Newcastle University ePrints

von Stosch M, Davy S, Francois K, Galvanauskas V, Hamelink JM, Luebbert A, Mayer M, Oliveira R, O'Kennedy R, Rice P, Glassey J. [Hybrid modeling for quality by design and PAT - benefits and challenges of applications in biopharmaceutical industry](#). *Biotechnology Journal* 2014, 9(6), 719-726.

Copyright:

This is the accepted version of the following article. von Stosch M, Davy S, Francois K, Galvanauskas V, Hamelink JM, Luebbert A, Mayer M, Oliveira R, O'Kennedy R, Rice P, Glassey J. [Hybrid modeling for quality by design and PAT - benefits and challenges of applications in biopharmaceutical industry](#). *Biotechnology Journal* 2014, 9(6), 719-726, which has been published in final form at <http://dx.doi.org/10.1002/biot.201300385>

Further information on publisher website: <http://onlinelibrary.wiley.com/>

Date deposited: 09-02-2015

Version of file: Accepted Version

This work is licensed under a [Creative Commons Attribution-NonCommercial 3.0 Unported License](#)

ePrints – Newcastle University ePrints

<http://eprint.ncl.ac.uk>

Perspective

Hybrid modelling for quality by design and PAT – benefits and challenges of applications in biopharmaceutical industry

Moritz von Stosch¹, Steven Davy², Kjell Francois³, Vytutas Galvanauskas⁴, Jan-Martijn Hamelink⁵, Andreas Luebbert⁶, Martin Mayer⁷, Rui Oliveira^{1,8}, Ronan O’Kennedy⁹, Paul Rice⁵, Jarka Glassey^{10*}

- 1) REQUIMTE, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal
- 2) NNE Pharmaplan, Gentofte, Denmark
- 3) Siemens NV, Huizingen, Belgium
- 4) Process Control Department, Kaunas University of Technology, Kaunas, Lithuania
- 5) GlaxoSmithKline Vaccines, Laval, Canada
- 6) University Halle-Wittenberg, Germany
- 7) evon GmbH, Gleisdorf, Austria
- 8) IBET/ITQB, Oeiras and Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal
- 9) Fujifilm Diosynth Biotechnologies, United Kingdom
- 10) CEAM, Newcastle University, Newcastle upon Tyne, United Kingdom

*Corresponding author:

Dr Jarka Glassey, CEAM, Merz Court, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK.

jarka.glassey@ncl.ac.uk

Keywords: Hybrid modelling, Biopharmaceuticals, Bioprocess development, Quality by design, Process analytical technology

Abbreviations: AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion; CPP, Critical Process Parameters; CQA, Critical Quality Attributes; DCS, Distributed Control Systems; DoE, design of experiments, EMA, European Medicines Agency; FDA, United States’ Food and Drug Administration; FTIH, First Time in Humans; ICH, International Conference on Harmonization; M³C, measurement, monitoring, modelling and control; MVDA, multivariate data analysis; PAT, process analytical technology; MSE, Mean Squared Error; POC, Proof of Concept; R&D, research and development; QbD, quality by design;

Abstract

This report highlights the drivers, challenges and enablers of the hybrid modelling applications in biopharmaceutical industry. It is a summary of an expert panel discussion of European academics and industrialists with relevant scientific and engineering backgrounds. Hybrid modelling is viewed in its broader sense, namely as the integration of different knowledge sources in form of parametric and nonparametric models into a hybrid semi-parametric model, for instance the integration of fundamental and data-driven models. A brief description of the current state-of-the-art and industrial uptake of the methodology is provided. The report concludes with a number of recommendations to facilitate further developments and a wider industrial application of this modelling approach. These recommendations are limited to further exploiting the benefits of this methodology within PAT applications in biopharmaceutical industry.

1. Background and purpose of the panel report

Traditionally, bioprocess models were developed by translating the fundamental knowledge about the process into mathematical equations, subsequently testing and improving the model hypotheses in comparison to experiments. A model formulated this way can be classified as a parametric model and its development is typically laborious. In recent years, data-driven approaches have increasingly been applied for bioprocess modelling, as those approaches require limited knowledge about the process and their development is systematic and easy. Those data-driven approaches can be classified as nonparametric models. This does not imply that they do not have parameters, only that the structure of those models is determined from data. The combination of parametric and nonparametric models results in hybrid semi-parametric models [1, 2], although typically referred to as hybrid models. Hybrid modelling in a broader sense can be viewed as a way of integrating and structuring process knowledge and information gained from various sources. A typical example is the modelling of kinetic rate expressions using data-driven techniques (frequently using neural networks), which are embedded in a dynamic material balance parametric framework [3, 4]. Since hybrid models can exploit a broader knowledge base, they offer a cost-effective methodology for the solution of complex problems.

