Trends in Analytical Chemistry 181 (2024) 118051

Contents lists available at ScienceDirect

Trends in Analytical Chemistry

journal homepage: www.elsevier.com/locate/trac

Navigating the complexity: Managing multivariate error and uncertainties in spectroscopic data modelling

Barbara Giussani [a,*], Giulia Gorla [b], Jokin Ezenarro [c], Jordi Riu [c], Ricard Boqué [c]

[a] Dipartimento di Scienza e Alta Tecnologia, Università degli Studi dell'Insubria, Via Valleggio 9, 22100, Como, Italy
[b] Department of Analytical Chemistry, Faculty of Science and Technology, University of the Basque Country UPV/EHU, Sarriena s/n, 48940, Leioa, Basque Country, Spain
[c] Universitat Rovira i Virgili. Department of Analytical Chemistry and Organic Chemistry, Carrer Marcel·lí Domingo 1, 43007, Tarragona, Spain

ARTICLE INFO

Keywords:
Uncertainty estimation
Multivariate measurement error
Multivariate calibration
Multivariate classification
Exploratory analysis
Spectroscopy

ABSTRACT

Spectroscopy and chemometrics, supported by computer science, have yielded promising outcomes, as evidenced by trends observed in literature searches. However, while researchers meticulously construct chemometric models for exploratory, quantitation and classification purposes, the investigation of data quality, particularly error analysis, remains less frequent. Understanding and quantifying measurement errors is crucial for robust spectroscopic modeling and uncertainty estimation. By unraveling complexities related to multivariate errors and uncertainties in spectroscopic data, the scientific community is empowered to extract reliable information from spectroscopic analyses, paving the way for enhanced analytical practices. This review underscores the necessity for the scientific community to integrate error analysis and uncertainty estimation into multivariate analysis methods, offering tailored solutions for diverse data types and analysis objectives.

1. Introduction

Spectroscopic techniques have become indispensable tools in diverse scientific disciplines, offering profound insights into the molecular composition of materials (see e.g. Refs. [1–5] and references therein). Alongside traditional benchtop spectroscopy, many applications have been developed in the last years with portable spectroscopic instrumentation [6–8]. However, in some cases, portability, may entail a lack of selectivity, requiring efforts to determine the analytes of interest by separating their contribution from interferents [9]. Portable sensors can be deployed at-line, on-line and in-line, and the massive amount of data collected requires appropriate analysis methods that have to be continuously improved [10].

Multivariate data generated by a spectrophotometer require appropriate techniques for analysis, and chemometric methods are certainly among the most suitable for this purpose. The coupling of infrared spectroscopy and chemometrics emerged as a compelling and potentially enduring duo, with the assistance of a third party: computer science. For helping visualizing the trajectories of the fields, the results obtained in Scopus by using the queries: "(TITLE-ABS-KEY ((IR W/2 spectroscopy) OR (infrared W/2 spectroscopy))), "(infrared W/1 spectroscopy) AND (chemometrics OR (multivariate W/2 analysis))", "(TITLE-ABS-KEY(Chemometrics))","(TITLE-ABS-KEY ((chemometrics OR (multivariate W/2 analysis)) AND (infrared W/1 spectroscopy)))" are reported in Fig. 1. Similar trends were observed from the results offered for the same keywords on Web of Science.

The figure reveals that chemometrics experienced significant success in tandem with the rise of spectrometers and the availability of computers. The trend of publications over the years shows similar tendencies for chemometrics and chemometrics applied in spectroscopy by highlighting the correlation between their spreading.

Researchers working in the spectroscopic field are very careful to construct and develop increasingly suitable and reliable chemometric models for obtaining predictions, classifications, or even just for visualizing data. However, less often, the data quality in terms of the error associated with the data is investigated. This occurs because the errors associated with the measurements are often small, and the assumptions made when applying chemometric methods are generally valid. However, this is not always true, and it needs further investigation [11–13]. The results of a chemometric model heavily depend on the data quality and the understanding of measurement errors [14]. The diverse nature of multivariate errors in spectroscopic data has a major impact on data preprocessing. By aligning preprocessing strategies with the nature of multivariate errors, researchers can enhance the signal-to-noise ratio,

* Corresponding author.
E-mail address: barbara.giussani@uninsubria.it (B. Giussani).

https://doi.org/10.1016/j.trac.2024.118051
Received 30 April 2024; Received in revised form 14 October 2024; Accepted 9 November 2024
Available online 12 November 2024

leading to more accurate and interpretable spectroscopic models [15]. For instance, if instrumental noise dominates the dataset, specific denoising methods tailored to the spectral characteristics of the instrument may be more effective. Conversely, if variations in sample composition contribute significantly to the errors, normalization techniques specific to the sample type may be warranted.

The evaluation of the uncertainties of the results is another key point, and this concerns both quantitative and qualitative methods [16]. It is also important to mention that there is a degree of confusion about the term 'uncertainty', and it is frequently used in the literature to refer to different related concepts such as precision, error or confidence interval, when uncertainty is a term that encompasses all the sources of error, both random but also systematic, of an analytical method. Let us consider the impact that studying uncertainty has on estimating the figures of merit of an analytical method. Any analytical method requires the assessment of figures of merit, and multivariate calibration methods are not an exception. In this context, metrology and chemometrics have found themselves working in synergy [17–22]. When it comes to industrial processes, uncertainty can be integrated into process monitoring and control strategies, providing very interesting insights [23–25]. And in the case of routine methods based on spectroscopic techniques, the significance of understanding measurement uncertainty is primary [26]. In all these cases, authors rarely discuss the uncertainty of their multivariate results [27] and, if they do, they infrequently use this information to optimize their models [28]. It is worth recalling that errors in the reference values (the **Y** vector in case of regression models) also affect uncertainty estimation [29]. The success of spectroscopic data modeling should rely not only on the prediction estimates but also on their associated uncertainty. The most popular option for the calculation of uncertainty, although not the only one, relies on error-propagation equations. These equations, designed to estimate and propagate uncertainties throughout the modeling process, must be tailored to the multivariate nature of errors. Understanding how errors interact across multiple dimensions allows for the formulation of more accurate equations that capture the complexity of uncertainties. This ensures that the propagated uncertainties in classification and prediction models are not oversimplified, providing a more realistic assessment of the reliability of analytical results [30].

Thus, everything is interconnected. Estimating uncertainty depends on certain assumptions, such as the error in measurement being independent and identically distributed (iid). However, this may not always be valid, as mentioned earlier, so studying the error structure of multivariate data is a key step in estimating the uncertainty of the results. In addition, the correct estimation of uncertainty is essential for accurately estimating the figures of merit of the calculated models or other important parameters, e.g. Q-residuals and Hotelling-T^2 statistics [30,31]. Back in 1997, Paul de Bièvre wrote a very strong statement concerning the importance of providing the uncertainty of analytical results. "So, a result without reliability (uncertainty) statement cannot be published or communicated because it is not (yet) a result. I am appealing to my colleagues of all analytical journals not to accept papers anymore which do not respect this simple logic." [32]. This statement also holds for results obtained from multivariate calibration or classification models. Knowing the structure of errors allows for the development of approaches to calculate the uncertainties of classification and prediction outcomes, as well as calibration transfer strategies [33–37].

Multivariate errors in spectroscopic data arise from multiple sources, encompassing instrumental noise, variations in sample composition, and fluctuations in experimental conditions. These perturbations are not necessarily independent from each other [38]. These errors are not unidimensional; rather, they manifest as complex interactions of multiple factors that challenge the traditional paradigms of error correction. Recognizing the nature of these errors is a fundamental step towards developing robust strategies for their identification, characterization, and effective management.

One of the distinctive features of multivariate errors in spectroscopic data is the presence of heteroscedasticity, where the variance of errors varies across different regions of the spectrum and/or among spectra [11,39–41]. This phenomenon can be attributed to factors such as the varying sensitivity of the instrument at different wavelengths or fluctuations in experimental conditions. In turn, correlated errors, where the errors in one part of the spectrum are systematically related to errors in another part, pose another layer of complexity in spectroscopic data modelling. These correlations may arise from instrumental drift, environmental factors, or other unmodeled sources of variability.

Effectively managing heteroscedastic and correlated errors is crucial for accurate model development and to develop more accurate uncertainty estimations. Ignoring these errors can lead to inaccurate (biased) parameter estimates and underestimated uncertainties [42].

Until a few years ago, it could be said that methods for the study of multivariate error and the calculation of uncertainty were not commonly used for several reasons: they require instrumental and

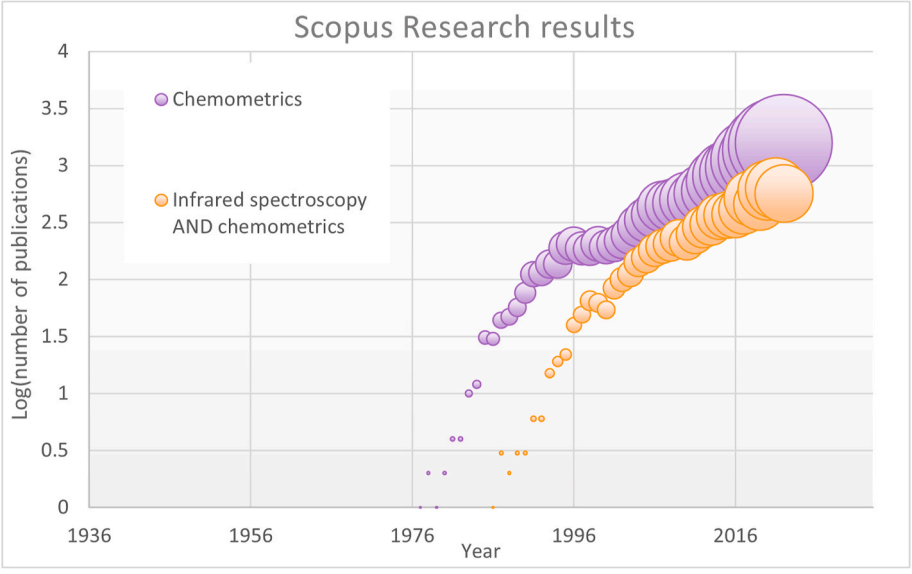

Fig. 1. Publications per year related to infrared spectroscopy, chemometrics and the combination of the two on Scopus. Representation in logarithmic scale. Different colors are intended for different keywords. The dimensions of the balls represent the numbers in their real scales.

analytical replicates, which were not always easy to obtain, especially with the expensive benchtop techniques prevalent years ago. Additionally, they certainly require computational effort, which could have been a hindrance a few years ago, but now is overcome by modern computers [43]. Today, these issues can be considered overcome in most cases. However, even though methods for modelling errors in data and uncertainty in results are known, there is no unanimity among scientists in the field, and often the proposed methods lack statistical validation [44,45].

This review embarks on a comprehensive exploration of the strategies and methodologies employed in the management of multivariate errors and uncertainties in spectroscopic data modelling. From addressing baseline shifts and instrumental noise to accommodating variations introduced by diverse sample matrices, the intricacies of handling uncertainties are both a theoretical and practical necessity. Understanding the nature of these errors and the uncertainties in classification and prediction models are critical for advancing the accuracy of analytical results. This article aims to provide researchers, analysts, and scientists with a roadmap for navigating the challenges posed by multivariate errors in spectroscopic data. As we unravel the layers of complexity inherent in data modelling, we aspire to empower the scientific community with the knowledge and tools necessary to extract meaningful and reliable information from spectroscopic analyses.

2. The nature of errors in analytical chemistry

As suggested by Booksh and Kowalski in 1994 [46], analytical chemistry, as the science of chemical measurements, is inherently linked with the concept of measurement error both in qualitative [16] and quantitative analysis [47]. From this, emerges the need for acknowledging figures of merit as essential both for characterizing a method and for method's comparison. Errors could originate from several sources of variability introduced throughout the analysis process, spanning from the sampling step to data analysis, encompassing sample preparation and instrumental measurement as well. In the case of spectroscopic measurements, the different sources of variability depend on how the experiment is conducted and on the equipment used. For example, in the case of portable near-infrared instrumentation, the variability introduced by sample preparation is often reduced; however, this implies that the heterogeneity of the sample has often not been addressed.

The use of molecular spectroscopy has spread in the laboratories through the 80's and since then, attention has been raised regarding the precision of spectrometers considering the different sources of variability involved [48–51]. The measurement precision was described theoretically as dependent upon a number of factors: the blank and dark references solutions, the light optical path, and the sample positioning. Over time, the technologies and instrumentations have evolved, and while the main errors are still present, their characteristics could be considered to have changed as well. For example, the instrumentation for near infrared has been pushed in recent years towards portability and real-time implementations through miniaturization. In this context, several studies on different applications [52] have been published to investigate issues related to strategies for use and improvement of the technologies [53,54].

When aiming at gaining insights into data, classification, authentication, process monitoring and prediction, spectroscopy is nowadays often coupled with chemometrics. Measurements can be easily acquired, and they can be presented and organized in the forms of vectors, matrices, and higher-order structures depending on the problem one wants to solve. The inherent complexity in instrumental data was systematically classified by Sanchez and Kowalski [55], employing tensor algebra. In this structured framework, tensors emerge as entities that encapsulate data, with their order indicating the minimum number of indices necessary for meaningful data organization. Zero- and first-order tensors correspond to scalars and vectors, respectively, while second-order tensors, represented by matrices with elements adhering to specific relationships, extend this classification. This systematic approach spans not only the data itself but also encompasses the instruments responsible for data delivery and the methodologies applied in subsequent analyses. The order of instrumentation linked to the form of data acquired is shown in Fig. 2.

Within this context, "error" means the quantitative deviation between a measured value (x) and its "true" value (μ) as in Equation (1). "Uncertainty" assumes the role of statistically characterize this disparity, typically apparent in the context of replicated measurements and expressed through parameters such as variance, confidence intervals, or standard deviation. Conversely, "noise" conveys an organized sequence of errors, each possessing distinct characteristics, such as photomultiplier noise or drift noise. A univariate measurement inherently constitutes a singular sampling of this sequential noise structure [56]. Other important definitions related to metrology in analytical chemistry can be found in Ref. [57].

$$e = x - \mu \qquad \text{[Eq. 1]}$$

Table 1 provides some possible classifications for measurement errors and their descriptions as reported by Wentzell in Ref. [11]. As pointed out in the article, it is important to note that these classifications are not mutually exclusive, and multiple types of noise are typically observed in any given system.

When analyzing spectroscopic data from a multivariate perspective, the most frequently used tools rely on simplified assumptions regarding the characteristics of errors. The statistical representations of these errors are usually based on the assumption of homoscedasticity, which is characterized by a constant variance across a set of variables. On the contrary, heteroscedastic errors (when not all the variables are characterized by a constant variance) are often encountered when measuring with spectrometers. Independent errors arise when the error covariances within a group of variables are all zero whereas non-zero values are indicative of correlated errors.

