We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

BOOK CITATION INDEX
CLARIVATE ANALYTICS
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Factorial Design and Machine Learning Strategies: Impacts on Pharmaceutical Analysis

Marwa S. Elazazy

Additional information is available at the end of the chapter

Abstract

Pharmaceutical analysis is going through an expeditious progress as the perception of 'multivariate data analysis' (MVA) becomes gradually more assimilated. Pharmaceutical analysis comprises a range of processes that covers both chemical and physical assessment of drugs and their formulations employing different analytical techniques. With the revolution in instrumental analysis and the huge amount of information produced, there must be an up-to-date data processing tool. The role of chemometrics then comes up. Multivariate analysis (MVA) has the capability of effectively drawing a complete picture of the investigated process. Moreover, MVA reproduces the arithmetic influence of variables and their interactions through a smaller number of trials, keeping both efforts and capitals. Spectrophotometry is among the most extensively used techniques in pharmaceutical analysis either direct (single component) or derivative (multicomponent). In addition to these recognized benefits, using chemometrics in conjunction with spectrophotometry affects three vital characteristics: accuracy, precision and robustness. The impact of hyphenation of spectrophotometric analytical techniques to chemometrics (experimental design and support vector machines) on analytical laboratory will be revealed. A theoretical background on the different factorial designs and their relevance is provided. Readers will be able to use this chapter as a guide to select the appropriate design for a problem.

Keywords: chemometrics, experimental design, machine learning strategies, support vector machines, pharmaceutical analysis, spectrophotometry

1. Introduction

Nowadays, an enormous amount of information is being generated by the state-of-the art analytical instrumentations, an issue that necessitates the presence of a potent data processing approach. Chemometry, a division of science that has seen a major progress in the past

few decades, depends on eliciting data and the development of a mathematical model that describes the relationship between the response signal and the process variables [1–3]. In simple words, chemometrics is the term that is used to describe the case when chemistry, biology and other branches of science meet with mathematics and computer science [4]. As a multidisciplinary science, chemometrics can be used to resolve many problems beyond the boundaries of chemistry, including medicine, pharmacy, environment and other domains of natural and applied sciences [5, 6].

Chemometric techniques, including both multivariate data analysis (MVA) and factorial designs, play a vital role in analysing systems that are both large and multidimensional, an issue that adds to the power of this methodology. Moreover, the growing in complexity from the conventional univariate data analysis (one-variable and a single response at a time) to multivariate data analysis (more than one factor and a single or multiple responses) is greatly reflected on the imperative analytical outcomes, for example, sensitivity and selectivity [7, 8]. Additionally, being a versatile approach, application of chemometry can offer several more advantages. At the simple level (first order, vector data), samples that cannot be signalled using the existent calibration setting can now be effectively modelled. At more sophisticated levels (second- or higher orders), and in addition to the accurate determination of the calibrated analyte, not only new sample constituents can be identified but also their impact on the entire response can be adequately modelled.

Pharmaceutical analysis is experiencing an expeditious growth as the concept of 'multivariate data analysis' becomes progressively integrated. As being known, pharmaceutical analysis encompasses both chemical and physical evaluation of drugs and their dosage forms using different analytical strategies. Yet, the common routine in most of analytical laboratories is to meditate only one-variable and one response at time. Measuring the impact of this variable on the analytical signal is the only source of any generated data [1]. Nevertheless, quality of collected information would be significantly improved if the impact of more than one-variable, their linear, second- and third-order interactions on a single or multiple responses was defined through an arithmetic model [9].

Incorporation of 'design of experiments' (DOE) in any (or all) of the phases of drug development would be of a great effect, not only on the quality of data produced, but also on the analytical process itself in terms of better understanding and usage of generated data, as well as resources preservation.

This chapter focuses on the impact of using hyphenated chemometric-spectroscopic techniques in pharmaceutical analysis. Experimental designs as well as machine learning strategies, as essential parts of chemometrics, will be the main topic of the chapter. The reader does not need to be familiar with the complicated mathematical concepts. Rather, and for practicality and reader's advantageousness, a brief on the simple hypotheses needed to get DOE straightforward will be revealed.

Distinctive application of chemometrics in the field of drug analysis will be shown as we go forward. Material presented throughout the chapter will be of interest to students, chemometricians, drug manufacturers, quality control chemists and pharmacists.

2. Experimental design

Design of experiments (DOE) is a fundamental part of multivariate analysis techniques. However, DOE is comprehended to deal with a limited number of factors (determined according to the design used) in comparison to the other multivariate techniques.

Moreover, multivariate methods either bilinear such as partial least squares (PLS) and principal component analysis (PCA), or multi-way models such as Tucker-3 and parallel factor analysis (PFA), are commonly deemed as supplementary methodologies to DOE. Factors that were not considered in the initial set-up of DOE, as well as their effect, can now be recognized by the subsequent multivariate techniques [6, 10–12].

The typical scenario for setting DOE starts with deciding upon the experimental objective as well as the number of factors to be investigated. The most common objectives can be summarized as follows [13–16]:

- *Screening goal:* where all factors that might contribute to the response are considered and labelled as the main effects. Only factors proved to be significant will be considered for the second stage, which is known as optimization or fine tuning. In this phase, levels for each factor are adjusted to a narrower range to get the optimum response.

- *Response surface goal:* where main factors as well as factor-factor interactions (linear, quadratic, etc.) can be determined.

- *Optimization goal:* the experiment is designed in this case to get the best proportion for a factorial blend needed to get the optimum response (minimum or maximum).

Table 1 recaps the rules for selecting a design based on the number of factors and the envisioned goal of the experiment.

Up to now, the conventional approach for investigating the influence of several factors on a response depends on fixing the levels of all factors except the one to be investigated. This approach is known as one-variable at a time (OVAT). Although still being applied for analytical method development, OVAT usually confronts several difficulties.

One of the main limitations accompanying this rehearsal is the need for a big number of trials. Nevertheless, the resulting delineation of 'ideal conditions' and hereafter the system execution cannot be handled with a high extent of certainty. One reason for that is the absence of an evaluation for the variable-variable interactions in the paradigms premeditated using OVAT.

Number of factors	Screening goal	Response surface goal
Two to four factors	Full or fractional factorial designs (FFD)	Central composite (CCD) or Box-Behnken (BBD) designs
Five or more factors	Fractional factorial (FFD) or Plackett-Burman (PBD) designs	Preliminary assessment using the appropriate screening design is required to control the number of factors.

Table 1. Design selection rubric.

Multivariate data analysis (MVA) and its advantages mentioned earlier has the ability to replicate the arithmetical influence of the discrete factors and similarly their interactions through a reduced number of experimentations, saving both efforts and resources [16, 17].

The set-up of experimental design then can be viewed as 2–3 phases depending on the number of factors to be investigated and the objective of investigation: *screening*, *optimization* and *verification*.

