I N S T I T U T D E
S T A T I S T I Q U E

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N
P A P E R

0710

DEVELOPMENT OF A NEW PREDICTIVE MODELLING TECHNIQUE TO FIND WITH CONFIDENCE EQUIVALENCE ZONE AND DESIGN SPACE OF CHROMATOGRAPHIC ANALYTICAL METHODS

LEBRUN, P., GOVAERTS, B., DEBRUS, B., CECCATO, A., CALIARO, G., HUBERT, P. and B. BOULANGER

Development of a new predictive modelling technique to find with confidence equivalence zone and design space of chromatographic analytical methods

Pierre Lebrun [a,*], Bernadette Govaerts [b], Benjamin Debrus [a], Attilio Ceccato [e], Gabriel Caliaro [d], Philippe Hubert [a], Bruno Boulanger [c]

[a] *Université de Liège, Institute of Pharmacy, Laboratory of Analytical Chemistry, Av. de l'Hopital 1, Bât. B36 , 4000 Liège , Belgium*

[b] *Université catholique de Louvain, Institute of Statistics, Voie du Roman Pays 20, 1348 Louvain-la-Neuve, Belgium*

[c] *UCB Pharma SA, R&D Clinical Pharmacometrics, Belgium*

[d] *Orailac Quality Solutions, Belgium*

[e] *GlaxoSmithKline Biologicals, Belgium*

Abstract

A new method for modelling chromatographic responses is presented as a critical piece for the achievement of automated development of analytical methods. This methodology is based on four parts. First, we propose to use a very little set of statistical equations to create predictive models for retention time based responses as the apex, the width and the asymmetry of peaks. Second, an experimental design is set up to realize experiments. Third, using grid search over the domain, multi criteria decision is taken with respect to different local or global optimization criteria, used as desirability functions. This allows finding an optimal chromatogram. Fourth, we advice to investigate how the predictive error of the models propagates around optimal solution. This allows to give confidence in the optimal solution, in finding a set of zones that presumably will give an acceptable solution. Design spaces can be derived with a similar technique. The approach is exemplified with a real case and predictions of models at optimal analytical conditions are validated through new experiments. Flexibility is left over all the presented methodology.

Key words: HPLC, chromatography, design of experiments, multi-criteria optimization, statistical models, error propagation, design space

* Corresponding author.
Email address: plebrun@ulg.ac.be (Pierre Lebrun).

1 Introduction

In analytical chemistry, the chromatographic techniques are widely used in different fields of activity such as chemical, pharmaceutical, biomedical, environmental and food analysis. Thus, the selection of the most appropriate experimental conditions allowing the separation of compounds of interest in various matrices is a matter of a very particular interest. Pharmaceutical industries are of course concerned by these problems and are more especially interested by all new approach allowing to separate their compounds properly and quickly in order to quantify them. Indeed, the analytical step is a crucial phase during the development of new drugs since the different decisions are taken based on results generated by one or more analytical methods. Amongst the chromatographic techniques, Liquid Chromatography (LC) is probably the most common technique to fulfil this objective.

Nowadays, the development of analytical methods in LC is still time consuming and not always under the perfect control of analysts. This is due to the fact that most of the parameters to manage to obtain acceptable separation conditions have complex effects on the chromatogram. The problem becomes even more complicated when the matrix is complex and contains many compounds with physico-chemical properties that are not necessarily known.

A lot of contributions from various authors about optimization and methods development have been reported over the last 20 years. The use of design of experiments to find optimal conditions is now largely accepted in chromatography. However, it is often problematic to take into account the total complexity of the problems and it is probably the reason why actual softwares in chromatography still don't allow the optimization with many parameters.

Schoenmakers (1986) reported formulations to compute global or limited optimization criteria used for optimization [1]. Massart (1990) and Snyder (1997) illustrated different tools for methods development [2,3]. Vanbel (1998) summarized the need of adequate and flexible optimization criteria, adapted to practical situations, and the need to provide various and flexible experimental designs and modelling equations [4]. Dewé *et al.* (2004) also proposed a methodology to optimize several analytical conditions [5]. Both last references agree on the fact that modelling directly chromatographic criteria to find some optimum configuration is not good practice due to discontinuities in modelled responses. It is proposed to model retention times as characteristic of chromatograms instead of criteria. After, these criteria are derived from the retention times and a multi-criteria optimization can be envisaged. The main problems encountered in these methodologies is the assumed independence between responses. This leads to many equations, which is not recommendable, and errors can occur (see later). Finally, Vanbel and Dewé *et al.* completely ignore error of the models in their predicted optima. In this paper, these methodologies are generalized, allowing to optimize in an automated way as many parameters

and criteria as needed by the analysis.

Changing parameters of the LC leads to different retention times for one analyte in the output chromatograms, as shown in Figure 1. The purpose of this work is that, following the practice of design of experiments [6,7], it is possible to find some optimal parameters configurations, considering different criteria computed on chromatograms, such as the minimal separation, the total retention time or the asymmetry of the peaks, for instance.

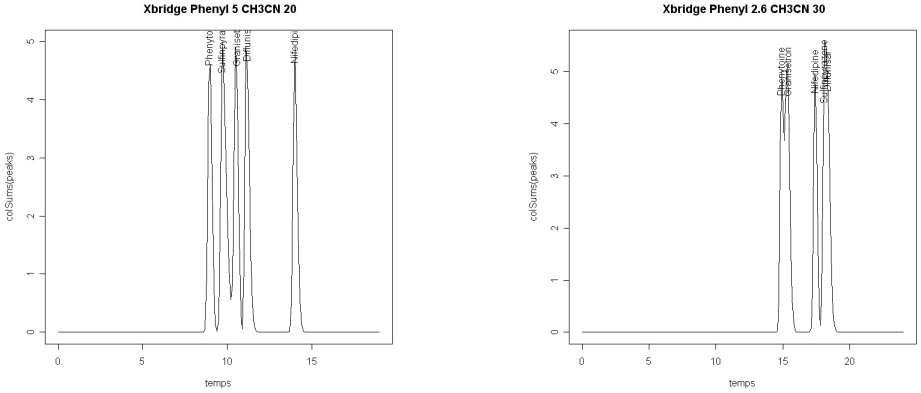

Figure 1. Example: different analytical conditions lead to different retention times for the analytes (reconstructed chromatograms, method from Dewé *et al.* (2004) [5]). Chromatographic parameters are the nature of the stationary phase, the mobile phase pH, the nature of the organic modifiers and the gradient slopes (min.) used in analysis. Left: Column XBridge Phenyl with acetonitrile in the buffer, pH 5 and, gradient 5%-95% of organic modifier realized in 20 minutes. Right: same column and solvent, pH 2.6, and gradient in 30 minutes

2 Objectives and main steps of the methodology

We present a methodology to predict, with a known level of confidence, the best *tuning parameters* of a HPLC in the range of possible analytical conditions in order to get the *best* chromatogram possible over the domain of potential mixtures. This global target can be subdivided into several smaller objectives.

The first objective aims at developing, from the results of experiments designed for this very purpose, predictive models of the peak retention times with respect to the HPLC tuning parameters. The second objective is to use these predictive models in order to find optimal chromatographic analytical conditions with respect to different criteria to characterize the quality of a chromatogram. This is a multicriteria optimization problem. The third objective aims at investigating the confidence of the predicted optimal conditions through a Monte-Carlo study of the propagation of

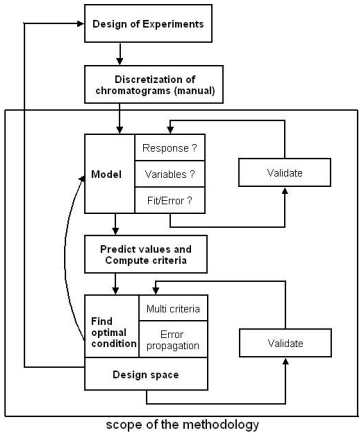

Figure 2. main steps of the presented methodology

the models predictive errors into the optimized criteria. Finally, this methodology allows to identify a design space if existing in the explored domain [8].

