different technologies. On one side, the factors that influence raw spectra are elucidated for each spectrometer, and the advantages and drawbacks of preprocessing are highlighted. The urgency for preliminary studies when dealing with a specific miniaturized spectrometer came out, and a guide for good practices was delineated to face the method set−up step. The identification of suitable acquisition strategies and systematic influences could allow the obtention of reliable results. On the other side, a method to investigate the sources of variability is presented. A way to identify the significant factors and interactions affecting spectral data is proposed. The background was proved to be one of the main factors to consider and to care about by the final user. Miniaturized NIR spectrometers were found to be very different to each other. The necessity for not treating them as a general class of instruments emerged due to significant technological differences in the characteristics and in the type of variation

associated to the spectra. Depending on the samples and their physical and chemical characteristics, the results could vary with each sensor. Results showed that preprocessing is not always needed and, depending on the sensor, enhances the influences in the spectra unrelated to the sample properties.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/chemosensors11030182/s1, Figure S1: Examples of spectrometer positioning during the analyses: (a) SCiO, (b) Viavi Micro-NIR, (c) AvaSpec with optical fibre, (d) AvaSpec with integrating sphere, (e) NeoSpectra Scanner, (f) NeoSpectra Development kit; Figure S2: Dataset structure scheme for SCiO; Figure S3: Dataset structure scheme for MicroNir OnSite-W; Figure S4: Dataset structure scheme for AvaSpec-Mini NIR; Figure S5: Dataset structure scheme for NeoSpectra Scanner; Figure S6: Dataset structure scheme for NeoSpectra development kit; Figure S7: Averaged spectra of each sample with the different spectrometers: (a) SCiO, (b) Viavi Micro-NIR, (c) AvaSpec with integrating sphere, (d) AvaSpec with optical fiber, (e) NeoSpectra Scanner, (f) NeoSpectra Development kit; Figure S8: PCA results on data preprocessed with MSC: Score plot for (a) SCiO, (b) Viavi Micro-NIR, (c) AvaSpec with integrating sphere, (d) AvaSpec with optical fibre, (e) NeoSpectra Scanner, (f) NeoSpectra Development kit; Figure S9: PCA results on data preprocessed with MSC: Loading plot for (a) SCiO, (b) Viavi Micro-NIR, (c) AvaSpec with integrating sphere, (d) AvaSpec with optical fiber, (e) NeoSpectra Scanner, (f) NeoSpectra Development kit; Figure S10: PCA results on data preprocessed with SNV: Score plot for (a) SCiO, (b) Viavi Micro-NIR, (c) AvaSpec with integrating sphere, (d) AvaSpec with optical fibre, (e) NeoSpectra Scanner, (f) NeoSpectra Development kit; Figure S11: PCA results on data preprocessed with SNV: Loading plot for (a) SCiO, (b) Viavi Micro-NIR, (c) AvaSpec with integrating sphere, (d) AvaSpec with optical fiber, (e) NeoSpectra Scanner, (f) NeoSpectra Development kit; Table S1: Resume of results for ASCA models of SCiO data. Effect in percentage which represent the contribution to the sum of squares of the data and cumulative eigenvalue. Models calculated on each type of sample according to the macro-chemical characteristics. In bold are reported factors that shown a *p*-value < 0.05 after 200 permutations. The factor "sample" included as sample levels respectively milled rice, brown rice, granulated sugar, sugar lump and only the replicate of the brown sugar acquired at the beginning of each session; Table S2: Resume of results for ASCA models of MicroNIR data. Effect in percentage which represent the contribution to the sum of squares of the data and cumulative eigenvalue. Models calculated on each type of sample according to the macro-chemical characteristics. In bold are reported factors that shown a *p*-value < 0.05 after 200 permutations. The factor "sample" included as sample levels respectively milled rice, brown rice, granulated sugar, sugar lump and only the replicate of the brown sugar acquired at the beginning of each session; Table S3: Resume of results for ASCA models of AvaSpec-mini-NIR data. Effect in percentage which represent the contribution to the sum of squares of the data and cumulative eigenvalue. Models calculated on each type of sample according to the macro-chemical characteristics. In bold are reported factors that shown a *p*-value < 0.05 after 200 permutations. The factor "sample" included as sample levels respectively milled rice, brown rice, granulated sugar, sugar lump and only the replicate of the brown sugar acquired at the beginning of each session; Table S4: Resume of results for ASCA models of NeoSpectra Scanner data. Effect in percentage which represent the contribution to the sum of squares of the data and cumulative eigenvalue. Models calculated on each type of sample according to the macro-chemical characteristics. In bold are reported factors that shown a *p*-value < 0.05 after 200 permutations. The factor "sample" included as sample levels respectively milled rice, brown rice, granulated sugar, sugar lump and only the replicate of the brown sugar acquired at the beginning of each session; Table S5: Resume of results for ASCA models of NeoSpectra development kit data. Effect in percentage which represent the contribution to the sum of squares of the data and cumulative eigenvalue. Models calculated on each type of sample according to the macro-chemical characteristics. In bold are reported factors that shown a *p*-value < 0.05 after 200 permutations. The factor "sample" included as sample levels respectively milled rice, brown rice, granulated sugar, sugar lump and only the replicate of the brown sugar acquired at the beginning of each session.

