Analytical Methods

PAPER

Cite this: *Anal. Methods*, 2016, **8**, 435

# Non-linear multivariate curve resolution applied to the spectrophotometric determination of cerium(III) in aqueous nitric acid solutions for process control†

Oxana Ye. Rodionova and Alexey L. Pomerantsev*

It is shown that the recently proposed Non-Linear Multivariate Curve Resolution (NL-MCR) method can be effectively employed to develop an accurate calibration of cerium(III) using spectrophotometry measurements of mixtures of rare earth elements in nitric acid. Spectroscopic techniques provide a unique opportunity for the in-line determination of critical concentrations rapidly and without serious risks to operating personnel and the environment. Cerium has no absorbance bands in the visual and near-infrared range. In the ultraviolet range cerium(III) has a unique large peak which overlaps with even larger peaks of nitric acid. In the case of in-line control, where conventional analytical means are limited, we encounter a peak flattening and, consequently, a Lambert–Beer law violation. Therefore, the conventional calibration methods, such as Partial Least Squares (PLS) and Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS), yield useless results. Our previous attempt to overcome Beer's law violation by introducing a non-linear constraint in the MCR-ALS procedure failed. Application of the NL-MCR method with specially selected transition function not only yields accurate cerium determination but also provides an opportunity to assess the unknown nitric acid concentration in new samples. It is shown that the established calibration models are stable to some extent for out-of-control cases.

Received 8th September 2015
Accepted 30th November 2015

DOI: 10.1039/c5ay02393a

www.rsc.org/methods

## 1. Introduction

Molecular spectroscopy in the 200–2500 nm range is the main tool in the frame of Process Analytical Technology (PAT).[1] This approach is of special interest in the nuclear industry, particularly in the field of spent nuclear fuel reprocessing.[2] A non-invasive on-line control provides a unique opportunity to determine the desired concentrations rapidly and without serious risk of radiation pollution.[3] Rare earth elements (REEs), and, in particular, cerium are in the focus of technologists' attention.[4]

It should be emphasized, that the spectrophotometric determination of cerium‡ is a hard job. Furthermore, if Ce is dissolved in nitric acid, it is a very hard job. Moreover, in case the determination is performed in-line, where the conventional analytical facilities are limited, this problem becomes a particularly arduous one.[5] It is known that cerium has no absorbance bands in the visual and near-infrared range, but has a large peak[6–8] near 250 nm. In the UV range, an aqueous solution of $HNO_3$ has two partly overlapping peaks around 247 nm and 295 nm.[9]

Example spectra are shown in Fig. 1a: the spectrum of 0.4 M nitric acid (blue curve 1), and the spectrum of 0.6 g $L^{-1}$ cerium (green curve 2). The spectrum of their mixture is presented by the black curve 3. All spectra shown in Fig. 1a are the 'ideal' ones because there exists no spectrometer that can measure them properly. All spectrometers have limited sensitivity of the detector, which cannot measure the low transmittance signal in highly absorptive media. The level of 3 AU (red dashed line 4) is a typical limit of a spectrophotometer's sensitivity and all absorbance peaks above this level are flattened.

Curves shown in Fig. 1b are the 'real' spectra that could be acquired by using a spectrometer with a saturation level of 3 AU. The black curve (3s) represents the distorted spectrum of a mixture of $HNO_3$ and Ce. The blue curve (1s) demonstrates the flattened spectrum of nitric acid. Finally, the green curve (2s) stands for the reconstructed spectrum of cerium that is obtained when the distorted spectrum of nitric acid (1s) is subtracted as a background from the distorted spectrum of the mixture (3s). Comparing the outcome (2s) with the 'ideal' spectrum of cerium (green dashed curve 2), we see that the peak height of (2s) is 1.2 AU instead of 2 AU, and the peak position shifted from 253 nm to 263 nm.

This reasoning clearly explains that the $HNO_3$ + Ce mixture does not comply with the Lambert–Beer law, which is a basic spectrophotometric principle. There are many ways to avoid such situations in a laboratory, *e.g.* by means of sample dilution, shortening of the optical path length, or, where possible,

*Semenov Institute of Chemical Physics RAS, Kosygin str. 4, 119991, Moscow, Russia. E-mail: forecast@chph.ras.ru*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c5ay02393a

‡ Hereinafter cerium means cerium(III).

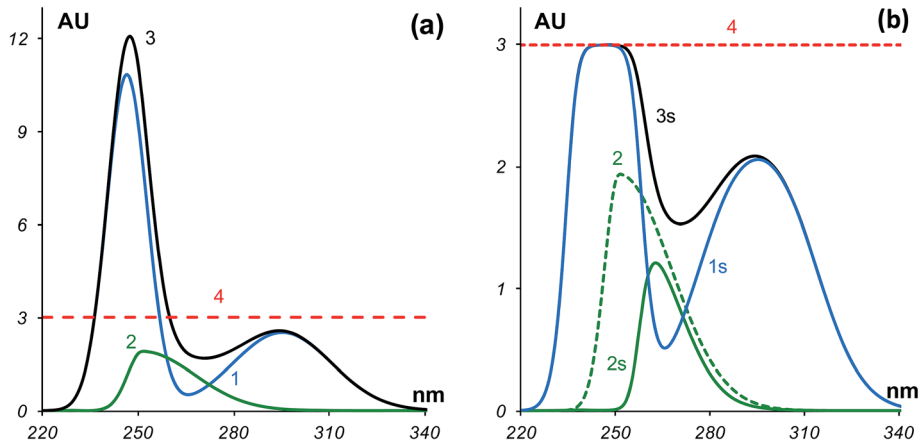


Fig. 1 (a) 'Ideal' absorption spectra: (1) 0.4 M $HNO_3$, (2) 0.6 g $L^{-1}$ Ce, and (3) mixture $HNO_3$ + Ce; (b) acquired saturated spectra: (1s) 0.4 M $HNO_3$, (3s) mixture $HNO_3$ + Ce; (2) ideal spectrum of 0.6 g $L^{-1}$ Ce, and (2s) the reconstructed spectrum of Ce. In both plots, the dashed line (4) represents the saturation level.

changing the spectral region. However, if the purpose of modeling is a PAT solution, the instruments and experimental set-up should be similar to those in the real process facilities. Practically this means that the spectral resolution and wavelength accuracy are subpar, the optical path length is fixed, no dilution is used, and no extra reagents are employed.

Another aspect is the variability of process parameters, which are often out of control in comparison with laboratory scale models. For these reasons, many appropriate laboratory scale PAT solutions have not been implemented in practice. To avoid such problems, the PAT aimed models should be, to some extent, insensitive to distorted data and to predictions out of the explored concentration range.

