

# Chemometrics for Raman Spectroscopy Harmonization

Applied Spectroscopy  
2022, Vol. 0(0) 1–21  
© The Author(s) 2022  
Article reuse guidelines:  
[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)  
DOI: 10.1177/00037028221094070  
[journals.sagepub.com/home/asp](https://journals.sagepub.com/home/asp)

Bastian Barton<sup>1</sup>, James Thomson<sup>2</sup>, Enrique Lozano Diz<sup>2</sup> and Raquel Portela<sup>3</sup>

## Abstract

Raman spectroscopy is used in a wide variety of fields, and in a plethora of different configurations. Raman spectra of simple analytes can often be analyzed using univariate approaches and interpreted in a straightforward manner. For more complex spectral data such as time series or line profiles (1D), Raman maps (2D), or even volumes (3D), multivariate data analysis (MVDA) becomes a requirement. Even though there are some existing standards for creation, implementation, and validation of methods and models employed in industry and academics, further research and development in the field must contribute to their improvement. This review will cover, in broad terms, existing techniques as well as new developments for MVDA for Raman spectroscopic data, and in particular the use associated with instrumentation and data calibration. Chemometric models are often generated via fusion of analytical data from different sources, which enhances model discrimination and prediction abilities as compared to models derived from a single data source. For Raman spectroscopy, raw or unprocessed data is rarely ever used. Instead, spectra are usually corrected and manipulated,<sup>1</sup> often by case-specific rather than universal methods. Calibration models can be used to characterize qualitatively and/or quantitatively samples measured with the same instrumentation that was used to create the model. However, regular validation is required to ensure that aging or incorrect maintenance of the instrument does not alter the model's predictions, particularly when applied in regulated fields such as pharmaceuticals. Furthermore, a model transfer may be required for different reasons, such as replacement or significant repair of the instrumentation. Modeling can also be used to consistently harmonize Raman spectroscopic data across several instrumental designs, accounting for variations in the resulting spectrum induced by different components. Data for Raman harmonization models should be processed in a protocolled manner, and the original data accessible to allow for model reconstruction or transfer when new data is added. Important processing steps will be the calibration of the spectral axes and instrument dependent effects, such as spectral resolution. In addition, data fusion and model transfer are essential for allowing new instrumentation to build on existing models to harmonize their own data. Ideally, an open access database would be created and maintained, for the purpose of allowing for continued harmonization of new Raman instruments using an outlined and accepted protocol.

## Keywords

Modelling, calibration, data processing, data fusion, model transfer, standard, signal processing, vibrational spectroscopy, statistical modeling, machine learning

Date received: 26 November 2021; accepted: 22 February 2022

## Introduction

In recent years, the application of Raman spectroscopy has been extended to a variety of fields. Amongst its unique features are non-destructiveness, capability to analyze underneath the sample surface, low interference by water, high spatial and spectral resolution, and material fingerprinting. These render Raman spectroscopy a powerful technique, for example, for pharmaceutical development,<sup>2</sup> as well as for life<sup>3</sup> and materials sciences. However, Raman users are still sometimes concerned with accuracy, reproducibility, and exchangeability of Raman data, particularly for chemometric

modeling. This highlights the importance of harmonization for Raman instrumentation, data acquisition, and processing.<sup>4</sup>

<sup>1</sup>Fraunhofer LBF, Darmstadt, Germany

<sup>2</sup>ELODIZ Ltd, High Wycombe, UK

<sup>3</sup>Institute of Catalysis and Petrochemistry, CSIC-ICP, Madrid, Spain

### Corresponding author:

Raquel Portela, Institute of Catalysis and Petrochemistry, CSIC-ICP, 2, Marie Curie, Madrid 28049, Spain.  
Email: [raquel.portela@csic.es](mailto:raquel.portela@csic.es)

**Table I.** Parameters influencing spectral S/N, together with typical values and limitations.

Parameter	Typical Values	Limitation By
N: number of acquisitions accumulated	2–10	Accumulation of CCD readout noise
$I_n$ : intensity of a single acquisition	Some 1000 photons per CCD pixel	CCD dynamic range, CCD readout noise
p: Average 2D laser power density on sample surface	<1 mW/ $\mu\text{m}^2$ (532 nm) <5 mW/ $\mu\text{m}^2$ (785 and 1064 nm) *20–150 mW/ $\mu\text{m}^2$	Laser power and wavelength, beam sensitivity and thermal conductivity of sample, transmissivity of sample
A: total illuminated area on sample	100–200 $\mu\text{m}^2$ *<1 $\mu\text{m}^2$	Focusing optics, sample homogeneity and surface topography
t: time of a single acquisition	0.5–10 s	Sample stability, instrumental drift

\*Values for micro-Raman

The quality of a model depends on the modeling procedure as well as on the quantity and quality of the data used for modeling. In this review, we analyze different approaches to robust and reliable chemometric models, describing the proposed steps, procedures, and protocols, with a particular focus on harmonization. The objective is to lead the reader through the various steps required to construct a chemometric model for calibration and harmonization of Raman data. Within each step, several techniques will be introduced. It is not the objective of this review to give detailed instructions on how to perform any or all of these techniques but merely to highlight their major differences and the need for them. Ultimately, the review of these steps and techniques will be related to their applicability to harmonization of Raman spectroscopy.

### Spectrum Collection

To ensure that different instruments can be harmonized, an initial step is to calibrate each instrument to an appropriate standard (see the Existing Standards, Guides, and Protocols section below). This creates a common ground to be used as a starting point for other instrumental factors to be accounted for. After instrument calibration, the acquisition protocol must be optimized. The performance of a prediction model from Raman spectra will be directly related to the quality of the data used for model generation. As such, the minimum acceptable data quality must be determined, which can be examined primarily by the signal to noise ratio (S/N). The parameters that determine the S/N, together with typical values and limitations, are listed in Table I. Despite the wide range of possible Raman instrument configurations, dispersive Raman spectrometers are characterized by three main parameters when collecting data through a specific optical path: laser power, integration time for each accumulation, and number of averaged accumulations in the spectrum. In case of a focusing optics, the laser focus will also have a significant impact. Along with sample properties such as Raman scattering cross-section, topography, or transmissivity, the incident light intensity will ultimately determine the S/N (Eq. 1) of a single spectrum<sup>5,6</sup>

$$S/N \propto \sqrt{I} = \sqrt{NI_n} = \sqrt{NpAt} \quad (1)$$

The model's quality depends as well on the quantity of data used for modeling. As for all other forms of statistical and predictive analytics, the general rule is that the larger the training set, the better the result. Specifically with regards to regression models, the number of observations should be at least one order of magnitude larger than the number of variables considered in the model, as a lower ratio of observations to variables will lead to misleadingly high covariance statistics.<sup>7</sup> Moreover, the quantity of data to be collected must take into account if the model validation method requires samples separated from the primary dataset, see the Model Validation (Performance) section below.

### Spectrum Processing

In order to be used for model generation, the input spectra must be comparable with one another. Thus, in general, multivariate data analysis should not be applied to unprocessed spectra, which can contain significant artifacts and structural noise. Models will be sensitive to these variations, which do not describe the analytes being modeled and therefore will lead to misinterpretation. As such, processing techniques should be applied to the data to remove erroneous and misleading variations, while maintaining the integrity of the data itself.<sup>8</sup> Methods of data processing for vibrational spectroscopy may be divided into two categories, namely, filtering and model-based methods. Filtering methods, like calculating second derivatives or performing vector normalization, simply transform spectra into a presumably “better” version of the same data, removing some undesired types of variation. Meanwhile, model-based methods allow for quantifying and separating different types of physical and chemical variations in the spectra, for example, when a fluorescent baseline is modeled (“guessed”) and separated from the remaining data.<sup>9,10,11</sup> Ryabchykov et al.<sup>12</sup> distinguish three phases of Raman data processing: (i) pre-treatment (including, e.g., data import/conversion, spike removal, and calibration), (ii) pre-

processing (including, e.g., smoothing, baseline separation, normalization, and nonlinear scaling), and (iii) processing or model generation. Following this scheme, several Raman spectrum data processing methods typically employed by users are described in the following sections. Note that a complete list is beyond the scope of this review.

### Pre-Treatment

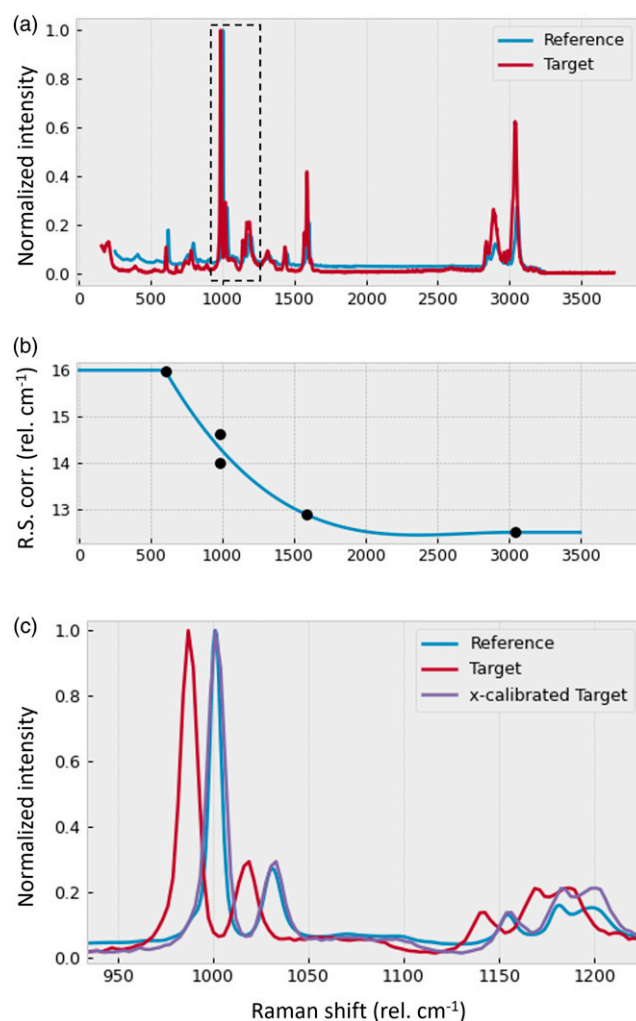
Pre-treatment of input spectra has a strong impact on harmonization models. During Raman data collection, raw data from the detector, usually a charge-coupled device (CCD) array, are converted into intensity versus primary Raman shift. In commercial devices, this first pre-treatment step, which often also applies the instrument calibration, is usually performed by the acquisition firmware and hidden from the user, although some corrections may be optional. Then, the obtained set of spectra must be checked for spikes and outliers before proceeding with (pre)processing.

### Instrument Calibration

Each detector and spectrometer in any given Raman device must be calibrated and verified. During the setup of a Raman device, well-established calibration procedures are performed that require a consistent and careful approach. Raman spectrometers must be calibrated independently for both axes, that is, the  $x$ - (wavelength calibration) and  $y$ -axis (intensity correction). Relevant calibrations require executing multiple, carefully designed steps to generate calibration data for the whole device. Hereby, Raman spectrometers and detectors are prepared in the same way as for other spectroscopy techniques, based on correlating detector pixels with the known signal generated by external references, such as gas emission lines.<sup>13</sup> Raman shift  $x$ -axis calibration is then performed by defining the incident laser wavelength/wavenumber as zero Raman shift, with the rest of the axis being redefined as the wavenumber difference between the pixel assignment and the laser line. There are several standard protocols and guides for the calibration of Raman instrumentation, and it is recommended to follow the appropriate one for the instrument's use (i.e., pharmacopeia for pharmaceutical instruments). Ideally, calibrations must be done not only per instrument, but rather per mode of measurement, depending, for example, on the excitation laser wavelength, sample mounting, sampling technique, or laser power output. However, this is not a standardized practice across system manufacturers and users. Since the instrumental response is subject to a steady and significant temporal drift, the calibrations—or at least their verifications—should be repeated frequently.<sup>14</sup>

### Wavenumber Calibration

Raman spectrometers are typically calibrated on the wavenumber ( $x$ ) axis using standard light sources, such as neon or



**Figure 1.** Raman  $x$ -axis calibration using a pair of target and reference spectra. (a) Spectra of the same polystyrene sample are recorded using the target and reference instruments. (b) The relative shift (RS) at a number of prominent, well-separated peaks is calculated (black dots), and the shifts fitted to a third-order polynomial (blue curve). (c) The resulting calibration can be used to compensate for the relative  $x$  shift, boxed region in (a). Note that the shifts calibration outside the range of measured  $x$  shifts in  $< \sim 500 \text{ cm}^{-1}$  and  $> \sim 3000 \text{ cm}^{-1}$  in (b) was set to the boundary values, since the polynomial rapidly diverges and may result in unreasonably large shift values.

mercury–argon lamps, with sharp emission lines between 250 and 1700 nm wavelength. The known line positions are compared to the measured peak positions, and a wavenumber calibration is generated accordingly. This procedure often requires technical modifications of the Raman setup, such as re-wiring of optical fibers, which sometimes cannot be performed by standard users. Furthermore, the procedures often bypass the optical path that the Raman-scattered light rays would normally follow, and thus deviations incurred by some components along this path may not be accounted for.

