

Review

The Sample, the Spectra and the Maths—The Critical Pillars in the Development of Robust and Sound Applications of Vibrational Spectroscopy

Daniel Cozzolino ^{1,2} 

¹ Centre for Nutrition and Food Sciences, Queensland Alliance for Agriculture and Food Innovation (QAAFI), The University of Queensland, Brisbane, Queensland 4072, Australia; d.cozzolino@uq.edu.au

² ARC Training Centre for Uniquely Australian Foods, Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Block 10, Level 1, 39 Kessels Rd, Coopers Plains Qld 4108, Australia

Received: 10 July 2020; Accepted: 7 August 2020; Published: 12 August 2020



Abstract: The last two decades have witnessed an increasing interest in the use of the so-called rapid analytical methods or high throughput techniques. Most of these applications reported the use of vibrational spectroscopy methods (near infrared (NIR), mid infrared (MIR), and Raman) in a wide range of samples (e.g., food ingredients and natural products). In these applications, the analytical method is integrated with a wide range of multivariate data analysis (MVA) techniques (e.g., pattern recognition, modelling techniques, calibration, etc.) to develop the target application. The availability of modern and inexpensive instrumentation together with the access to easy to use software is determining a steady growth in the number of uses of these technologies. This paper underlines and briefly discusses the three critical pillars—the sample (e.g., sampling, variability, etc.), the spectra and the mathematics (e.g., algorithms, pre-processing, data interpretation, etc.)—that support the development and implementation of vibrational spectroscopy applications.

Keywords: vibrational spectroscopy; sampling; multivariate data analysis; error; validation

1. Introduction

The last two decades have witnessed an increasing interest in the use of the so-called rapid analytical or high throughput techniques [1–8]. Most of these applications are based on the use of vibrational spectroscopy methods (near infrared (NIR), mid infrared (MIR), and Raman, visible (VIS)) in a wide range of samples (e.g., food ingredients, natural products, crops, animal and plant tissues, medical and pharmaceutical applications; etc.) [9–13]. The number of references using words such as “infrared”, “NIR”, “Raman”, “MIR”, “hyperspectral”, “green analytical methods”, “chemometrics” and “multivariate data analysis” [9–17], are evidence of this steady increase in the number of applications of these analytical methods.

In recent years, vibrational spectroscopy has been also considered for its potential as a high throughput phenotyping tool in both animals and plants, where novel applications related with plant breeding and selection, plant nutrition and physiology have been reported in the last 20 years [9–15]. More recently, vibrational spectroscopy (e.g., NIR, MIR, Raman and hyperspectral imaging systems) techniques have shown their ability to qualitatively (e.g., classifying, identifying, and monitoring) analyse several types of samples (e.g., wine, meat, coffee, condiments, etc.), targeting issues related with origin, traceability, and provenance of foods and food ingredients [9–17]. Concomitantly, recent developments in hardware (e.g., image techniques, optical sensors, handheld instrumentation, etc.) are adding new analytical possibilities to the potential users of these technologies, making them very

attractive to be implemented in the whole food value chain (e.g., the addition and use of objective tools in blockchain and food traceability) [18–20].

Another field where vibrational spectroscopy demonstrated to have a great impact is in the so-called process analytical technologies (PAT) [21–24]. This approach has not only been used to collect chemical information about the process (e.g., spatial and temporal information) to monitor the composition of the product, but also to provide information about the process itself, such as yield, energy input, faults and quality assurance [21–24]. The implementation of vibrational spectroscopy based on the utilization of different type of sensors has provided a platform for process data analysis and process sensor technology [21–24]. The data collected by the sensor could be also utilised to provide useful information about other aspects of the process, such as occupational safety, sustainable protection of employees, plant safety, hazardous operating conditions, and to assure environmental protection, providing feedback about the conditions of the industry [21–24]. The incorporation of these technologies and the development of applications of PAT has increased the demands for a knowledge-based approach [21–24]. According to the researchers in the field, the integration of vibrational spectroscopy and other sensing techniques with multivariate data methods and techniques (MVA) caused PAT to boost the multidisciplinary approach within the industry and research, where the design of state-of-the-art sensors with high specificity and resolution have improved the amount of data collected and therefore the information in order to manage the data generated by these approaches [21–24].

This approach is not entirely strange to the industry where applications of these techniques also attracted an increase in interest from the pharmaceutical, food and beverage industries, etc. [9–17]. The main reasons for the increasing use of this approach might be due to the main advantages that these methods and techniques possess when compared with other routine analytical techniques or methods, such as the non-destructive nature of these technologies, minimal or no sample preparation, no chemical reagents required, easy and ready to use instrumentation, and availability of inexpensive and portable devices [9–20].

One of the main analytical advantages of rapid analytical methods or high throughput techniques is that they can evaluate/measure the biochemical and/or chemical characteristics of any given organic compound [9–20]. This might be possible as chemical bonds present in the sample vibrate at specific frequencies or wavelengths depending of their mass of the constituent atoms, the shape of the molecule, and the stiffness of the bonds [9–20]. Therefore, the amount and the frequency of the absorbed light and the total reflected or transmitted light can be used to infer the chemical composition of the sample. The chemical and/or physical information derived from the use of vibrational spectroscopic methods resides in the manifestation of peaks, band positions, intensities, and shapes [9–20].

In modern routine chemical analysis, scientist are often confronted with so much data that the essential information may be not readily evident [11,25–31]. This is the case with spectral data for which many different observations (peaks or wavelengths) have been collected during the analysis of the sample. Each different measurement can be thought of as a different dimension [11,25–31]. Therefore, the success of the application will be highly dependent on the integration with the most appropriate multivariate data analysis (MVA) method, such as pattern recognition and modelling techniques, to develop the target application [11,25–31].

The advances and developments in modern analytical instrumentation and, in particular, those observed in vibrational spectroscopy, have determined the increasing growth in the so called high-dimensional data, where both the number of measured variables and samples can be high, together with the high variety of data (e.g., multiple data sources are available) and high speed during the collection of the data [32–34]. Thus, the increasing use of vibrational spectroscopy has determined an increase in the volume, variety and velocity of data collected determining the so-called “big data” [32–34]. The generation and use of big data becomes the reality in the routine life of analytical chemists and every researcher [32–34]. Contradictory, although the time dedicated to analysing a

single sample using vibrational spectroscopy has been reduced, the time dedicated to interpreting and mining the data has exponentially increased, depending on the dataset [32–34].

Classical statistics are not able to handle the current increase in the volume of data generated with this approaches. In this context, the scope of MVA is wide where its applications are found in many fields and where the number of the so-called toolboxes or methods is diverse [11,25–31]. The integration of MVA into vibrational spectroscopy provides the means to move the analysis beyond the one-dimensional (univariate) space, revealing constituents or properties that are important through the various interferences and interactions in the matrix [11,25–31]. Today, many modern instrumental measurement techniques are multivariate and based on indirect measurements of the chemical and physical properties of the sample [11,25–31]. Figure 1 shows the theoretical and practical links between the sample, the method or technique and the mathematics during the development of an application.

