HELMHOLTZ
RESEARCH FOR GRAND CHALLENGES

HELMHOLTZ MUNICH

Parameter estimation and prediction using mixed effects modelling method with different data types

Dantong Wang

February 2024

TECHNISCHE UNIVERSITÄT MÜNCHEN

TUM School of Computation, Information and Technology

Parameter estimation and prediction using mixed effects modelling method with different data types

Dantong Wang

Vollständiger Abdruck der von der TUM School of Computation, Information and Technology der Technischen Universität München zur Erlangung einer

Doktorin der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz:

Prof. Dr. Oliver Junge

Prüfende der Dissertation:

1. Prof. Dr.-Ing. Jan Hasenauer
2. Prof. Dr. Christina Kuttler
3. Prof. Dr. Nicole Radde

Die Dissertation wurde am 29.02.2024 bei der Technischen Universität München eingereicht und durch die TUM School of Computation, Information and Technology am 05.11.2024 angenommen.

Acknowledgments

I want to thank ...

... fist of all my supervisor Jan Hasenauer, for giving me the opportunity to work in his group, for showing me such a new and interesting research field, for all the new knowledge and help, for the great supervision, discussions, and for his patient.

... my mentor Carsten Marr for providing the chance to work in his group, for being the most supportive supervisor and for all the fruitful discussions.

... my former and present colleagues and friends at the ICB and the university of Bonn, Anna, Benjamin, Carolin, Clemens, Daniel, Dilan, Dominik, Elba, Emad, Emmanuel, Erika, Fabian, Jakob, Lea, Leonard, Lisa, Lorenzo, Manuel, Marc, Marlet, Melanie, Paul Jost, Paul Stapor, Philipp, Polina, Salome, Simon, Stephan, Sophia, Tingying, Vanessa, Valerio, Yannik, Yu for the discussions, events and the beautiful time with me.

... the Institute of Computational Health, with its director Fabian Theis, for providing an exciting environment to work.

... Anandhi Iyappan, Claudia Reske, Marianne Antunes, Nina Fischer, Elisabeth Noheimer, Sabine Kunz, Anna Sacher, and Meike Busch-Auf der Mauer for the support and organization at the ICB and University of Bonn.

... Christina Kuttler for being in my thesis advisory committee, and for the valuable discussions.

... my collaboration partners Stefan Kallenberger in the caspase-8 project, Frank Ziemann, Klaus Metzeler and Christian Rausch in the AML project for the fruitful discussions and the successful collaboration.

... my family and friends, especially my parents for all their supports in these years.

... my classmate, then my boyfriend, then my husband, Yuanwei Cao for being with me for all these year, for sharing all my happiness and sadness, for all your patience and supports.

... Nuoxi, my daughter for being born healthy, for all the happy time with me and for sharing your sweet laugh with me.

... Yuze, my son for being healthy and being with me to finish this thesis.

Abstract

Mechanistic mathematical models are widely used to study biological processes, which are usually modeled by ordinary differential equations (ODEs). They allow to improve our understanding of the latent causality and biological process in systems biology. ODE models usually depend on parameters, which cannot be measured directly from experiments, but need to be estimated by comparing the model simulation with measurement data.

It has been shown that even genetically equivalent cells can respond differently to the same stimuli. Furthermore, patients with the same disease also have different response to the same treatment. For this purpose, mixed effects models (MEMs) has been developed to describe the cell-to-cell and patient-to-patient variability. In MEMs, similarities and differences between cells are captured by fixed effects (which are shared between cells) and random effects (which differ between cells). Beyond the single-cell (patient) data, the development of experimental technology in recent years makes it possible to generate multiple types of data, which also give insights on the whole population level. However, parameter estimation of MEMs usually uses only single-cell or population data separately, but not an integration of the multiple data types. Furthermore, the covariance matrix of random effects are usually assumed to be diagonal, whereas some of the effects can be biologically correlated.

In this thesis, we first develop a method to integrate different types of data, including single-cell time-lapse, time-to-event, population average and single-cell snapshot data. By applying the joint likelihood function to a test and an application example, we show that by integrating the above mentioned data types in parameter estimation, both prediction accuracy and parameter identifiability are improved. Moreover, if one data type is left out, it cannot be predicted using the other data types. This means that each data type contains specific insight of the system, which cannot be replaced by other data types.

Secondly, we introduce a sampling method based on Cramér-von Mises Distance (CMD) to compute the simulated mean and standard deviation. We apply this method on four benchmark models and compared it with Monte Carlo, quasi-Monte Carlo and sigma point methods. We show that this method is of higher computation efficiency compared with Monte Carlo and quasi Monte Carlo methods as a small number of samples is enough to get the same accuracy. Moreover, compared with the sigma point methods, CMD methods is more flexible regarding the sample sizes. This method can be used to estimate parameters using population average and single-cell snapshot data.

Thirdly, we compare five different approaches to parametrize the covariance matrix of random effects, namely the Cholesky decomposition, matrix logarithm, spherical, Givens and Lie-algebraic parametrization approaches. By applying these method to five benchmark models and one application example of the JAK/STAT model, we show that the Lie-algebraic method outperforms the other methods regarding both convergence performance and the accuracy of parameter estimation.

Finally, we implement MEMs in a clinical application of modelling the acute myeloid leukemia treatment process. We developed ODE models to describe the production and transition of the cell-free tumor DNA (ctDNA). Time-resolved measurement data of blast cell numbers and ctDNA concentrations are taken from 10 acute myeloid leukemia (AML) patients, then the estimated parameters are used to infer biological insights. Results show that by using MEM methods, we are able to predict the blast cell numbers by using only the ctDNA data and population parameters estimated using other patient data. We are also able to predict the relapse situation with a model parameter related to mutation heterogeneity.

In conclusion, the contributions presented in this thesis allow for more efficient and accurate parameter estimation of MEMs compared to the previous methods. We anticipate that these contributions will improve the prediction ability of MEMs and broaden our understanding of complex biological systems.

Zusammenfassung

Mechanistische mathematische Modelle werden oft zur Untersuchung biologischer Prozesse verwendet, die in der Regel durch gewöhnliche Differentialgleichungen (ODEs) modelliert werden. Sie ermöglichen es, unser Verständnis latenter Kausalitäten und biologischer Prozesse in der Systembiologie zu verbessern. ODE-Modelle hängen in der Regel von Parametern ab, die nicht direkt durch Experimente gemessen werden können, sondern durch den Vergleich der Modellsimulation mit Messdaten geschätzt werden müssen.

Es hat sich gezeigt, dass selbst genetisch gleichartige Zellen unterschiedlich auf dieselben Reize reagieren können. Außerdem reagieren Patienten mit derselben Krankheit auch unterschiedlich auf dieselbe Behandlung. Zu diesem Zweck wurden Modelle mit gemischten Effekten (MEMs) entwickelt, um die Variabilität von Zelle zu Zelle und von Patient zu Patient anhand der oben genannten Datentypen zu beschreiben. In MEMs werden Ähnlichkeiten und Unterschiede zwischen den Zellen durch feste Effekte (die zwischen den Zellen gleich sind) und zufällige Effekte (die zwischen den Zellen unterschiedlich sind) erfasst. Bei der Parameterschätzung von MEMs werden jedoch in der Regel nur Einzelzell- oder Populationsdaten separat verwendet, nicht aber eine Integration der verschiedenen Datentypen. Außerdem wird die Kovarianzmatrix der zufälligen Effekte in der Regel als diagonal angenommen, während einige der Effekte biologisch korreliert sein können.

In dieser Arbeit entwickeln wir zuerst eine Methode zur Integration verschiedener Datentypen, einschließlich Einzelzell-Zeitablauf, Zeit-zu-Ereignis-, Populations-Durchschnitts- und Einzelzell-Schnappschuss-Daten. Durch Anwendung der gemeinsamen Likelihood-Funktion auf einen Test und ein Anwendungsbeispiel zeigen wir, dass durch die Integration der oben genannten Datentypen in die Parameterschätzung sowohl die Vorhersagegenauigkeit als auch die Identifizierbarkeit der Parameter verbessert werden. Wenn ein Datentyp weggelassen wird, kann er nicht durch die anderen Datentypen vorhergesagt werden. Dies bedeutet, dass jeder Datentyp spezifische Erkenntnisse über das System enthält, die nicht durch andere Datentypen ersetzt werden können.

Zweitens führen wir eine Stichprobenmethode ein, die auf der Cramér-von Mises-Distanz (CMD) basiert, um den simulierten Mittelwert und die Standardabweichung zu berechnen. Wir wenden diese Methode auf vier Benchmark-Modelle an und vergleichen sie mit Monte-Carlo-, Quasi-Monte-Carlo- und Sigma-Point-Methoden. Wir zeigen, dass diese Methode im Vergleich zu Monte-Carlo- und Quasi-Monte-Carlo-Methoden eine höhere Berechnungseffizienz aufweist, da eine geringe Anzahl von Stichproben ausreicht, um die gleiche Genauigkeit zu erzielen. Darüber hinaus ist die CMD-Methode im Vergleich zu den Sigma-Punkt-Methoden flexibler bzgl. der Stichprobengröße. Diese Methode kann zur Schätzung von Parametern unter Verwendung von Populationsdurchschnitts- und Einzelzell-Schnappschuss-Daten verwendet werden.

Drittens vergleichen wir fünf verschiedene Ansätze zur Parametrisierung der Kovarianzmatrix von Zufallseffekten, nämlich die Cholesky-Zerlegung, den Matrix-Logarithmus, sphärische, Givens- und Lie-algebraische Parametrisierung. Durch Anwendung dieser Methoden auf fünf Benchmark-Modelle und ein Anwendungsbeispiel des JAK/STAT-Modells zeigen wir, dass die Lie-algebraische Methode die anderen Methoden sowohl hinsichtlich der Konvergenz als auch der Genauigkeit der Parameterschätzung übertrifft.

Schließlich setzen wir MEMs in einer klinischen Anwendung zur Modellierung des Behandlungsprozesses der akuten myeloischen Leukämie ein. Wir haben ODE-Modelle entwickelt, um die Produktion und die Transition der zellfreien Tumor-DNA (ctDNA) zu beschreiben. Zeitaufgelöste Messdaten der Anzahl der Blastenzellen und der ctDNA-Konzentrationen wurden von 10 acute myeloid leukemia (AML)-Patienten entnommen, dann wurden die geschätzten Parameter ver-

wendet, um biologische Erkenntnisse abzuleiten. Die Ergebnisse zeigen, dass wir mit MEM-Methoden in der Lage sind, die Anzahl der Blastenzellen vorherzusagen, indem wir ausschließlich die ctDNA-Daten und die anhand anderer Patientendaten geschätzten Populationsparameter verwenden. Wir sind außerdem in der Lage, Rückfälle mit einem Modellparameter vorherzusagen, der mit der Mutationsheterogenität zusammenhängt.

Zusammenfassend lässt sich sagen, dass die in dieser Arbeit vorgestellten Beiträge eine effizientere und genauere Parameterschätzung von MEMs im Vergleich zu den bisherigen Methoden ermöglichen. Wir gehen davon aus, dass diese Beiträge die Vorhersagekraft von MEMs verbessern und unser Verständnis für komplexe biologische Systeme erweitern werden.

Contents

Chapter 1

Introduction

1.1 Research topics

1.1.1 Mixed effects model

Cell-to-cell variability is widely observed across physiologic processes and pathological conditions. Sources of this variability include noisy cellular processes (Elowitz et al., 2002), differences in cell cycle state (Buettner et al., 2015), the history of individual cell (Spencer et al., 2009), and spatio-temporal differences of the environment of cell (Snijder and Pelkmans, 2011). Phenotypic heterogeneity often leads to different fate determination, including cell growth, senescence, and death in response to both natural and artificial stimulus.

Patient-to-patient variability influences the processing of drugs and the response to treatments (Willmann et al., 2007), which makes it difficult to predict the effectiveness of therapy. The variability attributes to the heterogeneity in drug response from one cell to the next (Palmer and Sorger, 2017), the genetic variability (Ji et al., 2016) and even the difference of plasma concentrations (Leeder, 2019) . In clinical trials, it is usually hard to obtain enough data points for single patients because of measurement complexity, which leads to practical unidentifiability in parameter estimation using single-patient data. Furthermore, it can be difficult to find the common property of disease by computing a specific set of parameters for each single patient. Therefore, informing parameter estimation using data from multiple patients can possibly improve the predictive capability of the model.

Nonlinear mixed effects models (MEMs) are widely used to capture cell-to-cell (Karlsson et al., 2015; Llamosi et al., 2016; Fröhlich et al., 2018b) and patient-to-patient variability (Tornoe et al., 2004; Willmann et al., 2007; Bastogne et al., 2009). These models provide a hierarchical description of whole populations (Pinheiro, 1994). Variability is encoded in parameter values, which are composed of fixed effects and random effects. The fixed effects are the same for the whole population, while the random effects vary between individuals. Research by Lavielle and Aarons (2016) shows that, with data from multiple patients, a model can be practically identifiable at the population level even if it is not structurally identifiable at the single-patient level.

1.1.2 Different data types provide different levels of insights

In the study of cellular systems, commonly, different experimental techniques have to be combined in order to access different aspects of biological processes at different levels. The different types are provided by a broad spectrum of experimental techniques.

Time-lapse (TL) data (including time-to-event data as a special case) are for instance provided by time-lapse (fluorescence) microscopy or sequencing (Qiu et al., 2020). Also, patient data, which are measured in a sequence of days/time slots can also be regarded as TL data in MEMs. TL data facilitate various insights, from the assessment of developmental trajectories (Navin et al., 2011) to mechanistic insights of causal differences (Spencer and Sorger, 2011; Hasenauer et al., 2014; Loos et al., 2018b).

Snapshot (SH) data of single-cell can be obtained using, e.g., image and flow cytometry (Davey and Kell, 1996; Bodenmiller et al., 2012), single-cell sequencing (Buettner et al., 2015; Frei et al., 2016) and single-cell proteomics (Palit et al., 2019). When collecting long-term patients information (Coates et al., 2002), SH data are also more convenient compared to TL data while no tracking of individuals are needed. SH data usually contain the mean and (or) variance of a population. Unlike TL data, SH data are easier to obtain via experiments (for cells) and surveys (for patients) from a large number of population. These data can also be used to investigate model identifiability (Cinquemani, 2018), model reconstruction (Klimovskaia et al., 2016; Ocone et al., 2015) and also parameter estimation (Hasenauer et al., 2011a).

Population average (PA) data are available via methods such as bulk sequencing (Li and Wang, 2021), micro-array analysis (Daran-Lapujade et al., 2008), and western blotting (Kurien and Scofield, 2006), which provide the mean of cell or patient population. Same to SH data, the corresponding experiments or measurements can be done for a large number of cell or patient population. PA data have been used for the study of signaling effects (Keller et al., 2017). The methods offer different advantages regarding throughput, content, and visibility, and all of them are highly valuable in modern workflows.

1.1.3 Parameter inference for MEMs

Mechanistic models usually contain unknown parameters, which are essential when the models are used to predict the states, which cannot be measured directly. The parameter estimation process is often done by comparing the simulated data with measurement data, e.g. maximizing a likelihood function. In MEMs, parameters consist of two parts, the fixed effects and the covariance matrix of random effects.

In linear MEMs, where the output is a linear combination of fixed effects and random effects, parameter estimation can be done by joint methods, meaning that the fixed effects and random effects can be estimated simultaneously. These methods include Henderson's mixed model equations (Henderson et al., 1959), Goldberger's approach (Goldberger, 1962), and two-stage regression (Fröhlich et al., 2018b).

In contrast to linear MEMs, non-linear MEMs are commonly used to describe more complex biological processes. Parameter estimation of non-linear MEMs are commonly done by maxi-

mizing the likelihood, which is the probability of seeing the measurements given the parameters. Methods to compute likelihood function values can be chosen according to the type of data being used. The likelihood of models using TL data relies on numerically challenging integrals along random effects (Pinheiro, 1994). Methods have been developed to efficiently approximate it, including Laplacian (Tierney and Kadane, 1986), alternating (Lindstrom and Bates, 1990), first-order conditional approximation (Beal and Sheiner, 1992), importance sampling (Geweke, 1989), and Gaussian quadrature (Davidian and Gallant, 1992). These methods either sample from the random effect distribution or estimate the random effects first via an inner optimization problem. Regarding the PA and SH data, the population mean and variance first needs to be estimated for the model with specific parameters. The most intuitive way of computing the mean and variance of the population is to use Monte-Carlo (MC) sampling. However, this often requires a substantial number of model evaluations, which sometimes make the computation time-wise unaffordable. Sigma point (SP) methods (Julier et al., 1995; Menegaz et al., 2011; Lerner, 2002) require fewer model evaluations by deterministic sampling (van der Merwe, 2004), however, use a fixed number of samples, which does not allow to flexibly adjust the accuracy.

1.2 Challenges

With the improving capability of biological experimental technologies, more data are generated, which includes not only data from different cells, and patients, but also data of different types. This gives us the possibility to investigate biological systems in a more holistic way. From single-cell/patient data, cell-to-cell (patient-to-patient) variability can often be found. To estimate parameters using TL data, the standard two-stage approach estimates a set of parameters for each single-cell/patient separately. Whereas the parameters can be estimated for the whole population simultaneously in MEMs (Fröhlich et al., 2018b) by combining the fixed and random effects. The heterogeneity is encoded in the distributions of random effects. By using MEMs, TL data, and PA data can be combined to provide a better understanding of the biological system (Adlung et al., 2021) from different levels of insights. Furthermore, measurements from experiments with different perturbation (Fröhlich et al., 2018b) or that of different DNA modifications (Kochmanski et al., 2019) can also be modeled simultaneously in order to estimate both the similarity or the source of heterogeneity. Therefore, with the implementation of MEMs, a broader insight of the data is provided, while the reproducibility of the experimental findings is improved as well. Yet, there still exist countless open questions and challenges in MEMs, which we address the following in this thesis:

(i) Methods to generate different data types have different throughput, content, and visibility, and all of them are highly valuable in modern workflows, as they provide different insights into the cell population being measured. Researchers have used TL data under different experimental conditions (Almquist et al., 2015; Fröhlich et al., 2018b), or combined PA and SH data (Adlung et al., 2021) to estimate the parameters of mechanistic models. Results have shown that this data integration improves parameter fits and biological interpretations. However, it has so far not been investigated whether one can integrate more different data types, e.g. TL data (including event data), PA data, and SH data, and how different data types will affect the parameter estimation of models.

(ii) When estimating parameters in MEMs, an efficient way of computing the mean and variance

of the simulated population is necessary. To obtain robust estimates for the population statistics, e.g. mean and standard deviation, a substantial number of Monte-Carlo (MC) samples is required, causing a significant computational demand. SP methods are an alternative to MC methods (Julier et al., 1995; Julier and Uhlmann, 2004; Menegaz et al., 2011; Lerner, 2002; van der Merwe, 2004; Charalampidis and Papavassilopoulos, 2011). These methods aim to approximate the mean and covariance matrix of the parameters by sampling deterministically (van der Merwe, 2004). Then the approximated mean and covariance matrix of outputs are computed by propagating the samples through the model(Filippi et al., 2016; Loos et al., 2018b). However, the error is difficult to control for nonlinear models.

(iii) In many studies, the covariance matrix of the random effects in MEMs is assumed to be diagonal (Fröhlich et al., 2018b). However, it is common that some of the parameters are correlated in biological systems, hence the covariance values need to be estimated simultaneously with the variance values. Therefore, a crucial component of a MEM inference workflow is the parameterization of covariance matrices, for which there exist different approaches (Pinheiro and Bates, 1996; Lindstrom and Bates, 1990; Bates and Watts, 1988; Thisted, 2017). However, a comparison of these approaches regarding accuracy and efficiency is still needed.

Beside the aforementioned technical challenges, we also show in this thesis how can MEMs be applied in real patient data and what insight it provides as following:

(iv) Acute myeloid leukemia (AML) is a highly heterogeneous disease. During chemotherapy for patients with AML, not only the number of blast cells in bone marrow but also the amount of Cell-free tumor DNA (ctDNA) vary in different patients. When cell apoptosis happens, ctDNA is released from the blast cells, which can be detected in the blood. Models have been developed to investigate the ctDNA kinetics and showed that ctDNA can be used as a marker for the early assessment of therapy response and prognostic tool in AML patients (Holdenrieder et al. 2001; Mueller et al. 2006). However, it is still not clear how can we relate ctDNA detected in peripheral blood to the blast cell numbers in a patient's bone marrow using MEMs.

1.3 Contribution of this thesis

The overall aim of this thesis could be described as developing efficient and accurate methods to estimate the parameters using different data types in MEMs. We tackle various problems in the field of optimization for MEMs. Specifically, we address in this thesis the challenges outlined in Section 1.2 by the following contributions.

(I) *Joint likelihood function to combine three different data types.* This contribution addresses challenge (ii). We introduce a joint likelihood function to combine TL data, PA data, and snapshot data. Using a model of a simple conversion reaction, and an application example of the caspase-8 pathway model, we show that it is necessary to combine different data types, especially for more complex models, in order to ensure both estimation accuracy and parameter identifiability.

(II) *Dirac mixture distribution methods to approximate the continuous Gaussian distribution.* This contribution addresses challenge (iii). We minimize the Cramér–von Mises distance of Dirac mixture distributions and Gaussian distributions to obtain samples, which are then used

to compute the mean and variance of the distribution. We demonstrate superior performance regarding efficiency and flexibility of accuracy compared to the Monte Carlo method and SP methods, respectively.

(III) *Comparison of approaches to parametrize covariance matrix.* This contribution addresses challenge (iv). We compared five approaches of covariance matrix parametrization, including Cholesky, spherical, Givens, matrix-logarithm, and Lie algebraic approaches. By applying these approaches to benchmark models and an application example, we show that both the Lie algebraic and Givens approach perform comparable in benchmark models, while the Lie algebraic method performs best in the application example regarding the accuracy of parameter estimation.

(IV) *Modeling of chemotherapy using clinical data from NPM1 and IDH2 mutated AML patients.* This contribution addresses challenge (i). We use MEMs to model the chemotherapy process of AML patients and estimate the parameters by fitting to patient data. We show that by integrating information from various patients, it is possible to predict bone marrow blast cell numbers by using peripheral blood ctDNA data only.

Thus, (I) deals with the combination of different data types and demonstrates the effects of using only subsets of data. (II) focuses on an efficient way of parameter estimation using PA and SH data. (III) complements (I) and (II) by considering the possible correlation of parameters in MEMs. (Iv) shows an application of MEMs in clinical problems and demonstrates its ability to predict with heterogeneous single-patient data.

Some of these contributions are part of articles that have either been already peer-reviewed and published, are currently under peer review, or are in preparation. Therefore, parts of this thesis are based on and partly identical to the following first-author or shared first-author publications of the thesis author:

- **Wang, D**, Stapor, P, Hasenauer, J, (2019), Dirac mixture distributions for the approximation of mixed effects models. IFAC-PapersOnLine, 2019, 52(26), 200-206.
- **Wang, D**, Kallenberger, S, Frölich, F, Stapor, P, Schälte, Y, Eils, R, Hasenauer, Mixed-effect modelling enables integration of multiple data types. (in preparation)
- **Wang, D**, Stapor, P, Schälte, Y, Hasenauer, A lie-algebraic covariance parametrization method improves convergence for mixed effects models. (in preparation)
- **Wang, D**, Raush, C, Buerger, S, Tschuri, S, Rothenberg-Thurley, M, Schulz, M, Hasenauer, Ziemann, F, Metzeler, K, Marr, C, Modeling early treatment response in AML from cell-free tumor DNA. iScience, 2023, 26(12).

Furthermore, results of the following publications of the thesis author are used, albeit not discussed in detail:

- Buck M, Bast L, Hecker J, Rivière J, Rothenberg-Thurley M, Vogel L, **Wang D**, et al., Progressive disruption of hematopoietic architecture from clonal hematopoiesis to MDS. Iscience, 2023, 26(8).

- Schälte Y, Fröhlich F, Jost P, Vanhoefer J, Pathirana D, Stapor P, Lakrisenko P, **Wang D**, et al., pyPESTO: A modular and scalable tool for parameter estimation for dynamic models. arXiv preprint arXiv:2305.01821, 2023.
- Adlung L, Stapor P, Tönsing C, Schmiester L, Schwarzmüller L, Postawa L, **Wang D**, et al., Cell-to-cell variability in JAK2/STAT5 pathway components and cytoplasmic volumes define survival threshold in erythroid progenitor cells. Cell reports, 2021, 36(6).
- Schmiester L, Schälte Y, Bergmann F T, Camba T, Dudkin E, Egert J, Fröhlich F, Fuhrmann L, Hauber A, Kemmer S, Lakrisenko P, Loos C, Merkt S, Müller W, Pathirana D, Raimúndes E, Refisch L, Rosenblatt M, Stapor P, Städter P, **Wang D**, et al., PEtab—Interoperable specification of parameter estimation problems in systems biology. PLoS computational biology, 17(1): e1008646, 2021

1.4 Outline of this thesis

The remainder of this thesis is structured as follows. In Chapter 2 we introduce terminology and provide background knowledge on the MEM concept, with a focus on parameter estimation and likelihood computation with respect to different data types. Chapter 6-5 contain the main contributions of this thesis, the first one focuses on a clinical application example, while the remaining focus on methods development and comparison. In Chapter 6 we apply MEMs to real clinical data of AML patients and proved its ability of prediction. In Chapter 3, we introduce a joint likelihood function to combine three different data types. In Chapter 4, we introduce a sampling method based on the minimization of Cramér–von Mises distance, to efficiently compute the population mean and variance using MEMs. In Chapter 5, we compare different methods of covariance matrix parametrization regarding both accuracy and convergence using benchmark models and an application example. The thesis is concluded in Chapter 7 with a short summary and an outlook on potential further steps.

Chapter 2

Background

In this chapter, we introduce the terminology and provide the background on the methodologies, that are relevant throughout this thesis. We begin by introducing the notations of the modeling process, including mixed effects models (MEMs), observations, and noise models. We then give a general introduction to parameter estimation, including likelihood function and gradient-based optimization. As we are mainly looking at MEMs in this thesis, we further focus on the likelihood function of MEMs regarding different data types, including single-cell/patient time-lapse (TL), population average (PA), single-cell/patient snapshot (SH) data, and their gradient computation, respectively. Sampling methods are then introduced as this is relevant for the objective function value of PA and SH data. While this chapter provides a general and rather high-level overview, the subsequent main contribution chapters may contain further specific background relevant to the respective topics.

Parts of this chapter are based on and partly identical to the following publications of the thesis author:

- **Wang, D**, Stapor, P, Hasenauer, J, (2019), Dirac mixture distributions for the approximation of mixed effects models. IFAC-PapersOnLine, 52(26), 200-206.
- **Wang, D**, Kallenberger, S, Frölich, F, Stapor, P, Schälte, Y, Eils, R, Hasenauer, J, Mixed-effect modelling enables integration of multiple data types. (in preparation)
- **Wang, D**, Stapor, P, Schälte, Y, Hasenauer, J, A lie-algebraic covariance parametrization method improves convergence for mixed effects models. (in preparation)
- **Wang, D**, Raush, C, Buerger, S, Tschuri, S, Rothenberg-Thurley, M, Schulz, M, Ziemann, F, Metzeler, K, Marr, C, Modeling early treatment response of AML using cell-free tumor DNA kinetics. (in preparation)

2.1 Mechanistic modelling

In this section, we define the mathematical model including the process description via ordinary differential equations (ODEs), observations, noise model, and likelihood function. A key issue in

systems biology is to understand the interaction within and between cells, which determines their structure and the basic functions (Wolkenhauer et al., 2003; Degasperi et al., 2017). As typically not all quantities of interest can be measured directly via experiments, mathematical models are commonly employed to model the dynamics of a system in order to provide information on the non-measurable data. Futhermore, as mechanistic models seek to establish a mechanistic relationship between inputs and outputs, it can be leveraged for subsequent use in applications where experiments are either impossible or difficult to achieve once validated (Baker et al., 2018).

Mechanistic mathematical models are widely used to quantitatively describe the temporal dynamics of biological processes. For this purpose, a mechanistic model typically consists of interacting molecular components and a formal description of the possible mechanisms affecting the time evolution of the components (Intosalmi et al., 2016). They can also integrate prior knowledge on signaling pathways and the effects of somatic mutations and gene expression (Fröhlich et al., 2018). Furthermore, they allow for the description of individuals (Torregrosa and Garcia-Ojalvo, 2021; Falco et al., 2020) and populations (Gong and Sobie, 2018) and provide novel insights on both levels. While models for individuals are mostly informed by TL data (Albeck et al., 2013), models for populations often exploit SH (Hasenauer et al., 2011b) or PA data (Adlung et al., 2021).

These models have been used to identify drug targets (Schöberl et al., 2009) and the development of prognostic signatures(Fey et al., 2015; Eduati et al., 2017). Moreover, they also facilitated the investigation of oncogene addiction (Weinstein and Joe, 2006), synthetic-lethal phenotypes (Kaelin Jr, 2005), and many other relevant phenomena (Zhang et al., 2009).

2.1.1 Ordinary differential equations

To describe biological systems, many modeling approaches have been developed. Dynamic models consisting of ordinary differential equations (ODEs) are one of the frequently used approaches, which facilitates the analysis of the mechanism of the modeled process, as they allow us to get a quantitative insight of the systems. For example, the reaction of a cell pathway after specific stimuli and the response of patients to a certain therapy are common problems that can be simulated using ODE models.

The state variable vector of single cells is defined via its temporal evolution

$$\begin{aligned} \dot{x} &= f(x(t,\varphi),\varphi) \\ &= Nv(x,\varphi), \end{aligned} \tag{2.1}$$

where the state vector x is usually the concentration or amount of molecular or number of cells in patients. t in the modeled time points, and φ is the unknown parameter vector, which is usually the reaction rates of the biological systems or the initial values of the states. Parameters usually cannot be measured directly by experiments, but need to be estimated by comparing the simulated and experimental data. The right-hand side can usually be decomposed into a stoichiometry matrix N and reaction rate equations v of the molecular interactions (Heinrich and Schuster, 2012).

For most ODE models of practical interest in systems biology, it is non-trivial to find the solution in closed form. Therefore, it is necessary to use numerical methods to solve the Equation (2.1). These methods include Runge-Kutta, Adams-Moulton, and backward differentiation formula (Hindmarsh et al., 2005; Maiwald and Timmer, 2008; Mendes et al., 2009). (Städter et al., 2021) have found that the implicit methods are favorable because the biological problems often exhibit stiff dynamics.

2.1.2 Observation

The model parameters φ are usually unknown and cannot be measured directly via biological experiments. Therefore, it is necessary to do model calibration based on the observations which can be measured experimentally.

However, often experimental techniques do not provide direct measurements of the states, but only a subset, transformations, or combination of x with scaling factors or offsets. For example, time-lapse data are often scaled with parts or combinations of the true biological states. Therefore, we map states x to observations

$$y(t, \varphi) = h(x(t, \varphi), \varphi), \tag{2.2}$$

where $h()$ is the observable function. Sometimes the unknown parameters include also scaling factors and offsets. These parameters can be estimated with the other model parameters simultaneously, or a hierarchical approach can be used to analytically compute the scaling factors (Loos et al., 2018a).

2.1.3 Noise model

In real experimental and clinical trials, the observed data are in most cases not the exact observations, but are subject to noise. This yields noise-corrupted measurement data. The noise can come from numerous potential sources during data collection and data processing (Ghosh and Vogt, 2012). These include either technical limitations or human errors, for instance pipetting errors or incorrect labeling, which could potentially result in large errors (Motulsky and Christopoulos, 2003).

Different noise types have been employed in models, including additive normal noise, multiplicative log-normal noise, Poisson noise, Huber noise, Cauchy noise, and Laplace noise. It has been shown by Maier et al. (2017) that the heavy-tailed distributions, such as Laplace distributed noise is more stable with respect to measurement outliers. However, it is common and reasonable to assume normally distributed noise in biochemical assays. For parameter estimation additive normally distributed noise is also more convenient (Raue et al., 2013).

In the topics described in this thesis, we assume independent, additive normally distributed measurement noise, which can be loosely motivated by the Central Limit Theorem. So the measurement data $\bar{y}$ are given by

$$\bar{y}_{ij} = y_i(t_j, \varphi) + \sigma_{ij} \cdot \epsilon_{ij}, \quad \epsilon_{ij} \overset{iid}{\sim} \mathcal{N}(0,1), \quad i = 1, \ldots, d_y, \quad j = 1, \ldots, d_t, \tag{2.3}$$

where $\bar{y}_{ij}$ is the ith measurement at the time point t_j, d_y is the dimension of the measurements, and d_t the number of time points. σ_{ij} denotes the standard deviation of the normal distributed noise. It can be time and parameter dependent. For the models considered in this thesis, it is assumed to be constant at different time points for simplicity for synthetic models, while for the application models, there is no indication to assume time-dependent noise. Sometimes the noise standard deviation is not parameter-dependent and can be estimated by experimental replicates. Otherwise, it can be taken as a parameter and estimated with the other parameters simultaneously.

The above noise assumption yields the likelihood function

$$\mathcal{L}(\mathcal{D} \mid \varphi) = \prod_{i=1}^{d_y} \prod_{j=1}^{d_t} \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} \exp\left(-\frac{1}{2} \left(\frac{\bar{y}_{ij} - y_i\left(t_j, \varphi\right)}{\sigma_{ij}} \right)^2 \right). \tag{2.4}$$

2.2 Model cell and patient heterogeneity using mixed effects modelling based on different data types

Cellular heterogeneity widely exists even in a population of isogenic cells (Zhou et al., 2011; Raj and Van Oudenaarden, 2008; Spencer et al., 2009). The origins of this variability are believed to come partly from differences in cell phenotypes, from spatio-temporal variations of the cell environments, and from the intrinsic randomness of biochemical reactions (Llamosi et al., 2016). Similarly, patient-to-patient variability is also ubiquitous (Ancukiewicz et al., 2011; Detry and Ma, 2016). MEMs allow to account for cell-to-cell and patient-to-patient heterogeneity and is widely used in pharmacokinetics and pharmacodynamics (Sheiner and Beal, 1983). It can be combined with mechanistic modeling methods, for instance, ODEs as introduced in Section 2.1.1. This yields non-linear MEMs.

Unlike simple ODE models, parameters in MEMs are classified into two parts to incorporate both population and single-cell level information. Fixed effects describe the means of the model parameters, while random effects give information on single cells, meaning that they are different for each single cell. Random effects are usually assumed to follow certain distributions.

Throughout this thesis, we consider nonlinear MEMs as defined by Lindstrom and Bates (1990)

$$y(t, \varphi) = h(f(t, \varphi), \varphi), \tag{2.5}$$

where φ is the mixed effects vector (we call φ effects since here, to distinguish it from the parameters $\theta = [\beta, b]$, which need to be estimated in MEMs), which is given as a linear combination of fixed effects β and random effects b,

$$\varphi = A\beta + Bb, \quad b \sim \mathcal{N}(0, D), \tag{2.6}$$

with covariance matrix D, and design matrices A and B, where A is typically the identity matrix and B is typically a rectangular matrix with 0 or 1 entries that adds random effects to a subset of φ elements. When no correlation between the random effects is assumed for simplicity, the covariance matrix D can be written as

$$D = \exp(\text{diag}(\delta_m)), \quad m = 1, \cdots, d_\delta. \tag{2.7}$$

Here, δ is the variance vector, and the dimension of covariance matrix parameters d_δ is equal to the dimension of random effects d_r. This is the case in our thesis for Chapters 6, 3, and 4. However, in biological systems, some model parameters, e.g. initial concentrations, are known to be correlated. Thus not only the diagonal values but also covariance values of the covariance matrix D need to be estimated. This increases the dimension of covariance matrix parameters d_δ from the dimension of random effects d_r to $d_r(d_r - 1)/2$.

In the study of cellular systems as well as drug/treatment responses, commonly, different experimental and measurement techniques are employed to generate different types of data in order to access different aspects of biological processes. Therefore, we introduce different data types that can be measured in experiments and used for model parameter estimation, including time-lapse, time-to-event, PA, and SH data.

Techniques such as fluorescence microscopy (Pepperkok and Ellenberg, 2006; Miyashiro and Goulian, 2007) give time-lapse data following individual cells, while patient data, which are measured in a sequence of days/time slots can also be regarded as TL data (Endo et al., 2020). Information about events can also be provided, e.g. the time point when cell death or deterioration after treatment (Charton et al., 2020) happens. In contrast, in cytometry techniques (Davey and Kell, 1996; Bodenmiller et al., 2012) and single-cell RNA sequencing (Buettner et al., 2015; Frei et al., 2016), cells cannot be directly compared across cells, allowing to use for inference only distribution statistics such as mean and variance, of cellular measures at each snapshot. This is generally less informative than fitting single-cell trajectories and can however still provide important insights into cellular processes. Furthermore, the long-term patients information (Coates et al., 2002) are also more convenient to be collected as PA and SH data compared to TL data while no tracking of individuals are needed. We use the notation of $\bar{y}$ for measurement data and y for simulated data throughout this thesis. $\mathcal{D}$ denotes the data set.

Time-lapse data

Unlike population measurements, time-lapse data make it possible to track the behavior of individual cells or patients at continuous time points (Figure 2.1(a)). The measurement is usually either gene expression data or the concentration of the measured proteins. However, because of the experimental complexity, usually, only a limited number of single cells or patients can be measured using specific technologies. Therefore, we might not get the view of the whole population with only a small subset of the cells if the population is of high heterogeneity.

During the measurement of TL data,time-resolved measurements $\bar{y}$ are generated. Furthermore, event-resolved outputs $\bar{z}$ are also considered. Therefore, here we also introduce the event-resolved data, which are used in this thesis as well. Event-resolved data are time points $\bar{z}_i$ for a single

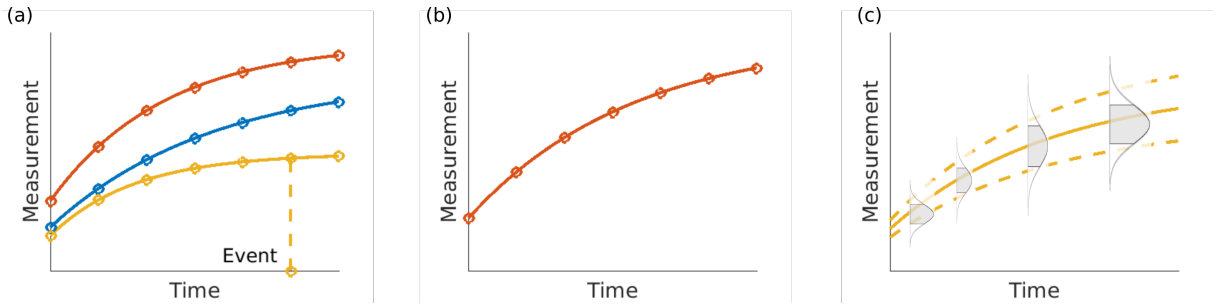

Figure 2.1: **Different data types can be measured in cellular systems.** (a) SCTL data with time-resolved data for each single-cell or patient. Time-to-event data can also be measured in this process. (b) PA data describes the mean of a whole population, therefore without single-cell resolution. (c) Single-cell snaptshot data contains mean and variance (or one of them) of the whole cell or patient population.

cell i, when a threshold $y_{threshold}$ is reached ($y(t = z_i) - y_{threshold} = 0$), e.g. when cell apoptosis happens. In our study, if the threshold is not reached until the last measured time point, z_i is set to the last time point and the root function value is calculated as $r_i = y(t = z_i) - y_{threshold}$ and used as regularization (Fröhlich et al., 2017b). In summary, the measured TL data set includes three parts; (i) the state output at different time points, (ii) the time point when the specific event happens, and (iii) the root value.

$$\mathcal{D}_{\text{TL}} = \left\{ \left\{ t_{jk}, \underbrace{\{\bar{y}_{ijk}\}_{i=1}^{d_y}}_{\text{output}} \right\}_{j=1}^{d_t}, \underbrace{\{\bar{z}_k\}}_{\text{event}}, \underbrace{\{\bar{r}_k\}}_{\text{root value}} \right\}_{k=1}^{d_c}, \tag{2.8}$$

where d_y is the number of different observations, d_t is the number of measured time point, d_c is the number of measured cells. i, j, k are the indices of measurement, time points, and single cells respectively.

Population average data

PA data provides the mean of a whole cell/patient population being measured at different time points (Figure 2.1(b)).

$$\bar{\mu}_{ij} = \sum_{k=1}^{d_c} \bar{y}_{ijk}, \tag{2.9}$$

The measurement data is given by

$$\mathcal{D}_{\text{PA}} = \left\{ \left\{ t_j, \{\bar{\mu}_{ij}\}_{i=1}^{d\mu} \right\}_{j=1}^{dt} \right\}, \tag{2.10}$$

where, d_μ is the number of mean measurements.

The measurement data is not of single-cell resolution, therefore cannot be compared across single cells. But because of the simplicity of the experimental methods, a large number of cells can be measured simultaneously. Therefore, PA data can give us more insight into the whole population. However, as the data are produced by averaging the properties of all individuals, if the population is highly heterogeneous, it may also result in misleading interpretations of the actual behaviors and the corresponding underlying mechanisms by only using PA data (Cohen-Saidon et al., 2009; Cheong et al., 2010; Love et al., 2010). In summary, although PA data can be less informative compared to SCTL data, they can still yield important insights into cellular processes.

Snapshot data

SH data provide the mean and variance (or one of them) of a whole cell/patient population at different time points (Figure 2.1(c)).

$$\bar{\Sigma}_{ij} = \frac{1}{d_c} \sum_{k=1}^{d_c} \left(\bar{y}_{ijk} - \mu_{ij} \right)^2 , \tag{2.11}$$

Same as it is for the PA data, SH data can also be obtained for hundreds to thousands of single cells, but can also potentially miss the individual properties.

The measurement data is given by

$$\mathcal{D}_{\mathrm{SH}} = \left\{ \left\{ t_j, \left\{ \bar{\Sigma}_{ij} \right\}_{i=1}^{d_\Sigma} \right\}_{j=1}^{dt} \right\} , \tag{2.12}$$

where d_Σ is the number of variance measurements.

