Review

About Model Validation in Bioprocessing

Vignesh Rajamanickam [1], Heiko Babel [2], Liliana Montano-Herrera [3], Alireza Ehsani [3], Fabian Stiefel [3], Stefan Haider [1], Beate Presser [3] and Bettina Knapp [3,*]

[1] Boehringer Ingelheim RCV GmbH & Co KG1, Biopharmaceuticals Austria, Dr. Boehringer Gasse 5-11, A-1121 Wien;

[2] Boehringer Ingelheim Pharma GmbH & Co.KG, Biopharmaceuticals Germany, Birkendorfer Strasse 65, D-88397 Biberach a. d. Riss, Germany;

[3] Boehringer Ingelheim Pharma GmbH & Co.KG, Development Biologicals Germany, Birkendorfer Strasse 65, D-88397 Biberach a. d. Riss, Germany;

* Correspondence: bettina.knapp@boehringer-ingelheim.com;

Abstract:

In bioprocess engineering the Qualtiy by Design (QbD) initiative encourages the use of models to define design spaces. However, clear guides on how models for QbD are validated are still missing. In this review we provide a comprehensive overview about validation methods, mathematical approaches and metrics currently applied in bioprocess modeling. The methods cover analytics for data used for modeling, model training and selection, measures for predictiveness and model uncertainties. We point out general issues in model validation and calibration for different types of models and put this into context of existing health authority recommendations. This review provides the start-point for developing a guidance for model validation approaches. There is no one-fits-all approach but this review shall help to identify the best fitting validation method or combination of methods for the specific task and type of bioprocess models that is developed.

Keywords: bioprocess models; model validation; model calibration; Quality by Design; mechanistical and statistical models; hybrid models; chemometric models; Biopharmaceutical engineering; regulatory guidance

1. Introduction

During the last years, the biopharmaceutical industry aims at developing biopharmaceutical products and the corresponding process in a QbD manner instead of a quality by testing (QbT) approach [1]. Process analytical technology (PAT) initiatives have also been proposed by the regulatory authorities to enhance process understanding and control [2]. In QbD and PAT, the ultimate goal is to gain model predictive control (MPC) of the process to improve process performance and control of the critical quality attributes (CQAs) by advanced monitoring and control (AM&C) of key and critical process parameters (KPPs and CPPs). CPPs are according to ICH Q8 process parameters whose variability within defined ranges have an influence on one or many CQAs [3]. KPPs show an influence on process performance parameters. To control that the active pharmaceutical ingredient is produced with the desired quality and performance, these parameters have to be monitored or controlled. To do this, well designed measurements (e.g. in a design of experiments (DoE) workflow) of KPPs and CPPs are necessary as well as a corresponding process model which describes the dependencies between CQAs with the process parameters. During upstream processing online sensors or offline measurements address the monitoring and testing precondition. One possibility of upstream online monitoring in procaryotes is via online sensors e.g. for turbidity or metabolite probes (e.g. with Raman [4, 5] or MIR [4, 6]). However, the dynamic behavior of the cells during upstream processing and the dependencies of the parameters are, especially in mammalian cell culture, rarely understood in detail. Many approaches exist to describe and model the biological behavior of cells in upstream biopharmaceutical manufacturing by using small experimental based models (e.g. in combinations with DoEs [7-9]), mechanistic models [10] or more big data driven models like machine learning models [11, 12] partly also in in combinations like in hybrid models [13, 14]. However, the quality of the models e.g. in terms of predictability and interpretability has to be evaluated early on. This shall ensure that also later in the context of commercialization the models are well validated to make decisions on the models such as classification of parameters into KPP or CPP or the definition of ranges. So far, no clear recommendations are outlined for model validation and a straight forward and comprehensive workflow is difficult to define. One reason for this might be the diversity of mathematics (e.g. statistical, mechanistical, hybrid, etc.) and the different nature of the underlying data (scale differences, batch versus perfusion mode, sample size differences, and many more). There are no gold standard data sets available as the biology is that diverse (e.g. different host cell lines, different targets, and so on). Thus, there is generally no clear protocol for bioprocess models available which in turn leads to a large diversity of model validation methods [15-17].

In this review, we describe model validation approaches which are used especially during upstream processing in the biopharmaceutical industry. We further discuss challenges and points to be considered when performing model validation. These challenges might arise from the type of the underlying data, the state of the model (e.g. model training and model selection) or the risk of overfitting. Furthermore, we state the regulatory view on model validation methods which is beside the diversity of academic research referring to a few methods only.

2. Model Validation Methods

In this review paper we focus primarily on three group of models: 1) statistical and chemometric models, 2) mechanistic models and 3) hybrid models. For the first group, often design of experiment (DoE) data is used to make statistical models (e.g. response surface models) which describe e.g. the relation of input parameters (i.e. factors such as CPPs and KPPs) on output parameters (i.e. responses such as CQAs) or which are used to find an optimum of a certain parameter (e.g. yield). Basis is mostly a very small and limited set of experimental data. One-factor-at-time (OFATs) experiments which have a sufficiently high statistical power (e.g. above 80%) can be used to model the relationship of CPPs/KPPs on CQAs or process performance parameters as well. Currently, validation of such models is performed mostly via validation experiments [18-20]. However, this is not the scope of the QbD approach at and this implies that quite a variety of experimental data would be required to set up the model, depending on the experimental design and the number of parameters which are evaluated.

Mechanistic models have been developed for different purposes and with different degrees of complexity ranging from simple systems of ordinary differential equations to genome-scale metabolic network models [21]. So-called unstructured mechanistic models describe cellular processes as a black-box and balance the conversion of metabolites into cells and products in the bioreactor using systems of ordinary differential equations [22, 23]. Here, the conversion and growth rates are modeled according to known or hypothesized mechanisms. These models can be naturally extended to balance also intracellular processes by the addition of intracellular compartments. In contrast to dynamic mechanistic models, metabolic network models are often static and valid for a distinct time-period during the bioprocess. Here the steady-state solutions of networks ranging from central-carbon metabolism to “genome”-scale are analyzed [24, 25]. The underlying assumption for these flux models is an intracellular pseudo-steady state i.e. that intracellular conversion rates are much faster than the growth rate or extracellular exchange rates.

Hybrid (semi-parametric) models combine statistical and mechanistic models for describing a system of study [26, 27]. It can be used when the process is too complex to be described mechanistically, or the process data is insufficient for data-driven approaches like response surface models. The mechanistic part of the model has a fixed structure given by knowledge while the other parts do usually not have a fixed structure, but a flexible one which is determined by experimental data [13, 28, 29]. The advantage of a hybrid model is to use data to fill or improve knowledge gaps in first principles or mechanistics. Setbacks for the implementation of hybrid models rely on the difficulty to establish algorithms for the parameter identification which are error prone and laborious. However, once a general hybrid modelling framework is implemented it is possible to reuse it [30].

