Machine Learning based optimisation of bioprocesses for the production of therapeutic proteins

João Medeiros Garcia Alcântara

Thesis to obtain the Master of Science Degree in

Biological Engineering

Supervisors: Prof. Dr. José Monteiro Cardoso de Menezes
Prof. Dr. Massimo Morbidelli

Examination Committee:

Chairperson: Prof. Dr. Jorge Humberto Gomes Leitão
Supervisor: Prof. Dr. José Monteiro Cardoso de Menezes
Member of the Committee: Prof. Dr. Ana Luísa Nobre Fred

November 2018

Preface

The work presented in this thesis was performed at the Morbidelli Group, at the Institute for Chemical and Bioengineering of ETH Zürich (Zürich, Switzerland), during the period February – October 2018, under the supervision of Prof. Dr. Massimo Morbidelli, M.Sc. Fabian Feidl and M.Sc. Moritz Wolf. The thesis was co-supervised at Instituto Superior Técnico by Prof. José Cardoso de Menezes.

Acknowledgments

Firstly, I would like to thank Prof. Morbidelli for the amazing opportunity to join his research group at ETH Zürich and for allowing to me embrace such a challenging and promising project. I would to express my gratitude to the group as a whole, who received me in the best way possible and made the stay in Zürich much more interesting, both academically and otherwise.

I would like to thank my direct supervisors, Moritz Wolf and Fabian Feidl that, in the best way possible, challenged me with their creative thinking and innovative ideas so that the project could move forward and create promising results. In addition, I would like to thank Dr. Alessandro Butté for the constructive discussions on data analysis and to Sebastian Vogg for the help in execution of reference analytics.

I would also to express my gratitude to Andrea Müller for the support in the execution of the laboratory protocols in my first few months at ETH and to Caterina Ruggeri, whose great help allowed me to stay a few weekends outside of the laboratory. Furthermore, I would like to thank Fabio Boniolo and Matevz Podobnik for the camaraderie and for all the relevant discussions. I have also to thank Fabio, Matevz, Caterina, Francesca, Matteo, Niccólo and, most of all, Mariana, for all the friendship during these months, it wouldn't have been the same without you.

In addition, I would like to thank my family for all the support during these 23 years and for always being there.

Finally, I would to express my regards to Prof. José Cardoso Menezes for introducing me to the field of chemometrics and for the motivation and support, both before and during my thesis.

Thank you all.

Dankeschön.

Grazie mille.

Obrigado.

Resumo

No seio da indústria biofarmacêutica, o desenvolvimento de bioprocessos ainda se afigura uma das maiores fontes de despesa. A optimização de meios de cultura, em particular, é uma actividade altamente trabalhosa e demorada, apresentando-se como o domínio perfeito para a aplicação de técnicas de aprendizagem automática. A aplicabilidade de optimização Bayesiana (BO) foi testada para este propósito.

A utilização de um modelo proprietário *in-silico* de um processo semi-contínuo permitiu a comparação das diferentes variações deste algoritmo. A melhor, em modo sequencial, era constituída por amostragem por hipercubo latino, como inicialização, e melhoria esperada como função de aquisição. BO por batelada for implementada via penalização local. Comparando com a anterior, a perda foi considerada desprezável, dadas as vantagens: possibilidade de correr várias experiências em paralelo, economizando tempo e recursos.

Devido a estas vantagens, BO por batelada foi aplicada na optimização de um protocolo de alimentação de um processo semi-contínuo baseado numa linha celular *Chinese Hamster Ovary*, produzindo uma proteína derivada de um anticorpo monoclonal. No total, foram realizadas 31 experiências em *spin-tubes*: 16 experiências na fase de inicialização e 4 bateladas de 4 experiências cada uma. Apesar de um problema com o método analítico *at-line*, o título foi aumentado em 35%. Ao invés, se tivesse sido utilizado um delineamento *full factorial* com 3 níveis, seria necessário realizar 81 experiências e o aumento referido não seria garantido. A adequação do modelo de redução de escala utilizado (*spin-tubes* de 50mL) foi validado em reactores à escala laboratorial, com resultados comparáveis.

No futuro, a melhor estratégia de alimentação pode ser utilizada para desenvolver um meio para culturas em perfusão e o algoritmo pode ainda sofrer melhorias, tal como a incorporação de conhecimento determinístico e o uso de bateladas em sobreposição.

Palavras-chave: Aprendizagem automática, Optimização Bayesiana, CHO, Anticorpo Monoclonal, meio, redução de escala, delineamento experimental

Abstract

Bioprocess development still presents one of the biggest cost drivers within the biopharmaceutical industry. Media optimisation, in particular, is a highly laborious, time-consuming activity, and thus the perfect setting for the application of machine-learning techniques. In this work, the applicability of Bayesian optimisation (BO) was tested for this purpose.

The use of an in-house developed *in-silico* model of a fed-batch process allowed the comparison of different variations of this algorithm. The best, in sequential mode, was composed of Latin hypercube sampling as initialisation and expected improvement as acquisition function. Batch BO via local penalisation was implemented. In this approach, the loss was found not to be significant, considering the advantages: possibility to run multiple experiments in parallel, saving time and resources.

Due to its advantages, batch BO was applied for the optimisation of an existing fed-batch protocol based on a Chinese Hamster Ovary cell line and producing a mAb-derived protein. In total, 31 spin-tube experiments were performed: 16 were part of an initialisation batch, plus 4 batches of 4 experiments. Although coping with an at-line analytics problem, an overall titre increase of 35% was achieved. Considering, instead, a full factorial design with 3 levels, 81 experiments would be required, and the reported increase would not be guaranteed. The suitability of the scale-down model (50mL spin-tubes) was then validated in benchtop bioreactors with comparable results.

In future work, the best feeding strategy can be implemented for perfusion media design, and the algorithm could be further improved, through the incorporation of deterministic information and the use of overlapping batches.

Key-words: Machine learning, Bayesian optimisation, CHO, monoclonal antibody, media, scale-down, DoE.

Contents

List of Figures

List of Tables

Nomenclature

Latin Symbols

$h_k(x)$ – equality constraint

$g_j(x)$ – inequality constraint

$x_{i,L}$ – lower bound

$x_{i,U}$ – upper bound

$h(x)$ – basis function

R – Resolution (of a design)

L – Local Lipchitz constant

$f(x^+)$ – target objective value

$f(x)$ – objective function

i – vector index

k – Kernel/covariance function

m – mean function

$p(\cdot)$ – probability

x – feature vector

Greek Symbols

H_ν – Bessel function of ν order

Γ – Gamma function

Φ – Cumulative Distribution Function of the standard normal distribution

ξ – Error variance

υ – Smoothness parameter of the Matérn class of kernels

φ – Probability Density Function of the standard normal distribution; Local penaliser of an acquisition function

Abbreviations

AI – Artificial Intelligence

BO – Bayesian Optimisation

CDF - Cumulative Distribution Function

CHO – Chinese Hamster Ovary

DoE – Design of Experiments

EDTA - Ethylenediaminetetraacetic acid

EI – Expected Improvement

ES – Entropy Search

GA - Genetic Algorithm

GP – Gaussian Process

HPLC - High-performance liquid chromatography

LDH – Lactate dehydrogenase

LHCS – Latin Hypercube Sampling

mAb – monoclonal antibody

ML – Machine Learning

MLR – Multiple Linear Regression

MSX – Methionine sulfoximine

PBS - Phosphate-buffered saline

PDF - Probability Density Function

PTFE – Polytetrafluoroethylene

R&D – Research and Development

RMSEP – Root Mean Square Error in Prediction

ST – Spin-tube

TS – Thompson Sampling

UCB – Upper Confidence Bound

UV – Ultra Violet

VCD – Viable Cell Density

WD – Working Day

1. Introduction

1.1. Bioprocessing for the production of therapeutic proteins

Therapeutic proteins, in particular monoclonal antibodies (mAbs) are the most successful class of biopharmaceuticals on sale today, with above 50 mAbs approved and target sales of 130-200B$ in 2022 [1]. After the approval of the first mAb in 1986, Orthclone OKT3, mAb development has seen a boom, with 300 more mAbs in the pipeline and three to five new approved products per year (2015 data) [2].

For a long time, fed-batch processes have represented the golden standard in biopharmaceutical production [3]. In similarity to batch processes, cells are inoculated into a basal media. However, in fed-batch mode, highly concentrated feeds are added along the culture time span at discrete times or in a continuous fashion. This is done in order to increase viable cell density in the culture and, therefore, titre. Nowadays, most of manufacturing capacity is built on this type of processes [4].

The current market growth in the field of biopharmaceuticals has driven the investment increase in research and development (R&D). The ever-increasing advances in the discovery of mechanisms of diseases has potentiated the creation of more complex and targeted products. In addition, the maturation of the industry has brought a higher cost pressure to manufacturers [5]. The conjunction of these factors demands for more flexible process platforms, with a lower process footprint [6], [7].

A current, however nonunanimous [8]–[10], change to face these challenges is to move from batch to continuous manufacturing. In particular, perfusion bioreactors present a suitable process mode to culture mammalian cells, in order to produce high value pharmaceuticals. In a classical continuous culture, the so called chemostat, fresh media is continuously added, while culture broth containing nutrients and the protein product, metabolites and cells are seamlessly removed at the same rate, in order to keep the reactor volume constant. The main limitation of this process mode is the maximum flow rate, dilution rate, at which it can operate. If this parameter is higher than the cellular growth rate, then the cells will be washed-out and the production of the molecule of interest will be compromised. This does not constitute a problem in microbial cultures, where the cellular doubling time is low. On the other hand, mammalian cells exhibit a relatively slow growth rate, severely restricting the flow rate of chemostat systems [11]. Perfusion bioreactors, in turn, allow the operation at higher dilution rates by making use of a cell retention/recycling device that permits that the cells remain in the bioreactor, increasing throughput through superior cell densities. Another advantage is that, in comparable dilution rates, perfusion is a much more stable operating mode than the chemostat [12]. Thus, perfusion technology enables the continuous addition of fresh media and the continuous removal of the protein product and other by-products, and thus allows the achievement of higher viable cell densities and increased volumetric productivities compared to fed-batch technology [13]–[15].

Media development and optimisation qualify as one of the most important steps in the process development cycle of perfusion processes and is critical in order to fully seize the potential of continuous cultures [16]. Despite the fact that perfusion processes have a higher media consumption relative to fed-batch, hence an increased need for optimisation [17], other challenges exist. These include a higher

complexity of the process, having to leverage different process parameters, and a lack of knowledge and methodologies to conduct the optimisation efforts. In addition, there also exists a lack of scale-down models which can reproduce the cell retention capability and the continuous media exchange, characteristic of perfusion processes [16]. The existing models are built on a semi-continuous operation (daily media exchanges) and make use of existing deep-well plates, spin-tubes and micro-scale bioreactors [18]. While it is true that since fed-batch and perfusion cultures have different media requirements and media should be developed separately for each culture technique [16], in practice, perfusion media is still based on fed-batch optimisation steps [19]. In addition, media screening experiments in perfusion are more labour intensive than in fed-batch mode, since the former require daily operations.

1.2. Machine Learning in Life Sciences

Machine learning (ML) can be seen as form of artificial intelligence (AI), a way for a system to learn from data instead of through explicit programming. In this fashion, ML employs a variety of algorithms, which iteratively learn from data to improve, to describe data and to predict outcomes. As the *learning* is data-driven, models are constantly being refined and becoming more precise [20].

In the last years, this topic has left the pure development phase and has been applied in smaller or bigger scale across all industries, from banking to automotive [21], [22]. McKinsey & Company, a consultancy, reckons the pharmaceuticals industry is where the adoption levels of AI are lower and where the willingness to pay to implement these technologies is higher [23]. Both these factors put together make this an industry where AI can have a high impact. In addition, one of the most potential fields for future application of these technologies is R&D, especially in the biopharmaceutical industry, where R&D is the primary profit driver [24]. In this context, machine learning could reverse the trend towards higher costs and longer development times [24]. These techniques have been proven successfully in drug discovery, predicting molecular target bonding, identifying new markers and discovering drug indications, being just a matter of time before being applied in many other areas [25], [26].

Across the development phase of biopharmaceuticals, there are increased potentialities for the use of machine-learning tools, due to the fact that this is a very labour-intensive area and with a panoply of conditions to be tested. The implementation of techniques, which could reduce the experimental burden, have the potential for cutting cost and saving time and even liberating the experimenter towards more creative tasks.

1.3. Aim of the work

The aim of this work is to optimise the feeding strategy of an existing fed-batch protocol for a Chinese Hamster Ovary (CHO) cell line, producing a mAb-derived protein. Due to being highly work-intensive and since cultures last for more than one week, it is the perfect setting to exemplify the advantages and broad potentialities of machine learning and its potential to speed up and cut costs in bioprocess development, specifically in the context of therapeutic protein production.

Firstly, based on the concept of sequential design of experiments, a Bayesian optimisation (BO) algorithm was developed and tested in an *in-silico* model of a fed-batch process. Based on current literature several variations of the algorithm were implemented and assessed against each other. The developed algorithm was then compared with common optimisation methodologies.

The optimised algorithm was then tested on the optimisation of an existing fed-batch feeding protocol, executing several sets (or batches) of Bayesian optimisation designed experiments in scale-down models. In the end, the results were validated in benchtop bioreactors.

2. Theoretical Background

2.1. Classical vs Sequential Design of Experiments

In the early days, experiments were not developed in a structured manner, instead they were run somewhat at random, in a non-principled way. Different experimental conditions were randomly tested or in the manner of one factor at a time optimisation, and then it was attempted to draw some conclusions; without even considering if there was a way to extract more information from the same number of experiments [27].

A different approach was then developed to design experiments in systematic and methodical way. Design of experiments (DoE) can be seen as a methodology for the investigation of the mathematical relationships between the input and output variables of a system, permitting the revelation of how the interactions between input parameters influence the output responses [28]. In a typical design of experiments (named here as classical DoE), the operational parameters are determined, the ranges to explore are fixed and a method is selected for designing the subsequent experiments. These experiments are then performed, and the results analysed (through some modelling step, e.g. Response surfaces, multiple linear regression, …). Hence, the size and composition of the samples are fully determined before the experimentation begins. Several media optimisation approaches have been developed based on this methodology [29]–[31].

Using this approach, one does not utilize the information retrieved from one experiment to design the following. Instead, the sequential design of experiments is such a more advanced methodology. After all, the size and composition of the samples/experiments are not fixed in advance, but are functions of the observations themselves [32]–[34]. In other words, the measurement data will provide the scientist with new insights into the system under investigation, and hence extend the existing knowledge. This will produce a new state of knowledge, which allows an optimised, data-driven choice, of the following experimental designs. Following this trend, this allows the implementation of protocols that use the data available until that point in time to design the latter parts of data collection [34]. Some authors make use of this methodology for media optimisation purposes [35]. Figure 2.1 schematically represents the different workflows that characterize these two approaches of design of experiments.

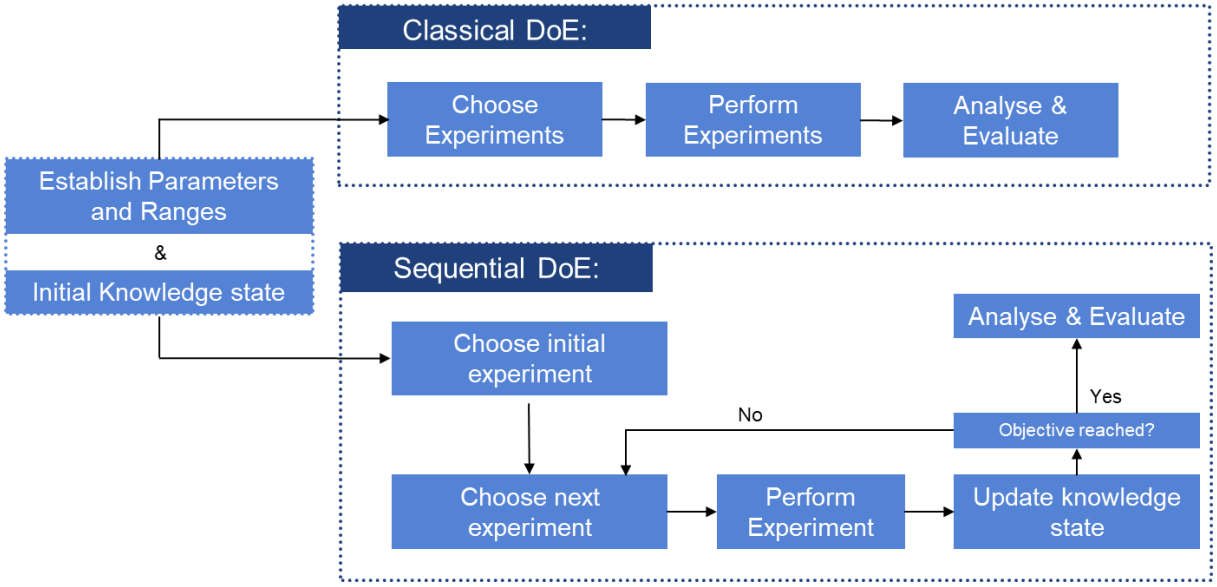

Figure 2.1 – Schematic representation of the different workflows that characterize classical and sequential design of experiments.

Geiges et al (2015) still describes a mixed approach, an adaptive DoE which has been applied in medical science [34]. In this approach, the experiments are initially designed in a globally optimal fashion and, since clinical studies last over an extended period of time and results are obtained sequentially, there are intermediate stages, where the remaining experiments are redesign according to the updated state of knowledge. Research towards the application of this mixed approach has been successfully carried out for media optimisation procedures [36].

2.2. Optimisation Algorithms

The methodology of sequential DoE can be applied towards the solution of an optimisation problem. In an optimisation problem the objective is normally the minimization or maximisation of an objective function (e.g. $f(x)$). The searchable design space is defined by upper ($x_{i,U}$) and lower ($x_{i,L}$) bounds of each design variable (vector x and i representing the position in such vector), referred to as side constraints. There are two types of optimisation problems: non-constrained and constrained. Non-constrained optimisation problems can have side constraints but do not have equality (e.g. $h_k(x)$) or inequality constraints (e.g. $g_j(x)$). A constrained problem has one or more equality and/or inequality constraints, with our without side constraints [37]. An example of a constrained optimisation problem is given (vector x represents the n design variables that are modified to obtain the optimum):

$$Maximize: f(x) \tag{1}$$

$$Subject\ to\ (e.g.): g_{j=1}(x) \leq 0, \quad (inequality\ constraint) \tag{2}$$

$$h_{k=1}(x) = 0, \quad (equality\ constraint\ ) \tag{3}$$

$$x_{i=1,L} \leq x_{i=1} \leq x_{i=1,U}, i = 1, (side\ constraint) \tag{4}$$

It should be noted that an equality constraint can be either violated or satisfied, while an inequality constraint can be violated, active or satisfied. For the given example an active inequality constraint would occur when $g_j(x) = 0$ [37].

The aim of optimisation algorithms is to find the best value of the objective function, according to the given constraints. However, one problem can have more than one optimum, hence several local optima. Following this trend, algorithms can also be divided between local optimisation algorithms and global ones [37]. Applied to the DoE methodology, local optima provide the highest utility in a certain neighbourhood of similar experiments, and the global optima provides the highest utility for the entire defined space of allowable experiments. Hence, in order to find the best solution of a given problem in the entire design space, the choice will fall on the utilisation of global search algorithms [34]. One must also consider that there is not a specific algorithm that can guarantee convergence on a global optimum [37].

It is important to point out that, in every optimisation algorithm, there is a duality or trade-off between exploitation and exploration. An increased exploitation will promote the search of areas of higher interest, or where local optima might be located, while an increased exploration will dedicate its search to different areas of the design space, taking a more global search approach [38], [39].

2.2.1. Evolutionary Algorithms

This class of algorithms is normally based on some natural phenomena and has the advantage of being very robust, sequential, and a higher probability of finding a near global optima and being simple to implement [37].

There are several algorithms that fall into this category, which are used in practice. These include: genetic algorithms, simulated annealing, tabu search, particle swarm optimisation, evolutionary programming, genetic programming, differential evolution, ant colony optimisation, harmony search, artificial bee colony, glow-worm swarm optimisation, cuck-oo search algorithm, among others [37], [38], [40].

Genetics algorithms (GA) are one of the most used algorithms in this group and are based on evolutionary biology, in Darwin's principle of the survival of the fittest. Hence, it mimics the process observed in nature where the strong tend to adapt and survive, while the weak tend to perish [37]. In this algorithm a new population is formed using specific genetic operators such as crossover, reproduction and mutation. A population is represented as a set of strings, or chromosomes, and in each new generation a new chromosome (a given member of the population) is created using the information from the fittest chromosomes of the last population. The first step of GA is then to create a randomized initial population. Each individual is then evaluated based in its fitness. This fitness parameter is calculated based on an objective function and, as specified earlier, fitter individuals have a higher probability of being selected. Is should be noted that a higher fitness value indicates a superior solution for maximisation and a low fitness value indicates a better solution for minimisation. An elemental GA is composed of five units: a random number generator, a fitness evaluation unit, a process of reproduction, a crossover process and a mutation operation. The reproduction process selects the fittest members of the population. The crossover combines the fittest chromosomes and passes the best genes to the offspring, while the mutation alters some of the genes in a chromosome, essential in order to explore

the design space. There are numerous modifications of this algorithm, including different mechanisms for the processes of crossover and mutation, which also has an impact on the duality between exploration and exploitation. There are also adaptive genetic algorithms where the characteristic parameters of the algorithm are constantly updated from iteration to iteration, hence optimizing the next steps [38], [40].

