Regular Article

ANALYTICAL AND BIOANALYTICAL CHEMISTRY RESEARCH

Published by the Iranian Chemical Society

Anal. Bioanal. Chem. Res., Vol. 9, No. 4, 351-362, September 2022.

Application of Different Interval Variable Selectors for Quantification of Spectrally Overlapping Pharmaceuticals by Multivariate Calibration

Yahya Al-Degs*, Jafar Abdelghani, Ghadeer Al-Galab, Zaher Algharaibeh and Hameed Al-Zawahreh

Department of Chemistry, Faculty of Science, The Hashemite University, P. O. Box: 330127, Zarqa 13133, Jordan

(Received 29 January 2022 Accepted 17 April 2022)

The novel application of interval variable iterative space shrinkage approach *i*VISSA and partial least squares PLS calibration for quantification of three overlapping pharmaceuticals (paracetamol, guaifenesin, and phenylephrine) is presented in this work. In addition to spectral overlapping, the drugs are available in the commercial tablet in varying proportions where paracetamol and guaifenesin are 20 to 50 times higher than phenylephrine. Net analyte signal calculations indicated that pH has an high influence on drug overlapping, and the optimum pH was at 12.0, with a total overlapping 57-77% among solutes. To eliminate the influence of excipients on PLS calibration, all standard mixtures were prepared with a 5% level of excipients. For Guaifenesin, variable selection by NAS resulted in better PLS prediction. *i*VISSA, which developed under the condition that locations, widths, and combinations of selected intervals are optimized simultaneously, resulted in the excellent prediction of phenylephrine by selecting 34 spectral data in which the solute became intensely absorbed. Analysis of commercial tablet (250 mg paracetamol, 100 mg guaifenesin, 5 mg phenylephrine) with the help of NAS revealed that GUA was predicted with high accuracy (98.4%) and precision (RSD 3.9) using the range 225-239 nm. For phenylephrine, the selected intervals were 200-207, 224-225, 232-246, 250-254, 263, 267-270, 275, and 278-280 nm by *i*VISSA, which resulted in accurate quantification with high accuracy (104.3%) and precision (RSD 0.5). For paracetamol, including the entire range (200-300 nm, 101 points) was necessary for better PLS prediction while variables selection by NAS or *i*VISSA negatively affected PLS calibration for that drug. The accuracy and precision of the proposed method were validated against liquid chromatography and both methods were found statically comparable.

Keywords: Interval variable selection methods, Spectral overlapping, PLS calibration, Pharmaceutical analysis

INTRODUCTION

Multivariate calibration substantially contributes to analytical chemistry and is often applied in many fields, including natural products, the pharmaceutical industry, the environment, and food research [1]. Multivariate calibration depends on statistical and mathematical tools to create quantitative relations between measured variables (*i.e.*, UV-Vis, IR, NMR signals) and physicochemical properties of the system under study, such as density, viscosity or concentration [1]. The main challenge in using multivariate calibration is the higher number of measured variables p compared to the number of samples n. This is known as a non-deterministic polynomial-time dilemma in statistical analysis [2,3]. For example, a modern IR instrument can generate a spectrum within the 4000-400 cm^{-1} range with a step of 2 cm^{-1} yielding 1801 variables for the measured sample. Hence, the number of variables is notably larger than the number of samples. The high-dimensional data represent a problem that most traditional statistical methods cannot cope with [2-4]. The standard regression estimators, including PLS and PCR, would produce unstable coefficient estimates with inflated standard errors when used with high-dimensional data (*i.e.*, $p >> n$), resulting in reduced

*Corresponding author. E-mail: yahyaaldeqs@yahoo.com

statistical power and incorrect conclusions about relationships between independent and dependent variables [5]. Furthermore, when p exceeds n, traditional estimating methods simply cannot be used to produce model parameter estimates [5].

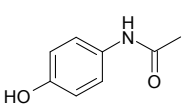
There are two strategies to deal with high-dimensional data: a) dimension reduction and b) variable selection [5,6]. The main aim of the earlier methods is to compress the dimensionality of the data by excluding the redundant variables while keeping the informative ones [5,6]. The primary function of dimension reduction methods is to represent the original variable space with a lower dimension space [6]. The most common dimension reduction method is partial least squares (PLS) which is based on latent variables and has many applications in different fields, including food analysis, spectroscopy, and many other research areas [1,7]. Although PLS is an efficient dimension reduction method, it suffers from weak interpretability and the inability to remove redundant/noisy variables [5,6]. The performance of PLS calibration is significantly improved after proper variables selection, and this issue was verified as outlined in the literature [5,6]. For spectroscopic methods, variables selection should identify informative wavelengths among the entire spectrum needed for adequate modelling [5]. Removing the uninformative variables can improve PLS prediction for new samples [8]. Accordingly, variable selection methods are significant for picking up the informative variables to end up with a safe dimension reduction of the collected variables. The appropriate selection of useful variables and removing the uninformative ones is essential due to: a) creating factor-based calibration methods (like PLS and PCR) using fewer numbers of factors, and b) enhancing the prediction output of the calibration model and especially for unknown samples [9,10].

Among spectral measurements, UV, Vis, and IR often contain many analytical details that should be carefully selected to get excellent PLS calibration. In addition, the appropriate selection of variables improves PLS performance toward spectrally overlapping analytical systems [11]. Variable-selection approaches are often categorized into a) single-variable selection approach which separately assesses the value of each variable, and the most applied methods are genetic algorithm, ant colony optimization, stepwise selection, variable interval space shrinkage approach VISSA, and regression coefficients [5,10], and b) interval-variable selection approach which picks up informative intervals as input for building models and the most adopted methods are moving window-partial least squares, net analyte signal regression plot, interval partial least squares, bootstrap-based methods, and interval successive projection algorithm [5,10]. In the second approach, the variables (spectral data in this work) are divided into equal-space intervals, and the interval(s) that improves model workability is chosen [9,5,11]. It is interesting to mention that searching combination moving window-partial least squares (SCMW-PLS) can monitor a combination of intervals but after being selected individually (not simultaneously) [10,11]. Methods which search for intervals are more practical for analytical applications as most compounds often absorb over wide bands, and it is rare to find a single wavelength absorber. Hence, picking up intervals of variables instead of a one-by-one strategy seems logical for better model prediction [12]. The aforementioned interval selection approaches, on the other hand, make it challenging to obtain optimized intervals since the widths and combinations of the intervals are not optimized at the same time [5,10].

