Chemometrics and Intelligent Laboratory Systems xxx (2012) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Extended multiplicative signal correction in vibrational spectroscopy, a tutorial

Nils Kristian Afseth [a,*], Achim Kohler [a,b]

[a] Nofima – Norwegian Institute of Food, Fisheries and Aquaculture Research, Osloveien 1, 1430 Ås, Norway
[b] CIGENE, Department of Mathematical Sciences and Technology, Norwegian University of Life Sciences, PO Box 5003, 1432 Ås, Norway

ARTICLE INFO

Article history:
Received 30 September 2011
Received in revised form 13 January 2012
Accepted 6 March 2012
Available online xxxx

Keywords:
Extended multiplicative signal correction
Preprocessing
Vibrational spectroscopy
Polynomial extensions
Chemical interference
Physical interference

ABSTRACT

Extended multiplicative signal correction (EMSC) is a model-based preprocessing technique that has been frequently used in vibrational spectroscopy. The framework of EMSC allows, apart from basic preprocessing as baseline correction and normalization, to separate and quantify different types of chemical and physical variations in vibrational spectra. EMSC thus enables the user to study these different effects separately. The aim of this tutorial is to give a comprehensive description of EMSC and its use in vibrational spectroscopy. Examples from FT-IR and Raman spectroscopy of biological materials will be used, and different extensions of EMSC and their properties will be discussed.

1. Introduction

In vibrational spectroscopy, the spectra are often affected by numerous phenomena other than the chemical components of interest, and these phenomena can create challenges in the subsequent qualitative and quantitative analyses. Depending on the kind of analysis, the phenomena can range from random measurement noise to systematic errors related to for instance non-linear instrument responses and interfering effects from undesired chemical and physical variations. Besides minimizing these challenges by sample preparation and analytical protocols, preprocessing methods are vital in order to correct for many of the unwanted phenomena.

Preprocessing methods in vibrational spectroscopy may be divided into two categories, namely filtering methods and model-based methods. The 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. Model-based methods, on the other hand, allow for quantifying and separating between the different types of physical and chemical variations in the spectra. Hence, the information filtered is not lost, but statistical estimates of the mathematical parameters involved in the filtering are obtained and can be used to study the different effects separately.

Multiplicative signal correction (MSC) is the starting point of all *model-based preprocessing* techniques that are used today in the field of vibrational spectroscopy. It was developed in the 1980s for applications in near-infrared (NIR) spectroscopy in food science [1–3]. MSC was originally used as an abbreviation for multiplicative *scatter* correction, but as one soon realized that the approach could serve as a generic tool for signal processing, the term multiplicative *signal* correction was introduced, and this is the term generally used today. In 1991, the basic idea behind MSC was extended (EMSC for extended MSC), and subsequently this approach has been extensively used in the field of NIR spectroscopy [4–7]. In 2003, an inverse version of EMSC, namely EISC or extended inverted signal correction, was introduced for the correction of NIR spectra [5]. EISC has found various applications in vibrational spectroscopy, but this technique will not be elaborated in this tutorial [8–10].

Within the last few years, EMSC has especially attracted attention in the field of FT-IR spectroscopy, not least due to its flexibility that allows more selective correction for various types of scattering and other unwanted variation effects than what is feasible using standard filtering techniques. EMSC has been shown to be a reliable tool to correct for additive baseline effects, multiplicative scaling effects, and interference effects. It has been applied to the correction of for instance sample thickness, water vapor, carbon dioxide, sample temperature and salt concentrations [11–14]. A method has also been presented based on EMSC for estimating and correcting the contribution of Mie scatter effects in FT-IR microspectroscopy of cells [15–17]. The basic EMSC algorithm has also been applied to Raman spectra by several researchers within the last few years [18–21], and recently EMSC has also been adapted to specific effects in Raman spectroscopy, such as to the correction of fluorescence effects [12,22].

The aim of this tutorial is to give a comprehensive description of EMSC and its use in vibrational spectroscopy, with a strong emphasis on extensions feasible for correction of specific effects frequently

* Corresponding author. Tel.: +47 64970100; fax: +47 64970333.
E-mail address: nils.kristian.afseth@nofima.no (N.K. Afseth).

0169-7439/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.chemolab.2012.03.004

encountered in FT-IR and Raman spectra. Examples from FT-IR and Raman spectroscopy of biological materials will be used, and different versions of EMSC and their properties will be discussed. Along with this tutorial, guidelines in the practical use of EMSC employing an in-house made Matlab code will be provided. This Matlab code is free of charge and can be downloaded together with a tutorial of its use at the website www.specmod.org. The basic principle of EMSC is patented [23]. Therefore the free algorithms on our homepage cannot be used commercially without any further agreements.

2. Deriving the EMSC model

Quantitative measurements in vibrational absorbance spectroscopy are based on the Lambert–Beer law, which also marks a convenient starting point for deriving the basic EMSC model. According to the Lambert–Beer law, an absorbance spectrum is directly proportional to the effective optical path length. For transparent samples containing a single light-absorbing chemical species, the absorbance $A(\tilde{\nu})$ is given by:

$$A(\tilde{\nu}) = k(\tilde{\nu}) \cdot c \cdot b \tag{1}$$

where $k(\tilde{\nu})$ is the characteristic *absorptivity* for a specific component at a specific wavenumber $\tilde{\nu}$, b is the *optical path length*, and c is the *concentration* of the light-absorbing chemical species in the sample. Spectral variations that are caused by variations in the optical path length are usually denoted 'multiplicative' variations.

Since biological samples are very complex we encounter many different species $k(\tilde{\nu})$, and many different biomolecule component spectra $k_j(\tilde{\nu})$ (where j denotes the different biomolecules from $j=1\ldots J$) have strongly overlapping characteristics. For J absorbing species, the Lambert–Beer law can therefore be written as a superposition of absorbances of several species:

$$A(\tilde{\nu}) = \left(\sum_{j=1}^{J} c_j \cdot k_j(\tilde{\nu}) \right) \cdot b \tag{2}$$

where $k_j(\tilde{\nu})$ is the spectrum and c_j the concentration of the constituent j. We assume here that the optical path length b is comparable for all constituents. This assumption may in general hold for sufficiently homogenous samples.