The Raman shift denotes the difference, in wavenumbers, between the excitation laser and the Raman-scattered light. As such, the Raman shift conversion of the resulting wavenumber axis is only possible if the excitation source's wavelength is accurately known. In addition to the initial wavelength calibration, x-axis calibration is required, which can be performed by typical users using standard samples (Fig. 1). The procedure can be described as follows:

- (i) A spectrum of a defined standard sample is measured with the Target instrument.
- (ii) A defined set of anchor points  $x_i$  is located, which are typically chosen as the positions of prominent peak maxima.
- (iii) For each anchor point, the wavenumber shift  $s_i(x_i)$  with respect to the Reference is calculated. The latter can be a list of known peak positions, or a reference spectrum recorded with a well-calibrated instrument.
- (iv) A continuous, parametrized curve  $s(x)$  is fitted through the graph of wavenumber-dependent shifts (e.g., a polynomial or spline). This wavenumber calibration curve is stored, assigned to the specific Target instrument and mode, and given a timestamp.
- (v) For all following measurements performed with the Target instrument, the original x-axis is substituted by the new, calibrated axis  $x'$ , with  $x' = x + s(x)$ .

Dörfer et al. reported that they could not eliminate instrument-dependent differences in Raman spectra of the same material even after applying a wavenumber calibration following the standard procedure described above.<sup>4</sup> They also indicate that the results could be improved by additional wavenumber adjustment using an alternative chemometric procedure based on cross-correlation (or HQI). This suggested procedure is described in the Spectral Alignment section below.

### Intensity Calibration

Calibration of the intensity (y) axis is performed to account for the variation in quantum efficiency and dynamic response of the detector across the spectrometers spectral range. This is especially important for Raman harmonization, as the spectral wavelength range varies depending on the excitation wavelength used by the Raman device. This leads to detectors with a wide variety of quantum efficiency curves which, without calibration, would make quantitative inter-instrument comparison impossible. Furthermore, the spectrometer's response to light depends on multiple factors, most significantly the wavelength of the light hitting the pixel. For instance, in a 532 nm Raman spectrometer the wavelength of the Raman scatter of  $3200\text{ cm}^{-1}$  is approximately 641 nm, while in a 785 nm Raman instrument it is approximately 1048 nm. The quantum efficiency of a CCD detector is significantly lower at 1048 nm than at 641 nm, and as such the intensity of the scatter appears reduced in that region.

A wavenumber calibration, as described in the previous section, and, if required, a baseline separation (see the Background and Baseline Correction section) must be performed prior to intensity calibration. Instead of a set of sharp peaks, a standard material for intensity calibration should generate an intense, constant signal over a wide wavenumber range, such as the standard SRM 2242.<sup>15</sup> A wavenumber-dependent intensity modulation is then calculated by dividing the Reference spectrum by the Target spectrum. This will exacerbate any noise in the regions of a spectrometer with a low response to light, which is particularly noticeable when correcting 785 nm based systems. One method, proposed to avoid this problem, is to apply the intensity correction in reverse to an intensity corrected library spectrum and then compare that with the spectrum collected on a separate instrument.<sup>16</sup>

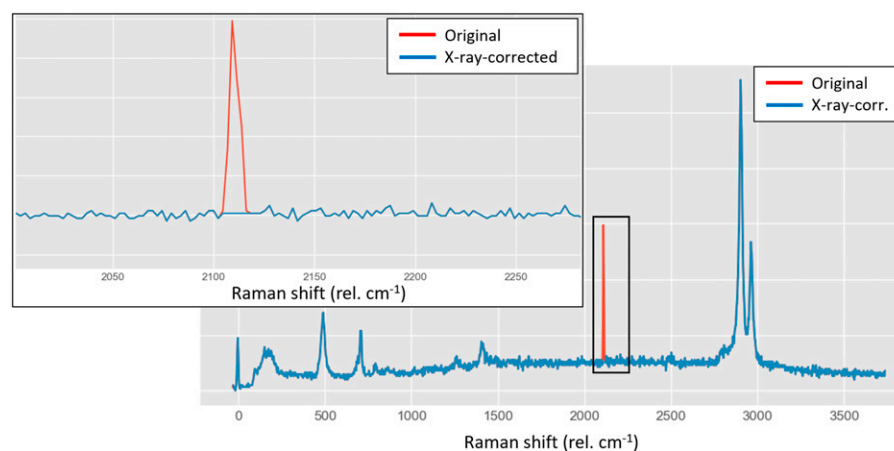
### Spike Removal

Spikes are a common artifact in Raman spectroscopy and appear as a consequence of cosmic rays (X-rays) hitting the detector during the acquisition of a spectrum. Cosmic rays can cause an intense and sharp signal burst at single pixel position on a CCD camera that typically spreads to a few neighboring pixels, thus giving rise to a corresponding spike in the digital Raman spectrum (Fig. 2). Spikes represent prominent features in an unprocessed Raman spectrum and, if uncorrected, can lead to errors in subsequent data processing steps such as normalization, peak search, calibration, or dimensionality reduction. Numerical spike correction is thus mandatory if these methods or other multivariate analyses are to be applied unless the section of the spectrum is ignored. After Ryabchikov et al.,<sup>12</sup> the correction can be performed by:

- (i) (Unspecific) filtering of the entire spectrum, for example, by median smoothing or wavelet (Gabor) transformation.
- (ii) Setting markers that locate channels affected by spikes, followed by replacing the large count values in the affected areas. Markers can be generated, for example, using nearest neighbor comparison through the discrete Laplace operator (numerical derivative, see Fig. 2) to the standard deviation.

### Outlier Detection/Rejection

An outlier is a single Raman spectrum being part of a larger dataset, such as a time series, which shows significant differences to the dataset average not related to sample properties. Outliers can occur for a multitude of reasons, including errors related to the user, instrument, sampling, or processing, and can have a significant effect on the predictive capability of a model. As such, their removal from the training dataset is key to creating a reliable model. Detection of outliers can be performed manually during data collection when obviously uncharacteristic features are observed in the spectrum. Besides, algorithmic methods can



**Figure 2.** Example of a simple, marker-based spike removal algorithm working on a noisy Raman spectrum of solid polyethylene. First, the spike is located via a threshold to the modulus of the numerical spectrum derivative along the x-axis (boxed area). Then, the y values in the spike interval (red), usually a few spectrum channels (four in this case), are substituted by the average counts of six channels adjacent to the spike interval.

also be employed to identify outlier data, which removes user bias and speeds up the review of large datasets. Examples are:

- (i) Parametric methods: RMSSR and SRVIV<sup>17</sup>
- (ii) Linear modeling methods: PCA and SVM<sup>18,19</sup>
- (iii) Proximity-based methods: NND, NNMD, cluster analysis, and local outlier factor<sup>17,20,21</sup>
- (iv) High-dimensional methods: HCS<sup>22</sup>

### Pre-Processing

Pre-processing is required to give relevance to data. Special care must be taken to avoid overprocessing, where the part of the data removed from the processed spectrum is of significance to the sample's response and not due to variance from other sources (e.g., instrumental or environmental effects). Over-processing can also lead to sample-dependent variations being unproportionally weighted within the model, and thus to over and underfitting (see the corresponding section) as well as misleading model classifications or predictions.<sup>23</sup> Some examples:

- (i) Over-smoothing can remove a shoulder peak from a larger peak.
- (ii) Rigorous baseline separation can make peak maxima shift away significantly from their original position.
- (iii) Multiple peaks can be fit to a single response peak, and while this may improve the residual of fitting, it may not be based on reality.

In the following sub-sections, four steps of Raman spectrum pre-processing are described, which are considered essential to data harmonization.

### Noise Reduction

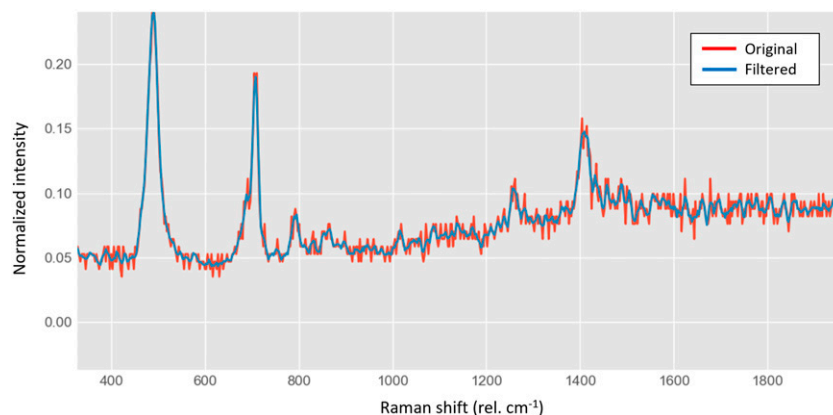
According to common literature on Raman spectroscopy, two main types of noise occur in digital Raman spectra:<sup>24</sup>

- (i) Additive noise with Gaussian distribution, originating mainly from detector readout and dark current.
- (ii) Intensity-dependent noise with a Poisson distribution.

Both types of noise can be reduced by spectrum smoothing, typically applying Savitzky–Golay (Fig. 3), mean, median, or Gauss filters.<sup>24</sup> Note that all smoothing algorithms will lead to some information loss, since they correspond to annual filters in Fourier space.<sup>26</sup> Therefore, smoothing should only be used where absolutely necessary, or for intermediate steps such as baseline model fitting. Information loss by smoothing can be mitigated by the implementation of maximum likelihood estimation (MLE).<sup>25</sup> By restricting the behavior of the smoothing algorithm based on the distribution of the noise, MLE is able to reduce the impact of smoothing on the Raman signal. A discrete wavelet transform (DWT) can also be used to reduce the noise.<sup>26</sup> DWT can detect and filter out gaussian noise based on a thresholding value but does not have a significant impact on the intensity-dependent noise. This process can be iterative, using different threshold values for each iteration to remove the noise.

An enhancement of the signal-to-noise ratio can be achieved without information loss through averaging of multiple spectra, for example, of a spectral time series. Here, the S/N will improve with  $N^{1/2}$ , where  $N$  is the number of spectra averaged. The same holds for many methods of dimensionality reduction relying on averaging, such as principal component analysis (PCA) or non-negative matrix factorization (NMF) (see previous Spectrum Processing section).





**Figure 3.** The effect of second-order Savitzky–Golay filtering with a window length of seven channels (blue), shown on an interval of a Raman spectrum of polyethylene (red; see Fig. 2).

### Background Correction and Baseline Separation

Spectral baseline is affected by several factors: environmental, instrumental, and sample-related. Background correction is the subtraction of any environmental and instrumental signatures detected by the spectrometer when there is no laser excitation. Background correction, if correctly applied, will improve the quality of the data. The background includes, but is not limited to:

- (i) Detector dark current and baseline offsetting.
- (ii) Environmental emissions detected by the spectrometer.
- (iii) Self-emission from the sample.

The baseline is also often altered by factors that are observed only under laser excitation. Relevant sources of baselines in a Raman spectrum are:

- (i) Rayleigh scattering (non-constant, more intense at small Raman shifts).
- (ii) Fluorescence (non-constant, low-frequency curve).
- (iii) Raman signatures originating from impurities, contamination, or substrates.

Baselines can interfere with the analysis of the underlying Raman spectrum, and sometimes even dominate the Raman peak intensity, such as in the case of fluorescent biological or polymer samples. Therefore, an essential pre-processing step is to separate the Raman signal, typically consisting of a set of sharp peaks, from its baseline. We deliberately avoid the term “baseline correction”, since all components of a Raman spectrum can potentially contain sample information and should thus be preserved in general. Therefore, we refer to “baseline separation”. Separating and preserving both the sharp Raman peaks and the baseline avoids information loss, since it allows for recovery of the original spectrum.

Baselines can often be modeled by a smooth function, which is then subtracted from the raw spectrum. Methods for separation of broad baselines include asymmetric least squares (ALS; Fig. 4),<sup>27</sup> advanced penalized least squares approaches,<sup>28</sup> sensitive nonlinear iterative peak (SNIP) algorithms,<sup>29</sup> Goldinec,<sup>30</sup> spline-fitting methods, such as morphology-based cubic p-spline fitting<sup>31</sup> or b-spline fitting,<sup>32</sup> piecewise polynomial fitting,<sup>32</sup> wavelet feature points,<sup>33</sup> convolutional neural networks,<sup>34</sup> morphological and mollifier-based baseline corrections,<sup>35</sup> numerical derivatives, and rolling ball techniques.<sup>36</sup>

A number of the listed methods are available as part of the SNIP,<sup>29</sup> ISREA,<sup>37</sup> and HyperSpec<sup>38</sup> software packages.

Like other pre-processing steps, baseline separation can drastically impact the performance of Raman-based parameter prediction models. Schulze et al. compared and evaluated several baseline separation techniques for automated implementation.<sup>39</sup> Guo et al. report a method of automated selection of a baseline algorithm that optimizes a specific model.<sup>40</sup> This potential model specificity of the baseline separation method underlines the need to preserve the original training data with any model to modify the baseline separation procedure if needed.

### Normalization

Intensity variations of Raman spectra can be dramatic and occur between laboratories, instruments, measurements, and even within one measurement (e.g., maps or time series). They can originate from instrumental or experimental factors, for example changes in laser focus or detector response. A global, that is, wavenumber-independent, scaling of Raman spectra can be mitigated by a number of normalization techniques, including<sup>41</sup> vector normalization, area normalization (i.e., normalization to the integrated spectral intensity), standard normal variate, and min–max scaling.