2.1. Screening

Usually, a consecutive investigation process starts with testing a relatively large number of prospective variables. Screening designs then are factorial designs that can be used to get the few utmost substantial variables affecting the response, **Table 1**. Several designs can be used for this purpose, which are mentioned the following section.

2.1.1. Two-level full factorial design (2^k-FFD)

This design can be used when the number of variables (k) is between 2 and 15. Each variable is set at two levels: low (−1) and high (+1). Therefore, for three factors, for example, eight runs will be conducted excluding the central points and replicates. **Table 2** presents the design table when three factors X_1, X_2, and X_3 are investigated using the proposed two-level full factorial design (FFD). **Figure 1** shows the pattern of experiments in a design for three factors, arrows illustrate the direction of increase of the factors.

2.1.2. Two-level fractional factorial design (2^{k-p})

Even when the number of factors is small, many runs are needed if an FFD is to be used. For example, for five factors, $2^5 = 32$ experiments are needed in the base run only. In case replicates are needed and central points are added, the number of runs becomes large and the objective of using the DOE to save time and efforts becomes meaningless. The only way out for such a

Run order	X_1	X_2	X_3
1	−1	−1	−1
2	1	−1	−1
3	−1	1	−1
4	1	1	−1
5	−1	−1	1
6	1	−1	1
7	−1	1	1
8	1	1	1

Note: Runs are shown in standard order.

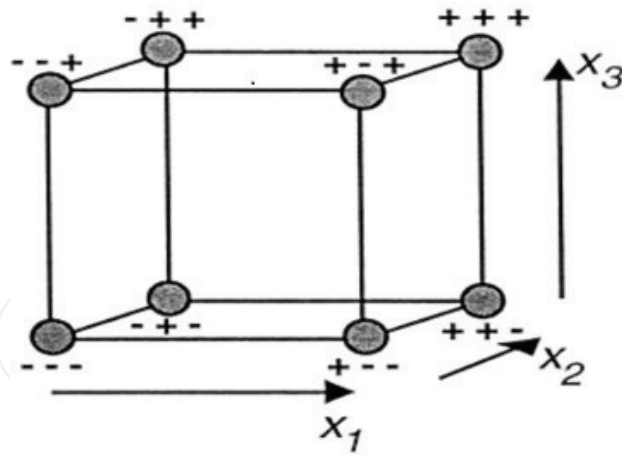**Table 2.** A two-level, full factorial design table for three factors.

Figure 1. Pattern of experiments in a 2^3 FFD.

case is to cautiously select a fraction (*p*) of the original runs proposed by the two-level FFD. For the previous example (3 factors), instead of performing 16 experiments (8 × 2 replicates) and by using a ½ fraction, only 8 runs will be performed in the 2 replicates.

Figure 2 shows a comparison between a full (2^k) and a fractional (2^{k-p}) factorial designs used to investigate three factors. While eight runs are needed in the first set-up, only four runs will be performed in the second arrangement, where main effects are confounded with the two-way interactions.

2.1.3. Plackett-Burman design (PBD)

This design has run numbers that are multiple of 4. Using this design allows performing a number of trials $N = 4n$ in order to investigate a number of factors $f = 4 (n - 1)$. PBD is an efficient approach when only main or large effects are of interest. In other words, this design can detect the most imperative factors affecting the experiment from a comparatively large number of factors (2–47) and without putting any concerns on interactions and non-linear effects. Minitab®, a commonly used software for this purpose, can generate a PBD for up to 47 factors.

PBD, in specific, is one of the commonly used approaches in robustness tests used in method validation compared to fractional factorial design, for example. The main reason for selecting PBD as a robustness test is that this design focuses only on the main effects, while factor-factor interactions are highly confounded with the large main effects, as previously mentioned [18–21].
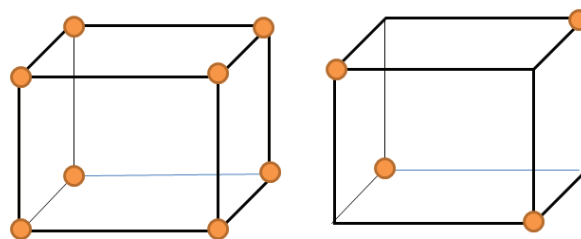

Figure 2. A 2^3 full factorial (left pane) and a 2^{3-1} fractional factorial designs (right pane) for three factors.

It is noteworthy to mention that, for any of the designs, identification of significant factors can be achieved using several tools. Pareto chart of standardized effects, normal and half-normal probability plots are among these tools.

2.2. Optimization

After selection of the most important factors from the previous screening process, levels of these factors need to be adjusted 'tuned' to identify the most suitable variable settings for optimizing a response. It is noteworthy to mention that significant factors can be also identified based on a former knowledge with the process under consideration. Another objective for this process is to assess the variable-variable linear interactions as well as the quadratic effects. This estimation gives an indication on how the response surface looks like. This approach is hence known as '*response surface methodology (RSM) designs*' [13].

Following the application of a response surface design, graphical representation of the developed polynomial mathematical model is assembled. Contour plots (2D) or response surface plots (3D) are used to graphically envisage the model.

2.2.1. Box-Behnken (BB) design

As a response surface design, BB design can capably determine the first- and second-order constants. BB design is simple, and independent with no contribution from a preceding factorial or fractional factorial design. Three levels for each factor are proposed; however, runs where all variables at their upper domains or all at lower domains are not included [22]. BB design is an economic choice since it involves less design points and hence a fewer number of runs compared to other RSM designs.

2.2.2. Central composite (CC) design

Unlike the BB design, CC designs usually contain in-built points from the factorial or fractional factorial designs (2^f trials) with added centre points that are enhanced with a group of axial points (2^f trials), **Figure 3**. Thus to scrutinize a number of factors = f, a number of experiments $N = 2^f + 2f + 1$ will be conducted. The design in such a configuration allows the estimation of data curvature. Furthermore, due to inclusion of data points from a prior screening design, CC design can be used in a consecutive experimental set-up. Classification of CC designs depends on the value of alpha (α) or the distance between the axial points and the centre. Three types of CC design then exist: *circumscribed (CCC)*, *inscribed (CCI)* and *face-centred (CCF)* [1, 13, 23–26].

