Parameter estimation and method of moments for multi dimensional population balance equations with application to vaccine production processes

Dissertation

zur Erlangung des akademischen Grades

**Doktoringenieur
(Dr.-Ing.)**

von Dipl.-Ing. Robert Dürr
geboren am 23.04.1985 in Torgau

genehmigt durch die Fakultät für Elektrotechnik und Informationstechnik
der Otto-von-Guericke-Universität.

Gutachter:

Prof. Dr.-Ing. habil. Achim Kienle
Prof. Dr.-Ing. habil. Dominique Thévenin

Promotionskolloquium am 23.11.2016

Kurzreferat

Partikelsysteme sind in einem breiten Spektrum verfahrenstechnischer Anwendungen von großer Bedeutung. Beispiele finden sich sowohl in der chemischen Verfahrenstechnik, z.B. Kristallisationsprozesse, als auch in der Bioverfahrenstechnik, z.B. bei der zellkulturbasierten Herstellung von Biopolymeren oder Impfstoffen. Dabei variieren die einzelnen Partikel, z.B. Kristalle oder Zellen, hinsichtlich bestimmter Eigenschaften wie der charakteristischen Größe, der Form und ihrer Zusammensetzung. Die daraus resultierende Eigenschaftsverteilung der Partikel hat einen wesentlichen Einfluss auf wichtige Produkteigenschaften. Neben experimentellen Untersuchungen ist eine mathematische Modellierung für eine vollständige Charakterisierung dieser Prozesse unabdingbar. Darüber hinaus stellt die Modellbildung einen wichtigen Schritt im Hinblick auf die Entwicklung geeigneter Strategien zur Optimierung und Führung der Prozesse dar.

Die Dynamik von Partikelsystemen lässt sich im Rahmen der populationsdynamischen Modellierung mithilfe partieller Differentialgleichungen beschreiben. Dabei entspricht die Ordnung der partiellen Differentialgleichungen der Anzahl der modellierten Partikeleigenschaften. Im allgemeinen können zwei Modellierungsstrategien unterschieden werden: Bei der *top-down*-Modellierung werden Messungen der Partikeleigenschaftsverteilungen direkt modelliert, was zu niedrig dimensionalen Populationsbilanzgleichungen führt. Die zeitliche Veränderung der Partikeleigenschaften wird dabei durch globale mechanistische Ansätze beschrieben. Diese müssen durch die Lösung eines inversen Problems an die vorhandenen Messdaten angepasst werden. Im Gegensatz dazu basiert die *bottom-up*-Modellierung auf einer detailierten Beschreibung der Mechanismen auf Einzelpartikelebene. Vor allem für biologische Prozesse beinhalten diese Einzelpartikelbeschreibungen oft eine große Anzahl von Partikelzuständen. Innerhalb einer populationsdynamischen Beschreibung übersetzt sich jede Eigenschaft direkt in eine unabhängige Koordinate der jeweiligen partiellen Differentialgleichung. Somit führt der *bottom-up*-Ansatz meist zu hochdimensionalen Problemstellungen. Diese erfordern den Einsatz effizienter numerischer Lösungsmethoden, da Standardansätze, wie zum Beispiel die Finite Volumen Methode, nicht ohne weiteres effektiv implementiert werden können.

Im Rahmen dieser Arbeit werden beide Modellierungsansätze auf einen zellkulturbasierten Produktionsprozess für Influenza A Impfstoffe angewendet. Dieser Prozess stellt aufgrund auftretender Varianzen innerhalb der Zellpopulation ein interessantes Anwendungsbeispiel für partikuläre Systeme dar.

Der erste Teil dieser Arbeit befasst sich mit der Parameterschätzung für ein in einer vorherigen Arbeit entwickeltes populationsdynamisches *top-down* Modell [81, 79].

Dieses beschreibt die Interaktion von Zellpopulation und Virus innerhalb eines Bioreaktors und erfasst zudem die Varianz der Zellen hinsichtlich der intrazellulären Menge des viralen Nucleoproteins (NP). Eine Anpassung des Modells an durchflusszytometrische Messungen erfordert die Bestimmung konstanter und funktioneller Modellparameter. Das daraus resultierende unendlich-dimensionale inverse Problem wird zunächst durch eine Hermite-Spline-Approximation der funktionellen Parameter in eine endliche Dimension überführt, um anschließend die Modellparameter mit der Methode der kleinsten Quadrate zu bestimmen. Das angepasste Modell ist in der Lage, die beobachteten dynamischen Effekte qualitativ nachzubilden. Des weiteren wird gezeigt, dass die Modellgüte durch die Verwendung komplexerer Approximationen der funktionellen Parameter verbessert werden kann. Dies geht jedoch mit einer Erhöhung des numerischen Aufwandes einher und eine biologisch sinnvolle Interpretation der geschätzten funktionellen Parameter wird erschwert.

Um diese Nachteile zu umgehen, bietet es sich an, mehr strukurelle Information über den Prozess in die Modellierung einzubeziehen. Eine entsprechende *bottom-up*-Modellierung basierend auf einer detaillierten Einzelzellbeschreibung der Influenzavirusreplikation [41] führt auf eine hochdimensionale Populationsbilanz. Hierfür wird im zweiten Teil dieser Arbeit eine effiziente approximative Momentenmethode zur numerischen Lösung vorgestellt. Diese ist eine Kombination aus der bereits existierenden direkten Quadraturmethode [70] und einer effizienten Wahl der Stützstellen auf Basis monomialer Kubaturen [107]. Anhand verschiedener generischer Beispiele wird gezeigt, dass die vorgestellte Methode einen guten Kompromiss zwischen numerischem Aufwand und Genauigkeit der Lösung darstellt.

Im letzten Teil dieser Arbeit wird die vorgestellte Methode auf das hochdimensionale populationsdynamische Modell der Influenzaimpfstoffproduktion angewendet. Dabei wird zunächst der Einfluss von steigender Varianz innerhalb der Zellpopulation auf die Virusausbeute untersucht. Hier zeigt sich, dass mit steigender Variabilität die Virusausbeute sinkt. Darüber hinaus wird die Verwendung genetisch modifizierter Zelllinien zur Verbesserung der Impfstoffproduktion diskutiert. In der Praxis werden diese Modifikationen beispielsweise mithilfe von Transduktionsverfahren vorgenommen. Jedoch werden nicht alle Zellen mit der gleichen Effizienz modifiziert, so dass eine Varianz der zellspezifischen Virusproduktionsraten zu erwarten ist. Eine modellbasierte Untersuchung zeigt, dass trotz dieser Heterogenitäten eine signifikante Steigerung der Virusausbeute erreicht werden kann und die Verwendung genetisch modifizierter Zelllinien eine gewichtige Rolle bei der Verbesserung des Impfstoffproduktionsprozesses spielen können.

Abstract

Particulate systems are found in many process engineering applications and play an important role in a large number of industries such as pharmaceutical and food industry. In general, the particles are not uniform and properties, like size, shape or internal composition appear distributed within the particle population. These property distributions may have significant effects on the overall product quality or processability. Examples are not only found in chemical process engineering, e.g. crystallization, but also in bioprocess engineering, e.g. cell culture based production of biopolymers or vaccines. Here, cell-to-cell variability with respect to size and intracellular composition is a common feature affecting the overall processes. A model based analysis of the observed effects supports an thorough experimental characterization of the process and is an important step in the design of model based process intensification and control strategies.

The dynamics of particulate processes can be characterized conveniently with the population balance modeling approach. Here, the resulting population balance equations represent partial differential equations and describe the temporal evolution of the properties distribution of the particle ensemble. The dimension of these equations increase with the number of modeled properties. To formulate a suitable population balance model, two general strategies can be followed. Within the top-down approach the measured distributed data is modeled directly. In result, low dimensional population balance equations characterizing the particles heterogeneity with respect to a moderate number of measured properties are obtained. To account for the temporal change of the particle properties, global mechanistic kinetics are used. These have to be adapted to the distributed measurements by solution of an inverse problem. In contrast, bottom-up modeling starts from a detailed description of the mechanisms on the single particle level. In particular for bioprocesses, these may include a large number of particle properties. As each modeled property directly translates to an independent coordinate of the corresponding population balance model formulation, high dimensional partial differential equations are obtained. These require the application of efficient numerical solution techniques as standard discretization based solution methods lack effective implementation for high dimensional problems.

Within this thesis, focus is on application of both modeling strategies to cell-culture based influenza A vaccine production processes. These represent an interesting example for particulate processes as cellular variability affects the overall virus yield.

The first part is concerned with parameter estimation for a top-down model of the process which was developed in a previous work [81, 79]. The model describes the interaction of cell population and virus within a bioreactor and is able to account for

cell-to-cell variability with respect to intracellular amount of viral nucleo protein (NP). As model adaption to flow cytometric data requires the estimation of both, constant and distributed parameters, an infinite dimensional inverse problem emerges. It is translated to a finite dimension by a suitable approximation of the distributed parameters. The parameters are estimated within a generalized least squares framework and the resulting model is able to qualitatively reproduce most of the dynamic effects, though not every trait of data is covered. It is also shown that the model fit can be improved when using more complex approximations of the distributed parameters. However, this comes along with an increased numerical effort and a biologically reasonable interpretation of the estimated parameters becomes more and more challenging.

This lack of interpretability motivates the application of a bottom-up modeling approach which allows to take into account more structural information. The model formulation is based on a detailed description of the viral replication mechanism on the single cell level presented in [41]. The resulting high dimensional population balance model requires the application of an efficient numerical solution technique. Therefor, in the the second part, an efficient approximate moment method is developed which based on a combination of the direct quadrature method of moments [70] and a specific choice of the abscissas applying monomial cubatures [107]. The method is evaluated for several benchmark processes which can be represented by low dimensional population balance models. Furthermore, performance is extensively discussed for a generic five dimensional model of viral replication in a multi cellular system. The technique represents a good trade off between computational effort and accuracy of the solution and thus enables an efficient numerical solution of high dimensional population balance models.

Finally, the method is applied to the high dimensional population balance model for influenza A virus replication in cell cultures. It is shown that an increasing cell-to-cell variability negatively affects the overall virus yield. In addition, the application of genetically modified cell lines for improvement of the vaccine production is discussed. In practice these can be achieved using e.g. transduction techniques. However, not all cells are modified with the same efficiency. Therefore, a certain degree of variance w.r.t. the cell specific virus production rate is expected. A model based analysis shows that even for a relatively large heterogeneity a significant increase in the overall virus yield is obtained.

"But still try, for who knows what is possible."
Michael Faraday

Foreword

Firstly, I want to express my gratitude to the head of the chair, Prof. Dr.-Ing. habil. Achim Kienle not only for his professional support and guidance but also for providing excellent working conditions and a comfortable working atmosphere in his group. I would also like to thank Prof. Dr.-Ing. habil. Dominique Thévenin for acting as the second referee for this thesis.

I also owe sincere gratitude to our collaboration partners from the the Bioprocess Engineering group and their head Prof. Dr.-Ing. habil. Udo Reichl for the many interesting and insightful discussions as well as for broadening my very limited knowledge of the underlying biological mechanisms. In particular, I would like to thank Britta Peschel, Tanja Laske, Mandy Bachmann, Stefan Heldt, Felipe Tapia and Timo Frensing.

I want to express my gratitude to my colleagues and friends in our group for their support, many fruitful, funny and interesting discussion on scientific and all-day topics as well as lots of off-work activities: Eugenia Schneider, Stefanie Duvigneau, Anke Göttert, Annett Bartels, Ilknur Disli-Kienle, Michael Mangold, Christian Kunde, René Schenkendorf, Carsten Seidel, Christoph Neugebauer, André Franz and Steffen Sommer. Special thanks go to Thomas Müller with whom I had the privilege of sharing an office for nearly six years, and to Stefan Palis and Andreas Bück, for giving me good advises and providing help and encouragement in countless situations over the last years.

Additionally, many thanks go to my friends for their constant efforts in reminding me that scientific work is not everything in life and for their constant support: Juliane and Philipp, Franziska and Jörg, Yujing and Benjamin, Dasha, Betty and Raik, Stephan, ... Most certainly this is just an excerpt of a much longer list which would fill an additional page of this thesis.

Last but not least, I would like to thank my family, especially my parents Ute and Holger, and my brother Oliver: Auch wenn ihr vieles von dem, was ich hier geschrieben habe wahrscheinlich nicht versteht, solltet ihr wissen, dass ich es ohne eure Unterstützung und Hilfe niemals bis hierher geschafft hätte. Danke für alles.

Contents

1. Introduction

1.1. Motivation

Particulate process systems are found in a broad field of process engineering applications in which single particles or individuals differ from each other with respect to certain characteristic properties. These variations arise in result of interaction with a temporal or spatially heterogeneous environment or stochastic behavior on the single particle scale.

Examples from chemical processes include agglomeration [89], granulation [115, 91], crystallization [8, 58, 63] and coating processes [37]. Here, nonuniform process parameters (e.g. spatial gradients of temperature or solute concentration) and nonuniform process steps, e.g. nucleation, grinding and sieving, can result in significant particle-to-particle variabilities with respect to physical properties like characteristic size and shape. Furthermore, interactions between particles, e.g. fusion of two particles or particle breakage, and effects at the single particle level, e.g. non ideal moisture distribution over a spherical particle, contribute to this heterogeneity. An example is depicted in Fig. 1.1a. It is clearly seen that the shown crystals exhibit a large variance w.r.t. characteristic size and shape.

Further examples for particulate process are found in bioprocess engineering applications where multi cellular systems are involved, e.g. cultivation of yeast [124, 122, 75], biopolymer production in microorganisms [28, 27] and vaccine production processes [103]. As in the aforementioned examples, nonuniformity of cells with respect to physical properties, like size and shape, but also with respect to intracellular composition is observed. Besides nonuniformity in the process conditions, unsynchronized cell cycles, age distributions [78], stochastic effects on the gene expression level and bistable behaviour on the single cell level (e.g. cell proliferation [14, 42] or apoptosis induction [24]) play a major role in the formation of these heterogeneities. A typical example for cell-to-cell variability, which is a common feature of multi cellular systems, can be seen in Fig. 1.1b. Here, bacteria can be distinguished with respect to their characteristic length.

The described variances within the particle ensemble may have significant effects on the processes and the quality of the overall products affecting e.g. storage ability or drug efficiency. A sophisticated mathematical description accompanied by thorough experimental work contributes to a better understanding of the overall processes and provide a basis for the design of suitable process intensification and process control schemes.

1
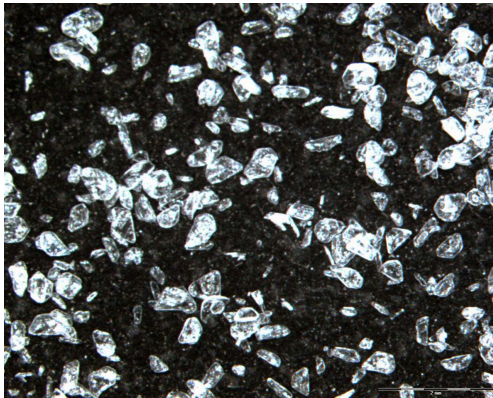
(a) *Microscopic picture of mandelic acid crystals; single crystals differing w.r.t. size and shape*
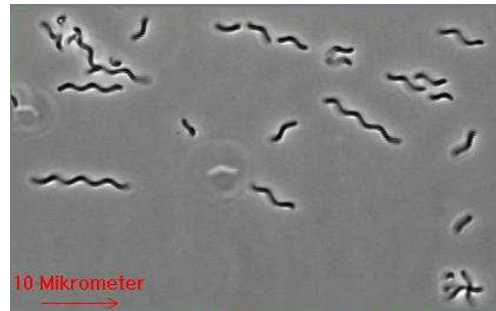
(b) *Microscopic picture of Rhodospirillum rubrum bacteria cells; cells differ in characteristic length*

Figure 1.1.: *Examples for heterogeneity within particle ensembles in chemical- and bioengineering processes*

1.2. Mathematical modeling of particulate systems

Each particle can be characterized by a set of properties of which each can undergo changes resulting from stochastic or deterministic effects during the process. These effects can be modeled on the microscopic scale using discrete event systems, ordinary differential equations and stochastic differential equations, respectively [31, 93]. Techniques like discrete particle methods [15, 17, 3], ensemble simulations [100] and Monte Carlo methods [123, 31] rely on a large number of realizations of the single particle dynamics. Yet, usually a large computational effort is involved as the number of single particle realizations has to be sufficiently high to guarantee reliable statements on the overall particle ensemble.

Alternatively, instead of treating every particle separately, particles with the same properties can be categorized within property classes. For infinitesimal classes a corresponding property distribution is obtained which characterizes the particle population. The dynamics can be described conveniently in the framework of population balance modeling (PBM) [92]. This technique does not only account for deterministic changes in the particle properties but can also capture random effects and discrete dynamics to a certain extent [93]. One of the first accounts of the application of PBM to a particulate process engineering system is found in the contributions of Hulburt and Katz [44] describing problems in statistical mechanics. For bioprocesses, one of the first efforts of using partial differential equations to describe a heterogeneous cell population is given by the von Foerster Equation [84] which was generalized to account for cell-to-cell variability in terms of intracellular composition [29]. From this on, the PBM concept was established further and found many applications in chemical process (see e.g. [10, 12, 61, 85, 97]) and bioprocess engineering (see e.g. [55, 25, 53, 82]).

Main focus in the present thesis is on multi cellular systems. For these, two alterna-
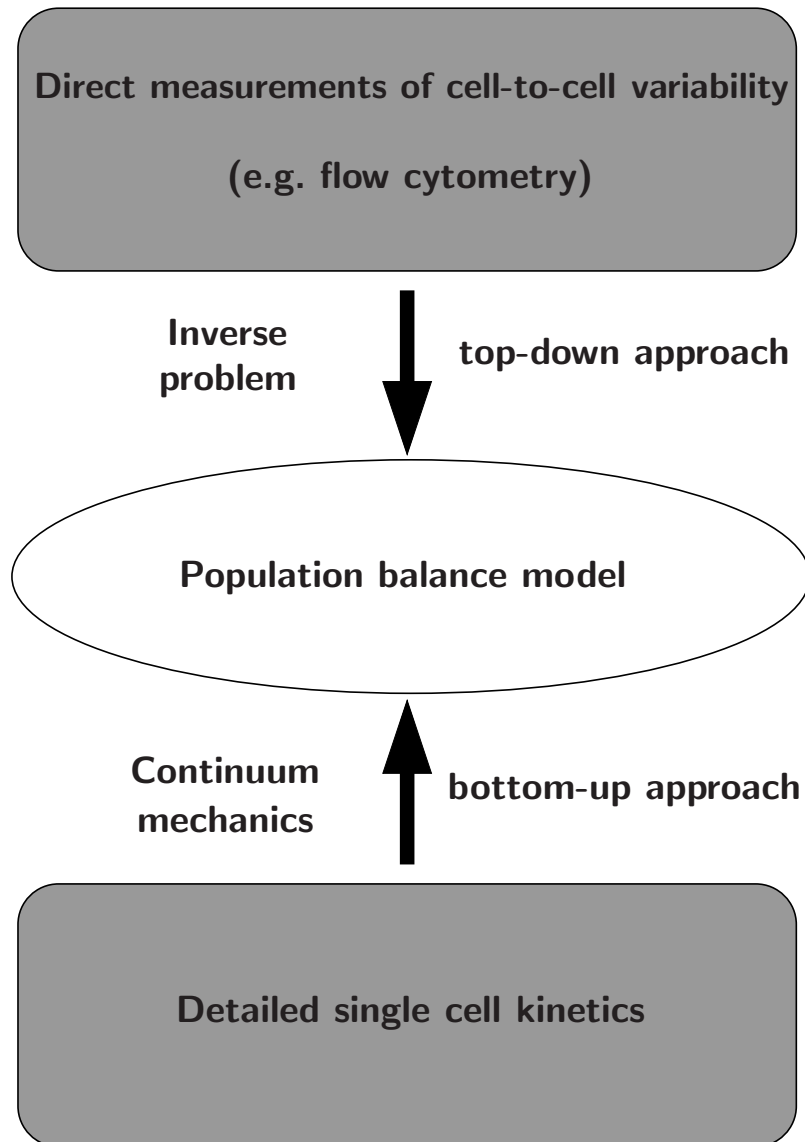
Figure 1.2.: *Schematic representation of two alternative population balance modeling approaches: In the top-down approach measurements of the cell-to-cell variability are directly modeled. In contrast, bottom-up modeling relies on a detailed formulation of the kinetics on the microscopic level*

tive modeling procedures can be followed for the formulation of a suitable population balance equation (PBE). A principle scheme of the modeling approaches is shown in Fig. 1.2. In the following, both approaches and the corresponding challenges will be discussed.

1.2.1. Top-down modeling approach

The *top-down modeling* approach relies on measurements of the cell-to-cell variability. These are available from sophisticated measurement techniques like video imaging or flow cytometry, in which a low number of properties, e.g. size, is quantified for a large number of individuals. These measurements are modeled directly and *unstructured* population balance models are obtained which represent low dimensional integro partial differential equations. Here, global mechanistic kinetics on the cell population level are used to account for the temporal evolution of the cellular properties.

To adapt these mechanistic kinetics to the available measurements an inverse problem has to be solved. As some of the mechanistic model parameters may exhibit functional dependencies on the particle properties, the inverse problem is generally set in an infinite dimension. This requires the application of elaborate solution techniques. In contrast to the estimation of constant model parameters for population balance models, for which a large number on contributions is found (see e.g. [36, 62, 9, 60]), the solution of infinite dimensional inverse problems has not received much attention (see e.g. [59, 5, 64, 65]).

1.2.2. Bottom-up modeling approach

Due to macroscopic mechanistic kinetics, interpretation, predictive capacity and accuracy of top-down models is limited. Alternatively, a *bottom-up* modeling strategy can be pursued. Initial point is a detailed description of the dynamics on the single cell scale represented by a system of ordinary differential equations. The size of this system depends on the number of considered cellular properties. In particular for bioprocesses, detailed information from the microscopic single cell level is available from sophisticated experiments and thus the size of the corresponding system of ordinary differential equations can become very large easily. To account for heterogeneity with respect to the cellular properties, the single particle description is transformed to a *structured* population balance model. As each particle property translates into an additional internal coordinate, high dimensional integro partial differential equations are obtained. Analytic solutions for these are only found for special cases requiring the application of numerical solution algorithms.

In contrast to the low dimensional case, standard discretization based methods like finite difference [66], finite volume [56] and finite element methods [68, 45] can not be implemented efficiently. This motivated the development of sophisticated extensions to ease the computational burden. These include alternative discretization schemes (e.g. [86, 34]), the application of spectral methods (e.g. [67, 112, 11]) and the use of parallel computing algorithms (e.g. [35]). However, application of discretization based algorithms is mostly limited to low dimensional population balance models with a maximum of three coordinates because of the increasing numerical effort. Alternatively, *moment methods* can be employed [44, 73]. Here, instead of the full particle property distribution only specific integral quantities of the distribution, so called moments, are tracked. However, in most cases a closed set of equations describing the moment
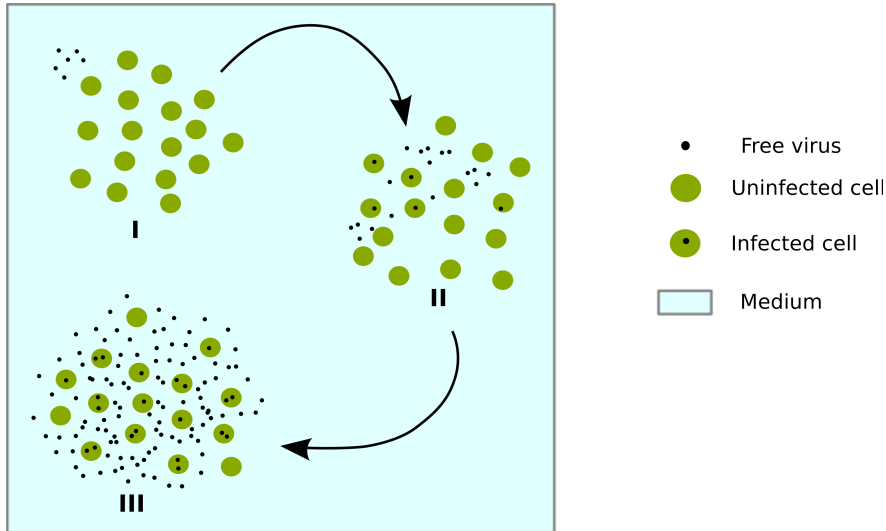
Figure 1.3.: *Basic scheme for viral replication in a multi cellular system: (I) initially, a small concentration of virions and a large concentration of uninfected cells, (II) virus particles infect the cells which release replicated virions to the medium, (III) finally, nearly all cells are infected and virions are highly concentrated in the medium*

dynamics can not be found and *approximate moment methods* have to be applied. Within this class the quadrature method of moments [73] is an important representative which has found many applications to one dimensional population balance models, e.g. crystallization [2] and precipitation [71]. Several extensions for multi dimensional population balance equations have been developed, e.g. [118, 119, 120, 121, 12]. Yet, applications to higher dimensional PBEs with more than three internal properties are rarely found. Another option for the approximation of moments is the direct quadrature method of moments [70] which can be implemented more efficiently for high dimensional problems as also illustrated in this thesis.

1.3. Scope and objectives

Within this thesis major focus is on influenza A virus replication in mammalian cell cultures for vaccine production. A basic scheme is depicted in Fig. 1.3. The virus replicates in the cells, is then released from the cells to the medium and infects still uninfected cells resulting in a spreading infection within the cell culture. The aim is the production of a large concentration of virus in the medium from a small initial concentration. The harvested virus is later used for the production of the vaccine. This process represents an interesting particulate system as the cells are characterized by a significant cell-to-cell variability with respect to the intracellular composition, i.e. the intracellular amount of viral compounds. Furthermore, the cell population interacts with a continuous phase accounting for free virus particles. The aforementioned modeling strategies

can be applied to come up with population balance models for this process and will both be discussed in this thesis.

Motivated by flow cytometric measurements a top-down population balance model was developed in a preceding work (see [81, 79]). Model adaption to available data involves the solution of an infinite dimensional inverse problem. The development of a suitable solution strategy is the aim of the first part of this thesis.

Based on a detailed description of the viral replication cycle on the single cell level a bottom-up population balance model of the vaccine production process can be formulated. For numerical solution, an efficient approximate moment method will be developed and evaluated in the second part of this thesis. Besides a generic model of the viral replication process benchmarks also include crystallization processes which demonstrates applicability to other particle processes.

In the third part, the method will be used for an efficient numerical solution of the multi dimensional population balance equation accounting for cell-to-cell variabilities during vaccine production. Here, specific interest lies on the modeling of genetically engineered cell lines which can be used to overcome bottlenecks of viral replication process. These modifications result in heterogeneity within the cell population affecting the overall production and thus require detailed analysis.

It has to be mentioned that parts of the results presented in this thesis have been published in advance and are listed in chronological order in the following:

- Thomas Müller, <u>Robert Dürr</u>, Britta Isken, Josef Schulze-Horsel, Udo Reichl, Achim Kienle. **Population balance modeling of influenza virus replication during vaccine production - Influence of apoptosis**. *Proceedings to 21st European Symposium on Computer Aided Process Engineering - ESCAPE 21 (2011)*, 1336-1340.

- <u>Robert Dürr</u>, Thomas Müller, Britta Isken, Josef Schulze-Horsel, Udo Reichl, Achim Kienle. **Distributed modeling and parameter estimation of influenza virus replication during vaccine production.** *Proceedings to 7th Vienna International Conference on Mathematical Modelling MATHMOD 2012 - Vienna (2012)*, 320-325.

- Thomas Müller, <u>Robert Dürr</u>, Britta Isken, Josef Schulze-Horsel, Udo Reichl, Achim Kienle. **Distributed modeling of human influenza a virus-host cell interactions during vaccine production.** *Biotechnology and Bioengineering 110*, 8 (2013), 2252–2266.

- <u>Robert Dürr</u>, Achim Kienle. **An efficient method for calculating the moments of multidimensional growth processes in population balance systems**. *The Canadian Journal of Chemical Engineering 92 (2014)*, 2088–2097.

- <u>Robert Dürr</u>, Thomas Müller, Achim Kienle. **Efficient DQMOM for multivariate population balance equations and application to virus replication in cell cultures.** *Proceedings to 8th Vienna International Conference on*

Mathematical Modelling MATHMOD 2015 - Vienna (2015), 29–34.

- <u>Robert Dürr</u>, Stefanie Duvigneau, Tanja Laske, Mandy Bachmann, Achim Kienle. **Analyzing the impact of heterogeneity in genetically engineered cell lines for influenza vaccine production using population balance modeling.** *6th IFAC Conference on Foundations of Systems Biology in Engineering FOSBE 2016 - Magdeburg (2016).*

- <u>Robert Dürr</u>, Thomas Müller, Stefanie Duvigneau, Achim Kienle. **An efficient approximate moement method for multi-dimensional population balance models - Application to virus replication in multi-cellular systems** *Chemical Engineering Science* (2016), in press.

1.4. Thesis outline

This thesis consists of six chapters which cover the tasks laid out in the motivation:

In the second chapter, a brief introduction to the framework of population balance modeling is provided. The corresponding mathematical formulations are presented and the notation of moments is introduced.

The third chapter covers the top down modeling of an influenza vaccine production process with mammalian cell cultures. The most important modeling steps which led to the final population balance model formulation (see [79]) will be presented. Furthermore, the transformation of the infinite dimensional inverse problem to a finite dimension using a Hermite spline representation of the functional parameters is discussed. Subsequently, the results of the overall model adaption will be discussed and the transition to the bottom-up modeling approach will be motivated.

In the fourth chapter, an efficient method for the approximate moment closure for multi dimensional population balance equations, which is based on the direct quadrature method of moments, will be presented. First, the general problem of moment closure will be explained and the classic formulation of the DQMOM will be given. Analytic solutions for the dynamics of abscissas and weights will be derived for processes which are characterized by particle growth and death. Furthermore, a discrete and a continuous reformulation is provided for processes including nucleation. Subsequently, different strategies for the construction of weight and abscissa sets are presented. The performance will be evaluated for several benchmark problems including crystallization and viral replication in cell cultures.

In the fifth chapter the developed efficient approximate moment method will be applied to a high dimensional PBE for influenza virus replication in mammalian cell cultures to analyze the effects of different levels of parameter heterogeneity on the overall process. Further, the algorithm is used to analyze the effects of nonuniform genetic modifications of the host cells on the overall process.

In the last chapter, the major contributions of this thesis are summarized and an outlook
to possible future research topics is given.

2. The framework of population balance modeling

In the following, the framework of population balance modeling is introduced as presented in [92]. In general, each particle or individual, like a single crystal or cell, can be characterized by its properties. These properties are divided into internal coordinates $\mathbf{x_i}$ (describing e.g. internal composition or particle morphology) and external coordinates $\mathbf{x_e}$ (i.e. spatial coordinates). Together they form the property state space $\mathbf{X} = \mathbf{X_i} \times \mathbf{X_e}$. Each individual is characterized by the overall property state vector $\mathbf{x} = [\mathbf{x_i}, \ \mathbf{x_e}]$. If one is interested in the number of particles within a certain volume of the property space $\widetilde{\mathbf{X}} \subseteq \mathbf{X}$, one has to compute the integral

$$N(t) = \int_{\widetilde{\mathbf{X}}} n(t, \mathbf{x}) \mathrm{d}\,\mathbf{x}\,. \tag{2.1}$$

Here, $n(t, \mathbf{x})$ is the number density function which describes the number of particles in an infinitesimal volume of the property space.

During a process, individual particle properties and the particle location can change. These changes correspond to a movement in the property state space. Thus, the number density function changes and so does in many cases the overall number of particles in $\widetilde{\mathbf{X}}$. In the following, spatial gradients are assumed to be neglible. This means that the overall system is well mixed, i.e. $dim(\mathbf{x_e}) = 0$.

The dynamics of the overall number of particles in the control volume is given by

$$\frac{\mathrm{d}N(t)}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t} \int_{\widetilde{\mathbf{X}}} n(t, \mathbf{x}) \mathrm{d}\mathbf{x} = F_{\mathbf{x}}(t) + P_{\mathbf{x}}(t) \tag{2.2}$$

where $F_{\mathbf{x}}(t)$ characterizes the net amount of particles entering or leaving the control volume over its surface of $\mathbf{S}_{\widetilde{\mathbf{X}}}$ and $P_{\mathbf{x}}(t)$ is the net production rate of particles in the control volume. Introducing transport and production densities, $f_{\mathbf{x}}(t, \mathbf{x})$ and $p_{\mathbf{x}}(t, \mathbf{x})$, (2.2) can be rewritten as

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{\widetilde{\mathbf{X}}} n(t, \mathbf{x}) \mathrm{d}\mathbf{x} = - \int_{\mathbf{S}_{\widetilde{\mathbf{X}}}} \langle f_{\mathbf{x}}(t, \mathbf{x}), \mathfrak{n} \rangle \, \mathrm{d}\mathbf{s}_{\widetilde{\mathbf{X}}} + \int_{\widetilde{\mathbf{X}}} p_{\mathbf{x}}(t, \mathbf{x}) \mathrm{d}\mathbf{x}\,. \tag{2.3}$$

Here $\mathfrak{n}$ represents the outward normal vector on $\mathbf{S}_{\widetilde{\mathbf{X}}}$. Applying Gauss' theorem, the surface integral can be transformed into a volume integral

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{\widetilde{\mathbf{X}}} n(t, \mathbf{x}) \mathrm{d}\mathbf{x} = - \int_{\widetilde{\mathbf{X}}} \nabla_{\mathbf{x}} \left\{ f_{\mathbf{x}}(t, \mathbf{x}) \right\} \mathrm{d}\mathbf{x} + \int_{\widetilde{\mathbf{X}}} p_{\mathbf{x}}(t, \mathbf{x}) \mathrm{d}\mathbf{x}\,. \tag{2.4}$$

This equation is also known as the integral formulation of the population balance equation (PBE). Assuming a constant control volume $\widetilde{X}$ the corresponding local formulation can be obtained

$$\frac{\partial n(t, \mathbf{x})}{\partial t} = -\nabla_{\mathbf{x}} \left\{ f_{\mathbf{x}}(t, \mathbf{x}) \right\} + p_{\mathbf{x}}(t, \mathbf{x}), \qquad (2.5)$$

characterizing the dynamics of the number density distribution. However, the actual expressions for the transport and production densities depend on the system under consideration.

2.1. Modeling of transport and production densities

As an example, an ensemble of particles (e.g. a cell culture in a stirred tank reactor) which differ in one specific property x (e.g. characteristic cell size) is considered. As a result of cell growth the cells size changes with a specific rate G. Thus, the change in size of an individual cell can be described by the following ordinary differential equation (ODE)

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = G(t, x). \qquad (2.6)$$

A change in size can be interpreted as convective transport from an infinitesimal control volume in the space of the cell size coordinate to the neighboring element. The value of the transport flux depends on the rate G and the number density n. It is therefore given by

$$f_{\mathbf{x}} = G(t, x) \, n(t, x). \qquad (2.7)$$

The growth velocity may have different effects on the number density distribution, depending on its functional form: For $G = const.$, all cells grow with the same velocity and the corresponding number density distribution moves through state space while keeping its initial shape (see Fig. 2.1a). If for example $G \sim x^2$, then large cells grow faster than small ones resulting in a broadening of the initial distribution with increasing time, which is illustrated in Fig. 2.1b.

As mentioned previously, the production density $p_{\mathbf{x}}(t)$ characterizes the net amount of newly born and vanished particles or cells. In general, the following subprocesses can be distinguished from another

- **aggregation/fusion**: two or more particles agglomerate and form a new particle (e.g. clumping of blood cells),

- **division/fission**: a particle splits and two or more new particles are formed (e.g. cell division),

- **injection/nucleation**: new particles are introduced into the considered population as a result of interaction with another species (e.g. infection of uninfected cell with virus injects new cells into the population of infected cells),

(a) $G = const.$ (b) $G \sim x^2$
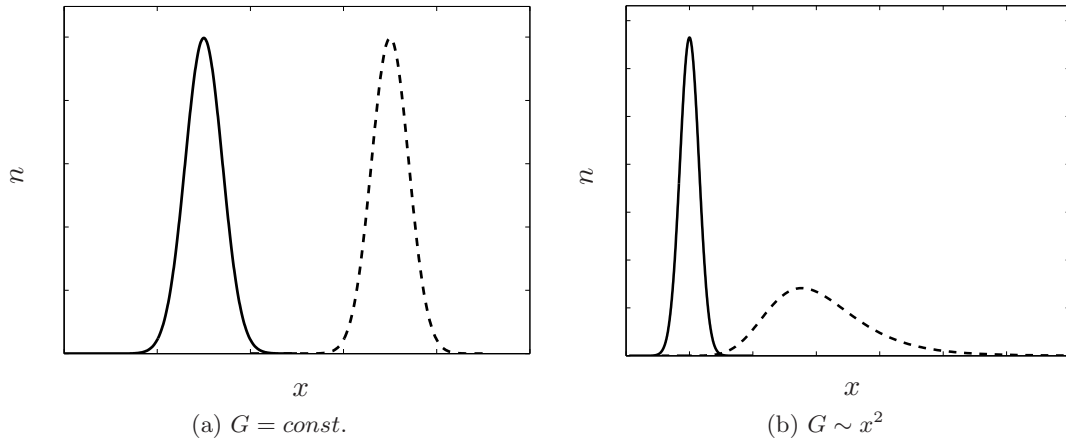
Figure 2.1.: *Effect of different growth functions $G(t, x)$ on the evolution of $n(t, x)$; initial number density distribution is given by the solid curves, final number density distribution is given by the dashed curves*

(a) $D = const.$ (b) $D \sim x^2$
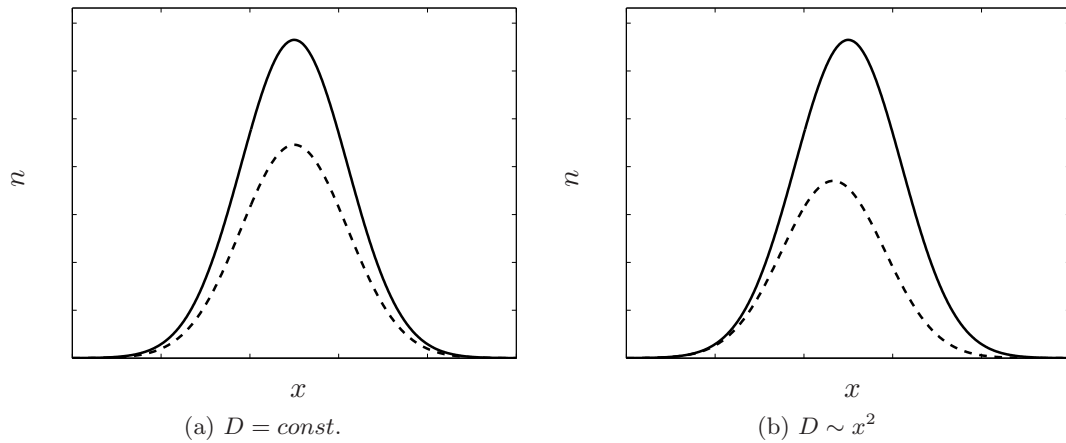
Figure 2.2.: *Effect of different death functions $D(t, x)$ on $n(t, x)$; the initial number density distribution is given by the solid curves, the final number density distribution is represented by the dashed curves*

- **death/withdrawal**: individuals vanish with a certain rate (e.g. cell death or withdrawal of cells from the population via outlet).

In the remainder, focus will be on systems characterized by growth, death and injection. However, to a certain extent, the methods presented in the following can also be applied to systems which include aggregation and division. For example cell division in a single PBE can be rewritten as a series of PBEs characterized by injection and death (see e.g. [57]).

The resulting production density can be expressed by

$$p_{\mathbf{x}}(t) = -D(t,x)\, n(t,x) + Nuc(t,x)\,, \tag{2.8}$$

where the structure of the death rate D can have effects on the number density distribution: If for example $D = const.$, all individuals die at the same rate (see Fig. 2.2a). The shape of n is not affected but with increasing time the magnitude is decreasing. On the other hand, if D depends on the characteristic cell size x (see Fig. 2.2b), some cells are "favored" for "withdrawal" over others, resulting in a changing shape over time. In contrast to the reproduction rate, the injection/nucleation rate is generally not proportional to n.

In result, the overall PBE in the local form for the example above is given by

$$\frac{\partial n(t,x)}{\partial t} + \frac{\partial}{\partial x}\left\{G(t,x)\, n(t,x)\right\} = -D(t,x)\, n(t,x) + Nuc(t,x)\,. \tag{2.9}$$

If more than one particle property is taken into account, a general multi dimensional PBE in local form can be obtained in a similar manner

$$\frac{\partial n(t,\mathbf{x})}{\partial t} + \nabla\left\{\mathbf{G}(t,\mathbf{x})\, n(t,\mathbf{x})\right\} = -D(t,\mathbf{x})\, n(t,\mathbf{x}) + Nuc(t,\mathbf{x})\,. \tag{2.10}$$

Here, $\mathbf{G}$ is the vector of growth rates that characterizes the change of properties for each particle.

2.2. Definition of moments

In particular, for the multi dimensional case, it can be difficult to interpret number density distributions. However, there are integral measures of the number density distributions, like mean and variance, which are much easier to handle. These quantities are related to the total moments of a distribution. For a density distribution with N_d internal properties a total moment is defined as

$$m_{l_1,\dots,l_{N_d}}(t) = \int\limits_{\mathbf{X}} x_1^{l_1} x_2^{l_2} \cdots x_{N_d}^{l_{N_d}}\, n(t,\mathbf{x})\mathrm{d}\mathbf{x}\,. \tag{2.11}$$

Unless it is stated otherwise, the notion "moment" instead of "total moment" is used in the following. The dynamics of a moment follows from integration of the general PBE in local form (2.5)

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} m_{l_1,\dots,l_{N_d}}(t) = -\sum_{k=1}^{N_d} \int\limits_{\mathbf{X}} x_1^{l_1} \dots x_{N_d}^{l_{N_d}} \frac{\partial}{\partial x_k}\left\{G_k(t,\mathbf{x})\, n(t,\mathbf{x})\right\} \mathrm{d}\mathbf{x} \\
- \int\limits_{\mathbf{X}} x_1^{l_1} \dots x_{N_d}^{l_{N_d}} D(t,\mathbf{x}) n(t,\mathbf{x})\, \mathrm{d}\mathbf{x} + \int\limits_{\mathbf{X}} x_1^{l_1} \dots x_{N_d}^{l_{N_d}} Nuc(t,\mathbf{x})\, \mathrm{d}\mathbf{x}\,.
\end{aligned}$$

$$\tag{2.12}$$

In the case of a one dimensional number density distribution with respect to particle length and under the assumption of a spherical particle shape, the overall number, mean length and mean surface of the particle ensemble are given by

$$N_{\text{over}} = m_0 = \int\limits_{\mathbf{X}} n(t, x) \, \mathrm{d}x \, ,$$

$$x_{\text{mean}} = \frac{m_1}{m_0} = \frac{\int\limits_{\mathbf{X}} x \, n(t, x) \, \mathrm{d}x}{N_{\text{over}}} \, ,$$

$$A_{\text{mean}} = \frac{A_{ges}}{m_0} = \frac{\int\limits_{\mathbf{X}} \pi \, x^2 \, n(t, x) \, \mathrm{d}x}{N_{\text{over}}} \, . \tag{2.13}$$

However, only for restricted classes of G, D and Nuc the moments can be computed from a finite set of equations in closed form. Alternatively, an approximate closure can be employed which will be discussed in detail in the fourth chapter.

2.3. Coupling to other species

Frequently, the dynamics of the number density distribution is coupled to a continuous phase whose dynamics are given by the system of ODEs

$$\frac{\mathrm{d}\mathbf{c}(t)}{\mathrm{d}t} = Dil \left(\mathbf{c}_{in} - \mathbf{c}(t) \right) + \mathbf{f_I}(t, \mathbf{c}(t), \mathbf{I}(t, n(t, \mathbf{x}))) \tag{2.14}$$

with Dil being the dilution rate and $\mathbf{f_I}$ characterizing the integral coupling of particle ensemble (i.e. dispersed phase) and continuous phase by means of moments or more general integral quantities $\mathbf{I}$ of the number density distribution. A simple example is found in bioprocess engineering: For the cultivation of cells in a defined environment, e.g. a stirred tank reactor, it is necessary to provide certain substrates (e.g. fructose and ammonium). The cell population (i.e. the dispersed phase) is taking up the substrates dissolved in the continuous phase and metabolizes them. In result, the cells increase in size or reproduce by means of cell division. The higher the concentration of cells in the reactor, the higher the uptake rate of substrates. If the necessary substrates are fully consumed and no new substrates are added to the reactor, the cells will no longer grow and divide.

Further, the particle ensemble may also be coupled to other cell populations, which can be described by additional number density distributions. An everyday example is found in the human body where different immune cell populations influence each other to fight pathogens. The dynamics of the resulting multi disperse process is characterized by a system of PBEs. Another example will be presented in the next chapter: Uninfected, infected and apoptotic cells are considered to account for cell-to-cell variability within an influenza vaccine production process.

3. Parameter estimation for top-down population balance model of influenza vaccine production process

The present chapter is concerned with parameter estimation for a population balance model describing an influenza virus replication in mammalian cell cultures which was proposed by Müller et al. [81, 79]. As the model features functional parameters model adaption involves the solution of an infinite dimensional inverse problem. A solution strategy will be developed and the resulting parameter estimates will be discussed. Results of the paramter estimation of this thesis were published in advance in [21].

3.1. Description of the process

Influenza is a highly contagious disease which is caused by infection with the influenza virus. This virus is classified into subtypes A, B and C. Among those the first is the most relevant affecting humans. It is characterized by a strong antigenic variability and can affect a wide range of possible hosts such as several mammal species and birds. It does not only occur in seasonal waves, but can also cause severe pandemic outbreaks. The treatment of both pose high demands to the public health care system. Vaccination is recognized as the best method to prevent infection and a possibly fatal course of the disease. The vaccine consists of inactivated virus particles which are traditionally produced in embryonated chicken eggs. This production process is well established and relatively cheap. However, there are several disadvantages e.g. a limited egg production capacity and less flexible production process, resulting in a delayed vaccine production in case of pandemic outbreaks.

Alternatively, cell cultures in bioreactors can be used instead of these "natural reactors" [30]. In combination with experimental work, mathematical modeling of the production processes is necessary to gain a deeper understanding of the underlying processes and to come up with suitable schemes for process intensification and process control.

In the following, the experimental setup will be briefly outlined. For a more detailed description see e.g. [102, 79]. Adherent *Madin-Darby-Canine-Kidney* (MDCK) cells are grown on microcarriers within a small-scale bioreactor. After the microcarriers are fully covered with a confluent layer of cells, fresh medium is added and the reactor is inoculated with virus seed. In general, the latter consists of infectious and non-infectious virus particles. Both have the ability to bind to the surfaces of uninfected cells. However,

the latter fail to initiate a successful viral replication cycle. After being infected, the cells start to synthesize viral compounds and subsequently fully assembled virus particles are released to the surrounding medium. Here, they can bind to still uninfected cells.

As in many bioprocesses, cell-to-cell variability is a common feature and has to be taken into account. In the current process, this heterogeneity is measured in terms of intracellular amount of viral nucleoprotein (NP). These can be measured with fluorescence microscopy and quantified with sophisticated techniques, e.g. flow cytometry. Such measurements have not only shown the existence of cell-to-cell variability as a matter of principle (see Fig. 3.1) but also revealed interesting dynamic phenomena which are depicted in Fig. 3.2. Most notable are the transient bimodality and the backshift of the distribution at late sample times. The experimental observations have been the
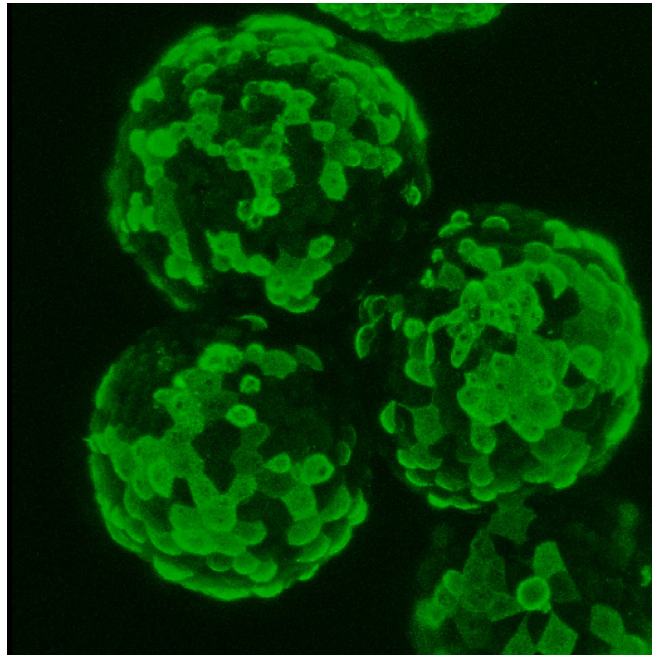
Figure 3.1.: *Fluorescence microscopic measurements showing cell-to-cell variability with respect to different degrees of fluorescences which are a measure of the intracellular amount of viral nucleoprotein (NP) [21]*

major motivation for the modeling efforts which are documented in the doctoral thesis of Thomas Müller [79]. There, a population balance model was developed to account for the described heterogeneous effects. Focus of [79] was on model formulation. Parameter identification, which is crucial to link the model with the experimental results was done within the present thesis and will be discussed in the following. Main results were published in advance together with Thomas Müller [80, 21, 81]. Experimental data were provided by the Bioprocess Engineering Group of the Max Planck Institute in Magdeburg.
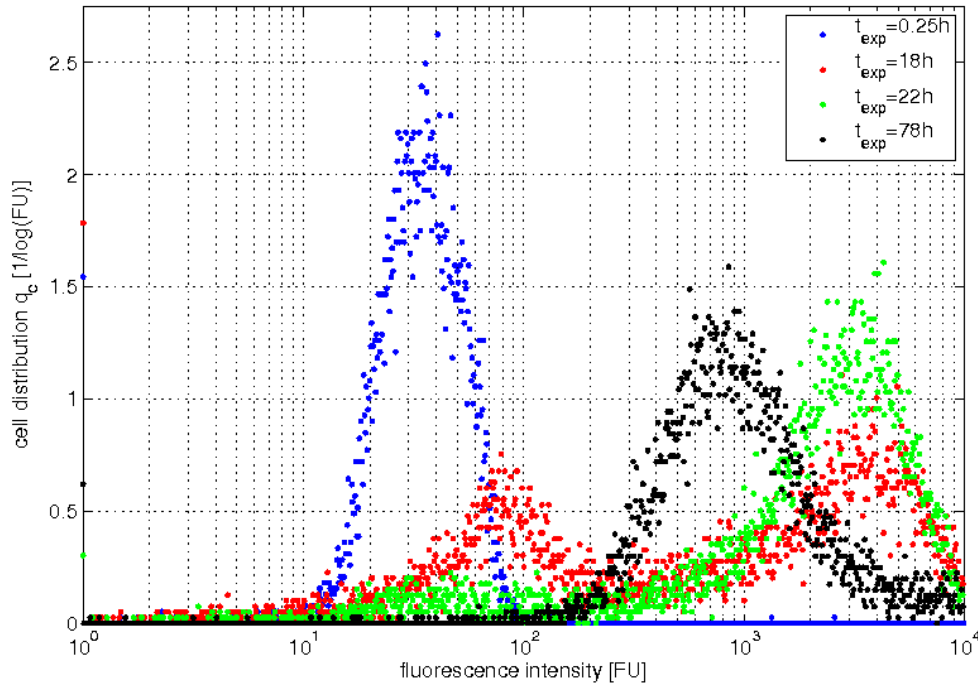
Figure 3.2.: *Number density distributions for selected sample times; At t = 18h a bimodal number density distribution is measured, which vanishes for later samples; A backshift is observed for late sample times*

3.2. Model Formulation

In the following, the development of the distributed model formulation will be summarized briefly following [79].

A first step towards the modeling of the influenza vaccine production processes was made by Möhler and coworkers [76]. Therein, an unstructured model of the process was derived characterizing the dynamics of uninfected cells, infected cells and free virus particles within a system of coupled ODEs. The basic scheme is depicted in Fig. 3.3. Virus particles are released from infected cells, which are "produced" by infection of uninfected cells. This model formulation constitutes a modification of the SIR models from classical epidemiology (see e.g. [83, 87]).

Next, the process description was extended additionally considering apoptosis [104, 102]. Apoptosis is a type of programmed cell death which is induced as a result of environmental stress or in course of the infection. It is known to be an integral part of the host cells defense against viral infections (see e.g. [101]). Nevertheless, more recent research indicates a more complex role within the virus replication cycle [54]. In fact, the genome of influenza A virus codes different types of proteins that can trigger both, apoptosis induction as well as apoptosis inhibition. Their interplay manipulates intracellular pathways which support an efficient replication of the virus within the cell
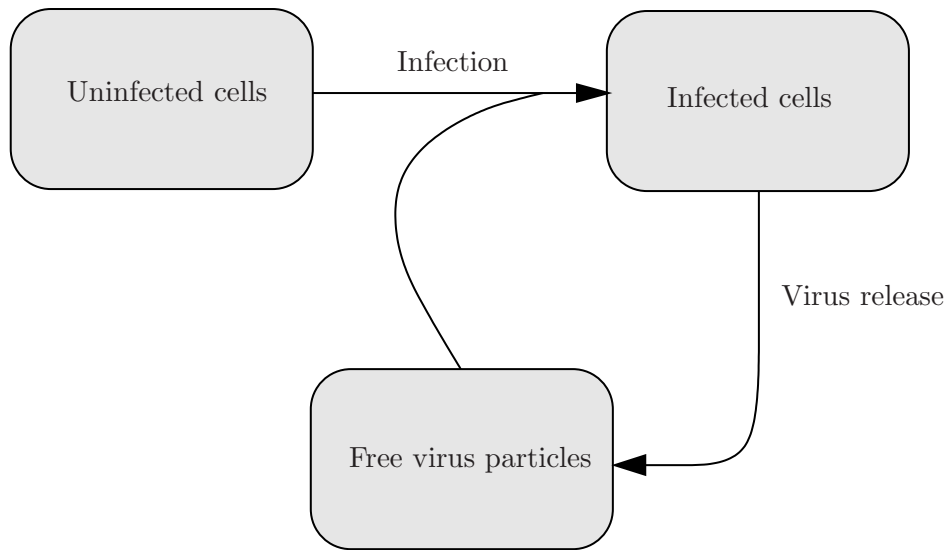
Figure 3.3.: *Basic reaction scheme for model of influenza virus replication in MDCK cell cultures developed by Möhler and coworkers [76]; Virus particles bind to uninfected cells; After intracellular replication cycle, new virus particles are released to the medium*

culture. A detailed description of apoptosis and its effects on the viral production cycle is found in [102] and [79]. In the model formulation of Schulze-Horsel and coworkers [104] an additional species of apoptotic cells was introduced. Furthermore, the free virus particles in the medium are subdivided into infectious virions and non-infectious virions. The corresponding scheme is shown in Fig. 3.4. Apoptotic cells can emerge from uninfected and infected cells. Within the model formulation it is assumed that apoptotic cells are in a late stage of apoptosis, in which most intracellular mechanisms for replication of viral proteins and release of virus to the medium are shut down. In result, they are not able to replicate the viral genome and release virus particles.

