AD-HOC BLOCKED DESIGN FOR THE ROBUSTNESS STUDY IN THE DETERMINATION OF DICHLOBENIL AND BAM IN ONIONS BY PROGRAMMED TEMPERATURE VAPORIZER-GAS CHROMATOGRAPHY-MASS SPECTROMETRY

Ana Herrero[1], Celia Reguera[1], M. Cruz Ortiz[1,*], Luis A. Sarabia[2], M. Sagrario Sánchez[2]
[1]*Dept. of Chemistry,* [2]*Dept. of Mathematics and Computation*
Faculty of Sciences, University of Burgos
Plaza Misael Bañuelos s/n, 09001 Burgos (Spain)

Abstract

An 'ad-hoc' experimental design to handle the robustness study for the simultaneous determination of dichlobenil and its main metabolite (2,6–dichlorobenzamide) in onions by Programmed Temperature Vaporizer-Gas Chromatography-Mass Spectrometry **(**PTV-GC-MS) is performed. Eighteen experimental factors were considered; 7 related with the extraction and clean up step, 8 with the PTV injection step and 3 factors related with the derivatization step. Therefore, a high number of experiments must be carried out that cannot be conducted in one experimental session and, as a consequence, the experiments of the robustness study must be performed in several sessions or blocks. The procedure to obtain an experimental design suitable for this task works by simultaneously minimizing the joint confidence region for the coefficient estimates and the correlation among them and with the block. In this way, the effect of the factors is not aliased with the block avoiding possible misinterpretations of the effects of the experimental factors on the analytical responses. The developed experimental design is coupled to the PARAFAC2 method, which allows solving some specific problems in chromatography when working with complex matrix such as co-elution of interferents (including silylation artifacts from the derivatization step) and small shifts in the retention time and, besides, the unequivocal identification of the target compounds according to document SANCO/12571/2013.

Keywords*: Robustness; ruggedness; ad-hoc blocked design; PARAFAC2; dichlobenil, PTV-GC-MS.*

1. Introduction

Checking robustness/ruggedness of an analytical method is a fundamental part of the method validation [1,2,3,4]. A worldwide adopted definition of robustness is that it is "*a measure of its capacity to remain unaffected by small, but deliberate variations in method parameters and provides an indication of its reliability during normal usage*" [5]. That is, this figure of merit refers to the effect on analytical results of small changes in the experimental conditions.

* Corresponding author. e-mail: mcortiz@ubu.es. Tel.: 34 947259571. Fax: 34 947258831.

The robustness of the experimental response to the factors can be then checked by slightly varying them above and below their ‘nominal’ values and seeing the effect on the response.

But analytical methods usually depend on several experimental factors; so many experiments are required to perform the robustness study, which might be an expensive and time consuming effort. This is especially relevant in procedures that include many steps (extraction, derivatization, chromatography, etc.) with experimental conditions involved so that a large number of factors has to be considered in the robustness testing.

The experiments to study the influence of these small changes are usually evaluated more efficiently by means of experimental design approaches with the factors at two levels, mostly by using screening designs, either a saturated fractional factorial or a Plackett-Burman design [6,7,8,9,10].

In some studies the number of experiments needed is so large that they cannot be carried out under homogeneous conditions, that is, in a single sequence or with the same GC liner, the same operator, etc. For example, when analytes are determined in complex matrices by means of PTV-GC (Programmed Temperature Vaporizer-Gas Chromatography), the validation or routine sequences typically imply a set of matrix-matched standards and samples, so in particular liners must be exchanged frequently to avoid that the response of the target analytes drops and/or to eliminate cross-contamination between sample runs.

Applications have been reported where liners are changed at a predetermined frequency (after 1, 5, 10, 20 injections) [11,12,13,14], and even automated liner exchange devices have been developed for this purpose. Therefore, in many cases the liner has to be exchanged in the course of the robustness study and this must be taken into account, otherwise the effect of changing the liner, if any, would be misattributed to the other experimental factors under study.

The case of the liner just exposed serves as an example of a general situation, confounding the so-called block effect (some factors difficult to control that remain under homogenous conditions in blocks of experiments but that may change from one block to another) with the effect of other controllable factors which are systematically and deliberately modified.

This issue can be approached by using blocked experimental designs, where the effect of variability that could arise from block changes is minimized and detected if it exists [15]. In this way the effect of the different blocks (different liners, sequences, etc.) can be studied separately from the effects on the response of the small changes in the experimental conditions. Adding a dummy variable in the model has been proposed to do this for response surface designs [16]. This binary variable (block variable) accounts for possible changes in the response and is related to a bias between the responses obtained in the different blocks.

But changes in variance and covariance of the estimated model coefficients are expected when blocking a design, therefore lack of correlation between the estimate of the coefficients of the blocking variable and the other coefficient estimates is a desirable property [16,17]. Sánchez et al. [18] developed an approach for blocking response surface and factorial designs so that they simultaneously have the largest possible D-value (which measures the joint precision of the coefficient estimates) with the smallest correlation between block and other coefficient estimates of the model and showed, through several examples, the effect of not considering a priori all these criteria on the properties of different experimental designs.

Screening saturated experimental designs are usual for robustness studies. In the present paper, the procedure of blocking in ref. [18] is extended for the first time to compute an ad-hoc screening saturated design for the problem at hand. The experimental design obtained is applied to the simultaneous determination of the herbicide dichlobenil (DIC) and its main metabolite, 2,6–dichlorobenzamide (BAM), in onions by PTV-GC-MS. The herbicide dichlobenil, despite its possible toxicity, is still used and is remarkably persistent in soil and, thus, it is possibly accumulated in the foods cultivated in them, such as onions. The analytical procedure is explained in detail in Ref. [19], where it can be seen that it depends on several variables.

In particular, eighteen experimental factors, related to the extraction/clean-up, derivatization and injection steps of the analytical procedure, are considered in this study. The liner of the PTV inlet is changed after each 15 injections in such a way that three different liners are used throughout the robustness study. Additionally to the experiments in the conditions stated for the experimental design, some more samples should be measured to evaluate recovery and also matrix-matched standards for calibration. Consequently, three experimental sessions (blocks) with different liner are needed.

The computed design is coupled to a Parallel Factor Analysis 2 (PARAFAC2) [20,21], a multiway technique which has proved to be very useful in solving common problems in GC-MS [22,23]. It is particularly helpful for determining compounds of interest in food commodities [19,24], for solving problems as small retention time shifts, severe interferences caused by unexpected derivatization artifacts or by co-eluents of the complex matrix which share m/z ratios with the target compounds. The second-order advantage of PARAFAC2 allows the determination of the target analytes in samples where unknown interferents are present without the need of calibrate them. In fact, if a three-way method had not been used in this case for extracting the contribution of the analytes to the signal, neither the unequivocal identification of dichlobenil and BAM nor their quantification could have been performed according to regulations in [25].

2. Theory

127 *2.1 Construction of the experimental design*
128
129 Although the procedure is completely general, for the sake of clarity, it is explained here only
130 for the robustness study at hand. As it has just mentioned, the experimental procedure under
131 study depends (or may depend) upon 18 experimental factors [19]. An experimental design is
132 set with the factors at two levels to perform the robustness study of the analytical procedure.
133 In this situation, a standard design, such as a Plackett-Burman [6] design, requires 20
134 experiments. In this work at most 15 experiments (15 injections of derivatized extracts of
135 complex matrix) can be performed under homogeneous conditions of the GC-MS system; i.e.
136 after 15 injections the GC system stops and the liner is changed. Consequently, the robustness
137 study designed in this way cannot be performed in a single session. That means that, to
138 consider the possible differences on the response due to external factors (and estimate its
139 effect, if any), the design must be blocked.
140
141 Additionally, 10 samples were planned for estimation of the variance of the method (two
142 series of five replicates each in two different sessions), and 9 more matrix-matched standards
143 needed for the subsequent quantification. Summing up all of them plus the, at least, 20
144 experiments of the robustness study for the validation process, at least three sessions (so three
145 blocks) would be necessary to perform all the experiments. Besides, it is necessary to take
146 into account when planning the design that the number of intended experiments in each block
147 is different: 6 experiments can be carried out in the first block (further to the 9 matrix-
148 matched standards), 10 experiments in the second one (plus 5 replicates to estimate variance)
149 and up to 10 in the third one (plus the other 5 replicates).
150
151 In terms of the model to be fitted to study the robustness of the response to (small) changes in
152 the experimental factors, we have 18 factors at two levels and the block at three levels, and
153 the goal is to decide if there is any significant effect on the response when moving the
154 experimental factors from low to high levels. Consequently, using the high level as the
155 reference level, we assume that the model that relates the response Y with the variation of
156 factors and blocks is written as in Eq. (1), where letter A represents the low level of each
157 experimental factor.
158

$$\begin{aligned} Y = {} & \beta_0 + \beta_{1A}x_{1A} + \beta_{2A}x_{2A} + \beta_{3A}x_{3A} + \beta_{4A}x_{4A} + \beta_{5A}x_{5A} + \beta_{6A}x_{6A} + \beta_{7A}x_{7A} + \\ & \beta_{8A}x_{8A} + \beta_{9A}x_{9A} + \beta_{10A}x_{10A} + \beta_{11A}x_{11A} + \beta_{12A}x_{12A} + \beta_{13A}x_{13A} + \beta_{14A}x_{14A} + \\ & \beta_{15A}x_{15A} + \beta_{16A}x_{16A} + \beta_{17A}x_{17A} + \beta_{18A}x_{18A} + \delta_1 x_{b1} + \delta_2 x_{b2} + \varepsilon \end{aligned} \quad (1)$$

159
160
161 In Eq. (1), ε denotes the experimental variability, which is assumed to follow a normal
162 distribution with the same variance σ^2 in all the experiments, and zero mean.
163

Moreover, x_{iA} (i = 1, …, 18) is an indicator variable, i.e. $x_{iA} = 1$ if factor x_i is at low level (level A) and 0 otherwise, so that coefficients β_{iA} (i = 1, …, 18) measure the effects on the response when the i-th experimental factor changes from the reference level to level A.

Regarding the blocking, it has been denoted differently in Eq. (1). Provided that the reference level is the high level (in this case, block 3) we use dummy-coding [15] so that, again, x_{b1} and x_{b2} are binary variables: x_{b1} is 1 only for the experiments in the first block, $x_{b2} = 1$ only in block 2 and the coefficients δ_i (i = 1, 2) take account of the possible shifts in the response due to differences among experimental sessions not attributable to the factors. Also, it is seen in Eq. (1) that the effect of the blocks on the response, if any, is additive.

After fitting the model to the experimental results, the significance of the experimental factors is decided with the corresponding coefficient estimate, so it seems clear that the coefficient estimates should be precise (small variance) and that correlation between the estimates of the coefficients related to blocks and of the remaining coefficients should be null, or at least, the closer to zero the better. This is so in order to guarantee that the interpretation of the effects of the experimental factors is independent of the block effect. Likewise, the uncorrelatedness among the coefficient estimates of β_{iA} is also a desirable property.

The key concept to bear in mind is that the precision of the estimated coefficients and the correlation between one another and with the blocking, apart from a constant factor, can be computed before doing the experiments.

For the particular case here, the so-called model matrix **X** is the matrix made up by the experiments to be carried out and adding three columns: a column with 1 in all positions, related to the independent term in the model of Eq. (1), β_0, and two columns for block, the dummy-coding we have already mentioned: (1, 0) for the experiments in block 1, (0, 1) for block 2 and (0, 0) for block 3.

Notice that the model in Eq. (1) has 21 coefficients to be estimated so in the general case, at least 21 experiments are needed. Denoting the number of experiments in the design by *N*, the model matrix is $N \times 21$ ($N \geq 21$).

The least squares estimator of the coefficients in Eq. (1) is given by $(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{y}$ and the variance-covariance matrix of the coefficient estimates is $(\mathbf{X}^t\mathbf{X})^{-1}\sigma^2$. Consequently, apart from the variance of the experimental error σ^2, the precision and correlation of the estimates depend upon matrix $(\mathbf{X}^t\mathbf{X})^{-1}$, which is known as the dispersion matrix and it is a symmetric matrix that is only determined by the planned design and the supposed model, Eq. (1) in this case.

Accordingly, in terms of precision of estimates and uncorrelatedness, the closer to zero the elements of the dispersion matrix the better. In particular, the main diagonal elements refer to the variance of the coefficient estimates and the off-diagonal terms refer to the covariance between one another. In practice, the variance of the coefficient estimates in the main diagonal of the dispersion matrix depends on the size of the experimental domain, so these values are usually scaled into the so-called Variance Inflation Factors (VIFs), so that we have a unique reference about the quality (in terms of precision) of the coefficient estimates: the best value for any VIF is one and the farther from that value, the worse the estimate. In any case, values above 4 indicate fully imprecise estimates. Details about this issue can be found, for example, in ref. [26].

As an illustration, without taking into account the blocking, a Plackett-Burman design with $N = 20$ experiments would be adequate, because the VIF of all coefficient estimates is one and it has the identity matrix I_{18} as correlation matrix among coefficient estimates (that means a square symmetric matrix of size 18×18 with ones in the main diagonal and zeros otherwise). In other words, all the coefficient estimates are uncorrelated to one another, the design is orthogonal.

However, the blocking of the design changes its properties [18]. In the present case, at first, the need of blocking the design implies that we have to move to a design with more than 21 experiments, for instance, the Plackett-Burman for $N = 24$, with 18 factors at two levels and the block at three levels. Table 1 contains the lower triangular part of the correlation matrix among coefficient estimates (it is always symmetric) for this design, d_1 and d_2 are the estimates for δ_1 and δ_2 and b_{iA} (i=1, …, 18) are the estimates for β_{iA} in Eq. (1). It is seen that the only non-null element, in bold in Table 1, is corr(d_1,d_2) which is equal to 0.58. That means that, except for the block coefficient estimates, the design is orthogonal and, in particular, orthogonally blocked. Besides, the VIF of all b_{iA} is 1, which guarantees the best possible precision for each one, whereas VIF for d_i is 1.5, still quite good.

Unfortunately, further to the need of doing 24 experiments instead of 21, the blocking is made in ‘regular’ blocks, as 6+12+6, i.e. 6 experiments are assigned to block 1, 12 to block 2 and the remaining 6 to block 3. In order to emphasize again the fact that the blocking of the design is a feature that should be considered from the beginning (when planning the experiment), if this same orthogonal design is blocked differently to adapt to our needs (6+10+8), just by changing two experiments (randomly selected) from block 2 to block 3, the resulting design is not orthogonal anymore nor either orthogonally blocked, as can be seen in the upper part of Table 1 that contains the upper triangular part of the new correlation matrix among coefficient estimates. Several coefficient estimates (from b_{1A} to b_{5A} and from b_{11A} to b_{15A}) are still uncorrelated to one another and with d_i (block coefficient estimates), small values of 0.05 (in absolute value) appear scattered in the remaining coefficient estimates but the main difference is in the correlation between the coefficient estimates d_i and $b_{iA,}$ some of which have increased up to 0.21 (in absolute value).

Besides, VIFs for some coefficient estimates, namely b_{iA} for i= 6, 8, 9, 10, 16 and 18, are now 1.05, VIF of d_1 is 1.40 and VIF of d_2 is 1.69. Although none of these values (correlation and VIFs) points to a strong correlation among coefficient estimates or a significant deterioration of the estimates themselves, they show the effect on the properties of the change of only two experiments.

