Exploiting the synergistic effect of concurrent data signals: low-level fusion of liquid chromatographic with dual detection data

Carla M. Teglia, Silvana M. Azcarate, Mirta R. Alcaráz, Héctor C. Goicoechea, María J. Culzoni

www.elsevier.com/locate/talanta

PII: S0039-9140(18)30447-8
DOI: https://doi.org/10.1016/j.talanta.2018.04.090
Reference: TAL18629

To appear in: *Talanta*

Received date: 6 February 2018
Revised date: 24 April 2018
Accepted date: 27 April 2018

Cite this article as: Carla M. Teglia, Silvana M. Azcarate, Mirta R. Alcaráz, Héctor C. Goicoechea and María J. Culzoni, Exploiting the synergistic effect of concurrent data signals: low-level fusion of liquid chromatographic with dual detection data, *Talanta,* https://doi.org/10.1016/j.talanta.2018.04.090

Exploiting the synergistic effect of concurrent data signals: low-level fusion of liquid chromatographic with dual detection data

Carla M. Teglia[a,c], Silvana M. Azcarate[b,c], Mirta R. Alcaráz[a,c], Héctor C. Goicoechea[a,c], María J. Culzoni[a,c,*]

[a]Laboratorio de Desarrollo Analítico y Quimiometría (LADAQ), Cátedra de Química Analítica I, Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral, Ciudad Universitaria, Santa Fe (S3000ZAA), Argentina.

[b]Facultad de Ciencias Exactas y Naturales, Universidad Nacional de La Pampa, and Instituto de Ciencias de la Tierra y Ambientales de La Pampa (INCITAP), Av. Uruguay 151, Santa Rosa (L6300CLB), La Pampa, Argentina.

[c]Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Godoy Cruz 2290 CABA (C1425FQB), Argentina.

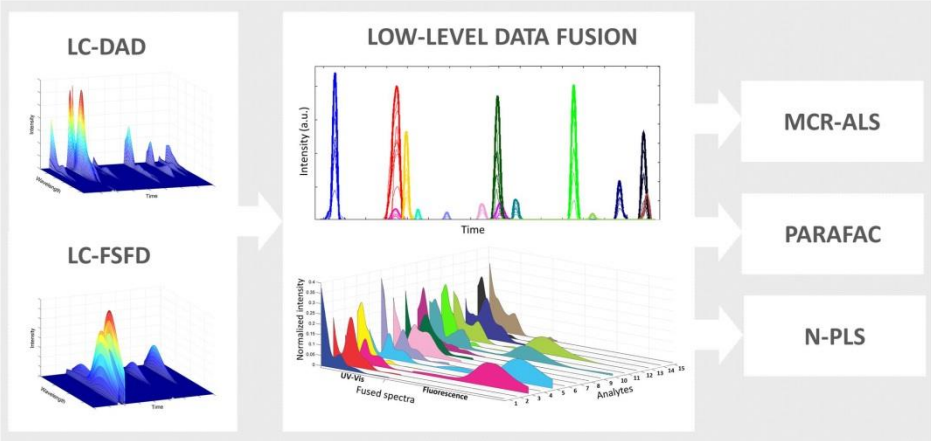
*Corresponding author: E-mail: mculzoni@fbcb.unl.edu.ar (M. J. Culzoni)

Abstract

A low-level data fusion strategy was developed and implemented for data processing of second-order liquid chromatographic data with dual detection, i.e. absorbance and fluorescence monitoring. The synergistic effect of coupling individual information provided by two different detectors was evaluated by analyzing the results

gathered after the application of a series of data preprocessing steps and chemometric resolution. The chemometric modeling involved data analysis by MCR-ALS, PARAFAC and N-PLS. Their ability to handle the new data block was assessed through the estimation of the analytical figures of merits achieved in the prediction of a validation set containing fifteen fluorescent and non-fluorescent veterinary active ingredients that can be found in poultry litter. Eventually, the feasibility of the application of the fusion strategy to real poultry litter samples containing the studied compounds was verified.

Graphical abstract



Keywords: Low-level data fusion; Liquid chromatography; Second-order data; Chemometrics; Dual detection

1. Introduction

Liquid chromatography (LC) is a versatile technique appropriate to analyze a wide variety of compounds [1]. In recent years, LC coupled to diode array detection (DAD) and/or fast-scanning fluorescence detection (FSFD) in association with multiway analysis have allowed improving the analytical performance, providing methods attaining highly informative results [2,3]. Aligned with this progression of gathering more and better information, the data fusion (DF) strategy allows obtaining accurate knowledge about a system and improving prediction results in comparison with those obtained using a single technique [4,5].

Concisely, the DF analysis consists in combining the outputs arisen from different sources and jointly processes them as a unique block [6,7]. In this regard, the possible alternatives for fusing the data are generally grouped into three strategies, depending on the way in which the fusion occurs. Low-level data fusion implies that the outputs describing the individual blocks are concatenated to build a single data array. On the other hand, in the mid-level fusion approach, the relevant features are firstly extracted from each original data source, and then concatenated evolving into a unique block. Eventually, either low- or mid-level fused data are processed in order to build a final model. On the contrary, the high-level strategy involves building separate models for the different data blocks and, then, the obtained responses are integrated into a single final response [8,9].

Recently, original multiway data acquired using the same instrument, i.e. a liquid chromatographer, and two different detection modes, i.e. DAD and FSFD, have been fused to improve the quantitation of endocrine disruptors. It has been demonstrated that the application of DF allowed achieving better results in comparison with those obtained from the analysis of the data provided by individual detectors [10]. Therefore, in an attempt to extract the hidden information from complex systems and improve the prediction ability of the analytical method, the fusion of highly informative data arising from different sources for a given sample becomes a strategy worth to be implemented. In this sense, Silvestri et al. proposed a mid-level fusion strategy in which the synergistic effect resulting from the modeling of merged data gathered from three different sources, i.e. nuclear magnetic resonance, fluorescence excitation–emission spectroscopy and LC-DAD, allowed acquiring the information that otherwise would not be possible to obtain [8]. Besides, some researchers have applied different types of DF to the same data set, with the purpose of comparing performances in terms of quality of the analytical results [7–9,11].

In order to exploit the empowering effect of attaching individual information provided by two different detectors by low-level DF, we developed an effective approach based on a series of data preprocessing steps that comprises the extraction of the main features of the system, which constitutes a significant improvement in the chromatographic DF analysis. In this case, the challenge consisted in finding the optimal combination of data preprocessing and data modeling that provide the best analytical figures. For that, LC-DAD and LC-FSFD second-order data matrices were firstly modeled by multivariate curve resolution-alternative least square (MCR-ALS) to eliminate the interferences. Then, the matrices were rebuilt only considering the analytes under study. To the best of our knowledge, this is the first time MCR-ALS is

applied as a data clean-up technique in the context of a DF strategy. In order to correct shifting and distortions of the chromatographic peaks of the analytes that contribute signal to both LC-DAD and LC-FSFD matrices, the correlation optimized warping (COW) [12] methodology was performed. Thereby, merged matrices free of time shifts, i.e. trilinear data, were generated [13,14]. Eventually, the fused matrices were submitted to three different chemometric modeling, i.e. MCR-ALS, parallel factor analysis (PARAFAC) [15,16] and N-way partial least squares (N-PLS) [17,18], and a comparative evaluation of their ability to handle the new data block with quantitative purposes was carried out through the estimation of the analytical figures of merit and the relative error of prediction (REP %) over a validation set. Eventually, the feasibility of the application of the DF strategy to real sample matrices containing the studied compounds was verified. As a case study, fifteen fluorescent and non-fluorescent veterinary active ingredients (AIs) were investigated in poultry litter samples. These AIs are orally administered and poorly absorbed in the animal gut; therefore, they are excreted unchanged in urine and feces [19]. Several studies have revealed the presence of numerous classes of veterinary residues in different environmental matrices close to agricultural areas presenting a negative impact on the environment [20,21].