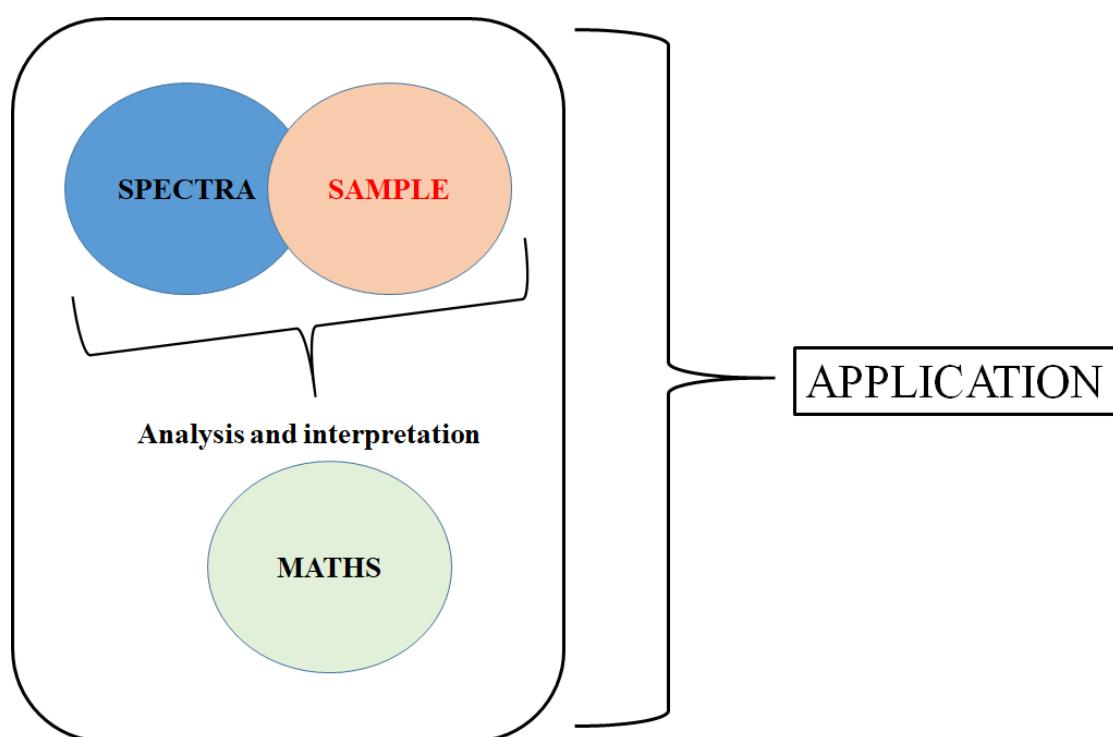


Figure 1. The integration or link between the sample, the method or technique and the mathematics during the development of applications based on vibrational spectroscopy.

Beyond the many advantages that the integration of vibrational spectroscopy with MVA offer, the ability of providing a holistic view of the system or sample analysed (e.g., fingerprint analysis) determines that these approaches are advantageous when compared with other analytical methods. In addition, the availability of modern and inexpensive instrumentation together with access to easy-to-use software is determining a steady growth in the number of applications of these technologies. Please note that this paper does not intend to be “another” review of multivariate data analysis and/or vibrational spectroscopy. The reader can find several excellent dedicated reviews already published in the scientific literature. Instead, the intention is to discuss and provide a guide of the main issues that can affect the successful implementation of these approaches.

Therefore, this paper underline and briefly discussed the three critical pillars—the sample (e.g., sampling, variability, etc.), the spectra and the mathematics (e.g., algorithms, pre-processing, data interpretation, etc.)—that support developments and implementations of vibrational spectroscopy applications.

2. The Source of Information—The Experiment and the Sample

2.1. The Theory of Sampling and Uncertainty

Regardless of all the care taken during sampling, the sample always differs in composition from the target intended [35–38]. Even the use of randomly replicated samples from the same target will differ among themselves, determining the so-called sampling uncertainty [35–38]. Understanding the uncertainty derived from both the sampling and the analysis will allow making rational decisions about a given process, classification or calibration results [35–38]. It is worth noting that the final application will be connected to making decisions about the target instead of about the sample [35–38].

Different authors have highlighted that one of the most important issues to be considered during sampling is related to how good the uncertainty depending on the purpose is [35–38]. One important issue to consider (and remember) is that the uncertainty of the measurement that arose from sampling is non-negligible [35–39]. This is even more significant when raw materials (e.g., food ingredients) and environmental samples (e.g., soil and water) are collected, where the uncertainty of the sampling exceeds the analytical contribution [35–39]. Therefore, the theory of sampling becomes highly relevant during the development of a given applications.

The theory of sampling (TOS) documents and details in a comprehensive means all aspects of the mechanical structure and chemical variation within a target in relation to the procedure for obtaining a primary sample from it [35–39]. Some of the main issues considered in the TOS are associated with the characteristics and/or properties of the target, including the size range of the particles comprising the target, the shapes of the particles, the compositional variation of the particles and the degree and style of the heterogeneity of the target, among others [35–39]. The method of collecting or extracting the sample and the degree of comminution/homogenisation/grinding at the different steps during the sampling process are important aspects included in the TOS [35–39]. All of these previously summarised issues and properties contributed to identifying the types of “error” of a given analysis or process [35–39].

The different sources and types of “errors” should be eliminated, and attention to detail will define the procedure or sampling protocol that will deliver the “correct” sample [35–39]. Researchers and practitioners in the field state that the interpretation of “correct” refers to “unbiased”, where sampling bias is avoided in the definition [35–39].

During the application of the TOS, it has been reported that sampling uncertainty is ignored and only the analytical uncertainty is considered [35–39]. The scientific literature in the field also suggested that the heterogeneity in the population and the ways of counteracting its adverse influence due to sampling/signal acquisition, sub-sampling and sample preparation/presentation processes, must be considered and evaluated before analysis [35–39].

In summary, the TOS is the main framework that must be used as a guide during meta-analysis of any application using vibrational spectroscopy [35–39]. It has been highlighted that the TOS emphasises the fundamental sampling principle (FSP), which states that all potential units from an original material must have an equal probability of being sampled in practice, and that samples are not altered in any way after sampling [35–39]. In the context of model development (e.g., calibration/validation and prediction), the main interactions between the sampling and the analysis (e.g., physical sampling), or the sampling and the on-line application, must be evaluated and understood in order to avoid inaccuracies and mistakes [35–39].

2.2. Samples

In any given application of vibrational spectroscopy, the sample itself plays an important role in defining the success of such application. However, the importance of both the sampling and the sample are usually overlooked. Two of the main characteristics or properties that define the success of a given application using vibrational spectroscopy are associated with both the perturbation and the observation of the sample [39–46]. The perturbation is usually associated with the experimental conditions used to develop the application (e.g., dry vs. wet sample, temperature, whole vs. powder,

etc.) while the observations/samples are associated with the sampling protocol and the property to be measured (e.g., limit of detection, range in concentration, standard error of the laboratory, number of samples etc.) [39–46].

2.3. Sample Properties and Pre-Processing

Preparing, pre-processing (e.g., grinding and homogenisation) and selecting the samples to be incorporated into the application is not a trivial task [35–39]. During the process of preparing and selecting samples for analysis, several inconsistencies or errors can be added into the overall error of the method (e.g., multiplicative effects) [39–46]. For example, different pre-processing steps, such as drying and grinding of the sample, can contribute to significantly exacerbating the analytical error [39–46]. This kind of interaction between the perturbation and the observation can be observed in most of the applications using analytical methods, and they will define the success or lack thereof of the application based on the systematic error [39–46].

2.4. Sample Variability

Probably one of the main questions asked during the development of the application is associated with the selection of the most suitable sample to be used during calibration development [47]. It has been agreed by several researchers that samples used to build a given calibration model have to be selected from samples similar to those that will be analysed in the future [39–48]. In addition, the samples have to be exposed to the same pre-processing and handling steps adopted, and this should be maintained when future samples are incorporated into the calibration. Samples used in calibration must be sourced from a wide-range composition, or at least considering the expected range of the composition [39–47]. All sources of possible variation to be encountered in the future must be considered and/or incorporated into the sample set [39–48]. If samples are used to represent a process all potential variations in the system, factors such as temperature, changes in particle size, physical changes in the sample, and equipment should be incorporated [39–47]. When dealing with biological materials (e.g., plants, animal muscle or tissues), other variations must be evaluated, such as harvest time and type of tissue (e.g., type of muscle), among others [39–48].

However, the selection of samples is not an arbitrary task and demands care. For example, during calibration development, the aim is to obtain homogenous and representative samples well distributed along the dataset. If there are too many samples available, it is recommendable to choose samples in order to develop a well balance dataset. Although randomisation is the preferred method to select samples to be included into the calibration, a better approach will be the utilization of robust techniques based in either Mahalanobis and Euclidean distances or the Kennard–Stone algorithm [49,50]. Recently, the use kernel distances have been reported as a robust method to objectively select samples [49,50].

3. Collecting the Information—The Spectra

A wide range of analytical methods and techniques based on vibrational spectroscopy are available in the market nowadays (e.g., NIR, MIR, Raman, lab bench and handheld instrumentation, hyperspectral imaging etc.) [51–53]. All of these techniques have in common the fact that they generate large amounts of data. Munck and collaborators stated that most instruments based on vibrational spectroscopy are extensively used a black box devices for the estimation of chemical compositions based on calibrations [51–53]. Very few scientist are aware that black box technology can be expanded for the physical–chemical characterisation of spectra [51–53]. Please note that it is not the objective of this paper to provide a comprehensive and detailed description of the different vibrational methods used as rapid or high throughput methods [54–64]. More detailed information about the different methods and techniques available as well the different technical characteristics or properties of the commercial instrumentation available in the market can be found elsewhere [54–64].

