Analytica Chimica Acta 652 (2009) 161–172

Contents lists available at ScienceDirect

Analytica Chimica Acta

journal homepage: www.elsevier.com/locate/aca

Experimental design in chemistry: A tutorial

Riccardo Leardi*

Department of Pharmaceutical and Food Chemistry and Technology, Via Brigata Salerno (ponte), University of Genoa, I-16147 Genoa, Italy

ARTICLE INFO

Article history:
Received 3 April 2009
Received in revised form 3 June 2009
Accepted 4 June 2009
Available online 10 June 2009

Keywords:
Experimental design
Optimization
Response Surface Methodology
Mixture design

ABSTRACT

In this tutorial the main concepts and applications of experimental design in chemistry will be explained. Unfortunately, nowadays experimental design is not as known and applied as it should be, and many papers can be found in which the "optimization" of a procedure is performed one variable at a time. Goal of this paper is to show the real advantages in terms of reduced experimental effort and of increased quality of information that can be obtained if this approach is followed. To do that, three real examples will be shown. Rather than on the mathematical aspects, this paper will focus on the mental attitude required by experimental design. The readers being interested to deepen their knowledge of the mathematical and algorithmical part can find very good books and tutorials in the references [G.E.P. Box, W.G. Hunter, J.S. Hunter, Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building, John Wiley & Sons, New York, 1978; R. Brereton, Chemometrics: Data Analysis for the Laboratory and Chemical Plant, John Wiley & Sons, New York, 1978; R. Carlson, J.E. Carlson, Design and Optimization in Organic Synthesis: Second Revised and Enlarged Edition, in: Data Handling in Science and Technology, vol. 24, Elsevier, Amsterdam, 2005; J.A. Cornell, Experiments with Mixtures: Designs, Models and the Analysis of Mixture Data, in: Series in Probability and Statistics, John Wiley & Sons, New York, 1991; R.E. Bruns, I.S. Scarminio, B. de Barros Neto, Statistical Design—Chemometrics, in: Data Handling in Science and Technology, vol. 25, Elsevier, Amsterdam, 2006; D.C. Montgomery, Design and Analysis of Experiments, 7th edition, John Wiley & Sons, Inc., 2009; T. Lundstedt, E. Seifert, L. Abramo, B. Thelin, Å. Nyström, J. Pettersen, R. Bergman, Chemolab 42 (1998) 3; Y. Vander Heyden, LC–GC Europe 19 (9) (2006) 469].

1. Introduction

To introduce the concepts of experimental design, let me tell you a story.

Once upon a time, in the Kingdom of Far Far Away, the King wanted to optimize a chemical reaction, finding the setting producing the highest yield. The task was quite simple, since only two variables (temperature and reaction time) were involved in the process and previous knowledge indicated that the ranges in which the optimum had to be looked for were 40–80 degrees Celsius and 20–60 min, respectively. It was also known that both high temperature and long reaction time could decompose the product; for both variables a maximum was therefore expected. Anyway, it was so important to optimize this reaction that the King decided that the person giving the best solution would have married his beautiful daughter or his beautiful son (depending on the sex of the winner...).

The first person to show up was a man who was very famous in the Kingdom for having "optimized" a lot of reactions in his long career.

As he always did, he started by performing the "cheapest" experiment (40 degrees, 20 min), with a yield of 20.6%. Then, to save time, he kept the time constant at 20 min and started increasing the temperature: at 50 degrees the yield was 32.3%, at 60 degrees it was 39.9%, at 70 degrees it was 43.5%, at 80 degrees it was 43.0%. Since the yield at 80 degrees was lower than at 70 degrees, this meaning that the best temperature was lower than 80 degrees, he performed a new experiment at 75 degrees, with a yield of 43.7%, which was the highest till then obtained. Therefore, he concluded that the best temperature had been found and that he could keep it constant and started changing the reaction time. This led him to do the following experiments, all of them at a temperature of 75 degrees: 30 min (50.2%), 40 min (52.1%), 50 min (49.5%). Since the yield at 50 min was lower than at 40 min, this meaning that the best time was lower than 50 min, he performed a new experiment at 45 min, with a yield of 51.4%. So, the best reaction time was 40 min.

The man ended his job and was very satisfied: since he had optimized at first temperature and then time, he was pretty sure that he had found the best conditions: when working at 75 degrees with a reaction time of 40 min a yield of 52.1% was obtained, which was the highest possible yield. The whole process required a total of ten experiments. So, becoming the new Prince had been quite easy...

But then a second man went to the King. He was a young man, not as famous as the first one, without any direct experience of that

* Tel.: +39 010 3532636; fax: +39 010 3532684.
E-mail address: riclea@dictfa.unige.it.

0003-2670/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.aca.2009.06.015

Table 1
The nine experiments suggested by the second man.

40 degrees, 20 min
40 degrees, 40 min
40 degrees, 60 min
60 degrees, 20 min
60 degrees, 40 min
60 degrees, 60 min
80 degrees, 20 min
80 degrees, 40 min
80 degrees, 60 min

specific problem, and the only thing he was proudly reporting on his CV was that he had followed a course of experimental design. Of course, the King did not know anything at all about experimental design, but he was liking this man so much that he gave him the possibility of trying to optimize his reaction. The man gave him a sheet of paper containing the list of the nine experiments that should have been performed (Table 1).

The King was quite upset by this unusual behaviour (having decided which experiments to perform even before actually doing the first one, and furthermore by changing both variables at the same time!) but since he was quite open-minded he allowed the man to do what he had planned.

The nine experiments gave the following results: 20.6%, 44.9%, 51.0%, 39.9%, 55.1%, 52.1%, 43.0%, 49.1%, 37.0%.

How shocked was the King when he saw that the fifth experiment (60 degrees, 40 min), gave a yield of 55.1%, higher than the 52.1% that the first expert had claimed as the highest possible value!

The young man, when called by the King, instead of staying content with the results he got, suggested to try a new experiment at 60 degrees and 45 min, predicting that the yield would have been even higher. The experiment was performed and the yield was indeed 56.1%!

Of course, the King was so happy that he immediately celebrated the wedding between this man and the beautiful Princess. The Princess was fascinated by his new husband and she whispered him: "Darling, how could you do that? Are you a magician?". And he replied: "My dear Princess, I'm not a magician, I'm just a chemometrician...". And the two of them lived happily ever after, with the young man optimizing every aspect of their life...

2. Interactions

Unfortunately, the very great majority of people behave as the first "expert" of the story. When browsing through the papers published in Analytica Chimica Acta in 2009 (from volume 631 to volume 645, plus the papers available in the "Articles in Press" section on June 3, 2009), I found 165 of them having the general title or a section title containing the words "optimization" or "development", or "improvement", or "effect of". Only in 11 papers (i.e., one out of 15...) a multivariate approach has been followed, while in the great majority of them the "optimization" was performed one variable at a time, sometimes with the titles of the subsections proudly remarking it ("3.1. Effect of pH", "3.2. Effect of temperature", "3.3. Effect of flow", and so on...). Five of them [9–13] have been published in the section "Extraction and Sample handling", two [14,15] in the section "Chemometrics", two [16,17] in the section "Electrochemistry" and two [18,19] in the section "Separation methods".

As the story has shown, the "optimization" performed OVAT (One Variable At a Time) does not guarantee at all that the real optimum will be hit. This because this approach would be valid only if the variables to be optimized would be totally independent from each other.

If somebody would ask you what is the best gear to ride a bike, your reply would surely be: "It depends".

"What is the best cooking time for a cake?" "It depends".

"What is the best waxing for your skis?" "It depends".

"What is the best setup for a racing car?" "It depends".

This means that you do not have "the best" gear, but the best gear depends on the levels of the other factors involved, such as the slope of the road, the direction and the speed of the wind, the level of the cyclist, how tired the cyclist is, the speed he wants to keep.

The same, when cooking a cake the best time depends on the temperature of the oven, the best waxing depends on the conditions of the weather and of the snow, the best setup for a racing car depends on the circuit, and so on.

Every time your reply is "it depends", then it means that you intuitively recognize that the effect of the factor you are talking about is not independent of the levels of the other factors; this means that an interaction among those factors is relevant and that not taking it into account can give terrible results.

So, it is evident that the housewife knows very well that there is a strong interaction between cooking time and oven temperature, a cyclist knows very well that there is an interaction between the gear and the surrounding conditions, and so on.

Of course, you will never hear a housewife using the word "interaction", but her behaviour demonstrates clearly that she intuitively understands what an interaction is.

Could you imagine somebody looking for the best gear on a flat course (i.e., changing gear while keeping all the remaining variables constant) and then using it on any other course simply because the first set of experiments demonstrated that it was the best?