2. Materials and methods

2.1. Materials and reagents

Trimethoprim (TMP), ceftiofur (CFT), pyrantel pamoate (PYR) and fenbendazole (FBZ) were purchased from Vetranal (Sigma-Aldrich Inc, St Louis, USA). Imidacloprid (IMD) and chlortetracycline (CTC) were acquired from Pestanal (Sigma-Aldrich Inc, St Louis, USA). Clenbuterol hydrochloride (CLB), chloramphenicol (CAP), flumenique

(FLU), albendazole (ABZ), flunixin meglumine (FXN), diazepam (DZP) and menbutone (MBT) were purchased from Sigma (Sigma-Aldrich Inc, St Louis, USA). Enrofloxacin (ENR) and difloxacin (DIF) were purchased from Fluka (Buchs, Switzerland). Sodium phosphate monobasic p.a. and phosphoric acid p.a. were provided by Cicarelli (San Lorenzo, Argentina). HPLC-grade acetonitrile (ACN) and methanol (MeOH) were obtained from Merck (Darmstadt, Germany). Dimethylformamide (DMF) was purchased from Cicarelli (San Lorenzo, Argentina).

Stock standard solutions were prepared by exactly weighing and dissolving an amount of each drug in MeOH (CFT, CLB, ENR, IMD, DIF, CTC, FLU, and DZP), ACN (TMP, MBT, and CAP), water (FXN) or DMF (PYR, FBZ and ABZ), as appropriate. The final concentration of each solution was 1.00 mg mL^{-1}.

2.2. Chromatographic conditions and data generation

The experiments were performed using an Agilent 1100 LC instrument (Agilent Technologies, Waldbronn, Germany), equipped with a quaternary pump, degasser membrane, autosampler, oven column compartment, UV-Vis DAD, FSFD and the Chemstation software package (Agilent Technologies, Waldbronn, Germany) to control the instrument, the data acquisition and the data analysis. The procedure separation was carried out on a 3.5 μm Zorbax Eclipse XDBC18 analytical column (4.6×50 mm) (Agilent Technologies, Waldbronn, Germany).

All the chromatographic experiments were performed in gradient mode, setting the column temperature at 40 °C and the flow rate at 0.65 mL min^{-1}. The initial composition of the mobile phase consisted of MeOH:ACN:phosphate buffer 10 mmol L^{-1} pH = 3.50 (2.5:2.5:95). Then, the gradient elution was performed as follow: 0-1 min: 2.5 % MeOH and 2.5 % ACN, 18-21 min: 40 % MeOH and 40 % ACN, and 23-

25 min: 2.5 % MeOH and 2.5 % ACN. The complete analysis was carried out in 25 min.

LC-DAD data matrices ($\mathbf{D}^0_{DAD}$) were registered in the spectral range from 200 nm to 500 nm, in the elution time from 0.0 min to 25.0 min. On the other hand, LC-FSFD data matrices ($\mathbf{D}^0_{FSFD}$) were registered in the emission spectral range from 310 nm to 600 nm, with the excitation wavelength fixed at 280 nm, which was selected as a compromise wavelength considering the excitation wavelength of maximum intensity of each analyte, in the same elution time range. In this way, $\mathbf{D}^0_{DAD}$ and $\mathbf{D}^0_{FSFD}$ consisted in 1868 × 151 and 940 × 291 data points for time and spectral dimensions, respectively.

2.3. Calibration and validation samples

A calibration set containing thirteen samples was prepared by triplicate immediately before analysis by spiking the appropriate volumes of each stock standard solution in 1.00 g of basal poultry litter. Final concentrations, given in μg g^{-1}, ranged from 0.27–10.96 for TMP, 0.06–0.59 for ENR, 0.31–10.95 for IMD, 0.44–10.39 for CLB, 0.04–0.56 for DIF, 0.57–10.34 for CFT, 0.57–10.30 for CAP, 0.38–14.53 for CTC, 0.20–2.54 for FLU, 0.37–13.18 for PYR, 0.28–10.76 for MBT, 0.17–5.51 for ABZ, 0.60–11.80 for FXN, 0.32–10.40 for DZP and 0.36–13.49 for FBZ. A 9-sample validation set was built considering random concentrations of AIs different from those used for calibration but spanning the same concentration range.

2.4. Poultry litter sampling and veterinary drug extraction

Five samples from commercial farms of different locations: PL1 and PL3 (Crespo, Entre Ríos, Argentina), PL2 y PL4 (Barrancas, Santa Fe, Argentina) and PL5 (Esperanza, Santa Fe, Argentina) were collected.

Optimum experimental conditions for extraction of the veterinary drugs in poultry litter were optimized in a previous work [22] and followed as stated. The extraction procedure was as follows: 1.00 g of calibration, validation or real sample was accurately weighed and placed into a 20 mL plastic centrifuge tube. Then, 5.00 mL of the extracting solution consisting in MeOH:phosphate buffer 10 mmol L^{-1} pH =3.50 (33:67) were added and the mixture was vortexed for 2 min, sonicated for 18 min, shaken for 6 min, and centrifuged at 3500 rpm for 10 min. Subsequently, the supernatant was filtered and 100 μL were injected into the LC system.

2.5. Chemometric algorithms and software

All the algorithms were implemented in MATLAB 9.2 (R2017a) [23]. MCR-ALS algorithms are available online at http://www.mcrals.info/. PARAFAC and N-PLS resolutions were carried out using MVC2 toolbox [24]. Homemade routines based on Eilers algorithm were applied to perform second-order data baseline correction [25]. COW routine [12] used to correct distortions for the chromatographic peaks is available at http://www.models.life.ku.dk/DTW_COW.

3. Results and discussion

3.1. General considerations

$\mathbf{D}^{0}_{DAD}$ and $\mathbf{D}^{0}_{FSFD}$ data sets with fifteen compounds and seventeen interferences were used to evaluate the performance of the DF strategy proposed in this work. First, the matrices for all samples were recorded from each detector separately, and non-fluorescent (TMP, IMD, CLB, PYR, CFT, CTC, CAP, DZP, MBT, FXN and FBZ) and fluorescent (DIF, ENR, FLU and ABZ) compounds were quantitated through their

absorption and fluorescent signals, respectively. As a proof of concept of the ability of DF to help improve the analytical performance, especially related to selectivity and sensitivity enhancement, the individual matrices were properly fused and concurrently analyzed.