In any case, we are looking for a design with good enough properties, blocked according to our needs, and with less experiments, ideally $N = 21$, that is, a design adapted to the problem and not forcing the problem to adapt to a specific design.

To look for such an 'ad-hoc' design, we use the algorithmic approach explained elsewhere [27], which is an evolutionary algorithm that evolves by searching experimental designs with optimal properties in several criteria defined by the user.

For the present case, the population is made up by experimental designs with N experiments for 18 factors at two levels all of them blocked in three blocks with 6, 10 and $N - 16$ experiments, and with good properties for fitting the model in Eq. (1). Specifically, the criteria to qualify the designs are related to both, the precision of the coefficient estimates and the correlation between one another, with special attention to the orthogonal (or near-orthogonal) blocking. Precisely, they are measured as:

i) The D-value that accounts for the joint precision of the coefficient estimates. It is defined for this case as

$$\frac{\det(\mathbf{X}^t\mathbf{X})}{N^{21}} \qquad (2)$$

In Eq. (2), *det* denotes the determinant of the matrix, **X** is the model matrix, and N is the number of experiments to estimate the 21 coefficients in the model in Eq. (1). The larger the D-value, the better the joint estimation, thus, during evolution, this number should be maximized.

ii) The maximum correlation (in absolute value) between coefficient estimates, precisely between b_{iA} and b_{jA} (i, j =1, …, 18, $i \neq j$) and between b_{iA} and d_j (j=1, 2, and i=1, …, 18). During evolution, this number should be minimized.

Due to the existence of more than one criterion, the optimum values (maximum or minimum) are not well or uniquely defined [28]. Therefore, the algorithm is a multi-objective evolutionary algorithm that evolves looking for 'optimal' solutions, in our case, preserving the so-called non-dominated solutions. For explanation, consider two solutions for a given problem, two different blocked experimental designs in the present case, B_1 and B_2, with N experiments in the vertices of an hypercube (18 dimensions), and qualified according to criteria i) and ii) as (c_{11}, c_{12}), (c_{21}, c_{22}) respectively. Design B_1 is said to dominate design B_2 if

B_1 is not worse than B_2 in the two objectives (i.e. $c_{11} \geq c_{21}$ and $c_{12} \leq c_{22}$) and at least it is strictly better in one of them (i.e. $c_{11} > c_{21}$ or $c_{12} < c_{22}$). Therefore, the non-dominated designs in a given set are those which are the best in at least one criterion. The Pareto-optimal front is the set of the non-dominated solutions for the entire search space.

With these conflicting criteria, for estimating the Pareto-optimal experimental designs, we use a genetic algorithm. For the implementation of the algorithm, experimental designs are unfolded into a single vector, then blocked with the proper size for each block and then qualified according to the criteria defined in the previous paragraph. Also, it is worth mentioning that only experimental designs with associated $\mathbf{X}^t\mathbf{X}$ matrix 'regular' enough are accepted in the population. This is quantified by using the condition number of the matrix, which is an indication of the accuracy of the results from matrix inversion when solving a system of linear equations. For details about the condition number, consult for example [29].

In each generation, off-springs are generated by systematically selecting two designs in the population (at random), and then double-point crossover and mutation with a predefined probability. The resulting off-springs with bad condition number are directly discarded. Furthermore, after computing fitness for the off-springs (their evaluation values according to the mentioned criteria), those whose fitness equals any other design already in the population, are also discarded.

When there are as many proper off-springs as individuals in the current population, the designs in the current population are merged together with the generated off-springs and the new enlarged population is sorted in levels of non-dominance and only the designs in the upper levels survive for the next generation (as many as needed to maintain population size along evolution). The main idea behind the procedure is that the non-dominance relation accounts for the designs which constitute the Pareto-optimal front of the actual population and, during evolution, it moves towards an estimate of the Pareto-optimal front of the whole set of designs.

Consequently, if the criteria are conflicting criteria, which is usually the case, the experimental designs in the final population (some of them) are an estimate of the Pareto-optimal front for the competing criteria, in such a way that moving among designs in the Pareto-optimal front improves one of the criteria by necessarily worsening another, but in the smallest possible amount.

2.2 PARAFAC2

Retention time shifts can occur in GC [21,30], but MS data rarely present alignment problems. Therefore, PARAFAC2 [20,21] is a valuable decomposition method for these signals. Precisely, if GC/MS data obtained for each compound of interest are arranged in a

three-way array or data tensor, $\underline{\mathbf{X}}$, PARAFAC2 decomposes the GC-MS data tensor into factors, according to the model:

$$\mathbf{X}_k = \mathbf{A}_k\mathbf{D}_k\mathbf{B}^T + \mathbf{E}_k = \mathbf{P}_k\mathbf{H}\mathbf{D}_k\mathbf{B}^T + \mathbf{E}_k\,,\ \ k = 1, \ldots, K \qquad (3)$$

where the matrix $\mathbf{X}_k$ is the *k*-th slab with dimension $I \times J$ (*J* ions monitored at *I* scans during the chromatographic elution of the analytes), $\mathbf{A_k}$ is the matrix of loadings of the chromatographic mode estimated for the *k*-th sample, $\mathbf{D}_k$ is a diagonal matrix that holds the *k*-th row of the sample mode, **B** is the loading matrix of the spectral mode, $\mathbf{E}_k$ is the matrix of the residuals, $\mathbf{P}_k$ is an orthogonal matrix of the same size as $\mathbf{A}_k$, and **H** is a small square matrix with dimension equal to the number of factors.

Unlike PARAFAC [31], PARAFAC2 does not assume that $\mathbf{A_k}$ is the same for all *k* but the cross-product matrix $\mathbf{A}_k{}^T\mathbf{A}_k$, which allows some deviation in the chromatographic mode and it can be shown that this constraint leads to the uniqueness of the model under mild conditions [32]. Therefore, the "second-order advantage" of the estimates is guaranteed, i.e. the analytical response of the compounds of interest is not masked by possible co-eluents.

In a PARAFAC model, an index useful to check somehow the validity of the trilinearity assumption is the core consistency diagnostic (CORCONDIA) developed by Bro and Kiers [33]. Recently, an approach for calculating a model diagnostic similar to core consistency but for PARAFAC2 models has been developed [34].

3. Experimental

3.1 Reagents

Ethyl acetate (SupraSolv) was obtained from Merck (Darmstadt, Germany). Dichlobenil and BAM (PESTANAL grade), and sodium sulphate anhydrous (p.a.) were purchased from Sigma-Aldrich (Madrid, Spain). Internal standards, 3,5–dichlorobenzonitrile (97%) (ISDIC) and 2,4–dichlorobenzamide (98%) (ISBAM) were purchased from Aldrich (Steinheim, Germany), and BSTFA from Supelco (PA, USA). 2 mL DisQuE clean-up tubes containing 150 mg anhydrous magnesium sulphate plus 50 mg PSA sorbent and 50 mg C_{18} were obtained from Waters (Milford, MA, USA).

3.2 Instrumental

All analyses were performed on an Agilent (Agilent Technologies, Wilmington, DE, USA) 7890A gas chromatograph coupled to an Agilent 5975 Mass Selective Detector (MSD). The injection system consisted of a septumless head and a PTV inlet (CIS 6 from Gerstel, Mülheim an der Ruhr, Germany) equipped with empty multi-baffled deactivated quartz

liners. Injections were carried out using a MultiPurpose Sampler (MPS 2XL from Gerstel) with a 10 μL syringe. Chromatographic separations were carried out on an Agilent DB-5ms (30 m × 0.25 mm i.d., 0.25 μm film thickness) column. A Velp Scientifica RX3 Vortex shaker (Milan, Italy) was used. The control of the temperature in the derivatization step was performed using a water bath equipped with a Digiterm 200 immersion thermostat (JP Selecta S.A., Barcelona, Spain). Extracts were centrifuged on a Sigma 2-16K refrigerated centrifuge (Osterode, Germany). A miVac DUO centrifugal concentrator (Genevac Ltd., Ipswich, UK) operating at low pressure was used for faster evaporation.

3.3 Experimental procedure

Stock solutions of DIC, BAM, ISDIC and ISBAM were prepared in ethyl acetate and stored in a refrigerator at 4ºC. Two sets of seven standard solutions were prepared so that each one contains the appropriate concentration of each compound before derivatization (5, 10, 15, 20, 25, 30 and 35 μg L^{-1} of the analytes and internal standards).

Samples for the robustness study (containing 20 μg L^{-1} of DIC, BAM, ISDIC and ISBAM), replicates to estimate variance (containing 20 μg L^{-1} of the four compounds) and matrix-matched standards (containing 10, 12.5, 15, 17.5, 20, 22.5, 25, 27.5 and 30 μg L^{-1} of DIC and BAM, and 20 μg L^{-1} of ISDIC and ISBAM) were prepared from onions purchased from a local food store. Each onion was cut with a knife, put into freezer overnight and blended while frozen until it reaches homogeneous texture. Next, 10 ± 0.1 g of the onion was transferred to a 50 mL centrifuge tube (tube 1) and appropriately fortified with the internal standards and the target compounds. Samples were extracted with 10 mL ethyl acetate in the presence of 10 g of sodium sulphate and followed by vortex mixing for 2 min (t_{mix1}). The homogenate was centrifuged at 3000 rpm (s_{centr1}) for 10 min (t_{centr1}) at 4ºC. A volume of 1.2 mL of the extract was transferred into a DisQue clean-up (tube 2). The tube 2 was shaken for 30 s (t_{mix2}) and next centrifuged at 10000 rpm for 1 min (t_{centr2}) at 4ºC. 0.8 mL of the supernatant was transferred into a 10 mL glass tube and evaporated to dryness under vacuum in a centrifugal concentrator during 10 min (t_{evap}) at 50ºC (T_{evap}). The residue was reconstituted with 0.8 mL of ethyl acetate before derivatization.

80 μL of each solution (standard solutions and reconstituted extracts) was derivatized in a 2 mL screw cap vial by addition of 56 μL of BSTFA (V_{BSTFA}). The vial was capped, shaken vigorously and allowed to stand at 44.5 ºC (D_{Temp}) for 42 min (D_{Time}) by placing the mixture in a water bath.

The derivatized solutions were injected into the GC-MS system; the PTV was operated in the solvent vent mode. Two microliter of each solution were injected at 50 μL s^{-1} (s_{inj}). During injection, the inlet temperature was held at 40 ºC ($T_{PTVinit}$) for 0.5 min ($t_{PTVinit}$), while the column head pressure was fixed at 9 psi (P_{init}) with a flow rate through the split vent of 100

mL min^{-1} ($vent_{flow}$). At 0.3 min ($vent_{time}$) the split valve was closed and next the temperature of the PTV was ramped at 10 ºC s^{-1} (r_{PTV}) up to 280 ºC (T_{PTVend}), which was held for 5 min. The split valve was re-opened 1 min after injection to purge the inlet at a vent flow of 60 mL min^{-1}.

The oven temperature was programmed as follows: the oven was maintained at 40ºC for 1 min and ramped at 120 ºC min^{-1} up to 120 ºC, held for 1 min and next ramped at 8 ºC min^{-1} to 200 ºC. A post-run step was performed for 4 min at 280 ºC. The transfer line temperature was set at 280 ºC, the ion source temperature at 230ºC, and the quadrupole at 150 ºC. The electron multiplier was set at 1424 V and the source vacuum at 10^{-5} torr. The solvent delay was fixed at 4.5 min. The mass spectrometer was operated in electron ionization mode at 70 eV. The acquisition was performed on selected ion monitoring (SIM) mode with two acquisition windows so that 5 ions (ion dwell time of 80 ms) were monitored for each compound: 100, 136, 171, 173 and 175 for DIC and ISDIC; and 145, 173, 175, 246, and 248 for BAM and ISBAM. The flow rate of the carrier gas was maintained at 1.1 mL min^{-1}.

3.4 Software

MSD ChemStation E.02.01.1177 (Agilent Technologies, Inc.) and Gerstel Maestro 1 (version 1.3.20.41/3.5) were used for GC-MS data acquisition and processing. Pareto-optimal experimental designs were computed with proper programs written in MATLAB version 7.10 (The MathWorks) and analyzed with NEMRODW [35]. PARAFAC2 models were performed with the PLS_Toolbox [36] for use with MATLAB. The calibration models were fitted and validated with STATGRAPHICS Centurion XVI [37].

4. Results and discussion

4.1. Experimental design

Starting with $N = 21$ experiments in the experimental design we look for, after some trials with different probability of mutation, the estimate of the Pareto-optimal front extracted from the final population with 100 individuals turned out to have 16 experimental designs, all of them blocked as 6+10+5. Their values in terms of the criteria being optimized are depicted in Fig. 1, on the horizontal axis the D-values (maximize) and the maximum of the absolute value of the correlation values is in the vertical axis (minimize).

The conflicting behavior of the two objectives is apparent in Fig. 1: to obtain larger D-values (better joint estimation of the coefficients) at least the correlation between two of the coefficient estimates (including block coefficients) should be larger; or vice versa, for decreasing all the correlation values some lost in D might be assumed. Also, the fact that it is

an estimate of the Pareto-optimal front gives a quantitative idea of the expected loss and gain in all criteria, which in this case is difficult to interpret specially for the D-value.

To have an idea about the differences seen in Fig. 1, we consider the two extremes of the Pareto-optimal front, namely the design with the least correlation values (the one with values at the bottom left corner of Fig. 1), let us denote it as D_1, and the one with best D value (at the top right part in Fig. 1), $D_{16.}$ After computing the individual VIFs and correlation matrices of their coefficient estimates, Fig. 2 is the bar graph of the VIFs, light red bars for D_1 and dark blue bars for $D_{16.}$

The VIFs for b_{iA} (i = 1, …, 18) in both designs are quite close to 1 with small differences between them, although it is stated again that the improvement of the overall precision via the D-criterion not necessarily implies that all individual VIFs are smaller (in general the dark blue bars are smaller but not in all the cases, above all for the VIFs of the coefficient estimates for the block effect). Regardless, the largest VIFs, also in both designs, are the ones related to d_1 and d_2 that, in any case, remain less than 2.

About the other criterion, Table 2 contains the correlation matrices for both designs. The lower part for D_1 and the upper part for D_{16}. In both cases, the designs are not orthogonal and neither orthogonally blocked, because small (in absolute value) correlation coefficients appear along the matrices, but not larger than 0.17 for D_1 and not larger than 0.22 for D_{16}, 0.21 if we only take into account the correlation between coefficient estimates to check the effect of the experimental factors, b_{iA}.

The numbers underlined in Table 2 refer to the worst case when studying only the correlation among b_{iA}, those in italics correspond to the largest (always in absolute value) correlation between b_{iA} and d_j (j = 1, 2, i = 1, …, 18,) related to the ‘orthogonal blocking’. These values, as well as VIFs of the coefficient estimates for the sixteen designs in the Pareto-optimal front are available in the supplementary material, Table S1.

Besides, to study if it is really worthy to move until a design with 24 experiments (such as the one whose properties are in Table 1), we apply the evolutionary algorithm, with the same criteria, to look for user-blocked experimental designs with coefficient estimates precise enough, small absolute values of correlation, and with N = 22, 23 and 24 experiments, so that they are blocked as 6 + 10 + (N – 16).

There were also sixteen Pareto-optimal designs for N = 22, none of them orthogonal nor orthogonally blocked; fifteen designs for N = 23 and, in addition to the design in Table 1, there are twenty-five more designs in the Pareto-optimal front for N = 24 that achieve lower correlation coefficients at the cost of slightly decreasing the D-value.