This paper is closely connected with our earlier publication.[5] Moreover, it can be considered as its natural continuation. In the former paper, we discussed the feasibility of simultaneous spectrophotometric determination of cerium, praseodymium, and neodymium using the UV-vis range. The predictability of two chemometric techniques, Partial Least Squares (PLS) regression and correlation constrained Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) were compared in order to evaluate the models' performance. It has been shown that both models are fairly similar; they are stable to some extent for out-of-control cases, and thus can be reliably used for in-line control of neodymium and praseodymium in nitric acid. In the case of cerium, we obtained a disappointing result. All models of Ce calibration demonstrated low quality and obvious non-linearity. It was also shown that the presence of Nd and Pr in the solutions does not influence the spectra in the UV range and, thus, does not hinder the quantification of Ce. The only restriction is a common experimental set-up that enables simultaneous quantification of REEs.

The aim of the current study is the development of a calibration model for the quantification of cerium in nitric acid solutions using UV-spectrophotometry and the recently proposed[9] Non-Linear Multivariate Curve Resolution (NL-MCR) method. For comparison, we employ the correlation constrained MCR-ALS method,[10,11] which is a standard calibration technique. The most popular PLS regression technique has been investigated in full in our previous publication,[5] therefore it is not considered here. In contrast to the former paper, the spectral data are considered in two different ways. In the first case, nitric acid is subtracted from the spectra as a background. In the second case, the background is preserved and treated as an additional component.

## 2. Theory

A calibration model relates two parts of data, namely, the $(I \times J)$ matrix $\mathbf{X}$ that contains spectra of $I$ samples recorded for $J$ wavelengths, and the $(I \times N)$ matrix $\mathbf{C}$ of concentrations. $N$ stands for the number of components. Direct calibration is based on the principle of linearity, aka the Lambert–Beer law,

$$\mathbf{X} = \mathbf{C}\mathbf{S}^{t}. \quad (1)$$

In this equation, $\mathbf{S}$ is the $(J \times N)$ matrix of pure spectra, also called the sensitivity matrix. If matrix $\mathbf{S}$ is known, the concentrations are calculated by an equation

$$\mathbf{C} = \mathbf{X}\mathbf{B}, \quad (2)$$

where $\mathbf{B} = \mathbf{S}(\mathbf{S}^{t}\mathbf{S})^{-1}$ is the $(J \times N)$ matrix. However, in practice, matrix $\mathbf{S}$ is usually unknown and should be reconstructed from a training subset of data.

The obtained model (*i.e.* matrix $\mathbf{B}$) is then applied to another subset $\mathbf{X}_{new}$ that can be a validation, or test, or new data set. The predicted concentrations

$$\mathbf{C}_{new} = \mathbf{X}_{new}\mathbf{B} \quad (3)$$

can be compared with the known reference values, or directly used to monitor the process.

Two methods employed in this paper are the correlation constrained MCR-ALS method, and the recently proposed NL-MCR.



 

### 2.1. Correlation constrained MCR-ALS

MCR-ALS is a well-established method.[12] This approach resolves spectroscopic data, and, then, it uses the obtained spectroscopic profiles for the quantitative calibration. MCR-ALS is based on the equation

$$\mathbf{X} = \mathbf{C}\mathbf{S}^{t} + \mathbf{E}. \quad (4)$$

Here $\mathbf{S}$ is the $(J \times N)$ matrix of pure spectra, $N$ is the number of components in the system, and $\mathbf{E}$ is the $(I \times J)$ matrix, which contains variations not explained by the model.

The procedure starts with evaluation of the initial estimate of matrix $\mathbf{S}$, or matrix $\mathbf{C}$.[13] In calibration, it is customary to use the known matrix $\mathbf{C}_{ref}$ for the initial step. Then, matrices $\mathbf{C}$ and $\mathbf{S}$ are found by the ALS method that iteratively minimizes the sum of the squared residuals

$$\|\mathbf{X} - \mathbf{C}\mathbf{S}^{t}\|^{2}. \quad (5)$$

The algorithm consists of two steps, the C-type step, and the S-type step, which are repeated until convergence. In the C-type step, matrix $\mathbf{S} \equiv \mathbf{S}_{hat}$ is fixed, and the $\mathbf{C}$ matrix is calculated using the unconstrained least squares (LS) estimator

$$\mathbf{C}_{in} = \mathbf{X}\mathbf{S}_{hat}(\mathbf{S}_{hat}^{t}\mathbf{S}_{hat})^{-1}. \quad (6)$$

Then, matrix $\mathbf{C}_{in}$ is transformed into matrix $\mathbf{C}_{hat}$ to incorporate the constraints for the concentration profile.

For the S-type step, matrix $\mathbf{C} \equiv \mathbf{C}_{hat}$ is fixed and matrix $\mathbf{S}$ is found using a similar formula

$$\mathbf{S}_{in} = \mathbf{X}^{t}\mathbf{C}_{hat}(\mathbf{C}_{hat}^{t}\mathbf{C}_{hat})^{-1}. \quad (7)$$

Subsequently, matrix $\mathbf{S}_{in}$ is transformed into matrix $\mathbf{S}_{hat}$ to incorporate the constraints for the spectral profile. The calculations, given in eqn (7) and eqn (8), and transformations subject to corresponding constraints are repeated till convergence.

Constraints are applied to give a physicochemical meaning to the LS estimates $\mathbf{C}_{in}$ and $\mathbf{S}_{in}$. For example, there are natural non-negativity restrictions that force the concentrations and spectra to be equal or greater than zero. In calibration problems, additional correlation constraints[10,11] are applied. They are based on regressions that relate the known reference concentration matrix, $\mathbf{C}_{ref}$, to matrix $\mathbf{C}_{in}$ obtained in the ALS procedure by eqn (6).

$$\mathbf{C}_{ref} = \mathbf{C}_{in}\mathbf{A} + \mathbf{G}. \quad (8)$$

Here $\mathbf{A}$ is the $(N \times N)$ matrix of slopes, and $\mathbf{G}$ is the $(I \times N)$ matrix of intercepts.[10] In a simple case, a univariate regression $\mathbf{c}_{n,ref} = a_n\mathbf{c}_{n,in} + g_n$ is developed for each component concentration vector $\mathbf{c}_n$.[11] Then, $\mathbf{A} = \mathrm{diag}(a_1,\ldots, a_n)$, and $\mathbf{G} = (g_1\mathbf{1},\ldots, g_n\mathbf{1})$, where $\mathbf{1}$ is the $(N \times 1)$ vector of units.

The estimated regression matrices $\mathbf{A}$ and $\mathbf{G}$ are used to find the adjusted concentration matrix, $\mathbf{C}_{hat}$, applying the following formula

$$\mathbf{C}_{hat} = \mathbf{C}_{in}\mathbf{A} + \mathbf{G}. \quad (9)$$

In MCR-ALS, the prediction is performed by means of the following formula

$$\mathbf{C}_{new} = \mathbf{X}_{new}\mathbf{S}_{hat}(\mathbf{S}_{hat}^{t}\mathbf{S}_{hat})^{-1}\mathbf{A} + \mathbf{G} \quad (10)$$

where $\mathbf{X}_{new}$ is a matrix of new (spectral) data, and matrices $\mathbf{A}$, $\mathbf{G}$, and $\mathbf{S}_{hat}$ have been obtained at the calibration stage.