The selection of the most appropriate measurement technique or sampling mode/method is also of importance. For example, the analysis of whole or powder samples (e.g., grains and forages) presents

a much greater challenge than liquids (e.g., milk, wine, juice, etc.) when vibrational spectroscopy methods are used (e.g., NIR and ATR-MIR), since the measurements are generally made in the reflectance mode [39–47]. This is because reflectance measurements have lower energy collection efficiency than transmission measurements [39–47]. In addition, when using NIR reflectance measurements, light scattering efficiency is higher in the long wavelength region than at shorter wavelengths, which helps to offset the less efficient light collection [39–47]. Most of the applications of reflectance use the scattered light or energy from the sample, and they are used in the collection of NIR spectra [39–47]. The spectral characteristics of the sample can be also dramatically altered due to the particle size. Other properties that can have a large influence on the spectra might be related to suspended particles (e.g., fruit homogenates); the shape, size and orientation of particles in powders; and the sample thickness [39–47].

4. Analysing and Interpreting the Information—The Maths

The use of vibrational spectroscopy generates large amounts of data, allowing for the simultaneous analysis/measurement of several parameters, which provides a rapid and non-destructive quantification of major components in many organic substances [65–78]. The integration of vibrational spectroscopy methods with MVA has been the key for the success of the application of these techniques in many fields [65–78].

It has been stated (and sometimes is the belief by some of the users of MVA) that if the data already contain information, then any MVA method will succeed [35–39]. Unfortunately, the data are not as clean as expected when sampling and instrument noise and typing mistakes, among others have a greater impact where the use or pre-processing or any other correction does not improve the accuracy of the analytical results (e.g., inaccuracies can never be modelled) [35–39]. Therefore, a word of caution: MVA is not a “black box” or “push button” approach where the modelling will automatically do the rest [35–39].

4.1. Data Pre-Processing

Before starting with the analysis, interpretation and model developing, data pre-processing is a critical stage, as it affects the performance of the algorithms used and therefore the results (e.g., calibration and classification) [79–83]. Different methods and/or techniques for data pre-processing have been applied and developed specifically to different types of data and experimental designs [79–83]. For example, pre-processing of the spectra using the first and second derivatives, smoothing, multiple scatter correction (MSC), standard normal variate (SNV) and other normalization techniques were reported in most of the applications using vibrational spectroscopy [79–83]. Details about these pre-processing methods and techniques can be found in reviews by other authors [79–83].

4.2. Mistakes and Error during Analysis and Interpretation of the Data

The analysis of large-scale data is a challenging task (e.g., big data), not so much because the amount of data is large, but because large-scale measurement technologies possess high inherent variability [81–84]. Sources of this variability contribute to defining the systematic errors (bias) and the so-called stochastic effects (noise) [81–86]. Systematic effects influence all measurements in a similar manner [81–86]. Therefore, they can be eliminated or reduced using different data normalisation or pre-processing techniques [81–86]. However, stochastic effects cannot be corrected by pre-processing, but can be quantified, in particular by the application of repeated measurements (replicates) [81–86]. Depending on the modelling approach, further data manipulations might be necessary [81–86]. Ultimately, pre-processing techniques used to remove any irrelevant information that cannot be handled by the regression techniques [81–86]. Several pre-processing methods have been proposed and developed for this purpose and several references can be found elsewhere [81–86].

4.3. Algorithms Used to Develop Models

The use of MVA, unlike classic statistics, can also allow for the simultaneous analysis of multiple variables and takes collinearity into account (the variation in one variable, or group of variables, in terms of co-variation with other variables) [65–78,87–95]. The analysis can mathematically describe

the co-variation (degree of association) between variables, or find a mathematical function (regression model) that calculates the values of the dependent variables from values of the measured (independent) variables [65–78,87–95].

The most commonly used data analysis algorithms for performing regression include partial least squares regression (PLS) and principal component regression (PCR) [65–78,87–95]. These regression methods are designed to avoid issues associated with noise and correlations (collinearity) in the data [65–78,87–95]. PLS has brought into the field an online analysis of plants for a variety of quality attributes [65–78,87–95]. Besides PLS and PCR regression, other multivariate data analysis methods have been applied either to explore datasets or to build calibration models, where principal component analysis (PCA); cluster analysis (CA); linear discriminant analysis (LDA) [65–78,87–96]; machine learning approaches, such as support vector machines, classification and regression (SVM) [97–101], artificial neural networks (ANN) and other non-linear techniques [102–105]; and pattern recognition methods are just few examples [63–76,85–105].

4.4. Validation

In practice, several applications of vibrational spectroscopy available in the scientific literature report the use of cross-validation as the main technique used to test the models [105–121]. One of the most important steps during the implementation of a calibration into a real-life situation is the process of verification, validation and required testing [105–115]. What appears to have improved in the last decades is the capability to manage the quality control, equation updates, and data analysis [105–115,122,123]. As mentioned above, in order to assess the accuracy of a calibration model and to avoid overfitting, validation procedures have to be applied; a calibration model without validation is nonsense [105–115]. Although in feasibility studies cross-validation can be the best practical method to demonstrate that a model can predict the measured property, the actual accuracy must be estimated with an appropriate test set or validation set [105–115]. For feasibility studies, different cross-validation techniques can be used [105–124]. For example, in leave-one-out cross-validation, one sample is removed from the dataset, and a calibration model is constructed for the remaining subset [104–114]. The removed samples are then utilised to calculate the prediction residual [105–115]. The process is repeated with other subsets until every sample has been left out once, and in the end, the variance of all prediction residuals is estimated. In multifold cross-validation, a well-defined number of samples ('segment') is left out instead of one [105–115]. In internal validation, the dataset is split into a calibration set and a validation set. Calibration models are determined to be robust when the prediction accuracy is relatively insensitive towards unknown changes of external factors [105–115,122–124].

A good validation method should include a dataset of a completely excluded set of samples (not included in cross-validation) sourced from a separate sample regime with separate analysis. An independent testing of the calibration models on an excluded validation set eludes several of the most frequent mistakes in MVA, such as model overfitting [105–115,120–124].

Validation of classification models (e.g., discrimination) derived from the application of hyperspectral imaging have their own challenges [105,117]. A recent tutorial revised the different validation methods used in hyperspectral imaging analysis [105,117]. One of the main issues encountered is related with the samples used to develop the models. If too many samples are used (e.g., oversampling), unconstrained bootstrap and k-fold cross-validation might yield inaccurate results, failing to provide a realistic estimate of the predictive performance of the model [105,117]. Factors that can have a large influence during the analysis might be related to the range of data points (e.g., wavenumbers) used, the size of the image, the distribution of pixels from the different classes in the image and the number of pixels included in the training set [117]. The authors of the tutorial indicated that better results were obtained when randomised samples were used to develop the calibration and validation datasets [117].

The development of discriminant models utilising image data acquired from a single sample is highly risky, as the models might not take into consideration the effect of several sources inducing variation in the IR signal (e.g., age, body mass index, collection dates, sample storage or instrument

performance) [105,114]. Therefore, validation using an external validation set is necessary in order to avoid overoptimistic results [105,116,117]. Other validation methods have been proposed during the integration of discriminant approaches to hyperspectral image analysis [116,122–124]. A summary of these applications can be found in a review by Guaita and collaborators [116,122–124].

4.5. Data Interpretation

One of the main issues is the comparison of results from the literature is usually complicated by variations in the population size and structure with respect to the attribute of interest. It is therefore critical to report the standard deviation (SD) of the population for the attribute of interest [28,40,41,46,48,78,109,110,113]. In general, a range of statistics is required to be reported in order to compare different calibrations, including the coefficient of correlation (R), root mean square for the standard error in cross-validation (RMSECV), standard error of prediction (SEP), SD, the number of samples used, the number of outliers removed, and the number of principal components [28,40,41,46,48,78,109,110,113,116]. The report of marginal gains in the standard of cross-validation or prediction after the use of several pre-processing methods should be avoided. The same can be applied when different algorithms are used with no real improvements in the predictive ability of the models. A summary of the main statistics to be considered during calibration interpretation and reporting can be found in the report by Williams and collaborators [112].