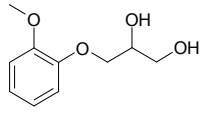
An advanced variable selection method was recently proposed for spectroscopic measurements, which often generate tremendous variables for the sample [5]. The method is known as interval variable iterative space shrinkage approach *i*VISSA and was developed under the condition that locations, widths and combinations of selected intervals are optimized simultaneously [10]. None of the published methods has such an advantage. *i*VISSA is a modified form of the old VISSA algorithm that performs global and local searches simultaneously to intelligently optimize the locations, widths, and combinations of variable intervals to improve calibration outputs [5,10]. In a systematic study, both single and interval variable selection methods including *i*PLS, MW-PLS, GA-PLS, and *i*VISSA were tested to extract informative IR spectra bands to quantify different ingredients in food and drug samples [10]. In fact, the earlier study indicated the high performance of *i*VISSA to extract informative spectra intervals which finally improved PLS performance. At the same time, the performance of GA-PLS outperformed the common MW-

PLS [10]. It should be mentioned that GA can be used as a single and interval variable selector as outlined in the literature [10].

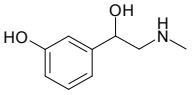
Many pharmaceutical combinations have intense spectral overlapping, requiring proper input variables before PLS calibration [11,13]. So far, a net-analyte signal regression plot was effectively applied to many pharmaceutical formulations to pick up the informative spectral regions before modelling [14,15]. Paracetamol PAR, Guaifenesin GUA, and Phenylephrine PHE is a relatively new formulation and has an intense overlapping between GUA and PHE [16]. Moreover, PHE is present in a lower fraction than PAR and GUA, which is an extra challenge for PLS calibration. Paracetamol (N-(4-hydroxyphenyl) acetamide) is a common pain reliever and fever reducer [16]. Guaifenesin ((2RS)-3-2-methoxyphenoxy)) is an expectorant drug; it assists in the bringing up of phlegm from the airway in acute respiratory tract infections [17]. Phenylephrine (1R)-1-(3-hydroxyphenyl)-2-(methylamino)ethanol Hydrochloride) is a selective α1-adrenergic receptor agonist and used as a decongestant [16]. The structural formulae of the tested compounds are provided in the following scheme:



Paracetamol (pK_a = 9.4)



Guaifenesin (pK_a = 13.6)



Phenylephrine (pK_a = 9.1)

The combination PAR-GUA-PHE is commercially available under different trade names since 2016 with levels of 250, 100, and 5 mg/tablet [17]. Limited research papers have been reported to quantify PAR-GUA-PHE using multivariate calibration [16] compared to liquid chromatography [18]. Although the tested ternary drug system was quantified using wavelet transform method [19] and PLS calibration [20], the application of *i*VISSA along with PLS calibration has not been reported so far which adds more value to the current research.

The present research aims to apply *i*VISSA as an advanced interval variable selection methodology combined with PLS calibration to measure the strongly overlapping combination PAR-GUA-PHE. For more assessment, variable selection is also carried out using the net-analyte signal methodology that received great attention in pharmaceutical analysis. The performance of both methodologies will be addressed taking into account the intense spectral overlapping among drugs.

THEORY

Symbols: The following notations are used in this section. Matrix is represented by a capital letter in boldface and vector is represented by a lowercase letter in boldface. A scalar is represented as a lowercase italic letter. Within the text: data matrix A$m\times n$ is contained absorbance values of m calibration samples measured at n wavelengths, the pure analyte spectrum k is given as s_k of $1\times n$ dimension, $c_k(m\times 1)$ is a vector collected calibration concentrations of analyte k, and $r(1\times n)$ is the vector collected absorbance data of the unknown sample.

PLS Regression

Partially least-squares calibration is a two-block predictive method applied to estimate latent variables needed to correlate two data sets simultaneously: spectral data and physical/chemical properties. PLS regression or calibration is used to develop a linear model that allows for predicting a preferred characteristic (c) from a measured spectrum (a). In matrix and vector forms, the following linear model is proposed: c = Ab where b has the calibration sensitivities collected in the calibration step [1]. The calibration performance of PLS can be notably improved after running a suitable variable selection approach [5].

Net Analyte Signal Methodology

The net analyte signal NAS of k is the portion of its spectrum that is orthogonal to the space encompassed by the spectra of all other analytes or interferences in a particular mixture [21-25]. Lorber's approach for inverse multivariate calibrations was used in this research [29-30]. At the start, PLS calibration was carried out using A & c_k using optimum PLS-factors. Matrix A was then reconstructed to (Â). Matrix $\hat{A}_{-k}$ collecting absorbance signals of all drugs

expect *k* was calculated using the rank annihilation approach [21]. The projection matrix, which is orthogonal to $\hat{A}_{-k}$, is estimated as [21,22]:

$$H = I - (\hat{A}_{-k})^{+} \hat{A}_{-k} \quad (1)$$

where "+" donates the pseudo-inverse matrix while I is the identity matrix. Multiplying H by r (the spectra of the mixture) gives the net analyte signal vector of solute k (r^*_k) [24]:

$$r^*_k = Hr \quad (2)$$

In the above equation, if spectrum r is substituted by the pure spectrum of solute *k* (**s**k), then the left-hand side of Eq. (2) will produce the net pure-spectrum of *k*, i.e. the portion of *k*'s pure spectrum that is orthogonal to other absorbing solutes' spectra ($\mathbf{s}*_k$). Analytical figures of merit such as sensitivity, selectivity, and detection limit are estimated using NAS methodology [23,24]. These measures are used to evaluate the effectiveness of a particular analytical methodology.