Before starting the MCR-ALS procedure, it is necessary to determine the number of chemical components, *i.e.* the $N$ value.

### 2.2. Non-linear MCR

The majority of chemometric method utilizes eqn (1) in the optimization problem given in eqn (5). In practice, due to the numerous unavoidable problems (detectors, noise, baseline, *etc*), matrix $\mathbf{X}$ is always distorted. Consequently, $\mathbf{X}$ is improved using preprocessing methods, such as smoothing, baseline corrections, *etc.*, which are aiming to make data closer to the model. As a result, we come to the following optimization task

$$\underset{\mathbf{C},\mathbf{S}}{\text{minimize}} \|G(\mathbf{X}) - \mathbf{C}\mathbf{S}^{t}\|^{2} \quad (11)$$

where $G$ is an improving filter.

Instead of improving $\mathbf{X}$, we propose to impair matrix $\mathbf{C}\mathbf{S}^{t}$ by applying filter $F$ with a similar goal – to draw the model nearer to the data. Such a perverse view of linearity turns out to be useful in the analysis of complex data. The first example belongs to Richard Harshman *et al.*[14] who employed the shift operator as an $F$ filter in factor analysis.

As discussed above, bands with high absorbance produce saturated spectra. This causes a loss of linearity between the peak height and concentration, and, thus, leads to a violation of the Lambert–Beer law. In general, we can suppose that acquired spectral matrix $\mathbf{X}$ relates to product $\mathbf{C}\mathbf{S}^{t}$ through an equation

$$\mathbf{X} = F(\mathbf{C}\mathbf{S}^{t}) + \mathbf{E}, \quad (12)$$

where filter $F$ is a row-wise non-linear function that converts each 'ideal' spectrum $\mathbf{c}_i\mathbf{S}^{t}$ into 'real' spectrum $\mathbf{x}_i$. In a previous paper[9] we introduced an empirical transition function that accounts for variations in the shape of saturated peaks. It is given by the following formula

$$F(\mathbf{x}) = \{\mathbf{x}\}_{s,p} = s\left[\tanh\left(\frac{x^{p}}{s^{p}}\right)\right]^{1/p}. \quad (13)$$

Here $\mathbf{x}$ is the ideal spectrum and $\{\mathbf{x}\}_{s,p}$ is its saturated modification. Function $\tanh(\cdot)$ stands for the hyperbolic tangent, $p$ is the peak shape parameter, and $s$ is the saturation level, *i.e.* the upper value above which an instrument cannot measure absorbance accurately.

The transition function $\{\mathbf{x}\}_{s,p}$ possesses the following evident properties. At small absorbance values ($x \ll s$), it is close to $x$, that is $\{x\}_{s,p} = x$; and at large absorbance values ($x \gg s$), it assumes a constant value $\{x\}_{s,p} = s$. In the proximity of the transition area ($x \approx s$) the function shape is defined by the

parameter $p$: the larger the parameter, the sharper the transition.

The NL-MCR method is based on eqn (12), which contains the above explained nonlinear filter $F$. Clearly, the linear LS method cannot be used to solve eqn (12), and we arrive at a nonlinear optimization problem

$$\underset{\mathbf{C},\mathbf{S}}{\text{minimize}} \|\mathbf{X} - F(\mathbf{C}\mathbf{S}^{\mathrm{t}})\|^2 \quad (14)$$

the solution of which can be obtained using the following modified ALS algorithm.

The C-step (compared with Eq. (6)) is now presented by a problem

$$\underset{\mathbf{C}}{\text{minimize}} \|\mathbf{X} - F(\mathbf{C}\mathbf{S}^{\mathrm{t}})\|^2 \quad \text{subject to } \mathbf{S} \equiv \mathbf{S}_{\mathrm{hat}}. \quad (15)$$

The S-step (see Eq.(7)) is performed *via* optimization

$$\underset{\mathbf{S}}{\text{minimize}} \|\mathbf{X} - F(\mathbf{C}\mathbf{S}^{\mathrm{t}})\|^2 \quad \text{subject to } \mathbf{C} \equiv \mathbf{C}_{\mathrm{hat}}. \quad (16)$$

The correlation constraint presented by eqn (8), where $\mathbf{C}_{\mathrm{in}}$ is now a solution of the problem given by eqn (15), is also used for calibration problems. Other restrictions (non-negativity, unimodality, *etc.*) for concentrations and spectra are also very important in the non-linear case. These constraints either may be applied in the same manner as in the classical procedure, or they can be included in a nonlinear optimization procedure.

In NL-MCR, the prediction is performed *via* a one-step optimization

$$\underset{\mathbf{C}}{\text{minimize}} \|\mathbf{X}_{\mathrm{new}} - F(\mathbf{C}\mathbf{S}^{\mathrm{t}})\|^2 \quad \text{subject to } \mathbf{S} \equiv \mathbf{S}_{\mathrm{hat}} \quad (17)$$

where $\mathbf{X}_{\mathrm{new}}$ is a matrix of new (spectral) data. Matrix $\mathbf{C}_{\mathrm{in}}$ that provides the minimum of eqn (17) is then converted into matrix $\mathbf{C}_{\mathrm{new}}$ with the aim to incorporate the constraints, *e.g.* $\mathbf{C}_{\mathrm{new}} = \max(\mathbf{0}, \mathbf{A}\mathbf{C}_{\mathrm{in}} + \mathbf{G})$. It is important that for prediction, all matrices $\mathbf{S}_{\mathrm{hat}}$, $\mathbf{A}$ and $\mathbf{G}$ are set equal to the values, which have been obtained at the calibration stage.

In some aspects, NL-MCR is very similar to the conventional MCR-ALS. The latter approach estimates matrices $\mathbf{C}$ and $\mathbf{S}$ whose product $\mathbf{C}\mathbf{S}^{\mathrm{t}}$ fits matrix $\mathbf{X}$ better. The non-linear MCR also seeks for matrices $\mathbf{C}$ and $\mathbf{S}$, but the aim is to fit in the distorted product $F(\mathbf{C}\mathbf{S}^{\mathrm{t}})$ with matrix $\mathbf{X}$. From a computational point of view, NL-MCR is more complex than MCR-ALS, because in each C- or S-step we have to solve a non-linear optimization problem, instead of direct matrix calculations. Previously, NL-MCR was used for spectrophotometric determination of aqueous nitric acid solutions.[9]

### 2.3. Figures of merit

We use the root mean square errors (RMSEs) of calibration, RMSEC, validation, RMSEV, and prediction, RMSEP, which are calculated by the formula

$$\mathrm{RMSE} = \sqrt{\sum_{i=1}^{I} (c_i - \hat{c}_i)^2 \Big/ I} \quad (18)$$

where $c_i$ are the reference concentration values, $\hat{c}_i$ are the estimated values, and $I$ is the number of samples in the calibration, or validation, or prediction set. RMSEC characterizes the quality of calibration. RMSEV is used for validation, namely it helps to select the optimal complexity of the model. RMSEP is utilized to assess the model stability with respect to the anticipated irregularities.

