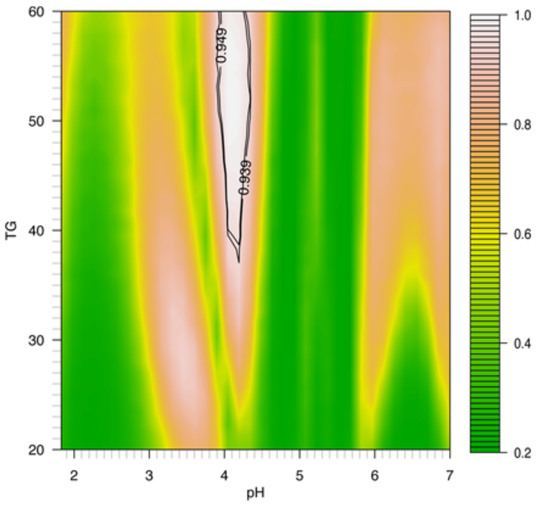
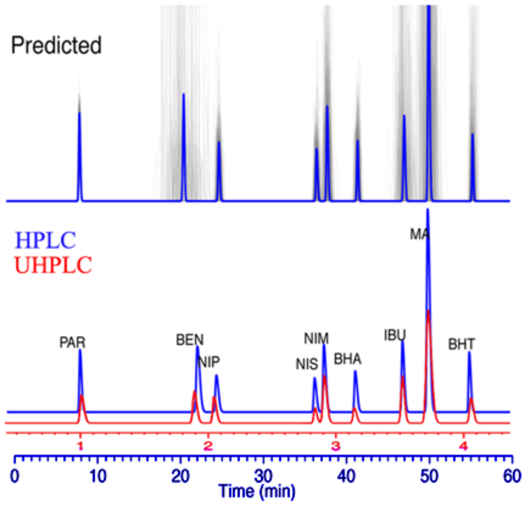
UNIVERSITE DE LIÈGE
Service de Chimie Analytique, Faculté de Médecine
Département de Mathématiques, Faculté des Sciences

Bayesian Design Space applied to Pharmaceutical Development





Thèse présentée en vue de l'obtention du grade de
Docteur en Sciences, orientation statistique, par:
Pierre Lebrun

Membres du Jury

Prof. Adelin Albert
Dr. Bruno Boulanger
Prof. Paul Eilers
Prof. Bernadette Govaerts
Prof. Gentiane Haesbroeck
Prof. Philippe Hubert
Prof. Philippe Lambert
Dr. John Peterson

Liège, May 2012

Abstract

Given the guidelines such as the Q8 document published by the International Conference on Harmonization (ICH), that describe the "Quality by Design" paradigm for the Pharmaceutical Development, the aim of this work is to provide a complete methodology addressing this problematic. As a result, various Design Spaces were obtained for different analytical methods and a manufacturing process.

In Q8, Design Space has been defined as the "the multidimensional combination and interaction of input variables (e.g., material attributes) and process parameters that have been demonstrated to provide assurance of quality" for the analytical outputs or processes involved in Pharmaceutical Development. Q8 is thus clearly devoted to optimization strategies and robustness studies.

In the beginning of this work, it was noted that existing statistical methodologies in optimization context were limited as the predictive framework is based on mean response predictions. In such situations, the data and model uncertainties are generally completely ignored. This often leads to increase the risks of taking wrong decision or obtaining unreliable manufactured product. The reasons why it happens are also unidentified. The "assurance of quality" is clearly not addressed in this case.

To improve the predictive nature of statistical models, the Bayesian statistical framework was used to facilitate the identification of the predictive distribution of new outputs, using numerical simulations or mathematical derivations when possible.

By use of the improved models in a risk-based environment, separation analytical methods such as the high performance liquid chromatography were studied. First, optimal solutions of separation of several compounds in mixtures were identified. Second, the robustness of the methods was simultaneously assessed thanks to the risk-based Design Space identification. The usefulness of the methodology was also demonstrated in the optimization of the separation of subsets of relevant compounds, without additional experiments.

The high guarantee of quality of the optimized methods allowed easing their use for their very purpose, i.e., the tracing of compounds and their quantification. Transfer of robust methods to high-end equipments was also simplified.

In parallel, one sub-objective was the total automation of analytical method development and validation. Some data treatments including the Independent Component Analysis and clustering methodologies were found more than promising to provide accurate automated results.

Next, the Design Space methodology was applied to a small-scale spray-dryer manufacturing process. It also allowed the expression of guarantees about the quality of the obtained powder.

Finally, other predictive models including mixed-effects models were used for the validation of analytical and bio-analytical quantitative methods.

Résumé

Afin de répondre aux exigences dictées par les documents traitant de la problématique du “Quality by Design”, tels ICH Q8 publié par la Conférence Internationale d’Harmonisation (ICH), l’objectif de ce travail est de fournir une méthodologie adéquate permettant le calcul du Design Space pour les différentes méthodes analytiques et procédés de fabrication liés au développement de médicaments.

Le Design Space (ou Espace de Conception) a été défini comme l’ensemble des conditions opératoires d’une méthode ou d’un procédé qui permettent de garantir une haute qualité du résultat ainsi optimisé, que ce soit une réponse analytique, un produit fini ou intermédiaire dans le processus de fabrication.

En analysant les méthodologies existantes permettant une telle optimisation, il a été noté que l’aspect prédictif des modèles statistiques, utilisés notamment en planification expérimentale, était généralement basé sur la prédiction de réponses moyennes. L’analyse de l’incertitude présente dans les données et les modèles n’est malheureusement pas souvent faite. Cela a pour conséquence l’obtention de résultats peu fiables dont les causes sont méconnues. Les “garanties de qualité” ne sont clairement pas considérées dans ces conditions.

Dans cette optique, la statistique Bayésienne a été utilisée afin de fournir un cadre de travail dans lequel l’analyse de l’incertitude prédictive est facilitée. Différents modèles statistiques ont ainsi été définis et la distribution de leurs prédictions a été dérivée mathématiquement, ou en utilisant des méthodes de simulations numériques.

En combinant ces modèles statistiques prédictifs avec une approche basée sur l’analyse du risque, les méthodes analytiques séparatives –telle la chromatographie liquide à haute performance– ont fait l’objet de différentes études. D’une part, il a ainsi été possible d’optimiser les conditions de séparation de composés de plusieurs mélanges. D’autres part, l’analyse de la robustesse de ces méthodes a pu être faite simultanément grâce à l’analyse du risque dans le Design Space. L’utilisation de sous-ensembles de données comprenant des composés particuliers a également permis l’optimisation de nouvelles méthodes rapides et robustes sans réaliser aucune expérience supplémentaire.

A terme, cela a permis l’application de ces méthodes analytiques pour analyser de nombreux composés, tout en simplifiant considérablement l’utilisation de ces méthodes pour leur quantification. Le transfert de méthode vers des équipements plus performants a également été abordé et largement simplifié grâce à la robustesse prédite et observée lors de l’étape d’optimisation basée sur l’analyse du risque.

Parallèlement, la problématique de l’automatisation du développement des méthodes séparatives a pu être envisagée grâce à l’utilisation de l’Analyse en Composantes Indépendantes, combinées à divers algorithmes de partitionnement de données.

Par ailleurs, le calcul du Design Space a pu être appliqué à un procédé pilote de fabrication de poudre pharmaceutique au moyen d’un équipement de laboratoire. Ceci a permis d’exprimer des garanties sur la qualité future du produit fini.

Finalement, les modèles prédictifs ont également été utilisés dans le contexte de la validation de méthodes analytiques et bio-analytiques quantitatives.

Ce travail est le fruit d'une collaboration étroite que j'ai eu la chance d'avoir avec de nombreuses personnes. Leur clairvoyance et leur soutien ont été maintes fois décisifs pour aboutir à l'écriture de ce manuscrit.

Mes premiers remerciements s'adressent tout d'abord aux Professeurs Bruno Boulanger et Philippe Hubert, co-instigateurs et véritables moteurs de ce projet passionnant. Je les remercie également pour m'avoir fourni un cadre de travail accueillant et agréable au sein du laboratoire de Chimie Analytique et au sein de la société Arlenda. J'éprouve une profonde reconnaissance pour l'opportunité d'avoir pu travailler avec eux. Leurs idées novatrices, leurs avis sur des sujets pointus et leur ouverture d'esprit m'ont permis d'apprendre tous les jours de nombreuses choses, tant au niveau de la recherche, qu'au niveau humain.

Il y a six ans de cela, ce projet est né d'un partenariat entre la société Eli Lilly and Company et le laboratoire de Chimie Analytique du département de Pharmacie de l'université de Liège. Merci au Docteur Attilio Ceccato et à Monsieur Gabriel Caliaro d'avoir également imaginé ce projet et d'avoir toujours cru en ses chances d'aboutissements.

Un des buts ultimes de ce projet était de réaliser une recherche combinant des aspects novateurs de la statistique et de la chimie analytique. Cela s'est traduit en débutant simultanément deux projets de recherche. Je désire remercier tout particulièrement le Docteur Benjamin Debrus pour l'étroite collaboration qui, je l'espère, nous a permis de démontrer que l'innovation survient bien souvent à la croisée de deux Sciences. Son travail passionné m'a permis de découvrir les diverses facettes de la chimie analytique, qui m'étaient alors inconnues. Il me serait difficile de lui témoigner toute ma gratitude en si peu de mots.

Je remercie également Bernadette Govaerts et Philippe Lambert de m'avoir fait profiter de leurs connaissances dans la recherche en statistiques. Bien souvent, leur grande expérience a été décisive pour me permettre d'avancer dans les étapes successives de cette recherche.

Je désire également faire parvenir mes chaleureux remerciements aux membres du service de Chimie Analytique avec qui j'ai pu collaborer sur de nombreux projets ou simplement passer d'agréables moments de bonne humeur: Mesdames Amandine Dispas, Iolanda Nistor et Sabah Houari; les Docteurs Roland Marini, Eric Rozet, Eric Ziemons et Jérome Mantanus; et Messieurs Frédéric Lecompte, Cédric Hubert, Sébastien Lhoest et Jérémie Mbinze Kindenge. Je garderai toujours un excellent souvenir de ces années de recherche.

Mes remerciements s'adressent également à la société Arlenda, pour avoir contribué au financement ce projet de recherche. Je remercie particulièrement Messieurs François Moonen, Didier Vandermaesen et Mark Denham pour leur aide administrative et technique. La région Wallonne et l'université de Liège sont également remerciés pour le soutien financier de cette recherche.

I would like to thank Gentiane Hasbroeck, Paul Eilers and Adelin Albert, members of the thesis committee, for the time devoted to the follow-up of this research and for the administrative management with the university. I also would like to thanks the jury to spend the time to read and judge this manuscript.

Il m'est difficile de remercier personnellement tous mes amis, rencontrés lors de mes études d'informatique et de statistiques, ou au sein de la 42SV unité scoute de Louvain-la-Neuve, pour leur soutien inconditionnel et les excellents moments passés ensembles. Que tous ceux qui, de prêt ou de loin, se reconnaissent dans ces quelques lignes soient sincèrement remerciés.

Finalement, je tiens à remercier ma famille, mes parents et mes frères pour leur soutien et leur patience durant ces années de recherche. Particulièrement, ma compagne Marie et notre fille Anaëlle ont toujours été du plus grand réconfort. Je les remercie pour ces années d'amour et de bonheur, en attendant les suivantes avec impatience.

Contents

Introduction

During the development and manufacturing of drug or the quality control of pharmaceutical formulations registered on the market, analytical methods play a prominent role. To understand the context of the present work, it is worth to give some hints about the drug development process and the drug analysis context.

Drug discovery and clinical phases

During the drug development process, analytical and bio-analytical methods are used intensively to obtain a thorough knowledge about the development and follow up of the involved molecules, vaccines, antibodies, genes, stem cells, hormones, etc.

The aims are multiple and include the understanding of the biological mechanisms in action during a treatment, the analysis of the effect of the dosing of the drug on its efficacy and its safety, and the ability to prove that a drug is compliant with the regulatory texts given by authorities. Indeed, regulatory bodies such as the US Food and Drug Administration (FDA), the European Medicines Agency (EMA), the Ministry of Health, Labour and Welfare (in Japan), etc. strongly regulates the development of drugs. The drug development process is schematized on Figure 1, which illustrates the different phases and time needed to bring a new drug on the market.

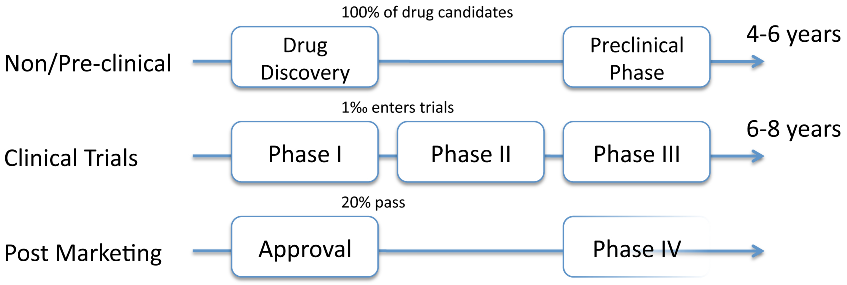


Figure 1: Drug development process. (Chang, 2011; DiMasi et al., 2003).

The results generated by analytical methods are directly used to make the critical decisions during all the phases of the drug development process (drug discovery, preclinical phase, clinical trials phases, etc.), or to provide useful working material in this direction. For instance, the kind of information and decision based on analytical methods are:

- the conformity of the drug, i.e. the precise dosage of its active pharmaceutical ingredients (API), as well as the determination of impurities,
- the establishment of the biodisponibility and of the pharmacokynetic (PK) and pharmacodynamic (PD) parameters and models, by dosing the drug in biological samples
- some clinical laboratory measures to determine safety during preclinical and clinical developments,
- the optimization of a dose and the concentration-effect modeling, usually using a biomarker to be quantified,
- the optimization of the dosage form,
- the release of a production batch,
- the stability studies to determine shelf life and the determination of potentially toxic impurities,
- the optimization and quality control (QC) of a manufacturing process based on quantitative and qualitative measurements of the outputs,
- etc.

The use of these analytical results can lead to the premature ending of preclinical or clinical studies, due to safety or lack of efficacy issues, or due to unstable production processes. On the other side, it could lead to continue a trial whereas hidden problems could threaten the safety of the subjects. Considering the time and the costs needed for the drug development process, the experimenters want the involved analytical methods to be as reliable as possible to avoid the risks to take wrong decisions. From the many drugs candidates that can be identified and/or created by research and development (R&D) laboratories, only a very small proportion will succeed to pass the successive and laborious steps (1/5000, Figure 1) allowing them to be commercial drugs. More information about drug development phases can be found in the references. See ICH E8 Expert Working Group (1998); Food and Drugs Administration (2006); Pocock (2004); Chang (2011).

Drug analysis

Analytical methods are also developed to control drugs on the market. Manufacturers or governments often asks independent laboratories to analyze drugs. The aim is generally to obtain independent results in order to assess the drug quality. In some cases, the analytical method to carry out the analysis is known (e.g. the pharmacopeia or international publications) or provided by the manufacturer. In other cases, no analytical method exists and the development of precise and rapid methods is mandatory. This second case is mainly explored in this manuscript.

A particular case of drug analysis concerns the fake detection in which more and more laboratories are involved. Counterfeit and fake drugs have adverse consequences for public health (Panusa et al., 2007). The World Health Organization (WHO) reported 6% of drugs worldwide are counterfeit and the Food and Drug Administration (FDA, USA) estimated this proportion to be 10% (Mazières, 2007). This proportion varies from one country to another. In some African countries, Marini et al. (2010a) confirmed that up to 80% of medical products are poor quality medicines. To ensure the quality of drugs and to help battle fake and counterfeit medicines, the development of analytical methods, that can simultaneously trace many of the most commonly used molecules in various therapeutic classes, is an important effort.

Different forms of frauds are observed. One of the most frequent is the low dosage of the drug's API compared to the nominal amount stated by the manufacturer. Of course, the drug might simply not contain the API that treats the disease. Another form of fraud occurs when a malicious manufacturer provides a fake drug in the box of another manufacturer. This is referred as counterfeit. Often, even the visual aspect of the counterfeit drug looks similar to the real drug, and precise analytical methods remain the only tools able to try to fight against these illegalities.

Logically, it is expected that the malicious manufacturer is selling the drugs without complete analysis to detect non-compliances such as impurities, with the objective to drastically reduce costs. Unfortunately, the worldwide market of fraudulent medicines is one of the most profitable, with lower risks for the counterfeiters than other drug crimes (Marini et al., 2010a).

A last type of fraud is related to poor storage conditions, such as incorrect temperature or direct sunlight on drugs. This might also cause potentially dangerous impurities to appear, even if the drug shelf life is respected.

To protect the consumer and to provide effective treatment against widespread diseases, the need for accurate method to analyze concurrently many compounds is high. During this work, innovative screening and quantitative methods for the analysis of various anti-paludic drugs (Debrus et al., 2011c) and non-steroidal anti-inflammatory drugs (NSAIDS) (Krier et al., 2011; Mbinze Kindenge et al., 2011) have been developed. Some results are presented in Chapters 7 and 8. The proposed analytical methods are illustrated with some excursions in drug quality, that may potentially help to combat fraud and counterfeit.

A last field where analytical methods are decisive is the analysis of forbidden products such as narcotics and doping product in biological samples. See for instance the joint works with De Backer et al. (2009) and Debrus et al. (2011a).

Analytical methods

In order to browse a majority of the existing analytical methods, a small survey is shown in Figure 2. The data represents the proportion of talks and posters concerning specific analytical methods, that have been presented during a recent and important symposium (Drug Analysis, 2010). It illustrates, among others, the prominence of the *separation techniques*, including the liquid chromatography (LC), the capillary electrophoresis (CE), and the gaz chromatography (GC). The proportion of the communications given during all the symposium is 50% for the LC methods only. This proportion is also representative of the place of the LC in the pharmaceutical industry.

The separation techniques are generally coupled with performant detectors such as ultra-violet diode-array detector (UV-DAD) or mass spectroscopy detector (MS).

Other methods such as the spectroscopy (including near infrared, infrared and Raman spectroscopy), or the nuclear magnetic resonance (NMR) are also wide areas of research and this survey should only be taken as a quick snapshot of these rapidly evolving fields.

Concerning separation techniques, they generally share the same objectives : to be able to separate the relevant compounds of various matrices in the shortest analysis time. These matrices includes pharmaceutical formulations, plant extracts, plasma samples, etc. This separation allows further quantitative and qualitative analysis. The next section briefly introduces the liquid chromatography separation technique.

High performance liquid chromatography

The birth of chromatography is attributed to the first works of Tswett (1906). Subsequent development of techniques by Martin and Synge (1941) during the 1940s and 1950s allowed the definitions of the basic principles of partition chromatography. This work led them to receive the Nobel prize in Chemistry as it permitted the

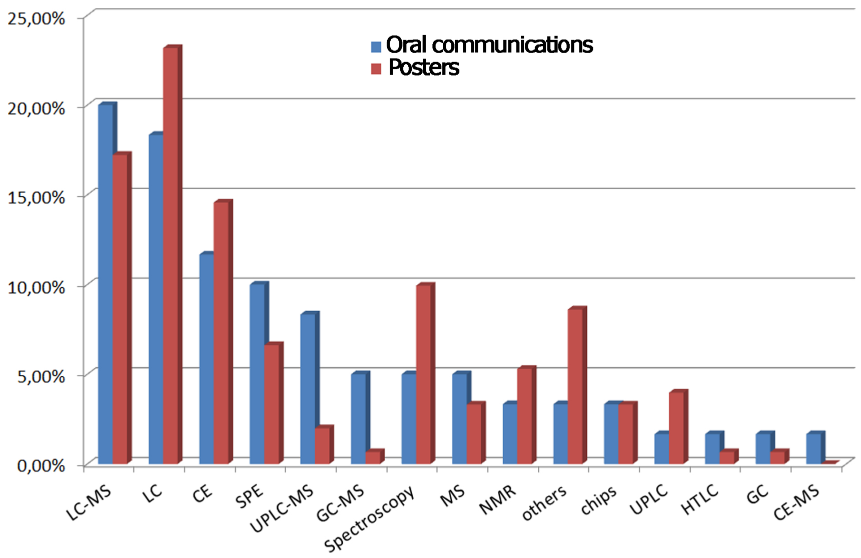


Figure 2: Proportion of talks and posters for several analytical techniques, as observed at the International Symposium of Drug Analyis (2010). (Courtesy of B. Debrus)

rapid development of several separation techniques, including the high performance liquid chromatography (HPLC) developed in the 1960s. Since these inventions, impressive technological advances have been performed. For instance, the ultra HPLC (UHPLC) is able to perform analysis in a very short run time while gaining in efficiency.

Nowadays, one of the most common chromatography technique is the HPLC, schematically presented on Figure 3. Within the device, the sample (or mixture) is driven by a liquid *mobile phase* (solvents and buffer) through a *stationary phase* (analytical column), and the physico-chemical interactions between both phases and the sample allow or not the separation of the compounds of the sample. This separation can be observed thanks to a detector that generates and records a plot called a chromatogram. Each peak within the chromatogram corresponds to a detected compound during the analysis time. The ability to detect low-concentrated or low-detectable compounds from background noise is referred as the *sensitivity*. The terms *high performance* were added to differentiate from older devices that were not able to deliver high pressure in a constant flow rate.

The affinities of the compounds to remain in the mobile or stationary phases make a separation possible. Indeed, given the nature of both phases, certain compounds will move through the stationary phase quicker or slower, before reaching the detector. With regards to this separation, the *selectivity* factor describes the ability of the separation of the compounds by the mobile and stationary phases. This selectivity provides the *specificity* of the method, which is its capability to deliver the observed

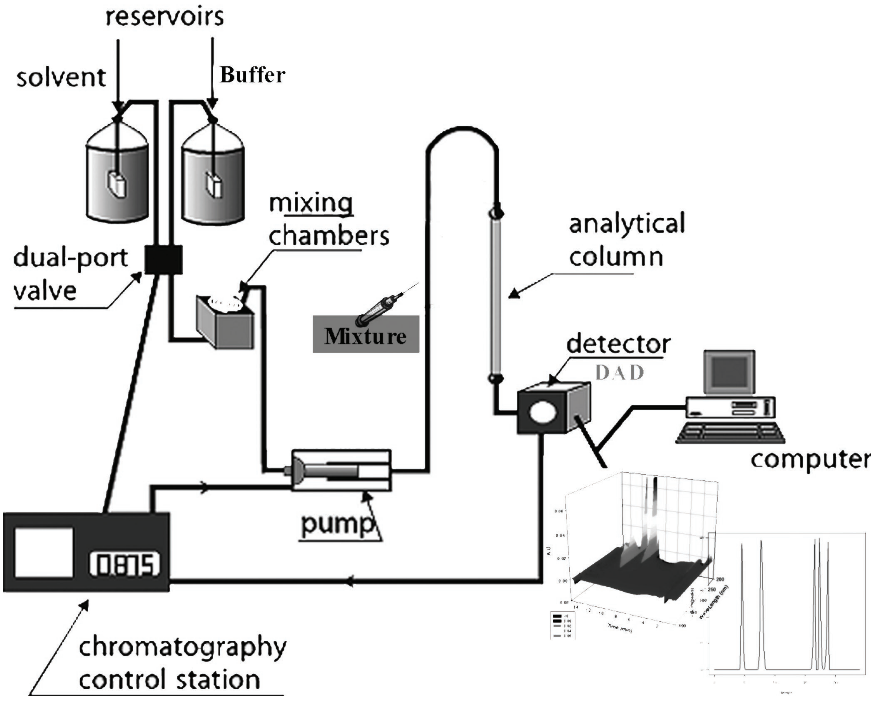


Figure 3: Schematic view of a HPLC system.

signals of the compounds with no interference of other compounds or artifacts.

The nature of the stationary phase (or analytical column) is generally chosen by the analyst. It is a complex task regarding the number of existing columns, that have different physico-chemical ways of action. For certain new and experimental columns, there is a limited knowledge about the compounds-columns interactions. However, information on the compounds and use of standard or generic columns can help this arduous process. Of course, the nature of a stationary phase can be included in a designed experiment to select an optimal one during a screening process.

Several factors interacting with the column are known to have significant impact on the separation: the temperature, the pressure given by the pump, etc. Furthermore, the nature of the mobile phase (the solvents and buffer) provides a very flexible way to improve separation, and many different *operating conditions* can be set. For instance, the pH, the percentage of the organic modifier (that may be changed during run time), the nature of the organic modifier, the nature of acid, the flow rate, etc., can generally be considered as having an influence on the separation and on the total run time of the method. More details about HPLC and similar analytical methods can be found in Snyder et al. (1997, 2010); Meyer (2004).

As previously mentioned, the output of the HPLC is named a chromatogram, which is a plot recorded by the detector during the run time, as shown on Figure 4. A bad results occurs when peaks are overlapping, or in chromatographic terms, when they are *coeluted* (peaks 1-4 and 5-6 on Figure 4, A). As stated previously, a quality output will show well separated peak in a reasonable analysis time (Figure 4, B).

Specific problems such as the method calibration for quantitative measurements, the inverse prediction and the validation of the analytical methods, are presented in the application sections (Part II).

While it is observed that the diversity and the quality of analytical methods have evolved exponentially allowing substantial gains in selectivity, sensitivity, and repeatability of the results, there is still a lack for a rationale towards the development of robust separation methods in a systematic way.

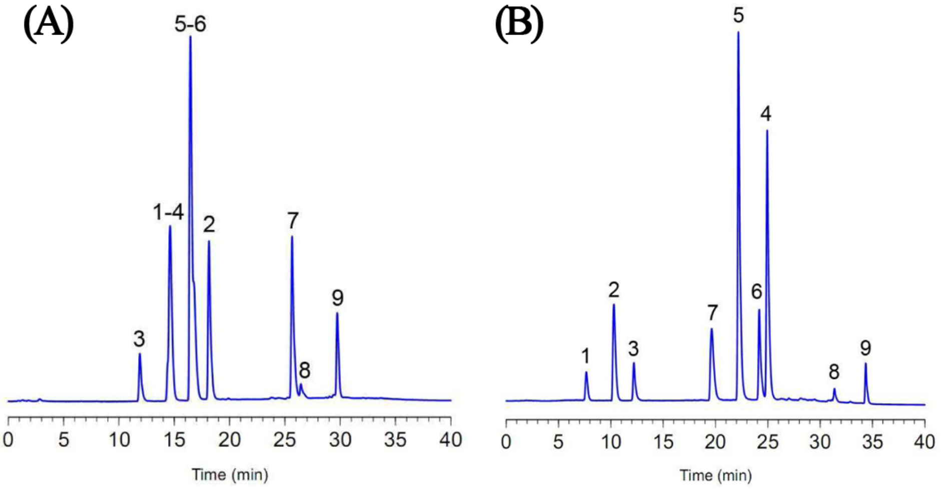


Figure 4: Example of HPLC outputs. (A) Typical results of a non optimal method with coeluting peaks. (B) Example of chromatograms showing nice properties of separation.

Design of experiments

Given that number of possible combinations of operating conditions is considerably high, it is natural to envisage design of experiments (DoE) when optimizing an analytical separation method. However, despite the fact that DoE methodologies exist and are successfully applied in many fields since more than 40 years, their generalization in pharmaceutical sciences, and especially in analytical chemistry, is far from being achieved.

The reasons are linked to the complexity of the processes under investigation. Indeed, the multivariate nature of experiments and responses that are encountered leads to complex designs and statistical models, slowing down the adoption of any structured statistical methodology. Also, the strong believe that years of practical experience can not be "replaced" by mathematics is still very common in pharmaceutical laboratories. Finally, many analysts see the statistical methodologies as black boxes and are not comfortable with the interpretation of the results.

In this text, statistical methodologies are proposed, extending more conventionnal DoE methodologies. They are not envisaged to "replace" the analytical skills and expertise, but rather to help and enhance the understanding and mastering of the processes and methods involved.

Notice that several softwares already exist for the optimization of HPLC conditions. The most known are Drylab,[1] Chromsword,[2] ACD/LC simulator[3], Osiris,[4] and Fusion AE.[5] These softwares are based on solvophobic and linear solvent strength theories (Horváth et al., 1976; Snyder et al., 1979, 1989; Molnar, 2002; Snyder et al., 2010) and generally allow for drastically reduced number of experiments to perform. However, they all have two major limitations. First, they only allow the optimization of two or three HPLC factors simultaneously. These factors are often the most prominent to achieve a separation but they are chosen by the softwares developers. Unfortunately, flexibility is not left to the analysts to choose other relevant factors. This can be problematic in many specific cases. The second reason is more critical as these softwares generally rely on mean response optimization, which will be explained as non sufficient in the next chapter. As the uncertainty of prediction is not taken into account, dissimilarities between the softwares' predictions and the experimental observations do occur, without clear explanation nor understanding. Notice however that some software developers tend to give more importance to uncertainty, as it is the case with Fusion AE which now relies on some Monte-Carlo simulations to assess the robustness.

Quality by Design

Improving and providing guarantees about the overall quality of a process is at the heart of Quality by Design (QbD). QbD has been first applied in processes within the industries such as the automotive industry, see for instance Deming (1986); Juran (1992). The aim was to improve the quality of the products by understanding as best as possible the products and the processes setup to build it. If quality controls

[1] http://www.molnar-institute.com

[2] http://www.chromsword.de

[3] http://acdlabs.eu/products/com_iden/meth_dev/lc_sim

[4] http://www.datalys.net/index.htm

[5] http://www.smatrix.com

(QC) remain mandatory, the improvements provided by QbD should lead to the production of less "*bad* products" since the mechanisms leading to this quality are known.

In the pharmaceutical industry, the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), that regroups experts from the pharmaceutical field and regulatory authorities from Europe, Japan and United States, has clearly understood the gains of such approaches when applied to pharmaceutical processes, in which analytical methods are obviously involved. The ICH releases regulatory documents that are adopted by countries as laws or guidelines. The work of ICH is classified in four topics: the quality (Q), the safety (S), the efficiency (E), and some multidisciplinary (M) topics.

In the present work, the focus is put on the first topic, the quality, with the objective to facilitate the analytical method development in a QbD environment. It will be shown how the ICH requirements can be applied for several applications involving analytical methods.

In particular, the document Q8 presents the concept of **Design Space** (DS) as a tool to achieve QbD. Quoting the website of ICH (2010) "*[...] the document Q8 shows how concepts and tools (e.g., design space) [...] could be put into practice by the applicant for all dosage forms. Where a company chooses to apply quality by design and quality risk management (Q9), linked to an appropriate pharmaceutical quality system, then opportunities arise to enhance science- and risk-based regulatory approaches (Q10)*". This underlines how the topics Q8, Q9 and Q10 – development, risk and quality – are written to provide ways to improve knowledge. These documents have been subject to questions and answers and the US Food and Drug Administration (2011) has now provided the procedures to effectively apply the guidances. This indicates it is a growing and important field within the pharmaceutical companies. However, given the broad nature of their applications, ICH Q8, Q9 and Q10 do not propose any detailed solution to apply QbD. Instead, it gives clearly the principles with some questionable examples of the proposed tools based on mean responses. Thus, there is a clear need from the industries to have methodologies to answer the overall QbD problematic.

In summary, *Quality by Design* should allow to design products and processes to ensure they meet specific objectives, in contrast to the classical *Quality by Testing*, where some batches or runs are tested to observe their performance *post-hoc.* Following the Food and Drugs Administration (2009), QbD is two-fold:

1. it must increase the scientific understanding of products and processes,
2. it must include a risk-based assessment of the quality of products and processes.

Design Space

ICH Q8 (2009) introduces the DS 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*". In other words, QbD for ICH aims at understanding and gaining knowledge about a process or a method to find a parametric region of reliable robustness for its future performance.

Figure 5 illustrates the DS concept. The DS is the region formed by the operating conditions $\mathbf{x}$, where the corresponding predicted outputs $\boldsymbol{Y} = f(\mathbf{x})$ will show acceptable quality level. The operating conditions can obviously be explored with a designed experiments strategy. The concept of acceptance on the outputs of the process is defined with specifications, namely, some acceptance limits (λ's) that are used to give some values of minimal acceptable quality. The critical step is to quantify the assurance of this quality, i.e. the guarantee to have outputs within specifications, or the risk to be out of specifications. This is the risk-based assessment as defined by Food and Drugs Administration (2009).

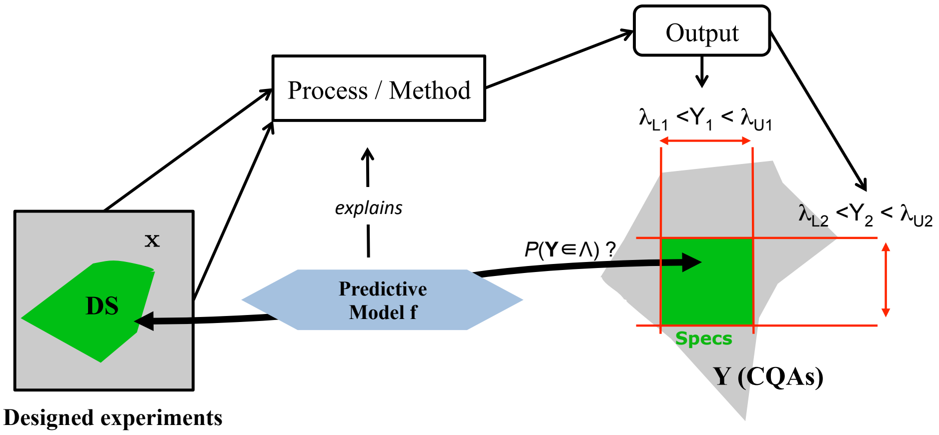

Output
Process / Method
$\lambda_{L1} < Y_1 < \lambda_{U1}$
explains
$\lambda_{L2} < Y_2 < \lambda_{U2}$
x
$P(\mathbf{Y} \in \Lambda)$?
DS
Predictive Model f
Specs
Y (CQAs)
Designed experiments

Figure 5: Illustration of the Design Space (DS) as the region of the operating conditions $\mathbf{x}$ for which there is guarantee that the related responses $\boldsymbol{Y} = f(\mathbf{x})$ are within acceptance limits (in red).

The guideline states that "*Working within the Design Space is not considered as a change. Movement out of the Design Space is considered to be a change and would normally initiate a regulatory post approval change process. (...)*". An interpretation of this is the following : if the DS is large with respect to operating conditions, the solution or the process can be considered as robust. This makes sense in the context of analytical method development, where robustness is mandatory since the variability of the inputs and of the analytical process can be important for a given time period. Robustness is also desired to ease the transfer of the method from an R&D department to a manufacturing site.

The overall aim is to provide "*assurance of quality*" to the management or to regulatory authorities. Regarding manufacturing, the producer must be confident that high quality products will, in the future, be obtained by reliable processes, i.e. that are designed proactively to become reliable. Regarding quantitative methods, the analyst want to be confident in the results that will be provided to take important decisions.

The "*quality*" is expressed in terms of quality criteria or responses (namely, the critical quality attributes, CQAs) that must lie within acceptance limits that are specified *a priori* and that reflect the minimal desired settings (e.g. yield $> 95\%$ and processing time ≤ 5 min.). Those specifications are, or at least should, be driven by the customer of the product or the results. As an example, safety considerations from phase I clinical studies should be used to define specifications for the maximum tolerated dose, i.e. the maximum limit to the quantity of an active pharmaceutical ingredient (API) that can be found in a drug. The processes must be precise enough to ensure this specification is never exceeded, and the analytical methods must be accurate enough in order to detect a poorly manufactured drug.

The concept of "*assurance*" refers to the ways to provide guarantees that, given uncertainty of processes and measurements, it is very likely that the quality will be met in the future as long as specific operating conditions are satisfied. In that perspective, to give the necessary assurance of quality, the manufacturer may try to demonstrate that tight operating conditions are fully under control, e.g. using Statistical Process Control (SPC), i.e., Quality by Testing. Unfortunately, the practicalities make this option difficult and expensive to achieve. This approach also fails to give indications about the optimality or robustness of the process.

Another option is the one dictated by ICH Q8: the manufacturer must show evidence that the quality of the outcome or the product remains within acceptable limits (λ's), even in presence of changes in operating conditions occurring within identified limits (i.e. operating conditions lying within the DS). This second option, referred as Quality by Design, is more efficient because it integrates the unavoidable variability of the process or method, which may more or less impact on the quality. It is then the ideal way to cope with uncertainty and to become less sensitive to normal range of disturbances in the operating conditions. It could also provide a broader set of conditions that allows achieving the intended quality, if robustness is demonstrated. Both producers and customers would benefit from this approach. As previously explained, QC remains mandatory but improved quality would ensure a lower rejection rate.

Objectives

It has been shown that several guidelines explain the ways to apply Quality by Design in estimating the guarantee to achieve the objectives of processes or analytical methods. However, when this work has begun no practical methodology yet existed with the ability to provide an answer for this problematic within the pharmaceutical industry. **The general objective is then to translate the requirements of the regulatory documents into statistical terms and to develop a complete methodology addressing Quality by Design** for analytical methods and processes.

Liquid chromatography is one of the most widespread analytical technique. While the chromatographic technologies have evolved permitting higher quality outputs in terms of selectivity and sensitivity, no generic way exists for the rapid development of complex analytical methods in a QbD environment. One identified reason was the complex nature of multivariate experiments and outputs. Several softwares for chromatographic optimization exist but they have some limitations that does not allow them to be as generic, flexible and QbD compliant, as needed. **The first objective is then to define a complete methodology, specific for the development of analytical separation methods**, compliant with the Quality by Design movement and allowing the achievement of separation for complex mixtures of possibly unknown compounds.

To provide guarantee of quality, this methodology must take into account the error from the models, the processes their measurements. One way to achieve this is to work with statistical models accounting for the multivariate error of prediction of every future result. Parallel to the development of the methodology, **the second objective consists in building predictive statistical models and to review the multi-criteria optimization strategies** accordingly.

To obtain a rapid and robust framework for method development, automation is added to the list of requirements. **The third objective is to use multivariate deconvolution techniques and automated identification of compounds to automatically read the analytical method outputs**, which is a time consuming and non-error-free manual process. This should allow for a fully **automated development of analytical methods**, from the very optimization to the method validation and routine analysis.

The three previous objectives are intended to show that the development of rapid and robust analytical methods is a reality for the pharmaceutical industry and for the laboratories involved in the drug analysis.

A fourth and final objective is to transpose and apply the developed

methodology to small-scale manufacturing processes and other steps of the analytical method life cycle, such as the validation of the quantitative results. The idea is then to use the developed statistical models and optimization strategies in different contexts to illustrate their applicability and confirm their generic nature.

Structure of the manuscript

Part I

Chapter 1 underlines the statistical needs for the application and the computation of Design Space, based on the ICH guidelines. Particularly, the objective is to point out several concerns arising when using classical tools for prediction. In this chapter, it is shown that neglecting the uncertainty and dependencies during the predictions is an obstacle to the QbD application.

These statistical needs are mainly centered about the predictions of future runs of the process or the methods. Chapter 2 generically introduces the Bayesian framework to derive the distribution of such predictions. This framework is then used to obtain the predictive distribution of the standard multivariate regression (Chapter 3), that is flexible enough to model the behavior of the liquid chromatographic methods. An application of this model to a small-scale manufacturing process is also envisaged in Part II. This multivariate model is helpful to optimize the quality of the outputs and to assess the uncertainty of the predictions. Next, Chapter 4 presents a Markov chain Monte Carlo (MCMC) sampling perspective to obtain the predictive distributions of several univariate mixed models including the random ANOVA, the hierarchical linear regression and the mixed-effects four-parameters logistic (4PL) regression with a heteroskedastic variance proportional to a power of the mean. These models are used for the validation of analytical methods and for the calibration and validation of analytical and bio-analytical methods, using the uncertainty of the back-calculated prediction.

Chapter 5 presents the concepts of multi-criteria optimization (MCO) and multi-criteria decision methods (MCDM) in the predictive context. This is useful when several CQAs must be optimized jointly in presence of competing responses. It is shown how a risk-based approach can be combined with MCDM, using joint posterior probabilities of acceptance or predictive desirability methodologies.

Finally, the Chapter 6 introduces the independent component analysis (ICA), that separates a matrix-like signal into a sum of independent sources. In many circumstances, the ICA can be used to help and to fasten the analysis of UV-DAD chromatograms. The objective of the proposed methodology is to automatically find and identify the peaks, and possibly to remove much of the noise and disturbances that are often observed. Innovative criteria are developed to assess if a source is relevant or not. A last step is the identification and clustering of the relevant sources ("which chromatographic peak corresponds to which compound?"), on the basis of their reconstructed spectral signature.

Part II

The second part of the manuscript illustrates the use of the methodologies with real laboratory applications. All examples except one are based on real data.

Chapter 7 introduces the development of a HPLC method to screen anti-paludic drugs in the QbD context while explaining the different steps of the methodology. Chapter 8 illustrates the power of a unique design of experiments to optimize several methods aiming at screening and quantifying an exhaustive list of non steroidal anti-inflammatory drugs (NSAIDS). The predicted robustness of the developed methods allowed for a transfer to a high-end equipment and eased the validation of quantitative methods. These excursions in drugs quality show the potentialities of the developed analytical methods to identify substandard drugs and potentially to combat fake and poor quality medicines.

Next, the transposition of the methodologies is made from analytical methods to a small-scale manufacturing process. An application of QbD and DS computation is shown for the optimization of a spray-dryer in Chapter 9. It allowed quantifying the guarantees about the quality of the future products manufactured by the optimized process.

The last chapters discuss the analysis and validation of quantitative methods using predictive methodologies. First, Chapter 10 explains how the ICA can be used to fully automate the treatment of chromatograms in order to obtain data for method validation. Automatic computations to derive calibration curves and validation results are also carried out. To assess the performance of ICA, the data have been created in such way it is nearly impossible to obtain any relevant results manually. Finally, Chapter 11 treats the problem of bio-analytical methods such as ligand-binding assays, where calibration curves are non-linear. It allows providing Critical Quality Attributes about the assay future performance.

Part I

Theory

Chapter 1

General context

1.1 Quality by Design

The concept of *Quality by Design* (QbD) is currently one of the most recurrent idea in the pharmaceutical literature. It mainly touches the drug discovery, method development and production areas. However, this concept is not new.

It has been emphasized by Deming (1986) and Juran (1992) in the end of the 80s, but the roots are older. For instance, Juran (1951) had worked on quality management from the beginning of the 50s. His idea was that the quality of a product can be rigorously planned in the ways this product is built. If each production processes can demonstrate great quality achievements, most problems related to the manufacturing of the product could be avoided. A common successful application of this idea was done with the automotive industry in Japan. It is well recognized that the overall quality of the production chains has allowed Japanese industries to manufacture low-cost and high quality cars. The same applies for high technologies such as the design and miniaturization of microprocessors, cellphones, etc. Due to its success, the QbD concept has naturally propagated around the world and the various industries.

More recently, the Food and Drug Administration (FDA) and the International Conference for Harmonization (ICH) have seen the opportunity to apply QbD to gain knowledge and understanding about the products and processes of the pharmaceutical industry (see the following regulatory document and guidelines: Food and Drugs Administration, 2007, 2009; ICH Q8, 2009; ICH Q9, 2005; ICH Q10, 2008). These guidelines rely on the use of information and prior knowledges gathered during pharmaceutical development studies to provide a scientific rationale about the manufacturing process of a product (Yu, 2008).

A risk-based approach. QbD emphases on the products (API and excipients), on the process understanding and on the process control. It advocates a systematic approach for their development based on clear and predefined objectives. Particularly, the *risks* that a product or a process will not fulfill the quality requirements must be assessed. Regulatory actions now begin to go in this risk-based direction and the pharmaceutical industries will have to be fully compliant with these guidelines in order to prove they are able to provide high quality medicines. See for instance the recent manual of policies and procedures (MaPP) published by the Food and Drug Administration (2011). Of course, the interest of the industries is high as QbD could finally allow more success in quality control (QC). Indeed, as the minimal required quality is proved during all the production steps, the QCs will be easier to fulfill, although still mandatory. Less products will be rejected, which is a noticeable gain for both the consumers and the producers, in terms of risk management (ICH Q9, 2005).

1.2 Method development

Let envisage how QbD can be implemented in order to achieve the quality for a general method to be developed. The reader will notice that the very general following explanations apply for analytical methods as well as any process that can be encountered in production, and also in other areas than the pharmaceutical field.

Figure 1.1 describes a generic scheme for a method or a process. It represents the knowledge available even if any experiment is carried out, and without any mathematical and statistical consideration. Generally, an operating condition defined by various input variables $\mathbf{x} = (x_1, x_2, x_3, ...)$ must be assigned to Critical Process Parameters (CPP) for the process to run. CPPs are the factors whose variation has an impact on the output. They should then be monitored or controlled to ensure that the process or method runs appropriately.

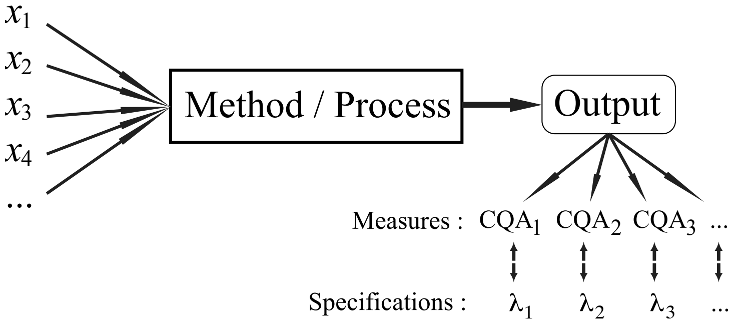

x_1
x_2
x_3
x_4
...
Method / Process
Output
Measures : CQA$_1$ CQA$_2$ CQA$_3$ ...
Specifications : λ_1 λ_2 λ_3 ...

Figure 1.1: Schematic representation of an analytical method or a process.

What matters is the output of the method, as it is used to make important decisions, such as the release of a production batch, a stability study, the optimization of the dose of a drug based on some biomarker information, etc. Thus, the analytical method must provide quality outputs. Otherwise, the risks to take a wrong decision will be high.

An output is analyzed through its Critical Quality Attributes (CQAs). According to ICH Q8, "*a CQA is a physical, chemical, biological, or microbiological property or characteristic that should be within an appropriate limit, range, or distribution to ensure the desired product quality*". A CQA can thus be the output of the method, or any properties that can be observed on this output. For a chromatographic separation method, the output is a chromatogram, and the CQAs can be the separation or/and the resolution between two or all peaks, or the total run time of the method. For a quantitative method, this can be its dosing range, its accuracy within this dosing range, etc. For a process, it can be the yield, the solidity, the taste of the product, etc. Clearly, CQAs must be measurable quantities. What is important is to define and choose the CQAs in accordance to the decisions that will be taken.

The ICH Q8 definition focuses also on *appropriate limits*. Equivalents of "appropriate limits" are "acceptance limits" or simply 'specifications" (the λ's, on Figure 1.1). These specifications are set up by the domain experts in order to define specifically what is a (minimal) satisfactory quality. For instance, in the chromatographic world, a separation of at least 0 min between two peaks means they are non coeluting (overlapping). For a quantitative method, one may want to have a dosing range of at least 0.1-500 mg/mL for a compound of interest, with a lower limit of detection of 0.05 mg/mL. For a process, it can be mandatory to have a yield higher than 95%, to be profitable enough. Usually, those specification should largely be driven by customer needs, or derived from customers requirements. Some could also be defined for efficacy purpose or good science practice.

Difficulties are sometimes encountered about the definition of these specifications. In most cases however, the inability to properly define these limits is a consequence that the method or process is not well understood. In this case it is clearly harder to give guarantees and risks about the quality of the outputs. By experience, little discussions are mandatory but generally sufficient to have a clearer view.

1.3 Design of experiments

Classically, the CQAs of a method are (or can be) optimized. The setting of the operating conditions is chosen so that the method has the best CQAs, possibly jointly, using multi-criteria approaches. Statistical tools such as Design of Experiments (DoE), including screening and response surface models, and multi-criteria decision methods (MCDM) are possible choices to achieve this objectives (Cox and Cochran, 1957; Harrington, 1965; Montgomery, 2009). The advantages of DoE methodology against other optimization strategies (e.g. *one-factor-at-a-time* optimization) are well known. Briefly, DoE is a tool to obtain the maximum information with the minimum number of experiments. It also allows providing knowledge about a whole well-defined multivariate experimental domain for the operating conditions.

Let us define the very general model $(Y_j \mid \mathbf{x}) = f_j(\mathbf{x}; \boldsymbol{\theta}_j) + \varepsilon_j$, for each of the m pertinent responses Y_j, with $j = 1, ..., m$. The measured responses can be some descriptions of the output or directly the CQAs. They are observed at different operating conditions $\mathbf{x}$ belonging to an experimental domain denoted χ. A common assumption is that the error follows a Normal distribution, $\varepsilon_i j \sim N(0, \sigma_j^2)$.

The responses must be well chosen to allow good model fitting properties. In parallel, the identification of the most relevant model(s) to be used can be done. These most relevant models f_j represent the assumed but unknown links between the CPPs and the responses. Finally, if the responses are not directly the CQAs, they must allow the computation of the CQAs.

The parameters $\boldsymbol{\theta}_j$ and σ_j^2 are unknown and are estimated using the data gathered through experiments using e.g. the ordinary least-square, maximum likelihood or maximum a posteriori estimate. The point estimate value of the parameters is noted $\hat{\boldsymbol{\theta}}_j$ and the mean predicted responses are obtained on every points $\mathbf{x}$ using these estimates: $\hat{y}_{j|\mathbf{x}} = \hat{E}[Y_j \mid \mathbf{x}] = f_j(\mathbf{x}; \hat{\boldsymbol{\theta}}_j)$. The plan defined by $\hat{y}_{j|\mathbf{x}}$, $\forall \mathbf{x} \in \chi$, is called a response surface.

1.3.1 Mean responses optimization

Let assume the univariate response Y_j ought to be optimized (e.g. maximized) over χ, as it is one CQA of interest. Classically, this optimization is carried out using the mean predicted response (i.e. the response surface) at new points $\tilde{\mathbf{x}}$ included in the experimental domain. One will look after the operating condition $\mathbf{x}^*$ such that

$$\mathbf{x}^* = \underset{\tilde{\mathbf{x}} \in \chi}{\operatorname{argmax}}\, \hat{E}[Y_j \mid \tilde{\mathbf{x}}] = \underset{\tilde{\mathbf{x}} \in \chi}{\operatorname{argmax}}\, f_j(\tilde{\mathbf{x}}; \hat{\boldsymbol{\theta}}_j).$$

Graphically, the results of such optimization process can be nicely represented if the dimensionality is moderated, as illustrated on Figure 1.2. Notice the data used to create all the graphs hereafter are artificial.

The majority of statistical packages are able to provide such optimization results. However, there is no clue that using the operating condition $\mathbf{x}^*$ will produce a satisfactory output. Indeed, at optimum, the result will be on average better than

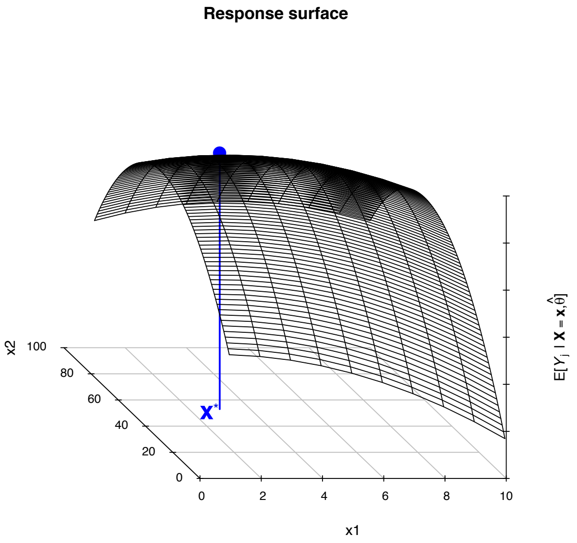


Figure 1.2: Illustration of a response surface model with an optimal solution $\mathbf{x}^*$.

with any other operating condition (in the explored domain χ). But this does not imply a high quality. Also, if the result is subject to noise, caution should be taken regarding the optimum. The analysis of the model parameters and classical tools such as residuals analysis, Q-Q plots, etc. associated with statistics such as R^2 or RMSEP, etc. may prevent the analyst to use the (mean) results provided by the model.

When envisaging multiple response optimization, the desirability methodology are appealing to aggregate the various (mean) responses into one index representing the quality of the solution, as proposed by Harrington (1965); Derringer and Suich (1980). More details on this subject are given in Chapter 5.

1.3.2 Sweet Spot

A better answer to assess the quality is to define some specification(s), say $\mathbf{\Lambda} = (\Lambda_1, ..., \Lambda_m)$, for the m responses or CQAs envisaged. In this context, not only the optimal solution is of interest, but instead the set of operating conditions giving outputs with mean CQAs $\hat{E}[Y_j \mid \tilde{\mathbf{x}}]$ within the specifications Λ_j. This is the concept of Sweet Spot (Anderson and Whitcomb, 1998; Peterson and Lief, 2010). Formally, for an univariate problem with one CQA Y_j, it is defined as:

$$\begin{aligned} \text{Sweet Spot} &= \left\{ \tilde{\mathbf{x}} \in \chi \;\middle|\; \hat{E}[Y_j \mid \tilde{\mathbf{x}}] \in \Lambda_j \right\}, \\ &= \left\{ \tilde{\mathbf{x}} \in \chi \;\middle|\; f_j(\tilde{\mathbf{x}}; \hat{\boldsymbol{\theta}}_j) \in \Lambda_j \right\}. \end{aligned} \tag{1.1}$$

It is possible to represent graphically the sweet spot when dimensionality is limited, as shown in Figure 1.3. In this example, the assumed specification Λ_j is to obtain $Y_j \geq \lambda$, i.e. the j^{th} CQA must be higher than a specified value. The zone of the experimental domain where the mean response surface is higher than λ defines the Sweet Spot (in red). The interpretation is as follows: in this zone, the response is, on average, within specifications. Of course, an *optimal* solution may still have sense to define exactly the operating condition to use.

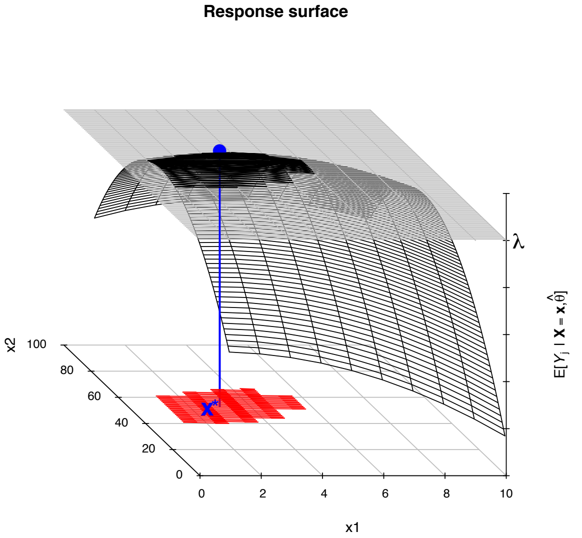


Figure 1.3: Illustration of a univariate Sweet Spot. In red is the subpart of χ providing an expected CQA better than λ (dashed plan).

When considering a multivariate responses process, Sweet Spot methodology is generally used on the overlapping mean response surfaces. The idea is then to find a subpart of χ where every mean response is located within its specifications. This can be written as follows:

$$\begin{aligned} \text{Sweet Spot} &= \left\{ \tilde{\mathbf{x}} \in \chi \;\middle|\; \hat{E}[\boldsymbol{Y} \mid \tilde{\mathbf{x}}] \in \boldsymbol{\Lambda} \right\} \qquad (1.2) \\ &= \left\{ \tilde{\mathbf{x}} \in \chi \;\middle|\; \hat{y}_{1|\mathbf{x}} \in \Lambda_1 \bigcap \ldots \bigcap \hat{y}_{m|\mathbf{x}} \in \Lambda_m \right\}, \end{aligned}$$

where $\hat{E}[\boldsymbol{Y} \mid \tilde{\mathbf{x}}]$ is the multivariate mean response surface and $\boldsymbol{\Lambda}$ is the set of specifications for the responses.

The sad news about Sweet Spot is that the provided solution gives limited guarantees that the quality will be observed in the future use of the method, under the presence of inevitable uncertainties (Peterson, 2009; Peterson and Lief, 2010; Schofield, 2010). Indeed, in the Sweet Spot, considering a symmetric distribution for Y_j, and ε_j being an additive error, the mere interpretation is that there is at least 50% of chance to observe the univariate response within specification, that is, $P(Y_j \geq \lambda_j \mid \tilde{\mathbf{x}}, \hat{\boldsymbol{\theta}}_j) \geq 0.5$. At the border of the Sweet Spot, there is one chance out of 2 to be out of specification when dealing with an univariate problem, if the model can be assumed correct.

Finally, correlation structures that might exist between the simultaneous responses are completely ignored, which further increases the risk to take wrong decision. As an example, considering two responses Y_1 and Y_2, *assumed independent* and predicted at an operating conditions $\tilde{\mathbf{x}}$, situated at the border of each univariate Sweet Spot, such that $P(Y_1 \geq \lambda_1 \mid \tilde{\mathbf{x}}, \hat{\boldsymbol{\theta}}) = 0.5$ and $P(Y_2 \geq \lambda_2 \mid \tilde{\mathbf{x}}, \hat{\boldsymbol{\theta}}) = 0.5$. The only guarantee about the joint acceptance of both specifications is then, using the rule of product for independent variables, $P(Y_1 \geq \lambda_1, Y_2 \geq \lambda_2 \mid \tilde{\mathbf{x}}, \hat{\boldsymbol{\theta}}) = 0.5^2 = 0.25$. In other words, there is $1 - 0.25 = 0.75$ probability that the actual outputs are not within both specifications, although the solution lies within the Sweet Spot. If correlations were present and taken into account, this result would be different.

Because the responses/CQAs dependencies are not taken into account, the use of (overlapping) mean responses and Sweet Spot may certainly give unexpected and unexplained results for the future use of the method or process.[1]

1.4 Design Space

As explained in the previous sections, the basic use of DoE for optimization is generally not sufficient to achieve the risk-based perspective advocated for the QbD approach. The issue comes from the use of mean responses (CQAs) derived from the statistical model. Hereafter are summarized the flaws of mean responses.

Mean responses does not provide sufficient clues about the method reliability. Assuming a statistical model is appropriate to describe data, the only interpretation of an univariate mean response is that the results are at least as good as the predicted one in about 50% of the runs (i.e., on average). Conversely, one can expect 50% of future results to be not that good ! The obvious solution is to work on

[1] Notice the examples in the appendix of ICH Q8 (2009) are all about Sweet Spot, which is believed non compliant with a fully QbD strategy. However, the document states that proven (univariate) acceptable ranges continue to be acceptable from the regulatory perspective but are not considered as a design space (Section 2.4.5 and Question and Answers B.1.Q8.)

the complete distribution of the responses instead of point estimates. Uncertainty and dependencies between responses are the keys to assess reliability. The previous section has also explained the problem when comparing mean responses to minimal specifications.

Mean responses give limited information on the future performance. The purpose of optimizing or validating a method is to give evidences that it will perform appropriately in its future use, i.e. most of the time, the outcome will meet quality criteria. The use of prediction intervals can be thought as a nice tool for this. Indeed, intervals are practical to express the uncertainty of the responses in a comprehensive way. But unfortunately, they do not *quantify* the guarantees or risks to be within or outside specifications, respectively. Intervals are then less appropriate when envisaging a risk-based approach, even if they integrate the various sources of uncertainty.

1.4.1 Definition

To explain the QbD practice, ICH Q8 guideline defines the important concept of Design Space (DS), central in this manuscript. The DS is "*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 objective of ICH Q8 is to improve the way to understand and gain knowledge about a process or method to find a parametric region of robustness for future performance of this process or method, in order to provide the guarantees of quality. To compute these guarantees, the idea is to replace the expectations in Equations (1.2) with a probability measure to obtain the results within specifications (Chiao and Hamada, 2001). Mathematically, a simplistic but pragmatic DS definition is as follows, for a univariate or a multivariate response:

$$\begin{aligned} \text{Design Space} &= \left\{ \tilde{\mathbf{x}} \in \chi \mid P(Y_j \in \Lambda_j \mid \tilde{\mathbf{x}}, \hat{\boldsymbol{\theta}}) \geq \pi_j,\ j = 1, ..., m \right\}, \\ &= \left\{ \tilde{\mathbf{x}} \in \chi \mid P(\boldsymbol{Y} \in \boldsymbol{\Lambda} \ \mid \tilde{\mathbf{x}}, \hat{\boldsymbol{\theta}}) \geq \pi \right\}. \end{aligned} \tag{1.3}$$

The main differences between Equations (1.2) and (1.3) is that the latter is about an acceptance probability that is compared to a minimal quality level π_j (marginally) or π (jointly). In the frequentist statistical framework, a common solution to obtain this probability estimate is to use the assumed distribution of the error ε_j in repeated sampling, with the parameters assumed known (Normal distribution) or estimated using the available data (Student's distribution). Obviously, the joint distribution of the responses must be considered when envisaging the computation of joint probabilities.

Figure 1.4 illustrates the results of such computations. First, Subfigure A depicts an hypothetical univariate distribution of a response Y_j for a given $\mathbf{x}^*$, assumed

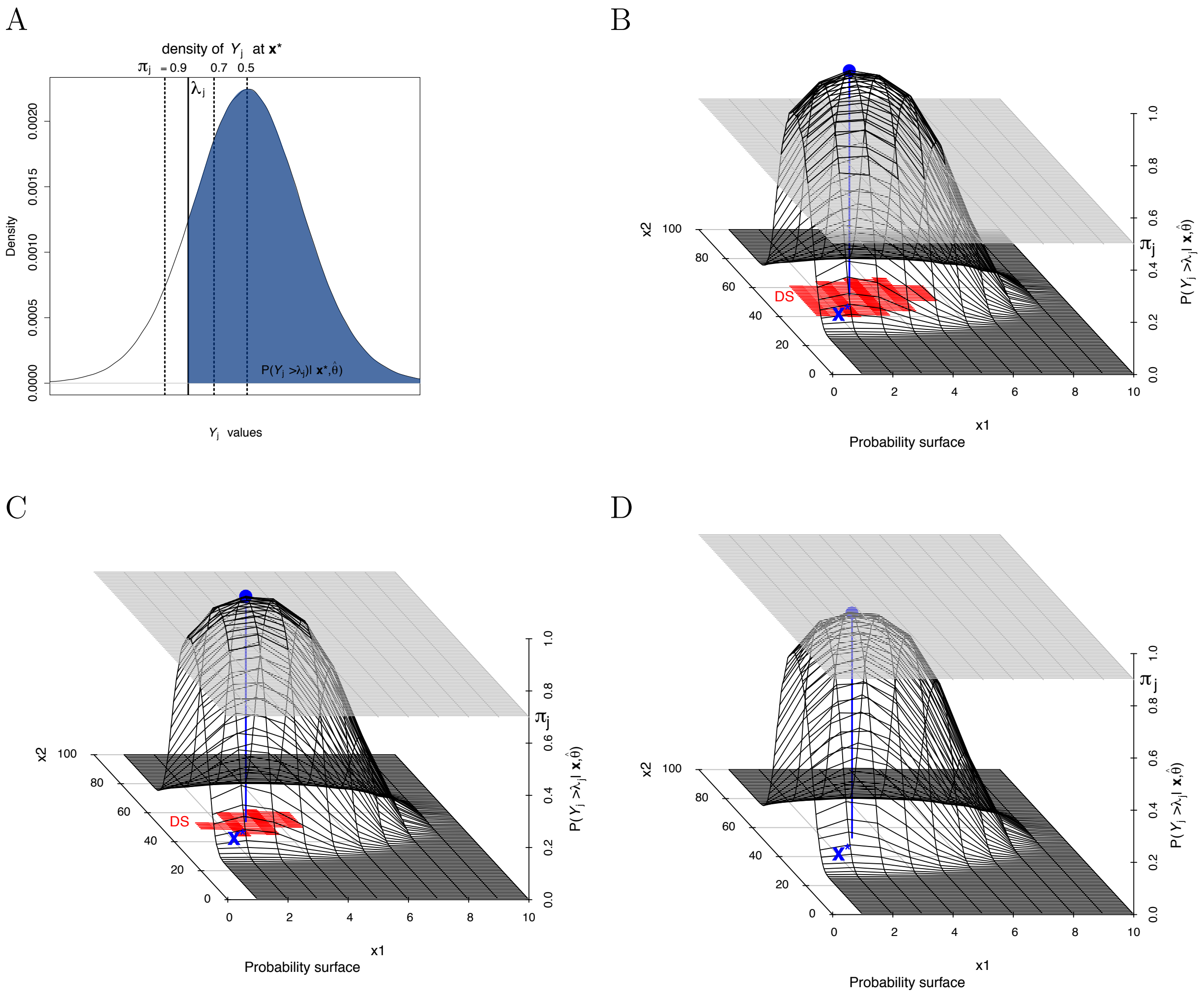


Figure 1.4: (A) Density at optimal condition. Shaded zone: Probability that $Y_j \geq \lambda_j$. (B-C-D) Surface of the probability that $Y_j \geq \lambda_j$ for every operating conditions, and Design Space (in red) at various quality levels. (B) Quality level of $\pi_j = 50\%$. Similar results than Sweet Spot. (C) Quality level of $\pi_j = 70\%$. (D) Quality level of $\pi_j = 90\%$, no Design Space is identified.

optimal. On this basis, the proportion of this distribution that is higher than a value λ_j can be computed (blue). It expresses the guarantee (i.e. the probability) to observe $\boldsymbol{Y} \in \boldsymbol{\Lambda}$. In this example, let's assume we are interested in $Y_j \geq \lambda_j$.

Repeating the operation on every point $\tilde{\mathbf{x}} \in \chi$, it is then possible to draw a map of these probabilities, as illustrated on the Subfigures B,C and D. The difference between these three images is the level of π_j. On Figure 1.4 (B), the minimal quality level π_j is set to 50% (shaded grey plan), and the operating conditions satisfying this quality level are said to belong to the Design Space (red) with $\pi_j = 50\%$. As it could be expected with a symmetrical distribution, choosing π_j =50% leads to the same set of operating conditions as the Sweet Spot (see Figure 1.3). However, when the desired quality level is higher (i.e. lower risks), such as $\pi_j = 70\%$ (C), the DS is becoming smaller and may potentially not exist anymore, as in Figure 1.4 (D) that illustrates $\pi_j = 90\%$. Indeed, in this example, the probability $P(Y_j \geq \lambda_j \mid \mathbf{x}^*, \hat{\boldsymbol{\theta}})$ at optimal condition $\mathbf{x}^*$ is merely 0.86, so it is not possible to achieve 90% quality level.

As with the Sweet Spot, an optimum still makes sense. Thus, one will look after the operating condition $\mathbf{x}^*$ such that

$$\mathbf{x}^* = \underset{\tilde{\mathbf{x}} \in \chi}{\operatorname{argmax}}\, P(Y_j \geq \lambda \mid \tilde{\mathbf{x}}, \hat{\boldsymbol{\theta}}).$$

If the DS is able to provide a set of satisfying operating conditions, the optimum still defines the best quality within χ, which a customer or a producer will definitely benefits. If high quality level is observed, this optimum could become the future working operating condition.

Figure 1.5 summarizes the concept of DS. This flowchart can be compared to Figure 1.1. The use of an acceptance probability is the major advance of DS against the classical use of Sweet Spot that is based on mean responses. As the joint distribution of responses is used, this method relies on the data covariance structure to improve results. However, to obtain results that are valid for the prediction of individual future outcomes, the DS definition must finally be based on a *predictive* distribution, as presented in the next subsection. To do so, the idea is to take into account the uncertainty of the model parameters instead of their point estimates (Peterson, 2004).

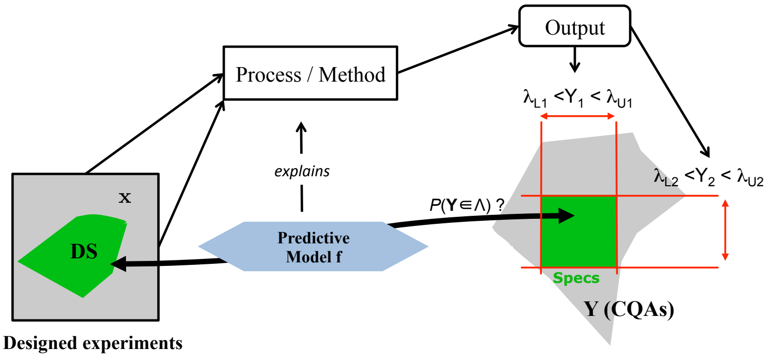

Output
Process / Method
$\lambda_{L1} < Y_1 < \lambda_{U1}$
explains
x
$\lambda_{L2} < Y_2 < \lambda_{U2}$
$P(\mathbf{Y} \in \Lambda)$?
DS
Predictive Model f
Specs
Y (CQAs)
Designed experiments

Figure 1.5: Flow chart representing the concept of Design Space.

1.4.2 Risk-based predictive approach - a Bayesian choice

From the previous explanations, it is enviable to obtain a general solution for multivariate problems, with a risk-based emphasis that is based on a predictive distribution of new outcome. In this context, the Bayesian framework is well adapted to obtain it. Among other qualities, the Bayesian analysis allows obtaining the predictive distribution for any problem, by means of simulations or analytical derivation. In the Bayesian predictive context, Equation 1.3 can be restated as (Peterson, 2004):

$$\text{Design Space} = \{\tilde{\mathbf{x}} \in \chi \mid P(\boldsymbol{Y} \in \boldsymbol{\Lambda} \mid \tilde{\mathbf{x}}, \text{data}) \geq \pi\}\,. \tag{1.4}$$

The parameters and CQAs uncertainties and their dependencies are fully taken into account in the predictive distribution of $\boldsymbol{Y}$, allowing computing the (posterior) predictive probability to observe $\boldsymbol{Y}$ within specifications $\boldsymbol{\Lambda}$. In this sense, even a poor model is able to provide some results: a large model uncertainty will result in a wide predictive distribution. This however still allows deriving an expression of the guarantees of future quality, or the risks not to meet the specifications, without relying on mean and univariate responses, and without ignoring the dependencies.

An analysis of the model performance using the predictive distribution is finally possible (Gelman et al., 2004). This can lead to the identification of problems such as poor sample size or noisy process or parameters. This may give hints on the direction to tackle the problems.

The Bayesian predictive distributions is described for several types of statistical models in the Chapters 2, 3 and 4. The Chapter 5 introduces the multi-criteria decision methodology using such predictive distribution. Applications are then given in the Part II.

1.5 Conclusion

DoE has been stressed as an appropriate tool to achieve QbD. However, it is clear that DoE, as classically understood, *is not* inherently QbD. It is used to explore an experimental domain efficiently and to provide some statistics of interest. One important step in the QbD initiative is to understand the impact of the parameters (CPP) on the outputs quality (CQA). DoE and classical statistics are an empirical ways to gather this understanding.

However, when it comes to prediction using statistical models, handling the uncertainty is necessary to appropriately assess the risk not observing the outputs within well-defined specifications. In summary, a QbD approach should answer the following questions:

- Are there parameters with strong or poor influence on the CQAs?
- What are the sources of uncertainties? Is there dependencies between the parameters and/or between CQAs?

- Which parameter settings will provide satisfactory outputs? What are the guarantees and risks?
- What are the possible improvements?

A predictive approach was shown to be the way to obtain acceptable process or method parameter ranges, i.e. a risk-based Design Space, in which there are guarantees that it will perform appropriately.

If no Design Space is found, two options are possible. Either the statistical models are not able to explain the behavior of the process precisely; or the process or method is not able to provide quality with high assurance. Both options will result in a large predictive uncertainty that can however help understanding why one run of the process or method does differ from others. This can give insights on the reasons of poor quality and be the starting point for improvements.

Chapter 2

Bayesian methodologies

From the publication of the initial formulation of the Bayes' theorem by Thomas Bayes (1763), there have been various opinions regarding its use. However, it has now become clear that the underlying statistical methodologies are more than a simple alternative to classical (frequentist) statistics.

These lasts decades, some factors has led to the increased application of Bayesian statistics. Probably the most important is the discovery of the Markov-chain Monte-Carlo method (MCMC) that allows overcoming many of mathematical problems by means of computer intensive simulations (Metropolis et al., 1953). The gain in power of computers explains the very recent explosion of Bayesian statistics, as the time needed to get results is now generally reasonable, even with complex models.

In the context of prediction, frequentist statistics suffers from various issues, as pointed out by Aitchison (1964) : "*in the theory of statistical tolerance regions, as usually presented in Frequentist terms, there are inherent difficulties of formulation, development and interpretation*". These issues are essentially related to predictions. The core of the problem lies in the definition of the probability in frequentist statistics that is the frequency of an event. When departing from asymptotic or other assumptions (e.g. Normality, independence of errors), the frequentist solutions often become rather intricate. This definition of a probability as a frequency of observed event is also clearly not adapted to some problems, when there is no data available yet today (e.g. stock exchanges, etc.). In Bayesian statistics, the probability is seen as a degree of plausibility that an event will occur. It provides a general framework for every statistical problem, using the Bayes' theorem as a central point.

The ability of Bayesian statistics to propose a general framework to obtain the (posterior) predictive distribution of the data is very practical as this distribution contains the necessary information about the future new data and their uncertainty

(Guttman, 1988).

The first sections of this chapter describes the basis of Bayesian analysis in very general terms. Chapters 3 and 4 illustrate some applications of Bayesian methodologies to get a predictive distribution for the examples that are used in this manuscript: the standard multivariate regression, the one-way random ANOVA, the hierarchical linear regression and the mixed-effect non-linear regression with model for the variance.

2.1 Bayes' theorem

In the Bayesian framework, the quantities of interest are assumed to be random values that follow a probability distribution. If $\mathbf{y} = (y_1, ..., y_n)$ is the vector of n observations of the random value Y, that depends upon some parameters $\boldsymbol{\theta} = (\theta_1, ..., \theta_p)$, then it has a probability density function $p(\mathbf{y} \mid \boldsymbol{\theta}, I)$. I represents any pertinent information included, translated into assumptions about the distribution. For simplicity, I might be ignored in the notations to shorten equations.

The uncertainty about the unknown parameters $\boldsymbol{\theta}$ is also expressed using a distribution $p(\boldsymbol{\theta} \mid I)$. Applying the rule of products:

$$p(\mathbf{y} \mid \boldsymbol{\theta}) \ . \ p(\boldsymbol{\theta}) = p(\mathbf{y}, \boldsymbol{\theta}) = p(\boldsymbol{\theta} \mid \mathbf{y}) \ . \ p(\mathbf{y}). \tag{2.1}$$

Isolating the distribution of $\boldsymbol{\theta}$ conditional to the data $\mathbf{y}$, the following result is obtained:

$$\begin{aligned} p(\boldsymbol{\theta} \mid \mathbf{y}) &= \frac{p(\mathbf{y} \mid \boldsymbol{\theta}) \ \ p(\boldsymbol{\theta})}{p(\mathbf{y})} \\ &= \frac{p(\mathbf{y} \mid \boldsymbol{\theta}) \ \ p(\boldsymbol{\theta})}{\int_{\boldsymbol{\theta}} p(\mathbf{y} \mid \boldsymbol{\theta}) \ p(\boldsymbol{\theta}) \ d\boldsymbol{\theta}} \\ \text{Posterior} &= \frac{\text{Likelihood} \ \times \ \text{Prior}}{\text{Marginal likelihood}}. \end{aligned} \tag{2.2}$$

This is the well-known *Bayes' theorem.* $p(\boldsymbol{\theta} \mid \mathbf{y})$ is called the *posterior density* of $\boldsymbol{\theta}$, expressing how $\boldsymbol{\theta}$ is distributed given the data. It is often written $p(\boldsymbol{\theta} \mid \text{data})$. $p(\boldsymbol{\theta})$ is referred as its *prior density.* It expresses what is known about $\boldsymbol{\theta}$ before any look on the data.

$p(\mathbf{y} \mid \boldsymbol{\theta})$ is the likelihood function and is also written $\mathcal{L}(\boldsymbol{\theta} \mid \mathbf{y})$. $p(\mathbf{y})$ is the marginal likelihood and is a normalizing constant depending only on the data. Its main utility is to ensure the posterior density integrates to 1. $p(\mathbf{y})$ being a constant, Equation (2.2) is often simplified into

$$\begin{aligned} p(\boldsymbol{\theta} \mid \mathbf{y}) &\propto \mathcal{L}(\boldsymbol{\theta} \mid \mathbf{y}) \ . \ p(\boldsymbol{\theta}) \\ \text{Posterior} &\propto \text{Likelihood } \times \text{ Prior}, \end{aligned} \tag{2.3}$$

where $\propto$ stands for "equals up to a constant". The posterior density of $\boldsymbol{\theta}$ is the prior knowledge of $\boldsymbol{\theta}$ that is updated through the likelihood function.

The prior distribution of $\boldsymbol{\theta}$ can be the expression of ignorance about $\boldsymbol{\theta}$. In this case, a *non-informative* prior distribution, or vague prior distribution, is used. This distribution is generally very flat over the domain of $\boldsymbol{\theta}$ (notice that giving a vague *a priori* is considered as a pertinent information I about $\boldsymbol{\theta}$). On the opposite, if previous experiments give clues about $\boldsymbol{\theta}$, or if general knowledge about the application domain provides useful information (i.e. location, spread), it can be incorporated in the definition of $p(\boldsymbol{\theta})$ as an informative *a priori*.

The use of a prior distribution for the parameters is what makes the difference with the frequentist approach, and is sometimes viewed as a strong argument against the Bayesian approach, because it is mandatory to set these prior values. This argument can however be reverted as it would be harmful not incorporating any useful and valid information about $\boldsymbol{\theta}$. In addition, when non-informative *a priori* are used, this leads to results similar to frequentist statistics: the posterior distribution has the same form than the likelihood.

The difficulty to set up priors increases with the dimensionality of parameters (Gelman et al., 2008; Kerman, 2011). This is a common problem for the Bayesian analysis that can be addressed through some sensitivity studies of the effect that the prior information has on the posterior distribution.

2.2 Posterior distribution of the parameters

In Bayesian statistics, the Equations (2.2) and (2.3) must be solved. This can be done analytically or using numerical methods. The most desirable situation arises when it is possible to find the analytical form of $p(\boldsymbol{\theta} \mid \mathbf{y})$, and to identify its underlying distribution. This option is realistic when working on simple problems (small dimension, linearity, classical distribution assumptions, simple priors etc.) but often, the full joint posterior distribution of the parameters remains unidentified. In that case, numerical solutions based on sampling methodologies might be envisaged. The reasons why the posterior distribution can remain unidentified are of various natures.

Reason 1. The distributions of the parameters are too different in nature. Regarding the various forms of parameters, it might be simpler to identify the conditional or marginal distribution of each of them, or of several subsets of them. For instance, assuming the joint posterior density of two parameters $p(\boldsymbol{\theta} \mid \mathbf{y}) = p(\theta_1, \theta_2 \mid \mathbf{y})$ is non identifiable as a standard distribution, one may rather try to identify the conditional $p(\theta_1 \mid \theta_2, \mathbf{y})$ and the marginal $p(\theta_2 \mid \mathbf{y})$. Using the rule of product, one still gets the joint posterior density:

$$p(\boldsymbol{\theta} \mid \mathbf{y}) = p(\theta_1, \theta_2 \mid \mathbf{y}) = p(\theta_1 \mid \theta_2, \mathbf{y})\ p(\theta_2 \mid \mathbf{y}). \tag{2.4}$$

Slicing the problem this way can obviously be easier. Regarding $p(\theta_1 \mid \theta_2, \mathbf{y})$ alone, θ_2 is simply considered as a constant. This constant may be dropped out of the computations. For the second term, $p(\theta_2 \mid \mathbf{y})$, θ_1 must be integrated out of the joint posterior density:

$$p(\theta_2 \mid \mathbf{y}) = \int_{\theta_1} p(\theta_1, \theta_2 \mid \mathbf{y}) d\theta_1. \tag{2.5}$$

This is feasible on various class of densities. Of course, it is possible to reverse the problem to search for the identification of $p(\theta_2 \mid \theta_1, \mathbf{y})$ and $p(\theta_1 \mid \mathbf{y})$, that also fully describes the joint posterior density by rule of product.

Reason 2. The joint posterior density may not be integrable w.r.t. one, some or all the parameters. The solution proposed in the previous paragraph is then not usable. However, if a full set of conditional posterior distributions can be identified (in the previous example, $p(\theta_2 \mid \theta_1, \mathbf{y})$ and $p(\theta_1 \mid \theta_2, \mathbf{y})$), Gibbs sampling algorithm may be used to draw samples from the joint posterior (Geman and Geman, 1984; Gelfand and Smith, 1990). This is presented in Appendix C.4.

Reason 3. Sometimes, very little can be done to identify the conditional or marginal distributions of the parameters. An elegant solution to obtain samples that follows any (joint posterior) density is the Markov-chains Monte-Carlo (MCMC) sampling algorithm due to Metropolis et al. (1953). It works with a random walk that runs on the unidentified joint posterior density of the parameters. Basis of MCMC sampling algorithms are presented in Appendix C.2 and C.3.

2.2.1 Credible interval and region

When the posterior distribution of parameters has been identified or when samples have been drawn from the posterior distribution, credible intervals or regions can be derived. Inspecting the (marginal) posterior distribution of an univariate parameter θ, any interval $[\delta_1, \delta_2]$ that contains $\beta.(100)\%$ of the distribution define a credible interval at level β (Edwards et al., 1963). For two-sided intervals, it is common to look for the shortest interval $[\delta_1, \delta_2]$. Under the assumption that θ is not uniform, the interval can be formally defined as

$$\delta_1, \delta_2 \text{ such as } \int_{\delta_1}^{\delta_2} p(\theta \mid \mathbf{y}) d\theta = \beta, \text{ with } p(\delta_1 \mid \mathbf{y}) = p(\delta_2 \mid \mathbf{y}) \tag{2.6}$$

It is also named the highest posterior density (HPD) interval (Box and Tiao, 1973). It is the bayesian equivalent of the classical confidence interval, but the interpretation is fairly different. A $\beta.(100)\%$ credible interval for the random variable θ means that the probability that θ lies in the interval $[\delta_1, \delta_2]$ is β. On the opposite, a frequentist $\beta.(100)\%$ confidence interval $[\delta_1, \delta_2]$ about an unknown but fixed θ means that with a large number of repeated sampling from the original population, $\beta.(100)\%$ of the calculated confidence intervals would include the true value of the parameter (De Gryze et al., 2007). In this case β is a level of confidence and not a coverage nor a probability.

In most cases, the credible and confidence intervals are different. The common frequentist assumption that error parameters have a value assumed known, leads to a different ways to manage the uncertainty, directly impacting the interval. The incorporation of prior information can also lead to obvious differences.

Notice that for a multivariate density, the credible interval is named the credible region.

2.3 Predictive distribution of new responses

Prediction in a risk-based environment is a major statistical problem since all the uncertainties that may impact on the outcome must be included to assess what will happen in the future. One of the strength of the Bayesian paradigm is to be able to provide the predictive distribution of some random values of interest, in many situations. This predictive distribution is at the heart of the Design Space definition.

The density of a future observation $\tilde{y}$ of Y is denoted $p(\tilde{y} \mid \mathbf{y})$. Its predictive density is formally defined as

$$\begin{aligned} p(\tilde{y} \mid \mathbf{y}) &= \int_{\boldsymbol{\theta}} p(\tilde{y}, \boldsymbol{\theta} \mid \mathbf{y})\ d\boldsymbol{\theta}, \\ &= \int_{\boldsymbol{\theta}} p(\tilde{y} \mid \mathbf{y}, \boldsymbol{\theta})\ p(\boldsymbol{\theta} \mid \mathbf{y})\ d\boldsymbol{\theta}. \end{aligned}$$

Assuming $\tilde{y}$ and $\mathbf{y}$ are conditionally independent given $\boldsymbol{\theta}$, the predictive density can be written as

$$p(\tilde{y} \mid \mathbf{y}) = \int_{\boldsymbol{\theta}} p(\tilde{y} \mid \boldsymbol{\theta})\ p(\boldsymbol{\theta} \mid \mathbf{y})\ d\boldsymbol{\theta}, \tag{2.7}$$

where $p(\tilde{y} \mid \boldsymbol{\theta})$ is given by the model for given values of the parameters and $p(\boldsymbol{\theta} \mid \mathbf{y})$ is the posterior density of these parameters given the data.

Different options are available to solve Equation (2.7). The ideal case is to carry out the integral and identify a known distribution for the prediction of $\tilde{y}$. When this is possible, inference about the future behavior of $\tilde{y}$ is made simple.

However, this integral might not be analytically tractable. In this case, Monte-Carlo simulations are a way to propagate the posterior uncertainty of the parameters to the model responses. To obtain samples from the predictive distribution, Equation (2.7) suggests to do the following:

- draw $\boldsymbol{\theta}^{(s)}$ from the joint posterior density $p(\boldsymbol{\theta} \mid \mathbf{y})$,
- draw $\tilde{y}^{(s)}$ from the model $p(\tilde{y} \mid \boldsymbol{\theta}^{(s)})$,

with $s = 1, ..., n^*$ is the number of samples to draw.

Apart from its direct implication in Design Space computations, the predictive density can be used to check the model quality, as suggested firstly by Rubin (1981, 1984) and discussed by Gelman et al. (2004). Basically, a good model should obviously have a good fit that can be assessed through the use of classical displays of the observed and mean predicted data in graphs (fit, residuals, etc.). Furthermore, the predictive uncertainty should be as limited as possible. The data should naturally be compliant with the predictive distribution (i.e. the model), and conversely.