The norm of the s^*_k vector is used to define sensitivity [23].

$$SEN = \|s_k*\| \quad (3)$$

On the other hand, selectivity SEL is related to the quantity of signal needed for prediction. Moreover, SEL measures the relative quantity of signal not affected by spectral overlapping and the higher value the lower overlapping of the solute. SEL is estimated as [22,23]:

$$SEL = \|s_k*\|/\|s_k\| \quad (4)$$

The magnitude of spectral overlapping (%) is estimated as: 100 × (1 - SEL). Detection limit DL is often estimated using different NAS approaches, however, the most practical one is derived from the classical univariate calibration [16]:

$$DL = 3||\varepsilon||/\|s_k*\| \quad (5)$$

where $||\varepsilon||$ is a quantity of the instrumental noise [23]. Calculation of quantification limit QL was carried out using the above equation but using factor 10 instead of 3 [25].

Wavelength Selection by NASRP

In this procedure, wavelengths selection was carried out by assessing the error indicator EI (as a function of a moving window) for the new samples, which is based on the NAS regression plot NASRP [13,26,27]. NASRP is a plot derived from sample vector r_k* and s_k*. The strategy of moving window was performed by altering the position of the first variable and the variable range [13,27], EI is given as [26,28]:

$$EI = \frac{[s_{fit}^2(1+(1+\frac{N^2 s_{fit}^2}{4\|r_k^*\|^2})]^{1/2}}{4\|r_k^*\|} \quad (6)$$

where $\|r_k*\|$, s_{fit} and N are the norm of the net analyte spectrum obtained for the unknown sample, the standard deviation of the best-fitted straight line to NASRP in the selected wavelength region, and the number of spectral data in NASRP. Both r_k* and s_k* were obtained from PLS model [26-28].

Variables Selection by iVISSA

The *i*VISSA is an iterative method with global and local search operations that alternately run to improve interval locations, widths, and combinations [10,25]. *i*VISSA is applied for selecting meaningful spectral intervals by generating two matrices known as binary matrix sampling BMS and weighted binary matrix sampling WBMS [10]. *i*VISSA can be used in either sample or variable dimensions, with the former being the more popular. Initially, a binary matrix M1 of dimension $k \times n$ is formed, containing either '0' or '1'. The number of sampling points is *k*, set to 4, and the number of spectral variables is *n*. Each column is allotted the same number of '1' and '0' (0.5*k*). Each column is permuted to produce M2, a new binary matrix. Each row of the new binary matrix M2 represents a random sampling technique, with "1" denoting the variable to be modelled while "0" denoting the variable to be removed from calibration. The values '0' and '1' in each column remain the same, guaranteeing that all variables in the sub-models have the same net frequency. WBMS modified BMS by allocating different weights (*w*) to distinct variables. According to the weights of the variables, various

numbers of "1" and "0" were assigned to each column in the binary matrix. After that, each column was permuted to generate a random set of variables. The values '0' and '1' in each column remain constant, guaranteeing that variables with higher weights have a greater chance of being selected into subgroups. *i*VISSA generated a tremendous number of intervals, and model population analysis MPA was used to pick the best intervals for calibration. MPA is a broad approach for developing new chemometrics algorithms. Due to the uncertainty of a single model, the basic notion of MPA is to extract information from a large population of sub-models rather than a single model. Using the subsets acquired by WBMS, a population of sub-models (*e.g*., 1000) is constructed in this study. The best sub-models with the lowest root-mean-square error of cross-validation (RMSECV) are then extracted, and each variable's frequency is counted [10,25].

EXPERIMENTAL METHODOLOGIES

Apparatus

Using a double beam spectrophotometer (Thermo scientific. Genesys 10S UV-VIS, USA), UV spectra were measured using a 1.0 cm quartz cell. The spectra were obtained in Excel format for running numerical calculations. PLS-1 calibration and all NAS calculations were carried out using MVC1® program [29]. The MATLAB codes for *i*VISSA are available in the literature [10]. MVC1 and *i*VISSA were performed using MATLAB 7.0 (MathWorks Inc. USA) and run on a Pentium VI personal computer with Windows XP operating system.

Reagents

The experiments were carried out using high-quality reagents and chemicals. Standard solutions (500 mg l^{-1}) of PAR, GUS and PHE were prepared by dissolving 0.50 g of the drug in distilled water and diluted to 1.0 Litre. Dilute solutions were then prepared from the standards using distilled water after adjusting pH by 0.1 M NaOH or 0.1 M HCl, and the pH was rechecked and adjusted after dilution.

Effect of Solution pH on UV Absorption: NAS Calculations

The UV spectrum was recorded at pH 6.0, 9.0, and 12.0 at 10 mg l^{-1} for each solute. Multivariate selectivity, multivariate sensitivity and spectral overlap were measured for each solute at each pH using NAS methodology (Eqs. (3) and (4)). As will be shown later, the optimum pH for analysis was 12.0. Accordingly, the solutions of calibration, validation, and formulation samples were prepared at pH 12.0.