2.2.2. Bayesian Optimisation

Bayesian optimisation [41] is another algorithm used for these purposes that, fundamentally, is a sequential model-based approach to solving a problem. In general, a prior belief on the function to evaluate is prescribed, a model, which is sequentially refined as new data is acquired via Bayesian posterior updating. This posterior, hence, represents the updated beliefs on the likely objective function that is being optimised. The other component is the acquisition function, which guides the search of the design space, deciding the next evaluation of the function [42]. This algorithm is called Bayesian because of its foundation on the widely known Bayes Theorem:

$$p(w|D) = \frac{p(D|w)p(w)}{p(D)} \tag{5}$$

Where $p(w)$ represents the *a priori* distribution of w (an unobserved quantity), this captures the prior beliefs for w before the function evaluation. Given the observed data D and a likelihood model $p(D|w)$, it is possible to infer an *a posteriori* distribution $p(w|D)$, which represents the posterior beliefs on the objective function [42].

In other words, the prior is a surrogate model of the actual function, which it is used to carry out the optimisation. This model is then continuously updated through the introduction of new data points selected by the use of an acquisition function (which will try to satisfy some optimality criteria). Figure 2.2 schematically represents the Bayesian optimisation workflow.

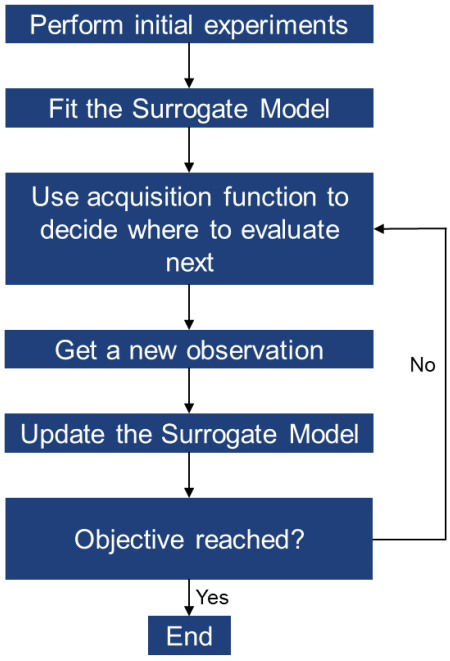

Figure 2.2 – Schematic representation of the Bayesian optimisation workflow.

2.2.2.1. Initialisation Methods

An important part of this algorithm is the initialisation. Different strategies are available, like one point in the centre of the domain, uniformly selected random locations, Latin hypercube sampling (LHCS), Halton sequences and determinant point processes. Independently from the choice made, it is important to start at a given location that minimises the initial model uncertainty [43]. However, these hypotheses are the most used in the machine-learning domain and specifically in the context of Bayesian optimisation, being seldom used in life sciences, where other DoE methodologies are more common. These last might include: full and fractional designs and all the types of optimal DoEs.

On that account, two distinct initialisation methods were implemented, one from each field: a fractional factorial design of resolution IV and Latin hypercube sampling. Fractional factorial designs are frequently used in screening experiments, or when there are not enough resources to carry out full factorial design, or even when it is possible to assume that higher order interactions are not significant. In this way, a fractional factorial design might be defined as a type of orthogonal array design that empowers experiments to study main effects and some interaction effects in a limited number of experiments [44]. Consider a system with n design variables: a design of resolution R means that no n −factor interaction is confounded with any other effect which contains less that $R - n$ factors. Thus, a design of resolution IV means that main effects are confounded with three-factor interactions and two-factor interactions are confounded with others of the same type [44], [45]. A design of resolution V is considered to be excellent, while IV might be adequate and III are only useful as economical screening designs [46].

It has been reported [47] that Latin hypercube design of experiments is especially appropriate for Gaussian process modelling and has become particularly prominent amidst other strategies for computer experiments. Consider a matrix where each column represents a parameter and each row represents a sample; in addition, the matrix contains as many rows as the number of experiments available. Dividing each of the parameters into equal levels, as many as there are rows, a Latin hypercube design is built by placing only one experiment at each level. In this way, the samples are distributed across the bounds of each parameter and throughout the search space [47].

2.2.2.2. Gaussian Processes

The determination of the specific type of surrogate model to utilize in the algorithm is of paramount importance. The default choice is Gaussian processes (GP). Although alternatives have been published, such as Random Forest [48], t-Student processes [49] and Sparse Spectrum GPs [42], most Bayesian optimisation articles make use of the traditional GPs [42], given the fact that they can handle nicely the uncertainty about the system in evaluation [50]. Similarly to a Gaussian distribution, which is a distribution over a random variable, specified by its mean and covariance, a GP is a distribution over a function, completely specified by its mean (m) and covariance/kernel functions (k) [51]:

$$f(x) \sim GP(m(x), k(x, x')) \tag{6}$$

Figure 2.3 shows a graphical representation of a Gaussian process prediction: the prior consists of all the possible distributions over the function (infinite, since are no sample points), while in the posterior, only the distributions which correspond to the obtained data points remain, and finally the prediction with a certain level of uncertainty.

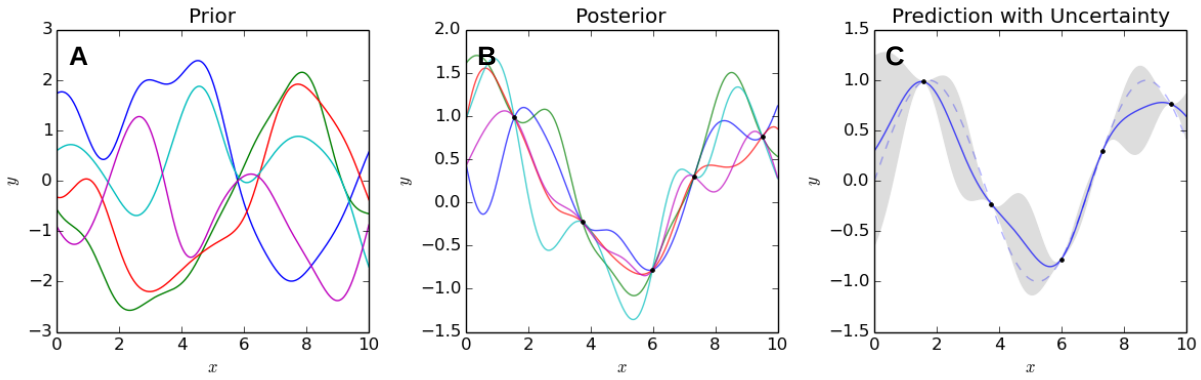

Figure 2.3 – Gaussian process model prediction, using a squared exponential kernel. A: draws from the prior distribution of functions, B: draws from the posterior predictive distribution, C: Mean prediction (continuous line) with one standard deviation shaded, actual function (dotted line). Reproduced with permission from: [52].

Basis Function

Normally, in Gaussian process regression, explicit mean functions are not used, due to the difficulty in specifying them. Instead, a zero-mean GP with a non-zero covariance function is considered and the following model is used:

$$h(x)^T\beta + f(x), \qquad where\ f(x) \sim GP(0, k(x, x')) \tag{7}$$

Where $h(x)$ represents a set of fixed basis functions and β additional parameters. In this way, the expressed model might be considered as a global linear model, where the residuals are modelled through the use of a GP. In other words, the basis function transforms the starting feature vector x in R^d into a new vector in R^p, hence β is a p-by-1 vector of basis function coefficients. These coefficients are derived from data and are subjected to optimisation [53], [54].

In the present work, several basis functions were considered: the non-existence of such function, $h(x) = 1$, a linear function and one with quadratic terms.

Kernel Function

The choice of the covariance or kernel function is decisive, since it determines the smoothness properties of the samples drawn from it [51]. In other words, it encodes assumptions about the similarity between data points, since two nearing data points should have similar responses, training points near to a test sample should be more informative about the prediction at that point [53].

For example, one very popular kernel is the squared exponential function:

$$k(x_i; x_j) = \exp\left(-\frac{1}{2}\left\|x_i - x_j\right\|^2\right) \tag{8}$$

This function approaches the value of 1 as two values become closer and 0 as they get further apart. However, in this kernel the divergence in all parameters of x affect the covariance equally, which does not hold true in most real cases, therefore more complicated kernels were developed [51].

One very popular class of covariance functions are the Matérn kernels [55], [56]. These kernels are parameterised by a smoothness parameter ν, this is due to the fact that samples from a GP with this kernel are differentiable $(\nu - 1)$ times [42]:

$$k(x_i; x_j) = \frac{1}{2^{\nu-1}\Gamma(\nu)}\left(2\sqrt{\nu}\|x_i - x_j\|\right)^{\nu} \mathrm{H}_{\nu}\left(2\sqrt{\nu}\|x_i - x_j\|\right) \tag{9}$$

where $\Gamma(\cdot)$ is the Gamma function and $H\nu(\cdot)$ is the Bessel function of order ν. The squared exponential kernel, defined previously, is the special case when $\nu \to \infty$. Another popular kernel is the exponential one, equivalent to the Matérn kernel when $\nu = 1/2$. Other prominent kernels belonging to this class are the Matérn 32 and Matérn 52, with $\nu = 3/2$ and $5/2$, respectively.

All of the mentioned kernels can include automatic relevance determination (ARD) parameters, which include an amplitude and length scale hyperparameters, as many as existing dimensions in the feature vector. The inclusion of these hyperparameters allow the attribution of more or less relevance to the different dimensions [42], [51].

2.2.2.3. Acquisition Function

Asides from the choice of the surrogate model, one must also choose the acquisition function that is most suitable to the problem at hand. This choice is important given that it has a high impact on the success of the simulation [57]. The decision on which function to implement implies a trade-off between exploration of the search space and exploitation of the most promising regions [42].

Several types of acquisition functions are known. These can be grouped in improvement-based, optimistic and information-based policies. The first group favours evaluations at points that are likely to improve a given target. One of the first functions to be developed was simply the probability of improvement, which can trigger an excessive exploitation of the search space. Hence, another function was developed, the expected improvement (EI), which also incorporates the amount of improvement. In the group of optimistic policies, an acquisition function called the upper confidence bound (UCB) criterion was developed and has been classified as a popular way of trading-off exploration and exploitation. Information-based policies were designed to consider the posterior distribution over an unknown minimiser. There are two policies in this class: Thompson sampling (TS) and entropy search (ES). TS aims to draw a continuous function from the posterior GP and in turn optimise this function in order to produce a sample point. For continuous spaces this must be done through spectral sampling techniques, which allow the drawing of a sample from the posterior that can be evaluated at any point. However, this function seems to deteriorate in high-dimensional spaces, through aggressive exploration. ES techniques aim to diminish the uncertainty in the location of the next evaluation point by selecting the point that is expected to cause the largest reduction in entropy of posterior distribution over the unknown minimiser. Explained in a different way, these techniques measure the expected information gain from querying an arbitrary evaluation point and select the point that offers the most information about the unknown evaluation point. For continuous search spaces this technique employs a discretisation or an alternative technique, much more used for these purposes, such as predictive entropy search can be utilised. Some authors [58] consider that no single acquisition function is best at all problem settings, hence, have proposed the use of a portfolio of acquisition functions. Thus, deciding

on which point to evaluate next, based on the choices given by the various acquisition functions, through a portfolio function, one of the most recent is the entropy search portfolio [42].

However, in the present work only the two most popular and well-described acquisition functions were implemented: expected improvement and upper confidence bound.

The analytical expression for EI is as follows [59], [60]:

$$EI(x) = \begin{cases} (m(x) - f(x^+) - \xi)\Phi(Z) + k(x)\phi(Z), & k(x) > 0 \\ 0, & k(x) = 0 \end{cases} \quad (10)$$

$$Z = \frac{\mu(x) - f(x^+)}{k(x)} \quad (11)$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ represent the probability density function (PDF) and the cumulative distribution function (CDF), respectively, of the standard normal distribution and $\mu(\cdot)$ and $k(\cdot)$ represent the mean and covariance of the GP posterior. Here the target objective value, $f(x^+)$, was adaptively implemented as the best observed value [42] and ξ, a measure by which one wishes to improve, equal to the error variance [47].

The UCB function, in its form as defined in [61] is:

$$UCB(x) = \mu(x) + \sqrt{\nu\tau_t}k(x) \quad (12)$$

implemented according with [51], [61] and where ν and τ_t are characteristic hyperparameters of this function.

Since, the experiments to be performed have the levels corresponding to the maximum point of the acquisition function, it is necessary to maximise this function. In this work, the DIRECT optimiser [62] was used, a deterministic, derivative-free optimiser, the most used algorithm for this objective [51], [63]. Unlike the unknown objective function, the acquisition function is not expensive to sample, hence any global search optimisation could in principle be used.

2.2.2.4. Batch Bayesian Optimisation

Bayesian optimisation aims to design the next experiment always based on the most current and up-to-date beliefs of the function/search space in question. However, in certain situations, it is profitable to run several experiments simultaneously. Be it the ability to run several simulations at the same time, making use of parallel computing capabilities, or even performing several wet-lab experiments at the same time, either due to time constrains or if the cost of doing one experiment is the same as doing several. This is the setting for the appearance of Bayesian optimisation algorithms which are able to select sets of experiments to be performed [51].

Sequential approaches have a fundamental advantage over batch BO. They are able to use more accurate *a priori* information when choosing the next experiment, making a more informed choice. Hence, batch Bayesian optimisation algorithms aim to minimise this potential loss, trying to design a batch policy that matches the sequential mode behaviour. Several algorithms have been developed to this end [64]–[66].

In the present work, batch selection via local penalisation was implemented [66]. This algorithm has the benefit of having a very intuitive foundation: after the selection of the first element of the batch through maximisation of the acquisition function, the same is penalised in the maximum point, creating an exclusion zone and allowing the selection of a different experiment for the next element of the batch.

The implementation of this algorithm followed the pseudo-code:

$\boldsymbol{Input}$: $dataset\ D = \{x_i; y_i\}_{i=1}^{n}, batch\ size\ n_b, iteration\ budget\ m$
$\boldsymbol{for}\ t = 1\ to\ m$:
  $Fit\ the\ GP\ to\ D$
  $Build\ acquisition\ function\ \alpha(\cdot)\ on\ the\ current\ GP$
  $\boldsymbol{for}\ j = 1\ to\ n_b$:
    $batch\ element\ j = argmax\big(\alpha(\cdot)\big)$
    $\alpha(\cdot) = \alpha(\cdot)\varphi(\cdot, L)$
  $\boldsymbol{end}$
$run\ batch\ B$
$add\ responses\ to\ D$
$\boldsymbol{end}$

where φ represents a local penaliser of the acquisition function and L the local Lipchitz constant.

The current setting has certain characteristics that influence the choice of the optimisation algorithm. Experiments involving therapeutic proteins are naturally “expensive”. In other words, they are very laborious and go on for a long time, even weeks or months. Comparing evolutionary algorithms to the last one described – Bayesian Optimisation – the latter needs much less observations to achieve convergence. In fact, the Bayesian optimisation algorithm is especially dedicated to expensive functions [51], [59], [67] and has also been applied towards biological systems [68]–[70].

3. Materials and Methods

3.1. Fed-batch *In Silico* Model

The used *in silico* model aims to describe the behaviour of a certain animal cell culture, in fed-batch mode, with temperature and pH control. This model was developed in-house, based on [71], [72], with a wide number of modifications and additions. In order to use the given model, an experiment with the levels for the 13 parameters was inputted. As output the model provided day-to-day information on six parameters (viable cell density and glucose, glutamine, ammonium, lactate and product concentration) until the end of the culture. The output provided was not deterministic, attempting to mimic real experiments, by the insertion of some error in the output.

3.2. Fed-Batch Media Optimisation

3.2.1. Cell line

An mAb-derived protein producing Chinese Hamster Ovary (CHO) industrial cell line (Merck Serono S.A., Corsier-sur-Vevey, Switzerland) was used for all experiments.

3.2.2. Media Solutions and Ranges

Expansion media consisted of a chemically defined, serum-free medium (Merck Serono S.A., Switzerland), supplemented with 100 μM methionine sulfoximine (MSX). Production media consisted of the same chemically defined media without the addition of MSX.

The concentrated feed was a chemically defined, proprietary medium (Merck Serono S.A., Switzerland). The other feeds were an enriched Cysteine/Tyrosine (L-Cysteine; Merck Serono S.A., Coinsins, Switzerland, Tyrosine, Sigma-Aldrich Co., St. Louis, USA) feed (pH 11.2), a high concentration asparagine (Sigma-Aldrich Co., St. Louis, USA) feed, and a high concentration glucose (D-(+)-Glucose monohydrate, Sigma-Aldrich Co., St. Louis, USA) feed.

All solutions were sterilized through filtration (Filtermax “rapid” 0.2 μm PES, TubeSpin; TPP, Trasadingen, Switzerland).

Ranges of the search space (Table 3.1) for the optimisation procedure were defined through expert guidance, based on the existing protocol (see Section 3.2.4).

Table 3.1 – Initial boundaries of the search space for the wet-lab media optimisation experiments and baseline process levels.

Component	Lower bound	Upper bound	Baseline Process
Glucose (g/L)	5.0	8.4	7
Cysteine/Tyrosine (% total volume)	0	0.36	0.15
Concentrated Feed (% total volume)	1.0	6.0	3
Asparagine (% total volume)	0	0.6	0

3.2.3. Expansion procedure

Cells, belonging to a previously constructed working cell bank and stored in liquid nitrogen, were thawed and diluted in expansion medium. Cells were centrifuged for 5 min at 800 g (Centrifuge 5810R, Vandaux-Eppendorf AG, Schönenbuch, Switzerland), supernatant discarded, and the pellet was resuspended in fresh expansion media, targeting a Viable Cell Density (VCD) of 0.3 Mcells/mL. The cell suspension was kept in 50 mL STs (TubeSpin; TPP, Trasadingen, Switzerland) at 36.5 °C, 320 rpm, 5% CO_2 and 90% humidity for 1 week in a shaking incubator (Adolf Kühner AG, Birsfelden, Switzerland). Cells were passaged every two days, when the volume of the suspension surpassed 100 mL, cells were kept in 2L roller bottles (Corning, New York, USA), these were in turn stored in a shaking incubator (Adolf Kühner AG, Birsfelden, Switzerland) with the shaking speed reduced to 130 rpm.

3.2.4. Spin-tube bioreactor operation

The scale-down bioreactor operation was performed in 50mL ST, using a 20mL initial working volume. Cells were inoculated on WD00 with an initial VCD of 0.5 Mcells/mL, by mixing the stock cell suspension with an appropriate amount of fresh production media. The STs were then kept in a shaking incubator in the same conditions mentioned in Section 3.2.3. As to prevent cell debris formation and accumulation due to the high shear stress, the cell culture was transferred to a fresh ST every 5 days [73].

Feedings were done according to the process schedule shown in Figure 3.1, in which the arrows represent the various feedings. The size of the arrows shows qualitatively the difference in amount of feed between the different days. The amount of solution fed to the ST was predetermined for the asparagine, cysteine/tyrosine and concentrated solution, based on a percentage of the total volume of the suspension. The amount of glucose added was based on the actual glucose concentration. Hence, a certain quantity of glucose solution was added in order to achieve a defined posterior glucose concentration.

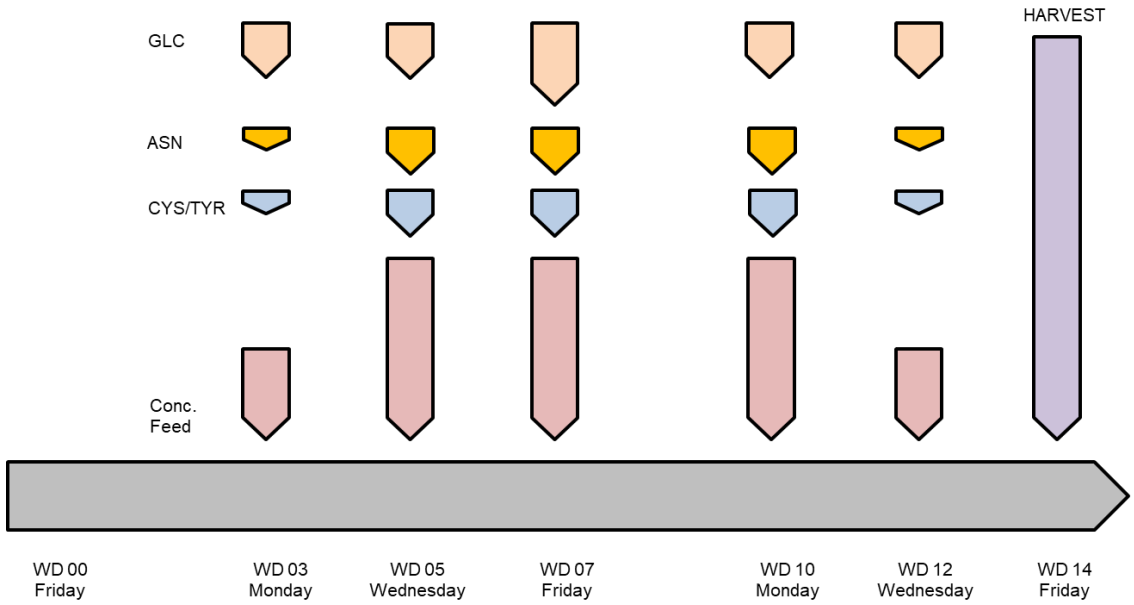

Figure 3.1 – Baseline process flow. Feeding are done every other day, not accounting for the weekend and starting on WD 3. The process lasts in total for 14 days, when the culture broth is harvested. The different size of shapes shows qualitatively the difference in feedings between the several days.

Samples were taken every day, by removing 0.5mL or 0.15mL of cell suspension in a feeding or non-feeding day, respectively. All the samples were analysed for glucose, lactate, target product, ammonia and lactate dehydrogenase (LDH) concentration. VCD and viability were only determined on days where feedings occurred.

On WD14 the culture broth was harvested through centrifugation (800g, 5min), filtered (Filtropur S 0.2 µm, Sarsted; Aktiengesellschaft & Co., Nümbrecht, Germany) and stored at -20°C for further analysis.

