

Dynamic parameter estimation and prediction over  
consecutive scales, based on moving horizon estimation -  
applied to an industrial cell culture seed train

Tanja Hernández Rodríguez<sup>1</sup>, Christoph Posch<sup>2</sup>,  
Ralf Pörtner<sup>3</sup>, Björn Frahm<sup>1\*</sup>

<sup>1</sup>Biotechnology & Bioprocess Engineering, Department of Life Sciences Technologies,  
Ostwestfalen-Lippe University of Applied Sciences and Arts, Lemgo, Germany

<sup>2</sup>Novartis Technical Research and Development, Sandoz GmbH, Schafftenau, Austria

<sup>3</sup>Institute for Bioprocess- and Biosystems Engineering, Hamburg University of Technology, Germany

\*Corresponding author

**E-mail:** bjoern.frahm@th-owl.de

---

## Abstract

Bioprocess modeling has become a useful tool for prediction of the process future with the aim to deduce operating decisions (e.g. transfer or feeds). Due to variabilities, which often occur between and within batches, updating (re-estimation) of model parameters is required at certain time intervals (dynamic parameter estimation) to obtain reliable predictions. This can be challenging in the presence of low sampling frequencies (e.g. every 24 hours), different consecutive scales and large measurement errors, as in the case of cell culture seed trains. In this contribution, two estimation techniques, which differ in terms of their objective functions, were investigated regarding robustness concerning the aforementioned challenges and the required amount of experimental data (estimation horizon). A common weighted least squares estimation (WLSE) and a moving horizon estimation (MHE), which takes prior knowledge about model parameters into account, were applied for re-estimation of model parameters over three consecutive cultivation scales (40 - 2,160 L) of an industrial cell culture seed train. It is shown how the proposed MHE can deal with the aforementioned difficulties by integration of prior knowledge, even if only data at two sampling points are available, outperforming the classical WLSE approach. A workflow illustrating the required steps is presented.

**Keywords:** Dynamic parameter estimation; bioprocess, cell cultures; moving horizon estimation; prior knowledge ■

---

# 1 Introduction

Mathematical models are playing an important role for simulation and prediction of bioprocesses. Model-based methods are applied in the context of model-assisted Design of Experiments, the design, layout and optimization of production processes (Pörtner et al., 2017, Möller et al., 2019a, Kuchemüller et al., 2020, Kroll et al., 2017, Xing et al., 2010, Narayanan et al., 2019b,a, Schenkendorf et al., 2020). Furthermore, they are used as predictive models, enabling prediction of the process future, featuring the development of decision making, optimization and control strategies (Frahm et al., 2002, Pörtner et al., 2017, Hass et al., 2001, Möller et al., 2019b). The performance mainly depends on the prediction accuracy of the model, which in turn depends on data quality (Sommeregger et al., 2017), the complexity of the model and the ability to address batch-to-batch variabilities (Xie and Schenkendorf, 2019) concerning cell growth, viability, uptake and production rate, a challenging task, especially in the case of mammalian cell cultures. These are the most frequent host for many biopharmaceuticals (e.g. antibodies & proteins for diagnostic and therapeutic purposes) (Walsh, 2018).

An iterative model updating procedure, meaning that model parameters are updated (re-estimated) after certain time steps, is required to take possible variabilities into account. Figure 1 illustrates the role of this model updating procedure within the production process steps. Data from the realized ongoing process is used in order to generate a digital twin of the process, that means a virtual representation of the process, which provides prediction of the remaining future. This prediction is then used to deduce operating decisions, e.g. concerning cell passaging (Frahm, 2007, Hernández Rodríguez and Frahm, 2020, Kern et al., 2016) or to control the process, e.g. through an Open-Loop-Optimal-Control (OLFO) method for the control of optimal feeding (Hass et al., 2001, Mears et al., 2017) or to control pH and temperature shifts (Paul et al., 2019). As soon as new process data becomes available model updating, meaning re-estimation of model parameters over a growing estimation horizon (window), is performed (= dynamic parameter estimation).

In literature, different approaches for re-estimation of model parameters over a growing data base can be found (Haseltine and Rawlings, 2005, Hedengren et al., 2014, Hedengren and Eaton, 2017, Jewaratnam et al., 2012). They differ in terms of the objective function, which has to be optimized, and in terms of the considered estimation horizon (number of data points used for parameter estimation). This can be growing or moving with a fixed horizon length. In this study, two estimation techniques are applied and their robustness and impact on prediction performance are investigated, a classical weighted least squares estimation (WLSE) approach and a moving horizon estimation (MHE) technique. While the objective function of the WLSE only considers information from data points within the estimation horizon, the objective function of the MHE approach includes a term containing prior knowledge concerning the parameter values and a penalty term for the change in parameter values (Rao et al., 2001, Haseltine and Rawlings, 2005, Hedengren and Eaton, 2017).

While typical applications found in literature cover cultivations in one scale (Jewaratnam et al., 2012, Haseltine and Rawlings, 2005), no recommendations can be found for re-estimation and prediction of cultivations over several consecutive scales as in the case of cell culture seed trains. However, these are required for the production of biopharmaceuticals in suspension culture to increase the cell number from cell thawing up to production scale (Frahm, 2007, Kern et al., 2016). Cells are cultivated in cultivation systems of different volumes, starting with small volumes (typically with shake flasks) and they are passaged every 3-4 days into the next bigger one (using bioreactor scales in larger scales). It has been shown that cultivation conditions during the seed train have a significant impact on cell performance in production scale (Le et al., 2012). Consequently, simplifying or enabling better decision making for seed train operations is required. Moreover, monitoring, control and development of optimization strategies play an important role.