Well, chemists optimizing their procedures OVAT behave in the very same way!

Why the same people answering "it depends" on a lot of questions about their everyday life, when entering a lab and working as chemists will never give the same answer?

Why, when looking for the best pH, chemists usually behave as the foolish cyclist described before, by changing pH and keeping constant all the remaining variables instead of thinking that the "best pH" may depend on the setting of the other variables?

In the story narrated in the Introduction this was absolutely evident: the young man found a better yield because there was a strong interaction between the two variables. As a consequence, the "best" temperature found by the first man was simply the best temperature when the reaction time was 20 min, while each reaction time has a different optimal temperature (the longer the reaction time, the lower the optimal temperature).

3. Modeling

Another extremely important difference between the two approaches described in the story is the fact that the first man decided which experiment to do next on the basis of the outcome of the previous experiments, while the second man had already well clear in his mind the whole set of experiments (it was a simple grid covering the whole experimental domain).

As a consequence of that, in the first case only a local knowledge was obtained, this meaning that only the results of the experiments actually performed could be known, each of them with an incertitude corresponding to the experimental error.

Instead, from the results obtained by the second man a simple mathematical model could be obtained, relating the response with the experimental conditions:

$$Y = b_0 + b_1X_1 + b_2X_2 + b_{12}X_1X_2 + b_{11}X_1^2 + b_{22}X_2^2$$

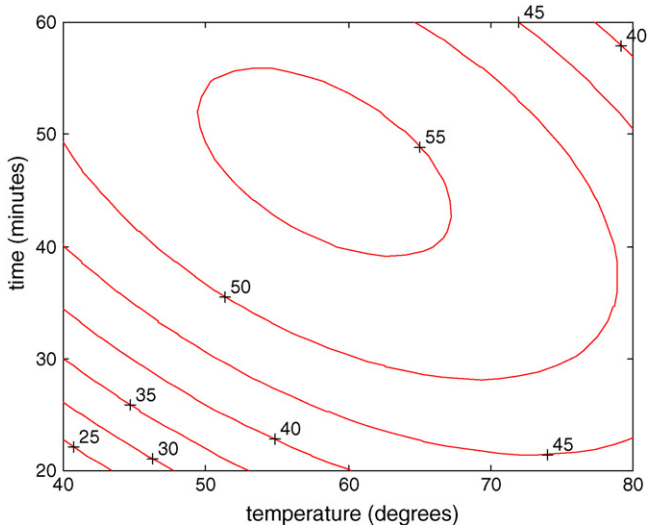

Fig. 1. Isoresponse plot (yield of the chemical reaction).

After having estimated the constant and the coefficients of the two linear terms, of the interaction and of the two quadratic terms (any software for multiple linear regression will do it), by simply replacing X_1 and X_2 with actual values it will be possible to predict the response for any possible setting (i.e., for any point of the experimental domain), even for those experiments that have not been actually performed.

Since the response in any point of the experimental domain can be predicted, then a graphical representation can be easily obtained. In Fig. 1, the isoresponse curves connect all the points having the same predicted response, in exactly the same way as in a topographical map the isohypses connect all the points having the same altitude. By looking at this plot, the general behaviour of the phenomenon can be easily understood, and the conditions corresponding to the global maximum can be easily found. This is how the future Prince could suggest a further experiment after the first series, leading to the real maximum!

By looking at the plot, it is also easy to understand that the best temperature changes according to the reaction time (and the other way round), and this is the physical interpretation of the interaction between the two variables (without any interaction, the axes of the ellipses would be parallel to the axes of the plot).

But there is something else... Fig. 2 shows the leverage plot of the experimental design of Table 1. The leverage can be computed in every point of the experimental domain (it depends on the experimental matrix and on the postulated model), and its value, multiplied by the experimental variance, corresponds to the variance of the estimate of the response in that point. Therefore, a leverage of 1 means that the response can be predicted with the same precision of the actual experiment, while a leverage < 1 means that the response can be predicted better than if one real experiment would be performed in the same point. Since in the whole experimental domain the leverage is always much lower than 1, this means that the response in any point will be known with a better precision by using the prediction from the model obtained by the responses of the nine experiments of the experimental design than by performing an actual experiment in the same point. This was the reason why the future Prince was so confident in suggesting a setting corresponding to an experiment he never performed!

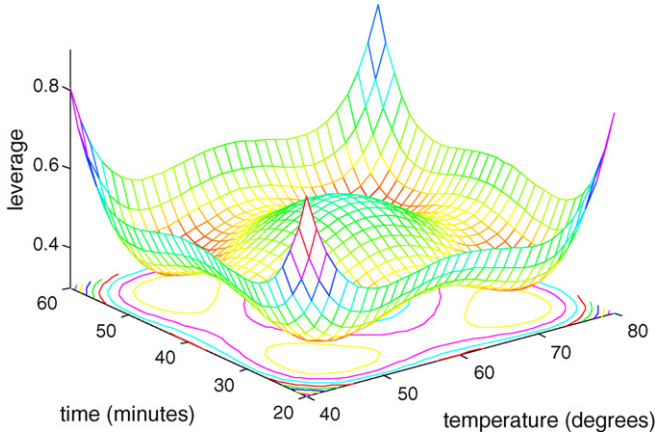

Fig. 2. Leverage plot (chemical reaction).

Another important point is that the leverage depends only on the experimental design and not on the outcome of the experiments. This means that, since it can be computed before starting to do the experiments, it is possible to know in advance whether the precision of the estimate will be acceptable or not (of course, to have this knowledge the experimental variance should be known).

4. A different way of thinking

As we have seen, the quality of the information, i.e. the precision of the estimate, depends both on the experimental variance and on the leverage. The great majority of the chemists try with any means to reduce the experimental variance (e.g., buying a more precise instrument, or doing the experiments as carefully as possible), but since they do not do any experimental design they do not care about the leverage. While the experimental variance cannot be reduced very much (perhaps by an order of magnitude), the leverage can jump to extremely high values.

By comparing the information obtained by an OVAT approach with the information obtained by an experimental design we can say that:

- the experimental design takes into account the interactions among the variables, while the OVAT does not;
- the experimental design provides a global knowledge (in the whole experimental domain), while the OVAT gives a local knowledge (only where the experiments have been performed);
- in each point of the experimental domain, the quality of the information obtained by the experimental design (leverage always <1) is higher than the information obtained by the OVAT (leverage = 1, and only for those points where the experiments have been performed);
- the number of experiments required by an experimental design is smaller than the number of experiments performed with an OVAT approach.

Summarizing, it should be well clear that:

- the quality of the results depends on the distribution of the experiments in the experimental domain;
- the optimal distribution of the experiments depends on the postulated model;
- given the model, the experimental limitations and the budget available (= maximum number of experiments), the experimental design will detect the set of experiments resulting in the highest possible information.

People should also be aware that building the experimental matrix (i.e., deciding which experiments must be performed) is the easiest part of the whole process. What is difficult is instead the def-

inition of the problem (Which are the factors to be studied? Which is the domain of interest? Which model? How many experiments?).

To perform an experimental design, the following five steps must be considered:

(1) Define the goal of the experiments. Though it can seem totally absurd, many people start doing experiments without having clear in their minds what the experiments are done for. This is a consequence of the general way of thinking, according to which once you have the results then you can anyway extract information from them (and the more experiments have been performed, the better).
(2) Detect all the factors that can have an effect. Particular attention must be given to the words "all" and "can". This means that it is not correct to consider a predefined number of factors (e.g., let us take into account only three factors) and that saying that a factor "can" have an effect is totally different from saying that we think that a factor has an effect. One of the most common errors is indeed that of performing what has been defined a "sentimental screening", often based only on some personal feelings rather than on scientific facts.
(3) Plan the experiments. Once the factors have been selected, their ranges have been defined and the model to be applied has been postulated, this step requires only a few minutes.
(4) Perform the experiments. While in the classical way of thinking this is the most important part of the process, in the philosophy of experimental design doing the experiments is just something that cannot be avoided, in order to get results that will be used to build the model.
(5) Analyse the data obtained by the experiments. This step transforms data into information and is the logical conclusion of the whole process.

Very often one single experimental design does not lead to the solution of the problem. In those cases the information obtained at point 5 is used to reformulate the problem (removal of the non-significant variables, redefinition of the experimental domain, modification of the postulated model), after which one goes back to step 3.

Since the possibility of having to perform more than one single experimental design must always be taken into account, it is wise not to invest more than 40% of the available budget in the first set of experiments.

