s c c h .

software competence center
hagenberg

FUZZY.LOGIC.LAB.LINZ

Advances in Knowledge-Based Technologies

Proceedings of the
Master and PhD Seminar
Winter term 2017/2018, part 1

Softwarepark Hagenberg
SCCH, Room 0/2
22 November 2017

Program

Session 1. Chair: Bernhard Moser

9:00	F. Sobieczky:
	Statistical Depth
9:30	R. Nikzad-Langerodi:
	Domain – Invariant Partial Least Squares Regression: A Novel Calibration Transfer Paradigm
10:00	W. Zellinger:
	Unsupervised Domain Adaptation for Neural Networks via Moment Alignment

Session 2. Chair: Susanne Saminger-Platz

10:45	U. Anlauf:
	The Steiner Tree Problem Considering Obstacles
11:00	A-M. Meder:
	Optimization of Electrical Drives Using Deep Learning Techniques
11:15	T. Natschläger:
	Workflow for Criticality Assessment Applied in Biopharmaceutical Process Validation Stage 1

Statistical Depth

Florian Sobieczky – scch

2017-11-13

Overview

Idea of Statistical Depth

For a given statistical model $\langle S, \mathcal{F}, \mathcal{P} \rangle$ to give an estimate of the probability density of the distribution from which a sample $X \in \mathbb{R}^{N \times d}$ is drawn.

- Tukey, 1975 [1]

1. Definition of Halfspace Depth

Statistical Depth of a point in a sample space of a statistical model is relative to a given sample $X \in \mathbb{R}^{n \times d}$:

- ▶ supposed to give an approximation of the density function of the distribution from which the sample has been drawn.

1. Location Depth (Halfspace-Depth, [2]):

$$\mathrm{dep}_n(x, X) = \min |\{v \in \mathbb{R}, y\mathrm{row}(X) \mid \|v\| = 1, \ \langle (x - y), v \rangle \geq 0\}|$$

- ▶ Restriction on $\mathcal{P}$: Unimodal Distributions with convex contourplanes
- ▶ High Robustness: Breakpoint is fraction of removable data points without estimator turning 'bad'

1. Other Depth Notions

Convex Peeling:

- ▶ Remove consecutive convex hulls
- ▶ Not robust: Outer Contours depend heavily on configuration of data points

Mahalanobis Depth:

- ▶ Remove consecutive *furthest* outliers with resp. to Mahalanobis distance
- ▶ Not Robust: 'Mean' instead of Median is used

Oja Depth (1983): Volume of simplices data point x falls into

Simplicial Depth (Liu, 1990): Number of simplices x falls into

2. Robustness (Now classical results)

- Halfspace depth has *high breakpoint* (Donoho and Gasko[2]):

1.)
$$\liminf_{n \to \infty} \max \left(\mathrm{dep}_n(x)/n \right) \geq \frac{1}{d+1}$$

2.) For centrosymmetric underlying distributions:

$$\liminf_{n \to \infty} \max \left(\mathrm{dep}_n(x)/n \right) \geq \frac{1}{2}$$

More precisely: $\max \left(\mathrm{dep}_n(x)/n \right) > \frac{1}{2} - O\left(\frac{1}{\sqrt{n}} \right)$

3.) For distributions ϵ-close to centrosymmetry:

$$\max \left(\mathrm{dep}_n(x)/n \right) > \frac{1}{2}(1 - \epsilon)$$

3. Implementation

Rousseew and Struyf have found fast algithms for computing $\mathrm{dep}(x, X)$ in

- ▶ $O(N \log N)$ steps for $d = 2$,
- ▶ $O(N^2 \log N)$ steps for $d = 3$.

Moreover, they define the *regression depth* as the number of sign flips of residuals in a linear regression to be performed when shifting approximating hyperplane to lie 'outside' of data cloud. This can be computed in

- ▶ $O(N^2 \log N)$ steps for arbitrary dimenion.
- ▶ Generalizing concept of depth which encompasses location depth and regression depth: [4].

3. R-Package 'depth'
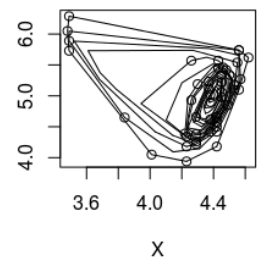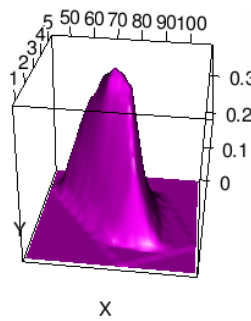

```
library(depth)
A<-matrix(rnorm(100), ncol=2)
depth::depth(c(0,0), A)
```

```
## [1] 0.34
```

Literature

[1] J. Tukey: 'Mathematicss and the picturing of data', 1975, Proc. Int. Congr. Mathem., 2, Vancouver, pp. 523-531

[2] D.L. Donoho, M. Gasko: 'Breaking Properties of Location Estimates Based on Halfspace Depth', 1992, Ann. Stat, 20, 4, 1803

[3] P. J. Rousseeuw, A. Struyf: 'Computing location depth and regression depth in higher dimensions', 1998, Statist. Comput., 8, pp 193-203

[4] M. Hubert, P. J. Rousseeuw, Van Alst: 'Similarities between location depth and regression depth', 2001, in: Statistics in Genetics and in the Environmental Sciences, Birkhauser, pp. 159-172

[5] P. J. Rousseeuw, M. Hubert: 'Statistical depth meets computational geometry: a short survey', arxiv 1508.03828

Domain – Invariant Partial Least Squares Regression: A Novel Calibration Transfer Paradigm

Ramin Nikzad-Langerodi [1], Werner Zellinger [1], Edwin Lughofer [1], Thomas Reischer[2] , Susanne Saminger-Platz [1]

[1] Department of Knowledge-Based Mathematical Systems, Johannes Kepler University, Linz, Austria, Ramin.Nikzad-Langerodi@jku.at, [2] Metadynea GmbH, Krems, Austria.

Calibration transfer (CT) is an important and widely studied problem in chemometrics. However, most of the current CT methods (e.g. DS, PDS) are limited in terms of application scope, i.e. they learn a mapping between the signals of calibration standards recorded on two devices in order to correct for the difference in the instruments' response [1]. In the absence of calibration standards, adaptation of calibration models is usually achieved by acquiring additional reference measurements in the target domain in order to account for any source of new variation [2]. In the current contribution, we introduce a novel extension to partial least squares (PLS) regression involving domain regularization in order to enforce that source (master) and target (slave) domain data are aligned in the latent variable subspace (Figure 1). Notably, the proposed extensions allow incorporation of labeled and unlabeled data (i.e. data without reference values) from source and target domains (semi-supervised learning) with the option to adapt source calibration models completely unsupervised, without the need for reference measurements in the target domain. The main focus of this contribution will be set on basic concepts of domain regularization, technical aspects of domain-invariant latent variable extraction and model selection. Application of domain-invariant PLS (di-PLS) will be exemplified on a real-world near infrared spectroscopy dataset from a Melamine Formaldehyde resin production plant, where the aim is to adapt a source calibration model to a target domain with altered MF composition.
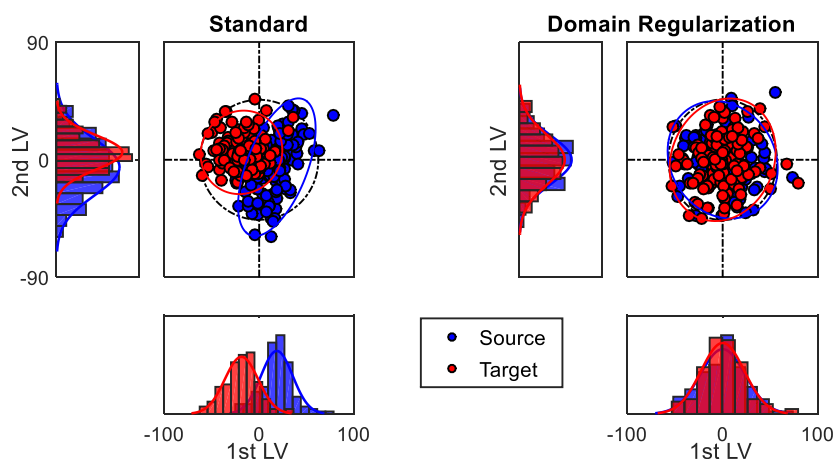
Figure 1. Distribution of the projections (scores) from source (blue) and target (red) data in a two dimensional latent variable (LV) subspace without (left) and with (right) domain regularization.