This expert panel report focuses on highlighting the progress in applying hybrid modelling in biopharmaceutical industries, briefly compared to other industrial sectors, and the challenges to be addressed for effective application within the Quality by Design (QbD) and PAT frameworks outlined in guidance documents [5-7]. The advances in the application of hybrid modelling in other subject areas and industrial sectors are not addressed in detail in this report.

Applications of hybrid models to bioprocesses focused so far on upstream processes, namely on monitoring [1, 3, 4, 8-11], control [1, 4, 11-12] and optimisation [1, 3, 4, 11, 13-15] of fermentation processes. In comparison to strictly nonparametric approaches, hybrid models usually provided a significantly increased performance, subsequently improving process operation [1]. Despite the advantages, hybrid models have received relatively little attention in biopharma as compared to classical pharmaceutical processes [16] but especially as compared to the chemical [17] and petrochemical sector [18]. In these sectors more advanced hybrid modelling approaches have been developed and their application to control [19], optimisation [20-21] and also scale-up [22] penetrated to the industrial sector. However, the introduction of the PAT initiative in 2004 provides the opportunity for advanced modelling and process control to contribute to improved bioprocess performance [5]. With squeezing profit margins and the patent life of a new chemical entity or formulation not producing as much return on investment as previously enjoyed [23], the impact of hybrid models would be more profound.

2. Business drivers and regulatory pressures

Process understanding, continuous quality improvement and risk analysis are the cornerstones of a robust scientific approach for modern drug development and manufacturing, and of the requirements of the ICH and FDA guidelines [5, 7, 24-25]. These regulatory and quality concerns combined with the commercial imperative are significant drivers for developments in the biopharmaceutical industry.

The basis of QbD is to understand the sources of variability in the process and product and to understand the linkages so that variability can be controlled [24-25]. In particular, the

identification and assessment of Critical Quality Attributes (CQAs) and Critical Process Parameters (CPPs) requires an appropriate scientific model of the process. Such a model maintains required process knowledge through the lifecycle of the drug [24].

The diverse nature of biopharmaceutical products translates into diverse business drivers, but common themes exist across the sector. New candidates can be identified by discovery R&D, in-licensed or acquired from a third party. Also, new therapeutic indications for an existing candidate can be identified. During development, the emphasis is on achieving fast to First time in human (FTIH) and Proof of Concept (POC) clinical studies, to be 'First-in-Class', i.e. it is common for several companies to be simultaneously pursuing promising new targets [26], therefore being first in a particular new class of drugs can distinguish between commercial success and failure. Short timelines leave little room for establishing detailed models of the process. Complications are added by scale-up and process transfer between facilities. Mitigating the risk of clinical failure often means realizing *FTIH* clinical studies under timeline and resource constraints.

Later entrants to a drug class market ('biosimilars') may aim to be 'Best-in-Class' ('biosuperiors'), providing further competition [26]. Biosimilars as a whole are expected to retain a considerable price, estimated at 65% to 85% of the original [27]. The biosimilars market is estimated to reach \$7-8 billion by 2020, exciting industry interest and concern [28]. A number of key industry players have, or have announced their intention, to enter biosimilar manufacturing (e.g. <http://www.gabionline.net/Biosimilars/News/Samsung-makes-biosimilars-deal-with-Biogen-Idex>) [29].

Biopharmaceuticals are entering a competitive era, in which a lack of process optimisation means that the expiration of a patent will lead to diminishing profits. The opportunity is for biosimilar manufacturers with process modelling and process understanding to produce a variety of biosimilars, and be less vulnerable to fluctuations in their pipeline.

These drivers demand modelling methods that enable effective process optimisation, monitoring and control, and minimise sub-optimal product selection or process implementation. Ultimately, this is to the benefit of patient safety.

3. Current applications and future potential

3.1 Process development

Early biopharmaceutical process development often starts with minimal or no experimental data, but rarely with a complete absence of previous knowledge. Traditional approaches to process development have often relied on combining this knowledge with established models that can be limited in scope.