Studying the nature of multivariate measurement errors offers several benefits. First, it provides insights into the origins of errors specific to a particular instrument or measurement system. This knowledge, in turn, can be used to enhance measurement quality by addressing the limiting sources of error. Second, understanding the characteristics of errors allows for the design of data analysis tools that can optimally handle or minimize these errors to more efficiently extract chemical information [11,15], such as for instance through specific data preprocessing [58]. Finally, the inherent error structure in multivariate data can be propagated through various pre-processing and data analysis steps to gauge its impact on the final result. Knowledge on measurement errors has been proved useful also for other scopes as to compare models [59–61] and final results, and to predict the performance of instruments [36]. Other areas such as wavelength selection and estimation of sample set sizes can also benefit from investigating uncertainty [29]. In section 3, multivariate error calculation and modeling concerning spectroscopic data will be described.

The recognition of different error sources, including errors in sampling, instrumental errors, and errors introduced during modeling, allows for a more comprehensive evaluation of the overall measurement of uncertainty [62]. Uncertainty in measurement refers to the lack of exact knowledge about the true value of a quantity being measured. It encompasses both systematic errors, which are consistent and predictable deviations from the true value, and random errors, which are unpredictable fluctuations in measured values. It also includes the concept of the errors one could make through the estimation phase. Estimating uncertainty involves quantifying both the systematic and random components of error to provide a range of values within which the true value is likely to lie. This can be done through methods such as error propagation, Monte Carlo simulations, resampling methods, or statistical analysis of repeated measurements. By accounting for uncertainty, researchers can communicate the reliability of their measurements and ensure that conclusions drawn from the data are appropriately qualified.

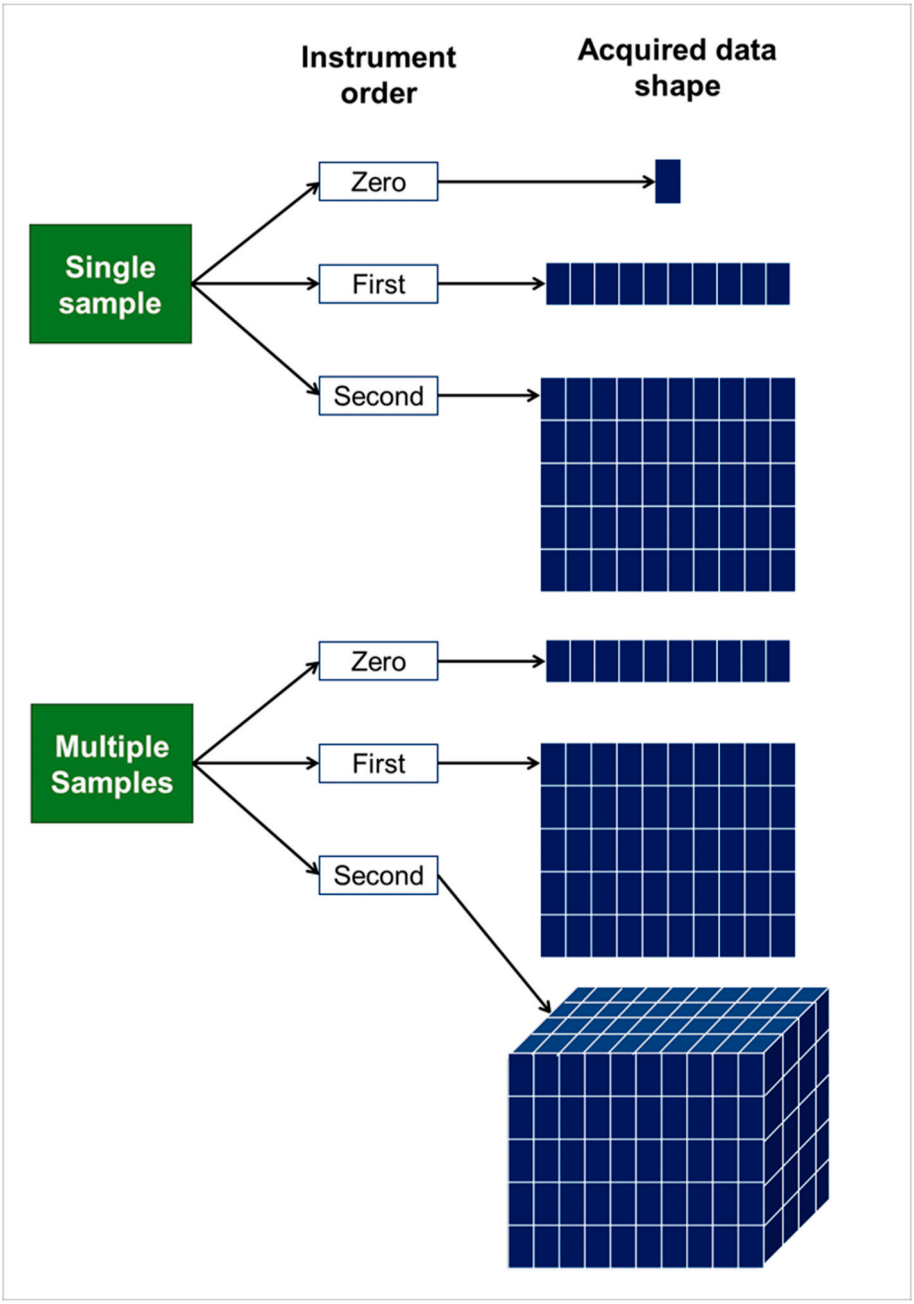

Fig. 2. Representation of the relationship between the nature of data and instrumental characteristics.

In the analytical process each step contributes to the overall uncertainty of models, and careful consideration of these steps is essential for accurate and meaningful chemical analyses. For instance, when calculating a regression model for property prediction, the error from the reference technique would be incorporated into the sources of variance for the models. Depending on the order of the data produced by the instrument, the applicable statistical methods differ, as do the figures of merit to investigate, since they are related to the model being calculated. The selection of the calculation approach should be thoughtful and reasonable, primarily contingent on the statistics of measurements, including the distribution of repeated measurements, the nature of measurement noise (homoscedastic or heteroscedastic), and the processing method [63]. In the case of spectroscopic data, which are intrinsically multivariate measurements, the characterization of uncertainties assumes heightened significance and is essential for successful data analysis. To describe uncertainty, it is necessary to provide the measurement variance for each variable and the covariance between measurement channels, as well as other model parameters. In other words, the inter-variable relationships within the error structures are crucial as well as the error associated with each variable [11,64]. In section 4, uncertainties related to models, including those concerning the final results of the models, will be discussed.

3. Estimation and study of multivariate errors

In the literature, errors are commonly determined by utilizing the mean ($\bar{x}$) of multiple measurements as a substitute for the "true" value (μ). In this context, the definition of replication becomes crucial and the establishment of the level of replication influences the components of error variance that would be captured through experimentation. Indeed, the variance one would include in the data is related to the type of replicates acquired: sampling error, preparation error, technical error,

Table 1
Classification and description of typical experimental measurement errors in spectroscopy. The table has been readapted from Ref. [11].

Type of noise	Description
Independent errors/ uncorrelated errors	Characterized by error covariance equal to zero.
Correlated errors	Characterized by error covariance different from zero.
Homoscedastic errors	Errors with a uniform variance.
Heteroscedastic errors	Errors with a non-uniform variance.
White noise	A vector of uncorrelated measurement errors.
Pink noise (or 1/*f* noise)	A type of low-frequency noise in which the errors in adjacent measurements are more correlated than for measurements that are farther apart.
Drift noise	Low-frequency or correlated noise which implies a slow change in measurement conditions, such temperature.
Source flicker noise	Low-frequency noise that is specifically associated with variations in source signal intensity.
Proportional noise	Heteroscedastic noise in which the standard deviation of the error is proportional to the magnitude of the signal.
Additive noise or offset noise	Correlated noise that randomly shifts the entire signal up or down by a fixed amount.
Multiplicative offset noise	Correlated noise that randomly shifts the entire signal up or down by an amount proportional to the magnitude signal.
Baseline noise	Variance introduced by a variable that displaces displacement of the baseline or the variance of the noise in the baseline regions where there is no signal.
Shot noise	Heteroscedastic noise where the noise standard deviation is proportional to the square root of the signal (photomultipliers).
Digitization noise or quantization noise	Arises from finite precision of analog-to-digital converters.

and/or instrumental measurement error. It is worthwhile to recall that some of these errors may be random errors and some other ones systematic errors. Therefore, careful consideration of how replicates are defined is essential when seeking to characterize measurement errors. In the case where the use of replicate measurements is not an option or it is inappropriate, the multivariate error could be obtained through theoretical calculations but under several assumptions [65]. In the past decades, a theoretical approach that allows for multivariate error simulation was presented [44] and an analytical measurement noise simulation software was published as MATLAB software [66].

The squared sum of all the variance terms included represents the overall variance of the system. For example, Equation (2) explains the variance included when considering instrumental and experimental replicates of the same sample:

$$\sigma^2_{\mathrm{TOTAL}} = \sigma^2_{\mathrm{instrumental}} + \sigma^2_{\mathrm{experimental}} \qquad \text{[Eq. 2]}$$

where instrumental replicates could be intended as multiple scan acquisitions without sample displacement, and experimental acquisition as scans of the sample in which the sample holder is emptied and refilled between scan acquisitions, respectively. In the same scenario, experimental replicates also represent independent aliquots of the same sample.

A widely used approach for characterizing multivariate measurement errors is the error covariance matrix (ECM). The literature describes three distinct methods for estimating this matrix: experimental replication, theoretical prediction, and empirical modeling [11]. The experimental estimation can be easily performed by conducting calculations once a sufficient number of sample replicates is obtained. In contrast, theoretical prediction requires a solid understanding of error sources beforehand, while empirical modeling strikes a balance between directly using replicate measurements and theoretical predictions.

In spectroscopy, if replicate measurements are possible, the procedure to characterize the multivariate measurement errors starts with identifying what would be considered as the true sample spectrum depending on the replicates, r, considered. Subsequently, the residuals vector $\mathbf{e}_i$ for each replicate can be obtained by subtracting from each spectrum ($\boldsymbol{x}_i$) the mean spectrum ($\overline{\boldsymbol{x}}$). A residuals matrix ($\hat{\mathbf{E}}$) can be constructed by collecting the residuals for all replicates. Then, the error covariance matrix ($\boldsymbol{\Sigma}$) can be calculated as the covariance between the residuals as in Equation (3).

$$\boldsymbol{\Sigma} = \frac{\sum_{i=1}^{r} \mathbf{e}_i^{\mathrm{T}} \mathbf{e}_i}{(r-1)} \qquad \text{[Eq. 3]}$$

where, $\mathbf{e}_i^{\mathrm{T}}\mathbf{e}_i$ is the outer product of the residual vector for each replicate spectrum, and gives a matrix that captures the variance and covariances between spectral variables.

From the error covariance matrix, the *error correlation matrix can be calculated* as in Equation (4). It contains the correlation coefficients of the elements of $\boldsymbol{\Sigma}$.

$$\boldsymbol{\Sigma}_{corr} = \frac{\boldsymbol{\Sigma}}{\sqrt{\boldsymbol{diag}(\boldsymbol{\Sigma}) \cdot \boldsymbol{diag}(\boldsymbol{\Sigma})^T}} \qquad \text{[Eq. 4]}$$

Estimates of variance based on experimental data are characterized by considerable uncertainty, primarily stemming from experimental limitations that typically impose constraints on the number of replicates. Therefore, to ensure a reliable estimation, it is crucial to either increase the number of replicates or, alternatively, aggregate error covariance across various sample subsets and then interpret the resulting average covariance matrix ($\boldsymbol{\Sigma}_{\mathrm{avg}}$). The latter approach is generally feasible, provided that the measurement data exhibit minimal changes between samples, which is typically for example in the case of near-infrared spectra of samples from the same origin. A graphical explanation of ECM calculation is furnished in Fig. 3.

Once the variance-covariance and the correlation matrices are obtained, the visual interpretation can take place (Fig. 4). The error covariance matrix describes the correlation between the errors at the different wavelengths. The diagonal of the matrix gives an idea about the uniformity of the errors in the spectra (homoscedasticity), while non-uniform values indicate that errors are not constant along the spectra (heteroscedasticity). The off-diagonal elements give information on the covariance of the measurement errors. The covariance matrix shows the strength of relationships among errors, while the correlation matrix, derived from it, reveals the underlying structure, providing complementary insights. The correlation matrix indicates the structure of the relationship among errors independently from the scale and is a matrix containing numbers ranging between −1 and 1.

Different methods have been proposed to systematically investigate the error matrices [43,67] and to compare those obtained for different experimental conditions [68]. Since errors are often dominated by a bilinear structure, principal component analysis (PCA) and multivariate curve resolution (MCR) can be used to deduce the main structure by allowing simple interpretation related to the magnitude of the error components [69]. A possible quantitative approach to determine the important factors contributing to measurement error was proposed by Leger et al. [43], assuming the underlying idea of the bilinear structure. So, the number of PCA factors required to reconstruct the covariance matrix and their shape should offer clues on the nature of the underlying errors. Typically, there is a finite number of suspected physical factors (like constant offset, linear offset, multiplicative offset). Each target can be represented as a vector and can be projected in the subspace defined for the significant factors of the PCA model and then reconstructed back in the original space. The correlation between the test and projected vectors can be calculated to assess the closeness of the target vector to the subspace.

To assess the contribution of uncorrelated errors, the comparison within the cumulative variance for the PCA on the error covariance matrix, and the cumulative percentage variance for the original residual

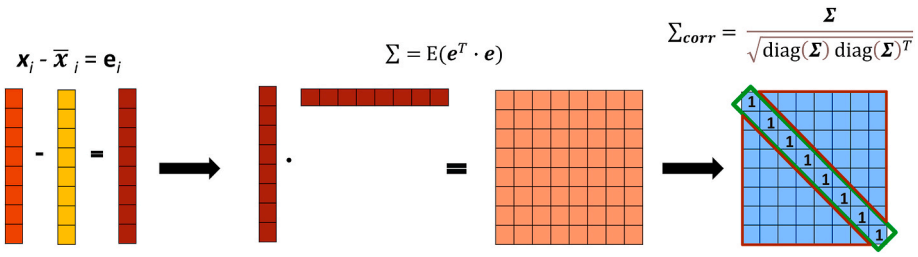

Fig. 3. Representation of the calculation procedure for multivariate error covariance and correlation matrices.