5. Factorial designs

A chemical company was producing a polymer, whose viscosity had to be >46.0×10^3 mPa s. As a consequence of the variation of a raw material, they got a final product rather different from the "original" product (being produced since several years), with a viscosity below the acceptable value. Of course, this was a very big problem for the company, since the product could not be sold anymore. The person in charge of the product started performing experiments OVAT, but after about thirty experiments he could not find any acceptable solution.

A young lady, who had just followed a three-day course in experimental design, decided to try to apply to this specific problem what she had learnt.

At first, she detected three variables that could have been relevant: they were the amounts of three reagents (let us call them A, B and C). The original formulation was 10 g of A, 4 g of B and 10 g of C. She decided to keep this experimental setting as starting point and to explore its surroundings. Since the number of possible experiments was quite limited, she decided to apply a 2^3 Factorial Design, requiring a total of eight experiments.

Table 2
The experimental matrix of the 2^3 factorial design.

Reagent A (X_1)	Reagent B (X_2)	Reagent C (X_3)
−1	−1	−1
1	−1	−1
−1	1	−1
1	1	−1
−1	−1	1
1	−1	1
−1	1	1
1	1	1

The 2^k factorial designs are the simplest possible design, requiring a number of experiments equal to 2^k, where k is the number of variables under study. In these designs each variable has two levels, coded as −1 and +1, and the variables can be either quantitative (e.g., temperature, pressure, amount of an ingredient) or qualitative (e.g., type of catalyst, type of apparatus, sequence of operations).

The experimental matrix is reported in Table 2, and it can be seen that it is quite easy to build it (no software required!). The matrix has eight rows (2^3, each row corresponding to an experiment) and 3 columns (each column corresponding to a variable); in the first column the −1 and +1 alternate at every row, in the second column they alternate every second row, in the third column they alternate every fourth row. The same procedure can be used to build any factorial design, whatever the number of variables.

From a geometrical point of view, as shown in Fig. 3, a factorial design explores the corners of a cube (if the variables are more than three it will be a hypercube; our mind will no more be able to visualize it, but from the mathematical point of view nothing will change).

Contrary to what happens in the OVAT approach, in which variable 1 is changed while variables 2 and 3 are kept constant, in the factorial designs variable 1 is changed while variables 2 and 3 have different values (of course the same happens for all the variables). This means that the factorial design is suitable for estimating the interactions between variables (i.e., the difference in changing variable 1 when variable 2 is at its higher level or at its lower level, and so on).

The mathematical model is therefore the following:

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_{12}X_1X_2 + b_{13}X_1X_3 + b_{23}X_2X_3 + b_{123}X_1X_2X_3$$

As a consequence, with just eight experiments it is possible to estimate a constant term, the three linear terms, the three two-term interactions and the three-term interaction.

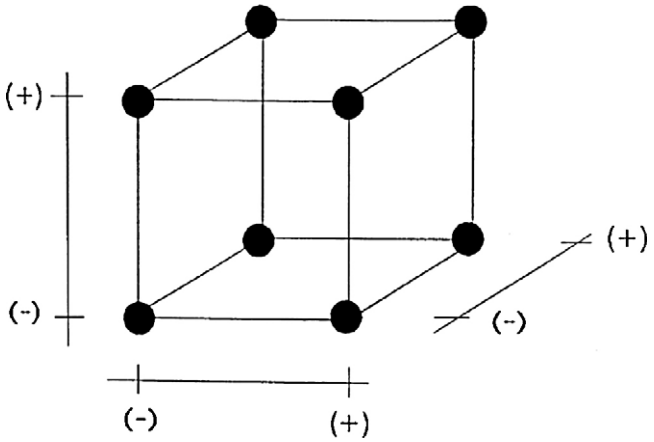

Fig. 3. A graphical representation of a 2^3 factorial design.

Table 3
The experimental plan and the responses of the 2^3 factorial design.

Reagent A (g)	Reagent B (g)	Reagent C (g)	Viscosity (mPa s) × 10^3
9	3.6	9	51.8
11	3.6	9	51.6
9	4.4	9	51.0
11	4.4	9	42.4
9	3.6	11	50.2
11	3.6	11	46.6
9	4.4	11	52.0
11	4.4	11	50.0

The next step was to define the levels of the variables and to write down the experimental plan.

As said before, she decided to keep the original recipe as center point and to set the levels −1 and +1 of each variable symmetrically to the original value (9 and 11 for reagents A and C, 3.6 and 4.4 for Reagent B), this leading to the experimental plan reported in Table 3. As it can be seen, while the experimental matrix contains the coded values (−1 and +1), the experimental plan reports the real values of the variables and therefore can be understood by anybody.

After having performed the eight experiments (in random order, not to introduce unwanted systematic effects) and having recorded the responses, it was immediately clear that in several cases the viscosity was much higher than the minimum acceptable value! How was it possible not to have found those solutions in more than 30 previous experiments?

To compute the coefficients, we must go from the experimental matrix to the model matrix (Table 4). While the former has as many rows as experiments and as many columns as variables, the latter has as many rows as experiments and as many columns as coefficients and can be easily obtained in the following way: the first column (b_0) is a column of +1, the columns of the linear terms are the same as the experimental matrix, the columns of the interactions are obtained by a point to point product of the columns of the linear terms of the variables involved in the interaction (e.g., the column b_{12} of the interaction between variables 1 and 2 is obtained by multiplying point to point the column b_1 by the column b_2).

Computing the coefficients is very simple (again, no software required!). For each of them, multiply point to point the column corresponding to the coefficient that has to be estimated by the column of the response, and then take the average of the results. For instance, for estimating b_1 (the linear term of X_1), just do $(-51.8+51.6-51.0+42.4-50.2+46.6-52.0+50.0)/8=-1.8$.

An interesting thing to notice is that, since every column of the model matrix has four −1 and four +1, every coefficient will be computed as half the difference between the average of the four experiments with positive sign and the average of the four experiments with negative sign. This means that each coefficient is computed with the same precision, and that this precision, being the difference of two averages of four values, is much better than that of an OVAT experiment, where the difference between two experiments (one performed at higher level and one performed at lower level) is usually computed. Once more, it can be seen how the experimental design can give much more information (the interaction terms) of much higher quality (higher precision of the coefficients).

Table 4
Model matrix and responses of the 2^3 factorial design.

b_0	b_1	b_2	b_3	b_{12}	b_{13}	b_{23}	b_{123}	Viscosity
+1	−1	−1	−1	+1	+1	+1	−1	51.8
+1	+1	−1	−1	−1	−1	+1	+1	51.6
+1	−1	+1	−1	−1	+1	−1	+1	51.0
+1	+1	+1	−1	+1	−1	−1	−1	42.4
+1	−1	−1	+1	+1	−1	−1	+1	50.2
+1	+1	−1	+1	−1	+1	−1	−1	46.6
+1	−1	+1	+1	−1	−1	+1	−1	52.0
+1	+1	+1	+1	+1	+1	+1	+1	50.0

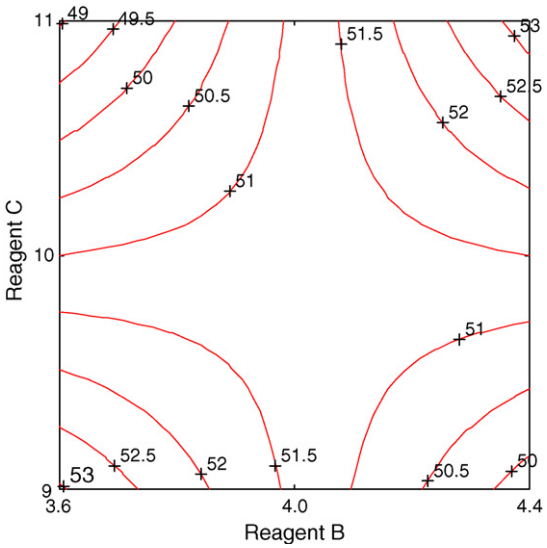

Fig. 4. Isoresponse plot (viscosity of the polymer); Reagent A at 9 g.

The following model has been obtained:

$$Y = 49.4 - 1.8X_1 - 0.6X_2 + 0.2X_3 - 0.8X_1X_2 + 0.4X_1X_3 + 1.9X_2X_3 + 1.2X_1X_2X_3$$

Since eight coefficients have been estimated with eight experiments (and therefore no degrees of freedom are available) and since the experimental variability is not known it is impossible to define a statistical significance of the coefficients. Anyway, those of the linear term of X_1 (Reagent A) and the interaction X_2–X_3 (Reagent B–Reagent C) have absolute values larger than the other ones.

