Novel calibration design improves knowledge transfer across products for characterisation of pharmaceutical bioprocesses

Laura M. Helleckes[1,2], Claus Wirnsperger[3], Jakub Polak[3],
Gonzalo Guillén Gosálbez[4], Alessandro Butté[3,4], Moritz von Stosch[3]*

[1] Institute of Bio- and Geosciences, IBG-1: Biotechnology, Forschungszentrum Jülich GmbH, Jülich, Germany
[2] Institute of Biotechnology, RWTH Aachen University, Aachen, Germany
[3] DataHow AG, Zürich, Switzerland
[4] Institute for Chemical and Bioengineering, Department of Chemistry and Applied Biosciences, ETH Zürich, Zürich, Switzerland
* Correspondence: Moritz von Stosch, DataHow AG, Hagenholzstrasse 111, 8050 Zürich, m.vonstosch@datahow.ch

Keywords: bioprocess development, cell culture, Gaussian process regression, knowledge transfer, meta learning

Running title: Calibration design for bioprocesses

Laura M. Helleckes	https://orcid.org/0000-0001-7825-7998
Claus Wirnsperger	https://orcid.org/0009-0000-8511-1428
Jakub Polak	https://orcid.org/0009-0008-5668-7071
Gonzalo Guillén Gosálbez	https://orcid.org/0000-0001-6074-8473
Alessandro Butté	https://orcid.org/0000-0003-3506-3792
Moritz von Stosch	https://orcid.org/0000-0001-7912-7992

Abstract

Modern machine learning has the potential to fundamentally change the way bioprocesses are developed. In particular, horizontal knowledge transfer methods, which seek to exploit data from historical processes to facilitate process development for a new product, provide an opportunity to rethink current workflows.

In this work, we firstly assess the potential of two knowledge transfer approaches, meta learning and one-hot encoding, in combination with Gaussian process (GP) models. We compare their performance with GPs trained only on data of the new process, i.e. local models. Using simulated mammalian cell culture data, we observe that both knowledge transfer approaches exhibit test set errors that are approximately halved compared to those of the local models when two, four or eight experiments of the new product are used for training.

Subsequently, we address the question whether experiments for a new product could be designed more effectively by exploiting existing knowledge. In particular, we suggest to specifically design a few runs for the novel product to calibrate knowledge transfer models, a task that we coin *calibration design*. We propose a customised objective function to identify a set of calibration design runs, which exploits differences in the process evolution of historical products. In two simulated case studies, we observed that training with calibration designs yields similar test set errors compared to common Design of Experiments approaches. However, the former requires approximately four times fewer experiments.

Overall, the results suggest that process development could be significantly streamlined when systematically carrying knowledge from one product to the next.

1 Introduction

Manufacturing processes for biopharmaceuticals must deliver high-quality products while maintaining economic feasibility. The development of these processes needs to be fast, exhibit low technical risk, require limited investment and produce sufficient process understanding. Three approaches have been introduced in the past to answer to these needs: platform processes, (miniaturised) high-throughput experimentation and Quality by Design (QbD).

(Miniaturised) High-throughput experimentation aims at generating experimental evidence at economically attractive small scales. Experimental results produced with small-scale, high-throughput devices have been shown to be representative of larger scale production process (Bareither & Pollard, 2011; Fink et al., 2021; Hemmerich et al., 2018; Kim et al., 2012). Consequently, these devices have gained popularity in industry for their ability to execute several runs in parallel, thereby reducing project timelines. One factor contributing to their widespread adoption is the integration of statistical Design of Experiments (DoE) (Politis et al., 2017), which facilitates the design of multiple experiments to achieve a planned outcome. Advances in automation such as liquid handling platforms have also fostered the developments in high-throughput experimentation.

Quality by Design (QbD) is defined as 'a systematic approach to development that begins with predefined objectives and emphasises product and process understanding and process control, based on sound science and quality risk management' (guidelines Q8/9/10/11 of the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use). Since its adaption in 2009, QbD found wide attention in the biopharmaceutical industry (Newcombe, 2014; ter Horst et al., 2021). Within QbD, DoE has been recognized as a central component (Rathore & Winkle, 2009), allowing for the systematic investigation of factors and their impact on process responses.

Platform processes seek to provide a template for process development. Typically, a proprietary host cell line, base and feed media as well as process conditions are developed once, and process conditions are only adapted slightly to improve quality or titer when changing the product. In this approach, process development is streamlined and fast. A body of knowledge about how the process behaves is acquired, while technical risk is significantly reduced. Regardless, both early-stage (involving cell line selection and screening for optimal conditions) as well as late-stage process development (focused on optimising quality and meeting economic target, as well as scaling up) will necessitate the execution of multiple process runs (Bradl et al., 2016; Rehberger et al., 2013; Xu et al., 2020, 2022). Astonishingly, the knowledge that is carried over from the development of the process of one product to the next, is limited to personal experience and risk based approaches, as promoted by QbD.

Apart from addressing process development needs, these approaches in combination produce data of sufficient quantity and quality to render possible a fundamental change in the way that processes are being developed (von Stosch et al., 2021). A key concept in this paradigm shift is the application of machine learning (ML) methods that have seen advances in data gathering, management and data science (Darmont et al., 2022; Jin et al., 2015; Narayanan et al., 2020b; Sarker, 2021). These ML methods can leverage data from processes of many molecules to streamline the process design of new molecules- a process known as

horizontal knowledge transfer (Hutter et al., 2021).

An efficient way to address the challenge of horizontal knowledge transfer are transfer and meta learning techniques. In transfer learning, the goal is to exploit knowledge that was gained from one task in a source domain to solve another task in a target domain (Weiss et al., 2016), e.g. by preserving several layers in a pre-trained neural network and only adapting the rest for unseen data. In the context of process development, a task is to find a process model for a specific product, such that the target domain would be a new product that is not explored yet. In contrast, meta learning, often associated with the term *learning to learn*, focuses on solving many source tasks in a joint fashion, usually to learn hyperparameters that can improve the performance of the learning algorithm (Upadhyay et al., 2023).

Regarding the choice of ML models for dynamic systems, Gaussian processs (GPs) gained popularity over the past decades (Deisenroth et al., 2009). GPs are non-parametric, Bayesian ML models; as such, they are simple to set up, work well with noisy and small data sets and naturally incorporate a measurement of uncertainty in the predictions (Kocijan, 2016). These attributes make GPs a valuable tool in biopharma, where they were demonstrated in applications such as modelling of bolus fed-batch cultures for antibodies (Cruz-Bournazou et al., 2022) or process monitoring with spectroscopic data (Tulsyan et al., 2021).

While many specialised transfer and meta learning methods for neural networks exist in the field of ML (e.g. reviewed in Hospedales et al., 2022; Tan et al., 2018), these paradigms are less explored for GPs, particularly in the context of bioprocesses. A simple approach to achieve horizontal knowledge transfer is joint training on historical and new product data using one-hot enconding (OHE) (Ashenden et al., 2021). This involves representing categorical features by several binary features. Recently, this approach was extended to vector embeddings, where Hutter et al. showed that transfer learning approaches slightly outperformed models that were trained without historical bioprocess data (Hutter et al., 2021). However, applications of meta learning, where an inductive bias is learned from historical data, remain scarce. Interestingly, meta learning for GPs was recently explored in the work of Rothfuss et al.; in particular, the study found a framework that addresses the challenge of overfitting to the meta-training tasks for small data set and allows generalisation as well as good scaling (Rothfuss et al., 2021).

Besides these modelling challenges, the related question of how to identify suitable experimental conditions to understand the behaviour of new processes has been addressed in the past. DoE is an old field that goes back to Sir Ronald A. Fisher in the 1920s (Politis et al., 2017). On a high level, the two classical approaches used in bioprocess development are I) screening designs, which aim at identifying critical design factors and their main effects, and II) response surface designs, which aim to optimise these critical factors (Beg & Swain, 2021). Examples for screening designs are fractional factorial designs or definitive screening designs (DSDs), the latter being suitable for screenings with many design factors that have confounding effects (Jones & Nachtsheim, 2011). The group of response surface designs include central composite, full factorial or Doehlert designs.

Besides classical DoE approaches, Bayesian optimisation became of interest for the field, a concept that was named one of the most important statistical ideas of the past 50 years by Gelman and Vehtari, 2021 in the context of adaptive decision analysis. In experimental designs, Bayesian optimisation has fostered significant improvements in the number of experiments that are required to optimise a process (Greenhill et al., 2020). However, initial designs for new processes, whose data is required to fit an initial model for Bayesian optimisation, are usually chosen by simple, uninformed methods: space-filling strategies like latin hypercube sampling (LHS) or classical DoE methods like factorial designs. As an alternative, we hypothesise that his-

torical data contains information on experimental designs that reveal the underlying process dynamics more efficiently, thus leading to less required experiments to fit a new process model compared to state-of-the-art methods.

In this work, we show that horizontal knowledge transfer improves model performance and that the choice of initial experiments to start process optimisation can be improved by using existing knowledge. In the first part, we investigate different methods of knowledge transfer. To efficiently handle small data sets, which are present at the start of new product development campaigns, GPs are applied as the model of choice. In a first step, we adopt the meta learning approach by Rothfuss et al., called PAC-optimal hyper-posterior (PACOH), to a biopharmaceutical process model based on GPs. In a case study with historical data sets and one novel product, we benchmark the model performance of GPs trained by the new algorithm compared to those trained by other knowledge transfer models and models that are only trained on data of the novel product.

In the second part of this study, we explore how to use transfer learning to calibrate process models to a new, completely unseen product. We refer to this novel experimental design as *calibration design* and determine the process conditions based on a customised objective that takes the differences in the process evolution of historical products into account. To further highlight the connection between both part of the study, an overview of the underlying processes is given in Fig 1.

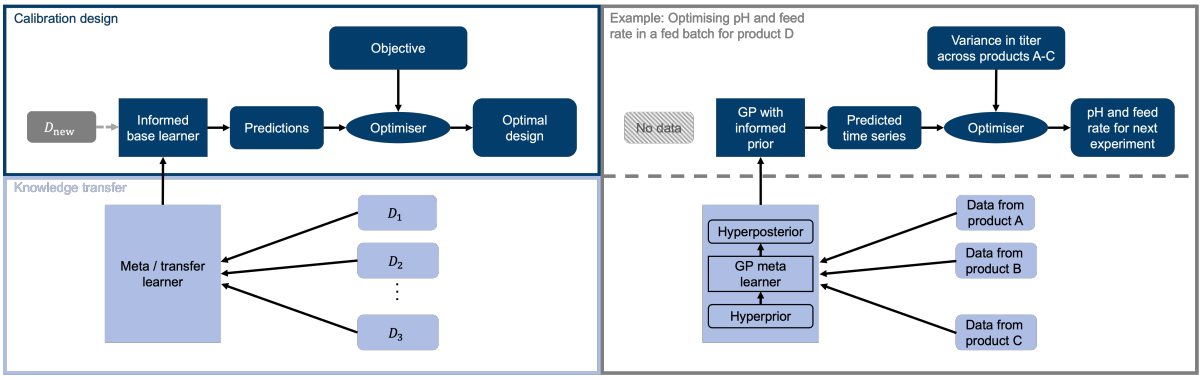

Figure 1: Link between knowledge transfer (light blue box) and calibration design (dark blue box), highlighted by a bioprocess example (grey box).

