HELMHOLTZ
RESEARCH FOR GRAND CHALLENGES

HELMHOLTZ
MUNICH

Efficient Bayesian parameter estimation applied to mechanistic models of gastric cancer signaling

Elba Raimúndez Álvarez

January 2022

TECHNISCHE UNIVERSITÄT MÜNCHEN

TUM School of Computation, Information and Technology

Efficient Bayesian parameter estimation applied to mechanistic models of gastric cancer signaling

Elba Raimúndez Álvarez

Vollständiger Abdruck der von der TUM School of Computation, Information and Technology der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzende(r): Prof. Dr. Johannes Müller

Prüfer der Dissertation:

1. Prof. Dr. Jan Hasenauer
2. Prof. Dr. Nicole Radde
3. Prof. Dr. Christiane Fuchs

Die Dissertation wurde am 17.01.2022 bei der Technischen Universität München eingereicht und durch die TUM School of Computation, Information and Technology am 21.07.2022 angenommen.

Para os meus avós
Lucinda e Manolo

Acknowledgments

First of all, I would like to thank Jan Hasenauer for many things: his great supervision and guidance, for giving me the possibility to join his research group and to learn so many new things, for providing me much help and support, for being very patient, kind and understanding, for fruitful discussions, for giving me the chance to move to Munich and Bonn, and of course for his humor. I very much enjoyed my time in the group and the amazing working atmosphere.

I especially want to thank all my former and present colleagues and friends at the ICB and the University of Bonn, as well as the thesis party team. In particular, Aaron, Akshaya, Anna F., Anna D., Benjamin, Caro, Daniel, Dantong, Dilan, Emad, Erika, Fabian, Fan, Jakob, Johanna, Katharina, Lara, Lea, Leonard, Lorenzo, Manuel, Marc, Michael, Paul J., Paul S., Polina, Sabrina, Simon, Stephan, Vannesa, Yannik ... and all who I forgot to list here. I had a lot of fun working with you all!

Next, I want to thank my collaborators, Karolin Ebert, Gwen Zwingenberger, Simone Keller, Birgit Luber, Dieter Maier, Ivonne Haffner, Florian Lordick and the SYS-Stomach consortium, all PEtab members, and Alejandro F. Villaverde and Julio R. Banga. I learned a lot and enjoyed many fruitful scientific discussions.

Thanks to the ICB, with its director Fabian Theis, and the University of Bonn for providing exciting environments to work. Big thanks to Claudia Reske, Christiane Stuke, Marianne Antunes, Nina Fischer, Elisabeth Noheimer, Sabine Kunz, and Anna Sacher for their support, help and patiente in all organizational matters at the University of Bonn and the ICB.

Furthermore, I want to thank my thesis advisory committee members Anna Marciniak-Czochra and Christina Kuttler, for valuable inputs and discussions.

I thank the TRA1 Modelling program of the University of Bonn, all reviewers and Julia Belau, for funding my first shared project UNCOVer, Dilan Pathirana for co-leading the project with me, and of course all our students who worked on it.

Last but not least, I want to thank my friends and family, especially my mamá, papá y hermana. What would I have done without you and your continuous support! I know that having a family member abroad is not easy. Always being my rock, os quiero!

Abstract

In systems biology, mathematical models are widely used as they deepen the understanding of biological processes. These models rely on model parameters, such as kinetic rate constants, and observable-related parameters, such as scaling and noise parameters. However, these parameters are often unknown *a priori* and have to be inferred from experimental data. The ever-increasing amount of biological knowledge and more advanced measurement techniques result in the development of larger and more comprehensive models, which ultimately pose new challenges in parameter inference.

In this thesis, I first developed a collection of ordinary differential equation (ODE) models to describe the dynamics of two gastric cancer cell lines: a cetuximab responder and a non-responder. To unravel causal differences between the two cell lines beyond mutation patterns, the model collection comprised different mechanistic hypotheses at distinct parts of the underlying signaling pathways. I performed model calibration and selection using a large dataset, which revealed an optimal candidate model in the considered set. On this model, I conducted uncertainty quantification and validated the model predictive power. The latter, I carried out under a wide range of experimental conditions. All validations of the model were successful, highlighting the value and potential of the proposed model.

Afterwards and in parts motivated by the models previously introduced, I focused on the dimensionality problem for Bayesian parameter inference in mathematical models. To this end, I derived a marginalization-based approach to reduce the dimensionality of the considered models focusing on observation-related parameters. The proposed approach is shown particularly in the context of ODE models with relative measurement data, where observable and noise parameter are integrated out of the estimation problem. Using Markov chain Monte Carlo sampling, I applied this approach to a collection of published models – including the aforementioned gastric cancer model. I found this approach to be not only more efficient than standard methods (where the full parameter space is sampled), but to also enable sampling-based parameter inference and uncertainty quantification for large models where standard approaches failed.

The concepts and methods presented in this thesis enable an efficient Bayesian parameter inference of large and comprehensive mechanistic models, and expand the knowledge of biological systems – with a focus on gastric cancer – providing new mechanistic insights.

Zusammenfassung

In der Systembiologie sind mathematische Modelle weit verbreitet, da sie das Verständnis für biologische Prozesse vertiefen. Diese Modelle beruhen auf Modellparametern, wie kinetischen Geschwindigkeitskonstanten, und beobachtungsbezogenen Parametern, wie Skalierungs- und Rauschparametern. Diese Parameter sind jedoch häufig *a priori* unbekannt und müssen aus experimentellen Daten abgeleitet werden. Die ständig wachsende Menge an biologischem Wissen und fortschrittlichere Messverfahren führen zur Entwicklung größerer und umfassenderer Modelle, die neue Herausforderungen bei der Inferenz von Parametern mit sich bringen.

In dieser Arbeit habe ich zunächst eine Reihe von Modellen basierend auf gewöhnlichen Differentialgleichungen (ODE) entwickelt, um die Dynamik von zwei Magenkrebs-Zelllinien zu beschreiben: ein Cetuximab-Responder und ein Non-Responder. Um die kausalen Unterschiede zwischen den beiden Zelllinien über die Mutationsmuster hinaus zu entschlüsseln, umfasste die Modellsammlung verschiedene mechanistische Hypothesen an unterschiedlichen Stellen der zugrunde liegenden Signalwege. Ich führte eine Modellkalibrierung und -auswahl anhand eines großen Datensatzes durch, die einen optimalen Modellkandidat in der betrachteten Gruppe ergab. Anhand dieses Modells führte ich eine Quantifizierung der Unsicherheit durch und validierte die Vorhersagekraft des Modells. Letzteres habe ich unter einer Vielzahl von Versuchsbedingungen durchgeführt. Alle Validierungsversuche des Modells waren erfolgreich, was den Wert und das Potenzial des vorgeschlagenen Modells unterstreicht.

Anschließend konzentrierte ich mich, zum Teil motiviert durch die zuvor vorgestellten Modelle, auf das Dimensionalitätsproblem bei der bayesianischen Parameterinferenz in mathematischen Modellen. Zu diesem Zweck leitete ich einen auf Marginalisierung basierenden Ansatz zur Reduzierung der Dimensionalität der betrachteten Modelle ab, der sich auf beobachtungsbezogene Parameter konzentriert. Die Vorteile des vorgeschlagenen Ansatzes werden insbesondere im Kontext von ODE-Modellen mit relativen Messdaten aufgezeigt, bei denen beobachtbare und Rauschparameter aus dem Schätzproblem integriert werden. Unter Verwendung von Markov chain Monte Carlo Sampling wandte ich diesen Ansatz auf eine Sammlung von veröffentlichten Modellen an – einschließlich des oben erwähnten Magenkrebsmodells. Ich stellte fest, dass dieser Ansatz nicht nur effizienter ist als Standardmethoden (bei denen der gesamte Parameterraum abgetastet wird), sondern auch eine

stichprobenbasierte Parameterinferenz und Unsicherheitsquantifizierung für große Modelle ermöglicht, bei denen Standardansätze versagten.

Die in dieser Arbeit vorgestellten Konzepte und Methoden ermöglichen eine effiziente Bayesianische Parameterinferenz großer und holistischer mechanistischer Modelle und erweitern das Wissen über biologische Systeme, wobei der Schwerpunkt hier auf Magenkrebs liegt und neue mechanistische Erkenntnisse liefert.

Contents

Chapter 1

Introduction

What is mathematics? It is only a systematic effort of solving puzzles posed by nature.

—Shakuntala Devi

Interdisciplinary research has emerged over the last decades and it has been shown as a key factor to reach a high impact in the scientific community (Okamura, 2019). Likewise, the synergy between biology and mathematics has extended both fields greatly in the last centuries. For example, many well-known mathematical methods, such as Markov chains (Robert & Casella, 2004), initially arose from biological problems (Cohen, 2004). This constant interplay has originated the research field of systems biology (Ideker et al., 2001; Kitano, 2002; Klipp et al., 2005).

1.1. Mathematical modeling of biological systems

In systems biology, mathematical models – especially mechanistic models – are widely used to gain causal descriptions and insights about underlying biological processes. They allow to extend and circumvent possible limitations arising from only using biological experiments (Kitano, 2002). Additionally, they can predict the dynamics and latent variables of a system of interest (Adlung et al., 2017). Models based on ordinary differential equations (ODEs) are among the most popular mechanistic models (Ingalls, 2013; Klipp et al., 2005). Applications of ODE models range from intracellular signal transduction (Bachmann et al., 2011; Kholodenko, 2007; Swameye et al., 2003), gene regulation (Kühn et al., 2009), cell cycle (Lloyd, 2013), metabolism (Smallbone & Mendes, 2013), and apoptosis (Spencer & Sorger, 2011); to diseases such as cancer (Fey et al., 2015; Hass et al., 2017), and infectious diseases in human populations (Giordano et al., 2020; Kermack et al., 1927; Zhao & Chen, 2020). Furthermore, mechanistic models can be used in the area of personalized medicine (Ogilvie et al., 2015; Saez-Rodriguez & Blüthgen, 2020). The ability of ODE models to describe and provide further mechanistic insights, within the broad spectrum of plausible applications, elucidate the widespread interest in using them to understand and research various biological processes and diseases.

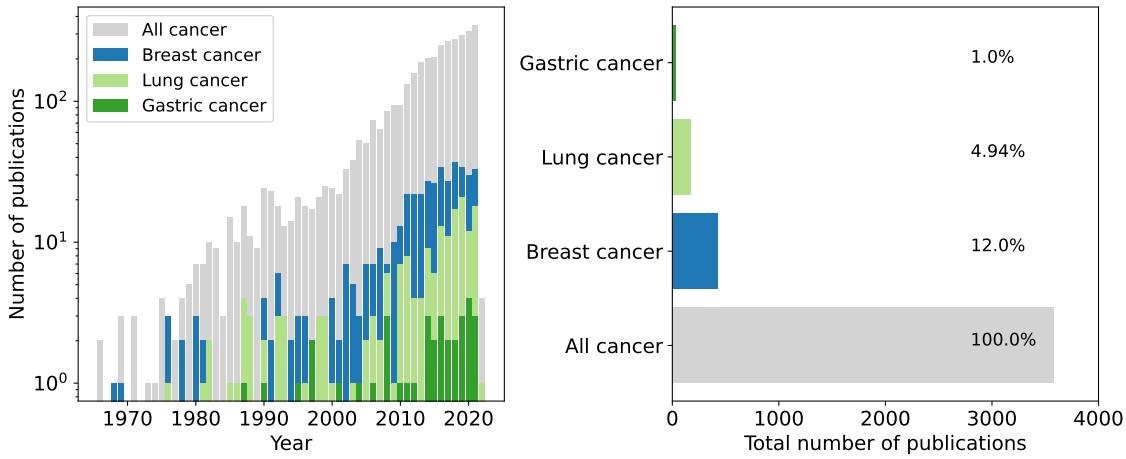

Figure 1.1.: Published studies using mathematical models in cancer research. In the left panel, yearly PubMed query results for mathematical/mechanistic/computational model for cancer (without distinction of types), and specific to female breast, lung, and gastric/stomach cancer are shown. In the right panel, the fraction of the total amount of publications across the years is depicted. Results were extracted on the 3$^{\text{rd}}$ of January, 2022. The search query terms are listed in Annex A.

Mathematical modeling in cancer research Cancer is a major leading cause of death worldwide (Bray et al., 2021). Lung, gastric/stomach and female breast cancer are among the most deadly cancers (Sung et al., 2021). Over the last decades, cancer research has benefited increasingly from mathematical and computational approaches, as evidenced by the constantly increasing number of publications (Figure 1.1). Altrock et al. (2015) provide an overview of various types of mathematical models to describe cancer at different scales. Yet, only a small fraction from all available modeling studies has focused specifically on female breast cancer (12%), lung cancer (5%), and gastric cancer (1%) as indicated in Figure 1.1. Among others, this is an obstacle to define personalized treatments, or to discover suitable biomarkers for patient stratification, in these particular cancer types. This is especially relevant for gastric cancer, which is clearly underrepresented in the literature (Figure 1.1). Therefore, more comprehensive mechanistic models tailored to specific cancer types are needed to pave the way towards personalized medicine.

1.2. Parameter inference

Mechanistic models based on ODEs usually comprise unknown parameters, such as kinetic rate constants, which in most cases cannot be directly measured experimentally. Instead, these need to be inferred from experimental data (Tarantola, 2005). This is frequently done by optimizing an objective function, e.g. a likelihood or least-squares function, which

quantitatively evaluates the discrepancy of model simulation and measurement data. In a Bayesian setting, this is combined with prior knowledge (Fröhlich et al., 2019).

The overall goal of parameter estimation is the identification of optimal point estimates, e.g. using optimization methods, and uncertainty quantification, e.g. using profile likelihoods or Markov chain Monte Carlo (MCMC) sampling (Raue et al., 2013b). Computationally speaking, parameter inference can be a challenging task, even for small-size models (Kapfer et al., 2019). This is often caused by the non-convex and multi-modal, i.e. with multiple modes, nature of the objective function. Therefore, efficient and scalable approaches are needed.

Mostly, the experimental data used to calibrate a model is subject to measurement errors. These can be quantified if experimental replicates are available, however they are often unknown and have to be inferred along with the model parameters. On top of this, commonly used experimental techniques – such as Western blotting (Renart et al., 1979), fluorescence microscopy (Sanderson et al., 2014) or flow cytometry (Herzenberg et al., 2006) – provide in the first place only relative information about the absolute concentrations of interest. All this comes with a cost, as for the mapping between model simulation and measurements, the incorporation of additional parameters is required (Loos et al., 2018). Consequently, this increases the complexity of the inference problem and raises the need for tailored methods. In the context of relative data, either data-driven normalization (Degasperi et al., 2017), or additional observable parameters have been applied. For the latter, hierarchical approaches were developed to reduce the resulting increased dimensionality of the parameter space by exploiting the problem structure (Loos et al., 2018; Schmiester et al., 2019; Weber et al., 2011). These approaches have shown to substantially improve optimizer efficiency for models of various sizes, including large-scale models (Schmiester et al., 2019). However, these methods cannot be applied if a rigorous Bayesian uncertainty quantification using MCMC methods is intended.

1.3. Overview and contributions of this thesis

In this thesis, I focus on the following open problems and challenges:

(i) **Limited availability of mechanistic models of gastric cancer.** Despite of gastric cancer being the fifth most common cancer and fourth leading cause of death from cancer worldwide (Sung et al., 2021), the number of available mechanistic models developed for this disease is very small constituting only 1% of all available models in cancer research (Figure 1.1). This implies that model reusability in this context is

very limited. Therefore, the development of gastric cancer mechanistic models using standardized formats is needed.

(ii) **Use of relative measurement data for model calibration (substantially) increases the dimension of the estimation problem.** Mechanistic models often comprise not only parameters which govern the dynamics, but also observable and noise parameters which relate to the measurement process. Mostly, these additional parameters need to be estimated along with the dynamic parameters increasing – in cases, substantially – the dimension of the problem. Approaches to tackle this exist (Loos et al., 2018; Schmiester et al., 2019; Weber et al., 2011), however none of them are suitable for uncertainty quantification using Bayesian sampling-based approaches.

(iii) **Convergence failure of available sampling algorithms when applied to large mechanistic models.** It is well-known that MCMC sampling methods tend to suffer from the curse of dimensionality when the number of sampled dimensions rises (Bellman, 1961). The more comprehensive a mechanistic model becomes, the more parameters are needed. However, changes in parameter dimensionality are not always related to growing model complexity. For example, in mechanistic models calibrated with relative measurement data, the more data is available, the more observable (e.g. scaling constants) and noise parameters are needed. All these contribute to hinder the convergence to the posterior distribution.

Especially problems (ii) and (iii) arise when modeling various biological processes and impede a reliable computational analysis of the system of interest. In this thesis, I provide a basis to address the following biological question:

(iv) **In gastric cancer, the patient response to cetuximab treatment is heterogeneous.** This phenomenon has prohibited the use of cetuximab in the clinical practice. Therefore, the development and discovery of predictive biomarkers to enable patient stratification is needed.

To address the aforementioned issues, these are my main contributions in this thesis:

(1) **Development of a gastric cancer model which captures different cell lines.** I developed a mechanistic model based on ODEs describing the MET, EGFR, ERK and AKT signaling pathways for two distinct cell lines. The model also accounts for different mutation patterns present in the considered cell lines, and for dynamical differences due to changes in the cell culture media. Furthermore, it incorporates available knowledge about the dynamical parameters and it is implemented in the

standardized SBML (Hucka et al., 2003) and `PEtab` format for parameter estimation (Schmiester et al., 2021). This contribution addresses problem (i) and (iv).

(2) **Unraveling causal differences of cetuximab responder and non-responder cell lines.** The gastric cancer model that I developed was applied to study the dynamics of cetuximab responder and non-responder cell lines. To model differences in the cell lines, I built a collection of eight mathematical models considering different parts of the proposed model to be cell line specific. Performing model selection, I found that the model assuming cell line specific dynamics at the EGFR turnover level seems to be most reasonable. Lastly, I performed uncertainty analysis and validated the predictions of the best candidate model. This contribution addresses problems (i) and (iv).

(3) **Efficient sampling of mechanistic models using relative measurements employing a marginalization-based approach.** I developed an approach based on the analytical marginalization of the posterior to reduce the dimensionality of the parameter space. The parameters subject to marginalization include those not influencing the dynamics of the considered model, i.e. observable and noise parameters. The marginalization-based approach for sampling achieved a higher performance in terms of effective sample size per unit of time, especially for multi-modal posteriors, than the standard approach, where the full posterior is sampled. This contribution addresses problem (ii).

(4) **Application of the novel marginalization-based sampling approach to the gastric cancer model.** I applied the introduced framework of the marginalization-based approach for sampling to assess parameter uncertainty in the gastric cancer model that I developed in (1). Due to the model and dataset size, the standard approach failed. Yet, the marginalization-based approach robustly converged to the posterior. This enabled the global assessment of parameter uncertainty using MCMC sampling for larger models. This contribution addresses problem (iii).

Some of these contributions are already part of peer-reviewed publications, currently submitted to peer-reviewed journals or in preparation. Parts of the work in this thesis thus correspond or are to some extent similar to the following publications:

- **Raimúndez, E.**, Keller, S., Zwingenberger, G., Ebert, K., Hug, S., Theis, F. J., Maier, D., Luber, B., and Hasenauer, J. (2020). Model-based analysis of response and resistance factors of cetuximab treatment in gastric cancer cell lines. *PLoS Computational Biology*, 16(3):e1007147. `doi:10.1371/journal.pcbi.1007147`.

- **Raimúndez, E.**, and Hasenauer, J. Posterior marginalization accelerates Bayesian inference for dynamical systems, *in preparation.*

Further, results of the following publications in which I was involved are used although not discussed in detail:

- **Raimúndez, E.**, Dudkin, E., Vanhoefer, J., Alamoudi, E., Merkt, S., Fuhrmann, L., Bai, F., and Hasenauer, J. (2021). COVID-19 outbreak in Wuhan demonstrates the limitations of publicly available case numbers for epidemiological modelling. *Epidemics*, 100439. doi:10.1016/j.epidem.2021.100439.

- Schmiester, L.*, Schälte, Y.*, Bergmann, F. T., Camba, T., Dudkin, E., Egert, J., Fröhlich, F., Fuhrmann, L., Hauber, A. L., Kemmer, S., Lakrisenko, P., Loos, C., Merkt, S., Pathirana, D., **Raimúndez, E.**, Refisch, L., Rosenblatt, M., Stapor, P. L., Städter, P., Wang, D., Wieland, F., Banga, J. R., Timmer, J., Villaverde, A. F., Sahle, S., Kreutz, C., Hasenauer, J., and Weindl, D. (2021). PEtab – Interoperable specification of parameter estimation problems in systems biology. *PLoS Computational Biology*, 17(1):e1008646. doi:10.1371/journal.pcbi.1008646.

- Hass, H.*, Loos, C.*, **Raimúndez-Álvarez, E.**, Timmer, J., Hasenauer, J.†, and Kreutz, C.† (2019). Benchmark problems for dynamic modeling of intracellular processes. *Bioinformatics*, 17(1):3073-3082. doi:10.1093/bioinformatics/btz020.

- Villaverde, A. F., **Raimúndez, E.**, Hasenauer, J., and Banga, J. R. (2019). A comparison of methods for quantifying prediction uncertainty in systems biology. *IFAC-PapersOnLine*, 52(26):45-51. doi:10.1016/j.ifacol.2019.12.234.

Other contributions of my doctoral research which are not included in this thesis are:

- Contento, L., Castelletti, N., **Raimúndez, E.**, Le Gleut, R., Schälte, Y., Stapor, P. L., Hinske, L. C., Hoelscher, M., Wieser, A., Radon, K., Fuchs, C., Hasenauer, J., and the KoCo19 study group (2021). Integrative modelling of reported case numbers and seroprevalence reveals time-dependent test efficiency and infection rates. *medRxiv*. doi:10.1101/2021.10.01.21263052

- Haffner, I., Schierle, K., **Raimúndez, E.**, Geier, B., Maier, D., Hasenauer, J., Luber, B., Walch, A., Kolbe, K., J. R. Knorrenschild, A. Kretzschmar, B. Rau, L. F. Weikersthal, Ahlborn, M., Siegler, G., Fuxius, S., Decker, T., Wittekind, C.,

*,† denotes equal author contribution

and Lordick, F. (2021). HER2 expression, test deviations, and their impact on survival in metastatic gastric cancer: Results from the prospective multicenter VARIANZ study. *Journal of Clinical Oncology*, 20.02761. doi:10.1200/JCO.20.02761.

- Villaverde, A. F., **Raimúndez, E.**, Hasenauer, J., and Banga, J. R. Assessment of prediction uncertainty quantification methods in systems biology, *in preparation*.

- Ebert, K., Haffner, I., Zwingenberger, G., Keller, S., **Raimúndez, E.**, Geffers, R., Wirtz, R., Barbaria, E., Hollerieth, V., Arnold, R., Walch, A., Hasenauer, J., Maier, D., Lordick, F., and Luber, B. Combining gene expression analysis of gastric cancer cell lines and tumor specimens to identify biomarkers for anti-HER therapies – The role of HAS2, SHB and HB-EGF, *under review*.

- Ebert, K., Zwingenberger, G., Barbaria, E., Keller, S., Heck, C., Arnold, R., Hollerieth, V., Mattes, J., Geffers, R., **Raimúndez, E.**, Hasenauer, J., and Luber, B. (2020). Determining the effects of trastuzumab, cetuximab and afatinib by phosphoprotein, gene expression and phenotypic analysis in gastric cancer cell lines. *BMC Cancer*, 20(1):1-19. doi:10.1186/s12885-020-07540-7.

In addition to the contribution to these research articles, I also contributed to the development of several toolboxes which enable the reusability of the here developed methods:

- **pyPESTO** (https://github.com/icb-dcm/pypesto, see Schälte et al. (2021)): A Python-based parameter estimation toolbox providing an integrated platform for different optimization and uncertainty quantification algorithms.

- **PEtab** (https://github.com/petab-dev/petab, see Schmiester et al. (2021)): A standardized, interoperable format to define parameter estimation problems based on SBML (Hucka et al., 2003) and tab-separated value files, alongside a Python-based library.

