Elsevier Editorial System(tm) for Analytica Chimica Acta
Manuscript Draft

Manuscript Number: ACA-12-1367R1

Title: On the construction of experimental designs for a given task by jointly optimizing several quality criteria: Pareto-optimal experimental designs

Article Type: Full Length Article

Section/Category: CHEMOMETRICS

Keywords: Experimental design; design optimality; multicriteria optimization; Pareto-optimal front; Evolutionary algorithms

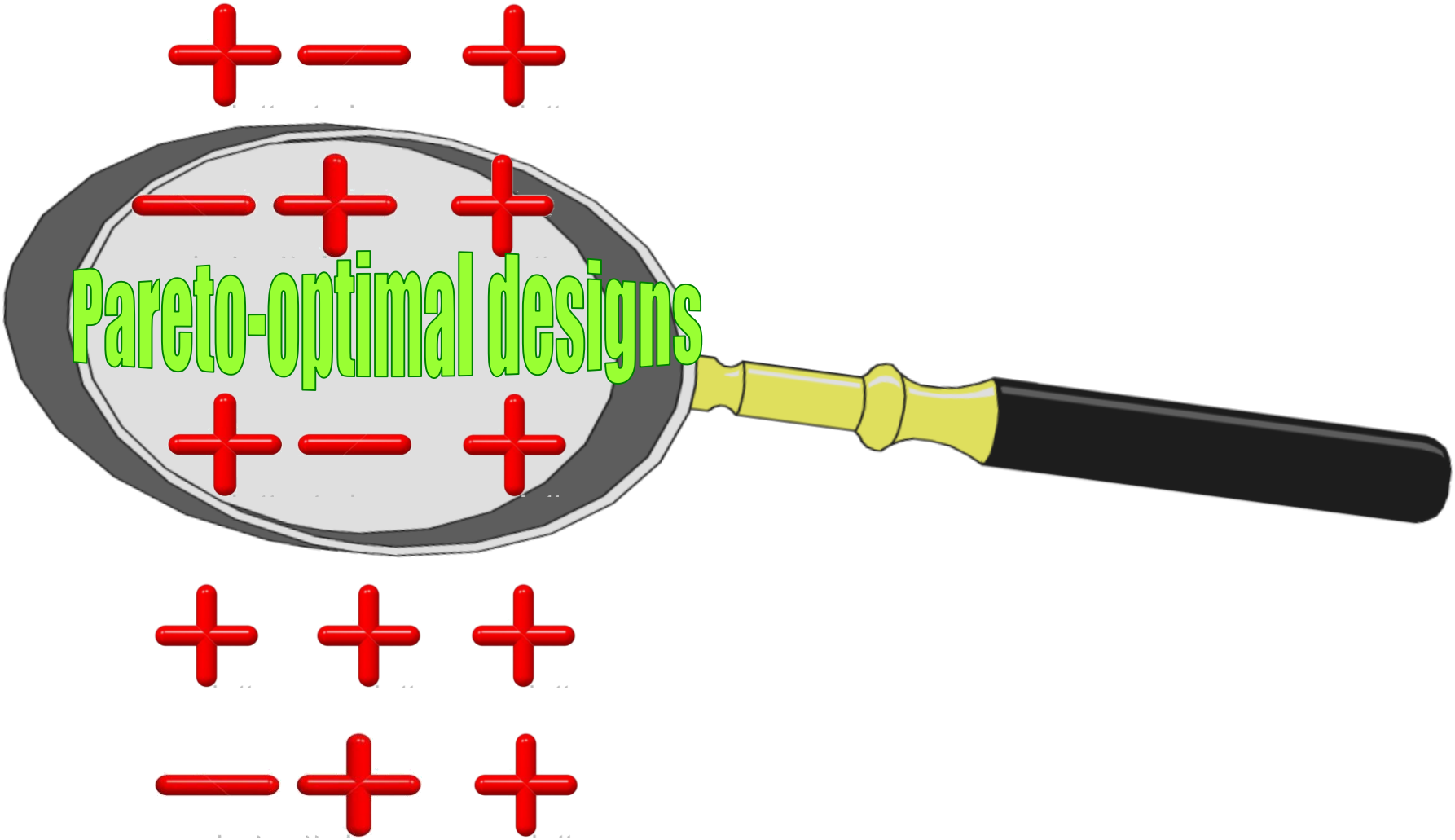
Corresponding Author: Dr. M.C. Ortiz,

Corresponding Author's Institution: University of Burgos, Faculty of Sciences

First Author: M. Sagrario Sánchez, Dra

Order of Authors: M. Sagrario Sánchez, Dra; Luis A. Sarabia, Dr; M.C. Ortiz

*Graphical Abstract (for review)



***Highlights (for review)**

A methodological approach to compute Pareto-optimal experimental designs

Pareto-optimal designs are a useful tool in Q&D (Quality by Design)

A family of optimal designs is computed by jointly handling several quality criteria

Ad hoc experimental designs are computed for a given number of experiments, domain, and model

Using genetic algorithms allows the search in both discrete and continuous spaces

***Novelty Statement**

An algorithmic approach is presented that allows the computation of the Pareto-optimal front for any criteria that a user may define to qualify an experimental design, indented to solve a specific problem. Complementary to similar approaches to the problem, the methodology presented here is more general because the search of the design can be made in both continuous and discrete spaces and there is not theoretical limit to the number of criteria under consideration. So, the user may reduce the cost without loosing quality of the design.

On the construction of experimental designs for a given task by jointly optimizing several quality criteria: Pareto-optimal experimental designs

M.S. Sánchez[1], L.A. Sarabia [1], M.C. Ortiz [21]

[1]Department of Mathematics and Computation, [2]Department of Chemistry
Faculty of Sciences, University of Burgos, Pza. Misael Bañuelos s/n, Burgos, Spain.

Abstract

Experimental designs for a given task should be selected on the base of the problem being solved and of some criteria that measure their quality. There are several such criteria because there are several aspects to be taken into account when making a choice. The most used criteria are probably the so-called alphabetical optimality criteria (for example, the A-, E-, and D-criteria related to the joint estimation of the coefficients, or the I- and G-criteria related to the prediction variance). Selecting a proper design to solve a problem implies finding a balance among these several criteria that measure the performance of the design in different aspects. Technically this is a problem of multi-criteria optimization, which can be tackled from different views.

The approach presented here addresses the problem in its real vector nature, so that *ad-hoc* experimental designs are generated with an algorithm based on evolutionary algorithms to find the Pareto-optimal front. There is not theoretical limit to the number of criteria that can be studied and, contrary to other approaches, no just one experimental design is computed but a set of experimental designs all of them with the property of being Pareto-optimal in the criteria needed by the user. Besides, the use of an evolutionary algorithm makes it possible to search in both continuous and discrete domains and avoids the need of having a set of candidate points, usual in exchange algorithms.

Keywords

Experimental design, design optimality, multicriteria optimization, Pareto-optimal front, evolutionary algorithms.

1. Introduction

It is known that the quality of the information extracted from an experiment depends upon the experimental conditions more than upon the response obtained from the experiment itself. Experimental design or design of experiments (DOE) is a methodology intended to obtain the best possible information from experiments. The relevance of the DOE is well known, even the American FDA's (Food and Drug Administration) Process Analytical Technology (PAT)

[1] Corresponding author, E-mail: mcortiz@ubu.es; Fax number: +34947258831

[1] puts statistical DOE and response surface methodologies (RSM) in the first place among multivariate mathematical approaches which should be used for PAT benefit.

