

Dissertation

Advancements in Integrated Modelling of Biopharmaceutical Processes: Incorporating Random Effects, Holistic Design of Experiments and Design Space Exploration

carried out for the purpose of obtaining the degree of
Doctor technicae (Dr. techn.)

submitted at TU Wien

Faculty of Mechanical and Industrial Engineering

Thomas Oberleitner, MSc

Mat.No.: 0816972

- Supervisor: Univ.Prof. Dipl.-Ing. Dr.nat.tech. **Oliver Spadiut**
Institute of Chemical, Environmental and Bioscience Engineering
TU Wien, Getreidemarkt 9, 1060 Vienna, Austria
- Co-Supervisor: Univ.Prof. Dipl.-Ing. Dr.techn. **Peter Filzmoser**
Institute of Statistics and Mathematical Methods in Economics
TU Wien, Wiedner Hauptstr. 8-10, 1040 Vienna, Austria
- Reviewer: Univ.Prof.in **Efstathia Bura**, PhD
Institute of Statistics and Mathematical Methods in Economics
TU Wien, Wiedner Hauptstr. 8-10, 1040 Vienna, Austria
- Reviewer: Univ.Prof. **Massimiliano Barolo**, PhD
Department of Industrial Engineering
University of Padova, Via Francesco Marzolo, 9, Padova, Italy

Vienna, September 2023

Affidavit

I declare in lieu of oath, that I wrote this thesis and carried out the associated research myself, using only the literature cited in this volume. If text passages from sources are used literally, they are marked as such.

I confirm that this work is original and has not been submitted for examination elsewhere, nor is it currently under consideration for a thesis elsewhere.

I acknowledge that the submitted work will be checked electronically-technically using suitable and state-of-the-art means (plagiarism detection software). On the one hand, this ensures that the submitted work was prepared according to the high-quality standards within the applicable rules to ensure good scientific practice "Code of Conduct" at the TU Wien. On the other hand, a comparison with other student theses avoids violations of my personal copyright.

Vienna, September 2023

Acknowledgements

First and foremost, I would like to thank Thomas Zahel at Körber Pharma Austria GmbH for his continuing support throughout the last three and a half years. As a lateral entry to the world of biotech, he helped me through my early difficulties, showed me the ropes in a domain that was entirely alien to me, provided advice when appropriate and encouragement when needed. This dissertation would not have been possible without him.

I thank my former advisor Christoph Herwig for giving me the chance to conduct this PhD thesis as part of a funded research project and introducing me to CHASE and Körber. My thanks also go to Oliver Spadiut for taking over as my supervisor and giving me helpful advice for the final stretches of my PhD. I thank Peter Filzmoser for agreeing to co-advise me in my PhD studies and his enlightening lectures on statistics.

I also need to express my gratitude to Competence Center CHASE for taking me in and for valuable insights into how professional research is conducted. Furthermore, I want to thank the team at its company partner, Körber Pharma Austria, for giving me a warm welcome and always lending me an ear for occasional problems.

Finally, I thank my girlfriend Astrid for her support during my PhD studies, for putting up with all the hours and my nonsense in general.

This work was conducted within the COMET Centre CHASE, funded within the COMET - Competence Centers for Excellent Technologies programme by the BMK, the BMDW and the Federal Provinces of Upper Austria and Vienna. The COMET programme is managed by the Austrian Research Promotion Agency (FFG).

Abstract

Process validation, characterization, and optimization are crucial in biopharmaceutical manufacturing to ensure the quality and consistency of the final product. However, even though these are standard practices in the industry, oftentimes they are not implemented in an effective manner due to a unit operation (UO)-centric view of the process that ignores interdependencies, resulting in unnecessary experimental effort. Similarly, recommendations by regulatory authorities like the ICH or FDA, e.g., incorporating all sources of variation or providing a robust design space are in many cases not considered due to a lack of knowledge or tools.

This rather applied dissertation presents a set of methods to address these issues and thereby contributes to advancing biopharmaceutical process development and validation. Furthermore, the work provides insights to the scientific community by highlighting the practical application and effectiveness of statistical methodologies, as well as introducing novel solutions to issues commonly encountered in the field. Specifically, it is comprised of the following three key subthemes: (i) incorporating random effects in the calculation of proven acceptable ranges (PAR) for biopharmaceutical process control strategies, (ii) developing a novel experimental design approach that captures the interplay of multiple unit operations to gain insights into final product quality, and (iii) proposing a method for defining design spaces using linear combinations of process parameters. The methods are strongly related considering ICH and FDA recommendations for robust biopharmaceutical manufacturing: By incorporating random effects into the calculation of PARs and utilizing a holistic experimental design approach, manufacturers can effectively address the requirements for process validation and control strategy development as emphasized by ICH Q8. These methodologies enable a more comprehensive understanding of process variability, identification of critical process parameters, and the establishment of PARs that accurately capture the impact of random effects on critical quality attributes (CQAs). Furthermore, the proposed method for defining design spaces aligns with the flexibility and adaptability emphasized in ICH Q8 as well as lifecycle change management plans outlined in ICH Q14 and enables manufacturers to explore and define design spaces in a scientifically rigorous manner.

The first part of the research highlights the significance of considering random effects in a control strategy for biopharmaceutical processes. Through the application of linear mixed models, the impact of random effects on CQAs is quantified, demonstrating their substantial contribution to process variation. The study reveals that neglecting random effects can lead to overly optimistic proven acceptable ranges and potentially misleading statements about the CQA distribution.

In the second part, a holistic design of experiments (hDoE) methodology is introduced to address the limitations of traditional methods that focus on investigating one process step at a time. This approach leverages an integrated process model comprised of regression models for each unit operation, interconnected by passing responses to subsequent unit operations as process parameters. By minimizing the simulated out-of-specification rate at the final step of the process, the proposed approach optimally places runs at the appropriate UO, providing valuable information about the impact of individual UOs on CQA variability and significantly reducing the number of experiments required.

The third part addresses the challenge of defining design spaces that align with acceptance criteria for process parameter deviations. A novel method is presented, leveraging a numeric optimizer and regression models to calculate the largest design space within the parameter space, ensuring CQA boundaries remain within acceptable limits. This approach offers an efficient alternative to discretization-based methods, enabling fast evaluations even in higher-dimensional parameter spaces. Additionally, a weighting scheme is proposed to prioritize certain process parameters, facilitating a dynamic approach to design space definition and exploration.

This dissertation contributes to advancing biopharmaceutical process development and validation by introducing random effects, holistic experimental design, and design space optimization into the original concept of integrated process models. The methodologies presented offer valuable insights and practical tools for biopharmaceutical manufacturers aiming to ensure consistent product quality throughout the manufacturing lifecycle.

Kurzfassung

Prozessvalidierung, Charakterisierung und Optimierung sind entscheidend für die biopharmazeutische Herstellung, um die Qualität und Konsistenz des Endprodukts sicherzustellen. Obwohl dies branchenübliche Praktiken sind, werden sie oft nicht effektiv umgesetzt, da ein auf einzelne Unit Operations (UOs) ausgerichteter Ansatz angewendet wird, der die Wechselwirkungen zwischen den Schritten ignoriert und oft mit unnötigem experimentellem Aufwand verbunden ist. Ähnlich werden Empfehlungen von Zulassungsbehörden wie der ICH oder FDA oft nicht wahrgenommen, wie z.B. die Berücksichtigung aller Quellen der Variation oder die Bereitstellung eines robusten Design Spaces (DS).

Diese anwendungsorientierte Dissertation präsentiert eine Reihe von Methoden, die sich mit der Lösung solcher Probleme beschäftigen und so zur Weiterentwicklung der biopharmazeutischen Prozessentwicklung und Validierung als auch des wissenschaftlichen Bereiches der CMC-Statistik (Chemistry, Manufacturing and Controls) beitragen. Die Arbeit ist in drei Hauptbereiche unterteilt: (i) Einbeziehung von Random Effects in die Berechnung von Proven Acceptable Ranges (PAR) für Kontrollstrategien biopharmazeutischer Prozesse, (ii) Entwicklung eines neuartigen experimentellen Designansatzes, der das Zusammenspiel mehrerer UOs erfasst indem auf die Out-of-specification Rate des Gesamtprozesses optimiert wird, und (iii) eine Methode zur Definition eines DS unter Verwendung linearer Kombinationen von Prozessparametern.

Die Methoden stehen in enger Verbindung zu den Empfehlungen der ICH und FDA für eine robuste biopharmazeutische Herstellung. Durch die Einbeziehung von Random Effects in die Berechnung von PARs und die Verwendung eines ganzheitlichen experimentellen Designansatzes können Hersteller die Anforderungen an die Prozessvalidierung und die Entwicklung von Kontrollstrategien effektiv erfüllen. Diese Methoden ermöglichen ein umfassenderes Verständnis der Prozessvariabilität, die Identifizierung kritischer Prozessparameter und die Festlegung von PARs, die den Einfluss von Random Effects auf Critical Quality Attributes (CQAs) genau erfassen. Des Weiteren ermöglicht die vorgeschlagene Methode zur multivariaten Design Space Berechnung und Untersuchung Herstellern die schnelle, numerische Definition von gültigen Parameterbereichen konform zu den ICH Q8 und Q14 Guidelines.

Der erste Teil der Dissertation hebt die Bedeutung der Berücksichtigung von Random Effects in der Kontrollstrategie für biopharmazeutische Prozesse hervor. Durch die Anwendung von Linear Mixed Models wird der Einfluss von Random Effects auf CQAs quantifiziert, was ihren erheblichen Beitrag zur Prozessvariation zeigt. Eine Fallstudie belegt, dass die Vernachlässigung von Random Effects zu überoptimistischen Kontrollstrategien führen kann und zu potenziell irreführende Aussagen über die Verteilung der CQAs.

Im zweiten Teil wird holistic Design-of-Experiments (hDoE) vorgestellt, eine Methode zur optimierten Errechnung von experimentellem Aufwand zur Charakterisierung von Prozessen. Im Gegensatz zu traditionellen Methoden, die sich auf die Untersuchung eines Prozessschritts nach dem anderen konzentrieren, nutzt dieser Ansatz ein integriertes Prozessmodell. Durch Minimierung der simulierten Out-of-specification Rate dieses holistischen Modells werden experimentelle Läufe optimal auf die einzelnen UOs verteilt. Der Vorgang liefert wertvolle Informationen über den Einfluss einzelner UOs auf die Variabilität von CQAs und reduziert signifikant die Anzahl der benötigten Experimente.

Der dritte Teil befasst sich mit der numerischen Berechnung von multivariaten Design Spaces. Ein DS im Kontext eines biopharmazeutischen Prozesses ist die Definition jener Parameterbereiche, die zu CQAs innerhalb gewisser Akzeptanzkriterien führen. Im Gegensatz zu PARs beinhaltet diese Definition auch alle Kombinationen von Prozessparametern. Es wird eine neue Methode vorgestellt, die numerische Optimierung und Regressionsmodelle nutzt, um den größten Designraum im Parameterbereich zu berechnen und sicherzustellen, dass die Grenzen der CQAs innerhalb der Akzeptanzkriterien bleiben. Dieser Ansatz bietet eine effiziente Alternative zu diskretisierungsbasierten Methoden und ermöglicht eine schnelle Berechnung auch in höherdimensionalen Parameterräumen. Darüber hinaus wird ein Gewichtungsschema vorgeschlagen, um bestimmte Prozessparameter priorisieren zu können. Dies verbessert die Flexibilität der Methode und ermöglicht das dynamische Untersuchen des DS, basierend auf Anforderungen unterschiedlicher Anwendungsfälle.

Diese Dissertation trägt zur Weiterentwicklung der biopharmazeutischen Prozessentwicklung und Validierung bei, indem Kontrollstrategien mit Random Effects, ganzheitliche experimentelle Designs und die Optimierung von Design Spaces in das ursprüngliche Konzept des IPMs integriert werden. Die vorgestellten Methoden bieten wertvolle Einblicke und praktische Werkzeuge für Hersteller, die konsistente Produktqualität über den gesamten Prozess Lifecycle sicherstellen möchten.

Table of Contents

1	Introduction	12
1.1	Data Science in Biopharmaceutical Manufacturing	12
1.2	Regulatory Requirements	12
1.3	Integrated Process Models	14
1.3.1	CQA Simulation	15
1.3.2	Extrapolation	15
1.4	Regression Models	16
1.5	Tolerance Intervals	16
1.6	Optimization	17
1.7	Design of Experiments	18
1.8	Implementation in Software	18
2	Dissertation Objectives	20
3	Thesis Structure	23
4	Results	24
4.1	Incorporating Random Effects in Biopharmaceutical Control Strategies	24
4.1.1	Research Question	24
4.1.2	Problem Statement	24
4.1.3	Results	25
4.1.4	Conclusion	28
4.1.5	Publication	29
4.1.6	Contribution	29
4.2	Holistic Design of Experiments Using an Integrated Process Model	30
4.2.1	Research Question	30
4.2.2	Problem Statement	30
4.2.3	Results/Method	30
4.2.4	Conclusion	34
4.2.5	Publication	35
4.2.6	Contribution	35
4.3	Identifying Design Spaces as Linear Combinations of Parameter Ranges for Bio- pharmaceutical Control Strategies	36
4.3.1	Research Question	36
4.3.2	Problem Statement	36
4.3.3	Results/Method	36
4.3.4	Conclusion	40
4.3.5	Manuscript	40
4.3.6	Contribution	40
5	Conclusion	41
6	Bibliography	43

7	Appendix A	47
7.1	Scientific publication I	47
7.2	Scientific publication II	61
7.3	Scientific publication III	77
8	Appendix B: Models and Intervals	96
8.1	Models	96
8.1.1	Ordinary Least Squares Models	96
8.1.2	Linear Mixed Models	96
8.2	Statistical Intervals	97
8.2.1	Confidence Intervals	97
8.2.2	Prediction Intervals	98
8.2.3	Tolerance Intervals	98
9	Appendix C	100
9.1	Curriculum vitae	100

Acronyms

CI Confidence Interval

CMC Chemistry, Manufacturing and Controls

COBYLA Constrained Optimization by Linear Approximation

CQA Critical Quality Attribute

DoE Design of Experiments

DS Design Space

EMA European Medicines Agency

FDA Food and Drug Administration

GMP Good Manufacturing Practice

hDoE Holistic Design of Experiments

IAC Intermediate Acceptance Criteria

ICH International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use

IoT The Internet of Things

IPC Inter-process Controls

IPM Integrated Process Model

ISPE International Society for Pharmaceutical Engineering

LMM Linear Mixed Model

ML Maximum Likelihood

MODR Method Operational Design Regions

OFAT One-factor-at-a-time

OLS Ordinary Least Squares

OOS Out-of-specification

PAR Proven Acceptable Range

PAT Process Analytical Technology

PCS Process Characterization Study

PI Prediction Interval

PP Process Parameter

PRT Probability-Ratio-Threshold

QbD Quality by Design

REML Restricted Maximum Likelihood

SLSQP Sequential Least Squares Programming Algorithm

TI Tolerance Interval

UO Unit Operation

1 Introduction

1.1 Data Science in Biopharmaceutical Manufacturing

Automated production processes based on realtime data collection, evaluation and control, generally referred to as Industry 4.0+, represents the current state of the art in many industries. To apply these practices to the biopharmaceutical domain, the International Society for Pharmaceutical Engineering (ISPE) established the Pharma 4.0 operating model [1], where particularities of biopharma production are taken into account, such as the need for regulatory approval and Good Manufacturing Practices (GMPs) [2, 3]. The main aspects of this transformation are categorized into resources, information systems, organization and culture, whereas the prerequisite and fundamental component connecting all of these features is digitization. In the context of Pharma 4.0, digitization is often associated with The Internet of Things (IoT), realtime data collection and interpretation as well as automated control. While the latter is not yet implemented in the majority of biopharma production, the collection, standardization and analysis of data is well established and required for regulatory filings. These activities fall into the broad category of data science or, depending on the exact use case, into the domain of clinical or CMC statistics (Chemistry, Manufacturing and Controls).

Clinical statistics is concerned with the design of clinical trials and the analysis of results, and its tasks as well as many of the methods involved are firmly anchored in regulatory guidelines [4]. On the other hand, the role of CMC statistics is more general, spanning from supporting the development of new drug products and processes to process optimization and validation [5, 6]. Regulatory guidelines recommend that statistical evidence for these activities should be provided, but do not specify the exact methods. This paves the way for the utilization of a comprehensive array of data science methods, along with techniques employed in the modeling of chemical or physical systems, such as mechanistic models. The contributions presented in this cumulative dissertation find their place within the realm of CMC statistics. They have been primarily formulated for process development, i.e., the initial stage in the process validation lifecycle as defined by the FDA's process validation guideline [7]. The lifecycle concept as well as other regulatory considerations are outlined in the next section.

Another important part of the Pharma 4.0 digitalization scheme are digital twins. While the term lacks a precise definition and its interpretation varies across domains [8], within biopharmaceutical manufacturing, it commonly refers to a comprehensive model of the entire production process [9, 10]. Oftentimes these definitions include a realtime component, i.e., Process Analytical Technology (PAT) and automated control, which is not considered in this dissertation as it relates to statistical methods alone. Here, the digital twin is realized as the Integrated Process Model (IPM), an ensemble model comprised of submodels for unit operations that facilitates process validation and optimization [11]. The IPM provides the context and framework for the individual parts of this dissertation.

1.2 Regulatory Requirements

Biopharmaceutical manufacturers are required by regulatory authorities such as the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH),

the Food and Drug Administration (FDA) or the European Medicines Agency (EMA) to provide descriptions of production processes and substantiate design decisions and control strategies through reports rooted in sound scientific principles.

“The aim of pharmaceutical development is to design a quality product and its manufacturing process to consistently deliver the intended performance of the product.” [12]

These organizations supply guideline texts that offer general recommendations on how to provide evidence for the robustness of a process. For example, they describe basic principles for utilizing statistical techniques to quantify the relationship between Process Parameters (PPs) and Critical Quality Attributes (CQAs).

“Design of Experiment (DoE) studies can help develop process knowledge by revealing relationships, including multivariate interactions, between the variable inputs (e.g., component characteristics or process parameters) and the resulting outputs (e.g., in-process material, intermediates, or the final product).” [7]

Those models can then be used to define valid ranges in which a parameter can deviate, either univariately in the form of Proven Acceptable Range (PAR) (see section 4.1.2) or multivariately by calculating a full Design Space (DS) (see section 7.3), while CQA specifications are still met. This is one way of providing a control strategy that constitutes evidence of a robust process and the first step in the process validation lifecycle introduced by the FDA in its 2011 guideline text, illustrated in figure 1.1.

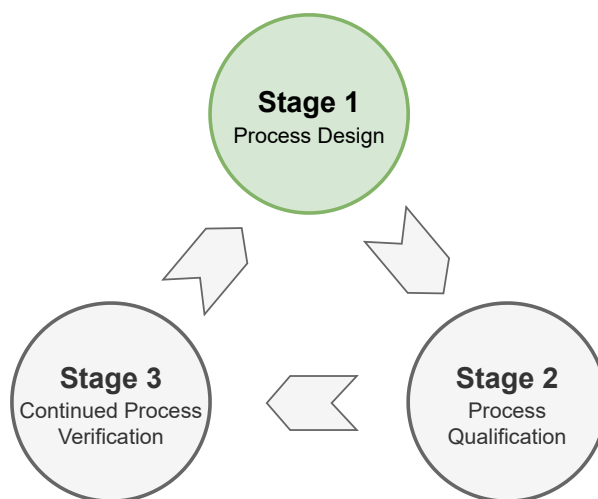


Fig. 1.1: The process validation lifecycle as defined by the FDA [7].

Process design, the first stage in this lifecycle is the most expensive for manufacturers and arguably the one that can derive the greatest benefit from statistical methods. Generally, the need for providing quantitative evidence for robustness gives rise to the scientific field of Chemistry, Manufacturing and Controls (CMC) statistics, the application of statistical methods to problems in biopharmaceutical manufacturing [5] and also the particular domain this dissertation expands upon.

While CMC statistics is a very active area of research, official process validation guidelines evolve quite slowly and mention only elementary statistical concepts. The level of detail

presented in the above quoted example is never exceeded. Furthermore, the proposed methods are recommendations rather than hard requirements.

“FDA’s guidance documents, including this guidance, do not establish legally enforceable responsibilities. Instead, guidances describe the Agency’s current thinking on a topic and should be viewed only as recommendations, unless specific regulatory or statutory requirements are cited.” [7]

Thus, manufacturers might be tempted to seek the path of least resistance when providing evidence for regulatory filings. For example, Intermediate Acceptance Criteria (IACs) or Inter-process Controls (IPC) are oftentimes derived from historical data of an Unit Operation (UO) by deviating two or three standard deviations from the mean, an approach that is flawed for several reasons and can lead to incorrect control strategies [13]. While guidelines advocate the consideration of all sources of variation, oftentimes only PPs are included in the analysis [14]. UOs tend to be investigated separately and the link between them is ignored, which can result in overly optimistic model predictions on one hand and unnecessary experimental effort on the other [15]. These are some of the issues than can be resolved by the application of appropriate statistical methods in the context of an IPM.

1.3 Integrated Process Models

The IPM is the fundamental framework for the methods introduced in this dissertation. It encapsulates either a chain of UOs or an entire production process. The data-driven variant used here was originally described by Zahel et al. [11, 16]. By concatenating UO models, the impact of PP changes on CQA distributions can be simulated not only within the corresponding UO, but also for all consecutive UOs, up to drug substance/product formulation, where predicted distributions can be compared against drug specifications. This is illustrated in figure 1.2.

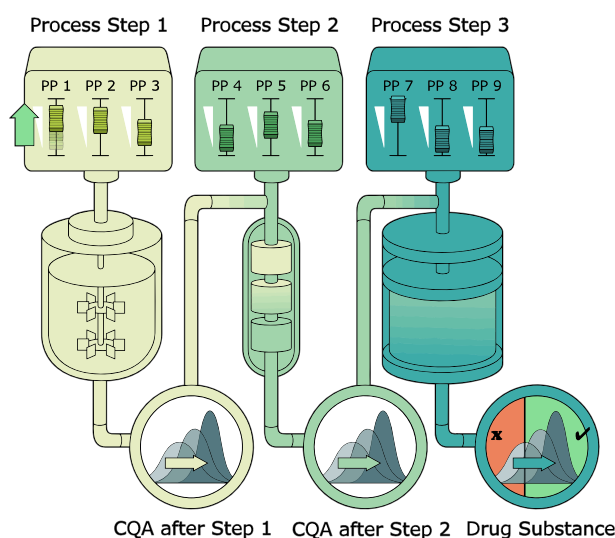


Fig. 1.2: An illustration of how PP changes affect CQAs and how the their distribution shifted along all UOs [11].

In recent years the value of holistic and end-to-end process models is increasingly recognized. They play part in determining robustness, optimization and control and represent an important tool for manufacturers to meet required Quality by Design (QbD) criteria [17].

1.3.1 CQA Simulation

After fitting UO models on representative data, CQA distributions are simulated using Monte Carlo sampling of PPs [18]. To this end, for each mean predicted by the model, an artificial data point is sampled from a normal distribution derived from a tolerance interval around that prediction. As explained in section 1.5, this results in a data point from the population distribution to a nominal level of confidence. The set of simulated CQA data points is then propagated to the next UO as the starting material, also called the *load parameter*. The procedure is repeated for each UO in the sequence up to the final step, which in most cases represents drug formulation. As a tool relevant to patient safety, special attention is given to extrapolation of the load parameter, described in the next section.

1.3.2 Extrapolation

As this is relevant for CQA simulations within the IPM in general, and for the proposed Holistic Design of Experiments (hDoE) method in particular (see section 4.2), the extrapolation scheme of load parameters is briefly explained. In the context of biopharmaceutical processes, downstream UOs either increase the purity of a product or decrease the impurity, depending on the modelled CQA. Based on this physical interpretation, a conservative extrapolation scheme of the simulated load parameter is implemented as follows. The domain of parameter values is segmented into three parts: (i) the range of experimentally investigated load parameters, (ii) the range between those load boundaries and the minimum/maximum CQA value observed in data and (iii) the range beyond both of these bounds. Each range is represented by a linear relationship between load and CQA, resulting in a piecewise linear model [19]. The remaining steps are explained for purities, i.e., CQAs where higher values represent better product quality, though a similar principle applies to impurities. Within the boundaries of investigated load values, a typical least-squares procedure is used to fit a linear relationship. Between these boundaries and the maximum value of the CQA found in data (minimum in case of impurities), impurity is not improved beyond that, i.e., even when the linear relationship would suggest a more pure product, the model predicts the maximum value in the training data. Beyond the ranges of either load or CQA data, no purification takes place, described in the regression model as a coefficient of one. Figure 1.3 shows the piecewise regression for purities and highlights the range that is affected by hDoE. For further details, please refer to [16].

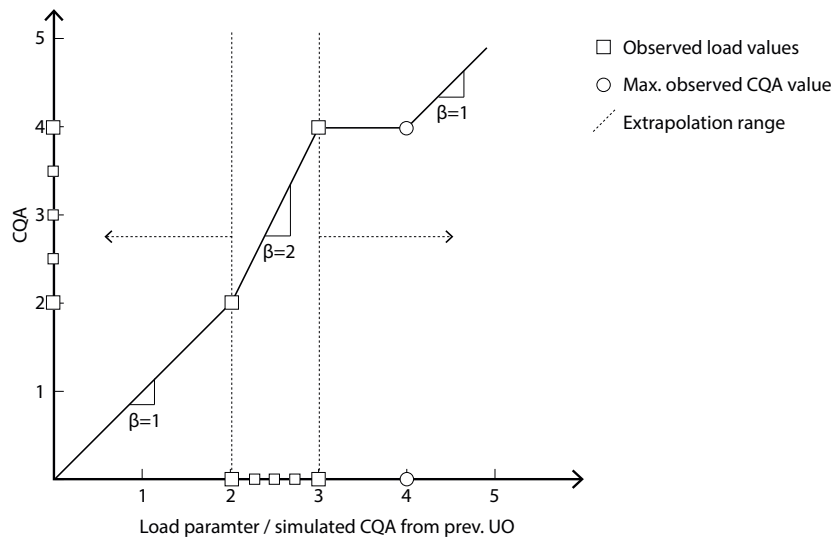


Fig. 1.3: The piecewise linear regression model for conservative extrapolation of the load parameter. Note that the dotted extrapolation range can be reduced by performing spiking studies as described in section 4.2.

1.4 Regression Models

Regression models are used to describe the relationship between PPs and CQAs per UO. In principle, the IPM framework is agnostic to the specific model type, which is also the case for the methods proposed in this dissertation. Depending on the properties of the data available, the particular IPM application or modelling preferences, models range from Ordinary Least Squares (OLS), Linear Mixed Models (LMMs), censored data models to time-dynamic approaches like mechanistic or hybrid models. The prerequisites for any model used in this implementation of the IPM are (i) the capacity to predict a single, numeric CQA value for a given PP configuration, which might result in the need for auto-integration schemes in dynamic models, and (ii) the support for uncertainty intervals around the prediction. For the studies conducted in this work, OLS models and LMMs were employed. Detailed information can be found in section 8.

1.5 Tolerance Intervals

Uncertainty quantification is of particular interest in the biopharmaceutical domain. In the light of patient safety, conservative estimates of the predicted CQA distribution are vital, although oftentimes the technical details of such intervals are poorly understood. For example, inferences on the distribution might be based on *Confidence Intervals (CIs)* which indicate the range of an estimator given a nominal confidence level. However, as regression models generally predict a mean, a CI relates directly to the distribution of the mean, not to the distribution of CQA values. *Prediction Intervals (PIs)* are an improvement on the CI as they cover the distribution of a *single*, predicted CQA, not its mean, to a nominal level of confidence. In the context of simulating CQA distributions in the IPM though, the range of interest should cover not a single value but a proportion of the population. This is where *Tolerance Intervals (TIs)* come into play. TIs cover a given proportion of the population, called coverage, to a nominal level of confidence. Detailed information about different types of intervals be found in section 8.

TIs are comprised of different sources of variation, such as parameter uncertainty and residual model error. This is important for the workflow proposed to incorporate random factors into control strategies, because the LMM-specific implementation of TIs additionally contains variation introduced by random effects, which tend to broaden the interval [20].

1.6 Optimization

Numeric optimization, or the minimization of an objective function can be found in all three of the workflows and methods introduced in this dissertation.

The OLS model $y = X\beta + \epsilon$ is a special case, as the minimization of the objective function, equivalent to estimating the vector of coefficients β , can be calculated by the closed-form expression $\hat{\beta} = (X^T X)^{-1} X^T y$, where X is the design matrix of observations, y the modelled response and ϵ the model error [19]. This is no longer the case for the Linear Mixed Models (LMMs) used to incorporate random effects into control strategies (section 4.1). The model equation $y = X\beta + Z\gamma + \epsilon$ includes an additional design matrix Z that assigns the observations in X to random effect blocks and the vector of random effect parameters γ [21]. The parameters β , γ , as well as variance components in ϵ need to be estimated simultaneously, which involves an iterative process that minimizes either Maximum Likelihood (ML) or Restricted Maximum Likelihood (REML) objective functions. As a well-defined optimization problem, a variety of optimization algorithms can be applied [22–24].

The optimization of Out-of-specification (OOS) rates in Holistic Design of Experiments (hDoE) (section 4.2) is performed on a stochastic Monte Carlo simulation and can, in this form, not be reduced to a single objective function. Instead, different types of experimental runs are simulated in each UO and optimization steps are reduced to picking the run that shows the largest improvement to the OOS rate from a table of simulation results. However, a problem related to optimization in hDoE is that of exploration vs. exploitation. As a consequence of iteratively improving UO models based on their own predictions, the process is biased towards effects that are already part of the model. In this instance, DoE runs represent the exploration aspect, as additional runs from an optimal design can uncover new effects and interactions [25], while spiking runs exploit the existing model structure. To encourage exploration, a decision scheme adapted from the Metropolis-Hastings algorithm is employed [26], which drastically improves convergence properties of the optimization (see section 7.2 for details).

The most obvious application of numeric optimization is found in the Design Space (DS) computation described in section 7.3. Here, the difficulty lies in the formulation of an objective function and a set of inequality constraints that can be minimized to find the largest possible DS spanned by linear combinations of parameter ranges, i.e., the largest possible hyper-rectangle within the parameter space. For a conservative evaluation of points in this space, tolerance interval boundaries around predictions from a regression model are compared against acceptance limits. As this can be a computational bottleneck, a polynomial approximation of the interval is computed beforehand. This speeds up the optimization procedure significantly, which in this case is performed by the Constrained Optimization by Linear Approximation (COBYLA) algorithm. As a simplex-based method, COBYLA is gradient-free and shows reasonable robustness against falling into local minima, while its support for inequality constraints facilitate the formulation of the requirements for a valid DS [27]. Results can be further improved with a second optimization pass that uses the actual TI instead of an approximation, for which a gradient-based Sequential Least Squares Programming Algorithm (SLSQP) algorithm [28] is recommended.

1.7 Design of Experiments

Design of Experiments (DoE) is a standard procedure to conduct experiments in many applied sciences. While there are various DoE variants, all of them aim to maximize the information gained about a system by systematically deviating its parameters. This can be achieved combinatorially, in the form of factorial or fractional factorial designs [25], or by maximizing different kinds of information in optimal designs, e.g. A-optimal, D-optimal, I-optimal, etc. [29]. While factorial or fractional factorial designs are considered the ideal approach for many experimental setups, specifically in situations with no constraints on the number of runs, the design region or the model, optimal designs offer a level of flexibility that renders them the tool of choice in many more constrained applications. This especially true in the biopharmaceutical domain where data is scarce and experimental effort associated with time and cost. The case study conducted for the investigation of the impact of random effects on control strategies was done mostly on DoE data from D-optimal designs, supplemented with some historical process data. hDoE uses D-optimal designs to augment the existing experimental effort performed on a UO, which is only possible in design approaches that do not require a fixed number of runs, such as optimal designs. Finally, the DS optimization procedure uses a rotatable central composite design [30] in the polynomial approximation of the tolerance interval, as this is the most efficient to characterize the quadratic form of the interval function and scales reasonably well over the number of parameters.

1.8 Implementation in Software

The applied nature of this dissertation is underlined by the fact that the presented methods are not only disseminated as articles in scientific journals but also as part of a Python software package. This code library is used by the Körber Pharma Austria GmbH data science consulting department as well as by their proprietary software PAS-X Savvy [31]. The Python package was developed as part of the research project that funded this dissertation and over the course of the PhD programme, the IPM as well as hDoE and other IPM applications were integrated into the software.

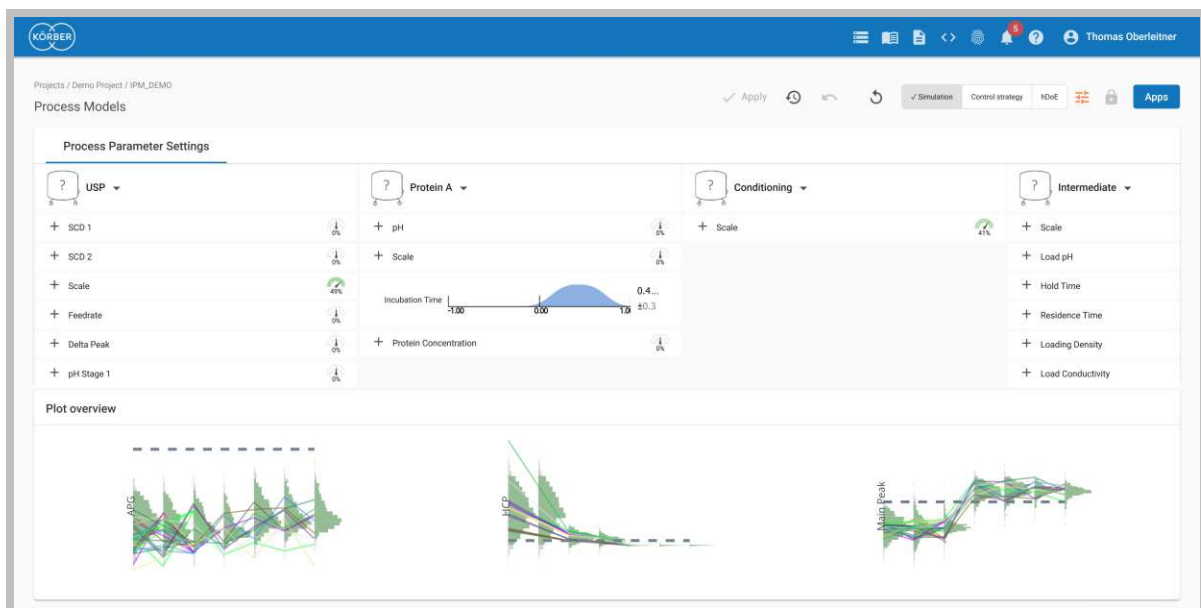


Fig. 1.4: Software implementation of the IPM in PAS-X Savvy 2023.06.

This empowers end-users, such as process engineers, to harness the complete capabilities of holistic ensemble models. As of version 2023.06, they can investigate the effects of PP changes over several UOs, conduct sensitivity analyses, and calculate the IACs based on OOS rates.

2 Dissertation Objectives

A reason for the lack of progress in the implementation of the lifecycle approach by manufacturers might be inadequate understanding of technology and science to mitigate risks [32]. The methods introduced in this thesis aim to provide manufacturers concrete solutions for problems encountered in the real world and their effectiveness is demonstrated in case and simulation studies. They supply scientific evidence for the robustness of a process in compliance with ICH and FDA recommendations. Furthermore, insights gained from these studies and the novel approaches proposed contribute to the field CMC statistics. Specifically, three contributions are proposed:

- (a) A workflow for incorporating random effects into biopharmaceutical control strategies.
- (b) A holistic Design of Experiments procedure to efficiently allocate experimental runs.
- (c) A design space optimization scheme that works in high-dimensional parameter spaces, such as the one representing an entire manufacturing process.

In recent years, advantages of holistic process models, i.e., ensemble models comprised of unit operation sub-models that represent the entire production chain of a drug substance/product, are acknowledged and encouraged by regulatory authorities [17, 33]. In the context of such end-to-end modelling approaches, the proposed methods can be seen as components as well as extended use cases of the IPM originally proposed by Zahel et al. [11]. Note that the method for incorporating random effects into process characterization as well as the design space optimization procedure can be used outside the context of an IPM. However, their full utility for regulatory filings is exploited as part of a holistic process model.