1.4. Outline of this thesis

The remainder of this thesis is organized as follows: In Chapter 2, I introduce the notation and background on ODE modeling of biological processes, as well as parameter inference and uncertainty analysis. Additionally, I provide a background to model selection. Chapters 3 and 4 gather my main contributions in this thesis. In Chapter 3, I present the development of a mechanistic model of gastric cancer to study response and resistance factors for cetuximab treatment. The model captures the MET, EGFR, ERK and AKT signaling

pathways in two gastric cancer cell lines. I calibrate the model using a comprehensive dataset, and provide an assessment of parameter and prediction uncertainties. I show that the proposed model facilitates the identification of causal differences between the cell lines, and provides predictions for the responses to multiple validation experiments. In Chapter 4, I propose a marginalization-based approach that analytically integrates over observable and noise parameters yielding a lower-dimensional posterior, and derive the distributions to retrieve these parameters subject to marginalization. I show the benefit of this newly developed approach on various application examples, including the gastric cancer model that I developed in Chapter 3. In Chapter 5, I conclude this thesis with a summary of my main contributions and an outlook on prospective further developments.

The following parts of this thesis are written from the *we*-perspective.

Chapter 2

Background

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

- **Raimúndez, E.**, and Hasenauer, J. Posterior marginalization accelerates Bayesian inference for dynamical systems, *in preparation.*
- **Raimúndez, E.**, Dudkin, E., Vanhoefer, J., Alamoudi, E., Merkt, S., Fuhrmann, L., Bai, F., and Hasenauer, J. (2021). COVID-19 outbreak in Wuhan demonstrates the limitations of publicly available case numbers for epidemiological modelling. *Epidemics*, 100439. doi:10.1016/j.epidem.2021.100439.

In this chapter, we introduce the notation and describe the relevant methods on which this thesis is founded. Firstly, we describe the modeling of biological systems using ordinary differential equations, and how the link between model output and measurements is constructed. Secondly, we describe how the inference of the unknown parameters – including uncertainty analysis – in these models is performed from two different perspectives: frequentist and Bayesian. Lastly, we outline different methods to perform model selection.

2.1. Mathematical modeling of biological systems

Mathematical models are very valuable tools to gain insights about the underlying mechanisms of biological processes. A wide range of modeling techniques, depending on the studied data type and research question, are available. Among others, examples are: Markov jump processes, reaction rate equation models, and agent-based models. In this thesis, we focus on deterministic reaction rate equation models, which are frequently used to model population averages (Klipp et al., 2005). In the following, we present their structure, as well as how the link to the experimental data required for parameter inference is constructed.

2.1.1. Ordinary differential equation models

Reaction rate equation models are based on ordinary differential equations (ODEs) of the form

$$\dot{x}(t,\theta,u) = f(x(t,\theta,u),\theta,u), \quad x(t_0,\theta,u) = x_0(\theta,u), \tag{2.1}$$

in which the vector field $f : \mathbb{R}^{n_x} \times \mathbb{R}^{n_\theta} \times \mathbb{R}^{n_u} \to \mathbb{R}^{n_x}$ denotes the temporal evolution of the state variables $x = x(t,\theta,u) \in \mathbb{R}^{n_x}$ at time $t \in [t_0, T] \subset \mathbb{R}$. We assume that the vector field f is uniformly Lipschitz continuous in x, then, following Picard's existence theorem (Coddington & Levinson, 1955) the local existence and uniqueness of the solution of (2.1) is assured. The state vector $x(t,\theta,u) \in \mathbb{R}^{n_x}$ describes the amount of the modeled species which can range, e.g., from mRNA and protein concentrations to number of individuals in a population. The vector of unknown **model parameters**, which is estimated from the measurements, is denoted as $\theta \in \Theta \subset \mathbb{R}^{n_\theta}$, where Θ is the domain of biologically reasonable parameter values. Usually, θ includes reaction rate constants and initial amounts of species. The vector of known inputs, i.e. different experimental conditions such as drug concentrations or gene mutations, is denoted as $u = u(t) \in \mathbb{R}^{n_u}$. We assume model and input parameters to stay constant over time. Here and in the following, $n_x \in \mathbb{N}$ is the total number of modeled species, $n_\theta \in \mathbb{N}$ the total number of model parameters, and $n_u \in \mathbb{N}$ the total number of known inputs.

2.1.2. Constructing the observables

The unknown **model parameters** θ are inferred from the measurements. For that, an appropriate link between the model outputs and experimental data is needed. Often, single measurements provide information on a set of species, such as the total amount of a phosphorylated protein, rather than on a single species. The link is constructed by introducing an observation function $h : \mathbb{R}^{n_x} \times \mathbb{R}^{n_\theta} \times \mathbb{R}^{n_u} \to \mathbb{R}^{n_y}$, which maps the states, model parameters and inputs, to observables $y(t,\theta,u) \in \mathbb{R}^{n_y}$ via

$$y(t,\theta,u) = h(x(t,\theta,u),\theta,u),$$

where $n_y \in \mathbb{N}$ is the total number of observables. The observables are the measured properties of the model. Frequently, measurements are noise corrupted. Hence, to obtain the measurements $\bar{y}$ the model also must capture experimental errors by defining a noise model.

For the most part, independent and additive Gaussian-distributed noise is assumed

$$\bar{y}_{i_y,i_t,i_e} = y_{i_y}(t_{i_t}, \theta, u^{i_e}) + \varepsilon_{i_y,i_t,i_e}, \quad \text{with } \varepsilon_{i_y,i_t,i_e} \sim \mathcal{N}(0, \sigma^2_{i_y,i_t,i_e}), \tag{2.2}$$

with observable index i_y, time index i_t, experimental condition index i_e, and **noise parameters** $\sigma \in \mathbb{R}^{n_\sigma}$. The total number of noise parameters is denoted by $n_\sigma = n_y \times n_e \times n_t$, with $n_t \in \mathbb{N}$ as the total number of time points and $n_e \in \mathbb{N}$ as the total number of distinct experimental conditions. In this particular noise model, the noise parameters correspond to the standard deviations of the Gaussian distribution. In addition to this, other common noise assumptions include: (i) log-normally-distributed multiplicative noise (Kreutz et al., 2007), and (ii) Laplace-distributed noise. The first is often assumed when noise levels are proportional to measurement values (Raue et al., 2013b), while the latter has shown to be more robust especially in the presence of measurement outliers (Maier et al., 2017). Depending on the measurement process, the noise parameters may be unknown for all or for a subgroup of the observables. For example, when experimental replicates are available, the noise parameters can be directly obtained. Yet, often they are unknown.

We denote the group of all measurements as

$$\mathcal{D} = \{\bar{y}_{i_y,i_t,i_e}\}^{\substack{i_y=1,\dots,n_y \\ i_t=1,\dots,n_t \\ i_e=1,\dots,n_e}}.$$

Most measurement techniques only provide relative information about the absolute concentrations of interest. Consequently, we must consider an additional measurement process function $g : \mathbb{R}^{n_y} \times \mathbb{R}^{n_\xi} \to \mathbb{R}^{n_y}$

$$y(t, \theta, u, \xi) = g(h(x(t, \theta, u), \theta, u), \xi)$$

with **observable parameters** $\xi \in \mathbb{R}^{n_\xi}$, and total number of observable parameters $n_\xi \in \mathbb{N}$. These parameters are often unknown and, therefore, also need to be estimated along with the unknown model and noise parameters. One of the most common measurement process function is a linear transformation of the form

$$g(h(x(t, \theta, u), \theta, u), s) = s \cdot h(x(t, \theta, u), \theta, u) \tag{2.3}$$

with observable parameters s, often denoted (for this specific case) as scaling constants. This structure is employed in Chapters 3 and 4. For additional examples of measurement process functions see (Schmiester et al., 2019).

2.2. Frequentist parameter inference

The mathematical models previously introduced rely on **model parameters**, e.g. kinetic rate constants and initial conditions, **noise parameters**, e.g. shape parameters related to the assumed noise distribution, and **observable parameters**, e.g. scaling and offset constants. All these parameters usually cannot be directly measured, hence to obtain them we need to solve an inverse problem. In other words, the unknown parameters are inferred from experimental data. In this section, we describe how models introduced in Section 2.1 can be calibrated to measurements from a frequentist perspective (Raue et al., 2013b) and, subsequently, how uncertainty in the inferred parameters is quantified.

2.2.1. Likelihood function

The likelihood $p(\mathcal{D}|\theta, \xi, \sigma)$ is the conditional density of observing the measurement data $\mathcal{D}$ given a model and its corresponding model parameters θ, observable parameters ξ and noise parameters σ. Let us denote the set of unknown θ, ξ and σ, as $\boldsymbol{\psi} = (\boldsymbol{\theta}, \boldsymbol{\xi}, \boldsymbol{\sigma})$ with $\boldsymbol{\psi} \in \mathbb{R}^{n_\psi}$, and the total number of unknown parameters $n_\psi = n_\theta + n_\xi + n_\sigma$. Then, if we assume independent and additive Gaussian-distributed noise, the likelihood function is in the form of

$$p(\mathcal{D}|\boldsymbol{\psi}) = \prod_{i_y=1}^{n_y} \prod_{i_t=1}^{n_t} \prod_{i_e=1}^{n_e} \frac{1}{\sqrt{2\pi}\boldsymbol{\sigma}_{i_y,i_t,i_e}} \exp\left(-\frac{1}{2} \left(\frac{\bar{y}_{i_y,i_t,i_e} - y_{i_y}(t_{i_t}, \boldsymbol{\theta}, u^{i_e}, \boldsymbol{\xi})}{\boldsymbol{\sigma}_{i_y,i_t,i_e}} \right)^2 \right). \tag{2.4}$$

The maximum likelihood estimate (MLE) $\hat{\boldsymbol{\psi}}$ maximizes the likelihood function,

$$\hat{\boldsymbol{\psi}} = \arg\max_{\boldsymbol{\psi} \in \Psi} p(\mathcal{D}|\boldsymbol{\psi}), \tag{2.5}$$

with the domain $\Psi \subset \mathbb{R}^{n_\psi}$ as the Cartesian product of Θ (Section 2.1.1) and the sets of plausible observable and noise parameters. Hence, Ψ contains all physically plausible parameters. For convenience, often the logarithm of the likelihood (2.4) is considered. This favors numerical stability (Raue et al., 2013a) and, as the logarithm is a monotonic function, the location of the optimum is conserved. Furthermore, since most optimization algorithms are minimization routines, the **negative log-likelihood** is used instead. This is given by

$$\begin{aligned} \mathcal{J}(\boldsymbol{\psi}) &= -\log p(\mathcal{D}|\boldsymbol{\psi}) \\ &= \frac{1}{2} \sum_{i_y=1}^{n_y} \sum_{i_t=1}^{n_t} \sum_{i_e=1}^{n_e} \log\left(2\pi\boldsymbol{\sigma}_{i_y,i_t,i_e}^2\right) + \left(\frac{\bar{y}_{i_y,i_t,i_e} - y_{i_y}(t_{i_t}, \boldsymbol{\theta}, u^{i_e}, \boldsymbol{\xi})}{\boldsymbol{\sigma}_{i_y,i_t,i_e}} \right)^2. \end{aligned} \tag{2.6}$$

Additionally, when the noise parameters $\boldsymbol{\sigma}_{i_y,i_t,i_e}$ are known from the measurements, the negative log-likelihood is linked to the weighted least-squares function

$$\mathcal{J}^{\mathrm{WLS}}(\boldsymbol{\psi}) = \sum_{i_y=1}^{n_y} \sum_{i_t=1}^{n_t} \sum_{i_e=1}^{n_e} \boldsymbol{w}_{i_y,i_t,i_e} \left(\bar{y}_{i_y,i_t,i_e} - y_{i_y}(t_{i_t}, \boldsymbol{\theta}, u^{i_e}, \boldsymbol{\xi})\right)^2 \tag{2.7}$$

with weights $\boldsymbol{w}_{i_y,i_t,i_e} = 1/2\sigma^2_{i_y,i_t,i_e}$. The constant terms can be disregarded, as they do not alter the optima of the function.

On top of this, in the field of systems biology parameters have mostly non-negative values, consequently these can be log-transformed. In (Hass et al., 2019) we showed substantial benefits in the optimization, e.g. increased convexity, due to this transformation.

2.2.2. Parameter optimization

Optimization typically consists of minimizing an objective function $\mathcal{J} : \Psi \to \mathbb{R}$, such as (2.6) or (2.7), given some constraints on the parameters. In the particular case of systems biology, plausible domains Ψ of the parameters $\boldsymbol{\psi}$ are considered, e.g. non-negativity, in the optimization problem. Frequently, these are simple box-constraints of the form

$$\begin{aligned} &\min_{\boldsymbol{\psi}\in\Psi} \mathcal{J}(\boldsymbol{\psi}) \\ &\text{s.t. } \mathbf{l} \leq \boldsymbol{\psi} \leq \mathbf{u} \end{aligned} \tag{2.8}$$

with lower bounds $\mathbf{l} \in \mathbb{R}^{n_\psi}$, upper bounds $\mathbf{u} \in \mathbb{R}^{n_\psi}$ and with $\mathbf{l}, \mathbf{u} \in \Psi$. Assuming that $\mathcal{J}$ is continuous in the box $[\mathbf{l}, \mathbf{u}]^{n_\psi}$, then the existence of at least one minimum value (and one maximum value) is guaranteed according to the extreme value theorem. Optimality conditions provide some structural information about the properties of optimal solutions. Under the assumption of sufficiently large box boundaries or no bounds present, the following theorems for optimality conditions are introduced.

Theorem 2.2.1 (First-order necessary condition for (local) optimality)**.** *If ψ^* is an unconstrained local minimizer of a differentiable function $\mathcal{J} : \Psi \to \mathbb{R}$, then we must have*

$$\nabla \mathcal{J}(\psi^*) = 0.$$

Thus, the first-order optimality is a necessary condition, but it is not sufficient since a point with first-order optimality equal to zero is not necessarily a minimum.

Theorem 2.2.2 (Second-order necessary condition for (local) optimality)**.** *If ψ^* is an unconstrained local minimizer of a twice differentiable function $\mathcal{J} : \Psi \to \mathbb{R}$, then, in addition to $\nabla\mathcal{J}(\psi^*) = 0$, we must have*

$$\nabla^2\mathcal{J}(\psi^*) \geq 0,$$

where $\nabla^2\mathcal{J}(\psi^*)$ denotes the Hessian matrix of $\mathcal{J}$ at ψ^* and the inequality means that it is positive semidefinite.

Therefore, the second-order necessary condition for optimality is not sufficient since a point with first-order optimality equal to zero and a positive semidefinite Hessian matrix is not necessarily a minimum (e.g. a saddle point).

Theorem 2.2.3 (Second-order sufficient condition for (local) optimality)**.** *Let us suppose $\mathcal{J} : \Psi \to \mathbb{R}$ is twice differentiable, $\nabla\mathcal{J}(\psi^*) = 0$, and*

$$\nabla^2\mathcal{J}(\psi^*) > 0,$$

where $\nabla^2\mathcal{J}(\psi^*)$ denotes the Hessian matrix of $\mathcal{J}$ at ψ^* and the inequality means that it is positive definite. Then, ψ^* is a strict local minimum of $\mathcal{J}$.

Here, a point with first-order optimality condition sufficed and a positive definite Hessian matrix is sufficient although not necessary for local optimality.

Accordingly, under the assumption of a continuous twice differentiable objective function $\mathcal{J}$, necessary and sufficient optimality conditions can be derived. If ψ^* is an interior point of $[\mathbf{l}, \mathbf{u}]^{n_\psi}$ and a local minimizer of (2.8), it holds that a minimum is found when $\nabla\mathcal{J}(\psi^*) = 0$ and $\nabla^2\mathcal{J}(\psi^*) > 0$. However, in the case that all or some elements of ψ^* lie at the boundaries of the box, it holds that the necessary optimality condition for a local minimum is given by

$$\frac{\partial\mathcal{J}(\psi^*)}{\partial\psi_j^*} \begin{cases} \geq 0, & \text{if } \psi_j^* = \mathbf{l}_j \\ = 0, & \text{if } \mathbf{l}_j < \psi_j^* < \mathbf{u}_j \\ \leq 0, & \text{if } \psi_j^* = \mathbf{u}_j \end{cases}$$

with parameter index $j = 1, \ldots, n_\psi$ (Boyd & Vandenberghe, 2004). Alternatively, more general inequality constraints can occur, therefore the Lagrangian of the optimization problem can be considered (Nocedal & Wright, 2006). Several well-established methods exist to find the optimal parameters ψ^* (Nocedal & Wright, 2006). In the following, we

briefly discuss the most commonly used in the field of systems biology: local and global parameter optimization methods.

Local parameter optimization

To find the optimal parameters of the problem (2.8), optimization has to be performed. To achieve this, various methods were developed (Nocedal & Wright, 2006). Local optimization aims at finding the **locally optimal parameters** for a specific region of the search space. In case of problems that have only global optima, this will actually return the global optimum. Various algorithms exist for local parameter optimization (Nocedal & Wright, 2006) and popular examples of these include: the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm (Fletcher & Powell, 1963; Goldfarb, 1970; Nocedal, 1980), the Nelder-Mead algorithm (Nelder & Mead, 1965), and the Hill-Climbing algorithm (De La Maza & Yuret, 1994). The BFGS method belongs to the class of Quasi-Newton methods, which means that it is a type of second-order optimization algorithm that approximates and uses the second-order derivative (i.e. the Hessian) of the objective function using the first-order derivative (i.e. the gradient). Various other optimization algorithms use derivative information (which can be of first-, second-, or higher-order) for local optimization. These can be arranged into two main groups: (i) **line-search methods**, which first fix a search direction of descent, and then determine the step-size ensuring sufficient decrease (Wächter & Biegler, 2006), and (ii) **trust-region methods**, which instead define a trust-region around the current position in parameter space, then simultaneously determine the search direction and step-size, and update the trust-region radius for the next step (Coleman & Li, 1996). In an own study (Hass et al., 2019) we have shown the latter to be more efficient.

However, for certain objective functions the first-order derivative may be unavailable, unreliable or impractical to obtain. To circumvent this, **gradient-free** local optimization methods exist including the aforementioned Nelder-Mead and Hill-Climbing algorithms. Using these may be helpful in case gradient not possible. Yet, in the context of biologically motivated ODE models gradients are usually available, therefore gradient-based optimization algorithms are preferred (Schälte et al., 2018).

Global parameter optimization

Global optimization is distinguished from local optimization by its focus in finding the **globally optimal parameters**. Finding these is not straightforward since analytical methods are often not applicable, and the use of numerical algorithms usually leads to very hard challenges. To globally optimize a multi-modal and non-linear objective function, multiple

approaches exist including evolutionary algorithms (Bäck, 1996; Runarsson & Yao, 2000), Bayesian optimization (Mockus, 2011), and simulated annealing (Kirkpatrick et al., 1983).

However, global optimization can also be pursued in a simpler manner by using **multi-start local optimization** in which multiple randomly initialized local optimizations take place (Raue et al., 2013b). From the set of all individual local searches, the resulting locally optimal points are sorted by the final objective function value. Then, if sufficiently many points are employed, the best local optimum is assumed to be a potentially global optimum. The more frequently one ends up at the same potentially global optimum, the more reliable the assumption of globality becomes. Moreover, this approach can provide insights into the structure of the different local optima of the objective function. Furthermore, hybrid approaches exist which combine global and local optimization approaches. A very well-known hybrid method is scatter-search (Egea et al., 2007; Penas et al., 2015; Villaverde et al., 2012), which fuses a genetic algorithm and multi-start local optimization. In this thesis, we focus on the multi-start local optimization strategy for all optimization tasks.

There are several studies that benchmark several of these methods aiming to identify the most efficient with a focus on ODE models in systems biology (Hass et al., 2019; Schälte et al., 2018; Villaverde et al., 2018). Moreover, in (Hass et al., 2019) we introduced a collection of 20 standardized benchmark models, which can be used to evaluate the performance of existing and future methods. A subset of these models are used in Chapter 4.

2.2.3. Uncertainty analysis

While the MLE (2.5) provides a single point estimator for the optimal parameter vector of all unknown parameters ψ^*, information about how uncertain and reliable these estimates are is missing. Uncertainty analysis of the parameter estimates is especially important for the assessment of how trustworthy model predictions are. The most common frequentist approaches to quantify this are by local approximations with the Fisher information matrix and profile likelihoods.

Fisher information matrix

The Fisher information matrix (FIM) (Fisher, 1922) provides an approximation to the Hessian of the negative log-likelihood function and information about (local) parameter uncertainty (Chis et al., 2011; Cobelli & DiStefano, 1980). Under the assumption of

Gaussian noise and known σ, the FIM can be calculated as

$$\mathrm{FIM}(\boldsymbol{\psi}) = \sum_{i_y=1}^{n_y} \sum_{i_t=1}^{n_t} \sum_{i_e=1}^{n_e} \left(\frac{\partial y_{i_y}(t_{i_t}, \boldsymbol{\theta}, u^{i_e}, \boldsymbol{\xi})}{\partial(\boldsymbol{\theta}, \boldsymbol{\xi})} \right) \frac{1}{\sigma^2_{i_y, i_t, i_e}} \left(\frac{\partial y_{i_y}(t_{i_t}, \boldsymbol{\theta}, u^{i_e}, \boldsymbol{\xi})}{\partial(\boldsymbol{\theta}, \boldsymbol{\xi})} \right)^T, \tag{2.9}$$

where $\frac{\partial y_{i_y}(t_{i_t}, \boldsymbol{\theta}, u^{i_e}, \boldsymbol{\xi})}{\partial(\boldsymbol{\theta}, \boldsymbol{\xi})}$ are the sensitivities of the observables with respect to the set of model and observable parameters $(\boldsymbol{\theta}, \boldsymbol{\xi})$. To have a glimpse on parameter uncertainty of the model, the eigenvalues of the FIM can be examined. Small eigenvalues indicate large uncertainties for the parameter combinations (given by the associated eigenvector), meaning that large variations in these do not alter the behavior of the system. On the contrary, large eigenvalues indicate stiff parameter combinations. A model is considered "sloppy" when the spectrum of FIM eigenvalues is spread over many orders of magnitude (Gutenkunst et al., 2007). Sloppiness provides a first hint about the most or least identifiable directions in parameter space, as modifying the sloppy parameter combinations has relatively little effect on the model behavior. This approach is particularly useful for high-dimensional problems, i.e. containing hundreds to thousands of parameters, since the methods later introduced may not be computationally tractable (Kapfer et al., 2019). However, sloppiness analysis cannot replace structural or practical identifiability analysis (Chis et al., 2011). Therefore, when possible, it should be complemented with other methods, like profile likelihoods (Raue et al., 2009), since the FIM-based confidence intervals cannot be asymmetric, i.e. are insensitive to practical non-identifiabilities, and vary under re-parameterizations of the model, e.g. logarithmic transformation of the parameters (Wieland et al., 2021).

Profile likelihood

The **profile likelihood** is given by

$$\mathrm{PL}\left(\psi_j\right) = \max_{\psi_{k \neq j}, \boldsymbol{\psi} \in \Psi} p(\mathcal{D}|\boldsymbol{\psi}),$$

where j is the parameter index and ψ_j is the parameter value of the j-th parameter (Raue et al., 2009). This method consists of maximizing the likelihood function with respect to all parameters $\psi_{k \neq j}$ for a fixed value for ψ_j. The calculation of the profiles can be done with optimization-based methods (Raue et al., 2015) or with integration-based methods (Hass et al., 2016; Stapor et al., 2018). This is generally more computationally expensive but also more accurate than the aforementioned local approximations by the FIM (Stapor et al., 2018).

From the profile likelihoods, the finite sample **confidence intervals** (CIs) for a given significance level α can be derived as

$$\mathrm{CI}_{j,\alpha} = \left(\psi_j \middle| \frac{\mathrm{PL}\left(\psi_j\right)}{p(\mathcal{D}|\hat{\boldsymbol{\psi}})} > \exp\left(-\frac{\Delta_\alpha}{2}\right) \right),$$

with Δ_α denoting the α-th percentile of the χ^2-distribution with one degree of freedom (Meeker & Escobar, 1995). Confidence intervals provide information about practical identifiability (Raue et al., 2009). Briefly, parameters are: (i) **practically identifiable** when their confidence intervals are finite, (ii) **practically non-identifiable** when their confidence intervals are only finitely-bounded on one side while the other is infinite, or (iii) **structurally non-identifiable** when their confidence intervals are infinite. For further details we refer to Raue et al. (2009).