Nevertheless, predictive modeling in case of only few available data points are rarely considered in literature. For various fields, such as many microbial and yeast processes, a relatively high density of data point is available, typically from on-line measurements and in combi-

nation with soft sensors, whereas cell culture processes are usually characterized by a low data density and samples are often taken only once or twice a day. In this contribution it is investigated how to integrate current knowledge into dynamic optimization in order to address possible variabilities, especially focusing on the impact of the estimation horizon (data points used for parameter estimation) and the objective function on model parameters and prediction performance. It is shown how changes in cell growth behaviour can be detected through changes in model parameter values after updating model parameters. A workflow, containing the recommended steps was developed and applied to an industrial CHO cell culture seed train.

## 2 Materials and methods

In this study different scenarios concerning estimation horizon and updating strategy are considered. Without loss of generality, they are explained for application to cultivation data over several consecutive scales (a seed train) and assuming that data from several consecutive seed trains are collected. Furthermore, two different estimation techniques having different objective functions (weighted least squares estimation (WLSE) and moving horizon estimation (MHE)), were applied.

### 2.1 Estimation and prediction horizon

Re-estimation of model parameters (or model updating) is performed at different points in time during the process. Therefore, the time span  $t_1, \dots, t_n$  of the process (here explained for a seed train) is divided at the current point in time  $t_i$  into an estimation horizon of length  $N$ ,  $t_{i-N}, \dots, t_i$ , containing the  $N$  most recent data points, and the prediction horizon,  $t_{i+1}, \dots, t_n$ . During prediction of one seed train, the right bound (the end) of the prediction horizon is kept fixed while the left bound (the beginning) of the prediction horizon moves forward with new measurements (receding prediction horizon).

At the same time, the right bound of the estimation horizon moves forward (*moving estimation horizon*). For the left bound of the estimation horizon different scenarios are applied (see Figure 2 A): i) Fixed point in time at the beginning of the current seed train or at the beginning of each cultivation scale (described in this work as *growing estimation horizon*), ii) moving left bound, meaning that as soon as a new data point becomes available, the oldest data point is excluded from the estimation horizon (described in this work as *moving estimation horizon of fixed length*).

## 2.2 Objective functions

The two applied objective functions (estimation techniques) in this study are presented below. In order to avoid confusion concerning the terminologies containing 'moving horizon' it should be noted, that the '*moving estimation horizon*' as described above refers to the data used directly within parameter estimation, whereby '*moving horizon estimation*' refers to an estimation technique characterized through a specific objective function, which will be explained below.

**Weighted Least Squares Estimation (WLSE)** *Weighted non-linear least square methods* are widely applied estimation methods for static optimization problems (Hedengren et al., 2014, Deppe et al., 2020). The aim is to minimize the squared deviation between  $n$  measured and simulated data points for  $k$  variables,  $y_{m,j} = (y_{m,j1}, \dots, y_{m,jn})$  and  $y = (y_{j1}, \dots, y_{jn})$  respectively for a variable  $y_j$ , multiplied by a weighting factor  $w_{m,ij}$  for  $i = 1, \dots, n$  and  $j = 1, \dots, k$  (e.g. to address measurement deviations and to compare quantities of different dimensions or giving more weights to specific quantities) for a fixed period of time  $t = (t_0, \dots, t_n)$ . Here, the weighting factor for variable  $y_j$  was determined through division by the maximum experimental value  $y_{j,max}$  over time span  $t$ , thus  $w_{m,ij} = \frac{1}{y_{j,max}}$ . This can be formulated as an static optimization problem, consisting in minimizing the following

objective function  $J$  over the space of model parameters  $p = (p_1, \dots, p_k)$ :

$$J(p) = \sum_{j=1}^k \sum_{i=1}^n w_{m,ij} \cdot (y_{m,ij} - y_{ij}(p))^2 \quad (1)$$

This notation is similar to  $J = (y_m - y)^T W_m (y_m - y)$ , also found in literature. This estimation method is simple to apply but suffers from sensitivities to outliers and high measurement deviations, when applied during dynamic optimization (Hedengren et al., 2014).

**Moving horizon estimation (MHE)** Now, if more and more data are collected (e.g. from several seed trains) and past and present data are used for parameter estimation (dynamic parameter estimation), the computational cost of a simple least square estimation increases. As stated in Haseltine and Rawlings (2005), an adequate approach to prevent computational limitations is to formulate the parameter estimation over a fixed size estimation horizon and to include information obtain from past data through a so called *arrival cost term*. This method, presented in (Rao et al., 2001), is known as a moving horizon estimation (MHE) and is divided into two main parts, state estimation and arrival cost estimation. A fixed estimation horizon is defined containing measurements used for the state estimation. Every time new measurements are supplied, old measurements are discarded and passed from the estimation horizon to the so called *arrival cost window* (or historical window), whenever they exceed a fix estimation horizon length. Information from the past (up to the estimation horizon) is now summarized within the arrival cost term and through a penalty term for the change in parameters. Figure 2 B illustrates how cultivation data from various seed trains can be divided iteratively into an arrival cost window, an estimation horizon and a prediction horizon.

