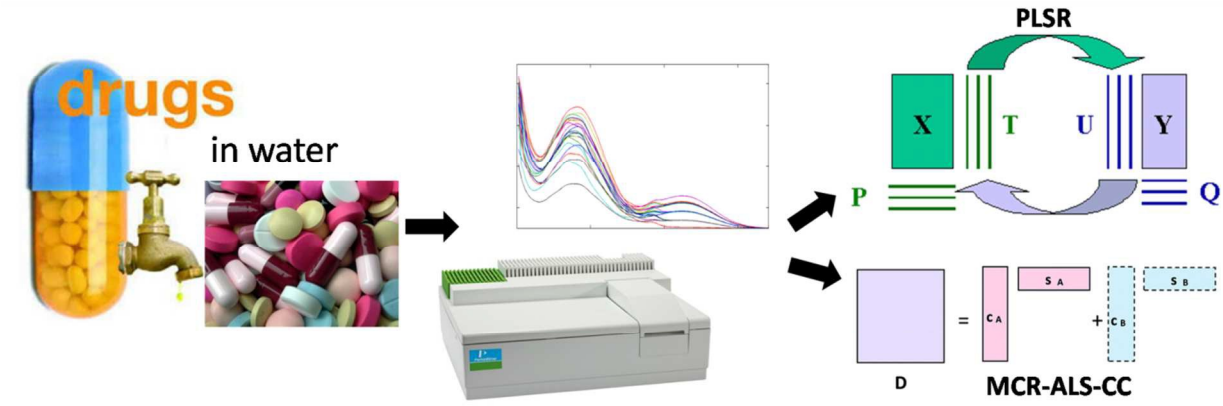
RSC Advances

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/advances

Graphical Abstract



Comparative study of partial least squares and multivariate curve resolution for simultaneous spectrophotometric determination of pharmaceuticals in environmental samples

Hadi Parastar*, Hamidreza Shaye

Department of Chemistry, Sharif University of Technology, P.O. Box 11155-3516, Tehran, Iran

* Author to whom correspondence should be addressed;

Tel.: +98- 21- 66165306; fax: +98- 21- 66005718 (Hadi Parastar)

E-mail: h.parastar@sharif.edu, h.parastar@gmail.com

Abstract

The potential of partial least squares regression (PLSR) and multivariate curve resolution alternating least squares (MCR-ALS) is evaluated for simultaneous determination of diclofenac (DCF), naproxen (NAP), mefenamic acid (MEF) and carbamazepine (CBZ) as target analytes and gemfibrozil (GEM) as interference in synthetic and real environmental samples. The analysis of first-order UV-Vis spectra is performed using PLSR with different variable selection methods including variable importance in projection (VIP), recursive partial least squares (rPLS), regression coefficient (RV) and uninformative variable elimination (UVE) and using MCR-ALS with correlation constraint (MCR-ALS-CC). The obtained statistical parameters in terms of relative error (RE), regression coefficient (R^2) and root mean square error (RMSE) were satisfactory for calibration and validation sets. Furthermore, in real environmental samples, the obtained statistical parameters of PLSR using VIP and rPLS and also MCR-ALS-CC were reasonable by considering the heavy overlap of target analytes and complex samples matrices. In general, PLSR showed better performance for determination of analytes in samples which are free of interference or contains calibrated interference(s). On the other hand, MCR-ALS-CC allowed for the accurate determination of analytes in the presence of unknown interference and more complex sample matrices.

Keywords: Multivariate calibration; Partial least squares; Multivariate curve resolution; Pharmaceuticals; Spectrophotometry.

1. Introduction

Analytical chemistry involves samples that are far from simple and often contain many components or a few sought analytes in the presence of other chemical interferences. Pharmaceuticals are synthetic chemicals which may react differently in the environment and some of them enter into the environment and are persistent, therefore, determination of this new class of pollutants which are used in a large volume every year is of prime importance.[1-4]

A critical aspect in quantitative analysis with first-order data (e.g., UV-Vis spectra) is the occurrence of overlapping spectra and matrix effects, which may lead to a significant difference in the response of an analyte in a sample as compared to a pure standard solution.[5, 6] Multivariate calibration methods have been proposed to overcome fundamental mathematical challenges occurred during analysis of these mixtures. Among different multivariate calibration methods, partial least squares regression (PLSR) and multivariate curve resolution-alternating least squares (MCR-ALS) have attracted attention in chemistry in recent years.[7]

PLSR has become the most popular multivariate calibration method because of the quality of its calibration models and the ease of its implementation.[8] PLSR has been frequently used in the spectrophotometric analysis of pharmaceuticals in complex biological and environmental samples.[9-12] However, the performance of PLSR method is strongly depends on the quality of variable selection methods. If most significant variables are included in the model, thus, the potential of the model in prediction of the desired properties in unknown samples will increase.[13] Popular variable selection methods include: variable importance in projection (VIP)[14, 15], regression vector (RV), uninformative variable elimination (UVE)[16, 17] and recursive weighted PLS (rPLS).[18] To

the best of our knowledge, there is no clear study showing the pros and cons of different variable selection methods in spectroscopic data.

In this study, the aforementioned PLSR techniques are directly compared with MCR-ALS, which tries to solve the mixture analysis problems into a bilinear model of meaningful pure component contributions.[19] MCR-ALS has been frequently used to resolve and provide pure concentration and spectral profiles of target compounds indifferent types of complex processes and mixtures for determination of pharmaceuticals in environmental matrices.[20, 21] The MCR-ALS with correlation constraint (MCR-ALS-CC) can be used as a useful tool for quantitative spectroscopic measurements. This constraint is extended to quantitative analysis of spectral data in the simultaneous analysis of different analytes in samples of increasing complexity, including pharmaceutical and agricultural samples.[20, 22-24] Generally, the predictive capability of MCR-ALS-CC for determination a particular analyte in unknown mixtures and natural samples especially in presence of interferences is comparable to the results obtained by PLSR calibration approaches using proper variable selection methods. The main advantage of using MCR-ALS instead of PLSR is the possible recovery of the spectral information of target analytes and unknown interferences.[20]

In the present contribution, simultaneous analysis of pharmaceuticals in aquatic media in the presence of interferences and matrix effects was considered as an example for evaluation of the potential of MCR-ALS and PLSR combined with various variable selection methods for simultaneous spectrophotometric determination of constituents with heavy overlap in the spectral direction. In this regard, performance of different algorithms

were compared using statistical parameters, such as relative error (RE, %), root-mean square error (RMSE) and coefficient of multiple determination (R^2).