The definition of constituent spectra in Eq. (2) for biological samples is in general only of theoretical interest, since the biochemical constituent spectra are often not known and often even difficult to obtain. However, in general the overall shapes of infrared absorbance spectra obtained from biological samples are very similar. This means that the average spectrum of a sample set is a very good approximation for each spectrum. A measured absorbance spectrum $A(\tilde{\nu})$ may thus be expressed as the mean $\bar{x}(\tilde{\nu})$ of all spectra in a given data set plus deviations $\Delta k_j(\tilde{\nu})$ from this mean:

$$k_j(\tilde{\nu}) = \bar{x}(\tilde{\nu}) + \Delta k_j(\tilde{\nu}). \tag{3}$$

Inserting Eq. (3) into Eq. (2) we obtain:

$$A(\tilde{\nu}) = \left(\sum_{j=1}^{J} c_j \cdot k_j(\tilde{\nu}) \right) \cdot b = \left(\sum_{j=1}^{J} c_j \cdot \bar{x}(\tilde{\nu}) + \sum_{j=1}^{J} c_j \cdot \Delta k_j(\tilde{\nu}) \right) \cdot b. \tag{4}$$

In order to bring the model in Eq. (4) to its final form, we require that the pure biochemical spectrum $z(\tilde{\nu})$, which we are about to estimate, can be represented as the sum of all theoretically available constituent spectra:

$$z(\tilde{\nu}) = \sum_{j=1}^{J} c_j \cdot k_j(\tilde{\nu}) \tag{5}$$

where the sum of concentrations sums up to 1 according to:

$$\sum_{j=1}^{J} c_j = 1. \tag{6}$$

This condition is meaningful, since it simply means that the constituent spectra are normalized such that, if multiplied with the respective concentrations, they sum up to 100% in each measured spectrum. Inserting the condition in Eq. (6) into Eq. (4), we finally obtain:

$$A(\tilde{\nu}) = \left(\bar{x}(\tilde{\nu}) + \sum_{j=1}^{J} c_j \cdot \Delta k_j(\tilde{\nu}) \right) \cdot b. \tag{7}$$

In all cases where the overall shape of the measured spectrum $A(\tilde{\nu})$ is close to the average spectrum $\bar{x}$, we can replace Eq. (7) by the statistical model:

$$A(\tilde{\nu}) = \bar{x}(\tilde{\nu}) \cdot b + e(\tilde{\nu}) \tag{8}$$

where the residual $e(\tilde{\nu})$ is defined as:

$$e(\tilde{\nu}) = b \cdot \sum_{j=1}^{J} c_j \cdot \Delta k_j(\tilde{\nu}). \tag{9}$$

The multiplicative signal correction (MSC) model is an extension of the Lambert–Beer-type model shown in Eq. (7) by an additive effect given by the constant baseline a. The basic MSC model thus writes:

$$A(\tilde{\nu}) = a + \bar{x}(\tilde{\nu}) \cdot b + e(\tilde{\nu}). \tag{10}$$

The unknown parameters in Eq. (10) are estimated by ordinary or weighted least squares regression. When the parameters a and b are estimated, spectra are corrected according to:

$$A_{corr}(\tilde{\nu}) = (A(\tilde{\nu}) - a)/b. \tag{11}$$

We will in the following subsequently extend the model in Eq. (10) to situations where Eq. (10) is oversimplified and an approximation of each spectrum by its average spectrum and a constant baseline is not sufficient. It is interesting to note that in Eq. (8) the interesting part, namely the biochemical differences among the spectra, is expressed by the residual. This shows the philosophy of the standardization procedure by MSC and later by its extensions: by the estimation of the parameters in the model in Eq. (10) and the subsequent correction by these parameters we make all spectra as close to the average or a reference as possible. The part of the spectra that is not taken into account by the model is assigned to the term $e(\tilde{\nu})$ in Eq. (10). Thus, the residual $e(\tilde{\nu})$ contains the interesting chemical differences between the samples. As mentioned before, in absorption spectroscopy the multiplicative effect is proportional to the effective optical path length. For scattering techniques like Raman spectroscopy, on the other hand, multiplicative effects are related to e.g. variations in the sampling volume, intensities of the excitation laser and even the positioning or focusing of the sample.

In many situations, as for example in Raman spectroscopy, baseline variations are present in the spectra that cannot be represented by a straight line. When for example fluorescence phenomena are present in Raman spectra, the MSC model in Eq. (10) can be extended by a baseline with an arbitrary slope, a quadratic term or terms of higher polynomial order, leading to basic extensions of MSC, the so-called extended multiplicative signal correction:

$$A(\tilde{\nu}) = a + \bar{x}(\tilde{\nu}) \cdot b + d_1\tilde{\nu} + d_2\tilde{\nu}^2 + \cdots + d_n\tilde{\nu}^n + e(\tilde{\nu}). \tag{12}$$

Please cite this article as: N.K. Afseth, A. Kohler, Extended multiplicative signal correction in vibrational spectroscopy, a tutorial, Chemometr. Intell. Lab. Syst. (2012), doi:10.1016/j.chemolab.2012.03.004

As for MSC, the unknown parameters are estimated using an ordinary or weighted least squares estimation, and the spectra are corrected according to:

$$A_{corr}(\tilde{\nu}) = \frac{A(\tilde{\nu}) - a - d_1\tilde{\nu} - d_2\tilde{\nu}^2 - \cdots - d_n\tilde{\nu}^n}{b}. \quad (13)$$

The EMSC model where the polynomial in Eq. (12) is extended up to the quadratic term is often referred to as the *basic EMSC model*.

3. Extending the basic EMSC model with polynomial extensions

In this section we will illustrate the use of a basic EMSC model and subsequently the polynomial extensions of an EMSC model on an example taken from Raman spectroscopy of salmon oils [24]. The data set consists of 3 repeated measurements of in total 45 salmon oil samples. Raw spectra are shown in Fig. 1a, whereas mean-centered raw spectra are displayed in Fig. 1d. In both figures dominant curved baseline variations in the Raman spectra are visible. These features arise from laser-induced fluorescence, most likely due to lipid oxidation products in the salmon oils. In addition, multiplicative features are observable, especially in the high Raman shift region of the spectra. In Fig. 1b Raman spectra that were corrected by the basic EMSC model, are shown. The corresponding mean-centered spectra are shown in Fig. 1e.