2.3 Parameter estimation

Mechanistic models are usually subject to unknown parameters θ, which cannot be measured directly. Such parameters often include kinetic rates, scaling factors, or noise terms. In the case of combining MEMs as in this thesis, the above-mentioned parameters, according to their biological definition, might also consist of two parts. One for fixed effect, which is the mean of the parameter, and one for random effect, which gives information about the within-population variability. Usually, we are in the first place interested in optimal parameters, which give us the best fit of the given data. With these optimum parameters, one can make quantitative predictions using the model. Such predictions can usually provide information about the mechanism of cell pathways (Klipp and Liebermeister, 2006) or treatment performance of patients (Heinzel et al., 2014).

In this section, we introduce optimization methods used for parameter estimation, including local and global optimization methods. Furthermore, we introduced methods to compute gradient, as well as likelihood computation of MEMs in terms of the above-mentioned three different data types.

2.3.1 Maximum likelihood estimation

In order to estimate the unknown parameters, commonly maximum likelihood (ML) estimation is used, this method aims to find parameters that maximize the probability of seeing the measured observations given the parameters (Fisher, 1922), based on the definition in Equation (2.4):

$$\theta^{ML} = \arg\max_{\theta} \{\mathcal{L}_{\mathcal{D}}(\theta)\} \tag{2.13}$$

In order to get a limited search space, the parameter vector θ is usually subject to so-called box-constrains, which are typically based on the prior knowledge of the biological system.

Because of the computation of the product involved in the likelihood function, it can be numerically unstable. Thus, instead usually the negative log-likelihood

$$\theta^{ML} = \arg\min_{\theta} \{-\log(\mathcal{L}_{\mathcal{D}}(\theta))\} \tag{2.14}$$

is minimized. This formulation also has the advantage that the location of optimum is conserved as log is a monotonic function. Besides, the function is often more concave, therefore easier to be maximized.

2.3.2 Likelihood of MEMs

The negative log-likelihood with additive Gaussian noise is written as follows,

$$J(\theta) = -\log(\mathcal{L}_{\mathcal{D}}(\theta)) = \frac{1}{2}\sum_{i=1}^{d_y}\sum_{j=1}^{d_t}\left[\log\left(2\pi\sigma_{ij}^2\right) + \left(\frac{\bar{y}_{ij} - y_i\left(t_j,\theta\right)}{\sigma_{ij}}\right)^2\right]. \tag{2.15}$$

In our mixed effects model, the fixed effects β and the covariance matrix entries of the random effects $\delta = [\delta_{ij}]$ together define the parameters vector θ, so the unknown parameter vector

$$\theta = [\beta; \delta] \tag{2.16}$$

needs to be estimated by minimizing the negative log-likelihood. Therefore, compared with the classical models, where parameters are constants instead of distributions, the MEM approach extends the parameter dimension by the number of parameters used to parametrize the random effects covariance matrix D.

In this section, we introduce the negative log-likelihood functions of different data types used throughout this thesis.

Single-cell/patient time-lapse data

As the parameters φ of MEMs are latent, we have to integrate out the random effects in order to compute the likelihood of TL data.

$$\mathcal{L}_{\mathcal{D}_{\mathrm{TL}}}(\beta, D) = \int p(\mathcal{D}_{\mathrm{TL}}|\varphi)p(\varphi|\beta, D)\, d\varphi \quad = \prod_{k=1}^{d_c} \int p(\mathcal{D}_{\mathrm{TL}}|\varphi_k = A\beta + Bb_k)p(b_k|D)db_k, \quad (2.17)$$

where the single-cell/patient likelihood $p(\mathcal{D}_{\mathrm{TL}}|\varphi_k = A\beta + Bb_k)$ is given via equation (2.15). Here we assume that the design matrices A is an identity matrix and B is a rectangular matrix with 0 or 1 entries that adds random effects to a subset of φ elements:

It is hard to compute the exact result of $\mathcal{L}_{\mathcal{D}_{\mathrm{TL}}}(\beta, D)$ because of the integrals over b_k. For this purpose, different approximations have been proposed. For instance, alternating approximation (Lindstrom and Bates, 1990), Laplacian approximation (Tierney and Kadane, 1986), importance sampling (Tokdar and Kass, 2010), Gaussian quadrature (Davidian and Gallant, 1992) and first order conditional approximation approaches (Stuart Beal, 1980) have been developed to approximate the integration with respect to random effects of the negative log-likelihood of TL data. Some of these approximation methods are based on a first order Taylor expansion of the model function f around the expected random effects values, or around the conditional modes of the random effects, for instance, the alternating approximation method.

Markov Chain Monte Carlo sampling could alternatively be used to numerically approximate the exact result, with vanishing approximation error a towards large sample size. Besides, importance sampling method provides a simple and efficient way of performing Monte Carlo (MC) integration. However, it is still highly time demanding, as it requires many samples for an accurate approximation.

In the following chapters, we use a first order conditional approximation to approximate the negative log-likelihood. This method avoids the integration with respect to the random effects and simplifies the computation via a local approximation:

$$\hat{b}_k(\beta, D) = \arg\min_{b^{(i)}} \psi^{(i)}(b^{(i)}, \beta, D). \quad (2.18)$$

with

$$\begin{aligned} \psi_k(b_k, \beta, D) &= -\log\big(p(\mathcal{D}_{\mathrm{TL}}|\beta, b_k)\, p(b_k|D)\big) \\ &\approx \psi_k(\hat{b}_k, \beta, D) + \frac{1}{2}\,(b_k - \hat{b}_k)\, H_k(\hat{b}_k, \beta, D)\, (b_k - \hat{b}_k)^T. \end{aligned} \quad (2.19)$$

Where $H_k(b_k, \beta, D) = \frac{\partial^2 \psi_k}{\partial b\, \partial b^T}\Big|_{b_k, \beta, D}$. The gradient $\frac{\partial}{\partial b} \log(p(\mathcal{D}_{\mathrm{TL}}^{(i)}|\beta, b)\, p(b|D))$ is zero at the estimate $\hat{b}_k(\beta, D)$. Furthermore, the Hessian $H_k(\beta, D)$ is, due to the term $p(\hat{b}_k|D)$, almost always positive definite.

Then an approximation of the integral is obtained by integrating the Taylor series expansion of the integrand. As this Gaussian integral can be solved analytically, we obtain

$$\int p(\mathcal{D}_{\mathrm{TL}}|\beta, b_k)\, p(b_k|D)\, \mathrm{d}b_k \approx \exp\big(-\psi_k(\hat{b}_k, \beta, D)\big)\, \det\big(H_k(\hat{b}_k, \beta, D)\big)^{-1/2}. \quad (2.20)$$

The approximation of the likelihood for a collection of cells/patients is given by the product of the approximations for the individual cells/patients,

$$p(\mathcal{D}_{\mathrm{TL}}|\beta, D) \approx \prod_{k=1}^{n_c} \exp\big(-\psi_k(\hat{b}_k, \beta, D)\big)\ \det\big(H_k(\hat{b}_k, \beta, D)\big)^{-1/2}, \tag{2.21}$$

and the corresponding, approximated negative log-likelihood function is

$$J_{\mathrm{TL}}(\theta) \approx \sum_{k=1}^{n_c} \psi_k(\hat{b}_k, \beta, D) + \tfrac{1}{2}\log\big(\det(H_k(\hat{b}_k, \beta, D))\big). \tag{2.22}$$

Single cell/patient parameter estimation

After population parameters are estimated, we are also interested in the single-cell fittings. Therefore, we estimated the single-cell parameters by optimizing the random effects for each individual cell/patient respectively, while giving the population parameters as prior:

$$\hat{b}_k(\beta, D) = \mathrm{argmax}_{b_k}(p(\Big\{t_{jk}, \{\bar{y}_{ijk}\}_{i=1}^{d_y}\Big\}_{j=1}^{d_t}|\beta, b_k) + p(b_k|D)). \tag{2.23}$$

Here, i, j, k are the indices of measurement, time point, and cell/patient respectively.

Population average data

Simulated mean and variance need to be computed to compare with the measured PA and SH data. For this purpose, MC methods, sigma point (SP) methods (Julier et al., 1995; Julier and Uhlmann, 2004; Menegaz et al., 2011; Lerner, 2002; van der Merwe, 2004 and Charalampidis and Papavassilopoulos, 2011) have been developed.

MC sampling is the most intuitive way used to calculate the simulated mean using MEMs. A large number of samples are generated according to the covariance matrix D, and the ODE model is simulated for each sample. The simulated mean is then computed using the outputs. By combining the PA data and the simulated mean, the negative log-likelihood of the PA data can be obtained. However, MC sampling is computationally expensive as a large number of simulations need to be done. As an alternative, SP methods can be used.

To compute the simulated mean, n_s samples (MC, SP or CMD samples), $s_m \sim \mathcal{N}(0, \boldsymbol{I}), \quad m = 1, ..., n_s$ are drawn from a standard normal distribution. Random effects samples $b \sim \mathcal{N}(0, D)$ are then calculated as $b_m = D^{\frac{1}{2}} s_m$, where $D^{\frac{1}{2}}$ is the Cholesky matrix square-root of D. The simulated means μ are given as

$$\mu_{ij} = \frac{1}{n_s} \sum_{m=1}^{n_s} y_i\left(t_j, \varphi = A\beta + Bb_m\right). \tag{2.24}$$

The negative log-likelihood of PA data is given by

$$\begin{aligned} J_{\mathrm{PA}}(\theta) &= -\log(p(\mathcal{D}_{\mathrm{PA}}|\beta, D)) \\ &= \frac{1}{2}\sum_{i=1}^{d_\mu}\sum_{j=1}^{d_t}\left(\frac{\bar{\mu}_{ij} - \mu_i\left(t_j, \varphi\right)}{\sigma^{\mu}_{ij}}\right)^2 + \log\left(2\pi\sigma^{(\mu)}_{ij}\right), \end{aligned} \tag{2.25}$$

According to the central limit theorem, the measured PA data are approximately normally distributed for $N \gg 1$, with the variance of σ^{μ} (Fröhlich et al., 2016), which can be differ between time points and can be measured by experimental repetitions.

Snapshot data

Same to the negative log-likelihood calculation of PA data, sampling methods can be used for the sing-cell snapshot data. The simulated variances Σ are given as

$$\Sigma_{ij} = \frac{1}{n_s}\sum_{m=1}^{n_s}\left(y_i\left(t_j, \varphi = A\beta + Bb_m\right) - \mu_{ij}\right)^2, \tag{2.26}$$

Thus the negative log-likelihood can be computed as follows,

$$\begin{aligned} J_{\mathrm{SH}}(\theta) &= p(\mathcal{D}_{\mathrm{SH}}|\beta, D) \\ &= \frac{1}{2}\sum_{i=1}^{d_\Sigma}\sum_{j=1}^{d_t}\left(\frac{\bar{\Sigma}_{ij} - \Sigma_i\left(t_j, \varphi\right)}{\sigma^{\Sigma}_{ij}}\right)^2 + \log\left(2\pi\sigma^{(\Sigma)}_{ij}\right). \end{aligned} \tag{2.27}$$

2.3.3 Optimization methods

As the first step of parameter inference, the model parameters need to be estimated using optimization methods, which aim to find the best solution to a problem.

Local optimization methods

Local optimization methods sequentially refine a single initial point locally until it reaches an approximate minimum. Many methods exist for solving local optimization problems. These can be classified as derivative-free and derivative-based methods. We used a derivative-based method throughout this thesis. The usage of derivative provides information that helps to search for a good optimization direction and step size (if Hessian is provided together with gradient) (Nocedal and Wright, 2006). Therefore, we employed derivative-based methods for parameter optimization throughout this thesis.

We used the interior-point method implemented in the MatLab build-in function `fmincon` in the following chapters. One of the first interior-point is presented by (Clement, 1991). This method

used a logarithmic barrier function to accommodate inequality constraints, and Newton's method to solve the Karush-Kuhn-Tucker conditions (KKT). It has been shown that the interior-point methods typically have superior performance on large-scale problems (Andersen et al., 1996; Koh et al., 2007).

Global optimization methods

Many approaches have been developed for the global optimization of biologically motivated ODE models (Fröhlich et al., 2019). The most intuitive idea is the multi-start local optimization method, which generates a set of randomly sampled start points and performs local optimization for each point. The convergence performance can be concluded from the generated waterfall plot if enough number of start points finished at the end of optimization. The best local optimum is assumed to be the global optimum. The start points can be generated from a uniform distribution or using the Latin-hypercube method (Raue et al., 2013).

Besides the multi-start local optimization method, many other methods are also developed for global optimization, for example, the swarm-based methods (Kennedy, 2011; Vaz and Vicente, 2007), genetic algorithms (Hansen and Ostermeier, 1996), and simulated annealing method (Kirkpatrick et al., 1983).

2.3.4 Gradient computation

Derivative-based methods are commonly used for local optimization, as they provide valuable information about the search direction and step length. Usually, three approaches are commonly employed to compute derivatives of the objective function, including the finite difference method, forward sensitivity, and adjoint sensitivity analysis.

In this section, we introduce the finite difference (Raue et al., 2013) and forward sensitivity methods, which are used in this thesis.

Finite difference

Finite difference methods approximate the slope of a tangent by a secant of small step size, which can be defined flexibly,

$$\frac{\mathrm{d}J}{\mathrm{d}\theta_k} \approx \frac{J(\theta - h_1 e_k) - (\theta - h_2 e_k)}{h_1 + h_2}, \tag{2.28}$$

where e_k is the kth unit vector and h_1, $h_2 \geq 0$ are the step sizes. θ_k denotes the kth parameter, with respect to which the derivative is computed.

Three different ways of choosing the step sizes are usually employed, including the forward ($h_1 > 0 = h_2$), central ($h_1 = h_2 > 0$), and backward ($h_1 = 0 < h_2$) finite difference. The approximation order of the central finite difference method is $\mathbf{O}(h^2)$ compared to $\mathbf{O}(h)$ of the

other two methods. However, the computation time is also doubled. According to our research in Chapter 3, the approximation of finite difference is not robust regarding different step sizes. Furthermore, the computation time scales up linearly with the dimension of parameters.

Forward sensitivity analysis

Consider an ODE model defined as Equation (2.1), the derivative of negative log-likelihood J with respect to parameter vector φ is

$$\begin{aligned}\Delta J(\varphi) &= \sum_{i=1}^{d_t}[\frac{\partial J_i}{\partial \varphi} + \frac{\partial J_i}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}\varphi}] \\ &= \sum_{i=1}^{d_t}[\frac{\partial J_i}{\partial \varphi} + \frac{\partial J_i}{\partial y}\frac{\partial h}{\partial \varphi} + \frac{\partial J_i}{\partial y}\frac{\partial h}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}\varphi}],\end{aligned} \tag{2.29}$$

where d_t is the number of time points, and

$$\frac{\mathrm{d}y}{\mathrm{d}\varphi} = \frac{\partial h}{\partial \varphi} + \frac{\partial h}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}\varphi}. \tag{2.30}$$

Define

$$s^x = \frac{\mathrm{d}x}{\mathrm{d}\varphi}. \tag{2.31}$$

Differentiating s^x by time t, we get

$$\begin{aligned}\dot{s}^x(t,\varphi) &= \frac{\mathrm{d}h}{\mathrm{d}\varphi} \\ &= \frac{\partial h}{\partial \varphi} + \frac{\partial h}{\partial x}s^x, \\ s^x(t_0,\varphi) &= \frac{\mathrm{d}x_0}{\mathrm{d}\varphi}(\varphi),\end{aligned} \tag{2.32}$$

this can be solved alongside with Equation (2.1), which yields an ODEs with the dimension of $d_x(1+d_\varphi)$. Where d_x and d_φ are the dimension of states x and parameters, respectively.

2.3.5 Adjoint sensitivity analysis

By introducing the adjoint state p, the adjoint sensitivity analysis circumvents the evaluation of s^x altogether (Plessix, 2006). For ODE models defined as in Equation (2.1), from which follows

$$\frac{d\dot{x}}{d\theta} - \frac{\partial f}{\partial \theta} - \frac{\partial f}{\partial x}\frac{dx}{d\theta} = 0 \tag{2.33}$$

Defining $J_0 = 0$, we can write the $j = 0, \dots, d_t - 1$ by adding the zero-term as defined above,

$$\begin{aligned}\frac{dJ_j}{d\theta} &= \frac{\partial J_j}{\partial y}\frac{\partial h}{\partial x}\frac{dx}{d\theta} + \int_{t_j}^{t_{j+1}} p^T(\frac{d\dot{x}}{d\theta} - \frac{\partial f}{\partial \theta} - \frac{\partial f}{\partial x}\frac{dx}{d\theta})dt \\ &= \frac{\partial J_j}{\partial y}\frac{\partial h}{\partial x}\frac{dx}{d\theta} + (\lim_{t\to t_{j+1}} p^T\frac{fx}{d\theta} - \lim_{t\to t_j} p^T\frac{dx}{d\theta}) - \int_{t_j}^{t_{j+1}} (\dot{p}^T + p^T\frac{\partial f}{\partial x})\frac{dx}{d\theta}dt - \int_{t_j}^{t_{j+1}} p^T\frac{\partial f}{\partial \theta}dt,\end{aligned} \tag{2.34}$$

using integration by parts in the second step. It is then helpful to define the adjoint system as follows,

$$\dot{p} = -\frac{\partial f^T}{\partial x}p, \tag{2.35}$$

with the terminal condition $p(t_{j+1}, \theta) = \lim_{t\to t_{j+1}} p(t, \theta) - \frac{\partial J_{j+1}}{\partial y}\frac{\partial h}{\partial x}$, where $p = 0$ for $t > t_{d_t}$. By imposing them on the likelihood function, we get

$$\begin{aligned}\frac{dJ_j}{d\theta} &= \frac{\partial J_j}{\partial y}\frac{\partial h}{\partial x}\frac{dx}{d\theta} + p^T\frac{dx}{d\theta}|_{t=t_{i+1}} - (p^T + \frac{\partial J_j}{\partial y}\frac{\partial h}{\partial x})\frac{dx}{d\theta}|_{t=t_j} - \int_{t_j}^{t_{j+1}} p^T\frac{\partial f}{\partial \theta}dt \\ &= p^T\frac{dx}{d\theta}|_{t=t_{j+1}} - p^T\frac{dx}{d\theta}|_{t=t_j} - \int_{t_j}^{t_{j+1}} p^T\frac{\partial f}{\partial \theta}dt.\end{aligned} \tag{2.36}$$

Summing over $j = 0, \dots, d_t - 1$, we get

$$\frac{J}{d\theta} = \sum_{j=1}^{d_t}[\frac{\partial J_j}{\partial \theta} + \frac{\partial J_j}{\partial y}\frac{\partial h}{\partial \theta}] - p^T\frac{dx_0}{d\theta}|_{t=t_0} - \int_{t_0}^{t_{n_t}} p^T\frac{\partial f}{\partial \theta}dt. \tag{2.37}$$

It has been shown that the computation time of adjoint sensitivity analysis hardly increases with the dimension of parameters (Fröhlich et al., 2017a; Özyurt and Barton, 2005). Therefore, the adjoint sensitivity method has been shown to be the most efficient method for large-scale systems (Fröhlich et al., 2017a; Fröhlich et al., 2018a).

2.3.6 Parameter identifiability

It is an important problem in parameter estimation, whether model parameters can be inferred from the observations. In this section, we introduced structural and practical identifiability.

Structural identifiability

Assuming perfect experimental data, structural identifiability regards the possibility of determining unique values for parameters from the available observations. A parameter θ is structurally identifiable, if for any parameter values in the constraint, when

$$h(t,\theta) = h(t,\theta'), \tag{2.38}$$

we get $\theta = \theta'$ (Chis et al., 2011; Ligon et al., 2018). It is to be noted that this does not take the measurement data into account. Therefore, structural identifiability is a theoretical property of the model structure, which depends only on the system dynamics and the observation (Walter et al., 1997). If a system is structural identifiable, the distribution of θ^{ML} tends to the Gaussian distribution when the sample size increases, $\theta^{ML} \sim \mathcal{N}(\theta^{true}, FIM^{-1})$, where FIM is the Fisher information matrix,

$$FIM(\theta) = \sum_{i=1}^{d_y}\sum_{j=1}^{d_t} \frac{1}{\sigma_{ij}^2(\theta)} (\frac{dy_i(t_j,\theta)}{d\theta})^T \frac{dy_i(t_j,\theta)}{d\theta}. \tag{2.39}$$

Under the assumption that $\bar{y}_{ij} - y_i(t_j,\theta)$ is small when θ is close to θ^{ML}, the Hessian matrix $H()$ can be approximated as following,

$$H(\theta) = \frac{\partial^2 J}{\partial\theta_l \partial\theta_m} = \sum_{i=1}^{d_y}\sum_{j=1}^{d_t} \frac{1}{\sigma_{ij}^2(\theta)} \frac{\partial y_i(t_j,\theta)}{\partial\theta_l} \frac{\partial y_i(t_j,\theta)}{\partial\theta_m} - (\bar{y}_{ij} - y_i(t_j,\theta) \frac{\partial^2 y_i(t_j,\theta)}{\partial\theta_l \partial\theta_m}) \approx FIM(\theta). \tag{2.40}$$

Practical identifiability

Unlike the definition of structural identifiability, practical identifiability takes measurement data $\mathcal{D}$, noise model, and prior knowledge into account. A parameter θ is practically identifiable from a data set $\mathcal{D}$, if its confidence interval is finite or sufficiently tight.

Profile likelihood and confidence interval

Profile likelihood computation is a commonly used method to assess the parameter uncertainty by computing the confidence interval. This method is of high accuracy but also computationally expensive. Research has shown that profile likelihood computation is a robust method also when the system is not structurally identifiable (Fröhlich et al., 2014). A profile is the maximum projection of the likelihood to a chosen parameter axis. For a parameter θ_r in the parameter vector θ, the profile value at $\theta_r = c$ is

$$PL_{\theta_r}(c) = \max \mathcal{L}_{\mathcal{D}}(\theta) \tag{2.41}$$

A $(1-\alpha)\%$ confidence interval is defined as an observed interval that frequently includes the parameter of interest if the experiment is repeated in $(1-\alpha)\%$ of the replicates. When the profile likelihood computation is feasible, the $(1-\alpha)\%$ confidence interval of parameter θ_r based on likelihood can be calculated as following,

$$\begin{aligned} CI^{\alpha}_{\theta_r} &= \left\{ \theta_r | \exists \theta : \frac{\mathcal{L}_{\mathcal{D}}(\theta)}{\mathcal{L}_{\mathcal{D}}(\theta^(ML))} > \exp(-\frac{\Delta_\alpha}{2}) \right\} \\ &= \left\{ \theta_r | \exists \theta : 2(J(\theta) - J(\theta^{ML})) < \Delta_\alpha \right\}, \end{aligned} \tag{2.42}$$

where Δ_α is the α-th percentile of the χ^2-distribution with d_θ degrees of freedom.

The computation time of profile likelihood scales up linearly with respect to the number of parameters, as it has to be done for each parameter, respectively. Although profiles provide substantial insights of the model, it is quite computationally expensive than a single local optimization problem, as the optimization needs to be done for each single step for each parameter.

Asymptotic confidence intervals

Although it is robust, the profile likelihood computation is extremely computationally expensive, especially for complex biological systems. Therefore, when it is not feasible to use such an assessment, the asymptotic confidence intervals are computed instead (Brown et al., 2002; Greselin and Pasquazzi, 2009; Tang et al., 2011). Consider the Taylor series of negative log-likelihood at θ^{ML},

$$\begin{aligned} J(\theta) &\approx J(\theta^{ML}) + \frac{1}{2}(\theta - \theta^{ML})^T H(\theta^{ML})(\theta - \theta^{ML}) \\ \mathcal{L}_{\mathcal{D}}(\theta) &\approx \exp(-J(\theta^{ML})) \exp\left\{ -\frac{1}{2}(\theta - \theta^{ML})^T H(\theta^{ML})(\theta - \theta^{ML}) \right\}. \end{aligned} \tag{2.43}$$

Thus,

$$\begin{aligned} CI^{\alpha}_{\theta_r} &= \left\{ \theta_r | \exists \theta : \exp\left\{ -\frac{1}{2}(\theta - \theta^{ML})^T H(\theta^{ML}) \right\} > \exp(-\frac{\Delta_\alpha}{2}) \right\} \\ &= \left\{ \theta_r | \exists \theta : (\theta - \theta^{ML})^T H(\theta^{ML}) < \Delta_\alpha \right\} \\ &= \left(\theta_r^{ML} - \sqrt{\Delta_\alpha H^{-1}_{rr}}, \theta_r^{ML} + \sqrt{\Delta_\alpha H^{-1}_{rr}} \right), \end{aligned} \tag{2.44}$$

where the Hessian matrix $H()$ can be substituted with $FIM()$ as introduced in Section 2.3.6, if assuming normally distributed noise.

Chapter 3

Integration of different data types

Experimental methods for assessing single-cells/patient and populations have tremendously improved over the last decades. The resulting multiple data types provide different insights of the biological systems being investigated. For instance, time-lapse (TL) data provide time-resolved data for each cell/patient, while population average (PA) and snapshot (SH) data give information on the whole population. The improvement of data collection methods allows for the development of quantitative mechanistic models for cellular or drug response processes. Yet, these mechanistic models are mostly based on single data types, as statistical methods for model-based data integration are missing.

In this chapter, we present an approach for integrating TL, SH, PA, and time-to-event data. The underlying formulation based on nonlinear mixed-effect modeling allows for the mapping to different data types, with and without single-cell/patient resolution. Our study shows that the approach can reliably integrate different data types, thereby improving parameter identifiability and prediction accuracy. Applying it to a model of extrinsic apoptosis reveals that the simultaneous consideration of multiple data types can be necessary due to experimental constraints. The proposed approach is widely applicable and might serve to further refine our understanding of biological processes.

3.1 Introduction

Different types of data are provided by a broad spectrum of experimental techniques. TL data and time-to-event data, which is the time point when a specific event happens, are for instance captured by time-lapse (fluorescence) microscopy or sequencing (Qiu et al., 2020). SH data can be obtained using image and flow cytometry (Davey and Kell, 1996; Bodenmiller et al., 2012), single-cell sequencing (Buettner et al., 2015; Frei et al., 2016) or single-cell proteomics (Palit et al., 2019). PA data are available via methods such as bulk sequencing (Li and Wang, 2021), microarray analysis (Daran-Lapujade et al., 2008), and immunoblotting (Kurien and Scofield, 2006). SH and PA data provide statistics of large populations of individuals. While these methods are easier to implement and data collection is cheaper, they provide no time-course for individuals. Time-lapse experiments track single cells/patients over time but use usually far fewer individuals.

The experimental techniques differ in throughput and content, yet, all are highly valuable in modern workflows, as they provide different insights into the cell population being studied. Therefore, it is desirable to be able to integrate information provided by different techniques. Researchers have used TL data under different experimental conditions (Almquist et al., 2015; Fröhlich et al., 2018b), or combined PA and SH data (Adlung et al., 2021) to estimate the parameters of mechanistic models. Results have shown that this data integration improves parameter fits and biological interpretations (Adlung et al., 2021). However, it has so far not been investigated whether one can integrate all of the above-mentioned data types, i.e. TL data (including event data), PA data, and SH data, in a statistically coherent way.

In this study, we are proposing a data integration framework based on MEMs which provides a description of the population based on models for the individuals. This allows for the modelling and assessment of cell-to-cell and patient-to-patient variability (Dharmarajan et al., 2019; Loos et al., 2018b) but also the population mean and variance. Regarding TL data, the integral of the likelihood is computed according to Section 2.3.2. For population-level data, the statistical moments of the population dynamics described by the MEM need to be evaluated. This can be done using Monte-Carlo (MC) sampling methods, which have high accuracy but is compensated by long computation time. Alternatively, computationally more efficient approximations are provided by the sigma point (SP) methods, which can be found in detail in Chapter 4.

In this work, we present an approach to fit a model jointly on information from TL and time-to-event data, as well as PA and SH measurements. To achieve this, we formulate a joint likelihood function for all data types which can be evaluated using established methods. As the resulting parameter estimation problem is computationally challenging, we derive and implement the respective objective function gradient for efficient optimization and uncertainty analysis. The proposed approach is evaluated using a conversion reaction model as a test problem, and an extrinsic apoptosis pathway model as an application problem. The test cases confirm the correctness of the approach and demonstrate its benefit compared to methods which could only exploit a subset of the datasets. The application of datasets for extrinsic apoptosis strengthens these findings and demonstrates that the proposed approach can be applied to large models which many fixed and random effects.

3.2 Joint likelihood to integrate three data types in mixed effects models

To take cell-to-cell variability into account, we consider non-linear MEMs as defined by Lindstrom and Bates (1990) (Equation (2.5)). We consider three different data types, namely TL data (including event data), PA data, and SH data. To use the information from these different data modalities in parameter estimation, we define a joint likelihood function or corresponding negative log-likelihood based on Equations (2.22), (2.25) and (2.27)

$$J(\theta) = J_{TL}(\theta) + J_{PA}(\theta) + J_{SH}(\theta) \tag{3.1}$$

with gradient

$$\frac{dJ(\theta)}{d\theta} = \frac{dJ_{TL}(\theta)}{d\theta} + \frac{dJ_{PA}(\theta)}{d\theta} + \frac{dJ_{SH}(\theta)}{d\theta}. \tag{3.2}$$

That is, we compute the negative log-likelihood and gradient values for the different data types separately, and then sum them up, which corresponds to assuming independence of the different data types. We use the first-order conditional approximation method to numerically solve the integration of the likelihood function for TL data and the Monte Carlo (MC) sampling method with 10,000 samples to approximate the mean and variance. The detailed equations can be found in Section 2.3.2.

As ODE problems usually possess many local optima and the use of gradient information has been shown to be advantageous (Schälte et al., 2018; Villaverde et al., 2019), we used gradient-based multi-start local optimization with 600 Latin-hypercube start points (Raue et al., 2013). Simulations were performed, and likelihood and gradient values were calculated, as mentioned above and outlined in Figure 3.1 (c).

To access the contribution of different subsets of experimental data to the parameter estimation, we performed the parameter optimization using different data set combinations. When one or two data types are left out, we tried to predict them by using the parameters estimated from the other data types (Figure 3.1 (a)). Besides, we also predicted the behavior of one cell line based on a model calibrated with data obtained from another cell line (wild-type HeLa and CD95-HeLa cells).

To evaluate parameter identifiability, we computed the profile likelihood of the conversion reaction model parameters (Raue et al., 2013). Because of computation time limitations, we accessed the confidence interval of the extrinsic apoptosis model parameters by computing the eigenvalues of the Hessian matrices, which were computed using the finite difference method.

3.3 Application to test problem and extrinsic apoptosis model

To illustrate the properties of our approach and the effects of excluding different data types for both simple and complex models, we used a conversion reaction model $x_1 \rightleftharpoons x_2$ with the reaction rate φ_1 and φ_2 given via ODEs (Equations (3.3)) as a test problem, and then applied it to a more complex model of extrinsic apoptosis, a form of programmed cell death (Figure 3.2).

Conversion reaction model

Conversion reaction model is a widely used model in systems biology as it is a common motive in biological processes. The conversion reaction model is given via the ODEs,

$$\begin{aligned} \dot{x}_1 &= -\varphi_1 x_1 + \varphi_2 x_2, \qquad x_1(0) = \frac{\varphi_1}{\varphi_1 + \varphi_2}, \\ \dot{x}_2 &= \varphi_1 x_1 - \varphi_2 x_2, \qquad x_2(0) = \frac{\varphi_2}{\varphi_1 + \varphi_2}. \end{aligned} \tag{3.3}$$

To insure the heterogeneity of the initial values, as well as to keep the parameter number small, we set the initial values of both x_1 and x_2 to be combinations of the reaction rate. Artificial data of the above-mentioned types, namely TL data, PA data and SH data, were generated as listed

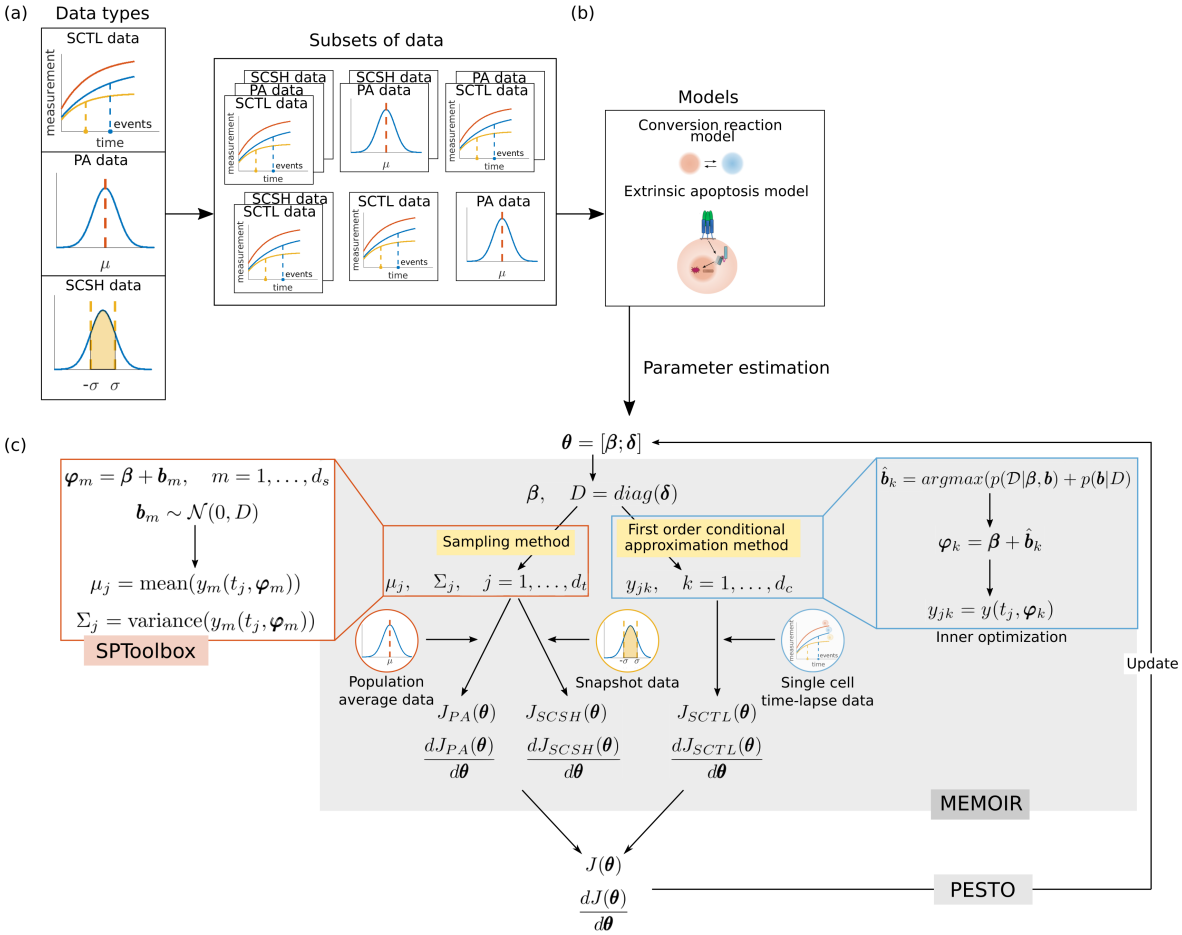

Figure 3.1: **Parameter estimation and comparison of integrating different subsets of data types.** (a) Three types are used in this study. Subsets of data are used for parameter optimization separately. (b) Two models are used to access the effects of different data combinations, the conversion reaction model and the extrinsic apoptosis model. The two icons are used in this paper to illustrate which model is used for analysis in the specific figures or sub-figures. (c) Illustration of the optimization routine. The MatLab toolbox MEMOIR (https://github.com/ICB-DCM/MEMOIR) is used to compute the negative log-likelihood values and the gradients. The first-order conditional approximation method is used for the TL data, while MC sampling is used for the PA and SH data, which is implemented in the MatLab toolbox SPtoolbox (https://github.com/ICB-DCM/SPToolbox). We used the PESTO toolbox (Stapor et al., 2018) for gradient-based multi-start local optimization.

later for the conversion reaction model, assuming only x_2 to be observed. As in real experiments, absolute measurements of TL data are often hard to obtain, we included a scaling factor φ_3 in the corresponding observation. The time points $\bar{z}$ when the measurement reached a certain value $x_1(0) \cdot \varphi_4$ is computed for each single cell and applied as event data. Additionally, the mean and variance of the absolute values of x_2 are also computed. In line with the experiment complexity, TL data are generated using 30 cells, with the other population data generated for 100 cells.

The assumption of noise is as follows: the measurement noise and the noise of measured event data $\epsilon \sim \mathcal{N}(0, 0.06^2)$, the noise of measured mean $\epsilon^{\mu} \sim \mathcal{N}(0, 0.03^2)$, the noise of measured variance $\epsilon^{\Sigma} \sim \mathcal{N}(0, 0.001^2)$. In detail, the following data types and parameterizations were considered:

- Time-lapse data (including event data)
  $d_c = 30, \quad \bar{y} = \varphi_3 x_2 + \epsilon, \quad \bar{z}_k = t_j + \epsilon$, when $\bar{y}_k(t_j) = x_1(0) \cdot \varphi_4, \quad k = 1, \ldots, d_c$
  $\varphi_3 = \exp(\beta_3), \quad \varphi = [\varphi_1, \varphi_2, \varphi_3, \varphi_4]^T$
  $\beta = [\log(0.6), \log(0.2), \log(2.5), \log(1.8)]^T, \quad \delta = [\log(0.1), \log(0.1), \log(0.1)]^T$

- Population average data
  $d_c = 100, \quad \bar{y} = x_2 + \epsilon, \quad \bar{\mu} = \frac{1}{d_c} \sum_{k=1}^{d_c} \bar{y_k} + \epsilon^{\mu}$
  $\varphi = [\varphi_1, \varphi_2]^T$
  $\beta = [\log(0.6), \log(0.2)]^T, \quad \delta = [\log(0.1), \log(0.1)]^T$

- Snapshot data
  $d_c = 100, \quad \bar{y} = x_2 + \epsilon, \quad \bar{\Sigma} = \frac{1}{d_c} \sum_{k=1}^{d_c} (\bar{y_k} - \bar{\mu})^2 + \epsilon^{\Sigma}$
  $\varphi = [\varphi_1, \varphi_2]^T$
  $\beta = [\log(0.6), \log(0.2)]^T, \quad \delta = [\log(0.1), \log(0.1)]^T$

where d_c is the number of cells. The unspecified parameters φ_i, $i = 1, 2, 4$, contain both fixed and random effect, $\varphi_i = \exp(\beta_i + b_i)$. $\bar{y}$ and $\bar{z}$ denote the measured observable and event time respectively. Vector δ contains the variance values of the random effect b

Extrinsic apoptosis model

A problem where cellular variability becomes immediately evident is apoptosis, a form of programmed cell death. When a cell population is exposed to a pro-apoptotic stimulus, some cells undergo apoptosis on a timescale of hours to days, while other cells survive and continue to proliferate. This "fractional killing" illustrates a main problem of cancer chemotherapy or radiotherapy, in which a surviving fraction of tumor cells may replenish the tumor cell population (Kim and Tannock, 2005). Previously, a model of extrinsic apoptosis is developed by our collaboration partener to study the mechanism of caspase-8 activation and describe heterogeneity in cell death kinetics based on a collection of single-cell and PA data sets (Kallenberger et al., 2014). The apoptosis model describes the activation of caspase-8 at death inducing signaling complexes (DISCs) followed by cleavage of the proapoptotic Bcl-2 family member BID to tBID, which causes mitochondrial outer membrane permeabilization (MOMP) resulting in apoptosis. The study motivated us to further develop strategies for combining different types of experimental measurements for model fitting in order to accurately characterize cell-to-cell variability and analyze sources of cellular heterogeneity.

Extrinsic apoptosis is initiated by extracellular death ligands, such as CD95 ligand (CD95L, also known as Fas ligand) or TRAIL. The binding of CD95L to CD95 (Fas) receptors induces the formation of death inducing signaling complexes (DISCs) (Kischkel et al., 1995). DISCs serve as a platform for the activation of initiator caspases, caspase-8 and caspase-10, that cleave and activate the effector caspases, caspase-3, and caspase-7, and cleave the proapoptotic Bcl-2 family member BID into tBID. Accumulation of tBID induces mitochondrial outer membrane permeabilization (MOMP) that irreversibly triggers activation of effector caspases and cell death (Kallenberger et al., 2014). In type I cells, effector caspase activation by initiator caspases is sufficient for apoptosis, whereas type II cells require indirect effector caspase activation by MOMP (Scaffidi et al., 1998).

In the studied model, CD95L binds to the cell death receptor CD95, thus leading to the formation of active death receptors (Figure 3.2) and binding of FADD, which results in the formation of DISCs. At DISCs, dimerized procaspase-8 (p55) is cleaved into the fragments p43 or p30 Hoffmann et al. (2009). Subsequently, p43 and p30 are cleaved to p18, which then leaves the DISC and enters the cytosol. The forms of caspase-8 that are catalytically active towards proteins not associated with the DISC, p43, and p18, can cleave the protein BID (BH3 interacting-domain death agonist). Truncated BID (tBID) then accumulates in mitochondria. In case, a certain threshold concentration of tBID is exceeded, mitochondrial outer membrane permeabilisation (MOMP) is caused resulting in apoptosis.

