Analytica Chimica Acta 811 (2014) 13–22

Contents lists available at ScienceDirect

Analytica Chimica Acta

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

The Successive Projections Algorithm for interval selection in trilinear partial least-squares with residual bilinearization

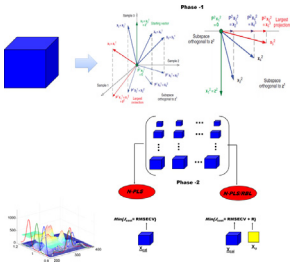
Adriano de Araújo Gomes [a], Mirta Raquel Alcaraz [b], Hector C. Goicoechea [a,b,*], Mario Cesar U. Araújo [a,**]

[a] Laboratório de Automação e Instrumentação em Química Analítica e Quimiometria (LAQA), Universidade Federal da Paraíba, CCEN, Departamento de Química, Caixa Postal 5093, CEP 58051-970, João Pessoa, PB, Brazil
[b] Laboratorio de Desarrollo Analítico y Quimiometría (LADAQ), Cátedra de Química Analítica I, Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral-CONICET, Ciudad Universitaria, 3000 Santa Fe, Argentina

HIGHLIGHTS

- The Successive Projections Algorithm (SPA) is proposed to interval selection to multiway data.
- Discarding of non-informative variables using SPA was proposed for trilinear PLS.
- The feasibility of the proposed method was demonstrated using simulated study and two sets of real data.
- The iSPA-N-PLS models provide better predictions as compared to N-PLS full model and GA-N-PLS.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 21 September 2013
Received in revised form
12 December 2013
Accepted 18 December 2013
Available online 27 December 2013

Keywords:
Multiway data
Variable selection
Second order calibration
Second order advantage

ABSTRACT

In this work the Successive Projection Algorithm is presented for intervals selection in N-PLS for three-way data modeling. The proposed algorithm combines noise-reduction properties of PLS with the possibility of discarding uninformative variables in SPA. In addition, second-order advantage can be achieved by the residual bilinearization (RBL) procedure when an unexpected constituent is present in a test sample. For this purpose, SPA was modified in order to select intervals for use in trilinear PLS. The ability of the proposed algorithm, namely *i*SPA-N-PLS, was evaluated on one simulated and two experimental data sets, comparing the results to those obtained by N-PLS. In the simulated system, two analytes were quantitated in two test sets, with and without unexpected constituent. In the first experimental system, the determination of the four fluorophores (L-phenylalanine; L-3,4-dihydroxyphenylalanine; 1,4-dihydroxybenzene and L-tryptophan) was conducted with excitation-emission data matrices. In the second experimental system, quantitation of ofloxacin was performed in water samples containing two other uncalibrated quinolones (ciprofloxacin and danofloxacin) by high performance liquid chromatography with UV–vis diode array detector. For comparison purpose, a GA algorithm coupled with N-PLS/RBL was also used in this work. In most of the studied cases *i*SPA-N-PLS proved to be a promising tool for selection of variables in second-order calibration, generating models with smaller RMSEP, when compared to both the global model using all of the sensors in two dimensions and GA-NPLS/RBL.

* Corresponding author at: Laboratorio de Desarrollo Analítico y Quimiometría (LADAQ), Cátedra de Química Analítica I, Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral-CONICET, Ciudad Universitaria, 3000 Santa Fe, Argentina. Tel.: +54 342 4575206x190; fax: +54 342 4575205.
** Corresponding author. Tel.: +55 83 3216 7438; fax: +55 83 3216 7437.
E-mail addresses: hgoico@fbcb.unl.edu.ar (H.C. Goicoechea), laqa@quimica.ufpb.br, mariougulino@gmail.com (M.C.U. Araújo).

0003-2670/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.aca.2013.12.022

1. Introduction

In recent years, the dramatic advances in analytical instrumentation allow obtaining a large amount of data per sample in a short time interval. Hyphenated analytical techniques like HPLC-DAD/MS, HPLC–MS-MS, GC–GC–TOFMS can be cited as good examples [1–3]. However, sometimes the large amount of acquired data would not be useful to build calibration models, and part of the recorded signal may be strongly correlated [4,5]. Moreover, some sensors may not provide relevant information for the system under consideration, and in the latter case their use may even compromise the model performance in terms of precision and accuracy [6].

In this context, variable selection as a previous stage in the construction of calibration models can be a tool to improve the predictive capability and robustness, eliminating non-informative variables [6–8]. In the context of first-order calibration, variable selection has been successfully employed building models for a variety of analytical techniques, for example, near-infrared spectroscopy [9,10], UV–vis [11] and plasma emission spectrometry [12]. Several strategies were reported in the literature, such as genetic algorithm (GA) [13], colony of ants (CA) [14], tabu search (TS) [15], simulated annealing (SA) [16] and many others [17–19]. These methods of selection of variables are usually coupled to regression models like multiple linear regressions (MLR) and partial least squares (PLS) [20,21].

Since the past decade, the application of multiway methods [22] has become increasingly important, demonstrating its importance as chemometric tool. Rapid quantitation of several chemical species in complex matrices with minor chemical treatment of the sample has been reported in the literature [23–25]. However, in the context of multiway calibration, variable selection techniques have been scarcely explored. To the best of our knowledge only four papers have been published in this regard: (a) Wu and collaborators presented a method using GA for variable selection together with data modeling by parallel factor analysis (PARAFAC) [26]; (b) Lopes and Menezes applied GA in variable selection to model the performance of industrial fed-batch fermentation process by trilinear PLS [27]; (c) Gourvénec et al. proposed the use GA for variable selection in multivariate curve resolution (MCR) to improve the quality of the on-line monitoring of batch processes [28]; and (d) Carneiro et al. developed a method based on GA for variable selection for a bilinear least squares (BLLS) whit residual bilinearization (RBL) procedure to determination of pesticides and metabolites in wine [29]. Interestingly, the process of variable selection led to improved results in all the mentioned works.

In the field of variable selection, Araújo et al., proposed the Successive Projections Algorithm (SPA) [30,31]. This algorithm was initially implemented as a tool for variable selection in MLR. Afterward, it was modified to work in several areas of chemometrics such as variable selection combined with Uninformative Variables Elimination (UVE) [32], classification [33], calibration transfer [34], sample selection [35] and selection of intervals for PLS regression [36]. Several reports of successful applications of SPA can be found in the literature [37].

The present paper proposes a modification of SPA for the selection of intervals of variables to be used in a trilinear PLS model. The proposed algorithm, namely *i*SPA-N-PLS, combines the properties of trilinear PLS with the possibility of discarding non-informative variables in SPA and the second-order advantage, which is achieved by RBL procedure when unexpected constituents occur in a test sample. In order to evaluate the ability of the new method, one simulated and two experimental data sets were studied. In the simulated system, two analytes were quantitated in two test sets, with and without unexpected constituent. In the first experimental system, the determination of the four fluorophores (L-phenylalanine, L-3,4-dihydroxyphenylalanine, 1,4-dihydroxybenzene and L-tryptophan) was conducted with excitation-emission data. In the second experimental system, quantitation of ofloxacin was performed in water samples containing two other uncalibrated quinolones (ciprofloxacin and danofloxacin) by high performance liquid chromatography with UV–vis diode array detector.

2. Background and theory

2.1. Notation

In what follows, three-way arrays, matrices, vectors and scalars will be denoted by bold capital letters underlined, bold capital letters, bold lowercase letters and italic characters, respectively. The *T* superscript indicates the transpose of a vector or matrix.

2.2. Trilinear PLS

The trilinear PLS or more generally N-PLS, has been proposed by Bro [38] in 1996, although other studies have previously reported the use PLS for trilinear data [39,40]. N-PLS has been proposed as an alternative to the U-PLS (unfolded-PLS), wherein the three-dimensional data structure is unfolded in a two-dimensional matrix [38–40]. When considering the structure of the trilinear data, N-PLS models present more stability and less complexity when compared to U-PLS models [38]. In the other hand, from the point of view of computational effort, outperforms PARAFAC, since N-PLS is based on solving a problem of eigenvectors [38].