Calibration, Validation, and Commercial Solutions

A training set of 12 mixtures of variable proportions of PAR-GUS-PHE were prepared for calibration (Table 1). A three-level full factorial design was created to build the calibration set with the concentrations of each solute lying within the linear absorbance-concentration range. Several synthetic samples were prepared containing different proportions of PAR-GUS-PHE taking into account their variability in the actual formulation (250:100:5). As claimed by the manufacturer, the added excipients were starch potassium sorbate, talc, and stearic acid. Among added

Table 1. Calibration and Validation Mixtures (mg l^{-1}) Used for *i*VISSA-PLS Regression

Calibration	PAR	GUA	PHE
C1	5.0	8.0	10.0
C2	2.0	12.0	18.0
C3	9.0	12.0	10.0
C4	9.0	8.0	18.0
C5	5.0	12.0	4.0
C6	9.0	4.0	4.0
C7	2.0	4.0	10.0
C8	2.0	8.0	4.0
C9	5.0	12.0	18.0
C10	2.0	4.0	4.0
C11	5.0	9.0	10.0
C12	9.0	12.0	18.0
Validation	PAR	GUA	PHE
V1	3.0	7.0	15.0
V2	8.0	5.0	4.0
V3	3.0	11.0	4.0
V4	3.0	11.0	13.0
V5	8.0	11.0	3.0
V6	50.0	25.0	1.0

excipients, potassium sorbate and stearic acid were the most soluble in water and would affect UV spectra of drugs. Accordingly, calibration and validation solutions were prepared in a 5% mixture of potassium sorbate and stearic acid so that the analytical signal can account for the added excipients.

As indicated in Table 1, the concentrations of V1-V5 solutions were within ranges selected in the calibration set, while V6 was prepared in relative amounts identical to those present in commercial formulation. The commercial formulation was obtained from a local pharmacy (Panadol® COLD+FLU ALLINONE, SmithKline Beecham, Spain). It was prepared as follows: Ten capsules were ground, and a portion equivalent to the mass of one capsule was treated with 50 ml distilled water, sonicated for 15 min and centrifuged. The supernatant was appropriately diluted, the final pH was adjusted to 12.0, and scanned over 200-300 nm with 1.0 nm step (*i.e.*, 101 variables/sample). The contents of PAR-GUA-PHE were determined using PLS after variables selection by NAS or *i*VISSA.

RESULTS AND DISCUSSION

Effect of Solution pH on the Spectral Behaviour of Pharmaceuticals: NAS Calculations

Solution pH strongly influences the spectra of pharmaceuticals, which affect sensitivity SEN, selectivity SEL, and spectral overlapping. Hence, multivariate SEN/SEL and spectral overlapping of PAR, GUA, and PHE were estimated at pH 6.0, 9.0, and 12.0 using NAS calculations (Eqs. (3) and (4)). The UV spectra of pharmaceuticals at different pHs are provided in Fig. 1, while NAS calculations are provided in Table 2.

In fact, pK_a values of examined drugs are 9.1, 9.4, and 13.6 PHE, PAR, and GUA. To investigate the influence of pH on spectra of drugs, different pHs were selected to cover pK_a of solutes except for GUA which has an extremely high acidity constant. As shown in Fig. 1, the UV spectra of pharmaceuticals were sensitive to pH but with different magnitudes. For PAR, better light absorption was observed at pH 6.0 and over the range (230-250 nm) as indicated in

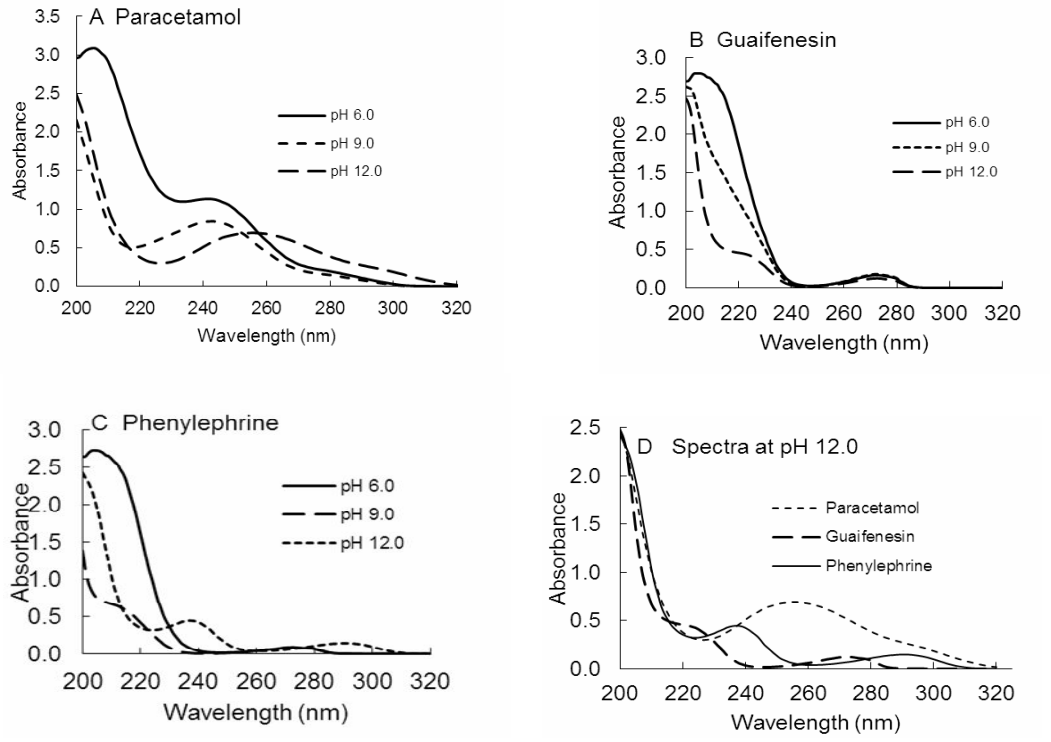


Fig. 1. UV spectra of pharmaceuticals (10.0 mg l^{-1}) measured at different pH values (A-C) and spectra of solutes at pH 12.0 (D).