[1] Yongdong. Wang, David J. Veltkamp, and Bruce R. Kowalski, *Analytical Chemistry* **1991** *63* (23), 2750-2756

[2] John H. Kalivas, Gabriel G. Siano, Erik Andries, Hector C. Goicoechea, *Applied Spectroscopy* **2009,** 63 (7), 800 - 809

Unsupervised Domain Adaptation for Neural Networks via Moment Alignment

Werner Zellinger, Bernhard A. Moser, Thomas Grubinger, Edwin Lughofer,
Thomas Natschläger, and Susanne Saminger-Platz

Abstract

A novel approach for unsupervised domain adaptation for neural networks is proposed that relies on a metric-based regularization of the learning process. The metric-based regularization aims at domain-invariant latent feature representations by means of maximizing the similarity between domain-specific activation distributions. The proposed metric results from modifying an integral probability metric in a way such that it becomes translation-invariant on a polynomial reproducing kernel Hilbert space. The metric has an intuitive interpretation in the dual space as sum of differences of central moments of the corresponding activation distributions. As demonstrated by an analysis on standard benchmark datasets for sentiment analysis and object recognition the outlined approach shows more robustness w. r. t. parameter changes than state-of-the-art approaches while achieving even higher classification accuracies.

Index Terms

transfer learning, domain adaptation, neural networks, moment alignment, integral probability metric

I. Outline

THE collection and preprocessing of large amounts of data can be time consuming and expensive. This can limit the application of many state-of-the-art methods like deep neural network architectures that require large amounts of data. Often, collecting unlabeled data is cheap and large amounts of labeled data with a different but related distribution are already available. One important example is in the sentiment analysis of product reviews [1] where only unlabeled data is available for a specific product category, e. g. kitchen appliances, but labeled data often exists for related categories, e. g. books. A second example is the training of image classifiers on unlabeled real images by means of nearly-synthetic images that are fully labeled but which have a distribution that is different [2], [3]. Another example is the content-based depth range adaptation of unlabeled stereoscopic videos by means of labeled data from movies [4], [5].

The problem of training a machine learning model in the presence of different training and test distributions is known as *domain adaptation* [3], [6]–[9]. The goal of domain adaptation is to build a model that performs well on a *target* distribution while it is trained on a different but related *source* distribution.

It is shown, that a classifier's error on the target domain can be bounded in terms of its error on the source domain and a difference between the source and the target domain distribution [10]. This motivated many approaches in the past, to first extract features that overcome the distribution difference and subsequently minimize the source error [8], [11]–[13]. With the recent developments in representation learning, approaches have been developed that embed domain adaptation in the feature learning process. One way for doing so is by minimizing a combined objective that ensures both, a small source error and feature representations that overcome the domain difference [3], [14], [15], see Fig. 1.

While much research has been devoted to the question of how to supervised minimize the source error [16], [17], relatively little is known about objectives that ensure domain-invariant feature representations. In this contribution we focus on the latter question. In particular, we deal with the task of *unsupervised domain adaptation* where no information about the target labels is available. However, the proposed approach is also applicable under the presence of target labels (*semi-supervised domain adaptation*).
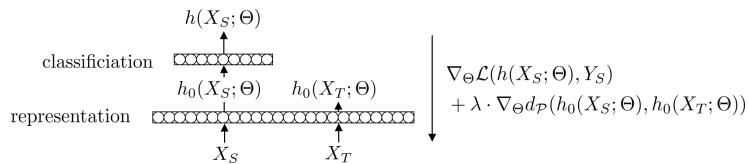
Fig. 1: Schematic sketch of a feed-forward neural network $h(X; \Theta)$ with parameters Θ optimized via gradient descent based on the minimization of a source loss $\mathcal{L}(h(X_S; \Theta), Y_S)$ and the minimization of a distance $d_{\mathcal{P}}$ between the activations $h_0(X_S; \Theta)$ and $h_0(X_T; \Theta)$ of the source sample X_S and the target sample X_T, where Y_S denotes the labels in the source domain. The minimization of $d_{\mathcal{P}}$ ensures domain-invariant representations. ∇_{Θ} refers to the gradient w. r. t. Θ.

We aim for a robust objective function. That is, (a) the convergence of our learning algorithm to sub-optimal solutions should guarantee similar domain-specific activation distributions and (b) the accuracy of our learning algorithm should be insensitive to changes of the hyper-parameters. The latter property is especially important in the unsupervised problem setting since the parameters have to be selected without label information in the target domain and the application of parameter selection routines for hierarchical representation learning models can be computationally expensive.

One simple idea to approach both properties is to minimize an integral probability metric [18] between the domain-specific hidden activation distributions that is based on a polynomial function space [19]–[22]. However, it can be shown that instability issues arise when higher order polynomials are considered. We solve these issues by modifying the integral probability metric in a way such that it becomes translation-invariant on a polynomial reproducing kernel Hilbert space. We call the metric the Central Moment Discrepancy (CMD). The CMD has an intuitive representation in the dual space as sum of differences of central moments of the corresponding distributions.

We propose a robust domain adaptation algorithm for the training of neural networks that is based on the minimization of the CMD. The classification performance and accuracy sensitivity w. r. t. parameter changes is analysed on artificial data as well as on benchmark datasets for sentiment analysis of product reviews [12] and object recognition [23].

The main contributions of this work are as follows.

- We propose a novel approach for unsupervised domain adaptation for neural networks that is based on a metric-based regularization of the learning process. We call the metric the Central Moment Discrepancy (CMD).
- We prove several appealing properties of the CMD including its computationally efficiently implementable dual representation, a relation to weak convergence of distributions and a strictly decreasing upper bound for its moment terms.
- Our algorithm outperforms comparable approaches on most domain adaptation tasks on two standard benchmark datasets for sentiment analysis of product reviews and object recognition.

In addition, our approach is robust w. r. t. the following aspects.

- Our approach overcomes instability issues of the learning process by solving the problem of mean over-penalization that arise in the application of integral probability metrics based on polynomial function spaces.
- Our algorithm can be used with a fixed weighting parameter for the domain adaptation regularization. In contrast to state-of-the-art approaches, our results on the object recognition dataset are obtained without data augmentation and without tuning of the learning rate.
- A post-hoc parameter sensitivity analysis shows that the classification accuracy of our approach is not sensitive to changes of the number of moments parameter and changes of the number of hidden nodes.

References

[1] X. Glorot, A. Bordes, and Y. Bengio, "Domain adaptation for large-scale sentiment classification: A deep learning approach," in *International Conference on Machine Learning*, pp. 513–520, 2011.

[2] B. Sun and K. Saenko, "From virtual to reality: Fast adaptation of virtual object detectors to real domains.," in *British Machine Vision Conference*, 2014.

[3] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *Journal of Machine Learning Research*, vol. 17, no. Jan, pp. 1–35, 2016.

[4] W. Zellinger and B. Moser, "Improving visual discomfort prediction for stereoscopic images via disparity-based contrast," *Journal of Imaging Science and Technology*, vol. 60, no. 1, pp. 1–8, 2016.

[5] W. Zellinger, B. Moser, A. Chouikhi, F. Seitner, M. Nezveda, and M. Gelautz, "Linear optimization approach for depth range adaption of stereoscopic videos," *Stereoscopic Displays and Applications XXVII, IS&T Electronic Imaging*, 2016.