In comparison, statistical methods of experimental analysis, such as those commonly applied in association with Design of Experiments (DoE) principles, e.g. MVDA, disregard any previous knowledge and generate a model based purely on fitting defined relationships to the

experimental data. For establishing statistical significance this demands a non-trivial experimental data set, which may not be available at the beginning of early process development, and the generation of which can represent a considerable resource investment.

Hybrid semi-parametric modelling allows incorporating the existing body of knowledge into a structured framework that adapts the projected relationships according to experimental data as it is collated. As such it promises an economic method for building practical and flexible models.

Whilst literature reports the optimisation of product titres in the upstream process [1], downstream unit operations are often more likely to be on the critical path for process development. Reports of applications of hybrid modelling to address product quality issues are currently limited and its suitability to downstream process unit operations such as chromatography optimisation, product purity, oligomerization or aggregation, has yet to be demonstrated fully [30].

The complexity of integrating multiple operations in process development is another significant challenge that may be suited to the flexible framework provided by hybrid semi-parametric models.

High throughput systems are being implemented across process development disciplines to better address the challenges of rapid product/ process optimisation. These systems are being increasingly integrated with statistical DoE and MVDA packages to manage the large data-sets generated. They also constitute suitable platforms for the incorporation of hybrid modelling software tools exploiting the available process knowledge (hybrid model) to iteratively design a series of experiments aiming either at the exploration of the space or the exploration of the optimal process conditions [1, 14-15]. These model-based experimental design approaches generally allow either to reach the optimal process conditions faster than with standard DoEs or to improve a particular model for a specific operation space.

3.2 Manufacturing operations

The objective during manufacturing operations is to delimit process variations as much as possible. In the QbD and PAT context, the intention is to reduce variability by identifying and, where possible, eliminating the sources of variability and ultimately to counteract the remaining variations by using advanced process control. Thus models that enable analysing the process, tracing back variability and applying advanced process control are required. The development of such models is challenging since the biopharmaceutical manufacturing processes are typically highly complex and not well understood. Different approaches can be applied to model the process, namely fundamental models, data-driven models and hybrid models with advantages and limitations of each approach shown in Table 1. The application of hybrid models is particularly efficient when the model can link different sources of knowledge, since the requirements on the data (both quantity and quality) can decrease and the models inter- and extrapolation properties may improve (Figure 1 and Table 2). This is especially important for process optimisation since

the model should accurately predict offline how the process will perform for conditions that have not yet been tested experimentally, which requires a rather global process understanding, see Figure 1c. Global process understanding, which ranges beyond the process space characterized by design of experiments, requires the model to have good extrapolation properties, a property that is typically associated with more fundamental knowledge (Figure 1b). Hybrid models have proven effective for the process optimisation of a range of manufacturing scales and for the manufacturing of various products [1, 3, 4, 11, 13-15], e.g. Preusting et al. [31] developed a hybrid model to optimize the operational procedure of a large-scale penicillin production process.

For online applications, such as monitoring or control, slightly different model properties are required. For monitoring and supervision the model should typically represent the fermentation state/cultivation state based on the online measurements. During production, the process will normally run within the operation space, which is well characterized by DoE and the local process understanding is sufficient (Figure 1c). In this space data-driven approaches can perform well (Figure 1b). However, while it is feasible to correlate e.g. the biomass concentration to pH, dissolved oxygen concentration or/and to dielectric probe signal, the correlation to product concentrations or even product quality is not straightforward, one reason being the relatively low product titres. Integrating fundamental knowledge with data-driven techniques may reduce these limitations. Several successful applications of hybrid models for process monitoring have been reported [1, 3, 4, 8-11]. Process control requires robust models that can describe the process dynamics and have appropriate extrapolation properties, particularly for frequency extrapolation. This becomes even more critical when moving towards online predictive or optimal control since the process has to be understood beyond the region in which the process was validated in order to describe its behaviour at the limits (Figure 1c). As hybrid models are typically inherently dynamic and as they can predict the process state accurately even for novel process conditions [1], they offer a very suitable method for accurate process control beyond prior tested conditions [12]. Moreover they are more transparent than e.g. data-driven approaches, allowing scrutinizing and correcting the control action, where necessary [12]. Very few applications for process supervision and control have so far been reported [1, 4, 11], but e.g. Dors et al. [32] showed that hybrid models can supervise a production-scale mammalian cell culture process operated in a repeated fed-batch mode and predict the best times for harvesting and the restart of the next fed-batch phase.