The validation of any model should depend on the purpose of the model and the type of the used model. One purpose might be to understand the dependencies of input parameters on output parameters to define KPPs and CPPs. However, sometimes the dependencies cannot be understood and modeled in detail, but still it is possible to make predictions for future batches. This is useful when e.g. computing the probability of having out of specification (OOS) runs in future manufacturing.

As mentioned above, the simplest way to validate models is based on data (data-driven validation). This is currently the most often used way as it is also accepted by health authorities (see Section 4). The methods which are widely used during validation of upstream models are listed and described in the following subsections.

2.1 **Crossvalidation**, such as leave-one-out crossvalidation (**LOOCV**) and **leave-multiple-out CV**

Crossvalidation (CV) is an internal re-sampling method where the original dataset is split into training and validation datasets. These datasets are simultaneously used for model building and validation. The model is trained or built with the training dataset by leaving out a part of the original dataset. The validation dataset can be a part of the original dataset or an external dataset used only for validating the model. The trained model is then used to predict the responses based on the validation dataset [31]. CV is often applied to smaller datasets (e.g. sample size < 10) during initial model development and validation. CV at different levels provide important insights regarding model validation (e.g. source of variation, comparability) and to assess the main sources of variation. It is important with CV to include samples at different levels (e.g. scales) in the training and validation datasets to ensure model reliability. Validated R^2 (see Section 2.4) from a smaller CV sample (e.g. LOOCV) is necessary for the initial training validation but not sufficient to ensure predictive performance in the validation and implementation stages [32]. Especially with chemometric and statistical models the underlying stratification, from splitting the original dataset into training and validation, plays an important role and must be taken into account for remediation and improving model performance.

Monte Carlo CV (MCCV) has been shown in applications to be more consistent in comparison to conventional CV approaches [33]. The authors compared three different methods, namely MCCV, LOOCV and k-fold CV, to determine the optimal number of model variables to achieve predetermined prediction accuracy. Model validation using LOOCV renders unnecessarily high number of model variables due to overfitting, which ultimately reduced the predictive capability of the model. Furthermore, when k-fold CV is used the computational power increases exponentially to determine the optimal number of model variables to achieve the predetermined prediction accuracy. Optimization of model parameters has been recently proposed using a two-layered (internal and external) cross validation approach for chemometric models [34]. The model

parameters are optimized using an internal CV approach, whereas the generalized model evaluation was done using an external CV approach. Since preprocessing is a crucial step in chemometric model development, preprocessing parameters can also be optimized inside the internal CV iterative loop. Other re-sampling approaches such as Jacknife, holdout and bootstrapping were also tested and compared to the two-layered CV approach. Finally, k-fold and k-replicate CVs were used to analyze the difference between the calibration and test sets and to account for reproducibility between replicate samples [34].

For selecting subsets for model calibration and validation for leave-multiple-out Crossvalidation there are several resampling methods, namely Kennard and Stone algorithm (maximin criterion) [35], Duplex (modification of Kennard Stone) [36], D-optimality criterion (maximise determinant of the information matrix) [37], K-means or Kohonen mapping (latter: used in neural networks extensively) [38]. The resampling methods determine how the original dataset is split for validation. A comparison of various resampling methods have been reviewed elsewhere [39, 40].

2.2. **Specificity** and **Sensitivity** (True Negatives and True Positives) and ROC

Classification models, yet not used for predicting process variables, give a qualitative overview of the process performance and can be used to e.g. identify different phases in a process. Spectroscopic sensor data is used to build such classification chemometric models to monitor the process performance. The performance of such models is assessed with true negative and true positive rates (TNR and TPR) using internal CV and external validation approaches [41]. True positive and true negative means that the model classifies the observation into the class where it actually belongs to. The TPR is computed as TP/(TP+FN) where TP are true positives and FN are false negatives. TNR is computed as TN/(TN+FP) where TN are true negatives (positive events wrongly categorized as negative) and FP are false positives (negative events wrongly categorized as positive). The false positive rate (FPR=FP/(FP+TN)) is the ration of the number of FP and the total number of actual negative events (regardless of classification). Plotting the TPR against the FPR results in a receiver operating characteristic (ROC) curve and illustrates the ability of the classification if the discrimination is varied. The area under the ROC curce (AUROC) is perfect if it is equal to one and only as good as a random classification if it is equal to 0.5.

Specificity and Sensitivity are measures for the proportion of true negatives respectively true positives that are correctly identified by the model. Thus, it is similar to the TNR and TPR measures. The sensitivity and specificity can be used for binary outcomes or classifications models for example in equivalence testing [42].

2.3**. Accuracy** (closeness of prediction to real value) and **precision** (random error of model predictions comparable with reproducibility of real data)

Accuracy is computed as (TP+TN)/(TP+TN+FP+FN) [43] and precision (sometimes also called positive predictive value) as TP/(TP+FP) [44]. Ideally, accuracy should be assessed by comparing the results obtained with the computational method (simulated in vitro data) with the results of an

independent set of real data (not used in training for example historical data). Then, for both data sets accuracy and precision can be computed and an acceptance criterion (computed e.g. on the historical data) can be determined to see whether the results are comparable. For analytical procedures there are different layers of precision (reproducibility, intermediate precision, analysis repeatability (replicates), systems repeatability (repeats) [45]). For computational models there are no such layers, however one should be aware of the precision which shall be reached by the model. Repeats have a narrower standard deviation than the reproducibility (e.g. long-term measurements). And not every model will aim at the same precision depending on its application, according to Box: *"all models are wrong, but some are useful. However, the approximate nature of the model must always be borne in mind...." [46].*

2.4 **R^2** (the coefficient of determination) and **RMSE** (Root Mean Squared Error)

The coefficient of determination (R^2) is in its most general definition computed by [47]:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

With SS_{res} being the sum of squares of residuals for measurements y_i and mean of observed data ($\bar{y}$):

$$SS_{res} = \sum_i (y_i - \bar{y})^2$$

And SS_{tot} being the total sum of squares of residuals:

$$SS_{tot} = \sum_i (y_i - f(y_i))^2$$

Where $f(y_i)$ is the model at point y_i. R^2 is commonly used in statistical models e.g. in response surface models generated with data performed during DoE runs to determine whether a model is adequate or not. To avoid overfitting, one should use the $R^2_{adjusted}$ which adjusts for the number of explanatory terms in the model in relation to the number of data points, as the R^2 is increasing with an increasing number of factors in the models. The $R^2_{predicted}$ is computed by using the model for predictions of data which has not been used in training the model.