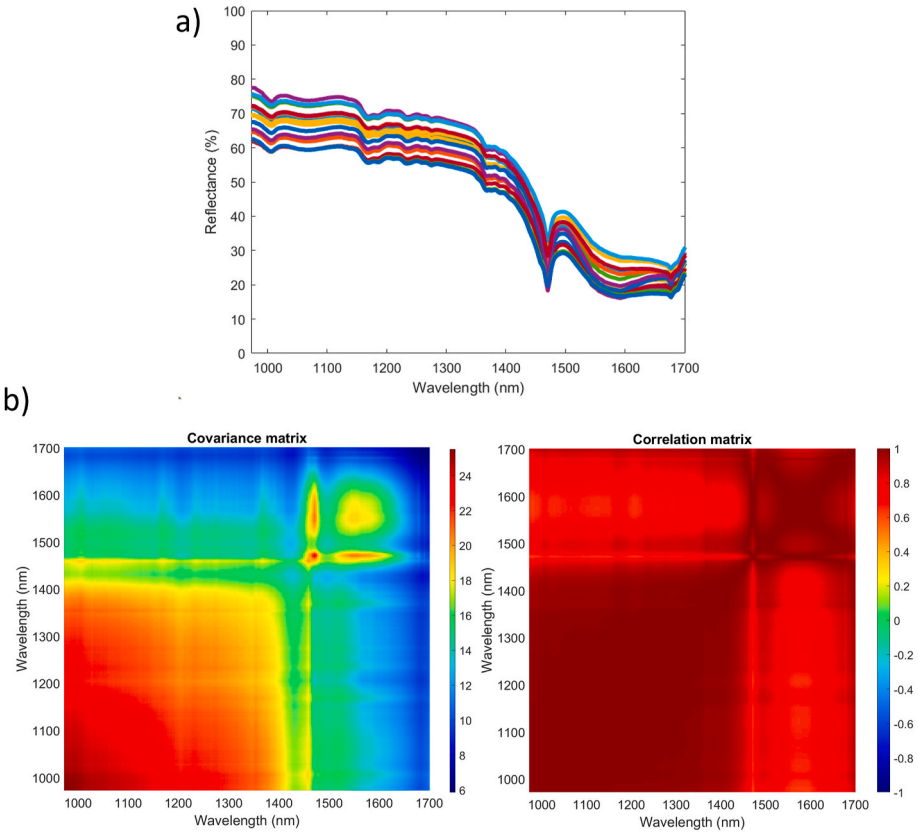

Fig. 4. a) Example of 15 experimental replicates acquired with a NIR portable instrument on a white compact opaque pill for dietary supplementation. The instrument used is an AvaSpec-Mini-NIR (Avantes, Apeldoorn, The Netherlands) coupled with AvaLight-HAL-S-Mini2 source and a reflection fiber probe. b) Measurement error covariance and correlation matrices.

matrix ($\hat{\mathbf{E}}$), can provide insights. The correlations matrices can be resumed and studied through *K* index [70] and image analysis [68].

4. Uncertainty of final results

How the error of the experimental data is incorporated into the final result has been one of the most controversial topics in chemometrics for many years. We mean by final result the scores of a PCA model, the predicted concentration of a PLS regression model or the predicted class in a PLS-DA model, for instance. In this section, more emphasis will be put on prediction and classification results, as they are by far the most usual in the chemometric literature. As stated in the introduction, an analytical result should not be reported without an estimation of its uncertainty. Otherwise, results cannot be compared with established product specifications or even regulatory limits.

4.1. Exploratory models

Data analysis often starts with an exploratory or descriptive analysis, and multivariate spectroscopic data, for instance, are not an exception. Principal component analysis (PCA) is a cornerstone of multivariate exploratory analysis: reducing data dimensionality usually helps unraveling complex datasets and identifying underlying patterns and trends [71]. PCA is not free from sources of uncertainty. However, uncertainties associated to PCA results are seldom estimated. Nevertheless, when it is used as a dimensionality reduction algorithm for subsequent clustering, classification or process control tools, for instance, the propagation of the errors and consequent uncertainty calculation becomes paramount [72].

Using resampling techniques is often the simplest way to estimate uncertainties associated to PCA. Resampling techniques involve the repeated draw of samples from a set and the recalculation of the model of interest on each subset of samples to obtain additional information from the model. In this way, with all these different subsets and therefore different submodels, it is possible to calculate the precision associated to different model parameters (e.g. scores and/or loadings), similar to how the precision of the result for a given sample can be estimated from replicates of that sample. These precision estimates will

then be used to calculate the uncertainty of the desired parameters. For instance, to estimate the variability of the fit of a set of samples to a PCA model, one can draw different subsets of samples from the original data set, fit a PCA model to each new subset and then examine the variation of the different models. By doing this, additional information not available from fitting the model to the original set of samples can be obtained.

Resampling methods can be computationally expensive since they involve fitting the same model many times using the different subsets of the original data. Nevertheless, thanks to the recent strides in computing power, the computational demands of resampling methods are typically affordable [73]. Resampling techniques are versatile and applicable to various statistical problems without assuming a specific data distribution and are able to cope with complicated statistics where no analytical formula is available [74].

The most important resampling methods in multivariate modelling are bootstrap, jackknife and cross-validation [75]. These methods can be mainly used for two purposes: to estimate the uncertainty of the parameters of the calculated models or to assess the reliability or the quality of the models (for instance estimating the mean squared error). In this paper resampling methods will be primarily used in the estimation of the uncertainty of the parameters of the model, although a few comments about the use of resampling methods in assessing the quality of the models will be given.

Bootstrap involves drawing multiple samples with replacement [76] from the original dataset to estimate the variability of a model property or parameter, that can go from scores and loadings to prediction errors. Jackknife systematically leaves out one observation at a time from the dataset and fits a model for each subsample [76]. This process is repeated for each observation in the dataset, allowing for a comprehensive evaluation of the model stability and performance. Jackknife is particularly valuable for estimating bias and variance in statistical estimators, providing insights into the robustness of models and their sensitivity to individual data points.

These resampling techniques have been used to estimate the uncertainty of scores or Hotelling's T^2 values of a PCA model. As reviewed by Castura et al. [72], several authors have used resampling techniques on PCA: for example, Josse et al. [77] from a theoretical perspective, or Preisner et al. [26] for classifying different bacteria using infrared spectroscopy, have proposed using jackknifing and bootstrapping to obtain a population of scores and their uncertainty (an estimated variance of their error).

Another resampling technique, widely used in chemometrics, is cross-validation, which is mostly used to assess the predictive performance of statistical models and less to estimate the uncertainty associated with the models. In cross-validation, the dataset is divided into multiple subsets, and the model is trained and tested iteratively on different combinations of these subsets. When the validation subsets are composed by only one sample, the method is called leave-one-out cross-validation (LOOCV) and when they include two or more samples it is called *k*-fold or segmented cross-validation.

While all three resampling methods described (bootstrap, jackknife and cross-validation) share the goal of assessing the reliability and variability of a model, they differ in their specific approaches. Bootstrap and jackknife focus on sampling from the dataset, with bootstrap using random sampling with replacement and jackknife systematically omitting observations. Cross-validation, on the other hand, emphasizes model evaluation by partitioning the dataset into training and testing sets. Although bootstrap (and also jackknife) is mainly used to estimate the uncertainty or the variability associated to the models, a specific byproduct of bootstrap, out-of-bootstrap (OOB) [78], provides data points not included in a particular bootstrap sample, what is useful for performance evaluation: in each bootstrap sample, certain data points are not selected because of the random sampling with replacement. These data points that were not sampled in a given bootstrap iteration are referred to as "out-of-bootstrap" samples. These OOB samples can be used to evaluate the performance of a model, providing a form of cross-validation.

Focusing briefly on estimating the reliability or the performance of a model, the reviewed resampling techniques are often used in model aggregation (aggregating models based on bootstrap resampling is termed *bagging*): the combination of predictions of a number of different submodels to improve the models, which may be especially useful in case of small sample sizes [79]. Aggregation is used in bootstrap and cross-validation, but also in jackknife [78,80].

Of the resampling methods reviewed so far in this paper, at first glance, LOOCV and jackknife appear to be very similar. It is therefore worthwhile to make a comparison between them, a comparison that encompasses the two main goals of resampling techniques: estimating the variability or uncertainty of the model parameters, and assessing the reliability or performance of the model. Even if LOOCV and jackknife share the commonality of systematically leaving out individual observations for analysis, their primary objectives and methodologies differ. LOOCV, mainly used to estimate the performance of a model, follows the following steps:

a) Dataset splitting: given a dataset with n observations, LOOCV involves creating n different training sets. Each training set consists of $n-1$ observations, leaving one observation out for testing.
b) Model training: a model is trained on each of the $n-1$ observation sets.
c) Model testing: the trained model is tested on the single observation that was left out.
d) Performance aggregation: the performance metric (e.g., mean squared error) is calculated for each of the n tests. The final performance estimate is the average of these n values.

Jackknife, that is mainly used for estimating the bias and variance of statistical estimators, works in the following way:

a) Dataset splitting: similar to LOOCV, jackknife involves creating subsets of the data by systematically leaving out one observation at a time. Given a dataset with n observations, n different subsets are created, each containing $n-1$ observations.
b) Estimator calculation: for each subset, a statistical estimator (e.g., mean or variance) is calculated.
c) Bias and variance estimation: the jackknife estimator of the parameter of interest is obtained by averaging these n estimators. The bias and variance of the estimator can also be derived from these n calculations.

Summarizing, jackknife focuses on estimating bias and variance in statistical estimators, while LOOCV (or cross-validation, generally speaking) is geared towards evaluating predictive models.

A final important remark in resampling techniques is how to deal with correlation between samples. Standard resampling methods, such as for instance standard bootstrapping, cross-validation or jackknife, assume independent observations (i.e., independent rows in the data matrix). However, in real-world scenarios where measurements are correlated due to experimental design or inherent structure (e.g., biological replicates, agricultural data, or spatial clusters), failing to account for these dependencies can lead to misleading conclusions. When dealing with correlated data in resampling methods, it is essential to use techniques that account for the dependencies between observations. Methods like block bootstrap [81], cluster bootstrap [82], stratified bootstrap [83] or multilevel/hierarchical bootstrap [84] for bootstrap resampling, or cluster jackknife [85] or block jackknife [86] for jackknife resampling, are specifically designed for such cases. Methods that assume independent rows, like standard resampling techniques, should not be used in cases with correlation between samples, as they can lead to biased estimates of variability and model performance.

Other methods where the uncertainty of results may play a major

role are soft modelling methods, such as multivariate curve resolution (MCR), often used to explore the data. The particularity of these methods is that they have intensity and rotational ambiguity and the solution they provide is not unique: they offer a range of possible solutions that are equally valid, therefore introducing an uncertainty in the results, as boundaries of the so called feasible bands [87]. Various approaches have been proposed to estimate the range of the feasible solutions, as reviewed by Golshan et al. [88]. Based on this solution uncertainty and the measurement error or noise, Dadashi et al. [89] proposed using the error propagation approach to estimate the uncertainty of results provided by MCR.

4.2. Predictive models

As with exploratory models, the importance of assessing reliability, variability, and figures of merit is of great interest in predictive models as well. In fact, in predictive (and classification) models, the significance of evaluating and characterizing models through validation steps is crucial. Proper validation is fundamental to assess both the robustness of the calibration model and its predictive power when applied to future samples [90,91]. Although validation has been extensively discussed in the literature and goes beyond the scope of this review, it is important to mention that there is a general consensus that robustness—or model reliability—refers to a model's ability to maintain stable performance under varying conditions, such as changes in data quality, noise, or minor input variations. Assessing robustness is key, and internal validation techniques, such as cross-validation or bootstrapping, are commonly employed to evaluate this aspect [92]. Additionally, evaluating model performance on external test sets is essential for assessing predictive power. It is crucial to carefully consider the sample selection method, as findings and chemical insights can be influenced by the type of validation and data analysis methods used.

In multivariate calibration methods, such as multiple linear regression (MLR), principal component regression (PCR) and partial least squares (PLS), errors in the training data impact the estimation of regression coefficients and model parameters. These errors can arise from variations in sample composition, instrumental noise, or other sources. The model is trained to establish the relationship between spectroscopic features and the target variable. When the trained model is applied to new data for prediction (new data that are also affected by errors), the multivariate errors from the training phase are carried forward. The propagated errors manifest as uncertainty in the predicted values. This uncertainty may be expressed as a confidence interval, providing a range within which the true value is likely to fall. Understanding and communicating this uncertainty is crucial for offering adequate analytical methods [45] and for decision-making based on the predicted values.

If the uncertainty associated with a predicted result is very high, reasonable doubts may arise about the validity of this predicted value. Probably no doubts would stem from the same predicted result using the same prediction model if the uncertainty is not calculated. In this way, a predicted value of '30 ppm' would be accepted, but some red flags could be crossed with a result of '30 ± 20 ppm'. There are two main approaches to assess the uncertainty of prediction results from regression models: error propagation and resampling.

Often rooted in statistical principles, error propagation quantifies and propagates uncertainties from input variables to the final predicted values (Fig. 5). These sources of error may include the inputs to the model or the modelling itself [42]. The error propagation approach leads to closed-form expressions, which are highly convenient to specifically consider or neglect different error sources, for the calculation of sample-specific prediction uncertainty, or other figures of merit such as detection and quantification limits [93].

Researchers already started working on error propagation in multivariate methods in the late 1980s. One of the first proposals using error propagation to estimate the prediction uncertainty in regression models (first Ordinary Least Squares, OLS, then extended to PCR and PLS), as described by Bano et al. [25], is the back-propagation approach of

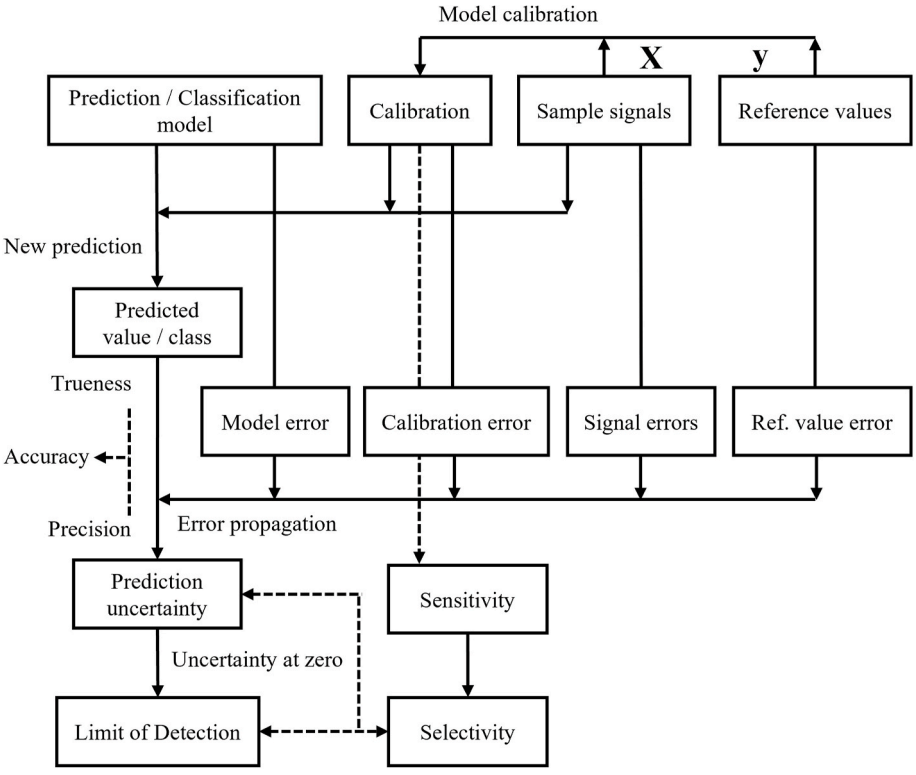

Fig. 5. *Scheme showing how the errors introduced in the model calibration step propagate to the prediction of a new sample, its uncertainty and other figures of merit. Adapted from Bauer* et al. [94].

uncertainty. This is, estimating a new **X**-set for the target **y**-values using the model, and taking the estimation of the uncertainty to propagate it again to the **y**-values. This was done by first estimating the uncertainties in the loadings and scores of the new **X**-set and then extending them to the prediction error including other error sources. However, as estimating both scores and loadings uncertainties at the same time can be complicated (as they are inter-dependant), one of them could be neglected. Another pioneering option is averaging the uncertainties of the scores and the loadings, as CAMO did with the version 5.5 of their 'The Unscrambler' software, using an empirical correction [95]. This approach was criticized by De Vries and Ter Braak [96] for being over-optimistic, and they proposed a new correction, included in version 7.0 of The Unscrambler. However, this proposal was further criticized by Faber and Kowalski [97], as measurement errors in the response and predictor variables (e.g. concentration and spectra, respectively) were neglected, and they proposed an error-in-variables approach that considers them.