In the first part of this work, different strategies for knowledge transfer are compared, e.g. meta learning. To identify optimal designs to calibrate a process model for a new process, initial data from so-called excitation designs (D_{new}) is usually required. However, meta or transfer learners already inform the base learner and its parameters from historical data. This allows to find optimal designs to calibrate a process model even without data from the new process. This task of generating initial optimal designs for an unseen product is coined *calibration design*. In the example (grey box), meta learning is used to train a GP. Based on data from three historical cell lines, the optimal pH and feed rates regarding for a new process with cell line D can be identified. The objective for optimisation of calibration designs should facilitate to differentiate the new cell line from the historical ones. A suitable customised objective is further introduced in Section 2.3.2.

The figure shows the typical scheme of knowledge transfer methods (light blue). In particular, these learners can inform the model for a new process, the so-called base learner, during training. The informed base learner can be subsequently used to identify optimal designs to calibrate a model during calibration design (dark blue). The example of a fed-batch process with pH and feed rate (grey box) is illustrating the connection further. Starting from the top left of Fig 1, data from a process D_{new}, which includes time series data of viable cell density (VCD), glucose etc., as well as information about control variables, is used to train a

process model, e.g. a GP as base learner. Once trained, the base learner can be used to make predictions of time series given the initial conditions. During knowledge transfer (Fig 1, bottom left), overarching training schemes leverage historical data and thus inform the base learner. In case of the fed-batch example (Fig 1, grey), time series data from different products A-C is available. Note that time series data (in contrast to end-point measurements) is important as an input when having the goal to predict process evolution and to increase process understanding, e.g. for regulatory approval. A meta learner for GPs uses these data sets to learn the expectation for GP parameters for different products, which can be transferred to the new product D. As a result, even if no data is available for product D yet, a base learner can be trained based on the historical data of products that have been observed in the past. This approach contrasts with common DoE strategies, which require an initial excitation design of the novel product to calibrate the base learner. Note that meta learning might be substituted by other methods of knowledge transfer, as later demonstrated in this study.

Once trained, calibration design (Fig 1, dark blue) can exploit the base learner to generate predictions of the process evolution for different process designs, i.e. initial conditions such as initial VCD, initial pH etc. An optimiser utilises these predictions to determine the most suitable design given an objective. In calibration design, this objective is to identify designs which will be most informative to characterise the new process, e.g. by taking the points with most variance in historical data into account. Further details on the choice of objective will be discussed in Section 2.3.2 and Section 3.3. In the fed-batch example with different cell lines (Fig 1, upper right), the suggested initial pH and feed rates are iterated until the optimal design for the new process is found. These settings would then be tested in a wet-lab experiment and the respective data could be used to retrain the GP base learner for the next round.

In this study, we explore both knowledge transfer and calibration design. The paper is structured as follows. In Section 3.1, we present the publicly available benchmarking *in silico* data set that was used throughout the study. In Section 3.2, we benchmark different knowledge transfer methods for GPs, including the PACOH meta learning approach. Afterwards, Section 3.3 introduces a novel objective for calibration design and compares the model performance of resulting GP models to those trained with common DoE designs or LHS. Finally, we shed light on potential use-cases as well as future directions for the approach (Section 3.5).

2 Material and methods

In this paper, two different methodologies are targeted: knowledge transfer methods and calibration design (identifying optimal initial experimental designs of a new product for model calibration). Both approaches require an underlying process model. GPs are the model of choice in this work due to their capability to generalise from small data sets and their ability to handle noisy experimental data, which is relevant for biopharmaceutical processes. More precisely, we use GP regression with mass balances to model mass evolution over time, resulting in a hybrid model. Further details on the hybrid GP model for cell culture processes can be found in Narayanan et al., 2019; Polak et al., 2024. In this section, the mathematical description of GP regression for cell culture processes is briefly summarised, followed by an overview of knowledge transfer approaches and introduction of the novel task of calibration design.

2.1 GP regression

GPs are Bayesian machine learning models used throughout various regression problems. In a simple description, consider n m-dimensional input and n scalar output data points. Given some input data $\mathbf{X} \in \mathbb{R}^{n \times m}$ and corresponding output data $\mathbf{y} \in \mathbb{R}^n$ and given a new data point $\mathbf{x}' \in \mathbb{R}^m$, a GP implements the predictive conditional distribution of the output variable y':

$$y' \sim \mathcal{N}(\mathbf{x}'|\mathbf{X}, \mathbf{y}) \equiv \mathcal{N}(m(\mathbf{x}', \mathbf{X}, \mathbf{y}), \sigma^2(\mathbf{x}', \mathbf{X}, \mathbf{y})) \tag{1}$$

where $m(\mathbf{x}', \mathbf{X}, \mathbf{y})$ and $\sigma^2(\mathbf{x}', \mathbf{X}, \mathbf{y})$ describe the predictive mean and variance for a new observation $\mathbf{x}'$, given the data $\mathbf{X}$ and $\mathbf{y}$. A detailed description of GP regression can be found in Rasmussen and Williams, 2006.

To model the time evolution of the process variables, we utilise GP regression in an autoregressive fashion to propagate the process dynamics along the time dimension. To facilitate this, the process is discretised along the time dimension and for all modelled species their generic mass balances are parameterised as follows:

$$\frac{d(c \cdot V)}{dt} = R(x) \cdot V + u_f \tag{2}$$

where c is a vector of concentrations of the modelled process dynamics variables, V the culture volume and u_f a vector of mass feed rates (nonzero only for compounds that are fed). $R(x)$ describes a vector containing the rate of production (or consumption when negative) of the modelled species as a function of all process variables x at a certain time t input to the GP model. The input x is given by a combination of uncontrolled process dynamics variables and controlled process variables (such as pH or temperature). Discretisation of the process in time allows us to utilise the forward-difference formula for the time-derivative of the concentration, where Eq 2 now becomes:

$$\frac{d(c \cdot V)}{dt} = V\frac{dc}{dt} + c\frac{dV}{dt} \cong V\frac{c(t_{i+1}) - c(t_i)}{t_{i+1} - t_i} + c\frac{dV}{dt} = R(x_i)V + u_f \tag{3}$$

where $x_i \equiv x(t_i)$ denotes the vector of all input process variables at time step t_i. Further simplification of Eq 3 allows us to lump the mass balances into a single effective rate of production $\tilde{R}$, as can be seen below in Eq 4.

$$\frac{c(t_{i+1}) - c(t_i)}{t_{i+1} - t_i} \cong R(x_i) + \frac{1}{V}\left(u_f - c\frac{dV}{dt}\right) \stackrel{!}{=} \tilde{R}(x_i, u_f, V) \tag{4}$$

To approximate this effective rate of production $\tilde{R}$ we can utilise any machine learning model. In our case, this approximation is parameterised through the predictive mean of a GP regression model: $GP(x, u_f, V) \approx \tilde{R}(x, u_f, V)$. Here we use the notation $GP(x, u_f, V)$ interchangeably with the predictive mean of the GP. Through this discretisation of the process mass balances in time, as well as the approximation of the effective rate $\tilde{R}$ through the predictive mean of a GP regression model, we can autoregressively propagate the state of the process along discrete steps in time:

$$c(t_{i+1}) \approx GP(x_i, u_f, V) \cdot (t_{i+1} - t_i) + c(t_i). \tag{5}$$

Throughout this work, a squared exponential (SE) kernel with automatic relevance determination (ARD) was used, with the kernel hyperparameters and noise parameters determined from the training data using

maximum likelihood estimations. To improve model robustness and predictive performance, a simple mean averaging ensemble approach was chosen with 30 GP models comprising the ensemble, each sub-sampling 50% of the training data experiments. The high number of models in the ensemble, together with the chosen sub-sampling percentage, ensure adequate coverage of the training data while improving the robustness against over-fitting. Further background is given in Pinto et al., 2019.

2.2 Knowledge transfer approaches

In Section 3.2, we compare different models for knowledge transfer with individual-product models as a benchmark. In particular, we use two knowledge transfer approaches, OHE and meta learning. OHE (also referred to as dummy variables) is perhaps the easiest and most widespread approach to transfer knowledge across different categories (Hutter et al., 2021) and therefore adopted. Meta learning approaches have shown great promise for knowledge transfer in other disciplines, whereas they have not yet been applied to biopharmaceutical processes to the best of our knowledge. In particular, meta learning frameworks for GPs have not been studied yet. Hence, this study will allow us to compare these two knowledge transfer approaches. By adopting two different methods, we also aim at evaluating whether knowledge transfer is useful for bioprocess modelling independently of the method.
All models are evaluated based on their performance of the predicted effective rate (Eq 4) on a test data set with 100 simulated experiments (Section 2.4). Similarly to the performance metric given in Section 2.3.4, we calculated the root-mean-square error (RMSE) of the predicted rates for VCD, glucose, glutamine, ammonia, lactate and titer. In the following, the methodology of the various models is detailed.

2.2.1 Individual-product models

In the simplest approach, only the data for the product of interest is used to train a GP regression model as described in Section 2.1. This serves as a benchmark for meta learning and one-hot encoding models, which are expected to outperform individual-product models in terms of the test set error (Section 2.3.4) if few experiments are provided. We varied the number of available experiments between one and 20, which were randomly sampled from the available training data generated as described in Section 2.4.