Figure 2 presents the main steps of our methodology. First, a designed experiment is set up in order to explore as best as possible the ranges of possible HPLC parameters. These experimental conditions are applied to a given complex mixture and supply related chromatograms. Second, for each chromatogram, the retention time of each compound is identified from the observed peaks. Third, statistical models are developed to predict the retention times of each mixture from the analytical conditions. Models have to be precise enough in order to have good predictive capabilities. This includes an adequate choice of the responses to be modelled, a variables selection step and the application of adequate statistical techniques to quantify the quality of the fit and validate the models. Then, criteria are defined to characterize what is a "good" chromatogram and predictive models are used to predict expected chromatograms and related criteria in the domain of explored analytical conditions. Derringer desirability functions and index are proposed to summarize the criteria to be optimized. Finally, optimal conditions are searched out of these predictions and their accuracy are analysed by Monte-Carlo propagation of the models prediction errors into the calculated criteria. The same technique can be applied on each point of the experimental domain to find a design space, if it exists.

Note that height of peaks of chromatograms has a limited interest for optimizing separation and is not concerned in this study. W. Dewé *et al.* (2004) [5] expose some formulations to reconstruct the height of peaks from their estimated area. It can easily be included in our methodology.

3 Details of the methodology

3.1 Design of experiments

Let's define F continuous or discrete tuning parameters of interest $(\boldsymbol{x}_1, ..., \boldsymbol{x}_f, ..., \boldsymbol{x}_F)$ for the HPLC device under study. Each continuous factor (e.g. pH) is defined over a domain of interest $[L_f, U_f]$ and each discrete factor (e.g. column) by a set of n_f levels. Let's note χ the experimental domain of $\boldsymbol{x}$'s.

Design of experiments and response surface methodology provide several methods to explore such domain χ according to the anticipated complexity of the factors effects on the responses (the peaks positions) [9]. Full factorial designs may be appropriate when the experiments are robotized. Central composite or D-Optimal designs are also well adapted in this context. Let's note $\boldsymbol{X}$ the resulting $(N \times F)$ design matrix.

The N experiments consist of applying each design factor setting $\boldsymbol{x}_i = (x_{i1}, ..., x_{iF})$ to a chosen (complex) mixture of M compounds $\boldsymbol{c} = (c_1, ..., c_j, ..., c_M)$ and provide N chromatograms with, for each of them, M more or less separated peaks.

Factors and levels to be used have to be identified by the analyst depending on the nature of the analytical method and the compounds to be separated.

3.2 Chromatogram discretization

Figure 3 illustrates the discretization process of the chromatograms. This is the main **manual** part of the methodology although current studies show that an automation of this process is possible [10],[11]. From each chromatogram, three retention times can be extracted: the retention times at the beginning, at the apex and at the end of the peaks at baseline-height (B, A and E, respectively). T_0 denotes the dead time of the system, associated to the analytical column. The retention times for the N chromatograms and the M components can be stored in three vectors B, A and E of size $(N.M)$. These retention times are ordered as N blocks of M components in these vectors.

Currently, it is still the task of the analyst to identify manually peaks in chromatograms. Patterns can be found in the DAD chromatogram, looking at the absorbance values for a specific time (1-D spectrum). For one peak corresponds one pattern. Difficulties of identifications can arise, e.g. when there are co-eluted peaks or when there are impurities, which possess nearly the same absorbance spectrum than relevant analytes.

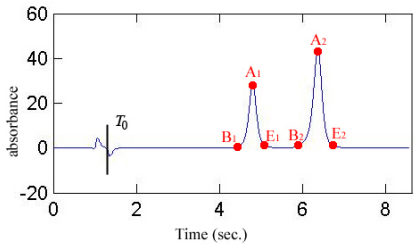

Figure 3. Raw chromatogram with the positions of discretized points

3.3 Models responses definition

The first step in the development of predictive models of the retention times with respect to HPLC analytical conditions, is to decide which responses to model. Dewé *et al.* [5] propose to create separate models for each retention time B, A and E and for each compound. Unfortunately, with this approach, uncertainty of the prediction of the models can lead to inversion in the predicted positions within a peak. For instance, it is undesirable that the predicted apex of a peak has a smaller retention time than its beginning, which is physically impossible. Moreover, there is little interest to try to make one model for each retention times vector because pairwise correlations between them are very close to 1. Modelling B, A and E leads to model three times non independent information. Deriving separate models for each retention times vector and each compound can also be heavy to handle in the subsequent steps of the methodology.

This paper proposes two enhancements to this classical approach. First, it suggests to build, for each response of interest, one single global model involving all F experimental factors and all M compounds. In this framework, it is suited to transform retention times B, A and E in retention factors k'. They are computed as follows:

$$k'_A = \frac{A - T_0}{T_0},\ k'_B = \frac{B - T_0}{T_0},\ k'_E = \frac{E - T_0}{T_0}$$

Notice that it is common to work with the logarithmic form of the k'.

Second, it suggests to transform the three original responses k'_B, k'_A and k'_E to three new responses which represent three independent parts of the information included in the original responses. Possible responses are the **position of the apex**, the **width of the peak** and its **asymmetry**. We define f_p as the transformation of the original responses to the p^{th} new response $Y^{(p)}$ (Equation 1). Examples will be given below. Note that, for simplicity, we consider that the transformations of retention times to retention factors are included in the functions f_p's.

$$Y^{(p)} = f_p(B, A, E), \quad (1 \leq p \leq P) \tag{1}$$

Then, $Y^{(p)}$ is the vector of size $N.M$ containing the observed values of $Y^{(p)}$ for the

N experiments and M compounds. The responses are ordered as N blocks of M components in this vector: $Y^{(p)} = (y_{11}^{(p)}, y_{12}^{(p)}, ..., y_{1M}^{(p)}, y_{21}^{(p)}, ..., y_{NM}^{(p)})$.

The inverse transformation function f^{-1} must exist in order to reconstruct the (predicted) retention times from transformed responses. This will be useful to compute retention time-based criteria to assess the quality of chromatograms.

$$(B, A, E) = f^{-1}(Y^{(1)}, ..., Y^{(P)}). \tag{2}$$

Such approach will prevent peak inversion in model prediction and clearly highlight, in the statistical modelling, the part of information available in each response.

3.4 Predictive models building

The ultimate goal of the model building phase is to be able to predict, for given settings $\boldsymbol{x}_i$ of the HPLC tuning parameters, the retention times B, A and E for each compound c_j of the mixture of interest. This will be achieved through the development of one model for each transformed response $Y^{(p)}$, $p = 1, ..., P$.

For given experimental conditions $\boldsymbol{x}_i$ and a compound c_j, a very general theoretical model for $y_{ij}^{(p)}$ can be written as:

$$y_{ij}^{(p)} = g_p(\boldsymbol{x}_i, c_j\ ;\ \boldsymbol{\beta}_p) + \epsilon_{ij}^{(p)} \tag{3}$$

where the experimental errors $\epsilon_{ij}^{(p)}$ are supposed to be independent, identically distributed with mean 0 and common variance σ_p^2. The independence hypothesis is taken at three levels : between responses, between experiments (chromatograms) and between compounds within chromatograms. In design of experiments applications, the function g_p is often defined as a polynomial model and estimated by classical least squares.