There are several well known 'standard' experimental designs that are the best possible choice depending on the kind of problem to be solved and on the experimental domain under study. For example, factorial designs in cubic domains when the interest is mainly in the relevance of the factors, their possible interactions and how they affect the response; or central composite designs for spherical domains when the focus is prediction in the experimental domain, mainly for optimization.

Nevertheless, sometimes the standard designs are not affordable. To take an obvious example, when increasing the number of factors, the number of experiments in a factorial design increases geometrically. In such cases, designs with less number of experiments are needed and criteria to select among them. The main idea when selecting a proper design for a given task is to reduce the experimental effort (and not less relevant, the economical cost) without loosing quality on the information extracted.

Several criteria can be used to measure the quality of a design, each one representing different aspects of performance. For example, the D-criterion is related to the volume of the joint confidence region of the estimated coefficients. Very close to it, the A- and E-criteria relate to the 'sphericity' of the same region. The G- and I-criteria, on its part, focus on the variance of the predicted response in the experimental domain. Orthogonality, uniform variance, or protection against misspecification of model, are also of interest when establishing the quality of a design.

Some of these criteria are complementary in some sense but other can be competing criteria that should be balanced to obtain a good design for a specific situation. The choice of the design can be made sequentially, by prioritizing the criteria and deciding accordingly (in such a case the D-criterion is usually considered the primary criterion; they are the D-optimal designs). However, sometimes it would be useful achieving a compromise among several criteria to adapt the design to the specific scientific context of each problem.

This balance among several criteria can be accomplished either by weighting different criteria into a single objective function (e.g., a desirability function) which should be optimized, or computing the Pareto-optimal front defined by the competing criteria. The Pareto-optimal front contains the designs that are the best for a specific criterion while maintaining the others in their best allowable values, so that it permits identification of the trade-offs among criteria. Besides, the Pareto-optimal front gives a more complete picture of multi-objective problems than using weighting strategies [2].

The last approach is the one presented here, using an evolutionary algorithm to compute the Pareto-optimal front. In that way, for a specific problem, the study of the designs in the front allows wiser decisions among different possible designs (all of them with the property of

being Pareto-optimal). It is also an answer to the increasing demand of specific experimental designs, for example, the manufacturing though Quality-by-Design principles requires a design space [3] as opposed to classical nominal operating ranges. Therefore, the need of *ad hoc* ('fit for purpose') experimental designs with increasing number of factors is rising rapidly.

Lu et al. [4] report the estimation of the Pareto frontier for competing criteria in discrete spaces (vertices of hypercubes) although with a modification of an exchange algorithm. Genetic algorithms are used in [5] to construct D-optimal designs, and in [6,7] for supersaturated experimental designs. A- and D-optimal designs are computed in [8] with semi-definite programming. Park et al. [9] evaluate design performance in second order response surface problems and explore some trade-offs by using graphical methods. Also graphical methods are reported in [10] to asses sensitivity of response surface designs to model misspecification.

Comparing to these approaches, specially [4], the use of a genetic algorithm makes it possible to perform the search of the experimental points (the design) in both discrete and continuous spaces, that is, it can be used for selecting optimal designs for first order models usual in screening designs (which is the most common situation) but also in the context of RSM when second order models are more usual and the interest is also in the prediction variance.

2. Background

2.1 Some quality criteria in experimental design

The DOE methodology often involves the (least squares) fitting of a multilinear regression model of the form

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_{p-1} x_{p-1} + \varepsilon \qquad (1)$$

where y is the response (measured variable), x_i (i = 1, 2, …, k, $k \leq p - 1$) are the experimental factors (or their combinations) written in coded form and ε is a random variable which is supposed to follow a normal distribution with mean 0 and standard deviation σ. It is usual to write the model in eq. (1) in matrix form as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad (2)$$

where matrix **X,** called the model matrix, contains the information about the experiments to be done (the design) and the model to be fitted.

The least squares estimator of $\boldsymbol{\beta} = \left(\beta_0, \beta_1, ..., \beta_{p-1}\right)^t$ is

$$\mathbf{b} = \left(\mathbf{X}^t\mathbf{X}\right)^{-1}\mathbf{X}^t\mathbf{y} \tag{3}$$

And the variance-covariance matrix of the estimates is

$$Cov(\mathbf{b}) = \left(\mathbf{X}^t\mathbf{X}\right)^{-1}\sigma^2 \tag{4}$$

This is why matrix $\left(\mathbf{X}^t\mathbf{X}\right)^{-1}$ is called the 'dispersion matrix'. Also these expressions (and other like eq. (5) and (7)) highlight the importance of the dispersion matrix. The methodology of the experimental design includes the construction of experimental matrices so that the dispersion matrix is well enough. In that sense, different measures can be used to characterize the estimation and prediction capability of a design. Detailed expressions and discussions can be consulted in [11, 12].

Focusing on the precision of the estimated coefficients, and provided that the error variance σ^2 in eq. (4) is constant, the elements of the main diagonal of the dispersion matrix determine the quality (in terms of precision) of the estimated coefficients, and the remaining elements of the matrix are the covariances between each pair of coefficients. The closer to zero the elements of this matrix, the more precise and less correlated the estimates are.

Therefore, two different designs can be compared regarding their precision in the estimation of the individual coefficients b_i. To avoid dependence on the size of the experimental domain, a standardized value for the precision of each coefficient is used, the so-called Variance Inflation Factors (VIFs), which all have a minimum value of 1. Therefore, the best allowable precision for a coefficient is achieved when its corresponding VIF is equal to one.

When the estimates are jointly considered, the (1 – α) x 100 % joint confidence ellipsoid for the coefficients is determined by the set of vectors β such that

$$\left(\boldsymbol{\beta}-\mathbf{b}\right)^t\mathbf{X}^t\mathbf{X}\left(\boldsymbol{\beta}-\mathbf{b}\right) \le p\hat{\sigma}^2 F_{\alpha,p,N-p} \tag{5}$$

where p is the number of estimated coefficients, N the number of experiments in the design, $\hat{\sigma}^2$ is the variance of the residuals (estimate of σ^2) and $F_{\alpha;\, p,\, N-p}$ is the corresponding upper percentage point of an F distribution with p and $N - p$ degrees of freedom.

Again it is clear that the region in eq. (5) is defined by matrix $\mathbf{X}^t\mathbf{X}$ (the information matrix), so it only depends on the design and the model. When the estimates are jointly considered, the interest is on the 'smallest' joint confidence region. The D-criterion takes account of the behavior of the volume of this region. It is usually computed as

$$D = \frac{\left|\mathbf{X}^t\mathbf{X}\right|}{N^p} \qquad (6)$$

where the vertical lines denote the determinant of the matrix.

A design is said to be D-optimal when it achieves the maximum value of D in eq. (6), which means the minimum volume of the joint confidence region, so the most precise joint estimation of the coefficients.

The A and E criteria are related to the shape of the confidence region (the more spherical the region, the less correlated the estimates). Some more details about these criteria, properties and some modifications can be consulted in [5, 13].