4. Future challenges

4.1 New theories and software

The hybrid semi-parametric modelling framework is at present not fully established and several open questions exist [1]. With respect to PAT-applications in (bio)pharmaceutical industry, major concerns are i) model validation; ii) compliance with FDA software regulations [7, 33] for software use in process monitoring or closed-loop control. With respect to the above outlined benefits, hybrid modelling tools can be introduced step-wise, starting from off-line utilization for process analysis, optimisation or scale-up towards their on-line use for monitoring, supervision or closed-

loop control, thus supporting the knowledge transfer from process development to manufacturing.

From an end-user perspective, hybrid modelling tools must be user-friendly and flexible with respect to the integration of different knowledge sources [34] and also to data acquisition and pre-processing (<http://ht.ly/hpoX5>), see Figure 2. Data treatment can be time-consuming, subject to errors and it can significantly impact on the model performance. The tools should be integrated into existing industrial Distributed Control Systems (DCS) and PAT data management platforms, extensively documented and the interfaces standardized [35]. To our knowledge, there are no hybrid modelling tools available yet that meet all mentioned requirements and could be applied in industry.

4.2 Model development process

The development process of hybrid models to date mostly relies on the modeller's experience and no systematic and integrated model development approach has so far been proposed. Central to the development of a hybrid model is its subsequent use. For example, a model for control will aim to capture the transient behaviour (frequency extra-/interpolation) more so than a model for optimisation. In line with this model objective, the available knowledge is assessed, deciding how and what to integrate (Figure 2). For serial hybrid models (Figure 3) it is important considering that the parametric model (the *a priori* incorporated fundamental knowledge) possesses an inductive bias for the subsequent identification of the nonparametric model [1, 3]. Therefore, an iterative model building approach should be employed, starting with the most generic model (containing limited fundamental knowledge) and subsequently incorporating more fundamental knowledge. For each of the structures the identification of the nonparametric model is necessary, which may be challenging for complex models [1]. Model performance criteria, such as BIC, AIC or MSE [36], allow the performance of different models to be compared against the experimental data. Assessing the nonparametric model outputs over time and over the inputs (or at least a subset of inputs) and analysing the model validity space can provide insights into further model improvements. The validity space is assumed to be mostly delimited by the input values of the nonparametric model [1, 37], since the data-driven models do not extrapolate well (Figure 1b). The form of this space can be manipulated through the choice of input variables and their value ranges. The value ranges can be systematically varied through the design of experiments and subsequently this design can be used for the segregation of the experimental data into training, validation and test sets.

Once a model is accepted (Figure 2), the modelling report should detail the model development history and testing. While a model can subsequently be improved by augmenting the existing data set with new data, strict version control should be applied when making any changes. Regular evaluation of model performance should be carried out and documented.

A model impact risk assessment should be carried out prior to any equipment changes such as sensor recalibration, equipment servicing or modification. Changes to process feed stocks and

raw materials can also have an impact on model performance and a model may require controlled updates.

5 Recommendations

The application of hybrid models in (bio)pharmaceutical industry can be significantly enhanced by the following initiatives:

1. *Increased number of industrial case studies describing the optimisation of business relevant process variables.* For instance, the incorporation of knowledge about glycosylation mechanism into the hybrid process model could potentially enable the optimisation of the process with respect to desired glycosylation patterns, and as such represent an important business case. Similarly, the optimization of the complete production strategy with respect to the process bottleneck while ensuring product quality represents an important business driver. The more accurately the model can describe the bottleneck unit, the greater the potential impact when optimising the overall process. As outlined in several places in this article, hybrid models have been shown to offer improved prediction performance compared to other modelling methods.
2. *Case studies reporting the use of hybrid modelling approaches in on-line process control of product related critical process parameters concentrating on processes relevant to the industry.* These would ideally include industry relevant studies that report on (predictive) on-line control of product quality and the influence on critical process parameters upon CQAs.
3. *Knowledge transfer from research to industry, communicating it in different forms to the various hierarchical levels, in order to sensitise and increase awareness for hybrid modelling approaches and their benefits.* This comprises the sketching of scenarios under which hybrid modelling approaches could produce benefits for more management oriented personal and education for staff that will have to understand how to implement the approaches to generate the benefits.
4. *User-friendly software hybrid model solutions, fully integratable within existing systems and easily validatable.*
5. *Introduction of a systematic and integrated approach for hybrid model development.*
6. *Development of a procedure that can determine the limitations of hybrid models with respect to changes in certain variables, thus delimiting the process operation and design region for which the model can faithfully describe the process.* This procedure will help validating the process when its performance critically depends on monitoring and control action that in turn was calculated using hybrid models.