The R^2 is more independent than the RMSE since it does not depend on the unit and thus it can also be used to compare models trained on different data sets. Nevertheless, the R^2 should never be looked at independently and the relation with the unexplained variance should be considered (see 3.2). There might be models which have high R^2 and $R^2_{adjusted}$ values, thus, they describe a lot of variance in the data, but the RMSE is far too high (e.g. in relation to historical data or method variability). This may indicate that some effects have been missed. On the other hand, if the RMSE is too small, overfitting might be a problem (see Section 3.2). To determine whether the RMSE is reasonable the analytical method variation can be used e.g. derived by control charts or historical data can be used.

The RMSE is computed by taking the square root of the mean square error [48]:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(f(y_i) - y_i)^2}{dof}}$$

With dof being the degrees of freedom. The RMSE is usually used to judge the performance of the trained model and to analyze the predictive power of the model e.g. with the validation dataset.

RMSE of validation should be compared to the measurement error (reference method, reproducibility error) to avoid overfitting/underfitting (see 3.2).

Alternatively to the RMSE, the mean absolute error of prediction (MAE) can be used [48]. If divided by the standard deviation of the experimental values the normalized MAE (nMAE) and normalized RMSE (nRMSE) are unbiased measurements for model predictions.

2.5. **Maximum Likelihood**

Likelihood is the probability density function that a dataset is observed given a parameter-set θ [12]:

$$\mathcal{L}(\theta\,|x) = p_\theta(x) = P_\theta(X = x)$$

Wherein X is a random variable with probability mass function p depending on θ.

The formulation of a likelihood limits its application to parametric models, however there are extensions dealing with non-parametric likelihood approaches [49]. The likelihood function is the basis for objective functions and used to derive the so-called maximum likelihood estimator [50]. Under the assumption of normal distributed error terms this directly leads to the residual sum of squares (RSS) weighted by the standard deviation of the error term [51]. The maximum likelihood estimator is a widely used objective function for bioprocess models. The obtained likelihood value for a model with a dataset can be used for model comparison. Here the Likelihood ratio test is performed (see Likelihood ratio), but also the Akaike Information Criterion (AIC) (see 2.6) can be used. Likelihood values are also computed for model diagnostics [22] and for model uncertainty (see 2.11).

2.6. **Information criteria** (Akaike information criterion(AIC), Bayesian information criterion (**BIC**))

When fitting a mechanistic or statistical model, it is possible to increase the model fit by adding parameters, which can result in overfitting. In this sense, AIC, AICc (corrected AIC used for small sample sizes) [52] and BIC are the most widely used selection criteria for modelling and identification of upstream systems to achieve the simplest model with the least variables but with greatest explanatory power. Both AIC [53] and BIC measure the trade-off of model fit [54] (quantified in terms of the log-likelihood (see 2.5)) with model complexity (a penalty for using the sample data to estimate the model parameters):

$$AIC = -2logL + 2K$$

$$BIC = -2\log L + K\log N$$

where L is the likelihood, K is the number of model parameters, and N is the number of data points used to train a model (computed on the joint training/ validation data). A model with better fit has smaller AIC or BIC, and while AIC and BIC penalize a model for having many parameters, BIC penalizes a model more severely compared to AIC [55-57]. Therefore, BIC could be more suitable in selecting a correct model while the AIC is more apt in finding the best model for predicting future observations for a given data set.

2.7. **Goodness-of-fit**

The goodness-of-fit (GOF) of a set of results of a model describes how this set of simulated results fits the observation dataset. When multiple models of a process are available, the GOF gives an assessment of relative model fit and provides information on selecting the superior model. Therefore, in the context of model validation the GOF can be used for two purposes: i) validation of the simulation results of a single model and ii) relative validation of different models' simulations.

The two most popular standard GOF statistics are Pearson's statistics (χ^2) [58],

$$\chi^2 = \sum_{c=1}^{C} \frac{(p_c - \hat{\pi}_c)^2}{\hat{\pi}_c}$$

and the likelihood ratio,

$$G^2 = 2N \sum_{c=1}^{C} p_c ln \frac{(p_c)}{\hat{\pi}_c}$$

Where c is the contingency table, π_c is the probability of the c, p_c is the observed proportion and $\hat{\pi}_c$ is the probability of the cell c under the model.

In order to use the GOF methods in the model selection studies, two most popular GOF indices, AIC and BIC can be used (section 2.6).

2.8. **Repeatability, intermediate precision**

In validation of analytical procedures quality testing is necessary to confirm that the analytical procedure is suitable for its intended use. To the same direction the validation of a computational model should confirm that the model is valid to describe the intended problem at hand [59, 60].

For the analytical validation intermediate precision and repeatability (intra-assays precision) are to be determined. Intermediate precision of analytical procedures considers several different factors such as date of test, test analyst, apparatus etc. In terms of a computational model this relates to factors such as start parameters, data (data split in CV) and so on. Repeatability should be done independently for example using CV. There, one can have homogenous samples but also

heterogeneous samples. Especially replicates within the underlying data should be handled with care (e.g. use all replicates either within test or within validation set in CV).

2.9. **Homogeneity of variance**

Homogeneity of variance (HOV) is an assumption for the validity of many parametric tests such as the t-test and ANOVA that rely on the assumption that the true population variance for each group is the same as in the observed sample. Testing for HOV is also useful to compare two data sets that do not come from the same source or that have other intrinsic differences to decide if the data sets are comparable and can be used together for model training. For instance, in bioprocesses to compare daily average substrate consumption rates and product formation efficiencies for two different sets of bioreactors which were initiated using a different inoculum. If the assumption of HOV is not met, it might be problematic to use the datasets to develop models right away. In this case, data transformation (such as log transformation) of the response variable can be helpful. The most common test to check for HOV are Levene's test [61], Bartlett's test, Brown & Forsythe's [62], Welch test and F-max test [63].

2.10. **Credibility score** and **continuous testing**

Model credibility accounts for the risk associated with the decisions made on the computational model. Based on a risk assessment quantitative and qualitative levels of credibility which need to be achieved have to be determined prior model building. This is a so called risk-informed credibility assessment framework used in the American Society of Mechanical Engineers (ASME) Norm for computational models of medical devices [64]. It goes in line with an in vitro (e.g. bench testing) and or in vivo testing (e.g. experiments) to demonstrate valid predictions of the computational model. The concept can be extended to other types of models such as statistical and machine learning models and their application on pharmaceutical and biological products as well [65, 66].

