A Bayesian Design Space for analytical methods based on multivariate models and predictions.

Authors: Pierre Lebrun [a,†], Bruno Boulanger [c], Benjamin Debrus [a], Philippe Lambert [b], Philippe Hubert [a]

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

[b] *Université de Liège, Institute of Social Sciences, Quantitative Methods, Bld. du Rectorat 7, Bât. B31 , 4000 Liège , Belgium*

[c] *Arlenda S.A, Avenue de l'Hôpital 1, 4000 Liège, Belgium*

August 1, 2012

Abstract

The International Conference for Harmonization (ICH) has released regulatory guide-lines for Pharmaceutical Development. In the document ICH Q8, The Design Space of a process is presented as the set of factor settings providing satisfactory results. How-ever, ICH Q8 does not propose any practical methodology to define, derive and compute Design Space. In parallel, in the last decades, it has been observed that the diversity and the quality of analytical methods have evolved exponentially allowing substantial gains in selectivity and sensitivity. However, there is still a lack for a rationale towards the development of robust separation methods in a systematic way.

Applying ICH Q8 to analytical methods provides a methodology for predicting a region of the space of factors in which results will be reliable. Combining design of experiments and Bayesian standard multivariate regression, an identified form of the predictive distribution of a new response vector has been identified and used, under

---

[†]Corresponding author.
Email address: plebrun@ulg.ac.be (Pierre Lebrun).

non-informative as well as informative prior distributions of the parameters. From the responses and their predictive distribution, various critical quality attributes can be easily derived.

This Bayesian framework was then extended to the multi-criteria setting to estimate the predictive probability that several critical quality attributes will be jointly achieved in the future use of an analytical method.

An example based on a high-performance liquid chromatography (HPLC) method is given. For this example, a constrained sampling scheme was applied to ensure the modeled responses have desirable properties.

Key words: Design Space, design of experiments, multi-criteria decision methods, multivariate linear models, Bayesian statistics, uncertainty, analytical method development, robust optimization

---

1   Introduction

In 2006, the International Conference for Harmonization (ICH) released documents that describe regulatory guidelines for Pharmaceutical Development (ICH Q8, 2006). These guidelines rely on the use of information and prior knowledge gained during pharmaceutical development studies to provide a scientific rationale about the manufacturing process of a product. In particular, the concept of Design Space (DS) is defined as "the multidimensional combination and interaction of input variables (e.g., material attributes) and process parameters that have been demonstrated to provide assurance of quality".

The aim is to provide "assurance of quality". Indeed, one must be confident that high quality products or analytical methods will be obtained by reliable processes. The "quality" is expressed in terms of several critical quality attributes (CQAs) that must lie within accep-

tance limits. These limits are decided before initiating the set of optimization experiments and reflect the desired specification settings (e.g. critical resolution > 2 and processing time ≤ 5 min.). To improve quality, the naive solution would be to use overlapping mean response surfaces methodology to define a Sweet Spot of input conditions where the mean responses are satisfactory. It has been well discussed how this approach is not recommendable (Peterson and Lief, 2010).

The concept of "assurance" refers to the ways to provide guarantees that, given uncertainty of processes, models and measurements, it is likely that the CQAs will be within specifications in the future, in the presence of specific input conditions. Two options are identified for the manufacturer to give the required assurance of quality. First, he or she should demonstrate that the operating conditions are fully under control, without statistical modeling of his process. Second, evidence should be shown that the quality of the outcome or the product remains within acceptable limits, for changes in input variables within identified limits. The first option is generally difficult or too expensive to achieve. The second option is more realistic because it considers the inevitable variability in the achievement of quality; it is then the ideal way to cope with uncertainty, to become less sensitive to normal ranges of disturbances in the input conditions and to achieve the intended quality under a broader set of conditions that are easier to guarantee.

The second option is what is referred as "Quality by Design": a satisfactory process must provide quality outputs, and the reasons why this quality is achieved (or not) are known. However, given the broad nature of its applications, ICH Q8 does not propose a technically detailed solution to derive and identify the Design Space, but instead, gives clearly the principles and associated (yet controversial) examples.

Given the lack of ICH Q8-compliant methodology for the development of analytical meth-ods, this paper proposes an extension of the methodologies presented in previous works (Dewé et al., 2004, Lebrun et al., 2008) for the development of chromatographic methods. Robust liquid chromatography aims at obtaining chromatograms that show separated peaks in the shortest time. In this desirable case, the retention times (t_R) and areas of the peaks can be quickly used for qualitative or quantitative assessment. However, it has been common practice to model discontinuous and non-linear CQAs for the optimization of chromato-graphic conditions. One example of such CQA is the *critical resolution* (R_s) , a criterion illustrating the distance between the closest pair of peaks in a chromatogram, and that is computed as the combination of several variables (Vanbel, 1999, see also Equation (24), below). Its complex form and behavior increase the risk of having poor statistical models. Other CQAs for chromatographic quality assessment also exist, such as the width of the peaks, their asymmetry, the signal to noise ratio, etc. (Schoenmakers, 1988, Massart et al., 1990, Snyder et al., 1997).

To face the huge amount of daily data coming from R&D, a general approach is then needed to separate many compounds with unknown physico-chemical properties. The is-sue is rather complicated when the matrix to analyze is complex. In that case, classical optimization techniques (e.g. One-Factor-At-a-Time approach) generally fails in providing satisfactory results and certainly do not answer the ICH Q8 requirements.

In the literature, an approach has been proposed to model the t_R instead of complex criteria (Dewé et al., 2004). The t_R have continuous and (log)-linear properties that allow the use of simple linear regression models. Furthermore, from these modeled t_R, it is easy to derive any CQA, such as R_s for instance. However, in this last reference, dependencies between

responses are completely ignored, as well as the uncertainty of prediction.

More recently, it was proposed to propagate the uncertainties of the modeled responses to several CQAs using Monte-Carlo simulations (Lebrun et al., 2008). Several response transformations were made to obtain more independent responses. This allowed finding a set of acceptable solutions based on confidence intervals. However, if this did increase the quality of the results, some dependencies between response predictions have still been neglected. To obtain more informative statistical models, the covariance structure of the responses must be taken into account in a multivariate model.

In a general Design Space framework, a multivariate Student's distribution was proposed as the predictive density of the responses, precisely stressing the fact that correlations among responses must be taken into account when a simultaneous optimization is carried out to find the DS (Peterson, 2004, Miró-Quesada et al., 2004). A similar approach was used to compute a DS that also satisfies other desirable manufacturing attributes, such as cost and throughput (Stockdale and Cheng, 2009). These works were extended in a more flexible way using Seemingly Unrelated Regression (SUR) models (Peterson et al., 2009). This is naturally a good idea, as SUR allows a different number of model parameters for each modeled response. However, the derivation of an identified predictive distribution with SUR is not analytically tractable. In the present study, the identified predictive distribution of the standard multivariate regression (SMR) has been used to obtain a clear exposition of the influence of prior assumptions. Notice that the proposed methodology is also valid for the application of other regression models such as SUR, using efficient Gibbs sampling algorithms when necessary.

The applicability of Bayesian Design Space for analytical methods has also been illus-

trated (Peterson and Yahyah, 2009). In this work, the direct modeling of CQAs such as R_s has been found suitable for a robustness study. However, this might be problematic in an optimization context, considering that the critical pair of peaks might change. This affects R_s in a way that may be highly discontinuous and difficult to model properly.

2 Theory

2.1 Design Space definition

Suppose a method has k input conditions $\mathbf{X} = (\mathbf{x}_1, ..., \mathbf{x}_k)$ belonging to an experimental domain χ. Denote $\tilde{\mathbf{x}} = (\tilde{x}_1, ..., \tilde{x}_k)$, a certain combination of these inputs. A Bayesian definition of the Design Space has been proposed as:

$$\text{DS} = \{\tilde{\mathbf{x}} \in \chi \mid P(\boldsymbol{Y}_{\tilde{\mathbf{x}}} \in \boldsymbol{\Lambda} \mid \tilde{\mathbf{x}}, \text{data}) \geq \pi\}, \tag{1}$$

where $\boldsymbol{Y}_{\tilde{\mathbf{x}}}$ is a vector of responses that is likely to fall within an acceptance region ($\boldsymbol{\Lambda}$) with a minimal quality level π (Peterson, 2008). Generally, acceptance limits are chosen *a priori*, i.e. before experiments and data analysis are performed, and reflect the objectives (e.g., total separation, analytical target profile). The DS is then the set of combinations of input conditions where the joint posterior probability that future responses lie within acceptance limits is sufficiently high.