The RMSE is not an appropriate measure for a mixture calibration, because the component concentrations can be presented in very different scales. For example, the cerium concentration is given in g $\mathrm{L}^{-1}$, while the traditional unit for the acid concentration is M = mol $\mathrm{L}^{-1}$. For this reason the relative error, RE, is applied. Presented in percents (RE%), it is calculated as follows

$$\mathrm{RE\%} = 100\sqrt{\sum_{i=1}^{I} (c_i - \hat{c}_i)^2 \Big/ \sum_{i=1}^{I} c_i^2} \quad (19)$$

where $c_i$ are the reference concentration values, $\hat{c}_i$ are the estimated values, and $I$ is the number of samples in the calibration, validation, or test sets.

The 'predicted ($\hat{\mathbf{c}}$) *vs.* reference ($\mathbf{c}$)' plot is constructed for model performance evaluation. The additional measures of modeling quality that arise from this plot are slope, offset, and squared correlation coefficient $R^2$. They characterize the calibration performance and thus help to compare various methods.

In the MCR method, we are also interested in the quality of modeling of the spectral matrix $\mathbf{X}$. For this purpose we use the residual sum of squares (RSS) at calibration (RSSC), at validation (RSSV), and at prediction (RSSP), calculated by a formula

$$\mathrm{RSS} = \sum_{i=1}^{I} \sum_{j=1}^{J} (x_{ij} - \hat{x}_{ij})^2 / I / J \quad (20)$$

where $x_{ij}$ are the elements of the spectral matrix $\hat{\mathbf{X}}$, $\hat{x}_{ij}$ are the elements of matrix $\hat{\mathbf{X}} = F(\mathbf{C}\mathbf{S}^{\mathrm{t}})$ that approximates $\mathbf{X}$, $J$ is the number of wavelengths, and $I$ is the number of samples in the calibration, validation, or prediction.

Figures of merit are calculated separately for each method and each data set.

### 2.4. Computing

Non-linear optimization in eqn (14) can be a complex task, as the larger the number of parameters to be fitted the more difficult the task. Especially it concerns the S-step, where the size of sought spectra ($J$) can be very large. Therefore, it may be advantageous to parameterize vector $\mathbf{S}$. We use a linear combination of Gaussian peaks, so the dimension of the search space is equal to $M \cdot G$, where $M$ is the number of peaks, and $G$ is the number of peak parameters. The similar techniques were used in ref. 15–17.

All calculations are made using Microsoft Excel software. For optimization in the NL-MCR method, the standard Solver Add-in is used. Chemometric algorithms (*e.g.* PCA) are performed using Chemometrics Add-in. The details are explained in ref. 20. The typical time of calculations is about 30 s with 3.40 GHz CPU.

# 3. Experimental

## 3.1. Sample preparation

Stock solutions of 50.00 g $L^{-1}$ of cerium(III), neodymium(III), and praseodymium(III) are prepared using nitrate hexahydrate oxides $Nd(NO_3)_3 \cdot 6H_2O$, $Pr(NO_3)_3 \cdot 6H_2O$, and $Ce(NO_3)_3 \cdot 6H_2O$ provided by Merck (extra pure), and 0.7 M $HNO_3$ provided by Merck (MSD, Moscow, Russia). Working solutions for the calibration and validation data sets are prepared by aliquot dilution and mixing of these stock solutions. Samples for the test set have the same concentration of the REEs dissolved in nitric acid of various concentrations. The 4.0 M $HNO_3$ is made by diluting 33 ml of $HNO_3$ in 100 ml of water using a volumetric flask; other nitric acid solutions are prepared diluting this solution further. Double-distilled water is used. The test set samples are prepared using accurate weighing of each REE, which are then dissolved in the appropriate $HNO_3$ solution.

## 3.2. Instrumentation and experimental measurements

The samples are subjected to UV-vis spectroscopy in the transmittance mode, and then converted to the absorbance units. Spectra are collected with a wavelength increment of 1 nm among consecutive measurements over the range of 200–1000 nm using spectrophotometer UNICO SQ-2800 (Unico, Dayton, NJ) with the photometric range of 0.01–3 AU. The acquisition time is eight minutes. A 10 mm path length quartz cuvette is used.

## 3.3. Data subsets

Concentrations of cerium between 0 and 5 g $L^{-1}$ and nitric acid, 0.7 M, were defined by the customer. The goal of the study is to investigate feasibility to develop a calibration model within this range and to verify the model stability with respect to variation in concentrations. The design of experiments and detailed description of the data are presented in ref. 5. The whole concentration ranges vary for cerium from 0 to 5 g $L^{-1}$, for neodymium from 0 to 30 g $L^{-1}$ and for praseodymium from 0 to 15 g $L^{-1}$. Concentrations of Ne and Pr are not indicated in Table 1 because, as shown in ref. 5, the latter REEs have no peaks in the UV range and do not affect the quantitative determination of Ce and $HNO_3$. The whole dataset is shown in Table 4. It can be seen (Table 1) that both the calibration, 12 samples, and validation, 14 samples, subsets contain samples with different cerium concentrations dissolved in the same $HNO_3$ solvent, 0.7 M. On the contrary, the test subset includes samples that have the same cerium concentration, but the acid concentration varies. This helps us to evaluate the stability against alternating nitric acid concentration.

For quantification of Ce and $HNO_3$ only the range of 220–350 nm is used. The range of 200–210 nm is very noisy and, therefore, not used in calibration. Ce and $HNO_3$ have no peaks in the range of 360–1000 nm, so the latter range is also excluded from consideration. The acquired UV spectra are preprocessed using the baseline offset method (adjusting the data to the minimum point in the data) and the Savitzky–Golay smoothing with seven points and a polynomial order of two. The results are collected in the first dataset labeled in what follows as $B^+$. It contains the spectra of mixtures of Ce and $HNO_3$. To develop the second dataset, the background spectra of $HNO_3$ at concentrations of 0.1 M, 0.4 M, 0.7 M, 1.8 M, 3.0 M, and 4.0 M are acquired, preprocessed as explained above, and then subtracted from the sample spectra with corresponding nitric acid concentration. The resulting dataset is further designated as $B^-$. For illustration purposes, the validation set spectra for datasets $B^+$ and $B^-$ are shown in Fig. 2. The non-linear effects in spectra are emphasized by the dashed lines. They demonstrate the evolution of the peak's position, which shifts for the reasons explained in the Introduction.