Calibration models are often evaluated and/or reported using a combination of some of the statistics presented above. However, the sole interpretation and evaluation of statistics is not enough, and the loadings or coefficients of regression must be interpreted in the context of the property or the measured chemical analyte [28,40,41,46,50,78,109,112,113]. For example, if a calibration was developed to measure or predict protein, it is expected that wavelengths or frequencies that contain information about the N–H bonds will be prevalent. In real-life applications of vibrational spectroscopy, the calibration or model must be judged or considered in relation to their fit-for-purpose criterion [28,40,41,46,50,78,109,112,113].

5. Outliers, Overfitting and Underfitting

Typing errors; file transfer; interface errors; sensor malfunctions; and fouling, bad or incorrect sampling or sample presentation of the instrument, among other factors [117–124], may induce outliers. Samples can be considered as outliers according to the spectra, reference, or both [117–124]. During calibration development, outliers can be visualised using a principal component (PCA) scores plot [117–124]. Outlier samples should be kept during the initial steps of calibration until further investigation into their origin, and only the sample outliers that have an effect on the regression model are to be removed [117–124]. In any case, excessive pruning of the dataset for outliers should be avoided [117–124].

During the application of any of the MVA techniques presented above, it is important to select the appropriate number of components or latent variables (optimization) [117–120]. In this process, there is a delicate balance: if too many are used, there is too much redundancy in the independent variables used during the development of the model, causing the model to become overfitted [117–120]. In this case, the calibration model will be very dependent on the dataset and might provide poor prediction results [117–124]. On the other hand, using too few components will cause underfitting and the model will not be large enough to capture the variability in the data [117–120]. This “fitting” effect is strongly dependent on the number of samples used to develop the model and, in general, more samples give rise to more accurate predictions [117–124].

Overall, the use of MVA has the risk of overfitting (over-parameterization) determining a potential increase in the risk of false discovery [121]. Overfitting can be reduced during exploratory applications of vibrational spectroscopy by the use of rank optimization (e.g., based on pragmatic cross-validation), or by the use of double cross-validation (cross-model validation) [121]. These approaches, although not ideal, can be used until large, representative and independent test sets are obtained [121].

The steps needed to develop an application combining the sample, the spectra and the reference data are summarised in Figure 2.

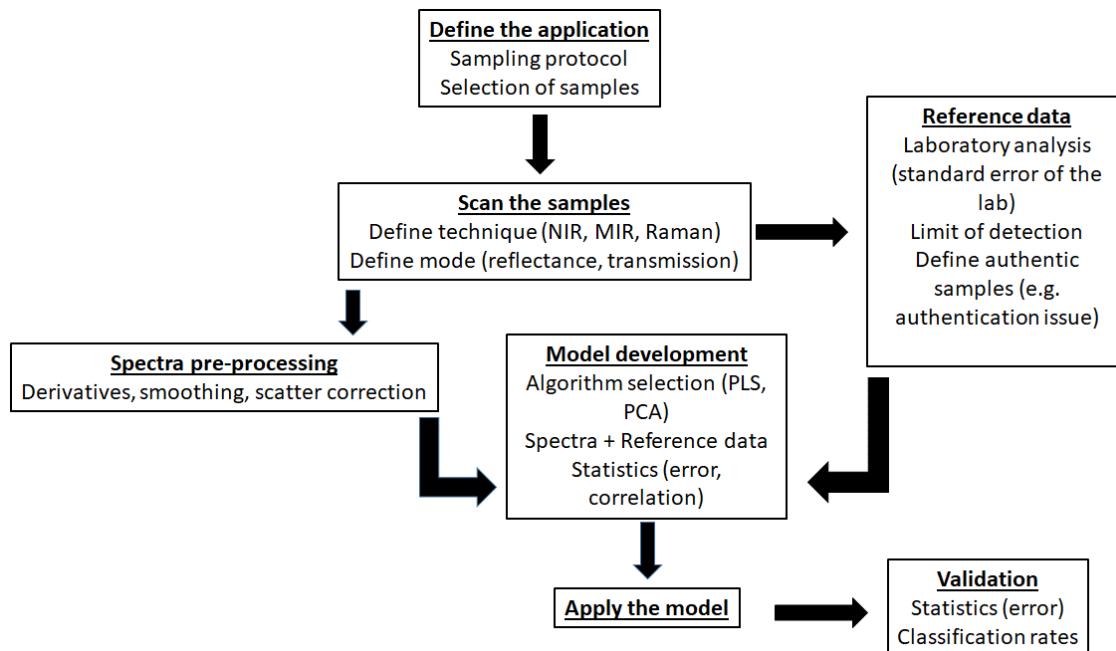


Figure 2. Steps needed to develop an application combining the sample, the spectra and the reference data.

6. Concluding Remarks

The integration of vibrational spectroscopy with MVA to develop analytical applications (e.g., calibration and classification) can be considered by the non-expert purely as a mathematical or statistical exercise. This, however, could not be further from the truth—calibration development is a complex process that implies the understanding of a system created by the sample and its inherent characteristics (e.g., physical and chemical properties, variability, origin, pre-processing, etc.), the origin of the spectra (e.g., instrument characteristics, sample collection mode, etc.) and all the aspects of the multivariate data analysis (e.g., pre-processing, selection of samples for calibration and validation, linear and non-linear algorithms, outliers, etc.).

These developments require a basic understanding of the different variables that contribute to the system and they include the sample, fundamentals of spectroscopy, data processing and analysis, sampling protocols, and limit of detection (see Figure 3). The adaptation of vibrational spectroscopy to efficiently and reliably contribute to the expansion in the number of applications related to analytical chemistry, process analytical technologies, traceability of food ingredients, and natural products, makes them an ideal set of methodologies towards sustainability along the food value chain. An increasing number of research groups have investigated the use of vibrational spectroscopy, as shown in several applications reported in the literature. However, commercial implementation of these techniques is still under development in some industries.

Even though several articles have been published in the scientific literature, most of them describe feasibility or potential applications of vibrational spectroscopy, where small datasets containing few samples are analysed and cross-validation, rather than an independent dataset, is used to validate the developed models (e.g., calibration). Adding to this is the little in-depth understanding of the reference lab (e.g., standard error of the laboratory method). Most of the application of vibrational spectroscopy are considered correlative methods, and their accuracy depends on the error of the reference method. Therefore, knowledge of the extent to which results are repeatable using wet chemistry or biochemical procedures is of paramount importance in judging the reliability calibration.

It is important to remember that the wet chemistry or reference data with all their known inadequacies are used to assess the performance of the calibrations; thus, before assessing the accuracy of a calibration or model, the error associated with the reference method should be known, and this is a fact that is often ignored. The lack of interpretation of loadings, significance of coefficients of regression, and inter-correlations among measured variables and chemical compounds are usually missing from the interpretation.

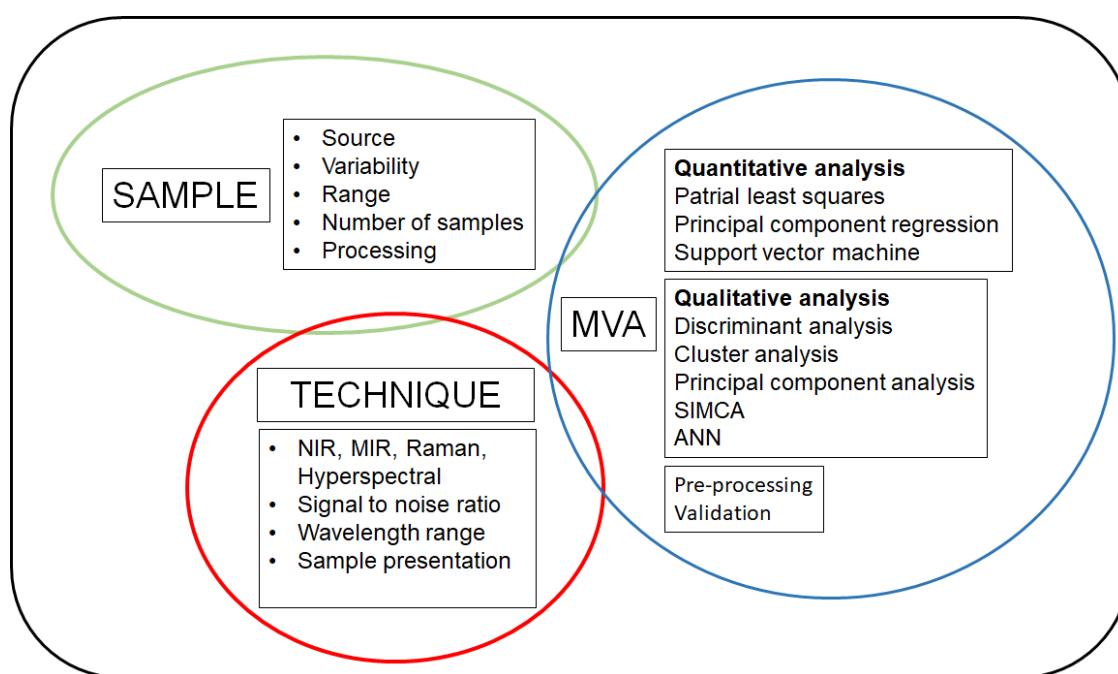


Figure 3. A schematic representation of the main components/variables that affect the sample, technique and data analysis.