Subsequently, the model formulation was further extended by segregation of the cells with respect to the intracellular amount of viral nucleoprotein NP. This quantity can be interpreted as a degree of infection. While focus was first on equine influenza A virus strains [82], it was shifted later to human influenza A strains [80, 81]. The resulting model scheme which is adopted from [79] can be seen in Fig. 3.5. Here, the cells are distinguished with respect to their specific degree of fluorescence φ. This quantity is proportional to the amount of intracellular NP and can be measured directly with flow cytometry. A more detailed description of the relation between the degree of fluorescence φ and the amount of intracellular NP will be given in the next section. The main principle is the following: If a cell synthesizes a certain amount of NP, its fluorescence increases. If NP binds to other viral proteins or is released from the cells within a newly assembled virus particle, the fluorescence decreases.
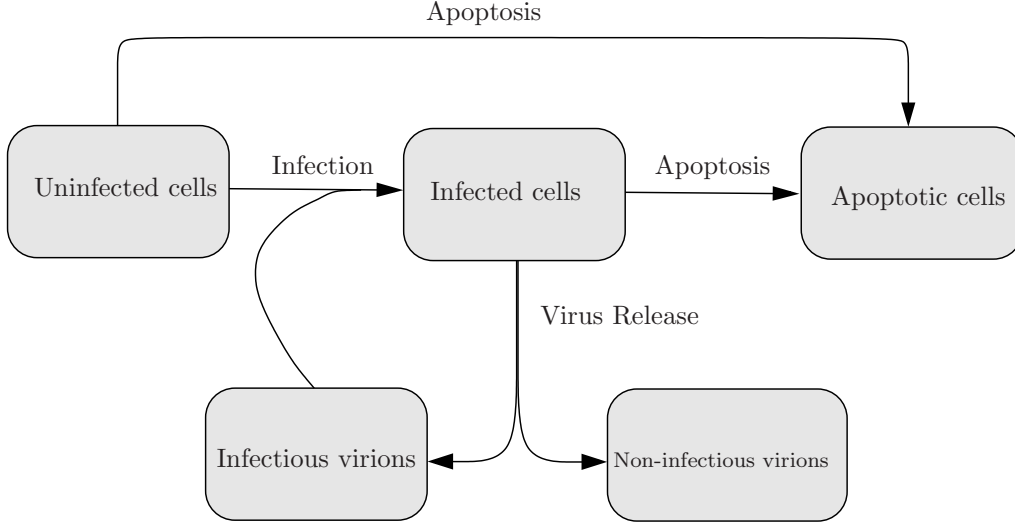
Figure 3.4.: *Basic scheme for the model formulation developed by Schulze-Horsel and coworkers [104]; in addition to Fig. 3.3, apoptotic cells are considered and infectious and non-infectious virions are distinguished*

Flow cytometric measurements of cells in very early stages after inoculation show a distribution at low degrees of fluorescence (see Fig. 3.2) which is caused by an unspecific antibody binding during the measurement procedure. This observation is also included in the model formulation by differentiation of the uninfected cells with respect to their degree of fluorescence. The dynamics of the uninfected cells are dominated by the infection of uninfected cells with active (i.e. infectious) virions as uninfected cells can not replicate viral proteins without being infected. Furthermore, cell death is neglected as the cell death rate of uninfected cells is considered to be much lower than the cell death rate of infected cells. In addition, growth can also be neglected for this specific type of adherent MDCK cells (see [79] for a detailed explanation). With this assumptions the temporal evolution of the number density distribution of the uninfected cells is described by

$$\frac{\partial u_c(t, \varphi)}{\partial t} = -k_{vi} u_c(t, \varphi) \ V_{ac}(t) \,. \tag{3.1}$$

After infection (k_{vi} denotes the infection rate coefficient), newly infected cells undergo a latent phase of length τ_{lat} in which the intracellular amount of NP is not increasing. Thus, the cells retain their specific fluorescence. This lag phase lumps together several intracellular processes which happen prior to the synthesis of NP, e.g. virus entry into the cell via endocytosis, uncoating of the viral nucleic acid, nuclear import and RNA transcription. The dynamics of the latent cell number density distribution is given by

$$\frac{\partial l_c(t, \varphi)}{\partial t} = k_{vi} u_c(t, \varphi) \ V_{ac}(t) - k_{vi} u_c(t - \tau_{lat}, \varphi) \ V_{ac}(t - \tau_{lat}) \,. \tag{3.2}$$

It has to be mentioned, that cell death of these cell species is not considered within the model formulation.
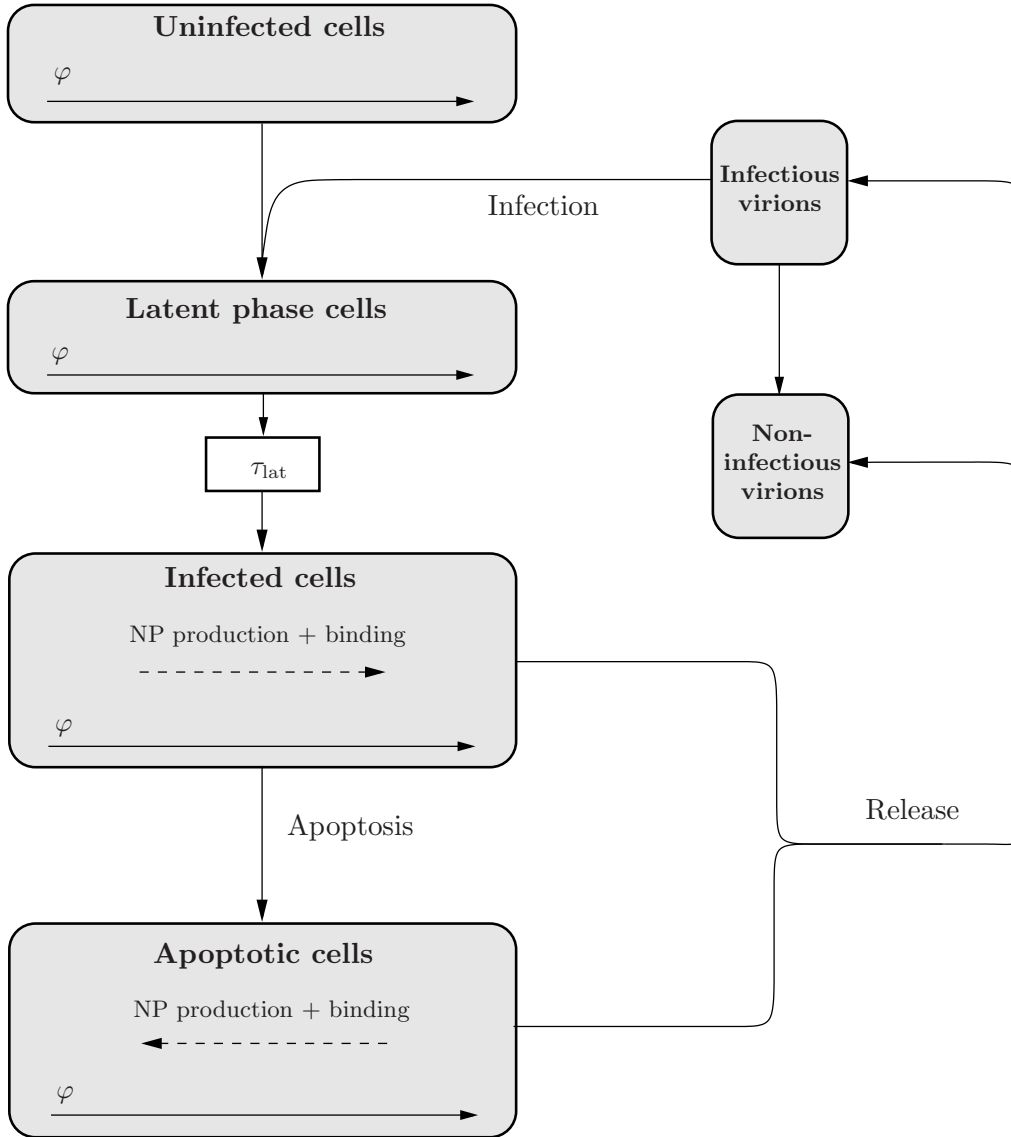
Figure 3.5.: *Detailed scheme for model developed by Müller et al. [81]; all cell species are segregated with respect to the specific fluorescence degree; virus particles are released by both, infected and apoptotic cells*

Following the latent phase, infected cells start to produce viral NP with rate coefficient $k_{\mathrm{prod}}(\varphi)$ thereby increasing the degree of fluorescence. Simultaneously, NP binds with coefficient $k_{\mathrm{bind}}(\varphi)$ during the formation of viral RNP which causes a decreasing degree of fluorescence Thus, the net rate coefficient is given by

$$k_{\mathrm{net}}(\varphi) = k_{\mathrm{prod}}(\varphi) - k_{\mathrm{bind}}(\varphi) \, . \tag{3.3}$$

Here, it is assumed that the net rate coefficient depends on the specific degree of fluorescence φ. As NP production is assumed to be faster than NP binding for infected

cells, the corresponding net coefficient is always assumed positive. In result, the degree of fluorescence generally increases for infected cells. The overall dynamics of the infected cells number density distribution is characterized by

$$\frac{\partial i_c(t, \varphi)}{\partial t} = k_{\text{vi}} \ u_c(t - \tau_{\text{lat}}, \varphi) \ V_{\text{ac}}(t - \tau_{\text{lat}}) - \frac{\partial}{\partial \varphi} \{k_{\text{net}}(\varphi) \ i_c(t, \varphi)\} - k_{\text{apo}}(\varphi) \ i_c(t, \varphi) \, .$$

$$(3.4)$$

In course of the infection and the viral replication cycle, infected cells become apoptotic. The apoptosis rate is proportional to the coefficient $k_{\text{apo}}(\varphi)$ which is also a function of the fluorescence degree. Note, that in contrast to the model formulation of Schulze-Horsel and coworkers [104] the apoptotic cells are assumed to be in an early stage of apoptosis. These are still able to synthesize viral compounds and release virus particles. However, apoptosis is assumed to affect the viral replication cycle resulting in an increased formation rate of RNP complexes and a decreased production of NP. Hence, the overall net rate coefficient

$$k_{\text{net,apo}}(\varphi) = k_{\text{prod,apo}}(\varphi) - k_{\text{bind,apo}}(\varphi) \tag{3.5}$$

characterizing the rate of change in fluorescence of apoptotic cells is assumed to be negative. Furthermore, lysis of apoptotic cells is accounted for by the coefficient $k_{\text{cd}}(\varphi)$. In result, the following PBE describing the apoptotic cell number density dynamics is obtained

$$\frac{\partial a_c(t, \varphi)}{\partial t} = k_{\text{apo}}(\varphi) i_c(t, \varphi) - \frac{\partial}{\partial \varphi} \left\{ k_{\text{net,apo}(\varphi)} \ a_c(t, \varphi) \right\} - k_{\text{cd}} \ a_c(t, \varphi) \, . \tag{3.6}$$

Both, infected and apoptotic cells are assumed to release two types of virus particles, active (i.e. infectious) and inactive (i.e. non-infectious). The production ratio between the two is regulated via an efficiency factor $P_{\text{eff}} \in [0, 1]$. Furthermore, active virions degrade to inactive ones with coefficient k_{deg}. Assuming a well mixed medium, balancing yields the following dynamics of the concentrations of active virions

$$\frac{\partial V_{\text{ac}}(t)}{\partial t} = P_{\text{eff}} \ k_{\text{rel}} \int\limits_{\varphi_{\text{min}}}^{\varphi_{\text{max}}} \{i_c(t, \varphi) + a_c(t, \varphi)\} \, \mathrm{d}\varphi$$

$$- k_{\text{vi}} \ V_{\text{ac}}(t) \int\limits_{\varphi_{\text{min}}}^{\varphi_{\text{max}}} u_c(t, \varphi) \ \mathrm{d}\varphi - k_{\text{deg}} \ V_{\text{ac}}(t) \tag{3.7}$$

and inactive virions

$$\frac{\partial V_{\text{inac}}(t)}{\partial t} = (1 - P_{\text{eff}}) \ k_{\text{rel}} \int\limits_{\varphi_{\text{min}}}^{\varphi_{\text{max}}} \{i_c(t, \varphi) + a_c(t, \varphi)\} \, \mathrm{d}\varphi + k_{\text{deg}} \ V_{\text{ac}}(t) \, . \tag{3.8}$$

3.3. Numerical solution of the process model

The presented model equations represent a coupled system of PDEs and ODEs. For the numerical solution this system was reduced to a set of ODEs as presented in [79]. In the following, the major points of the numerical solution method are summarized.

At first, the dynamic equations characterizing uninfected and latent cells were transformed to decrease the computational effort. It is assumed, that in cells from these species viral NP is neither produced nor bound or released and the infection coefficient k_{vi} does not depend on φ. Thus, the corresponding fluorescence degrees are neither increasing nor decreasing. All necessary information for the fluorescence distributions of these cell species is contained in the initial distribution of the uninfected cells, which is time invariant. To obtain time variant distributions of uninfected and latent phase cells, the time invariant initial distribution has to be multiplied with the time variant overall numbers of uninfected and latent cells. The resulting dynamics of the overall number of uninfected cells is given by an ODE. In contrast, the dynamics of the overall number of latent cells is characterized by a delay-differential equation which can be represented by a transport system. For the numerical solution, this transport system was reduced to a set of $N_{\mathrm{CV,lat}} = 50$ ordinary differential equations using a Finite Volume Method.

In the next step, the partial differential equations representing the dynamics of infected (3.4) and apoptotic cells (3.6) were discretized using a finite volume approach with $N_{\mathrm{CV}} = 128$ control volumes on a logarithmic grid. The virus release integral in (3.7) and (3.8) was approximated using the rectangular rule.

Including the dynamics of active and inactive virus particles, the overall size of the resulting system of ordinary differential equations is given by

$$N_{\mathrm{ODE}} = 1 + N_{\mathrm{CV,lat}} + 2 \cdot N_{\mathrm{CV}} + 2 = 309 \,. \tag{3.9}$$

This system was solved numerically using *ode15s* in MATLAB.

3.4. Measurement data

All experiments and the corresponding measurements were performed by Josef Schulze-Horsel. In the following the measurement setup will be briefly reviewed. For a more detailed description of the measurement procedure see [104] and [102].

3.4.1. Process measurements

To obtain information about the intracellular amount of viral nucleoprotein, sampled cells are stained with fluoresceine-conjugated monoclonal antibodies against NP. Thereby, a certain amount of NP is assumed as a virus equivalent and can thus be interpreted as degree of infection. Antibodies binding to their specific antigens cause

an increase in the measured fluorescence intensity F of the considered cell. In [103] it is shown, that the change of the intracellular amount of viral NP is linked linearly to the change in fluorescence intensity. A certain amount of binding antibodies is necessary to increase the fluorescence intensity by one fluorescence unit FU. In addition, there is an unspecific antibody binding which affects all types of cells, uninfected, infected and apoptotic. This leads to the introduction of a more general degree of fluorescence φ as an internal coordinate [79]

$$\varphi = \frac{F}{F_{\mathrm{VE}}} \, . \tag{3.10}$$

It can be defined as the ratio between the fluorescence intensity of a cell and the change in fluorescence intensity per virus equivalent F_{VE} which is given by [102]

$$F_{\mathrm{VE}} = 2.661 \ FU \, . \tag{3.11}$$

Depending on the individual fluorescence, sampled cells are categorized into 1024 classes during flow cytometric analysis. Thus, for each sample time t_k the counts for each class

$$Z_i(t_k) \, , \ \text{for } i = 1, \dots, 1024 \tag{3.12}$$

are available. This implies, that a measured distribution is a sum signal merging the measurement information about uninfected, latent, infected and apoptotic cell species. A detailed analysis of experimental data suggests that a simple heuristic differentiation into these classes as often applied in biology is not suitable. In addition to flow cytometric measurements, the overall cell concentration $C(t_k)$ was quantified via cell counts.

The concentration of active virus particles V_{ac} is quantified by TCID_{50} titration [43] and overall virus concentration V_{tot} was measured with a hemagglutination assay (HA) [52].

3.4.2. Transformation of flow cytometric data

As the model is solved on a logarithmic grid with N_{CV} control volumes, the available flow cytometric measurement information has to be transformed to fit the same grid. Furthermore, it has to be taken into account, that flow cytometric samples only give information about the fluorescence distribution of a representative number of cells ($\approx 10^4$ cells). To map the samples number density distribution to the whole cell culture within the reactor, each flow cytometric sample has to be scaled with the total concentration of cells at the corresponding sample time.

Taking both facts into account, the overall concentration of cells within one control

volume $c_i^*(t_k)$ at sample time t_k is given by

$$c_i^*(t_k) = \frac{\sum\limits_{j=\widetilde{N}(i-1)}^{\widetilde{N}i} Z_j(t_k)}{\sum\limits_{j=1}^{1024} Z_j(t_k)} C(t_k) \quad , \text{ for } i = 1, \ldots, N_{\mathrm{CV}} \quad , \widetilde{N} = 1024/N_{\mathrm{CV}} \;. \quad (3.13)$$

3.5. Parameter estimation

The model of the process (3.1)–(3.8) was adapted to experiments with different virus strains. To determine the unknown model parameters, an inverse problem has to be solved by minimizing the weighted squared error between data $\mathbf{y}^*$ and model prediction $\mathbf{y}$ [116]

$$\min J = \min \left\{ (\mathbf{y}^* - \mathbf{y})^T \cdot W \cdot (\mathbf{y}^* - \mathbf{y}) \right\}$$
$$\text{s.t. } \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = f(\mathbf{x}, \mathbf{p})$$
$$\mathbf{y} = g(\mathbf{x}, \mathbf{p}) \quad (3.14)$$

Therein, f represents the dynamics of the vaccine production process and the operator g relates the model states to the predicted model output. The model states $\mathbf{x}$ correspond to the discretized number density distributions for the distributed cell species and the active and inactive virus particle concentrations. The vector of unknown model parameters $\mathbf{p} = [\mathbf{p}_{\mathrm{const}}, \mathbf{p}_{\mathrm{funct}},]$ consists of constant rate coefficients characterizing the length of the latent phase, the infection rate, the efficiency factor, the virus release rate and the degradation of active virions

$$\mathbf{p}_{\mathrm{const}} = [\tau_{\mathrm{lat}}, \, k_{\mathrm{vi}}, \, P_{\mathrm{eff}}, \, k_{\mathrm{rel}}, \, k_{\mathrm{deg}}] \quad (3.15)$$

and functional rate coefficients characterizing net viral protein replication/release, apoptosis and cell lysis

$$\mathbf{p}_{\mathrm{funct}} = [k_{\mathrm{net}}(\varphi), \, k_{\mathrm{apo}}(\varphi), \, k_{\mathrm{net,apo}}(\varphi), \, k_{\mathrm{cd}}(\varphi)] \;. \quad (3.16)$$

The latter depend on the intracellular coordinate φ. In result, the inverse problem is set in an infinite dimension and has to be transformed to a finite dimension which will be discussed in the following.

3.5.1. Transformation of the inverse problem to a finite dimension

To transform the inverse problem to a finite dimension a finite dimensional representation of the functional parameters has to be found. In general, two solution approaches can be applied. In simple cases, some standard functional dependency of the functional parameter on the intracellular coordinate can be applied (e.g. a Gaussian, see [105]).

If no prior information on the shape of the functional parameter is available, a general approximation in terms of basis functions is more reasonable [59, 4, 5, 64, 65].

$$k_\circ(\varphi) = \sum_{k=1}^{N_k} a_k \ f_k(\varphi) . \tag{3.17}$$

Simple examples for these types of approximations include power series, sums of Gaussians but also piecewise constant approximations. Alternatively, more complex approximation schemes like the piecewise cubic Hermite spline approximation can be applied, as presented in [59]. Here, cubic polynomials are defined locally on an interval $[\varphi_{i-1}, \ \varphi_i]$. This specific approximation is characterized by a continuous first derivative with respect to φ (see Fig. 3.6 for a schematic representation). In result, each functional parameter is represented by a finite set of value pairs $\{\varphi_i, \ k_\circ(\varphi_i)\}$ which determine the locations and values of the so called spline nodes. If the locations of the nodes are fixed a priori,
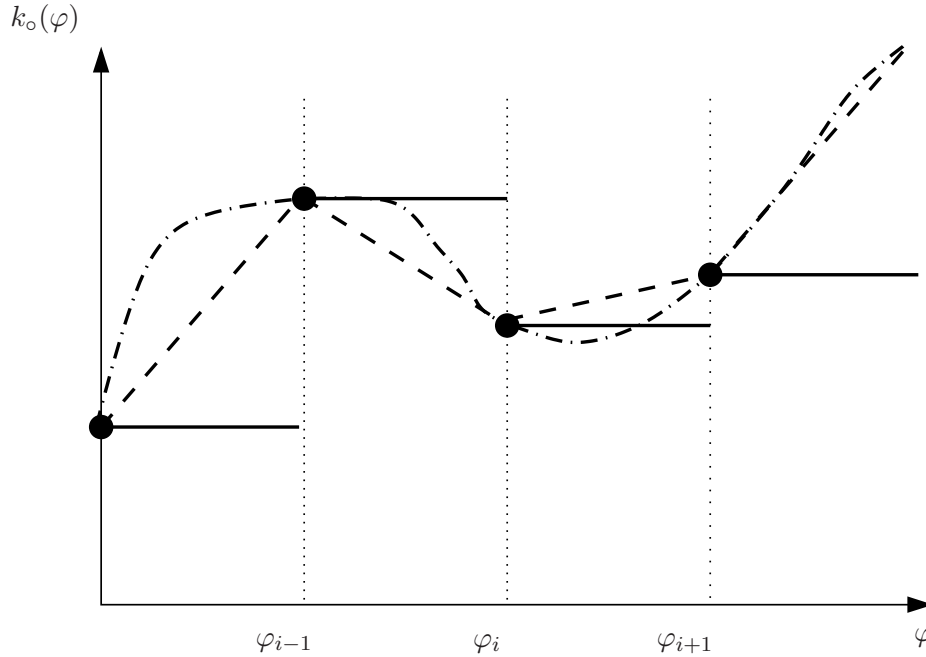
Figure 3.6.: *Schematic representation of piecewise constant (solid), piecewise linear (dashed) and piecewise cubic hermite spline interpolation (dash-dotted)*

the overall set of unknown parameters is further reduced to

$$\mathbf{p} = [\tau_{\mathrm{lat}}, \ k_{\mathrm{vi}}, \ P_{\mathrm{eff}}, \ k_{\mathrm{rel}}, \ k_{\mathrm{deg}}, \ k_{\mathrm{net}}(\varphi_1), \ldots, k_{\mathrm{net}}(\varphi_{i_{\mathrm{max}}}), \ k_{\mathrm{apo}}(\varphi_1), \ldots,$$
$$k_{\mathrm{apo}}(\varphi_{i_{\mathrm{max}}}), \ k_{\mathrm{net,apo}}(\varphi_1), \ldots, k_{\mathrm{net,apo}}(\varphi_{i_{\mathrm{max}}}), \ k_{\mathrm{cd}}(\varphi_1), \ldots, k_{\mathrm{cd}}(\varphi_{i_{\mathrm{max}}})] . \tag{3.18}$$

The degree of freedom of the corresponding inverse problem can be adjusted by increasing or decreasing the number of nodes characterizing the functional parameters. In general, a larger degree of freedom in terms of a large number of overall model

parameters promotes a better fit of the model prediction to the experimental data. Nevertheless, in doing so one does not only increase the computational effort of the corresponding optimization problem but probably also worsens the robustness of the parameter estimation setup. Small deviations in the measurement data would result in large differences in the estimated parameters. This also poses the question of the optimal degree of freedom which should be used in the model formulation.

3.5.2. Overall parameter estimation setup

The unknown kinetic parameters are estimated within a nonlinear weighted least squares framework by minimizing the cost function

$$\min J = \min \sum_{k=1}^{k_{\max}} e^T(t_k) \cdot W(t_k) \cdot e(t_k) \; . \tag{3.19}$$

Therein, the error between measurement $y^*(t_k)$ and model output $y(t_k)$ is given by

$$e(t_k) = y^*(t_k) - y(t_k) = \begin{pmatrix} c_1^*(t_k) \\ \vdots \\ c_{N_{\mathrm{CV}}}^*(t_k) \\ V_{\mathrm{ac}}^*(t_k) \\ V_{\mathrm{tot}}^*(t_k) - V_{\mathrm{ac}}^*(t_k) \end{pmatrix} - \begin{pmatrix} c_1(t_k) \\ \vdots \\ c_{N_{\mathrm{CV}}}(t_k) \\ V_{\mathrm{ac}}(t_k) \\ V_{\mathrm{tot}}(t_k) - V_{\mathrm{ac}}(t_k) \end{pmatrix} \; . \tag{3.20}$$

As explained previously, during flow cytometric measurement no distinction is made between the cell species. Hence, the corresponding model output, i.e. the simulated number of all cells with a specific degree of fluorescence, is given by

$$c_i(t_k) = U_{c,i}(t_k) + L_{c,i}(t_k) + I_{c,i}(t_k) + A_{c,i}(t_k) \quad \text{for} \quad i = 1, \dots, N_{\mathrm{CV}} \tag{3.21}$$

The weighting matrix was chosen according to the measurement variance

$$W(t_k) = W = \mathrm{diag}\left(\sigma_i^{-2}\right) \tag{3.22}$$

with

$$\begin{aligned} \sigma_{1,\dots,N_{\mathrm{CV}}}(t_k) &= \max_{\forall k, \ i=1,\dots,N_{\mathrm{CV}}} \left(c_i^*(t_k)\right) \\ \sigma_{N_{\mathrm{CV}}+1}(t_k) &= \max_{\forall k} \left(V_{\mathrm{ac}}^*(t_k)\right) \\ \sigma_{N_{\mathrm{CV}}+2}(t_k) &= \max_{\forall k} \left(V_{\mathrm{tot}}^*(t_k) - V_{\mathrm{ac}}^*(t_k)\right) \; . \end{aligned} \tag{3.23}$$

Thereby, it is assumed that the error variance is proportional to the maximum value of the corresponding measurement and does not scale with the current measurement.

The optimization was solved numerically using the MATLAB optimization toolbox function *fmincon* with the *active-set* optimization strategy. During the parameter estimation, a high sensitivity of the overall estimation setup on the parameter characterizing the duration of the latent phase τ_{lat} was revealed. For this reason, a two step

parameter estimation setup was implemented. At first, τ_{lat} was kept fixed and all other parameters were estimated. In the second step, the optimized parameters from the first step were kept fixed and the duration of the latent phase was estimated. The overall estimation setup may trap in local minima. For this reason, the parameter estimation was performed several times with different initial parameter guesses.

3.6. Results

The mathematical model of the process, represented by (3.1)–(3.8) has been adopted to measurements with different influenza virus strains using the previously described numerical solution approach (see [81] and [79]). In the following, focus will be on experiments with a human influenza A/Wisconsin/67/2005 (H3N2) virus strain.

The experiments with H3N2 were performed with an initial uninfected cell concentration of $U_c(t = 0) = 1.65 \ 10^6 (ml^{-1})$. The initial multiplicity of infection (MOI) characterizing the ratio of active virus particles and uninfected cells is given by 0.025. The initial fluorescence distributions of the uninfected cells were adopted from the flow cytometric sample at $0.25 \ h$ post infection (p.i.), i.e. the time after inoculating the reactor with virus seed. At this short period after seeding, it is reasonable to assume that the virus particles have not yet altered the cells degree of fluorescence.

3.6.1. Increasing complexity of the functional parameter approximation

At first, it was investigated how an increasing complexity of the functional parameter approximation improves the fit of the model prediction to the measurement data. The initial value for the length of the latent phase was chosen as $t_{\text{lat}} = 5 \ h$.

As presented in Section 3.5.1, the functional parameters characterizing the synthesis and release of viral NP, the apoptosis and the cell lysis were approximated by Hermite splines. In result, the unknown functional parameters are represented by sets of constant node values. At first, each of the functional parameters was represented by two nodes. The nodes are located at the minimum and the maximum value of the degree of fluorescence. Thereby, the functional parameters exhibit a linear dependency on the degree of fluorescence, which was already presented in [81]. The simulation results with the estimated parameter set for the fluorescence distributions are shown and compared to flow cytometric data in Fig. 3.7. The measurement data show only a slight increase in the fraction of cells with a higher degree of fluorescence for the sample at $14 \ h$ p.i. which is also observed for the simulation results. At $18 \ h$ p.i. a significant bimodality can be seen with one peak in a region of lower degree of fluorescence and one in a region of higher degree of fluorescence. The simulation results show, that the bimodal behavior can be reproduced qualitatively. However, the adapted model matches the second peak much better than the first one. For late samples at $34 \ h$ and $42 \ h$ p.i. the bimodal features have nearly vanished and are only vaguely perceivable. Again, the model reference matches the measurement data qualitatively. From $54 \ h$ p.i. on,
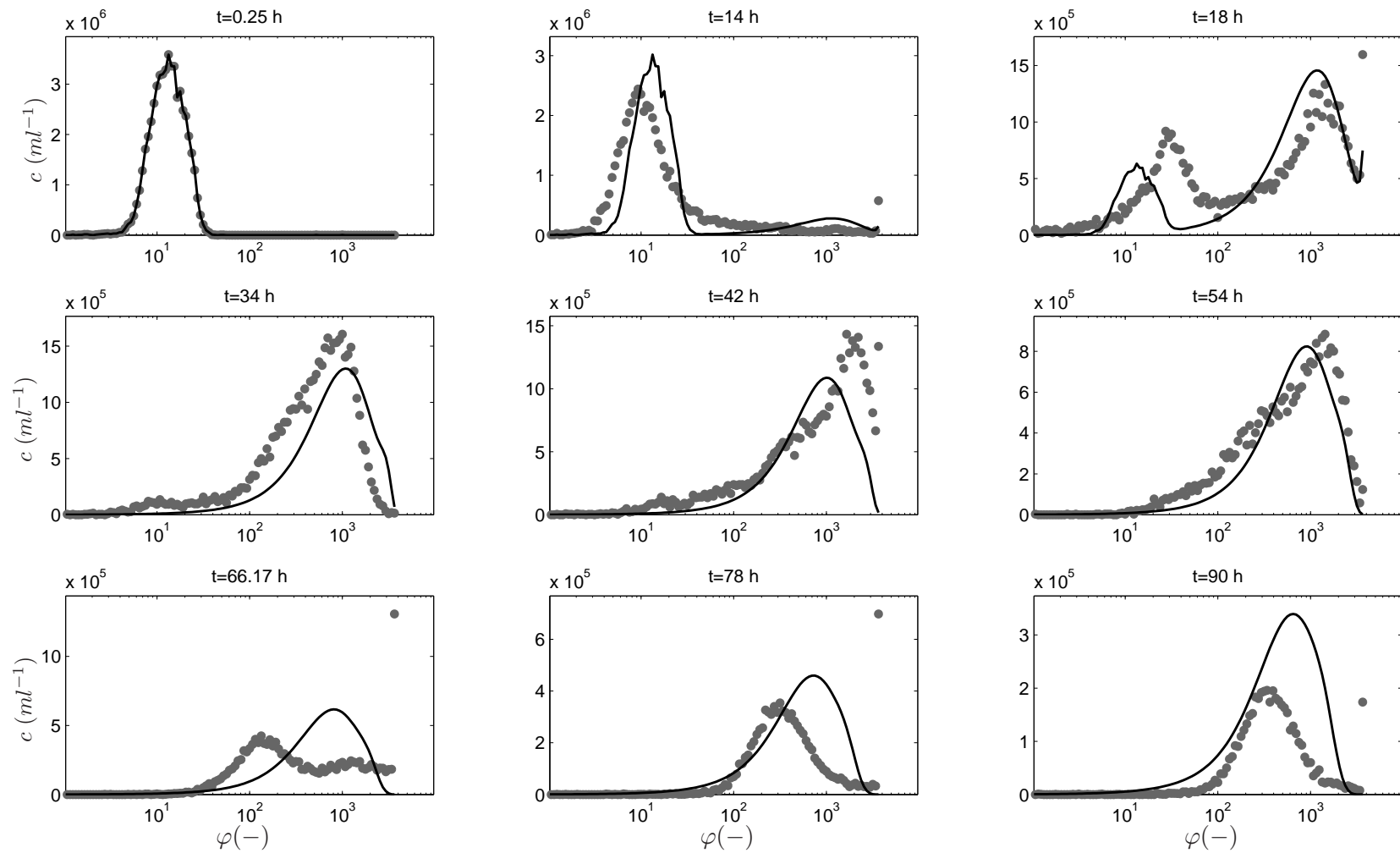
Figure 3.7.: *Influenza A/H3N2: Selected snapshots of flow cytometric data (dotted) in comparison to simulation results for two node scenario with estimated parameter set given in Table 3.1(solid curves)*

Figure 3.8.: *Influenza A/H3N2: Measurements of active and inactive virus particle concentration (dots) and simulation results for two node scenario; Estimated parameter set is given in Table 3.1 (solid curves)*

a backshift of the distribution to regions of lower degree of fluorescence can be seen. Though the backshift can be reproduced qualitatively, there is a significant discrepancy between measurement and model for later sample times.

In Fig. 3.8 measurement data of the active and inactive virus particle concentrations in the medium are shown along with the corresponding simulation results with the estimated parameter set. For the active virions, the model shows good agreement to the experimental data with the exception of the last sample at $90\ h$ p.i. Furthermore, at $26\ h$ p.i., the measurements show an unreasonable drop in the active virus particle concentration. This data point was considered to be an outlier and thus not included into the estimation procedure. In case of the inactive virus particle concentrations the model captures the data quite accurate. Note, that the detection limit for the measurement of the total virus particle concentration is given by $2\ 10^7 ml^{-1}$. Hence, concentrations below that value can not be detected. This explains the three samples at 6, 10 and $14\ h$ p.i. which hit the lower limit. The initial value of the inactive virus concentration was computed from the difference between the initial value of total virus concentration and the initial concentration of active virions measured by $TCID_{50}$.

In the next step, the number of nodes for the description of the unknown functional parameters was increased to three. In result, each functional parameter is now represented by a Hermite spline on two equidistant intervals between the three nodes. Thus, the functional dependency is not linear anymore. The remainder of the overall parameter estimation setup was kept as presented above. As initial values for the constant parameters, the resulting estimates which were obtained within the two node scenario were used. For the functional parameters, initial values for the nodes were interpolated from the two node scenario. As in the previous case, the estimation routine was repeated several times with different initial conditions.

Table 3.1.: *Estimated parameters for different degrees of discretization of functional parameters*

Parameter	Two nodes	Three nodes
$k_{\mathrm{vi}}\ (ml\ h^{-1})$	$6.337\ 10^{-7}$	$8.775\ 10^{-7}$
$k_{\mathrm{rel}}\ (h^{-1})$	597.2	589.1
$k_{\mathrm{deg}}\ (h^{-1})$	0.143	0.138
$P_{\mathrm{eff}}\ (-)$	$3.568\ 10^{-2}$	$3.451\ 10^{-2}$
$\tau_{\mathrm{lat}}\ (h)$	5	5
J	4.4947	3.7460

In Fig. 3.9 selected snapshot data of the fluorescence distributions are shown in comparison to the simulation results with the estimated parameter set. It can be seen, that the fit of the distributions has improved for snapshots at $14\ h$ p.i.. For $18\ h$ p.i. the overall deviation between model and data has decreased. However, for the next three samples an improvement in the fit is not obvious at first sight. In contrast, it can be seen easily that the model shows better agreement for later samples in comparison to the two node scenario. The corresponding simulation results for the active and inactive virus particle concentrations are depicted in Fig. 3.10. As for the cell species, the fit for the active virus particle concentrations is improved, in particular for late samples.

In Table 3.1 the estimated values for the constant parameters are summarized for both model adaption scenarios. It can be seen that similar values for the parameters are obtained. Furthermore, the value of the cost function J decreases from 4.4947 to 3.7460 showing an improved model fit. In Fig. 3.11 the estimation results for the functional parameters are shown. Similar trends are observed for both estimation setups. It can be seen that the coefficient characterizing the change of the fluorescence degree in infected cells k_{net} increases with increasing degree of fluorescence. This can be interpreted as a kind of auto-catalysis: the more viral templates are available in the cell, the more copies can be produced. In result, the virus production rate in infected cells increases until they become apoptotic. The apoptosis parameter is characterized by a very similar functional dependency. Thus, the probability of an infected cell to become apoptotic increases the more viral NP is accumulated in the cell. However, for the estimation of the coefficients for the change of fluorescence and cell death in the apoptotic cell species, different results are obtained. Using the linear approximation, the net coefficient $k_{\mathrm{net,apo}}(\varphi)$ has its highest absolute value at the upper limit of the degree of fluorescence. The more the apoptotic cells fluorescence decreases, the more does the net coefficients absolute value decrease. Hence, the loss of fluorescence intensity is slowed down. Furthermore, the cell death coefficient $k_{\mathrm{cd}}(\varphi)$ increases the more the apoptotic cells move to regions of lower fluorescence intensity. For interpretation of this behavior it has to be kept in mind that the model formulation assumes that virus release is larger than virus replication in apoptotic cells resulting in a generally negative net rate $k_{\mathrm{net,apo}}(\varphi)$. Thus the estimated results suggest that apoptotic cells are more likely to die the longer they are already apoptotic.
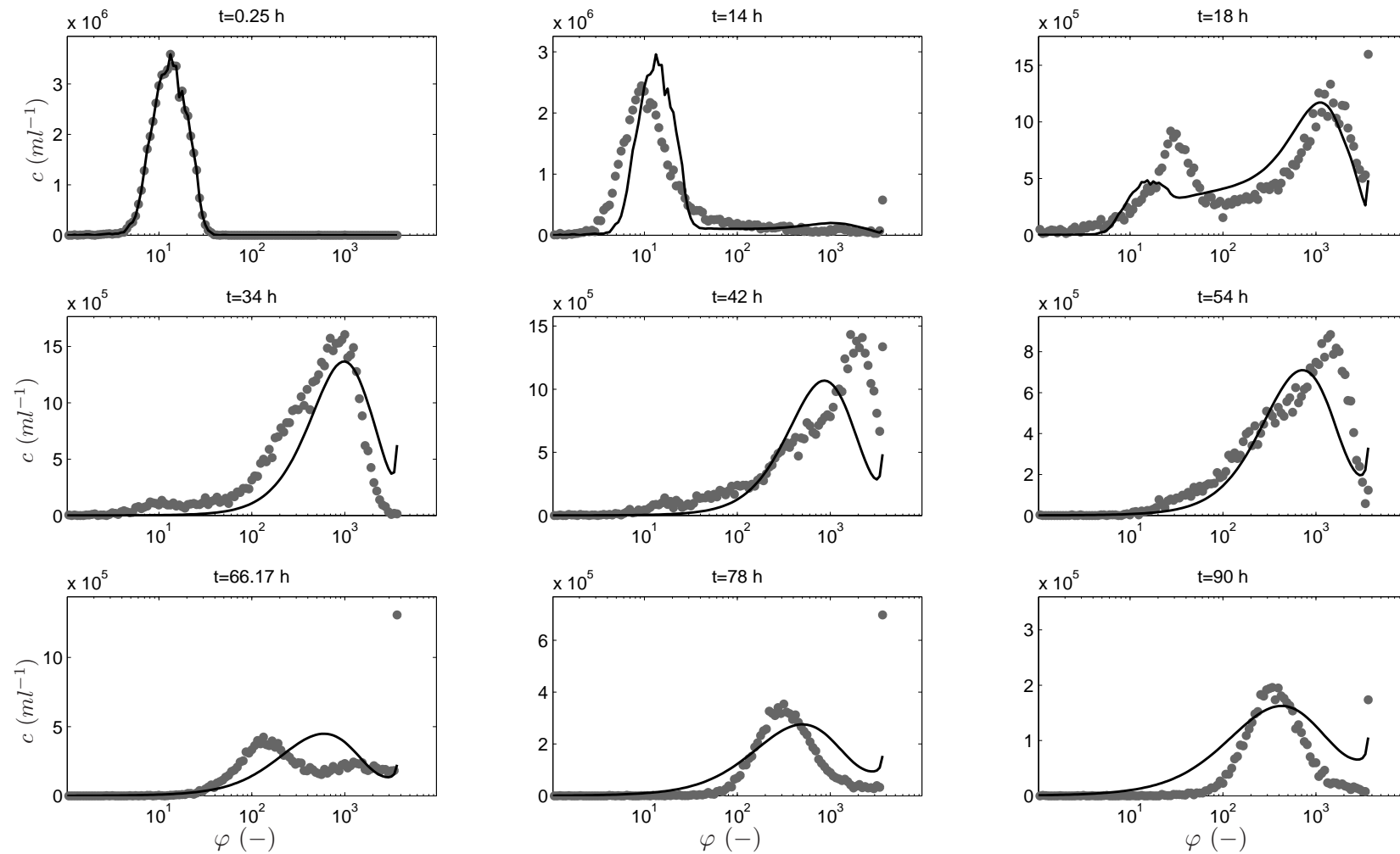
Figure 3.9.: *Selected snapshots of the overall number density distribution for simulation of model with estimated parameters for three node spline discretization; parameters are given in Table 3.1*
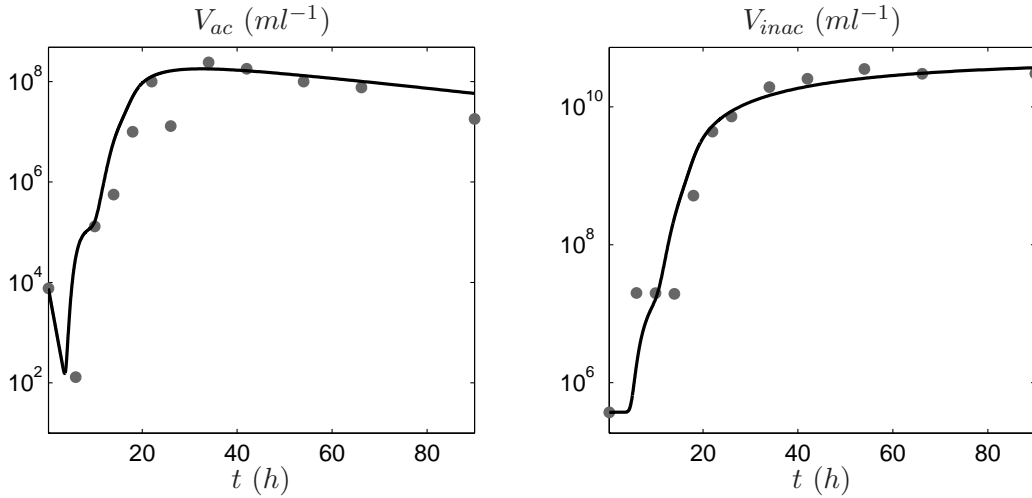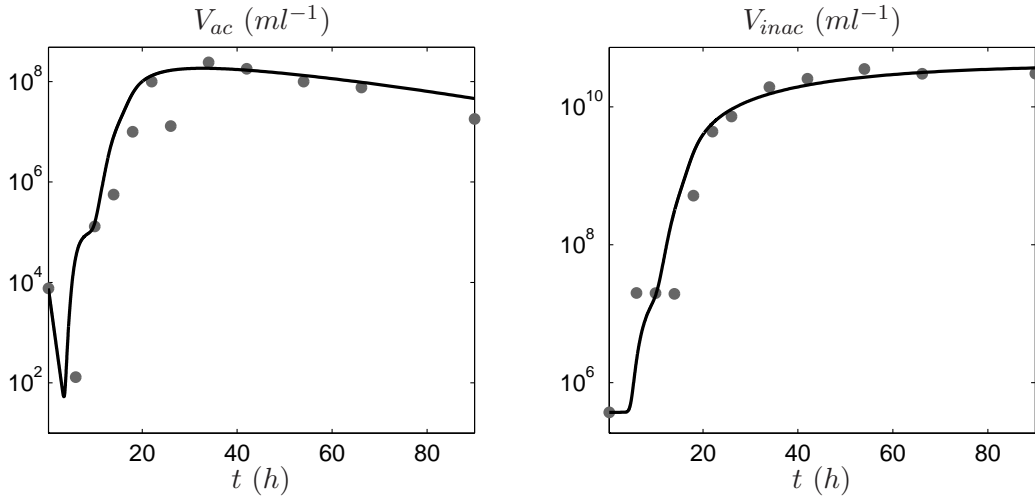
Figure 3.10.: *Influenza A/H3N2: Measurements of active and inactive virus particle concentration (dots) and simulation results for three node scenario; Estimated parameter set is given in Table 3.1 (solid curves)*

For the three node scenario the parameter estimation brings up a parabolic shape for the net coefficient of the apoptotic cells. This means, that cells with either a low or a very high intracellular amount of NP protein do only move very slowly to a region of lower φ. Furthermore, the estimated cell death coefficient shows an increased value for high degrees of fluorescence and cells with medium level of fluorescence are less likely to die than cells with a very high or very low level. One possible explanation may be that apoptotic cells which are characterized by a very high degree of fluorescence tend to die instead of staying apoptotic and keeping to release more virus progeny into the medium.

3.6.2. Variable latent phase times

The described two step parameter estimation revealed, that the estimation of the latent phase length parameter τ_{lat} differs only marginally from the initial estimate. This indicates, that the initial choice of the latent phase length fixes the other parameters. For this reason, the effect of different initial values on the overall parameter estimation was analyzed. For the two node scenario the overall estimation procedure was repeated for different values of τ_{lat}.

In Fig. 3.12 and Fig. 3.13 the corresponding simulation results are depicted. In case of the fluorescence distributions, the model predictions for different latent phase times τ_{lat} show only a slight difference, indicating, that the distributed measurements can be reproduced with a similar accuracy, independent of the choice of the latent phase length. This is due to the fact that no measured fluorescence distributions are available within the considered range of τ_{lat}. However, for the viral dynamics the model predictions show a larger variance. The largest differences are found for early samples of the active virus
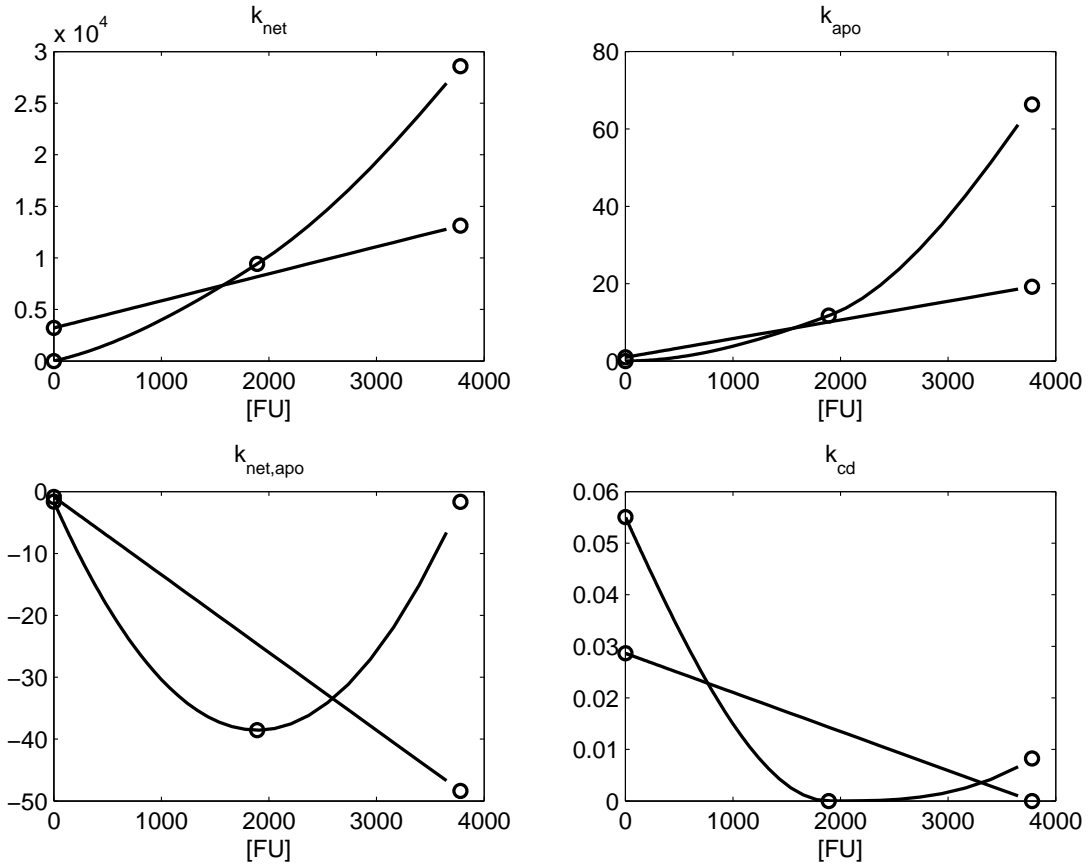
Figure 3.11.: *Influenza A/H3N2: Functional parameters characterizing change of fluorescence level and apoptosis of infected cells as well as change of fluorescence level and cell death of apoptotic cells for two and three node discretization*

concentrations. It is noticeable, that the strong decrease shortly after infection is not reproduced sufficiently using low values for the latent phase length ($\tau_{\mathrm{lat}} = 2, 3, 4\ h$). For later sampling times ($t > 30\ h.$p.i.) the model predictions differ only marginally.