Among the 73 computed designs, the worst VIF among coefficient estimates for the main factors is 1.3, which increases up to 2 when looking at the estimates for the blocking variables. Regarding the correlation, the maximum absolute value among coefficient estimates b_{iA}, $\left|corr\left(b_{iA}, b_{jA}\right)\right|$, is 0.22 for N = 21; 0.21 for N = 22; 0.25 for N = 23 and 0.19 for N = 24; whereas in relation to the orthogonal blocking, the maximum value for $\left|corr\left(b_{iA}, d_j\right)\right|$ is 0.22, 0.22, 0.20, and 0.21 for N equal to 21, 22, 23, and 24, respectively. The detailed properties of all these designs are also in Table S1 of the supplementary material.

Finally, the improvement of the criteria when increasing the number of experiments seemed not justify the increased experimental effort so the experimental design to perform the robustness study was chosen among the 16 designs with 21 experiments, precisely D_1, the most orthogonally blocked design in this front (i.e. the least correlation between block and the coefficient estimates for the factors), which is the first design corresponding to 21 experiments in Table S1 (first row) with correlation matrix between coefficient estimates in the lower triangular part of Table 2.

4.2 Robustness study

Before analyzing the robustness, it is necessary to guarantee that the responses (to be used in the design) correspond unequivocally to the analytes under study. The unequivocal identification of the compounds is performed according to Document SANCO/12571/2013 [25], where tolerances for retention time and relative abundance for diagnostic ions are established for pesticide residues analysis in food. This regulation requires that at least three relative ion abundances must be within the tolerance intervals when working with a standard mass resolution detector (in the SIM mode) for determining compounds for which a maximum residue limit (MRL) has been established, which is the case of DIC (with MRL equal to 20 μg kg^{-1}). In addition, the retention time must correspond to the one of a reference standard with a tolerance of ±0.2 min.

With the aim of obtaining the tolerance intervals, a set of 7 reference standards was prepared. Then, a PARAFAC2 model was obtained for each compound and the tolerance intervals for relative abundances were calculated following the procedure described in ref. [19]. The tolerance intervals obtained for the relative abundances are shown in Table 3, fourth column.

Next, the robustness study was performed according to the experimental plan in Table 4 to fit the model of Eq. (1). The levels considered for each experimental factor are shown in Table 4. The factors were slightly varied taking into account a reasonable variability range and the possibility of changes of the devices used. In all cases the experimental variables were set above and below their 'nominal' values shown in Section 3.3, except the PTV initial temperature, $T_{PTVinit}$, because the GC system only enables variations of ±1ºC; this last variable ranges from the nominal value (40ºC) to 41ºC since 2 ºC would be an excessively

large temperature interval. In the case of the block, two binary variables are used to codify the three blocks according to Eq. (1).

Injections of derivatized extracts of onion fortified with 20 $\mu g\ L^{-1}$ of both analytes and internal standards were carried out according to the experimental plan. Experiments were performed in random order within each block. In addition to the 21 experiments of the robustness study, 19 determinations were made so that, samples were distributed in blocks as follows:

Block 1: 9 matrix-matched standards together with the 6 experiments corresponding to the first block in the design (run 1-6 of Table 4), that is, 15 injections of complex matrix performed with the same liner;
Block 2: a first set of 5 replicates with the 10 experiments belonging to the second block of the design (run 7-16 in Table 4);
Block 3: a second set with 5 additional replicates with the 5 remaining experiments of the design (run 17-21 of Table 4).

In addition to the 40 samples detailed in the preceding paragraphs, another set of 7 standards prepared in ethyl acetate were also measured. This is so because the estimates of three-way models have proven [38] to be more precise when analyzing complex matrices if both, standards and fortified samples (matrix-matched standards), are included in the decomposition step.

GC-MS data were arranged in a data tensor of dimension $I \times 5 \times 47$ for each compound. I refers to the number of scans acquired around the retention time of the corresponding chromatographic peak (I was 17, 21, 17 and 8 for DIC, BAM, ISDIC and ISBAM respectively), 5 are the number of diagnostic ions monitored for each compound and 47 is the number of analyzed samples. Next, the four data tensors were decomposed using the PARAFAC2 method by applying the ALS algorithm to each tensor with unimodality and non-negativity constrains in the chromatographic mode and non-negativity constraint in spectral and sample modes respectively. Models with different number of factors were built and explained variance, CORCONDIA index, degree of agreement of the loadings of the sample mode with the ones expected and the unequivocal identification of each compound according to regulations were compared for choosing the number of factors of the final models.

Concerning the models for BAM and DIC two factors were necessary, while for ISBAM and ISDIC three factors were required. No outliers were found in the models built considering the Q residual and Hotelling's T^2 indices at the 99% confidence level. The models explained 99.67, 99.49, 99.87 and 99.10% of variance and have CORCONDIA index of 100, 99.99, 99.95 and 99.64 for BAM, DIC, ISBAM and ISDIC, respectively. The CORCONDIA index

is always greater than 99.6, concluding that the unequivocal identification of the analytes is guaranteed.

The loadings obtained for the chromatographic, spectral and sample modes for BAM are shown in Fig. 3 (chromatographic mode is referred to loadings scaled by the last mode loadings). The loadings of the 2^{nd} factor (green dashed lines and bars) of the chromatographic and spectral modes in Figs. 3a and 3b are coherent with BAM. It is confirmed that the retention time of the chromatographic profile obtained for each sample of the robustness study is within the tolerance interval estimated from the reference standards.

Besides, the loadings of the spectral mode of this factor (green dashed bars) matched the spectrum obtained from the reference standards, i.e. the relative abundances are within the corresponding tolerance intervals; 5^{th} and 6^{th} columns in Table 3 show that there is just one ion with relative abundance outside the corresponding tolerance interval. In any case, at least three ions are inside the intervals so the requirements are fulfilled and the unequivocal identification is guaranteed in the terms of regulations. Therefore the 2^{nd} factor of the model is unequivocally related to BAM, while the 1^{st} factor (in blue continuous line in Fig. 3a, blue solid bars in Fig. 3b, and blue filled triangles in Fig. 3c) is related to an interferent that overlaps the chromatographic peak of BAM and, despite the fact that it has a different MS spectrum, shares some of the diagnostic ions of BAM. That is, the three-way technique is capable of successfully extracting the contribution of the target analyte to the signals. If a three-way method would not have been used, the unequivocal identification of BAM could not have been performed according to regulations because interferents greatly contributed to many of the five diagnostic ions acquired, i.e. the diagnostic ions (their relative abundances) might have been outside the tolerance intervals.

As regards the sample mode, Fig. 3c, the loadings of the 2^{nd} factor, green circles, calculated for the standards of BAM (both matrix-matched standards, samples 31 to 40, and standards in ethyl acetate, samples 41 to 47) are fairly close to the expected pattern, the higher the concentration the larger the loading. The loadings of the 1^{st} factor, blue triangles, are only significant for those samples which imply injection of complex matrix in the chromatographic system, while for standards in ethyl acetate, last 7 samples in Fig. 3c, these loadings are almost null; clearly this factor is related to an interferent from the matrix. That is, the second-order advantage of the PARAFAC2 model has allowed the unequivocal identification of BAM in these samples where unknown interferents are present without the need to calibrate them too.

The corresponding PARAFAC2 decomposition for ISBAM required three factors. Their corresponding loadings are shown in Figs. 3d-f. The 1^{st} factor in this case, in blue, was unequivocally related to ISBAM taking into account the corresponding tolerance intervals for retention time and relative abundances despite the significant interference of co-eluents in both chromatographic (Fig. 3d) and spectral ways (the other two factors share ions with

ISBAM, Fig. 3e). The loadings of the sample profile of the 1st factor (blue triangles in Fig. 3f) shows the expected pattern, the same as in Fig. 3c, whereas the loadings on the second factor (green circles) are related to some interferent from the matrix, because they are null in the standards (last seven samples). The third factor can be attributable to some derivatization artifact because its loadings are non-null in all the samples, standards included.

These loadings were used to standardize the loadings obtained previously for BAM (Fig. 3c); the standardized loadings are shown in Fig. 4a. Dividing the loadings of BAM by the loadings of ISBAM corrects variations in injection and other factors of the system. When comparing Figs. 3c and 4a, it is noticeable that the high variability in the loadings of replicates in Fig. 3c has been significantly reduced as expected when standardization is performed. The standardized loadings in Fig. 4a (samples 1 to 21) are the responses used to fit the model in Eq. (1), in this case, for BAM, last column in Table 4.

Similarly, from the analogous PARAFAC2 decompositions performed for DIC and ISDIC, standardized loadings were obtained for DIC (using the loadings of the sample mode of ISDIC for standardization). Previous to the standardization, again, the unequivocal identification was checked: table 3 shows that for DIC, all ion ratios are within the tolerance intervals, and for ISDIC one ion has 19.53% relative abundance that is outside but really close to the lower endpoint of the tolerance interval, 19.56%, but the remaining three are well inside the corresponding interval.

Like for BAM, the first 21 standardized loadings shown in Fig. 4b are now the response of the design for the robustness study for DIC, 'penultimate column in Table 4.

The standardized loadings were used to fit the model of Eq. (1) for BAM and DIC. The two models are significant at 5% significance level (*p*-values were 0.03 and 0.01 for BAM and DIC respectively; null hypothesis: the linear regression model is not significant). To compute these probabilities an external estimate of the residual variance has been used, the one obtained from the standardized loadings of the two independent sets of replicates, each of size 5, conducted in the third and second experimental sessions, number 22-31 in Fig. 4. To do it, a two-sided *F*-test is carried out to test the null hypothesis ‘the variances between both sets of replicates are equal’. Their associated *p*-values (0.35 for BAM and 0.47 for DIC) allow concluding that the variances are statistically equal and, thus, a pooled variance was obtained: 0.0066 and 0.1814 (with 8 degrees of freedom) for BAM and DIC, respectively.

Fig. 5 graphically shows the coefficients estimated for model in Eq. (1) for all the factors and for the block. The significant coefficients (at 5% significance level) are those with light orange color in Fig. 5: the effect of changing the time of shaking tube 2 (x_7 or t_{mix2}), the time of centrifugation of tube 2 (x_8 or t_{centr2}), the temperature and time of evaporation (x_9 or T_{evap} and x_{10} or t_{evap}, respectively) are significant for BAM, and changes on the response due to the variations made in time of vortex mixing of tube 1 (x_4 or t_{mix1}), in speed of centrifugation (x_6

or s_{centr1}), in time of centrifugation of tube 2 (x_8 or t_{centr2}), in the initial temperature of the PTV (x_{11} or $T_{PTVinit}$) and in the vent flow rate (x_{14} or $vent_{flow}$) are significant for DIC.

Therefore, the procedure is robust in both compounds for several of the factors studied (10 out of 18). However, analyst must be very cautious with the remaining eight that should be carefully controlled, in particular, the time of centrifugation of tube 2 (x_8) that affects both compounds. It is noticeable that if the decision is made at 1% significance level, then no factor is critical for BAM (except for the block) and only the effects of x_8, x_{11} and x_{14} are non-null for DIC, which is not surprising because DIC is more volatile than BAM, so small changes in $T_{PTVinit}$ and $vent_{flow}$ (x_{11} and x_{14} respectively) have a significant effect only for the first compound.

The effect of the block is significant in both cases, i.e. changing the liner and performing the measures in a different session significantly affect the analytical responses of both compounds. Therefore, a misinterpretation of the effects of the experimental factors considered in the robustness study might be made if an appropriate blocked design had not been used.

To persist in the idea about the need of properly choosing an experimental design and to see the effect on the estimated coefficients (and thus in the conclusions reached from them) of a poor selection of the experimental design, suppose that the theoretical -unknown- model is the one in fig. 5b written in full in eq. (4) below and, with the estimated variance, blocking and factors x_4, x_6, x_8, x_{11} and x_{14} have a significant effect on the response, at 5% significance level.

In this situation (which, to make easy the comparison, is the one obtained here although, in general, this is not known before doing the experiments), and for different experimental designs, responses are simulated emulating the ones that would be obtained after experimentation.

In detail, for twenty times, the simulation consists of:

1. Random selection of 21 experiments, out of the 24 from the Hadamard (Plackett-Burman) design with properties in Table 1.
2. Block the selected design as 6+10+5. To maintain the same conditions as the ones in the population evolution, designs with an ill-conditioned information matrix are directly discarded and no coefficient estimates are computed with them. However, in a general situation, unless explicitly computed, the user could not be aware of this fact, which would lead directly to almost any value (disproportionately large in general) for the coefficient estimates.
3. For 1000 times, compute (simulate) experimental responses. To do it, the model is applied with the corresponding experimental conditions and some ‘random noise’ is

added by using a normal distribution with variance 0.1814 (the one estimated for DIC).

4. For each of the 1000 sets of estimated coefficients, the significance of every individual one is decided at 95% confidence level, and the number of non-significance decisions is counted.

Table S2 in the supplementary material contains the resulting proportions of non-significance, per design and coefficient. Additionally, the last two columns contain the corresponding minimum and maximum per coefficient. The same proportions are depicted in figure 6 for each coefficient, the first 18 for the b_{iA}, the last two for the blocks.

Because of the way these values are computed, it is expected that 95% of the times the coefficient is non-significant, the right decision is made, whereas a proportion of 0.05 of rejection is expected for the truly significant coefficients, namely b_4, b_6, b_8, b_{11}, b_{14} and the coefficients of the blocks.

Looking at fig. 6, it is clear that the true null coefficients (b_1, b_2, b_3, b_5, b_7, b_9, b_{10}, b_{12}, b_{13}, and from b_{15} to b_{18}) are declared non-significant approximately 95% of the times. However, when it comes to the non-null coefficients in the model, the decision clearly depends on the design used. For instance, b_4 is erroneously concluded as non-significant between 22% and 93% of the times, depending on the design; b_6 between 45% and 93% or, more importantly, the last two coefficients d_1 and d_2 (related to blocks), are erroneously considered non-significant more than 70% of the times.

4.3 Recovery rate

The recovery rate was calculated from the 10 replicates but taking into account that they were not measured all together, but 5 of them were measured with the experiments of the second block and the other 5 in the third block.

The independence between the coefficients of the block and the other estimated coefficients of the model also guarantees that the model can be used independently for correcting the additive effect of the block on the standardized loadings of the replicates. In this way, the standardized loadings of the matrix-matched standards, measured in the first block, can be corrected and used to estimate the concentration of the ten replicates, once these replicates have been corrected too.