2.2.2 Meta learning for GPs

In the usual training setup of a GP regression model, that is training the *base learner* for a new product (data set D^*), the training is initialised with pre-defined *prior probabilities* P for the mean and kernel functions, which are usually broad and uninformative if no specific process knowledge is available. In meta learning for GPs, the idea is to obtain informed prior probabilities for a base learner from a *hyperposterior* Q_{hyper}; that is the hyperposterior is learned by a *meta learner* trained on several historical data sets D_i with $i = 1, ..., N_{\text{historical}}$. This concept is illustrated in Fig 2.
Meta learning in this study was implemented by combining the PACOH framework by Rothfuss et al., 2021 with a stepwise GP described in Section 2.1. Essentially, the mean and kernel functions for the GP are defined as parametric functions, more precisely neural networks, whose parameters are meta-learned in training. As for other ML models, hyperparameters such as the number of layers and neurons in the neural network can determine the predictive output. As done in the original study, we thus performed a hyperparameter optimisation using `Hyperopt` (Bergstra et al., 2013). More details and the optimised parameters can be found

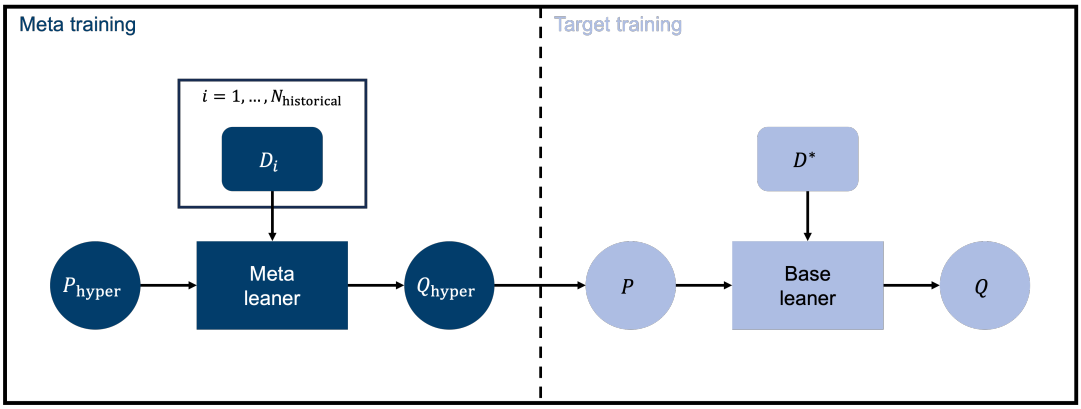

Figure 2: Overview of meta learning concept for GPs.
In a standard GP setup, a base learner is initialised with prior probabilities and the hyperparameters of the GP kernel are learned based on a data set D^*, resulting in their posterior distribution. The GP, i.e. the posterior distribution Q of the hyperparameters, can then be used for prediction. This conventional setup is shown in the right panel as *target training*. In meta learning, the prior probabilities are obtained from a hyperposterior Q_{hyper}, which is learned on historical data sets D_i with $i = 1, ..., N_{\text{historical}}$ during *meta training*. This step leads to an informed prior distribution for target training based on historical data sets. This plot is a simplified version of Figure 1 presented in Rothfuss et al., 2021, not showing *meta testing* and *target testing* as additional stages.

in the Supplementary Information. For further details on meta learning and the optimisation guarantees, we refer to Rothfuss et al., 2021. Meta learning models are benchmarked against local models and OHE models, which are described in the following subsection.

2.2.3 One-hot encoding models

OHE is a common binary encoding strategy to turn categorical variables into numerical feature vectors. The encoding turns any categorical feature c with N_c potential categories into N_c additional binary features, also referred to as dummy variables, with a 1 denoting the presence of the corresponding category for the data points respectively. Encoding the product information this way, the combined data shared across all products can be utilised to train a model. Keeping the SE kernel with ARD, OHE variables introduce additional similarity conditions to the kernel. A detailed explanation of the effects of binary encoded variables onto the kernel values can be found in Hutter et al., 2021.

2.2.4 Choice of training data for benchmarking

For benchmarking of knowledge transfer methods, we use the six data sets generated as described in Section 2.4, mimicking different products. We assume one to be the new product while the other five serve as historical data for knowledge transfer, e.g. meta training. This results in a minimal data set during training that comprises 100 experimental runs, 20 of each historical product. A varying number of simulated runs of the new product is added to the minimal data set as indicated in the respective figures. In addition, meta learning and OHE models are compared to local, i.e. individual-product, models, which only use data from the new product for training.

2.3 Calibration design of process conditions

The term *calibration design* refers to the task of identifying a suitable experimental design for a determined number of runs, with the aim to calibrate an existing model on process data of a new product or cell line of interest (in the following, we in short refer to *process data of a new product of interest* as *new product*). The idea for calibration design is to maximise the dissimilarity in the final titer among the historical data sets. The reasoning is that certain experimental designs are better suited to reveal differences in cell line behaviour than others. As a hypothesis, choosing these experimental designs, together with knowledge transfer models, leads to model calibration requiring less experiments and resulting in better model performance compared to standard DoE approaches, thus aiding process understanding.

As shown in Fig 1, the base learner can be used to make predictions of the process evolution based on a given design, i.e. the initial conditions, feed rates etc. These predictions in turn can be used in an optimisation step to find optimal designs for model calibration. In this study, we use a novel, customised objective function to identify the best process conditions to calibrate a OHE-GP regression model to the new product, which was initially trained on data of other products. For this, the initial OHE model is trained with 20 experiments of three historical products, designed with LHS over the design space. This step is shown as the meta / transfer learner in Fig 1. The trained OHE-GP model (informed base learner in Fig 1) is the basis for further generation of experimental designs for the new, completely unseen product, using the objective described in Section 2.3.2 in an optimisation detailed in Section 2.3.1.

In the simplest case, we designed the conditions for a single experimental run $n_\mathrm{e} = 1$, e.g. the setup of a single bioreactor. Here, solely the objective from Section 2.3.2 was optimised. Additionally, designs for parallelised bioreactor setups ($n_\mathrm{e} > 1$), which are common for screening, were explored. The design of parallel experiments with Bayesian optimisation is an ongoing field of research (González & Zavala, 2023), in which popular strategies include the *constant liar* method (Ginsbourger et al., 2010) or a distance-based optimisation of points (De Luca et al., 2023). In the distance-based approach, the first experiment is determined by Bayesian optimisation, which inherently uses a surrogate model to approximate the objective function over the design space. In this case, the design of the first experiment is identified purely by optimising the respective objective for similarity. For further experiments, the objective function for the first experiment is augmented by a term that is regulating the distance between suggested designs, thus jointly optimising the objective and the distance.

As the constant liar technique as implemented in `skopt` led to locally clustered suggestions for parallel experiments (supporting information), the distance-based approach was chosen for calibration design. Hence, the designs of the next experiments are sequentially identified (prior to experimentation) by taking the objective function $f(\mathbf{u})$ with design factors $\mathbf{u}$ (Section 2.3.2), as well as the distance of these factors to the previously suggested design d:

$$f_\mathrm{parallel}(\mathbf{u}) = \alpha \cdot f(\mathbf{u}) + (1-\alpha) \cdot \frac{1}{D} \sum_{d}^{D} ||\mathbf{u} - \mathbf{u}_d|| \tag{6}$$

with $\mathbf{u}$ as the vector of design factors that are optimised for the current experiment, D as the total number of experiments that have been optimised so far and the scalar α as the weight of the objective compared to the distance. In equation Eq 6, $||\bullet||$ denotes the euclidean distance over all factors, which are normalised by the boundaries of the design space to yield values between 0 and 1 before calculating the euclidean distance.

This approach is iteratively repeated until the desired number of parallel experiments is obtained.

2.3.1 Optimisation framework

Bayesian optimisation of different objective functions was performed with scikit-optimize in version 0.9.2, using the `Optimizer` class with a gradient-boosted regression tree as surrogate model. A LHS with 100 samples was used to initialise the surrogate model and at least 1400 iterations were run for each suggested experiment to ensure convergence. All optimisations were repeated at least five times with different seeds to subsequently assess model performance.
As an alternative to the distance-based design of parallel experiments (Eq 6), scikit-optimize offers parallel optimisation using the *constant liar* (Chevalier & Ginsbourger, 2013) methodology. In short, the surrogate model, which is an inherent part of the optimiser to approximate the objective function, is first queried for one optimal data point. Subsequently, this data point is added to the observations made so far with a dummy response, assuming either the mean or the minimum so far seen in the surrogate model as the corresponding function value. The surrogate model is then queried for the new optimum and the process is repeated as often as parallel points are to be designed. Afterwards, the true objective function is evaluated at all points in the batch simultaneously and the surrogate model is updated. Further details on the implementation can be found in the scikit-optimize documentation (Shcherbatyi et al., n.d.).

2.3.2 Objective function

In our customised objective for the optimisation step in calibration design, we reason that data of historical products can inform the choice of new designs that enhance the predictive power of the process model. As a hypothesis, certain experimental designs are better suited to reveal differences in cell line behaviour than others. Using knowledge transfer models, these differences in relevant process parameters such as titer are to be inferred using historical data. For this purpose, a novel, customised objective function is introduced, using the same notation as De Luca et al., 2023 for comparability.
In the case study, we chose to optimise based on the predicted interval of final titer $T_i(t_{\text{end}}, \mathbf{u})$ of historical product i, given a set of control variables $\mathbf{u}$. We focus on titer as its optimisation (along with product quality attributes) typically guides the development activities. Let $\tilde{T}_i(t_{\text{end}}, \mathbf{u})$ denote the median prediction and $\sigma_i^2(t_{\text{end}}, \mathbf{u}) = \text{Var}\left[\tilde{T}_i(t_{end}, \mathbf{u})\right]$ the variance of the predicted interval for product i out of N total historical products. We suggest the following objective that is calculated pairwise between products:

$$f(t_{\text{end}}, \mathbf{u}) = \sum_{i}^{N} \sum_{j,j>i}^{N} \frac{\left|\tilde{T}_i(t_{\text{end}}, \mathbf{u}) \cdot \sigma_j^2(t_{\text{end}}, \mathbf{u}) - \tilde{T}_j(t_{\text{end}}, \mathbf{u}) \cdot \sigma_i^2(t_{\text{end}}, \mathbf{u})\right|}{\sigma_i^2(t_{\text{end}}, \mathbf{u}) + \sigma_j^2(t_{\text{end}}, \mathbf{u})} \tag{7}$$

where t_{end} represents the end of cultivation and $\mathbf{u}$ a set of control variables, including initial VCD, that can be optimised. In its core, the objective is looking at the pairwise difference in predicted final titer scaled by the width of the predicted interval; it is thus trying to evaluate designs with factors $\mathbf{u}$ that maximise this difference while taking prediction uncertainty into account.

2.3.3 Comparison to further DoE strategies

To evaluate the customised objective proposed in Section 2.3.2, benchmarking of GP models trained on the calibration design runs obtained by optimisation is performed. In particular, we compare to models trained

on designs based on LHS or DSD. For LHS, experiments were generated with different random seeds within the design space, not using center points. DSD is a design strategy widely applied in bio-manufacturing and pharma (Dodds et al., 2022). To generate the design, we used the `definitive-screening-design` package with version 0.4.0 (Ongari, 2023). If not indicated otherwise, the training data of historical products is augmented by the suggested experiments for training a OHE model. In this case, the minimal data set for training are 20 runs of each historical product, generated as described in Section 2.4. The index "local", however, indicates that the models were trained on the experiments of the novel product alone, which is a standard for most DoE methods.

2.3.4 Performance metric for GP models

A test set of $n_e = 100$ unseen experimental conditions for the new product (Section 2.4) was used to assess the model performance, i.e. the predictive quality of models after training, which was determined with data from experiments that were simulated according to the conditions planned with the different design strategies. These predictions and the test data set are used to calculate the relative RMSE of each product:

$$\mathrm{RMSE} = \frac{1}{\sigma_T}\sqrt{\frac{1}{n_e}\frac{1}{n_t}\sum_{j=1}^{n_e}\sum_{t=1}^{n_t}\Big(T(t,j) - \tilde{T}(t,j)\Big)^2} \tag{8}$$

where n_t is the number of observations of titer made over time, $T(t,j)$ are the measured titer values at time t from the test set experiment j, and $\tilde{T}(t,j)$ the respective predicted median titer. To standardise the error, it is divided by the standard deviation in the actual measurements across the 100 experiments in the test data set, σ_T. Values of the RMSE well below 1.0 represent a good model performance while values above 1.0 indicate that a constant mean prediction over time would outperform the GP model prediction. For the RMSEs in Section 3.2, titer is substituted by the respective process variables.