2.3. Statistical validation

Following the last step, generated models can be statistically assessed using conventional approaches such as 'analysis of variance' (ANOVA). In this approach, variances are used to decide whether the means are different. For ANOVA to be properly conducted, the response variable has to be continuous and at least one of the investigated variables is categorical. For a factor to be significant, the *p*-value is usually less than α of 0.05 [1, 23–26].
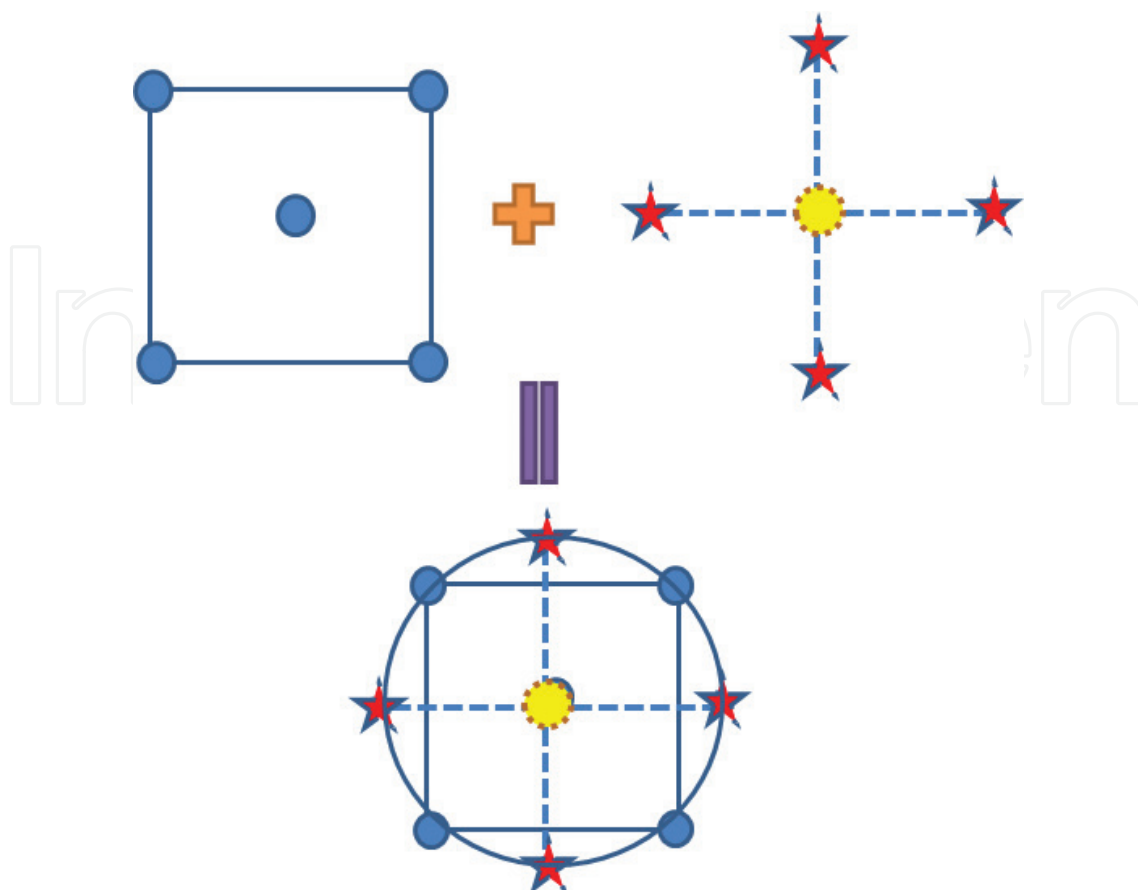
Figure 3. Central composite (CC) design for two factors.

Another model-fitting approach is the residual analysis. Residual plots are generally used to scrutinize the goodness of fit in regression and ANOVA. Examples of residual plots given by Minitab® include normal probability plots, residual versus fits, histograms and residuals versus order plots.

3. Support vector machines (SVMs)

SVM is a prevalent classification tool which was proposed by Vapnik [27]. As a kernel-based technique, support vector machines (SVMs) have seen a major development in the past few years. During such a short period, SVMs have found several applications in pharmacy, medicine and drug development industry. For example, SVMs have been used in finding the relation between drug structure and its activity 'structure-activity relationships (SAR)'. Moreover, SVMs with a capability of differentiating various drug substrates and classifying them as drugs or non-drugs are widely applied in drug design [28]. Fields of applications of SVMs extend to chemometrics, biosensors, computational biology and industrial modelling processes. Though being famous for the treatment of non-linear data, their application in handling linear models is still conceivable [27–32].

4. Pharmaceutical analysis and chemometrics

As mentioned earlier in this chapter, drug analysis covers all features related to both in- and after process (quality control) assay of drug substances. Details of these aspects include processes starting with drug synthesis, testing of physico-chemical properties, SAR and mechanism of drug action [28, 33, 34]. Quality control assays include stability testing of both raw and formulated drug materials, content homogeneity, solubility and dissolution properties. Nonetheless, drug assays are not circumscribed to the pure materials and the dosage forms, but the practice extends to include all complicated matrices (biological, foods, drinks, etc.). Moreover, analyses do not consider the active constituents only, but also look for the additives, degradation products and the impurities.

Different analytical techniques have been proposed for the determination of drugs (pure form, pharmaceutical formulations, biological fluids, etc.). For established drugs, standard analytical techniques can be obtained from compilations such as pharmacopoeias. The presence of almost daily new produces, however, requires constructing an appropriate analytical design. This design should inaugurate sufficient data on the analytical process and the product of concern. Data obtained should also be valid throughout the entire process of drug development and the procedure itself needs to be robust and applicable, when needed, in different laboratories.

These specifications do not mean that there is a need for a sophisticated technique such as chromatography. Yet, spectrophotometry might be an equivalent choice in the case being linked to an arithmetic backbone [16, 35–39]. Both single and multicomponent analyses (derivative spectrophotometry (DS)) can be readily linked to chemometry. Furthermore, analysis of a single response (e.g. absorbance) or multiple responses (at different wavelengths) can be better controlled using mathematical modelling [35–42].

Many challenges face the pharmaceutical analyst especially when trying to develop a new analytical method, inaugurate a drug stability study and establish automation into the laboratory. Handling these challenges using chemometrics will be revealed in the coming subsections.

Spectroscopic techniques have been used for long in pharmaceutical analysis. Ultraviolet and visible (UV-vis), infrared (IR), spectrofluorometry and near infrared (NIR) spectroscopy are among the most popular techniques in this concern. The application of techniques such as spectrophotometry in pharmaceutical analysis, though being simple, rapid, cost-effective and suitable for routine analysis, confronts many problems. A major problem that hinders the applicability of this technique is the lack of selectivity. Even in the analysis of a mixture of two or more components, the inability to select the most appropriate wavelength would have a negative impact on sensitivity, selectivity and reproducibility as well. Chromatography, though being a well-developed modern technique that is widely used in pharmaceutical analysis, suffers also from similar glitches. Inappropriate chemical deviations such as peaks from the matrix, alterations of mobile phase concentrations, baseline drift and shifts in retention times would greatly influence the cogency of the obtained results.