In development and validation, the process parameters, say $\boldsymbol{\theta}$, remain unknown and their values are estimated with uncertainty. The predictive posterior probability is then computed given the estimates and the uncertainty of $\boldsymbol{\theta}$.

Practically, an option is to model the CQAs of interest by (linear) response surface models

6

(RS), as classically envisaged when optimizing a process with Design of Experiments (DoE). For this study, it is stressed that a particular CQA could either be directly modeled as the response of a RS model, or either be the function of several modeled responses. This is the case for R_s (see Equation (24), third line, below), for instance.

The next section presents the SMR model. In DS computation with large number of parameters and responses, it is desirable to obtain a closed form for the predictive distribution for efficiency reasons. The constraints with SMR is that the same regressors are applied to each responses. It could be relaxed using SUR models, for instance.

2.2 Bayesian standard multivariate regression

Two main quantities are of great interest in DoE context. First, the posterior distribution of the parameters θ is useful to assess the effects of the factors. Second, the predictive distribution of responses Y is the core result used to compute the DS and its variability provides clue on both the process and the model qualities.

Assume Y has size m, i.e., m responses $\mathbf{Y} = (\mathbf{y}_1, ..., \mathbf{y}_m)$ have been observed. Each individual response $\mathbf{y}_j$ is a vector of size n containing the observations when the input conditions are changed. k common input conditions $\mathbf{X} = (\mathbf{x}_1, ..., \mathbf{x}_k)$ are further assumed to have strong influence on the responses.

A RS model in the form of a polynomial with p parameters is envisaged for all the responses. Then, $\mathbf{Z}$ is defined as the $(n \times p)$ model matrix that specifies a polynomial model that may be estimated over $\mathbf{X}$. The columns of $\mathbf{Z}$ generally consist of a constant term (intercept), main (qualitative and quantitative) factor effects, quadratic or higher order effect terms for quantitative factors and interactions. To simplify the analyse of factor effects, quanti-

7

tative factors can be centered and scaled in the interval $[-1, 1]$ before being included in $\mathbf{Z}$. Qualitative factors can be coded into dummy variables.

A set of m multiple linear regression equations is then developed,

$$\mathbf{y}_1 = \mathbf{Z}\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}_1,$$

$$\vdots$$

$$\mathbf{y}_j = \mathbf{Z}\boldsymbol{\beta}_j + \boldsymbol{\varepsilon}_j, \tag{2}$$

$$\vdots$$

$$\mathbf{y}_m = \mathbf{Z}\boldsymbol{\beta}_m + \boldsymbol{\varepsilon}_m,$$

where the $\boldsymbol{\beta}_j$ are $(p \times 1)$ model parameters for the jth responses and elements inside each $\boldsymbol{\varepsilon}_j = (\varepsilon_{1j}, ..., \varepsilon_{ij}, ..., \varepsilon_{nj})'$ are independent and identically distributed (i.i.d.). Regrouping the model parameters into the $(p \times m)$ matrix $\mathbf{B} = (\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_j, ..., \boldsymbol{\beta}_m)$, allows writing Equation (2) in matrix form:

$$\mathbf{Y} = \mathbf{Z}\mathbf{B} + \mathbf{E}. \tag{3}$$

The error matrix $\mathbf{E}$ is of size $(n \times m)$ with $\mathbf{0}$-mean and positive-semidefinite covariance matrix $\boldsymbol{\Sigma}$. $\boldsymbol{\varepsilon}_j$ refers to the jth column of $\mathbf{E}$ while $\boldsymbol{\varepsilon}_i$ refers to the ith line of $\mathbf{E}$. Classical model assumption is:

$$\boldsymbol{\varepsilon}_i \sim N_m (\mathbf{0}, \boldsymbol{\Sigma}). \tag{4}$$

The likelihood is then defined as:

$$\mathcal{L}(\mathbf{B}, \boldsymbol{\Sigma} \mid \mathbf{Y}) = (2\pi)^{\frac{-mn}{2}} |\boldsymbol{\Sigma}|^{\frac{-n}{2}} . \exp\left(-\frac{1}{2} \sum_{i=1}^{n} \left[(\mathbf{y}_i - \mathbf{z}_i \mathbf{B}) \boldsymbol{\Sigma}^{-1} (\mathbf{y}_i - \mathbf{z}_i \mathbf{B})' \right] \right)$$

$$\text{with} \quad -\infty < y_{ij} - \mathbf{x}_i \boldsymbol{\beta}_j = \varepsilon_{ij} < \infty, \quad i = 1, \ldots, n, \quad j = 1, \ldots, m.$$

$$\propto |\boldsymbol{\Sigma}|^{\frac{-n}{2}} . \exp\left(-\frac{1}{2} tr \left[\boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \mathbf{Z}\mathbf{B})' (\mathbf{Y} - \mathbf{Z}\mathbf{B}) \right] \right) \tag{5}$$

8

2.2.1 Non-informative prior distributions

A Bayesian solution to obtain the joint posterior density of the parameters $\theta = (\mathbf{B}, \boldsymbol{\Sigma})$ has been well described using non-informative prior distributions (Geisser, 1965, Box and Tiao, 1973). Given the following joint non-informative prior distribution:

$$p(\mathbf{B}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{1}{2}(m+1)}, \tag{6}$$

the posterior distribution of $\mathbf{B}$ given $\boldsymbol{\Sigma}$ is expressed as:

$$(\mathbf{B} \mid \boldsymbol{\Sigma}, \text{data}) \sim N_{(p \times m)}\left(\hat{\mathbf{B}}, \boldsymbol{\Sigma}, (\mathbf{Z}'\mathbf{Z})^{-1}\right). \tag{7}$$

It is then a matrix-variate Normal distribution of size $(p \times m)$, with location parameter $\hat{\mathbf{B}}$ and variance matrices $\boldsymbol{\Sigma}$ and $(\mathbf{Z}'\mathbf{Z})^{-1}$. $\hat{\mathbf{B}}$ is the least-square estimator of $\mathbf{B}$:

$$\hat{\mathbf{B}} = (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'\mathbf{Y}.$$

Next, the marginal posterior distribution of $\boldsymbol{\Sigma}$ has been computed as:

$$(\boldsymbol{\Sigma} \mid \text{data}) \sim W_m^{-1}(\mathbf{A}, \nu), \quad \nu > 0, \tag{8}$$

This is the m-dimensional inverse-Wishart with $\nu = n - (m + p) + 1$ degrees of freedom (d.f.). The $\mathbf{A}$ matrix is the $(m \times m)$ symmetric positive-semidefinite scale matrix defined as $(\mathbf{Y} - \mathbf{Z}\hat{\mathbf{B}})'(\mathbf{Y} - \mathbf{Z}\hat{\mathbf{B}})$. $\mathbf{A}$ is then proportional to the sample covariance matrix $\boldsymbol{\Sigma}$.

Notice that the analytical form of the inverse-Wishart distribution used by Geisser or Box and Tiao is slightly different from the one that may be found in other publications and that is usually implemented in software such as R (Dawid, 1981). If the former Wishart distribution has, say, ν d.f. ($\nu > 0$), the later have an equivalent of $\nu + m - 1$ d.f. ($\nu > m - 1$). This is of particular importance when comparing results with original references.

It is then possible to draw samples from the joint posterior distribution of the parameters using Equations (8) and (7).

Often, only the marginal posterior density of **B** is of interest as it contains the factor effects. It may be obtained by integrating **Σ** out of the joint posterior density and is:

$$(\mathbf{B} \mid \text{data}) \sim T_{(p \times m)}\left(\hat{\mathbf{B}}, \mathbf{A}, \left(\mathbf{Z}'\mathbf{Z}\right)^{-1}, \nu\right), \tag{9}$$

i.e. a matrix-variate Student's distribution with location $\hat{\mathbf{B}}$, scale matrices **A** and $\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}$ and $\nu = n - (m + p) + 1$ d.f. (Box and Tiao, 1973). The marginal density for a single line or column of (**B** | data) is a simpler multivariate Student's distributions (Gupta and Nagar, 1999).