2. Theory

2.1 Partial least squares (PLS)

PLSR is a multivariate method that finds a linear regression model by projecting the predicted variables and the observable variables to a new space. In the original PLS method all variables can be used.[8] However, to have a robust and stable model, only significant variables should be kept in the model. In this regard, various variable selection methods, such as variable importance in projection (VIP)[15], regression vector (RV) [8], uninformative variable elimination (UVE) [16]and recently developed recursive weighted partial least squares (rPLS) [18] have been proposed. Among these methods, rPLS which iteratively reweights the variables using the PLS regression coefficients can be considered as a new idea for variable selection. More details regarding the PLSR model and performance of different variable selection methods in PLS can be found elsewhere.[15, 25-27] However, a brief description of rPLS method is presented owing to its novelty.

In rPLS, two different models of independent variables, $\mathbf{X}$, are used, $\widehat{\mathbf{X}}_{\mathbf{M}}$ the original model and $\widehat{\mathbf{X}}_{\mathbf{A}}$ the alternative model. In this regard,$\widehat{\mathbf{X}}_{\mathbf{M}}$is an approximation of the original data and $\widehat{\mathbf{X}}_{\mathbf{A}}$is an approximation of $\widehat{\mathbf{X}}_{\mathbf{M}}$:

$$\widehat{\mathbf{X}}_{\mathbf{M}} = \mathbf{X}\mathbf{W}\mathbf{W}^{\mathbf{T}} \qquad (1)$$

$$\widehat{\mathbf{X}}_{\mathbf{A}} = \mathbf{X}\mathbf{A}\mathbf{A}^{\mathbf{T}} \qquad (2)$$

where $\mathbf{W}_{(J\times F)}$ and $\mathbf{A}_{(J\times 1)}$ denote the loading weights in the original model and the alternative model, respectively. The difference between the alternative model and the original model is calculated as follows:

$$\mathbf{E} = \widehat{\mathbf{X}}_{\mathbf{A}} - \widehat{\mathbf{X}}_{\mathbf{M}} = \mathbf{XAA^T}\text{-}\mathbf{XWW^T} \quad (4)$$

The optimization is performed using the goal of minimization of the objective function which is the square of the Frobenius norm of residual matrix, **E** as follows:

$$\mathbf{min_A}\|\mathbf{XAA^T} - \mathbf{XWW^T}\|_{\mathbf{F}}^{\mathbf{2}} \quad (5)$$

After each weighting, some variables have been amplified, and others have become attenuated. This is the basis of variable selection by rPLS method.[18]

2.2 Multivariate curve resolution-alternating least squares (MCR-ALS)

MCR is based on a bilinear additive model which can be expressed by the following expression:

$$\mathbf{X}=\mathbf{CS^T}+\mathbf{E} \quad (6)$$

where **X**(I×J) is a data matrix containing the UV–Vis spectra. The **C** matrix contains the concentration profiles of all components and $\mathbf{S}^T$ the corresponding pure spectra. Also, **E** contains the unmodelled part of the data not explained by the bilinear model.[19] The readers encourage reading references 5, 19 and 20to get more information about MCR-ALS.

The number of components can be obtained by singular value decomposition (SVD).[19] To start ALS optimization, an initial estimate of concentration or spectral profiles based on variable selection methods, such as simple-to-use interactive self-modelling mixture analysis (SIMPLISMA) is needed.[28] Various constraints can be used during ALS optimization, such as non-negativity [29], spectral normalization, known spectral profiles (equality constraint), and correlation constraint. Correlation constraint builds local internal univariate calibration models between reference concentration values in

calibration samples and values obtained by MCR.[20, 22] Using this local model the concentration of calibration samples is calculated and then, in order to predict the concentration of target analytes in prediction samples, the parameters of calibration model are used. In the same way, each ALS iteration is completed after updating the obtained values of prediction. Also, this constraint can be used to correct matrix effects.[23, 24] In detail, $\mathbf{c}_{\mathbf{cal}}^{\mathbf{ALS}}$ the concentrations estimated by ALS for the calibration set are regressed against $\mathbf{c}_{\mathbf{cal}}^{\mathbf{ref}}$, the known values (reference concentrations), in each iteration:

$$\mathbf{c}_{\mathbf{cal}}^{\mathbf{ALS}} = \mathrm{b}\mathbf{c}_{\mathbf{cal}}^{\mathbf{ref}} + \mathrm{b}_0 \qquad (7)$$

where the slope b and offset b_0 are obtained by fitting regression model. The predicted concentration vector, $\hat{\mathbf{c}}_{\mathbf{test}}$, obtained by:

$$\hat{\mathbf{c}}_{\mathbf{test}} = \frac{\mathbf{c}_{\mathbf{test}}^{\mathbf{ALS}} - \mathbf{b}_0}{\mathbf{b}} \qquad (8)$$

These models are then used to predict concentration in validation and test samples.

3. Experimental

3.1 Chemicals and solvents

Analytical standards of diclofenac (DCF), naproxen (NAP), mefenamic acid (MEF), carbamazepine (CBZ) and gemfibrozil (GMF) were provided by RouzDarou Pharmaceutical Company (Tehran, Iran).Selected pharmaceuticals as target analytes and their physic-chemical properties are shown in Table 1.

Table 1 near here

Spectroscopy grade methanol was purchased from Merck (Darmstadt, Germany). Also, formic acid (analytical grade) for sample preparation and pH adjustment was provided from Merck (Darmstadt, Germany).