The following steps are recommended to develop these predictive models.

3.4.1 Full model matrix definition

For the $(N \times F)$ design matrix $\boldsymbol{X}$ defined in Section 3.1, let's denote $\boldsymbol{Z}$ as the $(N \times R)$ model matrix containing the higher polynomial model terms that may be estimated with $\boldsymbol{X}$. $\boldsymbol{Z}$ will typically contains a constant term, main (qualitative and quantitative) factor effects, quadratic or higher order terms for quantitative factors and interactions. Quantitative factors should be centred and scaled in the $[-1, 1]$ interval before being included in $\boldsymbol{Z}$. Qualitative factors should be coded into dummy variables.

In order to build predictive models for $\boldsymbol{x}_i$ and c_j, matrix $\boldsymbol{Z}$ should be expanded with all the interactions between c and polynomial terms in $\boldsymbol{Z}$. This full model matrix is built as $\boldsymbol{S} = \boldsymbol{Z} \bigotimes \boldsymbol{I}_M$ where $\bigotimes$ is the Kronecker product and $\boldsymbol{I}_M$ the $M \times M$ identity matrix. $\boldsymbol{S}$ is thus a $((N.M) \times (R.M))$ matrix containing all the terms which may potentially explain $Y^{(p)}$, the $(N.M)$ vector of observed values for the p^{th} response. $(R.M)$ is potentially high and variables selection techniques will be necessary to achieve good predictive models.

3.4.2 Model estimation and dimension reduction

Multiple linear regression (MLR) or other statistical techniques such as partial least squares (PLS) are natural model estimation techniques in this framework. Good introductions can be found in Martens (2001) [12]. If MLR is applied, a variables selection technique as forward or stepwise will be necessary to select, for each response $Y^{(p)}$, the terms that are the most informative and avoid model overfitting. Let's note by $\boldsymbol{S}^{(p)}$ the $(N.M) \times q^p$ submatrix of $\boldsymbol{S}$ coming out of the variables selection step for response $Y^{(p)}$ ($q^p \leq R.M$). The least squares parameters estimators are then given by :

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{p}} = (\boldsymbol{S}^{(p)'}\boldsymbol{S}^{(p)})^{-1}\boldsymbol{S}^{(p)'}\boldsymbol{Y}^{(p)}$$

.

For MLR, the number of terms kept in the model will typically be chosen simply by optimizing some criterion on the training set like the AIC or the Adjusted R^2 or, more heavily, through cross-validation and optimization of a criterion like the RMSE.

3.4.3 Model Validation

Models must be validated before being used for prediction. A residual analysis will be appropriate to check model adequacy, detect outliers or heteroscedasticity. Appropriate $X - Y$ scatter plots will allow to check visually the independence assumptions. If available, lack of fit tests will allow verifying if the residual variance is close to the experimental variance.

3.4.4 Prediction

For given factor setting $\boldsymbol{x}_0$, the estimated models supply predictions $\hat{y}_{ij}^{(p)}$'s for each response p ($p = 1, ..., P$) and each compound c_j. In MLR, one must first build the vector of model terms $\boldsymbol{s}_{0j}^{(p)}$ of size q^p associated to $\boldsymbol{x}_0$, c_j and response $Y^{(p)}$. The

predicted responses are given by:

$$\begin{aligned}\hat{\boldsymbol{y}}_{0j}^{(p)} &= \boldsymbol{s}_{0j}^{(p)}\hat{\boldsymbol{\beta}}_p \\ &= \boldsymbol{s}_{0j}^{(p)}(\boldsymbol{S}^{(p)'}\boldsymbol{S}^{(p)})^{-1}\boldsymbol{S}^{(p)'}\boldsymbol{Y}^{(p)} \\ &= \hat{E}[\boldsymbol{y}_{0j}^{(p)} \mid \boldsymbol{x}_0, c_j] \quad p = 1, ..., P.\end{aligned} \tag{4}$$

One can then use the f^{-1} function to get original predicted responses $\hat{B}_{0j}$, $\hat{A}_{0j}$ and $\hat{E}_{0j}$ from the $\hat{\boldsymbol{y}}_{0j}^{(p)}$'s ($p = 1, ..., P$) and thus, predict the complete chromatogram retention times for the condition $\boldsymbol{x}_0$.

3.5 Criteria to assess the quality of a chromatogram

Several criteria can be defined to express quantitatively the quality of a chromatogram. Schoenmakers (1986) summarize some definitions of various useful criteria in [1]. In the framework of the methodology presented in this paper, we favour criteria which assess globally (for the whole chromatogram) a given characteristic of the chromatogram. Criteria are based on retention times only. Possible characteristics of interest are the resolution, separation, peak width, asymmetry or maximum elution time.

More formally, let's B_j, A_j and E_j, $j = 1, ..., M$ denote the retention times of the M peaks of a given chromatogram and $B_{(j)}, A_{(j)}$ and $E_{(j)}$, $j = 1, ..., M$ the ordered ones (with respect to the retention time of the apex). Each criterion cr_z can then be defined as a specific function t_z of these retention times:

$$cr_z = t_z(B_j, A_j, E_j; j = 1, ..., M).$$

Under these notations, the following interesting criteria may be defined: cr_1, the longer elution time which should be minimum; cr_2, the minimum separation between two subsequent peaks which should be maximum; cr_3, the maximum peak width which should be minimum; cr_4, the minimum peak resolution which should be maximum and cr_5, the maximum peak asymmetry which should be minimum. A way to express them formally is as follows:

$$\begin{aligned}
cr_1 &= max(A_j), \quad j = 1, ..., M \\
cr_2 &= min(B_{(j+1)} - E_{(j)}), \quad j = 1, ..., M-1 \\
cr_3 &= max(E_j - B_j), \quad j = 1, ..., M \\
cr_4 &= min\left(\frac{2 * (A_{(j+1)} - A_{(j)})}{(E_{(j+1)} - B_{(j+1)}) + (E_{(j)} - B_{(j)})}\right), \quad j = 1, ..., M-1 \\
cr_5 &= max\left(\frac{\left|E_j + B_j - 2A_j\right|}{E_j + B_j}\right), \quad j = 1, ..., M
\end{aligned}$$

Thus, each global criterion is defined as the worst value of a calculated characteristic in a given chromatogram. This ensures that all other computed values, for other peaks or between other pairs of peaks, are at least better.

The methodology described in this paper is, of course, applicable to other criteria. For example, Vanbel [4] and Dewé *et al.* [5] show the use of limited optimization criteria (e.g. the separation of only 2 peaks of interest) and robustness criteria.

3.6 Definition of a global optimization criterion

Finding optimal chromatographic analytical conditions according to several criteria as defined in Section 3.5 is a multicriteria optimization problem. A common methodology to approach such question has been introduced by Harrington [13] and Derringer and Suich [14]. They propose to aggregate the criteria of interest in one global optimization criterion in two steps.

First, each original criterion cr_z ($z = 1, ...Z$) is transformed into a desirability value $d_z(cr_z)$ through a desirability function d_z. d_z takes its values between 0 and 1 where 1 corresponds to a highly "desirable" value for cr_z and 0 to a non acceptable value. Values increasing between 0 and 1 express an increase of the "desirability" of the criterion. Second, all desirability values $d_1(cr_1), ..., d_Z(cr_Z)$ are aggregated in one global desirability index $D(d_1(cr_1), ..., d_Z(cr_Z))$ to be optimized. This desirability index is also restricted to the [0,1] interval.