The distribution of a new response vector $\tilde{\mathbf{y}}_{\tilde{\mathbf{x}}}$ at one new point $\tilde{\mathbf{x}} \in \chi$ is identified as a multivariate Student's distribution (Press, 1972):

$$(\tilde{\mathbf{y}} \mid \tilde{\mathbf{x}}, \text{data}) \sim T_m\left(\tilde{\mathbf{z}}\hat{\mathbf{B}}, \left(1 + \tilde{\mathbf{z}}'(\mathbf{Z}'\mathbf{Z})^{-1}\tilde{\mathbf{z}}\right)\mathbf{A}, \nu\right), \tag{10}$$

where $\tilde{\mathbf{z}}$ is the effect vector created from $\tilde{\mathbf{x}}$ and $\left(1 + \tilde{\mathbf{z}}'(\mathbf{Z}'\mathbf{Z})^{-1}\tilde{\mathbf{z}}\right)\mathbf{A}\nu^{-1}$ is the estimated spread matrix of the multivariate distribution, with ν d.f. ($\nu > 0$).

2.2.2 Informative prior distributions

In many domains, available data and expertise can be included in Bayesian statistical models through the use of informative prior distributions. For instance, in liquid chromatography, knowledge is high about the behavior of the responses (t_R) even when very little is known about the modeled compounds themselves. For instance, the correlations between the responses describing one chromatographic peak can be used as prior assumptions. This would decrease uncertainty of model parameters and highlights the fact that similar information is

modeled several times. It may also be known that some parameters do have a specific effect on the responses (e.g. the *gradient time* is known to increase the values of the t_R).

Prior distributions. In this section, the natural conjugate prior distributions are used (Press, 1972, §8.6.2). The prior distribution of $(\mathbf{B} \mid \Sigma)$ is defined as the $(p \times m)$-dimensional matrix-variate Normal distribution with mean $\mathbf{B}_0$ (same size than $\mathbf{B}$) and covariance matrices Σ and Σ_0.

$$\mathbf{B} \mid \Sigma \sim N_{(p \times m)}\left(\mathbf{B}_0, \Sigma, \Sigma_0\right). \tag{11}$$

The dependency on Σ leads to the restriction that the $(p \times p)$ prior covariance matrix Σ_0 is common for every m responses, i.e., all the corresponding regressors $\boldsymbol{\beta}_{1,\dots,j,\dots,m}$ have a similar prior covariance.

Equation (11) can be written as a classical multivariate Normal distribution, using the operator vec() applied on the matrix $\mathbf{B}$. vec($\mathbf{B}$) stacks the columns of $\mathbf{B}$ into a vector of length pm:

$$vec\left(\mathbf{B} \mid \Sigma\right) \sim N_{(pm)}\left(\text{vec}\left(\mathbf{B}_0\right), \Sigma \otimes \Sigma_0\right). \tag{12}$$

Next, a natural prior distribution for Σ is the following inverse-Wishart distribution:

$$\Sigma \sim W_m^{-1}(\mathbf{\Omega}, \nu_0), \tag{13}$$

where $\mathbf{\Omega}$ is the *a priori* responses scale matrix. ν_0 is the number of degrees of freedom of the prior distribution. The value of ν_0 indicates the certainty about $\mathbf{\Omega}$. For a simple interpretation of the prior distribution, ν_0 is defined in the same form as ν. That is, $\nu_0 = n_0 - (m + p) + 1$, with n_0 the number of virtual observation injected *a priori* in the model. Handling ν_0 (or n_0) allows the tuning of the prior subjectivity.

Notice that Equations (11) and (13) can be used to reduce the influence of unnecessary covariates. For instance, if one factor z_s is known to have a low influence on a particular response $\mathbf{y}_j$, the corresponding parameter β_{sj} can be given a prior with mean zero and small variance ($s = 1, ..., p$). However, the structure of the covariances of $\mathbf{B}$ is not convenient for prior parametrization (see Equation (12)).

Posterior distribution. Given these priors and applying the Bayes' theorem, the joint posterior density of $(\mathbf{B}, \boldsymbol{\Sigma})$ is:

$$p\left(\mathbf{B}, \boldsymbol{\Sigma} \mid \text{data}\right) \propto \mathcal{L}\left(\mathbf{B}, \boldsymbol{\Sigma} \mid \mathbf{Y}\right) p\left(\mathbf{B} \mid \boldsymbol{\Sigma}\right) p\left(\boldsymbol{\Sigma}\right) \tag{14}$$

$$\propto |\boldsymbol{\Sigma}|^{\frac{-n}{2}} \exp\left(-\frac{1}{2}tr\left[\boldsymbol{\Sigma}^{-1}\left(\mathbf{Y} - \mathbf{ZB}\right)^{'}\left(\mathbf{Y} - \mathbf{ZB}\right)\right]\right)$$

$$. |\boldsymbol{\Sigma}_0|^{\frac{-m}{2}} |\boldsymbol{\Sigma}|^{\frac{-p}{2}} \exp\left(-\frac{1}{2}tr\left[\boldsymbol{\Sigma}^{-1}\left(\mathbf{B} - \mathbf{B}_0\right)^{'} \boldsymbol{\Sigma}_0^{-1}\left(\mathbf{B} - \mathbf{B}_0\right)\right]\right)$$

$$. |\boldsymbol{\Omega}|^{\frac{\nu_0}{2}} |\boldsymbol{\Sigma}|^{-\frac{(\nu_0 + 2m)}{2}} \exp\left(-\frac{1}{2}tr\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{\Omega}\right]\right).$$

After some rearrangements, the posterior conditional distribution of $\mathbf{B}$ given $\boldsymbol{\Sigma}$ can be identified as the matrix-variate Normal:

$$\left(\mathbf{B} \mid \boldsymbol{\Sigma}, \text{data}\right) \sim N_{(p \times m)}\left(\mathbf{M}_{\mathbf{B}\text{post}}, \boldsymbol{\Sigma}, \left(\mathbf{Z}^{'}\mathbf{Z} + \boldsymbol{\Sigma}_0^{-1}\right)^{-1}\right), \text{ or equivalently,}$$

$$\left(\text{vec}\left(\mathbf{B}\right) \mid \boldsymbol{\Sigma}, \text{data}\right) \sim N_{(pm)}\left(\boldsymbol{\mu}_{\mathbf{B}\text{post}}, \boldsymbol{\Sigma} \otimes \left(\mathbf{Z}^{'}\mathbf{Z} + \boldsymbol{\Sigma}_0^{-1}\right)^{-1}\right), \tag{15}$$

where $\mathbf{M}_{\mathbf{B}\text{post}} = \left(\mathbf{Z}^{'}\mathbf{Z} + \boldsymbol{\Sigma}_0^{-1}\right)^{-1}\left(\mathbf{Z}^{'}\mathbf{Z}\hat{\mathbf{B}} + \boldsymbol{\Sigma}_0^{-1}\mathbf{B}_0\right)$, and $\boldsymbol{\mu}_{\mathbf{B}\text{post}}^{'} = \text{vec}(\mathbf{M}_{\mathbf{B}\text{post}})$.

Integrating $\mathbf{B}$ out of Equation (14) allowed retrieving the marginal density of $\boldsymbol{\Sigma}$, identified as the following inverse-Wishart distribution:

$$\left(\boldsymbol{\Sigma} \mid \text{data}\right) \sim W_m^{-1}\left(\boldsymbol{\Omega} + \mathbf{A}^*, \nu + n_0\right), \tag{16}$$

with $\mathbf{A}^* = \mathbf{Y}'\mathbf{Y} + \mathbf{B}_0'\boldsymbol{\Sigma}_0^{-1}\mathbf{B}_0 - (\mathbf{Z}'\mathbf{Z}\hat{\mathbf{B}} + \boldsymbol{\Sigma}_0^{-1}\mathbf{B}_0)'(\mathbf{Z}'\mathbf{Z} + \boldsymbol{\Sigma}_0^{-1})^{-1}(\mathbf{Z}'\mathbf{Z}\hat{\mathbf{B}} + \boldsymbol{\Sigma}_0^{-1}\mathbf{B}_0)$. When using uniform prior distributions, $\mathbf{A}^*$ is equal to the $\mathbf{A}$ matrix presented by Geisser, Box and Tiao. Section 3 will provide hints on the way to define the prior matrix $\boldsymbol{\Omega}$.