All algorithms listed above are based on spectrum modification by two constants  $a$  and  $b$ , according to Eqs. 2–4, such that each spectrum channel  $y_i$  is modified as follows:

$$y_i^* = \frac{y_i - a}{b} \quad (2)$$

For min–max scaling,

$$a = \min_i(y_i) \quad (3)$$

$$b = \max_i(y_i - a) \quad (4)$$

Spectrum normalization is widely applied to improve comparability of Raman data, and will also improve the performance of many prediction and decomposition models. However, it cannot account for wavenumber-dependent gain artifacts, where different scaling applies to different Raman peaks. Such effects must be compensated by appropriate intensity calibration (see Intensity Calibration section above).

### Non-Linear Scaling

Raman spectrometers will output data on a linear scale, assuming an intensity corrected spectrum (see “Intensity calibration”). For quantitative analysis, it is often assumed that the relative amplitude of a substance-characteristic peak is proportional to its relative content within the sample. However, the relationships of peaks and their behavior do not always act linearly, as can be seen in the Raman peak intensities from graphene.<sup>42</sup> Raman spectra can be scaled in a non-linear fashion to increase the relative weights of some spectral features, such as through a logarithmic, square root or power transformation. For example, if  $(x, y)$  represents a Raman spectrum, then  $(x, \log(y))$  is taken as the processed spectrum. This is often done to increase the performance of prediction models, such as partial least squares (PLS) regression.<sup>43</sup>

### Processing and Model Generation

Chemometric models are statistical, or machine learning models typically implemented in the form of numerical structures (classes), that is, computer programs, which predict sample properties or parameters from digital Raman spectra. In general, those models can be generated through supervised or unsupervised learning.<sup>44</sup> Models should be trained on a sufficiently large number of known Raman spectra, the “training set” (also referred to as fitting; see Model Validation (Performance) section). Afterward, the trained model is saved as a data structure, which can be later used for predicting properties of unknown samples from their Raman spectra (queries). This process is also referred to as model deployment. To consistently harmonize Raman spectroscopic data across different instrumental designs a chemometric model must account for the effect of the component variations on the resulting spectrum.

Predictions can be categorical, in the case of a classification (e.g., recognition of bacterial species), or numerical, in the case of a regression (e.g., for predicting the composition of a polymer blend). A recently reported example of classification for Raman spectra is the SIMCA method, a supervised pattern recognition method in which a PCA model is constructed for each material class; for classification of a spectrum, its residual difference with each class is calculated.<sup>45,46</sup> Particularly for pharmaceutical applications, the classification criterion HQI > 0.950 is often used as a simple model for material verification (classification).<sup>47</sup>

As an example, calibration and deployment of Raman models for pharmaceutical online process monitoring has been described to involve four main steps:<sup>43</sup> (i) Collection of well-characterized spectral and analytical data for the parameters of interest, either from existing databases (“historical” data) or from new measurements, (ii) spectral data pre-processing for S/N improvement, (iii) correlation of pre-processed spectra to the parameters of interest by statistical methods. The resulting model must be validated, and (iv) construction of a visual tool to use the calibrated model online for real-time monitoring and control applications.

A large number of statistical and machine learning models have been developed, and Fig. 5 shows some of the most relevant. Most of them are readily available within libraries of established scientific programming languages, such as Python’s Scikit-Learn library,<sup>48</sup> Matlab, R, and C++.

Unsupervised models are used if no training data with known properties (targets) are available, for example clustering; vertex component analysis;<sup>49</sup> or unmixing, for example, using NMF, PCA, independent component analysis (ICA), modified alternating least squares,<sup>50</sup> and multivariate curve resolution alternating least squares (MCR-ALS).

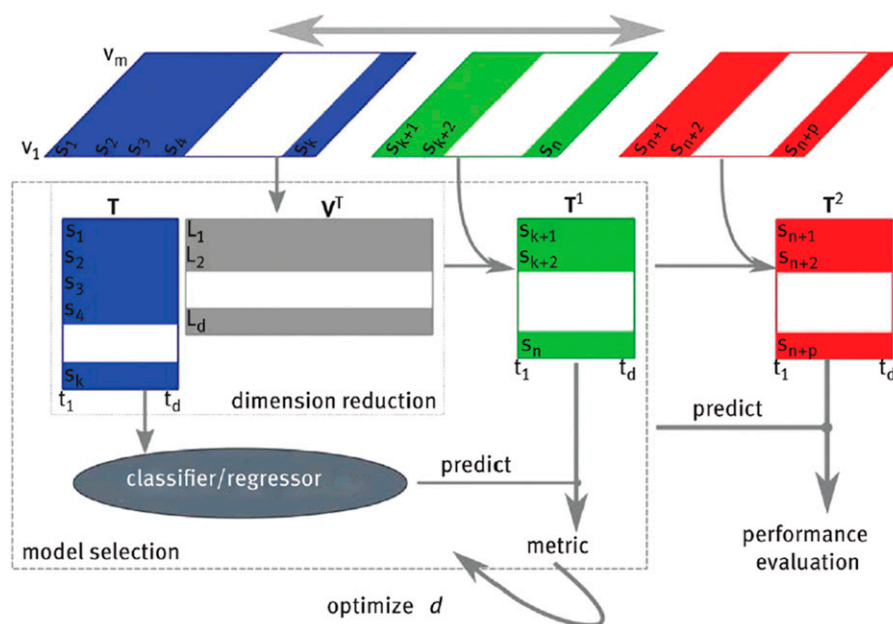
Supervised machine learning algorithms, like regression or classification models, can be applied if training data with targets are available to extract high-level information, such as linear classification and regression methods (e.g., multilinear regression or PLS regression); kernel support vector machines (SVMs);<sup>51</sup> artificial neural networks;<sup>52</sup> random forest (RF);<sup>53</sup> and linear discriminant analysis.<sup>54</sup>

A supervised model (e.g., Fig. 6) is typically generated in several steps (Fig. 7) that will be commented on in the following sections:

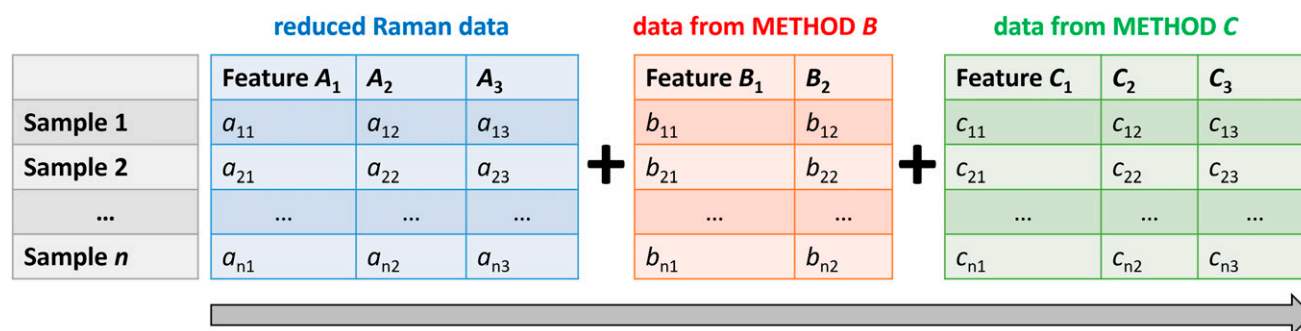
- (i) (Centralized) data fusion: The complete dataset is converted to a matrix composed of the individual spectra as lines. This requires appropriate pre-processing and interpolation of all spectra to a common x-axis. The columns of the resulting matrix are referred to as “features”, while its lines are “instances” (measurements).
- (ii) Dimensionality reduction: Use of algorithms that reduce the number of features (Fig. 10) and thus enable a robust prediction.







**Figure 7.** Schematics of a supervised model generation from a set of Raman spectra  $s_i$ . The authors suggest an additional step of optimization for dimensionality reduction, which requires data split into a third subset (evaluation). Reproduced with permission from Ryabchykov et al.<sup>12</sup>



**Figure 8.** Scheme for high-level data fusion of Raman data with data from other (hyphenated) methods. Raman data is subjected to dimensionality reduction to three scores (features), and then combined with data from two other analytic methods carried out on the same samples. The features are combined as columns of a fused data matrix, which can then be used for analysis and model generation. In this case, the result is an  $n \times 8$  data matrix.

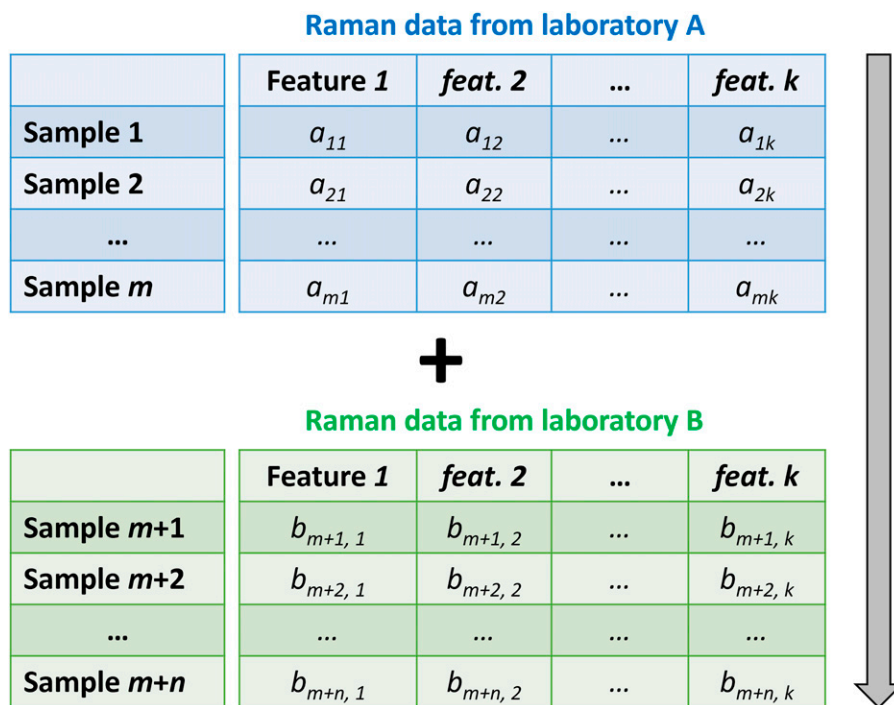
data from sources other than Raman spectroscopy. In principle, one can differentiate two types of data fusion:

- (i) (Horizontal) DF: The same set of samples is analyzed by Raman plus other analytical methods.<sup>56,57</sup> In this case, only columns/features are added, while the samples or data instances, that is, the matrix rows, are the same for all datasets. After appropriate dimensionality reduction or feature extraction, the resulting method-specific data matrices are merged (concatenated) horizontally to obtain a unified data matrix (Fig. 8).
- (ii) Data augmentation:<sup>58</sup> A Raman dataset is merged

with another Raman dataset either from the same or from a different set of samples. In this case, the features, that is, matrix columns, are the same for all datasets. This requires some overlap on the Raman shift ( $x$ ) axis and either dimensionality reduction or interpolation of spectra to a common  $x$ -axis (Fig. 9).

Furthermore, Ryabchykov et al.<sup>12</sup> describe two levels of data fusion:

- (i) Centralized DF is carried out directly after pre-processing, possibly even before dimensionality reduction,<sup>57</sup> by merging data from different sources into



**Figure 9.** Scheme for data augmentation (centralized or high-level) of Raman data. For unprocessed data, features can be the detector counts for a specified set of Raman shifts, while for processed data they may, for example, be weights of PCA components (scores). The result is an  $(m + n) \times k$  data matrix.

a single data matrix with subsequent simultaneous analysis.

- (ii) High-level DF, also called distributed DF, analyses each data type separately and combines the scores at the final step.<sup>59</sup>

In any case, the fusion of data following either of the described schemes will lead to some information loss, resulting either from dimensionality reduction or from x-axis interpolation and rescaling. Therefore, fused datasets should be temporal and only generated for a specific analytic purpose, for example, the training of a chemometric model. The original data should always be preserved alongside the fused dataset.

### Peak Fitting

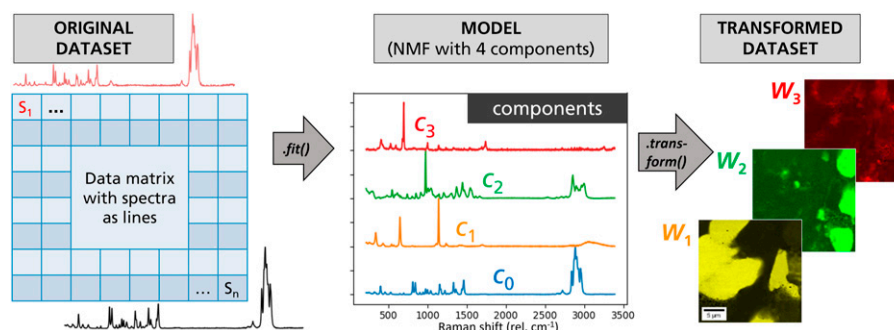
A common mode of quantitative analysis is fitting of a mathematical model function to one or multiple peaks (bands) in a Raman spectrum.<sup>60</sup> Typically, Gaussian, Lorentzian, Voigt, or Pearson-type profiles are used, since they often yield a good fit to experimental data.<sup>61</sup> For some materials, inhomogeneous peak broadening occurs, which requires asymmetric models for peak fitting.<sup>62</sup> The model functions yield a number of univariate peak features such as amplitude, maximum position, centers of mass, or full width half-maximum. All features derived from the intensity ( $y$ ) axis should either be measured in a normalized spectrum or given relative to a

reference peak. Note that the representation of a Raman spectrum by univariate peak features, such as intensity ratios or parameters of a fitted Gaussian profile, represents a form of dimensionality reduction (see the corresponding section).