Table 2. Multivariate Sensitivity and Selectivity Measured at D pH Values Using NAS Methodology

	PAR			GUA			PHE		
pH	SEN	SEL	SO%	SEN	SEL	SO%	SEN	SEL	SO%
6.0	4.91	0.35	65	1.01	0.08	92	1.02	0.09	91
9.0	4.02	0.57	43	1.42	0.16	84	0.58	0.17	83
12.0	3.32	0.43	57	1.61	0.27	73	1.58	0.23	77

SEN: Sensitivity SEL: Selectivity SO: Spectral Overlapping

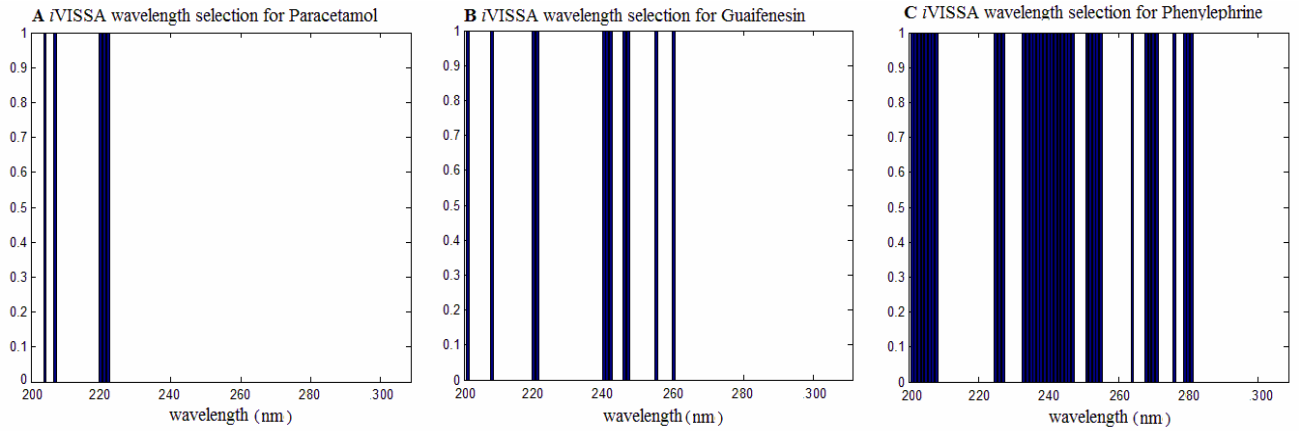


Fig. 2. Selected spectral variables for drugs by *i*VISSA methodology.

Fig. 1A. A significant bathochromic peak shift was reported at pH 12.0, and the solute has high absorption at 255 nm which is attributed to solute ionization (solution pH > pK_a).

Among tested pharmaceuticals, GUA spectra were also insensitive to pH over the range (250-285 nm) as shown in Fig. 1B. However, a strong hyperchromic effect was observed at wavelength <240 nm and pH 6.0 which could be attributed to solute protonation (solution pH << pK_a). The most interesting case was reported for PHE, as many changes were observed in its spectrum (Fig. 2C). Although the spectrum recorded at pH 6.0 indicated a high light absorption at wavelength <240 nm, the spectrum at pH 12 .0 indicated the creation of new peaks that appeared at 235 and 290 nm which indicated the ionization of the solute (solution pH > pK_a). The spectrum obtained at pH 9.0 was not practical as the solute does not show high absorption at wavelength > 230 nm. The spectra of the three drugs at pH 12.0 are shown in Fig. 1D. As can be observed, significant light absorption of PAR with an intense overlap in their spectra makes univariate calibration ineffective for drugs quantification. For commercial formulation, direct or derivative spectrometry would be applicable for PAR and GUA quantification but not for PHE as it is available in a small proportion. Hence, PLS is frequently used to achieve the latter goal, and it has been extensively applied to the analysis of multicomponent pharmaceutical combinations with different proportions [26,27]. The performance of PLS calibration for PHE in the tablet would be questionable, taking the high level of excipients even after dilution. Furthermore, effective spectral selection can improve PLS calibration performance for ingredients of a small proportion [27]. To check the influence of pH on multivariate SEN and SEL of the drugs, NAS was applied to find the significant part of UV spectra needed for analytical measurement, as summarized in Table 2.

As indicated in Table 2, PAR has the best sensitivity (3.32-4.91) among drugs and over the tested pHs. The high multivariate SEN of PAR was expected due to its significant light absorption over the whole range (Fig. 1A). However, SEL of solutes was notably affected by pH. The best pH is

the one that achieved the highest SEN/SEL and minimum overlapping. For PAR, the optimum pH was 9.0 because high SEN and SEL (and lower overlapping) was reported at this pH. The high sensitivity at pH 6.0 (4.91) was not enough to run calibration as spectral overlapping was high (65%) at this pH. The same is true at pH 12.0 with spectral overlapping 57%. For GUA, the analysis revealed that the optimum pH was 12.0 but with significant spectral overlapping (73%). For PHE, the optimum pH was 12.0 but also with a serious overlapping with other pharmaceuticals. Despite the low spectral overlap of PAR at pH 9.0 compared to 12.0, PLS calibration was carried out at pH 12.0 considering the high sensitivity of this drug at all pHs. NAS methodology is often applied to estimate sensitivity and selectively of overlapping drug mixtures [22,30].

Variable Selection by NAS and iVISSA

PLS, NAS-PLS, and *i*VISSA-PLS were tested to quantify PAR-GUS-PHE in synthetic mixtures and this was necessary due to: a) assessing the extent of spectral overlapping on the calibration performance of the models in the absence of excipients that are available in commercial tablet, and b) comparing the selected variables generated by both methods taking into account their different mechanisms for variables extraction. In NAS, the variables are obtained from NASRP which gives one fixed-width interval [27,28]. However, *i*VISSA intelligently picks up all possible intervals considering locations, widths, and combinations that are optimized simultaneously [10,28]. Hence, *i*VISSA can give separate intervals, and this advantage is not available in the NAS methodology. On the other hand, variables selection by NAS is based on the net analyte signal of the solute, which would result in more accurate outputs when analyzing commercial tablets. The final PLS calibration results before and after variables selection are provided in Table 3. Moreover, the selected