3.2.5. Benchtop bioreactor operation

In the benchtop reactor runs, the same process flow was used as described for the ST operation. The expanded cells were inoculated in a 3L working volume stirred tank reactor (DASGIP, Vandaux-Eppendorf AG, Schönenbuch, Switzerland). The cell cultivation was performed at pH 7.1, 36.5°C and 50% dissolved oxygen tension (DO). These process parameters were kept in a controlled state thorough the use of the DASGIP control system and pH, DO and temperature probes. Temperature was adjusted by the use of a heating mantel and DO and pH through the control of oxygen and carbon dioxide in the gas inlet, respectively. Mixing was achieved through the use of one Ruston impeller mounted at the bottom of the agitation shaft and pitched blade impeller mounted above, blades inclined by 30° from the horizontal plane [74]. Aeration was supplied to the reactor using an open-pipe sparger, with 7 holes. Anti-foam was added on demand to prevent excessive foaming (FoamAway, Life Technologies Corporation, NY, USA). The reactors were placed on balances which allowed the control of the added feedings, pumped by DASGIP peristaltic pumps, in manual mode.

Samples were taken every day, by removing 10mL of culture broth. All samples were analysed for glucose, lactate, product, ammonia and LDH concentration and VCD. The samples were stored at -20°C for further analysis.

3.2.6. Analytical Methods

3.2.6.1. Cell-related Parameters

Viable cell density and viability were determined by the trypan blue exclusion method, automatically performed using a Cedex HiRes Analyser (Roche Diagnostics, Basel, Switzerland).

3.2.6.2. Metabolites

Glucose, lactate, ammonia, LDH and mAb concentrations were determined by a CedexBio Analyser (Roche Diagnostics, Basel, Switzerland). Measurements are performed by means of an absorbance photometer. Samples consisted of a cell-free supernatant, due to previous centrifugation of the original sample for 5min at 800g.

3.2.9.3. Protein A affinity chromatography

The reference determination of mAb concentration was performed by Protein A affinity chromatography. The samples, stored at 20°C, were thawed and filtered (0.2 µm syringe filters) prior to analysis. A Chromolith WP300 Protein A column (25-4.6mm, Merck KGaA, Darmstadt, Germany), placed on an Agilent High Performance Liquid Chromatography (HPLC) system (Santa Clara, CA, USA), combined with a ultraviolet (UV) detector (280nm), was used for the analysis. The binding buffer

consisted of 100mM Na-phosphate aqueous buffer at pH 7.4 and the elution buffer of 100mM Na-phosphate aqueous buffer at pH 2.5. A flow rate of 2 mL/min and injection volume between 25-100μL was used. The elution was made in an isocratic fashion. The calculation of the antibody concentration was made according to the following formula:

$$[mAb] = \frac{Elution\ Peak\ Area \times Flow\ Rate}{Injection\ Volume \times Calibration\ Factor} \quad (13)$$

The calibration factor used was predetermined by injection of a sample with a know antibody concentration.

3.2.7. Machine Learning/Modelling

All algorithms were implemented in MATLAB R2017b (The MathWorks Inc., MA, USA). All individual methods were implemented within the work and, in certain cases, using the already existing MATLAB functions and using default settings (see Figure 3.2). The fitting of the Gaussian process and the optimisation of its structure and hyperparameters was done using the following specific options: 'FitMethod' as 'exact', 'DistanceMethod' as 'accurate' and 'PredictMethod' as 'exact'; particularly, for the hyperparameter and structure optimisation, repartitioning of the cross-validation sets at every iteration was used. Random search was developed within the present work, following the methodology of pure random search [75].

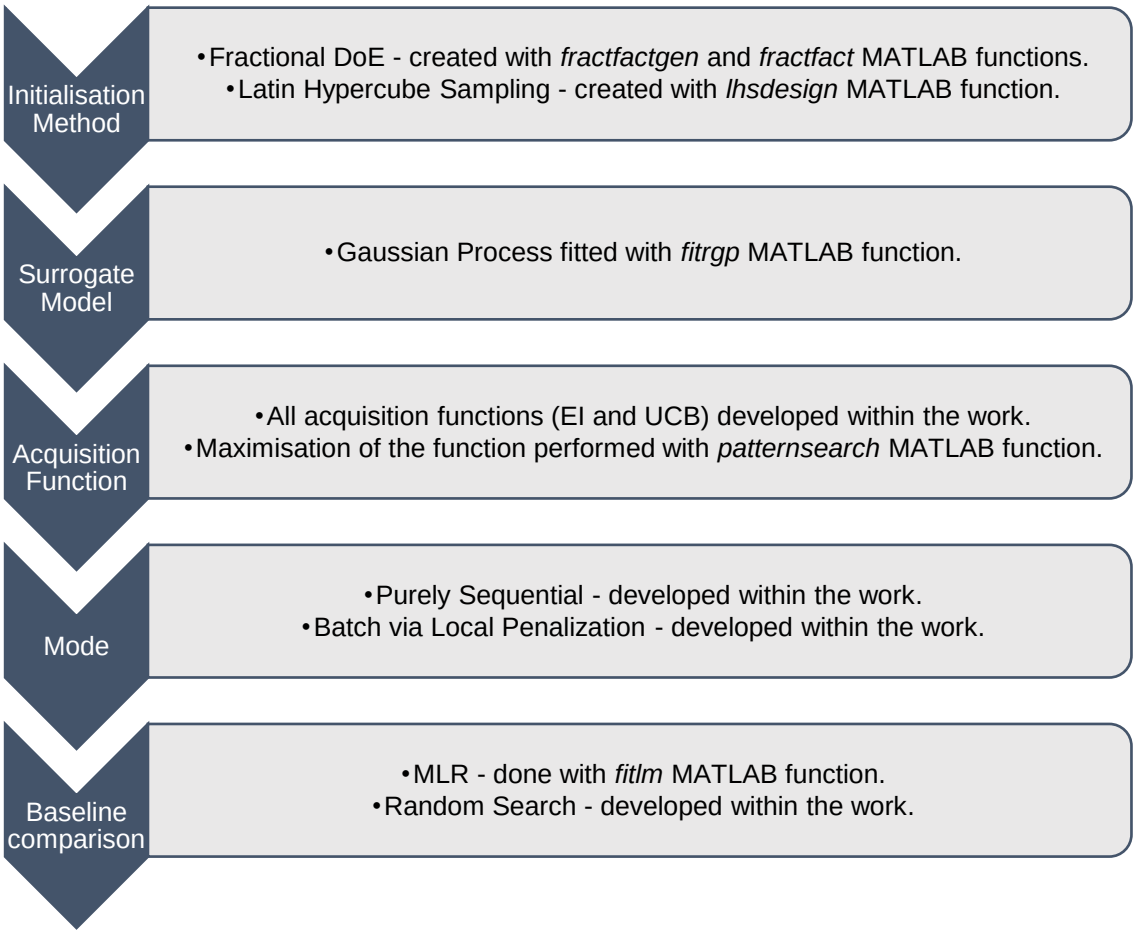

Figure 3.2 – Overview of the algorithms implemented within the work and, in certain cases, the MATLAB functions used.

4. Results and Discussion

4.1. Fed-Batch *In Silico* Experiments

Firstly, the developed algorithms were tested in an *In Silico* fed-batch model (see Section 3.1), allowing the comparison of the different algorithmic options and with established methods.

4.1.1. Initialisation Method

The initialisation phase of the Bayesian optimisation workflow aims to define a set of initial experiments, which will explore the search space and determine possible interesting regions. Two different methods were studied: a fractional DoE of IV resolution and Latin Hypercube Sampling.

The maximum titre achieved versus the number of experiments performed for BO using both methods is shown in Figure 4.1. In the initialisation phase, 33 experiments were done, because this was number of experiments necessary to achieve a resolution IV design with 13 parameters. The individual results of these experiments are not shown, due to the fact that the experiments chosen by the tested methods are part of a single design. Thus, the titre results emerge as the ones of a group and not of the individual experiments. Following this logic, experiment number zero represents the maximum titre achieved in the initialisation phase as a whole. Both algorithms were repeated 10 times, to safeguard against differences caused by the randomization/probabilistic steps inside the creation of the designs.

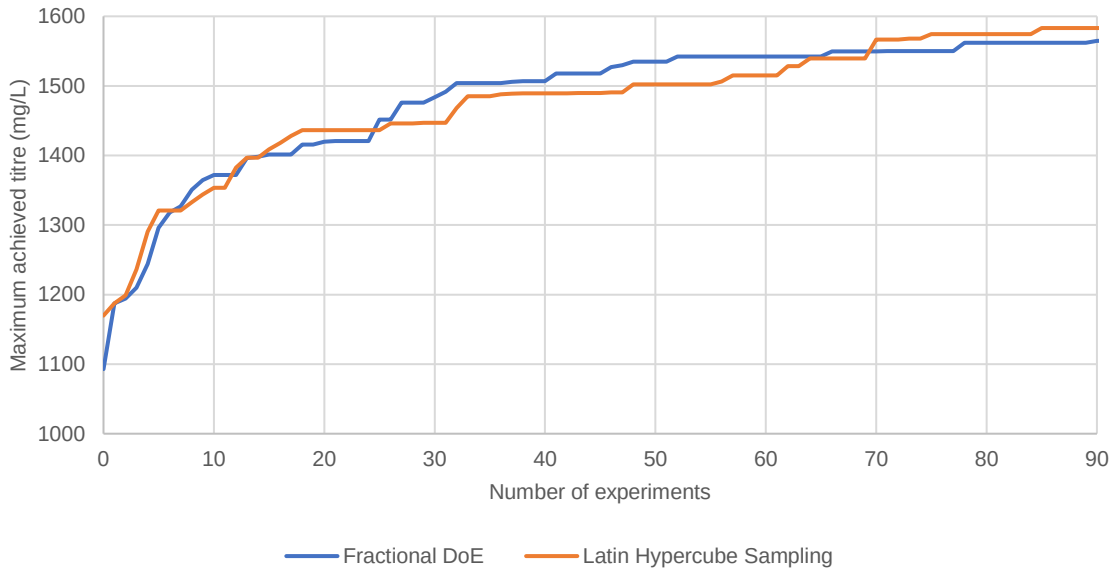

Figure 4.1 – Comparison between the two implemented initialisation methods for Bayesian optimisation (IV Fractional DoE and Latin Hypercube Sampling) in terms of maximum titre achieved vs. *number of experiments performed. For both cases the remainder of the BO algorithm was implemented using expected improvement as acquisition function and experiments chosen in a purely sequential fashion. Experiments belonging to the initialisation phase (33 experiments) are not shown. Experiment zero represents the maximum titre achieved in initialisation. Results shown are averaged over 10 repetitions of the BO algorithm.*

Comparing only the titre achieved at the end of the 33 experiments, which compose the initialisation phase (experiment 0 in Figure 4.1), the Latin hypercube sampling is superior (1.1 vs 1.18 g/L). After the initialisation phase, results show that there is not a significant difference in the gain in titre, between the two initialisation methods tested, essentially depending on the number of experiments performed which one is better.

The difference in titre out of the initialisation phase might be due to the fact that LHCS is a space filling algorithm, distributing the experiments evenly across the search space. On the other hand, the fractional factorial design will try to place samples in the corners and might miss some of the most interesting regions. Although in theory, the fractional DoE (in this case of resolution IV) might be more useful, since there is the theoretical guarantee of having no main effects aliased with other main effects or with 2-factor interactions, while in LHCS there no such theoretical assurances [44], [46]. It has also been reported [47] that Latin hypercube sampling is more appropriate for Gaussian process modelling approaches. In addition, LHCS has increased flexibility. While in factorial designs, if some dimensions are eliminated, sampling points collapse one into the other, reducing the size of the sample, this does not happen in LHCS [47]. Hereinafter, for the exposed reasons, LHCS would be used as initialisation algorithm.

4.1.2. Acquisition Function

Following the workflow of the Bayesian optimisation algorithm, two of the most popular acquisition functions were implemented and tested: expected improvement and upper confidence bound. Figure 4.2 shows the comparison of these different algorithms in terms of maximum titre achieved versus the number of experiments performed after the initialisation phase.

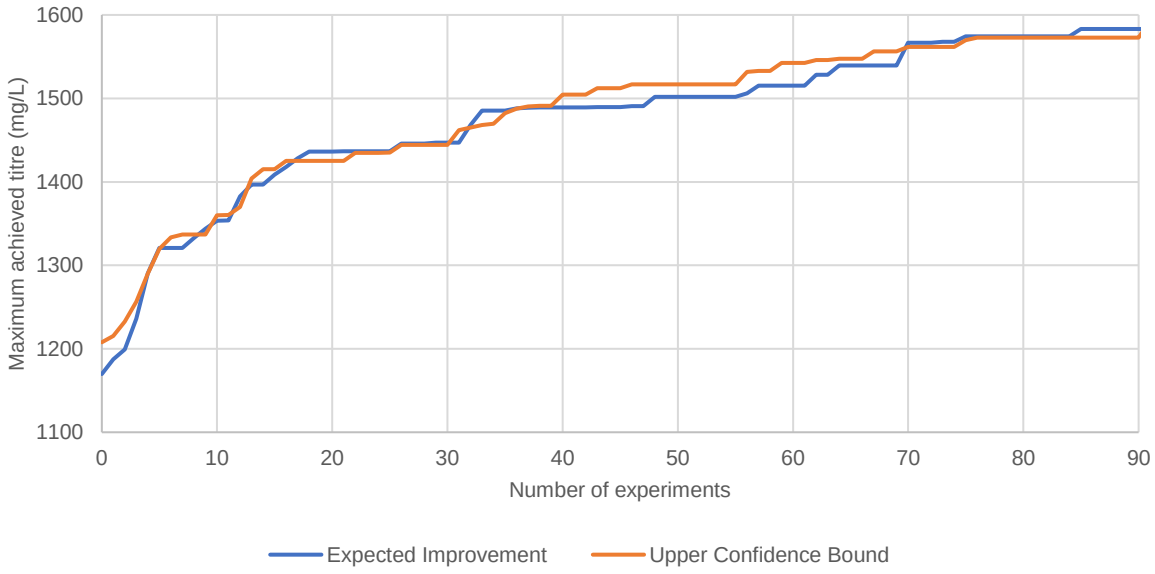

Figure 4.2 - Comparison between the two implemented acquisition functions (Expected Improvement and Upper Confidence Bound) in terms of maximum titre achieved vs. *number of experiments performed. For both cases the remainder of the BO algorithm was implemented using Latin hypercube sampling as initialisation method and experiments chosen in a purely sequential fashion. Experiments belonging to the initialisation phase (33 experiments) are not shown. Experiment zero represents the maximum titre achieved in initialisation. Results shown are averaged over 10 repetitions of the BO algorithm.*

Both algorithms show similar behaviours, simply depending on the number of experiments which one is better. Although it has been described [51] that the EI algorithm is more greedy and UCB more explorative. Since EI does not require its own tuning parameter (see Section 2.2.2.3), this is the function chosen for the next steps.

4.1.3. Comparison with established methods

A comparison with established DoE methods was performed (Figure 4.3). Such methods include: pure random search, multiple linear regression (MLR) based on LHCS defined experiments and MLR based on D-optimal DoE experiments. In these last two methods, the experiments defined by the design were performed and MLR was applied between the levels and the titre results.

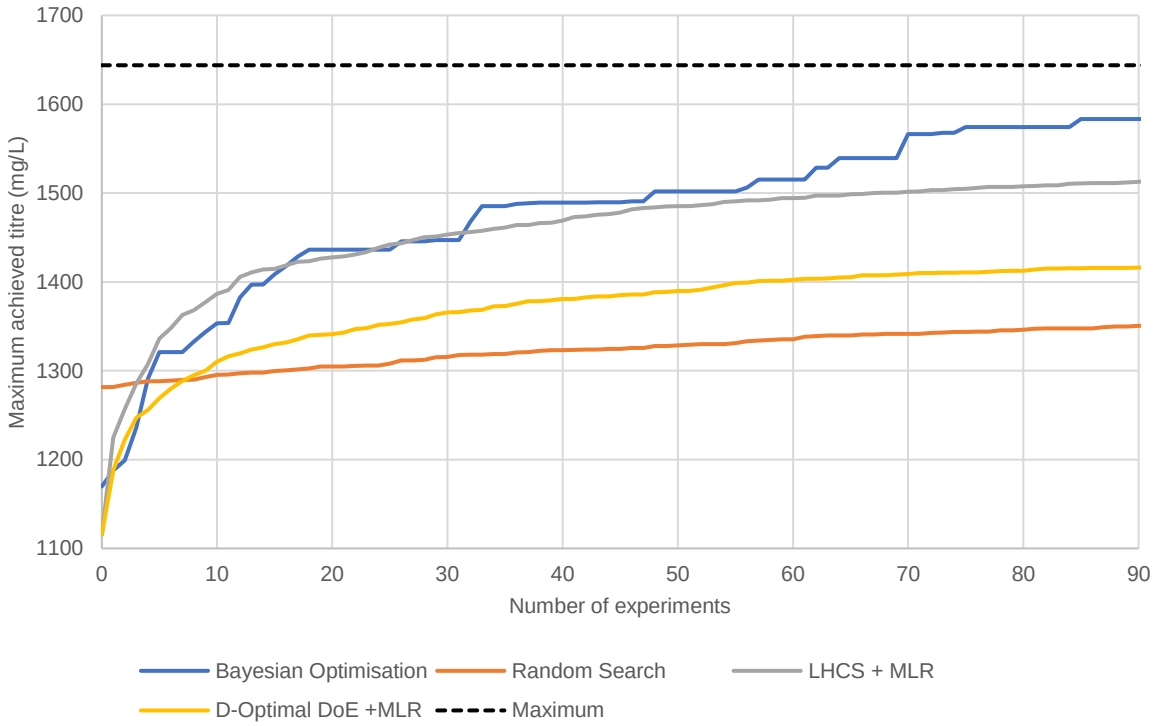

Figure 4.3 - Comparison between the best Bayesian optimisation algorithm (Latin hypercube sampling, expected improvement and purely sequential mode) with several established methods: random search, D-optimal DoE, multiple linear regression based on D-optimal DoE experiments and Latin hypercube sampling. Experiments belonging to the initialisation phase (33 experiments) are not shown. Experiment zero represents the maximum titre achieved in such phase. Results shown are averaged over 10 repetitions of the BO algorithm and 100 of the remaining methods.

The defined Bayesian optimisation algorithm greatly surpasses the performance of the most traditional baselines, as random search and MLR based on D-optimal DoE experiments. However, MLR based on LHCS experiments can perform as well as Bayesian optimisation for a small number of experiments. This might be due to the fact that while D-optimal DoE places experiments in theoretically optimal points, LHCS is a space-filling algorithm, hence distributes the available experiments throughout the search space, possibly capturing the most interesting regions. In conclusion, Bayesian optimisation continues to be a superior approach due to the fact that is truly sequential, allowing even the redesign of the ranges of the search space from experiment to experiment or from batch to batch. This is something that LHCS does not offer and, in some cases, might cause to do experiments which are not even effective.

Comparing to a more traditional DoE, the full factorial design, with which with 3 levels, would involve performing 13^3 experiments, BO has tremendous advantages, since it only required 90 experiments to reach very close to the maximum. In addition, one practical advantage for processes in the development phase is that Bayesian optimisation, in general, will suggest experiments that have

high gain (in this example, titre), given that experiments are chosen sequentially and focused on the most interesting regions. In this fashion, these experiments will produce more product than the ones with other methods, enabling the supply of more product to other steps of process development (e.g. downstream development).

4.1.4. Sequential *versus* Batch Bayesian Optimisation

Both in *in silico* and in a wet-lab, it is useful and time-saving to run several experiments at the same time. This can be achieved either, in one case, by making use of the possibilities of parallel computing, running different simulations in different processor cores, or in the other case, by saving time and effort through the preparation of reagents and other equipment for several experiments at the same time. Following this logic, batch Bayesian optimisation was implemented via local penalization (see Section 2.2.2.4). Figure 4.4 compares the maximum achieved titre *vs.*. the number of experiments performed after the initialisation in the referred new variation and in purely sequential mode.

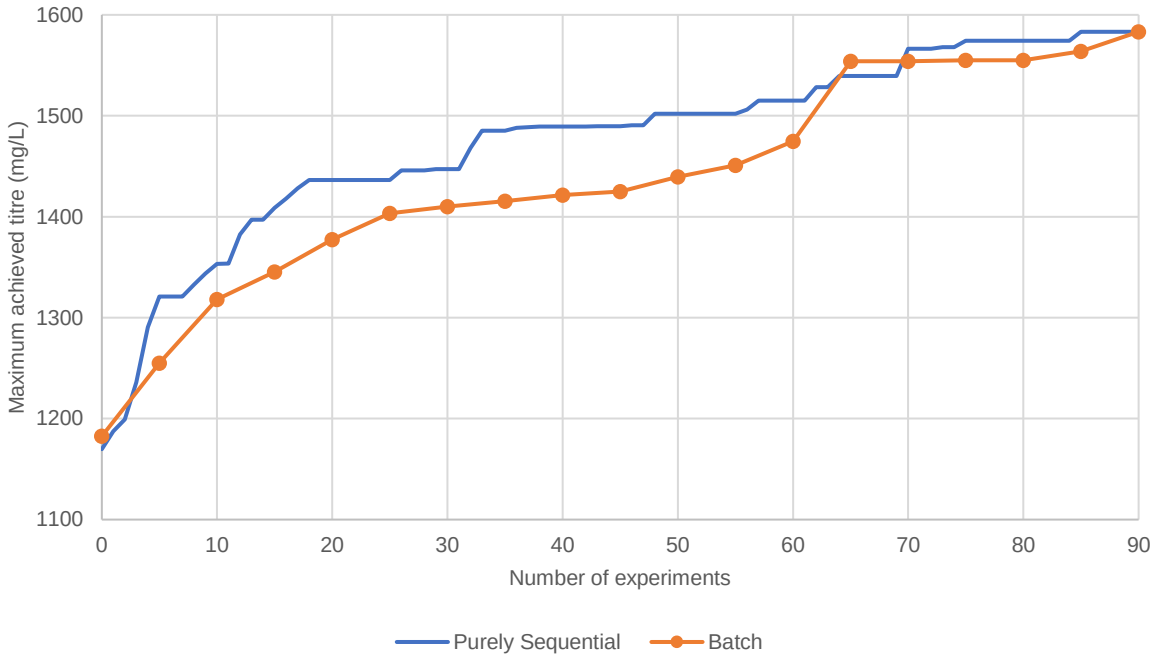

Figure 4.4 - Comparison between the two implemented Bayesian optimisation modes of operation (purely sequential and batch) in terms of maximum titre achieved vs *number of experiments performed. For both cases the remainder of the BO algorithm was implemented using Latin hypercube sampling as initialisation method and expected improvement as acquisition function. Batch Bayesian optimisation was done using batches composed of 5 experiments. Experiments belonging to the initialisation phase (33 experiments) are not shown. Experiment zero represents the maximum titre achieved in initialisation. Results shown are averaged over 10 repetitions of the BO algorithm. Markers represent the different performed batches.*

As would be expected, the sequential mode outperforms the batch mode for a significant number of experiments. Due to the fact that the next experiment is always defined with a superior amount of *a priori* information. However, for a higher number of experiments the behaviour of both methods is comparable, and the difference in optimised titre by using batch BO diminishes. These results are comparable to literature descriptions [66], albeit it should be pointed out that the comparison should be made in a model specific fashion. Some authors [65] have found that, depending on the model used, either policy could be better than the other. Another parameter, which causes difficulty in

generalizing such results, is the size of the batch. Hence, comparison should really be made on a case-by-case basis.