2.11. **Model uncertainty, model robustness**

There exists a parallel between uncertainty about the data and uncertainty about the model and its predictions (see Section 2.9). The typical standard errors and confidence intervals indicate uncertainty about the data, measuring how an estimate changes with sampling. However, a robust model should also account for model structure and its predictive capability in terms of nonlinear effects, and heterogeneities.

Data collection in upstream processes can be noisy, and susceptible to errors (i.e. corrupted sensors, errors in the measurement devices, etc.). Standard statistical models and data analytic techniques could fail under such scenarios, which can reduce their applicability. A robust model should handle various forms of errors as well as changes in the underlying data distribution in an automatic way, models that can be used in a reliable way, enabling their application even in complex applications such as biopharmaceutical bioprocess.

Many of the methods to characterize uncertainty in models apply equally to all types of models. However different assumptions are made and might be applicable for a certain model type and a certain dataset. In general, different approaches are utilized

Linear approximations of confidence intervals. These methods rely on the numerical estimation of a Jacobi matrix of the model with respect to the parameters (or weights for artificial neural networks (ANNs)). Uncertainties in the training data can directly be propagated by linear approximation [17, 67].

Bayesian approaches. Here the uncertainty in estimated parameters needs to be determined first (using e.g. Likelihood approaches). Model ensembles or distributions of outputs are then obtained by sampling the multi-dimensional parameter distributions using for example Markov-chain Monte Carlo methods [15, 68].

Bootstrapping: Here, the original data for model-training is bootstrapped. This results in a model ensemble that produces an according output distribution that depends directly on data uncertainty [69].

Mean Variance Estimation (MVE) Method. This method is unique to ANNs. Here the ANN is trained to learn an additional output which is the uncertainty in the prediction [17, 67].

Validation Profile Likelihood. This method is based on the Maximum Likelihood Estimator (see 2.5). Here Likelihood values of hypothetical data-points are calculated and using the X^2 distribution a confidence interval with level α can be determined [70].

2.12. Summary of validation method

In Figure 1 the model validation methods are summarized and an overview is given. Basically, four points should be considered when deciding which methods shall be used for model validation:

Nature of the dataset: Are there replicates given in the data? Is the variation in the data high (e.g. does the DoE allow to model the full design space or are only runs with the same settings available).

Sample size: Is the dataset which is used of low or high sample size (e.g. more than 10 samples)?

Model state: In which state is the model, i.e. model selection (e.g. identify whether a linear model versus a quadratic model should be used), training (e.g. perform the linear fit) or implementation (e.g. make linear fit during each campaign of commercial manufacturing)?

Model type: Is the model a statistical, a mechanistical or a hybrid model?

All these four points are interconnected. The definition of what “low” and “high” sample size numbers mean depends on the other three points. For example, for the selection of a mechanistic model higher samples sizes (e.g., $n > 10$) might be necessary than for the selection of statistical models. However, for model training mechanistic models may need lower sample sizes than

statistical models. On the other hand, if there is a high sample size given, but if all measurements are performed with the same settings, thus only replicates of a single experiment are given, this will not be helpful for model training. Thus, there is not one-fits-all approach and the definition of a single workflow is probably hard to define. Furthermore, the comparison of different models and their validation is difficult due for example the lack of gold standard data sets. Nevertheless, this summary should show that there are a plethora of different validation methods and for each model it should be critically assessed which method or combination of methods is suited best for validation.

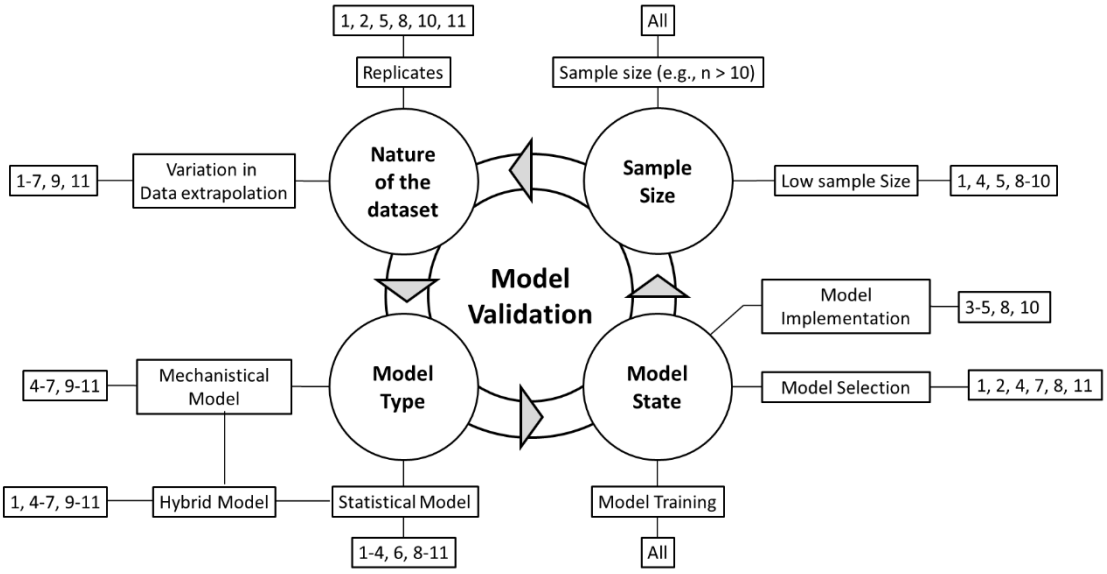

Figure 1. An overview of the different aspects to consider in order to choose a validation statistic or method. Section 2 sub-headers, namely 2.1 – 2.11 are denoted as 1 – 11.

3. Further points to consider

3.1. Calibration / model fitting

Model calibration plays a vital role in model validation and future of use of the calibrated model for prediction. The validation statistics could be interpreted only when the calibrated model and the data used for calibration are robust. Well calibrated models can be useful tools for studying the underlying mechanisms and rationalizing the results [12]. However, techniques used in model calibration also vary widely such as, subset or feature selection and data splitting approaches depending on the type of model (e.g. mechanistic vs chemometric). Therefore, clear calibration and validation protocols are not available [71]. The data needed to calibrate and validate a model is heavily reliant on the type of model in question. For example, a mechanistic model would require a much smaller dataset in comparison to an hybrid [13] or a chemometric model [72]. Therefore, validation statistics and approaches would also be dependent on the sample size of the datasets during model development. Nevertheless, use of sound validation statistics at each level of the

model calibration workflow would pave way for overall applicability and streamlining a model calibration protocol.