Finally, the marginal posterior density of $\mathbf{B}$ was obtained by integrating $\boldsymbol{\Sigma}$ out of the joint posterior density. In this case, it is:

$$(\mathbf{B} \mid \text{data}) \sim T_{(p \times m)}\left(\mathbf{M}_{\mathbf{B}\text{post}}, \boldsymbol{\Omega} + \mathbf{A}^*, \left(\mathbf{Z}'\mathbf{Z} + \boldsymbol{\Sigma}_0^{-1}\right)^{-1}, \, v + n_0\right) \tag{17}$$

The posterior d.f. is the sum of the d.f. coming from the likelihood (v), plus the number of virtual observations n_0 coming from the prior parameter distributions.

Predictive distribution. A new response vector $\tilde{\mathbf{y}}_{\tilde{\mathbf{x}}}$ at one new point $\tilde{\mathbf{x}} \in \chi$ can be obtained by drawing from the predictive distribution of $\tilde{\mathbf{y}}_{\tilde{\mathbf{x}}}$, whose density is:

$$p(\tilde{\mathbf{y}}_{\tilde{\mathbf{x}}} \mid \text{data}) = \int_{\mathbf{B},\boldsymbol{\Sigma}} p(\tilde{\mathbf{y}}_{\tilde{\mathbf{x}}} \mid \mathbf{B}, \boldsymbol{\Sigma}) p(\mathbf{B} \mid \boldsymbol{\Sigma}, \text{data}) p(\boldsymbol{\Sigma} \mid \text{data}) d(\mathbf{B}, \boldsymbol{\Sigma}) \tag{18}$$

One way to get samples from the predictive distribution of the responses is to draw samples from the joint posterior distribution of the parameters and to add a draw from the Normal distribution of the model:

For $s = 1$ to n^*

    Sample $(\mathbf{B}^{(s)}, \boldsymbol{\Sigma}^{(s)})$ from $p(\mathbf{B}, \boldsymbol{\Sigma} \mid data)$

    Sample $\tilde{\mathbf{y}}_{\tilde{\mathbf{x}}}^{(s)}$ from $p(\tilde{\mathbf{y}}_{\tilde{\mathbf{x}}} \mid \mathbf{B}^{(s)}, \boldsymbol{\Sigma}^{(s)}) \sim N(\mathbf{B}^{(s)}\tilde{\mathbf{z}}, \boldsymbol{\Sigma}^{(s)})$

End

where n^* is the total number of samples to be drawn. $\mathbf{B}^{(s)}, \boldsymbol{\Sigma}^{(s)}$ and $\tilde{\mathbf{y}}_{\tilde{\mathbf{x}}}^{(s)}$ are then the sth drawn values of the joint posterior distributions of $(\mathbf{B}, \boldsymbol{\Sigma})$ and from the distribution of $\tilde{\mathbf{y}}_{\tilde{\mathbf{x}}}$,

respectively. Notice this sampling scheme can also be used with samples of the posterior distributions of other models such as SUR, to get samples from the predictive distribution.

The integral in Equation (18) can be solved and the predictive distribution was identified as:

$$(\tilde{\mathbf{y}}_{\tilde{\mathbf{x}}} \mid \text{data}) \sim T_m \left(\mathbf{M}_{\mathbf{B}\text{post}} \tilde{\mathbf{z}}, \left(1 + \tilde{\mathbf{z}}'(\mathbf{Z}'\mathbf{Z} + \boldsymbol{\Sigma}_0^{-1})^{-1}\tilde{\mathbf{z}} \right)(\boldsymbol{\Omega} + \mathbf{A}^*), \nu + n_0 \right). \tag{19}$$

The multivariate RS model distribution is centered using the posterior mean of $\mathbf{B}$, instead of its least-square estimates. Notice that the spread matrix might be redefined given the specific implementation of the multivariate Student's distribution that will be used in practice. With uniform prior distributions, Equation (19) can be reduced to Equation (10).

Finally, the predictive distributions in both non-informative and informative cases are of β-expectation (Guttman, 1970). This means that the highest probability density (HPD) intervals derived from the predictive distribution are the β-expectation tolerance intervals (Hamada et al., 2004).

2.3 Multi-Criteria Decision Methods

Typically, when CQAs are complex combinations of several responses, their predictive distribution is not analytically tractable. In this case the technique of error propagation using Monte-Carlo simulations may be employed to obtain the distribution of the CQAs from the predictive distribution of the modeled responses.

With desirability functions. One of the most widespread techniques to combine several objectives into one value reflecting the quality of the solution is the desirability methodology (Harrington, 1965, Derringer and Suich, 1980). In general, this approach uses the desirability

concept without taking into account the uncertainty of parameters and measurements except by performing stability or sensitivity analysis around the optimum. This approach has been extended to focus on the distribution of the global desirability index under some classical distribution assumptions when desirability functions are used in response surface optimization (Steuer, 2000). It has also been shown that probabilistic derivations can be envisaged when the problem remains relatively simple (Le Bailly de Tilleghem and Govaerts, 2005). For more complex situations, simulations are practically preferred (Lebrun et al., 2008).

One benefit of the desirability approache is the possibility to define weights on each objective function (CQA) to be optimized. This allows a trade-off between each objective in the global desirability index. The deficiency of those approaches is that most of them still assume independent responses conditional to **X**. This assumption is justified in some cases, but, as in the example treated in the next section, there are many cases when this assumption can not be made and therefore ignoring dependencies can lead to misestimated risks. Again, the Bayesian framework provides an elegant solution to obtain the joint predictive distribution of the responses or CQAs to be included in a multi-criteria decision. In process optimization literature, it has been shown how to use desirability functions in a Bayesian framework using Monte-Carlo simulations (Peterson, 2004).

Without a desirability function. The Design Space definition in Equation (1) includes acceptance limits that are defined *a priori* on each CQA. Unlike the desirability approach, no trade-off assumption between the different objectives is made : all CQAs must achieve acceptance limits simultaneously. Obviously, if trade-off is allowed on the CQAs, or their acceptance limits, the interpretation of the results is more complicated.

15

For the present study, an appropriate optimization criterion is simply the joint predictive probability that acceptance limits are fully achieved. If this probability is higher than a minimal quality level (π), the solution can be said to belong to the DS. Let's consider the following simple example in which some acceptance limits λ_1 and λ_2 have been fixed for two arbitrary CQAs, o_1 and o_2:

$$\begin{cases} \text{Limit 1: } o_1 > \lambda_1, & \text{with } \lambda_1 = 0 \\[2ex] \text{Limit 2: } o_2 < \lambda_2, & \text{with } \lambda_2 = 3.95. \end{cases}$$

Using the SMR model and Monte-Carlo simulations, Assume the joint predictive distribution of the CQAs has been obtained for a certain point $\tilde{\mathbf{x}}$. Using samples from this predictive distribution, the proportion of them fulfilling the acceptance limits simultaneously is the Monte-Carlo estimate of the predictive probability of belonging to the DS:

$$P(o_1 > \lambda_1 \ \& \ o_2 < \lambda_2 \mid \tilde{\mathbf{x}}, \text{data}).$$

This is illustrated on Figure 1. In this case, the joint probability of achieving both criteria is roughly 0.4. With a minimum quality level of, say, $\pi = 0.8$, one cannot say $\tilde{\mathbf{x}}$ belongs to the DS.

3   Application

3.1   Introduction

In this intentionally small example, the proposed methodology is applied to chromatographic method development. The objective is to identify a set of factor settings where outputs (chromatograms) will have satisfactory CQAs. An ultraviolet (UV) chromatogram is a signal of
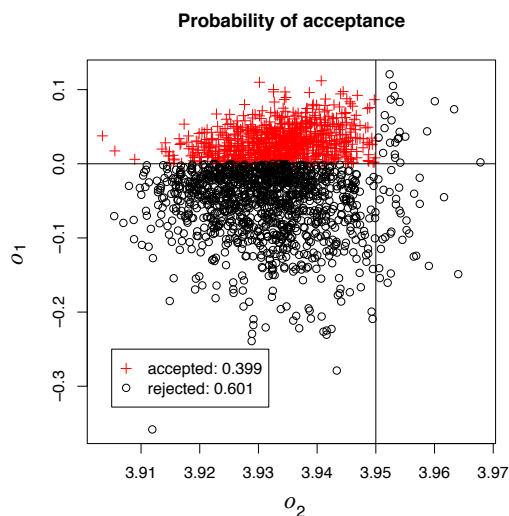
Figure 1: Example. Computation of the predictive probability of acceptance as the proportion of predicted samples lying within acceptance limits (+).

absorbance values that were recorded over a certain time period. Compounds are identified by absorbance peaks. If these peaks are separated, it means that the compounds have been physically separated, which is a desirable situation. A High Performance Liquid Chromatography equipment (HPLC) is a tunable device that carries out this separation when its tuning parameters are well adjusted. Many qualitative and quantitative tuning parameters can be used, such as pH of mobile phase, composition of buffer, type of analytical column or solvent used, etc. Designs of Experiments (DOE) provides an efficient plan to analyze the responses.