It should be noted that the removal or retention of the nitric acid background has a special meaning for the interpretation of the test subset prediction. Subtracting the background (data $B^-$), we assume that the $HNO_3$ concentration in a new sample is known by some other means. In the opposite case (data $B^+$), we are free from this impractical proposition. At the same time, when working with the designed training and validation subsets, we have full knowledge about the nitric acid concentration that always equals 0.7 M.

It is evident that the obtained spectra have flattened (saturated) peaks both in the mixtures and in the nitric acid background spectrum. The saturation level is about 3 AU. Using designation $\{\cdot\}$ introduced in eqn (13), the obtained datasets can be presented using the following mnemonic notations: $B^+ = \{Ce + HNO_3\}$, and $B^- = \{Ce + HNO_3\} - \{HNO_3\}$.

Table 1 Samples used for analysis

| | Calibration set | | | Validation set | | | Test set | | |
|---|---|---|---|---|---|---|---|---|---|
| No. | $HNO_3$ (M) | Ce (g $L^{-1}$) | Number of samples | $HNO_3$ (M) | Ce (g $L^{-1}$) | Number of samples | $HNO_3$ (M) | Ce (g $L^{-1}$) | Number of samples |
| 1 | 0.7 | $1 \times 10^{-4}$ | 1 | 0.7 | 0 | 5 | 0.1 | 0.3 | 1 |
| 2 | 0.7 | 0.03 | 3 | 0.7 | 0.03 | 1 | 0.4 | 0.3 | 1 |
| 3 | 0.7 | 0.3 | 2 | 0.7 | 0.3 | 1 | 1.8 | 0.3 | 1 |
| 4 | 0.7 | 0.6 | 3 | 0.7 | 0.5 | 2 | 3 | 0.3 | 1 |
| 5 | 0.7 | 2 | 1 | 0.7 | 0.6 | 1 | 4 | 0.3 | 1 |
| 6 | 0.7 | 5 | 2 | 0.7 | 2 | 3 | | | |
| 7 | | | | 0.7 | 5 | 1 | | | |

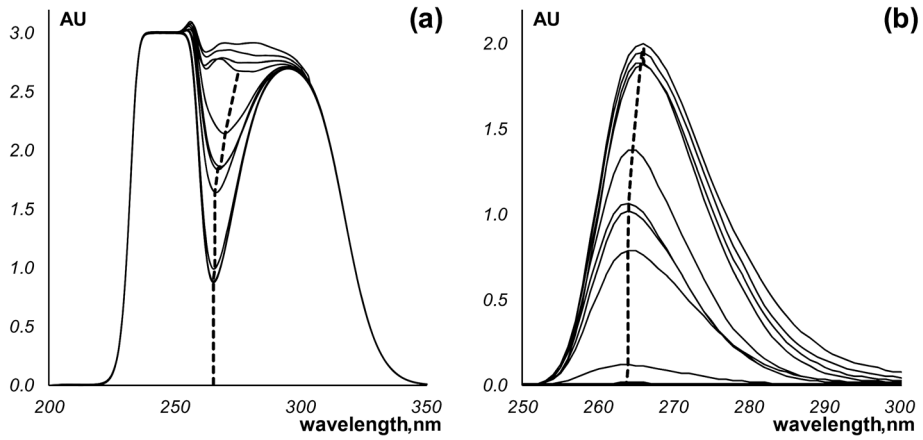


Fig. 2 Validation spectra: (a) data set $B^+$ with the $HNO_3$ background; (b) data set $B^-$ without the $HNO_3$ background. The dashed lines show the peak shifting due to non-linear effects.

## 4. Results and discussion

In this section, we demonstrate how to develop calibration models by means of the methods described above. Two data sets: $B^+$, where nitric acid background spectra are preserved, and $B^-$, which contains spectra without nitric acid background are considered in parallel.

The following methodological principles are applied in model development using each of the methods. The calibration (training) subset is used to estimate the model's parameters **B** that are given in eqn (2). The validation subset is employed for the assessment of model complexity that is the number of components. Then, the final models are applied to the test set data allowing us to make conclusions about the stability of each method.

### 4.1. MCR-ALS

In the calibration/validation subsets, the $B^+$ data differ from $B^-$ data by the background 0.7 M $HNO_3$ spectrum, which is common for all samples, *i.e.* $B^+ = B^- + \{0.7\ M\ HNO_3\}$. Therefore, it is apparent that mean-centered PCA applied to the calibration subsets $B^+$ and $B^-$ yields identical results. The first PC explains 97.4% of the total variance; the next two PCs add 2.3% and 0.3%, respectively. Therefore, it is reasonable to utilize the MCR procedure with one or two components. The results of the one-component case, presented in ref. 5, were worse. To resolve the non-linearity we introduce virtual components in the MCR model. In some cases such a technique helps to improve the calibration result.[18,19]

Now we consider two virtual components of cerium, *i.e.* Ce1 and Ce2. Since these components are not actual chemical constituents, the non-negativity constraint is applied to the sum $\mathbf{c}_{Ce1} + \mathbf{c}_{Ce2}$ only, but not to the virtual spectra and concentrations. The correlation constraint is presented by a regression $a_1\mathbf{c}_{Ce1} + a_2\mathbf{c}_{Ce2} + g = \mathbf{c}_{ref}$.

In the case of $B^+$ data, we should account for one more component that is nitric acid. This can be accomplished in two ways. First, we can include nitric acid in matrices **C** and **S** (see eqn (4)) as the third component. However, it is subject to a strict limitation: its spectrum is equal to the known $HNO_3$ background spectrum, and its concentration always equals 0.7 M. The second option is to perform the two-component MCR procedure, adding the $HNO_3$ background spectrum at the end of each S-step. Computations show that the first way of the MCR calculation is very similar to the second one. Namely, the obtained results differ in the third digit only. The latter method is eventually equivalent to the MCR modeling of the $B^-$ data set. Thus, we can conclude that in all possible methods of the MCR-ALS modeling, the $B^+$ and $B^-$ datasets provide the same results. The MCR-ALS results for the validation subset are shown in Fig. 3 by the red square marks.

### 4.2. NL-MCR

Before applying the NL-MCR method, we should select a non-linear filter $F$, introduced in eqn (12). In our case, the choice is evident. For the $B^+$ dataset it is

$$F = \left\{ \mathbf{c}_{Ce}\mathbf{s}_{Ce}^{t} + \mathbf{c}_{HNO_3}\mathbf{s}_{HNO_3}^{t} \right\}_{s,p}, \qquad (21)$$

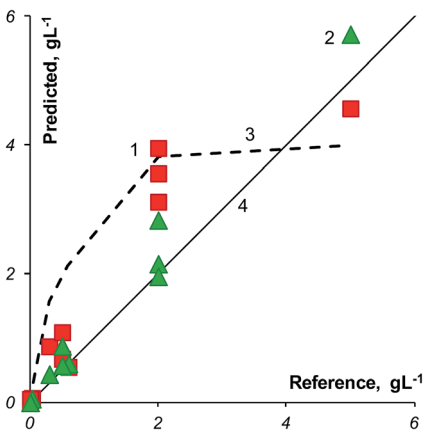


Fig. 3 Validation of cerium. Squares (1, ■) stand for MCR-ALS, triangles (2, ▲) represent NL-MCR-ALS, the dashed curve (3, arbitrary units) demonstrates the peak height evolution, and the line (4) is the target.