Interestingly, the algorithm proposed by Bro in essence is no different of PLS1, in both the decomposition of the independent variables are driven to maximize the covariance between **y** (dependent variable vector) and scores, but unlike the bilinear PLS1, the trilinear PLS1 decompose three-dimensional data arrays ($\underline{\mathbf{X}}_{i\times j\times k}$) in a set of triads. Each triad is comprised of a score vector **t** and two loading vectors, $\mathbf{w}^j$ and $\mathbf{w}^k$, which are the weights in the dimensions *j* and *k*, respectively. A triad can be expressed as shown in Eq. (1):

$$x_{ijk} = t_i w_j^j w_k^k \tag{1}$$

As in the bilinear PLS, in the trilinear model $\mathbf{w}^j$ and $\mathbf{w}^k$ are searched minimizing the squared residuals according to Eq. (2):

$$e^2 = (x_{ijk} - t_i w_j^j w_k^k) \tag{2}$$

The solution by the method of least squares is given by Eq. (3):

$$t_i = \sum_{j=1}^{j}\sum_{k=1}^{k}(z_{jk} w_j^j w_k^k) \tag{3}$$

where z_{jk} are the elements of a **Z** matrix of dimension $(j \times k)$ corresponding to the sum of each of the *i* matrix elements that comprise the three-dimensional data array ($\underline{\mathbf{X}}_{i\times j\times k}$), multiplied by the concentration of the analyte [41], as shown in Eq. (4).

$$\mathbf{Z} = \mathbf{X}_1 y_1 + \mathbf{X}_2 y_2 + \mathbf{X}_3 y_3 + \cdots + \mathbf{X}_I y_I \tag{4}$$

The next step is to determine $\mathbf{w}^j$ and $\mathbf{w}^k$, which is easily achieved by singular value decomposition (SVD) of the matrix **Z**, being **t** estimated using Eq. (3). In the next stage, the regression coefficient vector **v** is calculated employing Eq. (5).

$$\mathbf{v} = (\mathbf{t}^T\mathbf{t})^{-1}\mathbf{t}^T\mathbf{y} \tag{5}$$

The contribution of the factor *fth* is removed, and the next factor is computed in the remaining residue, where each sample $\mathbf{X}_i$ is replaced by $[\mathbf{X}_i - \mathrm{t}_i\mathbf{w}^j(\mathbf{w}^k)^T]$ and **y** by $(\mathbf{y} - \mathbf{Tv})$. The number of factors

calculated must be enough to describe the relationship between the independent variables and **y**, the optimal number of factors can be selected by means of techniques such as leave-one-out cross-validation [42].

2.3. Residual bilinearization

Once the model is built with the optimum number of factors, prediction of the concentration ($\hat{y}$) for a sample $\mathbf{X}_u$, is given by Eq. (6).

$$\hat{y}_u = \mathbf{t}_u^T \mathbf{v} \tag{6}$$

where $\mathbf{t}_u$ is the score vector corresponding to the test sample data matrix $\mathbf{X}_u$, obtained according Eq. (3) and $\mathbf{v}$ are the regression coefficients given by Eq. (5). However, the formulation of the N-PLS as described above is not suitable for the prediction of y_u when an unexpected component appears in the test sample. An alternative is to evaluate the test sample residue of the N-PLS model, which will be higher than the instrumental noise when an interfering agent is present [43]. The estimate of the residual variance for a test sample $\mathbf{X}_u$ is given by Eq. (9).

$$s_p = \frac{||\mathbf{E}_p||}{(JKI - f)^{1/2}} = \frac{||\mathbf{X}_u - reshape\{\mathbf{t}_u[(\mathbf{w}^j)| \otimes |(\mathbf{w}^k)]\}||}{(JKI - f)^{1/2}} \tag{9}$$

where $\mathbf{E}_p$ is the residue matrix, reshape is the operation of transforming a vector $JK \times 1$ in a matrix $J \times K$ (the inverse of the unfolding operation), $|\otimes|$ is the operator Kathri-Rao and $||\,||$ indicates the Euclidean norm. The bilinearization matrix $\mathbf{E}_p$ helps to circumvent the presence of interfering agents. $\mathbf{E}_p$ is decomposed by singular value decomposition as shown in Eq. (10).

$$\mathbf{B}_{unex}\mathbf{G}_{unex}(\mathbf{C}_{unex})^T = SVD(\mathbf{E}_p) \tag{10}$$

The term $\mathbf{B}_{unex}\mathbf{G}_{unex}(\mathbf{C}_{unex})^T$ containing information about the interferents is added to $\mathbf{X}_u$ rebuilt for f factors (see Eq. (11)). During the step of minimizing, the loadings obtained in the calibration stage are kept constant, and the scores $\mathbf{t}_u$ are adjusted to minimize the residual variance s_u (Eq. (12)) by the Gauss–Newton procedure [44].

$$\mathbf{X}_u = reshape\{\mathbf{t}_u[(\mathbf{w}^j)| \otimes |(\mathbf{w}^k)]\}|| + \mathbf{B}_{unex}\mathbf{G}_{unex}(\mathbf{C}_{unex})^T + \mathbf{E}_u \tag{11}$$

$$S_u = \frac{||\mathbf{E}_u||}{[(j-1)(k-1) - f]^{1/2}} \tag{12}$$

where l is the number of RBL factors. Once adjusted the scores ($\mathbf{t}_u$) of the test sample at the end of the RBL process, the analyte concentration is estimated using Eq. (4). It is worth remembering here that unlike methods such as PARAFAC and MCR-ALS, which estimates the pure profile of the active components of the system, in the RBL process, these profiles are obtained by svd and may have no resemblance to the true profiles [45].

Bortolato and coworkers [46] demonstrated that for complex test samples in which more than one unexpected constituent may be present, the RBL procedure could be stuck in local minima, resulting in inaccurate predictions. To circumvent this drawback, the authors propose adding a stochastic step based on particle swarm optimization (PSO) to find the global minimum during the RBL procedure [46]. Introduced by Kennedy and Eberhart [47], the PSO emerged from experiments with algorithms modeled from the observation of social behavior of certain species of birds. The particles considered by the algorithm behave like birds looking for food or place their nests, using own learning and the learning of the bunch. For be stochastic PSO is able to circumvent the presence of local minima, thus improving the RBL procedure. As proposed in [46] and used in this work, the PSO is used as a prior stage to generate good starting values for the minimization Gauss-Newton. A full description of the mathematical equations involved in the PSO is described elsewhere [47,48].

2.4. The Successive Projection Algorithm for MLR models

SPA-MLR can be divided into three phases [37]. In Phase 1, the instrumental responses of the calibration set are disposed in a mean-centered matrix $\mathbf{X}_{cal}$ of dimensions $(N_{cal} \times K)$ such that the kth variable x_k is associated to the kth column vector $\mathbf{x}_k \in \Re^{N_{cal}}$. These column vectors are subjected to sequence of projection operations that result in the creation of K chains of M variables, where $M = \min(N_{cal} - 1, K)$ is the maximum number of variables that can be included in an MLR model with mean-centered data. The kth chain is initialized with variable x_k and is progressively augmented with variables that display the least collinearity with the previous ones, according to the procedure described below. At the end of phase 1, these chains are stored in a matrix $\mathbf{SEL}(M \times K)$ such that $SEL(1, k), SEL(2, k), \ldots, SEL(M, k)$ correspond to the indexes of the M variables in the kth chain.

Step 1 (Initialization): Let

$\mathbf{z}^1 = \mathbf{x}_k$ (vector that defines the initial projection operations)

$\mathbf{x}_j^1 = \mathbf{x}_j, \quad j = 1, \ldots, K$

$SEL(1, k) = k$

$i = 1$ (iteration counter)

Step 2: Calculate the matrix $\mathbf{P}^i$ of projection onto the subspace orthogonal to $\mathbf{z}^i$ as

$$\mathbf{P}^i = \mathbf{I} - \frac{\mathbf{z}^i(\mathbf{z}^i)^T}{(\mathbf{z}^i)^T\mathbf{z}^i} \tag{13}$$

where $\mathbf{I}$ is a $(N_{cal} \times N_{cal})$ identity matrix.

Step 3: Calculate the projected vectors $\mathbf{x}_j^{i+1}$ as

$$\mathbf{x}_j^{i+1} = \mathbf{P}^i\mathbf{x}_j^i \tag{14}$$

for all $j = 1, \ldots, K$.