3.2. Overfitting and underfitting:

During model validation it is important to account for underfitting (i.e. oversimplifying e.g. by using a linear model instead of a quadratic model) but also overfitting should be avoided [38, 73]. In both scenarios the model will make error in predictions. In case of overfitting, measurement noise could be interpreted as being a process relevant effect, e.g. by fitting a quadratic function into a truly linear process. In case of underfitting, truly existing process effects might be missed. For example, if a linear fit is performed but the process is actually quadratic.

In data driven approaches (e.g. Machine Learning) this phenomenon is captured by the so-called Bias-Variance tradeoff, which is a decomposition of the Mean Squared Error into the "Bias" term that represents how well an average value is predicted and "Variance" term that increases when the model is overfitted [74]. To obtain statistical models with good generalization properties it is either common to apply early stopping in the learning phase [75] or to use regularization terms in the loss functions [76]. If the model is trained using cross validation a comparison of the R^2 values (see 2.4) for the training and test data (called Q2) can be informative to determine if overfitting occurs [77-79].

3.3. Limit of Detections and Limit of Quantification (LoD and LoQ)

In some analytical procedures there are detection and/or quantification limits. The LoD is the lowest possible distinction from noise of the analytical method [80]. The LoQ is the lowest quantitatively measurable amount with suitable accuracy and precision for the analytical procedure [59, 81]. These limits should be considered when using data to build a model as the model might not reflect the limits and thus bias results upon a certain limit. The LoD and LoQ are computed as follows [82]:

$$LoD = 3.3 * \frac{s_{noise}}{b}$$

$$LoQ = 10 * \frac{s_{noise}}{b}$$

With s_{noise} being the standard deviation of the calibration curve and b being the slope of the regression line.

3.4. Type of data

As part of model validation, the predictive accuracy of a model must be evaluated by comparing the model prediction against measured process data to ensure a model was built correctly. Certain modelling approaches such as statistical and chemometric models can require a large amount of data for their calibration and testing. Therefore, quality training and test data sets need to have some features such as enough information (variability in process inputs), sufficient number of experiments and observations and in the case of on-line data, a low signal to noise ratio is critical to

obtain reliable models. Further aspects to consider are the sampling timing points and the process intrinsic variability. On the other hand, if a large number of measurements is given, but they show only small variations in the settings (e.g. only replicates are given) then a model is hard to build [83].

Further, discrete data should be treated differently than continuous data. In general, it is recommended to use always the unrounded raw values when developing and validating a model.

Biopharmaceutical upstream manufacturing largely depends on batch processes, with data sets containing time dependent information with a typical three-dimensional shape of batches x variables x time. Unfolding procedures are used to reduce this three-dimension matrix to a two-dimensional format which is necessary for multivariate data analysis. The data set can be unfolded in different ways depending on the purpose of the analysis. Batch-wise unfolding, where each row in the matrix is a different batch, is used to analyze differences among batches by removing the dynamic behavior of the batch. Conversely, variable-wise unfolding is used to study the dynamic behavior of the batch relative to the mean of each variable [84, 85].

Moreover, in biopharmaceutical upstream processes there are different kinds of parameters such as input parameters which are highly controlled in a quite narrow range and others which are controlled only indirectly within a broad range. Some parameter settings cannot be tested in real experiments as they would lead to an edge-of-failure. For some parameter online measures give a high time resolution and thus a large set of data for other parameters only offline measures at certain time steps are available. This has to be taken into account during data cleaning and polishing and should be an important part during the whole model lifecycle to ensure that the data fits the problem and is suited for the model.

Intensified DoEs (iDoEs) have been recently used in upstream processing to develop hybrid models in an efficient manner with a small amount of data needed [14, 29, 86]. Independent of the number of samples measured, the nature of the data used for model generation plays also a role in model validation as described above. Furthermore, if there are replicates within a dataset they should be studied in detail. In CV it should be avoided to have replicates within the test and training set to not artificially improve validation measures such as RMSE.

3.5. State of model

Typical model states in bioprocesses are model development for generating base process knowledge, using developed model for process monitoring, prediction and optimization activities and, ultimately for continuous process improvement [87]. The amount of data needed during model development stages would be much higher than during implementation and maintenance stages. Maintenance of the model necessitates any improvements and changes in the model to be continuously assessed along the model lifecycle, namely through development, implementation and further maintenance. Based on the available data which are relevant for the model state, validation statistics would vary vastly.

3.6. Good Modeling Practice (GMoP)

Ideally, during model validation several methods are combined to account for over- and underfitting but also to assess different aspects of the model. For example, there might be models which have a very small RMSE but model predictability is rather poor due to effects which are not considered (e.g. equipment variability). Therefore, it is essential to adhere to a good modelling practice. Consequently, this will address all the aforementioned points to consider into context. Typically, it would start with a clear definition of an objective and necessary requirements for a specific model type. Based on the model nature, different assumptions are discussed, and model calibration is done. Thereon, the sensitivity analysis and an estimation of parameter uncertainty is done. Adhering to GMoP, would ensure a robust model capable of simulating and predicting outcomes [88].

4. Recommendations from health authorities

According to the ICH guidelines Q8, Q9 and Q10, model validation is an essential part of model development and implementation, and verification of such models must be carried out throughout the lifecycle of the product [89]. Furthermore, models should be categorized based on high, medium or low impact models and validation extent for such models must be considered based on their level of impact on the process. The following points must be considered for high impact models, namely setting acceptance criteria, comparison of accuracy using internal cross validation, validation of the model using external cross validation and verify prediction accuracy by parallel testing with the reference method throughout the lifecycle. Validation procedures should take into account any change in material attributes or analytical procedures and differences arising from scales.

In a PAT framework, validation procedures should consider analytical method validation and continuous quality assurance [45]. Statistical methods such as ANOVA to assess regression analysis (e.g. in a chemometric model), R (correlation coefficient) and R^2 or linear regression for linearity can be used to assess validation characteristics [81]. Computational and simulation models should be described to assess its validity and their prediction capability of the model outputs with validated analytical methods [90]. The sensitivity of the model outputs on the key model parameters must be described with a systematic analysis of the uncertainty. Although different validation procedures are presented in academic research, only a few are mentioned in the regulatory documents. Intesified use of modelling approaches to accelerate product approval stages would enable addition of robust validation approaches in regulatory documents.

Model implementation in production processes necessitates validation of the software, when the model is a part of the production process or the quality system [91].