The resulting corresponding parameter estimates for the constant parameters are summarized in Tab. 3.2. The corresponding final values of the cost function J show similar values for most cases, excluding $\tau_{\mathrm{lat}} = 7\ h$ where the value is larger. However, in general, the prior statement, which suggests, that the fits are similar independently from the value of the latent phase length, is confirmed. It can be observed that the estimated virus release rate does not change significantly for increasing latent phase time. Also the degradation rate coefficient k_{deg} stays within a narrow range. Solely for $\tau_{\mathrm{lat}} = 7\ h$ an increased value is estimated. In general, it can be seen, that the effective virus release P_{eff} increases with increasing latent phase length. However, the values remain in the same order of magnitude. In contrast, for the infection rate coefficient k_{vi} a more significant increase is observed which ranges over three decades.
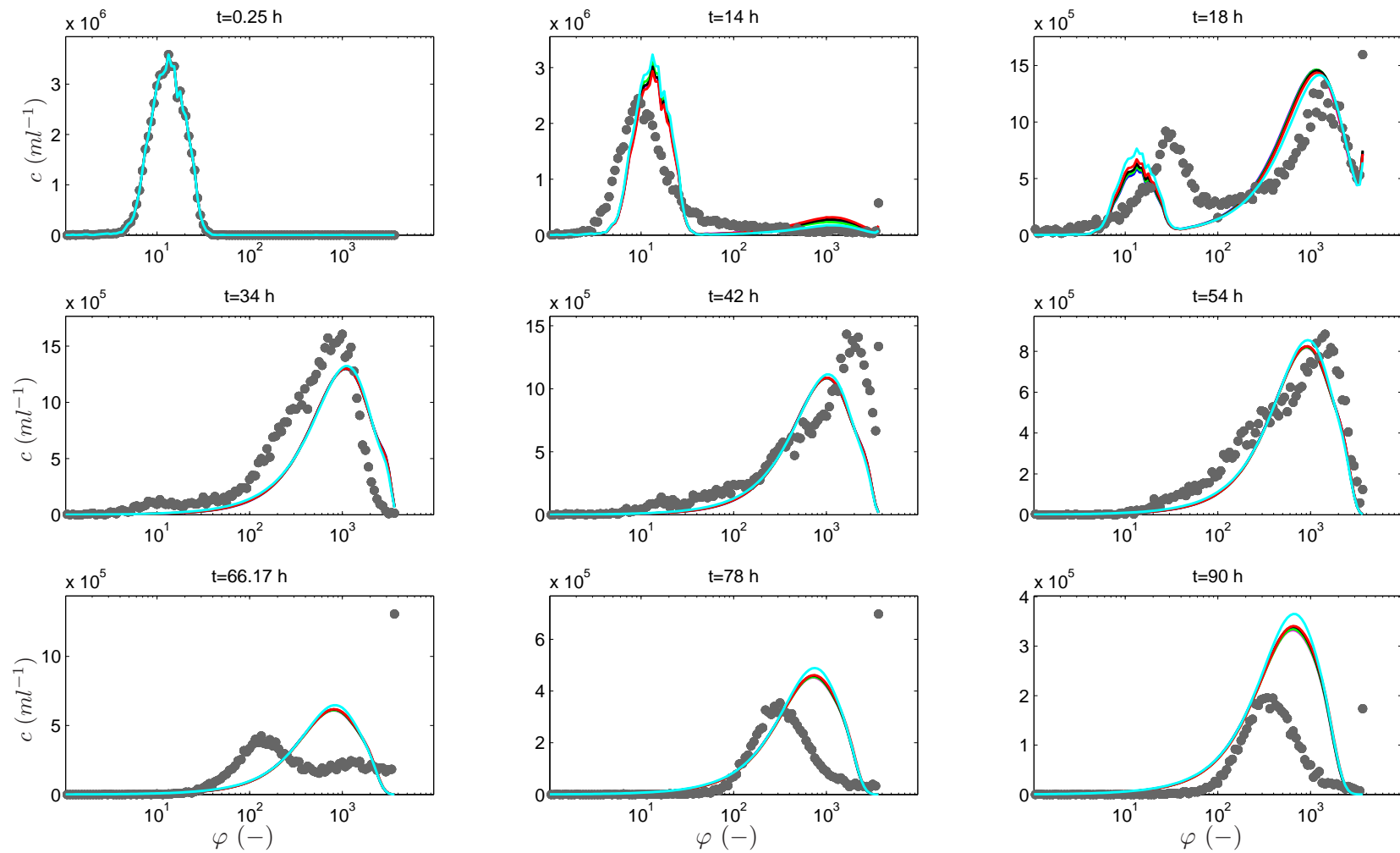
Figure 3.12.: *Simulated distributions of estimated parameter sets (see Tab. 3.2) for different values of τ_{lat}=2, 3, 4, 5, 6, 7 h (magenta/blue/green/black/red/cyan)*

Figure 3.13.: *Simulated virus concentrations of estimated parameter sets (see Tab. 3.2) for different values of $\tau_{\text{lat}}=2, 3, 4, 5, 6, 7$ h (magenta/blue/green/black-/red/cyan)*

Table 3.2.: *Estimation results for constant parameter vector $\mathbf{p}_{\text{const}}$ for different values of the latent phase length τ_{lat}*

τ_{lat} (h)	k_{vi} (ml h^{-1})	k_{rel} (h^{-1})	k_{deg} (h^{-1})	P_{eff} ($-$)	J
2	$5.314 \cdot 10^{-8}$	595.6	0.119	$2.956 \cdot 10^{-2}$	4.4901
3	$1.125 \cdot 10^{-7}$	594.2	0.115	$2.845 \cdot 10^{-2}$	4.5159
4	$2.376 \cdot 10^{-7}$	598.6	0.121	$3.031 \cdot 10^{-2}$	4.5752
5	$6.337 \cdot 10^{-7}$	597.2	0.143	$3.568 \cdot 10^{-2}$	4.4947
6	$8.874 \cdot 10^{-6}$	595.7	0.162	$4.069 \cdot 10^{-2}$	4.4402
7	$2.518 \cdot 10^{-5}$	592.0	0.363	$8.619 \cdot 10^{-2}$	4.8940
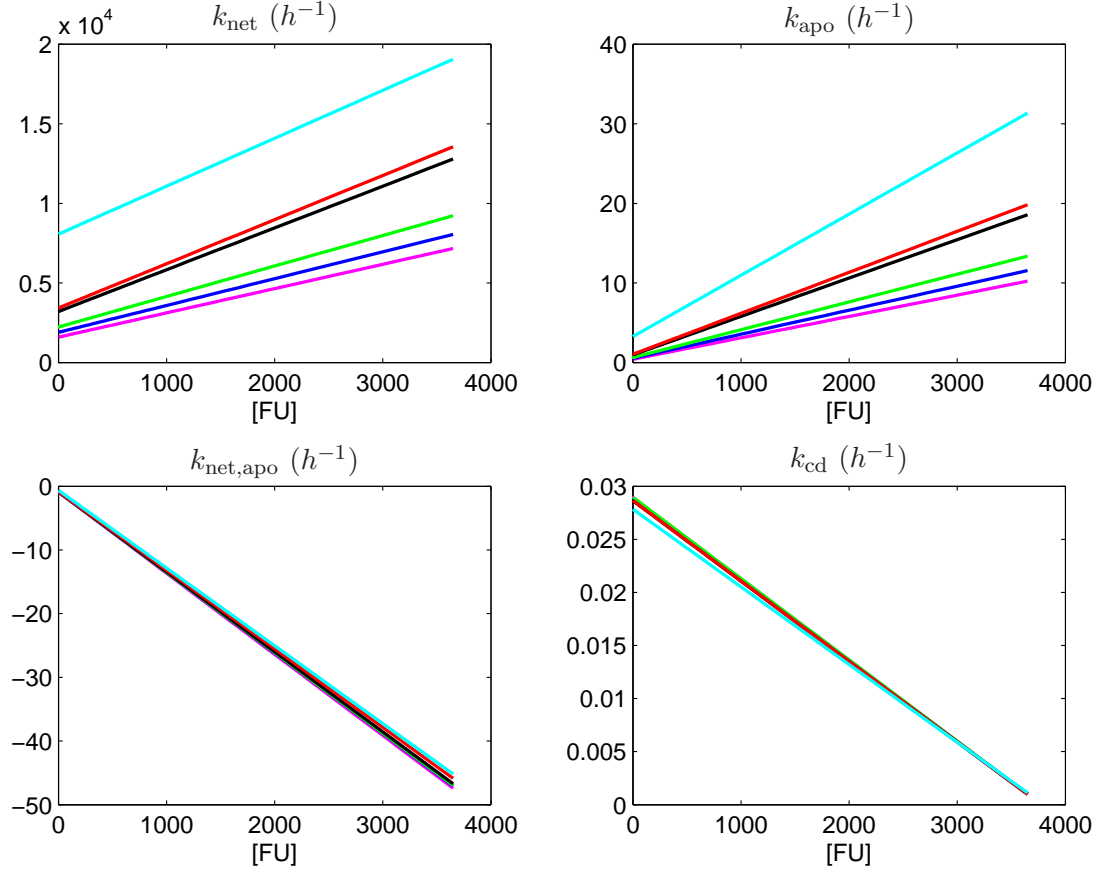
Figure 3.14.: *Functional parameters for different values of the latent phase length*
τ_{lat}=2, 3, 4, 5, 6, 7 h (magenta/blue/green/black/red/cyan)

In Fig. 3.14 the functional parameters are shown. It can be seen that estimates of the rate coefficients characterizing the change of fluorescence and the cell lysis of apoptotic cells, $k_{\mathrm{net,apo}}$ and k_{cd} are robust against changes of the latent phase length. In contrast, a clear trend is observed for the estimated rate coefficients characterizing the change of fluorescence in infected cells k_{net} and the apoptosis rate of infected cells k_{apo}. Though, the general functional dependency, i.e. the coefficients increase with increasing degree of fluorescence, is preserved, for increasing latent phase length the values of the node parameters increase.

The parameter estimation results can be explained as follows: the longer the newly infected cells remain in the latent phase, the later infected cells start to replicate viral compounds and thus the release of new virus is delayed. Thereby the overall spreading of the virus within the cell population is delayed. On the other hand, a larger virus infection rate means, that a larger amount of cells is infected resulting in a larger amount of newly replicated virions at an early stage of the infection. Thereby, the overall spreading of the virus within the cell population is accelerated. For this reason, both effects counteract each other in their effect on the overall virus release: an increase of τ_{lat} is compensated by a larger infection rate coefficient. In the same manner it can be

reasoned, that for a prolonged latent phase, the shift in the fluorescence distributions to larger degrees of fluorescence is delayed. This effect is obviously compensated by a larger net coefficient k_{net} and a larger apoptosis rate.

3.7. Discussion

This chapter was concerned with parameter estimation for a population balance model describing a vaccine production process as presented in full detail in [81] which was developed to account for cell-to-cell variability as observed by flow cytometric measurements of intracellular viral NP. The model distinguishes between four distributed cellular species, namely uninfected, latent phase, infected and apoptotic cells, and two viral species. Besides unknown constant model parameters, the model contains functional parameters which have been approximated by Hermite splines to come up with a finite dimensional inverse problem. The resulting distributed model is able to reproduce the observed dynamic phenomena, including the characteristic decrease of the viral concentrations at early stages as well as bimodality and backshift of the fluorescence distributions. Furthermore, it was shown that the model fit can be improved by increasing the complexity of the Hermite spline approximations. Although not every trait of the data is fully covered, the developed model represents a suitable tool for further analysis of the vaccine production process. Following [81], an important question is, which screws are to turn to improve the overall process. From a practical point of view, an increase of the maximum virus yield is most important which is significantly influenced by the virus release rate, as well as the length of the productive time of a host cell. Though not shown here, the general model structure applies to other human influenza A virus strains (see [81, 79]) and is expected to be adaptable to other production processes which show a similar dynamic behaviour (e.g. production of vaccinia virus).

To improve the model fit, more complex approximations of the functional parameters can be implemented. In result, the degree of freedom within the corresponding inverse problem formulation is increased (e.g. by further increasing the number of representative Hermite spline nodes). Although an increasing number of free model parameters is expected to result in an improved model fit, the numerical effort of the corresponding model parameter estimation setup is increased by means of complexity of the optimization problem. Furthermore, a larger set of free model parameters is likely to worsen condition of the inverse problem which requires more advanced estimation strategies. Thus, a good trade-off has to be found between the improved fit of data and model on the one hand and an increasing complexity of the estimation procedure. However, interpretation of the resulting functional parameter estimates becomes even more challenging in view of the underlying model assumptions and simplifications.

Furthermore, the results from Section 3.6.2 emphasize the limited biological interpretability of the parameter estimates. Here, the initial choice of the latent phase length τ_{lat} has significant effects on estimates of other model parameters while the corresponding values of the cost function are very similar. It was shown, that larger values of the

latent phase length are compensated basically by larger values for the infection rate coefficient k_{vi}, the rate coefficients in infected cells $k_{\mathrm{net}}(\varphi)$ and the apoptosis rate $k_{\mathrm{apo}}(\varphi)$. These results indicate, that the latent phase length can only be roughly identified from the given data. This situation may be improved using additional measurements, e.g. flow cytometric measurements at early stages of the process.

In addition to the problems listed above, there is a more basic issue restraining a detailed biological interpretation of the model results. In fact, the whole intracellular kinetics including the virus entry, replication of viral compounds, formation of protein complexes and release of virus progeny but also the complete host cell metabolism is lumped to only one internal coordinate, namely φ, and the length of the latent phase τ_{lat}. To improve the situation, more information on the intracellular state, like cell cycle state, or on further intracellular compounds like other viral proteins has to be taken into account to allow a more detailed characterization of the virus host cell interaction. In principle, these intracellular compounds can also be measured with techniques like flow cytometry (e.g. matrix protein M1, see [102]).

The most straightforward extension of the current model formulation is to develop an unstructured but segregated model formulation which categorizes cells with respect to the degree of infection in terms of a corresponding fluorescence φ_{NP} and degree of apoptosis in terms of a second fluorescence φ_{APO} (see [79] for a more detailed discussion). The latter can be quantified by staining of a suitable apoptosis signal. In consequence a two dimensional population balance equation would be obtained. Here, the net coefficients describing the change of the distributions during the process are two dimensional functional parameters $k_{\circ}(\varphi_{\mathrm{NP}}, \varphi_{\mathrm{APO}})$. To adapt the model to the data, the previously described solution approach for the inverse problem has to be extended. The generally unknown functional parameters can be represented by a sum of weighted basis functions (see e.g. [64]) thereby rendering the infinite dimensional inverse problem to a finite dimension. Yet, the numerical solution of the overall system of coupled partial and ordinary differential equations increases the computational effort by a large factor and a biological meaningful interpretation of the resulting coefficients may be difficult. Even if the these obstacles could be overcome for a two dimensional model formulation, further extension to a three or four dimensional model formulation may be necessary. Here, the numerical effort for model solution and adaption may become unreasonable large.

As an alternative to this top-down modeling approach a bottom-up modeling approach can be applied leading to multi dimensional structured PBEs of the process. These model formulations are based on the detailed description of the dynamics on the single cell level which are derived from single cell experiments. For the influenza virus replication in mammalian cells such a detailed model was developed by Heldt et al. [41]. In Fig. 3.15 the corresponding intracellular reaction scheme is depicted. Even without the knowledge of further details about the modeled compounds and their kinetics, it can be easily seen, that the single cell model is characterized by complex interactions of a high number of intracellular compounds. The corresponding population balance model represents a high dimensional PDE as each state of a single cell model directly translates to an internal coordinate. As already mentioned in the first chapter the numerical solution
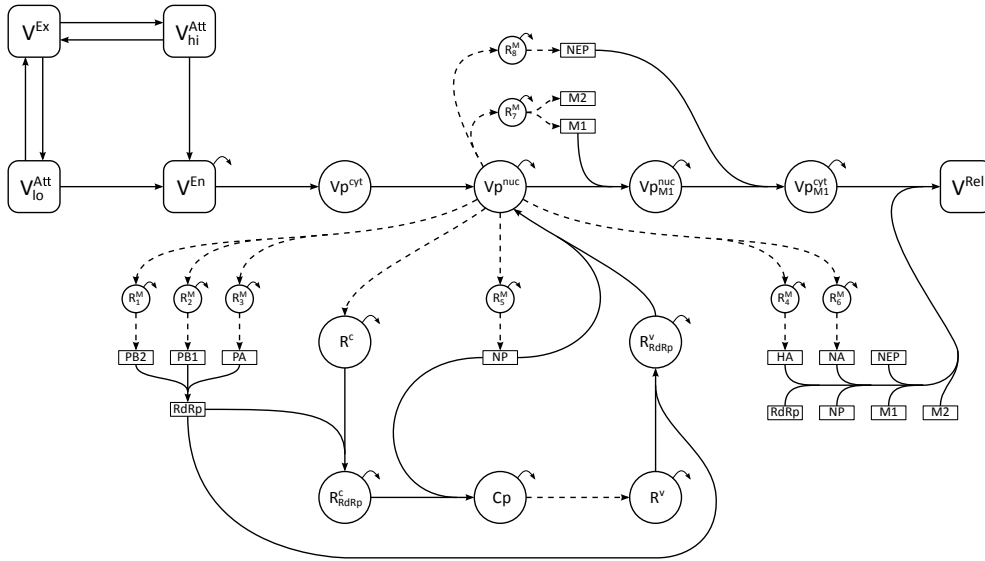
Figure 3.15.: *Schematic representation of the intracellular reactions for influenza virus replication according to [41]*

of the corresponding PBM is challenging. A first step into this direction for influenza was presented by Heldt et al. [40]: Within an age-structured model formulation it was assumed that all cells within the same time post infection have the same intracellular state. However, further cell-to-cell variability, e.g. with respect to viral RNA synthesis rates, was neglected. If cells with the same time post infection are assumed to exhibit cell-to-cell variability resulting from e.g. different gene expression levels, this approach can not be pursued. Instead, other numerical solution methods for the emerging multi dimensional PBEs have to be employed which can take into account this heterogeneity, e.g. moment methods as discussed in the second part of this thesis.

4. Efficient moment approximation for multi dimensional population balance equations resulting from bottom-up modeling

From a mathematical point of view the PBE (2.10) is a multi dimensional first order partial differential equation that is coupled to a set of ODEs characterizing the continuous phase (2.14). Analytical solutions can only be found for special cases. Thus, numerical solution techniques have to be employed. For high dimensional PBEs the application of discretization based solution techniques like the finite volume method is limited due to large computational costs. Alternatively, as introduced in Section 2.2, the PBEs can be reduced to a set of ODEs characterizing the moment dynamics. Only for special classes of problems, the dynamic moment equations are closing such that an approximate closure has to be applied in many cases. In the following, approximate moment methods will be discussed and applied to several benchmark problems.

4.1. Approximate closure for computation of moments

It was already mentioned that the moment dynamics (2.12) can be derived from integration of the general PBE in local form. In the following, the moment closure problem will be demonstrated for the simple one dimensional case with $Nuc = 0$ and $D = 0$. Here, the dynamics of arbitrary moments

$$m_l(t) = \int_0^\infty x^l n(t,x) \mathrm{d}x \tag{4.1}$$

with respect to the internal coordinate x, are given by

$$\frac{\mathrm{d}m_l(t)}{\mathrm{d}t} = -\int_0^\infty x^l \frac{\partial}{\partial x} \left\{ G(t,x)n(t,x) \right\} \mathrm{d}x \,. \tag{4.2}$$

Applying integration by parts, for a constant growth velocity $G = const.$, one obtains

$$\frac{\mathrm{d}m_l(t)}{\mathrm{d}t} = [G(t,x)n(t,x)]_0^\infty + l \, G \int_0^\infty x^{l-1} n(t,x) \, \mathrm{d}x = l \, G \, m_{l-1}(t) \,. \tag{4.3}$$

In the following, it will be assumed that there are no fluxes over the boundaries

$$[G(t,x)n(t,x)]_0^\infty = 0\,. \tag{4.4}$$

Thus the first integer moments are given by

$$\frac{\mathrm{d}m_0(t)}{\mathrm{d}t} = 0\,, \quad \frac{\mathrm{d}m_1(t)}{\mathrm{d}t} = G\,m_0(t)\,, \quad \frac{\mathrm{d}m_2(t)}{\mathrm{d}t} = 2\,G\,m_1(t)\,, \quad \ldots \tag{4.5}$$

In this case the moment dynamics are said to be closed as there can always be found a closed set of equations for the computation of the moment dynamics. Here, moment dynamics of arbitrary degree can always be described by functions of the lower order moments.

In contrast, if the growth velocity is a monomial function of the internal coordinate $G(t,x) = x^p$ one obtains the following moment dynamics

$$\frac{\mathrm{d}m_l(t)}{\mathrm{d}t} = l \int\limits_0^\infty x^{p+l-1} n(t,x)\,\mathrm{d}x = l\,m_{p+l-1}(t)\,. \tag{4.6}$$

Here, the moment dynamics are closed for $p = 0, 1$ but not for $p \geq 2$

$$\frac{\mathrm{d}m_0(t)}{\mathrm{d}t} = 0\,, \quad \frac{\mathrm{d}m_1(t)}{\mathrm{d}t} = m_p(t)\,, \quad \frac{\mathrm{d}m_2(t)}{\mathrm{d}t} = 2\,m_{p+1}(t)\,, \quad \ldots \quad \frac{\mathrm{d}m_l(t)}{\mathrm{d}t} = l\,m_{\hat{l}}(t)\,. \tag{4.7}$$

as the dynamics of an arbitrary moment m_l depend on a higher order moment $m_{\hat{l}}(t) = m_{p+l-1}(t)$. To compute the dynamics nevertheless, these higher order moments could be replaced by a suitable approximation in terms of lower order moments

$$m_{\hat{l}}(t) \approx A\{m_0(t),\,m_1(t),\ldots,\,m_l(t)\}\,, \tag{4.8}$$

resulting in the approximate moment dynamics

$$\frac{\mathrm{d}m_0(t)}{\mathrm{d}t} = 0\,, \quad \frac{\mathrm{d}m_1(t)}{\mathrm{d}t} = m_p(t)\,, \quad \frac{\mathrm{d}m_2(t)}{\mathrm{d}t} = 2\,m_{p+1}(t)\,, \quad \ldots$$
$$\frac{\mathrm{d}m_l(t)}{\mathrm{d}t} = l\,A\{m_0(t),\ldots,m_l(t)\}\,. \tag{4.9}$$

For a more general case, the approximate closure can be obtained by replacing all higher order moments within the moment dynamics system.

$$\begin{bmatrix} \frac{\mathrm{d}m_0}{\mathrm{d}t} \\ \frac{\mathrm{d}m_1}{\mathrm{d}t} \\ \vdots \\ \frac{\mathrm{d}m_l}{\mathrm{d}t} \end{bmatrix} = \begin{bmatrix} f_0\,(m_0,\,m_1,\ldots,\,m_{\hat{l}}) \\ f_1\,(m_0,\,m_1,\ldots,\,m_{\hat{l}}) \\ \vdots \\ f_l\,(m_0,\,m_1,\ldots,\,m_{\hat{l}}) \end{bmatrix} \approx \begin{bmatrix} f_0\,(m_0,\ldots,\,A\{m_0,\,m_1,\ldots,\,m_l\}) \\ f_1\,(m_0,\ldots,\,A\{m_0,\,m_1,\ldots,\,m_l\}) \\ \vdots \\ f_l\,(m_0,\ldots,\,A\{m_0,\,m_1,\ldots,\,m_l\}) \end{bmatrix}\,. \tag{4.10}$$

One possible way to find an approximation A is to use a weighted sum of abscissas

$$m_{\hat{l}}(t) \approx A\{m_0(t),\,m_1(t),\ldots,\,m_l(t)\} \approx \sum_{\alpha=1}^{N_\alpha} w_\alpha(t) x_\alpha^{\hat{l}}(t)\,. \tag{4.11}$$

Here, $x_\alpha(t)$ and $w_\alpha(t)$ are the abscissas and corresponding weights and N_α is the overall number of abscissas. At this point it has to be mentioned that this approximation is special case of the more general quadrature

$$\int_0^\infty n(t,x)f(t,x)\mathrm{d}x \approx \sum_{\alpha=1}^{N_\alpha} w_\alpha(t)f_\alpha(t,x_\alpha)\,, \qquad (4.12)$$

with $f_\alpha(t,x_\alpha) = x_\alpha^{\hat{l}}$. As presented in [73] the core idea of this quadrature based closure is that the weights and abscissas can be determined in terms of lower order moments

$$\mu_l = \sum_{k=0}^{N_\alpha} w_\alpha(t)x_\alpha^l(t)\,, \quad k = 0,\dots,2N_\alpha - 1\,. \qquad (4.13)$$

Thus values for the weights and abscissas have can be computed from the following system of nonlinear equations

$$m_0(t) = \sum_{\alpha=1}^{N_\alpha} w_\alpha(t)\,,$$

$$m_1(t) = \sum_{\alpha=1}^{N_\alpha} w_\alpha(t)x_\alpha(t)\,,$$

$$\vdots$$

$$m_{2N_\alpha-1}(t) = \sum_{\alpha=1}^{N_\alpha} w_\alpha(t)x_\alpha^{2N_\alpha-1}(t)\,. \qquad (4.14)$$

To obtain a numerical solution of the overall approximate moment dynamics a differential algebraic equation (DAE) system consisting of (4.10) and (4.14) has to be solved [33, 32]. Alternatively, the problem can be solved sequentially. The procedure is given by the following steps:

1. Calculate the initial moments from the given initial distributions

2. Solve the system of nonlinear equations (4.14) via algorithms like the Product Difference Algorithm [73] or the Wheeler-Algorithm [117] for given moments to obtain abscissas and weights

3. Approximate the right hand side of (4.10)

4. Solve (4.10) to obtain moments at next time step

5. Repeat steps 2. - 5. until t_{end}

This stepwise algorithm is also known as the quadrature method of moments (QMOM) [73]. A detailed discussion on convergence properties can be found f.e. in [77].

This formalism can be extended to the more general multi dimensional case using the

following approximation

$$m_{l_1,\ldots,l_{N_d}}(t) \approx \sum_{\alpha=1}^{N_\alpha} w_\alpha(t) x_{1,\alpha}^{l_1}(t) \ldots x_{N_d,\alpha}^{l_{N_d}}(t)\,. \tag{4.15}$$

The previously presented procedure for the approximate solution of the moment dynamics remains the same. However, the solution of the according nonlinear system of equations (4.14) is critical as there is no generalized solution concept as the Product Difference Algorithm for higher dimensional problems. In fact, it is not clear which moments are necessary to obtain a reasonable solution for a given number of weights and abscissas. Alternatively, the direct quadrature method of moments can be applied which can easily be extended to multi dimensional problems.

4.2. Moment approximation by direct quadrature method of moments

The main idea of the direct quadrature method of moments (DQMOM) is to track abscissas $\mathbf{x}_\alpha$ and weights w_α directly, instead of tracking moments as in the QMOM. Thus (4.14) has to be solved only once for the initial step but the repetitive solution would be omitted. The corresponding dynamic equations to characterize the temporal evolution of abscissas and weights are obtained from the PBE formulation and will be presented in the following.

4.2.1. Classical derivation for one dimensional PBEs

DQMOM was first introduced by Marchisio and Fox [70]. For reasons of simplicity the one dimensional case is considered first

$$\frac{\partial n(t,x)}{\partial t} = S(t,x) \tag{4.16}$$

where the right hand side comprises terms describing particle growth, nucleation and death processes. Basically, it is assumed that the number density distribution can be approximated by a weighted sum of delta distributions

$$n(t,x) = \sum_{\alpha=1}^{N_\alpha} w_\alpha(t)\delta(x - x_\alpha(t))\,. \tag{4.17}$$

The delta distribution is defined as

$$\delta(y) = \begin{cases} 0 & , \quad y \neq 0 \\ \infty & , \quad y = 0 \end{cases}, \tag{4.18}$$

with

$$\int_{-\infty}^{\infty} \delta(y)\mathrm{d}y = 1\,. \tag{4.19}$$

Thus (4.16) can be written as

$$\frac{\partial n(t,x)}{\partial t} = \sum_{\alpha=1}^{N_\alpha} \left[\frac{\partial w_\alpha(t)}{\partial t}\delta(x - x_\alpha(t)) \right.$$
$$\left. - \left(\frac{\partial}{\partial t}\{w_\alpha(t)x_\alpha(t)\} - x_\alpha(t)\frac{\partial w_\alpha(t)}{\partial t} \right) \delta'(x - x_\alpha(t)) \right] = S(t,x)\,. \tag{4.20}$$

In course of the process, the number density distribution $n(t,x)$ undergoes changes and thus the weights and abscissas change over time, accordingly. As already mentioned, the main idea of the DQMOM is the direct tracking of weights and abscissas. Introducing the following abbreviations for the dynamics of weights and weighted abscissas

$$\frac{\partial w_\alpha(t)}{\partial t} = a_\alpha, \qquad\qquad \frac{\partial}{\partial t}\{w_\alpha(t)x_\alpha(t)\} = b_\alpha\,, \tag{4.21}$$

the equation simplifies to

$$\sum_{\alpha=1}^{N_\alpha} \left[\left(\delta(x - x_\alpha(t)) + x_\alpha(t)\delta'(x - x_\alpha(t)) \right) a_\alpha - \delta'(x - x_\alpha(t))b_\alpha \right] = S(t,x) \,. \qquad (4.22)$$

Now the moment transform (2.11) can be applied resulting in

$$\int_0^\infty x^l \sum_{\alpha=1}^{N_\alpha} \left[\left(\delta(x - x_\alpha(t)) + x_\alpha(t)\delta'(x - x_\alpha(t)) \right) a_\alpha - \delta'(x - x_\alpha(t))b_\alpha \right] \mathrm{d}x$$

$$= \int_0^\infty x^l \, S(t,x)\mathrm{d}x \,. \qquad (4.23)$$

Exploiting the following features of the delta distribution

$$\int_{-\infty}^\infty x^l \delta(x - x_\alpha(t))\mathrm{d}x = x_\alpha^l(t) \,,$$

$$\int_{-\infty}^\infty x^l \delta'(x - x_\alpha(t))\mathrm{d}x = -l \, x_\alpha^{l-1}(t) \,, \qquad (4.24)$$

yields

$$\sum_{\alpha=1}^{N_\alpha} \left[(1 - l)x_\alpha^l(t)a_\alpha + lx_\alpha^{l-1}(t)b_\alpha \right] = \int_0^\infty x^l S(t,x)\mathrm{d}x \approx \widetilde{S}^l \,, \qquad (4.25)$$

where the integral on the right hand side can also be written in terms of a delta distribution. The equation can be derived for arbitrary moments $m_l(t) \,,\ l = 0, 1, \ldots, N_l$. In consequence, the following system of linear equations emerges

$$\begin{bmatrix} 1 & \ldots & 1 & 0 & \ldots & 0 \\ 0 & \ldots & 0 & 1 & \ldots & 1 \\ & \vdots & & & \vdots & \\ (1-l)x_1^l & \ldots & (1-l)x_{N_\alpha}^l & lx_1^{l-1} & \ldots & lx_{N_\alpha}^{l-1} \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ \vdots \\ a_{N_\alpha} \\ b_1 \\ \vdots \\ b_{N_\alpha} \end{bmatrix} = \begin{bmatrix} \widetilde{S}^0 \\ \vdots \\ \widetilde{S}^l \end{bmatrix} \,. \qquad (4.26)$$

The solution gives the required values a_α and b_α to be used in (4.21) for the dynamics of $w_\alpha(t)$ and $x_\alpha(t)$.

4.2.2. Alternative derivation omitting delta distributions

Alternatively, the previously introduced system of linear equations can be derived without the explicit use of delta distributions, as presented in [111]. This is done by directly

combining the approximation formula (4.11) with the moment dynamics (2.12)

$$
\int_0^\infty x^l S(t,x) \, \mathrm{d}x = \frac{\mathrm{d}m_l(t)}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t} \int_0^\infty x^l n(t,x) \, \mathrm{d}x \approx \frac{\mathrm{d}}{\mathrm{d}t} \sum_{\alpha=1}^{N_\alpha} x_\alpha^l(t) w_\alpha(t)
$$

$$
= \sum_{\alpha=1}^{N_\alpha} \left[x_\alpha^l(t) \frac{\mathrm{d}w_\alpha(t)}{\mathrm{d}t} + l \; w_\alpha(t) x_\alpha^{l-1}(t) \frac{\mathrm{d}x_\alpha(t)}{\mathrm{d}t} \right]
$$

$$
= \sum_{\alpha=1}^{N_\alpha} \left[x_\alpha^l(t) \frac{\mathrm{d}w_\alpha(t)}{\mathrm{d}t} + l \; x_\alpha^{l-1}(t) \left(\frac{\mathrm{d}w_\alpha(t)x_\alpha(t)}{\mathrm{d}t} - x_\alpha(t) \frac{\mathrm{d}w_\alpha(t)}{\mathrm{d}t} \right) \right]
$$

$$
= \sum_{\alpha=1}^{N_\alpha} \left[(1-l)x_\alpha^l(t) \frac{\mathrm{d}w_\alpha(t)}{\mathrm{d}t} + l \; x_\alpha^{l-1}(t) \frac{\mathrm{d}w_\alpha(t)x_\alpha(t)}{\mathrm{d}t} \right] . \tag{4.27}
$$

Introducing the same abbreviations for the dynamics of weights and weighted abscissas (4.21) as before yields (4.25).

4.2.3. Solution of the system of linear equations

In contrast to QMOM, the DQMOM does not involve the solution of a nonlinear algebraic equation system (4.14). Instead, a system of linear algebraic equations (4.26) has to be solved.

$$
A_{\mathrm{OV}} \cdot \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} = \widetilde{\mathbf{S}} . \tag{4.28}
$$

This can be done numerically by applying the inverse

$$
A_{\mathrm{OV}}^{-1} \widetilde{\mathbf{S}} = \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} \tag{4.29}
$$

where $A_{\mathrm{OV}} \in \mathbb{R}^{2N_\alpha \times 2N_\alpha}$ has to be regular for an unique solution. Alternatively, if the size of $\widetilde{\mathbf{S}}$ is larger than the number of unknowns (which means that $A_{\mathrm{OV}} \in \mathbb{R}^{N_k \times 2N_\alpha}$ is not quadratic anymore), an approximate solution can be computed using the Moore-Penrose-Pseudoinverse (see e.g. [90])

$$
\left[A_{\mathrm{OV}}^{\mathrm{T}} A_{\mathrm{OV}} \right]^{-1} A_{\mathrm{OV}}^{\mathrm{T}} \cdot \widetilde{\mathbf{S}} = A_{\mathrm{OV}}^* \cdot \widetilde{\mathbf{S}} = \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} . \tag{4.30}
$$

However, in some cases the unknowns can be computed analytically. For this reason, the right hand side is subdivided into parts characterizing the effects of growth, death and nucleation

$$
\widetilde{\mathbf{S}} = \widetilde{\mathbf{S}}_{\mathrm{G}} + \widetilde{\mathbf{S}}_{\mathrm{D}} + \widetilde{\mathbf{S}}_{\mathrm{N}}. \tag{4.31}
$$

In the same way the unknowns $\mathbf{a}$ and $\mathbf{b}$ are divided into corresponding solution parts. Thus, the solution of the overall system

$$
A_{\mathrm{OV}} \cdot \left(\begin{bmatrix} \mathbf{a}_{\mathrm{G}} \\ \mathbf{b}_{\mathrm{G}} \end{bmatrix} + \begin{bmatrix} \mathbf{a}_{\mathrm{D}} \\ \mathbf{b}_{\mathrm{D}} \end{bmatrix} + \begin{bmatrix} \mathbf{a}_{\mathrm{N}} \\ \mathbf{b}_{\mathrm{N}} \end{bmatrix} \right) = \widetilde{\mathbf{S}}_{\mathrm{G}} + \widetilde{\mathbf{S}}_{\mathrm{D}} + \widetilde{\mathbf{S}}_{\mathrm{N}} \tag{4.32}
$$

can be obtained by linear combination of solutions of the subproblems

$$A_{\mathrm{OV}} \cdot \begin{bmatrix} \mathbf{a}_{\mathrm{G}} \\ \mathbf{b}_{\mathrm{G}} \end{bmatrix} = \widetilde{\mathbf{S}}_{\mathrm{G}} \,, \quad A_{\mathrm{OV}} \cdot \begin{bmatrix} \mathbf{a}_{\mathrm{D}} \\ \mathbf{b}_{\mathrm{D}} \end{bmatrix} = \widetilde{\mathbf{S}}_{\mathrm{D}} \,, \quad A_{\mathrm{OV}} \cdot \begin{bmatrix} \mathbf{a}_{\mathrm{N}} \\ \mathbf{b}_{\mathrm{N}} \end{bmatrix} = \widetilde{\mathbf{S}}_{\mathrm{N}} \,. \tag{4.33}$$

In consequence, the overall dynamics of weights and weighted abscissas are given as sum of the corresponding dynamics of each single effect.

Solution for pure particle growth processes

If the right hand side of (4.26) is also approximated by the same sum of weighted abcissas (4.11), the system of linear equations

$$\begin{bmatrix} 1 & \dots & 1 & 0 & \dots & 0 \\ 0 & \dots & 0 & 1 & \dots & 1 \\ & \vdots & & & \vdots & \\ (1-l)x_1^l & \dots & (1-l)x_{N_\alpha}^l & lx_1^{l-1} & \dots & lx_{N_\alpha}^{l-1} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{a}_{\mathrm{G}} \\ \mathbf{b}_{\mathrm{G}} \end{bmatrix} = \begin{bmatrix} 0 \\ \sum\limits_{\alpha=1}^{N_\alpha} w_\alpha G_\alpha \\ \dots \\ \sum\limits_{\alpha=1}^{N_\alpha} w_\alpha l x_\alpha^{l-1} G_\alpha \end{bmatrix} \,. \tag{4.34}$$

A corresponding solution can be computed easily from the first two rows by comparing the coefficients

$$\frac{\mathrm{d}w_\alpha(t)}{\mathrm{d}t} = 0 \,, \qquad\qquad \frac{\mathrm{d}x_\alpha(t)}{\mathrm{d}t} = G_\alpha = G(t, x_\alpha(t)) \,. \tag{4.35}$$

Apparently, the dynamics of the abscissas correspond to the characteristic system of the corresponding PDE (see Appendix A) and the weights remain constant for pure growth.

Solution for pure particle death processes

As in the case of a pure particle growth process, the moments on the right hand side may be approximated by a weighted sum of abscissaas. In consequence,

$$\widetilde{\mathbf{S}}_{\mathrm{D}} = \left[\sum_{\alpha=1}^{N_\alpha} -w_\alpha D_\alpha \,, \quad \dots \quad , \sum_{\alpha=1}^{N_\alpha} -w_\alpha x_\alpha^l D_\alpha \right]^{\mathrm{T}} \,. \tag{4.36}$$

The solution of the corresponding subproblem, which can again be obtained by comparing the coefficients, is given by

$$\frac{\mathrm{d}w_\alpha(t)}{\mathrm{d}t} = -w_\alpha(t)D_\alpha = -w_\alpha(t)D(t, x_\alpha(t)) \,, \qquad\qquad \frac{\mathrm{d}x_\alpha(t)}{\mathrm{d}t} = 0 \,. \tag{4.37}$$

This implies, that the weights increase or decrease according to the particle death kinetics but the abscissas do not change within the process.

Solution for pure nucleation processes

Unfortunately, the solution for pure nucleation problems can not be computed in a similar easy manner. In fact, the unknown dynamics of weights and abscissas $\mathbf{a}_N$ and $\mathbf{b}_N$ are solution of corresponding linear system of equations

$$A_{OV} \cdot \begin{bmatrix} \mathbf{a}_N \\ \mathbf{b}_N \end{bmatrix} = \begin{bmatrix} \int\limits_0^\infty Nuc \ \mathrm{d}x \\ \int\limits_0^\infty x \ Nuc \ \mathrm{d}x \\ \vdots \\ \int\limits_0^\infty x^l \ Nuc \ \mathrm{d}x \end{bmatrix}. \tag{4.38}$$

As already mentioned, a solution can be computed by inversion of A_{OV}. As A_{OV} resembles a *Vandermonde* matrix, it may become ill conditioned easily, in particular if the number of abscissas N_α is large. To improve the condition, the abscissas may be normalized. Moreover, the usage of fractional order moments (i.e. $l \in \mathbb{R}$) is suggested in [69]. Furthermore, if two or more abscissas take the same value, A_{OV} becomes singular. One strategy to circumvent this is to add small random perturbations to the abscissas [70].

4.2.4. Extension to the multi dimensional case

The principle idea of the DQMOM can be extended directly to multiple dimensions [70]. Here, the number density distribution is represented by multivariate delta distributions

$$n(t, \mathbf{x}) = \sum_\alpha^{N_\alpha} w_\alpha(t) \prod_{i=1}^{N_d} \delta(x_i - x_{i,\alpha}(t)) = \sum_{\alpha=1}^{N_\alpha} w_\alpha(t) \delta\left(\mathbf{x} - \mathbf{x}_\alpha(t)\right). \tag{4.39}$$

Analogous to the one dimensional case, the following equation can be derived

$$\sum_{\alpha=1}^{N_\alpha} \left\{ (1 - \sum_{i=1}^{N_d} l_i) x_{1,\alpha}^{l_1} \ldots x_{N_d,\alpha}^{l_{N_d}} a_\alpha + \begin{bmatrix} l_1 \ b_{1,\alpha} \\ \vdots \\ l_{N_d} \ b_{N_d,\alpha} \end{bmatrix}^T \cdot \begin{bmatrix} x_{1,\alpha}^{l_1-1} x_{2,\alpha}^{l_2} \ldots x_{N_d,\alpha}^{l_{N_d}} \\ \vdots \\ x_{1,\alpha}^{l_1} x_{2,\alpha}^{l_2} \ldots x_{N_d,\alpha}^{l_{N_d}-1} \end{bmatrix} \right\}$$
$$= \int_{\mathbf{X}} x_1^{l_1} \ldots x_{N_d}^{l_{N_d}} \ S \ \mathrm{d}\mathbf{x}, \tag{4.40}$$

for an arbitrary moment $m_{l_1,\ldots,l_{N_d}}$. The dynamics of weights and weighted abscissas are defined as

$$\frac{\partial w_\alpha(t)}{\partial t} = a_\alpha, \qquad\qquad \frac{\partial}{\partial t} \{w_\alpha(t) \ x_{i,\alpha}(t)\} = b_{i,\alpha}. \tag{4.41}$$

As in the one dimensional case, for different choices of $\mathbf{l} = [l_1, l_2, \ldots, l_{N_d}]$, a system of linear equations can be constructed

$$
A_{\mathrm{MV}} \cdot
\begin{bmatrix}
\mathbf{a} \\
\mathbf{b}_1 \\
\vdots \\
\mathbf{b}_{N_d}
\end{bmatrix}
= \widetilde{\mathbf{S}} \, .
\tag{4.42}
$$

Again the corresponding unknowns can be found by inversion which requires

$$
A_{MV} \in \mathbb{R}^{N_{\mathrm{MV}} \times N_{\mathrm{MV}}} \, , \quad N_{\mathrm{MV}} = (N_d + 1) N_\alpha
\tag{4.43}
$$

to be regular. The set of moments chosen to build the linear system is of crucial importance for the stability and performance of the overall algorithm. In fact, the moments chosen for construction of the matrix are the moments which will be predicted the most accurate. Thus, moments representing important properties of the number density distribution should always be included. These could be for example the overall number of particles and the means with respect to particle properties which are represented by the zeroth and the first order moments of the distribution. However, the remaining moments to complete the set have to be chosen with care as some sets result in singular matrices. There are sophisticated rules for obtaining those sets for $N_d = 2, 3$ as reported for example in [69]. As in the one dimensional case the use of fractional moments is suggested to improve condition of the matrix, in particular for a large set of abscissas. However, no general rules for the choice of moments exist for higher dimensional cases with $N_d > 3$.

Alternatively, the unknowns can be determined using the Moore-Penrose-Pseudoinverse if A_{MV} is constructed from a moment set which is larger than the number of unknowns

$$
A_{\mathrm{MV}} \in \mathbb{R}^{N_{\mathrm{MV}}^* \times N_{\mathrm{MV}}} \, , \quad N_{\mathrm{MV}}^* > N_{\mathrm{MV}} \, , \quad N_{\mathrm{MV}} = (N_d + 1) N_\alpha \, .
\tag{4.44}
$$

As this strategy is equivalent to a solution of the linear system via linear regression by minimization of the quadratic error, all chosen moments are approximated equally good. However, the overall accuracy of the approximation always depends on the problem at hand.

Nevertheless, analytic solutions can be found analogously to the one dimensional case. For pure growth the solution is obtained by comparison of the coefficients for the zeroth and arbitrary first order moments

$$
\frac{\mathrm{d}w_\alpha(t)}{\mathrm{d}t} = 0 \, , \quad \frac{\mathrm{d}x_{i,\alpha}(t)}{\mathrm{d}t} = G_{i,\alpha} = G_i(t, \mathbf{x}_\alpha(t)) \, .
\tag{4.45}
$$

The corresponding dynamics of weights and abscissas for a pure particle death process are given by

$$
\frac{\mathrm{d}w_\alpha(t)}{\mathrm{d}t} = -w_\alpha(t) D_\alpha = -w_\alpha(t) D(t, \mathbf{x}_\alpha(t)) \, , \quad \frac{\mathrm{d}x_{i,\alpha}(t)}{\mathrm{d}t} = 0 \, .
\tag{4.46}
$$

The abscissa and weight dynamics of a process with growth and death are given by summation of (4.45) and (4.46). These solutions are also valid for higher order non

mixed moments. Furthermore, it can be shown that the solutions are also valid for arbitrary mixed moments.

In case the nucleation kinetics are not negligible, the corresponding dynamics of weights and abscissas have to be computed from a linear system of equations (4.42). Here, the critical points discussed for the one dimensional case also apply for the multi dimensional one: A_{MV} may possibly be ill-conditioned or even become singular in case two or more abscissas take the same value. Furthermore as already mentioned, it is not clear which moments have to be chosen on the right hand side of (4.42) to compute the dynamics of weights and abscissas. To circumvent the numerical solution of the linear system, an alternative approach which is based on the reformulation of the overall PBE will be presented next.

4.2.5. Reformulation of nucleation processes

In the following, an alternative problem formulation for the general multi dimensional PBE (2.10)

$$\frac{\partial n(t, \mathbf{x})}{\partial t} + \nabla_{\mathbf{x}} \left\{ \mathbf{G}(\mathbf{t}, \mathbf{x}) \, n(t, \mathbf{x}) \right\} = -D(t, \mathbf{x}) \, n(t, \mathbf{x}) + Nuc(t, \mathbf{x}) \,. \tag{4.47}$$

will be presented. Here, the nucleation $Nuc(t, \mathbf{x})$ is assumed to depend on the time and the internal coordinates. For simplicity, homogeneous boundary and initial conditions are assumed

$$n(t = 0, \mathbf{x}) = 0 \,, \quad n(t, \mathbf{x} = \mathbf{0}) = 0 \,. \tag{4.48}$$

Discrete reformulation

To circumvent the direct involvement of the nucleation kinetics, the PBE can be approximated by a series of initial value problems as presented in [57].

$$n(t, \mathbf{x}) \approx \sum_{k=1}^{N_{\mathrm{dis}}} n_k(t, \mathbf{x}) \,. \tag{4.49}$$

The principle idea is illustrated for a one dimensional example in Fig. 4.1. For this purpose, the nucleation kinetics is approximated by a sequence of pulses

$$
\begin{aligned}
Nuc(t, \mathbf{x}) &= \sum_{k=1}^{N_{\mathrm{dis}}} \left(\sigma_H(t - t_{k-1}) - \sigma_H(t - t_k) \right) \, Nuc(t_{k-1}, \mathbf{x}) \\
&= \sum_{k=1}^{N_{\mathrm{dis}}} \frac{\left(\sigma_H(t - t_{k-1}) - \sigma_H(t - t_k) \right)}{(t_k - t_{k-1})} \, Nuc(t_{k-1}, \mathbf{x}) \, (t_k - t_{k-1})
\end{aligned}
\tag{4.50}
$$

where the Heaviside step function is defined as

$$\sigma_H(t) = \begin{cases} 0 & , \quad t \leq 0 \\ 1 & , \quad t \geq 0 \end{cases} \,. \tag{4.51}$$
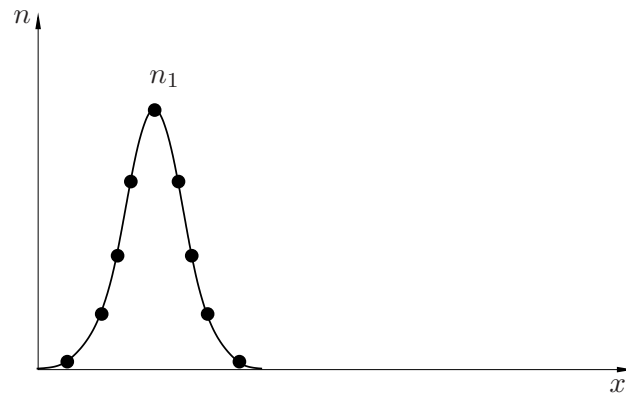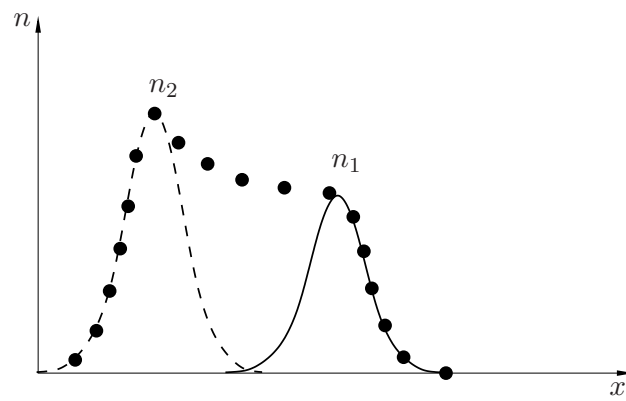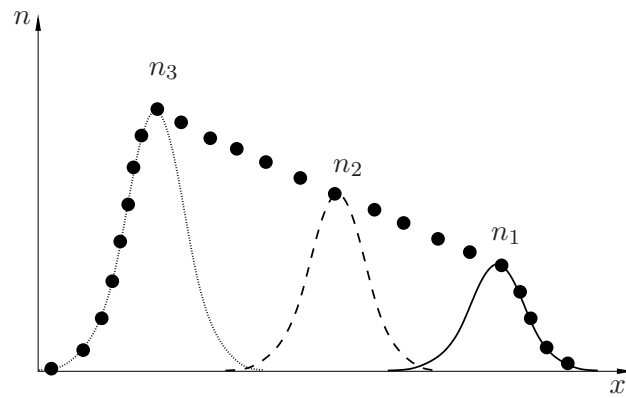
(a) t_1

(b) $t_2 > t_1$

(c) $t_3 > t_2$

Figure 4.1.: *Approximation of overall number density n (dotted) by series of initial value problems for one dimensional case*

In the remainder, the normalized pulses of finite width in (4.50)

$$\frac{\sigma_H\left(t - t_{k-1}\right) - \sigma_H\left(t - t_k\right)}{t_k - t_{k-1}} \tag{4.52}$$

are approximated by Dirac pulses $\delta\left(t - t_k\right)$, which are obtained in the limit $t_{k-1} \to t_k$. In result, the nucleation kinetics can be approximated by

$$Nuc(t, \mathbf{x}) \approx \sum_{k=1}^{N_{\mathrm{dis}}} \delta\left(t - t_k,\right) Nuc(t_{k-1}, \mathbf{x})\left(t_k - t_{k-1}\right) . \tag{4.53}$$

With (4.49) and (4.53) the population balance (4.47) reads

$$\frac{\partial}{\partial t} \sum_{k=1}^{N_{\mathrm{dis}}} n_k + \nabla_{\mathbf{x}}\left\{\mathbf{G} \sum_{k=1}^{N_{\mathrm{dis}}} n_k\right\} = -D \sum_{k=1}^{N_{\mathrm{dis}}} n_k + \sum_{k=1}^{N_{\mathrm{dis}}} \delta(t - t_k) Nuc(t_{k-1}, x)\left(t_k - t_{k-1}\right) . \tag{4.54}$$

The dynamics of each subpopulation $n_k(t, x)$ is given by

$$\frac{\partial n_k(t, x)}{\partial t} + \nabla_{\mathbf{x}}\left\{\mathbf{G}(t, \mathbf{x})\ n_k(t, x)\right\} = -D(t, x)\ n_k(t, x)$$
$$+ \delta(t - t_k) Nuc(t_{k-1}, x)\left(t_k - t_{k-1}\right)$$
$$n_k(t = t_k, \mathbf{x}) = 0 . \tag{4.55}$$

Alternatively, one may write

$$\frac{\partial n_k(t, x)}{\partial t} + \nabla_{\mathbf{x}}\left\{\mathbf{G}(t, \mathbf{x})\ n_k(t, x)\right\} = -D(t, x)\ n_k(t, x)$$
$$n_k(t = t_k, \mathbf{x}) = Nuc(t_{k-1}, \mathbf{x})\left(t_k - t_{k-1}\right) . \tag{4.56}$$

Applying the DQMOM formalism, each of these PBEs can be represented by a weighted sum of N_α delta distributions. As each PBE (4.56) is only affected by the growth and death kinetics, solutions of the corresponding system of linear equations (4.42)-(4.46) gives the dynamics of the corresponding weights and abscissas

$$\frac{\mathrm{d}w_{k,\alpha}(t)}{\mathrm{d}t} = -D_{k,\alpha} w_{k,\alpha}(t) = -D_k(t, \mathbf{x}_\alpha(t)) w_{k,\alpha}(t) , \quad \frac{\mathrm{d}\mathbf{x}_{k,\alpha}(t)}{\mathrm{d}t} = \mathbf{G}_{k,\alpha} = \mathbf{G}_k(t, \mathbf{x}_\alpha(t)) . \tag{4.57}$$

The initial conditions for the weights and abscissas are calculated from the moments of $n_k(t = t_k, \mathbf{x}) = Nuc(t_{k-1}, \mathbf{x})\left(t_k - t_{k-1}\right)$ and are given by

$$w_{k,\alpha}(t = t_k) = w_\alpha^{\mathrm{init}}(t_{k-1})\left(t_k - t_{k-1}\right) , \quad \mathbf{x}_{k,\alpha}(t = t_k) = \mathbf{x}_{k,\alpha}^{\mathrm{init}}(t_{k-1}) . \tag{4.58}$$

Integral quantities of the original problem formulation can now be approximated in terms of the series representation

$$\int_{\mathbf{X}} f(t, \mathbf{x}) n(t, \mathbf{x})\ \mathrm{d}\mathbf{x} \approx \sum_{k=1}^{N_{\mathrm{dis}}} \int_{\mathbf{X}} f(t, \mathbf{x}) n_k(t, \mathbf{x})\ \mathrm{d}\mathbf{x}$$
$$\approx \sum_{k=1}^{N_{\mathrm{dis}}} \sum_{\alpha=1}^{N_\alpha} f(t, \mathbf{x}_{k,\alpha}) w_{k,\alpha} . \tag{4.59}$$

In summary, the original problem was reformulated as series of subproblems which correspond to particle growth/death processes and the nucleation kinetics of the original problem is shifted to the corresponding initial conditions of the subproblems. The moments of each subproblem are approximated by weighted sums of abscissas with dynamics given by (4.57).

Continuous reformulation

Alternatively, a continuous reformulation of the original problem can be derived by introducing the age τ of a particle as an additional internal coordinate, as presented in [29]. The dynamics of the corresponding number density distribution $\widetilde{n}(t, \mathbf{x}, \tau)$ is given by

$$\frac{\partial \widetilde{n}(t, \mathbf{x}, \tau)}{\partial t} + \nabla_{\mathbf{x}} \{\mathbf{G}(t, \mathbf{x}) \cdot \widetilde{n}(t, \mathbf{x}, \tau)\} + \frac{\partial \widetilde{n}(t, \mathbf{x}, \tau)}{\partial \tau} = -D(t, \mathbf{x}) \, \widetilde{n}(t, \mathbf{x}, \tau) \qquad (4.60)$$

with

$$\begin{aligned} \widetilde{n}(t, \mathbf{x}, \tau = 0) &= Nuc(t, \mathbf{x}) \,, \\ \widetilde{n}(t, \mathbf{x}, \tau = \infty) &= 0 \,. \end{aligned} \qquad (4.61)$$

Here, $\widetilde{n}(t, \tau > t) = 0$ by definition. The original number density distribution is obtained by integration over all ages

$$n(t, \mathbf{x}) = \int_{\mathbf{X}} \widetilde{n}(t, \mathbf{x}, \tau) \, \mathrm{d}\tau \,. \qquad (4.62)$$

It can be verified easily that integration of (4.60) over τ results in the original PBE (4.47).