The large coefficient of X_1 indicates that by increasing the amount of Reagent A a decrease of the viscosity is obtained (the sign of the coefficient is negative), and therefore better results are obtained by reducing its amount. Since X_1 is not involved in any relevant interaction, we can conclude that this effect is present whatever the values of the other two reagents.

In what concerns the interaction Reagent B–Reagent C, it can only be interpreted by looking at the response surface shown in Fig. 4. Since we are plotting the response on the plane defined by two variables (think at a slice of the cube depicted in Fig. 3), we must define the level of variable 1 at which we want to represent the response (i.e., where to cut the slice. . .). The clear effect of Reagent A (the lower, the better), leads us in the choice of setting the value of X_1 at its lower level (−1, corresponding to 9 g).

The geometrical shape of a linear model without interactions is a plane (the isoresponse lines are parallel); if relevant interactions are present, then it becomes a distorted plane (the isoresponse lines are not parallel). This is the case of the response surface on the plane Reagent B–Reagent C. By looking at the plot, it can be seen that an increase of Reagent B decreases viscosity when Reagent C is at its lower level, while it has the opposite effect when Reagent C is at its higher level. In the same way, an increase of Reagent C decreases viscosity when Reagent B is at its lower level, while it has the opposite effect when Reagent B is at its higher level.

Looking at the plot, it can also be understood why the OVAT approach did not produce any good result. If you go to the center point (corresponding to the original formulation) and you change the amount of either Reagent B or Reagent C (but not both at the same time!), this meaning moving parallel to the axes, you

will realize that, whatever experiment you will do, nothing will change. Instead, owing to the strong interaction, you only have relevant variations when you change both variables at the same time.

Two combinations produce the same response: 3.6 g of Reagent B and 9 g of Reagent C, and 4.4 g of Reagent B and 11 g of Reagent C. Since a higher amount of reagents increases the speed of the reaction, and therefore the final throughput, the latter has been selected and therefore the best combination was: 9 g of Reagent A, 4.4 g of Reagent B and 11 g of Reagent C. All the experiments were performed at lab scale, and therefore this formulation had to be tested at the plant. When doing it, the results obtained in the lab were confirmed, with a viscosity in the range $50.0–52.0 \times 10^3$ mPa s, well over the acceptability value.

Happy but not totally satisfied, the lady tried one more experiment. The results of the experimental design showed that a decrease of Reagent A was leading to better products, and that this variable was not involved in interactions with the other variables. Of course, this behaviour was demonstrated only inside the experimental domain, but it could have been worthwhile to check if the effect was the same also outside it. The most logical development would have been to do a further experimental design centered on the new formulation, but she had not enough time to do eight more experiments. So, she just tried to further reduce Reagent A, and she tested the formulation having 7 g of Reagent A, 4.4 g of Reagent B and 11 g of Reagent C. This experiment was a total success, since the product obtained at the plant had a viscosity in the range $55.0–60.0 \times 10^3$ mPa s, well above the acceptable value.

Of course, everybody in the company was very happy with the result. Everybody except one person. Can you guess who? It was the expert in charge of the problem, who could not accept that somebody else could succeed with just nine experiments where he totally failed, in spite of having performed a huge number of experiments.

This not a fairy tale as the one in the Introduction. This is a real story [20]. The consequence of this success is that nowadays the same company regularly uses experimental designs whenever they can.

One more comment. The previous example is not an optimization. Probably, if more experiments would have been performed with more experimental designs, even better results could have been obtained. Anyway, the immediate goal of the company was not to find the optimum, but rather to get out of an embarrassing situation and to find a commercially valid solution as fast as possible, and the factorial design, the simplest of all the experimental designs, allowed to get a substantial improvement with a very limited experimental effort.

6. Face Centered Design

A company had to find the best operating conditions for a coal mill. Two variables were under study: the load of the mill (tons h^{-1}) and the position of the classifier (a curtain partially closing the mill). From previous knowledge, it was known that both variables had a quadratic behaviour and therefore a model containing also the quadratic terms was needed. For this reason it was not possible to use the factorial design shown in the previous example, since, having each variable only two levels, it is not suitable to estimate the quadratic terms.

It has then been decided to use a Face Centered Design (the same used by the future Prince), with the experimental matrix reported in Table 5.

Also in this case, building the experimental matrix is very simple. Having k variables, the first 2^k experiments are the same as for a Factorial Design. Then, the next $2 \times k$ experiments are obtained by keeping all the variables except one at their central level, with the other variable having the values of −1 and +1. Finally, the last n experiments (in our case it was just one) are performed with all the variables at their central value. It has to be noticed that by increasing the number of center points (up to three) an improvement of the mathematical characteristics of the model can be obtained (mainly, a lower leverage and therefore a better prediction in the region around the center); anyway, one should always evaluate if such an improvement justifies a higher experimental work.

Table 5
Experimental matrix, experimental plan and responses of the Face Centered Design.

X_1	X_2	Load	Classifier	Y_1	Y_2
−1	−1	5	1	79.00	0.05
+1	−1	15	1	66.58	1.44
−1	+1	5	3	63.50	0.03
+1	+1	15	3	58.42	0.50
−1	0	5	2	80.75	0.01
+1	0	15	2	69.22	0.88
0	−1	10	1	73.25	0.40
0	+1	10	3	63.31	0.08
0	0	10	2	77.31	0.12

Fig. 5 shows a graphical representation of a Face Centered Design on three variables. It can be seen that it spans a cubical domain and that it is composed by a Factorial Design (the points at the vertices of the cube) and by a Star Design. The name "Face Centered" derives from the fact that the location of the star points corresponds to the center of the faces of the cube.

More generally speaking, the Face Centered Design is a special case of a Central Composite Design, this name indicating the designs composed by a Factorial Design and a Star Design.

In the Circumscribed Central Composite Design the length of the "arms" of the star is the square root of the number of variables. In that case, each variable has five levels (vs. the three levels of the Face Centered Design), all the points are equidistant from the center and the design spans a spherical domain (Fig. 6).

The Central Composite Designs allows to estimate the constant, the linear terms, the interactions between variables and the quadratic terms, according to the following model (usually, the interactions among more than two terms are not taken into account):

$$Y = b_0 + b_1X_1 + b_2X_2 + b_{12}X_1X_2 + b_{11}X_1^2 + b_{22}X_2^2$$

After having decided the levels of the variables, the experimental plan reported in Table 5 has been obtained. The position of the classifier is a continuous variable. In the experimental plan it has been

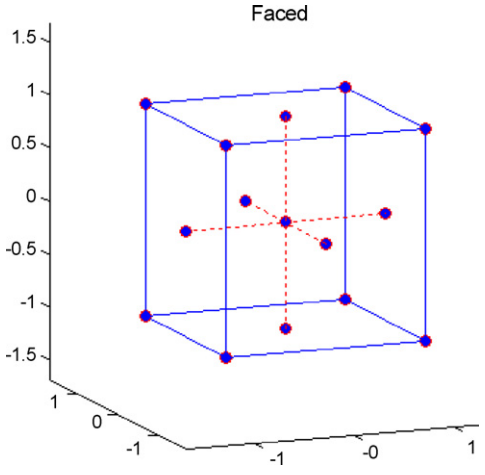

Fig. 5. A graphical representation of a Face Centered Design on three variables.

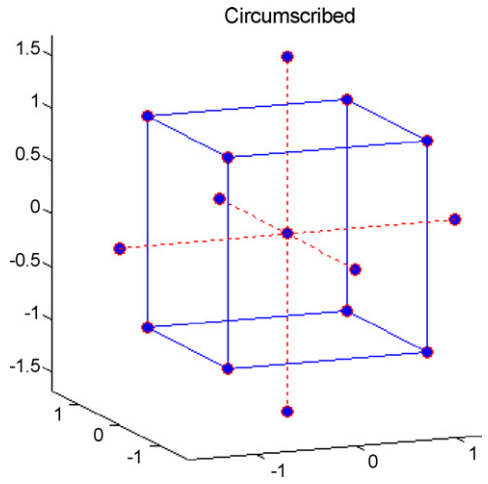

Fig. 6. A graphical representation of a Central Composite Design on three variables.

coded as 1 (closed), 2 (open at 50%) and 3 (open), but it can take any value.

For each experiment, two responses were measured: the percentage of "small" particles (with diameter <200 mesh) and the percentage of "large" particles (with diameter >50 mesh) (see Table 5). The goal of the study was to find the operating conditions leading to a product of acceptable quality ($Y_1 > 70$ and $Y_2 < 1$).

By comparing this study with the previous ones, we can immediately detect two relevant differences:

- more than one response is studied, this meaning that we must look at all of them at the same time, finding the best compromise;
- rather than looking for the "best" conditions, we are interested in identifying the region of the experimental domain corresponding to a product of "acceptable" quality; inside this region the most profitable setting will be selected (in our case, the one corresponding to the highest load).