3.2. Data preprocessing

In order to prepare the data for the fusion, chromatographic data preprocessing joint to extraction of relevant features was carried out. To favor a succinct explanation, the whole procedure, which is summarized in Fig. 1, was described considering its implementation for a single sample, although it was simultaneously applied to similar time-appended matrices. It consists in a series of actions to overcome issues present in both $\mathbf{D}^{\mathbf{0}}_{\mathrm{DAD}}$ and $\mathbf{D}^{\mathbf{0}}_{\mathrm{FSFD}}$ data matrices, and is explained in detail below.

3.2.1. Baseline correction and region selection

Second-order data baseline correction on $\mathbf{D}^{\mathbf{0}}_{\mathrm{DAD}}$ and $\mathbf{D}^{\mathbf{0}}_{\mathrm{FSFD}}$ matrices was performed applying a multidimensional extension of the asymmetric least-square method. Then, aiming at facilitating the chemometric resolution and extracting the more relevant features from each data source, the individual matrices were partitioned in the temporal mode into three sections of sizes 253 × 106, 341 × 106 and 211 × 106 for LC-DAD, and 120 × 181, 170 × 181 and 105 × 181 for LC-FSFD. As a result, the fused matrixes contained five analytes each (see Table 1).

3.2.2. Data complexity reduction by MCR-ALS

MCR-ALS performs a bilinear decomposition of a **D** matrix into three matrices according to the equation:

$$\mathbf{D} = \mathbf{C}\mathbf{S}^{T} + \mathbf{E} \quad (1)$$

where **C** is the matrix of the individual resolved chromatographic time profiles, **S** contains the spectral profiles of each component and **E** comprises the residuals [26].

The number of chemical components in the chromatographic segment under study was estimated by singular value decomposition (SVD) on augmented data matrices (see Table 1). The initial estimates of spectral profiles to start the ALS optimization were calculated using a method based on simple-to-use interactive self-modeling mixture analysis (SIMPLISMA) [27]. During the analysis, unimodality (in the temporal mode), non-negativity (in the temporal and spectral modes) and normalization (in the spectral mode) constraints were applied. As a result, the optimized time (**C**) and spectral (**S**) profiles for the analytes and the unknown compounds of each region were obtained. It is important to highlight that, from this point onwards, the contributions of the unknown components were dismissed and all the processing continued on data free of interferences.

3.2.3. Factor scaling and time interpolation

Previously to DF, each column of each $\mathbf{C}_{DAD}$ matrix was multiplied by a scaling factor to guarantee that their maximum intensities equal those of the related $\mathbf{C}_{FSFD}$ matrix, but maintaining the relation between the areas on the different samples, with the solely purpose of favoring the data visual inspection. Besides, it is mandatory to assure that the matrices to be fused comprise the same number of time sensors, since otherwise they cannot be fused. Therefore, since the time dimension of $\mathbf{D}_{DAD}$ and $\mathbf{D}_{FSDS}$ matrices is uneven due to the differences in the scanning frequency of the detectors, the time dimension of each $\mathbf{C}_{FSFD}$ matrix was interpolated using the MATLAB function *spline*.

3.2.4. Peak alignment and data fusion

It is worthwhile mentioning that lack of reproducibility in peak times and shapes is usually observed among chromatographic runs. Additionally, when dual detection is performed for a given sample, a constant time shift is encountered due to the time lag between the detectors. Thus, either bilinearity of fused single data or trilinearity of a third-order data array is not fulfilled. In this matter, several algorithms can be implemented to correct for peak distortions and turn the data into bi- and/or trilinear. COW is a mathematic procedure able to overcome shift and shape distortions in order to align by warping the chromatographic peaks with a reference, without modifying the relation between areas [12]. However, to the best of our knowledge, there is no alignment procedure capable of tackling the situation when highly overlapped signal or several peak distortions are present.

Under this scenario, a combination of MCR-ALS/COW [14] was here applied as a preprocessing alternative to reduce the complexity of the data by selective subtraction of the signals of the interferences and alignment of the target peaks. This strategy combines the essential characteristic of each algorithm, i.e. exploitation of the second-order advantage and correction for distortions of the chromatographic peaks. At this stage, it becomes crucial to remark that COW was applied to the MCR-ALS resolved time profiles before rebuilding the matrices for the subsequent chemometric modeling. The aligned concentration profiles were then multiplied by the transpose of $\mathbf{S}_{DAD}$ and $\mathbf{S}_{FSFD}$, as appropriate, to generate two individual matrices for each region, i.e. $\mathbf{D^1}_{DAD}$ and $\mathbf{D^1}_{FSFD}$, only including the target analytes. The purpose of applying the COW procedure is twofold: construct bilinear fused data matrices and build a trilinear data array ($\boldsymbol{D}^*$). Thus, for PARAFAC and N-PLS, both DAD-to-FSFD and run-to-run peak correction was performed, obtaining a trilinear data array. However, only DAD-to-FSFD peak

correction was required for MCR-ALS, since this algorithm can cope with lack of trilinearity (see Fig. 2).

Finally, DF was carried out concatenating on the time axis the individual matrices $\mathbf{D}^{1}_{\text{DAD}}$ and $\mathbf{D}^{1}_{\text{FSFD}}$ for each region to obtain fused single matrices $\mathbf{D}^{2}_{\text{DAD 1/FSFD 1}}$ (253 × 287), $\mathbf{D}^{2}_{\text{DAD 2/FSFD 2}}$ (341 × 287) and $\mathbf{D}^{2}_{\text{DAD 3/FSFD 3}}$ (211 × 287). Figure 2 E illustrates the spectral profiles of the analytes after data fusion.

3.3. Chemometric modeling

3.3.1. MCR-ALS predictive performance on individual vs. fused data

At first, augmented $\mathbf{D}^{0}_{\text{DAD}}$ and $\mathbf{D}^{0}_{\text{FSFD}}$ data matrices were modeled by MCR-ALS, and the performances of the resolutions were compared with those achieved for the fused data matrix ($\mathbf{D}^{*}$), which was built by appending on top of each other the fused $\mathbf{D}^{2}_{\text{DAD/FSFD}}$ matrices of each region, i.e. $\mathbf{D}^{2}_{\text{DAD 1/FSFD 1}}$, $\mathbf{D}^{2}_{\text{DAD 2/FSFD 2}}$ and $\mathbf{D}^{2}_{\text{DAD 3/FSFD 3}}$. Statistical figures were estimated according to Olivieri [28]. Tables 2 and 3 summarize the analytical figures of merits and REP %, respectively, computed for the non-fused and fused second-order data modeling. This algorithm was selected to cope with the loss of trilinearity found in the original matrices, since their constituent profiles in the time dimension were not constant from sample to sample. For this reason, data processing involved the construction of augmented matrices along the time mode.