Up to this point, the N_d dimensional PDE including nucleation and homogeneous boundary conditions (4.47) has been reformulated as an homogeneous $N_d + 1$ dimensional PDE (4.60) with nucleation in the boundary conditions (4.61). In the context of moment approximation via DQMOM this is advantageous as the reformulation (4.60) is only characterized by growth with respect to $\mathbf{x}$ and τ and withdrawal whereas the nucleation kinetics is shifted to the boundary condition (4.61). In the remainder, the basic idea of DQMOM is used to approximate age-distributed moments

$$\widetilde{m}_{\mathbf{l}}(t, \tau) = \int_{\mathcal{X}} x_1^{l_1} \cdots x_{N_d}^{l_{N_d}} \widetilde{n}(t, \mathbf{x}, \tau) \, \mathrm{d}\mathbf{x} \qquad (4.63)$$

in a similar fashion as presented in the upper part of this section. For simplicity, the following derivation will focus on the one dimensional case $N_d = 1$. Age-distributed moments are now approximated by a weighted sum of age-distributed abscissas

$$\widetilde{m}_l(t, \tau) = \int_0^\infty x^l \widetilde{n}(t, x, \tau) \, \mathrm{d}x \approx \sum_{\alpha=1}^{N_\alpha} \widetilde{x}_\alpha^l(t, \tau) \, \widetilde{w}_\alpha(t, \tau) \,. \qquad (4.64)$$

The corresponding moment dynamics are given by

$$\frac{\partial \widetilde{m}_l}{\partial t} = \frac{\partial}{\partial t} \int\limits_0^\infty x^l \widetilde{n} \; \mathrm{d}x \approx \frac{\partial}{\partial t} \sum_{\alpha=1}^{N_\alpha} \widetilde{x}_\alpha^l \; \widetilde{w}_\alpha$$

$$= \sum_{\alpha=1}^{N_\alpha} \left[\widetilde{x}_\alpha^l \frac{\partial \widetilde{w}_\alpha}{\partial t} + \widetilde{w}_\alpha \; l \; \widetilde{x}_\alpha^{l-1} \frac{\partial \widetilde{x}_\alpha}{\partial t} \right]$$

$$= \sum_{\alpha=1}^{N_\alpha} \left[(1-l)\widetilde{x}_\alpha^l \frac{\partial \widetilde{w}_\alpha}{\partial t} + l \; \widetilde{x}_\alpha^{l-1} \frac{\partial \widetilde{w}_\alpha \widetilde{x}_\alpha}{\partial t} \right]$$

$$= \sum_{\alpha=1}^{N_\alpha} \left[-\widetilde{x}_\alpha^l \widetilde{D}_\alpha \widetilde{w}_\alpha - \frac{\partial \widetilde{w}_\alpha \widetilde{x}_\alpha^l}{\partial \tau} + l \; \widetilde{x}_\alpha^{l-1} \widetilde{w}_\alpha \widetilde{G}_\alpha \right] , \qquad (4.65)$$

with

$$\widetilde{D}_\alpha = D(t, \widetilde{x}_\alpha(t,\tau)), \quad \widetilde{G}_\alpha = G(t, \widetilde{x}_\alpha(t,\tau)) . \qquad (4.66)$$

For the approximation of the zeroth order age-distributed moment ($l = 0$) one obtains

$$\sum_{\alpha=1}^{N_\alpha} \frac{\partial \widetilde{w}_\alpha}{\partial t} = \sum_{\alpha=1}^{N_\alpha} \left[-\widetilde{D}_\alpha \widetilde{w}_\alpha - \frac{\partial \widetilde{w}_\alpha}{\partial \tau} \right] \qquad (4.67)$$

and thus by comparison of the coefficients the dynamics of the age-distributed weights are found as

$$\frac{\partial \widetilde{w}_\alpha}{\partial t} + \frac{\partial \widetilde{w}_\alpha}{\partial \tau} = -\widetilde{D}_\alpha \widetilde{w}_\alpha , \quad \alpha = 1, \ldots, N_\alpha . \qquad (4.68)$$

Analogous, for approximation of the first order age-distributed moment ($l = 1$) one obtains

$$\sum_{\alpha=1}^{N_\alpha} \left[\frac{\partial \widetilde{w}_\alpha \widetilde{x}_\alpha}{\partial t} \right] = \sum_{\alpha=1}^{N_\alpha} \left[-\widetilde{x}_\alpha \widetilde{D}_\alpha \widetilde{w}_\alpha - \frac{\partial \widetilde{w}_\alpha \widetilde{x}_\alpha}{\partial \tau} + \widetilde{w}_\alpha \widetilde{G}_\alpha \right] \qquad (4.69)$$

and together with (4.68)

$$\sum_{\alpha=1}^{N_\alpha} \widetilde{w}_\alpha \frac{\partial \widetilde{x}_\alpha}{\partial t} = \sum_{\alpha=1}^{N_\alpha} \left[-\widetilde{w}_\alpha \frac{\partial \widetilde{x}_\alpha}{\partial \tau} + \widetilde{w}_\alpha \widetilde{G}_\alpha \right] . \qquad (4.70)$$

Comparison of the coefficients on the right and left hand side of this equation gives the dynamics of the age-distributed abscissas

$$\frac{\partial \widetilde{x}_\alpha}{\partial t} + \frac{\partial \widetilde{x}_\alpha}{\partial \tau} = \widetilde{G}_\alpha , \quad \alpha = 1, \ldots, N_\alpha . \qquad (4.71)$$

The boundary and initial conditions of abscissas and weights are calculated from moments of the nucleation distribution $Nuc(t,x)$

$$\int\limits_0^\infty x^l Nuc(t,x) \; \mathrm{d}x = m_l(t, \tau = 0) \approx \sum_{\alpha=1}^{N_\alpha} x_{\alpha,\mathrm{init}}^l(t) w_{\alpha,\mathrm{init}}(t) \qquad (4.72)$$

and are given by

$$\widetilde{w}_\alpha(t, \tau = 0) = w_{\alpha,\text{init}}(t) , \quad \widetilde{x}_\alpha(t, \tau = 0) = x_{\alpha,\text{init}}(t) . \tag{4.73}$$

Integral quantities like moments are approximated by integration over the age coordinate τ

$$m_l(t) = \int\limits_0^\infty \widetilde{m}_l(t, \tau)\mathrm{d}\tau = \int\limits_0^\infty \int\limits_0^\infty x^l \widetilde{n}(t, x, \tau)\mathrm{d}x \approx \int\limits_0^\infty \sum_{\alpha=1}^{N_\alpha} \widetilde{w}_\alpha(t, \tau)\widetilde{x}_\alpha(t, \tau)\mathrm{d}\tau \tag{4.74}$$

For a one dimensional population balance equation, a set of $2 \cdot N_\alpha$ one dimensional PDEs describing the dynamics of weights (4.68) and abscissas (4.71) has to be solved. It can be derived analogously, that for a N_d dimensional PBE a set of $(N_d + 1) N_\alpha$ one dimensional PDEs emerges

$$\frac{\partial \widetilde{w}_\alpha}{\partial t} + \frac{\partial \widetilde{w}_\alpha}{\partial \tau} = -\widetilde{D}_\alpha \widetilde{w}_\alpha ,$$

$$\frac{\partial \widetilde{x}_{1,\alpha}}{\partial t} + \frac{\partial \widetilde{x}_{1,\alpha}}{\partial \tau} = \widetilde{G}_{1,\alpha} ,$$

$$\vdots$$

$$\frac{\partial \widetilde{x}_{N_d,\alpha}}{\partial t} + \frac{\partial \widetilde{x}_{N_d,\alpha}}{\partial \tau} = \widetilde{G}_{N_d,\alpha} ,$$

$$\alpha = 1, \ldots, N_\alpha , \tag{4.75}$$

describing the temporal evolution of weights and abscissas.

Alternatively, integration of (4.60) over the internal coordinates yields the dynamics

$$\frac{\partial \widetilde{m}_0(t, \tau)}{\partial t} + \frac{\partial \widetilde{m}_0(t, \tau)}{\partial \tau} = \int\limits_{\mathbf{X}} -D(t, \mathbf{x}, \tau)\, \widetilde{n}(t, \mathbf{x}, \tau)\mathrm{d}\mathbf{x} \tag{4.76}$$

of the zeroth order age-distributed moment

$$\widetilde{m}_0(t, \tau) = \int\limits_{\mathbf{X}} \widetilde{n}(t, \mathbf{x}, \tau)\, \mathrm{d}\mathbf{x} \tag{4.77}$$

for

$$G \cdot n|_0^\infty = 0 . \tag{4.78}$$

If further, it is assumed that all particles of the same age share the same properties, than the dynamics in the multi dimensional property state space reduces to the following set of one dimensional PDEs as presented in [38]

$$\frac{\partial \widetilde{x}_i}{\partial t} + \frac{\partial \widetilde{x}_i}{\partial \tau} = \widetilde{G}_i , \quad i = 1, \ldots, N_d \tag{4.79}$$

It is worth noting that (4.76) and (4.79) are equivalent to (4.75) for $N_\alpha = 1$ because

$$\widetilde{m}_0(t, \tau) \approx \int\limits_{\mathbf{X}} \sum_{\alpha=1}^{1} \widetilde{w}_\alpha(t, \tau)\delta\left(\mathbf{x} - \widetilde{\mathbf{x}}_\alpha(t, \tau)\right)\, \mathrm{d}\mathbf{x} = \widetilde{w}_1(t, \tau) . \tag{4.80}$$

(a) *Overall Setup*

(b) *Age-distributed overall number of cells*
$\widetilde{m}_0(t,\tau)$

(c) *Age-distributed coordinate $\widetilde{x}_1(t,\tau)$*

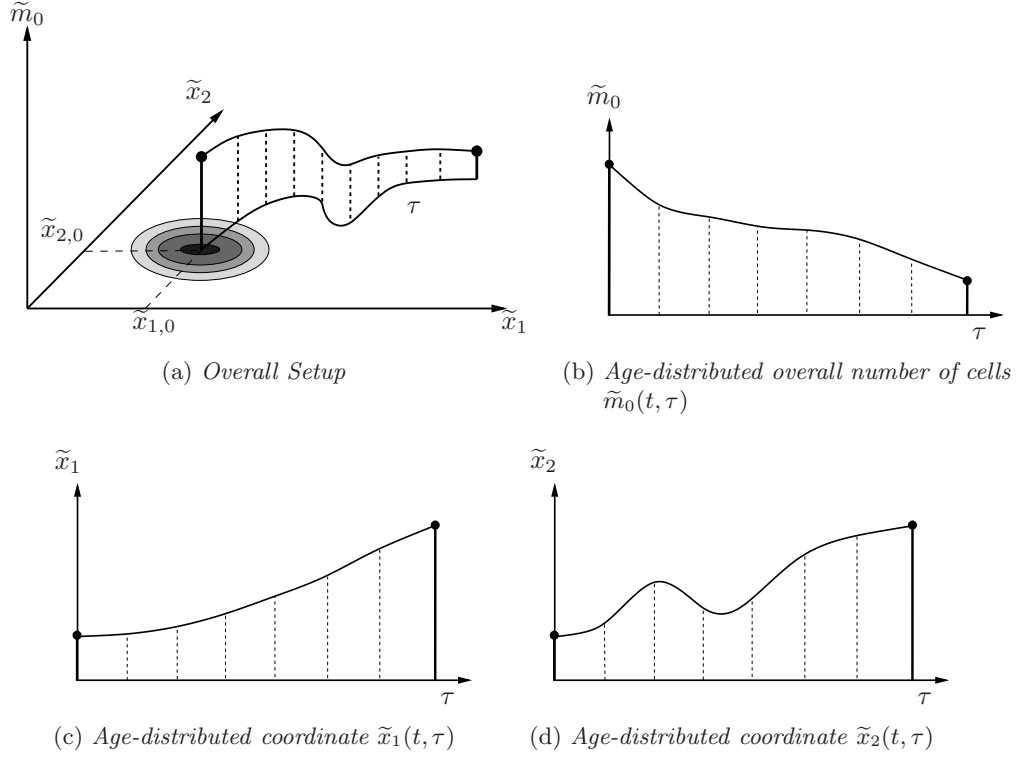(d) *Age-distributed coordinate $\widetilde{x}_2(t,\tau)$*

Figure 4.2.: *Schematic representation of age-distributed reformulation for two dimensional PBE at arbitrary time $t > t_0$; $Nuc(t,\mathbf{x})$ (grey) is approximated by one abscissa at $[\widetilde{x}_{1,0}, \widetilde{x}_{2,0}]$; age coordinate τ can be viewed as a curve parameter*

In Fig. 4.2, a schematic representation for a two dimensional example with time invariant nucleation kinetics $Nuc(t, x_1, x_2) = Nuc(x_1, x_2)$ is depicted. During the process the particles change their internal states and their number density distribution while new particles are introduced at $[\widetilde{x}_{1,0}, \widetilde{x}_{2,0}]$. This results in a ribbon like distribution along τ within the property state space for $t > t_0$. Similar to the method of characteristics (see Appendix A), the age coordinate τ can be viewed as a curve parameter parameterizing this ribbon. It can be seen easily that particles of the same age, i.e. the same position on the ribbon, have the same internal state.

4.3. Efficient choice of abscissas

In the previous section it was presented how to come up with dynamic equations for weights and abscissas for an approximate computation of moments. However, so far no word was lost on the initial values for the corresponding systems of ODEs (4.41). In the DQMOM context initial abscissas and weights are chosen based on moments of the initial number density distribution. The choice of this values represents an important factor in the performance of the overall moment approximation algorithm and will be discussed in the following.

The problem at hand is strongly related to the numerical approximation of multi dimensional integrals. The basic idea, which was already presented in the previous chapter, is to find a sufficiently accurate approximation of a general multi dimensional integral using a weighted sum of function evaluations

$$\int_{\mathbf{X}} f(\mathbf{x})p(\mathbf{x})\mathrm{d}\mathbf{x} \approx \sum_{\alpha=1}^{N_\alpha} w_\alpha f(\mathbf{x}_\alpha) \tag{4.81}$$

at the abscissas $\mathbf{x}_\alpha$. In the standard literature on numerical integration, p is also referred to as *weighting function* (see e.g. [16, 107]). In the special case of population balances, p corresponds to the number density distribution $n(t=0,\mathbf{x})$ and $f(\mathbf{x})$ is a monomial

$$m_{l_1,\ldots,l_{N_d}}(t=0) = \int_{\mathbf{X}} x_1^{l_1} \cdots x_{N_d}^{l_{N_d}} \; n(t=0,\mathbf{x}) \; \mathrm{d}\mathbf{x}$$

$$\approx \sum_{\alpha=1}^{N_\alpha} w_\alpha(t=0)x_{1,\alpha}^{l_1}(t=0) \cdots x_{\alpha,N_d}^{l_{N_d}}(t=0). \tag{4.82}$$

For the following explanations it is assumed that the initial distribution is normalized

$$\int_{\mathbf{X}} n(t=0,\mathbf{x}) \; \mathrm{d}\mathbf{x} = \int_{\mathbf{X}} n_0(\mathbf{x}) \; \mathrm{d}\mathbf{x} = 1 \tag{4.83}$$

without loss of generalization.

As already mentioned above, the quality of this initial approximation strongly affects the overall performance of the moment approximation algorithm: If moments of the initial distribution are approximated poorly by the initial abscissas and weights, one can hardly expect that moment approximations during the process are sufficiently accurate. Usually, a larger number of abscissas comes along with a higher degree of accuracy but also with increased numerical effort. Hence a good trade off between those requirements has to be found. The methods for finding the sets of weights and abscissas are often referred to as *cubature rules*. Those can be roughly classified into random based and deterministic rules, where the latter contain *product rules* and *non-product rules* (see e.g. [16, 107]. In the following, different formulas from those classes and their application for the overall moment approximation algorithm are discussed.

4.3.1. Random based rules

When applying random based rules, N_{MC} abscissas are determined by random sampling of the integration region. The corresponding numerical integration method is also known as *Monte Carlo Integration* [16]. Its popularity is based on the straightforward generation of abscissas. Furthermore, the accuracy of this method is independent of the dimension of the problem. In general, the error bound is given by

$$\frac{\sigma_{\text{MC}}}{\sqrt{N_{\text{MC}}}} \tag{4.84}$$

were N_{MC} represents the number of abscissas, σ_{MC}^2 is the variance of the estimated integral approximation. To increase the accuracy, one can either increase the number of samples or reduce σ_{MC}. If one increases the number of abscissas by a factor of 100, the error of the approximation is only reduced by a factor of 10. Alternatively, several variance reducing techniques are available [16]. In the most straightforward approach, initial abscissas are drawn using the initial distribution as a probability distribution. In this case all weights are identical

$$w_\alpha = \frac{1}{N_{\text{MC}}} \,. \tag{4.85}$$

This technique is also referred to as *Importance Sampling* [16], because samples are concentrated in regions which contribute most to the overall integral. However, in general a large number of samples is necessary to cover the whole integration region.

As the number of abscissas is generally very large for random based rules, they are not a good candidate for an efficient approximation algorithm. However, they will be used in the following to generate reference solutions to evaluate the accuracy of other cubature rules.

4.3.2. Product rules

Product rules are multi dimensional extensions of one dimensional quadrature rules. In the one dimensional case, Gaussian quadrature rules which are based on orthogonal polynomials can be applied to come up with appropriate sets of N_{GA} abscissas and weights, e.g. if the initial distribution corresponds to a Gaussian distribution, the Gauss Hermite rule may be applied [16]. The corresponding integral approximations are exact for polynomials up to degree $2N_{\text{GA}} - 1$ in f. These formulas can be extended to multi dimensional problems by using tensor products of one dimensional weight and abscissa sets $\mathcal{W}_i$ and $\mathcal{S}_i$, respectively.

$$\mathcal{S} = \mathcal{S}_1 \otimes \mathcal{S}_2 \otimes \ldots \otimes \mathcal{S}_{N_d}, \quad \mathcal{W} = \mathcal{W}_1 \otimes \mathcal{W}_2 \otimes \ldots \otimes \mathcal{W}_{N_d}. \tag{4.86}$$

In result, the overall number of abscissas is given by

$$N_{\text{GA}} = N_{\text{GA},1} \, N_{\text{GA},2} \, \ldots \, N_{\text{GA},N_d} \,. \tag{4.87}$$

In consequence, those rules suffer badly from the curse of dimensionality, as the sizes of the weight and abscissa sets generally increase exponentially with N_d. This is a major disadvantage for high dimensional applications and for $N_d > 5$ product rules are not considered as a reasonable choice for an efficient approximation of multi dimensional integrals.

4.3.3. Non-product rules

Non-product rules are advantageous in particular for high dimensional problem. In contrast to product rules, the initial sets of abscissas and weights are generated directly in the full problem space instead of tensoring low dimensional Gaussian quadrature formulas. The basic idea is to exploit special properties of the initial distribution and the weighting function respectively. In the case of symmetric weighting functions, formulas can be designed that scale polynomially, in the best case linearly with N_d which makes them favorable for high dimensional applications.

As an important concept, the notion of generator functions was introduced by Lerner [56]. These enable an efficient and direct generation of the abscissa and weight set using permutations and sign combinations. The first three generator functions are given by

1. $GF_I(\mathbf{0})$ changes no element in a given vector

2. $GF_{II}(\pm\theta)$ changes one element in a given vector

3. $GF_{III}(\pm\theta, \pm\theta)$ changes two elements in a given vector

The corresponding distribution of abscissas is depicted in Fig. 4.3 for a two dimensional example. Therein, the parameter θ can be viewed as a measure for the spread of the abscissas. The abscissas are used to approximate the integral by the already introduced weighted sum of abscissas. Thereby the abscissas are weighted depending on their corresponding generator function

$$\int\limits_{\mathbf{X}} f(\mathbf{x})n_0(\mathbf{x})\mathrm{d}\mathbf{x} \approx w_I GF_I + w_{II} \sum_{i=1}^{N_{II}} GF_{II} + w_{III} \sum_{i=1}^{N_{III}} GF_{III}. \qquad (4.88)$$

As for the general moment approximation (4.14), the unknown parameters w_i and θ are solutions of a system of nonlinear equations consisting of realizations of (4.88) for different moments of the initial distribution n_0. In case n_0 corresponds to a Gaussian distribution $\mathcal{N}(\mathbf{0}, \mathbf{I})$ the following nonlinear equation system has to be considered

$$\begin{aligned}
\int\limits_{\mathbf{X}} n_0 \mathrm{d}\mathbf{x} &= 1 &\approx\quad w_I &\quad +4w_{II} &\quad +4w_{III} \\
\int\limits_{\mathbf{X}} x_i^2 n_0 \mathrm{d}\mathbf{x} &= 1 &\approx\quad &\quad +2w_{II}\theta^2 &\quad +4w_{III}\theta^2 \\
\int\limits_{\mathbf{X}} x_i^4 n_0 \mathrm{d}\mathbf{x} &= 4 &\approx\quad &\quad +2w_{II}\theta^4 &\quad +4w_{III}\theta^4 \\
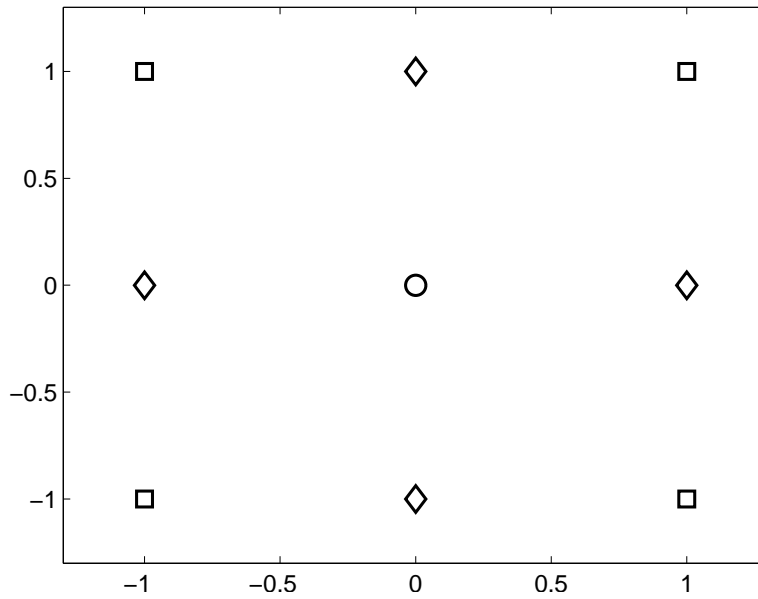&\;\;\vdots
\end{aligned} \qquad (4.89)$$

Figure 4.3.: *Location of different Generator Functions*; GF_I (circle), GF_{II} (rhomboids), GF_{III} (squares)

In comparison to product rules such non-product rules are characterized by a weaker approximation power. If only the first and the second order generator functions are used within the derivation, the resulting cubature rule is exact for monomials up to order 3 in f, for which a particular cubature formula will be discussed in the next chapter. When additionally taking into account GF_{III} the corresponding cubature rule will be exact for monomials up to order 5 in f [98].

4.3.4. Sigma points

In literature on numerical integration techniques (e.g. [16, 107]), a vast amount of non-product rules for different weighting functions can be found. One rule specifically tailored to Gaussian weighting functions has received large attention in the last two decades. The *sigma point rule* as first introduced by Julier and Uhlmann [51, 50] has been applied extensively for state and parameter estimation purposes for control systems within the *Unscented Kalman Filter* (see [48] and the references therein) as an alternative to the *Extended Kalman Filter*. Furthermore, applications include robust process optimization [88] as well as model discrimination and experimental design [26, 99].

For the use of the sigma point formula, the initial distribution $n(t = 0, \mathbf{x})$ is assumed to correspond to a multi dimensional Gaussian distribution $\mathcal{N}(\mu, \mathbf{\Sigma})$, where μ is the vector of mean values and $\mathbf{\Sigma}$ being the covariance matrix. The sets of abscissas and
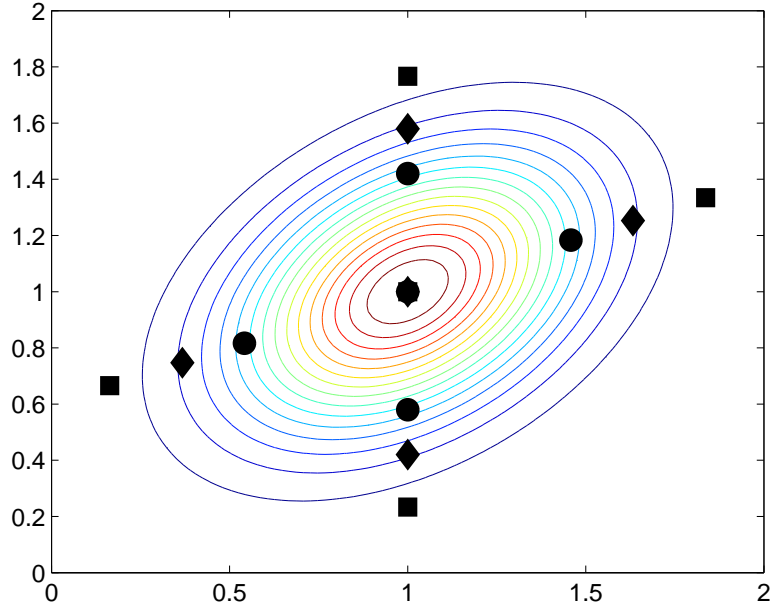
Figure 4.4.: *Location of Sigma-Points for* $\lambda_{\mathrm{SP}} = 0.1, 2, 5$

weights [113] are given by

$$
\begin{aligned}
\mathbf{x}_0 &= \mu \\
\mathbf{x}_i &= \mu + \sqrt{\lambda_{\mathrm{SP}} + N_d}\sqrt{\boldsymbol{\Sigma}}_i \\
\mathbf{x}_{i+N_d} &= \mu - \sqrt{\lambda_{\mathrm{SP}} + N_d}\sqrt{\boldsymbol{\Sigma}}_i
\end{aligned}
\tag{4.90}
$$

with $\sqrt{\boldsymbol{\Sigma}}_i$ being the i-th column of the square root of the covariance matrix and

$$
\begin{aligned}
w_0 &= \frac{\lambda_{\mathrm{SP}}}{\lambda_{\mathrm{SP}} + N_d} \\
w_i = w_{i+N_d} &= \frac{1}{2(\lambda_{\mathrm{SP}} + N_d)} \, .
\end{aligned}
\tag{4.91}
$$

In [113], the classical formulation was extended for additional design parameters

$$
\lambda_{\mathrm{SP}} = \alpha_{\mathrm{SP}}^2(\kappa_{\mathrm{SP}} + N_d) - N_d
\tag{4.92}
$$

which can be used to control the spread of the abscissas in the state space, as it is depicted for a two dimensional example in Fig. 4.4 for different values of λ_{SP}. This rule corresponds to a non-product formula which is exact for monomials up to order 3 and scales linearly with dimension $N_{\mathrm{SP}} = 2\,(N_d + 1)$. Several higher order sigma point formulas have been proposed to improve the accuracy of the approximation but also the design procedure for minimal number abscissa sets was shown in [48]. Nevertheless in this thesis focus will be on the above presented classic formula (4.90) - (4.91).

4.3.5. Using sigma points for non-Gaussian distributions

Sigma points can only be chosen directly for Gaussian initial conditions. However, if that assumption is not justified, the sigma point approach has to be modified. For some types of the initial distribution special transformation formulas exist. For example if n_0 corresponds to a one dimensional logarithmic normal distribution $\mathcal{L}(\mu_{\mathcal{L}}, \sigma_{\mathcal{L}})$ the transformed set can be computed as follows [98]

1. sigma points $\mathbf{x}_i$ are computed for $\mathcal{N}(0,1)$ according to (4.90)

2. the transformed sigma point set is then given by

$$x_{i,\mathcal{L}} = \exp\left(\mu_{\mathcal{L}} + \sigma_{\mathcal{L}_k} x_i\right), \quad i = 1, \ldots, N_\alpha \tag{4.93}$$

3. the set of weights is not changed within the transformation procedure and is given by (4.90)

Further examples for those transformation formulas include *Gamma-* and *t-Student* distributions [98].

Yet, no general transformations exists if n_0 contains multimodalities. To improve this situation a two step procedure is advantageous, which was applied in [96] and [95] to approximate uncertainties of biological models. At first, n_0 is approximated by a weighted sum of Gaussian distributions

$$n_0(\mathbf{x}) \approx \sum_{k=1}^{N_{\mathrm{GMD}}} w_k^{\mathrm{GMD}} \mathcal{N}(\mu_k, \boldsymbol{\Sigma}_k). \tag{4.94}$$

This approximation is also termed *Gaussian mixed density* (GMD) in some references. Afterwards, abscissas and weights are determined for each Gaussian applying the standard sigma point formula. For the overall approximation of the integrals the general formula (4.81) has to be adapted

$$\int_{\mathbf{X}} f(\mathbf{x}) n_0(\mathbf{x}) \mathrm{d}\mathbf{x} \approx \sum_{k=1}^{N_{\mathrm{GMD}}} w_k^{\mathrm{GMD}} \sum_{\alpha=1}^{N_\alpha} w_{k,\alpha} f(\mathbf{x}_{k,\alpha}). \tag{4.95}$$

The overall number of weights and abscissas is now given by $N_{\mathrm{GMD}} N_\alpha$. For more complex distributions (e.g. multi modal on a logarithmic scale) the previously mentioned approaches can be combined to come up with more accurate approximations.

4.4. One dimensional benchmark

First, the performance of the moment approximation algorithm is shown for a one dimensional growth process as presented in [20]. Focus is on the particle size distribution $n(t, x)$. Assuming that the process is dominated by particle growth and particle death/withdrawal, the following PBE characterizing the dynamics of the number density distribution can be derived

$$\frac{\partial n(t, x)}{\partial t} + \frac{\partial}{\partial x} \{G(t, x) \ n(t, x)\} = -D(t, x) \ n(t, x) \,. \tag{4.96}$$

Therein, the coefficients describing the growth and withdrawal of particles are nonlinear and time variant functions of the internal coordinate x

$$G(t, x) = \exp(g_t \ t) \sum_{l=0}^{3} g_i \ x^i \,,$$

$$D(t, x) = \exp(d_t \ t) \sum_{l=0}^{3} d_i \ x^i \,. \tag{4.97}$$

The initial particle distribution is assumed to be Gaussian

$$n(t = 0, x) = \mathcal{N}(\mu, \ \sigma) \,. \tag{4.98}$$

Furthermore, the boundary condition is homogeneous:

$$[G \ n(t, x = 0)] = 0 \,. \tag{4.99}$$

The dynamics of an arbitrary moment m_l are derived using (2.11)

$$\frac{\mathrm{d}m_l}{\mathrm{d}t} = (3 \ g_3 \ m_{l+2} + 2 \ g_2 \ m_{l+1} + 1 \ g_1 \ m_l) \exp(g_t \ t) +$$
$$(d_3 \ m_{l+3} + d_2 \ m_{l+2} + d_1 \ m_{l+1} + d_0 \ m_l) \exp(d_t \ t)$$
$$l = 0, 1, \ldots \quad . \tag{4.100}$$

It is easily seen, that the moment system is non-closing for non-zero coefficients a_i. The numerical values for the coefficients can be found in Table 4.1. Note, that this is a purely academic example and thus all parameters can be considered dimensionless.

To approximately compute the moments of the number density distribution the DQMOM is applied. For the system (4.96), the dynamics of the DQMOM abscissas and weights are given by (see Section 4.2.3):

$$\frac{\mathrm{d}x_\alpha(t)}{\mathrm{d}t} = G_\alpha = G(t, x_\alpha(t)) \,, \quad \frac{\mathrm{d}w_\alpha(t)}{\mathrm{d}t} = -w_\alpha(t) \ D_\alpha = -w_\alpha(t) \ D(t, x_\alpha) \,,$$
$$\alpha = 1, \ldots, N_\alpha \,. \tag{4.101}$$

As mentioned in Section 4.2, the abscissas move on the characteristic curves of the PDE (4.96). Based on the Gaussian initial distribution, the initial values for the abscissas were calculated according to the sigma point formulas (4.90) and (4.91) with $\lambda_{\mathrm{SP}} = 2$.

Table 4.1.: *Numerical values for coefficients and initial conditions for 1-dimensional benchmark example*

Parameter	Value	Parameter	Value
g_0	$2.40\ 10^{-2}$	d_0	$1.00\ 10^{-3}$
g_1	$4.80\ 10^{-4}$	d_1	$5.00\ 10^{-2}$
g_2	$1.44\ 10^{-2}$	d_2	$1.10\ 10^{-2}$
g_3	$9.60\ 10^{-4}$	d_3	$5.00\ 10^{-4}$
g_t	$2.50\ 10^{-2}$	d_t	$3.00\ 10^{-2}$
μ	4.00	σ	0.8
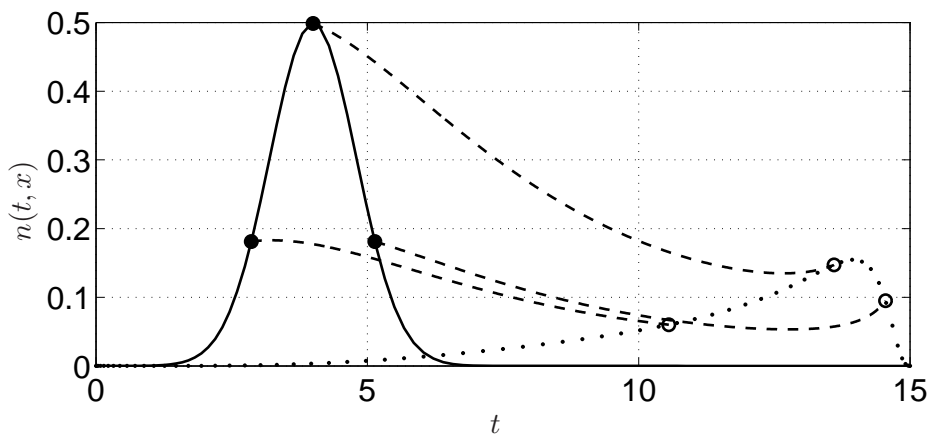
Figure 4.5.: *Dynamics of the particle number density distribution and the SP abscissas for one dimensional benchmark problem; initial number density (solid curve), final number density (dotted curve)*

The overall temporal evolution of the abscissas and the overall number density distribution is depicted in Fig. 4.5. Note, that the full solution was computed with a pseudo spectral collocation method (see for example [11]), which is described in more detail in Appendix D. This discretization based solution method for PDEs is characterized by a discretization of the internal coordinate which is based on Chebychev polynomials. In contrast to other discretization based methods as the finite volume method, a higher accuracy can be obtained for an equal degree of discretization. Vice versa, the same accuracy is achieved with a lower discretization effort, thereby drastically reducing the computational burden for the numerical solution of a PDE. In the current example 256 grid nodes are used and the reference values for the moments are calculated applying the Clenshaw - Curtis quadrature which provides a high degree of accuracy for the used pseudo spectral discretization scheme (see [112]). Furthermore, a standard QMOM approximation is implemented and compared to both, reference solution and approximation using DQMOM with sigma point abscissas.

In Fig. 4.6 the computed values of selected moments are shown. The corresponding
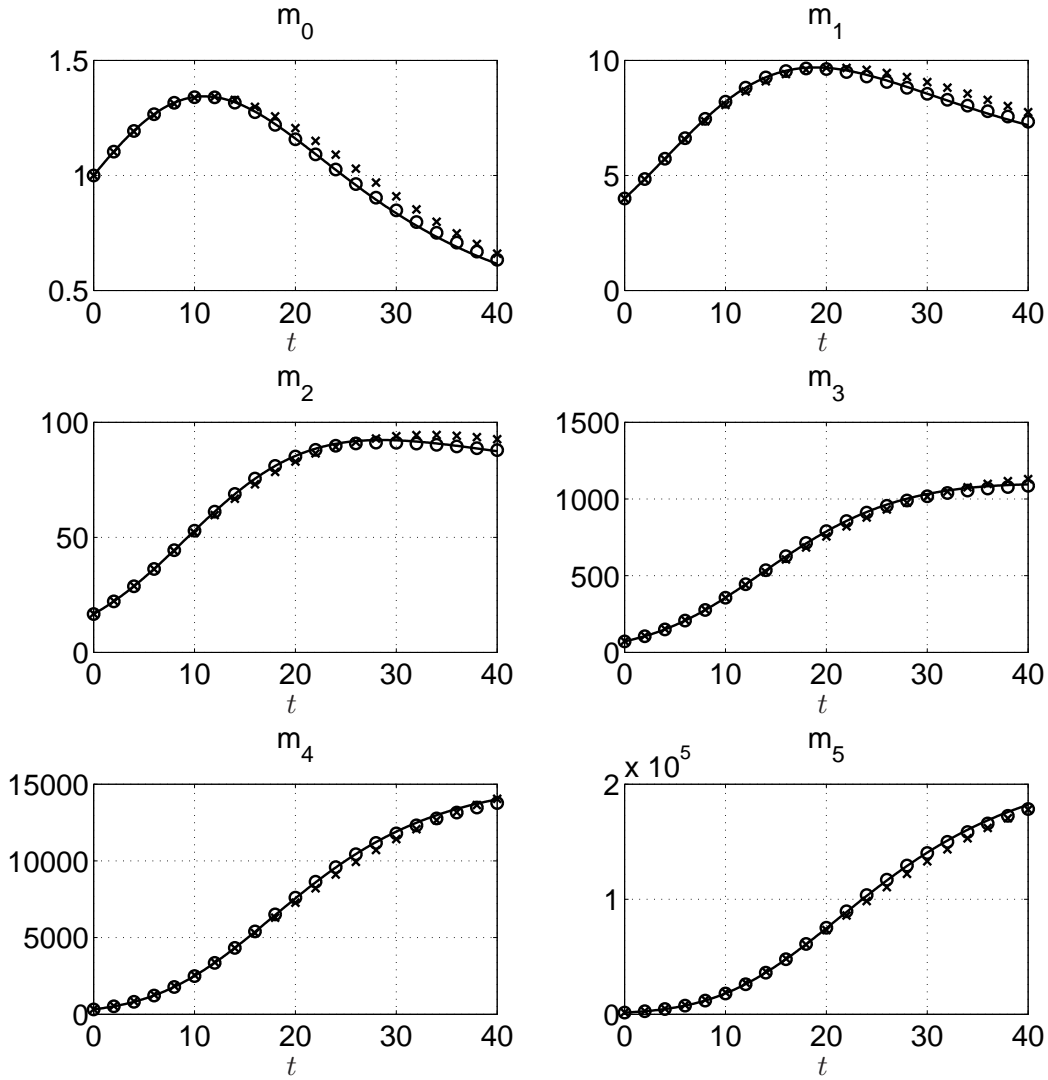
Figure 4.6.: *Moment Dynamics for sigma point approach ($\times$) compared to* QMOM ($\circ$)
and full solution with pseudo spectral collocation (solid)

relative errors

$$e_l(t) = \frac{m_l^{\mathrm{ref}}(t) - m_l^{\mathrm{approx}}(t)}{m_l^{\mathrm{ref}}(t)} \tag{4.102}$$

can be seen in Fig. 4.7. As it could be expected, both approximation approaches
show differences compared to the pseudo spectral collocation reference solution. The
QMOM approximation shows only a slight difference compared to the reference solution.
Despite not being as accurate, the approximation of the moments with the DQMOM
and sigma point abscissas is in good agreement to the reference solution and stays within
narrow error bounds ($e_l(t) < 10\%$). It can also be seen, that the error is increasing for
progressing simulation time in case of moments up to order 3.
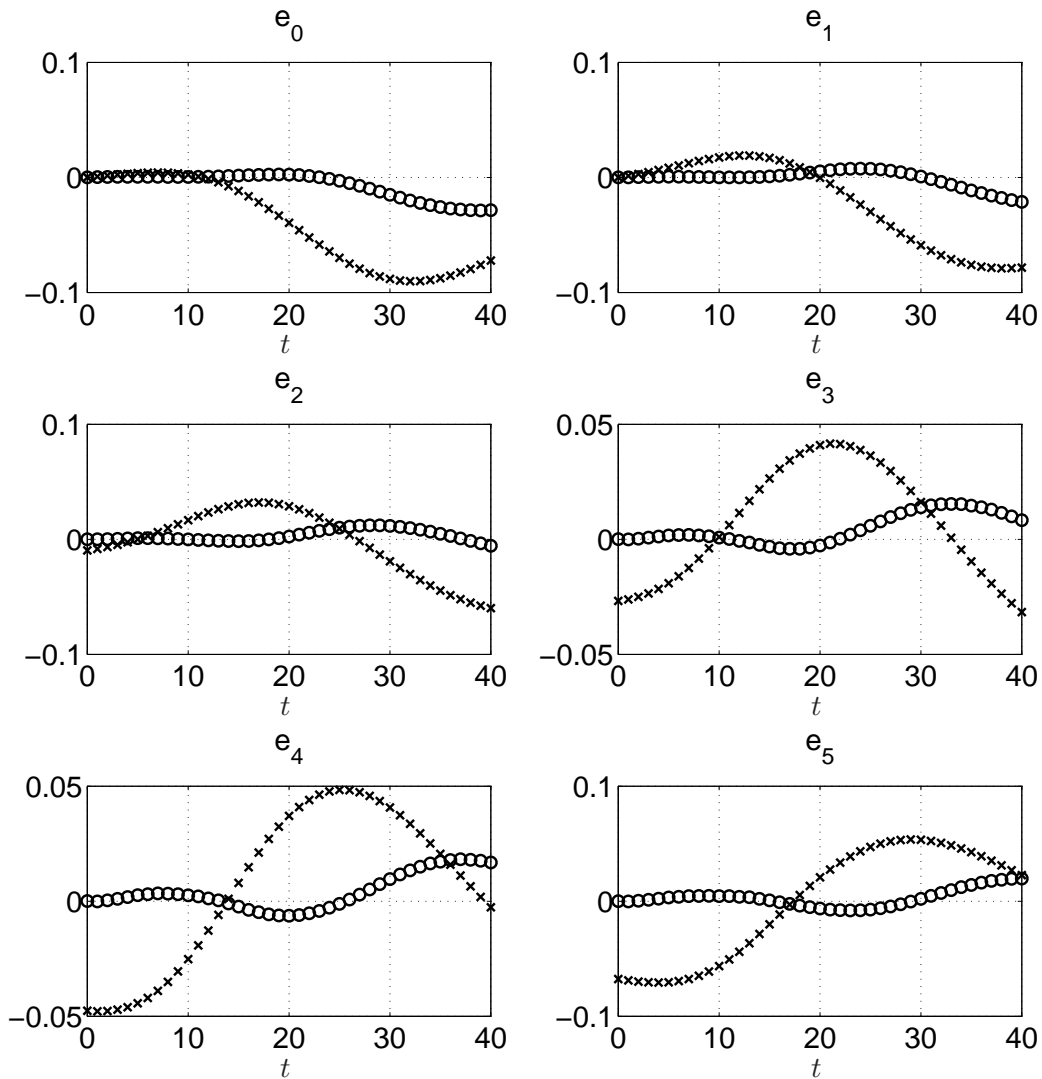
Figure 4.7.: *Error dynamics for sigma point approach (×) and* QMOM *(○) compared to full solution with pseudo spectral collocation*

4.5. Two dimensional crystallization

Next, application of the moment approximation technique is shown for a two dimensional population balance arising from the modeling of a crystallization process. Crystallization is an important class of production processes in chemical and pharmaceutical industries. It is used for manufacturing of a desired product in crystalline form from a liquid raw material [46]. Crystals are generated from the supersaturated solution and change their size and shape depending on the concentration of the solution. The crystals are heterogeneous with respect to specific properties like representative size or shape.

The structure of the following example is adopted from [58] and the example has also
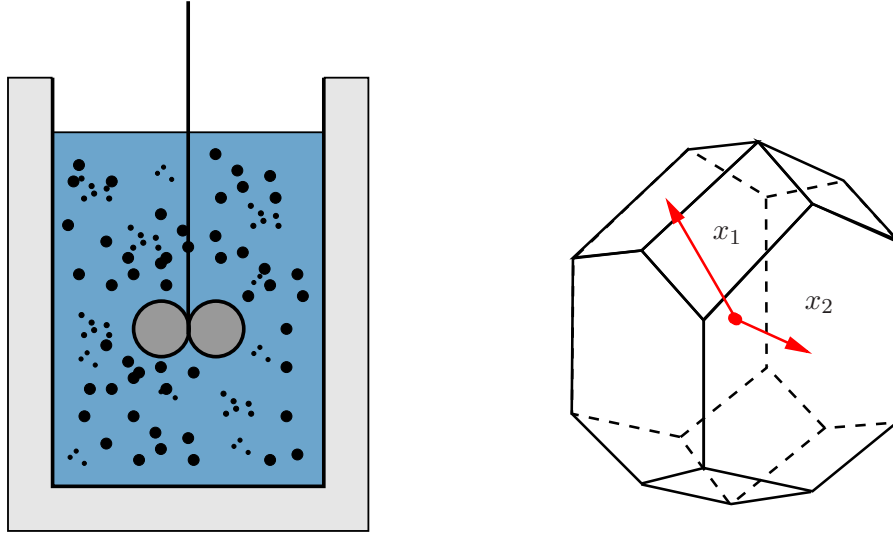
(a) *Schematic representation of batch crystal-lization in CSTR; crystals of different size (black) grow in solute (blue); temperature is kept constant by a cooling jacket (grey)*

(b) *Schematic representation of internal coor-dinates for typical crystal in lysozyme crys-tallization example with dynamics given by (4.103); x_1 and x_2 define the distance of the two representative crystal facets to the center*

Figure 4.8.: *Schematic representation of crystallization process and internal coordinates*

been presented in similar form in advance in [20]. In [58] a PBE has been derived, which describes the dynamics of the crystal morphology (or shape) for a lysozyme batch crystallization process. Therein heterogeneity of the crystals with respect to different distances of representative crystal facets from the center of the crystal is accounted for (see Fig. 4.8). Neglecting the effects of nucleation, dispersion and breakage, the temporal evolution of the crystal morphology can be characterized with the following PBE

$$
\begin{aligned}
\frac{\partial n(t, x_1, x_2)}{\partial t} = & -\frac{\partial}{\partial x_1} \left\{ G_1(t, x_1, x_2, c_{\text{Sol}}) \; n(t, x_1, x_2) \right\} \\
& -\frac{\partial}{\partial x_2} \left\{ G_2(t, x_1, x_2, c_{\text{Sol}}) \; n(t, x_1, x_2) \right\} .
\end{aligned} \tag{4.103}
$$

The growth kinetics of the facets have been modified compared to the original model formulation and are assumed to depend nonlinear on the facet lengths

$$
\begin{aligned}
G_1(t, x_1, x_2, c_{\text{Sol}}) &= k_{g,1} \; S_{\text{sat}}(t)^{p_1} \; \left(\sum_{i=0}^{3} g_{1,i} \; x_1^i \right) , \\
G_2(t, x_1, x_2, c_{\text{Sol}}) &= k_{g,2} \; S_{\text{sat}}(t)^{p_2} \; \left(\sum_{i=0}^{3} g_{2,i} \; x_2^i \right) .
\end{aligned} \tag{4.104}
$$

Table 4.2.: *Parameters and initial conditions for two dimensional crystallization example; all parameters in the dynamics of the crystal morphology distribution are scaled to the initial number of crystals*

Parameter	Value	Parameter	Value
$k_{g,1}$	$2.5 \ 10^0$	$k_{g,2}$	$4.1 \ 10^0$
p_1	$3.1 \ 10^0$	p_2	$3.3 \ 10^0$
$g_{1,0}$	$4.0 \ 10^{-1}$	$g_{2,0}$	$2 \ 10^{-1}$
$g_{1,1}$	$1.0 \ 10^{-1}$	$g_{2,1}$	$2 \ 10^{-1}$
$g_{1,2}$	$3.0 \ 10^{-1}$	$g_{2,2}$	$2 \ 10^{-1}$
$g_{1,3}$	$1.3 \ 10^0$	$g_{2,3}$	$1.4 \ 10^0$
ρ_c	$1.02 \ 10^0$	K_V	$6.0 \ 10^0$
μ	$10^{-1} \begin{pmatrix} 2.0 \\ 2.5 \end{pmatrix}$	$\mathbf{\Sigma}$	$10^{-2} \begin{pmatrix} 0.15 & 0.08 \\ 0.08 & 0.3 \end{pmatrix}$
$c_{\mathrm{Sol}}(t=0)$	$13.35 \ 10^0$		

and on the relative supersaturation

$$S(t) = \frac{c_{\mathrm{Sol}}(t) - S^*}{S^*} \,. \tag{4.105}$$

Here, S^* is the solubility which depends on the temperature of the solution. It is given by an empirical function

$$S^* = 0.0012096 \ T^3 - 0.010496 \ T^2 + 0.26159 \ T + 1.1408 \,. \tag{4.106}$$

In the studied example the reactor is assumed to be operated in an isothermal batch mode at $T = 20°C$. As no new solute is added to the reactor, the overall mass of lysozyme is constant. Thus, the concentration of the solute in the liquid phase will decrease if the solid phase (i.e. the crystal ensemble) increases its mass. In case of an ideally mixed liquid phase spatial gradients can be neglected. Hence, the corresponding dynamics of the solute concentration $c_{\mathrm{Sol}}(t)$ are given by the following ODE

$$\frac{\mathrm{d}c_{\mathrm{Sol}}(t)}{\mathrm{d}t} = - \rho_c \ K_V \int_0^\infty \int_0^\infty \left\{ G_1(t, x_1, x_2, c_{\mathrm{Sol}}) \ 8\sqrt{2} \ x_1 \ x_2^2 \right.$$
$$\left. + G_2(t, x_1, x_2, c_{\mathrm{Sol}}) \left[16\sqrt{2} \ x_1 \ x_2 - 12\sqrt{2} \ x_2^2 \right] \right\} n(t, x_1, x_2) \ \mathrm{d}x_1 \mathrm{d}x_2 \tag{4.107}$$

It is assumed, that the initial distribution of crystals can be approximated by a two dimensional Gaussian $\mathcal{N}(\mu, \mathbf{\Sigma})$, with μ and $\mathbf{\Sigma}$ being the vector of mean values and the covariance matrix, respectively. All process parameters used in the simulation are summarized in Table 4.2.

Analogous to the one dimensional example the dynamics of the abscissas and weights are given by

$$\frac{\mathrm{d}x_{1,\alpha}(t)}{\mathrm{d}t} = G_{1,\alpha}(t) \,, \quad \frac{\mathrm{d}x_{2,\alpha}(t)}{\mathrm{d}t} = G_{2,\alpha} \,, \quad \frac{\mathrm{d}w_\alpha(t)}{\mathrm{d}t} = 0 \,, \quad \alpha = 1, \ldots, N_\alpha \,. \tag{4.108}$$
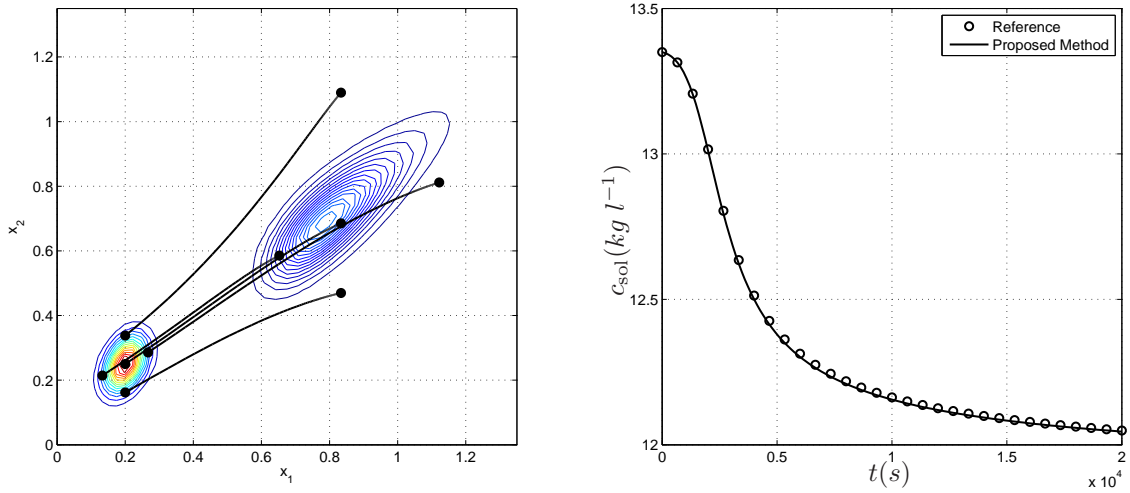
Figure 4.9.: *Left: Evolution of sigma point abscissas compared to overall evolution of $n(t, x_1, x_2)$ computed with pseudo spectral collocation; Right: Solute concentration computed with moment method compared to reference solution computed by pseudospectral collocation*

On basis of the crystals' Gaussian initial distribution the initial values for the abscissas and weights are determined from the sigma point formulas (4.90) and (4.91) with the tuning parameter $\lambda_{\mathrm{SP}} = 2$. The overall numerical effort in terms of number of ODEs is given by

$$N_{\mathrm{ODE}} = (N_d + 1) \, N_\alpha + 1 = 16 \, . \tag{4.109}$$

A reference solution for the overall number density distribution was computed using a two dimensional pseudo spectral collocation method as suggested in [11] with 64 grid points for each internal coordinate. The corresponding reference values for the integrals were computed using a two dimensional Clenshaw-Curtis quadrature, which was generated by tensoring of two one dimensional Clenshaw-Curtis quadratures. In Fig. 4.9, the temporal evolution of the abscissas can be seen in comparison to the reference solution of the number density distribution. As a matter of fact the abscissas move along the characteristic curves of (4.103). All integral quantities, like arbitrary moments or the integral characterizing the dynamics of the solute concentration (4.107) are approximated as a weighted sum of abscissas. It can be obtained from Fig. 4.9 that the solute concentration calculated from the approximation by the DQMOM shows good agreement to the reference solution. The same statement holds for the approximation of the moments. In Fig. 4.10 selected moments of the crystal number density distribution are shown for the approximation via the DQMOM and reference solution with pseudo spectral collocation, respectively. In addition, the dynamics of the corresponding relative errors are depicted in Fig. 4.11. It can be seen, that the DQMOM approximation with sigma point abscissas reproduces the moments accurately. The relative errors stay within small bounds ($e_{ij} < 10\%$). However, the error in the first order moment m_{10} and the second order moment m_{20} is increasing with increasing simulation time while e_{01} and e_{02} already seem to have reached a saturation. Additional figures of further

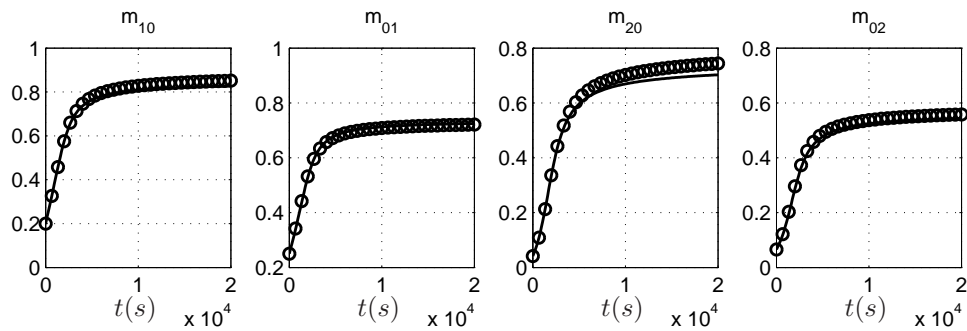Figure 4.10.: *Selected moment dynamics computed with sigma point approach compared to moments computed from solution by pseudo spectral collocation*
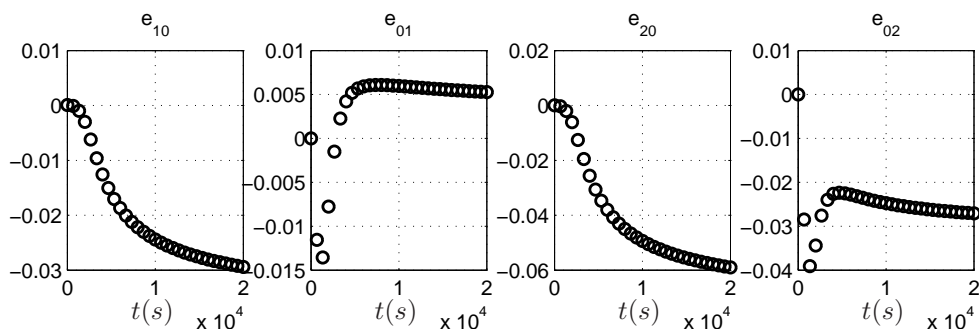

Figure 4.11.: *Relative error between moments computed with sigma point approach and full reference solution with pseudo spectral collocation*

moment dynamics and their corresponding relative error dynamics with respect to the reference solution can be found in Appendix B.