The following models have been obtained:

$$Y_1 = 76.90 - 4.84X_1(^{**}) - 5.60X_2(^{**}) + 1.84X_1X_2 - 1.71X_1^2 - 8.42X_2^2(^{**})$$

$$Y_2 = 0.15 + 0.45X_1(^{***}) - 0.21X_2(^{**}) - 0.23X_1X_2(^{**}) + 0.28X_1^2(^{**}) + 0.08X_2^2$$

Since nine experiments have been performed and six coefficients have been estimated, three degrees of freedom are available and therefore the significance of the coefficients can be estimated (the level of significance is indicated according to the usual convention: $^{*} = p < 0.05$, $^{**} = p < 0.01$, $^{***} = p < 0.001$). An easier way to look at the coefficients and at their relative magnitude is the bar plot reported in Fig. 7.

Fig. 8 shows the response surfaces. For the first response (to be maximized) it can be seen that it decreases when the load increases. The linear term of X_2 suggests that the response decreases when the classifier is opened, but the highly significant quadratic term indicates that the behaviour is not linear, as confirmed by the response surface, from which it can easily be seen that the best results are obtained with the classifier in an intermediate position. Anyway, the response is not acceptable only when the classifier is almost completely open and when the classifier is totally closed and the load is maximum.

The second response (to be minimized) increases with the load (very strong linear term for X_1), but this increase is particularly fast

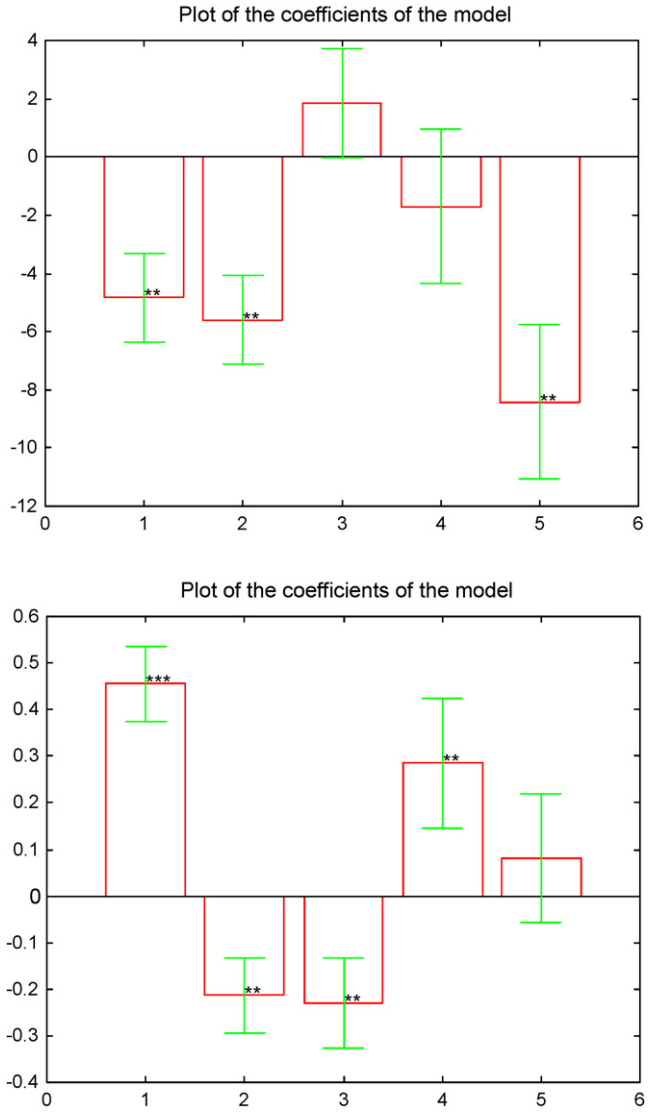

Fig. 7. A graphical representation of the coefficients of the models of the two responses of the Face Centered Design (top: Y_1, bottom: Y_2). The brackets correspond to the confidence intervals at $p = 0.05$; the stars indicate the significance of the coefficient ($^{*} = p < 0.05$, $^{**} = p < 0.01$, $^{***} = p < 0.001$).

at higher loads (as a consequence of the significant quadratic term for X_1). The more the classifier is opened, the better this response is (significant linear term for X_2). The interaction between the two variables can be seen by the fact that at higher loads the effect of the classifier is higher than at lower loads and that when the classifier is closed the effect of the load is higher than when the classifier is totally open. The response surface also shows that the response is always acceptable, except for a very small region at higher load and classifier closed, that was unacceptable also for the first response.

Therefore, the most profitable combination seems to be maximum load and classifier about 30% open.

This way of thinking would be correct if we would not have to cope with the experimental error. Indeed, the isoresponse curve at 70% connects all the points with a predicted response of 70%. But we also know that any prediction has a confidence interval, and when producing a product we must be reasonably sure that it has the desired characteristics. So, instead of looking at the predicted value as such, we should rather take into account the lower limit of the confidence interval if the response must be maximized (or the upper limit if the response has to be minimized).

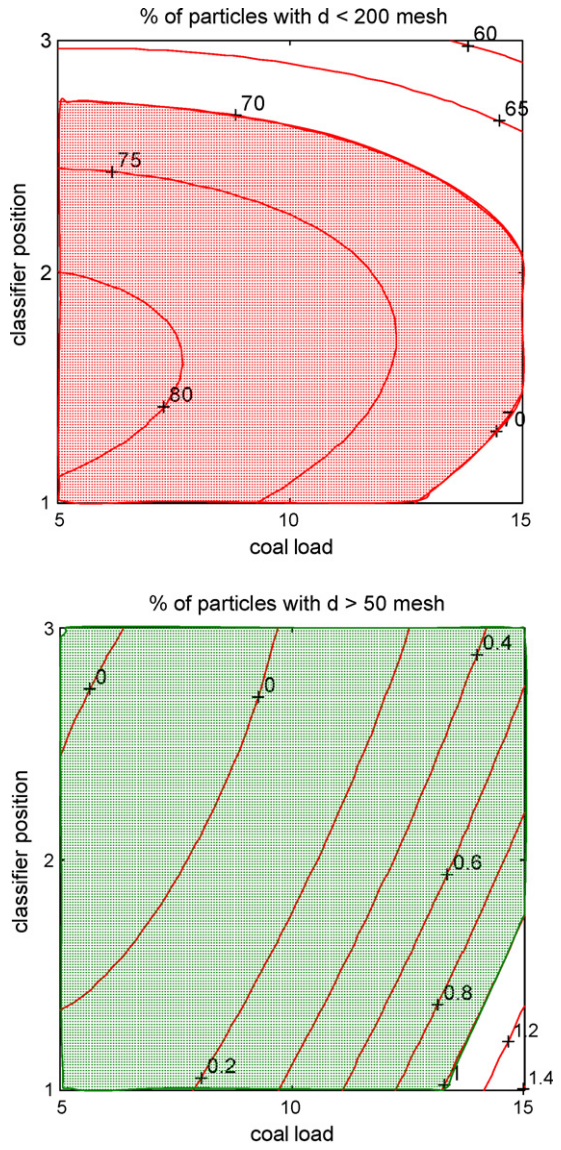

Fig. 8. Isoresponse plots of the two responses of the Face Centered Design (top: Y_1, bottom: Y_2). For each of them the region of acceptability is highlighted.