3.2 Preparation of standard and real samples

Stock standard solutions of pharmaceuticals were prepared in methanol-water (77:23) at concentration 1250.0 $\mu g mL^{-1}$ adjusted at pH=5.0 by formic acid buffer and were kept in refrigerator at 4 °C in the dark. The river, well and tap waters samples were filtered through a 0.45 μm filters. The composition of the real samples was fixed by mixing methanol to obtain 77:23% ratio. Thirty-eight synthetic mixtures were prepared in the concentration range of 1.0 and 30.0 $\mu g\ mL^{-1}$ and divided to calibration (22 samples), validation (7 samples) and test (9 samples) sets. In order to reduce the effects of uncontrolled factors, central composite design (CCD) [30] with five concentration levels was used to build the calibration set. The CCD was used to control the effects of uncontrollable factors, to have a randomized calibration set, and to have an orthogonal and rotatable calibration design. Concentration of validation set adopted randomly and test set fixed at three concentration levels. Standard solutions were obtained by adding required volumes of pharmaceutical stock solution. Table 2 shows the concentration matrix used to prepare calibration, validation and test set.

Table 2 near here

3.3 Instrumentation

Absorbance measurements were performed on a Lambda 25 spectrophotometer (PerkinElmer, USA) with the use of 1.0 cm glass cell. The spectra were acquired in a wavelength range from 200 to 411 nm with a resolution of 1 nm. The pH measurements were taken with a pH-meter AZ 86502 (Taichung, Taiwan). All data were recorded at room temperature. For the preparation of all samples, deionised water purified with cartridges from Millipore (Milli-Q) to a resistivity of 18.2 MΩ.cm was used.

3.4 Data analysis

Data analyses were performed on an Intel Core i7 based ASUS personal computer. All calculations were performed in MATLAB R2013a (Mathworks, Natick, MA, USA). The MCR–ALS and PLS toolboxes were used for building calibration models.[31, 32] Design-Expert 7.0.0 Trial (Stat-Ease Inc., Minneapolis, USA) was used for designing calibration and validation sets.[33] The MATLAB codes for rPLS and UVE used for building their calibration models were obtained from corresponding references.[18, 34, 35]

4. Results and discussion

4.1 Calibration set

The raw spectra of different samples of calibration set are shown in Figure 1inwavelength range 240–411 nm with 1nm resolution and in concentration range of 1.0 and 30.0 μg mL^{-1}. Figure 1(a)-(c) shows the raw spectra of different samples in calibration, validation and test sets, respectively. Also, Figure 1(d) shows pure spectra of four target analytes (CBZ, NAP, DIC and MEF) and GMF as calibrated interference at concentration of 15 μgmL^{-1}. The obtained UV-Vis spectra showed some problems, such as peak overlap, noise and

narrow band scanned wavelengths. Consequently, only wavelengths in ranges 200-225 and 241-411nm were kept and wavelengths with noisy absorbance and absorbance outside of linear range of instrument were eliminated. As shown in Figure 1(d), there is a heavy overlap between pure spectra of target analytes. In addition, GMF (interference) has low absorbance intensity compare to other analytes and also a heavy overlap with them. As it can be seen from pure spectra and mixtures, obtaining pure qualitative and quantitative results from target compounds using conventional tools seems impossible. As a result, multivariate calibration using PLSR and MCR-ALS may address this complex situation.

Figure 1 near here

At first, PLS was applied to develop a multivariate model using calibration set data. The **X**- and **Y**-block data sets were prepared before any PLS modelling using different preprocessing and variable selection methods. Various preprocessing methods including mean-centring, auto-scaling [8] and orthogonal signal correction (OSC)[13, 36] were tested and mean-centring was chosen as the most proper one because of the higher sensitivity and better statistical parameters than other preprocessing methods. As it has been already mentioned, variable selection methods of VIP[14, 15], RV, UVE[16, 17] and rPLS[18] were taken into account for PLS modelling in this study. In order to select the number of factors in PLS modelling, leave-one-out cross validation (LOO-CV)[37] was used and five latent variables (LVs) were chosen for UVE, VIP, RV and rPLS according to the minimum values of RMSE of cross-validation (RMSE-CV). It is important to point out that using LOO-CV helps us to select the significant number of components in the model and therefore, to avoid model overfitting. Table 3 shows obtained statistical parameters for

calibration set. Table 3 compares different PLSR variable selection methods in terms of R^2, RMSE and RE. The obtained statistical parameters of PLSR were RE 1.4 to 10.8%, $R^2$0.93 to 0.99and RMSE between 0.25 and 1.87. These parameters confirm better performance of rPLS than other methods almost for all analytes.

Table 3 near here

On the other hand, MCR-ALS was applied to develop a multivariate model using calibration set data under application of proper constraints, such as non-negativity[29], spectral normalization, known pure spectral profiles (equality constraint)[19], and correlation constraint. In addition, SIMPLISMA[28] was used for calculation of the initial estimate of spectral profiles. Furthermore, the pure spectral profiles of standard compounds were used as initial estimates to check the MCR-ALS solutions and also to reduce the effects of rotational ambiguities.[19] In order to test MCR-ALS-CC ability to predict the calibration set itself and also to make a comparison between the results obtained by PLSR modelling with MCR-ALS-CC, calibration data set was used as second subset to perform validation. It is important to note that the results obtained by MCR-ALS with the correlation constraint was used to correct sample matrix effects by using the regression model for calibration set as a reference for validation and test sets. Table 3 shows the statistical parameters for calibration set. The obtained statistical parameters for MCR-ALS-CC were 2.7 to 13.7% RE, R^2 between 0.89 and 0.99 and RMSE values 0.46 to 2.35 which are comparable with PLSR.