The use of MVA reveals interesting information about the system but important bits might remain undiscovered. The extent or the use of good MVA (e.g., new algorithms, new software, or mathematical pre-processing) is meaningless if we fail in evaluating the best sample presentation, processing or interactions of the sample collection and analysis.

One of the interesting aspects of the modern integration of these technologies is that it requires and sources information and knowledge from many fields (e.g., spectroscopy, analytical chemistry, data analysis, biology, physics, etc.). This determines the unique multidisciplinary characteristic of this approach. A close collaboration between several researchers is therefore critical for the application and development of the technology. It is also important that everyone involved in the process understands and agrees upon the goals and requirements of the study beforehand to reduce the risk of weak links in the study. The definition of protocols for reporting the outcomes and results of any given study is also important.

Knowing and understanding the reference laboratory method (such as the standard error of the lab method), the limitations of the method, the physics and chemical basis of the spectra, as well as knowing and interpreting the interactions that exist between the sample and the instrument, will allow the user to better interpret the calibration or obtained mathematical relationships. It is therefore important that the individual that developed such calibrations has this knowledge in order to produce a method that can be reliable.

Martens [121] has highlighted that the scientific process of boring into the solid “mountain of the unknown” never stops, and that it is continuous. The author suggested that statistically valid claims must be replicated independently, intuitive hunches should be chased and solid manmade theories should be assessed critically.

The advantages and ability of vibrational spectroscopy to predict multiple parameters and speed of analysis mean that we have a powerful tool that can revolutionise the way we produce foods. The future development of such applications will provide the industry with a very fast and non-destructive method to monitor composition or changes and to detect unwanted problems, providing a rapid means of qualitative rather than quantitative analysis. Moreover, the choice of measuring device(s) may benefit from the experience in, e.g., multichannel diffuse near infrared (NIR) spectroscopy measuring many properties—preferably more than necessary, (it usually does not cost much extra).

However, various hurdles still hinder the growth and development of vibrational spectroscopy applications. Among them is the reluctance to accept the incorporation of vibrational spectroscopy with new statistical tools, such as multivariate data analysis techniques, as routine analytical or quality control methods. Besides, most of the current courses and training programmes in food still focus on the so-called classical approach where several aspects related to the incorporation of new technologies, sensors and programming are not yet incorporated in the curricula. The same can be said regarding research and other aspects of informal training and extension. Together with the silo mentality that still exist in the food industry, this hinders the possibility of exploiting the full potential of these systems by the industry.

Finally, one of the most important and critical aspects of the development of vibrational spectroscopy is the need for an appropriate level of training. For example, although knowledge of the chemistry of a sample material is useful, routine analyses can be performed by analysts with a high-school education. On the other hand, calibration development (interpretation, application and monitoring) is by far the most critical aspect and thus requires a high level of expertise, particularly in multivariate data analysis, in order to make an application successful. Where methods based on vibrational spectroscopy have been applied in industry situations, the potential savings, reduction in time and cost of analysis have been demonstrated. These methods show promising potential for in-field and process analysis.

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflict of interest.

References

- Burlingame, B. Holistic and reductionist nutrition. *J. Food Compos. Anal.* **2004**, *17*, 585–586. [[CrossRef](#)]
- Castro-Puyana, M.; Mendiola, J.A.; Ibáñez, E. Strategies for a cleaner new scientific discipline of green foodomics. *TrAC Trends Anal. Chem.* **2013**, *52*, 23–35. [[CrossRef](#)]
- Capozzi, F.; Bordoni, A. Foodomics: A new comprehensive approach to food and nutrition. *Genes Nutr.* **2012**, *8*, 1–4. [[CrossRef](#)] [[PubMed](#)]
- Cevallos-Cevallos, J.M.; Reyes-De-Corcuera, J.I.; Etxeberria, E.; Danyluk, M.D.; Rodrick, G.E. Metabolomic analysis in food science: A review. *Trends Food Sci. Technol.* **2009**, *20*, 557–566. [[CrossRef](#)]
- Truong, V.K.; Dupont, M.; Elbourne, A.; Gangadoo, S.; Pathirannahalage, P.R.; Cheeseman, S.; Chapman, J.; Cozzolino, D. From Academia to Reality Check: A Theoretical Framework on the Use of Chemometric in Food Sciences. *Foods* **2019**, *8*, 164. [[CrossRef](#)] [[PubMed](#)]
- Gallo, M.; Ferranti, P. The evolution of analytical chemistry methods in foodomics. *J. Chromatogr. A* **2016**, *1428*, 3–15. [[CrossRef](#)]
- Fardet, A. New Approaches to Studying the Potential Health Benefits of Cereals: From Reductionism to Holism. *Cereal Foods World* **2014**, *59*, 224–229. [[CrossRef](#)]
- Garcia-Cañas, V.; Simo, C.; Herrero, M.; Ibáñez, E.; Cifuentes, A. Present and Future Challenges in Food Analysis: Foodomics. *Anal. Chem.* **2012**, *84*, 10150–10159. [[CrossRef](#)]
- Wang, H.; Peng, J.-Y.; Xie, C.; Bao, Y.; He, Y. Fruit Quality Evaluation Using Spectroscopy Technology: A Review. *Sensors* **2015**, *15*, 11889–11927. [[CrossRef](#)]
- Tahir, H.E.; Zou, X.; Jianbo, X.; Mahunu, G.K.; Jiyong, S.; Xu, J.-L.; Sun, D.-W. Recent Progress in Rapid Analyses of Vitamins, Phenolic, and Volatile Compounds in Foods Using Vibrational Spectroscopy Combined with Chemometrics: A Review. *Food Anal. Methods* **2019**, *12*, 2361–2382. [[CrossRef](#)]