Mostly, profile likelihoods are calculated for the subset of **model parameters** instead of all unknown parameters. The observable parameters and noise parameters can be disregarded since they are only related to the measurement process and do not provide any mechanistic insights (Loos et al., 2018). However, note that they are still part of the likelihood (2.4). This analysis provides a maximum projection of the likelihood on the individual parameters. Therefore, it is often complemented with Monte Carlo methods.

2.3. Bayesian parameter inference

As an alternative to frequentist approaches for parameter inference, there are Bayesian methods (Wilkinson, 2007). In Bayesian inference, the available knowledge of the process of interest is encoded as an *a priori* probability distribution. The prior distribution is updated with the data using Bayes' theorem to obtain the posterior distribution. The inference of the posterior distribution is often performed using Markov chain Monte Carlo (MCMC) algorithms. In this section, we introduce Bayes' theorem and shortly provide some basic theory of Markov chains before introducing the MCMC algorithms used throughout this thesis.

2.3.1. Bayes' theorem

According to Bayes' theorem, the **posterior distribution** is given by

$$p(\boldsymbol{\psi}|\mathcal{D}) = \frac{p(\mathcal{D}|\boldsymbol{\psi})p(\boldsymbol{\psi})}{p(\mathcal{D})} \tag{2.10}$$

in which $p(\mathcal{D}|\boldsymbol{\psi})$ denotes the likelihood (as introduced in (2.4)), $p(\boldsymbol{\psi})$ denotes the prior distribution and $p(\mathcal{D})$ denotes the marginal probability (or evidence). The latter is given by

$$p(\mathcal{D}) = \int_{\Psi} p(\mathcal{D}|\boldsymbol{\psi})p(\boldsymbol{\psi})d\boldsymbol{\psi}$$

and, in most cases, it is computationally expensive since normally $\boldsymbol{\psi}$ is high-dimensional. However, it can be neglected since it is a normalization constant independent of $\boldsymbol{\psi}$. The maximum a posteriori (MAP) is obtained, similarly to the MLE, by maximizing the posterior distribution. In order to approximate the posterior distribution $p(\boldsymbol{\psi}|\mathcal{D})$, sophisticated methods, such as MCMC sampling or variational inference, can be employed. While MCMC sampling is asymptotically exact, this does not always hold for variational inference (Blei et al., 2017). In other words, MCMC methods will – in the limit – exactly approximate the target distribution. This is not guaranteed for variational inference. In this thesis, we focus on MCMC algorithms.

2.3.2. Markov chain theory in a nutshell

Here, we shortly introduce some basic concepts about Markov chains. Further details can be found in (Robert & Casella, 2004) and references therein. In the following, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(S, \mathcal{S})$ a measurable space, i.e. the state space of the Markov chain. In general S can be continuous or discrete, but here we only consider the continuous space $(S, \mathcal{S}) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. A discrete-time Markov process is defined as:

Definition 2.3.1 (Markov chain)**.** A discrete-time stochastic process $\{X^{(t)}\}_{t\geq 0}$ is called a **Markov chain** with respect to the probability measure $\mathbb{P}$, if for all $t \in \mathbb{N}_0$, and all $x^{(0)}, \ldots, x^{(t+1)} \in S$, it holds that

$$\mathbb{P}\left(X^{(t+1)} = x^{(t+1)}|X^{(0)} = x^{(0)}, \ldots, X^{(t)} = x^{(t)}\right) = \mathbb{P}\left(X^{(t+1)} = x^{(t+1)}|X^{(t)} = x^{(t)}\right)$$

given that the conditional probabilities are well-defined, that is, if

$$\mathbb{P}\left(X^{(t+1)} = x^{(t+1)}|X^{(0)} = x^{(0)}, \ldots, X^{(t)} = x^{(t)}\right) > 0$$

Since $t \in \mathbb{N}_0$, a Markov chain is an entity with discrete indices. This can be generalized for each index of the continuous state space. A Markov chain can be viewed as, in some sense, memoryless since the conditional probability at any future state $X^{(t+1)}$, given the past states $X^{(0)}, \ldots, X^{(t-1)}$ and the present state $X^{(t)}$, only depends on the present state and it is independent of the past states. This is known as the Markov property.

Definition 2.3.2 (Time-homogeneous Markov chain)**.** A Markov chain $\{X^{(t)}\}_{t\in\mathbb{N}_0}$ is called **time-homogeneous** when for all t it holds that

$$\mathbb{P}\left(X^{(t+1)} = x | X^{(t)} = y\right) = \mathbb{P}\left(X^{(t)} = x | X^{(t-1)} = y\right).$$

This represents that the probability of moving from state y to state x is independent of *time*, with *time* being the chain index t. For the following, unless stated otherwise, we always assume time-homogeneous Markov chains.

The conditional probability can also be seen as a transition kernel, which is defined in the following.

Definition 2.3.3 (Transition kernel)**.** A **transition kernel** is a function P defined on a measurable space $(S, \mathcal{S})$ such that

(i) $\forall x \in S, P(x, \cdot)$ is a probability measure; and

(ii) $\forall A \in \mathcal{S}, P(\cdot, A)$ is $\mathcal{S}$-measurable.

If we put $P(x, y) := \mathbb{P}\left(X^{(t+1)} = x | X^{(t)} = y\right)$ for $x, y \in S$, we see how conditional probabilities relate to transition kernels.

The ultimate goal of Markov chain Monte Carlo methods is the inference of a distribution π, in our case the posterior $p(\boldsymbol{\psi}|\mathcal{D})$ (2.10), using samples from a Markov chain. For this, the key point is to define the conditions under which a Markov chain converges to π, as well as the conditions under which the stationary distribution π exists. The stationary distribution of a Markov chain denotes the distribution of $X^{(t)}$ after a sufficiently long time, where the distribution of $X^{(t)}$ does not change any longer.

Definition 2.3.4 (Stationary/invariant distribution)**.** A distribution π is called **stationary** (or invariant) for the transition kernel $P(\cdot, \cdot)$ if it satisfies

$$\pi(A) = \int_S P(x, A)\pi(dx), \qquad \forall A \in \mathcal{S}.$$

Therefore, if π is a stationary distribution and the Markov chain starts directly from π at $t = 0$, the chain will stay in π for all $t \geq 0$. To verify that in our case a stationary distribution exists and it is unique, the concepts of irreducibility and recurrence are needed.

Definition 2.3.5 (φ-irreducible Markov chain)**.** Given a measure φ on $\mathcal{S}$, a Markov chain $\{X^{(t)}\}_{t\in\mathbb{N}_0}$ with transition kernel $P(\cdot,\cdot)$ is called **φ-irreducible** if for every $A \in \mathcal{S}$ with $\varphi(A) > 0$, there exists a t such that

$$P^t(x, A) > 0, \qquad \forall x \in S$$

where $P^t(x, A) > 0$ denotes the kernel for t transitions defined inductively by $P^1(x, A) = P(x, A)$ and $P^t(x, A) = \int_S P^{t-1}(y, A)P(x, dy)$.

The property of **irreducibility** means that the transition kernel allows the Markov chain to move from any state to any other state within S using only transitions of positive probability regardless of what the starting value $x^{(0)}$ was.

Definition 2.3.6 (Recurrent Markov chain)**.** A Markov chain $\{X^{(t)}\}_{t\in\mathbb{N}_0}$ is **recurrent** if

(i) there exists a measure φ such that $\{X^{(t)}\}_{t\in\mathbb{N}_0}$ is φ-irreducible; and

(ii) for every $A \in \mathcal{S}$ such that $\varphi(A) > 0$,

$$\mathbb{E}_x\left[\sum_{t=1}^{\infty} \mathbb{1}\left(X^{(t)} \in A\right)\right] = \infty \qquad \forall x \in A,$$

where $\mathbb{1}(\cdot)$ denotes the indicator function.

This means that any set $A \in \mathcal{S}$ is reached infinitely many times. On top of this, a Markov chain is said to be **positive recurrent** when assumes the expected return time T to any state x to be finite.

Definition 2.3.7 (Positive recurrent Markov chain)**.** A recurrent Markov chain $\{X^{(t)}\}_{t\in\mathbb{N}_0}$ is **positive recurrent** if for all states $x \in S$, it holds that $\mathbb{E}_x[T(x)] < \infty$, where $T(x) = \min\{t \in \mathbb{N} : X^{(t)} = x\}$.

Following from Definitions 2.3.5 and 2.3.7, it can be shown that each irreducible, positive recurrent Markov chain has a unique stationary distribution (see Robert & Casella (2004)). However, this does not guarantee the convergence of a Markov chain to its stationary distribution. Fortunately, there also exist properties to ensure convergence: aperiodicity and ergodicity.

Definition 2.3.8 (Period of a Markov chain)**.** Let us suppose a Markov chain $\{X^{(t)}\}_{t\in\mathbb{N}_0}$ with transition kernel $P(\cdot,\cdot)$. The **period** of state x is

$$d(x) = \gcd\{t \geq 1 : P^t(x,x) > 0\},$$

where gcd denotes the greatest common divisor.

This means that the Markov chain can return to the starting x only at multiples of the period d, with d being the largest such integer. Therefore, a state x is **aperiodic** if $d(x) = 1$, and periodic if $d(x) > 1$. The property of **aperiodicity** ensures that there is no repeating structure in the sequence of a Markov chain.

Definition 2.3.9 (Ergodic Markov chain)**.** A Markov chain $\{X^{(t)}\}_{t\in\mathbb{N}_0}$ with transition kernel $P(\cdot,\cdot)$ and stationary distribution π is called **ergodic** if it converges to π regardless of its starting value $x^{(0)} \in S$.

This means, if a Markov chain is ergodic, it will converge to its stationary distribution. This is satisfied if all its states are positive recurrent and aperiodic.

Theorem 2.3.1 (Convergence of a Markov chain)**.** *Let $\{X^{(t)}\}_{t\in\mathbb{N}_0}$ be a Markov chain with transition kernel $P(\cdot,\cdot)$ and stationary distribution π, which is φ-irreducible, aperiodic and recurrent. Then*

(i) the Markov chain is positive recurrent;

(ii) π is its unique stationary distribution; and

(iii) P is ergodic for π.

This theorem yields a useful criterion for showing that an MCMC method is appropriate. Summarizing, the Markov chain should be φ-irreducible, aperiodic, positive recurrent, and should have a stationary distribution π. Then, distance measures like the total variation distance can be used to check closeness to the stationary distribution. For more details see Levin et al. (2006).

2.3.3. Markov chain Monte Carlo sampling algorithms

To construct a Markov chain with the desired posterior distribution as its stationary distribution, MCMC methods can be used. These belong to a class of Monte Carlo

methods suited for sampling from complicated, whose probability density is only known up to a multiplicative factor, and high-dimensional probability distributions. Examples of MCMC algorithms are **single-chain methods** like the Metropolis-Hastings (Hastings, 1970; Metropolis et al., 1953) and the adaptive Metropolis algorithms (Haario et al., 2001), or **multiple-chain methods** like the parallel tempering (Łacki & Miasojedow, 2015; Neal, 1996), the parallel hierarchical sampling (Rigat & Mira, 2012) and the region-based adaptive parallel tempering algorithms (Ballnus et al., 2018). Moreover, there exist MCMC algorithms that use the gradient information of the objective function, such as Hamilton Monte Carlo sampling (Graham & Storkey, 2017; Hoffman & Gelman, 2014) or the Metropolis-adjusted Langevin algorithm (Girolami & Calderhead, 2011).

Similarly to optimization, there exist efforts to establish standard workflows and to assess efficiency of available MCMC methods in the context of ODE models in systems biology (Ballnus et al., 2017). This standard pipeline is employed in Chapter 4. As mentioned earlier, alternatively (or complementary) to profile likelihoods, MCMC methods can be used to quantify parameter uncertainty. For this, credible intervals based on an MCMC sample can be derived (Hines et al., 2014). The simplest approach to calculate the credible interval of the j-th parameter ψ_j from a N-long posterior sample $\psi_1, \ldots, \psi_N$, is to (implicitly) marginalize over all other parameters $\psi_{k \neq j}$ and set

$$C_{j,\alpha} = \{\psi_j | \psi_j^{(\alpha/2)} \leq \psi_j \leq \psi_j^{(1-\alpha/2)}\},$$

where $\psi_j^{(\alpha)}$ denotes the α-quantile of the sample.

In the following, the adaptive Metropolis and parallel tempering algorithms, which are employed in Chapter 4, are introduced. Further methods are described in (Brooks et al., 2011).

Adaptive Metropolis algorithm

The selection of an effective proposal distribution for standard MCMC algorithms, e.g. the Metropolis-Hastings (MH) algorithm (Hastings, 1970; Metropolis et al., 1953), is key to obtain a reasonable exploration of the posterior in a limited amount of time. In most applications, the proposal distribution will be a Normal distribution centered at the current state. However, the covariance matrix of the Normal distribution has to be determined, which is often done by trial and error. This becomes more challenging for higher-dimensional problems as it is often not easy to find a good proposal distribution. To overcome this, algorithms with adaptation schemes have been introduced. One well-known method is the **adaptive Metropolis** (AM) algorithm as introduced by Haario et al. (2001).

Algorithm 1: Pseudo-code of the adaptive Metropolis sampling algorithm.

Input: Initial point ψ_0, initial covariance matrix Σ_0 and sample length N, covariance scaling factor η_{n_ψ}, index for starting the adaptation t_0, small constant $\epsilon > 0$, and identity matrix I_{n_ψ}

Output: Sample $\psi_1, \ldots, \psi_N$

$i = 1$
while $i \leq N$ **do**
  Sample $\psi_* \sim \mathcal{N}(\psi_{i-1}, \Sigma_{i-1})$ and $u \sim \mathcal{U}_{[0,1]}$
  // Evaluate the acceptance probability p_{acc}
  $p_{acc} \leftarrow \min\left(1, \frac{p(\psi_*|\mathcal{D})}{p(\psi_{i-1}|\mathcal{D})}\right)$
  if $u \leq p_{acc}$ **then**
    // Accept ψ_*
    $\psi_i \leftarrow \psi_*$
  else
    // Reject ψ_*
    $\psi_i \leftarrow \psi_{i-1}$
  end
  // Adaptation of the covariance matrix
  if $i > t_0$ **then**
    // Adapt Σ_i
    $\Sigma_i \leftarrow \eta_{n_\psi}\left(\text{cov}(\psi_0, \ldots, \psi_{i-1}) + \epsilon I_{n_\psi}\right)$
  else
    $\Sigma_i \leftarrow \Sigma_0$
  end
  // Move to next sample
  $i \leftarrow i + 1$
end

This method extends the MH algorithm to enable the continuous tuning of the proposal distribution based on the available samples' history. However, the acceptance scheme for newly proposed samples is conserved from the MH algorithm. As a result of including the adaptation scheme, an efficient sampling of high-dimensional spaces becomes more efficient by facilitating the convergence of the algorithm.

The AM algorithm employs a multivariate Normal distribution as proposal with a continuously updated covariance matrix based on the available samples via a recursion formula. The multivariate Normal distribution has its mean at the currently available sample ψ_{i-1} and the covariance $\Sigma_i = \Sigma_i(\psi_0, \ldots, \psi_{i-1})$, with i denoting the sample index. To regulate the dependence of the proposal covariance matrix on the history of the chain, a freely-chosen

index t_0 defines where the adaptation starts and Σ_i is set to

$$\Sigma_i = \begin{cases} \Sigma_0, & \text{if } i \leq t_0 \\ \eta_{n_\psi} \left(\text{cov}(\psi_0, \ldots, \psi_{i-1}) + \epsilon I_{n_\psi} \right), & \text{if } i > t_0 \end{cases}$$

with η_{n_ψ} being a scaling factor of the covariance matrix which depends on the dimension of the parameter space n_ψ, and I_{n_ψ} being the n_ψ-dimensional identity matrix. To ensure that the covariance matrix does not become singular, the constant $\epsilon > 0$ is set to a very small value. The initial covariance matrix Σ_0 has to be strictly positive definite and should be chosen according to the best available prior knowledge (even if poor) and, in a manner, the trust in Σ_0 is reflected on the size of t_0. According to the definition of the empirical covariance matrix and with $i \geq t_0 + 1$, the following recursion formula for the proposal covariance matrix is satisfied:

$$\Sigma_{i+1} = \frac{i-1}{i}\Sigma_i + \frac{\eta_{n_\psi}}{i} \left(i\bar{\psi}_{i-1}\bar{\psi}_{i-1}^{\mathrm{T}} - (i+1)\bar{\psi}_i\bar{\psi}_i^{\mathrm{T}} + \psi_i\psi_i^{\mathrm{T}} + \epsilon I_{n_\psi} \right). \tag{2.11}$$

Since the mean also follows a recursion formula $\bar{\psi}_i = \frac{1}{i}\sum_{m=1}^{i}\psi_m$, the computational cost of (2.11) is bearable. The value of the scaling parameter of the covariance matrix is adopted from Gelman et al. (1996) where it is given by $\eta_{n_\psi} = \frac{2.4^2}{n_\psi}$. The pseudo-code describing the method is shown in Algorithm 1.

Parallel tempering algorithm

Single-chain algorithms, such as adaptive Metropolis, explore the parameter space in a local manner. This is generally sufficient for sampling uni-modal posteriors, but frequently biologically motivated problems, on which this thesis focuses, result in posteriors with more complex shapes and multiple modes. Especially, single-chain algorithms struggle when the high probability posterior modes are separated by areas of low probability density. To circumvent this, **parallel tempering** (PT) algorithms have been introduced. PT algorithms use several Markov chains in parallel to sample from tempered versions of the posterior, i.e.

$$p(\psi|\mathcal{D})_{T_l} \propto p(\mathcal{D}|\psi)^{1/T_l} p(\psi) \tag{2.12}$$

with increasingly ordered temperatures $T_l \geq 1$, and chain index $l = 1, \ldots, L$ with L being the total number of chains (Miasojedow et al., 2013; Sambridge, 2013; Vousden et al., 2016). The tempered posteriors become flatter and broader, in comparison to the non-tempered posterior, which increases the transition probability between different posterior modes. Each auxiliary chain explores its tempered distribution $p(\psi|\mathcal{D})_{T_l}$ using an MCMC algorithm. The key point is that swaps can occur between adjacent pairs of chains, l and l', with

acceptance probability

$$p_{swap} = \min\left(1, \left(\frac{p(\mathcal{D}|\psi^{l'})}{p(\mathcal{D}|\psi^{l})}\right)^{\beta_l - \beta_{l'}}\right)$$

where $\beta_l \equiv 1/T_l$ is the inverse temperature of chain l. If a swap is accepted, the involved chains exchange positions in parameter space, i.e. the parameter samples swap between both chains. Consequently, the hottest chain T_{max} is the most likely to access all of the modes of $p(\psi|\mathcal{D})$ and propagate to colder chains facilitating to the coldest chain $T_1 = 1$ an efficient exploration of the whole posterior. Therefore, this algorithm is particularly well suited for problems containing multiple and well separated modes. The pseudo-code describing the method is shown in Algorithm 2. Moreover, some PT samplers can adapt the temperatures during sampling based on the swap acceptance probabilities (Miasojedow et al., 2013). In this thesis, we use adaptive PT in combination with adaptive Metropolis for each tempered chain available in the `pyPESTO` toolbox (Schälte et al., 2021). We would like to note that in the `pyPESTO` implementation the likelihood and prior are tempered, instead of only the likelihood as shown in (2.12).

2.3.4. Quality assessment of samples

Once a Markov chain is constructed using an MCMC algorithm of choice, the quality of the MCMC samples must be carefully analyzed. The quality assessment is often used as a measure to evaluate the performance of an MCMC method. An important criterion is the overall acceptance rate, i.e. the number of accepted samples divided by the number of total proposed samples, with an optimal value at 23% as shown by Roberts et al. (1997). Moreover, the concepts of burn-in period and effective sample size, which are employed in Chapter 4, are introduced in the following.

Testing for convergence

A common practice is to visually inspect the obtained MCMC chains to assess convergence to the stationary distribution (in our case to the posterior (2.10)), but there are also several statistical tests available (Brooks & Roberts, 1998). It is frequent to observe a distortion/bias in the MCMC chain due to initialization which is known as **burn-in period**. Even when initializating the chain at the MAP, burn-in may still occur, as observed for adaptive methods in (Calderhead, 2011). The burn-in related samples are often discarded as they do not represent the target distribution (Brooks & Roberts, 1998).

Algorithm 2: Pseudo-code of the parallel tempering sampling algorithm in combination with adaptive Metropolis (Algorithm 1) for each tempered chain.

Input: Number of chains L, sample length N, initial points ψ_0^l with $l = 1, \ldots, L$, initial temperatures $(1/T_1 > \cdots > 1/T_L) \equiv (\beta_1 > \cdots > \beta_L)$ with $T_1 = 1$, and adaptive Metropolis specific options (see Algorithm 1)
Output: Sample $\psi_1, \ldots, \psi_N$

$i = 1$
while $i \leq N$ **do**
  for $l \leftarrow 1$ **to** L **do**
    Sample $\psi_* \sim \mathcal{N}(\psi_{i-1}^l, \Sigma_{i-1}^l)$ and $u \sim \mathcal{U}_{[0,1]}$
    // Evaluate the acceptance probability p_{acc} on the tempered posterior
    $p_{acc} \leftarrow \min\left(1, \frac{p(\mathcal{D}|\psi_*)^{\beta_l} p(\psi_*)}{p(\mathcal{D}|\psi_{i-1}^l)^{\beta_l} p(\psi_{i-1}^l)}\right)$
    Update chain l according to Algorithm 1
  end
  // Swap samples between adjacent pairs of chains
  // starting from highest temperature (T_L) to lowest (T_1)
  for $l \leftarrow L$ **to** 2 **do**
    Sample $u \sim \mathcal{U}_{[0,1]}$
    // Evaluate the swap acceptance probability p_{swap}
    $p_{swap} \leftarrow \min\left(1, \left(\frac{p(\mathcal{D}|\psi_i^l)}{p(\mathcal{D}|\psi_i^{l-1})}\right)^{\beta_{l-1}-\beta_l}\right)$
    if $u < p_{swap}$ **then**
      // Accept swap
      $(\psi_i^{l-1}, \psi_i^l) \leftarrow (\psi_i^l, \psi_i^{l-1})$
    else
      // Do nothing
    end
  end
  [Optional: Temperature adaptation]
  // Move to next sample
  $i \leftarrow i + 1$
end

Convergence tests aim to check whether an MCMC chain has already converged to the target distribution (and thus to estimate the burn-in period). Similarly to hypothesis tests, they cannot prove actual convergence but only suggest a strong evidence against the null hypothesis that a chain has converged. Most well-known tests are the Geweke test (Geweke, 1992), and the Gelman-Rubin test (Brooks & Gelman, 1998; Gelman & Rubin, 1992), which we briefly review in the following.

The **Geweke test** was originally introduced to test for convergence of the Gibbs Sampler (Geweke, 1992). However, it also can be used with samples generated with other MCMC algorithms. In this test, the chain is split into two subsamples and, if the empirical mean of the two subsamples is very different, failure of convergence is detected. The Geweke's statistic is given by

$$z = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{\widehat{S}_A + \widehat{S}_B}}$$

with $\bar{X}_A$ and $\bar{X}_B$ denoting the empirical mean, and $\widehat{S}_A$ and $\widehat{S}_B$ the empirical standard deviation for the two different windows A and B of the Markov chain. Commonly, the sizes of windows A and B are respectively set to the first 10% and the last 50% of the corresponding chain (Geweke, 1992). As chain length goes to infinity, Geweke's statistic is asymptotically normally-distributed ($z \to \mathcal{N}(0,1)$). For high-dimensional Markov chains, the test is performed on each dimension separately. In this thesis, we use the Geweke test because of its easy applicability and low computational cost.