When predicting in a domain is of interest, the variance of the prediction should also be taken into account through the Prediction Variance. Precisely the variance of the response predicted for a given point **x** in the experimental domain, is given by

$$\mathrm{Var}\left(\hat{y}\left(\mathbf{x}\right)\right) = \mathbf{x}_{(m)}^t \left(\mathbf{X}^t\mathbf{X}\right)^{-1} \mathbf{x}_{(m)} \sigma^2 = d\left(\mathbf{x}\right)\sigma^2 \qquad (7)$$

where $\mathbf{x}_{(m)}$ is the point **x** expanded to model form, σ^2 is the experimental variance, which acts as a constant in eq. (7), so the factor to be decreased is the one denoted by $d(\mathbf{x})$, the variance function. Again, to compare designs with different size, N, the scaled prediction variance (SPV) is used, which is $N\,d\left(\mathbf{x}\right)$. Desirable designs are those with the smallest SPV over the design space [9, 11]. In that sense, the G-criterion measures the maximum of the variance function, $d_{\max}$, over the experimental domain:

$$G = N\,d_{\max} = N \max_{\mathbf{x}} \left\{ d\left(\mathbf{x}\right) \right\} \qquad (8)$$

A design is said to be G-optimal when it achieves the minimum value of G in eq. (8). The Q-criterion (also known [9,14] as IV-, V- and I-criterion) uses the average value of $N\,d\left(\mathbf{x}\right)$ obtained by integrating it over the domain, although Borkowski [15] advised about the different values under the name Average Prediction Variance.

Standard experimental designs for screening or to study the effect of factors are optimal in one or more of these criteria. For example, it has been proven that two-level full factorial designs, or fractional factorial designs (of at least resolution III) with a first order model are D-, G- and I-optimal. They have VIF = 1 for all the coefficients, the dispersion matrix is a diagonal matrix (i.e, the design is orthogonal) with the same value along the diagonal, $1/N$, which is the minimum possible. Therefore, the joint confidence region is perfectly spherical and with the minimum volume.

The variance function is always the product of $1/N$ by a sum of squares (or products of squares) so that the maximum is always p/N taken at the vertices of the cube (thus always it is less than 1). Nevertheless, for second-order models, this is no longer true even for standard designs: central composite designs (CCD) and Box-Behnken designs (BBD) have small D- and G- values, but they are not D- or G-optimal [14].

Along this work we will focus on the D-criterion and the VIFs (related to the variance of the estimates, jointly or individually respectively), and the G-criterion that is related to the prediction variance.

2.2 Evolutionary algorithms for computing the Pareto-optimal front

Finding an experimental design that balances different competing criteria is a problem of multi-objective optimization. In the present paper, the problem is tackled by looking for the Pareto-optimal front in the competing criteria.

To introduce the concept of Pareto-optimality in the case at hand, let ξ_1 and ξ_2 be two experimental designs, that is, two design matrices of dimension $N \times k$ (number of experiments by number of factors); and let **F** denote the vector function of criteria, i.e, for $C \geq 2$ criteria, $\mathbf{F}(\xi_i) = \left(F_1(\xi_i), F_2(\xi_i), ..., F_C(\xi_i)\right)$ contains the values of the criteria for the corresponding design. Finally, without loss of generality, let us suppose that all the criteria should be minimized.

Then, a solution (a design) ξ_1 is said to dominate another design ξ_2 if $F_j(\xi_1) \leq F_j(\xi_2)$ for all the criteria ($j = 1, \ldots, C$), and there is at least one criterion in which design ξ_1 is strictly better, that is, there exist i such that $F_i(\xi_1) < F_i(\xi_2)$. A solution is said to be non-dominated with respect to a set of solutions when there is no other that dominates it. Consequently, the non-dominated solutions are those designs which are not worse than the rest in all the criteria and are at least the best in one of them. The set of the non-dominated solutions of the entire (criteria) space is the Pareto-optimal front so that, inside it, no design can improve one criterion without worsening another. In that way, the Pareto-optimal front provides a clear idea about the trade-off among criteria, that is, how much I should raise one of the criteria to get a decrease in another (and in how much).

To compute the Pareto-optimal front for a given problem, an evolutionary algorithm is used. First, the problem should be fully defined in terms of the number of factors (k), domain, model to be fit (that determines the number of coefficients, p) and number of experiments (N, $N \geq p$) to do so, and also the criteria to be taken into account.

Each individual in the population is an experimental design (N x k design matrix), codified according to the search space and such that $\left|\mathbf{X}^t\mathbf{X}\right| \geq 0.01$. Every design is evaluated in terms of the criteria, so that the fitness associated to each individual is a vector.

For the implementation of the algorithm, the design matrices are unfolded and handled as vectors in the N *x* k space so that no distinction is made among different experiments. Precisely, P (population size) individuals are generated at random inside the search space all of them representing experimental designs such that $|X^tX| \geq 0.01$.

In each generation, by uniform selection, pairs of individuals are selected and double point cross-over is applied with the crossing positions randomly chosen also with a uniform distribution. Then, off-springs are mutated with a given probability (selected by the user) so that new designs are generated and evaluated (provided that they have $\left|\mathbf{X}^t\mathbf{X}\right| \geq 0.01$). The procedure is repeated until P new off-springs are generated.

After that, the populations of parents and off-springs are merged together and 'sorted' according to levels of dominance. The non-dominated solutions (level 1) in the enlarged population are selected to survive for next generation, then the non-dominated solutions (level 2) that appear when removing those of level 1, and so on until enough individuals were selected to survive. It may happen than inside a level there are more individuals than needed to complete population. In that case, only the most dispersal inside the level are chosen, according to the crowding distance [2, 16].

The algorithm stops when the population has evolved for a predefined number of generations.

3. Results and discussion

The applicability and interpretability of the proposed approach is shown by its application to some specific situations.

3.1 Study of factors (or screening designs)

In [17] eighteen experiments were done to study the effect of six factors (k = 6) and the interaction between two of them. The goal was to determine sulfathiazole in milk (substance that has a maximum residue limit established by the European Union) by using molecular fluorescence spectroscopy. Before proposing an analytical procedure the effect on the recovery (%) of i) type of milk (UHT or pasteurized), ii) volume of TCA:milk (v/v), iii) centrifugation speed (rpm), iv) centrifugation temperature (ºC), v) derivatization time (min), and vi) volume of fluorescamine, were studied along with the possible interaction between the derivatization time and the volume of fluorescamine (it is possible that a greater volume of fluorescamine needs less reaction time and vice versa).

The factors are at two levels so the domain is a discrete domain that contains the vertices of the hypercube (the $2^6 = 64$ experiments of the full factorial design). The model to be fitted is ($p = 8$):

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_{56} x_5 x_6 + \varepsilon \qquad (9)$$

and the interest is on the estimation of the coefficients, so eight criteria were considered: the seven VIFs (that should be minimized) and the D-value (that should be maximized).

Running the algorithm for 100 generations with 100 designs (with coordinates -1 or +1) and probability of mutation of 0.1, the whole final population constitutes the estimation of the Pareto-optimal front for these eight criteria.

