Journal Pre-proof

A new approach based on inversion of a partial least squares model searching for a preset analytical target profile. Application to the determination of five bisphenols by liquid chromatography with diode array detector

M.M. Arce, S. Ruiz, S. Sanllorente, M.C. Ortiz, L.A. Sarabia, M.S. Sánchez

PII: S0003-2670(21)00043-X

DOI: https://doi.org/10.1016/j.aca.2021.338217

Reference: ACA 338217

To appear in: *Analytica Chimica Acta*

Received Date: 28 October 2020

Revised Date: 8 January 2021

Accepted Date: 11 January 2021

Please cite this article as: M.M. Arce, S. Ruiz, S. Sanllorente, M.C. Ortiz, L.A. Sarabia, M.S. Sánchez, A new approach based on inversion of a partial least squares model searching for a preset analytical target profile. Application to the determination of five bisphenols by liquid chromatography with diode array detector, *Analytica Chimica Acta*, https://doi.org/10.1016/j.aca.2021.338217.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2021 Elsevier B.V. All rights reserved.

Journal Pre-proof

CREDIT AUTHOR STATEMENT:

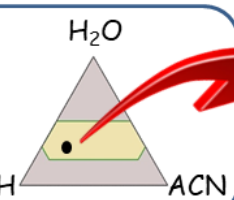
M.M. Arce:	Formal analysis, Investigation, Methodology; Writing - original draft, Writing - review & editing.
S. Ruiz:	Investigation, Methodology, Writing - original draft, Writing - review & editing.
M.C. Ortiz:	Conceptualization, Funding acquisition, Methodology, Supervision, Writing - original draft, Writing - review & editing.
S. Sanllorente:	Formal analysis, Investigation, Methodology, Supervision, Writing - review & editing.
L.A. Sarabia:	Conceptualization, Methodology, Supervision Writing - original draft, Writing - review & editing.
M.S. Sánchez:	Investigation, Software; Methodology, Supervision, Writing - original draft, Writing - review & editing.

Journal Pre-proof

AQbD: PLS2 COMPUTATIONAL INVERSION

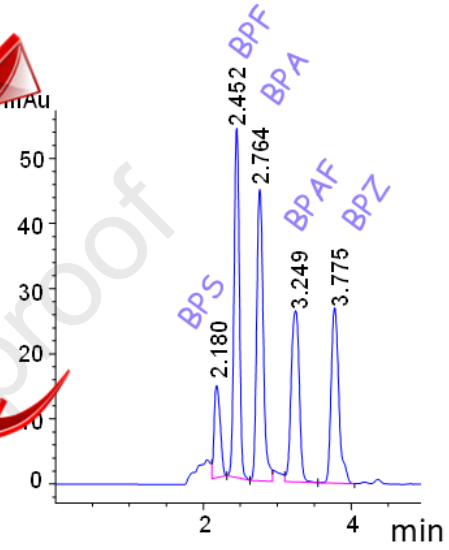
Control Method Parameters

22% water
58% methanol
20% acetonitrile
Flow rate 0.66 mL min^{-1}



Analytical Target Profile

Resolution between contiguous peaks:
R_{12}=1.88
R_{23}=2.12
R_{34}=2.64
R_{45}=2.67
Initial and final time: t_i=2.10, t_f=4.06



Journal Pre-proof

Journal Pre-proof

A NEW APPROACH BASED ON INVERSION OF A PARTIAL LEAST SQUARES MODEL SEARCHING FOR A PRESET ANALYTICAL TARGET PROFILE. APPLICATION TO THE DETERMINATION OF FIVE BISPHENOLS BY LIQUID CHROMATOGRAPHY WITH DIODE ARRAY DETECTOR

M.M. Arce[1], S. Ruiz[2], S. Sanllorente[1],M.C. Ortiz[1*], L.A. Sarabia[2], M.S. Sánchez[2]

[1]Dpt. Chemistry, Faculty of Sciences, Universidad de Burgos

[2]Dpt. Mathematics and Computation, Faculty of Sciences, Universidad de Burgos, Plaza Misael Bañuelos s/n, 09001 Burgos, Spain

Journal Pre-proof

Abstract

The paper shows a procedure for selecting the control method parameters (factors) to obtain a preset 'analytical target profile' when a liquid chromatographic technique is going to be carried out for the simultaneous determination of five bisphenols (bisphenol-A, bisphenol-S, bisphenol-F, bisphenol-Z and bisphenol-AF), some of them regulated by the European Union.

The procedure has three steps. The first consists of building a D-optimal combined design (mixture-process design) for the control method parameters, which are the composition of the ternary mobile phase and its flow rate. The second step is to fit a PLS2 model to predict six analytical responses (namely, the resolution between each pair of consecutive peaks, and the initial and final chromatographic time) as a function of the control method parameters. The third final step is the inversion of the PLS2 model to obtain the conditions needed for attaining a preset analytical target profile.

The computational inversion of the PLS2 prediction model looking for the Pareto front of these six responses provides a set of experimental conditions to conduct the chromatographic determination, specifically 22% of water, mixed with 58% methanol and 20% of acetonitrile, keeping the flow rate at 0.66 mL min^{-1}. These conditions give a

* Corresponding author. E-mail address: mcortiz@ubu.es

chromatogram with retention times of 2.180, 2.452, 2.764, 3.249 and 3.775 minutes for BPS, BPF, BPA, BPAF and BPZ, respectively, and excellent resolution among all the chromatographic peaks.

Finally, the analytical method is validated under the selected experimental conditions, in terms of trueness and precision. In addition, the detection capability for the five bisphenols were: 596, 334, 424, 458 and 1156 µg L^{-1}, with probabilities of false positive and of false negative equal to 0.05.

Keywords: Process Analytical Technology; Partial Least Squares; Pareto optimality; bisphenol A; HPLC-DAD, computational Latent Variable Model Inversion.

1. Introduction

The Process Analytical Technology (PAT) is a system for designing, analyzing, and controlling manufacturing through measurements of raw and in-process materials and processes, environmental properties, etc. (that constitute matrix **X**), with the goal of ensuring final product quality established by some values of the critical quality attributes (CQA) in matrix **Y**. Currently, PAT is a standard for the accreditation of pharmaceutical processes [1,2].

In the present work, the relation between **X** and **Y** is established by means of a latent variable regression model, precisely a PLS2 ('projection to latent space' o 'partial least squares' with more than one response) model that is obtained by computing '*a'* new set of orthogonal variables such that,

$$\mathbf{X} = \mathbf{T}_a \mathbf{P}_a^T + \mathbf{R}_X = \hat{\mathbf{X}} + \mathbf{R}_X \quad (1)$$

$$\mathbf{Y} = \mathbf{T}_a \mathbf{Q}_a^T + \mathbf{R}_Y = \mathbf{X}\mathbf{M} + \mathbf{R}_Y \quad (2)$$

Journal Pre-proof

In what follows, it is assumed that **X** (input variables) is representative of the 'process space', and **Y** (responses) of the 'quality space' of the product characteristics.

Like in many other data-driven situations, the intended CQA are known and can be defined as a vector $\mathbf{y}_{des}$, not necessarily in the training set. Whether this $\mathbf{y}_{des}$ can be attained (the product with these characteristics is feasible) depends on the existence of values for the process variables that would give this target product. Therefore, once the PLS2 model is built, its inversion is necessary to obtain the values of the input variables, $\mathbf{x}_{des}$, with which a preset quality $\mathbf{y}_{des}$ is obtained [1]. The study of the viability of this inversion, with the necessary constraints, to guarantee the existence of a solution is known as Latent Variable Model Inversion (LVMI).

In the literature, there are two alternatives to approach the LVMI. One of them is related to the inversion of the matrices in the decomposition (such as the ones in Eqs. (1) and (2)). A summary of this approach can be found in Refs. [3, 4]. Ref. [5] contains the conditions and necessary constraints, depending on the dimension of the spaces spanned by the latent, process and quality variables. In this scenario, no constraints can be directly imposed on the quality characteristics.

This is why the second alternative is to redefine the inversion and, for example, look for the scores that minimize the weighted squared difference between the predicted and desired characteristics, with a limit to the maximum value allowable for the Hotelling's T^2 statistics [6]. In this way, the problem is posed as a weighted, and possibly constrained, least squares problem, which is also applicable if the type of model is changed, by a non-linear PLS in Ref. [3], or with genetic programming in Ref. [7].

In both alternatives, the inversion 'goes' from the space **Y** to the subspace $\hat{\mathbf{X}}$ in **X,** Eq. (1), reconstructed from the space spanned by the latent variables. Consequently, it is implicitly assumed that the subspace $\hat{\mathbf{X}}$ somehow collects the whole correlation between

Y and **X**, and also that the residuals subspaces $\mathbf{R}_X$ and $\mathbf{R}_Y$ are uncorrelated to **X** and **Y**, respectively, which is not true, in general. Some attempts to tackle this problem include pretreatment of predictors, giving rise to the family of Orthogonal Signal Correction (OSC) procedures that is explained in Ref. [8], or modifications of the PLS method [9] without yet having an acceptable solution to obtain a subspace of **X** whose orthogonal complement is independent of **Y**.

The present work uses a third alternative with a computational approach for the model inversion [10]. The methodology relies on the properties of the Pareto optimal front computed when simultaneously minimizing the expected differences between the predicted quality characteristics (model predictions) and the target ones in $\mathbf{y}_{des}$. The advantage of this new approach is that it allows addressing the problem of the feasibility of a product with given characteristics from the multidimensional point of view, considering each quality characteristic but studying their joint behavior so as not to lose the correlation structure. In addition, the solution $\mathbf{x}_{des}$ for a target quality $\mathbf{y}_{des}$ is obtained in the space of the raw variables and not in the latent space, even though the relation between predictors and responses (process space and quality space) has been built using latent variables models.

Furthermore, the PLS2 model being inverted can include not only the input (measured) variables related to the process but also their cross-products (interactions) or any other possible transformations necessary to model non-linearities. In this case, the matrix with the predictor variables to fit the PLS2 model will contain the additional columns corresponding to these new ‘variables’ obtained from the original ones. Consequently, the direct inversion of the matrices is not possible, as shown in Ref. [11], but the solution obtained with the computational approach, $\mathbf{x}_{des}$, is still available and is always given in terms of the original variables.

The properties just summarized can be directly transferred to the analytical laboratory, where the 'process or input variables' are those that affect sample preparation or instrumental factors whereas the quality characteristics are related to analytical signals or figures of merit of a fit for purpose analytical procedure. In this context, the term Analytical Quality by Design (AQbD) has been coined referring to the version of the Quality by Design (QbD) concept applied to the development of an analytical method. The AQbD has aroused considerable interest, remarkable reviews on the subjects are Refs. [12,13]. It also appears in systematic chromatography reviews, for example, 47 citations out of the 158 references in Ref. [14] are about AQbD.