Table 3. Selection of Informative Spectral Variables by *i*VISSA and NAS Prior to PLS Calibration at pH 12.0

Method	Variables	PLS factors	REP%
	PAR		
PLS	**101** 200-300 nm	3	3.3
NAS/PLS[a]	**30** 229-258 nm EI 0.0012-0.0034	2	5.5
*i*VISSA-PLS[b]	**5** 203, 206, 219, 221, 222 nm	2	5.4
	GUA		
PLS	**101** 200-300 nm	4	7.2
NAS/PLS	**31** 220-250 nm EI 0.0011-0.0025	3	4.7
*i*VISSA/PLS	**11** 200, 207, 219-220, 239-240, 241, 245-246, 254, 259 nm	3	7.0
	PHE		
PLS	**101** 200-300 nm	4	4.2
NAS/PLS	**25** 220-244 nm EI 0.0010-0.0018	3	3.8
*i*VISSA/PLS	**34** 200-207, 224-225, 232-246, 250-254, 263, 267-270, 275, 278-280 nm	3	2.8

[a]The selected variables were estimated from NASRP of each solute in each mixture (V1-V5). The selected variables was the same for each solute in the tested mixture obviously due to absence of non-modelled compounds. [b]The selected variables were estimated by *i*VISSA by simultaneously adjusting widths, locations, and combinations of the variables that achieved the best prediction of calibrated solutes.

variables picked up by *i*VISSA and for each drug are provided in Fig. 2.

For PAR, including all variables (101) was necessary to give a high prediction with a REP of 3.3%. However, NAS-PLS and *i*VISSA-PLS ended up with good predictions (REP 5.4-5.5%) using fewer PLS variables (2 factors only). As shown in Fig. 1D, PAR has a high absorption over 240-280 nm, and the variables selected by NAS (229-258 nm) seem logical due to the reasonable closeness to the earlier spectral range. It should be mentioned that NAS methodology select variables based on the net analyte signal of the calibrated solute, which should be orthogonal with the other solutes. The four variables provided by *i*VISSA (203, 206, 219, 221, and 222 nm) were unexpected as all drugs intensely absorb at these wavelengths (Fig. 1D). However, with only four variables, *i*VISSA-PLS ended up with comparable performance with NAS-PLS, which selected 30 variables (229-258 nm) for PAR prediction. For GUA, the winner was NAS-PLS with a REP of 4.7% compared to 7.2% for PLS and 7.0% for *i*VISSA-PLS. The high performance of NAS-PLS was notably back to the proper selection of the interval (220-250 nm, 31 variables) as this solute strongly absorbs in this region (Fig. 1D). However, with only 11 spectral variables (200, 207, 219-220, 239-240, 241, 245-246, 254, 259 nm) *i*VISSA-PLS outperformed PLS with a REP% of 7.0. The interesting points regarding *i*VISSA are; a) selection of variables that all solutes significantly absorb (200 and 207 nm), b) unlike NAS, *i*VISSA picked up intervals that GUS strongly absorb including, 219-220 and 254 and 259 nm, and c) selection of 241, 245-246 nm by *i*VISSA was unpredicted due to the intense absorption of PAR in these regions (Fig 1D). The results of PHE indicated the effectiveness of *i*VISSA as variable selectors. *i*VISSA-PLS outperformed PLS and NAS-PLS for PHE prediction with REP down to 2.8%.

The high performance of *i*VISSA-PLS was attributed to the proper selection of the variables, especially 224-225, 232-246, 250-254, 267-270, 275, and 278-280 nm, where the drug notably absorb, as seen in Fig 1D. As was the case for PAR and GUS, the interval 200-207 nm was not expected as all solute strongly absorb in this region. The lower performance of NAS-PLS compared to *i*VISSA-PLS would be attributed to the selection of one interval (220-244 nm) only while excluding the region (280-300 nm) where the solute strongly absorb (Fig. 1D). For better solute prediction in commercial formulation, quantification of PAR, GUA, and PHE should be carried out by PLS, NAS-PLS, and *i*VISSA-PLS, respectively. However, using *i*VISSA-PLS is practical considering the low proportion of PHE in the commercial tablet. Both detection and quantification limits were estimated by NAS calculations (Eq. (5)) and at pH 12.0. Based on NAS calculations at pH 12.0, the drugs were detected down to 0.12 mg l^{-1} for PAR and 0.24 mg l^{-1} for GUA/PHE, while quantification limits were 0.40 and 0.79 mg l^{-1} for PAR and GUA/PHE, respectively.

Quantification of Pharmaceuticals in Commercial Tablet

As indicated earlier, NAS was efficient for selecting spectral range for GUA prediction and iVISSA was efficient for selecting a spectral intervals for PHE prediction. For the most sensitive drug (PAR), the best prediction was achieved using the whole spectrum. Hence, the performance of the methods was further checked to quantify drugs in the V6 mixture (Table 1) that contain the same proportion of the pharmaceuticals in the tablet. The results are provided in Table 4, along with the commercial tablet.

As for the synthetic solution, where there is very little PHE content compared to PAR and GUA, the drug has been determined with an accuracy of 106.9%. Determination of PHE with high accuracy is due to the proper selection of spectral intervals, as summarised in Table 3. PAR and GUA were quantified with high accuracy (98.6-102.5%), although their concentrations in the synthetic solution were outside the range used in the calibration stage (Table 1). For GUA, the selected interval was (220-250 nm, EI 0.0023), which was the same range reported for this drug in other validation solutions (V1-V5, Table 1). Dilution of the tablet solution was essential due to the high concentration of both PAR/GUA compared to PHE. The spectra of tablet solution were measured after dilution with water to different magnitudes and adjustment pH at 12.0, as shown in Fig. 3.