After the received critiques, CAMO reviewed the different proposals, as described by Høy et al. [98]. They compared the different methods by using Monte Carlo simulations on synthetic datasets and concluded that the original CAMO method [95] was indeed over-optimistic, but that the approach proposed by Faber and Kowalski [97] requires knowledge about the data that is not always available, such as the variance of the noise. They also concluded that the empirical correction proposed by De Vries and Ter Braak [96] and previously mentioned, worked reasonably well with low noise levels as an estimator of the prediction uncertainty, so they continued implementing it.

Based on the previous work of Lorber and Kowalski [9], as also did Karstang et al. [99], Faber and Kowalski [100] postulated a new and more comprehensive formulation for the estimation of the uncertainty or variance of prediction error (s_i^2), which was further expanded by Andersen and Bro [42] and is described in Fig. 6.

In Fig. 6, N is the number of samples in calibration, h_i, is the leverage of sample i, s_e^2 is the variance of the model error (which can be calculated as suggested by Faber et al. [101]), $s_{\Delta \mathbf{y}}^2$ is the variance of errors in the **y**-measurements, **b** is the regression coefficient vector and $s_{\Delta \mathbf{X}}^2$ is the variance of errors in the **X**-measurements (which are assumed to be iid). Andersen and Bro [42] proposed this calculation to differentiate between the actual prediction error (proposed by Faber and Kowalski [100]) and the apparent prediction error, which includes the error in **y**.

However, for the sake of practicality, usually not all sources of uncertainty are considered and the equation shown in Fig. 6 is simplified. For instance, Faber and Kowalski [100] considered a simplified calculation to estimate the uncertainty of a prediction. By neglecting the signal measurement errors and the measurement error in the reference concentration, the calculation only needs the root mean squared error of calibration (RMSEC) and the leverage of the unknown sample (h_i, scaled by the number of samples in calibration):

$$s_i^2 \approx RMSEC^2(1 + h_i) \qquad \text{[Eq. 5]}$$

Starting from this approach, the calculation of the confidence interval for a prediction can be extended, introducing the Student's t distribution, as described by Boqué et al. [102] and Faber et al. [103]: $y_i \pm t_{\alpha,\nu} \cdot s_i^2$, where y_i is the predicted value for the ith sample and $t_{\alpha,\nu}$ is the t-statistic for α confidence level and ν degrees of freedom ($\nu = N\text{-}r\text{-}1$, where r is the number of components of the model). Other simplifications that arrive at a similar estimation have also been proposed, such as the one proposed by Faber and Bro [104], which suggests not neglecting the reference concentration error. This simplification has been used in real infrared spectroscopy applications. For instance, Skou et al. [105] applied it on predictions of water quality using near infrared spectroscopy and PLS. Bu et al. [31] tested and compared several of the mentioned proposals on the prediction of pharmaceutical tablet quality using near infrared and PLS, stating the advantages and disadvantages of each method.

More recent studies on this topic have focused on measurement errors and their propagation. A unification of the available proposals for uncertainty estimation was published by the IUPAC (International Union of Pure and Applied Chemistry), in a technical report by Olivieri et al. [45]. Allegrini and Olivieri [17,19] took this report as a basis to propose other simplifications based on measurement errors and their heteroscedasticity or fulfilment of the iid assumption. These errors and their propagation were described more in depth by Allegrini et al. [30].

Using the error propagation approach, some authors have also proposed the calculation of different figures of merit. The figures of merit, metrics that quantitatively evaluate model performance, can be key parameters to assess the reliability of regression models, since they offer a deeper understanding of the predictive capabilities of a regression model [93]. For instance, an expression that addresses the measurement uncertainty in the calculation of sensitivity has been proposed by Fragoso et al. [21]. Additionally, expressions that consider the uncertainties of the model have been proposed for calculating the selectivity of a method for an analyte, as summarized by Valderrama et al. [106]. For more extensive description of figures of merit of a multivariate model, the reader is directed to recent reviews such as the ones by Olivieri [93, 107], Olivieri et al. [57] or Allegrini and Olivieri [17,108], that offer a holistic view of these metrics, their calculation and presentation.

However, the figure of merit that has been more prominently studied is the limit of detection (LOD) of a multivariate regression model, as it is a natural extension of the prediction uncertainty, but for zero concentration. The LOD is defined as the minimum detectable concentration for an analytical method and is considered "a fundamental performance characteristic of a chemical measurement process" by the IUPAC [45]. The LOD can be estimated in different ways considering the uncertainty of predictive models, as shown by Boqué and Rius [109]. These authors previously reviewed the diverse approaches for the LOD calculations up to 1996 [110]: the net analytical signal approach [111], the confidence interval of the predicted concentration approach (proposed by Faber and Kowalski [100]) and the error propagation approach.

Combining the use of the predicted concentration and error propagation approaches, Boqué et al. [102,112] defined the LOD by testing the hypothesis that the 0 value and the predicted concentration are statistically equal or statistically different (using a Student's t-test), determining the probabilities of type I and type II errors (Fig. 7). This method has been widely used in real spectroscopic applications, such as the ones described by Wu et al. [113] and Du et al. [27].

In a parallel way, based on the simplified calculation of the prediction uncertainty proposed by Faber and Bro [104], Alcalà et al. [114] extended these calculations to the LOD and applied it to a prediction

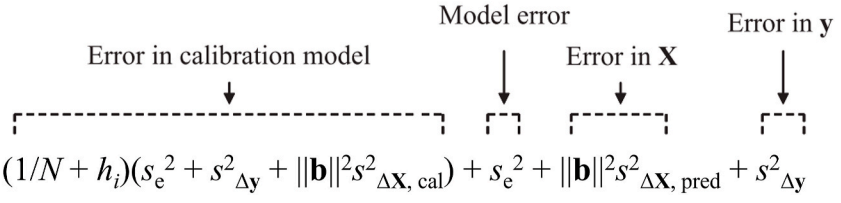

Fig. 6. *General equation for the estimation of the uncertainty of a prediction for a multivariate regression model, adapted from Andersen and Bro* [42].

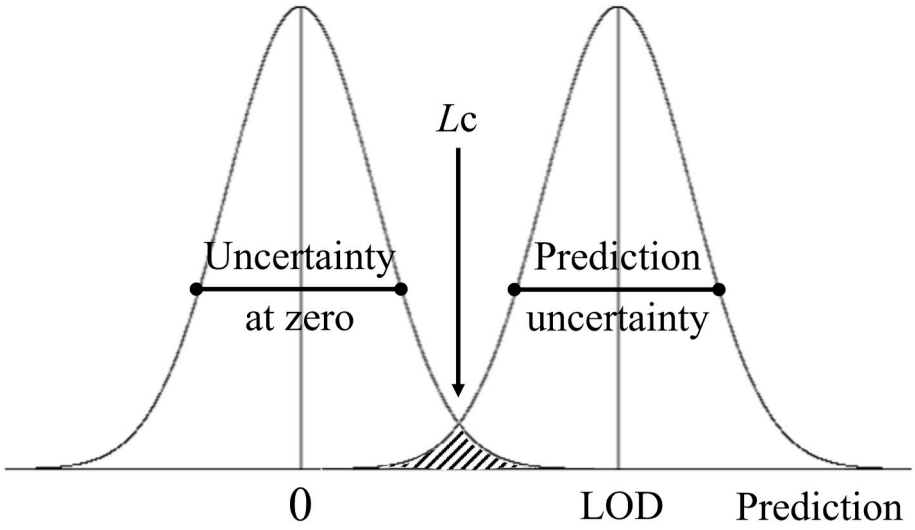

Fig. 7. *Estimation of the Limit of Detection (LOD) based on the minimal predicted value that is significantly different from zero considering both uncertainties, where Lc is the critical limit. Adapted from Olivieri* et al. [45].

method of a pharmaceutical tablet dose using near infrared spectroscopy. Following a similar method, Allegrini and Olivieri [17–19] extended the calculation of the LOD from the definition of prediction uncertainty proposed by Olivieri et al. [45].

The other approach to calculate the uncertainty of prediction results involve the use of resampling. As mentioned above, resampling techniques can cope with complex statistics where no analytical equation is available, and they have proven useful to estimate the uncertainty associated with predicted values [115]. Denham [116] used bootstrap and cross-validation to calculate the intervals of the predicted values obtained using a PLS model. The bootstrapped samples for the training set and for the test or validation set were constructed by bootstrapping the residuals from the PLS model calculated with the training set. Each set of bootstrapped samples was used to calculate a new PLS model, a new set of predictions of the future samples and therefore a new set of prediction errors (the difference between the new predictions of the future samples and the bootstrapped future samples). The empirical distribution function of the prediction errors ($\widetilde{G}$), which can be calculated in the bootstrapping process, is an estimate of the cumulative distribution function of the prediction errors for the *y*-predicted value. Finally, this cumulative distribution function of the prediction errors for a certain level of confidence is used to calculate the prediction interval for the *y*-values of new samples. In the same paper [116], cross-validation was used in a resampling procedure to estimate a studentized residuals [117] used in the prediction interval for the *y*-values of new samples. The authors compared the intervals of the predicted values obtained using bootstrap and cross-validation with naive prediction intervals (prediction intervals that are calculated without considering the uncertainty in the estimates of the model parameters [118]) and local linearization prediction intervals. It is important to mention that the choice of the number of PLS factors significantly affects the prediction intervals. Real and simulated data were used to compare the reliability of the different prediction intervals. The performance of bootstrap confidence intervals was globally poor. The performance was worse in cases with a few number of samples, which is not surprising due to the resampling procedure, since only a good approximation to the structure of the errors would be possible for moderately large number of observations. Cross-validation prediction intervals underestimated coverage for large number of factors while overestimated coverage for small number of factors. The article was published more than 25 years ago, at a time when computing power was much less developed than today, and some of the conclusions drawn refer to these computational difficulties, which would not be so relevant today.

Zhang and Garcia-Munoz [10] proposed a simpler procedure including bootstrap resampling. The authors used bootstrapping by residuals to generate $B = 10{,}000$ bootstrapped sets of **X** and **y** data to estimate a PLS regression coefficient vector $\widehat{\boldsymbol{\beta}}_{\boldsymbol{b}}$ for each one of the 10,000 bootstrapped data sets. After the B iterations, the covariance matrix of the estimate of the regression coefficient vector $\widehat{\boldsymbol{\beta}}$ can be calculated as:

$$var(\widehat{\boldsymbol{\beta}}) = \frac{1}{B-1}\sum_{b=1}^{B}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{b}} - \overline{\boldsymbol{\beta}})(\widehat{\boldsymbol{\beta}}_{\boldsymbol{b}} - \overline{\boldsymbol{\beta}})' \qquad \text{[Eq. 6]}$$

where $\overline{\boldsymbol{\beta}}$ is the average of the bootstrapped PLS values. The covariance matrix can then be used to calculate the variance of the PLS prediction error, and then, using the error propagation theory, the confidence interval (assuming a Student's *t*-distribution) can be found. The authors also mentioned the importance of estimating the correct number of degrees of freedom, since this is a key parameter in the calculation of the prediction uncertainty. Three approaches for the estimation of degrees of freedom were considered: naive approach (which globally estimated the degrees of freedom inaccurately), pseudo degrees of freedom and generalized degrees of freedom. The authors applied the bootstrapped prediction intervals, together with three other methods to four data sets from the pharmaceutical industry. The other three methods applied were an approximation method using ordinary least squares type expressions [97], linearization-based methods with three different proposals for the estimation of the Jacobian matrix [116,119,120] and the U-deviation method used in 'The Unscrambler' chemometrics software (CAMO, Trondheim, Norway) [95]. Globally speaking, and according to the authors [10], the best uncertainty estimates were obtained with the approximation method using ordinary least squares type expressions, with bootstrap uncertainties being slightly larger than expected. One of the reasons for this overestimation may be the estimation of the variance associated to the *y*-values. The authors used the RMSEC as this estimation, and they claim that RMSECs obtained from reference testing methods tend to overestimate this variance.

Faber [121] also used resampling techniques to estimate the uncertainty of the regression coefficients, which also play a significant role in the estimation of the uncertainty associated to the prediction results. An approximate variance expression was compared with four resampling methods: jackknife, bootstrapping objects, bootstrapping residuals and noise addition with the Monte-Carlo simulation technique, using real and simulated data sets. The best results were obtained with the approximate formula, bootstrapping residuals and noise addition, stressing that the best resampling methods for uncertainty estimation

are those that work with noise, not with objects.

4.3. Classification models

Multivariate classification models stand as powerful tools for discerning patterns, making predictions about the membership of a sample to a class, and categorizing complex datasets [122]. From medical diagnostics to food authentication or industrial quality control, the applications of these models are vast and impactful.

As in the case of multivariate calibration, the outcomes of multivariate classification models should also be accompanied by an estimation of their uncertainty, or reliability. The reliability or uncertainty of a classification result refers to the degree of confidence or doubt associated with the correctness of the predicted class assignment made by a classification model. Understanding the uncertainty of classification results is essential for making informed decisions based on model predictions, as it provides additional information beyond accuracy, especially in situations where misclassifications may have significant consequences.

However, unlike multivariate calibration, where the predictions are continuous values (e.g. concentrations), in multivariate classification the predictions are discrete, or categorical (e.g. class 1 vs class 2 in a two-class classification scenario). This leads to a different estimation of the classification uncertainty, usually in the form of a probability of correct classification.

Beyond global measures of classification performance like sensitivity, specificity, and accuracy [123], specific measures of classification uncertainty have not been so much studied in the chemometric literature. However, some approximations can be found, which can be divided, as in the case of multivariate calibration, in two groups: those using equations and those based on resampling methods.

In the first group, the most popular classification method is the one based on the Bayes theorem [124]:

$$P(\omega_c|\hat{y}) = \frac{p(\hat{y}|\omega_c) \times P(\omega_c)}{\sum_{c=1}^{C} p(\hat{y}|\omega_c) \times P(\omega_c)} \quad \text{[Eq. 7]}$$

where the summation goes from class $c = 1$ to C, the total number of classes. $P(\omega_c|\hat{y})$ is the (a posteriori) probability that a sample with prediction $\hat{y}$ belongs to the class ω_c. $p\ (\hat{y}|\omega_c)$ is the conditional probability, that is, the probability of observing a value of $\hat{y}$ given a sample from class ω_c. Finally, $P\ (\omega_c)$ is the prior probability, that is, the probability of observing class ω_c in the future, without any given conditions. For a binary classification (i.e. two classes: 0 and 1), the probability of classification of a sample in class 0 given a prediction $\hat{y}$ can be formulated as:

$$P(\omega_0|\hat{y}) = \frac{p(\hat{y}|\omega_0) \times P(\omega_0)}{p(\hat{y}|\omega_0) \times P(\omega_0) + p(\hat{y}|\omega_1) \times P(\omega_1)} \quad \text{[Eq. 8]}$$

Prior probabilities for each class, $P\ (\omega_c)$, are usually taken as the proportion of samples of each class in the calibration set, that is, the number of samples of a given class divided by the total number of samples in the calibration set. In some cases, if no prior information is available, it is common to assume all classes are equally likely. This is called a uniform prior, where all classes are assigned equal probability, that is, $P\ (\omega_c) = 1/C$. Therefore, to calculate the probability of a classification, $P(\omega_c|\hat{y})$, the problem reduces to calculate the conditional probabilities, $p\ (\hat{y}|\omega_c)$, for each class.