The main idea in AQbD is to develop the analytical method such that the intended quality is achieved. This intended quality is defined as an 'analytical target profile' (ATP), depending on the so-called control method parameters (CMP). In analogy with QbD, a model is built to predict the ATP from the CMP. The novelty in the present paper is that, then, the inversion of the fitted PLS2 model will give the particular values of the CMP, $\mathbf{x}_{des}$ with the notation in the previous paragraphs, to obtain a preset ATP, i.e., a predefined $\mathbf{y}_{des}$.

A bibliographical revision on the subject gives some papers [15,16] where this AQbD approach is used to optimize figures of merit of analytical procedures, although the authors have not found any references that include a PLS model inversion for the task.

The procedure proposed in the present work to guarantee an ATP by inverting a PLS2 model is a general methodological approach, though it is explained/applied to the determination of five bisphenols by means of liquid chromatography coupled with diode array detector (HPLC-DAD). The mobile phase composition from the ternary mixture (percentages of water, methanol and acetonitrile) together with its flow rate, constitute the CMP taken into account. The ATP, on the other hand, is related to the analytical characteristics of the resulting chromatogram. It is defined by means of six responses: four values of the resolution between consecutive chromatographic peaks (R_{12}, R_{23}, R_{34},

R_{45}), the initial time (t_i) to avoid the dead volume and the final time (t_f) related with the total runtime of the chromatogram.

A PLS2 model is fitted to simultaneously predict the six characteristics that define the ATP as function of the four CMP. PLS2 captures not only the correlation among the CMP of the analytical procedure but also the correlation among the ATP-related characteristics of the obtained chromatogram, modelling the predictive relation between CMP and ATP as well. Moreover, the use of the statistics Q and T^2 allows the 'control' of the process in the sense that the only points for which the PLS2 model is applicable are those with values of these statistics bellow some preset thresholds.

Besides being an application of the general methodology proposed, there are some other relevant reasons for the determination of these bisphenols. The endocrine disruptor nature of BPA, the social alarm generated by its use, the progressive substitution by other bisphenols in the manufacture of polycarbonate products for daily use, and the attention given by the regulatory bodies, have generated the need for methods for the simultaneous determination of these analytes, methods that need to be fast and efficient.

In 1999, The European Union approved a Community Strategy for Endocrine Disrupters [17], that included periodic reports from the European Commission on the progress made in the fields of research, international co-operation, communication to the public, and appropriate policy action.

The first report in 2001 [18] contains a provisional list of substances that already included BPA as substance with evidence of endocrine disruptor (potential or effective) (Category 1) and BPF as substance with insufficient data. Later, BPF was classified as substance with no or insufficient data on endocrine disruptor effects (Category 3a and 3b) [19].

Currently, there are regulations regarding endocrine disruptors in pesticides, biocides, sanitary products or water. In particular, BPA is subjected to recent specific regulations in toys [20], food contact materials [21], thermal paper (for example, tickets) [22], and cosmetics [23]. In fact, in 2019 the European Commission [24] required the prohibition of bisphenols in all materials in contact with food, expanding the ban that existed since 2011 for BPA on polycarbonate baby bottles [21]. Besides, it specifies that BPA is frequently replaced by other bisphenols, and suggests that substances with a similar chemical structure should be assumed to have toxicological properties just as harmful as those of the most toxic known substance of the group. There are several studies on the detrimental effects of compounds that are replacing BPA [25,26,27,28,29]. Alternatives include BPS, BPF, BPZ and BPAF, reason why these are the bisphenols selected in the present work.

In another bibliographic revision, summarized in Section 4.5, the separation by HPLC-DAD of these five bisphenols is generally done in gradient elution mode and with binary mixtures in the mobile phase, whether MeOH/H_2O or ACN/H_2O. In the present paper, a ternary mixture with both organic solvents is explored, whose composition, together with the flow rate of the mobile phase, constitute the selected CMP.

The paper is organized as follows. The analytical materials and methods are described in Section 2. Section 3 details the selection of the followed experimental design to obtain the training set for fitting the PLS2 model, and its computational inversion that gives the CMP that guarantee the ATP defined by the analyst in the determination of the five bisphenols. Results, validation of the method and its comparison with data already published in the literature are shown in Section 4.

Journal Pre-proof

Journal Pre-proof

2. Material and methods

2.1. Chemicals and reagents

Bisphenol-A (CAS no. 80-05-7), bis(4-hydroxyphenyl)methane (CAS no. 620-92-8) and 4,4'-cyclohexylidenebisphenol (CAS no. 843-55-0) were acquired in Sigma-Aldrich (Steinheim, Germany). 4,4'-(hexafluoroisopropylidene)diphenol (CAS no. 1478-61-1) was purchased by Alfa Aesar (Kandel, Germany). Bis(4-hydroxyphenyl)sulfone (CAS no. 80-09-1), acetonitrile (CAS no. 75-05-8; LiChrosolv® isocratic grade for liquid chromatography) and methanol (CAS no. 67-56-1; LiChrosolv® isocratic grade for liquid chromatography) were supplied by Merck (Darmstadt, Germany). Deionized water was obtained by using the Milli-Q gradient A10 water purification system from Millipore (Bedford, MA, USA).

2.2. Instrumental

Determination of the five bisphenols, BPA, BPS, BPF, BPZ and BPAF, was carried out using an Agilent 1260 Infinity HPLC chromatograph (Santa Clara, CA, USA) consisting of a quaternary pump (G1311C), a sampler (G1329B), a thermostatic column compartment (G1316 A), and a diode array detector (G7117C). A Kinetex EVO-C18 column (150 mm × 4.6 mm, 5 µm) was used for the separation. Deionized water (solvent A), methanol (solvent B) and acetonitrile (solvent C) were used as mobile phases.

The conditions for chromatographic analyses were programmed in isocratic mode. Mobile phase consists of different percentages of a mixture of water/methanol/acetonitrile (Z_1:Z_2:Z_3, v/v) and different mobile phase flow rate (U_4 mL min^{-1}), depending on the conditions in the experimental design followed, which is explained in Section 3.1. In all analyses, the temperature of the column compartment was 20°C and the injection volume was 10 µL. Diode array detector was programmed to measure the absorbance at a fixed wavelength of 225 nm.

2.3. Standard solutions and samples

Individual standard stock solutions of 500 mg L^{-1} were prepared by dissolving each standard in methanol. A mixture of 4 mg L^{-1} of each bisphenol was prepared from the individual stock solutions by dilution with methanol for the experiments carried out according to the D-optimal design (Section 3.1). Calibration standard solutions were prepared from the individual stock solutions by dilution with methanol, mixtures of 1 to 5 mg L^{-1} of each bisphenol (see Table 1), to build calibration lines. All solutions were stored protected from light at 4°C.

Table 1. Concentration levels (mg L^{-1}) of the standard mixtures used in the building of the calibration lines for the five bisphenols.

Code sample	BPA	BPS	BPF	BPZ	BPAF
Standard 1	1	2	3	4	5
Standard 2	5	1	2	3	4
Standard 3	4	5	1	2	3
Standard 4	3	4	5	1	2
Standard 5	2	3	4	5	1

2.4. Software

OpenLab CDS ChemStation software was used for acquiring data. The PLS2 models were fitted with the PLS_Toolbox [30]. The inversion of the PLS2 model and the Pareto optimal front were calculated with in-house programs written in MATLAB [31]. The regression models were fitted and validated using STATGRAPHICS Centurion 18 [32]. The experimental design is selected with NEMRODW [33]. The capability of detection (CCβ) was calculated using the DETARCHI program [34].

Journal Pre-proof

3. General procedure

Given an ATP, the whole procedure to obtain the corresponding CMP has the following steps:

1. To choose the appropriate experimental design for obtaining the training set for the PLS2 model. It includes setting the factors (CMP), their variation to define the experimental domain, a combined and reduced experimental design, and the precise definition of the responses of interest, that is, the definition of the ATP.
2. Selection of optimal experimental conditions. It includes the fitting and inversion of a prediction regression model, PLS2 model in this case, obtaining and exploring the Pareto front for the desired ATP and the final selection of a single set of CMP (the experimental conditions) to perform the chromatographic determination.
3. Validation of the analytical method with the selected CMP. It includes the experimental validation of the found conditions as well as the figures of merit (accuracy, decision limit and capability of detection) of the proposed analytical procedure to determine the five bisphenols.

Journal Pre-proof

3.1 Design of the experiments

There are four CMP (the ternary composition of the mobile phase and its flow rate), that can be varied and whose variation changes the resulting chromatogram.

Three of them define the composition of the mobile phase, namely proportions of water (Z_1), methanol (Z_2) and acetonitrile (Z_3), with some constraints: the composition of water in the mixture should be between 20% and 50%, and methanol and acetonitrile cannot exceed 70% of the mixture. The proportions used are selected following a mixture design in a restricted simplex.

The fourth factor, the flow rate of the mobile phase, is a continuous factor that varies between 0.6 and 1.0 mL min^{-1}. Table 2 summarizes the conditions for the four factors and their constraints. From a Design of Experiments (DOE) point of view, (Z_1, Z_2, Z_3) constitute the components of a mixture (varying on the restricted simplex) and factor U_4 is a continuous factor which is codified into [-1, 1]. As usual, X_4 refers to the coded variable and U_4 to the raw factor.

Table 2. CMP (experimental factors, or process variables) and their variation.

Factor	Lower bound	Upper bound	Centre	Step of variation
Z_1 Water	0.20	0.50		
Z_2 Methanol	0	0.70		
Z_3 Acetonitrile	0	0.70		
U_4 Flow rate (mL min^{-1})			0.80	0.20

In order to have a 'representative' training set that adequately covers the experimental domain, the experiments conducted followed an experimental design. As there are proportions of a mixture and a continuous factor, the design is a combined design (with mixture and process variables) in the domain defined in Table 2 and depicted in Fig. 1, which consists of the restricted simplex (for the mixtures) extended along the three levels considered for the flow rate.

Finally, the assumed model for each individual Y in the multiplicative mixture process design is quadratic in the continuous variable (flow rate), according to Eq. (3).

$$Y = \gamma_0 + \gamma_4 X_4 + \gamma_{44} X_4^2 \quad (3)$$

where γ_i represents also a quadratic dependence on the mixture composition (Z_1, Z_2, Z_3). For example, γ_4 in Eq. (3) means:

$$\gamma_4 = \beta_{41}Z_1 + \beta_{42}Z_2 + \beta_{43}Z_3 + \beta_{412}Z_1Z_2 + \beta_{413}Z_1Z_3 + \beta_{423}Z_2Z_3 \quad (4)$$

By substituting all the terms in Eq. (3), the model has 18 coefficients per each level of X_4.

$$\begin{aligned} Y = {} & \beta_1Z_1 + \beta_2Z_2 + \beta_3Z_3 + \beta_{12}Z_1Z_2 + \beta_{13}Z_1Z_3 + \beta_{23}Z_2Z_3 + \\ & \beta_{41}X_4Z_1 + \beta_{42}X_4Z_2 + \beta_{43}X_4Z_3 + \beta_{412}X_4Z_1Z_2 + \beta_{413}X_4Z_1Z_3 + \beta_{423}X_4Z_2Z_3 + \\ & \beta_{441}X_4^2Z_1 + \beta_{442}X_4^2Z_2 + \beta_{443}X_4^2Z_3 + \beta_{4412}X_4^2Z_1Z_2 + \beta_{4413}X_4^2Z_1Z_3 + \beta_{4423}X_4^2Z_2Z_3 \end{aligned} \quad (5)$$

The complete experimental design would require 19 experiments per level of X_4, which would imply the need of carrying out 57 chromatograms because the flow rate is at three levels, too many runs. In order to reduce the quantity of solvents (also reducing the impact in the environment and the time of analysis), a subset of experiments was selected, but without losing quality of the information extracted from the experiments. Appling the D-optimal criterion [35], nine out of the 19 experiments in each level of the flow rate were selected, maintaining good reliability properties because the maximum of the variance function in the experimental domain was 0.8. These 27 experiments are the ones in black in Fig. 1.