As the guidelines texts from regulatory authorities like the ICH, FDA or EMA are the common thread uniting the individual contributions of this cumulative dissertation, their connection to specific recommendations as well as the to stakeholders involved, that is, manufacturers and regulators, is illustrated in figure 2.1.

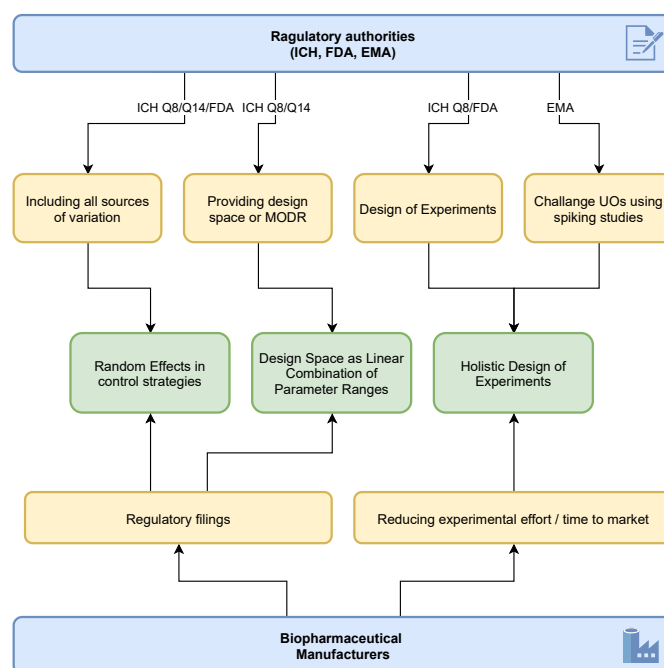


Fig. 2.1: Proposed methods in the context of regulatory authorities and manufacturers.

The consideration of different sources of variation is stressed in the ICH Q8 and FDA guidelines [7, 12]. However, the most explicit reference that motivates the inclusion of statistical random effects is given in a relatively recent text - the ICH guideline Q14 on analytical procedure development. It recommends that "... robustness should be built into the model by including relevant sources of variability related to materials, process, environment, instrumentation or other factors." [34]. These are classic examples of random effects. How to model these effects and how they can be incorporated into a control strategy is described in the the publication *Incorporating random effects in biopharmaceutical control strategies*. The full article is attached in section 7.1 and the results summarized in section 4.1.

Recommendations relevant to proposed, holistic experimental design approach (hDoE) can be found in ICH Q8 as well as the FDA guideline, where DoE is referenced as a standard method for formal experimental design [7, 12]. However, hDoE also includes a type of experiment not directly related to DoE: spiking runs. These types of experiments are suggested by the EMA to challenge downstream unit operations to determine purification capacities [35]. As described in the article *Holistic Design of Experiments Using an Integrated Process Model*, this is highly relevant for the optimization of experimental effort. Section 7.2 contains the full article and 4.2 a summary of result.

The utility of a Design Space (DS) or Method Operational Design Regions (MODR) is described in the ICH Q8 and Q14 guideline texts. A DS constitutes evidence that combinations of PPs result in CQAs that lie within acceptance limits, not to be confused with PARs that only consider the deviation of a single PP while others are kept at setpoint. For manufacturers this introduces as level of flexibility, as deviations within this space are not considered changes and do not need to be re-approved. State of the art implementations of DS calculators are generally based on grid-based algorithms and do not scale to higher dimensional problems. This is the incentive for the novel method for DS computation introduced in the revised manuscript *Identifying Design Spaces as Linear Combinations of Parameter Ranges for Biopharmaceutical Control Strategies*,

which can be found in section 7.3. The method and evaluation results are summarized in section 4.3.

3 Thesis Structure

This cumulative dissertation is comprised of three first-authored contributions. Two of them are published as articles in scientific journals while the third is, at the time of writing, a revised manuscript under review. The texts can be found in their entirety in section 7. Figure 3.1 presents an overview of the contents and corresponding publications. Section 7.1 introduces a workflow for incorporating random effects into control strategies and investigates the consequences in a case study conducted at a biopharmaceutical manufacturer (Boehringer Ingelheim). Section 7.2 introduces a method that leverages a holistic process model to optimize the allocation of experimental effort, called Holistic Design of Experiments (hDoE). The manuscript attached in section 7.3 describes a numeric optimization method to find a design space comprised of linear combinations of process parameter ranges that is compliant with ICH Q8 recommendations.

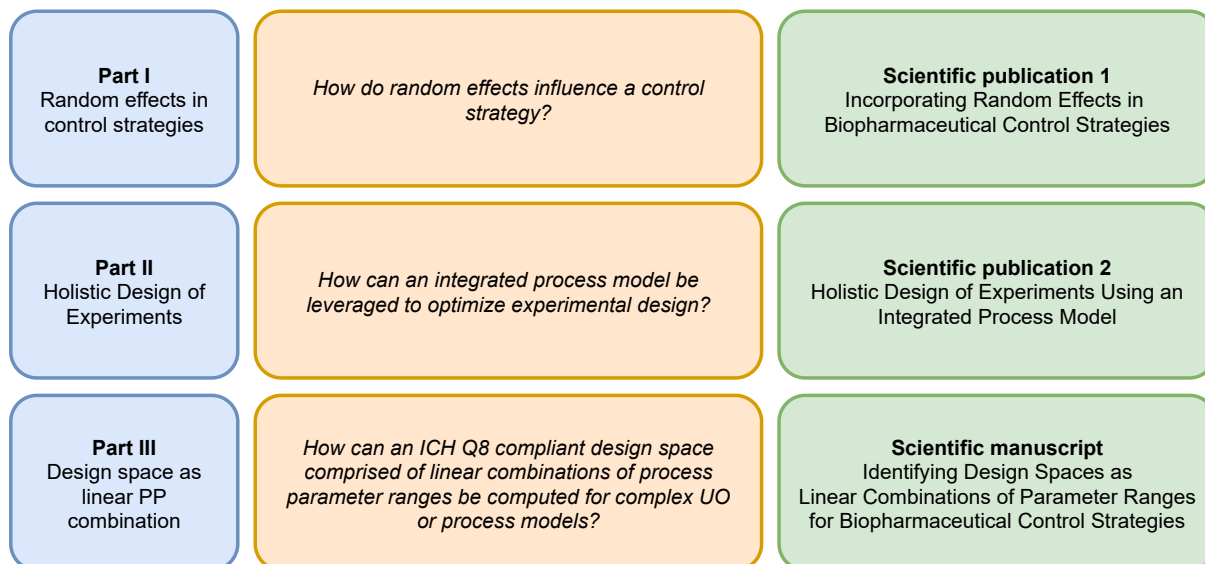


Fig. 3.1: Thesis structure with research questions and corresponding publications.

As regression models and statistical intervals are an integral part of all three parts, an overview of the exact mathematical formulations used and appropriate references are provided in section 8. This overview is published as supplementary information to the article *Incorporating Random Effects in Biopharmaceutical Control Strategies* [14].

4 Results

4.1 Incorporating Random Effects in Biopharmaceutical Control Strategies

4.1.1 Research Question

“How do random effects influence a control strategy?”

4.1.2 Problem Statement

Biopharmaceutical manufacturers are required to accurately describe the robustness of a new production process based on sound, scientific principles. An important part of a regulatory filing is the control strategy. The ICH describes the control strategy as the “...planned set of controls, derived from current product and process understanding that ensures process performance and product quality” and should contain “...at a minimum, control of the critical process parameters and material attributes” [12]. One way to define these controls is the proven acceptable range (PAR), where a single parameter is deviated while all others are kept at setpoint. A comprehensive control strategy should also consider sources of variability that are not immediately obvious: “Sources of variability that can impact product quality should be identified, appropriately understood, and subsequently controlled.” [12]. These two requirements and the connected statistical concepts are illustrated in the example for deriving a PAR in figure 4.1. The regression model’s mean prediction is shown as an orange line and the associated uncertainty interval as dashed lines around it. A PAR can be derived by finding the intersection points between the interval boundaries and the UO acceptance criteria or inter-process controls.

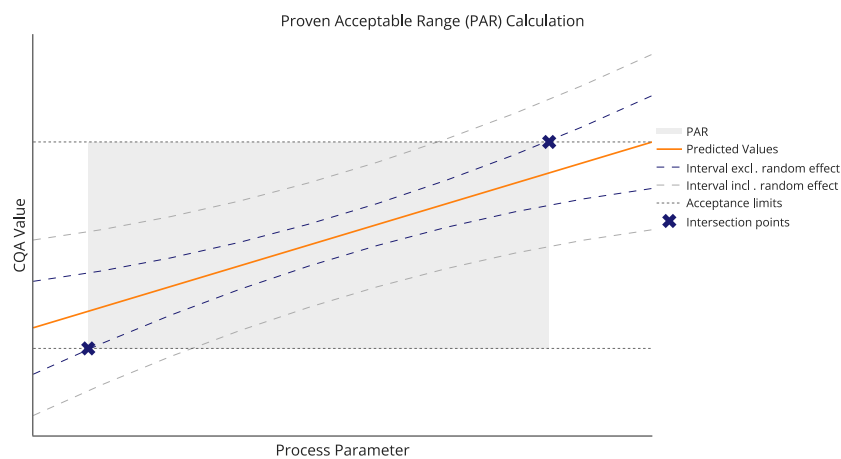


Fig. 4.1: An example for how the PAR of a process parameter can be calculated. The intersection points of a statistical interval and the CQA acceptance criteria define the lower and upper boundary. Note that the PAR that ignores the random effect is larger than the one that incorporates the random nature of the effect [14].

Uncertainty intervals are well defined for most regression models and include different sources of variation. For OLS models, the width of the interval comprises parameter as well as residual variability, i.e., the model error. However, considering the ICH recommendations, another source of variability should be considered: random effects. Random effects are well known in many domains of applied statistics and represent uncontrollable parameters that are not expected to take on a specific value when sampled repeatedly. In a biopharmaceutical context, such effects could be different raw materials, seed trains or variability introduced by different operators. For the case study presented here, several random effects are grouped into blocks labelled as week-to-week variability and modelled using linear mixed models (LMM). LMM-specific tolerance intervals proposed by Franzq et al. were implemented to quantify uncertainty around the prediction [20]. Aside from parameter uncertainty and model error, these intervals contain variability introduced by the random effect. The control strategy resulting from this approach is compared to one that ignores the random effect.

4.1.3 Results

To investigate the effect of incorporating random effects into the formulation of PARs, a Process Characterization Study (PCS) was conducted at a biopharmaceutical manufacturer (Boehringer Ingelheim). The production process contained eight typical up- and downstream UOs. Up to 22 different responses, i.e., CQAs, were modelled per UO with up to 26 effects found significant in model selection. Data was acquired from D-optimal DoE as well as One-factor-at-a-time (OFAT) experiments. The number of runs was chosen to provide adequate power to resolve interactions and detect effects within two or three standard deviations from its setpoint based on a full model, i.e., a model containing all two-level parameter interactions. The case study parameters are summarized in table 4.1.

Tab. 4.1: Parameters for process characterization study conducted at Boehringer Ingelheim [14].

UO	Power for 2 SD	Power for 3 SD	Runs DoE/OFAT	RE Levels	Significant Effects	Responses
UO 1	0.91	0.99	24/6	3	16	22
UO 2	1.00	1.00	27/24	4-5	30	20
UO 3	0.95	1.00	17/5	5	10	14
UO 4	0.71	0.93	18/3	6	10	11
UO 5	0.94	1.00	21/3	5	15	17
UO 6	0.99	1.00	42/11	3	36	13
UO 7	0.82	0.96	0/6	5	3	11
UO 8	0.89	0.98	0/13	5	8	15

Note that, at the time the experiments for this PCS were designed, the a-priori power to detect random effects was not considered. This was only verified when the data was analyzed and a variance ratio test was applied [36]. The test showed that the random effect was not significant in 18 out of the 123 models fit in the study. However, this did not change the fixed effects of the models, as p-values were checked and other good modelling practices were applied to ensure model integrity [37]. For the 13 models that failed the random effect significance test, a random effect size of zero was recorded in the results. Ideally, power to detect random effects should already be considered in the experimental design phase. Generally, methods to do so are simulation-based and available in several software packages [38, 39]. However, applicability in practice might be limited in biopharmaceutical development, as this type of power is mainly improved by increasing the number of blocks within the random effect, which can drastically increase experimental effort.

Based on data that allows for the identification of main effects and their interactions, i.e., to fit a second-order response surface model [40], a data-science-centric workflow is developed to arrive at a control strategy. The main purpose of this study was to determine the impact of including a random effect correctly, which is why two of those workflows are presented: one that includes the random effect as a fixed effect using OLS models, which can be considered the state-of-the-art in the industry (workflow A), and another workflow that uses LMMs to model the random component explicitly (workflow B). While both workflows are based on solid data-science practices, only the latter includes a proper estimate for the variance introduced by the random effect [41].

Of course, both workflows start by acquiring data, where a random effect is included as a categorical variable. For the LMM-based workflow a more detailed investigation is recommended at this stage, as multiple random effects can exert an effect on variation and they can be crossed or nested, which also changes the contribution to variance. For the OLS-based workflow, random effect blocks should be deviation-encoded to ensure that the effect's mean is at zero, i.e., that it does not bias the OLS prediction [42]. This is not necessary for LMM models as random effect blocks are handled separately from fixed effects. Correlation analysis should be performed in both cases without the block factor. As LMMs are fit by an optimization procedure as opposed to a closed-form expression in OLS, convergence criteria should be checked at this stage, such as the appropriate number of blocks and levels [43]. At this point, one can either choose to use the full model including all fixed effect interactions or to compute a more parsimonious model by employing variable selection [44–46]. While arguments can be made that variable selection should be applied cautiously [47], a smaller model is usually preferred due to its improved interpretability and expressiveness [48]. Efficient variable selection algorithms specific to LMMs are scarce at the time of writing and not commonly available in software packages, which is why a sensible workaround is to deviation-encode the random effect blocks and perform variable selection using OLS-models instead. The resulting model can then be transformed back into an LMM model. Finally, the full or parsimonious model is used to form tolerance intervals around the predicted CQAs based on the training data to capture the population distribution. For LMMs, this interval is formed using the approach recommended by Francq et al. and includes the variance introduced by random effects [20]. Figure 4.1 contains the parameter range indicated by the intersection points between interval and acceptance limits that represents the PAR and the control strategy for the analyzed process. The illustration suggests that acceptance limits are the other limiting factor when formulating control strategies in this way. Their exact impact on the process and how to use process understanding to optimize them is discussed in the next part of this dissertation that introduces Holistic Design of Experiments (hDoE). The two workflows described are summarized in figure 4.2.

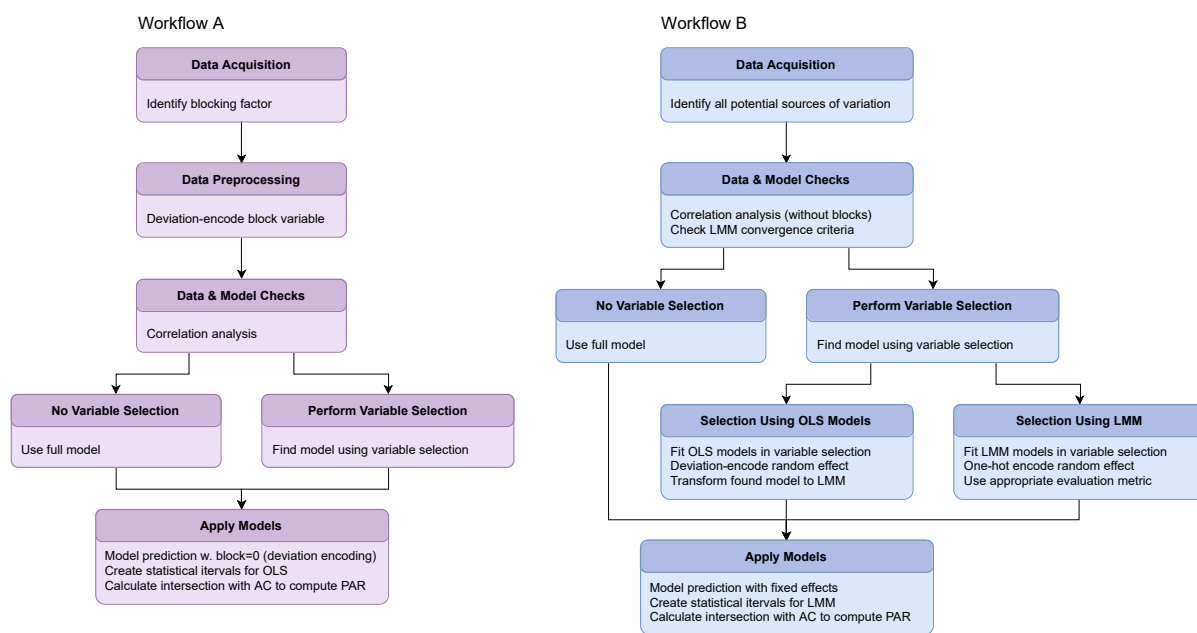


Fig. 4.2: Workflows for creating control strategies based on regression models. The left column describes an approach that uses OLS models for the estimation of PARs. A mixed-model-based workflow is summarized on the right. The differences in the steps involved are subtle but generally result in a more realistic estimation of variance and therefore a more robust control strategy [14].

Workflows A and B were applied to each unit operation in the case study and their results compared to highlight the effect of correctly including random effects in the analysis. Table 4.2 shows the ratio between LMM and OLS tolerance interval widths, i.e., the results from applying either workflow B vs. A for six CQAs common to all UOs.

Tab. 4.2: Ratio TI-width LMM/TI-width OLS for six CQAs common to all UOs. [14].

	UO 1	UO 2	UO 3	UO 4	UO 5	UO 6	UO 7	UO 8
CQA 1	1.85	3.80	2.67	3.18	3.01	1.57	1.17	1.08
CQA 2	1.62	1.65	6.15	3.36	4.59	3.37	2.38	5.15
CQA 3		1.70	4.99	3.43	4.06	2.43	2.34	1.41
CQA 4	1.85	1.64	1.77	1.12	4.38		8.95	2.84
CQA 5		1.80	4.58	1.37	8.34		2.92	3.10
CQA 6		1.67		1.25	9.02		3.46	2.98

Clearly, the intervals resulting from the LMM approach are much wider than those of the OLS-based workflow, in many cases resulting in a more conservative control strategy. To further illustrate this point, a PP was chosen randomly from the results and the PAR formulation illustrated in figure 4.3.

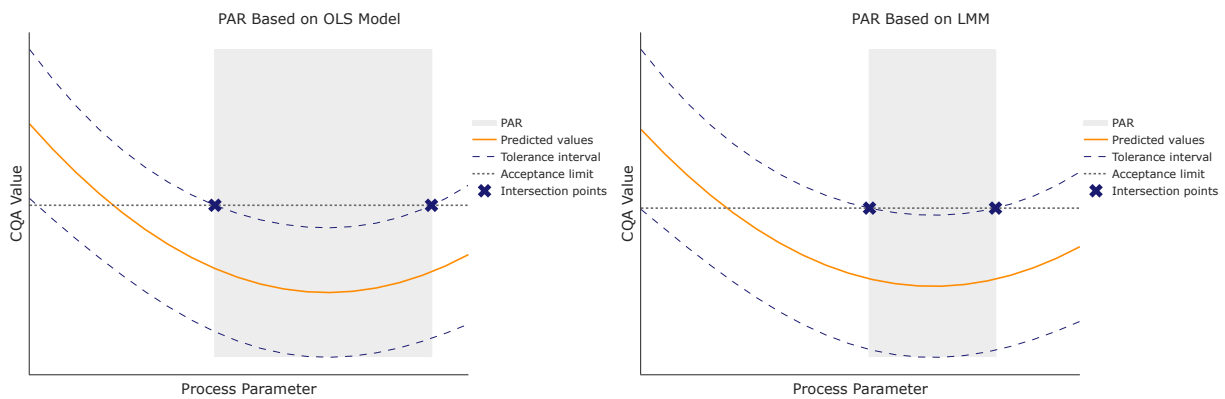


Fig. 4.3: PAR for a randomly picked parameter calculated from case study data. Due to the contribution of the random effect, the interval based on LMM variance components (right) is wider than its OLS counterpart (left). This results in smaller PAR (grey area) and a more conservative control strategy. In this example from the case study, the OLS PAR is 72% larger than the more conservative LMM PAR [14].

To highlight the magnitude of the random effect, normalized effect sizes were investigated in the models resulting from workflow B. Figure 4.4 provides an overview of the results. In half of the UOs, the median random effect was larger than the median fixed effect, resulting in a larger contribution to variance and wider tolerance intervals.

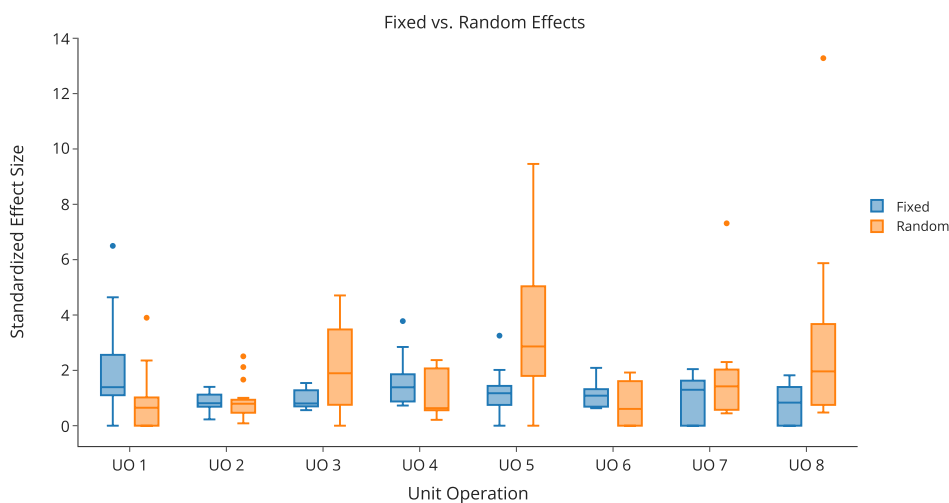


Fig. 4.4: Standardized fixed and random effect sizes are contrasted for each unit operation. A unit operation contains models for 11 – 22 CQAs and their respective fixed and random effect distributions are shown as box plots. To create comparable measures of effect size, normalized data were used to fit the models and the effects were divided by the RMSE. [14].

4.1.4 Conclusion

Random effects are oftentimes ignored in biopharmaceutical development when defining a control strategy. The results from the presented case study show that such effects can influence the PAR significantly and should be considered in any investigation of critical process parameters. To this

end, a workflow based on sound statistical methods is proposed. The workflow contains a set of tools for analysing and interpreting experimental data that is more compliant with the ICH and FDA recommendations than currently employed methods for process characterization and results in more robust processes, fewer out-of-specification events, and reduced patient risk.

4.1.5 Publication

T. Oberleitner et al. “Incorporating random effects in biopharmaceutical control strategies”. In: *AAPS Open* 9.1 (2023), pp. 1–13

4.1.6 Contribution

TO was the main contributor to the manuscript, implemented tolerance intervals for linear mixed models in Python (based on the method proposed by Franzq et al.), performed data analysis and the subsequent modeling. TZ contributed to the manuscript and was the primary advisor for statistics and data analysis methods. MK validated the manuscript and data and provided input from a manufacturing perspective. JT supervised the experiments during the process characterization study at Boehringer Ingelheim. CH provided guidance in writing of the manuscript as the PhD supervisor to TO.

4.2 Holistic Design of Experiments Using an Integrated Process Model

4.2.1 Research Question

“How can an integrated process model be leveraged to optimize experimental design?”

4.2.2 Problem Statement

Design of Experiments (DoE) methods like factorial [25], optimal [29] or definitive screening designs [49] are commonly used in process characterization and an integral part of the QbD approach [12, 50, 51]. The FDA guideline for process validation recommends to investigate the relationship between PPs and CQAs per unit operation, employing DoE combined with regression analysis to arrive at a control strategy for the process, i.e., the first stage in the validation lifecycle [7]. Integrated in an IPM, such models can be used to simulate the propagation of CQA distributions over the entire sequence of UOs, from fermentation to drug substance formulation. The IPM can not only be used to compare drug substance CQA distributions to specification limits, thus validating process robustness, but also to derive Intermediate Acceptance Criteria (IAC) or Inter-process Controls (IPC) [13]. The two steps involved in IPM setup and simulation, that is, (i) creating models based on DoE data and (ii) IPM integration, do not necessarily need to be performed independently. Instead, short, iterative cycles of experimentation and regression analysis over different unit operations can open up opportunities to radically increase efficiency while satisfying robustness requirements. This work introduces holistic Design of Experiments (hDoE), a method for systematically allocating experimental effort to unit operations.

4.2.3 Results/Method

As hDoE is mainly a tool for process design and the goal of this stage is to derive a robust control strategy, the method for deriving a PAR from section 4.1.2 is reiterated to stress the quantities and experimental runs involved. The elements affecting the definition of the PAR are the width of the tolerance interval and the acceptance limits, as illustrated in figure 4.5. Of course, effect sizes are another important factor, i.e., the slope of the line shown in the plot. However, this is a consequence of the experimental data rather than an aspect that can be controlled purely by the number and type of runs, which is why it is not directly considered in hDoE. In general, to increase the PAR, interval widths must be reduced or acceptance limits widened. The former can be achieved by reducing model uncertainty through simply fitting models on more data, ideally from DoE studies. The latter is more specific to the biopharmaceutical domain and involves spiking studies that challenge UOs with different quantities of the starting material [52, 53]. For example, in downstream UOs this typically means gaining knowledge about impurity clearance capabilities. This knowledge can then be used to update acceptance limits [13].

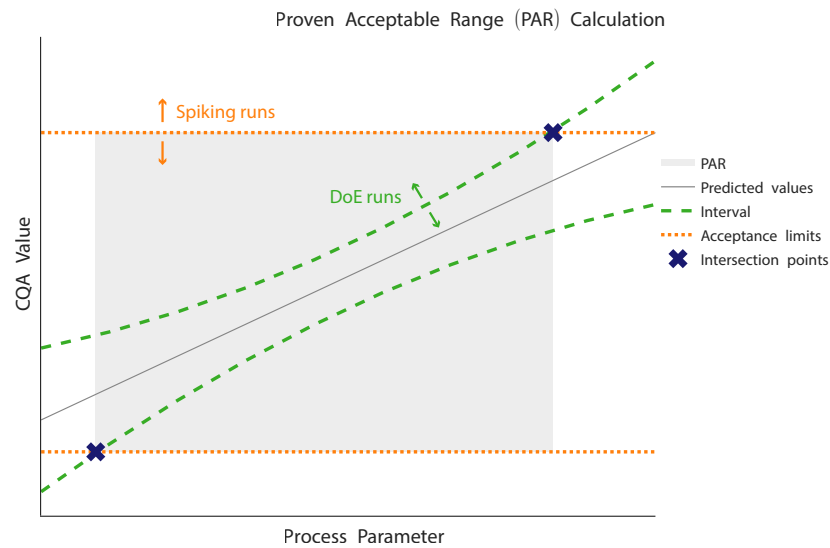


Fig. 4.5: Elements of the PAR definition affecting the control strategy and the type of experimental runs from which they are generally derived. Adapted from [15].

DoE and spiking experiments are the two run types recommended by hDoE to arrive at a robust control strategy in an efficient way. This is done as follows: the procedure starts out by fitting rough UO models to data from minimal, D-optimal designs. These models are integrated into an IPM and used to simulate CQA distributions by randomly sampling PP values [18] and deriving OOS rates in the final UO. Then, a set of experimental runs is recommended to minimize this OOS rate. Each proposed run is allocated to the specific UO that shows the largest reduction in the OOS rate. The experiments can be either DoE runs to reduce the width of the tolerance interval around the model prediction or spiking runs to widen acceptance limits, as previously described. For DoE runs, the next optimal run per UO is recommended based on standard augmentation algorithms used in optimal designs [54]. Recommendations for spiking runs are computed by deviating from the mean of the training data for the load parameter (the starting material) by different standard deviations. In practice, the starting material cannot be spiked indefinitely and limits have to be provided by a process expert. The set of recommended experiments are then conducted and the resulting data is fed back to the IPM models. This cycle is executed until the simulated OOS rate falls below the desired threshold and therefore satisfies the criteria for robustness, based on statistical evidence provided by regression models and conservative estimators for the CQA distribution in the form of tolerance intervals. Figure 4.6 provides a graphical overview of the steps involved in hDoE.

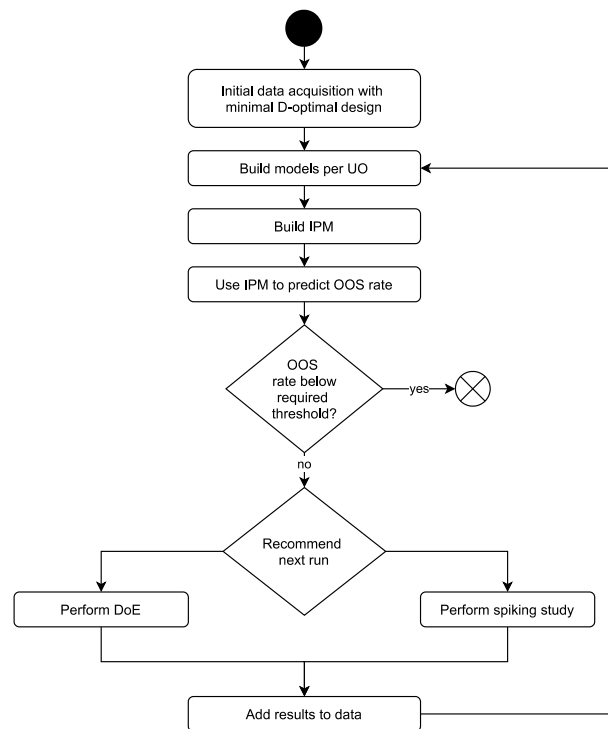


Fig. 4.6: The iterative hDoE recommender system for experimental runs [15].

As seen in the diagram, when new data is acquired it is used to set up the next iteration of the IPM. Part of this step is a variable selection procedure that finds the most suitable model based on the augmented data for the corresponding UO. DoE data, per definition, facilitates the detection of new effects and therefore variable selection could result in a new model. Spiking runs on the other hand focus on improving the load parameter estimator alone, which generally does not change the model structure. Recall that the recommender step picks the run that shows the largest improvement in the simulated OOS rate, whether it is a DoE or spiking study. Treating those types of experiment equally can be a problem, especially in later hDoE cycles where OOS-improvements might be the result of residual error or in situations where the OOS-simulation is stuck in local minima that could be overcome by exploring new factor interactions. Furthermore, run recommendations are based on the existing model structure of the IPM and therefore biased toward known effects. This can be seen as an exploration vs. exploitation problem, a concept generally known from reinforcement learning [55]: the recommender step can either exploit the current structure of the model to gain knowledge about load parameters or explore incorporating different factors and interactions into models and therefore gaining more general knowledge about the process. A similar problem is solved elegantly in the Metropolis–Hastings algorithm [26]. Essentially, the proposed decision scheme draws a random number from the interval $[0, 1]$ and only exploits the existing structure when this number exceeds the ratio $improvement_{exploit}/improvement_{explore}$, thereby promoting exploration. hDoE uses this method to overcome the problems described above and its effectiveness is shown in the simulation study results.

To validate the method and highlight applicability to a wide range of scenarios, simulation studies were conducted where hDoE was applied to different processes set up in-silico. For these processes, relations between PPs and CQAs are predefined and distorted by added noise. The effects serve as the ground truth for the characterization procedure performed with experiments recommended by hDoE. To compare results with the state-of-the-art, the simulation studies

were also performed with three predefined sets of D-optimal runs containing 24, 48, and 92 total experiments over all UOs. The simulated biopharmaceutical processes satisfy the conditions for the employed implementation of the IPM, i.e., (i) linear relationships between the the load parameter and the model response and (ii) effect heredity, i.e., the main factor of an interaction effect is always part of the model [56]. Other shared parameters of the studies are summarized in table 4.3.

Tab. 4.3: Shared parameters for simulation studies [15].

Parameter	Value
Number of UOs	4
Parameters per UO	5
CQA type	Impurity
hDoE start runs	6
Noise/std ratio for residual error	0.5 (0.9 in study D)
Number of runs recommended per cycle	1
Variable selection method	Bi-directional stepwise
p-value threshold for including effect	0.25
p-values threshold for excluding effect	0.05

These are the specific scenarios and associated ground truth effects simulated:

- (a) A typical biopharmaceutical process with common effect sizes and interactions as well as balanced load relationships.

$$\begin{aligned}
 y_{UO1} &= 8.0 + 0.7x_1 + 0.6x_2 + 0.5x_3 - 0.4x_4 + 0.9x_2^2 + 0.8x_3^2 \\
 y_{UO2} &= 3.4 + 0.5y_{UO1} + 0.5x_2 + 0.3x_3 + 0.5x_5 + 0.7x_2x_3 + 0.4x_2x_5 \\
 y_{UO3} &= 3.0 + 0.3y_{UO2} + 0.4x_1 - 0.3x_2 + 0.2x_4 - 0.2x_5 + 0.3x_1^2 - 0.7x_5^2 \\
 y_{UO4} &= 2.8 + 0.2y_{UO3} + 0.1x_1 + 0.2x_3 + 0.2x_5 + 0.3x_1x_3
 \end{aligned} \tag{4.1}$$

- (b) The same process as in (a), but with the load coefficient in UO3 set to one. This means that the UO does not affect the loaded material. It represents an unlikely scenario of a completely ineffective step in the process where additional information about the load parameter does not improve process knowledge, i.e., there is no point in performing spiking studies and hDoE is limited to allocating DoE runs. The same effect sizes were used as in (a) with the exception of UO3.

$$y_{UO3} = 3.0 + 1.0y_{UO2} + 0.4x_1 - 0.3x_2 + 0.2x_4 - 0.2x_5 + 0.3x_1^2 - 0.7x_5^2 \tag{4.2}$$

- (c) The load coefficients of all UOs are set to one in this process (for the sake of brevity, formulas are omitted). It is an even more extreme example of the effect described in (b) and is included in the study to demonstrate the worst case scenario for hDoE. Here, the entire process is ineffective regarding the start material and only DoE runs are allocated to the UOs. The purpose is to show that, at worst, the method is equivalent to the state-of-the-art in terms of experimental effort.
- (d) To highlight the effectiveness of the Probability-Ratio-Threshold (PRT) method from the Metropolis-Hastings algorithm described above, it is left out in this scenario. The recommender step simply selects the run with the largest improvement in the OOS rate.

Each simulation was repeated 100 times and the results are presented in terms of summary statistics. Figure 4.7 illustrates the mean improvement of the OOS rate over the number of experimental runs invested. It demonstrates how the experimental effort can be radically reduced when compared to SOTA methods for process characterization. As expected, scenario (c) is the only one where this does not hold true, as no information can be gained by investing spiking studies. Scenario (d) highlights the effectiveness of PRT, as omission leads to much higher variance in the results.