Step 4: Determine the index j^* of the largest projected vector and store this index in element $(i+1, k)$ of the **SEL** matrix:

$$j^* = arg \max_{j=1,\ldots,K}(\mathbf{x}_j^{i+1}) \tag{15}$$

$$SEL(i+1, k) = j^* \tag{16}$$

Step 5: Let $\mathbf{z}^{i+1} = \mathbf{x}_{j^*}^{i+1}$ (vector that defines the projection operations for the next iteration).

Step 6: Let $i = i + 1$. If $i < M$ return to Step 2.

Phase 2 consists of evaluating candidate subsets of variables extracted from the chains stored in matrix **SEL**. The candidate subset of m variables starting from x_k is defined by the index set $\{SEL(1, k), SEL(2, k), \ldots, SEL(m, k)\}$. Since m ranges from 1 to M and k ranges from 1 to K and, a total of $M \times K$ subsets of variables are to be evaluated. The best subset of variables is selected on the basis of a cost function related to the prediction ability of the resulting MLR model. Usually, this cost function is calculated as the root-mean-square error obtained by using either cross-validation or a separate validation set [37]. Phase 3 consists of a backward elimination procedure aimed at improving the parsimony of the model. More details concerning this final phase are provided elsewhere [37].

2.5. The Successive Projection Algorithm for N-PLS models

Recently, we showed that the SPA can eliminate uninformative variables in PLS regression through intervals selection, getting

better accuracy than the full spectrum PLS [36]. The algorithm proposed in this report is an extension of *i*SPA-PLS for three way data. In the *i*SPA-N-PLS algorithm, initially a three-way array ($\underline{\mathbf{X}}\mathrm{cal}_{J\times J\times K}$) for calibration samples is unfolded generating the matrix **uXcal**, with dimensions $IJ \times K$ (interval selection in second sensor) or $IK \times J$ (interval selection in first sensor), according to the choice of the user and the interval selection made on the columns of the **uXcal**. There is evidence that the choice of slightly more selectively intervals allows to obtain the best results. On the other hand, repeated use of interval selection in both sensors decreases the sensitivity (by using a very small amount of analytical channels). Thus, the best compromise between accuracy, selectivity and sensitivity should be taken into account for the choice of an instrumental way.

The **uXcal** matrix is partitioned into intervals. Assumed that the K variables $x_1, x_2, \ldots, x_K$ have been divided into s non-overlapped intervals of lengths $S_1, S_2, \ldots, S_w$. In general, the intervals will have the same length, but this is not a requirement. If K is not divisible by w, the remainder of the division can be distributed among the intervals so that $S_1 + S_2 + \cdots + S_w = K$.

The *i*SPA-N-PLS algorithm can be divided into two phases. In phase 1, the columns of **uXcal** are partitioned according to the intervals of variables previously defined. The column with the largest norm within each of the s intervals is taken as a representative element of that interval. The s representative columns obtained in that way are stored in a matrix $\mathbf{S}_{cal}(IJ \times s$ or $IK \times s)$. The projection operations described in Section 2.3 are then carried out by using the columns of $\mathbf{S}_{cal}$ instead of **uXcal**. Therefore, at the end of phase 1, the indexes in the resulting matrix **SEL** will correspond to the intervals under consideration.

In phase 2, N-PLS is employed to build models for each combination of intervals associated to the indexes stored in matrix **SEL**. The combination of m intervals starting from the kth interval is defined by the index set $\{SEL(1, k), SEL(2, k), \ldots, SEL(m, k)\}$. In this case, k ranges from 1 to w and m ranges from 1 to $(w-1)$. The best combination of intervals for each sample of the test set is then chosen on the basis of the smallest value of a cost function namely *Jcost* (see Eq. (17)). Wherein R represents the value associated with the RBL procedure, calculated as shown in Eq. (18).

$$j_{cost} = \sqrt{\frac{(y_u - \hat{y}_u)^2}{I}} + R \tag{17}$$

$$R = \left|1 - \frac{S_u}{S_{cal}}\right| \tag{18}$$

when s_u is the residual for test sample u, and s_{cal} is the residual for overall calibration set (see Eq. (12)). The cost function used in the selection of intervals by iSPA-N-PLS is composed of two terms: (a) the RMSECV ensures that the selected intervals have good correlation with the dependent variable (**y**); and (b) R allows to take into account the RBL procedure applied after calibration to achieve the second-order advantage. If no unexpected constituent occurs in the test sample (s_u approximately equal s_{cal}) the RBL is suppressed ($l = 0$) and Jcost becomes equal to RMSECV and a single interval set is selected for all test samples.

The R term has aimed to guide the interval selection in regions with smaller interferent contribution, and in this case for each sample u, of the test, it is possible to select different intervals, according to the sample composition. The structure of Eq. (18) aims to prevent that intervals with low instrumental noise be selected simply by not showing signal interference and analyte, ideally s_u as similar as possible to s_{cal}.

2.6. Software

The N-PLS and N-PLS/RBL algorithms and figures of merit were computed using the graphical interface MVC2 [44]. The *i*SPA-N-PLS algorithm was developed in MatLab [49], employing routine N-PLS available in the N-way Toolbox [50] (http://www.models.kvl.dk/algorithms) and RBL routines written by Olivieri available in http://www.chemometry.com/Index/Links%20and%20downloads/Programs/Olivieri/RBL.zip.

The GA algorithm for variable selection in second order calibration developed by Carneiro et al. [29], and used in this work involves five steps: (I) codification of the variables, (II) creation of the initial population, (III) evaluation of every chromosome, (IV) crossing and (V) mutation. The GA was carried out using binary coding in the two instrumental modes; the initial population is formed by 100 chromosomes randomly generated. The initial percentage of selected variables was 10 in each mode. The crossover probability was the 50% and mutation probability of the 1% and only 1% of the genes. For all cases 100 generations were employed. The best chromosome is selected based on the value of RMSEV. A full description GA implementation is described elsewhere [29].

The base line correction and noise removal were carried out for the simulated study according to the asymmetric least-squares methodology proposed by Eilers [51], adapted to second-order data [52], and the Savitzky–Golay smoothing [53], adapted for three-way data by the authors, respectively.

3. Experimental

3.1. Simulated data

In this section, a simulation study involving the determination of two analytes (identified here as A and B) is presented to illustrate the potential benefits of the proposed variable selection method. For this purpose second-order data were created starting from pure profiles for each analytes (see Fig. 1) and random numbers taken from a gaussian distribution were added to all calibration and test signals. The standard deviation of the gaussian noise was taken as 1% of the maximum calibration signal and noise added in concentration was of 0.01.

All matrices were of size 20×20 data points. A calibration set was built according to a full factorial design with concentrations ranging from 1 to 5 for both analytes. Two test sets were built, both containing 100 samples. The first set of test samples contained only the analytes A and B. The second set of test samples contained

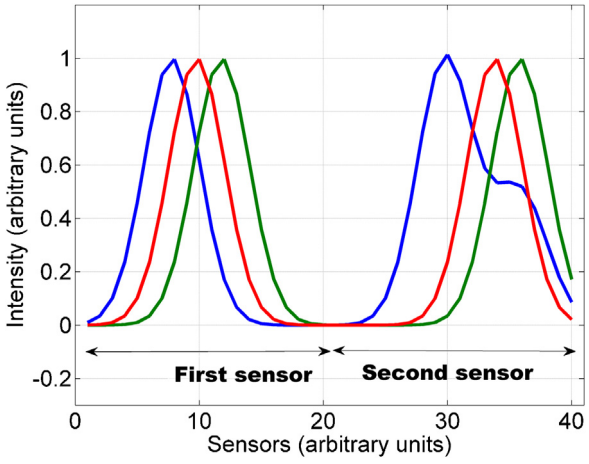


Fig. 1. Pure porfile of the ( ) A and ( ) B analytes and ( )interfering the first and second sensor.

an unexpected constituent as interferent. As can be appreciated in Fig. 1, the spectrum of the interfering agent overlaps the analytes A and B spectra, being the overlapping stronger for analyte B.