[6] J. Blitzer, R. McDonald, and F. Pereira, "Domain adaptation with structural correspondence learning," in *Conference on Empirical Methods in Natural Language Processing*, pp. 120–128, Association for Computational Linguistics, 2006.

[7] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineeringn*, vol. 22, no. 10, pp. 1345–1359, 2010.

[8] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Transactions on Neural Networks*, vol. 22, no. 2, pp. 199–210, 2011.

[9] S. Li, S. Song, and G. Huang, "Prediction reweighting for domain adaptation," *IEEE Transactions on Neural Networks and Learning Systems*, 2017.

[10] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. W. Vaughan, "A theory of learning from different domains," *Machine learning*, vol. 79, no. 1-2, pp. 151–175, 2010.

[11] S. J. Pan, X. Ni, J.-T. Sun, Q. Yang, and Z. Chen, "Cross-domain sentiment classification via spectral feature alignment," in *International Conference on World Wide Web*, pp. 751–760, ACM, 2010.

[12] M. Chen, Z. Xu, K. Weinberger, and F. Sha, "Marginalized denoising autoencoders for domain adaptation," *International Conference on Machine Learning*, pp. 767–774, 2012.

[13] K. Zhang, V. Zheng, Q. Wang, J. Kwok, Q. Yang, and I. Marsic, "Covariate shift in hilbert space: A solution via sorrogate kernels," in *International Conference on Machine Learning*, pp. 388–395, 2013.

[14] M. Long, Y. Cao, J. Wang, and M. Jordan, "Learning transferable features with deep adaptation networks," in *Proceedings of the International Conference on Machine Learning*, pp. 97–105, 2015.

[15] B. Sun and K. Saenko, "Deep coral: Correlation alignment for deep domain adaptation," in *Computer Vision–ECCV 2016 Workshops*, pp. 443–450, Springer, 2016.

[16] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[17] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.

[18] A. Müller, "Integral probability metrics and their generating classes of functions," *Advances in Applied Probability*, vol. 29, no. 2, pp. 429–443, 1997.

[19] B. K. Sriperumbudur, A. Gretton, K. Fukumizu, B. Schölkopf, and G. R. Lanckriet, "Hilbert space embeddings and metrics on probability measures," *Journal of Machine Learning Research*, vol. 11, no. 4, pp. 1517–1561, 2010.

[20] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, "A kernel two-sample test," *Journal of Machine Learning Research*, vol. 13, no. 3, pp. 723–773, 2012.

[21] Y. Mroueh, T. Sercu, and V. Goel, "Mcgan: Mean and covariance feature matching gan," *International conference on Machine Learning*, 2017.

[22] C.-L. Li, W.-C. Chang, Y. Cheng, Y. Yang, and B. Póczos, "Mmd gan: Towards deeper understanding of moment matching network," *Advances in Neural Information Processing Systems*, 2017.

[23] K. Saenko, B. Kulis, M. Fritz, and T. Darrell, "Adapting visual category models to new domains," in *European Conference on Computer Vision*, pp. 213–226, Springer, 2010.

The Steiner Tree Problem Considering Obstacles

Ulrike Anlauf

Knowledge-Based Mathematical Systems (KBMS) - Johannes Kepler University Linz

Abstract – A first glance at the Euclidean Steiner Tree Problem and its obstacle-avoiding variant by means of evolutionary computation techniques.

Optimization of Electrical Drives Using Deep Learning Techniques

Adela-Maria Meder

Knowledge-Based Mathematical Systems (KBMS) - Johannes Kepler University Linz

Abstract – In order to be effective in electrical drive-design use cases, multi-objective optimization algorithms must rely heavily on model-based surrogate evaluators (i.e., regression models) that replace the finite element simulations. Surrogates based on various machine learning paradigms (like shallow multi-layer perceptrons, support vector machines, radial basis functions) have been previously tested with mixed success. As recent types of deep structured neural networks have shown very promising results in several application fields, the goal is to test the potential of these advanced machine learning techniques in the context of existing electrical drive design frameworks.

Workflow for Criticality Assessment Applied in Biopharmaceutical Process Validation Stage 1

Thomas Zahel [1], Lukas Marschall [1], Sandra Abad [2], Elena Vasilieva [2], Daniel Maurer [2], Eric M. Mueller [3], Patrick Murphy [3], Thomas Natschläger [4], Cécile Brocard [2], Daniela Reinisch [2], Patrick Sagmeister [1] and Christoph Herwig [1,*]

[1] Exputec GmbH, Mariahilferstraße 147, Postcode Vienna, Austria; thomas.zahel@exputec.com (T.Z.); lukas.marschall@exputec.com (L.M.); patrick.sagmeister@exputec.com (P.S.)

[2] Boehringer Ingelheim RCV GmbH & Co KG, Doktor-Boehringer-Gasse 5-11, Postcode Vienna, Austria; sandra.abad@boehringer-ingelheim.com (S.A.); elena.vasilieva@boehringer-ingelheim.com (E.V.); daniel.maurer@boehringer-ingelheim.com (D.M.); cecile.brocard@boehringer-ingelheim.com (C.B.); daniela.reinisch@boehringer-ingelheim.com (D.R.)

[3] Versartis Inc., 4200 Bohannon Drive, Suite 250, Menlo Park, Postcode CA, USA; emueller@versartis.com (E.M.M.); pmurphy@versartis.com (P.M.)

[4] Software Competence Center Hagenberg, Softwarepark 21, 4232 Hagenberg, Austria; Thomas.Natschlaeger@scch.at (T.N.)

* Correspondence: christoph.herwig@exputec.com; Tel.: +43-1-58801-166400

Abstract: Identification of critical process parameters that impact product quality is a central task during regulatory requested process validation. Commonly, this is done via design of experiments and identification of parameters significantly impacting product quality (rejection of the null hypothesis that the effect equals 0). However, parameters which show a large uncertainty and might result in an undesirable product quality limit critical to the product, may be missed. This might occur during the evaluation of experiments since residual/un-modelled variance in the experiments is larger than expected a priori. Estimation of such a risk is the task of the presented novel retrospective power analysis permutation test. This is evaluated using a data set for two unit operations established during characterization of a biopharmaceutical process in industry. The results show that, for one unit operation, the observed variance in the experiments is much larger than expected a priori, resulting in low power levels for all non-significant parameters. Moreover, we present a workflow of how to mitigate the risk associated with overlooked parameter effects. This enables a statistically sound identification of critical process parameters. The developed workflow will substantially support industry in delivering constant product quality, reduce process variance and increase patient safety.

Keywords: retrospective power analysis; process characterization study; process validation stage 1; criticality assessment; control strategy; design of experiments

1. Introduction

Process validation of pharmaceutical processes aims to demonstrate the capability of the process to constantly deliver high product quality [1,2]. Most of the warning letters connected to process validation are raised due to flaws in stage 1 [3]. The aim of process validation stage 1 is to identify a robust process design that enables the ability to constantly deliver product quality. Therefore, it is key to identify critical process parameters (CPPs) that are likely to create risk to critical quality attributes (CQAs) and set up control strategies for these CQAs. Thereby it is possible to reduce out-of-specification (OOS) events, recalls, and ultimately risk to the patient. At process validation stage

1, it is of the highest priority not to overlook a CPP in the design of the process, which as a consequence might not be controlled properly.

