Bulgarian Chemical Communications, Volume 49, Number 2, (pp. 410 – 421) 2017

Assessment of different PLS algorithms for quantification of three spectrally overlapping drugs

G. Gergov[1] *, A. Alin[2], M. Doychinova[3], M. De Luca [4], V. Simeonov[5], Y. Al-Degs[6]

[1]*Department of Chemistry, Faculty of Pharmacy, Medical University, Dunav St.2, Sofia 1000, Bulgaria.*
[2]*Department of Statistics, Dokuz Eylul University, Izmir, Turkey*
[3]*Institute of Chemical Engineering, Bulgarian Academy of Sciences, Acad. G. Bonchev St., bl.103, Sofia 1113, Bulgaria*
[4]*Department of Pharmacy, Health and Nutritional Sciences, University of Calabria, 87036 Rende, Italy*
[5]*Laboratory of Chemometrics and Environmetrics, Faculty of Chemistry and Pharmacy, Sofia University "St. Kliment Ohridski", J. Bourchier Blvd. 1, Sofia 1164, Bulgaria.*
[6]*Chemistry Department, The Hashemite University, P.O. Box 150459, Zarqa 13115, Jordan*

Received July 20, 2016; Revised September 26, 2016

The primary aim of the present study was to compare the prediction power of different PLS algorithms as applied to the quantification of three spectrally overlapping drugs. Four variants of PLS were chosen for multivariate calibration and prediction of the three components of the drug formulation (paracetamol, propyphenazone and caffeine). NIPALS and SIMPLS algorithms were the most commonly used algorithms. The other tested algorithms were Kernel and Bidiagonalization which have been rarely applied in pharmaceutical analysis.

First-order data were created by measuring UV-spectra of drug mixtures over the range 190-300 nm with a resolution of 2.0 nm (i.e., 56 spectral points/sample). Reduced five-level full factorial design was used and the prediction power of PLS variants was tested for drugs levels outside the ranges selected in the calibration set.

De Luka method and the confirmative bootstrap method were helpful for the quick selection of the spectral regions.

The external prediction using the PLS-Kernel calibration model showed significant advantages in the analysis of the common marketed formulation SARIDON. The three drugs in SARIDON were quantified with mean recoveries and precisions of 96.4 (1.3), 95.1 (2.5) and 96.2 (2.9) for paracetamol, propyphenazone and caffeine, respectively. This turned out to be the optimal algorithm which could be successfully applied for the routine analysis of analgesic and antipyretic tablets in the pharmaceutical industry.

Keywords: PLS1 algorithms: NIPALS, SIMPLS, KERNEL, BIDIAGONALIZATION, Spectral overlapping; De Luca method, Bootstrap method, SARIDON formulation.

INTRODUCTION

Pharmaceutical formulations, in which one dominant component is combined with other drugs, are designated to enhance the final pharmacological effects of each substance and to cover a larger medical treatment [1]. Formulations containing paracetamol PAR, propyphenazone PRO, and caffeine CAF represent a commonly prescribed combination for pain relief [2]. Paracetamol is a common antipyretic and analgesic agent used as an alternative to aspirin (acetylsalicylic acid) in some countries [1]. Propyphenazone is derived from pyrazolone with analgesic, antipyretic and anti-inflammatory effects [2]. It is a non-steroidal antiinflammatory drug incorporated together with paracetamol into many analgesic combinations [2-3]. CAF, a methylated xanthine and potent stimulant of the central nervous system, has been added to PAR and PRO in various combinations [1,3]. Caffeine is also known to synergistically increase the analgesic effect of paracetamol and propyphenazone, providing relief from symptoms like headache, muscular aches, neuralgia, backache, joint pain, rheumatic pain, migraine, toothache and menstrual pain [2,4]. In Eastern Europe, very common marketed formulations containing PAR, PRO and CAF are known as SARIDON® and Pararemin® [3]. On the Bulgarian market the generic product SARIPHEZON® is also available.

Typical doses of the earlier preparations are 250 mg PAR, 150 mg PRO and 50 mg CAF, however, different levels are also available in the market. NeoOptalidon® is a common formulation with lower drug-doses 200 mg PAR, 125 mg PRO and 25 mg CAF, while Veramon® is available with higher PRO dosage: 200 mg PAR and 285 mg PRO [8].

The aforementioned ternary-drug formulations are of great challenge for pharmaceutical analysts to develop reliable and simple analytical procedures avoiding expensive chromatographic separation and characterized by low-consumption of organic solvents and short analysis time.

Quantification of PAR, PRO and CAF in different pharmaceutical preparations was carried

* To whom all correspondence should be sent:
E-mail: ggergov@abv.bg

out using liquid chromatography [3,5]. Derivative spectroscopic methods were also suitable for resolving such complex mixtures. Derivative ratio zero-crossing spectrophotometric determination of the three analytes was demonstrated by Dink *et al*. [5]. However, the applicability of derivative spectroscopic methods for handling overlapped spectra would be limited. More powerful multivariate calibration and UV spectrophotometry for accurate assaying marketed formulations has been well documented [1,8]. Among the multivariate calibration methods, classical least squares CLS, principal component regression PCR, and partial least squares regression PLS have been intensively applied [6].

Dink et al. proposed UV chemometric determination of a ternary mixture containing PAR, PRO and CAF in pharmaceutical preparations [7]. De Luca and co-workers have applied several chemometric methods using zero and derivative spectra to get better external prediction for PAR-PRO-CAF [8] .The same authors proposed a new procedure for wavelength selection based on the cumulative regression coefficients.

PLS modeling is the most important multivariate calibration method in many different fields including chemical and pharmaceutical analysis. PLS method has a very important advantage over other chemometric methods for using concentrations knowledge directly in calculations. PLS can deal with collinearity and offers an interactive diagnostic exploration of the data [9]. Modern instruments can generate a large number of data points per sample, which needs more advanced PLS-algorithms to end up with accurate results within minimum time , numerical stability and capacity [10,11]. PLS (particularly the nonlinear iterative partial least squares-NIPALS) is the most frequently adopted algorithm in chemical and pharmaceutical analysis [12,13]. Undoubtedly, the dramatic improvement in software production, modeling, and programming languages has positively reflected on the popularity of multivariate calibration. The intense applicability of PLS-NIPALS in pharmaceutical analysis is attributed to its availability in most commercial softwares like MVC1® and TOMCAT® [14,15]. For example, NIPALS is suitable for modeling many variables-**X** but it requires long computational time and more memory-storage [16]. PLS-SIMPLIS is proposed for increasing calibration speeding [17]. Along with this, there are two versions of the de Jong's algorithm (SIM-PLS and WIM-PLS), which are implemented in TOMCAT®[15]. WIM-PLS is specially designed for wide X matrices . Another approach, although not as fast, was presented and called the Kernel PLS algorithm . PLS-Kernel is considered as an adjustable algorithm which can be adopted for systems of many variables or even many mixtures by creating condensed and small matrices [18,19]. The kernel algorithms were improved by Dayal and MacGregor [20]. PLS-Bidiagonalization is an advanced version of another algorithm which decomposes the X matrix into three smaller matrices of orthonormal vectors [11] and this algorithm deserves investigation as it has no application in pharmaceutical analysis. It is known that the mentioned algorithms are different in their mechanisms for running chemical analysis [10,11,16].