The **Gelman-Rubin test** is especially suited for cases in which several Markov chains for the same target distribution are available. The Gelman-Rubin statistic compares the *between*- and *within*-chain variance which should be very similar if all chains have converged. Let L be the total number of available Markov chains. The multiple chains L are assumed to be initialized at different values $x_l^{(0)}$ with $l = 1, \ldots, L$ denoting the chain index. We define the number of samples after discarding the burn-in period as converged samples. Thus, let n be the number of converged samples. The *between*-chain variance B is defined by

$$B = \frac{n}{L-1} \sum_{l=1}^{L} (\bar{x}_l - \bar{x})^2 \qquad \text{with} \quad \bar{x}_l = \frac{1}{n} \sum_{t=1}^{n} x_l^{(t)} \qquad \text{and} \quad \bar{x} = \frac{1}{L} \sum_{l=1}^{L} \bar{x}_l,$$

and the *within*-chain variance W is defined by

$$W = \frac{1}{L} \sum_{l=1}^{L} \left(\widehat{S}_l\right)^2 \qquad \text{with} \quad \left(\widehat{S}_l\right)^2 = \frac{1}{n-1} \sum_{t=1}^{n} \left(x_l^{(t)} - \bar{x}_l\right)^2.$$

Then, the Gelman-Rubin statistic $\hat{R}$ is given by

$$\hat{R} = \sqrt{\frac{\widehat{\text{var}}}{W}} \qquad \text{with} \quad \widehat{\text{var}} = \frac{n-1}{n} W + \frac{1}{n} B.$$

When $\hat{R}$ is close to 1 it represents that the chains are well-mixed, and it can be assumed that the L Markov chains indeed are stationary. In practice, often a cut-off value of 1.1 for $\hat{R}$ is considered acceptable (Gelman et al., 2013).

Chain auto-correlation

MCMC chains naturally possess a certain degree of auto-correlation as the draws are not independent. Low auto-correlation is always desirable as it implicitly means that there is no need to generate longer chains in order to be representative. While the burn-in period is discarded, thinning a chain, i.e. only keeping every i-th element and discarding the rest, is often not advisable when calculating auto-correlation (Geyer, 1992). The two main reasons for high auto-correlation within a chain are: (i) new samples are only proposed by small step-sizes (generating consecutive samples close to each other), and (ii) the chain acceptance rates are low (low proportion of newly proposed samples are accepted). However, large step-sizes do not always relate with high acceptance rates (Link & Eaton, 2011). All these can result in many consecutive steps being even equal to each other and preventing the chain from efficiently exploring the target distribution. Therefore, highly auto-correlated chains do not contain the same amount of information as a chain with independent and identically distributed samples. To quantify this, the **effective sample size** can be used.

The **effective sample size** n_{eff} provides the number of independent draws from the target distribution that the correlated chain is effectively equivalent to. Thus, the larger the effective sample size, the better. It is defined by

$$n_{\text{eff}} = \frac{n}{1 + 2\sum_{\tau=1}^{\infty} \rho_\tau}$$

where n is the number of converged samples, and ρ_τ is the estimated auto-correlation at lag τ. Several methods differing in computational effort and statistical precision exist to approximate the auto-correlation, such as the periodogram- (Brockwell & Davis, 2016), Sacchi- (Sacchi et al., 1998b), and Sokal's adaptive truncated periodogram- (Sokal, 1997) estimator. In this thesis, we use Sokal's adaptive truncated periodogram-estimator (Haario et al., 2006; Sacchi et al., 1998a) available in the `pyPESTO` toolbox (Schälte et al., 2021). Note that, as this provides a univariate measure, the maximum of the auto-correlation across all dimensions of $\psi \in \mathbb{R}^{n_\psi}$ is considered in order to determine the effective sample size.

2.4. Model selection

While developing models, e.g. of biochemical processes or infectious diseases, it often occurs that multiple hypotheses about the underlying mechanisms are available. Consequently, in addition to the parameters, the structure of the ODE model (2.1) needs to be inferred from the measurements. In this section, we introduce likelihood-based model selection methods, such as the Akaike and Bayesian information criteria, and Bayesian model selection methods, such as the Bayes factor. The parts addressing the Akaike and Bayesian information criteria in Section 2.4.2 are modified versions of the corresponding sections of the thesis author's publication (Raimúndez et al., 2021).

2.4.1. Bayes factor

Given a set of candidate models $\mathcal{M} = \left(M^{(1)}, M^{(2)}, \ldots, M^{(n_M)}\right)$, with $n_M \in \mathbb{N}$ denoting the total number of model hypotheses, the goal is to obtain a model (or a set of them) which best describes the measurements at hand and at the same time generalizes to other datasets (Hastie et al., 2005). Hence, a balance between goodness-of-fit and model complexity needs to be considered as a quality measure for the selection of the optimal model(s). The choice of the optimal model(s) can be made with several selection criteria. A popular selection criteria based on Bayes' theorem (2.10) is the **Bayes factor** (Kass & Raftery, 1995). The posterior model probability for the m-th model is given by

$$p(M^{(m)}|\mathcal{D}) = \frac{p(\mathcal{D}|M^{(m)})p(M^{(m)})}{p(\mathcal{D})}$$

with model prior $p(M^{(m)})$, marginal probability $p(\mathcal{D}) = \sum_{m=1}^{n_M} p(\mathcal{D}|M^{(m)})p(M^{(m)})$, and marginal likelihood

$$p(\mathcal{D}|M^{(m)}) = \int_{\Psi^m} p(\mathcal{D}|\boldsymbol{\psi}^{(m)}, M^{(m)})p(\boldsymbol{\psi}^{(m)}|M^{(m)})d\boldsymbol{\psi}^{(m)} \tag{2.13}$$

where $\boldsymbol{\psi}^{(m)}$ is the vector of unknown parameters of model $M^{(m)}$. The Bayes factor is a likelihood ratio of the marginal likelihood of two competing models $M^{(1)}$ and $M^{(2)}$,

$$B_{12} = \frac{p(\mathcal{D}|M^{(1)})}{p(\mathcal{D}|M^{(2)})},$$

and describes how much more likely the assumption of the data being generated by $M^{(1)}$ instead of $M^{(2)}$ is. Following Jeffreys's scale of evidence, having $B_{12} > 100$ is often considered decisive for rejecting $M^{(2)}$ (Jeffreys, 1961). In contrast to the likelihood-ratio

test (cf. Section 2.4.3), the competing models are not required to be nested and, in case of being so, the Bayes factor does not equal the likelihood-ratio statistic as it intrinsically takes into account the dimension of the nested models by integrating over the whole parameter space of each model. Several methods exist to efficiently solve (2.13), the preferred of which is thermodynamic integration as it yields most robust and numerically stable results (Lartillot & Philippe, 2006). This approach uses the tempered posterior

$$p_\tau\left(\boldsymbol{\psi}^{(m)}|\mathcal{D}, M^{(m)}\right) = \frac{p(\mathcal{D}|\boldsymbol{\psi}^{(m)}, M^{(m)})^\tau p(\boldsymbol{\psi}^{(m)}|M^{(m)})}{\int_{\Psi^m} p(\mathcal{D}|\tilde{\boldsymbol{\psi}}^{(m)}, M^{(m)})^\tau p(\tilde{\boldsymbol{\psi}}^{(m)}|M^{(m)}) d\tilde{\boldsymbol{\psi}}^{(m)}}, \tag{2.14}$$

with $\tau \in [0, 1]$ being the temperature parameter. The tempered posterior (2.14) for $\tau = 0$ coincides with $p(\boldsymbol{\psi}^{(m)}|M^{(m)})$, and for $\tau = 1$ coincides with the "regular" (untempered) posterior. Then, integrating both sides with respect to τ yields (Lartillot & Philippe, 2006)

$$\log p(\mathcal{D}|M^{(m)}) = \int_0^1 \mathbb{E}_{p_\tau}[\log p(\mathcal{D}|\boldsymbol{\psi}^{(m)}, M^{(m)})] d\tau.$$

This integral can be numerically evaluated by choosing $0 = \tau_0 < \tau_1 < \cdots < \tau_{n_\tau - 1} < \tau_{n_\tau} = 1$ as discretization. Then, within the $n_\tau \in \mathbb{N}$ discretized values, the expectation for a specific τ can be obtained by Monte Carlo sampling and the integrand can be solved applying e.g. the Simpson's or trapezoidal rule. The number of temperatures n_τ can be subject to adaptation (Hug et al., 2016).

However, for high-dimensional or computationally demanding models, the calculation of Bayes factors becomes intractable, hence computationally cheaper model selection criteria, like Akaike and Bayesian information criteria, are preferred (Fröhlich et al., 2019).

2.4.2. Akaike and Bayesian information criteria

The **Akaike Information Criterion** (AIC) (Akaike, 1973) and **Bayesian Information Criterion** (BIC) (Schwarz, 1978) are based on the MLE and penalise model complexity. For model $M^{(m)}$ the AIC value is

$$\mathrm{AIC}_m = -2 \log p(\mathcal{D}|\hat{\boldsymbol{\psi}}^{(m)}, M^{(m)}) + 2n_\psi^m,$$

and the BIC value is

$$\mathrm{BIC}_m = -2 \log p(\mathcal{D}|\hat{\boldsymbol{\psi}}^{(m)}, M^{(m)}) + \log(n_\mathcal{D}) n_\psi^m,$$

in which m is the model index, $\hat{\boldsymbol{\psi}}^{(m)}$ is the MLE for the m-th model, and n_{ψ}^{m} is the total number of parameters of the m-th model, i.e. $\psi^{(m)} \in \mathbb{R}^{n_{\psi}^{m}}$. The total number of independent data points is denoted by $n_{\mathcal{D}}$. Low AIC and BIC values are favorable and a difference above 10 between AIC/BIC values of different models is considered as substantial (Burnham & Anderson, 2002). Note that none of these criteria provide information about the absolute quality of a model, but only the quality relative to other models. If prior information about the parameters is available, it can be treated as additional data points to still allow the use of BIC and AIC as they would be then considered to be part of the likelihood function. For this reason, they need also to be included in $n_{\mathcal{D}}$. In this thesis, we use the AIC and BIC for the model selection tasks presented in Chapter 3.

2.4.3. Likelihood-ratio test

Another method that is also based on the MLE is the **likelihood-ratio test** (Kirk et al., 2013). The models compared with this method need to be nested, in other words, the more complex model can be transformed into the simpler one by imposing parameter constraints on the complex model. It is a hypothesis test assuming as null hypothesis that the smaller model $M^{(1)}$ is the true model that generated the data, which is compared against the alternative hypothesis that the more complex model $M^{(2)}$ is the true model that generated the data. The test statistic is given by

$$\lambda_{LRT} = -2 \log \left(\frac{p(\mathcal{D}|\hat{\boldsymbol{\psi}}^{(1)}, M^{(1)})}{p(\mathcal{D}|\hat{\boldsymbol{\psi}}^{(2)}, M^{(2)})} \right)$$

with $\hat{\psi}^{(1)}$ and $\hat{\psi}^{(2)}$ denoting the MLE for $M^{(1)}$ and $M^{(2)}$ respectively. It often is used in its logarithmic form expressed as the difference between the log-likelihoods and it is asymptotically (under the null hypothesis) χ^2-distributed with degrees of freedom $n_{\psi}^{2} - n_{\psi}^{1}$ where n_{ψ}^{1} and n_{ψ}^{2} are the total number of parameters in $M^{(1)}$ and $M^{(2)}$ respectively. That means the λ_{LRT} can be tested for significance by comparing it to the corresponding χ^2-value of the desired statistical significance level.

Chapter 3

A mechanistic model of gastric cancer to describe multiple cell populations

This chapter is based on and in part identical to the following publication:

- **Raimúndez, E.**, Keller, S., Zwingenberger, G., Ebert, K., Hug, S., Theis, F. J., Maier, D., Luber, B., and Hasenauer, J. (2020). Model-based analysis of response and resistance factors of cetuximab treatment in gastric cancer cell lines. *PLoS Computational Biology*, 16(3):e1007147. doi:10.1371/journal.pcbi.1007147.

Mechanistic models based on ordinary differential equations (ODEs) are valuable tools to facilitate the quantitative understanding of biological processes. In particular, they enable the generation of new hypotheses and predictions, and they can be used for the design of new experiments (Altrock et al., 2015). In the context of cancer research, *in vitro* and *in vivo* models are dominant. However, in recent years the incorporation of *in silico* techniques based on mechanistic mathematical models has gained popularity and it has been shown to be beneficial (Bekisz & Geris, 2020; Jean-Quartier et al., 2018; Viceconti et al., 2016).

Several studies have applied mechanistic mathematical models for biomarker development (Kim & Schoeberl, 2015). Fey et al. (2015) predicted the survival of neuroblastoma patients using a mechanistic model of the c-Jun N-terminal kinase (JNK) pathway. Hass et al. (2017) predicted ligand dependence of solid tumors using a mechanistic multi-pathway model. Fröhlich et al. (2018) showed that large-scale mechanistic models facilitate the integration of large-scale data and enable the derivation and mechanistic interpretation of biomarkers. All these studies employ ODE models to describe the biochemical reaction networks involved in intracellular signal transduction. These models integrate (i) prior knowledge on the pathway topology derived over the last decades and available in databases, such as KEGG (Kanehisa et al., 2010), Reactome (Croft et al., 2011) and BioModels (Li et al., 2010), with (ii) heterogeneous experimental data for the process of interest. The

exploitation of prior knowledge constrains the search space and, in many cases, improves the predictive power. Furthermore, the chosen modeling approach facilitates the mechanistic interpretation of the findings.

In this chapter, we aim to provide a basis for understanding response and resistance mechanisms for cetuximab treatment in gastric cancer, to unravel likely causal differences between cetuximab responders and non-responders (Keller et al., 2017), and to possibly identify biomarkers for guiding targeted therapy by using a cell culture model.

3.1. Cetuximab in gastric cancer treatment

Gastric cancer is the fifth most common cancer and fourth leading cause of death from cancer worldwide (Sung et al., 2021). Treatment options include surgery, chemo- and radiation therapy. However, the overall survival rate remains unsatisfactory due to molecular and clinical heterogeneity (Lordick et al., 2014). Therefore, new treatment options are urgently required. Novel drugs targeting members of a family of receptor tyrosine kinases, including the epidermal growth factor receptor (EGFR), have shown mixed success in clinical trials (Bang et al., 2010; Lordick et al., 2013). Among others, the EGFR antibody cetuximab did not improve patient survival in a phase III clinical trial (Lordick et al., 2013).

EGFR is overexpressed in many cancer types and activated by a variety of ligands (Hass et al., 2017), such as the epidermal growth factor (EGF), the transforming growth factor-alpha (TGF-α), heparin-binding EGF-like growth factor (HBEGF), amphiregulin (AREG) and epiregulin (EREG). Although all these ligands bind to EFGR, they do not produce identical biological responses. These varying responses might be due to different ligand affinity, different ability to induce EGFR recycling or different ability to produce EGFR:ERBB2 heterodimers or EGFR:EGFR homodimers (Singh et al., 2016). Cetuximab is an antibody which binds EGFR (Li et al., 2005). The binding of cetuximab to EGFR inhibits the interaction with its ligands, thereby reducing survival and proliferation signals. In addition, cetuximab induces antibody-dependent cellular cytotoxicity by provoking immune cells to attack cancer (Hara et al., 2008).

A potential reason for the failure of cetuximab is the molecular heterogeneity of gastric cancer (Lordick et al., 2014). Due to this heterogeneity, only a small subgroup of patients might benefit from the targeted therapy. However, a suitable biomarker for patient stratification is currently not available. Conceptually, biomarkers for responsive patient subgroups can be identified using statistical approaches characterizing responder and non-responder subgroups within different molecular high-throughput methods. Unfortunately,

the necessary large-scale studies on the response of gastric cancer patients to cetuximab are missing. In addition, many proposed biomarkers from purely associative approaches have failed in clinical use (Poste, 2011). Even large-scale cancer cell line projects, such as the Cancer Cell Line Encyclopedia (CCLE) (Barretina et al., 2012) and the Genomics of Drug Sensitivity in Cancer (GDSC) (Yang et al., 2013) project, do not provide data for cetuximab response. Consequently, the limited amount of cell line and patient data prohibits the use of established statistical methods for biomarker development for cetuximab-responsive patient subgroups. Alternatively, mechanistic models can be employed.

In the following, we develop multiple candidate models for the EGFR, the Protein Kinase B (AKT) and the Extracellular-signal Regulated Kinase (ERK) signaling in cetuximab responder and non-responder cell lines. We calibrate the model collection using a comprehensive dataset and select the most appropriate model based on the AIC. We validate the optimal model candidate using newly collected data. To analyze the dependence of the cellular response on gene expression levels and (somatic) mutations, we interrogate the optimal model using simulation studies and *in silico* knockdown and knockout experiments. These simulation experiments suggest several intervention points and potential model-based biomarkers.

3.2. Experimental data

In this section, we describe the experimental data used to calibrate the models considered in this thesis chapter.

3.2.1. Data collection

Our experimental collaboration partners collected quantitative data from the human gastric cancer cell lines Hs746T and MKN1, which were used and cultured as reported by Keller et al. (2017). The data was obtained using the Western blot technique (Renart et al., 1979), i.e. measuring relative protein abundances. In an iterative process, published data (Keller et al., 2017; Kneissl et al., 2012) were complemented by newly collected data. The cells were treated for indicated times with 0.05, 0.1, 1, 10 or 50 μg/ml cetuximab and/or 5, 30 or 100 ng/ml EGF. Protein abundances were quantified for total EGFR, phosphorylated EGFR (pEGFR), phosphorylated ERK (pERK) and phosphorylated AKT (pAKT). This provided a time-resolved and dose reponse collection of 303 data points for the MKN1 cell line (Figure 3.1A) and 312 data points for the Hs746T cell line (Figure 3.1B).

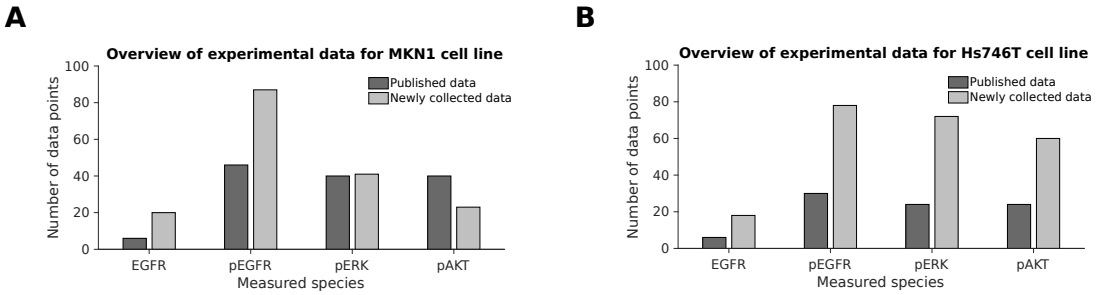

Figure 3.1.: The experimental data. Overview of the experimental data used for model calibration for (A) MKN1 and (B) Hs746T cell lines. The amount of data from the literature (published by Keller et al. (2017)) and newly collected data (published as part of this study) is distinguished. Adapted from Figure 1 of the thesis author's publication (Raimúndez et al., 2020).

To generate a transient knockdown of EGFR in the cell line MKN1 and a knockdown of MET in Hs746T cells, siRNA was used. The siRNA transfection was checked on protein level by Western blot analyses. Therefore, samples were collected 24 hours after the siRNA transfection. For analysis of EGFR, ERK and AKT activation after EGFR knockdown, cells were treated for 5 minutes with 5 ng/ml EGF or the combination of 1 μg/ml cetuximab plus 5 ng/ml EGF. For further details on the experimental methods and measurement techniques we refer the reader to Raimúndez et al. (2020).

The following sections present contributions made by the thesis author.

3.2.2. Data pre-processing

In Western blot data, as only intensity values on a single gel are comparable, the mapping of the data to the model simulation requires scaling constants as in (2.3) – see Loos et al. (2018) for a detailed discussion. Yet, due to the large number of Western blots, a large number of scaling constants would be necessary. These are often unknown and have to be inferred from the data on top of estimating the unknown model and noise parameters. Therefore, increasing further the dimensionality of the parameter estimation problem, which can cause convergence problems in the optimization and uncertainty analysis (especially when using MCMC sampling methods). In the field of optimization, the problem can be eliminated using novel hierarchical optimization methods (Loos et al., 2018; Schmiester et al., 2019). However, this approach is not supported by `Data2Dynamics` (Raue et al., 2015) – the modeling toolbox we used for this study.

Then, to reduce the number of scaling parameters, we pre-processed the experimental data by aligning replicates. Replicates are, in this case, the experimental data for different

Western blots capturing the same experimental conditions. Mathematically, we consider the experimental data

$$\bar{y}_{r,i} = s_r(\mu_i + \epsilon_{r,i}), \qquad \epsilon_{r,i} \sim \mathcal{N}(0, \sigma^2)$$

with $r = 1, \ldots, R$ indexing the replicate and $i = 1, \ldots, I$ indexing the experimental condition. The experimental condition is a combination of cell line, treatment and time point. The true level of a protein or phosphorylated protein in experimental condition i is denoted by μ_i, and corrupted by measurement noise $\epsilon_{r,i}$. The scaling constants associated to different replicates (i.e. Western blot gels) are denoted by s_r. As $\bar{y}_{1,i}$ and $\bar{y}_{2,i}$ should agree up to measurement noise and scaling constant, we inferred estimates for all but one scaling constant by solving the optimization problem

$$\begin{aligned} &\underset{s_2,\ldots,s_I,\sigma}{\operatorname{argmin}} \left(-\log(\mathcal{L}) := \frac{1}{2} \sum_{i=1}^{I} \log(2\pi\sigma^2) \left(\frac{\frac{1}{s_r}\bar{y}_{r,i} - \hat{\mu}_i}{\sigma} \right)^2 \right), \\ &\text{subject to} \quad \hat{\mu}_i = \sum_{r=1}^{R} \frac{1}{s_r} \bar{y}_{r,i}, \quad \text{and} \quad i = 1, \ldots, I \end{aligned}$$

with $s_1 = 1$ for $r = 1$. The result of the optimization are the pre-processed data $\hat{\mu}_i$ and the estimated standard deviation σ. The pre-processed data $\hat{\mu}_i$ aggregated the information of the replicates by compensating for the individual scales. Without loss of generality, the scale of the first replicate was considered as reference ($s_1 = 1$), hence, $\mu_i \approx s_i \hat{\mu}_i$ (up to measurement noise).

The pre-processing reduced the number of scaling constants for an experimental dataset from R to 1 and renders the optimization problem more tractable. This scaling constant effectively corresponds to the scaling constant for the first replicate s_1.

3.3. Mathematical model of individual cell lines

In this section, we introduce the pathway model we developed to describe the dynamics of the cetuximab responder and non-responder cell line. Firstly, we introduce the biology underlying the model. Secondly, we enumerate the considered model reactions as well as the parameter priors found from literature. Lastly, we calibrate the model individually for the responder and non-responder cell line. We found that the proposed model is a good candidate as the model provided a good description of most data points.

3.3.1. Description of the EGFR signaling pathway

We utilized the gastric cancer cell lines MKN1 and Hs746T as model systems to study response and resistance factors of cetuximab treatment. Based on previous results obtained by proliferation and motility analysis, the MKN1 cell line is characterized as a cetuximab responder, while Hs746T is characterized as a non-responder cell line (Keller et al., 2017; Kneissl et al., 2012). Different factors might contribute to this difference, including a 2.5 fold higher EGFR expression in MKN1 cells compared to Hs746T (Kneissl et al., 2012). Furthermore, MKN1 cells expresses the Phosphoinositide 3-Kinase (PI3K) mutant PI3K p.E545K and in Hs746T cells MET is mutated to MET p.L982_D1028del and amplified (den Dunnen et al., 2016). PI3K p.E545K (MPI3K) possesses an increased catalytic activity compared to wild-type PI3K (Kang et al., 2005) resulting in enhanced downstream signaling, while MET p.L982_D1028del (MMET) is assumed to be constitutively active, i.e. independent of its ligand.