To study the resulting Pareto-optimal front (8-dimensional) a parallel coordinates plot has been used. The parallel coordinates plot [18] is a graphical visualization of points in multidimensional spaces (usually more than three) which has proven to be useful in multi-response optimization [19]. It consists of as many parallel lines as coordinates of the point to be represented (8 in this case). The height in each line is the numerical value of the coordinate itself and all of them are joined together by broken lines.

For the purposes here, the corresponding values of the criteria were range-scaled in order to improve the visualization, and this is why maximum and minimum values were written at the top and bottom, respectively, of each coordinate.

Fig. 1 shows the resulting graph, i.e, the parallel coordinates plot of the scaled Pareto-optimal front which is linked to the 100 experimental designs in the final population. In fig. 1, the first coordinate is used for D (the larger the better) and the rest for the VIFs of the coefficients, in the same order as they appear in the model (recall that the best possible value for all of them is one).

Here fig. 1

The values of the determinant of the corresponding matrix (D in eq. (6)) range from 0.03 to 0.68 in the front. The VIFs on their part range from 1 to some large values greater than 4 (sometimes more than six) for at least one of the coefficients. Furthermore, the lines in the plot crossing each others, above all for the VIFs, indicate a conflicting behaviour among criteria.

Anyway, the first visual impression is that there is more density of lines at the bottom of the figure (good values for the VIFs) although it is not clear that they are linked with the highest values of D. Nevertheless, it is true that the designs which achieve the worst values of at least one of the VIFs are linked to low values of D.

To better see this effect and to show some of the possibilities of having the whole family of optimal designs and how to move inside it, let us suppose that values of VIF less than 3 are desired for all the coefficients. Consequently, the designs with at least one VIF greater than 3 are discarded, and the re-scaled parallel coordinates plot of the remaining 64 designs is in fig. 2, all of them with acceptable values, though different, for all the criteria.

Here fig. 2

To give an idea about the differences among the designs in fig. 2 from a practical point of view, the semi-length (radius) of the confidence intervals for the coefficients is computed with the designs in fig. 2. Although it is not really necessary to make comparisons, in the original paper [17] the standard deviation of the recovery was estimated to be $\hat{\sigma} = 9.43\%$. Using this value, the critical value (95% confidence) of the Student *t* and the elements of the main diagonal of the dispersion matrix of the corresponding design, the minimum expected radius for any coefficient is 5 in at least one of the designs, but the maximum can be (depending on the design chosen) 5.3 for b_1, 9.0 for b_2, 5.4 for b_3, 9.4 for b_4, 5.2 for b_5, 5.7 for b_6 or 5.7 for b_{56}. That means that, in this case, the precision of the estimates of b_2 and b_4 may be very different, the same coefficient can be estimated plus or minus 5, or plus or minus 9. Again to put the numbers in context, the effect of the temperature (4^{th} factor) was estimated to be $b_4 = 5.11$ so that its precision is really relevant to decide about the significance of the factor.

Additionally, in fig. 2, the values of the criteria for the best design in each criterion have been marked by using thicker lines. The corresponding numerical values are written in table 1, i.e, only the values of the criteria for the eight ~~'best'~~ design in the extremes of the Pareto-optimal front are written.

Here table 1

Fig. 1 and (more clearly) fig. 2 show that the D-optimal design (thicker blue line starting at the top of the first coordinate in fig. 2) has small values of VIF for all the coefficients although none of them is 1. They also show that there are designs with VIF = 1 (highlighted by continuous thicker lines in fig. 2) but not for all the coefficients simultaneously (in fact, for no more than one coefficient at a time, table 1). Again, it is clear that the criteria are competing criteria.

The first design, number 1 in table 1, is the D-optimal one, with determinant equal to 0.68, value that coincides with the corresponding one in the D-optimal design used by the authors in [17] that was computed with an exchange algorithm as implemented in NemrodW [20]. It is worth noting that if the interest is only in the D-optimal design there is not advantage in using the procedure proposed here as against an exchange algorithm. What the multiobjective

approach adds is the availability of some other designs that behave differently in the estimation of the individual coefficients.

In that sense, the second design in table 1 achieves VIF = 1 for b_1 at the cost of the volume of the joint confidence region (the D value decreases until 0.44) and the VIFs of the rest of coefficients that remain greater than 1.2 (except for b_4) reaching 1.7 for b_2. The best estimation for the latter coefficient is achieved in design 3, but this time a larger loss in D should be 'paid' and, overall, larger VIFs for the rest of the coefficients, although more similar (among them). Again, D decreases to achieve VIF(b_3) = 1 in design 4 with the VIF of b_4 raising to 1.94. To maintain b_4 in its best allowable precision, design 5 in table 1, the VIF of b_1 and b_3 are around their worst values (inside table 1), and if, say, it is the interaction that needs to be the most precise, then some precision has to be lost above all in the estimation of b_4, b_3 and b_2, with intermediate values for D.

Table 1 and figures 1 and 2 only show values of the criteria. Each point in this criteria space indeed corresponds to an experimental design. As an example, table 2 shows the experimental designs whose values are in table 1. The levels (- and +) can be of course reversed without altering the characteristics of the design. However, care must be taken if only some factors are reversed because of the interactions chosen in the model.

Here table 2

Attention must be paid when looking at table 2 to find out the differences among designs, differences that appear clear in the Pareto-optimal front depicted in figures 1 and 2. In any case, these values are deeply related to the design matrix and the model, and thus the structure of the corresponding information matrix $\mathbf{X}^t\mathbf{X}$. Just as an example, the information matrix of design 1 (the D-optimal design) is

$$\begin{pmatrix} 18 & -2 & 2 & -2 & 2 & 0 & 0 & 2 \\ & 18 & 2 & -2 & 2 & 0 & -4 & 2 \\ & & 18 & 2 & -2 & -4 & -4 & 2 \\ & & & 18 & 2 & 0 & 0 & -2 \\ & & & & 18 & 0 & 0 & 2 \\ & & & & & 18 & 2 & 0 \\ & & & & & & 18 & 0 \\ & & & & & & & 18 \end{pmatrix} \qquad (10)$$

whereas for design 8 (the most precise estimation of the coefficient of the interaction considered in the model) is

$$\begin{pmatrix} 18 & 0 & 4 & 0 & 0 & 0 & 0 & 2 \\ & 18 & -2 & -6 & -2 & 2 & 2 & 0 \\ & & 18 & 6 & -6 & -2 & 2 & 0 \\ & & & 18 & 6 & 2 & 2 & 0 \\ & & & & 18 & 6 & -2 & 0 \\ & & & & & 18 & 2 & 0 \\ & & & & & & 18 & 0 \\ & & & & & & & 18 \end{pmatrix} \quad (11)$$

Only the upper triangular part of the matrices has been written because they are symmetric matrices. Anyway, matrices in eq. (10) and (11) are rather different (notice the last column in eq. (11) which is almost null), much more different than they can appear in their design matrices in table 2.

Finally, returning to the discussion about the values of the criteria in the Pareto-optimal front, the results also suggest that the *D* value is not sensitive to changes in the VIFs, except that large values appear. Comparing to fig. 1, in fig. 2 the worst D-value is 0.09 instead of 0.03 while the VIF values are varying in a narrower range, from 1 to almost 3 (the worst values for VIF are half of the ones in figure 1 and this hardly improves the worst values of D). This effect is even more noticeable when looking, for instance, to the values in the front for the design with VIF(b_3) = 1 (design number 4 in table 1) and the one marked with a dashed line in fig. 2, they both have the same D-value and their VIFs are different, specially for b_1, b_6 and b_{56}.