3.2. Experimental data

3.2.1. Data set I

The data set I comprises 19 mixtures of four fluorophores, namely L-phenylalanine; L-3,4-dihydroxyphenylalanine; 1,4-dihydroxybenzene and L-tryptophan in different concentrations, starting from the respective stock solutions prepared using Milli-Q water as the solvent. Perkin-Elmer LS50B fluorescence spectrometer was used to measure fluorescence landscapes using excitation wavelengths between 200 and 350 nm with 5 nm intervals. The emission wavelength range was 200–750 nm. Excitation and emission monochromator slit widths were set to 5 nm, respectively. Scan speed was 1500 nm min^{-1}. This data set is available at http://www.models.kvl.dk/datasets, and for more information see [54,55]. For calibration purposes, a tree way data array was built with dimensions ($19 \times 121 \times 24$) and then unfolded in a matrix of 19×2904. Then the SPXY algorithm [56] was applied in order to generate calibration and external prediction sets with 12 and 7 samples, respectively. Then the three-dimensional structure data was reassembled. After that, for selecting all ranges in all cases, the tree way data were unfolded in the direction IJ × K, making the intervals selection on the sensor excitation.

3.2.2. Data set II

The second system consisted in the determination of ofloxacin in water samples in presence of two other uncalibrated quinolones (ciprofloxacin and danofloxacin) by high performance liquid chromatography with DAD detection. The chromatographic measurements were performed on an Agilent Model 1100 LC instrument (Agilent Technologies, Waldbronn, Germany) equipped with degasser, quaternary pump, autosampler, oven column compartment, UV–vis diode array detector (DAD) and the CHEMSTATION software package to control the instrument, data acquisition and data analysis. The analytical column used was a Zorbax Eclipse XDB-C18, dimensions 4.6 mm × 7.5 mm 3.5-μm (Agilent Technologies, Waldbronn, Germany).

The column temperature was controlled by an oven at 35 °C. The mobile phase consisted of a 10 mmol L^{-1} sodium acetate/acetic acid buffer (pH 4)–methanol–acetonitrile mixture (71:20:9, v/v). The stock standard solutions of ofloxacin were eluted in isocratic mode and the flow rate was maintained at 1.80 mL min^{-1} with injection volume of the 100 μL.

All standards and solvents were of analytical grade, ofloxacin (OFN) was provided by Sigma (Germany). ciprofloxacin (CPF) and danofloxacin (DNF) were purchased from Fluka (Switzerland). Methanol (MeOH) LC grade and Acetonitrile LC grade were obtained from J.T. Baker (Deventer, The Netherlands). Ultrapure water was obtained from a Milli-Q water purification system from Millipore (Bedford, MA, USA). Acetic acid (AcCl) analytical grade was purchased from Cicarelli (Santa Fe, Argentina) and trihydrate sodium acetate (NaAc) analytical grade was purchased from ANEDRA, (Research AG S.A., Argentina). Stock standard solutions of pharmaceuticals were prepared in MeOH with concentration levels of 200 mg L^{-1}, and were maintained under refrigeration at 4 °C in darkness condition. Working standard solutions were prepared in a daily basis by dilution of appropriate aliquots in water.

The calibration set consists of five standards at concentrations of 2.0, 4.0, 6.0, 8.0 and 10.0 mg L^{-1} of ofloxacin, and test set was built with nine water samples spiked with different concentrations of ofloxacin, ciprofloxacin and danofloxacin. The running time for all standards and samples was 2 min and the spectral range of detection was of 200 at 400 nm, generating matrices of time × absorbance (294×201). The range time was reduced for 51×201. On all case interval selection was carried out in wavelength dimension.

4. Results and discussion

4.1. Simulated data

Interval selection by iSPA-N-PLS was performed for both raw and smoothed and base line corrected data as indicated in Section 2.6. The results show small changes, and in all cases the same subset of intervals was selected by iSPA-N-PLS. However the best results in terms of accuracy were found for the raw data. This effect can be attributed to the fact that the data used in this study did not present baseline profile or strong noise. In addition, as discussed by Bro [57], preprocessing of three-way data is not a trivial task, and in most cases it is preferable to use the raw data.

The results obtained for the prediction of the two test sets are presented in Table 1. As can be observed, the RMSEP values computed when employing the variable selection were, in most cases, lower than those obtained without variable selection or selection of individual variables by GA. This effect is more evident for analyte B, whose selectivity (related to the grade of overlapping) in less than for analyte A in both validation sets.

The sensitivity corresponds to the fraction of the signal responsible for the addition of a unit concentration of the ownership

Table 1
Prediction results for simulated data set.

Set	Model	Analyte	RMSEP	Factors[a,b]	LOD	LOQ	SEN_n	γ^{-1}
I	N-PLS	A	0.0022	2/0	0.009	0.026	4.49	441
		B	0.0031	2/0	0.012	0.035	3.11	261
I	*i*SPA-N-PLS	A	0.0019	2/0	0.008	0.025	3.64	117
		B	0.0015	2/0	0.011	0.035	3.16	298
I	GA-N-PLS	A	0.0030	2/0	0.017	0.057	1.71	57.95
		B	0.0040	2/0	0.011	0.037	2.38	216.15
II[c]	N-PLS/RBL	A	0.0067	2/1	0.015	0.046	4.49	36
		B	0.0476	2/1	0.028	0.084	3.11	26
II[c]	*i*SPA-N-PLS/RBL	A	0.0060	2/1	0.020	0.060	1.68	14
		B	0.0140	2/1	0.036	0.107	1.74	15
II[c]	GA-N-PLS/RBL	A	0.0106	2/1	0.019	0.058	1.27	23.47
		B	0.0114	2/1	0.042	0.118	0.84	4.4808

[a] Number of N-PLS factors included in the model.
[b] Number of RBL factors.
[c] Test set with unexpected constituent.

interest [58], and in this work the N-PLS/RBL sensitivity (SEN_n) was estimated from the equation:

$$SEN_n = \frac{1}{\left\| (\mathbf{P}_{eff}^{+})^T \mathbf{v} \right\|} \quad (19)$$

where $\mathbf{P}_{eff}^{+}$ is defined as 'effective' latent matrix and $\mathbf{v}$ is the vector of regression coefficients, $\mathbf{P}$ is obtained with an idea similar to the net analyte signal (NAS) concept, proposed by Lorber [59], more details are provided elsewhere [60].

In order to estimate the figures of merit limit of detection (LOD) and limit of quantitation (LOQ), 100 blank samples were created for each analyte and the values of LOD and the LOQ were obtained using Eqs. (20) and (21) as suggested in [44]:

$$LOD = 3.3 s_y \quad (20)$$

$$LOQ = 10 s_y \quad (21)$$

where s_y is the standard deviation for prediction blank samples. For the first test set, similar values of all analytical figures of merit (AFOMs) were obtained, even employing a smaller number of sensors models *i*SPA-N-PLS. On the other hand, when computing AFOMs for models with RBL, a decrease in sensitivity was observed when employing the variable selection, with subsequent slightly higher LODs and LOQs. This fact was expected owing to that SEN is directly related to the number of sensors [60].

In order to have a better inside of the improvement in predictive ability when selecting variables, the elliptical joint confidence regions for all models are displayed in Fig. 2. As can be observed, these regions are smaller for models with variable selection by *i*SPA than those obtained for N-PLS models without variables selection, being the only exception the region computed for predictions of analyte B in test set II. Opposed, models carried out by GA-N-PLS present elliptical joint confidence regions which in most cases are larger than those obtained for iSPA-NPLS models, and in some cases being larger than those for N-PLS models. As shown by Höskuldsson [20], PLS models based in interval selection are more stable in prediction step, because scores vector based on intervals are larger than one based on individual variables.

Interestingly, the four regions computed for predictions with variable selection contain the ideal point of unit and zero for slope and intercept, respectively. The latter is not true for the region computed for analyte B without variable selection. This fact is indicative of a better modeling accuracy when employing *i*SPA coupled to modeling N-PLS.

Finally, the selected intervals and individual sensors selected by GA for each analyte (A and B) in every test set (I and II) are shown in the landscapes plotted in Fig. 3. For the test set II are presented only the intervals of the first test sample of both analytes. Every landscape plotted in Fig. 3 shows the full matrix and the selected by *i*SPA interval. In addition, pure profiles for components A, B and interfering are displayed.