In order to accomplish this goal, the following steps are commonly undertaken in industry to characterize process design following a risk-based approach:

1. Risk assessment: to identify potential influential/critical parameters for each unit operation. This is usually performed using tools such as failure mode and effect analysis (FMEA) [4,5]. Ranking of potential criticality is performed using expert knowledge, historical process data, and interdependencies identified in development data.
2. Scale down model establishment: Due to the costs related to large-scale experiments, in biopharmaceutical manufacturing it is necessary to develop appropriate scale down models (SDMs) that are appropriate to investigate the interdependency between process parameters and quality attributes.
3. Experimental designs: Design of Experiments are applied to quantify the impact of process parameters (PPs) on CQAs. Prior to conducting experiments, a priori power analysis is a good practice to evaluate if an effect that leads to a change in product quality—in the following defined as a critical effect—can be detected by the proposed design setting. Statistical power is defined as the probability that we are able to detect an effect if it is truly there [6]. This is done for a priori analysis by estimating the expected signal to noise ratio, which is thought to occur during the experiments [7]. As a result of this a priori power analysis, the number of required experiments, the intended screening range, or the design itself might be adjusted. After a sufficient power can be expected, potential influential/critical parameters are purposefully varied within experiments, which is done for each unit operation separately using the previously established SDMs.
4. Criticality assessment of process parameters by evaluating experimental designs: Identification of significant factors (rejection of the null hypothesis that the effect equals 0) at a desired significance level (typically $\alpha < 0.05$) is performed using Pareto charts and analysis of significance of regression coefficients by means of ANOVA. Misleadingly, this does not imply that for non-significant factors the null hypothesis is true and their effect is zero [8]. Rather, it indicates that the uncertainty around these factors in the range examined—often indicated by large confidence intervals around the effect—is large and critical levels cannot be excluded. Commonly, only significant factors that have been observed to impact product quality or process performance are defined as critical or key, respectively. Those which cannot be stated as significantly impacting are stated as non-critical or non-key, respectively.
5. Definition of control strategy: As a means to ensure all CQAs and quality specifications are met, a process control strategy for all critical and key process parameters must be put in place. Moreover, it has to be evaluated whether their mutual worst case setting would lead to acceptable product quality levels. Commonly for biopharmaceutical production, this is accomplished by setting normal operating ranges (NOR) and proven acceptable ranges (PAR).

Although all steps are equally important to design a robust process, we frequently observed that, in industry, steps 3, 4 and 5 are more difficult to accomplish in practice. The US Food and Drug Administration (FDA) and other agencies are not prescriptive but clearly state that statistics should be used within all stages of process validation [3]. Multiple statistical tools and software for step 3 (a priori power analysis and design of experiments) and step 4 (statistical analysis of significant parameters) exist, however, the approach of those steps as described above has two major drawbacks: (i) after making several assumptions about the expected noise in the a priori power analysis of step 4, those assumptions are not checked for validity after the experiments have been performed. Especially in biopharmaceutical engineering, reproduction and analytical variability from non-validated methods, which might be used during stage 1 of process validation, as well as unexpected non-linear effects (e.g., edge of failure experiments), may lead to increased noise in the conducted design of experiments (DoEs). (ii) Criticality and potential tightening of the NOR is only taken into account for significant parameters. This might not be sufficient since parameters with large uncertainty around

the estimated effect—those effects, which might be zero, but might be very large, too—can have severe effects on product quality as well.

The first of the mentioned drawbacks can be tackled by retrospective assessment of the actually received power. Although retrospective power analysis is controversially discussed when using the observed variance and observed effect size, it is an appropriate tool when comparing the observed variance in the experiments to a pre-specified critical effect [6,9]. Frequently, retrospective power is calculated using the observed effect size, which leads to uninformative results [10].

Both issues together might lead to situations where the process shows unexpected variability during routine manufacturing. Therefore, we want to present a workflow for criticality assessment that reduces the risk to overlook critical PPs. This is demonstrated based on a process characterization study of a novel long acting human growth hormone product. Exemplarily for two unit operations, we will address the following topics:

- Establishment of a methodology that prevents engineers, during process validation, from overlooking critical parameters;
- Setting a control strategy for critical and likely overlooked parameters that ensures a robust process design;
- A workflow that can be used during stage 1 process validation to assess PP criticality. Applying those guidelines, it will be possible to better understand potential process variability and provide an opportunity to reduce process variability, OOS events, and patient risk.

2. Methods

In the following sections, we describe the biopharmaceutical production process, selection of experimental designs to study the impact of PPs on CQAs (Section 2.1), calculation procedures for critical effects (Section 2.2), an a priori power analysis approach (Section 2.3) applied to assess the ability of the DoE to detect practically relevant (here critical) effects and their statistical evaluation (Section 2.4).

2.1. Description of Process and Design of Conducted Experiments

The workflow for criticality assessment will be presented for two unit operations from a biopharmaceutical manufacturing process producing a recombinant protein. The process consists of an Escherichia coli fermentation, cell lysis, precipitation (PR), clarification (depth filtration), and three subsequent preparative chromatographic columns (CC 1/CC 2/CC 3) for purification. Finally, ultrafiltration/diafiltration is performed to adjust product concentration. For the presented case study for criticality assessment, unit operations CC 1 and the precipitation step were exemplarily chosen.

Risk assessment (FMEA) conducted by process experts showed that five and four PPs respectively, had a high risk priority number and need to be studied experimentally in respect to their influence on CQAs for CC1 and PR, respectively (see Tables 1 and 2). Due to the number of studied PPs for both unit operations, a definitive screening design was chosen [11,12]. Except one parameter (Mixing [Yes/No] for precipitation), all DoE factors are numerically scaled. Small-scale experiments were used to conduct DoEs.

2.2. Calculation of Thresholds for Critical Effects

We formulate a critical gap (CG) as the difference between the performance at set-point conditions and the threshold for each response:

$$CG = theshold_{USL} - \bar{y}(x_{SP}) \tag{1}$$

where $\bar{y}(x_{SP})$ is the response value (here a specific concentration of an impurity) at set-point condition of manufacturing. Since we do not have lower specification limits for the studied impurities, the threshold, which must not be surpassed, is derived from the upper specification limit (USL) of drug substance (DS) specifications. The studied unit operations are at an intermediate stage

of the process. We therefore, calculate the specification limit for the investigated unit operation by multiplying the final DS specifications times the mean specific clearance factors from the manufacturing scale of all unit operations in between. This approach might be refined by including knowledge on increased impurity clearance, e.g., due to spiking studies. Choosing the approach with mean specific clearances might seem conservative, however, it is desirable to reduce the risk of overestimating the specific impurity clearance. The specific clearance factors for each unit operation are defined by:

$$Specific\ Clearance = SC = \frac{c_{CQA,load}}{c_{CQA,pool}} \qquad (2)$$

where cCQA, load and cCQA, pool are the specific concentrations (mg CQA per mg product) of the respective CQA prior to and after the unit operation.

$$threhsold_{USL} = USL * \prod_{u=k}^{U} SC_u \qquad (3)$$

where *u* = *k*, ..., *U* is counting the unit operations from the studied kth unit operation until the last unit operation (U) which equals DS.

2.3. A Priori Power Analysis

We want to investigate if the residual error during evaluation of experimental designs (DoEs) masks effects to an extent such that they could collectively surpass a critical threshold (e.g., specification limit of a specific CQA concentration) within normal operating ranges (see section 0 for calculation of thresholds). Since we are dealing with a multivariate problem, we need to identify how many parameters and to what extent each of those parameters contributes to surpassing such a critical threshold. From a sparsity assumption, it is unlikely that all effects that can be studied using a certain design (e.g., all main effects and interactions effects) are truly present. Therefore, it is a common assumption applied to many statistical packages to study only power of the total number of main effects [13].

Moreover, in multivariate analysis ($p > 1$), infinite combinations of effects of multiple parameters exist that lead to such a critical threshold being surpassed, e.g., the full effect to surpass the critical threshold might be explained solely by the first parameter (P_1) and no effect is present from the residual parameters (P_r), or a fraction of the entire effect is explained by P1 (e.g., 10%) and the residual 90% is equally explained by P_r. Overall, we are interested in the mean chance to detect any of those combinations. Per default, classical statistical software such as JMP (SAS Institute Inc.) or DesignExpert (Stat-Ease, Inc.) only allow for fixed effect power calculation [10,13]. Here, we propose a more general method based on the assumption that the effects are randomly distributed over all parameters. Therefore, we assigned weights to the parameters and varied the fraction/weight of the entire effect that is explained by each parameter gradually between 0.0 and 1.0 (we used a step size of 0.01 in our experiments, i.e., 100 steps) and split the residual effect equally under the residual parameters: $w_i = a$, $w_{j \neq i} = (1 - a)/(p - 1)$, for $a = 0, ..., 1$ and $i = 1, ..., p$. Hence all the weights w_i sum up to 1. In total, we obtain $C = p * 100$ combinations of possible effect distributions and the resulting power values. The mean for each parameter of these recorded power values was taken as the power for this experimental design (see step 6 of the a priori workflow present below).