In both cases (and probably for other analytical techniques), the application of chemometrics to interpret the obtained data would be an ideal solution if the approach is able to account for all variations in the obtained data as well as get quantitative data from the tested samples. In addition, the used approach should be able to reduce the effects of these variations on the anticipated response.

In the coming subsections, we will consider the impacts of linking chemometry on pharmaceutical analytical techniques. More details will be given in the recent advances that have been made in this field and how spectrophotometry in specific has been affected.

4.1. Spectrophotometry

Spectrophotometric techniques are, as mentioned before, among the most widely used approaches in pharmaceutical analysis. Direct application of spectrophotometric analysis is only possible if the selected wavelength is not affected by another concomitant analyte. As an approach, application of spectrophotometry entails a study of a variety of factors affecting a single response or multiple responses [37–39].

With the advent of chemometrics, data processing programs and user-friendly software, the outdated OVAT approach is being gradually replaced with MVA in the analytical laboratories. In general, in addition to the known advantages of using chemometrics in conjunction with spectrophotometry, three crucial performance features are usually assessed with this hyphenation; accuracy, precision and robustness.

DOE and SVM are among the widely used chemometric approaches in spectrophotometric analysis of drugs and formulations. The main idea behind implementing these chemometric techniques is to establish the concept of thinking before doing, arrange and perform a controlled experiment, interpret the obtained results, and hence maximize the efficiency of used technique and obtained data. Generally, preservation of resources and conducting the fewest number of experiments are taken into consideration. This comprehensive knowledge and control of the running process are represented by a multi-aspect assembly of input variables together with method parameters, in other words, the 'design space'. The outcome of application of 'design space' is reflected on a pledge of quality as defined by International Conference on Harmonisation (ICH) tripartite rules [43].

As we mentioned earlier, DOE can be used in many stages of the pharmaceutical industry. For example, while screening designs can be used at the early stages of method development, optimization and testing of robustness are used just before the discharge of the finalized product [44].

Several other examples exist in the literature showing the application of DOE and SVM in the pharmaceutical industry. For instance, a two-level full factorial design (2^3-FFD) was used to decide upon the most substantial factors in the formulation of ascorbic acid tablets that are resistant to oxidative degradation using hydrophilic polymers. Measured responses were the tensile strength, disintegration time and the release features of these tablets [45]. In another application, Plackett-Burman design was employed to investigate the impact of seven factors on the release of theophylline from hydrophilic vehicles. According to the proposed model,

12 experiments were performed and a polynomial model was generated. Out of the seven variables, only two were proved to be significant [46].

In many cases of drug analysis, chemical pre-treatment of the analyte(s) prior to measurement of the anticipated response is sometimes needed. Usually, this preceding treatment would serve to correct for lack of sensitivity and selectivity encountered using direct spectrophotometry. Practices that are now ordinarily used in this concern are condensation, ion-pairing, charge transfer complexation, metal ion chelation, diazotization and redox reactions. With this pre-treatment, the process becomes technically more complicated and requires an investigation of a larger number of factors. A compelling solution in this case is provided by chemometrics. The literature now shows a huge amount of records on the hyphenation of factorial designs to spectrophotometric drug analysis, compared to the situation earlier.

For example, the Hantzsch condensation reaction was used for the derivatization of sodium alendronate, an inhibitor of bone resorption that is commonly used for management of osteoporosis, and which does not have any chromophore. Analysis of sodium alendronate was done both in its pure form and in oral solutions. Plackett-Burman screening design was used to investigate the effect of seven factors on the absorbance of the resulting condensation product. Only four factors were proved to be important and this finding was verified by ANOVA testing. Tuning of factors' levels was done using a circumscribed central composite design (CCCD). Moreover, data obtained from the CCCD including both variables and responses were treated with Statsoft® software employing artificial neuron network (ANN). A network of the multi-layer perceptron type (MLP) that has three hidden layer neurons gave the best results. Similarly, data from the CCCD were processed using different SVM kernels. Best results were obtained using a radial-basis function (RBF) kernel [37].

Chemical derivatization of midodrine hydrochloride both as per se and in formulations (tablets and oral drops) was performed using the Hantzsch reaction accompanied by a two-level 2^4-FFD. Variables proved to be significant ($p < 0.05$) were warily attuned utilizing a response surface methodology (RSM) with a face-centred central composite design. The suggested model represented a perfect example for probing the efficiency of factorial designs in optimizing the reaction conditions and maximizing the output [38]. Statistical validation of the proposed technique was performed by using ANOVA in two successive steps. Moreover, D-optimality design was chosen to minimize the variance in the regression coefficients of the fitted model. **Table 3** shows the screened factors and the response domains employing the proposed screening design.

A suitable approach in finding the most significant variables for screening designs and the optimal locations following an optimization design is usually the graphical representation of the data or the generated model. This feature is usually implemented in chemometrics' software such as Statsoft® and Minitab®. The outcome of screening designs is customarily represented by the Pareto chart of standardized effects, where factors passing the reference line are considered significant. Similar conclusions can be drawn using normal and half-normal probability plots. **Figure 4** shows a Pareto chart showing the significant factors obtained after screening of all factors affecting the formation of a charge transfer complex between p-synephrine and p-chloranil employing a full factorial design.

Screened factor	Symbol	Level		Maximum absorbance of the product (Y)
		Low (−)	High (+)	
Temperature (°C)	X_1	25.0	100.00	0.602
Reaction time (min.)	X_2	5.00	30.00	0.495
Reagent volume (mL)	X_3	0.10	1.00	0.489
pH of acetate buffer	X_4	2.40	5.60	0.493
Response	Y	Target		

Table 3. Screened factors and response domains for a two-level (2^4) full factorial design (FFD) premeditated for Hantzsch reaction (reproduced from author's own work [38] with permission from the Royal Society of Chemistry).
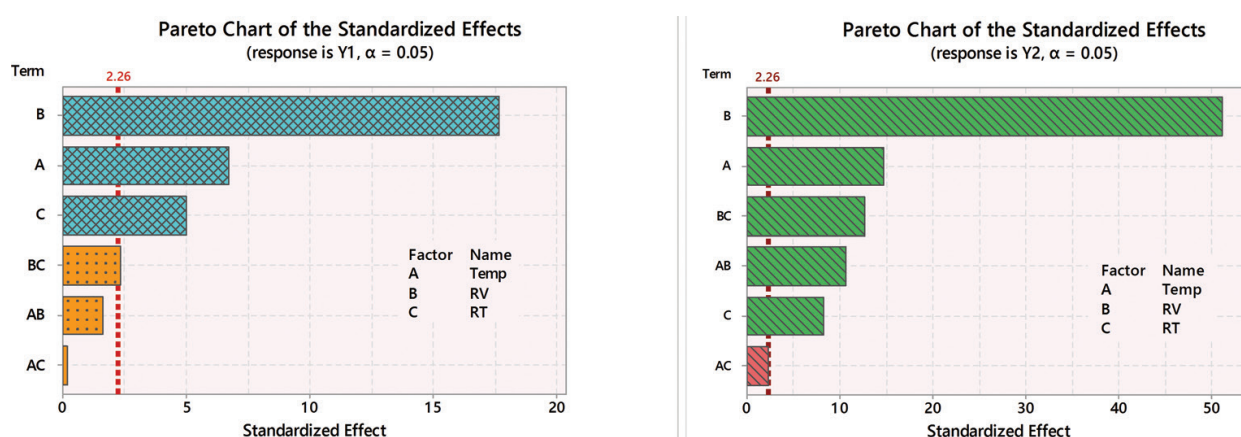

Figure 4. Pareto chart of standardized effects (reproduced from author's own work [39] with permission from the Royal Society of Chemistry).