Fig. 1b and e shows that a satisfactory correction of the Raman spectra can be obtained by the basic EMSC model. The baseline fluctuations have been significantly reduced, and the multiplicative effects as seen in the high Raman shift region of the spectra have been minimized. It is important to note that along with the corrected spectra, estimates for the parameters a, b, d_1 and d_2 according to Eq. (12) are available for all spectra. The use of information contained in these parameters will be the subject of later discussions.

Fig. 1 also reveals that even though the baseline fluctuations of the Raman spectra have been reduced due to EMSC correction, each spectrum still show a significant baseline. Since we used the mean spectrum as reference spectrum in Eq. (12), all corrected spectra will end up with approximately the same baseline as the average. This will, however, have no effect on a subsequent multivariate data analysis such as principal component analysis or partial least squares regression. Therefore the use of a 'baseline-free' reference spectrum is simply a question of esthetics; it does not affect the qualitative and quantitative information contained in the data set. In Fig. 1c and f basic EMSC corrected spectra and corresponding mean-centered spectra are shown using a reference spectrum with a reduced baseline slope. We observe that the corrected spectra also exhibit substantially reduced baseline slopes, but the corresponding mean-centered spectra in Fig. 1f are identical with the ones obtained when the average spectrum was used as reference spectrum as shown in Fig. 1e.

Even though most of the baseline variations could be removed by the correction of a basic EMSC model, mean-centered spectra in Fig. 1e and f still show broad baseline variation structures differing clearly from the sharp chemical Raman peaks that still remain in the spectra. Apparently, the basic EMSC model including linear and quadratic effects is not sufficient to model baseline variations in Raman spectra, which are due to fluorescence phenomena. In Fig. 2a and b results are shown for the correction by a sixth order polynomial EMSC model. Corrected spectra are shown in Fig. 2a while mean-centered corrected spectra are shown in Fig. 2b. The results clearly demonstrate that the polynomial extension of the EMSC model improves the correction, although there are still some minor un-modeled baseline variations visible in the mean-centered spectra. Usually in baseline correction of Raman spectra, polynomials up to the sixth or seventh order are found sufficient for estimating the present baseline slopes [18,25]. The polynomial extension of the EMSC model up to e.g., the seventh order will not pose any over-fitting problems in the

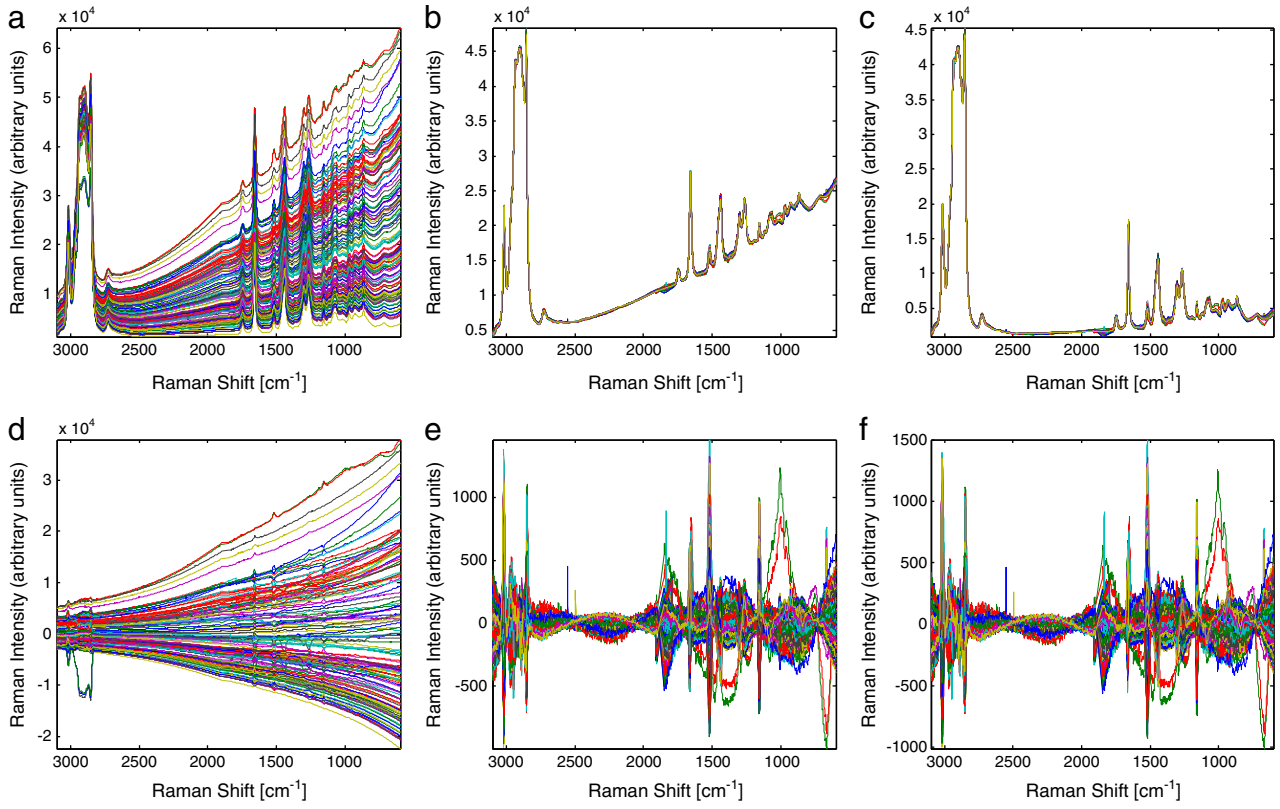

Fig. 1. Raw Raman spectra of salmon oils (a), and corresponding spectra subjected to basic EMSC (b) and basic EMSC using a reference spectrum with a reduced baseline slope (c). Corresponding mean-centered data sets are provided in the lower rows (d–f), respectively.