As the reference level in the fit model was block 3, the samples measured in the other two blocks have to be corrected by using the corresponding coefficients. The procedure is based on the computed model according to Eq. (1). To illustrate the procedure, consider the model

for response Y_1, that is, the standardized loadings of DIC, which, reordering the terms to highlight the estimates for the block, is:

$$\begin{aligned} Y_1 = {} & 0.962x_{b1} + 1.156x_{b2} \\ & +3.86 - 0.33x_{1A} + 0.19x_{2A} - 0.13x_{3A} + 0.63x_{4A} - 0.37x_{5A} - 0.52x_{6A} - 0.27x_{7A} \\ & -0.93x_{8A} - 0.13x_{9A} + 0.05x_{10A} + 1.11x_{11A} - 0.09x_{12A} - 0.27x_{13A} - 0.77x_{14A} \\ & -0.37x_{15A} - 0.14x_{16A} - 0.41x_{17A} - 0.31x_{18A} \end{aligned} \quad (4)$$

As x_{b1} is 1 only for the experiments in the first block, substituting in Eq. (4) for experiments in block 1, we have

$$\begin{aligned} Y_1 - 0.962 = {} & 3.86 - 0.33x_{1A} + 0.19x_{2A} - 0.13x_{3A} + 0.63x_{4A} - 0.37x_{5A} - 0.52x_{6A} - 0.27x_{7A} \\ & -0.93x_{8A} - 0.13x_{9A} + 0.05x_{10A} + 1.11x_{11A} - 0.09x_{12A} - 0.27x_{13A} - 0.77x_{14A} \\ & -0.37x_{15A} - 0.14x_{16A} - 0.41x_{17A} - 0.31x_{18A} \end{aligned} \quad (5)$$

Whereas for block 2 is $x_{b1} = 0, x_{b2} = 1$ so the following holds

$$\begin{aligned} Y_1 - 1.156 = {} & 3.86 - 0.33x_{1A} + 0.19x_{2A} - 0.13x_{3A} + 0.63x_{4A} - 0.37x_{5A} - 0.52x_{6A} - 0.27x_{7A} \\ & -0.93x_{8A} - 0.13x_{9A} + 0.05x_{10A} + 1.11x_{11A} - 0.09x_{12A} - 0.27x_{13A} - 0.77x_{14A} \\ & -0.37x_{15A} - 0.14x_{16A} - 0.41x_{17A} - 0.31x_{18A} \end{aligned} \quad (6)$$

Finally, for block 3 is $x_{b1} = 0, x_{b2} = 0$ and thus

$$\begin{aligned} Y_1 = {} & 3.86 - 0.33x_{1A} + 0.19x_{2A} - 0.13x_{3A} + 0.63x_{4A} - 0.37x_{5A} - 0.52x_{6A} - 0.27x_{7A} \\ & -0.93x_{8A} - 0.13x_{9A} + 0.05x_{10A} + 1.11x_{11A} - 0.09x_{12A} - 0.27x_{13A} - 0.77x_{14A} \\ & -0.37x_{15A} - 0.14x_{16A} - 0.41x_{17A} - 0.31x_{18A} \end{aligned} \quad (7)$$

Consequently, it is clear that correcting samples measured in block 1 is simply subtract the corresponding coefficient d_1 (0.962 for DIC as in Eq. (5)) to the standardized loadings, while d_2 (1.156 in Eq.(6)) should be subtracted for correcting samples of block 2. Similarly, the coefficients are $d_1 = 0.077$ and $d_2 = 0.272$ for BAM (Fig. 5).

Table 5 shows the standardized loadings of the replicates and their corresponding corrections. To compute the concentration, two calibration models were fit by means of least squares (LS) regression with the corrected standardized loadings of the matrix-matched standards for both analytes of interest. By comparing the LS and MAD (minimize absolute deviations) regression lines, two outliers were detected for DIC (standards containing 17.5 and 25 μg L^{-1}) and three for BAM (standards with 15, 20 and 30 μg L^{-1}). After removing those points, the resulting calibration models were Y = –1.800 + 0.170 *x* (with correlation coefficient, ρ, equal to 0.990, and standard error of estimation, $s_{y/x}$, equal to 0.196) for DIC and Y= –0.312 + 0.079 *x* (ρ = 0.995, $s_{y/x}$ = 0.062) for BAM.

These calibration lines were used to calculate the concentration of BAM and DIC of the replicates, all of them with a nominal concentration of 20 μg L^{-1}. The values found are also shown in Table 5, as well as the recovery rates reached with the analytical procedure (column 6). Mean recovery rates and the semi-length of their 95% confidence intervals are also computed for each compound and group of replicates (column 7). Notice that this quantification is possible because of the proposed procedure.

5. Conclusions

With present analytical instrumentation, there is a large number of factors that can affect the analytical response. Consequently, when doing a robustness study a high number of experiments must be conducted, in such a way that the experimentation cannot be carried out under homogeneous experimental conditions (same session, liner, etc.). In that case, if the fact that the experiments are performed in different blocks is significant (that is, if the effect of the block is statistically significant) and it is not taken into account in the study, the effects of the experimental factor considered will be misattributed leading to wrong conclusions.

Using the methodology proposed in this work, an ‘ad-hoc’ experimental design which simultaneously minimizes the volume of the joint confidence region for the coefficient estimates and the correlation between one another (including the block), aliased factors with the block and possible misinterpretations of effects are avoided. This methodology also gives a tool for correcting different signals measured in the same experimental sessions or blocks, which provides the possibility of having additional samples to estimate figures of merit of the analytical procedure, such as recovery rate in this work, in the same experimental study.

The validity of the data to implement the proposed procedure relies on the unequivocal identification of each compound, which is possible due to the use of PARAFAC2 as decomposition method, because of the second order advantage.

Extraction vortex mixing time, clean-up centrifugation time and speed, initial PTV temperature, temperature and time of evaporation, and solvent vent flow were the critical factors found in the robustness study performed for the determination of BAM and DIC in onions by PTV-GC-MS.

6. Acknowledgements

The authors thank the financial support provided by Ministerio de Economía y Competitividad and Junta de Castilla y León under projects CTQ2011-26022 and BU108A11-2, respectively.

812
813
814 **7. References**

[1] Commission Decision 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. L221 of 17 August 2002.

[2] S.L.R. Ellison, A. Williams (Eds.), Eurachem/CITAC guide: Quantifying Uncertainty in Analytical Measurement, 3rd ed., 2012. Available from www.eurachem.org.

[3] I. Taverniers, M. De Loose, E. Van Bockstaele, Trends in quality in the analytical laboratory. II. Analytical method validation and quality assurance, TRAC Trends Anal. Chem. 23 (2004) 535-552. DOI: 10.1016/j.trac.2004.04.001

[4] L. Cuadros-Rodríguez, R. Romero, J.M. Bosque-Sendra, The Role of the Robustness/Ruggedness and Inertia Studies in Research and Development of Analytical Processes, Crit. Rev. Anal. Chem. 35 (2005) 57-69. DOI:10.1080/10408340590947934

[5] ICH harmonised tripartite guideline prepared within the third international conference on harmonisation of technical requirements for the registration of pharmaceuticals for human use (ICH), Text on Validation of Analytical Procedures, 1994.

[6] Y. Vander Heyden, A. Nijhuis, J. Smeyers-Verbeke, B.G.M. Vandeginste, D.L. Massart, Guidance for robustness/ruggedness tests in method validation, J. Pharmaceut. Biomed. 24 (2001) 723-753. DOI: 10.1016/S0731-7085(00)00529-X

[7] B. Dejaegher, Y. Vander Heyden, Ruggedness and robustness testing, J. Chromatogr. A 1158 (2007) 138-157. DOI: 10.1016/j.chroma.2007.02.086

[8] E. Karageorgou, V. Samanidou, Youden test application in robustness assays during method validation, J. Chromatogr. A 1353 (2014), 131-139. DOI: 10.1016/j.chroma.2014.01.050

[9] R.L. Plackett and J.P. Burman, The Design of Optimum Multifactorial Experiments, Biometrika 33 (1946) 305-325.

[10] J. Goupy, What kind of experimental design for finding and checking robustness of analytical methods?, Anal. Chim. Acta 544 (2005) 184-190. DOI: 10.1016/j.aca.2005.01.051

[11] K. MacNamara, M. Lee, A. Robbat Jr., Rapid gas chromatographic analysis of less abundant compounds in distilled spirits by direct injection with ethanol–water venting and mass spectrometric data deconvolution, J. Chromatogr. A, 1217 (2010) 136-142. DOI: 10.1016/j.chroma.2009.11.010

[12] F. David, B. Tienpont, C. Devos, O. Lerch, P. Sandra, Increasing productivity for the analysis of trace contaminants in food by gas chromatography–mass spectrometry using automated liner exchange, backflushing and heart-cutting, J. Chromatogr. A, 1313 (2013) 147- 156. DOI: 10.1016/j.chroma.2013.07.038

[13] O. Fiehn, G. Wohlgemuth, M. Scholz, T. Kind, D.Y. Lee, Y. Lu, S. Moon, B. Nikolau, Quality control for plant metabolomics: reporting MSI-compliant studies, The Plant Journal 53 (2008) 691-704. DOI: 10.1111/j.1365-313X.2007.03387.x

[14] W. Zou, V.V. Tolstikov, Probing genetic algorithms for feature selection in comprehensive metabolic profiling approach, Rapid Commun. Mass Spectrom. 22 (2008) 1312-1324. DOI: 10.1002/rcm.3507

[15] W.F. Smith, Experimental design for formulation, ASA-SIAM Series on Statistical and Applied Probability, SIAM, Philadelphia, ASA, Alexandria, VA, 2005.

[16] G.E.P. Box, N.R. Draper, Response surfaces, mixtures, and ridge analysis, 2nd edition, John Wiley & Sons Inc., Hoboken 2007.

[17] R.H. Myers, D.C. Montgomery, Response surface methodology, process and product optimization using designed experiments, Wiley, NY, 2nd ed., 2002

[18] M.S. Sánchez, M.C. Ortiz, L.A. Sarabia, Selection of nearly orthogonal blocks in 'ad-hoc' experimental designs, Chemometr. Intell. Lab. 133 (2014) 109-120. DOI: 10.1016/j.chemolab.2013.12.008

[19] A. Herrero, C. Reguera, M.C. Ortiz, L.A. Sarabia, Determination of dichlobenil and its major metabolite (BAM) in onions by PTV-GC-MS using PARAFAC2 and experimental design methodology, Chemometr. Intell. Lab. 133 (2014) 92-108. DOI: 10.1016/j.chemolab.2013.12.001

[20] H.A.L. Kiers, J.M.F. ten Berge, R. Bro, PARAFAC2–Part I. A direct fitting algorithm for the PARAFAC2 model, J. Chemometrics 13 (1999) 275-294. DOI: 10.1002/(SICI)1099-128X(199905/08)13:3/4<275::AID-CEM543>3.0.CO;2-B

[21] R. Bro, C.A. Andersson, H.A.L. Kiers, PARAFAC2–Part II. Modeling chromatographic data with retention time shifts, J. Chemometrics 13 (1999) 295-309. DOI: 10.1002/(SICI)1099-128X(199905/08)13:3/4<295::AID-CEM547>3.0.CO;2-Y

[22] M.C. Ortiz, L.A. Sarabia, Quantitative determination in chromatographic analysis based on n-way calibration strategies, J. Chromatogr. A 1158 (2007) 94-110. DOI: 10.1016/j.chroma.2007.04.047

[23] J.M. Amigo, T. Skov, R. Bro, J. Coello, S. Maspoch, Solving GC-MS problems with PARAFAC2, Trends Anal. Chem. 27 (2008) 714-725. DOI: 10.1016/j.trac.2008.05.011

[24] A. Herrero, M.C. Ortiz, L.A. Sarabia, D-optimal experimental design coupled with parallel factor analysis 2 decomposition a useful tool in the determination of triazines in oranges by programmed temperature vaporization-gas chromatography-mass spectrometry when using dispersive-solid phase extraction, J. Chromatogr. A 1288 (2013) 111-126. DOI: 10.1016/j.chroma.2013.02.088

[25] Document nº SANCO/12571/2013, Guidance document on analytical quality control and validation procedures for pesticide residues analysis in food and feed, EU, Brussels, 2013.

[26] L. A. Sarabia and M. C. Ortiz, Response Surface Methodology. in: S. Brown , R. Tauler, B. Walczak (Eds.), Comprehensive Chemometrics, volume 1, Elsevier, Oxford, 2009, pp. 345-390.

[27] M.S. Sánchez, L.A. Sarabia, M.C. Ortiz, On the construction of experimental designs for a given task by jointly optimizing several quality criteria: Pareto-optimal experimental designs, Anal. Chim. Acta 754 (2012) 39- 46. DOI: 10.1016/j.aca.2012.10.014

[28] L.A. Sarabia, M.S. Sánchez, M.C. Ortiz, Introduction to Ranking Methods, in: Manuela Pavan and Roberto Todeschini, editors, Data Handling in Science and Technology. Elsevier, 2008, p. 1

[29] K. Atkinson, Elementary numerical analysis, John Wiley & Sons, Inc., New York, 2nd ed., 1993.

[30] G. Tomasi, R. Bro, Multilinear Models: Iterative Methods, Response surface methodology, in: S.D. Brown, R. Tauler, B. Walczak (Eds.), Comprehensive Chemometrics, Vol. 2, Elsevier, Oxford, 2009, pp. 411-451.

[31] R. Bro, PARAFAC. Tutorial and applications, Chemom. Intell. Lab. Syst. 38 (1997) 149-171. DOI: 10.1016/S0169-7439(97)00032-4

[32] J.M.F. ten Berge, H.A.L Kiers. Some uniqueness results for PARAFAC2. Psychometrika 61 (1996) 123–132. DOI: 10.1007/BF02296962

[33] R. Bro, H.A.L. Kiers, A new efficient method for determining the number of components in PARAFAC models, J. Chemom. 17 (2003) 274-286. DOI: 10.1002/cem.801

[34] M. H. Kamstrup-Nielsen, L. G. Johnsen and R. Bro, Core consistency diagnostic in PARAFAC2, J. Chemometrics 27 (2013) 99–105. DOI: 10.1002/cem.2497

[35] D. Mathieu, J. Nony, R. Phan-Than-Luu, NemrodW (Version 2007_03), L.P.R.A.I. Marseille, France, 2007.

[36] B.M. Wise, N.B. Gallagher, R. Bro, J.M. Shaver, W. Windig, R.S. Koch, PLS Toolbox 5.8.2. Eigenvector Research Inc., Manson, WA, 2010.

[37] STATGRAPHICS Centurion XVI (version 16.1.11), StatPoint Technologies, Inc. Herndon, VA, 2010.

[38] M.C. Ortiz, L.A. Sarabia, I. García, D. Giménez, E. Meléndez, Capability of detection and three-way data, Anal. Chim. Acta 559 (2006) 124-136. DOI: 10.1016/j.aca.2005.11.069

Table 1 Correlation matrices among coefficient estimates for a Plackett-Burman with N = 24 experiments, 18 factors, 3 blocks. The lower triangular part corresponds to the blocking as 6+12+6; the upper triangular part is for the design blocked as 6+10+8.

Coeff.	b_{1A}	b_{2A}	b_{3A}	b_{4A}	b_{5A}	b_{6A}	b_{7A}	b_{8A}	b_{9A}	b_{10A}	b_{11A}	b_{12A}	b_{13A}	b_{14A}	b_{15A}	b_{16A}	b_{17A}	b_{18A}	d_1	d_2
b_{1A}		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
b_{2A}	0.00		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
b_{3A}	0.00	0.00		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
b_{4A}	0.00	0.00	0.00		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
b_{5A}	0.00	0.00	0.00	0.00		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
b_{6A}	0.00	0.00	0.00	0.00	0.00		0.00	-0.05	0.05	-0.05	0.00	0.00	0.00	0.00	0.00	-0.05	0.00	0.05	0.12	0.21
b_{7A}	0.00	0.00	0.00	0.00	0.00	0.00		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
b_{8A}	0.00	0.00	0.00	0.00	0.00	0.00	0.00		-0.05	0.05	0.00	0.00	0.00	0.00	0.00	0.05	0.00	-0.05	-0.12	-0.21
b_{9A}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		-0.05	0.00	0.00	0.00	0.00	0.00	-0.05	0.00	0.05	0.12	0.21
b_{10A}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		0.00	0.00	0.00	0.00	0.00	0.05	0.00	-0.05	-0.12	-0.21
b_{11A}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
b_{12A}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
b_{13A}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		0.00	0.00	0.00	0.00	0.00	0.00	0.00
b_{14A}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		0.00	0.00	0.00	0.00	0.00	0.00
b_{15A}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		0.00	0.00	0.00	0.00	0.00
b_{16A}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		0.00	-0.05	-0.12	-0.21
b_{17A}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		0.00	0.00	0.00
b_{18A}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		0.12	0.21
d_1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		0.54
d_2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.58**	

Table 2 Correlation matrices among coefficient estimates for the experimental designs with N = 21 experiments in the extremes of the Pareto-optimal front. The lower triangular part corresponds to the design with least correlation among coefficient estimates; the upper triangular part is for the design with the best D-value. Bold underlined numbers refer to the worst case when studying only the correlation among b_{iA}; those in bold italics correspond to the largest correlation between b_{iA} and d_j (j = 1, 2, i = 1, …, 18,) related to the 'orthogonal blocking'.