Different types of desirability functions are proposed in the literature. Harrington introduced the first desirability functions, using exponential functions. Derringer and Suich based their desirability functions on a power of a linear transformation of the responses (criteria). Recently, le Bailly de Tilleghem and Govaerts [15,16] proposed functions based on the Normal cumulative distribution function (see Figure 4). These functions present no discontinuities, keeping the strict order of criteria. More formally, they define $d_z()$ as :

$$\begin{aligned} d_z(cr_z) &= \Phi(\frac{cr_z - a_z}{b_z}) \text{ if } cr_z \text{ has to be maximized,} \\ d_z(cr_z) &= 1 - \Phi(\frac{cr_z - a_z}{b_z}) \text{ if } cr_z \text{ has to be minimized,} \end{aligned} \tag{5}$$

where Φ is the cumulative distribution function (CDF) of the standard Normal variable defined as:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} exp(\frac{-t^2}{2})dt. \tag{6}$$

a_z and b_z are respectively localisation and dispersion parameters to be fixed by the analyst according to the context and the criterion. Changes in the parameter a_z imply left or right shifts of the curve. Increasing the parameter b_z will make the curve less stiff.

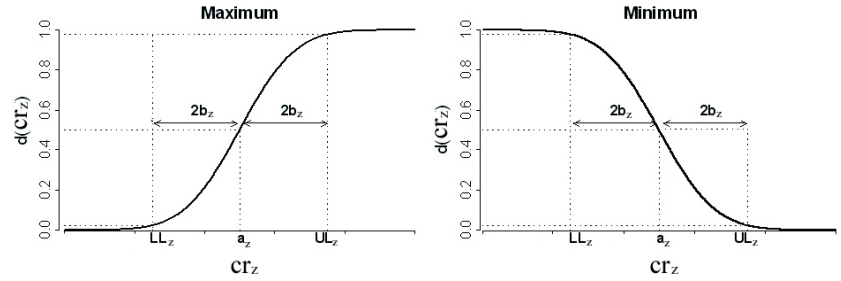

Figure 4. Desirability functions based on the standard Normal distribution function. The first graph illustrates a criterion to be maximized while the second stands for a criterion to be minimized.

Several global desirability indices are possible. The most popular is the weighted geometric mean of all individual desirability values:

$$D(\boldsymbol{cr}) = \prod_{z=1}^{Z} (d_z(cr_z))^{w_z} \text{ with } \sum_{z=1}^{Z} w_z = 1. \tag{7}$$

w_z values are fixed by the analyst according to the importance he wants to give to each chromatogram quality criterion in the global desirability index. Geometric mean is particularly adapted because one non fulfilled criterion will lead to very bad global desirability. A good global desirability ensures to have all the (weighted) criteria as good as needed.

3.7 Search for optimal chromatographic conditions

The experimental design, the predictive models, the quality criteria and the desirability index described in the previous sections provide the necessary elements to reach the goal of this paper: optimize chromatographic analytical conditions.

The models allow first to predict, for any values of the HPLC tuning parameters $\boldsymbol{x} = (x_1, ..., x_F)$ in the experimental domain χ, the retention times B_j, A_j and E_j for the M compounds of the reference mixture of interest. These predictions can then be transformed to (estimated) chromatographic quality criteria $(\hat{cr}_1, \hat{cr}_2, ..., \hat{cr}_Z)$ and the global quality of the chromatogram summarized in an (estimated) desirability index $\hat{D}(d_1(\hat{cr}_1), ..., d_Z(\hat{cr}_Z))$. Estimated optimal chromatographic conditions can then be obtained by searching a value $\boldsymbol{x}^*$ such that D is maximized:

$$\boldsymbol{x}^* = \max_{\boldsymbol{x}_0 \in \chi} D(\hat{\boldsymbol{cr}}) = \max_{\boldsymbol{x}_0 \in \chi} \prod_{z=1}^{Z} (d_z(\hat{cr}_z))^{w_z}$$

$$= \max_{\boldsymbol{x}_0 \in \chi} \prod_{z=1}^{Z} (d_z(t_z(f^{-1}(g_1(\boldsymbol{x}_0,\ \boldsymbol{c},\ \hat{\boldsymbol{\beta}}_1), ..., g_P(\boldsymbol{x}_0,\ \boldsymbol{c},\ \hat{\boldsymbol{\beta}}_P)))))^{w_z} \tag{8}$$

Optimal solution can be found using grid search. Other methods exist, such as gradient descend or simplex algorithm but grid search is the most appropriate tool in our context because the dimensionality (number of factors to simultaneously optimize) is generally limited. Furthermore, it avoids falling into local optima and gives a global map of the evolution of the desirability index over the experimental domain χ.

3.8 Model prediction and error propagation

When optimal conditions are derived from statistical model predictions, it is crucial to study the impact of the model prediction error on the reliability of the solution found. This evidence is however rarely highlighted in the multiresponse optimisation design of experiments literature, but has been discussed recently by Trautmann and Weihs [17] and le Bailly de Tilleghem and Govaerts [16].

These authors propose different approaches to quantify the incertitude of the desirability functions values and of the desirability index. They introduce also the notion of equivalence zone around the optimum. For simple cases, exact or approximate analytical solutions exist to implement these concepts but for more complex situations, Monte Carlo simulations are recommended. The situation of this paper must be considered as complex due to the introduction of the chromatogram quality measures in the calculation of the desirability index.

This section proposes a Monte Carlo approach to establish the distributions of the desirability functions and of the global desirability index. It shows then how these results allow to derive an equivalence zone for this optimum. Let's take $\boldsymbol{x}_0$, a given value for the HPLC tuning parameters, and c_j, a given compound of interest in the chromatogram. The prediction models described in Section 3.4.4 provide, for each response $Y^{(p)}$ and each compound c_j, an estimate of the expected response: $\hat{\mu}^{(p)}_{Y_{0j}} = \hat{E}[y^{(p)}_{0j} \mid \boldsymbol{x}_0, c_j]$. Let's then suppose that an estimate $\hat{\sigma}^2(\hat{\mu}^{(p)}_{Y_{0j}})$ is available for the variance of this estimator. In the MLR framework, it is given by (see equation 4 for notations) :

$$\hat{\sigma}^2(\hat{\mu}^{(p)}_{Y_{0j}}) = \hat{\boldsymbol{\sigma}}^2_{\epsilon_p} \boldsymbol{s}^{(p)}_{0j} (\boldsymbol{S}^{(p)\prime} \boldsymbol{S}^{(p)})^{-1} \boldsymbol{s}^{(p)\prime}_{0j} \ , \tag{9}$$

where the residual variance estimator $\hat{\boldsymbol{\sigma}}^2_{\epsilon_p}$ can be estimated from the training data as

$$\hat{\boldsymbol{\sigma}}^2_{\epsilon_p} = \frac{1}{(NM - q^p)} \sum_{i=1}^{N} \sum_{j=1}^{M} (y^{(p)}_{ij} - \hat{y}^{(p)}_{ij})^2 \ . \tag{10}$$

One can then generate, for each response $Y^{(p)}$ and each compound c_j, a large set

($i^* = 1, ...N_{sim}$) of simulated "predicted" responses in $\boldsymbol{x}_0$ according to:

$$Y_{0i^*j}^{(p)} = \hat{\mu}_{Y_{0j}}^{(p)} + \epsilon_{0i^*j}^{(p)} \ ,$$

where the $\epsilon_{0i^*j}^{(p)}$ are random Normal variables with mean 0 and variance $\sigma^2(\hat{\mu}_{Y_{0j}}^{(p)})$. In the MLR context, this normality assumption is common.

For each simulation i, $(M \times P)$ response values are thus generated. Then, original responses (B_j, A_j and E_j), quality criteria (cr_z) and the global desirability index D are derived. This allows to establish Monte Carlo distributions for the quality criteria, and for the global desirability index, which give an idea of the impact of the models prediction error on the uncertainty of the resulting estimations.