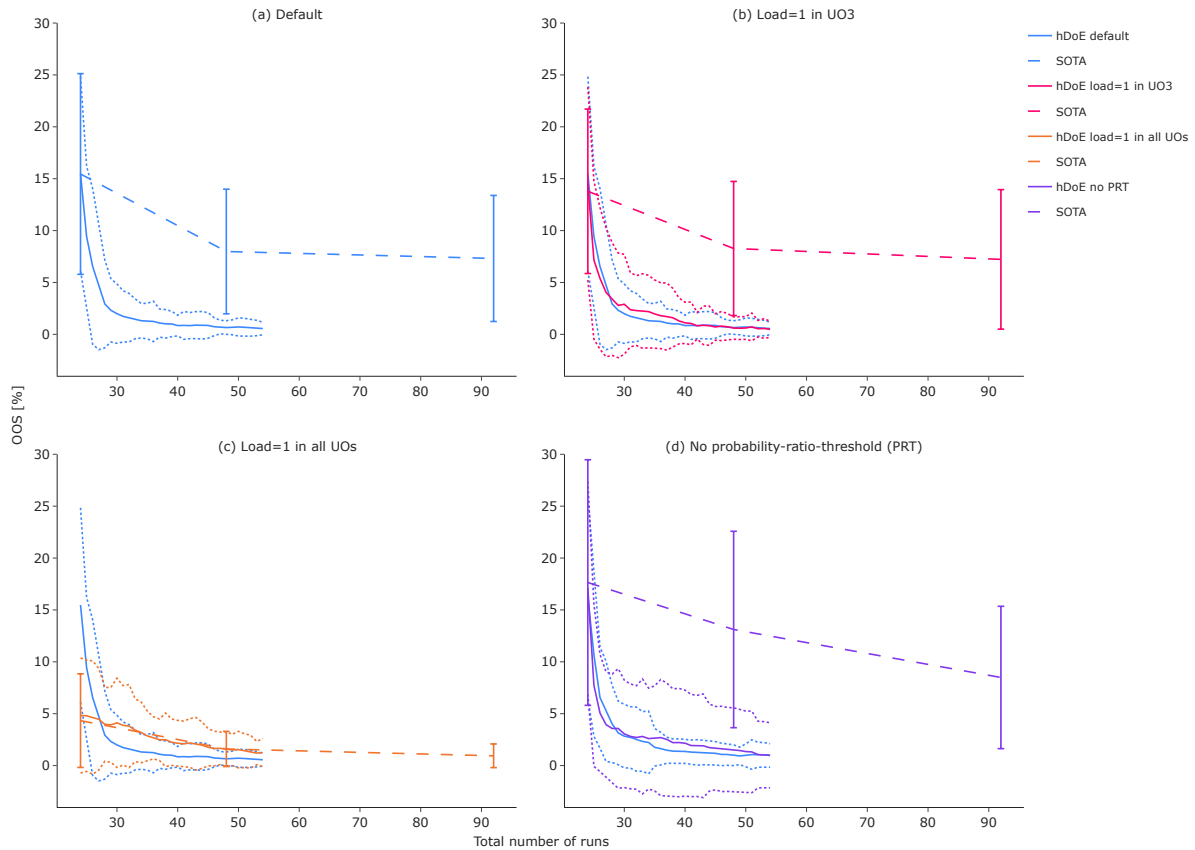


Fig. 4.7: The decrease in OOS probability over 30 hDoE steps, repeated 100 times. The solid line indicates the mean OOS probability while the dotted lines show the standard deviation in the 100 repetitions of that step. For the reference method, OOS probabilities are plotted as dashed lines and were calculated at 24, 48 and 92 total runs [15].

4.2.4 Conclusion

The IPM enables the simulation of CQA distributions based on conservative estimators such as tolerance intervals. These intervals adequately describe the uncertainty attributed to the model due to a lack of data. In hDoE iterations, this means that the interval starts out large in each UO and gets smaller as more data is added to the model. Combined with conservative extrapolation of the starting material described in the method section, this establishes evidence of a robust process, as described in the FDA guideline for process validation [7] while giving biopharmaceutical manufacturers an incentive to leverage holistic process models to reduce experimental effort and thus development cost and time-to-market.

hDoE can also be integrated easily into existing workflows for process characterization for easier adoption [57–59]. Figure 4.8 illustrates such a workflow and the specific activities affected by hDoE cycles. This suggests that hDoE transforms process characterization into a procedure that involves iterations between experimental design and evaluation, which is naturally associated with organisational overhead. However, the increased efficiency demonstrated should be a compelling reason for the adoption of such an approach.

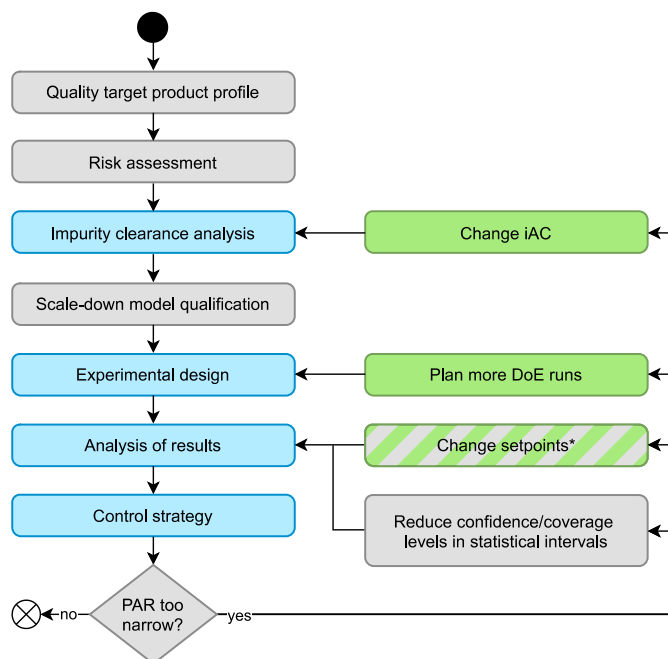


Fig. 4.8: The left column represents common steps involved in process characterization while the right column contains possible courses of action when no appropriate control strategy can be found. Green boxes are actions recommended by hDoE and steps affected by hDoE are colored blue. *Changing PP setpoints could be a viable option for process optimization in the future [15].

4.2.5 Publication

T. Oberleitner et al. “Holistic Design of Experiments Using an Integrated Process Model”. In: *Bioengineering* 9.11 (2022)

4.2.6 Contribution

TO was the main contributor to the manuscript and responsible for the methodology, implementation of simulation studies, data analysis, curation and visualization. TZ was responsible for conceptualization and contributed to the methodology and the manuscript. BP was also involved in conceptualization and contributed to data visualization. CH provided guidance in writing of the manuscript as the PhD supervisor to TO.

4.3 Identifying Design Spaces as Linear Combinations of Parameter Ranges for Biopharmaceutical Control Strategies

4.3.1 Research Question

“How can an ICH Q8 compliant design space comprised of linear combinations of process parameter ranges be computed for complex UO models or process models?”

4.3.2 Problem Statement

The ICH defines a design space (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.” [12]. Submitting a DS as part of regulatory filings can be a valuable asset for biopharmaceutical manufacturers, as deviations within this space are not considered changes to the process and do not trigger a post-approval procedure. Note that a DS is not the same as a PAR (see section I), in that it considers simultaneous, multivariate deviations of PPs rather than changing a single PP while keeping all others at setpoint. Therefore, finding a DS is a considerably harder computational problem that suffers from the curse of dimensionality [60]. Furthermore, independent PP ranges are desired for operational simplicity in biopharmaceutical manufacturing. Mathematically they can be described as linear combinations of PP ranges that satisfy certain acceptance criteria, or as hyper-rectangles when considered geometrically. Figure 4.9 highlights the difference between a linear and non-linear DS description.

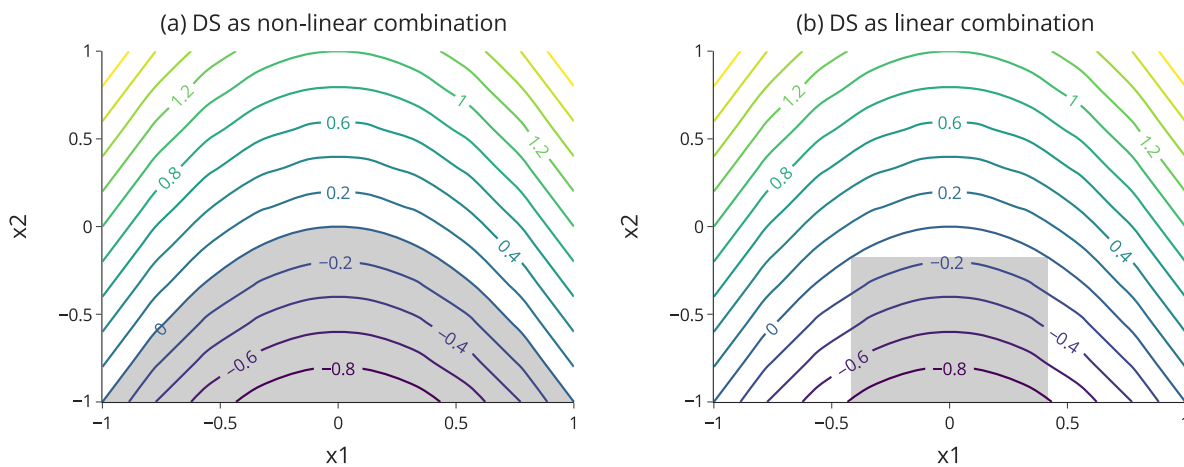


Fig. 4.9: The design space for the function $f(x_1, x_2) = x_1^2 + x_2$, where $f(x_1, x_2) \leq 0$, shown in the contour plot as a non-linear (a) and linear (b) combination of input parameters x_1 and x_2 [61].

This problem combines two subproblems: (i) finding the DS by evaluating points within the parameter space and (ii) fitting the largest possible, axis-aligned hyper-rectangle within that space [62]. The third part of this dissertation introduces a method that incorporates both into an optimization problem that can be solved efficiently, even for a large number of PPs.

4.3.3 Results/Method

The initial part of this contribution concerns the formulation of the tasks described above as an optimization problem subject to inequality constraints. With the formulation in place, one can

choose from a selection of well-studied algorithms to perform the actual optimization procedure. Here, the objective function to be maximized is simply the hyper-rectangle volume:

$$\begin{aligned} & \text{maximize} && \prod_{i=1}^p (x_{p+i} - x_i)w_i, && x \in \mathbb{R}^{2p} \\ & \text{subject to} && c_1(x), \dots, c_m(x) \geq 0 \end{aligned} \quad (4.3)$$

In this optimization problem, x is the vector containing the lower DS range per parameter in its first p elements and the upper DS range in the second half. Thus, the range of the parameter at index i can be extracted from the elements x_i and x_{p+i} . The volume also contains a weighting term w_i to stress the importance of certain parameters over others. This adds to the flexibility of the tool and gives users some control over how the result is calculated. The inequality constraints c_j contain most of complexity involved in finding an appropriate DS, such as ensuring that the setpoint is included in the result and that optimization is performed within the parameter space, as not all optimization algorithms support bounded minimization. Most importantly, the inequality constraints evaluate the corner points of the hypercube. This is done by using the underlying regression model to predict a CQA and check whether the tolerance interval boundaries around this prediction lie within acceptance limits.

$$\begin{aligned} c_l(x) &= \hat{t}_l(x\hat{\beta}) - a_l, & i = 1, 2, \dots, 2^p \\ c_u(x) &= a_u - \hat{t}_u(x\hat{\beta}), & i = 1, 2, \dots, 2^p \end{aligned} \quad (4.4)$$

Here, $x\hat{\beta}$ is the regression model prediction, $\hat{t}_l(x)$ and $\hat{t}_u(x)$ the functions returning the lower and upper tolerance interval boundary respectively, and a_l and a_u the corresponding acceptance limits. Note that, depending on the model used, calculating a tolerance interval can be a computationally expensive process that is performed on a large number of points in each iteration of the optimizer. To resolve this bottleneck, a second-order polynomial approximation of the interval is calculated beforehand and boundary checks can be performed with simple vector multiplication.

$$\begin{aligned} \hat{t}_i(\hat{y}) &= \hat{\beta}_0 + \hat{\beta}_1\hat{y} + \hat{\beta}_2\hat{y}^2 \\ \hat{t}_l(\hat{y}) &= \hat{y} - \hat{t}_i(\hat{y}) \\ \hat{t}_u(\hat{y}) &= \hat{y} + \hat{t}_i(\hat{y}) \end{aligned} \quad (4.5)$$

Another issue related to regression models are categorical factors. They are commonly used in UO models to include non-continuous factors such as discrete machine settings or scale. Their presence in a model generally limits the choice of optimizer to a more complex class of algorithms, such as mixed-integer programming, branch-and-bound, genetic algorithms, etc. [63–66]. However, given the potential involvement of a considerable number of factors in these models and the priority placed on computational efficiency, an alternative method is suggested. This approach involves treating categorical factors as if they were continuous variables, and is illustrated in figure 4.10. Individual levels of a deviation-encoded, categorical factor are simply offsets from the mean prediction of a regression model. This suggests that there is a range r_c between the minimum and maximum offset in which the factor can be treated as continuous for the purpose of optimization. After a valid range within r_c is found, the result is transformed back into a valid subset of categorical levels by "dropping" to the next valid level. For example, if the upper boundary of the valid range in figure 4.10 is found somewhere between the green and red dashed line, the level suggested by the red line is included in the results while the level

indicated by the green line is not. This treatment of categorical factors facilitates the use of the simple and efficient optimization algorithms used here.

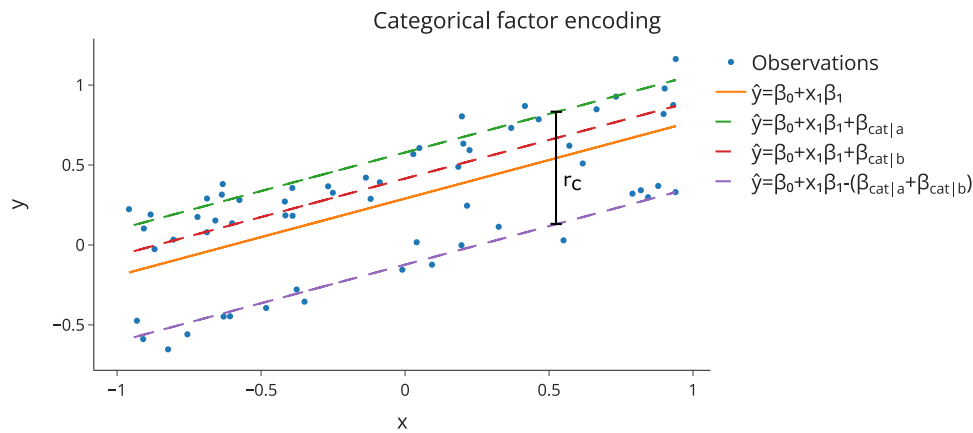


Fig. 4.10: The offset to the mean prediction of a regression model introduced by a deviation-encoded, categorical factor [61].

COBYLA is chosen as the main optimizer for the problem for its support of inequality constraints and resilience against local minima [27]. A second, gradient based SLSQP optimizer can be used to fine tune the results from the first pass [28]. The actual implementation of tolerance intervals can be used in this step instead of the approximation to further increase accuracy.

The described optimization procedure, including the tolerance interval approximation and categorical factor handling, is summarized in figure 4.11.

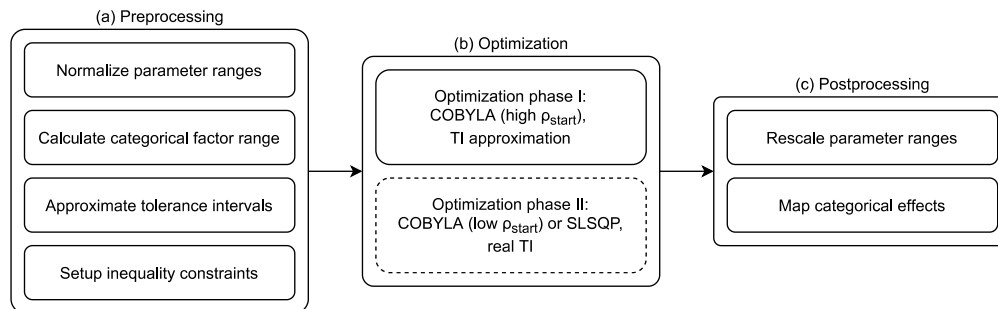


Fig. 4.11: Overview of the steps involved in the proposed design space calculation scheme [61].

The design space optimization method was validated in terms of effectiveness, measured as the maximum DS volume of the result, as well as efficiency, represented as the computing time. For a bivariate DS, effectiveness is shown by comparing the result from the optimization algorithm to a mathematically derived ground truth. The largest possible rectangle is inscribed in a unit circle, thus its area is known to be 2. Due to the long computation time involved in its derivation, no ground truth could be set up mathematically for high-dimensional problems. Instead, the DS volume and computation time of the optimizer-based method was only compared to the grid-based brute-force algorithm.

As shown in figure 4.12, effectiveness is generally improved in comparison to grid-based approaches that discretize the parameter space, as the resulting design space is not constrained to the finite set of points in the grid. Results for the higher dimensional DS in figure 4.12 (b) highlight the flexibility introduced by the weighting vector. The original DS returned by the optimizer (orange line) exhibits a relatively small range for factor x_6 . By increasing the weights for this factor and rerunning the optimizer, a more balanced, albeit smaller DS is returned (orange, dotted line), which might be desirable for operational flexibility. This example also illustrates the effect of the second optimization pass that slightly increases the DS (green line).

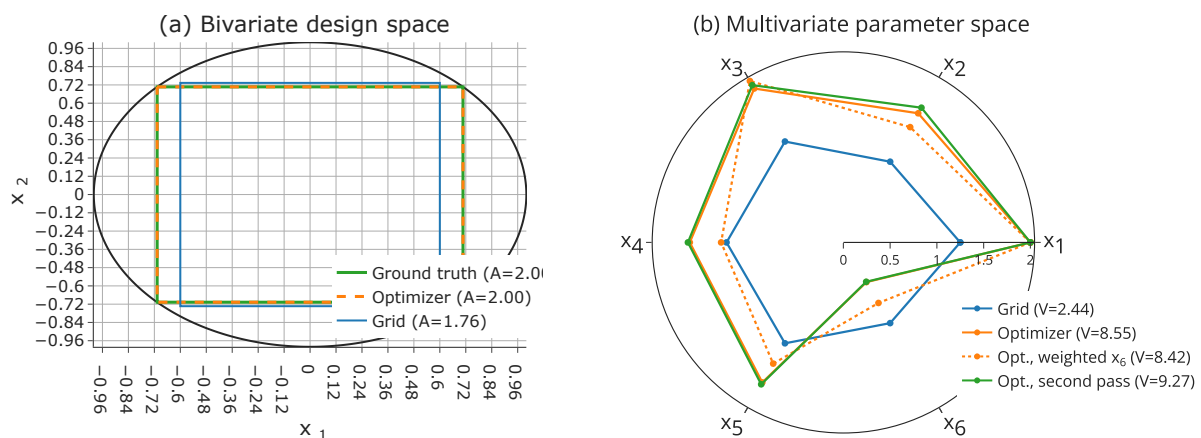


Fig. 4.12: (a) Bivariate parameter space with an upper tolerance interval boundary shown as the contour. The design spaces found by the optimizer and the grid method are marked as rectangles. (b) A multivariate parameter space illustrated as a spider plot with ranges per parameter shown in the axes and DS plotted as polygons [61].

Arguably the biggest advantage of the proposed, optimizer-based method is how computation time scales over the number of investigated factors. As the optimizer is able to pass over large parts of the parameter space, as opposed to performing a full scan of a discretized version of it, computing time for 10 factors+ is improved by several orders of magnitude, as indicated by the log-scaled figure 4.13. As the parameter space to be scanned grows exponentially over the number of parameters, so does runtime for grid-based methods (blue line). Note that these results were extrapolated from seven factors onwards, as it was no longer feasible to calculate them on a standard computer (blue, dotted line). This is very different from the optimizer results that show much better scaling (orange line). Note that performing a second optimization pass or using the actual tolerance interval calculation instead of the approximation only adds a fixed offset to computation time (green line).

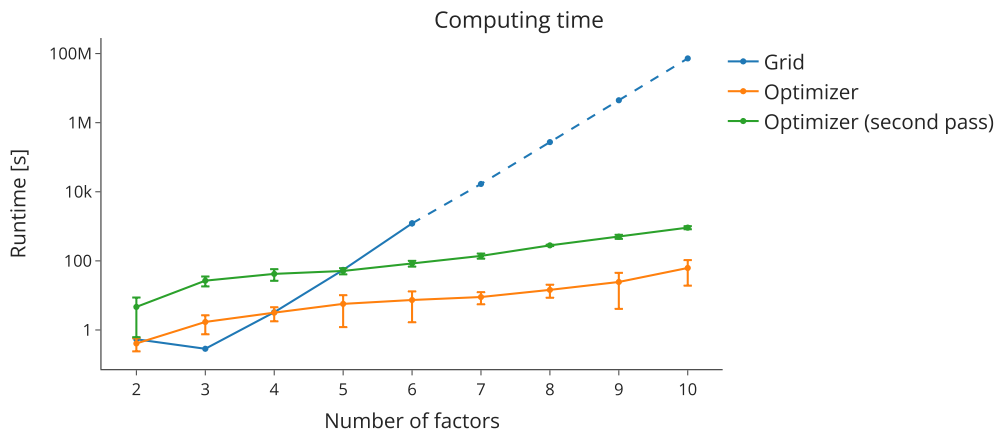


Fig. 4.13: Performance comparison of optimizer and grid-based DS optimization. [61].

4.3.4 Conclusion

Including a DS in regulatory filings aids to operational flexibility for biopharmaceutical manufacturers, as deviations within this space are not considered changes. Appendix 2 of the ICH Q8 guideline contains examples for how to define and represent a DS graphically [12]. While this is achieved easily in a bivariate parameter space, the problem of representation as well as computation grows exponentially with the number of parameters involved. Thus, numeric methods are required to solve this problem. Here, a novel method for finding and exploring a DS in higher dimensions is introduced that satisfies the requirements in the biopharmaceutical domain, i.e., independent parameter ranges and conservative evaluation of the PP space using tolerance intervals. Improved computing time compared to SOTA grid-based approaches as well as the PP weighting further add to the utility by facilitating efficient DS exploration.

4.3.5 Manuscript

T. Oberleitner et al. “Identifying Design Spaces as Linear Combinations of Parameter Ranges for Biopharmaceutical Control Strategies (under review)”. In: *arXiv preprint arXiv:2304.14666* (2023)

4.3.6 Contribution

TO was the main contributor to the manuscript and responsible for the methodology, implementation, validation, data analysis, curation and visualization. TO and TZ were responsible for conceptualization. TZ contributed to the manuscript and to implementation. CH provided guidance in writing of the manuscript as the PhD supervisor to TO.

5 Conclusion

This cumulative dissertation contributes to the field of CMC statistics by introducing tools that facilitate the provision of statistical evidence for robustness, aligning with recommendations for biopharmaceutical process validation by regulatory agencies. The rough descriptions of statistical approaches provided by such agencies are supplemented by concrete methods, which are validated and evaluated in case and simulation studies.

The first publication presents a workflow to incorporate random effects into biopharmaceutical process control strategies. While model predictions might not be affected by these effects, statistical intervals that consider the random variance components can increase drastically. This changes PARs derived from the interval boundaries and, in turn, the control strategy. A case study was conducted that highlights this effect. Results show that the investigated random effect was not only significant but in many cases stronger than the average fixed effect. For manufacturers, this implies that more sources of variation need to be considered when gathering and evaluating data, even if those sources cannot be controlled. The presented workflow for data gathering and analysis offers manufacturers recommendations on how to effectively accomplish this using statistically sound methods.

The second contribution introduces a recommender system for experimental runs, called Holistic Design of Experiments (hDoE). Based on minimal, initial data and an IPM, either DoE or spiking runs are recommended for specific UOs to minimize the Out-of-specification (OOS) rate in the drug substance formulation step. This emphasises process robustness, represented by the OOS rate, early on in development. Compared to state-of-the-art approaches that characterize UOs individually using designs with a fixed number of runs, robustness can be proven more efficiently with reduced experimental effort. Furthermore, due to conservative statistical methods such as the sampling from tolerance intervals and load extrapolation described in section 1.3.2, this does not compromise patient safety and complies with recommendations of ICH, FDA and EMA.

An efficient method for the computation of Design Spaces (DSs) is described in the third part of this dissertation. A DS defines how much PPs can deviate multivariately while certain CQA criteria are still met. Manufacturers can provide a DS definition as part of regulatory filings, whereupon changes of PPs within this space do not trigger a post-approval procedure. Here, this is defined as the range of a regression model parameter that results in predicted CQA values with tolerance interval boundaries that lie within acceptance limits. The multivariate space spanned by those ranges, or rather its volume, is maximized and yields the DS. An additional requirement for operational simplicity in biopharmaceutical manufacturing is, that PP ranges should not be conditioned on each other, i.e., they should be linearly independent or a hyper-rectangle within the parameter space. The proposed method finds a DS that satisfies those criteria by formulating a minimization problem subject to inequality constraints, which is then solved by a COBYLA optimizer. This procedure finds a DS in a much more computationally efficient way compared to grid-based methods currently found in statistical software, and thus enables manufacturers to define and explore a DS efficiently in higher dimensions.

Results from this applied dissertation might serve as an incentive for biopharmaceutical manufacturers to leverage a wider range of statistical methods in process development and validation. As illustrated by the results, providing statistical evidence for process robustness in

light of patient safety and reducing experimental effort, development cost and time-to-market need not be opposing goals.

6 Bibliography

- [1] ISPE. *Pharma 4.0 Operating Model*. 2023. URL: <https://ispe.org/initiatives/pharma-4.0> (visited on 08/10/2023).
- [2] I. GMDP. “The rules governing medicinal products in the European Union”. In: *Volume 4* (2014), pp. 2022–08.
- [3] U. CDER. “Current Good Manufacturing Practice (CGMP) Regulations”. In: *FDA: Silver Spring, MD, USA* (2018), p. 5.
- [4] I. E. E. W. Group et al. “ICH harmonised tripartite guideline: statistical principles for clinical trials”. In: *Stat Med* 18.15 (1999), pp. 1905–1942.
- [5] R. K. Burdick, D. J. LeBlond, L. B. Pfahler, J. Quiroz, L. Sidor, K. Vukovinsky, and L. Zhang. *Statistical applications for chemistry, manufacturing and controls (CMC) in the pharmaceutical industry*. Vol. 10. Springer, 2017.
- [6] L. Zhang, M. Kuhn, I. Peers, and S. Altan. *Nonclinical statistics for pharmaceutical and biotechnology industries*. Springer, 2016.
- [7] FDA. *Process Validation: General Principles and Practices*. 2011.
- [8] F. Tao, B. Xiao, Q. Qi, J. Cheng, and P. Ji. “Digital twin modeling”. In: *Journal of Manufacturing Systems* 64 (2022), pp. 372–389.
- [9] Y. Chen, O. Yang, C. Sampat, P. Bhalode, R. Ramachandran, and M. Ierapetritou. “Digital twins in pharmaceutical and biopharmaceutical manufacturing: a literature review”. In: *Processes* 8.9 (2020), p. 1088.
- [10] M. Sokolov, M. von Stosch, H. Narayanan, F. Feidl, and A. Butté. “Hybrid modeling—a key enabler towards realizing digital twins in biopharma?” In: *Current Opinion in Chemical Engineering* 34 (2021), p. 100715.
- [11] T. Zahel, S. Hauer, E. M. Mueller, P. Murphy, S. Abad, E. Vasilieva, D. Maurer, C. Brocard, D. Reinisch, P. Sagmeister, et al. “Integrated process modeling - a process validation life cycle companion”. In: *Bioengineering* 4.4 (2017), p. 86.
- [12] ICH. *ICH guideline Q8 (R2) on pharmaceutical development*. 2017.
- [13] L. Marschall, C. Taylor, T. Zahel, M. Kunzelmann, A. Wiedenmann, B. Presser, J. Studts, and C. Herwig. “Specification-driven acceptance criteria for validation of biopharmaceutical processes”. In: *Frontiers in Bioengineering and Biotechnology* (2022), p. 1768.
- [14] T. Oberleitner, T. Zahel, M. Kunzelmann, J. Thoma, and C. Herwig. “Incorporating random effects in biopharmaceutical control strategies”. In: *AAPS Open* 9.1 (2023), pp. 1–13.
- [15] T. Oberleitner, T. Zahel, B. Pretzner, and C. Herwig. “Holistic Design of Experiments Using an Integrated Process Model”. In: *Bioengineering* 9.11 (2022).
- [16] C. Taylor, B. Pretzner, T. Zahel, and C. Herwig. “Architectural and Technological Improvements to Integrated Bioprocess Models towards Real-Time Applications”. In: *Bioengineering* 9.10 (2022), p. 534.

- [17] J. Smiatek, A. Jung, and E. Bluhmki. “Towards a Digital Bioprocess Replica: Computational Approaches in Biopharmaceutical Development and Manufacturing”. In: *Trends in Biotechnology* 38.10 (2020). Special Issue: Therapeutic Biomanufacturing, pp. 1141–1153. ISSN: 0167-7799. DOI: <https://doi.org/10.1016/j.tibtech.2020.05.008>. URL: <https://www.sciencedirect.com/science/article/pii/S0167779920301426>.
- [18] C. Z. Mooney. *Monte carlo simulation*. 116. Sage, 1997.
- [19] D. C. Montgomery, E. A. Peck, and G. G. Vining. *Introduction to linear regression analysis*. John Wiley & Sons, 2021.
- [20] B. G. Francq, D. Lin, and W. Hoyer. “Confidence, prediction, and tolerance in linear mixed models”. In: *Statistics in medicine* 38.30 (2019), pp. 5603–5622.
- [21] H. Brown and R. Prescott. *Applied mixed models in medicine*. John Wiley & Sons, 2015.
- [22] M. J. Lindstrom and D. M. Bates. “Newton—Raphson and EM algorithms for linear mixed-effects models for repeated-measures data”. In: *Journal of the American Statistical Association* 83.404 (1988), pp. 1014–1022.
- [23] R. Wolfinger, R. Tobias, and J. Sall. “Computing Gaussian likelihoods and their derivatives for general linear mixed models”. In: *SIAM Journal on Scientific Computing* 15.6 (1994), pp. 1294–1310.
- [24] N. Laird, N. Lange, and D. Stram. “Maximum likelihood computations with repeated measures: application of the EM algorithm”. In: *Journal of the American Statistical Association* 82.397 (1987), pp. 97–105.
- [25] D. C. Montgomery. *Design and Analysis of Experiments*. 2017.
- [26] W. K. Hastings. “Monte Carlo sampling methods using Markov chains and their applications”. In: (1970).
- [27] M. J. Powell. *A direct search optimization method that models the objective and constraint functions by linear interpolation*. Springer, 1994.
- [28] D. Kraft. “A software package for sequential quadratic programming”. In: *Forschungsbericht-Deutsche Forschungs- und Versuchsanstalt für Luft- und Raumfahrt* (1988).
- [29] R. T. Johnson, D. C. Montgomery, and B. A. Jones. “An expository paper on optimal design”. In: *Quality Engineering* 23.3 (2011), pp. 287–301.
- [30] R. H. Myers, D. C. Montgomery, and C. M. Anderson-Cook. *Response surface methodology: process and product optimization using designed experiments*. John Wiley & Sons, 2016.
- [31] K. P. A. GmbH. *PAS-X Savvy*. 2023. URL: <http://pasx-savvy.koerber-pharma.com/> (visited on 07/21/2023).
- [32] P. Hal Baseman. *Process Validation Lifecycle Approach: A Return to Science*. PDA & ISPE Joint New England Chapter, Process Validation Event. 2015.
- [33] ICH. *ICH guideline Q13 on continuous manufacturing of drug substances and drug products*. 2023.
- [34] ICH. *ICH guideline Q14 on analytical procedure development*. 2022.
- [35] C. for Medicinal Products for Human Use. “Guideline on process validation for the manufacture of biotechnology-derived active substances and data to be provided in the regulatory submission”. In: *EMA: London, UK* (2016).

- [36] S. Nakagawa and H. Schielzeth. “A general and simple method for obtaining R^2 from generalized linear mixed-effects models”. In: *Methods in ecology and evolution* 4.2 (2013), pp. 133–142.
- [37] C. Chatfield. “Model uncertainty, data mining and statistical inference”. In: *Journal of the Royal Statistical Society Series A: Statistics in Society* 158.3 (1995), pp. 419–444.
- [38] L. Kumle, M. L.-H. Vö, and D. Draschkow. “Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R”. In: *Behavior research methods* 53.6 (2021), pp. 2528–2543.
- [39] J. S. D. LLC. “Design of Experiments Guide”. In: *Cary, NC: JMP Statistical Discovery LLC* (2022).
- [40] A. I. Khuri and S. Mukhopadhyay. “Response surface methodology”. In: *Wiley Interdisciplinary Reviews: Computational Statistics* 2.2 (2010), pp. 128–149.
- [41] B. Govaerts, B. Francq, R. Marion, M. Martin, M. Thiel, et al. “The essentials on linear regression, ANOVA, general linear and linear mixed models for the chemist”. In: *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering* (2020), pp. 431–463.
- [42] U. S. C. Group. *R Library Contrast Coding Systems for categorical variables*. <https://stats.oarc.ucla.edu/r/library/r-library-contrast-coding-systems-for-categorical-variables/>. Accessed: 2023-07-25. 2021.
- [43] D. Bates, M. Mächler, B. Bolker, and S. Walker. “Fitting linear mixed-effects models using lme4”. In: *arXiv preprint arXiv:1406.5823* (2014).
- [44] P. Peduzzi, R. Hardy, and T. R. Holford. “A stepwise variable selection procedure for nonlinear regression models”. In: *Biometrics* (1980), pp. 511–516.
- [45] G. M. Furnival and R. W. Wilson. “Regressions by leaps and bounds”. In: *Technometrics* 42.1 (2000), pp. 69–79.
- [46] G. Heinze, C. Wallisch, and D. Dunkler. “Variable selection—a review and recommendations for the practicing statistician”. In: *Biometrical journal* 60.3 (2018), pp. 431–449.
- [47] G. Heinze and D. Dunkler. “Five myths about variable selection”. In: *Transplant International* 30.1 (2017), pp. 6–10.
- [48] M. Z. I. Chowdhury and T. C. Turin. “Variable selection strategies and its importance in clinical prediction modelling”. In: *Family medicine and community health* 8.1 (2020).
- [49] B. Jones and C. J. Nachtsheim. “A class of three-level designs for definitive screening in the presence of second-order effects”. In: *Journal of Quality Technology* 43.1 (2011), pp. 1–15.
- [50] ICH. “ICH guideline Q9 on quality risk management”. In: (2005).
- [51] ICH. “ICH guideline Q10 on pharmaceutical quality system”. In: (2008).
- [52] A. Darling. “Considerations in performing virus spiking experiments and process validation studies.” In: *Developments in biological standardization* 81 (1993), pp. 221–229.
- [53] A. A. Shukla, C. Jiang, J. Ma, M. Rubacha, L. Flansburg, and S. S. Lee. “Demonstration of robust host cell protein clearance in biopharmaceutical downstream processes”. In: *Biotechnology progress* 24.3 (2008), pp. 615–622.
- [54] R. K. Meyer and C. J. Nachtsheim. “The coordinate-exchange algorithm for constructing exact optimal experimental designs”. In: *Technometrics* 37.1 (1995), pp. 60–69.
- [55] R. S. Sutton and A. G. Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

- [56] N. H. Choi, W. Li, and J. Zhu. “Variable selection with the strong heredity constraint and its oracle property”. In: *Journal of the American Statistical Association* 105.489 (2010), pp. 354–364.
- [57] ISPE. *Successful Process Characterization – A How-to-Guide in 7 steps*. 2020. URL: <https://ispe.org/pharmaceutical-engineering/ispeak/successful-process-characterization-how-guide-7-steps> (visited on 08/16/2023).
- [58] B. Technologies. *Process Characterization/Validation*. 2021. URL: <https://www.biolyotech.com/index.php/live-bacterial-product-services/process-characterization-validation/> (visited on 08/16/2023).
- [59] B. International. *Process Characterization Essentials: Process Understanding and Health Authorities Guidance*. 2017. URL: <https://www.biopharminternational.com/view/process-characterization-essentials-process-understanding-and-health-authorities-guidance> (visited on 08/16/2023).
- [60] R. Bellman. “Dynamic programming”. In: *Science* 153.3731 (1966), pp. 34–37.
- [61] T. Oberleitner, T. Zahel, and C. Herwig. “Identifying Design Spaces as Linear Combinations of Parameter Ranges for Biopharmaceutical Control Strategies (under review)”. In: *arXiv preprint arXiv:2304.14666* (2023).
- [62] K. Daniels, V. Milenkovic, and D. Roth. “Finding the largest area axis-parallel rectangle in a polygon”. In: *Computational Geometry* 7.1-2 (1997), pp. 125–148.
- [63] M. Munoz Zuniga and D. Sinoquet. “Global optimization for mixed categorical-continuous variables based on Gaussian process models with a randomized categorical space exploration step”. In: *INFOR: Information Systems and Operational Research* 58.2 (2020), pp. 310–341.
- [64] C. Vanaret. “A global method for mixed categorical optimization with catalogs”. In: *arXiv preprint arXiv:2104.03652* (2021).
- [65] K. Badran and P. Rockett. “Integrating categorical variables with multiobjective genetic programming for classifier construction”. In: *European Conference on Genetic Programming*. Springer. 2008, pp. 301–311.
- [66] P. Saves, N. Bartoli, Y. Diouane, T. Lefebvre, J. Morlier, C. David, E. Nguyen Van, and S. Defoort. “Bayesian optimization for mixed variables using an adaptive dimension reduction process: applications to aircraft design”. In: *AIAA SCITECH 2022 Forum*. 2022, p. 0082.