Similar analyses can be made with the rest of solutions in fig. 2 where there are 64 different possibilities to choose among them with the advantage of knowing exactly the extent of the conflicts that appear, which allows wiser decisions. In any case, usual practitioners of experimental design would accept any of them for the study at hand, particularly any of the eight designs in table 1.

3.2 Second-order models (response surface designs)

In the aim of some computations about D-optimal designs made by M.J. Box and Draper as reported by Atkinson and Donev [11], second order models for two factors (k = 2) varying in the square [-1, 1] x [-1, 1] are to be fitted, by using 6, 7, 8 or 9 experimental points (N). The selection of a two-dimensional experimental domain is also deliberated to depict the designs.

For all N, the model is defined in the following equation (12), it has p = 6 coefficients and the search space is continuous (any point inside the square).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + \varepsilon \quad (12)$$

In this case, concern focuses on the estimation of coefficients, and in the prediction variance. This is quantified by using two criteria, namely D- and G-criteria, equations (6) and (8) respectively. The value of D should be maximized, and the one of G should be minimized.

Several trials show that the Pareto-optimal front is very well populated in all the cases, and always the design in the Pareto-optimal front with the largest D value, the estimation of the D-optimal design for each N, coincides (except for rotations and symmetries) with the corresponding D-optimal design referred to in [11].

The results discussed in the following were obtained with 200 designs as population size, 0.1 of probability of mutation, and evolving for 1000 generations. Although the variance function is a quartic polynomial in the factors x_1 and x_2 for any design with N points chosen in the square, the computation of d_{max} is not straightforward, so it is estimated as the largest value in a uniform grid in the square.

This time there are two criteria so that the Pareto-optimal front can be plotted in the two-dimensional criteria space. This is Fig. 3 that depicts all the estimated Pareto-optimal fronts. Except for $N = 9$ that there is no conflict between criteria (in that case, the Pareto-optimal front reduces to a single solution which is the best in the two criteria), the rest of the fronts show the conflict: the increase of D implies an increase in G and vice versa, if G needs to be decreased it is at the cost of D. Nevertheless, the trade-offs between criteria that can be obtained are different depending on N.

Overall, the addition of experiments moves the fronts to the right (better D-values) and down (better G-values). In particular, doing 7 experiments, green asterisks in figure 3, instead of 6, blue crosses on the left of figure 3, clearly improves the D-criterion in all the designs but not so much the G-criterion; while adding another experiment, red pluses on the right of fig. 3, does not have such remarkable effect on D but the G values are clearly better in almost half of the designs with 8 experiments. Comparing the designs with 6 experiments to the designs with 8, both D and G are clearly improved.

Here fig. 3

Figure 3 can be misleading because of the definition of G in eq. (8). Most of the practitioners look directly for the value of d_{max} for evaluation of the prediction variance. None of the designs with $N = 6$ experiments reaches $d_{max} < 1$ and neither do the designs with $N = 7$. It is necessary to do at least $N = 8$ experiments to maintain the variance function below 1 in the whole domain and only for the designs whose G values in the Pareto-optimal front in fig. 3 are below 8 in the ordinate axis. The Pareto-optimality (the non-dominance) implies that, for these cases, G values less than 8 can be obtained but for values of D no greater than $8.9\ 10^{-3}$.

There are two factors, so the designs can be plotted in the two-dimensional experimental domain. Figure 4 shows all the experimental points whose Pareto-optimal front is in fig. 3;

fig. 4a) is for the designs with $N = 6$, fig. 4b) for $N = 7$, fig. 4c) when $N = 8$, and finally fig. 4d) contains a single design, the corresponding to $N = 9$, which is the standard central composite design in the cubic domain or the 3^2 factorial design with levels -1, 0, 1. Additionally, the two extremes of each Pareto-optimal front (the best design according to the D criterion and the best design with the G-criterion) are marked with different symbols, squares for the D-optimal and circles for the G-optimal. Also, they have been detailed numerically in table 3.

Here fig. 4

Apart from the two mentioned designs, figure 4 does not allow the distinction among the different 'intermediate' designs but shows a systematic structure in the selection of points: covering the sides (specially the vertices), the centre and, when there are enough experiments, the 'principal axes' of the domain ending in the standard structure for $N = 9$. It is noteworthy that rotations of these designs are equally qualified but different rotations do not appear often in the population. This is probably because the individuals are obtained mostly by cross-over of existing designs and because the algorithm evolves searching for improved, dispersal, non-dominated values for the criteria.

Here table 3

Moreover, this is not so clear in figure 4 but for $N = 6$ the 200 designs contain the two vertices of the square corresponding to $x_1 = 1$ (a single point is seen in fig. 4a) in these positions). Starting from the D-optimal design, the squares in fig. 4a), with coordinates in the first block of table 3, it contains three of the vertices of the domain, two points to the right (-0.57, 1) and bottom (-1, 0.26) of the fourth vertex and a near central point. As we change the design to obtain better values of G (and consequently worse values of D), that is, when moving in the fronts in fig. 3 from top to bottom, without considering the two vertices that are in all the designs, the other four points move themselves around, near the centre or following the corresponding side of the square, $x_2 = 1$ with x_1 moving from -0.56 to -0.65 and $x_1 = -1$ with x_2 slightly varying around 0.26, being the most distinctive characteristic the point that moves horizontally away from the vertex (-1,-1) to (-0.78, -1).

For $N = 7$, fig. 4b), all the designs contain the opposite vertices (1, -1) and (-1, 1). In fact, the D-optimal design and those 'near' it have the four vertices of the square, also a point near the centre and two points in the middle of two of the sides of the square, precisely (0.08, -1) and (1, -0.09) for the D-optimal design in the second block of table 3. Then, as D decreases, the central point moves up and left, the points in the middle of the sides move slightly around their positions and the two other opposite vertices move themselves along the side up and left to achieve (-1, -0.89) and (0.89, 1) in the G-optimal design.

For $N = 8$, fig. 4c), all 200 designs contain the four vertices. The D-optimal design, third block in table 3, contains also three points in the middle of three of the sides of the square

and an almost central point (nearer to the side without point in its middle, (0.22, 0)). When improving G, these four points move themselves to the middle of the upper and bottom side ($x_1 \approx 0$, with $x_2 = \pm 1$) and (last block in table 3) from (-1, 0) to (-0.78, 0) and from (0.22,0) to (0.88, 0), i.e, around $x_1 = \pm 0.8$ with $x_2 = 0$.

These arrangements find a no-conflicting situation when there are $N = 9$ points, fig. 4d), the four vertices, the centre and the four axial points in the sides of the square.

4. Conclusions

The proposed algorithmic approach makes it possible to address the computation of *ad hoc* experimental designs with the property of being optimal in one or several criteria stated by the user.

For some well-known and usual properties in discrete spaces (e.g. D-, A-, or E-criteria), the approach here is an alternative to the usual exchange algorithms. Besides, it is also applicable when the search space is a continuous space, situation in which exchange algorithms are no longer valid.

Acknowledgements

Financial support is acknowledged under projects CTQ2011-26022 (Spanish Ministerio de Economía y Competitividad) and BU108A11-2 (Junta de Castilla y León).