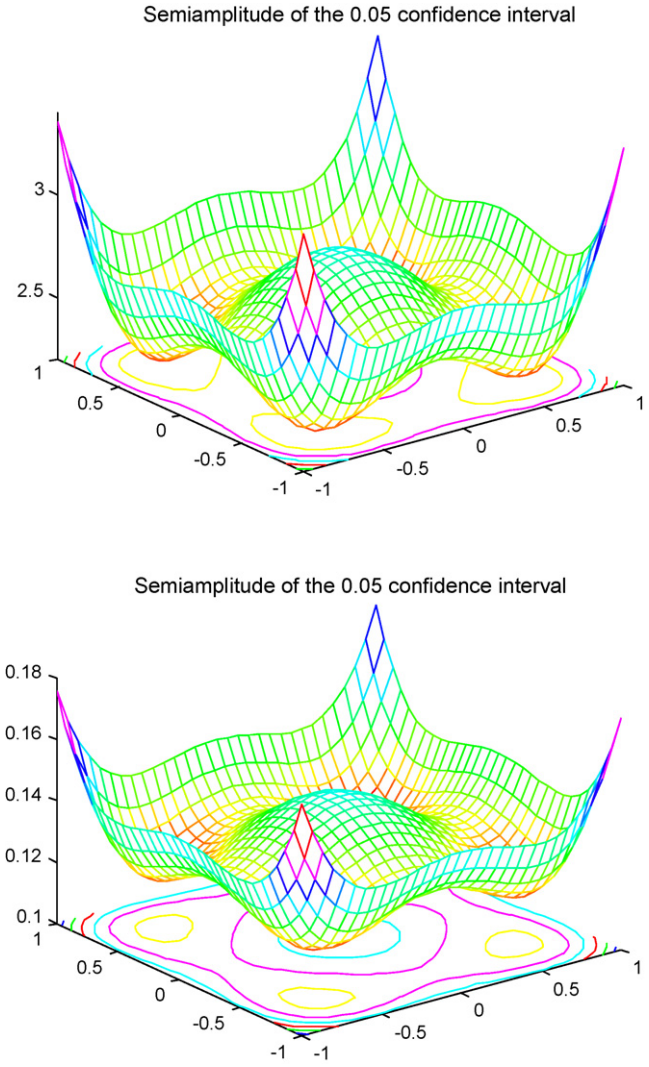

Fig. 9. Semiamplitude of the 95% confidence interval for the two responses of the Face Centered Design (top: Y_1, bottom: Y_2).

How to compute the confidence interval of the prediction? We have already seen that if we multiply the leverage by the experimental variance we obtain the variance of the prediction. We also know that it is possible to compute the leverage for any point of the experimental domain (do not worry if you do not know how to do it mathematically: a software will do it for you. . .). About the experimental variance, in this case we do not have replicates but, since we have degrees of freedom, we can estimate it from the residuals of the regression: in our case the standard deviation of the residuals is 1.18 for Y_1 and 0.06 for Y_2, both computed with 3 degrees of freedom.

Therefore if we multiply the leverage of a point by 1.18^2 we have the variance of the prediction of Y_1 in that point. By doing the square root we have the standard deviation of the prediction, and if we multiply it by 3.18 (the t at $p = 0.05$, with 3 degrees of freedom) we have the semiamplitude of the 95% confidence interval in the same point.

The leverage of this design is the same shown in Fig. 2. The relatively higher leverage in the center depends by the fact that no replicates have been performed at the center point. By adding more replicates, this bump would have been smaller: for instance, with three center points (i.e., two more experiments) a very wide flat region would have been obtained. Of course, when designing a set of experiments the best compromise between quality of information and experimental effort must be looked for. Two more experiments would have improved our design, but would it have been worthwhile to perform them (i.e., what would have been the extra cost in terms of time and money)? This is a decision that only the people in charge of the specific problem can take!

Fig. 9 shows the plots of the semiamplitude of the 95% confidence interval for each of the two responses. These plots, derived from the leverage plot, are anyway much more easily interpretable since they give a direct idea of the uncertainty of the predicted value. So, we can immediately see that for the first response the uncertainty is between about ±2 and about ±3, while for the second response it is in the range ±0.1–±0.2. It is very important to notice that the confidence interval changes according to the position in the experimental domain, and its shape depends on the distribution of the experiments in the experimental domain. Again, what is important is which experiments are performed, not how many!

For any point of the experimental domain both the predicted responses and their confidence intervals are known. This allows us to draw the plots reported in Fig. 10. As expected, the regions in which it is "safe" to work are a little bit more reduced than when

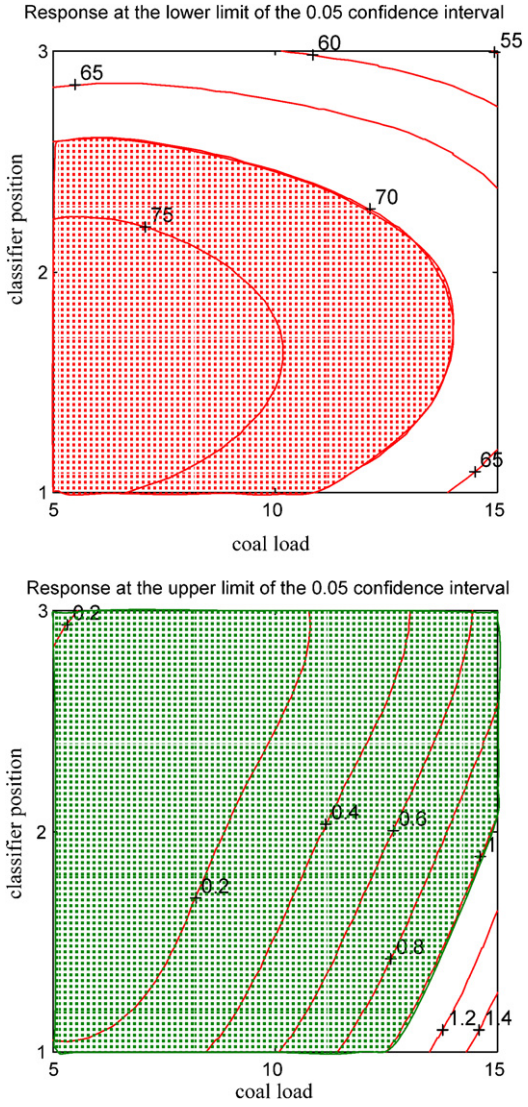

Fig. 10. Isoresponse plots of the two responses of the Face Centered Design, each at the limit of the confidence interval of interest (top: Y_1, bottom: Y_2). For each of them the region of acceptability is highlighted.

taking into account the predicted responses. Furthermore, there are two more comments that can be done.

While the isoresponse curves are very nicely shaped ellipses, in this case the curves are not completely regular. This depends on the fact that, as shown in Fig. 9, the size of the confidence interval is not the same in any point.

When looking at the isoresponse curves of the second response (Fig. 8), it can be seen that some part of the domain has a predicted response <0. Of course, this is totally nonsense from the chemical point of view, since it would mean that a negative amount of large particles has been produced! Anyway, the predicted values of this region are not significantly different from 0. And indeed, when taking into account the upper limit of the confidence interval as shown in Fig. 10, no points in the experimental domain have a value <0.

Fig. 11, showing the overlapped plots of the confidence limits of interest, can be used to take the final decision. It can be seen that the region of acceptability of the first response is a subregion of that of the second response. If the maximum possible coal load is the main goal of the study, then a load of about 14 tons h^{-1} with the classifier in position 1.6 (30% open) will give a product having both the responses at acceptable levels. It can also be seen that by reducing a little bit the load the percentage of large particles will

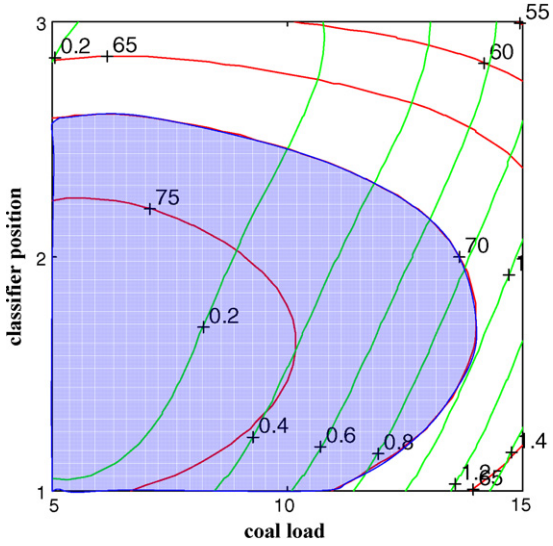

Fig. 11. Overlapped isoresponse plot of the two responses of the Face Centered Design, each at the limit of the confidence interval of interest. The region of joint acceptability is highlighted.

decrease quite fast. So, for instance, when working with a load of 12 tons h^{-1} and the classifier in position 2.4 (70% open), the percentage of large particles will decrease to no more than 0.4%, and therefore the quality of the product will be much higher.

Now, the question is: is it worthwhile to produce 2 tons h^{-1} less in order to have such an increase in the product quality? Of course, nobody else than the person in charge of the plant can answer. This shows how experimental design gives very good tools making easier to understand the problem; anyway, the final decision must rely on a practical basis.

7. Mixture design

A pharmaceutical company must find the best formulation for a tablet. The composition is the following: drug substance 4.0%, disintegrant 0.2%, lubricant A 4.0%, lubricant B 1.0%, ligands 90.8%. Three possible ligands are available: monohydrated alfa-lactose (X_1), anhydrous beta-lactose (X_2) and modified rice starch (X_3). The goal is to find the best mixture of the three ligands in order to have the best compromise between breaking load, i.e., the maximum force a tablet can sustain without breaking (Y_1), and dissolution rate (Y_2). In this case, it is logical to think that it would not be possible to optimize both responses at the same time, since higher breaking loads will correspond to lower dissolution rates (and vice versa).