7 Appendix A

7.1 Scientific publication I

RESEARCH

Open Access



Incorporating random effects in biopharmaceutical control strategies

Thomas Oberleitner¹ , Thomas Zahel², Marco Kunzelmann³, Judith Thoma³ and Christoph Herwig^{4*}

Abstract

Objective Random effects are often neglected when defining the control strategy for a biopharmaceutical process. In this article, we present a case study that highlights the importance of considering the variance introduced by random effects in the calculation of proven acceptable ranges (PAR), which form the basis of the control strategy.

Methods Linear mixed models were used to model relations between process parameters and critical quality attributes in a set of unit operations that comprises a typical biopharmaceutical manufacturing process. Fitting such models yields estimates of fixed and random effect sizes as well as random and residual variance components. To form PARs, tolerance intervals specific to mixed models were applied that incorporate the random effect contribution to variance.

Results We compared standardized fixed and random effect sizes for each unit operation and CQA. The results show that the investigated random effect is not only significant but in some unit operations even larger than the average fixed effect. A comparison between ordinary least squares and mixed models tolerance intervals shows that neglecting the contribution of the random effect can result in PARs that are too optimistic.

Conclusions Uncontrollable effects such as week-to-week variability play a major role in process variability and can be modelled as a random effect. Following a workflow such as the one suggested in this article, random effects can be incorporated into a statistically sound control strategy, leading to lowered out of specification results and reduced patient risk.

Keywords Biopharmaceutical manufacturing, Process validation, Process characterization study, Random effects, Mixed-effects model, Likelihood model

Introduction

Biopharmaceutical manufacturers have a regulatory need to accurately describe production processes and to underpin design choices and control strategies with reports based on sound science. The registration application for new drug substances and their corresponding

products includes a detailed description of the manufacturing process and a justification for the proposed control strategy (ICH, 2011; ICH, 2008). The ICH defines a control strategy as follows:

A planned set of controls, derived from current product and process understanding, that assures process performance and product quality. The controls can include parameters and attributes related to drug substance and drug product materials and components, facility and equipment operating conditions, in-process controls, finished product specifications, and the associated methods and frequency of monitoring and control.

*Correspondence:

Christoph Herwig
christoph.herwig@tuwien.ac.at

¹ Competence Center CHASE GmbH, Vienna, Austria

² Körber Pharma Austria GmbH, PAS-X Savvy, Vienna, Austria

³ Boehringer Ingelheim Pharma GmbH & Co. KG, Analytical Dev.

Biologicals, Biberach an der Riss, Germany

⁴ TU WIEN Research Area Biochemical Engineering, Vienna, Austria



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

The development of this control strategy, or *process characterization*, constitutes a major part in the first stage of the FDAs process validation guideline (FDA, 2011) and is often the most elaborate and critical phase of development for a new drug product. Ideally, this process should yield detailed knowledge about the individual parts of the process, i.e., critical quality attributes (CQA), impact of process parameters (PP), and sources of variability. The level of understanding of the product and its production process also affects the regulatory process, as stated in the ICH Q8 guideline (ICH, 2017):

A greater understanding of the product and its manufacturing process can create a basis for more flexible regulatory approaches. The degree of regulatory flexibility is predicated on the level of relevant scientific knowledge provided in the registration application.

The FDA's 2011 guide defines process validation as “the collection and evaluation of data” over the life cycle of the product, from product development to commercial production, in order to establish scientific evidence that “a process is capable of consistently delivering quality product.” Part of this is detecting and understanding different sources of variation affecting the production process (FDA, 2011). This is especially important in stage 1 of the process validation activities, i.e., in the design phase where the effects of process parameters (PPs)/material attributes (MAs) and their impact on product quality are quantified. Design of experiments is recommended as an effective tool to achieve the following:

Design of Experiment (DOE) studies can help develop process knowledge by revealing relationships, including multivariate interactions, between the variable inputs (e.g., component characteristics or process parameters) and the resulting outputs (e.g., in-process material, intermediates, or the final product).

DOE followed by linear regression for modelling relationships between PPs, MAs, and CQAs are common tools employed in biopharmaceutical development (ICH, 2017). The assumption of linear relationships, i.e., models being linear in their parameters and not necessarily linear in their prediction of a factor, is generally valid for sufficiently small regions around a known working point (Montgomery, 2017). However, this should be carefully evaluated, e.g., by performing residual analysis of the derived models.

Process parameters are generally modelled as fixed effects that are assumed to be distributed around the true parameter value, i.e., $E(\hat{\beta}_j) = \beta_j$. Typically, any-

thing actively controlled by an operator might be considered a fixed effect, e.g., the temperature within a reactor, pH values, and feeding rates.

Random effects, in contrast, are parameters that are not controllable in such a way. Their future setting in an experiment or run cannot be predicted beforehand. However, they can still impact product quality and should therefore be considered when identifying possible sources of variability. MAs can fall within this category. Examples for random effects are changing raw material attributes, transition conditions, or biological variability in seed trains or even variability introduced by different operators.

The FDA guide (FDA, 2011) states the following:

The functionality and limitations of commercial manufacturing equipment should be considered in the process design, as well as predicted contributions to variability posed by different component lots, production operators, environmental conditions, and measurement systems in the production setting.

In most cases, a random effect affects a group of runs, which is called a block and the random effect a “block factor” (Montgomery, 2017). A typical example could be an experimental setup using a specific raw material lot for a set of runs. Statistical analysis is used to investigate the impact of those blocking effects. The resulting measure of variance is sometimes called *inter-block variability*, as opposed to *intra-block variability*, which constitutes the residual error term of the individual observations.

When processes are modelled in silico as regression models, block factors are usually incorporated as deviation-encoded fixed effects. In its most common form, deviation encoding describes individual blocking effects by their distance from the overall mean of the response (Alkharusi, 2012). When no block information is provided for a prediction, its numeric value is set to zero, and the mean response for the “average” block is computed, which is often the desired behavior. This, however, does not account for inter-block variability, and thus, the overall variability of the model is underestimated. As estimators of variability are used to compute proven acceptable ranges (PARs), this underestimation can have a large effect on the accuracy of such measures.

To illustrate this point, consider the method for calculating the PAR shown in Fig. 1. The plot shows CQA values as function of the process parameter screening range. The slope of the line in the center indicates the effect of the parameter on the CQA value. A measure of variability around the predicted mean is given by a statistical interval (confidence, prediction, or tolerance interval) shown as dashed lines. The range of the PAR (gray area) can be calculated by finding intersection points between the

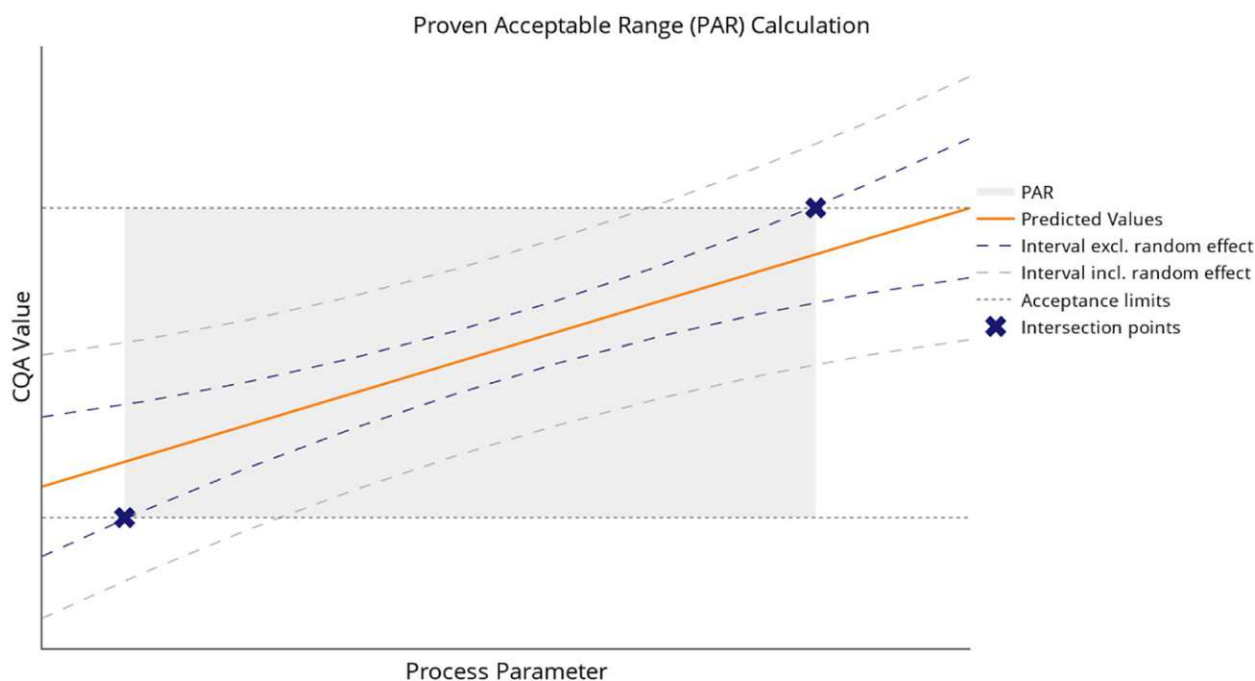


Fig. 1 An example for how the PAR of a process parameter can be calculated. The intersection points of a statistical interval and the CQA acceptance criteria define the lower and upper boundary. Note that the PAR that ignores the random effect is larger than the one that incorporates the random nature of the effect

acceptance limits and the statistical interval. When taking into account the additional variance introduced by random effects, the interval gets wider, and consequently, the PAR gets smaller. To find PARs for each critical process parameter in relation to the CQA acceptance criteria in drug substance, either each unit operation can be analyzed individually or integrated process modelling, enabling a holistic control strategy (Zahel et al., 2017), can be employed.

To accurately describe the random nature of those blocks, linear mixed models (LMM) can be employed to incorporate multiple sources of variation. In particular, the variation of random blocks can be computed separately and added to the overall variation of the model's prediction. Burdick et al. briefly illustrated the statistical methods behind LMMs and how they could be applied in process validation stage 1 (process design) in general (Burdick et al., 2017).

Goos et al. contrasted the conclusions drawn from OLS and LMM models in industrial split-plot designs and provides some guidance on analyzing experiments that involve random effects. The paper gives a motivating example and highlights some technical details of the method, like the proper choice of degrees of freedom (Goos et al., 2006). While pitfalls in improper experimental design and analysis are explained in a general manner, our work puts, for the first time, the problem in the

context of biopharmaceutical process development and provides a detailed workflow that considers characteristics specific to the domain.

Usually, the potential impact of fixed effects such as pH, temperature, or oxygen concentration is assessed in risk assessment such as failure mode and effects analysis (FMEA) and then investigated experimentally. The main contribution of this article is to illustrate in a case study how strong random effects can be in comparison with these fixed effects. Moreover, we introduce a workflow to incorporate random effects into process characterization in a statistically sound manner.

Following a workflow such as the one proposed is important in order to increase knowledge for the next round of risk assessment and experimental planning. Therefore, we aim to investigate the relative importance of random effects using a real-world data example of a full process characterization data set generated at Boehringer-Ingelheim over multiple unit operations of a drug substance production process including upstream and downstream.

Materials and methods

Models

To contrast ordinary least squares (OLS) and linear mixed model (LMM)-based approaches for creating a control strategy, both model types were fit to the process

characterization study (PCS) data. For OLS models, the random effect was treated as categorical fixed effect. While mean predictions are equivalent in this application, only LMMs decompose variance into a random and residual part, enabling the calculation of more accurate statistical intervals. This is especially important in the context of control strategies where patient safety could be at risk as a consequence of intervals that are too narrow, i.e., optimistic. A detailed description of OLS models and LMMs can be found in (Montgomery et al., 2021) and (SAS Institute Inc., 2010), and key differences and formulas are summarized in Additional file 4.

Statistical intervals

Statistical intervals represent an important and widely used tool to calculate and visualize uncertainty in data, estimators, or predictions in regression models. The most well-known type of interval is the confidence interval, which expresses uncertainty around the models' response, i.e., a confidence boundary around the predicted mean that contains the true population mean to a nominal level of confidence. Prediction intervals expand on this idea and add a standard deviation to the interval to define the region where a single, new observation is expected to fall within. To cover a nominal percentile of the actual distribution rather than a single observation, a third type of interval is used: the tolerance interval. Tolerance intervals cover the area that contains a predefined proportion of the true distribution of a response, often called coverage, to a nominal level of confidence. As we are interested in this true distribution of a modelled response, here in the form of critical quality attributes, tolerance intervals are used for the definition of the control strategy. Different techniques can be found in

literature to account for variance components in the calculation of intervals. Here, we use the method proposed by Francq et al. (Francq et al., 2019) to include the random contribution to variance estimated by linear mixed models. See Additional file 4 for a more detailed description of intervals and formulas used in the evaluation of the case study data.

Note that a tolerance interval converges to a prediction interval as the degrees of freedom increase. Figure 2 illustrates this effect while comparing the widths of the different types of intervals on simulated data. For this example, OLS-based intervals were computed. However, the relative widths and the effect of the degrees of freedom are the same when using LMM-based intervals.

Manufacturing process

The case study was performed at Boehringer-Ingelheim as a PCS of a monoclonal antibody process. The process consists of typical steps such as fermentation of a cell culture, harvesting, protein A column, intermediate and polishing column, and an UFDG step. Generally, the pool (output) of a unit operation is used as the load (input) of the next unit operation, so that the overall production process can be seen as a sequential chain of operations. The result of this chain is the actual drug substance, i.e., the product whose critical quality attributes are expected to fall within a predefined range, the so-called drug substance specifications.

Case study and data analysis

In a PCS conducted at Boehringer-Ingelheim, we investigated the impact of process parameters on 11–22 CQAs in each of the eight unit operations (UO). Models were created that regress a CQA in a unit operation onto

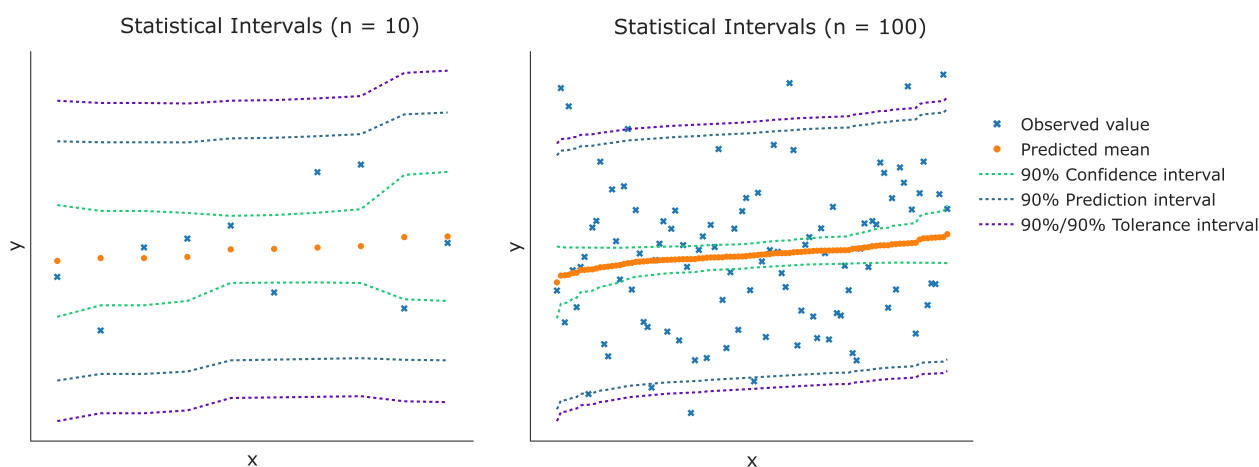


Fig. 2 Comparison of different statistical intervals and sample sizes. For conceptual and mathematical reasons, prediction intervals are always wider than confidence intervals when the same confidence level is assumed. Note that the tolerance interval converges to a prediction interval as the degrees of freedom are increased

factors found significant in the model selection process. The models were fit using data acquired in one-factor-at-a-time (OFAT) and design of experiments (DOE) runs using bench-scale experiments that are representative for the manufacturing process (see Additional file 1). Representativeness has been achieved via a pre-conducted small-scale qualification. During these activities, scale independent geometrical and engineering principles have been ensured as well as the absence of major performance differences between the scales.

The designs for DoE runs were planned in a way that minimizes aliasing and correlation between parameters (D-optimal), and that makes sure that effects can be detected with adequate statistical power. The DoE design and power analysis were conducted for fixed effects using the statistical software JMP (version 14.0.0). For the a priori power analysis, all runs (DoE + OFAT) were used. Moreover, a full model that includes all interaction and quadratic effects was assumed, which represents the worst possible case in terms of power. By convention, a power value of at least 0.8 is recommended. The power values for finding significant effects within two or three standard deviations of the residual error around the set-point are reported in Table 1. A common critical value of $\alpha=0.05$ was used as the significance threshold. Note that at the time the PCS was conducted, the power analysis was not explicitly conducted to incorporate random effects; see “Random effects in power analysis” section for an explanation of our approach. We performed variable selection (best subset selection or stepwise bidirectional) to eliminate nonsignificant effects and create more parsimonious models. Data for each unit operation consisted of one random effect describing the week-to-week variability across the experiments.

Good modelling practice was employed to check for model quality after variable selection. This was done by residual analysis to check for normality of residuals, inspecting model parameter p -values to determine if they

exceed a threshold of 0.1, checking whether the RMSE is within expected reproduction variability and thereby mitigating the risk of overlooking effects as well as overfitting, leave-one-out cross-validation to exclude biasing the model via single runs. These measures increase the confidence that neither a substantial type 1 error (including effects that are not significantly different from 0) nor type 2 error (overlooking effects) has been made. The latter also implies that no aliasing is expected, which may bias a fixed/random effect. In general, we followed the approach to data analysis and model creation outlined as “workflow B” in the “Workflow B: modelling random effects using linear mixed models” section. The evaluation of effects in “Effect sizes and variances” section is summarized as a series of box plots that show the fixed and random effect sizes as well as variance ratios for each model (Figs. 3 and 4). To create a comparable measure of effect size, the original data used to fit the model were min-max normalized based on the parameter screening ranges. This means that all the input parameter values lie within the interval $[-1, 1]$, and their effects after fitting are comparable within the model.

$$x^* = 2 \frac{x - \min(x)}{\max(x) - \min(x)} - 1 \tag{1}$$

As the response values in the training data were not normalized and used in their original scales, the effects were additionally divided by the root-mean-square error (RMSE) to make them comparable between models. For each model, an average measure of fixed and random effects was computed using their absolute values.

$$\bar{\beta}_{CQA} = \frac{1}{p-1} \sum_{i=1}^{p-1} \left| \frac{\hat{\beta}_i}{\hat{\sigma}_\epsilon} \right| \tag{2}$$

$$\bar{\gamma}_{CQA} = \frac{1}{m} \sum_{i=1}^m \left| \frac{\hat{\gamma}_i}{\hat{\sigma}_\epsilon} \right| \tag{3}$$

Table 1 Average statistical power over all effects to detect an effect within 2 or 3 standard deviations from the set point. Note that those values represent the worst case that assumes a full model, i.e., a model that includes all quadratic and interaction effects

UO	Power for 2 SD	Power for 3 SD	Runs DoE/OFAT	RE levels	Significant effects	Responses
UO 1	0.91	0.99	24/6	3	16	22
UO 2	1.00	1.00	27/24	4-5	30	20
UO 3	0.95	1.00	17/5	5	10	14
UO 4	0.71	0.93	18/3	6	10	11
UO 5	0.94	1.00	21/3	5	15	17
UO 6	0.99	1.00	42/11	3	36	13
UO 7	0.82	0.96	0/6	5	3	11
UO 8	0.89	0.98	0/13	5	8	15

where $p - 1$ is the number of parameters minus the intercept, m the number of levels of the random effect investigated, and $\hat{\sigma}_\epsilon$ represents the estimator of the residual variance, i.e., the RMSE. The intercept is excluded in the calculation of $\bar{\beta}_{CQA}$ as only the effect of actual model parameters should be measured. A distribution of those values per unit operation is illustrated as box plots in Fig. 3. The figure highlights the random effects contribution to overall variance and compares it to the fixed effect contribution.

A similar approach was taken with the variance ratios in Fig. 4. Per CQA model, the ratio $\hat{\sigma}_y / \hat{\sigma}_\epsilon$ was calculated, and the distribution of values is shown as box plots per unit operation.

Results

In the analysis of the case study data of real industrial data from a PCS, we contrast the OLS and LMM-based method to forming statistical intervals. As a random effects contribution to observed variance is proportional to its effect size, we first compare normalized estimators of fixed and random effects. This helps us to identify how strong random effects are in comparison with well-known fixed effects, such as pH and temperature. We then show how this random contribution increases the tolerance intervals and, in turn, reduces the PAR in an example picked from the case study.

Effect sizes and variances

The random effect investigated in the analysis of the case study data was week-to-week variability. To show effect sizes of the random effect predictors relative to fixed effects and their contribution to variability, LMMs were fit to the data. Figures 3 and 4 show the effects and variances per unit operation. It can be seen that the random effect is even greater than the fixed effect in some of the unit operations. Ignoring the random effects' impact on variability would underestimate the size of statistical intervals and result in an inappropriate control strategy. In four of the eight unit operations (UO3, UO5, UO7, and UO8), the median standardized random effect size was larger than the median standardized fixed effect size (see Fig. 3). In the other UOs, the random effect size is approximately the same as the fixed effect. Moreover, for six out of eight UOs (UO2, UO3, UO4, UO5, UO7, UO8), the median variance ratio of random versus residual variance is equal or larger to one (see Fig. 4). Effect sizes are of course dependent on the experimental design the data is based on. However, while variation of random effect estimates might be larger than those of fixed effects, a general trend is clearly discernible in our results (see Fig. 3), and significance of both random and fixed effects was checked in the model selection process.

LMM random effect predictors are often described as the empirical best linear unbiased predictors (EBLUP)

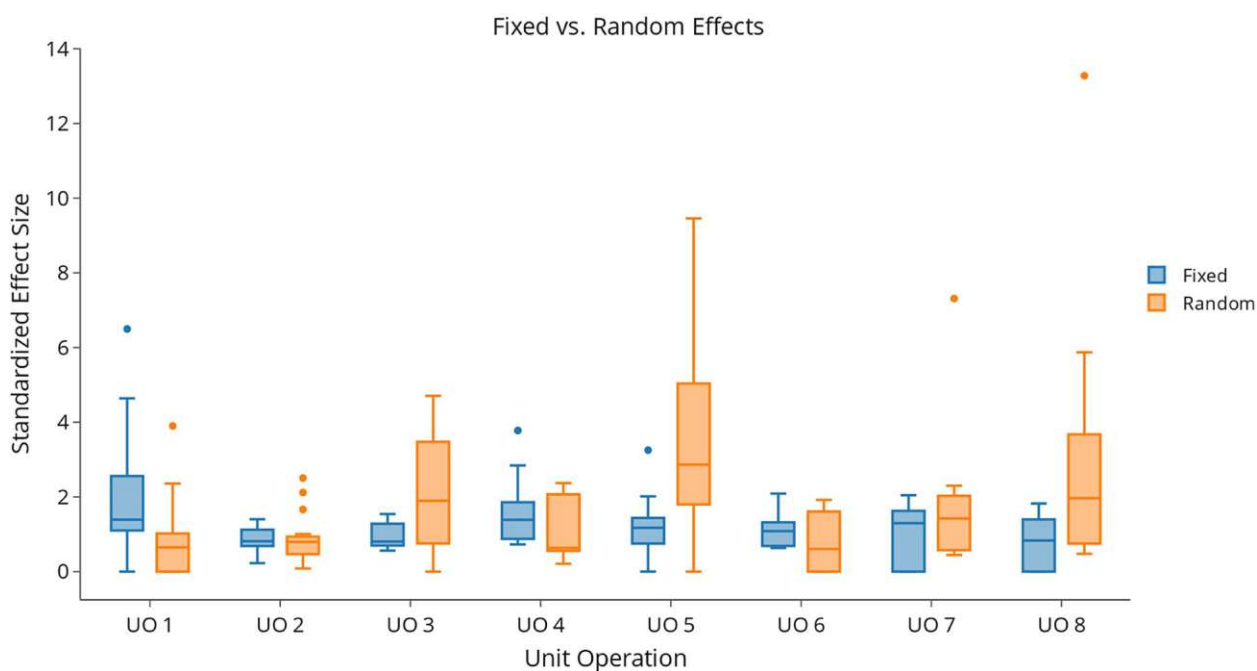


Fig. 3 Standardized fixed and random effect sizes are contrasted for each unit operation. A unit operation contains models for 11–22 CQAs, and their respective fixed and random effect distributions are shown as box plots. To create comparable measures of effect size, normalized data were used to fit the models, and the effects were divided by the RMSE. Note that for several unit operations, the median random effect is even larger than the median fixed effect

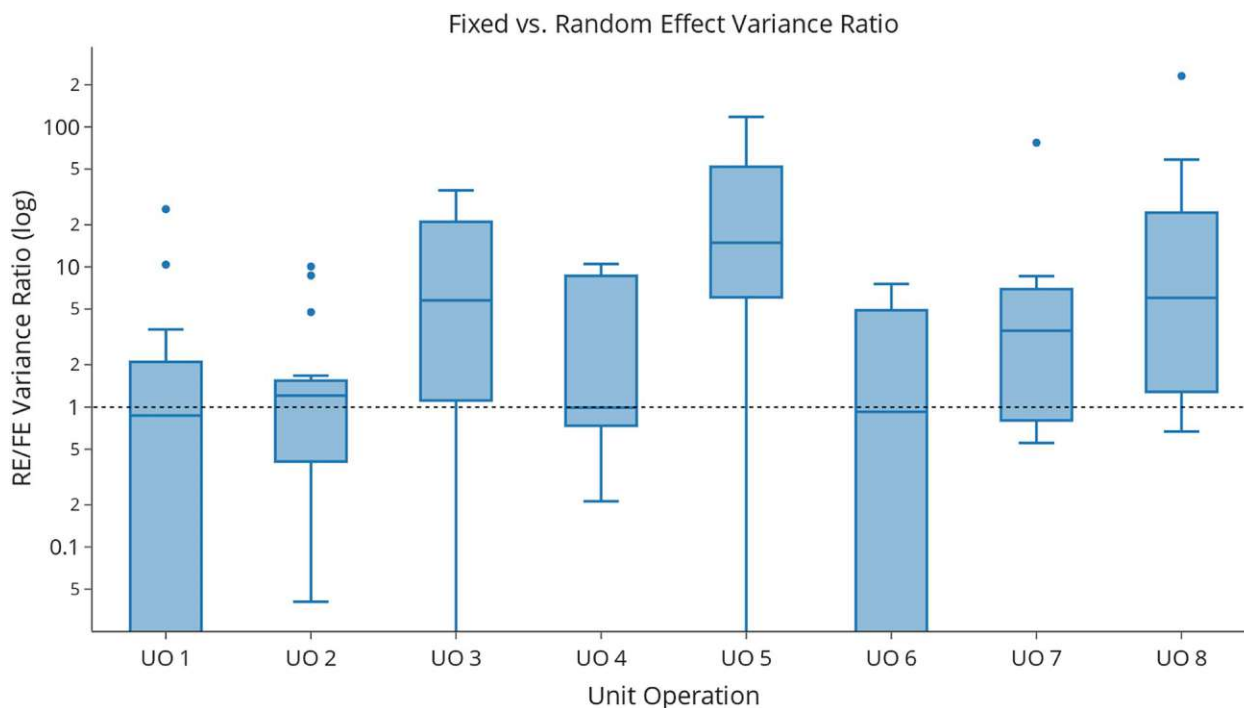


Fig. 4 Variance ratios (random variance/residual variance) are shown per unit operation on a logarithmic scale. For each model in a unit operation, the ratio between random and residual variance is calculated and the resulting distribution illustrated as a box plot. As random effect size increases, so does its contribution to variance — in some cases, the random contribution to variance is many times as large as the residual variance

in literature and yield more accurate effect sizes when compared to those obtained by modelling them as fixed effects using OLS (Govaerts et al., 2020). Due to the way they are calculated in mixed models (see Additional file 4), they tend to be closer to zero. This should be considered when comparing random and fixed effect sizes in Fig. 3. The amount by which their effect size is “shrunk” is inversely proportional to the associated variance component, i.e., the smaller the random effects variance, the larger the amount of shrinkage and vice versa. Figure 4 shows that in our case study, random effect variance is quite large relative to that of fixed effects, indicating that effect sizes based on EBLUPs should not differ substantially from those of obtained from modelling random effects as fixed effects using OLS. Moreover, our overall message that random effects are equally or more influential in a representative process characterization would be even more pronounced calculating out the shrinkage.

Tolerance intervals

Modeling a random effect as a categorical, fixed effect using OLS models is an often-employed practice in biopharmaceutical manufacturing. Here, we show the implications of this approach in a representative example picked from a real-world case study. Assuming a normal distribution of residuals, the chosen tolerance interval

should contain at least 90% of observations in 50% of repeated samplings. However, this was not the case. As illustrated in Fig. 5, in extreme cases, the OLS tolerance interval almost never included the value observed in the runs. This was due to variability introduced by different blocking factors in the production process. The larger the blocking effect, the larger its influence on variability — a quantity that is ignored in the OLS case. Incorporating random effect variability by employing LMM models and appropriate interval calculation methods solves this problem, which can be seen in the outer interval in Fig. 5.

Further analysis revealed that this observation was not an exception, but that the data for most CQAs included significant blocking effects that would result in a tolerance interval too narrow when ignored. Table 2 gives an overview of interval width ratios $r = (TI_{LMM, upper} - TI_{LMM, lower}) / (TI_{OLS, upper} - TI_{OLS, lower})$ for the most common CQAs at setpoint conditions. Depending on the random effect size, the LMM intervals can be several multiples as wide as their OLS counterparts, when ignoring the random effect.

PAR and control strategy

The general increase in tolerance interval width when incorporating random effects and LMMs reported in the previous section can have a considerable impact on

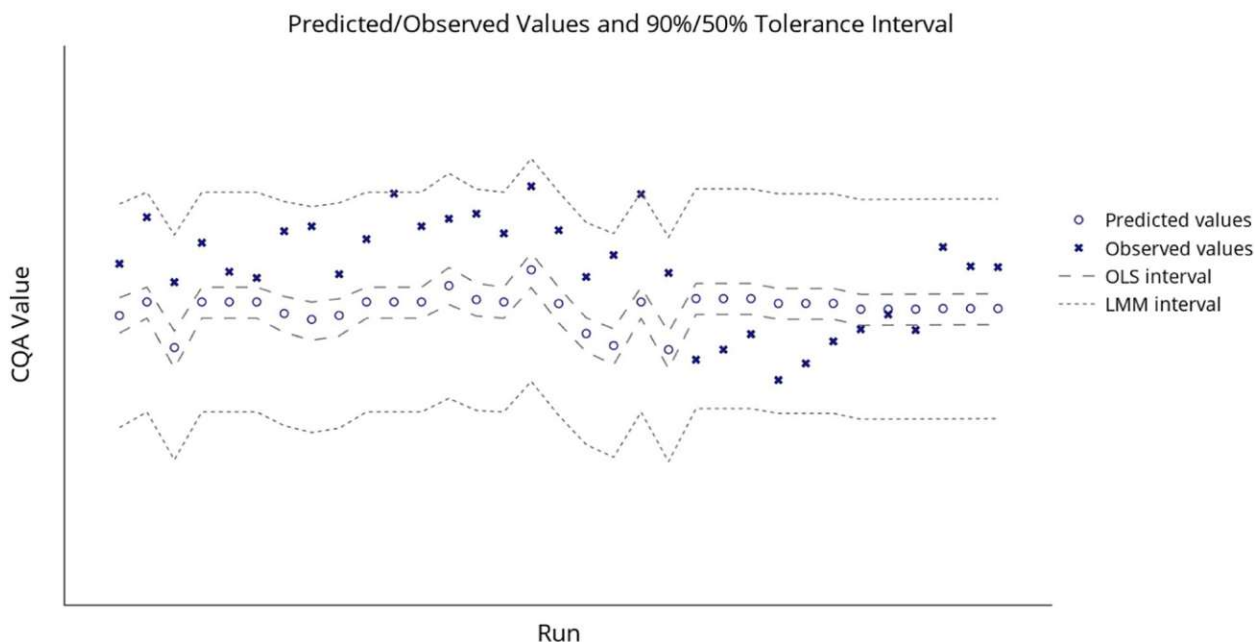


Fig. 5 A 90%/50% tolerance interval is created around the mean. By definition, it should include 90% of the data in 50% of cases, which is obviously not the case when using an OLS model. However, the interval computed using variance information from the LMM model does indeed cover at least 90% of observations

Table 2 LMM/OLS tolerance interval width ratios for 6 of the most common CQAs per unit operation. Due to the strong random effect, the LMM interval is generally much wider than the OLS interval

	UO 1	UO 2	UO 3	UO 4	UO 5	UO 6	UO 7	UO 8
CQA 1	1.85	3.80	2.67	3.18	3.01	1.57	1.17	1.08
CQA 2	1.62	1.65	6.15	3.36	4.59	3.37	2.38	5.15
CQA 3		1.70	4.99	3.43	4.06	2.43	2.34	1.41
CQA 4	1.85	1.64	1.77	1.12	4.38		8.95	2.84
CQA 5		1.80	4.58	1.37	8.34		2.92	3.10
CQA 6		1.67		1.25	9.02		3.46	2.98

the control strategy. Again, a representative example is selected from the case study in the form of a process parameter. For the chosen parameter, both an OLS model and LMM was fit, and tolerance intervals were calculated using the corresponding methods (for models and data, see Additional file 3). As can be seen in Fig. 6, when using the intersection points of the interval with the upper acceptance limit, the resulting PAR for OLS is indeed wider than the one based on the LMM.

Depending on the size of the fixed effect and the chosen acceptance limits, the reduction of the PAR might be more or less severe. Generally, for PARs formed with

the method illustrated, its size can only decrease with the increase of the interval width as indicated in Fig. 6.

Discussion

Workflows to establish a control strategy

As shown in the case studies presented in the “Results” section, random effects can have a large effect on statistical intervals, the PAR, and consequently the control strategy. Here, we propose a workflow for establishing a control strategy that incorporates random effects at various stages. We first suggest an OLS-based workflow typically used in the industry and then contrast this strategy with one that incorporates random effects using linear mixed models.