References

[1] US Department of Health and Human services: food and drug administration, guidance: PAT-A Framework for innovative pharmaceutical development, manufacturing and quality assurance, 2004

[2] K. Deb, Multiobjective optimization using evolutionary algorithms, Wiley, Chichester, 2001

[3] International conference on harmonisation, ICH (R2) Q8: Pharmaceutical development: Step 4, 2009.

[4] L. Lu, C.M. Anderson-Cook, T.J. Robinson, Technometrics 53 (2011) 353-365.

[5] A. Heredia-Langer, W.M. Carlyle, D.C. Montgomery, C.M. Borror, G.C. Runger, J. Qual. Technol. 35 (2003) 28-46.

[6] R. Cela, E. Martínez, A.M. Carro, Chemometr. Intell. Lab. 52 (2000) 167-182.

[7] R. Cela, E. Martínez, A.M. Carro, Chemometr. Intell. Lab. 57 (2001) 75-92.

[8] A. Babapour Atashgah, A. Seifi, Optim. Eng. 10 (2009) 75-90.

[9] Y.J. Park, D.E. Richardson, D.C. Montgomery, A. Ozol-Godfrey, C.M. Borror, C.M. Anderson-Cook, J. Qual. Technol. 37 (2005) 253-266.

[10] C. M. Anderson-Cook, C. M. Borror and B. Jones, Technometrics 51 (2009) 75-87.

[11] A.C. Atkinson, A.N. Donev, Optimum experimental designs, Oxford University Press, New York, 1992.

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

[13] G. Box and friends, Improving almost anything. Ideas and essays, Wiley, Hoboken (NY), 2006.

[14] C.A. Anderson-Cook, C.B. Borror, D.C. Montgomery, J. Stat. Plan. Infer. 139 (2009) 629-641.

[15] J.J. Borkowski, J. Qual. Technol. 35 (2003) 70-77.

[16] M.C. Ortiz, L. Sarabia, A. Herrero, M.S. Sánchez, Chemometr. Intell. Lab. 83 (2006) 157-168.

[17] R. Morales, M.C. Ortiz, L.A. Sarabia, M.S. Sánchez, Anal. Chim. Acta 707 (2011) 38-46.

[18] A. Inselberg, Parallel coordinates: visual multidimensional geometry and its applications, Springer, New York, 2008.

[19] M.C. Ortiz, L.A. Sarabia, M.S. Sánchez, D. Arroyo, Anal. Chim. Acta 687 (2011) 129-136.

[20] D. Mathieu, J. Nony, R. Phan-Tan-Luu, NemrodW, version 2000-D, LPRAI, Marseille, France.

Figure captions

Figure 1. Scaled parallel coordinates plot of the Pareto-optimal front in the eight criteria

Figure 2. Scaled parallel coordinates plot of the reduced Pareto-optimal front. The thicker lines mark the best values for at least one of the criteria.

Figure 3. Pareto-optimal fronts in the criteria space for $N = 6$ (blue crosses, x), 7 (green asterisks, *), 8 (red pluses, +) and 9 (cyan star)

Figure 4. Experimental points for the designs in the Pareto-optimal front for a) $N = 6$, b) $N = 7$, c) $N = 8$, and d) $N = 9$. The D-optimal design is marked with squares and the G-optimal design with circles.

Table

Table 1. Values of the eight criteria for the experimental designs which are the best in each one. The best possible values are underlined.

Number	D	VIF(b_1)	VIF(b_2)	VIF(b_3)	VIF(b_4)	VIF(b_5)	VIF(b_6)	VIF(b_{56})
1	<u>0.68</u>	1.12	1.18	1.08	1.08	1.06	1.11	1.05
2	0.44	<u>1.00</u>	1.70	1.57	1.09	1.21	1.22	1.25
3	0.39	1.44	<u>1.00</u>	1.38	1.35	1.25	1.20	1.48
4	0.29	1.39	1.27	<u>1.00</u>	1.94	1.10	1.69	1.44
5	0.28	1.54	1.33	1.74	<u>1.00</u>	1.08	1.18	1.44
6	0.30	1.38	1.13	1.77	1.50	<u>1.00</u>	1.43	1.28
7	0.47	1.27	1.11	1.29	1.33	1.13	<u>1.00</u>	1.55
8	0.38	1.18	1.54	1.71	1.71	1.18	1.09	<u>1.00</u>

Table 2. Experimental designs linked to the values in table 1

	design 1						design 2						design 3						design 4					
#	x_1	x_2	x_3	x_4	x_5	x_6	x_1	x_2	x_3	x_4	x_5	x_6	x_1	x_2	x_3	x_4	x_5	x_6	x_1	x_2	x_3	x_4	x_5	x_6
1	-	-	-	-	-	+	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	+	-
2	-	-	-	-	+	-	-	-	-	+	-	-	-	-	-	-	+	+	-	-	-	+	-	+
3	-	-	-	+	+	-	-	-	+	+	-	+	-	-	+	+	-	-	-	-	-	+	+	-
4	-	-	+	-	+	+	-	-	+	+	+	-	-	-	+	+	+	+	-	-	+	-	-	-
5	-	-	+	+	+	+	-	+	-	-	-	-	-	+	-	-	-	-	-	+	-	-	-	+
6	-	+	-	-	-	+	-	+	-	-	-	-	-	+	-	-	+	-	-	+	-	-	+	-
7	-	+	-	+	+	+	-	+	+	-	+	-	-	+	-	+	+	+	-	+	+	-	-	-
8	-	+	+	-	-	-	-	+	+	+	+	+	-	+	+	-	+	-	-	+	+	-	+	-
9	-	+	+	+	-	-	-	+	+	+	+	+	+	-	-	-	-	-	-	+	+	-	+	+
10	-	+	+	+	-	+	+	-	-	-	+	-	+	-	-	-	+	-	-	+	+	+	-	+
11	+	-	-	+	-	-	+	-	-	+	+	-	+	-	-	+	-	+	+	-	+	-	+	-
12	+	-	-	+	+	+	+	-	-	+	+	+	+	-	+	-	-	+	+	-	+	+	-	+
13	+	-	+	+	-	+	+	-	+	-	+	+	+	-	+	+	+	-	+	+	-	-	-	-
14	+	+	-	-	-	-	+	+	-	+	-	-	+	+	-	-	-	-	+	+	-	-	-	-
15	+	+	-	-	+	+	+	+	+	-	-	-	+	+	-	+	-	-	+	+	-	+	+	+
16	+	+	-	+	-	-	+	+	+	-	-	+	+	+	+	-	-	+	+	+	-	+	+	+
17	+	+	+	-	+	-	+	+	+	+	-	-	+	+	+	+	-	+	+	+	+	+	+	-
18	+	+	+	+	+	-	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+