Two types of graphs are commonly used to 'pinpoint' the optimal conditions; the response surface (3D) and contour (2D) plots. As shown in **Figure 5** [39], contour lines are produced when points that have the same absorbance are connected. On the other hand, 3D surface plots (figure is not shown) provide a stronger idea on interactions compared to contour plots. Both representations reveal a good matching with the obtained results, employing the polynomial equation.

Analysing one response is a simple task where analysis of each paradigm would merely identify zones of anticipated results. Conversely, concurrent optimization of two or more responses as a function of *n* variables is not that plausible. Different strategies are usually followed for this purpose; overlaid contour plots and global desirability function are among the commonly used approaches [39].

Overlaid contour plots are executed only if few responses are of concern (usually two responses). Simply, higher and lower bounds for each response are outlined. Contours for response boundaries versus variables under analysis are then displayed. A region that ensures both responses is recognized as the 'feasible' area [47, 48]. The plot usually shows the feasible regions where compromised optimum values for both responses meet. However, when more than one factor is involved and considering more than one response, a large number of graphs are requested, an issue that makes the procedure of pictorial observation tiresome.
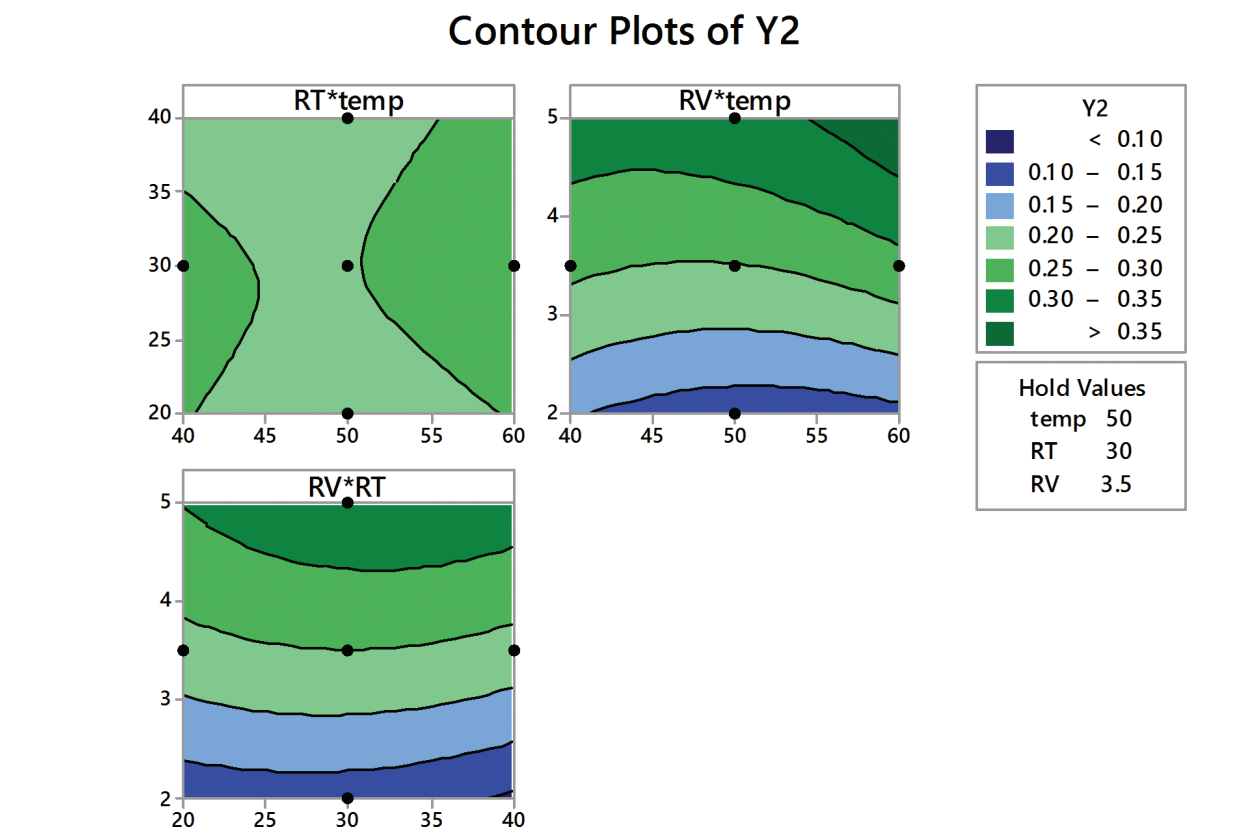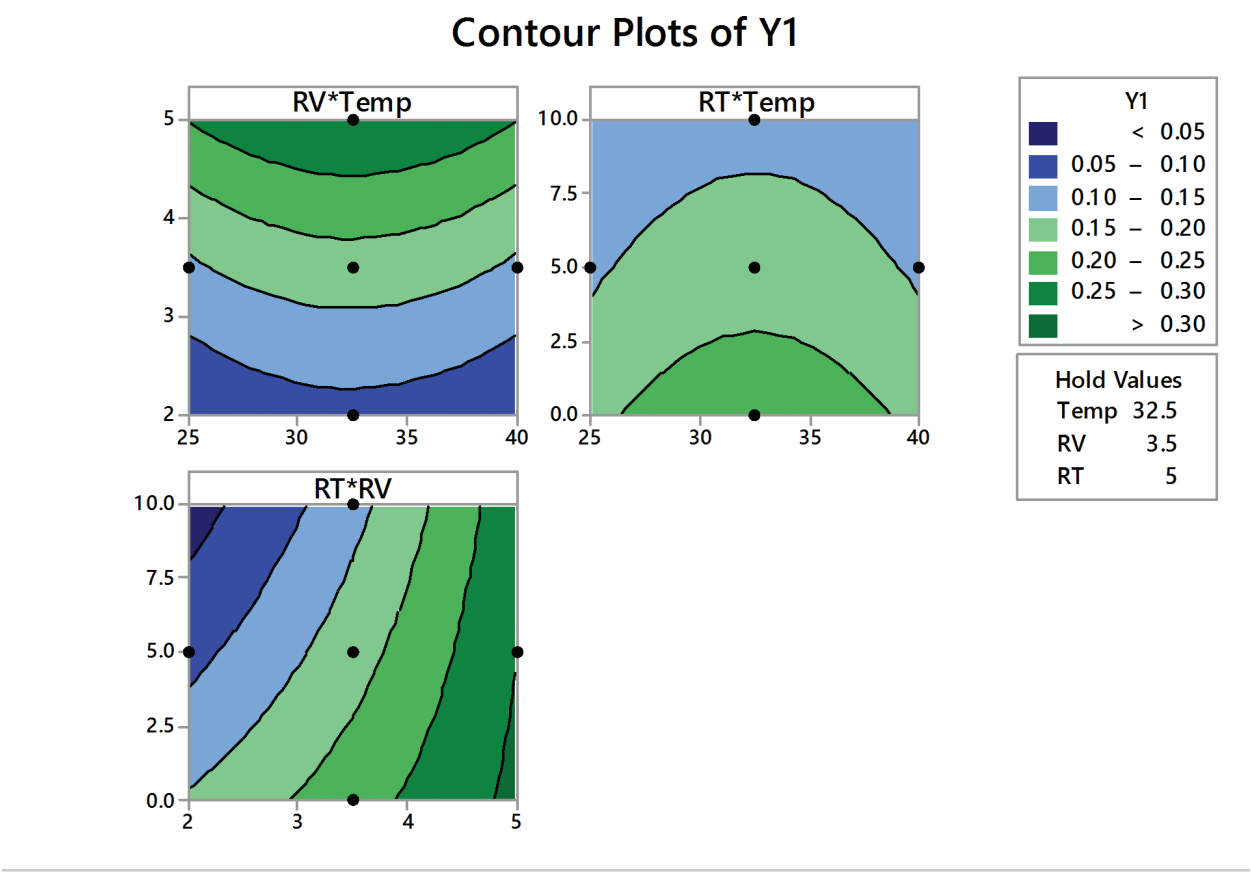
Contour Plots of Y1