The analytical sensitivity (γ) calculated for the MCR-ALS analysis of $\mathbf{D}^{0}_{\text{DAD}}$ and $\mathbf{D}^{0}_{\text{FSFD}}$ were in the ranges of 15-837 g μg^{-1} and 83-1623 g μg^{-1}, respectively, whereas those calculated from DF were significantly higher, i.e. in the range of 34-3427 g μg^{-1}. In addition, LOD and LOQ values reached with fused data are 1.7 to 8-folds lower than those achieved for the modeling of individual non-fused data matrices. This improvement could be ascribed to the enhancement in the sensitivity of the method,

possibly, by virtue of the pre-processing procedure conducted prior to DF (removal of interferences, noise reduction, etc.), concurrently with the improvement in the selectivity achieved after DF modeling for those analytes that present both absorption and fluorescent signal. Besides, the REP values in the range of 2.5-5.6 % for fused data were also notably better than those obtained for individual data (between 3.1-8.1 % for $\mathbf{D}^0_{DAD}$ and 7.5-9.3 % for $\mathbf{D}^0_{FSFD}$ data). Contrarily to this favorable tendency, in the particular case of CLB, the REP value increases from 3.7 to 3.8 for non-fused and fused data, respectively. This slight enhancement can be linked to the fact that CLB and CAP present highly similar spectra, and this does not constitute a drawback when the CLB modeling is carried out in non-fused $\mathbf{D}^0_{DAD}$ since the analytes belong to different regions. In light of the aforementioned facts, it can be stated that the analytical performance of DF, evaluated through figures of merit, shows an outstanding improvement in comparison to single data arrays.

With the purpose of assessing accuracy and precision of the predicted concentrations for the individual and fused data obtained by MCR-ALS, the elliptical joint confidence region (EJCR) [29] test was performed. As can be seen in Fig. 3, all ellipses include the theoretically expected point (1,0), suggesting that both approaches are appropriate for resolving the analytes under study. The sizes of the elliptical regions for fused data compared to non-fused data present evident dissimilarities between non-fluorescent (TMP, IMD, CLB, CFT, CTC, CAP, PYR, MBT, FXN, DZP and FBZ) and fluorescent (FLU, ENR, DIF and ABZ) compounds. There is no size variation among ellipses for non-fluorescent compounds for both non-fused and fused data sets, which allows infering about similar precision in their determination. However, the sizes of the ellipses for fluorescent compounds that underwent DF are significantly smaller than those acquired from data provided by individual detectors, revealing an enhancement in

the precision due to its application. These clear differences for fluorescent compounds have also been verified in a previous work [10].

3.3.2. Predictive performances of MCR-ALS, PARAFAC and N-PLS on low-level DF

The fused data was modelled by three well-known second-order calibration algorithms in order to assess their prediction performance. Details about analytes and interferences, size of each section, matrices scaling and fused matrix $\mathbf{D}^*$ are displayed in the Table 1.

After DAD-to-FSDS and run-to-run COW alignment, the fused data sets were arranged in three-way arrays fulfilling trilinearity conditions [30]. To appraise the potential of diversely structured algorithms in comparison to the bilinear-based model, the samples were analyzed by means of PARAFAC and N-PLS, and their results were compared with those provided by MCR-ALS. In PARAFAC modeling, direct trilinear decomposition was used as initialization method. In addition, the non-negativity constraint was applied to the three modes. The selection of the optimum number of components was performed using the core consistency analysis (CORCONDIA) [15]. The estimated number of components for the validation samples was fifteen, which was in accordance with the presence of fifteen different signals corresponding to each studied AI. The same phenomenon occurred for N-PLS analysis, in which the optimum number of latent variables was estimated by cross-validation.

The prediction results reported in Table 3 for the validation samples obtained through the application of the three algorithms were evaluated considering the root mean square error of prediction (RMSE) and the REP %. The comparison of their predictive performances shed light on the fact that the developed data preprocessing strategy allowed not only resolving the fused data with other algorithms than MCR-

ALS, but also achieving slightly better results, in this particular case with PARAFAC. This better performance can be due to both its inherent cubic structure and its unicity property. As an example, a significant REP % improvement was achieved for the analysis of FBZ, which is the analyte with fewer differences between samples in terms of changes in both retention time and peak shape. As an evident exception, the analysis of CLB still remained more favorable when conducted by MCR-ALS probably because the peak distortion correction was not efficient due to the magnitude of the differences found between samples, leading to a non-trilinear data array. Regarding N-PLS, the modeling demonstrated to be superior to cope with the high spectral collinearity between CAP and CLB, providing the best analytical results for CAP.

With respect to the fluorescent analytes, the main differences were observed between the results obtained with and without DF application, independently of the multivariate algorithm employed in the quantitation step, except for the case of DIF, in which PARAFAC and N-PLS also showed better performance than MCR-ALS.

Moreover, in most of the cases, the plots of the EJCR test for accuracy and precision assessment of the results provided by PARAFAC and N-PLS lead to smaller elliptical surfaces compared to the MCR-ALS results, which also contains the ideal point (1,0).

3.5. Analysis of the real samples

Five real samples acquired from different poultry farms of Entre Ríos and Santa Fe provinces were analyzed in order to prove the applicability of the proposed methodology. As can be observed in Table 4, the prediction results obtained by MCR-ALS for the non-fused matrices are in agreement with those provided by the three algorithms for the fused matrices. Nevertheless, it is important to highlight the fact that

FBZ, detected in sample PL3, and IMD, detected in sample PL5, are below the detection threshold when the analysis is separately performed with data obtained from individual detectors. Conversely, the synergist effect resulting from the combination of independent absorbance and fluorescence signals demonstrated that detection may also become possible in cases of extremely low analyte concentration.

4. Conclusions

The herein proposed strategy based on low-level DF combined to a thorough data preprocessing demonstrated to be an effective and powerful tool, which contributes to the analysis of complex second-order chromatographic data gathered by different detectors. The preprocessing strategy helps overcome particular issues such as baseline/background contribution, chromatographic peaks distortions and peak overlaps in the first place. Among its principal benefits, it allows making the data suitable for the application of a variety of algorithms in view of exploiting their intrinsic characteristics to improve the quality of the analytical results. As far as the performance of the models is concerned, low-level DF not only increases the global prediction ability but also decreases the uncertainty of each individual result.

5. Acknowledgment

The authors are grateful to Universidad Nacional del Litoral (Project CAI+D 2016-50120150100110LI), CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas, Project PIP-2015 Nº 0111) and ANPCyT (Agencia Nacional de Promoción Científica y Tecnológica, Project PICT 2014-0347) for financial support. C.M.T, S.M.A. and M.R.A. thank CONICET for their fellowship.

Reference

[1] A.K. Malik, C. Blasco, Y. Picó, Liquid chromatography‑mass spectrometry in food safety, J. Chromatogr. A 1217 (2010) 4018–4040

[2] H. Parastar, Fundamentals and Analytical applications of multiway calibration, in: Elsevier (Eds.), Data handling in Science and technology, Amsterdam, 2015, pp 314−324.

[3] G.M. Escandar, H.C. Goicoechea, A. Muñoz de la Peña, A.C. Olivieri, Second- and higher-order data generation and calibration: A tutorial, Anal. Chim. Acta 806 (2013) 8-26.

[4] S. Mas, R. Tauler, A. de Juan, Chromatographic and spectroscopic data fusion analysis for interpretation of photodegradation processes, J. Chromatogr. A 1218 (2011) 9260-9268.