Since the amount of the other components is fixed and cannot be changed, one can consider only the relative amount of the ligands, taking then into account that the sum of the ligands is 90.8% of the total.

In the cases studied till now (experimental designs for independent variables) each variable could be set at any value inside its range, independently of the value taken by the other variables. Instead, in a mixture we must cope with the implicit constraint that the sum of all the components must be 1 (or 100%). This means that the components of a mixture cannot be varied independently, since by varying the percentage of one component also the percentages of the other components will be changed.

Another relevant difference with the design for independent variables is that the object of the study in these problems is not the effect of the variation of the absolute quantity of the variables, but the effect of the variation of the ratios among the variables (a cocktail made by 10 mL of component A, 20 mL of component B and

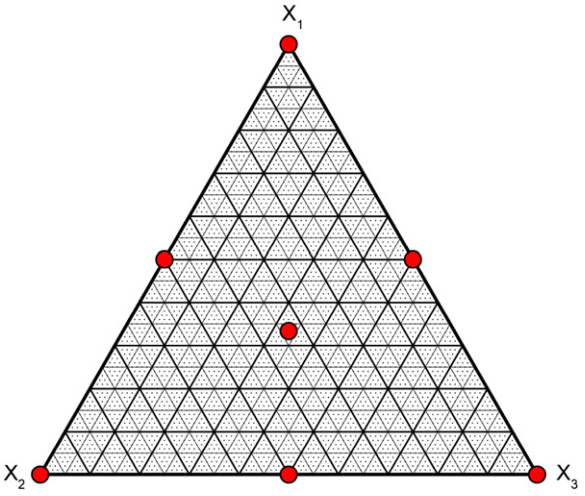

Fig. 12. The representation of a 3-component mixture, with the lines of the grid drawn at steps of 10% (dark solid lines), 5% (light solid lines) and 1% (dotted lines). The points of the 7 experiments of the mixture design are reported.

25 mL of component C will be exactly the same as a cocktail made by 1 L of A, 2 L of B and 2.5 L of C).

As a consequence of that, it is not possible to apply to the problems of mixtures the experimental designs previously studied.

Fig. 12 shows the graphical representation of a three-component mixture. It is an equilateral triangle, in which the vertices correspond to the pure components, the sides to the binary mixtures and the internal points to the ternary mixtures. In such a representation, there is a biunivocal correspondence between the plot and the composition, since to each composition corresponds one and only one point of the plot, and vice versa. The same holds for any number of components of a mixture, whose domain will be the regular figure having as many vertices as components, lying in the space having dimensionality equal to the number of components minus one (the equilateral triangle is the regular figure having three vertices in a two-dimensional space; the tetrahedron is the regular figure having four vertices in a three-dimensional space). For more than four components, we cannot visualize the whole domain since it lies in a space having more than three dimensions, but this is just a problem of our limited mind: from the mathematical point of view, the number of dimensions does not make any difference.

Table 6 reports the seven experiments that have been performed, together with the responses that have been obtained, measured on arbitrary units (the experiments are also highlighted in Fig. 12).

By doing these experiments, the coefficients of the following model can be estimated:

$$Y = b_1X_1 + b_2X_2 + b_3X_3 + b_{12}X_1X_2 + b_{13}X_1X_3 + b_{23}X_2X_3 + b_{123}X_1X_2X_3$$

Comparing it with the model for independent variables, it can immediately be seen that the constant is not present. This appears quite logical if we think that the constant corresponds to the response when all the variables have level 0. In the case of the designs for independent variables, it has a practical meaning, since it corresponds to the response at the center point. In the case of mixtures, since the sum of all the components must be 1, it is not possible to have a condition in which all the variables have level 0.

Table 6
The experimental matrix and the responses of the mixture design.

X_1	X_2	X_3	Y_1	Y_2
1	0	0	31	44
0	1	0	113	26
0	0	1	38	52
0.50	0.50	0	42	41
0.50	0	0.50	39	67
0	0.50	0.50	70	30
0.33	0.33	0.33	60	55

The model for the first response is:

$$Y = 31X_1 + 113X_2 + 38X_3 - 120X_1X_2 + 18X_1X_3 - 22X_2X_3 + 370X_1X_2X_3$$

It can be noticed that the coefficients of the linear terms correspond to the response obtained with the pure components.

The coefficients of the two-term interactions indicate the synergic effect of the two components. In the example described above, the experiment with pure X_1 gave a response of 31, while the experiment with pure X_2 gave a response of 113. If no synergic effect would be present, the mixture made by 0.5 X_1 and 0.5 X_2 would give a response of 72 (the average of 31 and 113). The response of this mixture is instead 42 (30 units less), meaning that a negative synergic effect is present. This can be found in the term X_1X_2, whose coefficient is −120. It can be seen that the magnitude of the synergic effects of the two-component mixtures is given by the coefficients divided by 4.

In the same way, the coefficient of the three-terms interaction, divided by 27, corresponds to the magnitude of the synergic effect of the three components (this is the reason why the coefficients of the higher interactions are usually very large).

From the isoresponse plot shown in Fig. 13 it can be seen that X_2 has the greatest effect on Y_1 (breaking load). To understand the effect of each variable one must follow how the response changes when going from a composition without that component to the pure component (without changing the relative amounts of the other components). On the plot, start from the center of the edge opposite to the component under study (i.e., a mixture made by the two remaining components, at 50% each) and go to the vertex representing the pure component. In the case of X_2, the response at the starting point (no X_2) is 39, and then it goes on increasing quite regularly up to 113 for pure X_2.

Doing the same for X_1, it can be seen that it has a negative effect: the response is 70 without it, and then regularly decreases down to 31 for pure X_1.

X_3 is the component with the lowest effect and, owing to the three-component interaction, it has a different behaviour: starting from 42 for the mixture made by 50% X_1 and 50% X_2, it goes down to 38 for pure X_3, but the addition of X_3 at first increases the response (up to 60 for the mixture made by 33.3% of each of the components), then decreases it.

The model for the second response is:

$$Y = 44X_1 + 26X_2 + 52X_3 + 24X_1X_2 + 76X_1X_3 - 36X_2X_3 + 206X_1X_2X_3$$

By looking at the coefficients of the interactions we can understand that the strongest synergic effect is the positive synergy between X_1 and X_3. By looking at the isoresponse plot of Fig. 13 it can be seen that the highest response corresponds to the mixture made approximately by 50% X_1 and 50% X_3.

Also for this response X_2 is the component with the highest effect, since the response decreases from 67 (no X_2) to 26 (pure X_2). When starting adding X_1 the response increases from 30 up to more than 55 (at approximately 50% of X_1), then it decreases to 44 (pure X_1). X_3 is the component with the smallest effect, since the response goes from 41 (no X_3) up to about 50 (at approximately 20% of X_3), to stay almost constant till the pure X_3, having a response of 52.

In the experimental designs for independent variables we saw that the effect of the factors and the relative importance of the

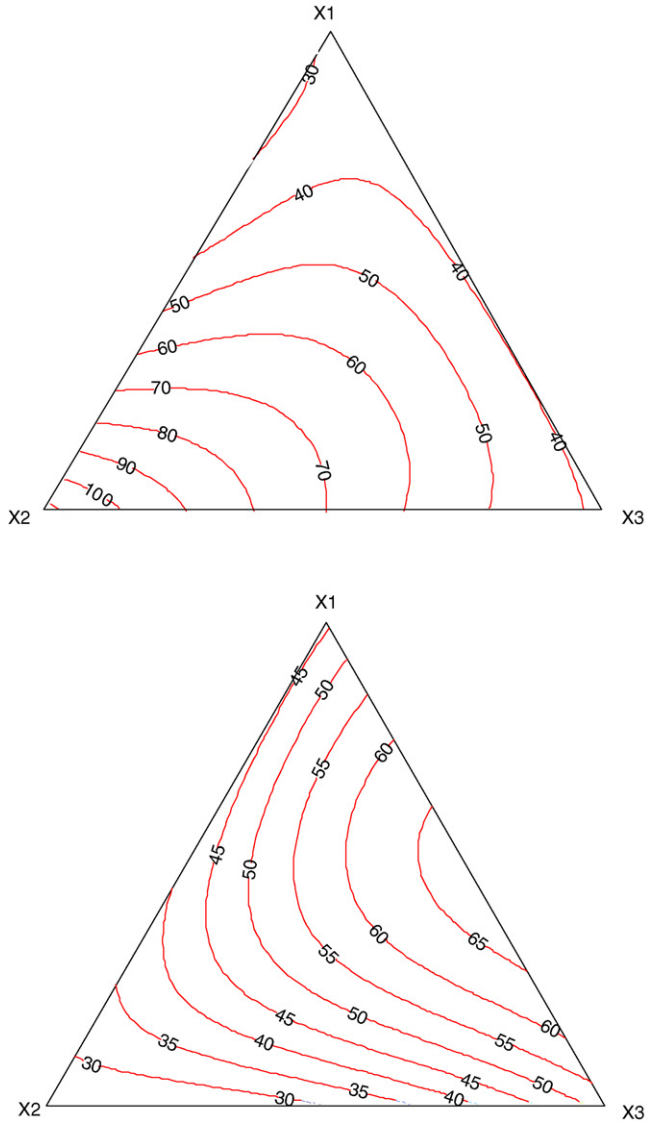

Fig. 13. Isoresponse plots of the two responses of the mixture design (top: breaking load, bottom: dissolution rate).

terms can be immediately understood just looking at the model (the higher the coefficient, the more important the term).