It was decided to model a simple log transformation of the chromatographic retention times (t_R), i.e. the times peaks are observed, because they are generally log-linear in their tuning parameters (Massart et al., 1990, Snyder et al., 1997). It was noted that they provide sufficient information to derive CQAs of interest (Dewé et al., 2004, Lebrun et al., 2008).

3.2 Software utilized

Computer programs have been developed for *R* 2.14 (R Development Core Team, 2011, Martin et al., 2011, Genz et al., 2011). Empower V2.0 for Windows was used to control a Waters 2695 separation module equipped with a Waters 996 Photodiode array detector and to gather and interpret chromatograms. Further information on HPLC, chemicals, samples preparation and experiments can be found in the reference, including the definition of the factors and their levels (Debrus et al., 2011a).

3.3 Design and responses definition

A DOE consisting of *n* experiments was conducted, consisting of applying each of the designed input conditions $\mathbf{x}_i = (x_{i1}, \ldots, x_{ik})$ to a mixture of compounds. Experiments provided *n* chromatograms, each with several overlapped or separated peaks.

A full-factorial design comprising three factors ($k = 3$) was set up to separate three compounds ($f = 3$) : the maleate, the phenilephrine and the saccharin. No prior knowledge was assumed on these compounds. Two *solvents* were used as a qualitative factor (Methanol or Acetonitrile) while the *pH* of mobile phase and *gradient time* were both quantitative factors. From past experiments and literature, it is known that the *pH* effect can be locally approximated by a quartic effect on the responses (*pH* has theoretically a sigmoid effect on retention times), and that *gradient time* generally have a quadratic effect on the responses (Snyder et al., 1997). The DOE has been chosen accordingly to estimate these effects.

The selected design resulted in 30 experiments. Replicates were added at center points for each level of qualitative factor, leading to $n = 32$ experiments. In practice, this high number

of experiments is not mandatory. It may be wise to select smaller ranges for factors having a complex effect on the responses (e.g. *pH*), and to use more appropriate designs such as central composite or Doehlert designs. For chromatographic applications, it has also been shown that the analysis of many compounds simultaneously permit the reduction of the cost *by compound*, since the experiments are more informative (Debrus et al., 2011b).

Figure 2 illustrates a schematic view of a chromatogram with two peaks at a certain condition, and with t_0, the dead time of the HPLC system. For each peak, three times can be extracted: t_R, i.e., the time at the maximum, known as the apex (A), the times at the beginning (B) and at the end (E) of the peaks at baseline-height. These observed times form three vectors of size n for each of the f peaks. They can be stored in a response matrix $\mathbf{Y}_{tR} = (\mathbf{Y}_1^B, \mathbf{Y}_1^A, \mathbf{Y}_1^E, ..., \mathbf{Y}_f^B, \mathbf{Y}_f^A, \mathbf{Y}_f^E)$. $\mathbf{Y}_{tR}$ is then of size $(n \times m)$, with $(m = 3 * f)$. Peaks-to-compounds identification is a crucial step and was mainly manually performed, although automated reading of chromatograms might be achieved by signal processing techniques (Tee-Won, 2001, Wang et al., 2008, Debrus et al., 2009).
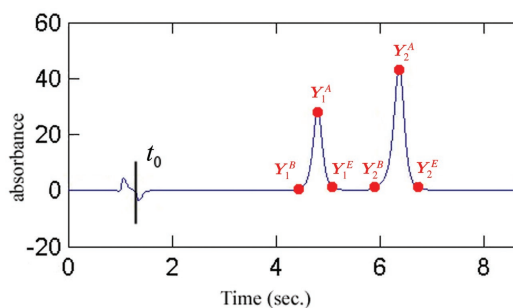

Figure 2: Raw chromatogram with the positions of recorded times. t_0 is the dead time of the HPLC system.

To account for log-linearity, it is common to use responses that are the logarithm of retention factors (Snyder et al., 1997), computed as $\mathbf{Y}_{resp} = \log\left(\frac{\mathbf{Y}_{tR} - t_0}{t_0}\right)$.

To simplify notations, assume $\mathbf{Y}$ stands for $\mathbf{Y}_{resp}$ in the subsequent sections. Residuals correlations can be found among the columns of the response matrix $\mathbf{Y}$, as shown in Table 1 (bold). There are two possible reasons for this. The main reason is that the three parts of one peak j, (Y_j^B, Y_j^A, Y_j^E) describe the same entity. As the peak is exposed to different input conditions, one can expect that its three parts move together. The second reason is the similar chromatographic behavior observed with highly similar, yet different compounds such as isomers or enantiomers. This is particularly frequent when working on pharmaceutical formulations or impurities, but this is not observed in this study. The informative SMR model makes use of these correlations through the $\mathbf{A}^*$ and $\mathbf{\Omega}$ matrices.

	$\mathbf{Y}_1^B$	$\mathbf{Y}_1^A$	$\mathbf{Y}_1^E$	$\mathbf{Y}_2^B$	$\mathbf{Y}_2^A$	$\mathbf{Y}_2^E$	$\mathbf{Y}_3^B$	$\mathbf{Y}_3^A$	$\mathbf{Y}_3^E$
$\mathbf{Y}_1^B$	**1.00**	**0.99**	**0.95**	0.13	0.14	-0.06	0.06	0.09	0.11
$\mathbf{Y}_1^A$	**0.99**	**1.00**	**0.93**	0.12	0.09	-0.01	0.12	0.15	0.18
$\mathbf{Y}_1^E$	**0.95**	**0.93**	**1.00**	0.27	0.29	0.00	-0.02	0.01	-0.03
$\mathbf{Y}_2^B$	0.13	0.12	0.27	**1.00**	**0.90**	**0.72**	0.36	0.36	0.11
$\mathbf{Y}_2^A$	0.14	0.09	0.29	**0.90**	**1.00**	**0.43**	0.22	0.20	-0.03
$\mathbf{Y}_2^E$	-0.06	-0.01	0.00	**0.72**	**0.43**	**1.00**	0.56	0.60	0.48
$\mathbf{Y}_3^B$	0.06	0.12	-0.02	0.36	0.22	0.56	**1.00**	**0.98**	**0.89**
$\mathbf{Y}_3^A$	0.09	0.15	0.01	0.36	0.20	0.60	**0.98**	**1.00**	**0.91**
$\mathbf{Y}_3^E$	0.11	0.18	-0.03	0.11	-0.03	0.48	**0.89**	**0.91**	**1.00**

Table 1: Residual responses correlations

3.4 Residual correlations and constrained sampling

When drawing new samples from the predictive distribution of the responses, nothing ensures that, for each peak j and for each point $\tilde{\mathbf{x}} \in \chi$, the following constraints are respected:

$$\mathbf{Y}_j^B < \mathbf{Y}_j^A \mid \tilde{\mathbf{x}},$$

$$\text{and } \mathbf{Y}_j^A < \mathbf{Y}_j^E \mid \tilde{\mathbf{x}}. \tag{20}$$

When not fulfilled, *ill-formed* peaks are drawn. Notice that high response correlations do not imply the constraints to be respected.

Naive rejection sampling can be employed: samples of the distribution not achieving the constraints are discarded. However, if there are a lot of peaks, and if they are tight, the ratio of rejected to accepted samples can be very high. Thus, the application of such constraints can be time consuming.

Another possibility is to use response transformations, as proposed previously in the chromatographic context (Lebrun et al., 2008). However, to explain the definition of possible prior assumptions, these transformations were not carried out here.

Improvement of the naive rejection sampling can be obtained when including an informative prior distribution for $\boldsymbol{\Sigma}$ in the regression model. Generally, priors are used to reduce the uncertainty of prediction, but a further interest here is to decrease the number of rejected draws not respecting the constraints. In the non-informative case, rejection might frequently occur when few observations are available to compute the $\mathbf{A}$ or $\mathbf{A}^*$ matrices, giving a poor representation of the actual covariance structure of the data.