The Bayes' rule is closely connected to two important figures of merit in classification, the positive predicted value (PPV), also known as precision, and the negative predicted value (NPV). PPV measures the proportion of positive predictions (i.e. predicted as class 1) that are actually correct. NPV measures the proportion of negative predictions (i.e. predicted in class 0) that are actually correct. These metrics directly measure the probability of a predicted class being correct and can be seen as the application of Bayes' rule for updating probabilities based on test results. PPV tells us how likely a positive test result is correct given the prior probability and the performance (sensitivity and specificity) of the test. NPV tells us how likely a negative test result is correct given the same factors. In many practical scenarios, Bayes' rule allows to adjust these predictive values based on prior knowledge. For example, in cases with very low prevalence (low prior probability of the positive class, rare conditions), even a test with high sensitivity and specificity can have a low PPV, meaning a positive result is not very reliable.

By using Bayes' rule, predictive values (PPV, NPV) can be updated to reflect changes in prevalence or other factors affecting prior probabilities, leading to a more accurate understanding of test results or model predictions.

Following with the binary classification problem and if we take a very well-known classification method, PLS-DA, the problem can be stated as follows. Given two classes, ω_0 and ω_1, the PLS-DA model seeks to separate the two classes using a y-block where samples of class ω_0 are assigned values of zero and samples of ω_1 are assigned values of one. After building and validating the PLS-DA model, y-values predicted ($\hat{y}$) on the calibration set for each class are obtained. In the absence of outliers, and if the classification model is correct, those values range around zero and one, respectively. In the simplest scenario, we can fit those predicted values using two separate Gaussian distribution functions, by taking the mean and standard deviation of the $\hat{y}$ values for each class. Those Gaussian distributions allow to compute the conditional probabilities of an individual $\hat{y}$ value for classes, ω_0 ($p\ (\hat{y}|\omega_0)$) and ω_1 ($p\ (\hat{y}|\omega_1)$, respectively, with the assumption that the Gaussian distributions are representative of the true distributions of all samples in the populations of ω_0 and ω_1. The approach described is the one used in the PLS_Toolbox (Eigenvector Inc.) for MATLAB (MathWorks Inc.) [125].

For binary classifications, a refinement of the previous method was proposed by Pérez et al. [126], who calculated a potential (kernel) function for each calibration sample, with the shape of a Gaussian curve, a commonly used kernel function, centered at $\hat{y}$ (the fitted value of each calibration sample predicted by the model) and with standard deviation (smoothing parameter) equal to SEP_i (what corresponds to s_i in equation (5)). Then, the probability density function (PDF) of each class, $p\ (\hat{y}|\omega_c)$ was obtained by averaging the potential functions of the I_0 and I_1 training samples of class ω_0 and class ω_1, respectively. Finally, the reliability of the classification was calculated from the area under the curve $p\ (\hat{y}_u|\omega_c) \times P\ (\omega_c)$, taking into account the standard error of prediction of $\hat{y}_u$, where $\hat{y}_u$ is the prediction for the unknown sample calculated with the optimal PLS-DA model.

This method was adapted by Botella et al. [127], who used microarray data and PLS-DA with the option to reject classification of biological samples. The method used kernel-based PDFs and the Bayes rule to classify samples while keeping the option of not classifying a sample if this could not be done with sufficient confidence. With this approach, only those samples having the highest probability of being correctly classified are indeed classified, whereas doubtful samples are rejected. The reject option tackles situations where the strict application of the Bayes rule may be questioned. In addition, the optimal model was found by simultaneously minimizing the misclassification and rejection costs, so the methodology involves evaluating the probability of each classification together with the overall cost.

Finally, the method was extended to multi-class classification [128]. In this case, the multi-classification problem was split into binary classification problems using probabilistic PLS-DA models. The results of these models were combined to obtain the final classification using the strategy one-against-one and the principle of winner-takes-all. The classification criterion used the position of an object in the multivariate space and its prediction uncertainty to estimate the reliability of the classification.

A somewhat similar approach but using the spectra instead of the predicted $\hat{y}$ values, has recently been proposed by Fearn et al. [129]. There, multivariate normal distributions were used to fit the spectral data, previously reduced in dimension using PCA. The procedure was

used to classify animal feed ingredients using NIR spectroscopy. More recently, the same authors performed an interesting comparison between four discriminant methods that produce classification probabilities to quantify the uncertainty of the results: Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Kernel Bayes and Logistic Regression [130]. Logistic regression directly models the probability, expressed as log odds, of class membership as a linear function of the vector of spectral data [131]. Kernel Bayes builds within-class distribution models by centering a Gaussian distribution (a kernel) on each data point of the training set and then averaging these distributions over each class to construct the probability distributions [129]. The authors applied the different methods to the in situ authentication of Iberian pig carcasses using NIR spectroscopy, and showed that LDA was the best classification model providing accurate classification uncertainties.

Toher et al. [132] compared different model-based classification methods, including LDA and QDA, to PLS-DA to classify pure and adulterated honey samples, and concluded that both types of discriminant analysis methods showed good classification results.

Another popular classification method is SIMCA (Soft Independent Modelling of Class Analogy). The first attempt to make SIMCA probabilistic, that is, able to provide the uncertainty of classification, was made by Van der Voet and Doornbos [133]. The authors applied kernel density estimation to the scores of the PCA class models and then transforming the probability densities to posterior probabilities for class membership using the Bayes equation.

In SIMCA, two distance metrics are usually calculated to evaluate the membership of a new object to a given class: the orthogonal distance (*OD* or *Q*), that is, the squared Euclidean distance from the measurement vector (i.e. spectrum) to its projection onto the PCA class subspace and the score distance (*SD*), that is, the squared Mahalanobis distance between this projection and the center of the PCA class model [134]. In other words, each sample is characterized by two statistics, Q and Hotelling T^2, which measure the information not included or included in the model, respectively. Class limits for both statistics can be calculated for each predefined class at a specific significance level (α). From these metrics, whether taken individually or combined, SIMCA computes the probability of each sample to belong to each possible class. From a sample to be classified, the confidence level associated to its T^2 and/or Q values is estimated from the distributions in the calibration data and then the confidence limits are converted into probabilities of classification. Different options exist, depending on the metric used, as for instance implemented in the PLS_Toolbox (Eigenvector Inc.) [125].

At this point, some words are needed to highlight the importance of sample sizes (both in training and validation sets) when evaluating the performance of classification models. The topic has been studied by Beleites et al. [135], who evaluated performance using learning curves, which graphically represent how a classification model's performance (i.e. sensitivity in their study) improves or stabilizes as the amount of training and/or test data increases. Learning curves are crucial for diagnosing model behavior, assessing capacity, and determining whether more data or parameter tuning is required for optimal performance. The authors also calculated the necessary test sample sizes for various testing scenarios, focusing on defining acceptable confidence interval widths for true sensitivity and the number of test samples required to demonstrate the superiority of one classifier over another. The other large group of methods for calculating the probability of classification includes those that use resampling. Resampling techniques have been described in the previous section and we will not give more details here. The most relevant applications for different classification techniques will simply be reviewed and discussed. A difference of using resampling methods for the estimation of the uncertainty associated with classification models is that resampling methods usually provide a confidence interval associated to the classification results, and not a probability of correct classification.

An interesting work was proposed back in 2008 by Preisner et al. [26]. The authors compared three classical methods: PCA, PLS-DA and SIMCA, for bacteria discrimination based on FT-IR spectra (despite PCA not being a classification method but an exploratory method). In all cases, the uncertainty of classification was calculated using two non-parametric resampling methods: jackknife and bootstrap.

For PLS-DA it is worth remarking on the work done by Almeida et al. [136], who discriminated between authentic and counterfeit banknotes using Raman spectroscopy and calculated the uncertainty of classification using residual bootstrap. The residuals are bootstrap generated from random substitutions with replacement of the initial values and added to the predicted $\hat{y}$ values to generate new y values, from which a new PLS-DA model is calculated and new residuals obtained. The process is repeated B times and the percentiles of the F-distribution, for a given significant level α, are used to estimate the confidence interval of the predicted $\hat{y}$ values.

A similar approach was applied by Rocha and Sheen [137] for a QSAR (Quantitative Structure–Activity Relationship) model to predict biodegradability for a set of substances, and by Rocha et al. [138] to estimate the uncertainty of the scores of PCA models and the uncertainty of classification of PLS-DA models to separate and classify environmental metabolomics NMR spectra. Finally, Morais et al. [139] used the same approach to estimate misclassification probabilities for PCA-LDA, PCA-QDA and PCA-SVM models.

For *k*NN (*k*-nearest neighbors) classification, the most common way to estimate the probability of a given object to belong to class ω_c is to calculate the ratio k_c/k, where k_c is the number of nearest neighbors of class ω_c and k is the number of neighbors considered for classification [124]. An improvement of the method was developed by Villa et al. [140], who proposed the probabilistic bagged *k*NN (PB*k*NN) method, combining *k*NN and bootstrap, to calculate the reliability of classification as a posterior probability. The method was applied to different public datasets and successfully compared to standard methods such as Bayes rule and LDA.

5. Turning errors into insights: redefining model development through error

As presented before, literature provides evidence that heteroscedastic and correlated measurement errors are widespread in analytical measurements, with the extent of deviation being a significant factor. Therefore, while it is common practice to assume homoscedastic noise in measurements, this assumption does not hold true for analytical instruments in many instances and also the assumption of independence in measurement errors is often violated since there is correlation and covariance between channels [34,64].

When measurement errors follow a normal distribution and are independent and identically distributed, models that maximize total variance are usually best for pinpointing chemical sources of variance, assuming no other information is available. However, in cases where errors are more complex including information about the type and the magnitude of the errors in the models could help extract chemical information better by focusing on variance components less likely to be noise-related. This idea of dealing with the measurement error before applying classical chemometrics methodology is sort of hinted when data are preprocessed. Indeed, preprocessing is often used to remove sources of variability in the data that could not be attributed to the study goal [29]. Preprocessing methods that consider error structures have been proposed [58]. One such method, introduced by Paatero et al. in 1993, involves weighted data scaling before PCA modeling [141]. This preprocessing approach has also been investigated for multivariate curve resolution (MCR) analysis of data affected by heteroscedastic noise [39].

Even if in practical terms, it is often assumed that the variance stemming from analytical uncertainty is negligible compared to the variance across samples, leading to its omission in data analysis, alternative data analysis methods have been suggested to account for the

noise structure within the data [142]. Especially throughout the 1990s, various modified algorithms emerged in the literature, some of which have been applied to chemical data [47]. However, the diffusion of such methods is limited since, up to date there is no such software solutions that enable to estimate exploration or calibration models with methods that could take into account different errors. Some of the articles cited in this paragraph include MATLAB codes for using the models proposed. Some codes and a multivariate noise simulation software are present at the website: http://groupwentzell.chemistry.dal.ca/software.html.

In exploratory analysis, maximum likelihood principal component analysis (MLPCA) has been proposed as an extension of PCA for data exhibiting non-homoscedastic noise structures [47]. MLPCA requires prior knowledge of the noise structure of the data. Mathematically, MLPCA has also been shown to be equivalent to total least squares (TLS) [143]. From a practical perspective, MLPCA is not a single method but rather a set of methods applicable to different scenarios. Several applications were explored for these methods [37]. In a general manner, PCA typically provides more reliable estimates when measurement error variance is uniform (homoscedastic noise), whereas MLPCA performs better when the error covariance matrix is explicitly known. A comparison within results obtained with PCA and with MLPCA in presence of heteroscedastic error can be seen in Fig. 8.

Other factor analysis methods, such as maximum likelihood common factor analysis (MLFA) and principal axis factoring (PAF), are valuable for handling errors. They offer the advantage of providing information about measurement uncertainty and can adapt to situations involving unknown heteroscedastic errors, eliminating the need for scaling [145,146]. An example of application is presented in Fig. 9. The dataset represented consists of spectral data with high degree of heteroscedasticity (a). In Fig. 9(b) scores plots are represented by four methods assuming three factors' models: MLPCA (using the error variance estimated from replicate measurements), PAF, MLFA *and PC*A. Not surprisingly, PCA performed poorly in this regard since the heteroscedastic noise had a significant influence on the results. However, it is surprising that PAF also did not reproduce the design matrix well, given the similarity of its results with MLFA in the simulations.

Traditionally, when developing multivariate calibration models, it is common to assume that the instrumental noise structure follows an iid pattern, although this assumption seems to be more of an exception than a rule. Maximum likelihood principal components regression (MLPCR) serves as the calibration counterpart to MLPCA. MLPCR differs from standard PCR models in two main aspects: first, it employs maximum likelihood principal component analysis (MLPCA) rather than PCA for decomposing the calibration data matrix. Second, it utilizes a maximum likelihood projection instead of an orthogonal projection into the PCA subspace during the prediction step. By incorporating information from the measurement error covariance matrix, MLPCR can effectively distinguish chemical variance from other sources, resulting in a more dependable model [41]. In fact, MLPCR has demonstrated notably superior predictive capability compared to PCR and PLS, especially in spectroscopic applications where heteroscedastic and/or correlated measurements are prevalent [147].

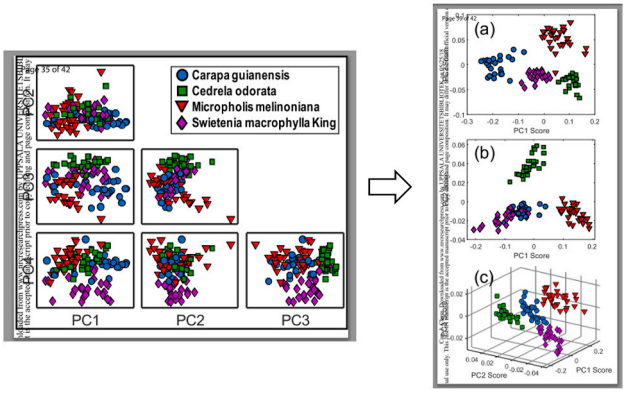

Fig. 8. *Left) Paired scores plots from principal components analysis of sample mean spectra after column mean-centering, with species identified as in the legend. Right) Scores plots from maximum likelihood principal components analysis (MLPCA) of NIR spectra. (a) Rank 2 MLPCA results using class specific error covariance matrices (ECMs). (b) Rank 2 MLPCA results using a global average ECM. (c) Rank 3 MLPCA results using a global average ECM. Reproduced with permission from Ref.* [144].

In 2018, Allegrini et al. [35] introduced and evaluated a new penalized regression model that integrates the error covariance matrix as information in the calibration algorithm. This approach, called error covariance penalized regression (ECPR), employs the error covariance matrix (ECM) as a penalization term, and it has been tested on both simulated and experimental datasets. ECPR outperforms traditional first-order multivariate methods like ridge regression (RR), principal component regression (PCR), and partial least-squares regression (PLS), particularly under non-iid conditions.