In conclusion, one might use batch Bayesian optimisation, with the advantages it carries, with minimal loss of information, at least in the present case. Furthermore, one might even argue that if experiments in are designed in batch mode, it will be possible to carry out more, in the same time frame, than in a sequential fashion. In this case, definitely surpassing the titre gain from sequential experiments. This last conclusion can be drawn, if the limiting factor is time. However, if instead the number of experiments one can perform is limiting, the experiments should be carried out in sequential mode. This is because, every experiment will carry a bigger information gain, due to it being more accurately selected based on a higher degree of prior knowledge. In batch mode, since experiments are chosen based on assumptions of the posterior, they are not so accurately selected.

4.2. Wet-lab Fed-Batch Media Optimisation

As a wet-lab validation of the proposed algorithm, the feed strategy of an mAb producing CHO culture was optimised. Initially, the end concentration of protein for the baseline process (see Table 3.1) was established. This experiment was followed by an initialisation batch and by 4 sequential batches, defined through the proposed Bayesian optimisation unit. All these cultures were performed in spin tubes. Finally, to validate the ST optimisation step, the baseline process and an optimised condition were performed in lab-scale bioreactors.

Unless otherwise stated, all time resolved titre information is based on Cedex Bio measurements, while all the titre values at the end of the culture are based on Protein A analytics (please see Section 4.2.4).

4.2.1. Spin-tube bioreactor experiments

4.2.1.1. Baseline Process & Initialisation Batch

Firstly, the cultures corresponding to the baseline process and to the initialisation batch were performed. This initialisation batch comprised 14 experiments defined by the used DoE (Latin Hypercube Sampling, see Chapter 2.2.2.1) and two centre points, whose levels are shown in Table 4.1.

*Table 4.1 – Overview of the levels for the experiments performed in the initialisation batch, defined by the applied initialisation method – LHCS. *Levels with replicate.*

Experiment	Glucose (g/L)	Cysteine / Tyrosine (% total vol.)	Concentrated Feed (% total vol.)	Asparagine (% total vol.)
1	5	0,07	1,69	0,33
2	5,4	0,31	1,73	0,47
3	5,49	0,3	2,62	0,23
4	5,95	0,11	2,81	0,07
5	6,21	0,27	4,66	0,17
6	6,27	0,16	5,48	0,49
7	6,69	0,05	4,2	0,54
8	6,83	0,01	1	0,1
9	7,01	0,36	4,39	0,42
10	7,19	0,09	2,11	0,26
11	7,48	0,15	5,77	0,21
12	7,83	0,24	5,18	0,56
13	8,14	0,2	3,69	0,04
14	8,31	0,22	3,36	0,35
Centre Point*	6,7	0,18	3,5	0,3
Baseline*	7	0,15	3	0

In Figure 4.5, the evolution of the following parameters during the duration of the culture is shown: VCD, viability and the concentrations of glucose, lactate, ammonia and target product (LDH concentration evolution is available in Appendix A).

As expected, since there are no differences between the cultures from inoculation until WD03, all cultures are very comparable until this point. After the first feeding the variability is noticed even from the following day. The maximum VCD that each culture achieved varied between 6.1 and 9.7 Mcells/mL, corresponding to experiments 8 and 2, respectively. While for most experiments the cell viability was kept above 90% throughout the culture, in 4 out of 14 this parameter started dropping towards the end. This trend was more pronounced for experiments 1, 7 and 8. One possibility is the depletion of the amino acids cysteine or tyrosine, given that these are experiments are the ones that had the lowest supply of this feed. On the other end of the spectra, the experiments which achieved the highest maximum VCD were 2, 3, 4, and the centre point.

The evolution of the glucose concentration, since it is part of the feeding strategy, varies immensely from experiment to experiment after WD03. However, in most of the cultures the decay in concentration after feeding is comparable. This observation indicates that the rate of glucose consumption, equivalent to the slope of the concentration of glucose *vs.* time, is equal for the majority of the STs. Deviation from this trend is experiment 8, which manifests a lower glucose consumption rate, possibly to due to its lower VCD and viability. In addition, in the last days of the culture, experiment 14 exhibits a higher glucose consumption rate than the rest of the experiments. However, since there is no other graspable difference from the remainder of cultures, it might be classified as an outlier.

Concerning the lactate production, it is possible to observe a direct trend between lactate accumulation and the cysteine/tyrosine feeding.

There also appears to be a direct relation between higher levels of asparagine and concentrated feed and increased ammonia production. Since more asparagine is fed to the culture, increased ammonia production occurs through the conversion of asparagine into aspartate. In addition, as reported before [76], increased amino acid levels, whose provenance is attributed to the discussed feed, could lead to increased catabolic breakdown, culminating in a surge in ammonia levels. In principle, through catabolism, the cell can utilize the carbon backbones for the formation of citric acid cycle intermediates, to be used in the central metabolic pathways. However, when supplied in excess, cellular metabolism lead to the formation of by-products, mainly ammonia. Depending on the amino acids, they can undergo direct deamination, or be transformed in glutamate through a transamination reaction. Glutamate can, consecutively, be subjected to deamination [77].

In terms of titre, the baseline process produced 1.3 ± 0.36 g/L at the end of the culture (two standard deviations used). So, considering the variability of the process, there was not a significant titre improvement in the initialisation batch. It should be pointed out that the slight decrease in titre between day 7 and day 10 is due to the change in analytical tests in the Cedex Bio device. In this device, two different tests exist for the determination of the target protein concentration: a basic and a diluted one. Since the ranges of these tests overlap, there is difference in measured concentrations in this area causing the observed decreased.

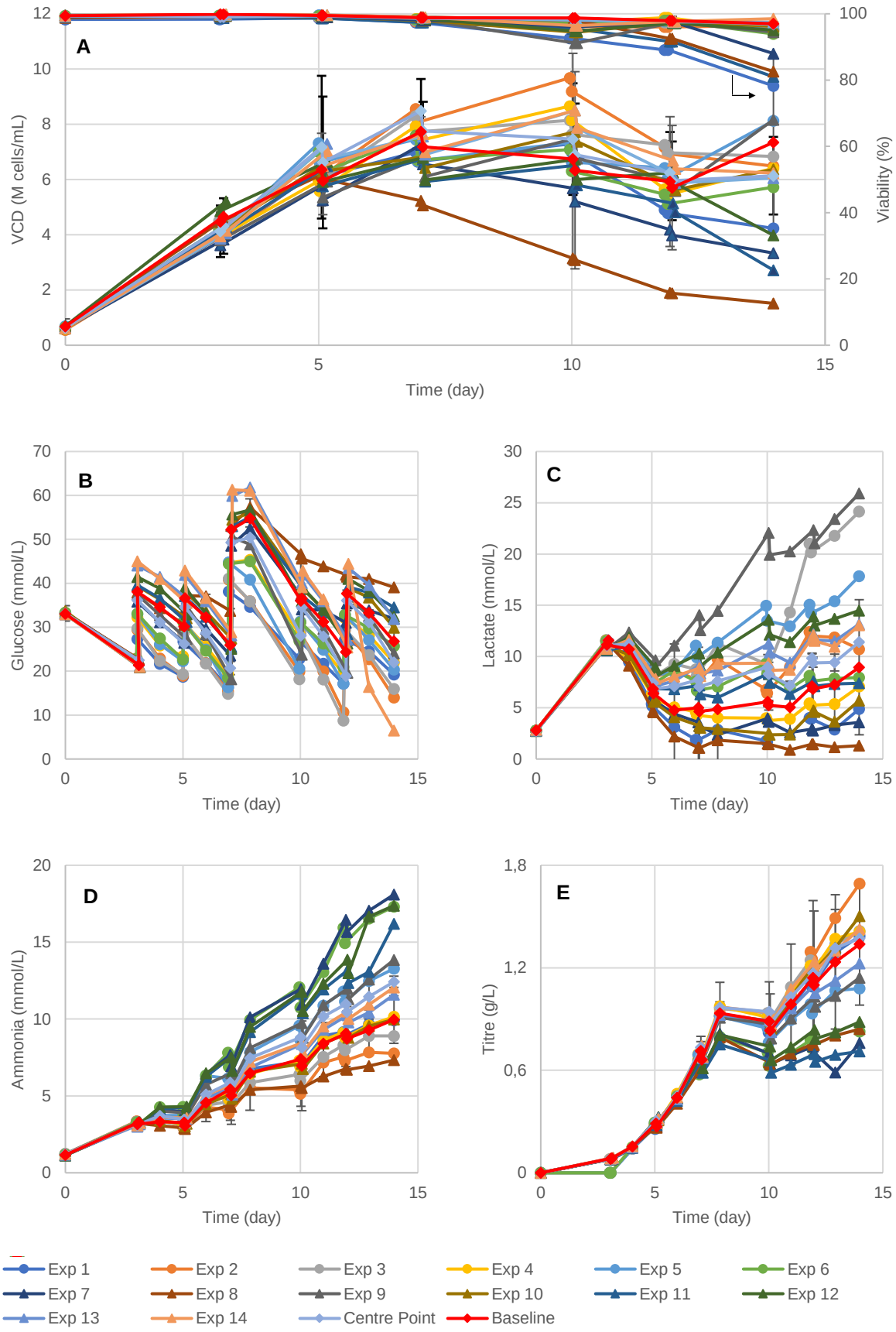

Figure 4.5 – Time evolution of the monitored parameters of the ST cultures which compose the initialisation batch and baseline establishment: Viable cell density and viability (A), glucose (B), lactate (C), ammonia (D) and product (E) concentration. Error bars represent two standard deviations of n=2.

4.2.1.2. 1st BayesOpt Batch

Using the data of the initialisation step, the first batch of experiments was chosen based on Bayesian optimisation. Inserted in the workflow of the algorithm, the structure of the surrogate model, a Gaussian process, was optimised and its hyperparameters determined. These details are exposed on Table 4.2.

The structure of the used Gaussian process included as kernel function the ARD Matérn 32 ($\nu = 3/2$), belonging to a class of covariance functions frequently used in machine learning [51]. The specific case of this kernel function represents a trade-off between the somewhat smooth, for practical optimisation problems, squared exponential (equivalent to the Matérn kernel with $\nu = 1/2$) and the ARD Matérn 52 ($\nu = 5/2$). This last probably is the most used kernel with Bayesian optimisation [42], [53], [78]. In terms of basis function, the one chosen was a constant scalar, indicating no change in dimensionality between the original feature space and the new one [54].

Table 4.2 – Details of the structure of the Gaussian process used in the Bayesian optimisation workflow used in the determination of the 1st batch of experiments.

		Hyperparameters	
Kernel Function	ARD Materné 32	Length Scale 1	3.1028×10^4
		Length Scale 2	0.4017
		Length Scale 3	2.5204
		Length Scale 4	1.1412
		Amplitude	0.4272
Basis Function	Constant	Beta	0.8521
Error Variance	0.0090		

Based on the specified model, a set of four experiments to be performed were chosen, exposed in Table 4.3. In this set, the only parameter, which had some variations intra-batch was glucose, with both the cysteine/tyrosine and asparagine feed in the upper extreme of the search space's range, albeit in the last case only for 3 out of the 4 experiments.

Table 4.3 – Levels performed in the first batch of experiments designed with Bayesian optimisation.

Experiment	**Glucose (g/L)**	**Cysteine / Tyrosine (% total vol.)**	**Concentrated Feed (% total vol.)**	**Asparagine (% total vol.)**
15	5.00	0.36	1.82	0.60
16	6,88	0.36	1.82	0.60
17	8.40	0.36	1.82	0.60
18	5.94	0.36	1.82	0

In Figure 4.6, the time-evolution of the different parameters monitored during the cell cultures corresponding to this batch is shown (LDH concentration evolution is available in Appendix A). Experiments 15, 16 and 17 are the most similar experiments, with only the level of glucose changing. In this sense, they have very comparable behaviours, in all parameters. Possibly the only major difference, was a slight decrease in VCD in the end of the culture for experiment 15. One hypothesis is that, since it has the lowest level of glucose, a certain degree of limitation might have occurred. The most

contrasting experiment was number 18, with decreased VCD, titre and ammonia levels and a high accumulation of lactate, in comparison with the remaining experiments. Since asparagine carries an amide group, it can generate ammonia in lower pH solutions, and might explain the higher levels of ammonia in the other 3 experiments [79]. In addition, is a source of nitrogen for the cells, hence, has an impact on ammonia formation, increasing it [80]. The increased lactate accumulation, might also be explained by the lack of asparagine feed, having been reported before [80]. Given that this amino acid is an essential one, if it is depleted cells cannot metabolise glucose further than the glycolysis step. Thus, the citric acid cycle cannot occur, leading to lactate accumulation. Finally, several authors have observed [79]–[81] that the addition of asparagine leads to an increased productivity in CHO cells, explaining the decline in titre for experiment 18. Also, in terms of mAb production, the titre in this batch was reached 2.00 g/L (Cedex Bio measurement), surpassing the levels reached in the initialisation batch.

Table 4.4 shows the comparison between the predicted titre by the surrogate model, the Gaussian process, and the titre achieved in the actual experiments (Cedex Bio measurement). Given, that the Gaussian process predictions for this first batch were only based on the data belonging to the initialisation phase, the root mean square error in prediction (RMSEP) is still high, standing at 0.31 g/L.

Table 4.4 – Comparison between the titre predicted by the surrogate model and the titre really achieved (Cedex Bio measurement) in the first batch defined by Bayesian optimisation. RMSEP of 0.31 g/L.

Experiment	Predicted Titre (g/L)	Titre achieved (g/L)
15	1.71	2.00
16	1.71	1.97
17	1.71	1.86
18	1.62	1.15
RMSEP = 0.31 g/L		

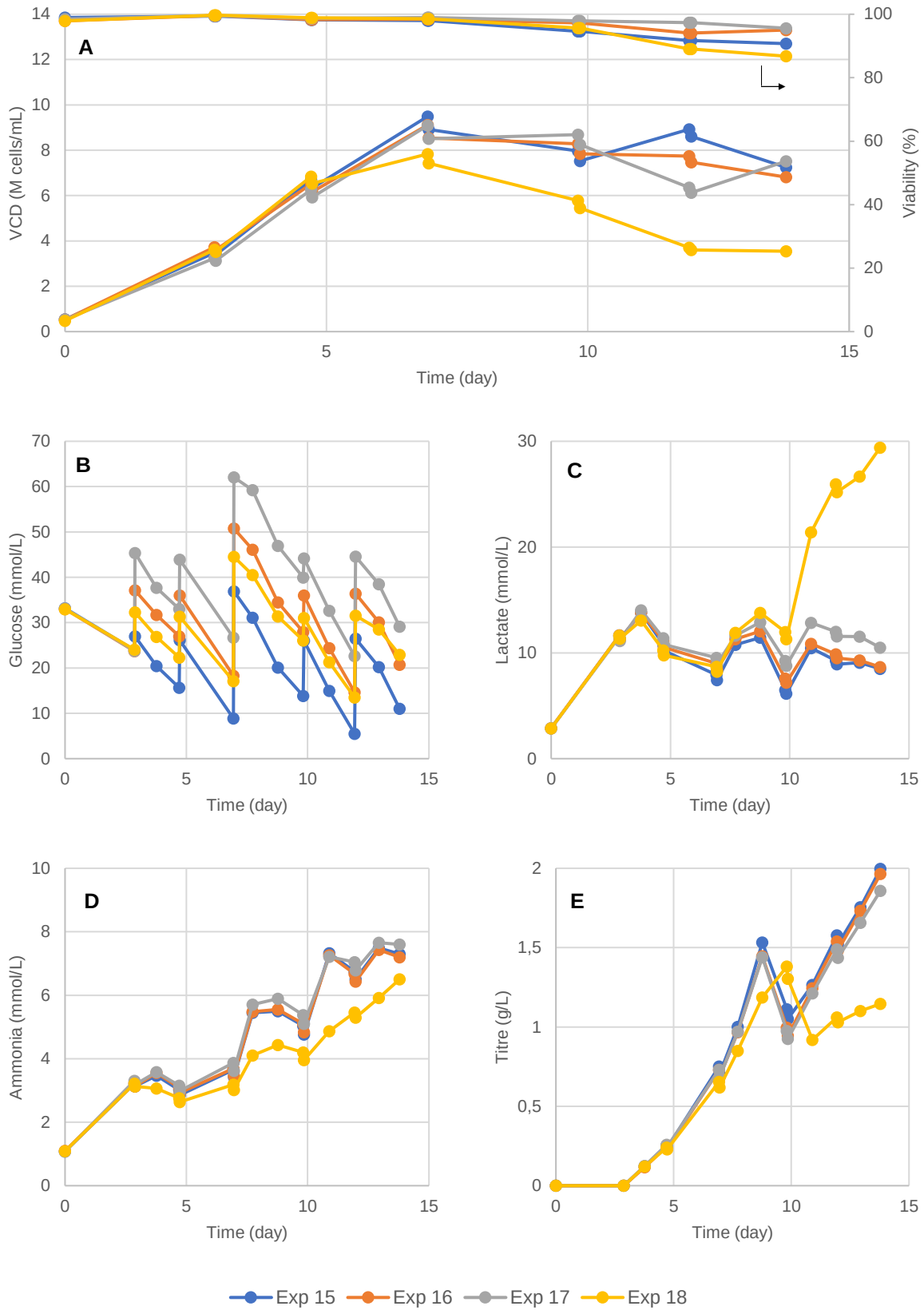

Figure 4.6 - Time-wise evolution of the monitored parameters of the ST cultures which compose 1st batch with Bayesian optimisation: Viable cell density and viability (A), glucose (B), lactate (C), ammonia (D) and product (E) concentration.

4.2.1.3. 2nd BayesOpt Batch

Based on the previous collected data, the Gaussian process, which aims to describe the search space, was updated. The structure remained the same, the only changes occurred in the hyperparameters (see Table 4.5). This maintenance of the overall structure is reasonable, since the architecture of the data and space, which the surrogate model has to emulate, has not changed. In this fashion, just the hyperparameters are truly optimised to better match the additional data and with each step reduce uncertainty. In fact, the variance of the error has diminished between the previous batch and the current one.

Table 4.5 - Details of the structure of the Gaussian process used in the Bayesian optimisation workflow used in the determination of the 2nd batch of experiments.

		Hyperparameters	
Kernel Function	ARD Materné 32	Length Scale 1	15.090
		Length Scale 2	0.6043
		Length Scale 3	1.9518
		Length Scale 4	0.6240
		Amplitude	0.4116
Basis Function	Constant	Beta	1.0459
Error Variance	0.0037		

Building on the updated mode, the Bayesian optimisation algorithm defined the next batch of experiments to be carried out (Table 4.6). The levels of the cysteine/tyrosine and asparagine feeds remained the same, in the upper extreme of the ranges, and the levels of the concentrated feed intra-batch, in practical terms, do not vary. However, the glucose levels have been narrowed, possibly indicating that the algorithm found the most interesting region for this parameter.

Table 4.6 - Levels performed in the second batch of experiments designed with Bayesian optimisation.

Experiment	**Glucose (g/L)**	**Cysteine / Tyrosine (% total vol.)**	**Concentrated Feed (% total vol.)**	**Asparagine (% total vol.)**
19	5.00	0.36	2.14	0.60
20	5.56	0.36	2.13	0.60
21	6.03	0.36	2.15	0.60
22	5.27	0.36	2.13	0.60

Given the slender changes between the four experiments, the differences in the monitored parameters inside the batch are inconspicuous (Figure 4.7, LDH concentration evolution is available in Appendix A), asides from the glucose differences. Besides this observation, there is a meagre difference on the VCD behaviour towards the end of the culture amongst the experiments. Both the experiments with the highest levels of the glucose (20 and 21) saw their VCD rise, while in experiments 19 and 22 this parameter kept constant. Comparing the present batch with experiments 15, 16 and 17 of the previous one, it is possible to observe that, in the current one, there was no change from lactate accumulation to consumption. Given that the difference in levels between the mentioned sets experiments is only noticeable for the concentrated feed, this must be the cause. This change in levels, in a general sense, might also be cause for the increase in ammonia (as mentioned earlier) and VCD.

According to Cedex Bio measurements, the maximum titre was not surpassed. Table 4.7 shows the titre achieved in the experiments belonging to this batch (Cedex Bio measurement) and the titre predicted by the surrogate model. Comparing to the results of the previous batch (Table 4.4), it is possible to observe that the RMSEP has been lowered. This is concordant with the above drawn conclusion the structure of the Gaussian process, in which only the hyperparameters were changed.