2.4 Generation of benchmarking data sets

The *in silico* model used to generate data is based on the macro-kinetic model for a fed-batch chinese hamster ovary (CHO) cell culture as described in Craven et al., 2013 and Xing et al., 2010, with adaptations to impose complex non-linearities in growth rate dependencies and to account for pH and temperature shifts (Narayanan et al., 2020a). The *in silico* model has been presented and utilised in past works (e.g. De Luca et al., 2023; Hutter et al., 2021). We simulated six different cell lines (which are associated with six different products, A-F), by using the *in silico* model and changing for each of the six cell lines the model parameters, invoking different process behaviours. This *in silico* data of the six products is used for benchmarking the knowledge transfer methods. Note that in the following, the terms *cell line*, *product* and *processes* are used interchangeably to differentiate between the six data sets. For all six products, parameters for biomass growth associated to glucose and glutamine, the maximum growth rate, product formation rate, inhibition by lactate and ammonia, as well as pH and temperature tolerance were varied, in total leading to a variation of 13 internal parameters for each product. To enable a subset that is more homogeneous and one that is more heterogeneous, we additionally divided the six cell lines in two groups: cell lines A, B, E and F are able to consume lactate, while cell lines C and D lack this ability. Further investigation of the resulting phenotypes will be presented in Section 3.1.

For each of the six cell lines, the control inputs over a fixed cultivation time of 14 days were determined to

follow the subsequently described set point profiles:

Table 1: Overview of process parameters varied for data set generation and optimisation
For benchmarking of meta learning vs. OHE, 15 parameters were varied as shown in column **Knowledge transfer range**. For calibration design, seven parameters were varied as shown in column **Optimisation range**. In cases where two ranges are given, the first was used for lactate-consuming products A, B, E and F and the second range for products C and D. The first value in the optimisation range was used for the case study with products A, B, E, F (lactate-consuming only), the second value for the case study with products A, B, C, D.

Process parameter	Description	Knowledge transfer range	Optimisation range	Unit
StirrerSpeed	constant value	[150 , 250]	[200 , 200]	-
DissolvedOxygen	constant value	[30 , 80]	[40 , 40]	% air saturation
pH-Phase1	pH profile	[6.5 , 7.5]	[6.5 , 7.5]	-
pH-TimeSwitch	(before, switch	[6 , 14]	[7 , 7]	d
pH-Phase2	time, after)	[6 , 7]	[6 , 7]	-
Temp-Phase1	temp. profile	[36 , 38]	[36 , 38]	°C
Temp-TimeSwitch	(before, switch	[6 , 14]	[7 , 7]	d
Temp-Phase2	time, after)	[35 , 37]	[35 , 37]	°C
BolusGlucose	concentration &	[2 , 6] / [0.5 , 1]	[2 , 6] / [0.5, 6]	mM
BolusGlutamine	time of daily	[6 , 8] / [0.5 , 4]	[6 , 8] / [0.5 , 8]	mM
FeedStart	bolus feeds for	[1 , 3]	[3 , 3]	d
FeedEnd	glucose/glutamine	[9 , 13]	[13 , 13]	d
VCD		[0.2 , 2]	[0.2 , 2]	10^{-3} cells/mL
Glc	initial conditions	[2 , 6] / [2 , 5]	[4 , 4]	mM
Gln		[4 , 6] / [2 , 4]	[4 , 4]	mM

The variables whose name contains Phase represent the value of the set point during the respective phase, those with TimeSwitch indicate when the switch from one phase to the other occurred. Initial concentration values of VCD, glucose and glutamine were assumed to be controllable (i.e. they can be optimised), those of lactate, ammonia and titer were fixed. Bolus feeds were simulated as a volumeless addition of the respective concentration described in Table 1 to the current concentrations of glucose and glutamine in the reactor, occurring once per day between the start and end day. This resulted in a total of 15 factors that can be varied for the process data generation and subsequently process optimisation.
The *in silico* model was used to simulate the evolution of six process variables, namely VCD, glucose, glutamine, ammonia, lactate, and titer, where the data is obtained as daily measurements of process variable concentrations c. In the case of noisy measurements of concentrations $\widetilde{c}$, they were corrupted with relative Gaussian noise of 2% and with absolute Gaussian noise applied as follows:

$$\widetilde{c} = c * (1 + 0.02 * \mathcal{N}(0,1)) + \mathcal{N}(0, \sigma_i) \tag{9}$$

In Eq 9, σ_i is corresponding to default standard deviations of the experimental error for different variables (VCD: $0.03 \times 10^6\,\mathrm{cells/mL}$, glucose: $0.5\,\mathrm{mM}$, glutamine: $0.1\,\mathrm{mM}$, ammonia: $0.1\,\mathrm{mM}$, lactate: $0.3\,\mathrm{mM}$, titer: $10\,\mathrm{mg\,L^{-1}}$).
Initial conditions of VCD, glucose, and glutamine were additionally corrupted only in form of an additive, absolute error consisting of Gaussian noise with standard deviations as defined above. Noisy measurements were included in the simulations to facilitate the comparison of different strategies under more realistic con-

ditions. However, run-to-run variation might impact the model performance and the width of the prediction interval, which can be expected to increase with increasing run-to-run variation.

Subsequently, the data were used to fit the stepwise GP model, as detailed in Section 2.1. To generate historical data sets for each product, which can be used for the different knowledge transfer strategies (Section 2.2), LHS was used to vary the 15 process parameters in the ranges described in Table 1. Data sets with 20 experiments for training and 100 experiments for testing were thereby generated, aiming at covering the entire process parameter space. For the calibration design, a subset of seven parameters was chosen, which are initial VCD, `pH-Phase1`, `pH-Phase2`, `TempPhase1`, `TempPhase2`, `BolusGlucose` and `BolusGlutamine`. The process data for the six products with and without noise is published in an accompanying repository (Helleckes et al., 2024).

3 Results and discussion

3.1 Design of benchmarking data set

The prerequisite to compare knowledge transfer methods and optimisation procedures for experimental design is the availability of suitable benchmarking data sets. For this study, we simulated process data for six different products A-F with the framework described in Section 2.4. Since the benchmarking data sets should represent the case of different historical process development campaigns, it is important that the different simulated product cases, that is cell lines, show a high degree of inter-product variability. Visually this can be assessed in Fig 3, where the process behaviour of the different cell lines for identical process conditions and initial values is shown.

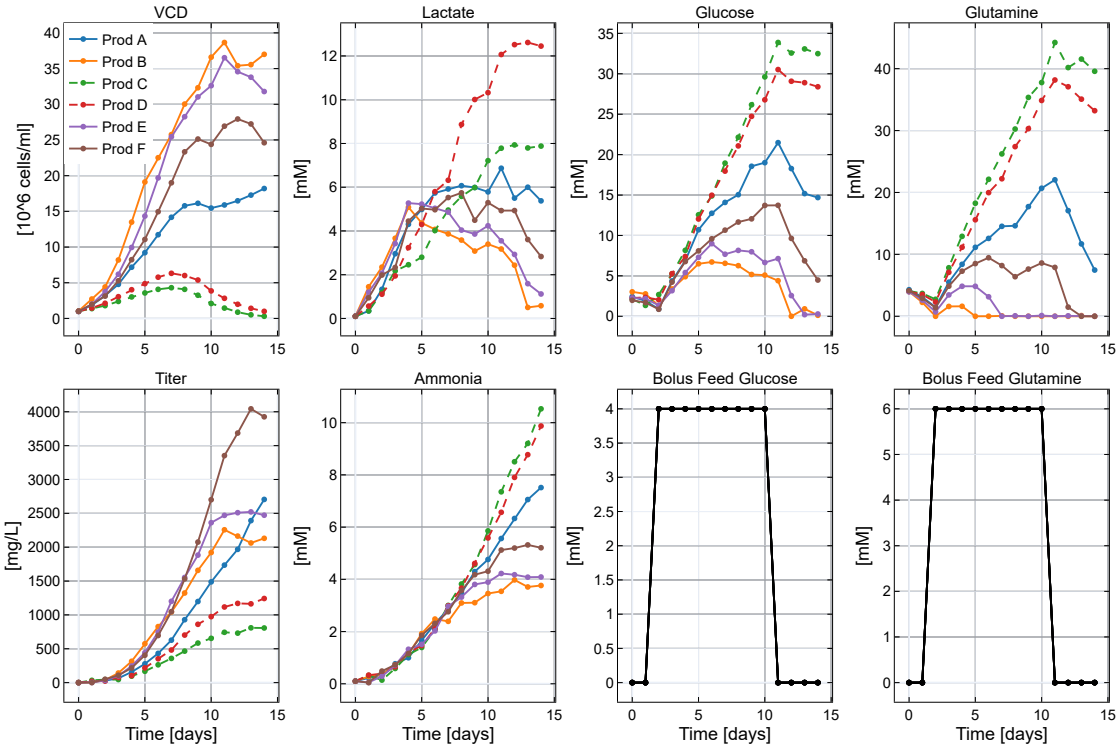

Figure 3: Inter-product variability for identical initial process conditions.
All processes of the six products were simulated using identical initial process conditions and product-specific kinetic parameters (Table S1). In particular, processes of products C and D mimic cell lines without lactate consumption (dashed lines). A relative measurement error of 2% along with an absolute error of one standard deviation was added to all simulations. The markers represent the daily measurements, which are used to train the model. It becomes evident that the processes of the six different products have different behaviour when exposed to the same process conditions.

Time series data for 14 days was simulated, assuming daily measurements of VCD, glucose, glutamine, ammonia, lactate and the product titer as indicated by the markers (and as common in industry). All products were simulated with the same identical conditions stated at Table S1, with a measurement error according to Eq 9. Fig 3 shows the high variance between the different products in simulation, introduced by different parameter settings in the *in silico* model. Most prominently, the ability to consume lactate can

be seen for cases A, B, E and F, while lactate is accumulating in cases C and D (dashed lines). Among the lactate-consuming products, the cases of B and E show higher degree of similarity than other products, which can be seen from the similar trajectories in Fig 3. However, VCD and titer significantly differ for the same process conditions, thus posing a suitable challenge for meta learning.