3.9 Optimum equivalence zone

The Monte Carlo distribution of the global desirability index gives also the necessary information to define a zone of the experimental domain which can not be stated to give significantly worse results than the optimum $\boldsymbol{x}^*$ found. The *equivalence zone* will simply be defined as the set of $\boldsymbol{x}$'s of the experimental domain χ such that the (estimated) desirability index $D(\boldsymbol{x})$ is greater than the 5^{th} percentile of the Monte Carlo distribution of the desirability index at the optimum $\boldsymbol{x}^*$. This definition is not perfectly correct because there is also an uncertainty on $D(x)$, but it is easy to implement and is sufficiently informative in the context of this paper. The size of the equivalent zone will give an idea of the real interest of the optimization process. A large zone indicates that the optimization process could not really differentiate between the quality of different analytical conditions. This may be due to the fact that there is effectively no real difference between the experimental conditions over the explored domain, or that the uncertainty in the predictive models is too large to be able to highlight the potential differences.

From a practical point of view, note that, when the number of quantitative factors of interest is small, the equivalence zone can elegantly be represented on graphs of contour plots of the predicted desirability index over the experimental domain for fixed values of the qualitative factors. This will be illustrated in the next section.

3.10 Design space

The last section showed how to propagate the error of the statistical models, in the optimal factors configuration, to find an equivalence zone. A similar approach can be used to find, with a known confidence, a design space [8]. Design space is defined as the established range of process parameters and formulation attributes that have been demonstrated to provide assurance of quality. A design space is

particularly useful. It can be seen as a zone of robustness defined in the experimental domain because working within is not considered as a change in the analytical method. At opposite, working out of the design space is considered to be a change and normally initiates new steps of validation.

The desired minimum quality must be fixed *a priori*. For instance, it is desirable to known if there is a zone, in the experimental design, where it is likely to have a separation of at least m_1 minutes and a total processing time not exceeding m_2 minutes. So, one will look, for each point of the experimental domain, under the propagated error, how many hits of sufficiently good chromatograms are achieved. More formally, we have (for criteria to be maximized):

$$[\forall x_0 \in \chi \mid P(cr_z > \lambda_z) > \gamma\%] \ \forall\, z \text{ in } 1, ..., Z \tag{11}$$

Similar formulation can be written for criteria to be minimized. An unified equation for maximized and minimized criteria can be written in the desirability space:

$$[\forall x_0 \in \chi \mid P(d_z(cr_z) > d_z(\lambda_z)) > \gamma\%] \ \forall\, z \text{ in } 1, ..., Z \tag{12}$$

Figure 5 illustrates this formulation, for a point $x_0 \in \chi$, and represents simulated distribution of a desirability function under the propagated error of the models. A limit value for the criteria $d_z(\lambda_z)$ is first chosen. The interest is to observe if $\gamma\%$ of the distribution is higher than $d_z(\lambda_z)$.

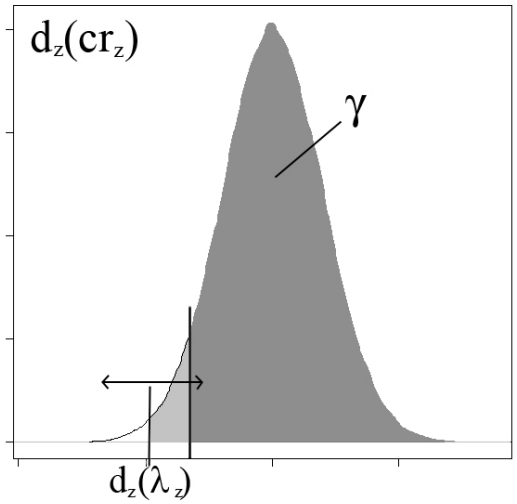

Figure 5. Simulated distribution of a criteria (desirability function) under propagated error of the models. The interest is to look if $\gamma\%$ of the distribution of $d_z(cr_z)$ is higher than a chosen limit $d_z(\lambda_z)$.

Thus, a point x_0 belongs to the design space if all the criteria of interest are fulfilled with a certain level a confidence $\gamma\%$. A wide design space, defined with strong constraints on the criteria (high $d_z(\lambda_z)$), and with a high level of confidence γ, is the most enviable situation in the experimental domain.

As for the equivalence zones, if the number of quantitative factors of interest is small, the design space can be represented on graphs of contour plots.

4 Application

4.1 Experimental

The objective of the separation is the evaluation of the methodology. Consequently, the compounds under investigation were not selected on the basis of an analytical problem nor for a practical application.

4.1.1 Chemical and reagents

A mixture of five commercially available compounds (Diflunisal, Granisetron, Nifedipine, Phenytoine, Sulfinpyrazone) was analysed using the design described here after (section 4.2). These five compounds were obtained from the Eli Lilly pharmaceutical company (Indianapolis, USA). Methanol and acetonitrile of HPLC grade were purchased from Sigma-Aldrich (Steinheim, Germany). Ultra pure water was obtained from an Academic A10 Milli-Q system (Millipore, Eschborn, Germany). Acetic acid (>98%) was purchased from Fluka (Steinheim, Germany), ammonium formate (99%) was purchased from Sigma-Aldrich.

4.1.2 Apparatus

All analysis were performed on a Waters 2695 separation module coupled to a Waters selector valve 7678 and a Waters 996 Photodiode array detector (Waters, Eschborn, Germany). The Empower 1.0 software was used to manage chromatographic data. Five analytical columns were used: C18, C8, RP18, Phenyl XBridge columns (Waters) (100x2.1 mm i.d. ; particle size 3.5 μm) and a C18 Cogent Bidentate column (100x2.1 mm i.d. ; particle size 4.0 μm) (Microsolv, Villecresnes, France).

4.1.3 Chromatographic conditions

The elution gradients were performed in 10, 20 or 30 minutes from 5% to 95% of organic modifier (methanol or acetonitrile) in the adequate buffer at a constant flow rate of 0.25 ml/minute at 30 °C. The buffer solutions were adjusted at the desired level of pH:

- pH 2.6: 0.1% concentrated Formic acid (99%) in water,

- pH 4.0, 5.0 and 7.0: Ammonium Formate 10 mM in water adjusted with concentrated formic acid and/or ammonia aqueous solution (35%).

The analytes were monitored photometrically at 240 nm although chromatographic data were recorded from 210 to 400 nm for all the analytical conditions investigated. The analytes were dissolved in an acetonitrile/water mixture (50:50, v/v). The injection volume was 2.0 or 5.0 μL.

4.1.4 Software

We used the statistical language *R* 2.4.0 for Windows to implement the methodology presented in the last section.

4.2 Design of experiments

We applied a full factorial design on 4 factors : the *pH* of the mobile phase, the *time* used for the gradient, the *solvent* used in the composition of the buffer, and the analytical *column*. It is clear these factors highly affect the position of peaks in the chromatograms. We used gradient mode because the behaviour of peaks elutions is not as known as in isocratic conditions. The goal was to stress the methodology using less repeatable experiments with gradient mode with less known behaviour of peaks elutions.

Moreover, the analytes have been chosen to cover a large range of pK_a and log P. Gradient mode gives more chance to obtain all the analytes eluted on the chromatogram in a constrained time.

Temperature, column diameter, injection volume or other chromatographic parameters were not include in the design example. Their values were then fixed as explained in section 4.1.

Factors and levels of factors were chosen to validate the methodology. The Figure 6 shows the values of the quantitative factors (pH and gradient time) over the experimental domain. 3 levels of pH and gradient were investigated in a full factorial design. Intermediate points were added to validate the methodology.