4.6. Two dimensional example with nucleation

The benchmark problems shown so far have all been dominated by the particle growth and withdrawal kinetics while nucleation has been neglected. It was derived in Section 4.2 that in case nucleation is non negligible, the DQMOM requires the solution of a linear system of equations which may cause numerical difficulties. In result, the accuracy of the approximation can not be guaranteed. To improve this situation, the corresponding PBE can be reformulated as presented in Section 4.2.5. Application of the DQMOM formalism yields the dynamics of age-distributed abscissas and weights. The performance of this approach will be demonstrated in the following for a two dimensional example.

Consider the previously presented crystallization example and assume that the super-saturation $S_{\text{sat}}(t)$ is kept constant by applying a suitable control mechanism. Hence, the dynamics of the solute concentration do not have to be accounted for anymore. Furthermore, it is assumed that the new crystal seeds are introduced to the system by a suitable dosing mechanism and that the particle size distribution of the seeds is normally distributed. The dynamics of the crystal number density distribution is then given by

$$\frac{\partial n(t, x_1, x_2)}{\partial t} = -\frac{\partial}{\partial x_1}\left\{G_1(x_1, x_2)\ n(t, x_1, x_2)\right\} - \frac{\partial}{\partial x_2}\left\{G_2(x_1, x_2)\ n(t, x_1, x_2)\right\}$$
$$+k_{\text{Nuc}}(t)\mathcal{N}(\mu, \boldsymbol{\Sigma}) \quad (4.110)$$

with the following boundary and initial conditions

$$n(t = 0, \mathbf{x}) = 0, \quad n(t, \mathbf{x} = \mathbf{0}) = 0. \quad (4.111)$$

The growth kinetics are given by

$$G_1(x_1, x_2) = \sum_{k=0}^{3} g_{1,k}\ x_1^k, \quad G_2(x_2, x_2) = \sum_{k=0}^{3} g_{2,k}\ x_2^k, \quad (4.112)$$

with parameters given in Table 4.3. As already mentioned above, it is assumed the scaled number density distribution of the seed crystals corresponds to a Gaussian $\mathcal{N}(\mu, \boldsymbol{\Sigma})$. The time variant amount of seeds is given by

$$k_{\text{Nuc}}(t) = \begin{cases} 0.1 & , \quad \text{if } 0.5 \geq t \geq 2.5 \\ 0 & , \quad\quad\quad \text{else} \end{cases} \quad (4.113)$$

To calculate the moments, the original problem is reformulated by introducing the age τ of the crystals (see Section 4.2.5). Following the extended DQMOM formalism presented in Section 4.2.5 the dynamics of age-distributed weights and age-distributed

Table 4.3.: *Parameters and initial conditions for two dimensional crystallization example with nucleation; all parameters in the dynamics of the crystal morphology distribution are scaled to the initial number of crystals*

Parameter	Value	Parameter	Value
$g_{1,0}$	$2.0 \ 10^{-2}$	$g_{2,0}$	$1.0 \ 10^{-2}$
$g_{1,1}$	$6.0 \ 10^{-2}$	$g_{2,1}$	$5.0 \ 10^{-2}$
$g_{1,2}$	$3.0 \ 10^{-1}$	$g_{2,2}$	$5.0 \ 10^{-2}$
$g_{1,3}$	$1.3 \ 10^{0}$	$g_{2,3}$	$2.5 \ 10^{-1}$
μ	$10^{-1} \begin{pmatrix} 2.0 \\ 2.5 \end{pmatrix}$	Σ	$10^{-2} \begin{pmatrix} 0.15 & 0.08 \\ 0.08 & 0.3 \end{pmatrix}$

abscissas are given by the following set of PDEs

$$\frac{\partial \widetilde{w}_\alpha(t,\tau)}{\partial t} + \frac{\partial \widetilde{w}_\alpha(t,\tau)}{\partial \tau} = 0 \,,$$

$$\frac{\partial \widetilde{x}_{1,\alpha}(t,\tau)}{\partial t} + \frac{\partial \widetilde{x}_{1,\alpha}(t,\tau)}{\partial \tau} = \sum_{k=0}^{3} g_{1,k} \ \widetilde{x}_1^k \,,$$

$$\frac{\partial \widetilde{x}_{2,\alpha}(t,\tau)}{\partial t} + \frac{\partial \widetilde{x}_{2,\alpha}(t,\tau)}{\partial \tau} = \sum_{k=0}^{3} g_{2,k} \ \widetilde{x}_2^k \,,$$

$$\alpha = 1, \ldots, N_\alpha \tag{4.114}$$

with homogeneous initial conditions

$$\widetilde{w}_\alpha(t=0,\tau) = \mathbf{0}, \qquad \widetilde{x}_{1,\alpha}(t=0,\tau) = \mathbf{0}, \qquad \widetilde{x}_{2,\alpha}(t=0,\tau) = \mathbf{0} \,. \tag{4.115}$$

The corresponding boundary conditions are computed by application of the sigma point formula (4.91) and (4.90) to the nucleation distribution $\mathcal{N}(\mu, \Sigma)$

$$\widetilde{w}_\alpha(t,\tau=0) = k_{\mathrm{Nuc}}(t) \ w_\alpha^{\mathrm{init}} \,,$$

$$\widetilde{x}_{1,\alpha}(t,\tau=0) = x_{1,\alpha}^{\mathrm{init}} \,,$$

$$\widetilde{x}_{2,\alpha}(t,\tau=0) = x_{2,\alpha}^{\mathrm{init}} \,,$$

$$\alpha = 1, \ldots, 5 \tag{4.116}$$

In summary, the approximation of the age-distributed moments requires the solution of $(N_d + 1) \, N_\alpha = (2+1) \, 5 = 15$ first order PDEs. These can be solved efficiently with a pseudo spectral collocation method [112]. In the current case, the PDEs were solved with a pseudo spectral collocation on 128 elements. A reference solution for the moments was computed from the full solution of (4.110) using a two dimensional pseudo spectral collocation on 64×64 elements.

In Fig.4.12 selected snapshots of the weight and abscissa dynamics are depicted. It can be seen, that the age-distributed weight dynamics follow the given nucleation kinetics

(4.113), which corresponds to the boundary condition (4.116). The age-distributed abscissas undergo change depending on their age τ. The moments approximated from the numerical solution of (4.114)-(4.116) are calculated by

$$m_{l_1,l_2}(t) = \int_{\mathbf{X}} x_1^{l_1} x_2^{l_2} n(t, x_1, x_2) \, \mathrm{d}x_1 \mathrm{d}x_2 \approx \int_0^\infty \sum_{\alpha=1}^{5} \widetilde{w}_\alpha(t, \tau) \, \widetilde{x}_{1,\alpha}^{l_1}(t, \tau) \, \widetilde{x}_{2,\alpha}^{l_2}(t, \tau) \mathrm{d}\tau \quad (4.117)$$

For the integration over τ, the Clenshaw-Curtis quadrature rule [112] was used. In Fig. 4.13 and 4.14 the moment dynamics and the corresponding relative errors are depicted. It can be seen that the approximation using the age-distributed formulation reproduces the moment dynamics accurately. However, for the approximation of higher order moments relative errors increase compared to the approximation of lower order moments.
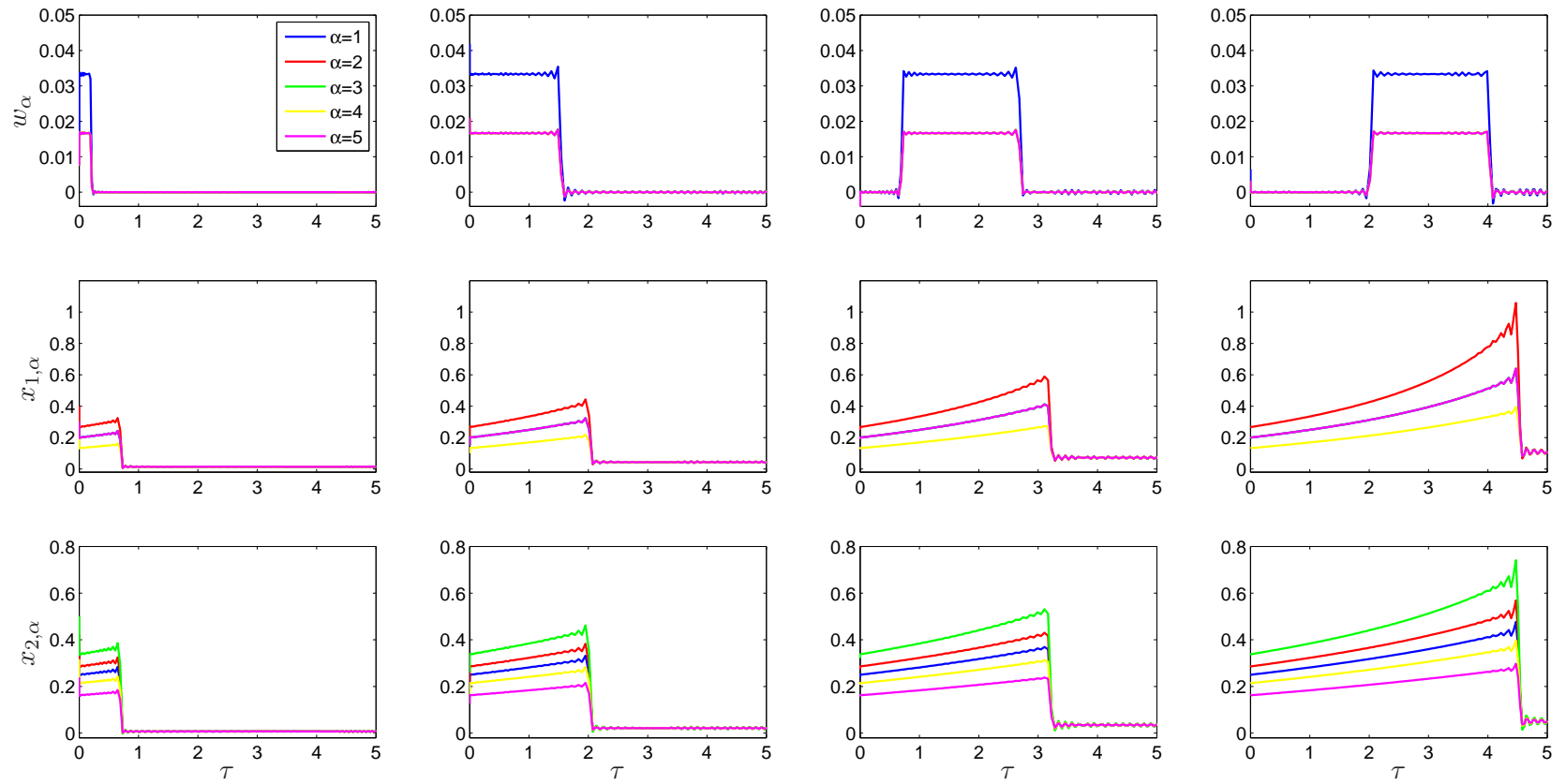
Figure 4.12.: *Selected snapshots from simulation of age-distributed abscissas and weights* (4.114) *with pseudo spectral collocation method*
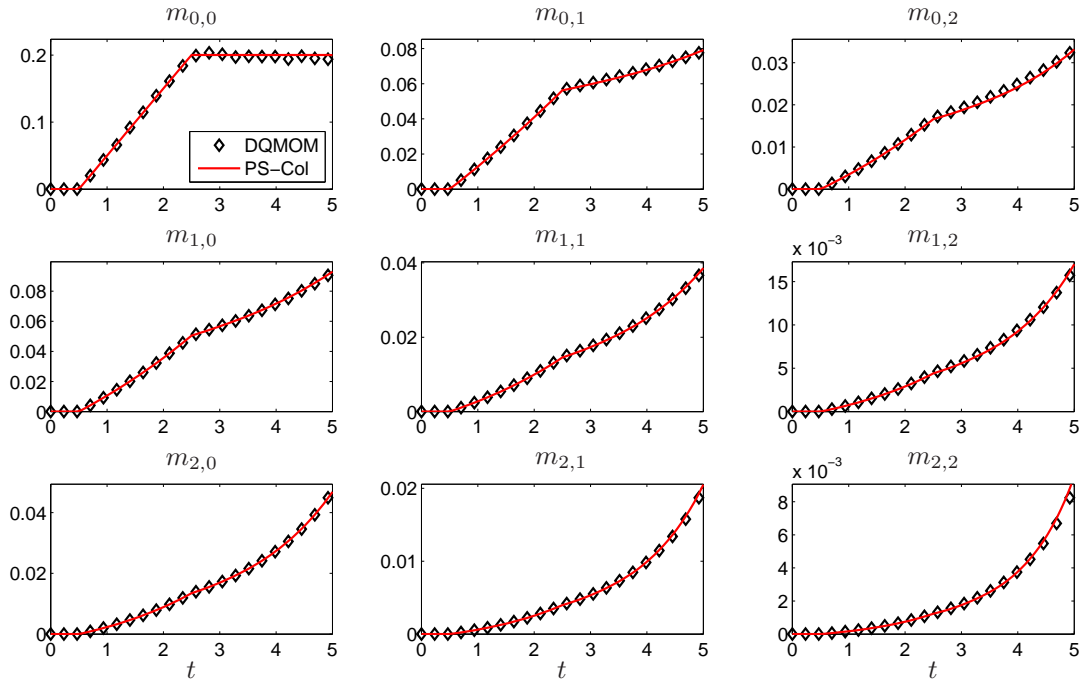
Figure 4.13.: *Moment dynamics computed from full reference solution of (4.110) with pseudo spectral collocation (solid curves) compared to approximation based on age-distributed abscissas and weights according to (4.117)*
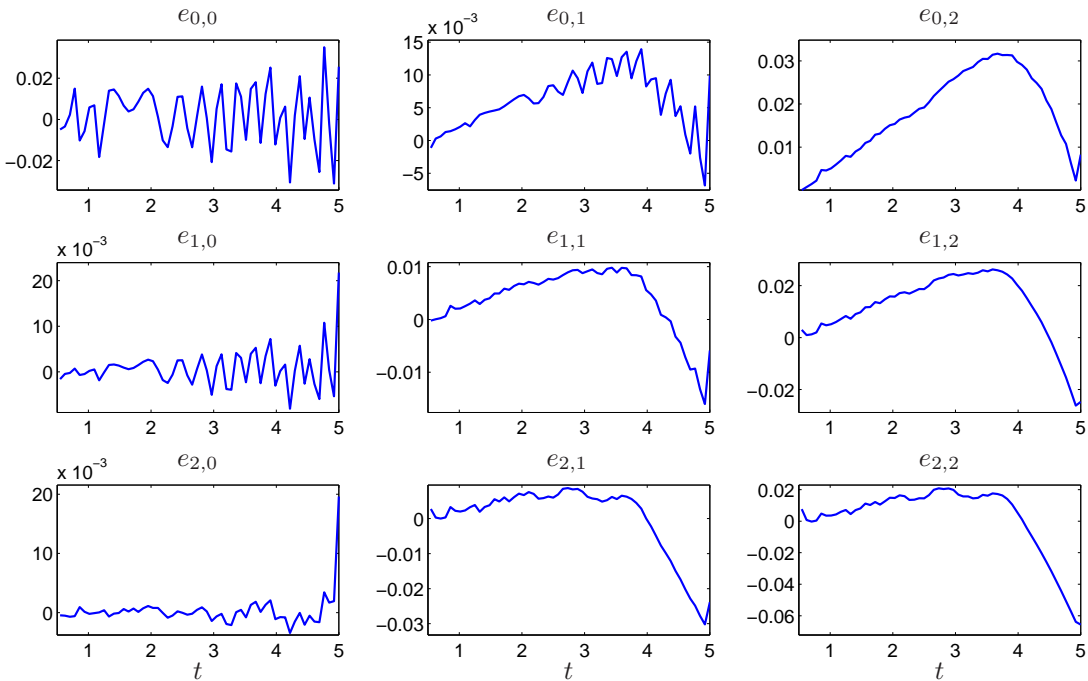
Figure 4.14.: *Relative error of approximation based age structured abscissas and weights according to (4.117) to reference solution with pseudo spectral collocation; relative errors for $t < 0.5$ can not be computed, as $m_{l_1,l_2} = 0$*

4.7. Five dimensional benchmark: A structured model of virus replication in cell cultures

In the previous sections the principle feasibility of the moment approximation technique was demonstrated. However, only one and two dimensional examples have been used. Furthermore, the assumed kinetics are relatively simple. In a preliminary work [20] a five dimensional nonlinear benchmark example was used to evaluate the performance of the moment approximation algorithm. Numerical results were compared to a reference solution which was based on a Monte-Carlo approximation of the integrals. Nevertheless, no couplings between the growth rates and other states were incorporated.

The presented moment approximation algorithm is developed for application to bottom-up population balance models of the influenza vaccine production process. For these, the corresponding high dimensional single cell dynamics are expected to be characterized by a strong coupling of the intracellular compounds (see scheme in Fig. 3.15).

For this reason in this section, performance of the presented technique will be analyzed for a more complex example which is characterized by both, coupling as well as non-linearities. The DQMOM is used to approximate the moments for a generic process model describing the spread of a viral infection within a cell culture. The model was adapted from [38]. Therein a description of the single cell viral replication dynamics is coupled to the dynamics of the cell culture on a macroscopic scale. Note that part of what follows has been published in advance [22].

In contrast to the top-down model for virus replication in cell cultures which was topic of the first part of this thesis, the intracellular replication mechanism processes is considered in more detail. The dynamic model on the single cell scale comprises the most important parts of the viral replication kinetics, while avoiding a detailed description of the subprocesses. The basic scheme is depicted in Fig. 4.15 and will be explained in the following. Virus particles bind to the surface of uninfected cells. After a virus particle has surpassed the cell membrane, the virus genome is uncoated and thus a certain amount of viral genetic information $[gen]$ is injected to the cell. From this, a template viral genomic acid $[tem]$ is produced. It delivers the blueprints for the replication of further viral genomic acid and viral structural protein $[str]$. It is assumed, that the corresponding production rates are catalyzed by intracellular enzymes v_1 and v_2. It has to be mentioned that in addition to the original model developed by Haseltine and coworkers [38, 39] a degradation of the virus template and the structural protein is modelled. Virus progeny is formed by binding of structural protein and viral genome. In a final step virus progeny is released from the cell. Balancing of the subspecies yields the following dynamics on the single cell level in terms of the following coupled system
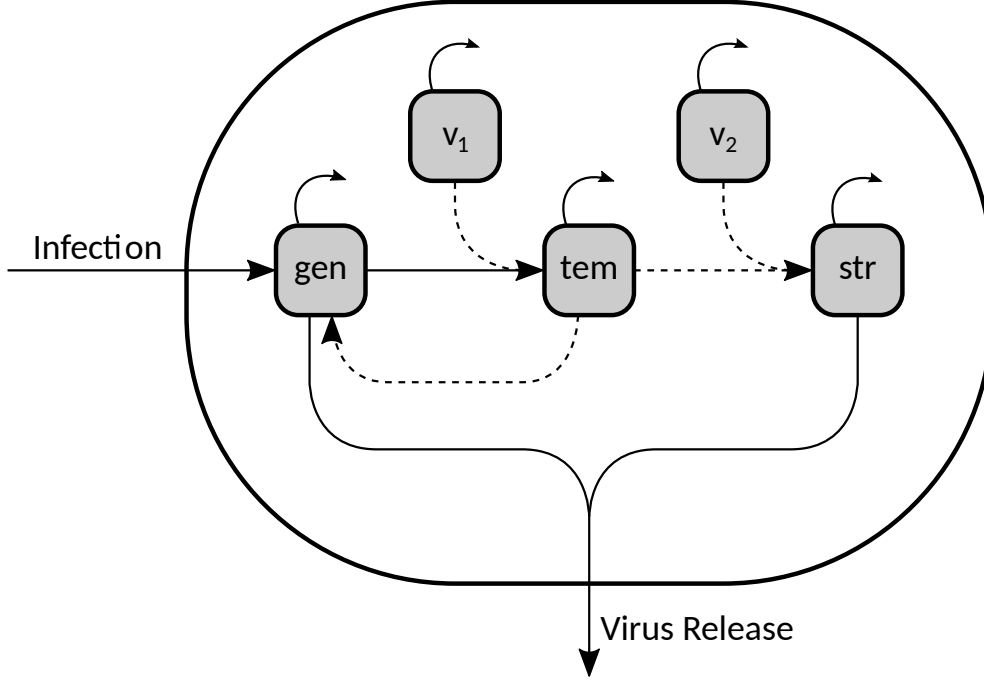
Figure 4.15.: *Scheme of the intracellular viral replication mechanism adapted from [38]; after infection, template viral genomic acid [tem] is produced from genomic acid [gen]; the production of viral structural protein [str] is catalyzed by cellular enzymes [v_1] and [v_2]*

of ODEs

$$
\dot{\mathbf{x}} = \frac{\mathrm{d}}{\mathrm{d}t}
\begin{pmatrix}
[tem] \\
[gen] \\
[str] \\
[v_1] \\
[v_2]
\end{pmatrix}
= \mathbf{h}(\mathbf{x}) =
\begin{pmatrix}
k_1 \, [v_1] \, [gen] - k_6 \, [tem] \\
k_3 \, [tem] - k_1 \, [v_1] \, [gen] - k_5 \, [gen] \, [str] - k_7 \, [gen] \\
k_2 \, [tem] \, [v_2] - k_4 \, [str] - k_5 \, [gen] \, [str] \\
f_{v_1} \\
f_{v_2}
\end{pmatrix} .
$$

(4.118)

Here, $f_{v_1} = f_{v_2} = 0$ characterize the enzyme dynamics. Alternatively, they can be used to describe different kinetic processes like enzyme production or degradation, which could be induced by cellular events like viral infection and apoptosis [22]. Furthermore, host cell resource limitation can be included choosing a suitable description of the enzyme dynamics.

To transform the description to the macroscopic scale within the framework of population balance modeling, the single cell states directly translate into internal coordinates of a corresponding PBE. The dynamics of the number density distribution of infected cells is thus given by

$$
\frac{\partial i_c(t, \mathbf{x})}{\partial t} = \underbrace{-\nabla_{\mathbf{x}} \{ \mathbf{h}(\mathbf{x}) \, i_c(t, \mathbf{x}) \}}_{\text{intracellular reactions}} + \underbrace{k_{\inf} \, U_c(t) \, V(t) \, \mathcal{I}(\mathbf{x})}_{\text{infection (nucleation)}} - \underbrace{k_{\mathrm{cd}, i_c}(\mathbf{x}) i_c(t, \mathbf{x})}_{\text{cell death}} . \quad (4.119)
$$

Here, it is assumed, that infected cells are produced by binding of free virus particles V to uninfected cells U_c. In view of the introduced definitions in Chapter 2, infection corresponds to a nucleation process. Cell-to-cell variability for newly infected cells is taken into account by spreading infected cells into the space of internal properties according to a normal distribution

$$\mathcal{I}(\mathbf{x}) = \mathcal{N}\left(\mu, \mathbf{\Sigma}\right) \tag{4.120}$$

with mean μ and covariance Σ. In result of the infection the cells undergo apoptosis and lysis. These cellular processes are summarized as "cell death" and are represented by the coefficient $k_{\mathrm{cd},i_c}(\mathbf{x})$. Its functional dependency on the vector of internal coordinates is motivated by the common assumption, that apoptosis and lysis can be induced by specific viral proteins.

In contrast to the infected cells, uninfected cells U_c are not differentiated with respect to their intracellular composition. In result, their dynamics is represented by the following ODE

$$\frac{dU_c(t)}{dt} = -k_{\mathrm{inf}}\ U_c(t)V(t) + k_{\mathrm{gro},U_c}\ U_c(t) - k_{\mathrm{cd},U_c}\ U_c(t)\,. \tag{4.121}$$

Therein, $k_{\mathrm{gro}}\ U_c(t)$ and $k_{\mathrm{cd},U_c}\ U_c(t)$ characterize the growth and cell death rates of the uninfected cells, respectively.

At the end of a successful replication, virus particles are released from the infected cells to the medium with individual rates. These depend on the intracellular state of each individual cell and is given by

$$r_{\mathrm{rel}}(\mathbf{x}) = k_5\ [gen]\ [str]\,. \tag{4.122}$$

according to the single cell dynamics.

In addition to the dynamics of the cell species, the dynamics of the virus particles in the medium has to be taken into account. Thereby, an ideally mixed medium is considered. The overall dynamics of active virus particles are thus determined by the following ordinary differential equation

$$\frac{dV(t)}{dt} = \int_{\mathbf{X}} r_{\mathrm{rel}}\left(\mathbf{x}\right)\ i_c(t,\mathbf{x})\ d\mathbf{x} - k_{\mathrm{inf}}\ U_c(t)\ V(t) - k_{\mathrm{deg},V}\ V(t) \tag{4.123}$$

where the rate coefficient k_{deg} characterizes the degradation and inactivation of free virus particles. The integral term on the right hand side represents the virus release of all individual cells. The simulation results shown in the following are based on the set of parameter values given in Table 4.4. In contrast to the formulation of Haseltine and coworkers [38], an unspecified time unit (tu) is used.

4.7.1. Single infection cycle

At first, a single infection cycle is considered. Here, it is assumed, that all cells are infected and no free virus is present in the medium at the beginning. Thus, the corre-

Table 4.4.: *Parameter values for adapted Haseltine model* (4.118) - (4.123)

Parameter	Value		Parameter	Value	
k_1	$3.13\ 10^{-4}$	$\#^{-1}tu^{-1}$	k_{inf}	$1\ 10^{-8}$	$\frac{[\text{Vol}]^2}{\text{cells virions } tu}$
k_2	$25.00\ 10^{-0}$	$\#^{-1}tu^{-1}$	k_{cd,i_c}	$5\ 10^{-3}$	tu^{-1}
k_3	$7.00\ 10^{-1}$	tu^{-1}	k_{cd,U_c}	$1.5\ 10^{-3}$	tu^{-1}
k_4	$2.00\ 10^{-0}$	tu^{-1}	k_{gro,U_c}	$3\ 10^{-3}$	tu^{-1}
k_5	$7.50\ 10^{-6}$	$\#^{-1}tu^{-1}$	$k_{\text{deg},V}$	$9\ 10^{-1}$	tu^{-1}
k_6	$1.00\ 10^{-1}$	tu^{-1}			
k_7	$1.40\ 10^{-1}$	tu^{-1}			

sponding initial conditions for uninfected and infected cells are given by

$$U_c(t=0) = 0\frac{\text{cells}}{[\text{Vol}]}, \quad I_c(t=0) = I_{c,0}\frac{\text{cells}}{[\text{Vol}]}. \tag{4.124}$$

In view of the influenza vaccine production process which was discussed in the first part of this thesis, this scenario is similar to a small scale experiment where the initial concentration of virus particles exceeds the concentration of uninfected cells by a multiple factor. In result, a high initial MOI would be obtained resulting in a quasi-instantaneous infection of all cells. Thereby, the major part of cellular heterogeneity would be an effect of initial cell-to-cell variability while heterogeneity resulting from the delayed infection process is negligible. For this reason, a similar experiment is often used to determine single cell parameters.

This setup is also very useful to analyze the performance of the presented DQMOM algorithm. As it is assumed that all cells are initially infected, the infection process

$$\underbrace{k_{\text{inf}}\ U_c(t)\ V(t)\ \mathcal{I}(\mathbf{x})}_{\text{infection (nucleation)}} = 0 \tag{4.125}$$

can be neglected. Thus the dynamics of the corresponding infected cell number density distribution is now given by

$$\frac{\partial i_c(t,\mathbf{x})}{\partial t} = -\nabla_\mathbf{x}\left\{\mathbf{h}(\mathbf{x})\ i_c(t,\mathbf{x})\right\} - k_{\text{cd},i_c}(\mathbf{x})i_c(t,\mathbf{x}) \tag{4.126}$$

and initial condition following from the above representations

$$i_c(t,\mathbf{x}) = \mathcal{I}(\mathbf{x}) = I_{c,0}\ \mathcal{N}(\mu,\ \mathbf{\Sigma}), \tag{4.127}$$

where

$$\mu = [1,\ 10,\ 1,\ 80,\ 40]^{\text{T}}, \quad \mathbf{\Sigma} = 0.05\ \text{diag}\left(\mu^2\right). \tag{4.128}$$

In the following, focus will be on the approximation of integral quantities of the multi-dimensional PBE characterizing infected cells and balancing of the free virus particles (4.123) is neglected for the single cycle infection.

Next, the accuracy of the presented moment approximation technique will be evaluated. For the given setup, the initial number of infected cells is normalized to $I_{c,0} = 1$. Thus, integral values computed in the following can be viewed as normalized quantities. Now, arbitrary integral quantities of the infected cell number density with respect to the internal coordinates can be approximated by (4.81)

$$\int\limits_{\mathbf{X}} f(\mathbf{x}) \; i_c(t,\mathbf{x}) \; \mathrm{d}\mathbf{x} \approx \sum_{\alpha=1}^{N_\alpha} w_\alpha(t) \; f(\mathbf{x}_\alpha(t)) \,. \tag{4.129}$$

As presented in Section 4.2, the dynamics of the weights and abscissas are given by

$$\dot{\mathbf{x}}_\alpha(t) = \frac{\mathrm{d}}{\mathrm{d}t} \begin{pmatrix} [tem]_\alpha \\ [gen]_\alpha \\ [str]_\alpha \\ [v_1]_\alpha \\ [v_2]_\alpha \end{pmatrix} = \mathbf{h}(\mathbf{x}_\alpha) \,, \quad \dot{w}_\alpha(t) = -k_{\mathrm{cd},i_c}(\mathbf{x}_\alpha) \, w_\alpha(t) \,, \quad \alpha = 1, \ldots, N_\alpha \,.$$

$$\tag{4.130}$$

As the initial condition is Gaussian, the sigma point formulas (4.90)-(4.91) can be applied to determine the initial values for the abscissas and weights $w_\alpha(t = 0)$ and $\mathbf{x}_\alpha(t = 0)$. The corresponding numerical effort in terms of complexity of the underlying ODE system is given by $N_{\mathrm{ODE,SP}} = (N_d + 1) \; N_\alpha$.

At first, the effect of the sigma point formula tuning parameter λ_{SP} (i.e. the parameter controlling the spread of the sigma point abscissas in state space, see Section 4.3.4) on the accuracy of the moment approximation is analyzed. When applied within the unscented Kalman filter, the tuning parameter has been reported to have a crucial influence on the performance of the resulting state estimation (see e.g. [113]). This well known fact is discussed only vaguely within the majority of contributions in the fields of unscented Kalman filtering and approximated uncertainty propagation, respectively. In most publications it is stated that means are always approximated with sufficient accuracy. To the best of the authors knowledge, no general optimization strategy for the tuning parameters exists. In [98], it is suggested to treat the tuning parameter as an additional uncertainty affecting the approximation accuracy. For the presented example the difference in the approximation is shown for four different choices of the tuning parameter λ_{SP} affecting the initial location of the $N_\alpha = 2N_d + 1 = 11$ abcissas. To determine the overall accuracy, approximation via sigma point abscissas are compared to a reference solution obtained from an approximation with the DQMOM with 10^4 random abscissas. Their initial values have been determined by an importance sampling strategy (see Section 4.3.1).

In Fig. 4.16 the approximations of the virus release rate integral

$$\int\limits_{\mathbf{X}} r_{\mathrm{rel}}(\mathbf{x}) \; i_c(t,\mathbf{x}) \; \mathrm{d}\mathbf{x} \approx \sum_{\alpha=1}^{N_\alpha} k_5 \; [gen]_\alpha \; [str]_\alpha \; w_\alpha \tag{4.131}$$

for the different choices of λ_{SP} are depicted. It is easily seen, that the initial choice of

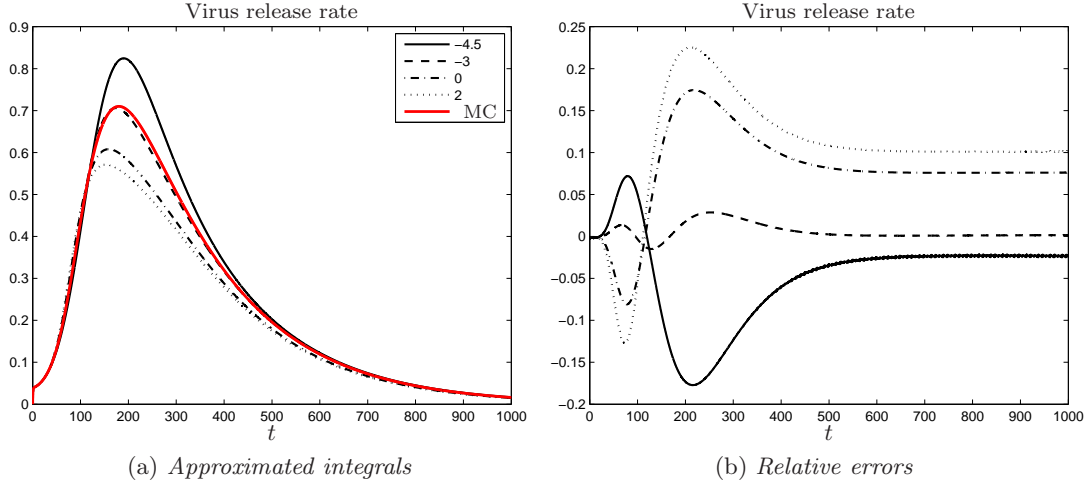(a) *Approximated integrals*   (b) *Relative errors*

Figure 4.16.: *Approximated virus release rate integrals for sigma points using different* $\lambda_{\mathrm{SP}} = -4.5, -3, 0, 2$ *values compared to Monte-Carlo approximation (MC) with* 10^4 *abscissas*

the sigma points in terms of distance from the nominal mean value μ has an significant effect on the approximation accuracy. For $\lambda_{\mathrm{SP}} = 2$, the initial abscissas have a wide spread in state space. The corresponding relative error between approximation and MC reference evaluation shows, that the relative error exceeds 0.2 at $t \approx 200\ tu$. When reducing the distance, the approximation quality improves up to a certain value around $\lambda_{\mathrm{SP}} = -3$. Here, the relative error is below 0.03 for the whole simulation time. If one further decreases λ_{SP}, the sigma points move near the mean value μ. This means, that the multi dimensional integral is approximated only in a very small local region around the mean. For the analyzed virus release integral, the approximation worsens for values below $\lambda_{\mathrm{SP}} = -3$. In fact, for $\lambda_{\mathrm{SP}} = -4.5$, the maximum absolute error is nearly as large as for $0 < \lambda < 2$. The specific choice of $\lambda_{\mathrm{SP}} = 0$ is worth an additional comment. Here, the moment approximation is not affected by the sigma point $\mathbf{x}_0$, which is located at the mean initially as the corresponding weight is given by $w_0 = 0$. Thus the approximation of moment dynamics using the sigma point abscissas is based on $2\ N_\alpha$ abscissas and weights.

However, the "optimal" value of λ_{SP} seems to depend on the specific integral expression for which an approximation is sought. A certain value of λ_{SP} resulting in the accurate approximation of one integral may not be the best choice for other integrals, which will be shown in the following. In Fig. 4.17 and Fig. 4.18, the first and second order moments with respect to the most important intracellular states $[tem], [gen], [str]$ and the corresponding relative errors are depicted. When looking at the second moment with respect to $[tem]$ the approximation with sigma points designed for $\lambda_{\mathrm{SP}} = -4.5$ shows the best accuracy. In contrast, for any other depicted moment the approach with $\lambda_{\mathrm{SP}} = -3$ outperforms the all other designs. This indicates that the initial location of the DQMOM abscissas has to be chosen with care. A "good" choice may allow an accurate approximation of one certain integral but in the meantime approximation
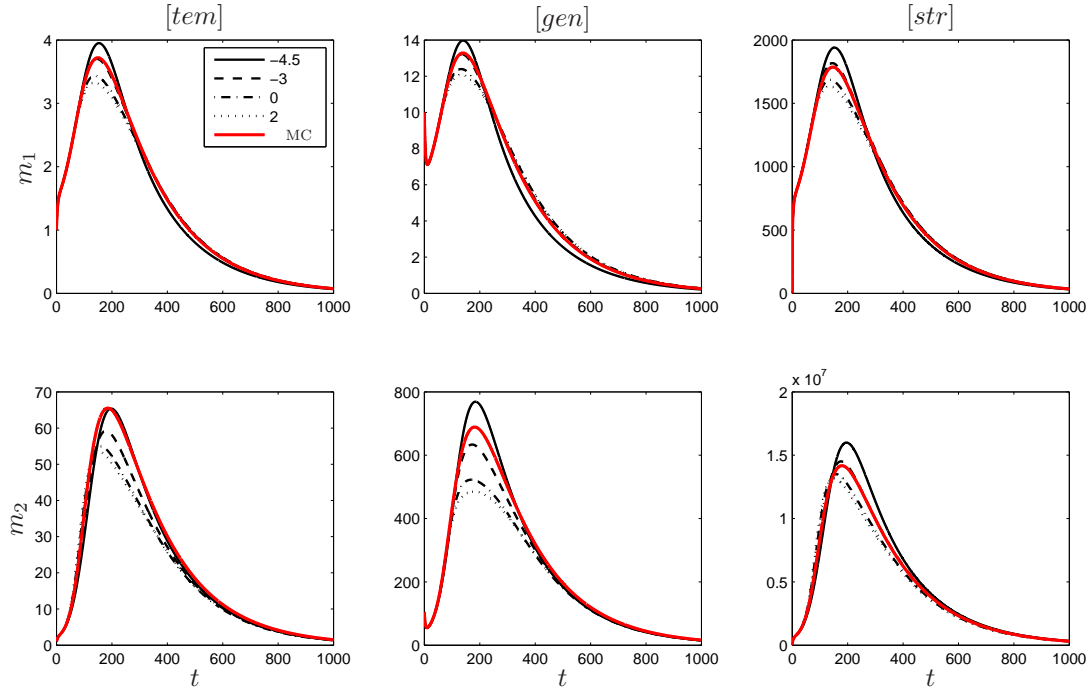
Figure 4.17.: *First and second order moments of $i_c(t, \mathbf{x})$ with respect to $[tem], [gen], [str]$ for $\lambda_{\mathrm{SP}} = -4.5, -3, 0, 2$ values compared to Monte Carlo approximation (MC) with 10^4 abscissas*

performance with respect to other integral quantities may degrade. As a rule of thumb, accuracy with respect to the most important quantities, should be ensured instead of finding a medium accurate solution for a larger class of integral quantities. For the current case, this means that the virus release integral and the mean values should be replicated accurately. Furthermore, the simulation results also imply that a good approximation of initial integrals may not automatically correspond to an accurate one for later time points. Thus, the tuning factors have to be optimized manually depending on the problem at hand. Additionally, accuracy of the approximations should be evaluated against a reference solution if possible.

All of the above mentioned properties can result in massive problems when the proposed approximate moment method is applied within a more complex framework, e.g. a process optimization or parameter estimation setup, in which the algorithm is applied successively for varying process models. Here, an underlying optimization of the tuning parameter λ_{SP} in each successive optimization or estimation step is not recommended.

To improve the situation, one may follow the alternative strategy presented in Section 4.3.5. In a first step, the initial distribution is replaced by a sum of N_{GMD} Gaussian distributions. Note, that in this connection it is not important if the initial distribution is already Gaussian or has a more general shape. For each of the sub-Gaussian distributions, individual sigma point abscissas are designed for an arbitrary "non critical" tuning parameter. Here, $\lambda_{\mathrm{SP}} = 0$ or $\lambda_{\mathrm{SP}} = 3 - N_d$ is recommended. This means for
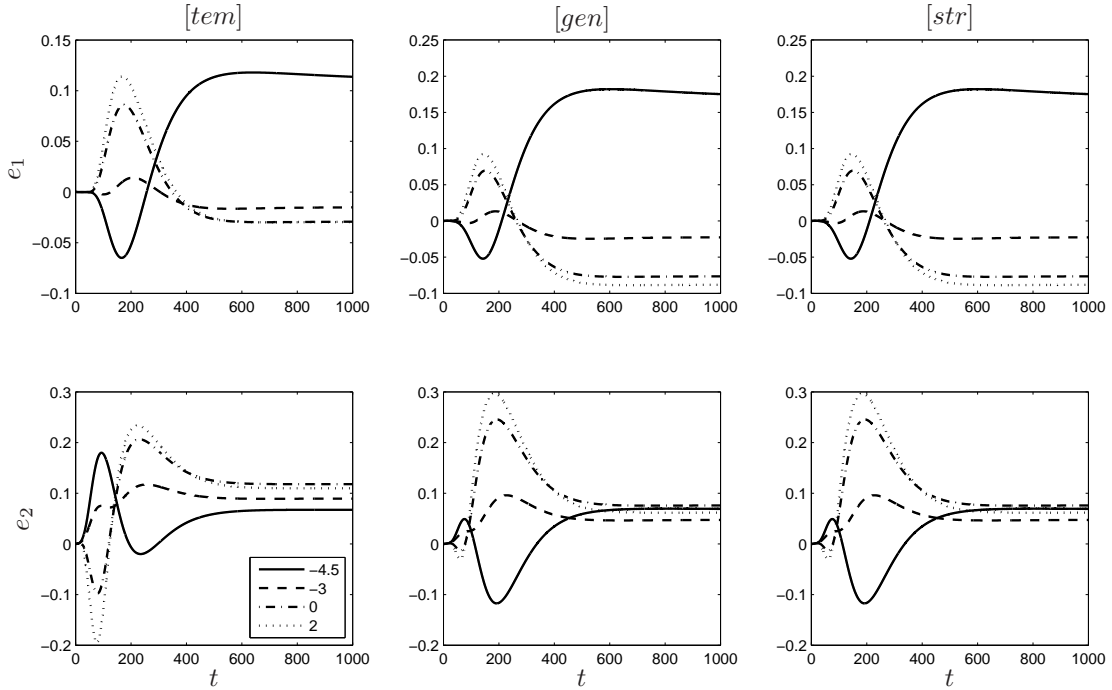
Figure 4.18.: *Relative errors of first and second order moments of $i_c(t, \mathbf{x})$ with respect to $[tem], [gen], [str]$ for $\lambda_{\mathrm{SP}} = -4.5, -3, 0, 2$*

the approximation procedure, that the problem of finding an optimal tuning parameter is traded for an increased numerical effort in terms of a larger set of abscissas. In general, a larger number of abscissas is related to an improved accuracy which is also demonstrated with the following simulation example.

Here, the approximation of the initial distribution by a sum of Gaussians was determined using the expectation maximization algorithm [74] which is implemented within the MATLAB function *fitgmdist* [1]. Alternatively, alternative strategies can be applied to reduce the number of overall abscissas (see e.g. [95]). In the following, simulation results are shown for integral approximations based on GMD approximations with increasing number of GMDs. The numerical complexity for the three shown cases are summarized as follows

$$N_{\mathrm{ODE}} = (N_d + 1) \times N_{\mathrm{GMD}} \times N_\alpha = 6 \times \{5, \ 11, \ 21\} \times 11 = \{330, \ 726, \ 1386\} \quad (4.132)$$

In Fig. 4.19 the dynamics of the overall released virus particles and the corresponding relative errors are shown. It can be seen that the accuracy of the approximation generally increases if N_{GMD} is increased. The same is observed for the moment dynamics which are depicted in Fig. 4.20. Furthermore, the corresponding relative errors are shown in Fig. 4.21. The first moments with respect to $[tem]$, $[gen]$ and $[str]$ are approximated with an excellent accuracy as the relative errors do not exceed $\pm 3\%$. In contrast, the relative error becomes quite large when using $N_{\mathrm{GMD}} = 5$. This is also seen for the approximation quality of the second order pure moments where the relative error exceeds 20% for the lowest shown N_{GMD}. An interesting effect can be seen for the

(a) *Approximated values*   (b) *Relative errors*
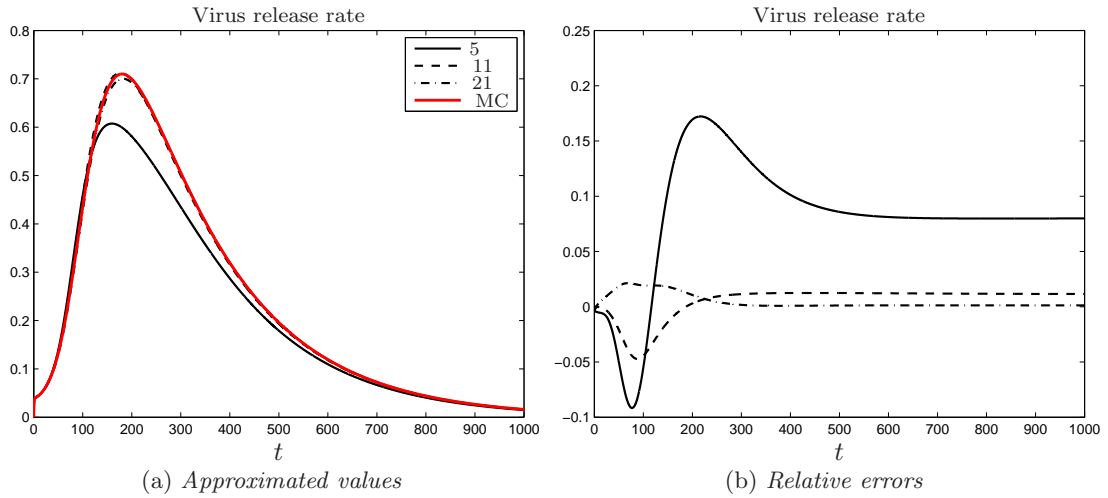
Figure 4.19.: *Approximated virus release rate integrals on basis of combined GMD-sigma-point approach compared to reference solution computed with Monte-Carlo approximation*
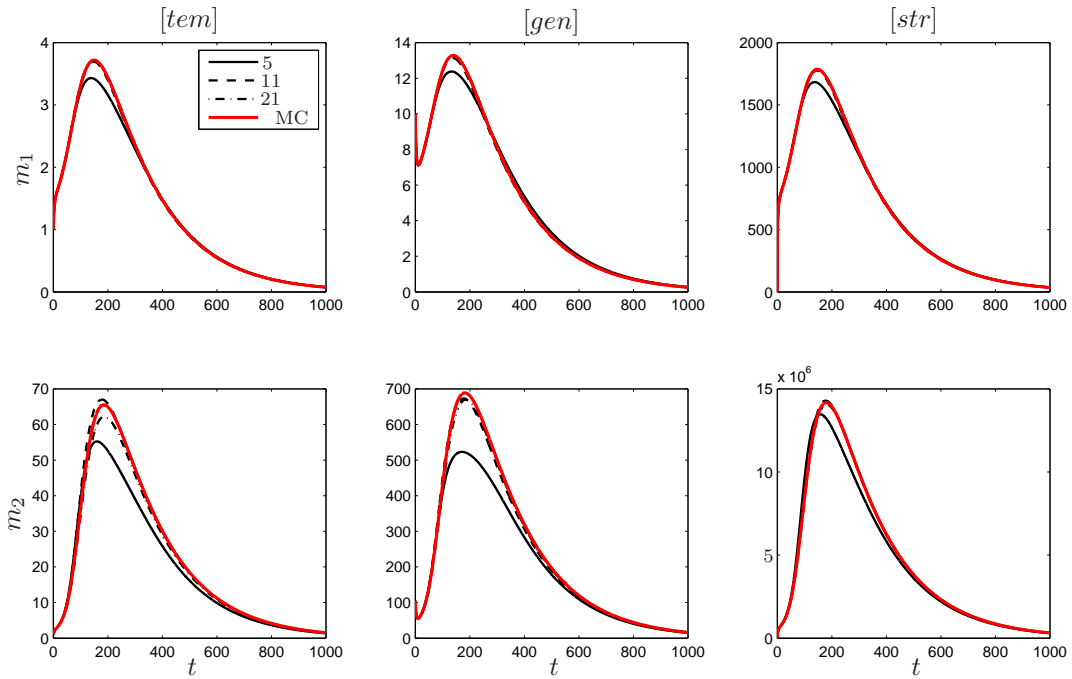
Figure 4.20.: *Moment dynamics for different N_{GMD} and comparison to Monte-Carlo approximation with 10^4 abscissas*

second order moment with respect to [*tem*]. Here, the accuracy of the approximation based on 21 GMDs is worse compared to 11 GMDs, yet the maximum error is larger for the latter one. Furthermore, the overall approximation in terms of an integral error is smaller when using 21 GMDs.

Figure 4.21.: *Relative errors of moment dynamics for different N_{GMD} w.r.t. to Monte-Carlo approximation with 10^4 abscissas*

For a very large number of Gaussians N_{GMD} the approximation is likely to converge to a solution obtained by choosing the initial DQMOM abscissas based on a Monte-Carlo sampling strategy. Expanding this thought, the following strategy can be applied to come up with a reasonable N_{GMD} to obtain an accurate approximation quality: The number of Gaussians can be 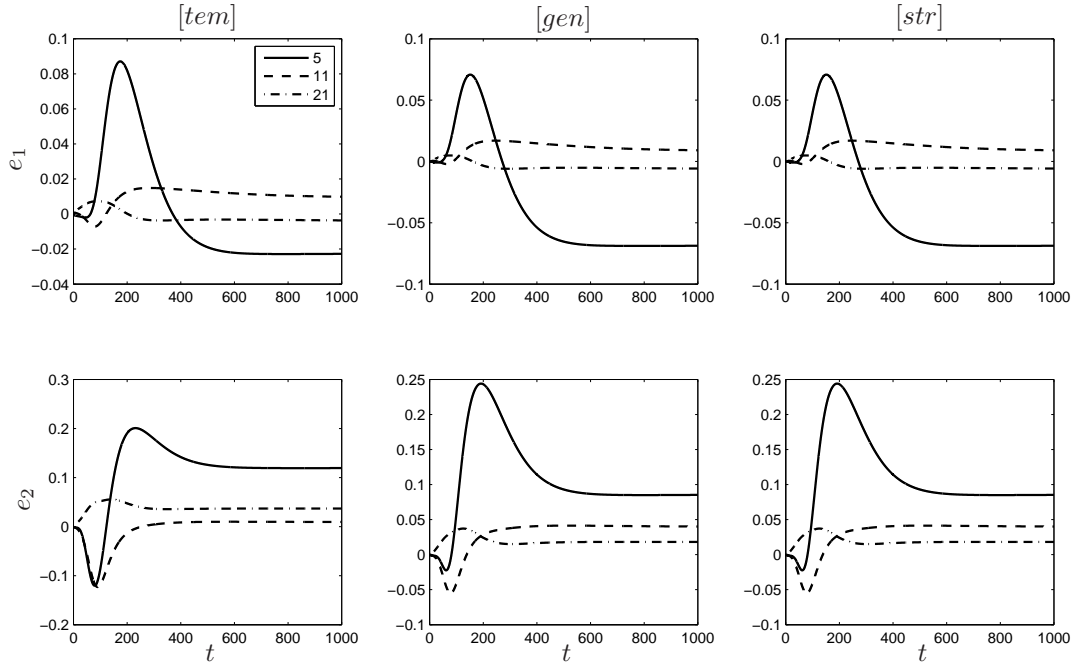increased step by step until a reasonable approximation accuracy is obtained. It is motivated by the same strategy used to find a reasonable discretization when solving PDEs numerically with the finite volume method (see e.g. [66, 34, 35]). On the one hand, an increasing number of finite volumes comes along with a better accuracy but on the other hand, the numerical effort increases. Thus, a good trade off between the two has to be found and it has to be decided for the problem at hand, if a larger number of abscissas is justified in relation to the improved accuracy which is expected.

As a final point of this section the Gaussian mixed density approach and the standard sigma point approach for the choice of the initial DQMOM abscissas will be compared to other cubature formulas. Details on all formulas are briefly summarized in the following:

1. *Mean approximation (ME):* Here the $i_c(t = 0, \mathbf{x})$ is represented by a single abscissa which is located initially at the mean μ. The corresponding initial weight is given by $w_\alpha = 1$. Thus the overall numerical effort in terms of number of ODEs is given by $N_{\mathrm{ODE,ME}} = N_\alpha(N_d + 1) = 6$.

2. *Sigma point formula (SP):* The sigma point abscissas and weights are determined by (4.90) and (4.91). The number of ODEs is given by $N_{\mathrm{ODE,SP}} = 66$. For the

present simulation, the tuning factors were chosen based on the recommendations given in [49]. Therein, for a Gaussian initial condition the following choice of tuning parameters is suggested

$$\lambda_{\mathrm{SP}} = 3 - N_d = -2\,, \quad \alpha_{\mathrm{SP}} = 1\,, \quad \beta_{\mathrm{SP}} = 0\,. \tag{4.133}$$

3. *Higher order non product formula (HONP):* The abscissas have been determined with a fifth order monomial cubature rule, which can be found in [98]. Details on the formula are given in Appendix E. Here the numerical effort scales quadratically with dimension and thus $N_{\mathrm{ODE,HONP}} = N_\alpha(N_d+1) = (2\,N_d^2+1)\,(N_d+1) = 306$.