Coeff.	b_{1A}	b_{2A}	b_{3A}	b_{4A}	b_{5A}	b_{6A}	b_{7A}	b_{8A}	b_{9A}	b_{10A}	b_{11A}	b_{12A}	b_{13A}	b_{14A}	b_{15A}	b_{16A}	b_{17A}	b_{18A}	d_1	d_2
b_{1A}		-0.06	-0.04	0.05	-0.02	0.14	-0.09	-0.01	0.09	-0.14	0.00	0.07	-0.01	0.00	0.16	0.05	-0.08	-0.06	0.10	0.19
b_{2A}	0.03		0.15	0.04	0.16	-0.06	-0.05	-0.03	0.06	0.05	0.00	0.05	0.14	0.06	-0.01	-0.11	-0.03	-0.05	0.04	0.02
b_{3A}	-0.14	-0.05		0.07	-0.06	0.15	0.09	-0.03	0.05	0.07	0.21	0.05	**0.21**	0.08	0.14	0.08	0.17	-0.08	-0.05	-0.06
b_{4A}	0.02	0.08	0.08		0.03	0.04	-0.14	0.02	-0.05	-0.05	0.05	-0.03	0.09	0.00	0.07	-0.04	0.06	0.03	-0.17	-0.05
b_{5A}	-0.14	-0.02	-0.07	-0.11		-0.08	-0.05	0.18	0.08	0.07	-0.04	0.08	-0.01	0.02	0.01	0.05	-0.08	-0.01	0.10	0.20
b_{6A}	0.03	-0.01	0.05	-0.03	-0.13		-0.08	-0.06	0.07	0.01	0.19	0.10	0.02	0.07	-0.02	0.04	0.00	-0.07	0.10	0.01
b_{7A}	0.16	0.06	0.02	0.15	0.00	0.15		0.11	0.02	0.06	-0.06	-0.14	-0.07	-0.13	-0.02	0.06	-0.04	-0.02	-0.07	-0.07
b_{8A}	0.00	-0.09	-0.09	-0.07	0.12	0.05	-0.12		0.10	0.07	0.11	0.06	0.00	0.01	0.17	0.11	-0.08	0.14	0.11	***0.22***
b_{9A}	0.10	0.08	0.01	0.11	0.16	0.10	0.01	-0.10		-0.05	0.08	0.17	0.09	-0.06	0.08	-0.02	-0.17	0.02	0.14	0.17
b_{10A}	0.07	-0.05	-0.06	0.02	-0.15	0.06	-0.12	-0.01	0.08		-0.10	-0.04	0.06	0.14	0.07	0.14	0.04	0.05	0.02	0.04
b_{11A}	0.11	-0.13	0.08	0.14	-0.16	-0.06	-0.02	-0.17	-0.02	0.04		0.12	0.16	0.06	0.01	0.06	0.15	-0.06	0.18	0.08
b_{12A}	0.07	-0.08	0.13	0.01	-0.11	0.11	0.01	0.15	0.06	0.01	0.02		0.07	-0.04	-0.09	0.13	0.03	0.04	0.12	0.13
b_{13A}	-0.03	0.04	0.02	-0.04	0.07	-0.09	-0.02	-0.01	-0.05	-0.17	0.11	-0.14		0.07	0.19	0.05	-0.01	-0.07	-0.01	0.11
b_{14A}	0.06	-0.01	-0.08	0.05	0.14	0.01	0.03	0.17	0.07	0.00	-0.02	-0.03	0.04		0.04	-0.03	0.08	0.04	-0.05	-0.13
b_{15A}	0.15	0.09	0.01	0.15	0.02	-0.06	0.17	-0.16	0.01	0.03	0.02	-0.17	0.13	0.08		0.08	-0.06	-0.05	0.07	0.21
b_{16A}	0.16	-0.13	-0.15	-0.11	0.04	0.15	0.15	0.10	-0.05	-0.14	-0.04	0.07	0.11	-0.08	-0.02		0.07	0.06	-0.03	0.06
b_{17A}	0.05	-0.02	-0.15	-0.04	0.07	0.00	-0.03	0.06	0.12	0.08	0.11	0.03	-0.03	0.06	0.11	0.14		-0.04	-0.10	-0.14
b_{18A}	0.04	**0.17**	-0.05	0.06	0.00	-0.15	-0.10	-0.02	0.06	-0.07	-0.14	0.08	0.06	0.11	0.03	-0.03	-0.14		-0.08	-0.07
d_1	-0.04	0.11	-0.02	0.00	-0.03	-0.08	0.02	-0.13	0.17	0.02	0.07	0.02	***0.17***	-0.07	-0.04	-0.09	-0.08	0.10		0.63
d_2	-0.05	0.13	-0.14	0.11	0.03	-0.09	0.11	-0.05	0.08	0.05	-0.03	-0.11	0.07	0.03	-0.02	-0.10	-0.09	0.11	0.60	

Table 3 Diagnostic ions (the base peak is in bold), relative abundances and tolerance intervals estimated from the reference standards and relative abundances calculated from the loadings of the spectral mode of the PARAFAC2 models built for DIC, BAM, ISDIC and ISBAM.

Compound	Ion	Reference standards		Loadings of PARAFAC2 models	
		Relative abundance (%)	Tolerance interval (%)	Relative abundance (%)	Verified compliance
DIC	100	22.80	(19.38, 26.22)	24.98	yes
	136	19.22	(15.38, 23.07)	18.93	yes
	171	**100.00**	-	-	-
	173	64.01	(57.61, 70.41)	62.06	yes
	175	10.15	(8.12, 12.18)	10.91	yes
BAM	136	9.99	(4.99, 14.98)	5.65	yes
	173	24.68	(20.97, 28.38)	17.05	no
	175	13.37	(10.69, 16.04)	11.03	yes
	246	**100.00**	-	-	-
	248	68.50	(61.65, 75.35)	69.54	yes
ISDIC	100	23.01	(19.56, 26.46)	19.53	no
	136	19.64	(15.71, 23.57)	21.35	yes
	171	**100.00**	-	-	-
	173	63.10	(56.79, 69.41)	63.45	yes
	175	9.98	(4.99, 14.97)	9.72	yes
ISBAM	136	9.46	(7.57, 11.35)	5.96	no
	173	13.98	(11.19, 16.78)	15.12	yes
	175	7.00	(3.50, 10.50)	9.93	yes
	246	**100.00**	-	-	-
	248	66.77	(60.10, 73.45)	70.12	yes

Table 4 Experimental plan for the robustness study. Factors: x_1 (D_{Temp}), x_2 (D_{Time}), x_3 (V_{BSTFA}), x_4 (t_{mix1}), x_5 (t_{centr1}), x_6 (s_{centr1}), x_7 (t_{mix2}), x_8 (t_{centr2}), x_9 (T_{evap}), x_{10} (t_{evap}), x_{11} ($T_{PTVinit}$), x_{12} ($t_{PTVinit}$), x_{13} (P_{init}), x_{14} ($vent_{flow}$), x_{15} ($vent_{time}$), x_{16} (r_{PTV}), x_{17} (T_{PTVend}), x_{18} (s_{inj}); x_{b1} and x_{b2} are the block variables. The responses are in the last two columns that contain the standardized loadings of the corresponding compound.

Run	Derivatization			Extraction							Injection								Block		Responses	
	D_{Temp}	D_{Time}	V_{BSTFA}	t_{mix1}	t_{centr1}	s_{centr1}	t_{mix2}	t_{centr2}	T_{evap}	t_{evap}	$T_{PTVinit}$	$t_{PTVinit}$	P_{init}	$vent_{flow}$	$vent_{time}$	r_{PTV}	T_{PTVend}	s_{inj}	x_{b1}	x_{b2}	BAM	DIC
	°C	min	µL	min	min	rpm	s	min	°C	min	°C	min	psi	mL min^{-1}	min	°C s^{-1}	°C	µL s^{-1}				
1	44.4	42.1	54	1.9	9.9	2900	33	0.9	49	9.9	41	0.51	8.95	102	0.305	10.2	279	48	1	0	1.12	2.09
2	44.4	41.9	54	1.9	10.1	3100	27	0.9	49	10.1	41	0.51	9.05	98	0.295	9.8	281	52	1	0	1.24	2.57
3	44.6	42.1	58	2.1	10.1	2900	27	0.9	51	10.1	40	0.51	8.95	102	0.295	10.2	281	48	1	0	1.08	3.26
4	44.4	41.9	58	2.1	9.9	3100	33	1.1	51	9.9	40	0.51	9.05	98	0.305	9.8	281	52	1	0	1.19	4.56
5	44.6	41.9	58	1.9	9.9	3100	27	1.1	51	10.1	41	0.49	9.05	102	0.305	10.2	279	48	1	0	1.22	4.19
6	44.6	42.1	54	1.9	9.9	2900	33	1.1	51	10.1	40	0.49	9.05	98	0.295	9.8	279	52	1	0	1.11	3.75
7	44.6	41.9	54	2.1	9.9	2900	33	1.1	49	10.1	40	0.51	9.05	98	0.295	10.2	281	48	0	1	1.37	3.70
8	44.6	42.1	58	1.9	10.1	2900	33	1.1	49	10.1	41	0.51	9.05	102	0.305	9.8	281	52	0	1	1.33	4.85
9	44.4	41.9	54	1.9	9.9	2900	27	1.1	51	9.9	40	0.49	8.95	102	0.305	10.2	281	52	0	1	1.43	5.01
10	44.6	41.9	54	1.9	10.1	3100	33	0.9	51	9.9	40	0.51	9.05	102	0.295	9.8	279	48	0	1	1.35	4.70
11	44.4	42.1	54	2.1	9.9	3100	33	0.9	51	10.1	41	0.49	9.05	102	0.295	10.2	281	52	0	1	1.16	2.79
12	44.4	42.1	58	1.9	9.9	2900	27	0.9	51	10.1	40	0.51	9.05	98	0.305	10.2	279	48	0	1	1.43	2.83
13	44.4	41.9	54	2.1	10.1	2900	33	1.1	51	10.1	41	0.49	8.95	98	0.305	9.8	279	48	0	1	1.14	2.23
14	44.4	42.1	58	1.9	9.9	3100	33	1.1	49	10.1	40	0.49	8.95	102	0.295	9.8	281	48	0	1	1.45	4.74
15	44.6	41.9	54	2.1	9.9	3100	27	0.9	49	10.1	40	0.51	8.95	102	0.305	9.8	279	52	0	1	1.20	3.67
16	44.6	42.1	58	2.1	10.1	3100	27	1.1	49	9.9	41	0.49	8.95	98	0.295	10.2	279	52	0	1	1.64	2.76
17	44.6	42.1	54	1.9	9.9	3100	27	1.1	51	9.9	41	0.51	8.95	98	0.305	9.8	281	48	0	0	1.11	2.28
18	44.4	41.9	58	1.9	9.9	2900	33	1.1	51	10.1	41	0.51	8.95	102	0.295	10.2	279	52	0	0	1.04	2.40
19	44.4	42.1	54	2.1	10.1	3100	27	1.1	49	10.1	40	0.51	9.05	102	0.305	10.2	279	48	0	0	1.34	3.40
20	44.6	41.9	58	1.9	10.1	3100	33	0.9	49	10.1	40	0.49	8.95	98	0.305	10.2	281	52	0	0	1.10	3.60
21	44.6	41.9	58	2.1	9.9	2900	27	0.9	49	9.9	40	0.49	9.05	102	0.295	9.8	281	48	0	0	1.06	2.07

Table 5 Standardized and corrected (by subtracting the block effect) loadings of the sample mode of the PARAFAC2 models and calculated concentration for replicates. Mean recovery rates and the semi-length of their 95% confidence intervals.

Compound	Replicate	Standardized loading	Corrected standardized loading	Calculated concentration ($\mu g\ L^{-1}$)	Recovery rate (%)	Mean recovery (%)
BAM	1 (block 3)	1.1647	1.1647	18.62	93.09	91.47 ± 4.66
	2 (block 3)	1.2132	1.2132	19.23	96.15	
	3 (block 3)	1.1323	1.1323	18.21	91.05	
	4 (block 3)	1.1348	1.1348	18.24	91.20	
	5 (block 3)	1.0500	1.0500	17.17	85.86	
	1 (block 2)	1.3448	1.0726	17.46	87.29	80.43 ± 7.70
	2 (block 2)	1.2834	1.0112	16.68	83.41	
	3 (block 2)	1.2786	1.0064	16.62	83.11	
	4 (block 2)	1.1758	0.9036	15.33	76.63	
	5 (block 2)	1.0975	0.8253	14.34	71.70	
DIC	1 (block 3)	2.1339	2.1339	23.13	115.64	129.45 ± 12.31
	2 (block 3)	2.6132	2.6132	25.95	129.73	
	3 (block 3)	2.4971	2.4971	25.26	126.32	
	4 (block 3)	3.0632	3.0632	28.59	142.96	
	5 (block 3)	2.7103	2.7103	26.52	132.58	
	1 (block 2)	3.5565	2.4009	24.70	123.49	117.15 ± 18.21
	2 (block 2)	3.8640	2.7084	26.51	132.53	
	3 (block 2)	3.6345	2.4789	25.16	125.78	
	4 (block 2)	2.6567	1.5011	19.41	97.04	
	5 (block 2)	2.9919	1.8364	21.38	106.90	

FIGURE CAPTIONS

Fig. 1 Pareto-optimal front of the criteria being optimized for looking for a blocked design with 21 experiments for the robustness study.