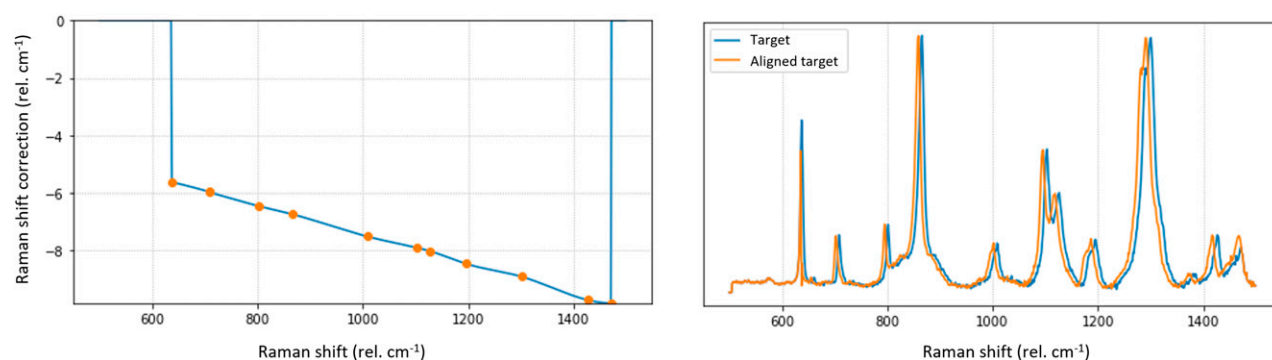
### Dimensionality Reduction (Model-Based)

Raman spectroscopy often produces large datasets that can have high dimensionality. For example, a dataset containing 10 Raman maps of  $100 \times 100$  spectra each and 1600 channels per spectrum will have four dimensions and consist of 160 million numbers. In order to make such datasets accessible to human observers, and enable a robust analysis and model generation, the dimensionality must thus be reduced, while preserving as much dataset information as possible. A schematic example is shown in Fig. 10. In many cases, dimensionality reduction (DR) is a prerequisite for fusion of Raman data from different sources, or Raman data with data from other analytical methods (see Data Fusion section above). The most common DR algorithms are unsupervised feature extraction methods, such as PCA,<sup>63</sup> independent component analysis (ICA),<sup>64</sup> NMF,<sup>65</sup> and multivariate curve resolution alternating least squares (MCR-ALS).<sup>66</sup>

Factor methods reduce dimensionality in terms of some factorization of the original data matrix, which has the individual  $n$  Raman spectra with  $k$  channels each as lines.<sup>12</sup> After DR, the original  $n \times k$  matrix is factorized into the  $m \times k$  matrix of  $m$  spectral components and the  $n \times m$  matrix of weights or scores, with  $m \ll n, k$ . A visualization of the  $m$  components is



**Figure 10.** Schematics of dimensionality reduction using a factor method (NMF) for a confocal Raman map, an example of a large Raman dataset (left). First, all Raman spectra included in the map are stacked as lines of a common matrix. To the latter, a NMF model is fitted, which contains four components (center): the polymer matrix (blue) and three types of flame retardant additives (orange, green, and red). The transformed dataset (right) has reduced dimensions, and each spectrum of the original dataset is represented by four scores  $W_n$ , corresponding to the components  $c_n$ . If the original dataset is a Raman map, the transformed dataset can be represented as a series of images (scale bar: 5  $\mu\text{m}$ ).<sup>68</sup>



**Figure 11.** Demonstration of chemometric spectrum adjustment implemented in a CHARISMA software prototype.<sup>97</sup> (left) A wavenumber-dependent negative shift (top hat plus ramp) was applied to a Raman spectrum. The shift is correctly detected by maximizing the HQI of original and shifted spectrum through simulated annealing (orange circles). A continuous shift function is modeled using a cubic spline (blue curve). (right) The target (shifted) spectrum is adjusted (calibrated) using the shift model.

usually called a loading plot (see Fig. 10). Once the factorization model is fitted, any Raman spectrum with  $k$  channels can be transformed into  $m$  scores, indicating the weights resulting from the decomposition of the target spectrum into the  $m$  inherent components of the model. In particular, NMF and MCR-ALS can provide physically and chemically meaningful decompositions, where loading plots correspond to Raman spectra of the pure components.

An alternative route to dimensionality reduction is variable or feature selection, where a subset of spectrum channels is chosen bearing the highest relevance for a subsequent prediction or decomposition model, for example, by means of a genetic algorithm.<sup>67</sup> Dataset dimensionality can also be reduced by extraction of a number of univariate features, such as (fitted) peak amplitudes, widths, or ratios.

### Over- and Under-Fitting

An overfitted model is defined as a “model that contains more unknown parameters than can be justified by the data”.<sup>69</sup> This

means that the model is too precisely tailored to the training dataset, so that the model cannot fit new data or reliably make predictions. The model is therefore hardly accounting for the defining variables but rather for artifacts and/or the noise.

An underfitted model is defined as a “model where some parameters or terms that would appear in a correctly specified model are missing either by mistake or by design.”<sup>69</sup> This means that the model is excluding variables that would better define the observations.

Both overfitting and underfitting will increase the error in classification or prediction performed by a model. By following the principle of parsimony, the user should find the adequate parameters to define a model. One can both loosen and tighten the parameters used, comparing the error of the model, to find the lowest possible number of parameters. The parameter analysis and model creation can be guided by theory; it can help streamlining the process. Tools such as Akaike’s information criterion can be used to compare several models constructed from the same dataset and indicate which is best in terms of statistical inference.<sup>8</sup> In the case of

supervised models, over- or under-fitting can be mitigated, for example, by grid search optimization over relevant model parameters, such as the number of components in PLS. For unsupervised approaches, the application of chemical or physical constraints can be appropriate, such as setting the number of PCA or NMF components equal to the number of expected chemical components in the sample.

### Data Structure of Chemometric Models

In contemporary scientific programming languages such as Python, models can be generated, trained, saved, and used in the form of class instances. Classes are data structures that provide a means of bundling data and functionality together.<sup>70</sup> Each class instance can have attributes attached to it for maintaining its state. Class instances can also have methods (defined by its class) for modifying its state. A model class for prediction from Raman spectra typically has a `fit()` method for training and a `predict()` method for prediction, plus attributes storing a numeric representation of the trained model, for example, a linear regression or neural network classifier. An instance of a model class can, for example, be serialized and saved to disk as a binary file. A similar implementation exists for other programming languages such as Matlab or R.

### Model Validation (Performance)

Validation and continued monitoring of a model are important to its performance and longevity, respectively. For harmonization, an increased longevity of the model is desired, which can be achieved by the reliable maintenance of the instruments throughout their lifetimes. Therefore, the model performance must be constantly monitored after deployment, and re-training must be applied if required.

Models can be quantitatively validated and scored using one of the following sampling schemes for a test/train split of the dataset:<sup>71</sup>

- (i) Cross-validation splits the data into a given number of folds, usually 5 or 10,<sup>72</sup> holding out one fold at a time for testing; the model is trained using the other folds. This process is repeated until all the folds were held out.
- (ii) Leave one out is similar, but it holds out just one instance at a time, training the model with all the others and then testing it with the held-out instance. This validation process is very stable and reliable, but also slow.
- (iii) Random sampling randomly splits the data into a training and a testing set in a given proportion (e.g., 65:35) for a specified number of times.

Using an appropriate sampling, one or more scores are then calculated following established metrics for regression and classification.<sup>71</sup>

In regression analyses, many metrics for regression performance are derived from the error of a single prediction,

defined here as the difference between the true value of the dependent variable and the value predicted by the model. First, the mean error of prediction (MEP) is the average error of predictions for the test dataset. It is equal to the bias of a model. A model is biased if it makes predictions with a non-zero MEP. Second, the mean absolute error of prediction (MAEP) is the arithmetic average of the absolute errors (how close forecasts or predictions are to eventual outcomes). The MEAP corresponds to an  $l_1$  norm and is generally less sensitive than the root mean standard error of prediction (RMSEP). Third, the mean squared error of prediction or deviation (MSEP or MSD) is the average of the squares of the prediction errors for the test dataset. Finally, the root mean square error of prediction or deviation (RMSEP or RMSD) is the square root of the MSEP. It is equal to the Euclidean or  $l_2$  norm, and therefore more sensitive towards large deviations as compared to the MAEP. The RMSEP is the most common measure of performance for regression models. The coefficient of determination ( $R^2$ ) is interpreted as the proportion of the variance in the dependent variable that is predictable from the independent variable.

In classification analyses, the area under the receiver-operating characteristic (ROC) curve<sup>73</sup> is a measure of sensitivity and specificity for a binary classifier, that is, a model that distinguishes two classes. A perfect classifier would generate a ROC with an area of 1.0, whereas totally random predictions will result in an area of 0.5. It should be noted that an appropriate external validation set, ideally acquired with instruments or parameters other than those of the training data needs to be chosen to test a classifier using the ROC. If discrimination is based on instrumental or other artifacts rather than on chemical differences, the resulting ROC will probably be misleading. Classification accuracy is the proportion of correctly classified examples.<sup>74</sup> The accuracy can be misleading, especially if classes are unequally represented in the training dataset. For example, a trivial model that always predicts the same class will achieve a high precision if 90% of the training instances belong to this class. The balanced F-score, or  $F_1$ , is a weighted harmonic mean of precision (positive predictive value) and recall (sensitivity),<sup>75</sup> which are the proportion of true positives among instances classified as positive<sup>76</sup> and among all positive instances, respectively. The F-score favors classifiers for which precision and sensitivity are similar. This may not always represent the best metric, especially for models that require either a particularly high sensitivity or precision.

For classification, combined metrics like the sum of ranking difference have been applied that can allow for more stable model evaluations than the single metrics listed above.<sup>77</sup>

### Model Transfer

Without model transfer, chemometric harmonization of multiple Raman instrumentation would be impossible. The requirement for transferability of a prediction model is best

illustrated using a typical use case for chemometric modeling with Raman spectroscopy data:

- (i) A prediction model for a material property is trained during a measuring campaign, where a large set of training spectra are recorded by some laboratory, using a defined set of instruments. This data, upon which the model is generated, is referred to as the primary dataset (PD). For training, the primary dataset is split into training and test datasets.
- (ii) The performance of the model is validated using the primary test dataset. Then, the trained model is stored in a standardized format and made available for future use (deployed).
- (iii) At some point in the future, a second laboratory intends to use the deployed model to predict the sample properties using its own current data (secondary dataset, SD), acquired using instruments and modes different from those used to record the primary dataset.

When a trained chemometric model (built using PD) is applied to newly measured data (SD), predictions are often less precise than those done based on the original test data (PD). This issue gets more important if the new data is significantly different to training data. Large, undesirable spectral variations often pose great challenges for chemometrics, manifested as a failed prediction of the trained model for newly measured datasets. Even if both primary and secondary datasets are stored in the same data format with a common  $x$ -axis, which in principle makes them mutually comparable, the prediction may fail for a secondary dataset that has been measured under different conditions compared with the conditions of the training set.<sup>78</sup> The reason for this is often wavenumber ( $x$ ) shifts and intensity ( $y$ ) variations due to instrumental artifacts, or changes of experimental conditions such as laser wavelength, power, or sample temperature.<sup>79</sup> Moreover, the established calibration procedures yield calibration models that are reliable only in the operating conditions they were calibrated in (see Data Fusion section above). Typically, it will not be possible for the secondary laboratory to reproduce a training dataset that allows for the generation of a new model, since this requires a large number of samples, which may be expensive or even impossible. Therefore, reliable model transfer procedures are necessary to tackle problems related to the above-mentioned undesirable variations (model transfer and calibration transfer are often used synonymously in the literature).<sup>12,80</sup> A number of model transfer approaches has been proposed,<sup>80–83</sup> such as discrete wavelet transformation (DWT)<sup>86</sup> and direct standardization (DS).<sup>85</sup> For near infrared spectra, piecewise standardisation<sup>86,87</sup> and finite impulse response filters<sup>88</sup> have proven to enable the successful transfer of calibration models.