Please cite this article as: N.K. Afseth, A. Kohler, Extended multiplicative signal correction in vibrational spectroscopy, a tutorial, Chemometr. Intell. Lab. Syst. (2012), doi:10.1016/j.chemolab.2012.03.004

EMSC modeling. This is related to the fact that the chemical information in Raman shifts is represented by very sharp bands that cannot be modeled by polynomials of a degree in this low range.

4. Extending the EMSC model by constituent spectra

So far we have captured chemical variation by the residual $e(\tilde{\nu})$ in the EMSC model. In the following we will consider an example where chemical variation is modeled by the EMSC model. Although part of the chemical variation in the data set is modeled, an un-modeled part is represented by a residual. Chemical variation can be integrated into EMSC models by using deviations of constituent spectra $\Delta k_j(\tilde{\nu})$ as defined in Eq. (3). This means that we preferably do not add the constituent spectra themselves to the EMSC model; instead we model the constituent spectra around the average spectra as expressed in Eq. (3). Mathematically, this means that we reduce confounding problems between constituent spectra and average spectra, which is requisite in order to make a least-squares estimate of the EMSC parameters [13]. For defining constituent spectra in the EMSC model we start with Eq. (7), where the residual part can be reduced by modeling one or few constituent difference spectra Δk_j:

$$A(\tilde{\nu}) = b \cdot \bar{x}(\tilde{\nu}) + \sum_{j=1}^{I} h_j \cdot \Delta k_j(\tilde{\nu}) + e(\tilde{\nu}) \tag{14}$$

where:

$$e(\tilde{\nu}) = \sum_{j=I+1}^{J} b \cdot c_j \cdot \Delta k_j(\tilde{\nu}) \tag{15}$$

with $I<J$ and $h_j = b \cdot c_j$. As in the previous section we may add polynomial expansions to model large baseline variations. Bertram et al. [26] used EMSC to correct FTIR microscopy spectra of meat tissue sections for variations of tightly bound water in the tissue sections. The variations due to the tightly bound water were caused by different amounts of water in the surrounding air introducing a strong day-to-day measurement variation. The thin tissue sections are in equilibrium with the surrounding air and the amount of tightly bound water in the tissue section is therefore directly depending on the amount of water vapor in the air. The water variation can be seen clearly in the O–H stretching region of the FTIR spectra between 3500 cm^{-1} and 3400 cm^{-1} as shown in Fig. 3a, where the spectra are shown after correction by a basic EMSC model. Fig. 3a shows clearly that the basic EMSC model cannot correct for these variations. The water variations also strongly affected the amide I region (1700 cm^{-1} to 1600 cm^{-1}) due to the O–H bending vibration which occurs around 1640 cm^{-1}. A clear day-to-day variation could be seen in a PCA of the amide I region (see Fig. 4a). In order to estimate a constituent difference spectrum that can correct for the water variations an additional experiment was performed. Spectra from the same position of a meat tissue section were acquired over time under different water vapor conditions. The different water vapor conditions were simulated using a purge box around the measurement compartment. A constituent difference spectrum was obtained as a difference spectrum between a spectrum obtained from the meat tissue section under high water vapor content in the air and a spectrum from the same meat tissue section (same position) with low water vapor content in the air. The difference spectrum is shown in Fig. 3b. The water vapor content in the spectra was kept at a minimum by acquiring background spectra in each environment. The difference spectrum shows a strong signature in the O–H stretching region between 3500 cm^{-1} and 3400 cm^{-1}, and, in particular, at the O–H bending vibration around 1640 cm^{-1}. Other regions all over the spectrum are also affected, which might be expected since the change in the amount of water in the tissues section may potentially lead to changes all over the spectrum. When adding the obtained constituent difference spectrum to Eq. (14) and correcting the spectra with the obtained model, the water variation could be removed completely as shown in Fig. 3c. Since the water constituent difference spectrum also shows a strong signature in the O–H bending region, we expect that the EMSC correction also affected the amide I bands. When a PCA was performed in the amide I region it turned out that an initially present day-to-day variation was considerably reduced

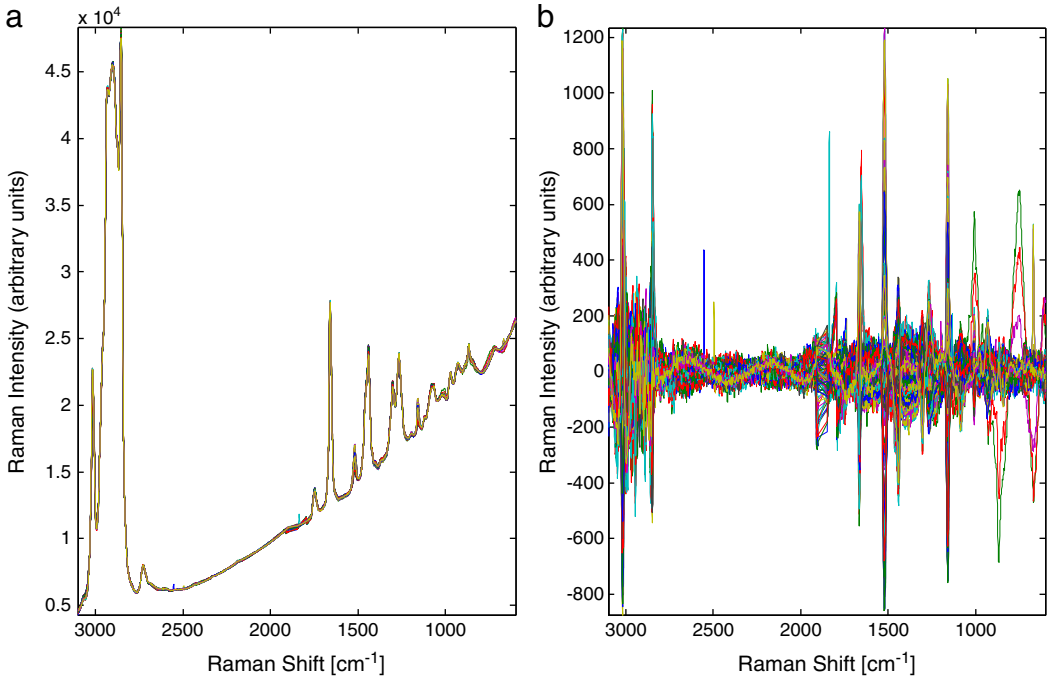

Fig. 2. Raman spectra of salmon oils corrected by a sixth order polynomial EMSC model (a) and corresponding mean-centered spectra (b).