References

1. Giussani, B.; Gorla, G.; Riu, J. Analytical chemistry strategies in the use of miniaturised NIR Instruments: An overview. *Crit. Rev. Anal. Chem.* **2022**, 1–33. [CrossRef]
2. Beć, K.B.; Grabska, J.; Huck, C.W. Principles and applications of miniaturized Near-Infrared (NIR) spectrometers. *Chem.–A Eur. J.* **2020**, *27*, 1514–1532. [CrossRef]
3. Kademi, H.I.; Ulusoy, B.H.; Hecer, C. Applications of miniaturized and portable near infrared spectroscopy (NIRS) for inspection and control of meat and meat products. *Food Rev. Int.* **2018**, *35*, 201–220. [CrossRef]
4. Yakubu, H.G.; Kovacs, Z.; Toth, T.; Bazar, G. The recent advances of near-infrared spectroscopy in dairy production—A review. *Crit. Rev. Food Sci. Nutr.* **2020**, *62*, 810–831. [CrossRef]
5. Zhu, C.; Fu, X.; Zhang, J.; Qin, K.; Wu, C. Review of portable near infrared spectrometers: Current status and new techniques. *J. Near Infrared Spectrosc.* **2022**, *30*, 51–66. [CrossRef]
6. Beć, K.B.; Grabska, J.; Huck, C.W. Miniaturized NIR Spectroscopy in Food Analysis and Quality Control: Promises, Challenges, and Perspectives. *Foods* **2022**, *11*, 1465. [CrossRef]
7. Pasquini, C. Near infrared spectroscopy: A mature analytical technique with new perspectives–A review. *Anal. Chim. Acta* **2018**, *1026*, 8–36. [CrossRef]
8. Acosta, J.J.; Castillo, M.S.; Hodge, G.R. Comparison of benchtop and handheld near-infrared spectroscopy devices to determine forage nutritive value. *Crop Sci.* **2020**, *60*, 3410–3422. [CrossRef]
9. Mayr, S.; Beć, K.B.; Grabska, J.; Schneckenreiter, E.; Huck, C.W. Near-infrared spectroscopy in quality control of Piper nigrum: A comparison of performance of benchtop and handheld spectrometers. *Talanta* **2020**, *223*, 121809. [CrossRef]
10. Wang, Y.-J.; Li, T.-H.; Li, L.-Q.; Ning, J.-M.; Zhang, Z.-Z. Micro-NIR spectrometer for quality assessment of tea: Comparison of local and global models. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2020**, *237*, 118403. [CrossRef]
11. Yu, H.; Liu, H.; Wang, Q.; van Ruth, S. Evaluation of portable and benchtop NIR for classification of high oleic acid peanuts and fatty acid quantitation. *LWT* **2020**, *128*, 109398. [CrossRef]
12. Mishra, P.; Passos, D.; Marini, F.; Xu, J.; Amigo, J.M.; Gowen, A.A.; Jansen, J.J.; Biancolillo, A.; Roger, J.M.; Rutledge, D.N.; et al. Deep learning for near-infrared spectral data modelling: Hypes and benefits. *TrAC Trends Anal. Chem.* **2022**, *157*, 116804. [CrossRef]
13. Gorla, G.; Fumagalli, S.; Jansen, J.J.; Giussani, B. Acquisition strategies for fermentation processes with a low-cost miniaturized NIR-spectrometer from scratch: Issues and challenges. *Microchem. J.* **2022**, *183*, 108035. [CrossRef]
14. Giussani, B.; Escalante-Quiceno, A.T.; Boqué, R.; Riu, J. Measurement Strategies for the Classification of Edible Oils Using Low-Cost Miniaturised Portable NIR Instruments. *Foods* **2021**, *10*, 2856. [CrossRef]
15. Bertinetto, C.G.; Schoot, M.; Dingemans, M.; Meeuwsen, W.; Buydens, L.M.; Jansen, J.J. Influence of measurement procedure on the use of a handheld NIR spectrophotometer. *Food Res. Int.* **2022**, *161*, 111836. [CrossRef]
16. Cabassi, G.; Cavalli, D.; Fuccella, R.; Gallina, P.M. Evaluation of four NIR spectrometers in the analysis of cattle slurry. *Biosyst. Eng.* **2015**, *133*, 1–13. [CrossRef]
17. Riu, J.; Gorla, G.; Chakif, D.; Boqué, R.; Giussani, B. Rapid Analysis of Milk Using Low-Cost Pocket-Size NIR Spectrometers and Multivariate Analysis. *Foods* **2020**, *9*, 1090. [CrossRef]
18. Yang, S.; Zhao, Z.-N.; Yan, H.; Siesler, H.W. Fast detection of cotton content in silk/cotton textiles by handheld near-infrared spectroscopy: A performance comparison of four different instruments. *Text. Res. J.* **2022**, *92*, 2239–2246. [CrossRef]
19. Chai, J.; Zhang, K.; Xue, Y.; Liu, W.; Chen, T.; Lu, Y.; Zhao, G. Review of MEMS Based Fourier Transform Spectrometers. *Micromachines* **2020**, *11*, 214. [CrossRef]
20. Antila, J.; Tuohiniemi, M.; Rissanen, A.; Kantojärvi, U.; Lahti, M.; Viherkanto, K.; Kaarre, M.; Malinen, J. *MEMS- and MOEMS-Based Near-Infrared Spectrometers*; Wiley: Hoboken, NJ, USA, 2000; ISBN 9780470027318.
21. Strani, L.; Grassi, S.; Casiraghi, E.; Alamprese, C.; Marini, F. Milk Renneting: Study of Process Factor Influences by FT-NIR Spectroscopy and Chemometrics. *Food Bioprocess Technol.* **2019**, *12*, 954–963. [CrossRef]
22. Grassi, S.; Lyndgaard, C.B.; Rasmussen, M.A.; Amigo, J.M. Interval ANOVA simultaneous component analysis (i-ASCA) applied to spectroscopic data to study the effect of fundamental fermentation variables in beer fermentation metabolites. *Chemom. Intell. Lab. Syst.* **2017**, *163*, 86–93. [CrossRef]