Herein, the following workflow for a priori power analysis can be formulated:

1. Estimate the mean ($\bar{y}_{SP}$) and variance (σ_{SP}) of the response variable from small-scale or pilot-scale experiments at set point conditions of manufacturing. We assume that residual error in the model is only due to process- and analytical variance. The latter estimate will be used to calculate the expected sum of squares of the residuals ($\widetilde{SS_{res}}$):

$$\widetilde{SS_{res}} = (n - 1) * \sigma^2{}_{SSP} \qquad (4)$$

2. For each of the combinations (c) described above, we calculate critical effects for each parameter using its weight $w_i^{(c)}$:

$$\beta^{(c)}{}_{crit,i} = \frac{w_i^{(c)} * CG}{\max(\text{NORU}_i - sp_i, sp_i - \text{NORL}_i)} \tag{5}$$

In order to estimate the individual coefficient for the *i*-th parameters, from a risk-based approach, we divide by the longest distance from the set-point (sp_i) to the nearest NOR border: where NORU_i is the upper boundary of the NOR and NORL_i is the lower boundary of the NOR of the parameter i. Note that this works for a symmetric as well as asymmetric NOR.

3. Using the design matrix X, obtained for a specific experimental design, we can simulate possible $\tilde{y}$ values at the screening range using:

$$\tilde{y}^{(c)} = X\beta_{crit}^{(c)} \tag{6}$$

4. From that, the total sum of squares can be estimated:

$$\widetilde{SS_{tot}}^{(c)} = \sum_i^n (\widetilde{y_i y}^{(c)} - mean(\tilde{y}^{(c)}))^2 \tag{7}$$

Together with the sum of squares of the residuals, the expected coefficient of variance can be calculated:

$$\tilde{R}^{2(c)} = 1 - \frac{\widetilde{SS_{res}}}{\widetilde{SS_{tot}}^{(c)}} \tag{8}$$

5. Using Choen's effect size (f), the non-centrality parameter λ and the critical F value (F_{crit}), the a priori power for the combination c of effects that no parameter has been overlooked can be calculated [7]:

$$f^{2(c)} = \frac{\tilde{R}^{2(c)}}{1 - \tilde{R}^{2(c)}} \tag{9}$$

$$\lambda^{(c)} = f^{2(c)} * v \tag{10}$$

6. Confidence intervals for the a priori power for the combination c were calculated according to

$$\lambda_{upp}^{(c)} = \lambda^{(c)} * c_{crit}(1 - \alpha|v)/v \tag{11}$$

$$\lambda_{low}^{(c)} = \lambda^{(c)} * c_{crit}(\alpha|v)/v \tag{12}$$

where $c_{crit}(\alpha|v)$ is the $100 * \alpha$ percentile from a χ^2 distribution with v degrees of freedom.

$$F_{crit} = F_{inv}(1 - \alpha| u, v) \tag{13}$$

7.

$$power_{apriori}{}^{(c)} = 1 - F_{nc}(F_{crit}|, u, v, \lambda^{(c)}) \tag{14}$$

where F_{nc} is the non-central F distribution with $u = p$ (number of DoE parameters) and $v = n - u - 1$, where n is the number of observations in the DoE.

8. The mean power over all combinations of effects was estimated as the arithmetic mean of all $power_{apriori}{}^{(c)}$:

$$power_{apriori} = \frac{\sum_{c=1}^{C} power_{apriori}{}^{(c)}}{C} \tag{15}$$

2.4. Evaluation of DoEs

Multiple linear models were used to identify the relationship of the studied PPs (DoE factors, *X*) on the response variable (*y*), representing a CQA or KPI of the process, up to a residual error (*ε*):

$$y = \beta_0 + X\beta + \varepsilon \qquad (16)$$

where *X* is a (*n* × *p*) dimensional design matrix for n DoE runs and p DoE factors which are studied, β_0 is the intercept, β are the true effects of the DoE factors, and ε is the residual, un-modelled error vector. The un-modelled error vector describes the analytical and process variance as well as non-linear effects which cannot be accounted for in the model structure. Identification of significant parameters was done using stepwise regression within the multiple linear regression (MLR) tool of inCyght software (inCyght version 2017.03, Exputec GmbH). Parameters showing a partial *p*-value below 0.05 were allowed to enter the model. Those which showed a *p*-value larger than 0.1 were excluded from the model. Starting with the most significant parameter, this including/excluding procedure was applied iteratively and was repeated till the model structure did not change any more and the optimal model was achieved by this approach; identified significant parameters and their respective *p*-value are shown in Tables 1 and 2 for CC 1 and PR, respectively. The normalized raw data are given in the Supporting Information Tables S1 and S2.

Table 1. *p*-values of significant process parameters that were used in the statistical models for each critical quality attributes (CQA) of CC 1. Normal operating ranges and thresholds are given for each process parameter or critical quality attribute, respectively. Non-significant parameters are indicated with "-". Also, the ratio of standard deviation of raw residuals of the model by the standard deviation at set-point ($\sigma_{residues}/\sigma_{SP}$) is given for each CQA.

		End Pooling [CV]	Elution Strength [mM]	Wash Strength [mM]	Column Loading Density [g/L]	pH [–]	$\dfrac{s_{residues}}{s_{SDM}}$
CQA	NOR [1]	−1.1–0	−1.1–0.65	−1.1–1.1	−0.51–1.1	−0.55–0.55	
	Threshold						
Process impurity 2 clearance	0.85	-	-	0.059	0.099	-	7.79
Product impurity 1 clearance	1.08	0.028	-	0.098	0.089	0.027	18.12
Product impurity 2 clearance	0.1	-	-	-	-	-	256.06

[1] NOR was normalized by the screening range.

Table 2. *p*-values of significant process parameters that were used in the statistical models for each CQA of precipitation (PR). Normal operating ranges or thresholds are given for each process parameter or critical quality attribute. Non-significant parameters are indicated with "-". Also, the ratio of standard deviation of raw residuals of the model by the standard deviation at set-point ($\frac{\hat{\sigma}_{residues}}{\hat{\sigma}_{SP}}$) is given for each CQA.

		Temperature [°C]	Time [Hours]	Mixing [Yes/No]	pH [–]	$\dfrac{\hat{\sigma}_{residues}}{\hat{\sigma}_{SP}}$
CQA	NOR [1]	−1.71–0.41	0.33–0.41	−0.95–0.95	−0.61–0.61	
	Threshold					
Process impurity 1 concentration specific	9×10^5	9×10^{-5} *	-	-	0.07	64.89
Process impurity 2 concentration specific (prior filtration)	9×10^4	-	-	-	-	2.68
Process impurity 2 concentration specific (post filtration)	784.7	-	-	-	0.021	0.55

[1] NOR was normalized by the screening range. * A quadratic effect was modelled for temperature and the shown p-value corresponds to the quadratic effect.

3. Results and Discussion

Experiments performed in biotechnological studies might contain data that violate the statistical assumptions of parametric tests (i.e., normality, homogeneity of variances and independence of errors). Moreover, with a limited number of experiments and a large number of unknown parameters, such assumptions are hard to assess. Consequently, nonparametric approaches bear potential and we want to present a novel permutation test to assess the power of individual DoE factors in a multivariate regression model.