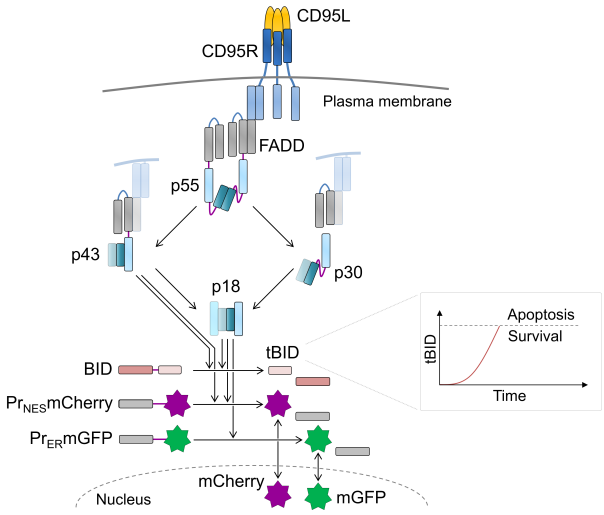

Figure 3.2: **Model of extrinsic apoptosis.** Binding of CD95L to CD95R results in assembly of complexes with FADD and dimers of procaspase-8 (p55). Cutting at two cleavage sites, drawn in purple, results in p43 or p30, and finally in p18. Membrane-bound p43 and cytosolic p18 are catalytically and can cleave BID to tBID as well as the cytosolic probe $Pr_{NES}mCherry$. The ER-bound probe $Pr_{ER}mGFP$ can only be cleaved by p18. After cleavage of probes, mCherry and mGFP can enter the nucleus, which is detected by confocal microscopy. Inlay: In case, a certain threshold of tBID is exceeded, a cell undergoes apoptosis.

The model of extrinsic apoptosis links CD95L concentrations to apoptosis kinetics and fractional killing in a heterogeneous cell population. In this chapter, the extrinsic apoptosis pathway model is used as an application problem to illustrate the property of our method and the effects of combinations of different data types. To quantitatively study this process, experimental data sets were recorded in wild-type HeLa cells and in stably CD95-overexpressing HeLa cells (CD95-HeLa) Kallenberger et al. (2014). Using fluorescent cleavage probes, cleavage activities of p18 and p43 were quantified in single cells. The two cell lines were treated with CD95L at four different concentrations per cell line resulting in different cell death kinetics. Compared to the previous study, here, the established model fitting approaches were applied in a larger collection of experimental data. More single-cell data sets from CD95-HeLa cells, 30 instead of 10 cells per ligand concentration, were used for model calibration. In these cells, due to the accelerated cell death kinetics, more single-cell data sets could be recorded and used for model fitting as compared to wild-type HeLa cells. Accordingly, the MEM was fitted to a total of 160 single-cell data sets. Additionally, single-cell apoptosis times, quantitative immunoblots, and SH data were

used for model fitting. In-depth descriptions of the applied experimental methods can be found in Kallenberger et al. (2014). Collectively, 14 datasets, including 8 TL datasets, 4 PA datasets, and 2 SH datasets, were combined for model fitting. An overview of all data sets can be found in Table 3.1.

Table 3.1: Overview of data

Dataset	1	2	3	4	5	6	7
Data type	TL	TL	TL	TL	TL	TL	TL
Cell line	CD95	CD95	CD95	CD95	wild-type	wild-type	wild-type
Cell number	30	30	30	30	10	10	10
CD95L concentration	50ng/ml	500ng/ml	5000ng/ml	10000ng/ml	500ng/ml	2000ng/ml	5000ng/ml
Dataset	8	9	10	11	12	13	14
Data type	TL	PA	PA	PA, quantitative	PA, quantitative	SH	SH
Cell line	wild-type	CD95	CD95	CD95	wild-type	CD95	wild-type
Cell number	10	$2,04 \cdot 10^5$	$2,04 \cdot 10^5$	$2,04 \cdot 10^5$	$2,04 \cdot 10^5$	$5 \cdot 10^5$	$5 \cdot 10^5$
CD95 concentration	10000ng/ml	50ng/ml	500ng/ml	0ng/ml	0ng/ml	0ng/ml	0ng/ml

By combining experimental data of single-cell caspase-8 activities with population measurements of caspase-8 fragments from immunoblotting experiments, an effective mechanism of caspase-8 activation was determined, in which the cleavage reactions from p55 to p43 and from p30 to p18 are described as interdimeric “trans” reactions, catalyzed by neighbored DISC-bound caspase-8 intermediates, and the reactions from p43 to p18 and from p55 to p30 are intra-dimeric “cis” cleavage reactions, reflected by uni-molecular reaction kinetics Kallenberger et al. (2014) as illustrated in Figure 1 (b). The catalytically active fragments p43 and p18 can cleave BID to tBID, which is assumed to induce cell death at a certain threshold concentration. Model equations are given as follows, where $[*]$ are concentrations and k_* are reaction.

$$[CD95^*] = \frac{[CD95_{tot}]^3[CD95L]k_{DL}^2}{([CD95L]+k_{DL})([CD95_{tot}]^2k_{DL}^2+k_{DL}[CD95L]^2+2k_{DR}[CD95L]k_{DL}+k_{DR}k_{DL}^2}$$

$$\frac{d[CD95^*]}{dt} = -k_{on,FADD}[CD95^*][FADD]+k_{off,FADD}[DISC]$$

$$\frac{d[FADD]}{dt} = -k_{on,FADD}[CD95^*][FADD]+k_{off,FADD}[DISC]$$

$$\frac{d[p55]}{dt} = -k_{on,p55}[DISC][p55]$$

$$\begin{aligned}\frac{d[DISC]}{dt} &= k_{on,FADD}[CD95^*][FADD]-k_{off,FADD}[DISC]\\ &+k_{cl,D374,trans,p55}[p30]([DISC\_p55]+[p30])+k_{cl,D374,trans,p43}[p30][p43]\\ &+k_{cl,D216,cis}[p43]\end{aligned}$$

$$\begin{aligned}\frac{d[DISC\_p55]}{dt} &= k_{on,p55}[DISC][p55]-k_{cl,D216,cis}[DISC\_p55]\\ &-k_{cl,D374,trans,p55}[DISC\_p55]\,([DISC\_p55]+[p30])\\ &+k_{cl,D374,trans,p43}[DISC\_p55][p43]\end{aligned}$$

$$\begin{aligned}\frac{d[p30]}{dt} &= k_{cl,D216,cis}[DISC\_p55]-k_{cl,D374,trans,p55}[p30]([DISC\_p55]+[p30])\\ &-k_{cl,D374,trans,p43}[p30][p43]\end{aligned}$$

$$\begin{aligned}\frac{d[p43]}{dt} &= k_{cl,D374,trans,p55}[DISC\_p55]\,([DISC\_p55]+[p30])\\ &+k_{cl,D374,trans,p43}[DISC\_p55][p43]-k_{cl,D216,cis}[p43]\end{aligned}$$

$$\begin{aligned}\frac{d[p18]}{dt} &= k_{cl,D374,trans,p55}[p30]([DISC\_p55]+[p30])+k_{cl,D374,trans,p43}[p30][p43]\\ &+k_{cl,D216,cis}[p43]-k_{p18,inactive}[p18]\end{aligned}$$

$$\frac{d[tBID]}{dt} = k_{cl,BID}[BID]\,([p43]+[p18])$$

$$\frac{d[Pr_{ER}mGFP]}{dt} = -k_{cl,probe}[Pr_{ER}mGFP][p18]$$

$$\frac{d[Pr_{cpl}mCherry]}{dt} = -k_{cl,probe}[Pr_{cpl}mCherry]\,([p43]+[p18])$$

In the model, it was assumed that cell-to-cell variability can be exclusively attributed to differences in initial protein concentrations, reflected by a multi-variate log-normal distribution and that reaction parameters are equal for all cells. Therefore, random effects were included for the initial protein concentrations reaction. Additionally, random effects were described for cell volumes and cellular events (apoptosis time points). Reaction parameters were exclusively described by fixed effects.

In total, the model contains 24 effects, among which 13 effects contain only fixed effects, and 11 effects contain both fixed and random effects, resulting in 35 estimated parameters. The measurement noise variances of single-cell, mean and variance was approximated via experimental replicates.

3.3.1 Gradient computed by forward sensitivity method is accurate and efficient

As we use a gradient-based method for optimization, it is important to make sure that the gradient is accurate. Therefore, we compared the gradients calculated via forward sensitivities based on Equation (2.32), which were used during optimization, to finite differences based on Equation (2.28) (Raue et al., 2013) for both tests (Figure 3.3 (a), (b), (c)) and application (Figure 3.3 (d), (e), (f)) examples. We found that the forward sensitivity gradient values agree well with the values computed using the forward finite difference method (Figure 3.3 (a), (d)). While the finite difference method is sensitive to the step size (Figure 3.3 (b) (e)), the results using forward sensitivities are more stable, as the error can be controlled during integration. Moreover, the computation time using forward finite differences, especially for the TL data in the application problem (Section 3.3) with higher complexity, was two orders of magnitude faster (Figure 3.3 (f)). While in the simple model as the conversion reaction model (Section 3.3), the computation time using forward sensitivity is not significantly improved (Figure 3.3 (c)). Yet, the forward sensitivity method is still preferable to avoid step size dependency. Therefore, we used the forward sensitivity method in this chapter for both test and application problems.

3.3.2 Inner optimization loop has good convergence performance

In the routine of computing the log-likelihood of the TL data using the first-order conditional approximation method, an inner optimization loop is implemented to compute the optimal value of random effects $\hat{b}$ for each single cell (Figure 3.1 (c), Equation (2.18)). In this inner multi-start optimization, we used 20 different start points.

To test the convergence performance of the inner optimization, in the extrinsic apoptosis model, we increased the number of start points from 20 to 500, with fixed values of fixed effects β and covariance matrix D, and did the optimization for all 8 TL datasets separately. Two scenarios are tested, 1) the fixed values equal to the optima. In this case, over all 8 datasets in the worst case 188 out of 500 starts converged. 2) the fixed values equal to a random start point. In this case, 61 starts converged in the worst case of 8 datasets. This implies that in the 20 different start points of the inner loop, the probability of having at least one point that converges to the optimum was 92.6%, which appears reasonably high.

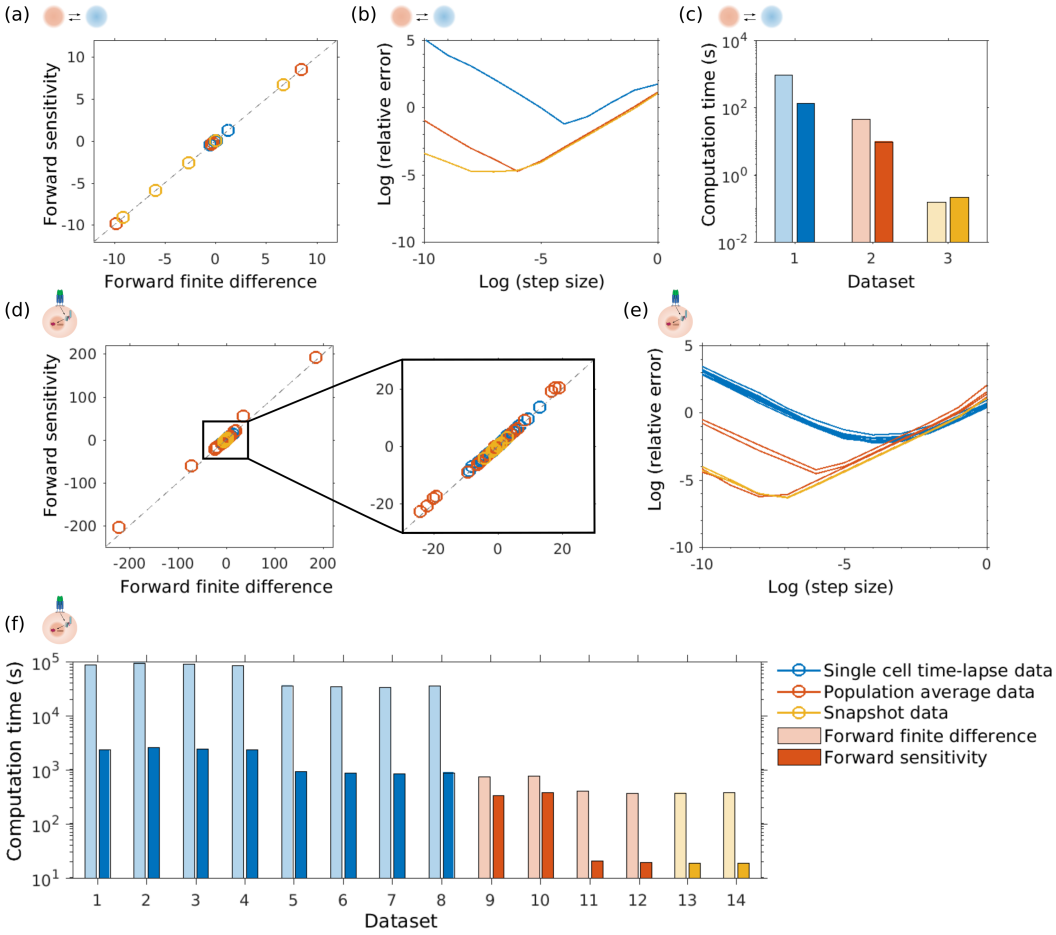

Figure 3.3: **Comparison of computing gradient values using forward sensitivity and forward finite difference method.** (a) Comparison of gradient values computed by forward sensitivity and forward finite difference method with the step size 10^{-4} for the conversion reaction model. (b) Relative error of gradient computed by forward finite difference method with different step sizes, the ground truth is computed using forward sensitivity method. (c) Computation time of the two methods for 3 data sets with different data types. (d) Comparison of gradient values computed by forward sensitivity and forward finite difference method with the step size 10^{-2} for caspase-8 model. (e) Relative error of gradient computed by forward finite difference method with different step sizes. (f) Computation time of the two methods for all 14 data sets.

3.3.3 Monte Carlo sampling method estimates the mean and variance with better accuracy

Different methods have been developed to compute the mean and variance of the data, with a trade-off between approximation error and computation time. The SP method is an alternative method to improve the computational efficiency compared to MC sampling methods, but it does not allow to improve accuracy, because of the fixed number of samples. To quantitatively investigate whether the SP method can be used to improve the efficiency of our model, we computed the relative errors of mean and variance using the SP method and MC sampling with different sample numbers.

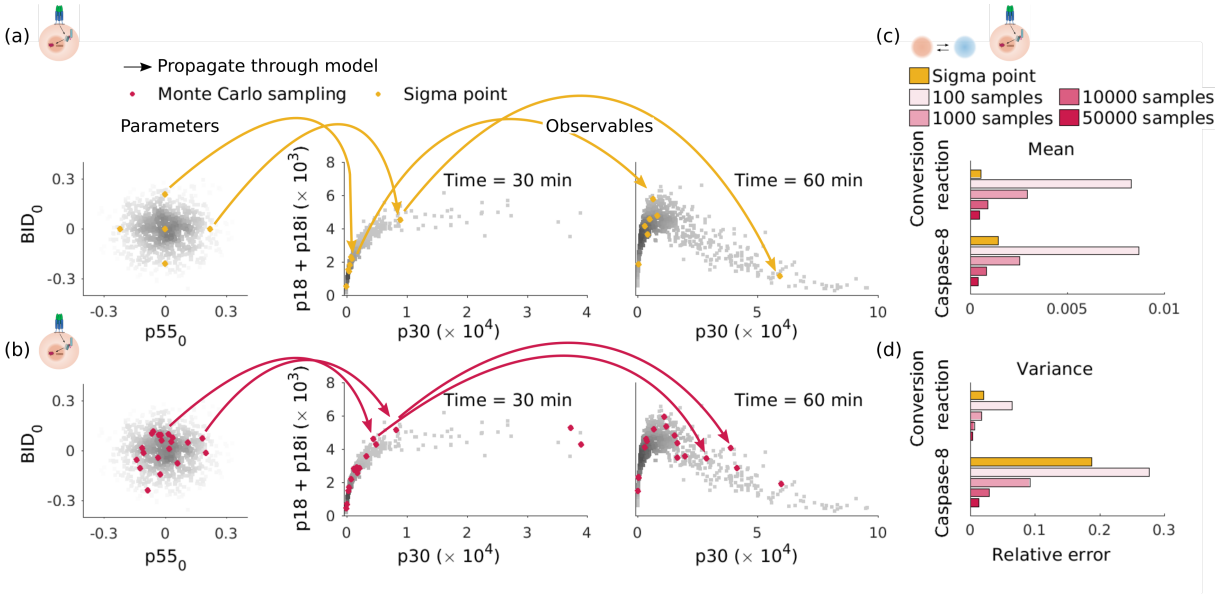

Figure 3.4: **Comparison of SP and MC sampling method.** (a) SP (yellow) samples of parameters propagate through the models. We use the two parameters BID_0 and $p55$ as an example. The five SP samples are propagated through the model and the corresponding observables are plotted. The correlation of two observables is shown using 1,000 MC samples (grey) at two time points, 30min and 60min for comparison. (b) MC samples (red) of parameters propagate through the model. To avoid overlapping each other, here only 20 MC samples are shown to compare with the distribution predicted by SP methods. The relative error of the mean (c) and variance (d) was predicted using SP and MC sampling methods with different sample sizes.

For comparison, we approximated the ground truth m^{true} using MC sampling with 100,000 samples. The relative error r is computed by

$$r = \frac{\sum_{i=1}^{d_m} \sum_{j=1}^{d_t} |m_{i,j} - m_{i,j}^{\text{true}}|}{\sum_{i=1}^{d_m} \sum_{j=1}^{d_t} |m_{i,j}^{\text{true}}|},$$

where m is the result computed by the SP or the sampling method (mean or variance), and d_m is the dimension of m.

To better understand the difference between the SP and MC sampling method, we visualized the scatter plot of the simulated data computed by MC samples and the SP samples at different time points (Figure 3.4 (a), (b)). We observed that because of the limitation of fixed sample number, the SP samples do not allow for a good approximation of the observable samples.

Regarding the conversion reaction model, the error of the SP method is comparable to the MC sampling method with 10,000 samples (0.0005 and 0.0009 for mean and 0.02 and 0.0059 for variance). Therefore, the SP method is an alternative for smaller models, in which a smaller number of samples is already enough to capture the shape of observations. However, for a more complex application such as the extrinsic apoptosis pathway model, the error of the SP method, especially for the variance (0.0014 and 0.0008 for mean and 0.1870 and 0.0292 for variance), is much larger than the MC sampling method with 10,000 samples. Considering that the computation time scales linearly with the number of samples, and the fact that using 50,000 samples does not have a large improvement (from 0.0008 to 0.0004 for mean and from 0.0292 to

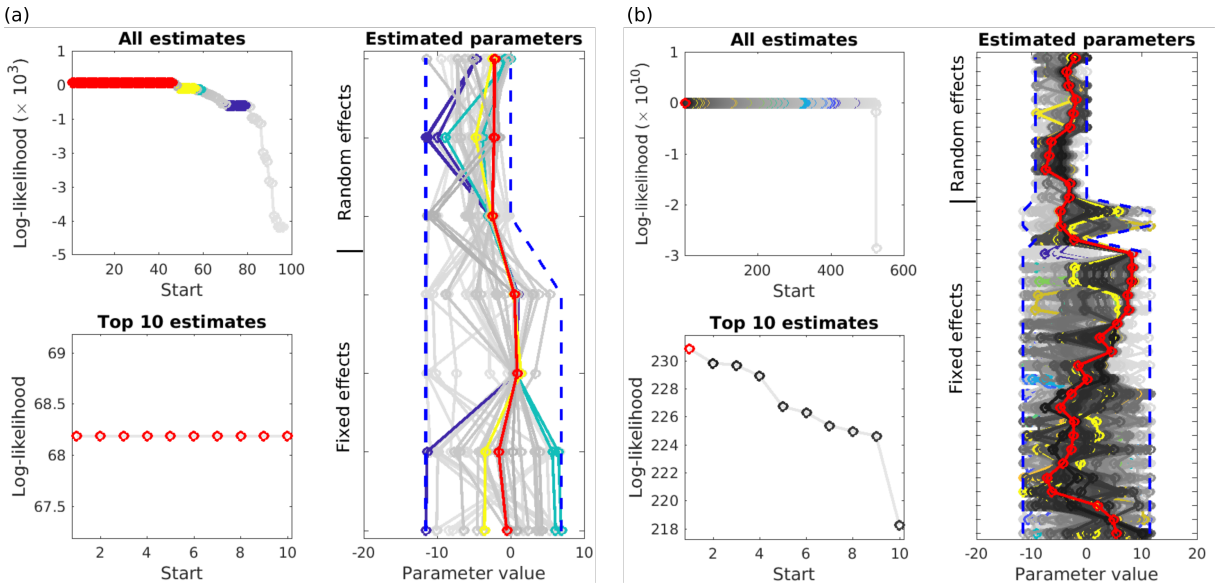

Figure 3.5: **Analysis of reproducibility of optimization results for (a) conversion reaction model and (b) caspase-8 model.** In each subplot, the left two plots show all/top 10 log-likelihood values across all optimization runs. The highest objective function values are colored in red. For other optimization runs, coloring is only applied if multiple optimization runs yield objective function values within 0.01 of each other, otherwise, a grey color is applied. Crashed optimization runs are not shown.

0.0129 for variance) (Figure 3.4 (c), (d)), we used the MC sampling method with 10,000 samples in this chapter.

3.3.4 Left-out data cannot be predicted by other data types

To analyze the relative relevance of individual datasets, we performed parameter optimization using different dataset combinations. When one or two data types are left out, we tried to predict them by using the parameters estimated from the other data types (Figure 3.1 (a)). Besides, we also predicted the behavior of one cell line based on a model calibrated with data obtained from another cell line (wild-type HeLa and CD95-HeLa cells).

For this purpose, we defined seven different data type combinations: (1) all data sets, (2) TL data left out, (3) PA data left out, (4) SH data left out, (5) both PA and SH data left out, (6) both TL and SH data left out, (7) wild-type cell line left out (for the apoptosis model). We performed the optimization using the included data sets and predicted the left-out data sets. Due to higher model complexity, the number of converged start points is less for the apoptosis model compared to the conversion reaction model (Figure 3.5).

For the conversion reaction model, we considered scenarios (1)-(6) (Figure 3.6). In Figure 3.6 (a), each row corresponds to one scenario, while each column corresponds to a data type. The green color means that the data type is predicted using the other types in the same row, while the other colors mean that the corresponding data types are used for parameter estimation.

Comparing the artificial data (grey) and the simulated results (colored) of the conversion reaction

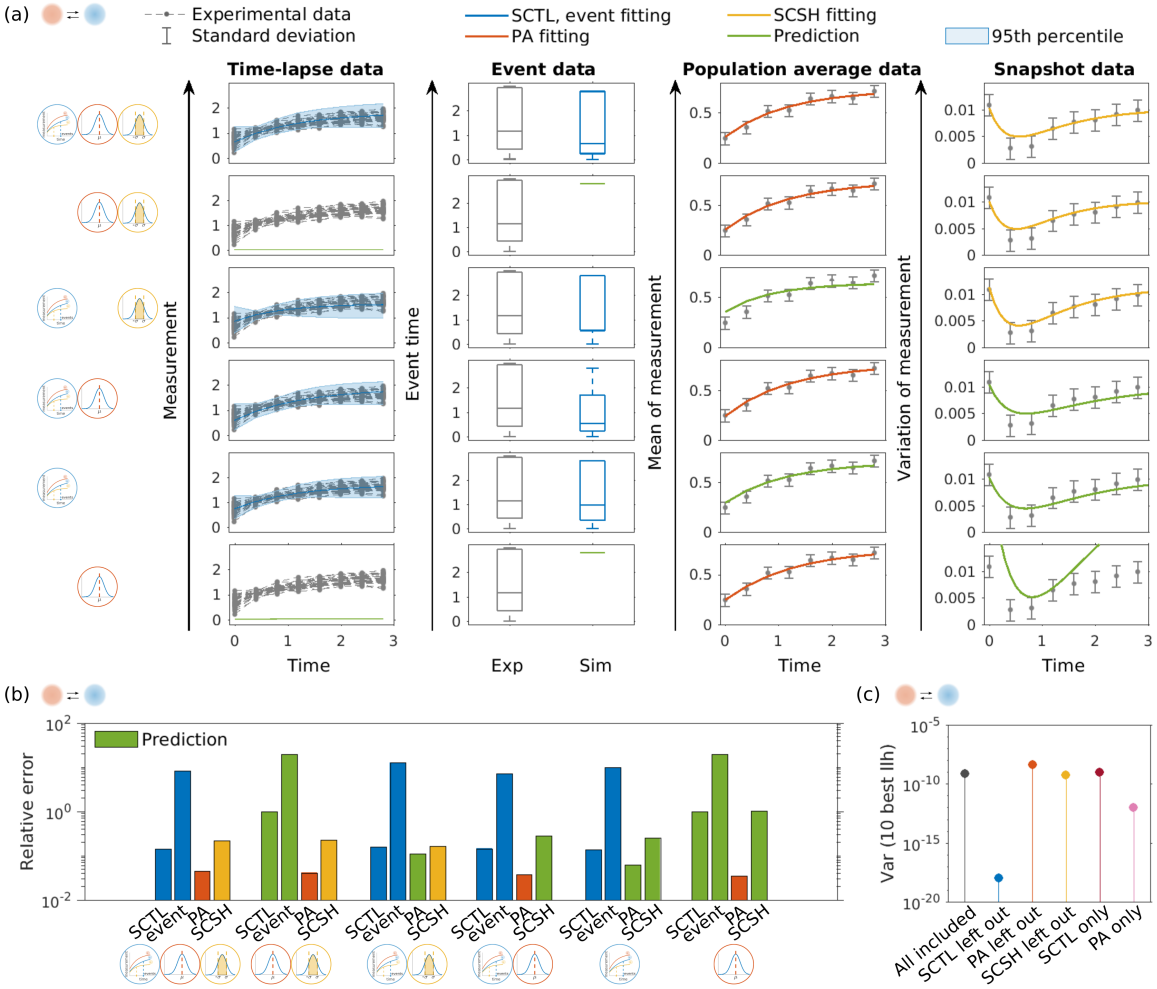

Figure 3.6: **Comparison of different data combinations for conversion reaction model.** (a) Fitting the conversion reaction model using different data type combinations. Top to bottom: Different scenarios of including/excluding data in the optimization. Left to right: Measured data (black), fits (data type specific colors), and predictions (green) for TL, time-to-event, PA, and SH data. The standard error of the PA data (second to right) and the SH data (right) are shown in black. The interval between the 5th and the 95th percentile of the errors across 1000 realization of single cells are depicted (left) in blue shadow. (b) Relative error of fitted (data type specific colors) and predicted (green) data sets compared to the artificial data. (c) Variance of 10 best log-likelihood values of different data combinations.

model for TL (blue, first column), time-to-event (blue, second column), PA (red, third column), and SH (yellow, last column) data, we generally observed good agreement for the training data (Figure 3.6 (a), scenarios as rows). We take scenario 1, with all data types included for parameter estimation, as a baseline. The relative errors are TL: 0.15, time-to-event: 8.24, PA: 0.06, SH: 0.22.

Generally, when one data type is excluded, the error between predicted (green) and artificial data becomes larger (Figure 3.6 (b), scenarios on x-axis). In cases where not all data sets were included for the optimization, we found that the prediction was the most inaccurate when the TL data (including time-to-event data) were left out (scenario 2, with a relative error value of 0.99 for the TL data and 19.7 for the time-to-event data), because the scaling factor and

the threshold cannot be estimated from the other two data types (See also Section 3.3.5 and Figure 3.9). When the scaling factor is provided as a known value to the model, TL data can still be predicted. Therefore, it is still possible to do the prediction if one can get information on the scaling factor from other sources. The prediction of the other two data types agreed well (with the maximum relative error value of 0.29 for excluded data types) with the artificial data scenario 3-5, implying that the TL data contains most of the information of the other two data types in the conversion reaction model. Leaving out PA data (scenario 3), we could observe that the prediction of event data worsens compared to using only TL data (scenario 5). This might be due to possible inconsistencies between different data sets, as the single-cell data sets consisted of comparably small cell numbers per experimental condition. This results in a large cell-to-cell variability, which can be detected when obtaining PA data. We also performed optimization using only single data types, the TL data scenario 5 or PA data scenario 6. Compared to settings with two included data types, the agreement of the fitted data was the best because of no data inconsistency (relative errors: PA data: 0.0423 in scenario 6 compared to 0.05 in scenario 2 and 0.045 in scenario 4; TL data: 0.14 in scenario 5 compared to 0.16 in scenario 3 and 0.15 in scenario 4). With the TL data alone, we could still predict the PA and SH data (relative errors: PA: 0.08, SH: 0.25 compared with 0.06 and 0.22 in scenario 1, with all data types included). However, with only PA data, the prediction of the other two data types was off (relative errors: TL: 0.98, time-to-event: 19.7, SH: 1.02).

Optimization convergence was best when excluding the TL data (Figure 3.6 (c)), indicating that these were the most challenging data type in the optimization. Notwithstanding, as seen above, their inclusion is important when aiming to capture single-cell dynamics.

Given the computational complexity of the apoptosis model and the experience with the conversion reaction model, it is unlikely that two data types can be predicted by using only one for the optimization as in scenarios 5 and 6. Therefore, we considered only scenarios (1)-(4) and (7) (Figure 3.7). Same as in the conversion reaction case, each row corresponds to one scenario, while each column corresponds to a data type. The green color means that the data type is predicted using the other types in the same row. As for the conversion reaction model, the agreement of the simulated data with the data used for optimization was good, as well as the single-cell fittings (Figure 3.8). Estimated parameter values can be found in Table 3.2. We also take the relative error of scenario 1, where all data types are used for parameter estimation, as a baseline for comparison. The relative errors are TL for CD95-overexpressing HeLa cell line: 0.35, TL for wild-type HeLa cell line: 0.30, time-to-event data: 0.41, PA: 0.12, SH: 0.12.

However, when leaving out any of the data types, the corresponding data could only be predicted with a relatively large error (scenario 2: TL for CD95-overexpressing HeLa cell line: 40.0, TL for wild-type HeLa cell line: 50.2, time-to-event data: 0.64; scenario 4: SH: 13.4) (Figure 3.7 (b)). In scenarios 3 and 5, the excluded data could not be predicted using the remaining data types (Figure 3.7 (a) (b)). Because some parameters are not shared between models for different data types or cell lines. When the SH data were left out, the fitting of event data became worse. Finally, when leaving out wild-type HeLa data (Figure 3.7 (a), scenario 5), the parameters could not be predicted, arguably because of cell line-specific parameters for initial protein concentrations. In principle, if one can get information of those parameters from other sources, data on the corresponding cell line can be predicted.

In line with the conversion reaction model, when TL data were left out, the optimization conver-

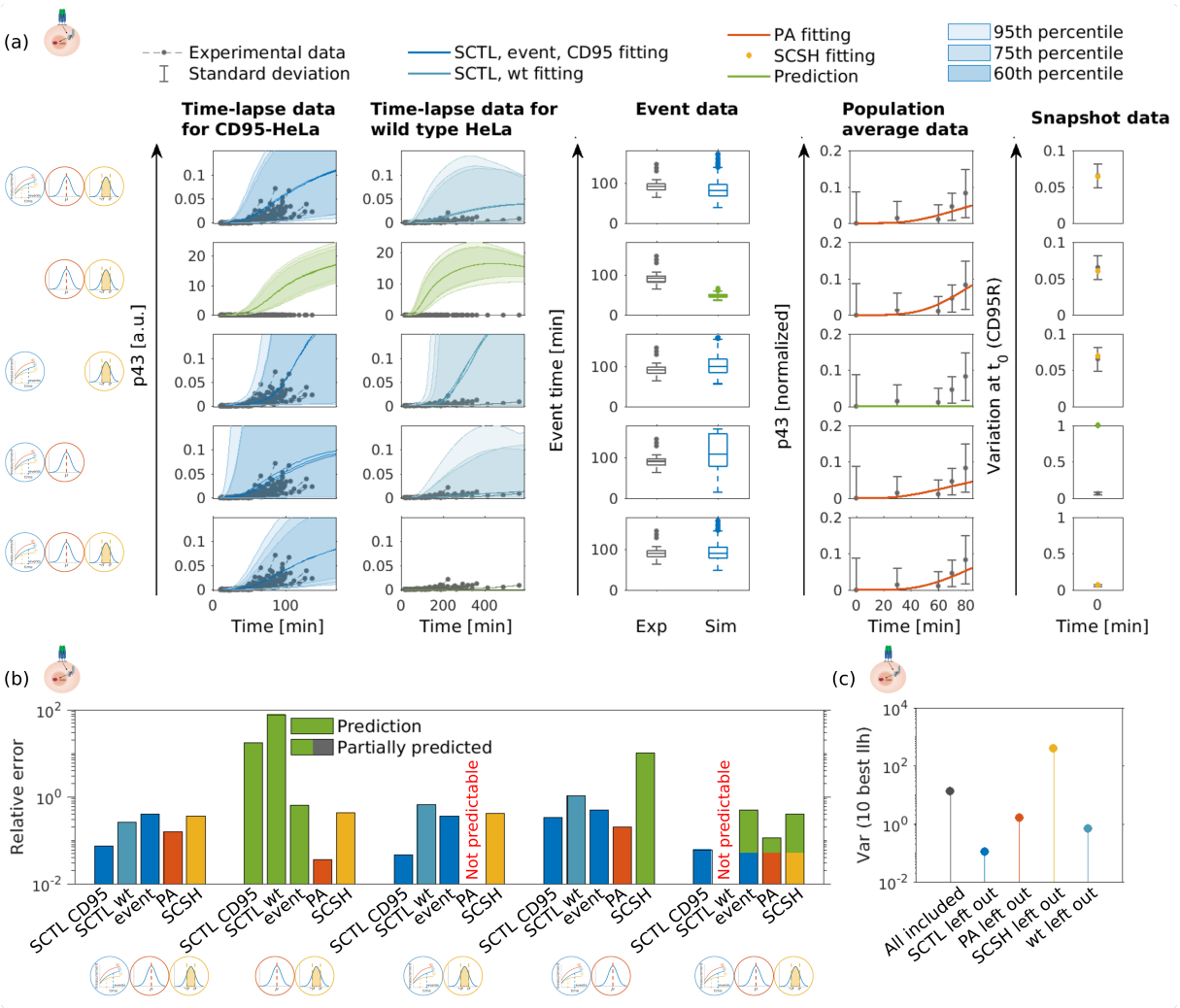

Figure 3.7: **Comparison of different data combinations for extrinsic apoptosis model.** (a) Fitting of the apoptosis model using different data type combinations. Top to bottom: Different scenarios of including/excluding data in the optimization. Left to right: Measured data (black), fits (data type specific colors), and predictions (green) for TL, time-to-event, PA, and SH data. The standard error of the PA data (second right) and the SH data (right) are shown in black. The interval between the 20th and 80th, 12.5th and 87.5th, 5th and the 95th percentile of the errors across 1000 realization of single cells are depicted (left for the CD95-HeLa cell line and second left for the wild-type HeLa cell line) in blue shadow. (b) Relative error of fitted (data type specific colors) and predicted (green) data sets compared to the artificial data. (c) Variance of 10 best log-likelihood of different data combinations.

gence performance was best (Figure 3.7 (c)) because of the reduced data complexity. When SH data were left out, the convergence performance became worse, which was also consistent with Figure 3.7 (a) (scenario 4). On the one hand, this demonstrates that different data types provide orthogonal information about the system, on the other hand, it clearly shows that including more data improves the predictive performance of the model.

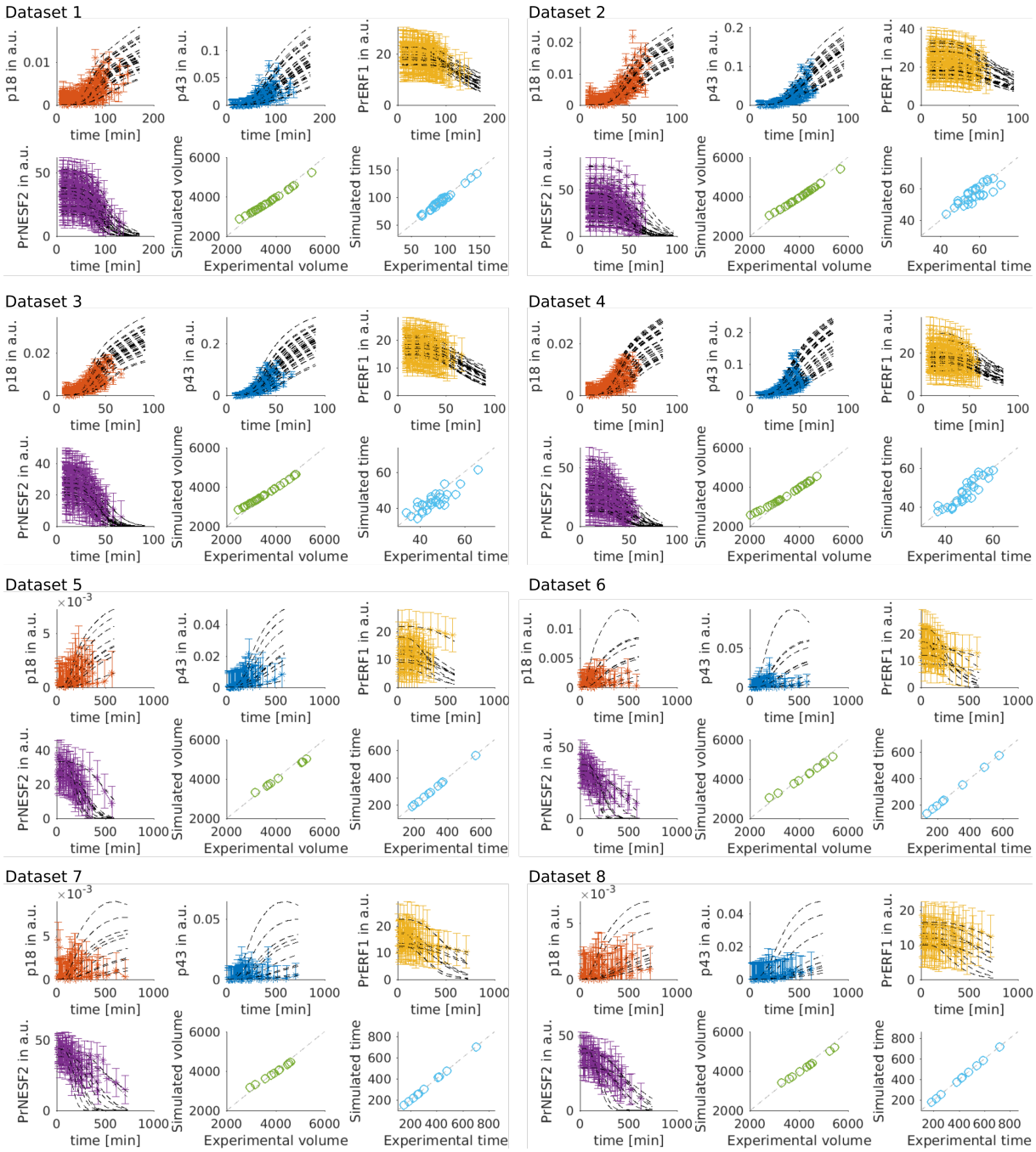

Figure 3.8: **Single-cell fittings of caspase-8 model.** Different colors are applied for different measurements in one experimental condition. 30 cells are measured for the first four datasets (experimental conditions), while 10 cells are measured for the last four datasets.

3.3.5 Integrating all data types improves parameter identifiability

To investigate how different data combinations affect the parameter identifiability, we assessed parameter uncertainty of the conversion reaction model by computing profile likelihoods (Raue et al., 2013) for all parameters (Figure 3.9, (Equation 2.41)). When the TL data were included, all parameters were identifiable (scenarios 1, 3, 4, 5). The parameters estimated using all data types

Table 3.2: **Estimated parameter values and confidence interval of the extrinsic apoptosis model.** (where / indicates that the specific parameter is not practically identifiable)

Parameter	K_{DR}	K_{DL}	$k_{on,FADD}$	$k_{off,FADD}$	$k_{on,p55}$	$k_{cl,D216,cis}$	$k_{cl,D374,trans,p55}$
Value	203.3	133.8	6.8	0.002	7×10^{-4}	0.01	0.06
95% CI	/	±0.4	/	±1.5	±0.9	±0.2	±1.0
Parameter	$k_{cl,D374,trans,p43}$	$k_{p18,inactive}$	$k_{cl,BID}$	$k_{cl,probe}$	$CD95_{tot,over}$	$CD95_{tot,wt}$	$FADD_0$
Value	0.09	0.1	0.008	0.06	1.2	0.2	91.3
95% CI	±0.8	±0.4	±0.6	±0.8	/	/	±0.1
Parameter	$p55_0$	BID_0	$Pr_{ER}mGFP_{over}$	$Pr_{ER}mGFP_{wt}$	$Pr_{cpl}mCherry_{over}$	$Pr_{cpl}mCherry_{wt}$	V_{c0}
Value	10.0	230.7	1951.6	1615.3	3595.6	3896.4	3770.5
95% CI	±0.6	±0.1	±0.3	±0.2	±0.5	±0.4	±0.02
Parameter	$tBID_{threshold}$	s_G	s_C	$\mathrm{Var}(CD95_{tot,over})$	$\mathrm{Var}(CD95_{tot,wt})$	$\mathrm{Var}(FADD_0)$	$\mathrm{Var}(p55_0)$
Value	0.1	0.009	0.008	0.04	0.05	5×10^{-4}	0.001
95% CI	±0.6	±0.3	±0.4	±0.2	±0.2	/	±0.8
Parameter	$\mathrm{Var}(BID_0)$	$\mathrm{Var}(Pr_{ER}mGFP_{over})$	$\mathrm{Var}(Pr_{ER}mGFP_{wt})$	$\mathrm{Var}(Pr_{cpl}mCherry_{over})$	$\mathrm{Var}(Pr_{cpl}mCherry_{wt})$	$\mathrm{Var}(V_{c0})$	$\mathrm{Var}(tBID_{threshold})$
Value	0.002	0.04	0.09	0.1	0.04	0.03	0.1
95% CI	±1.1	±1.0	±1.3	±1.0	±1.1	/	±1.2

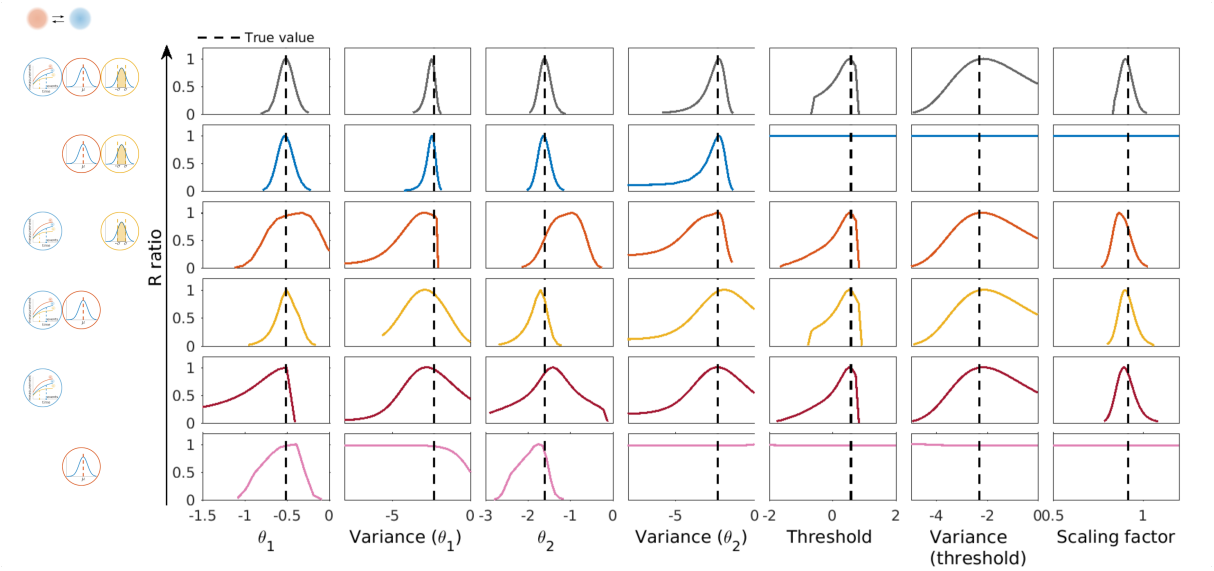

Figure 3.9: **Profile likelihood of all parameters optimized using different data type combinations.** Top to bottom: Different scenarios of including/excluding data in the optimization. Left to right: Model parameters. The true parameter values are indicated by black vertical lines. Curves with the data type specific colors indicate the calculated profiles, which were initialized at the optimized parameter values (usually at the peak of the curves).

are close to the true parameter values with narrow confidence intervals (scenario 1), compared to the other data type combinations. With only the TL data left out, the confidence interval of parameter $\theta = [\theta_1, \theta_2]$ was roughly the same as when using all data types (scenario 2), but the event threshold and scaling factor were totally unidentifiable, which is consistent with the predictions in Figure 3.6. Leaving out the SH data yielded better parameter identifiability than leaving out the population data (scenarios 3 and 4), indicating that the PA data contains more information about the parameters. With only TL data included, all parameters were still identifiable, but with much wider confidence intervals (scenario 5).