Contour Plots of Y2

Figure 5. Two-dimensional contour plots for FCCD showing Y1 and Y2 as a function of different variable interactions (reproduced from author's own work [39] with permission from the Royal Society of Chemistry).

Additionally, the overlaying process is not that practicable as the best regions for each response are a bit far from each other.

Derringer function is another approach that can be used in this case. Individual desirability for each response is used to calculate the global desirability employing the following function:

$$D = (d_1^{r1} \ d_2^{r2}....d_m^{rm})^{1/\Sigma ri} = \left(\Pi_{i=1}^{n} d_i^{ri}\right)^{\frac{1}{\Sigma ri}} \tag{1}$$

where D is the overall desirability, d is the single desirability, r is the significance of each response compared to the other and m is the number of responses to be optimized [49, 50]. In general, as the value of D gets closer to 1.0000, the desirability of this variable arrangement on the proposed response gets higher. **Figure 6** shows the desirability function plot following the optimization employing an FCCD approach. The horizontal dashed lines represent current response values. The vertical solid lines show the optimal value for each variable.

A serious drawback that hinders drawing useful data, either assessable or qualitative, from spectrophotometry is the overlapping of absorption bands. This overlapping might be arising
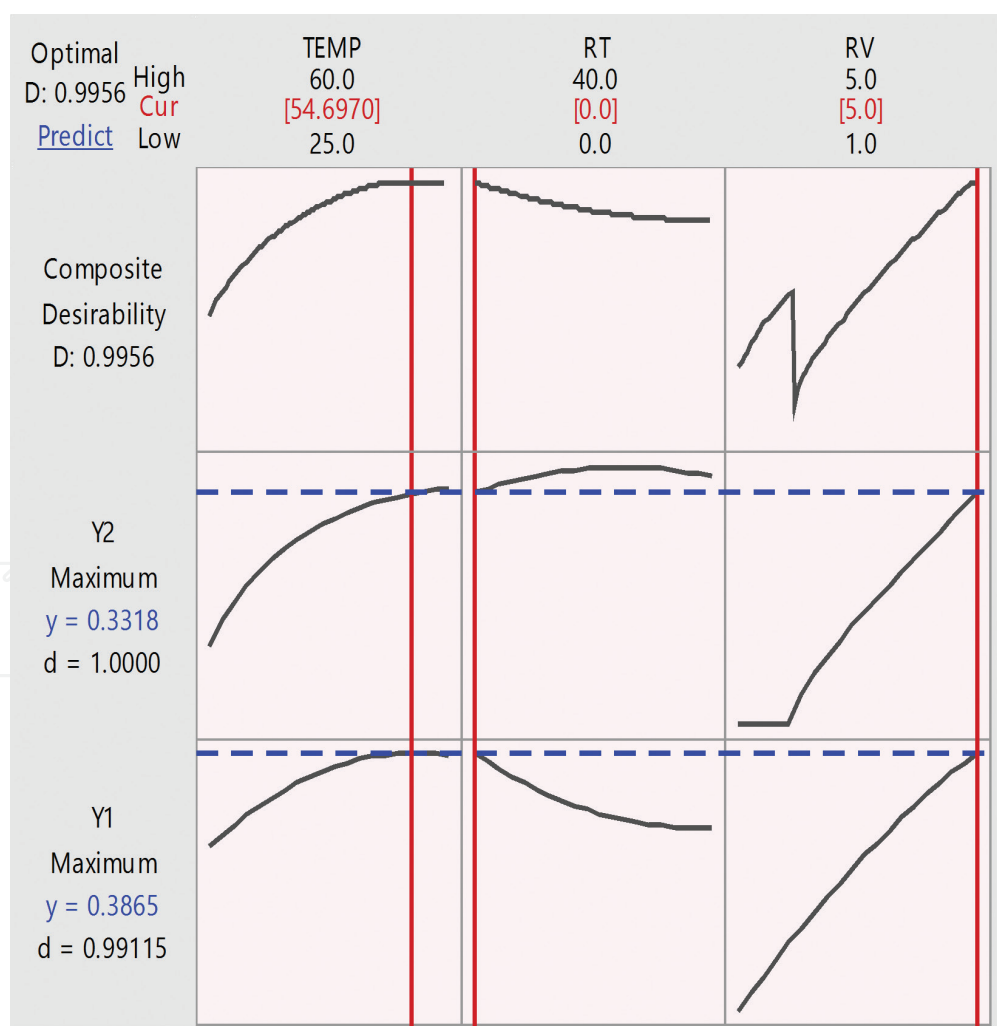

Figure 6. Desirability function plot for the FCC design (reproduced from author's own work [39] with permission from the Royal Society of Chemistry).

from the presence of drug or non-drug impurity, the presence of more than one component in the target formulation or due to the presence of degradation products. The presence of these components in one formulation at unequal concentration levels augments the problem. A compulsive solution to this problem is using derivative spectrophotometry (DS). This approach depends on differentiation of the regular absorption spectrum using arithmetical transformation into a first-order derivative or a higher order derivative. Several advantages are achieved using DS including but not limited to an improvement in resolution, reduction of noise level, elimination of interferences, augmentation of sensitivity and selectivity, and accordingly an improvement in separation efficiency [51–54].