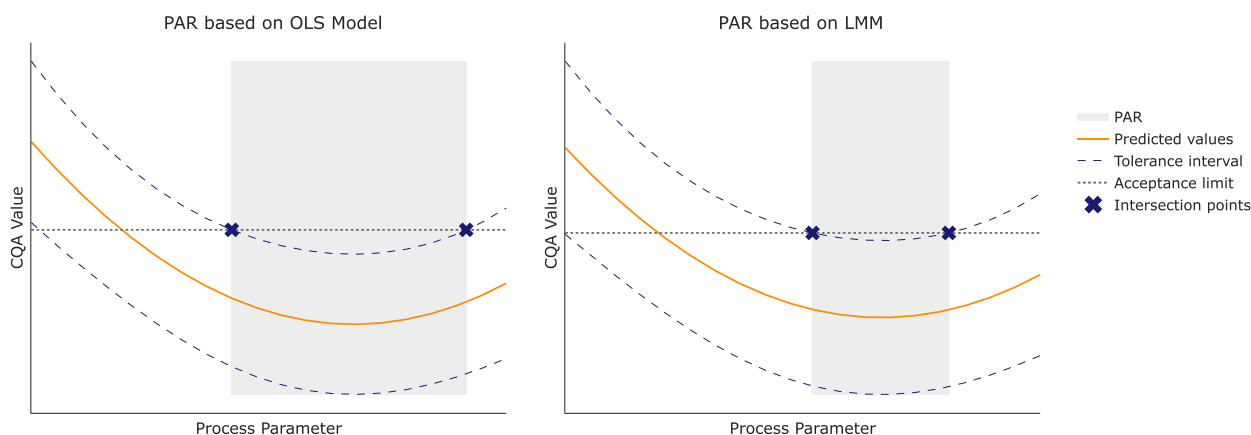


Fig. 6 PAR for a randomly picked parameter calculated from case study data. Due to the contribution of the random effect, the interval based on LMM variance components (right) is wider than its OLS counterpart (left). This results in smaller PAR (gray area) and a more conservative control strategy. In this example from the case study, the OLS PAR is 72% larger than the more conservative LMM PAR

Workflow A: modelling random effects as fixed effects using OLS

In the first step, the data that constitutes the basis for the regression model is acquired. We assume that data originates from a design that ensures desired properties for analysis, such as minimal correlation, minimal aliasing, and maximal power. In this phase, random blocking factors are identified alongside all other factors that might influence the response of the process, and their values are aggregated into a single data source that enables convenient analysis. This is followed by a preprocessing step where those data are cleaned up and response data possibly transformed to a form that satisfies OLS model assumptions (normality of residuals). Random effects are treated as categorical fixed effects and deviation-encoded so that the reference for the individual block coefficients is the grand mean of the response. This enables to set the blocks to zero for predictions, which results in a “mean block” prediction of the response. At this stage, a “full model” can be created by adding quadratic and interaction effects for each main effect. Given the available number of observations, use case, or preference, this full model can be used directly. Alternatively, the list of effects can be used as the input for a variable selection procedure to find a parsimonious model that explains the response while eliminating insignificant parameters. Such procedures are commonly based on estimators of prediction error, for example, the Akaike information criterion (AIC), or on p -values of model parameters. The implementation of such estimators depends on the type of model used, as they are different for OLS and LMM. Blocking factors might be found insignificant in the variable selection process and removed from the model. In the last step, either the full or optimal model

is used to compute the predicted values for the training data, whereby the predictor variables for the block are set to zero. Around those predictions, a tolerance interval is formed that contains a proportion of the population (coverage) with a certain probability (confidence). This should be reflected by the observed values of the training data contained in the interval. The PAR of the parameter is formed by the intersections of the interval with the acceptance criteria (see Fig. 1). However, such a tolerance interval based on OLS models does not incorporate the variance introduced by random effects correctly and might lead to control strategy that is too optimistic.

Workflow B: modelling random effects using linear mixed models

As it was the case in the first workflow, the LMM-based procedure starts by identifying both fixed and uncontrollable, random effects. Special attention is given to the latter as often multiple random factors are involved, which can be nested or crossed, both of which influences variance calculations in different ways. In addition to correlation analysis of fixed effects, some data prerequisites specific to LMMs should be checked to make sure the likelihood optimization converges, though this depends on the statistical software or library employed. Statistical significance of individual blocks potentially affects convergence and can be examined beforehand by deviation encoding them as described in in workflow A and investigating their effects using p -values obtained from an OLS fit. The levels of the random block variable as well as the number of intra-block observations are also factors in the optimization algorithm as highly imbalanced blocks can be the source of convergence problems. After making sure that the data meets all the criteria for applying a

LMM, a full model that includes quadratic and interaction effects can be created. Again, this full model can be used directly or as the input for variable selection where insignificant model parameters are eliminated in each step of the algorithm until the optimal model is found. In this workflow, variable selection can be performed in two different ways: one option is to deviation-encode random blocks and fit OLS models which are then used for evaluation in each step — essentially the same process as in the first workflow. The fixed effects from the final model are then used to transform the OLS model into an LMM. This can be a sensible workaround in situations where one is constrained by software lacking variable selection procedures that incorporate random effects. However, note that this approach might not be possible in some experimental designs. The second option is to use LMM-specific evaluation criteria in each step of the variable selection process. While this might be the most obvious approach, it is also not universally applicable, depending on the algorithm, criteria, or performance constraints. After a satisfactory model is found, predictions can be computed. For LMM, this means that only the matrix of fixed effects needs to be provided for the prediction as the model automatically assumes the “mean block” for the results. This is different from OLS models where blocks need to be set to zero explicitly for the prediction. The notable difference between the models is in

how model variance is computed and partitioned, which is important in the next step: the calculation of statistical intervals. Here, mixed models include a measure of variance of both fixed and random effects which, depending on the magnitude of the random effect, can widen the interval and therefore reduces the acceptable range of the process parameter.

Figure 7 summarizes the OLS- and LMM-based approaches to establishing a control strategy and provides an overview of their main differences.

The workflows outlined here represent two methods for computing PARs using common data-science techniques. Workflow A shows a common approach employed in biopharmaceutical manufacturing, while workflow B represents our proposal for an extended version that incorporates random variance correctly into statistical intervals. Given the unlikely scenario of a process being not affected by random effects at all, workflows A and B would result in the same control strategy.

Modelling scale impact

For the process characterization study described in this article, only data from bench-scale DoE and OFAT experiments were used, as no large-scale data was available at that point in time. Typically, manufacturing data is supplemented in the analysis to investigate the effects of scale. In regression models, this can be done by simply

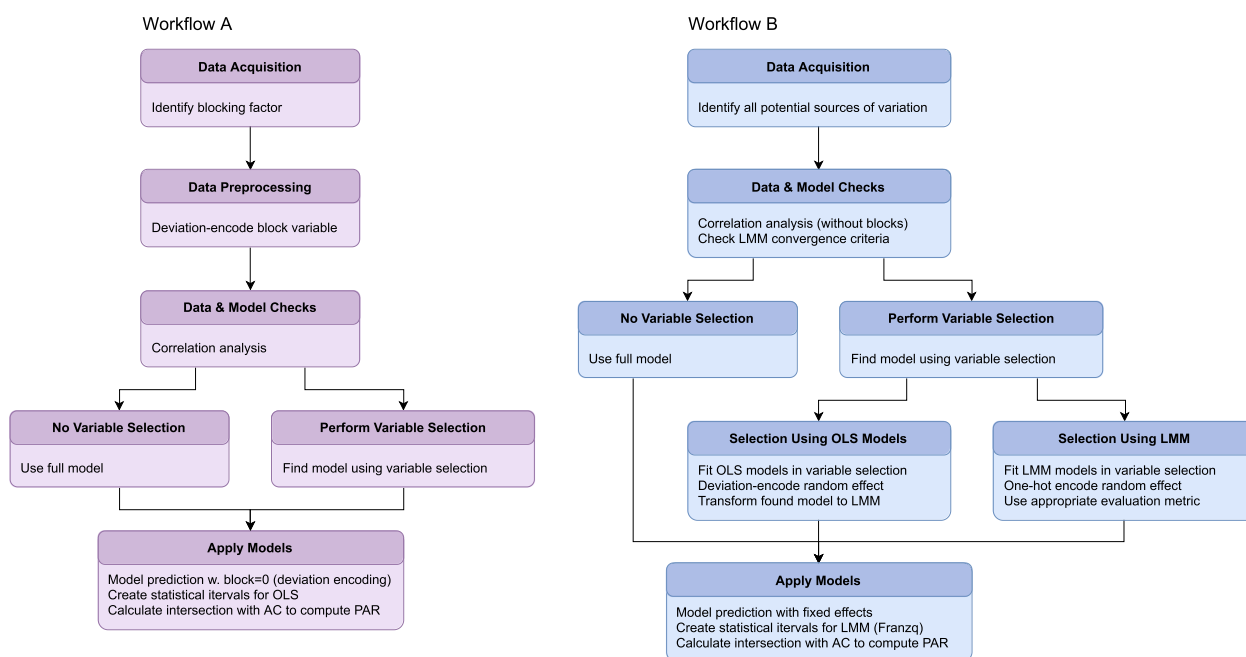


Fig. 7 Workflows for creating control strategies based on regression models. The left column describes an approach that uses OLS models for the estimation of PARs. A mixed-model-based workflow is summarized on the right. The differences in the steps involved are subtle but generally result in a more realistic estimation of variance and therefore a more robust control strategy

adding a categorical factor to the model with one level per data source (e.g., “large scale” and “DoE”). As a regular, fixed effect, such a factor can be subject to variable selection and might be removed from the model when deemed insignificant. Relationships between scale and other effects in the model can be explored by creating scale-interaction effects prior to variable selection, provided enough degrees of freedom are available to detect them. Admittedly, this requires off-setpoint runs at large scale which are unlikely to be available in a data set.

Random effects in power analysis

At the time the experimental runs for the PCS were planned and the power values in Table 1 were calculated, the importance of random effects was not known to its full extent. Therefore, a priori power analysis that considers the random variance component explicitly was not performed. Simulation-based power calculation methods that incorporate a random variance contribution are available in some software packages. This might be considered in future experimental planning. However, how does this affect our claim that random effects are strong throughout most UOs?. When considering the random effect levels in Table 1, one could argue that the number of levels might not be sufficient in terms of statistical power to detect all active random effects. However, it should be noted that the effect sizes shown in Fig. 3 have been obtained using the variable selection method described in workflow B (“Workflow B: modelling random effects using linear mixed models” section), which controls via a p -value threshold for the false-positive rate/type 1 error, even though actual effect sizes might be smaller due to the shrinkage effect described in “Effect sizes and variances” section. The random effect was then checked for significance in the resulting models using variance ratio tests (Nakagawa & Schielzeth, 2013). In only 18 of the 123 models, the random effect was found to be not significant, in which case a value of zero was used for the data points shown in Fig. 3. Moreover, averages of random effect predictor sizes found by LMM over all models are strong throughout all unit operations. This supports our finding that, overall, the random effect is oftentimes larger than the fixed effects. While the lack of statistical power might lead to overly conservative tolerance intervals, the PARs of this study have been found to be practically acceptable for manufacturing.

Implications for the biopharmaceutical industry

Workflow B proposed in the “Workflow B: modelling random effects using linear mixed models” section puts the method for considering random effects in process design (stage 1) proposed by Burdick et al. (Burdick et al., 2017) into the context of a workflow that includes

variable selection. This aligns with the ICH8 recommendations for including all potential sources of variation into the computation of the control strategy (ICH, 2017).

Ignoring a random effect or modelling it as a fixed effect can change effect and variance estimates notably. Goos et al. (Goos et al., 2006) demonstrated in a simulation study that this is the case for improperly analyzed split-plot designs, and our results show that it holds true for the analysis of a process characterization data in biopharmaceutical manufacturing. The statistical implications of inappropriately choosing an OLS model over an LMM for the calculation of intervals is shown in “Tolerance intervals” section.

Large tolerance intervals and pronounced random effect sizes indicate that an effect affecting the process is poorly understood, and its true root cause should be investigated. By identifying the source of random variation and controlling it, it can essentially be resolved into a fixed effect.

For example, vendor-to-vendor variability of a raw material might lead to a large random effect, i.e., an unexpected random source of variation. Consequently, a set of experiments can be conducted to identify the true root cause of this variation, e.g., a supplement of the raw material. Provided the manufacturer is able to control this supplement, it can be incorporated into a model as a fixed effect. If this is not feasible or planned for a later point in time, LMM tolerance intervals can be used to estimate the distribution of critical quality attributes more accurately and to find a conservative control strategy for the fixed effects, thus reducing out-of-specification events.

In general, we recommend the following:

- Investigate the practical significance of the random effect (e.g., does its variance take up a large fraction of the CQA acceptance limits/drug substance specification and hence is a risk to the patient?).
- If feasible, conduct experiments to identify causes of random variation and re-evaluate experimental data.
- If it turns out that the effect can be modelled and controlled as a fixed effect, implement changes in the process to control the root cause.
- Uncontrollable effects can still be modelled as random effects in LMMs for conservative tolerance intervals.

Note that process validation is a risk-based approach starting with risk assessment conducted to identify potentially impacting factors (fixed and random effects). Of course, if this initial step overlooks one of the important factors, they will not be assessed experimentally. In case those underrated factors from the risk assessment

are not controlled well in manufacturing, the control strategy established through a PCS might be insufficient. In that case, stage 3 of process validation (continued process verification — CPV) steps in and aims at identification of special cause variation possibly raised from one of the underrated factors. When special cause variation can be detected, it may trigger a new round of risk assessment and experimental planning and analysis, bringing birth to a true life cycle, which FDA proposes in its 2011 PV guideline.

Conclusion

In this article, the role and impact of random effects on setting the control strategy of a biopharmaceutical process were investigated in a real-world case study conducted at Boehringer Ingelheim. Data from a production process comprised of eight up- and downstream unit operations were analyzed in a case study. Although this contribution is based on an extensive process characterization of a monoclonal antibody process and the results are believed to be representative for similar processes, we encourage researchers to conduct similar case studies with other processes and random variables. Here, inter-week batch variability was chosen as the random effect. Such an effect, if not ignored entirely, is commonly incorporated in an OLS model as a categorical fixed effect. For the case study, however, the factor was modelled as a random effect using linear mixed models where the segmentation of variance components into random and fixed components enables a more accurate calculation of statistical intervals. The results show that the random effect not only increases the width of the statistical intervals used to compute PARs, but also exceeds even in several unit operations the average size of the fixed effect. Those findings are confirmed by the number of observations contained within the tolerance interval, which agrees with the nominal coverage level for LMMs but not for OLS models. As random effects might have such a strong impact and even stronger impact than fixed effects, they should be incorporated into risk assessments and included into experimental studies. If tolerance intervals derived from LMM models are too large, further investigations should be performed to resolve random effects into fixed effects, e.g., by identifying the underlying root cause of the variation and controlling it. However, until this state is reached, the random variance should at least be accounted for in the model prediction uncertainty as described in this contribution.

Furthermore, we presented a workflow commonly used for creating a control strategy using OLS models. In this workflow, one of the standard implementations of tolerance intervals in a multiple regression setting is utilized, and the intersection points with acceptance

criteria are computed to arrive at the acceptable range for each process parameter. This constitutes the control strategy for the process. As an alternative, we proposed an LMM-based workflow that performs similar actions but touches upon certain characteristics of random effects and mixed models. This mainly manifests in the variable selection process and in the computation of statistical intervals where the variance introduced by fixed and random effects is incorporated appropriately. We suggest the use of tolerance intervals based on the sum of expected mean squares proposed by Franzq et al. (Franzq et al., 2019). Depending on the group structure and the available degrees of freedom, the interval produced by this method tends to be wider than its OLS counterpart.

Identifying and incorporating random effects are vitally important when defining the control strategy of a process and adjacent tasks like experimental planning and risk assessment. Employing methods described in the proposed workflow, e.g., linear mixed models and corresponding tolerance intervals, leads to a more conservative and appropriate control strategy, which ultimately facilitates more robust processes, patient safety, and fewer out-of-specification events.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s41120-022-00070-5>.

Additional file 1. Anonymized PCS data.

Additional file 2. Estimated effect sizes and variance ratios.

Additional file 3. Motivating Example for PAR Calculation. As an introductory example for how different models and interval calculation methods affect PAR calculation, this plot was shown in the article.

Additional file 4. Models and Intervals.

Acknowledgements

This work was conducted within the COMET Centre CHASE, funded within the COMET — Competence Centers for Excellent Technologies program by the BMK, the BMDW, and the federal provinces of Upper Austria and Vienna. The COMET program is managed by the Austrian Research Promotion Agency (FFG). The authors acknowledge TU Wien Bibliothek for financial support through its Open Access Funding Program.

Authors' contributions

TO was the main contributor to the manuscript, implemented tolerance intervals for linear mixed models in Python (based on the method proposed by Franzq et al.), and performed the data analysis. TZ contributed to the manuscript and was the primary advisor for statistics and data analysis methods. MK validated the manuscript and data and provided input from a manufacturing perspective. JT supervised the experiments during the process characterization study at Boehringer Ingelheim. CH provided guidance in writing of the manuscript as the PhD supervisor to TO. The authors read and approved the final manuscript.

Funding

Open access funding provided by TU Wien (TUW). This work was supported by the Austrian Research Promotion Agency (FFG) (grant number: 844608) and within the framework of the Competence Center CHASE GmbH, funded

by the Austrian Research Promotion Agency (grant number 868615) as part of the COMET program (Competence Centers for Excellent Technologies) by BMVIT, BMDW, and the federal provinces of Upper Austria and Vienna.

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 31 May 2022 Accepted: 18 December 2022

Published online: 01 February 2023

References

- Alkharusi H (2012) Categorical variables in regression analysis: a comparison of dummy and effect coding. *Int J Educ* 4:202–210
- Burdick RK, LeBlond DJ, Pfahler LB, Quiroz J, Sidor L, Vukovinsky K, Zhang L (2017) "Process design: stage 1 of the FDA process validation guidance," in *Statistical Applications for Chemistry, Manufacturing and Controls (CMC) in the Pharmaceutical Industry*, Springer, pp 115–154
- FDA, *Process validation: general principles and practices*, 2011.
- Francq BG, Lin D, Hoyer W (2019) Confidence, prediction, and tolerance in linear mixed models. *Stat Med* 38(30):5603–5622
- Goos P, Langhans I, Vandebroek M (2006) Practical inference from industrial split-plot designs. *J Qual Technol* 38(2):162–179
- Govaerts B, Francq B, Marion R, Martin M, Thiel M (2020) The essentials on linear regression, ANOVA, general linear and linear mixed models for the chemist. *Ref Module Chem Mol Sci Chem Eng*:431–463
- ICH, *ICH pharmaceutical quality system Q10*, 2008.
- ICH, *ICH guideline Q11 on development and manufacture of drug substances*, 2011.
- ICH, *ICH guideline Q8 (R2) on pharmaceutical development*, 2017.
- Montgomery DC (2017) *Design and analysis of experiments*. Wiley
- Montgomery DC, Peck EA, Vining GG (2021) *Introduction to linear regression analysis*. Wiley
- Nakagawa S, Schielzeth H (2013) A general and simple method for obtaining R² from generalized linear mixed-effects models. *Methods Ecol Evol* 4:133–142
- SAS Institute Inc. (2010) *SAS/STAT 9.22 User's Guide*. SAS Institute Inc., Cary
- Zahel T, Hauer S, Mueller EM, Murphy P, Abad S, Vasilieva E, Maurer D, Brocard C, Reinisch D, Sagmeister P (2017) Integrated process modeling - a process validation life cycle companion. *Bioengineering* 4(4):86

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)

7.2 Scientific publication II

Article

Holistic Design of Experiments Using an Integrated Process Model

Thomas Oberleitner ¹, Thomas Zahel ², Barbara Pretzner ² and Christoph Herwig ^{2,3,*}

¹ Competence Center CHASE GmbH, Ghegastraße 3, Top 3.2, 1030 Vienna, Austria

² Körber Pharma Austria GmbH, PAS-X Savvy, Mariahilferstraße 88A/1/9, 1070 Vienna, Austria

³ Research Area Biochemical Engineering, Vienna University of Technology, Gumpendorferstrasse 1a, 1060 Vienna, Austria

* Correspondence: christoph.herwig@tuwien.ac.at

Abstract: Statistical experimental designs such as factorial, optimal, or definitive screening designs represent the state of the art in biopharmaceutical process characterization. However, such methods alone do not leverage the fact that processes operate as a mutual interplay of multiple steps. Instead, they aim to investigate only one process step at a time. Here, we want to develop a new experimental design method that seeks to gain information about final product quality, placing the right type of run at the right unit operation. This is done by minimizing the simulated out-of-specification rate of an integrated process model comprised of a chain of regression models that map process parameters to critical quality attributes for each unit operation. Unit operation models are connected by passing their response to the next unit operation model as a load parameter, as is done in real-world manufacturing processes. The proposed holistic DoE (hDoE) method is benchmarked against standard process characterization approaches in a set of in silico simulation studies where data are generated by different ground truth processes to illustrate the validity over a range of scenarios. Results show that the hDoE approach leads to a >50% decrease in experiments, even for simple cases, and, at the same time, achieves the main goal of process development, validation, and manufacturing to consistently deliver product quality.

Keywords: design of experiments; holistic experimental design; integrated process model; optimal designs; process characterization; biopharmaceutical process validation

1. Introduction

The goal of process characterization in biopharmaceutical development is to establish scientific evidence that a process is able to consistently deliver quality products. An important part of this procedure is to determine the effect of process parameters (PP) on critical quality attributes (CQA [1,2]). Design of experiments (DoE) is a well-established tool to design experimental runs that yield such information and is oftentimes followed by data analysis and inference based on regression models [3,4]. DoE variants such as factorial or optimal designs facilitate the detection of effects by minimizing or eliminating correlation, and they are comprised of all possible combinations of effect levels or a subset thereof [5]. These experiments are then conducted in small-scale models for each unit operation (UO), and results can be used to create mathematical models that quantify the impact of effects.

Of particular interest when defining a control strategy is the range in which PPs can safely operate while keeping CQA concentrations within acceptable boundaries. Those proven acceptable ranges (PAR) are part of the control strategy a manufacturer might submit to a regulatory authority [6]. One way of defining a parameter's PAR is illustrated in Figure 1. Using this approach, the PAR is affected by two components: first, the model prediction, including the statistical intervals and second, the acceptance limits, as the intersection between the two defines the PAR's range. Another option could be to alter the setpoint condition of other PPs (see Section 4.2). This leads to a change in the

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar. The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.



Citation: Oberleitner, T.; Zahel, T.; Pretzner, B.; Herwig, C. Holistic Design of Experiments Using an Integrated Process Model. *Bioengineering* **2022**, *9*, 643. <https://doi.org/10.3390/bioengineering9110643>

Academic Editors: Sundeeep Singh, Roderick Melnik and Esther Pueyo

Received: 14 October 2022

Accepted: 1 November 2022

Published: 3 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

univariate prediction plot by shifting the prediction vertically (orange line in Figure 1). However, optimization by changing setpoint conditions is usually not the focus of process characterization. The PAR is required to allow for sufficient process and operator variability while being conservative enough to keep CQAs within acceptance limits. If the PAR is too narrow for adequate operability, one can either try to reduce model uncertainty by investing DoE runs or change acceptance limits by performing spiking runs.

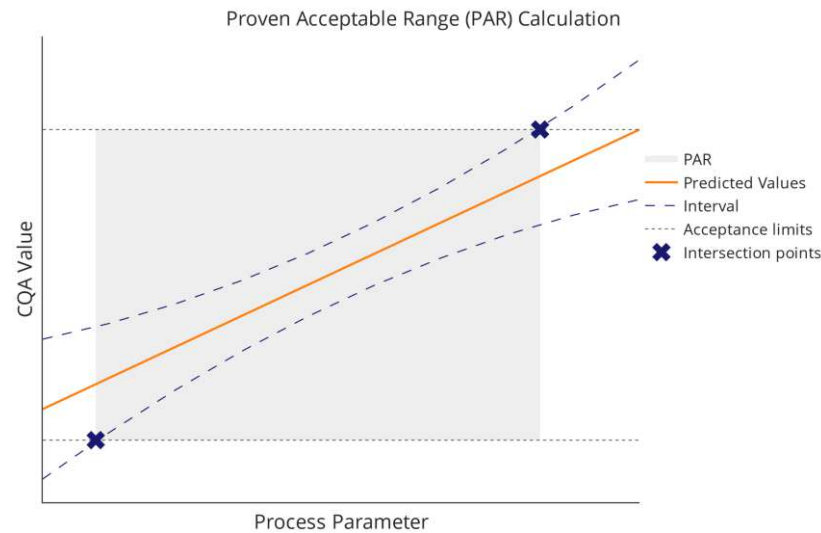


Figure 1. An example for how the PAR of a process parameter can be calculated. The predicted mean of the CQA as a function of the PP is shown in orange and the statistical interval around these predictions is illustrated as dashed lines. Lower and upper PAR boundaries can be defined by the intersection points of a statistical interval and the CQA acceptance criteria, marked as x.

1.1. Option A: Improving Model Estimates via DoE

To illustrate the effect of DoE runs on model uncertainty, consider the formula for calculating tolerance intervals for a normally distributed population [7]:

$$\hat{y} \pm z_{\frac{1+\psi}{2}} \sqrt{\frac{\nu \left(1 + \left(\frac{1}{N}\right)\right)}{\chi_{\alpha, \nu}^2}} \sigma \tag{1}$$

\hat{y} is the mean prediction of the model, $1 - \psi$ the nominal proportion of the population covered by the interval, and α the confidence level. Disregarding the critical value for the normal distribution $z_{\frac{1+\psi}{2}}$ and the standard deviation σ , the dominant factor in this formula is the square root term, which includes the lower α quantile of the χ^2 distribution in its denominator and converges toward one as N increases. The residual degrees of freedom in a regression model are calculated as $\nu = N - p$, where p is the number of model parameters. Note that this is a simplified version of the tolerance interval, and other methods might be used in a regression setting [8–10]. The graph of the square root term is shown in Figure 2 over a range of possible values of N .

The figure illustrates the strong decrease in this factor for the first values of N before the curve starts to flatten. While other measures of model quality, e.g., the standard deviation or parameter covariance are contributing factors in a regression setting, this effect is representative of the behavior of an interval as the number of observations increases. For the experimental effort invested in improving parameter estimates and model quality this means that at some point no large improvements can be achieved in the interval width and, in turn, the PAR. Then, tackling the second decisive element, the intermediate acceptance criteria, might be more rewarding.

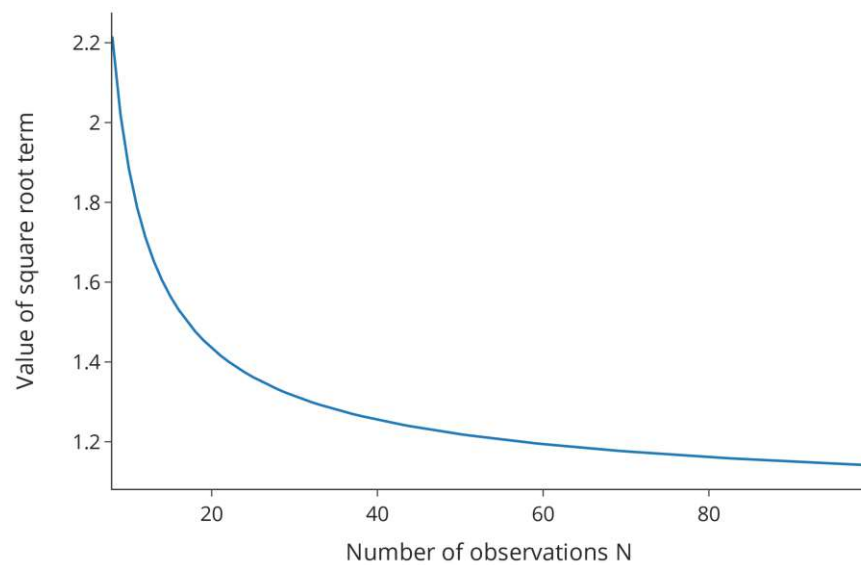


Figure 2. Value of the term $\sqrt{\frac{\nu(1+(\frac{1}{N}))}{\chi^2_{\alpha,\nu}}}$ over a range of values for N .

1.2. Option B: Improving Acceptance Limits via Spiking Studies

In typical biopharmaceutical process development and characterization, unit operations are studied individually, and acceptance criteria need to be defined for each UO. Those intermediate acceptance criteria (iAC) are the second component affecting the PAR calculation, as shown in Figure 1. A frequently followed but flawed approach to setting iACs is to calculate this range using three standard deviations (SD) of manufacturing scale runs [11–13]. A much more scientifically sound method has recently been published where the only requirement is to have drug substance/product specification. In this approach, an IPM is used, and iACs for all UOs can be calculated inversely, starting from the specifications [11]. The same approach is shown here in an illustrative manner: one can correlate the inputs/loads and outputs/pools of each unit operation, as shown in Figure 3. If the slope of this correlation equals one, all of the load will be found in the pool, which is not desirable in a downstream UO. If the slope equals zero, the same (low) pool value will be achieved regardless of the load values, which is an ideal and robust scenario of a downstream UO. We can now calculate the iAC by backpropagating the iAC of the next UO (starting with the DS specifications) through those models. By iteratively applying this technique, the iACs for the entire process can be calculated. As all models are data-based, conservative extrapolation needs to be taken into account when making predictions outside the observed training space of the explanatory variables (here, the input material of each model). As discussed in [14], for impurities, we assume that every additional amount of loaded impurity beyond the observed range will be directly propagated into the output/pool of the UO, which can be mathematically seen as a piecewise regression model with a slope of one (see Figure 3).

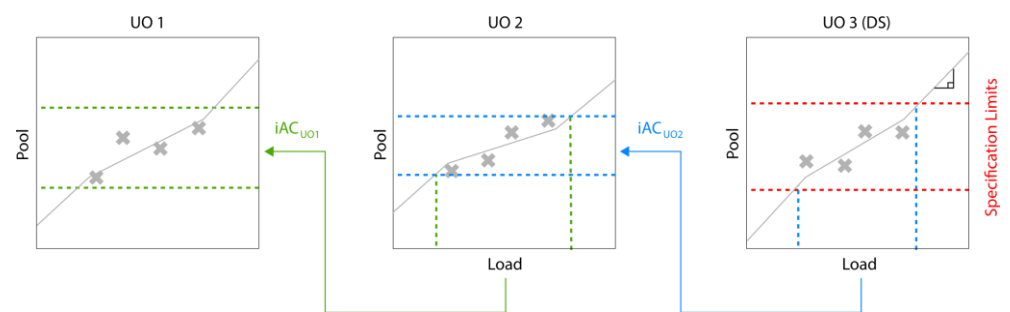


Figure 3. Inverse iAC calculation, beginning with drug specifications at the last UO.

Figure 4 illustrates the effect of adding a spiking run that shows successful clearance. The data point is added at UO 3 and the observed trend will be extrapolated, leading to an increase in iAC in UO 2. Since the iAC of UO 2 is used to calculate the iAC of UO 1, its iAC will be increased as well, etc. Hence, introducing a spiking run at one UO will potentially lead to increase in iACs of all previous UOs. Of course, the addition of spiking runs at extreme levels will be limited by the clearance capacity of the downstream UO. Note that Figures 3 and 4 show a simplified version for calculating iACs and that the actual method can involve more advanced statistical methods such as tolerance intervals and Monte Carlo sampling (for details, see [11]). Furthermore, while a linear correlation is assumed here, any mathematical model can be used to describe the dependency between input and output of individual UOs.

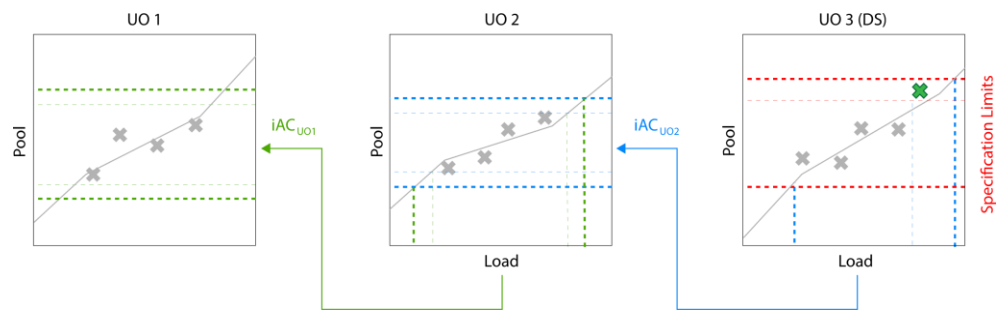


Figure 4. The effect of adding an additional data point on the iAC calculation.

Using spiking runs is not a novelty and commonly used in the industry [15,16]. However, results are usually only reported in documents and in our experience do not find entrance into mathematical modeling that also accounts for the uncertainty around individual experiments. Marschall et al. describes how any manufacturing or small-scale data can be used to calculate iACs and how spiking runs are included in that procedure [11].

At this point, we have demonstrated that both the addition of DoE runs to decrease model uncertainty as well as the addition of spiking runs to increase iACs can help to gain process understanding, which helps to increase PARs and facilitates a more flexible control strategy. However, it remains unclear which combination of DoE or spiking runs would give the maximum gain in PAR. Therefore, we want to:

- Develop a recommender system, called the holistic design of experiments (hDoE), that suggests the optimal runs (DoE or spiking) at specific UOs that lead to the fastest increase in process understanding. Here we define process understanding as the accuracy and precision of the (unknown) true relation between all PPs and CQAs, as well as the input/output relation of individual UOs. We describe this method in Section 3.1;
- Demonstrate that using such a recommender system can lead to a significant reduction in the required number of total runs of a process characterization study (PCS) using state-of-the-art workflows. We verify this in a set of simulation studies presented in Section 3.2;
- Identify an accelerated workflow for PCS using hDoE that can be applied in practice; see Section 4.1.

2. Materials and Methods

2.1. Optimal Designs

Factorial or fractional factorial designs are generally considered the ideal approach to creating experimental designs that yield the most information about how process parameters affect the response. An exhaustive account of such designs can be found in [5]. In practice, however, they are not universally applicable as they often require many runs and cannot incorporate existing data. Optimal designs constitute a more flexible alternative [17–20]. The number of runs required is not a consequence of the chosen type of

design, e.g., 2^k , for a factorial design with two levels and k parameters but can be chosen more flexibly. Furthermore, optimal designs can be used to augment an existing set of runs. Both of those properties are important for their application in hDoE, as the procedure starts off with a minimal set of experiments far smaller than a full factorial design, which gets augmented in each experiment/evaluation cycle. Based on the working set of already performed experiments, new runs based on optimal designs are recommended. Optimal designs optimize specific properties of the design matrix X . For example, D-optimal designs minimize the variance of parameter estimates in a model. As $Var(\hat{\beta}) = \sigma^2(X^T X)^{-1}$, this is equivalent to maximizing $|X^T X|$, the determinant of the squared design matrix and one can see that this is maximized when the columns of X are orthogonal. However, in contrast to factorial designs, strict orthogonality is not required and one consequence of that is that the number of runs in X can be set freely depending on the use case. The rows in X are then chosen by exchange algorithms from a candidate set of runs, which generally consists of all possible level combinations for the main effects defined in the model. We used augmented D-optimal designs to generate the experiments recommended by hDoE.

2.2. Integrated Process Model

An integrated or holistic process model is an in-silico representation of a manufacturing process comprised of multiple steps or unit operations. While there are many approaches to constructing process models (see [21] for an overview), here we consider the IPM as an empirical ensemble model implemented as a sequence of UO models that enable predictions of different CQAs as a function of process parameters. To simulate CQA concentrations, the predicted values are passed on to the next UO as a process parameter in a Monte Carlo simulation that randomly draws parameter values [22]. This is done over the entire chain of UOs in the process, from upstream operations to the final drug substance. For a comprehensive description of the method, please consult [14].

An important aspect of this approach is extrapolation. Because parameter values are drawn randomly from their corresponding distribution, CQA predictions of one UO regularly exceed the range of values observed in the training data of the predecessor UO. This is a problem for the conservative prediction of CQA concentrations, as UOs are modeled as data-driven regression models. Such models are generally only valid within the range of the training data and extrapolation can lead to highly biased results [23].

2.3. Simulation Study

We investigated the performance of hDoE in a set of simulation studies illustrating different situations and processes. Results are compared to a state-of-the-art (SOTA) process characterization workflow that consists of conducting experiments based on optimal designs to investigate the impact of PPs on CQAs per UO. For the reference method, a D-optimal design was chosen with 6, 12, and 23 runs per UO, which leads to 24, 48, or 92 runs overall in 4 UOs.