2.3.1 Predictive interval and region

Similarly to the credible interval, the predictive interval is defined as the shortest interval $[\delta_1, \delta_2]$ that contains $\beta(100)\%$ of the predictive density. It identifies the values of $\tilde{y}$ that have the highest (predictive) density support. They can formally be defined as

$$\delta_1, \delta_2 \text{ such as } \int_{\delta_1}^{\delta_2} p(\tilde{y} \mid \mathbf{y}) d\tilde{y} = \beta, \text{ with } p(\delta_1 \mid \mathbf{y}) = p(\delta_2 \mid \mathbf{y}) \tag{2.8}$$

The Bayesian predictive interval is not the equivalent of the classical prediction interval, but is directly related to the β-expectation tolerance interval (Guttman, 1988). It is expected that a specified proportion (or coverage) β of $\tilde{y}$ will fall within $[\delta_1, \delta_2]$.

This gives clue that, at least from a simulation point of view, the β-expectation tolerance intervals are simple to obtain in the Bayesian framework, while they are often complicated to derive in the frequentist framework, even for simple models.

For a multivariate predictive density, the predictive interval becomes the predictive region.

2.4 Risk-based approach

This last decade, it has been observed that the use of probabilities (e.g. p-values) has been progressively replaced by the use of statistical intervals (Bland, 2010). They provide a simpler interpretation, with values that are in the domain of the statistical variable of interest. They are particularly informative when analyzing univariate variables. For multivariate variables, they are less practical to describe as the intervals become ellipsoids or more complex regions.

Moreover, intervals or regions are not appropriate to provide risk-based information about the responses or some critical quality attributes (CQAs), i.e. when comparing their distribution with some values of interest. The risk-based approach might be summarized with the two following questions.

1. What is the guarantee to have my CQA(s) within some specification(s) ?
2. What is the risk to have my CQA(s) out of specification(s) ?

A probabilistic approach using the predictive distribution is the answer to these questions. Consider the following example with two specifications $\mathbf{\Lambda} : Y \geq \lambda_1$ and $Y \leq \lambda_2$. It is possible to estimate the guarantee to observe a future realization $\tilde{y}$ of Y within specifications:

$$\begin{aligned} P_{\mathbf{\Lambda}} &= P(Y \in \mathbf{\Lambda} \mid \mathbf{y}) \\ &= P(\lambda_1 \leq Y \leq \lambda_2 \mid \mathbf{y}) \\ &= \int_{\lambda_1}^{\lambda_2} p(\tilde{y} \mid \mathbf{y}) d\tilde{y}. \end{aligned} \tag{2.9}$$

$P_{\mathbf{\Lambda}}$ is an expression of the guarantee of quality, and is then a probability measure to obtain $\tilde{y}$ within specifications. A one-sided computation is achieved defining $\lambda_1 = -\infty$ or $\lambda_2 = +\infty$. The risk to fall outside $\mathbf{\Lambda}$ is expressed as $1 - P_{\mathbf{\Lambda}}$. Equation (2.9) can also be easily extended to deal with multivariate variables or CQAs (see Chapter 5). It is particularly elegant when the joint predictive distribution of the multivariate CQAs is available.

2.5 Conclusion

The Bayesian framework is able to provide the posterior and predictive densities of a (set of) variable(s) of interest. Either these densities can be identified as known distributions, so it is easy to work on; either the densities remain unidentified. In that case, sampling procedures such as Gibbs sampling, Markov-chains Monte-Carlo methods or Monte-Carlo simulations theoretically allow obtaining samples that follow any densities.

The predictive distribution can then be used to find Bayesian predictive intervals, known as β-expectation tolerance intervals in frequentist statistics, through the use of the probability density function or through Monte-Carlo simulations. The predictive distribution can also be used for a risk-based approach, i.e. to provide the guarantee that the variable(s)/CQA(s) will be within some specification(s) in the future.

The following chapters describe the Bayesian methodologies for the standard multivariate regression, the random one-way ANOVA model, the hierarchical simple linear regression and the mixed-effects non-linear regression using a four parameters logistic regression with model for the variance.

Chapter 3

Bayesian standard multivariate regression

This chapter presents the Bayesian solution for the distributions of a multivariate variable and the associated parameters in the linear regression context. Suppose that realizations of the multivariate variable $\mathbf{Y} = (\mathbf{y}_1, ..., \mathbf{y}_j, ..., \mathbf{y}_m)$ have been observed n times. Each $\mathbf{y}_j$ is then a vector of size n containing observations made on a response Y_j when some operating conditions are changed. Each of the n operating conditions is defined by k adjusted critical process parameters (CPP, also referred as factors or input variables) that are common for each response. Changing the values of the CPPs induces changes in the responses.

It is generally assumed that $n \gg k$. A response surface model in the form of a polynomial with p explanatory variables is envisaged for all the responses. Then, $\mathbf{X} = (\mathbf{x}_1, ..., \mathbf{x}_p)$ is the $n \times p$ model matrix that specifies a polynomial model that may be estimated over the k factors ($p \geq k$). The p columns of $\mathbf{X}$ generally consist of a constant term (intercept), main (qualitative and quantitative) factor effects, quadratic or higher order effect terms for quantitative factors and interactions. Because they are defined over the CPPs, these p effects are supposed to describe the main variations observed in the responses. Quantitative factors are usually centered and scaled in the $[-1, 1]$ interval before being included in $\mathbf{X}$, to facilitate the interpretation of the effects. Qualitative factors are coded into dummy variables.

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

$$\begin{aligned}\mathbf{y}_1 &= \mathbf{X}\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}_1,\\ &\vdots\\ \mathbf{y}_j &= \mathbf{X}\boldsymbol{\beta}_j + \boldsymbol{\varepsilon}_j,\\ &\vdots\\ \mathbf{y}_m &= \mathbf{X}\boldsymbol{\beta}_m + \boldsymbol{\varepsilon}_m,\end{aligned} \tag{3.1}$$

where $\boldsymbol{\beta}_j$ are the $(p \times 1)$ model parameters for the j^{th} responses and elements of $\boldsymbol{\varepsilon}_j = (\varepsilon_{1j}, ..., \varepsilon_{ij}, ..., \varepsilon_{nj})'$ are independent and identically distributed (i.i.d.) as a Normal. As the model matrix $\mathbf{X}$ is common for each response, it is possible to write the multivariate model in matrix form:

$$\underset{(n\times m)}{\mathbf{Y}} = \underset{(n\times p)}{\mathbf{X}} \; \underset{(p\times m)}{\mathbf{B}} + \underset{(n\times m)}{\mathbf{E}}, \tag{3.2}$$

where $\mathbf{E}$ is the $(n \times m)$ matrix for the errors, with $\mathbf{0}$-mean and semi-definite positive covariance matrix $\boldsymbol{\Sigma}$. $\mathbf{B}$ is the $(p \times m)$ matrix containing the regression parameters of the multivariate model.

3.1 Likelihood

From Equation 3.2 and assumptions about the error, each of the n observed response vector $\mathbf{y}_i$, of size m, $i = 1, ..., n$, is also i.i.d. and assumed to follow a multivariate Normal distribution given the parameters $\mathbf{B}$ and $\boldsymbol{\Sigma}$. Then,

$$\mathbf{y}_i \sim N_m(\mathbf{x}_i\mathbf{B}, \boldsymbol{\Sigma}), \;\; i = 1, ..., n, \tag{3.3}$$

where $\mathbf{x}_i$ is the line i of $\mathbf{X}$. The joint density of the n vectors of error $\boldsymbol{\varepsilon}_i = \mathbf{y}_i - \mathbf{x}_i\mathbf{B}$ defines the likelihood function and is

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

or, more conveniently,

$$\mathcal{L}(\mathbf{B}, \boldsymbol{\Sigma} \mid \mathbf{Y}) \propto |\boldsymbol{\Sigma}|^{\frac{-n}{2}} . \exp\left(-\frac{1}{2}tr\left[\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{XB})'(\mathbf{Y} - \mathbf{XB})\right]\right) \tag{3.4}$$

The next two sections present the posterior distributions and the predictive distributions of new responses obtained when using different prior distributions. Section 3.2 gives the well known solutions for these distributions when a non-informative prior distribution is chosen for the parameters. In Section 3.3 a solution is proposed when using proper informative and conjugate prior distributions for the parameters.

3.2 Solution with non-informative priors

With non-informative priors, the posterior density of the parameters has been well described by Geisser and Cornfield (1963), Geisser (1965) and Box and Tiao (1973). In the same non-informative context, the predictive distribution of a new (or several) response vector(s) has been described as a multivariate Student distribution by Zellner and Chetty (1965), Johnson (1987) or Press (2003). In the Design Space context, the benefits to use such a predictive distribution have already been shown by Miró-Quesada et al. (2004) and Peterson (2004).

3.2.1 Prior distributions

The following joint non-informative prior distribution has been proposed by Geisser and Cornfield (1963) to express ignorance about the parameters:

$$p\left(\mathbf{B}, \boldsymbol{\Sigma}\right) \propto |\boldsymbol{\Sigma}|^{-\frac{1}{2}(m+1)} . \tag{3.5}$$

Notice this assumes the independence of the parameters $\mathbf{B}$ and $\boldsymbol{\Sigma}$ *a priori*, which has been advocated by Jeffreys (1961) and Savage (1962) when little is known about both parameters. This distribution has the advantage to be invariant under parameter transformation.

3.2.2 Posterior distributions

Combining the prior distribution of Equation (3.5) with the likelihood using Bayes' theorem yields the joint posterior distribution $p\left(\mathbf{B}, \boldsymbol{\Sigma} \mid \text{data}\right) \propto \mathcal{L}\left(\mathbf{B}, \boldsymbol{\Sigma} \mid \mathbf{Y}\right). p\left(\mathbf{B}, \boldsymbol{\Sigma}\right)$. However, the joint posterior density is unpractical to work with so the marginal and conditional distributions of the parameters have been derived as described in the beginning of Section 2.2, page 31.

Conditional distribution of B given Σ

The conditional posterior distribution of $\mathbf{B}$ given $\boldsymbol{\Sigma}$ is expressed as the following matrix-variate Normal distribution (see Appendix D.4 for the distribution definition):

$$\mathbf{B} \mid \boldsymbol{\Sigma}, \text{data} \sim N_{p\times m}\left(\hat{\mathbf{B}}, \boldsymbol{\Sigma}, \left(\mathbf{X}'\mathbf{X}\right)^{-1}\right), \tag{3.6}$$

or, equivalently,

$$vec(\mathbf{B} \mid \boldsymbol{\Sigma}, \text{data}) \sim N_{pm}\left(vec(\hat{\mathbf{B}}), \boldsymbol{\Sigma} \otimes \left(\mathbf{X}'\mathbf{X}\right)^{-1}\right). \tag{3.7}$$

The operator *vec*, applied on a $(p \times m)$ matrix, stacks its columns into a vector of length pm. This is rather useful for implementation purpose as the matrix-variate Normal is usually not available in softwares. The matrix-variate Normal has location parameter $\hat{\mathbf{B}}$ of dimension $(p \times m)$ and covariance matrices $\mathbf{\Sigma}$ and $(\mathbf{X}'\mathbf{X})^{-1}$. $\hat{\mathbf{B}}$ is the least-square estimator of $\mathbf{B}$:

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

Marginal distribution of B

When envisaging linear modeling, it is often interesting to focus on the analysis of the effects of the input variables and their interpretation. For instance, in a screening experimental design, the distribution of the regression parameters is central. Integrating $\mathbf{\Sigma}$ out of the joint posterior, the marginal posterior distribution of $\mathbf{B}$ is given by:

$$\mathbf{B} \mid \text{data} \sim T_{p\times m}\left(\hat{\mathbf{B}}, \mathbf{A}, \left(\mathbf{X}'\mathbf{X}\right)^{-1}, \nu\right), \tag{3.8}$$

i.e. a matrix-variate Student's distribution with location $\hat{\mathbf{B}}$, scale matrices $\mathbf{A}$ and $\left(\mathbf{X}'\mathbf{X}\right)^{-1}$ and $\nu = n - (m+p) + 1$ degrees of freedom (See Appendix D.6). The $\mathbf{A}$ matrix is the $(m \times m)$ symmetric semi positive definite scale matrix defined as $(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})'(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})$. $\mathbf{A}$ is proportional to the estimated sample covariance matrix $\mathbf{\Sigma}$.

Marginal distribution of $\mathbf{\Sigma}$

The marginal posterior distribution of $\mathbf{\Sigma}$ is obtained by integrating $\mathbf{B}$ out of the joint posterior, and is:

$$\mathbf{\Sigma} \mid \text{data} \sim W_1^{-1}(\mathbf{A}, \nu), \quad \nu > 0, \tag{3.9}$$

This is an m-dimensional inverse-Wishart with ν degrees of freedom (see Appendix D.2). Notice that the analytical form of the inverse-Wishart distribution W_1^{-1} used by Geisser (1965) or Box and Tiao (1973) is slightly different than the one that may be found in more recent works or that is usually implemented in softwares such as R, WinBUGS or SAS (Dawid, 1981). The inverse-Wishart distribution W_1^{-1} has, say, ν d.f. $(\nu > 0)$, while the one in R or WinBUGS, noted W_2^{-1}, has an equivalent of $\nu + m - 1$ d.f. with $\nu > m - 1$ (e.g. package `MCMCpack`, Martin et al., 2010). This is of particular importance to compare results with the ones that may be found in the references. In summary,

$$\mathbf{\Sigma} \mid \text{data} \sim W_1^{-1}(\mathbf{A}, \nu) = W_2^{-1}(\mathbf{A}, \nu + m - 1). \tag{3.10}$$

Still for implementation purpose, as the inverse-Wishart distribution is sometimes unavailable, one can use the following equivalence:

$$\mathbf{\Sigma} \mid \text{data} \sim W_1^{-1}(\mathbf{A}, \nu) \;\; \Leftrightarrow \;\; \mathbf{\Sigma}^{-1} \mid \text{data} \sim W_1(\mathbf{A}^{-1}, \nu), \tag{3.11}$$

A more detailed presentation of the Wishart and inverse-Wishart distributions is available in Appendix D.1 and D.2.

It is then possible to draw samples from the joint posterior distribution of the parameters using Equations (3.9), (3.10) or (3.11) followed by (3.6). When analysing the regression effects, Equation (3.8) provides an easier way to obtain samples or statistics from $\mathbf{B} \mid$ data.

3.2.3 Predictive distribution of a new response vector

A new response vector $\tilde{\mathbf{y}}$ at one new point $\tilde{\mathbf{x}}$ included in the experimental domain is obtained from the predictive distribution, defined as,

$$p\left(\tilde{\mathbf{y}} \mid \tilde{\mathbf{x}}, \text{data}\right) = \int_{\mathbf{\Sigma}} \int_{\mathbf{B}} p\left(\tilde{\mathbf{y}} \mid \tilde{\mathbf{x}}, \mathbf{B}, \mathbf{\Sigma}\right).p\left(\mathbf{B}, \mathbf{\Sigma} \mid \text{data}\right).d\mathbf{B}.d\mathbf{\Sigma} \tag{3.12}$$

In the particular case of standard multivariate regression, this can be solved and the predictive distribution of $\tilde{\mathbf{y}} \mid \tilde{\mathbf{x}}$ is identified as a multivariate Student's distribution (Press, 1972; Kibria, 2006):

$$\tilde{\mathbf{y}} \mid \tilde{\mathbf{x}}, \text{data} \sim T_m\left(\tilde{\mathbf{x}}\hat{\mathbf{B}}, \left(1 + \tilde{\mathbf{x}}'(\mathbf{X}'\mathbf{X})^{-1}\tilde{\mathbf{x}}\right)\mathbf{A}, \nu\right), \tag{3.13}$$

where $\left(1 + \tilde{\mathbf{x}}'(\mathbf{X}'\mathbf{X})^{-1}\tilde{\mathbf{x}}\right)\mathbf{A}/\nu$ is the estimated scale or spread matrix of the multivariate distribution, with ν degrees of freedom ($\nu > 0$). For more information about the multivariate Student's distribution, refer to Appendix D.5.

3.3 Solution with informative priors

There is little literature existing on the use of informative prior distribution in multivariate regression problem. Concerning univariate multiple linear regression, Marriott and Spencer (2001) have shown the methodology to derive the posterior distribution of the parameters and the predictive distribution of a new responses. They used conjugate informative priors and illustrated the simplicity of updating prior information. In this section, only the solutions are presented. Detailed derivations can be found in Appendix A.

3.3.1 Prior distributions

For multivariate (multiple linear) regression, a classical conjugate joint prior for $(\mathbf{B}, \mathbf{\Sigma}^{-1})$ is the Normal-Wishart distribution. In this case, Guttman (1988); Press (2003) showed that the posterior distribution of $(\mathbf{B}, \mathbf{\Sigma}^{-1})$ also follows a Normal-Wishart. Similarly, it is possible to use the following decomposition of the prior distribution: $p(\mathbf{B}, \mathbf{\Sigma}) = p(\mathbf{B} \mid \mathbf{\Sigma})p(\mathbf{\Sigma})$ (Press, 1972). This is a simplified form for the Normal-inverse-Wishart as in Aitchison and Dunsmore (1975). One can see the similarity of such prior construction with the posterior distributions defined in Equations (3.6) and (3.9).

Conjugate prior distributions for both $p(\mathbf{B} \mid \mathbf{\Sigma})$ and $p(\mathbf{\Sigma})$ can be given. First, the prior distribution of $\mathbf{B} \mid \mathbf{\Sigma}$ is defined as the $(p \times m)$-dimensional matrix-variate Normal distribution with mean $\mathbf{B}_0$ (same dimension than $\mathbf{B}$) and covariance matrices $\mathbf{\Sigma}$ and $\mathbf{\Sigma}_0$, for the columns and the rows of $\mathbf{B}$, respectively. That is,

$$\mathbf{B} \mid \mathbf{\Sigma} \sim N_{p\times m}\left(\mathbf{B}_0, \mathbf{\Sigma}, \mathbf{\Sigma}_0\right). \tag{3.14}$$

The dependency on $\mathbf{\Sigma}$ leads to the restriction that the $(p \times p)$ prior matrix $\mathbf{\Sigma}_0$ is common for every m responses, i.e. all the corresponding regressors $\boldsymbol{\beta}_{1,\ldots,j,\ldots,m}$ have a similar prior covariance. Nevertheless, this restriction eases the identification of the posterior and predictive distributions.

Second, the prior distribution for $\mathbf{\Sigma}$ is chosen as an inverse-Wishart distribution (defined as in Box and Tiao, 1973):

$$\mathbf{\Sigma} \sim W_1^{-1}(\mathbf{\Omega}, \nu_0), \tag{3.15}$$

where $\mathbf{\Omega}$ is the *a priori* responses scale matrix, that has the same interpretation as a sum of squared errors. For instance, it might be $\mathbf{\Omega} = \mathbf{\Sigma}_{\text{prior}}.\nu_0$, where $\mathbf{\Sigma}_{\text{prior}}$ is a covariance matrix estimated over previous experiments. $\nu_0 > 0$ is the number of degrees of freedom of the prior distribution. The value of ν_0 indicates the certainty that one may have in $\mathbf{\Omega}$. For a simple interpretation of the prior distribution, we define ν_0 in the same form than ν. That is, $\nu_0 = n_0 - (m + p) + 1$. It is advised to keep ν_0 (or n_0) as low as possible to moderate the prior subjectivity. Sensitivity analysis can be done as well. n_0 can be seen as the number of virtual observations of the prior distribution.

3.3.2 Posterior distributions

Given the likelihood and the prior distributions (Equations (3.14) and (3.15)), and applying Bayes' theorem, the joint posterior density of the parameters is

$$\begin{aligned} p\left(\mathbf{B},\boldsymbol{\Sigma}\mid \text{data}\right) \propto & \mathcal{L}\left(\mathbf{B},\boldsymbol{\Sigma}\mid \mathbf{Y}\right).p\left(\mathbf{B}\mid\boldsymbol{\Sigma}\right).p\left(\boldsymbol{\Sigma}\right) \qquad (3.16)\\ & \propto|\boldsymbol{\Sigma}|^{\frac{-n}{2}}.\exp\left(-\frac{1}{2}tr\left[\boldsymbol{\Sigma}^{-1}\left(\mathbf{Y}-\mathbf{XB}\right)'\left(\mathbf{Y}-\mathbf{XB}\right)\right]\right)\\ & .|\boldsymbol{\Sigma}|^{\frac{-p}{2}}.\exp\left(-\frac{1}{2}tr\left[\boldsymbol{\Sigma}^{-1}(\mathbf{B}-\mathbf{B}_0)'\boldsymbol{\Sigma}_0^{-1}(\mathbf{B}-\mathbf{B}_0)\right]\right)\\ & .|\boldsymbol{\Sigma}|^{\frac{-\nu_0-2m}{2}}.\exp\left(-\frac{1}{2}tr\left(\boldsymbol{\Omega}\boldsymbol{\Sigma}^{-1}\right)\right). \end{aligned}$$

A metropolis-hasting algorithm can be implemented to explore this posterior distribution. However, to avoid efficacy trouble when dimensionality is moderate or high, it advised to use a built-in sampler such as WinBUGS (Lunn et al., 2000). Unfortunately, WinBUGS cannot be used directly to sample from the proposed joint posterior distribution. First, to be able to carry the computations, the matrix of regression parameters **B** must be converted into its vector form (see the equivalence in Equation (3.7)). Second, the dependence between **Σ** and **B** would imply a Kronecker product in the multivariate Normal prior for *vec*(**B**). WinBUGS has been found unable to deal with such operation.

To permit the computation, the WinBUGS code then slightly departs from the assumptions in that it forces the prior for **Σ** and **B** to be independent (i.e. $p(\mathbf{B},\boldsymbol{\Sigma}) = p(\mathbf{B}).p(\boldsymbol{\Sigma})$). This has a direct implication on the degrees of freedom (d.f.). Indeed, $p(\mathbf{B})$ no longer depends on **Σ**. As a result, $|\Sigma|^{\frac{-p}{2}}$ would then be missing in the prior density $p(\mathbf{B})$ (see the third line of Equation (3.16)). For this reason, p d.f. are lost due to the independent prior. To compare results, one may simply add these p d.f. to the prior $p(\boldsymbol{\Sigma})$ in WinBUGS, as shown in listing 3.1.

To use this code with a non-informative *a priori* on **B**, "InvSigxSig0" is defined as a low precision ($pm \times pm$) matrix. For the regression part (function inprod(...)), the elements of *vec*(**B**) that corresponds to each response are selected, which translates the matrix product **XB**. Finally, the Wishart distribution (function dwish(...)) is adapted to be in accordance with Box and Tiao (1973) (see comment on page 40). The degrees of freedom are then $(\nu_0 + p) + m - 1$. Finally, notice the different implementation of the Wishart distribution in Winbugs, that uses **Ω** instead of $\boldsymbol{\Omega}^{-1}$.

The sampler performs well when the dimensionality of the problem is low to moderate, but the computational burden remains high when the number of responses and parameters increases. Convergence can also be slow to achieve. Thus, it is preferable to identify the posterior distribution of the parameters. Fortunately, in this case, it is possible to identify them.

Listing 3.1: BUGS code for the multivariate regression.

```
 model
 {
  #For each observation
  for(i in 1:n) {
   Y[i,1:m] ~ dmnorm(mu.Y[i,1:m],invSigma[,]) #likelihood

   #For each response
   for(j in 1:m){
     mu.Y[i,j] <- inprod(vecB[((p*(j-1))+1):(j*p)], X[i,1:p])
   }
  }

  #prior distribution
  vecB[1:(m*p)] ~ dmnorm(vecB0[1:(m*p)],InvSigxSig0[,])
  invSigma[1:m,1:m] ~ dwish(Omega[,], nu0 + p + m -1)

  #convert precision matrix into covariance matrix
  Sigma[1:m,1:m] <- inverse(invSigma[,])
}
```

Conditional distribution of B given $\boldsymbol{\Sigma}$

The conditional posterior distribution of $\mathbf{B}$ given $\boldsymbol{\Sigma}$ can be identified as the following matrix-variate Normal:

$$\mathbf{B} \mid \boldsymbol{\Sigma}, \text{data} \sim N_{p\times m}\left(\mathbf{M}_{\mathbf{B}\text{post}}, \boldsymbol{\Sigma}, \left(\mathbf{X}'\mathbf{X} + \boldsymbol{\Sigma}_0^{-1}\right)^{-1}\right) \tag{3.17}$$

where

$$\mathbf{M}_{\mathbf{B}\text{post}} = \left(\mathbf{X}'\mathbf{X} + \boldsymbol{\Sigma}_0^{-1}\right)^{-1}\left(\mathbf{X}'\mathbf{X}\hat{\mathbf{B}} + \boldsymbol{\Sigma}_0^{-1}\mathbf{B}_0\right).$$

One can see that the posterior mean $\mathbf{M}_{Bpost}$ is a linear combination of the least-square estimation of $\mathbf{B}$ ($\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$) and its prior mean $\mathbf{B}_0$, weighted by their respective precision matrix (inverse of covariances). The posterior row covariance is the inverse of the sum of the matrix ($\mathbf{X}'\mathbf{X}$) and of the prior row precision matrix $\boldsymbol{\Sigma}_0^{-1}$.

Marginal distribution of B

It is possible to integrate $\boldsymbol{\Sigma}$ out of the joint posterior distribution in order to identify a simple form for the marginal posterior distribution of $\mathbf{B}$. It follows a matrix-variate Student's distribution.

$$\mathbf{B} \mid \text{data} \sim T_{p\times m}\left(\mathbf{M}_{\mathbf{B}\text{post}}, \boldsymbol{\Omega} + \mathbf{A}^*, \left(\mathbf{X}'\mathbf{X} + \boldsymbol{\Sigma}_0^{-1}\right)^{-1},\ \nu + n_0\right) \tag{3.18}$$

with $\mathbf{A}^* = \mathbf{Y}'\mathbf{Y} + \mathbf{B}_0'\boldsymbol{\Sigma}_0^{-1}\mathbf{B}_0 - (\mathbf{X}'\mathbf{X}\hat{\mathbf{B}} + \boldsymbol{\Sigma}_0^{-1}\mathbf{B}_0)'(\mathbf{X}'\mathbf{X} + \boldsymbol{\Sigma}_0^{-1})^{-1}(\mathbf{X}'\mathbf{X}\hat{\mathbf{B}} + \boldsymbol{\Sigma}_0^{-1}\mathbf{B}_0)$.

When the prior distributions are uniform, $\mathbf{A}^*$ is reduced to $\mathbf{A} = (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})'(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})$, which is the matrix presented by Geisser (1965) and Box and Tiao (1973) for the non-informative case.

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

Marginal distribution of $\boldsymbol{\Sigma}$

Integrating $\mathbf{B}$ out of (3.16) gives the marginal density of $\boldsymbol{\Sigma}$, identified as the following inverse-Wishart distribution:

$$(\boldsymbol{\Sigma} \mid \text{data}) \sim W_1^{-1}\left(\boldsymbol{\Omega} + \mathbf{A}^*, \nu + n_0\right) = W_2^{-1}\left(\boldsymbol{\Omega} + \mathbf{A}^*, \nu + n_0 + m - 1\right), \qquad (3.19)$$

Samples from the joint posterior $(\mathbf{B}, \boldsymbol{\Sigma} \mid \text{data})$ can be obtained by using Equation (3.19) followed by Equation (3.17).

3.3.3 Predictive distributions of a new response vector

In the informative case, the predictive distribution becomes:

$$\tilde{\mathbf{y}} \mid \tilde{\mathbf{x}}, \text{data} \sim T_m\left(\tilde{\mathbf{x}}\mathbf{M}_{\mathbf{B}\text{post}}, \left(1 + \tilde{\mathbf{x}}'(\mathbf{X}'\mathbf{X} + \boldsymbol{\Sigma}_0^{-1})^{-1}\tilde{\mathbf{x}}\right)(\boldsymbol{\Omega} + \mathbf{A}^*), \nu + n_0\right). \qquad (3.20)$$

That is, in the informative case, the multivariate Student's distribution is centered around the posterior mean regression surface, instead of its least-square estimates. With uniform prior distributions, Equation (3.20) naturally reduces to Equation (3.13).

Notice that for a joint prediction of a set of $\tilde{n}$ input conditions $\tilde{\mathbf{x}}_1, ..., \tilde{\mathbf{x}}_{\tilde{n}}$, the predictive distribution nicely extends to a matrix-variate Student's distribution, as presented in Appendix A.

With such results, Gibbs sampling or MCMC techniques are not required and predictions can be obtained directly from this predictive distribution. This allows a potentially immediate computation of Design Space when the modeled responses are the Critical Quality Attributes of interest.

3.4 Equivalence between sampling algorithms, a small simulation study

A small simulation study has been conducted. It involved a set of $m = 3, 5, 10$ responses modeled jointly, each with $p = 4, 6$ regression parameters. The aim was to assess the equivalence between different sampling schemes for the posterior and the predictive distributions. For each scenario, the number of virtual observations envisaged for the inverse-Wishart prior distribution were $n_0 = 3, 8, 17$.

Simulated data

The simulated responses were created as if they were recorded during a designed set of experiments. The design consisted of a 3^2 full factorial design with the two factors normalized between -1 and 1, with 3 repetitions added randomly on different points of the design, so that it is not always balanced. The model contained an intercept, two main effects, one interaction and two quadratic effects (for $p = 6$). In this example, $\mathbf{X}$ is then a (12×4) or a (12×6) matrix. The true parameters were randomly generated following an Uniform distribution between -40 and 40. A multivariate error vector of size m was then generated following a multivariate Normal $N(\mathbf{0}, \mathbf{\Sigma}_{true})$ for each of the 12 observations. Different variances and covariance structure were given for $\mathbf{\Sigma}_{true}$, to simulate correlated and uncorrelated data (matrices not presented).

Only vague prior distributions $(\mathbf{\Omega}, \mathbf{\Sigma}_0^{-1}, \mathbf{B}_0 = 0)$ are used in order to make the comparisons possible when using Winbugs. Indeed, the structure of the prior used in Winbugs are different and it is difficult to provide a similar information using $p(\mathbf{B}, \mathbf{\Sigma}) = p(\mathbf{B} \mid \mathbf{\Sigma}).p(\mathbf{\Sigma})$ or $p(\mathbf{B}, \mathbf{\Sigma}) = p(\mathbf{B}).p(\mathbf{\Sigma})$.

18 simulations were carried out for each assessed covariance matrices $\mathbf{\Sigma}_{true}$. Some simulations were not possible when n_0 was too low, as the inverse-Wishart distributions must have its d.f. strictly higher than $m-1$. In this case, n_0 has been increased so that the prior and posterior inverse-Wishart distributions are always identified. In each simulation and each sampling scheme, 200.000 samples were drawn, allowing to fit kernel densities and to record credible or predictive intervals.

Posterior distribution of the regression parameters

The envisaged sampling schemes to obtain samples from the marginal posterior $p(\mathbf{B} \mid \text{data})$ are the following (see explanations in the previous sections for d.f. assignment):

1. Winbugs (MCMC sampling) is used to draw samples from the unidentified joint posterior $p(\mathbf{B}, \mathbf{\Sigma} \mid \text{data})$, the d.f. of the prior Wishart distribution W_2 for $\mathbf{\Sigma}^{-1}$ being $(\nu_0 + p) + m - 1 = n_0$ (Equation (3.16)). $\mathbf{\Sigma}^{-1}$ is then inverted to obtain $\mathbf{\Sigma}$.

2. The identified $p(\mathbf{B} \mid \mathbf{\Sigma}, \text{data})$ and $p(\mathbf{\Sigma} \mid \text{data})$ are used to draw the samples, the d.f. of the posterior inverse-Wishart distribution W_2^{-1} being $\nu + n_0 + m - 1$ (Equations (3.17) and (3.19));

3. The marginal posterior $p(\mathbf{B} \mid \text{data})$ is used to draw the samples with $\nu + n_0$ d.f. (Equation (3.18)).

Figure 3.1 illustrates the comparison of the kernel densities of the marginal posterior distribution of $\mathbf{B}$ for $m = 5, p = 4, n_0 = 5$ (A) and $m = 2, p = 6, n_0 = 30$ (B). In blue, the curves are the densities obtained with scheme 1 (Winbugs sampling). The red densities are obtained using scheme 2 (sampling from the identified posterior distributions). The black densities are obtained with the scheme 3 (direct sampling from the marginal posterior distribution of $\mathbf{B}$). For each parameters, the 99% credible interval are graphically provided. They are in total agreement whatever the sampling scheme.

Predictive distribution of the responses

The next simulations aimed at exploring the predictive distribution of a new response vector, from which samples are obtained using three different sampling schemes.

The two firsts assume that n^* samples of the joint posterior distribution $p(\mathbf{B}, \mathbf{\Sigma} \mid \text{data})$ are available. To predict a new response vectors from the posterior samples, the following equation isapplied, that numerically solve Equation (3.12) to obtain samples from the predictive distribution.

$$(\tilde{\mathbf{y}}^{(s)} \mid \mathbf{Y}, \mathbf{x}_0, \text{data}) \sim N(\mathbf{x}_0 \mathbf{B}^{(s)}, \mathbf{\Sigma}^{(s)}), \tag{3.21}$$

with $(\mathbf{B}^{(s)}, \mathbf{\Sigma}^{(s)}) \sim p(\mathbf{B}, \mathbf{\Sigma} \mid \text{data})$, $s = 1, ..., n^*$.

Finally, the third sampling perspective is the draw of samples from the predictive distribution directly. In summary, the three sampling schemes envisaged are:

1. The joint posterior $p(\mathbf{B}, \mathbf{\Sigma} \mid \text{data})$ from Winbugs (Equation (3.16)) is used, followed by Equation (3.21);

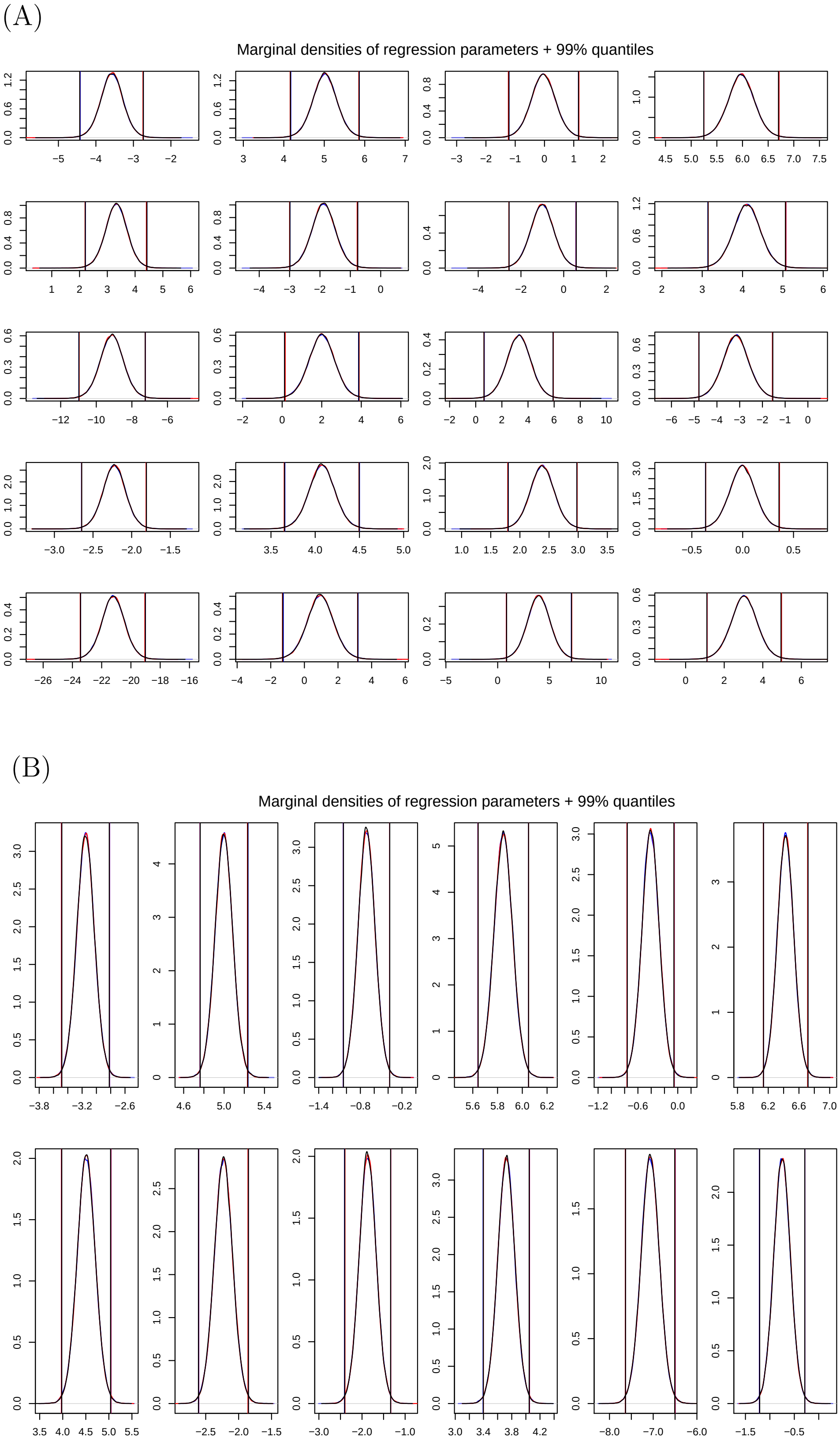


Figure 3.1: Marginal posterior distributions of the parameters in two sampling simulations. The results of 3 different sampling schemes are superimposed.

2. The joint posterior obtained as a draw from the identified $p(\mathbf{\Sigma} \mid \text{data})$, then a draw from $p(\mathbf{B} \mid \mathbf{\Sigma}, \text{data})$, is used (Equations (3.19) and (3.17)), followed by Equation (3.21);
3. The identified predictive distribution (with $\nu + n_0$ d.f.) is directly used (Equation (3.20)).

Figure 3.2 illustrates the comparison of the kernel densities of the marginal predictive distribution for $m = 10, p = 6, n_0 = 10$ (A) and $m = 5, p = 4, n_0 = 5$ (B). The blue curves are the marginal densities obtained with scheme 1 (Winbugs). The red densities are obtained from the sampling scheme 2 (identified posterior distributions). The green densities are obtained using scheme 3 (predictive distribution). For each response, the 99% predictive interval are graphically provided. They are also in agreement whatever the sampling scheme.

3.5 Linear constraint on the responses

Assume that some constraints apply on the responses. For instance, in the chromatographic models presented in Chapters 7 and 8, the responses representing one peak (k_B, k_A and k_E) must be well-ordered. Indeed, it is physically mandatory to have $k_B < k_A < k_E$. Not only mean responses must satisfy the linear constraints, but the whole (joint) distribution as well (i.e., every joint sample). Notice that a strong correlation between the responses does not ensure the constraints to be fulfilled.

Naive rejection sampling can be employed: samples of the distribution not achieving the constraints are discarded. However, in complex cases, the ratio of rejected to accepted samples can be very high. In this case, response transformations might be preferred, as described in Chapter 7.

Another possibility is to use a truncated version of the multivariate Student's distribution to directly generate samples that include the constraints as proposed by Geweke (1991). This technique is briefly presented in Appendix E.3.

The constraints must be explicitly given on the form of a matrix $\mathbf{C}$ and two vectors $\mathbf{a}$ and $\mathbf{b}$, giving the structure of the linear constraints and the limiting values, respectively. Formally, if the m-sized vector $\boldsymbol{Y}$, distributed as a $T_m(\boldsymbol{\mu}, \mathbf{\Sigma}, \nu)$, must satisfy $\mathbf{a} \leq \mathbf{C}\boldsymbol{Y} \leq \mathbf{b}$, it is possible to define $\mathbf{C}$ as a full-rank $m \times m$ matrix expressing the linear constraints, and the elements of $\mathbf{a}$ and $\mathbf{b}$ can be any reals between $-\infty$ and $+\infty$ ($\mathbf{a}_j < \mathbf{b}_j$, j= 1,...,m). Using this notation, a maximum of m linear restrictions can be applied on $\boldsymbol{Y}$, which is generally sufficient for many problems.

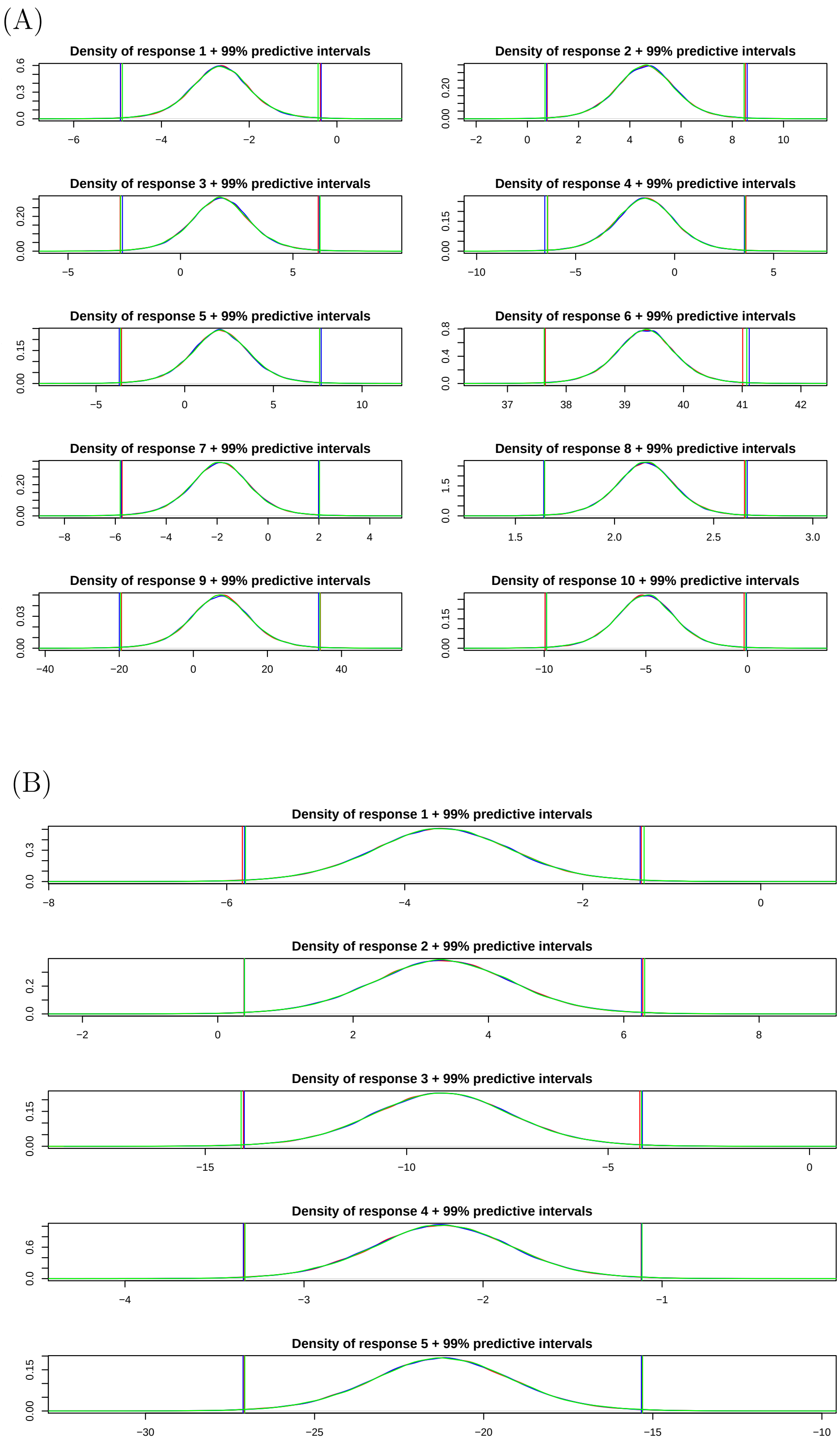


Figure 3.2: Marginal predictive distributions in two sampling simulations. The results of 3 different sampling schemes are superimposed.

Chapter 4

Bayesian Hierarchical models

4.1 Introduction

An univariate response Y is observed n times as $\mathbf{y} = (y_1, ..., y_n)$, at different combinations of k factors (input variables). This allows defining a model matrix specifying p variables explaining the effects of the k factors on the response. $\mathbf{X} = (\mathbf{x}_1, ..., \mathbf{x}_p)$ is this $n \times p$ model matrix. In many situations, the relation between $\mathbf{y}$ and $\mathbf{X}$ is well explained by a linear model of the form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

In this simple expression of linear model, $\boldsymbol{\varepsilon}$ is assumed to be the main source of error.

However, when the data are collected in such way that different observations are taken on related individuals (e.g. same person, same batch), clustered into some *series*, slight differences will be observed on the parameters of the linear model, for each of the different series of data. Then, the data are nested or hierarchically structured. In this case, the assumptions on the error $\boldsymbol{\varepsilon}$ (Gaussian, i.i.d.) are too simple to analyze the different source of variations that occur in the model.

This is customary when an analytical or biological quantitative method is *validated* to prove that its results will be accurate in its future routine use. For this purpose, the data are recorded in series that may represent different changes such as operator, batches of samples or analytical devices. The aim of this sampling scheme is to reproduce the real life of the analytical method. In practice, these different changes induce an additional source of error that will affect the results of the methods.

A mixed-effects modeling allows including this type of series effect as one or several additional random parameters, that are well suited to account for the modeling of such nested data. The general form of the model becomes:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}, \tag{4.1}$$

where $\boldsymbol{\beta}$ is the vector of the so-called fixed effects and $\boldsymbol{\alpha}$ is the vector of random effects parameters. $\mathbf{Z}$ is a known design matrix for the random effects parameters. The role of $\boldsymbol{\alpha}$ is to extend the way the error is described by the model. For instance, correlations between measurements can be estimated. For a general approach of mixed models, see for instance Henderson (1990); Searle et al. (1992); Verbeke and Molenberghs (1997); Demidenko (2004).

Further assumptions are that both $\boldsymbol{\epsilon}$ and $\boldsymbol{\alpha}$ are 0-centered and they are mutually independent. In the Normal case, the covariance structure of $\boldsymbol{\alpha}$ is $\mathbf{G}$ and the one of $\boldsymbol{\varepsilon}$ is $\boldsymbol{\Sigma}$.

Chapter structure

First, in Section 4.2, the one-way ANOVA random model for repeated measurements is presented as a special case of Equation (4.1). $\mathbf{X}$ specifies an intercept and $\mathbf{Z}$ contains dummy variables that code one categorical variable of interest (the series). $\mathbf{G}$ describes the variances in a diagonal structure, and $\boldsymbol{\Sigma}$ contains the classical error assumption $\boldsymbol{\Sigma} = \sigma_{\varepsilon}^2 \mathbf{I}_n$, $\mathbf{I}_n$ being a $(n \times n)$ identity matrix. Generally, the interest is to analyze the variances and covariances in the model, see e.g. Box and Tiao (1973); Hill (1965); Krishnamoorthy and Mathew (2010). However, in this work, the Bayesian predictive distribution remains the main interest. This distribution describes the predictive uncertainty of every single new data. This uncertainty comes from both the measurements and the series-to-series variation.

Second, Section 4.3 illustrates the problem of simple hierarchical linear regression in linear calibration problems. In this hierarchical univariate regression problem, the matrices $\mathbf{X} = \mathbf{Z}$ are identical and allow the estimation of an intercept and a parameter corresponding to a slope coefficients for one input variable. Both parameters are estimated for the fixed and random effects. The Bayesian predictive distributions of the response $(\tilde{y} \mid \tilde{x}, \text{data})$ and of the back-calculated input variable $(\tilde{x} \mid \tilde{y}, \text{data})$ will be the main result of this section as they are useful to assess the uncertainty of a linear calibration.

Finally, the last section introduces non-linear mixed-effects models in a similar calibration problem. These models are useful when a linear model is not able to fit the analytical data properly. In this case, a non-linear model is often appropriate to fit the calibration curve of ligand-binding assays with serial dilutions (e.g.

ELISA test). Typically, a 4-parameters logistic regression with a variance modeled as proportional to a *power of the mean* is used as a calibration curve. Some fixed parameters might have a random counterpart while others might be assumed fixed.

4.2 Bayesian one-way ANOVA random model

Assume that the observations of Y are repeated in m different experimental units such as series (either batches, individuals, days, devices, etc.) with n_j observations per series. The data $\mathbf{y}_j$ observed on the j^{th} serie will present a certain correlation, as they come from the same or experimental unit ($j = 1, ..., m$). The complete data set may be written $\mathbf{y} = (\mathbf{y}_1, ..., \mathbf{y}_m)$.

A wide review of classical and Bayesian methodologies to estimate variance components can be found in Khuri and Sahai (1985), that particularly address the one-way ANOVA random model. For the frequentist results addressing predictions, the approximation developed by Mee (1984) is a reference in this domain.

To remain as general as possible and to be concise, the posterior and predictive distributions will be presented in a sampling perspective, although it is possible to obtain closed-forms for the posterior distributions in specific cases (see for instance the chapter 5 of Box and Tiao, 1973).

4.2.1 Model

To capture the correlation between the data of the same series, the series effect are modeled by a random effect parameter α_j. Assuming for the sake of generality that the data are unbalanced, the one-way ANOVA random model can be written as:

$$y_{ij} = \mu + \alpha_j + \varepsilon_{ij}, \quad \text{with } \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2) \text{ and } \alpha_j \sim N(0, \sigma_\alpha^2) \tag{4.2}$$
$$j = 1, ..., m, \quad i = 1, ..., n_j, \quad n = \sum_{j=1}^{m} n_j.$$

In this equation, μ is the overall mean. The aim is to estimate the variance components that quantify the *between-series* variability ($\sigma_\alpha^2 = 1/\tau_\alpha$) and the *within-series* variability ($\sigma_\varepsilon^2 = 1/\tau_\varepsilon$)[1] in order to find the total variability of the response (referred as the intermediate precision, i.p.). The two variance components are assumed independent so that $var_{\text{i.p.}}(Y) = \sigma_\alpha^2 + \sigma_\epsilon^2$.

[1] τ_α and τ_ε are called *precisions*, i.e. the inverse of the variances.

An alternative notation can be used to simplify the description of the model:

$$y_{ij} \sim N(\alpha_j, \sigma_\varepsilon^2), \text{ with } \alpha_j \sim N(\mu, \sigma_\alpha^2). \tag{4.3}$$

Notice that when $n_1 = n_j = n_m$, the design is balanced. The vector of parameters is then $(\mu, \sigma_\alpha^2, \sigma_\varepsilon^2)$, or equivalently, $(\mu, \tau_\alpha, \tau_\varepsilon)$.

4.2.2 Likelihood

The likelihood $\mathcal{L}(\mu, \sigma_\alpha^2, \sigma_\varepsilon^2 \mid \mathbf{y})$ can be expressed as the following density (see Hill (1965); Box and Tiao (1973))

$$\mathcal{L}(\mu, \sigma_\alpha^2, \sigma_\varepsilon^2 \mid \mathbf{y}) = \prod_{j=1}^{m} \int_{\alpha_j} p(\mathbf{y} \mid \alpha_j, \sigma_\varepsilon^2).p(\alpha_j \mid \mu, \sigma_\alpha^2).d\alpha_j \tag{4.4}$$

$$\propto (\sigma_\varepsilon^2)^{-\frac{n-m}{2}}.\exp\left[-\frac{\sum\limits_{i,j}(y_{ij}-\bar{y}_j)^2}{2\sigma_\varepsilon^2}\right] \prod_{j=1}^{m} (\sigma_\varepsilon^2 + n_j\sigma_\alpha^2)^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\sum_{j=1}^{m} \frac{n_j(\bar{y}_j-\mu)^2}{\sigma_\varepsilon^2 + n_j\sigma_\alpha^2}\right], \tag{4.5}$$

with $\sigma_\alpha^2 > 0$ and $\sigma_\varepsilon^2 > 0$. Equivalently, in terms of precisions, the likelihood becomes:

$$\begin{aligned}\mathcal{L}(\mu, \tau_\alpha, \tau_\varepsilon \mid \mathbf{y}) \propto \tau_\varepsilon^{\frac{n-m}{2}} \exp\left[-\frac{\tau_\varepsilon}{2}\sum_{i,j}(y_{ij}-\bar{y}_j)^2\right] \\ .\prod_{j=1}^{m}(\tau_\alpha + n_j\tau_\varepsilon)^{-\frac{1}{2}}(\tau_\varepsilon\tau_\alpha)^{\frac{1}{2}} \exp\left[-\frac{1}{2}\sum_{j=1}^{m}\frac{\tau_\varepsilon\tau_\alpha}{\tau_\alpha + n_j\tau_\varepsilon}n_j(\bar{y}_j-\mu)^2\right]\end{aligned} \tag{4.6}$$

The first line of (4.6) expresses the information concerning y_{ij}. The second line represents the information concerning the random parameters α_j.

The complexity of this density (mixture of Normal densities) gives clue on the difficulty one may have to work analytically on its basis. This also gives little information on the way to find conjugate prior distribution.

4.2.3 Prior distribution of the parameters

Independent prior distributions are assumed for the parameters μ, τ_ε and τ_α. To avoid negative variance estimations, that is a major concern in the mixed model estimation, Culver (1971) examined the prior distribution of the within-series and between-series variance components. He proposed to use two independent inverted-Gamma prior distributions for the variance components and one uniform prior distribution for the location parameter μ. Similarly, when being non-informative, it is also possible to use uniform prior distributions on the domain $[0, c]$, with c being a large positive value. The sampling approach described by Wolfinger (1998) is developed in the following sections. It extends the one of Culver by replacing the uniform distribution of the location with a Normal.

In term of precision, a possible prior distribution for τ_ε is then a Gamma distribution, that corresponds to the choice of an inverted-Gamma for σ^2_ε.

$$p(\tau_\varepsilon) = Gamma(k^0_\varepsilon, \delta^0_\varepsilon). \tag{4.7}$$

For a distribution representing a diffuse information, k^0_ε and δ^0_ε are chosen being very small positive values, such as 10^{-5}. In this case, the Gamma is very flat in the domain $[0, +\infty)$.

For the parameters μ and τ_α, the prior distributions can be chosen as Normal and Gamma distributions, respectively.

$$p(\mu) = N(\mu_0, \sigma^2_0) = N(\mu_0, 1/\tau_0) \tag{4.8}$$

$$p(\tau_\alpha) = Gamma(k^0_\alpha, \delta^0_\alpha), \tag{4.9}$$

Diffuse *a priori* for μ can be set up with a very low precision $\tau_0 = 10^{-5}$. Notice that if the precision is very low, the particular value of μ_0 is non significant. An usual choice is to give a value of $\mu_0 = 0$. With this Normal prior, flexibility is left to define an informative prior distribution if desired, with μ_0 and τ_0 being values that might be estimated over previous experimental data, for instance. For $p(\tau_\alpha)$, diffuse knowledge can be given fixing k^0_α and δ^0_α to very small positive values, e.g. 10^{-5}. Informative prior knowledge can be used as well, choosing different values.

When possible, a direct implication of the use of informative prior information in analytical method validation and calibration is to be able to provide the results with limited experiments. In this case, the validity of the prior information must be guaranteed to avoid harmful biases.

4.2.4 (Sampled) posterior distribution of the parameters

Applying Bayesian rule, the joint posterior density of the parameters is expressed as

$$p(\mu, \tau_\alpha, \tau_\epsilon \mid \text{data}) \propto \mathcal{L}(\mu, \tau_\alpha, \tau_\varepsilon \mid \mathbf{y}).p(\tau_\varepsilon).p(\mu).p(\tau_\alpha). \tag{4.10}$$

MCMC sampling is used to draw samples from the joint posterior distribution of (4.10), with the prior distribution defined in Equations (4.7)-(4.9), provided that it is easy to have enough samples in a short computation time. In this context, Rajagopalan and Broemeling (1983), Harville and Zimmermann (1996) and Gamerman (1997) illustrated the use of MCMC methods to characterize the posterior distribution of parameters for the mixed-effect model. The classical MCMC methods presented in the Appendix C can be used to carry out such a work. Specialized softwares such as WinBUGS (Lunn et al. (2000)) and the MCMC procedure from SAS software's (SAS/STAT® 9.2.1 User's Guide, SAS Institute Inc. (2010)) provide nice interfaces to MCMC methods. An example of BUGS code for the unbalanced random one-way ANOVA is given in listing 4.1.

Listing 4.1: BUGS code for unbalanced random one-way ANOVA.

```
model{
        for(j in 1:m){
                for(i in n[j]:(n[j+1]-1)){
                    y[i] ~ dnorm(a[j], tau.e) #likelihood
                }
                a[j] ~ dnorm(mu, tau.a)
        }

        #flat prior distributions
        mu ~ dnorm(0, 0.0001)
        tau.a ~ dgamma(0.0001,0.0001)
        tau.e ~ dgamma(0.0001,0.0001)

        #convert precision into variance
        sigma2.e <- 1/tau.e
        sigma2.a <- 1/tau.a
  }
```

In the previous piece of code, the observations are stacked in one vector y and n[j] is a value of some offsets that indicates the series for each observation.

Notice that WinBUGS uses Normal distribution (dnorm()) whose dispersion is parametrized with a precision instead of a variance. The location of the posterior distributions of σ^2_α and σ^2_ε are useful to identify the major source of variance. Their spread may be used to assess the quality of these variance estimates.

The focus is now put on the sampled predictive distribution of a new observation of Y, that includes within-series and between-series sources of variability.

4.2.5 (Sampled) predictive distribution of one new response

In the frequentist predictive approach, the main works about prediction have been done on the computation of tolerance intervals (Wilks, 1941; Wald and Wolfowitz, 1946; Wallis, 1951; Ellison, 1964; Mee, 1984; De Gryze et al., 2007). Many complications arise when sufficient statistics for the model can not be analytically described. This is mostly the case when departing from Normality assumptions and when the data are unbalanced (Krishnamoorthy and Mathew, 2010).

From a Bayesian perspective, the link between the predictive distribution and tolerance intervals is direct, as pointed out by Guttman (1970). Following his idea, the β.(100)% HPD quantiles of a predictive distribution of a new response are simply the β-expectation tolerance intervals (see also Guttman, 1988; Hamada et al., 2004). Tolerance intervals can then be computed from the predictive distribution using simulations (see Aitchison, 1964; Wolfinger, 1998; Krishnamoorthy and Mathew, 2010).

The predictive density of a new observation $\tilde{y}$ is computed as

$$p(\tilde{y} \mid \text{data}) = \int_{\mu}\int_{\tau_\alpha}\int_{\tau_\varepsilon} p(\tilde{y} \mid \mu, \tau_\alpha, \tau_\varepsilon).p(\mu, \tau_\alpha, \tau_\varepsilon \mid \text{data}).d\tau_\varepsilon.d\tau_\alpha.d\mu. \tag{4.11}$$

Difficulties arises in the analytical identification of the predictive distribution of $\tilde{y}$. Instead, a simple simulation procedure can be used to draw samples from $(\tilde{y} \mid \text{data})$. First, following the model, the random effect parameter α_j is distributed as a Normal with *a posteriori* location $(\mu|\text{data})$ and precision $(\tau_\alpha|\text{data})$ (Wolfinger, 1998). This allows drawing samples of new random parameters $\tilde{\alpha}^{(1)}, \ldots, \tilde{\alpha}^{(s)}, \ldots, \tilde{\alpha}^{(n^*)}$. Each $\tilde{\alpha}^{(s)}$ then represents the effect of a new predicted serie that follows the observed between-series variance component. Second, this series effect is included in the model, to draw a new response. As stated by the model, it also follows a Normal distribution with posterior location $\tilde{\alpha}^{(s)}$ and estimated within-series precision $(\tau_\varepsilon \mid \text{data})$: $\tilde{y}^{(s)} \sim N(\tilde{\alpha}^{(s)}, \tau_\varepsilon \mid \text{data})$. This gives the following algorithm:

```
For $s = 1$ to $n^*$
    1. sample $(\mu^{(s)}, \tau_\alpha^{(s)}, \tau_\varepsilon^{(s)})$ from $p(\mu, \tau_\alpha, \tau_\epsilon \mid \text{data})$,    see Eq. (4.10)
    2. sample $\tilde{\alpha}^{(s)}$ from $N(\mu^{(s)}, \sigma_\alpha^{2(s)})$, or from $N(\mu^{(s)}, 1/\tau_\alpha^{(s)})$,
    3. sample $\tilde{y}^{(s)}$ from $N(\tilde{\alpha}^{(s)}, \sigma_\varepsilon^{2(s)})$, or from $N(\tilde{\alpha}^{(s)}, 1/\tau_\varepsilon^{(s)})$.
End
```

At the end, samples $\tilde{y}^{(1)}, ..., \tilde{y}^{(s)}, ..., \tilde{y}^{(n^*)}$ following $p(\tilde{y} \mid \text{data})$ are obtained. A simplification of the sampling process is possible, replacing step 2 and 3 of the algorithm by the sampling of $\tilde{y}^{(s)}$ directly from $N(\mu^{(s)}, \sigma_{\alpha}^{2(s)} + \sigma_{\varepsilon}^{2(s)})$. The same predictive distribution is obtained, due to the independence assumption between the variance components.

4.2.6 Simulation results

When prior distributions tends to be uniform, one normally gets the same predictive distribution that the one that can be obtained (approximately) in the Frequentist framework. In this section, the β-expectation tolerance interval presented by Mee (1984) and applied in Hubert et al. (2007) is used and compared to the β-expectation tolerance interval obtained from the Bayesian predictive distribution.

The simulation set up is as follows: $m = 3, ..., 200$ series of data were created. For each serie, $n_j = 3$ or 10 replicates (e.g. measurements of the same batch) were available, following the generating model:

$$y_{ij} = \mu + \alpha_j + \varepsilon_{ij}, \quad \text{with } \varepsilon_{ij} \sim N(0, \sigma_{\varepsilon}^2) \text{ and } \alpha_j \sim N(0, \sigma_{\alpha}^2) \qquad (4.12)$$
$$j = 1, ..., m, \quad i = 1, ..., n_j, \quad n = \sum_{j=1}^{m} n_j.$$

μ was set to 100; the between variance σ_{α}^2 was assumed to be 5 and the within variance was defined from the ratio $\sigma_{\varepsilon}^2 = \sigma_{\alpha}^2/R$, where R was 1 or 20. When R was 20, the effect of the series is considerably higher than the residual error. To obtain unbalanced data, 3 observations were removed at random in different series, except when $m = 3$ and $n_j = 3$ $(j = 1, 2, 3)$, as the data set was already limited in this situation.

The $\beta-$expectation tolerance interval of Mee at $\beta = 95\%$ states that, on average, the proportion of future results that would be found within the interval is β. After being adapted for unbalanced data and expressed in relative scale, it can be defined as:

$$\Bigg[\text{bias}(\%) - t_{\frac{1+\beta}{2}}(\nu)\sqrt{1 + \frac{1}{n.B^2}} RSD_{IP} \ ;$$
$$\text{bias}(\%) + t_{\frac{1+\beta}{2}}(\nu)\sqrt{1 + \frac{1}{n.B^2}} RSD_{IP}\Bigg], \qquad (4.13)$$

where

- bias(%) $= \frac{\hat{\mu}-\mu}{\mu} \times 100$,

- μ is known from simulation set up (in practice, it may be assumed known from averaged measurements with negligible error); $\hat{\mu}$ is estimated from the data (grand mean),
- $RSD_{IP} = \frac{\hat{\sigma}_{IP}}{\mu} \times 100$, the relative standard deviation of intermediate precision, with
- $\hat{\sigma}^2_{IP} = \hat{\sigma}^2_{\alpha} + \hat{\sigma}^2_{\varepsilon}$, the estimate of the variance of the intermediate precision, i.e. the sum of the estimated within and between series variances, and
- $B = \sqrt{\frac{\hat{R}+1}{\frac{n}{m}\hat{R}+1}}$, with $\hat{R} = \frac{\hat{\sigma}^2_{\alpha}}{\hat{\sigma}^2_{\varepsilon}}$,
- $\nu = \frac{(\hat{R}+1)^2}{(\hat{R}+\frac{m}{n})^2/(m-1)+(1-\frac{m}{n})/n}$, (known as Satterthwaite (1941) approximation),
- $t_{\frac{1+\beta}{2}}(\nu)$, the $\frac{1+\beta}{2}$ quantile of the Student distribution with ν degrees of freedom.

In the Bayesian framework, samples $\tilde{y}^{(s)}$ of the predictive distribution can directly be used. According to Guttman (1970), the β-expectation tolerance interval is the 95% HPD quantiles of the predictive distribution. To compare to Equation (4.13) the samples from the predictive distribution were also put in relative scale:

$$\tilde{y}^{(s)}_r = \frac{\tilde{y}^{(s)} - \mu}{\mu} \times 100.$$

The HPD quantiles were then compute on $\tilde{y}^{(s)}_r$. μ was known from simulation set up.

Using the sampler presented in Section 4.2.4 with non-informative flat prior distributions, 10000 samples were recorded, after a burnin period of 5000 samples. Samples of the predictive distribution were then obtained as in Section 4.2.5. No convergency issue was detected in the simulations, using autocorrelation plots and Gelman-Rubin-Brooks test/plots provided by the `coda` R package (Brooks and Gelman, 1998; Plummer et al., 2010).

Results of the simulations are presented in Figure 4.1. It presents the β-expectation tolerance interval at 95% versus the number of series. The interval derived from Mee is presented in blue whereas the one obtained from the predictive distribution is shown in green. In the four cases (A, B, C and D), both intervals were very similar, particularly when the number of series was increased. When the number of series was at least $m = 30$, the total uncertainty remains stable, even if the number of repetition was low. With these scenarii, it seems more profitable to envisage enough series while limiting the number of repetitions.

In this manuscript, the β-expectation tolerance interval for the one-way ANOVA random model is applied to validate the results of a quantitative method. An example of application is given in Chapter 10.

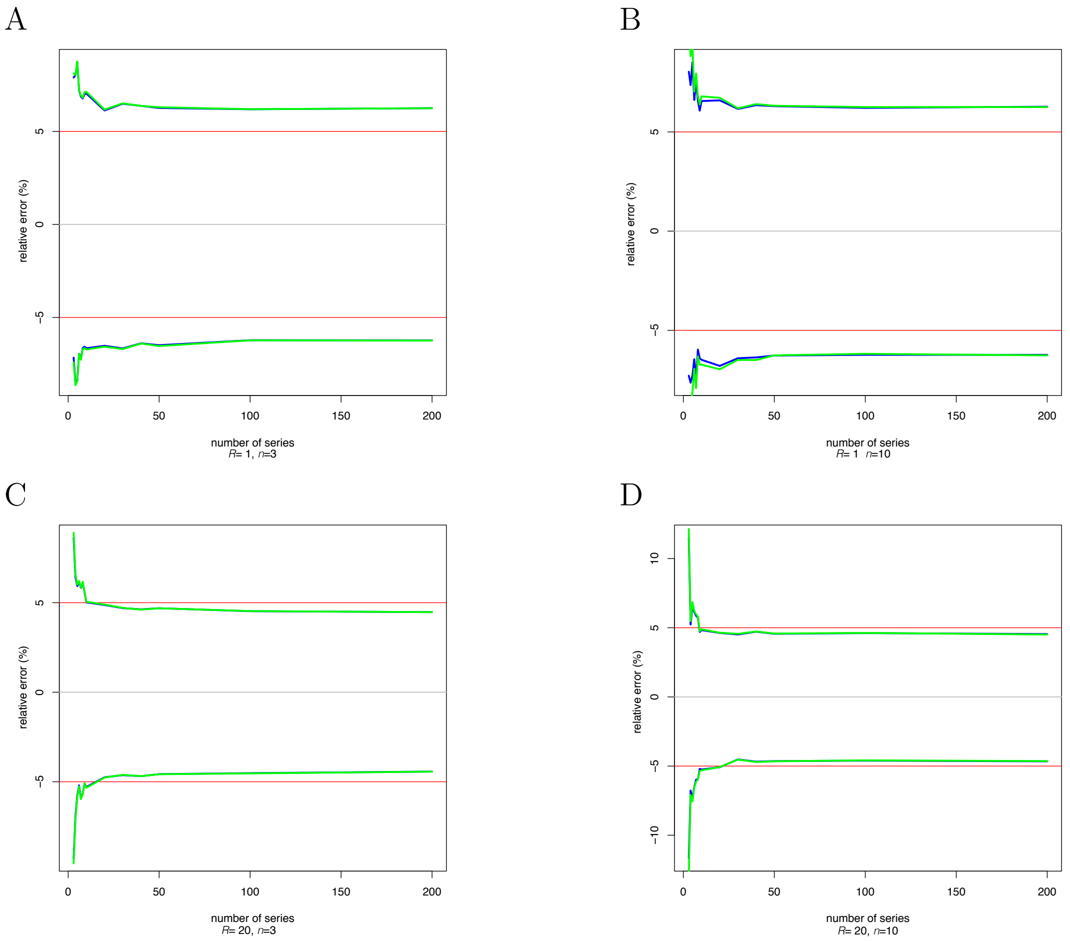


Figure 4.1: Evolution of the β-expectation tolerance interval when the number of series m is increased from 3 to 200. (Red) Reference total relative error specification fixed at + or - 5%. (Green) 95% HPD interval of the predictive distribution. (Blue) Frequentist tolerance interval as in Equation (4.13). (A) The ratio of the between and within variances, R, is 1 and the targeted number of repetitions, n_j, is 3. (B) $R = 1, n_j = 10$. (C) $R = 20, n_j = 3$. (D) $R = 20, n_j = 10$.

4.3 Bayesian hierarchical linear regression

The hierarchical regression is introduced in the context of the calibration of an analytical method. A comprehensive review of frequentist and Bayesian models for calibration is given in Osborne (1991).

In calibration problem, at a concentration of an analyte x_i corresponds a signal $Y \mid x_i$. The *standards* $\mathbf{x} = (x_1, ..., x_i, ..., x_n)$ are assumed known as they are determined by a supposed extremely accurate standard method. The error on $\mathbf{x}$ is then negligible. However, this standard method may also have some inconveniences such as running time or cost. The analytical method providing $\mathbf{y}$, the observations of Y, is generally more flexible, quicker and less expensive than the standard method. Nevertheless, this analytical method requires this calibration process.

After the calibration process, the analytical method is run (e.g. routine): a new signal is observed, and an inverse prediction is used to find the corresponding concentration. To mimic the real application of the method, different series of data may be used to create the calibration curve. Series effect can be included in the regression using random effect parameters. These series can represent different operators, different batches of samples, different devices carrying out the analytical method, etc. When the calibration curves are assumed linear, this model is useful to analyze the uncertainty one can expect for future runs.

4.3.1 Model

The proposed approach enriches the one described by Hunter and Lamboy (1981) through a mixed-effect modeling.

Assume that each measure of Y is repeated in m series, for different known concentration level $\mathbf{x}$. The general linear mixed-effect model equation is:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}.$$

In the simple linear case, this last equation can be written as:

$$y_{ij} = \beta_0 + \beta_1 x_i + \alpha_{0j} + \alpha_{1j} z_i + \varepsilon_{ij}, \qquad \text{with } j = 1, ..., m. \tag{4.14}$$

With hierarchical models, it is further assumed that $x_i = z_i$, and Equation (4.14) can be rewritten:

$$y_{ij} = (\beta_0 + \alpha_{0j}) + (\beta_1 + \alpha_{1j})x_i + \varepsilon_{ij}, \tag{4.15}$$

The classical assumptions are $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$, $\alpha_{0j} \sim N(0, \sigma_0^2)$ and $\alpha_{1j} \sim N(0, \sigma_1^2)$. As in the previous section, alternative notations can be used, in terms of variances or precisions:

$$\begin{aligned} &y_{ij} \sim N(\alpha_{0j} + \alpha_{1j} x_i, \sigma_\varepsilon^2), \text{ with } \alpha_{0j} \sim N(\beta_0, \sigma_0^2) \text{ and } \alpha_{1j} \sim N(\beta_1, \sigma_1^2), \text{ or },\\ &y_{ij} \sim N(\alpha_{0j} + \alpha_{1j} x_i, 1/\tau_\varepsilon), \text{ with } \alpha_{0j} \sim N(\beta_0, 1/\tau_0) \text{ and } \alpha_{1j} \sim N(\beta_1, 1/\tau_1). \end{aligned} \tag{4.16}$$

Notice that α_{0j} and α_{1j} are supposed independent. This is a simplistic assumption that can be improved using a multivariate Normal for their joint distribution, with a covariance matrix expressing their dependency.

For the non-balanced case, the total number of observations is the sum of all the data of each serie, $n = \sum_{j=1}^{m} n_j$.

4.3.2 Likelihood

With Normally distributed data in a mixed-effect model, the likelihood is decomposed as follows:

$$\begin{aligned} &\mathcal{L}(\beta_0, \beta_1, \tau_0, \tau_1, \tau_\varepsilon \mid \mathbf{y})\\ &= \prod_{j=1}^{m} \int_{\alpha_{0j}, \alpha_{1j}} \prod_{i=1}^{n_j} p(y_{ij} \mid \alpha_{0j}, \alpha_{1j}, \tau_\varepsilon) p(\alpha_{0j}, \alpha_{1j} \mid \beta_0, \tau_0, \beta_1, \tau_1) d(\alpha_{0j}, \alpha_{1j}),\\ &= \prod_{j=1}^{m} \int_{\alpha_{0j}, \alpha_{1j}} p(\mathbf{y} \mid \alpha_{0j}, \alpha_{1j}, \tau_\varepsilon) \qquad p(\alpha_{0j} \mid \beta_0, \tau_0) p(\alpha_{1j} \mid \beta_1, \tau_1) d\alpha_{0j} d\alpha_{1j}. \end{aligned} \tag{4.17}$$

The likelihood is then averaged over the possible values of the random parameters. This is complex to resolve analytically, but it is simple to code with the BUGS language as it will be shown later.

4.3.3 Prior distribution of the parameters

The following prior distributions $p(\beta_0, \beta_1, \tau_0, \tau_1, \tau_\varepsilon)$ can be defined, assuming the parameters are independent *a priori*:

$$\begin{aligned} p(\tau_\varepsilon) &= p(1/\sigma_\varepsilon^2) = Gamma(k_\varepsilon^0, \delta_\varepsilon^0),\\ p(\beta_0) &= N(\mu_{\beta_0}, 1/\tau_{\beta_0}),\\ p(\tau_0) &= p(1/\sigma_0^2) = Gamma(k_0^0, \delta_0^0),\\ p(\beta_1) &= N(\mu_{\beta_1}, 1/\tau_{\beta_1}),\\ p(\tau_1) &= p(1/\sigma_1^2) = Gamma(k_1^0, \delta_1^0). \end{aligned} \tag{4.18}$$

Vague information can be defined using small values for the hyper-parameters, for example:

$$\begin{aligned} k_\varepsilon^0, \delta_\varepsilon^0, k_0^0, \delta_0^0, k_1^0, \delta_1^0 &= 10^{-5}, \\ \tau_{\beta_0}, \tau_{\beta_1} &= 10^{-5}, \\ \text{and with } \mu_{\beta_0}, \mu_{\beta_1} &= 0. \end{aligned} \tag{4.19}$$

Informative prior distributions can be given in the same way, using estimates that come from previous experiments, if any.

4.3.4 (Sampled) posterior distribution of the parameters

Applying Bayesian rule, one obtains:

$$p(\beta_0, \beta_1, \tau_0, \tau_1, \tau_\varepsilon \mid \text{data}) \propto \mathcal{L}(\beta_0, \beta_1, \tau_0, \tau_1, \tau_\varepsilon \mid \mathbf{y}).p(\beta_0, \beta_1, \tau_0, \tau_1, \tau_\varepsilon). \tag{4.20}$$

This formula shares similarities with the one-way ANOVA random model of the previous section. The following BUGS code (listing 4.2) allows drawing samples from $p(\beta_0, \beta_1, \tau_0, \tau_1, \tau_\varepsilon \mid \text{data})$.

Listing 4.2: BUGS code for unbalanced hierarchical linear regression.

```
model{
       for(j in 1:m){
               for(i in n[j]:(n[j+1]-1)){
                   y[i] ~ dnorm(yj, tau.e) #likelihood
                   yj <- alpha0[j]+ alpha1[j]*x
               }
               alpha0[j] ~ dnorm(beta0, tau.0)
               alpha1[j] ~ dnorm(beta1, tau.1)
       }

  #hyper-priors #
  beta0 ~ dnorm(0, 0.0001)  #non-informative
  beta1 ~ dnorm(0, 0.0001)  #non-informative

  tau.0 ~ dgamma(0.0001,0.0001) #non-informative
  tau.1 ~ dgamma(0.0001,0.0001) #non-informative
  tau.e ~ dgamma(0.0001,0.0001)  #non-informative


  #convert precision into variance
  sigma2.e <- 1/tau.e
  sigma2.0 <- 1/tau.0
  sigma2.1 <- 1/tau.1
 }
```

4.3.5 (Sampled) predictive distribution of one new response

Given a new concentration $\tilde{x}$, the predictive distribution of a new observation $\tilde{y}$ is computed as

$$p(\tilde{y} \mid \tilde{x}, \text{data}) = \\ \int_{\beta_0,\beta_1,\tau_{b_0},\tau_{b_1},\tau_\varepsilon} p(\tilde{y} \mid \tilde{x}, \beta_0, \beta_1, \tau_{b_0}, \tau_{b_1}, \tau_\varepsilon).p(\beta_0, \beta_1, \tau_{b_0}, \tau_{b_1}, \tau_\varepsilon \mid \text{data}).d(\beta_0, \beta_1, \tau_{b_0}, \tau_{b_1}, \tau_\varepsilon). \tag{4.21}$$

It is possible to numerically solve this integral by doing the following computations (Wolfinger, 1998). First, obtain samples from the posterior distribution of the parameters. These samples are used to compute the values of regression parameters $\tilde{\alpha}_0^{(s)}$ and $\tilde{\alpha}_1^{(s)}$ that represent their future possible values for new predicted series $(s = 1, ..., n^*)$. Next, using these samples in the regression model, the residual uncertainty can be added to obtain samples from the predictive distribution of a new observation $\tilde{y}$, using Monte-Carlo simulations:

```
For $s = 1$ to $n^*$
    1. sample $(\beta_0^{(s)}, \beta_1^{(s)}, \tau_0^{(s)}, \tau_1^{(s)}, \tau_\varepsilon^{(s)})$ from $p(\beta_0, \beta_1, \tau_0, \tau_1, \tau_\varepsilon \mid \text{data})$,
    2. sample $\tilde{\alpha}_0^{(s)}$ from $N(\beta_0^{(s)}, 1/\tau_0^{(s)})$,
    3. sample $\tilde{\alpha}_1^{(s)}$ from $N(\beta_1^{(s)}, 1/\tau_1^{(s)})$,
    4. sample $\tilde{y}^{(s)} \mid \tilde{x}$ from $N(\tilde{\alpha}_0^{(s)} + \tilde{\alpha}_1^{(s)}\tilde{x}, 1/\tau_\varepsilon^{(s)})$.
End
```

As the variance parameters are assumed independent, it is also possible to sample directly as follows:

```
For $s = 1$ to $n^*$
    1. sample $(\beta_0^{(s)}, \beta_1^{(s)}, \tau_0^{(s)}, \tau_1^{(s)}, \tau_\varepsilon^{(s)})$ from $p(\beta_0, \beta_1, \tau_0, \tau_1, \tau_\varepsilon \mid \text{data})$,
    2. sample $\tilde{y}^{(s)} \mid \tilde{x}$ from $N(\beta_0^{(s)} + \beta_1^{(s)}\tilde{x}, 1/\tau_\varepsilon^{(s)} + 1/\tau_0^{(s)} + 1/\tau_1^{(s)})$,
                               $= N(\beta_0^{(s)} + \beta_1^{(s)}\tilde{x}, \sigma_\varepsilon^{2(s)} + \sigma_0^{2(s)} + \sigma_1^{2(s)})$.
End
```

4.3.6 (Sampled) distribution of the inverse prediction of one new result

When using the calibration line in practice, the interest is generally centered on the x values, i.e. the concentration of an analyte, given a observed signal $\tilde{y}$ provided by the analytical method.

In the bayesian framework, there is an opportunity to easily obtain samples from the predictive distribution of these concentrations given a new signal, using the (sampled) posterior distribution of the parameters obtained as in section 4.3.4. The idea is to use use the inverse of the linear calibration line (Hunter and Lamboy, 1981):

$$\tilde{x} = \frac{\tilde{y} - \alpha_0}{\alpha_1}, \tag{4.22}$$

Noticing that α_0 and α_1 are unknown, their posterior distributions must be used. The following algorithm shows how to proceed:

```
For $s = 1$ to $n^*$
    1. sample $(\beta_0^{(s)}, \beta_1^{(s)}, \tau_0^{(s)}, \tau_1^{(s)}, \tau_\varepsilon^{(s)})$ from $p(\beta_0, \beta_1, \tau_0, \tau_1, \tau_\varepsilon \mid \text{data})$,
    2. sample $\tilde{\alpha}_0^{(s)}$ from $N(\beta_0^{(s)}, 1/\tau_0^{(s)})$,
    3. sample $\tilde{\alpha}_1^{(s)}$ from $N(\beta_1^{(s)}, 1/\tau_1^{(s)})$,
    4. sample $\tilde{x}^{(s)} \mid \tilde{y}$ from $N(\frac{\tilde{y}-\tilde{\alpha}_0^{(s)}}{\tilde{\alpha}_1^{(s)}}, 1/\tau_\varepsilon^{(s)})$.
End
```

This predictive distribution can then be used to assess the calibration quality in a predictive fashion that takes into account the additional variability due to the series. Some examples of applications are the following: the derivation of the dosing range of the analytical method (the concentration for which the method will perform in a satisfactory manner), the illustration of the predictive uncertainty using precision profiles of the method, or the illustration of the risk not being within some predefined specifications. This is illustrated in Chapter 10.

4.4 Bayesian mixed-effects non-linear regression

When departing from the linearity of the calibration curve, two options are available. The first one is to reduce the concentration range so that the data looks locally linear, and apply the theory of Section 4.3. The second one is to select a model able to account for this non-linearity. The first option is of course rather disappointing due to the non ability to use all the possible data, i.e. every concentration level that could be recorded by the method. Fortunately, there is generally simple non-linear structures that can be adopted to model such data.

In this section, the focus is put on a particular bio-analytical method, the ligand-binding assay with serial dilution, used to quantify products such as proteins by the mean of biological properties. Without going into details, a heterogenous variance is common. In this case, Davidian and Giltinan (1995) proposed to use a model that accounts for this heterogeneity through a variance that is proportional to a power of the expected predicted responses, often referred as *power of the mean* (POM) variance.

With the problem of (inverse) prediction, some constraints will apply on the prior distribution of the parameters to obtain a predictive density of the signals (the concentration) that shares enviable properties, such as positivity and computability of the POM variance.

As an example, a mixed-effects four parameters logistic regression model with a POM variance is described, with some effects assumed fixed and some other that are random. As in the previous sections, the data are organized in m series.

4.4.1 Model

In the condition of assumed known concentrations $\mathbf{x} = (x_1, ..., x_i, ..., x_n)$, let the response Y be explained by the following logistic model:

$$y_{ij} = \alpha_{j1} + \frac{\beta_2 - \alpha_{j1}}{1 + (\frac{x_i}{\alpha_{j4}})^{\alpha_{j3}}} + \varepsilon_{ij} \text{ with } j = 1, ..., m. \tag{4.23}$$

The error is modeled as $\varepsilon_{ij} \sim N(0, 1/\tau_Y)$, with the following model for the variance, expressed in term of precision:

$$\tau_Y = \frac{\tau}{E[y_i]^\theta}.$$

Let define the set of fixed parameters $\boldsymbol{\beta} = (\beta_2, \tau, \theta)$ and the set of random parameters $\boldsymbol{\alpha}_j = (\alpha_{j1}, \alpha_{j3}, \alpha_{j4})$. For the interpretation, α_{j1} and β_2 are the top (for the j^{th} serie) and bottom asymptotes. The bottom asymptote has been chosen as fixed as its value should not depend upon the series. It is then estimated for all the series. This assumption states that a zero concentration sample should always result in a zero signal, whatever the series. Next, α_{j3} is the slope of the curve at the inflection point (C50) and C50 is modeled by α_{j4}. τ can be seen as an averaged precision, observed for the possible values of y_i. θ is used to inflate or deflate the precision with respect to the expected values of y_i. The choice to fit a parameter as random or fixed can also be driven by some predictive indices such as the Deviance Information Criterion (DIC) or other predictive checks.

As in the previous section, the number of observations for the j^{th} serie is n_j, and the total number of observations is $n = \sum_{j=1}^{m} n_j$.

The model also supposes the following independent distributions for the random variable of the calibration curve:

$$\begin{aligned} &\alpha_{j1} \sim \log N(\beta_1, 1/\tau_1), \\ &\alpha_{j3} \sim N(\beta_3, 1/\tau_3), \\ &\alpha_{j4} \sim \log N(\beta_4, 1/\tau_4) \end{aligned} \tag{4.24}$$

Finally, let's define $\boldsymbol{\zeta} = (\beta_1, \tau_1, \beta_3, \tau_3, \beta_4, \tau_4)$. $\boldsymbol{\zeta}$ is often referred as the set of hyper-parameters for the random parameters. These hyper-parameters explain how the series affect the values of $\boldsymbol{\alpha}$. $p(\boldsymbol{\alpha} \mid \boldsymbol{\zeta})$ is then the conditional density of the mixed variables.