Different approaches to estimate the arrival cost can be found in literature (Haseltine and Rawlings, 2005, Hedengren et al., 2014, Hedengren and Eaton, 2017, Ungarala, 2009). In this contribution, a the *squared-error moving horizon estimation (MHE)* was applied. The aim is to add penalty terms to the weighted squared errors (see eq. 1), taking into account

learned information from historical batches (cultivations) of a fixed arrival cost window size (e.g. one preceding seed train). One term contains the prior values  $\hat{y}_j = (\hat{y}_{j1}, \dots, \hat{y}_{jn})$  resulting from simulations (model values), based on the prior parameter vector  $\hat{p}$ , obtained after the previous estimation cycle. Then the deviation between the current model values  $y_j = (y_{j1}, \dots, y_{jn})$  and the prior model values  $\hat{y}_j = (\hat{y}_{j1}, \dots, \hat{y}_{jn})$  is computed and multiplied by a weighting factor  $w_{p,ij}$  for  $i = 1, \dots, n$  and  $j = 1, \dots, k$  for regulation of the prior's impact. This term is also called 'forgetting term' or 'forgetting penalty'. Furthermore, a penalty term  $\Delta p^T c_{\Delta p}$  for the change in parameters is added, where  $\Delta p^T$  describes the euclidean distance between the current (estimated)  $p$  and the prior parameter vector  $p_0$  (calculated by the euclidean norm of the differences) and  $c_{\Delta p}$  is a constant value to regulate the impact of the prior information. More details on this regularization technique can be found in Anane et al. (2019).

The resulting dynamic optimization problem is defined by minimizing the following objective function over the set of model parameters: (Hedengren et al., 2014):

$$J(p) = \sum_{j=1}^k \sum_{i=1}^n w_{m,ij} \cdot (y_{m,ij} - y_{ij}(p))^2 + w_{p,ij} \cdot (y_{m,ij} - \hat{y}_{ij})^2 + \Delta p^T c_{\Delta p}. \quad (2)$$

with  $\Delta p = \frac{\gamma}{n} \|p_0^s - p^s\|_2 = \frac{\gamma}{n} \sqrt{\sum_{i=1}^l (p_{0,i}^s - p_i^s)^2}$  for scale parameter vectors  $p_0^s$  and  $p^s$  of length  $l$  and constant tuning parameter  $\gamma$ . It should be noted that the squared-error MHE technique can also be found without the 'forgetting term' in equation 2 (Hedengren and Eaton, 2017).

### 2.3 Investigated suspension cell culture process

In this contribution, subject of investigation is an industrial CHO cell culture process containing a seed train comprising five shake flask scales and three bioreactor scales as well as the production scale, whereby the focus lies on the the bioreactor part of the seed train, which is composed by bioreactor 1 (N-3, 40 L), bioreactor 2 (N-2, 320 L) and bioreactor 3 (N-1, 2,160 L). From experimental data (offline measurements), taken once a day, time profiles for

viable cell density  $X_v$ , viability  $Via$ , concentrations of glucose  $c_{Glc}$ , glutamine  $c_{Gln}$ , lactate  $c_{Lac}$  and ammonia  $c_{Amm}$  have been used. In this work, data from 10 seed train cultivations from six campaigns with cultivation times between 72 - 96 hours per scale (meaning 4 to 5 measurement time points per scale including the measurement at time 0 (inoculation)) were used for investigation. Additional data sets have been generated for modeling purposes and first parameter estimations. Therefore, 12 batch cultivations in 4 flask scales (3 cultivations each) having filling volumes of 40, 70, 300 and 1,500 mL were provided. They cover cultivation time spans of 264 hours (11 days) each, meaning that also the stationary and death phase were included. Details, e.g. about controlled process parameters, analytics and data cleansing can be taken from Hernández Rodríguez et al. (2019).

## 2.4 Kinetic model and seed train prediction

The applied kinetic model, containing six mostly Monod-type algebraic equations (description of growth rate, death rate, substrate uptake and metabolite production kinetics) and 17 model parameters describing cell culture dynamics of total and viable cell density,  $X_t$  and  $X_v$ , as well as concentrations of glucose  $c_{Glc}$ , glutamine  $c_{Gln}$ , lactate  $c_{Lac}$  and ammonia  $c_{Amm}$ , can be taken from Hernández Rodríguez et al. (2019).

Seed train prediction starts over three consecutive bioreactor scales, containing two passing steps between them and is updated stepwise at each new sampling point. Evaluation of the prediction performance was realized calculating the relative prediction error in percent (= absolute deviation between predicted and afterwards added test data, divided by the test data). This criterion provides an intuitive assessment of the obtained prediction accuracy.

Since there are different overlapping factors having an influence on predicting several consecutive scales, (impact of estimated model parameters, e.g. maximum growth rate, and impact caused by possible deviations of the initial states at the beginning of each cultivation scale), the initial concentrations at the beginning of each cultivation scale are assumed to be known up to a certain experimental error. Otherwise, it would be unfeasible to make direct

conclusions about the impact of re-estimated model parameters on prediction performance. And the goal of this study is to learn about optimal updating strategies concerning model parameters and to develop an adequate workflow for model updating.

### 3 Results and discussion

The first goal of this study was to find out, how many data points (sampling points) are necessary for model parameter estimation to obtain high prediction accuracy of the predicted process future, using the widely applied weighted least squares estimation (WLSE) technique (see Section 3.1.1). In this context it was also investigated if old data points can be ignored for parameter estimation. Secondly, it has been investigated to what extend analysis of previous estimation steps concerning differences in model parameters can improve prediction performance of further seed trains of the same process (Section 3.1.2). The third goal was to investigate if the presented moving horizon estimation technique, which contains an arrival cost term (sometimes called *forgetting term*) and a parameter penalty term, is able to improve prediction performance (see Section 3.1.3). Finally, an iterative learning workflow is presented (see Section 3.2), containing the recommended steps for dynamic parameter estimation, based on the findings of this study.