Please cite this article as: N.K. Afseth, A. Kohler, Extended multiplicative signal correction in vibrational spectroscopy, a tutorial, Chemometr. Intell. Lab. Syst. (2012), doi:10.1016/j.chemolab.2012.03.004

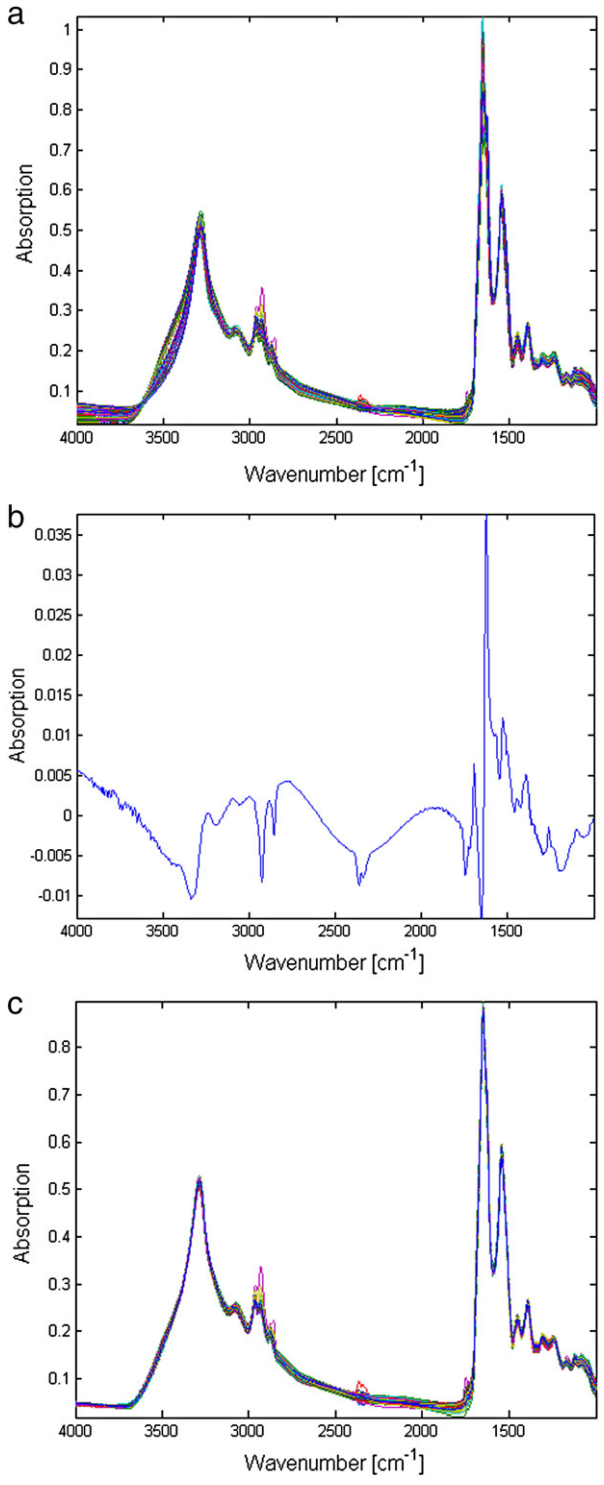

Fig. 3. An infrared constituent difference spectrum obtained from a meat tissue section under high and low water vapor contents, respectively, is shown (b) together with infrared spectra subjected to basic EMSC correction without (a) and with (c) employing the constituent difference spectrum.

after the EMSC correction with the water constituent difference spectrum (see Fig. 4b) [26].

5. Extending the EMSC model by orthogonal subspace models

We will now proceed to a more generic approach in extending the EMSC model by including estimations of scattering variations and other unwanted effects. In general, two different approaches might be

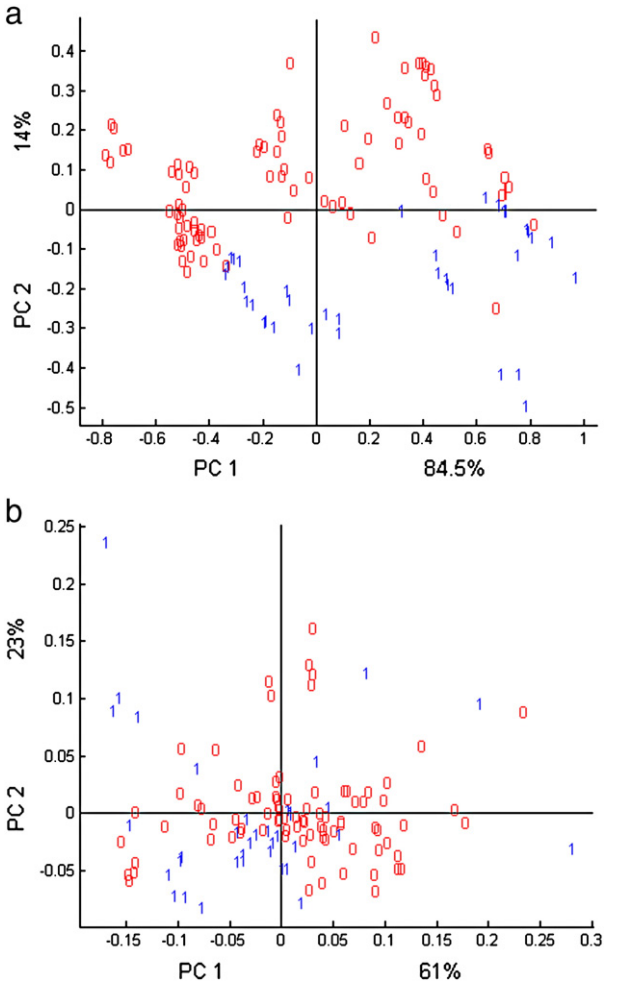

Fig. 4. PCA score plot of amide I region of EMSC corrected spectra before (a) and after (b) EMSC water correction. The samples are marked according to their water content (low water content samples are marked with 0, while high water content samples are marked with 1).

considered to accomplish this: estimating unwanted effects from the data itself, or including physical models to estimate the unwanted effects. In this section the former approach will be discussed.