4.4.2 Likelihood

Summarizing the model in a general form, one gets:

$$y_{ij} = m(x_i; \boldsymbol{\beta}, \boldsymbol{\alpha}_j \mid \boldsymbol{\zeta}) + \varepsilon_{ij}, \quad \text{with } \varepsilon_{ij} \sim N(0, 1/\tau_Y) \text{ and } j = 1, ..., m. \tag{4.25}$$

Conditionally to the parameters, each observation is distributed as:

$$y_{ij} \mid x_i, \boldsymbol{\beta}, \boldsymbol{\alpha}_j, \boldsymbol{\zeta} \sim N(m(x_i, \boldsymbol{\beta}, \boldsymbol{\alpha}_j \mid \boldsymbol{\zeta}), 1/\tau_Y), \tag{4.26}$$

Assuming the observations identically distributed, the likelihood is the product of the marginal densities of the observations conditional to the parameters. It can then be written similarly to the likelihood of the hierarchical model of the previous section.

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\zeta} \mid \mathbf{y}) = \prod_{j=1}^{m} \int_{\boldsymbol{\alpha}_j} p(\mathbf{y} \mid \boldsymbol{\beta}, \boldsymbol{\alpha}_j, \boldsymbol{\zeta}).p(\boldsymbol{\alpha}_j \mid \boldsymbol{\zeta}).d\boldsymbol{\alpha}_j. \tag{4.27}$$

In this application, $p(\boldsymbol{\alpha}_j \mid \boldsymbol{\zeta})$ is the product of the densities of Equation (4.24).

4.4.3 Prior distribution of the parameters

The regression parameters $(\alpha_{1j}, \beta_2, \alpha_{3j}, \alpha_{3j})$ of the logistic curve and the variance parameters are assumed independent *a priori*. This restriction is rather useful as all the proposed distributions are not Normal nor identical so it is difficult to extend them to multivariate distributions. The parameters for the bottom asymptote, top asymptotes and for the C50 (i.e. $\alpha_{j1}, \beta_2, \alpha_{j4}$) are assumed log-Normal to ensure their positivity. The support domain of the slope parameter (β_{j3}) is assumed to be $\mathbb{R}$. This may be controversial (see the discussion of Hill in Hunter and Lamboy, 1981), but it provides a more general model, for instance to be able to fit some data that shows a decreasing curve. In this case, β_{j3} could be negative.

The prior distribution of the fixed parameters $\boldsymbol{\beta}$ are as follows:

$$\begin{aligned} p(\beta_2) &= \log N(b_2, \tau_2) \\ p(\tau) &= Gamma(a, b), \\ p(\theta) &= Gamma(c, d), \qquad (4.28) \\ \text{then, } p(\boldsymbol{\beta}) &= p(\beta_2).p(\tau).p(\theta). \qquad (4.29) \end{aligned}$$

With these prior assumptions, τ and θ are restricted to positive value due to their Gamma prior. For vague *a priori*, the following values have been chosen:

$$\begin{aligned} b_2 &= 0, \\ \tau_2 &= 0.00001, \\ a, b, c, d &= 0.00001. \end{aligned}$$

For the random parameters $\boldsymbol{\alpha}$, *second-stage* priors are defined for the hyper-parameters $\boldsymbol{\zeta}$. They describe the distributions for their locations and scales, and are assumed independent.

$$\begin{aligned} p(\beta_1) &= N(\mu_{\beta_1}, \tau_{\beta_1}), \\ p(\tau_1) &= Gamma(a_1, b_1), \\ p(\beta_3) &= N(\mu_{\beta_3}, \tau_{\beta_3}), \\ p(\tau_3) &= Gamma(a_3, b_3), \\ p(\beta_4) &= N(\mu_{\beta_4}, \tau_{\beta_4}), \\ p(\tau_4) &= Gamma(a_4, b_4), \qquad (4.30) \\ \text{then } p(\boldsymbol{\zeta}) &= p(\beta_1).p(\tau_1).p(\beta_3).p(\tau_3).p(\beta_4).p(\tau_4). \end{aligned}$$

Vague priors can be given as follows:

$$\begin{aligned} \mu_{\beta_1}, \mu_{\beta_3}, \mu_{\beta_4} &= 0, \\ \tau_{\beta_1}, \tau_{\beta_3}, \tau_{\beta_4} &= 0.00001, \\ a_1, b_1, a_3, b_3, a_4, b_4 &= 0.00001. \end{aligned}$$

Informative prior distributions can be given in the same way, using estimates that come from previous experiments, if any.

4.4.4 (Sampled) posterior distribution of the parameters

The posterior density of the parameters $(\boldsymbol{\beta}, \boldsymbol{\zeta})$ is obtained using Bayes' theorem

$$p(\boldsymbol{\beta}, \boldsymbol{\zeta} \mid \text{data}) \propto \mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\zeta} \mid \mathbf{y}).p(\boldsymbol{\beta}).p(\boldsymbol{\zeta}) \tag{4.31}$$

Unfortunately, there is no analytical solutions to solve Equation 4.31. Instead of relying on complex approximations, MCMC simulations are used to obtain samples from the joint posterior distribution of the parameters. Listing 4.3 is a BUGS code implementing the model.

To implement the likelihood, BUGS coding allows making use of the variables $\boldsymbol{\alpha}_j$. This allows easily defining the likelihood without having to solve the multiple integral in (4.27). WinBUGS carries out the involved product of the marginal likelihood over the series.

Listing 4.3: Winbugs code for the mixed four parameters logistic regression with a variance proportional to a power of the mean (unbalanced case).

```
model{
for(i in 1:n){
  response[i] ~ dnorm(expect[i],powertau[i])

  #POM variance
  powertau[i] <- 1/(exp(theta*log(expect[i]))*sigma2.y)
  #mean
  expect[i]<-alpha1[series[i]]    +
     ( (beta2-alpha1[series[i]]) /
        (1+exp((log(x[i]/alpha4[series[i]]))*alpha3[series[i]]))    )
  }

for(j in 1:m){
  alpha1[j] ~ dlnorm(beta1,tau1)
  alpha3[j] ~ dnorm(beta3,tau3)
  alpha4[j] ~ dlnorm(beta4,tau4)
  }

#convert precision into var
sigma2.y <- 1/tau.Y

#priors on fixed parameters
tau.Y ~ dgamma(a,b)
theta ~ dgamma(c,d)
beta2 ~ dlnorm(b2,tau2)

#hyper-priors on mixed parameters
beta1 ~ dnorm(mubeta1,taubeta1)
tau1 ~ dgamma(a1,b1)
beta3 ~ dnorm(mubeta3,taubeta3)
tau3 ~ dgamma(a3,b2)
beta4 ~ dnorm(mubeta4,taubeta4)
tau4 ~ dgamma(a4,b4)
}
```

During MCMC simulations, the independence assumption might give poor sampling acceptance (high rejection of samples) with the Metropolis-Hasting algorithm as it will not be able to adapt the proposal to account for the dependencies. As a consequence, the sampler might slowly explore the posterior density, and can get stuck into local sub-optimal zones. The obtained chains of samples might also be strongly autocorrelated. A common solution, though computationally intensive, is to draw (many) more samples and to thin them.

In the listing, the unbalanced data are handled in a different way than in the previous sections. A vector series of size n was defined to identify the series for every observation i. This vector provides the index of the series for the variables that incorporate the series effect $(\boldsymbol{\alpha}_j)$.

It must be noticed that all the exponents including random variables were equivalently rewritten using the canceling exponentials property (i.e. $\exp(\log(x)) = x$). WinBUGS has been found unable to handle them properly. This permits using the power rule of the log operator, $log(x^d) = d \log(x)$, to get rid of the exponents that include variables.

This notation makes also clear why positiveness is essential for certain parameters under log operator. The POM variance constraints the mean sampled predicted values (expect[i]) to $\mathbb{R}^+$ such that the log operator is properly defined. The same fact is observed about the concentrations x[i] and the parameter alpha4[j]. To observe the positivity of the mean predicted, alpha1[j] must also be strictly higher than beta2, i.e. the calibration curve must be increasing. If expect[i] was defined over $\mathbb{R}$, one could use its absolute value when computing the POM variance, or use another model for the variance.

4.4.5 (Sampled) predictive distribution of one new response

Given a new concentration $\tilde{x}$, the predictive distribution of the signal is computed as

$$\begin{aligned}
&p(\tilde{y} \mid \tilde{x}, \text{data}) \\
&= \int_{(\boldsymbol{\beta},\boldsymbol{\zeta})} p(\tilde{y} \mid \tilde{x}, \boldsymbol{\beta}, \boldsymbol{\zeta}).p(\boldsymbol{\beta}, \boldsymbol{\zeta} \mid \text{data}).d(\boldsymbol{\beta}, \boldsymbol{\zeta}) \\
&= \int_{(\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{\zeta})} p(\tilde{y} \mid \tilde{x}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\zeta}).p(\boldsymbol{\alpha} \mid \boldsymbol{\zeta}).p(\boldsymbol{\beta}, \boldsymbol{\zeta} \mid \text{data}).d(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\zeta}). \qquad (4.32)
\end{aligned}$$

Using simulations, samples can be drawn from the predictive distribution of Y using the following algorithm.

$s = 1$ to n^*

1. Sample hyper-parameters $\boldsymbol{\zeta}^{(s)}$ from $p(\boldsymbol{\zeta} \mid \text{data})$,

2. Sample random parameters $\boldsymbol{\alpha}^{(s)} \mid \boldsymbol{\zeta}^{(s)} = (\tilde{\alpha}_1^{(s)}, \tilde{\alpha}_3^{(s)}, \tilde{\alpha}_4^{(s)})$ from $p(\boldsymbol{\alpha} \mid \boldsymbol{\zeta}^{(s)})$,

```
    3. Sample fixed parameters $\boldsymbol{\beta}^{(s)} = (\beta_2^{(s)}, \tau^{(s)}, \theta^{(s)})$ from $p(\boldsymbol{\beta} \mid \text{data})$,
    4. Compute mean $\tilde{\mu}_y^{(s)} = \tilde{\alpha}_1^{(s)} + \frac{\beta_2^{(s)} - \tilde{\alpha}_1^{(s)}}{1+(\frac{\tilde{x}}{\tilde{\alpha}_4^{(s)}})^{\tilde{\alpha}_3^{(s)}}}$ and precision $\tau_{\tilde{\mu}_Y}^{(s)} = \frac{\tau^{(s)}}{(\tilde{\mu}_y^{(s)})^{\theta^{(s)}}}$,
    5. Sample $(\tilde{y}^{(s)} \mid \tilde{x}, \text{data})$ from $N(\tilde{\mu}_y^{(s)}, \tau_{\tilde{\mu}_Y}^{(s)})$.
end
```

In step 1 and 3, the sampling from the posterior distribution of the parameters is done. Practically, these two steps are carried out in the same time as a draw from the joint posterior distribution.

4.4.6 (Sampled) distribution of the inverse prediction of one new result

When using the bio-analytical method during routine, the inverse calibration curve is used to translate an observed signal $\tilde{y}$ into the concentration of the corresponding sample. Assuming the parameters known, the equation of the inverse curve can be obtained from Equation (4.23) and is

$$\tilde{x} = \alpha_4 (\frac{\beta_2 - \tilde{y}}{\tilde{y} - \alpha_1})^{\frac{1}{\alpha_3}}. \tag{4.33}$$

As the parameters are not known, their (joint) posterior distribution is used to propagate their uncertainty to the inverse prediction. The following algorithm can be used to draw samples from the distribution of the inverse prediction.

```
$s = 1$ to $n^*$
    1. Sample hyper-parameters $\boldsymbol{\zeta}^{(s)}$ from $p(\boldsymbol{\zeta} \mid \text{data})$,
    2. Sample random parameters $\boldsymbol{\alpha}^{(s)} \mid \boldsymbol{\zeta}^{(s)} = (\tilde{\alpha}_1^{(s)}, \tilde{\alpha}_3^{(s)}, \tilde{\alpha}_4^{(s)})$ from $p(\boldsymbol{\alpha} \mid \boldsymbol{\zeta}^{(s)})$,
    3. Sample fixed parameters $\boldsymbol{\beta}^{(s)} = (\beta_2^{(s)}, \tau^{(s)}, \theta^{(s)})$ from $p(\boldsymbol{\beta} \mid \text{data})$,
    4. Compute precision $\tau_{\tilde{\mu}_Y}^{(s)} = \frac{\tau^{(s)}}{\tilde{y}^{\theta^{(s)}}}$,
    5. Sample $\tilde{y}^{(s)}$ from $N(\tilde{y}, \tau_{\tilde{\mu}_Y}^{(s)})$,
    6. Compute the inverse prediction $(\tilde{x}^{(s)} \mid \tilde{y}^{(s)}, \text{data}) = \tilde{\alpha}_4^{(s)} (\frac{\beta_2^{(s)} - \tilde{y}^{(s)}}{\tilde{y}^{(s)} - \tilde{\alpha}_1^{(s)}})^{\frac{1}{\tilde{\alpha}_3^{(s)}}}$.
end
```

4.5 Conclusion

In this chapter, different types of mixed models have been reviewed. The Bayesian analysis has been shown useful to easily obtain a predictive distribution for the response of interest. The simulation approach is rather systematic and can be used with small or complex models. In that simulation perspective, enough samples must be available to compute statistics such as a mean, an interval, a probability of acceptance, etc. The sampling size and the considerations of Appendix B for Monte-Carlo simulations then apply. Fortunately, modern computers are able to deal with long simulations and complex models within reasonable run time.

The distributions of the prediction and inverse prediction are of particular interest as often, the parameters estimates and their distribution give poor information about the model and the method quality. Instead, the predictive distribution provides a better insight on such qualities. It provides the necessary information towards the objective of the method, i.e. to know if it is capable to perform accurately, within known and well-defined specifications.

Chapter 5

Bayesian predictive multicriteria decision method

Introduction

Design Space (DS) is directly related to multi-criteria optimization (MCO) and multi-criteria decision making (MCDM). Regarding the factors influencing a process or a method, several Critical Process Parameters (CPP) are selected to analyze their concurrent effects on several outputs. Regarding the outputs, they are summarized by Critical Quality Attributes (CQA) that should show satisfactory properties concurrently. Design of Experiments (DoE) methodologies provide powerful tools to carry out such tasks.

In Chapter 1, the DS was presented as a risk-based tool that is built over DoE to jointly assess various CQAs in a predictive fashion. Chapters 2 – 4 showed some developments to obtain predictive solutions for various classes of statistical models.

In classical DoE context, the CQAs are random variables $\boldsymbol{Y} = (Y_1, ..., Y_j, ..., Y_m) \in \mathbb{R}^m$ that are the responses of a designed experiment defined over the domains of some CPPs, noted $\mathbf{x} = (x_1, ..., x_k)$. Changing these CPPs induces changes to each $Y_j \mid \mathbf{x}$ following a model f_j whose parameters $\boldsymbol{\theta}_j$ are estimated using the experimental data.

$$Y_j = f_j(\mathbf{x}; \boldsymbol{\theta}_j) + \varepsilon_j \tag{5.1}$$

Classically, the CQAs are considered independent and are often modeled separately. In addition, the j^{th} estimated model f_j is often assumed to be the *true* link function between the j^{th} CQA and the CPPs. The very basic optimization process looks for the CPP setting $\mathbf{x}^*$ such that the mean predicted responses $\hat{E}[Y_j \mid \mathbf{x}^*] = f_j(\mathbf{x}^*, \hat{\boldsymbol{\theta}}_j)$ are optimal jointly. Of course, it is unlikely that a unique condition would be optimal for every CQA. MCO methodologies should then be envisaged. The flaws of using mean responses have been presented in Chapter 1 but will be explained again hereafter.

In this chapter, the particular statistical models f_j are of limited interest. To be as general as possible, the CQAs are assumed to be either the multivariate response vector $\boldsymbol{Y}$, either transformations of these responses, say $\boldsymbol{Y}^* = O(\boldsymbol{Y})$. Simple transformations include log, logit or identity functions. More complex transformations include various combinations of responses, possibly with non continuous or non derivable operator such as min or max. In the chromatographic field, the *resolution* or the *separation* of the critical pair of peaks are good examples of combinations with discontinuities (see Chapter 7, page 120). To simplify, $\boldsymbol{Y}$ is assumed to be the vector containing the (possibly transformed) CQAs.

Structure of the chapter

First, Section 5.1 briefly explains some problems encountered with classical approaches based on mean responses. The issues are related to the uncertainty of the statistical models that is generally not taken into account, providing limited evidences about the predictions.

Next, Section 5.2 presents the classical solution for MCO in case where the CQAs are not optimal jointly at a unique operating condition. This is known as the *competing responses* problem. To answer this problem, methodologies have been developed to optimize one unitless index. It aggregates the different CQAs and indicates if the process performs well or not at a given operating condition, according to predefined targets to achieve. This is the well-known *desirability* concept: basically, each CQA is first transformed into a desirability index using a desirability function (DF). After, the indexes are aggregated into a global desirability index, reflecting how close the output is from the ideal objective.

However, the desirability approach can't be used directly for deriving a DS. The reasons are the following : first, it has been recognized that the use of mean predicted responses to compute a mean global desirability index often gives biased results (Steuer, 2000). Second, as with the Sweet Spot, using mean predictions does not provide any clue about the reliability of the solution. Recent developments of MCO tends to give a more important role to the uncertainty that is encountered in the statistical models, since uncertainty is crucial to assess risk. Some of them are presented in Section 5.3.

In Section 5.4, the desirability methodologies are linked to the risk-based approach presented in Chapter 1. It is shown how the risk-based DS approach is a particular case of the more flexible method based on desirability functions.

5.1 Classical vs. Bayesian methodologies

For years, the standard analysis of data coming from DoE has been done using the mean predicted CQAs, $\hat{E}[Y_j \mid \mathbf{x}]$. One reason is that most of the available statistical softwares having an MCO module (SAS/JMP, Statistica, Design Expert, Minitab, etc.) rely on mean responses only. To allow for a certain compliance with Quality by Design (QbD) approaches, the most sophisticated of them are able to compare the overlapped mean responses to some specifications, allowing the construction of some Sweet Spots or to compute a global optimum based on desirability functions (Harrington, 1965). Generally, the study of CQAs stops at this point.

In Chapter 1, Sweet Spot computation has been described as non sufficient: it does not take into account the model uncertainties and CQAs dependencies. As a consequence, it does not provide any indication about how well and how often the process or method can meet the specifications with respect to the investigated CQAs (Peterson, 2008). Thus solutions within the Sweet Spot are subject to give disappointing results for the future use of the process or method. In the context of the ICH Q8 (2009) guideline, this represents a major drawback since it is clearly asking to analyze the guarantees of quality, and not only a quality that would be observed *on average*. Clearly, desirability approaches based on mean responses suffer from the same problems.

Contrasting with the overlapped mean response approach, a Bayesian predictive approach to define the DS is able to take into account the parameters and CQAs uncertainties and their dependencies (Castagnoli et al., 2010; Lebrun et al., 2012b). This approach integrates these variabilities using the predictive multivariate error associated to the responses predictions. As the data and model parameters uncertainties are accounted, the Bayesian predictive approach may significantly improve the ability of models to provide predictions. This is certainly one of the most efficient way for "demonstrating assurance of quality" as requested by the ICH Q8 definition of the DS.

For the next parts of this chapter, a new observation of the CQAs $\boldsymbol{Y}$ observed at a CPP setting $\tilde{\mathbf{x}}$ will be noted $\tilde{\mathbf{y}}$. $\tilde{\mathbf{y}}$ is then a sample of $(\boldsymbol{Y} \mid \tilde{\mathbf{x}}, \text{data})$.

5.2 Classical multi-criteria methodologies using desirability functions

A flexible methodology for MCO has been proposed by Harrington (1965) to optimize only one value that aggregates together the relevant CQAs. In this way, the MCO is simplified in an univariate problem. Harrington based his methodology on two steps. First, the CQAs are scaled into a unitless value between 0 and 1, namely the desirability indices, using desirability functions that are parametrized according to each CQA and its specifications. A CQA with a desirability of 1 has the perfect achievement of quality. A desirability of 0 indicates a totally undesirable value for the CQA. Second, Harrington proposed to aggregate together the desirability indices into a global desirability index using a weighted geometric mean.

5.2.1 Desirability functions

Definition. For a CQA Y, let any function $d : \mathbb{R} \to [0, 1]$, $Y \mapsto d(Y)$, be a desirability function (DF). A desirability of $d(Y) = 0$ is considered a totally undesirable value for Y while a desirability of $d(Y) = 1$ represents a perfectly desirable value.

Three types of DFs are generally encountered, for the cases where Y must be maximized, minimized or must reach a target value. In the literature, various authors have proposed different implementations of the DFs, with the objective to provide functions that are flexible, understandable, or mathematically “well-behaved” (Harrington, 1965; Derringer and Suich, 1980; Gibb et al., 2001; Le Bailly de Tilleghem and Govaerts, 2005b).

5.2.2 Global desirability index

To reduce the MCO problem into a simpler univariate problem, the individual DFs, $d_1(Y_1), ..., d_m(Y_m)$, are aggregated into a global desirability index noted $D(\boldsymbol{Y})$. In general, a weighted geometric mean is used to accomplish the task, although a (weighted) arithmetic mean, a (weighted) harmonic mean or the minimum value between all individual DFs could also be envisaged. For m CQAs combined with a weighted geometric mean, $D(\boldsymbol{Y})$ is computed as:

$$D(\boldsymbol{Y}) = D(d_1(Y_1), ..., d_m(Y_m)) = \prod_{j=1}^{m} (d_j(Y_j))^{w_j} \text{ with } \sum_{j=1}^{m} w_j = 1, \tag{5.2}$$

where w_j values are fixed by an expert according to the relative importance he wants to give to each CQA in the global desirability index. Flexibility is the main advantage of this approach, while the loss of interpretation of each CQA within $D(\boldsymbol{Y})$ is the corresponding drawback.

5.2.3 Optimization

Classically, the optimization is done using the mean predicted values of the CQAs $\hat{E}[\boldsymbol{Y} \mid \mathbf{x}_0]$. Then, the optimal operating condition $\mathbf{x}^*$ is defined as:

$$\begin{aligned}\mathbf{x}^* = \underset{\tilde{\mathbf{x}} \in \chi}{\operatorname{argmax}}\, D(\hat{E}[\boldsymbol{Y} \mid \tilde{\mathbf{x}}]) &= \underset{\tilde{\mathbf{x}} \in \chi}{\operatorname{argmax}} \prod_{j=1}^{m} (d_j(\hat{E}[Y_j \mid \mathbf{x}_0]))^{w_j} \\ &= \underset{\tilde{\mathbf{x}} \in \chi}{\operatorname{argmax}} \prod_{j=1}^{m} (d_j(f_j(\mathbf{x}_0; \hat{\boldsymbol{\theta}}_j)))^{w_j},\end{aligned} \tag{5.3}$$

where χ is the experimental domain covered by the experiments.

As for any optimization problem, several algorithms exist. Grid search is advocated when dimensionality is small and when local optima might prevent to obtain the best solution. It also gives a global map of the evolution of the global desirability index over the experimental domain. Other methods might be used as well, such as gradient descend, simplex algorithm, or simulated annealing.

5.3 Improvements of the global desirability index

When optimal conditions are derived from statistical model predictions, it is fundamental to study the impact of the model prediction error on the reliability of the solution found. Unfortunately, this is rarely underlined in the literature. It has been discussed firstly by Steuer (2000). The obvious but hardly applied conclusion of this study is that the global desirability index must be taken as a random variable, and its full distribution must be used to find its expected value and other statistics. Steuer also noted that the classical use of mean predicted DFs gives biased results as most implementations of DFs are non linear. The technique of Steuer is to propagate the uncertainty of each predicted CQA (assumed known and Normally distributed, $\varepsilon_j \sim N(0, \sigma_j^2)$) using Monte-Carlo simulations. Thus, the maximization problem is modified to:

$$\mathbf{x}^* = \underset{\tilde{\mathbf{x}} \in \chi}{\operatorname{argmax}}\, \hat{E}[D(\boldsymbol{Y} \mid \tilde{\mathbf{x}})] = \underset{\tilde{\mathbf{x}} \in \chi}{\operatorname{argmax}}\, \hat{E}[\prod_{j=1}^{m} d_j(f_j(\tilde{\mathbf{x}}; \hat{\boldsymbol{\theta}}_j) + \varepsilon_j)^{w_j}] \tag{5.4}$$

The optimum is the one that maximize $\hat{E}[D(\boldsymbol{Y})]$, instead of the one that maximize the global desirability index computed from expected (mean predicted) CQAs: $D(\hat{E}[\boldsymbol{Y}])$.

To avoid the use of intensive Monte-Carlo simulations, further works have been done by Weber and Weihs (2003) and Trautmann and Weihs (2004, 2006), that discussed the impact of the model prediction error on the reliability of the optimal solution. They derived an approximation of the cumulative distribution function of $D(\boldsymbol{Y})$ in a bi-variate MCO (two responses to optimize jointly). The complexity of the density function of $D(\boldsymbol{Y})$ is however already high in this simple case. These improvements showed the way to better understand and use the desirability methods. Still, they do not provide a solution for correlated CQAs.

Le Bailly de Tilleghem and Govaerts (2005a,b); Le Bailly de Tilleghem (2007) also made progress in this field, firstly defining a new class of DFs, based on the Normal cumulative distribution function. As the DFs are continuous and strictly increasing/decreasing over their domain ($\mathbb{R}$), the Pareto optimality is guaranteed, unlike with other types of DFs. Their great flexibility and mathematical properties allow propagating analytically the uncertainty of regression parameters and the experimental error to $D(\boldsymbol{Y})$. This gave the way to define the classical prediction intervals for $D(\boldsymbol{Y})$ and an indistinguishable optimal zone in the space of factors, where one can not say it is significantly different than its optimal level. This work has also been extended to correlated models residuals.

Two gaps are identified on these improved approaches. First, when uncertainty is assessed, it is not based on the joint predictive distribution of $\tilde{\mathbf{y}}$. For instance, Le Bailly de Tilleghem and Govaerts (2005b) analyzed the variability of the global desirability index prediction using the 95% classical prediction intervals. However, to predict the behavior of individual new observations, Aitchison (1964) and Aitchison and Dunsmore (1975) advocated the generalization of the use of tolerance intervals. In this complex situation implying desirability functions, it is probably hopeless to succeed in deriving such intervals. The second gap is that none of the proposed solutions are risk-based.

In the next sections, the solution proposed by Steuer is envisaged using a more realistic distributional assumption on the CQAs. This would allow extending the methodology so that it is compliant with the ICH Q8 guideline and the Quality by Design concept. In this setting, Peterson (2004) successfully illustrated the use of the joint predictive distribution of CQAs to compute the predictive distribution of the Derringer's DFs and the global desirability index using Monte-Carlo simulations. However, the decision on the optimal solution is based on rather artificial specifications given on the global desirability index. Specifications are only used to define the parameters of the DFs, as proposed by Harrington (1965) and later by Derringer and Suich (1980).

5.4 Monte-Carlo simulations for MCDM

When envisaging the computation of a Design Space, Monte-Carlo simulations provide a generic framework to provide estimations of the guarantee to observe some CQAs within specifications. The next subsections illustrate the problem with a classical approach based on joint probabilities and with the desirability method.

The distribution of the improved global desirability index could be analytically obtained in some cases, with or without some approximations. However, Monte-Carlo (MC) simulations represent a great alternative as the solution is applicable in any situations, e.g. when CQAs are some discontinuous functions of some combinations of the responses.

Given that it is possible to obtain samples from the joint posterior predictive distribution of the CQAs, MCO can be simply envisaged using MC simulations. See Chapters 2–4 and Appendices A–C for generic and applied derivations of such predictive distribution.

Without a desirability function

For the classical risk-based approach not relying on desirability functions, the DS was defined as follows:

$$\text{Design Space} = \{\tilde{\mathbf{x}} \in \chi \mid P(\boldsymbol{Y} \in \boldsymbol{\Lambda} \mid \tilde{\mathbf{x}}, \text{data}) \geq \pi\}\,. \tag{5.5}$$

In other words, the DS is a region of the experimental domain χ where the posterior probability that the CQAs are within specifications $\boldsymbol{\Lambda}$, is higher than a specified quality level π. This posterior probability includes the uncertainty of the responses and of the parameters $\boldsymbol{\theta}$ and is defined conditionally to the observed data and prior knowledge. The expected posterior probability can be estimated through MC simulations. If $\tilde{\mathbf{y}}^{(s)}$ is a sample from $(\boldsymbol{Y} \mid \tilde{\mathbf{x}}, \text{data})$, it is:

$$P(\boldsymbol{Y} \in \boldsymbol{\Lambda} \mid \tilde{\mathbf{x}}, \text{data}) \approx \frac{1}{n^*} \sum_{s=1}^{n^*} I(\tilde{\mathbf{y}}^{(s)} \in \boldsymbol{\Lambda}), \tag{5.6}$$

where $I(A)$ is 1 if A is true, 0 otherwise; $\boldsymbol{\Lambda}$ is the set of specifications, e.g. $\{\mathbf{y} \mid \lambda_{jl} \leq y_j \leq \lambda_{ju},\ \ j = 1, ..., m\}$. Provided that n^* is high enough, this MC simulation can provide an accurate estimate of the acceptance probability. One-sided decision is obtained with λ_{jl} or λ_{ju} set to $-\infty$ or $+\infty$, respectively. More information about MC simulations can be find in Appendix B.

Figure 5.1 illustrates an example of distribution from the chromatographic world. Two CQAs are analyzed: the chromatographic run time (Y_1) and the minimal sepa-

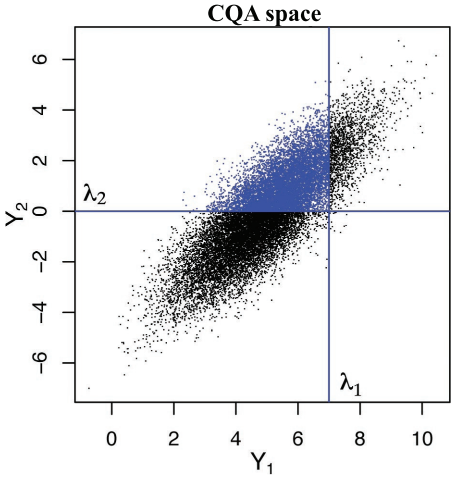


Figure 5.1: Predictive distribution of the CQAs $\boldsymbol{Y} = (Y_1, Y_2)$ at CPPs=$\tilde{\mathbf{x}}$, with λ_1 and λ_2, some specifications. The proportion of blue points is the MC estimate of the expected probability.

ration (Y_2), when predicted at CPPs=$\tilde{\mathbf{x}}$. A clear positive correlation was observed. Logically, when the run time is increased, there is more room to observe an higher separation between analytes. In blue are two one-sided specifications $\lambda_1 u$ and $\lambda_2 l$. Specifications are $\boldsymbol{\Lambda} = \{\mathbf{y} \mid y_1 \leq \lambda_{1u} = 7,\ y_2 \geq \lambda_{2l} = 0\}$. An MC simulation on $n^* = 20000$ samples drawn from the joint posterior predictive distribution of the two CQAs was envisaged.

The proportion of blue points is the MC estimate of $P((Y_1, Y_2) \in \boldsymbol{\Lambda} \mid \tilde{\mathbf{x}}, \text{data})$. The joint predictive probability to achieve simultaneously both limits is computed using Equation (5.6):

$$P((Y_1, Y_2) \in \boldsymbol{\Lambda} \mid \tilde{\mathbf{x}}, \text{data}) \approx 0.42. \tag{5.7}$$

The marginal expected posterior probabilities were also computed for both specifications separately and were $P(Y_1 \leq 7 \mid \tilde{\mathbf{x}}, \text{data}) = 0.92$ and $P(Y_2 \geq 0 \mid \tilde{\mathbf{x}}, \text{data}) = 0.49$. The joint probability is lower than the marginal ones because logically, two specifications are always harder to achieve than one, and also because the correlation structure between the CQAs is not favorable.

Applying the MC simulations on every points of a grid created over χ would allow recording, for every operating condition, the expected posterior probability of achievement of the specifications. Eventually, over this grid, the operating conditions where the probability is higher than the minimal quality level π could be identified. This set of operating conditions is the risk-based DS. It is also possible to make a complete optimization of the process and to chose the optimal operating condition $\mathbf{x}^*$:

$$\mathbf{x}^* = \underset{\tilde{\mathbf{x}} \in \chi}{\operatorname{argmax}}\, P(\boldsymbol{Y} \in \boldsymbol{\Lambda} \mid \tilde{\mathbf{x}},\ \text{data}). \tag{5.8}$$

With desirability functions

From the predictive distribution, it is easy to transform each sample $\tilde{\mathbf{y}}^{(s)} = (\tilde{y}_1^{(s)}, ..., \tilde{y}_j^{(s)}, ..., \tilde{y}_m^{(s)})$ into the desirability space using m DFs $d_j(\tilde{y}_j^{(s)})$. After, the desirability indices are usually aggregated into a global desirability index $D(\tilde{\mathbf{y}}^{(s)})$ using the classical weighted geometric mean. Thus, obtaining samples from the predictive distribution of $D(\boldsymbol{Y} \mid \tilde{\mathbf{x}}, \text{data})$ is straightforward using Monte-Carlo simulations.

The approach presented in the next paragraphs extends the Normal error propagation method of Steuer (2000) or the more convincing posterior predictive approach of Peterson (2004). In this last reference, it has been noted that a rather artificial specification for the global desirability index was used. It was either propose to base the decision process on the recommendations of Harrington (1965) on the expected (mean) global desirability index, and to identify the DS using a percent change from optimal global desirability index. Here, it is proposed to use the real specifications $\boldsymbol{\Lambda}$ that apply on the CQAs to provide a risk-based approach more similar to the one presented in the previous section.

For the continuation of the example, the DFs of Le Bailly de Tilleghem and Govaerts (2005a) have been chosen. Other DFs' definitions might be used as well without restriction. These DFs are presented in Figure 5.2.

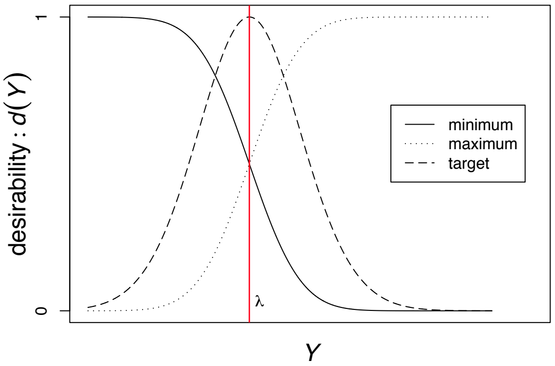


Figure 5.2: Desirability function of Le Bailly de Tilleghem and Govaerts (2005a). λ is a specification used to set up the desirability functions used in maximization and minimization problems.

.

The proposed methodology is explained for one-sided specification, as follows: for each CQA Y_j, the DF is centered on an acceptance limit λ_j (i.e., λ_{jl} for a CQA to maximize and λ_{ju} for a CQA to minimize). Next, the predictive distributions of the CQAs and the specifications are rescaled to the desirability space. Suppose that Y_1

ought to be minimized and Y_2 to be maximized. This gives:

$$\begin{aligned} d_1(\tilde{y}_1^{(s)}) &= 1 - \Phi(\frac{\tilde{y}_1^{(s)} - \lambda_1}{s_1}) \qquad && \text{and } d_1(\lambda_1) = 1 - \Phi(\frac{\lambda_1 - \lambda_1}{s_1}) = 0.5 \\ d_2(\tilde{y}_2^{(s)}) &= \Phi(\frac{\tilde{y}_1^{(s)} - \lambda_2}{s_2}) \qquad && \text{and } d_2(\lambda_2) = \Phi(\frac{\lambda_2 - \lambda_2}{s_2}) = 0.5 \end{aligned} \tag{5.9}$$

where s_1 and s_2 are parameters that allow changing the stiffness of the DF curves, and Φ is the cumulative distribution function of the standard Normal variable defined as:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} exp(\frac{-t^2}{2})dt. \tag{5.10}$$

Preferably, the parameters (λ_j, s_j) should depend on the objectives, and not on the data. For instance, Le Bailly de Tilleghem and Govaerts (2005a) proposed a computation based on lower and upper limits of interest for the CQAs. A manual tuning (related to the process to optimize) is also possible. In the example, λ_j were defined to be the objectives $(j = 1, 2)$. Thus, if a (mean) CQA barely achieves its specification so that $\hat{E}[Y_j] = \lambda_j$, its desirability would be 0.5. However, s_j was set up as the standard deviation of the observation of the CQA j, which may be non recommendable. For both CQAs, a desirability lower than 0.5 signifies a non achievement of the specification and a desirability higher than 0.5 represents a better achievement of quality.

Figure 5.3 continues the example of previous section. The sampled predictive distribution of the CQAs is presented on (A). Distribution and specifications were the same as in Figure 5.1. If the specifications are not too strict, a certain trade-off between the CQA quality is possible and the desirability methodology may be envisaged.

The joint distribution of $(d_1(\tilde{y}_1), d_2(\tilde{y}_2))$ is illustrated in Figure 5.3(B) with the univariate specifications $d_1(\lambda_1) = 0.5$ and $d_2(\lambda_2) = 0.5$ in the desirability space (blue). In the desirability space, every CQA has to be maximized. The red curve and red points are explained in the next paragraphs.

Next, the global desirability index is computed using the geometric mean with equal weights $w_j = 0.5$. It is done both on the CQA and on their specifications :

$$D(\tilde{\mathbf{y}}^{(s)}) = d_1(\tilde{y}_1^{(s)})^{0.5}.\ d_2(\tilde{y}_2^{(s)})^{0.5} \text{ and } D(\mathbf{\Lambda}) = d_1(\lambda_1)^{0.5}.\ d_2(\lambda_2)^{0.5} \tag{5.11}$$

A fitted density computed from the samples $D(\tilde{\mathbf{y}}^{(s)})$ of $D(\boldsymbol{Y} \mid \tilde{\mathbf{x}}, \text{data})$ is shown in Figure 5.3(C). The two specifications transformed into the global desirability space provide the univariate specification $D(\mathbf{\Lambda})$ (red line). In this case, $D(\mathbf{\Lambda}) = 0.5$. The portion of the density of $D(\boldsymbol{Y} \mid \tilde{\mathbf{x}}, \text{data})$ higher than $D(\mathbf{\Lambda})$ is also depicted in red.

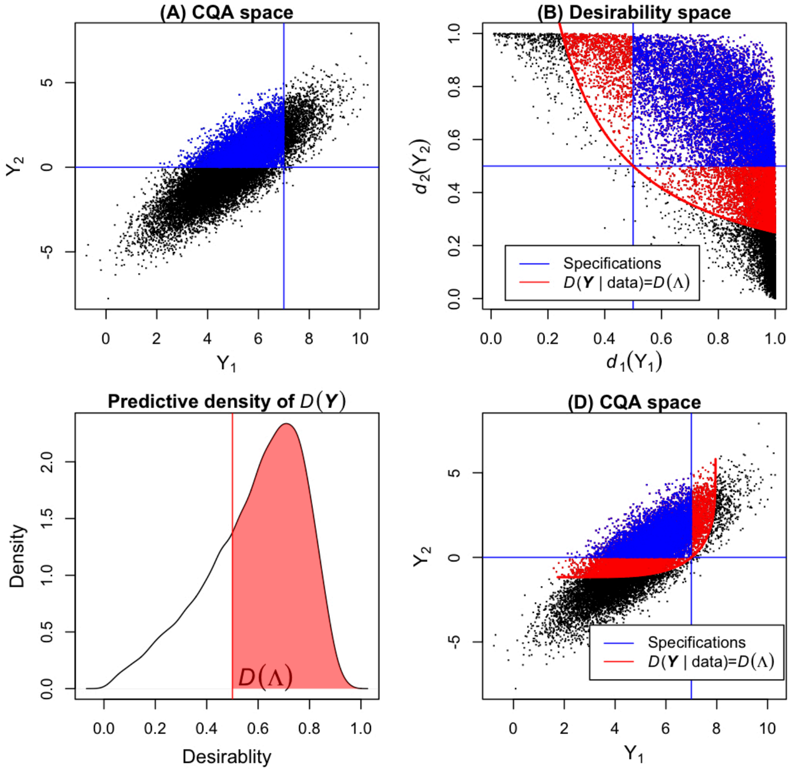


Figure 5.3: (A) : Predictive distribution of two CQAs ; (blue) samples within specifications. (B) predictive distribution scaled into the desirability space ; (blue) scaled specifications ; (red) samples with a global desirability index $D(\boldsymbol{Y} \mid \tilde{\mathbf{x}}, \text{data})$ higher than $D(\boldsymbol{\Lambda})$. (C) density of the $D(\boldsymbol{Y} \mid \tilde{\mathbf{x}}, \text{data})$; (red) proportion of sampled points such that $D(\boldsymbol{Y} \mid \tilde{\mathbf{x}}, \text{data}) \geq D(\boldsymbol{\Lambda})$. (D) predictive distribution of the two CQAs in their original space ; (red) points such that $D(\boldsymbol{Y} \mid \tilde{\mathbf{x}}, \text{data}) \geq D(\boldsymbol{\Lambda})$.

This portion is an expected probability of acceptance and may be computed via MC simulations:

$$P\left(D(\boldsymbol{Y}) \geq D(\boldsymbol{\Lambda}) \mid \tilde{\mathbf{x}}, \text{data}\right) = \frac{1}{n^*} \sum_{s=1}^{n^*} I\left(D(\tilde{\mathbf{y}})^{(s)} \geq D(\boldsymbol{\Lambda})\right) = 0.71. \qquad (5.12)$$

The higher predicted probability (0.71, compared to 0.42 in the previous section) is the result of the modification of the marginal univariate specifications into an acceptance border that integrates the trade-off between the CQAs.

This is illustrated in Figure 5.3(B). The red line is the iso-desirability curve corresponding to $D(\boldsymbol{\Lambda})$. The global desirability index computed on each point of this curve is the same than $D(\boldsymbol{\Lambda})$. Its equation is obtained from Equation 5.11 (right), writing $d(\lambda_2)$ as a function of $d(\lambda_1)$. The red points are the samples with a global desirability index higher than $D(\boldsymbol{\Lambda})$. The proportion of red points is also obtained using Equation 5.12.

Finally, it is possible to retrieve the iso-desirability curve in the CQAs space, using the inverse desirability functions (i.e. in this case, the Normal quantile functions Φ^{-1}). If the iso-desirability curve is represented by the set of points $(d_1(\mathbf{y}_1^*), d_2(\mathbf{y}_2^*))$ such that the global desirability index is exactly 0.5, the following set of points $(\mathbf{y}_1^*, \mathbf{y}_2^*)$ is the iso-desirability curve in the CQAs space:

$$\begin{aligned} \mathbf{y}_1^* &= \lambda_1 + s_1.\Phi^{-1}(1 - d_1(\mathbf{y}_1^*)) \\ \mathbf{y}_2^* &= \lambda_2 + s_2.\Phi^{-1}(d_2(\mathbf{y}_2^*)) \end{aligned} \tag{5.13}$$

The untransformed iso-desirability curve is shown on Figure 5.3(D, red), and illustrates the flexibility left for the specifications. Blue points corresponded to samples fulfilling jointly the specifications, while red points now correspond to samples providing a desirability better than $D(\mathbf{\Lambda})$.

Finally, regarding risk-based MCO, the expected probability to have both specifications *flexibly* accepted can be chosen as the univariate criteria to optimize. Thus, in this case, the optimal condition $\mathbf{x}^*$ is:

$$\mathbf{x}^* = \underset{\tilde{\mathbf{x}} \in \chi}{\operatorname{argmax}}\, P\left(D(\boldsymbol{Y}) \geq D(\mathbf{\Lambda}) \mid \tilde{\mathbf{x}}, \text{data}\right). \tag{5.14}$$

With or without desirability functions - link between both approaches

The desirability approach simply extends the approach based on joint probability. To show how it happens, the parametrization of the desirability functions will be changed progressively, with the stiffness parameters (s_j) approaching to 0. Doing so makes the cumulative Normal distribution a step function. This process is possible for the other forms of DFs.

When s_j tends to 0, the desirability function has a discrete form with the ordinate that can only be 0 or 1. The impact on the iso-desirability curves is illustrated on Figure 5.4. As previously, the DFs are kept centered on the specifications (i.e. the “steps” of the DFs are the specifications λ_j). In (A), s_j is as in the previous section. In (B), it has been divided by 6. Eventually, when the stiffness parameter comes very close to 0, as in (C), the DF tends to be a step function (in practice, s_j has been divided by 1000). The marginal desirability of a sampled point for the j^{th} CQA tends to 1 if it achieves the specification λ_j, 0 otherwise.

When each s_j takes a 0 value, no trade-off is allowed any longer. Samples from the distribution of the global desirability index will take value 0 when only one CQA has a desirability of 0. Conversely, it will be 1 only if every CQA has desirability 1.

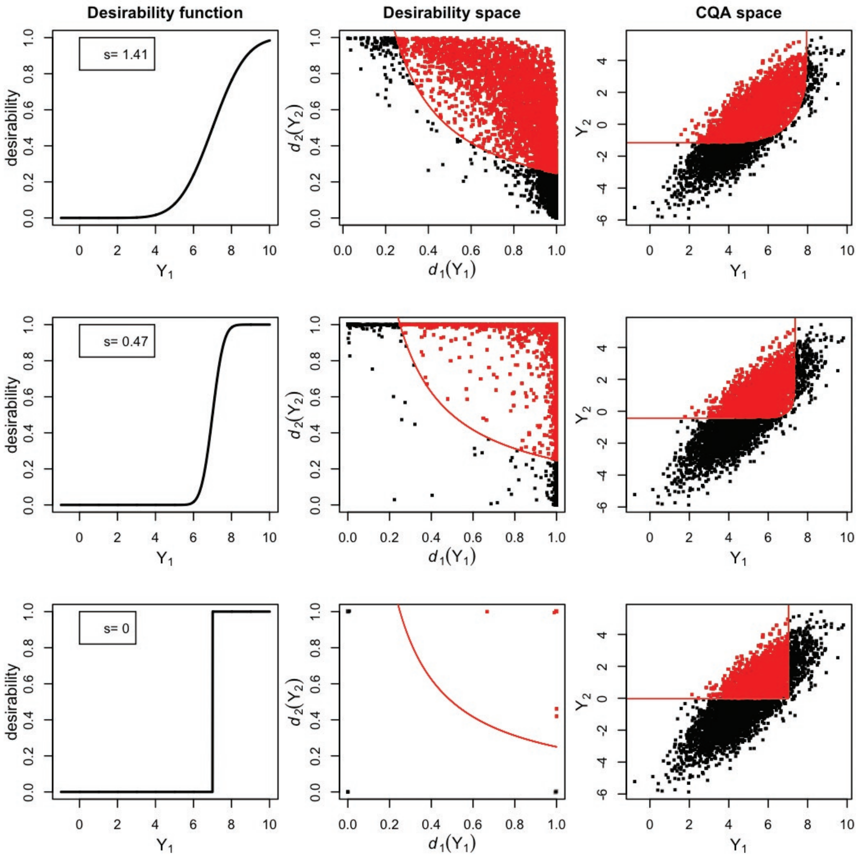


Figure 5.4: Behavior of the GDI when s_z tends towards 0.

Thus, the global desirability index is also discrete. Moreover, the results using this discrete global desirability index coincide with the results obtained using Equation (5.6), the global desirability values becoming the results of the indicator function $I()$ in Equation (5.6). This similarity is observed comparing the third plot of Figure 5.4 (C) with Figure 5.1).

5.5 Conclusions

In this chapter, the ways to compute the expected probability to observe a process or an analytical method running within specifications have been explained. Such predicted probability is at the heart of the Design Space computation.

A short review of multi-criteria decision method has been made to understand why it is important to base the decisions on the joint posterior predictive distribution of the Critical Quality Attributes.

When it is possible to envisage some trade-off between the critical quality attributes, desirability functions can be used as a flexible optimization tool. In this context, the decision must also be made on the basis of the predictive distribution of the global desirability index. Furthermore, it was shown it is possible to remain in the risk-based framework when using desirability indices.

Finally, it was illustrated how the desirability methodology simply extends the approach based on joint probability to be within specification.

Chapter 6

Independent component analysis and clustering methods to track chromatographic peaks in DoE context

6.1 Introduction

In the context of the development of separation methods such as the high performance liquid chromatography (HPLC), a matter of particular interest is to automate the data treatment of chromatograms. Indeed, when envisaging the computation of a Design Space for a separation method, or when obtaining the data for the validation of a chromatographic method, the identification of the peaks and their integration are generally tedious and time consuming.

This data treatment can be decomposed in two successive steps. The first step consists in the extraction of the peaks from the noise that is observed in the chromatograms. This process is referred as *peak picking*. The second step aims at discovering the compounds behind the identified peaks, and to track them on the different chromatograms. This is referred as *peak tracking*.

When envisaging design of experiments (DoE) for chromatographic method development, the manual peak picking and tracking can be very problematic as the amount of collected data might be large and the order of the apparition of the peaks (the elution order or the selectivity) changes from one chromatogram to another. A classical issue is the observation of several peaks at the same time. This phenomenon is known as *coelution*. All these facts make tedious the extraction of relevant information in complex chromatograms.

Fortunately, mathematical and statistical methodologies exist to improve knowledge from large arrays of data (Vivó-Truyols and al., 2005a,b; Hu et al., 2005; de Juan and Tauler, 2007). In the last decade, the independent component analysis (ICA) proposed by Hyvärinen and Oja (1997) has been found powerful to separate many types of signals. See also Hyvärinen (1999); Hyvärinen and Oja (2000); Delorme et al. (2007); Yamamoto et al. (2006). In analytical chemistry, the ICA even becomed a classical tool for data analysis. Recent reviews by Wang et al. (2008) and Parastar et al. (2011) show several applications and some limitations of the algorithm in this field.

In the domain of chromatographic method development, the usefulness of ICA has been demonstrated in Debrus et al. (2009, 2011b) and parts from these works and data are used in the present chapter. When human expertise is not sufficient to properly pick and track the peaks with diode array-detected (DAD) chromatograms, ICA may significantly improve the knowledge about the retention times of the peaks. It is aso able to reconstruct the UV-signatures of the compounds, useful for their identification.

Basically, ICA has the ability to *separate* signals recorded by different monitors, by demixing them in a number of independent sources. Considering a DAD-chromatogram (see Figure 6.1, left), the monitors are the recorded wavelengths obtained by the diode array detector. At each time point, a vector of information is provided. It consists in the absorbance of the UV light projected through the mobile phase to contains the mixture. Each change of absorbance corresponds to exogenous phenomenon such as compounds (observed as peaks), changes of mobile phase composition (e.g. baseline drift, dead-time perturbation), etc.

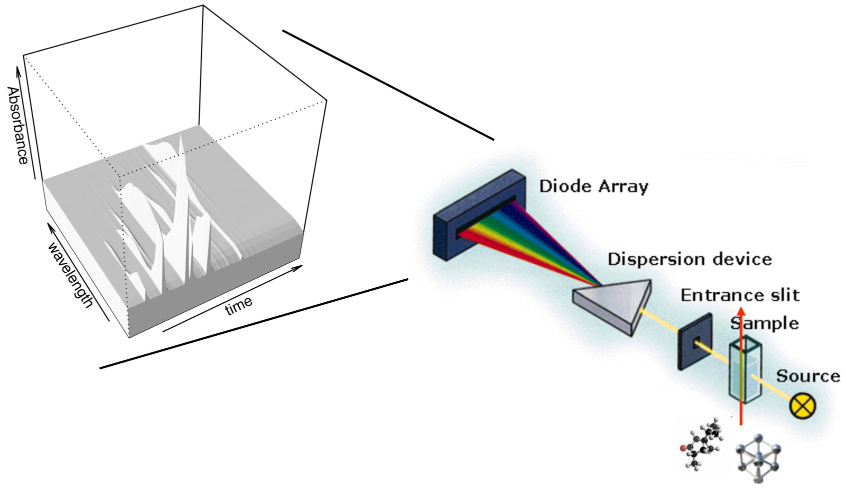


Figure 6.1: Recording of a DAD chromatogram.

From our previous works, ICA decomposition allowed the picking and manual identification of peaks, based on their reconstructed UV-signature (Debrus et al., 2009, 2011b). Nevertheless, the automated tracking of the peaks among various chromatograms has not been addressed so far. Various methods also exist to improve the peak matching.

Lankmayr et al. (1989) described a methodology based on fuzzy theory, that is applied on some criteria such as the area and the elution order of the peaks. Molnar et al. (1989) proposed tools based on the normalized band area of the peaks, allowing the matching of peaks between different runs where the HPLC gradient steepness is changed. See also Molnar (2002). Bogolomov and McBrien (2003) developed the mutual peak matching (MAP) approach. It uses abstract factor analysis (AFA) and iterative key set factor analysis (IKSFA) to detect the unknown number of peaks in a mixture. This last methodology also aims at providing the basis to develop an automated peak matching system. However, it is observed that these tools succeed to match peaks only for some specific cases, but they might fail the recognition with real or complex data.

Objectives, notations and structure of the chapter

Assume that the data consists in n DAD-chromatograms that were recorded on the **same mixture** at n **different chromatographic conditions**. Changes in these conditions includes modifications of selected factors such as the pH of the buffer, the temperature, or the time to modify linearly the proportion of organic modifier in the mobile phase from a low level to a higher level, known as gradient time. In this context, DOE is an efficient way to plan the way the chromatographic conditions are explored. Each DAD-chromatograms is an $(m \times t_{\text{tot}})$ matrix containing m values of absorbance recorded over a certain time period t_{tot} indexed as $1, ..., t, ..., t_{\text{tot}}$.

The objectives of this study are multiple. They are listed below:

- find the (assumed unknown) number of compounds in the mixture, c_{opt},
- optimize the number of ICA sources f_i used to separate the n DAD-chromatograms $(i = 1, ..., n)$,
- extract the c_{opt} compounds in the n DAD-chromatograms from the noise (peak picking),
- rebuild n ICA estimated DAD-chromatograms, cleaned from the noise,
- match the peaks among the n DAD-chromatograms (peak tracking),

- determine peaks parameters such as their elution times, their area under the curve (AUC), etc.

A general objective finally consists in the automation of the previous objectives.

To simplify the notation, t_{tot} must be understood as the total time of one specific DAD-chromatogram. However, t_{tot} is generally different from one DAD-chromatogram to another. Similarly, when studying one DAD-chromatogram, the number of ICA sources will be noted f.

To achieve the objectives, an innovative methodology is presented, based on classical tools of classification to cluster together the peaks corresponding to relevant compounds. Classification criteria are based on the peak UV-signatures that are reconstructed using ICA. Section 6.3 focuses on the ICA applied to chromatographic data, to provide a theoretical background. Section 6.4 describes how the number of sources f_i used to unmix the n original chromatograms can be chosen. The aim is to obtain the best results for peak picking. Section 6.5 proposes a methodology to classify together different peaks coming from different DAD-chromatograms, solving the problem of automated peak tracking. Additional results are discussed in Section 6.6.

6.2 Data

To illustrate the application of the methodology, an example of data is provided. It consists in $n = 17$ DAD-chromatograms realized on a blinded mixture provided by Eli Lilly & Company. In a DoE used to optimize the separation of the mixture, the chromatographic conditions were changed. This resulted in very different DAD-chromatograms. Data and optimization results are fully presented in Debrus et al. (2009). Figure 6.2 (top) illustrates two DAD-chromatograms. Several peaks in each chromatogram are observed. Some of them are very small, and their visibility may depend on the particular wavelength of observation (bottom chromatograms are observed at 240 nm). Furthermore, some peaks are mixtures of two (or more) peaks. Mixed signals are typical in DoE context, leading to a complex identification of the compounds and of their attributes (e.g. time of beginning, apex and end of the peaks, UV-signatures, etc.).

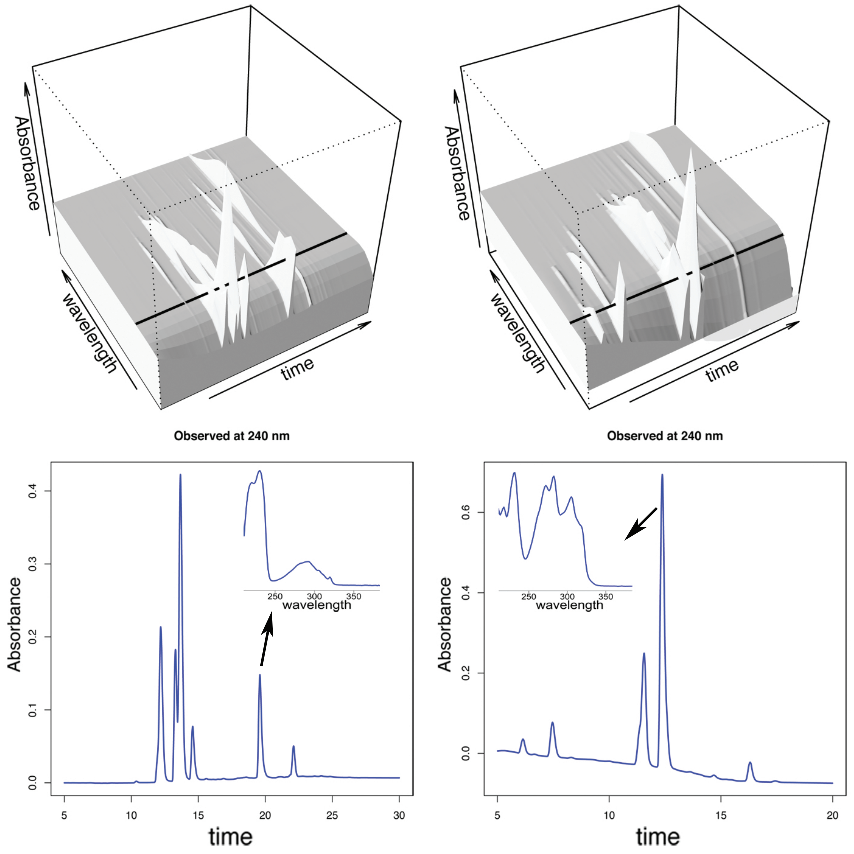


Figure 6.2: (top) Examples of DAD-chromatograms recorded at pH 2.6, gradient time 10 min (left), and pH 10, gradient time 20 min (right). (bottom) Chromatograms extracted at wavelength 240 nm. (left, insert) UV-signature of the peak observed at 19.75 min. (right, insert) UV-signature of the peak observed at 12.5 min. This spectrum is a mixture of the signatures of two coeluted compounds.

6.3 Sources extraction

6.3.1 Independent Component Analysis

ICA is a non supervised technique that is applicable to blind source separation problems. Different definitions can be given for ICA. Hereafter is presented the *noise-free* ICA, developed by Hyvärinen and Oja (1997).

The ICA of an $(m \times t_{\text{tot}})$ matrix $\mathbf{X}$ of m signals observed at t_{tot} different times, $(\mathbf{x}_1, ..., \mathbf{x}_m)'$, consists of estimating the following generative model for the data:

$$\mathbf{X} = \mathbf{AS} \tag{6.1}$$

where $\mathbf{A}$ is a constant $(m \times f)$ mixing matrix. f is the main parameter of ICA and requires fine tuning $(f \leq m)$. $\mathbf{S} = (\mathbf{s}_1, ..., \mathbf{s}_f)'$ is a $(f \times t_{\text{tot}})$ matrix containing the f independent sources of the signals. Both $\mathbf{A}$ and $\mathbf{S}$ must be estimated.

Before applying ICA, a preprocessing step is applied. It consists in centering and whitening the data. The concept of whitening is to transform the matrix $\mathbf{X}$ such that its lines are uncorrelated and unit variance. These two transformations provide a matrix $\mathbf{X}_w$ of dimension $(m \times t_{\text{tot}})$. A principal component analysis (PCA) is common to carry out this task. This can also allow for a first dimension reduction by selecting only the f most informative components.

Other possible pretreatments of the DAD-chromatograms can be envisaged before applying ICA. On one hand, useless data points can be removed, such as the beginning and the end of the chromatograms, if no peak is observed. On the other hand, some smoothing using low-pass filter can be done, using e.g. the algorithm of Savitzky and Golay (1964). This filter is particularly good at preserving original chromatographic shape while removing noise with high frequencies (R package `signal`, Short, 2011).

To accomplish the blind source separation, the ICA algorithm iteratively estimates an unmixing matrix $\hat{\mathbf{A}}_w'^{-1}$ from $\mathbf{X}_w$, in such way that the lines of $\hat{\mathbf{S}}_w$ are as statistically independent as possible. On the basis of $\hat{\mathbf{A}}_w'^{-1}$, the independent sources are estimated as follows.

$$\hat{\mathbf{S}}_w = \hat{\mathbf{A}}_w'^{-1} \mathbf{X}_w \tag{6.2}$$

A consequence of the preprocessing step with PCA is that ICA works on orthogonal and standardized data. However, most algorithms are able to retrieve the matrices $\hat{\mathbf{A}}^{-1}$ and $\hat{\mathbf{S}}$ in their original scale, from $\hat{\mathbf{A}}_w^{-1}$ and $\hat{\mathbf{S}}_w$. To apply the ICA, the package `fastICA` for R has been used (Marchini et al., 2010). The algorithm is based on the fact that a way to find independent sources is to search for an unmixing matrix $\hat{\mathbf{A}}_w^{-1}$ that maximizes the non-gaussianity of the sources, as proposed by Hyvärinen and Oja (1997). Finally, notice that the pseudo-inverse of $\hat{\mathbf{A}}_w^{-1}$ is computed to retrieve $\hat{\mathbf{A}}_w$.

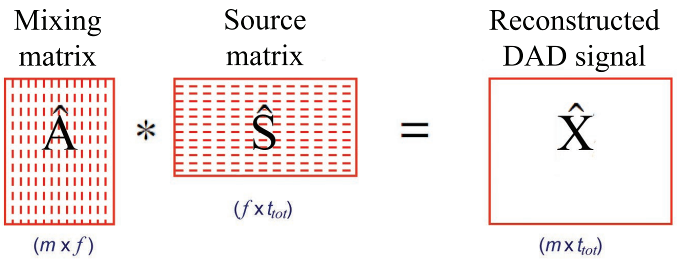


Figure 6.3: Illustration of the signal reconstruction after ICA.

After estimating the independent sources and their mixing process, it is possible to reconstruct the complete original data (assuming $f = m$), as illustrated on Figure 6.3. Furthermore, the g^{th} component $\hat{\mathbf{X}}_g$ of size $(m \times t_{\text{tot}})$ is defined as the matrix product between the g^{th} column of $\hat{\mathbf{A}}$ and the g^{th} line of $\hat{\mathbf{S}}$:

$$\hat{\mathbf{A}}_{,g}\hat{\mathbf{S}}_{g,} = \hat{\mathbf{X}}_g \text{ and } \sum_{g=1}^{f} \hat{\mathbf{X}}_g = \hat{\mathbf{A}}\hat{\mathbf{S}} = \hat{\mathbf{X}}. \tag{6.3}$$

At this stage, a selection can be done on the f estimated components. It is possible to discard the components containing noise (Bugli, 2006) or irrelevant artifacts from the summation. In this way, a cleaned version of $\hat{\mathbf{X}}$ can be obtained.

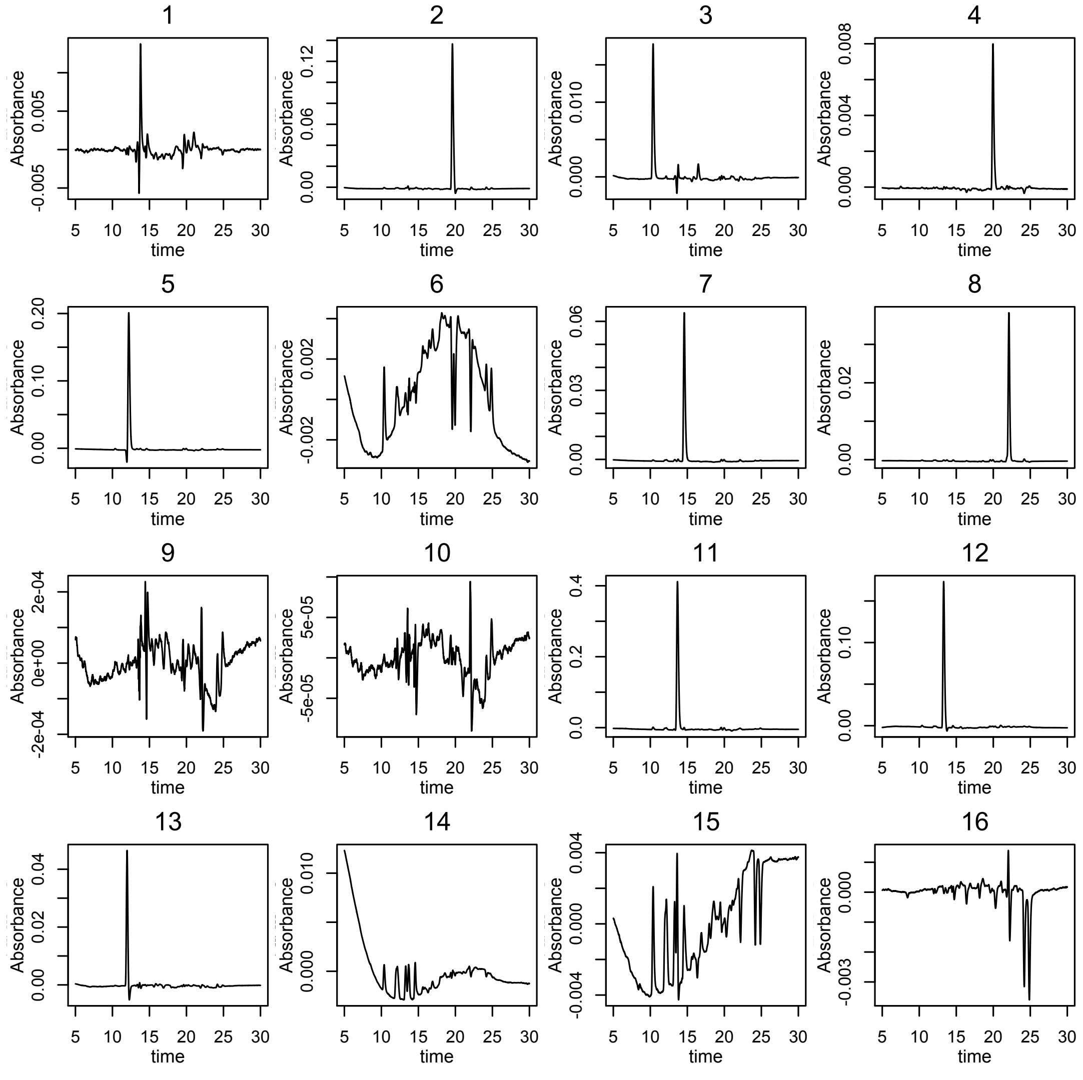


Figure 6.4: Components computed by ICA with $f = 16$, plotted at the wavelength of 240nm. Components 1–5, 7, 8, 10–13 are visually identified as peaks. Other components correspond to noise or irrelevant artifacts.

To illustrate the source separation, ICA was applied on a DAD-chromatogram from the data set (presented in Figure 6.2, left). On this example, f was manually tuned and a value of $f = 16$ was chosen. Figure 6.4 illustrates the estimated components $\hat{\mathbf{X}}_{g,240\text{nm}}$, observed at 240nm (i.e., each graph shows one line of each $\hat{\mathbf{X}}_g, g = 1, ..., f$). Unlike the results that can be obtained with PCA, the components are unordered. However, the ten components indexed as 1–5, 7, 8, 10–13 can be visually identified as chromatographic peaks. The other components seem to contain noise and non-relevant artifacts such as baseline drift (component 14). A possible way to manually assess the relevance of a component is to look at the scale (y-axis) of the plots.

Obviously, this process is non systematic and is subject to error of judgement. Furthermore, nothing indicates that the value of $f = 16$ is optimal and will be a good candidate for the other DAD-chromatograms under study. In the same way, there is low assurance that $c_{\text{opt}} = 10$ compounds are effectively present in the mixture. A last issue consists in the application of ICA for each of the n DAD-chromatograms, with the tedious and time-consuming selection of the appropriate number of sources f_i ($i = 1, ...n$), and followed by the manual peak picking. A methodology to automate this process is desirable and is the subject of the next section.

Finally, Figure 6.5 shows the corresponding chromatogram (observed at 240nm) that was reconstructed using only the ten selected components following Equation (6.3). The vertical grey lines show the result of the peak picking and are placed at the elution times of the apexes of the peaks that were identified in the 10 selected components. As the original data observed at 240nm did not show significant baseline drift or noise, the reconstructed DAD-chromatogram observed at the same wavelength is very similar to the one presented in Figure 6.2, left (bottom).

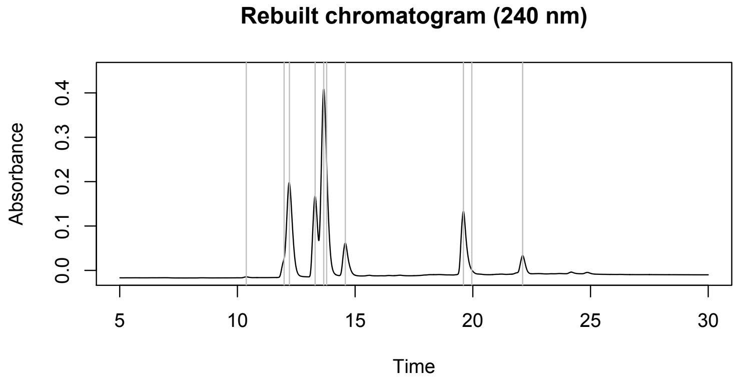


Figure 6.5: Chromatogram reconstructed with the ten selected components and observed at 240nm. (Grey) results of the peak picking.

6.4 Automated selection of relevant sources

This section details an automated methodology to optimize the number of sources f_i used to separate the n DAD-chromatograms ($i = 1, ..., n$), and to identify the unknown number of compounds c_{opt} in the mixture. The general idea is to derive some characteristics of the ICA components, computed on each of the DAD-chromatograms separately. These characteristics are then used to differentiate the peaks from noise and irrelevant artifacts. Next, the results obtained for all the DAD-chromatograms are analyzed concurrently to identify the optimal f_i and c_{opt}.

6.4.1 k-means algorithm to identify sources containing relevant information

Assuming that noise follows a Normal distribution (Gaussian white noise), the independent components given by ICA are investigated in order to find which one may be considered as relevant or contains only noise. It has been discussed that the kurtosis or other high-order statistics of a distribution are good criteria to characterize artifacts and noise (see McKeown et al., 1998; Delorme et al., 2007).

Thus, different statistics or moments can be used to check the normality of the distribution of each source. If normality is not observed, it should imply that the component is unlikely to be noise and therefore likely corresponds to an exogenous phenomenon such as a compound present in the mixture. In summary, (lines of the) components of interest are assumed to have a non-normal distribution, and will be identified as peaks or other relevant artifacts. The following statistics were found useful for this purpose : the kurtosis and the Shapiro-Wilks statistics

Kurtosis

The kurtosis is the fourth standardized moment of a random variable and is a measurement of the peakedness (or the flattening) of the distribution of this variable. It may be estimated on the components at a particular wavelength or simply on the sources provided by ICA. When computed on the g^{th} source of one DAD-chromatogram, it is defined as:

$$\hat{K}_g = \frac{\hat{m}_{g,4}}{\hat{m}^2_{g,2}} - 3 = \frac{\frac{1}{t_{\text{tot}}}\sum_{t=1}^{t_{\text{tot}}}(S_{g,t} - \bar{S}_g)^4}{\left(\frac{1}{t_{\text{tot}}}\sum_{t=1}^{t_{\text{tot}}}(S_{g,t} - \bar{S}_g)^2\right)^2} - 3. \tag{6.4}$$

$\hat{m}_{gk}$ represents the estimated moment of order k (about the mean) for source g, and $\bar{S}_g$ is the mean of source g. The kurtosis of the Normal distribution is zero. Lets also define $\hat{\mathbf{K}}_{i,f_i}$, the vector of f_i estimated kurtosis for the DAD-chromatogram i.

Shapiro–Wilk statistic

The non-parametric Shapiro–Wilk (S–W) test examines the null hypothesis that a sample $S_{g,1}, ..., S_{g,t}, ..., S_{g,t_{\text{tot}}}$ comes from a Normally distributed population. This hypothesis is rejected if the test statistic is too small. This test statistic can be used to characterize the distribution (or equivalently, the associated p-value). More details about this test can be found in Shapiro and Wilk (1965). S-W statistics can be computed on the components or on the corresponding sources as well. It is carried out on the sources for this application. Lets define $\widehat{\mathbf{SW}}_{i,f_i}$, the vector containing the f_i S-W statistics computed on the DAD-chromatogram i.

Kurtosis and S–W statistics can then be used to describe the degree of normality of a distribution and, hence, to identify sources containing noise. However, these are not the only possible criteria and other simple or complex statistics can also be used in this context, such as skewness of the distribution of the source, Kolmogorov–Smirnov Normality test, etc. Simple decision based on peak's height, range and variance of the source, area under the curve, etc. are also possible. It can allows discarding some identified peaks that are smaller than a certain threshold.

k-means clustering

The objective is to use the characteristics of the sources (or components) to discriminate relevant and irrelevant artifacts. k-means is a non-supervised method to cluster objects into k partitions on the basis of their attributes. For the DAD-chromatogram i, the objects are the f_i sources computed by ICA and the attributes are the estimated statistics computed on these sources, namely the kurtosis and Shapiro–Wilk statistic.

Before applying clustering, attributes are divided by their standard deviation computed from the f_i sources in order to give each attribute the same weight in the clustering decision process.

k-means works by comparing the Euclidian distance between the objects. A short distance (slight difference in the computed attributes) is a sign of closeness between objects. The closest objects are put together in the same cluster. Different implementations exist for k-means clustering. The Hartigan and Wong (1979) algorithm was used here (R package `stats`, R Development Core Team, 2010).

As the purpose is to separate the sources between two classes (relevant or noisy), k-means is set up with $k = 2$. The ideal situation arises when the clusters are as follows: the first with sources having a high Kurtosis and low S–W values, corresponding to the relevant cluster (c_r), and the second with low Kurtosis and high S–W value, corresponding to noise and irrelevant artifacts. The outcome of the k-means clustering algorithm may vary from one run to another because the k centers are randomly placed at the start of the algorithm. To avoid this dependency to initial conditions, it is a good practice to run the algorithm one hundred times. This is done for each of the n DAD-chromatograms, while changing initial values. For each DAD-chromatogram i, the clusters occurring with the highest frequency are selected.

Figure 6.6 (top) illustrates the results of the k-means algorithm applied on the chromatogram presented in Figure 6.2 (left), on the 16 sources computed by ICA. The cluster identified by the red cross contains the relevant artifacts. On the bottom, the apexes of the peaks in the sources identified as relevant are juxtaposed (sources 1–5, 7, 8, 10–13, grey lines) with the original chromatogram. The corresponding source numbers are given in red. The coelution between the peaks present in sources 13 and 5, and sources 2 and 4 have been resolved. Small peaks were also properly picked, such as the one observed in source 3. Notably, these automated results are the same than the one obtained previously with manual peak picking.

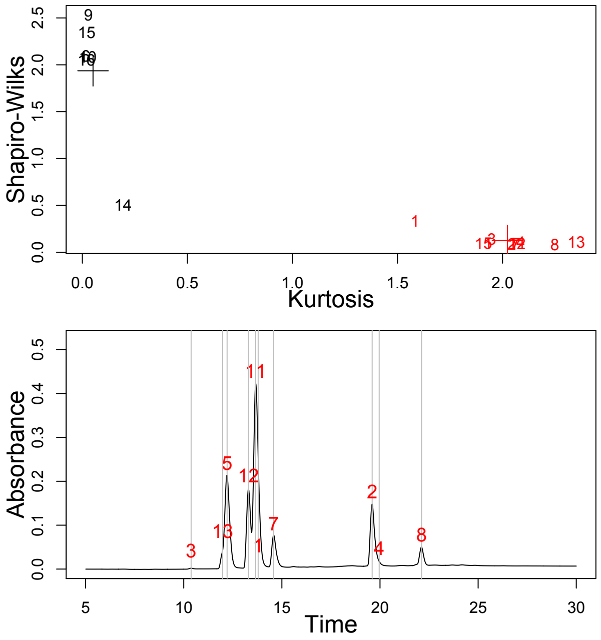


Figure 6.6: Results of k-means clustering. (top) Ten sources are clustered together as relevant artifacts (red). (bottom) Original chromatogram with the automatically identified apexes (grey lines).

6.4.2 Number of sources to identify the unknown number of peaks

In this section, the automated optimization of the number of sources f_i and automated identification of the unknown number of compounds c_{opt} are detailed, using the results of the k-means algorithm. The idea is to run ICA on the n DAD-chromatograms several times, while slowly increasing f_i until finding some convergency in the identified number of peaks. The treatment is the following :

1. Set f_i to a low value (e.g. $f_i = 3$)
2. For $i = 1, ..., n$

    - Compute ICA on the DAD-chromatogram i, with f_i sources

    - Compute the Kurtosis and S-W statistics of the f_i sources

    - Apply k-means algorithm on the Kurtosis and S–W cloud of points

    - Record $c_r(i, f_i)$, the number of sources found in the relevant cluster

    End
3. Define $c_r^*(f_i)$, the mode of the $c_r(i, f_i)$ distribution (for $i = 1, ..., n$), corresponding to the most probable number of compounds in the mixture, for a given f_i
4. $f_i <- f_i + 1$
5. Repeat steps 2–4 until $c_r^*(f_i)$ stabilizes
6. If $c_r^*(f_i)$ is stabilized, define $f^* = f_i$.

Thus, $c_r(i, f_i)$ is assumed to be the number of detected peak for the DAD-chromatogram i, for a given f_i. As previously stated, the same mixture is injected several times using different analytical conditions. Thus, each DAD-chromatogram should contain the same information, i.e., the same number of relevant peaks. The algorithm makes clear that finding an optimal value for f_i is thus similar to identifying the most plausible number of sources counted in the relevant clusters for all DAD-chromatograms. The number of iterations to declare that $c_r^*(f_i)$ is stabilized is still subject to subjectivity. In our various applications of the algorithm, adding 10 iterations after the first stable value was generally sufficient. Lets define $f_{\max}$, the maximum number of sources used for ICA.

Figure 6.7 illustrates the process on the complete set of data. The value of $c_r^*(f_i)$, computed from the $n = 17$ chromatograms, is plotted against f_i, ranging from 3 to 25. A stabilization of $c_r^*(f_i)$ to a value of 10 is observed when $f_i = 16, \forall i$. It allowed defining $c_{\text{opt}} = 10$, the number of relevant sources present in the mixture.

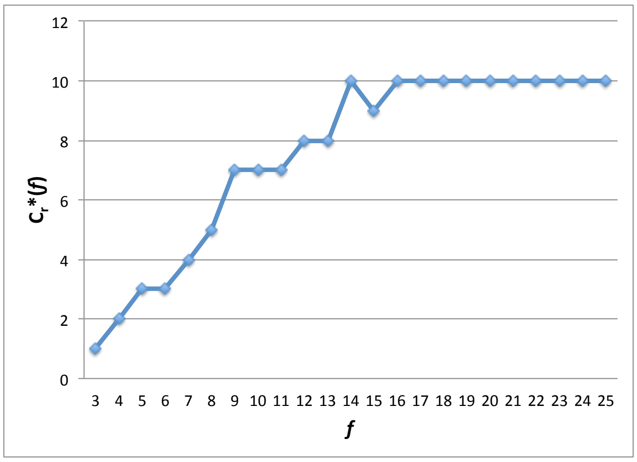


Figure 6.7: Optimization of f. Plot of the variation of $c_r^*(f)$ versus f.

Unfortunately, when fixing f^*, it is unlikely that the number of sources identified as relevant will be c_{opt} for every chromatogram. A solution is to use the recorded $c_r(i, f_i)$ to identify a value of f_i that can be used to obtain c_{opt} relevant sources for the chromatogram i ($i = 1, ..., n; f_i = 3, ..., f_{\max}$)[1].

For chromatograms where it is not possible to find c_{opt} sources for any f_i, another option is to use a ranking provided by the Kurtosis/S-W ratio and to chose the c_{opt} sources that have the highest ratio, keeping $f_i = f^*$. If the problem occurs for the DAD-chromatogram i, the ranking is then provided as the order of the following values: $\hat{\mathbf{K}}_{i,f^*}/\widehat{\mathbf{SW}}_{i,f^*}$.

For the chromatogram of Figure 6.2 (left) decomposed in $f^* = 16$ sources as in Figure 6.4, the ranking was the following:

Kurtosis/S-W	26.6	22.7	21.9	21.9	21.8	20.0	19.8	19.4	13.9	4.7	0.4	0.0	0.0	0.0	0.0	0.0
Source Index	8	2	13	7	12	11	5	4	3	1	14	10	9	16	15	6

For this example, the peak picking results were then unchanged when identifying the first c_{opt} sources with the highest ratio as relevant (c_{opt}=10).

After a comparison of the automated treatment and data that were treated manually, every peak of interest seemed to be automatically found among the 17 chromatograms, except one. Among the 17*10-1 = 169 sources identified as relevant, one source contained a baseline drift and 16 other sources contained artifacts around one compound (see sources 6 and 10 of Figure 6.6 (bottom)). In Debrus et al. (2009), these artifacts were deleted before further analysis, leading to a solution where 9 compounds were found in each chromatogram. For the next steps of this study, all the 169 relevant artifacts were intentionally kept to study how the automated peak

[1] To avoid non-deterministic execution of ICA, the random number generator seed must be fixed.

tracking can help to detect problems.

6.5 Peak tracking - classification

In the previous section, $s_{\text{tot}} = 169$ ICA sources and corresponding components were identified as relevant in the DAD-chromatograms of the experimental plan. The aim of this section is to develop a methodology to match which peak in one chromatogram corresponds to the same peak in the others. The s_{tot} reconstructed UV-signatures (i.e. one column of each independent components) are used to assess the similarity between peaks among the DAD-chromatograms.

To solve this problem, agglomerative clustering is proposed to create sets of objects (the components), that aims at finding groups that are the most homogeneous with regard to some predefined criteria computed on the UV-signatures recovered by ICA. A conventional and efficient method to find these sets is the hierarchical clustering algorithm developed by Ward (1963). The idea is to build a *dendrogram* representing a sequence of partitions, from the leaves (where s_{tot} clusters contain the components of every chromatogram), to the trunk (where 1 cluster contains s_{tot} components). At each iteration of the procedure, one cluster is nested within another based on a symmetrical $(s_{\text{tot}} \times s_{\text{tot}})$ dissimilarity (or distance) matrix computed from attributes of the objects.

Usually, an Euclidian distance matrix is provided to the Ward's algorithm to avoid local optima (and to preserve the interpretation of distance throughout the algorithm). However, it is possible to extend the application of the algorithm using dissimilarity matrices, although the strict order of "distances" among clusters will not be preserved (Duda et al., 2001). This is not a problem as these dissimilarities are not used as a metric for further analysis and the local optima generally does not prevent the algorithm to finish. The next section details how to compute several dissimilarity matrices that can be combined together to obtained the desired peak matching.

6.5.1 Distances and dissimilarities between objects

For each $(m \times t_{\text{tot}})$ component $\hat{\mathbf{X}}_s, (s = 1, ..., s_{\text{tot}})$ identified as relevant in the complete set of DAD-chromatograms, the lines are vectors containing an unmixed signal of a chromatogram, observed at specific wavelengths. Remember that this information was used to identify the components/sources as relevant or noisy. On the other hand, the columns of the $\hat{\mathbf{X}}_s$ are assumed to be the reconstructed UV-signature of a compound. It is further assumed that two components that have the same UV-signature represent the same compound. Hence, they should be grouped together. Usually, the elution time of the compounds does not influence the UV-signature.

Lets define $\mathbf{u}^{(s)}$, the column of $\hat{\mathbf{X}}_s$ that contains the UV-signature of the component s, observed at the time of the apex (maximum over the time) of the peak. Below are listed some interesting distances or dissimilarities that can be computed between two UV-signature vectors $\mathbf{u}^{(s_1)}$ and $\mathbf{u}^{(s_2)}$ $(s_1, s_2 = 1, ..., s_{\text{tot}})$.

Euclidean distance

The distance between the two points defined by the vectors $\mathbf{u}^{(s_1)}$ and $\mathbf{u}^{(s_2)}$ in the m-dimensional Euclidean space is defined as:

$$d_E(s_1, s_2) = \sqrt{\sum_{j=1}^{m} (u_j^{(s_1)} - u_j^{(s_2)})^2}. \tag{6.5}$$

The Euclidean distance between identical spectra is then 0.

Cosine of the angle between vectors

A classical method used to compare two vectors is the cosine of the angle formed between them in the m-dimensional space. It is computed as follow:

$$\cos\phi(s_1, s_2) = \frac{\sum_{j=1}^{m} (u_j^{(s_1)}.u_j^{(s_2)})}{\sqrt{\sum_{j=1}^{m} (u_j^{(s_1)})^2 . \sum_{j=1}^{m} (u_j^{(s_2)})^2}}. \tag{6.6}$$

When $\mathbf{u}^{(s_1)}$ and $\mathbf{u}^{(s_2)}$ are similar, the angle between them is equal to 0. $\cos\phi(s_1, s_2)$ is then equals to 1. $- \mid \cos\phi(s_1, s_2) \mid$ can be chosen as a distance between the two vectors.