#### 3.1 Application to an industrial CHO cell culture process - a case study

As a case study, a seed train data set (batch data set), consisting of three consecutive bioreactor scales were divided into the past (past data points, e.g. time interval  $t_1, \dots, t_3$ ) and the future (future data points, e.g. time interval  $t_4, \dots, t_{12}$ ) of the current seed train. The estimation horizon covers part of the past (or the whole past) and is used to re-estimate model parameters (see Figure 2). Each time new test data are added, the estimation horizon changes. Prediction performance of the remaining 'future' (= prediction horizon) is evaluated through the relative prediction error in percent, which is the discrepancy between measured

(observed) data and their expected (modeled) values, relative to the measured value. It was calculated for all six variables but since viable cell density is the main variable of interest only the results for viable cell density  $X_v$  are presented in the figures of this sections.

### **3.1.1 Impact of estimation horizon on prediction performance using weighted least squares estimation (WLSE)**

The following investigations deal with the question how many data points should be used for re-estimation of model parameters (= model updating) using the weighted least squares estimation (WLSE). First, the left horizon bound is kept fix meanwhile the right bound of the estimation horizon is growing after each iteration step.

Second, the estimation horizon length is kept fixed, i.e. adding a new data point means that the oldest data point of the estimation horizon is discarded (see Figure 2 A, top right).

Three different fix horizon lengths  $b$  were investigated,  $b = 2$ ,  $b = 3$  and  $b = 4$ . Figure 3 shows exemplary a problem that can arise, if only two data points are used for model updating. The prediction top left is based on model parameters determined from flask scale experiments. As mentioned before, the initial concentrations of every scale are assumed to be known in this investigation in order to directly see the influence of model parameters on prediction performance. It becomes clear, that viable cell density of scale 2 and scale 3 can be predicted with high accuracy (5% and 4.2% relative prediction error). Taking measurement deviations of approximately 5% into account it can be concluded, that for this example the prediction error of scale 2 scale 3 complies more or less with the amount of irreducible uncertainty. Prediction of scale one instead, shows a higher prediction error. The seed train prediction in Figure 3 top right grounds on model updating over two data points and it can be seen that prediction performance decreased for all three scales. If parameters change directly after model updating, two reasons could explain the decrease, a change in cell growth behaviour or a change in parameters due to measurement errors. Measurement deviations could get too much weight if the estimation is performed only over two data points, because

the estimation algorithm only focuses on minimizing the discrepancy between simulated and measured values at those two data points. Only, after addition of further data points it can be concluded which of both explanations may be appropriate. In the presented example, taking a third and fourth data point for model updating (bottom left and bottom right) indicates that the high prediction error after re-estimation over the first two data points (top right) probably grounds on measurement deviations.

In order to compare the different estimation techniques more systematically, the mean of the relative prediction error over all re-estimation steps was calculated per seed train, as well as the minimum and the maximum prediction error. This was performed for 10 seed trains/batches and the results are presented through boxplots in Figure 4 a) and b). The first group of boxplots within each diagram presents the mean prediction errors per batch over 10 seed trains, the second group shows the minimum values and the third group the maximum values.

In Figure 4 a) two estimation techniques are compared over all re-estimation steps from point in time  $t_2$  on (meaning that at the first model updating step, concentrations at sampling points  $t_1$  and  $t_2$  are known and the concentrations within the prediction horizon = remaining future,  $t_3, \dots, t_{end}$ , are predicted). A moving estimation horizon with a fixed horizon lengths of 2 data points is compared to a growing estimation horizon (keeping the left bound fixed at  $t_1$ ). It can be seen that taking only 2 data points for re-estimation of model parameters sometimes leads to low prediction accuracy (high maximum values between 21 and 66 %, left boxplot of maximum values). As mentioned earlier, this is not surprising because the measurements contain measurement deviations and those are getting too much weight if estimation is performed only over two data points.

For comparison of larger estimation horizon lengths (than 2 data points), the rel. prediction errors were considered from point in time  $t_4$  on to guarantee a fair comparison (i.e. the first considered prediction results were obtained for the prediction horizon  $t_5, \dots, t_{end}$ ). The results are presented in Figure 4 b). Taking at least three data points for estimation, typically

reduces the undesired effect described for two data points (lower maximum values, second boxplot of group maximum values). Taking the four recent data points for parameter estimation lead to a similar prediction performance as when taking three data points (averaged batch mean over 10 batches: 12.9% for horizon length = 2, 10.6% for lor horizon length = 3 and 10.3% for horizon length = 4). The best results are obtained if the whole past is used for model updating (9.99% averaged batch mean over 10 batches), especially concerning the maximum prediction error over one re-estimation cycle (fourth boxplot of group maximum values).

A detailed posterior analysis of the obtained model parameters at every updating step revealed further findings. If for example scale three should be predicted and model updating was performed including data from scale 1 and scale 2, the resulting prediction for scale 3 is not as good as prediction based on shake flask experiments. This is due to slight differences in some model parameters for scale 1 compared to scales 2 and 3 (maximum growth rate  $\mu_{max}$  shows a difference of 7% on average). So, prediction for scale 3, based on model updating over scale 1 and scale 2 is a compromise between the behaviour in both scales and therefore not optimal for scale 3. This observation is already mentioned in (Hernández Rodríguez et al., 2019). Obviously, the lower cell growth in reactor scale 1 occurs because mammalian cells have to adapt to different cultivation conditions (shaked system to stirred system). Variation of cell growth behaviour is a common challenge in cell cultivation and reported among others in (Xie and Schenkendorf, 2019). There may be differences in cell growth between different cultivation systems and also between different cultivation runs (seed trains). The latter can be taken into account through model parameter updating after adding new process data over a growing time horizon within a seed train. Differences in model parameters between cultivation scales could be detected as soon as one or more whole seed train data sets are collected. Then, model updating for further seed trains can be performed for each scale individually. In the next section, it was investigated, how model updating for each scale individually can change prediction accuracy for the remaining process future.