In general, DCF shows worst statistical parameters for all methods and MEF shows the best ones. In this regard, DCF shows RE values 6.7% for PLS (rPLS as variable selection) and 11.5% for MCR-ALS and R^2 values 0.97 for PLS and 0.92 for MCR-ALS,

respectively. The possible reasons for this trend can be interpreted using Figure 1(d).In this figure, the UV-Vis spectrum of DCF depicts low absorbance intensity, lack of selective region and heavy overlap with other analytes. In contrast to DCF, MEF shows RE values 1.4% for PLS (rPLS as variable selection) and 2.67% for MCR-ALS and R^2values 0.99 for both methods. In MEF UV-Vis spectrum in Figure 1(d), there is a selective region for this component in wavelength window 340-390 nm. In addition, despite the existence of other analytes, the MEF signal is simply recognisable in five concentration levels in calibration set spectra (Figure 1(a)).

On the other hand, inspection of statistical parameters of calibration set in terms of RE, R^2 and RMSE for different variable selection methods showed that rPLS method has a better performance in comparison to three other variable selection methods for accurate calibrating of analytes. However, the performance of VIP, RV and UVE are also acceptable at a specified level. In addition, MCR-ALS with correlation constraint showed reasonable statistical performance for calibration set but not better than PLS with four different variable selection methods.

In summary, PLS showed better performance than MCR for calibration samples where the effect of matrix and possible interfering components do not exist. Now, it is important to test the prediction ability of PLSR and MCR-ALS-CC models with external sets (i.e., validation and test sets) which may include different matrix effects. The sample matrix was changed to test the applicability of the developed calibration model in the cases with different sample matrices.

4.2 Validation set

The raw spectra of different samples of validation set are shown in Figure 1(b) in wavelength range 240–411 nm. Table 2 shows the different concentration of validation set which adopted randomly in seven different samples. As same as calibration set, the absorbance data of validation set was optimized before any PLS modeling by preprocessing and variables selection methods. Table 3 shows statistical parameters for validation set in terms of RE, R^2 and RMSE, which provided a basis for comparison between PLSR (with different variable selection methods) and MCR-ALS-CC. The obtained statistical parameters of PLSR using rPLS, VIP, UVE and RV are approximately the same, shows 5.9 to 22.0% RE, 0.42 to 0.99 R^2 and 0.67 to 4.71 RMSE. The MCR-ALS algorithm was performed under application of same constraints as calibration set. The obtained statistical parameters for MCR-ALS-CC were 6.0 to 24.3% RE, R^2 between 0.3 and 0.99 and RMSE values between 0.71 to 5.23. As it can be seen, the statistical parameters for validation set get worse and this is due to the change in composition and concentration levels of the included samples. Again, PLSR shows a better performance than MCR-ALS. However, the main priority of MCR-ALS over PLS is the ability to recover the pure resolved spectral profile for each component even interference(s). As an instance, Figure2 shows the resolved MCR–ALS spectral profiles for validation set. This is one of the most important features of MCR-ALS which compensate the worse statistical parameters of this method over PLS for calibration and validation samples. Furthermore, the resolved spectral profile of interference(s) can help for identification of these compounds (for example GEM in this study).

Figure 2 near here

4.3 Real environmental samples

To test the applicability of the developed models to quantify target analytes in complex samples, three different environmental samples (tap, well and farm waters) were used as test set. Figure 1(c) shows the raw UV-Vis spectra of these samples in wavelength range 240–411 nm. Also, Table 2 shows the different spiked concentrations of test set which fixed at three concentration levels include 5, 15 and 25 ppm. The same preprocessing methods of calibration and validation sets were used for test set. Table 4shows statistical parameters for test set, which again provided a comparison between different models. However, in this case, only the best variable selection methods for PLS confirmed by in previous steps are kept (i.e., VIP and rPLS).

Table 4 near here

The obtained statistical parameters of PLSR using VIP and rPLS and also MCR-ALS were satisfactory by considering the heavy overlap of target analytes and complex samples matrices. The vast range of errors clarifies the difference between these models.

Figure 3 near here

As shown in Figure3 the maximum absorbance of matrix interferences in test set samples especially in well water samples, occurs in 280 nm where the maximum absorbance of DCF occurred (Figure 1(d)). It can be concluded that this overlapping is the main source of DCF prediction error. In contrast to PLSR which is more sensible to the existence of interference, MCR-ALS has better performance due to its nature. Therefore, the results of MCR-ALS for DCF and CBZ are much better than PLS by VIP and rPLS as variable selection methods. For NAP, the results of three methods are comparable. However, in case of MEF, PLS give much better results than MCR-ALS. In this case, PLS-VIP has better performance than rPLS.

Although, there is no clear trend whether use PLSR or MCR-ALS-CC, but generally PLSR preferred for determination of analytes in samples which are free of interference and have simple matrixes using proper data preprocessing and variable selection. In this study PLS-VIP shows better performance than other analysed methods. On the other hand, MCR-ALS using correlation constraint allowed for the accurate determination of analytes in the presence of unknown interference and more complex matrix. Note that additional advantage of MCR-ALS was the possible recovery of the spectral information of analytes and of possible unknown interferences (GMF in this study).

It should be pointed out that combination of a simple spectrophotometric method and multivariate calibration for simultaneous analysis of a complex mixture of five pharmaceuticals is comparable with the performance of sophisticated instrumental techniques, such as high-performance liquid chromatography-diode array detector (HPLC-DAD).[21] As an example, in one of the recent studies, same pharmaceuticals were extracted using solid-phase extraction (SPE) from water matrices and then were analyzed using HPLC-DAD.[21] The predicted concentrations obtained for spiked river and well water samples showed 5.0 to 9.3% and 6.33 to 10.67% RE, respectively which are comparable with the results of this study (Table 4). Therefore, in spite of the heavy overlap of the spectral profiles of target compounds and the lack of selectivity, multivariate calibration methods can get unbiased results. RE is a quantitative measure of the bias of an analytical method. According to the results in Table 4 for different spiked test samples, the RE values are acceptable.