It has already been stressed the importance of exploring the effect of ternary mixtures in the mobile phase. With that purpose, the ten experiments discarded for the D-optimal design (in each level of flow rate) were used. Eight of them contain a ternary mobile phase and the other two consist of binary mixtures at some side midpoint of the reduced simplex. These ten experiments were distributed among the three levels of flow rate so that, in each level, there were at least two ternary mixtures. The resulting distribution is depicted with the red points in Fig. 1.

Finally, two additional replicates in the simplex centroid for each flow rate were included. Therefore, matrix **X** has 43 rows (27+10+6 chromatographic runs) and 4 columns corresponding to their CMP values.

<here Fig. 1>

After performing each chromatogram, the six ATP-related characteristics were computed, namely the resolution between consecutive peaks, R_{12}, R_{23}, R_{34} and R_{45}, the initial time, t_i, and the total (final) time, t_f, both in minutes. The resolution $R_{i,i+1}$ between the consecutive i-*th* and (i+1)-*th* chromatographic peaks is computed by means of Eq. 6.

$$R_{i,i+1} = \frac{2.35\left(t_{R,i+1} - t_{R,i}\right)}{2\left(w_{0.5,i+1} + w_{0.5,i}\right)} \tag{6}$$

where $t_{R,i}$ is the retention time and $w_{0.5,i}$ is the width at half height of the i-*th* chromatographic peak.

3.2 Selection of experimental conditions

With the notation stablished in the introduction, **X** is the matrix containing the values of CMP corresponding to the 43 experimental conditions selected with the experimental design, while **Y** is the matrix containing the six ATP-related values computed from the corresponding chromatograms. Assuming that **X** is noise free, the customary procedure in DOE is to fit separate multilinear regression (MLR) models for each response. Although the factors in the present experiment are controlled, and it is assumable at least that the variability in **X** is negligible compared to the one in **Y**, the correlation among responses is expected to be high. Therefore, to handle this correlation while fitting the six responses at once, a single PLS2 model will be built.

However, the intended PLS2 model contains more than the linear terms related to the columns already in **X**, that is, it does not depend exclusively on variables Z_1, Z_2, Z_3 and X_4 (main factors, the ones that can be modified inside the experimental domain) but also on several different cross-terms. This is so because interactions among factors are expected, not only among the mixture variables but also between the composition of the mixture and the flow rate in the mobile phase, further to possible quadratic effects among all the factors in the design. These interactions and non-linear effects will be

reflected in the model to be fitted with PLS2, in the present case, the model defined in Eq. (7), with sixteen coefficients.

$$\begin{aligned} Y \;=\; & \beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3 + \beta_4 X_4 + \\ & \beta_{12} Z_1 Z_2 + \beta_{13} Z_1 Z_3 + \beta_{23} Z_2 Z_3 + \beta_{41} X_4 Z_1 + \beta_{42} X_4 Z_2 + \beta_{43} X_4 Z_3 + \\ & \beta_{412} X_4 Z_1 Z_2 + \beta_{413} X_4 Z_1 Z_3 + \beta_{423} X_4 Z_2 Z_3 + \\ & \beta_{441} X_4^2 Z_1 + \beta_{442} X_4^2 Z_2 + \beta_{443} X_4^2 Z_3 \end{aligned} \tag{7}$$

This means that matrix **X** with the CMP is not the actual data matrix used to fit the model. On the contrary, each row $\mathbf{x}^T$ must be 'expanded' into a new vector $\mathbf{x}_E$ (in higher dimension) following the terms of the model to be fitted, similar to the construction of the model matrix from the design matrix. Hence, the new expanded vector $\mathbf{x}_E$ has 16 coordinates as expressed in Eq. (8).

$$\begin{aligned} \mathbf{x}^T &= \left(z_1, z_2, z_3, x_4\right) \\ \mathbf{x}_E^T &= \left(z_1, z_2, z_3, x_4, z_1 z_2, z_1 z_3, z_2 z_3, x_4 z_1, x_4 z_2, x_4 z_3, x_4 z_1 z_2, x_4 z_1 z_3, x_4 z_2 z_3, x_4^2 z_1, x_4^2 z_2, x_4^2 z_3\right) \end{aligned} \tag{8}$$

To avoid misunderstanding, the training set for fitting the PLS2 model will be denoted as $\mathbf{X}_E$ reflecting the fact that **X** is enlarged to $\mathbf{X}_E$, because the model is in fact built with the 16-dimensional vectors in $\mathbf{X}_E$, which are projected onto a latent space of lower dimension to predict **Y**.

Let *L* denote the model fitted from the data **X**-**Y** as a function of **u**, so that $\hat{\mathbf{y}} = L\left(\mathbf{u}\right)$ is the predicted response (a six-dimensional vector) for a four-dimensional **u** in the experimental domain in Fig. 1.

For applying *L*, each **u** is codified to obtain **x** and expand onto the corresponding 16-dimensional vector $\mathbf{x}_E$ as in Eq. (8). Then, it is scaled according to the pretreatment used when fitting the PLS2 model. As it has been already said, those resulting vectors $\mathbf{x}_S$ with

values of Q or T^2 statistics greater than the corresponding 95% confidence limits are discarded.

In the following, a *valid* point will be any four-dimensional $\mathbf{u} = (z_1, z_2, z_3, u_4)$ for which the resulting $\mathbf{x}_S$ complies with all the constraints to apply the model. The domain is in four dimensions, the model is fitted in 16 dimensions, consequently, the common inversion method for latent variable models (see for instance Refs. [3,4,5,6,36]) cannot be applied.

There is, though, an alternative [10] that, driven by an evolutionary algorithm, works by moving *valid* points inside the experimental domain (the four-dimensional experimental domain of the CMP) so that the predicted values of the responses be closer and closer to $\mathbf{y}_{des}$, the ideal vector with the desired characteristics for the chromatogram.

With this approach, the inversion of the PLS2 model is tackled as a multiobjective or multiresponse optimization problem. The most distinctive characteristic between optimizing a single response or multiple objectives at once is that, usually, in the latter case there is not a single solution that comply with all the requisites simultaneously.

A compromise among all the responses can be defined in the form of Pareto-optimal solutions, that is, those that cannot be improved in one response without worsening another. These solutions constitute the so-called Pareto front, which is reduced to a single point if there were no conflict among responses.

In a pseudo-code, looking for the Pareto front for the six responses consists of the following steps:

a) Start with a population of *ps* valid points (four-dimensional **u** vectors complying with the constraints).
b) Compute $\hat{\mathbf{y}} = L(\mathbf{u})$.
c) Compute the fitness function, defined as the six-dimensional vector of the absolute value of the individual differences between the predicted and the target values.

d) Apply selection, crossover and mutation operators to build new *ps* valid CMP, which are also evaluated in terms of the fitness function described in c).
e) Merge the old and newly generated populations.
f) Arrange the members of the extended population according to the Pareto order for multidimensional vectors [37] and select to survive for the next generation the non-dominated solutions. If there are more than *ps*, select the most dispersal along the front according to the crowding distance [38].
g) Repeat a)-f) for a given number of generations.

The final population is made up of different settings for the CMP expected to provide a chromatogram whose characteristics are the closest to the desired ones in at least one of the responses and, more interestingly, the estimate of the Pareto front describes the trade-off among the six responses that are being handled. As such, it is used to decide about the needed experimental conditions to perform the determination of the bisphenols.

The code for computation of the Pareto front, in a different context, can be found in the annex of ref. [39].

Journal Pre-proof

4. Results and discussion

4.1. Experimental data

Once conducted the 43 experiments, matrix **Y** collects the values of the responses computed from the obtained chromatograms. As we have anticipated, they are highly correlated, as can be seen in the pairwise correlation coefficients in Table 3. The first three responses are highly positively correlated but highly negatively correlated with R_{45}. This resolution also behaves oppositely when comparing with the time, both the initial and final, which in turn are positively correlated to each other.

Table 3. Correlation matrix of ATP variables, **Y**.

	R_{12}	R_{23}	R_{34}	R_{45}	t_i
R_{23}	0.9607				
R_{34}	0.9497	0.9874			
R_{45}	-0.6376	-0.7013	-0.7605		
t_i	0.5152	0.4919	0.5050	-0.3764	
t_f	0.8400	0.8117	0.8456	-0.7173	0.5765

It is not in the table, but there is also correlation among the CMP, even though the distribution of their values has been designed. As expected, the lowest correlation coefficients are between the flow rate X_4 and the proportions of the ternary mixture of the mobile phase (Z_1, Z_2, Z_3), but there is a high correlation (-0.899) between Z_2 and Z_3.

Furthermore, if the actual sixteen variables that are used to fit the PLS2 model (columns of $\mathbf{X}_E$) are taken into account, there are 120 correlation coefficients, 59 of which are significantly non-null (5% significance level) and 40, in fact, are greater than 0.8 in absolute value.

To take into account all these correlations, expected in the context of AQbD, a PLS2 model is fitted. Another reason is that the process control is exerted with the statistics Q and T^2, whose limits at 95% confidence level define the boundary of the region where the feasible solutions lie (provide that they belong to the experimental domain). In this way, only one set of limits are used instead of six different sets of threshold values that would be needed if a PLS1 model was fitted to each individual response.

4.2. PLS2 model

A PLS2 model is thus fitted with autoscaled predictor variables in $\mathbf{X_E}$ (43 × 16) and responses in **Y** (43 × 6). Responses R_{45} and t_f had to be transformed (monotonic transformation) to better fit them. In this case, $Y_4 = 2^{R_{45}}$ and $Y_6 = \log_{10}(t_f)$. The reason is

that the range of R_{45} is 2.98, very short compared to the other four resolutions whose range varies between 10 and 27.49. The transformation $2^{R_{45}}$ increased the range of this resolution, helping the joint fit with an explained variance greater than 90%. The opposite occurs with t_f, whose range is 73.70, so a logarithmic transformation was used.

With these six responses, seven latent variables were selected, with the characteristics in Table 4. The seven latent variables greatly explained the training matrices, 99.53% of the covariance in $\mathbf{X}_E$ with 96.97% of the total variance in **Y**.

Table 4. Variance captured in predictor $\mathbf{X}_E$ and response **Y** variables when adding latent variables in the PLS2 model.