Unfortunately, this is no more true in the mixture designs. For instance, in the case of the second response of the previous example, it is clear from the isoresponse surface that X_2 is by far the component with the most relevant effect, but its coefficient is the smallest one! This because, as previously explained, the coefficients of the linear terms correspond to the responses with pure components, and therefore have nothing to do with the effects of the components. When going to the interactions, in order to give them a practical meaning, we must remember to divide them by 4 (in case of a two-term interaction) or 27 (for three-term interactions), or 256 (for four-term interactions).

As a consequence, the only way to understand the meaning of a model is to look at the isoresponse surfaces. This is easy in the case of three components (the domain is just a triangle), it becomes more complex but still possible in the case of four components (the domain is a tetrahedron, and we must take into account "slices" obtained by cutting it with planes parallel to one of the faces, which means keeping one of the components constant), and it is extremely difficult with more than four components (with n components, in

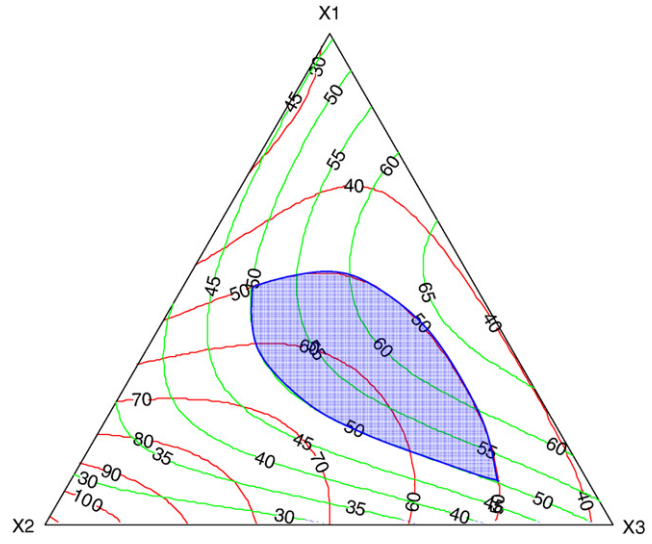

Fig. 14. Overlapped isoresponse plot of the two responses of the mixture design. The region of joint acceptability is highlighted.

order to have a plane, we should cut the experimental domain with hyperplanes obtained by keeping constant n-3 components).

This is the reason why the number of components under study in a mixture design usually is not higher than four. In case of a more complex system, it is suggested to split the whole domain into two or more subdomains, each of them with an "acceptable" number of components, and then to study each of them separately. Of course, by doing so, one has to realize that the interactions among the components of the different subdomains cannot be studied, but this is the price to be paid in order to increase the comprehensibility of the whole system.

Back to our problem, the pharmaceutical company was looking for the best compromise, with a minimum acceptable value of 50 for both responses. As expected, and as shown by the response surfaces, the two responses have opposite behaviour, and when one of them is increased the other one is decreased.

Fig. 14 shows the region of joint acceptability. Since it is totally internal to the domain, then only ternary mixtures can produce acceptable results. The producer can select the most convenient formulation on the basis of the responses but also taking into account the price of each of the components.

It has to be noticed that, apart from the pharmaceutical industry, mixture designs are extremely important in many fields, such as food products (e.g., formulation of a new recipe), beverages (e.g., formulation of a blend), paintings (e.g., formulation of a painting), method optimization (e.g., optimal mobile phase for HPLC).

8. Conclusions

The applications described in this paper clearly show the superiority of experimental design compared with the "classical" OVAT approach. In spite of the fact that its advantages are well known and well recognized since more than 70 years ago [21], experimental design is still applied only in a very small minority of the cases in which it could (and should) be used.

Therefore, one of the main efforts of the chemometric community should be to make it as known and as understandable as possible, in such a way that it could become the standard way of operating.

It is well recognized that the OVAT is not at all the correct way to perform an optimization or to study the effect of several variables, since it does not take into account the interactions among vari-

ables, it only gives a local knowledge of the phenomenon and often requires a much larger experimental effort; therefore, it would be advisable that the major journals would not accept anymore papers in which the OVAT approach is performed when an experimental design should instead be applied.

Acknowledgement

The author wishes to thank Paolo Oliveri for the careful revision of the manuscript and the many precious suggestions.

References

[1] G.E.P. Box, W.G. Hunter, J.S. Hunter, Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building, John Wiley & Sons, New York, 1978.
[2] R. Brereton, Chemometrics: Data Analysis for the Laboratory and Chemical Plant, John Wiley & Sons, New York, 1978.
[3] R. Carlson, J.E. Carlson, Design and Optimization in Organic Synthesis: Second Revised and Enlarged Edition, in: Data Handling in Science and Technology, vol. 24, Elsevier, Amsterdam, 2005.
[4] J.A. Cornell, Experiments with Mixtures: Designs, Models and the Analysis of Mixture Data, in: Series in Probability and Statistics, John Wiley & Sons, New York, 1991.
[5] R.E. Bruns, I.S. Scarminio, B. de Barros Neto, Statistical Design—Chemometrics, in: Data Handling in Science and Technology, vol. 25, Elsevier, Amsterdam, 2006.
[6] D.C. Montgomery, Design and Analysis of Experiments, 7th edition, John Wiley & Sons, Inc., 2009.
[7] T. Lundstedt, E. Seifert, L. Abramo, B. Thelin, Å. Nyström, J. Pettersen, R. Bergman, Chemolab 42 (1998) 3.
[8] Y. Vander Heyden, LC–GC Europe 19 (9) (2006) 469.
[9] C. Cortada, L. Vidal, S. Tejada, A. Romo, A. Canals, Anal. Chim. Acta 638 (1) (2009) 29.
[10] N. Negreira, I. Rodríguez, M. Ramil, E. Rubí, R. Cela, Anal. Chim. Acta 638 (1) (2009) 36.
[11] J. Schneider Carletto, R. Medeiros Luciano, G.C. Bedendo, E. Carasek, Anal. Chim. Acta 638 (1) (2009) 45.
[12] T. Alizadeh, M.R. Ganjali, P. Nourozi, M. Zare, Anal. Chim. Acta 638 (2) (2009) 154.
[13] S. Ozcan, A. Tor, M.E. Aydin, Anal. Chim. Acta 640 (1–2) (2009) 52.
[14] J.M. Conny, G.A. Norris, T.R. Gould, Anal. Chim. Acta 635 (2) (2009) 144.
[15] K. MacNamara, R. Leardi, F. McGuigan, Anal. Chim. Acta 636 (2) (2009) 190.
[16] W. de Jesus Rodrigues Santos, P. Rodrigues Lima, C.R. Teixeira Tarley, N.F. Höehr, L.T. Kubota, Anal. Chim. Acta 631 (1) (2009) 170.
[17] C.-W. Ye, J. Gao, C. Yang, X.-J. Liu, X.-J. Li, S.-Y. Pan, Anal. Chim. Acta 641 (1–2) (2009) 64.
[18] H. Liu, Y. Wen, F. Luan, Y. Gao, Anal. Chim. Acta 638 (1) (2009) 88.
[19] Y. Li, Y. Li, B. Zheng, L. Qu, C. Li, Anal. Chim. Acta 643 (1–2) (2009) 100.
[20] R. Leardi, F. Paganelli, K. MacNamara, Joint ENBIS-DEINDE 2007 Conference, Turin, 11–13 April 2007, Book of Abstracts, p. 37.
[21] R.A. Fisher, The Design of Experiments, Hafner Press, New York, 1935.