Conclusions

Combination of multivariate calibration methods to spectroscopic techniques can be an instrumentally straight-forward and rapid technique to improve the identification and quantification ability of these methods for simultaneous determination of target compounds in different sample matrices. In this study, PLSR and MCR-ALS were used for determination of five pharmaceuticals in different environmental samples. PLSR performance was evaluated using different data preprocessing and variable selection methods including VIP, RV, UVE and rPLS. Also, MCR-ALS was evaluated using correlation constraint. Evaluating these methods based on statistical parameters of RE, R^2 and RMSE showed that PLSR with VIP and rPLS as variable selection methods are appropriate for accurate determination of analytes in samples containing simple matrices. On the other hand, MCR-ALS-CC preferred for determination of analytes in more complex sample matrices owing to additional advantage of MCR-ALS method for possible recovery of the spectral information of analytes and of possible unknown interferences in the analysed samples.

Acknowledgements

The Research Council of Sharif University of Technology is acknowledged for financial support of this project. The authors would like to thank Dr. M.R. Hormozi-Nezhad from Department of Chemistry, Sharif University of Technology for providing access to UV-Vis spectrophotometer.

Notes and references

1. J. Santos, I. Aparicio, M. Callejón and E. Alonso, *J. Hazard. Mater.*, 2009, **164**, 1509-1516.
2. M. J. Gómez, M. Petrović, A. R. Fernández-Alba and D. Barceló, *J. Chromatogr. A*, 2006, **1114**, 224-233.
3. C. Hao, X. Zhao and P. Yang, *Trends Anal. Chem.*, 2007, **26**, 569-580.
4. D. Bratkowska, N. Fontanals, P. Cormack, F. Borrull and R. Marcé, *J. Chromatogr. A*, 2012, **1225**, 1-7.
5. A. de Juan and R. Tauler, *Crit. Rev. Anal. Chem.*, 2006, **36**, 163-176.
6. M. M. Galera, M. D. G. García and H. C. Goicoechea, *Trends Anal. Chem.*, 2007, **26**, 1032-1042.
7. H. Parastar and N. Akvan, *Anal. Chim. Acta*, 2014, **816**, 18-27.
8. S. Wold, M. Sjöström and L. Eriksson, *Chemom. Intell. Lab. Syst.*, 2001, **58**, 109-130.
9. Z. Rezaei, B. Hemmateenejad, S. Khabnadideh and M. Gorgin, *Talanta*, 2005, **65**, 21-28.
10. J. Arancibia, A. Olivieri and G. Escandar, *Anal. Bioanal. Chem.*, 2002, **374**, 451-459.
11. L. F. Capitán-Vallvey, N. Navas, M. del Olmo, V. Consonni and R. Todeschini, *Talanta*, 2000, **52**, 1069-1079.
12. M. A. Cantarelli, R. G. Pellerano, E. J. Marchevsky and J. M. Camiña, *Anal. Sci.*, 2011, **27**, 73.
13. S. Wold, H. Antti, F. Lindgren and J. Öhman, *Chemom. Intell. Lab. Syst.*, 1998, **44**, 175-185.
14. A. Oussama, F. Elabadi, S. Platikanov, F. Kzaiber and R. Tauler, *J. Am.Oil Chem. Soc.*, 2012, **89**, 1807-1812.
15. I. G. Chong and C. H. Jun, *Chemom. Intell. Lab. Syst.*, 2005, **78**, 103-112.
16. V. Centner, D.-L. Massart, O. E. de Noord, S. de Jong, B. M. Vandeginste and C. Sterna, *Anal. Chem.*, 1996, **68**, 3851-3858.
17. W. Cai, Y. Li and X. Shao, *Chemom. Intell. Lab. Syst.*, 2008, **90**, 188-194.
18. Å. Rinnan, M. Andersson, C. Ridder and S. B. Engelsen, *J. Chemom.*, 2014, **28**, 439-447.
19. A. de Juan, J. Jaumot and R. Tauler, *Anal. Methods*, 2014, **6**, 4964-4976.
20. T. Azzouz and R. Tauler, *Talanta*, 2008, **74**, 1201-1210.
21. N. Akvan and H. Parastar, *Chemom. Intell. Lab. Syst.*, 2014, **137**, 146-154.

22. M. Antunes, J. J. Simão, A. Duarte and R. Tauler, *Analyst*, 2002, **127**, 809-817.
23. L. B. Lyndgaard, F. van den Berg and A. de Juan, *Chemom. Intell. Lab. Syst.*, 2013, **125**, 58-66.
24. R. R. de Oliveira, K. M. G. de Lima, R. Tauler and A. de Juan, *Talanta*, 2014, **125**, 233-241.
25. T. Mehmood, K. H. Liland, L. Snipen and S. Sæbø, *Chemom. Intell. Lab. Syst.*, 2012, **118**, 62-69.
26. Z. Xiaobo, Z. Jiewen, M. J. W. Povey, M. Holmes and M. Hanpin, *Anal. Chim. Acta*, 2010, **667**, 14-32.
27. C. Abrahamsson, J. Johansson, A. Sparén and F. Lindgren, *Chemom. Intell. Lab. Syst.*, 2003, **69**, 3-12.
28. W. Windig and J. Guilment, *Anal. Chem.*, 1991, **63**, 1425-1432.
29. R. Bro and S. De Jong, *J. Chemom.*, 1997, **11**, 393-401.
30. P. W. Araujo and R. G. Brereton, *Trends Anal. Chem.*, 1996, **15**, 63-70.
31. L. M. Fang and M. Lin, *Chinese J. Anal. Chem.*, 2008, **36**, 815-818.
32. Z. Liu, W. Cai and X. Shao, *J. Chromatogr. A*, 2008, **1190**, 358-364.
33. G. Wang, Q. Ding and Z. Hou, *Trends Anal. Chem.*, 2008, **27**, 368-376.
34. L. Mutihac and R. Mutihac, *Anal. Chim. Acta*, 2008, **612**, 1-18.
35. H. Li, Q. Xu and Y. Liang, *LibPLS, www.libpls.info*.(Accessed July 2015).
36. T. Fearn, *Chemom. Intell. Lab. Syst.*, 2000, **50**, 47-52.
37. R. Kohavi, *Int. Joint Conf. Artif. Intell.*, 1995, **14**, 1137-1145.