Number	Variance captured in $\mathbf{X_E}$ (%)	Cumulative variance captured in $\mathbf{X_E}$ (%)	Variance captured in **Y** (%)	Cumulative variance captured in **Y** (%)
1	41.91	41.91	43.93	43.93
2	30.72	72.63	17.17	61.10
3	17.21	89.84	12.92	74.02
4	7.30	97.14	7.09	81.11
5	1.41	98.55	3.96	85.07
6	0.35	98.90	9.09	94.16
7	0.63	99.53	2.81	96.97

The six responses are more or less equally well fitted, except may be Y_4, as can be seen in Table 5 that contains the coefficient of determination and coefficient of determination in prediction, estimated with 10-fold cross-validation. The similarity of the explained variance in fitting and prediction points to models highly predictive, except again for Y_4 for which larger differences are observed.

Table 5. Coefficient of determination and coefficient of determination in prediction (estimated by cross-validation) for the six responses fitted with PLS2.

	$Y_1 = R_{12}$	$Y_2 = R_{23}$	$Y_3 = R_{34}$	$Y_4 = 2^{R_{45}}$	$Y_5 = t_i$	$Y_6 = \log_{10}(t_f)$
R^2 (%)	98.04	98.75	98.56	91.93	95.50	99.00

R^2_{pred} (%)	96.73	97.41	96.21	81.60	92.26	97.17

In addition, permutation tests were conducted to validate the seven latent variables model. Broadly speaking, they consist of fitting PLS2 models to random permutations of the true response values. If there are not significant differences between the model predicting the 'true' **Y** and the ones predicting 'any other' permuted values for the same $\mathbf{X}_E$, it would be an indication of the inadequacy of the PLS2 model originally fitted. For the six responses at hand, with 50 iterations, the probability of model significance as against the significance with the permuted samples is less than $5 \cdot 10^{-3}$.

4.3. PLS2 inversion to obtain the CMP for a desired ATP

After validation, we have a PLS2 model with seven latent variables, depending on sixteen input variables and predicting six responses. This is the model used to explore the availability of experimental conditions so that the obtained chromatogram has resolution between consecutive peaks of 1.1, with initial time of 2 minutes and final time of 4 minutes. Given the responses that needed to be transformed, the target vector (ATP) is:

$$\mathbf{y}_{des} = \left(1.1,\ 1.1,\ 1.1,\ 2^{1.1},\ 2.0,\ \log_{10}(4.0)\right)^T \quad (9)$$

Consequently, the fitness function to be minimized is:

$$\text{fitness}(\mathbf{u}) = \left(\left|\hat{y}_1 - 1.1\right|,\ \left|\hat{y}_2 - 1.1\right|,\ \left|\hat{y}_3 - 1.1\right|,\ \left|\hat{y}_4 - 2^{1.1}\right|,\ \left|\hat{y}_5 - 2.0\right|,\ \left|\hat{y}_6 - \log_{10}(4.0)\right|\right)^T \quad (10)$$

If one of such chromatograms existed, the fitness function of the corresponding experimental conditions would be the null vector. This is not the case here, as we have already anticipated due to the high correlations in **Y** (Table 3).

With *ps* = 50, uniform selection of individuals, uniform single point cross-over and probability of mutation of 0.1, the population evolves for 500 generations. From the final

population, 45 vectors constitute the estimate of the Pareto front for these six responses when trying to reach $\mathbf{y}_{des}$.

The fact that the front is not reduced to a single point is already an indication of the conflicting behavior of the responses, which is also seen in the last row of Table 6 that contains the individual minima of the fitness function for these 45 elements. According to Eq. (10), we are measuring the magnitude of each individual difference, so that the desired value for a response is obtained when the fitness is equal to zero.

Table 6. Experimental conditions to achieve at least one of the desired response values, with their predicted values (original scales for factors and all the responses).

Z_1 (H_2O)	Z_2 (MeOH)	Z_3 (ACN)	U_4 (mL min^{-1})	R_{12}	R_{23}	R_{34}	R_{45}	t_i (min)	t_f (min)
0.32	0.14	0.54	0.61	1.10*	2.28	4.15	3.08	2.21	4.51
0.33	0.03	0.64	0.95	0.05	1.14*	2.25*	2.85	1.32	2.28
0.50	0.49	0.01	0.87	10.00	12.49	27.58	1.11*	2.41	59.07
0.41	0.06	0.53	0.74	1.98	3.99	7.74	2.70	2.00*	4.95
0.28	0.18	0.54	0.63	0.76	1.74	2.98	3.11	2.15	4.01*
*Min. fitness				$2.5\ 10^{-5}$	$4.4\ 10^{-4}$	1.15	$8.2\ 10^{-5}$	$4.6\ 10^{-6}$	$1.4\ 10^{-4}$

This is the case, the minimum value is zero, in all the responses, except for R_{34}, so that the pursued goal is achieved in all but one response, although not simultaneously. This is clearer in the remaining rows in Table 6 that show the solutions achieving the corresponding minimum fitness of each individual response. The first four columns correspond to the experimental conditions and, rather than writing the fitness values of the solutions (that only provide information about the difficulty in reaching $\mathbf{y}_{des}$) the last six columns of Table 6 show the predicted responses themselves, including 'undoing' the transformations applied, to make the interpretation easier.

Therefore, the first row means that with a mixture of 32% water, 14% methanol and 54% acetonitrile, and flow rate at 0.61 mL min^{-1}, the desired value R_{12} = 1.1 is expected, with the remaining pair-wise resolution of at least 2.28, with 2.2 minutes of initial time and 4.5 min of final time.

The second row says that decreasing Z_2 and increasing both Z_3 and U_4, both R_{23} = 1.1 and the value closest to 1.1 for R_{34}, which is 2.25, are achieved. Moreover, in doing so, the resolution between the first two peaks, R_{12}, is practically null, not admissible, and there is less than a minute between final and initial time, the latter less than 2 min, also not admissible.

In the third row, with almost no acetonitrile (Z_3) and flow rate at 0.87 mL min^{-1}, the desired value of 1.1 for R_{45} (the resolution between fourth and fifth peaks) takes more than 59 minutes to finish the experiment and, besides, the values of the first three responses are very far from their desired values.

In the last two rows, with 41% of water (Z_1) and 6% of methanol (Z_2), the desired t_i = 2.0 min is obtained, whereas to get t_f = 4.0 min requires to increase Z_2 and decrease Z_1, though in that case, R_{12} is not admissible.

Although the previous analysis is reduced to the extremes points of the front (the best possible value for each individual response), it serves to illustrate the conflicting behavior among responses and introduces the utility of exploring the Pareto front. The conclusion so far is that the ideal chromatogram as defined in the ATP cannot be obtained, so a compromise is needed to select the experimental conditions for the determination of the bisphenols. This compromise is easier to reach knowing the extent of the conflict among responses, and the possibilities achievable.

For the problem at hand, to simplify the selection, all the solutions with any resolution less than 1.0, initial time less than 1.7 minutes or final time greater than 5.5 minutes are

discarded. Fig. 2 displays the remaining solutions in the form of a parallel coordinates plot, first the experimental conditions, then the predicted responses.

In the graph, the value of each coordinate is plotted against its position in the vector, so that each broken line in the plot represents a single vector. Additionally, the different scales make it difficult to follow the vectors, so the values have been scaled into a common range. As reference, the minimum and maximum values of each coordinate are at the bottom and top of the corresponding vertical line.

The conflicting behavior already mentioned is also seen in the graph: the first three resolutions R_{12}, R_{23}, R_{34} (pair-wise resolution between the first four peaks) increase or decrease simultaneously though in different proportion. However, it is clear that their behavior is opposed to the one of R_{45} that, in turn, is also opposed to the initial time t_i. Final time t_f and initial time also behave oppositely; this is an apparent contradiction with their positive correlation observed in Table 3. The reason is that the correlation coefficient in Table 3 (for the training data) means that, in general, the chromatogram needs more time to finish if it started later. The behavior observed in the solutions in the Pareto front says that trying to increase the initial time is an objective that is opposed to the objective of reducing the final time.

Journal Pre-proof

Here Fig. 2

Among the solutions depicted in Fig. 2, the interest lies in those complying with the criteria. Among them, the chosen conditions, solid blue line, correspond to the shortest experiment with final time of four minutes and more than two minutes for the initial time.

In summary, the optimal experimental conditions for the chromatographic determination of the five bisphenols were 22% of water, mixed with 58% methanol and 20% of acetonitrile, keeping the flow rate at 0.66 mL min^{-1}.

Fig. 3a) shows a chromatogram obtained in these selected CMP. To see the improvement, Fig. 3b) shows three overlapping peaks when the chromatogram is

obtained with a mobile phase that consists of a binary mixture (30% of water, 70% of acetonitrile) and a flow rate of 0.6 mL min^{-1}. With 0.8 mL min^{-1} of flow rate, in Fig. 3c) the first two peaks are still overlapping although a ternary mixture is used (26% water, 22% methanol and 52% acetonitrile).

<here Fig. 3>

4.4. Experimental verification of the CMP obtained

Ten determinations of a mixture of 4 mg L^{-1} of each bisphenol were performed with the control method parameters (experimental conditions for the ternary mixture and flow rate of the mobile phase) selected from the inversion of the prediction model.

The mean of the corresponding retention times (in minutes) of the five bisphenols were 2.182 for BPS, 2.456 for BPF, 2.769 for BPA, 3.256 for BPAF and 3.784 for BPZ.

For each of the six characteristics in the ATP, Table 7 shows the mean and 95% confidence intervals, comparing those obtained from the predicted values after inversion of the PLS2 model, and those computed from the ten experiments carried out. It is seen that the actual chromatographic data are included in the confidence intervals computed with the PLS2 model, except for the resolution R_{45} (between the fourth and fifth peaks of the chromatogram), which is, on the other hand, the response worst fitted by the model.

Table 7. 95% confidence intervals for the mean of the six ATP, a) theoretical data from the PLS2 model inversion and b) experimental data from HPLC-DAD analysis.

		R_{12}	R_{23}	R_{34}	R_{45}	t_i	t_f
a)	Lower limit	1.015	1.498	0.701	2.823	1.974	3.074
	Mean	1.545	2.093	2.280	3.047	2.114	3.999
	Upper limit	2.074	2.688	3.858	3.241	2.254	4.698
b)	Lower limit	1.870	2.121	2.638	2.670	2.099	4.055
	Mean	1.876	2.125	2.645	2.677	2.105	4.063
	Upper limit	1.882	2.129	2.652	2.684	2.111	4.070

After experimentally confirming that the computed CMP provide the expected ATP, the analytical procedure is validated in terms of trueness, decision limit (CCα) and detection capability (CCβ) for given fixed values of the risks of false positive and false negative (with probabilities α and β, respectively). The decision limit is described as "the value of the net concentration the exceeding of which leads, for a given error probability α, to the decision that the concentration of the analyte in the analyzed material is larger than that in the blank material" [40,41]. Moreover, the detection capability is "the true net concentration of the analyte in the material to be analyzed, which will lead, with probability 1-β, to the correct conclusion that the concentration in the analyzed material is larger than that in the blank material".