Correlation coefficient

The cloud of points created by plotting one UV-signature vector against one other will lay on a straight line if the vectors are similar. In this case the correlation coefficient is about 1. The Pearson correlation is computed as:

$$r(s_1, s_2) = \frac{\sum_{j=1}^{m} \left(u_j^{(s_1)} - \bar{u}^{(s_1)}\right)\left(u_j^{(s_2)} - \bar{u}^{(s_2)}\right)}{(m-1)s_{\mathbf{u}^{(s_1)}} s_{\mathbf{u}^{(s_2)}}}, \tag{6.7}$$

where $\bar{u}^{(s)}$ and $s_{u^{(s)}}$ are the estimated mean and standard deviation of $\mathbf{u}^{(s)}$, respectively. $-\mid r(s_1, s_2) \mid$ can be chosen as a distance between the two vectors.

Mutual information

In information theory, the mutual information quantifies the mutual dependence between two random variables. It is then a measure of the information shared by $\mathbf{u}^{(s_1)}$ and $\mathbf{u}^{(s_2)}$. It can be defined as:

$$MI(s_1, s_2) = \int_{\mathbf{u}^{(s_1)}} \int_{\mathbf{u}^{(s_2)}} p(\mathbf{u}^{(s_1)}, \mathbf{u}^{(s_2)}) \log\left(\frac{p(\mathbf{u}^{(s_1)}, \mathbf{u}^{(s_2)})}{p(\mathbf{u}^{(s_1)})\, p(\mathbf{u}^{(s_2)})}\right) d\mathbf{u}^{(s_1)}\, d\mathbf{u}^{(s_2)}, \tag{6.8}$$

where $p(x, y)$ is the joint density function of x and y, and $p(x)$ and $p(y)$ are the marginal density functions of x and y. If $\mathbf{u}^{(s_1)}$ and $\mathbf{u}^{(s_2)}$ are independent (dissimilar), $MI(s_1, s_2) = 0$. Specialized packages can compute the mutual information efficiently (e.g. R package `minet`, Meyer et al., 2008).

Length of correlation line

An innovative yet simple criterion is the length of the correlation line that is created by plotting one UV-signature vector against another one. The idea is then related to the correlation coefficient. This dissimilarity is computed as follows:

$$d_{cl}(s_1, s_2) = \sum_{j=1}^{m-1} \sqrt{(u_{j+1}^{(s_1)} - u_j^{(s_1)})^2 + (u_{j+1}^{(s_2)} - u_j^{(s_2)})^2} \tag{6.9}$$

This criterion makes a better use of the particular shape of the UV-signatures and is not too sensitive to their scales. It has been created to attempt to solve the fact that other criteria generally attribute too high similarities between vectors that are obviously different (e.g. it is too “easy” to obtain a correlation coefficient of 0.99).

Other criteria

It is difficult to be exhaustive in the definition of interesting criteria to quantify the closeness of agreement between UV-signatures. Among others, the difference between area under the curve (AUC) of two peaks, or the difference between the values of absorbance at apex (i.e. peak's height) for a specific wavelength, are attractive criteria if the injected solution keep identical the quantities of compounds. However, this might not be the case during validation experiments where the concentrations of compounds are generally changed. This results in changes of peak's AUC and height.

Usually, once a HPLC method is well developed so that all the compounds are separated, there is no reason to modify the input parameters. In this case, the times at which the peaks are observed will not change (except a small natural noise), and a last relevant criterion can be derived from the time of apexes of each peak observed in the relevant components. The elution order would also be an appealing criterion in this case. See Chapter 10 for an application of such criteria in a method validation context.

6.5.2 Artificial penalty distance

To avoid the possibility to match together several components coming from the same chromatogram, a special artificial distance is given between these objects. It is simply defined as:

$$\begin{aligned} \text{if } \mathbf{u}^{(s_1)} \text{ and } \mathbf{u}^{(s_2)} \in i^{\text{th}} \text{ DAD-chromatogram} : d_{art}(s_1, s_2) = c, \qquad i = 1, ..., n, \\ \text{else} : d_{art}(s_1, s_2) = 1, \end{aligned} \tag{6.10}$$

where c is a large number (e.g. 100).

To create an appropriate dissimilarity matrix for the Ward's algorithm, the penalty distance is combined with one or several other distance matrices using a matrix sum or a matrix dot product.

6.5.3 Example

Further data treatments

It is known that the pH or changes of organic modifier may have an impact on DAD-chromatogram for the lowest wavelengths around 220nm. The UV-signatures of the compounds are affected by these changes. Unfortunately, the ICA is generally not able to perfectly remove these disturbance. To avoid problem when tracking the peaks, a good practice is to remove low wavelengths when computing the dissimilarities, i.e. cutting in the matrices of components. Hereafter, the data from 210 to 240 nm were removed. Notice this preprocessing can also be done before applying ICA.

Clustering

The application of the agglomerative clustering is carried out on the $s_{\text{tot}} = 169$ relevant components found after ICA. In this example, the following dissimilarity matrix $\mathbf{D}$ was used, combining the artificial penalty distance and the length of the correlation line–distance with a sum:

$$\mathbf{D} \equiv \{d_{s_1,s_2}\} = d_{art}(s_1, s_2) + d_{cl}(s_1, s_2).$$

The corresponding dendrogram obtained with Ward's algorithm is presented in Figure 6.8. The leaves of the tree are the components (characterized by their UV-signature) while the branch represents the merging of the components into clusters. A low grouping height stands for similar components (peaks) while higher heights identify dissimilar components.

In order to terminate the tracking of peaks, the cutting height of the dendrogram must be chosen. As $c_{\text{opt}} = 10$ relevant components were found in each chromatogram, it would be logical to cut the dendrogram at a height where 10 clusters are found.

However, in this example, it was decided to cut it at a height where only 9 clusters are found. Indeed, Figure 6.8 suggests that the components grouped in the cluster 3 share some similarities, although the artificial penalty distance should put away the components belonging to the same DAD-chromatogram. The UV-signatures were then assumed similar for these compounds. As explained, the reason is that cluster 3 corresponds to a compound that was erroneously split in two sources by ICA in most (15 out of 17) chromatograms. The reasons of this poor results are unidentified. An option to overcome this problem is to sum the components found in the same cluster and belonging to the same chromatogram. This would simply reconstruct the peak that was split.

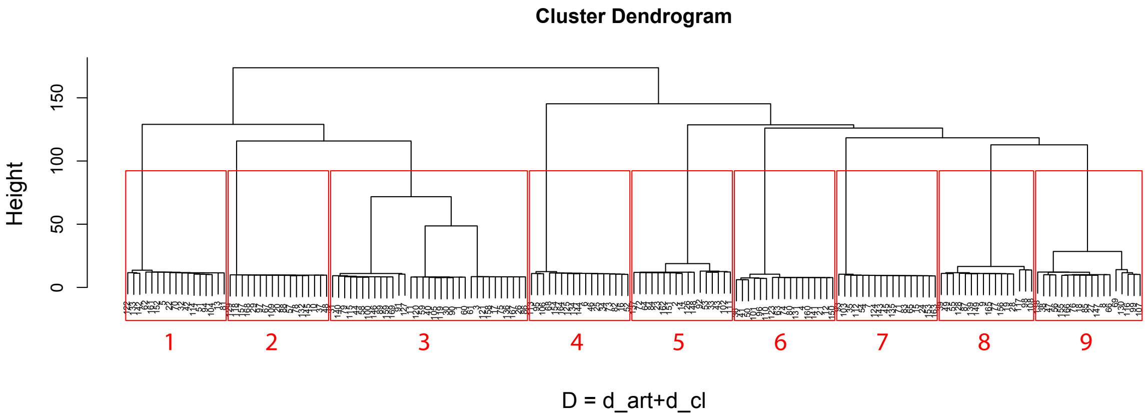


Figure 6.8: Dendrogram of the matching of peaks. The leaves are the objects (the components). Red rectangles are the clusters. Group 3 is bigger than the others as it contains the information of an ICA-duplicated peak.

Figure 6.9 (A-D) shows the UV-signatures of the compounds placed in the clusters 1, 3, 6 and 8, respectively. From the n =17 chromatograms, 17 components were properly classified in the cluster 1 (A), in spite of the clear perturbations observed at wavelengths 210-250 nm (note these wavelength were removed prior to computations). As explained, it was due to the pH changes across the experiments. Next, cluster 3 (B) contains the components of the peak (naproxen) that has been split by ICA. Fortunately, the matching allowed easily retrieving all the sources. Among the 33 components present in this cluster, one also corresponded to a wrongly matched peak (box). Next, 17 components were properly classified in cluster 6 (C). Finally, Clusters 8 (D) contained only 16 components. The missing peak was not picked as a relevant component after the ICA. This peak could be easily retrieved manually.

All the other clusters tracked perfectly the peaks except cluster 9 where one supplementary component was found. Based on the UV-signature, it was easy to identify it as a solvent artifact.

Finally, the compounds composing the unknown mixture were revealed and analytical expertise was used to match the real compound names and the clusters. Table 6.1 shows the correspondence between the cluster numbers and the real names of the compounds. The compounds that was erroneously split by ICA was the Naproxen. No other compound had a similar UV-signature.

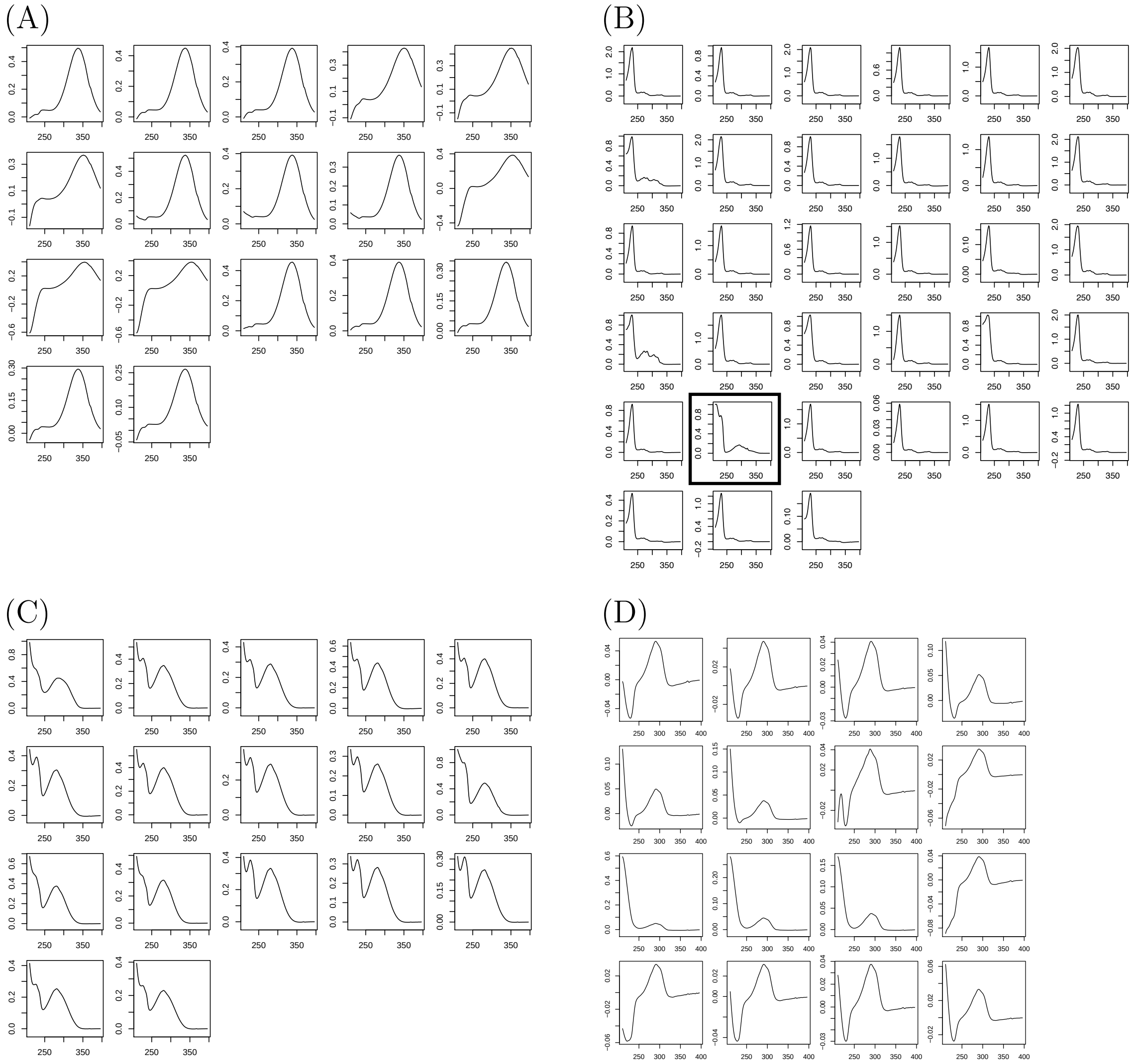


Figure 6.9: Example of matching of peaks based on the UV-signature of the ICA-reconstructed components. x-axis: wavelength; y-axis: absorbance. (A) cluster 1; (B) cluster 3; (C) cluster 6; (D) cluster 8.

Cluster number	Compound
1	Retinoic acid
2	Warfarin
3	Naproxen
4	Atenolol
5	Pindolol
6	Indoprofen
7	Impurity of retinoic acid
8	Unknown impurity
9	Propanolol

Table 6.1: Correspondence between cluster numbers and real compounds.

6.6 Additional results: accurate recording of retention times and other attributes

When peaks are severely coeluted, the manual recording of their time attributes is often intricate. The statistical models created to optimize and find Design Space of chromatographic method rely on these times and an accurate determination is mandatory. In several of our applications including the ones presented in the two next chapters, the models and their predictions were found better with the data processed by ICA.

Figure 6.10 compares the recording process on the first peak at 12 min. in Figure 6.2 (left) and on the chromatogram decomposed by ICA. Actual chromatographic management softwares generally proceed as follow when a coelution is detected: the retention times of the apexes ($\mathrm{A}_1, \mathrm{A}_2$) of the peaks are accurately recorded (top) but the times of beginning and end of the peaks are not. Generally, these times are assumed similar for peak 1 and 2. Thus, they are recorded here as B_{12} and E_{12}. Thus, they are recorded as if they were part of only one peak. Sometimes, if a valley is visible between the peaks, the time of this valley could be taken for the end of the first peak (E_1) and the beginning of the second (B_2). The problem of such practice is that the separation ($S_{crit} = \mathrm{B}_2 - \mathrm{E}_1$) computed from these times will be 0, although peaks are not separated. Furthermore, other information such as the area under the curve (AUC) of both peaks is simply not computable.

The solution provided by ICA is definitively better as it allows recording properly the retention times (bottom). As the peaks are numerically separated (atenolol (blue), warfarin (red)), one can work on the components to easily identify the beginning, apex and end of each peak.

Finally, it is noted that ICA may results in a slight change of the shape of the

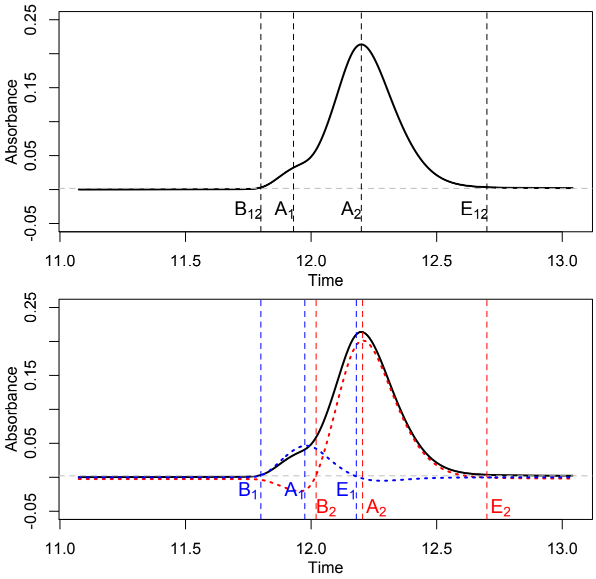


Figure 6.10: Example of recording of retention times (observed at 240 nm). (top) Inaccurate manual recording made on the raw chromatogram. (bottom) Improvement of the recording after ICA decomposition of the chromatogram. Black thick line: original chromatogram; blue: ICA-reconstructed peak corresponding to atenolol (peak 1); red: ICA-reconstructed peak corresponding to warfarin (peak 2).

sources corresponding to coeluted compounds. Indeed, in Figure 6.10 (bottom) the negative absorbance at the extremities of the peaks is not present in the real signals and is due to a perfectible unmixing. This has fortunately little importance on the better identification of the retention times, but could be more or less problematic when computing an AUC (see Chapter 10).

6.7 Conclusions

The automated peak picking is a very crucial step in the automated development of analytical methods. A new and original approach combining ICA, high-order statistics and clustering methodologies was successfully used for the data treatment of a real and blinded test mixture of pharmaceutical compounds in the framework of a DOE methodology. The present approach can also be envisaged in high throughput screening experiments consisting in the analysis of complex matrices with numerous and unknown compounds, which is often problematic. Moreover, it does not require expensive equipment to detect the compounds, such as a mass spectrometer, as long as the compounds absorb in the UV and have different UV-signatures.

The first clustering method allows the efficient separation of the noise components from the relevant ones, using adequate summary statistics. The technique to find an optimal number of sources is very convenient although time consuming. Fortunately, the time needed for the numerical data treatments is much smaller than the mixture analysis time. Therefore, this gives the opportunity to easily implement the numerical data treatments in a semi-concurrent mode, i.e. just after each chromatogram recording.

On the other hand, this process can also be performed only on sub-parts of a complex DAD-chromatogram. For instance, searching for coeluted impurities only around peaks of interest would allow for a significant reduction in the time devoted to the computational process. ICA can also be used as a simple noise removal algorithm. Rebuilding the DAD-chromatogram using only the relevant components allows the improvement of the signal-to-noise ratio and the cleaning of artifacts such as the baseline drift.

Finally, a strategy was proposed to automatically identify the peaks in unknown mixtures (peak tracking), solely based on the ICA rebuilt UV-signature of the compounds. Again, clustering methodologies carried out on some similarities/dissimilarities were found helpful to automatically and accurately group together the similar peaks.

Eventually, the combination of the original aforesaid strategies described in this chapter provides a powerful tool and opens new perspective for the automated development of LC–UV-DAD methods. Chromatographic management softwares and analysts would definitely improve their skills and capacities using such methodology, even in using this methodology in a semi-automatic way.

The only restriction to use this methodology is the following: the different compounds must have different UV-signatures to allow the tracking to work properly. Improvement can be obtained if a low number of UV-signatures are identical (e.g. enantiomeric compounds). This requires a different post-processing of the components that contains two or several peaks.

Part II

Applications

Chapter 7

Generic optimization of chromatographic methods to separate anti-paludic drugs

Parts of this chapter have been issued as an original publication by Debrus, Lebrun et al. (2011c). This chapter introduces the chromatographic method development with simple statistical models to illustrate the main concepts.

7.1 Introduction

Malaria remains one of the most extended illnesses worldwide. According to the World Health Organization (2010), almost 3.3 billion peoples scattered across hundreds of countries (mostly in the inter-tropical belt) are at risk of various species of plasmodium. Despite numerous active antimalarial molecules, several reasons are behind malaria resurgence. First, over past decades, the improper use of antimalarial drugs contributed to widen resistance against malaria parasite to several drugs (see Laufer and Plowe, 2004; Cowman and Foote, 1990). Nevertheless, artemisinin-based combination therapies (ACT) bring new hopes in the fight against malaria (Mutabingwa, 2005; Obonyo et al., 2007; Valecha et al., 2009; Singh et al., 2011). The second reason is the counterfeit of medecines. Marini et al. (2010a) observed that, in some African countries, up to 80% of medical products are counterfeit. In this context, analytical chemistry and especially chromatographic methods can help to fight this problem.

Screening analytical methods are usually used for several purposes. They can be used to confirm if a targeted active antimalarial ingredient (AAI) is present or ab-

Compounds	**pKa**		**logP**
	Most acidic	**Most Basic**	
Amodiaquine	9.43±0.50	5.62±0.50	3.126±0.840
Arteether	NA	NA	3.330±0.864
Artemether	NA	NA	2.820±0.864
Artemisinin	NA	NA	2.269±0.680
Artesunate	4.28±0.17	NA	3.291±0.883
Atovaquone	5.01±0.10	NA	6.465±0.729
Chloroquine	NA	10.47±0.25	4.412±0.758
Cinchonine	12.98±0.20	9.33±0.70	2.788±0.415
Dihydroartemisinin	12.61±0.70	NA	2.190±0.859
Halofantrine	13.57±0.20	9.44±0.50	8.902±1.302
Lumefantrine	13.44±0.20	8.71±0.50	8.671±0.405
Mefloquine	12.81±0.20	9.24±0.10	2.197±1.149
Piperaquine	NA	8.92±0.50	6.796±1.413
Primaquine	NA	10.38±0.10	2.740±0.255
Proguanil	NA	11.15±0.10	2.485±0.263
Pyrimethamine	NA	7.18±0.10	2.750±0.328
Quinine	12.80±0.20	9.28±0.70	2.823±0.431
Sulfadoxine	6.16±0.50	2.18±0.10	0.460±0.419
Sulfalene	6.61±0.40	1.46±0.10	0.880±0.456

Table 7.1: Pka and logP (at 25 °C) found on SciFinder®, calculated using Advanced Chemistry Development (ACD/Labs©) Software V11.02.

sent in a pharmaceutical formulation and to settle if an unappointed AAI is present or absent in this formulation. In this perspective, a screening HPLC method could be very helpful justifying the development of a generic method for the follow-up of the various antimalarial drugs available on the market. The proposed HPLC-UV method was thus developed for the screening of 19 active antimalarial ingredients (AAI)s: amodiaquine (AQ), arteether (AE), artemether (AM), artemisinin (ART), artesunate (AS), atovaquone (AT), chloroquine (CQ), cinchonine (CC), dihydroartemisinin (DHA), halofantrine (HF), lumefantrine (LF), mefloquine (MQ), piperaquine (PPQ), primaquine (PQ), proguanil (PG), pyrimethamine (PM), quinine (QN), sulfadoxine (SD) and sulfalene (SL). Chemical structures are presented on Figure 7.1. Calculated pKa and logP are given in Table 7.1.

Nowadays, HPLC method development can be achieved using different methodologies. Some have already led to some commercial softwares (e.g. Drylab, ACD/LC simulator, Chromsword, Osiris). These softwares make use of chromatography-based theory such as solvophobic theory, linear solvent strength relationship, etc., to optimize the separation of sample mixtures while maintaining the number of test experiments to a minimum. See Horváth et al. (1976); Carr et al. (1993); Nagrath et al.

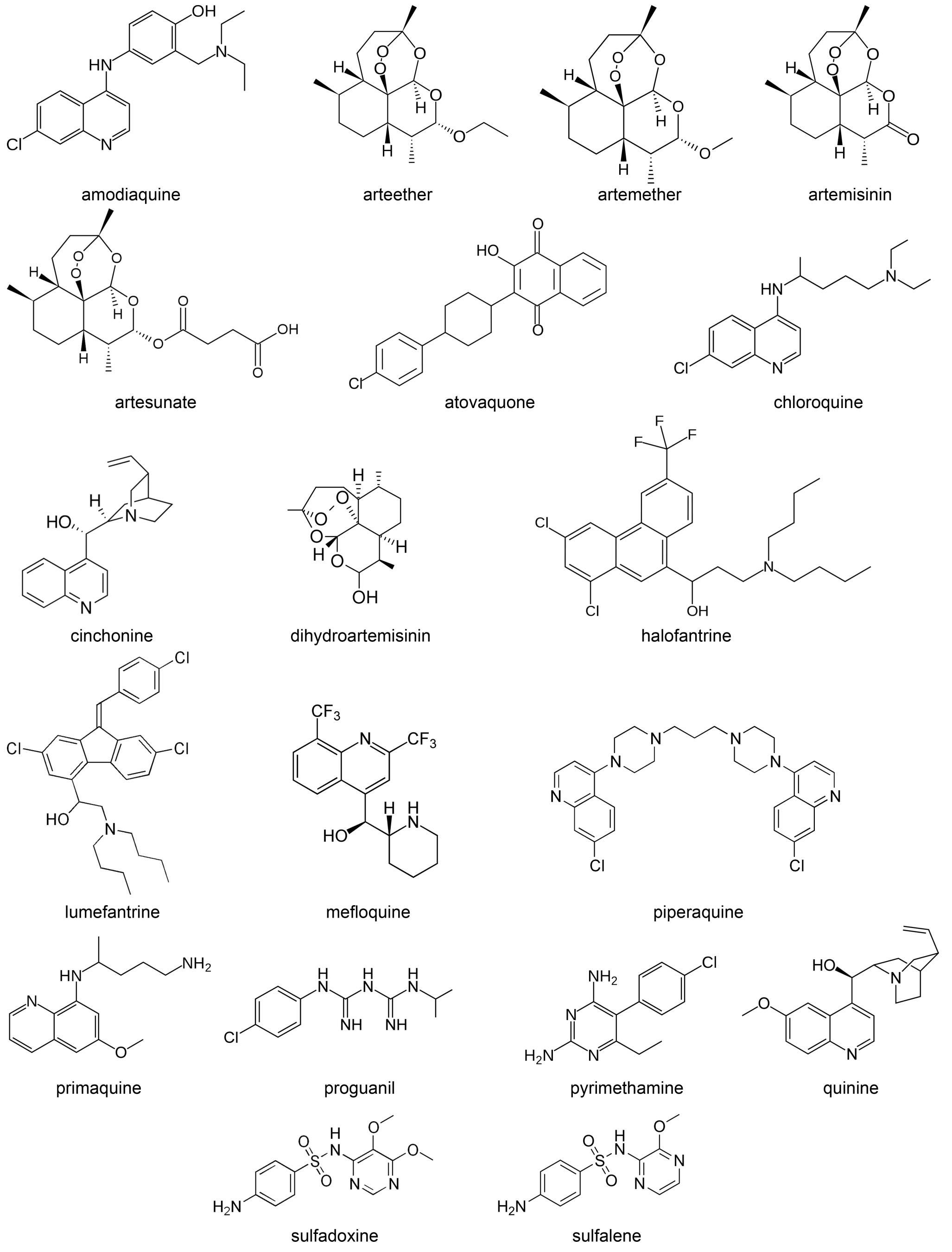


Figure 7.1: Chemical structures of the 19 studied antimalarial drugs.

(2011); Snyder et al. (1979, 1989); Nikitas and Pappa-Louisi (2009); Vivó-Truyols et al. (2003). These strategies are generally fast and efficient. Nevertheless, in the current trend of Quality by Design (QbD), these softwares do not provide the ability to advisedly compute or estimate the robustness, also sometimes called ruggedness in some regulatory documents. Consequently, Dejaegher and Vander Heyden (2007), Ragonese et al. (2000) and Hund et al. (2000) proposed classical robustness tests that have to be carried out at the end of the method validation phase to estimate the method ability to remain unaffected by small, but deliberate variations in method parameters.

In the present work, a distinct and innovative methodology combining design of experiments (DoE), Independent Component Analysis (ICA) and ICH Q8 (2009) Design Space (DS) was used to simultaneously optimize the separation and estimate the method robustness over the whole experimental domain instead of around the optima only. In the ICH pharmaceutical development guidelines Q8(R2), the 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". Therefore, the multidimensional combination and interaction of input variable should correspond to a subspace, the DS, where assurance of quality has been proved. Thus, the DS is necessarily encompassed within the experimental domain which is the multidimensional space formed by the factor ranges used during method development. The main concepts lying behind the ICH Q8(R2) definition of DS are assurance of quality and quality risk management. Hence, a HPLC method development process which do not take into account the errors from the process, measurements and models in order to manage the risk cannot be considered as QbD compliant and will not allow the identification or computation of the DS.

Even if ICH Q8 is the most appropriated guideline when discussing about QbD and DS applied to pharmaceutical sciences, the given DS definition and the examples shown in the Appendices of the document are divergent. From previous explanation, the identification of the multidimensional zone where a (mean) predicted Critical Quality Attribute (CQA) is within its specification does not define the DS which have to provide assurance of quality. In other terms, in LC, to predict the multidimensional region where the mean predicted resolution is acceptable (e.g. $R_s \geq 1.5$), does not define a DS as only quality is predicted but not assurance of quality. Indeed, the risk assessment of not being within the specifications is not carried out. On the other hand, using the posterior probability for a given or several CQA(s) to be in specifications is a better way to define DS (e.g. $P(R_s \geq 1.5 \mid \text{data})$). When computing such a probability, the quality risk management is carried out. Interesting discussions about QbD and DS applied in LC and in pharmaceutical development were already published in the scientific literature. See for instance the works of Peterson (2004); Peterson et al. (2009); Peterson (2008); Stockdale and Cheng (2009)). Some DoE-DS LC applications were also previously published by

Lebrun et al. (2008); De Backer et al. (2009); Krier et al. (2011); Debrus et al. (2011b).

The determination of the DS for a LC method development implies to consider the error on the studied responses and CQAs in a predictive framework. The variability of the retention times have to be taken into account during the development phase. These considerations hold completely with the QbD definition. Furthermore, ICH Q8 guideline states that "working within the Design Space is not considered as a change. Movement out of the Design Space is considered to be a change and would normally initiate a regulatory post approval change process". Consequently, the DS should define a zone of robustness as no significant changes in terms of separation quality should be observed on the resulting chromatograms.

DoE strategy can be considered, by some pure chromatographists, as a "black-box". Indeed, fitted mathematical models are only an approximation of the "true" chromatographic behavior of investigated compounds. Nevertheless, these models are very useful to identify and study the chromatographic behavior of compounds under investigation which can be unknown molecules. The most interesting advantage is that the DoE strategy is an overall data-driven methodology which does not necessarily imply preliminary knowledge of chromatographic behavior of compounds under consideration and/or the understanding of their chromatographic parameters before starting the optimization process. In some cases, the application of chromatographic theories will lead to very good results. But when molecules are unknown or when their chromatographic behaviors are hard to interpret, they can lead to unpredicted and inoperable results.

In this work, an innovative methodology using DoE and ICA was used to optimize the separation and identify the DS for the above mentioned AAIs. The present study is a useful and relevant application of complementary strategies previously published (Lebrun et al., 2008; Debrus et al., 2009). The first objective of the present work was to demonstrate the ability of the DoE-ICA-DS methodology to provide optimal and robust HPLC method. The second objective was to apply the methodology for the development of a useful method for the screening of 19 antimalarial drugs. This was also inscribed in the framework of the fight against counterfeit medicines.

7.2 Experimental

Sections "Chemicals and reagents", "Standard samples preparation" and details about "Experiments" can be found in Debrus et al. (2011c).

7.2.1 Design of experiments

In order to model the chromatographic behavior of each peak, a full factorial design was selected. It comprised three factors: pH of the aqueous part of the mobile phase (*pH*), gradient time (t_G) to linearly modify to proportion of methanol from 5% to 95%, and column temperature (T). Factors and their respective levels are summarized in Table 7.2. A total of 45 experimental conditions (5×3^2) were defined by this full factorial design. In the present case, a full factorial design was used to simultaneously optimize the method, estimate its robustness and evaluate the adequacy between chromatographic behaviors predicted by the theory of liquid chromatography and those obtained by the mathematical models. On the other hand, if method optimization is the unique objective, lighter DoE can be envisaged (e.g. fractional factorial or central composite designs).

Factors	**Levels**				
pH	2.5	4	6	8	10
Gradient time (tG, min)	20	40	60		
Temperature (T, °C)	25	30	35		

Table 7.2: Factors and levels of the full factorial design

The temperature was investigated to assess the robustness of the developed methods w.r.t this factors. It is known to have a limited separation effect. The reason is that the resulting screening HPLC method is intended to be used in different laboratories that frequently have no column temperature control. Moreover, the temperature should not be higher than 35 °C to avoid degradation of some unstable molecules.

Gradient time and pH ranges were expanded as much as possible in order to widen the experimental domain and to minimize the risk of not finding any separation within it. XBridge C18 analytical columns can sustain pH from 1 to 12. pH range was slightly narrowed from 2.5 to 10 in order to maintain suitable column lifetime. Gradient time range was also wide (from 20 to 60 min). These factors were selected to test out their effects on the selectivity and to shorten the time of analysis.

The experiments at a same pH were carried out in row for evident practical reasons. Then, the within pH blocks experiments were conducted in a random order. It is preferable to carry out the experiments in a totally random order to avoid experimental biases. Nevertheless, the column equilibration and conditioning times when constantly changing mobile phase pH drastically increase the time devoted to achieve the DoE results. Furthermore, the pH measurement error can be assumed to be equal to 0.1%. Other error sources (i.e. mobile phase composition during gradient, temperature, etc.) generated higher response errors and the pH blocking did not lead to poor predictive errors. pH blocking effect was thus considered as negligible in the present study.

The central point (i.e. $pH = 6.0$, $t_G = 40$ min, $T = 30$ °C) was independently run three times, including the preparation of new buffer and fresh mobile phase. The central points for lower and higher temperatures (i.e. $pH = 6.0$, $t_G = 40$ min, $T = 35$ °C and $pH = 6.0$, $t_G = 40$ min, $T = 25$ °C, respectively) were also carried out twice.

7.2.2 Independent Component Analysis and retention times recording

ICA is a statistical method allowing the numerical separation of sources maximizing the independence between them based on non-Gaussianity (Hyvärinen et al., 2001). In chromatography, ICA was already used by Debrus et al. (2009, 2011b) to numerically separate coeluting peaks in order to estimate their retention times (i.e. the times at the beginning and end of a peak). See Chapter 6 for details about ICA. Indeed, for coeluting peaks, when using a drop-line valley separator, the estimation of integration limits is highly biased. Then, the modeling of responses based on these biased times could lead to poor prediction accuracy. In order to avoid this situation, ICA was used to numerically separate coeluting peaks of antimalarial drugs.

The retention times of non–coeluting peaks were obtained manually on the chromatograms, as illustrated on Figure 7.2. From each peak in each chromatogram, three retention times can be extracted: the times at the beginning, at the apex and at the end of the peaks at baseline-height (t_B, t_R and t_E, respectively). T_0 denotes the dead time of the system, associated to the device and to the analytical column. The retention times for the n chromatograms and the m compounds can be stored in vectors $\mathbf{t}_{B,j}$, $\mathbf{t}_{R,j}$ and $\mathbf{t}_{E,j}$ of size n, $j = 1, ..., m$

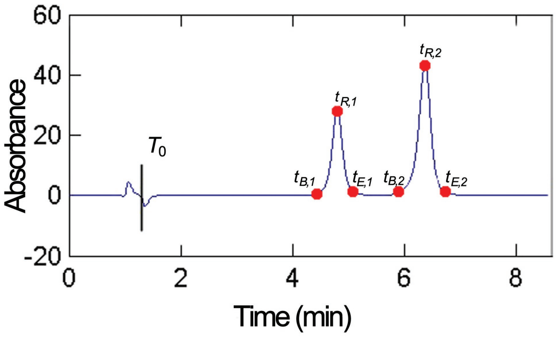


Figure 7.2: example of (artificial) chromatogram with the positions of discretized points: the retention times.

7.2.3 Critical Quality Attributes

Various attributes may be used to assess the chromatographic quality of an output of the method. $t_{B,j}$, $t_{R,j}$ and $t_{E,j}$ denote the retention times of the m peaks of a given chromatogram and $t_{B,(j)}$, $t_{R,(j)}$ and $t_{E,(j)}$, $j = 1, ..., m$ the ordered ones (with respect to the retention time of the apex). Each criterion cr can then be defined as a function of these retention times:

$$\begin{aligned}
cr_1 &= \text{Analysis time} = \max(t_{R,j}),\ \ j = 1, ..., m \\
cr_2 &= S_{crit} = \min(t_{B,(j+1)} - t_{E,(j)}), \quad j = 1, ..., m-1 \\
cr_3 &= \max(t_{E,j} - t_{B,j}),\ \ j = 1, ..., m \\
cr_4 &= R_{s,crit} = \min\left(\frac{2(t_{R,(j+1)} - t_{R,(j)})}{(t_{E,(j+1)} - t_{B,(j+1)}) + (t_{E,(j)} - t_{B,(j)})}\right),\ \ j = 1, ..., m-1 \\
cr_5 &= \max\left(\frac{\mid t_{E,j} + t_{B,j} - 2t_{R,j} \mid}{t_{E,j} + t_{B,j}}\right),\ \ j = 1, ..., m
\end{aligned} \tag{7.1}$$

Under these notations, the following interesting criteria are 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. They are expressed as follows:

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. Other criteria are also possible. In the next parts, only cr_1, cr_2 and cr_4 will be used.

7.2.4 Modeling and optimization methodology

Retention times modeling

Dewé et al. (2004) and Lebrun et al. (2008) provided a new approach for retention times modeling and DS computation. In these works, they showed that the modeling of the resolution can lead to poor prediction caused by its non-linear and non-continuous behavior when selectivity drastically changes.

The studied responses were the logarithm of the retention factor and the logarithms of both half-widths, computed as follows:

$$\mathbf{k}_{t_{R,j}} = \log\left(\frac{\mathbf{t}_{R,j} - T_0}{T_0}\right),\ \mathbf{w}_{l,j} = \log\left(\mathbf{t}_{R,j} - \mathbf{t}_{B,j}\right),\ \mathbf{w}_{r,j} = \log\left(\mathbf{t}_{E,j} - \mathbf{t}_{R,j}\right)$$

These $3m$ responses were modeled by individual multiple linear regressions using a stepwise approach to maximize the adjusted coefficient of determination (R^2_{adj}). The following model is then applied $3 \times m$ times. The reference to the particular response has been removed to simplify the notation.

$$\begin{aligned}\mathbf{y} &= \beta_0 + \beta_1 pH + \beta_2 pH^2 + \beta_3 pH^3 + \beta_4 t_G + \beta_5 t_G^2 + \beta_6 T + \beta_7 T^2 + \\ &\quad \beta_8 pH\ t_G + \beta_9 pH\ T + \beta_{10} t_G\ T + \beta_{11} pH\ t_G\ T + \boldsymbol{\varepsilon}, \\ &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},\end{aligned} \tag{7.2}$$

where $\mathbf{y}$ is either $\mathbf{k}_{t_{R,j}}$, $\mathbf{w}_{l,j}$ or $\mathbf{w}_{r,j}$. $\beta_0, \ldots, \beta_{11}$ are the regression parameters and $\boldsymbol{\varepsilon}$ are the residuals of the model, assumed Gaussian and i.i.d.. $\mathbf{X}$ is an $n \times p$ matrix containing the effects to be estimated. Least squares can be used to compute $\hat{\boldsymbol{\beta}}$, the estimator of $\boldsymbol{\beta}$.

Specification, error propagation and Design Space computation

A Critical Quality Attribute (CQA) was selected to assess the quality of a separation. Lebrun et al. (2008) proposed to use the separation criterion (S_{crit} or cr_2) defined as the difference between the beginning of the second peak and the end of the first peak of the critical pair. This separation criterion is easy to compute and to interpret. If $S_{crit} \geq 0$, the critical pair of peaks is baseline-resolved. An equivalent quality assessment is given by a critical resolution $R_{s,crit} \geq 1.5$, although most analysts generally agree that a minimum of $R_{s,crit} = 2$ is needed to ensure baseline-resolved peaks in case of asymmetry.

After the responses modeling, the responses were predicted using Equation (7.2). Predictions were then back-transformed into the original scale of the responses, i.e. the retention times.

For a new operating condition defining $\tilde{\mathbf{x}}$, included in the experimental domain χ, the predictive distribution obtained for the responses can be obtained from the marginalization properties of the multivariate Student's distribution proposed in Chapter 3 and Appendix D. Under non-informative priors, it is then defined as:

$$\tilde{y} \mid \tilde{\mathbf{x}}, \text{data} \sim t(\tilde{\mathbf{x}}\hat{\boldsymbol{\beta}}, a\left(1 + \tilde{\mathbf{x}}'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\tilde{\mathbf{x}}\right), n-p), \tag{7.3}$$

with $a = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$. This is related to the distribution obtained in the frequentist framework to compute prediction intervals, such as in Neter et al. (1990).

Practically, Monte-Carlo simulations were used to obtain, for a given operating condition, the predictive distribution of $\tilde{S}_{crit}$ from the predictive distributions of $\tilde{k}_{t_{R,j}}$, $\tilde{w}_{l,j}$ or $\tilde{w}_{r,j}$. Finally, using the distribution of $\tilde{S}_{crit}$, the posterior probability for $\tilde{S}_{crit}$ to be higher than a selected specification was used to determine the DS. In mathematical terms, the DS can be defined as

$$\text{Design Space} = \left\{ \tilde{\mathbf{x}} \in \chi \mid P(\tilde{S}_{crit} \geq 0 \mid \mathbf{X} = \tilde{\mathbf{x}}, \text{data}) \geq \pi \right\}. \tag{7.4}$$

where π is the quality level. In other words, the DS defines a subspace of the experimental domain where the posterior probability to obtain baseline-resolved peaks (i.e. $S_{crit} > 0$ min) is higher than a predefined quality level (e.g. $\pi = 65\%$).

In practice, it might be difficult to choose a value for π. No regulatory document yet provides guidelines on how to compute or estimate the Design Space quality level. Obviously, in this example, finding a DS with a quality level of 95% would be desirable to induce the achievement of a robust chromatographic method. A high π could only be obtained with a satisfactory mean separation as well as a relatively small predictive error on the involved responses.

To be able to draw DS even when the optimal π is not as high, the minimal quality level has been chosen a percent change of 85% of its optimal value. First, π^* is computed as the optimal value on every point of χ. Then, π is defined as 0.85 of π^*. For instance, if $\pi^* = 70\%$, the quality level will be $\pi = (0.85.\pi^*) = 59.5\%$.

Prediction of optimal separation

The experimental domain was investigated with a grid search method. A multi-dimensional grid was defined over the experimental domain. Then, the predictive distribution of the CQA was computed for each of the experimental condition defined by the grid. The optimum was selected as the point $\mathbf{x}^*$ of the grid giving the highest probability value (π^*), i.e.

$$\mathbf{x}^* = \max_{\tilde{\mathbf{x}} \in \chi} P(\tilde{S}_{crit} \geq 0 \mid \mathbf{X} = \tilde{\mathbf{x}},\ \text{data}).$$

In practice, the number of points was set as high as possible while keeping the total computing time lower than 12 hours (i.e. one night computation). On each point of the grid, 2500 Monte-Carlo simulations have been done to estimate the posterior probability.

7.2.5 Software

An in-house computer program was developed to perform the retention times modeling with stepwise multiple linear regressions, the Monte-Carlo simulations and the grid search method. The coding was carried out with R 2.11.1 (R Development

Core Team, 2010). ICA-based numerical separations were performed using FastICA algorithm implemented by Marchini et al. (2010).

7.3 Results and Discussions

7.3.1 Peak detection and peak matching

As the compounds from the artemisinin group (i.e. AE, AM, AS and DHA) present very similar and non-specific UV-signatures, these five molecules were injected individually to identify and match them. Then, for the rest of the compounds, in case of coelution, ICA was used to determine the times at the beginning (t_B), the apex (t_R, the retention time) and the end (t_E) of each peak. Finally, for the non-coeluting peaks, these times were manually read on the chromatograms.

Furthermore, at alkaline pH and high temperature ($pH = 10$, $T = 35$°C), DHA peak was split in two coeluting peaks and several unidentified peaks were also observed. These results suggested that DHA degraded in alkaline conditions at 35°C. Therefore, the experimental domain was reduced removing all the pH 10 experiments to avoid problems when recording the retention times.

7.3.2 Retention times modeling

Retention times modeling were achieved by stepwise multiple linear regressions which selected the terms of Equation (7.2) to maximize R^2_{adj}. As three times (t_B, t_R and t_E) of 19 peaks were modeled, 54 models were obtained. The estimated model parameters ($\beta_0, \ldots, \beta_{11}$), their complete chromatographic interpretation and a summary of the models comprising R^2_{adj} were presented in the original publication (Debrus et al., 2011c) but have been cut here for brevity.

Models adequacy

In order to visualize the models adequacy, Figure 7.3 displays the appropriateness between predicted and experimental data (a) as well as the corresponding residuals (b), represented in their original scale. The three rows of the Figure represent the retention times at the apex, the end and the beginning of the peaks. The m compounds are confounded and represented by different colors and shapes.

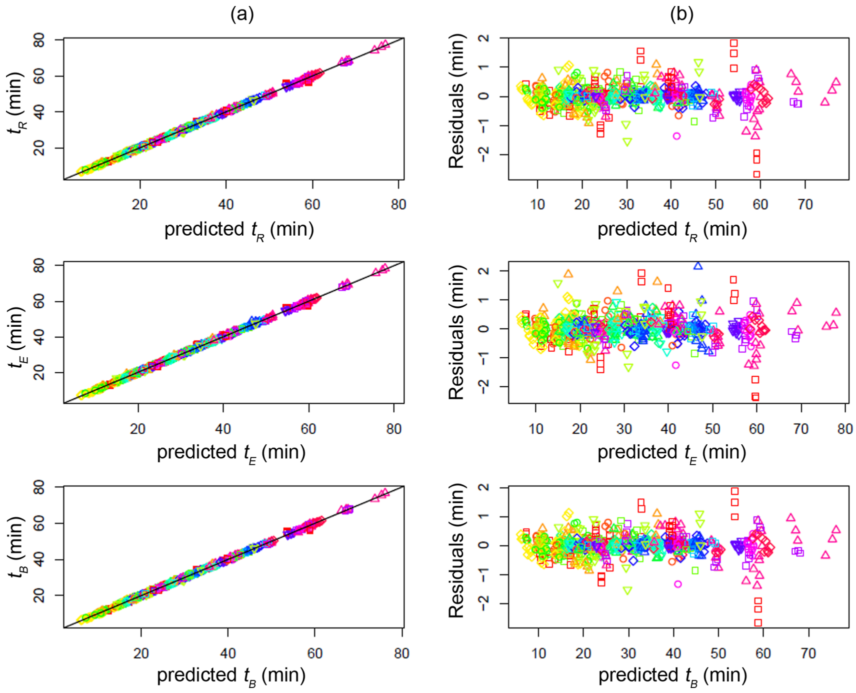


Figure 7.3: Predicted vs. observed plots and corresponding residuals. (a) Predicted versus experimental values for $\mathbf{t}_R$, $\mathbf{t}_E$ and $\mathbf{t}_B$. (b) Corresponding residuals plots. Compound assignation: (red square) PPQ, (red circle) CC, (orange head up triangle) CQ, (yellow diamond) SL, (green head down triangle) AQ, (green square) QN, (green circle) SD, (green head up triangle) PM, (green diamond) PQ, (blue head down triangle) PG, (blue square) MQ, (blue circle) ART, (blue head up triangle) DHA, (blue diamond) AS, (blue head down triangle) AM, (purple square) HF, (purple circle) AE, (purple head up triangle) LF, (purple diamond) AT

As observed on (b), the residuals are distributed between –2 and 2 min. The residuals standard deviations ($\hat{s}$) computed on the three sets of residuals were 0.36, 0.41 and 0.36 min, respectively. It is thus reasonable to assume that the averaged error on the retention times is about 0.8 min (i.e. 2 × standard deviation) over the whole experimental domain. Shapiro-Wilk Normality tests were also carried out on the residuals. The p-values were all higher than 0.05 which meant that the residuals are in agreement with a Normal distribution.

pH, Gradient time and temperature effects

The mean predicted t_R (back-transformed in their original scale) for the 19 anti-paludic drugs are shown on Figure 7.4 (top). For each subfigure, one operating condition is changed while the two others are fixed. As expected, pH is the factor that had the most significant effect on selectivity (crossing of peaks). The identifi-

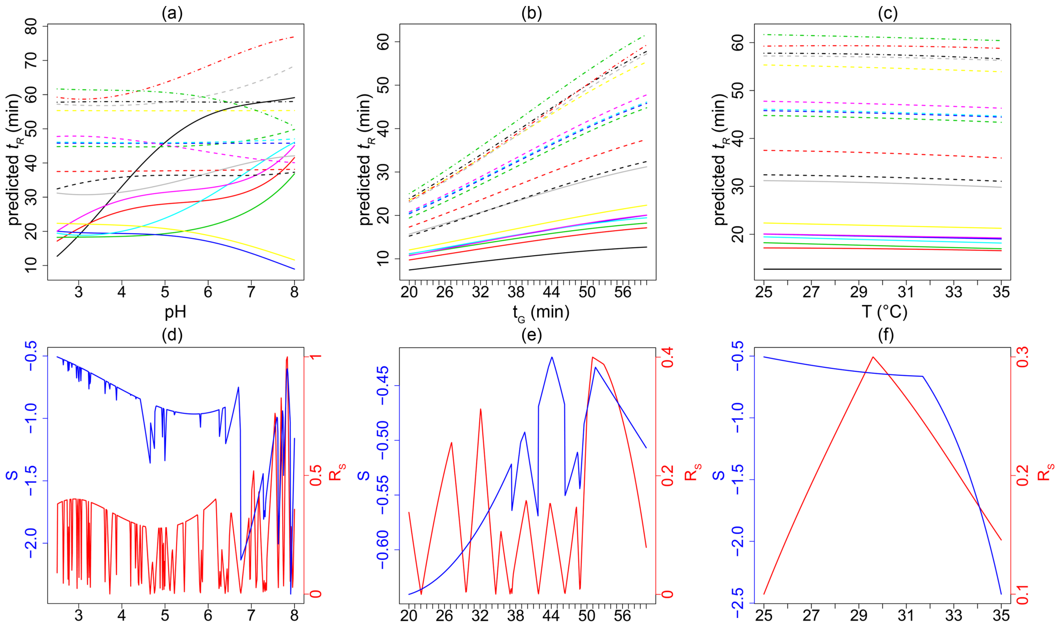


Figure 7.4: Predicted mean retention times (t_R) with respect to DoE's factors. (a) Predicted t_R (min) vs. pH – with $t_G = 60$ min and $T = 25$ °C. (b) Predicted t_R (min) vs. t_G (min) – with $pH = 2.5$ and $T = 25$°C. (c) Predicted t_R (min) vs. T (°C) – with $pH = 2.5$ and $t_G = 60$ min. Compound assignation: (black line) PPQ, (red line) CC, (green line) CQ, (blue line) SL, (cyan line) AQ, (magenta line) QN, (yellow line) SD, (grey line) PM, (dashed black line) PQ, (dashed red line) PG, (dashed green line) MQ, (dashed blue line) ART, (dashed cyan line) DHA, (dashed magenta line) AS, (dashed yellow line) AM, (dashed grey line) HF, (dotted-dashed black line) AE, (dotted-dashed red line) LF and (dotted-dashed green line) AT. (d,e,f) Corresponding predicted Critical Quality Attributes S_{crit} (blue line, left axis) and $R_{S,crit}$ (red line, right axis)

cation of neutral, acidic or basic compounds is also easy. Neutral compounds show no t_R variation with respect to pH. Acidic and basic compounds have a respective decreasing or increasing variation with respect to pH.

The t_G also induces changes in the retention time (b). However, the changes in selectivity is more limited. The temperature has the lowest effect (c). These low t_R variations with respect to T underline the method robustness while changing T as expected in routine use in Africa.

Notice that mean predicted t_B and t_E were also computed to predict the CQA at given operating conditions. (d), (e) and (f) show the behavior of the predicted S_{crit} (blue), compared to the more classical criterion of resolution $R_{S,crit}$ (red).

These non-continuous curves clearly illustrate the impossibility to envisage these attributes as model responses. This non-continuity is due to the selection of the critical pair of peaks (by min and max operators, see the Equations (7.1)), that might be different at each operating condition. The mean predicted CQAs are without appeal concerning a total separation of the 19 anti-paludic drugs. The best mean separation is about -0.45, i.e. there is at least two coeluting peaks.

7.3.3 Critical Quality Attributes and Design Space computation

The CQA separation (S_{crit}) was computed over the whole experimental domain. A grid of 42875 points (i.e. 35×35×35) was defined and S_{crit} was computed for each operating condition. The predictive distributions of the responses was used to generate samples whose uncertainty was propagated to S_{crit} using Monte-Carlo simulations. Then, the results were presented as probability surfaces (i.e. the posterior probability for the CQA to to be higher than its specification) rather than response surfaces.

The very similar chromatographic behavior between ART and DHA prevented obtaining a separation of the 19 AAIs. Nevertheless, at some operating conditions, they were the only two coeluting peaks. Thus, two groups were formed. The first contained 17 AAIs and ART (group 1) and the second contained the same 17 AAIs and DHA (group 2). These two groups are justified from a therapeutic point of view because ART and DHA are never present in the same pharmaceutical formulation.

The optimization process was then repeated for each group independently. The probability surfaces for $P(S_{crit} \geq 0)$ for group 1 and group 2 are presented in Figure 7.5 and Figure 7.6, respectively. Low optimal quality levels DS with $\pi = 28\%$ and $\pi = 8\%$ are depicted in red. They represent a better achievement of quality than in any other zone of the experimental design. It is important to keep in mind that a chromatogram with slightly coeluted peaks is still relevant when envisaging the screening of many compounds. Furthermore, these probabilities are representative of the separation of the critical pair of peaks solely. Then, every other separation will be at least better.

For a given quality level π, the DS shape is directly linked to the method robustness against factors variation. Therefore, the DS shapes on Figure 7.5 and Figure 7.6 allowed concluding that the resulting screening methods are still relatively robust with respect to modifications of t_G (from 54 to 60 min for group 1 and from 58 to 60 min for group 2) and T (from 25 to 25.8 °C for group 1 and from 25 to 26 °C for group 2). Conversely, these screening methods seems far less robust with respect to pH (from pH 4 to pH 4.1 for both groups). However, the pH measurement variability is estimated to 0.1%. Consequently, the relatively poor method robustness with respect to pH should therefore not be problematic when care is taken during the buffers preparation and pH measurements.

7.3.4 Prediction of optimal separation

As shown on Figure 7.5 and Figure 7.6, the optimization process gives quite identical optimal condition. The difference between quality levels for group 1 and 2 came from the shorter DHA retention times compared to ART. DHA slightly coeluted with MQ (Peak 12 on Figure 7.7) explaining the much lower π for group 2. Finally, an optimal operating condition was selected, lying inside both DS and allowing the separation of the 18 AAIs of both groups independently. At operating condition $pH = 4.05$, $t_G = 56.2$ min and $T = 25$ °C, the optimal $P(S_{crit} > 0) = \pi^*$ was 33% for group 1 and $\pi^* = 9.6\%$ for group 2.

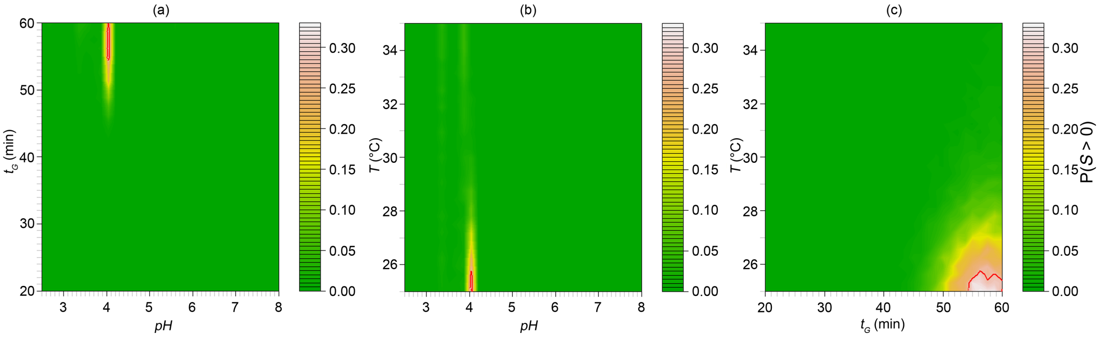


Figure 7.5: Posterior probability surfaces ($P(S_{crit} > 0)$) for group 1 separation. (a) Gradient time (min) vs. pH, (b) temperature (°C) vs. pH and (c) temperature (°C) vs. gradient time (min). The DS ($\pi = 28\%$) is encircled by a red line.

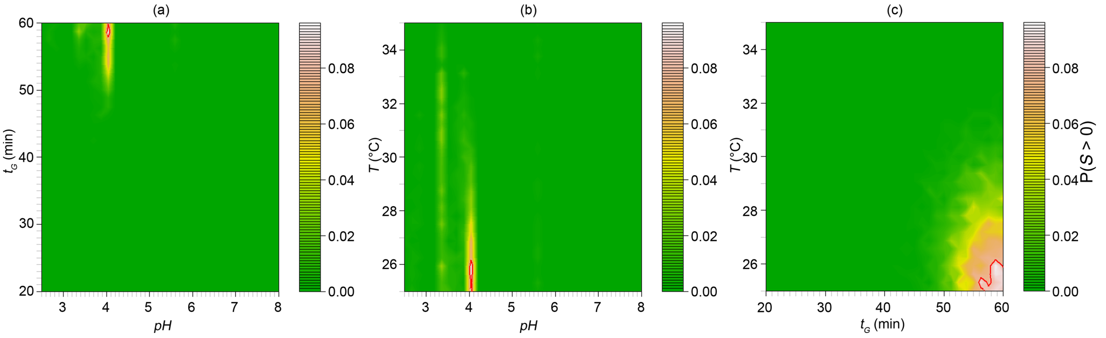


Figure 7.6: Posterior probability surfaces ($P(S_{crit} > 0)$) for group 2 separation. (a) Gradient time (min) vs. pH, (b) temperature (°C) vs. pH and (c) temperature (°C) vs. gradient time (min). The DS (π= 8%) is encircled by a red line.

Chromatograms recorded for the group 2 at the optimal operating condition are very close to those displayed in Figure 7.7 for group 1. Despite the inability to separate ART and DHA (inducing the creation of 2 groups), it was easy to identify them even if coeluting because their retention times were slightly different (i.e. $t_R =$ 44.5 min for ART and $t_R = 44.9$ min for DHA). Thus, it is still possible to separately inject a reference solution of each compound in working conditions to confirm their identification.

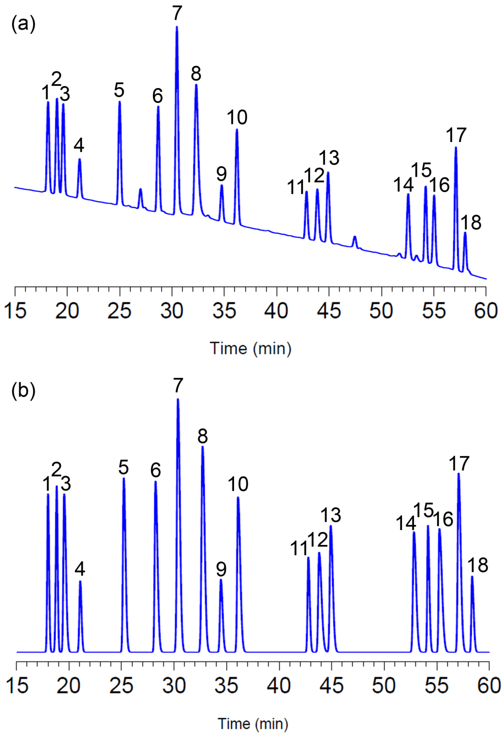


Figure 7.7: (a) Experimental chromatogram recorded at $pH = 4.05$, $t_G = 56.2$ min and $T = 25$ °C with group 1. (b) Predicted chromatogram at the same condition. Peak numbering: 1 = CQ, 2 = SL, 3 = AQ, 4 = SD, 5 = CC, 6 = QN, 7 = PM, 8 = PP, 9 = PQ, 10 = PG, 11 = MQ, 12 = ART, 13 = AS, 14 = AM, 15 = HF, 16 = AE, 17 = LM and 18 = AT.

7.3.5 Sub-mixture

One can observe that the screening method (Figure 7.7) has a quite long analysis time. It can be also observed that the compounds eluted between 15 and 60 min. The first 15 min are not "used" to separate the compounds in a shorter time.

Nevertheless, the DoE-DS methodology can also be used to develop methods aiming at reducing the analysis time while optimizing the separation of specific sub-mixtures of compounds. Some compounds (related to a given pharmaceutical formulation) were therefore selected to test out this opportunity without performing any additional experiments. Indeed, once models are available to explain the compound chromatographic behaviors, the optimization process for the separation of any specific compound combinations can be carried out. The resulting DS would directly depend on the selected compound involved in the computations. The selected sub-mixture contained AS, PM and SL. This combination is representative of a pharmaceutical formulation present on the Democratic Republic of the Congo market (Arte-Plus®).

In order to minimize the time of analysis while simultaneously optimizing the separation of these 3 compounds, a multi-criteria optimization was carried out. The selected CQAs were the critical separation S_{crit} (cr_2) and the time of analysis (cr_1). The specification were placed at 0.1 min for S_{crit} and 20 min for the total run time.

The DS is defined as the subspace where the posterior probability $P(\tilde{cr}_2 \geq 0.1, \tilde{cr}_1 \leq 20\text{min} \mid \text{data})$ is higher than the selected quality level. Here, $\pi = 97.5\%$, as shown on Figure 7.8, and $\pi^* \approx 1$.

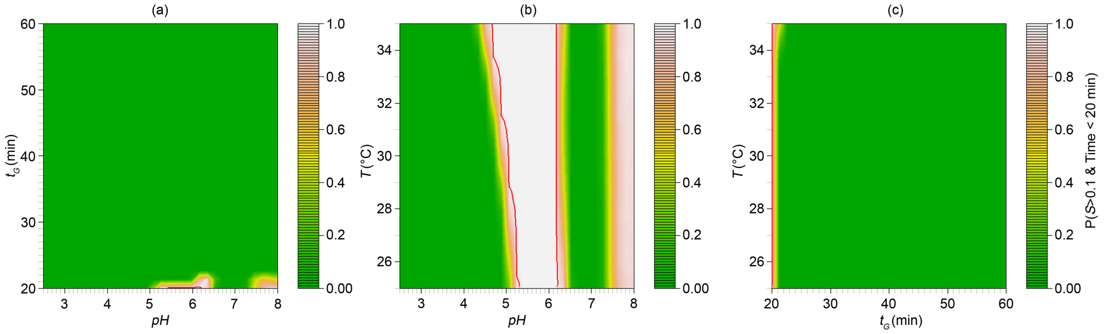


Figure 7.8: Posterior probability surfaces (i.e. $P(S_{crit} \geq 0.1\text{min}$ & analysis time $\leq 20\text{min}$)) for sub-mixture separation. (a) Gradient time (min) vs. pH, (b) temperature (°C) vs. pH and (c) temperature (°C) vs. gradient time (min). The DS ($\pi = 97.5\%$) is encircled by a red line.

One can observe that the DS is really small in the gradient time t_G space. The reason behind this observation is the closeness of the last peak with the analysis time specification (i.e. 20 min). The lowest t_G must be selected as it has the highest effect on compounds retention times. For the other factors, the DS is bigger. The specific method is then very robust in terms of separation. The optimal operating condition is $pH = 5.4$, $t_G = 20$ min and $T = 35$ °C. The recorded chromatogram at the optimal condition is displayed in Figure 7.9 (a). Notice that analysis time could certainly be reduced by decreasing t_G. In this case, the operating conditions would be outside the experimental domain, which is not advised.

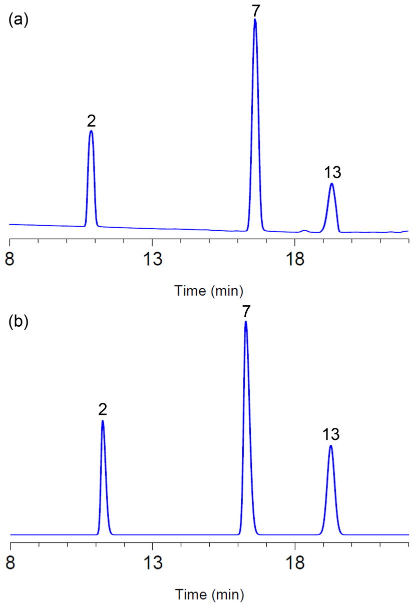


Figure 7.9: (a) Experimental chromatogram recorded at pH 5.4, t_G = 20 min and T = 35 °C with sub-mixture compounds. (b) Corresponding predicted chromatogram. Peak numbering: 2 = SL, 7 = PM, 13 = AS.

Figure 7.9 presents the adequacy between the predicted (b) and observed (a) chromatograms. It is quite obvious that a method development (using already available commercial softwares or a lighter full factorial design) for the separation of these three compounds would be able to find operating condition giving a good separation in a shorter analysis time. However, the present method optimization dedicated to the pharmaceutical formulation (Arte-Plus®) was carried out from the same data that those used for the optimization of the screening method. It underlined the fact that the DoE-DS methodology is generic. Here an innovative methodology was used to optimize the separation while simultaneously estimating the robustness of either a general screening method or specific methods.

7.4 Conclusions

In the current trend of being QbD compliant, it is of first importance to develop methodologies that provide robust optimal separations defined by a DS. With regards to this objective, DoE-ICA-DS allowed obtaining optimal screening methods for the separation of 19 AAIs. The methods robustness was evaluated thanks to the DS quality level, DS shape and the assessment of the factors effects. It resulted that the obtained screening methods were robust against temperature modification. This result is very important when one of the final aims is a method transfer to African laboratories where column ovens are not always available.

Theses screening methods can be considered as a step forward for the fight against counterfeit medicines. The present work also allowed demonstrating the ability of the DoE-ICA-DS methodology to encounter optimal separation for complex mixtures (i.e. containing compounds with very similar structures and physico-chemicals properties). This study also demonstrated the ability of fitted mathematical models to be used to identify and corroborate theoretical chromatographic behaviors of studied compounds (not presented here). It highlighted the fact that DoE strategy can be a very useful tool for chromatographists in order to develop or refine the understanding of some chromatographic behavior.

Furthermore, the separation of a three compound mixture was also carried out without performing any additional experiments. The resulting method was also very robust to temperature changes. Finally, the results presented in this manuscript strengthen the fact that DoE-ICA-DS can be considered as a generic QbD compliant methodology for the optimization and the robustness assessment of new chromatographic methods for the analysis of pharmaceutical formulations or more complex matrices such as plant or biological materials.

Chapter 8

Application of an innovative Design Space optimization strategy to the development of LC methods to combat potentially counterfeit NSAIDS

8.1 Introduction

Non-steroidal anti-inflammatory drugs (NSAIDs) are used against pain, fevers of various origins and inflammation (Ibrahim et al., 2007; Dinç et al., 2002; McMahon et al., 1998). Although they are widely prescribed throughout the world, many of them are associated with side effects. These drugs are often used in self-medication, as their purchase is also unrestricted over the Internet. The risk of administration of uncontrolled medicines is thus naturally greater.

Furthermore, NSAIDs are often subject to the practice of counterfeit medicines, which is gaining increasing momentum in the world and particularly in low-income populations. This has adverse consequences for public health (Panusa et al., 2007). The World Health organization (WHO) reported 6% of drugs worldwide are counterfeit and the Food and Drug Administration (FDA, USA) estimated this proportion to be 10% (Mazières, 2007). This proportion varies from one country to another. In some African countries, Marini et al. (2010a) estimated that up to 80% of medical products are counterfeit (see also Marini et al., 2010b). To ensure the quality of drugs and to help battle counterfeit medicines, the development of screening methods that can simultaneously trace many of the most commonly used molecules is an important effort. In this context, separation techniques are the usual option when planning to explore a substantial number of molecules. Conversely, while techniques such as near infrared or Raman spectroscopy are simpler and quicker procedures and require no sample preparation, their potential still remains limited when dealing with the sample analysis of complex mixtures of active ingredients or their impurities at low concentration levels.

Several liquid chromatographic (LC) methods are described in the literature for analyzing NSAIDs (Acuña et al., 2004; Ji et al., 2005; Dinç et al., 2002; Boonkerd et al., 1995; Altun et al., 2001; Franeta et al., 2002; Kartal, 2001; Šafra and Pospíšilová, 2008; Chen and Wu, 2005; Santini et al., 2007; Polásek et al., 2000) However, none of these includes an exhaustive list of the molecules present in NSAIDs. Thus, they are of limited use in the context of complex or unknown mixtures screening. For instance, Iuliani et al. (2010) optimized a LC for separating only seven NSAIDs regardless of major pharmaceuticals products such as naproxen, diclofenac, etc., or other major molecules often associated to NSAIDs such as paracetamol, caffeine, etc.

In the present study, several HPLC separation conditions were optimized for targeted subsets of 27 molecules used alone or in combination. The first objective was the optimization of the separation conditions for these 27 molecules among which were 18 NSAIDs: ibuprofen (IBU), diclofenac (DIC), mefenamic acid (MA), ketoprofen (KTO), nimesulide (NIM), dextropropoxyphene (DEX), niflumic acid (NA), tenoxicam (TE), piroxicam (PI), sulindac (SUL), phenylbutazone (PHE), flurbiprofen (FU), suprofen (SUF), naproxen (NAP), tiaprofenic acid (TA), phenoprofen (i.e. fenoprofen, PF), indomethacin (IDO) and acetylsalicylic acid (AA). Molecules often associated with NSAIDs were added to the list: chlorzoxazone (CHL), caffeine (CAF), paracetamol (i.e. acetaminophen, PAR) and salicylic acid (SAL), a degradation product of AA. Finally, five pharmaceutical conservatives found in syrups or suspensions were concurrently analyzed with the rest: nipagine (i.e. methylparaben, NIP), nipasol (i.e. propylparaben, NIS), butylated hydroxyanisole (BHA), butylated hydroxytoluene (BHT) and sodium benzoate (BEN).

To achieve this objective, design of experiments (DoE) was used to establish a Design Space (DS), extending the works in Lebrun et al., 2008.

According to ICH Q8 (2009) the DS is “the multidimensional combination and interaction of input variables (e.g. material attributes) and process parameters that have been demonstrated to provide assurance of quality”. Thus, the DS is a subspace of the experimental domain where assurance of quality has been proven. In the present study, the DS defines the space of HPLC operating conditions that will ensure quality outputs. The minimal expected quality might be described by acceptance criteria ($\mathbf{\Lambda}$) that apply on some critical quality attributes (CQAs), i.e. values or indices that provides some indications about the overall achievement of the analytical method. In chromatographic terms, CQAs may be the resolution ($R_{s,crit}$) or the separation (S_{crit}) of a critical pair of peaks, and the run time of the method (t_{tot}), while the acceptance criteria may be $S_{crit} > 0$ and $t_{tot} < 45$ min considered concurrently.

In this context, a result given as a predictive probability that the CQAs will fall within acceptance criteria allows the quantification of the assurance advocated by ICH Q8 (R2) (Debrus, Lebrun et al., 2011c). This leads to a risk-based definition of the DS that may be expressed as:

$$\text{Design Space} = \{\tilde{\mathbf{x}} \in \chi \mid P(\mathbf{CQAs} \in \mathbf{\Lambda} \mid \tilde{\mathbf{x}}, \text{data}) \geq \pi\}\,. \tag{8.1}$$

In other words, the DS is a region of an experimental domain χ (often called knowledge space) where the posterior probability that the CQAs are within acceptance criteria $\mathbf{\Lambda}$, is higher than a specified quality level π , conditionally on the available data. By the use of the posterior predictive distribution of the CQAs, the posterior probability accounts for the parameter uncertainties and interactions estimated by the statistical model, as well as residual variability (Peterson, 2008).

Note that if a large and high quality DS is identified within the experimental HPLC parameter ranges, the corresponding optimized method may be considered robust, as deliberate changes in the operating conditions (included in the defined DS) will not negatively impact the quality of the output.

In order to provide faster analysis, a transfer to ultra high performance liquid chromatography (UHPLC) was the second part of the planned study. This leads to a shortened run time and reduced solvent consumption, which results in reduced time and cost to identify substandard or counterfeit medicines in laboratories where UHPLC systems are available (Sacré et al., 2011). For that purpose, the robustness of the methods firstly developed with conventional HPLC system was found mandatory to permit the geometric transfer. Indeed, the induced variability when changing from one LC system to another may lead to small changes in the retention times of the analytes. Logically, this occurrence could be more pronounced when moving from a conventional LC system to a UHPLC system. Also, it is noted that the UHPLC method may suffer from a small loss in peak efficiency due to the drastic reduction of the analysis time. Consequently, the objective in this context was also to demonstrate that the knowledge coming from the built DS (i.e., robustness area) could facilitate the geometric transfer even if several facts might potentially impact on the separation quality of the transferred analytical methods.

To demonstrate the DS ability, the third objective was to validate one of the developed HPLC methods, using the accuracy profile as decision tool for the determination of four compounds (Hubert et al., 2004; Boulanger et al., 2003). A common NSAIDs combination marketed in some African countries was used. It consisted of capsules containing paracetamol, ibuprofen, caffeine and potentially 4-aminophenol, a well known impurity of paracetamol allowing to obtain information on the storage conditions.

Finally, the quantitative method was applied to analyze five drugs marketed in the Democratic Republic of Congo (DRC).

8.2 Theory

When CQAs show a highly nonlinear and discontinuous mathematical behavior with respect to changes of operating conditions, modeling of responses that allow a satisfactory fit with simple regression models has been recommended (Peterson, 2004; Dewé et al., 2004). This condition is likely to be observed with CQAs such as the resolution or the separation of a critical pair of peaks due to the inversions of elution order that may be observed. After their modeling, the selected responses must be sufficient to derive the CQAs in order to assess the quality of the HPLC system output.

In this case the retention times at the beginning ($\mathbf{t}_{B,j}$), at the apex ($\mathbf{t}_{R,j}$) and at the end ($\mathbf{t}_{E,j}$) of every j^{th} peak (j =1,. . . ,m) result in $3m$ vectors of n data for the n chromatograms resulting from the designed experiment. Next, the logarithms of defined retention factors are taken as modeled responses:

$$\mathbf{k}_{B,j} = \log\left(\frac{\mathbf{t}_{B,j} - t_0}{t_0}\right), \; \mathbf{k}_{R,j} = \log\left(\frac{\mathbf{t}_{R,j} - t_0}{t_0}\right), \; \mathbf{k}_{E,j} = \log\left(\frac{\mathbf{t}_{E,j} - t_0}{t_0}\right), \tag{8.2}$$

where t_0 is the dead time of the LC system.

This defines $\mathbf{Y} = (\mathbf{k}_{B,1}, \mathbf{k}_{R,1}, \mathbf{k}_{E,1}, ..., \mathbf{k}_{B,m}\mathbf{k}_{R,m}\mathbf{k}_{E,m})$, the $(n \times 3m)$ matrix containing the responses that are modeled jointly using a multivariate regression model:

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E}, \tag{8.3}$$

where the vector ε_i, the i^{th} line of $\mathbf{E}$ is assumed independent and identically distributed as multivariate Normal, i.e., $\varepsilon_i \sim N(\mathbf{0}, \mathbf{\Sigma})$, $i = 1, ..., n$. $\mathbf{X}$ is the $(n \times p)$ centered and reduced design matrix containing the p effects to be included in the model (see later) and $\mathbf{B}$ is the $(p \times 3m)$ matrix containing the p parameters for each of the $3m$ responses. The modeled effects in $\mathbf{X}$ are then similar for every response and should be chosen according to the particular experimental design. $\mathbf{\Sigma}$ is the covariance matrix of the residuals.

In order to account for the variability of the parameters $\mathbf{B}$ and $\mathbf{\Sigma}$, a posterior predictive density of new predicted responses can be obtained in the Bayesian framework. Considering an informative prior distribution of the parameters $p(\mathbf{B}, \mathbf{\Sigma}) = p(\mathbf{B} \mid \mathbf{\Sigma})p(\mathbf{\Sigma})$, with $p(\mathbf{\Sigma})$ distributed as an inverse-Wishart W_1^{-1} and $p(\mathbf{B} \mid \mathbf{\Sigma})$ distributed as a $(p \times 3m)$ matrix-variate Normal $N(\mathbf{B}_0, \mathbf{\Sigma}, \mathbf{\Sigma}_0)$, the predictive density of a new predicted set of responses $\tilde{\mathbf{y}} \mid \tilde{\mathbf{x}},$ data for a given operating condition $\tilde{\mathbf{x}} \in \chi$ follows a multivariate Student's distribution (Minka, 2001; Lebrun et al., 2012a, Appendix A):

$$\tilde{\mathbf{y}} \mid \tilde{\mathbf{x}}, \text{data} \sim T_m \left(\tilde{\mathbf{x}}\mathbf{M}_{\mathbf{B}\text{post}}, \left(1 + \tilde{\mathbf{x}}'(\mathbf{X}'\mathbf{X} + \mathbf{\Sigma}_0^{-1})^{-1}\tilde{\mathbf{x}}\right)(\mathbf{\Omega} + \mathbf{A}^*), \nu + n_0\right), \tag{8.4}$$

where $\mathbf{M}_{\mathbf{B}\text{post}} = \left(\mathbf{X}'\mathbf{X} + \mathbf{\Sigma}_0^{-1}\right)^{-1}\left(\mathbf{X}'\mathbf{X}\hat{\mathbf{B}} + \mathbf{\Sigma}_0^{-1}\mathbf{B}_0\right)$; $\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ and $\mathbf{A}^*$ is as follows:

$$\mathbf{A}^* = \mathbf{Y}'\mathbf{Y} + \mathbf{B}_0'\mathbf{\Sigma}_0^{-1}\mathbf{B}_0 - (\mathbf{X}'\mathbf{X}\hat{\mathbf{B}} + \mathbf{\Sigma}_0^{-1}\mathbf{B}_0)'(\mathbf{X}'\mathbf{X} + \mathbf{\Sigma}_0^{-1})^{-1}(\mathbf{X}'\mathbf{X}\hat{\mathbf{B}} + \mathbf{\Sigma}_0^{-1}\mathbf{B}_0). \tag{8.5}$$

Notably, when $\mathbf{\Sigma}_0^{-1}$ tends to 0, $\mathbf{A}^*$ is simply the residual sum of squares and $\mathbf{M}_{\mathbf{B}\text{post}}$ becomes $\hat{\mathbf{B}}$ the least square estimate of $\mathbf{B}$. Furthermore, when $\mathbf{\Omega} = \mathbf{0}$ and $n_0 = 0$, the predictive density is similar to that which is obtained using a simpler form of non-informative prior (see e.g. Press, 2003). The definition of the parameters of the prior distributions, i.e. the particular values that may be given to $\mathbf{\Omega}, n_0, \mathbf{\Sigma}_0^{-1}$ and $\mathbf{B}_0$ will be discussed later.

The proposed predictive distribution is of particular interest because it describes how the HPLC method will perform during future use, given the data and the prior information. This prior information should be as uninformative as possible when limited knowledge is available.

For this study, the CQAs computed from the modeled responses are the separation and the total run time. Using $\mathbf{t}_{B,j}$, $\mathbf{t}_{R,j}$ and $\mathbf{t}_{E,j}$ $(j = 1, \ldots, m)$ to denote the retention times of the j^{th} peak of a given chromatogram, and $\mathbf{t}_{B,(j)}$, $\mathbf{t}_{R,(j)}$ and $\mathbf{t}_{E,(j)}$ to denote the retention times of the j^{th} peak with respect to the time of the apexes $\mathbf{t}_R$, the CQAs are defined as follows:

$$\begin{aligned}\mathbf{t}_{tot} &= \text{Analysis time} = \max(\mathbf{t}_{R,j}), \ \ j = 1, ..., m,\\ \mathbf{S}_{crit} &= \text{Critical Separation} = \min(\mathbf{t}_{B,(j+1)} - \mathbf{t}_{E,(j)}), \quad j = 1, ..., m-1.\end{aligned} \tag{8.6}$$

Both CQAs are expressed in minutes. Actually, the run time $\mathbf{t}_{tot}$ is defined as the retention time of the most retained compound for the different operating conditions. Both CQAs are determined 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 as good than these values.

Using Monte-Carlo simulations, one can sample from the predictive distribution of the responses, obtain $\mathbf{t}_{B,j}$, $\mathbf{t}_{R,j}$ and $\mathbf{t}_{E,j}$ by back-transforming the responses to their original scale, and compute samples from the predictive distribution of the CQAs. This forms the basis for assessing the quality of the LC method. A Monte-Carlo estimate of the expected probability (i.e. assurance) of observing CQAs falling within acceptance criteria in future runs is obtained from the predictive distribution. For example, in chromatography it may be desirable to obtain a high predictive probability $P(S_{crit} > 0 \mid \text{data})$, or a high joint predictive probability $P(S_{crit} > 0, t_{tot} < 45 \mid \text{data})$.

8.3 Experimental section

8.3.1 Materials

Ketoprofen (99.7%), diclofenac (99.7%), naproxen (>98%), piroxicam (>99%), nimesulide (100.0%), sulindac (98%), suprofen (99.1%), sodium benzoate (99.9%), 4-aminophenol (98%), flurbiprofen (batch F8514-5G), phenoprofen (batch 029K-1043), tenoxicam (batch T0909-5G) and mefenamic acid (batch 36H0945) were purchased from Sigma-Aldrich (Antwerp, Belgium). Caffeine (100.1%), paracetamol (99.5%), ibuprofen (99.6%), indomethacin (99.2%), nipagin (100.1%), nipasol (101.9%), chlorzoxazon (99.1%), acetylsalicylic acid (batch 08G31-B28-232951), salicylic acid (batch 06K14 – B09 – 216351), and butylated hydroxytoluene (batch 05E31-B05) were purchased from Fagron N.V. (Waregem, Belgium). Dextropropoxyphen (batch 203100), phenylbutazone (batch 00951QA) and butylated hydroxyanisole (batch 511527) were purchased from Federa (Brussels, Belgium). Tiaprofenic acid was purchased from Erfa S.A (Brussels, Belgium). Niflumic acid (batch 0411545-2) was purchased from Cayman Chemical Company (Lansing, Michigan, USA). Lactose (batch 70756355176) was purchased from DMV Fronterra Excipients (Goch, Germany). Methanol (HPLC gradient grade), hydrochloric acid (37%), ammonium hydroxide (32%) and ammonium hydrogen carbonate (99%) were purchased from Merck (Darmstadt, Germany). Ammonium formate (99%) was provided by Alfa Aesar (Karlsruhe, Germany). Trifluoroacetic acid (batch 1001007) was purchased from Fisher Scientific Bioblock (Tournai, Belgium). Ultrapure water was obtained from a Milli-Q Plus 185 water purification system (Millipore, Billerica, MA, USA). For the preparation of validation standards, a matrix formulation of capsules containing 200 mg of paracetamol, 400 mg of ibuprofen, 40 mg of caffeine and 10 mg of lactose was kindly provided by Zenufa (Kinshasa, DRC)

The 27 materials were divided into 5 groups as presented in Table 8.1. These groups were based on the pharmaceutical form of the NSAIDs and were intended to expedite the determination of the method DS.

Group	Subgroup	Molecules
Group 1 (Compounds often presented in combination in tablet or capsule)	-	PAR, AA, IBU, DIC, CHL, DEX, NIM, KTO, MA, SAL, CAF
Group 2 (Compounds presented in combination in syrup and suspension)	-	PAR, IBU, NIM, MA, NIP, NIS, BEN, BHA, BHT
Group 3 (NSAIDs found alone in tablet or capsule)	-	IDO, TE, PI, FU, TA, NAP, SUF, PHE, PF, NA
Group 4 (Pharmaceutical combinations presented in tablet or capsule)	1	PAR, AA, CAF
	2	PAR, IBU
	3	PAR, DIC
	4	PAR, DIC, CHL
	5	**PAR, IBU, CAF**
	6	PAR, MA
	7	PAR, DEX
	8	PAR, DEX, CAF
Group 5 (Compounds presented in syrup and suspension)	1	PAR, NIP, NIS, BEN, BHA, BHT
	2	PAR, IBU, NIP, NIS, BEN, BHA, BHT
	3	IBU, NIP, NIS, BEN, BHA, BHT
	4	NIM, NIP, NIS, BEN, BHA, BHT
	5	MA, NIP, NIS, BEN, BHA, BHT

Table 8.1: Groups of compounds studied in this work. (Bold) submixture used for the validation experiments.

8.3.2 Standard sample preparation

Mixture preparation groups

1mg/mL stock solutions of each of the 27 studied materials were prepared in methanol. Mixture solutions were obtained by diluting stock solutions in methanol-water (50:50, v/v) in such a way as to obtain a working concentration of 50 μg/mL for HPLC analyses. Solutions injected into the UHPLC were 10 μg/ mL for each material. Aliquots of these solutions were filtered with 0.20 μm PTFE syringe filtration disks into vials for injection in the HPLC and UHPLC systems.

Solution used for calibration and validation

A stock solution of PAR, IBU, CAF and 4-aminophenol was prepared by dissolving 100 mg of each material in 100mL methanol. A stock solution of lactose was prepared by dissolving 100 mg of lactose in 100 mL of water (1 mg/mL). Heating and ultrasonic bath were necessary to ensure a complete dissolution.

For the calibration standards (CS), dilutions were performed in methanol-water (50:50, v/v) in order to obtain solutions at concentration levels of 200 μg/mL, 400 μg/mL and 600 μg/mL, except for 4-aminophenol where a dilution was made to obtain a concentration of 0.5 μg/mL (i.e. 0.1% of 500 μg/mL of paracetamol, being the reference concentration of 100%). For PAR, IBU and CAF, three concentration levels were sufficient to generate different regression models for the calibration, while for 4-aminophenol, a one-level calibration was made as advised in the European Pharmacopoeia monograph of paracetamol for the determination of impurities (European Pharmacopoeia, 2011c).

For validation standards (VS), independent stock solutions of PAR, IBU, CAF, 4-aminophenol were prepared in the same way as described for the CS. For the matrix, the same lactose solution was added into each working solution to obtain an amount of lactose of 4% relative to the amount of IBU. Subsequent dilutions in methanol-water (50:50, v/v) were carried out in order to obtain 5 solutions at different concentration levels (200 μg/mL, 300 μg/mL, 400 μg/mL, 500 μg/mL and 600 μg/mL) of PAR, IBU and CAF. For the 4-aminophenol, only one concentration level of 0.5 μg/mL was tested as described previously. The VS were independently prepared in the matrix, simulating as much as possible the formulation and its future routine analysis.