[5] D.L Hall, J. Llinas, An introduction to multisensor data fusion, J. Proceedings in IEEE 85 (1997) 6-23.

[6] Y. Liu, S.D. Brown, Wavelet multiscale regression from the perspective of data fusion: new conceptual approaches, Bioanal. Chem. 380 (2004) 445−452.

[7] A. Bajoub, S. Medina-Rodríguez, M. Gómez-Romero, E.A. Ajal, M.G. Bagur-González, A. Fernández-Gutiérrez, A. Carrasco-Pancorbo, Assessing the varietal origin of extra-virgin olive oil using liquid chromatography fingerprints of phenolic compound, data fusion and chemometrics, Food Chem. 215 (2017) 245-255.

[8] M. Silvestri, A. Elia, D. Bertelli, E. Salvatore, C. Durante, M. Li Vigni, A.Marchetti, M. Cocchi, A mid level data fusion strategy for the varietal classification of Lambrusco PDO wines, Chemom. Intell. Lab. Syst.137 (2014) 181–189.

[9] K.A. Obisesan, A.M. Jiménez-Carvelo, L. Cuadros-Rodriguez, I. Ruisánchez, M.P. Callao, HPLC-UV and HPLC-CAD chromatographic data fusion for the authentication of the geographical origin of palm oil, Talanta 170 (2017) 413-418.

[10] R.B. Pellegrino Vidal, G.A. Ibañez, G.M. Escandar, Advantages of data fusion: first multivariate curve resolution analysis of fused liquid chromatographic second-order data with dual diode array-fluorescent detection, Anal. Chem. 89 (2017) 3029−3035.

[11] X. Hong, J. Wang, Detection of adulteration in cherry tomato juices based on electronic nose and tongue: Comparison of different data fusion approaches, J. Food Eng. 126 (2014) 89–97.

[12] N.P.V. Nielsen, J.M. Carstensen, J. Smedsgaard, Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping, J. Chromatogr. A 805 (1998) 17–35.

[13] J.A. Arancibia, P.C. Damiani, G.M. Escandar, G.A. Ibañez, A.C. Olivieri, A review on second- and third-order multivariate calibration applied to chromatographic data, J. Chromatogr. B 910 (2012) 22– 30.

[14] H. Parastar, N. Akvan, Multivariate curve resolution based chromatographic peak alignment combined with parallel factor analysis to exploit second-order advantage in complex chromatographic measurements, Anal. Chim. Acta 816 (2014) 18–27.

[15] R. Bro, PARAFAC. Tutorial and applications, Chemom. Intell. Lab. Syst. 38 (1997) 149-171.

[16] K.R. Murphy, C.A. Stedmon, D. Graeberc, R. Bro, Fluorescence spectroscopy and multi-way techniques. PARAFAC, Anal. Methods 5 (2013) 6557–6566.

[17] J. Öhman, P. Geladi, S. Wold, Residual bilinearization. Part 1: Theory and algorithms, J. Chemometrics 4 (1990) 79-90.

[18] A. García Reiriz, P.C. Damiani, A.C. Olivieri, Residual bilinearization combined with kernel-unfolded partial least-squares: A new technique for processing non-linear second-order data achieving the second-order advantage, Chemom. Intell. Lab. Syst. 100 (2010) 127–135.

[19] M. Solliec, A. Roy-Lachapelle, M.-O. Gasser, M. Coté, C. Généreux, S. Sauvé, Fractionation and analysis of veterinary antibiotics and their related degradation products in agricultural soils and drainage waters following swine manure amendment, Sci. Total Environ. 543 (2016) 524–535.

[20] M. Gbylik-Sokorska, A. Posyniak, T. Sniegocki, J. Zmudzki, Liquid chromatography–tandem mass spectrometry multiclass method for the determination of antibiotics residues in water samples from water supply systems in food-producing animal farms, Chemosphere 119 (2015) 8–15.

[21] K.-R. Kim, G. Owens, S.-I. Kwon, K.-H. So, D. Lee, Y.S. Ok, Occurrence and environmental fate of veterinary antibiotics in the terrestrial environment, Water Air Soil. Pollut. 214 (2011) 163–174.

[22] C.M. Teglia, P.M. Peltzer, S.N. Seib, R.C. Lajmanovich, M.J. Culzoni, H.C. Goicoechea, Simultaneous multi-residue determination of twenty one veterinary drugs in poultry litter by modeling three-way liquid chromatography with fluorescence and absorption detection data, Talanta 167 (2017) 442–452.

[23] MATLAB 9.2. The MathWorks Inc., Natick, Massachussets, USA, 2017.

[24] A.C. Olivieri, H.L. Wu, R.Q. Yu, MVC2: A MATLAB graphical interface toolbox for second-order multivariate calibration, Chemom. Intell. Lab. 96 (2009) 246–251.

[25] P.H.C. Eilers, Parametric time warping, Anal. Chem. 76 (2004) 404–411.

[26] S.C. Rutan, A. de Juan, R. Tauler, 2.15 - Introduction to multivariate curve resolution, in: S.D. Brown, R. Tauler, B. Walczak (Eds.) Comprehensive chemometrics, Elsevier, Oxford, 2009, pp. 249-259.

[27] W. Windig, J. Guilment, Interactive self-modeling mixture analysis, Anal. Chem. 63 (1991) 1425–1432.

[28] A. C. Olivieri, Analytical figures of merit: from univariate to multiway calibration, Chem. Rev. 114 (2014) 5358–5378.

[29] González, A. G.; Herrador, M. A.; Asuero, A. G. Intra-laboratory testing of method accuracy from recovery assays, Talanta 48 (1999) 729−736.

[30] A.C. Olivieri, G.M Escandar, A. Muñoz de la Peña, Second-order and higher-order multivariate calibration methods applied to non-multilinear data using different algorithms, TRAC-Trends Anal. Chem. 30 (2011) 607–617.

Table 1. Specific details about relevant data extraction and low-level DF

Region	Scaling Factor	Analytes	Matrix size	Components: analytes + interferences[a]
1	DAD*10	TMP, IMD,CLB, ENR, DIF	253 × 106	5 + 3 = 8
	FSFD	ENR, DIF	120 × 181	2 + 2 = 4
2	DAD*4	CTC, CAP, CFT, PYR, FLU	341 × 106	5 + 6 = 11
	FSFD	FLU	170 × 181	1 + 2 = 3
3	DAD*4.4	MBT, FBZ, DZP, FXN, ABZ	211 × 106	5 + 2 = 7
	FSFD	ABZ	105 × 181	1 + 2 = 3
Data Fusion				
D*	DAD-FSFD	TMP, IMD, CLB, ENR, DIF, CTC, CAP, CFT, PYR, FLU, MBT, FBZ, DZP, FXN, ABZ	805 × 287	15 + 0 = 15[b]

[a] Number of interferences estimated by SVD
[b] Interferences were removed before DF