For the validation, a calibration line was fitted for each bisphenol (five in total) with the data obtained with six standard solutions in the range from 0 to 5 mg L^{-1} (see Table 1 in Section 2.3). Then, accuracy lines (that is, predicted concentration *vs* true concentration) were built and used to compute decision limit (CCα) and detection capability (CCβ), when both the probability of false positive and of false negative are set to 0.05.

Table 8 shows the details of both the calibration and accuracy lines, as well as CCα and CCβ for the five bisphenols. The p-values in row 6 correspond to the test for significance of the corresponding regression line (with null hypothesis, H_0: the model does not explain the variability of the response). As they are all less than 10^{-4}, the conclusion is that all the regression models are significant (5% significance level).

Table 8. Performance criteria of the analytical method. Parameters of calibration and accuracy lines (s_{yx} is the standard error of the regression). Decision limit and detection capability for α = β= 0.05.

	BPA	BPS	BPF	BPZ	BPAF
Calibration line					
Intercept	36.946	0.368	-1.069	-0.007	-0.735
Slope	55.768	19.644	61.757	41.228	48.886
Correlation coefficient	0.997	0.995	0.998	0.981	0.997
s_{yx}	7.652	3.785	6.662	15.411	7.233
P-value (significance of regression)	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
Accuracy line					
Intercept	$-2.14\cdot10^{-5}$	$4.76\cdot10^{-6}$	$9.52\cdot10^{-6}$	$1.19\cdot10^{-5}$	$-2.38\cdot10^{-6}$
Slope	1.000	1.000	1.000	1.000	1.000
s_{yx}	0.137	0.193	0.108	0.374	0.148
CCα (mg L^{-1})	0.217	0.305	0.171	0.591	0.234
CCβ (mg L^{-1})	0.424	0.596	0.334	1.156	0.458

The property of trueness was tested by computing the 95% joint confidence ellipse for the intercept and the slope of each accuracy line that shows that they are significantly equal to 0 and 1, as can be seen in the five ellipses depicted in Fig. 4 that contain the point (0,1). Therefore, the trueness is fulfilled. CCα and CCβ calculated through the accuracy lines, allow concluding that, with probabilities of false positive (α) and false negative (β) equal to 0.05, the analytical procedure enables one to determine, 0.424 mg L^{-1} of BPA, 0.596 mg L^{-1} of BPS, 0.334 mg L^{-1} of BPF, 1.156 mg L^{-1} of BPZ and 0.458 mg L^{-1} of BPAF.

Here Fig. 4

4.5. Comparative analysis

The validated procedure represents an advantage compared to similar works. To support this affirmation, Table 9 shows the 29 papers found when searching for analytical determination of the bisphenols by HPLC-DAD in the last decade. Except for two papers (the one codified as number 19 and the present work, which is in the last row of Table 9), all of them used binary mixtures in the mobile phase: MeOH/H_2O in 11 papers, ACN/H_2O in the remaining 17. About the elution mode, most of them (23 out of 29) used gradient, only in six of the papers the isocratic elution mode is utilized.

Table 9. Review of HPLC-DAD methods for the determination of BPA, BPS, BPF, BPZ and BPAF.

Code	Mobile phase	Elution mode	% Organic solvent	Reference
1	MeOH/H_2O	gradient	40 - 100	[42]
2	ACN/H_2O	gradient	50 - 70	[43]
3	MeOH/H_2O	isocratic	40	[44]
4	ACN/H_2O	gradient	42 - 85	[45]
5	ACN/H_2O	gradient	42 - 85	[46]
6	MeOH/H_2O	gradient	68 - 90	[47]
7	ACN/H_2O	gradient	42 - 85	[48]
8	MeOH/H_2O	gradient	35 - 100	[49]
9	MeOH/H_2O	gradient	35 - 100	[50]
10	MeOH/H_2O	gradient	35 - 100	[51]
11	MeOH/H_2O	gradient	35 - 100	[52]
12	MeOH/H_2O	gradient	35 - 100	[53]
13	ACN/H_2O	gradient	50 - 100	[54]
14	MeOH/H_2O	isocratic	60	[55]
15	ACN/H_2O	gradient	15 - 55	[56]
16	ACN/H_2O	gradient	50 - 80	[57]
17	ACN/H_2O	gradient	40 - 100	[58]
18	ACN/H_2O	gradient	50 - 80	[59]
19	MeOH/iPrOH/H_2O	gradient	60/5 - 80/2	[60]
20	ACN/H_2O	gradient	23 - 95	[61]
21	ACN/H_2O	gradient	15 - 55	[62]

22	ACN/H_2O	isocratic	50	[63]
23	ACN/H_2O	isocratic	60	[64]
24	ACN/H_2O	gradient	42 and 85	[65]
25	MeOH/H_2O	gradient	60 - 100	[66]
26	ACN/H_2O	isocratic	40	[67]
27	ACN/H_2O	gradient	40 - 70	[68]
28	ACN/H_2O	isocratic	35	[69]
29	MeOH/H_2O	gradient	40 - 90	[70]
30	MeOH/ACN/H_2O	isocratic	58/20	Present work

Finally, Fig. 5 depicts the retention time of BPS, BPF, BPA, BPAF and BPZ in the 29 published papers together with the ones obtained in the present paper. The works are identified with the number of the row they occupy in Table 9. In the present work, last position (number 30), the chromatogram takes only 4.06 min (total time), the shortest among those in Table 9.

Looking individually to each bisphenol, the minimum retention times (in min) listed in the bibliography of Table 9 are: 2.00 (number 28), 3.25 (number 28), 4.20 (number 7), 5.80 (number 7) and 6.30 (number 25) for BPS, BPF, BPA, BPAF and BPZ, respectively. With the exception of the BPS, these times are greater than those obtained in this work, already written in Section 4.4: 2.182, 2.456, 2.769, 3.256 and 3.784 min, respectively.

Here Fig. 5

5. Conclusions

The methodology developed for the inversion of PLS models is integrated into the scope of AQbD and enables the analysis of the conflict among six correlated responses that are used to define the ATP for a chromatographic determination of five bisphenols.

A systematic use of experimental designs, for a combined design and its reduction via the D criterion provide a representative training set for fitting the PLS2 model. After the fitting and validation of the prediction model, the study of the Pareto-optimal solutions provided by the inversion makes it possible the selection of the CMP to perform the chromatographic determination with the preset characteristics.

With these CMP, the analytical procedure implemented and validated represents an improvement with respect to similar works.

The developed procedure is general and can be applied to handle any other instrumental applications.

Conflict of interest

The authors declare no competing interests.

Acknowledgments

This work was supported by Spanish MINECO (AEI/FEDER, UE) and Consejería de Educación de la JCyL through projects CTQ2017-88894-R and BU052P20, co-financed with European Regional Development Funds. M.M. Arce wishes to thank Junta de Castilla y León and Fondo Social Europeo for her predoctoral grant.

References

[1] FDA, PAT- A Framework for Innovative Pharmaceutical Development, Manufacturing and Quality Assurance, Guidance for industry, Pharmaceutical CGMPs, Rockville, September 2004. https://www.fda.gov/media/71012/download

[2] European Commission, EU Guidelines for Good Manufacturing Practice for Medicinal Products for Human and Veterinary Use, Volume 4, Annex 15, Brussels,

2015. Last visit May 22, 2020. https://ec.europa.eu/health//sites/health/files/files/eudralex/vol-4/2015-10_annex15.pdf

[3] S. García-Muñoz, T. Kourti, J.F. MacGregor, F. Apruzzese, M. Champagne, Optimization of Batch Operating Policies. Part I. Handling Multiple Solutions, Ind. Eng. Chem. Res. 45 (2006) 7856–7866. https://doi.org/10.1021/ie060314g

[4] E. Tomba, M. Barolo, S. García-Muñoz, General framework for latent variable model inversion for the design and manufacturing of new products, Ind. Eng. Chem. Res. 51 (2012) 12886–12900. https://doi.org/10.1021/ie301214c

[5] M. Ottavian, E. Tomba, M. Barolo, Advanced process decision making using multivariate latent variable methods, in: M. Ierapetritou, R. Ramachandran (Eds.), Process simulation and data modeling in solid oral drug development and manufacture, Humana Press, Springer, New York, 2016, pp. 159-189. https://doi.org/10.1007/978-1-4939-2996-2_6

[6] F. Yacoub, J.F. MacGregor, Product optimization and control in the latent variable space of nonlinear PLS models, Chemometr. Intell. Lab. 70 (2004) 63–74. https://doi.org/10.1016/j.chemolab.2003.10.004

[7] S. Lakshminarayanan, H. Fujii, B. Grosman, E. Dassau, D.R. Lewin, New product design via analysis of historical databases, Comput. Chem. Eng. 24 (2000) 671–676. https://doi.org/10.1016/S0098-1354(00)00406-3

[8] U.G. Indahl, The O-PLS methodology for orthogonal signal correction -is it correcting or confusing?, J. Chemometr. 34 (2020) e2884. https://doi.org/10.1002/cem.2884

[9] S. Yin, S.X. Ding, P. Zhang, A. Hagahni, A. Naik, Study on modifications of PLS approach for process monitoring, IFAC Proceedings Volumes 44:1 (2011) 12389-12394. https://doi.org/10.3182/20110828-6-IT-1002.02876

[10] S. Ruiz, M.C. Ortiz, L.A. Sarabia, M.S. Sánchez, A computational approach to partial least squares model inversion in the framework of the process analytical technology and quality by design initiatives, Chemometr. Intell. Lab. 182 (2018) 70–78. https://doi.org/10.1016/j.chemolab.2018.08.014

[11] S. Ruiz, L.A. Sarabia, M.C. Ortiz, M.S. Sánchez, Residual spaces in latent variables model inversion and their impact in the design space for given quality characteristics, Chemometr. Intell. Lab. 203 (2020) 104040. https://doi.org/10.1016/j.chemolab.2020.104040

[12] R. Peraman, K. Bhadraya, Y.P. Reddy, Analytical Quality by Design: A tool for regulatory flexibility and robust analytics, Int. J. Anal. Chem. 868727 (2015) 1-9. https://doi.org/10.1155/2015/868727

[13] P. Das, A. Maity, Analytical Quality by Design (AQbD): A new horizon for robust analytics in pharmaceutical process and automation, International Journal of Pharmaceutics and Drug analysis 5 (2017) 324-337. http://www.ijpda.com/admin/uploads/4VdJA6.pdf

[14] T. Tome, N. Žigart, Z. Časar, A. Obreza, Development and Optimization of Liquid Chromatography Analytical Methods by Using AQbD Principles: Overview and Recent Advances, Org. Process Res. Dev. 23 (2019) 1784−1802. https://doi.org/10.1021/acs.oprd.9b00238

[15] S. Orlandini, S. Pinzauti, S. Furlanetto, Application of quality by design to the development of analytical separation methods, Anal. Bioanal. Chem. 405 (2013) 443–450. https://doi.org/10.1007/s00216-012-6302-2

[16] S. Orlandini, B. Pasquini, C. Caprini, M. Del Bubba, L. Squarcialupi, V. Colotta, S. Furlanetto, A comprehensive strategy in the development of a cyclodextrin-modified microemulsion electrokinetic chromatographic method for the assay of diclofenac and its impurities: Mixture-process variable experiments and quality by

Journal Pre-proof

design, J. Chromatogr. A 1466 (2016) 189–198. https://doi.org/10.1016/j.chroma.2016.09.013

[17] COM (1999) 706. Communication from the Commission to the Council and the European Parliament. Community Strategy for Endocrine Disrupters, a range of substances suspected of interfering with the hormone systems of humans and wildlife. Commission of the European Communities, Brussels, 1999.