There are two main goals of this work: a) assessment of the resolving power of four common PLS algorithms (NIPALS, SIMPLS, Kernel, and Bidiagonalization) for quantification of three spectrally overlapping drugs, and b) quick quantification of the active ingredients (PAR, CAF, and PRO) in the highly consumed marketed drug SARIDON® with minimum sample clean up. The application of the newly proposed De Luca procedure and bootstrap methods for selecting informative spectral regions before multivariate calibration is evaluated.

Theoretical background

PLS is an efficient tool for developing a quantitative relationship between several predictor variables **X** (spectral measurements in this work) and a property of interest **Y** (the independent variables or drugs content in this work). Mathematically, the relationship between **X** and **Y** or **y** (for one single independent variable) is given as [16]: **y=Xb,** where **y, X,** and **b** are drug standard concentrations in the calibration samples arranged in a vector**,** the data matrix containing the absorbances of standard solutions that are measured at different wavelengths, and the calibration sensitivity which is necessary for estimating drug content in the extracts of SARIDON®. PLS is an efficient numerical tool to find **b** which is often accomplished using different variants of PLS [11]. In general, the dimensions of the mentioned quantities are **X** (*I* samples×*J* variables) and **Y** (*I* samples×*k* solutes), **y** (*I* samples×1), and **b** (*J* variables×1). The tested PLS-variants are NIPALS, SIMPLS, Kernel, and Bidiagonalization. In the tests of comparing algorithms only PLS1(one dependent variable) was considered. The best selection of the optimum number of PLS-factors (A) is carried out by using leave-one-out cross-validation technique [21].

A brief summary on the PLS-variants is provided in this section.

PLS-NIPALS

This classical algorithm is suitable for modeling different sizes of X matrices which containing the explanatory variables. Therefore, for matrices of $I>J$ or $J>I$ but the algorithm reported to be not suitable for many variables **X** matrices [10-11]. This algorithm decomposes **X** and **y** (or **Y**) into smaller matrices and vectors to estimate calibration vector **b** to be used in the next prediction stages. The general steps of the algorithm are [10-11,16]:

w: PLS-weight for **X**: $\mathbf{w}^t=\mathbf{u}^t\mathbf{X}/(\mathbf{u}^t\mathbf{u})$

t: PLS-score for **X:** $\mathbf{t}=\mathbf{Xw}$

q: PLS-loading for **y**: $q=\mathbf{t}^t\mathbf{y}/(\mathbf{t}^t\mathbf{t})$

u: PLS-score for **y**: $q=\mathbf{yq}/(\mathbf{q}^t\mathbf{q})$

p: PLS-loading for **X**: $\mathbf{p}=\mathbf{t}^t\mathbf{X}/(\mathbf{t}^t\mathbf{t})$

X and optionally **y** are then deflated before repeating the above steps for the new PLS-variable:

$\mathbf{X}_1=\mathbf{X}-\mathbf{t}_1\mathbf{p}_1^t$ and $\mathbf{y}_1=\mathbf{y}-\mathbf{t}_1\mathbf{q}_1^t$

The next component is estimated using $\mathbf{X}_1$ and $\mathbf{y}_1$ and proceeding with $\mathbf{X}_2$,$\mathbf{y}_2$…. $\mathbf{X}_A$, $\mathbf{y}_A$ until an adequate model is established. Once the earlier vectors are estimated using the optimum number of factors (A), then **b** is estimated as:

$\mathbf{b}=\mathbf{W}^t(\mathbf{PW}^t)^{-1}\mathbf{q}$

where **W** is the weights matrix for **X**, **P** is the loadings matrix of **X**, and **q** is the loading vector for **y**. t and -1 stand for transpose and inverse operations, respectively. Once **b** is estimated by PLS-NIPALS, prediction of the target drug from the unknown spectrum $\mathbf{a}_{un}$ is carried out as following:

$c_{un}=\mathbf{a}_{un}\mathbf{b}$

PLS-SIMPLS

This algorithm is faster than PLS-NIPALS but it is not recommended for many variables-X matrices. To find the useful calibration vector b, the following quantities are computed [16, 17]:

$\mathbf{s}=\mathbf{X}^t\mathbf{y}$

r: PLS-loading for **y**: $\mathbf{r}=\mathbf{s}$

t: PLS-score for **X:** $\mathbf{t}=\mathbf{Xr}$

p: PLS-loading for **X**: $\mathbf{p}=\mathbf{X}^t\mathbf{t}$

q: PLS-loading for **y**: $\mathbf{q}=\mathbf{y}^t\mathbf{t}$

The quantities **r**, **t**, **p**, and q are stored in **R**, **T**, **P** and **q**, respectively. Before estimating the next PLS-variable s is projected on a subspace of **P**. The above algorithm is stopped once all PLS-variables are estimated as outlined earlier. Regression vector is calculated as [17]:

$\mathbf{b}=\mathbf{Rq}$

Prediction of the target drug in the new sample (extract of formulation) is estimated as shown above. PLS-SIMPLS is faster than classical PLS-NIPALS as it proceeded without deflation of **X** and **y** and fewer matrices (to find **b**) are used [16].

PLS-Bidiagonalization

Basically, this advanced algorithm is started by decomposing **X** into three matrices [11]:

$\mathbf{X}=\mathbf{URV}^t$

Where, $\mathbf{U}(I\times J)$ and $\mathbf{V}(I\times J)$ are matrices with orthonormal columns (i.e., $\mathbf{U}^t\mathbf{U}=\mathbf{V}^t\mathbf{V}=1$) and $\mathbf{R}(J\times J)$ is the bidiagonal matrix. It is imprint to mention that in the earlier algorithms all columns in the generated matrices must be orthogonal (i.e., $\mathbf{W}^t\mathbf{W}=\mathbf{P}^t\mathbf{P}=0$). Once the **U**, **R** and **V** matrices are estimated with the optimum PLS-variables, the calibration vector is estimated as [11]:

$\mathbf{b}=\mathbf{VR}^{-1}\mathbf{Uy}$

PLS-Kernel

There are two common variants of Kernel algorithm so far [11, 16, 18, 19]. The first one can handle matrices of many samples, i.e. I is larger than J ($I>>J$), and the other one (which is suitable for the current drug system) was proposed for many variables **X**-matrices ($J>>I$). In all kernel algorithms, condensed matrices are created from **X** and **Y** (or **y**) which is an essential step. In the adopted algorithm, two condensed matrices are created $\mathbf{XX}^t$ and $\mathbf{YY}^t$ or $\mathbf{yy}^t$. Kernel matrix is then estimated as: $\mathbf{XX}^t\mathbf{YY}^t$. The main steps of the algorithm are [19]:

1. The eigenvector of the kernel matrix is taken as the first **X** score vector $\mathbf{t}_1$. The Y score vector is then estimated as: $\mathbf{u}_1=\mathbf{YY}^t\mathbf{t}_1$
2. The next step is to update the association matrices by eliminating the explained variable as follows:

$\mathbf{G}_1=\mathbf{I}-\mathbf{t}_1\mathbf{t}_1^t$ (I identity matrix)

$\mathbf{X}_1\mathbf{X}_1^t=\mathbf{G}_1\mathbf{XX}^t\mathbf{G}_1$

$\mathbf{Y}_1\mathbf{Y}_1^t=\mathbf{G}_1\,\mathbf{YY}^t\,\mathbf{G}_1$

The above operations save us from going back to the original large matrices and calculation of association matrices which are necessary at the start of the algorithm. As can be seen, the matrices

involved in Kernel algorithm have lower dimensions than the original matrices.

3. The next **t** and **u** vectors are estimated as outlined above using the updated matrices. The calibration matrix (containing the calibration vectors for the target solutes) are estimated from weight and loading matrices (**W, P** and **Q**) as follows:

$\mathbf{W} = \mathbf{X}^t\mathbf{U}$

$\mathbf{P} = (\mathbf{T}^t\mathbf{X})(\mathbf{T}^t\mathbf{T})^{-1}$

$\mathbf{Q} = (\mathbf{T}^t\mathbf{Y})(\mathbf{T}^t\mathbf{T})^{-1}$

Step 3 is repeated until the optimum number of PLS-variables is estimated.

It should be mentioned that all vectors in **W** should be normalized before creating the **b** vector [18, 19]:

$\mathbf{b} = \mathbf{W}(\mathbf{P}^t\mathbf{W})^{-1}\mathbf{Q}^t$

The solutes could be predicted from the spectrum $\mathbf{a}_{un}$ of the sample as follows:

$\mathbf{C}_{un} = \mathbf{a}_{un}\mathbf{b}$

De Luca wavelength selection method

The main steps of the De Luca method [8] are:

1. Firstly, the optimal number of factors (A) for components in the mixture is found.

2. Secondly, regression vectors for every component using A are estimated from PLS - algorithms: $\mathbf{B} = \mathbf{W}(\mathbf{P}^t\mathbf{W})^{-1}\mathbf{Q}^t$.

The regression coefficients for every component have different values at each wavelength:

$$C = b_0 + b_1\lambda_1 + b_2\lambda_2 + \cdots + b_n\lambda_n$$

where C is the analyte concentration, b the regression coefficients and λ is the wavelengths.

3. The sum of the absolute values of regression coefficients gives a new vector called curve of cumulative coefficients (B):

$$B_{\lambda i} = \sum_{i=1}^{n} |b_i|$$

4. Finally, the mean of cumulative coefficients is estimated and so the cutoff values are obtained:

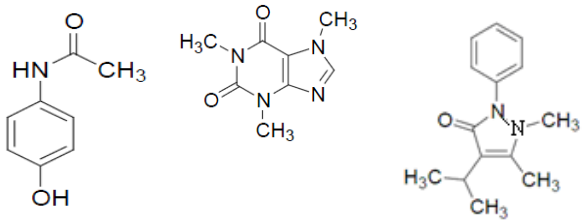
$$\bar{B} = \frac{1}{n}\sum_{i=1}^{n} B_i$$

5. The appropriate wavelength range includes the crossing points between cutoff value ($\bar{B}$) and the curve of cumulative coefficients (B).

EXPERIMENTAL

Chemicals and reagents

The drugs (Paracetamol, Propyphenazone and Caffeine) as shown in Fig 1 with maximum purity (> 99%) were purchased from Sigma–Aldrich. A 100.0 mg/L standard solution of each drug was prepared by dissolving 100 mg (±0.0001 g) from the corresponding pure materials in doubly distilled water in a 1.0 L volumetric flask. Due to the modest solubility of PAR and PRO in water, the initial solutions were mildly heated (50 °C). The calibration mixtures (22 solutions) and validation mixtures (12 solutions) were directly prepared from the stock solutions by appropriate dilution using distilled water.



Paracetamol Caffeine Propyphenazone

Fig. 1. Structural formulae of the drugs

Apparatus and software

The absorbance measurements were obtained using a quartz cuvette of 1.0 cm optical path, by a HP8452A diode-array UV–Vis spectrometer. The spectra of drugs were recorded over the wavelength range of 190–300 nm and the digitalized absorbance values were exported to MATLAB® for further analysis. Numerical solutions are calculated using MatLab®2013a (The Mathworks, Natick, MA, USA). PLS-variants (NIPALS, SIMPIS, Kernel and Bidiagonalization) were runned using home-made matlab® codes based on the algorithms outlined in the former section. The matlab® codes are available upon request from the authors. Independently, mathematical calculations were carried out using MVC1® [14] and TOMCAT® [15] programs which are freely available. Cross-validation procedure was carried out using cross-validation.m function from TOMCAT®, which is modified to be applicable for SIMPLS, Kernel and Bidiagonalization algorithms.

Calibration and validation sets

There are many strategies for building a calibration set which is necessary to run the calibration model. In pharmaceutical analysis, full factorial design is often adopted to end up with accurate multivariate calibration analysis. For *n* concentration levels for *k* solutes, the number of calibration mixtures *I* that should be prepared is n^k. Finally, a large reduction in the number of mixtures is obtained by applying Brereton`s table [22]. It is proposed for multilevel multifactor (multi-solute) systems. According to Brereton`s design, the number of mixtures is $I = n^2$ and the maximum

number of analyzed solutes is *I*-1. In our work, five-level full factorial design was adopted. Following this design, 25 mixtures should be prepared and up to 24 solutes analyzed [22]. In this work, 22 mixtures are found representative and prepared according to the levels provided in Table 1. The levels of drugs in the calibration set were randomly selected. As can be seen from Table 1, the levels of the drugs were carefully chosen to be comparable to those present in the marketed formulation (250 mg PAR-150 mg PRO-50 mg CAF per tablet). For PAR and PRO, the selected levels are 0, 4.0, 6.0, 8.0, and 10.0 mg/L, while for CAF, the levels were 0, 1, 2, 4, and 8 mg/L. For the three drugs, the concentration levels were coded as: -2 for the lower concentration and +2 for the higher concentration. In general, the design would show that the levels of drugs are orthogonal and span a large variation in levels. The orthogonal design is necessary for building a stable and robust PLS model. A 12-mixture validation set was prepared in order to test the performance of the models. For a validation set, new concentration levels were created for the three solutes, and the prediction power of the PLS model was tested for drugs levels outside the ranges selected in the calibration set. For example, 12.0 mg/L (for PAR and PRO) and 16.0 mg/L (for CAF) were selected and these levels are out of the calibration range.