Table 2. Figures of merit computed for second-order data modeling

	Figures of merit														
	γ (g μg^{-1})					LOD (μg g^{-1})					LOQ (μg g^{-1})				
	Non-fused		Fused			Non-fused		Fused			Non-fused		Fused		
Analyte	MCR-DAD	MCR-FSFD	MCR	PARAFAC	N-PLS	MCR-DAD	MCR-FSFD	MCR	PARAFAC	N-PLS	MCR-DAD	MCR-FSFD	MCR	PARAFAC	N-PLS
ENR	837	1623	3427	3843	3872	0.0200	0.0006	0.0002	0.002	0.00008	0.0061	0.0020	0.0005	0.006	0.003
DIF	528	1582	2422	2900	2895	0.0270	0.0055	0.00007	0.001	0.00001	0.0080	0.0010	0.0011	0.003	0.00003
FLU	62	114	437	536	539	0.0920	0.0400	0.0021	0.024	0.002	0.236	0.110	0.059	0.068	0.007
ABZ	15	83	365	335	319	0.0920	0.0400	0.0021	0.024	0.002	0.225	0.140	0.033	0.027	0.047
TMP	102		135	132	133	0.0400		0.0012	0.026	0.005	0.120		0.034	0.075	0.0133
IMD	36		49	47	47	0.1200		0.0155	0.026	0.005	0.3600		0.0440	0.064	0.016
CL	30		36	38	4	0.		0.	0.0	0.	0.		0.	0.1	0.

B					0	1	04	62	02	4	1	91	07
						4	2		3	6	3		2
						0				0	0		
CT	126		12	152	1	0.	0.	0.0	0.	0.	0.	0.0	0.
C			3		5	0	00	06	00	1	0	20	01
					6	3	8		5	0	2		6
						0				0	7		
CA	20		34	30	2	0.	0.	0.0	0.	0.	0.	0.1	0.
P					8	1	05	57	04	5	1	79	14
						9	5		6	9	7		1
						0				0	4		
CF	26		52	60	6	0.	0.	0.0	0.	0.	0.	0.0	0.
T					1	1	01	30	00	4	0	88	02
						6	7		7	6	4		1
						0				0	9		
PY	77		79	108	1	0.	0.	0.0	0.	0.	0.	0.0	0.
R					1	0	01	15	01	1	0	47	03
					3	5	8		0	6	5		1
						0				0	6		
M	85		95	97	9	0.	0.	0.0	0.	0.	0.	0.0	0.
BT					8	0	02	10	00	1	0	29	00
						5	2		3	5	6		9
						0				0	6		
FB	82		10	103	1	0.	0.	0.0	0.	0.	0.	0.0	0.
Z			8		1	0	01	09	01	1	0	26	04
					4	5	3		6	5	3		7
						0				0	9		
DZ	82		11	137	1	0.	0.	0.0	0.	0.	0.	0.0	0.
P			0		3	0	01	10	00	1	0	30	00
					5	5	2		2	6	3		5
						0				0	8		
FX	39		42	51	5	0.	0.	0.0	0.	0.	0.	0.0	0.
N					3	1	00	21	03	3	1	63	10
						1	36		6	3	1		8
						0				0	0		

γ: analytical sensitivity, LOD: limit of detection, and LOQ: limit of quantitation estimated according to Ref. [28]

Table 3. Predictions on validation samples by modeling of non-fused and fused second-order data

Analytes	RMSE (μg g^{-1}) [a]				REP%[b]			
	Non-fused	Fused			Non-fused	Fused		
	MCR-ALS	MCR-ALS	PARAFAC	N-PLS	MCR-ALS	MCR-ALS	PARAFAC	N-PLS
ENR	0.02	0.01	0.01	0.01	7.5	4.0	4.0	4.3
DIF	0.03	0.01	0.01	0.01	9.3	5.1	2.8	3.5
FLU	0.11	0.06	0.04	0.05	7.5	3.7	2.8	3.2
ABZ	0.18	0.08	0.09	0.08	6.7	3.1	3.5	3.2
TMP	0.17	0.14	0.29	0.28	3.1	2.5	5.2	5.1
IMD	0.21	0.16	0.16	0.22	3.7	3.0	2.8	3.9
CLB	0.17	0.17	0.18	0.21	3.7	3.8	4.0	4.7
CTC	0.33	0.31	0.31	0.24	3.7	3.5	3.5	2.7
CAP	0.29	0.26	0.27	0.23	5.6	5.0	5.2	4.5
CFT	0.36	0.16	0.23	0.20	8.1	3.5	5.2	4.7
PYR	0.30	0.24	0.19	0.21	6.2	4.9	4.0	4.5
MBT	0.29	0.25	0.24	0.22	6.6	5.6	5.4	5.5
FBZ	0.31	0.21	0.21	0.18	6.9	4.7	4.6	4.0

DZP	0.22	0.18	0.20	0.19	4.8	4.0	4.5	4.3
FXN	0.27	0.23	0.20	0.20	6.1	5.1	4.4	4.5

[a] RMSE: root mean square error, $\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{I}}\sum_1^{\mathrm{I}}\left(\mathrm{c_{nom}} - \mathrm{c_{pred}}\right)^2}$, where I = 9

[b] REP%: relative error of prediction, $\mathrm{REP\%} = 100 \times \frac{\mathrm{RMSE}}{\bar{\mathrm{c}}}$, where $\bar{\mathrm{c}}$ is the mean calibration concentration

Table 4. Determination of AIs in real samples collected from five commercial farms

		Non-fused	Fused		
Sample	Analyte (μg g^{-1})[a]	MCR-ALS	MCR-ALS	PARAFAC	N-PLS
PL1	ENR[b]	0.81 (0.05)	0.80 (0.01)	0.79 (0.03)	0.77 (0.03)
	IMD	0.54 (0.02)	0.55 (0.01)	0.52 (0.01)	0.533 (0.009)
	FXN	0.63 (0.05)	0.63 (0.03)	0.65 (0.02)	0.64 (0.02)
PL2	TMP	0.79 (0.06)	0.80 (0.02)	0.78 (0.04)	0.79 (0.02)
	ABZ	0.24 (0.01)	0.247 (0.003)	0.251 (0.008)	0.246 (0.004)
PL3	ABZ	0.86 (0.06)	0.83 (0.01)	0.86 (0.04)	0.87 (0.04)
	CAP	0.70 (0.02)	0.71 (0.02)	0.73 (0.04)	0.73 (0.04)
	FBZ	-	0.055 (0.004)	0.059 (0.005)	0.055 (0.005)
PL4	ABZ	0.21 (0.01)	0.213 (0.002)	0.219 (0.007)	0.212 (0.003)
	IMD	0.64 (0.04)	0.66 (0.01)	0.64 (0.04)	0.67 (0.01)
	FLU	0.15 (0.01)	0.156 (0.002)	0.159 (0.006)	0.155 (0.004)
	CAP	0.64 (0.05)	0.64 (0.02)	0.63 (0.03)	0.66 (0.01)
PL5	ENR[b]	1.73 (0.05)	1.730 (0.003)	1.72 (0.02)	1.706 (0.004)
	FLU	1.39 (0.07)	1.34 (0.01)	1.343(0.01)	1.32 (0.01)
	IMD	-	0.114 (0.008)	0.113 (0.007)	0.112 (0.008)