[18] COM (2001) 262. Communication from the Commission to the Council and the European Parliament on the implementation of the Community Strategy for Endocrine Disrupters, a range of substances suspected of interfering with the hormone systems of humans and wildlife, COM (1999) 706. Commission of the European Communities, Brussels, 2001.

[19] SEC (2007) 1635. Commission Staff Working Document on the implementation of the Community Strategy for Endocrine Disrupters, a range of substances suspected of interfering with the hormone systems of humans and wildlife, (COM (1999) 706), (COM (2001) 262) and (SEC (2004) 1372). Commission of the European Communities, Brussels, 2007.

[20] Commission Directive (EU) 2017/898 of 24 May 2017 amending, for the purpose of adopting specific limit values for chemicals used in toys, Appendix C to Annex II to Directive 2009/48/EC of the European Parliament and of the Council on the safety of toys, as regards bisphenol A. Off. J. Eur. Union L 138 (2017) 128.

[21] Commission Regulation (EU) 2018/213 of 12 February 2018 on the use of bisphenol A in varnishes and coatings intended to come into contact with food and amending Regulation (EU) No 10/2011 as regards the use of that substance in plastic food contact materials. Off. J. Eur. Union L 41 (2018) 6.

[22] Commission Regulation (EU) 2016/2235 of 12 December 2016 amending Annex XVII to Regulation (EC) No 1907/2006 of the European Parliament and of the

Council concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) as regards bisphenol A. Off. J. Eur. Union L 337 (2016) 3.

[23] Regulation (EC) No 1223/2009 of the European Parliament and of the Council of 30 November 2009 on cosmetic products. Off. J. Eur. Union L 342 (2009) 59.

[24] 2019/C 404/07. Opinion of the European Committee of the Regions — Towards a comprehensive EU framework on endocrine disruptors. Reference document: Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions — Towards a comprehensive European Union framework on endocrine disruptors, COM(2018) 734 final. Off. J. Eur. Union C 404 (2019) 34.

[25] A. Ullah, M. Pirzada, S. Jahan, H. Ullah, N. Turi, W. Ullah, M.F. Siddiqui, M. Zakria, K.Z. Lodhi, M.M. Khan, Impact of low-dose chronic exposure to bisphenol A and its analogue bisphenol B, bisphenol F and bisphenol S on hypothalamo-pituitary-testicular activities in adult rats: A focus on the possible hormonal mode of action, Food Chem. Toxicol. 121 (2018) 24-36. https://doi.org/10.1016/j.fct.2018.08.024

[26] J. Moreman, O. Lee, M. Trznadel, A. David, T. Kudoh, C.R. Tyler, Acute toxicity, teratogenic, and estrogenic effects of bisphenol A and its alternative replacements bisphenol S, bisphenol F, and bisphenol AF in Zebrafish embryo-larvae, Environ. Sci. Technol. 51 (2017) 12796-12805. https://doi.org/10.1021/acs.est.7b03283

[27] G. Russo, A. Capuozzo, F. Barbato, C. Irace, R. Santamaria, L. Grumetto, Cytotoxicity of seven bisphenol analogues compared to bisphenol A and relationships with membrane affinity data, Chemosphere 201 (2018) 432-440. https://doi.org/10.1016/j.chemosphere.2018.03.014

[28] A. Usman, M. Ahmad, Computational study suggesting reconsideration of BPA analogues based on their endocrine disrupting potential estimated by binding

affinities to nuclear receptors, Ecotox. Environ. Safe 171 (2019) 154-161. https://doi.org/10.1016/j.ecoenv.2018.12.071

[29] K. Owczarek, B. Kudłak, V. Simeonov, Z. Mazerska, J. Namieśnik, Binary mixtures of selected bisphenols in the environment: their toxicity in relationship to individual constituents, Molecules 23 (2018) 3226. https://doi.org/10.3390/molecules23123226

[30] B.M. Wise, N.B. Gallagher, R. Bro, J.M. Shaver, W. Winding, R.S. Koch, PLS Toolbox 8.8.1, Eigenvector Research Inc., Wenatchee, WA, USA, 2020.

[31] MATLAB, version 9.7.0.1190202 (R2019b), The Mathworks, Inc., Natick, MA, USA, 2019.

[32] STATGRAPHICS Centurion 18 Version 18.1.12, Statpoint Technologies, Inc., Herndon, VA, USA, 2020.

[33] D. Mathieu, J. Nony, R. Phan-Than-Lu, NEMRODW (Version 2015), L.P.R.A.I., Marseille, France, 2015.

[34] L.A. Sarabia, M.C. Ortiz, DETARCHI. A program for detection limits with specified assurance probabilities and characteristic curves of detection, TrAC-Trend. Anal. Chem. 13 (1994) 1-6. https://doi.org/10.1016/0165-9936(94)85052-6

[35] L.A. Sarabia, M.C. Ortiz, M.S. Sánchez, Response Surface Methodology, in: S. Brown, R. Tauler, B. Walczak (Eds.), Comprehensive Chemometrics. Chemical and Biochemical Data Analysis, second ed., Elsevier, 2020, pp. 287-326. https://doi.org/10.1016/B978-0-12-409547-2.14756-0

[36] C.M. Jaeckle, J.F. Macgregor, Product design through multivariate statistical analysis of process data, AIChE Journal 44 (1998) 1105-1118. https://doi.org/10.1002/aic.690440509

[37] L.A. Sarabia, M.S. Sánchez, M.C. Ortiz, Introduction to ranking methods, in: M. Pavan, R. Todeschini (Eds.), Scientific Data Ranking Methods: Theory and

Applications, Data Handl. Sci. Techn., Volume 27, 2008, pp. 1-50. https://doi.org/10.1016/S0922-3487(08)10001-6

[38] K. Deb, Multi-Objective Optimization using Evolutionary Algorithms, Wiley Interscience Series in Systems and Optimization, Wiley, Chichester, 2001.

[39] M.M. Arce, S. Sanllorente, M.C. Ortiz, L.A. Sarabia, Easy-to-use procedure to optimise a chromatographic method. Application in the determination of bisphenol-A and phenol in toys by means of liquid chromatography with fluorescence detection, J. Chromatogr. A 1534 (2018) 93-100. https://doi.org/10.1016/j.chroma.2017.12.049

[40] International Organization for Standardization, ISO 11843, Capability of Detection, Part 1: Terms and Definitions and Part2: Methodology in the Linear Calibration Case, Genève, Switzerland, 2000.

[41] Commission Decision (EC), No 2002/657/EC Of 12 August 2002 Implementing Council Directive 96/23/EC Concerning the Performance of Analytical Methods and the Interpretation of Results, Off. J. Eur. Commun. L 221 (2002) 8.

[42] L. Wang, Q. Li, L. Zhang, A convenient approach for the determination of multiple trace BPs using an in-syringe-assisted solid phase microextraction system packed with elastic spongy graphene rods coupled with HPLC, Anal. Methods 9 (2017) 2673–2681. https://doi.org/10.1039/c7ay00352h

[43] Q. Zhou, Z. Jin, J. Li, B. Wang, X. Wei, J. Chen, A novel air-assisted liquid-liquid microextraction based on in-situ phase separation for the HPLC determination of bisphenols migration from disposable lunch boxes to contacting water, Talanta 189 (2018) 116-121. https://doi.org/10.1016/j.talanta.2018.06.072

[44] L. Wang, D. Zhang, X. Xu, L. Zhang, Application of ionic liquid-based dispersive liquid phase microextraction for highly sensitive simultaneous determination of

three endocrine disrupting compounds in food packaging, Food Chem. 197 (2016) 754-760. https://doi.org/10.1016/j.foodchem.2015.11.042

[45] Y. Li, C. Yang, J. Ning, Y. Yang, Cloud point extraction for the determination of bisphenol A, bisphenol AF and tetrabromobisphenol A in river water samples by high-performance liquid chromatography , Anal. Methods 6 (2014) 3285-3290. https://doi.org/10.1039/C3AY42191K

[46] Y. Li, Y. Liao, Y. Guoc, Y. Yang, Determination of bisphenol-A, 2,4-dichlorophenol, bisphenol-AF and tetrabromobisphenol-A in liquid foods and their packaging materials by vortex-assisted supramolecular solvent microextraction/high-performance liquid chromatography, Anal. Methods 5 (2013) 5037–5043. https://doi.org/10.1039/C3AY40586A

[47] J. Yin, Z. Meng, Y. Zhu, M. Song, H. Wang, Dummy molecularly imprinted polymer for selective screening of trace bisphenols in river water, Anal. Methods 3 (2011) 173–180. https://doi.org/10.1039/C0AY00540A

[48] D. Yang, G. Li, L. Wu, Y. Yang, Ferrofluid-based liquid-phase microextraction: Analysis of four phenolic compounds in milks and fruit juices, Food Chem. 261 (2018) 96-102. https://doi.org/10.1016/j.foodchem.2018.04.038

[49] X. Sun, J. Wang, Y. Li, J. Jin, J. Yang, F. Li, S.M. Shah, J. Chen, Highly class-selective solid-phase extraction of bisphenols in milk, sediment and human urine samples molecularly using well-designed dummy molecularly imprinted polymers, J. Chromatogr. A 1360 (2014) 9-16. https://doi.org/10.1016/j.chroma.2014.07.055

[50] X. Sun, J. Wang, Y. Li, J. Jin, B. Zhang, S.M. Shah, X. Wang, J. Chen, Highly selective dummy molecularly imprinted polymer as a solid-phase extraction sorbent for five bisphenols in tap and river water, J. Chromatogr. A 1343 (2014) 33-41. https://doi.org/10.1016/j.chroma.2014.03.063

[51] J. Yang, Y. Li, J. Wang, X. Sun, R. Cao, H. Sun, C. Huang, J. Chen Molecularly imprinted polymer microspheres prepared by Pickering emulsion polymerization for selective solid-phase extraction of eight bisphenols from human urine samples, Anal. Chim. Acta 872 (2015) 35-45. https://doi.org/10.1016/j.aca.2015.02.058

[52] J. Yang, Y. Li, J. Wang, X. Sun, S.M. Shah, R. Cao, J. Chen, Novel sponge-like molecularly imprinted mesoporous silica material for selective isolation of bisphenol A and its analogues from sediment extracts, Anal. Chim. Acta 853 (2015) 311-319. https://doi.org/10.1016/j.aca.2014.09.051

[53] H. Sun, Y. Li, J. Yang, X. Sun, C. Huang, X. Zhang, J. Chen , Preparation of dummy-imprinted polymers by Pickering emulsion polymerization for the selective determination of seven bisphenols from sediment samples, J. Sep. Sci. 39 (2016) 2188-2195. https://doi.org/10.1002/jssc.201501305