5. Concluding Remarks

In this review, validation methods currently used in bioprocess modelling are described. Motivated by QbD and the necessity of speed-to-clinic approaches makes modeling more and more important in biopharmaceutical industry. Beside the fact that extensive experimental set-ups are expensive

and time consuming, modeling approaches often allow a better process understanding, optimization and control. However, the reliability of the given model needs to be ensured before answering the questions at hand (e.g. optimization of titer or the definition of KPPs and CPPs). As there are many different model types (statistical, mechanistical, hybrid, and so on) a unique way for model validation is difficult. This is also due to the fact that there is no clear protocol for model generation in bioprocessing. Even health authorities are referring to some basic approaches like R and R^2.

There exists a plethora of methods for model validation and in consequence to the points mentioned above, using a combination of the methods mentioned above is recommended. This shall better allow to judge the reliability and predictability of the model. Furthermore, it helps to account for under/overfitting and to address the type of the model even throughout the modeling lifecycle like model calibration, validation and implementation. In any case, data is needed to set-up a model and thus, the data should be checked very carefully before model generation and validation. For example, the sample size, the variation within the data but also the number of replicates are to be considered.

Ideally, beside using model validation *in silico* methods, models should undergo a lifecycle which includes continuously testing during a product lifecycle which is also recommended from health authorities [87]. Here also the purpose of the model shall be considered e.g. using a model for having a better process understanding in early stage development, for defining CPPs in late stage development or for process control during manufacturing. If the model has a high impact on the process, the model validation should be more detailed then for models with lower impact. We expect that the validation methods in bioprocess development will become more important as the models itself are getting widely accepted. Using all available and suitable modelling validation approaches will help to judge the models about their predictability and reliability. Reliable models will help to understand bioprocesses in their entire lifecycle (lab to market) and thereby, making them more robust and flexible.

Author Contributions: Conceptualization and writing (original draft preparation): V.R. and B.K. with important contributions from H.B., L.M., A.E. and revisions at multiple stages; writing - review and editing: all authors; visualization: V.R. All authors have read and agreed to the published version of the manuscript."

Funding: "This research received no external funding".

Acknowledgments: We would like to thank Christina Yassouridis, Martina Berger, Albert Paul, Ogsen Gabrielyan, Joachim Baer, Erich Bluhmki and Jens Smiatek for valuable discussions and useful hints.

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

References

1. Sommeregger, W., et al., Quality by control: Towards model predictive control of mammalian cell culture bioprocesses. Biotechnol J, 2017. **12**(7).

2. U.S. Department of Health and Human Services Food and Drug Administration, *PAT — A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance*, U.S.D.o.H.a.H.S.F.a.D. Administration, Editor. 2004: Silver Spring, MD, USA.

3. ICH, ICH guideline Q8 (R2) on pharmaceutical development, in EMA/CHMP/ICH/167068/2004, ICH, Editor. 2009: United Kingdom.

4. Bhatia, H., et al., In-line monitoring of amino acids in mammalian cell cultures using raman spectroscopy and multivariate chemometrics models. Engineering in life sciences, 2017. **18**(1): p. 55-61.

5. Rafferty, C., et al., Analysis of chemometric models applied to Raman spectroscopy for monitoring key metabolites of cell culture. Biotechnology Progress, 2020. **36**(4): p. e2977.

6. Sivakesava, S., J. Irudayaraj, and D. Ali, Simultaneous determination of multiple components in lactic acid fermentation using FT-MIR, NIR, and FT-Raman spectroscopic techniques. Process Biochemistry, 2001. **37**(4): p. 371-378.

7. Ramırez, J., H. Gutierrez, and A. Gschaedler, Optimization of astaxanthin production by Phaffia rhodozyma through factorial design and response surface methodology. Journal of Biotechnology, 2001. **88**(3): p. 259-268.

8. Torkashvand, F., et al., Designed Amino Acid Feed in Improvement of Production and Quality Targets of a Therapeutic Monoclonal Antibody. PLOS ONE, 2015. **10**(10): p. e0140597.

9. Möller, J., et al., Model-assisted Design of Experiments as a concept for knowledge-based bioprocess development. Bioprocess and Biosystems Engineering, 2019. **42**(5): p. 867-882.

10. Ehsani, A., et al., Towards Model-Based Optimization for Quality by Design in Biotherapeutics Production, in Computer Aided Chemical Engineering, A.A. Kiss, et al., Editors. 2019, Elsevier. p. 25-30.

11. Schmidberger, T., et al., Progress toward forecasting product quality and quantity of mammalian cell culture processes by performance-based modeling. Biotechnology Progress, 2015. **31**(4): p. 1119-1127.

12. Smiatek, J., A. Jung, and E. Bluhmki, Towards a Digital Bioprocess Replica: Computational Approaches in Biopharmaceutical Development and Manufacturing. Trends in Biotechnology, 2020. **38**(10): p. 1141-1153.

13. Narayanan, H., et al., A new generation of predictive models: The added value of hybrid models for manufacturing processes of therapeutic proteins. Biotechnology and Bioengineering, 2019. **116**(10): p. 2540-2549.

14. Bayer, B., G. Striedner, and M. Duerkop, Hybrid Modeling and Intensified DoE: An Approach to Accelerate Upstream Process Characterization. Biotechnol J, 2020: p. e2000121.

15. Möller, J., et al., Model uncertainty-based evaluation of process strategies during scale-up of biopharmaceutical processes. Computers & Chemical Engineering, 2020. **134**: p. 106693.

16. Ulonska, S., et al., Workflow for Target-Oriented Parametrization of an Enhanced Mechanistic Cell Culture Model. Biotechnology Journal, 2018. **13**(4): p. 1700395.

17. Anane, E., et al., Output uncertainty of dynamic growth models: Effect of uncertain parameter estimates on model reliability. Biochemical Engineering Journal, 2019. **150**: p. 107247.

18. Wang, Y.-H., et al., Optimization of medium composition for the production of clavulanic acid by Streptomyces clavuligerus. Process Biochemistry, 2005. **40**(3): p. 1161-1166.

19. Fricke, J., et al., Designing a fully automated multi-bioreactor plant for fast DoE optimization of pharmaceutical protein production. Biotechnology Journal, 2013. **8**(6): p. 738-747.

20. Adinarayana, K., et al., Response surface methodological approach to optimize the nutritional parameters for neomycin production by Streptomyces marinensis under solid-state fermentation. Process Biochemistry, 2003. **38**(11): p. 1565-1572.

21. Koutinas, M., et al., BIOPROCESS SYSTEMS ENGINEERING: TRANSFERRING TRADITIONAL PROCESS ENGINEERING PRINCIPLES TO INDUSTRIAL BIOTECHNOLOGY. Computational and Structural Biotechnology Journal, 2012. **3**(4): p. e201210022.

22. Kroll, P., et al., Workflow to set up substantial target-oriented mechanistic process models in bioprocess engineering. Process Biochemistry, 2017. **62**: p. 24-36.