As can be appreciated in Fig. 3a and b, which corresponds to data set I, the selected intervals match with regions in which less overlapping between the analytes is observed. On the other hand, Fig. 3c and d, which corresponds to data set II, it can be seen that the intervals are selected keeping a compromise between analytical information for the target analyte, discarding regions of strong overlap, especially with the interfering component, on the other hand GA selects variables with low signal to noise ratio and uninformative.

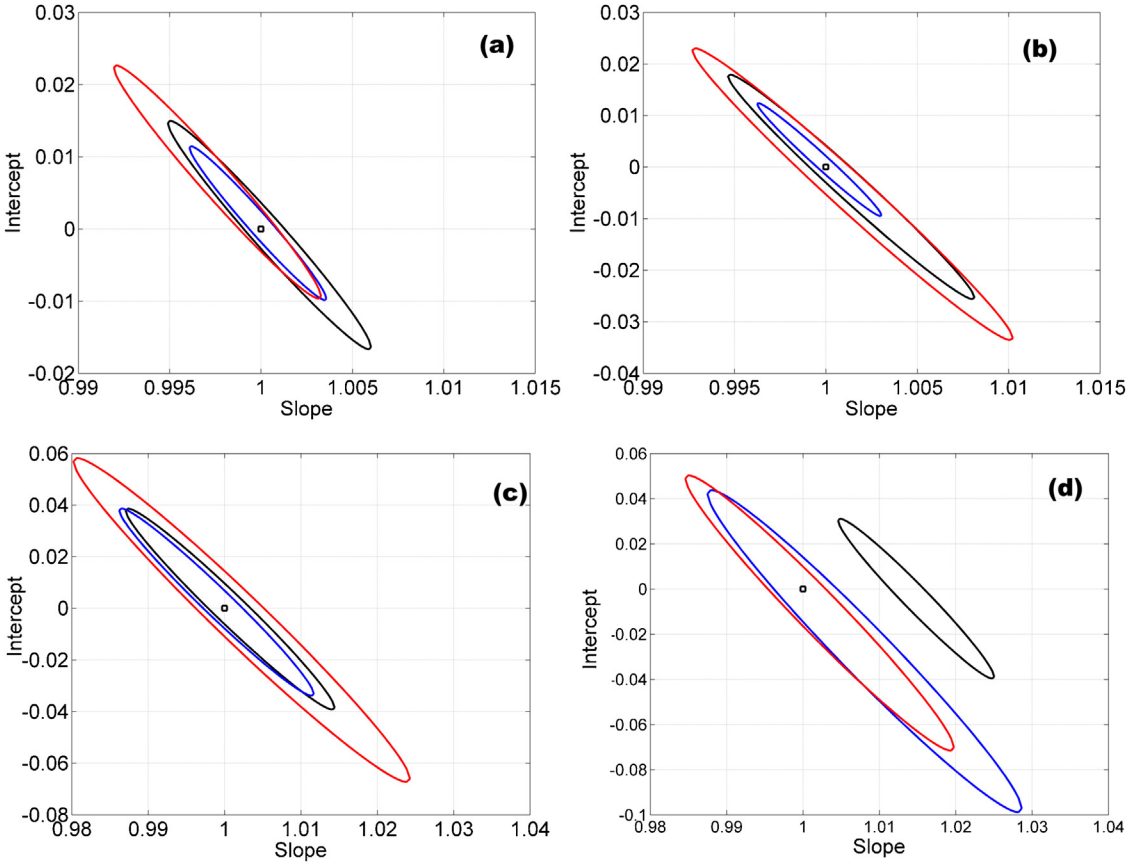


Fig. 2. Elliptical joint confidence regions for the slope and intercept of the regression of predicted concentrations vs. reference values for (—) N-PLS, (▬) GA-PLS and (▬) iSPA-N-PLS: (a) analyte A test set I, (b) analyte B test set I, (c) analyte A test set II and (d) analyte B test set II.

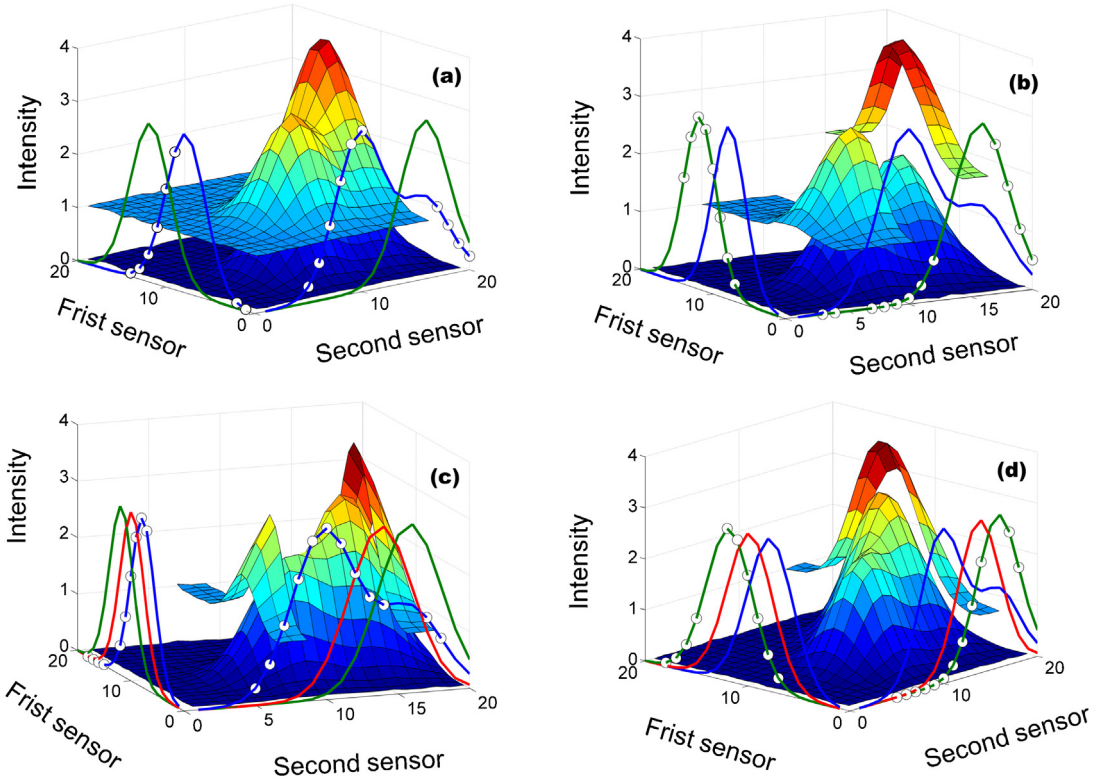


Fig. 3. Simulated porfile landscapes for firt test samples of both test sets and landscapes corresponding to selected intervals shifted in the *y*-axis by an offset value for (a) analyte A test set I, (b) analyte B test I, (c) analyte A test set II (d) analyte B test set II. The lines are the pure profiles for ( ) A, ( ) B, ( ) Interferent and (o) are selected sensors by GA.

4.2. Experimental data set I

This experimental system was studied in order to evaluate the benefits of the variable selection by *i*SPA in a data set in which it is not necessary second order advantage. Initially Rayleigh and Raman scattering effects present in fluorescence excitation-emission matrix (EEM) were adequately addressed, not to compromise the drawings of calibration models. If these effects are not properly treated and eliminated, prediction results can be enormously affected [61].

The optimum number of latent variables in each N-PLS model was found using full cross-validation when modeling the calibration data set. The prediction results for the analytes L-phenylalanine, L-3,4-dihydroxyphenylalanine, 1,4-dihydroxibezene and L-trytophan are shown in Table 2. As can be observed, the prediction of all the four analytes present smaller RMSEP values using the method proposed when compared to conventional N-PLS and GA-N-PLS. This better predictive ability can be corroborated by observation of the elliptical joint confidence regions for the slope and intercept of the regression of predicted concentration versus reference values obtained by bivariate least squares displayed in Fig. 4.