Initial analysis indicated that the spectrum recorded at 1:10 dilution was convenient for pharmaceuticals prediction. The spectra recorded at 1:3 and 1:5 dilutions were excluded due to the high interferences of excipients available in the tablet (Fig. 3). As indicated in Table 4,

Table 4. Results Obtained by Applying PLS, PLS-NAS, and PLS-*i*VISSA Analysis to both Synthetic and Real Ternary Mixtures of Pharmaceuticals (RSD, $n = 4$)

System	Drug	Added	Predicted	Interval	Predicted	Recovery (%)
Synthetic solution[b]	PAR	50.0	PLS	200-300 nm, 101 variables	50.3(±1.4)	102.5
	GUA	25.0	NAS-PLS	220-250 nm, 31 variables EI 0.0023	24.7(±2.3)	98.6
	PHE	1.0	*i*VISSA-PLS	202-207, 229-236, 250-254, 268-275 nm, 27 variables	1.1(±4.3)	106.9
Panadol® COLD+FLU (ALLIN**ONE**)[c] PAR 250 mg GUA 100 mg PHE 5 mg	PAR	257.6(±4.3)[d]	PLS	200-300 nm, 101 variables	247.8(±6.5) $t^e = 1.65$ $F^e = 3.23$	99.7
	GUA	103.4(±5.4)[d]	NAS-PLS	225-239 nm, 15 variables EI 0.0131	98.5(±7.9) $t = 2.23$ $F = 3.67$	98.4
	PHE	5.4(±5.1)[d]	*i*VISSA-PLS	202-207, 229-236, 250-254, 268-275 nm, 27 variables	5.7(±8.2) $t = 2.04$ $F = 4.67$	104.3

[a]The provided concentrations represent the mean of four identical measurements per sample ($n = 4$, ± relative standard deviation). [b]Directly prepared from the standard solutions (low level of PHE) and measured at pH 12.0. [c]The tablet solution was prepared by grinding 10 capsules and dissolving an equivalent mass of one capsule in 1.0 litre water. The spectra were measured by spectrometer after dilution 1:10 with water and at pH 12.0. [d]Concentration of active ingredients were independently measured by liquid chromatography as outlined in the literature [17]. [e]t and F values were estimated from the results of the proposed multivariate method and standard HPLC. Tabulated t (0.05, 6) and F values (0.05, 3, 3) are 2.45 and 15.44, respectively.

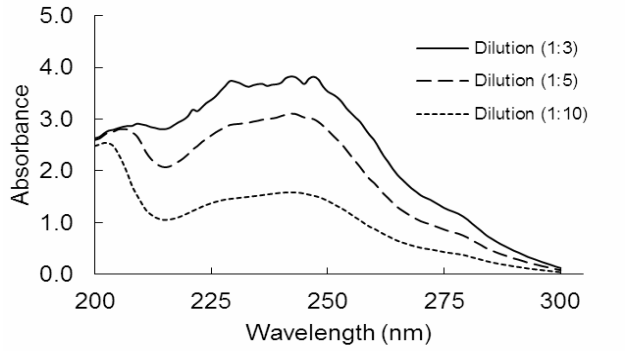


Fig. 3. UV spectra of the commercial tablet at different dilutions at pH 12.0.

active ingredients were assayed in the tablet with high recovery (98.4-104.3%) and precision (RSD 0.5-3.9). Unlike *i*VISSA, the NAS method can search the optimum spectral interval for any new sample. Determination of PAR, GUA, and PHE by the independent chromatographic method was also included in Table 4. Both t and F tests were carried out to check if the results of both methods were statistically comparable as indicated in Table 4. In all cases, calculated t values were lower than tabulated ones, indicating no significant difference between both methods. Moreover, F-calculated values were also lower than tabulate one, which confirmed that both methods were of comparable precision (Table 4). The reported NASRP for GUA prediction in the tablet is provided in Fig. 4.

As shown in Fig. 4B, the best spectral interval for GUA was (225-239 nm, 15 variables) with EI of 0.0131. The selected region was different from the one selected for GUA prediction in synthetic mixtures (220-250 nm, Table 3). The presence of uncalibrated solutes in the tablet affected the function of NAS for the extraction of variables. Hence, NAS method will search for regions that contain the maximum spectral information of GUA. *i*VISSA-PLS

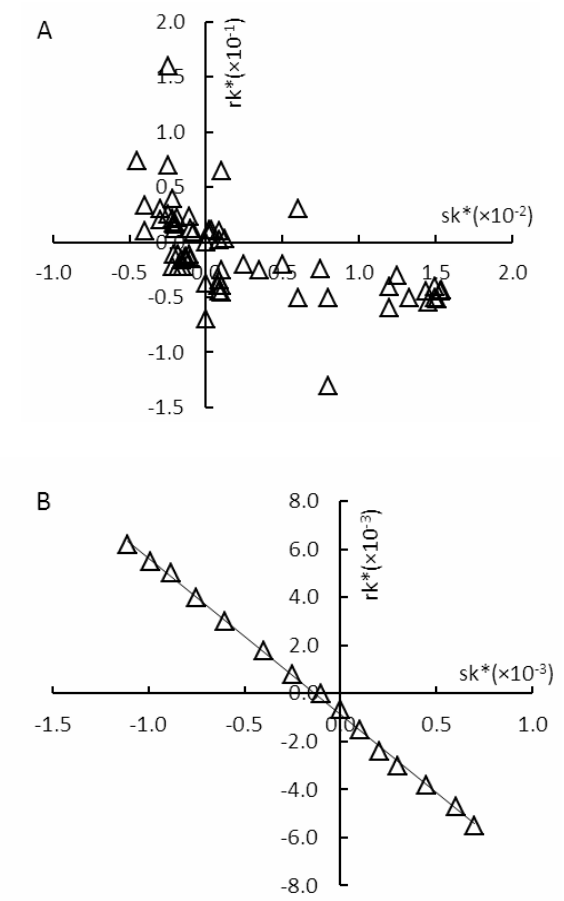


Fig. 4. (A) NASRP for GUA in commercial sample using the full wavelength range 200-300 nm. (B) NASRP in the range 225-239 nm, as predicted from the analysis of the EI parameter. The solid line is the best linear fit to the calculated points.

managed to quantify PHE in commercial tablets with high accuracy by selecting the intervals (200-207, 224-225, 232-246, 250-254, 263, 267-270, 275, 278-280 nm) where the solute efficiently absorb.