The situation is not complicated if no chemical interaction among the components, and their spectra are only partially overlapped. In such a case, an acceptable resolution can be achieved employing first derivative spectra. Depending on the spectral characteristics of components to be analysed and the nature of interventions in multicomponent samples, chemometric algorithms have been proved to be a powerful tool in resolving binary (or more) mixture. Approaches such as principal component regression (PCR) and partial least squares (PLSs) have been widely applied both for zero- or higher- order spectra. A combination of MVA and derivative spectral data is highly beneficial where features such as easiness of application and reliability of obtained results are greatly improved [55–58].

5. Conclusion

Pharmaceutical analysis involves generation of a large amount of data. A pharmaceutical analyst then has an apparently intimidating task and needs to choose from a plethora of methods for handling the obtained data.

Chemometry has started to realize its potential. Assimilation of chemometric modelling (experimental design, artificial neuron networking, support vector machines, principal component analysis, etc.) to different analytical methods (spectrophotometry, chromatography, etc.) with the purpose of optimizing the analytical objectives is the novel trend followed by researchers nowadays. For every analytical process, the principal role of the analyst is to optimally obtain informative data. Unfortunately, best usage of data cannot be accomplished using the traditional univariate analysis. Multivariate analysis, in contrary, would be the golden solution, where a reasonable amount of information would be obtained through a fewer number of experiments, reduced effort and smaller amount of chemicals. As such, application of 'design of experiments (DOE)' becomes a need, and integration of DOE in any analytical procedure would be a must.

Author details

Marwa S. Elazazy

Address all correspondence to: marwasaid@qu.edu.qa

Department of Chemistry and Earth Sciences, College of Arts and Sciences, Qatar University, Doha, Qatar

References

[1]   Massart DL, Vandeginste BGM, Deming SN, Michotte Y, Kaufman L. Chemometrics: A Textbook. Amsterdam: Elsevier; 1988. Chapter 2

[2]   Wold S. Chemometrics, why, what and where to next? Journal of Pharmaceutical and Biomedical Analysis. 1991;**9**(8):589-596

[3]   Mocák J. Chemometrics in Medicine and Pharmacy. Nova Biotechnologica et Chimica. 2012;**11**(1): 11-25

[4]   Lopes JA, Costa PF, Alves TP, Menezes JC. Chemometrics in bioprocess engineering: Process analytical technology (PAT) applications. Chemometrics and Intelligent Laboratory Systems. 2004;**74**:269-275

[5]   Krantz-Rülcker C, Stenberg M, Winquist F, Lundström I. Electronic tongues for environmental monitoring based on sensor arrays and pattern recognition: A review. Analytica Chimica Acta. 2001;**426**(2):217-226

[6]   Singh I, Juneja P, Kaur B, Kumar P. Pharmaceutical applications of chemometric techniques. ISRN Analytical Chemistry. 2013;**2013**:1-13

[7]   Miller CE. Chemometrics in process analytical chemistry. In: Bakeev KA, editor. Process Analytical Technology. Oxford, UK: Blackwell Publishing Ltd.; 2005

[8]   Olivieri AC. Perspective analytical advantages of multivariate data processing. One, Two, Three, Infinity? Analytical Chemistry. 2008;**80**:5713-5720

[9]   Fisher RA. The Design of Experiments. New York: Haffner Press; 1935

[10] Bro R. Multivariate calibration what is in chemometrics for the analytical chemist? Analytica Chimica Acta. 2003;**500**:185-194

[11] Martens H, Martens M. Multivariate Analysis of Quality: An Introduction. Chichester, UK: Wiley; 2000

[12] Huang J, Kaul G, Cai C, Chatlapalli R, Hernandez-Abad P, Ghosh K, Nagi A. Quality by design case study: An integrated multivariate approach to drug product and process development. International Journal of Pharmaceutics. 2009;**382**:23-32

[13] Box GEP, Draper NR. Response Surfaces, Mixtures, and Ridge Analyses. 2nd ed. Hoboken, NJ, USA: Wiley; 2007. ISBN 978-0-470-05357-7

[14] Bruns R, Scarmiano I, Neto B. Statistical Design — Chemometrics. 1st ed. Elsevier Science; Volume 25 (Data Handling in Science and Technology); Amsterdam: Elsevier; 2006

[15] Carlson R. Design and Optimization in Organic Synthesis. 3rd ed. Amsterdam: Elsevier; 1991

[16] Leardi R. Experimental design in chemistry: A tutorial. Analytica Chimica Acta. 2009; **652**:161-172

[17] Eiroa AA, Diévart P, Dagaut P. Improved optimization of polycyclic aromatic hydrocarbons (PAHs) mixtures resolution in reversed-phase high-performance liquid chromatography by using factorial design and response surface methodology. Talanta. 2010;**81**:265-274

[18] Plackett RL, Burman JP. The design of optimum multifactorial experiments. Biometrika. 1946;**33**:305-325

[19] Morgan E. Chemometrics: Experimental Design. Analytical Chemistry by Open Learning. Chichester: Wiley; 1991. pp. 118-188

[20] Box GEP, Hunter W, Hunter J. Statistics for Experimenters, An Introduction to Design, Data Analysis and Model Building. New York: Wiley; 1978. pp. 306-418

[21] Vander Heyden Y, Massart DL. Review of the use of robustness and ruggedness in analytical chemistry. In: Smilde A, de Boer J, Hendriks M, editors. Robustness of Analytical Methods and Pharmaceutical Technological Products. Amsterdam: Elsevier; 1996. pp. 79-147

[22] Box GEP, Behnken DW. Some new three level designs for the study of quantitative variables. Technometrics. 1960;**2**:455-475

[23] Dejaegher B, Vander Heyden Y. The use of experimental design in separation science. Acta Chromatographia. 2009;**21**:161-201

[24] Dejaegher B, Durand A, Vander Heyden Y. Experimental design in method optimization and robustness testing. In: Hanrahan G, Gomez FA, editors. Chemometric Methods in Capillary Electrophoresis. New Jersey: John Wiley & Sons; 2010. pp. 11-74

[25] Montgomery DC. Design and Analysis of Experiments. 4th ed. New York: John Wiley; 1997