### **3.1.2 One parameter set for all scales vs. individual parameter sets per scale - Impact on prediction performance using WLSE**

Figure 5 shows the results for the relative prediction error if model parameters are re-estimated for each scale individually over a growing estimation horizon, keeping the model parameters for the other two reactor scales fixed. These results are compared to the previously described parameter estimation technique using only one parameter set for all scales with a growing estimation horizon.

Figure 5 a) shows exemplary for one seed train how the relative prediction error develops within one sample-to-sample cycle. It becomes clear, that especially at the beginning of scale 1 (when only two or three data points are available), prediction performance of the remaining future can be improved by using individual parameter sets per scale.

Figure 5 b) shows the results over 10 seed trains. It can be seen that prediction performance improves in terms of mean and minimum values, but especially concerning the maximum prediction error values. Therefore, it is recommended to analyze the updated model parameter values concerning differences between scales, as soon as seed train data are collected. Nevertheless, relative prediction errors of nearly 20% (maximum values) are still not entirely satisfying and another estimation technique, which take more information into account was investigated. The results are presented in the following sections.

### **3.1.3 Comparison: Parameter estimation including prior knowledge - Moving horizon estimation (MHE) vs. weighted least squares estimation (WLSE).**

While, up to know, the sample-to-sample updating cycles were subject of investigation, now model updating is also investigated regarding the batch-to-batch updating cycles, meaning that knowledge learned within one sample-to-sample cycle is passed to the next sample-to-sample cycle. More specifically, this means the re-estimation of model parameters is performed using data of the current seed train (up to the current point in time) as well as information about model parameters from the last seed train. Technically, this can be

realized using the moving horizon estimation technique described in section 2.2. The objective function contains a term describing the discrepancy between the modeled values (e.g. for viable cell density) based on the prior model parameters (estimated during the previous seed train) and a penalty term for the deviation between prior and current model parameter values. The tuning parameter for this penalty term and the weighting factor for the prior model values were chosen as follows:  $c_{\Delta p} = \frac{1}{n}$  and  $w_p = \frac{4}{n}$ , with  $n$ : number of data points within the estimation horizon. The dependence on  $n$  was introduced to reduce the influence of prior information when the data base of the current seed train grows.

Figure 6 shows the prediction results comparing the WLSE technique using 1 parameter set for all scales, using 1 individual parameter set per scale and the MHE technique using also 1 individual parameter set per scale and including prior knowledge. The diagram in Figure 6 (left side) shows exemplary the development of the relative prediction error for one seed train and all three methods, showing that especially during the first model updating steps, the WLSE techniques, using only one parameter set for all scales, shows a higher prediction error ( $> 30\%$ ) than using individual parameter sets (prediction error  $< 15\%$ ) or using the MHE approach (prediction error  $< 10\%$ ). An improvement can be achieved using individual parameter sets, one per batch. However, the best result is obtained using the MHE technique, even during the first updating steps when only very few data points are available for parameter estimation.

In order to obtain more representative results, the prediction error was evaluated over 10 seed trains. These results are presented in Figure 6 (right side). It can be seen that the MHE technique leads to lower mean and maximum values of relative prediction error than the compared techniques (see Figure 6 (right side) first three boxplots and last three boxplots). Moreover, it should be highlighted that there is a very low variation of the prediction error during an sample-to-sample updating cycle, which stands for a robust estimator. In section 3.1.1 it has been shown that estimation over two data points can lead to low prediction accuracy (sometimes more than 100%) in the presence of measurement deviations.

The MHE approach turned out not to be sensitive to measurement deviations or outliers. Prediction errors of more than 20 % were not observed over 10 tested seed trains and the mean comes to 9%.

It has been shown that a possible bias caused by measurement deviations can be prevented this way. Furthermore, this benefit can be achieved without high computational costs, because learned information is included without increasing the data base used for parameter estimation. Another benefit is, that batch-to-batch variabilities and variabilities between cultivation systems can be taken into account.

Furthermore, it was investigated if it is necessary to go 1, 2 or 3 seed trains back and to learn from these cultivations in order to receive an optimal prediction.

The following results show what happens if three consecutive batch-to-batch cycles are performed. This means that learned knowledge from seed train one is used for re-estimation and prediction of seed train 2. The obtained information is used for seed train 3 and the updated knowledge after seed train 3 is used for seed train 4 (see Figure 7 a)). This could be also described as a arrival cost window size of 3 seed trains / batches. The results presented in Figure 7 indicate that it is not necessary to include more than one historical seed train. The example in Figure 7 a) show that there is hardly any difference in prediction performance between the three arrival cost window sizes and at all points in time, a relative prediction error less than 10 % has been achieved. Although, there are differences between the investigated seed trains, Figure 7 b) shows that most seed trains show a mean prediction error of approximately 9 % and maximum values of approximately 14 %, not exceeding 18%.

### **3.2 Developed iterative learning workflow**

Based on the findings obtained in this study, a workflow for seed train prediction was deduced illustrating the required steps to account for variabilities and possible differences between scales. Simply spoken, the concept of the proposed workflow (see Figure 8) is to iteratively include knowledge gained from 'new'/current data, without discarding the knowledge already

obtained during the previous steps. After providing a kinetic model, model parameters from previous experiments and a process model (e.g. containing equations for the computation of passing between two scales/vessels) a new seed train for given initial concentrations is considered. Samples are taken in specific time steps and the following 'sample-to-sample' updating cycle is performed. At every new sampling point, the model is updated, meaning that model parameters are re-estimated through the moving horizon estimation technique taking model parameters from previous experiments into account in form of additional terms within the objective function (see equation 2).