11. Amodio, M.L.; Chaudhry, M.M.A.; Colelli, G. Spectral and Hyperspectral Technologies as an Additional Tool to Increase Information on Quality and Origin of Horticultural Crops. *Agronomy* **2019**, *10*, 7. [[CrossRef](#)]
12. Cortés, V.; Blasco, J.; Aleixos, N.; Cubero, S.; Talens, P. Monitoring strategies for quality control of agricultural products using visible and near-infrared spectroscopy: A review. *Trends Food Sci. Technol.* **2019**, *85*, 138–148. [[CrossRef](#)]
13. Roberts, J.; Cozzolino, D. An Overview on the Application of Chemometrics in Food Science and Technology—An Approach to Quantitative Data Analysis. *Food Anal. Methods* **2016**, *9*, 3258–3267. [[CrossRef](#)]
14. Wang, L.; Sun, D.-W.; Pu, H.; Cheng, J.-H. Quality analysis, classification, and authentication of liquid foods by near-infrared spectroscopy: A review of recent research developments. *Crit. Rev. Food Sci. Nutr.* **2016**, *57*, 1524–1538. [[CrossRef](#)] [[PubMed](#)]
15. Alander, J.T.; Bochko, V.; Martinhauppi, B.; Saranwong, S.; Mantere, T. A review of optical non-destructive visual and near infrared methods for food quality and safety. *Int. J. Spectrosc.* **2013**, *2013*, 341402. [[CrossRef](#)]
16. Manley, M. Near-infrared spectroscopy and hyperspectral imaging: Non-destructive analysis of biological materials. *Chem. Soc. Rev.* **2014**, *43*, 8200–8214. [[CrossRef](#)]
17. Sorak, D.; Herberholz, L.; Iwascek, S.; Altinpinar, S.; Pfeifer, F.; Siesler, H.W. New Developments and Applications of Handheld Raman, Mid-Infrared, and Near-Infrared Spectrometers. *Appl. Spectrosc. Rev.* **2012**, *47*, 83–115. [[CrossRef](#)]
18. Callao, M.; Ruisánchez, I. An overview of multivariate qualitative methods for food fraud detection. *Food Control* **2018**, *86*, 283–293. [[CrossRef](#)]
19. Cozzolino, D. The role of vibrational spectroscopy as a tool to assess economically motivated fraud and counterfeit issues in agricultural products and foods. *Anal. Methods* **2015**, *7*, 9390–9400. [[CrossRef](#)]
20. Power, A.; Cozzolino, D. How Fishy Is Your Fish? Authentication, Provenance and Traceability in Fish and Seafood by Means of Vibrational Spectroscopy. *Appl. Sci.* **2020**, *10*, 4150. [[CrossRef](#)]
21. Eisen, K.; Eifert, T.; Herwig, C.; Maiwald, M. Current and future requirements to industrial analytical infrastructure—Part 1: Process analytical laboratories. *Anal. Bioanal. Chem.* **2020**, *412*, 2027–2035. [[CrossRef](#)] [[PubMed](#)]
22. Eifert, T.; Eisen, K.; Maiwald, M.; Herwig, C. Current and future requirements to industrial analytical infrastructure—Part 2: Smart sensors. *Anal. Bioanal. Chem.* **2020**, *412*, 2037–2045. [[CrossRef](#)] [[PubMed](#)]
23. Herwig, C. Applied basic science in process analytics and control technology. *Anal. Bioanal. Chem.* **2020**, *412*, 2025–2026. [[CrossRef](#)] [[PubMed](#)]
24. Rolinger, L.; Rudt, N.; Hubbuch, J. A critical review of recent trends, and a future perspective of optical spectroscopy as PAT in biopharmaceutical downstream processing. *Anal. Bioanal. Chem.* **2020**, *412*, 2047–2064. [[CrossRef](#)]
25. Bendell, A.; Disney, J.; McCollin, C. The Future Role of Statistics in Quality Engineering and Management (with Discussion). *J. R. Stat. Soc. A. Stat.* **1999**, *48*, 299–326. [[CrossRef](#)]
26. Brereton, R.G. *Applied Chemometrics for Scientists*; Johns Wiley & Sons Ltd.: Chichester, UK, 2007.
27. Brereton, R.G. Introduction to multivariate calibration in analytical chemistry. *Analyst* **2000**, *125*, 2125–2154. [[CrossRef](#)]
28. Buco, S.M. How good are your results? An approach to qualitative and quantitative statistical analysis for food monitoring and process control systems. *Food Control* **1990**, *1*, 40–46. [[CrossRef](#)]
29. Granato, D.; Calado, V.M.D.A.; Jarvis, B. Observations on the use of statistical methods in Food Science and Technology. *Food Res. Int.* **2014**, *55*, 137–149. [[CrossRef](#)]
30. Nunes, C.A.; Alvarenga, V.; Sant’Ana, A.D.S.; Santos, J.S.; Granato, D. The use of statistical software in food science and technology: Advantages, limitations and misuses. *Food Res. Int.* **2015**, *75*, 270–280. [[CrossRef](#)]
31. Wold, S. Chemometrics; what do we mean with it, and what do we want from it? *Chemom. Intell. Lab. Syst.* **1995**, *30*, 109–115. [[CrossRef](#)]
32. Bronson, K.; Knezevic, I. Big Data in food and agriculture. *Big Data Soc.* **2016**, *3*, 1–5. [[CrossRef](#)]
33. Ahearn, M.C.; Armbruster, W.; Young, R. Big Data’s Potential to Improve Food Supply Chain Environmental Sustainability and Food Safety. *Int. Food Agribus. Man.* **2016**, *19*, 155–171. [[CrossRef](#)]
34. Szymańska, E. Modern data science for analytical chemical data—A comprehensive review. *Anal. Chim. Acta* **2018**, *1028*, 1–10. [[CrossRef](#)] [[PubMed](#)]
35. Esbensen, K.H.; Wagner, C. Theory of sampling (TOS) versus measurement uncertainty (MU)—A call for integration. *Trends Analyt. Chem.* **2014**, *57*, 93–106. [[CrossRef](#)]

36. Wagner, C.; Esbensen, K.H. Theory of Sampling: Four Critical Success Factors before Analysis. *J. AOAC Int.* **2015**, *98*, 275–281. [[CrossRef](#)]
37. Esbensen, K.H.; Swarbrick, B. Sampling for spectroscopic analysis: Consequences for multivariate calibration. *Spectrosc. Eur.* **2019**, *3*, 22–28.
38. Analytical Methods Committee, AMCTB No. 71. Sampling theory and sampling uncertainty. *Anal. Methods* **2015**, *7*, 10085–10087. [[CrossRef](#)]
39. Saeys, W.; Trong, N.N.D.; Van Beers, R.; Nicolaï, B.M. Multivariate calibration of spectroscopic sensors for postharvest quality evaluation: A review. *Postharvest Biol. Technol.* **2019**, *158*, 110981. [[CrossRef](#)]
40. Cozzolino, D.; Shah, N.; Cynkar, W.; Smith, P. A practical overview of multivariate data analysis applied to spectroscopy. *Food Res. Int.* **2011**, *44*, 1888–1896. [[CrossRef](#)]
41. Badertscher, M.; Pretsch, E. Bad results from good data. *TrAC Trends Anal. Chem.* **2006**, *25*, 1131–1138. [[CrossRef](#)]
42. Murray, I. Forage analysis by near infrared spectroscopy. In *Sward Management Handbook*; Davies, A., Baker, R.D., Grant, S.A., Laidlaw, A.S., Eds.; British Grassland Society: Cheshire, UK, 1993; pp. 285–312.
43. Murray, I. Hirschfeld Award. NIR spectroscopy of food: Simple things, subtle things and spectra. *NIR News* **1999**, *10*, 10–12. [[CrossRef](#)]
44. Murray, I.; Cowe, I. Sample preparation. In *Near Infrared Spectroscopy in Agriculture*; Roberts, C.A., Workman, J., Reeves, J.B., Eds.; American Society of Agronomy, Crop Science Society of America, Soil Science Society of America: Madison, WI, USA, 2004; pp. 75–115.
45. Cozzolino, D. Sample presentation, sources of error and future perspectives on the application of vibrational spectroscopy in the wine industry. *J. Sci. Food Agric.* **2014**, *95*, 861–868. [[CrossRef](#)] [[PubMed](#)]
46. Williams, P.C. Implementation of Near-Infrared technology. In *Near Infrared Technology in the Agricultural and Food Industries*; Williams, P.C., Norris, K.H., Eds.; American Association of Cereal Chemist: St. Paul, MN, USA, 2001; pp. 145–169.
47. Nicolai, B.M.; Beullens, K.; Bobelyn, E.; Peirs, A.; Saeys, W.; Theron, K.I.; Lammertyn, J. Non-destructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. *Postharvest Biol. Tec.* **2007**, *46*, 99–118. [[CrossRef](#)]
48. Brereton, R.G.; Jansen, J.; Lopes, J.; Marini, F.; Pomerantsev, A.L.; Rodionova, O.Y.; Roger, J.-M.; Walczak, B.; Tauler, R. Chemometrics in analytical chemistry—Part II: Modeling, validation, and applications. *Anal. Bioanal. Chem.* **2018**, *410*, 6691–6704. [[CrossRef](#)]
49. Gani, W.; Limam, M. A kernel distance-based representative subset selection method. *J. Stat. Comput. Simul.* **2015**, *86*, 135–148. [[CrossRef](#)]
50. Morais, C.L.M.; Santos, M.C.D.; Lima, K.M.G.; Martin, F.L. Improving data splitting for classification applications in spectrochemical analyses employing a random-mutation Kennard-Stone algorithm approach. *Bioinformatics* **2019**, *35*, 5257–5263. [[CrossRef](#)]
51. Munck, L. A new holistic exploratory approach to Systems Biology by Near Infrared Spectroscopy evaluated by chemometrics and data inspection. *J. Chemom.* **2007**, *21*, 406–426. [[CrossRef](#)]
52. Munck, L.; Nørgaard, L.; Engelsen, S.B.; Bro, R.; Andersson, C. Chemometrics in food science—A demonstration of the feasibility of a highly exploratory, inductive evaluation strategy of fundamental scientific significance. *Chemom. Intell. Lab. Syst.* **1998**, *44*, 31–60. [[CrossRef](#)]
53. Munck, L.; Møller, J.B.; Rinnan, Å.; Fast, S.H.; Møller, E.M.; Nørgaard, L.; Balling, E.S. A physicochemical theory on the applicability of soft mathematical models—Experimentally interpreted. *J. Chemom.* **2010**, *24*, 481–495. [[CrossRef](#)]
54. Khakimov, B.; Bak, S.; Engelsen, S.B. High-throughput cereal metabolomics: Current analytical technologies, challenges and perspectives. *J. Cereal Sci.* **2014**, *59*, 393–418. [[CrossRef](#)]
55. Cattaneo, T.; Stellar, A. Review: NIR Spectroscopy as a Suitable Tool for the Investigation of the Horticultural Field. *Agronomy* **2019**, *9*, 503. [[CrossRef](#)]
56. Pasquini, C. Near infrared spectroscopy: A mature analytical technique with new perspectives—A review. *Anal. Chim. Acta* **2018**, *1026*, 8–36. [[CrossRef](#)] [[PubMed](#)]
57. Karoui, R.; Downey, G.; Blecker, C. Mid-Infrared Spectroscopy Coupled with Chemometrics: A Tool for the Analysis of Intact Food Systems and the Exploration of Their Molecular Structure–Quality Relationships—A Review. *Chem. Rev.* **2010**, *110*, 6144–6168. [[CrossRef](#)] [[PubMed](#)]