Table 4.7 - Comparison between the titre predicted by the surrogate model and the titre really achieved (Cedex Bio measurement) in the second batch defined by Bayesian optimisation. RMSEP of 0.27 g/L.

Experiment	Predicted Titre (g/L)	Titre achieved (g/L)
19	1.98	1.75
20	1.98	1.68
21	1.96	1.69
22	1.98	1.69
RMSEP = 0.27 g/L		

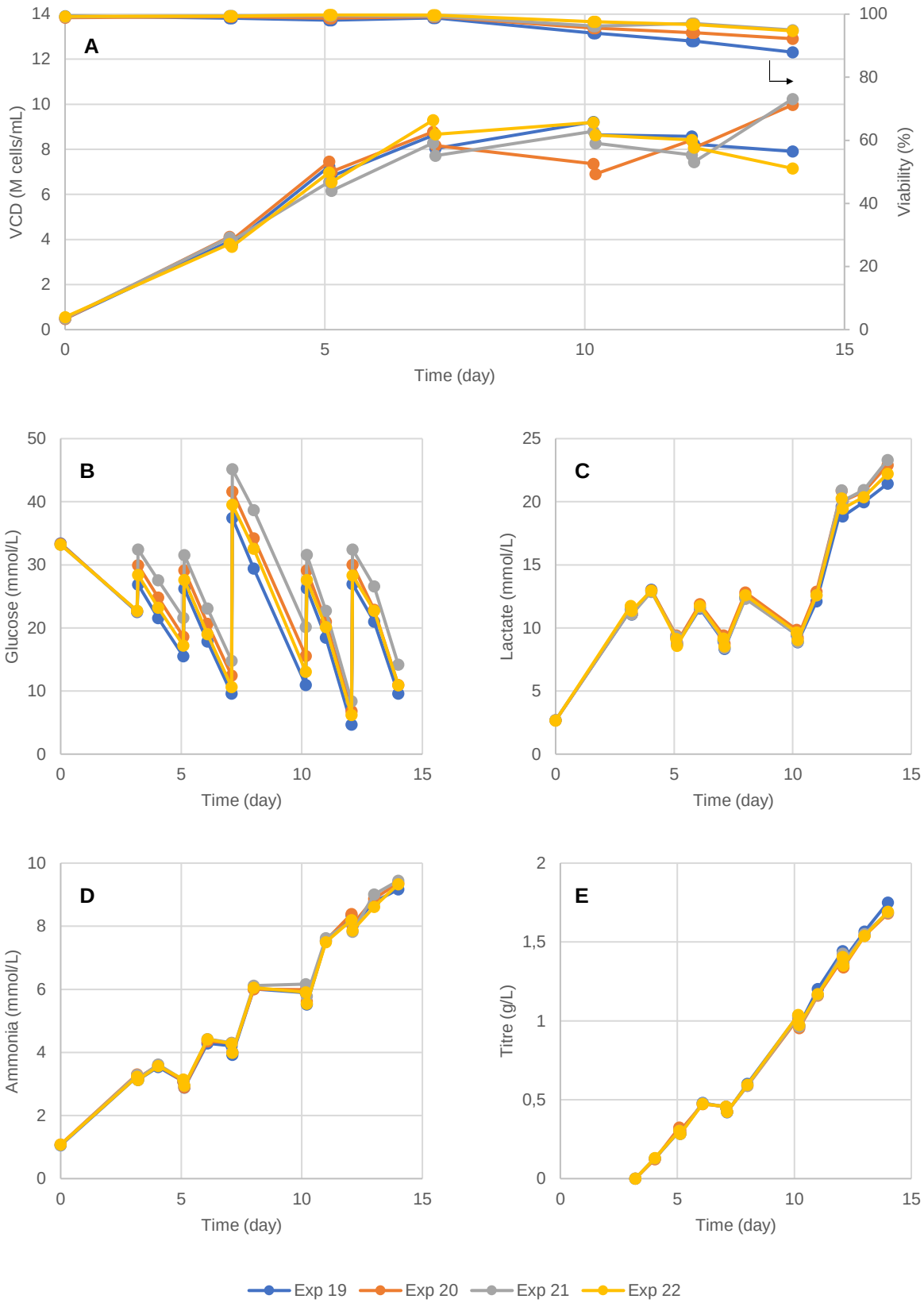

Figure 4.7 - Time-wise evolution of the monitored parameters of the ST cultures which compose 2nd batch with Bayesian optimisation: Viable cell density and viability (A), glucose (B), lactate (C), ammonia (D) and product (E) concentration.

4.2.1.4. 3rd BayesOpt Batch

Due to the fact that in the previous two batches, the levels of both the cysteine/tyrosine and asparagine feeds were in the upper extreme of its ranges, it was possible that the defined search space was too constrained. In other words, it is possible that the actual global titre maximum is located outside of the defined ranges for these two parameters. Following this chain of thought, the upper bounds for these parameters were increased by 30% (Table 4.8). It should be highlighted that this possibility of reviewing the boundaries of the search space is characteristic of BO, and would not be possible in baseline methods analysed previously (see Section 4.1.3).

Table 4.8 – Comparison between the previously established upper bounds for the cysteine/tyrosine and asparagine feeds and the ones applied for the 3rd and 4th Bayesian optimisation batches, an increase of 30% between the two.

	Previous upper bounds	**Increased upper bounds**
Cysteine/Tyrosine (% total volume)	0.36	0.47
Asparagine (% total volume	0.6	0.78

Hinged on the collected data, the hyperparameters of the Gaussian process were optimised (Table 4.9) and the new levels for the 3rd batch of experiments defined (Table 4.10). As it is possible to observe, the structure remains the same, and the difference between the presented hyperparameters and the ones determined for the 2nd batch is much smaller than the difference between the latter and the 1st batch. This indicates that the adjustment that the model needs to do mimic the search space is increasingly smaller and, in consequence, the information gain from each of the experiments decreases.

Table 4.9 - Details of the structure of the Gaussian process used in the Bayesian optimisation workflow used in the determination of the 3rd batch of experiments.

		Hyperparameters	
Kernel Function	ARD Materné 32	Length Scale 1	15.090
		Length Scale 2	0.6047
		Length Scale 3	1.9518
		Length Scale 4	0.6240
		Amplitude	0.4116
Basis Function	Constant	Beta	1.0459
Error Variance	0.0037		

As with the previous batch of experiments, the most significant difference between the levels lies with the glucose parameters, while the remaining are more or less consistent intra-batch. Noticeable is the increase in the cysteine/tyrosine and asparagine levels, due to the increase in the bounds of the search space.

Table 4.10 - Levels performed in the third batch of experiments designed with Bayesian optimisation. The construction of this batch of experiments was done with the extended range of percentage of total volume to feed of cysteine/tyrosine and asparagine.

Experiment	Glucose (g/L)	Cysteine / Tyrosine (% total vol.)	Concentrated Feed (% total vol.)	Asparagine (% total vol.)
23	5.00	0.41	1.90	0.74
24	5.98	0.41	1.90	0.74
25	6.83	0.43	1.90	0.78
26	5.35	0.41	1.90	0.74

The results show that there are not intra-batch significant differences in the monitored parameters of such cultures (Figure 4.8, LDH concentration evolution is available in Appendix A). An exception is the decline in VCD and viability observed in experiment 23, the one with the lowest glucose level. This observation had already been made for similar experiments (19 and 22) and can confirm the previously made hypothesis: such low glucose levels can have a limiting effect in cell growth. mAb production in this batch resumed the elevated levels achieved in the 1st Bayesian optimisation batch, to 1.9 g/L (Cedex Bio measurement). Table 4.11 shows the comparison between the titre achieved and the predicted by the Gaussian process. Once again, the RMSEP is lower than in the previous batch. This is another proof that the surrogate model, from batch to batch, is able to better adapt to the data, emulating the search space, refining the hyperparameters and reducing error in predictions. In addition, the RMSEP is lower even if the bounds of the search space were increased, indicating the capacity of the Gaussian process to extrapolate beyond he boundaries of the training data. However, it should be pointed out that the objective of the algorithm and of its underlying model is it not to obtain a perfect capture of the search space, but to design the best performing experiments. Hence, although the reduction of the RMSEP is a welcome attribute of the algorithm, it is not the main objective.

Table 4.11 - Comparison between the titre predicted by the surrogate model and the titre really achieved (Cedex Bio measurement) in the third batch defined by Bayesian optimisation. RMSEP of 0.16 g/L.

Experiment	Predicted Titre (g/L)	Titre achieved (g/L)
23	2.05	1.84
24	2.04	1.83
25	1.99	1.89
26	2.05	1.95
RMSEP = 0.16 g/L		

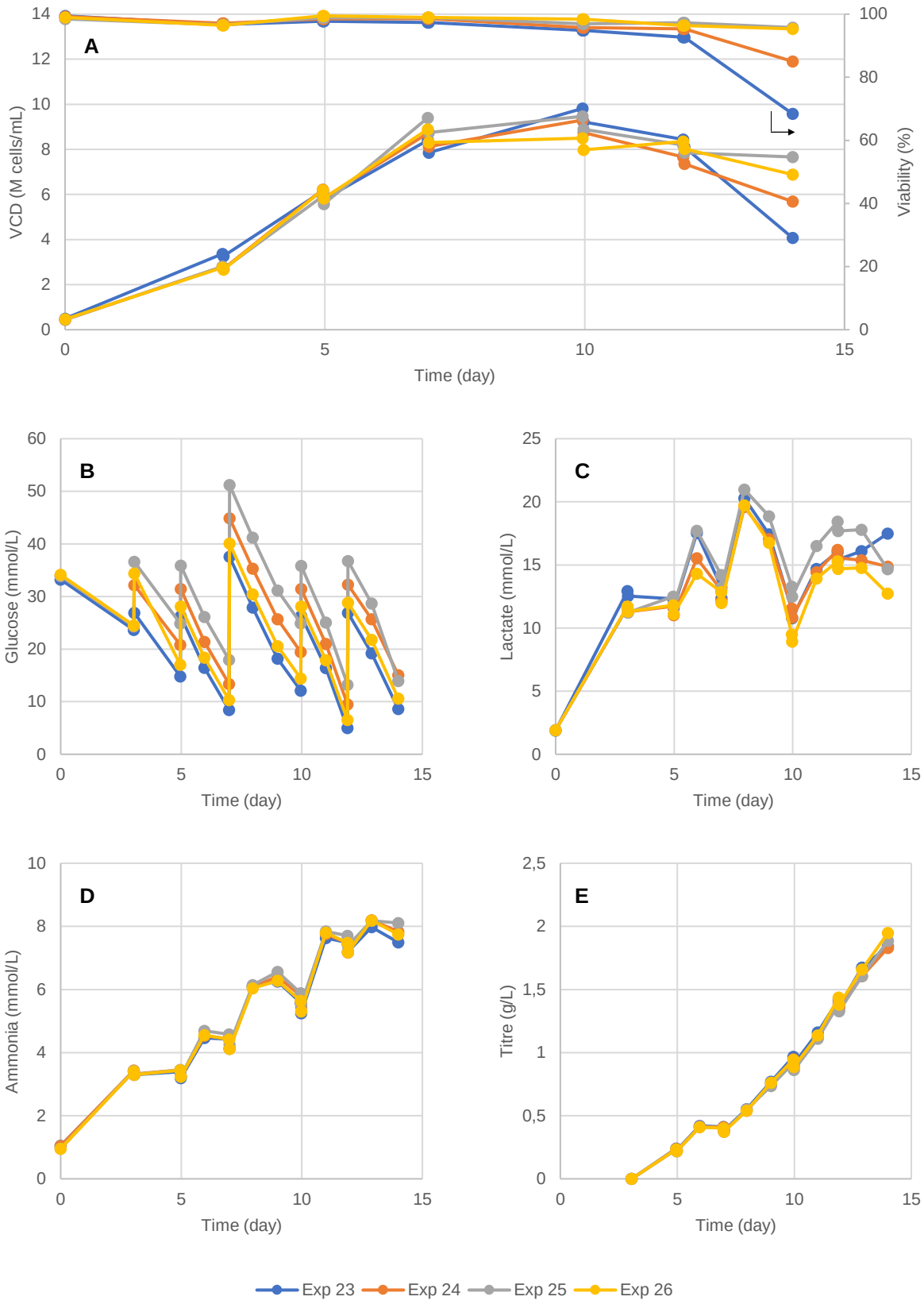

Figure 4.8 - Time-wise evolution of the monitored parameters of the ST cultures which compose 3rd batch with Bayesian optimisation: Viable cell density and viability (A), glucose (B), lactate (C), ammonia (D) and product (E) concentration.

4.2.1.5. 4th BayesOpt Batch

Again, the structure and hyperparameters of the surrogate model was optimised. While, in the previous batches, the structure did not change, in this case the basis function was eliminated and, interestingly, the error variance increased. The cause of the first change might be the fact that since the ranges of the search space were partly revised, the architecture of the data had some slight changes. In other words, previously unknown frontiers had to be incorporated into the model, which might make the structure of model change marginally and could also be responsible for the increase in the variance of the error.

Table 4.12 - Details of the structure of the Gaussian process used in the Bayesian optimisation workflow used in the determination of the 4th batch of experiments.

		Hyperparameters	
Kernel Function	ARD Materné 32	Length Scale 1	4.1889
		Length Scale 2	2.5714
		Length Scale 3	3.6010
		Length Scale 4	0.8045
		Amplitude	0.9066
Basis Function	None		
Error Variance	0.0039		

Leveraging this updated model, the last batch of experiments was defined (see Table 4.13). The levels of glucose were even more narrowed, in comparison with the previous set of experiments, while the cysteine/tyrosine feed suffered an increase and the asparagine one, a slight reduction in part of the experiments. In relation to the enriched feed, three of the experiments show levels similar to the ones who led to high titres in the first batch, while the remaining one explores a bit into the upper boundaries of this parameter.

Table 4.13 - Levels performed in the fourth batch of experiments designed with Bayesian optimisation. The construction of this batch of experiments was done with the extended range of percentage of total volume to feed of cysteine/tyrosine and asparagine.

Experiment	**Glucose (g/L)**	**Cysteine / Tyrosine (% total vol.)**	**Concentrated Feed (% total vol.)**	**Asparagine (% total vol.)**
27	5.00	0.47	6.00	0.78
28	5.34	0.47	1.06	0.68
29	5.52	0.47	1.00	0.69
30	5.06	0.47	1.14	0.66

As regards to the monitored parameters of the four cultures (Figure 4.10, LDH concentration evolution is available in Appendix A), it is important to point out that, due to the sharp levels of amino acid feedings, precipitation of the media components occurred in the last two days of the protocol, for all experiments (Figure 4.9). Hence, this batch was considered unsuccessful, and, as a result, the boundaries of the search space would have to be revised in further iterations.

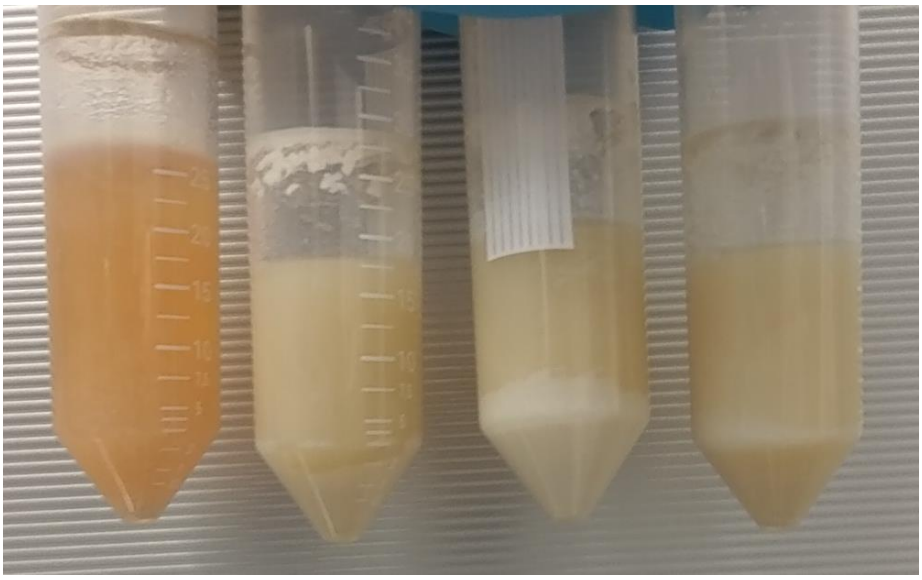

Figure 4.9 – Photography of the four spin-tubes that constituted the 4th batch with Bayesian optimisation, on the last day of culture. Precipitation of the media components occurred due to the elevated feeding levels.

Due to this event, VCD and viability was not analysed in the final day of the culture. However, in relation to these parameters, it is possible to observe that the maximum VCD levels achieved in this batch, was much lower (*circa* 6 Mcells/mL) than in previous batches. One possibility is that the prominent levels of feedings resulted in a high osmolarity and led to a inhibitory effect, impeding growth. A direct result of this lower VCD is the decrease in productivity, explaining the low levels in titre encountered in this batch.

Comparing the experiments intra-batch, the behaviour of three out of the four is very comparable (experiments 28, 29 and 30). The exception is experiment 27, which differs from the other due to the increased level in concentrated feed, leading to heightened lactate and ammonia accumulation and even lower titre levels, as mentioned previously.

The occurrence pictured in Figure 4.9, evidence of the challenge in dealing with highly concentrated feeds, is not uncommon [16]. In particular, given that cysteine is not stable at neutral pH and the solubility is of tyrosine is very low, these components can exhibit the described solubility constraints. The use of alternative compounds, such as sulfocysteine and phosphotyrosine sodium salts, has been reported and proven to preserve the performance of the cell cultures [82], [83].

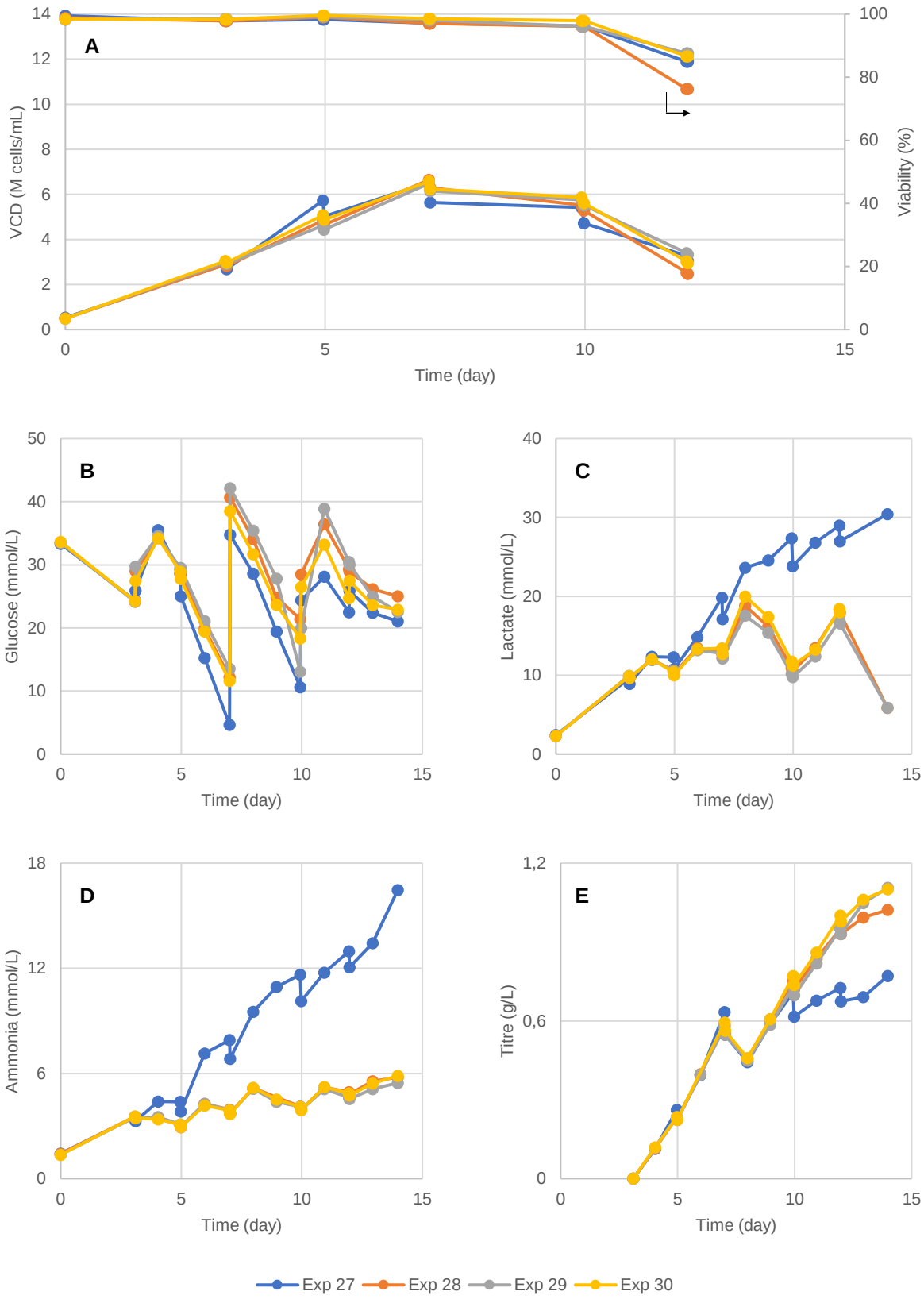

Figure 4.10 - Time-wise evolution of the monitored parameters of the ST cultures which compose 4th batch with Bayesian optimisation: Viable cell density and viability (A), glucose (B), lactate (C), ammonia (D) and product (E) concentration. VCD and viability values no shown for WD14 due to precipitation of media components, disabling the use of such analytics.