Standard deviations between parentheses

[a] Concentrations below the corresponding LOD are indicated with dashes

[b] The samples were diluted (1/5) in order to reach concentrations between the limits of the calibration curve

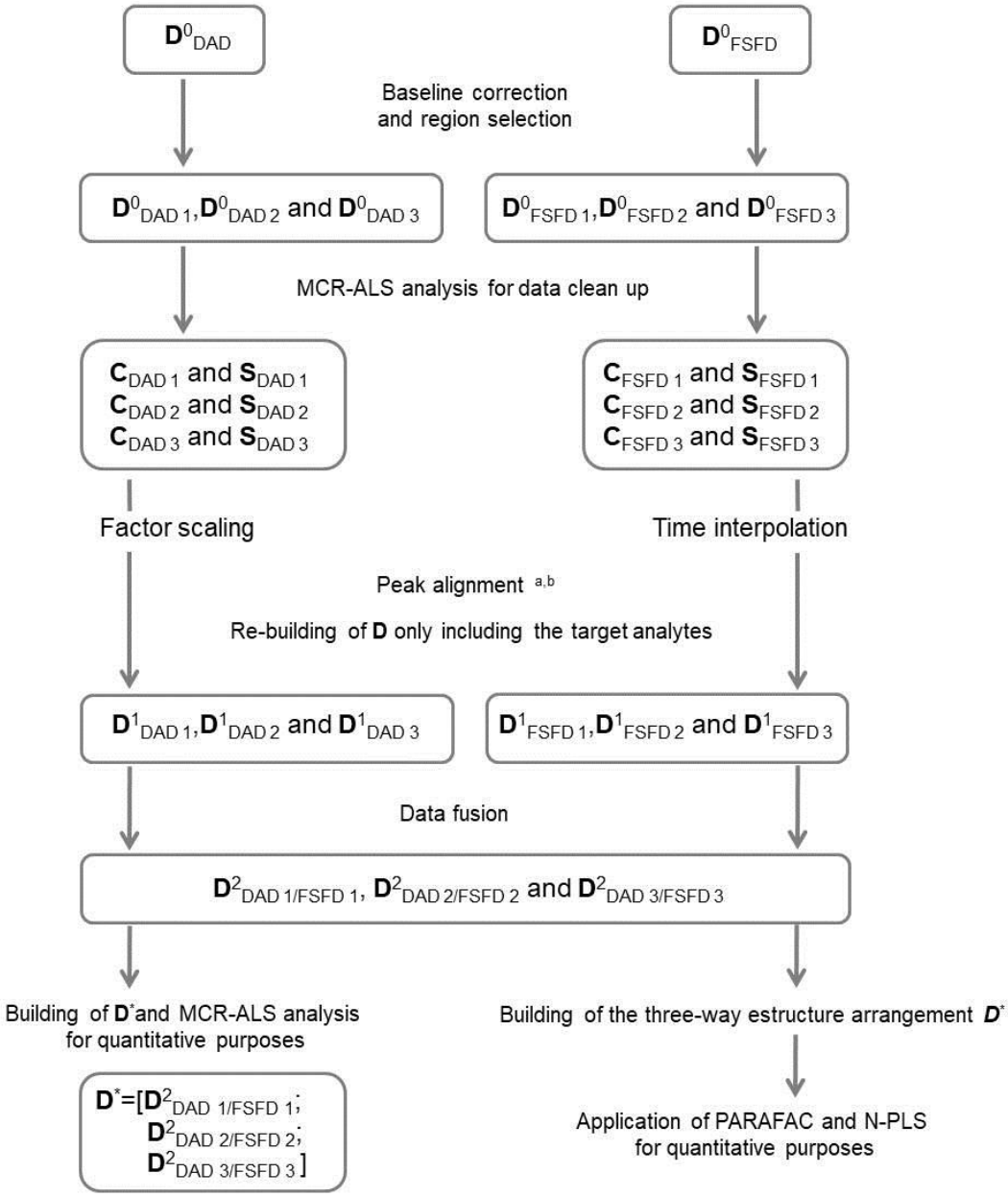
Figure captions:

Figure 1. Data pretreatment carried out for a single sample to reach DF and modeling by chemometric algorithms: baseline correction and region selection, MCR-ALS processing for data clean up, factor scaling and time interpolation, peak correction, rebuilding of **D** only including the target analytes, data fusion, building of $\mathbf{D}^*$ and MCR-ALS for quantitative purposes, building of the three-way structure arrangement $\boldsymbol{D}^*$, application of PARAFAC and N-PLS for quantitative purposes. $\mathbf{D}^0$ indicates the original matrices, while $\mathbf{D}^1$ and $\mathbf{D}^2$ are those obtained after the first and second preprocessing steps, respectively.

Figure 2. Time profiles and fused spectra for the calibration samples having 1, TMP; 2, IMD; 3, ENR; 4, CLB; 5, DIF; 6, CFT; 7, CTC; 8, CAP; 9, PYR; 10, FLU; 11, MBT; 12, ABZ; 13, FXN; 14, DZP; 15, FBZ: (A and B) data obtained after DAD-to-FSFD and run-to-run COW alignment for each analyte in each sample, (C and D) the same data after DAD-to-FSFD and run-to-run COW alignment for each analyte from sample to sample, and (E) fused analyte spectra.

Figure 3. Elliptical joint confidence regions (at 95% confidence level) for slope and intercept of the regression of MCR-ALS non-fused data (green line), MCR-ALS fused data (red line), PARAFAC (blue line), N-PLS (violet line). The black dot marks the theoretical (intercept = 0, slope = 1) point.

Figure 1



[a] Peak alignment of each analyte between C_{DAD} and C_{FSFD} before MCR-ALS analysis
[b] Peak alignment of each analyte from sample to sample before PARAFAC and N-PLS analysis