	design 5						design 6						design 7						design 8					
#	x_1	x_2	x_3	x_4	x_5	x_6	x_1	x_2	x_3	x_4	x_5	x_6	x_1	x_2	x_3	x_4	x_5	x_6	x_1	x_2	x_3	x_4	x_5	x_6
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	+
2	-	-	-	-	+	+	-	-	-	-	+	+	-	-	-	+	-	+	-	-	+	+	-	-
3	-	-	-	-	+	+	-	-	-	-	+	+	-	-	+	+	+	-	-	-	+	+	+	+
4	-	-	-	+	-	-	-	-	+	+	+	+	-	+	-	-	-	-	-	+	-	-	-	-
5	-	-	-	+	-	+	-	-	+	+	+	+	-	+	-	+	-	+	-	+	-	-	-	+
6	-	-	-	+	+	-	-	+	-	-	-	+	-	+	-	+	+	-	-	+	+	-	+	-
7	-	-	+	-	-	+	-	+	-	+	-	-	-	+	+	-	+	+	-	+	+	+	-	-
8	-	+	-	-	-	+	-	+	-	+	+	-	-	+	+	+	-	-	-	+	+	+	+	-
9	-	+	+	-	-	-	-	+	+	-	+	-	-	+	+	+	-	+	-	+	+	+	+	+
10	+	-	-	-	-	-	+	-	-	-	+	-	+	-	-	-	-	+	+	-	-	-	-	-
11	+	-	+	-	-	+	+	-	+	+	-	+	+	-	+	-	-	-	+	-	-	+	+	-
12	+	-	+	-	+	-	+	-	+	+	-	+	+	-	+	+	-	-	+	-	-	+	+	-
13	+	-	+	-	+	-	+	-	+	+	+	-	+	-	+	+	+	+	+	-	-	+	+	+
14	+	-	+	+	-	+	+	+	-	-	+	-	+	+	-	-	-	-	+	+	-	-	-	-
15	+	+	-	-	-	-	+	+	-	-	+	-	+	+	-	-	+	+	+	+	-	-	+	+
16	+	+	+	-	-	+	+	+	+	-	-	-	+	+	-	+	+	-	+	+	+	-	-	+
17	+	+	+	+	-	-	+	+	+	+	+	+	+	+	+	-	+	-	+	+	+	-	+	+
18	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	-	+

Table 3. Experimental designs that correspond to the extremes of the Pareto-optimal front.

		D-optimal		G-optimal	
	#	x_1	x_2	x_1	x_2
$N = 6$	1	-1	-1	-1	0.32
	2	-1	0.26	-0.78	-1
	3	-0.57	1	-0.65	1
	4	0.08	-0.17	0.21	0
	5	1	-1	1	-1
	6	1	1	1	1
		x_1	x_2	x_1	x_2
$N = 7$	1	-1	-1.00	-1	-0.89
	2	-1	1	-1	1
	3	-0.09	0.06	-0.18	0.21
	4	0.08	-1	0.06	-1
	5	1	-1	0.89	1
	6	1	-0.09	1	-1
	7	1	1	1	-0.07
		x_1	x_2	x_1	x_2
$N = 8$	1	-1	-1	-1	-1
	2	-1	0	-1	1
	3	-1	1	-0.78	0
	4	-0.09	-1	-0.07	-1
	5	-0.08	1	-0.06	1
	6	0.22	0	0.88	0
	7	1	-1	1	-1
	8	1	1	1	1