CONCLUSIONS

Interval variable selection approaches such as *i*VISSA and NAS have been effective to handle spectral overlapping while selecting the informative intervals for the best PLS calibration. NAS calculation was practical to determine the optimum pH needed to analyze the mixture of pharmaceuticals. The pharmaceutical combination (PAR-GUA-PHE) showed high spectral overlapping (57-77%) at pH 12.0, and variables selection by NAS & *i*VISSA was found useful to get high PLS prediction. For PAR, the entire range (200-300 nm, 101 spectral points) was necessary to be included. However, NAS calculations revealed that GUA could be predicted over the interval (220-250 nm) with high accuracy. For PHE, *i*VISSA outperformed NAS method for picking up the informative variables needed for PLS calibration. The selected variables for PHE were 200-207, 224-225, 232-246, 250-254, 263, 267-270, 275, 278-280 nm. The ternary drug system was quantified in commercial tablet with high accuracy (99.7-104.3%) and precision (RSD 0.5-3.9) using full-spectrum PLS calibration (for PAR), NAS-PLS (for GUA), and *i*VISSA-PLS (for PHE).

ACKNOWLEDGEMENTS

We would like to thank Mr Bassem Nasrallah for samples preparation and spectral analysis.

LIST OF ABBREVIATIONS

DL	Detection limit
EI	Error indicator
*i*VISSA	Interval Variable Iterative Space Shrinkage Approach
Guaifenesin	GUA
Paracetamol	PAR
PLS	Partial Least Squares
NAS	Net Analyte Signal
NASRP	Net Analyte Signal Regression Plot
Phenylephrine	PHE
Selectivity	SEL
Sensitivity	SEN
QL	Quantification limit

REFERENCES

[1] R. Brereton, Applied Chemometrics for Scientists, John Wiley & Sons Ltd, England, 2007.
[2] I. Johnstone, D. Titterington, Philos. Trans. A Math. Phys. Eng. Sci. 367 (2009) 4237.
[3] S. Wold, M. Sjostrom, L. Eriksson, Chemom. Intell.

Lab. Syst. 58 (2001) 109.
[4] J. Fan, J. Lv, J.R. Stat. Soc. B. 70(2008) 849.
[5] L. Wang, Y. Lin, X. Wang, N. Xiao, Y. Xu, H. Li, Q. Xu, Chemom. Intell. Lab. Syst. 172 (2018) 229.
[6] J. Fan, J. Lv, Stat. Sin. 20 (2010) 101.
[7] N. Chavoshi, B. Hemmateenejad, Anal. Bioanal. Chem. Res. 9 (2022) 183.
[8] X. Zou, J. Zhao, M. Povey, M. Holmes, H. Mao, Anal. Chim. Acta 667 (2010) 14.
[9] E. Thomas, Anal. Chem. 66 (1994) 795.
[10] B. Deng, Y. Yun, P. Ma, C. Lin, D. Ren, Y. Liang, Analyst 140 (2015) 1876.
[11] H. Goicoechea, A. Olivieri, J. Pharm. Biomed. Anal. 20 (1999) 255
[12] J. Jiang, R. Berry, H. Siesler, Y. Ozaki, Anal. Chem. 74 (2002) 3555.
[13] M. Ribone, A. Pagani, A. Olivieri, J. Pharm. Biomed. Anal. 23 (2000) 591.
[14] A. Hoskuldsson, Chemom. Intell. Lab. Syst. 55 (2001) 23.
[15] I. Fasfous, Y. Al-Degs, A. Mallah, Inddian Drugs 53 (2016) 26.
[16] G. Gergov, A. Alin, P. Katsarov, V. Simeonov, D. Yankov, Y. Al-Degs, Bulg. Chem. Comm. 50 (2018) 265.
[17] A. Yehia, H. Mohamed, Spectrochim. Acta A 152 (2016) 491.
[18] P. Dicpinigaitis, Y. Gale, Chest 124 (2003) 2178.
[19] M. Khoshayand, H. Abdollahi, A. Ghaffari, M. Shariatpanahi, H. Farzanegan, Daru. 18 (2010) 292.
[20] M. Ebrahimi, M. Sohrabi, F. Motiee, M. Davallo, Optik 230 (2021) 166323.
[21] B. Hemmateenejad, R. Ghavami, R. Miri, M. Shamsipur, Talanta 68 (2006) 1222.
[22] B. Hemmateenejad, S. Yousefinejad, Anal. Bioanal. Chem. 394 (2009) 1965.
[23] B. Salameh, Y. Al-Degs, K. Abu-Safieh, A. AL-Zghool, Food Measure 14 (2020) 78.
[24] H. Goicoechea, A. Olivieri, Chemom. Intell. Lab. Syst. 56 (2001) 73.
[25] Y. Al-Degs, B. Abu-Ghazaleh, J. Abdelghani, O. Zaid, Food Measure (2022).
[26] H. Goicoechea, A. Olivieri, Chemom. Intell. Lab. Syst. 56 (2001) 73.
[27] H. Goicoechea, A. Olivieri, Analyst 124 (1999) 725.
[28] Y. Al-Degs, A. El-Sheikh, A. Saleh, A. Al-Reyahi, Food Measure 15 (2021) 2562.
[29] A. Olivieri, H. Goicoechea, F. Inon, Chemom. Intell. Lab. Syst. 73 (2004) 189.
[30] K. Siddareddy, M. Reddy, J. Sreeramulu, Res. J. Pharm. Bio. Chem. Sci. 7 (2016) 2274.