Acknowledgements

The authors would like to thank the Faculty of Engineering, University of Porto and the Faculty of Science and Technology, University Nova de Lisboa for their support. The author Moritz von Stosch gratefully acknowledges financial support through the Fundação para a Ciência e a Tecnologia (Ref: SFRH / BPD / 84573 / 2012).

Conflict of interest statement

The authors declare no commercial or financial conflict of interest.

References

- [1] von Stosch M., Oliveira R., Peres J. and Feyo de Azevedo S., Hybrid Modelling in Process Systems Engineering: Past, Present and Future. *Comput. Chem. Eng.*, Engineering, 2014, 60, 86-101.
- [2] Thompson M. L., Kramer M. A., Modeling chemical processes using prior knowledge and neural networks, *AIChE J.*, 1994, 40, 1328–1340.
- [3] Psychogios D. C., Ungar L. H., A hybrid neural network-first principles approach to process modeling, *AIChE J.*, 1992, 38, 1499–1511.
- [4] Schubert J., Simutis R., Dors M., Havlik I., et al, Bioprocess optimization and control: Application of hybrid modelling, *J. Biotechnol.*, 1994, 35, 51–68.
- [5] United States Federal Food and Drug Administration (USA). Guidance for industry, process analytical technology, FDA, 2004.
- [6] European Medicines Agency, EMA-FDA pilot program for parallel assessment of Quality by Design applications. Document EMA/172347/2011, 2011.
- [7] International conference on Harmonization (ICH) Document ICH Q8–Q10, 2005–2010. Available at www.ich.org.
- [8] Gnoth S., Jenzsch M., Simutis R., Luebbert A., Product formation kinetics in genetically modified e. coli bacteria: inclusion body formation, *Bioproc. Biosyst. Eng.*, 2008, 31, 41–46–46.
- [9] von Stosch M., Oliveira R., Peres J., Feyo de Azevedo S., A hybrid modeling framework for pat: Application to *Bordetella pertussis* cultures, *J. Biotechnol. Progress*, 2011, 28, 284–291.
- [10] Boareto A. J. M., De Souza M. B., Valero F., Valdman B., A hybrid neural model (HNM) for the on-line monitoring of lipase production by *Candida rugosa*, *J. Chem. Technol. Biotechnol.*, 2007, 82, 319–327.
- [11] Galvanauskas V., Simutis R., Luebbert A., Hybrid process models for process optimisation, monitoring and control, *Bioproc. Biosyst. Eng.*, 2004, 26, 393-400.
- [12] von Stosch M., Oliveira R., Peres J., Feyo de Azevedo S., A general hybrid semi-parametric process control framework, *J. Process Contr.*, 2012, 22, 1171-1181
- [13] Laursen S. O., Webb D., Ramirez W. F., Dynamic hybrid neural net- work model of an industrial fed-batch fermentation process to produce foreign protein, *Comput. Chem. Eng.*, 2007, 31, 163–170.
- [14] Teixeira A., Cunha A. E., Clemente J. J., Moreira J. L., et al, Modelling and optimization of a recombinant bhk-21 cultivation process using hybrid grey-box systems, *J. Biotechnol.*, 2005, 118, 290–303.
- [15] Ferreira A.R., Dias J.M.L., von Stosch M., Clemente J., et al., Fast development of *Pichia pastoris* GS115 Mut+ cultures employing batch-to-batch control and hybrid semi-parametric modeling; *J. Bioproc. Biosys. Eng.*, 2013 DOI:10.1007/s00449-013-1029-9.