Each process in this simulation study consists of a sequence of unit operations represented by ground truth equations that map PPs to a CQA. The first UO can be interpreted as the fermentation step, followed by three downstream UOs. The equations that describe these UOs satisfy IPM conditions by adhering to the heredity principle [24] and having linear load dependencies. We then try to find effect coefficients in the presence of added noise, employing both hDoE and the SOTA method that uses a predefined number of runs per UO, as described above. For hDoE, we start with a minimal design of 6 runs per UO (total 24 runs for 4 UOs) and add an additional 30 runs, chosen by the recommender system. Note that for the simulation study results reported here, only a single run was recommended per cycle, though results are similar for larger sets of run recommendations. To calculate OOS rates, an upper drug substance specification for the output/pool of UO 4 was set as three standard deviations above the mean of the ground truth process. Hence, when adding an infinite number of runs, a minimum OOS rate of 0.00135 ($[1-0.9973]/2$) can be achieved. As the OOS rate simulated by the IPM is based on drawing random

values from a PP distribution, mean and variance must be specified. In this normalized setting, each PP's setpoint, i.e., the mean of the distribution, was chosen to be zero and the variance was set to be the same as that of the observed ground truth data that were used to derive specification limits. This was kept constant over all simulated hDoE steps to avoid a misleading optimization trajectory that improves OOS rates by simply reducing PP variances without increasing process knowledge. Each simulation scenario was repeated 100 times with different random seeds. A summary of simulation parameters is provided in Table 1.

Table 1. Parameters for the simulation study.

Parameter	Value
Number of UOs	4
Parameters per UO	5
CQA type	Impurity
hDoE start runs	6
Noise/std ratio for residual error	0.5 (0.9 in study D)
Number of runs recommended per cycle	1
Variable selection method	Bi-directional stepwise
<i>p</i> -value threshold for including effect	0.25
<i>p</i> -values threshold for excluding effect	0.05

The number of hDoE steps, repetitions of the simulation studies and the bi-directional stepwise variable selection method [25] for (re)fitting IPM models were chosen as a compromise between accuracy of results and simulation runtime.

For demonstration purposes but without loss of generality, we employ all simulation studies only for one CQA of the product. However, the methodology is not limited to the number of investigated CQAs. In practice, one would focus on the CQA, which shows the highest OOS rate.

2.3.1. Study A: Baseline

This simulation study represents a typical biopharmaceutical process with some quadratic and interaction effects and coefficients commonly found in characterization studies. In our experience, approximately 20–40% of all possible effects are practically significant in a model. A total of 5 factors lead to 20 effects (main, 2-factor interaction, and quadratic). In the ground truth, we have chosen 4–6 active effects, which equals 20–30% of all possible effects and is, therefore, within the expectation of a representative biopharmaceutical process. Moreover, we assume linear load dependencies and heredity between main and higher-order effects, which is also representative of the authors' experience. The ground truth equations for this scenario are as follows:

$$y_{UO1} = 8.0 + 0.7 x_1 + 0.6 x_2 + 0.5 x_3 - 0.4 x_4 + 0.9 x_2^2 + 0.8 x_3^2 \quad (2)$$

$$y_{UO2} = 3.4 + 0.5 y_{UO1} + 0.5 x_2 + 0.3 x_3 + 0.5 x_5 + 0.7 x_2 x_3 + 0.4 x_2 x_5 \quad (3)$$

$$y_{UO3} = 3.0 + 0.3 y_{UO2} + 0.4 x_1 - 0.3 x_2 + 0.2 x_4 - 0.2 x_5 + 0.3 x_1^2 - 0.7 x_5^2 \quad (4)$$

$$y_{UO4} = 2.8 + 0.2 y_{UO3} + 0.1 x_1 + 0.2 x_3 + 0.2 x_5 + 0.3 x_1 x_3 \quad (5)$$

2.3.2. Study B: Load Effect Set to One

As hDoE leverages the UOs dependency on the load, we investigated its behavior when the load coefficient is set to one, and its values are passed directly to the output of UO 3, provided other PPs are at setpoint. This mimics the situation where a full propagation of the CQA through this UO is expected, and no clearance takes place. This is, of course, not

the desired behavior of a downstream unit operation. In this setting, the load coefficient in UO 3 was set to one:

$$y_{\text{UO}3} = 3.0 + 1.0 y_{\text{UO}2} + 0.4 x_1 - 0.3 x_2 + 0.2 x_4 - 0.2 x_5 + 0.3 x_1^2 - 0.7 x_5^2 \quad (6)$$

2.3.3. Study C: All Load Effects Set to One

Here we set the load effects of all UOs to one. Every UOs output is directly propagated to the next UO, which means that there is no information about the load that could be detected by spiking runs. This a very untypical scenario as usually, we expect some clearance activity of the downstream UOs (UO 2–4). The purpose of this simulation is to show that, in the worst case, hDoE performs similarly to standard approaches using a predefined set of runs.

$$y_{\text{UO}2} = 3.4 + 1.0 y_{\text{UO}1} + 0.5 x_2 + 0.3 x_3 + 0.5 x_5 + 0.7 x_2 x_3 + 0.4 x_2 x_5 \quad (7)$$

$$y_{\text{UO}3} = 3.0 + 1.0 y_{\text{UO}2} + 0.4 x_1 - 0.3 x_2 + 0.2 x_4 - 0.2 x_5 + 0.3 x_1^2 - 0.7 x_5^2 \quad (8)$$

$$y_{\text{UO}4} = 2.8 + 1.0 y_{\text{UO}3} + 0.1 x_1 + 0.2 x_3 + 0.2 x_5 + 0.3 x_1 x_3 \quad (9)$$

2.3.4. Study D: Disabled Probability-Ratio-Threshold

To highlight the importance of the probability-ratio-threshold (PRT) decision scheme, a method borrowed from the Metropolis–Hastings algorithm that encourages the detection of new effects (described in Section 3.1), we repeat the baseline study without PRT. Here, the decision logic simply recommends the type of run that leads to the largest reduction in OOS. As PRT is most effective in situations where no clear decision can be made due to high residual error, the error/standard deviation ratio in the ground truth was increased from 0.5 to 0.9.

3. Results

3.1. Holistic Design of Experiments

In this contribution, we propose a new tool for process development and characterization: holistic design of experiments (hDoE), an iterative approach to experimental design and evaluation that minimizes the number of runs invested while maximizing the overall process understanding, as defined in the introduction. As all UOs of the process contribute to the generation of DS material, we can boil down process understanding to how well the true distribution of DS product quality is known. To express this in a single measure that can be used for optimization, we chose the out-of-specification (OOS) rate based on known DS specifications.

As described in Section 2.2, the IPM connects UOs by passing the output of a UO to the next one as a load parameter. While the output might be affected by different parameters and interactions, the univariate relationship between load and output is assumed to be linear inside the range of observed load values in the training data. However, in the Monte Carlo simulation of a CQA distribution, simulated load values might exceed this observed range, in which case the CQA value is handled conservatively to avoid predictions that are too optimistic (see piecewise load model in Taylor et al. [14]). This means that the simulated CQA distribution in drug substances is highly dependent on the range of load parameter values in the training data. Small variations in the training data will lead to a broadened CQA distribution, accounting for the uncertainty due to extrapolation. However, in many cases, missing information about load parameters can be supplemented by performing spiking studies where PPs are kept at setpoint, and only the load is varied. This is in contrast to classic DoE studies, which examine specific combinations of factors and disregard the load.

Spiking and DoE runs are the two different types of the experiment proposed by hDoE in this contribution. Assuming model parameters do not change from one iteration to the next, DoE runs are expected to improve general process parameter estimates, while

spiking studies improve estimates of the model's load coefficient and reduce extrapolation in the IPM simulation. However, the assumption of unchanging model parameters is regularly violated when variable selection on newly acquired data results in a new model. As a consequence, the simulated OOS probability is not guaranteed to be improved in every iteration.

hDoE starts out by roughly characterizing the process with a minimal D-optimal design per UO that facilitates fitting the initial regression models. Based on the information acquired in this first step, a set of runs is proposed. This set consists of DoE and spiking runs and includes the target UOs in which to perform them. The runs are chosen by how much they would reduce the OOS probability calculated by the IPM. After conducting the proposed experiments, the IPM data are supplemented with new information, and a variable selection step updates models where appropriate.

As OOS predictions and therefore run suggestions are based on models found in previous steps, the process is biased toward already detected effects. Additionally, the decision rule is susceptible to noise, especially in early steps, and might consider the value of adding spiking or DoE runs equivalent. Of course, this is generally not true, as only DoE experiments enable the detection of new or interaction effects. To mitigate this bias and encourage the detection of new effects, we employ a technique based on a decision scheme used in the Metropolis–Hastings algorithm [26]. Let X be the design matrix of the data already incorporated into the IPM and x_{DOE} and $x_{spiking}$ be new DoE and spiking samples, respectively, chosen from a set of sample candidates that result in the lowest OOS probability. We then calculate the ratio of those probabilities $\alpha = P(x_{DOE}|X)/P(x_{spiking}|X)$, draw a number from a uniform distribution, $u \in [0, 1]$, and only suggest a spiking run when $u \geq \alpha$. This means that DoE runs are always recommended if $P(x_{DOE}|X) \geq P(x_{spiking}|X)$. Spiking runs, however, are only suggested when the improvement in OOS probability considerably exceeds that of a DoE run. In the following, we term this the probability-ratio-threshold (PRT) approach. In our evaluation of the algorithm, PRT generally circumvents the problem of selecting spiking runs unnecessarily or overlooking effects (see Section 3.2). Figure 5 illustrates the individual steps and decision processes involved in hDoE.

3.2. Simulation Results

3.2.1. Out-of-Specification Rates

Figure 6 shows that in most scenarios, a high process understanding (quantified as a low OOS rate) can be achieved with a much smaller number of experiments compared to the SOTA method that uses a fixed number of DoE runs (in this case, 24, 48 and 92). This effect is most pronounced in study A, where the mean OOS rate drops to ~2.5% after only six additional runs recommended by hDoE (30 total), whereas 92 D-optimal runs calculated beforehand result in an OOS rate of ~7.5% due to the lack of exploration of load-to-pool dependencies. At first glance, this might appear as an unfair comparison, but workflows applied in the industry generally do not incorporate spiking runs in a mathematical framework to achieve a specific goal, e.g., establishing a control strategy. The effect of an UOs dependency on the load can also be seen in the results of study B, where variation around the OOS rate is larger due to setting the load coefficient to one in UO 3 in the ground truth. This means that no additional information about the load can be acquired in this UO, and the advantage of systematically recommending spiking studies at this UO is softened. Study C represents the worst case for hDoE, where the load has no effect in any unit operation, resulting in OOS rates close to that of the reference method. The larger variation shown in study D is due to increased noise in the ground truth. This, of course, affects hDoE as well as the reference method. However, note that variation is drastically increased when PRT is disabled (purple, dotted lines) compared to the recommended procedure that uses it when deciding on runs (blue, dotted lines).

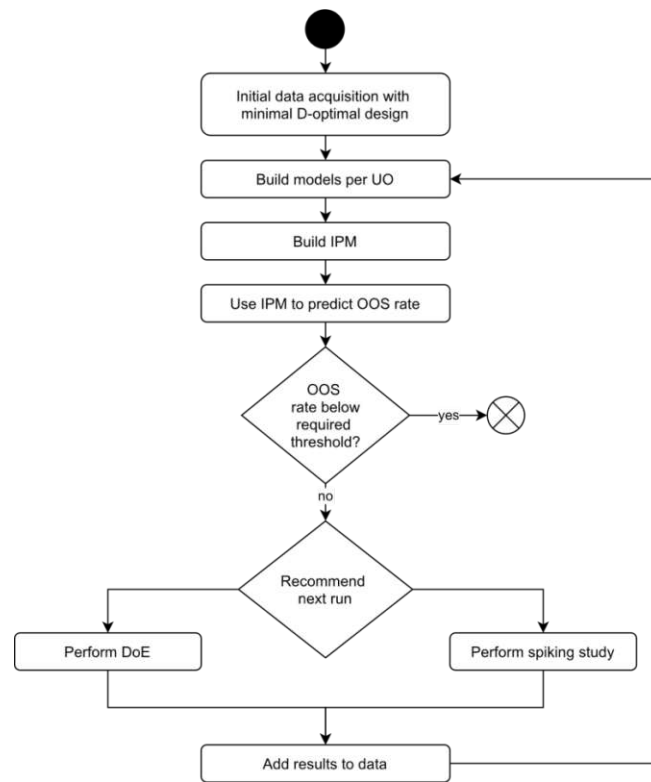


Figure 5. The hDoE recommender system that minimizes OOS by iteratively augmenting data and predicting optimal experiments.

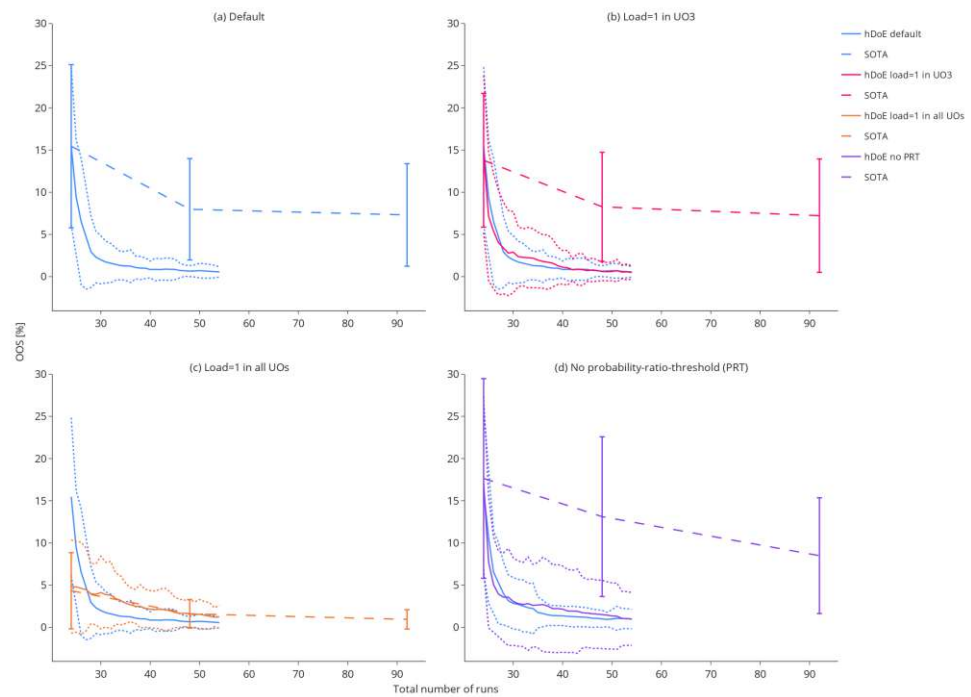


Figure 6. The decrease in OOS probability over the 30 hDoE steps for the different simulation scenarios. Each scenario was repeated 100 times. The solid line indicates the mean OOS probability, while the dotted lines show the standard deviation in the 100 repetitions of that step. For the reference method, OOS probabilities are plotted as dashed lines and were calculated at 24, 48, and 92 total runs.

3.2.2. Run Allocation

The drastic decrease in OOS probability over the number of experimental runs invested, shown in Figure 6, is achieved by performing spiking studies at the right UO. Generally, spiking runs are favored in the early steps of the procedure, as they eliminate extrapolation in the IPM (see Section 2.2) and therefore lead to the largest reduction in OOS early on. After this initial phase, larger OOS improvements can be achieved by improving parameter estimates, prompting hDoE to suggest more DoE runs. Figure 7 illustrates the allocation of different run types to the four unit operations of the simulation study. In the *y*-axis, the plots show the cumulative number of allocated runs over 100 repetitions of a simulation study, while the corresponding hDoE step can be seen in the *x*-axis. As the first UO (e.g., fermentation) is not affected by a load parameter, no spiking runs (dotted lines) are allocated. In the three simulation studies where the load influences a UO, spiking runs at UO 4 are recommended in the early steps and, in many of the 100 iterations, also in UO 3. This makes sense, as the OOS probability in drug substance, i.e., the last UO, is the main driver of the recommender system and its load coefficient directly affects the CQA distribution. In the absence of load effects, DoE runs are distributed approximately equally, while some spiking runs are accumulated due to noise (orange lines). The effect of increased noise and the absence of PRT are illustrated by the purple line. Much more spiking runs are invested, as the OOS simulation for the two types of experiment yields similar results, and choices are made more randomly.

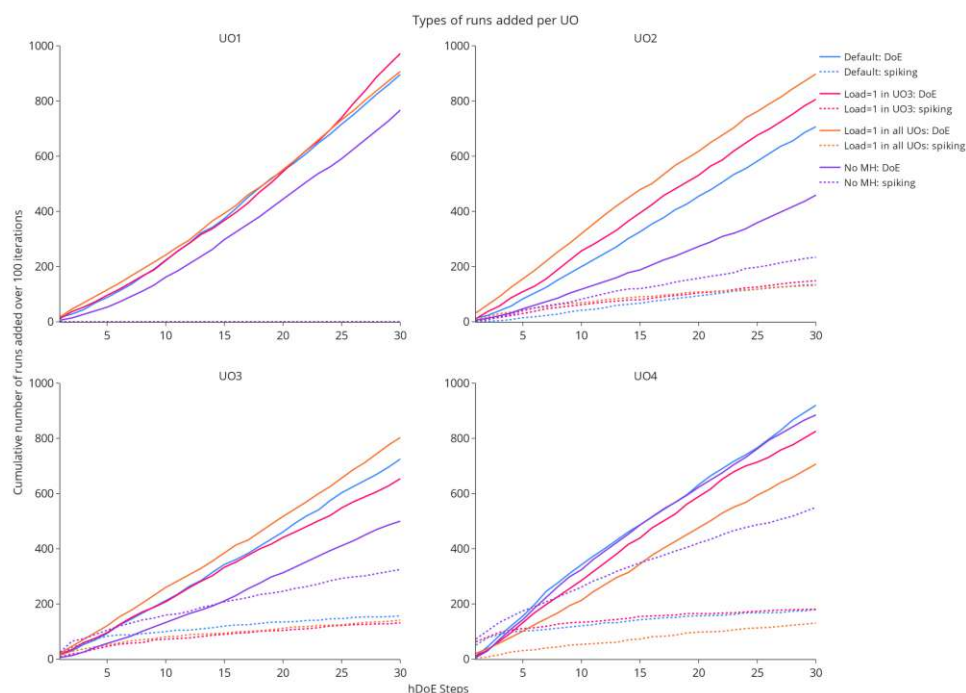


Figure 7. The cumulative allocation of either spiking or DoE runs in 100 repetitions of the simulation is shown over each step taken by the hDoE procedure (after the initial 24 runs). In total, each study distributes $100 \times 30 = 3000$ runs of any type to the four unit operations.

3.2.3. Parameter Estimates

A low OOS rate alone does not indicate correct models, as it does not account for aliasing effects in the parameters. Figure 8 shows the distributions of effects identified in the variable selection procedure at the last hDoE step (54 total runs invested). Each data point represents the effect size in a particular iteration of the simulation. We added an estimator of the residual variance in the form of the root mean squared error (RMSE). The mean estimated value (blue dashed line of each boxplot) converges to the ground truth (red solid line of each boxplot) for most parameters. Higher-order effect estimates in unit operations two and three are biased toward zero, i.e., they were not detected in variable

selection. Note that stepwise variable selection was used for performance reasons in this simulation, which is known to eliminate effects prematurely [27], and that some of this bias could be mitigated by using more modern approaches such as leaps and bounds or other exhaustive algorithms. However, as tolerance intervals are used in the estimation of model uncertainty [14], overlooking individual effects, which results in larger estimates of the RMSE, is accounted for correctly in the uncertainty interval. Of course, the quality of parameter estimates is a direct consequence of the number of DoE runs invested, which in turn is dependent on when the hDoE procedure is terminated.

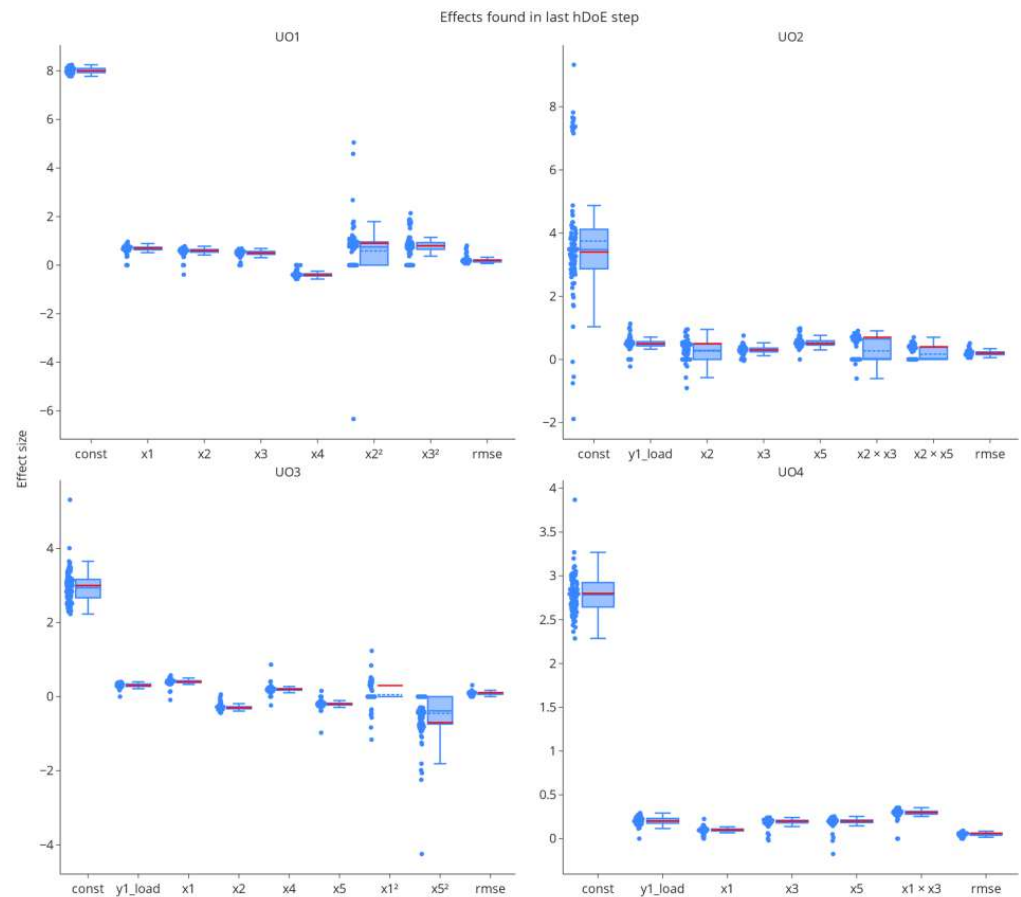


Figure 8. The distribution of effects found by stepwise variable selection at the last hDoE step. Each point represents the effect size in one of the 100 repetitions of the simulation study. The RMSE was added as an estimator of residual error.

Parameter estimates, especially for higher-order effects, are naturally not on par with those from a full D-optimal design with 23 runs per unit operation, see Figures 8 and 9. However, due to spiking studies invested in the characterization of the process, hDoE results in the improved estimation of load and intercept effects.

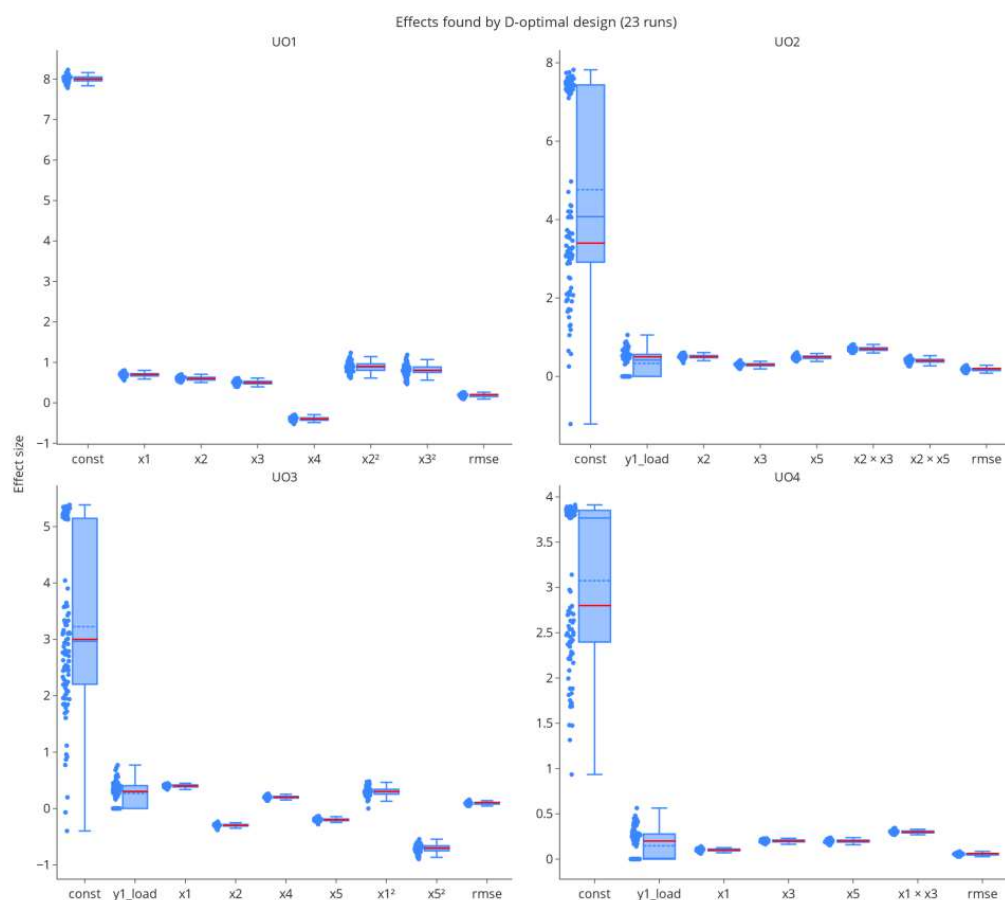


Figure 9. Parameter estimates for the full D-optimal design with 23 runs per UO. Variance in estimates is smaller than in hDoE. However, intercept and load effects are biased due to missing load data.

4. Discussion

4.1. hDoE in Process Characterization

Figure 10 shows how hDoE affects common steps in process characterization. The result of this procedure is a control strategy, of which PARs are an essential component. We describe in Section 1 the different courses of action when a PAR is too narrow to be part of an appropriate control strategy, steps that are also reflected by the hDoE recommender system. A third option that is currently not incorporated into the recommender system is to change the setpoint of other PPs that are active in the UO model. We present this approach as an outlook in Section 4.2. Finally, PAR ranges can be increased by reducing confidence/coverage levels of the statistical interval, although in most cases, this is not recommended and only mentioned here for the sake of completeness. We consider the workflow shown in Figure 10 as an extended version of the state-of-the-art workflow for process characterization (left column) that incorporates hDoE (right column, green boxes). This aligns with both the FDA and EMA guidelines for process characterization, as the former recommends DoE to increase process knowledge [3] and the latter impurity spiking challenges for downstream operations [28]. hDoE employed in the context of this extended workflow integrates both types of experiments and provides a systematic method for when and where to perform them.

4.2. Outlook: Changing PP Setpoints to Increase the PAR

An important aspect of the PAR calculation illustrated in Figure 4 is the univariate nature of this method. Only the screening range of the current PP is considered, while all other PPs are kept at their setpoint. As the UO model output, i.e., the CQA, is in most cases affected by multiple PPs, their setpoint can influence the offset of the univariate mean

prediction of the target PP (orange line in the figure) significantly. Consequently, a change in the setpoint of another PP can push the predicted CQA distribution inside/outside the acceptance limits and change the OOS rate. Similarly, interaction effects with other PPs can also influence the target PPs effect.

The optimization of PP setpoints is already available in some statistical software [29], and hDoE could be easily extended to include such recommendations based on their effect on the OOS rate of the process. While this is generally not the focus of process characterization studies, the FDA recommends optimization based on setpoint shifts in the continuous verification phase [3]. This would advance the applicability of the hDoE recommender system into the domain of process optimization.

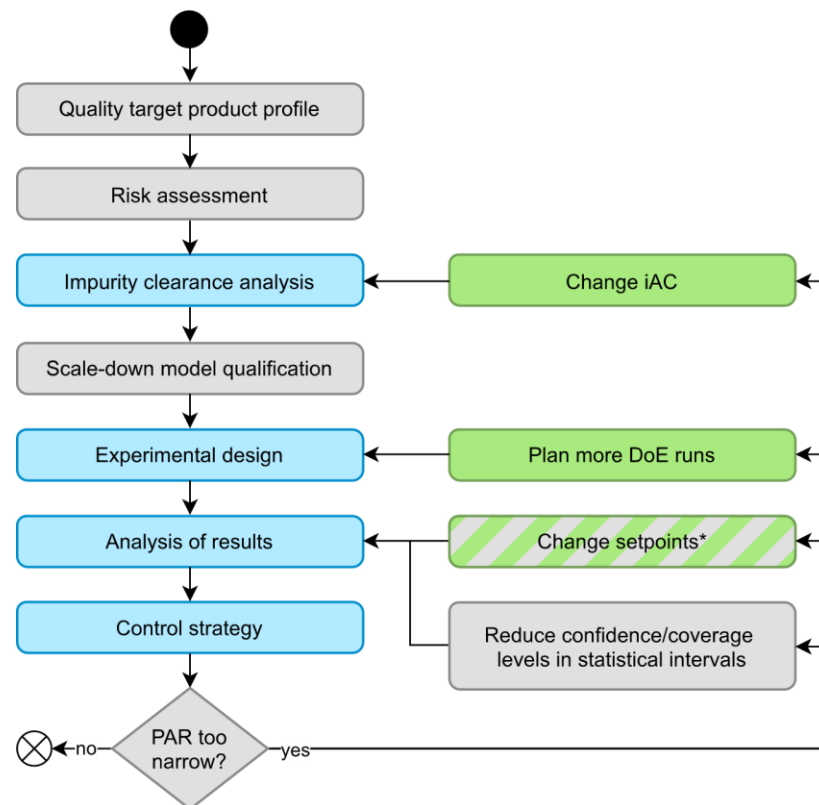


Figure 10. The left column represents common steps involved in process characterization, while the right column contains possible courses of action when no appropriate control strategy can be found. Green boxes are actions recommended by hDoE, and steps affected by hDoE are colored blue. * Changing PP setpoints could also be a viable, recommended action by future iterations of hDoE (see Section 4.2).

5. Conclusions

In this article, we introduced hDoE, an iterative tool for process development and characterization that facilitates a more effective way of gaining process understanding related to final product quality. This is essential, as it reduces experimental effort and time to market. To demonstrate the benefits of this approach quantitatively, we presented results from simulation studies where we chose the OOS rate as a measure of process understanding, which should be close to the OOS rate of the true (usually unknown) process. The benefit of this measure is that it includes both the mean and variability of the final product quality distribution as well as practically relevant limits (drug specifications). Other measures, such as the Kullback–Leibler divergence [30], also could have been used. However, we believe the OOS rate represents a more practically relevant measure and might be more tangible for process experts. We have demonstrated that hDoE leads to better overall process understanding with more than a 50% reduction in the number of

experiments performed for simple scenarios. The reduction of experimental costs can even be increased for specific cases. hDoE starts out with an initial, minimal set of D-optimal runs on which the first set of unit operation regression models is fitted. Used as a recommender tool, either DoE or spiking runs are added in an iterative fashion guided by improvements in the predicted OOS rate. As the process is biased toward effects already found, we are using a recommender scheme akin to the one used in the Metropolis–Hastings algorithm to promote the detection of unknown effects and to improve parameter estimates. Of course, the overall quality of effect estimates is influenced by the number of runs available to the algorithm. However, our simulation studies show that a compromise between effects detected and runs invested can be found using a relatively low number of hDoE runs.

hDoE leverages the link between UOs as modeled by the IPM and thereby improves OOS rates by strategically recommending spiking studies at specific process steps. It provides valuable information to biopharmaceutical manufacturers about which type of experiment to perform next, and in which UO, and can decrease the time and money invested in experimental design. Potentially increased parameter uncertainty due to a lower number of samples is accounted for in the IPM's OOS prediction using conservative sampling and estimation methods such as tolerance intervals. We believe that hDoE is a viable option for experimental design that yields robust estimates of process properties while providing better control of the resources invested. This will lead to a substantial reduction of development costs and time to market, ultimately leading to more affordable drugs.

Author Contributions: Conceptualization, T.Z. and B.P.; methodology, T.Z. and T.O.; software, T.O. and B.P.; validation, B.P. and T.O.; formal analysis, T.O.; investigation, T.O.; data curation, T.O.; writing—original draft preparation, T.O. and T.Z.; writing—review and editing, T.O., T.Z., B.P. and C.H.; visualization, T.O.; supervision, C.H. and T.Z.; project administration, T.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Austrian Research Promotion Agency (FFG) (grant number: 844608) and within the framework of the Competence Center CHASE GmbH, funded by the Austrian Research Promotion Agency (grant number 868615) as part of the COMET program (Competence Centers for Excellent Technologies) by BMVIT, BMDW, and the Federal Provinces of Upper Austria and Vienna. Open Access Funding by TU Wien.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data generated or analyzed during this study are included in this published article.

Acknowledgments: This work was conducted within the COMET Centre CHASE, funded within the COMET—Competence Centers for Excellent Technologies program by the BMK, the BMDW and the Federal Provinces of Upper Austria and Vienna. The COMET program is managed by the Austrian Research Promotion Agency (FFG). The authors acknowledge TU Wien Bibliothek for financial support through its Open Access Funding Program.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. ICH. *ICH Guideline Q8 (R2) on Pharmaceutical Development*; EMA: London, UK, 2017.
2. Burdick, R.; LeBlond, D.; Pfahler, L.; Quiroz, J.; Sidor, L.; Vukovinsky, K.; Zhang, L. *Statistical Applications for Chemistry, Manufacturing and Controls (CMC) in the Pharmaceutical Industry*; Springer: Cham, Switzerland, 2017.
3. FDA. *Process Validation: General Principles and Practices*; US FDA: Rockville, MD, USA, 2011.
4. Montgomery, D.C.; Peck, E.A.; Vining, G.G. *Introduction to Linear Regression Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 2021.
5. Montgomery, D.C. *Design and Analysis of Experiments*; John Wiley & Sons: Hoboken, NJ, USA, 2017.
6. EMA. *Questions and Answers: Improving the Understanding of NORs, PARs, DSp and Normal Variability of Process Parameters*; EMA: London, UK, 2017.
7. Howe, W. Two-sided tolerance limits for normal populations—Some improvements. *J. Am. Stat. Assoc.* **1969**, *64*, 610–620.
8. Krishnamoorthy, K.; Mathew, T. *Statistical Tolerance Regions: Theory, Applications, and Computation*; John Wiley & Sons: Hoboken, NJ, USA, 2009.