To account for known correlations, the matrix $\boldsymbol{\Omega}$ was defined from a similarly sized $\boldsymbol{\Omega}_{\text{cor}}$ that contains prior correlations. It was then rescaled so that it is proportional to $\mathbf{A}^*$. The

following prior parameters have been defined:

$$\mathbf{\Omega}_{\text{cor}} = I_P \otimes \begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{bmatrix}, \tag{21}$$

where $\otimes$ is the Kronecker product. Strong correlation links are assumed for the parts of every

peak. Accordingly, ρ was chosen to be 0.95. Finally, the elements of $\mathbf{\Omega} \equiv \{\Omega_{ij}\}$ were defined

as:

$$\{\Omega_{ij}\} = \frac{\sqrt{\{a_{ii}^*\}} \cdot \{\Omega_{\text{cor},ij}\} \cdot \sqrt{\{a_{jj}^*\}}}{\nu} * \nu_0, \quad \text{with } i, j = 1, ..., m, \tag{22}$$

with $\{a_{ii}^*\}$ being the ith element of the diagonal of $\mathbf{A}^*$. To remain as objective as possible, ν_0

was kept low, with a value of 3. The $\mathbf{A}^*$ matrix is then assumed correctly estimated, thanks

to the high number of experiments.

A last possibility in order to avoid naive rejection sampling is to use a truncated version

of the multivariate Student's distribution to directly generate well-formed samples (Geweke,

1991). The problem is then to draw samples from the predictive distribution $\tilde{\mathbf{y}}_{\tilde{\mathbf{x}}} \sim T_m(...)$

under the constraints $\mathbf{a} \leq \mathbf{D}\tilde{\mathbf{y}}'_{\tilde{\mathbf{x}}} \leq \mathbf{b}$, with:

$$\mathbf{a} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ -\infty \\ -\infty \\ -\infty \end{pmatrix}, \quad \mathbf{D} = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} +\infty \\ +\infty \\ +\infty \\ +\infty \\ +\infty \\ +\infty \\ +\infty \\ +\infty \\ +\infty \end{pmatrix} \tag{23}$$

The first six rows of $\mathbf{a}$, $\mathbf{D}$ and $\mathbf{b}$ make explicit the applied constraints while the last three lines

represent no constraints and are defined to avoid singular matrix problems in the sampling

algorithm of Geweke.

However, Geweke insisted on the fact that this sampler can have convergence issues when the correlations between responses are high, as is the case in our application. A careful comparison of the results using the truncated multivariate Student's distribution and classical rejection sampling should be employed, e.g. by comparing marginal densities or tolerance intervals at various level of confidence. A *C* implementation of the sampler for the *R* language was developed and is available on demand.

3.5   Summary of the fit

The matrix $\mathbf{Z}$ was created to contain an intercept, the main effects, the second and third order interactions and the polynomial effects. This gave a matrix $\mathbf{Z}$ of size $(n \times p) = (32 \times 12)$. Prior information for $\mathbf{B} \mid \mathbf{\Sigma}$ was set to be non-informative ($\mathbf{B}_0$ and $\mathbf{\Sigma}_0^{-1}$ were defined as 0 everywhere), and the previously described prior assumption for $\mathbf{\Sigma}$ was used.

Figure 3 shows the actual versus mean predicted plot and the residuals plot for all the responses combined. A brief look at these plots shows good adjustment of the model. Moreover, adjusted R^2 for each response is greater than 0.95, allowing the multivariate model to be well suited for prediction.

Notice that the multivariate distribution of the model errors should be checked, as the probabilistic statement of the Design Space is sensitive to their supposed multivariate normal assumptions. Obviously, Q–Q plots of the marginal errors are a first step towards normality check. However, this is generally non sufficient when dealing with multivariate variables. For instance, clusters or structures in the residuals could be found whereas the univariate Normality tests would pass. Del Castillo et al. (2012) discussed about these issues. Timm

(2002) (pp. 118-130) or Royston (1995) proposed various multivariate normality tests. Such tests are not performed in this text for brevity.
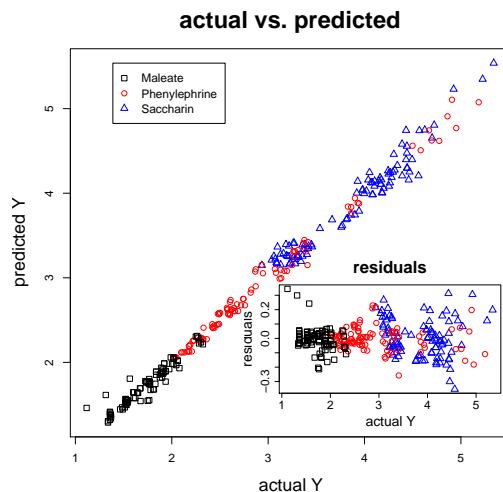

Figure 3: Plot of actual values of the responses versus the mean predicted ones. Included is the plot of residuals.

Figure 4 aims at representing the uncertainty of the responses prediction, following Equation (19). One limitation is given by the dimensionality of the problem that only permits the representation of marginal distributions or bivariate plots to observe the effects of correlations. Figures 4(A) and 4(B) exemplify correlated responses while Figures 4(C) and 4(D) shows no significant correlation.

3.6 Objective functions for MCDM - CQAs definition

Various objective functions (CQAs), derived from the untransformed responses Y, were used to assess the quality of the chromatograms. Here, the general objective was to observe if the peaks will be separated in a short analysis time. Some examples of CQAs are listed in Equation (24): o_1 is an objective function representing the *total run time*, to be minimized; o_2 and o_3 are respectively the chromatographic *critical separation* and *critical resolution* (R_s),

that both must be maximized. Notice that o_z is a simplified notation for $(o_z \mid \tilde{\mathbf{x}}, \text{data})$, the posterior distribution of the zth objective function at the point $\tilde{\mathbf{x}}$ ($z = 1, 2, 3$). The envisaged CQAs were:

$$o_1 = \max_{1 \leq j \leq f} (Y_j^A) \qquad\qquad < \lambda_1,$$

$$o_2 = \min_{1 \leq j \leq f-1} (Y_{(j+1)}^B - Y_{(j)}^E) \qquad\qquad > \lambda_2,$$

$$R_s = o_3 = \min_{1 \leq j \leq f-1} \left(\frac{2 * (Y_{(j+1)}^A - Y_{(j)}^A)}{(Y_{(j+1)}^E - Y_{(j+1)}^B) + (Y_{(j)}^E - Y_{(j)}^B)} \right) \qquad > \lambda_3,$$

$$\tag{24}$$

where $Y_{(j)}^B$, $Y_{(j)}^A$ and $Y_{(j)}^E$ represent the parts of the jth ordered peak with respect to the mean predicted times of the apexes. All proposed acceptance limits are unilateral. Bilateral limits can be used as well but are not useful in the present case. o_3 is constrained on positive real values. It is worth noting that if the constraints applied on $\mathbf{Y}$ are not respected, o_3 would
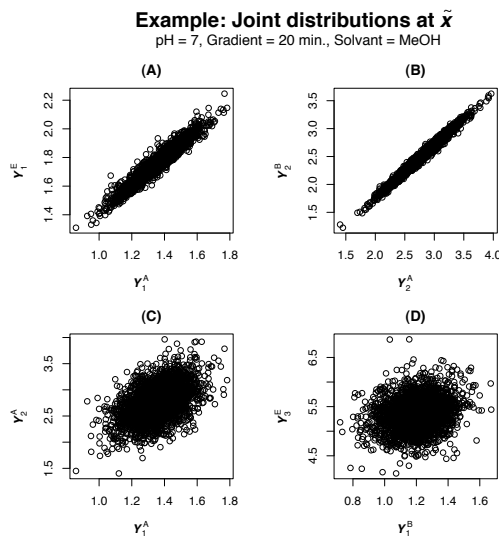

Figure 4: Illustration of the bivariate joint predictive distribution for some responses. (A) and (B) show correlation between responses while (C) and (D) show non-significant correlation.

25

show an unusual behavior, especially if its denominator is negative or very close to 0. o_1 and o_2 would suffer less from constraint problems and are generally preferred in an optimization context.