For the extrinsic apoptosis model, profile calculation would have been prohibitively computationally expensive for the apoptosis model. Instead, we quantified the parameter uncertainty via

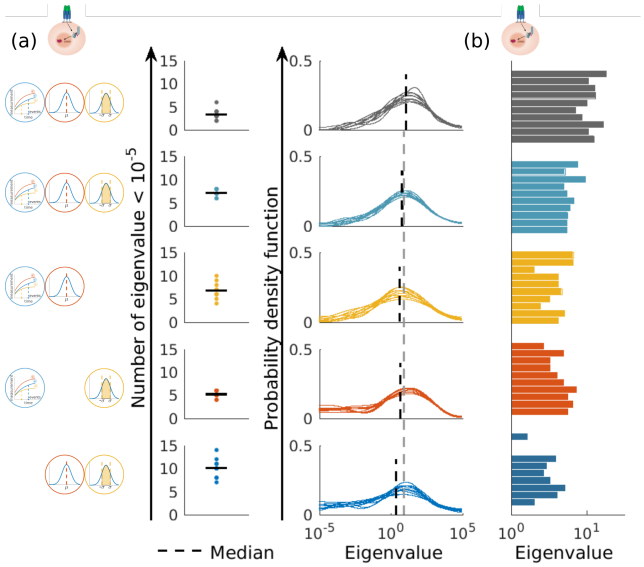

Figure 3.10: **Parameter identifiability for the caspase-8 model assessed via Hessian eigenvalues** (a) Eigenvalue probability density of the Hessian matrices. Frequency of eigenvalues below 10^{-5} (out of 35 eigenvalues) are plotted on the left. The probability density functions of the best 10 optimization runs for different data type combinations are shown in the right column, restricted to values above 10^{-5}, using density plots. The median values of the frequency and the probability density function are depicted as black (dotted and solid) lines. Smaller eigenvalues correspond to larger uncertainties. (b) Median of the eigenvalues of the 10 best runs.

eigenvalues of the Hessian matrices at the optimal parameter vector based on Equation (2.44), using a local approximation (Fröhlich et al., 2018a). Therefore, the larger the eigenvalues are, the smaller are the confidence intervals. Estimated confidence intervals can be found in Table 3.2. In Figure 3.10 (a) (right column), the spectrum of the 35 eigenvalues per model was visualized via a kernel density estimator, with small eigenvalues indicating higher uncertainties. As before, the uncertainty was the smallest when all data types were included in the optimization process (1st row). When the wild-type cell line was left out, the number of the eigenvalues smaller than 10^{-5} was larger, as the cell line specific parameters could not be optimized at all. However, the median of the eigenvalue spectrum was not substantially smaller than the median, when both cell lines are included, as still all the data types were used for fitting (2nd row). Similar to the conversion reaction model, leaving out the PA data yielded more eigenvalues numerically close to 0 ($< 10^{-5}$) than leaving out the SH data (3rd and 4th row), meaning that the PA data are more informative. When we left out the TL data, the parameter uncertainty was the largest. In line with Figure 3.7 (c), when we left out SH data, the fluctuation of the median values is the highest compared to all other scenarios, meaning that the estimated parameter values are largely different between optimization runs.

In conclusion, it is necessary to include all three data types to ensure the best parameter identifiability. By comparing different combinations using the two models, we show that the number of identifiable parameters is lowest when TL data are excluded while leaving out the PA data yields the second worst identifiability.

3.4 Discussion

3.4.1 Summary and conclusion

Cell-to-cell variability is widely observed in various biological experiments. Non-linear MEMs are a powerful approach to account for this variability in mechanistic models. Depending on the experimental method, different types of data can be measured, giving measurements resolved on the single-cell level, or only population statistics. In this chapter, we demonstrated that different data types can be jointly integrated into a mixed effects modeling framework. Specifically, we combined TL data, PA data, and SH data.

To fit parameters, we employed a gradient-based optimization approach. We compared gradients computed by our in-house MatLab toolbox PESTO with the finite difference method. Results show that our computation method based on forward sensitivity is more robust as it does not depend on the step size, and the computational efficiency is also higher as the complexity of the model increases.

We applied the method to two problems, a simple synthetic conversion reaction model, and a model of extrinsic apoptosis. We demonstrated substantially improved performance in terms of data fits when combining all three data types, compared to only a subset. We comparede the identifiability of different dataset combinations by computing the profile likelihoods of the conversion reaction model and the eigenvalues of parameters for the extrinsic apoptosis pathway model. We observed by including all data types, most parameters were identifiable.

Furthermore, we also estimated the parameters of two cell lines simultaneously in the same model in the application problem. We show that it is possible to define both cell line specific and shared parameters in MEMs using our group-developed MatLab toolbox MEMOIR. Furthermore, we show that without knowledge of the cell line specific parameters, data of the corresponding cell line cannot be predicted.

In summary, we developed an approach to integrate different data types in a mixed effects modeling framework. In contrast to using only one data type, combining all data types led to improved fits of both our test and application models to the data. Additionally, we could increase the number of identifiable parameters. Therefore, by including more data types in a mechanistic model, more insights from different aspects can be gained. We anticipate that such approaches will improve our understanding of complex biological pathways by leveraging information from different sources.

3.4.2 Outlook

A limitation of our approach is that we did not specifically set weight values for any of the data types when computing the objective function as shown in Equation (3.1), meaning that data types with a larger number of observations have a high impact on the estimated parameters. In our optimization routine, it is possible to estimate parameters shared and independent in different cell lines simultaneously. The same method of defining parameters can also be used for single-cell measurements with subgroups. Moreover, it is also possible to weight different cell

lines (subgroups) differently. An interesting direction of further research would be to investigate methods that choose weights to balance the impact of e.g. different data types and cell lines (Schälte and Hasenauer, 2023) as follows,

$$J(\theta) = \sum_{l=1}^{d_l} w_{TL}^l J_{TL}^l(\theta) + \sum_{l=1}^{d_l} w_{PA}^l J_{PA}^l(\theta) + \sum_{l=1}^{d_l} w_{SH}^l J_{SH}^l(\theta), \tag{3.4}$$

with w represents the wights, l and d_l represent the cell lines and the corresponding dimension. However, a fair weighting of heterogeneous data is an open problem.

Furthermore, the covariance matrix D of the random effects is diagonal, i.e. there is no correlation between random effects, meaning that we assume all initial values are independent in our biological model. In general, to further analyze covariances in biological systems, it would be possible to add covariance parameters in D (Adlung et al., 2021). This would increase the parameter space but is conceptually independent of the data integration approach presented in this work.

Our method was applied to two models with different complexity, the simple model of a conversion reaction, and the extrinsic apoptosis model. To ensure the approximation accuracy of the PA and SH data, we used MC sampling method with 10,000 samples instead of SP methods. However, for the simple conversion reaction model, the SP method can potentially decrease the computation time while keeping the approximation accuracy high at the same time (Figure 3.4). Regarding the TL data, our approach scales linearly with respect to the number of single cells, the application to higher throughput experiments can still be problematic because of the computation time for hundreds of cells. One could investigate methods using e.g. subsampling or mini-batch approaches to scale to such scenarios.

Chapter 4

Cramér-von Mises Distance based method to approximate moments

Mixed effect modeling is widely used to study cell-to-cell and patient-to-patient variability. When population data is used in the model, the population statistics is usually approximated using Monte Carlo (MC) methods or alternatively sigma point (SP) methods. However, while better accuracy can be achieved via MC methods with a larger number of samples, the computation time is also considerably high. On the other hand, SP methods are computationally more efficient with much smaller sample sizes, they do not allow for flexible handling of accuracy because of the fixed number of samples. Both MC methods and SP methods can be considered as Dirac mixture distributions, which is a collection of Dirac components that can be weighted.

In this Chapter, we propose the use of a method based on the Cramér-von Mises Distance, which has been introduced in the context of filtering. We assess the accuracy of the different methods using several problems and provide the first scalability study for the Cramér-von Mises Distance (CMD) method. Our results indicate that for a given number of points, the method based on the modified Cramér-von Mises Distance method tends to achieve a better approximation accuracy than MC and quasi Monte-Carlo (QMC) methods with a smaller sample size. Furthermore, in contrast to SP methods, the method based on the modified Cramér-von Mises Distance allows for a flexible number of points and a more accurate approximation for nonlinear problems.

This chapter is based on and partly identical to the following publication, in which implementation of the CMD methods, the application models and data analyses are performed by the thesis author:

- **Wang, D**, Stapor, P, Hasenauer, J, Dirac mixture distributions for the approximation of mixed effects models. IFAC-PapersOnLine, 2019, 52(26), 200-206.

4.1 Introduction

When population average and snapshot data are used for parameter estimation, to simulate MEMs, the parameters of individual cells/patients are sampled and the nonlinear function is evaluated for these sampled parameters. The sampling essential yields a Dirac mixture distribution, which is simply a collection of Dirac components that can be weighted. To obtain robust estimates for the population statistics, e.g. mean and standard deviation, a substantial number of Monte-Carlo (MC) samples is required, causing a significant computational demand. To reduce the required number of samples, alternative Dirac mixture distributions can be employed. One alternative is the use of QMC methods, which generate a set of deterministic samples, thus resulting in reduced sample sizes while keeping higher accuracy. However, to the best of our knowledge, this has not been published and used in the context of MEMs. In contrast to standard MC methods, QMC methods such as Sobol (Sobol, 1967) and Halton (Halton, 1964) are based on low discrepancy sequences (Niederreiter, 1978). The use of low discrepancy sequences reduces the randomness and improves the robustness and convergence order.

SP methods are also an alternative to MC methods (Julier et al., 1995; Julier and Uhlmann, 2004; Menegaz et al., 2011; Lerner, 2002; van der Merwe, 2004; Charalampidis and Papavassilopoulos, 2011). These methods aim to approximate the mean and covariance matrix of the parameters by sampling deterministically (van der Merwe, 2004). Then the approximated mean and covariance matrix of outputs are computed by propagating each of the samples through the model(Filippi et al., 2016; Loos et al., 2018b). For linear models, usually, the approximated moments are the true values. However, a problem is that the error is difficult to control when models are nonlinear. Besides, for most SP methods, the number of samples is fixed or follows a certain formula, which makes it hard to flexibly improve the accuracy by increasing the number of samples, as for the MS and QMC methods.

In this chapter, we propose the approximation of population statistics with Dirac mixture distributions constructed by minimizing a modified Cramér-von Mises Distance (CMD). The use of the CMD has been proposed in the context of state estimation, also known as filtering (Gilitschenski and Hanebeck, 2013). Here, we provide a first scalability analysis for CMD methods as well as a comparison with MC, QMC, and SP methods for the analysis of nonlinear MEMs. Results show that with the same number of samples, the approximated moments are more accurate than the ones approximated by MC and QMC methods. Furthermore, compared to the a SP methods, the accuracy of CMD methods can be flexibly controlled by adjusting the number of samples used.

4.2 Background

4.2.1 Dirac mixture distributions

Dirac distribution is also known as the unit impulse, whose value is zero everywhere except at $x = 0$, and whose integral over the entire real line is equal to one. In the case of one dimension:

$$\delta(x) = \begin{cases} +\infty, & x = 0 \\ 0, & x \neq 0 \end{cases} \tag{4.1}$$

In n-dimensional case, with $x = (x_1, x_2, \ldots, x_n)$, the Dirac distribution is the product of the 1-dimensional Dirac distributions in each direction, separately,

$$\delta(x) = \delta(x_1)\,\delta(x_2)\ldots\delta(x_n). \tag{4.2}$$

Therefore, Dirac distribution can be considered as a discrete point distribution.

Taking a 2-dimensional case as an example, according to the equations above, a Dirac mixture distribution at the point (m_x, m_y) is defined as

$$\delta(x - m_x, y - m_y) = \delta(x - m_x)\,\delta(y - m_y). \tag{4.3}$$

4.2.2 Cramér-von Mises distance based on localized cumulative distribution

In order to approximate the difference between two distributions, Cramér-von Mises distance is used in this study. Cramér-von Mises distance is defined as the difference between the corresponding cumulative distributions. However, for Dirac mixture distributions with dimension higher than 1, there are several possible definitions of the cumulative distribution. Therefore, we introduce in this section the definition of localized cumulative distribution and Cramér-von Mises distance based on this definition.

Localized cumulative distribution

We still consider the 2-dimensional distribution as an example. Unlike the case in 1-dimension, there are four different possible definitions of the cumulative distribution of $f(X, Y)$ (Hanebeck and Klumpp, 2008):

- $F_1(X, Y) = p(x \leq X, y \leq Y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u, v)\,\mathrm{d}v\mathrm{d}u$
- $F_2(X, Y) = p(x > X, y > Y) = \int_{x}^{\infty}\int_{y}^{\infty} f(u, v)\,\mathrm{d}v\mathrm{d}u$
- $F_3(X, Y) = p(x \leq X, y > Y) = \int_{-\infty}^{x}\int_{y}^{\infty} f(u, v)\,\mathrm{d}v\mathrm{d}u$
- $F_4(X, Y) = p(x > X, y \leq Y) = \int_{x}^{\infty}\int_{-\infty}^{y} f(u, v)\,\mathrm{d}v\mathrm{d}u$

In order to avoid this problem, Hanebeck and Klumpp (2008) proposed the localized cumulative distribution as follows,

$$F\left(x, b_l\right) = p\left(|\bar{x} - x| \leq b\right) \tag{4.4}$$

where $\bar{x}$ is a random vector that can be characterized by a probability density function f. b_l defines the range around point x where the cumulative distribution is computed.

Modified Cramér-von Mises distance

Cramér-von Mises distance D^c defines the distance between two distributions $f_1(x, y),\ f_2(x, y)$. This is computed by the difference of the cumulative distributions $F_1(x, y),\ F_2(x, y)$ multiplied by a weight function ω:

$$D^c\left(m_x, m_y\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \omega\left(x, y\right)\left(F_1\left(x, y\right) - F_2\left(x, y\right)\right)^2 \mathrm{d}x\mathrm{d}y \tag{4.5}$$

Based on the definition of localized cumulative distribution, Hanebeck and Klumpp (2008) the proposed modified Cramér-von Mises distance D^{cm}, which is suitable for comparing higher dimensional distributions,

$$D^{cm} = \int_{\mathbb{R}^N}\int_{\mathbb{R}_+^N} \omega\left(x, b_l\right)\left(F_1\left(x, b_l\right) - F_2\left(x, b_l\right)\right)^2 \mathrm{d}b\mathrm{d}x. \tag{4.6}$$

4.3 CMD and sampling based approximation method

4.3.1 MC method

According to MC methods, the random effects $b^{(i)}$ can be obtained by randomly sampling over a multi-variate normal distribution $\mathcal{N}(0, D)$. The weights are the same for each MC sample, with $w_\mu^{(i)} = w_\Sigma^{(i)} = \frac{1}{n}$, with n denotes the sample size.

MC methods yield unbiased results because of random and representative sampling. However, this comes at the cost of reduced computational efficiency. It is well known that estimates obtained using small sample sizes n possess a large variance and that the convergence rate is only $\mathcal{O}(n^{-1/2})$. However, if the number of samples n is limited, e.g. due to the computational complexity of model simulation, even biased approaches might be beneficial.

4.3.2 Quasi-Monte Carlo methods

One alternative to the standard MC methods is the use of QMC methods. QMC methods address problems of MC methods by using low discrepancy sequences. These sequences are more regular than random samples used by MC methods (Figure 4.1 (a). This improves the robustness and convergence rate to $\mathcal{O}(n^{-1})$ (Caflisch, 1998). The points of the sequence are used in the same way

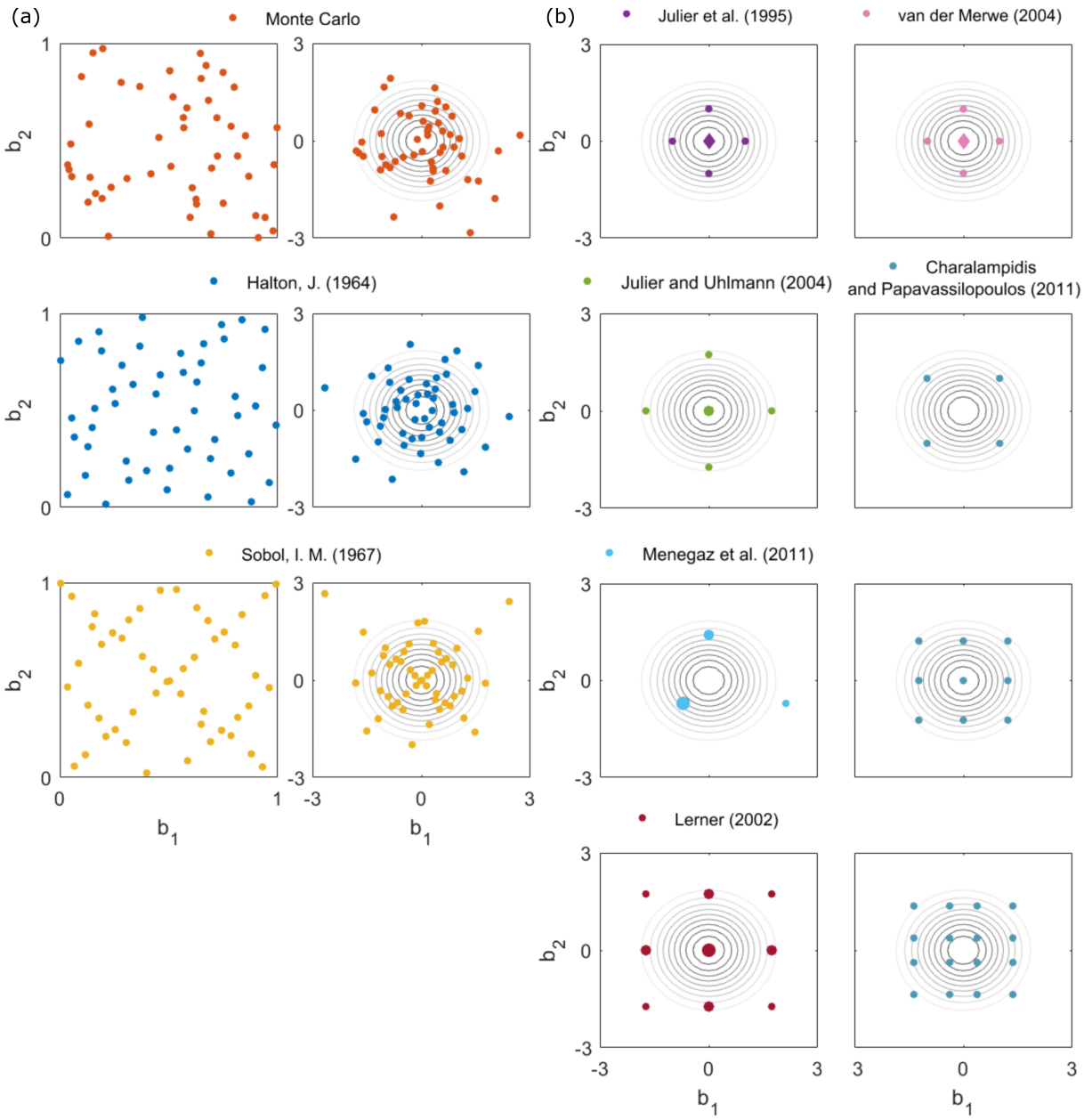

Figure 4.1: **Illustration of MC, QMC, and SP methods in two dimensions.** (a) Random sequence used by the MC method, and deterministic sampling of Halton and Sobol sequence used by QMC methods: (left) sequences in the unit cube, and (right) transformed sequences to approximate standard normal distribution with a mean of 0 and variance of 1. As a reference, level sets of the standard normal distribution are shown as gray contour lines. (b) Point distributions and weights for different SP methods. Positive point weights are indicated by a circle, while negative weights are by a diamond. The absolute values of the weights is encoded by the marker size (small markers for small absolute values, and large markers for large absolute values). Among others, it is possible for the Charalampidis method to have different sample sizes: $n = \theta^{d_r}$, where θ can be any integer and d_r is the dimension. Here we show the cases with $\theta = 2, 3, 4$

as MC samples and the weights are also the same for each sample. In this section, we consider two QMC methods, Halton (1964) and Sobol (1967) method.

The Halton method generates a sample sequence constructed according to the deterministic method that uses co-prime numbers. The nth number of the Halton sequence is the number n written in binary representation, inverted, and written after the decimal point. In our implementation of high dimensional sequences, the bases are prime numbers starting from 2.

The original motivation of the Sobol method is to construct a sequence s_n in I^{d_r}, where $I^{d_r} = [0,1]^{d_r}$ is the d_r-dimensional unit hypercube, so that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f(x_i) = \int_{I^{d_r}} f(x)\,\mathrm{d}x \tag{4.7}$$

where f is a real integrable function over I^{d_r}.

4.3.3 Sigma point methods

In this section, we consider six different SP methods (Julier et al., 1995; Julier and Uhlmann, 2004; Menegaz et al., 2011; Lerner, 2002; van der Merwe, 2004; Charalampidis and Papavassilopoulos, 2011). These are in our opinion the most widely used SP methods. The methods differ in the number of points n, or their locations and/or weights (Figure 4.1(b)). For details on the selection of the point locations, we refer to the original publications. A comparison of the SP methods and CMD method regarding the number of samples and the moments which can be exactly approximated is provided in Table 4.1.

Table 4.1: **Properties of SP and CMD methods.** The references, number of points n (for an d_r-dimension random effect vector), and moments of Gaussian distribution which are matched precisely are listed.

Method	Number of points	Exact moments
Julier et al. (1995)	$2d_r + 1$	1st, 2nd
Julier and Uhlmann (2004)	$2d_r + 1$	1st, 2nd
Menegaz et al. (2011)	$d_r + 1$	1st, 2nd
Lerner (2002)	$2d_r^2 + 1$	1st, 2nd
van der Merwe (2004)	$2d_r + 1$	1st, 2nd
Charalampidis and Papavassilopoulos (2011)	c^{d_r}	1st, 2nd
CMD	flexible	1st

In the following, we introduce the different SP methods in detail, including the sample methods and the weights of generated mean $w_{\mathbb{E}}$ and variance $w_{\mathbb{C}}$. Where $S = D^{\frac{1}{2}}$ is the Cholesky decomposition of the covariance matrix.

Julier et al. (1995) samples:

$$s = \left[\mathbf{0}, \sqrt{d_r + \kappa}S, -\sqrt{d_r + \kappa}S\right], \tag{4.8}$$

with $\kappa = 0$. The weights are

$$w_{\mathbb{C}} = w_{\mathbb{E}} = \left[\frac{\kappa}{d_r + \kappa}, \left[\frac{1}{2\left(d_r + \kappa\right)}\right]_{1\times 2d_r}\right]. \tag{4.9}$$

Julier and Uhlmann (2004) samples:

$$s = \left[\mathbf{0}, \sqrt{\frac{d_r}{1 - w_0}}S, \sqrt{\frac{d_r}{1 - w_0}}S\right], \tag{4.10}$$

with $w_0 = \frac{1}{3}$. The weights are

$$w_{\mathbb{C}} = w_{\mathbb{E}} = \left[w_0, \left[\frac{1 - w_0}{2d_r}\right]_{1\times 2d_r}\right]. \tag{4.11}$$

Menegaz et al. (2011) samples:

$$s = \left[S \cdot C\left(\left(w_i \cdot \boldsymbol{I}\right)^{\frac{1}{2}}\right)^{-1}, -\sqrt{\frac{1 - w_l}{d_r}}S\left(w_l \cdot \boldsymbol{I}\right)\right], \tag{4.12}$$

with

$$\begin{aligned} w_l &= 0.5 \\ C &= \left(\boldsymbol{I} - \frac{1 - w_l}{d_r}\left[1\right]_{d_r\times d_r}\right)^{\frac{1}{2}} \\ w_i &= \operatorname{diag}\left(C^{-1}w_l\right)\frac{1 - w_l}{d_r}\left[1\right]_{d_r\times L}C^{-1}. \end{aligned} \tag{4.13}$$

The weights are

$$w_{\mathbb{C}} = w_{\mathbb{E}} = \left[w_i', w_l\right]. \tag{4.14}$$

Lerner (2002) samples:

$$
\begin{aligned}
G_1 &= \left[\text{perms}\left(\left[\sqrt{3}, \mathbf{0}_{1\times(d_r-1)}\right]\right)', \text{perms}\left(\left[-\sqrt{3}, \mathbf{0}_{1\times(d_r-1)}\right]\right)'\right] \\
G_2 &= \left[\text{perms}\left(\left[\sqrt{3}, \sqrt{3}, \mathbf{0}_{1\times(d_r-1)}\right]\right)', \text{perms}\left(\left[\sqrt{3}, -\sqrt{3}, \mathbf{0}_{1\times(d_r-1)}\right]\right)',\right. \\
&\quad \left.\text{perms}\left(\left[-\sqrt{3}, \sqrt{3}, \mathbf{0}_{1\times(d_r-1)}\right]\right)', \text{perms}\left(\left[-\sqrt{3}, -\sqrt{3}, \mathbf{0}_{1\times(d_r-1)}\right]\right)'\right] \\
G &= \text{unique}\left([G_1, G_2, \mathbf{0}_{d_r\times 1}]\right)' \\
s &= S \cdot G,
\end{aligned}
\tag{4.15}
$$

with perms() generates all possible permutations of the input values, unique() generates the unique points of the input values and deletes all repetitions. The weights are

$$
w_{\mathbb{C}} = w_{\mathbb{E}} = \left[\left[\frac{4-d_r}{18}\right]_{[1\times 2d_r]}, \left[\frac{1}{36}\right]_{[1\times(2d_r^2-2d_r)]}, \frac{d_r^2-7d_r}{18}+1\right] \tag{4.16}
$$

van der Merwe (2004) samples:

$$
s = \left[\mathbf{0}_{d_r\times 1}, \sqrt{d_r+\lambda}S, -\sqrt{d_r+\lambda}S\right], \tag{4.17}
$$

with

$$
\begin{aligned}
\lambda &= a^2\left(d_r+\kappa\right) - d_r \\
a &= 0.7 \\
\kappa &= 0.
\end{aligned}
\tag{4.18}
$$

The weights are

$$
\begin{aligned}
w_{\mathbb{E}} &= \left[\frac{\lambda}{d_r+\lambda}, \left[\frac{1}{2\left(d_r+\lambda\right)}\right]_{[1\times 2d_r]}\right] \\
w_{\mathbb{C}} &= \left[\frac{\lambda}{d_r+\lambda} + \left(1-a^2+b\right), \left[\frac{1}{2\left(d_r+\lambda\right)}\right]_{1\times 2d_r}\right],
\end{aligned}
\tag{4.19}
$$

with $b = 2$.

Charalampidis and Papavassilopoulos (2011) samples:

$$
\begin{aligned}
G &= \text{unique}\left(\text{perms}\left(\underbrace{\{x_1\}, ..., \{x_i\}}_{d_r}, d_r\right)\right) \\
s &= SG
\end{aligned}
\tag{4.20}
$$

with perms (X, d_r) generates all possible combinations of taking L items from X. Furthermore,

$$\begin{aligned} p_i &= \frac{2i-1}{2c}, i = 1, 2, ..., c \\ y_i &= F^{-1}(p_i) \\ x_i &= \sqrt{\frac{c}{\sum_{i+1}^{c} y_i^2}} y_i. \end{aligned} \tag{4.21}$$

The weights are

$$w_{\mathbb{C}} = w_{\mathbb{E}} = \left[\frac{1}{c^{d_r}}\right]_{\left[1 \times c^{d_r}\right]}. \tag{4.22}$$

4.3.4 CMD methods

In this section, we propose CMD methods for the approximation of statistical moments of nonlinear MEMs. These methods construct a representative set of points by minimizing the modified CMD,

$$\min_{\{m^{(i)}\}_{i=1}^{n}} \text{CMD}\left(m^{(1)}, \ldots, m^{(n)}\right). \tag{4.23}$$

The modified CMD has been introduced by Gilitschenski and Hanebeck (2013) and is based on a local cumulative distribution function (Hanebeck and Klumpp, 2008). It measures the discrepancy between a Dirac mixture distribution and a multi-variate standard normal distribution. By minimizing the CMD, Dirac mixture distribution can be used to approximate a standard normal distribution. To samples with a given covariance matrix D, the Mahalanobis transform is applied to the solution of Equation (4.23) to obtain $b^{(i)}$, $i = 1, \ldots, n$.

To the best of our knowledge, CMD methods are not widely used and have not been applied in MEMs. Yet, CMD methods possess several advantageous properties: (1) CMD methods allow – similar to MC and QMC methods – for any discrete number of samples (Figure 4.2), which enables control of approximation accuracy; and (2) they provide robust approximations at low sample sizes compared to MC methods.

Unlike the SP methods, CMD methods consider the distribution and not particular moments. Accordingly, even for linear models, the mean might be off. To address this problem, we enforced a zero mean by estimating only $n-1$ points and setting $m^{(1)} = -\sum_{i=2}^{n} m^{(i)}$. Hence, the CMD point are:

$$CMD_{points} = \{m^{(1)}, m^{(2)}, \ldots, m^{(n-1)}, -\sum_{i=1}^{n-1} m^{(i)}\} \tag{4.24}$$

For the first step, we generate CMD points by using the same weight for all samples. Furthermore,

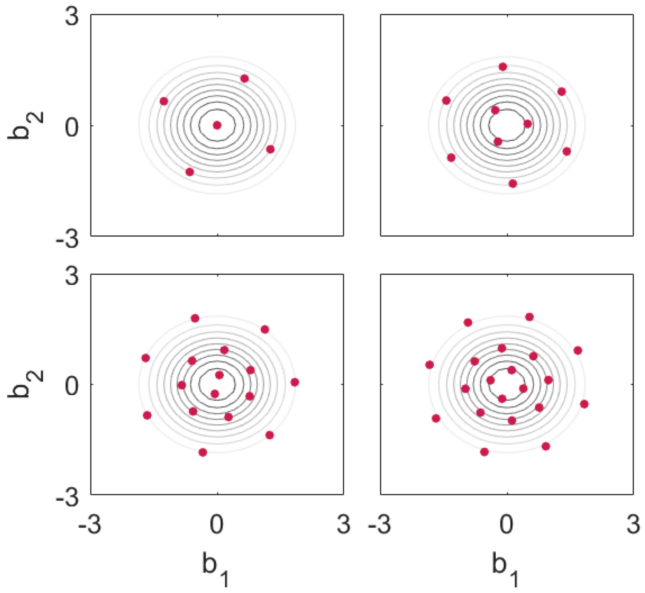

Figure 4.2: **Location of point sets obtained using the CMD method** with the sample size $n = 5$, $n = 9$, $n = 16$, $n = 20$.

we generalized CMD methods by Gilitschenski and Hanebeck (2013) to allow for non-uniform weights, as this might improve the approximation accuracy. The CMD distance and gradients with respect to CMD_{points} is denoted by D_n and $G_n = \{G^{(1)}, G^{(2)}, G^{(n)}\}$, respectively. The equations to compute them can be found in Gilitschenski and Hanebeck (2013).

Therefore, for the generalized optimization routine of $n-1$ points, we can calculate the distance and gradients as following:

$$D_{n-1} = D_n \tag{4.25}$$

$$G_{n-1} = \{G^{(1)} - G^{(n)}, G^{(2)} - G^{(n)}, \ldots, G^{(n-1)} - G^{(n)}\}. \tag{4.26}$$

4.4 Methods assessment and application to test problem

4.4.1 QMC methods outperforms MC method

In this section, we consider Halton and Sobol sequences. As these sequences are defined on the unit cube, we transformed them into representative sequences for the standard multi-variate normal distribution using the corresponding cumulative distribution function. The application of the Mahalanobis transform – using the matrix square root of D – to these sequences yields samples for the random effects $b^{(i)}$ with the desired covariance. The Halton and Sobol sequences perform similarly (Figure 4.3), and both have better accuracy than the MC method, especially when sample numbers get larger. Therefore, in the following investigation, we only show the results of Halton sequences for simplicity.

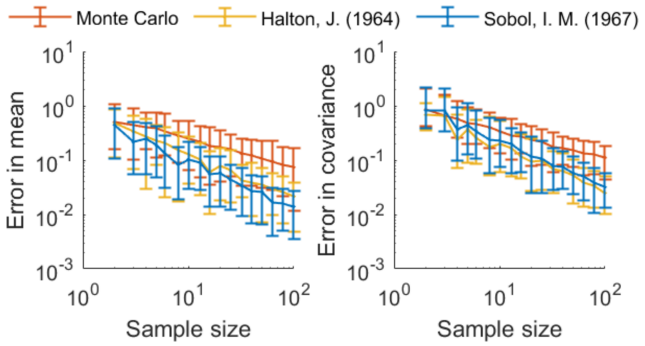

Figure 4.3: **Absolute difference between true moments (mean and covariance) and sample moments.** The results of MC and QMC methods for the standard normal distribution are depicted. The interval between the 5th and the 95th percentile of the errors across 1000 realizations is depicted.

4.4.2 Computation time of CMD samples scales linearly with sample size and dimension

As the calculation of random effects in the CMD method requires the solution of an optimization problem, the scalability of the approach is unclear. To assess it, we evaluate the average computation time per local optimization (Villaverde et al., 2019), which is the overall computation time divided by the number of local optimization runs achieving the best objective function value. We computed the corresponding computation time for different sample sizes in the same dimension, as well as multiple dimensions with a fixed number of sample sizes. We found that the computation time scales roughly linearly with the number of points n (Figure 4.4(a)), and also linearly with the dimension of the random effect vector (Figure 4.4(b)). Accordingly, the computation time required for optimization increases substantially with both problem dimensions.

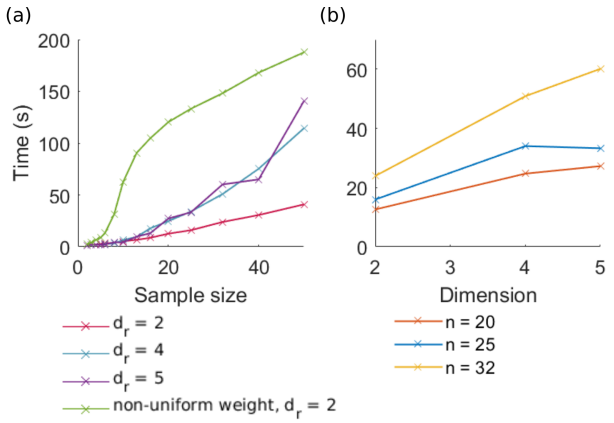

Figure 4.4: **Scalability of the CMD method.** Average computation time per converged start for (a) different number of points and (b) different dimensions is depicted.

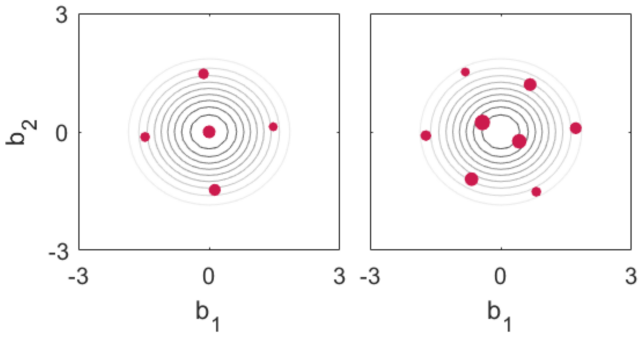

Figure 4.5: **Location of weighted CMD points** with sample size $n = 5$ and $n = 8$. The values of the weights are encoded by the marker sizes.

4.4.3 Multi-start local optimization outperforms partical-swarm method

As an alternative to multi-start local optimization, we also considered the particle-swarm optimization algorithm `PSwarm` (Vaz and Vicente, 2009) as a global optimization method. We computed CMD sample locations using both methods with a smaller (smaller than 10) and a larger sample sizes (larger than 50). We found that for low numbers of CMD samples, the optimized objective function values of multi-start local optimization and PSwarm were comparable. For large particle numbers, PSwarm did not achieve the same cost function values as multi-start local optimization (for 50 CMD points, multi-start: 0.00088, PSwarm: 0.0051).

4.4.4 Non-uniform weighting marginally improves CMD sampling

Non-uniform weighting of the points was considered to reduce the number of points by potentially increasing of approximation accuracy. Therefore, we optimized points and weights together, using analytically derived expressions for the gradients. We found that point locations change (see Figure 4.2 vs. 4.5) and that points close to the center of the distributions tend to have higher weights (Figure 4.5). The value of the modified cost function improved marginally by introducing the weights (for $n = 8$: uniform CMD $= 0.0261$ and non-uniform CMD $= 0.0236$). However, this improvement comes at a substantially increased computation time (Figure 4.4(a)).

4.4.5 Test and application examples

In this section, we considered Dirac mixture distributions generated using different methods to evaluate the statistical moments of the response vector, in particular its mean $\mathbb{E}[y]$ and covariance $\mathbb{C}[y]$. For MEMs, these methods use a set of random effects $b^{(i)}$, $i = 1, \ldots, n$, for which the corresponding mixed effects $\phi^{(i)}$ and responses $y^{(i)}$ are evaluated. When we consider only the moments of the responses, the estimators of the mean and covariance are

$$\mu = \sum_{i=1}^{n} w_{\mathbb{E}}^{(i)} f\left(\phi^{(i)}\right) \tag{4.27}$$

$$\Sigma = \sum_{i=1}^{n} w_{\mathbb{C}}^{(i)} \left(f\left(\phi^{(i)}\right) - \mathbb{E}\right)\left(f\left(\phi^{(i)}\right) - \mathbb{E}\right)^T, \tag{4.28}$$

with weights $w_{\mathbb{E}}^{(i)}$ and $w_{\mathbb{C}}^{(i)}$. In the following, we outline the considered methods.

Remark: As measurement noise has no influence on the mean ($\mathbb{E}[y] = \mathbb{E}[h]$) and an additive influence on the covariance ($\mathbb{C}[y] = \mathbb{C}[h] + \Lambda$, when the weights sum up to 1), it can be handled analytically. For the purpose of this section, we set it to zero ($\Lambda = 0$) to avoid that approximation errors are hidden.

In order to test the accuracy of the above-mentioned different methods regarding models with different complexity, we provide an assessment of the methods using two test problems:

- Quadratic function (M1): $y_i = \varphi_i^2$ for $i = 1, 2$
  with $\varphi = \beta + b$, $\beta = [0, 0]^T$ and $b \sim \mathcal{N}(0, I_2)$

- Hill-type function (M2): $y_i = \frac{\phi_i^4}{1+\phi_i^4}$ for $i = 1, 2$
  with $\varphi = \beta + b$, $\beta = [0, 0]^T$ and $b \sim \mathcal{N}(0, I_2)$

and two application problems. The first application problem is a model of a conversion reaction model (M3):

- Conversion reaction model (M3): $y = x_2$
  $\dot{x}_1 = -e^{\varphi_1} x_1 + e^{\varphi_2} x_2, \qquad x_1(0) = 2$
  $\dot{x}_2 = e^{\varphi_1} x_1 - e^{\varphi_2} x_2, \qquad x_2(0) = 0$
  with $\varphi = \beta + b, \beta = [0, 0.3]^T$ and $b \sim \mathcal{N}(0, I_2)$

Conversion reactions are common motives in biological processes. The second application example is a model for JAK/STAT signaling (M4) based on Swameye et al. (2003). This model is widely used for method evaluation (Raue et al., 2009; Hass et al., 2016; Fröhlich et al., 2016) and possesses 9 state variables, 16 parameters, and 3 observations. We assume that the 5 initial concentrations out of the 16 model parameters are random effects, meaning that they differ between single cells.