9. Wallis, W.A. Tolerance intervals for linear regression. In Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA, USA, January 1951.
10. Francq, B.G.; Lin, D.; Hoyer, W. Confidence, prediction, and tolerance in linear mixed models. *Stat. Med.* **2019**, *38*, 5603–5622. [[CrossRef](#)]
11. Marschall, L.; Taylor, C.; Zahel, T.; Kunzelmann, M.; Wiedenmann, A.; Presser, B.; Studts, J.; Herwig, C. Specification-driven acceptance criteria for validation of biopharmaceutical processes. *Front. Bioeng. Biotechnol.* **2022**, *10*, 1010583. [[CrossRef](#)]
12. Seely, R.; Munyakazi, L.; Haury, J. Statistical tools for setting in-process acceptance criteria. *Dev. Biol.* **2003**, *113*, 17–26.
13. Wang, X.; Germansderfer, A.; Harms, J.; Rathore, A. Using statistical analysis for setting process validation acceptance criteria for biotech products. *Biotechnol. Prog.* **2007**, *23*, 55–60. [[CrossRef](#)] [[PubMed](#)]
14. Taylor, C.; Pretzner, B.; Zahel, T.; Herwig, C. Architectural & Technological Improvements to Integrated Bioprocess Models towards Real-Time Applications. *MDPI Bioeng.* **2022**, *9*, 534.
15. Darling, A. Considerations in performing virus spiking experiments and process validation studies. *Dev. Biol. Stand.* **1993**, *81*, 221–229. [[PubMed](#)]
16. Shukla, A.; Jiang, C.; Ma, J.; Rubacha, M.; Flansburg, L.; Lee, S. Demonstration of robust host cell protein clearance in biopharmaceutical downstream processes. *Biotechnol. Prog.* **2008**, *24*, 615–622. [[CrossRef](#)]
17. Johnson, R.; Montgomery, D.; Jones, B. An Expository Paper on Optimal Design. *Qual. Eng.* **2011**, *23*, 287–301. [[CrossRef](#)]
18. de Aguiar, F.; Bourguignon, B.; Khots, M.; Massart, D.; Phan-Thau-Luu, R. D-optimal designs. *Chemom. Intell. Lab. Syst.* **1995**, *30*, 199–210. [[CrossRef](#)]
19. Goos, P.; Jones, B.; Syafitri, U. I-optimal design of mixture experiments. *J. Am. Stat. Assoc.* **2016**, *111*, 899–911. [[CrossRef](#)]
20. Jones, B.; Allen-Moyer, K.; Goos, P. A-optimal versus D-optimal design of screening experiments. *J. Qual. Technol.* **2021**, *53*, 369–382. [[CrossRef](#)]
21. Velayudhan, A. Overview of integrated models for bioprocess engineering. *Curr. Opin. Chem. Eng.* **2014**, *6*, 83–89. [[CrossRef](#)]
22. Mooney, C. *Monte Carlo Simulation*; Sage: Thousand Oaks, CA, USA, 1997.
23. Hahn, G. The hazards of extrapolation in regression analysis. *J. Qual. Technol.* **1977**, *9*, 159–165. [[CrossRef](#)]
24. Hamada, C.; Hamada, M. All-subsets regression under effect heredity restrictions for experimental designs with complex aliasing. *Qual. Reliab. Eng. Int.* **2010**, *26*, 75–81. [[CrossRef](#)]
25. Desboulets, L.D.D. A review on variable selection in regression analysis. *Econometrics* **2018**, *6*, 45. [[CrossRef](#)]
26. Hastings, W.K. *Monte Carlo Sampling Methods Using Markov Chains and Their Applications*; Oxford University Press: Oxford, UK, 1970.
27. Olusegun, A.M.; Dikko, H.G.; Gulumbe, S.U. Identifying the limitation of stepwise selection for variable selection in regression analysis. *Am. J. Theor. Appl. Stat.* **2015**, *4*, 414–419. [[CrossRef](#)]
28. Committee for Medicinal Products for Human Use. *Process Validation for the Manufacture of Biotechnology-Derived Active Substances and Data to Be Provided in Regulatory Submissions*; EMA: London, UK, 2016.
29. SAS Institute Inc. *JMP®16 Profilers*; SAS Institute Inc.: Cary, NC, USA, 2020–2021.
30. Joyce, J. Kullback-Leibler Divergence. In *International Encyclopedia of Statistical Science*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 720–722.

7.3 Scientific publication III

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.



Identifying Design Spaces as Linear Combinations of Parameter Ranges for Biopharmaceutical Control Strategies

Thomas Oberleitner,^a Thomas Zahel,^b Christoph Herwig,^{b,c}

^a Competence Center CHASE GmbH, Ghegastraße 3, Top 3.2, 1030 Vienna, Austria

^b Körber Pharma Austria GmbH, PAS-X Savvy, Mariahilferstraße 88A/1/9, 1070 Vienna, Austria

^c TU WIEN Research Area Biochemical Engineering, Getreidemarkt 9, 1060 Vienna, Austria,
christoph.herwig@tuwien.ac.at

Abstract

According to ICH Q8 guidelines, the biopharmaceutical manufacturer submits a design space (DS) definition as part of the regulatory approval application, in which case process parameter (PP) deviations within this space are not considered changes and do not trigger a regulatory post approval procedure. A DS can be described by non-linear PP ranges, i.e., the range of one PP conditioned on specific values of another. However, independent PP ranges (linear combinations) are often preferred in biopharmaceutical manufacturing due to their operation simplicity, as mentioned in the guideline. While statistical software such as Modde supports the calculation of a DS comprised of linear combinations, their algorithms are generally based on discretizing the parameter space - an approach that suffers from the curse of dimensionality as the number of PPs increases. Here, we introduce a novel method for finding linear PP combinations using a numeric optimizer to calculate the largest design space within the parameter space that results in critical quality attribute (CQA) boundaries within acceptance criteria, predicted by a regression model. A precomputed approximation of tolerance intervals is used in inequality constraints to facilitate fast evaluations of this boundary using a single matrix multiplication. The correctness of the method was validated by comparing results to that of a grid-based approach and, in a simple case, to an analytically defined ground truth. In the examples investigated, the volume of the resulting DS was significantly larger than that of the grid method, with the improvement being proportional to the granularity of the grid and the number of parameters involved. Furthermore, computational time for the optimization-based approach is several orders of magnitude faster in higher dimensions. In addition, a proposed weighting scheme can be used to favor certain PPs over others and therefore enabling a more dynamic approach to DS definition and exploration. The increased PP ranges of the larger DS provide greater operational flexibility for biopharmaceutical manufacturers.

Keywords: design space; linear combination of process parameters; biopharmaceutical development; ICH Q8; numeric optimization; parameter space

1. Introduction

The ICH Q8 guideline for pharmaceutical development defines the design space (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” [1]. The process parameters (PP) described here are generally identified in the risk assessment or process development phases and are considered critical when sufficient evidence was found that they affect the output of a unit operation, i.e., a critical quality attribute (CQA). A design space is comprised of the ranges of these process parameters that result in CQA values within acceptable limits. For the biopharmaceutical manufacturer a DS definition can be submitted as part of the regulatory approval application, in which case PP deviations within this space are not considered a change and therefore do not trigger a regulatory post approval procedure. For operators, the DS constitutes a valuable guideline document for controlling a process. While ICH Q8 does not recommend a specific form or

method for calculating a DS, it provides examples for how to present the DS as non-linear and linear combinations of parameter ranges in the form of contour plots (appendix 2c in guide). Non-linear combinations describe the DS as a set of rules, or parameter ranges conditioned on other parameters, e.g., “PP1 is allowed to move between -1 and 1 if PP2 is lower than 0.5”. Linear combinations of parameter ranges on the other hand are independent of each other. While the former description generally represents a larger space to operate in and methods for computing it can be found several publications [2, 3], the latter might be preferred due to its operational simplicity and is the subject of this contribution [1]. Figure 1 shows the different types of design space graphically.

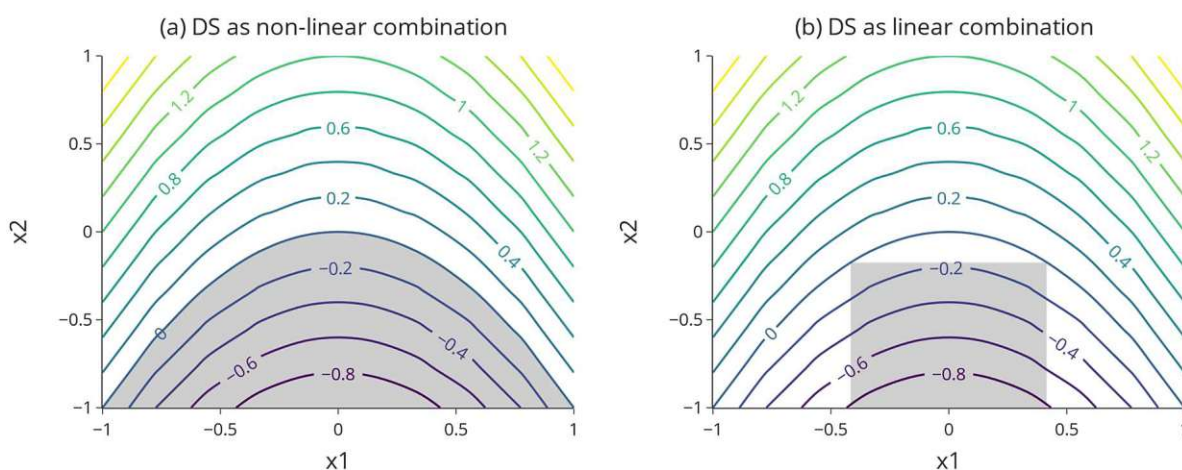


Figure 1: The design space for the function $f(x_1, x_2) = x_1^2 + x_2$, where $f(x_1, x_2) \leq 0$, shown in the contour plot as a non-linear (a) and linear (b) combination of input parameters x_1 and x_2 , as defined in [1].

In these examples, the definition of a design space does not incorporate any measure of statistical uncertainty, i.e., the contour shown in Figure 1 directly represents the predicted mean CQA values from the model. In the context of biopharmaceutical process validation, we suggest a more conservative approach. To accurately quantify uncertainty inherent in the regression model due to analytical and process variability, we replace predicted CQA values with the upper and lower boundary of a tolerance interval (TI) that incorporates nominal levels of both confidence and coverage. The statistical relevance of tolerance intervals in biopharmaceutical control strategies has been discussed earlier [4]. These boundaries are then used to validate acceptance limits in a conservative manner.

As the visual definition of a DS is only feasible in a low-dimensional parameter space, e.g., the bivariate contour plots in Figure 1, computational methods are required to find the exact parameter ranges for typical biopharmaceutical models containing 10 parameters or more. To the knowledge of the authors, this is currently not possible in state-of-the-art statistical software used in process development, such as Modde, because the common approach is to discretize the parameter space and to evaluate points on the resulting grid [5]. This method suffers from the curse of dimensionality as the parameter space grows exponentially over the number of parameters in the model [6]. Segmenting the range of 10 parameters into 10 parts would result in a parameter space of 10^{10} points, for which all possible combinations would need to be evaluated to find the largest possible DS – an exceptionally computationally expensive task. As a result, it is practically impossible to calculate design spaces for a larger number of PPs, even though this is commonly required for process development and characterization (see case study in [4]). Furthermore, the design space found by grid-based methods is generally not the one with the largest possible multidimensional volume, or

“hypervolume”, because the solution space is limited to discrete points that are distributed over the parameter range. We illustrate these drawbacks in section 4. The consequence of these problems is that no valid DS can be computed in many Quality by Design (QbD) projects and the manufacturer goes back to univariate individual controls, which leads to a loss of flexibility and prohibits process optimization.

There is extensive research on the identification of a DS that incorporates uncertainty. Employing the categorization from Kusumo et al., approaches generally differ in terms of how model uncertainty is derived as well as how the resulting DS should be represented [3]. Uncertainty can be expressed (i) by confidence levels from frequentist estimators of parameters [7, 8, 9] or (ii) as distributions derived from sampling methods or Bayesian posteriors [10, 3, 11, 12]. On the other hand, complex DS contours can be described by a set of points that satisfy acceptance criteria [3, 13] or as simple shapes, such as rectangles [1, 9]. Depending on the use case, any combination of these options can be viable and informs the choice of approach to identifying the DS. Arguably the simplest algorithms discretize the parameter space and evaluate an underlying model at each point [14]. Even though such schemes are computationally expensive, they are employed in software due to their simplicity [5]. Bayesian methods use more refined sampling methods such as Markov Chain Monte Carlo to update their estimators for the posterior distribution of the CQA [3, 12, 15]. Finally, optimization-based methods formulate the problem of finding a DS in way that facilitates the use of standard optimization algorithms [9, 16].

The methods referenced here are generally based on the idea of a probabilistic design space, defined by Peterson as the set $\{x: \Pr[y \in A|x, data] \geq R\}$, where x is the vector of PPs, y the modelled response, A the acceptance region and R the probability threshold to be cleared [11]. This is not the type of probability employed by the method introduced in this article. Instead, regression model tolerance intervals are used, which explicitly to quantify model uncertainty as well as the population distribution of the predicted response. Consequently, uncertainty quantification is decoupled from the optimization procedure, facilitating the use of any model, provided a TI can be defined, and a wide range of optimization algorithms. Furthermore, TIs include a nominal threshold for population coverage in addition to a confidence level that encapsulates model uncertainty, adding a further option for estimating CQA distributions conservatively, as desired when defining biopharmaceutical control strategies. Another distinguishing feature is that the objective function for the optimization procedure is directly based on the linear combination of PP ranges, i.e., no post-processing of sampled points [17], nested optimization [9] or visual inspection of the results [1] is necessary. This effectively combines the problems of identifying the DS and inscribing the largest possible hyperrectangle into a single optimization problem.

In this contribution, we present a novel method for finding a design space comprised of linearly independent parameter ranges while treating CQA predictions conservatively by evaluating the tolerance interval boundaries around them, i.e., checking whether those boundaries fall within acceptance criteria. We propose a method based on numeric optimization rather than the screening of a grid, which results in improved computational times and, in many cases, a larger volume of the DS. An approximation for tolerance intervals is presented as well as a method for incorporating categorical model factors in a computationally efficient manner. Finally, a weighting scheme to favor certain PPs over others facilitates increased flexibility in the calculation of the results. This facilitates a method for DS computation and exploration that is computationally more efficient and provides greater flexibility than the state-of-the-art, both in terms of controlling the DS computation as well as greater flexibility for operators due to increased PP ranges.

2. Methodology

2.1. Regression Models in Biopharmaceutical Development and Manufacturing

Due to their simplicity and optimal statistical properties, regression models are a popular choice in biopharmaceutical development and manufacturing, where they are used to express the relationship between PPs and CQAs. One of its more basic representatives is the ordinary least-squares (OLS) model, which assumes a polynomial relationship between parameters and response and normally distributed residuals [18]. The model equation takes the form:

$$y = X\beta + \epsilon \quad (1)$$

Where y is the model output (here, the CQA), X an $n * p$ matrix of n observations comprised of p parameter settings (PPs). The formula also includes the vector of model coefficients β that assigns each parameter a numeric value (a parameter's effect) and the residual error $\epsilon \sim N(0, \sigma^2)$. The least-squares fit is obtained by finding an estimator for the model coefficients. In the case of OLS, can be calculated in the following way:

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad (2)$$

Under the assumption that $\epsilon \sim N(0, \sigma^2)$, this definition yields the best linear unbiased estimator (BLUE) for β . Note that in practice, the assumption of normally distributed residuals might not be satisfied, in which case other types of regression models might be employed where $\hat{\beta}$ is acquired using numerical optimization rather than a closed-form expression [19] [20] [21]. The method for DS calculation proposed in this article is generally agnostic about the type of model used for representing the relationship between PP and CQA as long as tolerance intervals can be defined for its predictions. For the sake of simplicity and their frequent application in biopharmaceutical process development, we chose OLS models for the simulation studies in 4.

2.2. Tolerance Intervals

Tolerance intervals (TI) are used to quantify uncertainty of the predicted mean of a model. Its purpose is to estimate the population distribution of the model's predicted response, given the uncertainty associated with the modelling and sampling process. Using such conservative estimators is especially important in the biopharmaceutical domain where a majority of the future population of runs needs to be within process or specification limits - hence the usage of confidence or prediction intervals is not recommended [4]. As defined in [8], tolerance intervals estimate the following probability:

$$\Pr[\Pr[y|x_{n+1} \in I_T(x_{n+1})] \geq 1 - \psi] = 1 - \alpha \quad (3)$$

The inner probability expresses whether the model's response y for a new observation x_{n+1} is contained within the true population distribution $I_T(x_{n+1})$ in a least $100(1 - \psi)\%$ of repeated samplings from the reference distribution, whereas the outer probability represents the confidence level of $100(1 - \alpha)\%$. Consequently, tolerance intervals contain nominal parameters for both the level of coverage (ψ) as well as confidence (α).

Depending on the underlying model, the computation of tolerance intervals can be quite complex and might involve numeric optimization, bootstrapping or other computationally expensive methods (for a comprehensive overview, see [22]). This is not the case for OLS models, where tolerance intervals can be calculated using a ratio of critical values of the χ^2 distribution [23]:

$$\hat{y} \pm \sigma \sqrt{\frac{(n-p)\chi_{1;\psi}^2 \left(\frac{1}{n_i^*}\right)}{\chi_{n-p;\alpha}^2}} \quad (4)$$

The term $\chi_{1;\psi}^2$ in the numerator is the critical value of the χ^2 distribution at probability ψ , 1 degree of freedom and the noncentrality parameter set to $1/n_i^*$ where n_i^* is the vector of “effective number of observations” $n_i^* = \frac{\hat{\sigma}^2}{se(\hat{y}_i)^2}$. The distribution in the denominator is evaluated at probability α and uses $n - p$ degrees of freedom. While not strictly necessary in the case of OLS, we use this definition for illustrating the approximation of intervals in section 3.3.

2.3. Optimization Algorithms

The approach to DS computation presented here is based on finding the largest possible set of linearly independent PP ranges that satisfy acceptance criteria. In other words, it maximizes the volume of a hypercube within the parameter space. Here, this is formulated as a continuous optimization problem. Such problems find specific values of the parameter vector x that minimizes a function $f(x)$ subject to a set of inequality constraints $g_i(x) \leq 0, i = 1, \dots, m$ where m is the number of constraints. The DS volume problem can be formulated as an optimization objective function without equality constraints, only using inequality constraints (see section 3 for details). Those relatively minor requirements enable us to choose from a multitude of well researched and widely available optimization algorithms (for benchmarks and an overview see [24]).

COBYLA (Constrained Optimization BY Linear Approximation) was chosen as the main optimization algorithm to maximize the DS volume, as it meets the requirement for inequality constraints and, as a gradient-free method, shows reasonable robustness against converging in local minima [25]. The algorithm repeatedly evaluates the objective function at the corners of a “simplex”, i.e., at $p + 1$ points, p being the number of variables in x . One of the most widely used variants of simplex-based optimization algorithms is the one proposed by Nelder and Mead [26]. The Nelder-Mead algorithm, however, can run into situations where an incorrect minimum of the objective function is found, as shown by MacKinnon [27]. COBYLA expands upon Nelder-Mead by interpolating the vertices of the simplex using linear polynomials and introducing a trust region, which in turn are used to find the next vertex candidate. This new vertex is different from all current vertices in the simplex, thus circumventing the problem of incorrect convergence. The trust region radius $\Delta > 0$ represents the boundary for finding a new vertex, i.e., an improved point \hat{x} in the vicinity of the current point x_0 is found by minimizing the objective function $f(\hat{x})$ subject to $\|\hat{x} - x_0\| \leq \Delta$. The trust region radius is adjusted automatically and associated with a lower boundary ρ . This boundary starts with a predefined value ρ_{start} and gets smaller in later iterations to avoid local minima. The optimization parameter ρ_{start} can be set by the user and is relevant for the method proposed in the following sections, as it controls the granularity of the optimization process.

An intermediate step in our proposed approach concerns the search for minima and maxima of the objective function $f(x)$ within the boundaries of a hyperrectangle, i.e., a nested optimization problem with boundaries and no constraints. The L-BFGS-B algorithm was chosen due to its support for simple bounds and performance properties, especially in higher dimensions [24, 28]. This quasi-Newton method uses a limited-memory version of the approximation of the Hessian proposed in the original BFGS algorithm to guide optimization. Of course, any optimization algorithm that supports box constraints can be used for this step, such as Nelder-Mead [26], TNC [29] or SLSQP [30].

We propose an optional, second optimization pass for finetuning results, which can be executed using a COBYLA optimizer with a smaller value of ρ_{start} or SLSQP (sequential least squares programming), a quasi-Newton optimization algorithm [30]. This step uses the same constraints as in

the main optimization step, except that the true TI calculation is used as opposed to an approximation. The rationale behind this multi-step approach is that the first pass is expected to converge within the vicinity of the global optimum, while the second pass refines results and eliminates potential errors introduced by the TI approximation.

3. Finding a Design Space Comprised of Linear Combinations of Parameter Ranges

3.1. Overview

The method proposed in this contribution aims to solve the accuracy and computational time problem outlined in the introduction. This is achieved by employing numerical optimizers instead of grid screening methods commonly found in state-of-the-art software. Section 3.2 describes the objective function and inequality constraints used by those optimizers. As the minimization of computational time is of paramount importance, we furthermore introduce a quadratic approximation of tolerance intervals in section 3.3 as well as a fast method for incorporating categorical effects in section 3.4. These steps, as well as the use of an optimization algorithm, require some pre- and post-processing procedures, shown in Figure 2.

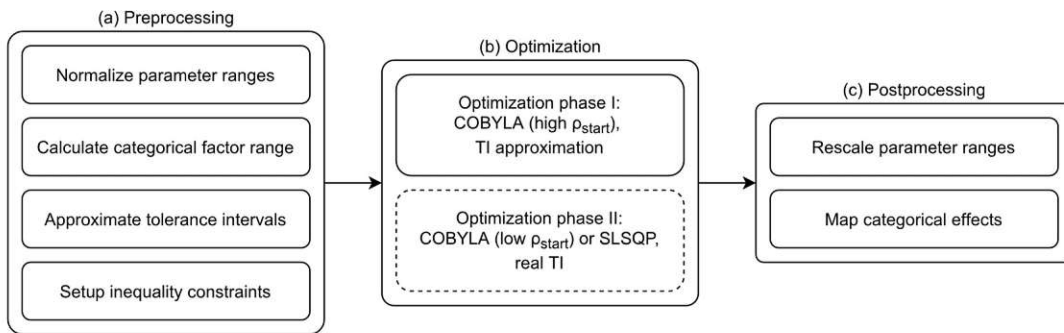


Figure 2: Overview of the steps involved in the DS calculation using an optimizer.

In the preprocessing step, each dimension of the parameter space is normalized to the range $[-1, 1]$ to improve speed and accuracy of the optimization algorithms [31]. Categorical factor ranges are determined, and the tolerance interval approximation is set up. This information and appropriate inequality constraints are then used in the optimization step. The primary optimization phase tries to find the region of the global optimum using COBYLA and large, initial step sizes expressed by the parameter ρ_{start} . Optionally, the results from the first phase can be improved in a second phase using either COBYLA with a reduced step size ρ_{start} or, depending on the underlying model and dimensionality, a gradient-based SLSQP optimizer. After a satisfactory DS is found, the results are transformed into their original scale and the continuous boundaries for categorical effects are mapped to valid levels (for details, see next sections).

3.2. Optimization Problem and Constraints

We define the search for a rectangular design space with the maximum volume as a continuous optimization problem with an objective function that represents the hyperrectangles volume:

$$\text{maximize} \quad f(x) = \prod_{i=1}^p (x_{p+i} - x_i) w_i, \quad x \in \mathbb{R}^{2p} \quad (5)$$

$$\text{subject to} \quad c_1(x), \dots, c_{2p+1+5p+2}(x) \geq 0 \quad (6)$$

Where x is the vector of parameter values that is varied in the optimization process, comprised of the lower parameter boundaries in the first p elements and the upper boundaries in the next p elements, with p being the number of factors in the model. One can see that (5) simply maximizes the volume of the hypercube spanned by lower and upper parameter ranges. The vector w contains predefined weights per parameter which can be used to favour one parameter range over. By default, this a vector of ones. The effects of weighting are illustrated in the example in section 4.2.1.

To meet all requirements for a valid design space, a total of $2^{p+1} + 5p + 2$ inequality constraints is defined. First, the parameter space to be searched is constrained by ensuring that a parameter's lower range boundary is smaller than its upper boundary.

$$c_i(x) = x_{p+i} - x_i, \quad i = 1, 2, \dots, p \quad (7)$$

Oftentimes the design space is required to contain each parameter's setpoint, so that $x_i \leq s_i \leq x_{p+1}$. This is expressed as the inequality constraint:

$$c_i(x) = s_i - x_i, \quad i = 1, 2, \dots, p \quad (8)$$

$$c_i(x) = x_i - s_i, \quad i = p + 1, \dots, 2p \quad (9)$$

Similarly, optimization of the parameter space should only be performed within the screening range boundaries b_l and b_u . As our main optimization algorithm does not support natural boundaries, this is implemented as inequality constraints:

$$c_i(x) = x_i - b_{l,i}, \quad i = 1, 2, \dots, p \quad (10)$$

$$c_i(x) = b_{u,i} - x_i, \quad i = p + 1, \dots, 2p \quad (11)$$

The remaining constraints address the evaluation of the TI. To that end, the approximation generated in the pre-optimization step is used to calculate boundaries around CQA predictions that capture model uncertainty. These boundaries are then compared against the lower and upper acceptance limits a_l and a_u in each of the 2^p corner points of the hyperrectangle:

$$c_i(x) = \hat{t}_l(x\hat{\beta}) - a_l, \quad i = 1, 2, \dots, 2^p \quad (12)$$

$$c_i(x) = a_u - \hat{t}_u(x\hat{\beta}), \quad i = 1, 2, \dots, 2^p \quad (13)$$

Here, x denotes a point that is taken from all corners of the current DS candidate and $x\hat{\beta}$ yields the predicted mean that is passed to the TI approximation. The tolerance interval approximation functions $\hat{t}_l(x)$ and $\hat{t}_u(x)$ are described in section 3.3.

Evaluating the TI at corners alone does not guarantee a valid design space, as curvature in the response surface might lead to parameter ranges between corner points that exceed acceptance limits. To resolve this problem, a final inequality constraint uses a nested optimization step to find minima and maxima of the TI boundaries inside the hypercube:

$$c_i(x) = \min_{x \in DS} \hat{t}_l(x\hat{\beta}) - a_l \quad (14)$$

$$c_i(x) = a_u - \max_{x \in DS} \hat{t}_u(x\hat{\beta}) \quad (15)$$

As described in section 2.3, the L-BFGS-B optimization algorithm is used to solve this nested problem [28]. While these constraints might seem to make equations (12) and (13) redundant, having both TI checks in place can improve convergence of the optimizer in certain scenarios. Furthermore, one or the other can be deactivated in practice, depending on the type of optimization problem.

3.3. Tolerance Interval Approximation

As mentioned in section 2.2, the calculation of tolerance intervals can be computationally expensive, depending on the type of model used for prediction. Furthermore, the inequality constraints described in section 3.2 evaluate intervals at several points in each iteration of the optimization algorithm, turning them into a potential bottleneck. Therefore, we present an approximation method that turns the calculation of a tolerance interval into a simple vector multiplication that can be carried out in a computationally efficient manner using standard linear algebra libraries. Although not necessarily required in the case of OLS, we use the tolerance interval definition in (4) as an illustrative example. To derive a parsimonious approximation, we exploit a property common to regression models, that is, that the least-squares projection $\hat{y} = X\hat{\beta}$ always goes through the multivariate mean of X and therefore parameter uncertainty associated with predictions around this point is smaller than at the boundaries of the parameter space. As tolerance intervals incorporate parameter uncertainty, this means that the interval is smaller in the center. This is true for the factors in X as well as the mean prediction \hat{y} , as illustrated in Figure 3. Here, observations in X as well as a vector of coefficients β were randomly generated to create the response y with some added noise. After fitting an OLS model to the data, tolerance intervals were calculated for the predicted values \hat{y} and their widths plotted on the y-axis. Note that this kind of curvature can be observed independently of parameter ranges or effect sizes.

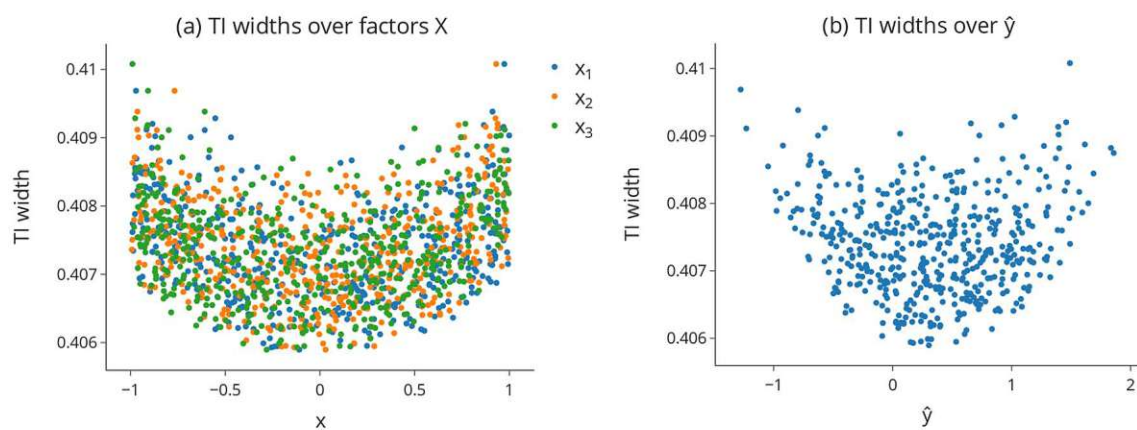


Figure 3: Tolerance interval (TI) widths over ordered (a) factor values and (b) mean predictions.

While the shape can be affected by strong quadratic and interaction effects, the relationship between \hat{y} and the tolerance interval width can be roughly approximated as a quadratic polynomial – a fact that is utilized in the proposed approximation. In a preprocessing step performed before DS optimization, the original CQA model is used to predict means and TI ranges for a design matrix X specifically designed to detect quadratic trends in data – a central composite design (CCD) [32]. The design is composed of $2^p + 2p + 5$ rows in X , giving it reasonable scalability over the number of factors typically used in biopharmaceutical models. The data is then used to regress the predicted means \hat{y} onto the TI widths assuming a second order polynomial:

$$\hat{t}i(\hat{y}) = \hat{\beta}_0 + \hat{\beta}_1\hat{y} + \hat{\beta}_2\hat{y}^2 \quad (16)$$

This regression model of TI widths is subsequently used in the optimization process. Furthermore, lower and upper boundaries are defined around the predicted mean:

$$\hat{t}_l(\hat{y}) = \hat{y} - \hat{t}(\hat{y}) \quad (17)$$

$$\hat{t}_u(\hat{y}) = \hat{y} + \hat{t}(\hat{y}) \quad (18)$$

To evaluate the accuracy of the approximation, it was compared to the actual tolerance interval over a range of randomly generated values for p , σ^2 , α and ψ as well as different interaction and quadratic effects. A measure of relative error was calculated by dividing the difference between real and approximated TI boundaries by the range of the model response. For OLS models, the mean error over 1000 iterations was 1.49%, normalized over the model response range. To highlight that the approximation method is largely model agnostic, the simulation was repeated with linear mixed models (LMM) and data containing a single random effect. In each iteration the random effect variance and BLUP values were varied while the data was evenly split into four random blocks. For calculating tolerance intervals, we used the method proposed by Franzq et al. [7]. Here, the relative approximation error is even smaller at 0.70%, as the range of the model response used for normalization is inflated due to additional random effect variance. Figure 4 contains histograms of the error distribution for both the OLS and the LMM simulation.

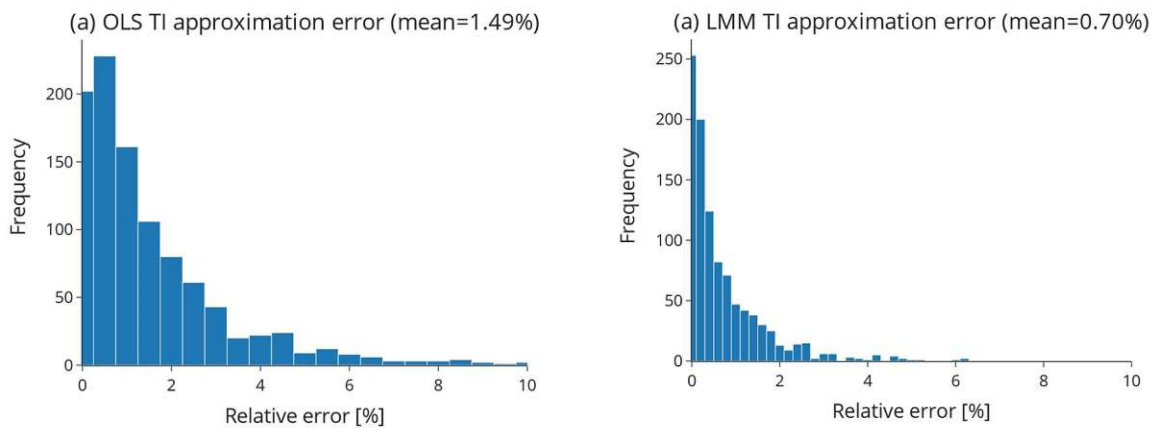


Figure 4: Relative error of the TI approximation for (a) OLS and (b) LMM, calculated as $err = (TI_{approx} - TI_{true}) / (y_{max} - y_{min})$. The relative error for LMMs is even smaller due to inflated variance introduced by the random effect.

While this level of accuracy cannot be expected for more complex model types, we believe that it yields a reasonable estimate of the TI to guide the optimization process. Furthermore, inaccuracies due to the approximation are mitigated by the second pass of the optimizer that uses the actual TI calculation method, as described in the following sections.

3.4. Categorical Factors

Many different approaches to optimization problems involving categorical factors can be found in literature, e.g., mixed-integer programming [33], branch-and-bound tree searching [34], genetic algorithms [35], Bayesian methods [36], etc.

Here we describe how to include categorical variables as continuous parameters of an optimization problem by exploiting how they are represented in a regression model. Specifically, we assume that such factors enter the model as sum-coded columns of the design matrix X [37]. A categorical factor with k levels is encoded into $k - 1$ columns and their values indicate that the level corresponding to the observation is active ("1"), inactive ("0") or that the last level, which is not encoded as a separate column, should be applied ("-1"). An example for this schema is given in Table 1 and Table 2.

Table 1: Original levels of the categorical factor.

Category
A
A
B
B
C
C

Table 2: Columns representing individual levels of the original column as a result of sum-encoding.

Category A	Category B
1	0
1	0
0	1
0	1
-1	-1
-1	-1

When representing the training data for the model in this way, the least-squares solution yields coefficients for the first $k - 1$ levels of the categorical factor, hereafter denoted as the vector $\hat{\beta}_c \in \mathbb{R}^{k-1}$. The coefficient for the last level, not represented as a column in the training data, can be calculated by $-\sum_{j=1}^{k-1} \hat{\beta}_{c,j}$. Consequently, the coefficients sum to zero and their values correspond to offsets from the mean of level means, or the intercept in the case of OLS models. In other words, the effect of a particular level in a categorical factor is simply added to the model response $\hat{y} = X\hat{\beta}$. Figure 5 shows the vertical shift of the regression line caused by three categorical levels "a", "b" and "c", whereas "c" is encoded as the negative sum of all other level effects.

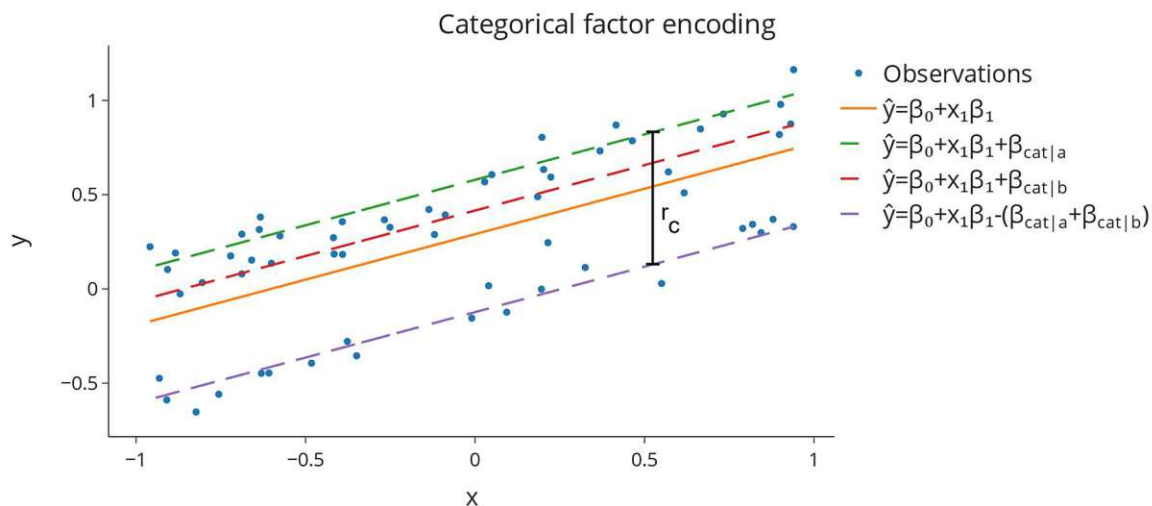


Figure 5: The effect of different levels of a sum-coded, categorical factor. The original mean prediction without the categorical factor is shown as the orange line, while the effects of the different levels are shown as dashed lines. We use the range r_c indicated by the black line for optimization.