23. De Luca, S.; De Filippis, M.; Bucci, R.; Magrì, A.D.; Magrì, A.L.; Marini, F. Characterization of the effects of different roasting conditions on coffee samples of different geographical origins by HPLC-DAD, NIR and chemometrics. *Microchem. J.* **2016**, *129*, 348–361. [CrossRef]

24. Firmani, P.; Vitale, R.; Ruckebusch, C.; Marini, F. ANOVA-Simultaneous Component analysis modelling of low-level-fused spectroscopic data: A food chemistry case-study. *Anal. Chim. Acta* **2020**, *1125*, 308–314. [CrossRef] [PubMed]

25. Fan, S.; Zhang, B.; Li, J.; Huang, W.; Wang, C. Effect of spectrum measurement position variation on the robustness of NIR spectroscopy models for soluble solids content of apple. *Biosyst. Eng.* **2016**, *143*, 9–19. [CrossRef]

26. Blanco, M.; Villarroya, I. NIR spectroscopy: A rapid-response analytical tool. *TrAC Trends Anal. Chem.* **2002**, *21*, 240–250. [CrossRef]

27. Wold, S.; Esbensen, K.; Geladi, P. Principal Component Analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37–52. [CrossRef]

28. 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]

29. Oliveri, P.; Malegori, C.; Simonetti, R.; Casale, M. The impact of signal pre-processing on the final interpretation of analytical outcomes–A tutorial. *Anal. Chim. Acta* **2018**, *1058*, 9–17. [CrossRef]

30. Smilde, A.K.; Hoefsloot, H.C.; Westerhuis, J.A. The geometry of ASCA. *J. Chemom.* **2008**, *22*, 464–471. [CrossRef]

31. Bertinetto, C.; Engel, J.; Jansen, J. ANOVA simultaneous component analysis: A tutorial review. *Anal. Chim. Acta X* **2020**, *6*, 100061. [CrossRef]

32. Nieuwoudt, M.K.; Giglio, C.; Marini, F.; Scott, G.; Holroyd, S.E. Routine Monitoring of Instrument Stability in a Milk Testing Laboratory With ASCA: A Pilot Study. *Front. Chem.* **2021**, *9*, 733331. [CrossRef] [PubMed]

33. Schoot, M.; Alewijn, M.; Weesepoel, Y.; Mueller-Maatsch, J.; Kapper, C.; Postma, G.; Buydens, L.; Jansen, J. Predicting the performance of handheld near-infrared photonic sensors from a master benchtop device. *Anal. Chim. Acta* **2022**, *1203*, 339707. [CrossRef]