8.3.3 Instrumentation and chromatographic conditions

The optimization, validation and routine analysis were performed on a HPLC system comprised of a Waters 2695 separation module coupled to a Waters selector valve 7678 and a Waters 996 Photodiode array (PDA) detector (Waters, Eschborn, Germany). The analytical column was an XBridge C18 (250 mm x 4.6 mm i.d., particle size 5 μm), preceded by a guard column XBridge guard C18 (20 mm × 4.6 mm i.d., particle size 5 μm), both from Waters. The HPLC method was transferred to a UHPLC system Acquity ultra performance liquid chromatography (UPLC™) system from Waters, comprised of a binary solvent manager, an autosampler with a 10 μL loop, operating in the partial loop with needle overfill injection mode, and a PDA detector. The UHPLC system was equipped with an Acquity BEH C18 column (50 mm×2.1 mm i.d., particle size 1.7 μm) from Waters. XBridge and Acquity BEH columns are made with same stationary phase chemistry, providing an equivalent selectivity allowing for a geometrical transfer. The analytes were monitored photometrically at 220 nm while chromatographic data were recorded from 210 to 400 nm for all the studied conditions. For the HPLC system, the injection volume was 10 μL and the mobile phase flow rate was 1 mL/min. For the UHPLC system, the injection volume and the mobile phase flow rate were reduced geometrically to 2 μL and 613 μL/min, respectively. After each injection, the HPLC system was reconditioned for 30 min and the UHPLC system for 2 min.

The buffer solutions consisted of 20 mM ammonium formate, except for pH higher than 5 for which 20 mM ammonium hydrogen carbonate was used. The pH was adjusted with hydrochloric acid and ammonium hydroxide except for pH 1.85 where a solution of 0.1

8.3.4 Software

Empower 2.0 for Windows was used to control the HPLC and the Acquity UPLC™ systems, and to record and interpret the chromatograms.

An algorithm was set up to develop the Bayesian model and to compute the DS. The algorithm was written in R 2.13, which is available as freeware for most operating systems (R Development Core Team, 2010).

HPLC calculator v3.0 was used to carry out the necessary computations to identify the UHPLC conditions from the HPLC conditions using gradient geometric transfer methodology (Guillarme et al., 2007, 2008).

The accuracy profiles as well as the statistical calculations including the validation results and uncertainty estimates were obtained using e-noval® V3.0 software (Arlenda, Belgium).

8.4 Results and discussions

8.4.1 Design of experiments

For the optimization of HPLC conditions, an augmented central composite design was generated using the following factors: the pH of the aqueous part of mobile phase (pH), the gradient time needed to linearly modify the proportion of methanol from 15% to 95% of methanol (TG), and the column temperature (Temp). Experiment at the center of the experimental domain (i.e. at pH = 4.43, TG = 40 min and Temp = 27.5°C) was repeated trice, including the preparation of new buffer solutions. Factors values are presented in Table 8.2. The data augmentation consisted of the addition of 8 vertices of the cuboid domain and of 4×2 intermediate support points to obtain better estimates the pH effect (red) at the central levels of TG (green) and Temp (blue), leading to 32 experimental conditions.

Factors	Levels						
pH	1.85	2.42	3.14	4.42	5.71	6.42	7
Gradient time (TG, min)	-	20	24.5	40	55.5	60	-
Temperature (Temp, °C)	-	20	21.7	27.5	33.3	35	-

Table 8.2: Factors and corresponding levels of the augmented central composite design

An isocratic elution step with 95% methanol for 10 min was applied after the gradient to ensure the elution of all the tested molecules. For each experimental run, the three retention times of each peak (apex, begin and end) were recorded. When coelutions prevented the data treatment to be carried out properly, an independent component analysis (ICA) was used to accurately determine the retention times of coeluted peaks (Debrus et al., 2009). Furthermore, individual injections were made when strong coelutions of many compounds with too similar UV spectra did not allow peak identification and tracking.

8.4.2 Model

Responses from the experimental data were modeled using the following multivariate linear model,

$$\begin{aligned}
&\mathbf{Y} = \mathbf{XB} + \mathbf{E}, \\
&\mathbf{y}_j = \beta_{0,j} + \beta_{1,j}.pH + \beta_{2,j}.pH^2 + \beta_{3,j}.pH^3 + \beta_{4,j}.pH^4 + \beta_{5,j}.TG \\
&\quad + \beta_{6,j}.TG^2 + \beta_{7,j}.Temp + \beta_{8,j}.Temp^2 + \beta_{9,j}.pH.TG + \beta_{10,j}.pH.Temp \\
&\quad + \beta_{11,j}.Temp.TG + \beta_{12,j}.pH.temp.TG + \boldsymbol{\varepsilon}_j, \qquad (8.7)
\end{aligned}$$

where $\mathbf{y}_j$ is the jth column of $\mathbf{Y}$ ($j = 1, \ldots, 3m$). As commonly practiced, quantitative factors were centered and scaled to $[-1, 1]$ before being included in $\mathbf{X}$.

8.4.3 Prior information and model quality

As stated previously, when limited prior information on responses is available a priori, the parameters of the prior distributions should be as uninformative as possible. This is the case for the prior parameters $\mathbf{B}_0$ and $\boldsymbol{\Sigma}_0^{-1}$, which are defined as 0 everywhere. However, it is known that the correlations among the three responses describing one chromatographic peak are high. Indeed, as these responses are measures of the same compound, a strong dependence could be assumed. The matrix $\mathbf{A}^*$ might not account for this dependence accurately because data are limited. In order to induce a stronger correlation, the prior parameter $\boldsymbol{\Omega}$ is determined through a correlation matrix $\boldsymbol{\Omega}_{\text{cor}}$, defined as:

$$\boldsymbol{\Omega}_{\text{cor}} = \mathbf{I}_m \otimes \begin{pmatrix} 1 & 0.95 & 0.95 \\ 0.95 & 1 & 0.95 \\ 0.95 & 0.95 & 0 \end{pmatrix},$$

where $\mathbf{I}_m$ is an identity matrix of size m and $\otimes$ is the Kronecker product. Finally, the elements of $\boldsymbol{\Omega}_{\text{cor}} \equiv \{\omega_{\text{cor},ij}\}$ were rescaled towards the estimated matrix $\mathbf{A}^*$ as follows:

$$\boldsymbol{\Omega} \equiv \{\omega_{ij}\} = \sqrt{a^*_{ii}}.\omega_{\text{cor},ij}.\sqrt{a^*_{jj}}\frac{n_0}{n}, \qquad (8.8)$$

where a^*_{ii} is the i^{th} element of the diagonal of $\mathbf{A}^*$. Thus, in practice, $\boldsymbol{\Omega}$ is scaled to $\mathbf{A}^*/n$ before being multiplied by n_0 in order to be properly scaled (i.e. the scale of a so-called scale matrix with $\nu_0 = n_0 - (m + p) + 1$ d.f.). Since the experimental design provided sufficient data to estimate the correlation structure, n_0 has been fixed to 3 to limit the influence of the prior information.

Goodness of fit is evidenced by the relationship between observed and predicted responses, and the corresponding residuals in Figure 8.1. As seen on the graph (bottom), the majority of the residuals are distributed in the [-1, 1] min interval and the p-values of the Shapiro-Wilk normality test of residuals are all above 0.05 except for the responses of phenylbutazone. Given the satisfactory fit of the model, it can be used to predict the chromatographic behavior of each compound and to compute DS.

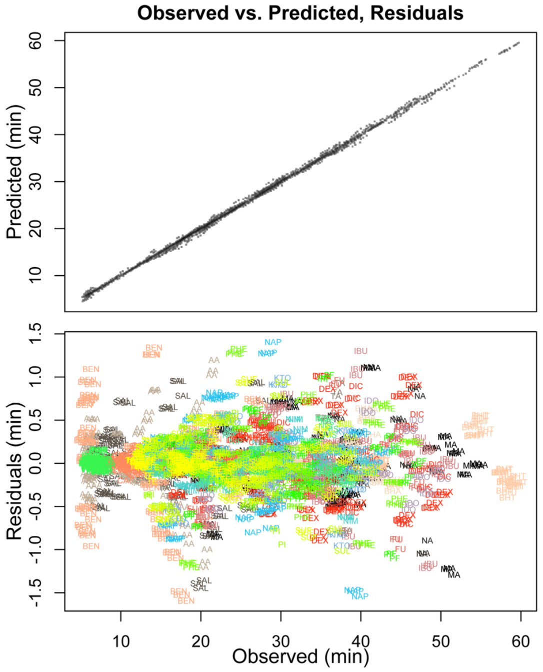


Figure 8.1: Modeling results. (Top) observed vs. predicted responses. (Bottom) residuals.

8.4.4 Design Space

As described in Section 8.2, the optimization process was repeated for each of the five groups of materials. For each group, the model was simplified in order to account only for the included materials, with the effect of reducing the size of the response matrix **Y** (Gupta and Nagar, 1999, Appendix D.6). The computed Monte-Carlo probability surfaces for $P(S_{crit} > 0)$ for groups 1, 2, and 3 are presented in Figure 8.2 (top, middle and bottom, respectively). In Figure 8.3 (top), the probability maps are the product of the probabilities $P(S_{crit} > 0, t_{tot} < 25)$, for each subgroup of group 4. In this case, the DS consisted in the intersection of the individual DSs. The shaded grey zones illustrate the part of the experimental domain where the quality level was not achieved. Thus a unique optimal condition and an unique DS was identified for all the subgroups (white area within red contour lines). Similar computation is depicted in Figure 8.3 (bottom, group 5) using the joint probability $P(S_{crit} > 0, t_{tot} < 40min)$ for optimization, and the DSs intersection of the five submixtures.

The DS shapes depicted in Figures 8.2 and 8.3 show broad regions with respect to variations in TG ([40, 60] min for group 1 and 2, [57, 60] min for group 3, [20, 25] min for group 4 and [34, 36] for group 5), in Temp ([20, 35]°C for group 1, 2 and 4, [20, 27]°C for group 3 and from [23, 30]°C for group 5) and pH (from [3.90, 4.30] for group 2 and [1.83, 4] for group 4). These broad regions are key results as they represent robustness w.r.t. changes of the operating conditions if quality level is high. Robustness across a wide range of temperature is particularly important in the case of using these methods in laboratories where the control of temperature is difficult. This can be the case for some developing countries. Conversely, three of the developed methods are far less robust with respect to pH ([2.80, 3.10] for group 1, [6.90, 7.00] for group 3 and [6.05, 6.20] for group 5). Fortunately, the relatively poor method robustness with respect to pH should not be problematic since this is easier to control, and care can be taken during the buffer preparation and subsequent pH measurements.

The results of optimal conditions and operating ranges are summarized in Table 8.3 for each of the 5 groups. For the three factors, the operating range is obtained as the interval in which the probability to achieve a satisfactory quality is higher than the specified quality level π. π is selected for each method in order to allow the identification of a risk-based DS. Except for groups 3, this quality level is generally high, indicating guarantees of quality for future use of the methods.

Optimal conditions	Optimal $P(S > 0)$	π	pH	TG (min)	Temp (°C)
Group 1	67.00%	0.63	3.05 (2.80-3.10)	49.30 (40-60)	34.5 (20-35)
Group 2	∼100.0%	0.95	4.05 (3.90-4.30)	53.14 (40-60)	23.0 (20-35)
Group 3	23.00%	0.2	7.00 (6.90-7.00)	60.00 (57-60)	21.7 (20-27)
Group 4	∼100.0%	0.95	3.04 (1.83-4)	20.00 (20-25)	27.5 (20-35)
Group 5	∼100.0%	0.95	6.14 (6.05-6.20)	35.00 (34-36)	29.4 (23-30)

Table 8.3: Optimal conditions and operating range within DS for the separation of the 5 groups of tested materials.

To support the ability of the DS to predict analytical conditions that permit chromatographic separation for the 5 groups, the different optimal conditions were tested twice (involving the preparation of new buffer solutions) to assess repeatability. Results are presented in Figures 8.4–8.6.

The grey area in the predicted chromatograms (top) is generated from 200 Monte-Carlo simulated chromatograms. This provides information about the uncertainty

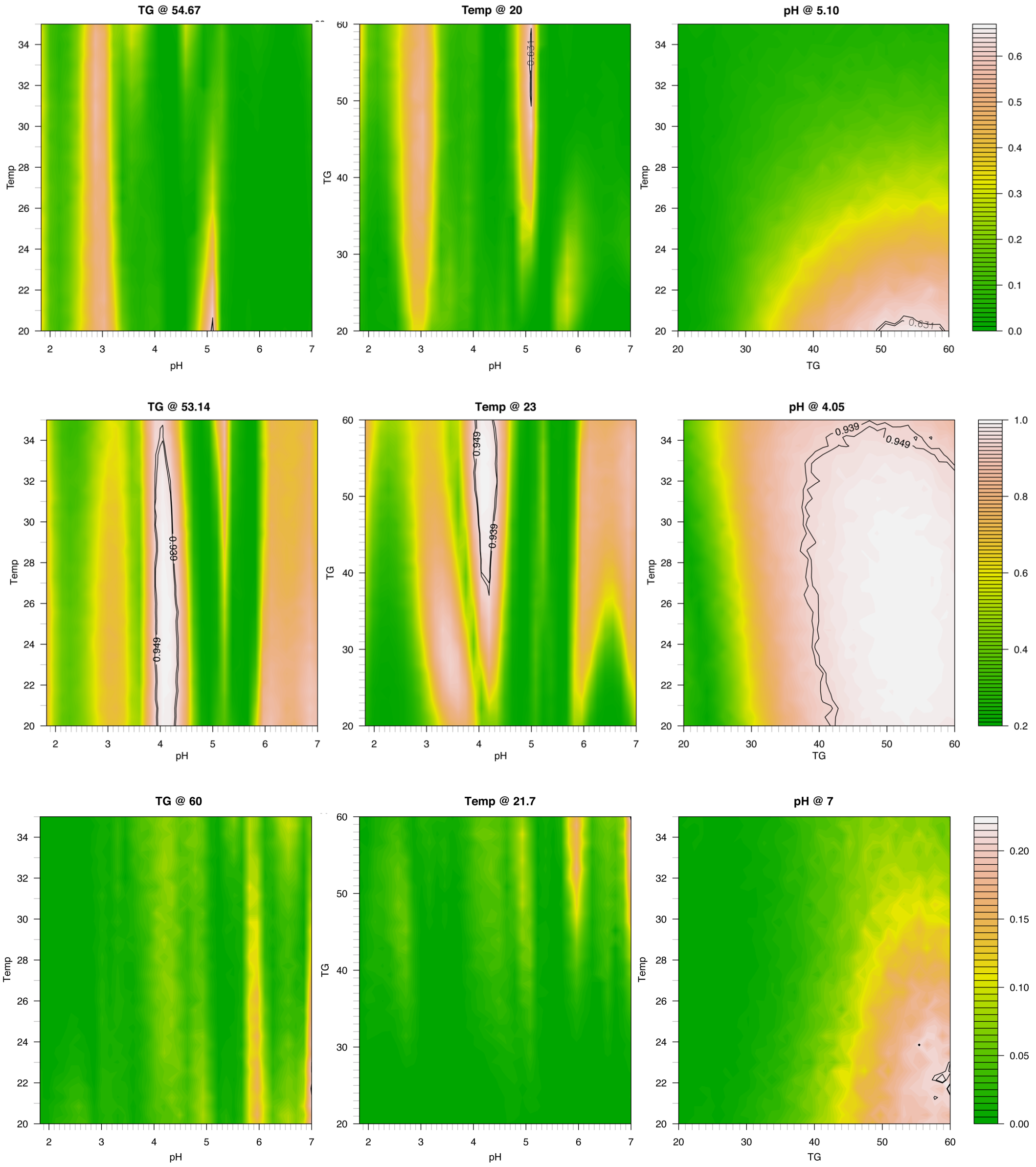


Figure 8.2: (Top) probability surfaces $P(S_{crit} > 0)$ for group 1. The DS is located in the white region with minimum quality level of $\pi = 0.63$. (Middle) probability surfaces $P(S_{crit} > 0)$ for group 2 with minimum quality level of π= 0.95. (Bottom) probability surfaces $P(S_{crit} > 0)$ for group 3 with minimum quality level of $\pi = 0.22$.

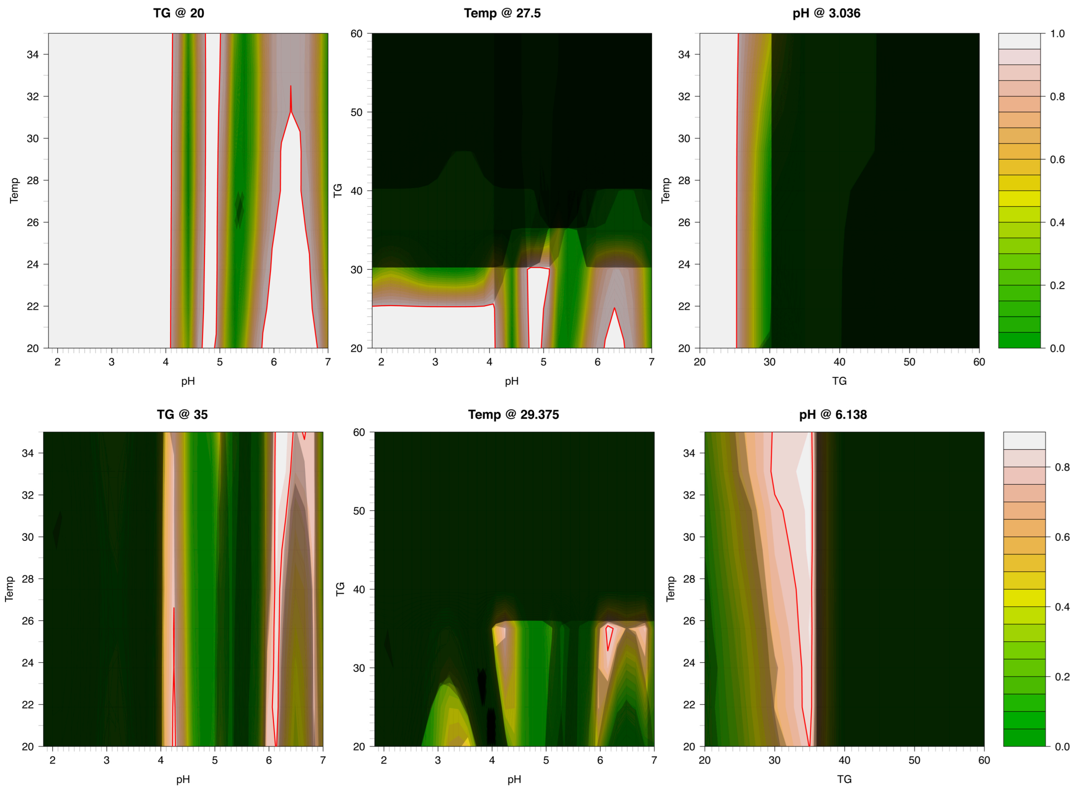


Figure 8.3: (Top) Design Space identification $P(S_{crit} > 0, t_{tot} < 25)$ for group 4. (Bottom) Design Space identification $P(S_{crit} > 0, t_{tot} < 45)$ for group 5. In both cases, the DSs consist in the intersection of the DSs ($\pi = 0.95\%$) of the proposed submixtures.

of the prediction. Given uncertainty of prediction, it is easy to visualize prediction quality while assessing robustness. Figure 8.4 illustrates the quality of the predicted optimal condition for the screening methods of groups 1 (left) and 3 (right). By the same way, Figure 8.5 illustrates the prediction quality of a unique method to identify and quantify eight combinations of NSAIDs in tablet. Figure 8.6 shows the results of a unique method to analyze five NSAIDs in suspension or syrup, in the presence of several adjuvants. As can be seen on the Figures 8.4–8.6, the chromatograms generated from materials tested under optimal conditions (bottom) are in close agreement with the corresponding predicted chromatograms (top) since the chromatographic peaks are accurately predicted, as shown by the limited uncertainty (grey). These results are corroborated in Table 8.4. The difference between the predicted and the observed critical separation was always negligible (less than one minute). Notably, this is consistent with the pattern of residuals (Figure 8.1).

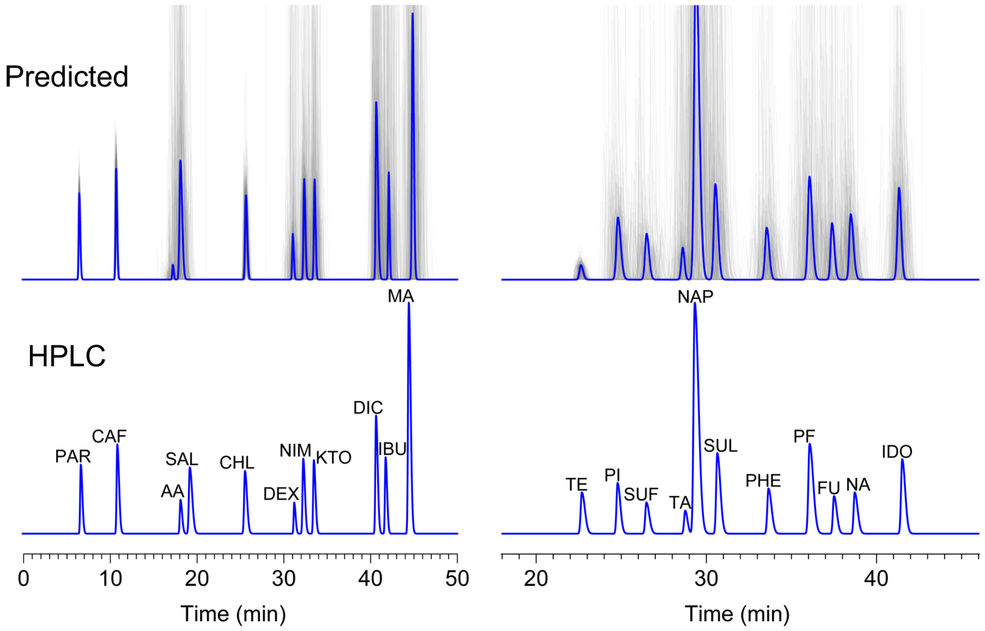


Figure 8.4: (Left) optimal condition for group 1. (Right) optimal condition for group 3. (Top) predicted chromatograms including simulations showing the uncertainty of predictions (grey). (Bottom) observed chromatograms.

Groups	Subgroup	Predicted Separation (min)	Observed Separation (min)	Error (min)
1	-	0.259	0.516	-0.257
2	-	0.894	0.317	0.577
3	-	0.115	0.016	0.099
4	1	1.946	2.034	-0.088
	2	15.193	15.827	-0.634
	3	14.636	14.475	0.161
	4	9.168	9.175	-0.007
	5	1.946	2.041	-0.095
	6	16.35	15.975	0.375
	7	1.946	2.033	-0.087
	8	10.186	10.483	-0.297
5	1	0.616	0.225	0.391
	2	0.616	0.267	0.349
	3	0.597	1	-0.403
	4	1.558	1.734	-0.176
	5	0.923	0.675	0.248

Table 8.4: Predicted and observed critical chromatographic separations.

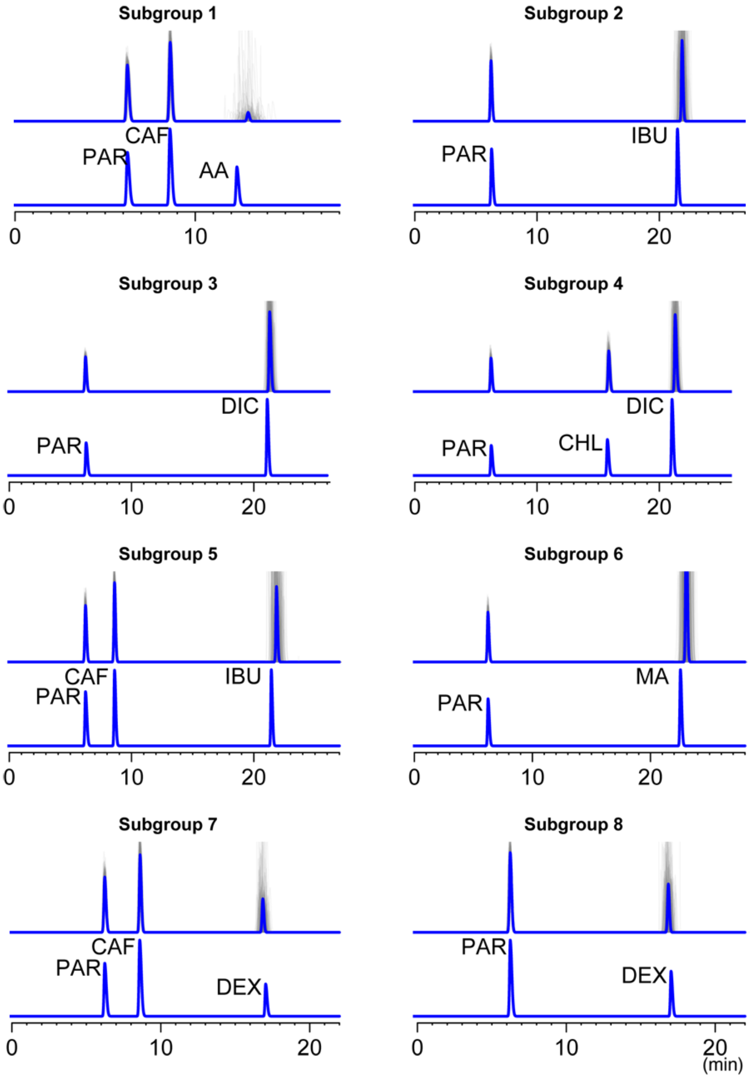


Figure 8.5: Optimal condition for the height submixtures of group 4. For each subfigure: (top) predicted chromatogram; (grey) simulations showing the uncertainty of prediction; (bottom) observed chromatogram.

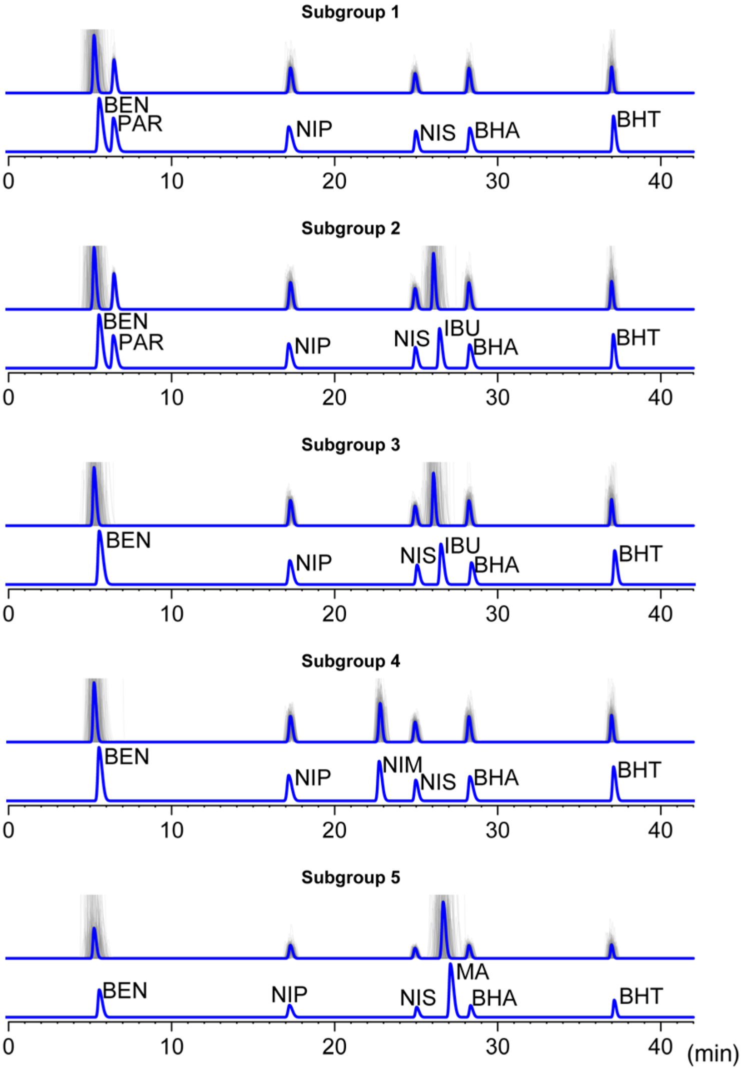


Figure 8.6: Optimal condition for the five submixtures of group 5. For each subfigure: (top) predicted chromatogram; (grey) simulations showing the uncertainty of prediction; (bottom) observed chromatogram.

4-aminophenol was not part of the experimental plan. It always elutes close to the column dead time. However, it was tested under conditions of optimal separation for all groups containing paracetamol, and was well separated from other compounds in every mixture and submixture. For instance, its retention time was 3.4 min at the optimal condition of group 4.

8.4.5 Transfer

One of the other objectives of the present study was the reduction of analysis time. This has important implications in the identification of substandard or counterfeit medicines since rapid analytical results provide fast decisions about suspected medicines. For this reason the screening method using HPLC (Figure 8.7 (a-b)) was transferred to UHPLC (Figure 8.7 (c)) using geometric transfer methodology (Guillarme et al., 2007, 2008). Analytical conditions were similar on HPLC and UHPLC, except those described in Section 3.3 and the gradient time, which was set to 3.52 min. As illustrated in Figure 8.7, the methods yielded very similar optimal separation. This is associated with a 15-fold reduction in analysis time from screening method, with 25 times less consumption of mobile phase. Because the UHPLC column geometry has been chosen to maximize the reduction of analysis time, a slight loss in peak efficiency was observed as predicted by the chromatographic theory. Detailed results are presented in Table 8.5. Relative retention times were used to compare elution performance of the two LC systems. This was achieved by dividing every retention time by the retention time of the last eluting material. Relative predicted and observed retention times were close. Similar results were obtained for the other groups of molecules, confirming the adequate geometric transfer of the methods from HPLC to UHPLC.

Moreover, these results clearly show that the variability induced by the transfer has not overly degraded the chromatographic separation. Thus, it also permitted the demonstration of the high robustness of the developed methods by means of the proposed DS optimization strategy.

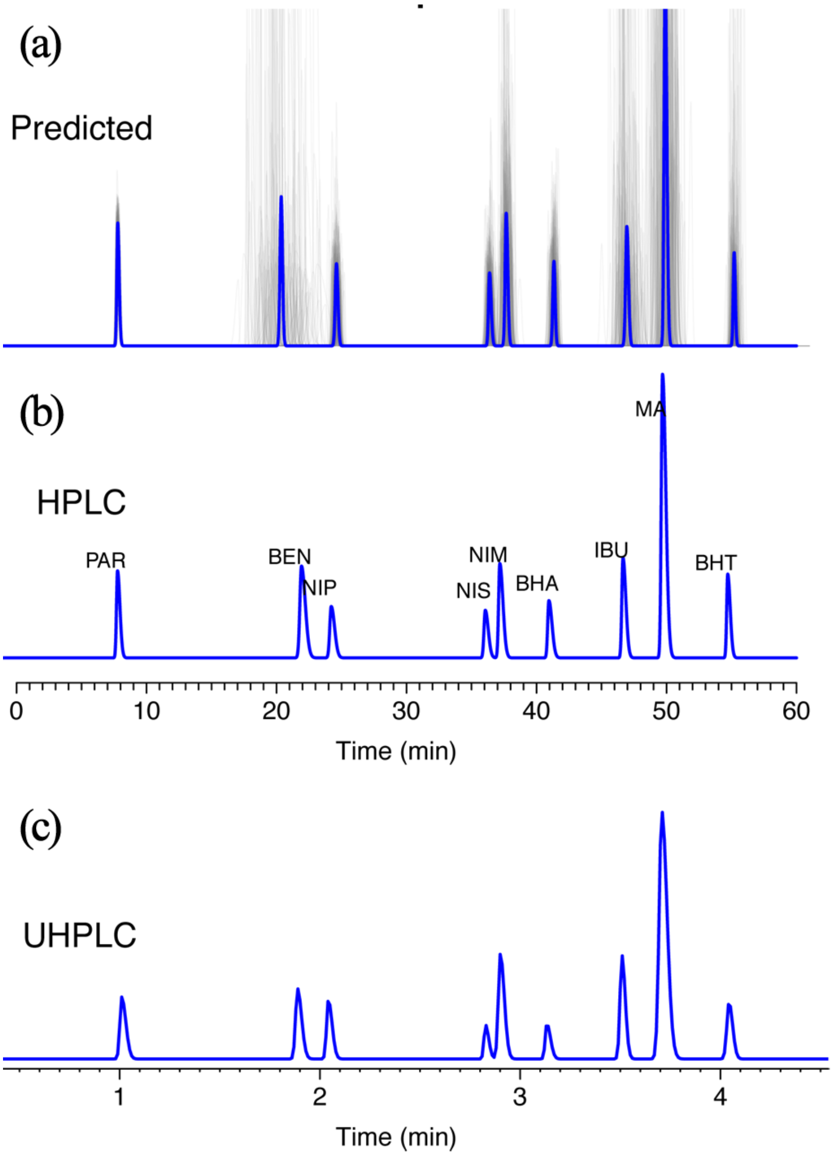


Figure 8.7: Optimal condition for group 2. (a) predicted chromatogram. (Grey) simulations showing the uncertainty of prediction. (b) observed chromatogram (HPLC). (c) observed chromatogram resulting of the transfer to UHPLC.

Name	Pred t_R	Lower Pred Int	Upper Pred Int	Obs t_R	**Pred t_R (rel)**	Lower Pred Int (rel)	Upper Pred Int (rel)	**Obs t_R (rel)**	Obs t_R UHPLC	**Obs t_R UHPLC (rel)**
PAR	7.79	7.67	7.91	8.01	**0.141**	0.135	0.139	**0.146**	0.54	**0.15**
BEN	20.36	17.91	23.01	22.28	**0.369**	0.318	0.41	**0.406**	1.42	**0.396**
NIP	24.61	24.26	24.98	24.54	**0.446**	0.434	0.447	**0.447**	1.57	**0.438**
NIS	36.39	36.06	36.71	36.36	**0.659**	0.649	0.661	**0.662**	2.35	**0.657**
NIM	37.67	37.15	38.21	37.45	**0.682**	0.668	0.687	**0.682**	2.43	**0.68**
BHA	41.34	40.98	41.7	41.23	**0.749**	0.737	0.751	**0.75**	2.66	**0.743**
IBU	46.95	45.74	48.13	46.92	**0.85**	0.824	0.867	**0.854**	3.04	**0.849**
MA	49.9	48.51	51.11	49.98	**0.904**	0.876	0.923	**0.91**	3.25	**0.908**
BHT	55.21	54.76	55.64	54.95	**1**	0.988	1.004	**1**	3.58	**1**

Table 8.5: Results of the transfer from HPLC to UHPLC. (Columns 2-5) predicted retention times (Pred t_r), predictive intervals (Pred Int) and observed (Obs) retention times for HPLC experiments. (Columns 6-9) relative (rel) predicted and observed retention times from HPLC experiments. (Columns 10-11) observed and relative observed retention times for UHPLC experiments. (Bold) comparison of relative retention times.

8.4.6 Method validation

After the optimization process, it is necessary to demonstrate that an analytical method provides accurate quantification results. This is carried out through a method validation. In this study, a quantitative method for capsules containing PAR, IBU and CAF (HPLC method for group 4, subgroup 5) was validated by applying the concept of total error represented by an accuracy profile (Hubert et al., 2007, 2008). As the capsules contain PAR, a quantitative method for 4-aminophenol impurity was developed concurrently. According to the European Medicines Agency, a formulation is declared compliant if its active molecules are within 5% of the nominal content (The European Medicines Agency, 1996). The accuracy profiles were used to assess the ability of the analytical methods to accurately quantify these three active ingredients, with acceptance limits that were set at 5% of the targeted concentration of the analytes (i.e. 5% relative total error is tolerated). The objective was thus to establish the dosing range in which the method is providing accurate results. To adequately estimate the total error of the quantitative methods under investigation and to mimic routine use of the method, three independent replicates were made for each concentration level. The process was also repeated independently during three days to estimate intermediate precision.

When a method is validated using the total error approach, and following the accuracy profile methodology, the validation parameters designated in ICH Q2 (precision, accuracy, linearity) are simultaneously combined to define a concentration range over which there is high probability to obtain future analytical results within the predefined acceptance limits (ICH Q2(R1), 2005; Feinberg et al., 2004). Accuracy profiles are presented in Figure 8.8. For PAR, IBU and CAF, a simple linear regression model was determined to be suitable for calibration. For 4-aminophenol, a one-level calibration was found appropriate. Individual validation parameters are presented in Table 8.6.

Trueness

Trueness is reported as the mean bias observed between the series of measurements and the targeted concentrations. Using the calibration curve of each material the concentrations of the VS were back-calculated and expressed in terms of absolute bias (μg/mL) and relative bias (%) (Hubert et al., 2004; ICH Q2(R1), 2005) The trueness of the developed methods was satisfactory, while the relative biases were close to 0 and were less than or equal to 1.03%.

Precision

Precision refers to the ability of the method to provide proximate results from multiple measurements of the same samples, under the same analytical conditions.

Precision is expressed as relative standard deviation (RSD%), and is reported for repeatability and intermediate precision at each targeted concentration. As shown in Table 8.6, precision was acceptable. The RSD% values never exceeded 1.21%.

Accuracy

Accuracy was assessed using the 95% β-expectation tolerance interval in order to analyze the closeness of agreement of individual measurements (ICH Q2(R1), 2005) and the assumed true value of the associated measurement. This combines the uncertainties associated with trueness and precision and is expressed as measured values and as a percentage of the targeted concentration (Table 8.6). The methods was found to provide accurate results, as the lower and upper tolerance bounds are included within the acceptance limits for all the targeted concentration levels

Validation criteria	Conc. (μg/mL)	PAR	IBU	CAF	4-aminophenol
Trueness: Absolute bias (μg/mL) (Relative bias (%))	50				<0.01 (0.21)
	200	1.72 (0.86)	2.06 (1.03)	-0.71 (-0.36)	
	300	1.30 (0.44)	-0.29 (-0.09)	-3.01 (-1.01)	
	400	-0.71 (-0.18)	-1.15 (-0.29)	-3.64 (-0.91)	
	500	3.10 (0.62)	1.46 (0.29)	-0.44 (-0.09)	
	600	1.27 (0.21)	-1.32 (-0.22)	-3.65 (-0.61)	
Precision: Repeatability (%) / Intermediate precision (%)	50				1.19 / 1.35
	200	0.20 / 0.36	0.18 / 0.67	0.36 / 0.59	
	300	0.16 / 0.33	0.35 / 0.35	0.75 / 0.75	
	400	0.17 / 0.40	0.38 / 0.55	1.08 / 1.08	
	500	0.16 / 0.55	0.32 / 0.63	1.21 /1.21	
	600	0.26 / 0.48	0.50 / 0.56	1.16 / 1.16	
Accuracy: 95% β-exp. tol. int. (μg/mL) (Rel. 95% β-exp. tol. int (%))	50				0.47–0.51 (-3.34–3.75)
	200	198.3–203.2 (-0.37–2.09)	195.0–207.1 (-2.02–4.09)	194.4–202.2 (-2.32–1.62)	
	300	296.2–303.4 (-0.77–1.64)	295.7–300.8 (-0.95–0.76)	290.0–301.0 (-2.84–0.83)	
	400	391.2–403.4 (-1.72–1.37)	390.1–403.6 (-1.98–1.40)	383.9–404.8 (-3.55–1.72)	
	500	488.4–512.8 (-1.83–3.08)	487.2–510.7 (-2.07–2.65)	482.4–11.8 (-3.04–2.87)	
	600	587.9–608.7 (-1.53–1.95)	587.1–604.3 (-1.66–1.22)	576.4–610.3 (-3.45–2.22)	
Linearity:	Slope	1.001	1.001	1.002	
	Intercept	0.979	-1.555	-4.071	
	R^2	0.998	0.997	0.99	

Table 8.6: Summary of the validation criteria for PAR, IBU, CAF and 4-aminophenol.

(Figure 8.8), thus assuring that each future result will fall within the acceptance range with a probability of at least 95% (Boulanger et al., 2003). Moreover, the relative 95% β-expectation tolerance intervals are generally within a range of [-3, +3]%.

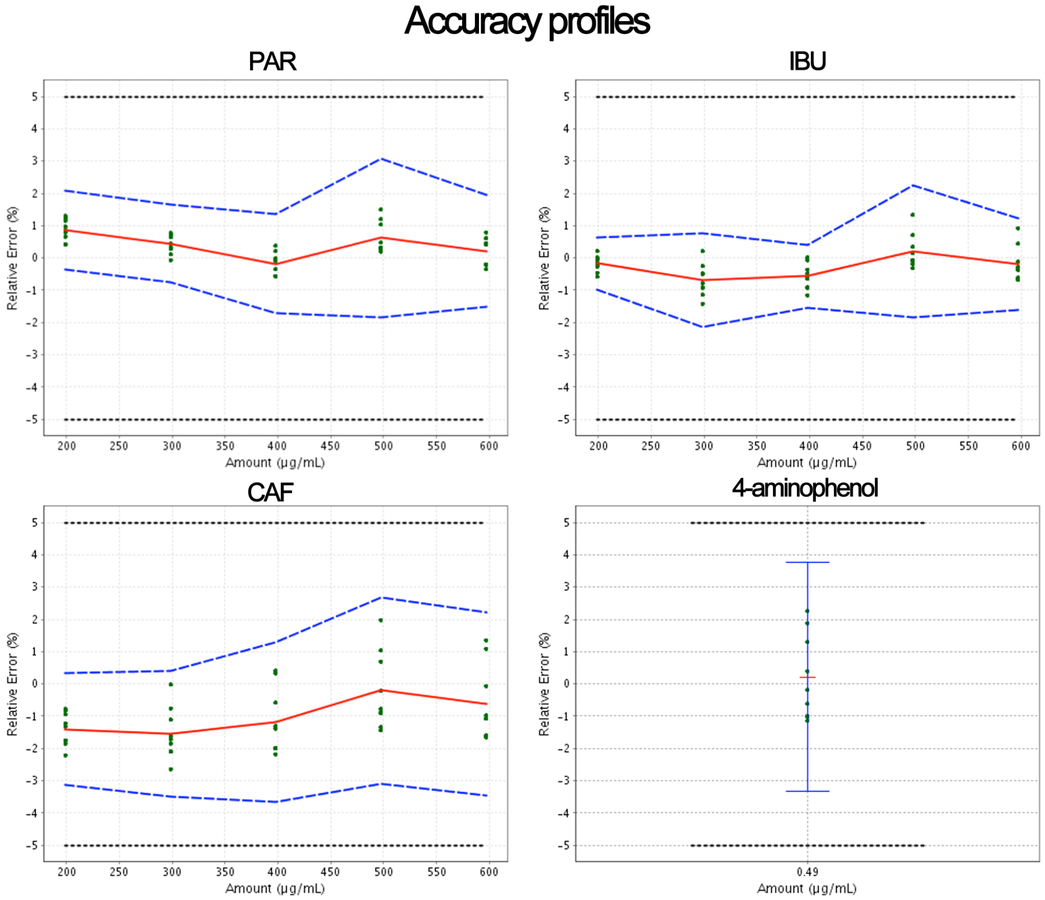


Figure 8.8: Accuracy profiles for quantitative methods validation (PAR=paracetamol, IBU=ibuprofen, CAF=Caffeine). (Red) bias (%). (Black) acceptance limit (±5%). (Blue) 95% β-expectation tolerance interval. (Green) individual measures. For the 4-aminophenol, a one-level calibration is used.

Linearity

The linearity of the results expresses the ability of the methods to produce results directly proportional to the concentrations. A simple regression model was adjusted to the observed vs. targeted concentration results to measure the linearity of the results. The coefficient of determination (R^2) obtained for the three compounds were all higher than 0.999 thus supporting the adequacy of the linear model adjusted. In addition, the linearity of the results was illustrated by the slopes of these regression models that are close to 1, ranging from 1.001 to 1.002 for PAR, IBU and CAF. This demonstrates the linearity of the results for the developed method.

Finally, for each concentration level of the VS, the 95% β-expectation tolerance intervals were all within ±5% of the targeted concentration of the analytes studied.

8.4.7 Application

The validated method was applied to the identification and assay of the three active ingredients (PAR, IBU, CAF). As an example, five different brands of pharmaceutical drugs coded A, B, C, D and E were tested. These were purchased in capsule form in the DRC and are mainly of Indian origin. Dilutions of the drugs were adapted so that the concentrations fell within the assay ranges.

Drug	PAR Content	CAF Content	IBU Content
A	325 mg 98.4 ± 0.41 %	30 mg **90.7 ± 1.49** %	200 mg 103.7 ± 0.74 %
B	325 mg 100.0 ± 0.35 %	40 mg **94.7 ± 0.63** %	200 mg 103.0 ± 0.58 %
C	200 mg **90.4 ± 0.22** %	40 mg **85.2 ± 0.79** %	400 mg **91.1 ± 0.73** %
D	325 mg **78.2 ± 0.39** %	40 mg **74.5 ± 0.44** %	400 mg **77.9 ± 0.15** %
E	325 mg **78.9 ± 0.28** %	40 mg **75.9 ± 0.31** %	400 mg **80.6 ± 0.35** %

Table 8.7: Assay results of five pharmaceuticals marketed in DRC. Results consist in the mean percentage of claimed nominal content and the standard deviation computed on 3 independent samples. Specifications are set to 95%–105% of the claimed nominal content (mg). (Bold) non-compliant results for the tested tablets.

The five drugs contained the three active ingredients but, as shown in Table 8.7, most of the products were in one way or another non-compliant. For example, product A had a declared nominal amount of caffeine equal to 30 mg, while only 90% of this amount was measured. For product B, the measured amount of caffeine was about 95% of the claimed nominal amount of 40 mg. From the low number of experiments carried out, this product is however considered non compliant. Finally, products C, D and E were not compliant since the contents of the 3 active ingredients were below the acceptance criteria.

The determination of the impurity of PAR (i.e. 4-aminophenol) was also established because of its potential toxicity. The European Pharmacopoeia places a limit of no more than 0.005% of 4-aminophenol in 100% paracetamol raw material (European Pharmacopoeia (2011c)). With pharmaceutical formulations, slightly higher concentrations might be tolerated due to the possible natural degradation of PAR during the manufacturing process. However, to our knowledge, no precise specification exists. The method for 4-aminophenol was determined to be valid to quantify as low as 0.1% of 4-aminophenol (0.5 μg/mL) in 100% paracetamol (500 μg/mL); however further experiments have shown that the limit of detection for 4-aminophenol is about 0.1 μg/mL. This would represent 0.02% of 4-aminophenol in 100% paracetamol, equivalent to four times the acceptance criterion of the European Pharmacopoeia.

The presence of 4-aminophenol was investigated in the five tested drugs and was not detected. Therefore, this seems to show that the low levels of active ingredients are linked to an insufficient dosing of these medicines rather than poor storage conditions.

8.5 Conclusions

The main objective of this work was to develop generic methods able to trace, screen and determine multiple non-steroidal anti-inflammatory molecules and common associated molecules, in order to help detect the potential counterfeiting of these drugs.

Using an experimental design based on three analytical factors (temperature, pH and gradient time), HPLC methods for five groups of NSAIDs and molecules of interest were developed in an innovative predictive risk-based framework. As an outcome of this original methodology, DSs were identified. This approach was then very helpful to optimize the separations of the tested molecules, allowing, for instance, their further quantification. The experiment showed that only the pH and gradient time had significant effects on peak separations within the explored experimental domain. The effect of temperature on quality was assessed and found to be limited. This may be due to the narrow range of temperatures investigated, but may suggest that using these methods in laboratories with no or insufficient temperature control is acceptable.

In order to concurrently support the robustness demonstrated using the computed DSs and to provide faster analytical methods with less solvent consumption, geometric transfer was used to adapt the optimized HPLC method to a UHPCL system. Faster methods are critical for laboratories involved in the control of drugs and counterfeits due to the increasing demands of analysis by legal authorities.

As an example of validation and application of their use in routine testing, a selected method was used for the determination of paracetamol, ibuprofen, caffeine and one impurity of paracetamol (4-aminophenol). The method was successfully validated using the total error approach and accuracy profile methodology. Finally, the method was effectively applied to analyze 5 brands pharmaceuticals marketed in the Democratic Republic of Congo. On the basis of the dramatic results obtained, it was confirmed that substandard and counterfeit medicines remain a crucial problem on public health in low-income countries.

Chapter 9

Design Space approach in the optimization of the spray-drying process

This chapter has been published in the European Journal of Pharmaceutics and Biopharmaceutics (Lebrun et al., 2012b).

9.1 Introduction

Nowadays, there is an increasing demand from regulatory authorities calling for the pharmaceutical industries to gain a comprehensive understanding of their manufacturing processes together with an accurate estimation of their robustness and reliability. Instead of providing solutions to meet these demands and requirements, authorities such as the International Conference on Harmonization (ICH) have published guidelines establishing the overall methodology to achieve these expectations. In the ICH Q8 (2009) guideline on pharmaceutical development, the emphasis is put on the "Quality by Design" (QbD) concept, stating that quality should not be tested into products, but should be built in (Yu, 2008). The Design Space (DS) concept is also introduced in this guideline, which is "the multidimensional combination and interaction of input variables (e.g., materials attributes) and process parameters that have been demonstrated to provide assurance of quality." Furthermore, ICH indicates that as long as the process and formulation parameters are kept within the defined DS, no regulatory post-approval change is needed. Thus, the DS of a process must also guarantee its reliability and robustness. In the US, the Food and Drug Administration (2011) has released the manual of policies and procedures for the effective application of several guidelines including ICH Q8. This is a strong indicator that industries must now be fully compliant with these QbD approaches.

Pulmonary delivery is an attractive administration route for the treatment or prophylaxis of airways diseases. It is also a reliable alternative to the subcutaneous and intravenous administration routes, especially for proteins and peptides. Indeed, the large surface area of the alveolar epithelial, the abundance of capillaries, and the low thickness of the air-blood barrier enable a drug delivery with systemic activity (Taylor, 2001).

Optimal drug deposition in the lungs requires several criteria to be fulfilled in terms of morphological aspects and ventilatory parameters. In addition, the particle's size and geometry aspects are also crucial (Groneberg et al., 2007). The optimum aerodynamic particle size for delivery is in the range of 1–5 μm (Hickey, 1996). Among the different particle processing techniques, spray-drying is known to produce particles that well fulfill the requirements for the pulmonary administration route. This processing technique offers many advantages, the first one being that the drying time of a droplet is only a fraction of a second, with a fast evaporation avoiding droplet overheating. The second one is that the final product has a large surface area and a uniform and controllable particle size (Maltesen et al., 2008). Furthermore, spray-drying is a continuous drying process consuming less energy than a freeze-drying process, for example (Masters, 2002). All the previous advantages make spray-drying (see Figure 9.1) an attractive manufacturing process for the pharmaceutical industry.

Figure 9.1: Mini spray-drier.

The aerodynamic properties of the powders obtained by spray-drying are determined by the particle size, density, and shape, which are influenced by spray-drying process parameters such as the inlet/outlet temperature, the air flow rate, and the feed flow rate (Cabral-Marques and Almeida, 2009). Facing the previous considerations, it is obvious that a holistic approach is needed to map the process parameters interactions. In this context, design of experiments is perfectly adapted to gather the data and translate how the combination of Critical Process Parameters (CPPs) affects the product Critical Quality Attributes (CQAs). It will eventually help defining the combinations of CPPs that will keep the product performance within the specifications with a quantified guarantee for the future use of the process: the Design Space.

In this context, the DS being identified is a region of reliable robustness, into the knowledge space. To provide guarantees of future quality, DS can be defined in a risk-based framework. The approach would finally be compliant with the QbD expectations. The results focus on the assessment of quality and on the guarantees (risks) that this quality could (could not) be achieved. Formally, Design Space is defined as

$$\mathrm{DS} = \{\tilde{\mathbf{x}} \in \chi \mid P(\mathbf{CQAs} \in \boldsymbol{\Lambda} \mid \tilde{\mathbf{x}}, \mathrm{data}) \geq \pi\} \tag{9.1}$$

In other words, the DS is a region of an experimental domain χ (often called knowledge space) where the posterior probability that the CQAs are within specifications $\boldsymbol{\Lambda}$, is higher than a specified quality level π.

Predictive posterior probability is central when dealing with concepts such as Design Space, as it allows quantifying the guarantees and risks that specifications will (or will not) be met in the future runs of the process, given the today's information. Specifications express the minimal satisfying quality that the experimenters want to obtain.

The optimization of multiple response surfaces usually involves the overlapping mean responses approach, which can be computed with commercially available software packages such as Modde, SAS-JMP, Minitab, Statistica. This approach is performed as follows: if, for example, a process response is influenced by three process parameters, then specific pieces of software can generally display the mean predicted process responses for any combination of the three process parameters within the defined parameter range. However, such approach does not take into account the model uncertainty: it will not provide any indication about how well and how often the process can meet the specifications with respect to the investigated CQAs, as stated by Peterson (2008). This represents a major drawback since ICH Q8 is clearly asking for a level of assurance guaranteeing the product specifications will be met.

In contrast to the overlapped mean response approach, a Bayesian predictive approach to define the DS takes into account the uncertainty of the process and of the analytical methods used to determine the CQAs and the uncertainties and correlations between the envisaged responses and the derived CQAs (Castagnoli et al., 2010). This approach integrates the uncertainty of parameters and the correlations between CQAs by propagating the multivariate error associated with the responses prediction. Consequently, the Bayesian predictive approach will significantly improve the model's prediction ability. We believe that this is the most efficient way for "demonstrating assurance of quality" as requested by the ICH Q8 definition of the DS. Within the pharmaceutical industry, the application of Bayesian statistics now begins to gain more interest and to be well accepted by authorities, especially in the field of clinical trials, as advocated by the Food and Drugs Administration (2010). Among other qualities, the Bayesian approaches allow the incorporation of prior information into models, if available, and may ease the solutions toward the predictive risk assessment.

In a previous work, Baldinger et al. (2001) investigated the influence of the processing parameters inlet temperature, spray flow rate, and feed rate on the following Critical Quality Attributes: yield, moisture content, particle size, and flowability by means of a design of experiment. However, high uncertainty was observed on most of the model parameters and no DS was identified. In this context, however, the question is to know to what extend the uncertainty could impact the prediction reliability. Within the framework of that previous work, the aim of the present study was to extend the approach of Baldinger et al. (2001) to define the spray-drying process DS according to ICH Q8. To our knowledge, this represents the first truly QbD-compliant approach to a spray-drying manufacturing process.

9.2 Materials and methods

9.2.1 Materials

D (−)-mannitol and D (+)-trehalose dihydrate were purchased from BDH Prolabo (Leuven, Belgium). Spray-drying was performed using 100 mL of an aqueous solution containing 10 g of a mixture of mannitol and trehalose in a mass ratio 90/10. Products were stored in closed vials at 5% relative humidity at room temperature.

9.2.2 Spray-drying

A Büchi Mini Spray-Dryer B-290 (Büchi Labortechnik AG, Flawil, Switzerland) with a 0.7-mm two-fluid nozzle was used. The solution was sprayed in a co-current flow with air as drying medium. Relevant spray-drying parameters were varied as stated in Section 9.3. The spray-dried particles were separated from the drying air by an improved cyclone (Maury et al. (2005)). Other key parameters were kept constant: The aspirator rates were set at 100% in all experiments, leading to a drying air flow of approximately 35 m^3/h. Spray-dried powders were collected, weighed, and stored in capped glass vials.

9.2.3 Particle size measurement by laser diffraction

The particle size distribution of the dry powders was measured using a laser diffractometer Mastersizer 2000 connected with a Scirrocco 2000 powder feeder (both: Malvern Instruments, Malvern, UK). For the measurement of the particles in air, a dispersion pressure of 1 bar was used.

9.2.4 Thermogravimetric analysis

The residual moisture content of the samples was investigated directly after spray-drying by using a TGA 7 (Perkin Elmer, Norwalk, CT). Powder samples between 3 and 12 mg were loaded onto a platinum sample pan and heated from 25 to 150 °C at a rate of 10 °C/min.

9.2.5 Bulk and tapped density

Bulk density and tapped density were obtained by following the European Pharmacopoeia (2011a) procedure 2.9.34. Due to the small amount of sample, a 10-mL tarred graduated cylinder was used. The bulk volume used for the calculation of the bulk density was directly read from the cylinder. Triplicates were made, and the mean value has been taken to define the bulk density.

$$\text{Bulk density (g/ml)} = \text{(weight of powder)} / \text{(bulk powder volume)}$$

The tapped density is obtained by mechanically tapping a graduated measuring cylinder containing the powder sample. The tapped density is read after 1250 taps corresponding to 5 min at a tapping height of 3 mm. The mean value of three replicates is recorded along with the observed variances among the experiments.

$$\text{Tapped density (g/ml)} = \text{(weight of powder)} / \text{(tapped powder volume)}$$

9.2.6 Softwares

An in-house computer program was developed to perform the statistical analysis. The coding was done with R 2.12, freely distributed at http://www.r-project.org and available for most operating systems (R Development Core Team, 2010). The package mvtnorm developed by Genz et al. (2011) has been used in order to sample from a multivariate Student's distribution.

9.3 Design of experiments

9.3.1 Critical Process Parameters

Three CPPs have been identified as having an impact on product quality. They are the inlet temperature, the spray flow rate, and the feed rate. Their range and unit are presented in Table 9.1.

Critical Process Parameters	Abbreviation	Low level	High level
Inlet temperature (°C)	IT	110	220
Spray flow rate (L/h)	SFR	439	1744
Feed rate (ml/min)	FR	2.5	7.5

Table 9.1: The Critical Process Parameters, their abbreviations and ranges.

For these three factors, a central composite face-centered design has been chosen, leading to $n = 17$ experiments comprising a center point in (independent) triplicates (Montgomery, 2009). Other key process parameters such as the drying air flow, the aspirator rate, the product variables (raw material characteristics) are kept constant.

9.3.2 Critical Quality Attributes

On every experiment, CQAs are recorded or derived from other attributes. They are defined in order to allow numerical assessment of the quality of the output. For every CQA, a specification ($\mathbf{\Lambda}$) is given that indicates a minimal satisfactory level of quality. The knowledge about these specifications can be a hard task, but it is the key toward a thorough and sound understanding of the process. Economical, efficiency, and safety reasons help in the definition of specifications. The five CQAs that will be taken into account for further analysis are reviewed hereafter.

Yield

The yield of the process is taken as a first quality attribute. It is computed as the percentage of obtained powder to the use of raw material. The experimenter will naturally look for a high yield. A specification limit can be derived from economical reason, but also from practical process management, and leads to a yield that must not be inferior to 80 %. A yield that is lower than this limit means that the use of the raw material is not optimal. In this case, the proportion of raw material that is not present in the obtained powder may be lost in the apparatus and more cleaning would be needed.

Moisture content

The residual moisture content of the obtained powder was analyzed using a thermogravimetric analysis. However, one can be easily convinced that a spray-dried powder has generally a very low level of moisture. Precision of the thermogravimetric apparatus is low in this case. For obvious quality reasons (conservation, non-aggregation of the powder), residual moisture of no more than 1% must be observed.

Aerodynamic particle size – inhalable fraction

The optimum aerodynamic particle size distribution for most inhalation aerosols has generally been recognized to be in the range of 1–5 µm (Hickey (1996)). Aerosols outside this range generally do not deposit in the lungs. The actual data consist of a mean of two analyses from the Mastersizer. The specification for this CQA is set to a minimum proportion of 60% of the particles should have a size between 1 and 5 μm.

Compressibility index and Hausner ratio

Two final CQAs are taken into account, the compressibility index (sometimes referred as Carr's index) and the Hausner ratio. They quantify the flowability of the obtained powder. They are both computed on the basis of the bulk and tapped density of the obtained powder:

Compressibility index (%) = 100 x (tapped density – bulk density) / tapped density

$$\text{Hausner ratio} = \text{tapped density} \;/\; \text{bulk density}$$

Since the compressibility index and the Hausner ratio are the combinations of random variables, we do not envisage their direct modeling. Instead, it is preferable to model the tapped density and the bulk density and to derive the two CQAs from these two responses. Table 9.2 illustrates some specifications about the compressibility index and the Hausner ratio (European Pharmacopoeia, 2011b):

Flowability	**Compressibility index** (%)	**Hausner ratio**
Excellent	0–10	1.00–1.11
Good	10–15	1.12–1.18
Fair	16–20	1.19–1.25
Passable	21–25	1.26–1.34
Poor	26–31	1.35–1.45
Very poor	32–37	1.46–1.59
Very, very poor	> 38	>1.60

Table 9.2: Specification for compressibility index and Hausner ratio.

Specifications for the compressibility index and the Hausner ratio have been chosen such as to have a good flowability, that is, the compressibility index must be lower 15% and the Hausner ratio lower than 1.18.

The five CQAs are summarized in Table 9.3 together with their specifications $\mathbf{\Lambda}$. An optimized process should provide outputs satisfying all the five specifications simultaneously, with the highest level of quantified guarantee possible.

CQA	Specification
Yield	≥ 80
Moisture	≤ 1
Fraction [1-5]μm	≥ 60
Compressibility index	≤ 15
Hausner ratio	≤ 1.18

Table 9.3: The CQAs and their specifications.

9.4 Results and discussion

9.4.1 Model

In this section, the statistical model and the related results are detailed.

Responses

From the analysis of the CQAs, five model responses are envisaged to allow analyzing the quality of the output (obtained powder):

- $Yield$
- $Moisture$
- $Fraction$ (Aerodynamic particle size - inhalable fraction)
- $Bulk$ (density)
- $Tapped$ (density)

The yield and the inhalable fraction are percentage values, and should have a range constrained to a domain [0-100]%. A logit transformation is applied to ensure that this property will be valid during the predictions:

$$LYield = \log\left(\frac{Yield}{100 - Yield}\right),$$

$$LFraction = \log\left(\frac{Fraction}{100 - Fraction}\right).$$

For the 3 other variables, log transformations are applied to ensure positivity.

$$LMoisture = \log(Moisture),$$

$$LBulk = \log(Bulk),$$

$$LTapped = \log(Tapped).$$

Notice also that a good practice is to constrain the value of bulk and tapped density so that the tapped density is always higher than the bulk density. Otherwise, the computation of the compressibility index and the Hausner ratio could lead to a negative value or value smaller than 1, respectively. Constraints can be applied during a Monte-Carlo simulation step, discarding the samples that do not fulfill them.

Multivariate multiple linear regression

To account for the correlations that will be observed between the responses, a multivariate multiple linear regression (MMLR) is adopted. Other statistical models are possible but MMLR has the advantage of simplicity for the identification of its predictive distribution. This model is fitted for every response jointly

$$\boldsymbol{Y} = (LYield, LFraction, LMoisture, LBulk, LTapped).$$

Let the following model be applied on the $m = 5$ responses ($j = 1, ..., m$),

$$\begin{aligned} \mathbf{y}_j = {} & \beta_{0,j} + \mathrm{IT}.\beta_{1,j} + \mathrm{IT}^2.\beta_{2,j} + \mathrm{FR}.\beta_{3,j} + \mathrm{FR}^2.\beta_{4,j} + \mathrm{SFR}.\beta_{5,j} + \\ & \mathrm{IT.SFR}.\beta_{6,j} + \mathrm{IT.FR\ .SFR}.\beta_{7,j} + \boldsymbol{\varepsilon}_j, \\ \mathbf{Y} = {} & \mathbf{XB} + \mathbf{E}, \end{aligned} \tag{9.2}$$

with the lines of $\mathbf{E}$, $\boldsymbol{\varepsilon}_i$, assumed to be i.i.d. as a multivariate Normal distribution, $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}), i = 1, \ldots, n$, with n the number of experiments. $\mathbf{X}$ is then the $(n \times p)$ centered and reduced design matrix and $\mathbf{B}$ is the $(p \times m)$ matrix containing the p effects for each of the m responses. The modeled effects have been chosen so that the model has the best properties for every response, jointly. $\boldsymbol{\Sigma}$ is the covariance matrix of the residuals. In order to account for the variability of the parameters $\mathbf{B}$ and $\boldsymbol{\Sigma}$, a predictive density of new predicted responses can be obtained in the Bayesian framework, considering the non-informative prior distribution $p(\mathbf{B}, \boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{-(m+1)/2}$(see Box and Tiao, 1973; Peterson, 2004). In this context, the predictive posterior density of a new predicted set of responses $(\tilde{\mathbf{y}} \mid \mathbf{X} = \tilde{\mathbf{x}}, \text{data})$ at a new operating condition $\tilde{\mathbf{x}} \in \chi$, is identified as a multivariate Student's distribution, defined as follows:

$$(\tilde{\mathbf{y}} \mid \mathbf{X} = \tilde{\mathbf{x}}, \text{data}) \sim T_m \left(\tilde{\mathbf{x}}\hat{\mathbf{B}}, \mathbf{A}. \left(1 + \tilde{\mathbf{x}}(\mathbf{X}'\mathbf{X})^{-1}\tilde{\mathbf{x}}' \right), \nu \right), \tag{9.3}$$

where $\hat{\mathbf{B}}$ is the least squares estimate of $\mathbf{B}$, $\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$, and $\mathbf{A} = (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})'(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})$ is a scale matrix and $\nu = n - (m + p) - 1$ is the degrees of freedom.

Effects analysis

The analysis of the parameters effects is done using the marginal posterior density of the parameters $\mathbf{B}$, as in Box and Tiao (1973). This density is centered on the ordinary least squares estimates $\hat{\mathbf{B}}$ and provides the credible intervals of the parameters, as shown on Figure 9.2. It illustrates the high uncertainty observed on most of the regression parameters. The parameters for which the marginal 95% credible intervals contain 0 are said to be nonsignificant (red) whereas the others differ significantly from 0 (green).

Briefly looking at Figure 9.2, one may observe that the increase in feed rate and spray flow rate have a positive impact on the moisture content of the product. However, as the moisture remains limited whatever the condition, this may not have

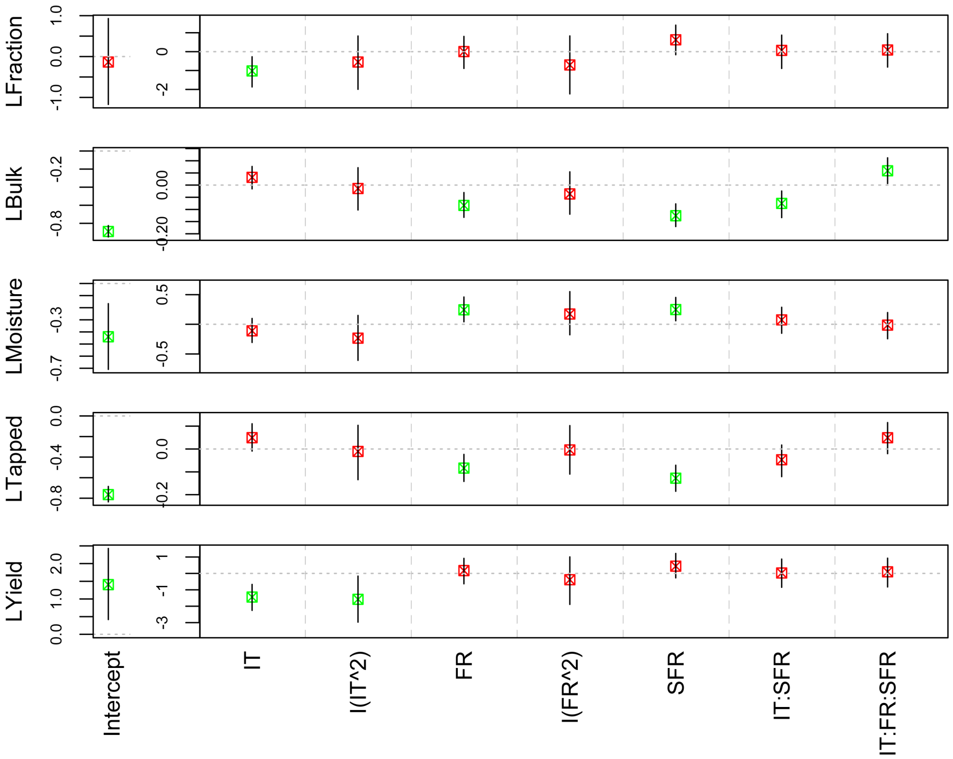


Figure 9.2: Parameter estimates and 95% credible intervals. (Green) Parameters significantly different from 0. (Red) Parameters non-significantly different from 0.

a strong impact on the overall quality. The same CPPs show a negative impact on both bulk and tapped densities. Next, the inlet temperature shows a negative impact on the yield. It might be due to the fact that more agglomerates are created at higher temperature, leading to a lower yield (Baldinger et al., 2001). The same observation is done about the inhalable fraction. These types of results are the first keys toward a better understanding of the process. Finally, Figure 9.2 clearly emphasis the necessity in considering the uncertainties of the parameters, when envisaging a risk-based approach. If checking their mean values and their distributions provide good insights into the process, it does not provide any direct valuable information on the quality of the output of the process, nor on the guarantee that this quality would be achieved.

Predicted vs. observed and residual analysis

Other model checks allow a better understanding of the model's capabilities. Firstly, it is advised to visualize the model suitability by plotting the mean (untransformed) predicted responses against the observed ones. The residuals have been graphically also checked, as illustrated in Figure 9.3.

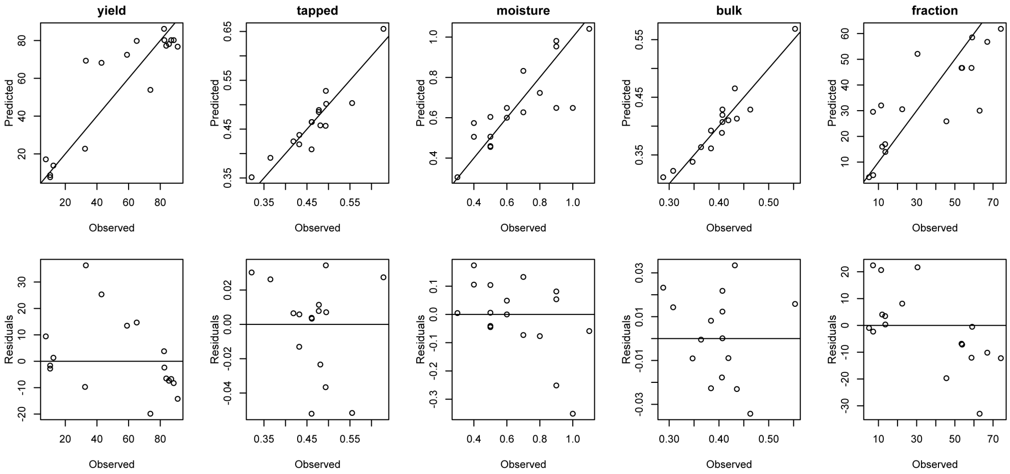


Figure 9.3: Predicted vs. observed plots (top) and residuals plots (bottom) for every response.

Figure 9.3 illustrates the low quality of the multivariate model. This has clear explanations for responses like the moisture content (low precision of the measuring device for very low concentrations) or the tapped/bulk density (that are experimentally carried on small quantities of powder). Again, taking into account these residual uncertainties allows giving risk-based results even in the presence of poor model fit. Finally, Q–Q plots are drawn to see whether the model residuals do not depart from Normality assumption. This is shown in Figure 9.4. The hypothesis of Normality of the residuals seems acceptable.

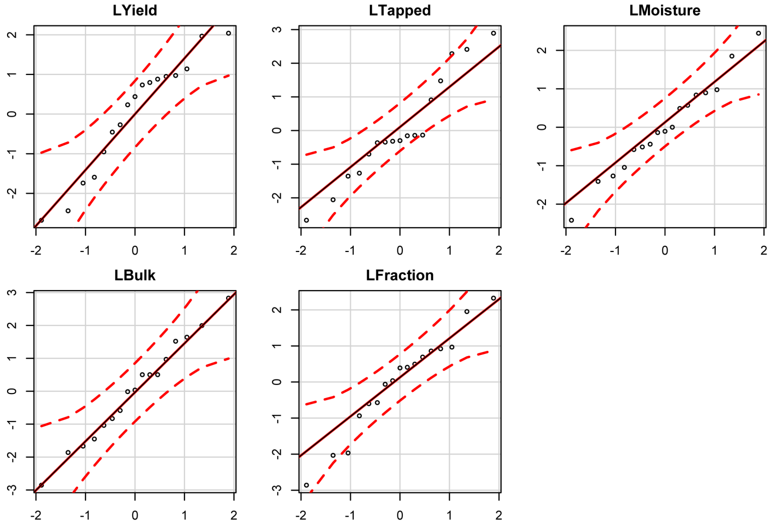


Figure 9.4: Marginal Normal Q-Q plots of the residuals.

Response correlations

From the model, one can estimate the correlations that exist between the responses by rescaling the matrix **A** into an estimated correlation matrix, by dividing its elements with the appropriate row and column standard deviations (Draper and Smith, 1998). The correlations are given in Table 9.4.

	$LYield$	$LTapped$	$LMoisture$	$LBulk$	$LFraction$
$LYield$	1				
$Ltapped$	-0.05	1			
$Lmoisture$	0.16	0.04	1		
$LBulk$	0.30	0.90	0.31	1	
$LFraction$	0.83	0.32	0.28	0.60	1

Table 9.4: Estimated residuals correlations between the responses.

A strong correlation (0.90) is observed between the bulk and the tapped densities meaning that these two values generally have a similar behavior. These similarities might be a result of the tapping process carried out on small quantities of powder. The observation is as follows: if the bulk density is low (high), this will give a proportionally low (high) tapped density. From a statistical point of view, not taking this correlation into account when deriving the compressibility index and the Hausner ratio would be harmful. Next, there is an agreement between the yield and the inhalable fraction. In the present process, the higher the yield, the higher the inhalable fraction will be. In addition, a correlation of 0.60 between the inhalable fraction and the bulk density can be observed. From a physical point of view, this can be explained by the fact that the lower the particle size is, the higher the bulk density will be. Moreover, the tapped density being correlated with the bulk density, the tapped density is also slightly correlated with the inhalable fraction. Nevertheless, this correlation is lower than the one between the inhalable fraction and bulk density because the inhalable fraction determination is based on the bulk powder and not on the tapped one

Design Space computation

Basically, the way to compute the DS is to use the joint predictive distribution of the CQAs, derived from Equation (9.3) in every point of the experimental domain. When CQAs are transformations and/or combinations of the responses, Monte-Carlo simulations are envisaged to propagate the predictive uncertainty and interactions/correlations of the responses to the CQAs. Randomly looking at a point $\mathbf{x}_{sub}$ of the knowledge space, for instance, Inlet Temperature = 180 °C, Feed Rate = 2.5 ml/min, Spray Flow Rate = 1744 L/h, one can analyze the sampled CQAs derived from the responses drawn from the multivariate Student's t distributions, as shown on Figure 9.5.

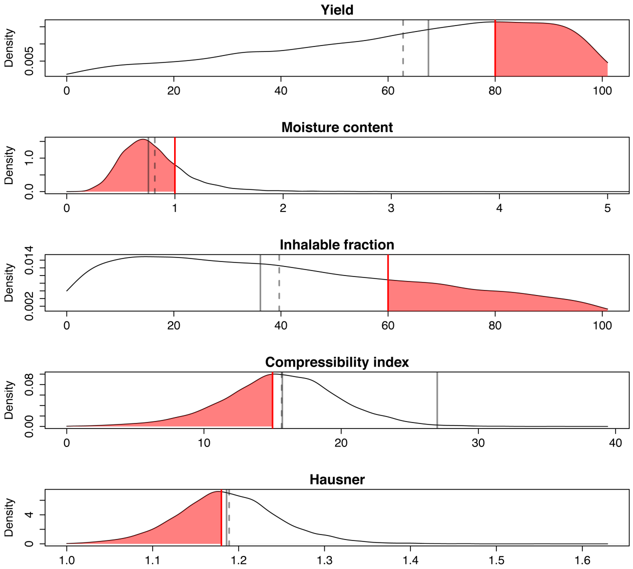


Figure 9.5: Marginal predictive kernel density estimations of the Critical Quality Attributes on a point $\mathbf{x}_{sub}$ of the experimental domain. The red lines are the specification limits and the red regions are the estimated probabilities of achieving the specifications. The black lines are the medians (plain) and the means (dashed) of the distributions.

The red lines indicate the specifications for each CQA, while the red regions of the predictive densities illustrate the proportion of simulated points (i.e., the estimated predictive probability) that are within specifications. For instance, at operating condition $\mathbf{x}_{sub}$ and for the CQA $Yield$, the proportion is about 30%, meaning that there is a probability of about 0.3 to have, in the future, a yield higher than 80% ($P(Yield \geq 80\%) = 0.3$).[1] This represents a high risk of $1 - 0.3 = 0.7$ of being outside the specification. Taking into account the yield alone, this condition may thus be considered outside of the DS with a satisfying quality level. Regarding the $Moisture$ content, the probability to have this CQA within the specification is 0.8, so there is a risk of 0.2 of being outside specifications. Envisaging this CQA alone, $\mathbf{x}_{sub}$ could belong to the "design space" with a specified minimal quality level π of, say, 0.6. However, going multivariate deteriorate the results. The joint probability to accept all the specifications is lower than 0.001. The $\mathbf{x}_{sub}$ input condition is then clearly not within any DS.

Next, the same computations are done, on every point $\tilde{\mathbf{x}}$ of the experimental domain χ. To do so, a grid search is applied. For each operating condition, the

[1] i.e., $P(Yield \geq 80\% \mid data) = 0.3$. The conditioning on data is removed to simplify the notations.

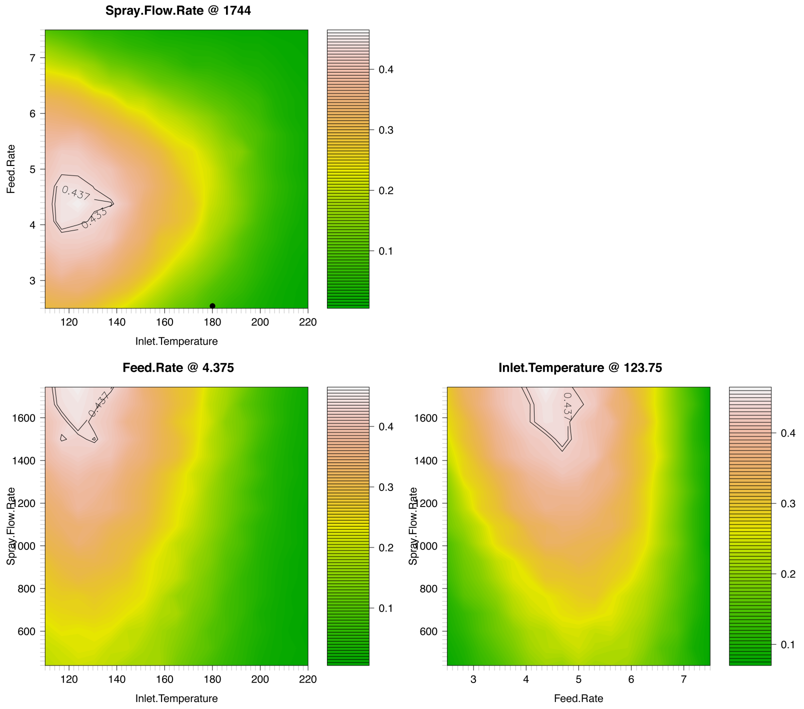


Figure 9.6: Posterior probability map that the CQAs satisfy the five specifications presented in Table 9.3. Inner black lines define the DS for the minimal quality level π=0.437 The black point is the sub-optimal condition $\mathbf{x}_{sub}$ presented in Figure 9.5.

estimated expected probability that the 5 CQAs are jointly within specifications is recorded. A probability map is then drawn, as shown in Figure 9.6. A clear similarity with response surface is seen, but the interpretation of such maps is quite different. In each operating condition represented, the map gives the joint expected probability (i.e., the guarantees) to observe the process within specifications, on a future run. On the cuboid knowledge space, only three slices that include the optimal condition are represented.

The black-contoured region is the DS, i.e., the operating conditions where the joint probability to achieve all specifications on the CQAs is the highest over the experimental domain. A cuboid can be extracted, and the limit values for its vertices are given in Table 9.5.

Critical Process Parameters	DS range
Feed Rate (ml/min)	[4.2–4.8]
Spray Flow Rate (L/h)	[1614–1744]
Inlet temperature (°C)	[118–125]

Table 9.5: Design Space of the process.

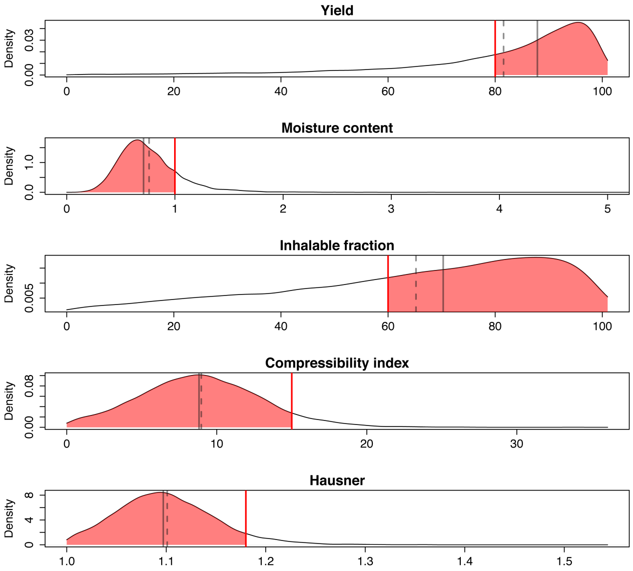


Figure 9.7: Marginal predictive kernel density estimations of the Critical Quality Attributes on at the optimum. See Figure 9.5 for colors and legend.

In this application, a DS is found, for a specified minimal quality level of π=0.437, chosen as 95% of the quality level of the optimal solution. However, one can be suspicious about the quality of the results. Indeed, the optimal expected probability is only about 0.45. Thus, within the DS, there is a risk of 0.55 not to be within all the specifications concurrently. A good insight for a better comprehension of what happens is to have a look at the marginal predictive distributions at the optimal point, which is: Inlet Temperature = 123.75 °C, Feed Rate = 4.69 ml/min and Spray Flow Rate = 1744 L/h, as shown in Figure 9.7.

Regarding Figure 9.7, it is obvious that, marginally, the estimated expected probabilities for every CQA are quite satisfactory. The acceptance probability is higher than 0.7 for the yield ($P(Yield \geq 80\%) = 0.71$), 0.78 for the moisture content ($\leq 1\%$), 0.62 for the inhalable fraction ($\geq 60\%$), and 0.85 for both the compressibility index (≤ 15) and the Hausner ratio (≤ 1.18). Then, except for the inhalable fraction, the model provides us a satisfying confidence toward the future performance of the process.

From a probabilistic perspective, the addition of univariate specifications in a multivariate analysis logically leads to a decrease of the joint predictive probability of acceptance (Ekins et al., 2002). At optimal condition, the following decreasing probabilities illustrate this situation.

$P(Yield \geq 80\%) = 0.71,$
$P(Yield \geq 80\% \text{ and } Moisture \leq 1\%) = 0.56,$
$P(Yield \geq 80\% \text{ and } Moisture \leq 1\% \text{ and } Fraction \geq 60\%) = 0.48,$
$P(Yield \geq 80\% \text{ and } Moisture \leq 1\% \text{ and } Fraction \geq 60\% \text{ and } Hausner \leq 1.18) = 0.45$

The definition of multivariate specifications may be seen as a remedy to this. In this context, desirability functions can be envisaged to aggregate the values of every individual predicted CQA into a single value, namely the desirability index, representing the desirability of the solution (Harrington, 1965; Derringer and Suich, 1980; see also Chapter 5). Steuer (2000) has shown how Monte-Carlo simulations can be used to propagate the predictive uncertainty and the correlations of the CQAs (or the responses) to the desirability index. This index allows for certain trade-offs between the CQAs. A slightly bad result for one CQA could be compensated by a very satisfactory result for another.

In this 5-CQAs study with univariate specifications, it may not be surprising to observe the optimal joint estimated expected probability of acceptance being about 0.45. Of course, finding a DS with a higher minimal quality level and even stronger specifications would be an even more desirable situation.

Some estimates for each CQA are provided in Table 9.6, computed from the distribution presented in Figure 9.7. The mean values (dashed lines) or the medians (plain line) are the values expressing the central tendency one can expect to observe. Additionally, the 75% and 95% Bayesian predictive intervals are also provided as valuable information about the uncertainty of prediction.

CQA	Lower 95%	Lower 75%	**Median**	**Mean**	Upper 75%	Upper 95%
Yield (%)	42	75	**88**	**81**	94	100
Moisture content (%)	0.26	0.57	**0.71**	**0.76**	0.89	1.31
Inhalable fraction (%)	17	49	**70**	**65**	85	100
Compressibility index	0.4	6.2	**8.8**	**9**	11.5	16.1
Hausner ratio	1	1.07	**1.09**	**1.1**	1.13	1.19

Table 9.6: Statistics on the CQAs at the optimal input condition.

For instance, the 75% predictive interval around the CQA Inhalable $fraction$ is very large ([49–85]%). Then, the model is poorly informative regarding this CQA. A similar conclusion was reached when looking at the marginal acceptance probability for this CQA at the optimum, which was only 0.62.

9.4.2 Validation

The optimal solution has been carried out three times independently on the same apparatus to observe how the process performs within its 0.45 quality level DS. Table 9.7 summarizes the experimental results. They reinforce the statistics observed during the optimization process.