To compute the ground truth of the statistical moments for the test and application problems, we used MC sampling with 100,000 points. Since outputs for the application examples are multi-dimensional, average values of absolute differences to true mean and true variance were computed.

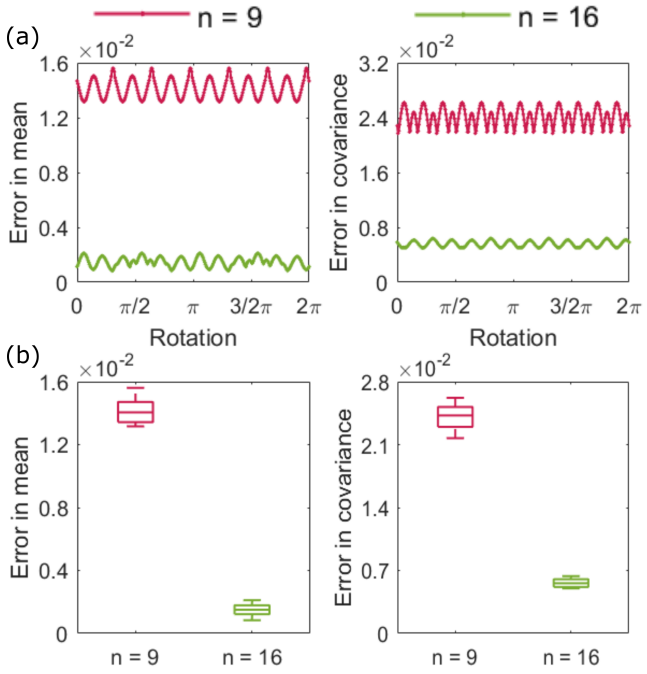

Figure 4.6: **Influence of rotation angle for problem M2.** (a) Absolute difference between true mean and variance and the approximations obtained using the CMD method with rotated point sets. (b) Box plot showing the variability induced by the rotation.

4.4.6 Effects of CMD points swap and rotation is marginal

An analysis of the optimization problem (4.23) reveals that the optimal point is not unique, meaning that samples with different locations may result in the same objective function values. Optimized point properties can be swapped and the point sets can be rotated around the origin without changing the objective function values. While points swaps can not have any influence on the approximation of the moments, the rotation angle can have an influence when f is nonlinear. However, our analysis shows that the standard deviation of the approximation error was smaller than its mean (Figure 4.6(a), (b)), suggesting that the effect is not too important. As the point sets are structured, the approximation error shows a periodic behavior. The period length depends on the number of points (Figure 4.6(a)).

4.4.7 Approximation properties of Dirac mixture distributions

To assess the approximation accuracy of the MC, QMC, SP, and CMD (including uniform and non-uniform weighting) methods, we applied them to the test and application problems as mentioned in Methods. It is to be noted that, because of the higher dimension of M4 and the double scaled optimization problem of computing non-uniform weights, we only applied the CMD method with uniform weighting to M4. For the MC, QMC, and CMD methods, as well as the SP method by Charalampidis and Papavassilopoulos (2011), multiple point numbers were considered to study the convergence rates. For all problems, we determined the absolute differences of mean and variance compared to the ground truth, which is computed using the MC method with 100,000 samples. As each point requires an evaluation of the output, the computation time

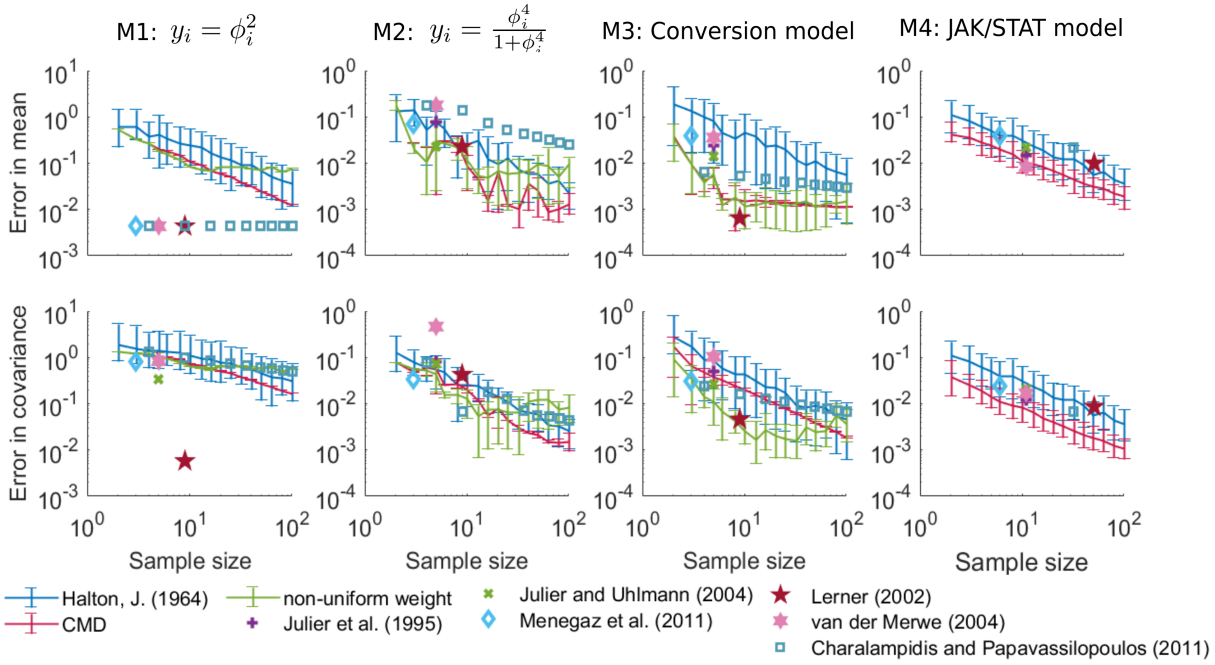

Figure 4.7: **Assessment of the approximation accuracy for test and application problems.** The absolute difference between approximated moments and ground truth is depicted: (upper) mean and (lower) covariance.

increases linearly with the number of points. Hence, a high approximation accuracy with a low point number is desirable.

For M1, the mean computed by the SP methods is exact (by construction), and the CMD method outperforms the Halton QMD method. For M2-M4, the CMD method with uniform weights achieves better accuracy than the Halton QMC method and SP methods (Figure 4.7). In addition, compared to the Halton method, CMD methods show (1) similar or better convergence rates and (2) smaller confidence intervals. The latter implies an improved stability of the approximation and independence from the rotation angle. Furthermore, CMD methods with uniform and non-uniform weighting achieve similar accuracy. In the comparison of M2 and M3, we see an increase in approximation error by using the CMD method with non-uniform weighting. This will need further investigation.

4.5 Discucssion

4.5.1 Summary and conclusion

In this chapter, we perform an evaluation using different Dirac mixture distributions to approximate normal distributions, namely MC, QMC, SP, and CMD methods. For CMD methods – which are not commonly applied for MEMs – we perform a first scalability study, investigated the effects of point swap and rotation, and assess the influence of non-uniform weights.

It is shown that our newly introduced CMD method scales linearly with both sample size and

parameter dimension. However, in the case of normal distribution, the sample locations only need to be optimized once for a standard normal distribution with the needed sample size and dimension and stored. Mahalanobis transformation can then be employed to transform it according to the corresponding covariance matrix. Therefore, no repetitive optimization is necessary even for the simulation of high-dimensional models.

Furthermore, we applied the above-mentioned methods to two tests and two application examples. Results show that the CMD method outperforms SP and QMC methods in three out of four test cases regarding both mean and covariance. compared to the SP methods, CMD methods are more flexible since a continuous refinement is possible, and possess a better accuracy when the model is highly nonlinear. Compared to MC and QMC methods, CMD methods have better convergence and lower variance, especially for more complicated models, e.g. ODE models in systems biology. We also show that using non-uniform weight in the CMD method does not improve the approximation accuracy, while substantially increasing the computation time.

Overall, we introduced CMD methods to discretely approximate normal distributions. CMD methods outperform MC, QMC, and SP methods regarding both accuracy and flexibility, especially for high nonlinear models. Our results demonstrate the potential usage of CMD methods for MEM of heterogeneous cell and patient populations in systems biology and systems medicine.

4.5.2 Outlook

In this chapter, we compared the approximation accuracy of CMD methods with different sample sizes. Results show that with the increasing of sample numbers, both accuracy and computation time are also increased. Therefore, it is important to find a balance between necessary accuracy and an affordable computation time.

We used a multi-start local optimization method to compute the locations of CMD samples. Furthermore, this method is compared with the partical swarm method and results show that it achieves better objective function values when the sample size is large. However, other global optimization methods might be able to improve the convergence performance and the computation efficiency, for instance, the simulated annealing and ant colony algorithm. This requires further investigation.

We applied our newly introduced CMD methods in normal distributions and showed a substantial improvement compared to MC, QMC, and SP methods regarding both accuracy and flexibility. Therefore, we would anticipate that CMD methods can also be applied to other commonly used assumptions of distributions in MEMs, for instance, log-normal distribution and Laplacian distribution.

Chapter 5

Benchmarking covariance matrix parametrization methods

Often, the covariance structure of random effects is unknown in MEMs and has to be inferred. Sometimes it is assumed that only variance values are non-zero for simplicity, meaning that there is no correlation between random effects. However, in real biological experiments, some correlations are clearly known and the effects have to be considered. Parametrizing the corresponding covariance matrices is non-trivial. There exist different approaches, e.g. via the matrix logarithm, the Cholesky decomposition, the spherical parametrization, and the Givens parametrization. The Givens parametrization is based on a spectral decomposition and thus allows to directly bound eigenvalues. However, it parametrizes rotation matrices as a product of elementary rotations, which depends on the choice of ordering of the rotations.

In this chapter, we introduce a novel approach that is first developed by Stapor (2021) and uses ideas from Lie theory to parametrize the rotation matrices. We compare the performance of the different parametrizations in parameter estimation on five benchmark models and one more complex application example with different methods of parameter initialization. Our newly introduced Lie-algebraic method performs similarly well to the Givens parametrization on the benchmark models and outperforms all established approaches on the application example, regarding both accuracy and computational robustness.

5.1 Introduction

For simplicity, the covariance matrix of random effects D is in many applications assumed to be diagonal (Fröhlich et al., 2018b). However, correlations may be essential, as they encode information about the modeled biological system such as co-regulations of proteins or coupled synthesis and degradation rates (Llamosi et al., 2016). Hence, it is important to not restrict the model to diagonal covariance matrices but to also take block-diagonal or more general covariance structures into account.

MEMs usually depend on unknown parameters that need to be inferred by comparing model outputs to experimental data and maximizing a likelihood function. Besides unknown fixed effect

and noise level parameters, often also the covariance structure is unknown. However, parametrizing the corresponding covariance matrix is non-trivial. Desirable properties of a parametrization – especially in the context of an application to ordinary differential equation (ODE) MEMs – are:

1. Favorable convergence properties during parameter estimation.

2. Meaningful parameter bounds for the parameter estimation problem.

3. Inferring structured covariance matrices without constraints, in order to ensure positive-semi definiteness.

In the literature, four different parametrization approaches have been proposed (Pinheiro and Bates, 1996): The matrix logarithm (Pinheiro and Bates, 1996), the (upper) Cholesky decomposition (Lindstrom and Bates, 1990), the spherical parametrization (Bates and Watts, 1988), and the Givens parametrization (Thisted, 2017).

Although the Cholesky and matrix logarithm parametrization methods are conceptionally and computationally simple, they do not allow to constrain the variances of D by box-constrain directly (Pinheiro, 1994). This make the definition of parameter boundaries complicate. Besides, the spherical parametrization method allows to constrain the eigenvalues of D by directly parametrizing the variance and defining parameters for angles on sphere. However, if D is high-dimensional, the eigenvalues can become large, which may lead to unfavorable initial values of the ODE at an early stage in the optimization. Finally, the Givens parametrization method constrainss the eigenvalues directly. Therefore, it is favorable for ODE MEMs as the probability of integration failure is reduced at the initial point. However, the parametrization of the orthogonal matrices is rather involved. This may explain the relative slow convergence observed for this method.

In this chapter, we consider an alternative novel parametrization approach (Stapor, 2021), which relies on Lie theoretic ideas, and perform an in-depth evaluation. We discuss it in the context of ODE MEMs, comparing the convergence performance of the newly proposed and the established methods on five benchmark covariance matrices with different complexity. Furthermore, we compared their robustness and prediction accuracy on a model of JAK/STAT signaling. Results show that the Givens and Lie algebraic parametrizations have overall the best convergence performance. Moreover, the Lie algebraic method has better performance regarding prediction accuracy.

5.2 Background

In this chapter, we compare the above-mentioned five covariance matrix parametrization approaches. Among these approaches, the Cholesky decomposition and spherical parametrization approaches are based on the Cholesky factorization, while the matrix logarithm and Givens parametrization approaches are based on the spectral decomposition.

Cholesky decomposition

As the covariance matrix D can be written as $D = L^T L$ for lower triangular L, we can then parametrize L entry-wise. Furthermore, we use exponential transformation of the values of L to ensure the uniqueness of L.

Matrix logarithm

Since D is positive definite, it has positive eigenvalues λ. Letting U denote the orthogonal matrix of orthonormal eigenvectors of D and $\Lambda = \text{diag}(\lambda)$, then

$$D = U\Lambda U^T. \tag{5.1}$$

The matrix logarithm of D

$$\log(D) = U\log(\Lambda)U^T. \tag{5.2}$$

is a symmetric matrix. Therefore, the covariance matrix can be parametrized entry-wise by the upper triangular elements.

Spherical parametrization

The spherical parametrization approach parametrizes the vectors of the upper Cholesky decomposition by spherical coordinates. Therefore, it allows for box constraints on the variance values.

Let L_i denote the ith column of L as in the Cholesky decomposition of D, and l_i denote the spherical coordinates of the non-zero elements of L_i. Then

$$\begin{aligned}
[L_i]_1 &= [l_i]_1 \cos([l_i]_2) \\
[L_i]_2 &= [l_i]_1 \sin([l_i]_2)\cos([l_i]_3) \\
&\dots \\
[L_i]_{i-1} &= [l_i]_1 \sin([l_i]_2)\dots\cos([l_i]_i) \\
[L_i]_i &= [l_i]_1 \sin([l_i]_2)\dots\sin([l_i]_i)
\end{aligned} \tag{5.3}$$

Givens parametrization

Givens parametrization approach is based on the fact that the eigenvector matrix U can be represented by $q(q-1)/2$ angles, where q is the dimension of covariance matrix D.

$$U = G_1 G_2 \dots G_{q(q-1/2)}, \tag{5.4}$$

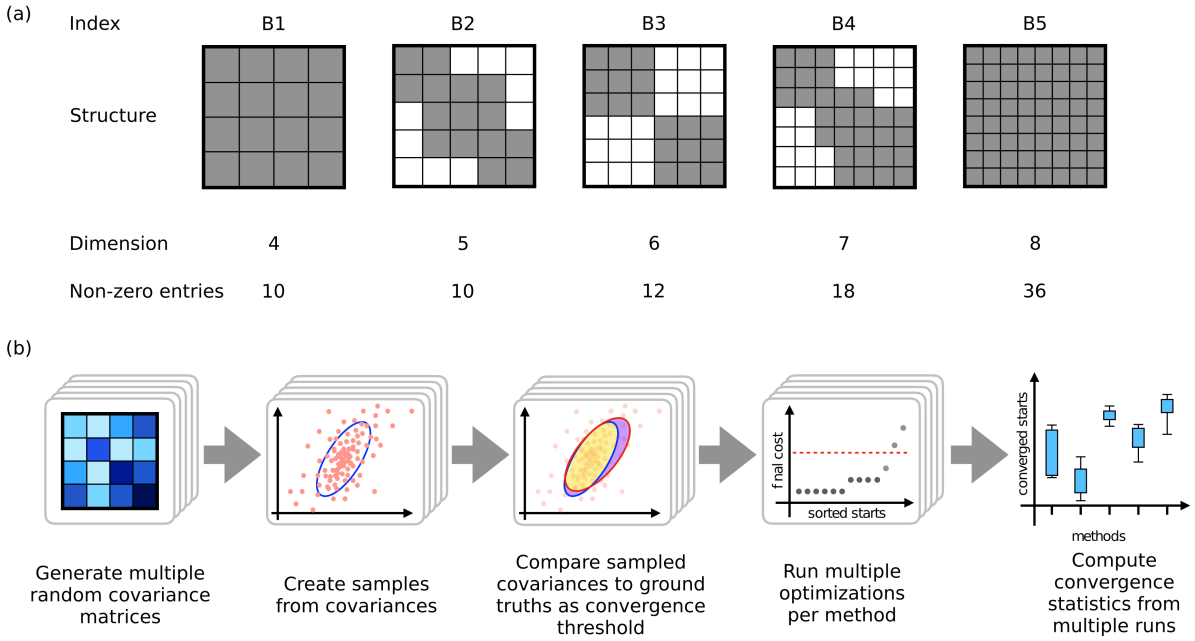

Figure 5.1: **Concept of benchmark study for covariance parametrizations.** (a) Correlation patterns of benchmark models. Grey fields denote non-zero entries. (b) Work flow of benchmark study.

where G_i are Givens rotation matrices. For details of this method, we refer to Thisted (2017).

5.3 Lie-algebraic method and benchmark models

5.3.1 Lie-algebraic method

As the Givens parametrization method depends on the choice of an ordering of these elementary rotations. This may cause dependencies between the rotation angles. To account for this problem, we propose a novel approach that relies on Lie theoretic ideas.

Similar to the Givens parametrization, the proposed Lie-algebraic parametrization δ decomposites a covariance matrix D by a rotation matrix and its eigenvalue. This rotation matrix is then parametrized via an antisymmetric matrix, which is then exponentiated to generate the actual rotation matrix. The parametrization of the antisymmetric matrix happens via coefficients $\alpha_1, \ldots, \alpha_{n_A}$, which are multiplied with a basis $A_1, \ldots, A_{n_A}$ of the antisymmetric matrices (with $n_A = n(n-1)/2$). Therefore, we have:

$$\delta_n \colon (\lambda_1, \ldots, \lambda_n, \alpha_1, \ldots, \alpha_{n_A}) \longmapsto \exp\left(\sum_{m=1}^{n_A} \alpha_m A_m\right) \exp\left(\operatorname{diag}\left(\lambda_1, \ldots, \lambda_n\right)\right) \exp\left(\sum_{m=1}^{n_A} -\alpha_m A_m\right) \tag{5.5}$$

5.3.2 Benchmark models

In order to compare the parametrization methods mentioned in Section 5.2 regarding their convergence and accuracy performance, a benchmark setup was created based on 5 test cases (B1 - B5, Figure 5.1 (a)) of increasing complexity, with $d = 4, \ldots, 8$ and $d \times d$ describing the shape of the matrix as used in Stapor (2021). Each benchmark study was repeated five times. For benchmark $B1, \ldots, B5$ and repetition $1, \ldots, 5$, we performed multi-start local optimization as follows: A covariance matrix D_B^R was randomly generated based on a previously defined correlation pattern (Figure 5.1 (a)). This covariance matrix was considered to be the ground truth, from which a population with $10000 \cdot d$ individuals were sampled. From this population, the empirical covariance $\hat{D}_B^R$ of the sample was computed to mimic a certain level of random measurement noise (Figure 5.1 (b)). Based on $\hat{D}_B^R$, the following objective function was defined:

$$J_{\hat{D}_B^R}(\delta) = \frac{1}{2}\left(\left\|\hat{D}_B^R - d_B^{\text{sim}}(\delta)\right\|_F^2 + \left\|\text{diag}\left(\hat{D}_B^R - D_B^{\text{sim}}(\delta)\right)\right\|_F^2\right), \tag{5.6}$$

where $\|\cdot\|_F$ denotes the Frobenius norm, diag$(\cdot)$ reduces a matrix to its diagonal, and D_B^{sim} is the simulated covariance matrix. Therefore, the objective function is an adaptation of the Frobenius norm. As an extension of the initial analysis by Stapor (2021), we initialized the covariance matrix values in two different ways: 1) uniformly sampled in the constraints; 2) uniform sample of the variance values in the constraints, and set all the other entries as zero. An optimization run was considered to be converged if its final objective function value reached a threshold T_B^R defined by the ground truth:

$$J_{\hat{D}_B^R}(\delta) < T_B^R = \frac{1}{2}\left(\left\|\hat{D}_B^R - D_B^R\right\|_F^2 + \left\|\text{diag}\left(\hat{D}_B^R - D_B^R\right)\right\|_F^2\right) \tag{5.7}$$

5.3.3 Lie algebraic method performs comparably with Givens parametrization regarding convergence

Uniform initialization of all matrix parameters

Regarding the final objective function values for all repetitions and benchmarks, the overall convergence for each method was similar across repetitions (Fig 5.2 (a)). Furthermore, the quality of optimization decreases when the dimension of the benchmark problem increases. While for benchmark B1 with the lowest dimension, at least some optimization runs reached low objective function values in each repetition for all methods, convergence was substantially worse for benchmark B5 with the highest dimension.

When analyzing the total number of converged starts and the variability across repetitions, the matrix-logarithm parametrization method performed worst for all benchmarks (Figure 5.2 (b)). The Cholesky and the spherical Cholesky parametrization still yielded favorable results for the smallest benchmark B1, but performed considerably worse for the benchmarks B2 to B5. The Givens parametrization approach performed best overall, considering the benchmark models

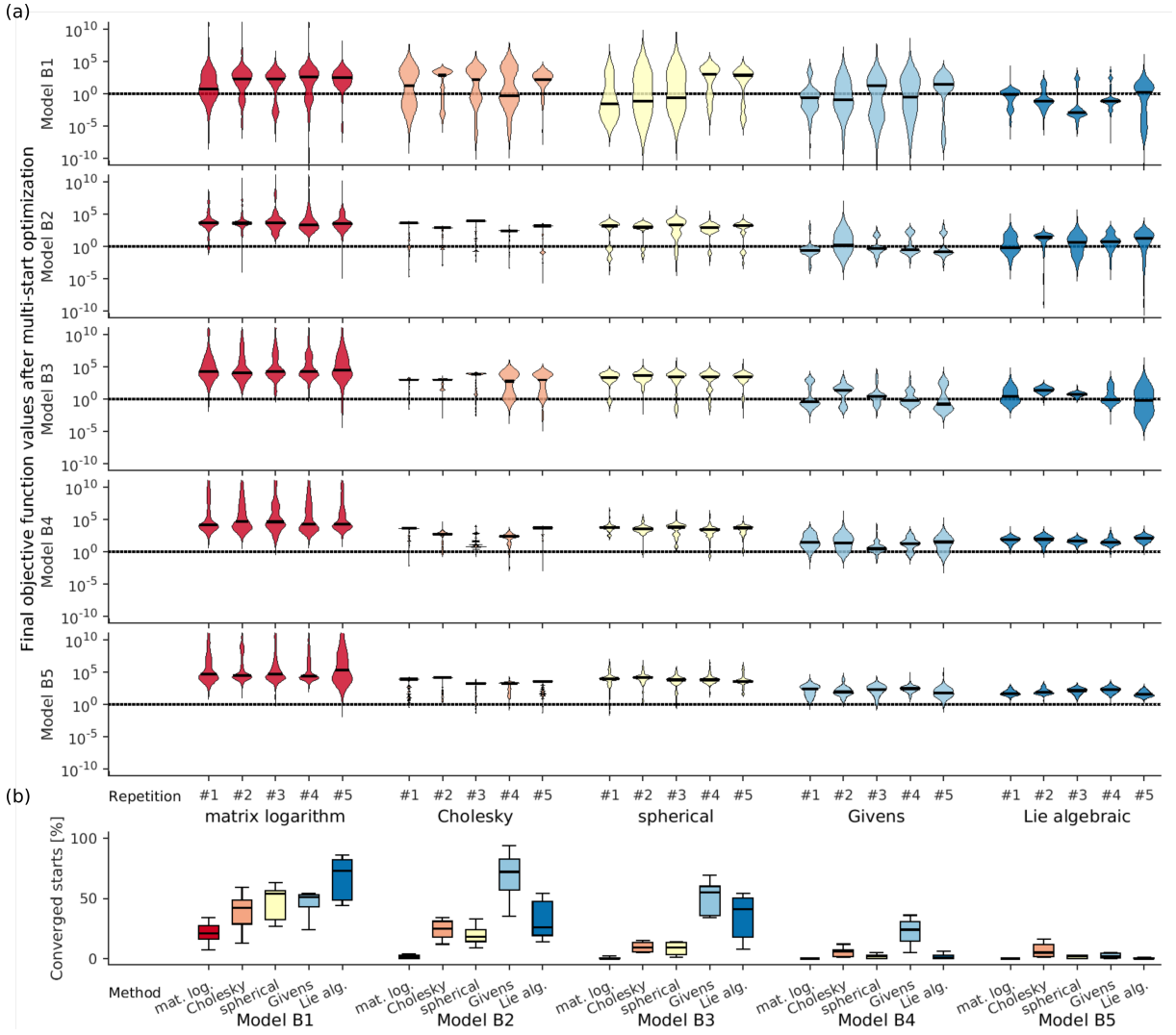

Figure 5.2: **Comparison of convergence properties of different covariance parametrization approaches with uniformly sampled initial values.** (a) Violin plots over final objective function values for all benchmarks and all repetitions across all methods. The solid lines in the violin plots indicate the medians. (b) Converged starts per benchmark model and method. Boxes show 25 and 75 percentiles computed over repetitions, whiskers show minima and maxima, medians are depicted as solid black lines.

from B2 to B4. While for B1, the Lie algebraic parametrization performed clearly best. For benchmark B5, all methods get extremely few converged starts.

Uniform initialization of only variance parameters

In order to test the effects of different ways of parameter initialization, we performed the same benchmark tests with uniformly sampled variance values, while other covariance matrix entries are set to zero.

Compared to the former scenario, where all the matrix entries are uniformly sampled, the new proposed initialization yields more favorable converged runs (Fig 5.3 (a)). For benchmark B1,

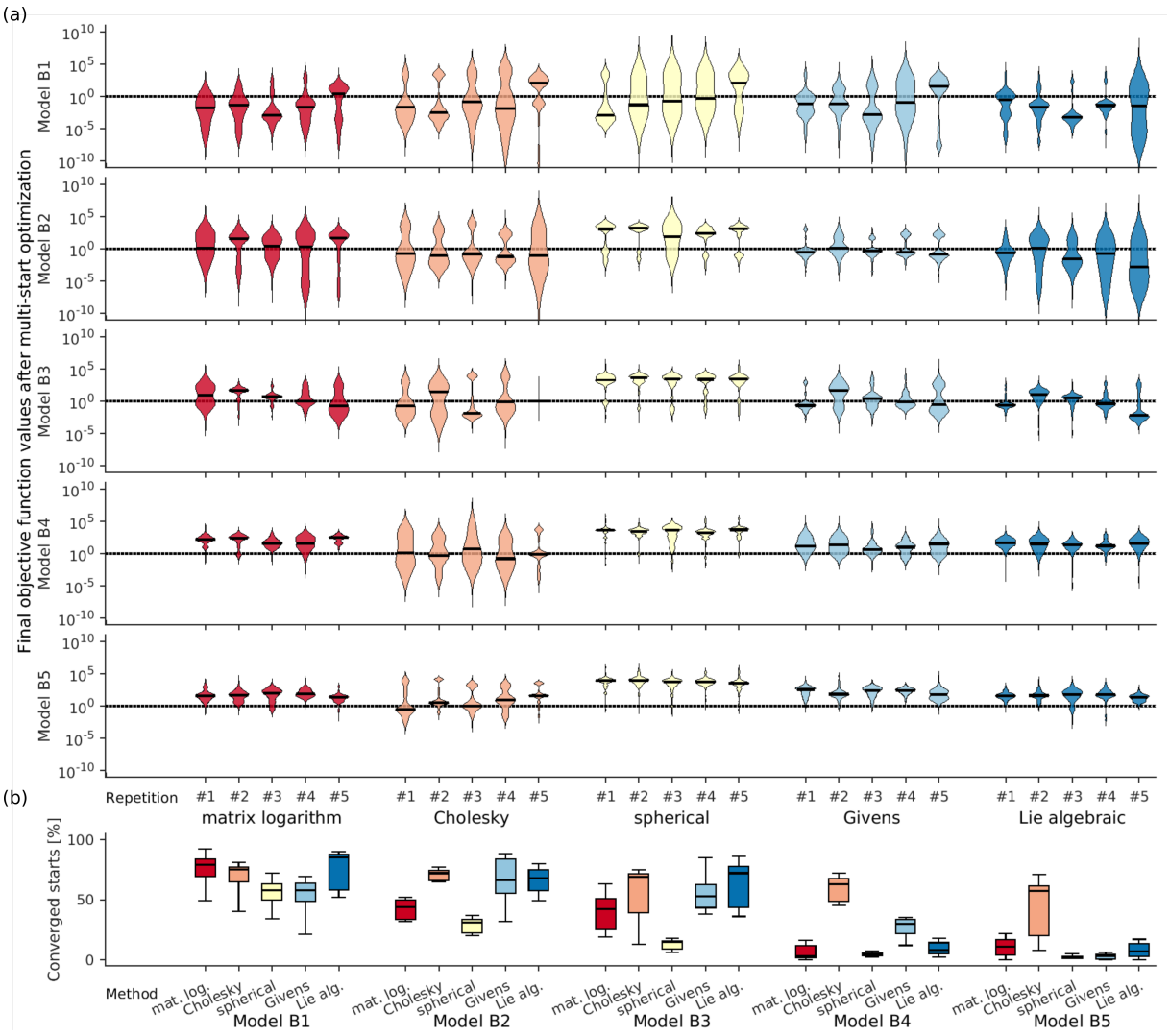

Figure 5.3: **Comparison of convergence properties of different covariance parametrization approaches with uniformly sampled variance values.** (a) Violin plots over final objective function values for all benchmarks and all repetitions across all methods. The solid lines in the violin plots indicate the medians. (b) Converged starts per benchmark model and method. Boxes show 25 and 75 percentiles computed over repetitions, whiskers show minima and maxima, medians are depicted as solid black lines.

the medians of most of the repetitions for all five methods are below the convergence threshold. But similarly as in the former scenario, when the complexity of benchmark models increases, the convergence gets more unlikely. However, with this initialization, even for the most complex benchmark B5, a small part of the local optimization runs still converge for all approaches.

When analyzing the total converged starts across the five approaches, we find that the Cholesky method outperforms the other methods in most of the cases, but in some cases with higher variance, e.g., for benchmark B3 and B5 (Fig 5.3 (b)). Except for benchmark B1, the matrix-logarithm and spherical approach performs the worst. For benchmark B1, the Lie algebraic method clearly outperforms the Givens parametrization approach, while for the other benchmarks, both approaches get a similar number of converged starts.

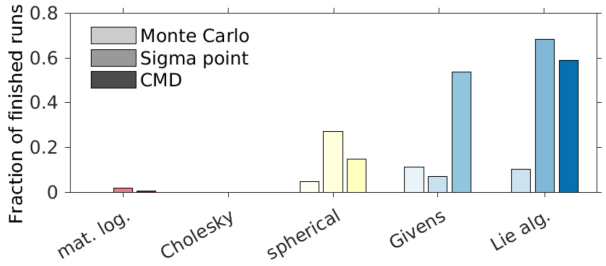

Figure 5.4: **Fraction of runs finished without error using MC, SP sampling methods, and CMD method with 30 samples.** The comparison was done for all five parametrization methods used in this study.

Overall, the Cholesky and Lie algebraic approach outperform the other parametrization methods: When comparing the median number of converged starts across repetitions, the Cholesky approach was the best method for three of the five benchmarks, while the Lie algebraic method performs best in the two models with the higher complexity.

5.4 Application to test problem

As an application example, we used a model for JAK/STAT signaling based on Swameye et al. (2003). For the definition of model parameters and number of states we refer to Section 4.4.5. All the parametrization approaches mentioned above are used for parameter estimation.

As artificial data, we generated three different types of data, the time-lapse (TL) data, population (PA) data and snapshot (SH) data. For the sake of computation time, the TL data are generated for five individual cells, which consist of the observables of each individual at time points 0, 2, ..., 50. The mean (PA) and variance (SH) of the observable are generated at the same time points for 1,000 individuals. The three types of data are fitted separately, meaning that in each multi-start local optimization problem, only one dataset is used. The covariance matrix is generated in the same way as described in section 5.3 for the benchmarks with a dimension of 5×5.

Use CMD method to avoid integration error in application example

To compute the simulated mean and variance and at the same time insure the accuracy, we first used MC sampling method with 10,000 samples. The optimization was performed using 600 start points. We found that most of the optimization runs for all five methods crashed because of integration errors (Fig 5.4 bars with lightest shade). This is because even one unreasonable sample in all the 10,000 samples of parameters could lead to unexpected crashing. Therefore, this problem can be overcome by using a smaller number of samples, but this has to be done while keeping the approximate accuracy.

The SP method is an alternative, but, as shown in Chapter 4, the SP method has a restriction of accuracy when it comes to complex ODE models such as the JAK/STAT model, because of its fixed number of samples. Therefore, we also used the CMD method proposed in Chapter 4

with a sample number of 30 based on Equation (4.24).

Results show that with all three sampling methods, all optimization runs still crashed for the Cholesky decomposition approach (Fig 5.4). The CMD method outperforms both MC and SP methods for Givens paramtrization approach. For the other parametrization approaches, the CMD method performs similarly as the SP method. Take the Lie algebraic method as an example: We found that almost 60% of the optimization runs can be finished by using the CMD method, which is much better than using MC sampling method with a successful rate at about 10%. Therefore, the optimization results computed using the CMD method with 30 samples were used to compare the performance of five approaches.

The optimization results of mean and variance show that the Lie algebraic method has the highest success rate, while the Givens parametrization method is only 5.3% less (Fig 5.4). None of the 600 starts using the Cholesky method finished, while only 4 using the matrix-logarithm method finished.

5.4.1 Uniform initialization of all matrix parameters

We fit the artificial data with three different data types separately using the five covariance matrix parametrization approaches. In this section, uniform sampling is used to generate the initial values of the covariance matrix.

As optimized parameters, we expect that all 30 samples generated using the parameters with the CMD method will not get any integration error. It is to be noted that only for the Lie algebraic approach, the MC sampling method with 10,000 samples can also be used for the fitting of mean and variance data, meaning that the estimated covariance matrix is more reasonable than the ones estimated with other approaches.

Taking the first out of all three measurements as an example (Fig 5.5 (a)), we see that only the Lie algebraic approach is able to fit the TL data (first column) with a reasonably accurate mean and variance. The Givens and Lie algebraic methods are able to fit both the mean (second column) and variance (third column), while the matrix logarithm approach can only fit the mean of the cell population.

The predicted error of the covariance matrix is plotted as a heat map in Fig 5.5 (b). The covariance matrix error of the TL data estimated by the Lie algebraic approach is on average two orders of magnitude smaller than the one estimated by the Givens approach. When using the mean data for parameter estimation, the errors computed by Givens and Lie algebraic method are similar with the value of 0.56 and 0.32. When using the variance data for parameter estimation, the covariance matrix error computed by the Lie algebraic method is the largest, while the predicted trajectory agrees the best with the artificial data, meaning that even a different covariance matrix of the random effects may give us the same prediction of variance measurements.

In conclusion, the Lie algebraic approach outperforms all the other parametrization approaches regarding data fitting. When TL data are used for parameter estimation, the Lie algebraic approach also yields the best accuracy in recovering the original covariance matrix, while it performs similarly as Givens approach and better than all others when the mean of the cell

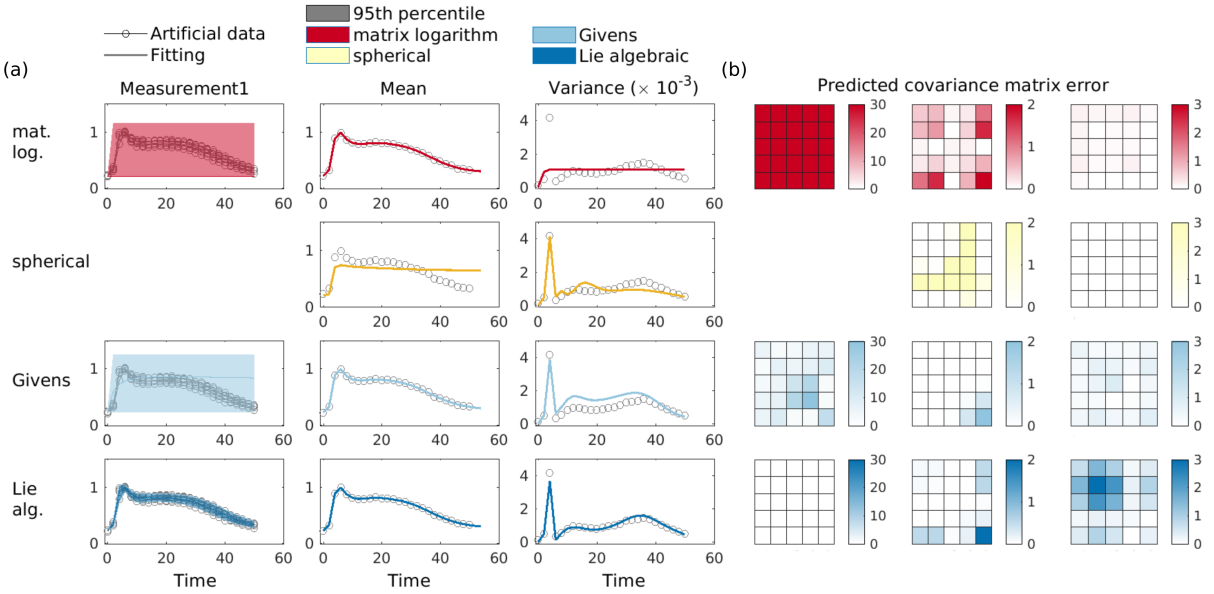

Figure 5.5: **Comparison of five parametrization approaches, with all initial values uniformly sampled, fitting to three different data types separately.** (a) Fitting of three different data types. All the runs for Cholesky decomposition and the spherical approach to fit TL data get integral errors. Therefore, no fitting is plotted. (b) Error of estimated covariance matrices compared with the ground truth.

population is used. Parameters estimated using the Lie algebraic method are less accurate than the ones estimated using all the other approaches when the variance of the population is used, this is arguably because of overfitting.

5.4.2 Uniform initialization of only variance parameters

To investigate different initialization methods, we also performed the parameter estimation using uniform sampling to generate the variance values, while the other entries of the covariance matrix are set to be zero.

Results in Figure 5.6 (a) shows that none of the parametrization approaches can fit the TL data (first column) with a reasonable mean and variance as at least some did in Section 5.4.1. Unlike for the uniform initialization, both Lie algebraic and Givens approaches are not able to fit the variance data (third column) well, while all approaches can fit the mean data (second column) reasonably. The heat maps of matrix errors show that the Lie algebraic approach still get the best accuracy when TL data are used for optimization (Fig 5.6 (b) first column), while Givens and Lie algebraic approaches get similar accuracy when mean (second column) is used with the average error value of 0.64 and 0.57. Similar to the uniform initialization, the Givens approach yields better accuracy when variance data (third column) are used.

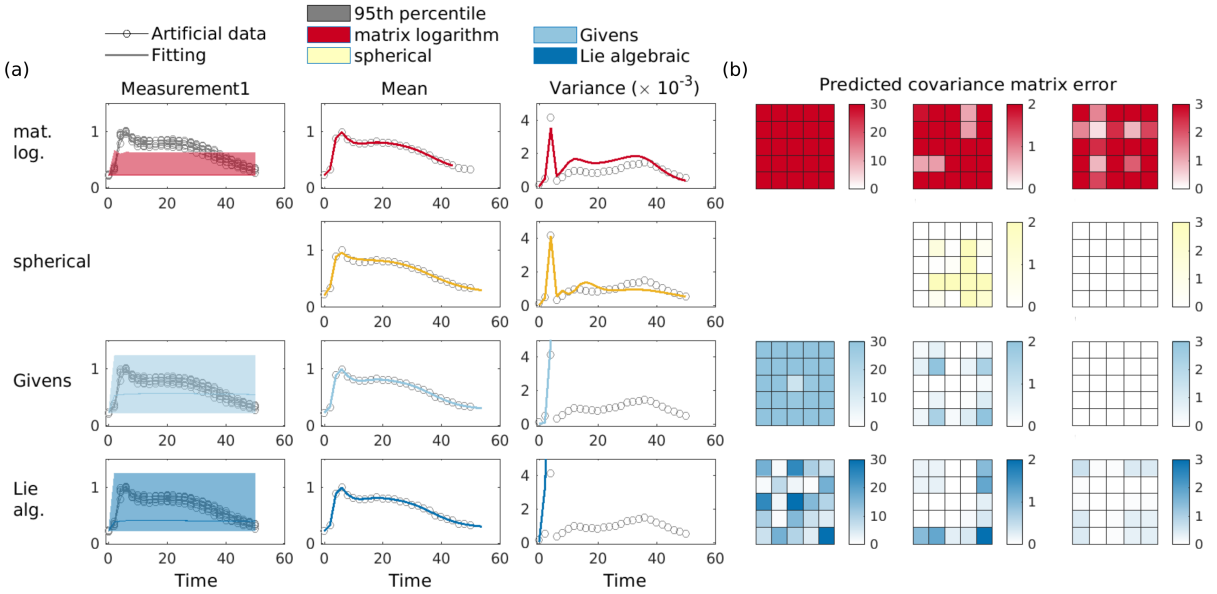

Figure 5.6: **Comparison of five parametrization approaches, with only variance values uniformly sampled, fitting to three different data types separately.** In this case, all the other entries of the initial covariance matrix are set to zero. (a) Fitting of three different data types. All runs for the Cholesky method and spherical method of fitting TL data get integral errors. Therefore, no fitting is plotted. (b) Error of estimated covariance matrices.