Preparation of the formulation before analysis

The drugs were safely extracted from SARIDON following a simple procedure. Pharmaceutical formulations were assayed by weighing the content of five tablets, grinding to a fine powder, and storing in a cold place. A great care was taken to ensure safe extraction of the three drugs from the complex extract of the formulation. Extraction by hot water was found useful, as will be shown later. An amount exactly corresponding to the average tablet weight was suspended in water or hot water (50 ^{o}C) and made up to a volume of 1000 ml. The final suspension was sonicated for 10 min and then filtered through a PTFE 0.45 μm membrane filter. Samples were scanned after proper dilution using a spectrophotometer over the spectral range 190-300 nm. The initial examination of the spectra indicated the existence of co-extracted components (mainly excipients) which needs application of multivariate calibration for more accurate quantification. Derivative spectroscopic methods were found of limited application for this complex system.

Table 1. Composition of calibration and validation sets

	Calibration set (mg/L)			Validation set (mg/L)		
Number	Paracetamol	Propyphenazone	Caffeine	Paracetamol	Propyphenazone	Caffeine
1	4.0 (-1)	6.0 (0)	2.0 (0)	12.0	6.0	2.0
2	6.0 (0)	6.0 (0)	2.0 (0)	10.0	12.0	2.0
3	8.0 (1)	6.0 (0)	2.0 (0)	10.0	6.0	8.0
4	4.0 (-1)	6.0 (0)	2.0 (0)	10.0	6.0	16.0
5	8.0 (1)	0 (-2)	0 (-2)	0	0	8.0
6	10.0 (2)	4.0 (-1)	0 (-2)	8.0	6.0	1.6
7	10.0 (2)	6.0 (0)	2.0 (0)	6.0	4.8	1.6
8	10.0 (2)	8.0 (1)	4.0 (1)	6.0	4.0	0.8
9	10.0 (2)	10.0 (2)	8.0 (2)	8.0	4.8	1.6
10	0 (-2)	8.0 (1)	0 (-2)	0	4.8	1.6
11	10.0 (2)	6.0 (0)	1.0 (-1)	8.0	4.8	0
12	10.0 (2)	6.0 (0)	2.0 (0)	8.0	0	1.6
13	10.0 (2)	6.0 (0)	4.0 (1)			
14	10.0 (2)	6.0 (0)	0 (-2)			
15	0 (-2)	6.0 (0)	2.0 (0)			
16	10.0 (2)	0 (-2)	2.0 (0)			
17	0 (-2)	0 (-2)	2.0 (0)			
18	10.0 (2)	6.0 (0)	1.0 (-1)			
19	10.0 (2)	6.0 (0)	2.0 (0)			
20	8.0 (-1)	0 (-2)	0 (-2)			
21	0 (-2)	8.0 (1)	0 (-2)			
22	0 (-2)	0 (-2)	8.0 (2)			

a. Five-level full factorial design according to orthogonal Brereton`s design [22]. Values in parentheses are the codes necessary for building orthogonal mixtures.

RESULTS AND DISCUSSION

Spectral overlapping

The UV absorption spectra of the three drugs along with the typical SARIDON extract are given in Fig 2.

As indicated in Fig 2, the drugs exhibited strong UV absorption over the studied range (190-300 nm). Both PAR and CAF showed their typical spectra with suitable wavelengths for detection at 244 and 274 nm, respectively. In fact, the absorption spectrum of PRO has irregular shape where the drug has stable absorption over the range 210-270 nm. The spectra of the drugs are highly overlapped over the studied range. Net analytical signal (NAS) is a suitable method to characterize the analytical figures of merit and spectral overlapping related to the multivariate calibration [23]. NAS analysis indicates that PAR, PRO and CAF would be detected down to 0.21, 0.37 and 0.15 mg/L^{-1} respectively. On the other hand, NAS calculations indicate a high degree of spectral overlap with other signals for PRO (80% spectral overlap). PAR and CAF showed lower spectral overlapping (71 and 55%). The absorption spectrum of the mixture indicated the additive nature of the individual signals of the drugs and the linearity of the current system. In fact, the experimental spectrum and the one estimated from the single-drug spectra were almost identical indicating the additive nature of the generated signals. The claimed values of the three drugs in the formulations indicated that PAR and PRO are more dominant than CAF (250 mg PAR, 150 mg PRO and 50 mg CAF per one tablet). This fact clearly reflected on the shape of the recorded spectrum (dilution factor 1:25) of drug extract where the final shape is very close to that of PAR. The similarity of the extract spectra with PAR would indicate that the extraction procedure was effective as the drugs were selectively obtained among other constituents like excipients.

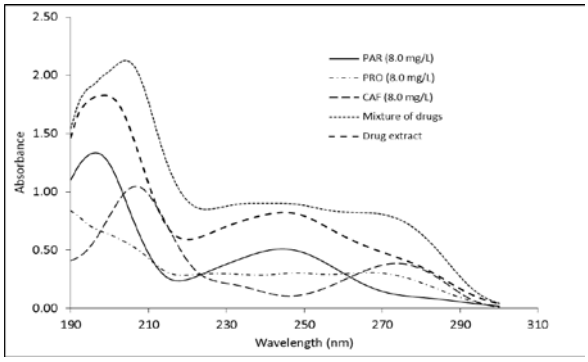


Fig. 2. Absorption spectra of the studied ternary drug system

In fact, simultaneous determination of the ternary drug system is not possible by univariate calibration which is due to intense spectral overlapping. In our view, the main analytical problem in the current system is the intense spectral overlap where the influence of unexpected interference is also high, as indicated from the spectrum of the drug extract.

In this study, three matrices were created: calibration matrix (22×56), validation matrix (12×56) and matrix of drug extracts (18×56).

Selection of the optimum spectral ranges before PLS calibration

In fact, the performance of multivariate calibration would be improved if calibration is carried out over certain informative ranges instead of the entire spectrum. There are many procedures for selecting those informative regions prior to multivariate calibration, including statistical analysis related to the external validation [24], genetic algorithm [25], changeable size moving-window [26], and De Luca's procedure [8]. In the latter procedure, the absolute values of the calibration vector **b** (obtained by the PLS model) for the three solutes were numerically summed to find the cumulative coefficient B. Another important line called cutoff line is estimated. This line is necessary for the final selection of the spectral ranges of the studied drugs. The De Luca plot is given in Fig 3.

As indicated in the plot, the best spectral ranges which are included in the calibration are those located under the De Luca line and above the cutoff line. The best regions for analyzing drugs are: 202-220, 234-254, and 266-284 nm. In fact, the purpose of the cutoff line was to help the analyst to select or locate the informative spectral ranges [8].