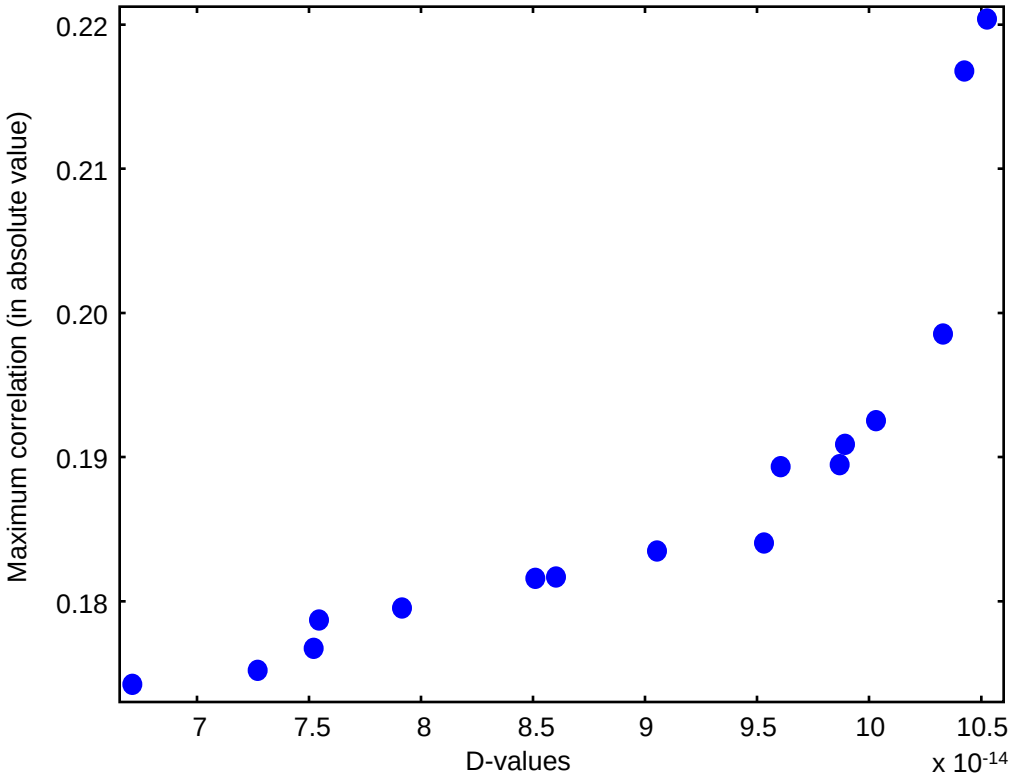
Fig. 2 Bar chart of the VIFs of the coefficient estimates of model in Eq. (1). Light red bars are for the design with the least correlation values; dark blue bars are for the design with the largest D-value in the Pareto-optimal front.

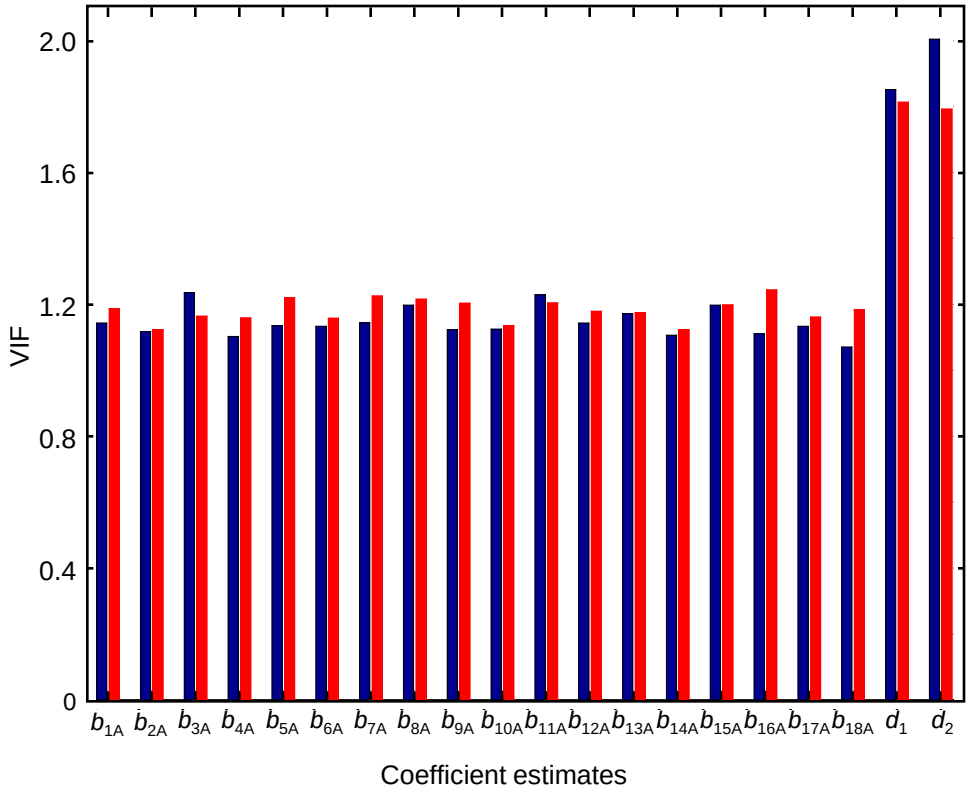
Fig. 3 Loadings of the chromatographic (a and d), spectral (b and e), and sample (c and f) modes of the PARAFAC2 models built for BAM (a, b and c) and ISBAM (d, e and f). Chromatographic loadings are scaled loadings. First factor is in blue continuous line (blue solid bars in the spectral mode and blue triangles in the sample mode), second factor is in green dashed line (green dashed bars in the spectral mode and green circles in the sample mode), and third factor is in red dotted line (red pointed bars in the spectral mode and red squares in the sample mode).

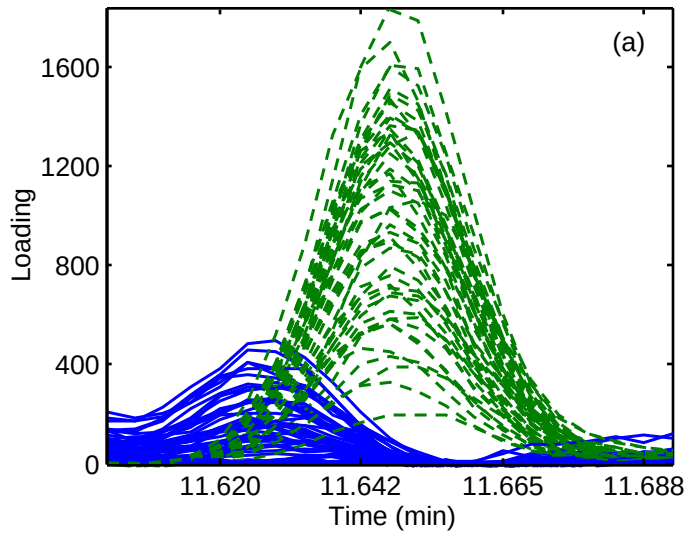
Fig. 4 Standardized loadings of the sample mode for BAM (a) and DIC (b).

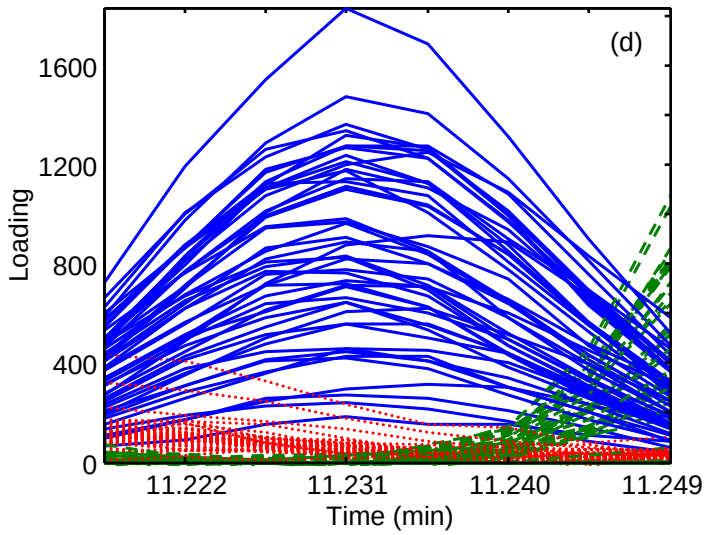
Fig. 5 Graphical analysis of the effects of the studied experimental factors on the responses. Light orange bars are for significant coefficients (5% significant level); dark blue bars are for the non-significant ones.

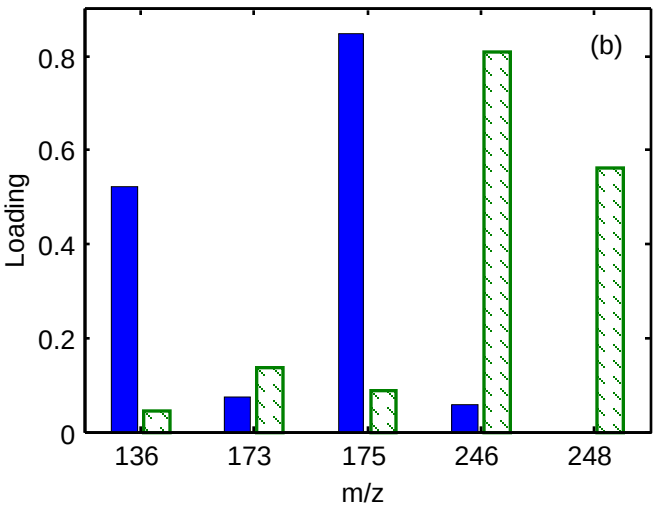
Fig. 6 Proportion of times that the corresponding coefficient (abscissa axis) is declared non-significant at 5% significance level. Twenty blocked designs randomly extracted from a three block Plackett-Burman design with 24 experiments.

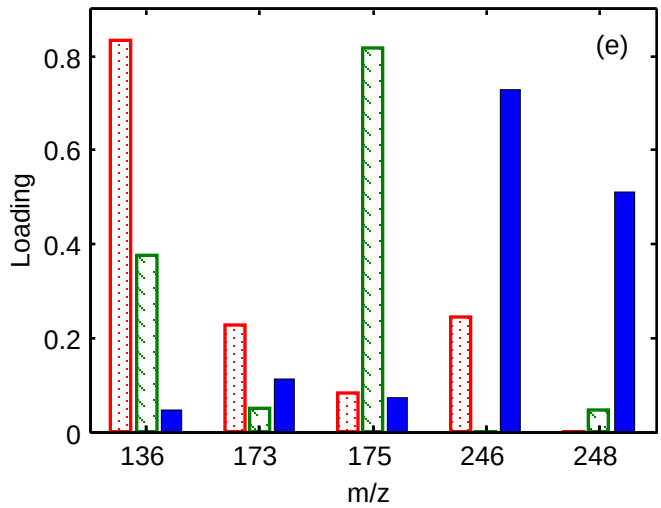

Maximum correlation (in absolute value)
0.22
0.21
0.20
0.19
0.18
7
7.5
8
8.5
9
9.5
10
10.5
D-values
x 10-14

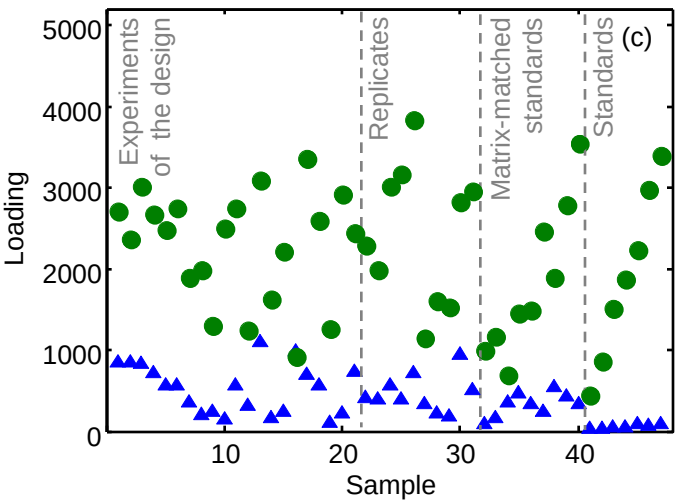

2.0
1.6
1.2
0.8
0.4
0
VIF
b_{1A} b_{2A} b_{3A} b_{4A} b_{5A} b_{6A} b_{7A} b_{8A} b_{9A} b_{10A} b_{11A} b_{12A} b_{13A} b_{14A} b_{15A} b_{16A} b_{17A} b_{18A} d_1 d_2
Coefficient estimates

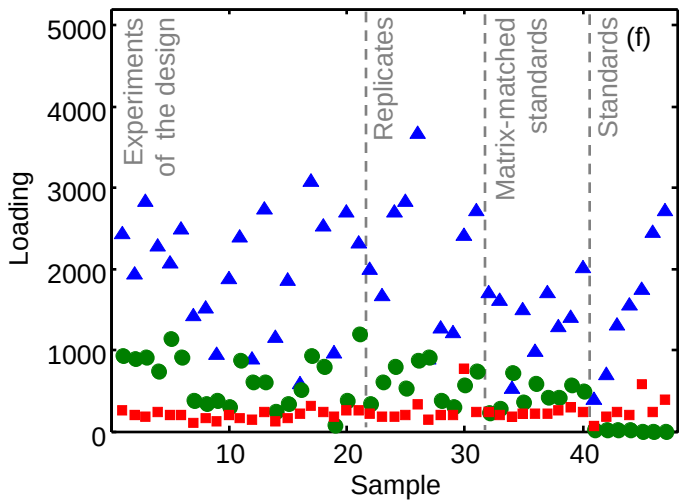

(a)
Loading
1600
1200
800
400
0
11.620
11.642
11.665
11.688
Time (min)

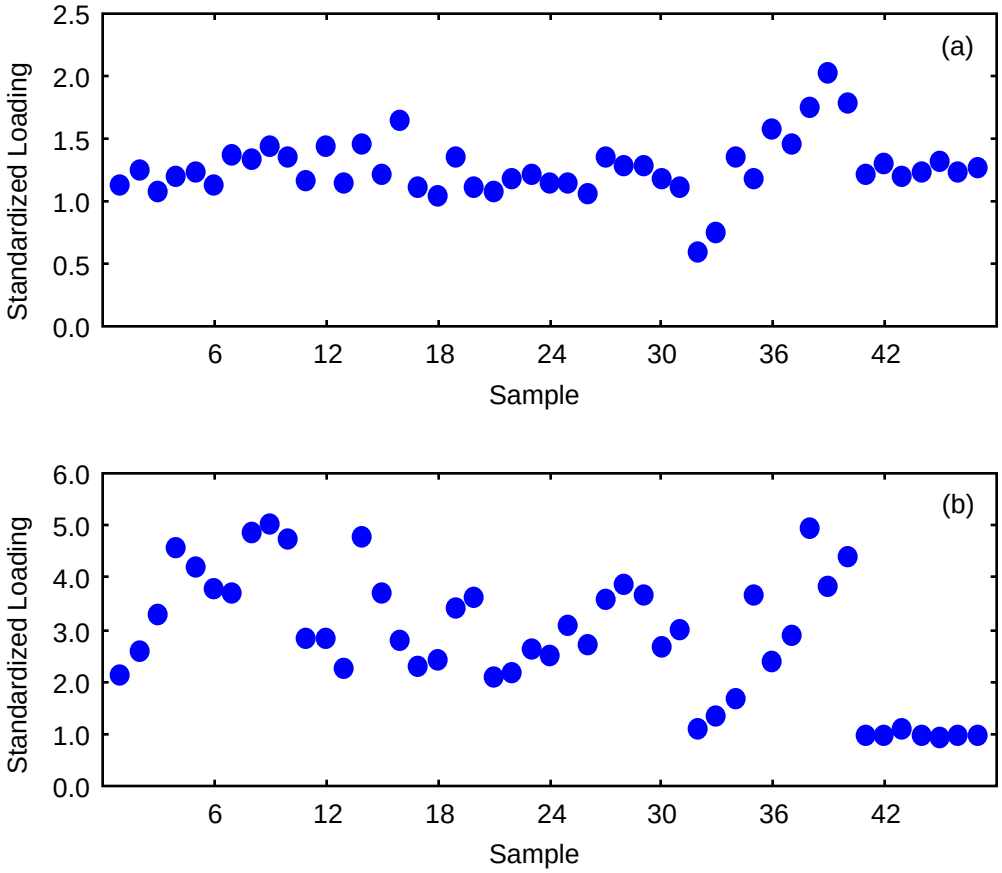

(d)
Loading
1600
1200
800
400
0
11.222
11.231
11.240
11.249
Time (min)