23. Brüning, S., et al., Modeling Suspension Cultures of Microbial and Mammalian Cells with an Adaptable Six-Compartment Model. Chemical Engineering & Technology, 2017. **40**(5): p. 956-966.

24. Popp, O., et al., A hybrid approach identifies metabolic signatures of high-producers for chinese hamster ovary clone selection and process optimization. Biotechnology and Bioengineering, 2016. **113**(9): p. 2005-2019.

25. Quek, L.-E., et al., *Metabolic flux analysis in mammalian cell culture.* Metabolic Engineering, 2010. **12**(2): p. 161-171.

26. Oliveira, R., Combining first principles modelling and artificial neural networks: a general framework. Computers & Chemical Engineering, 2004. **28**(5): p. 755-766.

27. Teixeira, A.P., et al., Hybrid semi-parametric mathematical systems: bridging the gap between systems biology and process engineering. (0168-1656 (Print)).

28. Simutis, R. and A. Lubbert, *Hybrid Approach to State Estimation for Bioprocess Control.* Bioengineering (Basel), 2017. **4**(1).

29. von Stosch, M., J.-M. Hamelink, and R. Oliveira, Toward intensifying design of experiments in upstream bioprocess development: An industrial Escherichia coli feasibility study. Biotechnology Progress, 2016. **32**(5): p. 1343-1352.

30. Von Stosch, M., et al., Hybrid semi-parametric modeling in process systems engineering: Past, present and future. Computers & Chemical Engineering, 2014. **60**: p. 86-101.

31. Brereton, R.G., et al., *Chemometrics in analytical chemistry—part II: modeling, validation, and applications.* Analytical and Bioanalytical Chemistry, 2018. **410**(26): p. 6691-6704.

32. Westad, F. and F. Marini, *Validation of chemometric models – A tutorial.* Analytica Chimica Acta, 2015. **893**: p. 14-24.

33. Xu, Q.-S. and Y.-Z. Liang, *Monte Carlo cross validation.* Chemometrics and Intelligent Laboratory Systems, 2001. **56**(1): p. 1-11.

34. Guo, S., Chemometrics and Statistical Analysis in Raman Spectroscopy-based Biological Investigations, in Chemisch-Geowissenchaftlichen Fakultät 2018, Friedrich Schiller University Jena Jena p. 182.

35. Morais, C.L.M., et al., Improving data splitting for classification applications in spectrochemical analyses employing a random-mutation Kennard-Stone algorithm approach. Bioinformatics, 2019. **35**(24): p. 5257-5263.

36. Snee, R.D., *Validation of Regression Models: Methods and Examples.* Technometrics, 1977. **19**(4): p. 415-428.

37. St. John, R.C. and N.R. Draper, *D-Optimality for Regression Designs: A Review.* Technometrics, 1975. **17**(1): p. 15-23.

38. Charaniya, S., W.-S. Hu, and G. Karypis, *Mining bioprocess data: opportunities and challenges.* Trends in Biotechnology, 2008. **26**(12): p. 690-699.

39. Molinaro, A.M., R.M. Simon R Fau - Pfeiffer, and R.M. Pfeiffer, *Prediction error estimation: a comparison of resampling methods.* (1367-4803 (Print)).

40. Kakumoto, K. and Y. Tochizawa, Comparison of Resampling Methods for Bias-Reduced Estimation of Prediction Error: A Simulation Study Based on Real Datasets from Biomarker Discovery Studies. Japanese Journal of Biometrics, 2017. **38**(1): p. 17-39.

41. Guo, S., et al., *Common mistakes in cross-validating classification models.* Analytical Methods, 2017. **9**(30): p. 4410-4417.

42. Quiroz, J. and R.K. Burdick, *Assessing Equivalence of Two Assays Using Sensitivity and Specificity.* Journal of Biopharmaceutical Statistics, 2007. **17**(3): p. 433-443.

43. Metz, C.E., *Basic principles of ROC analysis.* Seminars in Nuclear Medicine, 1978. **8**(4): p. 283-298.

44. Olson, D.L. and D. Delen, *Advanced Data Mining Techniques*. 2008: Springer Publishing Company, Incorporated.

45. FDA, Guidance for Industry, Q2B Validation of Analytical Procedures: Methodology, in FDA-1996-D-0169, C.f.B.E.a.R. Center for Drug Evaluation and Research, Editor. 1997.

46. Box, G.E.P. and N.R. Draper, *Empirical model-building and response surfaces*. Empirical model-building and response surfaces. 1987, Oxford, England: John Wiley & Sons. xiv, 669-xiv, 669.

47. Sachs, L. and J. Hedderich, *Angewandte Statistik. Methodensammlung mit R*. 12., vollst. neu bearb. Aufl. ed. 2006, Berlin: Springer.

48. Hyndman, R.J. and A.B. Koehler, *Another look at measures of forecast accuracy.* International Journal of Forecasting, 2006. **22**(4): p. 679-688.

49. Laird, N., *Nonparametric Maximum Likelihood Estimation of a Mixing Distribution.* Journal of the American Statistical Association, 1978. **73**(364): p. 805-811.

50. Hanomolo, A., et al., Maximum likelihood parameter estimation of a hybrid neural-classical structure for the simulation of bioprocesses. Mathematics and Computers in Simulation, 2000. **51**(3): p. 375-385.

51. Raue, A., et al., Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. Bioinformatics, 2009. **25**(15): p. 1923-1929.

52. Hurvich, C.M. and C.-L. Tsai, *Regression and time series model selection in small samples.* Biometrika, 1989. **76**(2): p. 297-307.

53. Akaike, H., *A new look at the statistical model identification.* IEEE Transactions on Automatic Control, 1974. **19**(6): p. 716-723.

54. Gideon, S., *Estimating the Dimension of a Model.* The Annals of Statistics, 1978. **6**(2): p. 461-464.

55. Deppe, S., et al., *Estimation of Process Model Parameters*, in *Animal Cell Biotechnology: Methods and Protocols*, R. Pörtner, Editor. 2020, Springer US: New York, NY. p. 213-234.

56. Li, B., J. Morris, and E.B. Martin, *Model selection for partial least squares regression.* Chemometrics and Intelligent Laboratory Systems, 2002. **64**(1): p. 79-89.

57. Cavanaugh, J.E. and A.A. Neath, The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. WIREs Computational Statistics, 2019. **11**(3): p. e1460.

58. Maydeu-Olivares, A. and C. García-Forero, *Goodness-of-Fit Testing*, in *International Encyclopedia of Education (Third Edition)*, P. Peterson, E. Baker, and B. McGaw, Editors. 2010, Elsevier: Oxford. p. 190-196.