As it can be observed in the latter figure, all the *i*SPA-N-PLS models furnish confidence regions containing the ideal point of unit and zero for slope and intercept, respectively. In addition, confidence regions are smaller than those obtained by N-PLS and GA-N-PLS, indicating that there is not significant bias for the prediction when *i*SPA was applied, i.e. best accuracy and precision was achieved by implementation of the novel algorithm presented herein.

On the other hand, as was expected based on the observations made analyzing the previous studied system (see above), the use of a smaller number of sensors impacts in a negative way on the sensitivity computed for the *i*SPA-N-PLS models, compared with those for N-PLS model (see Table 2). However, the decreasing on sensitivity is the price to be paid for using more selective sensors, disposing non-informative variables, facts that allow building N-PLS models which achieve better both accuracy and precision. For all cases the selected sensors by *i*SPA-N-PLS are presented in Fig. 5.

Table 2
Prediction results for phenylalanine, L-3,4-dihydroxyphenylalanine (L-DOPA), 1,4-dihydroxibezene and L-trytophan.

Model	Analyte	RMSEP (mmol L^{-1})	Factors (N-PLS)	LOD (mmol L^{-1})	LOQ (mmol L^{-1})	SEN_n	γ^{-1} (mmol L^{-1})
N-PLS	Phenylalanine	2.43	4	4.35	13.17	79.28	2.24
	L-DOPA	0.31	4	0.36	1.08	600.22	19.15
	1,4-Dihydroxibezene	323.13	4	159.88	484.47	0.69	0.03
	L-Trytophan	3.21	4	3.44	10.42	66.75	1.93
*i*SPA-N-PLS	Phenylalanine	0.62	4	0.15	0.45	51.04	20.06
	L-DOPA	0.26	4	0.23	0.70	331.56	38.02
	1,4-Dihydroxibezene	278.43	4	94.83	287.37	0.05	0.02
	L-Trytophan	2.74	4	1.30	3.96	13.91	2.42

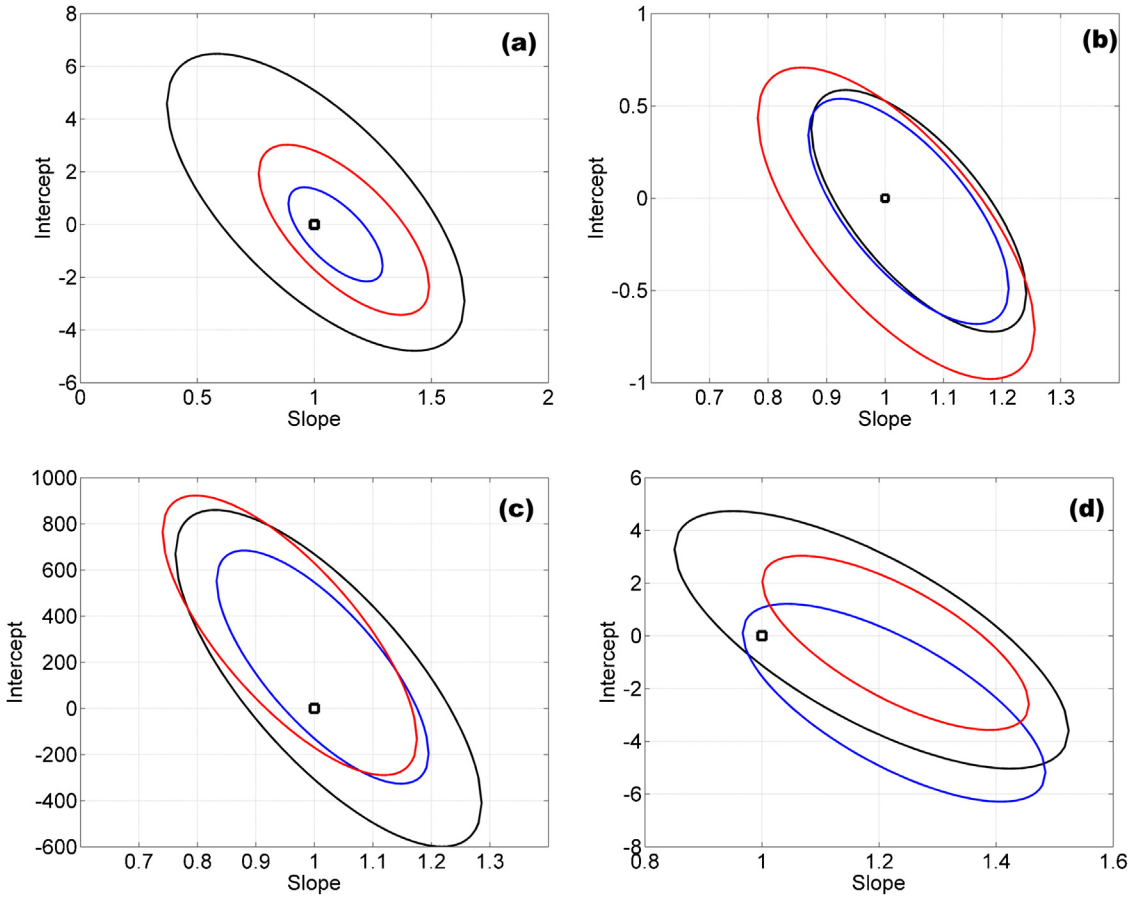


Fig. 4. Elliptical joint confidence regions for the slope and intercept of the regression of predicted concentrations vs. reference values for (—) N-PLS, (▬) GA-N-PLS and (▬) iSPA-N-PLS: (a) L-phenylalanine, (b) L-3,4-dihydroxyphenylalanine, (c) 1,4-dihydroxybenzene and (d) L-tryptophan.

It is important to remark that unlike the previous simulated system studied in this work, which present a soft background, in the experimental set 1, the removal of the noise by variable selection conducts to an increase in the analytical sensitivity, which is a function of sensitivity and instrumental noise [60]. This increased analytical explains the lower LOD and LOQ values obtained when selecting variables.

4.3. Experimental data set II

In this example, ofloxacin was determined in water samples containing two unexpected constituents: ciprofloxacin and danofloxacin. In order to achieve the second-order advantage RBL was implemented every time a test sample was predicted. Table 3 summarize the results for quantification of the ofloxacin.

Table 3
Prediction results for oflaxacin in water samples.

Sample	Nominal	N-PLS/RBL			*i*SPA-N-PLS/RBL			GA-N-PLS/RBL		
		Predicted	Rec (%)	Factors[a]	Predicted	Rec (%)	Factors[a]	Predicted	Rec (%)	Factors[a]
1	8.83	9.22	104.43	1/2	9.21	104.28	1/1	9.59	108.66	1/1
2	6.00	6.92	115.35	1/2	7.00	116.64	1/1	6.91	115.18	1/1
3	2.00	2.66	132.99	1/2	2.62	131.12	1/1	2.46	131.80	1/1
4	3.17	3.48	109.75	1/2	3.45	108.73	1/1	3.38	106.68	1/1
5	10.00	10.27	102.66	1/2	10.26	102.63	1/1	10.48	104.81	1/1
6	3.17	3.30	104.20	1/2	3.29	103.82	1/1	3.63	114.56	1/1
7	6.00	5.38	89.67	1/2	5.23	87.24	1/1	5.98	99.66	1/1
8	6.00	4.22	70.41	1/2	4.69	78.25	1/1	4.76	79.26	1/1
9	6.00	6.63	110.50	1/2	6.68	111.32	1/1	6.64	110.41	1/1
10	8.83	8.62	97.61	1/2	8.95	101.38	1/1	9.22	115.19	1/1
11	6.00	6.51	108.43	1/2	6.52	108.60	1/1	6.91	112.68	1/1
12	6.00	6.53	108.45	1/2	6.45	107.53	1/1	6.76		
RMSEP		0.72			0.64			0.70		
LOD ($g\,L^{-1}$)		0.05			0.08			0.07		
LOQ ($g\,L^{-1}$)		0.13			0.23			0.22		
SEN_n		556.74			119.04			191.22		
γ^{-1}		7.58			2.20			3.28		

[a] Number of calibration N-PLS factors/number of RBL factors.