Cetuximab targets the EGFR signaling pathway, which regulates growth, survival, proliferation, differentiation (Oda et al., 2005) and motility (Keller et al., 2017; Wells, 1999). Upon ligand binding, EGFR homodimerizes and auto-phosphorylates, promoting its catalytic activity. pEGFR is internalized and degraded or recycled. Membrane-bound and internalized pEGFR catalyzes the activation of the G-protein RAS and PI3K by phosphorylation. While RAS activates the mitogen-activated (MAPK) signaling pathway resulting in the phosphorylation of ERK, PI3K activity indirectly results in the phosphorylation of AKT. Both, pERK and pAKT are important regulators of DNA transcription.

Cetuximab binds EGFR and blocks the binding of EGF or other EGFR ligands, such as AREG and EREG (Li et al., 2005). This reduces – in the absence of resistance factors – the activity of EGFR and its downstream targets. Known resistance factors to EGFR-targeted therapies in other cancer types include mutations and overexpression of the receptor tyrosine kinases AXL (Hrustanovic et al., 2013) or MET (Zhao et al., 2017).

A visualization of the EGFR signaling pathway accounting for common adaptor proteins, cetuximab, as well as the aforementioned mutants of PI3K and MET is provided in Figure 3.2A. Based on this pathway information and the available literature on published ODE models (Fröhlich et al., 2018; Fujita et al., 2010; Hass et al., 2017; Sasagawa et al., 2005; Schöberl et al., 2009) and logical models (Flobak et al., 2015) – in particular the work of Sasagawa et al. (2005) and Fujita et al. (2010) – we constructed a simplified pathway model using systems of ODEs as introduced in Section 2.1.1. To limit the model complexity, adaptor proteins and multi-site phosphorylation of EGFR and ERK are disregarded, and

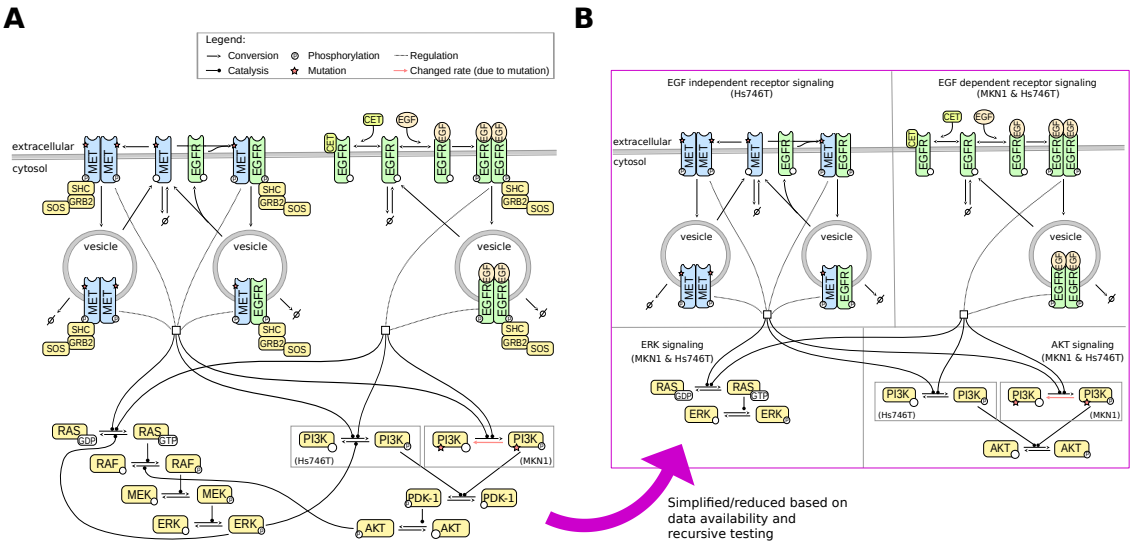

Figure 3.2.: The pathway model. (A) Illustration of the EGFR/MMET signaling pathway, focusing on the MAPK and AKT signaling cascades, used to build the mathematical model. (B) Illustration of the mathematical model indicating shared and cell line specific biochemical species/reactions. Adapted from Figure 1 of the thesis author's publication (Raimúndez et al., 2020).

the intermediate steps in the MAPK signaling pathway are lumped into a single reaction. The simplified pathway description is visualized in Figure 3.2B.

3.3.2. Model reactions

Here, the biochemical reactions implemented in the proposed pathway model (Figure 3.2B) are reported. After discussing with our experimental collaboration partners and building on previously published models (Fujita et al., 2010; Sasagawa et al., 2005), we grouped the model reactions into 5 modules: (i) EGF receptor dynamics, (ii) MMET receptor dynamics, (iii) EGFR:MMET heterodimer dynamics, (iv) RAS-MAPK downstream signaling, and (v) PI3K- and MPI3K-AKT downstream signaling. The parts that are exclusive for Hs746T cells are the MMET receptor dynamics and EGFR:MMET heterodimer dynamics. Moreover, PI3K is only expressed in Hs746T cells, while MPI3K is expressed only in MKN1 cells. The remaining reactions and biochemical species are common for both cell lines. All model reactions follow mass action kinetics.

We further classified each model reaction into (i) summarizing multi-step signaling processes and (ii) describing direct reactions. Examples of multi-step signaling processes include the indirect activation of ERK by RAS or of AKT by MPI3K/PI3K (Figure 3.2A), which

is described by one single reaction in the considered pathway model (Figure 3.2B). In contrast, examples of direct reactions are the binding of EGF to EGFR. This classification becomes relevant in Section 3.4 where modeling of cell line differences is considered.

EGF receptor dynamics

R1: $\emptyset \rightarrow$ EGFR
R2: EGFR $\rightarrow \emptyset$ $(*, \text{membrane turnover})$
R3: EGFR + CET $\rightarrow$ EGFR:CET
R4: EGFR:CET $\rightarrow$ EGFR + CET
R5: EGFR:CET $\rightarrow \emptyset$ $(*, \text{membrane turnover})$
R6: EGFR + EGF $\rightarrow$ EGFR:EGF
R7: EGFR:EGF $\rightarrow$ EGFR + EGF
R8: EGFR:EGF + EGFR:EGF $\rightarrow$ $(\text{EGFR:EGF})_2$
R9: $(\text{EGFR:EGF})_2 \rightarrow (\text{pEGFR:EGF})_2$
R10: $(\text{pEGFR:EGF})_2 \rightarrow (\text{pEGFR:EGF})_{2,\text{endocytic vesicle}}$ $(*)$
R11: $(\text{pEGFR:EGF})_{2,\text{endocytic vesicle}} \rightarrow$ EGFR + EGFR $(*)$
R12: EGFR:EGF $\rightarrow \emptyset$ $(*, \text{membrane turnover})$
R13: $(\text{EGFR:EGF})_2 \rightarrow \emptyset$ $(*, \text{membrane turnover})$
R14: $(\text{pEGFR:EGF})_{2,\text{endocytic vesicle}} \rightarrow \emptyset$ $(*)$

This module consists of 14 reactions describing – among other things – synthesis, ligand and cetuximab (CET) binding, dimerization, internalization and degradation of EGFR. In total 7 out of 14 reactions summarize multi-step processes, denoted by $(*)$. For the dynamics of R2, R5, R12 and R13, we assumed a common kinetic rate constant describing the membrane turnover. For R1, we considered in the EGFR silencing experiments that the synthesis rate of EGFR is multiplied with the fold-change observed after siRNA treatment (see Figure S12 from the thesis author's publication (Raimúndez et al., 2020)).

MMET receptor dynamics

R15: $\emptyset \rightarrow$ MMET
R16: MMET $\rightarrow \emptyset$
R17: MMET + inhibitor $\rightarrow$ MMET:inhibitor
R18: MMET:inhibitor $\rightarrow$ MMET + inhibitor
R19: MMET:inhibitor $\rightarrow \emptyset$
R20: MMET + MMET $\rightarrow \text{MMET}_2$
R21: $\text{MMET}_2 \rightarrow \text{pMMET}_2$
R22: $\text{pMMET}_2 \rightarrow \text{pMMET}_{2,\text{endocytic vesicle}}$

R23: $\text{pMMET}_{2,\text{endocytic vesicle}} \rightarrow \text{MMET} + \text{MMET}$
R24: $\text{MMET}_2 \rightarrow \emptyset$
R25: $\text{pMMET}_{2,\text{endocytic vesicle}} \rightarrow \emptyset$
R26: $\text{MMET} + \text{MMET:inhibitor} \rightarrow \text{MMET}_2\text{:inhibitor}$
R27: $\text{MMET:inhibitor} + \text{MMET:inhibitor} \rightarrow (\text{MMET:inhibitor})_2$
R28: $\text{MMET}_2 + \text{inhibitor} \rightarrow \text{MMET}_2\text{:inhibitor}$
R29: $\text{MMET}_2\text{:inhibitor} + \text{inhibitor} \rightarrow (\text{MMET:inhibitor})_2$
R30: $\text{MMET}_2\text{:inhibitor} \rightarrow \emptyset$
R31: $(\text{MMET:inhibitor})_2 \rightarrow \emptyset$

This module consists of 16 reactions describing – among other things – synthesis, ligand and MET inhibitor binding, dimerization, internalization and degradation of MMET. This signaling module is specific for Hs746T cells since MMET is not expressed in MKN1 cells. For R15, we considered in the MMET silencing experiments that the synthesis rate of MMET is multiplied with the fold-change observed after siRNA treatment (see Figure S11 from the thesis author's publication (Raimúndez et al., 2020)). For R17, R28 and R29, we considered that the K_D value of the MET inhibitor is known from Park et al. (2016) and the inhibitor concentrations are set in the experiments.

EGFR:MMET heterodimer dynamics

R32: $\text{MMET} + \text{EGFR} \rightarrow \text{MMET:EGFR}$
R33: $\text{MMET:EGFR} \rightarrow \text{pMMET:pEGFR}$
R34: $\text{pMMET:pEGFR} \rightarrow \text{pMMET:pEGFR}_{\text{endocytic vesicle}}$
R35: $\text{pMMET:pEGFR}_{\text{endocytic vesicle}} \rightarrow \text{MMET} + \text{EGFR}$
R36: $\text{MMET:EGFR} \rightarrow \emptyset$
R37: $\text{pMMET:pEGFR}_{\text{endocytic vesicle}} \rightarrow \emptyset$
R38: $\text{MMET:inhibitor} + \text{EGFR} \rightarrow \text{EGFR:MMET:inhibitor}$

This module consists of 7 reactions describing the dynamics for the EGFR and MMET heterodimers. This signaling module is specific for Hs746T cells since MMET is not expressed in MKN1 cells.

RAS-MAPK downstream signaling

R39: $\text{RAS-GDP} \rightarrow \text{RAS-GTP}\ (*)$
R40: $\text{RAS-GDP} + (\text{pEGFR:EGF})_2 \rightarrow \text{RAS-GTP} + (\text{pEGFR:EGF})_2\ (*, \text{EGFR induced activation})$
R41: $\text{RAS-GDP} + (\text{pEGFR:EGF})_{2,\text{endocytic vesicle}} \rightarrow \text{RAS-GTP} + (\text{pEGFR:EGF})_{2,\text{endocytic vesicle}}$
$(*, \text{EGFR induced activation})$

R42: RAS-GDP + pMMET:pEGFR → RAS-GTP + pMMET:pEGFR
R43: RAS-GDP + $\text{pMMET:pEGFR}_{\text{endocytic vesicle}}$ → RAS-GTP + $\text{pMMET:pEGFR}_{\text{endocytic vesicle}}$
R44: RAS-GDP + pMMET_2 → RAS-GTP + pMMET_2
R45: RAS-GDP + $\text{pMMET}_{2,\text{endocytic vesicle}}$ → RAS-GTP + $\text{pMMET}_{2,\text{endocytic vesicle}}$
R46: RAS-GTP → RAS-GDP (∗)
R47: ERK + RAS-GTP → pERK + RAS-GDP (∗)
R48: pERK → ERK

This module consists of 10 reactions describing – among other things – basal and receptor induced activation (including EGFR, MMET and their association), and downstream signaling of RAS. We excluded the reactions involving the biochemical species pMMET_2 and pMMET:pEGFR – also their corresponding internalized forms – as they are only expressed in Hs746T cells. This leads to a total of 5 out of 10 reactions summarizing multi-step processes, denoted by (∗). For the dynamics of R10 and R41, we assumed a common kinetic rate constant describing the EGFR induced activation of RAS.

PI3K- and MPI3K-AKT downstream signaling

R49: PI3K → pPI3K
R50: PI3K + $(\text{pEGFR:EGF})_2$ → pPI3K + $(\text{pEGFR:EGF})_2$
R51: PI3K + $(\text{pEGFR:EGF})_{2,\text{endocytic vesicle}}$ → pPI3K + $(\text{pEGFR:EGF})_{2,\text{endocytic vesicle}}$
R52: PI3K + pMMET:pEGFR → pPI3K + pMMET:pEGFR
R53: PI3K + $\text{pMMET:pEGFR}_{\text{endocytic vesicle}}$ → pPI3K + $\text{pMMET:pEGFR}_{\text{endocytic vesicle}}$
R54: PI3K + pMMET_2 → pPI3K + pMMET_2
R55: PI3K + $\text{pMMET}_{2,\text{endocytic vesicle}}$ → pPI3K + $\text{pMMET}_{2,\text{endocytic vesicle}}$
R56: pPI3K → PI3K (∗, †)
R57: MPI3K → pMPI3K
R58: MPI3K + $(\text{pEGFR:EGF})_2$ → pMPI3K + $(\text{pEGFR:EGF})_2$
R59: MPI3K + $(\text{pEGFR:EGF})_{2,\text{endocytic vesicle}}$ → pMPI3K + $(\text{pEGFR:EGF})_{2,\text{endocytic vesicle}}$
R60: MPI3K + pMMET:pEGFR → pMPI3K + pMMET:pEGFR
R61: MPI3K + $\text{pMMET:pEGFR}_{\text{endocytic vesicle}}$ → pMPI3K + $\text{pMMET:pEGFR}_{\text{endocytic vesicle}}$
R62: MPI3K + pMMET_2 → pMPI3K + pMMET_2
R63: MPI3K + $\text{pMMET}_{2,\text{endocytic vesicle}}$ → pMPI3K + $\text{pMMET}_{2,\text{endocytic vesicle}}$
R64: pMPI3K → MPI3K (∗, †)
R65: AKT + pPI3K → pAKT + pPI3K (∗, PI3K/MPI3K induced activation)
R66: AKT + pMPI3K → pAKT + pMPI3K (∗, PI3K/MPI3K induced activation)
R67: pAKT → AKT (∗)

This module consists of 19 reactions describing – among other things – basal and receptor induced activation, and downstream signaling of PI3K, which is expressed in Hs746T cells, and MPI3K, which is expressed in MKN1 cells. All kinetic rate constants defined for PI3K and MPI3K are common between the cell lines, except for the inactivation rate – due to PI3K mutation in MKN1 cells – indicated by (†). We excluded the reactions involving the biochemical species $pMMET_2$ and pMMET:pEGFR – also their corresponding internalized forms – as they are only expressed in Hs746T cells. This leads to a total of 4 out of 19 reactions summarizing multi-step processes, denoted by (∗). We assumed a common kinetic rate constant describing the PI3K (and MPI3K) induced activation of AKT.

The model accounts for 27 biochemical species. Protein abundances for RAS, ERK, PI3K, MPI3K and AKT were assumed to be conserved, which results in an ODE model with 22 state variables. As the absolute abundances of the conserved species are unknown, we consider as state the relative abundance with the concentration in MKN1 cells and rich culture media as reference point for RAS, ERK, MPI3K and AKT. As PI3K is not represented in MKN1 cells, we used the concentration in Hs746T cells and rich culture media as reference point. The changes in expression levels between the different culture media (rich and starvation medium) were estimated as proposed by Steiert et al. (2016) (cf. Section 3.4.1). All these, together with reaction rates and scaling constants, lead to a total of 57 unknown parameters for MKN1 cells and 115 for Hs746T cells. As initial conditions for the simulations of time and dose response data, we employed the steady state of the unperturbed system. This steady state is specific to the combination of cell line and culture medium.

For a more detailed description – including reaction rates – refer to the SBML model available in the `PEtab` benchmark collection (`https://github.com/Benchmarking-Initiative/Benchmark-Models-PEtab`).

3.3.3. Encoding available information about model parameters

We screened the literature for the considered model parameters, in particular for the kinetic rate constants. Using the obtained information, we defined prior distributions for individual rate constants (Table 3.1). As priors, we considered Gaussian distributions of the $\log_{10}$-transformed parameter values $\theta_{\log} := \log_{10}(\theta)$

$$p(\theta_{\log}) = \frac{1}{\sqrt{2\pi}\hat{\sigma}} \exp\left(-\frac{1}{2}\left(\frac{\theta_{\log} - \hat{\theta}_{\log})}{\hat{\sigma}}\right)^2\right),$$

Table 3.1.: Literature values used as prior information in the parameter estimation. The reported mean values were $\log_{10}$-transformed and used as priors for the parameterization of the model. Dissociation constants are denoted by K_{D} and forward reaction rates by k_f. Note that the same kinetic rate constant can appear in different reactions, e.g., receptor endocytosis in R10, R22 and R34. Formation of endocytic vesicles is denoted in short by "end. ves.". Adapted from Table S2 of the author's publication (Raimúndez et al., 2020).

Reaction	Mean value	Reference
R6, R7: EGFR + EGF $\rightleftharpoons$ EGFR:EGF	$K_{\mathrm{D}} = 2$ nM	Klein et al. (2004) (page 1)
R3, R4: EGFR + CET $\rightleftharpoons$ EGFR:CET	$K_{\mathrm{D}} = 0.39$ nM	Kim & Grothey (2008) (page 2)
R10: $(\mathrm{pEGFR{:}EGF})_2 \rightarrow (\mathrm{pEGFR{:}EGF})_{2,\text{end. ves.}}$ R22: $\mathrm{pMMET}_2 \rightarrow \mathrm{pMMET}_{2,\text{end. ves.}}$ R34: $\mathrm{pMMET{:}pEGFR} \rightarrow \mathrm{pMMET{:}pEGFR}_{\text{end. ves.}}$	$k_f = 0.25\ \mathrm{min}^{-1}$	Sorkin & Duex (2010) (page 19)
R67: pAKT $\rightarrow$ AKT	$k_f = (1.5 \times 60)\ \mathrm{min}^{-1}$	Schöberl et al. (2009) (kinetic rate constant $k_f 83$)

with $\hat{\theta}_{\log}$ denoting the mean and $\hat{\sigma}$ the standard deviation of the Gaussian-distributed prior. The Gaussian priors were centered around, i.e. their mean is equal to, the corresponding $\log_{10}$-transformed literature values (Table 3.1) with a standard deviation of 0.2. For the frequentist approaches used here for parameter estimation, these available estimates are treated as additional data points.

3.3.4. Model calibration

We calibrated the pathway model using a comprehensive dataset obtained using quantitative immunoblotting. The dataset contains time and dose responses for EGFR, pEGFR, pERK and pAKT. We considered 303 data points for the MKN1 cell line (Figure 3.1A) and 312 data points for the Hs746T cell line (Figure 3.1B).

The concentrations of biochemical species were linked to the observables as introduced in Section 2.1.2. As Western blot measurements provide relative information about the absolute molecular state, the model observables correspond to (2.3). To adapt the scale of different replicates and standard deviation of the experimental data, we pre-processed the experimental data employing mixture modeling (see Section 3.2.2). We assumed additive Gaussian-distributed noise with noise parameters fixed to the estimated standard deviations obtained during the data pre-processing step and therefore further reduced the complexity of the optimization problem (Loos et al., 2018).

We determined the maximum likelihood estimates of all unknown model parameters, including model and observable parameters, using multi-start local optimization (Section 2.2.2). The estimation was performed individually for MKN1 and Hs746T cells. We generated 1000 starting points for local optimization using latin hypercube sampling. Local optimization was performed using a trust-region based algorithm (see Section 2.2.2) implemented in the MATLAB least-square solver `lsqnonlin`, which exploits the least-squares like structure of the optimization problems (Coleman & Li, 1996), interfaced by the `Data2Dynamics` toolbox.

For MKN1 cells, the optimizer converged reliably. This is reflected in the clear plateau in the waterfall plot (Figure 3.3A). For the optimal parameters we observed a good agreement of MKN1 data and model fit (Figure 3.3B-D) (Pearson correlation $\rho = 0.95$). For Hs746T cells, the results of the multi-start optimization do not show clear plateaus, suggesting that the objective function possesses a large number of local optima or flat regions (Figure 3.4A). But even so, the best fit found provides an accurate description of the data (Figure 3.4B-D) ($\rho = 0.91$) and the best 100 fits achieve similar correlations (Annex Figure B.1). Furthermore, the best 10 fits are statistically not distinguishable from the best fit (Hross & Hasenauer, 2016). A potential source of the convergence problems for Hs746T cells is the low signal-to-noise ratio, which is caused by the limited response to EGF and cetuximab treatment (see, e.g., Figure 3.4C-D) since this is a non-responder cell line. Yet, despite the low signal-to-noise ratio, several changes are statistically significant (see analysis by Keller et al. (2017) for a part of the published data). As the model fits provided a good description of most data points, the parameter estimation confirmed that the proposed pathway model is a good candidate for further analysis.

3.4. Integrated mathematical model of multiple cell lines

In this section, we extend the pathway model to simultaneously describe the dynamics of the considered cell lines. We start with the description of how dynamical differences between the cell lines are encoded in the model. Afterwards, we build a collection of various model hypotheses and calibrate them to the available data to unravel causal differences between the cell lines. Using model selection, we found that the extended model with the most evidence considers cell line differences in the EGFR turnover dynamics.

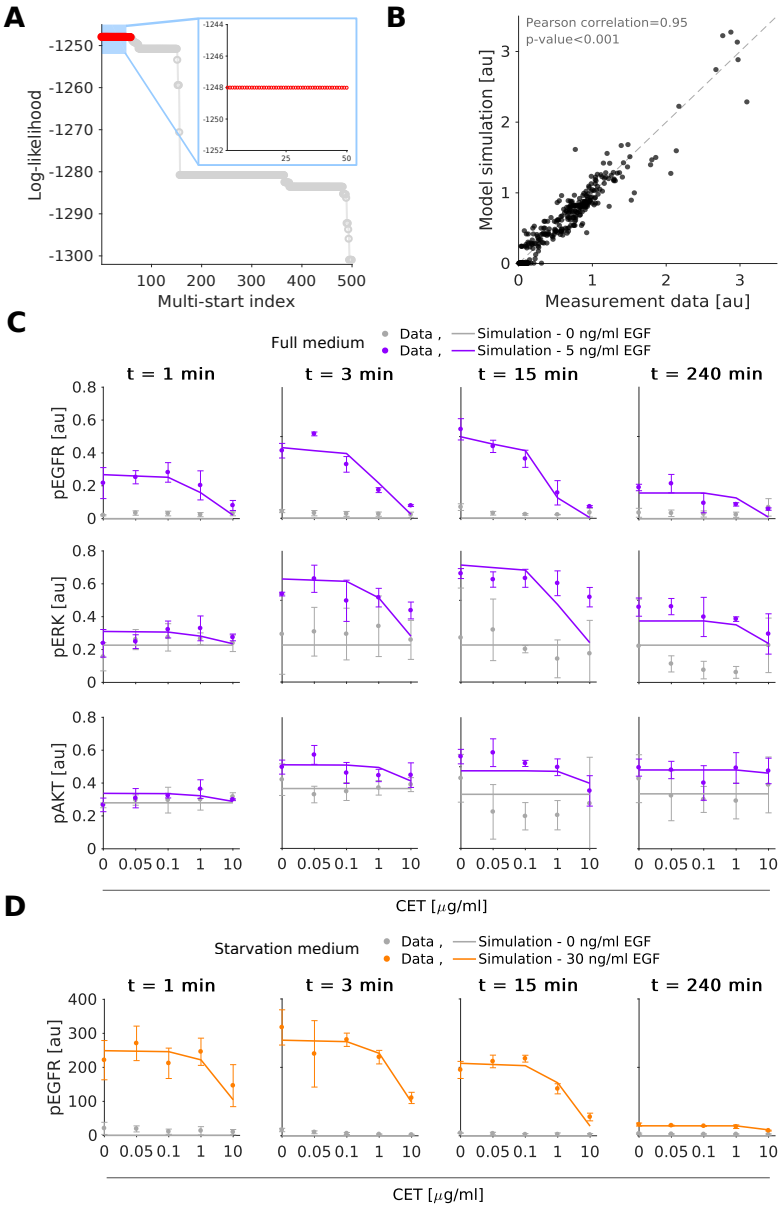

Figure 3.3.: Experimental data and model fit for the gastric cancer cell line MKN1. (A) Waterfall plot for multi-start local optimization. The best 500 out of 1000 runs are depicted from which a magnification of the best 50 multi-starts is indicated by the blue box. Red dots denote the starts converged to the global optimum within a small numerical margin. (B) Scatter plot for the overall agreement of experimental data and model fit. (C-D) Comparison of selected experimental data and model fits. Time and dose response data obtained using immunoblotting indicate the mean and standard deviation of three biological experiments. For visualization in (C) and (D), experimental measurements were scaled to model simulation using the estimated scaling constants to overlay the response to different experimental conditions. Adapted from Figure 2 of the thesis author's publication (Raimúndez et al., 2020). Additional data and model fits are provided in Annex Figure B.2.