Figure 2

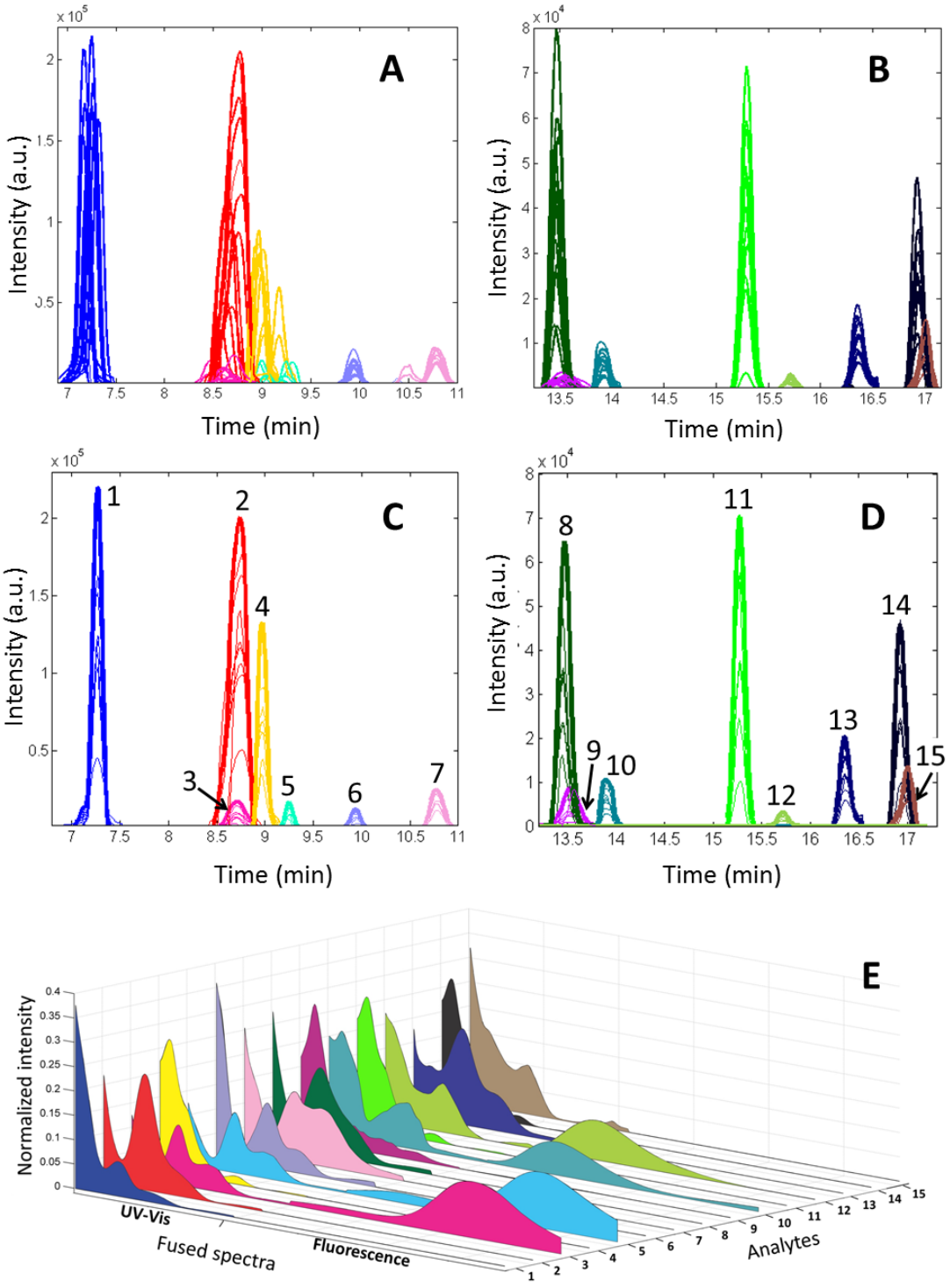
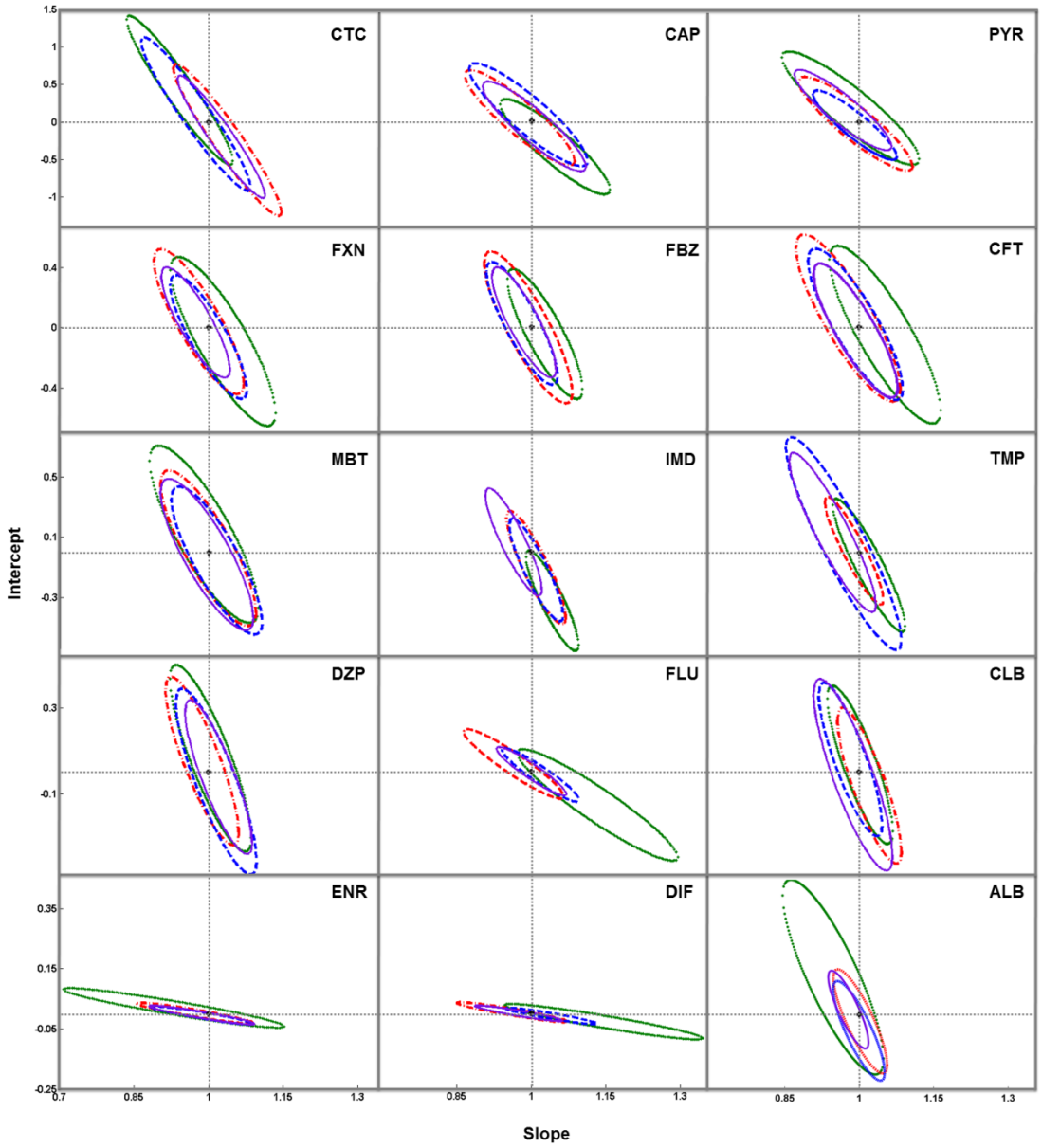

A
B
C
D
E
Intensity (a.u.)
Time (min)
x 10^5
x 10^4
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
Normalized intensity
UV-Vis
Fused spectra
Fluorescence
Analytes

Figure 3



Highlights

- Low-level data fusion strategy for second-order liquid chromatography data processing.
- Assessment of the synergistic effect of coupling the individual information provided by two different detectors.
- Chemometric data pretreatment and quantitative modeling by MCR-ALS, PARAFAC and N-PLS.
- Simultaneous analysis of fluorescent and non-fluorescent veterinary active ingredients that can be found in poultry litter.
- Application to real poultry litter samples containing the studied compounds.