5 analytical columns and 2 solvents have been selected as qualitative factors.

- column : *Bidentate NA*, *Xbridge C*18*, *Xbridge C*8, *Xbridge Phenyl** and *Xbridge RP*18,
- solvent : CH_3CN and *MeOH*.

Analytical columns were chosen for their different chemical interactions with so-

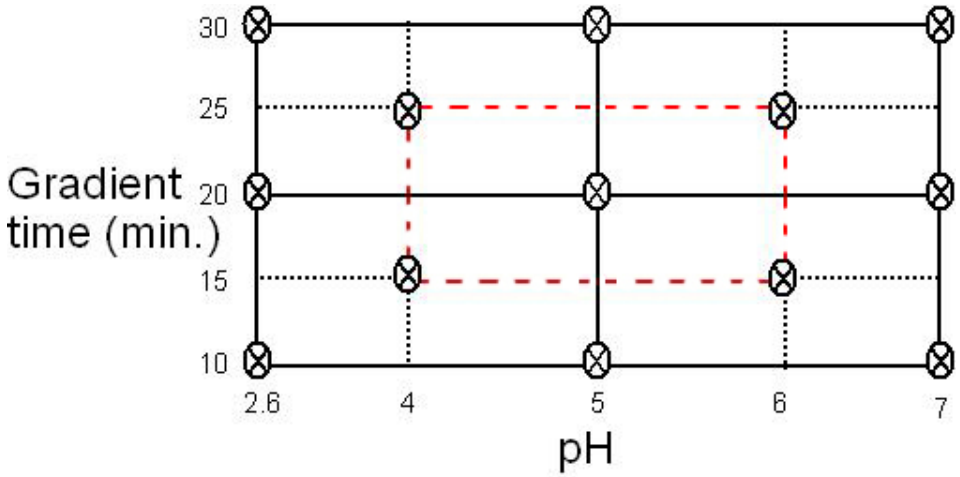

Figure 6. Full factorial design applied on the quantitative factors (plain lines). Intermediate points (pH=4, 6) have been added to the design (dashed lines) to validate the methodology.

lutes and also their robustness against large range of pH. The number of columns was a good compromise between domain investigation and overloading of experimental design.

Full factorial design on the 4 used factors gives 3 levels of pH $*$ 3 gradients $*$ 5 columns $*$ 2 solvents = 90 experiments that have been first realized. This is quite a lot of experiments, but it was useful to set up the models properly. The intermediate validation points were realized on both CH_3CN and $MeOH$ solvent, and on 2 columns: *Xbridge C*18 and *Xbridge Phenyl*. This gives $2*2*2*2 = 16$ experiments to generate a test set. Validation and test data give $90 + 16 = 106$ experiments. 5 compounds were analysed, giving $N * M = 106 * 5 = 530$ observations. Of course, final use of the methodology will lead to less experiments.

Fractional factorial design or D-optimal design are recommended when the number of factors to simultaneously optimize increases. Another improvement is to use a column selection before realizing experiments. A good practice is to choose columns which possess very different physico-chemical properties. For instance, reducing this number of columns to 2 would lead to 36 experiments instead of 90 in full factorial design. Using D-optimal design, number of experiments can be decreased to 24 or even 12. This is quite reasonable to develop a method with 4 factors to be simultaneously optimized.

4.3 Responses selection

We first modelled the retention factors k'. We assumed that the peaks were symmetric, which was suitable for this example. Thus, we modelled the apex and the width of the peaks. The supposed symmetry allows to define easily f^{-1} and thus to

reconstruct original retention times (see later). Formally,

$$Y^{(1)} = log(k'_A), \;\; Y^{(2)} = log(k'_E - k'_B) \tag{13}$$

Other responses definitions can also be used. We chose these responses because they possess the best modelling properties.

4.4 Models

We used polynomial models with a forward variables selection of included terms maximizing the Adjusted R^2 for model $Y^{(1)}$ and $Y^{(2)}$. This method gave the best predictive results.

Models possess a high number of parameters such that it is not possible to describe them completely. The Table 4.4 gives a brief summary of the models. The most impactful model, for the apex, seems good while the fit for the width is very limited. This can be seen as problematic, but the error on the widths estimation has less importance on the final estimation of the retention times (E and B). Moreover, this error is taken into account in the final steps of the methodology. Validation of the models are bypassed in this report.

	$Y^{(1)} = log(k'_A)$	$Y^{(2)} = log(k'_E - k'_B)$
adjusted R^2	0.96	0.51
# of parameters	180	68

Table 1. Summary of the fits (MLR) for the selected responses.

4.5 Prediction

So far, The models have been set up and can be used to predict new values of responses on the complete experimental domain χ. Retention times vectors can be reconstructed as follows :

$$\begin{gathered} \hat{k}'_A = e^{\hat{Y}^{(1)}}; \;\; \hat{k}'_B = \hat{k}'_A - \frac{e^{\hat{Y}^{(2)}}}{2}; \;\; \hat{k}'_E = \hat{k}'_A + \frac{e^{\hat{Y}^{(2)}}}{2} \\ \hat{A} = (T_0.\hat{k}'_A) + T_0; \;\; \text{idem for } \hat{B} \text{ and } \hat{E} \end{gathered} \tag{14}$$

This set of equations is the function f^{-1} presented in equation 2. Criteria can be computed from the retention times $\hat{A}$, $\hat{B}$ and $\hat{E}$. This allows to compute the desirability functions and global desirability index on each points of χ.

Then, a grid search was applied. The only parameter to tune is the value of the step of the quantitative variables. This defines a grid more or less dense. Computational explosion can arise if the step is too small. To avoid this, the search can be done in several step, increasing the precision around the optima found. Our selected step is 0.1. Values of quantitative factors are still normalized in the interval $[-1, 1]$ in this process, giving 21 points to be computed for each quantitative factors. The full factorial combination of factors then led to 21 levels of $pH * 21$ $gradients *$ 5 $columns * 2$ $solvents = 4410$ points to be predicted.

For each point, complete chromatogram, criteria, desirability values of criteria and global desirability index were easily computed, following the equation 8, page 12. Different global desirability indexes can be found using different weights w_z (see section 3.6).

The desired objective in our example was the best separation of all peaks in the minimum processing time. Only the criteria of minimal separation cr_2 and of maximum retention time cr_1 were then used. In this case, other criteria had little interest or were non informative (e.g. asymmetry).

We first considered default weights for both criteria ($w_1 = w_2 = 1/2$). However, Figure 7 (left) shows a non achieved separation between the fourth and the fifth peaks. To manage this problem, we elaborated a second solution, that tends to give less weight to the maximum retention time ($w_1 = 1/6$, $w_2 = 5/6$) in order to achieve a better separation.

The optimal points $\boldsymbol{x}^*$ for these two solutions are given in Table 4.6 (see equation 8, page 12).

	(Column, pH, Solvent, Gradient time)
Solution 1	(*Xbridge Phenyl*, 3.92, CH_3CN, 12)
Solution 2	(*Xbridge RP*18, 7, CH_3CN, 18)

Table 2. Values of optimal points $\boldsymbol{x}^*$. These points are the ones in χ that maximise the global desirability.

Figures 7 and 8 show, for the two solutions envisaged, the optimal predicted chromatogram and the corresponding contour plot of the global desirability index (qualitative factors are fixed to optimum). The second chromatogram has a better separation but takes more time to be processed. It is clearly shown that the weights allow to find an acceptable solution (without any more experiment). This give a lot of flexibility to this approach.