[54] T. Tuzimski, D. Pieniążek, G. Buszewicz, G. Teresiński, QuEChERS-Based Extraction Procedures for the Analysis of Bisphenols S and A in Breast Milk Samples by LC-QqQ-MS, Journal of AOAC INTERNATIONAL 102 (2019) 23–32. https://doi.org/10.5740/jaoacint.18-0297

[55] L. Wang, Z. Zhang, X. Xu,D. Zhang, F. Wang, L. Zhang, Simultaneous determination of four trace level endocrine disrupting compounds in environmental samples by solid-phase microextraction coupled with HPLC, Talanta 142 (2015) 97-103. https://doi.org/10.1016/j.talanta.2015.04.043

[56] E. Herrero-Hernández, R. Carabias-Martínez, E. Rodríguez-Gonzalo, Use of a bisphenol-A imprinted polymer as a selective sorbent for the determination of phenols and phenoxyacids in honey by liquid chromatography with diode array and tandem mass spectrometric detection, Anal. Chim. Acta 650 (2009) 195-201. https://doi.org/10.1016/j.aca.2009.07.043

[57] D.C. Morelli, G. Bernardi, L. Morés, M.E. Pierri, E. Carasek, A green - high throughput –extraction method based on hydrophobic natural deep eutectic solvent for the determination of emerging contaminants in water by high performance liquid chromatography – diode array detection, J. Chromatogr. A 1626 (2020) 461377. https://doi.org/10.1016/j.chroma.2020.461377

[58] T. Tuzimski, S. Szubartowski, Method Development for Selected Bisphenols Analysis in Sweetened Condensed Milk from a Can and Breast Milk Samples by HPLC–DAD and HPLC-QqQ-MS: Comparison of Sorbents (Z-SEP, Z-SEP Plus, PSA, C18, Chitin and EMR-Lipid) for Clean-Up of QuEChERS Extract, Molecules 24 (2019) 2093. https://doi.org/10.3390/molecules24112093

[59] P. Baile, J.Medina, L. Vidal, A. Canals, Determination of four bisphenols in water and urine samples by magnetic dispersive solid-phase extraction using a modified zeolite/iron oxide composite prior to liquid chromatography diode array detection, J. Sep. Sci. 43 (2020) 1808-1816. https://doi.org/10.1002/jssc.201901022

[60] H. Wang, S. Song, M. Shao, Y. Gao, C. Yang, Y. Li, W. Wang, Y. He, P. Li, Determination of bisphenol analogues in food-contact plastics using diode array detector, charged aerosol detector and evaporative light-scattering detector, Ecotox. Environ. Safe. 186 (2019) 109778. https://doi.org/10.1016/j.ecoenv.2019.109778

[61] M. Eckardt, T. J. Simat, Bisphenol A and alternatives in thermal paper receipts - a German market analysis from 2015 to 2017, Chemosphere 186 (2017) 1016-1025. https://doi.org/10.1016/j.chemosphere.2017.08.037

[62] E. Herrero-Hernández, E. Rodríguez-Gonzalo, M. S. Andrades, S, Sánchez-González, R. Carabias-Martínez, Occurrence of phenols and phenoxyacid herbicides in environmental waters using an imprinted polymer as a selective sorbent, Sci. Total Environ. 454–455 (2013) 299-306. http://doi.org/10.1016/j.scitotenv.2013.03.029

[63] Z. Liu, Z. Xu, Y. Liu, Y. Liu, B. Lu, L. Ma, Supramolecular imprinted polymeric stir bar sorptive extraction followed by high-performance liquid chromatography for endocrine disruptor compounds analysis, Microchem. J. 158 (2020) 105163. https://doi.org/10.1016/j.microc.2020.105163

[64] J. Liu, H. oiu, F. Zhang, Y. Li, Zeolitic imidazolate framework-8 coated Fe_3O_4@SiO_2 composites for magnetic solid-phase extraction of bisphenols, New J. Chem. 44 (2020) 5324-5332. https://doi.org/10.1039/d0nj00006j

[65] D. Yang, Y. Wang, H. Li, Y. Yang, Acid-base-governed deep eutectic solvent-based microextraction combined with magnetic solid-phase extraction for determination of phenolic compounds, Microchim. Acta 187 (2020) 124. https://doi.org/10.1007/s00604-020-4109-y

[66] M. Háková, L. Chocholoušová Havlíková, J. Chvojka, Jakub Erben, P. Solich, F. Švec, D. Šatínský, , Polycaprolactone nanofibers functionalized with a dopamine coating for on-line solid phase extraction of bisphenols, betablockers, nonsteroidal drugs, and phenolic acids, Microchim. Acta 186 (2019) 710. https://doi.org/10.1007/s00604-019-3846-2

[67] N. Li, j. Chen, Y.P. Shi, Magnetic nitrogen-doped reduced graphene oxide as a novel magnetic solid-phase extraction adsorbent for the separation of bisphenol endocrine disruptors in carbonated beverages, Talanta 201 (2019) 194-203. https://doi.org/10.1016/j.talanta.2019.04.002

[68] X. Han, J. Chen, H. Oiu, Y.P. Shi, Solid/liquid phase microextraction of five bisphenol-type endocrine disrupting chemicals by using a hollow fiber reinforced with graphene oxide nanoribbons, and determination by HPLC-PDA, Microchim Acta 186 (2019) 375. https://doi.org/10.1007/s00604-019-3498-2

[69] A. Kaleniecka, P. K. Zarzycki, Analysis of Selected Endocrine Disrupters Fraction Including Bisphenols Extracted from Daily Products, Food Packaging and Treated

Wastewater Using Optimized Solid-Phase Extraction and Temperature-Dependent Inclusion Chromatography, Molecules 24 (2019) 1285. https://doi.org/10.3390/molecules24071285

[70] G. Peng, Y. Lu, W. You, Z. Yin, Y. Li, Y. Gao, Analysis of five bisphenol compounds in sewage sludge by dispersive solid-phase extraction with magnetic montmorillonite, Microchem. J. 157 (2020), 105040. https://doi.org/10.1016/j.microc.2020.105040

Journal Pre-proof

Figure Captions

Figure 1. Experimental domain for control method parameters. For each flow rate of the mobile phase, the constrained simplex where the ternary mixtures can vary is marked in yellow. The experiments from the D-optimal design are in black, the 10 additional experiments are in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

Figure 2. Parallel coordinates plot of the reduced Pareto front, raw experimental conditions in the first four coordinates and predicted responses in the last six coordinates. All the values were range-scaled with the individual minima and maxima at the bottom and top, respectively, of each coordinate.

Figure 3. Chromatograms obtained with different control method parameters: a) the ones selected with the proposed procedure; b) 0.6 mL min^{-1} of flow rate with a binary mixture of 30% water and 70% acetonitrile; c) 0.8 mL min^{-1} of flow rate flow rate and a ternary mixture of 26% water, 22% methanol and 52% acetonitrile. Elution order: BPS, BPF, BPA, BPAF and BPZ.

Figure 4. 95% joint confidence region for intercept and slope of the accuracy lines. BPA in blue, BPF in green, BPAF in brown, BPS in yellow and BPZ in dark brown. + indicates the point (0,1). For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.

Figure 5. Retention time of analytes BPS, BPF, BPA, BPAF and BPZ from the 29 revised works in Table 9, whose row numbers serve as identification in the abscissa axis. Number 30 corresponds to the present work.