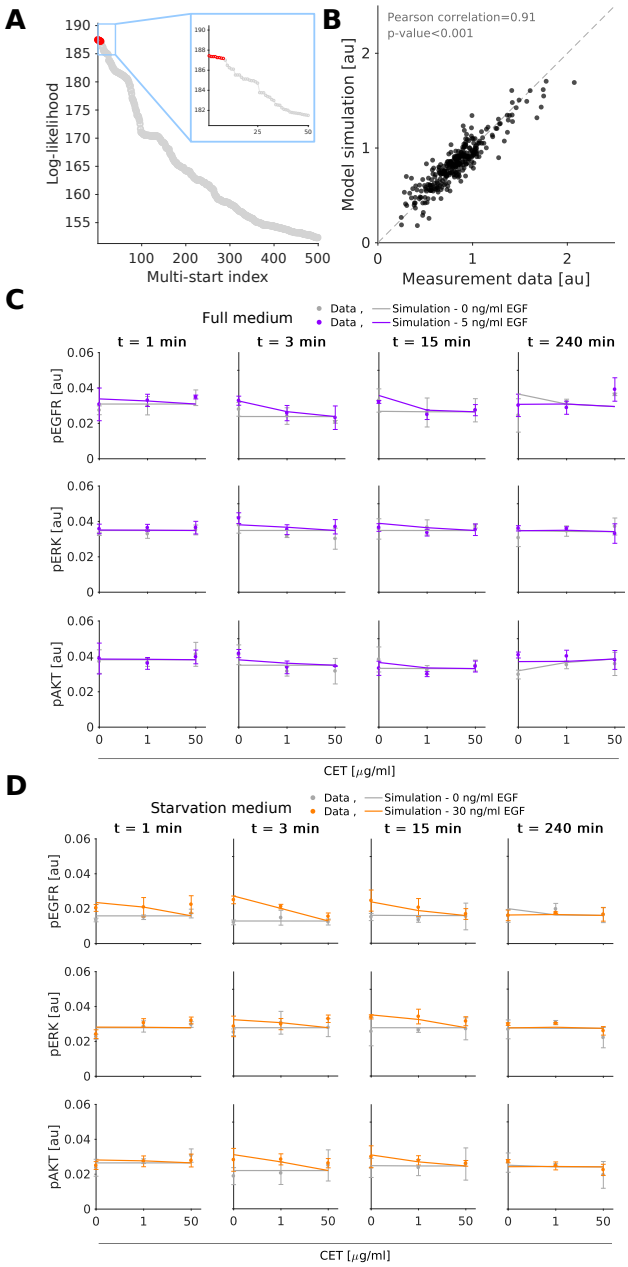

Figure 3.4.: Experimental data and model fit for the gastric cancer cell line Hs746T. (A) Waterfall plots for multi-start local optimization. The best 500 out of 2000 runs are depicted from which a magnification of the best 50 multi-starts is indicated by the blue box. Red dots denote the starts converged to the global optimum within a small numerical margin. (B) Scatter plot for the overall agreement of experimental data and model fits. (C-D) Comparison of selected experimental data and model fit. Time and dose response data obtained using immunoblotting indicate the mean and standard deviation of three biological experiments. For visualization in (C) and (D), experimental measurements were scaled to model simulation using the estimated scaling constants to overlay the response to different experimental conditions. Adapted from Figure 3 of the thesis author's publication (Raimúndez et al., 2020). Additional data and model fits are provided in Annex Figure B.3.

3.4.1. Encoding cell line differences into a single model

To infer causal differences between responder (MKN1) and non-responder (Hs746T) cell lines, we allowed for differences between the model parameters, i.e. rate constants and initial conditions, of the cell lines. Following the work of Steiert et al. (2016), we modelled the difference by the additional parameters $\boldsymbol{\phi_j}$,

$$\theta_{j,\text{Hs746T}} = \exp(\boldsymbol{\phi_j}) \cdot \theta_{j,\text{MKN1}}$$

with $\boldsymbol{\phi_j}$ denoting the log-fold change for the j-th model parameter. If $\boldsymbol{\phi_j}$ is zero, the parameters of the two cell lines are identical. We fixed $\boldsymbol{\phi_j}$ to zero for all model parameters which are associated with direct biochemical interactions, e.g. binding rates, which should be conserved between cell lines. Only for expression levels and indirect interactions, i.e. simplification of multi-step processes, we allowed for $\boldsymbol{\phi_j} \neq 0$ and therefore estimated $\boldsymbol{\phi_j}$ along with all unknown parameters $\boldsymbol{\psi} = (\boldsymbol{\psi}, \boldsymbol{\phi})$. On top of modeling differences between cell lines, we maintained the differences between cell culture media (Section 3.3.2).

3.4.2. Finding causal differences between two cell lines beyond mutations and expression patterns

The available experimental data (Figures 3.3 and 3.4, and Annex Figures B.2 and B.3) showed a pronounced difference in the response of the cell lines to EGF and cetuximab treatment. Potential sources of this behavior are differences in mutation patterns, protein expression levels/abundances and reaction fluxes (due to an influence of latent components between the cell lines). While we did not perform a comprehensive quantification of the protein abundances, selected measurements and transcriptome data indicated substantial differences between the cell lines.

To identify important differences in the reaction fluxes, we compared the parameter estimates obtained for the individual cell lines. These parameter estimates should reflect changes on the biochemical level between the cell lines. In the comparison, we only considered parameters associated with single model reactions summarizing multi-step signaling processes. In contrast, for direct reactions the reaction rate constant should be identical for both cell lines. For details on the classification of the reactions we refer to Section 3.3.2.

For the comparison we considered the best 100 parameter vectors obtained by multi-start local optimization for MKN1 and Hs746T cell lines (Annex Figure B.1). Statistical testing for cell line specificity of parameters was performed using one-way-ANOVA suggesting

Table 3.2.: Reaction rate constant candidates for cell line specificity. The following kinetic rate constants were identified as significantly different between the cell lines by the ANOVA statistical test. All kinetic rate constants belonging to the same module were simultaneously tested (not individually) resulting in 8 different model candidates. Formation of endocytic vesicles is denoted in short by "end. ves.". Adapted from Table S1 of the author's publication (Raimúndez et al., 2020).

Module	Reaction	Description
EGFR turnover	R10: $(\text{pEGFR:EGF})_2 \rightarrow (\text{pEGFR:EGF})_{2,\text{end. ves.}}$	Endocytosis of pEGFR
	R14: $(\text{pEGFR:EGF})_{2,\text{end. ves.}} \rightarrow \emptyset$	Lysosomal degradation of endocyted pEGFR
	R2: $\text{EGFR} \rightarrow \emptyset$ R5: $\text{EGFR:CET} \rightarrow \emptyset$ R12: $\text{EGFR:EGF} \rightarrow \emptyset$ R13: $(\text{EGFR:EGF})_2 \rightarrow \emptyset$	Basal membrane turnover
	R11: $(\text{pEGFR:EGF})_{2,\text{end. ves.}} \rightarrow \text{EGFR} + \text{EGFR}$	EGFR recycling endosomes
RAS-MAPK signaling	R39: $\text{RAS-GDP} \rightarrow \text{RAS-GTP}$	Basal activation of RAS
	R40: $\text{RAS-GDP} + (\text{pEGFR:EGF})_2 \rightarrow$ $\rightarrow \text{RAS-GTP} + (\text{pEGFR:EGF})_2$ R41: $\text{RAS-GDP} + (\text{pEGFR:EGF})_{2,\text{end. ves.}} \rightarrow$ $\rightarrow \text{RAS-GTP} + (\text{pEGFR:EGF})_{2,\text{end. ves.}}$	EGFR dependent activation of RAS
	R46: $\text{RAS-GTP} \rightarrow \text{RAS-GDP}$	Basal GTPase activity of RAS
	R47: $\text{ERK} + \text{RAS-GTP} \rightarrow \text{pERK} + \text{RAS-GDP}$	RAS dependent activation of MAPK
PI3K-AKT signaling	R65: $\text{AKT} + \text{pPI3K} \rightarrow \text{pAKT} + \text{pPI3K}$ R66: $\text{AKT} + \text{pMPI3K} \rightarrow \text{pAKT} + \text{pMPI3K}$	PI3K and MPI3K dependent activation of AKT
	R67: $\text{pAKT} \rightarrow \text{AKT}$	Basal dephosphorylation of pAKT
	R56: $\text{pPI3K} \rightarrow \text{PI3K}$ R64: $\text{pMPI3K} \rightarrow \text{MPI3K}$	Basal dephosphorylation of pPI3K and pMPI3K

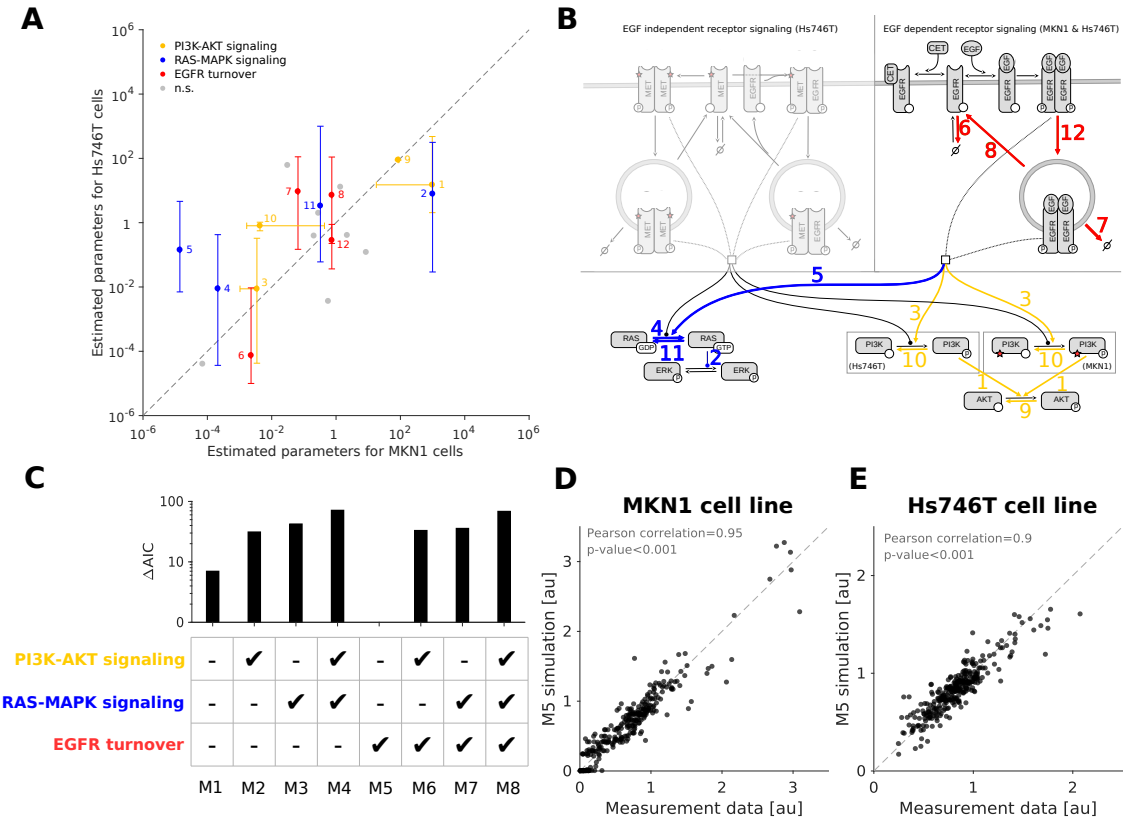

Figure 3.5.: Identification of cell line specific parameters. (A) RAS-MAPK and PI3K-AKT signaling pathways as well as EGFR turnover dynamics as possible cell line specificity candidates. Dots and bars depict the median and one standard deviation (68% percentile interval) of the top 100 $\log_{10}$-transformed parameter estimates obtained by the optimization. Values obtained from MKN1 and Hs746T cells are shown along the X and Y axis, respectively. Coloring indicates significantly different parameter pairs ($p < 0.05$). Gray points indicate non-significantly different and direct reactions (n.s.). (B) Highlighted cell line specificity candidates in the model overview. Coloring and numeric labeling correspond to Figure 3.5A. (C) Model selection using AIC shows that only changing the receptor turnover dynamics results in the best model. (D,E) Scatter plot for the overall agreement of experimental data and combined model fit for (D) the MKN1 cell line and (E) the Hs746T cell line. Results correspond to the best model (M5). Adapted from Figure 4 of the thesis author's publication (Raimúndez et al., 2020). The fits for the selected model (M5) are provided in Figure 3.6.

significant ($p < 0.05$) differences for 12 out of 20 estimated kinetic rate constants (Figure 3.5A). The mapping of the findings on the pathway visualization revealed potential differences in (I) RAS-MAPK signaling, (II) PI3K-AKT signaling, and (III) EGFR turnover including internalization, degradation and recycling (Figure 3.5B). A detailed list of the parameters and reactions is provided in Table 3.2. In addition, mutated MET and mutated PI3K can cause differences between cell lines as each is only present in one of the cell lines.

The comparison of the optimal parameter values found for the different cell lines provides a conservative evaluation. Indeed, not all differences between cell lines might be required to fit the data (see Steiert et al. (2016)). To assess the relevance of differences in modules (I)–(III), we fitted the datasets for the MKN1 and Hs746T cell lines simultaneously. Therefore, a collection of mathematical models was developed accounting for cell line specific mutations and protein abundances, as well as all possible combinations of differences of parameters belonging to modules (I), (II) and (III). All remaining kinetic rate constants were assumed to be identical for the cell lines. In total, eight different models were constructed accounting for all possible combinations. For each candidate model, a multi-start local optimization run was performed using the complete dataset and 100 starting points. The combination of molecular data for responder and non-responder cell lines provides additional constraints for the parameters and increases the number of data points per estimated parameter.

Model selection using AIC (Section 2.4.2), indicated that the model including cell line specific differences in only the EGFR turnover dynamics provides the best qualitative description of the experimental data (Figure 3.5C). This agrees with previous findings reporting the presence of FBXW7 p.R465C mutation in MKN1 cells, but not in Hs746T cells. FBXW7 ubiquitinates EGFR, leading to changes in the turnover/degradation of EGFR between the cell lines (Heindl et al., 2012). The resulting model provides an accurate description of the experimental data for the responder (Figure 3.5D) and the non-responder (Figure 3.5E) cell lines. Details on the model fit for the selected model are provided in Figure 3.6 and Annex Figures B.4 and B.5.

3.4.3. Testing additional model structures

We evaluated if the fitting of the best candidate model (M5 in Figure 3.5C) can be improved by including additional processes such as the negative feedback regulation of RAS activity by phosphorylated ERK (Lake et al., 2016) (Figure 3.7A). As the negative feedback is mediated by decreasing the activation of RAS, we can model this feedback by redefining

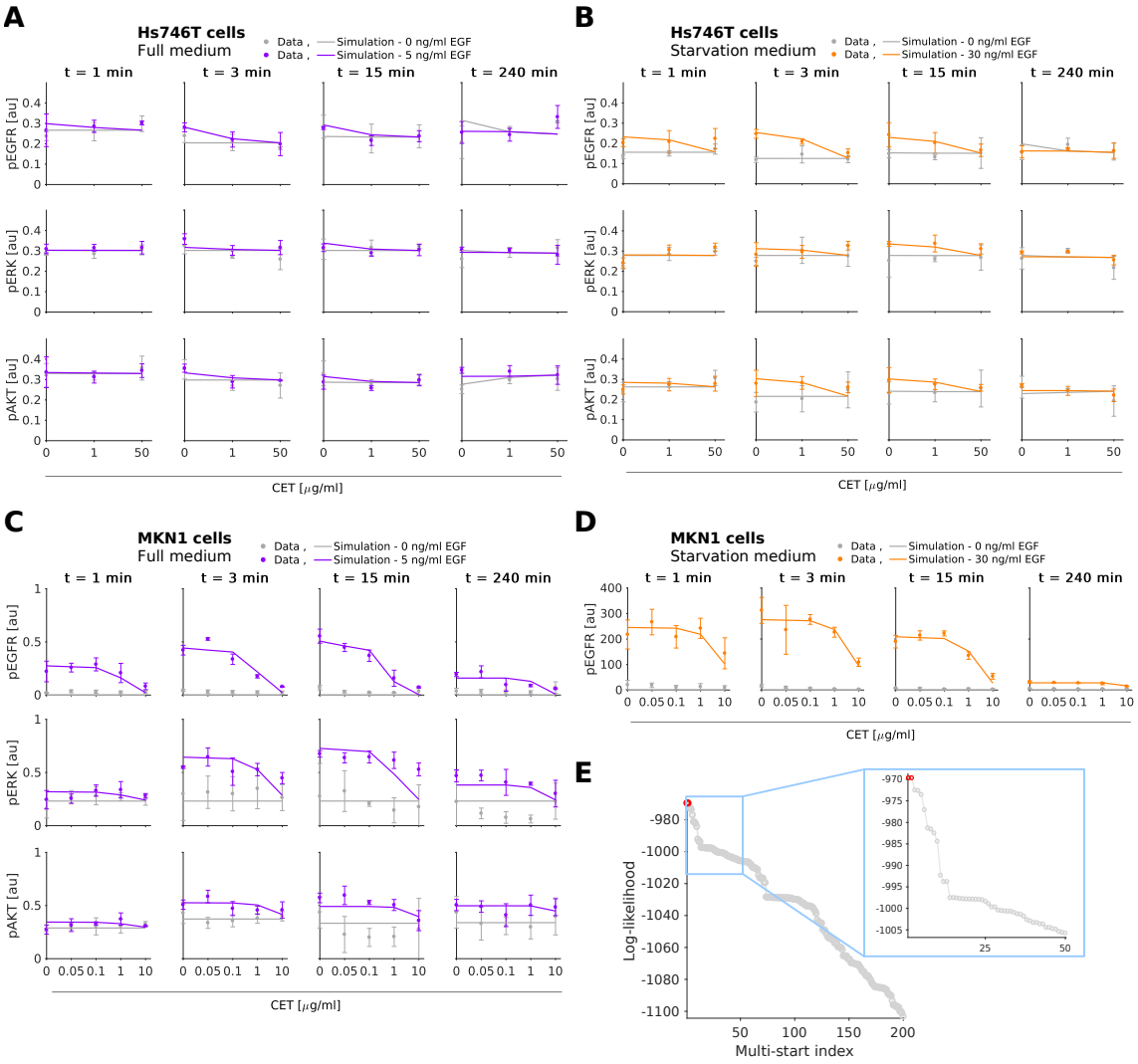

Figure 3.6.: Experimental data and combined model fit for the best model M5. (A-B) Comparison of selected experimental data and model fit for Hs746T cell line. (C-D) Comparison of selected experimental data and model fit for MKN1 cell line. (A-D) Time and dose response data obtained using immunoblotting indicate the mean and standard deviation of three biological experiments. Experimental measurements were scaled to model simulation using the estimated scaling factors. (E) Waterfall plots for multi-start local optimization. The best 200 out of 1000 runs are depicted from which a magnification of the best 50 multi-starts is indicated by the blue box. Red dots denote the starts converged to the global optimum within a small numerical margin. Additional data and model fits are provided in Annex Figures B.4 and B.5. Adapted from Figure S1 of the thesis author's publication (Raimúndez et al., 2020).

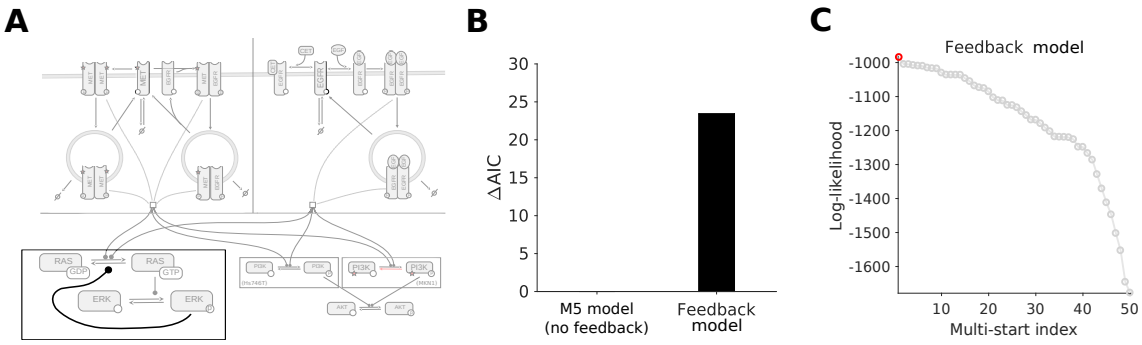

Figure 3.7.: Comparison of the model with and without feedback. (A) Schematic of the model including negative feedback regulation from ERK to RAS. (B) Differences of AIC values for the model and the best AIC. The parameter estimation results for both models were obtained using 300 local optimization runs. The analysis suggested that the model without feedback is more consistent with the experimental data. (C) Waterfall plot for multi-start local optimization for the model with the feedback loop. The best 50 out of 300 runs are depicted. The red dot indicates the only start that converged to the presumed global optimum, within a small numerical margin. Adapted from Figure S8 of the thesis author's publication (Raimúndez et al., 2020).

the model reaction R39 as

$$\text{RAS-GDP} \rightarrow \text{RAS-GTP} \quad \text{with rate} \quad k_a[\text{activated receptor}][\text{RAS-GDP}]\frac{K_i}{(K_i + [\text{pERK}])},$$

with K_i being the inhibition strength, [activated receptor] being the sum of concentration of $(\text{pEGFR:EGF})_2$, $(\text{pEGFR:EGF})_{2,\text{end. ves.}}$, pMMET:pEGFR, $\text{pMMET:pEGFR}_{\text{end. ves.}}$, pMMET_2 or $\text{pMMET}_{2,\text{end. ves.}}$, and k_a being the activation rate of RAS-GDP by a single activated receptor. The short form "end. ves." denotes the formation of endocytic vesicles.

We fitted all unknown model parameters of the extended model to the experimental data (Figure 3.1) using multi-start local optimization with 300 starting points. The AIC value of the extended model is higher than for the model without negative feedback (Figure 3.7B), which suggests that the model without the feedback loop is more consistent with the experimental data. We note that the difference observed between the AIC values is larger than two times the number of additional parameters. This implies that the model with the feedback loop provides a worse fit than the original model. This can be explained by the fact that (i) the model with feedback is not a special case of the model without feedback (as a rate of zero is not in the admissible interval) or (ii) that the multi-start optimization did not find the optimal parameter value (Figure 3.7C). Therefore, in the subsequent analyses we used the model without the feedback loop.

3.5. Assessment of uncertainty and predictive power of the integrated model

In this section, we assess the uncertainty in the estimated parameters of the best candidate of the integrated pathway model. Then, we perform model predictions that are validated with experimental data. Lastly, we interrogate the model to predict already established response and resistance factors. We found that, despite uncertainties in the parameters, the model is able to yield accurate predictions.