However, the used gradient (from 5% to 95% of organic modifier in the buffer) leads to suboptimal solutions. Indeed, the peaks elutes in 5 minutes from 10 to 15 minutes. The optimality of a solution is only valid within the experimental domain and it is still the responsability of the analyst to define the most interesting domain. Other gradient slopes or isocratic conditions would have been used and maybe, would have given better performance.

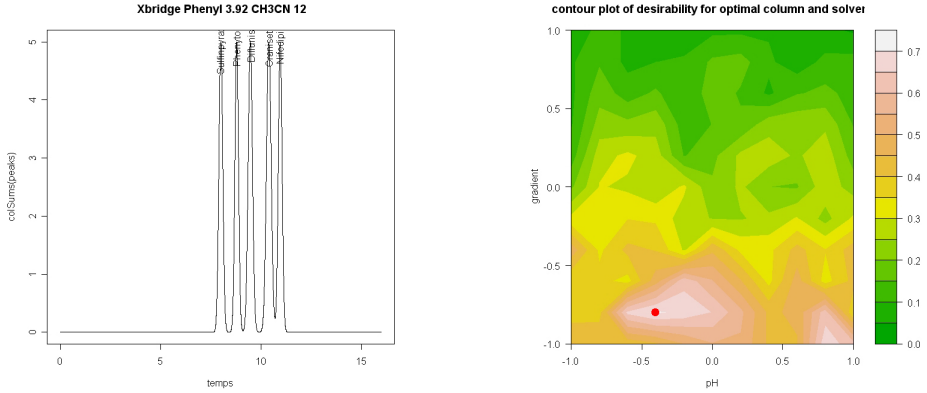

Figure 7. Solution 1. *Left*: Predicted optimized chromatogram with equal weights for criteria. *Right*: Corresponding contour plot of the global desirability index across the quantitative normalized domain (qualitative factors fixed to optimum). The red (dark) point shows optimum.

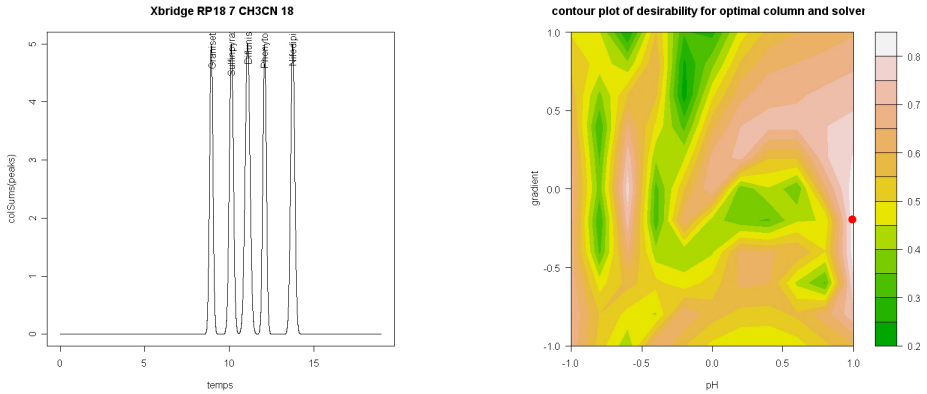

Figure 8. Solution 2. *Left*: Predicted optimized chromatogram with higher weight for separation. *Right*: Corresponding contour plot of the global desirability index across the quantitative normalized domain (qualitative factors fixed to optimum). The red (dark) point shows optimum.

It is important to be able to give confidence in these optimal solutions. Uncertainty of the expected estimated responses for the predicted optimal points can be propagated. One can look at the values of the desirability of criteria, under propagated error, with histograms of their distribution. The main interest is to look if the error of the models allows to have confidence in the values of desirability of criteria and also the global desirability index.

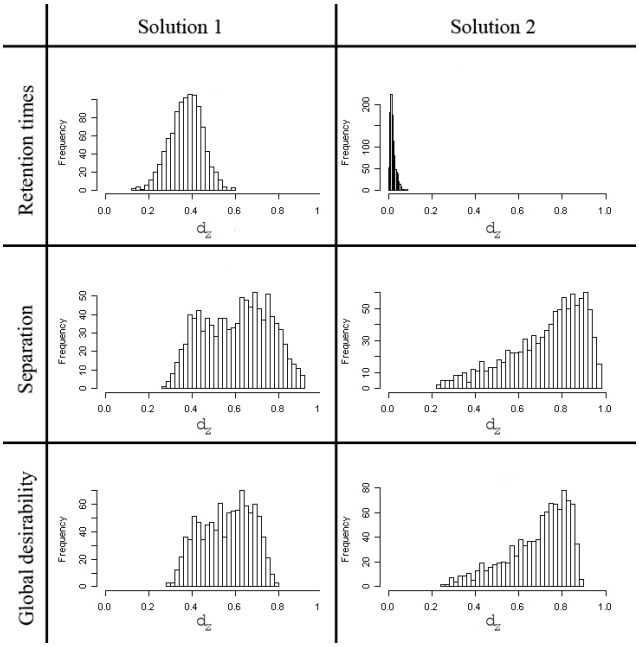

Figure 9. Distributions of the two estimated criteria (desirability functions of maximum retention times and minimal separation) and global desirability for the two solutions envisaged.

Figure 9 shows the error propagated on the considered criteria and on the global desirability, for the two solutions envisaged. In the first solution (left column), we wanted to maximise separation and to minimize total retention time of the chromatogram. In the second solution, we gave less relative weight to the retention time criterion. This leads to a worse distribution of the error propagated for this criterion (first row of the array), while the distribution for the separation possess a certain asymmetry underlying the better achievement of this criterion (second row). The same observation can be done on the global desirability (last row). Notice that the criteria are shown non weighted (rows 1 and 2 of the graph), but global desirability takes the weights into account (last row).

4.8 Equivalence zones

Figure 10 shows the equivalence zones at 95% on the contour plot of the figure 7 and 8. The first graph (left) shows clearly differentiated values. This leads to the conclusion that this optimal point has a limited equivalence zone. The optimal equivalence zone is better (in term of desirability) than elsewhere in the design of experiment. The second graph can not be used to give a similar conclusion.

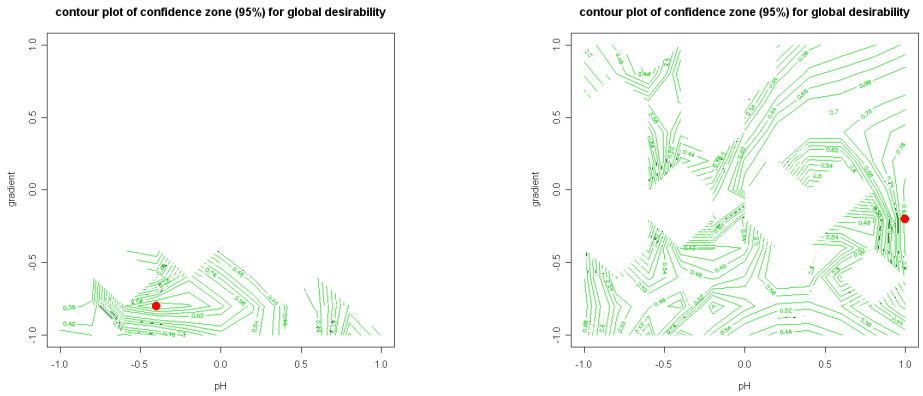

Figure 10. Contour plot showing the set of solutions giving an estimated desirability greater than the 5^{th} percentile of the simulated global desirability index under error of the models (at optimal solution). *Left*: solution 1. Very few conditions give desirability likely close to the optimal solution. *Right*: solution 2. Over the domain, most conditions are likely to give desirability index close to the optimal one.