Regrouping the CQAs in the vector $\boldsymbol{O} = (o_1, o_2, o_3)'$ and defining $\boldsymbol{\Lambda}$ as the region defined jointly by the acceptance limits: $\boldsymbol{\Lambda} = \{\tilde{\mathbf{x}} \in \chi \mid o_1 < \lambda_1 \cap o_2 > \lambda_2 \cap o_3 > \lambda_3\}$, DS definition can be applied to the CQAs instead of the responses:

$$\text{DS} = \{\tilde{\mathbf{x}} \in \chi \mid P(\boldsymbol{O}_{\tilde{\mathbf{x}}} \in \boldsymbol{\Lambda} \mid \tilde{\mathbf{x}}, \text{data}) \geq \pi\}, \tag{25}$$
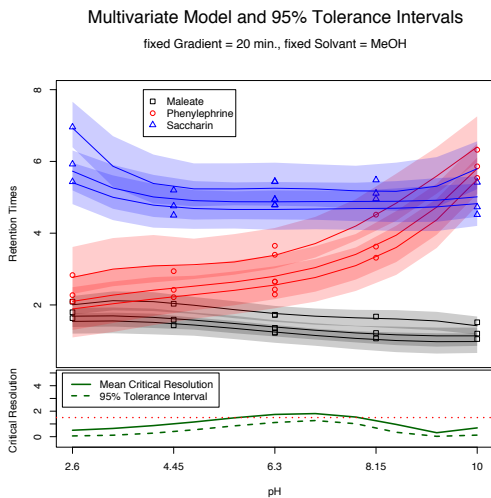
Figure 5: Fit of the multivariate responses and 95% β-expectation tolerance intervals. Top: fit of the original responses. Bottom: Objective function o_3. $\lambda_3 = 1.5$ is shown with a dotted red line.

The fit of the responses vs. the factor *pH* is shown on the top of Figure 5. The *Gradient time* is fixed at 20 minutes and the *Solvent* is Methanol. o_3, the critical resolution, is represented on the bottom. This CQA is maximized when the peaks are as separated as possible. Uncertainty of responses was directly propagated and allowed showing the predictive intervals on this CQA. Experts generally agree that a minimal resolution of at least 1.5 ($\lambda_3 = 1.5$)

26

is mandatory to have well separated peaks. It can be seen that the mean critical resolution is maximized around pH of 7 and is slightly better than 1.5. However, when observing the one-sided predictive interval, one cannot say that this resolution will be achieved with a 0.95 probability.

Similarly, a critical separation of 0 minutes ($\lambda_2 = 0$) with a maximum run time of 6 minutes ($\lambda_1 = 6$) was decided to be desirable situation. The minimal quality level was set at 85% ($\pi = 0.85$). Input conditions showing predicted CQAs jointly accepted with a probability π were declared to belong to the DS.

Drawing 2000 samples out of the joint predictive distribution of the CQAs permitted to draw Figure 6. On the left plot is the joint predictive distribution of criteria o_1 and o_2 conditional to $\tilde{\mathbf{x}} = (pH = 7, \textit{Gradient time} = 20, \textit{Solvent} = \text{MeOH})$. This plot shows a slight correlation between objective functions. On the right plot, a stronger dependency between o_2 and o_3 is observed. For this input condition, the estimated probability to have the objectives o_1, o_2 and o_3 jointly accepted was about 0.85. Not taking into account the correlations would increase the risk of making a wrong decision.
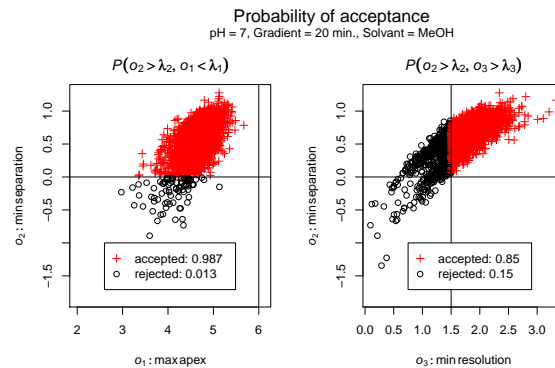

Figure 6: Left: joint predictive distribution of objective function o_1 and o_2. Right: joint predictive distribution of objective function o_2 and o_3. In red (+), the sampled points of predictive distribution achieving acceptance limits.
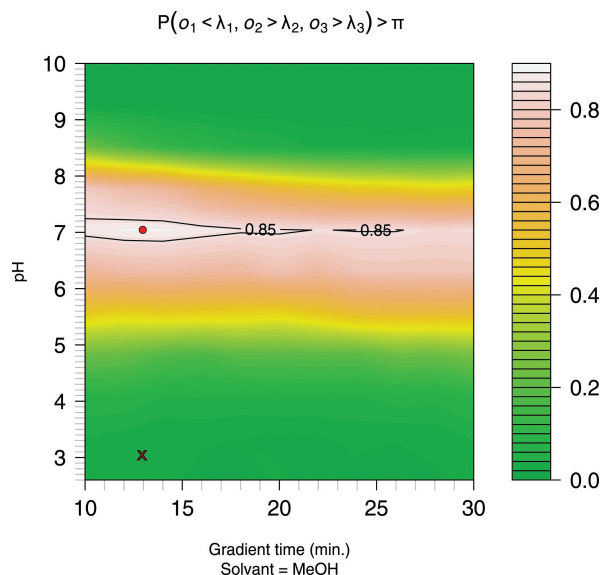
Figure 7: Contour plot of DS with qualitative factor *Solvent* fixed to MeOH. The red bullet (o) indicates the optimal probability to observe the CQAs into acceptance limits. This optimal probability reaches 0.89. Black lines are placed at $\pi = 0.85$. Cross (x) indicates a sub-optimal input condition.

3.7 Design Space

Reporting the values of estimated probability for each $\tilde{x} \in \chi$ allowed the drawing of the DS, as in Figure 7. When the number of quantitative factors of interest is small, the DS can elegantly be represented on graphs of contour plots of the expected joint probability of acceptance over the experimental domain. For multi-dimensional DS, treillis of contour plots still remains convenient to represent bivariate projections of the DS, but practicalities in using these graphs is low when the number of dimensions increases. It was shown that using simple sortable spreadsheets is a possible non-graphical alternative (Peterson, 2008).

The DS for the minimum quality level of $\pi = 0.85$ is identified by the black lines and is relatively large according to the *Gradient time* factor, but narrower for the *pH*. A more

desirable situation would occur if stronger acceptance limits could be identified with a higher probability of achieving them (e.g. $\pi = 0.95$). Such DS does not exist in this case. Notice that π should drive the decision process regarding the future use of the method and its expected quality. In the DS, the maximal risk of obtaining non satisfactory or out-of-specification results is also prominent and is simply computed as $1 - \pi$.

At the optimal input condition, $\tilde{\mathbf{x}}^* = (pH = 7.05,\ Gradient\ time = 13,\ Solvent = \text{MeOH})$, the probability to meet jointly the three CQAs is 0.89.

3.8 Validation

A small validation has been carried out including two additional experiments. The first is the optimal point lying within the DS. The second is a sub-optimal factor setting. Values for the three input factors are presented in Table 2.

	Gradient time	pH	Solvent
Optimal (o)	13	7.05	MeOH
Sub-optimal (x)	13	3	MeOH

Table 2: Validation experiments

The resulting chromatograms are presented in Figure 8. The top chromatograms are constructed from mean predicted untransformed responses ($\mathbf{Y}_{tR}$, see section 3.3). The bottom chromatograms are the ones observed with the new experiments. Left plots show the optimal solution (o) while right plots show a sub-optimal one (x). Comparing top and bottom plots illustrates the accuracy of the predictive multivariate model developed to fit the peak's retention times.
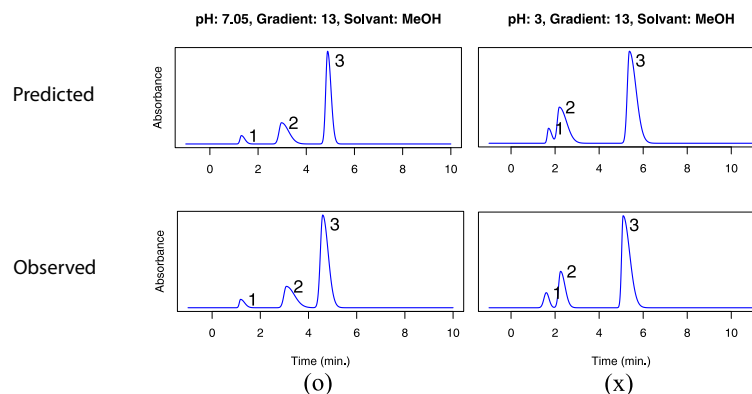
Figure 8: Validation results. Top: predicted chromatograms at optimal condition (o) and at sub-optimal condition (x). Bottom: experimental chromatograms at the same input conditions.