- [16] Kumar Akkisetty P., Lee U., Reklaitis G., Venkatasubramanian V., Population balance model-based hybrid neural network for a pharmaceutical milling process, *J. Pharm. Innov.*, 2010, 5, 161–168.
- [17] Mogk G., Mrziglod T., Schuppert A., Application of hybrid models in chemical industry, in: J. Grievink, J. van Schijndel (Eds.), *European Symposium on Computer Aided Process Engineering-12*, volume 10, Elsevier, 2002, 931–936.
- [18] Mahalec V., Sanchez Y., Inferential monitoring and optimization of crude separation units via hybrid models, *Comput. Chem. Eng.*, 2012, 45, 15–26.
- [19] Wei N. C., Hussain M. A., Wahab A. K. A., Control of a batch polymerization system using hybrid neural network - first principle model, *Can. J. Chem. Eng.*, 2007, 85, 936–945.
- [20] Ibrehem A. S., Hussain M. A., Ghasem N. M., Hybrid mathematical model and advanced control of a fluidized bed using a model-predictive controller, *Journal of Petroleum and Gas Engineering*, 2011, 2, 25-44.
- [21] Zhang S., Wang F., He D., Jia R., Batch-to-batch control of particle size distribution in cobalt oxalate synthesis process based on hybrid model, *Powder Technology*, 2012, 224, 253–259.
- [22] Bollas G. M., Papadokonstadakis S., Michalopoulos J., Arampatzis G., et al, Using hybrid neural networks in scaling up an FCC model from a pilot plant to an industrial unit, *Chem. Eng. Process.*, 2003, 42, 697–713.
- [23] Chew, W., Sharratt P., Trends in process analytical technology, *Analytical Methods*, 2010, 2, 1412-1438.
- [24] United States Federal Food and Drug Administration (USA). Process Validation: General Principles and Practises, FDA, 2011.
- [25] United States Federal Food and Drug Administration (USA). A Risk Based Approach, FDA, 2002.
- [26] Schulze U., Ringel M., What matters most in commercial success: first-in-class or best-in-class, *Nature Reviews Drug Discovery*, 2013, 12, 419-420.
- [27] Calo-Fernández B, Martínez-Hurtado J. "Biosimilars: Company Strategies to Capture Value from the Biologics Market". *Pharmaceuticals*, 2012, (12), 5 , 1393–1408.
- [28] Gal R., Biosimilars: Reviewing US Law and US/EU Patents; Bottom Up Model Suggests 12 Products and \$7-\$8B Market by 2020. Bernstein Research. 26 May 2011.
- [29] What You Need to Know About Biosimilar Medicinal Products. Consensus Information Paper, European Commission, 2013.
- [30] Gao Y., Kipling K., Glassey J., Willis M., et al., Application of agent-based system for bioprocess description and process improvement, *Proc Biochem.*, 2010, 26, 706-716
- [31] Preusting H., Noordover J., Simutis R., Luebbert A., The use of hybrid modeling for optimization of the penicillin fermentation process, *Chimia*, 1996, 50, 416-417.

- [32] Dors M., Simutis R., Luebbert A., Hybrid process modeling for advanced process state estimation, prediction and control exemplified in a production-scale mammalian cell culture, *ACS Symposium Series*, 1995, 613, 144-154.
- [33] Mandenius C.-F., Graumann K., Schultz T. W., Premsteller A., et al., Quality-by-Design (QbD) for biotechnology-related pharmaceuticals. *Biotechnol. J.* 2009, 4, 600–609.
- [34] Gernaey K.V., Gani R., A model-based systems approach to pharmaceutical product-process design and analysis. *Chemical Engineering Science*, 2010, 65, 5757-5769.
- [35] Glassey J., Gernaey K.V., Oliveria R., Striedner G. et al., PAT for biopharmaceuticals. *Biotechnol. J.* 2011, 6, 369–377.
- [36] Peres J., Oliveira R., de Azevedo S. F. Bioprocess hybrid parametric/nonparametric modelling based on the concept of mixture of experts, *Biochemical Engineering Journal*, 2008, 39, 190–206.
- [37] Teixeira A. P., Clemente J. J., Cunha A. E., Carrondo M. J. T., et al., Bioprocess iterative batch-to-batch optimization based on hybrid parametric/nonparametric models, *Biotechnology Progress*, 2006, 22, 247–258.
- [38] Eriksson L., Johansson E., Kettaneh-Wold N., Trygg J., et al., Model derivation, Interpretation and Validation, in Multi- and Megavariate Data Analysis Part II Advanced Applications and Method Extensions, 2006, Umetrics AB Sweden.
- [39] Patnaik P.R., Hybrid neural simulation of a fed-batch bioreactor for a non-ideal recombinant fermentation, *Bioproc. Biosyst. Eng.*, 2000, 24, 151–161.

Table 1. Conventional process model features vs. hybrid semi-parametric models.