4. *Gaussian mixed density approach (GMD):* The initial distribution is approximated by a sum of $N_{GMD} = 11$ Gaussian distributions using the *expectation maximization algorithm* in *Matlab*. Afterwards for each Gaussian sigma points are generated. In contrast to the pure sigma point approach, $\lambda_{\mathrm{SP}} = 0$ is chosen without further tuning. The overall numerical effort is given by $N_{\mathrm{ODE,GMD}} = 660$.

5. *Gaussian product formula (GA):* For the one dimensional case and $\lambda_{\mathrm{SP}} = 1$ the SP formula (4.90) enables to capture the kurtosis of a Gaussian initial condition [49]. For each internal coordinate a three point quadrature is thus derived using the sigma Point formulas. The sets of weights and abscissas are tensored into the five dimensional space to obtain the overall abscissa and weight set which was presented as *specialized product Gauss formula* in [6]. The numerical effort is given by $N_{\mathrm{ODE,GA}} = 1458$.

The reference solution was computed using Monte-Carlo approximation (MC) with $N_{\mathrm{MC}} = 10^4$ abscissas.

In a first scenario, it is assumed, that cell death does not depend on the intracellular composition. An example for this setup can be found when analyzing the effects of physical or chemical stimuli on the cell, like medium temperature, radiation or pH-level which affect all cells in the same manner. Furthermore, the enzyme levels remain constant in course of the infection process. In consequence, the corresponding rates are given by

$$k_{\mathrm{cd},i_c}(\mathbf{x}) = k_{\mathrm{cd},i_c}\,, \quad f_{v_1} = f_{v_2} = 0\,. \tag{4.134}$$

In Fig. 4.22, approximations of the overall number of released virions (4.131) are depicted for all listed approaches. It can be seen that most approaches reach a good accuracy with the GMD approach slightly outperforming the other approaches. In contrast, approximation with only one DQMOM abscissa located at the mean exhibits significant errors. In Fig. 4.23 the approximation of the infected cells zeroth order moment, i.e. the overall number of infected cells, is shown. The dynamics of arbitrary abscissa weights can be computed easily as the cell death coefficient is the same for all abscissas

$$w_\alpha(t) = w_{\alpha,0}\ \exp\left\{-k_{\mathrm{cd},i_c}\ t\right\}\,. \tag{4.135}$$

(a) *Computed integral approximations* (b) *Relative errors*

Figure 4.22.: *Approximations of virus release rate integral for different cubature formu-las and relative errors w.r.t. to Monte-Carlo approximation*

(a) *Computed integral approximations* (b) *Relative errors*
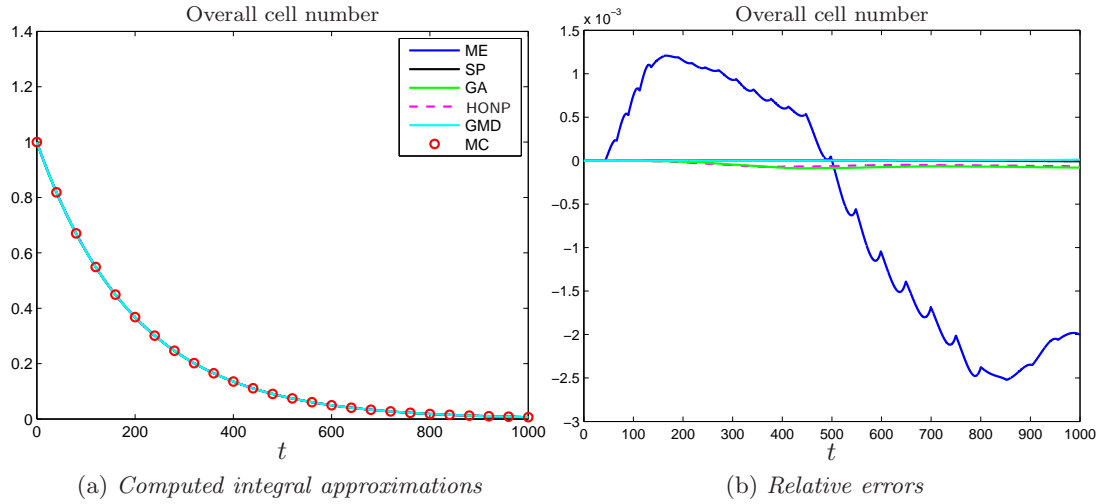
Figure 4.23.: *Overall number of infected cells for different cubature formulas and relative errors w.r.t. to Monte-Carlo approximation*

Hence the overall dynamics of the approximation is given by

$$m_{\mathbf{0}}(t) \approx \exp\left\{-k_{\mathrm{cd},i_c}\ t\right\} \sum_{\alpha=1}^{N_\alpha} w_{\alpha,0} = \exp\left\{-k_{\mathrm{cd},i_c}\ t\right\}\ I_{c,0}\,. \qquad (4.136)$$

This explains the trivial fact that, apart from some very small numerical differences which are an effect of the numerical solution of the ODEs, all approximations show the same accuracy. In contrast, the approaches show different accuracies when used for the approximation of the first and second order pure moments that are found in Fig. 4.24. Additionally, relative errors are depicted in Fig. 4.25. Similar to the approximation
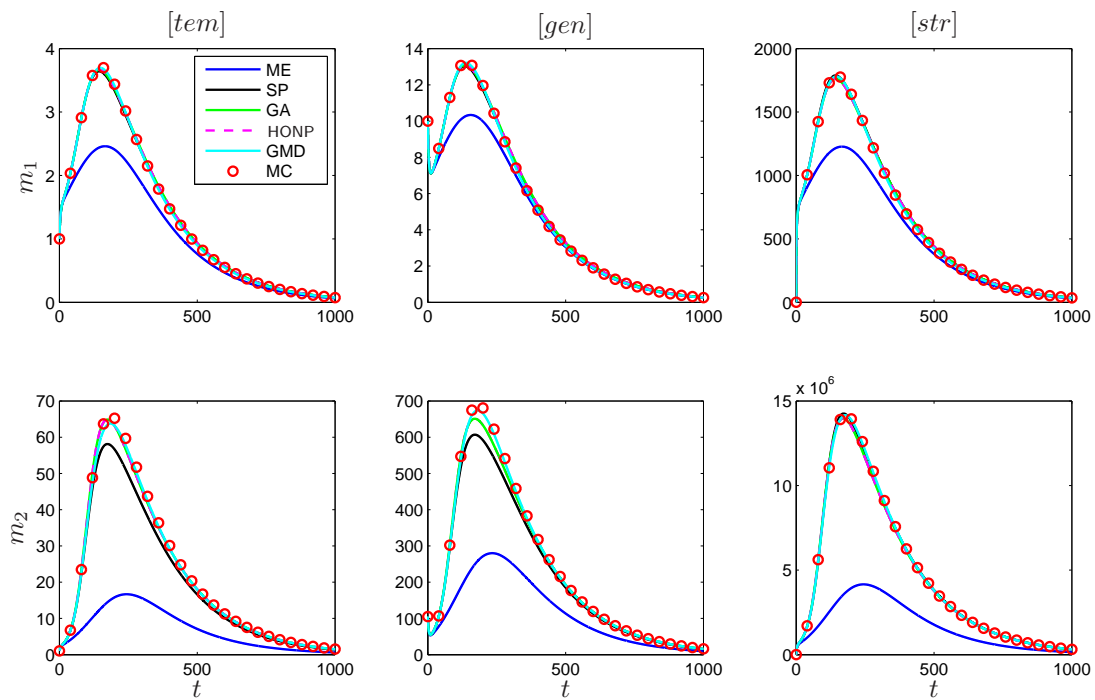
Figure 4.24.: *Approximations of first and second order moments for different approaches*
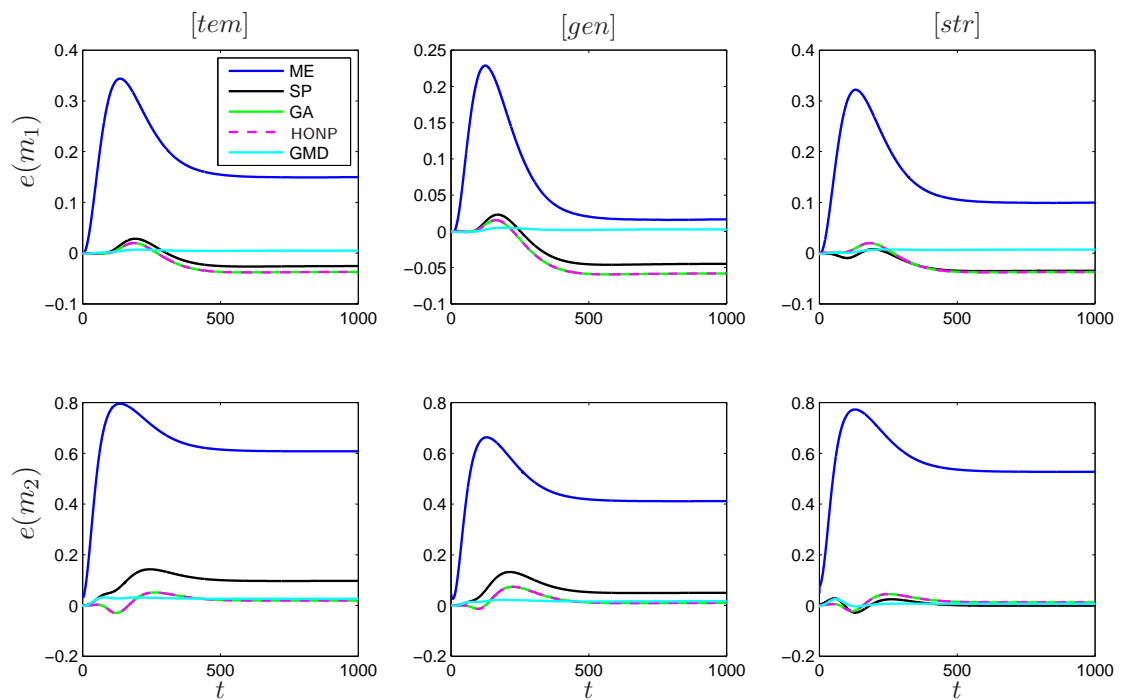
Figure 4.25.: *Relative errors for approximation of first and second order moments for different approaches w.r.t. Monte-Carlo approximation*

of the overall virus release integral, good accuracy is obtained for all approaches apart from the mean approach (highlighted with a blue curve) for the approximation of the first moments. However, the GMD approach (cyan) outperforms the other formulas with a maximum relative error around 1%. The maximum errors of the HONP, SP and GA approaches are characterized by similar relative errors. In contrast, the SP approach exhibits larger relative errors than the other approaches (apart from the ME approach). In particular, this can be seen for the second order moment with respect to $[tem]$ where the maximum relative error is around 10%. Again, the GMD approach outperforms the other approaches.

In the second scenario, it is assumed that each cell is characterized by an individual cell death rate depending on the intracellular composition

$$k_{\text{cd},i_c}(\mathbf{x}) = k_{\text{cd},i_c} \ r_{\text{rel}}(\mathbf{x}), \quad f_{v_1} = f_{v_2} = 0 \,. \tag{4.137}$$

This means, that cells which are characterized by an increased intracellular amount of viral compounds $[gen]$ and $[str]$ have a lower survival probability than cells that are characterized by a low amount. Examples for such behavior are found for example for apoptosis induction (see e.g. [23, 24]) where the amount of certain intracellular species triggers certain events of the cellular life cycle. Apoptosis is one form of programmed cell death and is for example induced as a reaction to environmental stress or to viral infection.

Approximations of the integral virus release for the different approaches can be seen in Fig. 4.26. The reference solution was again computed using Monte-Carlo integration with $N_{\text{MC}} = 10^4$ abscissas. It can be observed, that neither the standard sigma point

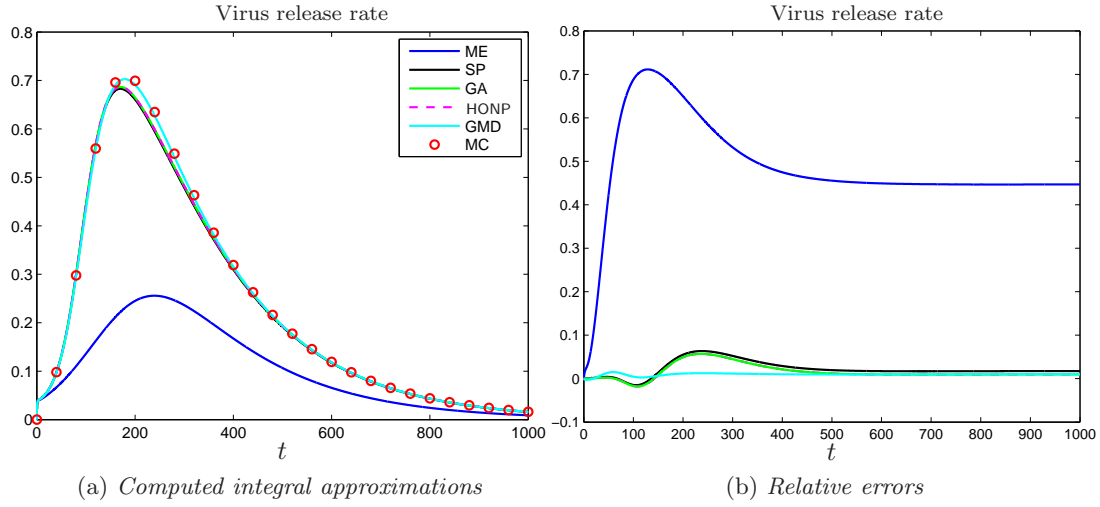(a) *Computed integral approximations*   (b) *Relative errors*
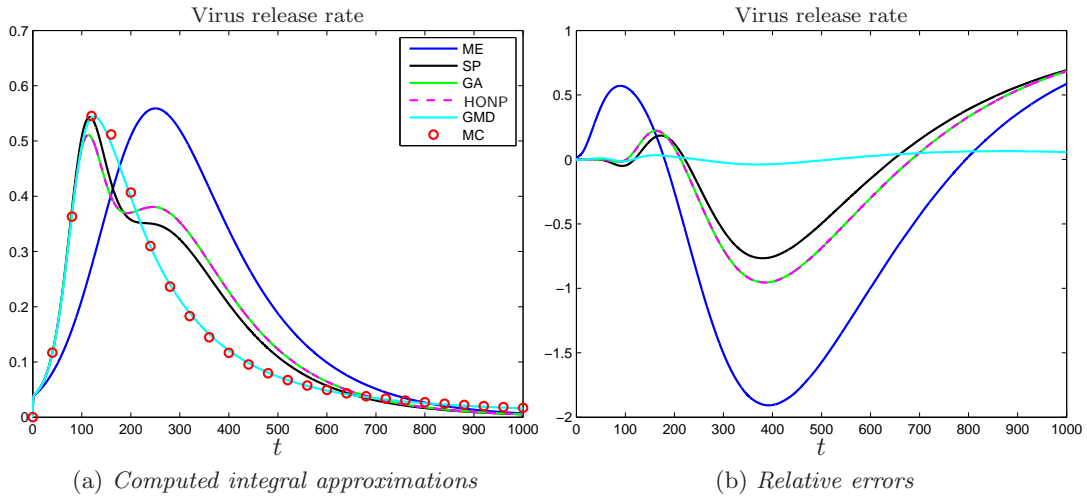
Figure 4.26.: *Approximated virus release rate integrals for different cubature formulas and relative errors w.r.t. Monte-Carlo approximation*

approach, nor the higher order non product cubature approach and the Gaussian approximation are able to approximate the overall virus release rate with a reasonable

accuracy. In contrast, the GMD approach shows good results. The corresponding rela-
tive error stays within narrow error bounds $|e| < 5$ %. The same is also observed from
Fig. 4.27 which shows the overall number of infected cells. Compared to the approxi-

(a) *Computed integral approximations*   (b) *Relative errors*
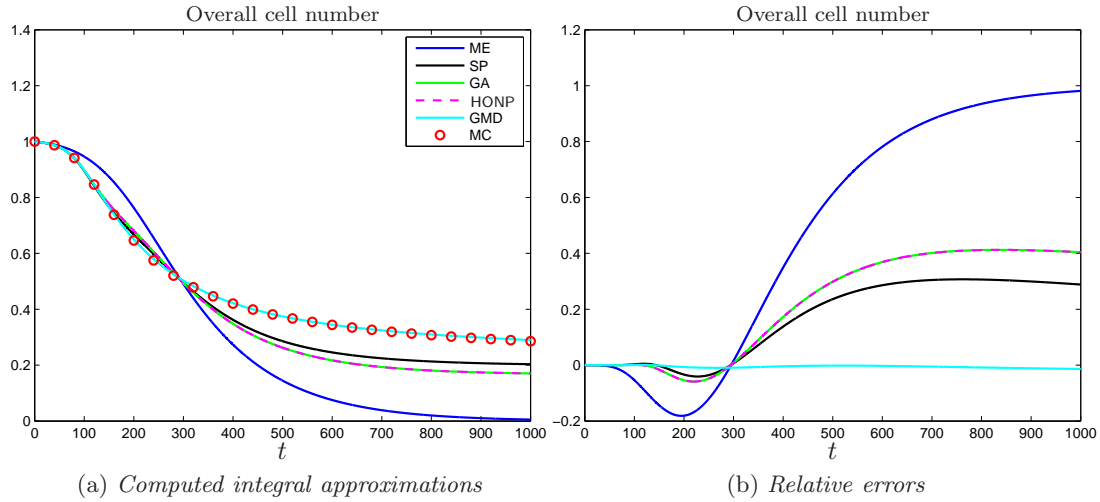
Figure 4.27.: *Overall number of infected cells for different cubature formulas and relative
errors w.r.t. Monte-Carlo approximation*

mation using the mean value as single abscissa, all other approaches show much better
performance. Again, the GMD approach outperforms the other approaches and stays
within very narrow error bounds ($|e| < 2$ %). In Fig. 4.28 and Fig. 4.29 approxima-
tions of the first and second order moments are depicted. As for the approximations
of the overall cell number and the overall virus release, all approaches except the GMD
approach result in significant approximation errors for the first order and the second
order moments.
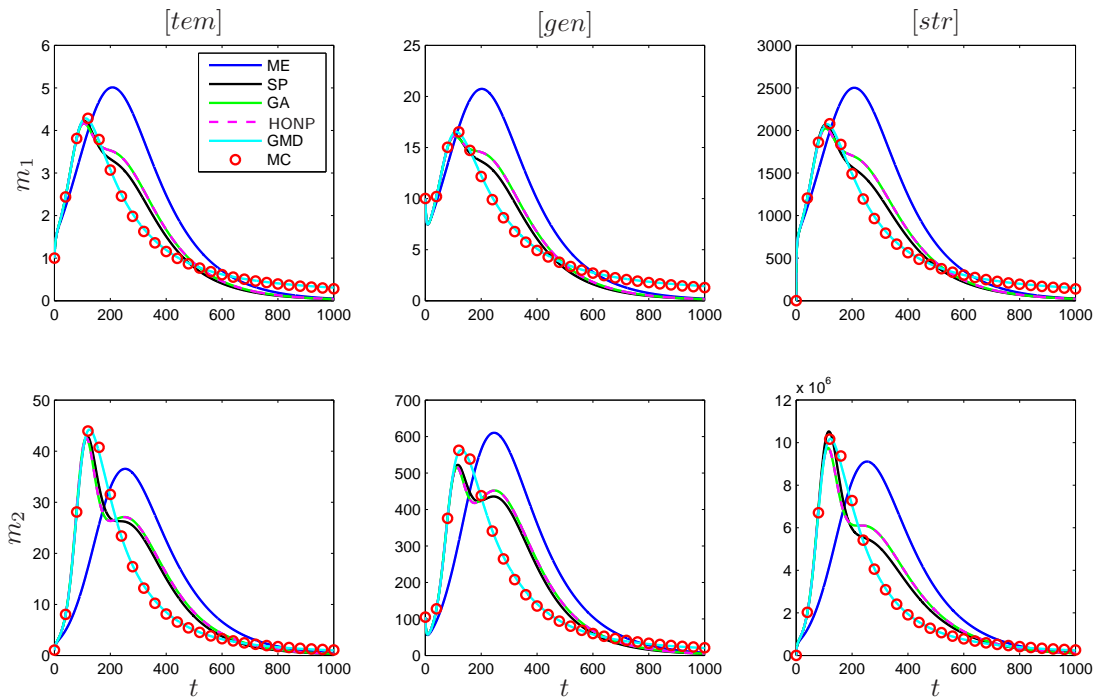
Figure 4.28.: *Approximations of first and second order pure moments for different approaches*
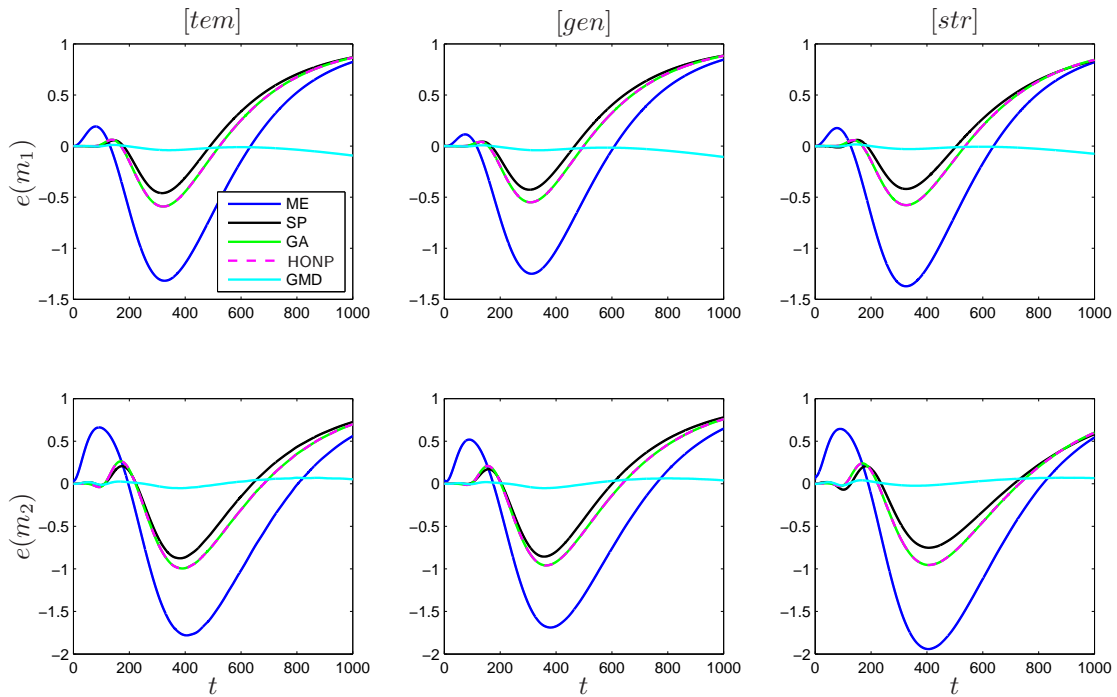
Figure 4.29.: *Relative errors for approximation of first and second order pure moments for different approaches w.r.t. Monte-Carlo approximation*

4.7.2. Multiple cycle infection

In the previous scenario, it was assumed, that all cells are infected initially. This is not the case for a large scale vaccine production process. In a real production process, the initial MOI is generally low: A low number of virus particles is used to inoculate the reactor to obtain a large harvest of replicated virus particles. In the following, the accuracy of the approximation approaches presented in the previous section will be shown for this multiple cycle infection scenario.

In contrast to the previous scenario, now initially only virions and uninfected cells are present and thus the initial conditions are given by

$$U_c(t = 0) = 10^8 \frac{\text{cells}}{[\text{Vol}]}, \quad V(t = 0) = 10^3 \frac{\text{virions}}{[\text{Vol}]}, \qquad I_c(t = 0, \mathbf{x}) = 0 \frac{\text{cells}}{[\text{Vol}]}. \quad (4.138)$$

In contrast to the single cycle infection setup, the full PBE (4.119) has to be taken into account to characterize the infected cell dynamics. To avert the expected numerical difficulties which may arise from the application of the DQMOM to problems with nucleation (see Section 4.2.3), the overall problem is reformulated as a series of initial value problems as presented in Section 4.2.5

$$\frac{\partial i_{c,k}(t, \mathbf{x})}{\partial t} = -\nabla_\mathbf{x} \{\mathbf{h}(\mathbf{x}) \, i_{c,k}(t, \mathbf{x})\} - k_{\text{cd},i_c}(\mathbf{x}) \, i_{c,k}(t, \mathbf{x}) \qquad (4.139)$$

with initial conditions given by

$$i_{c,k}(t = t_k, \mathbf{x}) = k_{\text{inf}} \, U_c(t_{k-1}) V(t_{k-1}) \, \mathcal{I}(\mathbf{x}) \, (t_k - t_{k-1})$$
$$= Inf(t_k)\mathcal{I}(\mathbf{x}) \,. \qquad (4.140)$$

Thereby, each of the initial value problems describes the dynamics of cells $i_{c,k}(t, \mathbf{x})$ which have been infected in the interval $[t_{k-1}, \, t_k)$. Comparing the obtained series reformulation to the single cycle infection formulation (4.126), it can be seen that the reformulated multiple cycle infection corresponds to finite number of single cycle infection scenarios. The number of subpopulations N_k results from the chosen temporal discretization

$$t = [t_0, \, t_1, \ldots, \, t_{k-1}, \, t_k, \, t_{k+1}, \ldots, t_{N_k}] \,. \qquad (4.141)$$

Here, a equidistant grid with $(t_k - t_{k-1}) = 0.5 \ tu$ and $N_k = 2000$ was used. Integral quantities (i.e. moments, overall virus release rate etc.) can be approximated by

$$\int\limits_\mathbf{X} f_i(\mathbf{x}) i_c(t, \mathbf{x}) \mathrm{d}\mathbf{x} \approx \sum_{k=1}^{N_k} \int\limits_\mathbf{X} f_i(\mathbf{x}) i_{c,k}(t, \mathbf{x}) \mathrm{d}\mathbf{x} \,. \qquad (4.142)$$

At this point it has to be emphasized that the intracellular virus kinetics are decoupled from the extracellular states (i.e. the single cell dynamics of the infected cells (4.118) do not depend on $U_c(t)$ and $V(t)$). Under these circumstances, the overall numerical procedure can be further simplified [39]. Instead of solving the full problem for the discrete

reformulation, the single cycle infection scenario is solved only once for a normalized initial condition

$$\frac{\partial i_c^*(t^*, \mathbf{x})}{\partial t^*} = -\nabla_{\mathbf{x}} \left\{ \mathbf{h}(\mathbf{x}) \; i_c^*(t^*, \mathbf{x}) \right\} - k_{\mathrm{cd},i_c}(\mathbf{x}) \; i_c^*(t^*, \mathbf{x})$$
$$i_c^*(t^* = 0, \mathbf{x}) = \mathcal{I}(\mathbf{x}) = \mathcal{N}(\mu, \boldsymbol{\Sigma}). \tag{4.143}$$

Arbitrary integral quantities can be approximated using the moment approximation algorithm presented previously

$$\int_{\mathbf{X}} f_i(\mathbf{x}) i_c^*(t^*, \mathbf{x}) \; \mathrm{d}\mathbf{x} \approx \sum_{\alpha=1}^{N_\alpha} f_i(\mathbf{x}_\alpha^*(t^*)) w_\alpha^*(t^*). \tag{4.144}$$

Afterwards each of those is multiplied with the corresponding initial condition, i.e. the number of newly infected cells in the interval $[t_{k-1}, \; t_k)$ (4.140) to obtain integral approximations for the subpopulations $i_{c,k}(t, \mathbf{x})$.

$$\int_{\mathbf{X}} f_i(\mathbf{x}) i_{c,k}(t, \mathbf{x}) \; \mathrm{d}\mathbf{x} = Inf(t_k) \int_{\mathbf{X}} f_i(\mathbf{x}) i_c^*(t - t_k, \mathbf{x}) \; \mathrm{d}\mathbf{x}$$

$$\approx Inf(t_k) \sum_{\alpha=1}^{N_\alpha} f_i(\mathbf{x}_\alpha^*(t - t_k)) w_\alpha^*(t - t_k) \tag{4.145}$$

Thus, the virus dynamics (4.123) can be written as

$$\frac{\mathrm{d}V(t)}{\mathrm{d}t} = \sum_{k=1}^{N_k} \int_{\mathbf{X}} r_{\mathrm{rel}}(\mathbf{x}) i_{c,k}(t, \mathbf{x}) \; \mathrm{d}\mathbf{x} - k_{\mathrm{inf}} \; U_c(t) \; V(t) - k_{\mathrm{deg}} \; V(t)$$

$$= \sum_{k=1}^{N_k} Inf(t_k) \sum_{\alpha=1}^{N_\alpha} r_{\mathrm{rel}} \left(\mathbf{x}_\alpha^*(t - t_k) \right) w_\alpha^*(t - t_k) - k_{\mathrm{inf}} \; U_c(t) \; V(t) - k_{\mathrm{deg}} \; V(t). \tag{4.146}$$

In the following, the same dependency of the cell death rate on the intracellular components as in the second single cycle infection scenario is assumed

$$k_{\mathrm{cd},i_c}(\mathbf{x}) = k_{\mathrm{cd},i_c} r_{\mathrm{rel}}(\mathbf{x}) = k_{\mathrm{cd},i_c} k_5 [gen][str]. \tag{4.147}$$

The overall uninfected and infected cell concentration dynamics are depicted in Fig. 4.30. It can be seen, that the concentration of infected cells increases significantly after a certain delay of round about 300 tu. This is a direct result of the low initial MOI: At the begin only a small number of uninfected cells gets infected and the infection takes a certain time to spread through the whole system. On the single cell scale, the viral replication mechanism and the resulting virus release rate (see Fig. 4.26) additionally contribute to the delay.

It can be seen that all cubature formulas show nearly the same results for the uninfected cells, only the approximation which is based on a sole cubature abscissa at the mean
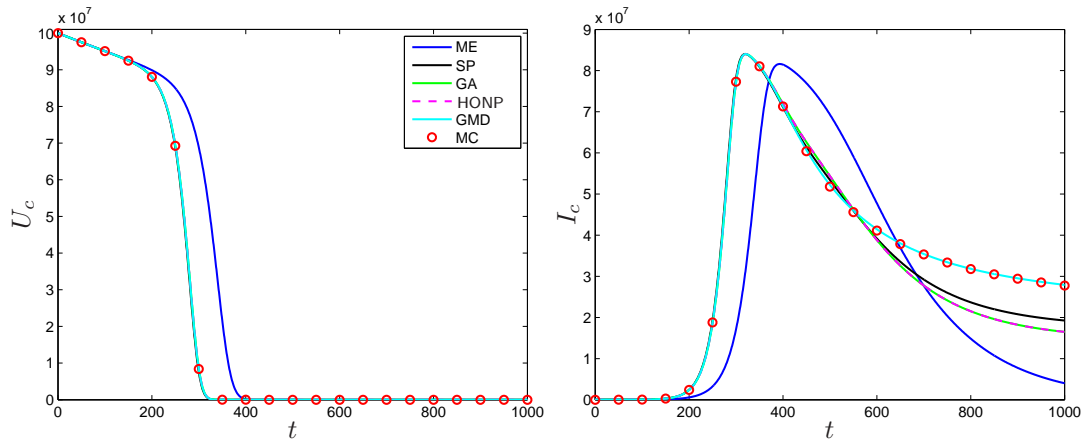
Figure 4.30.: *Multiple cycle infection: Dynamics of cellular species for different cubatures*
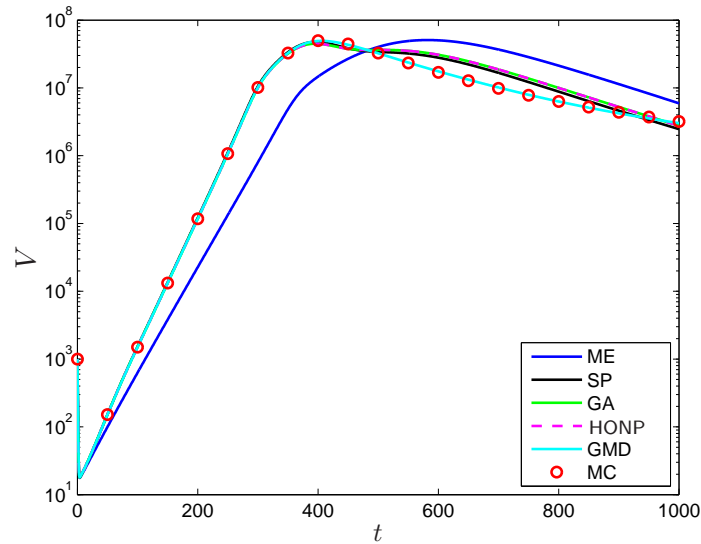
Figure 4.31.: *Multiple cycle infection: Virus concentration for different cubature formulas*

of the initial distribution yields significant errors. However, differences are obtained for the dynamics of the overall number of infected cells. Again, the mean abscissa approximation shows the worst performance. Other approximation approaches, based on the standard sigma points, the Gaussian cubatures and the higher order non-product formula are at least sufficiently accurate for $t < 600\ tu$, but their approximation quality worsens for larger simulation times. Again, the performance of the Gaussian mixed density approach stands out and stays very close to the reference computed with a large number of random abscissas. These statements on the approximation accuracy are also valid for the overall virus concentration dynamics which are shown in Fig. 4.31 Excluding the mean abscissa approach, all approaches show a good performance for $t < 500\ tu$ but significant errors emerge for larger simulation times where only the
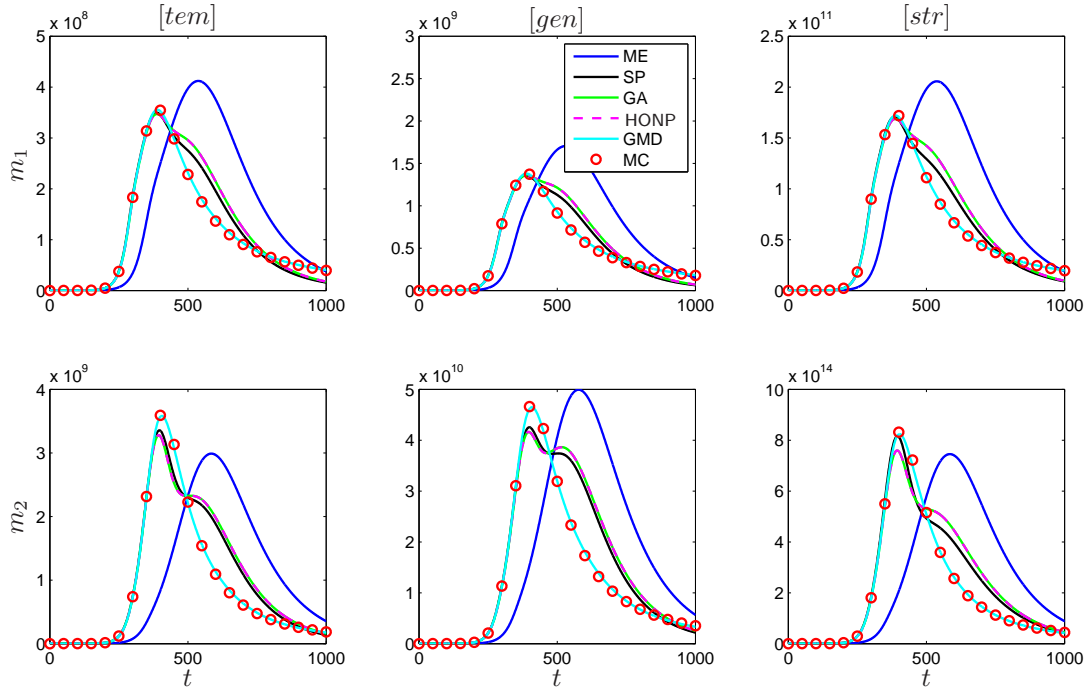
Figure 4.32.: *Multiple cycle infection: Pure moments of the distribution for different cubatures*

Gaussian mixed density approach gives accurate results. The same is observed for the first and second order pure moments which are depicted in Fig. 4.32. It can be seen that the approximations based on standard sigma points, the Gaussian cubature or the higher order non product cubature formula are accurate up to $t \approx 400\ tu$ but show significant errors for later time points. Again it can be seen, that the GMD stays very close to the reference solution and provides an accurate approximation.

4.8. Discussion

In the present chapter, it has been shown that a sufficient accuracy of multi dimensional moments can be achieved using a combination of the DQMOM and special cubature rules. For one and two dimensional benchmark problems without nucleation, the application of the sigma point cubature rule provides sufficient accuracy, yet the choice of "optimal" sigma points is not clear. For more complex dynamics and a higher number of internal coordinates, respectively, the application of the approach using the standard sigma point formula can result in a poor approximation performance. Alternative rules can be applied as shown for a five dimensional benchmark. Here, the Gaussian mixed density approach shows superior performance. It can be viewed as a direct extension of the sigma point approach: the problem of choosing a good sigma point distribution is omitted by cost of using more abscissas. For an increasing number N_{GMD} of Gaussian mixed densities, the overall approximation approaches the corresponding Monte-Carlo

approximation. Thus by successively increasing the number of applied GMDs, a good approximation accuracy can be easily evaluated. In the current example, trading off computational expense and accuracy, $N_{\mathrm{GMD}} = 11$ has shown the best performance. Systems with non-negligible nucleation can also be reformulated as a series of initial value problems (see Section 4.2.5) or as an age structured system as presented in Section 4.2.5, respectively. For a two dimensional example it was shown that the age-distributed reformulation of DQMOM can be used to accurately approximate two dimensional moments when the initial abscissas are chosen with the sigma point formula. However, for higher dimensional problems and more complex dynamics of the internal coordinates a larger number of abscissas may be necessary to obtain a sound approximation. As the age-distributed reformulation of DQMOM requires the numerical solution of $(N_d+1)N_\alpha$ one dimensional PDEs, a computationally efficient implementation is crucial.

5. Bottom-up modeling of influenza virus replication in cellular systems

Following the argumentation of Chapter 3 a bottom up model can be formulated based on a detailed description of the involved single cell dynamics to overcome the limited interpretability of the top down model.

In [41] a detailed single cell model characterizing the dynamics of the virus replication mechanism in the cells was presented. A slightly modified single cell model was later used within an age structured model to describe viral replication in cell cultures to analyze the effects of direct acting antivirals [40]. This approach is based on the crucial assumption that all cells with the same "age", i.e. time post infection, have the same internal state. Further cell-to-cell variability resulting from stochasticity on the gene expression level, non-synchronicity, or genetic modifications is neglected. These variances may however have significant effects on the overall process and thus require detailed analysis.

In this chapter, an extension of previous modeling approaches will be presented which is able to cope with cell-to-cell variability by means of heterogeneous production rates. It will be investigated, how different degrees of cell-to-cell variability affect the overall virus yield of unmodified and modified cell populations. For the solution of the emerging multi dimensional PBEs the previously presented approximate moment method is applied. It has to be mentioned that part of what follows is topic of a joint publication with Tanja Laske and Mandy Bachmann from the BPE group of the Max Planck Institute Magdeburg [18].

5.1. The single cell model of influenza virus replication

In the single cell model the interactions of a large number of viral components are considered. The basic scheme is depicted in Fig. 5.1. At this point, the full presentation of the mathematical model will be skipped, but a short overview of the intracellular viral reproduction process shall be given. For a detailed description see [41]. All model equations and parameters can also be found in Appendix F.

Free virus particles V^{Ex} attach to the cell surface at binding sites with either high or low affinity. Attached virus particles (V^{Att}_{hi} and V^{Att}_{lo}) either detach again or they are absorbed via endocytosis resulting in the enclosure of the virus particle V^{En} by an intracellular endosome. After fusion of the viral envelope with the endosomal mem-
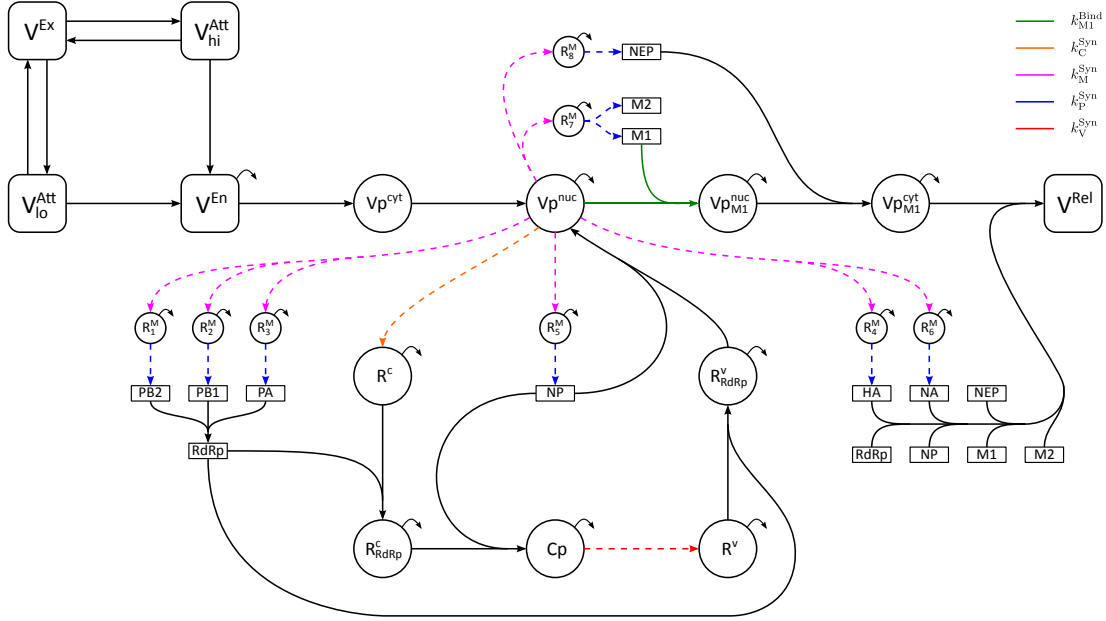
Figure 5.1.: *Basic scheme of the single cell kinetics with affected reaction steps according to [41]; dashed lines indicate translational and transcriptional steps whereas solid lines indicate binding reactions*

brane the segmented viral genome is released into the cytoplasm in the form of eight viral ribonucleoprotein complexes (vRNPs). In Fig. 5.1 these vRNPs are denoted as Vp^{cyt}. Subsequently, they are imported into the nucleus (denoted as Vp^{nuc}) where transcription into messenger RNAs (mRNAs denoted as $\text{R}^{\text{M}}_{1\ldots8}$) and complementary RNAs (cRNAs, denoted as R^{C}) takes place. Viral mRNAs migrate to the cytoplasm where they are translated into viral proteins PA, PB1, PB2, HA, NP, NA, NEP, M1, and M2 of which the first three form the RNA-dependent RNA polymerase RdRp. In the nucleus cRNAs are stabilized by consecutive attachment of RdRp and nucleoprotein NP forming cRNPs (Cp). In a second step freshly produced cRNPs are used to synthesize new vRNA molecules, which are also encapsidated by RdRp and NP forming stabilized progeny vRNPs. By binding of viral matrix protein M1 vRNPs are excluded from any further participation in the replication cycle ($\text{Vp}^{\text{nuc}}_{\text{M1}}$). Subsequent attachment of nuclear export protein NEP initiates the export to the cytoplasm ($\text{Vp}^{\text{cyt}}_{\text{M1}}$). Eventually, all necessary vRNP complexes and viral proteins assemble at the cell membrane to form progeny virions that bud from the surface into the surrounding medium as newly released virus particles V^{Rel}.

In [40] a slightly modified model was used within an age-structured model formulation to analyze the effects of direct acting antivirals within a low MOI setup. Though the model does not account for non infective virus particles, it can be used to characterize the influenza vaccine production process up to certain extent and will be used in the following within a population balance modeling framework.

5.2. The population balance model

The basic structure of the population balance model is similar to the structure of the previously introduced generic model for viral replication which was used in Section 4.7. The dynamics of the infected cells number density distribution are given by the following multidimensional PBE

$$\frac{\partial i_c(t, \mathbf{x}^*)}{\partial t} + \underbrace{\nabla_{\mathbf{x}^*} \{\mathbf{h}^*(\mathbf{x}^*) \ i_c(t, \mathbf{x}^*)\}}_{\text{Intracellular reactions}} = \underbrace{-(k_T^{\text{Apo}} + k_i^{\text{Apo}}) i_c(t, \mathbf{x}^*)}_{\text{Apoptosis}} + \underbrace{r^{\text{inf}}(t) \ T(t) \ \mathcal{I}(\mathbf{x}^*)}_{\text{Infection}},$$

(5.1)

where $\mathbf{h}^*(\mathbf{x}^*)$ represents the intracellular dynamics

$$\mathbf{h}^*(\mathbf{x}^*) = [\mathbf{h}(\mathbf{x}^*), \ \mathbf{0}]^T$$

(5.2)

of the augmented state vector

$$\mathbf{x}^* = [\mathbf{x}, \ \mathbf{k}]^T .$$

(5.3)

In the definition of the latter, $\mathbf{x}$ and $\mathbf{h}$ are the single cell state vector and the single cell dynamics according to [40] (see Appendix F), respectively. It is assumed that cell-to-cell variability of intracellular rates can directly be mapped to a distribution of the corresponding reaction rate parameters in the cell population. These parameters are summarized in the vector $\mathbf{k}$ and are assumed to be not affected by the infection dynamics. As in the generic example presented in Section 4.7, heterogeneity of newly infected cells is accounted for by $\mathcal{I}(\mathbf{x}^*)$ which describes the distribution of newly infected cells in the state space. In addition to infected cells, apoptotic infected cells are considered. These are assumed uniform with respect to their intracellular composition. Their dynamics are characterized by the following ODE

$$\frac{\mathrm{d}I_a(t)}{\mathrm{d}t} = \int_{\mathbf{X}^*} \left(k_T^{\text{Apo}} + k_I^{\text{Apo}}\right) i_c(t, \mathbf{x}^*) \ \mathrm{d}\mathbf{x}^* + r^{\text{Inf}}(t) \ T_a(t) - k^{\text{Lys}}(t) \ I_a(t) .$$

(5.4)

Furthermore, the concentrations of uninfected target cells $T(t)$ and their apoptotic counterparts $T_a(t)$ are characterized by the following dynamics

$$\frac{\mathrm{d}T(t)}{\mathrm{d}t} = g \ T(t) - r^{\text{inf}}(t) \ T(t) - k_T^{\text{Apo}} \ T(t) ,$$

$$\frac{\mathrm{d}T_a(t)}{\mathrm{d}t} = k_T^{\text{Apo}} \ T(t) - r^{\text{inf}}(t) \ T_a(t) - k^{\text{Lys}}(t) \ T_a(t) ,$$

(5.5)

with the growth rate defined by

$$g = \left[\frac{g_{\max}}{T_{\max}} \left(T_{\max} - T(t) - \int_{\mathbf{X}^*} i_c(t, \mathbf{x}^*) \ \mathrm{d}\mathbf{x}^*\right)\right] .$$

(5.6)

Virus particles are distinguished according to their location. There are free virus particles $V(t)$ (located in the extracellular medium), virus particles attached to the surface

of target cells $V^{\text{Att}}(t)$ and virions located in endosomes of target cells $V^{\text{Ex}}(t)$. The respective dynamics of the virus particle concentration in the extracellular medium $V(t)$ are given by,

$$\frac{\mathrm{d}V(t)}{\mathrm{d}t} = \int\limits_{\mathbf{X}^*} r^{\text{Rel}}(\mathbf{x})\ i_c(t, \mathbf{x}^*)\ \mathrm{d}\mathbf{x}^* - k_V^{\text{Deg}}\ V(t) + \sum_k \left[k_n^{\text{Dis}}\ V_n^{\text{Att}(t)} - k_{c,n}^{\text{Att}} B_n\ V(t) \right].$$

(5.7)

Furthermore, virions attached to the surface of the target cells can be described by

$$\frac{\mathrm{d}V_n^{\text{Att}}(t)}{\mathrm{d}t} = k_{c,n}^{\text{Att}}\ B_n\ V(t) - \left(k_n^{\text{Dis}} + k^{\text{En}}\right)\ V_n^{\text{Att}}(t) - \left(r^{\text{inf}}(t) + r^{\text{lys}}(t)\right)\ V_n^{\text{Att}}(t),\quad (5.8)$$

and the virions in the endosome are characterized by

$$\frac{\mathrm{d}V^{\text{En}}(t)}{\mathrm{d}t} = k^{\text{En}}\left(V_{hi}^{\text{Att}}(t) + V_{lo}^{\text{Att}}(t)\right) - k^{\text{Fus}}\ V^{\text{En}}(t) - \left(r^{\text{inf}}(t) + r^{\text{lys}}(t)\right)\ V^{\text{En}}(t),\quad (5.9)$$

with

$$B_n = B_n^{\text{tot}}\left(T(t) + T_a(t)\right) - V_n^{\text{Att}}(t),$$

$$k_n^{\text{Dis}} = \frac{k_{c,n}^{\text{Att}}}{k_{c,n}^{\text{Equ}}},$$

$$n \in \{lo,\ hi\}.$$

(5.10)

Two types of binding sites for the virus particles on the surface are considered: low affinity (lo) and high affinity (hi). A detailed description of the involved kinetic processes can be found in [41, 40]. The infection and lysis rates are defined as

$$r^{\text{inf}}(t) = \frac{F_{\text{inf}} k^{\text{Fus}} V^{\text{En}}(t)}{T(t) + T_a(t)},$$

$$r^{\text{lys}}(t) = \frac{k^{\text{lys}} T_a(t)}{T(t) + T_a(t)},$$

(5.11)

and the virus release rate depends on the amounts of viral compounds in each cell

$$r^{\text{Rel}}(\mathbf{x}) = k^{\text{Rel}} \frac{V p_{M1}^{cyt}}{V p_{M1}^{cyt} + 8\ K_{Vrel}} \prod_j \frac{P_j}{P_j + N_{P_j}\ K_{Vrel}},$$

$$P_j \in \{RdRp,\ HA,\ NP,\ NA,\ M1,\ M2,\ NEP\}.$$

(5.12)

All parameters for the population balance model were taken from [40] and can be found in Table 5.1.

Table 5.1.: *Parameters for population balance model characterizing influenza virus replication according to [40]*

Parameter	Value		Parameter	Value	
F_{inf}	1	$\frac{\text{cells}}{\text{virions}}$	k_I^{Apo}	$3.28\ 10^{-2}$	h^{-1}
k_T^{Apo}	$7.35\ 10^{-3}$	h^{-1}	k^{Fus}	$9.56\ 10^{-3}$	h^{-1}
k_{Lys}	$6.39\ 10^{-2}$	h^{-1}	k^{Rel}	586	$\frac{\text{virions}}{h}$
$k_{c,lo}^{\text{Att}}$	$1.85\ 10^{-10}$	$\frac{ml}{\text{sites } h}$	$k_{c,hi}^{\text{Att}}$	$3.32\ 10^{-8}$	$\frac{ml}{\text{sites } h}$
$k_{c,lo}^{\text{Equ}}$	$3.32\ 10^{-11}$	$\frac{ml}{\text{site}}$	$k_{c,hi}^{\text{Equ}}$	$4.48\ 10^{-9}$	$\frac{ml}{\text{site}}$
B_{lo}^{tot}	1000	$\frac{\text{sites}}{\text{cell}}$	B_{hi}^{tot}	150	$\frac{\text{sites}}{\text{cell}}$
T_{max}	$7\ 10^5$	$\frac{\text{cells}}{ml}$	μ_{max}	0.03	h^{-1}
k_V^{Deg}	0.1	h^{-1}	k^{En}	4.8	h^{-1}

5.3. Effects of cell-to-cell variability on unmodified cell line

In the following the previously presented approach for efficient moment approximation is applied and the effect of different levels of host cell heterogeneity with respect to certain production rate parameters on the overall virus yield is analyzed. Focus is on the production rates of viral proteins as well as viral mRNA, cRNA, vRNA and the binding rate of M1 which mark important steps in the viral replication kinetics. The corresponding distributed parameters are summarized in the following vector

$$\mathbf{k} = [k_{\text{Syn,M}},\ k_{\text{Syn,P}},\ k_{\text{Syn,C}},\ k_{\text{Syn,V}},\ k_{\text{Bind,M1}}]\ . \tag{5.13}$$

The intracellular reactions are highlighted in the single cell reaction scheme in Fig. 5.1.

In practice, the process is run with multiple infection cycles to obtain a high number of newly produced virions from a relative low number of seed virions. For the current example, the following initial conditions are assumed

$$T(t=0) = 4.9\ 10^5\ \frac{\text{cells}}{ml}, \quad T_a(t=0) = 9.49\ 10^3\ \frac{\text{cells}}{ml},$$

$$I_c(t=0) = I_a(t=0) = 0\ \frac{\text{cells}}{ml}, \quad V(t=0) = 6.9\ 10^4\ \frac{\text{virions}}{ml},$$

$$V_{lo}^{\text{Att}}(t=0) = V_{hi}^{\text{Att}}(t=0) = V^{\text{En}}(t=0) = 0\ \frac{\text{virions}}{ml}, \tag{5.14}$$

which correspond to a MOI of 0.14. Following [40], the virus entry dynamics is described by (5.8)-(5.9). Infection is defined as the time point, when the virus is transferred from the endosome to the cytoplasm.