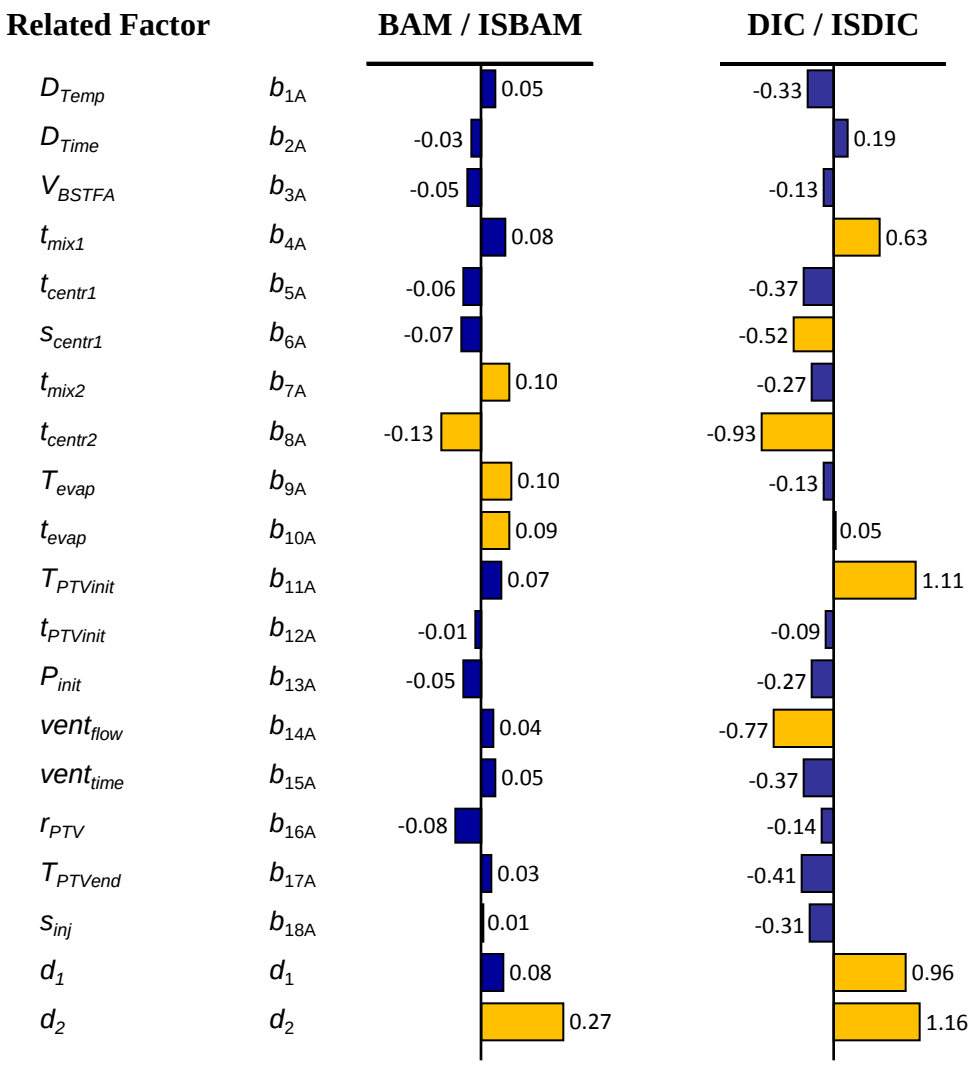

(b)
Loading
0.8
0.6
0.4
0.2
0
136
173
175
246
248
m/z


(e)
Loading
0.8
0.6
0.4
0.2
0
136
173
175
246
248
m/z


(c)
Experiments of the design
Replicates
Matrix-matched standards
Standards
Loading
5000
4000
3000
2000
1000
0
10
20
30
40
Sample


(f)
Experiments of the design
Replicates
Matrix-matched standards
Standards
Loading
5000
4000
3000
2000
1000
0
10
20
30
40
Sample


(a)
Standardized Loading
2.5
2.0
1.5
1.0
0.5
0.0
6
12
18
24
30
36
42
Sample
(b)
Standardized Loading
6.0
5.0
4.0
3.0
2.0
1.0
0.0
6
12
18
24
30
36
42
Sample


Related Factor
BAM / ISBAM
DIC / ISDIC
D_{Temp} b_{1A} 0.05 -0.33
D_{Time} b_{2A} -0.03 0.19
V_{BSTFA} b_{3A} -0.05 -0.13
t_{mix1} b_{4A} 0.08 0.63
t_{centr1} b_{5A} -0.06 -0.37
s_{centr1} b_{6A} -0.07 -0.52
t_{mix2} b_{7A} 0.10 -0.27
t_{centr2} b_{8A} -0.13 -0.93
T_{evap} b_{9A} 0.10 -0.13
t_{evap} b_{10A} 0.09 0.05
$T_{PTVinit}$ b_{11A} 0.07 1.11
$t_{PTVinit}$ b_{12A} -0.01 -0.09
P_{init} b_{13A} -0.05 -0.27
$vent_{flow}$ b_{14A} 0.04 -0.77
$vent_{time}$ b_{15A} 0.05 -0.37
r_{PTV} b_{16A} -0.08 -0.14
T_{PTVend} b_{17A} 0.03 -0.41
s_{inj} b_{18A} 0.01 -0.31
d_1 d_1 0.08 0.96
d_2 d_2 0.27 1.16

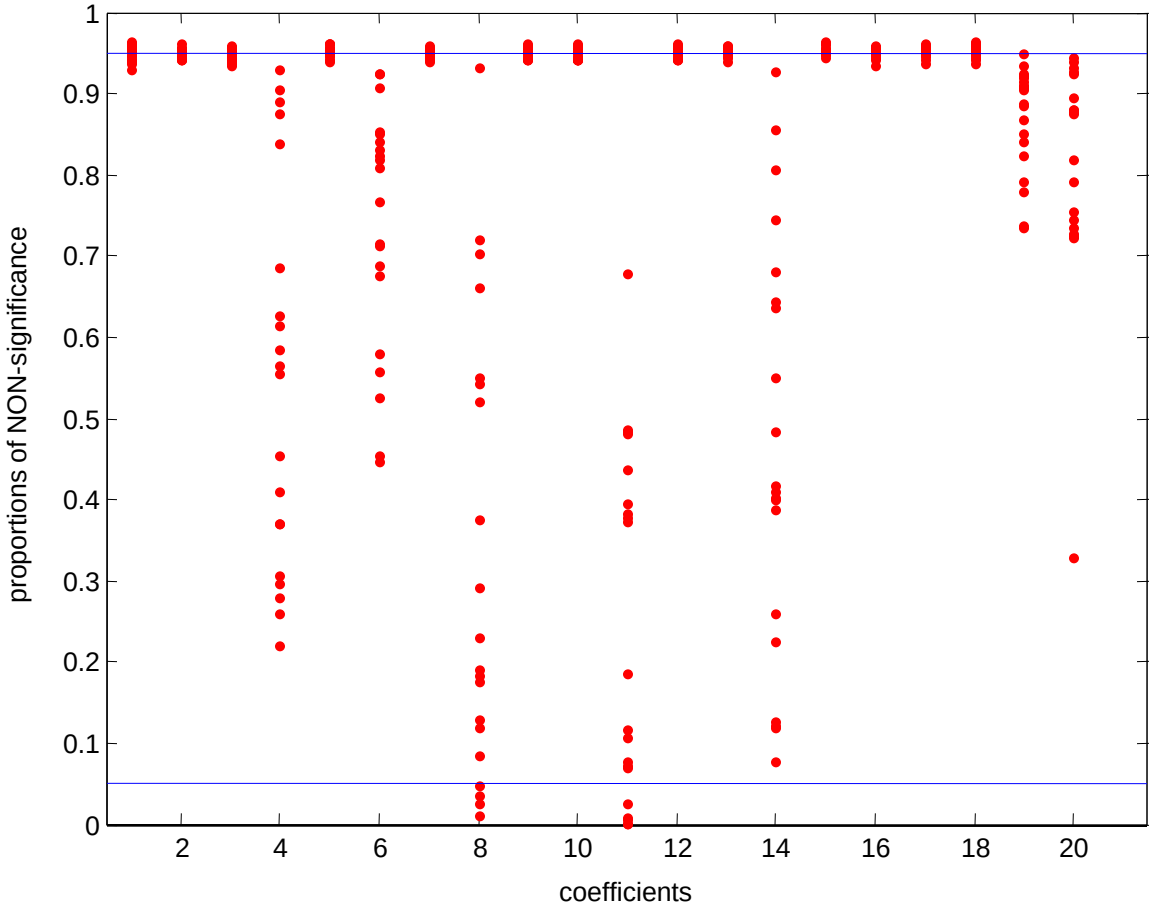

proportions of NON-significance
coefficients
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
2
4
6
8
10
12
14
16
18
20

Table S1 Properties of the designs in the corresponding Pareto-optimal fronts (increasing order of D): *N*, number of experiments in the design; maximum of the absolute value of the correlation coefficients among 'main' coefficients, $\left|corr\left(b_{iA}, b_{jA}\right)\right|$, and with block, $\left|corr\left(b_{iA}, d_j\right)\right|$; VIFs of the corresponding coefficient estimates.

N	Maximum values of		VIF																		
	$\left\|corr\left(b_{iA}, b_{jA}\right)\right\|$	$\left\|corr\left(b_{iA}, d_j\right)\right\|$	b_{1A}	b_{2A}	b_{3A}	b_{4A}	b_{5A}	b_{6A}	b_{7A}	b_{8A}	b_{9A}	b_{10A}	b_{11A}	b_{12A}	b_{13A}	b_{14A}	b_{15A}	b_{16A}	b_{17A}	b_{18A}	d_1	d_2
21	0.17	0.17	1.19	1.12	1.16	1.16	1.22	1.16	1.23	1.22	1.20	1.14	1.20	1.18	1.18	1.12	1.20	1.24	1.16	1.18	1.81	1.79
21	0.18	0.17	1.26	1.09	1.17	1.21	1.15	1.08	1.17	1.24	1.13	1.21	1.18	1.22	1.21	1.14	1.15	1.26	1.16	1.14	1.81	1.92
21	0.18	0.17	1.19	1.20	1.15	1.16	1.21	1.12	1.13	1.18	1.13	1.10	1.30	1.19	1.19	1.13	1.16	1.13	1.18	1.16	1.90	1.79
21	0.17	0.18	1.17	1.13	1.17	1.15	1.13	1.14	1.26	1.20	1.19	1.11	1.17	1.17	1.18	1.10	1.20	1.25	1.16	1.20	1.81	1.82
21	0.17	0.18	1.28	1.09	1.16	1.18	1.16	1.08	1.16	1.20	1.15	1.20	1.18	1.12	1.19	1.12	1.16	1.26	1.16	1.11	1.82	1.93
21	0.18	0.18	1.22	1.12	1.16	1.16	1.20	1.20	1.20	1.12	1.19	1.14	1.16	1.16	1.14	1.19	1.11	1.15	1.09	1.20	1.92	1.83
21	0.18	0.18	1.23	1.14	1.23	1.11	1.14	1.16	1.12	1.18	1.15	1.13	1.23	1.19	1.23	1.19	1.13	1.15	1.13	1.17	1.81	1.86
21	0.18	0.18	1.15	1.13	1.18	1.12	1.11	1.15	1.18	1.13	1.15	1.13	1.21	1.13	1.17	1.16	1.18	1.12	1.16	1.17	1.93	1.85
21	0.18	0.18	1.16	1.14	1.15	1.13	1.13	1.12	1.14	1.13	1.19	1.12	1.23	1.20	1.14	1.18	1.16	1.15	1.21	1.13	1.83	1.91
21	0.19	0.17	1.12	1.16	1.23	1.12	1.11	1.13	1.12	1.19	1.14	1.14	1.25	1.24	1.16	1.13	1.19	1.12	1.13	1.11	1.82	1.84
21	0.19	0.18	1.14	1.17	1.17	1.13	1.11	1.12	1.14	1.13	1.17	1.14	1.22	1.20	1.12	1.16	1.18	1.12	1.20	1.12	1.80	1.91
21	0.19	0.18	1.16	1.13	1.23	1.08	1.11	1.16	1.13	1.20	1.19	1.16	1.20	1.13	1.23	1.14	1.17	1.13	1.17	1.07	1.78	1.88
21	0.19	0.18	1.16	1.24	1.20	1.14	1.09	1.17	1.15	1.10	1.21	1.10	1.18	1.18	1.19	1.19	1.18	1.11	1.13	1.09	1.82	1.96
21	0.20	0.18	1.14	1.11	1.26	1.07	1.12	1.16	1.15	1.17	1.16	1.18	1.15	1.16	1.21	1.14	1.16	1.15	1.17	1.10	1.79	1.92
21	0.22	0.21	1.14	1.10	1.24	1.11	1.13	1.13	1.15	1.16	1.12	1.14	1.23	1.14	1.17	1.11	1.21	1.11	1.14	1.09	1.84	1.92
21	0.21	0.22	1.14	1.12	1.24	1.10	1.14	1.13	1.14	1.20	1.12	1.12	1.23	1.14	1.17	1.11	1.20	1.11	1.13	1.07	1.85	2.00
22	0.16	0.16	1.17	1.21	1.22	1.14	1.17	1.16	1.15	1.19	1.13	1.17	1.13	1.12	1.14	1.16	1.21	1.13	1.19	1.15	1.58	1.54
22	0.16	0.16	1.16	1.21	1.22	1.14	1.17	1.15	1.14	1.17	1.13	1.17	1.13	1.05	1.14	1.15	1.24	1.13	1.18	1.14	1.60	1.54
22	0.16	0.16	1.05	1.20	1.20	1.12	1.15	1.13	1.14	1.14	1.11	1.17	1.13	1.09	1.12	1.16	1.25	1.16	1.13	1.18	1.69	1.56
22	0.17	0.16	1.05	1.16	1.15	1.12	1.16	1.19	1.14	1.06	1.11	1.14	1.21	1.11	1.13	1.17	1.17	1.13	1.14	1.13	1.55	1.55
22	0.17	0.15	1.08	1.14	1.08	1.15	1.19	1.17	1.14	1.11	1.08	1.12	1.18	1.08	1.15	1.19	1.12	1.14	1.11	1.16	1.60	1.60