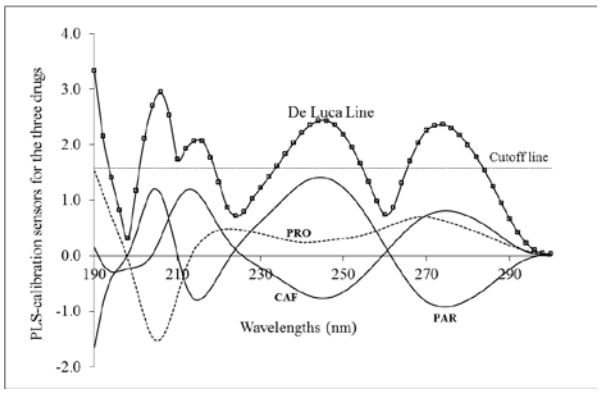


Fig. 3. De Luca plot generated from PLS-calibration

Table 2. Prediction of drugs in the validation set before and after applying De Luca`s procedure for selecting the informative spectral regions.

	Spectral region (nm) 190-300 nm (56 points/spectrum)[a]			De Luca wavelength selection (31 points/spectrum)[a]		
	PAR	PRO	CAF	PAR	PRO	CAF
PLS-NIPALS variables	8	10	7	8	6	7
REP%[b]	5.9	14.3	12.2	4.9	7.9	6.7
RMSEP[b]	0.45	0.78	0.40	0.37	0.67	0.39
Mean Recovery[b]	98.5	90.7	96.1	99.0	90.8	96.4

a. Based on De Luca procedure, the optimum spectral regions for drugs are: 202-220, 234-254, and 266-284 nm (31 spectral points/sample).

b. Statistical analyses were applied to the non-zero concentration of the drugs. REP%: Relative error of prediction and RMSEP: root mean square error of prediction.

The regions that left over would account for the serious overlapping between drugs (220-230 nm, 258-266 nm, as shown in Fig 3). Although the first region (190-194 nm) would be included in the regression, it was excluded, as it is close to the extreme limit of the instrument. One more interesting point in the procedure is including the maximum wavelengths of absorption of drugs like 244 nm for PAR and 208/274 nm for CAF. It was interesting to notice that 196 nm (a significant wavelength for PAR) was excluded by the procedure. Now, the numerical analysis by PLS (NIPALS variant) was repeated to quantify the drugs in the validation set using selected regions. The main results are compiled in Table 2.

Application of De Luca method for selecting the informative spectral regions has improved PLS regression in certain aspects. For PRO, the number of PLS-variables was reduced from 10 to 6 which will reflect on the computation time. Moreover, the prediction power of the model was improved by applying De Luca procedure. Another important method that would be used for finding the optimum spectral ranges is the bootstrap method [16]. In this method, bootstrap sampling is used to estimate the standard errors in the PLS-calibration vectors **b** (for each drug) and from these errors an assessment of each explanatory variable on modeling **y** is carried out. Usually 1000 bootstrap samples are enough to estimate the standard error in **b**. From bootstrap samples (taken from **X** and **y**), the standard errors of b_j, j= 1,…,A (σ_{bj}) are estimated and are used to calculate the standardized coefficients b_j / σ_{bj}. The standardized coefficient larger than 1.96 is considered important at 0.05 significant level meaning this variable is essential for modeling **y**. The bootstrap procedure was repeated for each drug. Bootstrap analysis indicated that the best spectral regions for PAR are 212-220, 230-252, and 258-282 nm. For PRO: 190, 194-196, 206, 212-228, 232-242, 248-280 nm. For CAF: 190, 194-198, 204-206, 212-228, 244-272, and 282-290 nm. For PAR, the results of De Luca were comparable to the bootstrap ranges. However, for the other two drugs, more spectral regions were obtained by bootstrap compared to De Luca. Accordingly, PLS-NIPALS calibration was repeated using the regions of bootstrap and the optimum PLS-variables were also estimated by the cross-validation technique. The overall results were comparable to those obtained by De Luca/PLS-NIPALS.

Comparison between PLS variants for drugs quantification

MVC1 and TOMCAT are the most commonly used software package in the pharmaceutical analysis [14,15]. MVC1 uses only NIPALS [14], but in TOMCAT, except for NIPALS, there are additional two variants of the SIMPLS algorithm: the WIM-PLS and SIM-PLS algorithms [15]. Nine PLS-algorithms are available for handling different kinds of data, but application of Kernel and Bidiagonalization is rather limited [11].

For the current drug system, the size of the **X** matrix is 22×56. From a practical point of view, the size of **X** is an adjustable parameter and is dependent on the system under investigation.

The current analytical system was subjected to different PLS-variants and the assessment strategy was based on two items: a) number of PLS-variants needed to build the model, and b) the prediction power of different PLS-variants. Model of lower PLS-variables with better prediction is the best choice. Other important criteria including computational time and memory-storage were not investigated [16].

For each drug, PLS-variables needed for calibration (A) were estimated by the leave-one-out method. The final PRESS-PLS-variable plots are shown in Fig 3a, 3b and 3c and the performance for drug prediction is summarized in Table 3.

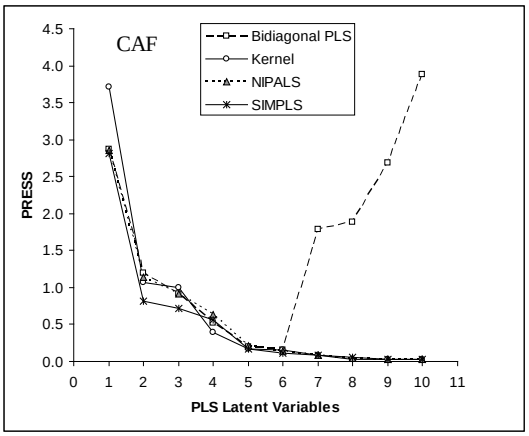


Fig 3 a) PRESS-Latent variables plots for different PLS-variants as obtained by cross-validation technique for CAF

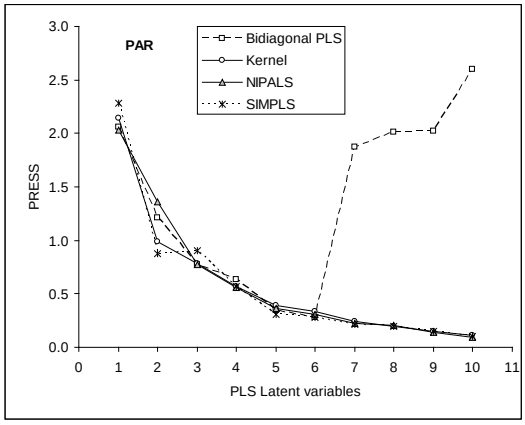


Fig 3 b) PRESS-Latent variables plots for different PLS-variants as obtained by cross-validation technique for PAR

In fact, typical PRESS-Latent variable plots were obtained for all variants (except Bidiagonalization). In all cases, a large reduction in PRESS is observed with increasing variables. Generally speaking, from 6 to 10 variables are needed in all variants to explain the variances in the data. Except for PLS-Bidiagonal, the models exhibited a stable performance at high latent variables. The interesting point in Fig 3 is the abrupt jump in PLS-Bidiagonalization behavior where PRESS is suddenly increased at 7 variables and this is true for the three drugs. The optimum number of factors needed for optimum prediction is shown in Table 3.