58. Ryan, D.; Robards, K. Analytical Chemistry Considerations in Plant Metabolomics. *Sep. Purif. Rev.* **2006**, *35*, 319–356. [[CrossRef](#)]
59. Beć, K.B.; Grabska, J.E.; Huck, C.W. Near-Infrared Spectroscopy in Bio-Applications. *Molecules* **2020**, *25*, 2948. [[CrossRef](#)]
60. Cozzolino, D.; Roberts, J. Applications and Developments on the Use of Vibrational Spectroscopy Imaging for the Analysis, Monitoring and Characterisation of Crops and Plants. *Molecules* **2016**, *21*, 755. [[CrossRef](#)]
61. Walsh, K.; McGlone, V.; Han, D. The uses of near infra-red spectroscopy in postharvest decision support: A review. *Postharvest Biol. Technol.* **2020**, *163*, 111139. [[CrossRef](#)]
62. Beć, K.B.; Huck, C.W. Breakthrough Potential in Near-Infrared Spectroscopy: Spectra Simulation. A Review of Recent Developments. *Front. Chem.* **2019**, *7*. [[CrossRef](#)]
63. Fu, X.; Chen, J. A Review of Hyperspectral Imaging for Chicken Meat Safety and Quality Evaluation: Application, Hardware, and Software. *Compr. Rev. Food Sci. Food Saf.* **2019**, *18*, 535–547. [[CrossRef](#)]
64. Crocombe, R.A. Portable Spectroscopy. *Appl. Spectrosc.* **2018**, *72*, 1701–1751. [[CrossRef](#)]
65. Adams, M.J. Chemometrics in Analytical Spectroscopy. In *Chemometrics in Analytical Spectroscopy*; Royal Society of Chemistry (RSC): Cambridge, UK, 2004; p. 216.
66. Blanco, M.; Bernardez, M. Multivariate calibration for quantitative analysis. In *Infrared Spectroscopy for Food Quality Analysis and Control*; Sun, D.W., Ed.; Elsevier: Oxford, UK, 2009; p. 448.
67. Esbensen, K.H. *Multivariate Data Analysis in Practice*; CAMO Process AS: Oslo, Norway, 2002.
68. Frank, I.E.; Friedman, J.H. A Statistical View of Some Chemometrics Regression Tools. *Technometrics* **1993**, *35*, 109. [[CrossRef](#)]
69. Geladi, P.; Kowalski, B.R. Partial least-squares regression: A tutorial. *Anal. Chim. Acta* **1986**, *185*, 1–17. [[CrossRef](#)]
70. Geladi, P. Chemometrics in spectroscopy. Part 1. Classical chemometrics. *Spectrochim. Acta Part B: At. Spectrosc.* **2003**, *58*, 767–782. [[CrossRef](#)]
71. Kumar, N.; Bansal, A.; Sarma, G.; Rawal, R.K. Chemometrics tools used in analytical chemistry: An overview. *Talanta* **2014**, *123*, 186–199. [[CrossRef](#)] [[PubMed](#)]
72. Haaland, D.M.; Thomas, E.V. Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. *Anal. Chem.* **1988**, *60*, 1193–1202. [[CrossRef](#)]
73. Mark, H.; Workman, J. *Statistics in Spectroscopy*, 2nd ed.; Elsevier BV: London, UK, 2003; 342p.
74. Martens, H.M.A.M. Multivariate Analysis of Quality. An Introduction. *Meas. Sci. Technol.* **2001**, *12*, 1746. [[CrossRef](#)]
75. Massart, D.L.; Vandegiste, B.G.M.; Deming, S.N.; Michotte, Y.; Kaufman, L. *Chemometrics: A Textbook*; Elsevier: Amsterdam, The Netherlands, 1988; 500p.
76. Naes, T.; Isaksson, T.; Fearn, T.; Davies, T. *A User-Friendly Guide to Multivariate Calibration and Classification*; NIR Publications: Chichester, UK, 2002; 420p.
77. Skov, T.; Honore, A.H.; Jensen, H.M.; Næs, T.; Engelsen, S.B. Chemometrics in foodomics: Handling data structures from multiple analytical platforms. *TrAC Trends Anal. Chem.* **2014**, *60*, 71–79. [[CrossRef](#)]
78. Bureau, S.; Cozzolino, D.; Clark, C. Contributions of Fourier-transform mid infrared (FT-MIR) spectroscopy to the study of fruit and vegetables: A review. *Postharvest Biol. Technol.* **2019**, *148*, 1–14. [[CrossRef](#)]
79. Duckworth, J. Mathematical data processing. In *Near Infrared Spectroscopy in Agriculture*; Roberts, C.A., Workman, J., Reeves, J.B., Eds.; American Society of Agronomy, Crop Science Society of America, Soil Science Society of America: Madison, WI, USA, 2004; pp. 115–132.
80. Engel, J.; Gerretzen, J.; Szymanska, 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](#)]
81. Mehmood, T.; Liland, K.H.; Snipen, L.; Sæbø, S. A review of variable selection methods in Partial Least Squares Regression. *Chemom. Intell. Lab. Syst.* **2012**, *118*, 62–69. [[CrossRef](#)]
82. Rinnan, Å.; Berg, F.V.D.; Engelsen, S.B. Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends Anal. Chem.* **2009**, *28*, 1201–1222. [[CrossRef](#)]
83. Mutihac, L.; Mutihac, R. Mining in chemometrics. *Anal. Chim. Acta* **2008**, *612*, 1–18. [[CrossRef](#)] [[PubMed](#)]
84. Leardi, R. Experimental design in chemistry: A tutorial. *Anal. Chim. Acta* **2009**, *652*, 161–172. [[CrossRef](#)] [[PubMed](#)]
85. DiFoggio, R. Guidelines for Applying Chemometrics to Spectra: Feasibility and Error Propagation. *Appl. Spectrosc.* **2000**, *54*, 94A–113A. [[CrossRef](#)]