3.5.1. Quantification of parameter uncertainty

The selected integrated pathway model describing the experimental data for the cell lines MKN1 and Hs746T possesses 136 parameters in total. From these, 57 parameters are associated with the reaction kinetics and protein abundances (for wild-type and mutant proteins), and 79 are parameters related to the observations (i.e. scaling constants). To assess the identifiability of the kinetic parameters as well as differences between cell lines and culture media, we computed the profile likelihoods (as introduced in Section 2.2.3). We did not calculate the profiles for the observable parameters as they differ between single experiments and are not relevant for the model dynamical response.

The analysis of the profile likelihoods revealed that 23 out of 57 parameters are practically identifiable, because the 90% confidence intervals are finite (Figure 3.8A). Most of the identifiable parameters take part in the EGFR dynamics module, involving processes such as internalization, degradation, ligand binding and dimerization. The remaining parameters are practically non-identifiable, meaning that lower and/or upper bounds could not be found for the defined confidence level (for details on parameter identifiability see Section 2.2.3). In particular, parameters related to MMET signaling are practically non-identifiable. This is not unexpected as MMET is not directly observed.

We complemented the profile likelihood calculation with an evaluation of the Fisher Information Matrix (FIM) at the maximum likelihood estimate (see Section 2.2.3). The eigenvalue spectrum of the FIM for the kinetic parameters and initial conditions spans many orders of magnitude (Figure 3.8B), implying sloppiness (Gutenkunst et al., 2007). Interestingly, there are 50 eigenvalues that differ substantially from numerical zero, meaning that the number of constrained directions in parameter space – given by the eigenvectors – is larger than the number of identifiable parameters. This can happen if individual parameters are non-identifiable but functions of several parameters (e.g. sums, differences or ratios) are identifiable. As the detailed analysis of this using profile likelihoods is

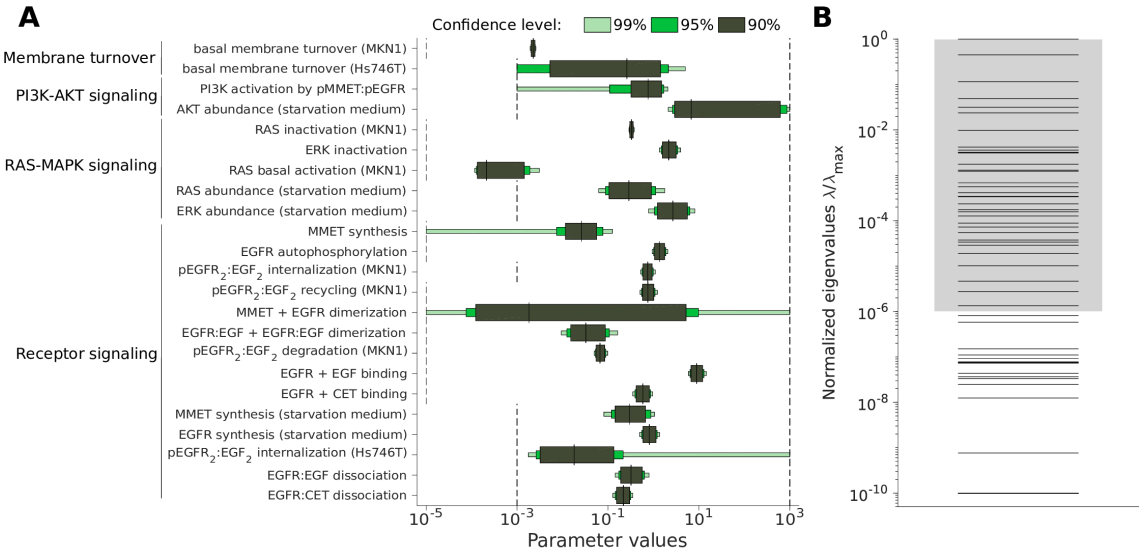

Figure 3.8.: Uncertainty of the parameter estimates for the integrated model. (A) Confidence intervals for the identifiable parameters. The confidence intervals corresponding to the confidence levels 90%, 95% and 99% are shown. Parameter bounds used for optimization are indicated in black dashed lines. (B) Eigenvalue spectrum of the FIM for the dynamical parameters. The spectra have each been normalized by their largest eigenvalue (λ_{max}). Gray shading indicates a spreading of less than six orders of magnitude. Adapted from Figure 5 of the thesis author's publication (Raimúndez et al., 2020).

computationally demanding, the analysis of the FIM provides a first glimpse of parameter uncertainty.

3.5.2. The integrated model yields reliable predictions

While the estimates of many parameters appear reliable, the insights that can be obtained by studying the parameter values are limited. As many parameters in the model represent multi-step processes, there is no clear biological counterpart. This is different for the state variables and outputs of the model, allowing for model-based predictions. To assess the predictive power of the model despite the sloppy eigenvalue spectrum and the non-identifiable parameters, we simulated the system in additional experimental conditions, i.e. in conditions not used for the fitting. Firstly, we evaluated whether the model can, despite the non-identifiabilities in the MMET signaling dynamics, predict published experimental data for the response of Hs746T cells to selected MET inhibitors. Therefore, the MET inhibitor was implemented in the model, and simulation was performed using the previously optimized parameters. The inhibition rate was set as reported by Park et al. (2016) (Section 3.3.2). We found that the model qualitatively predicts the reduction of

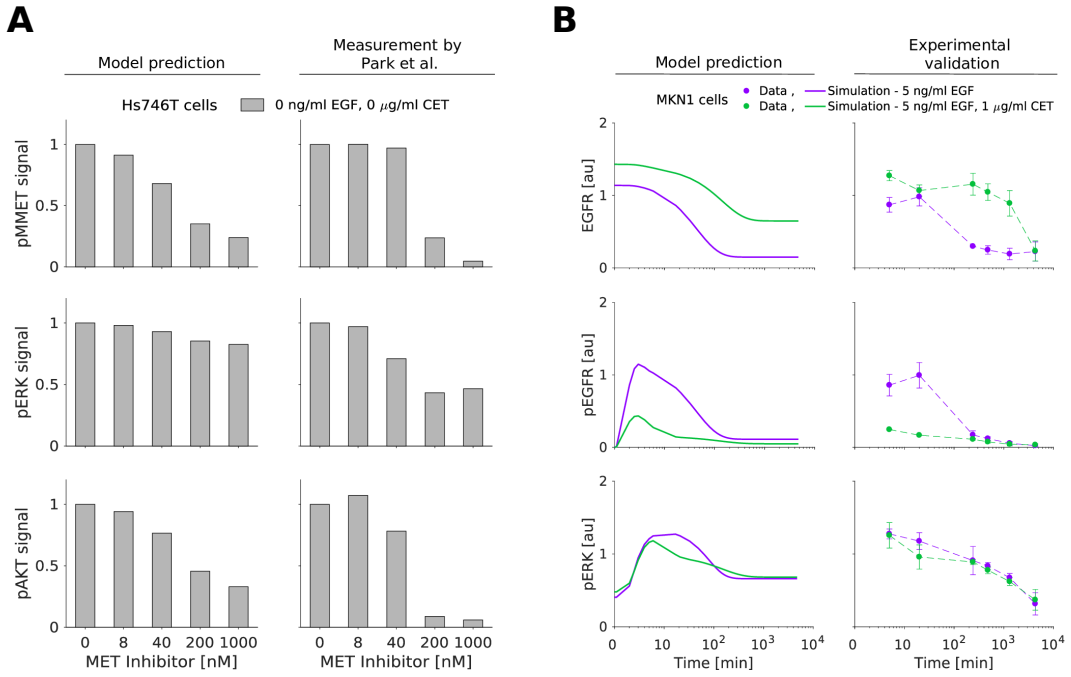

Figure 3.9.: Model predictions by the integrated model. (A) *In silico* prediction and experimental validation of MET inhibitor response in Hs746T cells. Validation data for the MET inhibitor KRC-00715 extracted from Park et al. (2016). The signal was normalized with respect to the untreated condition. (B) Model prediction (left) and experimental validation (right) of long-term response to EGF and the combination of EGF with cetuximab in MKN1 cells. Adapted from Figure 5 of the thesis author's publication (Raimúndez et al., 2020).

pMMET and pAKT levels observed by Park et al. (2016) (Figure 3.9A). Secondly, we predicted the state of MKN1 cells beyond the first 4 hours for which experimental data are available. We found that the model provides reasonable predictions for long-time response of EGFR, pEGFR and pERK (Figure 3.9B). Accordingly, our analysis showed that the model structure and parameter estimates are reasonable and that the model is able to predict the considered validation experiments.

3.5.3. Model prediction of response and resistance factors

The developed mathematical model provides a screening tool for the rapid assessment of potential response and resistance factors. Here, we demonstrate this by studying the validity of the predictions for several established factors (Figure 3.10A).

The evident differences between the responder cell line, MKN1, and the non-responder cell line, Hs746T, are the mutation patterns. The experimental data for MKN1 cells

showed that PIK3CA p.E545K is not a resistance factor. However, our model predicted the association of PIK3CA overexpression (due to p.E545K mutation) with an insensitivity of pAKT to cetuximab treatment (Annex Figure B.6). This is difficult to test experimentally, but consistent with the finding that PIK3CA mutation is a source of acquired cetuximab resistance in metastatic colorectal cancer patients (Xu et al., 2017). The model predictions suggested the same for gastric cancer, although *"PIK3CA mutations were not significantly associated with any clinical, epidemiologic, or pathologic characteristic"* (Harada et al., 2016) in gastric cancer patients obtaining non-targeted therapy.

The MET exon 14 juxtamembrane splice site mutation found in Hs746T cells inhibits the degradation of MET receptor, prolonging its oncogenic activity (Pilotto et al., 2017). MET activation is an established resistance factor for cetuximab treatment in gastric cancer (Kneissl et al., 2012). Indeed, our model predicted that a knockdown of mutant MET reduced Hs746T cell proliferation and survival signaling via ERK and AKT (Figure 3.10B). To validate the prediction, MMET was silenced and quantitative immunoblotting was performed. We found that the model provided accurate quantitative predictions for EGFR, pEGFR, pERK and pAKT (Figure 3.10C) ($\rho = 0.872$) although the down-regulation of pERK was slightly underestimated.

Beyond mutations, amplifications and expression changes have been reported as response and resistance factors. In particular the abundance of EGFR has been reported to be associated with cetuximab response in gastric cancer (Kneissl et al., 2012; Zhang et al., 2013). Our model predicted that reducing EGFR expression levels in MKN1 – a cell line overexpressing EGFR – decreased the levels of pEGFR, pERK and pAKT (Figure 3.10D). Interestingly, the effect on downstream signaling was predicted to be relatively small. We tested this prediction by silencing EGFR expression and found a good agreement with experimental data (Figure 3.10E) ($\rho = 0.915$). The result implies that the dependence of ERK and AKT activity on EGFR activity is limited.

Beyond the expression of EGFR, the abundance of the EGFR ligand AREG has been shown to correlate positively with cetuximab response (Kneissl et al., 2012). As the biochemical processes underlying AREG- and EGF-induced activation of EGFR signaling are similar (Wilson et al., 2009), we employed the same model for AREG and EGF treatment. Following the literature (Macdonald-Obermann & Pike, 2014), we assumed that AREG has an EGFR affinity about 50 times lower than EGF. We neglected to include that AREG stimulation leads to EGFR recycling while EGF promotes EGFR degradation (Roepstorff et al., 2009). The resulting model predicted – in the absence of MMET activation, i.e. in MKN1 cells – that for higher AREG levels cetuximab achieved a higher reduction in EGFR and ERK

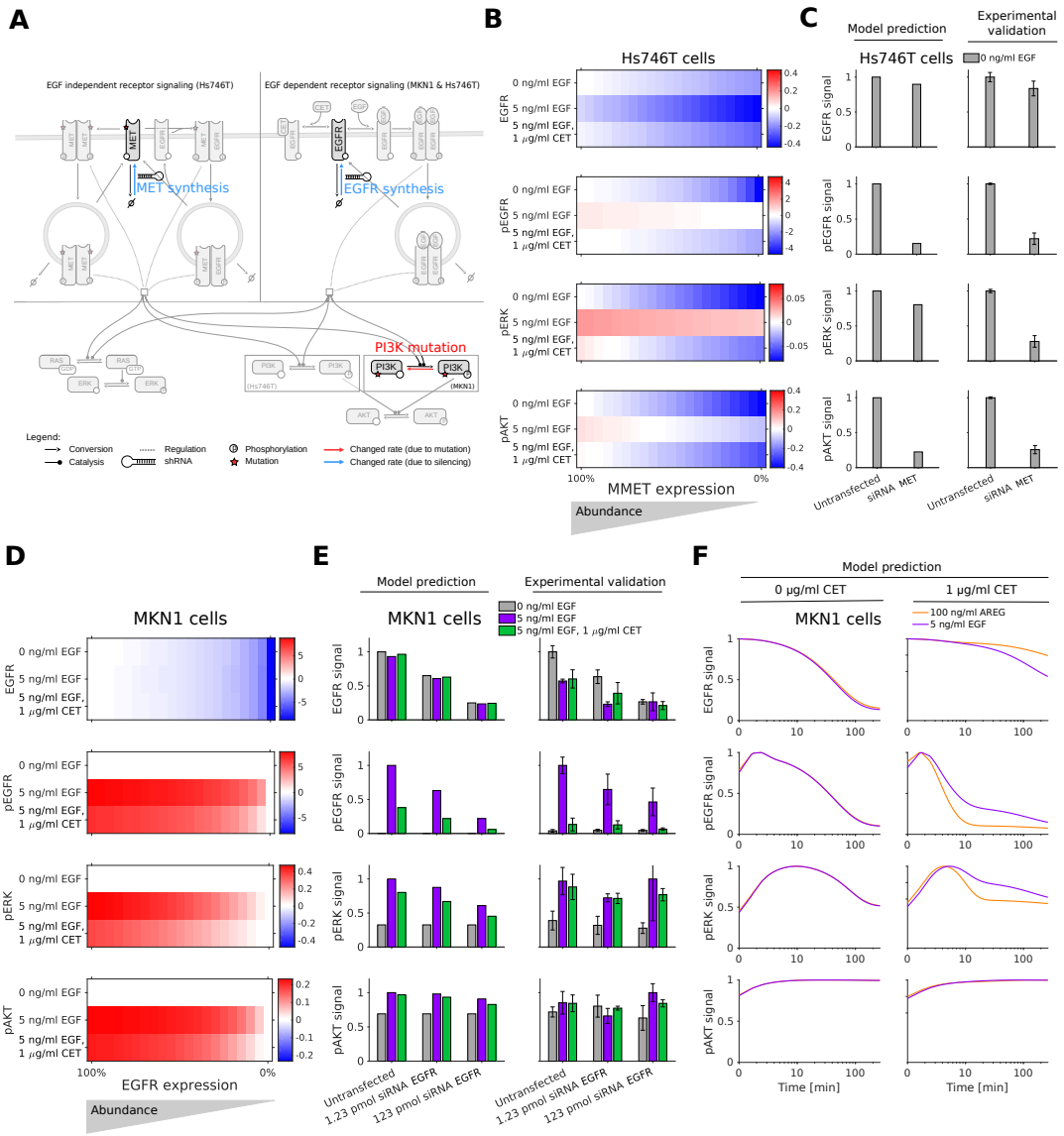

Figure 3.10.: Model prediction of response and resistance factors. (A) Model overview showing changed rates for EGFR and MMET synthesis, and PI3K mutation. (B) *In silico* screening for MMET expression silencing in Hs746T cell line for unstimulated, EGF, and EGF in combination with cetuximab treatment at 3 min. (C) Model prediction and experimental validation of MMET expression silencing using siRNA MET at 3 min. For the simulation, the MMET synthesis rate was scaled with the measured MMET reduction achieved by siRNA treatment in the untreated (no stimulation) condition. (D) *In silico* screening for EGFR expression silencing in MKN1 cell line for unstimulated, EGF, and EGF in combination with cetuximab treatment at 5 min. For the simulation, the EGFR synthesis rate was scaled with the measured EGFR reduction achieved by siRNA treatment in the untreated (no stimulation) condition. (E) Model prediction and experimental validation of EGFR expression silencing using siRNA EGFR at 5 min. (F) Model prediction of time response of AREG stimulation (black line) shows higher sensitivity to cetuximab treatment compared to EGF (red line) in MKN1 cells. (B-F) The signal is normalized with respect to the maximum activity level for each observed component. Adapted from Figure 6 of the thesis author's publication (Raimúndez et al., 2020).

phosphorylation (Figure 3.10F). As a consequence of PI3K mutation in MKN1 cells, AKT activation was insensitive to changes in the receptor signal. This model prediction is in line with the published experimental data in (Kneissl et al., 2012).

3.5.4. Validation of the parameter sets

To dissect the contribution of the selected model topology and the estimated parameter values on the model predictions (Figures 3.8 and 3.10), we evaluated the prediction accuracy for random parameters and optimized parameters. As all the model predictions were performed with the same model, we first looked at the differences in the likelihood value for the 20 best fits compared to the best likelihood estimate (Figure 3.11A). This already indicated a worse performance of the initial parameter guesses as they showed a larger likelihood ratio compared to the final optimal parameters. Additionally, we looked at the correlations between the model predictions and each validation dataset: MMET inhibition (Figure 3.11B), long-term response in MKN1 cells (Figure 3.11C), and EGFR and MMET expression silencing experiments (Figure 3.11D). We found that already random parameter values achieved a positive correlation of the model predictions and the validation data, implying that there is information in the model topology. However, the correlation was substantially higher for the fitted parameters, meaning that the parameter estimation improved the predictive power. Accordingly, model structure and estimated model parameters contributed to the observed prediction accuracy.

3.6. Summary and discussion

Mechanistic ODE models are widely used for the integration of experimental data and the analysis of causal relations. Furthermore, recent studies demonstrate their potential for the identification of biomarkers (Fey et al., 2015; Fröhlich et al., 2018; Hass et al., 2017). To utilize this approach for gastric cancer, in this chapter we developed a mechanistic mathematical model of signaling in gastric cancer. The model describes the dynamics of the EGFR, ERK and AKT signaling pathways in response to EGF and cetuximab treatment, accounting for mutation patterns and protein expression levels. The proposed model provides a more detailed description than the only available logical model (Flobak et al., 2015) by capturing individual biochemical reactions and encoding the wild-type and mutated proteins. To the best of our knowledge, the proposed model is the first mechanistic mathematical model tailored to gastric cancer signaling.

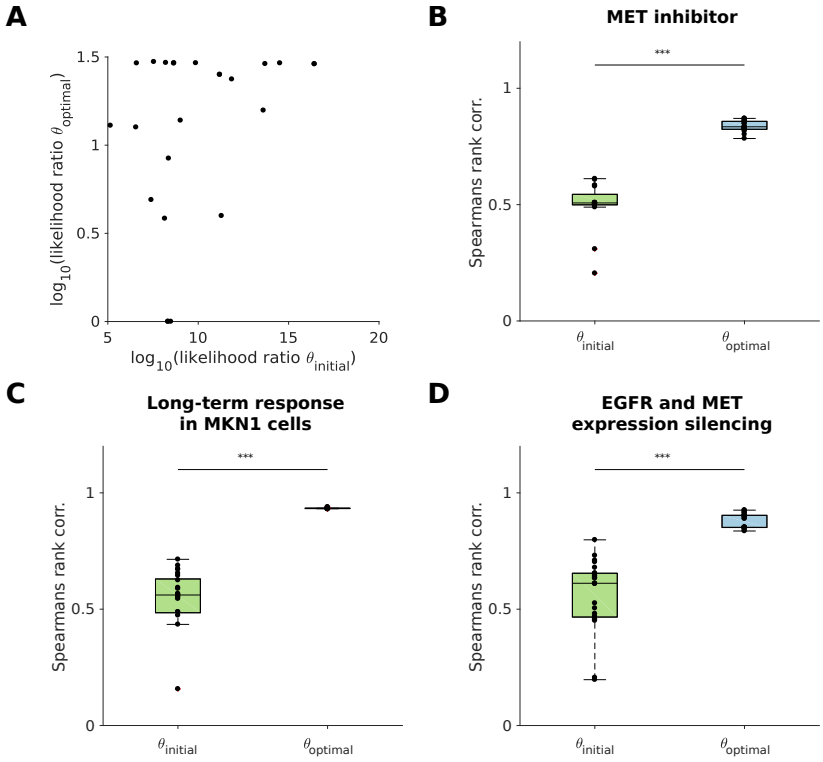

Figure 3.11.: Validation of the parameter sets. (A) Scatter plot of the likelihood ratios for the initial parameter guesses, $\theta_{initial}$, and their corresponding final optimized values, $\theta_{optimal}$. The likelihood ratio was calculated with respect to the best/maximum likelihood estimate found during the multi-start local optimization. The 20 best parameter sets are shown paired with their initial guesses. (B-D) Boxplots of the agreement between the validation data and model simulation for the initial parameter guesses, $\theta_{initial}$, and their corresponding final optimized values, $\theta_{optimal}$. The agreement is shown in terms of Spearman's correlation coefficients. The 20 best parameter sets are shown paired with their initial guesses. The validation datasets shown are (B) MMET inhibition, (C) long-term kinetic response in MKN1 cells, and (D) silencing of EGFR and MMET expression. Adapted from Figure S9 of the thesis author's publication (Raimúndez et al., 2020).

To assess the predictive power of the model, we performed a broad spectrum of analyses. First, we used the model to study causal differences between the cetuximab responder cell line, MKN1, and non-responder cell line, Hs746T. The analysis suggested the presence of a MET mutation as well as differences in receptor internalization and recycling as key factors for the response to cetuximab treatment. Second, we validated the predicted response to a MET inhibitor (KRC-00715) as well as to long-term EGF and EGF in combination with cetuximab treatment. Third, we employed the model to predict response and resistance factors. The predicted role of the EGFR abundance as well as MMET were experimentally confirmed by knockdown experiments. The array of successful validations suggests that the

model has the potential to be used as a hypothesis-generating tool for guiding biomarker discovery.

While all predictive tests were positive, the parameters and predictions of the proposed model are subject to uncertainties. Less than 50% of the parameters were practically identifiable, implying that additional experimental data is required to improve identifiability. In particular, measurements of MMET activity would be beneficial, as well as the absolute quantification of expression and phosphorylation levels. The complexity and size of the considered model impaired the global assessment of parameter uncertainties using Bayesian (sampling-based) approaches as introduced in Section 2.3, which failed for the available tools. Therefore, efficient Bayesian inference methods tailored to models of such complexity are needed.

Furthermore, a detailed experimental analysis of the impact of gene expression on cetuximab response using isogenic cell lines would provide valuable information about signal processing. Complementary to this, the model provides only a crude description of the core pathways and extensions might become necessary. Specifically, the adaptor proteins, intermediate kinases and the effect of multi-site phosphorylation could be included. Furthermore, additional members of the HER family and additional receptor tyrosine kinases, such as AXL, could be included. A template for the refinement could be provided by the Atlas of Cancer Signalling Networks (Kuperstein et al., 2015) or the large-scale mechanistic model introduced by Fröhlich et al. (2018). However, all model extensions should be counterbalanced with acquisition of additional experimental data for gastric cancer cell lines.

For model-based patient stratification, the model also needs to be extended to account for antibody-dependent cellular cytotoxicity (ADCC). ADCC is – besides the binding of cetuximab to EGFR and the inhibition of receptor dimerization and intracellular signal transduction – the second mode of action of cetuximab. ADCC provokes immune cell-mediated killing of tumor cells (Hara et al., 2008). Since the ADCC effect cannot be observed *in vitro*, patient samples would be required to study this effect.

In summary, the proposed model provides a potentially valuable tool for gastric cancer research in the context of cetuximab treatment. It aggregates the information of a comprehensive list of publications and describes a large set of data points. In the future, it might be used to integrate additional data, e.g. from different cell lines and drugs.

Chapter 4

A marginalization-based approach for Bayesian parameter inference in mechanistic models with relative data

This chapter is based on and in part identical to the following publication:

- **Raimúndez, E.**, and Hasenauer, J. Posterior marginalization accelerates Bayesian inference for dynamical systems, *in preparation.*

Mechanistic models based on ordinary differential equations (ODEs) have become important tools for understanding and unraveling the mechanisms underlying, among others, biological signaling processes (Kitano, 2002; Klipp et al., 2005; Schöberl et al., 2009), patient survival (Fey et al., 2015; Hass et al., 2017) and epidemic dynamics (Contento et al., 2021; Giordano et al., 2020; Zhao & Chen, 2020). However, the parameters of all these models are in general unknown *a priori* and need to be inferred from experimental data (Raue et al., 2013b).