Table 2 Results of calibration and validation of Ce by MCR-ALS (2 Comp.) and NL-MCR (1 Comp.)

| | MCR-ALS | NL-MCR |
|---|---|---|
| **Calibration** | | |
| RMSEC | 0.36 | 0.25 |
| Slope | 0.95 | 0.98 |
| Offset | 0.06 | 0.02 |
| $R^2$ | 0.96 | 0.98 |
| RE | 17% | 12% |
| RSSC | $2.2 \times 10^{-4}$ | $5.3 \times 10^{-4}$ |
| **Validation** | | |
| RMSEV | 0.78 | 0.32 |
| Slope | 1.09 | 1.14 |
| Offset | 0.32 | 0.03 |
| $R^2$ | 0.84 | 0.98 |
| RE | 47% | 19% |
| RSSV | $1.7 \times 10^{-4}$ | $5.7 \times 10^{-4}$ |

and for the $B^-$ dataset it is

$$F = \left\{ \mathbf{c}_{Ce}\mathbf{s}_{Ce}^{t} + \mathbf{c}_{HNO_3}\mathbf{s}_{HNO_3}^{t} \right\}_{s,p} - \left\{ \mathbf{c}_{HNO_3}\mathbf{s}_{HNO_3}^{t} \right\}_{s,p}. \quad (22)$$

In these equations, the operation $\{\cdot\}_{s,p}$ is defined by eqn (13). The saturation level $s = 3$ is evident from the data (see Fig. 2a) and is determined by the instrument employed. The shape parameter $p$ is also determined by the instrument in use. Therefore, based on the results obtained in ref. 9, the parameter $p$ is set to 1. In general, parameters $s$ and $p$ can be optimized for the minimum of RMSEV at the validation stage. However, since the optimized parameters, $s = 2.98$ and $p = 1.02$, are close to the above-mentioned values, in further calculations we use $s = 3$ and $p = 1$.

The nitric acid component is known in advance. The concentration vector $\mathbf{c}_{HNO_3}$ is given by the design of the calibration subset (see Table 1), and the spectral vector $\mathbf{s}_{HNO_3}$ was determined earlier in ref. 9. It is important to note that in the NL-MCR method we use the ideal spectrum of nitric acid,[9] whereas in the MCR-ALS method (described in the previous section) we utilize the saturated spectrum of nitric acid $\{HNO_3\}$.

Technically, concentrations of cerium and nitric acid can be determined simultaneously in one experiment. For this purpose, a calibration subset should be designed with different $HNO_3$ concentrations. In our case, the constant concentration of $HNO_3$ in the training set did not allow us to do so.

The cerium concentrations and spectra in eqn (21) and (22) are unknown, so vectors $\mathbf{c}_{Ce}$ and $\mathbf{s}_{Ce}$ can be found by solving the problem given in eqn (14). The dimensionality of vector $\mathbf{c}_{Ce}$ is 12, which corresponds to the number of samples in the calibration subset. Concentrations are computed at the C-step in order to find a minimum in eqn (15). Subsequently they are corrected to follow the correlation constraint given by regression $a\mathbf{c}_{Ce} + g = \mathbf{c}_{ref}$, and the non-negativity restriction.

The dimensionality of spectral vector $\mathbf{s}_{Ce}$ is 151, which equals the number of wavelengths. This is a rather large number, hence parameterized optimization seems most appropriate in this case. We use an asymmetrical Gaussian peak to model the cerium spectrum $\mathbf{s}_{Ce}$, *i.e.*

$$\mathbf{s}_{Ce} = h \exp\left[ -\left( \frac{\lambda - m}{\sigma} \right)^2 \right], \quad \sigma = \begin{cases} l_1, & \lambda \le m \\ l_2, & \lambda > m \end{cases}. \quad (23)$$

Here $\lambda$ is the wavelength. The shape parameters $h$, $m$, $l_1$ and $l_2$ are sought in the S-step (see eqn. (16)) of the NL-MCR procedure. The non-negativity restriction applied earlier to the spectra is not necessary for the functional form of the cerium spectrum defined by eqn (23).

The NL-MCR models developed for the $B^+$ and $B^-$ datasets are very similar. For example, for the $B^+$ data: RMSEC = 0.2599 and RMSEV = 0.3178, and for the $B^-$ data: RMSEC = 0.2488 and RMSEV = 0.3181. For this reason, we report the averaged results of determination that are presented in Fig. 3 by the green triangles.

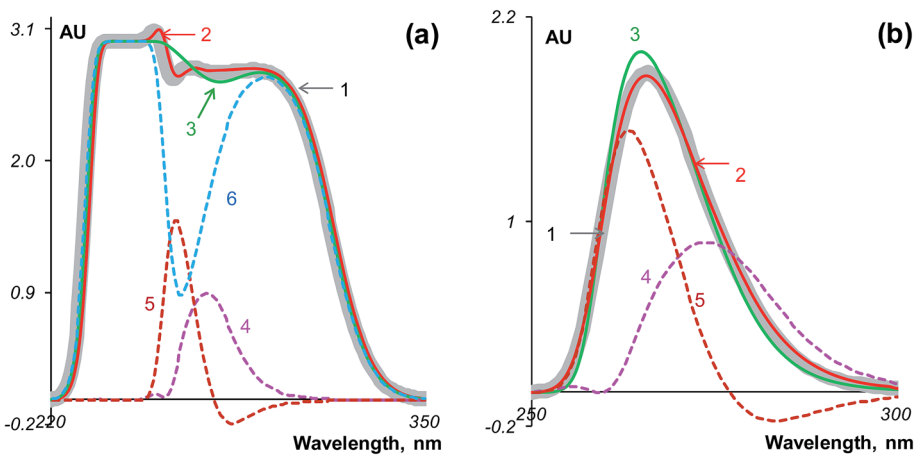


Fig. 4 Fitting of validation sample #6 (2 g $L^{-1}$ Ce) in datasets $B^+$ (a), and $B^-$ (b). In both plots: the wide grey strip (1) represents the data spectrum, the red line (2) stands for the conventional MCR solution, the green line (3) demonstrates the NL-MCR result, and dashed curves (4) and (5) represent virtual cerium components. In plot (a) the dashed curve (6) represents the saturated $HNO_3$ spectrum.