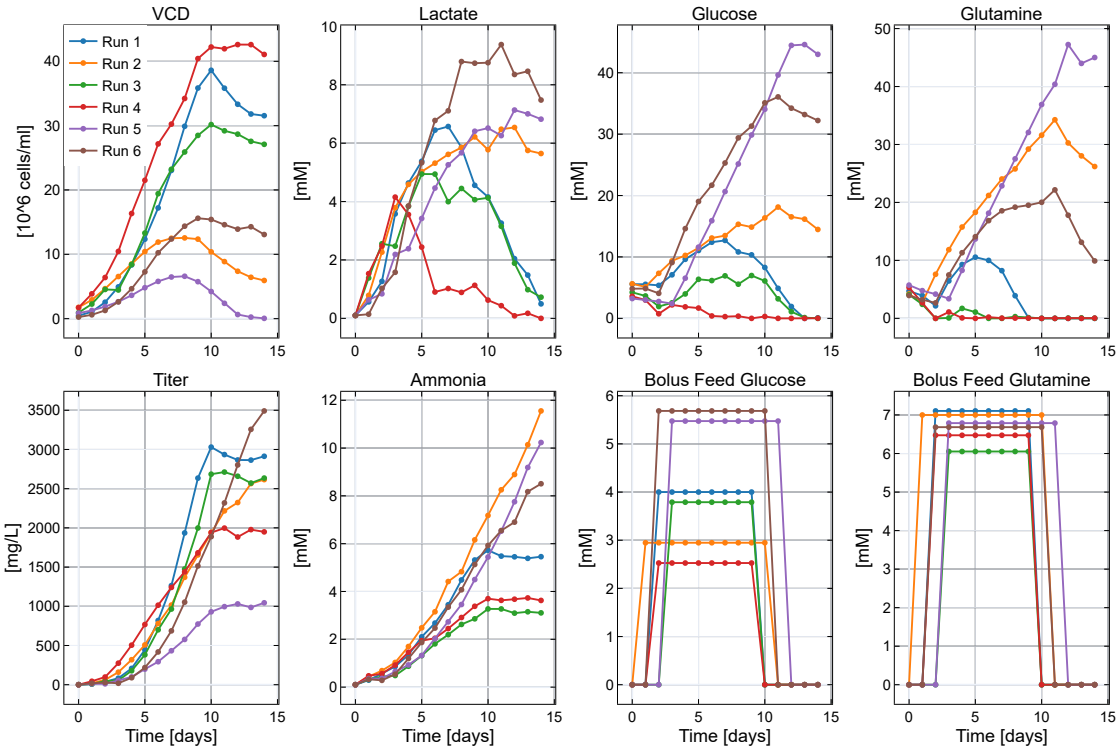

Figure 4: Intra-product variability for product F.
Six exemplary runs are shown to demonstrate the high variability of process conditions within the chosen design space. The data sets thus mimic a process development campaign in which the optimal conditions are investigated in a broad range without prior knowledge. A relative measurement error of 2% along with an absolute error of one standard deviation was added to all simulations. The markers represent the daily measurements, which are used to train the model.

Beside inter-product variability, the data set used in this paper also shows high intra-product variability when exposed to different process conditions, which is shown in Fig 4 for product F. Similarly to the inter-product variability, this characteristic is important to mimic real-world application scenarios with noisy measurements in bioreactors. Interesting non-linear behaviour can especially be seen for lactate and glutamine. While being used as a challenging benchmark for training of Gaussian processes in this study, the data sets are published in a dedicated Zenodo repository (Helleckes et al., 2024), thus being available for comparison in future methodological papers on knowledge transfer and design of experiments.

3.2 Comparison of models for rate prediction

Three different GPs approaches are compared regarding their prediction of the effective production rate $\tilde{R}$: a local, OHE and meta-learning approach. The production rate predicted by each approach is then used to

update concentrations in the stepwise model as described in Section 2.1. We shuffle through all possible combinations of the six cell lines, simulating each of them as the new product (which the capability of knowledge transfer is evaluated on), while utilising the other five as historical data sets, similar to the approach presented by Hutter et al., 2021. As a first benchmark, the GP is trained on a varying number of experiments of the new cell line alone, referred to as local model. In the second approach, we evaluate the OHE model described in Section 2.2.3 and train it on data from the new product plus additional twenty experiments from each of the other five historical products. Finally, we train a meta learning model using the PACOH approach described in Section 2.2.2, which uses several GPs as meta learner on the historical 20 experiments of each product. After meta training, the hyperposterior of the kernel parameters can be used to initialise a new GP for the novel product. As presented by Rothfuss et al., 2021, hyperparameters influence the meta learning performance for the PACOH approach. Similar to their approach, we thus performed hyperparameter optimisation with the `HyperOpt` Python package and used the determined hyperparameters to train the meta learning model.

Model performance is evaluated by calculating the relative RMSEs on a test data set with 100 experiments, using the predicted and simulated effective rates (Section 2.3.4). The available data from the new product is varied between two, four, eight and 20 experiments. To avoid bias by the choice of subsets in the training data, the selection was shuffled with 10 different seeds. Exemplary results for products A and C are shown as boxplots of RMSEs in Fig 5, the results for the remaining products can be found in Fig S2. The two examples were chosen since for product A, a large amount of data of processes that behave similarly (products B, E and F with lactate consumption) is provided in the corresponding training data set. For product C (without lactate consumption), a far smaller body of similar experiments is available in the training data set with the same size (only product D).

For product A (Fig 5a), it can be seen that both knowledge transfer approaches clearly outperform the local model (blue) trained only on the data from the new product. This observation is coherent with previous findings (Hutter et al., 2021; Kay et al., 2023). As also expected (and in line with previous findings such as Polak et al., 2024), using an increasing number of experiments from the new product in training is enhancing the model performance, which can be seen by the improved relative RMSEs for the predicted effective rates across various features. While for some variables such as VCD or titer, OHE (orange) outperforms the meta learning (green), the opposite is true for glucose and lactate. Overall, the RMSEs are well below 0.4, indicating a good performance for both OHE and meta learning. The boxplots in Fig 5a also show the variability in model performance, i.e. the spread of the RMSEs for the ten different subsets of two, four, and eight experiments that were randomly drawn from the training data. Here, variability is much higher for local models compared to both OHE and meta learning. For example, the RMSEs regarding VCD with four experiments are in the range from 0.5 to 0.95 in case of a local model, but only spread between 0.28 to 0.35 and 0.25 to 0.31 for the OHE and meta learning model, respectively. These results indicate that the knowledge transfer approaches are more robust towards the choice of designs for model calibration. Interestingly, while a local model with 20 experiments also leads to low RMSEs below 0.4, the knowledge transfer approaches still outperform it, indicating that similarities from historical products strongly enhance the model performance.

Regarding product C (Fig 5b), which is a case where a cell line without the lactate consumption was simulated, the observations are different. Here, meta learning is outperforming OHE for five of six compounds, namely VCD, glutamine, ammonia, lactate and titer predictions. Especially for VCD, glutamine, ammonia and lactate, the meta learning yields test set errors that are reduced by 25–50% compared to those obtained

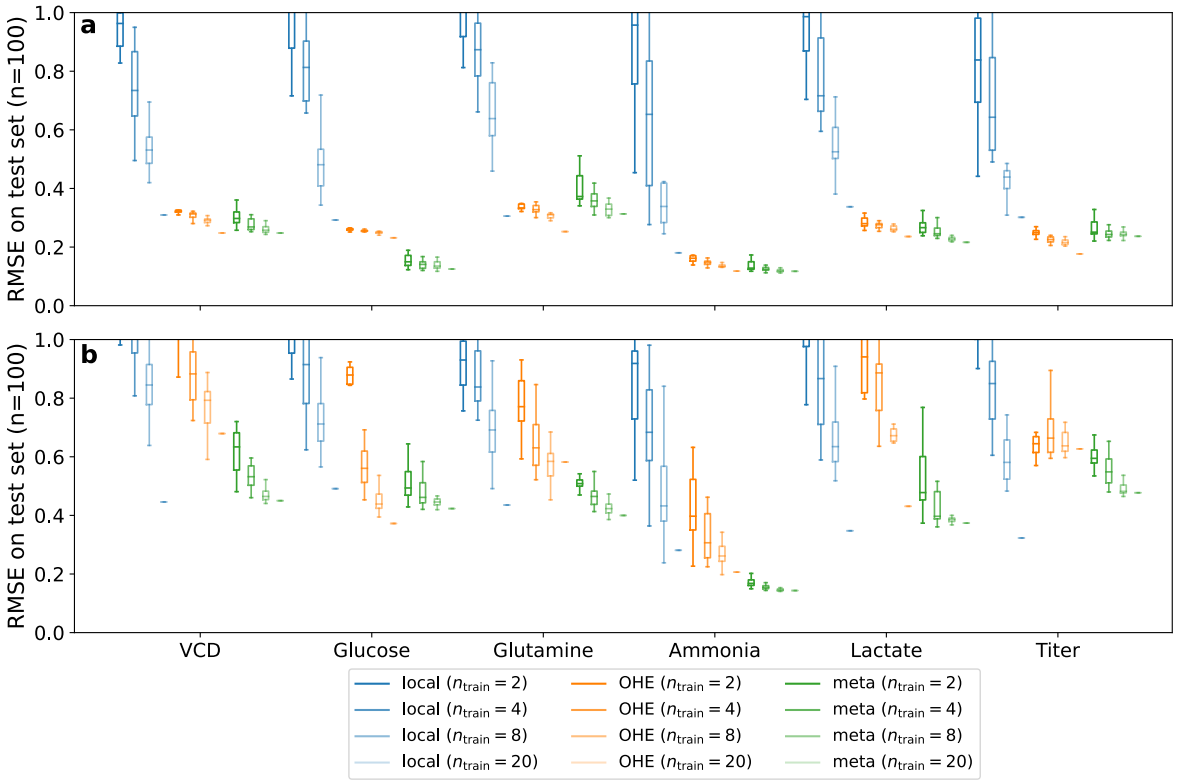

Figure 5: Prediction of rates for **a**: product A and **b**: product C. Local models (blue) are compared to OHE models (orange) and meta learning models (green), i.e. two different kinds of knowledge transfer models. Single GP instead of ensembles are compared to assess model performance on the learning problem directly.

with the OHE for two or four runs of the new product in training data. In case of glucose, test set errors are at comparable levels. The local model with 20 runs performs about equally well as meta learning for VCD, glucose, glutamine, ammonium and lactate, but significantly better for titer, indicating that the prediction quality for this product does not benefit as much from the historical data. This makes sense since fewer runs with similar process behaviour are present in the training data set, namely those of products C and D, simulating the case of missing lactate consumption. Since only products C and D simulate the case of missing lactate consumption in the cell line, these results indicate that knowledge transfer works best when similarity is high, which is to be expected. For heterogeneous data sets, meta learning outperforms OHE, for some compounds like ammonia even halving the test set error, which indicates a promising direction to improve GP regression models. This is in line with previous findings for transfer and meta learning with neural networks trained on heterogeneous data sets Félix et al., 2016; Tseng et al., 2021. As for product A, variability of RMSEs is lower for the knowledge transfer approaches, particularly for meta learning in comparison to local models. These results demonstrate the robustness towards the choice of designs for model calibration, which is important to reduce the number of required experiment during process development.

Generally, we observed that knowledge transfer methods outperform the local models. In the following, we therefore seek to understand whether the design of the experiments used to train the model can systematically improve the model performance regarding prediction of the novel product. Due to the comparable model performance and lower computation time compared to meta learning, the OHE model was used as a

benchmark for the following case studies of calibration design. However, it can be expected that the proposed approach will deliver similar results independent of the chosen knowledge transfer method.

3.3 Calibration design for parallel experiments

Related to the previously assessed methods for knowledge transfer, one can seek to increase the insight into process behaviour from the beginning of the process development activities. In particular, runs can be designed to calibrate knowledge transfer models to data of the new product in a systematic way, rather than just evaluating the process behaviour experimentally. We refer to this task as *calibration design*, which has the potential to reduce the number of experiments required for process model calibration and hence streamline early to mid-stage process development activities significantly (Section 2.3).