A feasible approach to estimate unwanted features in vibrational spectra involves a close inspection of the systematic between-replicate variation for all samples in a given system. This approach has been utilized in the so-called EMSC replicate correction, originally developed to reduce the between-replicate variation in FT-IR spectra of microorganisms [27]. In this approach, variations between technical replicates are summarized by a small number of components and built into the EMSC replicate model. The between replicate variation is estimated in the following way: 1) For each sample, a basic EMSC model (according to Eq. (12)) is established for all technical replicates. 2) Then, for each sample, all technical replicates are mean-centered. 3) Subsequently, all EMSC corrected mean-centered replicate spectra are merged into one data table. 4) Finally, principal component analysis is used to estimate the main between-replicate variation in the data set. The between-replicate variation spectra are thus summarized in a few orthogonal subspace components. Generally, two or three technical replicates for each sample are sufficient in order to obtain sound estimates of the between-replicate variation, as long as the data set consists of a sufficient amount of samples.

After having estimated the between-replicate variation in the data-set, a new global EMSC model can be calculated on the original spectra

Please cite this article as: N.K. Afseth, A. Kohler, Extended multiplicative signal correction in vibrational spectroscopy, a tutorial, Chemometr. Intell. Lab. Syst. (2012), doi:10.1016/j.chemolab.2012.03.004

including the most important principal components describing the replicate correction:

$$A(\tilde{\nu}) = a + \bar{x}(\tilde{\nu})\cdot b + d_1\tilde{\nu} + d_2\tilde{\nu}^2 + \sum_{k=1}^{A} g_k \cdot p_k(\tilde{\nu}) + e(\tilde{\nu}) \quad (16)$$

where $p_k(\tilde{\nu})$ refers to the A most important replicate variation loading vectors as estimated by principal component analysis. Each spectrum is subsequently corrected according to Eq. (17):

$$A_{corr}(\tilde{\nu}) = \frac{A(\tilde{\nu}) - a - d_1\tilde{\nu} - d_2\tilde{\nu}^2 - \sum_{k=1}^{A} g_k \cdot p_k(\tilde{\nu})}{b}. \quad (17)$$

Usually only a few components are needed or should be used for summing up most of the replicate variation.

We now turn back to the Raman salmon oil data set discussed in Section 2, consisting of spectra of 45 salmon oil samples with 3 replicate spectra for each sample. Since the oil samples are homogenous, we can expect that the replicate variation estimated from these spectra is largely related to physical effects and not to chemical differences. The results of the EMSC replicate correction approach are shown in Fig. 5. When comparing the mean-centered EMSC corrected spectra to the corresponding spectra of Figs. 1 and 2, we see that the broad underlying structures visible in the latter figures are completely removed in Fig. 5. In the modeling, four loadings were included in the EMSC replicate model. These are shown in Fig. 6 together with the model spectra and the reference spectrum chosen for the EMSC replicate model. For this procedure it should be kept in mind that unwanted effects of vibrational spectra are estimated from the data itself. Thus, it is very important to be restrictive in the number of components that is included in the final EMSC model. The number of components that could be included will mainly depend on the variation in the given dataset, and a visual inspection and validation to guide the choice of components is necessary in order to avoid removing any "wanted" information from the data set. In Fig. 6c, for instance, it is clear that the 4 subspace-model components included mainly explain broad baseline structures and noise structures, thus ensuring that no chemical information will been removed from the data set.

6. Discussion

The examples of this tutorial illustrate how EMSC acts as a general framework for preprocessing. The framework is flexible and allows the user to include known or estimated physical and chemical variations, depending on the application and nature of the data sets. As mentioned in Section 5, physical models can be included in the EMSC framework to estimate unwanted effects in the data. Recently, for instance, Mie scattering in single-cell FT-IR spectroscopy has been corrected by building mathematical models of this scattering phenomenon into the EMSC framework [15–17].

The user should in general bear in mind the possibility of rank problems when adding extra terms to the EMSC model. For the least squares estimation of the EMSC parameters, full rank of the model spectrum matrix is desired. If the model spectra do not have full rank, the estimated EMSC parameters may be erroneous. The rank of the matrix of model spectra may be estimated by principal component analysis. It is, however, important to note that the examples of this tutorial present no rank problems. The polynomial terms added to the EMSC model in Section 3 are linearly independent, and since they describe broad structures compared to the chemical peaks in Raman, they are also linearly independent from the reference spectrum. In Section 4, full rank of the model spectrum matrix is ensured by including the difference spectrum $\Delta k_j(\tilde{\nu})$. And the components added to the model in Section 5 are orthogonal by definition.

One important aspect of EMSC has not been discussed so far: the interpretation of the EMSC model parameters. EMSC modeling provides quantitative estimates of the different physical or chemical features included in the model. This is one of the major benefits of model-based preprocessing compared to the filtering approaches, and several authors have explored this possibility. Kohler et al. [13] related estimated EMSC parameters of FT-IR microscopy spectra of muscle sections directly to different textural properties of the muscle samples due to heat treatment. De Gelder et al. [20] used EMSC parameters to track metabolic products in Raman spectra of the bacterium *Cupriavidus metallidurans* LMG 1195 in five stages of its growth.

There are several software packages available that provide the use of EMSC. EMSC is for instance included as a standard tool for preprocessing in The Unscrambler® (Camo Process AS, Norway), and available source codes include the EMSC_Toolbox from Eigenvector Research (www.