Figure captions

Figure 1 The raw UV-Vis spectra of different samples. (a) Calibration set, (b) validation set, (c) test set (tap, well and farm waters at three concentration levels 5, 15 and 25 μgmL^{-1}) and (d) pure standards at concentration of 15.0 μgmL^{-1}.

Figure 2 Resolved MCR–ALS-CC spectral profiles for four target analytes DCF, NAP, MEF and CBZ and interference GEM in standard samples. The summation of the spectra for five components is also shown.

Figure 3 UV-Vis spectra for tap, well and farm water samples as test set without pharmaceuticals.

Figure 1

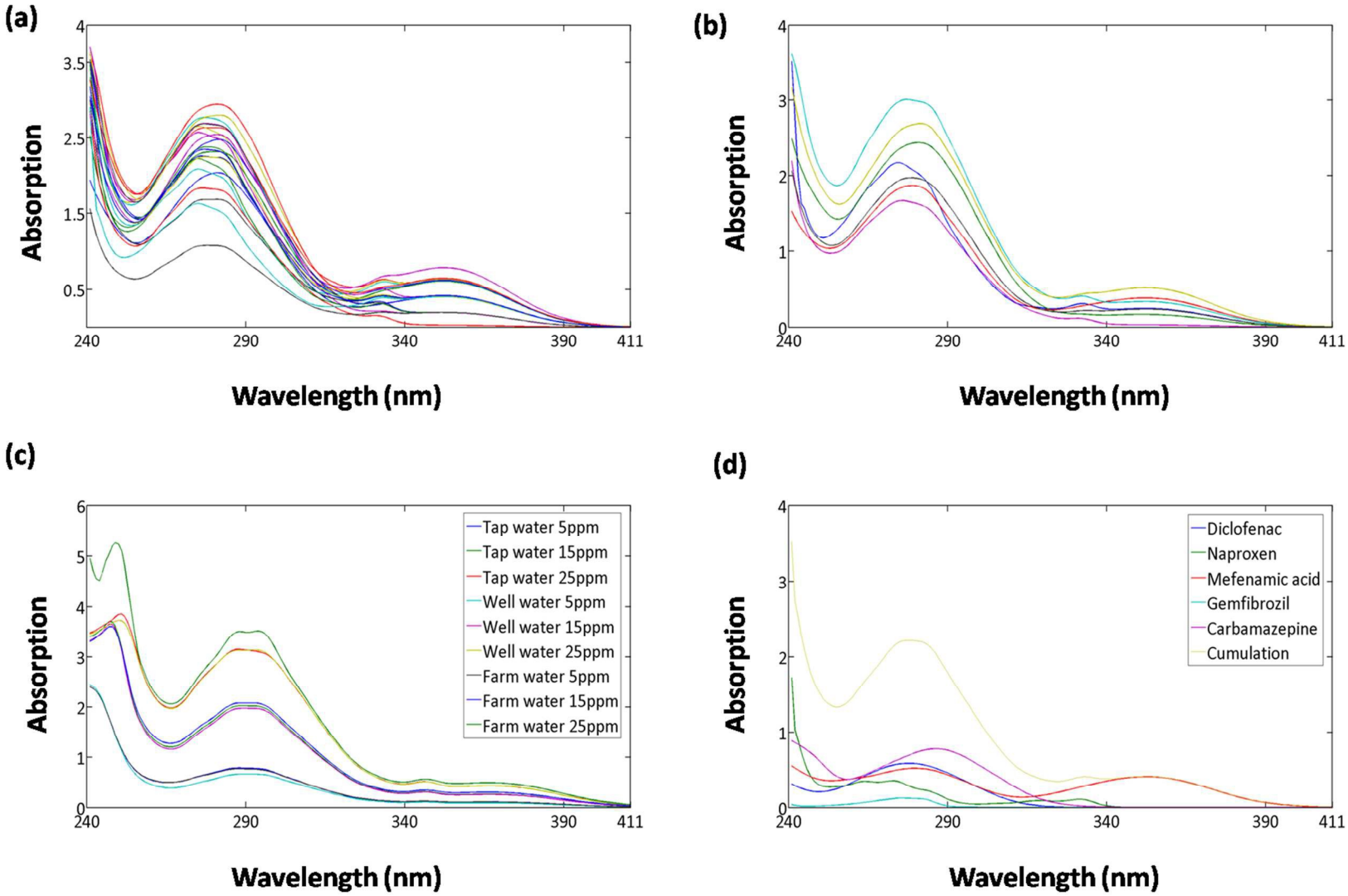


Figure 2

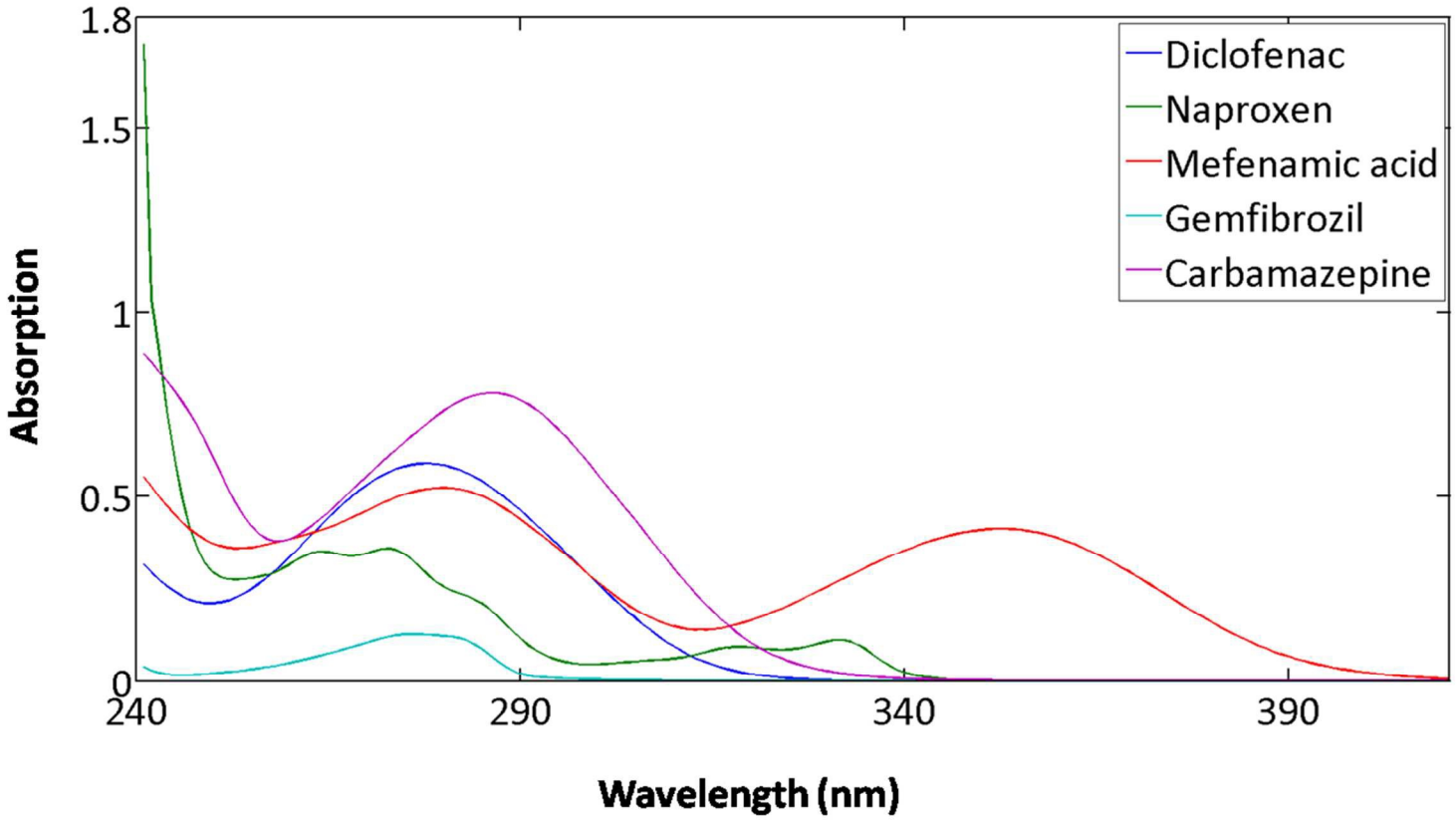


Figure 3

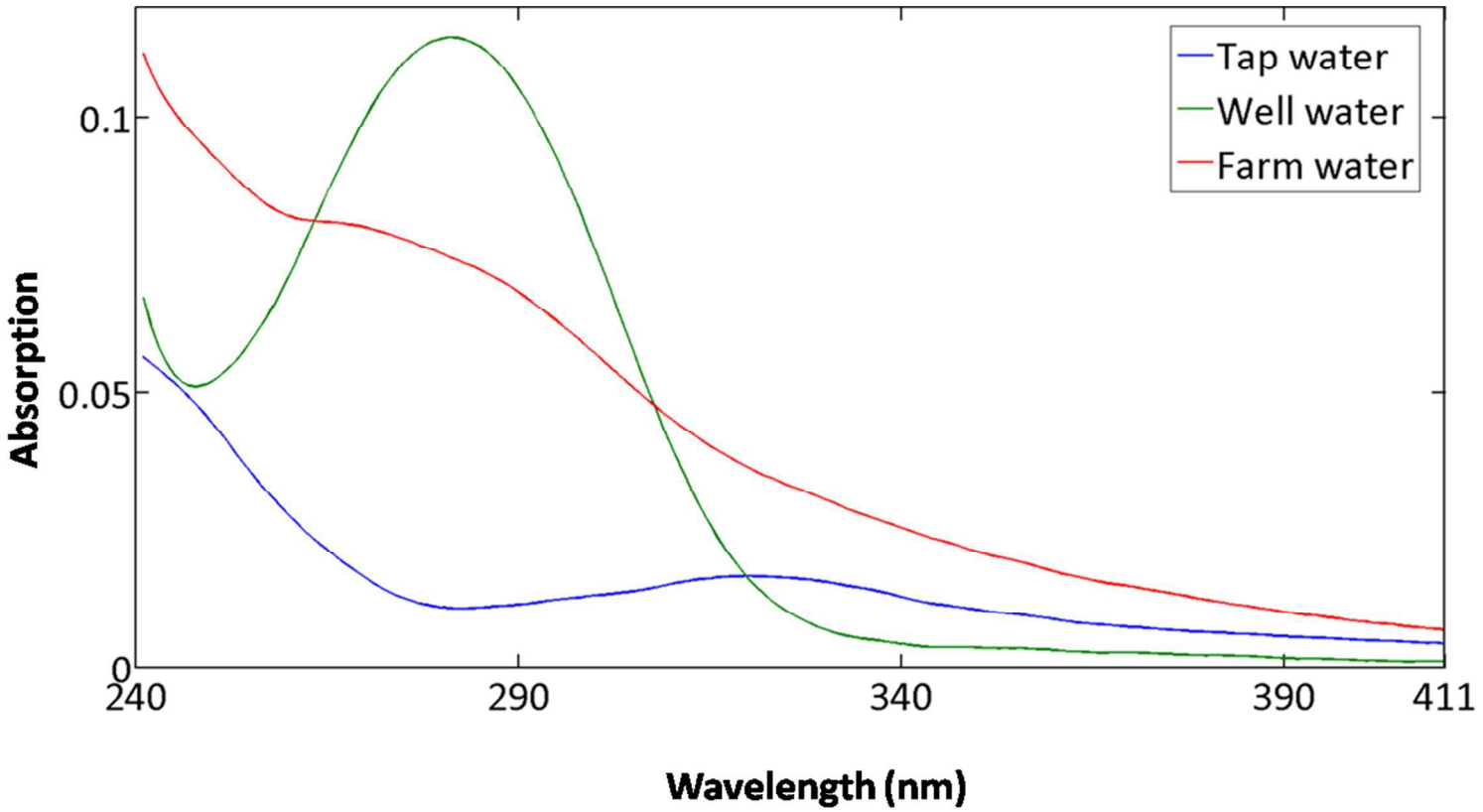


Table 1 Pharmaceuticals selected as target analytes and their physicochemical properties.