In the following, the results obtained with the customised objective are discussed using two case studies, which contain different combinations of similar and dissimilar historical data sets to evaluate the calibration design. For benchmarking, we compare the results to a process model that was trained on designs sampled by LHS, which was previously evaluated by Stosch, 2018 to be a well-suited excitation design for hybrid models. To obtain a feasible subset for optimisation, we chose four of our six simulated products and a subset of seven design factors that determine the calibration design (Section 2.4). In the first case study, all four products A, B, E and F represent lactate-consuming cell lines. Within this case, the processes of products B and E were designed to be similar to each other while A and B differ stronger in their dynamics (see also Section 3.1). Since all products were simulated as cell lines with lactate consumption, we chose a OHE model as the knowledge transfer model for calibration design, which is in line with the findings in Section 3.2 that OHE and meta learning perform equally well on these products regarding the RMSEs of model predictions.

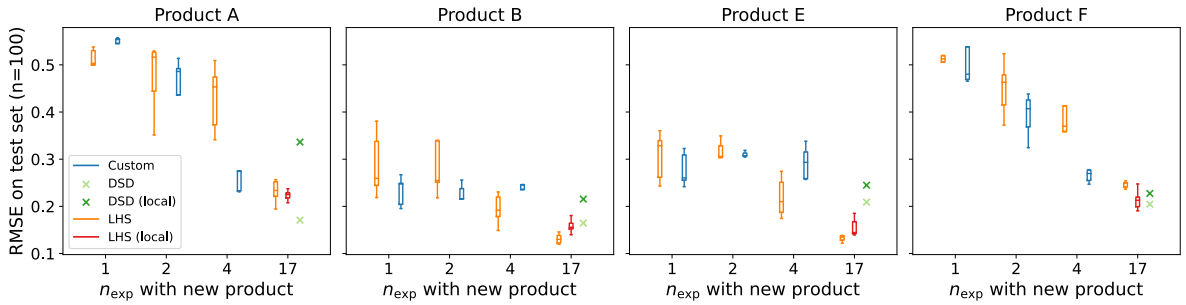

Figure 6: First case study. Relative RMSEs on test set for titer utilising a OHE model trained on new data obtained through various design paradigms: calibration design with custom objective (blue), definitive screening designs (green) and LHS designs (orange, red). OHE models were initially trained on historical 20 experiments from the other three products. Calibration and LHS designs were then performed to identify experiments to retrain the model to the novel product using one, two and four new experiments. All four chosen products mimic cell lines that can consume lactate. For 17 experiments, which is the number of experiments required for DSD with seven factors, we also trained local GP model on data from the new product alone. The weight between distance and objective in Eq 6 was chosen to be $\alpha = 0.0001$ so that both terms are in the same order of magnitude and have approximately equal weight.

The results for the calibration design of one, two and four experiments using the customised objective (Sec-

tion 2.3.2) or LHS are shown in Fig 6. More precisely, initial OHE models were retrained with experiments of calibration design to improve the model performance, which is benchmarked by the relative RMSEs of titer on test data sets of the new product. This measure was chosen since accurate prediction of process variables increases process understanding and is important for process characterisation, e.g. in the context of QbD. We shuffle through all possible combinations of historical data sets within the the case study for better generalisation.

Fig 6 shows that the relative RMSEs of titer on a test data set with 100 experiments is lower for products B and E as compared to products A and F. Since the OHE model used to evaluate the objective for calibration design is based on the other three products and B and E share a greater similarity, these lower RMSE values are to be expected. The customised objective based on dissimilarities of historical products outperforms a LHS calibration design for the dissimilar products A and F, especially when four experiments are designed. This indicates that the combination of the customised objective and the distance works well to spread experiments across the design space; at the same time, it optimises for those designs which reveal differences in process dynamics well. For product B and E, RMSE values lower than 0.4 can already be observed when using only one experiment for calibration design, which shows that the initial OHE has a high model performance due to the high similarity of historical products.

Subsequently, we tested the proposed calibration design strategy on two products with lactate consumption (A and B) and two without (C and D). The results are shown in Fig 7.

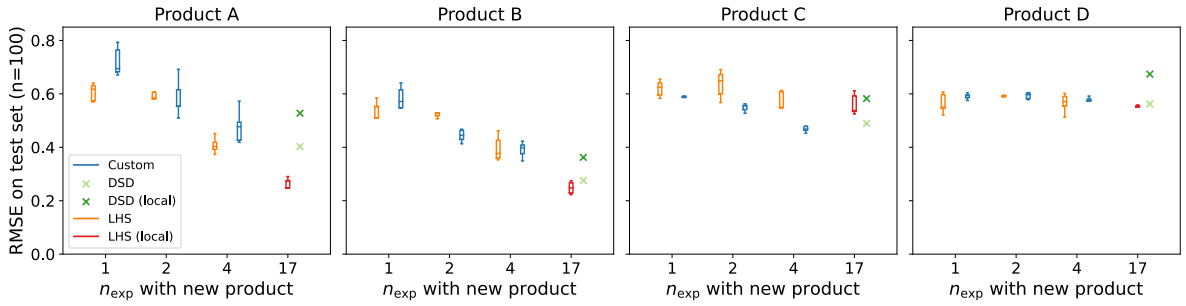

Figure 7: Second case study. Relative RMSEs on test set for titer of a OHE model utilising calibration design with custom objective (blue) and LHS designs (orange). OHE models were initially trained on historical 20 experiments from the other three products. In this case study, two products with lactate consumption (A and B) and two without lactate consumption (C and D) were chosen. Calibration design was performed to identify experiments to calibrate the model to the novel product. For one, two and four experiments, all data shows the result for OHE models, i.e. knowledge transfer approaches. For 17 experiments, which is the number of experiments required for DSD with seven factors, we also trained a local model on data from the new product alone, both for LHS (red) and DSD (dark green). The weight between distance and objective in Eq 6 was chosen to be $\alpha = 0.0001$ so that both terms are in the same order of magnitude and have approximately equal weight.

It can be seen that the relative RMSEs are generally greater than observed for the previous case, which is explained by the fact that a OHE model works better if data is more similar. Furthermore, it can be observed that LHS and the customised objective for calibration design result in models with similar model performance, i.e. RMSEs. Compared to the DSD, a well-established strategy for design of experiments during early process development, a similar or even better model performance is achieved already with four experiments, whereas

a DSD with seven design factors generates 17 experiments. The benefit of knowledge transfer models can also be seen here as even with 17 experiments in the training data, a OHE model with DSD (light green) outperforms a local model (dark green). However, LHS is superior to DSD for sampling 17 experiments, since a local LHS model without data from the historical data sets outperforms even the OHE-DSD model in three out of four cases and performs similarly well for the fourth case (product C). This is in line with observations that LHS designs are better suited for machine-learning model development than full or fractional factorial designs (or variations thereof), which investigate the edges of the process parameter space Arboretti et al., 2022. Most interestingly in the case of only four experiments for training, the customised objective for calibration design leads to similar (one out of four products) or better performance (three out of four products) compared to local DSD. Compared to LHS for one, two or four experiments, the test set performance for the models resulting from the customised objective is similar in most cases and better for product C.

In conclusion, calibration design is particularly useful for homogeneous data sets as shown in the first case study, but still performs well in case of heterogeneous historical data sets. However, the performance on heterogeneous historical data sets might improve with a different choice of knowledge transfer method, as it was seen before that meta learning works better than OHE for those cases. Nevertheless, these results outline the potential to exploit knowledge transfer in calibration design for process campaigns, even if historical training data is more dissimilar to the product of interest.

3.4 How could calibration designs change process development workflows?

For a CHO platform process, the first process development activities typically comprise clone selection and process condition screening (Kelley, 2020; Xu et al., 2022)[1]. Hence, for a new product one (or a limited number of) cell clone(s) are picked from a pre-selected number of clones (which were screened for performance after clone creation). This is typically accomplished by performing a few process runs at small scale using the "standard" platform process conditions or variations thereof to select the clone. Subsequently, few process runs in which the platform process conditions are slightly varied are performed to assess whether these conditions robustly deliver sufficient product for clinical trial phase 1. Variations in the process conditions are either designed based on experience (sophisticated guessing) or using screening designs, e.g. DSD.

The process development activities after a successful clinical trial phase 1 are typically focused on improving titer (or economic targets) and perhaps product quality before scale-up to larger scale (potentially after scale-down)[2]. These activities are usually supported through process models. For the development of these models, a series of process runs are executed, referred to as excitation design (De Luca et al., 2023; Ferreira et al., 2014; Huang et al., 2023). Traditionally, the excitation design is either determined by methods of DoE such as a full or fractional factorial designs (Freier et al., 2016; Polak et al., 2024), or space-filling designs

[1]Early stage process development typically has the goal to provide material for clinical phase 1 trials. The idea is to go fast into clinic and limit the investment into process development activities (as the chances of success are relatively low), while at the same time keeping the door open to optimise the process later (Farid et al., 2020). Dependent on the trade-off chosen by each company, the number of runs might range from tens to several tens of runs. However, the resulting process conditions selected to produce the material for the trials are typically sub-optimal, in the sense that the product titers could be increased significantly by further optimisation of the process conditions. Another pertinent question is whether the clone with the "best" performance at larger scale as been selected (Li et al., 2010).

[2]While moving towards late-stage development, process characterisation and qualification activities are executed to gather evidence of the process behaviour for the preparation of the process control strategy and process performance qualification (Xu et al., 2022). Depending on the product and platform, the average number of process development runs is in the tens in early stages (not considering clone selection), in the hundreds mid-stage and in the tens for later stages

such as LHS (Bader et al., 2023), Doehlert Designs (Pinto et al., 2019) or Sobol sequences (Siedentop et al., 2023). Irrespective of the design, these activities typically require the execution of several tens of runs before moving to larger scales. However, industrial process development campaigns might share certain similarities, in particular if clones are derived from the same cell line for protein production, or if similar design spaces around the standard platform process conditions are explored during optimisation.
Considering that four runs in this study suffice to gain a level of process understanding that is similar or superior to that obtained with 17 runs using the traditional approach (i.e. a local model trained with data from DSD), the magnitude of the opportunity should provide an incentive to re-think development workflows and fuel further research in this direction. Imagining that models could also be used to transfer much of the knowledge across the scales, the observed factor 4 reduction might be at the lower end. This opens opportunities to use the existing experimental capacity to investigate more clones and handle more projects at the same time, rather than characterising the process.