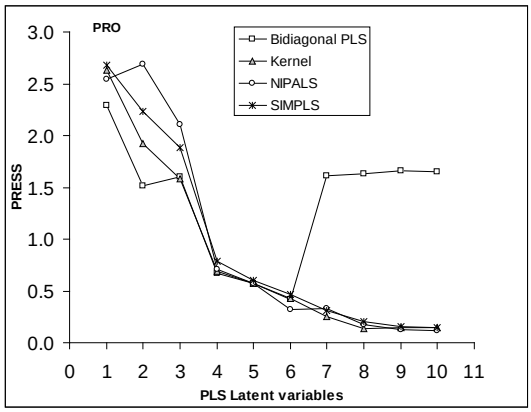


Fig 3 c) PRESS-Latent variables plots for different PLS-variants as obtained by cross-validation technique for PRO

The performance of PLS-variants was comparable for PAR prediction in the validation set with an overall recovery of 97.2-99.7 and excellent REP% 1.6-4.9. Although the NIPALS model used fewer variables (compared to SIMPLS and Kernel) for prediction, the model needs longer computation time, particularly for many variables-X matrices [16]. For all variants, poor prediction was observed for PRO and this is attributed to intense spectral overlap of this drug with other components. Due to its unstable performance and poor prediction for PRO (16.2%) and CAF(8.5%), PLS-Bidiagonalization (see Fig 3 c) was not appropriate algorithm. Kernel model, in fact, outperformed other variants for PAR, PRO and CAF prediction which is expected, as this algorithm is perfectly designed for the current analytical system. Compared to the other variants, PLS-kernel is known for its shorter computation time and less computer storage [18,19].

Table 3. Prediction of drugs in the validation set using several PLS-variants

PLS-variant[a]	PAR				PRO				CAF			
	A[b]	Mean recovery	REP%	RSD	A	Mean recovery	REP%	RSD	A	Mean recovery	REP%	RSD
NIPALS	8	99.0	4.9	3.4	6	90.8	7.9	2.9	7	96.4	6.7	4.0
SIMPLS	9	97.4	3.2	2.9	10	89.0	12.5	3.5	9	90.3	11.8	2.8
Kernel	9	97.2	3.3	2.8	8	95.5	5.4	2.5	9	96.4	5.5	2.7
Bidiagonal	6	99.7	1.6	1.1	6	82.0	16.2	6.5	6	109.8	13.7	8.5

a. See "Theoretical background" section for more details on the algorithms.
b. PLS-factors needed for optimum modeling was estimated using cross-validation technique [21].

Drugs quantification in SARIDON formulation and comparison with other methods

Although PLS-variants were workable for simultaneous analysis of PAR-CAF-PRO in their pure mixtures, the performance of the models was further tested for SARIDON® formulation. In real preparations, excipients are present along with the drug component which may negatively interfere with drugs quantification by PLS-regression. The extraction experiments indicated that distilled water is an efficient extractant for the three drugs. Accordingly, the consumption of expensive and toxic organic solvents is avoided. Moreover, hot water (50 °C) was also tested for drugs extraction. After extraction, the extracts were directly scanned by UV and the obtained spectra were analyzed by PLS calibration. As Kernel model was the optimum model for drugs prediction, the model was used for drugs quantification in real extracts. The overall results are summarized in Table 4.

The analytical performance of the Kernel-PLS method should be tested against a standard analytical procedure to asses the overall accuracy and precision. The levels of drugs in the marketed formulation were accurately measured by an independent chromatographic procedure and this was carried out by the manufacturer at earlier stages. The following main conclusions are drawn from Table 4: a) the proposed Kernel-PLS is workable for quantifying drugs with stable performance. No advanced cleaning procedures like solid phase or liquid-liquid extraction are applied in the current case and no chromatographic procedures are applied; b) extraction by hot water (50 °C) clearly ended up with better results and this is attributed to the enhanced solubility of drugs at higher temperature. Extraction at still higher temperatures is not recommended due to the unexpected influences on the chemical structure of drugs; c) the estimated recoveries ± RSD (96.4 (1.3), 95.1 (2.5) and 96.2 (3.0) for PAR, PRO and CAF, respectively) reflected the applicability of the Kernel-PLS method for drugs quantification. In fact, many analytical methods were proposed for PAR-PRO-CAF quantification in commercial pharmaceutical formulations. The reported methods extended from laborious matrix cleaning-liquid chromatography to non-separative ones including partial least squares PLS calibration.

For further assessment, the current method was compared with published methods as summarizes in Table 5.

Most of the reported methods are using either ethanol or methanol for drugs extraction, however, other aqueous solvents are also applicable. For chromatographic methods, the main steps of extract-cleaning were centrifugation and filtration. In general, all chromatographic procedures were found efficient for drugs quantification with acceptable accuracy and precision. The best detection of the ternary system was reported by Soponar and co-workers [3]. Using micellar electrokinetic capillary chromatography with DAD detection, the drugs were quantified in SARIDON® with very low detection limits of 42, 194, and 74 ng/ml for PAR, PRO and CAF, respectively [3].

Table 4. Quantification of active ingredients in marketed SARIDON® formulation by Kernel-PLS

Extraction [a]	PAR			PRO			CAF		
	Content (mg/ tablet)	Rec % [b]	RSD [b]	Content (mg/ tablet)	Rec %	RSD	Content (mg/tablet)	Rec %	RSD
water extraction (25 °C)									
S1	239.3			127.8			44.0		
S2	229.8	93.8	1.7	137.5	86.7	6.1	34.3	80.2	10.4
S3	233.8			120.0			42.0		
S4	235.2			135.0			40.0		
Hot water extraction (50 °C)									
S1	240.0			144.1			48.5		
S2	241.0	96.4	1.3	140.5	95.1	2.5	46.5	96.2	2.9
S3	237.5			139.0			48.5		
S4	245.0			146.8			49.3		

a. In all extractions, 5.0 g tablets of SARIDON® were grinded and a mass equivalent to one tablet was directly extracted with water. The extract was centrifuged and finally filtered through a 0.45 μm filter. Before spectral analysis, 1:25 dilution was carried out with distilled water. Four identical extractions were carried out (n=4).

b. The overall accuracy (mean recovery) and precision (RSD) were estimated against the actual or claimed values (250 mg PAR, 150 mg PRO and 50 mg CAF) provided by the manufacturer.