The now obtained model parameters are used as new prior model parameters in form of additional terms within the objective function for the next cultivation run and so on (batch-to-batch updating cycle).

Moreover, if differences between scales are observed it is advisable to use different model parameter sets for each individual scale and to update them individually (first model parameters for the first scale, then for the second scale and so on). Then the next cultivation run (e.g. the next seed train within the same production process) is considered. A new 'sample-to-sample' updating cycle is performed, including learned knowledge from the first sample-to-sample updating cycle.

This proceeding lead to the results presented in Subsection 3.1.3. Compared to a common weighted least squares approach, prediction accuracy for viable cell density could be improved this way from 16.7 % to 8.6 % mean values and from 38.8 % to 13.2 % maximum values. Moreover, the standard deviation of the relative prediction error has been reduced from 9.5% to 2.9 % relative prediction error. These results emphasize the robustness of the proposed estimation workflow.

### **3.3 Common features of MHE and Bayesian parameter estimation**

The estimation techniques applied in this study belong to the group of point estimates, meaning that for each future point in time within the prediction horizon, only one value

per state variable is predicted (e.g. one value for viable cell density at point in time  $t_1$ , one value at point in time  $t_2$  and so on). Meanwhile the Bayesian approach belongs to the group of interval estimates, meaning that for every future point in time a whole interval (e.g. a 90%-interval) containing the possible values for a state variable is predicted, including the specific probability of occurrence for each value. In other words, prediction can be made at every future point in time including information about the predictive uncertainty (possible deviation). Common features of Bayesian parameter estimation and the presented MHE approach are, that both techniques include prior knowledge concerning the possible model parameter values as a kind of 'memory', integrated in the parameter estimation process. Especially when only few data are available and these data contain measurement errors, this 'memory' helps not to propagate the change in model parameters caused by the measurement deviation onto the prediction of the future process.

Despite the added values of the Bayesian approach mentioned in (Hernández Rodríguez et al., 2019), the MHE-based workflow presented in this contribution may be useful for many applications, where it is desired to save computational cost and to use a parameter estimation process simple to implement.

## 4 Conclusion

Two different estimation techniques (weighted least squares estimation (WLSE) and moving horizon estimation (MHE), having different objective functions) were applied and compared as well as the data base used for re-estimation of model parameters.

When using the WLSE technique at least 3 data points are necessary to reduce the maximum values of relative prediction error for viable cell density to less than 25% for the applied set up. Moreover, if differences between cultivation scales are observed, individual model parameter sets should be updated per scale (see row three of the overview in Table 1). Concerning the compared estimation techniques WLSE and MHE, it turned out that the optimal

solution was obtained using the moving horizon estimation technique (MHE) and individual set of model parameters per scale in case of differences between scales. High prediction accuracy, represented through a relative prediction error of less than 15% maximum values was achieved from the beginning on as well as less than 10% on average. This estimation technique requires prior information (estimated model parameters based on previous cultivation data), whereby estimation results based on one preceding seed train turned out to be sufficient.

These results were then formulated in form of an iterative learning workflow, which is presented in section 3.2. The performance of this workflow is described by the relative prediction error shown in row 6 of table 1 (Arrival cost window: 1 previous seed train). A mean value of 8.6% relative prediction error and maximum values of 13.2% on average were achieved for viable cell density.

To properly assess these results concerning prediction error, the intermediate precision (within-lab reproducibility) of viable cell density should be taken into account. As stated in (Hernández Rodríguez et al., 2019) it is expressed as 4.7% coefficient of variation for the investigated process. Now, the prediction error has to be considered in relation to the intermediate precision, giving the precision-weighted prediction error. Values close to 1 mean, that the prediction error is more or less the amount of irreducible uncertainty, which stands for stable models representing the stochastic nature of the environment. Therefore, low values are desired, meaning that there is not much reducible uncertainty. As can be seen in Table 1 columns 6 and 8, applying the MHE lead to a precision-weighted prediction error of 1.8 - 1.9 on average in comparison to 2.3-2.9 on average for the WLSE, which shows the superiority of the MHE-based iterative learning workflow.

Advantages of this workflow are that process knowledge is generated and incorporated during the process and possible variabilities (e.g. batch-to-batch variabilities concerning maximum cell growth) can be taken into account, which could lead to improved decision making (e.g. regarding points in time for passaging or split ratios). Problems arising from re-estimation

over very few data points (e.g. bias due to measurement deviation when re-estimating over two sampling points) can be mitigated through the integration of prior knowledge within the objective function. Furthermore, the computational cost can be kept manageable because past data points from previous cultivation runs don not have to be used for re-estimation of model parameters. Even though, the learned knowledge from these cultivations flow in the re-estimation process. The general form of this framework allows the application to other bioprocesses as well as an implementation as part of predictive control methods.