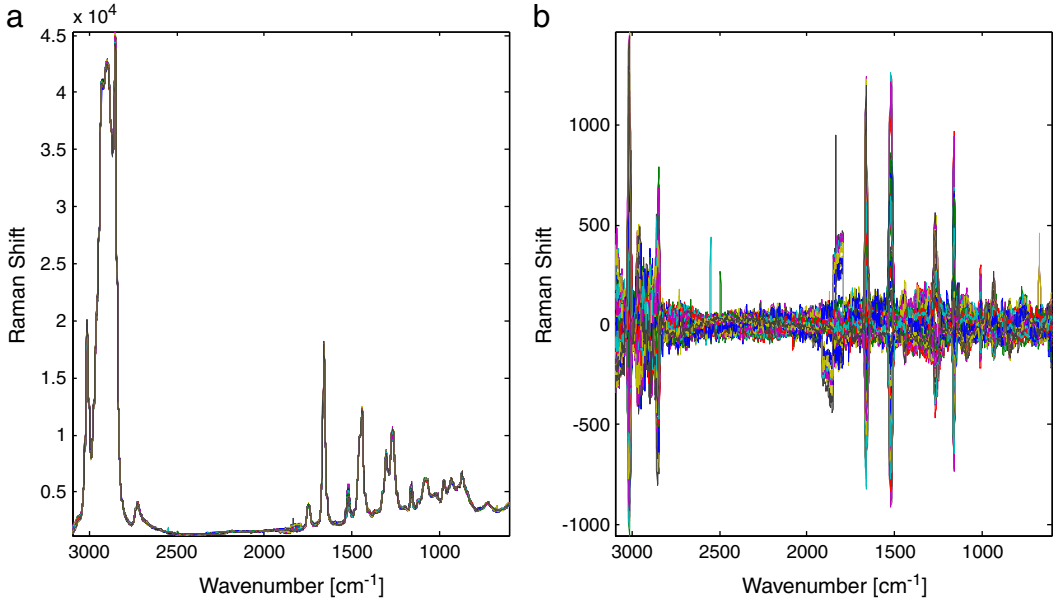

Fig. 5. Raman spectra of salmon oils corrected by a replicate EMSC model (a) and corresponding mean-centered spectra (b).

Please cite this article as: N.K. Afseth, A. Kohler, Extended multiplicative signal correction in vibrational spectroscopy, a tutorial, Chemometr. Intell. Lab. Syst. (2012), doi:10.1016/j.chemolab.2012.03.004

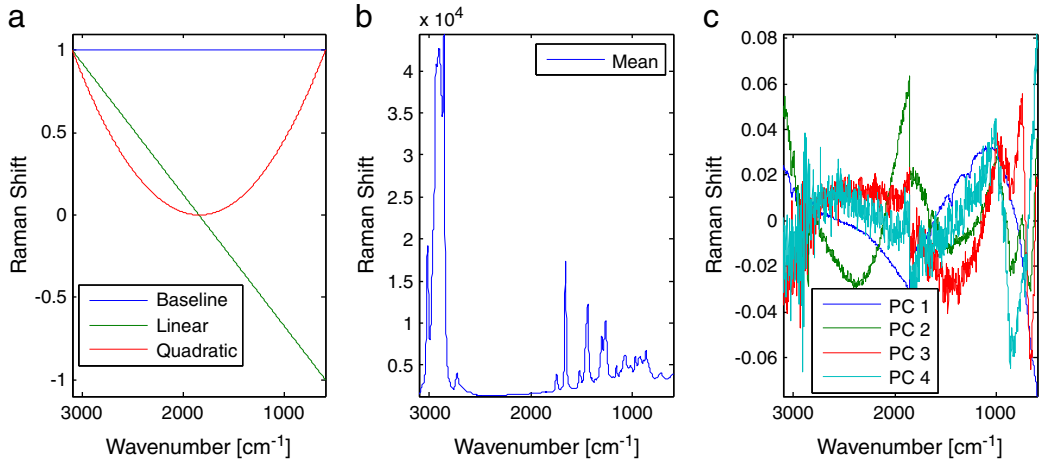

Fig. 6. The model spectra used for EMSC replicate correction of Raman salmon oil spectra: model spectra of constant, linear, and quadratic effects (a); the reference spectrum (b); the four principal component spectra obtained from a principal component analysis of replicate variation (c).

eigenvector.com) and the EMSC toolbox for MATLAB (www.models.life.ku.dk/EMSCtoolbox). For all these sources, the user may preprocess according to the basic EMSC model, and possibilities for adding difference spectra are also included in the packages. However, the replicate EMSC approach and the possibility for adding polynomial terms are not included. A versatile Matlab-based script covering all the aspects of EMSC preprocessing as described in this tutorial can be downloaded at the web site www.specmod.org.

7. Conclusion

EMSC is a model-based preprocessing framework that possesses a huge potential in the preprocessing of vibrational spectra. In addition to provide basic preprocessing steps related to baseline correction and normalization, EMSC also allows to separate and quantify different types of chemical and physical variations in vibrational spectra. EMSC thus contributes to achieving better interpretability of the vibrational spectra, in addition to making the subsequent calibration models simpler and statistically more robust. The EMSC approach and its extensions are adaptable to a standard calibration/validation scheme. In recent years the EMSC approach has been frequently used to estimate and correct for specific physical and chemical features in both infrared absorbance and Raman scattering spectra, and it is expected that the approach will develop into new important applications within vibrational spectroscopy in the years to come.

Acknowledgment

Financial support from the Agricultural Food Research Foundation of Norway and by the grants 199581/I10 and 178280/I10 financed by the Research Council of Norway is greatly acknowledged.