3.5 Limitations and future work

In this work, we explored the use of knowledge transfer models to improve prediction of pharmaceutical processes and suggest calibration designs to adapt the process for a new product. While meta learning showed comparable model performance for those products which exhibit similarity in the training data, it showed significant potential for cases with few similar products in the historical training data, where the test set RMSEs could be reduced compared to OHE and local models. This should be further investigated with heterogeneous data sets from industrial applications to assess the benefit for real-world applications. Moreover, approaches for similarity assessment between data sets might be exploited to further understand for which use-cases meta learning can be beneficial.
Regarding the calibration design, this study showed high potential of process models utilising knowledge transfer to characterise pharmaceutical processes. However, some aspects of this novel procedure should still be addressed. Firstly, local optimisers were used throughout this study, thus not guaranteeing convergence to a global optimum. In addition to the proposed objective, further objectives for calibration design based on dissimilarity could be investigated, for example by varying the weight of uncertainty in the prediction of historical data sets or introducing weights for individual historical data sets based on their similarity to the new product. Moreover, we focused on final titer as an important performance indicator of a process and work could thus be extended to account for the entire evolution and other variables. In the future, the customised objective could also be extended to a combination of various features, most importantly including VCD as a relevant influence. While the different case studies indicate generalisability of the calibration design approach, further investigation with real-world data and heterogeneous data sets are needed to investigate the versatility of the approach in industrial practice.

4 Conclusion

In this work, we compared two knowledge transfer modelling methods with local models developed on data of one product only, by using data of a simulated mammalian cell cultivation process, typically used for monoclonal antibody production in the pharmaceutical industry. Subsequently, we adopted the OHE knowledge transfer model to identify process conditions that are particularly suited to calibrate the model to data from a new product. To achieve this task, which we coined *calibration design*, we proposed a customised objective

function. The objective function favours those process conditions that lead to the most dissimilar predicted titers in the historical data sets, reasoning that those conditions are most informative to train a knowledge transfer model.

We observed that in a case study with six simulated products, both knowledge transfer methods, i.e. meta learning and OHE models, outperformed local models, especially if only few experiments of the new product were available. In particular, for the case that historical data exhibits more similarity to the new product, the test set errors of the knowledge transfer models for VCD, glucose, glutamine, ammonia, lactate and titer were at least half that of the local models, irrespective of whether they were trained with two, four, or eight experiments of the new product. Even for the cases with 20 experiments in training, knowledge transfer models showed a lower test set error than the local models.

In case the historical data is more heterogeneous, the test set error of meta learning models was still reduced by about 25–50% compared to the errors for local models for two, four and eight experiments and all compounds except titer. In contrast, OHE models performed more similarly to local models for all compounds. In case of 20 runs, the model performance of the local models matched that of both knowledge transfer models, except for titer, where the test set error was lower for the local model. Further, we observed that meta learning boosts the model performance of GP hybrid models for heterogeneous data sets. Specifically, for five of six compounds, the test set errors were reduced by approximately 25% compared to the errors of the OHE models. For future benchmarking and comparison to methods proposed by others, we provide the simulated data sets in a dedicated repository.

For the calibration design, we observed that the OHE models trained on experiments obtained through LHS or calibration design with the new customised optimisation objective yield similar test set RMSEs in the final predicted titer when compared to common approaches such as DSDs. However, four times fewer experiments were needed for calibration designs utilising the LHS or customised objective paradigms, making them a promising alternative. The customised objective for optimisation led to models that perform similarly or better than models based on LHS designs, both on homogeneous and heterogeneous data sets. Specifically, three out of four product cases showed a better performance with the customised objective for the homogeneous data set, while a better performance of the objective was still observed for one out of four products in the heterogeneous case. Importantly, even in cases where model performance with the customised objective was not improved, test set errors remained in the same range. The proposed objective function thus presents an interesting alternative with low risk for inferior model performance. While further studies on industrial data are needed to evaluate the practical application of this novel approach, it shows a new direction for process development workflows.

In the future, our customised objective function could be further extended to more features than only titer. In addition, the combination of meta learning and calibration design to boost the predictive performance should be investigated further. It could also be tested in scenarios where more process knowledge is integrated, i.e. using hybrid GP models that consider first principles. Since calibration design and knowledge transfer approaches seem to be able to reduce the number of experiments for process development significantly, they are of high research interest and could possibly lead to significant savings in biopharma and perhaps other bioprocess industries.

Abbreviations

ARD automatic relevance determination

CHO chinese hamster ovary

DoE Design of Experiments

DSD definitive screening design

GP Gaussian process

LHS latin hypercube sampling

ML machine learning

OHE one-hot enconding

PACOH PAC-optimal hyper-posterior

QbD Quality by Design

RMSE root-mean-square error

SE squared exponential

VCD viable cell density

Author's contributions

LH performed data analyses, prepared the figures and wrote the manuscript, with contributions from CW, JP and MVS. CW, JP, GG, AB and MVS contributed to project administration and supervision. All authors were involved in regular project discussions, read and approved the final manuscript.

Acknowledgements

This work was performed as part of the Helmholtz School for Data Science in Life, Earth and Energy (HDS-LEE).

Data availability

The data that was used for benchmarking throughout this study is available on Zenodo (DOI: 10.5281/zenodo.10630629).

Funding

This study was performed during a research stay funded by the German Academic Exchange Service (DAAD).

Conflict of interest

CW, JP, AB and MVS were employees of DataHow AG at the time of the study.

References

Arboretti, R., Ceccato, R., Pegoraro, L., & Salmaso, L. (2022). Design of experiments and machine learning for product innovation: A systematic literature review. *Quality and Reliability Engineering International*, *38*(2), 1131–1156. https://doi.org/10.1002/qre.3025

Ashenden, S. K., Bartosik, A., Agapow, P.-M., & Semenova, E. (2021, January). Chapter 2 - Introduction to artificial intelligence and machine learning. In S. K. Ashenden (Ed.), *The Era of Artificial Intelligence, Machine Learning, and Data Science in the Pharmaceutical Industry* (pp. 15–26). Academic Press. https://doi.org/10.1016/B978-0-12-820045-2.00003-9

Bader, J., Narayanan, H., Arosio, P., & Leroux, J.-C. (2023). Improving extracellular vesicles production through a Bayesian optimization-based experimental design. *European Journal of Pharmaceutics and Biopharmaceutics*, *182*, 103–114. https://doi.org/10.1016/j.ejpb.2022.12.004

Bareither, R., & Pollard, D. (2011). A review of advanced small-scale parallel bioreactor technology for accelerated process development: Current state and future need. *Biotechnology Progress*, *27*(1), 2–14. https://doi.org/10.1002/btpr.522

Beg, S., & Swain, S. (2021). Introduction to the Application of Experimental Designs in Pharmaceutical Product Development. In S. Beg (Ed.), *Design of Experiments for Pharmaceutical Product Development: Volume II : Applications and Practical Case studies* (pp. 1–17). Springer. https://doi.org/10.1007/978-981-33-4351-1_1

Bergstra, J., Yamins, D., & Cox, D. (2013). Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures. *Proceedings of the 30th International Conference on Machine Learning*, 115–123. Retrieved November 15, 2023, from https://proceedings.mlr.press/v28/bergstra13.html

Bradl, H., Bechmann, J., Mueller, M., Schulz, P., Wucherpfennig, T., & Greulich, B. (2016). Platform Approach Speeds Process Development. *BioPharm International*, *29*(4), 20–25. Retrieved August 29, 2023, from https://www.biopharminternational.com/view/platform-approach-speeds-process-development

Chevalier, C., & Ginsbourger, D. (2013). Fast Computation of the Multi-Points Expected Improvement with Applications in Batch Selection. In G. Nicosia & P. Pardalos (Eds.), *Learning and Intelligent Optimization* (pp. 59–69). Springer. https://doi.org/10.1007/978-3-642-44973-4_7

Craven, S., Shirsat, N., Whelan, J., & Glennon, B. (2013). Process model comparison and transferability across bioreactor scales and modes of operation for a mammalian cell bioprocess. *Biotechnology Progress*, *29*(1), 186–196. https://doi.org/10.1002/btpr.1664

Cruz-Bournazou, M. N., Narayanan, H., Fagnani, A., & Butté, A. (2022). Hybrid Gaussian Process Models for continuous time series in bolus fed-batch cultures. *IFAC-PapersOnLine*, *55*(7), 204–209. https://doi.org/10.1016/j.ifacol.2022.07.445

Darmont, J., Novikov, B., Wrembel, R., & Bellatreche, L. (2022). Advances on Data Management and Information Systems. *Information Systems Frontiers*, *24*(1), 1–10. https://doi.org/10.1007/s10796-021-10235-4

De Luca, R., Costa, G., Narayanan, H., Wirnsperger, C., Cruz Bournazou, M. N., Butte, A., & von Stosch, M. (2023). Comparison of strategies for iterative model-based upstream bioprocess development with single and parallel reactor set-ups. *Biochemical Engineering Journal*, *191*, 108813. https://doi.org/10.1016/j.bej.2023.108813

Deisenroth, M. P., Rasmussen, C. E., & Peters, J. (2009). Gaussian process dynamic programming. *Neurocomputing*, *72*(7), 1508–1524. https://doi.org/10.1016/j.neucom.2008.12.019

Dodds, M., Roberts, J., & Finrow, B. (2022). Improving combination drug trials using 'definitive screening designs'. *Nature Biotechnology*, *40*(12), 1720–1721. https://doi.org/10.1038/s41587-022-01521-w

Farid, S. S., Baron, M., Stamatis, C., Nie, W., & Coffman, J. (2020). Benchmarking biopharmaceutical process development and manufacturing cost contributions to r&d. *mAbs*, *12*(1), 1754999. https://doi.org/10.1080/19420862.2020.1754999

Félix, C., Soares, C., & Jorge, A. (2016). Can Metalearning Be Applied to Transfer on Heterogeneous Datasets? In F. Martínez-Álvarez, A. Troncoso, H. Quintián, & E. Corchado (Eds.), *Hybrid Artificial Intelligent Systems* (pp. 332–343). Springer International Publishing. https://doi.org/10.1007/978-3-319-32034-2_28

Ferreira, A. R., Dias, J. M., Von Stosch, M., Clemente, J., Cunha, A. E., & Oliveira, R. (2014). Fast development of Pichia pastoris GS115 Mut+ cultures employing batch-to-batch control and hybrid semi-parametric modeling. *Bioprocess and Biosystems Engineering*, *37*(4), 629–639. https://doi.org/10.1007/S00449-013-1029-9/FIGURES/7

Fink, M., Cserjan-Puschmann, M., Reinisch, D., & Striedner, G. (2021). High-throughput microbioreactor provides a capable tool for early stage bioprocess development. *Scientific Reports*, *11*(1), 2056. https://doi.org/10.1038/s41598-021-81633-6

Freier, L., Hemmerich, J., Schöler, K., Wiechert, W., Oldiges, M., & von Lieres, E. (2016). Framework for Kriging-based iterative experimental analysis and design: Optimization of secretory protein production in Corynebacterium glutamicum. *Engineering in Life Sciences*, *16*(6), 538–549. https://doi.org/10.1002/elsc.201500171

Gelman, A., & Vehtari, A. (2021). What are the Most Important Statistical Ideas of the Past 50 Years? *Journal of the American Statistical Association*, *116*(536), 2087–2097. https://doi.org/10.1080/01621459.2021.1938081