Newly infected cells are initialized with a complete set of 8 vRNP segments in the cytoplasm while the other intracellular compounds are zero. No variances with respect to the intracellular compounds are assumed for newly infected cells

$$\mathbf{x}_{\text{inf}} = \left[8, \mathbf{0}^{1 \times 26}\right]^T\ . \tag{5.15}$$

However, cell-to-cell variability of newly infected cells with respect to the rate coefficients (5.13) is taken into account. In this section, it is assumed that the parameters are distributed according to the following logarithmic Gaussian distribution

$$\mathbf{k}_{\text{inf}} = \mathcal{L}\left(\mathbf{k}_{\text{nom}}, p_{\text{var}}\right) = \mathbf{k}_{\text{nom}} \exp\left(\mathcal{N}(\mathbf{0}, p_{\text{var}}\mathbf{I})\right) \tag{5.16}$$

with nominal parameter values given as (see [40])

$$\mathbf{k}_{\text{nom}} = \begin{pmatrix} 8.53 \ 10^5 \text{ nucleotides } h^{-1} \\ 6.48 \ 10^4 \text{ nucleotides } h^{-1} \\ 5.29 \ h^{-1} \\ 32.18 \ h^{-1} \\ 2.43 \ 10^{-4} \text{ (molecule } h)^{-1} \end{pmatrix}. \tag{5.17}$$

Thereby, it is assumed that all parameters exhibit the same relative variance with respect to their nominal values which is scaled by the parameter p_{var}. Here, the resulting distribution of newly infected cells $\mathcal{I}(\mathbf{x}^*)$ to the augmented state space can be described by

$$\mathcal{I}(\mathbf{x}^*) = \begin{pmatrix} \mathbf{x}_{\text{inf}} \\ \mathbf{k}_{\text{inf}} \end{pmatrix}. \tag{5.18}$$

As described in Section 4.2.5, the PBE characterizing the infected cells number density distribution (5.1) can be reformulated as a series of initial value problems

$$\frac{\partial i_{c,l}(t, \mathbf{x}^*)}{\partial t} + \nabla_{\mathbf{x}^*}\left\{\mathbf{h}^*(\mathbf{x}^*) \ i_{c,l}(t, \mathbf{x}^*)\right\} = -(k_T^{\text{Apo}} + k_i^{\text{Apo}})i_{c,l}(t, \mathbf{x}^*), \tag{5.19}$$

with the corresponding initial conditions

$$i_{c,l}(t = t_l, \mathbf{x}^*) = r^{\text{inf}}(t_{l-1}) \ T(t_{l-1}) \ \mathcal{I}(\mathbf{x}^*)\Delta t, \quad \Delta t = (t_l - t_{l-1}). \tag{5.20}$$

Each initial value problem describes the dynamics of cells which have been infected within the interval Δt. As the intracellular dynamics do not depend on the dynamics of the extracellular species (i.e. target/infected apoptotic cells, non-apoptotic target cells), these can be solved in advance based on a normalization as it was described in Section 4.7.2. Thereby, the numerical burden can be reduced further.

For the solution of the normalized initial value problem by application of the efficient moment approximation algorithm, $\mathcal{I}(\mathbf{x}^*)$ has to be represented by a set of abscissas and weights. As the cells do not exhibit heterogeneity with respect to the intracellular compounds $\mathbf{x}$, all abscissas would have the same initial location with respect to these. In contrast, the abscissas will be different with respect to $\mathbf{k}_{\text{inf}}$. Thus, it is sufficient to choose initial abscissas as

$$\mathbf{x}_\alpha = \begin{pmatrix} \mathbf{x}_{\text{inf}} \\ \mathbf{k}_\alpha \end{pmatrix}, \quad \alpha = 1, \ldots, N_\alpha \tag{5.21}$$

where the abscissas $\mathbf{k}_\alpha$ are chosen as follows: At first, abscissas are chosen for the Gaussian distribution $\mathcal{N}(\mathbf{0}, p_{\text{var}}\mathbf{I})$ applying the cubature formulas presented in Section 4.3, resulting in the abscissa/weight pairs

$$\left[\mathbf{k}_{\mathcal{N},\alpha}, \ \mathbf{w}_\alpha\right]. \tag{5.22}$$
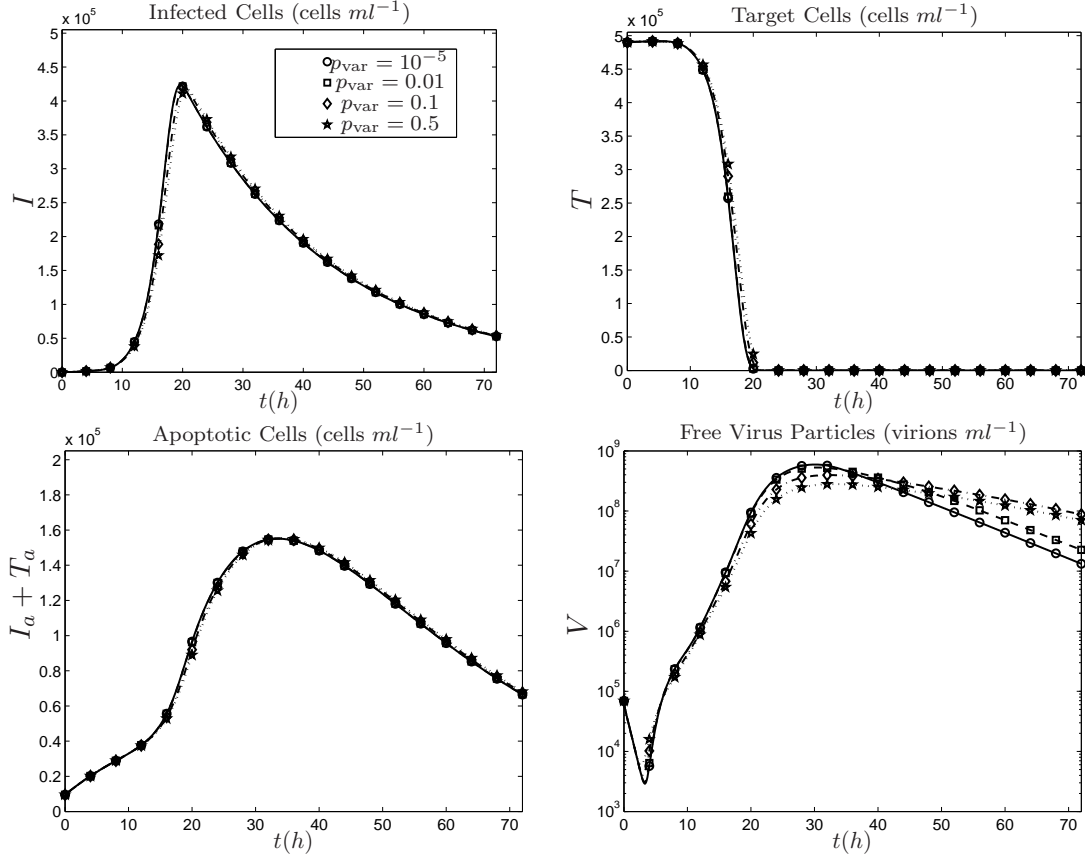
Figure 5.2.: *Simulation results for influenza vaccine production for different degrees of cell-to-cell variability*

Afterwards, the abscissas are transformed to a logarithmic Gaussian distribution

$$\mathbf{k}_\alpha = \mathbf{k}_{\text{nom}} \exp(\mathbf{k}_{\mathcal{N},\alpha}). \tag{5.23}$$

Now, the model is simulated for increasing degrees of cell-to-cell variability. This is implemented by choosing the value of p_{var} from the following vector

$$\mathbf{p}_{\text{var}} = \begin{bmatrix} 10^{-5}, \ 0.01, \ 0.1, \ 0.5 \end{bmatrix}. \tag{5.24}$$

The first case $p_{\text{var}} = 10^{-5}$ represents a very small cell-to-cell variability, such that the corresponding results can be compared to the ones presented in [40]. For the approximation of $\mathcal{I}(\mathbf{x}^*)$, 11 Gaussian mixed densities were used (see Appendix G for approximation accuracy). For each, abscissas and weights were determined using the sigma point rule. Thus the overall number of abscissas is given by

$$N_\alpha = N_{\text{GMD}} \left(2N_d + 1\right) = 121. \tag{5.25}$$

In Fig. 5.2 simulation results are shown for $\Delta t = 0.1 \ h$. It can be seen that the degree of cell-to-cell variability, represented by the parameter p_{var}, does not exert a significant

influence on the overall numbers of infected, uninfected and apoptotic cells. Only a slight delay in the formation of infected cells and apoptotic cells is observed. Simultaneously, the decrease in the target cells is slightly delayed. In contrast, the overall virus dynamics are significantly affected by an increased cell-to-cell variability. The difference is observed from around 20 h post infection (p.i.). It can be seen, that an increased cell-to-cell variability results in a decreased maximum virus concentration. In comparison to a nearly homogeneous cell population (i.e. $p_{\text{var}} = 10^{-5}$) an increased heterogeneity of $p_{\text{var}} = 0.5$ results in a decrease of the maximum concentration from $V_{\text{max}} \approx 6 \ 10^8$ to $V_{\text{max}} \approx 3 \ 10^8$ virions ml^{-1}. Having in mind that process for vaccine production is stopped at the peak of the virus concentration, the results indicate, that an increased cell-to-cell variability is a disadvantage for the overall production process. However, the simulation results show a sustained virus production for increasing parameter heterogeneity. This indicates that virus particles are released from the cell population over a longer time leading to a slower decrease of the virus concentration which may be advantageous for a continuous vaccine production process.

5.4. Effects of cell-to-cell variability on modified host cells

Besides optimization of the operating conditions, genetically modified cells can be used to improve the overall production process. One approach is to modify gene expression of host cell factors important for viral replication to establish a high yield production cell line. A comprehensive analysis of host cell factors affecting influenza virus polymerase activity demonstrate the importance of RNA synthesis rates in the mathematical model [108]. Furthermore, the viral matrix protein 1 (M1) is known to promote the nuclear export or viral RNP [72]. For this reason, in the following, focus is on modifications of the synthesis rates of viral mRNA, vRNA, cRNA, viral protein translation and the binding rate of viral protein M1. The affected kinetic reactions are colored in the basic scheme Fig. 5.1.

Genetic modifications, e.g. by lentiviral transduction, can be used in order to achieve an overexpression or knockdown of selected host cell factors to increase the cell specific virus production rate. However, upon lentiviral transduction the degree of overexpression or knockdown of host cell genes varies within the modified cell population [13]. Thus a considerable cell-to-cell variability in the viral production rates is expected which may significantly affect the overall vaccine production process.

5.4.1. Sensitivity analysis

First, a sensitivity analysis will be carried out to determine the most promising targets for an improvement of the virus yield. Simulations are evaluated with respect to maximum concentration of the extracellular virus particles and the respective peak time aiming at increasing the first and decreasing the latter. It is assumed, that genetic modifications of the cells can be mapped directly to a modification of the corresponding intracellular reaction rate parameters (5.13). The kinetic parameters are modified
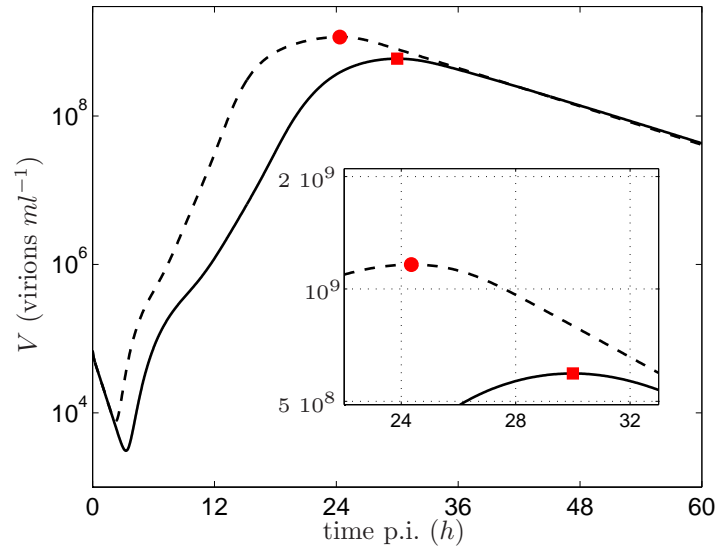
Figure 5.3.: *Virus concentration for inhibition of binding rate of M1 by factor* 0.2 *(dashed) compared to unmodified cell line (solid); filled red symbols mark the maximum concentrations*

by a factor of 0.2 or 5, respectively. Furthermore, for the sensitivity analysis it is assumed that all cells are modified with the same degree of knockdown/overexpression and thereby exhibit the same modified parameter set. Thus the distribution of newly infected cells $\mathcal{I}(\mathbf{x}^*)$ can be represented by a single abscissa in the augmented state space.

At first, only one parameter is changed while the others are kept constant. Thus 10 independent combinations have to be analyzed in view of the desired improvements stated above. Here, an inhibition of the binding rate of M1 $k_{Bind,M1}$ has the largest effect on the virus dynamics. The simulation results are depicted in Fig. 5.3. In contrast to the unmodified cell line (solid), the maximum virus yield approximately doubles from $5.94 \cdot 10^8$ to $1.16 \cdot 10^9$ virions ml^{-1}. Furthermore, the peak time has decreased from approximately 30 to 24.4 h p.i. In the second case, all reaction rates are subject to modification. Thus, an overall amount of $3^5 = 243$ combinations has to be analyzed. The best improvement is found for the combination, in which the synthesis rates for vRNA, mRNA and viral proteins are increased, the M1 binding rate is reduced while the cRNA synthesis rate is kept unmodified. In Fig. 5.4 the corresponding virus dynamics is depicted (dashed) in comparison to the unmodified cell population (solid). The maximum virus concentration of the modified cells increases by a factor of 2.49 to $1.48 \cdot 10^9$ virions ml^{-1}. Furthermore, the peak time decreases to 25.2 h p.i. and the virus concentration does not decrease as fast as in case of the previous modification of one reaction rate (see Fig. 5.3). This suggests that the cells produce virions for a longer time period which may also be advantageous for the overall vaccine production process: as harvesting at later time points still results in a considerably higher virus concentration.
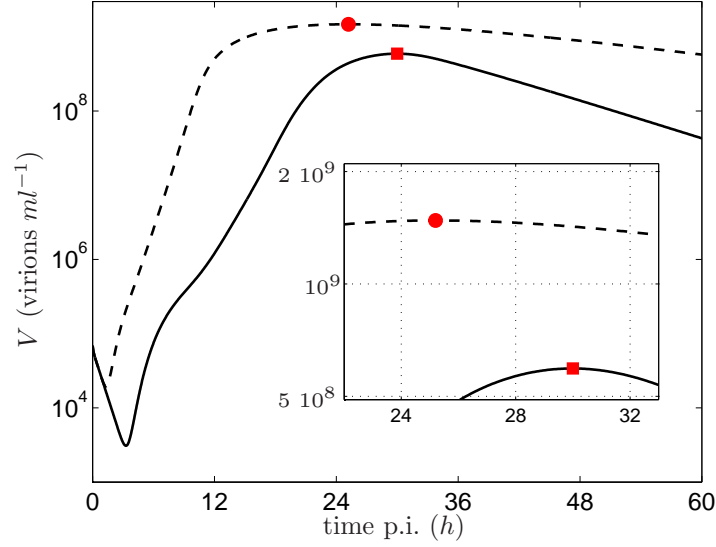
Figure 5.4.: *Virus concentration for modification of 5 rates $\left[k_{M1}^{Bind} \downarrow,\ k_C^{Syn} \rightarrow,\ k_M^{Syn} \uparrow, \right.$ $\left. k_P^{Syn} \uparrow,\ k_V^{Syn} \uparrow \right]$ (dashed) compared to unmodified cell line (solid); filled red symbols mark the maximum concentrations*

5.4.2. Effects of cell-to-cell variability on modified cell line

In the previous section it was shown that genetic modifications can be used in principle to improve the overall vaccine production process. However, it was assumed that all modified cells show the same level of gene expression which is a rather strong assumption. When cells are modified e.g. with the help of lentiviral vectors, the degree of overexpression and knockdown varies from cell to cell. Thus, the modified cell population will exhibit some degree of variance with respect to the viral production rates. A desirable transduction would result in a cell population with rate parameters distributed narrowly around the desired modification. Different degrees of heterogeneity resulting from lentiviral transduction can now be represented by broadening distributions of the kinetic parameters. In the following, the parameter distribution within the cell population is approximated by a weighted sum of five logarithmic Gaussian distributions to account for asymmetric distributions as illustrated in Fig. 5.5.

Modification of one reaction rate

At first, the inhibition of the M1 binding to viral genomes is considered while the other rates are kept unmodified. As mentioned above, it is assumed that a prior transduction results in a distribution of the parameter k_{Bind}^{M1} within the cell population which can described by a weighted sum of five logarithmic Gaussian distributions

$$k_{Bind}^{M1} \sim \sum_{l=1}^{5} a_l\ \mu_l\ e^{\mathcal{N}(0, \sigma_l)}\,. \tag{5.26}$$

(a) $a_{1,2,3,4,5} = (0.01, 0.03, 0.04, 0.06, 0.86)$ (b) $a_{1,2,3,4,5} = (0.04, 0.12, 0.19, 0.27, 0.38)$

(c) $a_{1,2,3,4,5} = (0.2, 0.2, 0.2, 0.2, 0.2)$
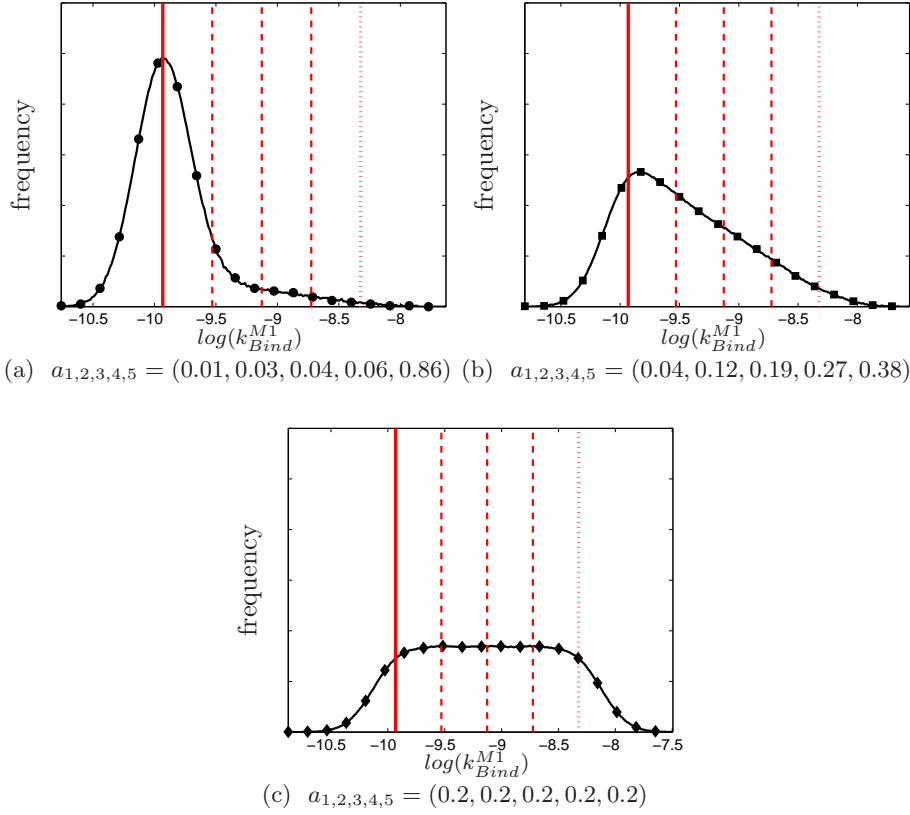
Figure 5.5.: *Parameter distributions for investigated scenarios representing different de-grees of cell-to-cell variability; mean values of the underlying logarithmic normal distributions are highlighted red*

Thus, the distribution of newly infected cells is given by

$$\mathcal{I}(\mathbf{x}^*) = \begin{pmatrix} \mathbf{x}_{\text{inf}} \\ k_{\text{Syn,M,nom}} \\ k_{\text{Syn,P,nom}} \\ k_{\text{Syn,C,nom}} \\ k_{\text{Syn,V,nom}} \\ k_{\text{Bind,M1}} \end{pmatrix}. \tag{5.27}$$

The mean values μ_l are logarithmically distributed between the nominal value $k_{Bind,\text{nom}}^{M1}$ and $0.2\, k_{Bind,\text{nom}}^{M1}$, which was determined in the prior sensitivity analysis. The variances of the logarithmic Gaussians are chosen as $\sigma_l = 0.05$, $l = 1, \ldots, 5$. By using different values for the coefficients a_l, different scenarios can be simulated. For the following simulation study, three different cases are investigated (see Fig. 5.5a - Fig. 5.5c). These map an increasing heterogeneity of the cell population ranging from a narrow distri-bution (good transduction, Fig. 5.5a) to a very broad distribution (bad transduction, Fig. 5.5c).

For the following numerical simulations $\mathcal{I}(\mathbf{x}^*)$ is first represented by 11 logarithmic
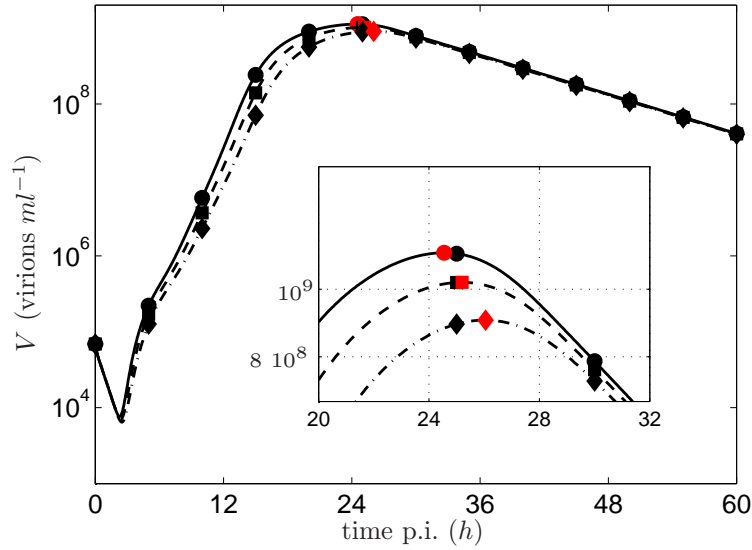
Figure 5.6.: *Virus concentration for different levels of host cell heterogeneity if the M1 binding rate is modified; the maxima are highlighted red for each scenario*

Gaussian distributions. For each, $2N_d + 1 = 3$ pairs of abscissas and weights were chosen using the sigma point rule (4.90)-(4.91) resulting in a total amount of $N_\alpha = 33$ abscissas. The simulation results are shown for the initial conditions which were already used for the unmodified case (5.14) and $\Delta t = 0.1$ h in Fig. 5.6. It can be seen that the maximum virus concentrations (flagged with red markers in Fig. 5.4) decrease with increasing degree of heterogeneity. For the analyzed scenarios (symbols corresponding to the ones in Fig. 5.5) the maximum virus concentrations are $1.13 \cdot 10^9$ virions ml^{-1} for scenario (a), $1.02 \cdot 10^9$ virions ml^{-1} for (b) and $9.03 \cdot 10^8$ virions ml^{-1} for (c). The peak time increases to 24.6 (a), 25.2 (b) and 26.1 h p.i.(c). The simulation results suggest that the overall production process using modified cell lines is relatively robust against host cell heterogeneity and an increase of the maximum virus titer is possible. However, for larger degrees of heterogeneity the desired aim of increasing the maximum virus concentration is reduced.

Modification of multiple reaction rates

In this section the influence of cell-to-cell variability is investigated, when up to five reaction rates are modified simultaneously. The best combination for a up/down regulation of five reaction rates was already found from the sensitivity analysis. Here, enhancement of mRNA, vRNA and protein synthesis rates in combination with a reduced M1 binding rate has shown the best improvement in maximum virus yield and peak time. Similar to the previous investigation, it is assumed that cell-to-cell variability resulting from transduction can be represented by a weighted sum of five logarithmic Gaussian

distributions. Those are now five dimensional, as five reaction rates are considered

$$\mathbf{k} = \sum_{l=1}^{5} a_l \ \mu_{\mathbf{l}} \cdot e^{\mathcal{N}(\mathbf{0}, \mathbf{diag}(\sigma_{\mathbf{l}}))} \ . \tag{5.28}$$

The mean vectors of the distributions $\mu_{\mathbf{l}}$ are logarithmically distributed between the nominal parameter vector $\mathbf{k}_{\mathrm{nom}}$ and the best combination $\mathbf{k}_{\mathrm{best}}$

$$\mathbf{k}_{\mathrm{nom}} = \left[k_C^{Syn}, \ k_M^{Syn}, \ k_P^{Syn}, \ k_V^{Syn}, \ k_{M1}^{Bind} \right]^T$$

$$\mathbf{k}_{\mathrm{mod}} = \left[k_C^{Syn}, \ 5 \ k_M^{Syn}, \ 5 \ k_P^{Syn}, \ 5 \ k_V^{Syn}, \ 0.2 \ k_{M1}^{Bind} \right]^T \ . \tag{5.29}$$

The variances are chosen as $\sigma_l = 0.05$. Furthermore, for the different scenarios, the weighting coefficients a_l are kept at the values reported in Fig. 5.5. It is worth mentioning that the cell population exhibits heterogeneity with respect to all considered reaction rates, even if they are not modified (the mean value of k_C^{Syn} remains at the original value). Thereby, a more general effect of genetic modifications is incorporated: viral and cellular factors are interconnected in a complex virus host cell interaction network. Consequently, the genetic modification of one host cell factor is likely to affect multiple steps of the viral replication cycle. Thus, it can be expected that an overexpression or knockdown of one host cell factor aiming at the enhancement or inhibition of a certain kinetic reaction rate also introduces some degree of heterogeneity in other reaction rates.

Three different scenarios are used to analyze the effects of heterogeneity within the cell population on the overall vaccine production process. For the representation of the distribution of newly infected cells $\mathcal{I}(\mathbf{x}^*)$ is first approximated by 15 Gaussian mixed density distributions for which abscissa/weight pairs were chosen with the sigma point rule (4.90)-(4.91) resulting in a total amount of $N_\alpha = 165$ abscissas. The simulation results are depicted in Fig. 5.7. As in the previously analyzed case, a broader parameter distribution correlates with a decreased maximum virus concentration and an increased peak time. The resulting values for the maximum virus concentrations (highlighted with red symbols in Fig. 5.7) are $1.44 \cdot 10^9$ virions ml^{-1} (a), $1.32 \cdot 10^9$ virions ml^{-1} (b) and $1.11 \cdot 10^9$ virions ml^{-1} (c). The corresponding values for the peak times are 25.5, 25.9 and 26.9 h p.i. In comparison to the simulation results for a homogeneous population as computed in Section 5.4.1, only around 40% of the increase in virus concentration is achieved which again emphasizes the impact of cell-to-cell variability on the overall process. However, even for the worst assumed scenario (c) a significant increase in the virus yield is predicted which suggests that the modification of certain reaction rates by knockdown/overexpression of host cell factors can be a suitable tool to overcome limitations in the overall process.
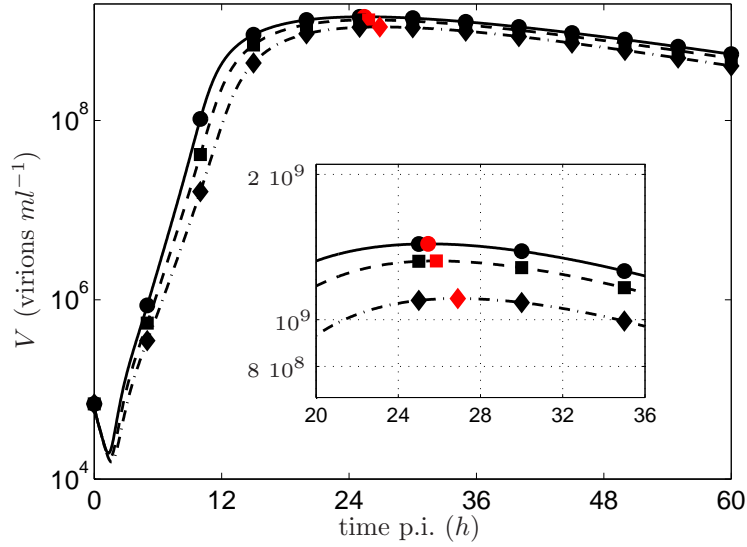
Figure 5.7.: *Virus concentration for different levels of host cell heterogeneity if five ki-netic parameters are modified (see (5.29)); the maxima are highlighted in red for each scenario*

5.5. Discussion

In this chapter application of the proposed approximate moment method to a high dimensional population balance is shown. The PBE is based on a detailed single cell model for influenza virus replication. At first, it is researched how an increasing variance of kinetic rate parameters in the cell population affects the overall dynamics of the process. It is shown, that an increasing cell-to-cell variability is attended by a decreased maximum virus yield. In the second part, the impact of heterogeneity in genetically engineered cell lines for influenza vaccine production is analyzed using the population balance model. A pragmatic sensitivity analysis shows that genetic modifications by means of overexpression and knockdown of host cell factors lead to an increased maximum virus yield and a reduced maximum virus yield peak time. Furthermore, it is shown the overall increase in maximum virus yield and the reduction of the virus yield peak time are relatively robust w.r.t. cell-to-cell variability as a result of genetic modifications via transduction techniques. Thus, the simulations suggest that geneti-cally engineered cell lines can be an important tool to improve the influenza vaccine production process.

6. Concluding remarks

6.1. Summary

Particulate systems are found in a wide range of processes from chemical and bioprocess engineering where a large number of individual particles, e.g. crystals or cells, are involved. In general, these are not uniform, but differ from each other with respect to characteristic properties, like size and internal composition. The dynamics of the particle ensemble can be described conveniently in the framework of population balance modeling which allows to account for particle-to-particle variabilities. Here, two alternative modeling strategies can be pursued: top-down and bottom-up modeling. While the first concentrates on a mechanistic modeling of available distributed measurements, the latter starts from a detailed description of kinetic processes on the single particle level. The obtained population balance models represent low dimensional partial differential equations for the top-down approach. Here, the main challenge is the adaption of the model equations to the distributed measurements which generally involves solution of an infinite dimensional inverse problem. In contrast, bottom-up modeling results in high dimensional partial differential equations. Here, the main issue is an efficient numerical solution.

Focus within this thesis was on influenza virus replication in cell cultures for vaccine production. This process represents an interesting particulate process as the cell population is characterized by large cell-to-cell variability. Both modeling approaches, top-down and bottom-up, are addressed.

The first part is concerned with top-down modeling of the process. For a previously developed model [79] a suitable scheme for the estimation of unknown parameters is presented. The model accounts for cellular variances with respect to the intracellular amount of viral nucleoprotein, which had been measured via flow cytometry. As the model features functional parameters which depend on the internal coordinate, the inverse problem is set in infinite dimension. The inverse problem is transformed to a finite dimension using Hermite spline approximations of different complexities for the functional parameters. Afterwards, the finite dimensional inverse problem is solved in a weighted least squares framework to adapt the model to flow cytometric and virus concentration measurements. The resulting model formulation is able to reproduce the dynamic phenomena observed in the experiments, including bimodality and backshift of the fluorescence distributions. It is also shown, that the model fit improves for a more complex spline representation. However, the increased degree of freedom is attended by increased numerical effort of the parameter estimation setup. Furthermore a biological meaningful interpretation of resulting parameter estimates is limited. Moreover, it is

shown that models with different assumptions on the length of the latent phase yield a similar model fitness, while the estimated parameters for the different models undergo drastic changes. This indicates, that identifiability of the model parameters is limited and the model parameters can therefore only be related roughly to underlying biological kinetics. Thus it has to be concluded that the top down model represents a convenient basis for qualitative analysis, yet further insight into the underlying biological effects of the process is limited. In fact, the whole intracellular viral replication mechanism is mapped to one internal coordinate (i.e. the intracellular amount of viral nucleoprotein) and the latent phase length.

To improve the situation more structured information has to be taken into account and a bottom-up model of the process can be formulated based on the detailed single cell kinetics. The emerging multi dimensional population balance equation can be solved numerically using approximate moment methods which require an efficient implementation. Thus, the second part of this thesis is concerned with the development, evaluation and application of an efficient moment approximation technique for multi dimensional population balance equations which are dominated by particle growth, particle death and nucleation. The presented technique is based on the direct quadrature method of moments. Here, moments are approximated by a weighted sum of abscissas. It is shown, how the dynamics of weights and abscissas can be derived analytically for pure particle growth and particle death. Furthermore, it is demonstrated how problems including nucleation can be reformulated either in form of a series of initial value problems or by introduction of an additional internal coordinate characterizing the particle age. Afterwards, an efficient choice of the abscissas and weights from Gaussian initial distributions applying a monomial cubature rule is presented. The latter is also known as sigma point rule and the number of abscissas and corresponding weights scales only linearly with the dimension of the problem. The overall method is applied to one and two dimensional population balance models. The comparison to reference solutions show that the approach using sigma point abscissas provides accurate approximations while simultaneously keeping the numerical cost at a low level.

Furthermore, the method was evaluated for a generic five dimensional population balance model of viral replication in cell cultures. At first, it is shown that the sigma point tuning parameters significantly affect the overall approximation accuracy. Next, an alternative to a problem specific manual optimization of the tuning parameters is proposed. Therein, the initial distribution is approximated by a sum of Gaussians and sigma point abscissas are chosen for each of these without optimization of the tuning parameter. It is shown, that the accuracy can be increased by increasing the number of Gaussians which are used to approximate the initial distribution. Thereby, the task of finding optimal sigma point abscissas is traded for an increased numerical effort. Both approaches, standard as well as alternative, are evaluated against other cubature formulas. It is demonstrated, that the alternative approach which combines Gaussian mixed densities and the sigma point cubature formulas outperforms all other approaches while keeping the computational costs on a reasonable scale.

Finally, application of the technique to a detailed population balance model describing influenza virus replication is shown. At first the effect of cell-to-cell variability in kinetic

reaction rate parameters on the overall virus yield is analyzed for an unmodified cell line. It is assumed, that protein synthesis rates of the cells are distributed on a logarithmic scale. The simulation results indicate, that an increased heterogeneity of the cells results in a lower maximal virus yield. On the other hand, the cell population shows a sustained virus production the larger the cell-to-cell variability. Furthermore, the influence of cell-to-cell variability on genetically engineered cell lines is analyzed. These are considered to be a helpful tool to overcome bottlenecks in influenza vaccine production processes. At first, a sensitivity analysis is conducted to identify the most promising kinetic parameters which can serve as a target for genetic modifications via transduction techniques. In practice, nonuniform gene expression is expected within the modified cell population when using these transduction techniques. Different degrees of heterogeneity are represented by broadening distributions of the kinetic parameters. Simulation results indicate that with modified cell lines the maximum virus yield can be increased and the maximum peak time can be reduced. In addition, the model based analysis shows that these improvements are relatively robust against moderate degrees of heterogeneity. Thus, the simulation results suggest that the use of genetically modified cell lines represents a promising option to overcome bottlenecks in vaccine production processes.

6.2. Future Perspective

The top-down approach and the proposed model adaption can be used to formulate population balance models for other vaccine production processes where flow cytometric measurements have shown similar dynamic effects of the cell culture, e.g. vaccinia virus replication [106, 47]. Furthermore, an extension of the top-down modeling approach to other process modes, e.g. continuous and multi stage process schemes [110, 109], are an interesting topic of future research. Here, additional phenomena, e.g. the accumulation of defective interfering particles [40], have to be taken into account and may require an extension of the parameter estimation technique. Moreover, the inverse problem solution technique may also be applied to problems which feature a similar model structure, e.g. crystallization processes, to estimate the crystal growth rates which depend on crystals properties, like size or shape.

So far the proposed moment approximation algorithm can only be applied to particulate systems in which kinetic processes like agglomeration, breakage and cell division are negligible. Thus extension to these processes is a major future research issue. Here, a problem reformulation similar to the technique presented in Section 4.2.5 or the successive generations approach [57],represents a promising alternative. Therein, the particle population is divided into subpopulations which are probably easier to handle. Furthermore, an extension of the moment approximation to spatially distributed problems is desirable to describe the interaction of the particulate phase with a non ideally mixed fluid phase or spatial separation of cells in an multi cellular environment (see e.g. [42] and the references therein).

Future applications of the algorithm include parameter estimation procedures for the

presented multi dimensional population balance model for influenza virus replication to multi dimensional data obtained from flow cytometric measurements. Furthermore, the proposed approximate moment method can be applied to multi dimensional population balance models derived by bottom-up modeling of other multi cellular biological processes. Examples are found in bioprocess engineering, e.g. biopolymer production in bacteria cell cultures [28, 27, 19], but also in the field of systems biology and systems medicine, e.g. modeling of cell proliferation [14, 7, 42] and cell metabolism [114]. Moreover, application of the developed method is not limited to examples from bioprocess engineering but can also be applied to multi dimensional PBEs modeling the evolution of crystal morphology (e.g. [8]).

Appendix

A. The method of characteristics

The method of characteristics (MOC) represents a solution technique for first order hyperbolic partial differential equations[94]. The basic idea is to transform the PDE to a set of ODEs by reparameterization. Without loss of generalization, in the following the special case of the scalar quasi-linear partial differential equation

$$A(t, x, n)\frac{\partial n}{\partial t} + B(t, x, n)\frac{\partial n}{\partial x} = C(t, x, n) \,, \tag{A.1}$$

is considered. The boundary and initial conditions are given by

$$n(t = 0, x) = n_{\text{IC}}(x) \,, \qquad n(t, x = 0) = n_{\text{BC}}(t) \,. \tag{A.2}$$

Here t and x are the independent variables and the overall solution of interest is $n(t, x)$.

In order to compute a solution using the MOC, the independent variables are parameterized by new variables θ and s

$$t = t(\theta, s) \,, \qquad x = x(\theta, s) \,. \tag{A.3}$$

The idea is now to find the solution of the PDE along these curves or characteristics using s to parameterize the initial point of a curve and using θ to parameterize the curve itself. The change of n along the curve (i.e. under variation of θ) can be computed applying chain rule:

$$\frac{\partial n}{\partial \theta} = \frac{\partial t}{\partial \theta}\frac{\partial n}{\partial t} + \frac{\partial x}{\partial \theta}\frac{\partial n}{\partial x} \,. \tag{A.4}$$

Comparison of (A.1) and (A.4) yields the following system of ODEs

$$\frac{\partial t}{\partial \theta} = A(\theta, s, n) \,,$$
$$\frac{\partial x}{\partial \theta} = B(\theta, s, n) \,,$$
$$\frac{\partial n}{\partial \theta} = C(\theta, s, n) \,, \tag{A.5}$$

which is also called the *characteristic system* of (A.1). Here, the first and the second ODE describe the dynamics of the characteristic curve whereas the third equation describes the evolution of the solution n along this curve. The principle scheme is visualized in Fig. A.1 for the solution of the initial value problem (IVP) and the boundary value problem (BVP), respectively. It can easily be seen that s parameterizes the initial curve $n(t = 0, x)$ while θ characterizes the position on the characteristic curve.

(a) *parametrised initial curve (IVP)*   (b) *parametrised boundary curve (BVP)*
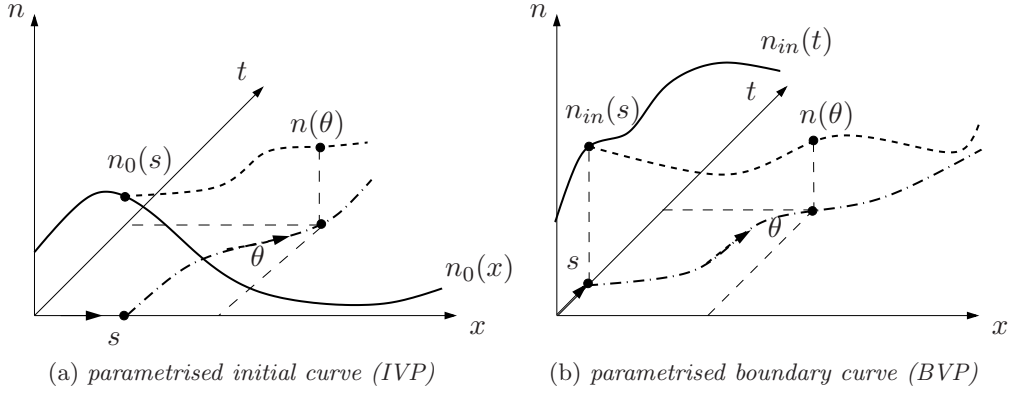
Figure A.1.: *Parameterization for initial value problem and boundary value problem*

For a solution of the IVP the corresponding initial values for the characteristic system are given as

$$
\begin{aligned}
t(\theta = 0) &= 0 \\
x(\theta = 0) &= s \\
n(\theta = 0) &= n_{IC}(s)
\end{aligned}
\tag{A.6}
$$

For the solution of the corresponding BVP, s is used to parameterize the boundary condition. A solution can be computed by solution of the characteristic system for the following initial conditions

$$
\begin{aligned}
t(\theta = 0) &= s \\
x(\theta = 0) &= 0 \\
n(\theta = 0) &= n_{BC}(s)
\end{aligned}
\tag{A.7}
$$

The analytical solutions of the IVP and the BVP yield expression for the parameterization of the characteristics. To obtain an expression for $n(t, x)$ the solutions have to be inverted

$$
s = s(t, x) , \qquad \theta = \theta(t, x)
\tag{A.8}
$$

which is not always possible.

The MOC is the standard solution approach for processes described by systems of first order hyperbolic partial differential equations and has many applications in chemical engineering (see e.g. [94]).

B. Two dimensional crystallization process

In Fig. B.1 the temporal evolution of moments up to the sixth order are shown for the proposed moment approximation algorithm based on the DQMOM in combination with an efficient choice of the initial abscissas and weights based on the sigma point rule. The corresponding relative errors are shown in Fig. B.2.
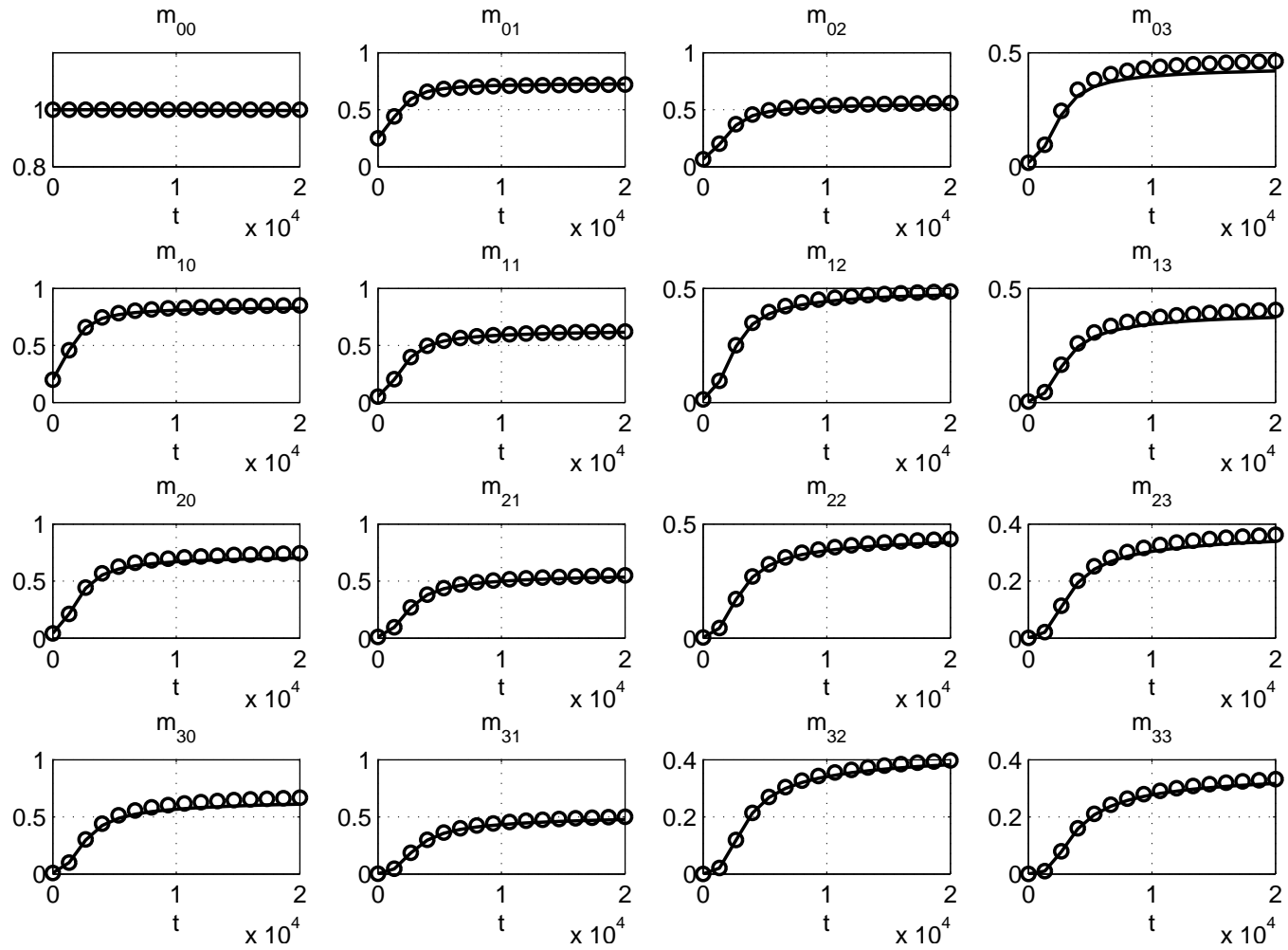
Figure B.1.: *Moments computed with sigma point approach (circles) and full reference solution with pseudo spectral collocation (solid)*
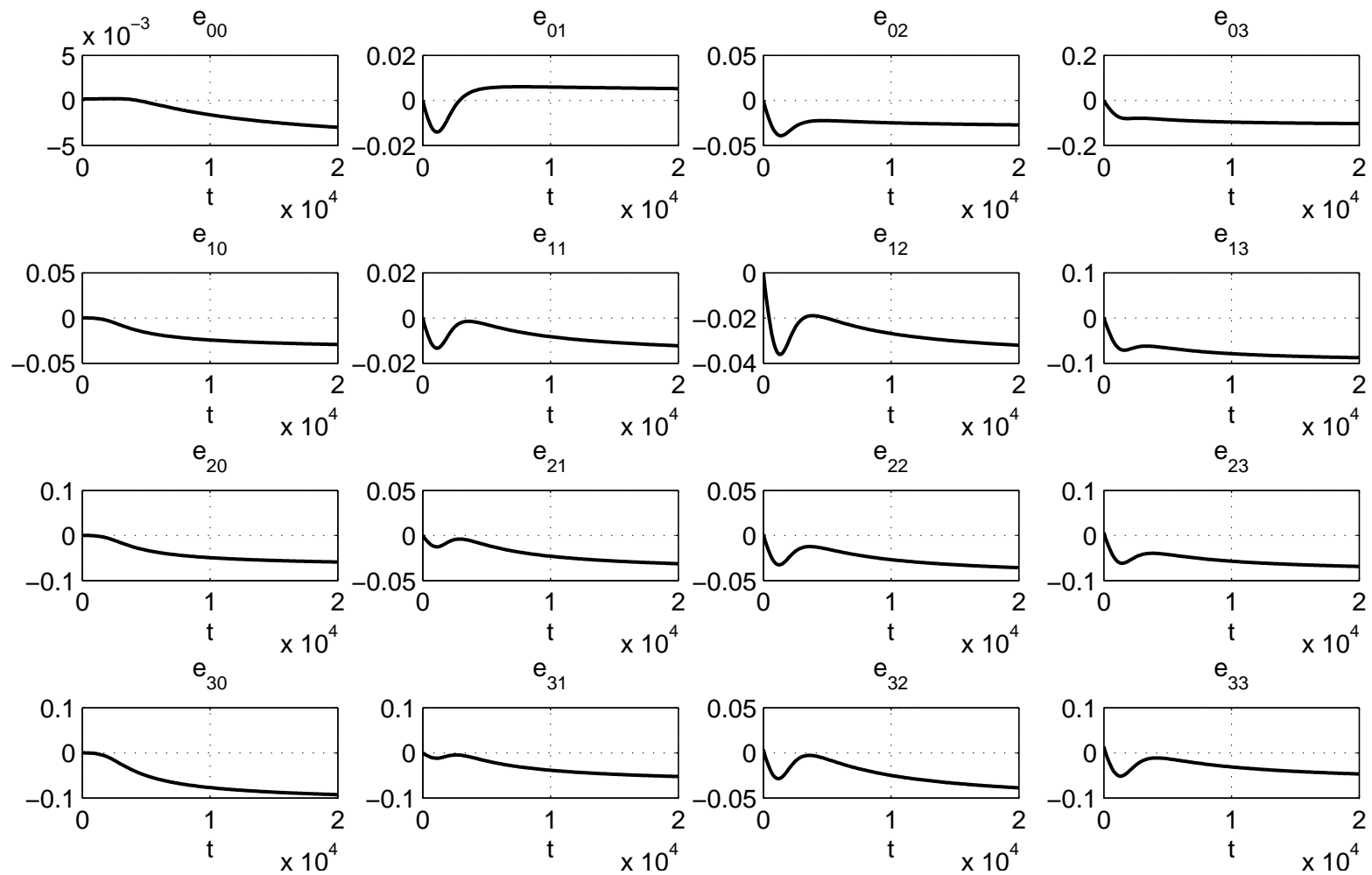
Figure B.2.: *Relative error between moments computed with sigma point approach and full reference solution with pseudo spectral collocation*

C. The finite volume method

The finite volume method is a discretization method which is frequently applied within the numerical solution of population balance equations [56]. In the following its principle application will be demonstrated for the one dimensional growth dominated PBE

$$\frac{\partial n(t,x)}{\partial t} = -\frac{\partial}{\partial x}\{G(t,x)\ n(t,x)\} - D\ n(t,x) \tag{C.1}$$

describing the number density distribution with respect to the general particle property x. The property coordinate is discretized into N_{CV} intervals or control volumes as depicted in Fig. C.1. Then the number density distribution is described on grid nodes
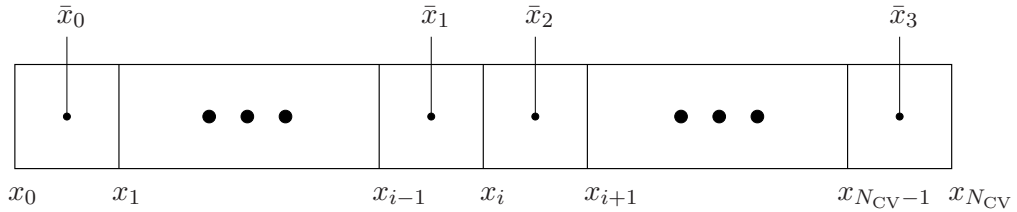
Figure C.1.: *Discretization of property coordinate x into control volumes*

$\bar{x}_i$ which lie in the control volumes (e.g. in the middle of the control volumes $\bar{x}_i = (x_i + x_{i+1})/2$). The dynamics of the number density distribution on a grid node can be derived by integration of (D.1) over the corresponding control volume

$$\int_{x_i}^{x_{i+1}} \frac{\partial n(t,x)}{\partial t}\mathrm{d}\ x = -\int_{x_i}^{x_{i+1}} \frac{\partial}{\partial x}\{G(t,x)n(t,x)\}\ \mathrm{d}x + \int_{x_i}^{x_{i+1}} D\ n(t,x)\mathrm{d}x\,. \tag{C.2}$$

Integration by parts for the first integral on the right hand side and and interchange of differentiation and integration on the left hand side of the equation yields

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{x_i}^{x_{i+1}} n\ \mathrm{d}x = -\left[G\ n\right]_{x_i}^{x_{i+1}} + \int_{x_i}^{x_{i+1}} D\ n(t,x)\ \mathrm{d}x \tag{C.3}$$

In the next step assumptions on the profiles of n and D in the interior of a control volume have to be made, e.g. piece wise constant profiles

$$n(t,x) = n(t,\bar{x}_i) = n_i(t)$$
$$D(t,x) = D(t,\bar{x}_i)$$
$$\text{for } x \in [x_i, x_{i+1})$$
$$i = 1,\ldots,N_{\mathrm{CV}} \tag{C.4}$$

Furthermore, the fluxes over the control volume boundaries are approximated using an upwind scheme

$$[G \ n]_{x_i}^{x_{i+1}} = (\left. G \ n \right|_{\bar{x}_i} - \left. Gn \right|_{\bar{x}_{i-1}}) \qquad \text{(C.5)}$$

resulting in the following set of ODEs

$$\frac{\mathrm{d}n_i(t)}{\mathrm{d}t} = \frac{1}{\Delta x_i} \left(\left. Gn \right|_{\bar{x}_{i-1}} - \left. Gn \right|_{\bar{x}_i} \right) + Dn_i(t). \qquad \text{(C.6)}$$

D. Pseudo spectral collocation

Spectral methods can be used for the solution of population balance models represented by PDEs. Similar to the FVM, the PDE is reduced to a set ODEs. In the following their application will be demonstrated for the general one dimensional case

$$\frac{\partial n(t,x)}{\partial t} = -\frac{\partial}{\partial x}\left\{G(t,x)\ n(t,x)\right\} + D\ n(t,x)\,. \tag{D.1}$$

For further information on spectral methods the interested reader is referred to [112, 67, 11] and the references mentioned therein.