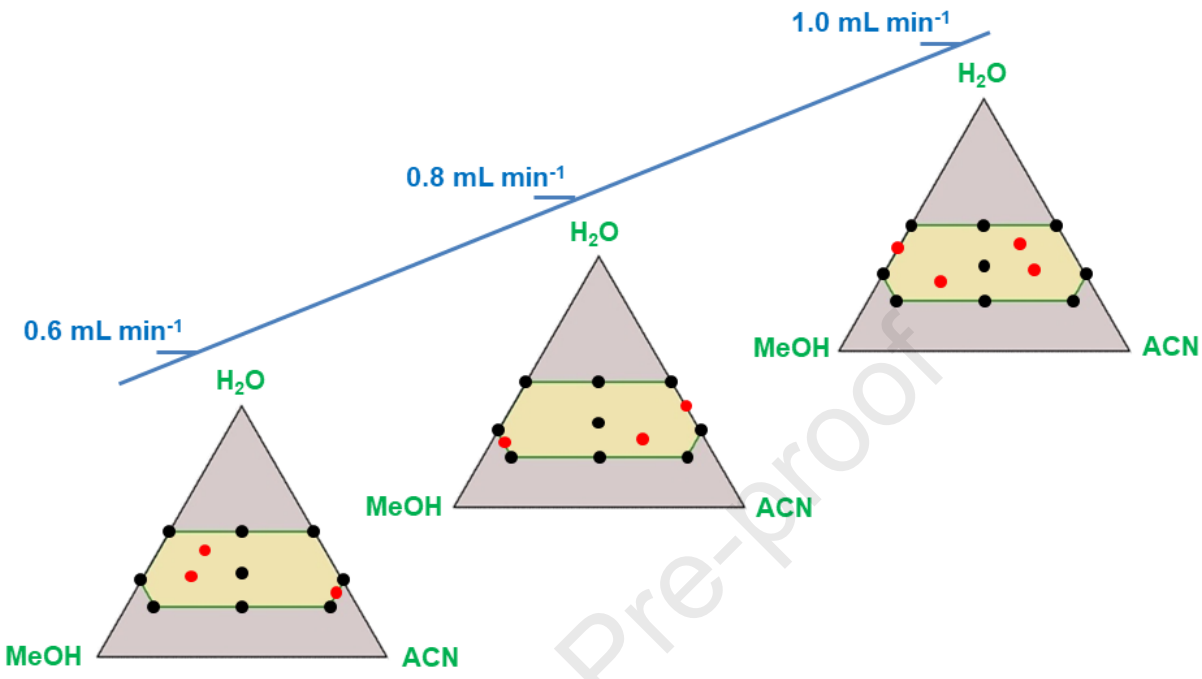


FIGURE 1

Journal Pre-proof

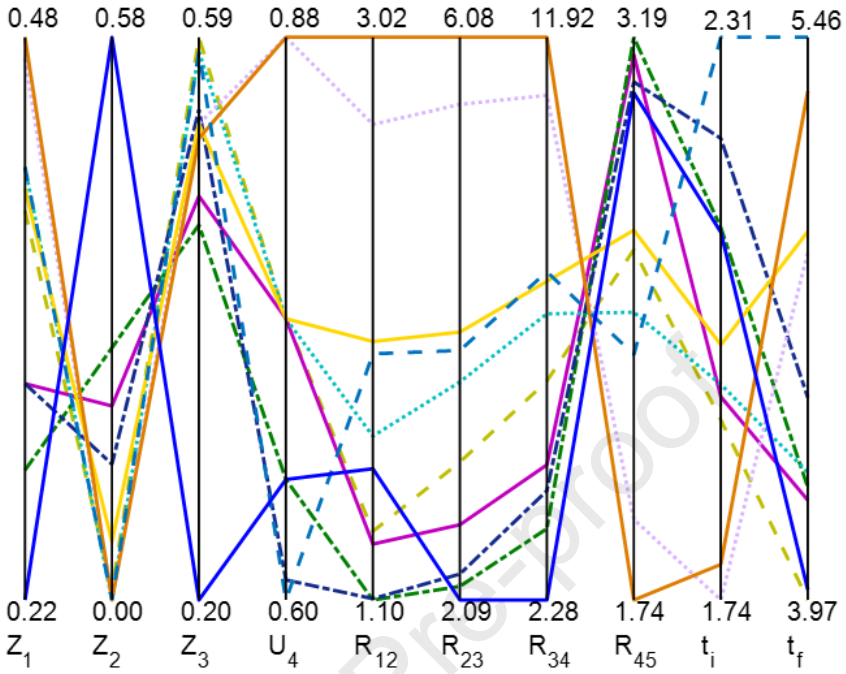


FIGURE 2

Journal Pre-proof

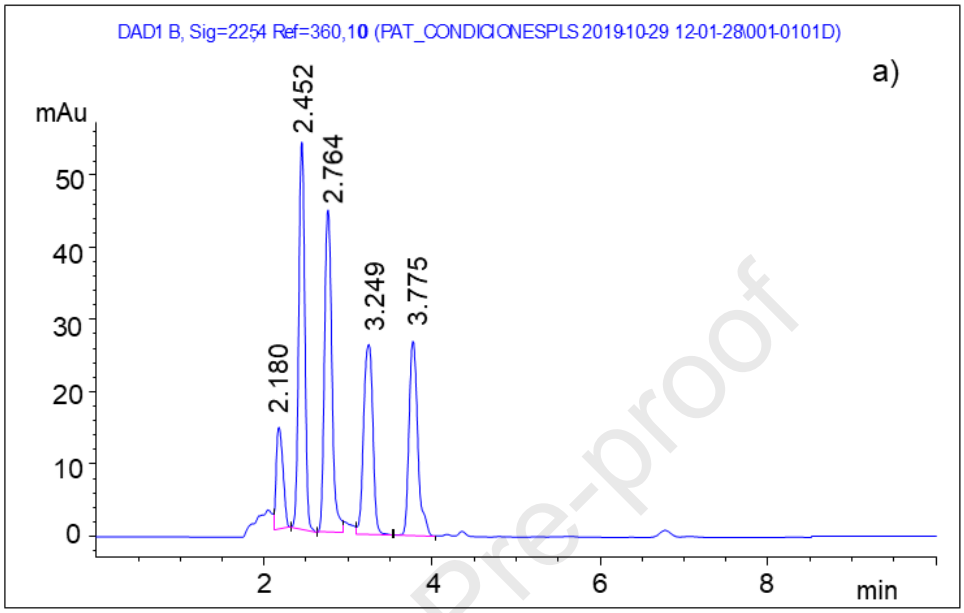


FIGURE 3A

Journal Pre-proof

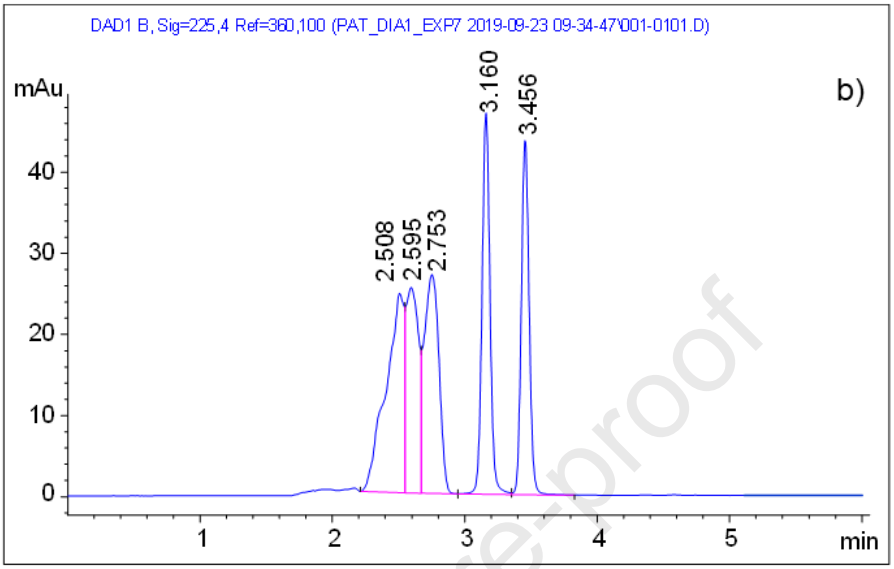


FIGURE 3B

Journal Pre-proof

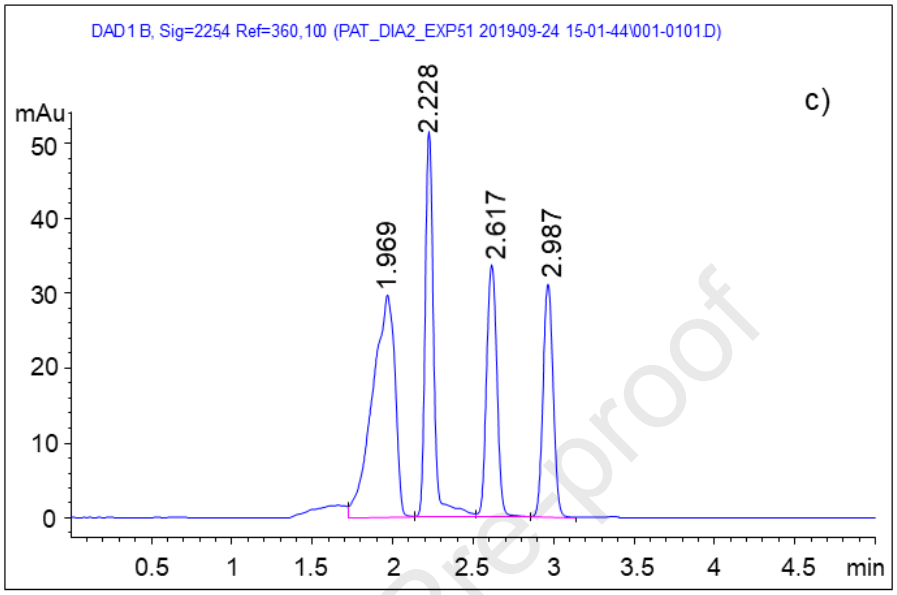


FIGURE 3C

Journal Pre-proof

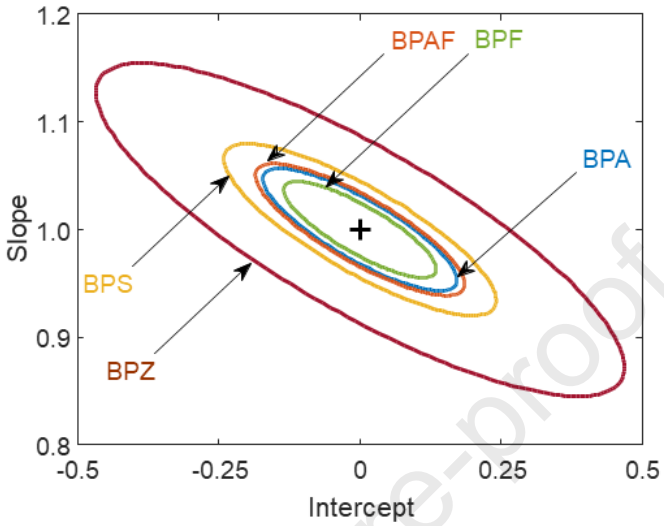


FIGURE 4

Journal Pre-proof

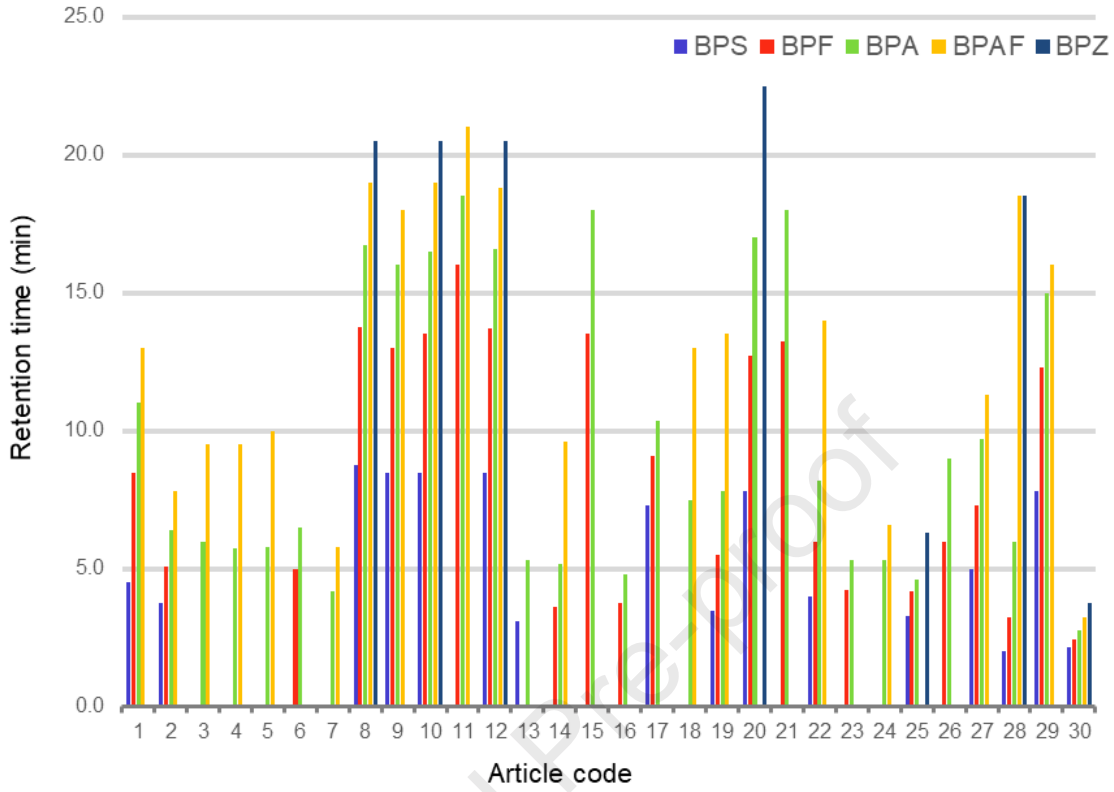


FIGURE 5

Journal Pre-proof

Journal Pre-proof

HIGHLIGHTS

Computational inversion of a PLS2 model inside the Analytical Quality by Design

Method to select chromatographic parameters to obtain an analytical target profile

Selection of representative chromatographic conditions using a D-optimal design

Fitting and inversion of a PLS2 looking for the Pareto front of several responses

Five bisphenols (including Bisphenol-A) determined by HPLC-DAD in less than 4 minutes

Journal Pre-proof

Journal Pre-proof

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Journal Pre-proof