Existing model transfer approaches can be categorized into three strategies:

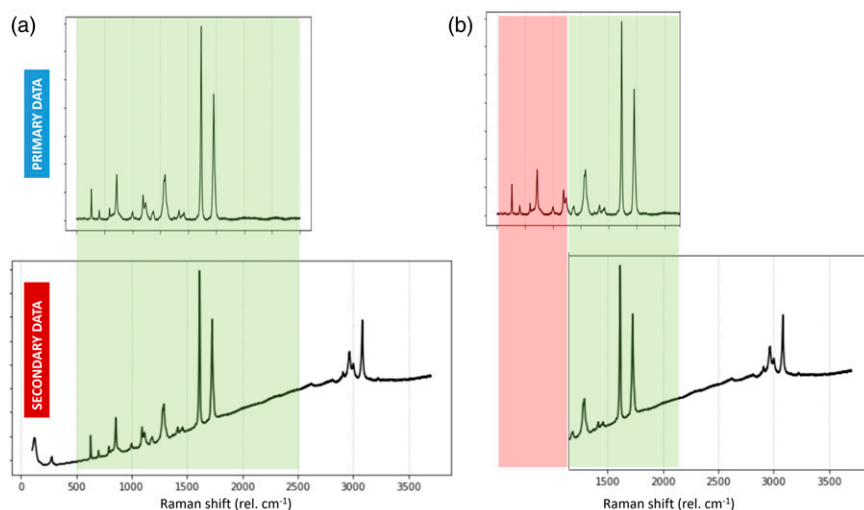
- (i) Build robust models: Raman spectral features are sensitive to measurement variations. To build a robust model, one option is to involve metadata of measurements into model construction, as in the generic solution proposed by Kalivas et al. to always enable the application of Tikhonov Regularization.<sup>79</sup>
- (ii) Standard-free spectral alignment: Use of numeric calculations to remove spectral variations between the primary and secondary data set. Examples for spectral standardization methods used in near-infrared spectroscopy, such as Procrustes analysis, orthogonal signal correction, parametric time warping, or piecewise direct standardization, are still limited in Raman spectroscopy due to the high noise they introduce. Moreover, as purely data-driven procedures, they may remove both sample- and setup-induced variations without distinction. In this context, methods based on bilinear modeling, such as replicate extended multiplicative signal correction<sup>89</sup> and ANOVA-simultaneous component analysis, make it possible to estimate the setup-induced spectral differences and remove them from the data while preserving the sample-induced variations. This can be a more reliable mechanism, although it requires the experiments to be well designed. Such a method is also limited if the sample- and setup-induced spectral variations are not independent of each other.<sup>90–92</sup>
- (iii) Model updating: The aim is to update the model introducing the differences between the primary and the secondary datasets to obtain a new model that tolerates interfering variations.

The following sections focus on spectral alignment and model updating, approaches that require model training based on modified primary and secondary datasets. Both model transfer strategies can be combined, which may optimize the model performance, as shown for the case of biological Raman datasets.<sup>82</sup>

### Spectral Alignment

Spectral alignment does not require the generation of new training data by the second laboratory, because the data, both primary and secondary, are processed in order to make them more similar. However, it does require the primary dataset to be stored with the model. Two methods of spectral alignment are possible: spectrum calibration to a standard, and standard-free spectrum adjustment: For spectrum calibration, the wavenumber ( $x$ ) and counts ( $y$ ) axes of both datasets are calibrated using a standard material, for example, 4-acetamidophenol ( $x$ -axis) and reference





**Figure 12.** Effect of spectral range on model transferability. If the primary range is included in the secondary range (case a), the original model may be used after appropriate truncation and re-sampling of the secondary dataset. However, the model cannot be used if the primary range is not fully included in the secondary (case b). In this case, an appropriate model transfer is mandatory, which requires the training data of the original model.

material SRM 2242 (y-axis);<sup>93</sup> for standard-free spectrum adjustment, both datasets can be adjusted to each other by using the total spectral average of both datasets combined as the reference, and the averages of each individual dataset (primary and secondary) as targets. Then, the targets are aligned to the reference.

In both cases, a set of peak maxima is used as anchor points to which local wavenumber shifts with respect to the reference are assigned. A wavenumber calibration function is then fitted to the anchor points, typically a third-order polynomial. As opposed to the standard calibration method, an optimization step is added for spectrum adjustment where the anchor point positions are refined using a maximization algorithm. Such an algorithm may use, for example, the hit quality index (HQI; Eq. 5):<sup>94</sup>

$$HQI = \frac{(\mathbf{s} \cdot \mathbf{s}_{ref})^2}{\mathbf{s}^2 \cdot \mathbf{s}_{ref}^2} \quad (5)$$

where the  $\mathbf{s}$  are the vector representations of the averaged spectra and  $\mathbf{s}_{ref}$  is the reference spectrum. The HQI is equivalent to the numerical cross-correlation (fig. 11). The refinement of local wavenumber shifts often leads to a significant improvement of the model accuracy. Parameter optimization methods that may be used for this task include simulated annealing<sup>95</sup> or downhill simplex.<sup>96</sup>

Notably, spectrum adjustment must be performed as an ensemble method equally for all spectra in the PD and SD, and not individually for each spectrum, since this may eliminate material-dependent effects. This requires modification of the primary data, and thus a new model training after adjustment. This step may be undesirable and could be avoided if instead of adjusting both datasets to their average, the secondary data is

chosen as the target and adjusted to the PD (reference), while leaving the PD unchanged. However, this is only possible if the PD range is contained in the SD range, so that the SD can be truncated and re-sampled to the PD (Fig. 12).

### Model Updating

Model updating has shown to be effective for prediction models based on Raman data. Rather than modifying the primary or secondary data, model updating requires a small number of training data acquired in the secondary laboratory. Using the primary (spectra  $\mathbf{X}$  and targets  $\mathbf{y}$ ) and secondary ( $\mathbf{L}$ ,  $\mathbf{y}^*$ ) training data, the model is re-generated, assigning an increased weight  $\lambda$  to the SD portion as compared to the PD. This can be achieved by Tikhonov regularization (Eq. 6):

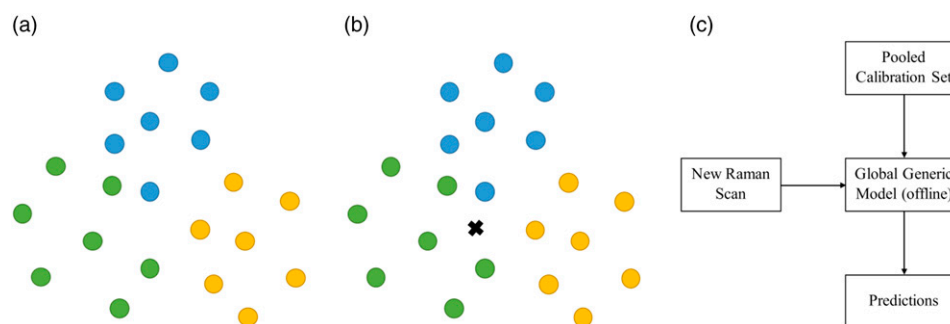
$$\begin{pmatrix} \mathbf{y} \\ 0 \\ \lambda \mathbf{y}^* \end{pmatrix} = \begin{pmatrix} \mathbf{X} \\ \epsilon \mathbf{I} \\ \lambda \mathbf{L} \end{pmatrix} \mathbf{b} \quad (6)$$

The weighted unity matrix  $\epsilon \mathbf{I}$  is inserted to avoid singularity of the matrices.  $\mathbf{b}$  is a matrix representation of the model, e.g., a multilinear regression.

However, the criteria that must be met for applying Tikhonov regularization often limit its utility. For instance, the method is not applicable if there are no secondary targets  $\mathbf{y}^*$  available.

### Other Approaches

Another approach for model generation based on heterogeneous Raman datasets is locally weighted modeling, such as



**Figure 13.** Avoiding model transfer and increasing the accuracy and robustness of global models by just-in-time learning (JITL). (a) The global training set (library) is composed of calibrated spectra from different laboratories (colored circles). (b,c) For each new query (black cross), for example, measured Raman spectrum for prediction, a local model is generated from a local training set. (d) Schematic process for JITL. Reproduced with permission from Tulsyan et al.<sup>45</sup>

localized regression or PLS.<sup>98</sup> In a nutshell, such algorithms identify those training spectra most similar to the one to be predicted, and build a new local model using these training instances only. Tulsyan et al.<sup>43</sup> propose the use of just-in-time learning (JITL) on global generic models. Here, an extensive training data set (library) is generated, including many spectra from different experiments, instruments, and laboratories. Fitting a generic model on this “incompatible” global dataset would result in poor model performance. Instead, JITL calibrates “local” models in real-time upon measurement of a Raman spectrum to be predicted. For local model fitting, the most relevant local calibration set is selected from the library based on the Euclidean distance, and then the model (e.g., PLS) is fitted with this local training set. Once the model predictions are available, the local model is discarded, and the JITL waits for a new Raman spectrum to calibrate a local model (Fig. 13). The authors report a substantial improvement in the online prediction of component concentrations during various pharmaceutical processes by Raman spectroscopy and mention a “paradigm shift in pharmaceutical manufacturing.”

## Existing Standards, Guides, and Protocols

There exist several protocols from multiple sources related to Raman spectroscopy.<sup>99</sup> The main source is the ASTM International organization,<sup>100–104</sup> which has published several standards on multivariate analysis and its use for qualifying and verifying spectrometers.<sup>17,105–108</sup> Besides, the main pharmacopoeias include specific chapters dedicated to Raman<sup>109,110</sup> and are also leading the guidance to chemometric data analysis.<sup>111,112</sup>

## Pharmacopoeias

The pharmacopoeias (mainly European and U.S.) have specific chapters fully dedicated to the use of chemometrics. The EU pharmacopoeia created the chapter 5.21 “Chemometric Methods Applied to Analytical Data” in 2016,<sup>111</sup> and the US pharmacopoeia created the chapter 1039 “Chemometrics” in 2017.<sup>112</sup> The former is a general text, which aims at guiding

the user towards good chemometric practice and introduces chemometric requirements and techniques. It explains the need during the creation of a model for error analysis by RMSEP and standard error of calibration (SEC), as well as outlining the considerations that should be made when selecting samples and processing methods to ensure a good quality model. It also outlines the need for maintenance of the models, with the parameters to be validated for both qualitative and quantitative models. While the EU pharma chapter recommends PCA, PCR, and PLS, it also explains the principles and potential uses for measures between objects, SIMCA, clustering, MCR, MLR, support vector machines, and artificial neural networks. The US pharmacopoeia is similar, but it also guides the user through calibration, validation, maintenance, and update methods in an integrated workflow for lifetime maintenance.

## ASTM International (ASTM)

While the MVDA-focused ASTM standards<sup>17,105–108,113,114</sup> are not always specific to Raman spectroscopy, they are all relevant and transferable to multivariate analysis performed on Raman spectroscopic data. The multivariate techniques covered by the standards are limited: the user must employ partial least squares (specifically the PLS-I algorithm), principal component regression (PCR), or multilinear regression (MLR) to be considered compliant with some of these standards. Moreover, the standards are also limited in their scope, as they do not describe the entire multivariate data analysis method and model creation process, and instead they recommend a subject matter expert to define the process steps that are not described, such as how to treat and process the spectra before model creation without compromising the integrity of the data.

Methodologies and mathematics for qualifying and validating models during development, implementation, and continual running, as well as after model updating are described. This includes flow charts, considerations for the model and for building a calibration set, as well as several

equations for error statistics to highlight issues in the validation as well as outliers:

The SEC and its derivative equation, the root mean standard error of calibration, which are used to assess the agreement in the predicted and actual result.

The *T*-value of the significance of validation bias, which informs the users whether they should use the standard error of validation (SEV) and the standard deviation of the validation residual (SDV).

The equations for the SEV and SDV are altered depending on the calibration model being used (i.e., single reference value and estimate, multiple reference, and single estimate, multiple of both).

The predicted residual error sum of squares, which is used within the standard error of cross validation to assess those models, utilizes a cross validation method. Several outlier statistics are described for assessing the spectra input in creating the model as well as those used to validate the model when implemented and after update. These are the leverage statistic, root mean sum of squares residual (RMSSR), nearest neighbor distance (NND), nearest neighbor Mahalanobis distance (NNMD), and the standard residual variance of the independent variables (SRVIV).

The procedures defined above are described using equations, but the standards also note other techniques for the purposes of model calibration and validation. It is recommended that validation is performed using external and internal data. Internal validation data are obtained from data used to create the model (primary dataset) or by statistical resampling of the analyte. External validation data are independent of the model, and thus can be neither part of the creation dataset, nor collected from the analytes used to train the model (secondary dataset).

The outlier statistics are described to allow the user to identify and exclude data which would be an extrapolation of the model and/or have unusually high leverage on the model. These statistics can be applied to data used to create, validate, or calibrate the model, as well as to data tested against the model.

## Outlook

For models to be developed, validated, and used for predictions, the data involved must comply to certain requirements, that is, there must be common variables across the data. When data is collected from multiple instruments, x-axis matching is easily achieved by interpolation to create a new universal x-axis across the data. However, if the x-axis is not properly calibrated then the model's results will be misleading, as well as in the case of variations in the resolution of the different spectrometers, as a spectrum can lose key features with lower resolution. Therefore, a harmonization effort should be made to account for these effects, and this relies on chemometrics. The implications for data handling and processing can be summarized as follows:

- (i) Calibration of any instrument should be carried out using a set of standards which generate well-defined features in their Raman spectrum, such as peak positions and relative peak intensities.
- (ii) Any data treatment or processing should follow a carefully defined and outlined scheme.

Data pre-treatment must be kept at a minimum. That is, outside of calibration, only spike removal should be performed unless further treatments are required to obtain usable data.

- (iii) Data (pre)processing should preserve the original data. Processing steps should be reversible, reproducible, and well described in terms of their algorithms (logged) within the data file.
- (iv) A common database should be built and made accessible for the intended users, to allow for models to be reconstructed for transfer.