The basic assumption is that the number density function can be approximated by a weighted sum of spectral functions $f_i(x)$

$$n(t,x) \approx \sum_{i=0}^{N_i} c_i(t) f_i(x) \tag{D.2}$$

where $a_i(t)$ are the spectral weights. The spectral functions $f_i(x)$ are chosen once for each problem and are on the whole region $x = [x_{\min}, x_{\max}]$. A common choice for the function set, which has to consist of mutually orthogonal elements, are Chebyshev polynomials. To determine the spectral weights, the integral of the weighted residual has to be zero

$$\int\limits_{x_{\min}}^{x_{\max}} w(x) \left[\frac{\partial n(t,x)}{\partial t} + \frac{\partial}{\partial x}\left\{G(t,x)n(t,x)\right\} - D\ n(t,x)\right] \mathrm{d}x = 0\,. \tag{D.3}$$

Here, $w(x)$ corresponds to the weighting function. However, if the weighting function is chosen as

$$\rho(x) = \delta(x - x_i) \tag{D.4}$$

the spectral approximation satisfies the population at certain grid points or collocation points x_i, for $i = 1, \ldots N_i$. The overall procedure is also known as spectral collocation. The choice of the collocation points depends on the spectral functions f_i. When using Chebyshev polynomials, the overall method is called pseudo spectral collocation and the collocation points are defined as

$$x_i = 0.5(x_{\max} - x_{\min})\tilde{x}_i + 0.5(x_{\max} + x_{\min}) \tag{D.5}$$

with the Chebychev points $\tilde{x}_i$ defined as

$$\tilde{x}_i = \cos\left\{\frac{i}{\pi}N_i\right\}\,, \ \text{for } i = 0, \ldots, N_i\,. \tag{D.6}$$
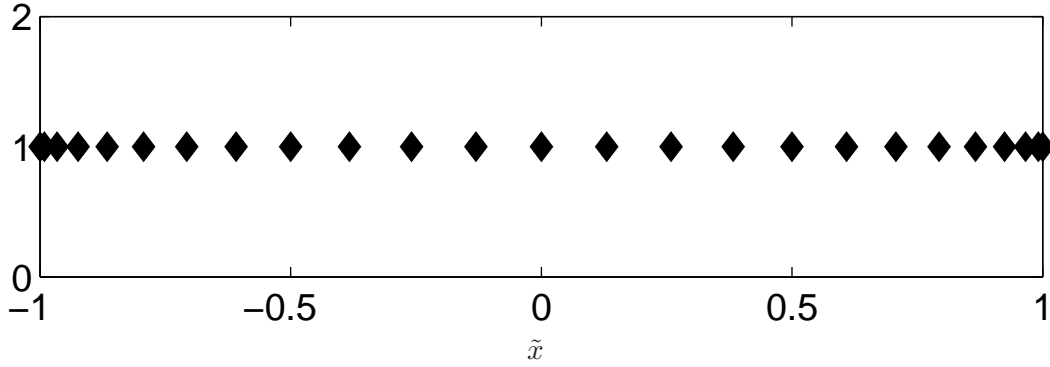
Figure D.1.: *Distribution of collocation points when using Chebychev polynomials*

The corresponding distribution is shown in Fig. D.1. The particular choice of the weighting function (D.4) enables the direct calculation of $n(t, x)$ at the grid nodes $n_i(t) = n(t, x_i)$. In result, the differentiation can be represented by the following matrix-vector multiplication

$$\frac{\partial n(t, x)}{\partial x} \approx D_x \cdot \mathbf{n}$$
$$\mathbf{n} = [n_1(t), \dots, n_{N_i}(t)]^T \tag{D.7}$$

Thereby, information on the spectral functions is incooperated in the generation of the differentiation matrix D_x. In contrast to the FVM, this differentiation matrix is densily filled [112]. The overall PDE is approximated by the finite set of ODEs

$$\frac{\mathrm{d}\mathbf{n}}{\mathrm{d}t} = D_x \cdot \mathbf{n} + D\mathbf{n} \tag{D.8}$$

As when using the finite volume method, the number of grid nodes has a significant effect on the overall accuracy. However, in general a considerably lower number of grid points is necessary when using pseudospectral collocation as the approximation of the derivatives is much more accurate compared to an FVM with the same number of grid nodes.

E. Higher order non product cubature rule

In the following the higher order non product cubature rule which is applied is adopted from [98] and applied in Section 4.7 will be presented.

Based on the moment approximation using the definitions of the generator functions

$$\int_{\mathbf{X}} f(\mathbf{x})p(\mathbf{x})\mathrm{d}\mathbf{x} \approx w_I GF_I(\mathbf{0}) + w_{II} \sum_{i=1}^{N_{II}} GF_{II}(\pm\theta) + w_{III} \sum_{i=1}^{N_{III}} GF_{III}(\pm\theta,\ \pm\theta)\,. \quad \text{(E.1)}$$

a non product cubature rule can be derived exploiting the properties of Gaussian distribution $n_0(\mathbf{x}) = \mathcal{N}(\mathbf{0}, \mathbf{I})$. The parameterisation of w_I and θ can be derived taking into account monomials up to degree 5 in $f(\mathbf{x})$

$$\int_{\mathbf{X}} n_0(\mathbf{x})\mathrm{d}\mathbf{x} = 1 \approx w_I + 2\, N_d w_{II} + 2\, N_d(N_d - 1)w_{III}\,,$$

$$\int_{\mathbf{X}} x_i^2 n_0(\mathbf{x})\mathrm{d}\mathbf{x} = 1 \approx 2\, N_d w_{II}\theta^2 + 4(N_d - 1)w_{III}\theta^2\,,$$

$$\int_{\mathbf{X}} x_i^4 n_0(\mathbf{x})\mathrm{d}\mathbf{x} = 3 \approx 2\, N_d w_{II}\theta^4 + 4(N_d - 1)w_{III}\theta^4\,,$$

$$\int_{\mathbf{X}} x_i^2 x_{j\neq i}^2 n_0(\mathbf{x})\mathrm{d}\mathbf{x} = 1 \approx 4w_{III}\theta^4\,. \quad \text{(E.2)}$$

The unknowns can be determined from the solution of the system

$$\theta = \sqrt{3}\,,$$

$$w_I = 1 + \frac{N_d^2 - 7N_d}{18}\,,$$

$$w_{II} = \frac{4 - N_d}{18}\,,$$

$$w_{III} = \frac{1}{36}\,. \quad \text{(E.3)}$$

F. Influenza A single cell kinetics

In the following, the dynamic equations describing influenza A virus replication cycle on the intracellular level as presented in [41] are listed. As described in [40], virus entry is described on the macroscopic scale and is thus cancelled out from the intracellular model.

F.1. Viral replication

$$\frac{\mathrm{d}Vp^{\mathrm{cyt}}}{\mathrm{d}t} = -k^{\mathrm{Imp}}Vp^{\mathrm{cyt}}, \tag{F.1}$$

$$\frac{\mathrm{d}Vp^{\mathrm{nuc}}}{\mathrm{d}t} = k^{\mathrm{Imp}}Vp^{\mathrm{cyt}} + k_{\mathrm{NP}}^{\mathrm{Bind}}P_{\mathrm{NP}}R_{\mathrm{RdRp}}^{\mathrm{V}} - \left(k_{\mathrm{M1}}^{\mathrm{Bind}}P_{\mathrm{M1}} + k_{\mathrm{RnP}}^{\mathrm{Deg}}\right)Vp^{\mathrm{Nuc}}, \tag{F.2}$$

$$\frac{\mathrm{d}R^{\mathrm{C}}}{\mathrm{d}t} = k_{\mathrm{C}}^{\mathrm{Syn}}Vp^{\mathrm{nuc}} - k_{\mathrm{RdRp}}^{\mathrm{Bind}}P_{\mathrm{RdRp}}R^{\mathrm{C}} - k_{\mathrm{R}}^{\mathrm{Deg}}R^{\mathrm{C}}, \tag{F.3}$$

$$\frac{\mathrm{d}R^{\mathrm{V}}}{\mathrm{d}t} = k_{\mathrm{V}}^{\mathrm{Syn}}Cp - k_{\mathrm{RdRp}}^{\mathrm{Bind}}P_{\mathrm{RdRp}}R^{\mathrm{V}} - k_{\mathrm{R}}^{\mathrm{Deg}}R^{\mathrm{V}}, \tag{F.4}$$

$$\frac{\mathrm{d}R_{\mathrm{RdRp}}^{\mathrm{C}}}{\mathrm{d}t} = k_{\mathrm{RdRp}}^{\mathrm{Bind}}P_{\mathrm{RdRp}}R^{\mathrm{C}} - k_{\mathrm{NP}}^{\mathrm{Bind}}P_{\mathrm{NP}}R_{\mathrm{RdRp}}^{\mathrm{C}} - k_{\mathrm{RRdRp}}^{\mathrm{Deg}}R_{\mathrm{RdRp}}^{\mathrm{C}}, \tag{F.5}$$

$$\frac{\mathrm{d}R_{\mathrm{RdRp}}^{\mathrm{V}}}{\mathrm{d}t} = k_{\mathrm{RdRp}}^{\mathrm{Bind}}P_{\mathrm{RdRp}}R^{\mathrm{V}} - k_{\mathrm{NP}}^{\mathrm{Bind}}P_{\mathrm{NP}}R_{\mathrm{RdRp}}^{\mathrm{V}} - k_{\mathrm{RRdRp}}^{\mathrm{Deg}}R_{\mathrm{RdRp}}^{\mathrm{V}}, \tag{F.6}$$

$$\frac{\mathrm{d}Cp}{\mathrm{d}t} = k_{\mathrm{NP}}^{\mathrm{Bind}}P_{\mathrm{NP}}R_{\mathrm{RdRp}}^{\mathrm{C}} - k_{\mathrm{Rnp}}^{\mathrm{Deg}}Cp, \tag{F.7}$$

$$\frac{\mathrm{d}Vp_{\mathrm{M1}}^{\mathrm{nuc}}}{\mathrm{d}t} = k_{\mathrm{M1}}^{\mathrm{Bind}}P_{\mathrm{M1}}Vp^{\mathrm{nuc}} - \left(k^{\mathrm{Exp}}P_{\mathrm{NEP}} + k_{\mathrm{Rnp}}^{\mathrm{Deg}}\right)Vp_{\mathrm{M1}}^{\mathrm{nuc}}, \tag{F.8}$$

$$\frac{\mathrm{d}Vp_{\mathrm{M1}}^{\mathrm{cyt}}}{\mathrm{d}t} = k^{\mathrm{Exp}}P_{\mathrm{NEP}}Vp_{\mathrm{M1}}^{\mathrm{nuc}} - 8r^{\mathrm{rel}} - k_{\mathrm{Rnp}}^{\mathrm{Deg}}Vp_{\mathrm{M1}}^{\mathrm{cyt}}. \tag{F.9}$$

F.2. Viral transcription and protein synthesis

$$\frac{\mathrm{d}R_i^{\mathrm{M}}}{\mathrm{d}t} = \frac{k_{\mathrm{M}}^{\mathrm{Syn}}}{L_i}\frac{Vp^{\mathrm{nuc}}}{8} - k_{\mathrm{M}}^{\mathrm{Deg}}R_i^{\mathrm{M}}, \quad i=1,\dots,8, \tag{F.10}$$

$$\frac{\mathrm{d}P_{\mathrm{PB1}}}{\mathrm{d}t} = \frac{k_{\mathrm{P}}^{\mathrm{Syn}}}{D_{\mathrm{Rib}}}R_2^{\mathrm{M}} - k^{\mathrm{RdRp}}P_{\mathrm{PB1}}P_{\mathrm{PB2}}P_{\mathrm{PA}}, \tag{F.11}$$

$$\frac{\mathrm{d}P_{\mathrm{PB2}}}{\mathrm{d}t} = \frac{k_{\mathrm{P}}^{\mathrm{Syn}}}{D_{\mathrm{Rib}}}R_1^{\mathrm{M}} - k^{\mathrm{RdRp}}P_{\mathrm{PB1}}P_{\mathrm{PB2}}P_{\mathrm{PA}}, \tag{F.12}$$

$$\frac{\mathrm{d}P_{\mathrm{PA}}}{\mathrm{d}t} = \frac{k_{\mathrm{P}}^{\mathrm{Syn}}}{D_{\mathrm{Rib}}}R_3^{\mathrm{M}} - k^{\mathrm{RdRp}}P_{\mathrm{PB1}}P_{\mathrm{PB2}}P_{\mathrm{PA}}, \tag{F.13}$$

$$\frac{\mathrm{d}P_{\mathrm{RdRp}}}{\mathrm{d}t} = k^{\mathrm{RdRp}}P_{\mathrm{PB1}}P_{\mathrm{PB2}}P_{\mathrm{PA}} - k_{\mathrm{RdRp}}^{\mathrm{Bind}}P_{\mathrm{RdRp}}\left(R^{\mathrm{V}}+R^{\mathrm{C}}\right)$$
$$- \left(N_{P_{\mathrm{RdRp}}}-8\right)r^{\mathrm{Rel}}, \tag{F.14}$$

$$\frac{\mathrm{d}P_{\mathrm{NP}}}{\mathrm{d}t} = \frac{k_{\mathrm{P}}^{\mathrm{Syn}}}{D_{\mathrm{Rib}}}R_5^{\mathrm{M}} - \frac{L_{\mathrm{V}}}{N_{\mathrm{NP}}^{\mathrm{Nuc}}}k_{\mathrm{NP}}^{\mathrm{Bind}}P_{\mathrm{NP}}\left(R_{\mathrm{RdRp}}^{\mathrm{V}}+R_{\mathrm{RdRp}}^{\mathrm{C}}\right)$$
$$- \left(N_{P_{\mathrm{NP}}}-8\frac{L_{\mathrm{V}}}{N_{\mathrm{NP}}^{\mathrm{Nuc}}}\right)r^{\mathrm{Rel}}, \tag{F.15}$$

$$\frac{\mathrm{d}P_{\mathrm{M1}}}{\mathrm{d}t} = \frac{k_{\mathrm{P}}^{\mathrm{Syn}}}{D_{\mathrm{Rib}}}\left(1-F_{\mathrm{Spl7}}\right)R_7^{\mathrm{M}} - \frac{L_{\mathrm{V}}}{N_{\mathrm{M1}}^{\mathrm{Nuc}}}k_{\mathrm{M1}}^{\mathrm{Bind}}P_{\mathrm{M1}}Vp^{\mathrm{Nuc}}$$
$$- \left(N_{P_{\mathrm{M1}}}-8\frac{L_{\mathrm{V}}}{N_{\mathrm{M1}}^{\mathrm{Nuc}}}\right)r^{\mathrm{Rel}}, \tag{F.16}$$

$$\frac{\mathrm{d}P_{\mathrm{NEP}}}{\mathrm{d}t} = \frac{k_{\mathrm{P}}^{\mathrm{Syn}}}{D_{\mathrm{Rib}}}F_{\mathrm{Spl8}}R_8^{\mathrm{M}} - \frac{L_{\mathrm{V}}}{N_{\mathrm{NEP}}^{\mathrm{Nuc}}}k^{\mathrm{Exp}}P_{\mathrm{NEP}}Vp_{\mathrm{M1}}^{\mathrm{Nuc}}$$
$$- \left(N_{P_{\mathrm{NEP}}}-8\frac{L_{\mathrm{V}}}{N_{\mathrm{NEP}}^{\mathrm{Nuc}}}\right)r^{\mathrm{Rel}}, \tag{F.17}$$

$$\frac{\mathrm{d}P_{\mathrm{HA}}}{\mathrm{d}t} = \frac{k_{\mathrm{P}}^{\mathrm{Syn}}}{D_{\mathrm{Rib}}}R_4^{\mathrm{M}} - N_{P_{\mathrm{HA}}}r^{\mathrm{Rel}}, \tag{F.18}$$

$$\frac{\mathrm{d}P_{\mathrm{NA}}}{\mathrm{d}t} = \frac{k_{\mathrm{P}}^{\mathrm{Syn}}}{D_{\mathrm{Rib}}}R_6^{\mathrm{M}} - N_{P_{\mathrm{NA}}}r^{\mathrm{Rel}}, \tag{F.19}$$

$$\frac{\mathrm{d}P_{\mathrm{M2}}}{\mathrm{d}t} = \frac{k_{\mathrm{P}}^{\mathrm{Syn}}}{D_{\mathrm{Rib}}}F_{Spl7}R_7^{\mathrm{M}} - N_{P_{\mathrm{M2}}}r^{\mathrm{Rel}}. \tag{F.20}$$

F.3. Parameters

Table F.1.: *Intracellular model parameters as listed in [40]*

Parameter	Value	Unit
k^{Imp}	6	h^{-1}
$k_{\text{NP}}^{\text{Bind}}$	$3.01 \cdot 10^{-4}$	$(molecule\ h)^{-1}$
$k_{\text{M1}}^{\text{Bind}}$	$2.43 \cdot 10^{-4}$	$(molecule\ h)^{-1}$
$k_{\text{RnP}}^{\text{Deg}}$	0.09	h^{-1}
$k_{\text{C}}^{\text{Syn}}$	5.29	h^{-1}
$k_{\text{RdRp}}^{\text{Bind}}$	1	$(molecule\ h)^{-1}$
$k_{\text{R}}^{\text{Deg}}$	36.36	h^{-1}
$k_{\text{V}}^{\text{Syn}}$	32.18	h^{-1}
$k_{\text{RRdRp}}^{\text{Deg}}$	4.25	h^{-1}
K_{Vrel}	300	$virions$
k^{Exp}	$1 \cdot 10^{-6}$	$(molecules\ h)^{-1}$
$k_{\text{M}}^{\text{Syn}}$	$8.53 \cdot 10^{5}$	$nucleotides\ h^{-1}$
L_1	2320	$nucleotides$
L_2	2320	$nucleotides$
L_3	2211	$nucleotides$
L_4	1757	$nucleotides$
L_5	1540	$nucleotides$
L_6	1392	$nucleotides$
L_7	1005	$nucleotides$
L_8	868	$nucleotides$
$k_{\text{M}}^{\text{Deg}}$	0.33	h^{-1}
$k_{\text{P}}^{\text{Syn}}$	64800	$nucleotides\ h^{-1}$
D_{Rib}	160	$nucleotides$
k^{RdRp}	1	$molecule^{-2}h^{-1}$
N_{RdRp}	45	$molecules\ virion^{-1}$
L_{V}	1700	$nucleotides$
$N_{P_{\text{NA}}}$	100	$molecules\ virion^{-1}$
$N_{P_{\text{HA}}}$	500	$molecules\ virion^{-1}$
$N_{P_{\text{M1}}}$	3000	$molecules\ virion^{-1}$
$N_{P_{\text{M2}}}$	40	$molecules\ virion^{-1}$
$N_{P_{\text{NEP}}}$	165	$molecules\ virion^{-1}$
$N_{P_{\text{NP}}}$	1000	$molecules\ virion^{-1}$
$N_{\text{M1}}^{\text{Nuc}}$	200	$nucleotides$
$N_{\text{NEP}}^{\text{Nuc}}$	1700	$nucleotides$
$N_{\text{NP}}^{\text{Nuc}}$	24	$nucleotides$
F_{Spl7}	0.02	–
F_{Spl8}	0.125	–

G. Moment approximation for multi dimensional population balance equation

For the analysis presented in Section 5.3 the moments were approximated with the efficient approximate moment method. Initial abscissas were chosen based on the combined Gaussian mixed density and sigma point approach with $N_{\text{GMD}} = 11$. It can be obtained from Fig. G.1 that thereby a sufficiently accurate approximation of the virus release rate can be guaranteed. In contrast, the direct sigma point approach yields a worse approximation.
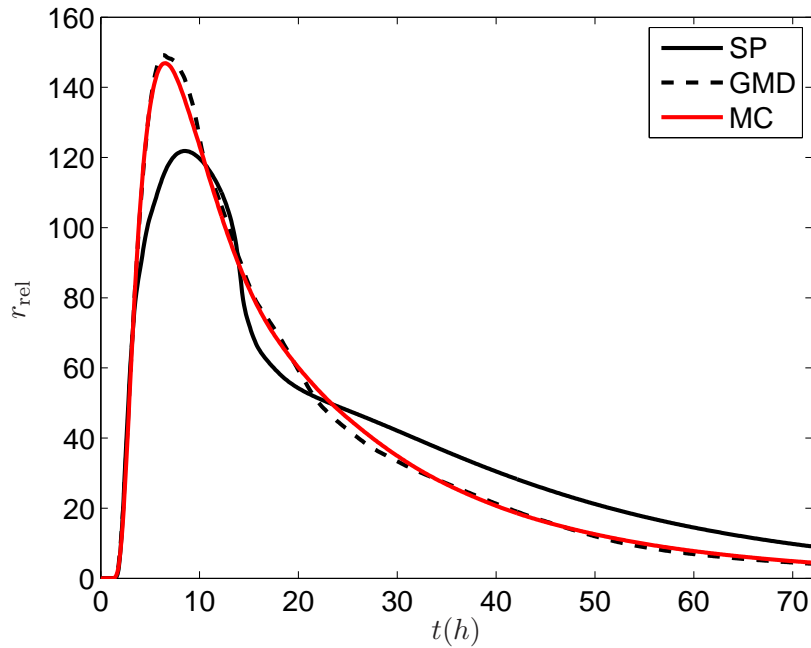

Figure G.1.: *Approximation of virus release rate for* p_{var} *for single infection cycle and different abscissa formulas; sigma points(SP) with* $N_\alpha = 11$ *, Gaussian mixed densities (GMD) with* $N_{\text{GMD}} = 11$, *Monte-Carlo(MC) with* $N_\alpha = 10^4$

Nomenclature

$a_c(t, \varphi)$	number density distribution of apoptotic cells
a_α	dynamics of abscissas
b_α	dynamics of weigths
$\mathbf{c}(t)$	states of the continuous phase
c_{Sol}	solute concentration
d_i	death rate coefficient
e_l	relative error
$f_{\mathbf{x}}(t, \mathbf{x})$	transport density
g_i	growth rate coefficients
$\mathbf{f_I}(t, \mathbf{x})$	integral coupling between particulate and continuous phase
$i_c(t, \varphi)$	number density distribution of infected cells
$k_{\{1,2,3,4,5,6,7\}}$	rate coefficients
$k_{\mathrm{apo}}(\varphi)$	apoptosis rate coefficient for infected cells
$k_{\mathrm{cd}}(\varphi)$	cell death rate coefficient for apoptotic cells
k_{cd,i_c}	cell death rate coefficient for infected cells
k_{cd,U_c}	cell death rate coefficient for uninfected cells
k_{deg}	virus degradation rate coefficient
$k_{\mathrm{deg},V}$	virus degradation rate coefficient
k_{gro,U_c}	growth rate coefficient for uninfected
k_{inf}	infection rate coefficient
$k_{\mathrm{net}}(\varphi)$	net rate coefficient for infected cells
$k_{\mathrm{net,apo}}(\varphi)$	net rate coefficient for apoptotic cells
$k_{\mathrm{Nuc}}(t)$	nucleation rate coefficient
k_{rel}	virus release rate coefficient
k_{vi}	infection rate coefficient
$l_c(t, \varphi)$	number density distribution of latent phase cells
$m_\circ(t)$	distribution moment of order $\circ$
$\widetilde{n}(t, x)$	number density function
$n(t, \mathbf{x})$	number density function
$n_0(\mathbf{x})$	initial number density function
$\mathbf{p}$	vector of unknown parameters
$\mathbf{p}_{\mathrm{const}}$	vector of unknown constant parameters
$\mathbf{p}_{\mathrm{funct}}$	vector of unknown functional parameters
$p_{\mathbf{x}}(t, \mathbf{x})$	production density
$r_{\mathrm{rel}}(\mathbf{x})$	virus release rate
t	time
$u_c(t, \varphi)$	number density distribution of uninfected cells
$w_\alpha(t)$	weight

$\widetilde{w}_\alpha(t,\tau)$	age distributed weight
x	characteristic particle property
$x_\alpha(t)$	abscissa
$\widetilde{x}_\alpha(t,\tau)$	age distributed abscissa
$\mathbf{x}$	vector of particle properties
$\mathbf{x_e}$	vector of external particle properties
$\mathbf{x_i}$	vector of internal particle properties
A	approximation operator
A_{OV}	DQMOM Matrix for one dimensional case
A_{MV}	DQMOM Matrix for multi dimensional case
$C(t)$	overall cell concentration
$D(t,\mathbf{x})$	death rate
$Dil(t,\mathbf{c})$	dilution rate
F	measured fluorescence intensity
$F_\mathbf{x}(t)$	flow of particles
F_{VE}	fluorescence intensity per virus equivalent
$G(t,x)$	growth rate
$\mathbf{G}(t,\mathbf{x})$	vector of growth rates
GF_i	generator function of order i
J	cost function
K_V	form factor
$L_c(t)$	overall number of latent phase cells
$\hat{L}_c(t,\lambda)$	latent phase cells in transport system
N_{CV}	number of control volumes
N_D	number of characteristic particle properties
$N(t)$	number of particles in a control volume
N_{GA}	number of abscissas for Gaussian cubature formula
N_{GMD}	number of Gaussian Mixed Density distributions
N_{MC}	number of Monte Carlo abscissas
N_{ODE}	number of ordinary differential equations
N_{SP}	number of sigma point abscissas
N_α	number of abscissas
$Nuc(t,x)$	number of particles in a control volume
$P_\mathbf{x}(t)$	net production rate of particles in control volume
P_{eff}	efficiency ratio of released active virus particles to total amount of released virus particles
$P_\mathbf{x}(t)$	net production rate of particles in control volume
$S(t,x)$	summarized growth, death and nucleation processes
$S_{\mathrm{sat}}(t)$	relative supersaturation
S^*	solubility
$\tilde{S}(t,x)$	approximation of $\widetilde{S}(t,x)$
$\tilde{S}_{\mathrm{G}}(t,x)$	approximation of growth processes
$\tilde{S}_{\mathrm{D}}(t,x)$	approximation of death processes
$\tilde{S}_{\mathrm{N}}(t,x)$	approximation of nucleation processes
T	temperature

$U_c(t)$	overall concentration of uninfected cells
$V_{ac}(t)$	concentration of active virus particles
$V_{inac}(t)$	concentration of inactive virus particles
$V_{tot}(t)$	total concentration of virus particles
W	weighting matrix
$\mathbf{X}$	property state space
$\mathbf{X_e}$	external property state space
$\mathbf{X_i}$	internal property state space
Z_i	number of cells within a certain range of fluorescence
$\mathcal{I}(\mathbf{x})$	Distribution of newly infected cells
$\mathcal{L}(\mu_\mathcal{L}, \mathbf{\Sigma}_\mathcal{L})$	logarithmic Gaussian normal distribution
$\mathcal{N}(\mu, \mathbf{\Sigma})$	Gaussian normal distribution
$\mathcal{S}$	set of abscissas
$\mathcal{W}$	set of weights
$[tem]$, $[gen]$, $[str]$, $[v_1]$, $[v_2]$	intracellular compounds
α_{SP}	tuning parameter for sigma point formula
$\delta(x - x_\alpha(t))$	delta distribution
λ	local coordinate for transport system
λ_{SP}	tuning parameter for sigma point formula
κ_{SP}	tuning parameter for sigma point formula
μ	mean value vector of Gaussian distribution
$\mu_\mathcal{L}$	mean value vector for logarithmic Gaussian distribution
ρ_c	solute density
τ	particle age
τ_{lat}	latent phase length
φ	degree of fluorescence
σ_{MC}	variance of the Monte Carlo estimation
$\Delta\varphi_i$	difference between two specific degrees of fluorescence
$\mathbf{\Sigma}$	covariance matrix of Gaussian distribution
$\mathbf{\Sigma}_\mathcal{L}$	covariance matrix of Log-Gaussian distribution

Bibliography

[1] Matlab documentation. *http://de.mathworks.com/help/stats/fitgmdist.html*. Accessed: 2016-5-30.

[2] Aamir, E., Nagy, Z. K., Rielly, C. D., Kleinert, T., and Judat, B. Combined Quadrature Method of Moments and Method of Characteristics approach for efficient solution of Population Balance Models for dynamic modeling and crystal size distribution control of crystallization processes. *Industrial & Engineering Chemistry Research 48*, 18 (2009), 8575–8584.

[3] Antonyuk, S., Palis, S., and Heinrich, S. Breakage behaviour of agglomerates and crystals by static loading and impact. *Powder Technology 206*, 1–2 (2011), 88–98.

[4] Banks, H. T., Sutton, K. L., Thompson, W. C., Bocharov, G., Doumic, M., Schenkel, T., Argilaguet, J., Giest, S., Peligero, C., and Meyerhans, A. A new model for the estimation of cell proliferation dynamics using CFSE data. *Journal of Immunological Methods 373*, 12 (2011), 143–160.

[5] Banks, H. T., Sutton, K. L., Thompson, W. C., Bocharov, G., Roose, D., Schenkel, T., and Meyerhans, A. Estimation of cell proliferation dynamics using CFSE data. *Bulletin of Mathematical Biology 73*, 1 (2011), 116–150.

[6] Bernardo, F. P., Pistikopoulos, E. N., and Saraiva, P. M. Integration and computational issues in stochastic design and planning optimization problems. *Industrial & Engineering Chemistry Research 38*, 8 (1999), 3056–3068.

[7] Bessonnard, S., De Mot, L., Gonze, D., Barriol, M., Dennis, C., Goldbeter, A., Dupont, G., and Chazaud, C. Gata6, Nanog and Erk signaling control cell fate in the inner cell mass through a tristable regulatory network. *Development 141*, 19 (2014), 3637–3648.

[8] Borchert, C., and Sundmacher, K. Morphology evolution of crystal populations: Modeling and observation analysis. *Chemical Engineering Science 70*, 0 (2012), 87–98.

[9] Braumann, A., Man, P. L. W., and Kraft, M. Statistical approximation of the inverse problem in multivariate population balance modeling. *Industrial & Engineering Chemistry Research 49*, 1 (2010), 428–438.

[10] Briesen, H. Simulation of crystal size and shape by means of a reduced two-dimensional population balance model. *Chemical Engineering Science 61*, 1 (2006), 104–112.

[11] BÜCK, A., KLAUNICK, G., KUMAR, J., PEGLOW, M., AND TSOTSAS, E. Numerical simulation of particulate processes for control and estimation by spectral methods. *AIChE Journal 58* (2012), 2309–2319.

[12] BUFFO, A., AND MARCHISIO, D. L. Modeling and simulation of turbulent polydisperse gas-liquid systems via the generalized population balance equation. *Reviews in Chemical Engineering 30*, 1 (2014), 73–126.

[13] BUSHMAN, F., LEWINSKI, M., CIUFFI, A., BARR, S., LEIPZIG, J., HANNENHALLI, S., AND HOFFMANN, C. Genome-wide analysis of retroviral DNA integration. *Nature Reviews Microbiology 3*, 11 (2005), 848–858.

[14] CHICKARMANE, V., TOEIN, C., NUBER, U. A., SAURO, H. M., AND PETERSON, C. Transcriptional Dynamics of the Embryonic Stem Cell Switch. *PLoS Computational Biology 2*, 9 (2006), e123.

[15] CUNDALL, P. A., AND STRACK, O. D. L. A discrete numerical model for granular assemblies. *Geotechnique 29*, 1 (1979), 47–65.

[16] DAVIS, P. J., AND RABINOWITZ, P. *Methods of Numerical Integration*, 2nd ed. Computer Science and Applied Mathematics. Academic Press, 1994.

[17] DEEN, N., VAN SINT ANNALAND, M., VAN DER HOEF, M. A., AND KUIPERS, J. A. M. Review of discrete particle modeling of fluidized beds. *Chemical Process Engineering 62*, 1-2 (2007), 28–44.

[18] DÜRR, R., DUVIGNEAU, S., LASKE, T., BACHMANN, M., AND KIENLE, A. Analyzing the impact of heterogeneity in genetically engineered cell lines for influenza vaccine production using population balance modeling. In *6th IFAC Conference on Foundations of Systems Biology in Engineering* (2016), p. submitted.

[19] DÜRR, R., FRANZ, A., AND KIENLE, A. Combination of limited measurement information and multidimensional population balance models. In *Proceedings to the 9th IFAC Symposium on Biological and Medical Systems BMS 2105, Berlin* (2015), vol. 48, pp. 261–266.

[20] DÜRR, R., AND KIENLE, A. An efficient method for calculating the moments of multidimensional growth processes in population balance systems. *The Canadian Journal of Chemical Engineering 92* (2014), 2088–2097.

[21] DÜRR, R., MÜLLER, T., ISKEN, B., SCHULZE-HORSEL, J., REICHL, U., AND KIENLE, A. Distributed modeling and parameter estimation of influenza virus replication during vaccine production. In *Proceedings to the 7th Vienna International Conference on Mathematical Modelling* (2012), pp. 320–325.

[22] DÜRR, R., MÜLLER, T., AND KIENLE, A. Efficient DQMOM for multivariate population balance equations and application to virus replication in cell cultures. In *Proceedings to the 8th Vienna International Conference on Mathematical Modelling* (2015), pp. 29–34.

[23] Eissing, T., Chaves, M., and Allgöwer, F. Live and let die–A systems biology view on cell death. *Computers & Chemical Engineering 33*, 3 (2009), 583–589.

[24] Eissing, T., Conzelmann, H., Gilles, E. D., Allgöwer, F., Bullinger, E., and Scheurich, P. Bistability analysis of a caspase activation model for receptor-induced apoptosis. *Journal of Biological Chemistry 279*, 35 (2004), 36892–36897.

[25] Fadda, S., Cincotti, A., and Cao, G. A novel population balance model to investigate the kinetics of in vitro cell proliferation: Part I. Model development. *Biotechnology and Bioengineering 109*, 3 (2012), 772–781.

[26] Flassig, R. J., and Sundmacher, K. Optimal design of stimulus experiments for robust discrimination of biochemical reaction networks. *Bioinformatics 28*, 23 (2012), 3089–3096.

[27] Franz, A., Dürr, R., and Kienle, A. Population balance modeling of biopolymer production in cellular systems. In *Proceedings of the 19th IFAC World Congress* (2014), vol. 19, pp. 1705–1710.

[28] Franz, A., Song, H.-S., Ramkrishna, D., and Kienle, A. Experimental and theoretical analysis of poly(-hydroxybutyrate) formation and consumption in Ralstonia eutropha. *Biochemical Engineering Journal 55*, 1 (2011), 49–58.

[29] Fredrickson, A. G., Ramkrishna, D., and Tsuchiya, H. M. Statistics and dynamics of procaryotic cell populations. *Mathematical Biosciences 1*, 3 (1967), 327–374.

[30] Genzel, Y., Frensing, T., and Reichl, U. Herstellung moderner Grippeimpfstoffe. *Chemie in unserer Zeit 47*, 1 (2013), 12–22.

[31] Gillespie, D. Stochastic simulation of chemical kinetics. *Annual Review of Physical Chemistry 58* (2007), 35–55.

[32] Gimbun, J., Nagy, Z. K., and Rielly, C. D. Simultaneous quadrature method of moments for the solution of population balance equations, using a differential algebraic equation framework. *Industrial & Engineering Chemistry Research 48*, 16 (2009), 7798–7812.

[33] Grosch, R., Briesen, H., Marquardt, W., and Wulkow, M. Generalization and numerical investigation of qmom. *AIChE Journal 53*, 1 (2007), 207–227.

[34] Gunawan, R., Fusman, I., and Braatz, R. D. High resolution algorithms for multidimensional population balance equations. *AIChE Journal 50*, 11 (2004), 2738–2749.

[35] Gunawan, R., Fusman, I., and Braatz, R. D. Parallel high-resolution finite volume simulation of particulate processes. *AIChE Journal 54*, 6 (2008), 1449–1458.

[36] Gunawan, R., Ma, D. L., Fujiwara, M., and Braatz, R. D. Identification of kinetic parameters in multidimensional crystallization processes. *International Journal of Modern Physics B 16*, 1 (2002), 367–374.

[37] Hampel, N., Bück, A., Peglow, M., and Tsotsas, E. Continuous pellet coating in a wurster fluidized bed process. *Chemical Engineering Science 86* (2013), 87–98.

[38] Haseltine, E. L., Rawlings, J. B., and Yin, J. Dynamics of viral infections: incorporating both the intracellular and extracellular levels. *Computers & Chemical Engineering 29*, 3 (2005), 675–686.

[39] Haseltine, E. L., Yin, J., and Rawlings, J. B. Implications of decoupling the intracellular and extracellular levels in multi-level models of virus growth. *Biotechnology and Bioengineering 101*, 4 (2008), 811–820.

[40] Heldt, F. S., Frensing, T., Pflugmacher, A., Gröpler, R., Peschel, B., and Reichl, U. Multiscale modeling of influenza a virus infection supports the development of direct-acting antivirals. *PLoS Computational Biology 9*, 11 (2013).

[41] Heldt, S. F., Frensing, T., and Reichl, U. Modeling the intracellular dynamics of influenza virus replication to understand the control of viral RNA synthesis. *Journal of Virology 86*, 15 (2012), 7806–7817.

[42] Herberg, M., and Roeder, I. Computational modelling of embryonic stem-cell fate control. *Development 142*, 13 (2015), 2250–2260.

[43] Hierholzer, J., and Killington, R. 2 - Virus isolation and quantitation. In *Virology Methods Manual*. Academic Press, London, 1996, pp. 25–46.

[44] Hulburt, H. M., and Katz, S. Some problems in particle technology: A statistical mechanical formulation. *Chemical Engineering Science 19*, 8 (1964), 555–574.

[45] John, V., and Suciu, C. Direct discretizations of bi-variate population balance systems with finite difference schemes of different order. *Chemical Engineering Science 106* (2014), 39–52.

[46] Jones, A. G. *Crystallization Process Systems*. Butterworth-Heinemann, 2002.

[47] Jordan, I., Horn, D., John, K., and Sandig, V. A Genotype of Modified Vaccinia Ankara (MVA) that Facilitates Replication in Suspension Cultures in Chemically Defined Medium. *Viruses 5*, 1 (2013), 321–339.

[48] Julier, S., and Uhlmann, J. Unscented filtering and nonlinear estimation. *Proceedings of the IEEE 92*, 3 (2004), 401–422.

[49] Julier, S., Uhlmann, J., and Durrant-Whyte, H. A new method for the nonlinear transformation of means and covariances in filters and estimators. *IEEE Transactions on Automatic Control 45*, 3 (2000), 477–482.

[50] JULIER, S. J., AND UHLMANN, J. K. A new extension of the Kalman Filter to nonlinear systems. In *The Proceedings of AeroSense: The 11th International Symposium on Aerospace/Defense Sensing, Simulation and Controls* (1997), pp. 182–193.

[51] JULIER, S. J., AND UHLMANN, J. K. A consistent, debiased method for converting between polar and cartesian coordinate systems. In *The Proceedings of AeroSense: The 11th International Symposium on Aerospace/Defense Sensing, Simulation and Controls* (1997), pp. 110–121.

[52] KALBFUSS, B., KNÖCHLEIN, A., KRÖBER, T., AND REICHL, U. Monitoring influenza virus content in vaccine production: Precise assays for the quantitation of hemagglutination and neuraminidase activity. *Biologicals 36*, 3 (2008), 145–161.

[53] KAVOUSANAKIS, M. E., MANTZARIS, N. V., AND BOUDOUVIS, A. G. A novel free boundary algorithm for the solution of cell population balance models. *Chemical Engineering Science 64*, 20 (2009), 4247–4261.

[54] KUROKAWA, M., KOYAMA, A., YASUOKA, S., AND ADACHI, A. Influenza virus overcomes apoptosis by rapid multiplication. *International Journal of Molecular Medicine 3*, 5 (1999), 527–530.

[55] LENCASTRE FERNANDES, R., CARLQUIST, M., LUNDIN, L., HEINS, A.-L., DUTTA, A., SØRENSEN, S. J., JENSEN, A. D., NOPENS, I., LANTZ, A. E., AND GERNAEY, K. V. Cell mass and cell cycle dynamics of an asynchronous budding yeast population: Experimental observations, flow cytometry data analysis, and multi-scale modeling. *Biotechnology and Bioengineering 110*, 3 (2013), 812–826.

[56] LEVEQUE, R. *Finite Volume Methods for Hyperbolic Problems*. Cambridge University Press, New York, NY, 2002.

[57] LIOU, J., SRIENC, F., AND FREDRICKSON, A. G. Solutions of population balance models based on a successive generations approach. *Chemical Engineering Science 52*, 9 (1997), 1529–1540.

[58] LIU, J. J., MA, C. Y., HU, Y. D., AND WANG, X. Z. Modelling protein crystallisation using morphological population balance models. *Chemical Engineering Research and Design 88*, 4 (2010), 437–446.

[59] LUZYANINA, T., ROOSE, D., AND BOCHAROV, G. Distributed parameter identification for a label-structured cell population dynamics model using CFSE histogram time-series data. *Journal of Mathematical Biology 59*, 5 (2009), 581–603.

[60] MA, C. Y., AND WANG, X. Z. Model identification of crystal facet growth kinetics in morphological population balance modeling of l-glutamic acid crystallization and experimental validation. *Chemical Engineering Science 70*, 0 (2012), 22–30.

[61] MA, C. Y., WANG, X. Z., AND ROBERTS, K. J. Morphological population balance for modeling crystal growth in face directions. *AIChE Journal 54*, 1 (2008), 209–222.

[62] MA, D. L., AND BRAATZ, R. D. Robust identification and control of batch processes. *Computers & Chemical Engineering 27*, 8–9 (2003), 1175–1184.

[63] MA, D. L., TAFTI, D. K., AND BRAATZ, R. D. Optimal control and simulation of multidimensional crystallization processes. *Computers & Chemical Engineering 26*, 78 (2002), 1103–1116.

[64] MAHONEY, A. W. *Inverse Problem Modeling of Particulate Systems*. PhD thesis, Purdue University Graduate School, 2001.

[65] MAHONEY, A. W., DOYLE, F. J., AND RAMKRISHNA, D. Inverse problems in population balances: Growth and nucleation from dynamic data. *AIChE Journal 48*, 5 (2002), 981–990.

[66] MANTZARIS, N. V., DAOUTIDIS, P., AND SRIENC, F. Numerical solution of multi-variable cell population balance models: I. Finite difference methods. *Computers & Chemical Engineering 25*, 11-12 (2001), 1411–1440.

[67] MANTZARIS, N. V., DAOUTIDIS, P., AND SRIENC, F. Numerical solution of multi-variable cell population balance models. II. Spectral methods. *Computers & Chemical Engineering 25*, 11-12 (2001), 1441–1462.

[68] MANTZARIS, N. V., DAOUTIDIS, P., AND SRIENC, F. Numerical solution of multi-variable cell population balance models. III. Finite element methods. *Computers & Chemical Engineering 25*, 11-12 (2001), 1463–1481.

[69] MARCHISIO, D. Quadrature method of moments for poly-disperse flows. In *Multiphase Reacting Flows: Modelling and Simulation*, D. Marchisio and R. O. Fox, Eds., vol. 492 of *CISM International Centre for Mechanical Sciences*. Springer Vienna, 2007, pp. 41–77.

[70] MARCHISIO, D. L., AND FOX, R. O. Solution of population balance equations using the direct quadrature method of moments. *Journal of Aerosol Science 36*, 1 (2005), 43–73.

[71] MARCHISIO, D. L., PIKTURNA, J. T., FOX, R. O., VIGIL, R. D., AND BARRESI, A. A. Quadrature method of moments for population-balance equations. *AIChE Journal 49*, 5 (2003), 1266–1276.

[72] MARTIN, K., AND HELENIUS, A. Nuclear transport of influenza virus ribonucleoproteins: the viral matrix protein (m1) promotes export and inhibits import. *Cell 67*, 1 (1991), 117–130.

[73] MCGRAW, R. Description of aerosol dynamics by the quadrature method of moments. *Aerosol Science and Technology 27*, 2 (1997), 255–265.

[74] MENG, X. L., AND RUBIN, D. B. Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika 80*, 2 (1993), 267–278.

[75] MHASKAR, P., HJORTSO, M. A., AND HENSON, M. A. Cell population modeling and parameter estimation for continuous cultures of saccharomyces cerevisiae. *Biotechnology Progress 18*, 5 (2002), 1010–1026.

[76] MÖHLER, L., FLOCKERZI, D., SANN, H., AND REICHL, U. Mathematical model of influenza a virus production in large-scale microcarrier culture. *Biotechnology and Bioengineering 90*, 1 (2005), 46–58.

[77] MOTZ, S. *Reduktion Populationsdynamischer Modelle.* PhD thesis, Universität Stuttgart, 2004.

[78] MÜLLER, S., HARMS, H., AND BLEY, T. Origin and analysis of microbial population heterogeneity in bioprocesses. *Current Opinion in Biotechnology 21*, 1 (2010), 100–113.

[79] MÜLLER, T. *Population balance modeling of influenza A virus replication in MDCK cells during vaccine production.* PhD thesis, Otto-von-Guericke University Magdeburg, 2015.

[80] MÜLLER, T., DÜRR, R., ISKEN, B., SCHULZE-HORSEL, J., REICHL, U., AND KIENLE, A. Population balance modelling of influenza virus replication during vaccine production - Influence of apoptosis. In *Proceedings to the 21st European Symposium on Computer Aided Process Engineering* (2011), Elsevier, pp. 1336–1340.

[81] MÜLLER, T., DÜRR, R., ISKEN, B., SCHULZE-HORSEL, J., REICHL, U., AND KIENLE, A. Distributed modeling of human influenza a virus-host cell interactions during vaccine production. *Biotechnology and Bioengineering 110*, 8 (2013), 2252–2266.

[82] MÜLLER, T., SCHULZE-HORSEL, J., SIDORENKO, Y., REICHL, U., AND KIENLE, A. Population balance modeling of influenza virus replication in mdck cells during vaccine production. In *Proceedings to the 18th European Symposium on Computer Aided Process Engineering* (2008), pp. 133–138.

[83] MURRAY, J. D. *Mathematical Biology I: An Introduction*, 3rd ed. Springer-Verlag Berlin, 2001.

[84] MURRAY, J. D. *Mathematical Biology II: Spatial Models and Biomedical Applications*, 3rd ed. Springer-Verlag Berlin, 2001.

[85] NAGY, Z. K. Model based robust control approach for batch crystallization product design. *Computers & Chemical Engineering 33*, 10 (2009), 1685–1691.

[86] NANDANWAR, M. N., AND KUMAR, S. A new discretization of space for the solution of multi-dimensional population balance equations. *Chemical Engineering Science 63*, 8 (2008), 2198–2210.

[87] Nowak, M. A., and May, R. M. *Virus Dynamics: Mathematical Principles of Immunology and Virology.* Oxford University Press, 2004.

[88] Paramasivan, G., and Kienle, A. Decentralized control system design under uncertainty using mixed-integer optimization. *Chemical Engineering & Technology 35*, 2 (2012), 261–271.

[89] Peglow, M., Kumar, J., Hampel, R., Tsotsas, E., and Heinrich, S. Towards a complete population balance model for fluidized-bed spray agglomeration. *Drying Technology 25*, 7-8 (2007), 1321–1329.

[90] Penrose, R. A generalized inverse for matrices. *Mathematical Proceedings of the Cambridge Philosophical Society 55* (1955), 406–413.

[91] Radichkov, R., Müller, T., Kienle, A., Heinrich, S., Peglow, M., and Mörl, L. A numerical bifurcation analysis of continuous fluidized bed spray granulation with external product classification. *Chemical Engineering and Processing: Process Intensification 45*, 10 (2006), 826–837.

[92] Ramkrishna, D. *Population Balances: Theory and Applications to Particulate Systems in Engineering.* Academic Press, San Diego, 2000.

[93] Ramkrishna, D., and Singh, M. R. Population balance modeling: Current status and future prospects. *Annual Review of Chemical and Biomolecular Engineering 5* (2014), 123–146.

[94] Rhee, H.-K., Aris, R., and Amundson, N. R. *First-Order Partial Differential Equations*, vol. 1. Dover Publications, Inc., Mineola, NY, 2002.

[95] Rossner, N. *Robuste modellgestützte Prozessführung auf Basis von Gaußschen Mischdichten am Beispiel der Bray-Liebhafsky-Reaktion und der autotrophen Kultivierung von Ralstonia eutropha H16.* PhD thesis, Fakultät III - Prozesswissenschaften der Technischen Universität Berlin, 2014.

[96] Rossner, N., Heine, T., and King, R. Quality-by-design using a Gaussian Mixture Density approximation of biological uncertainties. In *Proceedings of the 11th International Symposium on Computer Applications in Biotechnology* (2010), vol. 11, pp. 7–12.

[97] Sanyal, J., Marchisio, D. L., Fox, R. O., and Dhanasekharan, K. On the comparison between population balance models for cfd simulation of bubble columns. *Industrial & Engineering Chemistry Research 44*, 14 (2005), 5063–5072.

[98] Schenkendorf, R. *Optimal experimental design for parameter identification and model selection.* PhD thesis, Otto-von-Guericke University Magdeburg, 2014.

[99] Schenkendorf, R., and Kremling, A. Mangold, M. Optimal experimental design with the sigma point method. *IET Systems Biology 3*, 1 (2009), 10–23.

[100] Schliemann, M., Bullinger, E., Borchers, S., Allgöwer, F., and Scheurich, P. Heterogeneity reduces sensitivity of cell death for TNF-Stimuli. *BMC Systems Biology 5* (2011).

[101] SCHULTZ-CHERRY, S., KRUG, R. M., AND HINSHAW, V. S. Induction of apoptosis by influenza virus. *Seminars in Virology 8*, 6 (1998), 491 – 495.

[102] SCHULZE-HORSEL, J. *Zellphysiologische Charakterisierung von Zellkulturen in der Influenza-Impfstoffproduktion.* PhD thesis, Otto-von-Guericke University Magdeburg, 2011.

[103] SCHULZE-HORSEL, J., GENZEL, Y., AND REICHL, U. Flow cytometric monitoring of influenza A virus infection in MDCK cells during vaccine production. *BMC Biotechnology 8*, 45 (2008).

[104] SCHULZE-HORSEL, J., SCHULZE, M., AGALARIDIS, G., GENZEL, Y., AND REICHL, U. Infection dynamics and virus-induced apoptosis in cell culture-based influenza vaccine production - Flow cytometry and mathematical modeling. *Vaccine 27*, 20 (2009), 2712–2722.

[105] SHERER, E., TOCCE, E., HANNEMANN, R. E., RUNDELL, A. E., AND RAMKRISHNA, D. Identification of age-structured models: Cell cycle phase transitions. *Biotechnology and Bioengineering 99*, 4 (2008), 960–974.

[106] SMITH, G. L., VANDERPLASSCHEN, A., AND LAW, M. The formation and function of extracellular enveloped vaccinia virus. *Journal of General Virology 83*, 12 (2002), 2915–2931.

[107] STROUD, A. *Approximate Calculation of Multiple Integrals.* Prentice-Hall, Englewood Cliffs, N.J., 1971.

[108] TAFFOREAU, L., CHANTIER, T., PRADEZYNSKI, F., PELLET, J., MANGEOT, P. E., VIDALAIN, P.-O., ANDRE, P., RABOURDIN-COMBE, C., AND LOTTEAU, V. Generation and comprehensive analysis of an influenza virus polymerase cellular interaction network. *Journal of virology 85*, 24 (dec 2011), 13010–8.

[109] TAPIA, F., VÁZQUEZ-RAMÍREZ, D., GENZEL, Y., AND REICHL, U. Bioreactors for high cell density and continuous multi-stage cultivations: options for process intensification in cell culture-based viral vaccine production. *Applied Microbiology and Biotechnology 100*, 5 (2016), 2121–2132.

[110] TAPIA, F., VOGEL, T., GENZEL, Y., BEHRENDT, I., HIRSCHEL, M., GANGEMI, J. D., AND REICHL, U. Production of high-titer human influenza A virus with adherent and suspension MDCK cells cultured in a single-use hollow fiber bioreactor. *Vaccine 32*, 8 (2014), 1003–1011.

[111] THEIN, F. On the Efficiency and Condition of the Core Routine of the Quadrature Methods of Moments (QMOM). Diploma thesis, Otto von Guericke Universität Magdeburg, 2011.

[112] TREFETHEN, L. N. *Spectral Methods in MATLAB.* SIAM, Philadelphia, 2001.

[113] VAN DER MERWE, R. *Sigma-Point Kalman Filters for Probablilistic Inference in Dynamic State-Space Models.* PhD thesis, Oregon Health & Science University, 2004.

[114] VAN HEERDEN, J. H., WORTEL, M. T., BRUGGEMAN, F. J., HEIJNEN, J. J., BOLLEN, Y. J. M., PLANQUÉ, R., HULSHOF, J., O'TOOLE, T. G., WAHL, S. A., AND TEUSINK, B. Lost in Transition: Start-Up of Glycolysis Yields Subpopulations of Nongrowing Cells. *Science 343*, 6174 (2014).

[115] VREMAN, A., VAN LARE, C., AND HOUNSLOW, M. A basic population balance model for fluid bed spray granulation. *Chemical Engineering Science 64*, 21 (2009), 4389–4398.

[116] WALTER, E., AND PRONZATO, L. *Identification of parametric models: from experimental data.* Springer, Berlin, 1997.

[117] WHEELER, J. Modified moments and Gaussian quadratures. *Rocky Mountain Journal of Mathematics 4*, 2 (1974), 287–296.

[118] WRIGHT, D., MCGRAW, R., AND ROSNER, D. Bivariate extension of the quadrature method of moments for modeling simultaneous coagulation and sintering of particle populations. *Journal of Colloid and Interface Science 236*, 2 (2001), 242–251.

[119] YOON, C., AND MCGRAW, R. Representation of generally mixed multivariate aerosols by the quadrature method of moments: I. Statistical foundation. *Journal of Aerosol Science 35*, 5 (2004), 561–576.

[120] YOON, C., AND MCGRAW, R. Representation of generally mixed multivariate aerosols by the quadrature method of moments: II. Aerosol dynamics. *Journal of Aerosol Science 35*, 5 (2004), 577–598.

[121] YUAN, C., AND FOX, R. Conditional quadrature method of moments for kinetic equations. *Journal of Computational Physics 230*, 22 (2011), 8216–8246.

[122] ZHANG, Y., ZAMAMIRI, A. M., HENSON, M. A., AND HJORTSO, M. A. Cell population models for bifurcation analysis and nonlinear control of continuous yeast bioreactors. *Journal of Process Control 12*, 6 (2002), 721–734.

[123] ZHAO, H., MAISELS, A., MATSOUKAS, T., AND ZHENG, C. Analysis of four Monte Carlo methods for the solution of population balances in dispersed systems. *Powder Technology 173*, 1 (2007), 38–50.

[124] ZHU, G. Y., ZAMAMIRI, A., HENSON, M. A., AND HJORTSO, M. A. Model predictive control of continuous yeast bioreactors using cell population balance models. *Chemical Engineering Science 55*, 24 (2000), 6155–6167.

Curriculum vitae

Personal details

Name	Robert Dürr
Date of birth	April 23rd, 1985
Place of birth	Torgau, Germany

Education

1991-1995	Elementary school "An der Promenade", Torgau
1995-2003	Gymnasium "Johann Walter", Torgau

Military service

2003-2004	Basic military service

Higher education

2004-2009	Otto von Guericke University, Magdeburg
	Diploma studies "Systemtechnik und Technische Kybernetik"
	Academic degree: Diplomingenieur (Dipl.-Ing.)

Professional experience

2010-2016	Scientific assistant,
	Chair for Automation/Modeling,
	Otto von Guericke University, Magdeburg
	and
	Process Systems and Process Dynamics Group,
	Max Planck Institute for Complex Technical Systems, Magdeburg