Table 5. Comparison of the analytical characteristics of Kernel-PLS with published methods for PAR-PRO-CAF quantification in commercial pharmaceutical formulations

Formulation (mg/tablet)	Extraction solvent	Matrix purification	Analytical technique	LOD (mg L^{-1})	Spike Rec. (%)	RSD (%)	Ref.
Chromatographic methods							
Not provided PAR 250 PRO 150 CAF 50	Methanol	Centrifugation	HPLC-DAD	Not provided	PAR 100.2 PRO 99.8 CAF 99.3	PAR 0.2 PRO 0.2 CAF 0.2	27
Not provided PAR 250 PRO 150 CAF 50	Methanol	Centrifugation	HPLC-DAD	Not provided	PAR 102.4 PRO 97.5 CAF 99.6	PAR 1.4 PRO 1.2 CAF 0.8	28
Not provided PAR 250 PRO 150 CAF 50	Distilled water	Centrifugation and membrane filtration	Micellar electrokinetic capillary chromatography-MECK-DAD	PAR 0.6 CAF 0.8 PRO 0.8	PAR 100.3 PRO 99.9 CAF 100.0	PAR 0.2 PRO 0.2 CAF 0.4	29
SARIDON® PAR 250 PRO 150 CAF 50	Methanol	Filtration by pore-size filter paper (no centrifugation)	HPLC-UV (other related drugs were also detected)	PAR 0.042 PRO 0.194 CAF 0.074	97.2-102.3	0.5-1.1	3
Pararemin® PAR 250 PRO 150 CAF 50	Methanol	Filtration by pore-size filter paper (no centrifugation)	HPLC-UV (other related drugs were also detected)	PAR 0.042 PRO 0.194 CAF 0.074	97.9-101.1	1.8-2.6	3
Minoset® PAR 250 PRO 150 CAF 50	Methanol	Filtration and dilution	HPLC-UV	PAR 0.30 PRO 0.25 CAF 0.36	PAR 100.0 PRO 100.0 CAF 100.0	PAR 0.4 PRO 0.6 CAF 1.8	5
Non-Chromatographic methods							
Minoset® PAR 250 PRO 150 CAF 50	0.1 M HCl	Filtration and dilution	Derivative UV-Spectrophotometry	PAR 0.29 PRO 0.35 CAF 0.10	PAR 99.8 PRO 100.1 CAF 99.0	PAR 0.3 PRO 0.5 CAF 1.6	5
Minoset plus® PAR 250 PRO 150 CAF 50	gastric juice solution	Shaking, and filtration by 0.20pm membrane filter	Multivariate calibration PLS-NIPALS	Not provided	PAR 100.0 PRO 99.9 CAF 101.6	PAR 0.7 PRO 1.2 CAF 2.7	7
NeoOptalidon® PAR 200 PRO 125 CAF 25	Ethanol	Sonication and filtration by 0.45pm membrane filter. Final extract diluted (1000 time) by distilled water	Multivariate calibration PLS-NIPALS (absorbance)	Not provided	PAR 105.1 PRO 105.5 CAF 126.9	PAR 3.75 PRO - CAF -	8
SARIDON® PAR 250 PRO 150 CAF 25	Ethanol	Sonication and filtration by 0.45pm membrane filter. Final extract diluted (1000 time) by distilled water	Multivariate calibration PLS-NIPALS (absorbance)	Not provided	PAR 103.1 PRO 106.6 CAF 129.5	PAR 3.75 PRO - CAF -	8
Veramon® PAR 200 PRO 285	Ethanol	Sonication and filtration by 0.45pm membrane filter. Final extract diluted (1000 time) by distilled water	Multivariate calibration PLS-NIPALS (absorbance)	Not provided	PAR 95.9 PRO 88.5	PAR 3.75 PRO -	8
SARIDON® PAR 250 PRO 150 CAF 50	Hot water (50 ^{0}C)	Shaking, centrifugation, and filtration by 0.45 pm membrane filter	Multivariate calibration PLS-Kernel (absorbance spectra)	PAR 0.21 PRO 0.37 CAF 0.15	PAR 96.4 PRO 95.1 CAF 96.2	PAR 1.3 PRO 2.5 CAF 2.9	This work

The drugs were quantified after separation by C_{18} column with excellent recoveries 97.2-102.3 and precision 0.5-1.1 [3]. In general, the reported chromatographic methods manifested an excellent analytical performance for quantifying commercial tables containing PAR, PRO and CAF [5, 27-29].

As can be seen from table5, chromatographic procedures have shown a better precision compare to multivariate calibration methods. Interestingly and as depicted in this table , multivariate calibration methods do not apply any extra purification or preconcentration step and

manifested a comparable stable analytical performance to chromatographic methods. As indicated in Table 5, the proposed multivariate calibration has achieved better detection for drugs without using advanced instruments. Although derivative spectrophotometry was workable for analyzing PRO, CAF and PRO in commercial tablet (Minoset®), the overall precision seems to questionable [5]. As shown in Table 5, PLS-NIPALS is the most adopted algorithm for drugs quantification in commercial tables. Indeed, PLS-NIPALS was effective for analyzing the current ternary-drug system in different commercial formulations with optimum accuracy and precision. Dinç and co-workers showed that the results obtained by derivative spectrophotometry and liquid chromatography for PAR-CAF-PRO quantifications are statistically comparable [5]. De Luca and co-workers showed that application of PLS NIPALS on absorbance spectra gave good recovery only for PAR, but for PRO and CAF satisfactory quantifications were obtained using third derivative spectra [8]. Our results demonstrate that recovery can be improved especially for CAF using only absorbance spectra and Kernel-PLS algorithm. Besides the reported detection limits, accuracy, and precision of Kernel-PLS method is of comparable quality to laborious and time-consuming chromatographic methods. Another advantage of the proposed method is that no extensive matrix-cleaning procedures are adopted.

CONCLUSIONS

The following main conclusions are deduced from the current pharmaceutical-chemometric study:

- Modeling many variables-X matrices is less time-consuming and very safe using Kernel-PLS method.
- Selection of informative spectral regions by De Luca`s method has improved the overall regression which was also in agreement with bootstrap method.
- Extraction of the ternary drug mixture by hot water instead of using ethanol or other toxic organic solvents is a good practice.
- The reported recoveries±RSD [96.4 (1.3), 95.1 (2.5) and 96.2 (3.0) for PAR, PRO and CAF, respectively] proved the applicability of Kernel-PLS method for drugs quantification in commercial SARIDON®.

The proposed multivariate calibration procedure is applicable for other formulations and, in the same time indicated the adequacy of Kernel-PLS for pharmaceutical analysis.