## Spectrometer Calibration

Wavelength calibration of the spectrometer is not sufficient to calibrate the x-axis. The precise laser wavelength must also be known for an accurate conversion from wavelength in units of nm, to Raman shift in units of  $\text{cm}^{-1}$ . Even in cases where the laser wavelength is known with good precision, not all aspects of the instrumentation that can cause deviations can be fully accounted for. One of the reasons is that the spectrometer is usually calibrated independently, thus ignoring deviations due to other components within the Raman instrument, which should be considered as a whole to reduce the sources of discrepancy when comparing separate Raman instruments. Commonly, a practical solution is analyzing a Raman shift standard and using it to fine tune the Raman shift calibration. The spectrum adjustment by HQI described by Rodriguez et al.<sup>115</sup> may mitigate artifacts introduced into spectrum wavenumber calibration by unequal spectral resolution, which may initially lead to errors in peak locations, and thus in the Raman shift calibration model function, usually a polynomial.<sup>116</sup>

One calibration which is often overlooked as an insurmountable hurdle is that of the resolution, which should not be confused with interpolation of the x-axis for variable matching. The most straightforward way is smoothing the higher resolution spectrum until its resolution is equal to that of the lower resolution spectrum.<sup>47</sup> This will improve the comparability, but also cause some information loss. Alternatively, through use of a double digital projection slit (DDPS) it appears possible to in fact simulate higher resolution in low resolution data.<sup>80</sup> Using a virtual ideal spectrum of the calibration material, DDPS can improve matching between low- and high-resolution spectra, and visual inspection of the spectra shows deconvolution of peaks that were indistinguishable in the low-resolution spectrum. The reliance on a virtual ideal spectrum may prevent the use of DDPS on wide scale harmonization. The use of the HQI as target

function for spectral alignment takes all spectral values into account, rather than only the peak positions, and may also be able to correctly align spectra of unequal resolution.<sup>115</sup>

Currently there is no literature that demonstrates intensity correction outside of white light and standard fluorescent glass samples. The HQI can be affected by wavenumber-dependent intensity variations, caused by different excitation wavelengths, which impacts the intensity of Raman scatter observed at higher Raman shifts due to the spectrometer quantum efficiency as well as deviations due to resonance effects. This issue with HQI is studied by Park et al. to determine the applicability of modified versions of the HQI approach. It is believed by the authors that following the work of Guo et al.<sup>83</sup> describing the adjustment of wavenumbers (x-axis) using HQI, it may in principle also be used for intensity adjustment (y-axis).

### Data Fusion and Model Transfer

Data fusion and model transfer are core elements for the harmonization of Raman spectroscopy, especially when applied to a variety of materials. A high-level data fusion is implemented in harmonization models. For the longevity and integrity of the harmonization system, the original data should always be preserved alongside the fused dataset. A centralized data fusion can be implemented in the form of file groups, where a file group is generated each time a number of spectra are compared or used for model generation. This process should be computation-efficient, since it may have to be performed many times during search in a large database. While appropriate data fusion methods are a prerequisite for developing a broad database, the reusability of existing prediction models should be a core value in the databases creation. However, relatively little literature exists on the transfer of chemometric models based on Raman spectroscopy specifically. Hence, research activities should aim at developing the state-of-the art further and contribute to publications on this topic.

According to the typical use case for models described in the Model Transfer section, a model transfer will almost always be necessary prior to its application, which usually requires the primary dataset to be accessible for re-training. Therefore, the complete training data along with the pre-processing steps and fitting parameters should be stored with the model. In contrast, storage of only the (serialized) pre-fitted model object file will generally not permit model transfer. However, a pre-fitted, serialized model should also be stored to be routinely used once the transfer has been performed, because model transfer may be computationally expensive. It is therefore recommended that, wherever possible, data used in model building and transfer is maintained with all original information in a database to facilitate model reconstruction when required. Since Raman instrument calibrations are subject to steady drifts, the performance of chemometric models can be expected to deteriorate with time. Ryabchykov et al. advise instrument calibrations on a daily basis,<sup>12</sup> which, strictly speaking, would also

apply to model transfers. However, model transfer should not be performed with secondary data consisting of very few, or even a single, spectrum, since this may introduce a bias. Therefore, a compromise must be found between frequent model transfer (update) and completeness of secondary transfer data. In any case, model transfer must be re-done when predicting from data recorded with significant time difference, that is, days or weeks, to the secondary data used for transfer, or if stored models become obsolete due to changes in the programming language or the serialization libraries (such as Pickle for Python). It is recommended that serialization of models is avoided in the Raman data structure, but this is likely to be unavoidable for storage of fitted models, such as PCA classifiers or neural networks.

### Conclusion

The steps and considerations required to create, validate, and transfer a chemometric model using Raman spectra are outlined. With regard to harmonization, the pre-treatment step is of utmost importance as inaccurate calibration of the instrument removes any level of common ground to begin the harmonization. The systematic calibration of Raman instrumentation performed with a set of reference materials characterizes the instrument response and allows for digital calibration of subsequent acquired data. Second in significance is model validation and maintenance. The harmonization model needs to be long lived to make its adoption worthwhile. If laboratories need to adopt a new model regularly to maintain a harmonized standard, the benefit from being harmonized will not compensate its cost. However, pre-treatment and processing will also have a significant impact on the resulting model.

There already exists a large array of methods that can be used to harmonize Raman spectroscopic data, and more are being created by a very active chemometric community. Thus, it can be expected that a complete chemometric method tailored to build a robust and reliable system for Raman harmonization can be soon developed and should be adopted by the Raman community.

### Acknowledgments

The authors would like to thank Iván Moya (UNE) for the analysis of the ISO, IEC, CEN, and CENELEC standardization systems.

### Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the EU H2020 Project, "Characterisation and Harmonisation for Industrial Standardisation of Advanced Materials" (CHARISMA), under Grant Agreement no. 952921.

## ORCID iD

Raquel Portela  <https://orcid.org/0000-0002-1882-4759>

## Supplemental material

All supplemental material mentioned in the text is available in the online version of the journal.

## References

1. P. Mishra, D.N. Rutledge, J.M. Roger, K. Wali, H.A. Khan. "Chemometric Pre-Processing Can Negatively Affect the Performance of Near-Infrared Spectroscopy Models for Fruit Quality Prediction". *Talanta*. 2021. 229: 122303. doi:[10.1016/j.talanta.2021.122303](https://doi.org/10.1016/j.talanta.2021.122303).
2. Y. Roggo, K. Degardin, P. Margot. "Identification of Pharmaceutical Tablets by Raman Spectroscopy and Chemometrics". *Talanta*. 2010. 81(3): 988–995. doi:[10.1016/j.talanta.2010.01.046](https://doi.org/10.1016/j.talanta.2010.01.046).
3. N.M. Ralbovsky, I.K. Lednev. "Raman Spectroscopy and Chemometrics: A Potential Universal Method for Diagnosing Cancer". *Spectrochim. Acta, Part A*. 2019. 219: 463–487. doi:[10.1016/j.saa.2019.04.067](https://doi.org/10.1016/j.saa.2019.04.067).
4. T. Dörfer, T. Bocklitz, N. Tarcea, M. Schmitt, J. Popp. "Checking and Improving Calibration of Raman Spectra Using Chemometric Approaches". *Z. Phys. Chemie*. 2011. 225(6-7): 753–764. doi:[10.1524/zpch.2011.0077](https://doi.org/10.1524/zpch.2011.0077).
5. B.T. Bowie, D.B. Chase, P.R. Griffiths. "Factors Affecting the Performance of Bench-Top Raman Spectrometers. Part I: Instrumental Effects". *Appl. Spectrosc.* 2000. 54(5): 164A–173A. doi:[10.1366/0003702001949924](https://doi.org/10.1366/0003702001949924).
6. S. Barton. *Methods for Improving Signal to Noise Ratio in Raman Spectra*. [Doctor of Philosophy Dissertation]. Kildare, Ireland: Maynooth University, 2018.
7. D.A. Freedman, D.A. Freedman. "A Note on Screening Regression Equations". *Am. Stat.* 1983. 37(2): 152–155. doi:[10.1080/00031305.1983.10482729](https://doi.org/10.1080/00031305.1983.10482729).
8. K.P. Burnham, D.R. Anderson. *Model Selection and Multimodal Inference: Model Selection and Multimodal Inference*. New York: Springer, 2004. doi:[10.1007/b97636](https://doi.org/10.1007/b97636).
9. M. Féré, O. Piot, L.H. Liu, A. Beljebbar, et al. "Focus on Pre-Processing Step to Ensure the Clinical Transferability of Raman Data Acquired on Lymphocytes in Different Experimental and Instrumental Conditions". *Vib. Spectrosc.* 2019. 103: 102931. doi:[10.1016/j.vibspec.2019.102931](https://doi.org/10.1016/j.vibspec.2019.102931).
10. T. Bocklitz, A. Walter, K. Hartmann, P. Rösch, J. Popp. "How to Pre-Process Raman Spectra for Reliable and Stable Models?" *Anal. Chim. Acta*. 2011. 704(1-2): 47–56. doi:[10.1016/j.aca.2011.06.043](https://doi.org/10.1016/j.aca.2011.06.043).
11. X. Zhao, Z. Liu, Y. He, W. Zhang, L. Tong. "Study on Early Rice Blast Diagnosis Based on Unpre-Processed Raman Spectral Data". *Spectrochim. Acta, Part. A*. 2020. 234: 118255. doi:[10.1016/j.saa.2020.118255](https://doi.org/10.1016/j.saa.2020.118255).
12. O. Ryabchykov, S. Guo, T. Bocklitz. "Analyzing Raman Spectroscopic Data". *Phys. Sci. Rev.* 2019. 4(2): 1–16. doi:[10.1515/psr-2017-0043](https://doi.org/10.1515/psr-2017-0043).
13. H. Okajima, H.O. Hamaguchi. "Accurate Intensity Calibration for Low Wavenumber (–150 to 150 cm<sup>–1</sup>) Raman Spectroscopy Using the Pure Rotational Spectrum of N<sub>2</sub>". *J. Raman Spectrosc.* 2015. 46(11): 1140–1144. doi:[10.1002/jrs.4731](https://doi.org/10.1002/jrs.4731).
14. CHARISMA Consortium. "D3.1–List of Key Parameters and Samples". 2021.
15. National Institute of Standards and Technology (NIST). SRM 2242a. Relative Intensity Correction Standard for Raman Spectroscopy: 532 nm Excitation. Gaithersburg, Maryland: National Institute of Standards and Technology (NIST), 2019. <https://www-s.nist.gov/srmors/certificates/2242a.pdf> [accessed 15 October 2019].
16. J. Zhao, K. Frano, J. Zhou. "Reverse Intensity Correction for Raman Spectral Library Search". *Appl. Spectrosc.* 2017. 71(8): 1876–1883. doi:[10.1177/0003702817694381](https://doi.org/10.1177/0003702817694381).
17. ASTM International. ASTM E1655-17 Standard Practices for Infrared Multivariate Quantitative Analysis. West Conshohocken, Pennsylvania: ASTM International, 2017. doi:[10.1520/E1655-17](https://doi.org/10.1520/E1655-17).
18. W. Li, Y. Huang. "A Combined Method of Cross-Correlation and PCA-Based Outlier Algorithm for Detecting Structural Damages on a Jacket Oil Platform Under Random Wave Excitations". *Appl. Ocean Res.* 2020. 102: 102301. doi:[10.1016/j.apor.2020.102301](https://doi.org/10.1016/j.apor.2020.102301).
19. M. Baldomero-Naranjo, L.I. Martínez-Merino, A.M. Rodríguez-Chia. "A Robust SVM-Based Approach with Feature Selection and Outliers Detection for Classification Problems". *Expert Syst. Appl.* 2021. 178: 115017. doi:[10.1016/j.eswa.2021.115017](https://doi.org/10.1016/j.eswa.2021.115017).
20. A. Nowak-Brzezinska, C. Horyn. "Outliers in Rules: The Comparison of LOF, COF, and KMEANS Algorithms". *Procedia Comput. Sci.* 2020. 176: 1420–1429. doi:[10.1016/j.procs.2020.09.152](https://doi.org/10.1016/j.procs.2020.09.152).
21. M.M. Breunig, H.P. Kriegel, R.T. Ng, J. Sander. "LOF: Identifying Density-Based Local Outliers". *ACM SIGMOD Rec.* 2000. 29(2): 93–104. doi:[10.1145/335191.335388](https://doi.org/10.1145/335191.335388).
22. F. Keller, E. Muller, K. Bohm. "HiCS: High Contrast Subspaces for Density-Based Outlier Ranking". Paper presented at: IEEE 28th International Conference on Data Engineering. Arlington, Virginia; 1–5 April 2012. Pp. 1037–1048. doi:[10.1109/ICDE.2012.88](https://doi.org/10.1109/ICDE.2012.88).
23. J.M. Chalmers, P.R. Griffiths. *Handbook of Vibrational Spectroscopy*. Chichester, UK: John Wiley and Sons, Ltd., 2001. doi:[10.1002/0470027320](https://doi.org/10.1002/0470027320).
24. T. Dieing, O. Hollricher, J. Toporski. *Confocal Raman Microscopy*. Berlin; Heidelberg: Springer, 2011. doi:[10.1007/978-3-642-12522-5](https://doi.org/10.1007/978-3-642-12522-5).
25. S.J. Barton, T.E. Ward, B.M. Hennelly. "Algorithm for Optimal Denoising of Raman Spectra". *Anal. Methods*. 2018. 10(30): 3759–3769. doi:[10.1039/C8AY01089G](https://doi.org/10.1039/C8AY01089G).
26. D. Ozog. *Signal Processing*. Walla Walla, Washington: Whitman College, 2017. <https://www.whitman.edu/Documents/Academics/Mathematics/ozogdm.pdf> [accessed Mar 25 2022].
27. S. He, W. Zhang, L. Liu, Y. Huang, et al. "Baseline Correction for Raman Spectra Using an Improved Asymmetric Least Squares Method". *Anal. Methods*. 2014. 6(12): 4402–4407. doi:[10.1039/c4ay00068d](https://doi.org/10.1039/c4ay00068d).
28. P.J. Cadusch, M.M. Hlaing, S.A. Wade, S.L. McArthur, P.R. Stoddart. "Improved Methods for Fluorescence Background