References

[1] J.L. Ilari, M. Martens, T. Isaksson, Determination of particle size in powders by scatter correction in diffuse near-infrared reflectance, Applied Spectroscopy 42 (1988) 722–728.
[2] H. Martens, S.A. Jensen, P. Geladi, Multivariate linearity transformations for near infrared reflectance spectroscopy, in: O.H.J. Christie (Ed.), Nordic Symposium on Applied Statistics, Stokkland Forlag, Stavanger, Norway, 1983, pp. 205–234.
[3] P. Geladi, D. Macdougall, H. Martens, Linearization and scatter-correction for near-infrared reflectance spectra of meat, Applied Spectroscopy 39 (1985) 491–500.
[4] H. Martens, E. Stark, Extended multiplicative signal correction and spectral interference subtraction — new preprocessing methods for near-infrared spectroscopy, Journal of Pharmaceutical and Biomedical Analysis 9 (1991) 625–635.
[5] H. Martens, J.P. Nielsen, S.B. Engelsen, Light scattering and light absorbance separated by extended multiplicative signal correction. Application to near-infrared transmission analysis of powder mixtures, Analytical Chemistry 75 (2003) 394–404.
[6] A. Rinnan, F. van den Berg, S.B. Engelsen, Review of the most common pre-processing techniques for near-infrared spectra, TRAC Trends in Analytical Chemistry 28 (2009) 1201–1222.
[7] S. Ottestad, T. Isaksson, W. Saeys, J.P. Wold, Scattering correction by use of a priori information, Applied Spectroscopy 64 (2010) 795–804.
[8] D.K. Pedersen, H. Martens, J.P. Nielsen, S.B. Engelsen, Near-infrared absorption and scattering separated by extended inverted signal correction (EISC): analysis of near-infrared transmittance spectra of single wheat seeds, Applied Spectroscopy 56 (2002) 1206–1214.
[9] M. Dyrby, R.V. Petersen, J. Larsen, B. Rudolf, L. Norgaard, S.B. Engelsen, Towards on-line monitoring of the composition of commercial carrageenan powders, Carbohydrate Polymers 57 (2004) 337–348.
[10] N.B. Gallagher, T.A. Blake, P.L. Gassman, Application of extended inverse scatter correction to mid-infrared reflectance spectra of soil, Journal of Chemometrics 19 (2005) 271–281.
[11] S.W. Bruun, A. Kohler, A. I., G.D. Sockalingum, M. Manfait, H. Martens, Correcting attenuated total reflaction-Fourier transform infrared spectra for water vapour and carbon dioxide, Applied Spectroscopy 60 (2006) 1029–1039.
[12] A. Kohler, N.K. Afseth, H. Martens, Chemometrics in biospectroscopy, in: J.M. Chalmers, E. Li-Chan, P.R. Griffiths (Eds.), Applications of Vibrational Spectroscopy in Food Science, John Wiley & Sons Ltd., Chichester, UK., 2010
[13] A. Kohler, C. Kirschner, A. Oust, H. Martens, Extended multiplicative signal correction as a tool for separation and characterization of physical and chemical information in Fourier transform infrared microscopy images of cryo-sections of beef loin, Applied Spectroscopy 59 (2005) 707–716.
[14] H. Martens, S.W. Bruun, I. Adt, G.D. Sockalingum, A. Kohler, Pre-processing in biochemometrics: correction for path-length and temperature effects of water in FTIR bio-spectroscopy by EMSC, Journal of Chemometrics 20 (2006) 402–417.
[15] P. Bassan, A. Kohler, H. Martens, J. Lee, H.J. Byrne, P. Dumas, E. Gazi, M. Brown, N. Clarke, P. Gardner, Resonant Mie scattering (RMieS) correction of infrared spectra from highly scattering biological samples, Analyst 135 (2010) 268–277.
[16] P. Bassan, A. Kohler, H. Martens, J. Lee, E. Jackson, N. Lockyer, P. Dumas, M. Brown, N. Clarke, P. Gardner, RMieS-EMSC correction for infrared spectra of biological cells: extension using full Mie theory and GPU computing, Journal of Biophotonics 3 (2010) 609–620.
[17] A. Kohler, J. Sule-Suso, G.D. Sockalingum, M. Tobin, F. Bahrami, Y. Yang, J. Pijanka, P. Dumas, M. Cotte, D.G. van Pittius, G. Parkes, H. Martens, Estimating and correcting Mie scattering in synchrotron-based microscopic Fourier transform infrared spectra by extended multiplicative signal correction, Applied Spectroscopy 62 (2008) 259–266.
[18] N.K. Afseth, V.H. Segtnan, J.P. Wold, Raman spectra of biological samples: a study of preprocessing methods, Applied Spectroscopy 60 (2006) 1358–1367.
[19] A. Oust, T. Moretro, K. Naterstad, G.D. Sockalingum, I. Adt, M. Manfait, A. Kohler, Fourier transform infrared and Raman spectroscopy for characterization of *Listeria monocytogenes* strains, Applied and Environmental Microbiology 72 (2006) 228–232.
[20] J. De Gelder, K. De Gussem, P. Vandenabeele, P. De Vos, L. Moens, Methods for extracting biochemical information from bacterial Raman spectra: an explorative study on *Cupriavidus metallidurans*, Analytica Chimica Acta 585 (2007) 234–240.
[21] P. Heraud, B.R. Wood, J. Beardall, D. McNaughton, Effects of pre-processing of Raman spectra on in vivo classification of nutrient status of microalgal cells, Journal of Chemometrics 20 (2006) 193–197.
[22] K. Maquelin, M. Scholtes-Timmerman, H. Willemse-Erix, T.B. Schut, A. van Belkum, G. Puppels, A novel approach to correct variations in Raman spectra due to photobleachable cellular components, Analyst 134 (2009) 387–393.

Please cite this article as: N.K. Afseth, A. Kohler, Extended multiplicative signal correction in vibrational spectroscopy, a tutorial, Chemometr. Intell. Lab. Syst. (2012), doi:10.1016/j.chemolab.2012.03.004

[23] E.W. Stark, H. Martens, Multiplicative Signal Correction Method and Apparatus, in, US, , 1996.
[24] N.K. Afseth, J.P. Wold, V.H. Segtnan, The potential of Raman spectroscopy for characterisation of the fatty acid unsaturation of salmon, Analytica Chimica Acta 572 (2006) 85–92.
[25] C.A. Lieber, A. Mahadevan-Jansen, Automated method for subtraction of fluorescence from biological Raman spectra, Applied Spectroscopy 57 (2003) 1363–1367.
[26] H.C. Bertram, A. Kohler, U. Bocker, R. Ofstad, H.J. Andersen, Heat-induced changes in myofibrillar protein structures and myowater of two pork qualities. A combined FT-IR spectroscopy and low-field NMR relaxometry study, Journal of Agricultural and Food Chemistry 54 (2006) 1740–1746.
[27] A. Kohler, U. Bocker, J. Warringer, A. Blomberg, S.W. Omholt, E. Stark, H. Martens, Reducing inter-replicate variation in Fourier transform infrared spectroscopy by extended multiplicative signal correction, Applied Spectroscopy 63 (2009) 296–305.

Please cite this article as: N.K. Afseth, A. Kohler, Extended multiplicative signal correction in vibrational spectroscopy, a tutorial, Chemometr. Intell. Lab. Syst. (2012), doi:10.1016/j.chemolab.2012.03.004