4.2.1.6. Overview

Figure 4.11 shows the variation of the 4 feeding levels across all the experiments. It should be noted that the upper ranges of the cysteine/tyrosine and asparagine feeds (Figure 4.11 C and D, respectively) were increased for the 3rd and 4th batches. It is possible to observe that the glucose feed is the most widely varied parameter. Two hypotheses arise: either this feeding was the most influential parameter or is the one to which the model attributes a higher level of uncertainty, trying to vary the levels in order to obtain a clearer correlation between titre and the feed.

Figure 4.12 exhibits the final titre results for all scale-down experiments, in a scatter-plot matrix, according to the different feeding levels. Here, it is possible to observe the experiments performed in two ways: one feed at a time (diagonal) and for the combination of two feeds (rest of the matrix). It is possible to observe, from a qualitative standpoint, that the most successful experiments occurred with moderate to high Cys/Tyr levels, low glucose and concentrated feed and moderate to high asparagine feed. Thus, coherent with the levels of experiments 15 and 16, which showed the highest titre levels (CedexBio measurement). Observing the plots in the diagonal, one can observe quantitatively that the concentrated feed possesses the highest correlation between its values and titre, while for the other components, despite existing a relation it is not such a strong one. Observing the glucose *vs.* glucose levels, high titre exists for lower glucose levels, but also for high ones. This indicates that there is not a high correlation between titre and glucose, hence the second, of the hypotheses mentioned earlier, is the most logic one.

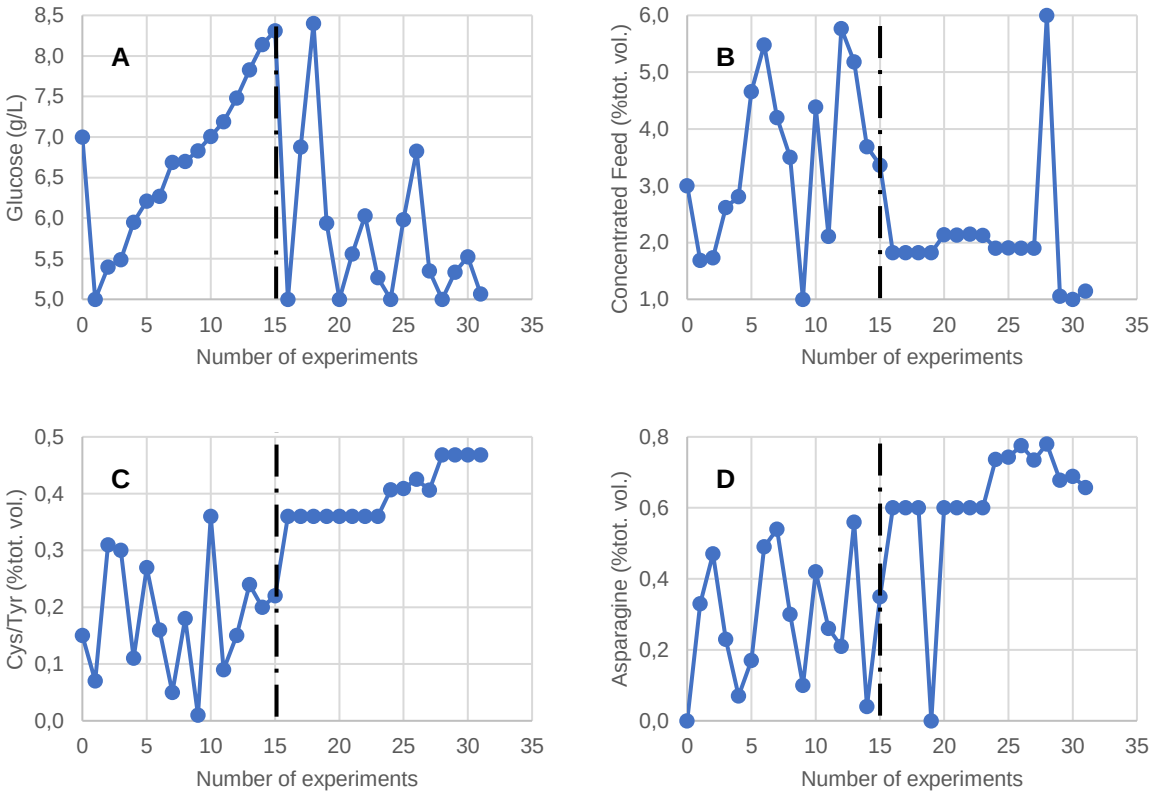

Figure 4.11 - Evolution of the feeding levels across the initialisation phase and the 4 batches defined through Bayesian optimisation. Feeds: glucose (A), concentrated (B), cysteine/tyrosine (C) and asparagine (D). Dotted vertical line indicates the transition from the initialisation to the batch-wise phase.

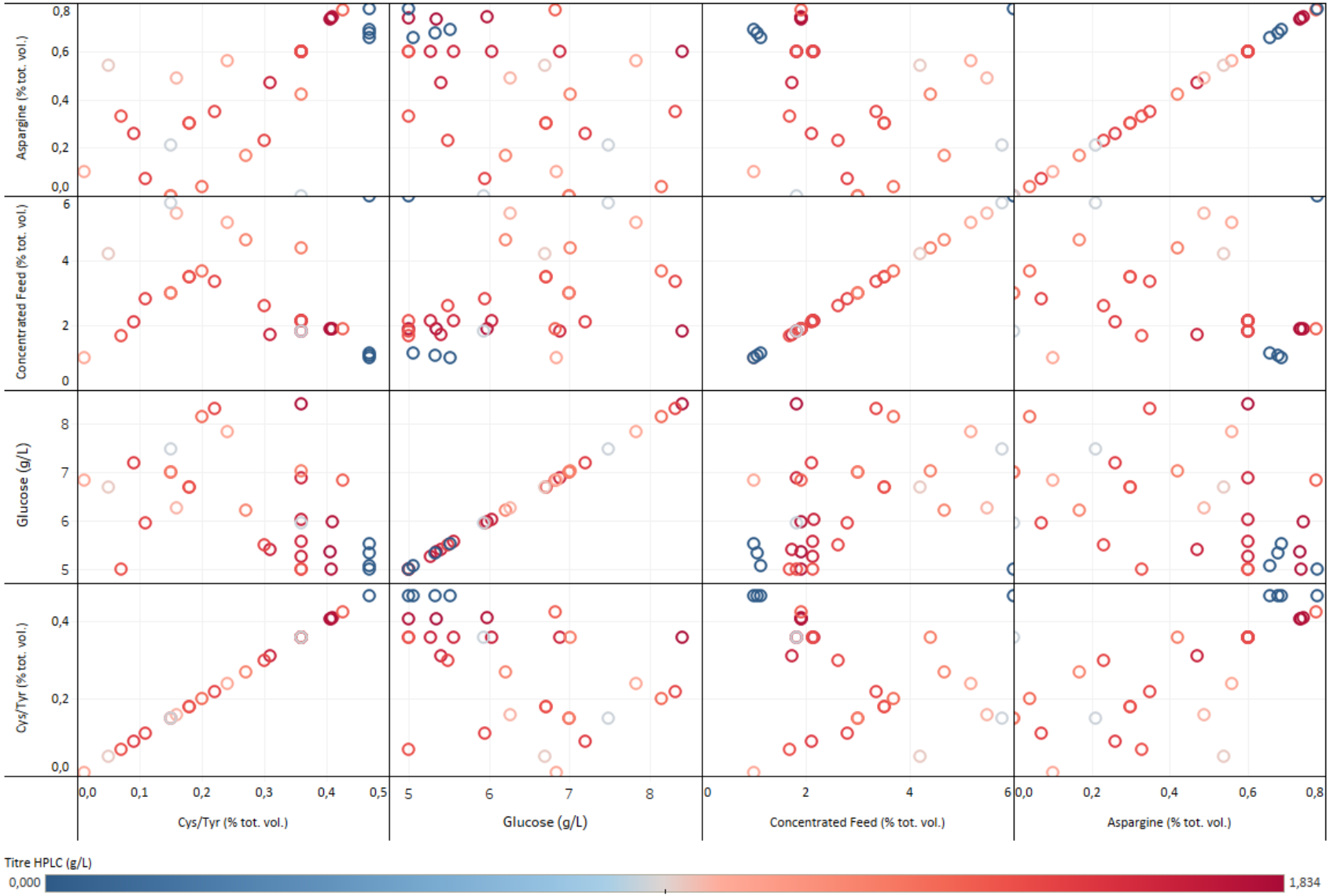

Figure 4.12 – Scatter-plot matrix of all the experiment carried out in spin-tubes, according to the different feeding levels. Titre values according to HPLC measurements.

In Figure 4.13, an overview of the titre level according to the individual feedings is shown, a different way of visualizing the data patent in the diagonal of Figure 4.12. Here it is possible to ascertain that the concentrated feed possesses the steepest correlation with titre, maximum to the lower levels of this parameter. In addition, it demonstrated that there is a limited, but direct, relation between titre and the cysteine/tyrosine and asparagine feeds. Nonetheless, for the glucose feed this relation does not seem so apparent, although titre does increase for moderate glucose levels, substantiating the previous comments.

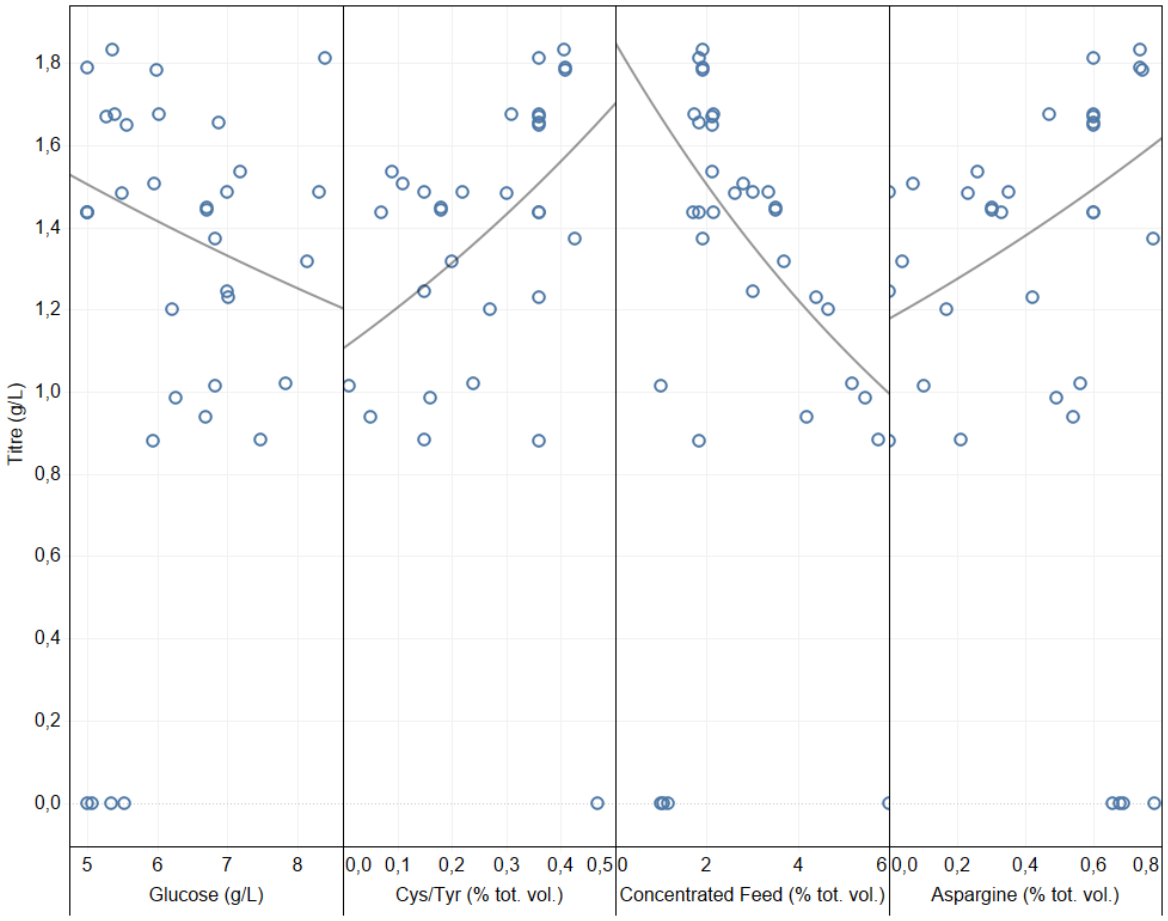

Figure 4.13 – Overview of the titre (HPLC measurement) achieved in all ST experiments, according to the individual feed levels. Trendline only valid as an assistance to the reader.

4.2.2. Benchtop bioreactor experiments

In order to validate the suitability of the optimisation process carried out in 50 mL spin-tubes, experiment 16 and the baseline process were performed in benchtop bioreactors with a 3L working volume.

In Figure 4.14 the time-evolution of different monitored parameters of the benchtop reactors cultures is shown (LDH concentration evolution is available in Appendix A). These include VCD, viability, and glucose, ammonia, target protein and lactate concentration. It is possible to observe that there is no major difference between the baseline and the tested condition. Given this conclusion there are two possible hypotheses: either the spin-tubes are not an adequate scale-down model of the benchtop bioreactors, or there was a problem with the monitoring of the spin-tubes. In the following section, a

comparison between the experiments carried out in the lab-scale bioreactors and the matching spin-tube ones is made.

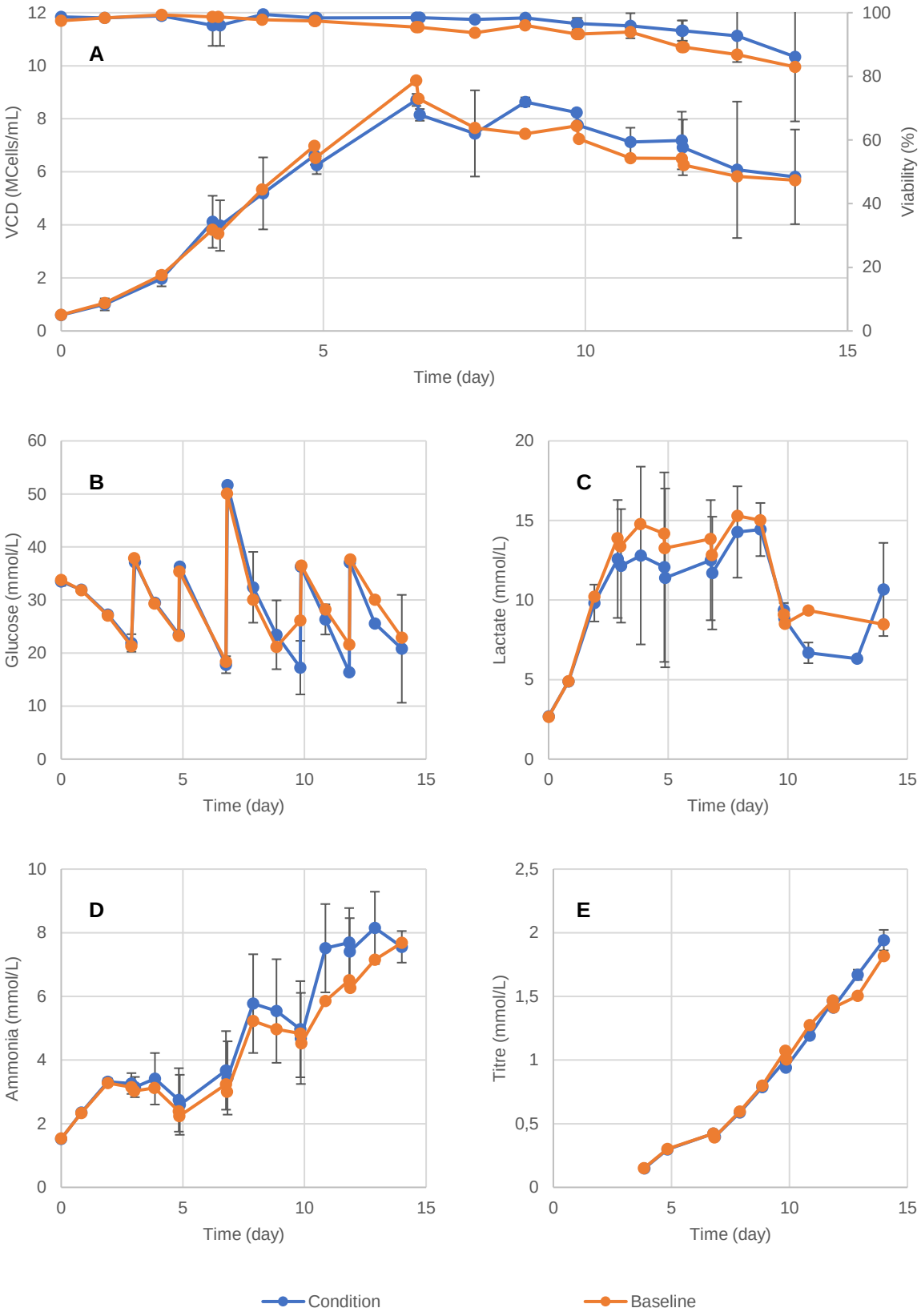

Figure 4.14 - Time-wise evolution of the monitored parameters of the benchtop bioreactor cell cultures: Viable cell density and viability (A), glucose (B), lactate (C), ammonia (D) and product (E) concentration. Error bars represent two standard deviations of n=2.

4.2.3. Comparison Spin-tube *vs.* Benchtop Bioreactors

In order to verify the validity of the 50 mL spin-tubes as scale-down models of the benchtop bioreactors the behaviour of these cultures were compared with the equivalent spin-tube cultures. Figure 4.15 shows this comparison for the baseline cultures and Figure 4.16 for the tested condition (LDH concentration evolution is available in Appendix A).

These results show that the cultures in benchtop reactors and in 50 mL spin-tubes are very comparable, asides from some slight differences in protein, lactate and glucose concentration. As mentioned earlier, the asymmetries in titre between working day six and ten are due to the change in tests in the Cedex Bio device. Lactate concentration was observed to be higher in the benchtop STRs than in the 50 mL spin-tubes. Two connected factors might have and influence in this case. Firstly, gas transfer is only done via head aeration through the existing holes on the ST cap and by way of a polytetrafluoroethylene (PTFE) membrane. This can lead to a deficient removal of carbon dioxide. The second factor is that the STs have no pH control system. Hence, the meagre carbon dioxide removal will generally originate lower pH values. In this case, cells might counterbalance the accumulation by producing less lactate to maintain pH at feasible levels [18]. The aforementioned observation was reported in literature [84], [85]. Also, in relation to the difference in lactate concentration, it is possible to observe that the shift between lactate production and consumption occurred later in the STR reactors. Although there is not an agreement in the literature [86], some have reported [87], [88] that that this change can occur when glucose is depleted, however since glucose feeding is frequent in this case, that is not the reason. However, other authors [89], [90] believe that this shift occurs concomitantly with a deviation to lower pH values. These last examples concur with the fact that since STs have no pH control, the shift occurs sooner than in the benchtop reactors. Lastly, in the benchtop reactors, from feeding to feeding, there was a higher glucose consumption. This is in agreement with the last observations, since that in the STs there was a higher lactate consumption, brought on by the decline in pH [85].

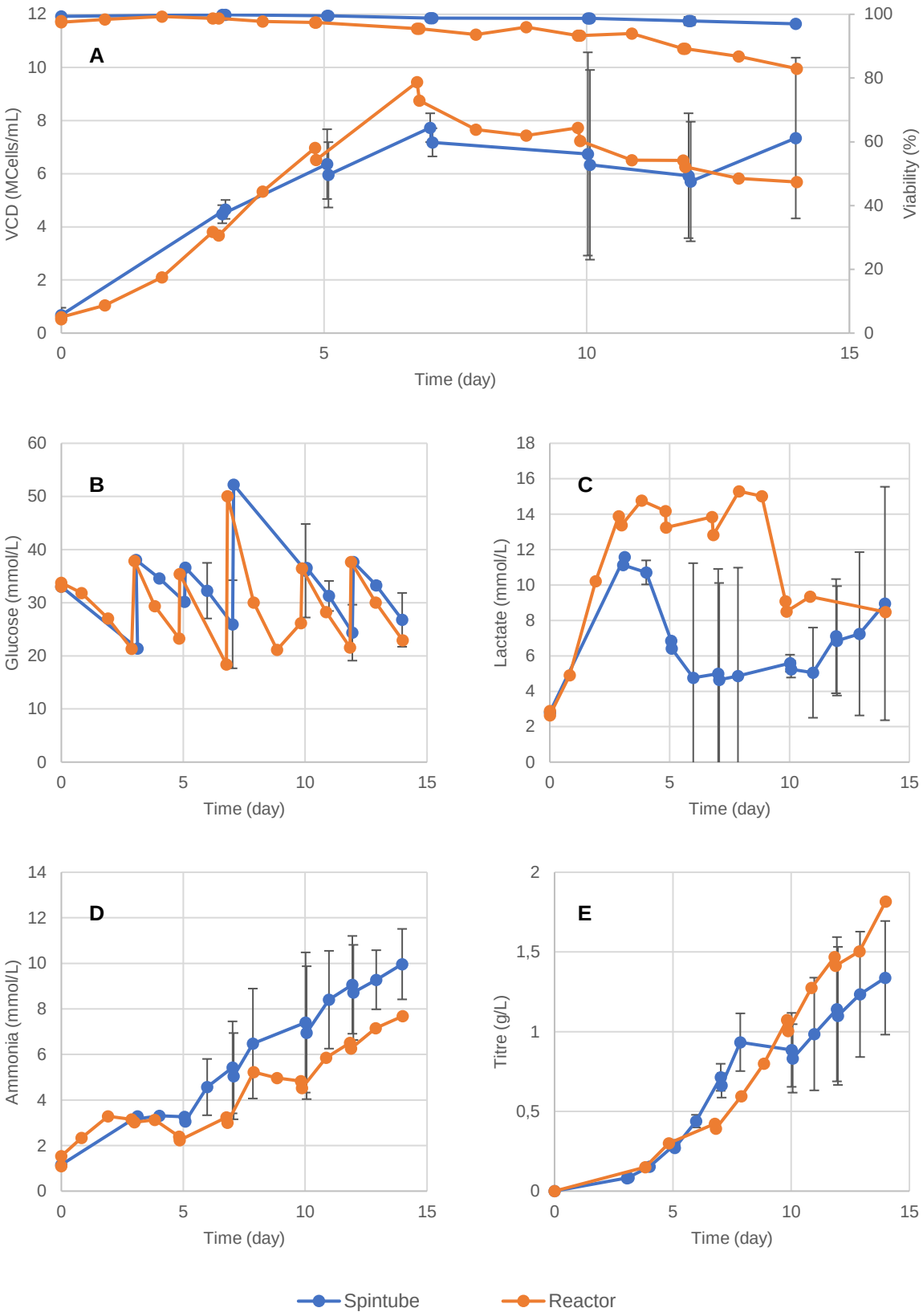

Figure 4.15 – Comparison of time-wise evolution of the different monitored parameters between the baseline spin-tube and equivalent benchtop bioreactor: Viable cell density and viability (A), glucose (B), lactate (C), ammonia (D) and product (E) concentration. Error bars represent two standard deviations of n=2.