86. Ahmadvand, M.; Parastar, H.; Sereshti, H.; Olivier, A.; Tauler, R. A systematic study on the effect of noise and shift on multivariate figures of merit of second-order calibration algorithms. *Anal. Chim. Acta* **2017**, *952*, 18–31. [[CrossRef](#)]
87. Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: A basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130. [[CrossRef](#)]
88. Wold, S.; Josefson, M.; Gottfries, J.; Linusson, A. The utility of multivariate design in PLS regression. *J. Chemom.* **2004**, *18*, 156–165. [[CrossRef](#)]
89. Hopke, P.K. The evolution of chemometrics. *Anal. Chim. Acta* **2003**, *500*, 365–377. [[CrossRef](#)]
90. Møller, S.F.; Von Frese, J.; Bro, R. Robust methods for multivariate data analysis. *J. Chemom.* **2005**, *19*, 549–563. [[CrossRef](#)]
91. Vogt, F. Information fusion via constrained principal component regression for robust quantification with incomplete calibrations. *Anal. Chim. Acta* **2013**, *797*, 20–29. [[CrossRef](#)]
92. Bro, R.; Smilde, A.K. Principal component analysis. *Anal. Methods* **2014**, *6*, 2812–2831. [[CrossRef](#)]
93. Szymanska, E.; Gerretzen, J.; Engel, J.; Geurts, B.; Blanchet, L.; Buydens, L.M. Chemometrics and qualitative analysis have a vibrant relationship. *TrAC Trends Anal. Chem.* **2015**, *69*, 34–51. [[CrossRef](#)]
94. Bevilacqua, M.; Bro, R.; Marini, F.; Rinnan, Å.; Rasmussen, M.A.; Skov, T. Recent chemometrics advances for foodomics. *TrAC Trends Anal. Chem.* **2017**, *96*, 42–51. [[CrossRef](#)]
95. Gromski, P.S.; Muhamadali, H.; Ellis, D.I.; Xu, Y.; Correa, E.; Turner, M.L.; Goodacre, R. A tutorial review: Metabolomics and partial least squares-discriminant analysis—A marriage of convenience or a shotgun wedding. *Anal. Chim. Acta* **2015**, *879*, 10–23. [[CrossRef](#)]
96. Tange, R.I.; Rasmussen, M.A.; Taira, E.; Bro, R. Benchmarking support vector regression against partial least squares regression and artificial neural network: Effect of sample size on model performance. *J. Near Infrared Spectrosc.* **2017**, *25*, 381–390. [[CrossRef](#)]
97. Balabin, R.M.; Lomakina, E.I. Support vector machine regression (SVR/LS-SVM)—An alternative to neural networks (ANN) for analytical chemistry? Comparison of nonlinear methods on near infrared (NIR) spectroscopy data. *Analyst* **2011**, *136*, 1703–1712. [[CrossRef](#)]
98. Balabin, R.M.; Smirnov, S.V. Interpolation and extrapolation problems of multivariate regression in analytical chemistry: Benchmarking the robustness on near-infrared (NIR) spectroscopy data. *Analyst* **2012**, *137*, 1604. [[CrossRef](#)]
99. Ni, W.; Nørgaard, L.; Mørup, M. Non-linear calibration models for near infrared spectroscopy. *Anal. Chim. Acta* **2014**, *813*, 1–14. [[CrossRef](#)]
100. Tange, R.; Rasmussen, M.A.; Taira, E.; Bro, R. Application of Support Vector Regression for Simultaneous Modelling of near Infrared Spectra from Multiple Process Steps. *J. Near Infrared Spectrosc.* **2015**, *23*, 75–84. [[CrossRef](#)]
101. Souihé, N.; Lindegren, A.; Eriksson, L.; Trygg, J. OPLS in batch monitoring—Opens up new opportunities. *Anal. Chim. Acta* **2015**, *857*, 28–38. [[CrossRef](#)]
102. Funes, E.; Allouche, Y.; Beltrán, G.; Jiménez, A. A Review: Artificial Neural Networks as Tool for Control Food Industry Process. *J. Sens. Technol.* **2015**, *5*, 28–43. [[CrossRef](#)]
103. Zupan, J.; Gasteiger, J. Neural networks: A new method for solving chemical problems or just a passing phase? *Anal. Chim. Acta* **1991**, *248*, 1–30. [[CrossRef](#)]
104. Liu, A.; Li, G.; Fu, Z.; Guan, Y.; Lin, L. Non-linearity correction in NIR absorption spectra by grouping modeling according to the content of analyte. *Sci. Rep.* **2018**, *8*, 8564. [[CrossRef](#)] [[PubMed](#)]
105. Brereton, R.G. Pattern recognition in chemometrics. *Chemom. Intell. Lab. Syst.* **2015**, *149*, 90–96. [[CrossRef](#)]
106. Fearn, T. Assessing calibrations: SEP, RPD, RER and R2. *NIR News* **2002**, *13*, 12–13. [[CrossRef](#)]
107. Fearn, T. Validation. *NIR News* **1997**, *8*, 7–8. [[CrossRef](#)]
108. Norris, K.H.; Ritchie, G.E. Assuring specificity for a multivariate near-infrared (NIR) calibration: The example of the Chambersburg Shoot-out 2002 data set. *J. Pharm. Biomed. Anal.* **2008**, *48*, 1037–1041. [[CrossRef](#)]
109. Walsh, K.B.; Kawano, S. Near infrared spectroscopy. In *Optical Monitoring of Fresh and Processed Agricultural Crops*; Zude, M., Ed.; CRC Press: Boca Raton, FL, USA, 2009; pp. 192–239.
110. Westad, F.; Marini, F. Validation of chemometric models—A tutorial. *Anal. Chim. Acta* **2015**, *893*, 14–24. [[CrossRef](#)]
111. Asuero, A.G.; Sayago, A.; Gustavo, A.G. The Correlation Coefficient: An Overview. *Crit. Rev. Anal. Chem.* **2006**, *36*, 41–59. [[CrossRef](#)]

112. Williams, P.; Dardenne, P.; Flinn, P. Tutorial: Items to be included in a report on a near infrared spectroscopy project. *J. Near Infrared Spectrosc.* **2017**, *25*, 85–90. [[CrossRef](#)]
113. Agelet, L.E.; Hurlburgh, C.R. A Tutorial on Near Infrared Spectroscopy and Its Calibration. *Crit. Rev. Anal. Chem.* **2010**, *40*, 246–260. [[CrossRef](#)]
114. 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* **2019**, *1058*, 9–17. [[CrossRef](#)] [[PubMed](#)]
115. Alewijn, M.; van der Voet, H.; van Ruth, S. Validation of multivariate classification methods using analytical fingerprints—Concept and case study on organic feed for laying hens. *J. Food Compost. Anal.* **2016**, *51*, 15–23. [[CrossRef](#)]
116. Perez-Guaita, D.; Kuligowski, J.; Lendl, B.; Wood, B.R.; Quint, G. Assessment of discriminant models in infrared imaging using constrained repeated random sampling-Cross validation. *Anal. Chim. Acta* **2018**, *1033*, 156–164. [[CrossRef](#)] [[PubMed](#)]
117. Kjeldahl, K.; Bro, R. Some common misunderstandings in chemometrics. *J. Chemom.* **2010**, *24*, 558–564. [[CrossRef](#)]
118. Hawkins, D.M. The problem of overfitting. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1–12. [[CrossRef](#)]
119. Wise, B.M.; Gallagher, N.B. The process chemometrics approach to process monitoring and fault detection. *J. Process. Control* **1996**, *6*, 329–348. [[CrossRef](#)]
120. Hibbert, D.B.; Minkkinen, P.; Faber, N.; Wise, B.M. IUPAC project: A glossary of concepts and terms in chemometrics. *Anal. Chim. Acta* **2009**, *642*, 3–5. [[CrossRef](#)]
121. Martens, H. Quantitative Big Data: Where chemometrics can contribute. *J. Chemom.* **2015**, *29*, 563–581. [[CrossRef](#)]
122. Oliveri, P. Class-Modelling Approaches: Advantages over Discriminant Methods for Compliance Verifications. *NIR News* **2016**, *27*, 29–30. [[CrossRef](#)]
123. Oliveri, P. Class-modelling in food analytical chemistry: Development, sampling, optimisation and validation issues—A tutorial. *Anal. Chim. Acta* **2017**, *982*, 9–19. [[CrossRef](#)] [[PubMed](#)]
124. Riedl, J.; Esslinger, S.; Fauhl-Hassek, C. Review of validation and reporting of non-targeted fingerprinting approaches for food authentication. *Anal. Chim. Acta* **2015**, *885*, 17–32. [[CrossRef](#)] [[PubMed](#)]



© 2020 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).