Figure1

Click here to download Figure: figure1.eps

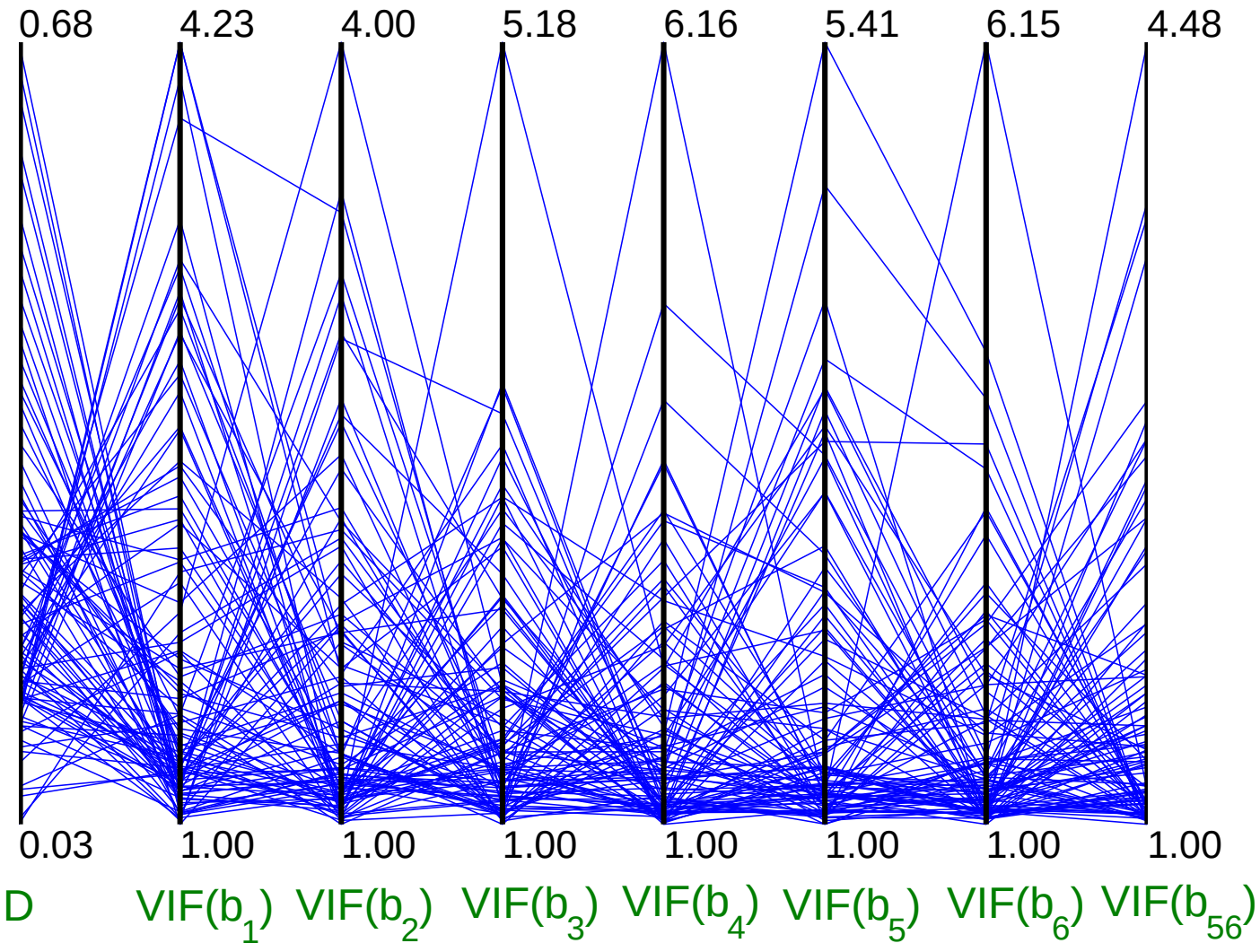


Figure2

Click here to download Figure: figure2.eps

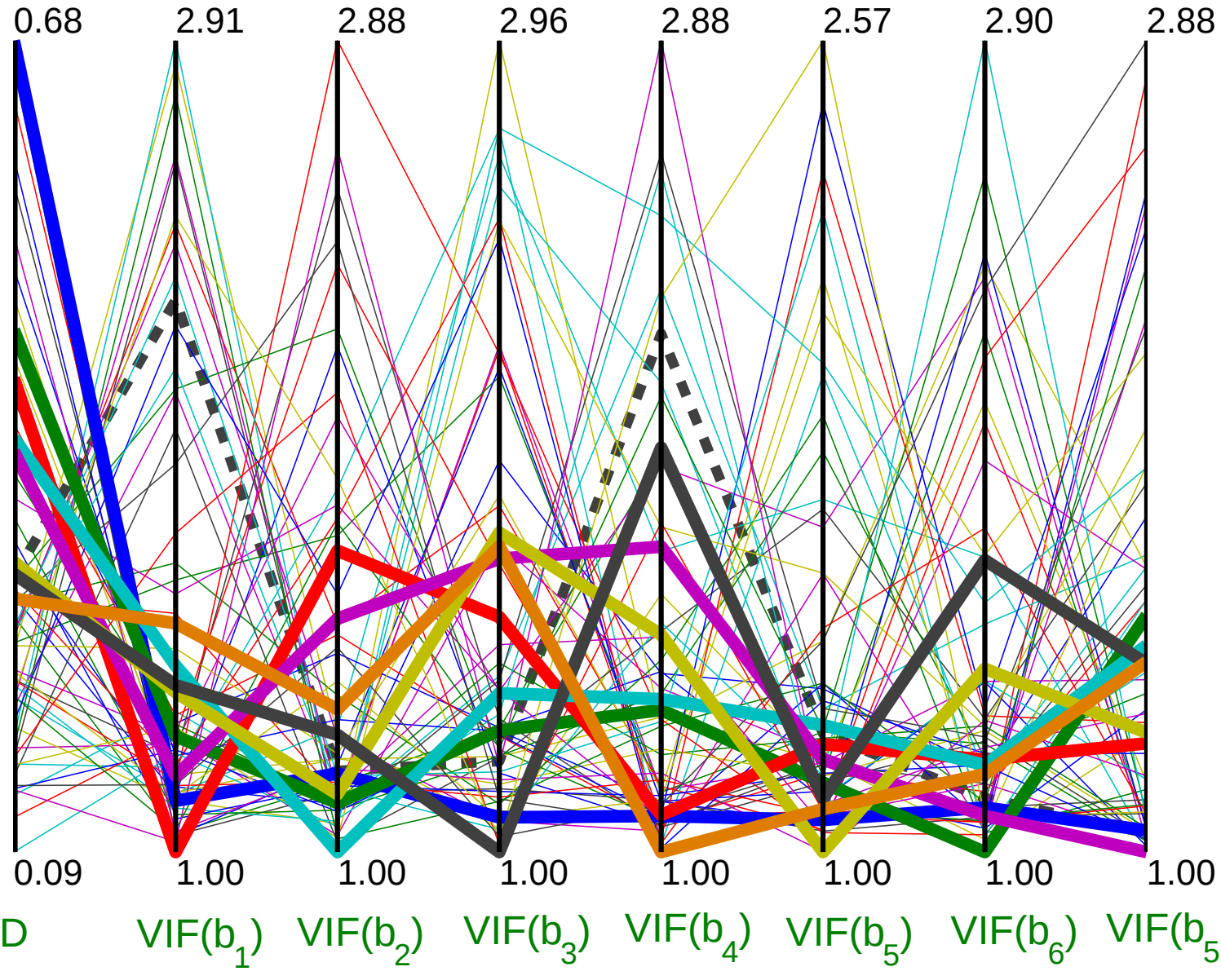


Figure3
Click here to download Figure: figure3.eps

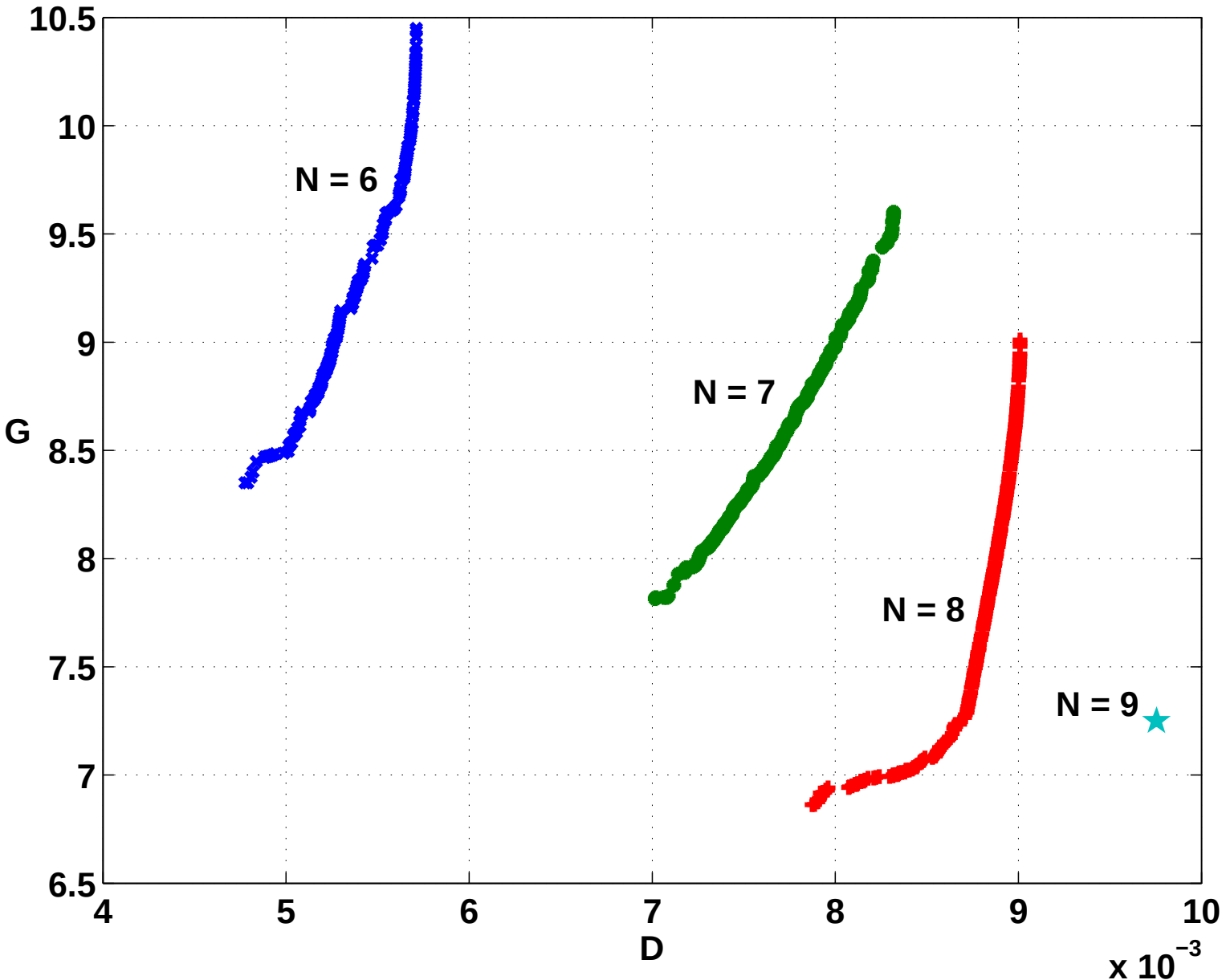


Figure 4

Click here to download high resolution image