59. ICH, ICH Q2 (R1) Validation of analytical procedures: text and methodology. 1997.

60. Kojima, S., Evaluation of intermediate precision in the validation of analytical procedures for drugs: from NDA Dossiers, . Pharm. Tech. Japan 2002. **18**: p. 695-704.

61. Fan, Y., et al., A multi-pronged investigation into the effect of glucose starvation and culture duration on fed-batch CHO cell culture. Biotechnology and Bioengineering, 2015. **112**(10): p. 2172-2184.

62. Sangari, R., Establish methodology for estimating process performance capability during the design phase for biopharmaceutical processes, in Sloan School of Management.; Massachusetts Institute of Technology. Institute for Data, Systems, and Society.; Massachusetts Institute of Technology. Engineering Systems Division.; Leaders for Global Operations Program. 2016, Massachusetts Institute of Technology: Massachusetts.

63. Lee, K.-M. and D.F. Gilmore, *Statistical experimental design for bioprocess modeling and optimization analysis.* Applied Biochemistry and Biotechnology, 2006. **135**(2): p. 101-115.

64. Assessing Credibility of Computational Modeling through Verification and Validation: Application to Medical Devices V&V 40. 2018. 60.

65. Viceconti, M., et al., In silico trials: Verification, validation and uncertainty quantification of predictive models used in the regulatory evaluation of biomedical products. Methods, 2021. **185**: p. 120-127.

66. Bideault, G., et al., Verification and Validation of Computational Models Used in Biopharmaceutical Manufacturing: Potential Application of the ASME Verification and Validation 40 Standard and FDA Proposed AI/ML Model Life Cycle Management Framework. Journal of Pharmaceutical Sciences, 2021.

67. Khosravi, A., et al., *Comprehensive Review of Neural Network-Based Prediction Intervals and New Advances.* IEEE Transactions on Neural Networks, 2011. **22**(9): p. 1341-1356.

68. Liu, Y. and R. Gunawan, *Bioprocess optimization under uncertainty using ensemble modeling.* Journal of Biotechnology, 2017. **244**: p. 34-44.

69. Pinto, J., et al., A bootstrap-aggregated hybrid semi-parametric modeling framework for bioprocess development. Bioprocess and Biosystems Engineering, 2019. **42**(11): p. 1853-1865.

70. Kreutz, C., A. Raue, and J. Timmer, Likelihood based observability analysis and confidence intervals for predictions of dynamic models. BMC Systems Biology, 2012. **6**(1): p. 120.

71. Saleh, D., et al., Straightforward method for calibration of mechanistic cation exchange chromatography models for industrial applications. Biotechnology Progress, 2020. **36**(4): p. e2984.

72. Kastenhofer, J., et al., Monitoring E. coli Cell Integrity by ATR-FTIR Spectroscopy and Chemometrics: Opportunities and Caveats. Processes, 2021. **9**(3).

73. Walsh, I., et al., DOME: Recommendations for supervised machine learning validation in biology. arXiv e-prints, 2020: p. arXiv:2006.16189.

74. Neal, B., et al., *A Modern Take on the Bias-Variance Tradeoff in Neural Networks.* arXiv e-prints, 2018: p. arXiv:1810.08591.

75. Kashani, M.N. and S. Shahhosseini, *A methodology for modeling batch reactors using generalized dynamic neural networks.* Chemical Engineering Journal, 2010. **159**(1): p. 195-202.

76. Fellner, M., A. Delgado, and T. Becker, *Functional nodes in dynamic neural networks for bioprocess modelling.* Bioprocess and Biosystems Engineering, 2003. **25**(5): p. 263-270.

77. Roussouly, N., F. Petitjean, and M. Salaun, *A new adaptive response surface method for reliability analysis.* Probabilistic Engineering Mechanics, 2013. **32**: p. 103-115.

78. Riley, R.D., et al., Penalization and shrinkage methods produced unreliable clinical prediction models especially when sample size was small. Journal of Clinical Epidemiology, 2021. **132**: p. 88-96.

79. Van Calster, B., et al., Regression shrinkage methods for clinical prediction models do not guarantee improved performance: Simulation study. Statistical Methods in Medical Research, 2020. **29**(11): p. 3166-3178.

80. Marson, B.M., et al., VALIDATION OF ANALYTICAL METHODS IN A PHARMACEUTICAL QUALITY SYSTEM: AN OVERVIEW FOCUSED ON HPLC METHODS. Química Nova, 2020. **43**: p. 1190-1203.

81. FDA, *Analytical Procedures and Methods Validation for Drugs and Biologics*, C.f.B.E.a.R. Center for Drug Evaluation and Research, Editor. 2015.

82. Thompson, M., S.L.R. Ellison, and R. Wood, Harmonized guidelines for single-laboratory validation of methods of analysis (IUPAC Technical Report). Pure and Applied Chemistry, 2002. **74**(5): p. 835-855.

83. Solle, D., et al., Between the Poles of Data-Driven and Mechanistic Modeling for Process Operation. Chemie Ingenieur Technik, 2017. **89**(5): p. 542-561.

84. Zhang, Y. and T. Edgar, Bio‐Reactor Monitoring with Multiway PCA and Model Based PCA. Omnipress, 2006

85. Borchert, D., et al., *Comparison of data science workflows for root cause analysis of bioprocesses.* Bioprocess and Biosystems Engineering, 2019. **42**(2): p. 245-256.

86. von Stosch, M. and M.J. Willis, *Intensified design of experiments for upstream bioreactors.* Engineering in Life Sciences, 2017. **17**(11): p. 1173-1184.

87. Kroll, P., et al., Model-Based Methods in the Biopharmaceutical Process Lifecycle. Pharm Res, 2017. **34**(12): p. 2596-2613.

88. Rischawy, F., et al., Good modeling practice for industrial chromatography: Mechanistic modeling of ion exchange chromatography of a bispecific antibody. Computers & Chemical Engineering, 2019. **130**: p. 106532.

89. FDA, Q8, Q9, & Q10 Questions and Answers -- Appendix: Q&As from Training Sessions (Q8, Q9, & Q10 Points to Consider), C.f.B.E.a.R. Center for Drug Evaluation and Research, Editor. 2012.

90. FDA, *Reporting of Computational Modeling Studies in Medical Device Submissions*, C.f.B.E.a.R. Center for Drug Evaluation and Research, Editor. 2016.

91. FDA, General Principles of Software Validation; Final Guidance for Industry and FDA Staff, C.f.B.E.a.R. Center for Devices and Radiological Health, Editor. 2002.