34. Ouyang, Q.; Liu, L.; Wang, L.; Zareef, M.; Wang, Z.; Li, H.; Yin, J.; Chen, Q. Rapid determination of free amino acids and caffeine in matcha using near-infrared spectroscopy: A comparison of portable and benchtop systems. *J. Food Compos. Anal.* **2022**, *115*, 104868. [CrossRef]

35. Catelani, T.A.; Santos, J.R.; Páscoa, R.N.; Pezza, L.; Pezza, H.R.; Lopes, J.A. Real-time monitoring of a coffee roasting process with near infrared spectroscopy using multivariate statistical analysis: A feasibility study. *Talanta* **2018**, *179*, 292–299. [CrossRef] [PubMed]

36. Ferrara, G.; Melle, A.; Marcotuli, V.; Botturi, D.; Fawole, O.A.; Mazzeo, A. The prediction of ripening parameters in Primitivo wine grape cultivar using a portable NIR device. *J. Food Compos. Anal.* **2022**, *114*, 104836. [CrossRef]

37. McVey, C.; Gordon, U.; Haughey, S.; Elliott, C. Assessment of the Analytical Performance of Three Near-Infrared Spectroscopy Instruments (Benchtop, Handheld and Portable) through the Investigation of Coriander Seed Authenticity. *Foods* **2021**, *10*, 956. [CrossRef]

38. Kirchler, C.G.; Pezzei, C.K.; Beć, K.B.; Mayr, S.; Ishigaki, M.; Ozaki, Y.; Huck, C.W. Critical evaluation of spectral information of benchtop vs. portable near-infrared spectrometers: Quantum chemistry and two-dimensional correlation spectroscopy for a better understanding of PLS regression models of the rosmarinic acid content in Rosmarini folium. *Analyst* **2016**, *142*, 455–464. [CrossRef]

39. Berzaghi, P.; Cherney, J.; Casler, M. Prediction performance of portable near infrared reflectance instruments using preprocessed dried, ground forage samples. *Comput. Electron. Agric.* **2021**, *182*, 106013. [CrossRef]

40. Gorla, G.; Taborelli, P.; Alamprese, C.; Grassi, S.; Giussani, B. On the Importance of Investigating Data Structure in Miniaturized NIR Spectroscopy Measurements of Food: The Case Study of Sugar. *Foods* **2023**, *12*, 493. [CrossRef]

41. Riu, J.; Vega, A.; Boqué, R.; Giussani, B. Exploring the Analytical Complexities in Insect Powder Analysis Using Miniaturized NIR Spectroscopy. *Foods* **2022**, *11*, 3524. [CrossRef]

42. Ryckewaert, M.; Gorretta, N.; Henriot, F.; Marini, F.; Roger, J.-M. Reduction of repeatability error for analysis of variance-Simultaneous Component Analysis (REP-ASCA): Application to NIR spectroscopy on coffee sample. *Anal. Chim. Acta* **2019**, *1101*, 23–31. [CrossRef]

43. Engel, J.; Gerretzen, J.; Szymańska, E.; Jansen, J.J.; Downey, G.; Blanchet, L.; Buydens, L.M. Breaking with trends in pre-processing? *TrAC Trends Anal. Chem.* **2013**, *50*, 96–106. [CrossRef]

44. Borer, M.W.; Zhou, X.; Hays, D.M.; Hofer, J.D.; White, K.C. Evaluation of key sources of variability in the measurement of pharmaceutical drug products by near infrared reflectance spectroscopy. *J. Pharm. Biomed. Anal.* **1998**, *17*, 641–650. [CrossRef] [PubMed]

45. Xu, X.; Xie, L.; Ying, Y. Factors influencing near infrared spectroscopy analysis of agro-products: A review. *Front. Agric. Sci. Eng.* **2019**, *6*, 105. [CrossRef]

46. Folch-Fortuny, A.; Vitale, R.; de Noord, O.E.; Ferrer, A. Calibration transfer between NIR spectrometers: New proposals and a comparative study. *J. Chemom.* **2017**, *31*, e2874. [CrossRef]

47. Pereira, L.S.; Carneiro, M.F.; Botelho, B.G.; Sena, M.M. Calibration transfer from powder mixtures to intact tablets: A new use in pharmaceutical analysis for a known tool. *Talanta* **2016**, *147*, 351–357. [CrossRef]

48. Workman, J.J. A Review of Calibration Transfer Practices and Instrument Differences in Spectroscopy. *Appl. Spectrosc.* **2017**, *72*, 340–365. [CrossRef] [PubMed]
49. Eady, M.; Payne, M.; Changpim, C.; Jinnah, M.; Sortijas, S.; Jenkins, D. Establishment of instrument operation qualification and routine performance qualification procedures for handheld near-infrared spectrometers used at different locations within a laboratory network. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2021**, *267*, 120512. [CrossRef] [PubMed]