Therefore, by comparing both ways of initialization, we found that sampling the covariance matrix initial values uniformly would yield better fitting and more accurate parameter estimation results than only sampling variance values and setting all other entries to zero. Both initialization methods show that the Lie algebraic method is more accurate when TL data are used for optimization, while the Givens approach performs better when variance data are used regarding parameter accuracy. However, fittings of the variance data estimated by the Givens approach is far off even with a good match of the covariance matrix values.

5.5 Discussion

5.5.1 Summary and conclusion

In biological systems, it is possible that the single-cell parameters are correlated. Therefore, it is important, when describing them as MEMs, to allow for correlation between random effects. In this chapter, we compared a novel Lie algebraic parametrization approach for the covariance matrix proposed by Stapor (2021) with the matrix logarithm, Cholesky, spherical, and Givens approaches. We applied the five methods to five benchmark models and one application example of a JAK/STAT signaling model. Two different initialization methods to sample the initial values of the covariance matrix are compared using the benchmark and application models.

By applying the above-mentioned methods to five benchmark cases with uniform initialization, we show that the Lie algebraic and Givens parametrization approaches outperform all the other

methods regarding both converged start numbers and the objective function values. When only the variance values are uniformly sampled, results show that the Cholesky approach yields the best convergence, while Lie algebraic and Givens approaches are the second best.

As an application example, we implemented a JAK/STAT model with five random effects and three outputs. Three different data types are used for parameter estimation separately, namely TL data with five cells, and mean and variance data generated using 1,000 simulated cells. Because of the high probability of getting integration errors using the MC sampling method, we use the CMD method with 30 samples instead for estimating the mean and variance during simulation. This ensures the number of runs can be finished, while keeping a certain accuracy of approximation at the same time.

Unlike in the benchmark models, we found that the Cholesky approach is the least robust approach in the application example, where none of the starts get finished even using the CMD method for all three data types. When uniform sampling is applied to all initial matrix values, the Lie algebraic approach performs best for all three data types regarding the quality of fitting. The Givens approach performs best regarding the covariance matrix error when using variance data, but only marginally better than the Lie algebraic method. It needs to be noted that the MC method with 10,000 samples can also be used for computing the mean and variance using parameters estimated by the Lie algebraic approach. This can possibly be explained by more reasonable covariance values estimated by the Lie algebraic method. When uniform sampling is only applied to the variance values and all other matrix entries are set to zero, none of the five approaches is able to fit the TL data.

In summary, we compared five parametrization approaches to estimate the covariance matrix in MEMs. By using five benchmark tests and one application problem, we demonstrated that the Lie algebraic approach proposed by Stapor (2021) outperforms all other parametrization methods regarding both convergence performance and fitting errors in almost all investigated use cases. We anticipate that such an approach will improve the accuracy of estimating correlations in a lot more biological pathway models.

5.5.2 Outlook

In this chapter, we assume non-zero values for all covariance matrix entries when estimating the covariance matrix values in the application example. However, usually, not all random effects are correlated in real biological problems. Therefore, an l_1-norm can be implemented in the likelihood function to ensure sparsity (Ning and Georgiou, 2011). Furthermore, in a real biological pathway model, it is usually possible to know which parameters are correlated, and which are not. Therefore, one can estimate only submatrices values and assume the other values to be zero (Adlung et al., 2021).

We used two different methods to initialize the values of the covariance matrix for optimization, uniform samples of all entries and only variance values. It is also possible to employ other initialization methods, for instance, the Latin-hypercube method. Therefore, we can test other methods of sampling the initial values in further investigation.

Although the Cholesky parametrization approach performs relative good in the test problem, it

is the least robust approach used in the application example. This is because that the randomly sampled parameter values may lead to integration errors of the ODEs. Therefore, the performance of the Cholesky approach can be further tested on ODE problems, which is less prone to get integration errors or with smaller parameter space.

Chapter 6

Modelling early treatment response in acute myeloid leukemia patients

Time-resolved monitoring of a patient's disease during acute myeloid leukemia (AML) intensive induction chemotherapeutic treatment is currently only possible via invasive bone marrow biopsies and is associated with substantial patient discomfort. In contrast, the residual disease can be easily assessed by measuring DNA fragments in the peripheral blood that contain tumor-specific mutations, called cell-free tumor DNA (ctDNA). Yet, it is unclear how ctDNA concentration is related to blast cell kinetics in the bone marrow, and if we can predict patients' risk of relapse after the chemotherapy. Besides, AML is a highly heterogeneous disease, with various mutations and different responses to chemotherapy. Therefore, mechanistic model parameters tend to be different for different patients. In this thesis, we use MEMs to describe the heterogeneity and approximate the variation of parameters.

In this chapter, we implemented three mechanistic models, given by ordinary differential equations and combined with MEMs, to model the response of blast cells to chemotherapy in 10 AML patients with two different AML-associated mutations (*NPM1* and *IDH2*). Results show that a three-step description coupled with mixed effects modeling fits the clinical data of blast cell and ctDNA numbers best. With the fitted model, numbers of blast cells in the bone marrow can be predicted using only ctDNA measurement from the blood. Furthermore, the scaling factor, which represents the heterogeneity of AML-related mutations, is potentially related to the relapse after chemotherapy.

This chapter is based on and partly identical to the following publication, in which all analyses, implementation of the code and models are performed by the thesis author:

- **Wang, D**, Raush, C, Buerger, S, Tschuri, S, Rothenberg-Thurley, M, Schulz, M, Hasenauer, Ziemann, F, Metzeler, K, Marr, C, Modeling early treatment response in AML from cell-free tumor DNA. iScience, 2023, 26(12).

6.1 Introduction

Acute myeloid leukemia (AML) is the most common acute leukemia in adults (Estey et al., 2008). In patients with AML, cell differentiation stops at the myeloblast stage (Afenya, 1996; Nair et al., 2021), leukemic myeloblasts fill the bone marrow niche, displace normal hematopoiesis, and lead to severe illness. Initial diagnosis of AML includes a bone marrow aspiration to determine the percentage of blast cells in the bone marrow. The current standard of care for fit AML patients involves an initial phase of intensive chemotherapy followed by post-remission treatment including additional chemotherapy and/or allogeneic stem cell transplantation (Schlenk, 2014). Within 16 to 21 days after the beginning of intensive chemotherapy, a second bone marrow sample is often taken to evaluate the reduction in the number of blast cells (blast clearance). Additional follow-up bone marrow biopsies are needed for disease monitoring during and after postremission therapy and are associated with considerable patient discomfort.

To reduce the burden on the patient, time-resolved monitoring of cell-free tumor DNA has been proposed (Colmenares et al., 2022). Cell-free tumor DNA (ctDNA) consists of small DNA fragments found in the blood plasma of cancer patients and can be used to detect tumor mutations (Alix-Panabières et al., 2012; Pantel and Alix-Panabières, 2019). Decreasing ctDNA levels during the initial phase of induction chemotherapy showed promise as a marker for the early assessment of therapy response and prognostic tool in AML patients (Holdenrieder et al., 2001; Mueller et al., 2006; Nakamura et al., 2018). However, it is still not clear how ctDNA detected in peripheral blood relates to the percentage of blast cells in the bone marrow, as clearance of leukemic blasts in the bone marrow after initial chemotherapy is a prerequisite for achieving complete remission, an important treatment milestone with relatively favorable prognosis for AML patients. A mathematical model can link these two quantities in measured time series.

By fitting ODE models to clinical time series, unknown parameters and unmeasured states of the model, for instance the rate constants (Palaniappan et al., 2013) and the amount of ctDNA detected in peripheral blood, can be estimated. Furthermore, for heterogeneous diseases such as AML (Martelli et al., 2013; Welch et al., 2012), the use of MEMs as described in Section 2.2 is particularly relevant. In our case, initial blast cell numbers and the concentration of ctDNA in the peripheral blood will differ between patients, because of individual effects, such as the time point of diagnosis. MEMs were thus employed to describe patient-to-patient variability and estimate the population parameters for all patients simultaneously.

In this chapter, we studied the percentage of bone marrow blast cells and ctDNA concentration in peripheral blood from ten AML patients undergoing initial chemotherapy (Rausch et al., 2021). We compared three hypotheses on ctDNA kinetics via ODE model implementation and MEM parameter fitting. We then proved that the best fitting model is structurally identifiable and able to predict the percentage of bone marrow blast cells from ctDNA data from peripheral blood as input. A model inherent scaling factor, which represents the heterogeneity of AML-related mutations, correlates with the risk of relapse.

6.2 Mechanistic modeling of AML chemotherapy process

6.2.1 Three mechanistic models and parameter assumption

Blast cell apoptosis happens during chemotherapy treatment, thus the small DNA fragments (ctDNA), which carries the tumor-specific mutations, are released and transits to the peripheral blood. To describe ctDNA and blast kinetics, we considered three models with increasing complexity (Figure 6.1):

- The one-step model ignores blast cells and assumes a simple exponential decay of ctDNA in peripheral blood (Figure 6.1(a)).
- The two-step model considers blast cells in the bone marrow and ctDNA in peripheral blood with a transition reaction between the two compartments (Figure 6.1(b)).
- The three-step model assumes that blast cells die and release ctDNA in the bone marrow (Figure 6.1(c)). ctDNA then transits to peripheral blood. Such an intermediate state allows for delayed arrival of ctDNA in the peripheral blood compartment.

The ODEs of the three models and their respective parameters are listed in Tables 6.1 and 6.2, respectively.

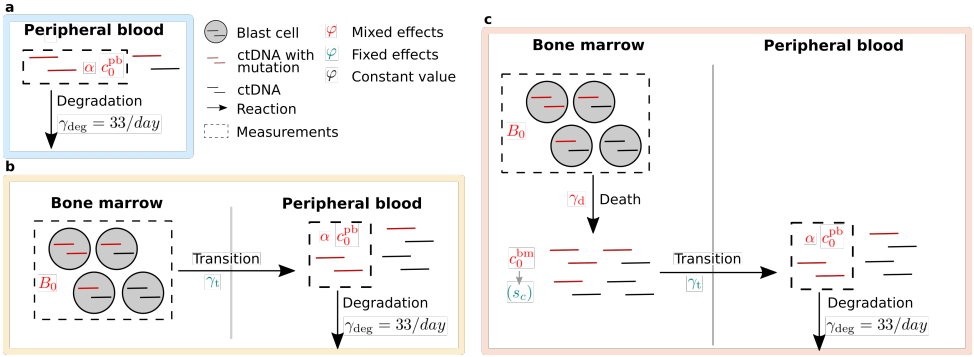

Figure 6.1: **Three-step model fits both population and single-patient data of blast cell and ctDNA kinetics.** (a) A one-step model describes the kinetics of ctDNA in peripheral blood as exponential decay. (b) A two-step model considers bone marrow and peripheral blood. When blast cells die upon chemotherapeutic treatment, their ctDNA appears in peripheral blood. (c) The three-step model includes an intermediate state: After blast cells die, their ctDNA is released in the bone marrow and then transits into peripheral blood at a corresponding rate.

In all models, we focus on the change in the number of ctDNA molecules in a patient's body, $\dot{c}^{pb}$, and on the measurement of a subset of ctDNA molecules with a specific mutation (*NPM1* or *IDH2*), y. In the two-step and three-step models, we also consider the changing number of blast cells (independent of their mutational profile) in the bone marrow B. The number of ctDNA (c^{pb}) and the number of blast cells (B) are independent of the specific mutation. Only the measurement y is specific for *NPM1* and *IDH2*, respectively. Thus, the scaling factor α couples (i) the fraction of ctDNA with the specific mutation in all ctDNA, and (ii) the fraction of ctDNA

Table 6.1: Definition of three models describing ctDNA kinetics with increasing complexity.

Model	One-step	Two-step	Three-step
Ordinary differential equation	$\dot{c}^{pb} = -\gamma_{deg} c^{pb}$ $c^{pb}(0) = c_0^{pb}$ $y = \alpha c^{pb}$	$\dot{B} = -\gamma_t B$ $\dot{c}^{pb} = \gamma_t B - \gamma_{deg} c^{pb}$ $B(0) = B_0$ $c^{pb}(0) = c_0^{pb}$ $y = \alpha c^{pb}$	$\dot{B} = -\gamma_d B$ $\dot{c}^{bm} = \gamma_d B - \gamma_t c^{bm}$ $\dot{c}^{pb} = \gamma_t c^{bm} - \gamma_{deg} c^{pb}$ $B(0) = B_0$ $c^{bm}(0) = c_0^{bm}$ $c^{pb}(0) = c_0^{pb}$ $c_0^{bm} = \frac{\gamma_{deg}}{\gamma_t} c_0^{pb} s_c$ $y = \alpha c^{pb}$
Fixed effects		γ_t	γ_t
Mixed effects	c_0^{pb}, α	B_0, c_0^{pb}, α	B_0, c_0^{pb}, γ_d, s_c, α
Measurement noise	σ_c	σ_c	σ_c
Number of population parameters	5	8	12
Number of patient specific parameters	2	3	5

that can be measured experimentally. Since the measurable fraction (ii) should be similar for all patients, we are comparing different fractions of mutated ctDNA when we compare α between patients. A low value of α thus indicates a low fraction of specifically mutated *NPM1* or *IDH2* ctDNA and a high fraction of ctDNA molecules with other mutations.

Table 6.2: Definition of model parameters.

Parameter	Description
c_0^{pb}	Total amount of all ctDNA in peripheral blood at start of treatment
s_c	Scaling factor of initial value of ctDNA in bone marrow to ensure positive gradient of $c^{pb}(0)$
B_0	Total amount of blast cells in bone marrow at the start of treatment
γ_d	Death rate of bone marrow blast cells during treatment in 1/days
α	Scaling factor of ctDNA in peripheral blood
γ_t	Transition rate of ctDNA from bone marrow to peripheral blood in 1/days. We assume this to be physiologically constrained and similar for all patients, thus handling it as a fixed effect
σ_c	Standard deviation of measured ctDNA

As the ctDNA measurement noise σ_c is unknown, we estimate it simultaneously with other parameters. The details of our parameter assumptions can be found in Table 6.2.

In order to consider the patient-to-patient variability, we implemented MEMs as defined in Section 2.2. Parameters are also shown in Figure 6.1. The initial values of the blast cells and ctDNA amount are assumed to be mixed effects as usually patients are diagnosed at different stages of AML. We also assume that the death rate of blast cells in the bone marrow is a mixed effect, because patients may respond differently even to the same clinical treatment. The transition rate γ_t is assumed to be a fixed effect, we assume that this is physiologically constrained, therefore same for each patient. As the ctDNA measurement noise σ_c is unknown, we estimate it simultaneously with other parameters.

To avoid an unphysiological drop of ctDNA in peripheral blood at the beginning of the treatment, we constrained our models. We want to ensure that the derivative of c^{pb} at time 0 to be non-negative,

$$\dot{c}^{pb}(0) = \gamma_t c_0^{bm} - \gamma_{deg} c_0^{pb} \geq 0 \tag{6.1}$$

Thus,

$$c_0^{bm} \geq \frac{\gamma_{deg}}{\gamma_t} c_0^{pb} \tag{6.2}$$

We can ensure this with a scaling factor s_c with the boundary $[1, \infty]$:

$$c_0^{bm} = \frac{\gamma_{deg}}{\gamma_t} c_0^{pb} \cdot s_c, \; s_c \geq 1 \tag{6.3}$$

Therefore, the gradient of at time 0 is ensured to be not negative.

Parameter assumptions

Because of the limited number of measurement data and the complexity of models, we constrained parameters using the information found in the literature, in order to decrease the number of parameters that need to be estimated to improve identifiability.

- As we do not have multiple measurements to assess measurement noise, we assume a multiplicative noise of 10% (Barnett et al., 1933).
- We assume the degradation rate of ctDNA to be 33[day^{-1}], which equals a half-life of 30 min (Yu et al., 2013).
- Bone marrow of AML patients is usually completely packed with leukemic blasts, we assume that every blast belongs to the AML (meaning is carrying mutations) and the amount of healthy bone marrow blasts is neglectable.

6.2.2 Parameter estimation

As the data are taken at different time points for each single patients in this chapter, we considered the likelihood computed using first order conditional estimation method as introduced in Section 2.3.2 for the time-lapse data.

Analytical solution and model identifiability

In order to check the identifiability of the most complex model considered in this chapter, we solved the ODEs of all three models analytically as follows:

- One-step model

$$c^{pb}(t) = c_0^{pb} \exp(-\gamma_{deg} t)$$

The observation is

$$y = \alpha c^{pb}$$

- Two-step model

$$B(t) = B_0 \exp(-\gamma_t t)$$

$$c^{pb}(t) = \begin{cases} \gamma B_0 t \exp(-\gamma t) + c_0^{pb} \exp(-\gamma t), & \text{if} \quad \gamma_{deg} = \gamma_t = \gamma \\ \frac{\gamma_t B_0}{\gamma_{deg} - \gamma_t} \exp(-\gamma_t t) + \left(c_0^{pb} - \frac{\gamma_t B_0}{\gamma_{deg} - \gamma_t} \right) \exp(-\gamma_{deg} t), & \text{if} \quad \gamma_{deg} \neq \gamma_t \end{cases}$$

The observations are

$$y_1 = \alpha c^{pb}$$
$$y_2 = B$$

- Three-step model

$$B(t) = B_0 \exp(-\gamma_d t)$$

$$c^{bm}(t) = \begin{cases} \gamma B_0 t \exp(-\gamma t) + C_0^{bm} \exp(-\gamma t), & \text{if} \quad \gamma_t = \gamma_d = \gamma \\ \frac{\gamma_d B_0}{\gamma_t - \gamma_d} \exp(-\gamma_d t) + \left(c_0^{bm} - \frac{\gamma_d B_0}{\gamma_t - \gamma_d} \right) \exp(-\gamma_t t), & \text{if} \quad \gamma_t \neq \gamma_d \end{cases}$$

$$
c^{pb}(t) = \begin{cases}
\left(\frac{1}{2}\gamma^2 B_0 t^2 + c_0^{pb}\right)\exp\left(-\gamma t\right), & \text{if} \quad \gamma_{deg} = \gamma_t = \gamma_d = \gamma \\
\left(\frac{\gamma^2 B_0}{\gamma_{deg}-\gamma}t - \frac{\gamma^2 B_0}{\left(\gamma_{deg}-\gamma\right)^2} + \frac{\gamma c_0^{bm}}{\gamma_{deg}-\gamma}\right)\exp\left(-\gamma t\right) & \\
+\left(c_0^{pb} - \frac{\gamma^2 B_0}{\left(\gamma_{deg}-\gamma\right)^2} - \frac{\gamma c_0^{bm}}{\gamma_{deg}-\gamma}\right)\exp\left(-\gamma_{deg} t\right), & \text{if} \quad \gamma_{deg} \neq \gamma_t, \gamma_t = \gamma_d = \gamma \\
\frac{\gamma_1\gamma_d B_0}{\left(\gamma_1-\gamma_d\right)^2}\exp\left(\gamma_d t\right) + \left(c_0^{bm} - \frac{\gamma_d B_0}{\gamma_1-\gamma_d}\right)t\exp\left(-\gamma_1 t\right) & \\
+\left(c_0^{pb} - \frac{\gamma_1\gamma_d B_0}{\left(\gamma_1-\gamma_d\right)^2}\right)\exp\left(-\gamma_1 t\right), & \text{if} \quad \gamma_{deg} = \gamma_t = \gamma_1, \gamma_t \neq \gamma_d \\
\frac{\gamma_t\gamma_2 B_0}{\gamma_t-\gamma_2}t\exp\left(-\gamma_2 t\right) + \left(\frac{c_0^{bm}}{\gamma_2-\gamma_t} + \frac{\gamma_2 B_0}{\left(\gamma_2-\gamma_t\right)^2}\right)\exp\left(-\gamma_t t\right) & \\
+\left(c_0^{pb} - \frac{c_0^{bm}}{\gamma_2-\gamma_t} - \frac{\gamma_2 B_0}{\left(\gamma_2-\gamma_t\right)^2}\right)\exp\left(-\gamma_2 t\right), & \text{if} \quad \gamma_{deg} = \gamma_d = \gamma_2, \gamma_d \neq \gamma_t \\
\frac{\gamma_t\gamma_d B_0}{\left(\gamma_{deg}-\gamma_d\right)\left(\gamma_t-\gamma_d\right)}\exp\left(-\gamma_d t\right) & \\
+\frac{\left(\gamma_t-\gamma_d\right)\gamma_t c_0^{bm} - \gamma_t\gamma_d B_0}{\left(\gamma_{deg}-\gamma_t\right)\left(\gamma_t-\gamma_d\right)}\exp\left(-\gamma_t t\right) & \\
+\left(c_0^{pb} + \frac{\gamma_t\gamma_d B_0}{\left(\gamma_t-\gamma_d\right)\left(\gamma_{deg}-\gamma_d\right)} - \frac{\gamma_t c_0^{bm}}{\gamma_{deg}-\gamma_t}\right. & \\
\left.+\frac{\gamma_t\gamma_d B_0}{\left(\gamma_{deg}-\gamma_t\right)\left(\gamma_t-\gamma_d\right)}\right)\exp\left(-\gamma_{deg} t\right), & \text{if} \quad \gamma_{deg} \neq \gamma_t \neq \gamma_d
\end{cases}
$$

The observations are

$$
\begin{aligned}
y_1 &= \alpha c^{pb} \\
y_2 &= B
\end{aligned}
$$

From the equations above, it can be seen that in the one-step model the parameters c_0^{pb} and α have the same effects on the output y. Therefore, the one-step model parameters are non-identifiable. We have also checked the identifiability of the models using the MatLab toolbox GenSSI 2.0 (Ligon et al., 2018). We see that if both, blast cell number and ctDNA concentration, are provided as measurements, the two-step model and three-step model are structurally identifiable according to Equation (2.38). If only ctDNA concentration is provided, then one of the parameters is not structurally identifiable. In order to improve the computation efficiency, we implemented gradient-based optimization method by providing the analytical gradients as listed in the Appendix.

Single patient parameter estimation

In this chapter, we investigated not only the population parameters of all patients included but also the parameters of a single patient. The single-patient parameters are used in two scenarios: 1) We used the single-patient parameters, which are estimated by using the population parameters of the whole patient cohort as prior, to evaluate the possibility of relapse. 2) When the blast cell measurements of one patient are excluded, the population parameters estimated using the

other patients are used as prior together with the ctDNA measurement from the specific patient to estimate the corresponding single-patient parameters.

Similar to the inner optimization loop introduce in Section 2.3.2, single-patient parameters are computed using empirical bayes estimation, i.e., the random effects for each patient are optimized separately, while giving the population parameters as prior:

$$\hat{b}_k(\beta, D) = \text{argmax}_{b_k} \left(p\left(\left\{ t_{jk}, \{\bar{y}_{ijk}\}_{i=1}^{d_y^{(k)}} \right\}_{j=1}^{d_t^{(k)}} | \beta, b_k \right) + p(b_k|D) \right). \tag{6.4}$$

Here, i, j, k are the indices of measurement, time point, and patient, respectively, β is the fixed effects, and D is the covariance matrix of random effects, t is the time point, and $\bar{y}$ is the measurement data. $d_y^{(k)}$ and $d_t^{(k)}$ are the dimension of measurements and time points, that are patient specific, respectively.

6.3 Application to patient data

6.3.1 Clinical data

We used data from 10 AML patients treated at the LMU Department of Hematology, as previously described (Rausch et al., 2021). Blast cell percentages in bone marrow were measured for all patients at initial diagnosis and 16-18 days after the beginning of induction therapy (Table 6.3, Figure 6.2(a)). The ctDNA concentration for two recurrent AML-related mutations, affecting the *NPM1* and *IDH2* genes, was measured between 2 and 14 times (mean = 8.7, std = 3.7) during the first 18 days after initiation of induction chemotherapy. During model fitting, we scaled measurements to absolute blast cell numbers and ctDNA numbers in the whole body instead of blast cell percentage and ctDNA concentration (Figure 6.2(b) and 6.2c). For this purpose, we assume the total blood volume to be 5L (Sharma and Sharma, 2023), and the nucleated cell number in the bone marrow to be 1.2×10^{12} (Cosgrove et al., 2021). Details of the underlying assumptions can be found in Section 6.3.2.

6.3.2 Data preprocessing

In order to decrease the number of parameters that need to be estimated, we constrained parameters using the information found in the literature.

- Our measured data in the bone marrow compartment is the percentage of blast cells in all nucleated cells, p^{pb}. To compute the total number of blast cells, B, we assume that the nucleated cell number in the bone marrow is 1.2×10^{12} (Cosgrove et al. 2021): $B = 1.2 \times 10^{12} p^{pb}$.
- ctDNA is measured as concentration $C^{pb}(1/L)$. To compute the total amount of ctDNA c^{pb} in peripheral blood, we assume the blood volume to be 5L (Sharma and Sharma, 2023).

Table 6.3: Clinical data of *NPM1* and *IDH2* mutated patients.

Patient ID	Age	Gender	Mutation	Relapse	# ctDNA measurements	# blast cell measurements
N1	28	Female	*NPM1*	No	9	2
N2	36	Female	*NPM1*	No	2	2
N3	46	Female	*NPM1*	No	9	2
N4	52	Female	*NPM1*	No	7	2
N5	59	Male	*NPM1*	No	5	2
N6	60	Female	*NPM1*	No	12	2
N7	67	Male	*NPM1*	Yes	12	1
I1	55	Female	*IDH2*	No	14	2
I2	67	Male	*IDH2*	Yes	6	2
I3	68	Male	*IDH2*	Yes	11	2

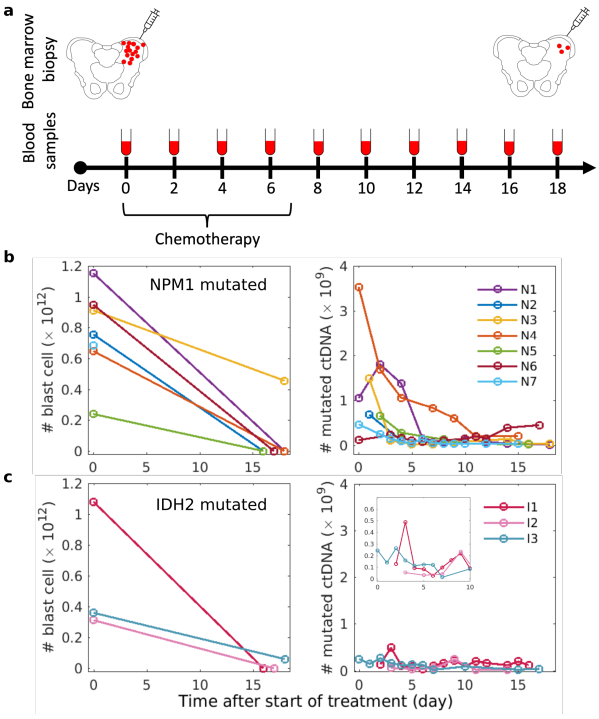

Figure 6.2: **Measurement scheme and blast cell and ctDNA kinetics of 10 AML patients under chemotherapy treatment.** (a) Overview of sample collection. Bone marrow aspiration is taken at initial diagnosis and 16 to 21 days after the start of chemotherapy, to assess leukemic blast count (red dots) within the bone marrow. Peripheral blood, for measuring ctDNA kinetics, was collected at least every second day during the first 21 days after the beginning of treatment. (b) Absolute number of bone marrow blast cells and ctDNA for seven *NPM1* mutated AML patients. (c) Absolute number of bone marrow blast cells and ctDNA for three *IDH2* mutated AML patients.

- The first measurement is often taken 1 to 20 days before the chemotherapy treatment. We assume that the blast cells do not proliferate dramatically, therefore we use the measured data as the initial point of the modelled period.

6.3.3 The three-step mixed effects model can fit both population and individual measurements

To assess the adequacy of the three proposed models for describing blast cell and ctDNA kinetics, their unknown parameters were inferred from the experimental data. Therefore, we fitted the MEMs using our internally developed MatLab toolbox MEMOIR (https://github.com/ICB-DCM/MEMOIR). The parameter optimization for all three models converged and provided reproducible results (see estimated parameter values Table 6.4, and waterfall plots in Figure 6.4). The assessment of these results revealed that the one-step model does not provide a satisfactory fit due to the high reported values of the degradation rate (Yu et al., 2013). Both the population-level fit and the fit for individual patients are inadequate (Figure 6.3(a), (b)). For the two-step model, we observed a very large variability in the population fit (Figure 6.3(c)), but reasonable single-patient fits (Figure 6.3(d)). When we use the three-step model, the population fitting agrees well with the patient data (Figure 6.3(e)), and also single-patient kinetics are fitted well (Figure 6.3(f)). Comparing the single-patient fits in more detail and zooming into the first 24 hours, we see a very steep decrease when using the two-step model (Figure 6.3(d), inset). In contrast, the three-step model shows a short increase of ctDNA in the peripheral blood before levels drop. Based on these model fits, we continue analyzing the three-step model.

6.3.4 Blast cell numbers can be predicted given estimated parameters and ctDNA data

To assess the value of the three-step model, we tested the prediction performance for the absolute bone marrow blast cell numbers from ctDNA measurements in peripheral blood. We assume that we have only measured ctDNA from one patient, and combine this with the population parameters estimated from the other patients as prior (Figure 6.5(a)). With that approach, the bone marrow cell numbers can be predicted and compared with the measured data of each patient, respectively. We then validated our method by excluding each patient separately (Figure 6.5(a)).

Results show that we hit the 10% variance of the measured blast cell numbers for four out of seven patients (Figure 6.5(b)). The disagreement of the prediction for patients N1 and N4 (Figure 6.5(b)) can be explained by the fact that the blast cell measurement of those two patients are upper and lower limits in our dataset (Figure 6.5(c)). We also fail to predict the second blast cell measurement of patient N3, 18 days after therapy started, probably because the kinetics is so different from other patients. We speculate that this measurement is an outlier to all other data points due to measurement errors. However, when patient N3 is included in parameter estimation, the corresponding single-patient measurement data can still be fitted, meaning that our model is flexible enough to fit the different kinetics. We expect that by including more data with similar kinetics, we would be able to predict the specific measurement as well.

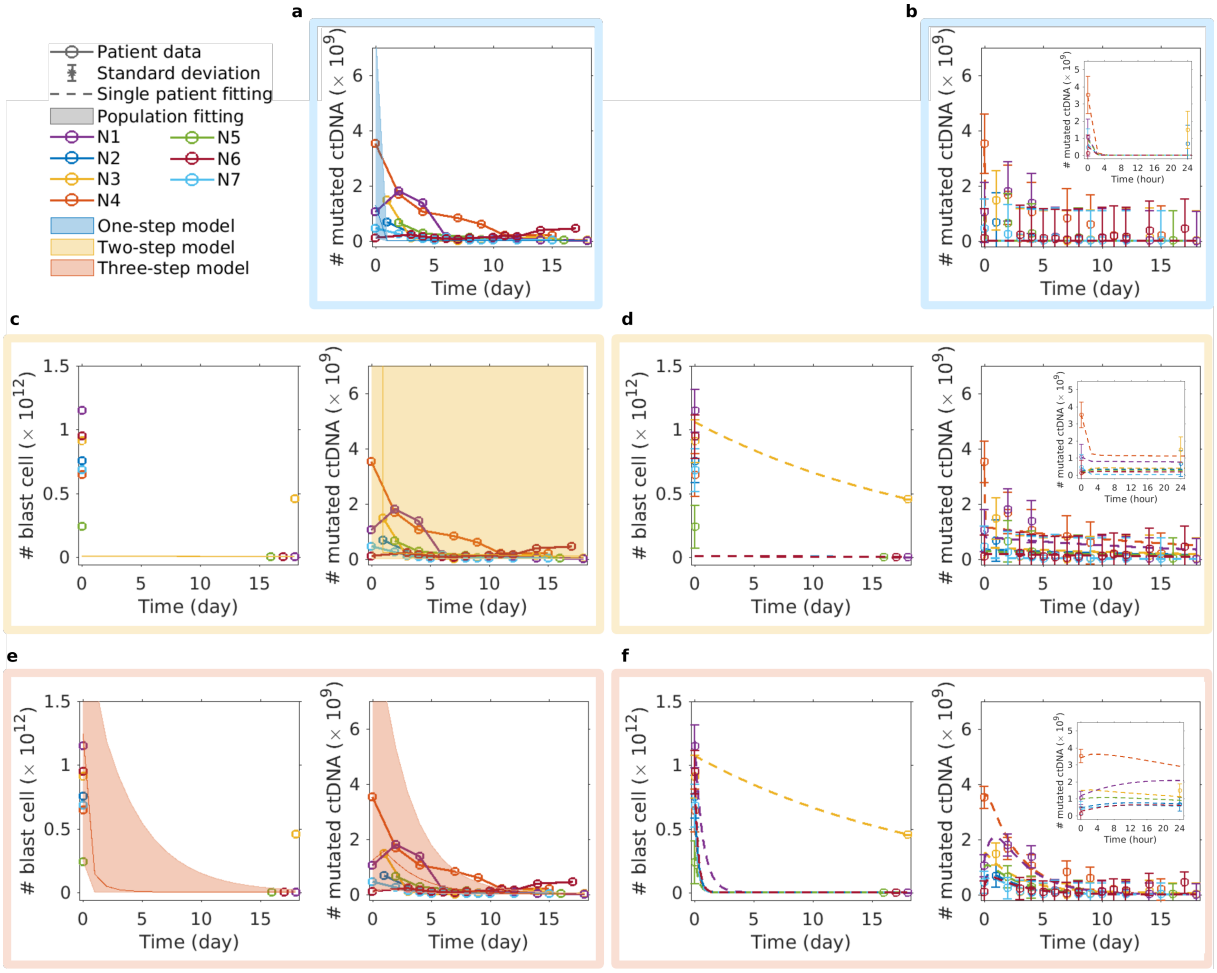

Figure 6.3: **Three-step model fits both population and single-patient data of blast cell and ctDNA kinetics.** (a) Population fitting with the one-step model fails to explain the slow decay of the number of mutated ctDNA within days. (b) The zoom into the first 24 hours (inset) of the single patient fits using the one-step model reveals a rapid model decay of ctDNA within hours instead of days, as observed in the data. (c) Two-step model population fits estimate an overly large variance of the number of mutated ctDNAs (yellow shading). (d) Two-step model single-patient fits agree well with single-patient data. However, the zoom into the first 24 hours (inset) shows an overly steep drop in the first four hours. The standard deviation of the blast cell count is assumed to be 10% of the measured values. (e) Three-step model population fits agree well with blast cell numbers and the number of mutated ctDNA. (f) Three-step model single-patient fits agree well with each single-patient. The zoom into the first 24 hours (inset) shows smooth kinetics. Again, the standard deviation of the blast cell count is assumed to be 10% of the measured values.

6.3.5 Kinetics of patients with *NPM1* and *IDH2* mutations can be fit using one set of parameters

The ctDNA measurements of *NPM1* and *IDH2* mutated patients are in the same scale (Figure 6.2). Therefore, we tried to fit the data from the two patient groups simultaneously. While only three patients with *IDH2* mutation are measured in our study, by combining both data sets, the parameters can be better constrained by data from *NPM1* mutated patients. Results show that data from both mutations can be fit simultaneously and that the population and single-patient

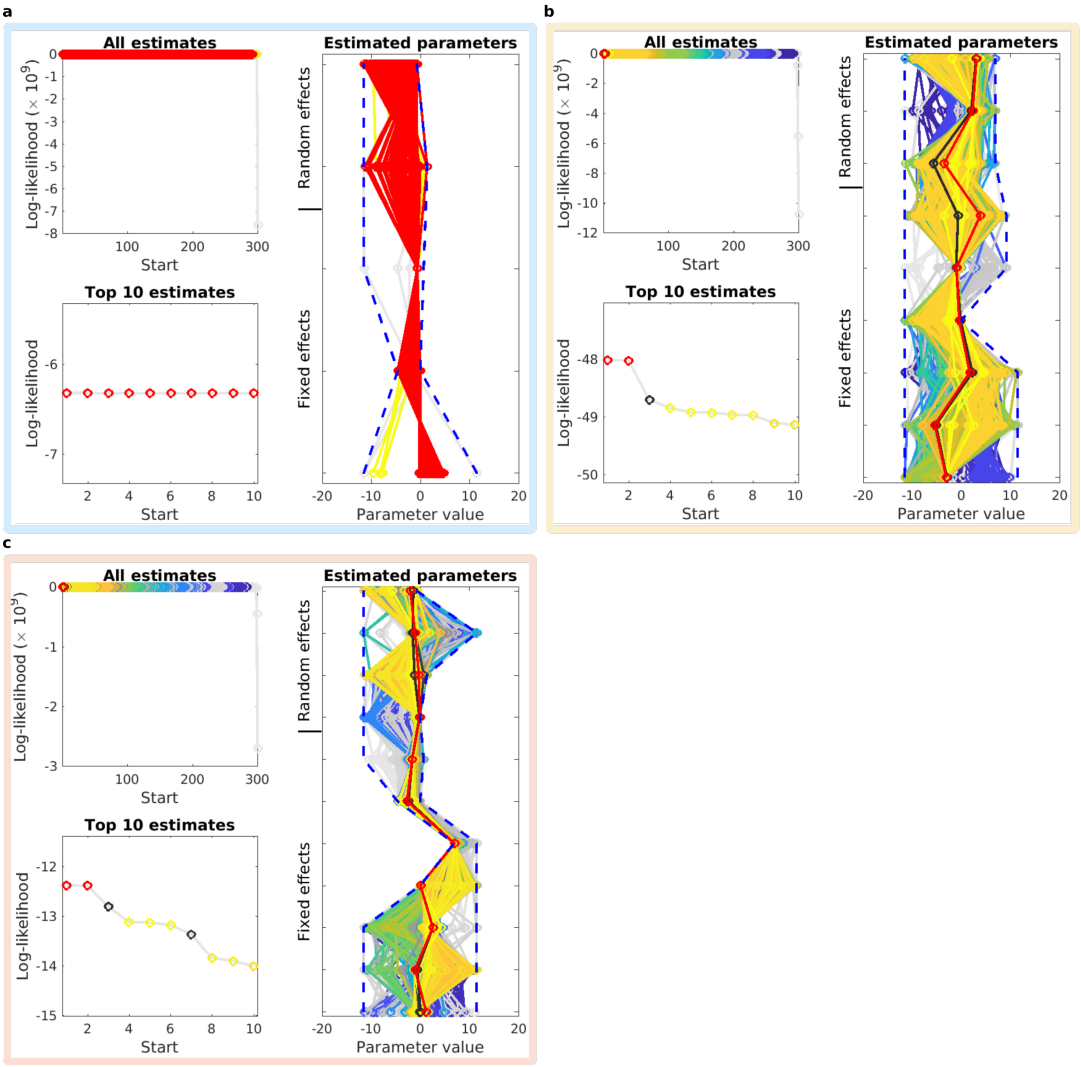

Figure 6.4: **Convergence performance and reproducibility of (a) one-step model, (b) two-step model and (c) three-step model.** Top left: waterfall plot of all 300 local optimization starts, the multiple red color dots show convergence of the corresponding model. Bottom left: log-likelihood values of the top 10 local optimization starts. Right: parameter values of all local optimization starts, the overlaid parameter values show producibility of the corresponding model.

Table 6.4: Estimated parameter values of the two-step and three-step model.

Two-step model	Parameter	$B_0 \times 10^9$	$c_0^{pb} \times 10^9$	γ_t	α	Var(B_0)	Var(c_0^{pb})	Var(α)
	Value	5.63	0.0045	0.047	0.66	10.06	0.030	22.33
Three-step model	Parameter	$B_0 \times 10^9$	$c_0^{pb} \times 10^9$	γ_d	γ_t	s_c	α	Var(B_0)
	Value	1082.85	12.78	3.37	0.36	1.06	0.067	0.35
	Parameter	Var(c_0^{pb})	Var(γ_d)	Var(α)				
	Value	0.77	0.94	0.14				

fitting also agree well with patient data (Figure 6.6). Compared to the population fittings using only *NPM1* mutated data, the variance of predicted ctDNA data is even smaller, because of the

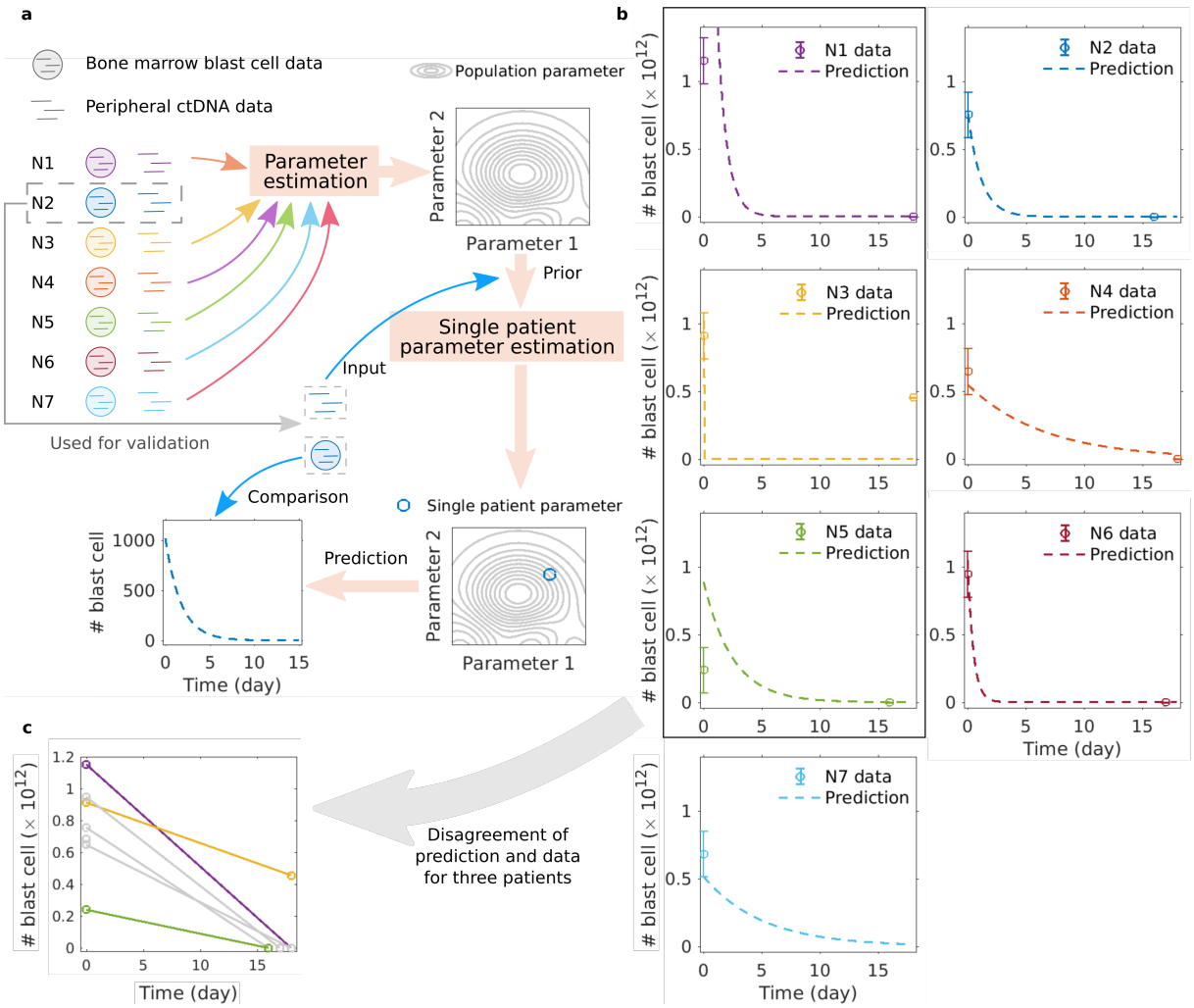

Figure 6.5: **Our three-step model predicts kinetics and scale of bone marrow blast cells during the first days of chemotherapy from peripheral blood ctDNA measurements.** (a) Predicting bone marrow blast cell numbers using only ctDNA data. One patient (N2 in this exemplary case) is excluded when estimating the population parameters of the three-step model. Next, the ctDNA data of N2 is used as input, combined with the population parameters as prior, and the single-patient parameters of N2 are estimated. For validation, the predicted blast cell number is compared with the true patient data. (b) Comparison of predicted bone marrow blast cell number and patient data. Bone marrow data of four out of seven patients can be predicted accurately. (c) N1 (violet) and N5 (green) constitute the upper and lower boundaries of the measured ctDNA numbers. The second blast cell measurement for N3 (yellow) is a potential outlier since all the other second measurements are 0.

lower scale of the measurement from *IDH2* mutated patients, which can also be seen from the mean of ctDNA measurement.