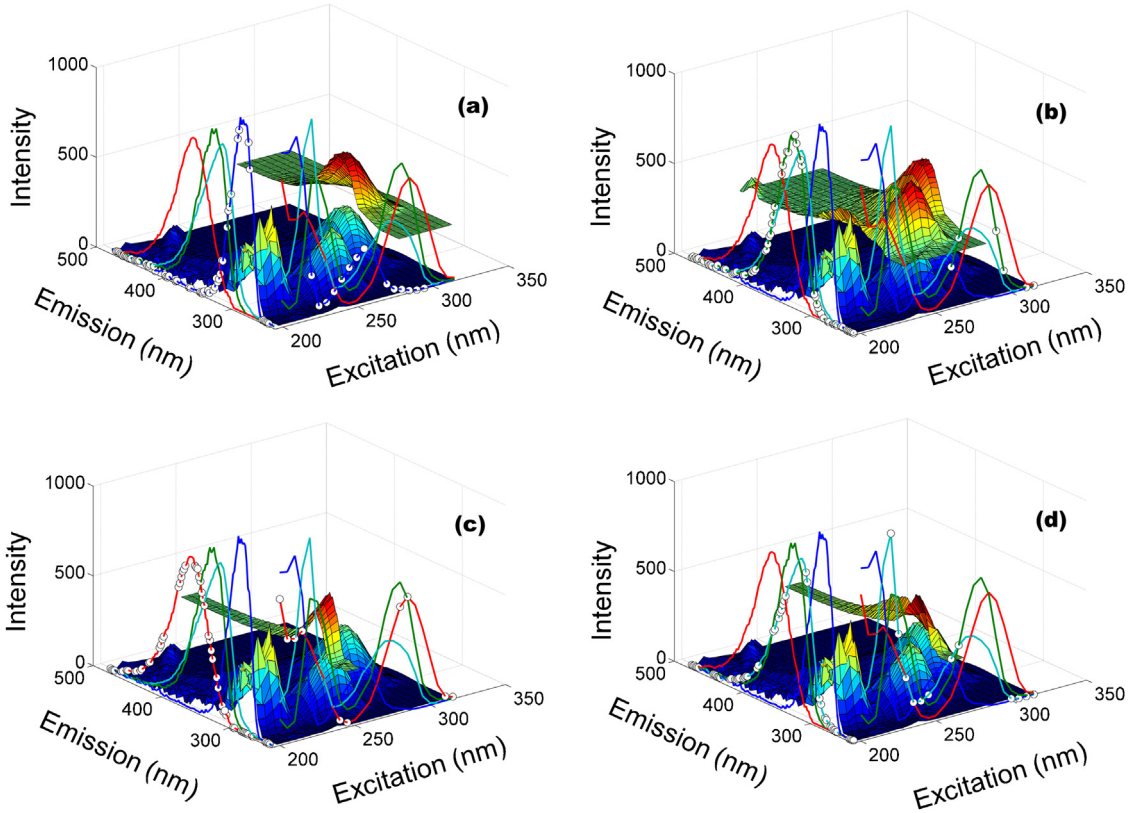


Fig. 5. Fluorescence landscapes of the samples containing four fluorophores and displaced by offset landscapes selected intervals for (a) ( ) L-phenylalanine, (b) ( ) L-3,4-dihydroxyphenylalanine, (c) ( ) 1,4-dihydroxybenzene and (d) ( ) L-tryptophan and (o) are selected sensors by GA.

As was found when analyzing the other cases, the *i*SPA algorithm allowed to obtain a lower prediction error. This fact can be corroborated by examination of the elliptical joint confidence regions presented in Fig. 6a. The elliptical regions plotted for both models contain the ideal points. However, the regions corresponding to the *i*SPA-N-PLS model is smaller than the one corresponding to modeling with N-PLS without variable selection. On the other hand, when compared with the GA-N-PLS/RBL model, the elliptical joint confidence regions are slightly larger for the *i*SPA-N-PLS/RBL model. However, the ideal point is located in a more central position for, indicating a lesser bias value.

Finally, the interval selected by *i*SPA and variables selected by GA for prediction of the ofloxacin samples can be seen in Fig. 6b. As was mentioned, the analyzed samples contained two interferents, thus two RBL factors should be used when building the models with the whole data for each sample. Interestingly, when the region was selected by the *i*SPA algorithm or individual variables by GA, better predictions were achieved using a single RBL factor. Consequently, it can be concluded that when the signal of the analyte is not completely overlapped, variable selection can reduce the number of RBL factors which are necessary to achieve second-order advantage.

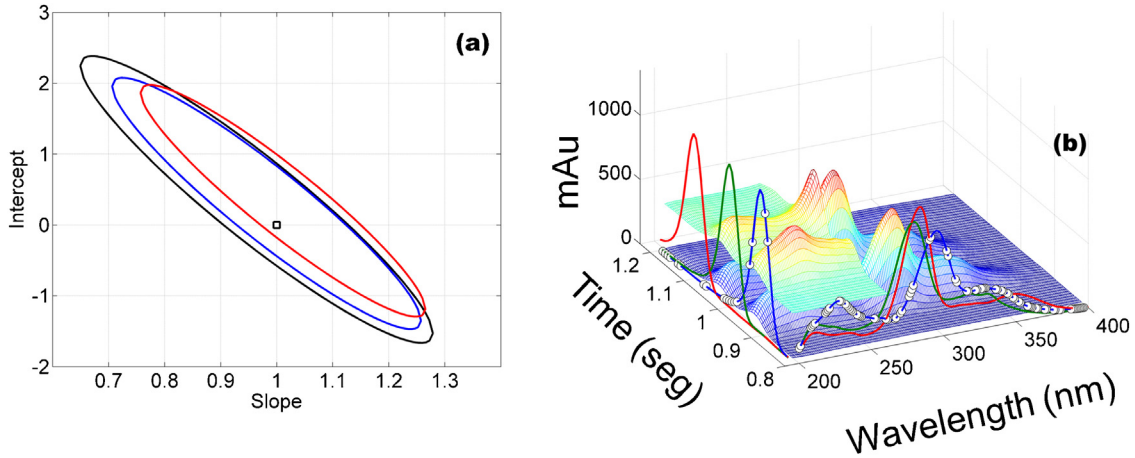


Fig. 6. Graphical results for ofloxacin quantification: (a) elliptical joint confidence regions for the slope and intercept of the regression of predicted concentrations vs. reference values for (—) N-PLS/RBL, ( ) GA-NPL-RBL and ( ) iSPA-N-PLS/RBL and (b) matrix time × absorbance for first test samples and pure profile for ( ) ofloxacin, ( ) ciprofloxacin and ( ) danofloxacin. Displaced by offset the selected interval by *i*SPA-N-PLS and (o) are selected sensors by GA.

5. Conclusion

The ability of the already known SPA algorithm for interval selection when extended for N-PLS was demonstrated. A simulated data set and two experimental examples involving fluorescence excitation-emission and high performance liquid chromatography UV–visible diode array detector were used to evaluate the benefits of the new method. The proposed algorithm allows building models which discards non-informative variables leading to a more accurate and precise prediction when compared with models using the full data. The results further suggest that as well as the bilinear PLS in trilinear PLS (N-PLS) selection of intervals also produces better results when bought in discrete variables.

Acknowledgments

The authors acknowledge the support of CAPES (PhD scholarship) and CNPq (research fellowships). The authors are also grateful to Universidad Nacional del Litoral (Project CAI + D 2011 No. 11-11), to CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas, Project PIP 2012-14 No. 455) and to ANPCyT (Agencia Nacional de Promoción Científica y Tecnológica, Project PICT 2011-0005) for financial support. MRA thanks CONICET for her fellowship. We are grateful to Prof. Dr. Retano Lajarin (UFSCar) for GA-N-PLS/RBL routine facilities.