3.1. Permutation Test for Retrospective Power Analysis

The following permutation approach is adapted from a permutation test aiming to investigate power retrospectively [14]. Here, we adapted this approach to study the significance of the alternative hypothesis that critical effects are present. The following steps are performed:

1. Using variable selection procedures, we select a significant regression model (all included effects are not 0 to a certain significance level):

$$y = \beta_0 + \beta_s * X_s + R_{y|X_s} \tag{17}$$

   where X_s denotes the s significant parameters selected from a variable selection procedure (e.g., stepwise variable selection) and $R_{y|X_s}$ are the residuals of the obtained model. A list of those significantly selected parameters for the case studies of this work can be found in Tables 1 and 2.

2. We define a critical gap (CG) that we must not surpass as the difference of the threshold and the worst case model prediction within the NOR ($x_{worst\ case_{NOR}}$), which is the parameter setting where the model prediction ($\hat{y}(x)$) is closest to the $threshold_{USL}$:

$$CG = theshold_{USL} - \hat{y}(x_{worst\ case_{NOR}}) \tag{18}$$

3. Similar to the approach discussed in section 0 for the a priori power analysis, for non-significant parameters, a variety of combinations (in total C) of effects for those parameters exist that lead to surpassing a critical threshold. In order to estimate the mean likelihood of not overlooking a specific parameter, we vary the relative impact on the threshold of each parameter gradually between 0 and 1 in 100 steps. The fraction of the CG which is attributed to the non-significant parameter i is expressed as the weight $w_i^{(c)}$ for the combination c. Equation (5) can be used to calculate the critical effect of the parameter i.

4. The residuals $R_{y|X_s}$ are permuted randomly, producing $R^*_{y|X_s}$.

5. New response values are calculated from the permuted residuals assuming that the critical effect is present under the alternative hypothesis (H_A):

$$y^* = \beta_0 + \beta_s * X_s + \beta^{(c)}_{crit} * Z + R^*_{y|X_s} \tag{19}$$

   where $\beta^{(c)}_{crit}$ is a vector of regression coefficients for the non-significant parameters and Z is the design matrix for all non-significant parameters.

6. Make a model for y^* based on X and Z and record significance of $\hat{\beta}_{crit}$ at a certain significance level (here $\alpha = 0.05$)

7. Repeat steps 4, 5 and 6 a large number of times (here 1000) and count the number of significant outcomes for each $\hat{\beta}_{crit,i}$ at a certain significance level (here $\alpha = 0.05$). The fraction of significant outcomes of all iteration cycles equals the retrospective power of parameter i.

3.2. Comparison of a Priori and Retrospective Power

If we apply the proposed retrospective power analysis permutation test of section 0 to experimental data recorded from two unit operations (CC1 and PR), we obtain power values for each PP/CQA combination from Tables 1 and 2, respectively.

Figure 1A shows a comparison of the retrospective and a priori power analysis for the CC1 unit operation. For all three studied CQAs at this stage ('process impurity 2 clearance', 'product impurity 2 clearance' and 'product impurity 1 clearance'), we obtain a priori estimates of 1 (rightmost bar group in Figure 1A). This indicates an ideal case to start with experiments since there is no chance of

overlooking a critical effect. Retrospective power analysis revealed that all investigated PPs power values are well below the common statistical practice cut-off value of 0.8. This can be explained by the fact that the residual variance in the model is much higher than the initial estimate at the set point, expressed by ratios of $\frac{\hat{\sigma}_{residues}}{\hat{\sigma}_{SP}}$ well above 1, as shown in Table 1. In general, multiple reasons for this discrepancy between the initial guess of expected variance and the actual residual variance in the model might exist. It could be a non-representative selection of set-point runs (e.g., runs conducted with different operators), unexpected increase of variance during experiments (e.g., it is more difficult to control experiments at unusual parameter settings) or even non-linear dependency which cannot be captured by the linear model structure. Although statistically good practice, our experience shows that such non-linear dependencies might not be obvious from analysis of residuals (e.g., investigation of plots of residual vs. DoE factors). In a DoE approach, each experiment is unique in its settings if we do not use replicates and thereby no redundancy is available to hinder the model from being leveraged by non-linear responses.

For the precipitation step (PR), a priori power analysis again suggested a power of 1 (Figure 1B). Retrospectively assessed power values match the results obtained from a priori analysis, indicating that the performed DoE had sufficient power to assess critical effects of process parameters on quality attributes. This is reasonable since ratios of $\frac{\hat{\sigma}_{residues}}{\hat{\sigma}_{SP}}$ are closer to 1 for this unit operation compared to CC 1, as shown in Table 1.
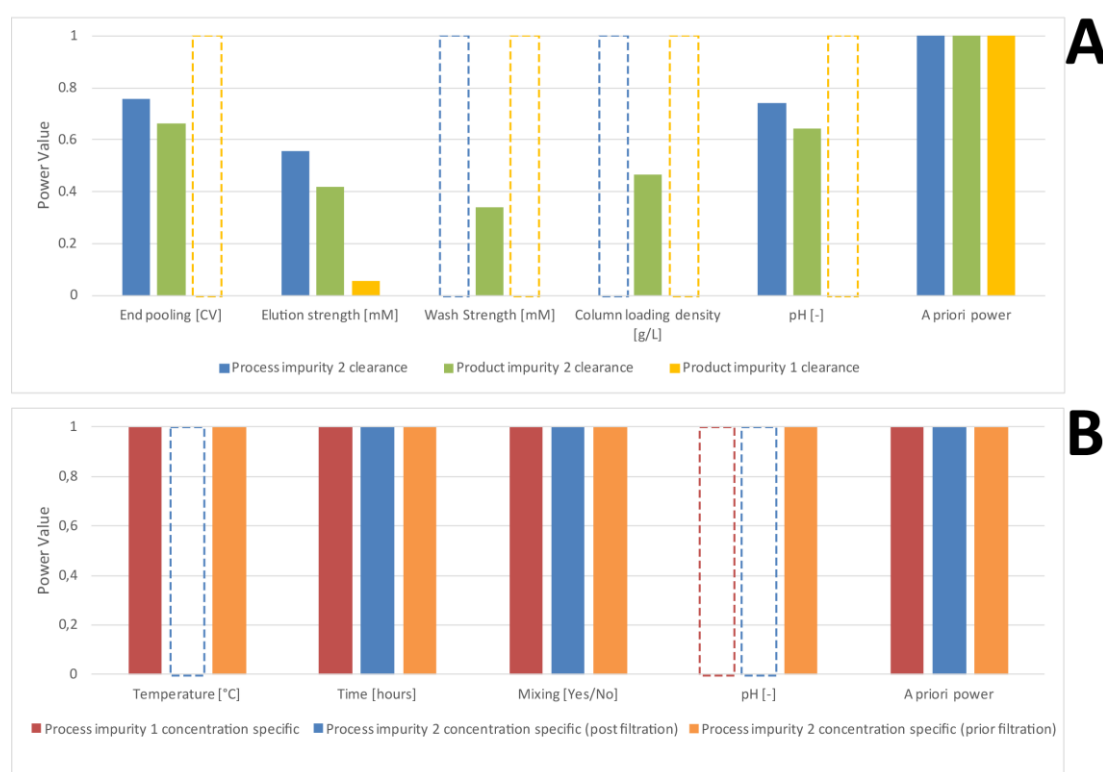

Figure 1. Power values for chromatographic column (CC) 1 (**A**) and PR (**B**) for each process parameter (PP) and CQA. Where significant process parameters were detected for a quality attribute, bars are marked grey. (**A**) Though a priori power analysis suggested a power of 100% for each investigated CQA for chromatography step 1, retrospective power analysis revealed that the power to detect a critical effect did not surpass 80% for any of the investigated process parameters. Strategies to tackle these low-power-situations are given in Figure 4. (**B**) For the precipitation step, a priori power analysis suggested a power of 100% for each investigated CQA as well. Retrospective power confirmed the findings that there is a 100% chance that we did not overlook a critical effect of the investigated process parameters on quality attributes.