- Subtraction from Raman Spectra". *J. Raman Spectrosc.* 2013. 44(11): 1587-1595. doi:[10.1002/jrs.4371](https://doi.org/10.1002/jrs.4371).
29. M. Caccia, A. Ebolese, M. Maspero, R. Santoro, et al. "Background Removal Procedure Based on the SNIP Algorithm for  $\gamma$ -Ray Spectroscopy with the CAEN Educational Kit". CAEN Tools Discovery Education Note ED3163. [https://www.caen.it/wp-content/uploads/2017/10/ED3163\\_gamma\\_spectroscopy\\_CAEN\\_edu\\_kit.pdf](https://www.caen.it/wp-content/uploads/2017/10/ED3163_gamma_spectroscopy_CAEN_edu_kit.pdf) [accessed Mar 25 2022].
30. J. Liu, J. Sun, X. Huang, G. Li, B.L. Goldindec: A Novel Algorithm for Raman Spectrum Baseline Correction". *Appl. Spectrosc.* 2015. 69(7): 834-842. doi:[10.1366/14-07798](https://doi.org/10.1366/14-07798).
31. J. José González-Vidal, R. Pérez-Pueyo, M.J. Soneira. "Automatic Morphology-Based Cubic P-Spline Fitting Methodology for Smoothing and Baseline-Removal of Raman Spectra". *J. Raman Spectrosc.* 2017. 48(6): 878-883. doi:[10.1002/jrs.5130](https://doi.org/10.1002/jrs.5130).
32. H. Hu, J. Bai, G. Xia, W. Zhang, Y. Ma. "Improved Baseline Correction Method Based on Polynomial Fitting for Raman Spectroscopy". *Photonic Sensors.* 2018. 8(4): 332-340. doi:[10.1007/s13320-018-0512-y](https://doi.org/10.1007/s13320-018-0512-y).
33. F. Qian, Y. Wu, P. Hao. "A Fully Automated Algorithm of Baseline Correction Based on Wavelet Feature Points and Segment Interpolation". *Opt. Laser Technol.* 2017. 96: 202-207. doi:[10.1016/j.optlastec.2017.05.021](https://doi.org/10.1016/j.optlastec.2017.05.021).
34. B.H. Kung, C.C. Huang, P.Y. Hu, S.Y. Lo, et al. "Baseline Correction and Denoising of Raman Spectra by Deep Residual CNN". Poster presented at: CLEO: Science and Innovations 2020. Washington, DC; 10-15 May 2020. doi:[10.1364/CLEO\\_AT.2020.JTu2G.28](https://doi.org/10.1364/CLEO_AT.2020.JTu2G.28).
35. M. Koch, C. Suhr, B. Roth, M. Meinhardt-Wollweber. "Iterative Morphological and Mollifier-Based Baseline Correction for Raman Spectra". *J. Raman Spectrosc.* 2016. 48(2): 336-342. doi:[10.1002/jrs.5010](https://doi.org/10.1002/jrs.5010).
36. S.R. Sternberg. "Biomedical Image Processing". Computer (Long Beach, Calif.). 1983. 16(1): 22-34. doi:[10.1109/MC.1983.1654163](https://doi.org/10.1109/MC.1983.1654163).
37. Y. Xu, P. Du, R. Senger, J. Robertson, J.L. Pirkle. "ISREA: An Efficient Peak-Preserving Baseline Correction Algorithm for Raman Spectra". *Appl. Spectrosc.* 2021. 75(1): 34-45. doi:[10.1177/0003702820955245](https://doi.org/10.1177/0003702820955245).
38. C. Beleites, V. Sergo. "Chemometric Analysis of Spectroscopic Data in R: HyperSpec". Poster presented at: 13th European Conference on the Spectroscopy of Biological Molecules (ECSBM). Palermo, Italy; 28 August-2 September 2009. <https://hyperspec.r-forge.r-project.org/blob/Beleites2009.pdf> [accessed Mar 25 2022].
39. G. Schulze, A. Jirasek, M.M.L. Yu, A. Lim, et al. "Investigation of Selected Baseline Removal Techniques as Candidates for Automated Implementation". *Appl. Spectrosc.* 2005. 59(5): 545-574. doi:[10.1366/0003702053945985](https://doi.org/10.1366/0003702053945985).
40. S. Guo, T. Bocklitz, J. Popp. "Optimization of Raman-Spectrum Baseline Correction in Biological Application". *Analyst.* 2016. 141(8): 2396-2404. doi:[10.1039/c6an00041j](https://doi.org/10.1039/c6an00041j).
41. R. Gautam, S. Vanga, F. Ariese, S. Umapathy. "Review of Multidimensional Data Processing Approaches for Raman and Infrared Spectroscopy". *EPJ Tech. Instrum.* 2015. 2(1). doi:[10.1140/epjti/s40485-015-0018-6](https://doi.org/10.1140/epjti/s40485-015-0018-6).
42. V. Giegold, L. Lange, R. Ciesielski, A. Hartschuh. "Non-Linear Raman Scattering Intensities in Graphene". *Nanoscale.* 2020. 12(9): 5612-5617. doi:[10.1039/C9NR10654E](https://doi.org/10.1039/C9NR10654E).
43. A. Tulsyan, G. Schorner, H. Khodabandehlou, T. Wang, et al. "A Machine-Learning Approach to Calibrate Generic Raman Models for Real-Time Monitoring of Cell Culture Processes". *Biotechnol. Bioeng.* 2019. 116(10): 2575-2586. doi:[10.1002/bit.27100](https://doi.org/10.1002/bit.27100).
44. J.C. Ang, A. Mirzal, H. Haron, H.N.A. Hamed. "Supervised, Unsupervised, and Semi-Supervised Feature Selection: A Review on Gene Selection". *IEEE/ACM Trans. Comput. Biol. Bioinf.* 2016. 13(5): 971-989. doi:[10.1109/TCBB.2015.2478454](https://doi.org/10.1109/TCBB.2015.2478454).
45. C.M. Gryniwicz-Ruzicka, J.D. Rodriguez, S. Arzhantsev, L.F. Buhse, J.F. Kauffman. "Libraries, Classifiers, and Quantifiers: A Comparison of Chemometric Methods for the Analysis of Raman Spectra of Contaminated Pharmaceutical Materials". *J. Pharm. Biomed. Anal.* 2012. 61: 191-198. doi:[10.1016/j.jpba.2011.12.002](https://doi.org/10.1016/j.jpba.2011.12.002).
46. R.G. Brereton. *Chemometrics for Pattern Recognition*. Chichester, UK: John Wiley and Sons, 2009. doi:[10.1002/9780470746462](https://doi.org/10.1002/9780470746462).
47. J.D. Rodriguez, B.J. Westenberg, L.F. Buhse, J.F. Kauffman. "Quantitative Evaluation of the Sensitivity of Library-Based Raman Spectral Correlation Methods". *Anal. Chem.* 2011. 83(11): 4061-4067. doi:[10.1021/ac200040b](https://doi.org/10.1021/ac200040b).
48. Scikit-Learn. "Machine Learning in Python: Scikit-Learn". <https://scikit-learn.org/stable/> [accessed Apr 14 2021].
49. M. Hedegaard, C. Matthäus, S. Hassing, C. Krafft, et al. "Spectral Unmixing and Clustering Algorithms for Assessment of Single Cells by Raman Microscopic Imaging". *Theor. Chem. Acc.* 2011. 130: 1249-1260. doi:[10.1007/s00214-011-0957-1](https://doi.org/10.1007/s00214-011-0957-1).
50. J.H. Wang, P.K. Hopke, T.M. Hanczewicz, S.L. Zhang. "Application of Modified Alternating Least Squares Regression to Spectroscopic Image Analysis". *Anal. Chim. Acta.* 2003. 476(1): 93-109. doi:[10.1016/S0003-2670\(02\)01369-7](https://doi.org/10.1016/S0003-2670(02)01369-7).
51. T. Bocklitz, M. Putsche, C. Stüber, J. Käs, et al. "A Comprehensive Study of Classification Methods for Medical Diagnosis". *J. Raman Spectrosc.* 2009. 40(12): 1759-1765. doi:[10.1002/jrs.2529](https://doi.org/10.1002/jrs.2529).
52. J. Acquarelli, T. van Laarhoven, J. Gerretzen, T.N. Tran, et al. "Convolutional Neural Networks for Vibrational Spectroscopic Data Analysis". *Anal. Chim. Acta.* 2017. 954: 22-31. doi:[10.1016/j.aca.2016.12.010](https://doi.org/10.1016/j.aca.2016.12.010).
53. L. Breiman. "Random Forests". *Mach. Learn.* 2001. 45(1): 5-32. doi:[10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324).
54. J. Hutchings, C. Kendall, N. Shepherd, H. Barr, N. Stone. "Evaluation of Linear Discriminant Analysis for Automated Raman Histological Mapping of Esophageal High-Grade Dysplasia". *J. Biomed. Opt.* 2010. 15(6): 66015. doi:[10.1117/1.3512244](https://doi.org/10.1117/1.3512244).
55. J. Demšar, A. Erjavec, T. Hočevár, M. Milutinovič, et al. "Orange: Data Mining Toolbox in Python". *J. Machine Learning Res.* 2013. 14(1): 2349-2353.
56. K. Rammelkamp, S. Schröder, S. Kubitz, D.S. Vogt, et al. "Low-Level LIBS and Raman Data Fusion in the Context of In Situ Mars