References

[1] L. Hu, M. Ye, X. Jiang, S. Feng, H. Zou, Anal. Chim. Acta 598 (2007) 193–204.
[2] G. Lespes, J. Gigault, Anal. Chim. Acta 692 (2011) 26–41.
[3] H. Parastar, J.R. Radovic, M. Jalali-Heravi, S. Diez, J.M. Bayona, R. Tauler, Anal. Chem. 83 (2011) 9289–9297.
[4] C.M. Andersen, R. Bro, J. Chemom. 24 (2010) 728–737.
[5] M. Forina, S. Lanteri, M.C.C. Oliveros, C.P. Millan, Anal. Bioanal. Chem. 380 (2004) 397–418.
[6] Z. Xiaobo, Z. Jiewen, M.J.W. Povey, M. Holmes, M. Hanpin, Anal. Chim. Acta 667 (2010) 14–32.
[7] R. Leardi, L. Nørgaard, J. Chemom. 18 (2004) 486–497.
[8] M. Shamsipur, V. Zare-Shahabadi, B. Hemmateenejad, M. Akhond, J. Chemom. 20 (2006) 146–147.
[9] L.F.B. Lira, M.S. Albuquerque, J.G.A. Pacheco, T.M. Fonseca, E.H.S. Cavalcanti, L. Stragevitch, M.F. Pimentel, Microchem. J. 96 (2010) 126–131.
[10] R.M. Balabin, S.V. Smirnov, Anal. Chim. Acta 692 (2011) 63–72.
[11] D.D.S. Fernandes, A.A. Gomes, G.B. da Costa, G.W.B. da Silva, G. Véras, Talanta 87 (2011) 30–34.
[12] R.K.H. Galvão, M.F. Pimentel, M.C.U. Araújo, T. Yoneyama, V. Visani, Anal. Chim. Acta 443 (2001) 107–115.
[13] R.J. Leardi, J. Chemom. 14 (2000) 643–655.
[14] F. Allegrini, A.C. Olivieri, Anal. Chim. Acta 699 (2011) 18–25.
[15] J.A. Hageman, M. Streppel, R. Wehrens, L.M.C. Buydens, J. Chemom. 17 (2003) 427–437.
[16] U. Hörchner, J.H. Kalivas, J. Chemom. 9 (1995) 283–308.
[17] J.A.F. Pierna, O. Abbas, V. Baeten, P. Dardenne, Anal. Chim. Acta 642 (2009) 89–93.
[18] M. Forina, C. Casolino, C.P. Millan, J. Chemom. 13 (1999) 165–184.
[19] R.F. Teófilo, J.P.A. Martins, M.M.C. Ferreira, J. Chemom. 23 (2009) 32–48.
[20] A. Höskuldsson, Chemom. Intell. Lab. Syst. 55 (2001) 23–38.
[21] U. Hörchner, J.H. Kalivas, Anal. Chim. Acta 9 (1995) 1–13.
[22] A.C. Olivieri, Anal. Methods 4 (2012) 1876–1886.
[23] R.M. Maggio, P.C. Damiani, A.C. Olivieri, Talanta 83 (2011) 1173–1180.
[24] H.C. Goicoechea, K. Calimag-Williams, A.D. Campiglia, Anal. Chim. Acta 717 (2012) 100–109.
[25] M.J. Culzoni, A.M. Llanos, M.M. Zan, A. Espinosa-Mansilla, F. Cañada-Cañada, A.M. Peña, H.C. Goicoechea, Talanta 85 (2011) 2368–2374.
[26] W. Wu, Q. Guo, D.L. Massart, C. Boucon, S. Jong, Chemom. Intell. Lab. Syst. 65 (2003) 83–95.
[27] J.A. Lopes, J.C. Menezes, Chemom. Intell. Lab. Syst. 68 (2003) 75–81.
[28] S. Gourvénec, X. Capron, D.L. Massart, Anal. Chim. Acta 519 (2004) 11–21.
[29] R.L. Carneiro, J.W.B. Braga, C.B.G. Bottoli, R.J. Poppi, Anal. Chim. Acta 595 (2007) 51–58.
[30] M.C.U. Araújo, T.C.B. Saldanha, R.K.H. Galvão, T. Yoneyama, H.C. Chame, V. Visani, Chemom. Intell. Lab. Syst. 57 (2001) 65–73.
[31] H.M. Paiva, S.F.C. Soares, R.K.H. Galvão, M.C.U. Araújo, Chemom. Intell. Lab. Syst. 118 (2012) 260–266.
[32] S. Ye, D. Wang, S. Min, Chemom. Intell. Lab. Syst. 91 (2008) 194–199.
[33] M.J.C. Pontes, R.K.H. Galvão, M.C.U. Araújo, P.N.T. Moreira, O.D. Pessoa Neto, G.E. José, T.C.B. Saldanha, Chemom. Intell. Lab. Syst. 78 (2005) 11–18.
[34] F.A. Honorato, R.K.H. Galvão, M.F. Pimentel, B. Barros Neto, M.C.U. Araújo, F.R. Carvalho, Chemom. Intell. Lab. Syst. 76 (2005) 65–72.
[35] H.A. Dantas Filho, R.K.H. Galvão, M.C.U. Araújo, E.C. Silva, T.C.B. Saldanha, G.E. José, C. Pasquini, I.M. Raimundo Jr., J.J.R. Rohwedder, Chemom. Intell. Lab. Syst. 72 (2004) 83–91.
[36] A.A. Gomes, R.K.H. Galvão, M.C.U. Araújo, G. Véras, E.C. Silva, Microchem. J. 110 (2013) 202–208.
[37] S.F.C. Soares, A.A. Gomes, A.R. Galvão Filho, M.C.U. Araujo, R.K.H. Galvão, Trends Anal. Chem. 42 (2013) 84–98.
[38] R. Bro, J. Chemom. 10 (1996) 47–61.
[39] S. Wold, P. Geladi, K. Esbensen, J. Øhman, J. Chemom. 1 (1987) 41–56.
[40] L. Stlhle, Chemom. Intell. Lab. Syst. 7 (1989) 95–100.
[41] R.G. Brereton, Analyst 15 (2000) 2125–2154.
[42] R. Bro, Multi-way analysis in the food industry (doctoral thesis), University of Amsterdam, Netherlands, 1998.
[43] V.A. Lozano, G.A. Ibañez, A.C. Olivieri, Anal. Chim. Acta 610 (2008) 186–195.
[44] A.C. Olivieri, H. Wu, R. Yu, Chemom. Intell. Lab. Syst. 96 (2009) 246–251.
[45] D.B. Gil, A.M. Peña, J.A. Arancibia, G.M. Escandar, A.C. Olivieri, Anal. Chem. 78 (2006) 8051–8058.
[46] S.A. Bortolato, J.A. Arancibia, G.M. Escandar, A.C. Olivieri, J. Chemom. 20 (2007) 1–10.
[47] J. Kennedy, R.C. Eberhart, Particle swarm optimisation, in: Proceedings of IEEE International Conference on Neural Networks, Perth, Australia, November, 1995, pp. 1942–1948.
[48] R. Thangaraj, M. Pant, A. Abraham, P. Bouvry, Appl. Math. Comput. 217 (2011) 5208–5226.
[49] MATLAB 6.0, The MathWorks Inc., Natick, MA, 2000.
[50] C.A. Andersson, R. Bro, Chemom. Intell. Lab. Syst. 52 (2000) 1–4.
[51] P.H.C. Eilers, Anal. Chem. 76 (2004) 404–411.
[52] P.H.C. Eilers, I.D. Currie, M. Durbán, Comput. Stat. Data Anal. 50 (2006) 61–76.
[53] A. Savitzky, M.J.E. Golay, Anal. Chem. 36 (1964) 1627–1639.
[54] D. Baunsgaard, Factors affecting 3-way modelling (PARAFAC) of fluorescence landscapes (internal report), Department of Dairy and Food Science, The Royal Veterinary and Agricultural University Denmark, 1999.
[55] J. Riu, R. Bro, Chemom. Intell. Lab. Syst. 65 (2003) 35–49.
[56] R.K.H. Galvão, M.C.U. Araújo, G.E. José, M.J.C. Pontes, E.C. Silva, T.C.B. Saldanha, Talanta 67 (2005) 736–740.
[57] R. Bro, Chemom. Intell. Lab. Syst. 38 (1997) 149–171.
[58] K.S. Booksh, B.R. Kowalski, Anal. Chem. 66 (1994) 782–791.
[59] A. Lorber, Anal. Chem. 58 (1986) 1167–1172.
[60] A.C. Olivieri, J. Chemom. 19 (2005) 253–265.
[61] M. Bahram, R. Bro, C. Stedmon, A. Afkhami, J. Chemom. 20 (2006) 99–105.