*3.3. How to Deal with Low-Powered Parameters?*

The most common approaches to tackle insufficient power values in screening designs are by increasing the sample size, reduction of measurement variance (either analytical or process), increasing the screening range if technically possible, or accepting the lack of power, however stating the parameter as key or critical. The latter strategy will have an impact on the extended monitoring of such parameters during a subsequent process performance qualification (PPQ) campaign and routine manufacturing. As seen in section 0, a priori power analysis suggested high power values for all investigated unit operations, however, drastically overestimated the power for CC1. In specific cases, retrospectively increasing the sample size or the screening range might not be possible due to shortage of starting material or technical limitations. A measurement method with less variance might not be at hand to re-measure backup samples. Another approach made possible by the presented method for power analysis is to narrow the NOR of some process parameters. If the threshold stays the same and the NOR is symmetrically located around the set point, for smaller NORs larger effect sizes are necessary to surpass the critical threshold as shown in Equation (5) (i.e., steeper slopes). As a demonstrating scenario, we have chosen the relatively low power for Product impurity 2 clearance on CC 1 (see Figure 1A). For this response, no significant parameter could be found. Figure 2 shows how a reduction of the NOR of the process parameter, "wash strength", impacts the power of all studied PPs of this unit operation. Upon reducing the initially defined NOR by 50% of its width, the power value for "wash strength" increases from 0.34 to 0.68. As seen in Figure 2, the power values of the residual process parameters' effects on the same quality attribute remained unaffected, neglecting the residual variation caused by the Monte Carlo approach in permutation.
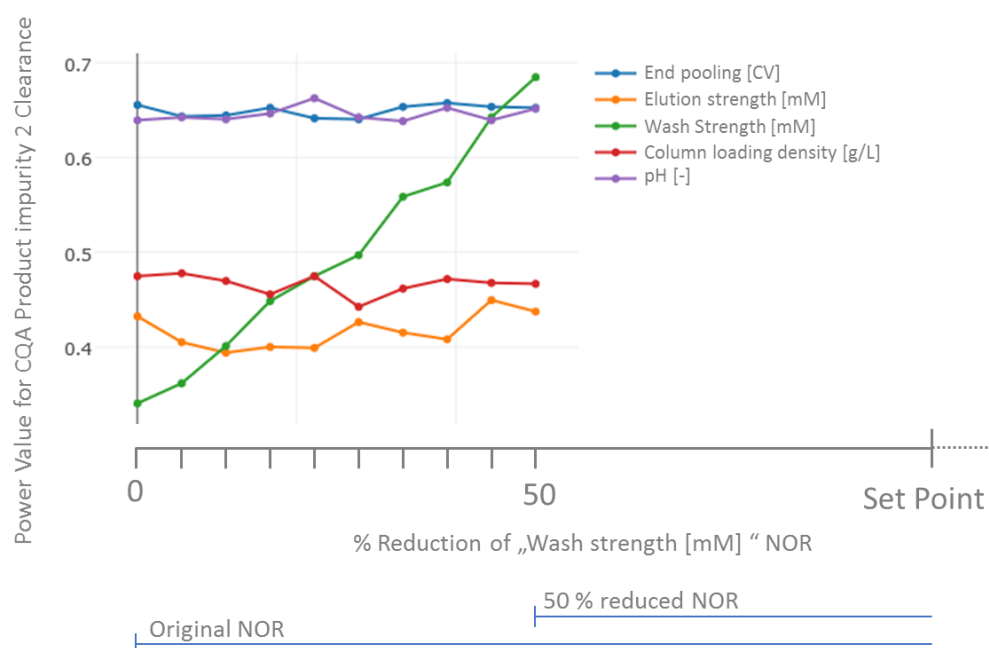
Figure 2. Retrospective power values for product impurity 2 clearance for unit operation CC1 as a function of tightened NOR of process parameter "wash strength". At the initially defined NOR, the power value is 0.34. Upon reducing the NOR symmetrically by 50%, the power value for this process parameter increases to 0.68. The power values of the residual process parameters remain unaffected. The visible variation can be attributed to the variance in the permutation test.

This provides an opportunity to implement a tighter control strategy though adjusting the NOR as an approach to ensure no critical effects have been overlooked. However, it may not be technically feasible or desirable for all process parameters to implement a tighter control strategy with narrower ranges, especially for a parameter that has not been confirmed to significantly impact a CQA. Since a process parameter is studied in respect to multiple CQAs, we want to note that the tightening of a NOR of a process parameter that significantly impacts one specific CQA will also increase the

capacity to not overlook this parameter regarding all other CQAs which have been studied in the same experiment. In contrast to changing the NOR of a non-significant parameter onto a CQA as shown for the combination 'wash strength' onto 'product impurity 2 clearance' in Figure 2, we investigated how the change of a significant parameter impacts power levels. This was exemplarily done for a decrease in NOR of 'wash strength' and we recorded power values for 'process impurity 2 clearance' of all non-significant parameters (here End pooling, elution strength and pH). We can see that due to the reduction of the NOR of a significant parameter, the power values of all non-significant parameters increase too. In detail, a 50% reduction of the NOR of the significantly impacting parameter "wash strength" increases the power of all non-significant parameters by approximately 10%. This can be explained by the fact that the worst case model prediction within the reduced NOR leads to a larger CG as defined in Equation (18). Thereby, the critical effects will also be larger (Equation (6)) and consequently the chances of overlooking larger critical effects will be reduced. In this way, an improved control strategy for a known significant parameter would improve the confidence that all residual non-significant parameters were not overlooked. This is potently a more desirable approach as improved control of known significant parameters is typically required and advantageous, if feasible.
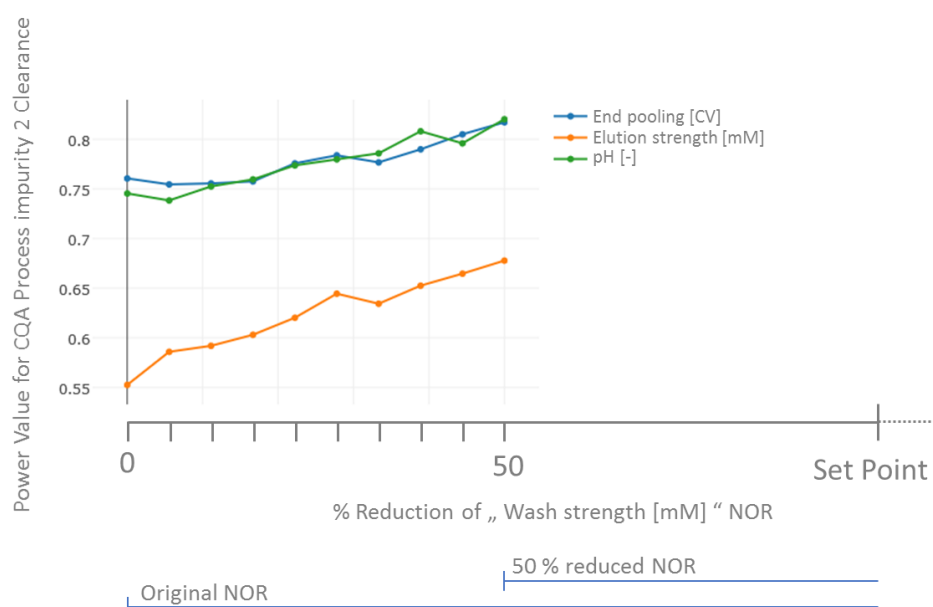

Figure 3. Retrospective power values for 'process impurity 2 clearance' for unit operation CC 1 as a function of tightened NOR of process parameter "wash strength". Since wash strength and column loading density are significant parameters in this model, the power was not assessed for those two parameters. Upon reducing the NOR symmetrically by 50% of the significant parameter "wash strength", power values of all other parameters increase since the critical gap is increased, too, due to a reduction of the worst case model prediction in the NOR (Equation (18)).