- Exploration". *J. Raman Spectrosc.* 2020. 51(9): 1682-1701. doi: [10.1002/jrs.5615](https://doi.org/10.1002/jrs.5615).
57. C. Márquez, M.I. López, I. Ruisánchez, M.P. Callao. "FT-Raman and NIR Spectroscopy Data Fusion Strategy for Multivariate Qualitative Analysis of Food Fraud". *Talanta*. 2016. 161: 80-86. doi: [10.1016/j.talanta.2016.08.003](https://doi.org/10.1016/j.talanta.2016.08.003).
  58. S. Guo, C. Beleites, U. Neugebauer, S. Abalde-Cela, et al. "Comparability of Raman Spectroscopic Configurations: A Large-Scale Cross-Laboratory Study". *Anal. Chem.* 2020. 92(24): 15745-15756. doi: [10.1021/acs.analchem.0c02696](https://doi.org/10.1021/acs.analchem.0c02696).
  59. E. Borràs, J. Ferré, R. Boqué, M. Mestres, et al. "Data Fusion Methodologies for Food and Beverage Authentication and Quality Assessment: A Review". *Anal. Chim. Acta*. 2015. 891: 1-14. doi: [10.1016/j.aca.2015.04.042](https://doi.org/10.1016/j.aca.2015.04.042).
  60. X. Yuan, R.A. Mayanovic. "An Empirical Study on Raman Peak Fitting and its Application to Raman Quantitative Research". *Appl. Spectrosc.* 2017. 71(10): 2325-2338. doi: [10.1177/0003702817721527](https://doi.org/10.1177/0003702817721527).
  61. T. Sundius. "Computer Fitting of Voigt Profiles to Raman Lines". *J. Raman Spectrosc.* 1973. 1(5): 471-488. doi: [10.1002/jrs.1250010506](https://doi.org/10.1002/jrs.1250010506).
  62. V.I. Korepanov, D.M. Sedlovets. "An Asymmetric Fitting Function for Condensed-Phase Raman Spectroscopy". *Analyst*. 2018. 143(11): 2674-2679. doi: [10.1039/c8an00710a](https://doi.org/10.1039/c8an00710a).
  63. Scikit-Learn. "Sklearn.Decomposition.PCA: Scikit-Learn 0.24.1 Documentation". <https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html> [accessed Mar 25 2022].
  64. Scikit-Learn. "Sklearn.Decomposition.FastICA: Scikit-Learn 0.24.1 Documentation". <https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.FastICA.html> [accessed Mar 25 2022].
  65. E. Malinowski, D. Howery. *Factor Analysis in Chemistry*. New York: Wiley-Interscience, 1980.
  66. T. Azzouz, R. Tauler. "Application of Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) to the Quantitative Analysis of Pharmaceutical and Agricultural Samples". *Talanta*. 2008. 74: 1201-1210. doi: [10.1016/j.talanta.2007.08.024](https://doi.org/10.1016/j.talanta.2007.08.024).
  67. B.K. Lavine, C.E. Davidson, A.J. Moores, P.R. Griffiths. "Raman Spectroscopy and Genetic Algorithms for the Classification of Wood Types". *Appl. Spectrosc.* 2001. 55(8): 960-966. doi: [10.1366/0003702011953108](https://doi.org/10.1366/0003702011953108).
  68. B. Barton, M. Zou, G. Geertz, E. Metsch-Zillingen, R. Brüll. "Raman Imaging of Additives in Polymers: Multivariate Decomposition of 3D Maps and Super-Resolution Imaging". Presented at: International Conference for Chemical Characterization and Imaging, Ulm, Germany; 23-25 September 2019.
  69. B.S. Everitt, A. Skrondal. *The Cambridge Dictionary of Statistics*. Cambridge; New York: Cambridge University Press, 2010.
  70. Python. "Classes: Python 3.9.12 Documentation". <https://docs.python.org/3.9/tutorial/classes.html> [accessed Mar 25 2022].
  71. J. Demšar, T. Curk, A. Erjavec, Č. Gorup, et al. "Orange: Data Mining Toolbox in Python". *J. Mach. Learn. Res.* 2013. 14: 2349-2353.
  72. A. Celisse. "Optimal Cross-Validation in Density Estimation with the L2-Loss". *Ann. Stat.* 2014. 42(5): 1879-1910. doi: [10.1214/14-AOS1240](https://doi.org/10.1214/14-AOS1240).
  73. T.G. Tape. "Interpreting Diagnostic Tests: The Area Under an ROC Curve". <http://gim.unmc.edu/dxtests/roc3.htm> [accessed Apr 13 2021].
  74. J.R. Taylor. *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*. Sausalito, California: University Science Books, 1999.
  75. D.M.W. Powers. "What the F-Measure Doesn't Measure: Features, Flaws, Fallacies and Fixes". ArXiv. 2015. <https://arxiv.org/ftp/arxiv/papers/1503/1503.06410.pdf> [accessed Mar 25 2022].
  76. T. Fawcett. "An Introduction to ROC Analysis". *Pattern Recognit. Lett.* 2006. 27(8): 861-874. doi: [10.1016/j.patrec.2005.10.010](https://doi.org/10.1016/j.patrec.2005.10.010).
  77. K. Héberger. "Sum of Ranking Differences Compares Methods or Models Fairly". *TrAC, Trends Anal. Chem.* 2010. 29(1): 101-109. doi: [10.1016/j.trac.2009.09.009](https://doi.org/10.1016/j.trac.2009.09.009).
  78. P. Shahbazikhah, J.H. Kalivas. "A Consensus Modeling Approach to Update a Spectroscopic Calibration". *Chemom. Intell. Lab. Syst.* 2013. 120: 142-153. doi: [10.1016/j.chemolab.2012.06.006](https://doi.org/10.1016/j.chemolab.2012.06.006).
  79. J.H. Kalivas, G.G. Siano, E. Andries, H.C. Goicoechea. "Calibration Maintenance and Transfer Using Tikhonov Regularization Approaches". *Appl. Spectrosc.* 2009. 63(7): 800-809. doi: [10.1366/000370209788701206](https://doi.org/10.1366/000370209788701206).
  80. Q. Li, X. Sun, X. Ma, B. Li, et al. "A Calibration Transfer Methodology for Standardization of Raman Instruments with Different Spectral Resolutions Using Double Digital Projection Slit". *Chemom. Intell. Lab. Syst.* 2019. 191(April): 143-147. doi: [10.1016/j.chemolab.2019.07.004](https://doi.org/10.1016/j.chemolab.2019.07.004).
  81. S. Guo, R. Heinke, S. Stöckel, P. Rösch, et al. "Model Transfer for Raman-Spectroscopy-Based Bacterial Classification". *J. Raman Spectrosc.* 2018. 49(4): 627-637. doi: [10.1002/jrs.5343](https://doi.org/10.1002/jrs.5343).
  82. S. Guo, R. Heinke, S. Stöckel, P. Rösch, et al. "Towards an Improvement of Model Transferability for Raman Spectroscopy in Biological Applications". *Vib. Spectrosc.* 2017. 91: 111-118. doi: [10.1016/j.vibspec.2016.06.010](https://doi.org/10.1016/j.vibspec.2016.06.010).
  83. D. Brouckaert, J.S. Uytersprot, W. Broeckx, T. De Beer. "Calibration Transfer of a Raman Spectroscopic Quantification Method from At-Line to In-Line Assessment of Liquid Detergent Compositions". *Anal. Chim. Acta*. 2017. 971: 14-25. doi: [10.1016/j.aca.2017.03.049](https://doi.org/10.1016/j.aca.2017.03.049).
  84. M. Kompany-Zareh, F. Van Den Berg. "Multi-Way Based Calibration Transfer between Two Raman Spectrometers". *Analyst*. 2010. 135(6): 1382-1388. doi: [10.1039/b927501k](https://doi.org/10.1039/b927501k).
  85. D. Brouckaert, J.S. Uytersprot, W. Broeckx, T. De Beer. "Calibration Transfer of a Raman Spectroscopic Quantification Method for the Assessment of Liquid Detergent Compositions between Two At-Line Instruments Installed at Two Liquid Detergent Production Plants". *Anal. Chim. Acta*. 2017. 984: 1-18. doi: [10.1016/j.aca.2017.07.044](https://doi.org/10.1016/j.aca.2017.07.044).
  86. Y. Wang, D.J. Veltkamp, B.R. Kowalski. "Multivariate Instrument Standardization". *Anal. Chem.* 1991. 63(23): 2750-2756. doi: [10.1021/ac00023a016](https://doi.org/10.1021/ac00023a016).

87. E. Bouveresse, D.L. Massart. "Improvement of the Piecewise Direct Standardisation Procedure for the Transfer of NIR Spectra for Multivariate Calibration". *Chemom. Intell. Lab. Syst.* 1996. 32(2): 201-213. doi:[10.1016/0169-7439\(95\)00074-7](https://doi.org/10.1016/0169-7439(95)00074-7).
88. H. Leion, S. Folestad, M. Josefson, A. Sparén. "Evaluation of Basic Algorithms for Transferring Quantitative Multivariate Calibrations between Scanning Grating and FT NIR Spectrometers". *J. Pharm. Biomed. Anal.* 2005. 37(1): 47-55. doi:[10.1016/j.jpba.2004.09.046](https://doi.org/10.1016/j.jpba.2004.09.046).
89. S. Guo, A. Kohler, B. Zimmermann, R. Heinke, et al. "Extended Multiplicative Signal Correction Based Model Transfer for Raman Spectroscopy in Biological Applications". *Anal. Chem.* 2018. 90(16): 9787-9795. doi:[10.1021/acs.analchem.8b01536](https://doi.org/10.1021/acs.analchem.8b01536).
90. C.E. Anderson, J.H. Kalivas. "Fundamentals of Calibration Transfer through Procrustes Analysis". *Appl. Spectrosc.* 1999. 53(10): 1268-1276. doi: [10.1366/0003702991945515](https://doi.org/10.1366/0003702991945515).
91. J. Sjöblom, O. Svensson, M. Josefson, H. Kullberg, S. Wold. "An Evaluation of Orthogonal Signal Correction Applied to Calibration Transfer of Near Infrared Spectra". *Chemom. Intell. Lab. Syst.* 1998. 44(1-2): 229-244. doi:[10.1016/S0169-7439\(98\)00112-9](https://doi.org/10.1016/S0169-7439(98)00112-9).
92. P.H.C. Eilers. "Parametric Time Warping". *Anal. Chem.* 2004. 76(2): 404-411. doi:[10.1021/ac034800e](https://doi.org/10.1021/ac034800e).
93. National Institute of Standards and Technology (NIST). SRM 2241 Relative Intensity Correction Standard for Raman Spectroscopy: 785 nm Excitation. Gaithersburg, Maryland: National Institute of Standards and Technology (NIST), 2022. <https://www-s.nist.gov/m-srmors/certificates/2241.pdf> [accessed March 11 2022].
94. J.D. Rodriguez, B.J. Westenberger, L.F. Buhse, J.F. Kauffman. "Standardization of Raman Spectra for Transfer of Spectral Libraries Across Different Instruments". *Analyst.* 2011. 136(20): 4232. doi:[10.1039/c1an15636e](https://doi.org/10.1039/c1an15636e).
95. SciPy.org. "SciPy.Optimize.Anneal: SciPy v0.14.0 Reference Guide". <https://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.optimize.anneal.html> [accessed Apr 9 2021].
96. SciPy.org. "SciPy.Optimize.Fmin: SciPy V1.6.2 Reference Guide". <https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.fmin.html> [accessed Apr 9 2021].
97. CHARISMA Consortium. DELIVERABLE D4.1. 2021.
98. M. Bevilacqua, F. Marini. "Local Classification: Locally Weighted-Partial Least Squares-Discriminant Analysis (LW-PLS-DA)". *Anal. Chim. Acta.* 2014. 838: 20-30. doi:[10.1016/j.aca.2014.05.057](https://doi.org/10.1016/j.aca.2014.05.057).
99. A. Ntziouni, J. Thomson, I. Xiarchos, X. Li, et al. "Review on Existing Standards, Guides and Practices for Raman Spectroscopy". *Appl. Spectrosc.* doi:[10.1177/00037028221090988](https://doi.org/10.1177/00037028221090988).
100. ASTM International, E13 Committee. "ASTM E1654-94 Standard Guide for Measuring Ionizing Radiation-Induced Spectral Changes in Optical Fibers and Cables for Use in Remote Raman Fiberoptic Spectroscopy". doi:[10.1520/E1654-94R13.2](https://doi.org/10.1520/E1654-94R13.2).
101. ASTM International. ASTM E1683-02 Standard Practice for Testing the Performance of Scanning Raman Spectrometers. West Conshohocken, Pennsylvania: ASTM International, 2014. doi:[10.1520/E1683-02R14E01.2](https://doi.org/10.1520/E1683-02R14E01.2).
102. ASTM International. ASTM E1840-96 Standard Guide for Raman Shift Standards for Spectrometer Calibration. West Conshohocken, Pennsylvania: ASTM International, 2014.
103. ASTM International. ASTM E2529-06 Standard Guide for Testing the Resolution of a Raman Spectrometer. West Conshohocken, Pennsylvania: ASTM International, 2014.
104. ASTM International. ASTM E2911-13 Standard Guide for Relative Intensity Correction of Raman Spectrometers. West Conshohocken, Pennsylvania: ASTM International, 2013.
105. ASTM International. ASTM D6122-20a Standard Practice for Validation of the Performance of Multivariate Online, At-Line, Field, and Laboratory Infrared Spectrophotometer, and Raman Spectrometer Based Analyzer Systems. West Conshohocken, Pennsylvania: ASTM International, 2021. doi:[10.1520/D6122-21](https://doi.org/10.1520/D6122-21).
106. ASTM International. ASTM E2056-04 Standard Practice for Qualifying Spectrometers and Spectrophotometers for Use in Multivariate Analyses, Calibrated Using Surrogate Mixtures. West Conshohocken, Pennsylvania: ASTM International, 2016. doi:[10.1520/E2056-04R16](https://doi.org/10.1520/E2056-04R16).
107. ASTM International. ASTM E2617-17 Standard Practice for Validation of Empirically Derived Multivariate Calibrations. West Conshohocken, Pennsylvania: ASTM International, 2017.
108. ASTM International. ASTM E2891-20 Standard Guide for Multivariate Data Analysis in Pharmaceutical Development and Manufacturing Applications. West Conshohocken, Pennsylvania: ASTM International, 2020. doi:[10.1520/E2891-20](https://doi.org/10.1520/E2891-20).
109. United States Pharmacopeial Convention. <1858> Raman Spectroscopy: Theory and Practice. Rockville, Maryland: United States Pharmacopeia, 2020. doi:[10.31003/USPNF\\_M8188\\_02\\_01](https://doi.org/10.31003/USPNF_M8188_02_01).
110. United States Pharmacopeial Convention. <858> Raman Spectroscopy. Rockville, Maryland: United States Pharmacopeia, 2020. doi:[10.31003/USPNF\\_M8188\\_02\\_01](https://doi.org/10.31003/USPNF_M8188_02_01).
111. European Directorate for the Quality of Medicines and Healthcare. "Chemometric Methods Applied to Analytical Data". *European Pharmacopoeia*. Strasbourg: Council of Europe, 2020. 10th ed. Chap. 5.21.
112. United States Pharmacopeial Convention. <1039> Chemometrics. Rockville, Maryland: United States Pharmacopeia, 2020.
113. ASTM International. ASTM E456-02 Standard Terminology Relating to Quality and Statistics. West Conshohocken, Pennsylvania: ASTM International, 2002.
114. ASTM International. ASTM E178-21 Standard Practice for Dealing with Outlying Observations. West Conshohocken, Pennsylvania: ASTM International, 2021.
115. J.D. Rodriguez, B.J. Westenberger, L.F. Buhse, J.F. Kauffman. "Standardization of Raman Spectra for Transfer of Spectral Libraries Across Different Instruments". *Analyst.* 2011. 136(20): 4232-4240. doi:[10.1039/c1an15636e](https://doi.org/10.1039/c1an15636e).
116. D. Hutsebaut, P. Vandenabeele, L. Moens. "Evaluation of an Accurate Calibration and Spectral Standardization Procedure for Raman Spectroscopy". *Analyst.* 2005. 130(8): 1204-1214. doi:[10.1039/b503624k](https://doi.org/10.1039/b503624k).