Calibration models using latent variables find extensive application in process monitoring scenarios [148]. Here, errors can exhibit highly complex structures due to various sources such as sampling error, sampling bias, operator variability, intricate analytical protocols, and instrumentation uncertainty that influence the overall errors. In response, other alternative approaches to PLS that account for errors have been developed by Reis et al. [13,23,33].

Subsequently, the maximum likelihood treatment of measurement errors has been extended to multiway methods like parallel factor analysis (MLPARAFAC) [149] and maximum likelihood via iterative least squares estimation (MILES) [150]. Nevertheless, these extensions become complex since additional orders are introduced, leading to an expansion of potential error covariance structures to be considered and requiring greater memory usage to manage error covariance matrices in unfolded data.

In a broader context, Bayesian statistics can significantly enhance the ability to incorporate domain-specific knowledge for obtaining more accurate and useful models, presenting numerous research opportunities along with challenges. While tutorial articles demonstrate the potential benefits for various experimental chemical data [151–153], widespread adoption remains challenging for the average user due to the lack of available software capable of incorporating error structures during the modeling phases.

6. Conclusions

Chemometrics is a field within Analytical Chemistry that focuses on the application of mathematical and statistical methods to multivariate chemical data. However, for its complete integration into Analytical Chemistry, it must adopt practices consistent with the field. This implies providing the uncertainty of analytical results and the figures of merit of the multivariate methods developed, which are often neglected in many chemometric applications documented in the literature.

Nowadays, many researchers and people in industries apply multivariate analysis methods. But a method without, for example, an estimate of its bias or limit of detection may be of no use. In Analytical Chemistry, these aspects are routine, but not in chemometrics. For chemometrics to incorporate to routine methods, reference methods or standardized methods, it must offer the possibility of providing everything that an analytical method based on univariate statistics does.

Furthermore, the study of the measurement error of multivariate raw data is often underestimated, while it is fundamental for the understanding of the data itself. Just as in analytical methods using univariate calibration models the measurement error is evaluated (e.g. by studying the residuals of the calibration line), when using multivariate methods this is even more important, since measurement errors can be heteroscedastic but also correlated. Not taking this into account can lead to the choice of an inappropriate multivariate method and, in the long run, to a loss of accuracy in the results.

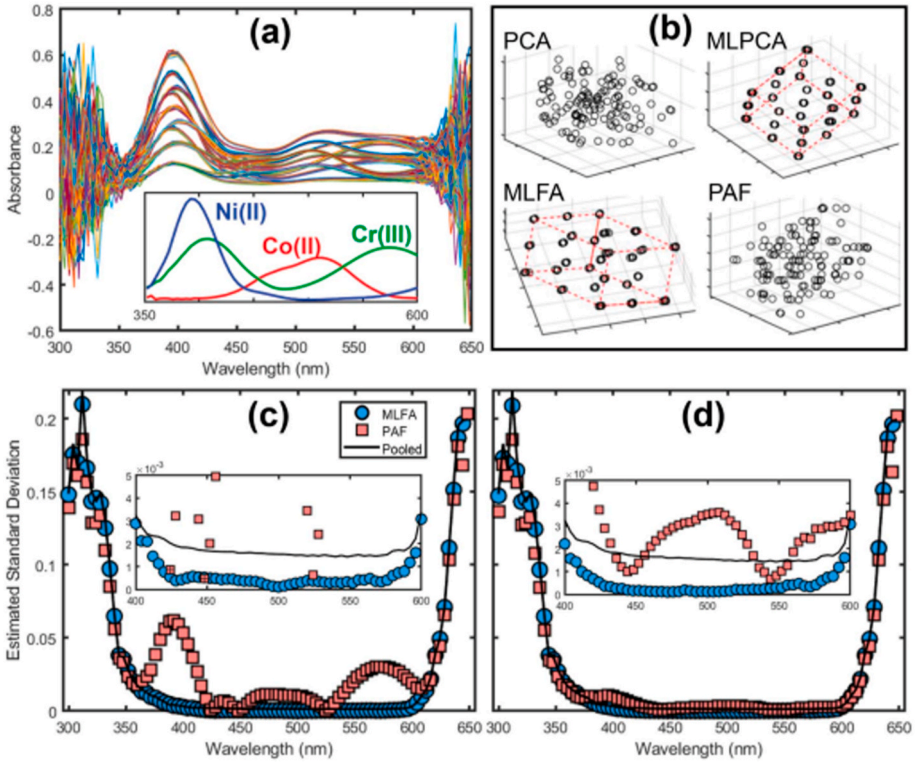

Fig. 9. *(a) Metal ion spectra, with pure component spectra inset (normalized to mean concentrations); (b) scores plots for four methods using a 3-factor model; (c) measurement error standard deviation estimates for MLFA and PAF compared to pooled estimates from replicates for a 3-factor model; (d) error estimates as in (c) for a 4-factor model. Reproduced with permission from Ref.* [146].

For all that, this review basically aims for two things. First, it urges scientific community that develops and applies multivariate methods of analysis to study the measurement error both in the predictor and predicted variables. Such errors have significant implications for preprocessing spectral data and selecting appropriate modelling methods. Second, it calls for the scientific community presenting analytical results obtained with multivariate methods somehow calculate the uncertainty of those results. Results without uncertainty lack comparability and undermine their significance.

This review offers different alternatives to achieve this, adapted to different types of errors in the data and to different types of multivariate analysis methods (exploratory, classification or quantification).

CRediT authorship contribution statement

Barbara Giussani: Writing – review & editing, Supervision, Methodology, Conceptualization. **Giulia Gorla:** Writing – original draft, Investigation. **Jokin Ezenarro:** Writing – original draft, Investigation. **Jordi Riu:** Writing – original draft, Validation, Investigation. **Ricard Boqué:** Writing – review & editing, Supervision, Methodology.

Declaration of competing interest

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

Acknowledgements

Chemometrics and Sensorics for Analytical Solutions (CHEMOSENS, ref.2021 SGR 00705, Departament de Recerca i Universitats, Generalitat de Catalunya). Grant URV Martí i Franqués – Banco Santander (2021PMF-BS-12; J. Ezenarro). J. Riu would like to acknowledge the financial support from MICIU/AEI/10.13039/501100011033/and from FEDER 'Una manera de hacer Europa' (project PID2022-136649OB-I00).

Data availability

No data was used for the research described in the article.

References

[1] A. Weber, B. Hoplight, R. Ogilvie, C. Muro, S.R. Khandasammy, L. Pérez-Almodóvar, S. Sears, I.K. Lednev, Innovative vibrational spectroscopy research for forensic application, Anal. Chem. 95 (2023) 167–205, https://doi.org/10.1021/acs.analchem.2c05094.

[2] F. Li, J. Zhang, Y. Wang, Vibrational spectroscopy combined with chemometrics in authentication of functional foods, Crit. Rev. Anal. Chem. 54 (2024) 333–354, https://doi.org/10.1080/10408347.2022.2073433.

[3] S.K. Pirutin, S. Jia, A.I. Yusipovich, M.A. Shank, E.Y. Parshina, A.B. Rubin, Vibrational spectroscopy as a tool for bioanalytical and biomonitoring studies, Int. J. Mol. Sci. 24 (2023), https://doi.org/10.3390/ijms24086947.

[4] R. Pandiselvam, N.U. Sruthi, A. Kumar, A. Kothakota, R. Thirumdas, S.V. Ramesh, D. Cozzolino, Recent applications of vibrational spectroscopic techniques in the grain industry, Food Rev. Int. 39 (2023) 209–239, https://doi.org/10.1080/87559129.2021.1904253.

[5] M.M. Jansson, M. Kögler, S. Hörkkö, T. Ala-Kokko, L. Rieppo, Vibrational spectroscopy and its future applications in microbiology, Appl. Spectrosc. Rev. 58 (2023) 132–158, https://doi.org/10.1080/05704928.2021.1942894.

[6] B. Giussani, G. Gorla, J. Riu, Analytical chemistry strategies in the use of miniaturised NIR Instruments: an overview, Crit. Rev. Anal. Chem. 54 (2024) 11–43, https://doi.org/10.1080/10408347.2022.2047607.

[7] LetíciaP. Foli, M.C. Hespanhol, K.A.M.L. Cruz, C. Pasquini, Miniaturized Near-Infrared spectrophotometers in forensic analytical science – a critical review, Spectrochim. Acta Mol. Biomol. Spectrosc. 315 (2024) 124297, https://doi.org/10.1016/j.saa.2024.124297.

[8] A. Amirvaresi, H. Parastar, Miniaturized NIR spectroscopy and chemometrics: a smart combination to solve food authentication challenges, Front. Anal. Sci. 3 (2023) 1–8, https://doi.org/10.3389/frans.2023.1118590.

[9] A. Lorber, B.R. Kowalski, Estimation of prediction error for multivariate calibration, J. Chemom. 2 (1988) 93–109, https://doi.org/10.1002/CEM.1180020203.

[10] L. Zhang, S. García-Muñoz, A comparison of different methods to estimate prediction uncertainty using Partial Least Squares (PLS): a practitioner's perspective, Chemometr. Intell. Lab. Syst. 97 (2009) 152–158, https://doi.org/10.1016/J.CHEMOLAB.2009.03.007.

[11] P.D. Wentzell, Measurement errors in multivariate chemical data, J. Braz. Chem. Soc. 25 (2014) 183–196, https://doi.org/10.5935/0103-5053.20130293.

[12] M.S. Reis, P.M. Saraiva, Integration of data uncertainty in linear regression and process optimization, AIChE J. 51 (2005) 3007–3019, https://doi.org/10.1002/aic.10540.

[13] M.S. Reis, P.M. Saraiva, A comparative study of linear regression methods in noisy environments, J. Chemom. 18 (2004) 526–536, https://doi.org/10.1002/cem.897.

[14] T. Feital, D.M. Prata, J.C. Pinto, Comparison of methods for estimation of the covariance matrix of measurement errors, Can. J. Chem. Eng. 92 (2014) 2228–2245, https://doi.org/10.1002/cjce.22063.

[15] G. Gorla, A. Taiana, R. Boqué, P. Bani, O. Gachiuta, B. Giussani, Unravelling error sources in miniaturized NIR spectroscopic measurements: the case study of forages, Anal. Chim. Acta 1211 (2022) 339900, https://doi.org/10.1016/j.aca.2022.339900.

[16] A. Pulido, I. Ruisánchez, R. Boqué, F.X. Rius, Uncertainty of results in routine qualitative analysis, TrAC, Trends Anal. Chem. 22 (2003) 647–654, https://doi.org/10.1016/S0165-9936(03)01104-X.

[17] F. Allegrini, A.C. Olivieri, Recent advances in analytical figures of merit: heteroscedasticity strikes back, Anal. Methods 9 (2017) 739–743, https://doi.org/10.1039/c6ay02916g.

[18] F. Allegrini, A.C. Olivieri, IUPAC-consistent approach to the limit of detection in partial least-squares calibration, Anal. Chem. 86 (2014) 7858–7866, https://doi.org/10.1021/ac501786u.

[19] F. Allegrini, A.C. Olivieri, Multi-way figures of merit in the presence of heteroscedastic and correlated instrumental noise: unfolded partial least-squares with residual multi-linearization, Chemometr. Intell. Lab. Syst. 158 (2016) 200–209, https://doi.org/10.1016/j.chemolab.2016.09.001.

[20] N. Akvan, G. Azimi, A systematic study on the effect of different error structures on pseudo-univariate and multivariate figures of merit, J. Chemom. 37 (2023) 1–10, https://doi.org/10.1002/cem.3410.

[21] W. Fragoso, F. Allegrini, A.C. Olivieri, A new and consistent parameter for measuring the quality of multivariate analytical methods: generalized analytical sensitivity, Anal. Chim. Acta 933 (2016) 43–49, https://doi.org/10.1016/j.aca.2016.06.022.

[22] F. Allegrini, A.C. Olivieri, Sensitivity, prediction uncertainty, and detection limit for artificial neural network calibrations, Anal. Chem. 88 (2016) 7807–7812, https://doi.org/10.1021/acs.analchem.6b01857.

[23] M.S. Reis, R. Rendall, S.T. Chin, L. Chiang, Challenges in the specification and integration of measurement uncertainty in the development of data-driven models for the chemical processing industry, Ind. Eng. Chem. Res. 54 (2015) 9159–9177, https://doi.org/10.1021/ie504577d.

[24] L. Zhou, Y.C. Chuang, S.H. Hsu, Y. Yao, T. Chen, Prediction and uncertainty propagation for completion time of batch processes based on data-driven modeling, Ind. Eng. Chem. Res. 59 (2020) 14374–14384, https://doi.org/10.1021/acs.iecr.0c01236.

[25] G. Bano, P. Facco, N. Meneghetti, F. Bezzo, M. Barolo, Uncertainty back-propagation in PLS model inversion for design space determination in pharmaceutical product development, Comput. Chem. Eng. 101 (2017) 110–124, https://doi.org/10.1016/j.compchemeng.2017.02.038.

[26] O. Preisner, J.A. Lopes, J.C. Menezes, Uncertainty assessment in FT-IR spectroscopy based bacteria classification models, Chemometr. Intell. Lab. Syst. 94 (2008) 33–42, https://doi.org/10.1016/j.chemolab.2008.06.005.

[27] C. Du, S. Dai, Y. Qiao, Z. Wu, Error propagation of partial least squares for parameters optimization in NIR modeling, Spectrochim. Acta Mol. Biomol. Spectrosc. 192 (2018) 244–250, https://doi.org/10.1016/j.saa.2017.10.069.

[28] E. Heid, C.J. McGill, F.H. Vermeire, W.H. Green, Characterizing uncertainty in machine learning for chemistry, J. Chem. Inf. Model. 63 (2023) 4012–4029, https://doi.org/10.1021/acs.jcim.3c00373.

[29] A.C. Olivieri, A simple approach to uncertainty propagation in preprocessed multivariate calibration, J. Chemom. 16 (2002) 207–217, https://doi.org/10.1002/cem.716.

[30] F. Allegrini, P.D. Wentzell, A.C. Olivieri, Generalized error-dependent prediction uncertainty in multivariate calibration, Anal. Chim. Acta 903 (2016) 51–60, https://doi.org/10.1016/j.aca.2015.11.028.

[31] D. Bu, B. Wan, G. McGeorge, A discussion on the use of prediction uncertainty estimation of NIR data in partial least squares for quantitative pharmaceutical tablet assay methods, Chemometr. Intell. Lab. Syst. 120 (2013) 84–91, https://doi.org/10.1016/J.CHEMOLAB.2012.11.005.

[32] P. De Bièvre, Measurement results without statements of reliability (uncertainty) should not be taken seriously, Accred Qual. Assur. 2 (1997) 269, https://doi.org/10.1007/s007690050147, 269.

[33] M.S. Reis, P.M. Saraiva, Heteroscedastic latent variable modelling with applications to multivariate statistical process control, Chemometr. Intell. Lab. Syst. 80 (2006) 57–66, https://doi.org/10.1016/j.chemolab.2005.07.002.