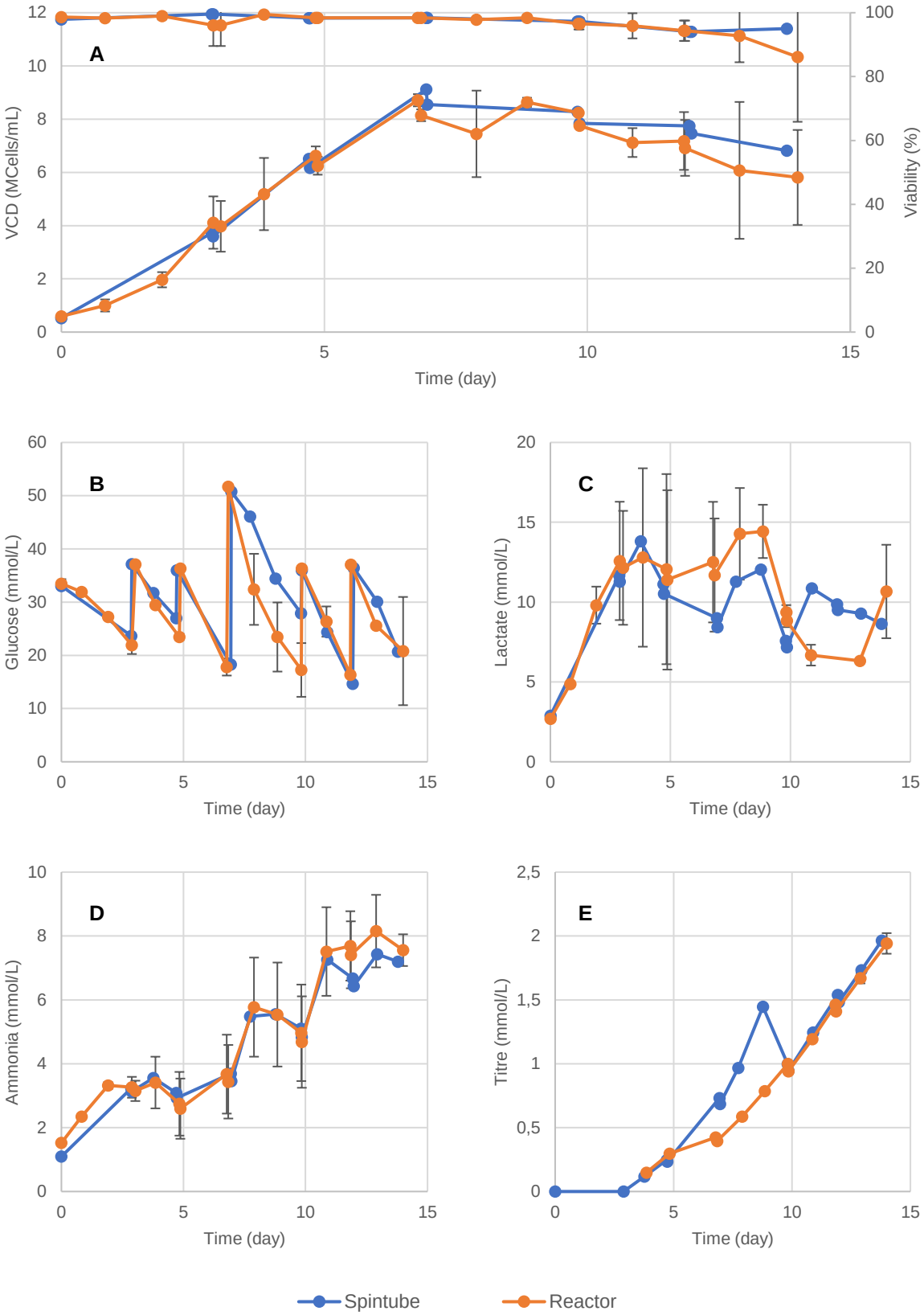

Figure 4.16 - Comparison of time-wise evolution of the different monitored parameters between the benchtop reactor where the optimal condition was applied and the equivalent spin-tube: Viable cell density and viability (A), glucose (B), lactate (C), ammonia (D) and product (E) concentration. Error bars represent two standard deviations of n=2.

4.2.4. Protein A Reference Analytics

Due to some slight discrepancies between measurements made by the Cedex Bio instrument all samples were analysed for product concentration using the reference analytical method – Protein A affinity chromatography. Results are shown in Figure 4.17.

Cedex Bio measurements of product concentration have a root mean square error (RMSE) of 0.19 g/L, in most cases, and especially at high concentrations, overestimating the actual results. Since the analytic method is turbidimetric based (imunoturbidimetric assay) [91], the precipitation behaviour of the target protein can be different than the one for which the method was validated and can even depend on the existing background. Given that the determination of the Bayesian optimisation experimental batches was made on these results, a high level of uncertainty was introduced into the algorithm. In specific terms, the used surrogate model had then a high error, subsequently deciding with high uncertainty on the next experiments. In addition, the experimental condition tested in the benchtop bioreactor also was selected through Cedex Bio measurements.

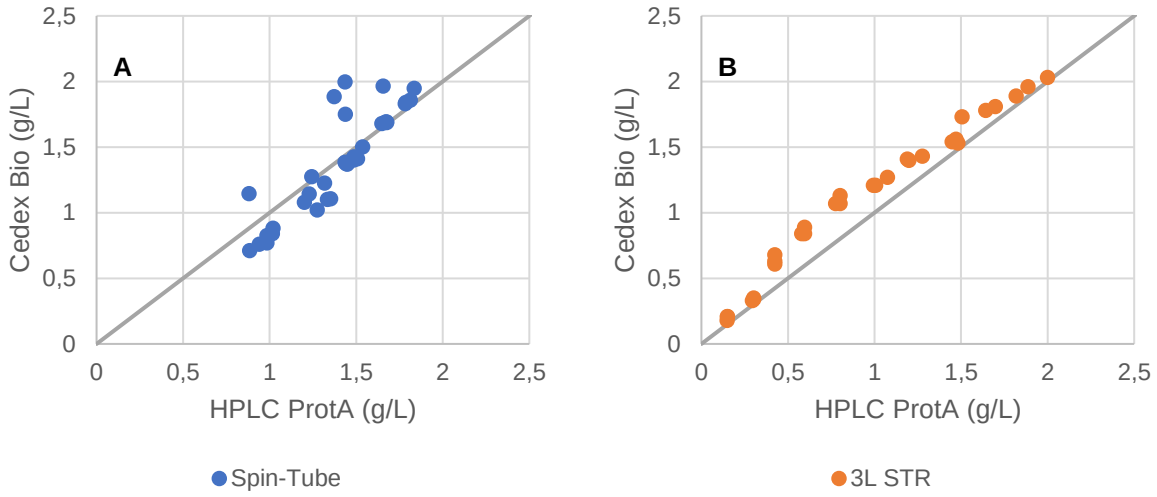

Figure 4.17 – Cedex Bio titre measurements vs. Protein A affinity chromatography reference analytics for ST (A) and 3L STR (B). Cedex Bio IgG analytics shows a root mean square error (RMSE) of 0.19 g/L.

In Figure 4.18 is shown both the end titre for each one of 28 spin-tube experiments made and its improvement across experiments. Comparing to the baseline process, Bayesian optimisation led to an increase in titre of 35%. However, given that the predictions were made resorting to Cedex Bio measurements, introducing error into the predictions one could expect even higher gains if the optimisation was done with more accurate and precise analytics.

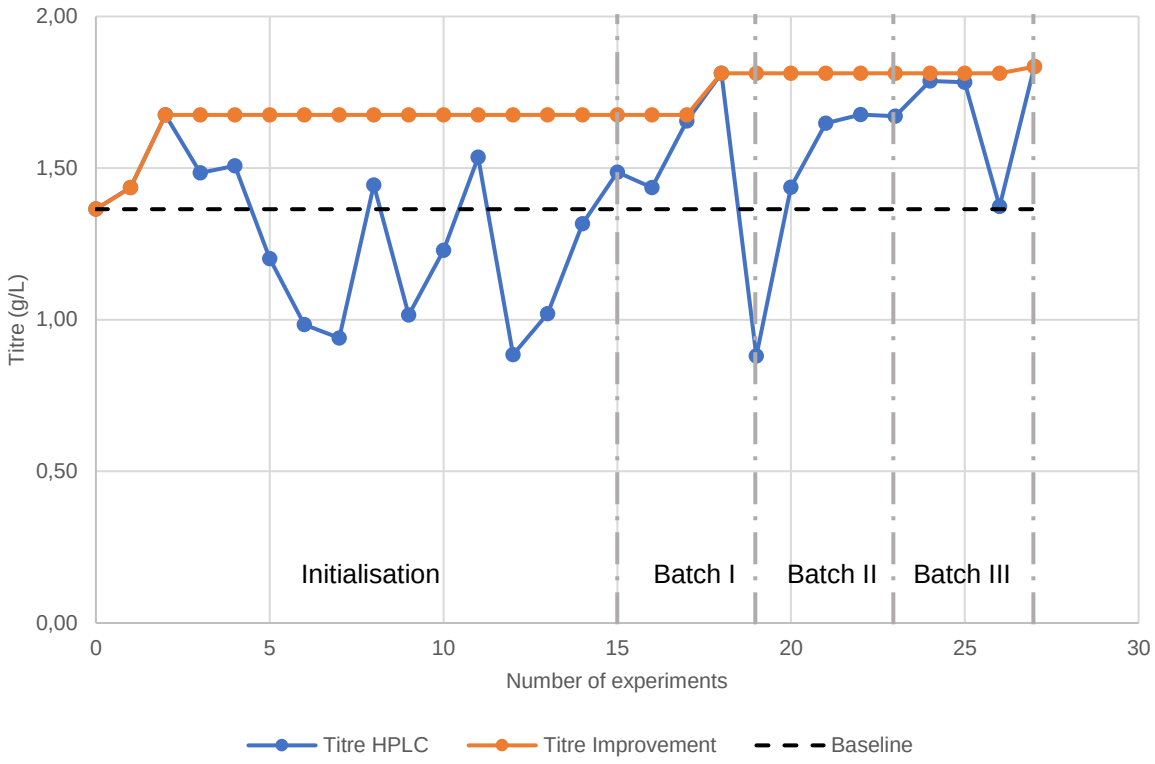

Figure 4.18 – Titre at end of the spin-tube cultures, measured with HPLC. In blue is shown the end titre for each of the 28 experiments made and in orange the maximum titre achieved vs. experiments made. Experiments 29 – 32 not shown due to precipitate formation in the end of such cultures. Experiment 0 represent the baseline process. Improvement from baseline to experiment 28 of 35%.

5. Conclusions and Future Work

Bayesian optimisation, a machine learning algorithm, was implemented and tested in both *in-silico* and wet-lab approaches, developed towards the media optimisation of an existing fed-batch protocol for a CHO cell line, producing a mAb derived protein. The perfect setting to observe the advantages of these advanced techniques, as it is a highly laborious process.

Firstly, several Bayesian optimisation variations were implemented and compared through computer simulations, using an *in silico* model of a fed-batch process. The best options were defined as: Latin hypercube sampling as initialisation method and expected improvement as acquisition function. Further, batch Bayesian optimisation via local penalisation [66] was implemented. Simulations showed that, given the convenience it possesses, the loss in comparison with the sequential policy is not significant. Both policies outperformed common baseline algorithms.

Based on an existing, un-optimised, fed-batch protocol of a CHO cell line producing a therapeutic protein the algorithm was tested. Specifically, the objective was to try to find the feed strategy among 4 different feeds which produced the highest titre at the end of the culture. Sixteen experiments were done as an initialisation phase and a further 4 batches of 4 experiments each were performed in 50 mL spin-tubes. These 31 experiments lead to the increase of 35% in titre, against the baseline process. Finally, the suitability of the scale-down models was confirmed in 3L benchtop bioreactors.

It is worth mentioning that this increase was achieved even with an operating problem with the at-line analytics, only discovered after the experiments had been carried out. Theoretically, Cedex Bio measurements are already pre-validated against reference analytics. Which were the same used in the present work: protein A affinity chromatography. In addition, the manufacturer indicates that no sample filtration or pre-treatment is required for the IgG assay [91]. Also, these tests have already been used with success in-house for other processes. However, as stated ICH's Q2 guideline [92], since the composition of the samples might be different in the current process, the analytic method should have been revalidated, in hindsight.

Hence, in future work the 4 sets of experiments defined through Bayesian optimisation will have to be repeated and defined based on HPLC titre measurements. Based on these new results, the best feeding strategy will be transferred for the design of a suitable medium composition for perfusion cultures. Here, concerns on the osmolarity of the media will have to be addressed and suitable operating parameters will be found through the use of perfusion scale-down models [18].

In algorithmic terms, there are some additions that could be made in order to even shorten the development time-scale further. Freeze-Thaw Bayesian optimisation [93] is a modification that introduces the concept of stopping an experiment that does not seem promising, or in other words that is "less optimal" than a previous experiment. This is important when evaluating expensive functions since conducting not profitable experiments to the end does not add value to the optimisation problem at hand [42]. An implementation of this could be done through the use of a step-ahead Gaussian process, in order to predict the next state of the culture based on the previous one and in turn the end state. It would then be possible to decide whether to continue with the experiment or not, or even do overlapping batches. In addition, since the structure of the surrogate model is with certainty the most

critical part of the Bayesian optimisation algorithm, it is possible to improve this, by trying other types of models, or even through the incorporation of deterministic knowledge.

It should be also pointed out that soon, through the development of high-throughput scale-down systems, such as deep-well plates [94] and AMBR (automated micro bioreactors) for perfusion [95], it will be possible to apply the developed algorithm directly in optimising media for perfusion without doing screening in fed-batch. Although scale down models for perfusion already exist, the use of these is still highly work-intensive, involving the exchange of the culture media on a daily basis [18].

References

[1] A. L. Grilo and A. Mantalaris, “The Increasingly Human and Profitable Monoclonal Antibody Market,” *Trends Biotechnol.*, Jun. 2018.

[2] D. M. Ecker, S. D. Jones, and H. L. Levine, “The therapeutic monoclonal antibody market.,” *MAbs*, vol. 7, no. 1, pp. 9–14, 2015.

[3] V. Warikoo *et al.*, “Integrated continuous production of recombinant therapeutic proteins,” *Biotechnol. Bioeng.*, vol. 109, no. 12, pp. 3018–3029, Dec. 2012.

[4] S. Xu, J. Gavin, R. Jiang, and H. Chen, “Bioreactor productivity and media cost comparison for different intensified cell culture processes,” *Biotechnol. Prog.*, vol. 33, no. 4, pp. 867–878, Jul. 2017.

[5] P. Soelkner, “A Changing Industry Brings New Challenges, but also Opportunities,” *PharmTech: Advancing Development and Manufacturing*, 2018. [Online]. Available: http://www.pharmtech.com/changing-industry-brings-new-challenges-also-opportunities. [Accessed: 01-Nov-2018].

[6] R. Otto, A. Santagostino, and U. Schrader, “Rapid growth in biopharma: Challenges and opportunities,” *McKinsey & Company*, 2014. [Online]. Available: https://www.mckinsey.com/industries/pharmaceuticals-and-medical-products/our-insights/rapid-growth-in-biopharma. [Accessed: 13-Mar-2018].

[7] J. Sterling, “Key Trends in 2018 in the Biopharmaceutical Market,” *Genetic Engineering & Biotechnology News*, 2018. [Online]. Available: https://www.genengnews.com/gen-exclusives/key-trends-in-2018-in-the-biopharmaceutical-market/77901038. [Accessed: 01-Nov-2018].

[8] M. S. Croughan, K. B. Konstantinov, and C. Cooney, “The future of industrial bioprocessing: Batch or continuous?,” *Biotechnol. Bioeng.*, vol. 112, no. 4, pp. 648–651, Apr. 2015.

[9] D. J. Karst, F. Steinebach, and M. Morbidelli, “Continuous integrated manufacturing of therapeutic proteins,” *Curr. Opin. Biotechnol.*, vol. 53, pp. 76–84, 2018.

[10] K. B. Konstantinov and C. L. Cooney, “White paper on continuous bioprocessing May 20-21, 2014 continuous manufacturing symposium,” *J. Pharm. Sci.*, vol. 104, no. 3, pp. 813–820, 2015.

[11] W.-S. Hu, “Overview of Cell Culture Technology,” in *Cell Culture Bioprocess Engineering*, 2012.

[12] K. B. Konstantinov, Y. -s. Tsai, D. Moles, and R. Matanguihan, “Control of Long-Term Perfusion Chinese Hamster Ovary Cell Culture by Glucose Auxostat,” *Biotechnol. Prog.*, vol. 12, no. 1, pp. 100–109, Feb. 1996.

[13] J. Bonham-Carter and J. Shevitz, “A Brief History of Perfusion Biomanufacturing,” *BioProcess International*, 2011. [Online]. Available: http://www.bioprocessintl.com/upstream-processing/bioreactors/a-brief-history-of-perfusion-biomanufacturing-322322/. [Accessed: 07-Mar-2018].

[14] D. J. Karst, E. Serra, T. K. Villiger, M. Soos, and M. Morbidelli, “Characterization and comparison of ATF and TFF in stirred bioreactors for continuous mammalian cell culture processes,”

Biochem. Eng. J., vol. 110, pp. 17–26, Jun. 2016.

[15] W.-S. Hu, “Cell Retention and Perfusion,” in *Cell Culture Bioprocess Engineering*, Wei-Shou Hu, 2012, pp. 249–262.

[16] J.-M. Bielser, M. Wolf, J. Souquet, H. Broly, and M. Morbidelli, “Perfusion mammalian cell culture for recombinant protein manufacturing – A critical review,” *Biotechnol. Adv.*, vol. 36, no. 4, pp. 1328–1340, Jul. 2018.

[17] J. Walther, R. Godawat, C. Hwang, Y. Abe, and A. Sinclair, “The business impact of an integrated continuous biomanufacturing platform for recombinant protein production,” *J. Biotechnol.*, vol. 213, pp. 3–12, Nov. 2015.

[18] M. K. F. Wolf, V. Lorenz, D. J. Karst, J. Souquet, H. Broly, and M. Morbidelli, “Development of a shake-tube-based scale-down model for perfusion cultures,” *Biotechnol. Bioeng.*, Jul. 2018.

[19] H. Lin, R. W. Leighty, S. Godfrey, and S. B. Wang, “Principles and approach to developing mammalian cell culture media for high cell density perfusion process leveraging established fed-batch media,” *Biotechnol. Prog.*, vol. 33, no. 4, pp. 891–901, 2017.

[20] J. Hurwitz and D. Kirsch, *Machine Learning for Dummies*, IBM Limite. Hoboken, NJ, USA: John Wiley & Sons, Inc, 2018.

[21] G. Press, “These Banks Are Using AI To Help Their Customers Manage Their Finances,” *Forbes*, 2018. [Online]. Available: https://www.forbes.com/sites/gilpress/2018/09/12/these-banks-are-using-ai-to-help-their-customers-manage-their-finances/#. [Accessed: 28-Oct-2018].

[22] M. Kasser, A. Padhi, A. Tschiesner, and W. Dominik, “Artificial intelligence as auto companies’ new engine of value,” *McKinsey & Company*, 2018. [Online]. Available: https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/artificial-intelligence-as-auto-companies-new-engine-of-value. [Accessed: 01-Nov-2018].

[23] G. Batra, A. Queirolo, and N. Santhanam, “Artificial intelligence: The time to act is now,” *McKinsey & Company*, 2018. [Online]. Available: https://www.mckinsey.com/industries/advanced-electronics/our-insights/artificial-intelligence-the-time-to-act-is-now. [Accessed: 01-Nov-2018].

[24] P. Gerbert, M. Hecker, Se. Steinhauser, and P. Ruwolt, “Putting Artificial Intelligence to Work,” *The Boston Consulting Group*, 2017. [Online]. Available: https://www.bcg.com/publications/2017/technology-digital-strategy-putting-artificial-intelligence-work.aspx. [Accessed: 28-Oct-2018].

[25] B. Hirschler, “Big pharma turns to AI to speed drug discovery, GSK signs deal,” *Reuters*, 2017. [Online]. Available: https://www.reuters.com/article/us-pharmaceuticals-ai-gsk/big-pharma-turns-to-ai-to-speed-drug-discovery-gsk-signs-deal-idUSKBN19N003. [Accessed: 28-Oct-2018].

[26] A. Sosna, “Machine learning finds purpose in commercial life sciences,” *Pharmaceutial Commerce*, 2018. [Online]. Available: http://pharmaceuticalcommerce.com/information-technology/machine-learning-finds-purpose-in-commercial-life-sciences/. [Accessed: 28-Oct-2018].