4.9 Design space

The final objective was to observe the existence, or non existence, of a design space in the experimental domain. A design space is considered as a zone of robustness in the experimental domain because it allows to tolerate variability of materials and slight changes in the process.

Criteria	**Value**
Separation	0.0 min. (minimum)
Total retention time	20s min. (maximum)
Probability γ	60%

Table 3. Design space: limit values for the criteria and confidence in design space.

The Table 3 shows the selected values for criteria to be fulfilled. According to these criteria, a small design space is found (see section 3.10 for mathematical details),

as shown in the Figure 11. It is certain that the used level of confidence and the chosen limits for the criteria are rather weak. The solution 2 is included in this design space.

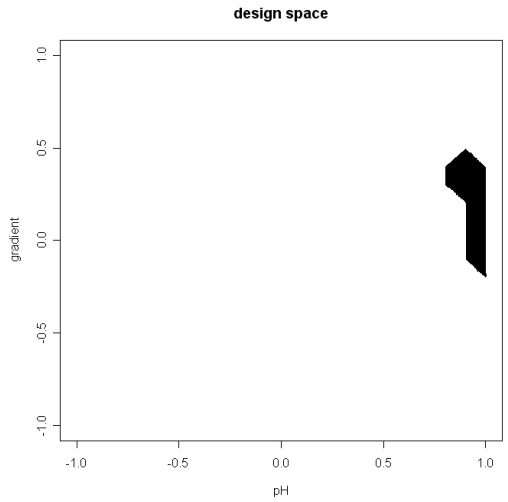

Figure 11. Contour plot of the design space (dark area) for the values presented in Table 3

4.10 Validation

The second solution has been validated. Figure 12 compares the chromatogram obtained with the proposed optimal conditions (right) with the one predicted using the solution 2 (left). The positions of the apexes of the peaks does not suffer from excessive imprecision but the width of the peaks are clearly not well predicted. This was foreseeable due to the bad fitting of the model for the width ($Y^{(2)}$).

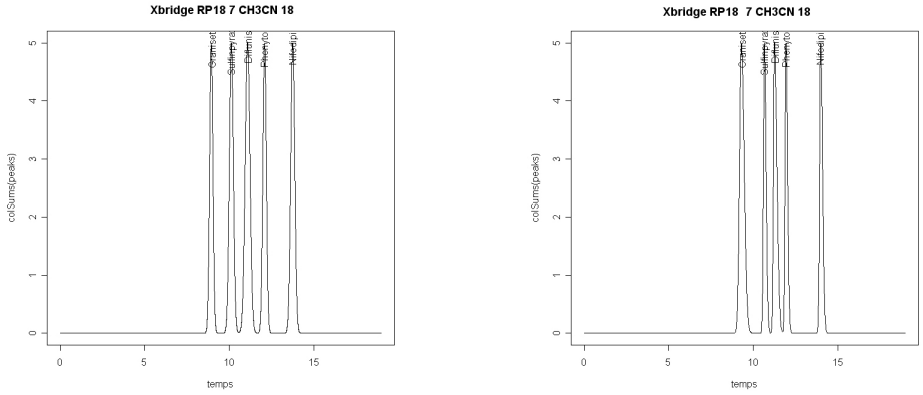

Figure 12. Validation of optimal solution 2. Left: predicted chromatogram. Right: Chromatogram obtained when using optimized HPLC parameters (Column *Xbridge RP*18, pH 7, Solvent CH_3CN, and gradient time 18 minutes).

5 Conclusions

A very flexible methodology for modelling chromatography has been proposed and shown in order to optimize analytical conditions. This is an important step to permit the automation of methods development. The discretization of retention times is so far the main manual part of the job but evidences show that this process can also be automated. Given the flexibility of the proposed methodology, it is now possible to envisage modelling for automation on real process.

The use of the Normal CDF is a very convenient and automated way to combine various criteria into a global desirability index. This also allows flexibility in regards of the chosen criteria. Furthermore, they can be weighted according to the need of the analysis.

Finally, assessing the way the uncertainty propagates into global desirability index is simple and effective to verify the confidence in optimal condition(s). Design space can also be computed with confidence, using a similar methodology of models error propagation, on each point of the experimental domain. Robustness can finally be assessed or validated analytically with new experiments in the design space around optimal solution(s).

6 Acknowledgements

This paper has been written for the ADAM project. The acronym ADAM stands for *Automated development of analytical methods*. The project is born from a cooperation between Eli Lilly company and the Institute of Pharmacy, laboratory of analytical chemistry, University of Liège. The authors thank the Eli Lilly company, the Walloon Region (FIRST Doctorant Entreprise International project ADAM n°516130) and the Interuniversity Attraction Pole (project n° P6/03) for their research grants. We also would like to thanks Professor Paul Eilers, University of Leiden, The Nederlands, for the precious advice given during this work.

References

[1] P. Schoenmakers, "Optimization of chromatographic selectivity: a guide to method development," *Analytica Chimica Acta*, vol. 208, pp. 357–358, 1988.

[2] L. Snyder, J. Kirkland, and J. Glajch, *Practical HPLC Method Development, second Edition*. Wiley-Interscience, 1997.

[3] D. Massart and al., *Chemometrics: A Textbook (Data Handling in Science and Technology)*. Elsevier Science, 1990.

[4] P. Vanbel, "Development of flexible and efficient strategies for optimizing chromatographic conditions," *J. of Pharmaceutical and Biomedical analysis*, vol. 21, pp. 603–610, 1998.

[5] W. Dewé, R. Marini, P. Chiap, P. Hubert, J. Crommen, and B. Boulanger, "Develoment of response models for optimizing HPLC methods," *Chemometrics and Intelligent Laboratory Systems*, vol. 74, pp. 263–268, 2004.

[6] G. Cox and W. Cochran, *Experimental Designs, 2nd Edition*. Wiley, 1957.

[7] B. Winer, *Statistical Principles in experimental design, 2nd Edition*. McGraw-Hill, 1962.

[8] ICH, "Q8 draft guidance on pharmaceutical development," 2004. Ver 4.3.

[9] A. Khuri and J. Cornell, *Response surfaces: designs and analyses*. Marcel Dekker, 1987.

[10] G. Vivó-Truyols, *New Strategies for optimisation and data treatment in HPLC*. PhD thesis, University of Valencia, 2005.

[11] B. Boulanger, "Utilisation de l'Analyse en Composantes Indépendantes (ICA) pour la séparation numérique des pics et la quantification automatique en CLHP-UV," conference Chimiométrie 2006, Paris, december 2006.

[12] H. Martens and M. Martens, *Experimental Designs, 2nd Edition*. Wiley, 2001.

[13] E. Harrington, "The desirability function," *Industrial Quality Control*, vol. 21, pp. 494–498, 1965.

[14] G. Derringer and R. Suich, "Simultaneous Optimization of Several Response Variables," *J. Qual. Tech.*, vol. 12(4), pp. 214–219, 1980.

[15] C. Le Bailly de Tilleghem and B. Govaerts, "Uncertainty Propagation in Multiresponse Optimization using a Desirability Index," Tech. Rep. 0532, Université catholique de Louvain, Louvain-la-Neuve, 2005.

[16] C. Le Bailly de Tilleghem and B. Govaerts, "Distribution of Desirability Index in Multicriteria Optimization using Desirability Functions based on the Cumulative Distribution Function of the Standard Normal," Tech. Rep. 0531, Université catholique de Louvain, Louvain-la-Neuve, 2005.

[17] H. Trautmann and C. Weihs, "Uncertainty of optimum influence factor levels in multicriteria optimization using the concept of desirability," Tech. Rep. 23/04, SFB 475, Dortmund University, 2004.