## References

- Anane, E., López C, D. C., Barz, T., Sin, G., Gernaey, K. V., Neubauer, P., and Cruz Bour-nazou, M. N. (2019). Output uncertainty of dynamic growth models: Effect of uncertain parameter estimates on model reliability. *Biochemical Engineering Journal*, 150:107247.
- Deppe, S., Frahm, B., Hass, V. C., Hernández Rodríguez, T., Kuchemüller, K. B., Möller, J., and Pörtner, R. (2020). Estimation of process model parameters. *Methods in molecular biology (Clifton, N.J.)*, 2095:213–234.
- Frahm, B. (2007). Seed train optimization for cell culture. In Pörtner, R., editor, *Animal cell biotechnology*, volume 1104 of *Methods in biotechnology*, pages 355–367. Humana Press, Totowa, NJ.
- Frahm, B., Lane, P., Atzert, H., Munack, A., Hoffmann, M., Hass, V. C., and Pörtner, R. (2002). Adaptive, model-based control by the open-loop-feedback-optimal (olfo) controller for the effective fed-batch cultivation of hybridoma cells. *Biotechnology progress*, 18(5):1095–1103.
- Haseltine, E. L. and Rawlings, J. B. (2005). Critical evaluation of extended kalman filtering and moving-horizon estimation. *Industrial & Engineering Chemistry Research*, 44(8):2451–2460.
- Hass, V. C., Lane, P., Hoffmann, M., Frahm, B., Schwabe, J.-O., Pörtner, R., and Munack, A. (2001). Model-based control of hybridoma cell cultures. *IFAC Proceedings Volumes*, 34(5):323–329.
- Hedengren, J. D. and Eaton, A. N. (2017). Overview of estimation methods for industrial dynamic systems. *Optimization and Engineering*, 18(1):155–178.
- Hedengren, J. D., Shishavan, R. A., Powell, K. M., and Edgar, T. F. (2014). Nonlinear mod-

- eling, estimation and predictive control in a monitor. *Computers & Chemical Engineering*, 70:133–148.
- Hernández Rodríguez, T. and Frahm, B. (2020). Design, optimization, and adaptive control of cell culture seed trains. *Methods in molecular biology (Clifton, N.J.)*, 2095:251–267.
- Hernández Rodríguez, T., Posch, C., Schmutzhard, J., Stettner, J., Weihs, C., Pörtner, R., and Frahm, B. (2019). Predicting industrial-scale cell culture seed trains—a bayesian framework for model fitting and parameter estimation, dealing with uncertainty in measurements and model parameters, applied to a nonlinear kinetic cell culture model, using an mcmc method. *Biotechnology and Bioengineering*, 116(11):2944–2959.
- Jewaratnam, J., Zhang, J., Hussain, A., and Morris, J. (2012). Batch-to-batch iterative learning control using updated models based on a moving window of historical data. *Procedia Engineering*, 42:206–213.
- Kern, S., Platas-Barradas, O., Pörtner, R., and Frahm, B. (2016). Model-based strategy for cell culture seed train layout verified at lab scale. *Cytotechnology*, 68(4):1019–1032.
- Kroll, P., Hofer, A., Ulonska, S., Kager, J., and Herwig, C. (2017). Model-based methods in the biopharmaceutical process lifecycle. *Pharmaceutical research*, 34(12):2596–2613.
- Kuchemüller, K. B., Pörtner, R., and Möller, J. (2020). Efficient optimization of process strategies with model-assisted design of experiments. *Methods in molecular biology (Clifton, N.J.)*, 2095:235–249.
- Le, H., Kabbur, S., Pollastrini, L., Sun, Z., Mills, K., Johnson, K., Karypis, G., and Hu, W.-S. (2012). Multivariate analysis of cell culture bioprocess data—lactate consumption as process indicator. *Journal of biotechnology*, 162(2-3):210–223.
- Mears, L., Stocks, S. M., Sin, G., and Gernaey, K. V. (2017). A review of control strategies for

- manipulating the feed rate in fed-batch fermentation processes. *Journal of biotechnology*, 245:34–46.
- Möller, J., Kuchemüller, K. B., Steinmetz, T., Koopmann, K. S., and Pörtner, R. (2019a). Model-assisted design of experiments as a concept for knowledge-based bioprocess development. *Bioprocess and biosystems engineering*, 42(5):867–882.
- Möller, J., Rodríguez, T. H., Müller, J., Arndt, L., Kuchemüller, K. B., Frahm, B., Eibl, R., Eibl, D., and Pörtner, R. (2019b). Model uncertainty-based evaluation of process strategies during scale-up of biopharmaceutical processes. *Computers & Chemical Engineering*, page 106693.
- Narayanan, H., Sokolov, M., Butté, A., and Morbidelli, M. (2019a). Decision tree-pls (dt-pls) algorithm for the development of process: Specific local prediction models. *Biotechnology progress*, 35(4):e2818.
- Narayanan, H., Sokolov, M., Morbidelli, M., and Butté, A. (2019b). A new generation of predictive models: The added value of hybrid models for manufacturing processes of therapeutic proteins. *Biotechnology and Bioengineering*, 116(10):2540–2549.
- Paul, K., Rajamanickam, V., and Herwig, C. (2019). Model-based optimization of temperature and ph shift to increase volumetric productivity of a chinese hamster ovary fed-batch process. *Journal of bioscience and bioengineering*, 128(6):710–715.
- Pörtner, Ralf, Barradas, P., Oscar, Frahm, Björn, Hass, and C, V. (01.01.2017). Advanced process and control strategies for bioreactors.
- Rao, C. V., Rawlings, J. B., and Lee, J. H. (2001). Constrained linear state estimation—a moving horizon approach. *Automatica*, 37(10):1619–1628.
- Schenkendorf, R., Gerogiorgis, D. I., Mansouri, S. S., and Gernaey, K. V. (2020). Model-based tools for pharmaceutical manufacturing processes. *Processes*, 8(1):49.