Batches	Yield (%)	Moisture (%)	Fraction (%)	Compressibility index	Hausner ratio
1	88	< 0.2	63	11.6	1.13
2	89	< 0.2	62	12	1.14
3	88	< 0.2	59	11.5	1.13
Mean	**88.7**	**< 0.2**	**61.18**	**11.76**	**1.13**
Standard Deviation	0.61	NA	1.82	0.22	0.01

Table 9.7: Results of the validation experiments.

As expected, the process performs according to the predictions. Most batches are within specifications. The inhalable fraction is seen as acceptable (higher than 60%) except in the third batch (red). However, on average (bold), the process corroborates the results of the joint expected probability, which was about 0.45. Obviously, a longer-term study would be necessary to plainly assess the routine performance of the process.

Finally, Table 9.7 provides the indication of the variability observed in the three independent batches. This variability is low compared to the predictive uncertainty that was observed (see Figure 9.7 and Table 9.6). This indicates that the residuals predictive uncertainty is not only due to the noise of the process. The poor model fit is also a concern. A possible explanation is that more complex interactions and higher order or non-linear effects are present. Unfortunately, the central composite face-centered design used in the experimental part is too light to detect such effects. Indeed, the design allows only the estimation of the main and quadratic effects and the principal interactions. This underlines the need to define more informative designs when little is known about the process, even if the price that must be paid is the carrying out of more experiments.

9.5 Conclusions

When setting up a QbD-compliant ICH Q8 Design Space for a process such as spray-drying, the use of the mean response surface optimization methodology is not recommended due to the inevitable uncertainties and interactions that are encountered. Accordingly, the data gathered through an experimental plan have been analyzed using a risk-based Bayesian predictive approach allowing the uncertainties and interactions to be integrated into a multivariate statistical model.

These variabilities result in a minimal quality level that has been kept relatively low in order to be able to define a Design Space, i.e., the guarantee of jointly observing the Critical Quality Attributes within their acceptance limits is low. Even with this situation, these guarantees are quantified along with the risks of not observing such quality, jointly or marginally. The specifications have been designed such as to provide a minimal satisfying quality for whole process. In this way, the quality of the resulting product is built in by the design and controlled setup of the spray-drying equipment.

Validation of the optimal condition within the Design Space has been carried out, and these experiments provided a product compliant with the predicted quality. To better assess how the guarantees of quality prediction perform, one would consider analyzing longer-term process data.

In addition, the validation experiments carried out independently provided supplementary information concerning the statistical model. Indeed, the good repeatability of the process seems to indicate that the causes of the poor model fit were not solely due to the noise present in the data. Instead, more complex interactions or non-linearity of the responses can be present. In cases where nothing or little is known about a specific process, defining a more informative though labor-intensive design of experiments should be envisaged.

Finally, the definition of a low guarantee Design Space could be seen as the very first step toward a Quality by Design methodology. The results presented are of great interest for the spray-drying manufacturers and experimenters in order to improve quality. For instance, the causes of variation could be identified, such as poorly controlled factors. Furthermore, the effect of the key process parameters that have been kept constant during this study could be analyzed in a more detailed way through a new experimental plan.

Chapter 10

Automated validation of a quantitative chromatographic method

10.1 Introduction

In the pharmaceutical field, the need to develop analytical methods able to quantify accurately compounds of interest is high. For instance, the determination of the conformity of samples before clinical trials or during the phases of a drug development is crucial. Drugs quality issues were also discussed in Chapters 7 and 8, where the screening and complete analysis of drugs were detailed, in order to detect and potentially fight against the poor quality medicines.

An analytical method such as the high performance liquid chromatography (HPLC) is a flexible tools to obtain these types of results. In order to become *quantitative*, the method needs first to be calibrated. This calibration links the results of interest (the concentration or amount of a compound) to the responses of the method, that are extracted from the chromatograms recorded with a diode array detector (DAD-chromatograms). Typically, the area under the curve (AUC) of the observed peaks can be recorded if the corresponding compounds are separated. For a given wavelength of observation, an AUC is proportional to the real concentration of a compound, following the Beer-Lambert law (1852). For HPLC system with an ultra-violet detector, the relation between the responses and the concentrations is generally well explained by a simple linear regression.

To prove that a method is able to provide **accurate** results for its **future** runs, the method must go through a process called *validation*. Following ICH Q2(R1)

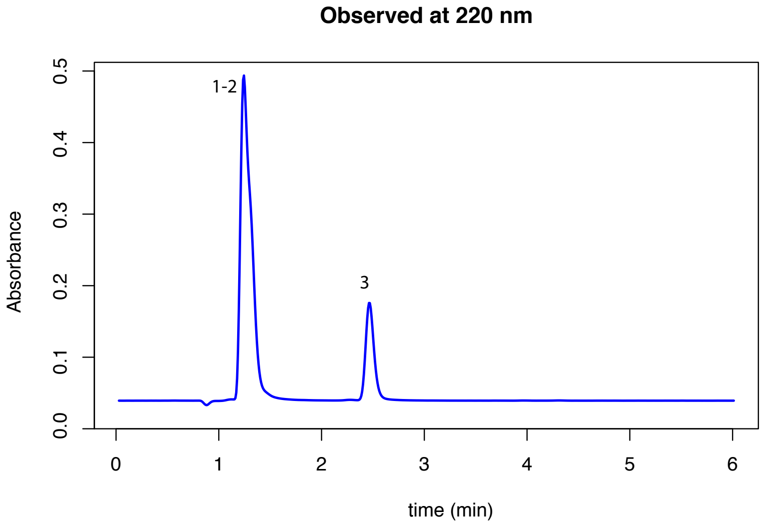


Figure 10.1: Example of chromatogram. For this study, suboptimal chromatograms were used, with the first peak being the sums of two peaks of sulfinpyrazone (1) and granisetron (2).

(2005), “the objective of validation of an analytical procedure is to demonstrate that it is suitable for its intended purpose”. See also ISO/CEI 17025 (2005). However, validation is a non productive process from an industrial point of view, with many experiments that must be carried out. One bottleneck remains the analysis of the chromatograms, comprising the identification, the integration of the peaks, and the reporting of the analytical responses comprising the AUC. In the best cases, softwares can help in this data processing, but it is far to be automated. Besides, even when it is made with great care, data manipulation problems or integration errors can also occur, as it will be shown.

In this chapter, a totally automated validation of quantitative method is presented based on real data obtained on a mixture of sulfinpyrazone, granisetron and phenytoine. In the proposed example, a suboptimal chromatographic condition was purposely chosen, resulting in non fully separated peaks. This type of results departs from the best case scenario, and generally, the method validation can not be envisaged with these poor separation properties. An example of data is provided in Figure 10.1. The two first peaks (sulfinpyrazone and granisetron) are coeluted and observed at 1.3 min. A third peak (phenytoine, at 2.5 min) is well separated from the others. During the experiments, the position of the peaks will not vary as the HPLC method is always the same. However, their heights, widths and resulting AUC will change because different concentrations of the compounds are injected into the HPLC system.

This chapter is an application of the independent component analysis (ICA) and of the clustering methodologies presented in Chapter 6. It also makes use of the predictive models presented in Chapter 4, illustrating an example of automated calibration and validation of quantitative methods.

Structure of the chapter

Section 10.2 describes the data. In Section 10.3, the independent component analysis (ICA) is used to process the original chromatograms and identify sources that contain the peak information. Next, Section 10.4 illustrates the method validation using firstly the distribution of the inverse prediction from the calibration, and secondly, the accuracy profile strategy. Both validation methodologies are intended to assess the quality of the quantitative methods for each of the three involved compounds.

10.2 Materials and methods

10.2.1 Instrumentation and chromatographic conditions

All experiments were carried out on a conventional HPLC system Alliance from Waters (Eschborn, Germany) equipped with an ultra-violet diode-array detector (UV-DAD), similar to the one presented in Chapter 7. The analytical column was an Xbridge C18 (100 mm x 4.6 mm i.d., particle size 5 μm) from Waters and was kept at 20°C during the experiments. An isocratic elution was envisaged. The mobile phase used for the analysis was a 38:62 (v/v) mixture of acetonitrile and ammonium hydrogen carbonate buffer (10 mM) with aqueous part adjusted to pH 7 with hydrochloric acid and ammonium hydroxide.

10.2.2 Samples preparation

Two sets of data were generated for the calibration of the quantitative method for the three compounds, and for its validation. They are referred as the calibration set and the validation set. Each set consisted in 45 DAD-chromatograms recorded during 3 days of 15 runs, as explained in the two next subsections.

Calibration set

From stock solutions of phenytoine, granisetron and sulfinpyrazone, dilutions were carried out in methanol-water (20:80, v/v) in order to obtain working solutions at three concentration levels of 30 μg/mL, 150 μg/mL and 300 μg/mL for each compound. For the calibration set, the three compounds were then present at the same concentration (either 30, 150 or 300 μg/mL) in the working solutions. Aliquots of the working solutions were filtered with 0.20 μm PTFE syringe filtration disks into vials for injection in the HPLC system. The injection volume was 10 μL and the flow of the mobile phase was 1 ml/min.

Three independent series of data were generated, including new preparation of the solutions on 3 different days. Each day, five replicates were carried out, for each of the three concentration levels. This provided $3 \times 5 \times 3 = 45$ chromatograms.

Validation set

From stock solutions of phenytoine, granisetron and sulfinpyrazone, dilutions were carried out in methanol-water (20:80, v/v) to obtain working solutions at three concentration levels of 30 μg/mL, 150 μg/mL and 300 μg/mL for each compound. However, for the validation set, each sample has been created with different concentrations of the compounds. In order to design the way these concentrations are mixed in the samples, an incomplete latin-square design was used (Cox and Cochran, 1957). The design is presented on Table 10.1 and was also replicated during three days, leading to 45 chromatograms. Filtration and injections of the solutions were made as mentioned for the calibration set.

	Experiments														
Compounds	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Sulfinpyrazone	2	1	3	1	3	1	3	2	1	2	1	3	2	3	2
Granisetron	1	2	2	3	3	1	1	2	3	3	1	1	2	2	3
Phenytoine	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3

Table 10.1: Incomplete latin square design for the concentration levels in the validation samples. 1: 30 μg/mL; 2: 150 μg/mL; 3: 300 μg/mL. Validation data resulted of three application of this design.

The latin square was used to create arduous and worst-case data. In this way, the coeluted compounds are often observed as a tall and a small peak. In some experiments where the concentrations have a ratio 1/10, the smaller peak is nearly non detectable when looking at the original chromatograms, as the peaks 1-2 in Figure 10.1.

10.2.3 Softwares

Empower V2.0 for Windows was used to control the HPLC system and to record the chromatograms. The R environment was used for all the analysis, including ICA, for which the package `fastICA` was used (R Development Core Team, 2010; Marchini et al., 2010).

10.3 Independent component analysis

The aim of the ICA algorithm is to estimate the independent components that are mixed in the observed signal, and to estimate the way they are mixed. Here, the relevant components should correspond to peaks of interest. Eventually, after of a successful ICA, it is possible to compute the AUC of each numerically separated peak by integrating the peaks observed in the relevant components.

In Chapter 6, the ICA was briefly described. On the basis of a DAD-chromatogram $\mathbf{X}$ consisting of m observed wavelengths (or signals) at t_{tot} recorded time points, the idea is to estimate a $(m \times f)$ mixing matrix $\hat{\mathbf{A}}$ and the f sources in $\hat{\mathbf{S}}\,(f \times t_{\text{tot}})$ as independent as possible, such that:

$$\mathbf{AS} = \mathbf{X}.$$

After the estimation, f components $\hat{\mathbf{X}}_g$ and a reconstructed DAD-chromatogram $\hat{\mathbf{X}}$ are obtained by computing:

$$\hat{\mathbf{A}}_{,g}\hat{\mathbf{S}}_{g,} = \hat{\mathbf{X}}_g \text{ and } \sum_{g=1}^{f} \hat{\mathbf{X}}_g = \hat{\mathbf{A}}\hat{\mathbf{S}} = \hat{\mathbf{X}}, \tag{10.1}$$

where $\hat{\mathbf{A}}_{,g}$ is the g^{th} column of $\hat{\mathbf{A}}$ and $\hat{\mathbf{S}}_{g,}$ is the g^{th} line of $\hat{\mathbf{S}}$. Notice that ICA is applied on each DAD-chromatogram separately.

In Section 6.4.2, a methodology to find the optimal number of sources f^* was proposed, based on the fact that a mixture containing the same compounds is injected several times. This methodology was then applied on the 45+45 DAD-chromatograms from the calibration and validation sets. For both sets of this example, a value of $f^* = 4$ for every DAD-chromatograms was identified to numerically separate the three peaks. For simple cases, in general, ICA performs well when the number of sources allows the separation of the components of interest (here, 3), plus one or more sources for the noise and other artifacts. $f^* = 4$ is then not a surprising result. On more complex cases, it was also shown in Chapter 6 how f^* can be adapted for each chromatogram.

10.3.1 ICA results

On the (45+45)*3=270 peaks, the percentage of correct peak picking is 100%. Furthermore, no irrelevant artifact was found. Figure 10.2 illustrates the components (observed at 220 nm) of two numerically separated DAD-chromatograms, with the apexes of the peaks marked in red. It is observed that the components from the validation set (B) seem more affected by disturbances. Indeed, some small artifacts generally remain in the relevant components (lines 2-3). For instance, an artifact is found in the second source (line 2, insert at 2.4 min). It was found that it might correspond to an impurity of granisetron, that has appeared in the samples of the validation set. This small impurity elutes at the same time than the compound phenytoine, which is referred as a *specificity* problem (see for instance ICH Q2(R1), 2005) that can cause some difficulties if care is not taken.

The very similar UV-signature of granisetron and of this impurity is the reason why they appear in the same ICA component. Finally, in the third source, a peak at 4 min. is also clearly visible, and it might be an impurity of phenytoine.

Thus, ICA successfully achieved the automated peak detection process. It also improved the identification of some impurities that were not detected manually.

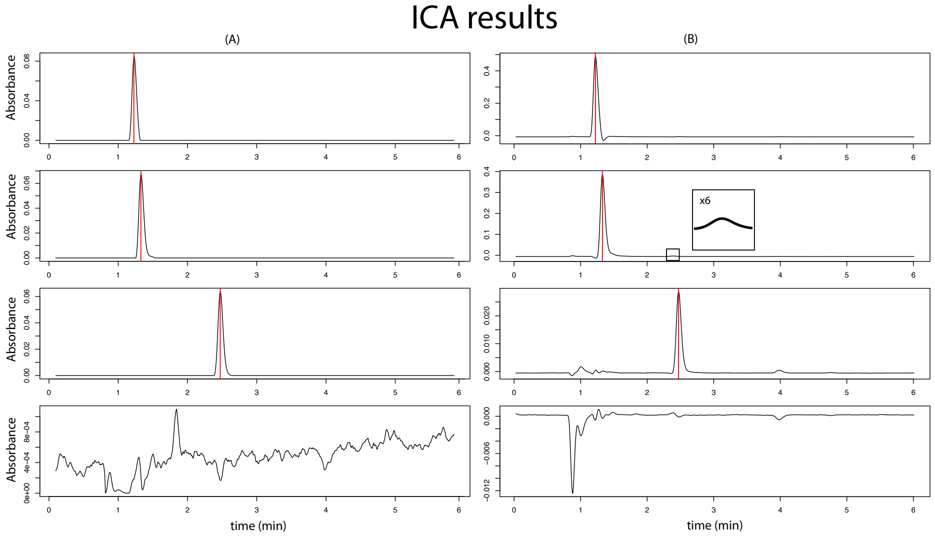


Figure 10.2: Numerical separation with ICA observed at 220 nm. (A) components from the calibration set. (B) components from the validation set. (Red) position of the apexes.

10.3.2 Peak tracking

Continuing the application of the methodology of Chapter 6, some dissimilarities computed on the sources and the components are derived to track the peaks among the DAD-chromatograms. In the present example, the process was simplified as all the similar peaks had similar retention times (times of the apexes). Assuming that s_{tot} relevant components $\hat{\mathbf{X}}_s(s = 1, ..., s_{\text{tot}})$ were identified in all the chromatograms, and using these components as in Equations 6.5–6.10 (page 101), let $t_R^{(s)}$ be defined as the time of the apex of the peak contained in the component $\hat{\mathbf{X}}_s$. An appropriate distance is then

$$d_t(s_1, s_2) = |t_R^{(s_1)} - t_R^{(s_2)}|, (\forall s_1, s_2 = 1, ..., s_{\text{tot}}) \tag{10.2}$$

with $|a|$ being the absolute value of a. The penalty distance d_{art} was used as a second metric to avoid matching together the components coming from the same DAD-chromatogram (see page 103). The Ward's algorithm was then applied on the dissimilarity matrix $\mathbf{D} \equiv \{d_{s_1,s_2}\} = d_{art}(s_1, s_2) + d_t(s_1, s_2)$, for the calibration set and the validation set separately.

The dendrograms for both the calibration and validation data are presented in Figure 10.3. The number of clusters was set to 3, the number of compounds. Every peak was matched accurately by the clustering algorithm with 100% of hit.

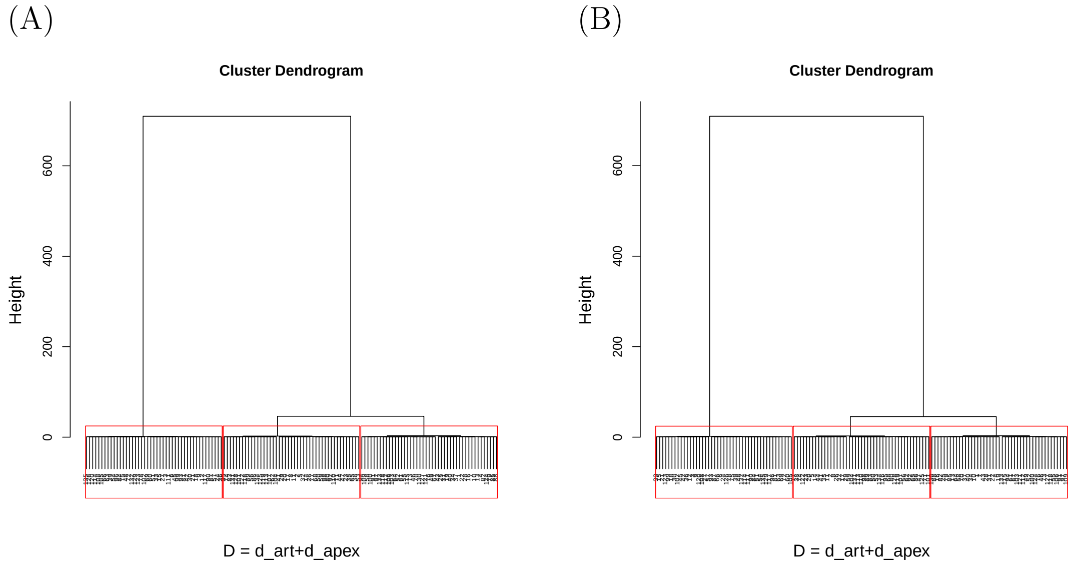


Figure 10.3: Dendrograms of the matching of peaks. (A) calibration set. (B) validation set. Red rectangles are the clusters for the 3 compounds.

10.3.3 Integration

As previously explained, the analytical results of the methods are the AUC of each peak in each chromatogram. To compute AUC, each of the (45+45)*3 numerically separated peaks was integrated individually. To make simple, the integration was done at one selected wavelength (here, 220 nm), i.e., one column of $\hat{\mathbf{X}}_s$. The wavelength was manually chosen so that all the compounds under investigation have an acceptable absorbance and a high signal-to-noise ratio.

For each component, the time of the apex $t_R^{(s)}$ was recorded, as shown in Figure 10.4. Next, locally around $t_R^{(s)}$, the beginning $t_B^{(s)}$ and the end $t_E^{(s)}$ of the peak were defined as the time where the absolute value of the first derivative of the chromatogram (red line) was close to 0. Simply summing the values of absorbance of the chromatogram between $t_B^{(s)}$ and $t_E^{(s)}$ would allow a first approximation of the AUC. However, an improvement was added because the baseline was not always perfectly horizontal. The integral of the linear baseline estimated from $t_B^{(s)}$ to $t_E^{(s)}$ (blue line) was simply removed from the previous estimation of AUC.

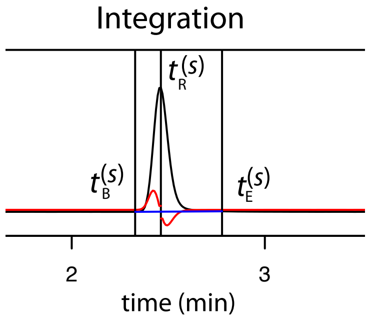


Figure 10.4: Example of peak integration. (Black) component observed at 220 nm. (Red) first derivative. (Blue) baseline estimation.

For both the calibration and the validation sets, the concentrations are known from the dilutions carried out when preparing the samples. At this stage, the data consists of the known concentrations and of the AUC computed on the ICA components observed at 220nm, for each of the 3 series, 5 replicates and 3 compounds, that were identified using Ward's algorithm.

10.4 Method validation

Results quality is a strongly regulated topic about which several guidances have been published by authorities (see ICH Q2(R1) (2005); Food and Drug Administration (2001); ISO/CEI 17025 (2005)). Indeed, authorities as well as the experimenter want to be confident about the method quality for its intended purpose.

To provide evidences of quality results for the future runs of the methods, two options to analyze their performances are presented using Bayesian predictive methodology and distributions. First, based solely on the calibration data, predictive precision and (posterior) probability profiles are derived from the distribution of the inverse predictions. Second, the strategy of accuracy profiles is applied using the data from both the calibration and the validation sets. In both options, series were made in such way to represent as close as possible the conditions that will be met during the routine analysis (change of operator or device). See Hubert et al. (2006).

10.4.1 Calibration model

The following hierarchical calibration model was applied separately for the three compounds using the data of the calibration set:

$$y_{ij} = (\beta_0 + \alpha_{0j}) + (\beta_1 + \alpha_{1j})x_i + \varepsilon_{ij}, \tag{10.3}$$

with y_{ij} being the logarithm of the observed response (AUC) for the i^{th} repetition and the j^{th} serie, and x_i is the logarithm of the results (concentrations). Logarithmic transformation is applied to ensure positivity of the results and of the responses and to provide an homogenous variance over the concentration range. β_0 and β_1 correspond to fixed effects. α_{0j} and α_{1j} represent the additional variability induced by the serie j and are random parameters. The Bayesian model and the related non-informative priors were presented in Chapter 4, page 61. The advantage of using such model is the ability to easily obtain the distributions of the predictions and of the inverse predictions, accounting for the effect of the series, and potentially using prior information if available. The generic use of prior information is unfortunately far from being adopted in the method validation context, although numerous data sets and analysis reports are available.

Using Monte-Carlo simulations from the joint posterior distribution of the parameters, the predictive distribution of the back-calculated concentrations $(\tilde{x} \mid \tilde{y}, \text{data})$ given a new response $\tilde{y}$ was thus computed, as presented in Figure 10.5 (blue, $\tilde{y} = 3.8$).

10.4.2 Validation of the method

Method quality can be analyzed using different tools. Among other, the precision profile and the probability profiles are interesting plots constructed using the calibration data set only. They can then be available without a validation set. Another

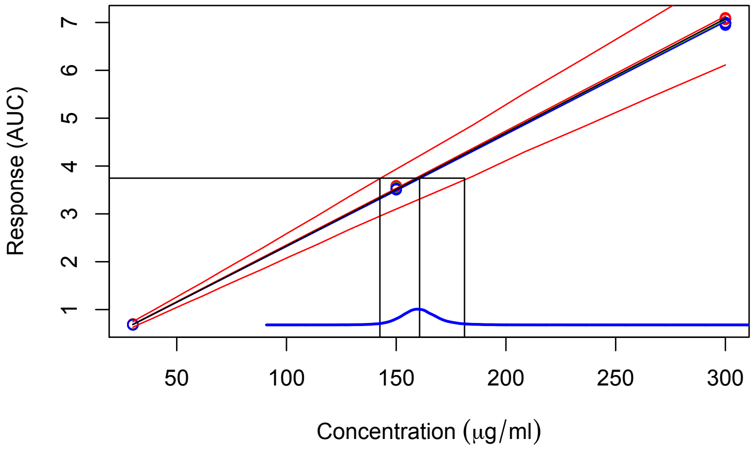


Figure 10.5: Inverse prediction using the hierarchical linear calibration model.

tool is the accuracy profile, that is computed using both the calibration set to create several calibration curves, and the validation set to assess the inverse prediction of new samples.

Precision profile

The precision profile is a graph of the (predictive) coefficient-of-variation (CV) of new concentrations versus the assumed true or mean concentrations (Dmitrienko et al., 2007). The predictive CV is calculated as follows:

$$\mathrm{CV}_{\tilde{x}} = \frac{100 \times \mathrm{sd}(\tilde{x} \mid \tilde{y}, \mathrm{data})}{\hat{E}(\tilde{x} \mid \tilde{y}, \mathrm{data})},$$

where $\mathrm{sd}(\tilde{x})$ is the estimated standard deviation of $\tilde{x}$ and $\hat{E}(\tilde{x})$ can be taken as the mean of the distribution of $\tilde{x}$. Both sd and $\hat{E}$ are computed from the Monte-Carlo samples of the predictive distribution of the inverse prediction (back-calculated results). In this example, $\tilde{y}$ is the log-AUC of a new sample.

The precision profile for the phenytoine is illustrated on Figure 10.6. A specification was set up as follows: the $\mathrm{CV}_{\tilde{x}}$ must be lower or equals to 10%. When the specification is achieved, the method quality is assumed sufficiently high. As observed on the precision profile, the range of concentrations between 29.95 ($\sim$ 30) and 298.85 ($\sim$ 300) μg/mL fulfill the specification and is then defined as the dosing range of the assay, i.e., the range of concentration for which the new measures will be satisfactory w.r.t. the specification.

Similar profiles were obtained for granisetron and sulfinpyrazone, but are not

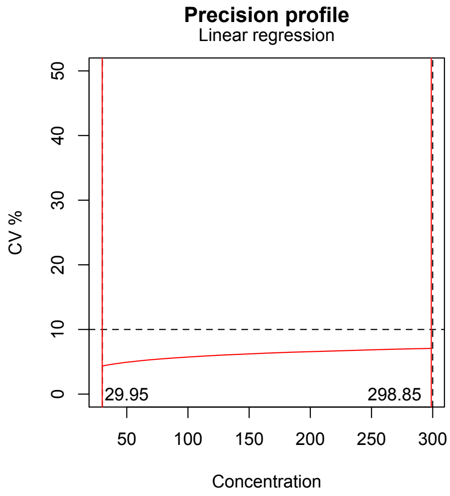


Figure 10.6: Precision profile of phenytoine. A specification is fixed as $\mathrm{CV}_{\tilde{x}} \leq 10\%$ (horizontal dashed line). Concentrations satisfying the specifications are in the dosing range of the method (vertical red lines).

presented here.

Probability profile

Another way to assess the performance of the method is to represent the guarantees that future runs will be within predefined specifications with regards to their uncertainty. These specifications also apply on the back-calculated concentrations $\tilde{x}$. Let the following specification by applied: the experimenter want to know the guarantee and the risk for the method to have a predictive uncertainty constrained into $\lambda\%$ of the mean predicted concentrations ($\lambda = 10$). Monte-Carlo simulations were used to compute the probability that the specifications will be satisfied:

$$P_{\tilde{x}} = P(\tilde{x} \in \Lambda \mid \tilde{y}, \text{data}) \quad \text{with } \Lambda : \hat{E}(\tilde{x} \mid \tilde{y}, \text{data})[100\% \pm \lambda]. \tag{10.4}$$

Repeating the computation for every concentration level, a probability profile was plotted and compared to a minimal quality level π (here, $\pi = 90\%$), as shown on Figure 10.7 for the phenytoine. The dosing range computed from the probability profile is then the concentration values for which $P_{\tilde{x}} \geq \pi$. On the basis of the calibration data, the dosing range is $[29.95, 234.57]$ μg/mL. Similar probability profiles were computed from the two other compounds but are not presented here.

A similar profile can be obtained using $1 - P_{\tilde{x}}$ instead of $P_{\tilde{x}}$. This defines a risk profile expressing the risk not being within specifications. The main interest of both

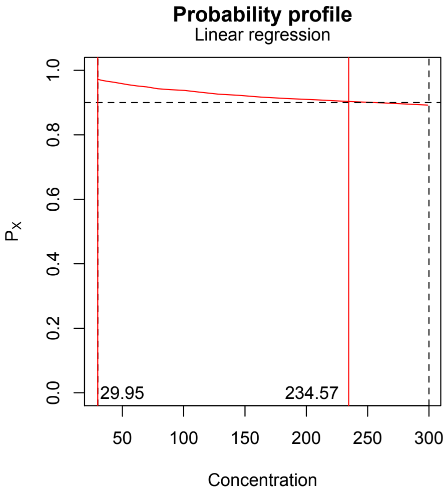


Figure 10.7: Probability profile to observe the measure within 10% of the mean concentration of phenytoine. Minimal quality level is set to 90% (horizontal dashed line). Vertical red lines identify the dosing range.

profiles is the ability to control the risk to make wrong decision by adapting π and λ with respect to the situation.

In this manuscript, there is a clear similarity between the computations of the dosing range in the present chapter, and the Design Space approach presented in the previous chapters. Indeed, for the quantitative analytical method, the dosing range is its univariate Design Space, i.e. the range where quality is demonstrated.

Accuracy profile

The two previous profiles were computed using the predictive uncertainty (i.e. the intermediate precision) of the analytical method, including two sources of variation, but lacked the inclusion of a bias estimation. To account for this bias, a possibility is to compute difference between the found and the true concentrations of the validation set.

For a given serie j, the concentration of a compound from the validation set $x_{j,val}$ is back-calculated from the response $y_{j,val}$ (log-AUC) using the mean calibration curves fitted with the calibration data of the same serie (Hubert et al., 2006). $x_{j,val}$ is then compared to the true results $\mu_{j,val}$, obtained using a reference method. In this way, the precision and bias of several $x_{val,11}, ..., x_{val,ij}, ..., x_{val,nm} = \mathbf{x}_{val}$ are both included in the validation process, when compared to the corresponding references $\boldsymbol{\mu}_{val}$.

According to Hubert et al. (2007) and Rozet et al. (2011), β-expectation tolerance intervals computed on the results in relative scale, $(\mathbf{x}_{val} - \boldsymbol{\mu}_{val})/\boldsymbol{\mu}_{val}$, are one of the most suitable solutions to simultaneously combine the systematic error (bias) and the random influences encountered in the data (spread or precision). The β-expectation tolerance interval allows predicting where, on average, a stated proportion β of future results will be found. From a Bayesian predictive perspective, tolerance intervals are simply computed from the posterior predictive distribution new results as demonstrated by Guttman (1970).

β-expectation tolerance intervals are then computed on the (relative) back-calculated results for each concentration level found in the validation set (here: 30, 150, 300 μg/ml). For this purpose, the data were modeled with a random one-way ANOVA to estimate the effect of the series, as introduced in Chapter 4.2.

For the following results, the frequentist and the Bayesian tolerance intervals were computed. These relative tolerance intervals were compared to acceptance limits that defines minimal quality (i.e. specifications). This allows drawing the accuracy profile. Acceptance limits must generally be defined following regulation requirements but might also be driven by the objectives of the quantitative method. For the following results, the acceptance limits were arbitrary set at ±15% of the true concentrations, given the complexity of the validation standards preparation.

The first result that concerns phenytoine is shown on Figure 10.8. For this compound, the results from the ICA computations (A) were compared to the results obtained when processing the chromatograms manually (B). Notice this comparison is only possible because phenytoine was completely separated from the two others compounds (see Section 10.3.1). The obtained Frequentist and Bayesian β-expectation tolerance intervals ($n^* = 20000$) are drawn in dark and light blue, respectively, and look equivalent.

Focusing on the levels of 150 and 300 μg/mL, the data from the automated and the manual peak integration showed similar uncertainty, with a total relative error (i.e., the relative β-expectation tolerance intervals) below 5%. However, at the lowest concentration level (30 μg/mL), the profiles were different. On one hand, the manually integrated data (B) seemed to concede a sufficient quality (i.e. accuracy) for the quantitative method, with a small bias. On the other hand, the automatically treated data (A) showed a higher negative bias, although the intermediate precision (i.e., the total uncertainty) looks similar in both profiles.

The reasons of this difference can be identified using the original data or the sources given by ICA. On the sources (see Figure 10.2 (B)), a specificity problem was identified for phenytoine, which elutes at the same time than an impurity of granisetron. Moreover, phenytoine degradation might have begun during the validation experiments. First, the degradation certainly induces a loss of precision for

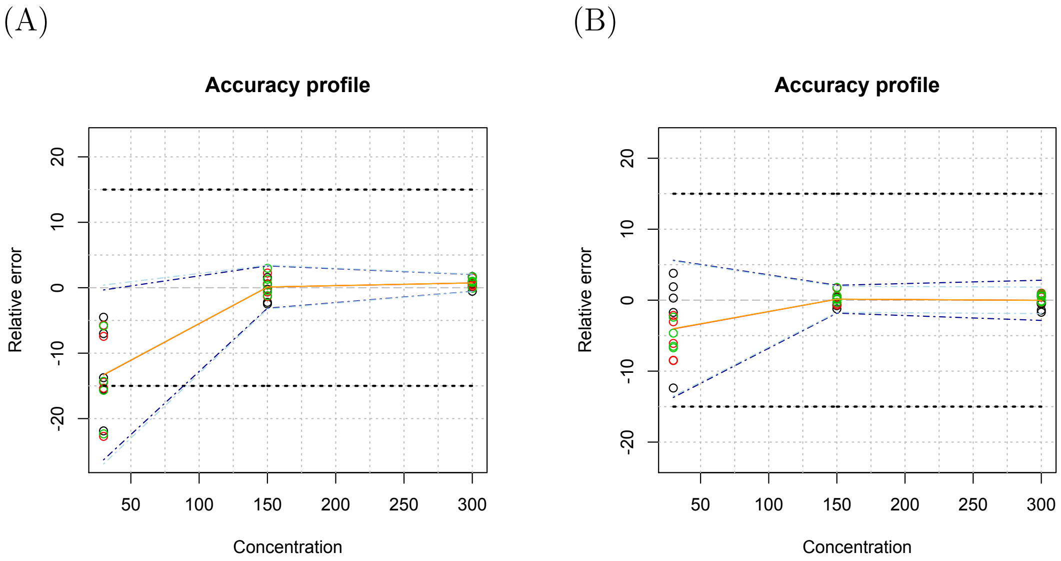


Figure 10.8: Accuracy profiles for phenytoine. (Dots) results in relative scale (colors identify series). Frequentist (dashed dark blue line) and Bayesian (dashed light blue line) β-expectation tolerance intervals expressed in relative scale. (Red) relative bias. (Dotted black) Acceptance limits. (A) data obtained using the ICA automated methodology. (B) data obtained by integrating peaks manually.

the manual and the automated methods, as these degradations might not be similar in every samples. Second, when envisaging the manual treatment, the impurity of granisetron was integrated with phenytoine, then (un-)fortunately reducing the bias on this peak, with a higher effect when the concentration of granisetron is high (inducing an higher concentration of the impurity) and the concentration of phenytoine is low.

A manual check allowed confirming these results: with the help of ICA to detect the position of the impurities, it was possible to identify them in the original chromatograms. The purity of the solutions were not tested because it was not expected to observe degradation during the short total run time of the experiments. As conclusion, the experimenter should choose the profile from ICA-treated data because specificity of the original data is not good, even if the peak was thought to be well separated. In summary, ICA permitted improving the specificity and, in the same time, showed that the results were biased at the lowest concentration level. This was probably due to a lower concentration of phenytoine than the expected one, caused by the apparition of the impurity during the validation experiments.

Finally, accuracy profiles for the two coeluted peaks were computed and drawn. They are illustrated on Figure 10.9 for sulfinpyrazone (A) and granisetron (B). As no sufficiently good manual technique to unmix these peaks was found, no comparison is done with manually treated data.

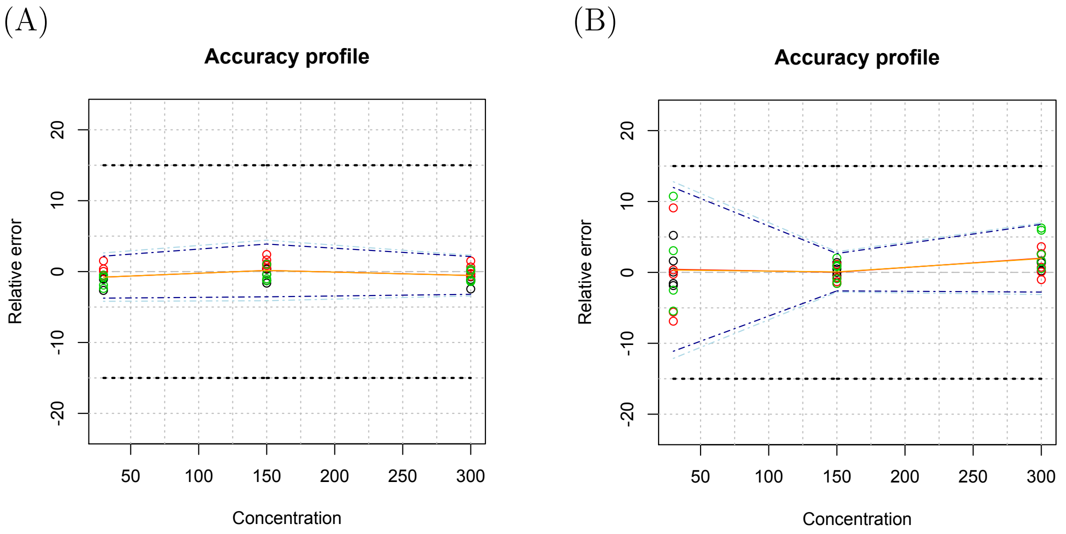


Figure 10.9: Accuracy profiles for coeluted sulfinpyrazone (A) and granisetron (B) automatically treated using ICA. Frequentist (dashed dark blue line) and bayesian (dashed light blue line) β-expectation tolerance intervals expressed in relative scale. (Red) relative bias. (Dotted black) Acceptance limits.

For both compounds, the specification was kept at 15% of the nominal concentrations. Both profiles were found acceptable for a dosing range covering all the concentrations used during the experiments. As previously, ICA was able to numerically separate the peaks in order to recover the specificity of both analytes. The detected impurity of granisetron does not seem to affect its profile, except on the lowest concentration level where the precision is close to the specifications.

Notice that to be fully compliant with ICH Q2, several other validation results must be given, such as the linearity of the calibration curve, the trueness, the detection and quantitation limits, the precision, etc. These results are cut here, for brevity.

10.5 Conclusion

An example of method validation for several compounds was presented. The proposed analytical method was a high-performance liquid chromatography with a diode-array detector, resulting in matrix-like signal. The input factors were chosen so that the method was suboptimal, with two strongly coeluting peaks. In this case, the method validation can generally not been carried out as method specificity is one of the first method quality to demonstrate following ICH Q2.

Independent component analysis was shown helpful in this case to simply locate all the peaks in the chromatograms while numerically separating them. From this first identification, an automated peak matching procedure was possible using the retention times at the apex of the ICA-recovered peaks, to discriminate them. This procedure was simple because the HPLC operating condition was always similar.

For each of the compounds, different quality profiles were used to analyze the quantitative results of the method. First, based solely on the calibration data, a precision profile and a probability profile have been computed. They allowed a first assessment of the results precision based on hierarchical Bayesian predictive models to account for a series effect. Second, the strategy of accuracy profile was used, allowing the inclusion of the observed bias in the results. Each profile allowed deriving critical quality attributes such as the dosing range of the method.

ICA definitively opens new perspectives toward the fully automated treatment of chromatograms. In this validation study, some of the results obtained after ICA were at least as good as the results obtained manually, while other results could not be obtained without an appropriate numerical separation.

The first analysis focused on a single peak. Comparing manual and automated integration after ICA, the superiority of the automated process using ICA was shown. Indeed, it permitted detecting and isolating some impurities, thus improving the specificity of the method. Second, the successful application of ICA shown that strong coelutions are not a barrier against method validation. Indeed, the results obtained for the two coeluted peaks also fully met the expectations. It would be difficult to reproduce such results without the help of ICA.

ICA and the proposed automated treatments could then be used to drastically solve the time-consuming problem of quantitative methods validation while increasing the quality of information extracted from the chromatograms. However, the approval of such methodology by the authorities and regulatory bodies might be doubtful given its complexity.

10.6 Further works

In this study, the integration of the peaks were made at one wavelength of interest (220 nm). This wavelength was chosen to obtain a one-dimensional signal on which it is easy to work. It is a usual analytical practice to have an acceptable absorbance for all the compounds under investigation, to improve the signal-to-noise ratio. With manual integration, it must clearly be chosen prior to further analysis.

The automated treatment opens the perspective to compute the profiles for every wavelengths. After this intensive computation, it is rather obvious to select the wavelength(s) providing the best profiles. This can allow the automated selection of a wavelength that shows i) acceptable absorbance, ii) no influence from solvent or other artifacts, iii) no influence of ICA estimation, and iiii), the best results given the purpose of the method.

For the point iii), it was observed that in some cases, ICA can have difficulties to properly unmix UV-signatures, mainly when they are too similar in some ranges of the UV spectra (see e.g. Figure 6.10 (Bottom), page 108). Improvements might be obtained by using more specific implementation of ICA, such as the non-negative ICA (see Yuan and Oja, 2004; Zheng et al., 2006).

Chapter 11

Validation and routine of the ligand-binding assay

This chapter has been written using materials from three talks made in international conferences (Lebrun and Boulanger, 2010; Lebrun et al., 2010; and Lebrun et al., 2011).

11.1 Introduction

Ligand-binding assays (LBA), such as the Enzyme-Linked Immuno-Sorbent Assay (ELISA), are widely used in life sciences. They allow the quantification of endogenous analytes such as proteins or peptides in biological samples. The biological properties of the analyte of interest are used to capture them by means of plates pre-coated with a specific antibody, that is able to link the analyte. The process is followed by an immunological detection of the specifically captured analyte by a conjugate enzyme. With the addition of a substrate, the enzyme provides a reaction resulting in some coloration. The coloration is related to the amount of enzymes and the amount of analyte. It can be detected through measurement of the optical density (signal) of the product within the plates.

Generally, the plates are some sets of wells that allows to make different measurements simultaneously, possibly with some operating conditions that may be changed. Figure 11.1 illustrates one plate with its wells having different optical densities.

The general purpose of LBA is to provide results that are used to make decisions such as the release of a production batch, stability studies, the PK of a drug (e.g. that might interact with some proteins used as biomarkers), the optimization of

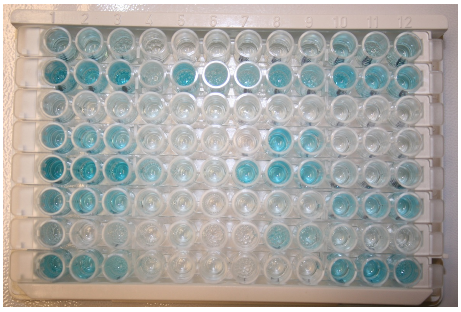


Figure 11.1: Plate for ELISA test.

the dose of a drug, the optimization of a process, etc. Therefore, the quantitative results given by the LBA must be as **accurate** as possible, otherwise the risk to make wrong decision or to provide low quality drugs to patients would be high.

In practice, an assay should first be validated and then used during routine. The objective of validation is to prove the quality of the results for the **future use** of the LBA (Khan and Findley, 2010, chapter 5).

11.1.1 Data

To make a LBA quantitative, calibration standards are prepared in appropriate matrices (e.g. blood sample, urine) from known stock solutions of the analyte. Various concentrations can be obtained using serial dilution of the initial solutions. François et al. (2004) analyzed the optimal design for different calibration models and with the inverse prediction problem. As conclusion, equidistant data points (on the log-concentrations scale) around the EC50 are close to optimal when envisaging a four-parameter logistic regression, as frequently occuring with LBA. Example of calibration data are given in Figure 11.2 (simulated data). For different concentrations, different signals would be obtained, and a calibration model is envisaged to describe the data.

11.2 Validation, setting of critical quality attributes

Assay quality is defined by the results quality. Several values that reflect the quality of the results may be defined as critical quality attributes (CQAs).

The CQAs should be set up to reflect the use of the LBA. When new experiments are carried out, new signals are obtained. Inverse prediction is done to provide an

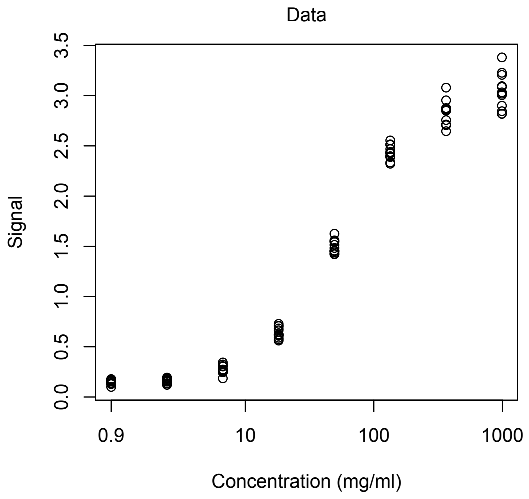


Figure 11.2: Example of data for the calibration of LBA.

estimation of the real concentrations of the analytes in the samples. Thus, the distribution of these back-calculated concentration seems appropriate when defining the CQAs.

11.2.1 Model

A four-parameter logistic (4PL) regression is generally well adapted to fit data of LBA assay (Findlay and Dillard, 2007). Often, a heteroskedastic variance is observed. The simulated data have been generated to have this property, as illustrated on Figure 11.2. To account for this particular variability, Davidian and Giltinan (1995) proposed that the model includes a variance (or a precision) that is modeled as a power of the expected values (POM). Thus, if y_i is an observed signal and x_i is the known analyte concentration, a simple fixed effects model for the 4PL regression is given by:

$$y_i = \beta_1 + \frac{\beta_2 - \beta_1}{1 + (\frac{x_i}{\beta_4})^{\beta_3}} + \varepsilon_i, \text{ and } \varepsilon_i \sim N(0, \tau_Y), \tag{11.1}$$

with $\tau_Y = \frac{\tau}{E[y_i]^\theta}$.

If different batches are used, it is possible to account for the induced variability using the model presented in Chapter 4.4, page 65. In this example, four batches ($m = 4$) were used with 3 replicates ($n_j = 3$), so the mixed-effects model holds. 200000 simulations using vague *a priori* were done and the chains were thinned to keep one sample out of ten in order to reduce autocorrelation. The chain for each

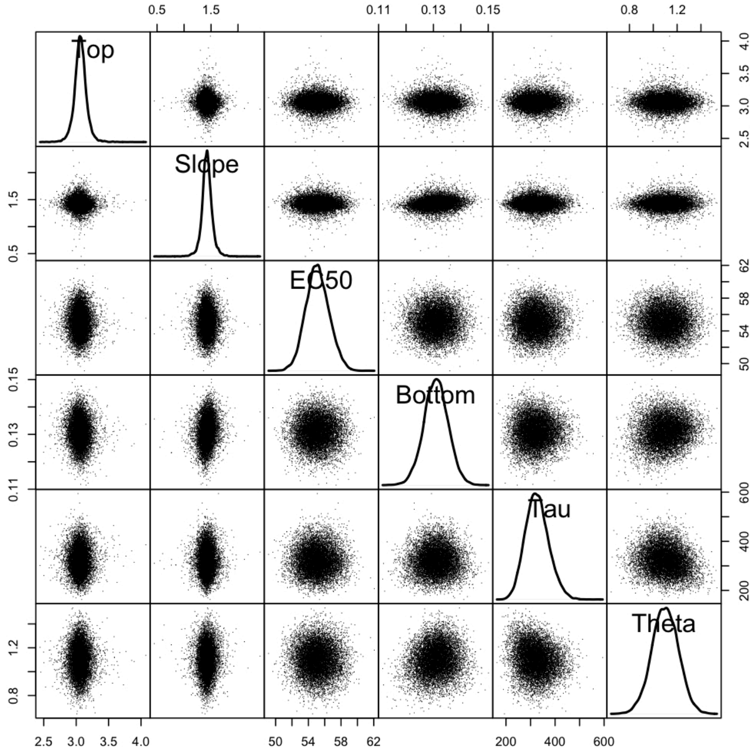


Figure 11.3: Bivariate joint distributions of the model parameters.

parameters then contained n^* =20000 samples. Figure 11.3 presents the bivariate joint distributions of the parameters. No special correlation structure are observed from the simulated data. The marginal posterior distributions of the top asymptote and of the slope are rather peaked. Due to the mixed-effects modeling, they certainly depart from the log-normality assumption of the prior.

From the parameter chains, Chapter 4.4 presented how to produce samples from the predictive distribution of an inverse prediction $\tilde{x}$ given a new observed signal $\tilde{y}$, i.e., samples following the density $p(\tilde{x} \mid \tilde{y}, \text{data})$. Monte-Carlo simulations were used to propagate the uncertainty of the joint posterior distribution of the parameters to the distribution of the inverse prediction.

Figure 11.4 illustrates the fit of the mixed model with POM variance. The 95% Bayesian predictive interval is represented in red and is equivalent to the β-expectation tolerance interval for a new signal Guttman (1970). The mean prediction over the 4 series is drawn in black.

Assume a first new signal is observed at $\tilde{y}_1 = 1.0$. In this case, the inverse prediction is close to the inflection point (C50) of the 4PL curve. The uncertainty of the distribution (blue densities) is rather limited. On the other hand, when a second signal is observed at $\tilde{y}_2 = 2.7$, the inverse prediction is closer to an asymptote. In that case, the uncertainty is logically higher. It is common to loose quality of prediction when an observed signal is close to the limits of the system.

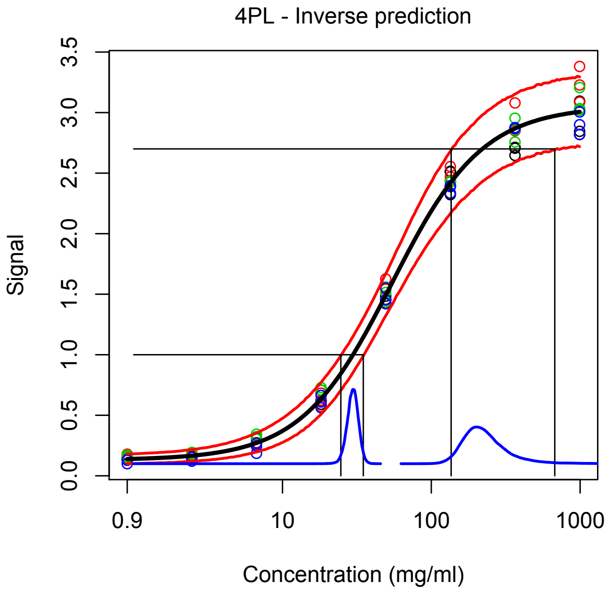


Figure 11.4: Inverse prediction using the Bayesian 4PL model.

11.2.2 Derivation of critical quality attributes

The objective of the validation of the assay is to give guarantee that each new measure will be close enough to the unknown true amount. CQAs are developed to summarize the quality of the assay. They must provide an appreciation of the original data and of the calibration curve. To do so, it is first envisaged to develop two predictive profiles, on which CQAs can be directly computed : the precision profile and the probability profile.

Precision profile

The precision profile is a characterization of the precision of the inverse predicted concentrations of new samples, using the calibration curve. It may be drawn by plotting the coefficient-of-variation (CV) of new predicted concentration versus the assumed true concentration on a log scale (Dmitrienko et al., 2007). As the true concentrations are generally unknown with the serial dilution assay, the expected predicted concentration may be used instead. The predictive CV is then calculated as follows,

$$\mathrm{CV}_{\tilde{x}} = \frac{100 \times \mathrm{sd}(\tilde{x} \mid \tilde{y}, \mathrm{data})}{\hat{E}(\tilde{x} \mid \tilde{y}, \mathrm{data})},$$

where $\mathrm{sd}(\tilde{x})$ is the estimated standard deviation of $\tilde{x}$ and $\hat{E}(\tilde{x})$ can be taken as the mean (or possibly the median or the mode) of the distribution of $\tilde{x}$. Both sd and $\hat{E}$ have been computed using Monte-Carlo simulations.

Figure 11.5 illustrates the precision profile computed from the data of Figure 11.4. A specification was fixed so that the CV must not be higher than 20%. The range of concentrations between 1.76 and 228.15 mg/ml is the dosing range of the assay, i.e. the range of concentration for which the new measures will be acceptable given the specification.

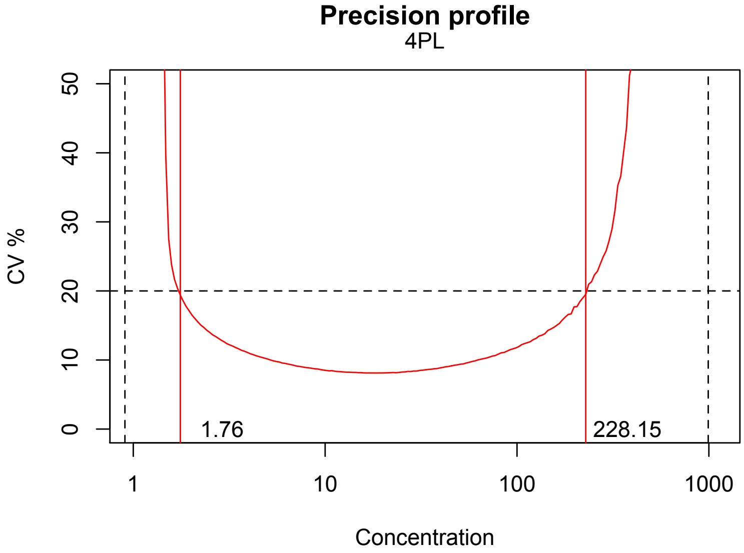


Figure 11.5: Precision profile. A specification is fixed as CV≤20% (horizontal dashed line). Concentrations satisfying the specifications are in the dosing range of the method (vertical red lines).

Probability profile

Monte-Carlo simulations can also be used to compute different risk-based results from the predictive distribution of the inverse prediction. For instance, Assume some specifications holds on the predicted concentrations. Next, simply compute the probability that the specifications will be satisfied (i.e. the guarantee), or the probability that the specifications will not be satisfied (i.e. the risk).

In the example, the specification was set so that the uncertainty around $\hat{E}(\tilde{x} \mid \tilde{y}, \text{data})$ must not be greater than $\lambda\%.E(\tilde{x} \mid \tilde{y}, \text{data})$, with $\lambda = 20$. It is then possible to compute the probability $P_{\tilde{x}}$ to achieve the specification:

$$P_{\tilde{x}} = P(\tilde{x} \in \Lambda \mid \tilde{y}, \text{data}) \quad \text{with } \Lambda : \hat{E}(\tilde{x} \mid \tilde{y}, \text{data})[100\% \pm \lambda]. \tag{11.2}$$

Repeating the operation for every signal in the range of interest, a probability profile can be plotted, as shown on Figure 11.6. In this example, a minimal quality level of $\pi = 90\%$ was defined. The dosing range is defined as the concentration range for which $P_{\tilde{x}} \geq 0.9$.

A similar profile can be obtained using $1 - P_{\tilde{x}}$ instead of $P_{\tilde{x}}$. This is a risk profile expressing the risk not being within specifications. The main interest of the probability or risk profiles is the ability to control the risk to make wrong decision by adapting λ with respect to regulations, if any, and π with respect to the situation.

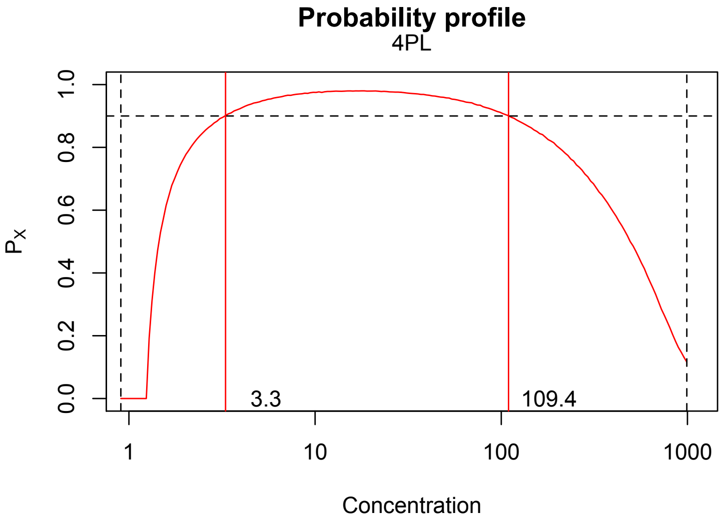


Figure 11.6: Probability profile to observe the measure within 20% of the mean concentration. Minimal quality level is set to 90%. Vertical red lines are the dosing range.

Example of CQAs

The profiles are good tools to control the assay quality. Different CQAs can be extracted from them:

- The size of dosing range,
- the area under the probability curve,
- the area under the precision curve,
- the lower limit of quantification,
- the average precision over the concentration range,
- etc.

The two first CQAs are values to be maximized while the three last are intended to be minimized. For each, various specifications can be given, driven by economical, efficiency or regulatory reasons. As they are based and/or represent similar information, each CQA might be strongly dependent on each other.

The assay will remain valid for operating conditions and concentration for which the envisaged CQAs are within specification. For instance, the dosing range was computed from the probability profile as the concentrations for which the uncertainty on the future results was limited with high guarantee. Furthermore, if the dosing range (i.e., the CQA) covers the actual values of concentration for which the assay is intended to be used (i.e., the specification), the assay is declared valid. The parallelism between the dosing range and the Design Space as presented in the previous chapters is direct.

With this in mind, CQAs might be used as responses for the screening or the optimization of the operating conditions of the LBA (Lebrun et al. (2010)). Obviously, it is necessary to estimate one calibration curve for each operating condition of the experimental plan. With such approach, the experimenter may want to reduce the cost. Fortunately, in an optimization context, it is generally less important to estimate random effects and it is often possible to use one LBA plates for several well-chosen operating conditions.

The model presented in the next section simplify the mixed-effects model presented in Chapter 4 to a simpler fixed-effects models. It may be used with non-informative prior distributions in optimization context, and can be used with informative prior in routine.

11.3 Routine of LBA

During routine, new calibration experiments are done for each new run. Some wells on the plates are often reserved for this purpose. As only one “series” of data is analyzed each time, calibration curves do not include random effects.

The costly data gathered during the validation experiments represents an opportunity to improve the models used in routine. Indeed, using the information from the posterior distributions of the parameters obtained during validation, it is possible to help the definition of prior distributions for the routine. The inclusion of prior knowledge may then lead to decrease the uncertainty of the parameters and then to improve the calibration models, while also decreasing the number of experiments for the new calibration curves that will be made.

Without the estimation of random effects, the model presented in Section 4.4 can be simplified and is defined as in Equation 11.1. Prior distributions are defined as follows:

$$\begin{array}{ll} p(\beta_1) = lN(b_1, \tau_1) & p(\beta_2) = lN(b_2, \tau_2) \\ p(\beta_3) = N(b_3, \tau_3) & p(\beta_4) = lN(b_4, \tau_4) \\ p(\tau) = Gamma(a, b) & p(\theta) = Gamma(c, d) \end{array} \tag{11.3}$$

11.3.1 Update of prior from validation experiments

Assuming the prior and posterior distribution are approximately similar and the correlation structure between the parameters is limited. It is then possible to use the information gathered during the validation experiments. Table 11.1 references the update rules to compute the prior parameters $b_{1-4}, \tau_{1-4}, a, b, c, d$ for the routine, from the posterior distributions obtained during validation. The first column is the assumption on the prior distribution of the parameters, while the second column is the way to compute the prior parameters from available information in the posterior distributions obtained in the validation step.

Prior distribution	**Parameters update**
$\beta \sim N(\mu, \tau)$	$\mu = E(\beta \mid data)$ $\tau = V(\beta \mid data)^{-1}$
$\beta \sim lN(\mu, \tau)$	$\mu = \log\left(E(\beta \mid data)\right) - 0.5 \log\left(1 + \frac{V(\beta \mid data)}{E(\beta \mid data, I)^2}\right)$ $\tau = \log\left(1 + \frac{V(\beta \mid data)}{E(\beta \mid data)^2}\right)^{-1}$
$\tau \sim Gamma(s, r)$	$s = E(\tau \mid data)^2 / \ V(\tau \mid data)$ $r = E(\tau \mid data) \ / \ V(\tau \mid data)$

Table 11.1: Update of the prior parameters from the posterior distributions of the parameters.

With the proposed 4PL mixed-effects model of the validation step, the prior distributions were not conjugate with the likelihood. Furthermore, the model for the routine is not identical to the one used for the validation. Thus, the marginal

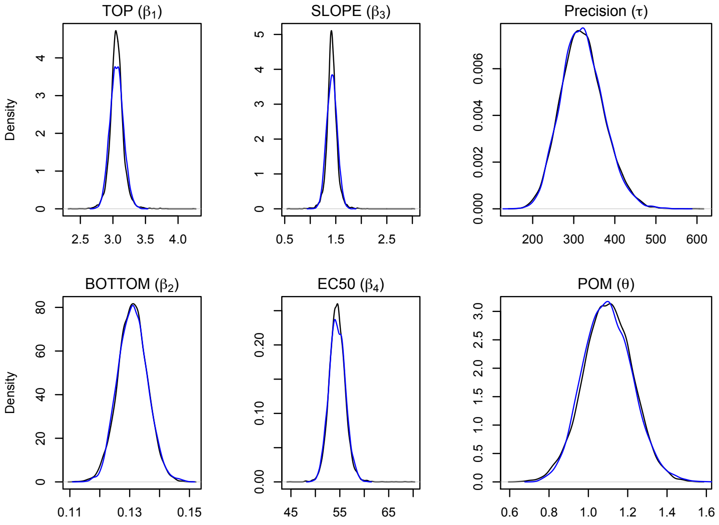


Figure 11.7: Comparison of the posterior distribution from validation (black) and the new informative prior distributions (blue).

posterior distributions are certainly not from the same distribution families. The proposed prior distributions in Equation (11.3) could however be considered as an approximation of the posteriors. Of course, dependencies that might exists between the parameters are lost during the update. However, this dependencies have been shown limited for this simple example, when observing the bivariate joint distributions of the parameters.

In this case, a careful comparison of the posterior densities (obtained via MCMC simulations) and of the updated prior distribution is advised. As shown on Figure 11.7, the agreement of the distributions for τ, θ and β_2 is satisfactory, while the three effects considered as random during validation experiments (β_1, β_3 and β_4) are not perfectly fitted by the updated prior.

Limited knowledge on the posterior, poor mixing of the chains, and small number of d.f. might be of concern. In this case, the d.f should be analyzed because, from the linear regression framework, it is assumed that the marginal posterior distribution of regression parameters follows some Student's distributions. However, it is not straightforward to obtain approximations of these d.f. for mixed-effects nonlinear models (Fong et al., 2010).

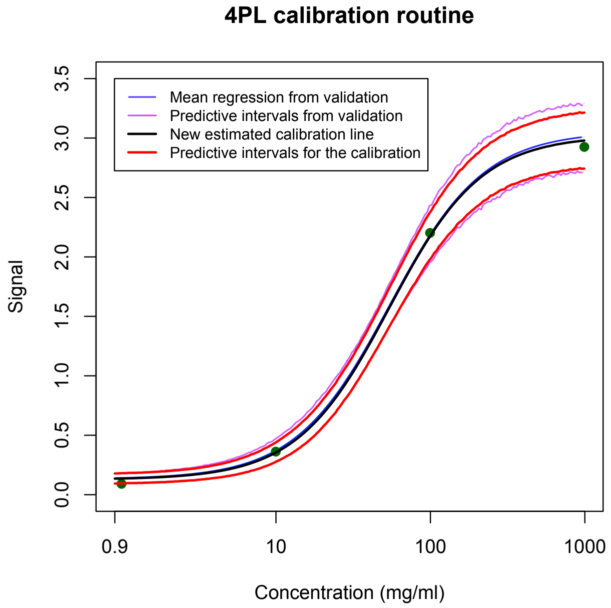


Figure 11.8: New calibration curve fitted using prior information from the validation experiments and four new data points.

11.3.2 New routine data

New data and updated prior distributions can be combined using Bayes' theorem and sampling methodologies to obtain the posterior distribution of the parameters for the routine. Thanks to the strong prior information coming from validation, it is possible to rely on very few data points. In this example, only one new data point was taken for each of the following concentration levels: 1, 10, 100 and 1000 mg/ml. [1] This is then a very light design for 4PL calibration and the model can not be adjusted without the prior knowledge. For instance, there is no support for the estimation of the additional parameters for POM variance.

The new calibration curve is presented in Figure 11.8 (thick black line) and is close to the mean regression line observed during the validation (grey line), while still appropriately adjusted to the 4 new data points. The 95% predictive interval of this new calibration (red) is included with the one computed during validation (light red). About 95% of future data of this particular new batch will be included in the new interval. Also, *at least* 95% of future data of this particular new batch will be included in the validation predictive interval. It is noted that the increased variance due to the series effect during validation has then a more limited effect for

[1] Notice also that if the validation data allowed getting a better idea of the EC50, the experimental design could be adapted accordingly.

the routine (although it is present in the updated prior distributions).

Regarding the uncertainty, The POM variance has a clear effect, although the new data points did not allow its identification.

Finally, depending on the number of new data points and of the prior precision, the new calibration curve might be biased towards the regression obtained during validation. This is not necessarily a problem if the validation was made in order to mimic the true routine of the assay. Small simulation studies might be carried out in order to assess the sensitivity of the updated prior and to help powering the new calibration with enough data points.

11.4 Conclusion

In the Bayesian setting, samples from the predictive distribution of the inverse prediction are simple to obtain using simulations, even with the (unbalanced) non-linear mixed-effects model and the addition of a heteroskedastic variance. It is then an appropriate way to gather the right information about the assay and its future performance.

The Bayesian analysis was then used for the validation of a ligand-binding assay and for new calibrations during its routine use. It allowed to predict the future behavior of the methods to ensure the predictive probability of accurate results is high. Information from validation was transferred to routine through the update of prior distributions.

Another use of the statistical models and the Critical Quality Attributes, mentioned in this chapter, is the optimization of the ligand-binding assay with regards to various operating conditions. This was the topic of two presentations focusing on a risk-based Design Space approach to simultaneously optimize and assess the robustness of the developed assay (Lebrun and Boulanger, 2010; Lebrun et al., 2010).

General conclusions

Given the guidelines and regulatory documents making authority for the Pharmaceutical Development, their is a pressing need from the pharmaceutical industry to obtain precise methodologies to apply the Quality by Design paradigm for their processes and methods. Indeed, documents such ICH Q8 explain the principles for drug developments in a quality environment, and the expectations of the regulatory authorities, but no precise methodology is proposed given the broadness of the application domain.

A first step was to understand the flaws of the classical optimization strategies, including Design of Experiments and Multi-Criteria Optimization techniques. Chapter 1 explained why the classical use of these tools is generally not sufficient, even if they aim at improving the quality and the understanding of processes or methods. The main reason is that the most of the softwares and toolboxes that address this problematic make use of the mean responses predicted by statistical models only, without taking into account the uncertainty and responses dependencies. If quality can be improved using classical methodologies, there is limited possibilities to provide *assurance of quality* in a predictive risk-based framework.

To integrate the uncertainty of the response predictions for the future use of the method or process under development, fully predictive statistical models were developed. In this context, the Bayesian statistical framework has been helpful to obtain the (posterior) predictive distribution of new predictions using simulations or mathematical derivation. This was made for various classes of statistical models that were presented in Chapters 2–4.

The proposed methodology was then not intended to provide optimal solutions based on *mean predictions*. Instead, it was developed to define the settings of the operating conditions where high quality can be guaranteed. The minimal required quality was expressed with specifications defined on Critical Quality Attributes (CQAs). It was illustrated that, when these high guarantee-settings allowed variations in the operating conditions, robustness of the methods and process can be easily demonstrated, which is the direct application of the Design Space defined by ICH Q8.

Finally, a last general topic was addressed: as the Critical Quality Attributes are multivariate by nature, the Multi-Criteria Optimization problematic was reviewed and extended in Chapter 5, to obtain risk-based results based on the predictive distribution. The classical approach of computing the joint probability of acceptance was shown to be easily carried out using Monte-Carlo simulations form the predictive distribution. This simulation approach was then extended to compute the probability to have specifications flexibly accepted using desirability methodologies. It was shown how the classical approach is a particular case of the desirability approach, when desirability functions are properly parametrized.

Analytical method development. High performance liquid chromatography is one of the most widespread analytical methods in the laboratories of the pharmaceutical industry. In this work, the first example of application of a Quality by Design compliant methodology was made on this particular process: the analytical separation methods (Chapters 7 and 8).

Considering the complex nature of the Critical Quality Attributes, it was decided to model simpler responses – the chromatographic retention times – that eventually allows the computation of the CQAs, i.e. the separation or the resolution of the critical pair of peaks, the total run time, etc. As no closed-form can be found for the predictive distribution of the future observations of CQAs, it was decided to use Monte-Carlo simulations to propagate the predictive error from the retention times to the CQAs. Monte-Carlo simulations were found powerful as the correlation structures of the multivariate error is preserved.

The Design Spaces computed with this methodology allowed an efficient separation for the screening and precise analysis of several sets of possibly unknown molecules, showing the way to a systematic development of robust analytical methods. This robustness allowed the methods to be simply transferred to high-end equipment and also eased the validation of their results.

In addition, to thematic of automatic development was addressed, using an automated reading of the chromatograms using the Independent Component Analysis (ICA, Chapter 6). The aim was to automatically separate from the noise the peaks, and track them among a set of chromatograms from the same experimental design. This was helpful as it allowed to achieve automation in most of the real case studies presented in this manuscript and references. When complete automation was not achieve, it permitted easing the data treatment, which was too difficult to be carried out manually with precision.

Manufacturing process. In order to evaluate the generic character of the approach, a transposition of the Design Space methodology to a small-scale manufacturing process was envisaged as a feasibility study (Chapter 9). The joint prediction of future outputs of the process was also the key to answer the objective dictated by ICH Q8: to provide assurance of quality.

For this application, the variabilities resulted in a minimal quality level that was kept relatively low in order to be able to define a Design Space. The risks of being out of specifications were identified higher than acceptable, while a mean response surface optimization might have wrongly stated that the process would run accordingly to its specifications.

One of the identified reasons was the inability of the multivariate linear model to fit part of the data properly. However, the experimental plan available was poorly designed to allow better modeling properties.

Analytical method validation. Finally, the Bayesian framework was used to derive in a simulation perspective the predictive distribution of univariate random and mixed-effect models. The choice of these models was made because they were useful for the calibration problem of assay with both linear and non-linear behaviors, and for the validation of the results provided by the analytical methods.

In Chapter 10, the validation of a chromatographic method to quantify three compounds was illustrated, in a completely automated way, using predictive models and the ICA as a deconvolution technique. Furthermore, the analytical conditions were set in such way this method validation would have been erroneous and even impossible with data treated manually.

With the proposed tools, several steps of the method life cycle were explored, including the development of robust separation methods and their validation, aiming at demonstrating the quality of their quantitative results. It was also noted that demonstrating the quality of the results of analytical methods is an important objective as these results are used to take the important decision during the drug development process, the batch releases, and also for post-marketing analysis.

In Chapter 11, a last insight was made on another type of method: the bioanalytical ligand-binding assay. The most noticeable difference with the classical analytical methods is that the calibration curves are non-linear. They follow a logistic form that must be use to take advantage of all the data. The Bayesian predictive framework also allowed providing Critical Quality Attributes about the assay future performance. It was also illustrated how to account for past experiments to set up informative prior distributions that result in very light experimental designs to collect data during routine.

Further works

The successful application of the methodologies proposed in the work leads to envisage new applications as further works. It might include other analytical separation methods such as particular liquid chromatography modes (chiral, ion exchange, HILIC, supercritical fluid, etc.), gas chromatography, capillary electrophoresis. Other manufacturing processes will also be considered in a near future, considering the very promising preliminary results obtained.

Next, it is recognized that, during the statistical process control (SPC) operated on a manufacturing process or the use if an analytical method, the use of prior information gathered during the experiments and modeling steps can improve the quality of the decision to accept or reject production batches. However, no SPC have been set up and carried out after the presented optimization processes, which could be the subject of future studies.

The methodology proposed in this work showed good results in terms of efficiency and flexibility. However, for each of the various applications, it was mandatory to code several new functions and to adapt the scripts in order to provide the Design Space and solutions.

One obvious further work is to propose the methodology as one or several toolbox(es). Fortunately, all the steps of the methodology are modular, from the model definition to the Critical Quality Attributes and Design Space computations. Specific modules can thus be coded once for every new applications.

Finally, a computer-based system can provide a way for the industry to retrieve and make use of its previous results easily, using a database system. Three advantages can be highlighted. First, if the data and the experimental design are sufficient, it can allow new optimizations without any additional experiments. For instance, the development of robust and specific methods for molecule screening and determination can then be made in a snap, as shown in Chapter 8. A second advantage is to ease the definition of informative priors that are based on previous experiments, including prior covariance structures. Finally, if changes of formulation cause new materials to be added in a drug, the number of experiences to analyze the new formulation might be drastically reduced as the experimental domain can readily be restrained to the Design Space of the previous formulations. Indeed, outside this Design Space, it was already demonstrated that the analytical method will not have a good quality, with several compounds that might be not fully separated.

Listing of publications

The complete list of publications including articles, conferences, and posters, can be found on http://orbi.ulg.ac.be/. Hereafter is a selection of peer-reviewed articles published in international journals.

Publications addressing the Design Space problematic

1. P. Lebrun, B. Boulanger, B. Debrus, Ph. Lambert, Ph. Hubert, "A Bayesian Design Space for analytical methods based on multivariate models and predictions", Accepted with minor corrections in the Journal of Biopharmaceutical Statistics (2012).

2. P. Lebrun, F. Krier, J. Mantanus, H. Grohganz, M. Yang, E. Rozet, B. Boulanger, B. Evrard, J. Rantanen, and Ph. Hubert, "Design space approach in the optimization of the spray-drying process", European Journal of Pharmaceutics & Biopharmaceutics, 2012, 80(1): 226–234. http://hdl.handle.net/2268/100593

3. M. Rafamantanana, B. Debrus, G. Raoelison, E. Rozet, P. Lebrun, S. Uverg-Ratsimamanga, Ph. Hubert, and J. Quetin-Leclercq, "Application of design of experiments and Design Space Methodology for the HPLC-UV Separation Optimization of Aporphine Alakaloïds from leaves of Spirospermum penduliflorum thouars", Journal of Pharmaceutical & Biomedical Analysis, 2012, 62:23–32. http://hdl.handle.net/2268/109321

4. B. Debrus, P. Lebrun, A. Ceccato, G. Caliaro, E. Rozet, I. Nistor, R. Oprean, F.J. Rupérez, C. Barbas, B. Boulanger and Ph. Hubert "Application of new methodologies based on design of experiments, independent component analysis and design space for robust optimization in liquid chromatography", Analytica Chimica Acta, 2011, 691:33–42. http://hdl.handle.net/2268/88084

5. B. Debrus, P. Lebrun, J. Mbinze Kindenge, F. Lecomte, A. Ceccato, G. Caliaro, J. Mavar Tayey Mbay, R. Marini, E. Rozet and Ph. Hubert, "Innovative 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", Journal of Chromatography A, 2011, 1218:5205–5215. http://hdl.handle.net/2268/93241

6. F. Krier, M. Brion, B. Debrus, P. Lebrun, A. Driesen, E. Ziemons, B. Evrard, and Ph. Hubert, "Optimisation and validation of a fast HPLC method for the quantification of sulindac and its related impurities", Journal of Pharmaceutical & Biomedical Analylis, 2011, 54:694–700. http://hdl.handle.net/2268/75222

7. I. Nistor, M. Cao, B. Debrus, P. Lebrun, F. Lecomte, E. Rozet, L. Angenot, M. Frederich, R. Oprean, and Ph. Hubert, "Application of a new optimization strategy for the separation of tertiary alkaloids extracted from extracted from Strychnos usambarensis leaves", Journal of Pharmaceutical & Biomedical Analysis, 2011 56:30–37. http://hdl.handle.net/2268/89785

8. R. Marini Djang'Eing'A, J. Mbinze Kindenge, M. L. A. Montes, B. Debrus, P. Lebrun, J. Mantanus, E. Ziemons, C. Rohrbasser, S. Rudaz, and Ph. Hubert, "Analytical tools to fight against counterfeit medicines", Chimica Oggi, Chemistry Today, 2010, 28(5):10-14. http://hdl.handle.net/2268/62379

9. B. De Backer, B. Debrus, P. Lebrun, L. Theunis, N. Dubois, L. Decock, A. Verstraete, Ph. Hubert, and C. Charlier, "Innovative development and validation of an HPLC/DAD method for the qualitative and quantitative determination of major cannabinoids in cannabis plant material" Journal of Chromatography. B : Analytical Technologies in the Biomedical & Life sciences, 2009, 877(32):4115–4124. http://hdl.handle.net/2268/4442

10. 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, 2009, 79:77–85. http://hdl.handle.net/2268/13236

11. 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", Chemometrics and Intelligent Laboratory system, 2008, 91:4–16. http://hdl.handle.net/2268/1640

Other publications

12. B. Debrus, J. Broséus, D. Guillarme, P. Lebrun, Ph. Hubert, J.-L. Veuthey, P. Esseiva, and S. Rudaz, "Innovative methodology to transfer conventional GC-MS heroin profiling to UHPLC-MS/MS", Analytical and Bioanalytical Chemistry, 2011, 399(8): 2719–2730. http://hdl.handle.net/2268/75667

13. E. Rozet, B. Govaerts, P. Lebrun, K. Michail, E. Ziemons, R. Wintersteiger, S. Rudaz, B. Boulanger, and Ph. Hubert, “Evaluating the reliability of analytical results using a probability criterion: a Bayesian perspective”, Analytica Chimica Acta, 2011, 705:193–206. `http://hdl.handle.net/2268/91261`

14. E. Ziemons, H. Bourichi, J. Mantanus, E. Rozet, P. Lebrun, E. Essassi, Y. Cherrah, A. Bouklouze, and Ph. Hubert, “Determination of binary polymorphic mixtures of fluconazole using near infrared spectroscopy and X-ray powder diffraction: A comparative study based on the pre-validation stage results”, Journal of Pharmaceutical & Biomedical Analysis, 2011, 55:1208–1212. `http://hdl.handle.net/2268/84986`

15. J. Mantanus, E. Ziemons, P. Lebrun, E. Rozet, R. Klinkenberg, B. Streel, B. Evrard, and Ph. Hubert, “Active content determination of non-coated pharmaceutical pellets by near infrared spectroscopy: Method development, validation and reliabilty evaluation”, Talanta, 2010, 80:1750–1757. `http://hdl.handle.net/2268/26616`

16. E. Ziemons, J. Mantanus, P. Lebrun, E. Rozet, B. Evrard, and Ph. Hubert, “Acetominophen determination in low-dose pharmaceutical syrup by NIR spectroscopy”, Journal of Pharmaceutical & Biomedical Analysis, 2010, 53:510–516. `http://hdl.handle.net/2268/60504`

17. J. Mantanus, E. Ziemons, P. Lebrun, E. Rozet, R. Klinkenberg, B. Streel, B. Evrard, and Ph. Hubert, “Moisture content determination of pharmaceutical pellets by near infrared spectroscopy: method development and validation”, Analytica Chimica Acta, 2009, 642(1-2):186-92. `http://hdl.handle.net/2268/18780`

Part III

Appendix

Appendix A

Use of informative prior with the multivariate linear regression

A.1 Description of the problem

The problem is a multivariate response surface model with m responses and k factors for each responses, using a common derivative matrix $\mathbf{X}$. The inference using non-informative priors distribution have been well discussed in Box and Tiao (1973), in Geisser and Cornfield (1963), and in Geisser (1965).

In this problem, n observations of the m-variate response vector are available. The multivariate model can then be written as

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E}, \tag{A.1}$$

where $\mathbf{Y}$ and $\mathbf{E}$ are $(n \times m)$ matrices for the m-variate responses and for the corresponding errors, respectively. $\mathbf{X}$ is the $(n \times p)$ design matrix ($p > k$, generally) and $\mathbf{B}$ is the $(p \times m)$ matrix containing the predictors of the model. One observation of the responses is described as $\mathbf{y}_i = (y_{i1}, \dots, y_{ij}, \dots, y_{im})$ and the expectation of its elements is

$$E(y_{ij}) = \sum_{l=1}^{p} x_{il}\beta_{lj}, \;\; i = 1, ..., n; \; j = 1, ..., m. \tag{A.2}$$

This equation can be written in matrix form for the m-variate response as

$$E(\mathbf{y}_i) = \mathbf{x}_i\mathbf{B} \tag{A.3}$$

with

$$\mathbf{x}_i = \left(x_{i1}, \ldots, x_{il}, \ldots, x_{ip}\right) \text{ and } \mathbf{B} = \begin{bmatrix} \beta_{11} & \cdots & \beta_{1j} & \cdots & \beta_{1m} \\ \vdots & \ddots & & & \vdots \\ \beta_{l1} & & \beta_{lj} & & \beta_{lm} \\ \vdots & & & \ddots & \vdots \\ \beta_{p1} & \cdots & \beta_{pl} & \cdots & \beta_{pm} \end{bmatrix}. \tag{A.4}$$

Thus, $\mathbf{x}_i$ is the i^{th} line of $\mathbf{X}$. Finally, it is assumed that each m-variate response follows a multivariate Normal distribution. Furthermore response vectors are independent and identically distributed (i.i.d.). That is,

$$\mathbf{y}_i \sim N_m\left(\mathbf{x}_i\mathbf{B}, \boldsymbol{\Sigma}\right), \; i = 1, ..., n. \tag{A.5}$$

A.2 Bayesian framework

Introduction. Bayesian methods are used to derive the posterior density of model parameters. Using Bayesian rule, the posterior density is the product of the likelihood and the prior densities for the parameters:

$$\begin{aligned} p(\mathbf{B}, \boldsymbol{\Sigma} \mid \text{data}) &\propto \mathcal{L}(\mathbf{B}, \boldsymbol{\Sigma} \mid \mathbf{Y}) \;.\; p(\mathbf{B}, \boldsymbol{\Sigma}) \\ &\propto \mathcal{L}(\mathbf{B}, \boldsymbol{\Sigma} \mid \mathbf{Y}) \;.\; p(\mathbf{B} \mid \boldsymbol{\Sigma}) \;.\; p(\boldsymbol{\Sigma}). \end{aligned} \tag{A.6}$$

Prior assumptions are the following: the density $p(\mathbf{B} \mid \boldsymbol{\Sigma})$ is assumed to be Normally distributed, and $p(\boldsymbol{\Sigma})$ follows an inverse-Wishart distribution. These are natural assumptions when inferring using proper informative prior distributions. Moreover, they are conjugate (Press, 1972, §8.6.2). Some results concerning these prior distributions have been derived in a note by Minka (2001). Alternatively, it is possible to use the Normal-Wishart prior distribution for the joint prior $p(\mathbf{B}, \boldsymbol{\Sigma}^{-1})$ as shown in Guttman (1970).

Likelihood. Following Equation A.5 and the i.i.d. assumptions on n observations, the likelihood for this model is given by:

$$\begin{aligned} \mathcal{L}\left(\mathbf{B}, \boldsymbol{\Sigma} \mid \mathbf{Y}\right) &= (2\pi)^{\frac{-mn}{2}} \left|\boldsymbol{\Sigma}\right|^{\frac{-n}{2}} . \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\left[\left(\mathbf{y}_i - \mathbf{x}_i\mathbf{B}\right)\boldsymbol{\Sigma}^{-1}\left(\mathbf{y}_i - \mathbf{x}_i\mathbf{B}\right)'\right]\right) \\ &\text{with} \quad -\infty < \varepsilon_{ij} = y_{ij} - \mathbf{x}_i\boldsymbol{\beta}_j < \infty, \quad i = 1, \ldots, n, \quad j = 1, \ldots, m. \end{aligned} \tag{A.7}$$

Or, more conveniently,

$$\mathcal{L}\left(\mathbf{B}, \boldsymbol{\Sigma} \mid \mathbf{Y}\right) = (2\pi)^{\frac{-mn}{2}} \left|\boldsymbol{\Sigma}\right|^{\frac{-n}{2}} . \exp\left(-\frac{1}{2}tr\left[\boldsymbol{\Sigma}^{-1}\left(\mathbf{Y} - \mathbf{X}\mathbf{B}\right)'\left(\mathbf{Y} - \mathbf{X}\mathbf{B}\right)\right]\right) \tag{A.8}$$

Prior distribution of $\mathbf{B} \mid \mathbf{\Sigma}$. Assume that the prior distribution of $\mathbf{B}$ given $\mathbf{\Sigma}$ is a matrix-variate Normal distribution of mean $\mathbf{B}_0$ and covariance matrices $\mathbf{\Sigma}$ (columns of $\mathbf{B}$) and $\mathbf{\Sigma}_0$ (rows of $\mathbf{B}$). The matrix-variate Normal distribution extends the classical multivariate Normal distribution to matrices of random values (instead of vectors).

$$\mathbf{B} \mid \mathbf{\Sigma} \sim N_{p \times m} \left(\mathbf{B}_0, \mathbf{\Sigma}, \mathbf{\Sigma}_0 \right). \tag{A.9}$$

$\mathbf{B}_0$ is then of the same size than $\mathbf{B}$, $\mathbf{\Sigma}$ is $(m \times m)$ and $\mathbf{\Sigma}_0$ is $(p \times p)$. By definition, the the matrix-variate Normal can be compute from a classical multivariate Normal using the following indentity (See Appendix D):

$$vec \left(\mathbf{B} \mid \mathbf{\Sigma} \right) \sim N_{pm} \left(vec \left(\mathbf{B}_0 \right), \mathbf{\Sigma} \otimes \mathbf{\Sigma}_0 \right), \tag{A.10}$$

where $\otimes$ is the Kronecker product. In this case, vectorized forms of $\mathbf{B}$ and $\mathbf{B}_0$ are used. This is rather useful for implementation purpose as the matrix-variate Normal is usually not available in softwares. The operator *vec*, applied on any matrix $\mathbf{B}$ of size $(p \times m)$, stacks its columns into a vector of length $p.m$:

$$vec(\mathbf{B}) = (\beta_{11}, \beta_{21}, \ldots, \beta_{p1}, \beta_{12}, \beta_{22}, \ldots, \ldots, \beta_{pm})'. \tag{A.11}$$

Notice that $\mathbf{\Sigma} \otimes \mathbf{\Sigma}_0$ is the $(pm \times pm)$ matrix defining the covariances between the pm parameters in $\mathbf{B}$. This covariance matrix is highly structured due to the Kronecker product. Note that, conditional to $\mathbf{\Sigma}$, the definition of the prior covariance matrix is only depending on $\mathbf{\Sigma}_0$. It follows that the prior information in $\mathbf{\Sigma}_0$ will be similar for each response.