Table 3 Results of the prediction of Ce in the test set by MCR-ALS (2 components) and NL-MCR (1 component)

| | MCR-ALS | NL-MCR |
|---|---|---|
| $B^+$ | | |
| RMSEP | 1.99 | 0.05 |
| RE | 660% | 16% |
| RSSP | $218155 \times 10^{-4}$ | $0.4 \times 10^{-4}$ |
| $B^-$ | | |
| RMSEP | 0.25 | 0.05 |
| RE | 84% | 16% |
| RSSP | $61 \times 10^{-4}$ | $0.5 \times 10^{-4}$ |

### 4.3. Quality of the calibration models

A summary of the cerium quantification is presented in Table 2.

The conventional model constructed with the help of the MCR-ALS method demonstrates poor quality caused by the obvious non-linearity. This effect can be seen on the predicted *vs.* reference plot (Fig. 3), where the MCR-ALS prediction (red squares) is similar to the peak height evolution (Fig. 2, dashed curve 3). Non-linearity also manifests itself *via* the large offset at the validation stage and through a large difference between the RMSEC and RMSEV values. It is interesting that MCR-ALS demonstrates superiority when modeling spectral data. This can be noticed from the values in rows titled RSS in Table 2. It is clear that such quality of spectral modeling is achieved by increasing the model complexity, *i.e.* the number of virtual components in MCR-ALS. The role of virtual components Ce1 and Ce2 is demonstrated by Fig. 4.

On the other hand, the NL-MCR method demonstrates a lower quality in the spectral modeling (by a factor 6), but doubles cerium calibration and validation accuracy. The NL-MCR method also demonstrates a clear linear trend in Fig. 3 and lower offsets.

### 4.4. The test set

**Stability assessment.** In this section, we present and discuss the outcomes obtained by application of the above calibration models to the test set data. As discussed earlier, for practical implementation it is important to assess the performance of the models in out-of-control cases. For this purpose we use a special test set, in which the samples have different concentrations of $HNO_3$ (see the last column of Table 1) but identical concentrations of cerium (0.3 g $L^{-1}$).

In contrast to the calibration/validation stage of modeling, the test set prediction drastically depends on the nitric acid background treatment. The data sets $B^+$ (the background is preserved) and $B^-$ (the background is subtracted) deliver very different results when using the MCR-ALS models. At the same time, the NL-MCR model demonstrates comparable outcomes regardless of the data type used. The overall results of prediction are summarized in Table 3 and Fig. 5. The details for each model are discussed below.

As shown in Table 3 and by the red squares (1) in Fig. 5, the MCR-ALS models provide poor results for both test sets $B^+$ and $B^-$. For example, for the last sample (4 M $HNO_3$) in test set $B^+$, the MCR-ALS model predicts the cerium concentration to be 3.12 g $L^{-1}$, instead of 0.3 g $L^{-1}$. It must be noted that for MCR-ALS, the $B^-$ model is substantially better than the $B^+$ model. However, an 84% relative error of prediction is not a result that can be used in practice. Such a poor result can be explained by the fact that MCR-ALS is designed to work under the condition of bi-linearity that is severely violated for the test set data. It is interesting that we observe a negative correlation ($R = -0.954$) between $B^+$ and $B^-$ prediction outcomes in MCR-ALS.

The NL-MCR results are of another kind – see Table 3 and Fig. 5, where these outcomes are shown by the green triangles (3). Both test sets, $B^+$ and $B^-$, are predicted with a similar accuracy that is also well suited for practical needs.

Moreover, $B^+$ data (the $HNO_3$ background is preserved) offer an additional opportunity to estimate the nitric acid concentration in each test sample. For this purpose, we extend the **C** matrix in eqn (17) by adding a set of 5 nitric acid concentrations, *i.e.* $\mathbf{C}_{new} = (\mathbf{C}_{Ce}, \mathbf{C}_{HNO_3})$ which are predicted together with cerium concentrations. The results are very satisfactory, and an RMSEP = 0.08 confirms this. The obtained outcome means that we can omit the impractical assumption that concentrations of

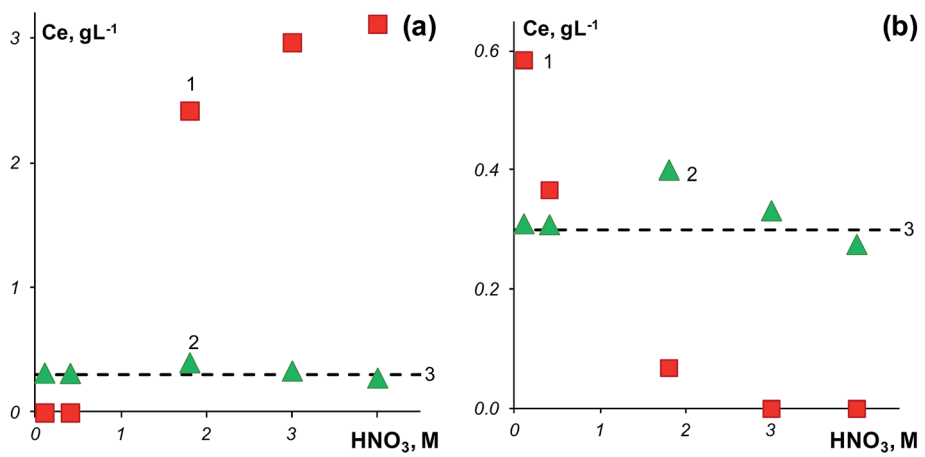


Fig. 5 Prediction of the test set samples. Datasets $B^+$ (a), and $B^-$ (b). In both plots: squares (1, ■) stand for MCR-ALS, triangles (2, ▲) represent NL-MCR, and the dashed line (3) demonstrates the reference value of Ce (0.3 g $L^{-1}$).

Table 4 Complete dataset used for analysis

| | Calibration set | | | | Validation set | | | | Test set | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $HNO_3$ (M) | Ce (g $L^{-1}$) | Nd (g $L^{-1}$) | Pr (g $L^{-1}$) | $HNO_3$ (M) | Ce (g $L^{-1}$) | Nd (g $L^{-1}$) | Pr (g $L^{-1}$) | $HNO_3$ (M) | Ce (g $L^{-1}$) | Nd (g $L^{-1}$) | Pr (g $L^{-1}$) |
| 1 | 0.7 | 0.60 | 0.60 | 10.00 | 0.7 | 0.60 | 12.00 | 10.00 | 0.1 | 0.30 | 2.00 | 3.00 |
| 2 | 0.7 | 0.60 | 12.00 | 0.30 | 0.7 | 0.03 | 0.60 | 0.30 | 0.4 | 0.30 | 2.00 | 3.00 |
| 3 | 0.7 | 0.60 | 0.60 | 0.30 | 0.7 | 0.30 | 2.00 | 3.00 | 1.8 | 0.30 | 2.00 | 3.00 |
| 4 | 0.7 | 0.03 | 12.00 | 10.00 | 0.7 | 0.50 | 10.00 | 8.00 | 3.0 | 0.30 | 2.00 | 3.00 |
| 5 | 0.7 | 0.03 | 12.00 | 0.30 | 0.7 | 2.00 | 3.50 | 1.00 | 4.0 | 0.30 | 2.00 | 3.00 |
| 6 | 0.7 | 0.03 | 0.60 | 10.00 | 0.7 | 0.00 | 0.04 | 0.01 | | | | |
| 7 | 0.7 | 0.30 | 2.00 | 3.00 | 0.7 | 5.00 | 30.00 | 1.00 | | | | |
| 8 | 0.7 | 0.30 | 2.00 | 3.00 | 0.7 | 2.00 | 30.00 | 1.00 | | | | |
| 9 | 0.7 | 5.00 | 5.00 | 15.00 | 0.7 | 2.00 | 0.00 | 0.00 | | | | |
| 10 | 0.7 | 2.00 | 5.00 | 15.00 | 0.7 | 0.50 | 0.00 | 0.00 | | | | |
| 11 | 0.7 | 5.00 | 0.30 | 5.00 | 0.7 | 0.00 | 0.00 | 1.00 | | | | |
| 12 | 0.7 | $3 \times 10^{-4}$ | $4 \times 10^{-4}$ | $1 \times 10^{-4}$ | 0.7 | 0.00 | 0.00 | 0.25 | | | | |
| 13 | | | | | 0.7 | 0.00 | 3.50 | 0.00 | | | | |
| 14 | | | | | 0.7 | 0.00 | 0.875 | 0.00 | | | | |