6.3.6 Single-patient parameters indicate disease relapse

To investigate the relationship between inferred parameter values and disease relapse, we compute single-patient parameters using the population parameter as a prior (Figure 6.7(a)). By

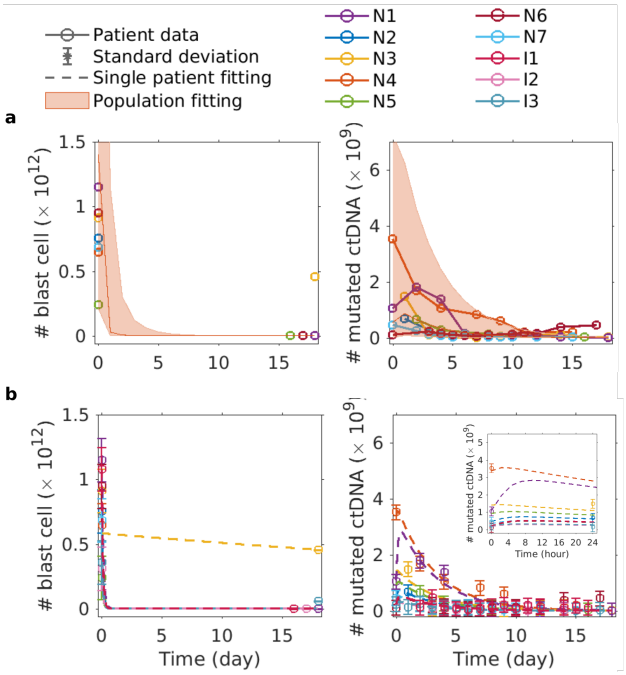

Figure 6.6: **The three-step model is able to fit *NPM1* and *IDH2* mutated patients simultaneously.** (a) Population fitting. (b) Single-patient fitting.

comparing all parameters (Figure 6.8), we found for the three relapsed patients (blue circles in Figure 6.7(b)) the scaling factor α to be smaller than for the seven non-relapsed patients. This holds when using only *NPM1* mutated data, or both mutations for inferring single patient parameters. Note that by giving the population parameters as a prior, the single-patient parameter α is identifiable (Figure 6.9, Equation (2.41)).

As α is the fraction of ctDNA with the specific mutation in the total amount of ctDNA, small α indicates that there is a large fraction of ctDNA which comes from a blast cell clone with mutations that we have not measured (not *NPM1* or *IDH2* in our case), i.e. a larger heterogeneity of AML related mutations. For instance, comparing patients N4 and N7 (Figure 6.7(d)), these are two patients with very similar bone marrow blast cell numbers and kinetics, we see that the ctDNA level of patient N7 is much lower than it for patient N4. As we specifically measured ctDNA with *NPM1* mutation for these two patients, the small measured fraction means that for patient N7, a large fraction of ctDNA released from blast cells contains other AML related mutations, which are not measured in our case. This agrees with the fact that patients with a heterogenous clone composition are more prone to relapse.

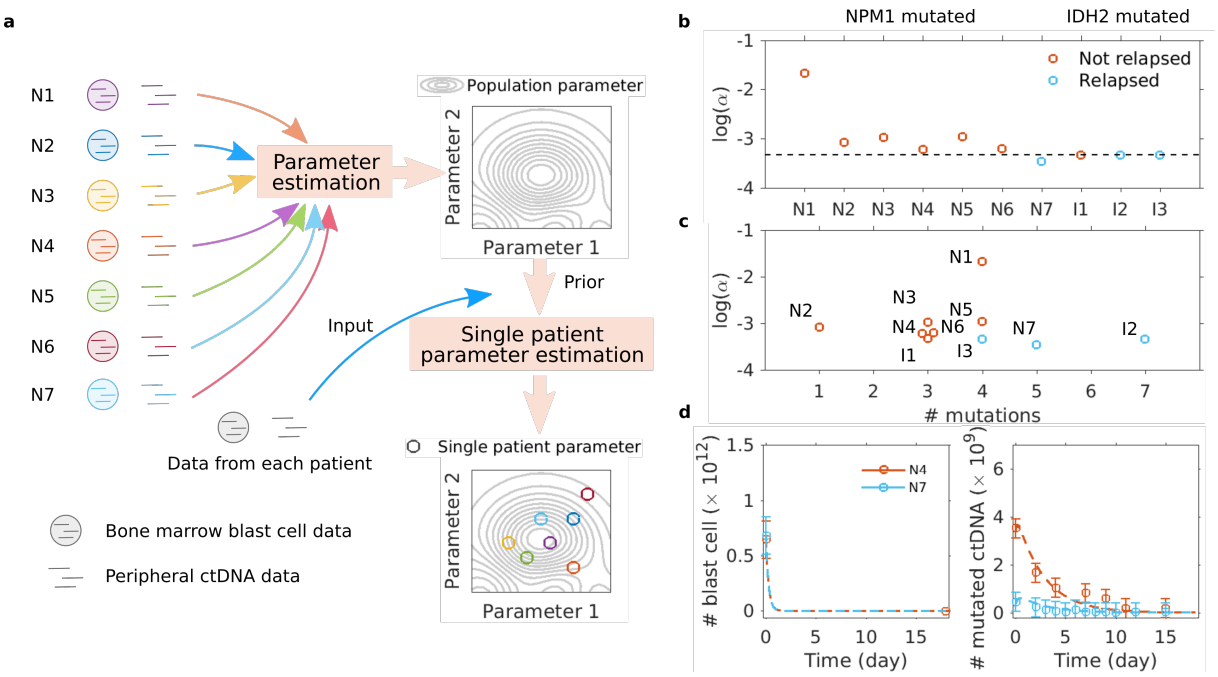

Figure 6.7: **Small fraction of measured mutations correlates with relapse.** (a) Pipeline of computing single-patient parameters. All patient data are used to estimate the population parameters, which are then given as a prior, and all single-patient data are used to estimate the corresponding parameters. (b) Values of measurement scaling factor α for each single patient. (c) Values of α versus number of mutations. (d) Single-patient fitting of patients N4 and N7 with similar blast cell kinetic and different ctDNA measurements.

6.4 Discussion

6.4.1 Summary and conclusion

Recent advances in AML have led to improved risk stratification and approval of new therapies to achieve complete remission (Döhner et al., 2022; Khoury et al., 2022). The best long-term outcome for intensively treated patients depends largely on early clearance of leukemic blasts in the bone marrow and negativity of measurable residual disease (MRD), which may have stronger prognostic significance than pretreatment genetic risk (Heuser et al., 2021; Blachly et al., 2022; Ihlow et al., 2022). For assessing treatment response, bone marrow biopsy remains the gold standard, as the sensitivity of MRD analyses from bone marrow is at least one log-fold higher than the same analysis from peripheral blood (Ivey et al., 2016; Rausch et al., 2021). Thus, 16-21 days after intensive induction chemotherapy, the percentage of remaining blasts in the bone marrow is determined. This timing is critical to guide further treatment, such as a second intensive induction chemotherapy. However, determining blast counts in bone marrow is a burden for the patient, and can be difficult in aplastic bone marrow due to aspicular samples, dilution with peripheral blood, or changes associated with early marrow recovery (Astle et al., 2017; Ihlow et al., 2022). Sometimes marrow analysis must be repeated after leucocyte regeneration. To address these challenges, we developed a mathematical model to predict blast cells in the bone marrow at this early time point by analyzing ctDNA from the peripheral blood of AML

patients. In 4 of 7 patients, our predictions are within 10% of the measured data. Our finding that 2 cases at the parameter limits cannot be predicted correctly gives us confidence that bone marrow prediction will work for the majority of patients in larger cohorts.

The potential of ctDNA to assess treatment response has been shown in a wide variety of solid tumors (Keller et al., 2021) and is a feasible tool to follow-up hematologic cancers, such as lymphoma and leukemia (Alig et al., 2021; Fernández-Miranda et al., 2023; Rausch et al., 2021; Nakamura et al., 2019). Combined with mathematical modeling, it outcompetes the standard, computed tomography imaging-based follow-up measurements in solid cancers (Avanzini et al., 2020). Using the percentage of bone marrow blast cells in combination with mutation-specific ctDNA measurements in peripheral blood, we estimated population parameters of a mixed effect model for a cohort of 7 *NPM1* and 3 *IDH2* mutated AML patients. The model was selected in a stepwise way, beginning with a simple one-step model to prevent overfitting (Lever et al., 2016), and increasing complexity. Our three-step model, where leukemic cells die within the bone marrow and release ctDNA that transits into the bloodstream where it degrades at a fixed rate, was able to fit the clinical data. One parameter of this model, a patient-specific scaling factor α, is small for relapsing patients, independent of the ELN risk group (Döhner et al., 2022). Since α quantifies the ratio of ctDNA with a specific mutation to the amount of total ctDNA, we assume that this parameter is informative about the tumor heterogeneity and might be suitable to identify patients who are at risk for relapse.

Interestingly, we are able to calculate the scaling factor α already when only 2 measurements of the ctDNA at early and late time points are performed. This excludes that the calculation of α is influenced by other sources of cell-free DNA (e.g. regeneration of healthy hematopoiesis after induction therapy). Thus our model might be used to refine risk assessment in AML.

In summary, we developed the first mechanistic model combining with MEM method to describe the early response to intensive induction chemotherapy in AML patients. We applied it to clinical data and showed that it is able to predict blast cell clearance from ctDNA data and might be able to predict relapse risk by simultaneously incorporating all available datasets. Using ctDNA measurements and disease modeling of AML might help to gain deeper insight into therapy response and clonal heterogeneity of AML. Our modeling approach might guide measurable disease monitoring and enhance risk stratification to identify patients at high risk for initial refractory disease or develop relapses.

Because only a limited number of clinical measurements at different time points and a total of only 10 patients were available, we had to restrict the model parameters to improve their identifiability. Therefore, we fixed the degradation rate for ctDNA to the value previously estimated by monitoring fetal cell-free DNA in women after delivery (Yu et al., 2013). Although the degradation of ctDNA and cell-free DNA should not differ, we cannot exclude cancer-specific effects due to inflammation or chemotherapy. Identifying ctDNA degradation rates in cancer patients would help validate our model for AML and previous models for solid cancers. Remarkably, all model parameters are identifiable. Another assumption of the model is that the number of bone marrow cells is constant. Although this number is certainly capped by physical constraints, it could be higher at the start of chemotherapy when blasts are abundant than at the second measurement over 2 weeks later. However, because blast measurements at this time are 0 in 8 of 10 patients, we believe that errors from this assumption affecting the parameter distributions of the two patients with nonzero measurements (patients N3 and I3) are negligible.

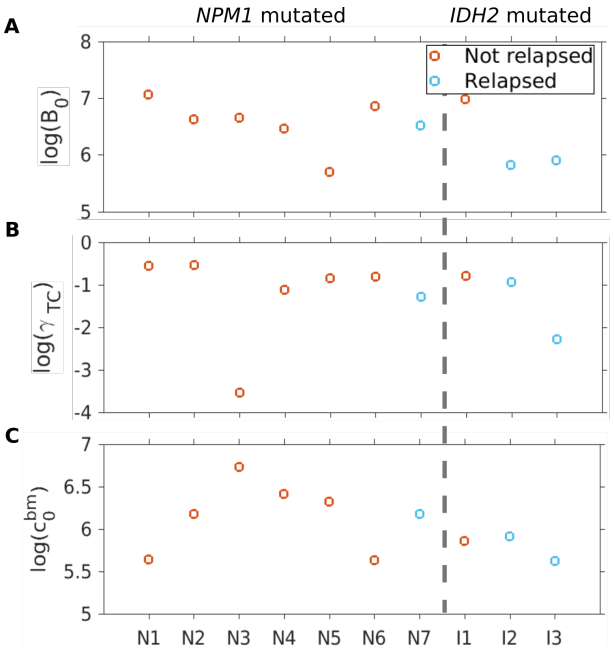

Figure 6.8: **Parameter estimates for the individual patients.** (a) Initial values of blast cell numbers for each patient. (b) Death rate of blast cells in the bone marrow for each patient. (c) Initial values of ctDNA in the bone marrow compartment for each patient.

6.4.2 Outlook

With our small cohort and only one mutation measured in ctDNA per patient we cannot conclude on the clonal trajectory of the AML, in particular as the clonal architecture of AML can be quite complex (Morita et al. 2020): Even mutations considered early in leukemogenesis, such as *NMP1*, may be preceded by other mutations and thus may not be present in all AML clones. While α is not informative about the mutations present at relapse, we found only minor changes in the mutational spectrum of our relapsed patients. Of note, the mutations we tracked (*NPM1* or *IDH2*) were still present at relapse because, as previously shown, *NPM1*-negative relapse in AML is quite rare when *NPM1* is present at initial diagnosis (Höllein et al., 2018). Because the presence of ctDNA is largely dependent on release from dying tumor cells and passage of ctDNA into the circulation, follow-up studies are needed to further investigate whether the clonal architecture of AML can be detected in ctDNA and whether different AML clones that respond differently to therapy also exhibit different ctDNA kinetics.

A larger cohort of patients could give us insight into the degree of variability between AML patients, especially with different mutation patterns, as we have only used ctDNA from two genes commonly mutated in AML, *NPM1* and *IDH2*. To analyze more than one mutation within the same patient over time a next-generation sequencing (NGS) panel from liquid biopsies would be required for AML, comparable to those commercially available for solid tumors and already used in precision oncology programs when tumor biopsies cannot be obtained (Pascual et al., 2022). Such a panel would also allow us to more reliably assess the impact of scaling factor α. It also has the potential to add relevant information to existing NGS panels from bone marrow aspirates and liquid biopsies, as it contains mutations that could only be identified using a whole genome sequencing approach (Zhao et al., 2014; Harris et al., 2019). A model based on parameters estimated from such a larger and more diverse group of patients could then be used

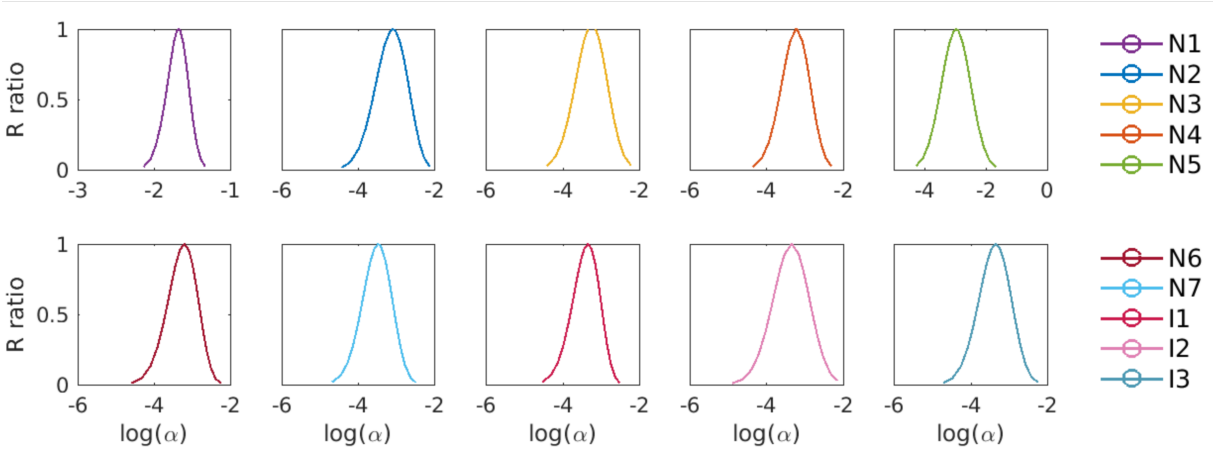

Figure 6.9: **Profile likelihood of α values for individual patient given population parameter values.** The α value ids identifiable for each patient.

in a multicenter cohort to predict treatment response and relapse risk. Because recent studies highlight the importance of MRD monitoring in AML patients regardless of treatment and the high risk of relapse for patients who do not achieve MRD negativity (Blachly et al., 2022), we believe that our model could also guide MRD monitoring by identifying relevant mutations that do not decrease during induction chemotherapy and should be followed in AML patients for MRD assessment.

Chapter 7

Discussion

7.1 Summary

Mechanistic models combined with MEMs are practical tools to consider cell-to-cell and patient-to-patient variability while at the same time giving us insights into the biological systems. As shown in this thesis, the models often depend on unknown parameters, which need to be estimated using optimization methods, which minimize the difference between the simulated and experimental data. As in MEMs, different methods are needed to compute the likelihood for different data types. Besides, parameter estimation for the covariance matrix is also challenging. In this thesis, we developed such methods to address the specific challenges as posed in Section 1.2, in the context of ordinary differential equation (ODE) models, and MEMs.

In Chapter 3, we developed a joint likelihood function to integrate three different data types, namely the TL data (including time-to-event data), PA data, and SH data. We used gradient-based optimization method to estimate parameters for different combinations of the subsets of data. By applying the method to a conversion reaction model with artificial data and a cell apoptosis pathway model with experimental data, we show that for the pathway model, none of the data types can be predicted by the other included data types. Furthermore, by including all data types, the parameter identifiability can be improved.

In Chapter 4, we developed a method based on the modified Cramér-von Mises Distance (CMD) to discretely approximate Gaussian distributions. We investigated the method regarding its scalability of computation time. Furthermore, we compared the method with Monte Carlo (MC) sampling methods and sigma point (SP) methods using four test problems. Results show that the method based on the modified Cramér-von Mises Distance can reliably approximate Gaussian distributions with a smaller number of samples compared with both MC and quasi-Monte Carlo methods. Besides, the approximation accuracy can be flexibly controlled by changing the number of CMD points, which is not possible with SP methods.

In Chapter 5, we introduced a Lie algebraic method to parametrize the covariance matrix of MEMs. For comparison, we implemented it together with the matrix logarithm, the (upper) Cholesky decomposition, the spherical, and the Givens parametrization methods in a group-developed MatLab toolbox SPToolbox. We compared the above-mentioned methods using five

forward benchmark problems and one reverse problem using JAK/STAT pathway model. Results show that for the benchmark models, our newly introduced Lie algebraic method performs similarly well as the Givens parametrization method. Regarding the parameter estimation problem using JAK/STAT pathway model, the Lie algebraic method outperforms all the other parametrization methods regarding both fitting results and the relative error of estimated parameters.

In Chapter 6, we showed how MEMs can be used to address patient-to-patient heterogeneity using single-patient time-lapse (TL) data. We used MEMs to model the chemotherapy process regarding the blast cell in the bone marrow and ctDNA in the peripheral blood for 10 patients with 2 different AML-related mutations (*NPM1* and *IDH2*). Results show that a three-step model can be used to fit both population and single-patient data. Furthermore, the blast cell number of an excluded single patient can be predicted using the ctDNA data combined with the population data estimated using the other patients.

Taking a more general perspective, our contributions can be summarized as improving the usability of MEMs in the field of systems biology: Firstly, we developed a method to integrate different data types, in order to combine different information sources to get as much insight into the biological system as possible. We showed this integration is not only necessary for data fitting but also can improve the model parameter identifiability. Secondly, we developed and introduced methods to improve the parameter estimation accuracy, robustness, and efficiency regarding the parameters in the covariance matrices of the random effects, and for PA and SH data. Thirdly, we showed the potential of using MEMs to describe the patient-to-patient variable and predict the further relapse situation of the AML patients and the clinical data, which is difficult to measure or cannot be measured as frequently as needed in real clinical trials.

7.2 Outlook

We discussed the results, implications, and possible extensions of the specific topics presented in this thesis already in Section 3.4, 4.5 and 5.5, 6.4, see these sections for details. Taking a more general perspective, ideas for further research include generalization, benchmarking, and integration of the methods mentioned in this thesis. Furthermore, larger datasets integrating multiple types of data are also expected.

All the methods mentioned in this thesis, e.g. in Chapter 3, 4, 5 and 6 are based on the assumption of normal or log-normal distributed random effects and noise. In systems biology, also other distributions are commonly used, for instance, Laplacian distribution, which is more stable for outliers.

Likelihood function for TL data in MEMs is introduced in Chapter 3 using a first order conditional estimation method. Furthermore, we use this method in the application examples in Chapter 5 and 6. Besides this, also other methods exist for likelihood estimation, e.g. SEAM, alternating approximation, Laplacian approximation, importance sampling and Gaussian quadrature. Therefore, further investigation to compare these methods regarding accuracy or computation efficiency is needed. Furthermore, it is also expected that these methods to be included in the corresponding toolbox.

In this thesis, the developed and introduced methods are applied on either test or application examples. However, in each chapter, we used parts of the methods separately to test different properties. Therefore, the performance of integrating all the methods in one model is still unknown, i.e. integrating different data types, using the CMD method for PA and SH data, while estimating the covariance matrix using the Lie algebraic approach.

To test the performance of our proposed methods, we used different benchmark models in this thesis, from the simple quadratic model in Chapter 4 to the complex apoptosis pathway model as in 3. Consider the different methods of MEMs considered or not considered in this thesis, e.g. methods to compute likelihood for different data types, sampling methods to approximate the random effect distributions, and methods to parameterize the covariance matrices, benchmark models to test the integration of all theses methods or subsets of the methods are required.

Appendix

Here we show the derivatives of the outputs of all three models used in Chapter 6 with respect to the model parameters.

- One-step model

$$\frac{dy}{dc_0^{pb}} = \alpha \exp(-\gamma_{deg} t)$$

$$\frac{dy}{d\alpha} = c_0^{pb} \exp(-\gamma_{deg} t)$$

$$\frac{dy}{d\gamma_{deg}} = -\alpha c_0^{pb} t \exp(-\gamma_{deg} t)$$

- Two-step model

$$\frac{dy_2}{dB_0} = \exp(-\gamma_t t)$$

$$\frac{dy_2}{d\gamma_t} = -B_0 t \exp(-\gamma_t t)$$

$$\frac{dy_1}{dB_0} = \begin{cases} \alpha\gamma_t t \exp(-\gamma_{deg} t), & \text{if} \quad \gamma_{deg} = \gamma_t = \gamma \\ \alpha(\frac{\gamma_t}{\gamma_{deg}-\gamma_t}\exp(-\gamma_t t) - \frac{\gamma_t}{\gamma_{deg}-\gamma_t}\exp(-\gamma_{deg} t)), & \text{if} \quad \gamma_{deg} \neq \gamma_t \end{cases}$$

$$\frac{dy_1}{dc_0^{pb}} = \begin{cases} \alpha \exp(-\gamma_{deg} t), & \text{if} \quad \gamma_{deg} = \gamma_t = \gamma \\ \alpha \exp(\gamma_{deg} t), & \text{if} \quad \gamma_{deg} \neq \gamma_t \end{cases}$$

$$\frac{dy_1}{d\alpha} = \begin{cases} \gamma B_0 t \exp(-\gamma t) + c_0^{pb}\exp(-\gamma t), & \text{if} \quad \gamma_{deg} = \gamma_t = \gamma \\ \frac{\gamma_t B_0}{\gamma_{deg}-\gamma_t}\exp(-\gamma_t t) + (c_0^{pb} - \frac{\gamma_t B_0}{\gamma_{deg}-\gamma_t})\exp(-\gamma_{deg} t), & \text{if} \quad \gamma_{deg} \neq \gamma_t \end{cases}$$

$$\frac{dy_1}{d\gamma_t} = \begin{cases} \alpha B_0 t \exp(-\gamma_{deg} t), & \text{if} \quad \gamma_{deg} = \gamma_t = \gamma \\ \alpha(-\frac{\gamma_t B_0}{\gamma_{deg}-\gamma_t} t \exp(-\gamma_t t) + \frac{\gamma_{deg}}{(\gamma_{deg}-\gamma_t)^2}\exp(-\gamma_t t) & \\ -\frac{\gamma_{deg}}{(\gamma_{deg}-\gamma_t)^2}\exp(-\gamma_{deg} t)), & \text{if} \quad \gamma_{deg} \neq \gamma_t \end{cases}$$

$$\frac{dy_1}{d\gamma_{deg}} = \begin{cases} \alpha(-\gamma_t B_0 t^2 \exp(-\gamma_{deg} t) - c_0^{pb} t \exp(-\gamma_{deg} t)), & \text{if} \quad \gamma_{deg} = \gamma_t = \gamma \\ \alpha(\frac{\gamma_t}{(\gamma_{deg}-\gamma_t)^2}\exp(-\gamma_t t) + \frac{\gamma_t}{\gamma_{deg}-\gamma_t} t \exp(-\gamma_{deg} t) & \\ -\frac{\gamma_t}{(\gamma_{deg}-\gamma_t)^2}\exp(-\gamma_{deg} t)), & \text{if} \quad \gamma_{deg} \neq \gamma_t \end{cases}$$

- Three-step model

$$\frac{dy_2}{dB_0} = \exp(-\gamma_d t)$$

$$\frac{dy_2}{d\gamma_t} = -B_0 t \exp(-\gamma_d t)$$

$$\frac{dy_1}{dB_0} = \begin{cases} \frac{1}{2}\alpha\gamma^2 t^2 \exp(-\gamma t), & \text{if} \quad \gamma_{deg} = \gamma_t = \gamma_d = \gamma \\ \alpha((\frac{\gamma^2}{\gamma_{deg}-\gamma} t - \frac{\gamma^2}{(\gamma_{deg}-\gamma)^2}) \exp(-\gamma t) & \\ +(\frac{\gamma^2}{(\gamma_{deg}-\gamma)^2}) \exp(-\gamma_{deg} t)), & \text{if} \quad \gamma_{deg} \neq \gamma_t, \gamma_t = \gamma_d = \gamma \\ \alpha(\frac{\gamma_1\gamma_d}{(\gamma_1-\gamma_d)^2} \exp(\gamma_d t) & \\ +\frac{\gamma_d}{\gamma_1-\gamma_d} t \exp(-\gamma_1 t) - \frac{\gamma_1\gamma_d}{(\gamma_1-\gamma_d)^2} \exp(-\gamma_1 t)), & \text{if} \quad \gamma_{deg} = \gamma_t = \gamma_1, \gamma_t \neq \gamma_d \\ \alpha(\frac{\gamma_t\gamma_2}{\gamma_t-\gamma_2} t \exp(-\gamma_2 t) + \frac{\gamma_2}{(\gamma_2-\gamma_t)2} \exp(-\gamma_t t) & \\ -\frac{\gamma_2}{(\gamma_2-\gamma_t)^2} \exp(-\gamma_2 t)), & \text{if} \quad \gamma_{deg} = \gamma_d = \gamma_2, \gamma_d \neq \gamma_t \\ \alpha(\frac{\gamma_t\gamma_d}{(\gamma_{deg}-\gamma_d)(\gamma_t-\gamma_d)} \exp(-\gamma_d t) & \\ -\frac{\gamma_t\gamma_d}{(\gamma_{deg}-\gamma_t)(\gamma_t-\gamma_d)} \exp(-\gamma_t t) & \\ +(\frac{\gamma_t\gamma_d}{(\gamma_t-\gamma_d)(\gamma_{deg}-\gamma_d)} + \frac{\gamma_t\gamma_d}{(\gamma_{deg}-\gamma_t)(\gamma_t-\gamma_d)}) \exp(-\gamma_{deg} t)), & \text{if} \quad \gamma_{deg} \neq \gamma_t \neq \gamma_d \end{cases}$$

$$\frac{dy_1}{dc_0^{pb}} = \begin{cases} \alpha \exp(-\gamma t), & \text{if} \quad \gamma_{deg} = \gamma_t = \gamma_d = \gamma \\ \alpha \exp(-\gamma_{deg} t), & \text{if} \quad \gamma_{deg} \neq \gamma_t, \gamma_t = \gamma_d = \gamma \\ \alpha \exp(-\gamma_1 t), & \text{if} \quad \gamma_{deg} = \gamma_t = \gamma_1, \gamma_t \neq \gamma_d \\ \alpha \exp(-\gamma_2 t), & \text{if} \quad \gamma_{deg} = \gamma_d = \gamma_2, \gamma_d \neq \gamma_t \\ \alpha \exp(-\gamma_{deg} t), & \text{if} \quad \gamma_{deg} \neq \gamma_t \neq \gamma_d \end{cases}$$

$$\frac{dy_1}{dc_0^{bm}} = \begin{cases} 0, & \text{if} \quad \gamma_{deg} = \gamma_t = \gamma_d = \gamma \\ \alpha(\frac{\gamma}{\gamma_{deg}-\gamma}\exp(-\gamma t) - \frac{\gamma}{\gamma_{deg}-\gamma}\exp(-\gamma_{deg} t)), & \text{if} \quad \gamma_{deg} \neq \gamma_t, \gamma_t = \gamma_d = \gamma \\ \alpha t \exp(-\gamma_1 t), & \text{if} \quad \gamma_{deg} = \gamma_t = \gamma_1, \gamma_t \neq \gamma_d \\ \alpha(\frac{1}{\gamma_2-\gamma_t}\exp(-\gamma_t t) - \frac{1}{\gamma_2-\gamma_t}\exp(-\gamma_2 t)), & \text{if} \quad \gamma_{deg} = \gamma_d = \gamma_2, \gamma_d \neq \gamma_t \\ \alpha(\frac{(\gamma_t-\gamma_d)\gamma_t}{(\gamma_{deg}-\gamma_t)(\gamma_t-\gamma_d)}\exp(-\gamma_t t) - \frac{\gamma_t}{\gamma_{deg}-\gamma_t})\exp(-\gamma_{deg} t), & \text{if} \quad \gamma_{deg} \neq \gamma_t \neq \gamma_d \end{cases}$$

$$\frac{dy_1}{d\alpha} = \begin{cases} (\frac{1}{2}\gamma^2 B_0 t^2 + \gamma c_0^{bm}) t \exp(-\gamma t) + (c_0^{pb} - \gamma c_0^{bm})\exp(-\gamma t), & \text{if} \quad \gamma_{deg} = \gamma_t = \gamma_d = \gamma \\ (\frac{\gamma^2 B_0}{\gamma_{deg}-\gamma} t - \frac{\gamma^2 B_0}{(\gamma_{deg}-\gamma)^2} + \frac{\gamma c_0^{bm}}{\gamma_{deg}-\gamma})\exp(-\gamma t) \\ +(c_0^{pb} - \frac{\gamma^2 B_0}{(\gamma_{deg}-\gamma)^2} - \frac{\gamma c_0^{bm}}{\gamma_{deg}-\gamma})\exp(-\gamma_{deg} t), & \text{if} \quad \gamma_{deg} \neq \gamma_t, \gamma_t = \gamma_d = \gamma \\ \frac{\gamma_1\gamma_d B_0}{(\gamma_1-\gamma_d)^2}\exp(\gamma_d t) + (c_0^{bm} - \frac{\gamma_d B_0}{\gamma_1-\gamma_d}) t \exp(-\gamma_1 t) \\ +(c_0^{pb} - \frac{\gamma_1\gamma_d B_0}{(\gamma_1-\gamma_d)^2})\exp(-\gamma_1 t), & \text{if} \quad \gamma_{deg} = \gamma_t = \gamma_1, \gamma_t \neq \gamma_d \\ \frac{\gamma_t\gamma_2 B_0}{\gamma_t-\gamma_2} t \exp(-\gamma_2 t) + (\frac{c_0^{bm}}{\gamma_2-\gamma_t} + \frac{\gamma_2 B_0}{(\gamma_2-\gamma_t)2})\exp(-\gamma_t t) \\ +(c_0^{pb} - \frac{c_0^{bm}}{\gamma_2-\gamma_t} - \frac{\gamma_2 B_0}{(\gamma_2-\gamma_t)^2})\exp(-\gamma_2 t), & \text{if} \quad \gamma_{deg} = \gamma_d = \gamma_2, \gamma_d \neq \gamma_t \\ \frac{\gamma_t\gamma_d B_0}{(\gamma_{deg}-\gamma_d)(\gamma_t-\gamma_d)}\exp(-\gamma_d t) \\ +\frac{(\gamma_t-\gamma_d)\gamma_t c_0^{bm} - \gamma_t\gamma_d B_0}{(\gamma_{deg}-\gamma_t)(\gamma_t-\gamma_d)}\exp(-\gamma_t t) \\ +(c_0^{pb} + \frac{\gamma_t\gamma_d B_0}{(\gamma_t-\gamma_d)(\gamma_{deg}-\gamma_d)} - \frac{\gamma_t c_0^{bm}}{\gamma_{deg}-\gamma_t} \\ +\frac{\gamma_t\gamma_d B_0}{(\gamma_{deg}-\gamma_t)(\gamma_t-\gamma_d)})\exp(-\gamma_{deg} t), & \text{if} \quad \gamma_{deg} \neq \gamma_t \neq \gamma_d \end{cases}$$

$$\frac{dy_1}{d\gamma_t} = \begin{cases} \alpha(\frac{1}{2}\gamma B_0 t^2 + c_0^{bm} t - \frac{1}{3}\gamma^2 B_0 t^3 - \frac{1}{2}c_0^{bm} t^3)\exp(-\gamma t), & \text{if} \quad \gamma_{deg} = \gamma_t = \gamma_d = \gamma \\ \alpha(\frac{2}{(\gamma_{deg}-\gamma)^3} + \frac{\gamma\gamma_{deg}B_0 t - 1}{(\gamma_{deg}-\gamma)^2} \\ -\frac{\gamma c_0^{bm} t - 0.5\gamma^2 B_0 t^2 + c_0^{bm}}{\gamma_{deg}-\gamma})\exp(-\gamma t), & \text{if} \quad \gamma_{deg} \neq \gamma_t, \gamma_t = \gamma_d = \gamma \\ \alpha(\frac{-\gamma_d^2 B_0}{(\gamma_1-\gamma_d)^3}\exp(-\gamma_d t) \\ +(c_0^{bm} t + \frac{\gamma_1\gamma_d B_0 t^2}{2(\gamma_1-\gamma_d)} - \frac{1}{2}\gamma_1 c_0^{bm} t^2)\exp(-\gamma_1 t)), & \text{if} \quad \gamma_{deg} = \gamma_t = \gamma_1, \gamma_t \neq \gamma_d \\ \alpha(\frac{-\gamma_2^2 B_0}{(\gamma_t-\gamma_2)^2} t\exp(-\gamma_2 t) \\ +(\frac{c_0^{bm}\gamma_2 - \gamma_2\gamma_t B_0 t + \gamma_2\gamma_t c_0^{bm} t + \gamma_t^2 c_0^{bm} t}{(\gamma_2-\gamma_t)^2} \\ +\frac{2\gamma_2\gamma_t B_0}{(\gamma_2-\gamma_t)^3})\exp(-\gamma_t t)), & \text{if} \quad \gamma_{deg} = \gamma_d = \gamma_2, \gamma_d \neq \gamma_t \\ \alpha(-\frac{\gamma_d^2 B_0}{(\gamma_{deg}-\gamma_d)(\gamma_t-\gamma_d)^2}\exp(-\gamma_d t) \\ +(\frac{c_0^{bm}\gamma_t}{(\gamma_{deg}-\gamma_t)^2(\gamma_t-\gamma_d)} - \frac{\gamma_d^2 B_0}{(\gamma_t-\gamma_d)^2(\gamma_{deg}-\gamma_d)} \\ +\frac{\gamma_d\gamma_t^2 B_0 - \gamma_d^2\gamma_{deg}B_0}{(\gamma_{deg}-\gamma_t)^2(\gamma_t-\gamma_d)^2})\exp(-\gamma_{deg} t) \\ +(\frac{\gamma_{deg}c_0^{bm}}{(\gamma_{deg}-\gamma_t)^2} + \frac{\gamma_{deg}\gamma_d^2 B_0 - \gamma_d B_0\gamma_t^2}{(\gamma_{deg}-\gamma_t)^2} \\ +\frac{\gamma_t^2 c_0^{bm} - \gamma_t\gamma_d c_0^{bm} - \gamma_t\gamma_d B_0}{(\gamma_{deg}-\gamma_t)(\gamma_t-\gamma_d)} t)\exp(-\gamma_t t)), & \text{if} \quad \gamma_{deg} \neq \gamma_t \neq \gamma_d \end{cases}$$

$$\frac{dy_1}{d\gamma_d} = \begin{cases} \alpha(\frac{1}{2}\gamma B_0 t^2 - \frac{1}{6}\gamma^2 B_0 t^3)\exp(-\gamma t), & \text{if} \quad \gamma_{deg} = \gamma_t = \gamma_d = \gamma \\ \alpha(\frac{\gamma\gamma_{deg}B_0 t - 1}{(\gamma_{deg}-\gamma)^2} - \frac{0.5\gamma^2 B_0 t^2}{\gamma_{deg}-\gamma} - \frac{1}{(\gamma_{deg}-\gamma)^3})\exp(-\gamma t), & \text{if} \quad \gamma_{deg} \neq \gamma_t, \gamma_t = \gamma_d = \gamma \\ \alpha((\frac{\gamma_1 B_0}{(\gamma_1-\gamma_d)^2} - \frac{\gamma_d B_0 t}{\gamma_1-\gamma_d})\exp(-\gamma_d t) \\ -\frac{\gamma_1 B_0}{(\gamma_1-\gamma_d)^2}\exp(-\gamma_1 t)), & \text{if} \quad \gamma_{deg} = \gamma_t = \gamma_1, \gamma_t \neq \gamma_d \\ \alpha(\gamma_t B_0(\frac{\gamma_t t}{\gamma_t-\gamma_2} - \frac{\gamma_2 t^2}{2(\gamma_t-\gamma_2)})\exp(-\gamma_2 t) \\ +\frac{\gamma_t^2 B_0}{(\gamma_t-\gamma_2)^3}\exp(-\gamma_2 t)), & \text{if} \quad \gamma_{deg} = \gamma_d = \gamma_2, \gamma_d \neq \gamma_t \\ \alpha((\frac{\gamma_{deg}\gamma_t^2 B_0 - \gamma_d^2\gamma_t B_0}{(\gamma_t-\gamma_d)^2(\gamma_{deg}-\gamma_d)^2} - \frac{\gamma_d\gamma_t B_0 t}{(\gamma_{deg}-\gamma_d)(\gamma_t-\gamma_d)})\exp(-\gamma_d t) \\ -\frac{\gamma_t^2 B_0}{(\gamma_{deg}-\gamma_t)(\gamma_t-\gamma_d)^2}\exp(-\gamma_t t) \\ +(\frac{\gamma_{deg}\gamma_t^2 B_0 - \gamma_d^2\gamma_t B_0}{(\gamma_t-\gamma_d)^2(\gamma_{deg}-\gamma_d)^2} + \frac{\gamma_t^2 B_0}{(\gamma_{deg}-\gamma_t)(\gamma_t-\gamma_d)^2})\exp(-\gamma_{deg} t)), & \text{if} \quad \gamma_{deg} \neq \gamma_t \neq \gamma_d \end{cases}$$

$$\frac{dy_1}{d\gamma_{deg}} = \begin{cases} \alpha(-\frac{1}{6}\gamma^2 B_0 t^3 + \frac{3}{2}\gamma c_0^{bm} t^2 + c_0^{pb} t + \gamma c_0^{bm} t)\exp(-\gamma t), & \text{if} \quad \gamma_{deg} = \gamma_t = \gamma_d = \gamma \\ \alpha(-\frac{1}{6}\gamma^2 B_0 t^3 + \frac{3}{2}\gamma c_0^{bm} t^2 + \gamma c_0^{bm} t + c_0^{pb} t)\exp(-\gamma_{deg} t), & \text{if} \quad \gamma_{deg} \neq \gamma_t, \gamma_t = \gamma_d = \gamma \\ \alpha(-\frac{\gamma_d B_0}{(\gamma_1-\gamma_d)^2}\exp(-\gamma_d t) + (c_0^{bm} - \frac{\gamma_d B_0}{\gamma_1-\gamma_d})t\exp(-\gamma_1 t)), & \text{if} \quad \gamma_{deg} = \gamma_t = \gamma_1, \gamma_t \neq \gamma_d \\ \alpha(-\frac{\gamma_2 B_0}{(\gamma_t-\gamma_2)^2}\exp((\gamma_t - 2\gamma_2)t) & \\ +(c_0^{bm} - \frac{\gamma_2 B_0}{\gamma_t-\gamma_2})t\exp(-\gamma_2 t)), & \text{if} \quad \gamma_{deg} = \gamma_d = \gamma_2, \gamma_d \neq \gamma_t \\ \alpha(-\frac{\gamma_t\gamma_d B_0}{(\gamma_{deg}-\gamma_d)^2(\gamma_t-\gamma_d)}\exp(-\gamma_d t) & \\ -\frac{\gamma_t^2 c_0^{bm} - \gamma_d\gamma_t c_0^{bm} - \gamma_t\gamma_d B_0}{(\gamma_{deg}-\gamma_t)^2(\gamma_t-\gamma_d)}\exp(-\gamma_t t) & \\ +(-c_0^{pb} t - \frac{\gamma_t\gamma_d B_0}{(\gamma_t-\gamma_d)(\gamma_{deg}-\gamma_d)^2} & \\ -\frac{\gamma_t\gamma_d B_0 t}{(\gamma_t-\gamma_d)(\gamma_{deg}-\gamma_d)} + \frac{\gamma_t c_0^{bm}}{(\gamma_{deg}-\gamma_t)^2} + \frac{\gamma_t c_0^{bm} t}{\gamma_{deg}-\gamma_t} & \\ -\frac{\gamma_t\gamma_d B_0}{(\gamma_{deg}-\gamma_t)^2(\gamma_t-\gamma_d)} - \frac{\gamma_t\gamma_d B_0 t}{(\gamma_{deg}-\gamma_t)(\gamma_t-\gamma_d)})\exp(-\gamma_{deg} t)), & \text{if} \quad \gamma_{deg} \neq \gamma_t \neq \gamma_d \end{cases}$$

Bibliography

L. Adlung, P. Stapor, C. Tönsing, L. Schmiester, L. E. Schwarzmüller, L. Postawa, D. Wang, J. Timmer, U. Klingmüller, J. Hasenauer, et al. Cell-to-cell variability in JAK2/STAT5 pathway components and cytoplasmic volumes defines survival threshold in erythroid progenitor cells. *Cell Reports*, 36(6):109507, 2021.