The density of $\mathbf{B} \mid \mathbf{\Sigma}$ is:

$$\begin{aligned} p \left(\mathbf{B} \mid \mathbf{\Sigma} \right) = {} & (2\pi)^{\frac{-pm}{2}} \left| \mathbf{\Sigma}_0 \right|^{\frac{-m}{2}} \left| \mathbf{\Sigma} \right|^{\frac{-p}{2}} \\ & . \exp \left(-\frac{1}{2} tr \left[\mathbf{\Sigma}^{-1} (\mathbf{B} - \mathbf{B}_0)' \mathbf{\Sigma}_0^{-1} (\mathbf{B} - \mathbf{B}_0) \right] \right). \end{aligned} \tag{A.12}$$

Prior distribution of $\mathbf{\Sigma}$. $\mathbf{\Sigma}$ is the $(m \times m)$ unknown covariance matrix that describes the multivariate residual error between the m column of $\mathbf{E}$. $\mathbf{\Sigma}$ is assumed to follow an inverse-Wishart distribution W_1^{-1} with a scale matrix $\mathbf{\Omega}$ and ν_0 degrees of freedom (See Appendix D):

$$\mathbf{\Sigma} \sim W_1^{-1}(\mathbf{\Omega}, \nu_0), \;\; \nu_0 > 0. \tag{A.13}$$

The density of $\mathbf{\Sigma}$ is:

$$\begin{aligned} p(\mathbf{\Sigma}) = {} & 2^{\frac{m(\nu_0+m-1)}{2}} \, \Gamma_m \left(\frac{\nu_0 + m - 1}{2} \right)^{-1} \left| \mathbf{\Omega} \right|^{\frac{\nu_0+m-1}{2}} \left| \mathbf{\Sigma} \right|^{-\frac{\nu_0+2m}{2}} \\ & . \exp \left(-\frac{1}{2} tr[\mathbf{\Sigma}^{-1} \mathbf{\Omega}] \right), \end{aligned} \tag{A.14}$$

where $\Gamma_m(.)$ is the multivariate gamma function. This is the definition used in Box and Tiao (1973). ν_0 has to be understood as the number of degrees of freedom of the prior distribution. It could be defined as follows: $\nu_0 = n_0 - (m + p) + 1$, with n_0 being the number of prior observations.

Complete joint posterior distribution of the parameters. Combining the likelihood and the proposed prior distributions as in Equation (A.6), the posterior density of the parameters can be written, by regrouping first, normalizing constants; second, terms comprising $\mathbf{\Sigma}$ alone; and finally, terms with $\mathbf{\Sigma}$ and $\mathbf{B}$:

$$\begin{aligned} p\left(\mathbf{B},\mathbf{\Sigma}\mid \text{data}\right) = {} & (2\pi)^{\frac{-mn}{2}}(2\pi)^{\frac{-pm}{2}}\left|\mathbf{\Sigma}_0\right|^{\frac{-m}{2}}2^{\frac{m(\nu_0+m-1)}{2}}\ \Gamma_m(\tfrac{\nu_0+m-1}{2})^{-1}\left|\mathbf{\Omega}\right|^{\frac{\nu_0+m-1}{2}} \\ & .\left|\mathbf{\Sigma}\right|^{\frac{-n}{2}}\left|\mathbf{\Sigma}\right|^{\frac{-p}{2}}\left|\mathbf{\Sigma}\right|^{-\frac{\nu_0+2m}{2}}.\exp\left(-\frac{1}{2}tr[\mathbf{\Sigma}^{-1}\mathbf{\Omega}]\right) \\ & .\exp\left(-\frac{1}{2}tr\left[\mathbf{\Sigma}^{-1}\left(\mathbf{Y}-\mathbf{X}\mathbf{B}\right)'\left(\mathbf{Y}-\mathbf{X}\mathbf{B}\right)\right]\right) \\ & .\exp\left(-\frac{1}{2}tr\left[\mathbf{\Sigma}^{-1}(\mathbf{B}-\mathbf{B}_0)'\mathbf{\Sigma}_0^{-1}(\mathbf{B}-\mathbf{B}_0)\right]\right). \end{aligned} \tag{A.15}$$

Simplification of the posterior density. Equation (A.15) is unpractical as the regression parameters are present in both third and fourth lines. The idea is to regroup those two exponents in one single term. The last two lines can be recognized to be proportional to the posterior distribution of $\mathbf{B}$ conditional to $\mathbf{\Sigma}$. They can be combined using exponential and trace properties:

$$\begin{aligned} & p\left(\mathbf{B}\mid\mathbf{\Sigma},\text{data}\right)\propto \\ & \exp\left(-\frac{1}{2}tr\left[\mathbf{\Sigma}^{-1}\left(\left(\mathbf{Y}-\mathbf{X}\mathbf{B}\right)'\left(\mathbf{Y}-\mathbf{X}\mathbf{B}\right)+(\mathbf{B}-\mathbf{B}_0)'\mathbf{\Sigma}_0^{-1}(\mathbf{B}-\mathbf{B}_0)\right)\right]\right). \end{aligned} \tag{A.16}$$

The quadratic terms are developed and isolated as follows:

$$\begin{aligned} & p\left(\mathbf{B}\mid\mathbf{\Sigma},\text{data}\right) \\ & \propto \exp\left(-\frac{1}{2}tr\left[\mathbf{\Sigma}^{-1}\left(\mathbf{Y}'\mathbf{Y}-2\mathbf{B}'\mathbf{X}'\mathbf{Y}+\mathbf{B}'\mathbf{X}'\mathbf{X}\mathbf{B}+\mathbf{B}'\mathbf{\Sigma}_0^{-1}\mathbf{B}-2\mathbf{B}'\mathbf{\Sigma}_0^{-1}\mathbf{B}_0+\mathbf{B}_0'\mathbf{\Sigma}_0^{-1}\mathbf{B}_0\right)\right]\right) \\ & \propto \exp\left(-\frac{1}{2}tr\left[\mathbf{\Sigma}^{-1}\left(\mathbf{B}'(\mathbf{X}'\mathbf{X}+\mathbf{\Sigma}_0^{-1})\mathbf{B}-2\mathbf{B}'(\mathbf{X}'\mathbf{Y}+\mathbf{\Sigma}_0^{-1}\mathbf{B}_0)+(\mathbf{Y}'\mathbf{Y}+\mathbf{B}_0'\mathbf{\Sigma}_0^{-1}\mathbf{B}_0)\right)\right]\right). \end{aligned} \tag{A.17}$$

Noticing first that $\mathbf{X}'\mathbf{Y}=\mathbf{X}'\mathbf{X}\hat{\mathbf{B}}$, where $\hat{\mathbf{B}}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ is the OLS estimation of $\mathbf{B}$; and second, the *a posteriori* covariance matrix for the rows of $\mathbf{B}$ is $(\mathbf{X}'\mathbf{X}+\mathbf{\Sigma}_0^{-1})^{-1}$ and the covariance matrix for the columns of $\mathbf{B}$ is $\mathbf{\Sigma}$, thus the quadratic form in $\mathbf{B}$ need to be completed (see Equation (A.18)). In the next steps, it is shown how the terms that do not influence this quadratic form, conditional to $\mathbf{\Sigma}$, are isolated (see

Equations (A.20)):

$$\begin{aligned}
&p\,(\mathbf{B} \mid \boldsymbol{\Sigma}, \text{data}) \\
&\propto \exp\Bigg(-\frac{1}{2} tr\Big[\boldsymbol{\Sigma}^{-1}\Big(\mathbf{B}^{'}(\mathbf{X}^{'}\mathbf{X}+\boldsymbol{\Sigma}_0^{-1})\mathbf{B} \\
&\qquad -2\mathbf{B}^{'}(\mathbf{X}^{'}\mathbf{X}+\boldsymbol{\Sigma}_0^{-1})(\mathbf{X}^{'}\mathbf{X}+\boldsymbol{\Sigma}_0^{-1})^{-1}(\mathbf{X}^{'}\mathbf{X}\hat{\mathbf{B}}+\boldsymbol{\Sigma}_0^{-1}\mathbf{B}_0) \\
&\qquad +(\mathbf{X}^{'}\mathbf{X}\hat{\mathbf{B}}+\boldsymbol{\Sigma}_0^{-1}\mathbf{B}_0)^{'}(\mathbf{X}^{'}\mathbf{X}+\boldsymbol{\Sigma}_0^{-1})^{-1}(\mathbf{X}^{'}\mathbf{X}\hat{\mathbf{B}}+\boldsymbol{\Sigma}_0^{-1}\mathbf{B}_0)\Big)\Big]\Bigg) \\
&.\exp\Bigg(-\frac{1}{2} tr\Big[\boldsymbol{\Sigma}^{-1}\Big((\mathbf{Y}^{'}\mathbf{Y}+\mathbf{B}_0^{'}\boldsymbol{\Sigma}_0^{-1}\mathbf{B}_0) \\
&\qquad -(\mathbf{X}^{'}\mathbf{X}\hat{\mathbf{B}}+\boldsymbol{\Sigma}_0^{-1}\mathbf{B}_0)^{'}(\mathbf{X}^{'}\mathbf{X}+\boldsymbol{\Sigma}_0^{-1})^{-1}(\mathbf{X}^{'}\mathbf{X}\hat{\mathbf{B}}+\boldsymbol{\Sigma}_0^{-1}\mathbf{B}_0)\Big)\Big]\Bigg) \qquad \text{(A.18)} \\
&\propto \exp\Bigg(-\frac{1}{2} tr\Big[\boldsymbol{\Sigma}^{-1}(\mathbf{B}-\mathbf{M}_{\mathbf{B}\text{post}})^{'}(\mathbf{X}^{'}\mathbf{X}+\boldsymbol{\Sigma}_0^{-1})(\mathbf{B}-\mathbf{M}_{\mathbf{B}\text{post}})\Big]\Bigg) \qquad \text{(A.19)} \\
&.\exp\Bigg(-\frac{1}{2} tr\Big[\boldsymbol{\Sigma}^{-1}\mathbf{A}^{*}\Big]\Bigg), \qquad \text{(A.20)}
\end{aligned}$$

with

$$\begin{aligned}
&\mathbf{M}_{\mathbf{B}\text{post}} = (\mathbf{X}^{'}\mathbf{X}+\boldsymbol{\Sigma}_0^{-1})^{-1}(\mathbf{X}^{'}\mathbf{X}\hat{\mathbf{B}}+\boldsymbol{\Sigma}_0^{-1}\mathbf{B}_0), \quad \text{and,} \\
&\mathbf{A}^{*} = \mathbf{Y}^{'}\mathbf{Y}+\mathbf{B}_0^{'}\boldsymbol{\Sigma}_0^{-1}\mathbf{B}_0 - (\mathbf{X}^{'}\mathbf{X}\hat{\mathbf{B}}+\boldsymbol{\Sigma}_0^{-1}\mathbf{B}_0)^{'}(\mathbf{X}^{'}\mathbf{X}+\boldsymbol{\Sigma}_0^{-1})^{-1}(\mathbf{X}^{'}\mathbf{X}\hat{\mathbf{B}}+\boldsymbol{\Sigma}_0^{-1}\mathbf{B}_0).
\end{aligned}$$

$\mathbf{M}_{\mathbf{B}\text{post}}$ is a linear combination of the OLS estimation of $\mathbf{B}$ and its prior mean $\mathbf{B}_0$, weighted by their respective variances. The terms $\mathbf{A}^{*}$ is simply a constant. When using uniform prior distributions, $\mathbf{A}^{*}$ is equal to the $\mathbf{A}$ matrix presented by Geisser (1965) or Box and Tiao (1973) (See later). $\mathbf{A}^{*}$ is also the sum of squares of the multivariate residual error, though it is not easy to visualize it due to the prior incorporation.

Naturally, given $\boldsymbol{\Sigma}$, Equation (A.20) is also a constant, and the density of Equation A.19 alone is proportional to $p\,(\mathbf{B} \mid \boldsymbol{\Sigma}, \text{data})$. It can readily be identified as a matrix-variate Normal distribution with mean $\mathbf{M}_{\mathbf{B}\text{post}}$ and covariance matrices $\boldsymbol{\Sigma}$ (for the columns) and $(\mathbf{X}^{'}\mathbf{X}+\boldsymbol{\Sigma}_0^{-1})^{-1}$ (for the rows):

$$\mathbf{B} \mid \boldsymbol{\Sigma}, \text{data} \sim N_{p\times m}\Big(\mathbf{M}_{\mathbf{B}\text{post}}, \boldsymbol{\Sigma}, \Big(\mathbf{X}^{'}\mathbf{X}+\boldsymbol{\Sigma}_0^{-1}\Big)^{-1}\Big) \qquad \text{(A.21)}$$

Or, more conveniently,

$$vec(\mathbf{B}) \mid \boldsymbol{\Sigma}, \text{data} \sim N_{pm}\Big(vec(\mathbf{M}_{\mathbf{B}\text{post}}), \boldsymbol{\Sigma} \otimes \Big(\mathbf{X}^{'}\mathbf{X}+\boldsymbol{\Sigma}_0^{-1}\Big)^{-1}\Big) \qquad \text{(A.22)}$$

Notice that the posterior row covariance is simply the (inverse of the) sum of the likelihood precision matrix ($\mathbf{X}'\mathbf{X}$) and of the prior row precision matrix ($\mathbf{\Sigma}_0^{-1}$).

Now that the two last lines of Equation (A.15) have been simplified, the joint posterior distribution can be restated as:

$$\begin{aligned} p\Big(\mathbf{B}, \mathbf{\Sigma} \mid \text{data}\Big) = \ & (2\pi)^{\frac{-mn}{2}} (2\pi)^{\frac{-pm}{2}} \left|\mathbf{\Sigma}_0\right|^{\frac{-m}{2}} 2^{\frac{m(\nu_0+m-1)}{2}} \ \Gamma_m(\tfrac{\nu_0+m-1}{2})^{-1} \left|\mathbf{\Omega}\right|^{\frac{\nu_0+m-1}{2}} \\ & . \left|\mathbf{\Sigma}\right|^{\frac{-n}{2}} \left|\mathbf{\Sigma}\right|^{\frac{-p}{2}} \left|\mathbf{\Sigma}\right|^{-\frac{\nu_0+2m}{2}} \\ & . \exp\left(-\frac{1}{2} tr\Big[\mathbf{\Sigma}^{-1}\mathbf{\Omega}\Big]\right) \exp\left(-\frac{1}{2} tr\Big[\mathbf{\Sigma}^{-1}\mathbf{A}^*\Big]\right) \\ & . \exp\left(-\frac{1}{2} tr\Big[\mathbf{\Sigma}^{-1}(\mathbf{B} - \mathbf{M}_{\mathbf{B}\text{post}})'(\mathbf{X}'\mathbf{X} + \mathbf{\Sigma}_0^{-1})(\mathbf{B} - \mathbf{M}_{\mathbf{B}\text{post}})\Big]\right). \end{aligned} \qquad \text{(A.23)}$$

Marginal posterior density of $\mathbf{\Sigma}$. The posterior distribution of $\mathbf{\Sigma}$ is computed as

$$p\left(\mathbf{\Sigma} \mid \text{data}\right) \propto \int_{\mathbf{B}} p\left(\mathbf{B}, \mathbf{\Sigma} \mid \text{data}\right) d\mathbf{B}. \qquad \text{(A.24)}$$

This integral is fortunately simple to resolve as only the last line of Equation (A.23) has terms including $\mathbf{B}$. All the other terms are then putted out of the integral. Noticing that, as in Geisser (1965) (Equation 4.6), the terms in $\mathbf{B}$ can be integrated out as:

$$\begin{aligned} & \int_{\mathbf{B}} \exp\left(-\frac{1}{2} tr\left[\mathbf{\Sigma}^{-1}\Big(\mathbf{B} - \mathbf{M}_{\mathbf{B}\text{post}}\Big)\left(\mathbf{X}'\mathbf{X} + \mathbf{\Sigma}_0^{-1}\right)\Big(\mathbf{B} - \mathbf{M}_{\mathbf{B}\text{post}}\Big)\right]\right) d\mathbf{B} \\ & \quad = (2\pi)^{\frac{pm}{2}} \ \ \left|\mathbf{\Sigma}\right|^{\frac{p}{2}} \ \ \left|\mathbf{X}'\mathbf{X} + \mathbf{\Sigma}_0^{-1}\right|^{\frac{-m}{2}}, \end{aligned} \qquad \text{(A.25)}$$

the marginal posterior density of $\mathbf{\Sigma}$ can be written as

$$\begin{aligned} p\left(\mathbf{\Sigma} \mid \text{data}\right) \propto & \left|\mathbf{\Sigma}\right|^{-\frac{\nu_0+n+p+2m}{2}} \exp\left(-\frac{1}{2} tr\Big[\mathbf{\Sigma}^{-1}(\mathbf{\Omega} + \mathbf{A}^*)\Big]\right) \\ & . \int_{\mathbf{B}} \exp\left(-\frac{1}{2} tr\left[\mathbf{\Sigma}^{-1}\Big(\mathbf{B} - \mathbf{M}_{\mathbf{B}\text{post}}\Big)\left(\mathbf{X}'\mathbf{X} + \mathbf{\Sigma}_0^{-1}\right)\Big(\mathbf{B} - \mathbf{M}_{\mathbf{B}\text{post}}\Big)\right]\right) d\mathbf{B} \\ \propto & \left|\mathbf{\Sigma}\right|^{-\frac{\nu_0+n+\cancel{p}+2m}{2}} \exp\left(-\frac{1}{2} tr\Big[\mathbf{\Sigma}^{-1}(\mathbf{\Omega} + \mathbf{A}^*)\Big]\right) \\ & .(\sqrt{2\pi})^{mp} \ \ \cancel{\left|\mathbf{\Sigma}\right|^{\frac{p}{2}}} \ \ \left|\mathbf{X}'\mathbf{X} + \mathbf{\Sigma}_0^{-1}\right|^{\frac{-m}{2}} \\ \propto & \left|\mathbf{\Sigma}\right|^{-\frac{(\nu_0+n)+2m}{2}} \exp\left(-\frac{1}{2} tr\Big[\mathbf{\Sigma}^{-1}(\mathbf{\Omega} + \mathbf{A}^*)\Big]\right) \end{aligned} \qquad \text{(A.26)}$$

This density can then be identified as the inverse-Wishart $W_1^{-1}(\mathbf{\Omega} + \mathbf{A}^*, \nu_0 + n)$. Note that if $\nu = n - (m+p) + 1$ and $\nu_0 = n_0 - (m+p) + 1$, then $\nu_0 + n = \nu + n_0$, and

$$\mathbf{\Sigma} \mid \text{data} \sim W_1^{-1}\left(\mathbf{\Omega} + \mathbf{A}^*, \nu + n_0\right). \qquad \text{(A.27)}$$

The scale matrix is then the sum of the prior scale matrix $\mathbf{\Omega}$ and $\mathbf{A}^*$. There is $\nu + n_0$ degrees of freedom (d.f.), i.e., the number of d.f. from the likelihood (ν) added by the number of virtual observations n_0 injected from the informative prior distribution of $\mathbf{\Sigma}$.

Here follows some results derived from the properties of the Wishart and inverse-Wishart distribution. First, the distribution of $\mathbf{\Sigma}^{-1}$ is a classical Wishart distribution (Box and Tiao, 1973) :

$$\mathbf{\Sigma}^{-1} \mid \text{data} \sim W_1\left((\mathbf{\Omega} + \mathbf{A}^*)^{-1}, \nu + n_0\right). \tag{A.28}$$

This is straightforward as the only difference between the inverse-Wishart and the Wishart distributions is the transformation from $\mathbf{\Sigma}$ to $\mathbf{\Sigma}^{-1}$, whose Jacobian is $|\mathbf{\Sigma}|^{m+1}$. This equivalence may be helpful for several purposes, such as the sampling using computer softwares, or the identification of the predictive distribution of the responses.

Second, great care should be taken when sampling or using statistics derived from other implementations of the Wishart or inverse-Wishart distributions that exist (referred as W_2^1 and W_2^{-1} respectively). Indeed, the distributions included in softwares such as R or WinBUGS may have a different definition of the d.f. than the ones proposed here. As discussed in Appendix D, if W_1^1 and W_1^{-1} both have ν_1 d.f. ($\nu_1 > 0$), then, W_2^1 and W_2^{-1} must have ν_2 d.f. ($\nu_2 > m - 1$) with the following relation: $\nu_1 + M - 1 = \nu_2$.

Marginal posterior density of B. The posterior distribution of $\mathbf{B}$ is computed as

$$p\left(\mathbf{B} \mid \text{data}\right) \propto \int_{\mathbf{\Sigma}} p\left(\mathbf{B}, \mathbf{\Sigma} \mid \text{data}\right) d\mathbf{\Sigma}. \tag{A.29}$$

As in the previous section, this integral is not hard to resolve if the terms in $\mathbf{\Sigma}$ are on the "right place". From Dickey (1967), the following relation is available:

$$\int_{\mathbf{U}} |\mathbf{U}|^{\lambda - \frac{m+1}{2}} . \exp\left(tr\Big[-\mathbf{UM}\Big]\right) d\mathbf{U} = \Gamma_m(\lambda)\, |\mathbf{M}|^{-\lambda}, \tag{A.30}$$

where $\mathbf{U}$ and $\mathbf{M}$ are symmetric and definite positive ($m \times m$) matrices and $\lambda > \frac{1}{2}(m-1)$. The idea is to transform the joint posterior density so that this relationship can be applied.

Thus, the joint posterior of Equation A.23 is simplified using trace and exponential properties, and only the relevant terms are selected:

$$\begin{aligned}
&p\Big(\mathbf{B}, \mathbf{\Sigma} \mid \text{data}\Big) \\
&\quad\propto |\mathbf{\Sigma}|^{-\frac{\nu_0+n+p+2m}{2}} . \exp\left(-\frac{1}{2} tr\Big[\mathbf{\Sigma}^{-1}\mathbf{\Omega}\Big]\right) . \exp\left(-\frac{1}{2} tr\Big[\mathbf{\Sigma}^{-1}\mathbf{A}^{*}\Big]\right) \\
&\quad . \exp\left(-\frac{1}{2} tr\Big[\mathbf{\Sigma}^{-1}(\mathbf{B}-\mathbf{M}_{\mathbf{B}\text{post}})'(\mathbf{X}'\mathbf{X}+\mathbf{\Sigma}_0^{-1})(\mathbf{B}-\mathbf{M}_{\mathbf{B}\text{post}})\Big]\right) \\
&\quad\propto |\mathbf{\Sigma}|^{-\frac{\nu_0+n+p+2m}{2}} \\
&\quad . \exp\left(-\frac{1}{2} tr\Big[\mathbf{\Sigma}^{-1}\Big((\mathbf{\Omega}+\mathbf{A}^{*}) + (\mathbf{B}-\mathbf{M}_{\mathbf{B}\text{post}})'(\mathbf{X}'\mathbf{X}+\mathbf{\Sigma}_0^{-1})(\mathbf{B}-\mathbf{M}_{\mathbf{B}\text{post}})\Big)\Big]\right)
\end{aligned} \tag{A.31}$$

To use Equation (A.30), the posterior must be restated. Using the Jacobian of the transformation of $\mathbf{\Sigma}$ to $\mathbf{\Sigma}^{-1}$, equal to $|\mathbf{\Sigma}|^{m+1}$ (Box and Tiao, 1973), the posterior is:

$$\begin{aligned}
&p\Big(\mathbf{B},\, \mathbf{\Sigma}^{-1} \mid \text{data}\Big) \\
&\quad\propto \Big|\mathbf{\Sigma}^{-1}\Big|^{\frac{\nu_0+n+p+m-1}{2}-\frac{m+1}{2}} \\
&\quad . \exp\left(tr\Big[-\mathbf{\Sigma}^{-1}\frac{(\mathbf{\Omega}+\mathbf{A}^{*}) + (\mathbf{B}-\mathbf{M}_{\mathbf{B}\text{post}})'(\mathbf{X}'\mathbf{X}+\mathbf{\Sigma}_0^{-1})(\mathbf{B}-\mathbf{M}_{\mathbf{B}\text{post}})}{2}\Big]\right).
\end{aligned} \tag{A.32}$$

With this arrangement, the posterior density now has a form compatible with the formula of Dickey in Equation (A.30). Assuming that $\mathbf{U} = \mathbf{\Sigma}^{-1}$, $p(\mathbf{B}, \mathbf{\Sigma}^{-1} \mid \text{data})$ is integrated with respect to $\mathbf{\Sigma}^{-1}$, and gives:

$$\begin{aligned}
p\left(\mathbf{B} \mid \text{data}\right) = &\frac{1}{2}\Gamma_m\big(\tfrac{(\nu_0+n)+p+m-1}{2}\big) \\
&. \Big|(\mathbf{\Omega}+\mathbf{A}^{*}) + (\mathbf{B}-\mathbf{M}_{\mathbf{B}\text{post}})'(\mathbf{X}'\mathbf{X}+\mathbf{\Sigma}_0^{-1})(\mathbf{B}-\mathbf{M}_{\mathbf{B}\text{post}})\Big|^{-\frac{(\nu_0+n)+p+m-1}{2}}
\end{aligned} \tag{A.33}$$

This last density is identified as the following $(p \times m)$ matrix-variate Student's distribution:

$$\mathbf{B} \mid \text{data} \sim T_{p\times m}(\mathbf{M}_{\mathbf{B}\text{post}}, \mathbf{\Omega}+\mathbf{A}^{*}, (\mathbf{X}'\mathbf{X}+\mathbf{\Sigma}_0^{-1})^{-1}, \nu_0+n). \tag{A.34}$$

Notice that the use of the Jacobian is similar to assume a Wishart distribution for $\mathbf{\Sigma}^{-1}$ instead of an inverse-Wishart distribution for $\mathbf{\Sigma}$.

A.3 Using non-informative prior distributions

Without being a justification of the previous calculation, it is interesting to see if the use of uniform priors leads to the same solutions than the classical non-informative distributions (Box and Tiao, 1973).

For $\mathbf{B} \mid \mathbf{\Sigma}$, a vague prior is given by any values of $\mathbf{B}_0$ (e.g. a $\mathbf{0}-$matrix), and a precision matrix $\mathbf{\Sigma}_0^{-1}$ being non significant with 0 everywhere. Clearly, the posterior variance for $vec(\mathbf{B})$ will be

$$\mathbf{\Sigma} \otimes \left(\mathbf{X}'\mathbf{X} + \mathbf{\Sigma}_0^{-1}\right)^{-1} = \mathbf{\Sigma} \otimes \left(\mathbf{X}'\mathbf{X}\right)^{-1} \tag{A.35}$$

which is the solution proposed by Box and Tiao. The simplification of the posterior mean in a non-informative case is as follows:

$$\begin{aligned}\mathbf{M}_{\mathbf{B}\text{post}} = \left(\mathbf{X}'\mathbf{X} + \mathbf{\Sigma}_0^{-1}\right)^{-1}\left(\mathbf{X}'\mathbf{X}\hat{\mathbf{B}} + \mathbf{\Sigma}_0^{-1}\mathbf{B}_0\right) &= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{X}\hat{\mathbf{B}}\right) \\ &= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{X}\right)\hat{\mathbf{B}} \\ &= \hat{\mathbf{B}},\end{aligned} \tag{A.36}$$

which is the ordinary least-squares estimation of $\mathbf{B}$.

Similarly, the scale parameter of the posterior distribution of $\mathbf{\Sigma}$ is similar to the one proposed in the non-informative case in the literature. Let $n_0 = 0$ and $\mathbf{\Omega}$ be 0 everywhere, the posterior distribution becomes:

$$\mathbf{\Sigma} \sim W_1^{-1}\left(\mathbf{\Omega} + \mathbf{A}^*, \nu + n_0\right) = W_1^{-1}\left(\mathbf{A}^*, \nu\right) \tag{A.37}$$

Below are listed the details to simplify $\mathbf{A}^*$ in the non-informative case:

$$\begin{aligned}\mathbf{A}^* &= \left(\mathbf{Y}'\mathbf{Y} + \mathbf{B}_0'\mathbf{\Sigma}_0^{-1}\mathbf{B}_0\right) - \left(\mathbf{X}'\mathbf{X}\hat{\mathbf{B}} + \mathbf{\Sigma}_0^{-1}\mathbf{B}_0\right)'\left(\mathbf{X}'\mathbf{X} + \mathbf{\Sigma}_0^{-1}\right)^{-1}\left(\mathbf{X}'\mathbf{X}\hat{\mathbf{B}} + \mathbf{\Sigma}_0^{-1}\mathbf{B}_0\right) \\ &= \left(\mathbf{Y}'\mathbf{Y}\right) - \left(\mathbf{X}'\mathbf{X}\hat{\mathbf{B}}\right)'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{X}\hat{\mathbf{B}}\right) \\ &= \left(\mathbf{Y}'\mathbf{Y}\right) - \left(\hat{\mathbf{B}}'\mathbf{X}'\mathbf{X}\hat{\mathbf{B}}\right) \\ &= \left(\mathbf{Y}'\mathbf{Y}\right) - 2\left(\hat{\mathbf{B}}'\mathbf{X}'\mathbf{X}\hat{\mathbf{B}}\right) + \left(\hat{\mathbf{B}}'\mathbf{X}'\mathbf{X}\hat{\mathbf{B}}\right) \\ &= \left(\mathbf{Y}'\mathbf{Y}\right) - 2\left(\hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y}\right) + \left(\hat{\mathbf{B}}'\mathbf{X}'\mathbf{X}\hat{\mathbf{B}}\right) \\ &= \left(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}\right)'\left(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}\right) \\ &= \mathbf{A},\end{aligned} \tag{A.38}$$

to compare with $\mathbf{A}$ in Geisser (1965) and Box and Tiao (1973), i.e., the sum of squares of the multivariate residual error. The simplification of the marginal posterior distribution of $\mathbf{B}$ under non-informative prior also leads to the same Student's distribution as in Box and Tiao (1973).

A.4 Simultaneous Predictions

Known $\mathbf{\Sigma}$. Assume that $\tilde{n}$ vectors of m responses create a $(\tilde{n} \times m)$ matrix $\tilde{\mathbf{Y}}$. These vectors are predicted simultaneously at different operating conditions $\tilde{\mathbf{X}} = (\tilde{\mathbf{x}}_1, ..., \tilde{\mathbf{x}}_{\tilde{n}})'$. Mean responses are obviously obtained using the regression model $E(\tilde{\mathbf{Y}} \mid \tilde{\mathbf{X}}, \text{data}) = \tilde{\mathbf{X}}\mathbf{M}_{\mathbf{B}\text{post}}$, where $\tilde{\mathbf{X}}$ is a $(\tilde{n} \times p)$ containing the design vector effects. Each line of $\tilde{\mathbf{X}}$ may be estimated in the experimental domain. The interest is to obtain the distribution of several independent new response vectors. To keep the text simple, the first results are presented conditionally to $\mathbf{\Sigma}$. All constant terms (terms that does not depend upon $\mathbf{B}$) can then be dropped.

$$p(\tilde{\mathbf{Y}} \mid \tilde{\mathbf{X}}, \mathbf{\Sigma}, \text{data}) = \int_{\mathbf{B}} p(\tilde{\mathbf{Y}} \mid \tilde{\mathbf{X}}, \mathbf{B}, \mathbf{\Sigma}).p(\mathbf{B} \mid \mathbf{\Sigma}, \text{data})d\mathbf{B} \tag{A.39}$$

$$\propto \int_{\mathbf{B}} |\mathbf{\Sigma}|^{\frac{-\tilde{n}}{2}} . \exp\left(-\frac{1}{2}tr\left[\mathbf{\Sigma}^{-1}(\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\mathbf{B})(\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\mathbf{B})'\right]\right)$$
$$. \exp\left(-\frac{1}{2}tr\left[\mathbf{\Sigma}^{-1}(\mathbf{B} - \mathbf{M}_{\mathbf{B}\text{post}})'(\mathbf{X}'\mathbf{X} + \mathbf{\Sigma}_0^{-1})(\mathbf{B} - \mathbf{M}_{\mathbf{B}\text{post}})\right]\right) d\mathbf{B}. \tag{A.40}$$

Grouping the two exponential within the integral (using trace and exponential properties) gives:

$$\begin{aligned} p(\tilde{\mathbf{Y}} \mid \tilde{\mathbf{X}}, \mathbf{\Sigma}, \text{data}) \propto \int_{\mathbf{B}} \exp\Bigg(- \frac{1}{2}tr\Big[\mathbf{\Sigma}^{-1}\Big((\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\mathbf{B})(\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\mathbf{B})' \\ + (\mathbf{B} - \mathbf{M}_{\mathbf{B}\text{post}})'(\mathbf{X}'\mathbf{X} + \mathbf{\Sigma}_0^{-1})(\mathbf{B} - \mathbf{M}_{\mathbf{B}\text{post}})\Big)\Big]\Bigg)d\mathbf{B} \\ \propto \int_{\mathbf{B}} \exp\left(-\frac{1}{2}tr\Big[\mathbf{\Sigma}^{-1}\Big(\mathbf{Q}\Big)\Big]\right) d\mathbf{B}. \end{aligned} \tag{A.41}$$

The quadratic form $\mathbf{Q}$ is developed hereafter,

$$\begin{aligned} \mathbf{Q} &= \tilde{\mathbf{Y}}'\tilde{\mathbf{Y}} - 2\mathbf{B}'\tilde{\mathbf{X}}'\tilde{\mathbf{Y}} + \mathbf{B}'\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\mathbf{B} \\ &+ \mathbf{B}'(\mathbf{X}'\mathbf{X} + \mathbf{\Sigma}_0^{-1})\mathbf{B} - 2\mathbf{B}'(\mathbf{X}'\mathbf{X} + \mathbf{\Sigma}_0^{-1})\mathbf{M}_{\mathbf{B}\text{post}} + \mathbf{M}_{\mathbf{B}\text{post}}'(\mathbf{X}'\mathbf{X} + \mathbf{\Sigma}_0^{-1})\mathbf{M}_{\mathbf{B}\text{post}}. \end{aligned} \tag{A.42}$$

To solve the integral in $\mathbf{B}$, it is enviable to group the terms that include $\mathbf{B}$ to simplify the equation:

$$\begin{aligned} \mathbf{Q} = \mathbf{B}'\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}} + (\mathbf{X}'\mathbf{X} + \mathbf{\Sigma}_0^{-1})\right)\mathbf{B} - 2\mathbf{B}'\left[(\mathbf{X}'\mathbf{X} + \mathbf{\Sigma}_0^{-1})\mathbf{M}_{\mathbf{B}\text{post}} + \tilde{\mathbf{X}}'\tilde{\mathbf{Y}}\right] \\ + \tilde{\mathbf{Y}}'\tilde{\mathbf{Y}} + \mathbf{M}_{\mathbf{B}\text{post}}'(\mathbf{X}'\mathbf{X} + \mathbf{\Sigma}_0^{-1})\mathbf{M}_{\mathbf{B}\text{post}}. \end{aligned} \tag{A.43}$$

Analyzing the first line of Equation A.43 a quadratic form in $\mathbf{B}$ might be found. The variance term would be the inverse of $(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}+(\mathbf{X}'\mathbf{X}+\mathbf{\Sigma}_0^{-1}))$. Let $\mathbf{V} = (\tilde{\mathbf{X}}'\tilde{\mathbf{X}}+(\mathbf{X}'\mathbf{X}+\mathbf{\Sigma}_0^{-1}))$ to shorten equations. Pre-multiplicating the crossed term (i.e. $-2\mathbf{B}'[...]$) by

$\mathbf{I} = \mathbf{V}\mathbf{V}^{-1}$ and completing the square form gives

$$\begin{aligned}\mathbf{Q} = \mathbf{B}^{'}\mathbf{V}\mathbf{B} - 2\mathbf{B}^{'}(\mathbf{V}\mathbf{V}^{-1})&\left[(\mathbf{X}^{'}\mathbf{X} + \mathbf{\Sigma}_0^{-1})\mathbf{M}_{\mathbf{B}_{\text{post}}} + \tilde{\mathbf{X}}^{'}\tilde{\mathbf{Y}}\right] \\ &+ \left[(\mathbf{X}^{'}\mathbf{X} + \mathbf{\Sigma}_0^{-1})\mathbf{M}_{\mathbf{B}_{\text{post}}} + \tilde{\mathbf{X}}^{'}\tilde{\mathbf{Y}}\right]^{'}\mathbf{V}^{-1}\mathbf{V}\mathbf{V}^{-1}\left[(\mathbf{X}^{'}\mathbf{X} + \mathbf{\Sigma}_0^{-1})\mathbf{M}_{\mathbf{B}_{\text{post}}} + \tilde{\mathbf{X}}^{'}\tilde{\mathbf{Y}}\right] \\ &+ \tilde{\mathbf{Y}}^{'}\tilde{\mathbf{Y}} + \mathbf{M}^{'}_{\mathbf{B}_{\text{post}}}(\mathbf{X}^{'}\mathbf{X} + \mathbf{\Sigma}_0^{-1})\mathbf{M}_{\mathbf{B}_{\text{post}}} \\ &- \left[(\mathbf{X}^{'}\mathbf{X} + \mathbf{\Sigma}_0^{-1})\mathbf{M}_{\mathbf{B}_{\text{post}}} + \tilde{\mathbf{X}}^{'}\tilde{\mathbf{Y}}\right]^{'}\mathbf{V}^{-1}\mathbf{V}\mathbf{V}^{-1}\left[(\mathbf{X}^{'}\mathbf{X} + \mathbf{\Sigma}_0^{-1})\mathbf{M}_{\mathbf{B}_{\text{post}}} + \tilde{\mathbf{X}}^{'}\tilde{\mathbf{Y}}\right].\end{aligned} \tag{A.44}$$

Although this is barely visible, the three first terms of $\mathbf{Q}$ are now a quadratic form in $\mathbf{B}$. Letting $\mathbf{L} = \left[(\mathbf{X}^{'}\mathbf{X} + \mathbf{\Sigma}_0^{-1})\mathbf{M}_{\mathbf{B}_{\text{post}}} + \tilde{\mathbf{X}}^{'}\tilde{\mathbf{Y}}\right]$, $\mathbf{Q}$ is simplified as follows:

$$\begin{aligned}\mathbf{Q} = &\left(\mathbf{B} - \mathbf{V}^{-1}\mathbf{L}\right)^{'}\mathbf{V}\left(\mathbf{B} - \mathbf{V}^{-1}\mathbf{L}\right) \\ &+ \tilde{\mathbf{Y}}^{'}\tilde{\mathbf{Y}} + \mathbf{M}^{'}_{\mathbf{B}_{\text{post}}}(\mathbf{X}^{'}\mathbf{X} + \mathbf{\Sigma}_0^{-1})\mathbf{M}_{\mathbf{B}_{\text{post}}} - \mathbf{L}^{'}\mathbf{V}^{-1}\mathbf{L}.\end{aligned} \tag{A.45}$$

Thus, the terms in $\mathbf{B}$ have been isolated. Returning back to Equation A.41, $\mathbf{Q}$ and $\mathbf{L}$ are kept as in Equation A.45. The constant values are putted out from the integral using trace and exponential properties:

$$\begin{aligned}p(\tilde{\mathbf{Y}} \mid \tilde{\mathbf{X}}, \mathbf{\Sigma}, \text{data}) &\propto \int_{\mathbf{B}} \exp\left(-\frac{1}{2}tr\left[\mathbf{\Sigma}^{-1}\left(\mathbf{Q}\right)\right]\right)d\mathbf{B} \\ &\propto \int_{\mathbf{B}} \exp\left(-\frac{1}{2}tr\left[\mathbf{\Sigma}^{-1}\left((\mathbf{B} - \mathbf{V}^{-1}\mathbf{L})^{'}\mathbf{V}(\mathbf{B} - \mathbf{V}^{-1}\mathbf{L})\right.\right.\right. \\ &\qquad\left.\left.\left. + \tilde{\mathbf{Y}}^{'}\tilde{\mathbf{Y}} + \mathbf{M}^{'}_{\mathbf{B}_{\text{post}}}(\mathbf{X}^{'}\mathbf{X} + \mathbf{\Sigma}_0^{-1})\mathbf{M}_{\mathbf{B}_{\text{post}}} - \mathbf{L}^{'}\mathbf{V}^{-1}\mathbf{L}\right)\right]\right)d\mathbf{B} \\ &\propto \int_{\mathbf{B}} \exp\left(-\frac{1}{2}tr\left[\mathbf{\Sigma}^{-1}\left((\mathbf{B} - \mathbf{V}^{-1}\mathbf{L})^{'}\mathbf{V}(\mathbf{B} - \mathbf{V}^{-1}\mathbf{L})\right)\right]\right).d\mathbf{B} \\ &.\exp\left(-\frac{1}{2}tr\left[\mathbf{\Sigma}^{-1}\left(\tilde{\mathbf{Y}}^{'}\tilde{\mathbf{Y}} + \mathbf{M}^{'}_{\mathbf{B}_{\text{post}}}(\mathbf{X}^{'}\mathbf{X} + \mathbf{\Sigma}_0^{-1})\mathbf{M}_{\mathbf{B}_{\text{post}}} - \mathbf{L}^{'}\mathbf{V}^{-1}\mathbf{L}\right)\right]\right) \\ &\propto (2\pi)^{\frac{pm}{2}}\left|\mathbf{\Sigma}\right|^{\frac{p}{2}}\left|\mathbf{V}\right|^{\frac{m}{2}} \\ &.\exp\left(-\frac{1}{2}tr\left[\mathbf{\Sigma}^{-1}\left(\tilde{\mathbf{Y}}^{'}\tilde{\mathbf{Y}} + \mathbf{M}^{'}_{\mathbf{B}_{\text{post}}}(\mathbf{X}^{'}\mathbf{X} + \mathbf{\Sigma}_0^{-1})\mathbf{M}_{\mathbf{B}_{\text{post}}} - \mathbf{L}^{'}\mathbf{V}^{-1}\mathbf{L}\right)\right]\right).\end{aligned} \tag{A.46}$$

The integration of $\mathbf{B}$ has been done as in Geisser (1965), Equation 4.6. These constants are dropped for the moments.

Next, the objective is to simplify and identify the form of the density for $\tilde{\mathbf{Y}}$. Let work on $\mathbf{Q}^*$, defined as:

$$\begin{aligned} p(\tilde{\mathbf{Y}} \mid \tilde{\mathbf{X}}, \boldsymbol{\Sigma}, \text{data}) &\propto \exp\Bigg(-\frac{1}{2} tr\Big[\boldsymbol{\Sigma}^{-1}\Big(\tilde{\mathbf{Y}}'\tilde{\mathbf{Y}} + \mathbf{M}'_{\mathbf{B}_{\text{post}}}(\mathbf{X}'\mathbf{X} + \boldsymbol{\Sigma}_0^{-1})\mathbf{M}_{\mathbf{B}_{\text{post}}} \\ &\qquad\qquad - \mathbf{L}'\mathbf{V}^{-1}\mathbf{L}\Big)\Big]\Bigg) \\ &\propto \exp\Bigg(-\frac{1}{2} tr\Big[\boldsymbol{\Sigma}^{-1}\Big(\mathbf{Q}^*\Big)\Big]\Bigg), \end{aligned} \tag{A.47}$$

A quadratic form in $\tilde{\mathbf{Y}}$ would look similar to $\mathbf{Q}^* = (\tilde{\mathbf{Y}} - \boldsymbol{\mu}_{\mathbf{y}})'\mathbf{C}(\tilde{\mathbf{Y}} - \boldsymbol{\mu}_{\mathbf{y}}) = \tilde{\mathbf{Y}}'\mathbf{C}\tilde{\mathbf{Y}} - 2\boldsymbol{\mu}'_{\mathbf{y}}\mathbf{C}\tilde{\mathbf{Y}} + \boldsymbol{\mu}'_{\mathbf{y}}\mathbf{C}\boldsymbol{\mu}_{\mathbf{y}}$. Basically, the idea is, again, to develop quadratic forms, to isolate relevant terms, and to identify variance and mean terms by completing the quadratic form with the necessary values. First, developing $\mathbf{L}'\mathbf{V}^{-1}\mathbf{L}$ gives:

$$\begin{aligned} \mathbf{Q}^* &= \tilde{\mathbf{Y}}'\tilde{\mathbf{Y}} + \mathbf{M}'_{\mathbf{B}_{\text{post}}}(\mathbf{X}'\mathbf{X} + \boldsymbol{\Sigma}_0^{-1})\mathbf{M}_{\mathbf{B}_{\text{post}}} \\ &\qquad - \Big[(\mathbf{X}'\mathbf{X} + \boldsymbol{\Sigma}_0^{-1})\mathbf{M}_{\mathbf{B}_{\text{post}}} + \tilde{\mathbf{X}}'\tilde{\mathbf{Y}}\Big]' \mathbf{V}^{-1} \Big[(\mathbf{X}'\mathbf{X} + \boldsymbol{\Sigma}_0^{-1})\mathbf{M}_{\mathbf{B}_{\text{post}}} + \tilde{\mathbf{X}}'\tilde{\mathbf{Y}}\Big] \\ &= \tilde{\mathbf{Y}}'\tilde{\mathbf{Y}} + \mathbf{M}'_{\mathbf{B}_{\text{post}}}(\mathbf{X}'\mathbf{X} + \boldsymbol{\Sigma}_0^{-1})\mathbf{M}_{\mathbf{B}_{\text{post}}} \\ &\qquad - \Big[\mathbf{M}'_{\mathbf{B}_{\text{post}}}(\mathbf{X}'\mathbf{X} + \boldsymbol{\Sigma}_0^{-1})\mathbf{V}^{-1}(\mathbf{X}'\mathbf{X} + \boldsymbol{\Sigma}_0^{-1})\mathbf{M}_{\mathbf{B}_{\text{post}}}\Big] \\ &\qquad - 2\Big[\mathbf{M}'_{\mathbf{B}_{\text{post}}}(\mathbf{X}'\mathbf{X} + \boldsymbol{\Sigma}_0^{-1})\mathbf{V}^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{Y}}\Big] - \Big[\tilde{\mathbf{Y}}'\tilde{\mathbf{X}}\mathbf{V}^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{Y}}\Big]. \end{aligned} \tag{A.48}$$

Regrouping the squared terms in $\tilde{\mathbf{Y}}$ (first and last terms of Equation (A.48)), and removing the two squared terms in $\mathbf{M}_{\mathbf{B}_{\text{post}}}$ (second and third terms; note they will simplify to $\mathbf{0}$ later on) provides:

$$\tilde{\mathbf{Y}}'(\mathbf{I} - \tilde{\mathbf{X}}\mathbf{V}^{-1}\tilde{\mathbf{X}}')\tilde{\mathbf{Y}} - 2\Big[\mathbf{M}'_{\mathbf{B}_{\text{post}}}(\mathbf{X}'\mathbf{X} + \boldsymbol{\Sigma}_0^{-1})\mathbf{V}^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{Y}}\Big]. \tag{A.49}$$

Let $\mathbf{C} = (\mathbf{I} - \tilde{\mathbf{X}}\mathbf{V}^{-1}\tilde{\mathbf{X}}')$. Notably, it is the inverse of the (predictive) variance. Pre-multiplying $\tilde{\mathbf{Y}}$ in the crossed term ($-2[...]$) by $\mathbf{I} = (\mathbf{C}^{-1}\mathbf{C})$ and completing the quadratic form allow obtaining:

$$\begin{aligned} &\tilde{\mathbf{Y}}'\mathbf{C}\tilde{\mathbf{Y}} - 2\Big[\mathbf{M}'_{\mathbf{B}_{\text{post}}}(\mathbf{X}'\mathbf{X} + \boldsymbol{\Sigma}_0^{-1})\mathbf{V}^{-1}\tilde{\mathbf{X}}'(\mathbf{C}^{-1}\mathbf{C})\tilde{\mathbf{Y}}\Big] \\ &\qquad + \Big[\mathbf{M}'_{\mathbf{B}_{\text{post}}}(\mathbf{X}'\mathbf{X} + \boldsymbol{\Sigma}_0^{-1})\mathbf{V}^{-1}\tilde{\mathbf{X}}'\mathbf{C}^{-1}\mathbf{C}\mathbf{C}^{-1}\tilde{\mathbf{X}}\mathbf{V}^{-1}(\mathbf{X}'\mathbf{X} + \boldsymbol{\Sigma}_0^{-1})\mathbf{M}_{\mathbf{B}_{\text{post}}}\Big] \\ &\qquad - \Big[\mathbf{M}'_{\mathbf{B}_{\text{post}}}(\mathbf{X}'\mathbf{X} + \boldsymbol{\Sigma}_0^{-1})\mathbf{V}^{-1}\tilde{\mathbf{X}}'\cancel{\mathbf{C}}^{-1}\cancel{\mathbf{C}}\mathbf{C}^{-1}\tilde{\mathbf{X}}\mathbf{V}^{-1}(\mathbf{X}'\mathbf{X} + \boldsymbol{\Sigma}_0^{-1})\mathbf{M}_{\mathbf{B}_{\text{post}}}\Big] \\ &= \Big[\tilde{\mathbf{Y}} - \mathbf{C}^{-1}\tilde{\mathbf{X}}\mathbf{V}^{-1}(\mathbf{X}'\mathbf{X} + \boldsymbol{\Sigma}_0^{-1})\mathbf{M}_{\mathbf{B}_{\text{post}}}\Big]' \mathbf{C} \Big[\tilde{\mathbf{Y}} - \mathbf{C}^{-1}\tilde{\mathbf{X}}\mathbf{V}^{-1}(\mathbf{X}'\mathbf{X} + \boldsymbol{\Sigma}_0^{-1})\mathbf{M}_{\mathbf{B}_{\text{post}}}\Big] \\ &\qquad - \Big[\mathbf{M}'_{\mathbf{B}_{\text{post}}}(\mathbf{X}'\mathbf{X} + \boldsymbol{\Sigma}_0^{-1})\mathbf{V}^{-1}\tilde{\mathbf{X}}'\mathbf{C}^{-1}\tilde{\mathbf{X}}\mathbf{V}^{-1}(\mathbf{X}'\mathbf{X} + \boldsymbol{\Sigma}_0^{-1})\mathbf{M}_{\mathbf{B}_{\text{post}}}\Big]. \end{aligned} \tag{A.50}$$

Assuming $\boldsymbol{\mu}_{\mathbf{y}} = \mathbf{C}^{-1}\tilde{\mathbf{X}}\mathbf{V}^{-1}(\mathbf{X}'\mathbf{X} + \boldsymbol{\Sigma}_0^{-1})\mathbf{M}_{\mathbf{B}_{\text{post}}}$, this last equation simplifies to:

$$\Big[\tilde{\mathbf{Y}} - \boldsymbol{\mu}_{\mathbf{y}}\Big]' \mathbf{C} \Big[\tilde{\mathbf{Y}} - \boldsymbol{\mu}_{\mathbf{y}}\Big] - \Big[\boldsymbol{\mu}'_{\mathbf{y}}\mathbf{C}\boldsymbol{\mu}_{\mathbf{y}}\Big],$$

which is the desired quadratic form in $\tilde{\mathbf{Y}}$. Considering the constant terms w.r.t $\tilde{\mathbf{Y}}$ are dropped, and that the density is conditional to $\boldsymbol{\Sigma}$, the density of Equation A.40 may then be simplified in

$$p(\tilde{\mathbf{Y}} \mid \tilde{\mathbf{X}}, \boldsymbol{\Sigma}, \text{data}) \propto \exp\left(-\frac{1}{2} tr\Big[\boldsymbol{\Sigma}^{-1}\Big((\tilde{\mathbf{Y}} - \boldsymbol{\mu}_{\mathbf{y}})^{'}\mathbf{C}(\tilde{\mathbf{Y}} - \boldsymbol{\mu}_{\mathbf{y}})\Big)\Big]\right). \tag{A.51}$$

Conditional to $\boldsymbol{\Sigma}$, this joint predictive distribution for $\tilde{n}$ new samples is identified as a matrix-variate Normal distribution:

$$\tilde{\mathbf{Y}} \mid \tilde{\mathbf{X}}, \boldsymbol{\Sigma}, \text{data} \sim N_{\tilde{n}\times m}\left(\boldsymbol{\mu}_{\mathbf{y}}, \boldsymbol{\Sigma}, \mathbf{C}^{-1}\right). \tag{A.52}$$

Some simplifications can be easily made. First, noticing that $\mathbf{C}^{-1} = (\mathbf{I} + \tilde{\mathbf{X}}(\mathbf{X}^{'}\mathbf{X} + \boldsymbol{\Sigma}_0^{-1})^{-1}\tilde{\mathbf{X}}^{'})$ (see e.g. Marriott and Spencer, 2001), $\boldsymbol{\mu}_{\mathbf{y}}$ directly simplifies into $\tilde{\mathbf{X}}\mathbf{M}_{\mathbf{B}_{\text{post}}}$, which was expected.

Second, regrouping all the terms in $\mathbf{M}_{\mathbf{B}_{\text{post}}}$, which were removed during the computations (from Equations (A.48) and (A.51)), these lasts can be shown to be equal to $\mathbf{0}$:

$$\begin{aligned}\mathbf{M}^{'}_{\mathbf{B}_{\text{post}}}(\mathbf{X}^{'}\mathbf{X} + \boldsymbol{\Sigma}_0^{-1})\mathbf{M}_{\mathbf{B}_{\text{post}}} - \Big[\mathbf{M}^{'}_{\mathbf{B}_{\text{post}}}(\mathbf{X}^{'}\mathbf{X} + \boldsymbol{\Sigma}_0^{-1})\mathbf{V}^{-1}(\mathbf{X}^{'}\mathbf{X} + \boldsymbol{\Sigma}_0^{-1})\mathbf{M}_{\mathbf{B}_{\text{post}}}\Big] \\ - \boldsymbol{\mu}^{'}_{\mathbf{y}}\mathbf{C}\boldsymbol{\mu}_{\mathbf{y}}.\end{aligned} \tag{A.53}$$

As $\boldsymbol{\mu}_{\mathbf{y}}$ as been defined as $\mathbf{C}^{-1}\tilde{\mathbf{X}}\mathbf{V}^{-1}(\mathbf{X}^{'}\mathbf{X} + \boldsymbol{\Sigma}_0^{-1})\mathbf{M}_{\mathbf{B}_{\text{post}}} = \tilde{\mathbf{X}}\mathbf{M}_{\mathbf{B}_{\text{post}}}$, the following relation holds:

$$\mu^{'}_{\mathbf{y}}\mathbf{C}\boldsymbol{\mu}_{\mathbf{y}} = \Big[\mathbf{M}^{'}_{\mathbf{B}_{\text{post}}}(\mathbf{X}^{'}\mathbf{X} + \boldsymbol{\Sigma}_0^{-1})\mathbf{V}^{-1}\tilde{\mathbf{X}}^{'}\tilde{\mathbf{X}}\mathbf{M}_{\mathbf{B}_{\text{post}}}\Big].$$

Finally, replacing $\boldsymbol{\mu}^{'}_{\mathbf{y}}\mathbf{C}\boldsymbol{\mu}_{\mathbf{y}}$ in Equation (A.53) gives:

$$\begin{aligned}&\mathbf{M}^{'}_{\mathbf{B}_{\text{post}}}(\mathbf{X}^{'}\mathbf{X} + \boldsymbol{\Sigma}_0^{-1})\mathbf{M}_{\mathbf{B}_{\text{post}}} - \Big[\mathbf{M}^{'}_{\mathbf{B}_{\text{post}}}(\mathbf{X}^{'}\mathbf{X} + \boldsymbol{\Sigma}_0^{-1})\mathbf{V}^{-1}(\mathbf{X}^{'}\mathbf{X} + \boldsymbol{\Sigma}_0^{-1})\mathbf{M}_{\mathbf{B}_{\text{post}}}\Big] \\ &\qquad - \Big[\mathbf{M}^{'}_{\mathbf{B}_{\text{post}}}(\mathbf{X}^{'}\mathbf{X} + \boldsymbol{\Sigma}_0^{-1})\mathbf{V}^{-1}\tilde{\mathbf{X}}^{'}\tilde{\mathbf{X}}\mathbf{M}_{\mathbf{B}_{\text{post}}}\Big] \\ &= \mathbf{M}^{'}_{\mathbf{B}_{\text{post}}}(\mathbf{X}^{'}\mathbf{X} + \boldsymbol{\Sigma}_0^{-1})\left[\mathbf{I} - \mathbf{V}^{-1}\Big((\mathbf{X}^{'}\mathbf{X} + \boldsymbol{\Sigma}_0^{-1}) + \tilde{\mathbf{X}}^{'}\tilde{\mathbf{X}}\Big)\right]\mathbf{M}_{\mathbf{B}_{\text{post}}} \\ &= \mathbf{M}^{'}_{\mathbf{B}_{\text{post}}}(\mathbf{X}^{'}\mathbf{X} + \boldsymbol{\Sigma}_0^{-1})\left[\mathbf{I} - \mathbf{V}^{-1}\Big(\mathbf{V}\Big)\right]\mathbf{M}_{\mathbf{B}_{\text{post}}} \\ &= \mathbf{M}^{'}_{\mathbf{B}_{\text{post}}}(\mathbf{X}^{'}\mathbf{X} + \boldsymbol{\Sigma}_0^{-1})\left[\mathbf{I} - \mathbf{I}\right]\mathbf{M}_{\mathbf{B}_{\text{post}}} \\ &= \mathbf{0}.\end{aligned} \tag{A.54}$$

Notice that, conditional to $\boldsymbol{\Sigma}$, the joint predictive distribution for $\tilde{n}$ new samples can be better expressed as:

$$\tilde{\mathbf{Y}} \mid \tilde{\mathbf{X}}, \boldsymbol{\Sigma}, \text{data} \sim N_{\tilde{n}\times m}\left(\tilde{\mathbf{X}}\mathbf{M}_{\mathbf{B}_{\text{post}}}, \boldsymbol{\Sigma}, \mathbf{I} + \tilde{\mathbf{X}}(\mathbf{X}^{'}\mathbf{X} + \boldsymbol{\Sigma}_0^{-1})^{-1}\tilde{\mathbf{X}}^{'}\right). \tag{A.55}$$

The non-zero constants are $|\mathbf{\Sigma}|^{\frac{-\tilde{n}}{2}}$ from Equation (A.40) and $(2\pi)^{\frac{pm}{2}}.\,|\mathbf{\Sigma}|^{\frac{p}{2}}.\,|\mathbf{V}|^{\frac{m}{2}}$ from Equation (A.46). These constants given $\mathbf{\Sigma}$ will be important in the next section.

Adding the uncertainty of $\mathbf{\Sigma}$. Until now, $\mathbf{\Sigma}$ has been assumed known when identifying the predictive distribution of new response vectors. With $\mathbf{\Sigma}$ unknown, its uncertainty is added by multiplicating the predictive density of the previous section by the posterior density of $\mathbf{\Sigma}$, and integrate everything over $\mathbf{\Sigma}$:

$$p(\tilde{\mathbf{Y}} \mid \tilde{\mathbf{X}}, \text{data}) \propto \int_{\mathbf{\Sigma}} p(\tilde{\mathbf{Y}} \mid \tilde{\mathbf{X}}, \mathbf{\Sigma}, \text{data}).p(\mathbf{\Sigma} \mid \text{data}).d\mathbf{\Sigma} \tag{A.56}$$

In Equation (A.40), $p(\tilde{\mathbf{Y}} \mid \tilde{\mathbf{X}}, \mathbf{\Sigma}, \text{data})$ was computed using only the last line of the joint posterior in Equation (A.31). Retrieving all the constants, the predictive density could be written as:

$$\begin{aligned}
p(\tilde{\mathbf{Y}} \mid \tilde{\mathbf{X}}, \text{data}) \propto \int_{\mathbf{\Sigma}} & |\mathbf{\Sigma}|^{\frac{-\tilde{n}}{2}} (2\pi)^{\frac{pm}{2}} |\mathbf{\Sigma}|^{\frac{p}{2}} |\mathbf{V}|^{\frac{m}{2}} \\
& .\exp\Bigg(-\frac{1}{2} tr\Big[\mathbf{\Sigma}^{-1}\Big((\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\mathbf{M}_{\mathbf{B}\text{post}})'\mathbf{C}(\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\mathbf{M}_{\mathbf{B}\text{post}})\Big)\Big]\Bigg) \\
& .(2\pi)^{\frac{-mn}{2}} (2\pi)^{\frac{-pm}{2}} |\mathbf{\Sigma}_0|^{\frac{-m}{2}} 2^{\frac{m(\nu_0+m-1)}{2}} \Gamma_m(\tfrac{\nu_0+m-1}{2})^{-1} |\mathbf{\Omega}|^{\frac{\nu_0+m-1}{2}} \\
& .\,|\mathbf{\Sigma}|^{\frac{-n}{2}} |\mathbf{\Sigma}|^{\frac{-p}{2}} |\mathbf{\Sigma}|^{-\frac{\nu_0+2m}{2}} \\
& .\exp\Big(-\frac{1}{2} tr\Big[\mathbf{\Sigma}^{-1}\mathbf{\Omega}\Big]\Big).\exp\Big(-\frac{1}{2} tr\Big[\mathbf{\Sigma}^{-1}\mathbf{A}^*\Big]\Big)\, d\mathbf{\Sigma}.
\end{aligned} \tag{A.57}$$

This last equation can be considerably simplified by dropping every term that does not comprise $\mathbf{\Sigma}$ or $\tilde{\mathbf{Y}}$. In this case, it looks as:

$$\begin{aligned}
p(\tilde{\mathbf{Y}} \mid \tilde{\mathbf{X}}, \text{ data}) \propto \int_{\mathbf{\Sigma}} & |\mathbf{\Sigma}|^{\frac{-\tilde{n}}{2}} |\mathbf{\Sigma}|^{\frac{p}{2}} |\mathbf{\Sigma}|^{\frac{-n}{2}} |\mathbf{\Sigma}|^{\frac{-p}{2}} |\mathbf{\Sigma}|^{-\frac{\nu_0+2m}{2}} \\
& .\exp\Bigg(-\frac{1}{2} tr\Big[\mathbf{\Sigma}^{-1}\Big((\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\mathbf{M}_{\mathbf{B}\text{post}})'\mathbf{C}(\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\mathbf{M}_{\mathbf{B}\text{post}})\Big)\Big]\Bigg) \\
& .\exp\Big(-\frac{1}{2} tr\Big[\mathbf{\Sigma}^{-1}\mathbf{\Omega}\Big]\Big).\exp\Big(-\frac{1}{2} tr\Big[\mathbf{\Sigma}^{-1}\mathbf{A}^*\Big]\Big)\, d\mathbf{\Sigma}. \\
\propto \int_{\mathbf{\Sigma}} & |\mathbf{\Sigma}|^{-\frac{\tilde{n}+n+\nu_0+2m}{2}} \\
& \exp\Big(-\frac{1}{2} tr\Big[\mathbf{\Sigma}^{-1}\Big(\mathbf{\Omega} + \mathbf{A}^* + (\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\mathbf{M}_{\mathbf{B}\text{post}})'\mathbf{C}(\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\mathbf{M}_{\mathbf{B}\text{post}})\Big)\Big]\Big)\, d\mathbf{\Sigma}.
\end{aligned} \tag{A.58}$$

To solve this integral (see Equation (A.30)), the transformation of $\mathbf{\Sigma}$ to $\mathbf{\Sigma}^{-1}$ is necessary, with Jacobian equals to $|\mathbf{\Sigma}|^{m+1}$. Again, this is similar to assume that $\mathbf{\Sigma}^{-1}$ follows a Wishart distribution. Indeed, the Wishart distribution is part of the generative process of a matrix-variate Student's distribution, as in Gupta and Nagar (1999). Multiplying by the Jacobian, the predictive density is expressed:

$$\begin{aligned}
p(\tilde{\mathbf{Y}} \mid \tilde{\mathbf{X}}, \text{data}) &\propto \int_{\boldsymbol{\Sigma}^{-1}} p(\tilde{\mathbf{Y}} \mid \tilde{\mathbf{X}}, \boldsymbol{\Sigma}^{-1}, \ \text{data}).p(\boldsymbol{\Sigma}^{-1} \mid \text{data}).d\boldsymbol{\Sigma}^{-1} \\
&\propto \int_{\boldsymbol{\Sigma}^{-1}} \left|\boldsymbol{\Sigma}^{-1}\right|^{\frac{(n+\nu_0)+\tilde{n}+2m}{2}} \left|\boldsymbol{\Sigma}\right|^{m+1} \\
\exp&\left(-\frac{1}{2}tr\Big[\boldsymbol{\Sigma}^{-1}\Big(\boldsymbol{\Omega} + \mathbf{A}^* + (\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\mathbf{M}_{\mathbf{B}\text{post}})'\mathbf{C}(\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\mathbf{M}_{\mathbf{B}\text{post}})\Big)\Big]\right) d\boldsymbol{\Sigma}^{-1} \\
&\propto \int_{\boldsymbol{\Sigma}^{-1}} \left|\boldsymbol{\Sigma}^{-1}\right|^{\frac{(n+\nu_0)+\tilde{n}+m-1}{2} - \frac{m-1}{2}} \\
\exp&\left(-\frac{1}{2}tr\Big[\boldsymbol{\Sigma}^{-1}\Big(\boldsymbol{\Omega} + \mathbf{A}^* + (\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\mathbf{M}_{\mathbf{B}\text{post}})'\mathbf{C}(\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\mathbf{M}_{\mathbf{B}\text{post}})\Big)\Big]\right) d\boldsymbol{\Sigma}^{-1}.
\end{aligned} \tag{A.59}$$

The integrals in $\boldsymbol{\Sigma}^{-1}$ is performed (Equation (A.30) with $\mathbf{U} = \boldsymbol{\Sigma}^{-1}$) to obtain

$$p(\tilde{\mathbf{Y}} \mid \tilde{\mathbf{X}}, \text{data}) \propto \left|\boldsymbol{\Omega} + \mathbf{A}^* + (\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\mathbf{M}_{\mathbf{B}\text{post}})'\mathbf{C}(\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\mathbf{M}_{\mathbf{B}\text{post}})\right|^{-\frac{(n+\nu_0)+\tilde{n}+m-1}{2}}, \tag{A.60}$$

which is recognized as a matrix-variate Student's distribution of size $(\tilde{n} \times m)$ with mean $\tilde{\mathbf{X}}\mathbf{M}_{\mathbf{B}\text{post}}$, scale matrices $(\boldsymbol{\Omega} + \mathbf{A}^*)$ and $\mathbf{C}^{-1}$, and $n + \nu_0$ d.f.

As classical implementations of matrix-variate Student's distribution make use of spread matrices (often wrongly referred as covariance matrices) instead of scale matrices, the *a posteriori* spread matrix can be computed as the scale matrix divided by the total d.f.: $(\mathbf{A}^* + \boldsymbol{\Omega})/(n + \nu_0)$ (See for instance Press, 2003, p 294). Particularly, this spread matrix corresponds to the covariance matrix of the matrix-variate Normal distribution that is used in the generative process of the matrix-variate Student's distribution. Finally, this covariance matrix of the Student's distribution is:

$$(\mathbf{A}^* + \boldsymbol{\Omega})/(n + \nu_0 - 2).$$

The predictive distribution for $\tilde{n}$ new samples $\tilde{\mathbf{Y}}$ can then be written as

$$\tilde{\mathbf{Y}} \mid \tilde{\mathbf{X}}, \text{data} \sim T_{\tilde{n}\times m}\left(\tilde{\mathbf{X}}\mathbf{M}_{\mathbf{B}\text{post}}, \boldsymbol{\Omega} + \mathbf{A}^*, \mathbf{C}^{-1}, n + \nu_0\right), \tag{A.61}$$

or, more conveniently

$$vec(\tilde{\mathbf{Y}}) \mid \tilde{\mathbf{X}}, \text{data} \sim T_{\tilde{n}m}\left((\boldsymbol{\Omega} + \mathbf{A}^*) \otimes \mathbf{C}^{-1}, n + \nu_0\right). \tag{A.62}$$

This result is related to the ones of Zellner and Chetty (1965) and Kibria (2006). Finally, the marginalization property of the matrix-variate Student's distribution can be used to retrieve the predictive distribution for the prediction of only one new response vector $\tilde{\mathbf{y}}$. $\tilde{\mathbf{y}}$ is then a vector of size m (i.e., a $(1 \times m)$ matrix) and, in this case, it has the form of a multivariate Student's distribution (See Appendix D.6).

$$vec(\tilde{\mathbf{y}}) \mid \tilde{\mathbf{X}}, \text{data} \sim T_m \left((\mathbf{\Omega} + \mathbf{A}^*).\mathbf{C}^{-1}, n + \nu_0 \right), \tag{A.63}$$

where $\mathbf{C}^{-1}$ is a scalar (i.e., a (1×1) matrix).

A.5 Matrix operations

Here are presented the basic matrix operations that are used in this text. Most equations are referenced in Petersen and Pedersen (2008).

$$\begin{aligned}
(\mathbf{AB})' &= \mathbf{B}'\mathbf{A}' \quad ,..., \quad (\mathbf{ABCD})' = \mathbf{D}'\mathbf{C}'\mathbf{B}'\mathbf{A}', \quad \textit{etc.} && \text{(A.64)}\\
(\mathbf{AB})^{-1} &= \mathbf{B}^{-1}\mathbf{A}^{-1} \quad ,..., \quad (\mathbf{ABCD})^{-1} = \mathbf{D}^{-1}\mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1}, \quad \textit{etc.} && \text{(A.65)}\\
tr(\mathbf{ABC}) &= tr(\mathbf{CAB}) = tr(\mathbf{BCA}) \qquad \text{cyclic property} && \text{(A.66)}\\
tr(c.\mathbf{A}) &= c.tr(\mathbf{A}) && \text{(A.67)}\\
\mathbf{A}(\mathbf{BC}) &= (\mathbf{AB})\mathbf{C} \qquad \text{associativity} && \text{(A.68)}\\
\mathbf{A}(\mathbf{B}+\mathbf{C}) &= \mathbf{AB} + \mathbf{AC} \qquad \text{distributivity} && \text{(A.69)}\\
(\mathbf{A}+\mathbf{B})\mathbf{C} &= \mathbf{AC} + \mathbf{BC} && \text{(A.70)}\\
\mathbf{ABA} + \mathbf{ACA} &= \mathbf{A}(\mathbf{B}+\mathbf{C})\mathbf{A} && \text{(A.71)}\\
c(\mathbf{AB}) &= (c\mathbf{A})\mathbf{B} = \mathbf{A}(c\mathbf{B}) = (\mathbf{A}c)\mathbf{B} = (\mathbf{AB})c = \mathbf{A}(\mathbf{B}c) && \text{(A.72)}\\
r(\mathbf{A}+\mathbf{B}) &= r\mathbf{A} + r\mathbf{B} && \text{(A.73)}\\
\mathbf{I}_{N\times N}\mathbf{A} &= \mathbf{AI}_{M\times M} = \mathbf{A} \qquad \text{size of } \mathbf{A} : (N \times M) && \text{(A.74)}
\end{aligned}$$

Another useful property:

$$\sum_{j=1}^{m} (\mathbf{a}_j)'(\mathbf{a}_j) = (\mathbf{A}'\mathbf{A}), \tag{A.75}$$

with a_j being the j^{th} column of the $\mathbf{A}$ matrix.

A.6 Kronecker product and *vec* operator

Let $\mathbf{A}$ be an $(n \times k)$ matrix and $\mathbf{B}$ be a $(m \times l)$ matrix. The Kronecker product $\otimes$ of both matrices is the following $(nm \times kl)$ matrix:

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & \cdots & a_{1k}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{n1}\mathbf{B} & \cdots & a_{nk}\mathbf{B} \end{pmatrix} \tag{A.76}$$

It is also referred as the direct product or the tensor product. Here are some properties of the Kronecker product:

$$(\mathbf{A} \otimes \mathbf{B}) \neq (\mathbf{B} \otimes \mathbf{A}) \quad \text{most of times} \tag{A.77}$$

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD} \tag{A.78}$$

$$(\mathbf{A} \otimes \mathbf{B})' = \mathbf{A}' \otimes \mathbf{B}' \tag{A.79}$$

$$(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1} \tag{A.80}$$

Intensive usage of *vec* operator is used in this text. The *vec* operator applied on a matrix $\mathbf{A}$ stacks the columns into a vector, as shown below for a 2×2 matrix:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \Leftrightarrow vec(\mathbf{A}) = (a_{11}\ a_{21}\ a_{12}\ a_{22})' \tag{A.81}$$

The following properties link the *vec* operator to the Kronecker product or trace.

$$(\mathbf{B}' \otimes \mathbf{A})vec(\mathbf{X}) = vec(\mathbf{AXB}) \tag{A.82}$$

$$tr(\mathbf{A}'\mathbf{B}) = vec(\mathbf{A})'vec(\mathbf{B}) \tag{A.83}$$

$$tr(\mathbf{A}'\mathbf{BCD}') = vec(\mathbf{A})'(\mathbf{D} \otimes \mathbf{B})vec(\mathbf{C}) \tag{A.84}$$

Appendix B

Monte-Carlo simulation methods

The Bayesian statistical analysis has been confronted to strong limitations during many years. Indeed, before the discovery of computer based simulation methods, the Bayesian statistician had to be able to produce a posterior distribution of some variables in a closed-form and to identify a well-known distribution to allow a practical work on their results. If this identification is possible with some simple sets of problems, in many others, this analysis is not tractable.

When deriving Critical Quality Attributes (CQAs) that are rather complex functions of several variables that may contain ratios or non-linear operators (such as min or max), the situation becomes even more intricate, and the use of approximations is of limited help. In this context, Monte-Carlo simulations are a popular strategy to propagate the uncertainty of some modeled responses (the variables) to the CQAs.

Assume the interest is the posterior density $p(\boldsymbol{\theta} \mid \text{data})$. $\boldsymbol{\theta}$ is either a variable, a set of variables or a set of responses with various distribution assumptions and possible correlations. Even if available in closed-form, $p(\boldsymbol{\theta} \mid \text{data})$ is rather complex and highly dimensional.

In this case, statistics and moments such as the mean, the variance, the quantiles of the distributions and the probabilities used to make risk-based decision, etc. remains unfortunately unavailable for the responses, and *a fortiori*, for the critical quality attributes of interest. The Monte-Carlo method provides a simple way to obtain estimates of such statistics by drawing samples that follow the posterior distribution.

B.1 Monte-Carlo estimates

Metropolis and Ulam (1949) proposed to learn from a distribution by sampling elements from it. Obviously, the only limitation of their approach is to be able to draw samples from the distribution. When this is not possible directly, a solution is to use the Markov-chain Monte-Carlo method, presented in Appendix C.

To make simple, the following results are made on a univariate posterior distribution of a random variable θ. Let $\theta^{(1)}, \theta^{(2)}, ..., \theta^{(n^*)}$ be some samples from $p(\theta \mid \text{data})$. The different statistics of interests can be *estimated* from these samples with a relative accuracy, depending on how many samples are drawn and the way they are drawn.

From the samples, one can easily compute the following Monte-Carlo estimates of the properties of $p(\theta \mid \text{data})$:

1. $\hat{E}(\theta \mid \text{data}) = \bar{\theta} = \frac{1}{n^*} \sum_{s=1}^{n^*} \theta^{(s)}$.
2. $\hat{Var}(\theta \mid \text{data}) = s^2_\theta = \frac{1}{n^*-1} \sum_{s=1}^{n^*} (\theta^{(s)} - \bar{\theta})^2$.
3. $\hat{\pi} = p(\theta \leq \lambda \mid \text{data}) = \frac{1}{n^*} \sum_{s=1}^{n^*} I(\theta^{(s)} \leq \lambda)$, $I(\text{A})$ being an indicator function being 1 if A is true, 0 otherwise.
4. $\hat{\lambda}$ such that $p(\theta \leq \hat{\lambda} \mid \text{data}) = \pi$. In this case, one will look after the value of λ such that the proportion of sampled $\theta^{(s)}$ lower or equal to λ is π, $(s = 1, 2, ..., n^*)$.
5. The shape of the density curve can be visualized using a histogram or a kernel/splines density estimate on $\theta^{(1)}, \theta^{(2)}, ..., \theta^{(n^*)}$.
6. etc.

For multivariate problem, the generalization of such method is direct. For instance, the computation of $\hat{\pi}$ for m variables $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_m)$ that must achieve different quantiles Λ (acceptation limits or criteria), i.e., that must lie within a region defined by some specifications, is done as follows:

$$\hat{\pi} = p(\boldsymbol{\theta} \in \boldsymbol{\Lambda} \mid \text{data}) = \frac{1}{n^*} \sum_{s=1}^{n^*} I(\boldsymbol{\theta}^{(s)} \in \boldsymbol{\Lambda}).$$

For instance, let $\boldsymbol{\Lambda} = \{\lambda_1 \leq \boldsymbol{\theta}_1 \leq \lambda_2, \;\; \boldsymbol{\theta}_2 \leq \lambda_3, \;\; \boldsymbol{\theta}_3 \geq \lambda_4, ...\}$.

When there is interest in several CQAs that are combinations of several variables $\boldsymbol{\theta}$, i.e. $\text{CQA}_j = O_j(\boldsymbol{\theta})$, the uncertainty of $\boldsymbol{\theta}$ may be propagated to the function O_j. For each sampled value $\boldsymbol{\theta}^{(s)}$, one basically computes the critical quality attribute of interest $\text{CQA}_j^{(s)} = O_j(\boldsymbol{\theta}^{(s)}),\ (s = 1, 2, ..., n^*)$. Doing so, the uncertainties and interactions present in $\boldsymbol{\theta}$ will be propagated in the newly sampled distribution of the CQA. The different Monte-Carlo estimates can then be computed on the sampled values of the CQA(s).

B.2 Monte-Carlo error

Following the law of large numbers, and assuming the sampling is I.I.D, it is possible to estimate the accuracy of the Monte-Carlo simulated statistics, conditional to n^*. For instance, if $\bar{\theta} = \frac{1}{n^*}\sum_{s=1}^{n^*}\theta^{(s)}$ is our statistics of interest, its variance computed from repeated Monte-Carlo simulations is $Var(\bar{\theta}) = \frac{\sigma_\theta^2}{n^*}$, with σ_θ^2 being the true variance of $p(\theta \mid \text{data})$. A Monte-Carlo standard error (MCSE) estimator is then:

$$\widehat{\text{MCSE}}(\bar{\theta}) = \frac{s_\theta}{\sqrt{n^*}}, \tag{B.1}$$

where s_θ is the standard deviation computed on the sampled values. This results can be used to find the value of n^* that must be chosen to get an MCSE of a particular value δ. Assume a first set of 1000 samples has been drawn, providing a s_θ of (say) 0.8, and assume a MCSE about δ =0.01 is envisaged. Inverting equation B.1 to have n^* in function of the other quantities gives:

$$\widehat{MCSE}(\bar{\theta}) = \delta = \frac{s_\theta}{\sqrt{n^*}} \iff n^* = \frac{s_\theta^2}{\delta^2}, \tag{B.2}$$

providing $n^* = 0.8^2/0.01^2 = 6400$. Notice that, for the same problem, reaching $\delta = 0.001$ would require 640000 samples ! Generally speaking, the reduction of the error by a factor 10 (one additional accurate decimal) requires the sample size to be increased by a factor 100 (Bauer, 1958; Hammersley and Handscomb, 1964). However, the magnitude of the error on $\bar{\theta}$ remains of order $1/\sqrt{n^*}$, whatever the dimension of the problem (Robert, 2007). The increasing speed of computing devices allows today the creation of millions of samples, and the computation of useful statistics takes less than a second. Nevertheless, simply increasing the sample size to decrease simulation error might still lead to high computational burden.

In the risk-based approach presented along the manuscript, intensive use is done of the estimate of the probability of acceptance $\hat{\pi}$. Similar computations can be done on the binary vector $I(\boldsymbol{\theta}^{(s)} \in \Lambda),\ (s = 1, ..., n^*)$ to estimate its MCSE and the corresponding n^*.

Finally, in the proposal of Metropolis and Ulam, nothing indicates that Monte-Carlo simulation methods must be carried out on i.i.d. samples. Indeed, the Monte-Carlo estimates of the statistics are also obtainable if the samples are dependent, such as the ones that can be obtained using Markov-chains (Appendix C). In this case, the samples are autocorrelated. Then, they do not provide as much information about $p(\theta \mid \text{data})$ than i.i.d. samples would. As a consequence, more dependent samples are needed to have a similar accuracy than with i.i.d. samples. If there is a significative first-order auto-correlation ρ in the Markov-chain of sampled $(\theta \mid \text{data})$ (i.e. the chain is assumed AR_1), the following adapted formula can be used to compute the MCSE or the desired n^*:

$$\widehat{\text{MCSE}}_{AR_1}(\bar{\theta}) = \frac{s_\theta}{\sqrt{n^*}} \cdot \sqrt{\frac{1-\rho}{1+\rho}}, \tag{B.3}$$

where ρ can be estimated as $\hat{\rho} = \frac{1}{s_\theta \cdot (n^*-1)} \sum_{s=2}^{n^*} (\theta^{(s)} - \bar{\theta})(\theta^{(s-1)} - \bar{\theta})$.

Other methods to estimate the Monte-Carlo error exist, for instance to take into account the autocorrelations at lag of order higher than 1. See for instance Ntzoufras (2009).

Appendix C

Markov-chains Monte-Carlo methods

C.1 Introduction

Markov chains Monte-Carlo methods (or Monte-Carlo simulation from Markov chains, MCMC) are a practical tool to generate samples from a distribution of interest. Particularly, the joint posterior distribution of random variables obtained using Bayes' theorem is of great interest.

Under the name MCMC, different methodologies have been proposed to create, from a predefined starting value $\boldsymbol{\theta}^{(0)}$, a chain of correlated elements $\boldsymbol{\theta}^{(s)}$, $(s = 1, 2, 3, ...)$. This takes the form of a random walk as each $\boldsymbol{\theta}^{(s)}$ is a drawn from a *transition* distribution, conditionally to the previous element of the chain, $\boldsymbol{\theta}^{(s)} \sim g_s(\boldsymbol{\theta}^{(s)}, \boldsymbol{\theta}^{(s-1)})$. The transition distribution is chosen accordingly so that the chain has a *stationary* distribution equivalent to the posterior distribution of interest. Then, for a sufficiently large number of iterations B, $\boldsymbol{\theta}^{(s^* > B)}$ $(s^* = B+1, ..., B+n^*)$ may be viewed as random samples from the distribution, presenting more or less autocorrelations.

The elements $\boldsymbol{\theta}^{(s)}$ with $s < B$ are called the burn-in elements (or period) of the sampled chains and are generally discarded. B should be cautiously specified to make sure the chains have converged around their stationary point. At the end, useful statistics can be computes from the draws $B + 1$ to $B + n^*$. For instance, the means, the posterior modes, the variances, some probabilities of acceptance, etc. can be computed with Monte-Carlo estimates (see Appendix B). As MCMC techniques provide dependent samples, this generally leads to a slower learning, i.e., more iterations are necessary to obtain the same information about the distribution than with a i.i.d. sampling. The advantage of MCMC methods is to allow the sampling from nearly any distribution, even when it is non-identified and highly dimensioned.

The following sections present three classical MCMC algorithm, in their chronological order of appearance: the Metropolis algorithm, the Metropolis-Hasting algorithm and the Gibbs sampler.

C.2 Metropolis algorithm

The Metropolis algorithm, due to Metropolis et al. (1953), is the foundation of MCMC, and still one of the most popular methodology, because it is simple but practical. Whatever the dimensionality or complexity of the distribution, the Metropolis algorithm is able to generate samples from it. Furthermore, the related density may only be known up to constant. This simplifies the analytical task as normalizing constants must not be computed.

Metropolis works with a transition function based on a symmetric proposal distribution

$$q(\boldsymbol{\theta}^{(s)} \mid \boldsymbol{\theta}^{(s-1)}) = q(\boldsymbol{\theta}^{(s-1)} \mid \boldsymbol{\theta}^{(s)}),$$

those draws are either accepted or rejected following a simple decision rule. Some restrictions must be observed on the proposal : q and p must have the same support, and there is a constant ϕ such as $p(\boldsymbol{\theta})/q(\boldsymbol{\theta}) \leq \phi$, $\forall\ \boldsymbol{\theta}$.