22	0.18	0.17	1.10	1.17	1.11	1.12	1.13	1.17	1.14	1.19	1.09	1.13	1.18	1.12	1.10	1.17	1.15	1.15	1.11	1.14	1.60	1.54
22	0.18	0.15	1.11	1.19	1.09	1.08	1.14	1.17	1.10	1.07	1.09	1.20	1.18	1.14	1.11	1.16	1.16	1.11	1.10	1.16	1.64	1.55
22	0.18	0.15	1.07	1.13	1.06	1.09	1.19	1.15	1.12	1.08	1.10	1.14	1.20	1.09	1.10	1.18	1.15	1.14	1.12	1.14	1.59	1.56
22	0.18	0.13	1.08	1.14	1.07	1.12	1.18	1.17	1.14	1.10	1.09	1.13	1.19	1.09	1.10	1.17	1.13	1.15	1.11	1.15	1.57	1.57
22	0.18	0.17	1.09	1.17	1.10	1.16	1.17	1.16	1.13	1.09	1.08	1.12	1.17	1.11	1.12	1.13	1.16	1.15	1.09	1.17	1.60	1.56
22	0.18	0.15	1.07	1.16	1.08	1.09	1.16	1.18	1.11	1.06	1.09	1.14	1.19	1.12	1.09	1.16	1.16	1.13	1.10	1.15	1.61	1.55
22	0.19	0.17	1.08	1.17	1.08	1.12	1.15	1.18	1.13	1.08	1.08	1.13	1.17	1.12	1.09	1.15	1.14	1.15	1.10	1.16	1.60	1.55
22	0.20	0.19	1.10	1.13	1.08	1.06	1.12	1.10	1.16	1.09	1.06	1.11	1.16	1.10	1.07	1.15	1.15	1.14	1.07	1.14	1.64	1.56
22	0.21	0.18	1.10	1.15	1.09	1.08	1.15	1.18	1.13	1.07	1.07	1.12	1.14	1.09	1.08	1.14	1.14	1.12	1.09	1.17	1.68	1.56
22	0.20	0.21	1.10	1.14	1.09	1.08	1.15	1.18	1.13	1.07	1.08	1.12	1.13	1.09	1.08	1.10	1.14	1.12	1.08	1.17	1.66	1.62
22	0.21	0.22	1.09	1.13	1.08	1.09	1.13	1.19	1.14	1.06	1.09	1.13	1.12	1.08	1.09	1.11	1.14	1.09	1.10	1.14	1.66	1.61
23	0.15	0.16	1.14	1.16	1.15	1.11	1.20	1.15	1.10	1.11	1.12	1.13	1.14	1.10	1.20	1.14	1.15	1.16	1.14	1.09	1.55	1.49
23	0.16	0.16	1.15	1.16	1.11	1.10	1.22	1.13	1.14	1.04	1.13	1.12	1.13	1.09	1.20	1.14	1.12	1.17	1.13	1.07	1.52	1.51
23	0.16	0.16	1.14	1.18	1.13	1.07	1.20	1.13	1.11	1.05	1.13	1.13	1.14	1.08	1.18	1.12	1.15	1.11	1.16	1.11	1.56	1.55
23	0.16	0.16	1.17	1.14	1.13	1.10	1.20	1.10	1.11	1.05	1.11	1.13	1.14	1.08	1.20	1.12	1.13	1.15	1.16	1.09	1.53	1.52
23	0.16	0.16	1.18	1.13	1.12	1.15	1.15	1.09	1.12	1.09	1.12	1.15	1.11	1.07	1.22	1.15	1.11	1.12	1.14	1.14	1.46	1.52
23	0.16	0.16	1.10	1.07	1.13	1.12	1.10	1.11	1.17	1.08	1.13	1.10	1.15	1.07	1.15	1.11	1.15	1.08	1.16	1.10	1.53	1.56
23	0.17	0.16	1.11	1.07	1.13	1.12	1.09	1.12	1.18	1.11	1.10	1.11	1.15	1.07	1.17	1.10	1.13	1.10	1.14	1.08	1.53	1.53
23	0.18	0.19	1.15	1.07	1.11	1.11	1.11	1.11	1.13	1.12	1.10	1.08	1.13	1.05	1.18	1.12	1.16	1.16	1.15	1.10	1.54	1.53
23	0.18	0.19	1.13	1.07	1.13	1.12	1.10	1.11	1.13	1.11	1.11	1.09	1.14	1.07	1.18	1.12	1.13	1.14	1.12	1.10	1.54	1.54
23	0.17	0.20	1.10	1.10	1.15	1.12	1.08	1.15	1.13	1.08	1.08	1.10	1.06	1.10	1.17	1.11	1.14	1.13	1.14	1.08	1.60	1.60
23	0.20	0.17	1.10	1.08	1.11	1.12	1.09	1.10	1.16	1.11	1.13	1.11	1.13	1.09	1.17	1.10	1.11	1.12	1.17	1.06	1.54	1.54
23	0.21	0.18	1.15	1.06	1.12	1.16	1.10	1.14	1.10	1.08	1.11	1.13	1.16	1.09	1.11	1.14	1.16	1.10	1.12	1.08	1.51	1.54
23	0.23	0.17	1.10	1.09	1.12	1.09	1.13	1.08	1.14	1.07	1.16	1.10	1.13	1.10	1.15	1.10	1.11	1.10	1.16	1.08	1.53	1.56
23	0.23	0.19	1.15	1.04	1.12	1.15	1.11	1.12	1.11	1.12	1.11	1.08	1.13	1.07	1.16	1.11	1.14	1.16	1.14	1.09	1.54	1.56
23	0.25	0.19	1.14	1.05	1.13	1.13	1.16	1.12	1.09	1.12	1.11	1.09	1.14	1.07	1.14	1.09	1.15	1.14	1.13	1.09	1.54	1.56
24	0.15	0.15	1.10	1.10	1.07	1.07	1.08	1.18	1.08	1.07	1.13	1.16	1.14	1.14	1.19	1.08	1.20	1.09	1.08	1.12	1.50	1.40
24	0.16	0.15	1.10	1.07	1.06	1.08	1.07	1.18	1.08	1.07	1.12	1.15	1.14	1.15	1.17	1.09	1.19	1.08	1.08	1.11	1.48	1.40
24	0.16	0.14	1.08	1.07	1.07	1.07	1.08	1.20	1.07	1.07	1.13	1.17	1.15	1.14	1.17	1.08	1.17	1.07	1.07	1.11	1.47	1.39

24	0.17	0.16	1.07	1.07	1.09	1.07	1.09	1.18	1.06	1.09	1.14	1.19	1.14	1.15	1.16	1.07	1.19	1.07	1.10	1.08	1.45	1.41
24	0.17	0.15	1.08	1.08	1.09	1.06	1.09	1.21	1.06	1.09	1.14	1.18	1.16	1.13	1.17	1.08	1.18	1.07	1.07	1.11	1.44	1.39
24	0.18	0.17	1.07	1.22	1.06	1.07	1.08	1.06	1.09	1.20	1.20	1.08	1.07	1.18	1.04	1.06	1.13	1.05	1.14	1.21	1.37	1.48
24	0.18	0.17	1.06	1.10	1.06	1.18	1.21	1.05	1.09	1.13	1.06	1.05	1.06	1.13	1.10	1.05	1.17	1.06	1.13	1.10	1.38	1.54
24	0.18	0.18	1.17	1.17	1.06	1.06	1.09	1.04	1.09	1.16	1.06	1.05	1.04	1.10	1.06	1.05	1.21	1.05	1.13	1.14	1.37	1.52
24	0.18	0.18	1.05	1.09	1.05	1.05	1.18	1.04	1.09	1.15	1.05	1.05	1.05	1.14	1.10	1.05	1.14	1.04	1.15	1.08	1.38	1.53
24	0.18	0.18	1.04	1.18	1.03	1.05	1.07	1.04	1.08	1.15	1.03	1.05	1.04	1.15	1.03	1.05	1.15	1.04	1.16	1.09	1.37	1.53
24	0.19	0.19	1.03	1.20	1.03	1.03	1.08	1.05	1.15	1.19	1.03	1.05	1.03	1.08	1.03	1.03	1.20	1.05	1.07	1.08	1.39	1.57
24	0.19	0.19	1.03	1.08	1.03	1.03	1.17	1.03	1.16	1.19	1.03	1.03	1.03	1.18	1.03	1.03	1.07	1.03	1.07	1.08	1.38	1.57
24	0.18	0.19	1.03	1.07	1.03	1.03	1.07	1.03	1.07	1.15	1.03	1.04	1.03	1.15	1.03	1.03	1.16	1.03	1.15	1.07	1.38	1.58
24	0.14	0.20	1.04	1.15	1.04	1.04	1.08	1.04	1.15	1.13	1.04	1.04	1.04	1.08	1.04	1.04	1.15	1.04	1.08	1.07	1.39	1.59
24	0.12	0.20	1.03	1.13	1.03	1.03	1.12	1.03	1.13	1.08	1.03	1.03	1.03	1.08	1.03	1.03	1.08	1.03	1.07	1.13	1.38	1.59
24	0.11	0.21	1.03	1.12	1.03	1.03	1.12	1.03	1.12	1.12	1.03	1.03	1.03	1.08	1.03	1.03	1.07	1.03	1.08	1.07	1.39	1.60
24	0.18	0.21	1.02	1.07	1.03	1.03	1.07	1.02	1.07	1.14	1.03	1.02	1.02	1.17	1.03	1.03	1.07	1.02	1.18	1.07	1.40	1.65
24	0.18	0.21	1.02	1.07	1.03	1.03	1.14	1.02	1.07	1.07	1.03	1.02	1.02	1.17	1.03	1.03	1.07	1.02	1.18	1.07	1.40	1.65
24	0.18	0.21	1.03	1.07	1.02	1.02	1.14	1.03	1.07	1.07	1.02	1.03	1.03	1.17	1.02	1.02	1.07	1.03	1.18	1.07	1.40	1.65
24	0.14	0.21	1.02	1.07	1.02	1.02	1.17	1.02	1.07	1.07	1.03	1.03	1.02	1.15	1.02	1.02	1.07	1.02	1.14	1.07	1.39	1.64
24	0.13	0.21	1.03	1.15	1.03	1.03	1.07	1.03	1.08	1.14	1.03	1.03	1.03	1.07	1.03	1.03	1.15	1.03	1.07	1.06	1.41	1.66
24	0.13	0.21	1.03	1.15	1.03	1.03	1.07	1.03	1.15	1.14	1.03	1.03	1.03	1.07	1.03	1.03	1.08	1.03	1.07	1.06	1.41	1.66
24	0.13	0.21	1.03	1.07	1.03	1.03	1.15	1.03	1.15	1.14	1.03	1.03	1.03	1.07	1.03	1.03	1.07	1.03	1.07	1.06	1.41	1.66
24	0.11	0.21	1.02	1.07	1.02	1.02	1.07	1.02	1.13	1.13	1.02	1.02	1.02	1.13	1.02	1.02	1.07	1.02	1.07	1.07	1.40	1.66
24	0.14	0.21	1.01	1.06	1.01	1.01	1.16	1.01	1.06	1.06	1.02	1.02	1.01	1.16	1.01	1.01	1.06	1.01	1.06	1.06	1.40	1.71
24	0.05	0.21	1.00	1.05	1.00	1.00	1.05	1.00	1.05	1.05	1.00	1.00	1.00	1.05	1.00	1.00	1.05	1.00	1.05	1.05	1.45	1.88

Table S2. Proportion of times the corresponding coefficient estimate (in rows) was non-significant at 5% significance level. Results of applying the design in columns: Twenty designs randomly blocked from a Plackett-Burman design. The last two columns contain minimum (Min) and maximum (Max) values per coefficient, respectively. Shaded rows correspond to the non-null coefficients in the model

Designs/ Coeff. estimates	Dis1	Dis2	Dis3	Dis4	Dis5	Dis6	Dis7	Dis8	Dis9	Dis10	Dis11	Dis12	Dis13	Dis14	Dis15	Dis16	Dis17	Dis18	Dis19	Dis20	**Min**	**Max**
b_{1A}	0.96	0.96	0.95	0.94	0.96	0.96	0.95	0.95	0.94	0.94	0.94	0.95	0.96	0.96	0.93	0.94	0.96	0.95	0.95	0.95	0.93	0.96
b_{2A}	0.95	0.95	0.95	0.96	0.95	0.96	0.96	0.94	0.95	0.95	0.94	0.94	0.95	0.95	0.95	0.94	0.96	0.95	0.95	0.96	0.94	0.96
b_{3A}	0.96	0.94	0.95	0.95	0.95	0.95	0.95	0.94	0.95	0.95	0.94	0.94	0.95	0.96	0.94	0.96	0.94	0.96	0.94	0.95	0.94	0.96
b_{4A}	0.37	0.57	0.26	0.45	0.31	0.37	0.30	0.41	0.59	0.84	0.61	0.88	0.69	0.90	0.28	0.56	0.22	0.63	0.93	0.89	0.22	0.93
b_{5A}	0.96	0.95	0.94	0.96	0.96	0.95	0.95	0.95	0.95	0.94	0.95	0.95	0.96	0.95	0.96	0.94	0.95	0.95	0.95	0.95	0.94	0.96
b_{6A}	0.71	0.85	0.83	0.85	0.69	0.68	0.81	0.53	0.45	0.83	0.58	0.45	0.84	0.82	0.72	0.92	0.77	0.56	0.93	0.91	0.45	0.93
b_{7A}	0.95	0.94	0.96	0.95	0.95	0.96	0.95	0.95	0.96	0.95	0.95	0.96	0.96	0.95	0.94	0.94	0.95	0.95	0.95	0.95	0.94	0.96
b_{8A}	0.03	0.70	0.18	0.72	0.66	0.08	0.12	0.19	0.18	0.54	0.13	0.23	0.52	0.38	0.01	0.04	0.29	0.05	0.93	0.55	0.01	0.93
b_{9A}	0.96	0.95	0.95	0.95	0.95	0.96	0.94	0.95	0.95	0.95	0.94	0.94	0.95	0.96	0.95	0.96	0.96	0.95	0.95	0.95	0.94	0.96
b_{10A}	0.96	0.94	0.96	0.95	0.94	0.95	0.96	0.96	0.96	0.94	0.96	0.94	0.95	0.95	0.94	0.94	0.95	0.95	0.96	0.94	0.94	0.96
b_{11A}	0.48	0.37	0.44	0.40	0.00	0.49	0.07	0.07	0.48	0.11	0.03	0.38	0.08	0.38	0.01	0.00	0.01	0.19	0.12	0.68	0	0.68
b_{12A}	0.96	0.96	0.94	0.95	0.94	0.96	0.95	0.95	0.95	0.94	0.95	0.94	0.96	0.95	0.95	0.96	0.95	0.95	0.95	0.96	0.94	0.96
b_{13A}	0.95	0.95	0.95	0.95	0.94	0.96	0.95	0.96	0.96	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.96	0.96	0.96	0.95	0.94	0.96
b_{14A}	0.39	0.81	0.12	0.42	0.26	0.23	0.48	0.08	0.64	0.40	0.40	0.74	0.13	0.12	0.86	0.55	0.64	0.41	0.93	0.68	0.08	0.93
b_{15A}	0.96	0.95	0.96	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.96	0.96	0.95	0.96	0.96	0.95	0.96
b_{16A}	0.95	0.95	0.94	0.94	0.95	0.95	0.96	0.96	0.96	0.95	0.95	0.96	0.95	0.96	0.94	0.96	0.95	0.96	0.95	0.96	0.94	0.96
b_{17A}	0.96	0.95	0.95	0.95	0.95	0.95	0.95	0.96	0.95	0.94	0.94	0.95	0.95	0.96	0.95	0.96	0.95	0.95	0.95	0.95	0.94	0.96
b_{18A}	0.95	0.94	0.95	0.95	0.96	0.96	0.95	0.96	0.94	0.96	0.95	0.95	0.96	0.96	0.95	0.96	0.96	0.96	0.95	0.96	0.94	0.96
d_1	0.91	0.92	0.84	0.92	0.89	0.78	0.92	0.79	0.87	0.85	0.82	0.91	0.92	0.89	0.74	0.91	0.91	0.74	0.95	0.93	0.74	0.95
d_2	0.88	0.94	0.88	0.82	0.33	0.76	0.88	0.72	0.79	0.88	0.74	0.73	0.90	0.73	0.93	0.92	0.93	0.75	0.94	0.88	0.33	0.94