- Sommeregger, W., Sissolak, B., Kandra, K., von Stosch, M., Mayer, M., and Striedner, G. (2017). Quality by control: Towards model predictive control of mammalian cell culture bioprocesses. *Biotechnology journal*, 12(7).
- Ungarala, S. (2009). Computing arrival cost parameters in moving horizon estimation using sampling based filters. *Journal of Process Control*, 19(9):1576–1588.
- Walsh, G. (2018). Biopharmaceutical benchmarks 2018. *Nature biotechnology*, 36(12):1136–1145.
- Xie, X. and Schenkendorf, R. (2019). Robust process design in pharmaceutical manufacturing under batch-to-batch variation. *Processes*, 7(8):509.
- Xing, Z., Bishop, N., Leister, K., and Li, Z. J. (2010). Modeling kinetics of a large-scale fed-batch CHO cell culture by markov chain monte carlo method. *Biotechnology progress*, 26(1):208–219.

## Tables

Table 1: Summary table containing the mean and maximum values of the relative prediction error (Rel. pred. error) and the precision-weighted prediction error (Precision-weighted pred. error) for different estimation techniques (WLSE/MHE), estimation horizon, arrival cost windows and number of parameter sets (Par.-sets), here 1 parameter set for all three scales vs. 3 parameter sets = one per scale).

Estimation technique			Adaption	Mean values		Max values	
WLSE / MHE	Estimation horizon	Arrival cost window	Par.- sets	Rel. pred. error [%]	Precision- weighted pred. error [-]	Rel. pred. error [%]	Precision- weighted pred. error [-]
WLSE	Whole past of current seed train	none	1	11.8	2.5	23.2	4.9
WLSE	The last 2 data points	none	1	13.7	2.9	38.8	8.3
WLSE	Past of current scale	none	3	10.6	2.3	18.9	4
MHE	Past of current scale	Shake flasks	3	9.0	1.9	14.2	3
MHE	Past of current scale	1 previous seed train	3	8.6	1.8	13.2	2.8
MHE	Past of current scale	2 previous seed trains	3	8.6	1.8	13.5	2.9
MHE	Past of current scale	3 previous seed trains	3	9.0	1.9	14.7	3

## Figure legends

Figure 1: Role of parameter estimation techniques within a biopharmaceutical production process

Figure 2: Scheme of different estimation horizons. **A** i) Growing estimation horizon and ii) moving estimation horizon of fixed length, both applied within the weighted least squares estimation (WLSE) technique; **B** Growing estimation horizon including an arrival cost window, applied within the moving horizon estimation (MHE) technique.

Figure 3: Viable cell concentration over time exemplary for seed train no. 1, before parameter estimation (top left), in contrast after re-estimation of model parameters based on data points at  $t_1$  and  $t_2$  (top right), after re-estimation of model parameters based on data points at  $t_1$ ,  $t_2$  and  $t_3$  (bottom left) and after re-estimation of model parameters based on data points at  $t_1$ ,  $t_2$ ,  $t_3$  and  $t_4$  (bottom right).

Figure 4: a) Prediction error for viable cell density over 10 seed trains using weighted least squares estimation; moving estimation horizon of length 2 vs. growing estimation horizon (from the second data point on). b) Prediction error for viable cell density over 10 seed trains using weighted least squares estimation; moving estimation horizon of length 2, 3 and 4 vs. growing estimation horizon (from the fourth data point on).

Figure 5: Comparison of prediction error for viable cell density using weighted least squares estimation; One parameter set over all scales (growing estimation horizon) vs. three parameter sets (one parameter set per scale, over growing estimation horizon). a) Exemplary for seed train no. 1, b) results of 10 seed trains.

Figure 6: Comparison: Prediction error for viable cell density based on weighted least squares estimation (WLSE), which does not include any arrival cost term containing information from a-priori performed parameter estimation vs. moving horizon estimation (MHE) (which includes an arrival cost term containing information from a-priori performed parameter estimations, here from shake flask scale). a) Prediction error (relative deviation between predicted and measured values) over sampling times exemplary for seed train no. 1; b) Prediction error over 10 seed trains.

Figure 7: Impact of historical (arrival cost) window size for MHE technique. a) Prediction error for viable cell density, exemplary for seed train no. 1, b) prediction error for viable cell density over 10 seed trains and c) the corresponding updating scheme.

Figure 8: Proposed iterative learning workflow. Incorporation of knowledge (combining previous knowledge and new knowledge) during the whole process progress through sample-to-sample updating cycle during a batch / seed train and batch-to-batch updating cycle between batches / seed trains.

## Nomenclature

$\gamma$	Tuning parameter
$\Delta p$	euclidean distance between estimated and prior parameter vector
$c_{\Delta p}$	Constant value to regulate the impact of the prior information
$c_{Amm}$	Ammonia concentration [mmol L <sup>-1</sup> ]
$c_{Glc}$	Glucose concentration [mmol L <sup>-1</sup> ]
$c_{Gln}$	Glutamine concentration [mmol L <sup>-1</sup> ]
$c_{Lac}$	Lactate concentration [mmol L <sup>-1</sup> ]
$J$	Objective function
$l$	Number of estimated parameter components
$n$	Number of simulated points in time within estimation horizon
$p$	Parameter vector
$p_0$	Initial parameter vector
$t$	Time [hr]
$X_t$	Total cell density [cells/L]
$X_v$	Viable cell density [cells/L]
$y$	Concentration values in general (data)
$y_0$	Initial concentration values
$\hat{y}$	prior model values (simulated values based on prior model parameters)

## **Abbreviations**

**MPC** Model Predictive Control

**MHE** Moving horizon estimation

**WLSE** Weighted Least Squares Estimation