Parametric / Fundamental	Nonparametric/ Data-driven
<ul style="list-style-type: none"> • Based on first-principles, mechanistic or empirical knowledge about the system • Parameters have a physical meaning • Model structure is fixed through system knowledge • Model development is laborious • Good extrapolation properties • Examples: <ul style="list-style-type: none"> - Material / Energy / Momentum laws - Thermodynamic laws - Reaction kinetics 	<ul style="list-style-type: none"> • Based on process data provided by the system • Parameters have no physical meaning • Model structure is identified from data • Models are quickly applicable • Descriptive quality within experimentally explored region only • Examples: <ul style="list-style-type: none"> - Artificial Neural Networks - Partial Least Square Regression - Splines, kernels, wavelets
<p style="text-align: center;">Hybrid semi-parametric SYNERGY BENEFITS</p> <ul style="list-style-type: none"> • Exploitation of advantages/ Reduction of disadvantages of particular modelling techniques • Improved modelling / prediction quality • More efficient model development • Better benefit / cost ratio when solving complex problems 	

Table 2: Application-specific requirements on modelling methods and respective (dis-)advantages of hybrid models

Application	Requirements	Advantages/ Disadvantage of hybrid models
Process Development	<ul style="list-style-type: none"> • Support maximising product yield • Integration of various knowledge sources should be feasible • Impact of several factors must be studied • Extrapolation properties 	<ul style="list-style-type: none"> • Potentially fewer experiments required than for the development of pure data-based models • More user-intensive than data driven approaches • Allows to study impact of certain variables without the execution of experiments, e.g. for the initial biomass concentration • May provide good extra- and interpolation properties • May require DoE for identification of the nonparametric model
Process Scale-up	<ul style="list-style-type: none"> • Support scale-up of the process to manufacturing scale • Extrapolation properties • Account for scale dependent phenomena, e.g.: the influence of transport phenomena on gross kinetic expressions at production scale 	<ul style="list-style-type: none"> • Experimental studies (partial DoEs) required to study the scale specific phenomena • Influence of transport phenomena on gross kinetic expressions at production scale can be accounted for, see e.g. [39] • May provide good extra- and interpolation properties
Process Optimisation	<ul style="list-style-type: none"> • Allow to increase process robustness • Extrapolation and Interpolation properties • Impact of several factors must be incorporated • Several sources of knowledge should be integrated 	<ul style="list-style-type: none"> • Potentially fewer experiments required • Allows to study impact of variables without executing complete DoE • Easier to incorporate the influence of factors • Convergence to the optimum is not ensured
Process Validation	<ul style="list-style-type: none"> • Demonstrate that the process is understood within the process operation space • Impact of factors on process and product quality (also changes with time) must be understood • Interpolation properties 	<ul style="list-style-type: none"> • Studies on certain factors can be more easily integrated than for fundamental models • Easier to capture the dynamic effects with hybrid modelling approaches than with data driven ones

Process Monitoring	<ul style="list-style-type: none"> • Enable reliable tracking of the process performance • Support decision making • Interpolation (and extrapolation) 	<ul style="list-style-type: none"> • Development of models takes slightly longer with increasing number of inputs • Model predictions may be bound [1], therefore e.g. no negative substrate concentrations predicted • More data can be used for calibration than for data-driven models [9] • Process prediction and online optimisation or decision support feasible with dynamic models
Process Control	<ul style="list-style-type: none"> • Ensure that the process stays within the approved operation space • Multivariate control to act on the remaining process variability • Cell dependent time restrictions for control action • Control action should be analysable 	<ul style="list-style-type: none"> • Easy to capture multivariate effects [12] • Multivariate control becomes possible (e.g. MPC, optimal control) • Better analysability of control action than for data-driven models [12] • May be difficult to assess stability, robustness

Figure Legends

Figure 1: Schematic diagrams for A) The comparison of modelling methods in relation to the available fundamental knowledge/ requirements on data; B) The comparison of modelling methods in relation to the process understanding and the expenses for model development; and C) The requirements of process operation and design approaches on process understanding for a certain level of performance.

Figure 2: Process flows for data acquisition prior to modelling and for model lifecycle. Additional steps in the modelling analysis cycle must be considered for creating hybrid models in comparison to creating multivariate models [38].

Figure 3. Schematic representation of hybrid model structures. A) Serial Structure; B) Serial structure with feed-back of state-variables; C) Parallel structure.

Figure 1

Figure 2

Figure 3