[27] D. R. Cox and N. Reid, *The Theory of the Design of Experiments*. Chapman & Hall/CRC, 2000.

[28] C.-F. Mandenius and A. Brundin, “Bioprocess optimization using design-of-experiments methodology,” *Biotechnol. Prog.*, vol. 24, no. 6, pp. 1191–1203, Nov. 2008.

[29] I. J. González-Leal *et al.*, “Use of a Plackett-Burman statistical design to determine the effect of selected amino acids on monoclonal antibody production in CHO cells,” *Biotechnol. Prog.*, vol. 27, no. 6, pp. 1709–1717, Nov. 2011.

[30] Y. Rouiller, A. Périlleux, N. Collet, M. Jordan, M. Stettler, and H. Broly, “A high-throughput media design approach for high performance mammalian fed-batch cultures,” *MAbs*, vol. 5, no. 3, pp. 501–511, May 2013.

[31] W. L. W. Ling, Y. Bai, C. Cheng, I. Padawer, and C. Wu, “Development and manufacturability assessment of chemically-defined medium for the production of protein therapeutics in CHO cells,” *Biotechnol. Prog.*, vol. 31, no. 5, pp. 1163–1171, Sep. 2015.

[32] H. Robbins, “Some Aspects of the Sequential Design of Experiments,” *Bull. Am. Math. Soc.*, vol. 58, no. 5, pp. 527–535, 1952.

[33] H. Chernoff, “Sequential Design of Experiments,” *Ann. Math. Stat.*, vol. 30, no. 3, pp. 755–770, 1959.

[34] A. Geiges, Y. Rubin, and W. Nowak, “Interactive design of experiments: A priori global versus sequential optimization, revised under changing states of knowledge,” *Water Resour. Res.*, vol. 51, no. 10, pp. 7915–7936, Oct. 2015.

[35] L. Freier, J. Hemmerich, K. Schöler, W. Wiechert, M. Oldiges, and E. von Lieres, “Framework for Kriging-based iterative experimental analysis and design: Optimization of secretory protein production in Corynebacterium glutamicum,” *Eng. Life Sci.*, vol. 16, no. 6, pp. 538–549, Sep. 2016.

[36] D. Brühlmann *et al.*, “Parallel experimental design and multivariate analysis provides efficient screening of cell culture media supplements to improve biosimilar product quality,” *Biotechnol. Bioeng.*, vol. 114, no. 7, pp. 1448–1458, Jul. 2017.

[37] G. Venter, Venter, and Gerhard, “Review of Optimization Techniques,” in *Encyclopedia of Aerospace Engineering*, Chichester, UK: John Wiley & Sons, Ltd, 2010.

[38] M. N. Ab Wahab, S. Nefti-Meziani, and A. Atyabi, “A comprehensive review of swarm optimization algorithms.,” *PLoS One*, vol. 10, no. 5, p. e0122827, 2015.

[39] X. Zheng, J. Ishikawa, T. Sugiyama, and Y. Maruyama, “Bayesian Optimization Analysis of Containment-Venting Operation in a Boiling Water Reactor Severe Accident,” *Nucl. Eng. Technol.*, vol. 49, no. 2, pp. 434–441, Mar. 2017.

[40] C. D. Lin, C. M. Anderson-Cook, M. S. Hamada, L. M. Moore, and R. R. Sitter, “Using Genetic Algorithms to Design Experiments: A Review,” *Qual. Reliab. Eng. Int.*, vol. 31, no. 2, pp. 155–167, Mar. 2015.

[41] H. J. Kushner, “A New Method of Locating the Maximum Point of an Arbitrary Multipeak Curve in the Presence of Noise,” *J. Basic Eng.*, vol. 86, no. 1, p. 97, Mar. 1964.

[42] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas, “Taking the Human Out of

the Loop: A Review of Bayesian Optimization," *Proc. IEEE*, vol. 104, no. 1, pp. 148–175, Jan. 2016.

[43] J. González, "Introduction to Bayesian Optimization," *Lancaster University*, 2017. [Online]. Available: gpss.cc/gpmc17/slides/LancasterMasterclass_1.pdf. [Accessed: 02-Mar-2018].

[44] J. Antony, "Fractional factorial designs," in *Design of Experiments for Engineers and Scientists*, Butterworth-Heinemann, 2003, pp. 73–92.

[45] The MathWorks Inc., "Fractional Factorial Designs," 2018. [Online]. Available: https://ch.mathworks.com/help/stats/fractional-factorial-designs.html. [Accessed: 27-Oct-2018].

[46] NIST/SEMATECH, "Process Improvement - NIST/SEMATECH e-Handbook of Statistical Methods," 2012. [Online]. Available: https://www.itl.nist.gov/div898/handbook/pri/section3/pri334.htm. [Accessed: 25-Oct-2018].

[47] F. A. C. Viana, "A Tutorial on Latin Hypercube Design of Experiments," *Qual. Reliab. Eng. Int.*, vol. 32, no. 5, pp. 1975–1985, Jul. 2016.

[48] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[49] A. Shah, A. G. Wilson, and Z. Ghahramani, "Student-t Processes as Alternatives to Gaussian Processes," in *Proceedings of the 17th International Conference on Articial Intelligence and Statistics (AISTATS)*, 2014, vol. 33.

[50] R. Bardenet and B. Kégl, "Surrogating the surrogate: accelerating Gaussian process-based global optimization with a mixture cross-entropy algorithm," in *Proceedings of the 26th International Conference on Machine Learning*, 2010.

[51] E. Brochu, V. M. Cora, and N. de Freitas, "A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modelling and Hierarchical Reinforcement Learning," *Dept. Comput. Sci., Univ. Br. Columbia, Vancouver, BC, Canada*, 2009.

[52] Cdipaolo96, "Gaussian Process Regression," *Wikipedia*, 2016. [Online]. Available: https://en.wikipedia.org/wiki/File:Gaussian_Process_Regression.png.

[53] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, 1st Editio. Cambridge, Massachusetts: MIT Press, 2006.

[54] The MathWorks Inc., "Gaussian Process Regression Models," 2018. [Online]. Available: https://www.mathworks.com/help/stats/gaussian-process-regression-models.html. [Accessed: 23-Oct-2018].

[55] M. L. Stein, *Interpolation of Spatial Data: Some Theory for Kriging*. New York, NY: Springer New York, 1999.

[56] B. Matérn, *Spatial Variation*, 2nd Editio. Springer-Verlag New York, 1960.

[57] M. W. Hoffman, B. Shahriari, and N. de Freitas, "Exploiting correlation and budget constraints in Bayesian multi-armed bandit optimization," Mar. 2013.

[58] M. Hoffman, E. Brochu, and N. de Freitas, "Portfolio allocation for Bayesian optimization," in *Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence*, 2011, pp. 327–336.

[59] D. R. Jones, M. Schonlau, and W. J. Welch, “Efficient Global Optimization of Expensive Black-Box Functions,” *J. Glob. Optim.*, vol. 13, no. 4, pp. 455–492, 1998.

[60] D. J. Lizotte, “Practical Bayesian optimization,” Library and Archives Canada, Bibliothèque et Archives Canada, 2008.

[61] N. Srinivas, A. Krause, S. Kakade, and M. Seeger, “Gaussian process optimization in the bandit setting: no regret and experimental design,” in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, 2010, pp. 1015–1022.

[62] D. R. Jones, C. D. Perttunen, and B. E. Stuckman, “Lipschitzian optimization without the Lipschitz constant,” *J. Optim. Theory Appl.*, vol. 79, no. 1, pp. 157–181, Oct. 1993.

[63] P. Morere, R. Marchant, and F. Ramos, “Sequential Bayesian optimization as a POMDP for environment monitoring with UAVs,” in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 6381–6388.

[64] C. Chevalier and D. Ginsbourger, “Fast Computation of the Multi-Points Expected Improvement with Applications in Batch Selection,” in *LION 2013: Learning and Intelligent Optimization. Lecture Notes in Computer Science, vol 7997*, G. Nicosia and P. Pardalos, Eds. Springer, Berlin, Heidelberg, 2013, pp. 59–69.

[65] J. Azimi, A. Fern, and X. Z. Fern, “Batch Bayesian Optimization via Simulation Matching,” in *Advances in Neural Information Processing Systems 23*, J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, and A. Culotta, Eds. Curran Associates, Inc., 2010, pp. 109–117.

[66] J. González, Z. Dai, P. Hennig, and N. D. Lawrence, *Batch Bayesian Optimization via Local Penalization*. 2015.

[67] J. Snoek *et al.*, “Scalable Bayesian Optimization Using Deep Neural Networks,” *Proc. Mach. Learn. Res.*, vol. 37, pp. 2171–2180, Jun. 2015.

[68] M. Mehrian *et al.*, “Maximizing neotissue growth kinetics in a perfusion bioreactor: An in silico strategy using model reduction and Bayesian optimization,” *Biotechnol. Bioeng.*, vol. 115, no. 3, pp. 617–629, Mar. 2018.

[69] S. Olofsson, M. Mehrian, L. Geris, R. Calandra, M. P. Deisenroth, and R. Misener, “Bayesian Multi-Objective Optimisation of Neotissue Growth in a Perfusion Bioreactor Set-Up,” *Comput. Aided Chem. Eng.*, vol. 40, pp. 2155–2160, Jan. 2017.

[70] D. Ulmasov, C. Baroukh, B. Chachuat, M. P. Deisenroth, and R. Misener, “Bayesian Optimization with Dimension Scheduling: Application to Biological Systems,” *Comput. Aided Chem. Eng.*, vol. 38, pp. 1051–1056, Jan. 2016.

[71] S. Craven, N. Shirsat, J. Whelan, and B. Glennon, “Process model comparison and transferability across bioreactor scales and modes of operation for a mammalian cell bioprocess,” *Biotechnol. Prog.*, vol. 29, no. 1, pp. 186–196, Jan. 2013.

[72] M. Gadgil, “Development of a mathematical model for animal cell culture without pH control and its application for evaluation of clone screening outcomes in shake flask culture,” *J. Chem. Technol. Biotechnol.*, vol. 90, no. 1, pp. 166–175, Jan. 2015.

[73] A. Villiger-Oberbek, Y. Yang, W. Zhou, and J. Yang, “Development and application of a high-

throughput platform for perfusion-based cell culture processes," *J. Biotechnol.*, vol. 212, pp. 21–29, Oct. 2015.

[74] B. Neunstoecklin, "Criteria for scale-up and scale-down of bioreactors for cultivation of mammalian cells," ETH Zurich, 2014.

[75] Z. B. Zabinsky, "Random Search Algorithms," in *Wiley Encyclopedia of Operations Research and Management Science*, Hoboken, NJ, USA: John Wiley & Sons, Inc., 2011.

[76] X. Pan, M. Streefland, C. Dalm, R. H. Wijffels, and D. E. Martens, "Selection of chemically defined media for CHO cell fed-batch culture processes.," *Cytotechnology*, vol. 69, no. 1, pp. 39–56, Feb. 2017.

[77] S. Pereira, H. F. Kildegaard, and M. R. Andersen, "Impact of CHO Metabolism on Cell Growth and Protein Production: An Overview of Toxic and Inhibiting Metabolites and Nutrients," *Biotechnol. J.*, vol. 13, no. 3, p. 1700499, Mar. 2018.

[78] J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian Optimization of Machine Learning Algorithms," 13-Jun-2012. [Online]. Available: https://arxiv.org/abs/1206.2944. [Accessed: 23-Oct-2018].

[79] P. Xu, X.-P. Dai, E. Graf, R. Martel, and R. Russell, "Effects of glutamine and asparagine on recombinant antibody production using CHO-GS cell lines," *Biotechnol. Prog.*, vol. 30, no. 6, pp. 1457–1468, Nov. 2014.

[80] T. M. Duarte, N. Carinhas, L. C. Barreiro, M. J. T. Carrondo, P. M. Alves, and A. P. Teixeira, "Metabolic responses of CHO cells to limitation of key amino acids," *Biotechnol. Bioeng.*, vol. 111, no. 10, pp. 2095–2106, Oct. 2014.

[81] L. Zhang *et al.*, "Responses of CHO-DHFR cells to ratio of asparagine to glutamine in feed media: cell growth, antibody production, metabolic waste, glutamate, and energy metabolism," *Bioresour. Bioprocess.*, vol. 3, no. 1, p. 5, Dec. 2016.

[82] A. Zimmer, R. Mueller, M. Wehsling, A. Schnellbaecher, and J. von Hagen, "Improvement and simplification of fed-batch bioprocesses with a highly soluble phosphotyrosine sodium salt," *J. Biotechnol.*, vol. 186, pp. 110–118, Sep. 2014.

[83] C. Hecklau *et al.*, "S-Sulfocysteine simplifies fed-batch processes and increases the CHO specific productivity via anti-oxidant activity," *J. Biotechnol.*, vol. 218, pp. 53–63, Jan. 2016.

[84] H. Zhou, J. Purdie, T. Wang, and A. Ouyang, "pH measurement and a rational and practical pH control strategy for high throughput cell culture system," *Biotechnol. Prog.*, vol. 26, no. 3, pp. 872–880, Dec. 2009.

[85] D. T. Monteil *et al.*, "A comparison of orbitally-shaken and stirred-tank bioreactors: pH modulation and bioreactor type affect CHO cell growth and protein glycosylation," *Biotechnol. Prog.*, vol. 32, no. 5, pp. 1174–1180, Sep. 2016.

[86] F. Hartley, T. Walker, V. Chung, and K. Morten, "Mechanisms driving the lactate switch in Chinese hamster ovary cells," *Biotechnol. Bioeng.*, vol. 115, no. 8, pp. 1890–1903, Aug. 2018.

[87] V. S. Martínez, S. Dietmair, L.-E. Quek, M. P. Hodson, P. Gray, and L. K. Nielsen, "Flux balance analysis of CHO cells before and after a metabolic switch from lactate production to

consumption," *Biotechnol. Bioeng.*, vol. 110, no. 2, pp. 660–666, Feb. 2013.

[88] C. Altamirano, A. Illanes, S. Becerra, J. J. Cairó, and F. Gòdia, "Considerations on the lactate consumption by CHO cells in the presence of galactose," *J. Biotechnol.*, vol. 125, no. 4, pp. 547–556, Oct. 2006.

[89] L. Liste-Calleja, M. Lecina, J. Lopez-Repullo, J. Albiol, C. Solà, and J. J. Cairó, "Lactate and glucose concomitant consumption as a self-regulated pH detoxification mechanism in HEK293 cell cultures," *Appl. Microbiol. Biotechnol.*, vol. 99, no. 23, pp. 9951–9960, Dec. 2015.

[90] D. Zalai, K. Koczka, L. Párta, P. Wechselberger, T. Klein, and C. Herwig, "Combining mechanistic and data-driven approaches to gain process knowledge on the control of the metabolic shift to lactate uptake in a fed-batch CHO process," *Biotechnol. Prog.*, vol. 31, no. 6, pp. 1657–1668, Nov. 2015.

[91] Roche Diagnostics, *Cedex Bio Analyser: Operator's Manual*, Version 2. 2016.

[92] International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, *Q2 - Validation of Analytical Prodedures: Text and Methodology*. 1994, pp. 1–17.

[93] K. Swersky, J. Snoek, and R. P. Adams, "Freeze-Thaw Bayesian Optimization," 15-Jun-2014. [Online]. Available: http://arxiv.org/abs/1406.3896. [Accessed: 06-Mar-2018].

[94] J.-M. Bielser, J. Domaradzki, J. Souquet, M. Morbidelli, and H. Broly, "Semi-continuous scale-down models for clone and operating parameter screening in perfusion bioreactors," *Submiss.*, 2018.

[95] B. Zoro and A. Tait, "Development of a novel automated perfusion mini bioreactor 'ambr® 250 perfusion," in *Integrated Continuous Biomanufacturing III*, 2017.

Appendix A: LDH Monitoring

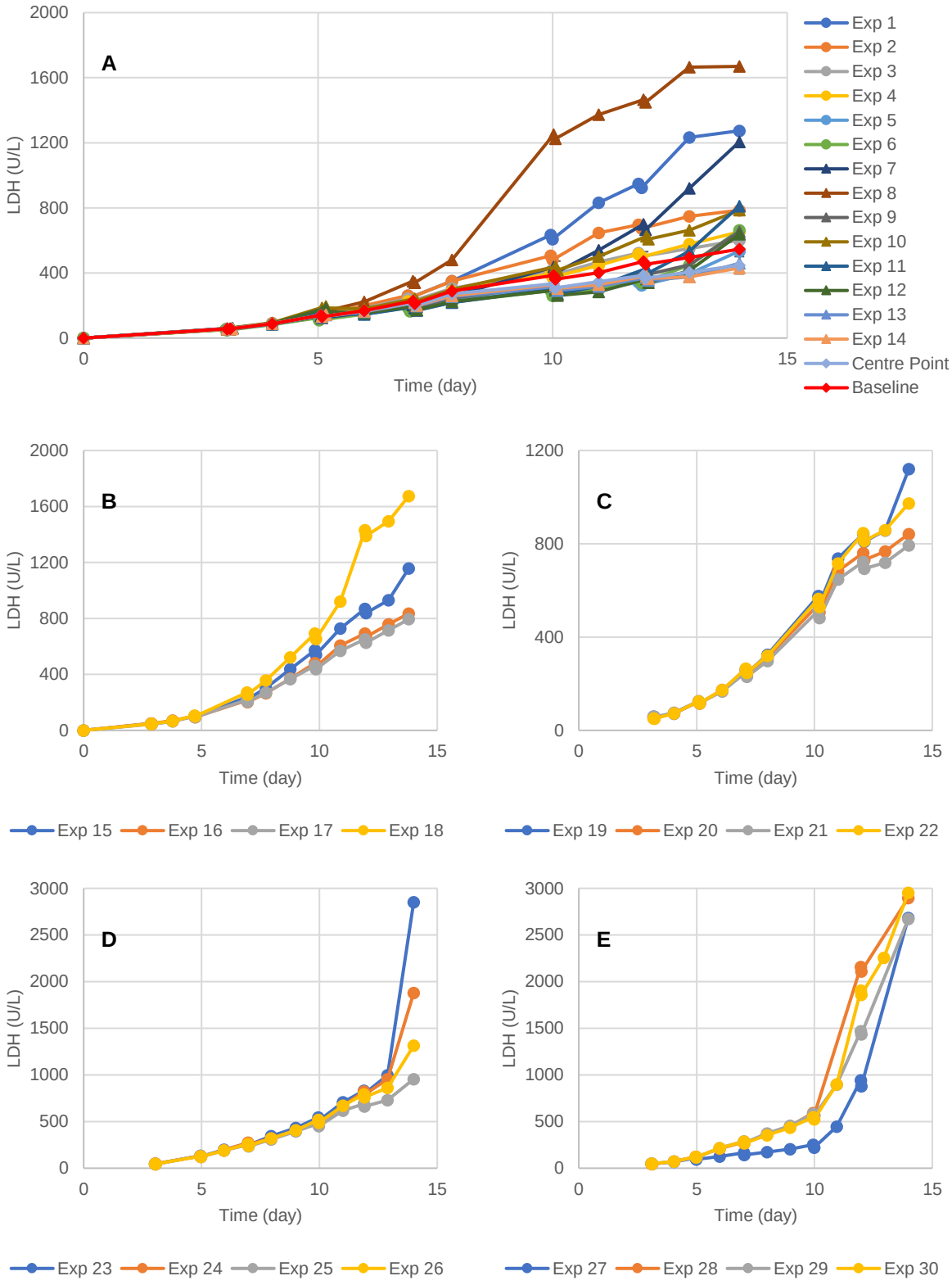

Figure A.0.1 – Time evolution of LDH concentration for the spin-tube experiments: initialisation phase (A), 1st batch (B), 2nd batch (C), 3rd batch (D) and 4th batch (E).

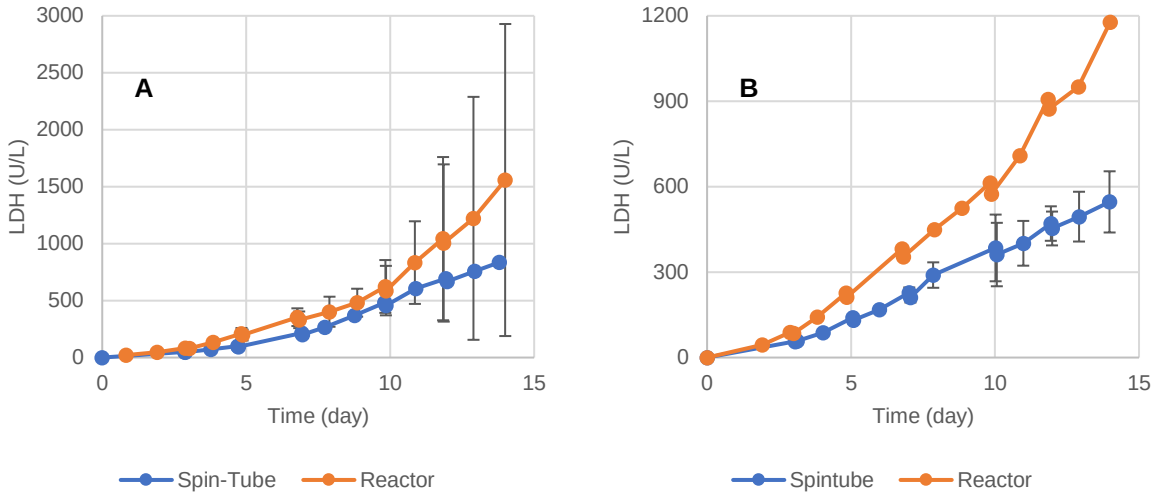

Figure A.2 – Comparison of the time evolution of LDH concentration in the benchtop bioreactor and the corresponding spin-tube: optimised feeding strategy (A), baseline (B).