[34] P.D. Wentzell, S. Hou, Exploratory data analysis with noisy measurements, J. Chemom. 26 (2012) 264–281, https://doi.org/10.1002/cem.2428.

[35] F. Allegrini, J.W.B. Braga, A.C.O. Moreira, A.C. Olivieri, Error Covariance Penalized Regression: a novel multivariate model combining penalized regression with multivariate error structure, Anal. Chim. Acta 1011 (2018) 20–27, https://doi.org/10.1016/j.aca.2018.02.002.

[36] M. Schoot, M. Alewijn, Y. Weesepoel, J. Mueller-Maatsch, C. Kapper, G. Postma, L. Buydens, J. Jansen, Predicting the performance of handheld near-infrared photonic sensors from a master benchtop device, Anal. Chim. Acta 1203 (2022) 339707, https://doi.org/10.1016/j.aca.2022.339707.

[37] D.T. Andrews, P.D. Wentzell, Applications of maximum likelihood principal component analysis: incomplete data sets and calibration transfer, Anal. Chim. Acta 350 (1997) 341–352, https://doi.org/10.1016/S0003-2670(97)00270-5.

[38] A. da R.D. Monteiro, T. de S. Feital, J.C. Pinto, Statistical aspects of near-infrared spectroscopy for the characterization of errors and model building, Appl. Spectrosc. 71 (2017) 1665–1676, https://doi.org/10.1177/0003702817704587.

[39] J. Mohammad Jafari, R. Tauler, H. Abdollahi, Balanced scaling as a pretreatment step in Multivariate Curve Resolution analysis of noisy data, Microchem. J. 160 (2021) 105738, https://doi.org/10.1016/j.microc.2020.105738.

[40] L. Blanchet, J. Réhault, C. Ruckebusch, J.P. Huvenne, R. Tauler, A. de Juan, Chemometrics description of measurement error structure: study of an ultrafast absorption spectroscopy experiment, Anal. Chim. Acta 642 (2009) 19–26, https://doi.org/10.1016/j.aca.2008.11.039.

[41] S.K. Schreyer, M. Bidinosti, P.D. Wentzell, Application of maximum likelihood principal components regression to fluorescence emission spectra, Appl. Spectrosc. 56 (2002) 789–796, https://doi.org/10.1366/000370202760076857.

[42] C.M. Andersen, R. Bro, Quantification and handling of sampling errors in instrumental measurements: a case study, Chemometr. Intell. Lab. Syst. 72 (2004) 43–50, https://doi.org/10.1016/j.chemolab.2003.12.014.

[43] M.N. Leger, L. Vega-Montoto, P.D. Wentzell, Methods for systematic investigation of measurement error covariance matrices, Chemometr. Intell. Lab. Syst. 77 (2005) 181–205, https://doi.org/10.1016/j.chemolab.2004.09.017.

[44] P.D. Wentzell, C.S. Cleary, M. Kompany-Zareh, Improved modeling of multivariate measurement errors based on the Wishart distribution, Anal. Chim. Acta 959 (2017) 1–14, https://doi.org/10.1016/j.aca.2016.12.009.

[45] A.C. Olivieri, N. Faber, J. Ferré, R. Boqué, J.H. Kalivas, H. Mark, Uncertainty estimation and figures of merit for multivariate calibration: (IUPAC technical report), Pure Appl. Chem. 78 (2006) 633–661, https://doi.org/10.1351/pac200678030633.

[46] K.S. Booksh, B.R. Kowalski, Theory of analytical chemistry, Anal. Chem. 66 (1994) 782–791, https://doi.org/10.1021/ac00087a001.

[47] P.D. Wentzell, The errors of my ways: maximum likelihood PCA seventeen years after bruce, ACS (Am. Chem. Soc.) Symp. Ser. 1199 (2015) 31–64, https://doi.org/10.1021/bk-2015-1199.ch003.

[48] J. Galbán, S. De Marcos, I. Sanz, C. Ubide, J. Zuriarrain, Uncertainty in modern spectrophotometers, Anal. Chem. 79 (2007) 4763–4767, https://doi.org/10.1021/ac071933h.

[49] J.D. Ingle, S.R. Crouch, Evaluation of precision of quantitative molecular absorption spectrometric measurements, Anal. Chem. 44 (1972) 1375–1386, https://doi.org/10.1021/ac60316a010.

[50] L.D. Rothman, S.R. Crouch, J.D. Ingle, Theoretical and experimental investigation of factors affecting precision in molecular absorption spectrophotometry, Anal. Chem. 47 (1975) 1226–1233, https://doi.org/10.1021/ac60358a029.

[51] J.D. Ingle, Spectrochemical Analysis, Prentice Hall, Englewood Cliffs, N.J., 1988.

[52] K.B. Beć, J. Grabska, C.W. Huck, Principles and applications of miniaturized near-infrared (NIR) spectrometers, Chem. Eur J. 27 (2021) 1514–1532, https://doi.org/10.1002/chem.202002838.

[53] C.G. Kirchler, C.K. Pezzei, K.B. Beć, S. Mayr, M. Ishigaki, Y. Ozaki, C.W. Huck, 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 Rosmarin, Analyst 142 (2017) 455–464, https://doi.org/10.1039/c6an02439d.

[54] S. Mayr, K.B. Beć, J. Grabska, E. Schneckenreiter, C.W. Huck, Near-infrared spectroscopy in quality control of Piper nigrum: a comparison of performance of benchtop and handheld spectrometers, Talanta 223 (2021) 121809, https://doi.org/10.1016/j.talanta.2020.121809.

[55] E. Sanchez, B.R. Kowalski, Tensorial resolution: a direct trilinear decomposition, J. Chemom. 4 (1990) 29–45, https://doi.org/10.1002/cem.1180040105.

[56] S.P. Driscoll, Exploration of Multivariate Chemical Data in Noisy Environments: New Algorithms and Simulation Methods, 2019.

[57] A.C. Olivieri, S. Bortolato, F. Allegrini, Figures of merit in multiway calibration, in: Data Handling in Science and Technology, Elsevier, 2015, pp. 541–575, https://doi.org/10.1016/B978-0-444-63527-3.00013-8.

[58] H. Martens, M. Høy, B.M. Wise, R. Bro, P.B. Brockhoff, Pre-whitening of data by covariance-weighted pre-processing, J. Chemom. 17 (2003) 153–165, https://doi.org/10.1002/cem.780.

[59] J.M. Bland, D.G. Altman, Measuring agreement in method comparison studies, Stat. Methods Med. Res. 8 (1999) 135–160, https://doi.org/10.1177/096228029900800204.

[60] B.G. Francq, B.B. Govaerts, Measurement methods comparison with errors-in-variables regressions. From horizontal to vertical OLS regression, review and new perspectives, Chemometr. Intell. Lab. Syst. 134 (2014) 123–139, https://doi.org/10.1016/j.chemolab.2014.03.006.

[61] C. Hartmann, J. Smeyers-Verbeke, W. Penninckx, D.L. Massart, Detection of bias in method comparison by regression analysis, Anal. Chim. Acta 338 (1997) 19–40, https://doi.org/10.1016/S0003-2670(96)00341-8.

[62] K.H. Esbensen, C. Wagner, Theory of sampling (TOS) versus measurement uncertainty (MU) - a call for integration, TrAC, Trends Anal. Chem. 57 (2014) 93–106, https://doi.org/10.1016/j.trac.2014.02.007.

[63] M. Belter, A. Sajnóg, D. Barałkiewicz, Over a century of detection and quantification capabilities in analytical chemistry - historical overview and

trends, Talanta 129 (2014) 606–616, https://doi.org/10.1016/j.talanta.2014.05.018.
[64] P.D. Wentzell, 2.25 - other topics in soft-modeling: maximum likelihood-based soft-modeling methods, in: S.D. Brown, R. Tauler, B. Walczak (Eds.), Comprehensive Chemometrics, Elsevier, Oxford, 2009, pp. 507–558, https://doi.org/10.1016/B978-044452701-1.00057-0.
[65] P.D. Wentzell, A.C. Tarasuk, Characterization of heteroscedastic measurement noise in the absence of replicates, Anal. Chim. Acta 847 (2014) 16–28, https://doi.org/10.1016/j.aca.2014.08.007.
[66] S. Driscoll, P.D. Wentzell, NoiseGen - analytical measurement error simulation software, Chemometr. Intell. Lab. Syst. 189 (2019) 155–160, https://doi.org/10.1016/j.chemolab.2019.04.011.
[67] Marc Norman Léger, Measurement Errors and Signal Preprocessing in Spectroscopy, Dalhousie university, 2004.
[68] G. Gorla, P. Taborelli, B. Giussani, A multivariate analysis-driven workflow to tackle uncertainties in miniaturized NIR data, Molecules 28 (2023), https://doi.org/10.3390/molecules28247999.
[69] F. Matinrad, M. Kompany-Zareh, N. Omidikia, M. Dadashi, Systematic investigation of the measurement error structure in a smartphone-based spectrophotometer, Anal. Chim. Acta 1129 (2020) 98–107, https://doi.org/10.1016/j.aca.2020.06.066.
[70] R. Todeschini, V. Consonni, A. Maiocchi, The K correlation index: theory development and its application in chemometrics, Chemometr. Intell. Lab. Syst. 46 (1999) 13–29, https://doi.org/10.1016/S0169-7439(98)00124-5.
[71] C.A. Meza Ramirez, M. Greenop, L. Ashton, I. ur Rehman, Applications of machine learning in spectroscopy, Appl. Spectrosc. Rev. 56 (2021) 733–763, https://doi.org/10.1080/05704928.2020.1859525.
[72] J.C. Castura, D.N. Rutledge, C.F. Ross, T. Næs, Discriminability and uncertainty in principal component analysis (PCA) of temporal check-all-that-apply (TCATA) data, Food Qual. Prefer. 96 (2022) 104370, https://doi.org/10.1016/J.FOODQUAL.2021.104370.
[73] G. James, D. Witten, T. Hastie, R. Tibshirani, J. Taylor, Resampling methods, in: An Introduction to Statistical Learning, Springer, Cham, Cham, Switzerland, 2023, https://doi.org/10.1007/978-3-031-38747-0_5.
[74] R. Wehrens, H. Putter, L.M.C. Buydens, The bootstrap: a tutorial, Chemometr. Intell. Lab. Syst. 54 (2000) 35–52, https://doi.org/10.1016/S0169-7439(00)00102-7.
[75] C.F.J. Wu, Jackknife, bootstrap and other resampling methods in regression analysis, Ann. Stat. 14 (1986) 1261–1295, https://doi.org/10.1214/aos/1176350142.
[76] J. Shao, D. Tu, The Jackknife and Bootstrap, Springer, New York, New York, NY, 1995, https://doi.org/10.1007/978-1-4612-0795-5.
[77] J. Josse, S. Wager, F. Husson, Confidence areas for fixed-effects PCA, J. Comput. Graph Stat. 25 (2016) 28–48, https://doi.org/10.1080/10618600.2014.950871.
[78] B. Efron, R.J. Tibshirani, An Introduction to the Bootstrap, Chapman & Hall, New York, 1993.
[79] C. Beleites, R. Salzer, Assessing and improving the stability of chemometric models in small sample size situations, Anal. Bioanal. Chem. 390 (2008) 1261–1271, https://doi.org/10.1007/s00216-007-1818-6.
[80] D.K. Barrow, S.F. Crone, Crogging (cross-validation aggregation) for forecasting - a novel algorithm of neural network ensembles on time series subsamples, in: The 2013 International Joint Conference on Neural Networks (IJCNN), IEEE, 2013, pp. 1–8, https://doi.org/10.1109/IJCNN.2013.6706740.
[81] Y.-W. Lin, B.-C. Deng, L.-L. Wang, Q.-S. Xu, L. Liu, Y.-Z. Liang, Fisher optimal subspace shrinkage for block variable selection with applications to NIR spectroscopic analysis, Chemometr. Intell. Lab. Syst. 159 (2016) 196–204, https://doi.org/10.1016/j.chemolab.2016.11.002.
[82] C.A. Field, A.H. Welsh, Bootstrapping clustered data, J R Stat Soc Series B Stat Methodol 69 (2007) 369–390, https://doi.org/10.1111/j.1467-9868.2007.00593.x.
[83] C. Beleites, R. Baumgartner, C. Bowman, R. Somorjai, G. Steiner, R. Salzer, M. G. Sowa, Variance reduction in estimating classification error using sparse datasets, Chemometr. Intell. Lab. Syst. 79 (2005) 91–100, https://doi.org/10.1016/j.chemolab.2005.04.008.
[84] L.C. Groff, J.N. Grossman, A. Kruve, J.M. Minucci, C.N. Lowe, J.P. McCord, D. F. Kapraun, K.A. Phillips, S.T. Purucker, A. Chao, C.L. Ring, A.J. Williams, J. R. Sobus, Uncertainty estimation strategies for quantitative non-targeted analysis, Anal. Bioanal. Chem. 414 (2022) 4919–4933, https://doi.org/10.1007/s00216-022-04118-z.
[85] J.G. MacKinnon, M.Ø. Nielsen, M.D. Webb, Fast and reliable jackknife and bootstrap methods for cluster-robust inference, J. Appl. Econom. 38 (2023) 671–694, https://doi.org/10.1002/jae.2969.
[86] S. Fang, G. Hemani, T.G. Richardson, T.R. Gaunt, G. Davey Smith, Evaluating and implementing block jackknife resampling Mendelian randomization to mitigate bias induced by overlapping samples, Hum. Mol. Genet. 32 (2023) 192–203, https://doi.org/10.1093/hmg/ddac186.
[87] J. Jaumot, R. Tauler, MCR-BANDS: a user friendly MATLAB program for the evaluation of rotation ambiguities in Multivariate Curve Resolution, Chemometr. Intell. Lab. Syst. 103 (2010) 96–107, https://doi.org/10.1016/j.chemolab.2010.05.020.
[88] A. Golshan, H. Abdollahi, S. Beyramysoltan, M. Maeder, K. Neymeyr, R. Rajkó, M. Sawall, R. Tauler, A review of recent methods for the determination of ranges of feasible solutions resulting from soft modelling analyses of multivariate data, Anal. Chim. Acta 911 (2016) 1–13, https://doi.org/10.1016/j.aca.2016.01.011.
[89] M. Dadashi, H. Abdollahi, R. Tauler, Error propagation along the different regions of multivariate curve resolution feasible solutions, Chemometr. Intell. Lab. Syst. 162 (2017) 203–213, https://doi.org/10.1016/j.chemolab.2017.01.009.
[90] K.H. Esbensen, P. Geladi, Principles of proper validation: use and abuse of re-sampling for validation, J. Chemom. 24 (2010) 168–187, https://doi.org/10.1002/cem.1310.
[91] E. Lopez, J. Etxebarria-Elezgarai, J.M. Amigo, A. Seifert, The importance of choosing a proper validation strategy in predictive models. A tutorial with real examples, Anal. Chim. Acta 1275 (2023) 341532, https://doi.org/10.1016/j.aca.2023.341532.
[92] F. Westad, F. Marini, Validation of chemometric models - a tutorial, Anal. Chim. Acta 893 (2015) 14–24, https://doi.org/10.1016/j.aca.2015.06.056.
[93] A.C. Olivieri, Analytical figures of merit: from univariate to multiway calibration, Chem. Rev. 114 (2014) 5358–5378, https://doi.org/10.1021/cr400455s.
[94] G. Bauer, W. Wegscheider, H.M. Ortner, Selectivity and limits of detection in inductively coupled plasma optical emission spectrometry using multivariate calibration, Spectrochim. Acta Part B At. Spectrosc. 47 (1992) 179–188, https://doi.org/10.1016/0584-8547(92)80017-B.
[95] S. Camo, The Unscrambler User's Guide, Trondheim, 1994, version 5.5.
[96] S. De Vries, C.J.F. Ter Braak, Prediction error in partial least squares regression: a critique on the deviation used in the Unscrambler, Chemometr. Intell. Lab. Syst. 30 (1995) 239–245, https://doi.org/10.1016/0169-7439(95)00030-5.
[97] K. Faber, B.R. Kowalski, Prediction error in least squares regression: further critique on the deviation used in the Unscrambler, Chemometr. Intell. Lab. Syst. 34 (1996) 283–292, https://doi.org/10.1016/0169-7439(96)00022-6.
[98] M. Høy, K. Steen, H. Martens, Review of partial least squares regression prediction error in Unscrambler, Chemometr. Intell. Lab. Syst. 44 (1998) 123–133, https://doi.org/10.1016/S0169-7439(98)00163-4.
[99] T.V. Karstang, J. Toft, O.M. Kvalheim, Estimation of prediction error for samples within the calibration range, J. Chemom. 6 (1992) 177–188, https://doi.org/10.1002/cem.1180060403.
[100] K. Faber, B.R. Kowalski, Propagation of measurement errors for the validation of predictions obtained by principal component regression and partial least squares, J. Chemom. 11 (1997) 181–238, https://doi.org/10.1002/(SICI)1099-128X(199705)11:3<181::AID-CEM459>3.0.CO;2-7.
[101] N.M. Faber, D.L. Duewer, S.J. Choquette, T.L. Green, S.N. Chesler, Characterizing the uncertainty in near-infrared spectroscopic prediction of mixed-oxygenate concentrations in gasoline: sample-specific prediction intervals, Anal. Chem. 70 (1998) 4877, https://doi.org/10.1021/AC9815608.
[102] R. Boqué, M.S. Larrechi, F.X. Rius, Multivariate detection limits with fixed probabilities of error, Chemometr. Intell. Lab. Syst. 45 (1999) 397–408, https://doi.org/10.1016/S0169-7439(98)00195-6.
[103] N. Faber, X.H. Song, P.K. Hopke, Sample-specific standard error of prediction for partial least squares regression, TrAC, Trends Anal. Chem. 22 (2003) 330–334, https://doi.org/10.1016/S0165-9936(03)00503-X.
[104] N. Faber, R. Bro, Standard error of prediction for multiway PLS 1. Background and a simulation study, Chemometr. Intell. Lab. Syst. 61 (2002) 133–149, https://doi.org/10.1016/S0169-7439(01)00204-0.
[105] P.B. Skou, T.A. Berg, S.D. Aunsbjerg, D. Thaysen, M.A. Rasmussen, F. van den Berg, Monitoring process water quality using near infrared spectroscopy and partial least squares regression with prediction uncertainty estimation, Appl. Spectrosc. 71 (2017) 410–421, https://doi.org/10.1177/0003702816654165.
[106] M. Otto, W. Wegscheider, Selectivity in multicomponent analysis, Anal. Chim. Acta 180 (1986) 445–456, https://doi.org/10.1016/0003-2670(86)80024-1.
[107] A.C. Olivieri, Practical guidelines for reporting results in single- and multi-component analytical calibration: a tutorial, Anal. Chim. Acta 868 (2015) 10–22, https://doi.org/10.1016/j.aca.2015.01.017.
[108] F. Allegrini, A.C. Olivieri, 2.20 - figures of merit, in: R. T, B.W. Steven Brown (Eds.), Comprehensive Chemometrics, second ed., Elsevier Inc., 2020, pp. 441–463, https://doi.org/10.1016/b978-0-12-409547-2.14612-8.
[109] R. Boqué, F.X. Rius, Computing detection limits in multicomponent spectroscopic analysis, TrAC, Trends Anal. Chem. 16 (1997) 432–436, https://doi.org/10.1016/S0165-9936(97)00048-4.
[110] R. Boqué, F.X. Rius, Multivariate detection limits estimators, Chemometr. Intell. Lab. Syst. 32 (1996) 11–23, https://doi.org/10.1016/0169-7439(95)00049-6.
[111] A. Lorber, Error propagation and figures of merit for quantification by solving matrix equations, Anal. Chem. 58 (1986) 1167–1172, https://doi.org/10.1021/AC00297A042/ASSET/AC00297A042.FP.PNG_V03.
[112] R. Boqué, N.M. Faber, F.X. Rius, Detection limits in classical multivariate calibration models, Anal. Chim. Acta 423 (2000) 41–49, https://doi.org/10.1016/S0003-2670(00)01101-6.
[113] Z. Wu, C. Sui, B. Xu, L. Ai, Q. Ma, X. Shi, Y. Qiao, Multivariate detection limits of on-line NIR model for extraction process of chlorogenic acid from Lonicera japonica, J. Pharm. Biomed. Anal. 77 (2013) 16–20, https://doi.org/10.1016/J.JPBA.2012.12.026.
[114] M. Alcalà, J. León, J. Ropero, M. Blanco, R.J. Romañach, Analysis of low content drug tablets by transmission near infrared spectroscopy: selection of calibration ranges according to multivariate detection and quantitation limits of PLS models, J. Pharmaceut. Sci. 97 (2008) 5318–5327, https://doi.org/10.1002/JPS.21373.
[115] R.A. Stine, Bootstrap prediction intervals for regression, J. Am. Stat. Assoc. 80 (1985) 1026–1031, https://doi.org/10.1080/01621459.1985.10478220.
[116] M.C. Denham, Prediction intervals in partial least squares, J. Chemom. 11 (1997) 39–52, https://doi.org/10.1002/(SICI)1099-128X(199701)11:1<39::AID-CEM433>3.0.CO;2-S.