The process goes as follows: first, an initial value $\boldsymbol{\theta}^{(0)}$ is chosen. From this value, the Metropolis algorithm is used to generate the elements of the chain $\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \boldsymbol{\theta}^{(\dots)}$, by successively repeating the following steps :

```
For $s = 1$ to $n^*$
    1. From a symmetric proposal distribution $q(\boldsymbol{\theta}^{(s)} \mid \boldsymbol{\theta}^{(s-1)})$, draw a new candidate vector $\boldsymbol{\theta}_t$,
    2. compute the acceptance probability: $P_a = \min\left(1, \frac{p(\boldsymbol{\theta}^{(s)}|\text{data})}{p(\boldsymbol{\theta}^{(s-1)}|\text{data})}\right)$,
    3. keep $\boldsymbol{\theta}^{(s)}$ with probability $P_a$ or assign the old value $\boldsymbol{\theta}^{(s)} = \boldsymbol{\theta}^{(s-1)}$ otherwise.
End
```

In step 1, the proposal distribution can be a Normal distribution $\boldsymbol{\theta}^{(s)} \sim N(\boldsymbol{\theta}^{(s-1)}, v)$. It is a good practice to choose a distribution that well reflects the (posterior) distribution of interest. In step 2, one just need to evaluate the value of the (posterior) density at the particular values of the variable $\boldsymbol{\theta}^{(s)}$ and $\boldsymbol{\theta}^{(s-1)}$. The ratio of the densities at iteration s and $s-1$ makes clear that the posterior density may only be known up to a multiplicative constant. Indeed, both constants of the numerator and denominator always simplify to 1 in the fraction. In step 3, the probability P_a can simply be compared with a draw from an Uniform distribution $U(0,1)$.

The variance of the proposal distribution (here, v) plays an important role in the sampling algorithm. In the Normal case, it is defined by$v$. If this variance is too small, the proposed set of parameters will be very close from the previous one, and will be accepted very often. Unfortunately, this would lead to high autocorrelations in the chain, and a poor visit of the distribution. Indeed, the chain will stay in small region of the variable space for long period. Notice that a large autocorrelation is symptomatic and allows identifying this problem.

At opposite, if the variance is too large, incongruous values of the variable would be drawn from the proposal. These values will be rejected too frequently and the parameters will then stay at the same place possibly for many iterations. Thus, this may also lead to a slow visit of the distribution. The monitoring of the sampled chains generally reveals such problems by the (excessive) succession of similar values for the parameters.

There exist some *golden rules* about the selection of the variance of the proposal. For instance, in an univariate problem with approximately Normal posterior distribution, an acceptance rate (the number of accepted $\boldsymbol{\theta}^{(s)}$ over the total number of iterations) should be between 0.15 to 0.4 to yield at least 80% of the maximum efficiency obtainable (see Gelman et al., 1996). Then, the variance of the proposal could be chosen accordingly to reach such acceptance rate.

It has also been proposed to update the variance parameter v at each iteration to automatically attain the target acceptance rate P_{target} (Haario et al., 2001). After each iteration, let

$$\sqrt{v_{s+1}} = h(\sqrt{v_s} + \frac{1}{s.(P_a(s) - P_{target})}) \tag{C.1}$$

with

$$h(x) = \begin{cases} c & \text{if} \quad x < c \\ x & \text{if} \quad x \in (c, A); \\ A & \text{if} \quad x > A; \end{cases} \tag{C.2}$$

where c is a very small positive value (e.g. 10^{-5}) and A is larger (e.g. 1). Both A and c constrain the variance of the proposal distribution. They must be chosen in accordance with the problem, and are subject to fine tuning. $P_a(s)$ is the acceptance rate after iteration s.

Finally, this type of procedure is extendable for multivariate sampling. In this case, not only the variances may be adapted, but also the covariances between the variables.

C.3 Metropolis-Hasting algorithm

In 1970, Hasting generalized the work of Metropolis et al. to use a proposal distribution that is not symmetric. More flexibility is left to use any proposal distribution. The only difference with the Metropolis algorithm is the acceptance probability that is computed as

$$P_a = \min\left(1, \frac{p(\boldsymbol{\theta}^{(s)} \mid \text{data})\ .\ q(\boldsymbol{\theta}^{(s-1)} \mid \boldsymbol{\theta}^{(s)})}{p(\boldsymbol{\theta}^{(s-1)} \mid \text{data})\ .\ q(\boldsymbol{\theta}^{(s)} \mid \boldsymbol{\theta}^{(s-1)})}\right)$$

The adaptative variance parameters of Haario et al. (2001) can be used as well.

C.4 Gibbs sampler

When the conditional distributions of some subsets $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ..., \boldsymbol{\theta}_m)$ are exactly known, one can use these distributions as the proposal distribution (Geman and Geman, 1984; Gelfand and Smith, 1990; Casella and George, 1992). As a consequence, every draws from the (conditional) proposal distributions are accepted ($P_a = 1$). The algorithm is then simplified as follows, for each iteration s:

```
For $j = 1$ to $m$
    Draw a sample from $\theta_j^{(s)} \sim p(\theta_j \mid \theta_1^{(s-1)}, \theta_2^{(s-1)}, ..., \theta_{j-1}^{(s-1)}, \theta_{j+1}^{(s-1)}, ..., \theta_m^{(s-1)}, \text{data})$,
End
```

C.5 Concluding remarks

Some basic MCMC methods have been presented in this Appendix. They allow drawing samples from a (posterior) distribution of interest. When no direct sampling is possible, the use of Gibbs sampling algorithm is of course the most desired situation as it remains computationally efficient.

Assuming $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, it is sometimes possible to identify the conditional posterior distribution for one subset of variables ($\boldsymbol{\theta}_1 \mid \text{data}$), but not for an other set ($\boldsymbol{\theta}_2 \mid \text{data}$). In this case, it is also possible to combine the different MCMC algorithms. For instance, in the iterative scheme, one will make a first Gibbs step to draw $\boldsymbol{\theta}_1$, followed by a Metropolis-Hasting step to generate a sample from $\boldsymbol{\theta}_2$.

To avoid the trouble to code complex MCMC sampler in the Bayesian context of learning from posterior distributions, it is possible to use softwares such as WinBUGS, OpenBUGS or JAGS, that allow easily dealing with MCMC techniques (Lunn et al., 2000, 2009; Plummer, 2011). Using the structured language BUGS (Spiegelhalter et al., 1996), it is easy to describe a likelihood and the related prior distributions of the variables of interest. Fed with some data, these samplers draw samples from the posterior distribution. Under the hood, different sampling algorithms are used, that include direct sampling from known distributions when possible, Gibbs sampling when the conditional posterior distributions are available in closed-form, or, when non available, using other MCMC techniques such as derivative-free adaptive rejection sampling, slice sampling (both not presented here, see Gilks, 1992; Neal, 1997), and the Metropolis-Hasting algorithm, which is one of the most polyvalent, but also one of the most inefficient.

The force of these automated sampler is to automatically select the most suitable sampling algorithm. Another great advantage is their ability to automatically tune the sampling parameters (such as the variance of the proposal distributions), which allows the user not being bothered with such considerations. The drawback related to these advantages is that this process is generally hidden within the softwares, and most often, the user does not know what has been chosen for him.

Recently, the SAS Institute Inc. has developed a MCMC sampler for the SAS system, under the MCMC procedure, which is certainly a major advance towards the massive adoption of simulation-based Bayesian statistics in the industry (SAS/-STAT® 9.2.1 User's Guide, SAS Institute Inc., 2010).

Appendix D

Multivariate densities

In this Appendix, several multivariate distributions that are used throughout this manuscript are detailed.

D.1 Wishart distribution

The Wishart distribution is a multivariate generalization of the gamma distribution. Suppose that $\mathbf{\Sigma}$ is a $(m \times m)$ positive definite symmetric matrix of random variables. The structure and the variances of $\mathbf{\Sigma}$ are defined by an $(m \times m)$ positive definite scale matrix $\mathbf{A}$ (related to a sum of squares of multivariate error) and ν degrees of freedom (d.f.).

Under classical Normality assumptions, the Wishart distribution is the conjugate prior distribution of the inverse of a covariance matrix, namely $\mathbf{\Sigma}$ (i.e. a precision matrix $\mathbf{\Sigma}^{-1}$).

Reviewing the literature, slightly different definitions of the Wishart density can be found. They differ in the way the d.f. are defined (Dawid, 1981). Two definitions are proposed hereafter.

The first one, that is widely used in the Bayesian multivariate regression domain (see Box and Tiao, 1973; Dawid, 1981), is defined as

$$
\begin{aligned}
\boldsymbol{\Sigma}^{-1} &\sim W_1(\mathbf{A}^{-1}, \nu_1) \\
p(\boldsymbol{\Sigma}^{-1}) &= 2^{\frac{m(\nu_1+m-1)}{2}} \Gamma_m\left(\frac{\nu_1+m-1}{2}\right)^{-1} \left|\mathbf{A}^{-1}\right|^{\frac{\nu_1+m-1}{2}} \left|\boldsymbol{\Sigma}^{-1}\right|^{\frac{\nu_1-2}{2}} \\
&\quad . \exp\left(-\frac{1}{2} tr[\mathbf{A}\boldsymbol{\Sigma}^{-1}]\right) \qquad \text{(D.1)}
\end{aligned}
$$

with d.f. $= \nu_1 > 0$ and $\Gamma_m(b)$ is the m-dimensional generalized gamma function

$$
\Gamma_m(b) = (\pi)^{\frac{m(m-1)}{4}} \prod_{j=1}^{m} \Gamma\left(b + \frac{j-m}{2}\right), \quad b > (m-1)/2. \qquad \text{(D.2)}
$$

$\Gamma(c)$ is the classical gamma function, with c a positive real or a complex number with a positive real part.

The second one, that tends to be more used these last decades (see Aitchison and Dunsmore, 1975; Gupta and Nagar, 1999; Gelman et al., 2004; Ntzoufras, 2009), is defined as

$$
\begin{aligned}
\boldsymbol{\Sigma}^{-1} &\sim W_2(\mathbf{A}^{-1}, \nu_2) \\
p(\boldsymbol{\Sigma}^{-1}) &= \left(2^{\frac{m\nu_2}{2}} \pi^{\frac{m(m-1)}{4}} \prod_{j=1}^{m} \Gamma\left(\frac{\nu_2+1-j}{2}\right)\right)^{-1} \left|\mathbf{A}^{-1}\right|^{\frac{-\nu_2}{2}} \left|\boldsymbol{\Sigma}^{-1}\right|^{\frac{\nu_2-m-1}{2}} \\
&\quad . \exp\left(-\frac{1}{2} tr[\mathbf{A}\boldsymbol{\Sigma}^{-1}]\right), \qquad \text{(D.3)}
\end{aligned}
$$

with d.f. $= \nu_2 > m - 1$.

This density has several interesting advantages, such as to be invariant under marginalization. This means that a submatrix $\boldsymbol{\Sigma}^{-1*}$ of $\boldsymbol{\Sigma}^{-1}$, of size $m^* \times m^*$, will simply follows a Wishart distribution $\boldsymbol{\Sigma}^{-1*} \sim W(\mathbf{A}^{-1*}, \nu)$, with the degrees of freedom kept unchanged, and $\mathbf{A}^{-1*}$ being a similar submatrix of $\mathbf{A}^{-1}$.

Notice W_2 is also the Wishart that is implemented in some R packages (Martin et al., 2010; Rossi, 2010) and in Winbugs (Lunn et al., 2000). This makes it very convenient to work with.

The obvious equivalence between W_1 and W_2 is easily obtain observing that $\nu_1 + m - 1 = \nu_2$. In other words, to implement the results presented for instance in Box and Tiao (1973), it is possible to use $W_2(\mathbf{A}^{-1}, \nu_1 + m - 1)$ instead of $W_1(\mathbf{A}^{-1}, \nu_1)$.

D.2 Inverse-Wishart distribution

The inverse-Wishart distribution is used as a natural conjugate prior for the $(m \times m)$ covariance matrix $\boldsymbol{\Sigma}$ in the context of multivariate Normal distribution

(Anderson, 1984). The structure and scale of the random variables in $\mathbf{\Sigma}$ are also described by an $(m \times m)$ positive definite scale matrix $\mathbf{A}$ and ν d.f.

Again, different forms of the inverse-Wishart density can be found in the literature, with various uses of the d.f. To be consistent with the presentation of the Wishart distribution, the inverse-Wishart found in Box and Tiao (1973) and Dawid (1981) is first presented:

$$\begin{aligned}\mathbf{\Sigma} &\sim W_1^{-1}(\mathbf{A}, \nu_1)\\ p(\mathbf{\Sigma}) &= 2^{\frac{m(\nu_1+m-1)}{2}} \Gamma_m\left(\frac{\nu_1+m-1}{2}\right)^{-1} |\mathbf{A}|^{\frac{\nu_1+m-1}{2}} |\mathbf{\Sigma}|^{-\frac{\nu_1+2m}{2}}\\ &\quad . \exp\left(-\frac{1}{2} tr[\mathbf{A}\mathbf{\Sigma}^{-1}]\right) \end{aligned} \tag{D.4}$$

with d.f. $= \nu_1 > 0$. This time, this density has the advantage to be invariant under marginalization.

As it is the case with the Wishart distribution, the second definition seems to make consensus in the recent literature, such as Gelman et al. (2004), and is also the one that can be found implemented in the R language (Martin et al., 2010; Rossi, 2010) and in the proc IML of SAS (SAS/STAT® 9.2.1 User's Guide, SAS Institute Inc., 2011). It is defined as

$$\begin{aligned}\mathbf{\Sigma} &\sim W_2^{-1}(\mathbf{A}, \nu_2)\\ p(\mathbf{\Sigma}) &= \left(2^{\frac{m\nu_2}{2}} \pi^{\frac{m(m-1)}{4}} \prod_{j=1}^{m} \Gamma\left(\frac{\nu_2+1-i}{2}\right)\right)^{-1} |\mathbf{A}|^{\frac{\nu_2}{2}} |\mathbf{\Sigma}|^{-\frac{\nu_2+m+1}{2}}\\ &\quad . \exp\left[-\frac{1}{2} tr(\mathbf{A}\mathbf{\Sigma}^{-1})\right], \end{aligned} \tag{D.5}$$

with $\nu_2 > m-1$. Note again the clear relation $W_2^{-1}(\mathbf{A}, \nu_1+m-1) = W_1^{-1}(\mathbf{A}^{-1}, \nu_1)$.

There exists at least one other form of the inverse-Wishart density, that can be found in Gupta and Nagar (1999) or Tiao and Zellner (1964). This rare form is the one that is used by SAS in proc MCMC (SAS/STAT® 9.2.1 User's Guide, SAS Institute Inc. (2010)).

D.2.1 Relation between Wishart and Inverse-Wishart distributions

A very convenient relation is expressed as follows:

$$\mathbf{\Sigma}^{-1} \sim W_1(\mathbf{A}^{-1}, \nu_1) \quad \text{iff} \quad \mathbf{\Sigma} \sim W_1^{-1}(\mathbf{A}, \nu_1) \tag{D.6}$$

$$\mathbf{\Sigma}^{-1} \sim W_2(\mathbf{A}^{-1}, \nu_2) \quad \text{iff} \quad \mathbf{\Sigma} \sim W_2^{-1}(\mathbf{A}, \nu_2) \tag{D.7}$$

To be convinced of the previous relationships, it is possible to account for the transformation of $\mathbf{\Sigma}$ (inverse-Wishart) to $\mathbf{\Sigma}^{-1}$ (Wishart), using the Jacobian of the transformation that is equal to $|\mathbf{\Sigma}|^{m+1}$ (Box and Tiao, 1973).

D.3 Multivariate Normal distribution

The multivariate Normal distribution is of much importance as it is a classical regression model assumption, and a typical prior distribution for the model parameters of a multiple linear regression. It is described by many textbook. In the Bayesian framework, two classical references are Box and Tiao (1973) and Gelman et al. (2004).

Let $\boldsymbol{\theta}$ be a vector of m random variables with mean location $\boldsymbol{\mu}$ and symmetric positive definite $(m \times m)$ covariance matrix $\mathbf{\Sigma}$. Then, if the distribution of $\boldsymbol{\theta}$ is multivariate Normal, it is defined as

$$\begin{aligned} &\boldsymbol{\theta} \sim N_m(\boldsymbol{\mu}, \mathbf{\Sigma}) \\ p(\boldsymbol{\theta}) = &(2\pi)^{-\frac{m}{2}} |\mathbf{\Sigma}|^{-\frac{1}{2}} . \exp\left[-\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\boldsymbol{\theta}-\boldsymbol{\mu})\right]. \end{aligned} \tag{D.8}$$

In this case, $E(\boldsymbol{\theta}) = mean(\boldsymbol{\theta}) = mode(\boldsymbol{\theta}) = \boldsymbol{\mu}$ and $var(\boldsymbol{\theta}) = \mathbf{\Sigma}$.

D.4 Matrix-variate Normal distribution

A generalization of the multivariate Normal distribution for a matrix of random variables is the matrix-variate Normal distribution. It commonly occurs as a proper prior distribution for the parameters of a multivariate linear regression, when expressed in matrix form, and conditionally to the variance $\mathbf{\Sigma}$. It can also be retrieved to be the posterior predictive distribution of several jointly predicted vector of responses of a multivariate linear regression, still assuming the variance $\mathbf{\Sigma}$ known.

Let $\mathbf{\Theta}$ be a $(p \times m)$ matrix of random variables with mean location $\mathbf{M}$ and two symmetric positive definite $(m \times m)$ and $(p \times p)$ covariance matrix $\mathbf{\Sigma}$ (for the columns of $\mathbf{\Theta}$) and $\mathbf{\Omega}$ (for the rows of $\mathbf{\Theta}$), respectively. Then the matrix-variate Normal density of $\mathbf{\Theta}$ is

$$\begin{aligned} &\mathbf{\Theta} \sim N_{p\times m}(\mathbf{M}, \mathbf{\Sigma}, \mathbf{\Omega}), \\ p(\mathbf{\Theta}) = &(2\pi)^{-\frac{pm}{2}} |\mathbf{\Omega}|^{-\frac{m}{2}} |\mathbf{\Sigma}|^{-\frac{p}{2}} . \exp\left[-\frac{1}{2}tr\left(\mathbf{\Sigma}^{-1}(\mathbf{\Theta}-\mathbf{M})'\mathbf{\Omega}^{-1}(\mathbf{\Theta}-\mathbf{M})\right)\right]. \end{aligned} \tag{D.9}$$

D.4.1 Link between the Multivariate and Matrix-Variate Normal distributions

When using the matrix-variate Normal distribution, it is generally easier to apply the following identity:

$$\boldsymbol{\Theta} \sim N_{p\times m}(\mathbf{M}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}) \quad \Leftrightarrow \quad vec(\boldsymbol{\Theta}) \sim N_{pm}(\mathbf{M}, \boldsymbol{\Sigma} \otimes \boldsymbol{\Omega}), \tag{D.10}$$

where $\otimes$ is the Kronecker product and the $vec(\mathbf{X})$ operator stacks the column of the $(p \times m)$ matrix $\mathbf{X}$ into a (column) vector of size pm. In other word, to draw a sample from a matrix-variate Normal distribution, it is possible to draw a sample from a multivariate Normal and unstack the resulting vector into the desired matrix, following the previous relation.

Proof. The proof is direct considering the three following relations applied on the density of Equation D.9, that allow to retrieve the density of Equation D.8 :

$$|\mathbf{A}|^{-1} = |\mathbf{A}^{-1}| \tag{D.11}$$

$$|\mathbf{A} \otimes \mathbf{B}| = |\mathbf{A}|^{m} |\mathbf{B}|^{p}, \text{ if } rank(A) = p \text{ and } rank(B) = m \tag{D.12}$$

$$tr(\mathbf{A}'\mathbf{B}'\mathbf{C}\mathbf{D}) = vec(\mathbf{B})'(\mathbf{C} \otimes \mathbf{A})vec(\mathbf{D}). \tag{D.13}$$

D.5 Multivariate Student's distribution

When working on limited samples (when d.f. is small), normality is often a too strong hypothesis. The Student's distribution (or t-distribution) is more adapted in this case. More specifically, it extends the univariate Student's distribution in the same way the univariate Normal distribution is generalized by the multivariate Normal.

Notice that different types of multivariate Student's distributions exist as different generalization of the univariate Student's distributions are used. Hereafter is presented one of the most common distribution, that is found and discussed in Johnson and Kotz (1972); Sutradhar (2006). This form allows for non-centrality (i.e., when the mean may be different from $\mathbf{0}$) and scaling. Notice this definition uses a scale matrix instead of a correlation matrix, as in Kotz and Nadarajah (2004), Equation (1.1)).

Let $\boldsymbol{\theta}$ be a vector of m random variables with mean location $\boldsymbol{\mu}$ and symmetric positive definite $(m \times m)$ *scale* matrix $\mathbf{A}$. Let also be ν d.f. If the distribution of $\boldsymbol{\theta}$ is a noncentral multivariate Student, it is generated by two variables $\mathbf{y} \sim N(0, \boldsymbol{\Sigma})$

and $w \sim \chi^2(\nu)$ such that $\boldsymbol{\theta} = \mathbf{y}\sqrt{\nu/w} + \boldsymbol{\mu}$. In this case, $\boldsymbol{\theta}$ has the following density:

$$\begin{aligned} &\boldsymbol{\theta} \sim T_m(\boldsymbol{\mu}, \mathbf{A}, \nu) \\ &p(\boldsymbol{\theta}) = \frac{\Gamma(\frac{\nu+m}{2})}{\Gamma(\frac{\nu}{2})}(\pi\nu)^{-\frac{m}{2}}\,|\mathbf{A}|^{-\frac{1}{2}}\left(1 + (\boldsymbol{\theta}-\boldsymbol{\mu})'\mathbf{A}(\boldsymbol{\theta}-\boldsymbol{\mu})\right)^{-\frac{\nu+m}{2}} \end{aligned} \tag{D.14}$$

if $m = 1$, the probability density function reduces to a noncentral scaled univariate Student's distribution. Further, if $\mathbf{A} = 1$ and $\boldsymbol{\mu} = \mu = 0$, the distribution of the univariate $\boldsymbol{\theta}$ reduces to the classical univariate Student's distribution.

To clarify a common confusion, note that $\mathbf{A}/\nu$ is the covariance matrix of the multivariate Normal distribution that is in the generative process of the multivariate Student's distribution. However it is not the covariance matrix of the Student's distribution. The covariance matrix of the multivariate Students is defined as $(\mathbf{A}/\nu).\nu/(\nu-2) = \mathbf{A}/(\nu-2)$, if $\nu > 2$. It is why, in the context of the multivariate Student's distribution, $\mathbf{A}/\nu$ is often called a *scale* or *spread* matrix, as in Sutradhar (2006) and Gupta and Nagar (1999). To avoid confusion with the scale matrices of the Wishart and inverse-Wishart distributions (i.e., $\mathbf{A}$), the terms spread matrix is preferably employed for $\mathbf{A}/\nu$. In summary,

- $\mathbf{A}$ is a *scale* matrix,
- $E(\boldsymbol{\Sigma}) = \mathbf{A}/\nu$ is a *covariance* matrix for Normal distribution, and a *spread* matrix for Student's distribution,
- $\mathbf{A}/(\nu-2)$ is a *covariance* matrix for a Student's distribution.

The covariance matrix of the multivariate Student's distribution is then undefined for $\nu \leq 2$. Some degenerative or generalized forms of the multivariate Student's distribution exist when $0 \leq \nu \leq 2$ and when $\nu \in \mathbb{R}$. In the context of drawing samples from a multivariate Student's distribution, as presented in Appendix E, the classical χ^2 distribution allows for positive and non integer degrees of freedom, $\nu \in \mathbb{R}_0^+$ (Johnson et al. (1995)).

D.6 Matrix-variate Student's distribution

As the matrix-variate Normal extends the multivariate Normal for matrix of random values instead of vector of random values, the matrix-variate Student's distribution extends the multivariate Student's distribution presented in the previous section. The matrix-variate Student's distribution is a very convenient form for the posterior distribution of regression parameters in multivariate multiple regression.

It is also the distribution of the joint prediction of several new responses in the same context.

Again, different forms exists, that differ for instance in the choice of the degrees of freedom (Dawid (1981)). The focus is kept on the definition proposed first by Kshirsagar (1961), discussed by Dickey (1967) and used by Box and Tiao (1973), as it directly generalizes Equation D.14. A complete discussion about the matrix-variate Student's distribution and its properties can be found in Gupta and Nagar (1999).

Let $\mathbf{\Theta}$ be an $(p \times m)$ matrix of random variables with mean location $\mathbf{M}$ and two symmetric positive definite $(m \times m)$ and $(p \times p)$ scale matrices $\mathbf{A}$ (for the columns of $\mathbf{\Theta}$) and $\mathbf{\Omega}$ (for the rows of $\mathbf{\Theta}$), respectively. Finally, $\mathbf{\Theta}$ has ν d.f. $(\nu > 0)$. Then $\mathbf{\Theta}$ follows a matrix-variate Student's distribution if

$$\begin{aligned}\mathbf{\Theta} &\sim T_{p\times m}(\mathbf{M}, \mathbf{A}, \mathbf{\Omega}, \nu),\\ p(\mathbf{\Theta}) &= \left(\Gamma(\tfrac{1}{2})^{pm} \frac{\Gamma_m(\frac{1}{2}\nu + m - 1)}{\Gamma_m(\frac{1}{2}\nu + m + p - 1)}\right)^{-1} |\mathbf{\Omega}|^{-\frac{m}{2}} \, |\mathbf{A}|^{-\frac{p}{2}}\\ &\qquad . \left|\mathbf{I}_m + \mathbf{A}^{-1}(\mathbf{\Theta} - \mathbf{M})'\mathbf{\Omega}^{-1}(\mathbf{\Theta} - \mathbf{M})\right|^{-\frac{1}{2}(v+m+p-1)} . \qquad \text{(D.15)}\end{aligned}$$

Notice that, in the exponent, $(v + M + F - 1)$ may simply be the number of observations n in regression context. When $m = 1$ or $p = 1$, the density reduces to a multivariate Student's distribution.

D.6.1 Marginal distribution

Assume that $\mathbf{\Theta} \sim T_{p\times m}(\mathbf{M}, \mathbf{A}, \mathbf{\Omega}, \nu)$ and that the interest lies in solely a partition of the values in $\mathbf{\Theta}$. Partitioning $\mathbf{\Theta}$ and $\mathbf{M}$ in columns, and $\mathbf{A}$ accordingly as

$$\mathbf{\Theta} = (\mathbf{\Theta}_1, \mathbf{\Theta}_2), \quad \mathbf{M} = (\mathbf{M}_1, \mathbf{M}_2), \text{ and } \quad \mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} \qquad \text{(D.16)}$$

with $\mathbf{\Theta}_1$ and $\mathbf{M}_1$ being $(p \times m_1)$ matrices $(1 \leq m_1 \leq m)$ and $\mathbf{A}$ is $(m_1 \times m_1)$, yields $\mathbf{\Theta}_1 \sim T_{p\times m_1}(\mathbf{M}_1, \mathbf{A}_{11}, \mathbf{\Omega}, \nu)$. This results is proven in Gupta and Nagar (1999).

The same partitioning apply also when splitting $\mathbf{\Theta}$ in lines. In this case, taking the corresponding part of $\mathbf{\Omega}$ allows retrieving a similar matrix-variate Student's distribution. This means that the matrix-variate Student's distribution is invariant under marginalization. At the limit, when only one line or one column (in this case, $m_1 = 1$) is taken from $\mathbf{\Theta}$, the multivariate Student's distribution holds.

This marginalization property is important as the sampling from a matrix-variate Student's might be intricate. Indeed, many different ways to sample random variables already exist only for the multivariate Student's distribution (see previous section), and the task is clearly harder in this matrix-variate case. The marginalization then allows obtaining marginal quantities easily (e.g., the marginal distribution of the regression parameters of the multivariate regression).

Appendix E

Sampling from a Student's distribution

The Student's distribution, or "t" distribution, has been introduced by Gosset (1908), under the pseudonym of Student. The Student's distribution is well adapted in the context of hypothesis testing for small samples, when Normality is difficult to prove. Indeed, Normal distribution fits well when the sample size from a population is very large, so that its parameters (mean and variance) can be accurately estimated. They can then be assumed known. When departing from this large sample hypothesis, the Student's distribution allows, thanks to its heavier tails, to better model the parameters uncertainty.

Logically, the Bayesian statistical framework often lead to Student's distribution, because uncertainty of the parameters is prominent in the analysis.

Here several methods are presented to draw samples from the univariate Student, the multivariate Student and the linear constrained multivariate Student's distributions. These results are finally extended to the matrix-variate Student's distribution. These samples can then be used in Monte-Carlo simulations.

E.1 Univariate Student's distribution

The univariate Student's distribution is defined as the probability distribution of the variable

$$X = \frac{Z}{\sqrt{V/\nu}} \tag{E.1}$$

where Z is Normally distributed as $N(0,1)$, V is chi-square distributed, with ν degrees of freedom (d.f.), and Z and V are independent. The ratio is said to follow the Student's distribution with ν d.f, and this distribution is noted $T(\nu)$.

If Z is distributed as $N(\mu, \sigma^2)$, this gives the more general case of the non-central, scaled, Student's distribution (also known as three-parameters Student distribution), $T(\mu, a, \nu)$, with a/ν being en estimator of σ^2.

To draw a sample from a Student's distribution, the previous explanations simply suggest to draw a sample from a Normal distribution and divide this sample by a square-rooted draw from a chi-square distributed variable with ν d.f., divided by ν:

- For $s = 1$ to n^*
- 1. Sample $z^{(s)}$ from $N(\mu, a/\nu)$
- 2. Sample $v^{(s)}$ from $\chi^2(\nu)$
- 3. The Student's sample $x^{(s)}$ equals $z^{(s)} / \sqrt{v^{(s)}/\nu}$
- End.

Normal samples can be obtained by inverse transform sampling or rejection sampling while chi-square samples may be obtained using rejection sampling.

E.2 Multivariate Student's distribution

The distribution of the multivariate data $\mathbf{X}$ of size m, that follow the multivariate (non-central, scaled) Student's distribution $T_m(\boldsymbol{\mu}, \mathbf{A}, \nu)$ is obtained by the ratio of a multivariate normal $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ to an independent chi-square variable $\sqrt{\chi^2(\nu)/\nu}$. An estimator of $\boldsymbol{\Sigma}$ is assumed to be $\mathbf{A}/\nu$.

Then, this simply extends the previous univariate example :

- For $s = 1$ to n^*
- 1. Sample $\mathbf{z}^{(s)}$ from $N(\boldsymbol{\mu}, \mathbf{A}/\nu)$
- 2. Sample $v^{(s)}$ from $\chi^2(\nu)$
- 3. The Student's sample $\mathbf{x}^{(s)}$ equals $\mathbf{z}^{(s)} / \sqrt{v^{(s)}/\nu}$
- End.

E.3 Truncated sampling

Let's assume that several constraints apply on the distribution of the m-sized vectors $\mathbf{x} \sim T(\boldsymbol{\mu}, \mathbf{A}, \nu)$. Geweke (1991) proposed a method to sample $\mathbf{x}$ subject to the constraints $\mathbf{a} \leq \mathbf{C}\mathbf{x} \leq \mathbf{b}$, where $\mathbf{C}$ is a full-rank $(m \times m)$ matrix and the elements of $\mathbf{a}$ and $\mathbf{b}$ can be any real to $-\infty$ (for $\mathbf{a}$), and to $+\infty$ (for $\mathbf{b}$). Then, maximum m linear restrictions can be applied on $\mathbf{x}$.

A naive rejection procedure could be applied, deleting each sample that does not fulfill one of the constraints. Doing so could be dramatic ! As the number of constraints can be high, and they can sometimes be hard to achieve, the massive number of samples (potentially all, almost surely) that would be deleted invalidates this naive solution for many applications.

The first subsection shows how to proceed with the (simpler) problem of truncated multivariate Normal distribution. Next, it is explained how to use the same methodology for the truncated multivariate Student's distribution in the second subsections.

E.3.1 Truncated multivariate Normal distribution

Geweke showed that the problem of constructing samples from $\mathbf{x} \sim N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ w.r.t $\mathbf{a} \leq \mathbf{C}\mathbf{x} \leq \mathbf{b}$ is equivalent to the sampling of

$$\mathbf{z} \sim N_m(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{C}}), \qquad \boldsymbol{\alpha} \leq \mathbf{z} \leq \boldsymbol{\beta}, \tag{E.2}$$

with

$$\boldsymbol{\Sigma}_{\mathbf{C}} = \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}', \quad \boldsymbol{\alpha} = \mathbf{a} - \mathbf{C}\boldsymbol{\mu}, \quad \boldsymbol{\beta} = \mathbf{b} - \mathbf{C}\boldsymbol{\mu},$$

$\mathbf{x}$ being equals to $\boldsymbol{\mu} + \mathbf{C}^{-1}\mathbf{z}$.

The method relies on the fact that each element of $\mathbf{z}$ is a truncated univariate Normal, conditional to all the other elements of $\mathbf{z}$. A Gibbs sampler can then be used to sample from these sub-distributions.

Suppose that, using the non-truncated (classical) Normal distribution $N(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{C}})$, we have

$$E[z_i \mid \mathbf{z}_{-i}] = \sum_{j \neq i} c_{ij} zj.$$

Now, using the truncated Normal distribution of (E.2), the distribution of $z_i \mid \mathbf{z}_{-i}$ ($\mathbf{z}_{-i}$ is the vector $\mathbf{z}$ with deleted element i) can be constructed as

$$z_i \mid \mathbf{z}_{-i} = \sum_{j \neq i} c_{ij} z_j + h_i \epsilon_i, \quad \epsilon_i \sim TN\left(\frac{\alpha_i - \sum_{j \neq i} c_{ij} z_j}{h_i}, \frac{\beta_i - \sum_{j \neq i} c_{ij} z_j}{h_i}\right),$$

where $TN(a,b)$ is the standardized univariate normal restricted to (a,b), that can be easily simulated with an inverse transform sampling where the limit of the uniform distribution are $\phi(a)$ and $\phi(b)$, $\phi(.)$ being the probability density function of a $N(0,1)$. The vector of coefficients in the conditional means $c_{i,-i} = (c_{i,1},...,c_{i,i-1},c_{i,i+1},...,c_{i,m})'$, $i = 1,...,m$ is defined as:

$$c_{i,-i} = -\frac{\Sigma^{-1}_{\mathbf{C};i,-i}}{(\Sigma^{-1}_{\mathbf{C};ii})},$$

where $\Sigma^{-1}_{\mathbf{C};ii}$ is the i^{th} diagonal element of $\mathbf{\Sigma}^{-1}_{\mathbf{C}}$, and $\Sigma^{-1}_{\mathbf{C};i,-i}$ is the i^{th} row of $\mathbf{\Sigma}^{-1}_{\mathbf{C}}$ with the i^{th} diagonal elements deleted. Finally, let define

$$h_i^2 = 1/\Sigma^{-1}_{\mathbf{C};ii}$$

In a Gibbs sampling scheme, one will construct at each iteration, the successive values $z_i^{(s)} \mid (z_1^{(s)},...,z_{i-1}^{(s)},z_{i+1}^{(s-1)},...,z_m^{(s-1)})$, and finally one get the samples of the truncated distribution doing $\mathbf{x}^{(s)} = \boldsymbol{\mu} + \mathbf{C}^{-1}\mathbf{z}^{(s)}$.

A burn-in period should be envisaged, and a comparison with a classical multivariate Normal sampling can be interesting, particularly if the variables in $\mathbf{x}$ are strongly correlated. In this case, it is advised to draw more samples. Indeed, as the z_i variables are sampled conditionally to the other variables in $\mathbf{z}$, but without accounting for any correlation structure, a slow exploration of the distribution will be observed, as well as high auto-correlations. The sample size for correlated samples is discussed in Appendix B.

Notice that other methods exist to generate samples from a constrained multivariate Normal distribution. Among other, the sampler presented by Rodriguez-Yam et al. (2004) seems to be much more efficient than the sampler of Geweke, as it provides nearly i.i.d. samples, even in presence of strong correlations between the variables. It also allows for any number of constraints.

E.3.2 Truncated multivariate Student's distribution

Now, if the samples are assumed to follow a multivariate Student's distribution with ν d.f., $\mathbf{x} \sim T_m(\boldsymbol{\mu},\mathbf{\Sigma},\nu)$, w.r.t to the linear constraints $\mathbf{a} \leq \mathbf{Cx} \leq \mathbf{b}$, the same idea is used. The Student construction is made by the ratio of a multivariate Normal and an independent chi-square $w \sim \sqrt{\chi^2(\nu)/\nu}$. This leads to the Gibbs sampling algorithm for the vector of parameters $(w, z_1, ..., z_M)$. The elements $\mathbf{x}$ are retrieved computing $\mathbf{x} = \boldsymbol{\mu} + \mathbf{C}^{-1}\mathbf{z}w^{-1}$.

At iteration s, $(w^{(s-1)}, z_1^{(s-1)}, ..., z_m^{(s-1)})$ are available. First, Geweke proposes to draw $w^{(s)} \sim \sqrt{\chi^2(\nu)/\nu}$, subject to the constraints

$$\alpha_j w^{(s)} \leq \mathbf{z}_j^{(s-1)} \leq \beta_j w^{(s)} \qquad (j = 1, ..., m)$$

with an acceptance/rejection procedure. Second, $\mathbf{z}$ can be drawn from the truncated multivariate Normal conditional to $w^{(s)}$, with the restrictions

$$\alpha_j w^{(s)} \leq \mathbf{z}_j^{(s)} \leq \beta_j w^{(s)}$$

doing

$$\begin{aligned} z_i \mid \mathbf{z}_{-i} &= \sum_{j \neq i} c_{ij} z_j + h_i \epsilon_i, \\ \epsilon_i &\sim TN\left(\frac{\alpha_i w^{(s)} - \sum_{j \neq i} c_{ij} z_j}{h_i}, \frac{\beta_i w^{(s)} - \sum_{j \neq i} c_{ij} z_j}{h_i}\right). \end{aligned}$$

At the end, one can compute $\mathbf{x}^{(s)} = \boldsymbol{\mu} + \mathbf{C}^{-1}\mathbf{z}^{(s)}(w^{(s)})^{-1}$.

The acceptance/rejection procedure for the sampling of w is fortunately very efficient and a low number of samples are generally rejected.

We developed the truncated multivariate Normal and Student algorithms with the R language (R Development Core Team, 2010), with intensive Gibbs sub-routine programmed in C language for efficiency. The C code, although not directly portable, has been successfully compiled for Windows XP (32 bits) and Mac OS 10.6 (64 bits). No problems should be noted for a Unix/Linux version.

E.4 Matrix-variate Student's t-distribution

In several application, the matrix-variate Student distribution might be used to draw samples for a set of $\tilde{n}$ new response vectors. As the rows of the generated matrix are independent, it is possible use the marginalization properties of the matrix-variate Student distribution. In this case, each row is distributed as a simpler multivariate Student.

However, for computational efficacy purpose, it may be practical to draw first a matrix-variate Normal (with lines that are independent) and, second, to add an independent chi-square draw for every line/prediction.

For the variable $\mathbf{X}$ following this matrix-variate Student $T_{\tilde{n}\times m}(\mathbf{M}, \mathbf{A}, \boldsymbol{\Omega}, \nu)$, this scheme can be used to obtain samples $\mathbf{X}^{(s)}$:

- For $s = 1$ to n^*
- 1. Draw $\mathbf{Z}^{(s)} \sim N_{\tilde{n}\times m}(\mathbf{M}, \frac{\mathbf{A}}{\nu}, \mathbf{\Omega})$, i.e. $vec(\mathbf{Z}^{(s)}) \sim N_{\tilde{n}m}(vec(\mathbf{M}), \frac{\mathbf{A}}{\nu} \otimes \mathbf{\Omega})$,
- 2. Draw $\tilde{n}$ samples $v_{\tilde{i}}^{(s)}$ from $\chi^2(\nu)$, $\tilde{i} = 1, ..., \tilde{n}$,
- 3. Divide each line of $\mathbf{Z}^{(s)}$, $\mathbf{z}_{\tilde{i}}^{(s)}$, by $\sqrt{v_{\tilde{i}}^{(s)}/\nu}$, to obtain $\mathbf{X}^{(s)}$.
- End.

If $\tilde{n}$ is large (with m potentially large as well), the size of the covariance matrix $(\mathbf{A}/\nu) \otimes \mathbf{\Omega}$ might be huge. In this case, a large amount of memory is necessary and computational overheads might be encountered. A proper number of simultaneous prediction $\tilde{n}$ might be chosen to optimize the computations.

Bibliography

J.A. Acuña, M.D. Vázquez, M.L. Tascón, and P. Sánchez-Batanero. Polarographic Behaviour of Aceclofenac, Tenoxicam and Droxicam in a Methanol-Water Mixture. *Journal of Pharmaceutical and Biomedical Analysis*, 36(1):157–162, Sep 2004. doi: 10.1016/j.jpba.2004.04.018.

J. Aitchison. Two papers on the comparison of bayesian and frequentist approaches to statistical problems of prediction: Bayesian tolerance regions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 26(2):161–175, 1964.

J. Aitchison and I.R. Dunsmore. *Statistical Prediction Analysis.* Cambridge University Press, 1975.

M.L. Altun, T. Ceyhan, M. Kartal, T. Atay, N. Oezdemir, and S. Cevheroglu. LC Method for The Analysis of Acetylsalicylic Acid, Caffeine and Codeine Phosphate in Pharmaceutical Preparations. *Journal of Pharmaceutical and Biomedical Analysis*, 25(1):93–101, 2001. doi: 10.1016/S0731-7085(00)00488-X.

M.J. Anderson and P.J. Whitcomb. Find the Most Favorable Formulations. *Chemical Engineering Progress*, pages 63–67, April 1998.

T.W. Anderson. *An Introduction to Multivariate Statistical Analysis.* Wiley, New-York, NY., 1984.

A. Baldinger, L. Clerdent, J. Rantanen, M. Yang, and H. Grohganz. Quality by Design Approach in the Optimization of the Spray-drying Process. *Pharmaceutical Development and Technology*, Early Online:1–9, 2001.

W.F. Bauer. The Monte Carlo Method. *Journal of the Society for Industrial and Applied Mathematics*, 6(4):438–451, 1958.

T.R. Bayes. An essay towards solving a problem in the doctrine of chances (reprinted in biometrika 45:293, 1958). *Phil. Trans. Roy. Soc. London*, 53:370–418, 1763.

A. Beer. Bestimmung der Absorption des rothen Lichts in farbigen Flüssigkeiten. *Annalen der Physik und Chemie*, 86:78–88, 1852.

M. Bland. Improving statistical quality in published research: the clinical experience. In *Non Clinical Statistics Conference*, Lyon (France), 2010.

A. Bogolomov and M. McBrien. Mutual peak matching in a series of HPLC-DAD mixture analyses. *Analytica Chimica Acta*, 490:41–58, 2003.

S. Boonkerd, M. Lauwers, M.R. Detaevernier, and Y. Michotte. Separation and Simultaneous Determination of The Components in An Analgesic Tablet Formulation by Micellar Electrokinetic Chromatography. *Journal of Chromatography A*, 695(1):97–102, 1995. doi: 10.1016/0021-9673(94)01183-F.

B. Boulanger, P. Chiap, W. Dewé, J. Crommen, and Ph. Hubert. An Analysis of The SFSTP Guide on Validation of Chromatographic Bioanalytical Methods: Progress and Limitations. *Journal of Pharmaceutical and Biomedical Analysis*, 32(4-5):753–765, Aug 2003.

G.E.P. Box and G.C. Tiao. *Bayesian Inference in Statistical Analysis.* Addison-Wesley Publishing Company, 1973.

S.P. Brooks and A. Gelman. General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*, 7:434–455, 1998.

C. Bugli. *Statistical tools for the analysis of event-related potentials in electroencephalograms.* PhD thesis, Université catholique de Louvain catholique de Louvain, Belgium, 2006.

H. Cabral-Marques and R. Almeida. Optimisation of Spray-drying Process Variables for Dry Powder Inhalation (DPI) Formulations of Corticosteroid/Cyclodextrin Inclusion Complexes. *European Journal of Pharmaceutics and Biopharmaceutics*, 73:121–129, 2009.

P.W. Carr, J. Li, A.J. Dallas, D.I. Eikens, and L.C. Tan. Revisionist Look at Solvophobic Driving Forces in Reversed-phase Liquid Chromatography. *Journal of Chromatography A*, 656:113–133, 1993.

G. Casella and E.I. George. Explaining the Gibbs sampler. *The American Statistician*, 46 (3):167–174, 1992.

C. Castagnoli, M. Yahyah, and J.J. Peterson Z. Cimarosti. Application of Quality by Design Principles for the Definition of a Robust Crystallization Process for Casopitant Mesylate. *Organic Process Research and Development*, 14:1407–1419, 2010.

Mark Chang. *Monte Carlo Simulation for the Pharmaceutical Industry.* Chapman and Hall/CRC, Taylor and Francis Group, 2011.

Y.-L. Chen and S.-M. Wu. Capillary Zone Electrophoresis for Simultaneous Determination of Seven Nonsteroidal Anti-Inflammatory Drugs in Pharmaceuticals. *Analytical and Bioanalytical Chemistry*, 381(4):907–912, 2005.

C. Chiao and M. Hamada. Analyzing Experiments with Correlated Multiple Responses. *Journal of Quality Technology*, 33:451–465, 2001.

A.F. Cowman and S.J. Foote. Chemotherapy and Drug Resistance in Malaria. *International Journal for Parasitology*, 20(4):503–513, 1990.

G.M. Cox and W. Cochran. *Experimental Designs, 2nd Edition.* Wiley, 1957. ISBN 0-471-16203-5.

D. H. Culver. *A Bayesian Analysis of the Balanced One-way Variance Components Model.* PhD thesis, University of Michigan, 1971.

M. Davidian and D.M. Giltinan. *Nonlinear Models for Repeated Measurement Data.* Monographs on Statistics and Applied Probability. Chapman and Hall, London, 1995.

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

B. De Backer, B. Debrus, P. Lebrun, L. Theunis, N. Dubois, L. Decock, A. Verstraete, Ph. Hubert, and C. Charlier. Innovative Development and Validation of an HPLC/DAD Method for the Qualitative and Quantitative Determination of Major Cannabinoids in Cannabis Plant Material. *Journal of Chromatography B*, 877:4115–4124, 2009. doi: http://hdl.handle.net/2268/4442.

S. De Gryze, I. Langhans, and M. Vendebroeck. Using The Correct Intervals For Prediction: A Tutorial on Tolerance Intervals For Ordinary Least-Squares Regression. *Chemometrics and Intelligent Laboratory Systems*, 87:147—154, 2007.

A. de Juan and R. Tauler. Factor analysis of hyphenated chromatographic data. Exploration, resolution and quantification of multicomponent systems. *Journal of Chromatography A*, 1158:184–195, 2007.

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*, 79:77–85, 2009. doi: http://hdl.handle.net/2268/13236.

B. Debrus, J. Broséus, D. Guillarme, P. Lebrun, Ph. Hubert, J-L. Veuthey, P. Esseiva, and S. Rudaz. Innovative Methodology to Transfer Conventional GC-MS Heroin Profiling to UHPLC-MS/MS. *Analytical and Bioanalytical Chemistry*, 399 (8):2719–30, Mar 2011a. doi: http://hdl.handle.net/2268/75667.

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 optimization in liquid chromatography. *Analytica Chimica Acta*, 691(1-2):33–42, 2011b.

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. Innovative 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. *Journal of Chromatography A*, 1218(31):5205–5215, 2011c. doi: http://hdl.handle.net/2268/93241.

B. Dejaegher and Y. Vander Heyden. Ruggedness and Robustness Testing. *Journal of Chromatography A*, 1158(1):138–157, 2007.

A. Delorme, T. Sejnowski, and S. Makeig. Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *NeuroImage*, 34(4):1443–1449, 2007.

E. Demidenko. *Mixed Models, Theory and Applications.* Wiley-Interscience, 2004.

W.E. Deming. *Out of the Crisis.* Cambrige, MA: Massachusetts Institute of Technology Center for Advanced Engineering Study, 1986.

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. Develoment of Response Models for Optimizing HPLC Methods. *Chemometrics and Intelligent Laboratory Systems*, 74:263–268, 2004.

J.M. Dickey. Matricvariate Generalizations of the Multivariate t Distribution and the Inverted Multivariate t Distribution. *The Annals of Mathematical Statistics*, 38(2):511–518, 1967.

J.A. DiMasi, R.W. Hansen, and H.G. Grabowski. The Price of Innovation: New Estimates of Drug Development Costs. *Journal of Health Economics*, 22:151–185, 2003.

E. Dinç, C. Yücesoy, and F. Onur. Simultaneous Spectrophotometric Determination of Mefenamic Acid and Paracetamol in a Pharmaceutical Preparation Using Ratio Spectra Derivative Spectrophotometry and Chemometric Methods. *Journal of Pharmaceutical and Biomedical Analysis*, 28(6):1091–1100, Jun 2002.

A. Dmitrienko, C. Chuang-Stein, and R. D'Agostino, editors. *Pharmaceutical Statistics Using SAS: a Practical Guide.* SAS Press, 2007.

N.R. Draper and H. Smith. *Applied Regression Analysis.* Wiley, New York, third edition, 1998.

Drug Analysis. *International Symposium on Drug Analysis.* Antwerpen, Belgium, 2010.

R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley-Interscience, New York, second edition, 2001.

W. Edwards, H. Lindman, and L. J. Savage. Bayesian Statistical Inference for Psychological Research. *Psychological Review*, 70(3):193–242, 1963.

S. Ekins, B. Boulanger, P. W. Swaan, and M. A. Z. Hupcey. Towards a New Age of Virtual ADME/tox and Multidimensional Drug Discovery. *Journal of Computer-Aided Molecular Design*, 5(4):381–401, 2002.

B.E. Ellison. On Two-Sided Tolerance Intervals for a Normal Distribution. *Annals of Mathematical Statistics*, 35(2):762–772, 1964.

European Pharmacopoeia. *(7.1th ed.) Bulk Density and Tapped Density of Powders*. Council of Europe, Strasbourg, France, 2011a.

European Pharmacopoeia. *(7.1th ed.) Powder flow*. Council of Europe, Strasbourg, France, 2011b.

European Pharmacopoeia. *(7.2th ed.) Monographs 1061900–1061901*. Council of Europe, Strasbourg, France, 2011c.

M. Feinberg, B. Boulanger, W. Dewé, and Ph. Hubert. New Advances in Method Validation and Measurement Uncertainty Aimed at Improving The Quality of Chemical Data. *Analytical and Bioanalytical Chemistry*, 380(3):502–14, Oct 2004. doi: 10.1007/s00216-004-2791-y.

J.W.A. Findlay and R.F. Dillard. Appropriate Calibration Curve Fitting in Ligand Binding Assays. *The American Association of Pharmaceutical Scientists Journal*, 9(2):260–267, 2007.

Y. Fong, H. Rue, and J. Wakefield. Bayesian Inference for Generalized Linear Mixed Models. *Biostatistics*, 11(3):397–412, 2010.

Food and Drug Administration. Guidance for Industry: Bioanalytical Method Validation. *US department of Health and Human Services, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER)*, May 2001.

Food and Drug Administration. Applying ICH Q8(R2), Q9, and Q10 Principles to CMC Review. *Chapter 5000 - Pharmaceutical Sciences*, MAPP 5016.1, February 2011.

Food and Drugs Administration. Guidance for Industry, Investigators and Reviewers. *Department of Health and Human Services*, January 2006.

Food and Drugs Administration. Pharmaceutical Quality for the 21st Century A Risk-Based Approach Progress Report. *Department of Health and Human Services*, 2007.

Food and Drugs Administration. Pharmaceutical cGMPS for the 21st Century – A Risk-Based Approach: Second Progress Report and Implementation Plan. *Department of Health and Human Services, accessed 08/17/2009*, 2009.

Food and Drugs Administration. Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials. *Department of Health and Human Services*, February 2010.

N. François, B. Govaerts, and B. Boulanger. Optimal Designs for Inverse Prediction in Univariate Nonlinear Calibration Models. *Chemometrics and Intelligent Laboratory Systems*, 74:283–292, 2004.

J.T. Franeta, D. Agbaba, S. Eric, S. Pavkov, M. Aleksic, and S. Vladimirov. HPLC Assay of Acetylsalicylic Acid, Paracetamol, Caffeine and Phenobarbital in Tablets. *Farmaco*, 57(9):709–713, Sep 2002.

D. Gamerman. Sampling from The Posterior Distribution in Generalized Linear Mixed Model. *Statistics and Computing*, 7:57–68, 1997.

S. Geisser. Bayesian Estimation in Multivariate Analysis. *The Annals of Mathematical Statistics*, 36 (1):150–159, 1965.

S. Geisser and J. Cornfield. Posterior Distributions for Multivariate Normal Parameters. *Journal of the Royal Statistical Society. Series B (Methodological)*, 25 (2): 368–376, 1963.

A.E. Gelfand and A.F.M. Smith. Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*, 85:398–409, 1990.

A. Gelman, G.O. Robert, and W.R. Gilks. Efficient Metropolis Jumping Rules. *Bayesian Statistics*, 5(599–607), 1996.

A. Gelman, Carlin J.B., H.S. Stern, and D.B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, 2004.

A. Gelman, A. Jakulin, M. G. Pittau, and Y-S. Su. A Weakly Informative Default Prior Distribution for Logistic and Other Regression Models. *The Annals of Applied Statistics*, 2(4):1360–1383, 2008.

S. Geman and D. Geman. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.

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* 23^{rd} *Symposium on the Interface, Seattle*, April 1991.

R.D. Gibb, W.H. Carter, and R.H. Myers. Incorporating Experimental Variability in The Determination of Desirable Factor Levels. Unpublished manuscrpipt, 2001.

W.R. Gilks. *Derivative-Free Adaptive Rejection Sampling for Gibbs Sampling.* In: Bernardo J.M., Berger J.O., Dawid A.P., and Smith A.F.M. (Eds.), Bayesian Statistics 4, Oxford University Press, Oxford, pp. 641–665, 1992.

W.S. Gosset. The Probable Error of A Mean. *Biometrika*, 6(1):1–25, 1908.

D.A. Groneberg, C. Witt, U. Wagner, K.F. Chung, and A. Fischer. Fundamentals of Pulmonary Drug Delivery. *Respiratory Medecine*, 97:382–387, 2007.

D. Guillarme, D.T.-T. Nguyen, S. Rudaz, and J.-L. Veuthey. Method Transfer for Fast Liquid Chromatography in Pharmaceutical Analysis: Application to Short Columns Packed with Small Particle. Part II: Isocratic Separation. *European Journal of Pharmaceutics and Biopharmaceutics*, 66(3):475–482, Jun 2007. doi: 10.1016/j.ejpb.2006.11.027.

D. Guillarme, D.T.-T. Nguyen, S. Rudaz, and J.-L. Veuthey. Method Transfer for Fast Liquid Chromatography in Pharmaceutical Analysis: Application to Short Columns Packed with Small Particle. Part II: Gradient Experiments. *European Journal of Pharmaceutics and Biopharmaceutics*, 68(2):430–440, Feb 2008. doi: 10.1016/j.ejpb.2007.06.018.

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.

I. Guttman. Tolerance regions, statistical. In *Encyclopedia of statistical sciences*, volume 9. Wiley, 1988.

H. Haario, E. Saksman, and J. Tamminen. An adaptative metropolis algorithm. *Bernouilli*, 7:223–242, 2001.

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

J.M. Hammersley and D.C. Handscomb. *Monte Carlo Methods.* Methuen, 1964.

E.C. Harrington. The Desirability Function. *Industrial Quality Control*, 21:494–498, 1965.

J.A. Hartigan and M.A. Wong. A k-means Clustering Algorithm. *Applied Statistics*, 28:100–108, 1979.

D.A. Harville and A.G. Zimmermann. The Posterior Distribution of the Fixed and Random Effects in a Mixed-Effects Linear Model. *Journal of Statistical Computation and Simulation*, 54:211–229, 1996.

W.K. Hasting. MonteCarlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, 57:97–109, 1970.

C. R. Henderson. *Statistical Method in Animal Improvement: Historical Overview, in Advances in Statistical Methods for Genetic Improvement of Livestock*, volume 1–14. Springer-Verlag, New York, 1990.

A.J. Hickey. *Pharmaceutical Inhalation Aerosol Technology.* Marcel Dekker Inc., New York, 1996.

B.M. Hill. Inference about variance components in the one-way model. *Journal of the American Statistical Association*, 60(311):806–825, 1965.

C. Horváth, W. Melander, and I. Molnár. Solvophobic Interactions in Liquid Chromatography with Nonpolar Stationary Phases (Solvophobic Theory of Reversed Phase Chromatography, Part I). *Journal of Chromatography A*, 125:129–156, 1976.

Y. Hu, W. Shen, W. Yao, and D.L. Massart. Using singular value ration for resolving peaks in HPLC-DAD data sets. *Chemometrics and Intelligent Laboratory Systems*, 77:97–103, 2005.

Ph. Hubert, J.-J. Nguyen-Huu, B. Boulanger, E. Chapuzet, P. Chiap, N. Cohen, P.-A. Compagnon, W. Dewé, M. Feinberg, M. Lallier, M. Laurentie, N. Mercier, G. Muzard, C. Nivet, and L. Valat. Harmonization of Strategies for The Validation of Quantitative Analytical Procedures, A SFSTP Proposal - Part I. *Journal of Pharmaceutical and Biomedical Analysis*, 36(3):579–586, 2004. doi: 10.1016/j.jpba.2004.07.027.

Ph. Hubert, J.-J. Nguyen-Huu, B. Boulanger, E. Chapuzet, N. Cohen, P.-A. Compagnon, W. Dewé, M. Feinberg, M. Laurentie, N. Mercier, G. Muzard, and L. Valat. Validation des Procédures Analytiques Quantitatives : Harmonisation des Démarches, Partie II - Statistiques. *STP Pharma Pratiques*, 16(1):1–31, 2006.

Ph. Hubert, J.-J. Nguyen-Huu, B. Boulanger, E. Chapuzet, N. Cohen, P.-A. Compagnon, W. Dewé, M. Feinberg, M. Laurentie, N. Mercier, G. Muzard, L. Valat, and E. Rozet. Harmonization of Strategies for The Validation of Quantitative Analytical Procedures, A SFSTP Proposal-Part III. *Journal of Pharmaceutical and Biomedical Analysis*, 45:82–96, 2007.

Ph. Hubert, J.-J. Nguyen-Huu, B. Boulanger, E. Chapuzet, N. Cohen, P.-A. Compagnon, W. Dewé, M. Feinberg, M. Laurentie, N. Mercier, G. Muzard, L. Valat, and E. Rozet. Harmonization of Strategies for The Validation of Quantitative Analytical Procedures, A SFSTP Proposal-Part IV. *Journal of Pharmaceutical and Biomedical Analysis*, 48(3):760–771, 2008.

E. Hund, Y. Vander Heyden, M. Haustein, D.L. Massart, and J. Smeyers-Verbeke. Robustness testing of a reversed-phase high-performance liquid chromatographic assay: comparison of fractional and asymmetrical factorial designs. *Journal of chromatography A*, 874(2):167–85, 2000.

W.G Hunter and W.F. Lamboy. A Bayesian Analysis of The Linear Calibration Problem (with discussion). *Technometrics*, 23(4):323–328, 1981.

A. Hyvärinen. Fast and Robust Fixed-Point Algorithms for Independent Component Analysis. *IEEE Transactions on Neural Networks*, 10(3):626–634, 1999.

A. Hyvärinen and E. Oja. A Fast Fixed-Point Algorithm for Independent Component Analysis. *Neural Computation*, 9(7):1483–1492, 1997.

A. Hyvärinen and E. Oja. Independent Component Analysis: Algorithms and Applications. *Neural Networks*, 13(4-5):411–430, 2000.

A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis.* Wiley, New York, 2001.

H. Ibrahim, A. Boyer, J. Bouajila, F. Couderc, and F. Nepveu. Determination of Non-Steroidal Anti-Inflammatory Drugs in Pharmaceuticals and Human Serum by Dual-Mode Gradient HPLC and Fluorescence Detection. *Journal of Chromatography B*, 857(1):59–66, Sep 2007. doi: 10.1016/j.jchromb.2007.07.008.

ICH E8 Expert Working Group. *ICH Topic E8, General Considerations for Clinical Trials.* CPMP/ICH/291/95. , March 1998.

ICH Q10. Guidance for Industry, Q10 Pharmaceutical Quality System. *International Conference on Harmonization (ICH) of Technical Requirements for registration of Pharmaceuticals for Human Use*, Geneva, 2008.

ICH Q2(R1). Topic Q2 (R1): Validation of Analytical Procedures: Text and Methodology. *International Conference on Harmonization (ICH) of Technical Requirements for registration of Pharmaceuticals for Human Use*, Geneva, November 2005.

ICH Q8. Guidance for Industry, Q8 Pharmaceutical Development. *International Conference on Harmonization (ICH) of Technical Requirements for registration of Pharmaceuticals for Human Use*, Geneva, 2009.

ICH Q9. Guidance for Industry, Q9 Quality Risk Management. *International Conference on Harmonization (ICH) of Technical Requirements for registration of Pharmaceuticals for Human Use*, Geneva, 2005.

ISO/CEI 17025. General Requirements for The Competence of Testing and Calibration Laboratories. *International Organization for Standardization (ISO)*, Geneva, 2005.

P. Iuliani, G. Carlucci, and A. Marrone. Investigation of The HPLC Response of NSAIDs by Fractional Experimental Design and Multivariate Regression Analysis. Response Optimization and New Retention Parameters. *Journal of Pharmaceutical and Biomedical Analysis*, 51(1):46–55, Jan 2010. doi: 10.1016/j.jpba.2009.08.001.

H. Jeffreys. *Theory of Probability.* Oxford: Clarendon Press, third edition, 1961.

H.Y. Ji, H.W. Lee, Y.H. Kim, D.W. Jeong, and H.S. Lee. Simultaneous Determination of Piroxicam, Meloxicam and Tenoxicam in Human Plasma by Liquid Chromatography With Tandem Mass Spectrometry. *Journal of Chromatography B*, 826(1-2):214–219, 2005. doi: 10.1016/j.jchromb.2005.08.023.

M.E. Johnson. *Multivariate Statistical Simulation.* John Wiley, New-York, NY, 1987.

N.L. Johnson and S. Kotz. *Distributions in Statistics: Continuous Multivariate Distributions.* Wiley, 1972.

N.L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous Univariate Distributions.* Wiley, 1995.

J.M. Juran. Directions for ASQC. *Industrial Quality Control*, 1951.

J.M. Juran. *Juran on Quality by Design: The New Steps for Planning Quality into Goods and Services.* Free Press, 1992.

M. Kartal. LC Method for The Analysis of Paracetamol, Caffeine and Codeine Phosphate in Pharmaceutical Preparations. *Journal of Pharmaceutical and Biomedical Analysis*, 26(5-6):857–864, 2001. doi: 10.1016/S0731-7085(01)00527-1.

J. Kerman. Default Prior Information in Models Estimating Rates and Proportions: The Case for Neutral Priors. In *The Second International Symposium on Biopharmaceutical Statistics*, Berlin, March 2011.

M.N. Khan and J.W.A. Findley, editors. *Ligand-Binding Assays.* Wiley, 2010.

A. I. Khuri and I. Sahai. Variance Components Analysis: A Selective Literature Survey. *International Statistical Review / Revue Internationale de Statistique*, 53 (3):279–300, 1985.

B.M.G. Kibria. The Matrix-t Distribution and its Applications in Predictive Inference. *Journal of Multivariate Analysis*, 97:785–795, 2006.

S. Kotz and S. Nadarajah. *Multivariate t Distributions and Their Applications.* Cambridge University Press, 2004.

F. Krier, M. Brion, B. Debrus, P. Lebrun, A. Driesen, E. Ziemons, B. Evrard, and Ph. Hubert. Optimisation and Validation of a Fast HPLC Method for the Quantification of Sulindac and Its Related Impurities. *Journal of Pharmaceutical and Biomedical Analysis*, 54:694–700, 2011. doi: http://hdl.handle.net/2268/75222.

K. Krishnamoorthy and T. Mathew. *Statistical Tolerance Regions: Theory, Applications and Computation.* Wiley, 2010.

A.M. Kshirsagar. Some Extensions of the Multivariate t-Distribution and the Multivariate Generalization of the Distribution of the Regression Coefficient. *Mathematical Proceedings of the Cambridge Philosophical Society*, 57:80–85, 1961.

E.P. Lankmayr, W. Wegscheider, J. Daniel-Ivad, I. Kolossvàry, G. Csonka, and M. Otto. Recent advances in fuzzy peak tracking in high-performance liquid chromatography. *Journal of Chromatography*, 485:557–567, 1989.

M.K. Laufer and C.V. Plowe. Withdrawing Antimalarial Drugs: Impact on Parasite Resistance and Implications for Malaria Treatment Policies. *Drug Resistance Updates*, 7:279–288, 2004.

C. Le Bailly de Tilleghem. *Statistical Contribution to The Virtual Multicriteria Optimisation of Combinatorial Molecules Libraries and to The Validation and Application of QSAR Models.* PhD thesis, Université catholique de Louvain, Louvain-la-Neuve, December 2007.

C. Le Bailly de Tilleghem and B. Govaerts. Uncertainty Propagation in Multiresponse Optimization using a Desirability Index. Technical Report 0532, Université catholique de Louvain, Louvain-la-Neuve, 2005a.

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. Technical Report 0531, Université catholique de Louvain, Louvain-la-Neuve, 2005b.

P. Lebrun and B. Boulanger. Optimization of Ligand-binding Assay in a QbD Environment. Use of Bayesian Non-linear Regression to Set Up Probability Profile as Quality Response. *NCS 2010, non clinical statistics conference*, Lyon, France, 27-29 sept 2010.

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. *Chemometrics and Intelligent Laboratory Systems*, 91:4–16, 2008. doi: http://hdl.handle.net/2268/1640.

P. Lebrun, B. Boulanger, and Ph. Hubert. How Can QbD Be Used and Implemented to Optimize Method Development and Validation? *Informa Life Sciences 9th annual biological assays conference*, London, UK, 3-4 Nov 2010.

P. Lebrun, B. Boulanger, and Ph. Hubert. Validation and Routine of Ligand-Binding Assays, a Bayesian Perspective. *Bayes 2011, second applied bayesian biostatistics workshop*, Louvain-la-Neuve, Belgium, 27-29 apr 2011. doi: http://www.bayes2011.org/.

P. Lebrun, B. Boulanger, B. Debrus, Ph. Lambert, and Ph. Hubert. A Bayesian Design Space for Analytical Methods Based on Multivariate Models and Predictions. *Submitted to Journal of Biopharmaceutical Statistics*, 2012a.

P. Lebrun, F. Krier, J. Mantanus, H. Grohganz, M. Yang, E. Rozet, B. Boulanger, B. Evrard, J. Rantanen, and P. Hubert. Design Space Approach in The Optimization of The Spray-Drying Process. *European Journal of Pharmaceutics and Biopharmaceutics*, 80(1):226–234, 2012b.

D.J. Lunn, A. Thomas, N. Best, and D. Spiegelhalter. WinBUGS – A Bayesian Modelling Framework: Concepts, Structure, and Extensibility. *Statistics and Computing*, 10:325–337, 2000.

D.J. Lunn, D. Spiegelhalter, A. Thomas, and N. Best. The BUGS project: Evolution, critique, and future directions. *Statistics in Medecine*, 28:3049–3067, 2009.

M. J. Maltesen, S. Bjerregaard, L. Hovgaard, S. Havelund, and M. van de Weert. Quality by Design – Spray-Drying of Insulin Intended for Inhalation. *European Journal of Pharmaceutics and Biopharmaceutics*, 70:828–838, 2008.

J. L. Marchini, C. Heaton, and B. D. Ripley. *fastICA: FastICA Algorithms to perform ICA and Projection Pursuit.* http://CRAN.R-project.org/package=fastICA, 2010.

R.D. Marini, J. Mbinze Kindenge, M.L.A. Montes, D. Debrus, P. Lebrun, J. Mantanus, E. Ziemons, S. Rudaz, and Ph. Hubert. Analytical Tools to Fight Against Counterfeit Medicines. *Chimica Oggi*, 28(5):10–14, 2010a.

R.D. Marini, E. Rozet, M.L.A. Montes, C. Rohrbasser, S. Roht, D. Rhème, P. Bonnabry, J. Schappler, J.-L. Veuthey, Ph. Hubert, and S. Rudaz. Reliable Low-Cost Capillary Electrophoresis Device for Drug Quality Control and Counterfeit Medicines. *Journal of Pharmaceutical and Biomedical Analysis*, 53(5): 1278–1287, Dec 2010b. doi: 10.1016/j.jpba.2010.07.026.

J.M. Marriott and N.M. Spencer. A Note on Bayesian Prediction from The Regression Model with Infortmative Priors. *Aust. N.Z.J. Stat.*, 43(4):473–480, 2001.

A.D. Martin, K.M. Quinn, and J. Hee Park. *MCMCpack: Markov chain Monte Carlo (MCMC) Package.* R package version 1.0-8, http://CRAN.R-project.org/package=MCMCpack, 2010.

A.J.P. Martin and R.L.M. Synge. A New Form of Chromatogram Employing Two Liquid Phases. *Biochemical Journal*, 35:1358–1368, 1941.

K. Masters. *Spray-drying in Practice.* SprayDry Consult International ApS, Copenhagen, Denmark, 2002.

M. Maury, K. Murphy, S. Kumar, L. Shi, and G. Lee. Effects of Process Variables on the Powder Yield of Spray-dried Trehalose on a Laboratory Spray-dryer. *European Journal of Pharmaceutical Sciences*, 59:565–573, 2005.

M. Mazières. *Pharmaceutiques, dossier contrefaçon.* 68–73, 2007.

J. Mbinze Kindenge, P. Lebrun, B. Debrus, F. Lecomte, J. Mavar Tayer Mbay, B. Boulanger, R.D. Marini, and Ph. Hubert. Application of An Innovative Design Space Optimization Strategy to The Development of LC Methods for Nsaids. *to be submitted in Analytical Chemistry*, 2011.

M. McKeown, T.P. Jung, S. Makeig, G. Brown, S. Kindermann, T-W. Lee, and T.J. Sejnowski. Spatially Independent Activity Patterns in Functional Magnetic Resonance Imaging Data During the Stroop Color-naming Task. In *Proceedings of the National Academy of Sciences, USA*, volume 95, pages 803–810, 1998.

G.P. McMahon, S.J. O'Connor, D.J. Fitzgerald, S. le Roy, and M.T. Kelly. Determination of Aspirin and Salicylic Acid in Transdermal Perfusates. *Journal of Chromatography B*, 707(1-2):322–327, Apr 1998.

R.W. Mee. Beta-Expectation and Beta-Content Tolerance Limits for Balanced One-Way ANOVA Random Model. *Technometrics*, 26(3):251–254, 1984.

N. Metropolis and S. Ulam. The Monte Carlo Method. *Journal of the American Statistical Association*, 44(247):335–341, september 1949.

N. Metropolis, A.W. Rosenbluth, M.N. Teller, and E. Teller. Equations of State Calculations by Fast Computing Machines. *Journal of Chemical Physics*, 21: 1087–1092, 1953.

P.E. Meyer, F. Lafitte, and G. Bontempi. MINET: An open source R/Bioconductor Package for Mutual Information based Network Inference. *BMC Bioinformatics*, 9, 2008.

V. Meyer. *Practical high-performance liquid chromatography.* Wiley, 2004.

P. Minka. Bayesian Linear Regression. *MIT Media Lab note*, 2001.

G. Miró-Quesada, E. Del Castillo, and J. Peterson. A Bayesian Approach for Multiple Response Surface Optimization in the Presence of Noise Variables. *Journal of Applied Statistics*, 31(3):251–270, 2004.

I. Molnar. Computerized design of separation strategies by reversed-phase liquid chromatography: development of Drylab software. *Journal of Chromatography A*, 965:175–194, 2002.

I. Molnar, R. Boysen, and P. Jekow. Peak tracking in high-performance liquid chromatography based on normalized band areas. *Journal of Chromatography*, 485:569–579, 1989.

D.C. Montgomery. *Design and Analysis of Experiments.* Wiley, 2009.

T.K. Mutabingwa. Artemisinin-based Combination Therapies (ACTs): best hope for malaria treatment but inaccessible to the needy! *Acta Tropica*, 95:305, 2005.

D. Nagrath, F. Xia, and S.M. Cramer. Characterization and Modeling of Nonlinear Hydrophobic Interaction Chromatographic Systems. *Journal of Chromatography A*, 1218(9):1219–1226, 2011.

R.M. Neal. Markov Chain Monte Carlo Methods Based on 'Slicing' The Density Function. Technical Report 9722, Dept. of Statistics, University of Toronto, 1997.

J. Neter, W. Wasserman, and M.H. Kutner. *Applied Linear Statistical Models, third edition.* Irwin, 1990. ISBN 0-256-08338-X.

P. Nikitas and A. Pappa-Louisi. Retention Models for Isocratic and Gradient Elution in Reversed-phase Liquid Chromatography. *Journal of Chromatography A*, 1216 (10):1737–1755, 2009.

I. Ntzoufras. *Bayesian Modeling Using WinBUGS.* John Wiley and Sons, Wiley Series in Computational Statistics, New Jersey, 2009.

C.O. Obonyo, E.A. Juma, B.R. Ogutu, J.M. Vulule, and J. Lau. Amodiaquine Combined with Sulfadoxine/Pyrimethamine versus Artemisinin-based Combinations for The Treatment of Uncomplicated Falciparum Malaria in Africa: a Meta-analysis. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 101 (2):117–126, 2007.

C. Osborne. Statistical Calibration: A Review. *International Statistical Review / Revue Internationale de Statistique*, 59(3):309–336, 1991.

A. Panusa, G. Multari, G. Incarnato, and L. Gagliardi. High-Performance Liquid Chromatography Analysis of Anti-Inflammatory Pharmaceuticals with Ultraviolet and Electrospray-Mass Spectrometry Detection in Suspected Counterfeit Homeopathic Medicinal Products. *Journal of Pharmaceutical and Biomedical Analysis*, 43(4):1221–1227, Mar 2007. doi: 10.1016/j.jpba.2006.10.012.

H. Parastar, M. Jalali-Heravi, and R. Tauler. Is Independent Component Analysis Appropriate for Multivariate Resolution in Analytical Chemistry ? *Trends in Analytical Chemistry*, To appear, 2011.

K.B. Petersen and M.S. Pedersen. *The Matrix Cookbook*. http://matrixcookbook.com, 2008.

J.J. Peterson. A Posterior Predictive Approach to Multiple Response Surface Optimization. *Journal of Quality Technology*, 36:139–153, 2004.

J.J. Peterson. A Bayesian Approach to The ICH Q8 Qefinition of Design Space . *Journal of Biopharmaceutical Statistics*, 18:959–975, 2008.

J.J. Peterson. Multivariate Predictive Distributions; A Risk-based Strategy for ICH Q8 Design Space Development. In *Nonclincial Biostatictics Conference*, Harvard School of Public Health, Boston, 2009.

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, R.D. Snee, P.R. McAllister, T.L. Schofield, and A.J. Carella. Rejoinder. *Journal of Quality Technology*, 41((2)):111, 2009.

M. Plummer. JAGS. *http://mcmc-jags.sourceforge.net/*, 2011.

M. Plummer, N. Best, K. Cowles, and K. Vines. *coda: Output analysis and diagnostics for MCMC*, 2010. URL `http://CRAN.R-project.org/package=coda`. R package version 0.14-2.

S.J. Pocock. *Clinical Trials: A Practical Approach*. John Wiley & Sons, 2004.

M. Polásek, M. Pospísilová, and M. Urbánek. Capillary Isotachophoretic Determination of Flufenamic, Mefenamic, Niflumic and Tolfenamic Acid in Pharmaceuticals. *Journal of Pharmaceutical and Biomedical Analysis*, 23(1):135–142, Aug 2000.

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

S.J. Press. *Subjective and Objective Bayesian Statistics: Principles, Models, and Applications*. John Wiley, New-York, NY, 2nd ed. edition, 2003.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, http://www.r-project.org edition, 2010.

R. Ragonese, M. Mulholland, and J. Kalman. Full and Fractionated Experimental Designs for Robustness Testing in the High-performance Liquid Chromatographic

Analysis of Codeine Phosphate, Pseudoephedrine Hydrochloride and Chlorpheniramine Maleate in a Pharmaceutical Preparation. *Journal of Chromatography A*, 870(1):45–51, 2000.

M. Rajagopalan and L. Broemeling. Bayesian Inference for The Variance Components in General Mixed Linear Models. *Communications in Statistics - Theory and Methods*, 12(6):701 – 723, 1983.

P.C. Robert. *The Bayesian Choice, from Decision-Theoretic Foundations to Computational Implementation.* Springer Texts in Statistics, New York, 2nd edition, 2007.

G. Rodriguez-Yam, R.A. Davis, and L.L. Scharf. Efficient Gibbs Sampling of Truncated Multivariate Normal with Application to Constrained Linear Regression. Technical report, Colorado State University, 2004.

P. Rossi. *bayesm: Bayesian Inference for Marketing/Micro-econometrics.* R package version 2.2-3, http://CRAN.R-project.org/package=bayesm, 2010.

E. Rozet, S. Rudaz, R.D. Marini, E. Ziémons, B. Boulanger, and Ph. Hubert. Models to Estimate Overall Analytical Measurements Uncertainty: Assumptions, Comparisons and Applications. *Analytica Chimica Acta*, 702(2):160–171, 2011.

D.B. Rubin. Estimation in Parallel Randomized Experiments. *Journal of Educational Statistics*, 6:377–401, 1981.

D.B. Rubin. Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician. *Annals of Statistics*, 12:1151–1172, 1984.

P.-Y. Sacré, E. Deconinck, P. Chiap, J. Crommen, F. Mansion, E. Rozet, P. Courselle, and J. De Beer. Development and Validation of a Ultra-High-Performance Liquid Chromatography-UV Method for The Detection and Quantification of Erectile Dysfunction Drugs and Some of Their Analogues Found in Counterfeit Medicines. *Journal of Chromatography A*, 1218(37):6439–6447, Sep 2011. doi: 10.1016/j.chroma.2011.07.029.

J. Šafra and M. Pospíšilová. Separation and Determination of Ketoprofen, Methylparaben and Propylparaben in Pharmaceutical Preparation by Micellar Electrokinetic Chromatography. *Journal of Pharmaceutical and Biomedical Analysis*, 48 (2):452–455, 2008.

A.O. Santini, H.R. Pezza, and L. Pezza. Development of A Potentiometric Mefenamate Ion Sensor for The Determination of Mefenamic Acid in Pharmaceuticals and Human Blood Serum. *Sensors and Actuators B: Chemical*, 128(1):117–123, 2007. doi: 10.1016/j.snb.2007.05.039.

SAS/STAT® 9.2.1 User's Guide, SAS Institute Inc. *www.sas.com*, 2010.

SAS/STAT® 9.2.1 User's Guide, SAS Institute Inc. *SAS/IML Function Modules for Multivariate Random Sampling*, 2011.

F.E. Satterthwaite. Synthesis of variances. *Psychometrika*, 6:309–315, 1941.

L. J. Savage. Bayesian Statistics. In *Decision and Information Process*. Macmillan and Co., 1962.

A. Savitzky and M. J. E. Golay. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8):1627–1639, 1964.

Tim Schofield. Multifactor Design of Experiments (DOE) in Bioassay Development and Validation. In *Informa Life Sciences' 9th Annual Biological Assays*, London, UK, 2010.

S. R. Searle, G. Casella, and C. E. McCulloch. *Variance Components*. John Wiley and Sons, New York, 1992.

S.S. Shapiro and M.B. Wilk. An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, 52:591–611, 1965.

Tom Short. *signal: Signal processing*. R package version 0.6-2, http://CRAN.R-project.org/package=signal, 2011.

N. Singh, M. M Shukla, G. Chand, P.K. Bharti, M. P Singh, M.K. Shukla, R.K. Mehra, R.K. Sharma, and A.P. Dash. Epidemic of Plasmodium Falciparum Malaria in Central India, an area where Chloroquine has been replaced by Artemisinin-based combination therapy. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 105(3):133–139, 2011.

L.R. Snyder, J.W. Dolan, and J.R. Gant. Gradient Elution in High-performance Liquid Chromatography : I. Theoretical Basis for Reversed-phase Systems. *Journal of Chromatography A*, 165(1):3–30, 1979.

L.R. Snyder, J.W. Dolan, and Lommen D.C. Drylab® Computer Simulation for High-performance Liquid Chromatographic Method Development : I. Isocratic Elution. *Journal of Chromatography A*, 485:91–112, 1989.

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

R.L. Snyder, J.J Kirkland, and J.W Dolan. *Introduction to Modern Liquid Chromatography, 3rd Edition*. Wiley, 2010.

D. Spiegelhalter, A. Thomas, N. Best, and W. Gilk. Bugs 0.5: Bayesian inference using gibbs sampling–manual (version ii). *Medical Research Council Biostatistics Unit, Cambridge*, 1996.

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. *Quality Technology and Quantitative Management*, 41:111, 2009.

B.C. Sutradhar. Multivariate t Distribution. In *Encyclopedia of statistical sciences.* Wiley, 2006.

K. Taylor. *The Science of Dosage Form.* Churchill Livingstone, Leicester, second edition, 2001.

The European Medicines Agency. *CPMP/QWP/486/95.* Note for Guidance on Manufacture of The Finished Dosage Form, April 1996.

The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. http://www.ich.org, 2010.

G.C. Tiao and A. Zellner. On the Bayesian Estimation of Multivariate Regression. *Journal of the Royal Statistical Society. Series B (Methodological)*, 26(2):277–285, 1964.

H. Trautmann and C. Weihs. Uncertainty of Optimum Influence Factor Levels in Multicriteria Optimization Using The Concept of Desirability. Technical Report 23/04, SFB 475, Univsersity of Dortmund, 2004.

H. Trautmann and C. Weihs. On The Distribution of The Desirability Index Using Harrington's Desirability Function. *Metrika*, 63:207–213, 2006.

M. Tswett. Physico-chemical Studies of Chlorophyll. The Adsorption. *Ber. dtsch botan. Ges.*, 24:316–323, 1906.

N. Valecha, H. Joshi, P.K. Mallick, S.K. Sharma, A. Kumar, P.K. Tyagi, B. Shahi, M.K. Das, B.N. Nagpal, and A.P. Dash. Low Efficacy of Chloroquine: Time to Switchover to Artemisinin-based Combination Therapy for Falciparum Malaria in India. *Acta Tropica*, 111(1):21–28, 2009.

G. Verbeke and G. Molenberghs. *Linear Mixed Models in Practice, a SAS-Oriented Approach.* Springer, 1997.

G. Vivó-Truyols and al. Automatic Program for Peak Detection and Deconvolution of Multi-Overlapped Chromatographic Signals. Part I : Peak Detection. *Journal of Chromatography A*, 1096:133–145, 2005a.

G. Vivó-Truyols and al. Automatic Program for Peak Detection and Deconvolution of Multi-Overlapped Chromatographic Signals. Part II : Peak Model and Deconvolution Algorithms. *Journal of Chromatography A*, 1096:146–155, 2005b.

G. Vivó-Truyols, J.R. Torres-Lapasió, and García-Alvarez-Coque M.C. Enhanced Calculation of Optimal Gradient Programs in Reversed-phase Liquid Chromatography. *Journal of Chromatography A*, 1018(2):183–196, 2003.

A. Wald and Wolfowitz. Tolerance Intervals for a Normal Distribution. *The Annals of Mathematical Sciences*, 17:208–215, 1946.

W.A. Wallis. Tolerance Intervals for Linear Regression. *in: J. Neyman (Ed.), Second Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Los Angeles, Berkeley, CA*, pages 43–51, 1951.

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.

J. H. Ward. Hierarchical Grouping to Optimize an Objective Function. *Journal of American Statistical Association*, 58(301):236–244, 1963.

H. Weber and C. Weihs. On The Distribution of The Desirability Index Using Harrington's Desirability Function. Technical Report Technical Report 2003/3, Univsersity of Dortmund, 2003.

S. Wilks. Determination of Sample Sizes For Setting Tolerance Limits. *Annals of Mathematical Statistics*, 12:91–96, 1941.

R.D. Wolfinger. Tolerance Intervals for Variance Component Models Using Bayesian Simulation. *Journal of Quality Technology*, 30(1):18–32, 1998.

World Health Organization. World Malaria Report. *http://www.who.int/malaria/world_malaria_report_2010/en/index.html*, 2010.

H. Yamamoto, K. Hada, H. Yamaji, T. Katsuda, H. Ohno, and H. Fukuda. Application of regularized alternating least squares and independent component analysis to HPLC-DAD data of Haematococcus pluvialis metabolites. *Biochemical Engineering Journal*, 32:149–156, 2006.

L. Yu. Pharmaceutical Quality by Design: Product and Process Development, Understanding, and Control. *Pharmaceutical Research*, 25:781–791, 2008.

Z. Yuan and E. Oja. *A FastICA Algorithm for Non-negative Independent Component Analysis*, chapter ICA, pages 1–8. C.G. Puntonet AND A. Prieto (eds.), Springer-Verlag, Berlin Heidelberg, 2004.

A. Zellner and V.K. Chetty. Prediction and Decision Problems in Regression Models from the Bayesian Point of View. *Journal of the American Statistical Association*, 60(310):608–616, 1965.

C.-H. Zheng, D.-S. Huang, Z.-L. Sun, M.R. Lyu, and T.-M. Lok. Nonnegative Independent Component Analysis Based on Minimizing Mutual Information Technique. *Neurocomputing*, 69:878–883, 2006.