nitric acid in new samples are known, and find these concentrations in-line in the same manner as we quantify cerium concentrations.

The NL-MCR method provides us with another interesting result being a pure cerium spectrum. This spectrum is presented by the asymmetrical Gaussian peak given in eqn (23). The shape parameters are estimated at the calibration stage of the NL-MCR procedure. Both calibration sets, $B^+$ and $B^-$, yield very similar results. Our estimation for the peak height is 453 $M^{-1}$ $cm^{-1}$ at $\lambda = 251$ nm. The extinction coefficient of cerium in sulfuric acid given in ref. 6 is 1.5 times greater. Due to a well-known scaling ambiguity,[13] the pure spectra found by MCR should be treated with care. However, we were not concerned with the cerium pure spectra evaluation, but aimed to solve calibration problems with an appropriate accuracy.

## 5. Conclusions

We consider that the presented study can be summarized in two types of conclusions. The practical one confirming that the non-linear multivariate curve resolution method is an appropriate tool for the on-line analysis of cerium in aqueous solutions of nitric acid. Both of the considered NL-MCR models, the first one where the nitric acid background spectra are preserved, and the second one, which uses the spectra without the nitric acid background, are stable to some extent for out-of-control cases. The accuracy of prediction, which is about 15%, looks suitable for practical implementation in PAT solutions. An additional advantage of the first NL-MCR model is the opportunity to assess nitric acid concentration in new samples with the accuracy of about 1%. This result implies that the developed model can be employed even for unstable manufacturing technology, when the concentration of nitric acid varies to a large degree. The possibility to assess nitric acid concentration in solutions of REEs has a particular practical importance.

The conventional calibration models, such as MCR-ALS, provide very bad results. This can be explained by the fact that these models are designed to work under the condition of bi-linearity that is severely violated in the case of nitric acid solution of cerium.

The more general conclusion is that optimization problems given in eqn (11) and (14) have an essential distinction. The former is bi-linear, but the latter is nonlinear regarding the matrices **S** and **C**. Here we can mark that the NL-MCR approach and the hard MCR modeling[15] have some similarity because both methods use the non-linear optimization. However, the hard MCR resolves non-linearity *inside* the bi-linear term $\mathbf{CS}^t$, while NL-MCR has non-linearity *outside* this term. In this sense, eqn (14) can be considered as a shelled optimization problem, which is still bi-linear inside the shell *F*. We suppose that the impairing modeling of multilinear data may result in a series of new methods such as 'shelled' PCA, PLS, *etc.*

## References


1 A. L. Pomerantsev and O. Y. Rodionova, *J. Chemometrics*, 2012, **26**, 299–310.
2 S. A. Bryan, T. G. Levitskaia, A. M. Johnsen, C. R. Orto and J. M. Peterson, *Radiochim. Acta*, 2011, **99**, 563–571.
3 D. Kirsanov, V. Babain, M. Agafonova-Moroz, A. Lumpov and A. Legin, *Radiochim. Acta*, 2012, **100**, 185–188.
4 I. D. Nickson, C. Boxall, A. Jackson and G. O. H. Whillock, *IOP Conf. Ser.: Mater. Sci. Eng.*, 2010, **9**, 012011.
5 O. Y. Rodionova, T. I. Tikhomirova and A. L. Pomerantsev, *Anal. Chim. Acta*, 2015, **869**, 59–67.
6 H. L. Greenhaus, A. M. Feibush and L. Gordon, *Anal. Chem.*, 1957, **29**, 1531–1534.
7 C. K. Jørgensen and J. S. Brinen, *Mol. Phys.*, 1963, **6**, 629–631.
8 L.-H. Chou, W. E. Cleland Jr. and C. L. Hussey, *Inorg. Chem.*, 2012, **51**, 11450–11457.
9 A. L. Pomerantsev, Y. V. Zontov and O. Y. Rodionova, *J. Chemom.*, 2014, **28**, 740–748.
10 R. R. de Oliveira, K. M. G. de Lima, R. Tauler and A. de Juan, *Talanta*, 2014, **125**, 233–241.

11 M. C. Antunes, J. E. J. Simao, A. C. Duarte and R. Tauler, *Analyst*, 2002, **127**, 809–817.
12 A. de Juan and R. Tauler, *Crit. Rev. Anal. Chem.*, 2006, **36**, 163–176.
13 A. de Juan, S. C. Rutan, M. Maeder, R. Tauler. Multivariate Curve Resolution Chapters in, *Comprehensive Chemometrics*, ed. D. Brown, R. Tauler and B. Walczak, Elsevier: Amsterdam, 2009, vol. 2, pp.325–344, 473-506.
14 R. A. Harshman, S. Hong and M. E. Lundy, *J. Chemom.*, 2003, **17**, 363–378.
15 P. Jandanklang, M. Maeder and A. C. Whitson, *J. Chemom.*, 2001, **15**, 511–522.
16 J. M. D. Cruz, J. Sanch, E. Chekmeneva, C. Arino and M. Esteban, *Analyst*, 2010, **135**, 1653–1662.
17 H. Keles, A. Naylor, F. Clegga and C. Sammon, *Analyst*, 2014, **139**, 2355–2369.
18 R. L. Carneiro, J. W. B. Braga, R. J. Poppia and R. Tauler, *Analyst*, 2008, **133**, 774–783.
19 V. Pomareda, A. V. Guamán, M. Mohammadnejad, D. Calvo, A. Pardo and S. Marco, *Chemom. Intell. Lab. Syst.*, 2012, **118**, 219–229.
20 A. L. Pomerantsev, *Chemometrics in Excel*, John Wiley & Sons: Hoboken NJ, 2014.