E. Afenya. Acute leukemia and chemotherapy: a modeling viewpoint. *Mathematical Biosciences*, 138(2):79–100, 1996.

J. G. Albeck, G. B. Mills, and J. S. Brugge. Frequency-modulated pulses of ERK activity transmit quantitative proliferation signals. *Molecular Cell*, 49(2):249–261, 2013.

S. Alig, C. W. Macaulay, D. M. Kurtz, U. Dührsen, A. Hüttmann, C. Schmitz, M. C. Jin, B. J. Sworder, A. Garofalo, M. Shahrokh Esfahani, et al. Short diagnosis-to-treatment interval is associated with higher circulating tumor DNA levels in diffuse large B-cell lymphoma. *Journal of Clinical Oncology*, 39(23):2605–2616, 2021.

C. Alix-Panabières, H. Schwarzenbach, K. Pantel, et al. Circulating tumor cells and circulating tumor DNA. *Annual Review of Medicine*, 63(1):199–215, 2012.

J. Almquist, L. Bendrioua, C. B. Adiels, M. Goksör, S. Hohmann, and M. Jirstrand. A nonlinear mixed effects approach for modeling the cell-to-cell variability of Mig1 dynamics in yeast. *PLoS ONE*, 10(4):1–32, 2015.

M. Ancukiewicz, T. A. Russell, J. Otoole, M. Specht, M. Singer, A. Kelada, C. D. Murphy, J. Pogachar, V. Gioioso, M. Patel, et al. Standardized method for quantification of developing lymphedema in patients treated for breast cancer. *International Journal of Radiation Oncology* Biology* Physics*, 79(5):1436–1443, 2011.

E. D. Andersen, J. Gondzio, C. Mészáros, X. Xu, et al. *Implementation of interior point methods for large scale linear programming.* HEC/Université de Geneve, 1996.

J. M. Astle, M. L. Xu, T. Friedman, and E. Brown. Limitations of poor bone marrow aspirations (for an accurate diagnosis) despite the multimodal analytical era: a longitudinal retrospective study. *American Journal of Hematology*, 92(10):E600–E602, 2017.

S. Avanzini, D. M. Kurtz, J. J. Chabon, E. J. Moding, S. S. Hori, S. S. Gambhir, A. A. Alizadeh, M. Diehn, and J. G. Reiter. A mathematical model of ctDNA shedding predicts tumor detection size. *Science Advances*, 6(50):eabc4308, 2020.

R. E. Baker, J.-M. Pena, J. Jayamohan, and A. Jérusalem. Mechanistic models versus machine learning, a fight worth fighting for the biological community? *Biology Letters*, 14(5):20170660, 2018.

C. Barnett et al. The unavoidable error in the differential count of the leukocytes of the blood. *The Journal of Clinical Investigation*, 12(1):77–85, 1933.

T. Bastogne, A. Samson, S. Mézières-Wantz, P. Vallois, S. Pinel, and M. Barberi-Heyob. System identification of tumor growth described by a mixed effects model. In *Proceedings of 15th IFAC Symposium on System Identification*, volume 15, pages 1277–1282. M. Basseville, IFAC-PapersOnline, 2009.

D. M. Bates and D. G. Watts. *Nonlinear regression analysis and its applications*. Wiley, 1988.

S. Beal and L. Sheiner. NONMEM users guide-Part VII: conditional estimation methods. *San Francisco: Division of Clinical Pharmacology, University of California at San Francisco*, 1992.

J. S. Blachly, R. B. Walter, and C. S. Hourigan. The present and future of measurable residual disease testing in acute myeloid leukemia. *Haematologica*, 107(12):2810–2822, 2022.

B. Bodenmiller, E. R. Zunder, R. Finck, T. J. Chen, E. S. Savig, R. V. Bruggner, E. F. Simonds, S. C. Bendall, K. Sachs, P. O. Krutzik, and G. P. Nolan. Multiplexed mass cytometry profiling of cellular states perturbed by small-molecule regulators. *Nature Biotechnology*, 30(9):858–867, 2012.

L. D. Brown, T. T. Cai, and A. DasGupta. Confidence intervals for a binomial proportion and asymptotic expansions. *The Annals of Statistics*, 30(1):160–201, 2002.

F. Buettner, K. N. Natarajan, F. P. Casale, V. Proserpio, A. Scialdone, F. J. Theis, S. A. Teichmann, J. C. Marioni, and O. Stegle. Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nature Biotechnology*, 33(2):155–160, 2015.

R. E. Caflisch. Monte Carlo and quasi-Monte Carlo methods. *Acta Numerica*, 7:1–49, 1998.

A. C. Charalampidis and G. P. Papavassilopoulos. Development and numerical investigation of new non-linear Kalman filter variants. *IET Control Theory & Applications*, 5:1155–1166, 2011.

E. Charton, B. Cuer, F. Cottone, F. Efficace, C. Touraine, Z. Hamidou, F. Fiteni, F. Bonnetain, M.-C. Woronoff-Lemsi, C. Bascoul-Mollevi, et al. Time to deterioration in cancer randomized clinical trials for patient-reported outcomes data: a systematic review. *Quality of Life Research*, 29:867–878, 2020.

R. Cheong, S. Paliwal, and A. Levchenko. Models at the single cell level. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 2(1):34–48, 2010.

O.-T. Chis, J. R. Banga, and E. Balsa-Canto. Structural identifiability of systems biology models: A critical comparison of methods. *PLoS ONE*, 6(11):e27755, 2011.

E. Cinquemani. Identifiability and reconstruction of biochemical reaction networks from population snapshot data. *Processes*, 6(9):136, 2018.

K. Clement. An interior point algorithm for weighted least absolute value power system state estimation. *91 WM 235-2 PWRS, New York*, 1991.

P. Coates, S. Vyakrnam, E. Eady, C. Jones, J. Cove, and W. Cunliffe. Prevalence of antibiotic-resistant propionibacteria on the skin of acne patients: 10-year surveillance data and snapshot distribution study. *British Journal of Dermatology*, 146(5):840–848, 2002.

C. Cohen-Saidon, A. A. Cohen, A. Sigal, Y. Liron, and U. Alon. Dynamics and variability of ERK2 response to EGF in individual living cells. *Molecular Cell*, 36(5):885–893, 2009.

R. Colmenares, N. Álvarez, S. Barrio, J. Martínez-López, and R. Ayala. The minimal residual disease using liquid biopsies in hematological malignancies. *Cancers*, 14(5):1310, 2022.

J. Cosgrove, L. S. Hustin, R. J. de Boer, and L. Perié. Hematopoiesis in numbers. *Trends in Immunology*, 42(12):1100–1112, 2021.

P. Daran-Lapujade, J.-M. Daran, A. J. van Maris, J. H. de Winde, and J. T. Pronk. Chemostat-based micro-array analysis in baker's yeast. *Advances in Microbial Physiology*, 54:257–417, 2008.

H. M. Davey and D. B. Kell. Flow cytometry and cell sorting of heterogeneous microbial populations: the importance of single-cell analyses. *Microbiological Reviews*, 60(4):641–696, 1996.

M. Davidian and A. R. Gallant. Smooth nonparametric maximum likelihood estimation for population pharmacokinetics, with application to quinidine. *Journal of Pharmacokinetics and Biopharmaceutics*, 20(5):529–556, 1992.

A. Degasperi, D. Fey, and B. N. Kholodenko. Performance of objective functions and optimisation procedures for parameter estimation in system biology models. *npj Systems Biology and Applications*, 3(1):20, 2017.

M. A. Detry and Y. Ma. Analyzing repeated measurements using mixed models. *Jama*, 315(4): 407–408, 2016.

L. Dharmarajan, H.-M. Kaltenbach, F. Rudolf, and J. Stelling. A simple and flexible computational framework for inferring sources of heterogeneity from single-cell dynamics. *Cell Systems*, 8(1):15–26, 2019.

H. Döhner, A. H. Wei, F. R. Appelbaum, C. Craddock, C. D. DiNardo, H. Dombret, B. L. Ebert, P. Fenaux, L. A. Godley, R. P. Hasserjian, et al. Diagnosis and management of AML in adults: 2022 recommendations from an international expert panel on behalf of the ELN. *Blood, The Journal of the American Society of Hematology*, 140(12):1345–1377, 2022.

F. Eduati, V. Doldàn-Martelli, B. Klinger, T. Cokelaer, A. Sieber, F. Kogera, M. Dorel, M. J. Garnett, N. Blüthgen, and J. Saez-Rodriguez. Drug resistance mechanisms in colorectal cancer dissected with cell type-specific dynamic logic models. *Cancer Research*, 77(12):3364–3375, 2017.

M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain. Stochastic gene expression in a single cell. *Science*, 297(5584):1183–1186, 2002.

M. Endo, S. Takahashi, N. Araki, H. Sugiura, T. Ueda, T. Yonemoto, M. Takahashi, H. Morioka, H. Hiraga, T. Hiruma, et al. Time lapse analysis of tumor response in patients with soft tissue sarcoma treated with trabectedin: a pooled analysis of two phase II clinical trials. *Cancer Medicine*, 9(11):3656–3667, 2020.

E. H. Estey, S. Faderl, and H. M. Kantarjian. *Hematologic malignancies: acute leukemias*. Springer, 2008.

M. M. Falco, M. Peña-Chilet, C. Loucera, M. R. Hidalgo, and J. Dopazo. Mechanistic models of signaling pathways deconvolute the glioblastoma single-cell functional landscape. *NAR Cancer*, 2(2):zcaa011, 2020.

I. Fernández-Miranda, L. Pedrosa, M. Llanos, F. F. Franco, S. Gómez, P. Martín-Acosta, F. R. García-Arroyo, J. Gumá, B. Horcajo, A. K. Ballesteros, et al. Monitoring of circulating tumor DNA predicts response to treatment and early progression in follicular lymphoma: results of a prospective pilot study. *Clinical Cancer Research*, 29(1):209–220, 2023.

D. Fey, M. Halasz, D. Dreidax, S. P. Kennedy, J. F. Hastings, N. Rauch, A. G. Munoz, R. Pilkington, M. Fischer, F. Westermann, W. Kolch, B. N. Kholodenko, and D. R. Croucher. Signaling pathway models as biomarkers: patient-specific simulations of JNK activity predict the survival of neuroblastoma patients. *Science Signaling*, 8(408), 2015.

S. Filippi, C. P. Barnes, P. D. W. Kirk, T. Kudo, K. Kunida, S. S. McMahon, T. Tsuchiya, T. Wada, S. Kuroda, and M. P. Stumpf. Robustness of MEK-ERK dynamics and origins of cell-to-cell variability in MAPK signaling. *Cell Reports*, 15(11):2524–2535, 2016.

R. A. Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A*, 222(594-604):309–368, 1922.

A. P. Frei, F.-A. Bava, E. R. Zunder, E. W. Y. Hsieh, S.-Y. Chen, G. P. Nolan, and P. F. Gherardini. Highly multiplexed simultaneous detection of RNAs and proteins in single cells. *Nature Methods*, 13(3):269–275, 2016.

F. Fröhlich, F. J. Theis, and J. Hasenauer. Uncertainty analysis for non-identifiable dynamical systems: profile likelihoods, bootstrapping and more. In *Computational Methods in Systems Biology: 12th International Conference, CMSB 2014, Manchester, UK, November 17-19, 2014, Proceedings 12*, Lecture Notes in Bioinformatics, pages 61–72. Springer International Publishing Switzerland, 2014.

F. Fröhlich, P. Thomas, A. Kazeroonian, F. J. Theis, R. Grima, and J. Hasenauer. Inference for stochastic chemical kinetics using moment equations and system size expansion. *PLoS Computational Biology*, 12(7):e1005030, 2016.

F. Fröhlich, B. Kaltenbacher, F. J. Theis, and J. Hasenauer. Scalable parameter estimation for genome-scale biochemical reaction networks. *PLoS Computational Biology*, 13(1):e1005331, 2017a.

F. Fröhlich, F. J. Theis, J. O. Rädler, and J. Hasenauer. Parameter estimation for dynamical systems with discrete events and logical operations. *Bioinformatics*, 33(7):1049–1056, 2017b.

F. Fröhlich, T. Kessler, D. Weindl, A. Shadrin, L. Schmiester, H. Hache, A. Muradyan, M. Schuette, J.-H. Lim, M. Heinig, F. Theis, H. Lehrach, C. Wierling, B. Lange, and J. Hasenauer. Efficient parameterization of large-scale mechanistic models enables drug response prediction for cancer cell lines. *Cell Systems*, 7(6):567–579.e6, 2018.

F. Fröhlich, T. Kessler, D. Weindl, A. Shadrin, L. Schmiester, H. Hache, A. Muradyan, M. Schütte, J.-H. Lim, M. Heinig, F. J. Theis, H. Lehrach, C. Wierling, B. Lange, and

J. Hasenauer. Efficient parameter estimation enables the prediction of drug response using a mechanistic pan-cancer pathway model. *Cell Systems*, 7(6):567–579.e6, 2018a.

F. Fröhlich, A. Reiser, L. Fink, D. Woschée, T. Ligon, F. J. Theis, J. O. Rädler, and J. Hasenauer. Multi-experiment nonlinear mixed effect modeling of single-cell translation kinetics after transfection. *npj Systems Biology and Applications*, 5(1), 2018b.

F. Fröhlich, C. Loos, and J. Hasenauer. Scalable inference of ordinary differential equation models of biochemical processes. In *Gene Regulatory Networks: Methods and Protocols*, volume 1883 of *Methods in Molecular Biology*, chapter 16, pages 385–422. Humana Press, 1 edition, 2019.

J. Geweke. Bayesian inference in econometric models using monte carlo integration. *Econometrica: Journal of the Econometric Society*, pages 1317–1339, 1989.

D. Ghosh and A. Vogt. Outliers: an evaluation of methodologies. In *Joint Statistical Meetings*, pages 3455–3460. American Statistical Association San Diego, CA, 2012.

I. Gilitschenski and U. D. Hanebeck. Efficient deterministic dirac mixture approximation of gaussian distributions. *IEEE*, 2013.

A. S. Goldberger. Best linear unbiased prediction in the generalized linear regression model. *Journal of the American Statistical Association*, 57(298):369–375, 1962.

J. Q. Gong and E. A. Sobie. Population-based mechanistic modeling allows for quantitative predictions of drug responses across cell types. *npj Systems Biology and Applications*, 4(1): 1–11, 2018.

F. Greselin and L. Pasquazzi. Asymptotic confidence intervals for a new inequality measure. *Communications in Statistics-Simulation and Computation*, 38(8):1742–1756, 2009.

J. H. Halton. Algorithm 247: radical-inverse quasi-random point sequence. *Communications of the ACM*, 7(12):701–702, 1964.

U. D. Hanebeck and V. Klumpp. Localized cumulative distributions and a multivariate generalization of the Cramér-von mises distance. In *2008 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pages 33–39. IEEE, 2008.

N. Hansen and A. Ostermeier. Adapting arbitrary normal mutation distributions in evolution strategies: the covariance matrix adaptation. In *Proceedings of IEEE International Conference on Evolutionary Computation*, pages 312–317. IEEE, 1996.

L. A. Harris, S. Beik, P. M. Ozawa, L. Jimenez, and A. M. Weaver. Modeling heterogeneous tumor growth dynamics and cell–cell interactions at single-cell and cell-population resolution. *Current Opinion in Systems Biology*, 17:24–34, 2019.

J. Hasenauer, N. Radde, M. Doszczak, P. Scheurich, and F. Allgöwer. Parameter estimation for the CME from noisy binned snapshot data: formulation as maximum likelihood problem. Extended abstract at *Conf. of Stoch. Syst. Biol.*, Monte Verita, Switzerland, 2011a.

J. Hasenauer, S. Waldherr, M. Doszczak, N. Radde, P. Scheurich, and F. Allgöwer. Identification of models of heterogeneous cell populations from population snapshot data. *BMC Bioinformatics*, 12(125), 2011b.

J. Hasenauer, C. Hasenauer, T. Hucho, and F. J. Theis. ODE constrained mixture modelling: a method for unraveling subpopulation structures and dynamics. *PLoS Computational Biology*, 10(7):e1003686, 2014.

H. Hass, C. Kreutz, J. Timmer, and D. Kaschek. Fast integration-based prediction bands for ordinary differential equation models. *Bioinformatics*, 32(8):1204–1210, 2016.

R. Heinrich and S. Schuster. *The regulation of cellular systems.* Springer Science & Business Media, 2012.

A. Heinzel, P. Perco, G. Mayer, R. Oberbauer, A. Lukas, and B. Mayer. From molecular signatures to predictive biomarkers: modeling disease pathophysiology and drug mechanism of action. *Frontiers in Cell and Developmental Biology*, 2:37, 2014.

C. R. Henderson, O. Kempthorne, S. R. Searle, and C. Von Krosigk. The estimation of environmental and genetic trends from records subject to culling. *Biometrics*, 15(2):192–218, 1959.

M. Heuser, S. D. Freeman, G. J. Ossenkoppele, F. Buccisano, C. S. Hourigan, L. L. Ngai, J. M. Tettero, C. Bachas, C. Baer, M.-C. Béné, et al. 2021 update on MRD in acute myeloid leukemia: a consensus document from the European LeukemiaNet MRD Working Party. *Blood*, 138(26): 2753–2767, 2021.

A. C. Hindmarsh, P. N. Brown, K. E. Grant, S. L. Lee, R. Serban, D. E. Shumaker, and C. S. Woodward. SUNDIALS: Suite of Nonlinear and Differential/Algebraic Equation Solvers. *ACM Transactions on Mathematical Software*, 31(3):363–396, 2005.

J. C. Hoffmann, A. Pappa, P. H. Krammer, and I. N. Lavrik. A new C-terminal cleavage product of procaspase-8, p30, defines an alternative pathway of procaspase-8 activation. *Molecular and Cellular Biology*, 29(16):4431–4440, 2009.

S. Holdenrieder, P. Stieber, H. Bodenmüller, M. Busch, J. VON PAWEL, A. Schalhorn, D. Nagel, and D. Seidel. Circulating nucleosomes in serum. *Annals of the New York Academy of Sciences*, 945(1):93–102, 2001.

A. Höllein, M. Meggendorfer, F. Dicker, S. Jeromin, N. Nadarajah, W. Kern, C. Haferlach, and T. Haferlach. NPM1 mutated AML can relapse with wild-type NPM1: persistent clonal hematopoiesis can drive relapse. *Blood Advances*, 2(22):3118–3125, 2018.

J. Ihlow, S. Gross, L. Busack, A. Flörcken, J. Jesse, M. Schwarz, N. R. Neuendorff, A.-C. Von Bruenneck, I. Anagnostopoulos, S. Türkmen, et al. Acute myeloid leukemia: negative prognostic impact of early blast persistence can be in part overcome by a later remission prior to post-induction therapy. *Haematologica*, 107(8):1773, 2022.

J. Intosalmi, K. Nousiainen, H. Ahlfors, and H. Läähdesmäki. Data-driven mechanistic analysis method to reveal dynamically evolving regulatory networks. *Bioinformatics*, 32(12):i288–i296, 2016.

A. Ivey, R. K. Hills, M. A. Simpson, J. V. Jovanovic, A. Gilkes, A. Grech, Y. Patel, N. Bhudia, H. Farah, J. Mason, et al. Assessment of minimal residual disease in standard-risk AML. *New England Journal of Medicine*, 374(5):422–433, 2016.

Y. Ji, J. M. Skierka, J. H. Blommel, B. E. Moore, D. L. VanCuyk, J. K. Bruflat, L. M. Peterson, T. L. Veldhuizen, N. Fadra, S. E. Peterson, et al. Preemptive pharmacogenomic testing for precision medicine: a comprehensive analysis of five actionable pharmacogenomic genes using next-generation DNA sequencing and a customized CYP2D6 genotyping cascade. *The Journal of Molecular Diagnostics*, 18(3):438–445, 2016.

S. J. Julier and J. K. Uhlmann. Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3):401–422, 2004.

S. J. Julier, J. K. Uhlmann, and H. F. Durrant-Whyte. A new approach for filtering nonlinear systems. In *Proceedings of 1995 American Control Conference - ACC'95*, volume 3, pages 1628–1632, 1995.

W. G. Kaelin Jr. The concept of synthetic lethality in the context of anticancer therapy. *Nature Reviews Cancer*, 5(9):689–698, 2005.

S. M. Kallenberger, J. Beaudouin, J. Claus, C. Fischer, P. K. Sorger, S. Legewie, and R. Eils. Intra- and interdimeric caspase-8 self-cleavage controls strength and timing of CD95-induced apoptosis. *Science Signaling*, 7(316):ra23, 2014.

M. Karlsson, D. L. I. Janzén, L. Durrieu, A. Colman-Lerner, M. C. Kjellsson, and G. Cedersund. Nonlinear mixed-effects modelling for single cell estimation: when, why, and how to use it. *BMC Systems Biology*, 9(1):52, 2015.

L. Keller, Y. Belloum, H. Wikman, and K. Pantel. Clinical relevance of blood-based ctDNA analysis: mutation detection and beyond. *British Journal of Cancer*, 124(2):345–358, 2021.

S. Keller, J. Kneissl, V. Grabher-Meier, S. Heindl, J. Hasenauer, D. Maier, J. Mattes, P. Winter, and B. Luber. Evaluation of epidermal growth factor receptor signaling effects in gastric cancer cell lines by detailed motility-focused phenotypic characterization linked with molecular analysis. *BMC Cancer*, 17(1):845, 2017.

J. Kennedy. Particle swarm optimization. In *Encyclopedia of Machine learning*, pages 760–766. Springer, 2011.

J. D. Khoury, E. Solary, O. Abla, Y. Akkari, R. Alaggio, J. F. Apperley, R. Bejar, E. Berti, L. Busque, J. K. Chan, et al. The 5th edition of the World Health Organization classification of haematolymphoid tumours: myeloid and histiocytic/dendritic neoplasms. *Leukemia*, 36(7): 1703–1719, 2022.

J. J. Kim and I. F. Tannock. Repopulation of cancer cells during therapy: an important cause of treatment failure. *Nature Reviews Cancer*, 5(7):516–525, 2005.

S. Kirkpatrick, C. D. Gelatt Jr, and M. P. M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

F. C. Kischkel, S. Hellbardt, I. Behrmann, M. Germer, M. Pawlita, P. H. Krammer, and M. E. Peter. Cytotoxicity-dependent APO-1 (Fas/CD95)-associated proteins form a death-inducing signaling complex (DISC) with the receptor. *The EMBO Journal*, 14(22):5579–5588, 1995.

A. Klimovskaia, S. Ganscha, and M. Claassen. Sparse regression based structure learning of stochastic reaction networks from single cell snapshot time series. *PLoS Computational Biology*, 12(12):e1005234, 2016.

E. Klipp and W. Liebermeister. Mathematical modeling of intracellular signaling pathways. *BMC Neuroscience*, 7:1–16, 2006.

J. Kochmanski, C. Savonen, and A. I. Bernstein. A novel application of mixed effects models for reconciling base-pair resolution 5-methylcytosine and 5-hydroxymethylcytosine data in neuroepigenetics. *Frontiers in Genetics*, 10:801, 2019.

K. Koh, S.-J. Kim, and S. Boyd. An interior-point method for large-scale l1-regularized logistic regression. *Journal of Machine Learning Research*, 8(Jul):1519–1555, 2007.

B. T. Kurien and R. H. Scofield. Western blotting. *Methods*, 38(4):283–293, 2006.

M. Lavielle and L. Aarons. What do we mean by identifiability in mixed effects models? *Journal of Pharmacokinetics and Pharmacodynamics*, 43(1):111–122, 2016.

J. S. Leeder. Who believes they are "just average": informing the treatment of individual patients using population data. *Clinical Pharmacology and Therapeutics*, 106(5):939, 2019.

U. N. Lerner. *Hybrid Bayesian Networks for Reasoning about Complex Systems*. PhD thesis, Stanford University, 2002.

J. Lever, M. Krzywinski, and N. Altman. Model selection and overfitting. *Nature Methods*, 13 (9):703–704, 2016.

X. Li and C.-Y. Wang. From bulk, single-cell to spatial RNA sequencing. *International Journal of Oral Science*, 13(1):1–6, 2021.

T. S. Ligon, F. Fröhlich, O. T. Chiş, J. R. Banga, E. Balsa-Canto, and J. Hasenauer. GenSSI 2.0: multi-experiment structural identifiability analysis of SBML models. *Bioinformatics*, 34 (8):1421–1423, 2018.

M. Lindstrom and D. Bates. Nonlinear mixed effects models for repeated measures data. *Biometrics*, pages 673–687, 1990.

A. Llamosi, A. M. Gonzalez-Vargas, C. Versari, E. Cinquemani, G. Ferrari-Trecate, P. Hersen, and G. Batt. What population reveals about individual cell identity: single-cell parameter estimation of models of gene expression in yeast. *PLoS Computational Biology*, 12(2):1–18, 2016.

C. Loos, S. Krause, and J. Hasenauer. Hierarchical optimization for the efficient parametrization of ODE models. *Bioinformatics*, 34(24):4266–4273, 2018a.

C. Loos, K. Moeller, F. Fröhlich, T. Hucho, and J. Hasenauer. A hierarchical, data-driven approach to modeling single-cell populations predicts latent causes of cell-to-cell variability. *Cell Systems*, 6(5):593–603, 2018b.

K. R. Love, V. Panagiotou, B. Jiang, T. A. Stadheim, and J. C. Love. Integrated single-cell analysis shows Pichia pastoris secretes protein stochastically. *Biotechnology and Bioengineering*, 106(2):319–325, 2010.

C. Maier, C. Loos, and J. Hasenauer. Robust parameter estimation for dynamical systems from outlier-corrupted data. *Bioinformatics*, 33(5):718–725, 2017.

T. Maiwald and J. Timmer. Dynamical modeling and multi-experiment fitting with PottersWheel. *Bioinformatics*, 24(18):2037–2043, 2008.

M. P. Martelli, P. Sportoletti, E. Tiacci, M. F. Martelli, and B. Falini. Mutational landscape of AML with normal cytogenetics: biological and clinical implications. *Blood Reviews*, 27(1): 13–22, 2013.

P. Mendes, S. Hoops, S. Sahle, R. Gauges, J. Dada, and U. Kummer. Computational modeling of biochemical networks using COPASI. In *Systems Biology*, pages 17–59. Springer, 2009.

H. M. Menegaz, J. Y. Ishihara, and G. A. Borges. A new smallest sigma set for the unscented transform and its applications on SLAM. In *Proceedings of IEEE Conference on Decision and Control and European Control Conference*, pages 3172–3177, Dec 2011.

T. Miyashiro and M. Goulian. Single-cell analysis of gene expression by fluorescence microscopy. *Methods in Enzymology*, 423:458–475., 2007.

H. Motulsky and A. Christopoulos. *Fitting models to biological data using linear and nonlinear regression: a practical guide to curve fitting.* GraphPad Software Inc., San Diego CA, 2003.

S. Mueller, S. Holdenrieder, P. Stieber, T. Haferlach, A. Schalhorn, J. Braess, D. Nagel, and D. Seidel. Early prediction of therapy response in patients with acute myeloid leukemia by nucleosomal DNA fragments. *BMC Cancer*, 6(1):1–10, 2006.

R. Nair, A. Salinas-Illarena, and H.-M. Baldauf. New strategies to treat AML: novel insights into AML survival pathways and combination therapies. *Leukemia*, 35(2):299–311, 2021.

S. Nakamura, K. Yokoyama, N. Yusa, M. Ogawa, T. Takei, A. Kobayashi, M. Ito, E. Shimizu, R. Kasajima, Y. Wada, et al. Circulating tumor DNA dynamically predicts response and/or relapse in patients with hematological malignancies. *International Journal of Hematology*, 108 (4):402–410, 2018.

S. Nakamura, K. Yokoyama, E. Shimizu, N. Yusa, K. Kondoh, M. Ogawa, T. Takei, A. Kobayashi, M. Ito, M. Isobe, et al. Prognostic impact of circulating tumor DNA status post–allogeneic hematopoietic stem cell transplantation in AML and MDS. *Blood, the Journal of the American Society of Hematology*, 133(25):2682–2695, 2019.

N. Navin, J. Kendall, J. Troge, P. Andrews, L. Rodgers, J. McIndoo, K. Cook, A. Stepansky, D. Levy, D. Esposito, L. Muthuswamy, A. Krasnitz, W. R. McCombie, J. Hicks, and M. Wigler. Tumour evolution inferred by single-cell sequencing. *Nature*, 472:90 EP –, 2011.

H. Niederreiter. Quasi-Monte Carlo methods and pseudo-random numbers. *Bulletin of the American Mathematical Society*, 84(6):957–1041, November 1978.

L. Ning and T. T. Georgiou. Sparse factor analysis via likelihood and l1-regularization. In *2011 50th IEEE Conference on Decision and Control and European Control Conference*, pages 5188–5192. IEEE, 2011.

J. Nocedal and S. Wright. *Numerical Optimization.* Springer Science & Business Media, 2006.

A. Ocone, L. Haghverdi, N. S. Mueller, and F. J. Theis. Reconstructing gene regulatory dynamics from high-dimensional single-cell snapshot data. *Bioinformatics*, 31(12):i89–i96, 2015.

D. B. Özyurt and P. I. Barton. Cheap second order directional derivatives of stiff ODE embedded functionals. *SIAM Journal on Scientific Computing*, 26(5):1725–1743, 2005.

S. K. Palaniappan, B. M. Gyori, B. Liu, D. Hsu, and P. Thiagarajan. Statistical model checking based calibration and analysis of bio-pathway models. In *International Conference on Computational Methods in Systems Biology*, pages 120–134. Springer, 2013.

S. Palit, C. Heuser, G. P. De Almeida, F. J. Theis, and C. E. Zielinski. Meeting the challenges of high-dimensional single-cell data analysis in immunology. *Frontiers in Immunology*, 10:1515, 2019.

A. C. Palmer and P. K. Sorger. Combination cancer therapy can confer benefit via patient-to-patient variability without drug additivity or synergy. *Cell*, 171(7):1678–1691, 2017.

K. Pantel and C. Alix-Panabières. Liquid biopsy and minimal residual disease—latest advances and implications for cure. *Nature Reviews Clinical Oncology*, 16(7):409–424, 2019.

J. Pascual, G. Attard, F.-C. Bidard, G. Curigliano, L. de Mattos-Arruda, M. Diehn, A. Italiano, J. Lindberg, J. D. Merker, C. Montagut, et al. ESMO recommendations on the use of circulating tumour DNA assays for patients with cancer: a report from the ESMO Precision Medicine Working Group. *Annals of Oncology*, 33(8):750–768, 2022.

R. Pepperkok and J. Ellenberg. High-throughput fluorescence microscopy for systems biology. *Nature Reviews Molecular Cell Biology*, 7(9):690–6, 2006.

J. C. Pinheiro. *Topics in mixed effects models*. Ph.d. thesis, University of Wisconsin, Madison, Madison, USA, 1994.

J. C. Pinheiro and D. M. Bates. Unconstrained parametrizations for variance-covariance matrices. *Statistics and Computing*, 6(3):289–296, 1996.

R.-E. Plessix. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, 167(2):495–503, 2006.

Q. Qiu, P. Hu, X. Qiu, K. W. Govek, P. G. Cámara, and H. Wu. Massively parallel and time-resolved RNA sequencing in single cells with scNT-seq. *Nature Methods*, 17(10):991–1001, 2020.

Raj and A. Van Oudenaarden. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell*, 135(2):216–26, 2008.

A. Raue, C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmüller, and J. Timmer. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(25):1923–1929, 2009.

A. Raue, M. Schilling, J. Bachmann, A. Matteson, M. Schelke, D. Kaschek, S. Hug, C. Kreutz, B. D. Harms, F. J. Theis, U. Klingmüller, and J. Timmer. Lessons learned from quantitative dynamical modeling in systems biology. *PLoS ONE*, 8(9):e74335, 2013.

C. Rausch, M. Rothenberg-Thurley, S. A. Buerger, S. Tschuri, A. Dufour, M. Neusser, S. Schneider, K. Spiekermann, K. H. Metzeler, and F. Ziemann. Double drop-off droplet digital pcr: a novel, versatile tool for mutation screening and residual disease monitoring in acute myeloid leukemia using cellular or cell-free DNA. *The Journal of Molecular Diagnostics*, 23(8):975–985, 2021.

C. Scaffidi, S. Fulda, A. Srinivasan, C. Friesen, F. Li, K. J. Tomaselli, K.-M. Debatin, P. H. Krammer, and M. E. Peter. Two CD95 (APO-1/Fas) signaling pathways. *The EMBO Journal*, 17(6):1675–1687, 1998.

Y. Schälte and J. Hasenauer. Informative and adaptive distances and summary statistics in sequential approximate bayesian computation. *PLoS ONE*, 18(5):e0285836, 2023.

Y. Schälte, P. Stapor, and J. Hasenauer. Evaluation of derivative-free optimizers for parameter estimation in systems biology. *IFAC-PapersOnLine*, 51(19):98–101, 2018.

R. F. Schlenk. Post-remission therapy for acute myeloid leukemia. *Haematologica*, 99(11):1663, 2014.

B. Schöberl, E. A. Pace, J. B. Fitzgerald, B. D. Harms, L. Xu, L. Nie, B. Linggi, A. Kalra, V. Paragas, R. Bukhalid, V. Grantcharova, N. Kohli, K. A. West, M. Leszczyniecka, M. J. Feldhaus, A. J. Kudla, and U. B. Nielsen. Therapeutically targeting ErbB3: a key node in ligand-induced activation of the ErbB receptor–PI3K axis. *Science Signaling*, 2(77):ra31, 2009.

R. Sharma and S. Sharma. *Physiology, Blood Volume*. StatPearls Publishing, Treasure Island (FL), 2023. Accessed: 2025 - Jan - ; PMID: 30252333.

L. B. Sheiner and S. L. Beal. Evaluation of methods for estimating population pharmacokinetic parameters. III. Monoexponential model: routine clinical pharmacokinetic data. *Journal of Pharmacokinetics and Biopharmaceutics*, 11(3):303–319, 1983.

B. Snijder and L. Pelkmans. Origins of regulated cell-to-cell variability. *Nature Reviews Molecular Cell Biology*, 12(2):119–125, 2011.

I. Sobol. Distribution of points in a cube and approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics*, 7(4):86–112, 1967. ISSN 0041-5553.

S. L. Spencer and P. K. Sorger. Measuring and modeling apoptosis in single cells. *Cell*, 144(6): 926–939, 2011.

S. L. Spencer, S. Gaudet, J. G. Albeck, J. M. Burke, and P. K. Sorger. Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nature*, 459(7245):428–433, 2009.

P. Städter, Y. Schälte, L. Schmiester, J. Hasenauer, and P. L. Stapor. Benchmarking of numerical integration methods for ODE models of biological systems. *Scientific Reports*, 11(1):1–11, 2021.

P. Stapor. *Efficient computational methods for parameter estimation of ordinary differential equation and mixed-effect models in systems biology*. PhD thesis, Technische Universität München, 2021.

P. Stapor, D. Weindl, B. Ballnus, S. Hug, C. Loos, A. Fiedler, S. Krause, S. Hross, F. Fröhlich, and J. Hasenauer. PESTO: Parameter EStimation TOolbox. *Bioinformatics*, 34(4):705–707, 2018.

L. S. Stuart Beal. The NONMEM system. *The American Statistician*, 34(2):118–119, 1980.

I. Swameye, T. G. Müller, J. Timmer, O. Sandra, and U. Klingmüller. Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by databased modeling. *Proceedings of the National Academy of Sciences of the United States of America*, 100(3):1028–1033, 2003.

N.-S. Tang, S.-F. Qiu, M.-L. Tang, and Y.-B. Pei. Asymptotic confidence interval construction for proportion difference in medical studies with bilateral data. *Statistical Methods in Medical Research*, 20(3):233–259, 2011.

R. A. Thisted. *Elements of statistical computing: numerical computation*. Routledge, 2017.

L. Tierney and J. B. Kadane. Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81(393):82–86, 1986.

S. T. Tokdar and R. E. Kass. Importance sampling: a review. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(1):54–60, 2010.

C. W. Tornoe, H. Agerso, E. N. Jonsson, H. Madsen, and H. A. Nielsen. Non-linear mixed-effects pharmacokinetic/pharmacodynamic modelling in NLME using differential equations. *Computers in Biology and Medicine*, 76(1):31–40, 2004.

G. Torregrosa and J. Garcia-Ojalvo. Mechanistic models of cell-fate transitions from single-cell data. *Current Opinion in Systems Biology*, 26:79–86, 2021.

R. van der Merwe. *Sigma-point Kalman filters for probabilistic inference in dynamic state-space models*. Ph.d. thesis, Oregon Health & Science University, Apr. 2004.

A. Vaz and L. Vicente. A particle swarm pattern search method for bound constrained global optimization. *Journal of Global Optimization*, 39(2):197–219, 2007.

A. I. F. Vaz and L. N. Vicente. PSwarm: a hybrid solver for linearly constrained global derivative-free optimization. *Optimization Methods & Software*, 24(4-5):669–685, 2009.

A. F. Villaverde, F. Fröhlich, D. Weindl, J. Hasenauer, and J. R. Banga. Benchmarking optimization methods for parameter estimation in large kinetic models. *Bioinformatics*, 35(5): 830–838, 2019.

E. Walter, L. Pronzato, and J. Norton. *Identification of parametric models from experimental data*, volume 1. Springer, 1997.

D. Wang, P. Stapor, and J. Hasenauer. Dirac mixture distributions for the approximation of mixed effects models. *IFAC-PapersOnLine*, 52(26):200–206, 2019.

D. Wang, C. Rausch, S. A. Buerger, S. Tschuri, M. Rothenberg-Thurley, M. Schulz, J. Hasenauer, F. Ziemann, K. H. Metzeler, and C. Marr. Modeling early treatment response in aml from cell-free tumor dna. *Iscience*, 26(12), 2023.

I. B. Weinstein and A. K. Joe. Mechanisms of disease: oncogene addiction—a rationale for molecular targeting in cancer therapy. *Nature Clinical Practice Oncology*, 3(8):448–457, 2006.

J. S. Welch, T. J. Ley, D. C. Link, C. A. Miller, D. E. Larson, D. C. Koboldt, L. D. Wartman, T. L. Lamprecht, F. Liu, J. Xia, et al. The origin and evolution of mutations in acute myeloid leukemia. *Cell*, 150(2):264–278, 2012.

S. Willmann, K. Hohn, A. Edginton, M. Sevestre, J. Solodenko, W. Weiss, J. Lippert, and W. Schmitt. Development of a physiology-based whole-body population model for assessing the influence of individual variability on the pharmacokinetics of drugs. *Journal of Pharmacokinetics and Pharmacodynamics*, 34(3):401–431, 2007.

O. Wolkenhauer, H. Kitano, and K.-H. Cho. Systems biology. *IEEE Control Systems Magazine*, 23(4):38–48, 2003.

S. C. Yu, S. W. Lee, P. Jiang, T. Y. Leung, K. A. Chan, R. W. Chiu, and Y. D. Lo. High-resolution profiling of fetal DNA clearance from maternal plasma by massively parallel sequencing. *Clinical Chemistry*, 59(8):1228–1237, 2013.

J. Zhang, P. L. Yang, and N. S. Gray. Targeting cancer with small molecule kinase inhibitors. *Nature Reviews Cancer*, 9(1):28–39, 2009.

B. Zhao, M. T. Hemann, and D. A. Lauffenburger. Intratumor heterogeneity alters most effective drugs in designed combinations. *Proceedings of the National Academy of Sciences*, 111(29): 10773–10778, 2014.

J. X. Zhou, L. Brusch, and S. Huang. Predicting pancreas cell fate decisions and reprogramming with a hierarchical multi-attractor model. *PLoS ONE*, 6(3):e14752, 2011.