[117] R. Butler, E.D. Rothman, Predictive intervals based on reuse of the sample, J. Am. Stat. Assoc. 75 (1980) 881–889, https://doi.org/10.1080/01621459.1980.10477567.
[118] R.J. Carroll, D. Ruppert, Prediction and tolerance intervals with transformation and/or weighting, Technometrics 33 (1991) 197–210, https://doi.org/10.1080/00401706.1991.10484807.
[119] A. Phatak, P.M. Reilly, A. Penlidis, An approach to interval estimation in partial least squares regression, Anal. Chim. Acta 277 (1993) 495–501, https://doi.org/10.1016/0003-2670(93)80461-S.
[120] S. Serneels, P. Lemberge, P.J. Van Espen, Calculation of PLS prediction intervals using efficient recursive relations for the Jacobian matrix, J. Chemom. 18 (2004) 76–80, https://doi.org/10.1002/cem.849.
[121] N.M. Faber, Uncertainty estimation for multivariate regression coefficients, Chemometr. Intell. Lab. Syst. 64 (2002) 169–179, https://doi.org/10.1016/S0169-7439(02)00102-8.
[122] F. Marini, Classification methods in chemometrics, Curr. Anal. Chem. 6 (2010) 72–79, https://doi.org/10.2174/157341110790069592.
[123] L. Cuadros-Rodríguez, E. Pérez-Castaño, C. Ruiz-Samblás, Quality performance metrics in multivariate classification methods for qualitative analysis, TrAC, Trends Anal. Chem. 80 (2016) 612–624, https://doi.org/10.1016/j.trac.2016.04.021.
[124] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, second ed., Wiley, New York, 2001.
[125] Eigenvector Research, Sample classification predictions, n.d. https://wiki.eigenvector.com/index.php?title=Sample_Classification_Predictions, 2024. April 1
[126] N.F. Pérez, J. Ferré, R. Boqué, Calculation of the reliability of classification in discriminant partial least-squares binary classification, Chemometr. Intell. Lab. Syst. 95 (2009) 122–128, https://doi.org/10.1016/j.chemolab.2008.09.005.
[127] C. Botella, J. Ferré, R. Boqué, Classification from microarray data using probabilistic discriminant partial least squares with reject option, Talanta 80 (2009) 321–328, https://doi.org/10.1016/j.talanta.2009.06.072.
[128] N.F. Pérez, J. Ferré, R. Boqué, Multi-class classification with probabilistic discriminant partial least squares (p-DPLS), Anal. Chim. Acta 664 (2010) 27–33, https://doi.org/10.1016/j.aca.2010.01.059.
[129] T. Fearn, D. Pérez-Marín, A. Garrido-Varo, J.E. Guerrero-Ginel, Classifying with confidence using Bayes rule and kernel density estimation, Chemometr. Intell. Lab. Syst. 189 (2019) 81–87, https://doi.org/10.1016/j.chemolab.2019.04.004.
[130] D. Pérez-Marín, T. Fearn, C. Riccioli, E. De Pedro, A. Garrido, Probabilistic classification models for the in situ authentication of iberian pig carcasses using near infrared spectroscopy, Talanta 222 (2021) 121511, https://doi.org/10.1016/j.talanta.2020.121511.
[131] D.W. Hosmer, S. Lemeshow, R.X. Sturdivant, Applied Logistic Regression, third ed., Wiley, Hoboken, 2013 https://doi.org/10.1002/9781118548387.
[132] D. Toher, G. Downey, T.B. Murphy, A comparison of model-based and regression classification techniques applied to near infrared spectroscopic data in food authentication studies, Chemometr. Intell. Lab. Syst. 89 (2007) 102–115, https://doi.org/10.1016/j.chemolab.2007.06.005.
[133] H. van der Voet, D.A. Doornbos, The improvement of SIMCA classification by using kernel density estimation, Anal. Chim. Acta 161 (1984) 115–123, https://doi.org/10.1016/S0003-2670(00)85783-9.
[134] R. Vitale, M. Cocchi, A. Biancolillo, C. Ruckebusch, F. Marini, Class modelling by soft independent modelling of class analogy: why, when, how? A tutorial, Anal. Chim. Acta 1270 (2023) 341304, https://doi.org/10.1016/j.aca.2023.341304.
[135] C. Beleites, U. Neugebauer, T. Bocklitz, C. Krafft, J. Popp, Sample size planning for classification models, Anal. Chim. Acta 760 (2013) 25–33, https://doi.org/10.1016/j.aca.2012.11.007.
[136] M.R. de Almeida, D.N. Correa, W.F.C. Rocha, F.J.O. Scafi, R.J. Poppi, Discrimination between authentic and counterfeit banknotes using Raman spectroscopy and PLS-DA with uncertainty estimation, Microchem. J. 109 (2013) 170–177, https://doi.org/10.1016/j.microc.2012.03.006.
[137] W.F.C. Rocha, D.A. Sheen, Classification of biodegradable materials using QSAR modelling with uncertainty estimation, SAR QSAR Environ. Res. 27 (2016) 799–811, https://doi.org/10.1080/1062936X.2016.1238010.
[138] W.F.C. Rocha, D.A. Sheen, D.W. Bearden, Classification of samples from NMR-based metabolomics using principal components analysis and partial least squares with uncertainty estimation, Anal. Bioanal. Chem. 410 (2018) 6305–6319, https://doi.org/10.1007/s00216-018-1240-2.
[139] C.L.M. Morais, K.M.G. Lima, F.L. Martin, Uncertainty estimation and misclassification probability for classification models based on discriminant analysis and support vector machines, Anal. Chim. Acta 1063 (2019) 40–46, https://doi.org/10.1016/j.aca.2018.09.022.
[140] J.L. Villa, R. Boqué, J. Ferré, Calculation of the probability of correct classification in probabilistic bagged k-Nearest Neighbours, Chemometr. Intell. Lab. Syst. 94 (2008) 51–59, https://doi.org/10.1016/J.CHEMOLAB.2008.06.007.
[141] P. Paatero, U. Tapper, Analysis of different modes of factor analysis as least squares fit problems, Chemometr. Intell. Lab. Syst. 18 (1993) 183–194, https://doi.org/10.1016/0169-7439(93)80055-M.
[142] B.R. Kowalski, Improving the reliability of factor analysis of chemical data by utilizing the measured analytical uncertainty, Anal. Chem. 48 (1976) 1986–1988.
[143] M. Schuermans, I. Markovsky, P.D. Wentzell, S. Van Huffel, On the equivalence between total least squares and maximum likelihood PCA, Anal. Chim. Acta 544 (2005) 254–267, https://doi.org/10.1016/j.aca.2004.12.059.
[144] P.D. Wentzell, C.C. Wicks, J.W.B. Braga, L.F. Soares, T.C.M. Pastore, V.T. R. Coradin, F. Davrieux, Implications of measurement error structure on the visualization of multivariate chemical data: hazards and alternatives, Can. J. Chem. 96 (2018) 738–748, https://doi.org/10.1139/cjc-2017-0730.
[145] M. Kompany-Zareh, P. Wentzell, B. Dalvand, M.T. Baharifard, Factor analysis for signal modeling and noise characterization in spectro-kinetic data, Chemometr. Intell. Lab. Syst. 240 (2023) 104916, https://doi.org/10.1016/j.chemolab.2023.104916.
[146] P.D. Wentzell, C. Giglio, M. Kompany-Zareh, Beyond principal components: a critical comparison of factor analysis methods for subspace modelling in chemistry, Anal. Methods 13 (2021) 4188–4219, https://doi.org/10.1039/d1ay01124c.
[147] P.D. Wentzell, D.T. Andrews, B.R. Kowalski, Maximum likelihood multivariate calibration, Anal. Chem. 69 (1997) 2299–2311, https://doi.org/10.1021/ac961029h.
[148] T. Kourti, J.F. MacGregor, Process analysis, monitoring and diagnosis, using multivariate projection methods, Chemometr. Intell. Lab. Syst. 28 (1995) 3–21, https://doi.org/10.1016/0169-7439(95)80036-9.
[149] L. Vega-Montoto, P.D. Wentzell, Maximum likelihood parallel factor analysis (MLPARAFAC), J. Chemom. 17 (2003) 237–253, https://doi.org/10.1002/cem.789.
[150] R. Bro, N.D. Sidiropoulos, A.K. Smilde, Maximum likelihood fitting using ordinary least squares algorithms, J. Chemom. 16 (2002) 387–400, https://doi.org/10.1002/cem.734.
[151] H. Chen, B.R. Bakshi, P.K. Goel, Toward Bayesian chemometrics-A tutorial on some recent advances, Anal. Chim. Acta 602 (2007) 1–16, https://doi.org/10.1016/j.aca.2007.08.044.
[152] N. Armstrong, D.B. Hibbert, An introduction to Bayesian methods for analyzing chemistry data. Part 1: an introduction to Bayesian theory and methods, Chemometr. Intell. Lab. Syst. 97 (2009) 194–210, https://doi.org/10.1016/j.chemolab.2009.04.001.
[153] D.B. Hibbert, N. Armstrong, An introduction to Bayesian methods for analyzing chemistry data. Part II: a review of applications of Bayesian methods in chemistry, Chemometr. Intell. Lab. Syst. 97 (2009) 211–220, https://doi.org/10.1016/j.chemolab.2009.03.009.