4 Conclusion

Following the recommendations of the ICH Q8 guideline to adopt a systematic quality-by-design approach for the early development of pharmaceutical processes, the proposed methodology translated the concept of Design Space to analytical separation methods.

When making prediction, it has been indicated that uncertainty of models should be taken into account to obtain plausible results. In this context, Bayesian multivariate multiple regression was shown to be powerful and flexible enough to deal with correlated responses and parameters. Moreover, the Bayesian framework and Monte-Carlo simulations allow for a straightforward computation of Design Spaces, even in complex situations when CQAs might be discontinuous functions of the input conditions.

In the particular context of analytical method development, Bayesian models applied in chromatography can take advantage from the long history of the domain to set up informative prior distributions on the parameters.

Constrained sampling is a consistent way of getting the predictive distribution of re-

30

sponses that must fulfill certain requirements. When constraints can be directly applicable, the use of a truncated multivariate Student's distribution might simplify the sampling.

5   Acknowledgements

References

G.E.P. Box and G.C. Tiao. Bayesian Inference in Statistical Analysis. Wiley Classic Library, 1973.

A.P. Dawid. Some Matrix-variate Distribution Theory: Notational Considerations and a Bayesian Application. Biometrika, 68(1):264–274, 1981.

B. Debrus, P. Lebrun, A. Ceccato, G. Caliaro, B. Govaerts, B.A. Olsen, E. Rozet, B. Boulanger, and Ph. Hubert. A new statistical method for the automated detection of peaks in UV-DAD chromatograms of a sample mixture. Talanta, accepted for publication, 2009.

B. Debrus, P. Lebrun, A. Ceccato, G. Caliaro, E. Rozet, I. Nistor, R. Oprean, F.J. Ruperez, C. Barbas, B. Boulanger, and Ph. Hubert. Application of new methodologies based on

design of experiments, independent component analysis and design space for robust opti-
mization in liquid chromatography. Anal. Chim. Acta, 691(1-2):33–42, 2011a.

B. Debrus, P. Lebrun, J. Mbinze Kindenge, F. Lecomte, A. Ceccato, G. Caliaro,
J. Mavar Tayey Mbay, B. Boulanger, R.D. Marini, E. Rozet, and Ph. Hubert. Innova-
tive High-performance Liquid Chromatography Method Development for the Screening
of 19 Antimalarial Drugs Based on a Generic Approach, Using Design of Experiments,
Independent Component Analysis and Design Space. J. Chrom. A, 1218(31):5205–5215,
2011b. doi: http://hdl.handle.net/2268/93241.

E. Del Castillo, B.M. Colosimo, and H. Alshraideh. Bayesian Modeling and Optimization
of Functional Responses Affected by Noise Factors. J. Qual. Tech., 44(2):117–135, 2012.

G.C. Derringer and R. Suich. Simultaneous Optimization of Several Response Variables. J.
Qual. Tech., 12(4):214–219, 1980.

W. Dewé, R.D. Marini, P. Chiap, Ph. Hubert, J. Crommen, and B. Boulanger. Development
of response models for optimizing HPLC methods. Chemom. Intell. Lab. Syst., 74:263–
268, 2004.

S. Geisser. Bayesian Estimation in Multivariate Analysis. Ann. Math. Statist., 36 (1):150–
159, 1965.

A. Genz, F. Bretz, T. Miwa, X. Mi, F. Leisch, F. Scheipl, and T. Hothorn. mvtnorm:
Multivariate Normal and t Distributions. http://CRAN.R-project.org/package=mvtnorm,
2011.

J. Geweke. Efficient Simulation from the Multivariate Normal and Student-t Distributions Subject to Linear Constraints and the Evaluation of Constraint Probabilities. In Computing Science and Statistics: the 23rd Symposium on the Interface, Seattle, April 1991.

A.K. Gupta and D.K Nagar. Matrix Variate Distributions. Chapman and Hall/CRC, 1999.

I. Guttman. Statistical Tolerance Regions: classical and bayesian. Griffin's statistical monographs and courses, 1970.

M. Hamada, V. Johnson, and L.M. Moore. Bayesian Prediction Intervals and Their Relationship to Tolerance Intervals. Technometrics, 46(4):452–459, 2004.

E.C. Harrington. The desirability function. Ind. Qual. Control, 21:494–498, 1965.

ICH Q8. Guidance for Industry, Q8 Pharmaceutical Development, 2006. International Conference for Harmonization.

C. Le Bailly de Tilleghem and B. Govaerts. Distribution of Desirability Index in Multi-criteria Optimization using Desirability Functions based on the Cumulative Distribution Function of the Standard Normal. Technical Report 0531, Université catholique de Louvain, Louvain-la-Neuve, 2005.

P. Lebrun, B. Govaerts, B. Debrus, A. Ceccato, G. Caliaro, Ph. Hubert, and B. Boulanger. Development of a new predictive modelling technique to find with confidence equivalence zone and design space of chromatographic analytical methods. Chemom. Intell. Lab. Syst., 91:4–16, 2008.

A.D. Martin, K. M. Quinn, and J. Hee Park. MCMCpack: Markov chain monte carlo in R. J. Statist. Soft., 42(9):22, 2011. URL http://www.jstatsoft.org/v42/i09/.

D.L. Massart, S.N. Deming, Y. Michotte, L. Kaufman, and B.G.M. Vandeginste. Chemometrics: A Textbook (Data Handling in Science and Technology). Elsevier Science, 1990. ISBN 978-0444426604.

G. Miró-Quesada, E. del Castillo, and J.J. Peterson. A Bayesian approach for multiple response surface optimization in the presence of noise variable. J. Appl. Statist., 13 (3): 251–270, 2004.

J.J. Peterson. A posterior predictive approach to multiple response surface optimization. J. Qual. Tech., 36:139–153, 2004.

J.J. Peterson. A Bayesian approach to the ICH Q8 definition of Design Space . J. Biopharm. Statist., 18:959–975, 2008.

J.J. Peterson and K. Lief. The ICH Q8 Definition of Design Space: A Comparison of the Overlapping Means and the Bayesian Predictive Approaches. Statistics in Biopharmaceutical Research, 2:249–259, 2010.

J.J. Peterson and M. Yahyah. A Bayesian Design Space Approach to Robustness and System Suitability for Pharmaceutical Assays and Other Processes. Statistics in Biopharmaceutical Research, 1(4):441–449, 2009.

J.J. Peterson, G. Miró-Quesada, and E. del Castillo. A Bayesian reliability approach to multiple response optimization with seemingly unrelated regression models . J. Qual. Technol. Quant. Management, 6:353–369, 2009.

S. J. Press. Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference. R.E. Krieger Pub. Co., 1972.

R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2011. URL `http://www.R-project.org/`. ISBN 3-900051-07-0.

J.P. Royston. Remark AS R94: A Remark on Algorithm AS 181: The W-Test for Normality. J. Roy. Statist. Soc. C (Applied Statistics), 44(4):547–551, 1995.

P. Schoenmakers. Optimization of chromatographic selectivity: a guide to method development. Anal. Chim. Acta, 208:357–358, 1988.

L.R. Snyder, J.J. Kirkland, and J.L. Glajch. Practical HPLC Method Development, second Edition. Wiley-Interscience, 1997. ISBN 978-0471007036.

D. Steuer. An Improved Optimisation Procedure for Desirability Indices. Technical Report 27/00, SFB 475, Dortmund University, 2000.

G.W. Stockdale and A. Cheng. Finding Design Space and a Reliable Operating Region using a multivariate Bayesian approach with experimental design. J. Qual. Technol. Quant. Management, 4:391–408, 2009.

L. Tee-Won. Independent Component Analysis, theory and applications. Kluwer Academic Publishers, 2001. ISBN 0-7923-8261-7.

N.H. Timm. Applied Multivariate Analysis. Springer-Verlag, New-York, N-Y, 2002.

P.F. Vanbel. Development of flexible and efficient strategies for optimizing chromatographic conditions. J. Pharm. Biomed. Anal., 21:603–610, 1999.

G. Wang, Q. Ding, and Z. Hou. Independent Component Analysis and its Applications in

Signal Processing for Analytical Chemistry. Trends in Analytical Chemistry, 37(4):368–

376, 2008.