REFERENCES

1. M.R. Khoshayand, H. Abdollahi, M. Shariatpanahi, A. Saadatfard, A. Mohammadi, *Spectrochim. Acta A*, **70**, 491 (2008).
2. H.G. Kraetsch, T. Hummel, J. Lötsch, R. Kussat, G. Kobal, *Eur. J. Clin. Pharmacol.*, **49,** 377 (1996).
3. F. Soponar, D. Staniloae, G. Moise, B. Szaniszlo, V. David, *Rev. Roum. Chim.*, **58**, 433 (2013).
4. J.R. Migliardi, J.J. Armellino, M. Friedman, D.B. Gillings, W. T. Beaver, *Clin. Pharmacol. Ther*., **56**, 576 (1994).
5. E. Dinç, G. Kökdil, F. Onur, *J. Pharm. Biomed. Anal*., **26**, 769 (2001).
6. Y.S. Al-Degs, A.H. El-Sheikh, A.A. Issa, M. Al-Ghouti, M. Sunjuk, *Water Sci. Technol.*, **66**, 1647 (2012).
7. E. Dinc, C. Serin, F. Tugcu-Demiroz, T. Doganay, *Int. J. Pharm.*, **250**, 339 (2003)
8. M. De Luca, F. Oliverio, G. Ioele, G. Ragno, *Chemometr. Intel. Lab*., **96**, 14 (2009).
9. P. C. Damiani, G. M. Escandar, A. C. Olivieri, H. C. Goicoechea, *Current Pharm. Anal*., **1**, 145 (2005).
10. J.A. Martins, R.F. Teófilo, M.M. Ferreira, *J. Chemom*., **24**, 320 (2010).
11. M. Andersson, *J. Chemom.*, **23**, 518 (2009).
12. H.W. Darwish, I.A. Naguib, *Drug Test. Anal.*, **5**, 325 (2013)
13. H.W. Darwish, S. A. Hassan, M.Y. Salem, B. A. El-Zeany, *Spectrochim. Acta A*, **122**, 744, (2014)
14. A.C. Olivieri, H.C. Goicoechea, F.A. Iñón, *Chemom. Intell. Lab. Syst*., **73**, 189 (2004).
15. M. Daszykowski, S. Serneels, K. Kaczmarek, P. Van Espen, C. Croux, B. Walczak, *Chemom. Intell. Lab. Syst*., **85**, 269 (2007).
16. A. Alin, *Stat. Papers*, **50**, 711 (2009).
17. S. de Jong, *Chemom. Intell. Lab. Syst*., **18**, 251 (1993).
18. S. Rännar, F. Lindgren, P. Geladi, S. Wold, *J. Chemom.*, **8**, 111 (1994).
19. S. Rännar, P. Geladi, F. Lindgren, S. Wold, *J. Chemom.*, **9**, 459 (1995).
20. B.Dayal, J.MacGregor. *J. Chemom.*,**11**, 7 (31997).
21. D.M. Haaland, E.V. Thomas, *Anal. Chem.*, **60**, 1193 (1988).
22. R. Brereton, *Analyst*, **122**, 1521 (1997).
23. S. Yousefinejad, B. Hemmateenejad, *Drug Test. Anal.*, **4,** 507 (2012).
24. M.S. Linares, J.M.G. Fraga, A.I. Jimenez, F. Jimenez, J.J. Arias, *Anal. Lett*., **32**, 2489 (1999).
25. J. Ghasemi, A. Niazi, R. Leardi, *Talanta*, **59**, 311 (2003).
26. Y.P. Du, Y.Z. Liang, J.H. Jiang, R.J. Berry, Y. Ozaki, *Anal. Chim. Acta*, **501**, 183 (2004).
27. M.G. Mamolo, L. Vio, V. Maurich, *J. Pharm. Biomed. Anal*., **3**, 157 (1985)
28. A.M. Di Pietra, R. Gatti, V. Andrisano, V. Cavrini, *J. Chromatogr. A*, **729**, 355 (1996).
29. D. Emre, N. Özaltýn, *J. Chromatogr. B*, **847**, 126 (2007).

ОЦЕНКА НА РАЗЛИЧНИ PLS АЛГОРИТМИ ЗА КОЛИЧЕСТВЕНО ОПРЕДЕЛЯНЕ НА ТРИ СПЕКТРАЛНО ПРИПОКРИВАЩИ СЕ ЛЕКАРСТВА

Г. Гергов[1] *, А. Алин[2], М. Дойчинова[3], М. Де Лука [4], В. Симеонов[5] , Я. Ал-Дегс[6]

[1]*Катедра „Химия“, Фармацевтичен факултет, Медицински университет София, 1000, ул. „Дунав“2, София, България.*

[2]*Катедра „Статистика“, Университет „Девети септември“, Измир, Турция*

[3]*Институт по инженерна химия, Българска Академия на Науките, ул. „Акад. Г. Бончев“, бл. 103, София, България*

[4]*Катедра „Фармация , Здравни науки и хранене“, Калабрийски университет, 87036 Ренде, Италия*

[5]*Лаборатория „Хемометрия и екометрия“, Факултет по химия и фармация, Софийски университет “Св. Климент Охридски”, 1164 София, бул. Дж. Баучър 1, София, България.*

[6]*Катедра „Химия“, Хашемитски университет, П.К. 150459, Ез-Зарка 13115, Йордания*

Получена 20 юли, 2016 г.; коригирана на 26 септември, 2016 г.

(Резюме)

Основната цел на настоящото изследване е да се сравни прогнозната мощ на различни PLS алгоритми, прилагани за количествено определяне на три спектрално припокриващи се лекарства. Бяха избрани четири варианта на PLS за многовариационно калибриране и предсказване на трите компонента на лекарствената форма (парацетамол, пропифеназон и кофеин). Алгоритмите NIPALS и SIMPLIS са най-често срещаните и използвани алгоритми. Другите тествани алгоритми са Kernel и Bidiagonalization, които се прилагат сравнително рядко във фармацевтичния анализ.

Данните от първи ранг бяха генерирани чрез измерване на UV-спектрите на лекарствените смеси в диапазона 190-300 nm с разделителна способност от 2,0 nm (т.е. 56 спектрални точки / проба). Използван бе редуциран пълен факториален дизайн на пет нива и прогнозната мощност на PLS вариантите беше тествана за концентрационни нива на лекарствата извън диапазоните, избрани в калибрирационната матрица.

Методът на Де Лука и потвърждаващия буутстрап метод бяха избрани като удачни за оптимална селекция на спектралните региони.

Външното валидиране използвайки калибрационния модел PLS-Kernel показа значителни предимства при анализа на таблетки SARIDON. Трите лекарства в SARIDON бяха количествено определени с аналитичен добив и прецизност 96.4 (1.3), 95.1 (2.5) and 96.2 (2.9) съответно за парацетамол, пропифеназон и кофеин. Това се оказва оптималния алгоритъм, който може успешно да се приложи за рутинен анализ на аналгетични и антипиретични многокомпонентни таблетки във фармацевтичната индустрия.