Ginsbourger, D., Le Riche, R., & Carraro, L. (2010). Kriging Is Well-Suited to Parallelize Optimization. In Y. Tenne & C.-K. Goh (Eds.), *Computational Intelligence in Expensive Optimization Problems* (pp. 131–162). Springer. https://doi.org/10.1007/978-3-642-10701-6_6

González, L. D., & Zavala, V. M. (2023). New paradigms for exploiting parallel experiments in Bayesian optimization. *Computers & Chemical Engineering*, *170*, 108110. https://doi.org/10.1016/j.compchemeng.2022.108110

Greenhill, S., Rana, S., Gupta, S., Vellanki, P., & Venkatesh, S. (2020). Bayesian Optimization for Adaptive Experimental Design: A Review. *IEEE Access*, *8*, 13937–13948. https://doi.org/10.1109/ACCESS.2020.2966228

Helleckes, L., Polak, J., Wirnsperger, C., & von Stosch, M. (2024, February). *Benchmarking in silico dataset for mammalian cell cultures* (Version v.1.0.0). Zenodo. https://doi.org/10.5281/zenodo.10630629

Hemmerich, J., Noack, S., Wiechert, W., & Oldiges, M. (2018). Microbioreactor Systems for Accelerated Bioprocess Development. *Biotechnology Journal*, *13*(4), 1700141. https://doi.org/10.1002/biot.201700141

Hospedales, T., Antoniou, A., Micaelli, P., & Storkey, A. (2022). Meta-Learning in Neural Networks: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(9), 5149–5169. https://doi.org/10.1109/TPAMI.2021.3079209

Huang, R., Fogelquist, J., & Lin, X. (2023). Reinforcement Learning of Optimal Input Excitation for Parameter Estimation With Application to Li-Ion Battery. *IEEE Transactions on Industrial Informatics*, *19*(11), 11160–11170. https://doi.org/10.1109/TII.2023.3244342

Hutter, C., von Stosch, M., Cruz Bournazou, M. N., & Butté, A. (2021). Knowledge transfer across cell lines using hybrid Gaussian process models with entity embedding vectors. *Biotechnology and Bioengineering*, *118*(11), 4389–4401. https://doi.org/10.1002/bit.27907

Jin, X., Wah, B. W., Cheng, X., & Wang, Y. (2015). Significance and Challenges of Big Data Research. *Big Data Research*, *2*(2), 59–64. https://doi.org/10.1016/j.bdr.2015.01.006

Jones, B., & Nachtsheim, C. J. (2011). A Class of Three-Level Designs for Definitive Screening in the Presence of Second-Order Effects. *Journal of Quality Technology*, *43*(1), 1–15. https://doi.org/10.1080/00224065.2011.11917841

Kay, S., Kay, H., Rogers, A. W., & Zhang, D. (2023). Integrating hybrid modelling and transfer learning for new bioprocess predictive modelling. In A. C. Kokossis, M. C. Georgiadis, & E. Pistikopoulos (Eds.), *33rd european symposium on computer aided process engineering* (pp. 2595–2600, Vol. 52). Elsevier. https://doi.org/10.1016/B978-0-443-15274-0.50412-1

Kelley, B. (2020). Developing therapeutic monoclonal antibodies at pandemic pace. *Nature Biotechnology*, *38*(5), 540–545. https://doi.org/10.1038/s41587-020-0512-5

Kim, B. J., Diao, J., & Shuler, M. L. (2012). Mini-scale bioprocessing systems for highly parallel animal cell cultures. *Biotechnology Progress*, *28*(3), 595–607. https://doi.org/10.1002/btpr.1554

Kocijan, J. (2016). Introduction. In J. Kocijan (Ed.), *Modelling and Control of Dynamic Systems Using Gaussian Process Models* (pp. 1–20). Springer International Publishing. https://doi.org/10.1007/978-3-319-21021-6_1

Li, F., Vijayasankaran, N., Shen, A. Y., Kiss, R., & Amanullah, A. (2010). Cell culture processes for monoclonal antibody production. *mAbs*, *2*(5), 466–479. https://doi.org/10.4161/mabs.2.5.12720

Liu, Q., & Wang, D. (2016). Stein Variational Gradient Descent: A General Purpose Bayesian Inference Algorithm. *Advances in Neural Information Processing Systems*, *29*. Retrieved November 15, 2023, from https://proceedings.neurips.cc/paper_files/paper/2016/hash/b3ba8f1bee1238a2f37603d90b58898d-Abstract.html

Narayanan, H., Behle, L., Luna, M. F., Sokolov, M., Guillén-Gosálbez, G., Morbidelli, M., & Butté, A. (2020a). Hybrid-EKF: Hybrid model coupled with extended Kalman filter for real-time monitoring and control of mammalian cell culture. *Biotechnology and Bioengineering*, *117*(9), 2703–2714. https://doi.org/10.1002/bit.27437

Narayanan, H., Luna, M. F., von Stosch, M., Cruz Bournazou, M. N., Polotti, G., Morbidelli, M., Butté, A., & Sokolov, M. (2020b). Bioprocessing in the Digital Age: The Role of Process Models. *Biotechnology Journal*, *15*(1), 1900172. https://doi.org/10.1002/biot.201900172

Narayanan, H., Sokolov, M., Morbidelli, M., & Butté, A. (2019). 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. https://doi.org/10.1002/bit.27097

Newcombe, A. R. (2014). The Evolution of Quality by Design (QbD) for Biologics. *PDA Journal of Pharmaceutical Science and Technology*, *68*(4), 320–322. https://doi.org/10.5731/pdajpst.2014.00989

Ongari, D. (2023, February). Definitive-screening-design: Definitive Screening Design. Retrieved November 20, 2023, from https://github.com/danieleongari/definitive_screening_design

Pinto, J., de Azevedo, C. R., Oliveira, R., & von Stosch, M. (2019). A bootstrap-aggregated hybrid semi-parametric modeling framework for bioprocess development. *Bioprocess and Biosystems Engineering*, *42*(11), 1853–1865. https://doi.org/10.1007/s00449-019-02181-y

Polak, J., Huang, Z., Sokolov, M., von Stosch, M., Butté, A., Hodgman, C. E., Borys, M., & Khetan, A. (2024). An innovative hybrid modeling approach for simultaneous prediction of cell culture process dynamics and product quality. *Biotechnology Journal*, *19*(3), 2300473. https://doi.org/10.1002/biot.202300473

Politis, S. N., Colombo, P., Colombo, G., & Rekkas, D. M. (2017). Design of experiments (DoE) in pharmaceutical development. *Drug Development and Industrial Pharmacy*, *43*(6), 889–901. https://doi.org/10.1080/03639045.2017.1291672

Rasmussen, C. E., & Williams, C. K. (2006). *Gaussian processes for machine learning* (Vol. 1). Springer.

Rathore, A. S., & Winkle, H. (2009). Quality by design for biopharmaceuticals. *Nature Biotechnology*, *27*(1), 26–34. https://doi.org/10.1038/nbt0109-26

Rehberger, B., Wodarczyk, C., Reichenbächer, B., Köhler, J., Weber, R., & Müller, D. (2013). Accelerating stable recombinant cell line development by targeted integration. *BMC Proceedings*, *7*(Suppl 6), P111. https://doi.org/10.1186/1753-6561-7-S6-P111

Rothfuss, J., Fortuin, V., Josifoski, M., & Krause, A. (2021, June). PACOH: Bayes-Optimal Meta-Learning with PAC-Guarantees. Retrieved August 24, 2023, from http://arxiv.org/abs/2002.05551

Sarker, I. H. (2021). Data Science and Analytics: An Overview from Data-Driven Smart Computing, Decision-Making and Applications Perspective. *Sn Computer Science*, *2*(5), 377. https://doi.org/10.1007/s42979-021-00765-8

Shcherbatyi, I., Kumar, M., Head, T., & Nahrstaedt, H. (n.d.). Parallel optimization — scikit-optimize 0.8.1 documentation. Retrieved November 16, 2023, from https://scikit-optimize.github.io/stable/auto_examples/parallel-optimization.html

Siedentop, R., Siska, M., Möller, N., Lanzrath, H., von Lieres, E., Lütz, S., & Rosenthal, K. (2023). Bayesian Optimization for an ATP-Regenerating In Vitro Enzyme Cascade. *Catalysts*, *13*(3), 468. https://doi.org/10.3390/catal13030468

Stosch, M. v. (2018). Hybrid Models and Experimental Design. In *Hybrid Modeling in Process Industries*. CRC Press.

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., & Liu, C. (2018). A Survey on Deep Transfer Learning. In V. Kůrková, Y. Manolopoulos, B. Hammer, L. Iliadis, & I. Maglogiannis (Eds.), *Artificial Neural Networks and Machine Learning – ICANN 2018* (pp. 270–279). Springer International Publishing. https://doi.org/10.1007/978-3-030-01424-7_27

ter Horst, J. P., Turimella, S. L., Metsers, F., & Zwiers, A. (2021). Implementation of Quality by Design (QbD) Principles in Regulatory Dossiers of Medicinal Products in the European Union (EU) Between 2014 and 2019. *Therapeutic Innovation & Regulatory Science*, *55*(3), 583–590. https://doi.org/10.1007/s43441-020-00254-9

Tseng, G., Kerner, H., Nakalembe, C., & Becker-Reshef, I. (2021). Learning to predict crop type from heterogeneous sparse labels using meta-learning. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1111–1120. https://doi.org/10.1109/CVPRW53098.2021.00122

Tulsyan, A., Khodabandehlou, H., Wang, T., Schorner, G., Coufal, M., & Undey, C. (2021). Spectroscopic models for real-time monitoring of cell culture processes using spatiotemporal just-in-time Gaussian processes. *AIChE Journal*, *67*(5), e17210. https://doi.org/10.1002/aic.17210

Upadhyay, R., Phlypo, R., Saini, R., & Liwicki, M. (2023, August). Sharing to learn and learning to share – Fitting together Meta-Learning, Multi-Task Learning, and Transfer Learning: A meta review. Retrieved August 24, 2023, from http://arxiv.org/abs/2111.12146

von Stosch, M., Portela, R. M., & Varsakelis, C. (2021). A roadmap to AI-driven in silico process development: Bioprocessing 4.0 in practice. *Current Opinion in Chemical Engineering*, *33*, 100692. https://doi.org/10.1016/j.coche.2021.100692

Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. *Journal of Big Data*, *3*(1), 9. https://doi.org/10.1186/s40537-016-0043-6

Xing, Z., Bishop, N., Leister, K., & 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. https://doi.org/10.1002/btpr.284

Xu, J., Ou, J., McHugh, K. P., Borys, M. C., & Khetan, A. (2022). Upstream cell culture process characterization and in-process control strategy development at pandemic speed. *mAbs*, *14*(1), 2060724. https://doi.org/10.1080/19420862.2022.2060724

Xu, J., Rehmann, M. S., Xu, M., Zheng, S., Hill, C., He, Q., Borys, M. C., & Li, Z. J. (2020). Development of an intensified fed-batch production platform with doubled titers using N-1 perfusion seed for cell culture manufacturing. *Bioresources and Bioprocessing*, *7*(1), 17. https://doi.org/10.1186/s40643-020-00304-y