This description of a categorical factor allows us to incorporate it into the optimization problem as a continuous factor bounded by $[\min(\hat{\beta}_c), \max(\hat{\beta}_c)]$. The range that should be considered by the optimizer is denoted as $r_c = \max(\hat{\beta}_c) - \min(\hat{\beta}_c)$, shown in Figure 5. As the optimizer operates in a normalized space, i.e., parameters are deviated within $[-1, 1]$, a mean-normalization rescaling scheme is incorporated into the evaluation of \hat{y} within the optimization procedure. To that end the mean prediction is expanded to $\hat{y} = X\hat{\beta} + x_c r_c + \bar{r}_c$ or, equivalently, by simply adding the terms

model data and parameters $\hat{y} = [x_1, \dots, x_p, x_c, 1_n]^T [\hat{\beta}_1, \dots, \hat{\beta}_p, \bar{r}_c]$. After optimization, a lower and upper range is returned that encloses all valid levels. These levels are added to the design space as possible values for the categorical factor.

While this approach is specific to polynomial regression models and sum-coded categorical factors, it is performant and poses fewer restrictions on the choice of optimization algorithm when compared to other methods.

4. Simulation Study

4.1. Setup and Investigated Cases

Accuracy and efficiency of the proposed method were evaluated in a set of simulation studies. The former was investigated in two studies with the purpose of showing (i) the general validity of the approach and to highlight differences in the resulting DS compared to a grid-based method and (ii) the applicability to a realistic scenario in the biopharmaceutical domain. Efficiency was evaluated by randomly creating regression models with different model sizes and measuring the computational time to compute a DS based on them. The results of both accuracy and efficiency are compared to a grid-based approach that performs a full scan of the parameter space by dividing it into a grid of equidistant points and evaluating all possible point combinations to find the largest hyperrectangle that contains only points satisfying the acceptance criteria. Note that due to the difference between the definition of a probabilistic design space used by other methods referenced in the introduction and the definition used here, i.e., a regression tolerance interval that includes model uncertainty and population coverage, results from these methods are not directly comparable and therefore not included in this evaluation.

The investigated cases are described in the following subsections. All studies were conducted on a workstation notebook with an Intel Core i7-8565U CPU, integrated graphics and 16 GB of DDR4 memory, in a Python 3.9 environment using the packages numpy 1.23.5 for algebra, scipy 1.8.0 for optimization and statsmodels 0.13.2 for modelling.

4.1.1. Case A: 2-Dimensional Parameter Space

To demonstrate that the method is capable of finding a ground truth in principle and to illustrate the difference to sampling from a predefined grid of 16x16 points graphically, the first simulation study identifies the DS in a 2-dimensional parameter space and inscribes the largest possible rectangle into a unit circle with the radius $r = 1$. The sides of this rectangle are of length $r\sqrt{2}$, thus its area is $(r\sqrt{2})^2$ or, for the unit circle, simply 2. Centered at the point $(0, 0)$, the acceptable parameter range for x_1 and x_2 is $\left[-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right]$. The form of the regression model used for the optimization-based as well as the grid-based method is a consequence of the unit circle problem:

$$y = x_1^2 + x_2^2 \quad (19)$$

Model uncertainty was not considered for the sake of results being comparable to the ground truth, i.e., model predictions were compared directly to acceptance limits and neither tolerance interval nor its approximation were applied. A single pass of the COBYLA optimization procedure was performed with a trust region boundary of $\rho_{start} = 0.01$.

4.1.2. Case B: 6-Dimensional Parameter Space

A second simulation study serves as an example that is more representative of models used in the biopharmaceutical domain. The regression model used here contains 6 main factors as well as interactions and quadratic effects:

$$\begin{aligned}
y = & -0.45 + 2.1x_1 - 0.93x_2 + 0.63x_3 + 0.42x_4 - 0.32x_5 + 0.28x_6 + 0.76x_2^2 - 0.21x_1x_2 \\
& - 0.34x_1x_5 + 0.22x_1x_6 + 0.27x_2x_3 + 0.24x_2x_4 - 0.25x_2x_5 - 0.19x_3x_6 \\
& - 0.32x_4x_6
\end{aligned} \quad (20)$$

Effect sizes are fairly balanced and common in models used to represent downstream unit operations [4]. Model uncertainty as well as population coverage is considered by a tolerance interval with the parameters $\alpha = 0.05$ and $\psi = 0.05$, i.e., the interval covers 95% of the population values with a confidence level of 95%. As it is not possible to define a ground truth for this example analytically given the model and the interval included in checking acceptance criteria, results are only compared to those of the grid-based method. The parameter space for the grid was discretized over 9 equidistant points per parameter, resulting in 9^6 evaluations overall. After the initial run of the optimization algorithm with a trust region boundary $\rho_{start} = 0.01$, a second pass is performed with $\rho_{start} = 0.001$ for finetuning and its effect is illustrated in the results. Finally, to showcase the utility of the weighting scheme, the results include a run in which the parameter x_6 is weighted up to identify a more balanced DS.

4.1.3. Case C: Performance Evaluation

Computational time of the optimization-based method was evaluated for up to 10 parameters. For each model size p , 100 regression models were generated by sampling parameter values and effects uniformly from the interval $[-1, 1]$, randomly adding interactions and adding residual error $\sigma \sim N(0, s)$, where s was uniformly drawn from $[0.1, 1]$. In each interaction, tolerance intervals around the fitted values of the training data were calculated and the 80% quantile of the upper boundaries was used as a reasonable upper acceptance limit for calculating the design space. The TI approximation procedure described in section 3.3 was used and the study was performed with and without a second optimization pass. Mean and standard deviation of the computational time for computing the DS in the 100 iterations per model size were recorded. Due to performance limitations, this approach was not possible for the grid-based method and only a single model was used to compute the DS on a grid of 8^p points per model size. For the same reason, the DS was computed only for up to 6 parameters and a projection is used for the rest. As the parameter space grows exponentially, this projection fits an exponential function to the observed data and extrapolates for a higher number of parameters.

4.2. Results

4.2.1. Accuracy

Figure 6a illustrates the results of case study A in a contour plot. The optimization-based method successfully identifies the true rectangle, with an area of 2, inside the unit circle, exhibiting negligible error. As the grid-based method can only evaluate points on the 16x16 grid, the resulting DS volume is smaller than that found by the optimizer, which is not subject to that limitation. Of course, the accuracy of the grid method can be improved by increasing the resolution. This, however, can drastically increase the computational time for higher dimensional spaces and it is not clear which grid resolution yields a DS with reasonable accuracy. The numeric results are summarized in Table 3.

Table 3: Results for simulation study A. The optimizer results match the ground truth with negligible error.

	x1	x1	x2	x2	volume
	from	to	from	to	
ground truth	-0.70710678	0.70710678	-0.70710678	0.70710678	2.0
optimizer	-0.70710642	0.70710642	-0.70710642	0.70710642	2.0+9.53 ⁻¹³
grid	-0.6	0.6	-0.73333334	0.73333334	1.76

Figure 6b shows the results for a case study B, comprised of six main factors and several interaction and quadratic effects. In this representation of the DS, valid parameter ranges are indicated by the axes of the spider plot. As in the previous example, the volume of the DS based on the 9^6 grid is much smaller than that of the optimizer (orange solid line in Figure 6b). It is, however, more balanced, giving each parameter a similar range. This is an effect that might be observed in optimization-based results for higher-dimensional parameter spaces, as the optimization objective only considers DS volume. Here, the range of x_6 is quite small compared to the grid method. In case such a result is not desired, the weight for x_6 in objective (5) can be increased to yield a more balanced parameter range while sacrificing some volume, indicated by the dashed orange line in Figure 6b. Both DS plotted in orange were calculated using the first phase of the proposed scheme only, i.e., with an approximated TI and $\rho_{start} = 0.01$. The unweighted DS was then passed to the second phase for further refinement using the true TIs and $\rho_{start} = 0.001$, which results in the largest volume. The parameter ranges of the DS are shown in green and numeric results are summarized in Table 4.

Table 4: Results from simulation study B, highlighting increased DS volume of the optimization-based method and the flexibility added by the weighting scheme.

	x1		x2		x3		x4		x5		x6		volume
	from	to	from	to	from	to	from	to	from	to	from	to	
grid	-1.00	0.00	0.00	1.00	-1.00	0.25	-0.75	0.50	-0.25	1.00	-0.25	1.00	2.44
optimizer	-1.00	-0.51	-0.63	0.97	-1.00	0.64	-1.00	0.73	-0.91	1.00	-1.00	1.00	8.55
opt., weighted	-1.00	-0.51	-0.69	0.98	-1.00	0.66	-1.00	0.76	-0.95	1.00	-1.00	1.00	8.42
opt., second pass	-1.00	-0.25	-0.43	1.00	-1.00	0.31	-1.00	0.50	-1.00	1.00	-1.00	1.00	9.27

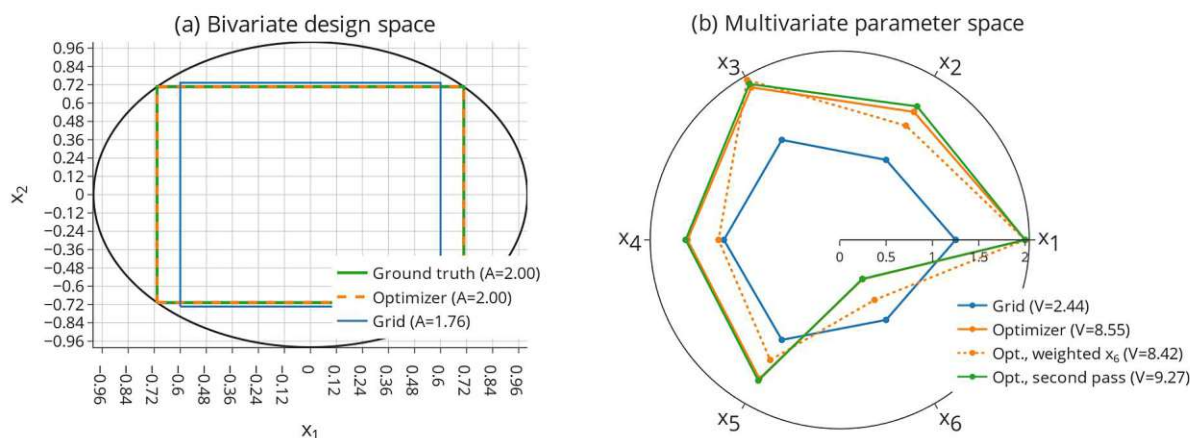


Figure 6: (a) Bivariate parameter space with an upper tolerance interval boundary shown as the contour. The design spaces found by the optimizer and the grid method are marked as rectangles. (b) A multivariate parameter space illustrated as a spider plot with ranges per parameter shown in the axes and DS plotted as polygons.

4.2.2. Computational Time

A major problem of grid-based approaches is poor scaling to higher dimensions of the parameter space. As the space to be searched grows exponentially with the number of parameters, so does computational time required to find the DS [6]. This renders the method infeasible for many use cases in the biopharmaceutical domain where ten main factors or more are not uncommon for unit

operation models (see case study in [4]). For example, the screening algorithm would need to evaluate 8^{10} points for parameter ranges segmented into eight parts and ten parameters while also computing all possible hyperrectangles within this space. Although optimization-based methods are not guaranteed to find the global maximum, the parameter space is screened in a more systematic way, ideally eliminating large parts of it in a few iterations. This is shown by the results of simulation study C illustrated in Figure 7, where the time for computing a DS is plotted against the number of model parameters of randomly generated models. As described in section 4.1.3, grid-based results are projected from 7 factors onwards, as it was not feasible to compute results. This is indicated by the dashed line in the plot.

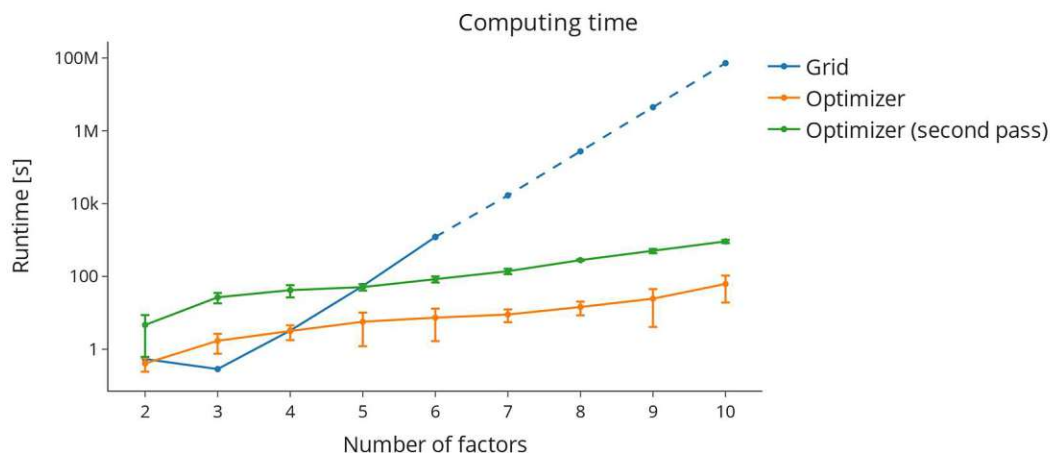


Figure 7: Performance evaluation results of the proposed method compared to the grid-based approach. The orange and green line indicate optimization-based results, where the latter results in a fixed offset in computational time. The blue line shows the grid-based results, which could only be calculated for up to 6 factors and exponential extrapolation was used to project results for a larger number of factors, indicated by the dashed line. Note that for a low number of factors, the grid-based method can be subject to random differences in initialization, which is why the computational time for 2 factors is shown to be lower than that of 3.

The number of iterations for the second optimization pass was capped at 100 and the computational time added to the calculation scales linearly with the number of parameters, shown as the green line in the plot. Overall, the optimization-based computation for 10 factors is six to seven orders of magnitude faster than the grid method, effectively enabling the exact computation of a DS in higher dimensions.

5. Discussion

We believe that the optimization-based approach to calculating a DS circumvents the problems associated with grid-based methods, i.e., poor scaling over the number of parameters and a lack of accuracy due to the limitation to grid points in the solution space. In contrast to existing Bayesian methods [10, 3, 11, 12] the resulting DS is deterministic, contains model uncertainty and consists of a linear combination of parameter ranges, facilitating operational simplicity for biopharmaceutical control strategies. However, as is the case for many optimization problems - especially in high-dimensional parameter spaces - there is a possibility of converging to local minima and not finding the largest possible DS. This problem can be mitigated by using different random seeds [38], different starting points for the algorithm [39] or different factor weights (section 4.1.2). Results from different starting conditions could be compared to see if they agree and thus gain confidence in the result or only the maximum DS from the set of candidates could be returned.

The weighting scheme in combination with fast computational times also facilitates the iterative exploration of the DS in higher dimensions, as shown in the results in Figure 6. One can investigate

the effect of weighing a parameter on the possible ranges of other parameters and thereby gain understanding about DS-related dependencies that cannot be directly derived from the model.

As mentioned in section 2.1 and 3.3, the optimization-based approach is largely model-agnostic, given that a tolerance interval can be defined and approximated, though the latter is only required for performance reasons. We have shown in section 3.3 that the quadratic approximation works well for LMMs. Future work could incorporate a wider range of models, such as GLS, GLM, censored data or mechanistic models.

6. Conclusion

In this contribution, we outlined a novel method for finding an ICH Q8 compliant design space comprised of linear combinations of PP ranges. The relationship between CQA and PPs is represented as a polynomial regression model and its prediction is used to evaluate whether CQAs meet acceptance criteria.

Conservative estimation methods are vital in the biopharmaceutical domain, which is why the boundaries of tolerance intervals are used for evaluation of ACs instead of the predicted mean CQA. As the TI calculation can be complex, an approximation is generated in a pre-optimization step that can be used to calculate the TI by performing a single matrix multiplication. COBYLA is used for the minimization of the main objective, and we suggest refining results with SLSQP or COBYLA with smaller ρ_{start} and using the true TI instead of an approximation.

Performance evaluations show that the proposed method results in design spaces with a larger volume when compared to PP space discretization methods, and that they can be calculated in a fraction of the time. This, for the first time, enables the calculation of design spaces for more than 10 process parameters of one model. We believe that this approach will facilitate a robust definition of the design space for biopharmaceutical development that reduces patient risk by employing conservative estimators while allowing manufacturers to maximize control ranges and operational flexibility. Furthermore, this increased flexibility, gained solely by an improved evaluation of existing data and models is a chance for manufacturers to optimize a process within the DS.

7. Acknowledgements

This work was conducted within the COMET Centre CHASE, funded within the COMET – Competence Centers for Excellent Technologies program by the BMK, the BMDW and the Federal Provinces of Upper Austria and Vienna. The COMET program is managed by the Austrian Research Promotion Agency (FFG).

The authors acknowledge TU Wien Bibliothek for financial support through its Open Access Funding Program.

8. Bibliography

- [1] ICH, *ICH guideline Q8 (R2) on pharmaceutical development*, EMA, 2017.
- [2] I. Y. Kim and B. M. Kwak, "Design space optimization using a numerical design continuation method," *International Journal for Numerical Methods in Engineering*, pp. 1979-2002, 2002.

- [3] K. Kusumo, L. Gomoescu, R. Paulen, S. García Muñoz, C. Pantelides, N. Shah and B. Chachuat, "Bayesian approach to probabilistic design space characterization: A nested sampling strategy," *Industrial & Engineering Chemistry Research*, vol. 59, no. 6, pp. 2396-2408, 2019.
- [4] T. Oberleitner, T. Zahel, M. Kunzelmann, J. Thoma and C. Herwig, "Incorporating random effects in biopharmaceutical control strategies," *AAPS Open*, vol. 9, pp. 1-13, 2023.
- [5] Sartorius Stedim Data Analytics AB, *MODDE® 12 User Guide*, 2017.
- [6] R. Taylor, "Dynamic programming and the curses of dimensionality," in *Applications of dynamic programming to agricultural decision problems*, CRC Press, 2019, pp. 1-10.
- [7] B. G. Francq, D. Lin and W. Hoyer, "Confidence, prediction, and tolerance in linear mixed models," *Statistics in medicine*, vol. 38, no. 30, pp. 5603-5622, 2019.
- [8] S. De Gryze, I. Langhans and M. Vandebroek, "Using the correct intervals for prediction: A tutorial on tolerance intervals for ordinary least-squares regression," *Chemometrics and intelligent laboratory systems*, vol. 87, no. 2, pp. 147-154, 2007.
- [9] D. Laky, S. Xu, J. Rodriguez, S. Vaidyaraman, S. García Muñoz and C. Laird, "An optimization-based framework to define the probabilistic design space of pharmaceutical processes with model uncertainty," *MDPI Processes*, vol. 7, no. 2, p. 96, 2019.
- [10] J. Tabora, F. Lora Gonzalez and J. Tom, "Bayesian probabilistic modeling in pharmaceutical process development," *AIChE journal*, vol. 65, no. 11, p. e16744, 2019.
- [11] J. Peterson, "A Bayesian approach to the ICH Q8 definition of design space," *Journal of biopharmaceutical statistics*, vol. 18, no. 5, pp. 959-975, 2008.
- [12] G. Bano, P. Facco, F. Bezzo and M. Barolo, "Probabilistic Design space determination in pharmaceutical product development: A Bayesian/latent variable approach," *AIChE Journal*, vol. 64, no. 7, pp. 2438-2449, 2018.
- [13] F. Boukouvala, F. Muzzio and M. Ierapetritou, "Methods and tools for design space identification in pharmaceutical development," in *Comprehensive Quality by Design for Pharmaceutical Product Development and Manufacture*, Wiley Online Library, 2017, pp. 95-123.
- [14] E. Polak, "An implementable algorithm for the optimal design centering, tolerancing, and tuning problem," *Journal of Optimization Theory and Applications*, vol. 37, no. 1, pp. 45-67, 1982.
- [15] S. Kucherenko, D. Giamalakis, N. Shah and S. García-Muñoz, "Computationally efficient identification of probabilistic design spaces through application of metamodeling and adaptive sampling," *Computers & Chemical Engineering*, vol. 132, p. 106608, 2020.
- [16] I. Grossmann and M. Morari, "Operability, resiliency, and flexibility: Process design objectives for a changing world," *Design Research Center, Carnegie-Mellon University*, 1983.
- [17] C. Knauer, L. Schlipf, J. Schmidt and H. Tiwary, "Largest inscribed rectangles in convex polygons," *Journal of discrete algorithms*, vol. 13, pp. 78-85, 2012.
- [18] D. C. Montgomery, E. A. Peck and G. G. Vining, *Introduction to linear regression analysis*, John Wiley & Sons, 2021.

- [19] H. Brown and R. Prescott, Applied mixed models in medicine, John Wiley & Sons, Inc., 2015.
- [20] W. Stroup, Generalized linear mixed models: modern concepts, methods and applications, CRC press, 2012.
- [21] C. Knudson, *Monte Carlo likelihood approximation for generalized linear mixed models*, University of Minnesota, 2016.
- [22] K. Krishnamoorthy and T. Mathew, Statistical tolerance regions: theory, applications, and computation, John Wiley & Sons, 2009.
- [23] W. Guenther, "Tolerance intervals for univariate distributions," *Naval Research Logistics*, vol. 19, pp. 309-333, 1972.
- [24] K. Varelas and M.-A. Dahito, "Benchmarking multivariate solvers of SciPy on the noiseless testbed," in *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, 2019, pp. 1946-1954.
- [25] M. Powell, "A direct search optimization method that models the objective and constraint functions by linear interpolation," in *Advances in optimization and numerical analysis*, Springer, 1994, pp. 51-67.
- [26] J. Nelder and R. Mead, "A simplex method for function minimization," *The computer journal*, vol. 7, no. 4, pp. 308-313, 1965.
- [27] K. McKinnon, "Convergence of the Nelder-Mead simplex method to a nonstationary point," *SIAM Journal on optimization*, vol. 9, no. 1, pp. 148-158, 1998.
- [28] R. Byrd, P. Lu, J. Nocedal and C. Zhu, "A limited memory algorithm for bound constrained optimization," *SIAM Journal on scientific computing*, pp. 1190-1208, 1995.
- [29] S. Nash, "Newton-type minimization via the Lanczos method," *SIAM Journal on Numerical Analysis*, vol. 21, no. 4, pp. 770-788, 1984.
- [30] D. Kraft, "A software package for sequential quadratic programming," *Forschungsbericht-Deutsche Forschungs- und Versuchsanstalt für Luft- und Raumfahrt*, 1988.
- [31] R. Horst and P. Pardalos, Handbook of global optimization, Springer Science & Business Media, 2013.
- [32] D. C. Montgomery, Design and Analysis of Experiments, John Wiley & Sons, 2017.
- [33] M. Muñoz Zuniga and D. Sinoquet, "Global optimization for mixed categorical-continuous variables based on Gaussian process models with a randomized categorical space exploration step," *INFOR: Information Systems and Operational Research*, pp. 310-341, 2020.
- [34] C. Vanaret, "A global method for mixed categorical optimization with catalogs," *arXiv preprint arXiv:2104.03652*, 2021.
- [35] K. Badran and P. Rockett, "Integrating categorical variables with multiobjective genetic programming for classifier construction," in *European Conference on Genetic Programming*, Berlin, Germany, 2008.

- [36] P. Saves, E. N. Van, N. Bartoli, T. Lefebvre, C. David, S. Defoort, Y. Diouane and J. Morlier, "Bayesian optimization for mixed variables using an adaptive dimension reduction process: applications to aircraft design," San Diego, United States, 2022.
- [37] UCLA Statistical Consulting Group, "R Library Contrast Coding Systems for categorical variables," 2021. [Online]. Available: <https://stats.oarc.ucla.edu/r/library/r-library-contrast-coding-systems-for-categorical-variables/>.
- [38] S. Bethard, "We need to talk about random seeds," *arXiv preprint arXiv:2210.13393*, 2022.
- [39] S. Brooks and B. Morgan, "Optimization using simulated annealing," *Journal of the Royal Statistical Society Series D: The Statistician*, vol. 44, no. 2, pp. 241-257, 1995.

8 Appendix B: Models and Intervals

8.1 Models

8.1.1 Ordinary Least Squares Models

One of the simplest and widely used regression models is the ordinary least squares (OLS) model. It is composed of a response vector y , a $n * p$ design matrix of observations X that contains different factors in its columns and a set of observations in its rows, a vector of coefficients β and a residual error vector ϵ :

$$y = X\beta + \epsilon \quad (8.1)$$

Fitting the model is synonymous with finding an estimate of the coefficient vector β . Due to the simplicity of OLS, a unique solution exists for this, and it can be obtained with a closed form expression, provided X is non-singular and therefore $X^T X$ positive definite:

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad (8.2)$$

Under the assumption that $\epsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$, $\hat{\beta}$ is the best linear unbiased estimator (BLUE) for β , where σ^2 is the measure of variance of residuals. The optimal properties of estimators and wide availability in statistical software make OLS models a popular tool in many applied sciences.

8.1.2 Linear Mixed Models

The mathematical formulation for linear mixed models (LMM) adds the Boolean design matrix Z that assigns each observation to a group. The general form of Z allows for the inclusion of multiple, random effects and more complex group structures like nested and crossed designs. An introduction to mixed models can be found in [1] and [2]. The mathematical notation for the model, variance components, estimators and likelihood function used here is taken from [3].

$$y = X\beta + Z\gamma + \epsilon \quad (8.3)$$

In addition to the design matrices X and Z , the model contains the vector of fixed effect coefficients β and random effect coefficients $\gamma \sim N(\mathbf{0}, \sigma_\gamma^2 \mathbf{I})$. The residual error is assumed to be normally distributed with $\epsilon \sim N(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I})$, as in the OLS model. Note that the mixed model splits up variance into a random component σ_γ^2 and a residual component σ_ϵ^2 . This enables a more detailed explanation of the variance observed in the data, i.e., how much of it is the result from fixed effects and how was introduced by random blocks. Using those different variance components, the overall variance in y can be expressed as $Var(y) = V = ZGZ^T + R$, with random effect variances in the diagonal elements of G and $R = \sigma_\epsilon^2 \mathbf{I}$.

Assuming both random blocks and residuals are normally distributed, the feasible general least-squares (FGLS) estimates of the coefficient vectors β and γ can be obtained by:

$$\hat{\beta} = (X^T \hat{V}^{-1} X)^{-1} X^T \hat{V}^{-1} y \quad (8.4)$$

$$\hat{\gamma} = \hat{G}Z^T\hat{V}^{-1}(y - X\hat{\beta}) \quad (8.5)$$

However, this requires the random and residual variances in V to be already known, which is usually not the case. A common approach to calculate unbiased estimates of variance components and the estimator for the variance matrix \hat{V} based on the data is restricted maximum likelihood (REML). In this method the matrices G and R are computed numerically by minimizing the likelihood function,

$$L_{REML}(G, R) = -\frac{1}{2}\log|V| - \frac{1}{2}\log|X^TV^{-1}X| - \frac{1}{2}r^TV^{-1}r - \frac{n-p}{2}\log(2\pi) \quad (8.6)$$

with r being the vector of residuals $r = y - X(X^TV^{-1}X)^{-1}X^TV^{-1}y$. Generally, statistical software uses some variant of the Newton-Raphson algorithm for optimization [3]. Of course, REML optimization followed by computing FGLS estimates for fixed and random effects is more expensive computationally than fitting an OLS model. Nevertheless, the difference is negligible in modern computing environments and the more accurate prediction of statistical intervals outweigh any concerns regarding performance in most use cases.

8.2 Statistical Intervals

Statistical intervals represent an important and widely used tool to calculate and visualize uncertainty and data, estimators, or predictions. The most widely used type of interval is the confidence interval. Other types like prediction or tolerance intervals are applied less frequently, though they might be more appropriate in use cases where uncertainty about the true distribution of drawn samples is expressed. The following sections describe intervals in the context of regression models.

8.2.1 Confidence Intervals

Confidence intervals in regression models are used to express uncertainty about the predicted mean of a model, or in other words, about the expected value of the response. Because this property of the population can only be estimated imperfectly with a limited set of observations, confidence intervals define a region that contains the true value with a predefined level of confidence $100(1-\alpha)\%$. Borrowing notation from [4], the probability that the predicted mean of the response is contained within the "true" interval I_c , based on the multivariate predictor $x_{(n+1)}$ is

$$Pr[\mu_{(y|x_{n+1})} \in I_c(x_{n+1})] = 1 - \alpha \quad (8.7)$$

This leads to the common form for two-sided, multivariate confidence intervals in both OLS models and LMM:

$$\hat{y} \pm t(n-p, \alpha/2) \sqrt{\sigma^2 x_{n+1}^T (X^T X)^{-1} x_{n+1}} \quad (8.8)$$

Where t is the probability density function of the T-distribution with $n-p$ degrees of freedom and the critical value $\alpha/2$. The square root term is called the standard error of the fit, resulting from the covariance matrix $(X^T X)^{-1}$ [1, 5].

8.2.2 Prediction Intervals

Oftentimes, when regression models are applied, we are not interested in the uncertainty of the predicted mean but in the distribution of actual values. The probability that a new observation $y|x_{n+1}$ is contained within the true interval I_p is

$$Pr[y|x_{n+1} \in I_p(x_{n+1})] = 1 - \alpha \quad (8.9)$$

The difference between confidence and prediction intervals is subtle and sometimes not well understood, which can lead to problematic outcomes when conservative model results are favorable, e.g., in biopharmaceutical manufacturing. Prediction intervals are wider than confidence intervals, as the residual model error is added to the standard error. An OLS prediction interval can be computed by

$$\hat{y} \pm t(n-p, \alpha/2) \sqrt{\sigma^2(1 + x_{n+1}^T (X^T X)^{-1} x_{n+1})} \quad (8.10)$$

A similar formula can be applied when using LMMs, where the covariance matrix for the OLS fixed effect estimator $\sigma^2(X^T X)^{-1}$ is replaced with its GLS counterpart $(X^T \hat{V}^{-1} X)^{-1}$ and the degrees of freedom for the t-distribution $n-p$ are replaced with the generalized Satterthwaite approximation proposed in [6].

8.2.3 Tolerance Intervals

Note that prediction intervals define a region into which a single, new prediction might fall with a predefined level of confidence. Using prediction intervals alone, one cannot express whether a certain proportion of responses will fall within a range. The appropriate tool to incorporate both, confidence in and coverage of the true population in a region around the prediction, is the tolerance interval. It illustrates the probability that the new prediction is contained within the true interval I_T in at least $(100P)\%$ of repeated samplings to a nominal confidence of $100(1-\alpha)\%$.

$$Pr[Pr[y|x_{n+1} \in I_T(x_{n+1})] \geq P] = 1 - \alpha \quad (8.11)$$

Several approximations for tolerance intervals in OLS models can be found in literature [7–9]. Here, we highlight the method employed by the R package "tolerance" [8, 10]:

$$\hat{y} \pm \sigma \sqrt{\frac{(n-p)\chi_{1;P}^2(\frac{1}{n_i^*})}{\chi_{n-p;\alpha}^2}} \quad (8.12)$$

where $n_i^* = \frac{\hat{\sigma}^2}{se(\hat{y}_i)^2}$ are the *effective number of observations* [11], P the coverage, α the confidence level and $\chi_{d;\alpha}^2(\delta)$ the probability density function with degrees of freedom d , confidence level α and an optional non-centrality parameter δ .

An approximation for tolerance intervals that includes random effects is proposed by Franzq et al. [6]. It is similar to the prediction interval, except that the standard deviation added to the fixed effect variance is replaced by the upper $100(1-\alpha)\%$ confidence bound, which in turn is computed by a sum of expected mean squares (EMS). EMS are used instead of the actual variances to allow for crossed and nested scenarios, where variance components are not independent and cannot be simply summed up to calculate the total variance.

$$\hat{y} \pm z(\psi/2) \sqrt{x_{n+1}^T (X^T V^{-1} X)^{-1} x_{n+1} + \sigma_T^2} \sqrt{1 + \left(\frac{1}{\hat{\sigma}_T^2}\right) \sqrt{\sum_{j=1}^q H_j^2 k_j^2 EMS_j^2}} \quad (8.13)$$

The total variance, σ_T^2 , is calculated as the sum of EMS, $H_j^1 = \frac{r_j}{\chi_{\alpha, r_j}^2} - 1$ with r_j being the number of replicates in a random block and $k_j^2 EMS_j^2$ linear combinations of expected mean squares. The tolerance intervals for LMMs used to incorporate random effects into control strategies were calculated using this formula.

References

- [1] D. C. Montgomery, E. A. Peck, and G. G. Vining. *Introduction to linear regression analysis*. John Wiley & Sons, 2021.
- [2] D. C. Montgomery. *Design and Analysis of Experiments*. 2017.
- [3] S. Institute. *SAS/STAT® 9.22 User's Guide*. SAS Institute Inc., 2010.
- [4] S. De Gryze, I. Langhans, and M. Vandebroek. “Using the correct intervals for prediction: A tutorial on tolerance intervals for ordinary least-squares regression”. In: *Chemometrics and intelligent laboratory systems* 87.2 (2007), pp. 147–154.
- [5] P. E. C. of Science. *PennState: Statistics Online Courses*. 2021. URL: <https://online.stat.psu.edu/stat501/lesson/3/3.2> (visited on 07/20/2023).
- [6] B. G. Francq, D. Lin, and W. Hoyer. “Confidence, prediction, and tolerance in linear mixed models”. In: *Statistics in medicine* 38.30 (2019), pp. 5603–5622.
- [7] W. Howe. “Two-sided tolerance limits for normal populations—some improvements”. In: *Journal of the American Statistical Association* 64.326 (1969), pp. 610–620.
- [8] K. Krishnamoorthy and T. Mathew. *Statistical tolerance regions: theory, applications, and computation*. John Wiley & Sons, 2009.
- [9] W. C. Guenther and W. C. Guenther. *Sampling Inspection in statistical quality control*. 37. C. Griffin, 1977.
- [10] D. S. Young. “Tolerance: an R package for estimating tolerance intervals”. In: *Journal of Statistical Software* 36 (2010), pp. 1–39.
- [11] W. A. Wallis. “Tolerance intervals for linear regression”. In: *Proceedings of the second Berkeley symposium on mathematical statistics and probability*. Vol. 2. University of California Press. 1951, pp. 43–52.

9 Appendix C

9.1 Curriculum vitae

Thomas Oberleitner

✉ tom.oberleitner@gmail.com

in <https://www.linkedin.com/in/thomas-oberleitner-a22832156/>

RG <https://www.researchgate.net/profile/Thomas-Oberleitner-2>



Employment History

- 2023 – ... **Research Scholar, Environmental Data Science**, International Institute for Applied Systems Analysis, Laxenburg, Austria.
- 2020 – 2023 **Researcher**, Competence Center CHASE GmbH, Vienna, Austria.
- 2019 – 2020 **Software and R&D Engineer**, LogicX consulting & workflow integration GmbH, Vienna, Austria.
- 2011 – 2019 **Principal Software Engineer, Backend**, Micado IT-Solutions, Austria.

Education

- 2020 – 2023 **Ph.D., CMC Statistics**
Vienna University of Technology, Dept. Chemical and Process Engineering
Thesis: *Advancements in Integrated Modeling of Biopharmaceutical Processes: Incorporating Random Effects, Holistic Design of Experiments and Design Space Exploration.*
- 2017 – 2019 **M.Sc. Software Engineering**
University of Applied Sciences Technikum Wien
Thesis: *Combining Deep Reinforcement Learning Optimizations to Perform Tasks in Doom.*
- 2014 – 2017 **B.Sc. Information and Communication Systems and Services**
University of Applied Sciences Technikum Wien
Thesis I: *Security in Open Source Content Management Systems*
Thesis II: *An Evaluation of Tools and Techniques for Source Code Analysis.*

Research Publications

Journal Articles

- 1 T. Oberleitner, T. Zahel, and C. Herwig, "A method for finding a design space as linear combinations of parameter ranges for biopharmaceutical control strategies (under review)," *arXiv preprint arXiv:2304.14666*, 2023.
- 2 T. Oberleitner, T. Zahel, M. Kunzelmann, J. Thoma, and C. Herwig, "Incorporating random effects in biopharmaceutical control strategies," *AAPS Open*, vol. 9, no. 1, pp. 1–13, 2023.
- 3 T. Oberleitner, T. Zahel, B. Pretzner, and C. Herwig, "Holistic design of experiments using an integrated process model," *Bioengineering*, vol. 9, no. 11, 2022.

Books and Chapters

- 1 T. Oberleitner, T. Zahel, and C. Herwig, "A method for finding a design space as linear combinations of parameter ranges for biopharmaceutical development," in *Computer Aided Chemical Engineering*, vol. 52, Elsevier, 2023, pp. 909–914.