Name	Purity (%)	Molecular Mass ($gmol^{-1}$)	Molecular Formula	pKa
Diclofenac	99.66	296.148	$C_{14}H_{11}Cl_2NO_2$	4.15
Naproxen	99.6	230.259	$C_{14}H_{14}O_3$	4.15
Mefenamic Acid	99.76	241.285	$C_{15}H_{15}NO_2$	4.18
Gemfibrozil	99.8	250.333	$C_{15}H_{22}O_3$	4.75
Carbamazepine	99.6	236.269	$C_{15}H_{12}N_2O$	13.94

Table 2 The concentration matrix used to prepare calibration, validation and test set.

Set	Number	DCF(ppm)	NAP(ppm)	MEF(ppm)	GMF(ppm)	CBZ(ppm)	Matrix
Calibration	1	15.5	1	15.5	15.5	15.5	-
	2	23.5	7.5	23.5	23.5	7.5	-
	3	7.5	23.5	23.5	7.5	23.5	-
	4	7.5	23.5	23.5	23.5	7.5	-
	5	15.5	15.5	30	15.5	15.5	-
	6	23.5	23.5	23.5	7.5	7.5	-
	7	1	15.5	15.5	15.5	15.5	-
	8	7.5	7.5	23.5	23.5	23.5	-
	9	15.5	15.5	15.5	1	15.5	-
	10	15.5	15.5	1	15.5	15.5	-
	11	30	15.5	15.5	15.5	15.5	-
	12	15.5	30	15.5	15.5	15.5	-
	13	15.5	15.5	15.5	15.5	30	-
	14	23.5	23.5	7.5	7.5	23.5	-
	15	7.5	23.5	7.5	23.5	23.5	-
	16	23.5	23.5	7.5	23.5	7.5	-
	17	23.5	7.5	23.5	7.5	23.5	-
	18	15.5	15.5	15.5	15.5	1	-
	19	23.5	7.5	7.5	23.5	23.5	-
	20	15.5	15.5	15.5	15.5	15.5	-
	21	7.5	7.5	7.5	7.5	7.5	-
	22	15.5	15.5	15.5	30	15.5	-
Validation	1	24	20	9	23.5	1	-
	2	19.5	4	6	7	24	-
	3	16.5	1	14	20.5	8	-
	4	26.5	21.5	12	18	23	-
	5	19	11.5	1	6.5	9	-
	6	15	8	19	14.5	25	-
	7	26	6	8.5	4.5	10	-
Test	1	5	5	5	5	5	Tap Water
	2	15	15	15	15	15	Tap Water
	3	25	25	25	25	25	Tap Water
	4	5	5	5	5	5	Well Water
	5	15	15	15	15	15	Well Water
	6	25	25	25	25	25	Well Water
	7	5	5	5	5	5	Farm Water
	8	15	15	15	15	15	Farm Water
	9	25	25	25	25	25	Farm Water

Table 3 Comparison of PLSR variable selection methods and MCR-ALS-CC AFOMs for modeling and quantification of calibration and validation set.

		Calibration					Validation				
		PLS				MCR-ALS	PLS				MCR-ALS
		UVE	VIP	RV	rPLS		UVE	VIP	RV	rPLS	
DCF	R^2	0.93	0.95	0.96	0.97	0.92	0.44	0.42	0.47	0.43	0.30
	RMSE	1.87	1.72	1.44	1.17	1.98	4.18	4.40	3.76	4.71	5.23
	RE (%)	10.76	9.88	8.27	6.68	11.49	19.58	20.59	17.62	22.06	21.07
NAP	R^2	0.97	0.98	0.94	0.98	0.89	0.97	0.96	0.92	0.98	0.85
	RMSE	1.24	1.12	1.70	0.95	2.35	1.37	1.32	2.19	1.40	3.49
	RE (%)	7.16	6.47	9.82	5.45	13.66	10.89	10.52	17.41	11.10	24.31
MEF	R^2	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
	RMSE	0.43	0.29	0.31	0.25	0.46	0.89	0.70	0.67	0.73	0.71
	RE (%)	2.44	1.67	1.79	1.41	2.67	7.84	5.93	5.97	6.45	5.99
CBZ	R^2	0.96	0.98	0.95	0.98	0.94	0.98	0.97	0.97	0.98	0.97
	RMSE	1.38	1.64	1.55	1.10	1.68	1.25	1.56	1.66	1.37	2.18
	RE (%)	7.94	9.42	8.92	6.31	9.71	7.43	9.30	9.89	8.15	14.16

Table 4 Comparison of PLSR variable selection methods and MCR-ALS AFOMs for modeling and quantification of test set (spiked tap, well and farm waters at three concentration levels 5, 15 and 25 µg.mL^{-1}).

		Tap water			Well water			Farm water		
		PLS - VIP	rPLS	MCR-ALS	PLS - VIP	rPLS	MCR-ALS	PLS - VIP	rPLS	MCR-ALS
DCF	R^2	0.99	0.99	0.99	0.98	0.48	0.99	0.99	0.93	0.99
	RE (%)	16.34	19.79	13.28	18.32	46.05	12.44	11.63	22.41	4.41
NAP	R^2	0.97	0.99	0.71	0.91	0.96	0.58	0.86	0.99	0.56
	RE (%)	12.30	8.40	32.46	21.70	16.85	37.07	22.62	9.46	37.09
MEF	R^2	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
	RE (%)	1.11	1.34	3.02	6.97	5.33	9.79	13.05	15.48	17.00
CBZ	R^2	0.99	0.97	0.99	0.99	0.85	0.99	0.99	0.99	0.99
	RE (%)	5.33	8.83	2.97	4.22	25.03	3.37	9.39	14.43	7.45