3.4. Workflow for Criticality Assessment

In order to summarize the knowledge obtained from the application of the proposed posterior power analysis on two unit operations, we present a workflow that should aid process engineers in assessment of critical parameters (Figure 4). After selection of design and appropriate experiment number, a priori power analysis identifies if it is likely that a critical effect will not be overlooked. Sufficient power levels are normally assumed at 0.8 to 0.9. In cases where sufficient power cannot be assumed, the number of experiments, type of design or screening range must be increased. Both add to the expected signal to noise ratio. When increasing the screening range, care must be taken not to incur failure in experiments due to technical limitations or likely interaction effects (edge of failure). In order to reduce the risk of edge of failure experiments, it is beneficial to conduct an expected

worst case scenario of the process parameters first and potentially revise the screening range afterwards.

In case sufficient power can be assumed, experiments can be conducted and regression modelling can be performed together with selection of significant DoE factors/parameters. After the "optimal" model was selected with its significant factors, retrospective power analysis, as shown in section 0, will estimate the chances that the residual non-significant factors might contribute to effects that surpass a pre-specified critical threshold. In case all non-significant parameters show power values well above 0.8 to 0.9, all of them can be stated as non-critical since the residual chance that they have been overlooked is only 20 to 10%, respectively. Otherwise, for those parameters that show insufficient power, analytical and/or reproducibility variance might be lowered by re-measurement of the samples or re-conducting of experiments, respectively. Another option is to narrow the NOR of potentially overlooked parameters which show large variability. This decreases their respective critical effect according to Equation (5). After one of those three countermeasures has been taken, retrospective power analysis can be repeated to ensure sufficient power values are reached and all parameters can be stated as non-critical. If none of the above three options is technically feasible or desirable, potentially overlooked parameters should be stated as critical and monitored during process performance qualification (PPQ) runs or routine manufacturing.
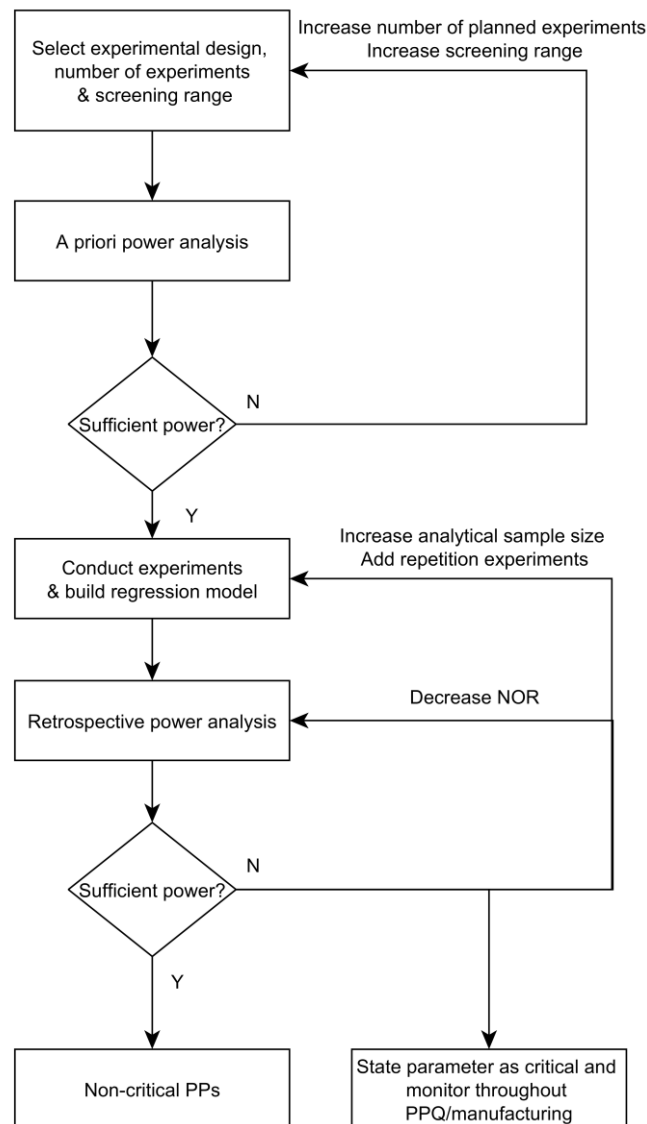
Figure 4. Workflow for criticality assessment of process parameters during process validation stage 1.

4. Conclusions

The goal of the contribution was to demonstrate the capability of a multivariate retrospective power analysis methodology to identify critical process parameters during pharmaceutical process validation stage 1.

We have shown in a case study that parameters that are non-significant in models, which were initially thought to be sufficiently powerful to identify critical effects, might still show effects that surpass a critical threshold due to increased analytical, process, or reproducibility variance. This leads to situations where the impact of those parameters on final drug product quality cannot be excluded. This was shown using a biopharmaceutical case study conducted at a world leading CMO. However, common practice is to state such parameters as non-critical and thereby overlook their potential harmful impact. Therefore, two missing parts have been introduced in this contribution: (i) a novel permutation methodology for multiple linear regression that estimates retrospective power (i.e., the chance of non-significant parameters to mutually combine to a critical effect) and (ii) a workflow for criticality assessment that shows strategies of how to mitigate the risk of low-powered parameters. Besides the well-known fact that an increase in experiments increases power, it could be shown that a reduction of the NOR of significant parameters increases the power of all non-significant parameters via a reduction of the worst case model predictions; a reduction of the NOR of a specific non-significant parameter increases power solely for this parameter. Additionally, if implementation of tighter NOR controls is practically infeasible, this methodology can, at a minimum, appropriately assess the process risk and increase awareness of the limitations of the initial classification, potentially suggesting that an improved control strategy is required.

Using both tools, it will be possible for process engineers during the design stage of a process validation (stage 1) to:

- reduce the chance of overlooking potential CPPs
- develop a control strategy for potentially overlooked CPPs in order to increase process robustness
- lower OOS events and finally contribute to increased patient safety.

References

1. Ahir, K.B.; Singh, K.D.; Yadav, S.P.; Patel, H.S.; Poyahari, C.B. Overview of Validation and Basic Concepts of Process Validation. *Sch. Acad. J. Pharm.* **2014**, *3*, 178–190.
2. *FDA Guidance for Industry*; 2011.
3. Katz, P.; Campbell, C. FDA 2011 process validation guidance: Process validation revisited. *J. GXP Compliance* **2012**, *16*, 18.
4. Mollah, A.H. Application of failure mode and effect analysis (FMEA) for process risk assessment. *BioProcess Int.* **2005**, *3*, 12–20.
5. ICH. *ICH Harmonised Tripartite Guideline. Pharmaceutical Development Q8 (R2)*; Current Step 4 Version; Publisher: City, Country, 2009.
6. Peres-Neto, P.R.; Olden, J.D. Assessing the robustness of randomization tests: Examples from behavioural studies. *Anim. Behav.* **2001**, *61*, 79–86.

7. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*; Revised Edition; Academic Press: New York, NY, USA, 1977; ISBN 978-0-12-179060-8.

8. Nickerson, R.S. Null hypothesis significance testing: A review of an old and continuing controversy. *Psychol. Methods* **2000**, *5*, 241.

9. Thomas, L. Retrospective Power Analysis. *Conserv. Biol.* **1997**, *11*, 276–280, doi:10.1046/j.1523-1739.1997.96102.x.

10. Thomas, L.; Krebs, C.J. A review of statistical power analysis software. *Bull. Ecol. Soc. Am.* **1997**, *78*, 126–138.

11. Jones, B.; Nachtsheim, C.J. A class of three-level designs for definitive screening in the presence of second-order effects. *J. Qual. Technol.* **2011**, *43*, 1.

12. Tai, M.; Ly, A.; Leung, I.; Nayar, G. Efficient high-throughput biological process characterization: Definitive screening design with the Ambr250 bioreactor system. *Biotechnol. Prog.* **2015**, *31*, 1388–1395.

13. Shari, K.; Pat, W.; Mark, A. *Handbook for Experimenters*; Publisher: City, Country, Year.

14. Freedman, D.; Lane, D. A Nonstochastic Interpretation of Reported Significance Levels. *J. Bus. Econ. Stat.* **1983**, *1*, 292–298.