[26] Lewis GA, Mathieu D, Phan-Tan-Luu R. Pharmaceutical Experimental Design. New York: Marcel Dekker; 1999

[27] Vapnik V. The Nature of Statistical Learning Theory. New York: Springer; 2000

[28] Korkmaz S, Zararsiz G, Goksuluk D. Drug/nondrug classification using support vector machines with various feature selection strategies. Computer Methods and Programs in Biomedicine. 2014;**117**:51-60

[29] Zararsiz G, Elmali F, Ozturk A. Bagging support vector machines for leukaemia classification. International Journal of Computer Science Issues. 2012;**9**:355-358

[30] Ivanciuc O. Applications of support vector machines in chemistry. In: Lipkowitz KB, Cundari TR, editors. Reviews in Computational Chemistry, Weinheim: Wiley-VCH; 2007. pp. 291-400

[31] Hamel L. Support vector machines. In: Larose DT, editor. Knowledge Discovery with Support Vector Machines. Hoboken, New Jersey, USA, John Wiley & Sons, Inc; 2009, pp. 89-132

[32] Naguib IA, Abdelaleem EA, Draz ME, Zaazaa HE. Linear support vector regression and partial least squares chemometric models for determination of hydrochlorothiazide and benazepril hydrochloride in presence of related impurities: A comparative study. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy. 2014;**130**:350-356

[33] Puzyn T, Leszczynski J, Cronin MT, editors. Recent Advances in QSAR Studies. Methods and Applications. Heidelberg, Germany: Springer; 2010. p. 414

[34] Merz KM, Ringe D, Reynolds CH. Drug Design: Structure- and Ligand-Based Approaches. New York: Cambridge University Press; 2010

[35] Hamel L. Support vector machines. In: Larose DT, editor. Knowledge Discovery with Support Vector Machines. Hoboken, New Jersey, USA, John Wiley & Sons, Inc; 2009, pp. 89-132

[36] Berridge JC, Jones P, Roberts-Mcintosh AS. Chemometrics in pharmaceutical analysis. Journal of Pharmaceutical and Biomedical Analysis. 199l;**9**:597-604

[37] Korany MA, Ragab MAA, Youssef RM, Afify MA. Experimental design and machine learning strategies for parameters screening and optimization of Hantzsch condensation reaction for the assay of sodium alendronate in oral solution. RSC Advances. 2015;**5**:6385-6394

[38] Elazazy MS. Determination of midodrine hydrochloride via Hantzsch condensation reaction: A factorial design based spectrophotometric approach. RSC Advances. 2015;**5**:48474-48483

[39] Elazazy MS, Ganesh K, Sivakumar V, Huessein YHA. Interaction of p-synephrine with p-chloranil: Experimental design and multiple response optimization. RSC Advances. 2016;**6**:64967-64976

[40] Boeris MS, Luco JM, Olsina RA. Simultaneous spectrophotometric determination of phenobarbital, phenytoin and methyl phenobarbital in pharmaceutical preparations by using partial least-squares and principal component regression multivariate calibration. Journal of Pharmaceutical and Biomedical Analysis. 2000;**24**:259-271

[41] Berridge JC. Chemometrics and method development in high-performance liquid chromatography. Part 1: Introduction. Chemometrics and Intelligent Laboratory Systems. 1988;**3:**175-188

[42] Berridge JC. Chemometrics and method development in high-performance liquid chromatography. Part 2: Sequential experimental designs. Chemometrics and Intelligent Laboratory Systems. 1989;**5**:195-207

[43] ICH, 2005. Q2 (R1), Validation of analytical procedures: text and methodology, ICH Harmonised Tripartite Guideline. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, Chicago, USA, 2005

[44] Eriksson L, Johansson E, Kettaneh-Wold N, Wikström C, Wold S. Design of Experiments: Principles and Applications. 3rd ed. Umeå, Sweden: MKS Umetrics AB; 2008

[45] Odeniyi MA, Jaiyeoba KT. Optimization of ascorbic acid tablet formulations containing hydrophilic polymers. Farmacia. 2009;**57**:157-166

[46] El-Malah Y, Nazzal S. Hydrophilic matrices: application of Placket-Burman screening design to model the effect of POLYOX–carbopol blends on drug release. International Journal of Pharmaceutics. 2006;**309**:163-170

[47] Sivertsen E, Bjerke F, Almøy T, Segtnan V, Næs T. Multivariate optimization by visual inspection. Chemometrics and Intelligent Laboratory Systems. 2007;**85**:110-118

[48] Hasniyati MR, Zuhailawati H, Sivakumar R, Dhindaw BK. Optimization of multiple responses using overlaid contour plot and steepest methods analysis on hydroxyapatite coated magnesium via cold spray deposition coated magnesium via cold spray deposition. Surface Coatings and Technology. 2015;**280**:250-255

[49] Derringer G, Suich R. Simultaneous optimization of several response variables. Journal of Quality Technology. 1980;**12**:214-219

[50] Minitab 17 Statistical Software. Computer software; 2010. State College, PA: Minitab, Inc. (www.minitab.com)

[51] Talsky G. Derivative Spectrophotometry. 1st ed. Weinheim: VCH; 1994

[52] Sanchez Rojas F, Ojeda CB. Recent development in derivative ultraviolet/visible absorption spectrophotometry: 2009-2011 a review. Microchemical Journal. 2013;**106**:1-16

[53] Hasan NY, Abdel-Elkawy M, Elzeany BE, Wagieh NE. Stability indicating methods for the determination of aceclofenac. Il Farmaco. 2003;**58**:91-99

[54] El-Saharty YS, Refaat M, El-Khateeb SZ. Stability-indicating spectrophotometric and densitometric methods for determination of aceclofenac. Drug Development and Industrial Pharmacy. 2002;**28**:571-582

[55] De Luca M, Oliverio F, Ioele G, Ragno G. Multivariate calibration techniques applied to derivative spectroscopy data for the analysis of pharmaceutical mixtures. Chemometrics and Intelligent Laboratory Systems. 2009;**96**:14-21

[56] Bautista RD, Jiménez AI, Jiménez F, Arias JJ. Simultaneous determination of drugs in concentration ratios above 40 1 by application of multivariate calibration to absorbency and derivative spectrophotometric signals. Fresenius Journal of Analytical Chemistry. 1997;**357**:449-456

[57] Dinç E, Üstündağ O. Chemometric resolution of a mixture containing hydrochlorothiazide and amiloride by absorption and derivative spectrophotometry. Journal of Pharmaceutical Biomedical Analysis. 2002;**29**:371-379

[58] Brown C, Vega-Montoto L, Wentzell P. Derivative preprocessing and optimal corrections for baseline drift in multivariate calibration. Applied Spectroscopy. 2000;**54**:1055-1068