To estimate unknown parameters and quantify their uncertainties, Bayesian approaches are commonly used (Contento et al., 2021; Hug et al., 2013; Raue et al., 2013a), which often rely on the sampling of the posterior distribution using Markov chain Monte Carlo (MCMC) algorithms (Section 2.3.3). Bayesian parameter inference provides information about parameter and prediction uncertainties, and thereby the reliability of models. Yet, the generation of representative posterior samples tends to become computationally challenging for high-dimensional problems (Bellman, 1961). This is a major bottleneck that leaves sampling methods on the edge of computational feasibility. Procedures to tackle this dimensionality issue have been applied in the field of cosmology, where integration over nuisance parameters is performed, i.e. they get marginalized out of the estimation problem, as these characterize effects for which an accurate understanding is lacking (Taylor & Kitching, 2010).

In this chapter, with a focus on the field of systems biology and particularly in the context of ODE models with relative measurement data, we present an approach based on the

marginalization over observation-related parameters that allows to reduce the dimensionality of the sampling problem. We present cases for which an analytical marginalization is possible and demonstrate the benefit of the marginalization-based approach for various application examples. Furthermore, we show the huge potential of this approach when successfully applying it to the gastric cancer model presented in Chapter 3, for which previous standard sampling methods failed.

4.1. Parameter inference for mechanistic models with relative data

In systems biology, most measurement techniques, e.g. Western blotting (Renart et al., 1979), flow cytometry (Herzenberg et al., 2006) and fluorescence microscopy (Sanderson et al., 2014), only provide relative information about the absolute molecular state. Therefore, as introduced in Section 2.1.2, to establish a proper statistical link between experimental data and a computational model, observable parameters (e.g. scaling and offset constants) and noise parameters need to be introduced (Degasperi et al., 2017; Raue et al., 2013b; Weber et al., 2011). Mostly, these are also unknown, hence they need to be estimated along with the model parameters, e.g. reaction rate constants and initial conditions, increasing the dimensionality of the estimation problem, in some cases, even doubling its size (Bachmann et al., 2011; Raimúndez et al., 2020).

In the context of frequentist parameter estimation, hierarchical optimization approaches have been developed (Loos et al., 2018; Schmiester et al., 2019; Weber et al., 2011). These methods exploit that optimal observable and noise parameters can be computed analytically for a given set of model parameters, and provide a substantial benefit in the optimization convergence. However, these concepts cannot be used for rigorous Bayesian parameter inference since, among others, these methods do not take into account prior distributions of the observable and noise parameters.

Here, we consider ODE models as introduced in Section 2.1.1 with observables constructed as in (2.3). Following Bayes' theorem for relative measurement data under the assumption of additive Gaussian noise, the posterior distribution $p(\theta, s, \sigma^2|\mathcal{D})$ is given by

$$p(\theta, s, \sigma^2|\mathcal{D}) = \frac{p(\mathcal{D}|\theta, s, \sigma^2)p(\theta)p(s, \sigma^2)}{p(\mathcal{D})}, \tag{4.1}$$

in which $p(\mathcal{D}|\theta, s, \sigma^2)$ denotes the likelihood function, $p(\theta)$ denotes the prior on the model parameters θ, $p(s, \sigma^2)$ denotes the prior on the noise variances σ^2 and scaling constants s, and $p(\mathcal{D})$ denotes the marginal probability of the data (being a normalization constant). Currently, the **standard approach** is to use MCMC sampling methods to construct repre-

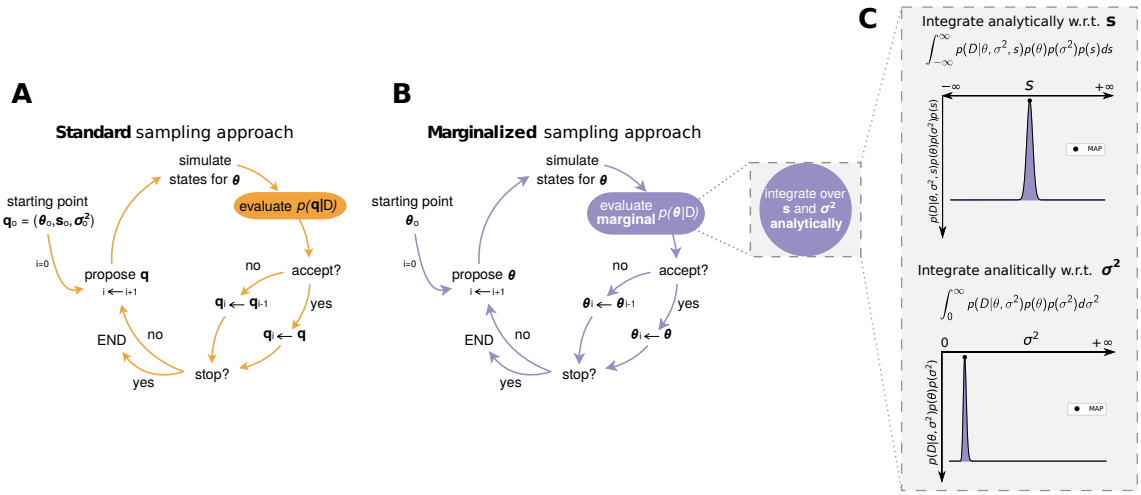

Figure 4.1.: Visualization of standard and marginalized sampling schemes. (A) Standard approach. (B) Marginalized approach. (C) Sequential integration of s and σ^2 to obtain $p(\mathcal{D}|\theta)$.

sentative samples from (4.1) for subsequent analysis (see Section 2.3.3). This means that all parameters are sampled jointly, disregarding their nature (Figure 4.1A). Standard sampling is often challenging, even unfeasible for models with large datasets, since the amount of observable and noise parameters can easily double the full parameter dimensionality, as seen in e.g. (Bachmann et al., 2011) and (Raimúndez et al., 2020).

In the following, focusing on additive Gaussian noise (2.2) – which is most commonly used in systems biology (Fröhlich et al., 2019) – combined with scaling constants (2.3), we determine an analytical marginalization of the posterior distribution with respect to scaling and noise parameters (Figure 4.1). This yields a lower dimensional posterior, which can be explored more easily using MCMC sampling or related approaches. We benchmark the proposed approach with a collection of published models to illustrate its properties. The method achieves a higher sampling efficiency as it reduces the auto-correlation of the samples and increases the transition rates between posterior modes. Lastly, we apply it to the gastric cancer model presented in Chapter 3, which successfully converges to the posterior, while sampling in the full parameter space failed.

4.2. Marginalization of observable and noise parameters for additive Gaussian-distributed noise

In this section, we introduce the essence of the proposed method. Firstly, we provide a detailed mathematical derivation of the analytical marginalization of the posterior

distribution with respect to scaling and noise parameters. Secondly, we show how the marginalized parameters can be retrieved. Lastly, we evaluate the performance of the proposed method in a toy model.

4.2.1. Analytical derivation of the marginal likelihood

For a mechanistic understanding of biological processes, only the vector of the model parameters θ is required, but not the scaling constants and noise variances (Loos et al., 2018). Therefore, we considered the marginalization of the posterior distribution (4.1) with respect to s and σ^2, yielding

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{p(\mathcal{D})} \tag{4.2}$$

with $p(\mathcal{D}|\theta)$ as the **marginal likelihood** given by

$$p(\mathcal{D}|\theta) = \int_0^\infty \int_{-\infty}^\infty p(\mathcal{D}|\theta, s, \sigma^2)p(s, \sigma^2)dsd\sigma^2. \tag{4.3}$$

For ease of notation, we defined our problem (4.3) in terms of precision $\lambda := 1/\sigma^2$, therefore with $\lambda \in \mathbb{R}^+$ and

$$p(\mathcal{D}|\theta) = \int_0^\infty \int_{-\infty}^\infty p(\mathcal{D}|\theta, s, \lambda)p(s, \lambda)dsd\lambda. \tag{4.4}$$

Notation: To simplify the notation introduced in Section 2.1.2, in the following we assume $n_y = 1$ and $n_e = 1$, and we skip some dependencies of h and write $h_k := h_{i_y}(x(t_k, \theta, u^{i_e}), \theta, u^{i_e})$.

The likelihood of data given θ for h_k under **additive Gaussian noise** is defined as

$$p(\mathcal{D}|\theta, s, \lambda) = \left(\frac{\lambda}{2\pi}\right)^{n_t/2} \exp\left(-\frac{\lambda}{2}\sum_{k=1}^{n_t}(\bar{y}_k - s \cdot h_k)^2\right),$$

where n_t denotes the number of time points, $\bar{y}_k$ denotes the data point and h_k the unscaled observable (i.e. simulation output) for time index k.

To tractably solve (4.4) for an additive Gaussian likelihood, the prior distribution $p(s, \lambda)$ has to be chosen appropriately. We assumed a Normal-Gamma-distributed prior for s and

λ, given by

$$\begin{aligned} p(s,\lambda) &= f(s,\lambda|\mu,\kappa,\alpha,\beta) = \mathcal{N}(s|\mu,1/(\kappa\lambda)) \cdot \Gamma(\lambda|\alpha,\beta) \\ &= \frac{\beta^\alpha\sqrt{\kappa}}{\Gamma(\alpha)\sqrt{2\pi}}\lambda^{\alpha-1/2}\exp\left(-\beta\lambda - \frac{\kappa\lambda(s-\mu)^2}{2}\right), \end{aligned} \tag{4.5}$$

with $\mu \in \mathbb{R}$ and $\kappa, \alpha, \beta \in \mathbb{R}_+$ denoting hyper-parameters of the prior distribution, and $\Gamma(\cdot)$ the Gamma function. The Normal-Gamma distribution is the conjugate prior of a Gaussian distribution with unknown mean and precision, therefore it gives a closed-form expression for the posterior.

Hence, the evaluation of the integral (4.4) is given by

$$\begin{aligned} p(\mathcal{D}|\theta) = \int_0^\infty \int_{-\infty}^\infty \left(\frac{\lambda}{2\pi}\right)^{n_t/2} \exp\left(-\frac{\lambda}{2}\sum_{k=1}^{n_t}(\bar{y}_k - s\cdot h_k)^2\right) \\ \cdot \frac{\beta^\alpha\sqrt{\kappa}}{\Gamma(\alpha)\sqrt{2\pi}}\lambda^{\alpha-1/2}\exp\left(-\beta\lambda - \frac{\kappa\lambda(s-\mu)^2}{2}\right) ds\, d\lambda. \end{aligned} \tag{4.6}$$

This two-dimensional integral can be computed employing numerical solvers, however this would lead to a computational overhead. To avoid this, analytical integration of (4.6) is preferred. For this, we sequentially integrated out (i) the scaling constants and (ii) the precision parameters (Figure 4.1C), which we describe in detail in the following.

Starting with (4.6) and after pulling out the constant terms, we obtain

$$\begin{aligned} p(\mathcal{D}|\theta) = \frac{\beta^\alpha\sqrt{\kappa}}{\Gamma(\alpha)(2\pi)^{(n_t+1)/2}} \\ \cdot \int_0^\infty \lambda^{\alpha+(n_t-1)/2}\exp\left(-\beta\lambda\right) \underbrace{\int_{-\infty}^\infty \exp\left(-\frac{\lambda}{2}\left(\sum_{k=1}^{n_t}(\bar{y}_k - s\cdot h_k)^2 + \kappa(s-\mu)^2\right)\right) ds}_{(*)}\, d\lambda. \end{aligned} \tag{4.7}$$

We firstly focused on the evaluation of the integral with respect to s, which is denoted by $(*)$

$$\begin{aligned}(*) &= \int_{-\infty}^{\infty} \exp\left(-\frac{\lambda}{2}\left(\sum_{k=1}^{n_t}(\bar{y}_k - s\cdot h_k)^2 + \kappa(s-\mu)^2\right)\right) ds \\ &= \int_{-\infty}^{\infty} \exp\left(-\frac{\lambda}{2}\left(\kappa + \sum_{k=1}^{n_t} h_k^2\right)s^2 + \lambda\left(\kappa\mu + \sum_{k=1}^{n_t}\bar{y}_k h_k\right)s - \frac{\lambda}{2}\left(\kappa\mu^2 + \sum_{k=1}^{n_t}\bar{y}_k^2\right)\right) ds.\end{aligned} \tag{4.8}$$

The integral (4.8) can be solved analytically, since it follows the structure of the Gaussian integral (Owen, 1980)

$$\int_{-\infty}^{\infty} \exp\left(-ax^2 + bx + c\right) dx = \sqrt{\frac{\pi}{a}} \exp\left(\frac{b^2}{4a} + c\right) \tag{4.9}$$

with constants

$$a := \frac{\lambda}{2}\left(\kappa + \sum_{k=1}^{n_t} h_k^2\right), \qquad b := \lambda\left(\kappa\mu + \sum_{k=1}^{n_t}\bar{y}_k h_k\right), \qquad \text{and} \qquad c := -\frac{\lambda}{2}\left(\kappa\mu^2 + \sum_{k=1}^{n_t}\bar{y}_k^2\right).$$

Therefore, we obtain

$$\begin{aligned}(*) &= \sqrt{\frac{\pi}{\frac{\lambda}{2}\left(\kappa + \sum_{k=1}^{n_t} h_k^2\right)}} \exp\left(\frac{\lambda^2\left(\kappa\mu + \sum_{k=1}^{n_t}\bar{y}_k h_k\right)^2}{4\frac{\lambda}{2}\left(\kappa + \sum_{k=1}^{n_t} h_k^2\right)} - \frac{\lambda}{2}\left(\kappa\mu^2 + \sum_{k=1}^{n_t}\bar{y}_k^2\right)\right) \\ &= \sqrt{\frac{2\pi}{\lambda\left(\kappa + \sum_{k=1}^{n_t} h_k^2\right)}} \exp\left(-\lambda\widehat{C}\right)\end{aligned}$$

with

$$\widehat{C} := -\frac{1}{2}\left(\frac{\left(\kappa\mu + \sum_{k=1}^{n_t}\bar{y}_k h_k\right)^2}{\left(\kappa + \sum_{k=1}^{n_t} h_k^2\right)} - \left(\kappa\mu^2 + \sum_{k=1}^{n_t}\bar{y}_k^2\right)\right).$$

Substitution of $(*)$ in (4.7) yields

$$\begin{aligned}p(\mathcal{D}|\theta) &= \frac{\beta^\alpha\sqrt{\kappa}}{\Gamma(\alpha)(2\pi)^{(n_t+1)/2}} \int_0^\infty \lambda^{\alpha+(n_t-1)/2}\exp\left(-\beta\lambda\right)\sqrt{\frac{2\pi}{\lambda\left(\kappa + \sum_{k=1}^{n_t} h_k^2\right)}}\exp\left(-\lambda\widehat{C}\right) d\lambda \\ &= \frac{\beta^\alpha\sqrt{\kappa}}{\Gamma(\alpha)(2\pi)^{(n_t+1)/2}}\sqrt{\frac{2\pi}{\left(\kappa + \sum_{k=1}^{n_t} h_k^2\right)}} \int_0^\infty \lambda^{\alpha+n_t/2-1}\exp\left(-\lambda C\right) d\lambda\end{aligned} \tag{4.10}$$

with $C := \beta + \widehat{C}$.

Using the transformation $\omega = \lambda C$, we get

$$\begin{aligned} p(\mathcal{D}|\theta) &= \frac{\beta^\alpha \sqrt{\kappa}}{\Gamma(\alpha)(2\pi)^{(n_t+1)/2}} \sqrt{\frac{2\pi}{\left(\kappa + \sum_{k=1}^{n_t} h_k^2\right)}} \int_0^\infty \left(\frac{\omega}{C}\right)^{\alpha+n_t/2-1} \exp(-\omega)\ \frac{1}{C} d\omega \\ &= \left(\frac{1}{C}\right)^z \cdot \frac{\beta^\alpha \sqrt{\kappa}}{\Gamma(\alpha)(2\pi)^{(n_t+1)/2}} \sqrt{\frac{2\pi}{\left(\kappa + \sum_{k=1}^{n_t} h_k^2\right)}} \int_0^\infty \omega^{z-1} \exp(-\omega)\ d\omega \end{aligned} \tag{4.11}$$

with $z := \alpha + n_t/2$. The Gamma function is defined by the integral in (4.11)

$$\int_0^\infty \omega^{z-1} \exp(-\omega)\, d\omega = \Gamma(z). \tag{4.12}$$

Therefore, we obtain

$$\begin{aligned} p(\mathcal{D}|\theta) &= \left(\frac{1}{C}\right)^z \cdot \frac{\beta^\alpha \sqrt{\kappa}}{\Gamma(\alpha)(2\pi)^{(n_t+1)/2}} \sqrt{\frac{2\pi}{\left(\kappa + \sum_{k=1}^{n_t} h_k^2\right)}} \cdot \Gamma(z) \\ &= \frac{(\beta/C)^\alpha}{\Gamma(\alpha)(2\pi C)^{n_t/2}} \cdot \sqrt{\frac{\kappa}{\kappa + \sum_{k=1}^{n_t} h_k^2}} \cdot \Gamma\left(\alpha + \frac{n_t}{2}\right). \end{aligned} \tag{4.13}$$

After substituting C, we can use the marginal likelihood (4.13) together with $p(\theta)$ to sample in the reduced space θ (Figure 4.1B). In the remainder of this thesis, we refer to this as the **marginalized approach**.

4.2.2. Retrieving the marginalized observable and noise parameters

After deriving $p(\theta|\mathcal{D})$, we can recover the marginalized observable and noise parameters s and λ. These can be obtained by sampling from the conditional distribution $p(s, \lambda|\theta, \mathcal{D})$. As we assume that the marginalized parameters are Normal-Gamma-distributed, we can derive the conditional distribution in closed-form.

Sampling the marginalized noise parameters

To construct the conditional probability of the noise parameters given θ and $\mathcal{D}$, we know from the integrand of (4.10) that λ has – up to a constant – the distribution of a Gamma distribution, mainly

$$\lambda \sim \text{Gamma}(\alpha' = \alpha + n_t/2, \beta' = C).$$

The distribution is well-defined if $\alpha', \beta' \in \mathbb{R}_+$, which holds for both parameters. While for α' this is clear by definition, for β' we consider the following

$$C \in \mathbb{R}_+ \Leftrightarrow \beta + \frac{1}{2}\left(\kappa\mu^2 + \sum_{k=1}^{n_t} \bar{y}_k^2\right) > \frac{1}{2}\frac{\left(\kappa\mu + \sum_{k=1}^{n_t} \bar{y}_k h_k\right)^2}{\left(\kappa + \sum_{k=1}^{n_t} h_k^2\right)}.$$

As $\beta > 0$ it is sufficient to show that

$$\left(\kappa\mu^2 + \sum_{k=1}^{n_t} \bar{y}_k^2\right) \geq \frac{\left(\kappa\mu + \sum_{k=1}^{n_t} \bar{y}_k h_k\right)^2}{\left(\kappa + \sum_{k=1}^{n_t} h_k^2\right)}. \tag{4.14}$$

Then, following Cauchy–Schwarz inequality

$$\left(\sum_{k=1}^{n_t} a_k \cdot b_k\right)^2 \leq \left(\sum_{k=1}^{n_t} a_k^2\right) \cdot \left(\sum_{k=1}^{n_t} b_k^2\right)$$

with $a_k = \overline{y}_k,\ b_k = h_k$ for $1 \leq k \leq n_t$ and $a_{n_t+1} = \sqrt{\kappa}\mu,\ b_{n_t+1} = \sqrt{\kappa}$, we receive

$$\left(\kappa\mu + \sum_{k=1}^{n_t} \bar{y}_k h_k\right)^2 \leq \left(\kappa\mu^2 + \sum_{k=1}^{n_t} \bar{y}_k^2\right) \cdot \left(\kappa + \sum_{k=1}^{n_t} h_k^2\right).$$

By dividing through $\left(\kappa + \sum_{k=1}^{n_t} h_k^2\right)$ we receive the desired result (4.14).

Sampling the marginalized observable parameters

To construct the conditional probability of the observable parameters s given θ, $\mathcal{D}$ and λ, we consider the integrand of (4.8). We receive

$$\begin{aligned}
s &\sim \exp\left(-\frac{\lambda}{2}\left(\kappa + \sum_{k=1}^{n_t} h_k^2\right) s^2 + \lambda\left(\kappa\mu + \sum_{k=1}^{n_t} \bar{y}_k h_k\right) s\right) \\
&\sim \exp\left(-\frac{\lambda}{2}\left(\kappa + \sum_{k=1}^{n_t} h_k^2\right)\left(s - \frac{\left(\kappa\mu + \sum_{k=1}^{n_t} \bar{y}_k h_k\right)}{\left(\kappa + \sum_{k=1}^{n_t} h_k^2\right)}\right)^2\right) \\
&\sim\ \mathcal{N}\left(\mu' = \frac{\left(\kappa\mu + \sum_{k=1}^{n_t} \bar{y}_k h_k\right)}{\left(\kappa + \sum_{k=1}^{n_t} h_k^2\right)}, \lambda' = \lambda\left(\kappa + \sum_{k=1}^{n_t} h_k^2\right)\right)
\end{aligned}$$

The distribution is well-defined if $\mu' \in \mathbb{R}$ and $\lambda' \in \mathbb{R}_+$, which holds for both and is clear by definition.

Similarly, as $\sigma^2 = 1/\lambda$, we conclude that

$$\sigma^2 \sim \text{Inv-Gamma}(\alpha' = \alpha + n_t/2, \beta' = C)$$

and

$$s \sim \mathcal{N}\left(\mu' = \frac{(\kappa\mu + \sum_{k=1}^{n_t} \bar{y}_k h_k)}{\left(\kappa + \sum_{k=1}^{n_t} h_k^2\right)}, (\sigma')^2 = \frac{\sigma^2}{\left(\kappa + \sum_{k=1}^{n_t} h_k^2\right)}\right).$$

4.2.3. Simulation example using a toy model

We performed a simulation study with a simple model having the minimal number of dimensions to apply the marginalized approach. We considered a model of a conversion reaction process (Figure 4.2A). This model describes a reversible chemical reaction, which converts a biochemical species A to a species B with rate k_1, and B to A with rate k_{-1} (Figure 4.2A). The corresponding ODEs are

$$\dot{x}_1 = -k_1 x_1 + k_{-1} x_2$$
$$\dot{x}_2 = k_1 x_1 - k_{-1} x_2$$

for which the state vector $x = (x_1, x_2)^T$ consists of the concentrations of A and B, respectively. We assumed that x_2 is measured yielding the observation model $y(t) = s \cdot h(x(t, \theta), \theta) = s \cdot x_2(t)$. For the evaluation of the proposed method, we generated one artificial dataset. The dataset was generated with initial conditions $x_0 = (1, 0.01)^T$, model parameters $\theta = (k_1, k_{-1})^T = (10^{-1.3}, 10^{-0.1})^T$, scaling constant $s = 10^{-0.1}$, and additive Gaussian noise with variance $\sigma^2 = 10^{-3.0}$. The dataset is depicted in Figure 4.2E.

To setup a minimal example, we fixed the forward rate k_1 to the value used for the synthetic data generation and only sampled the reverse rate k_{-1} (Table 4.1, Toy model). Firstly, we checked whether our analytical integration of the marginal likelihood (4.13) coincided with the numerical integration of (4.6). We observed a perfect match for a range of different parameter values (Figure 4.2B). As expected, the evaluation of the analytical marginal likelihood was substantially faster than the numerical integration (Figure 4.2C). Secondly, following the workflow presented by Ballnus et al. (2017), we performed 100 independent MCMC sampling runs using the adaptive Metropolis algorithm (Haario et al. (2001), Section 2.3.3). All the runs were initialized at the global optima found during multi-start optimization (Raue et al., 2013b) for the standard and marginalized approach. Here and in the following, we employ the MCMC sampling quality and convergence measures introduced in Section 2.3.4 to assess the performance of the standard and marginalized method. We found a superior performance of the marginalized approach, as the observed effective sample

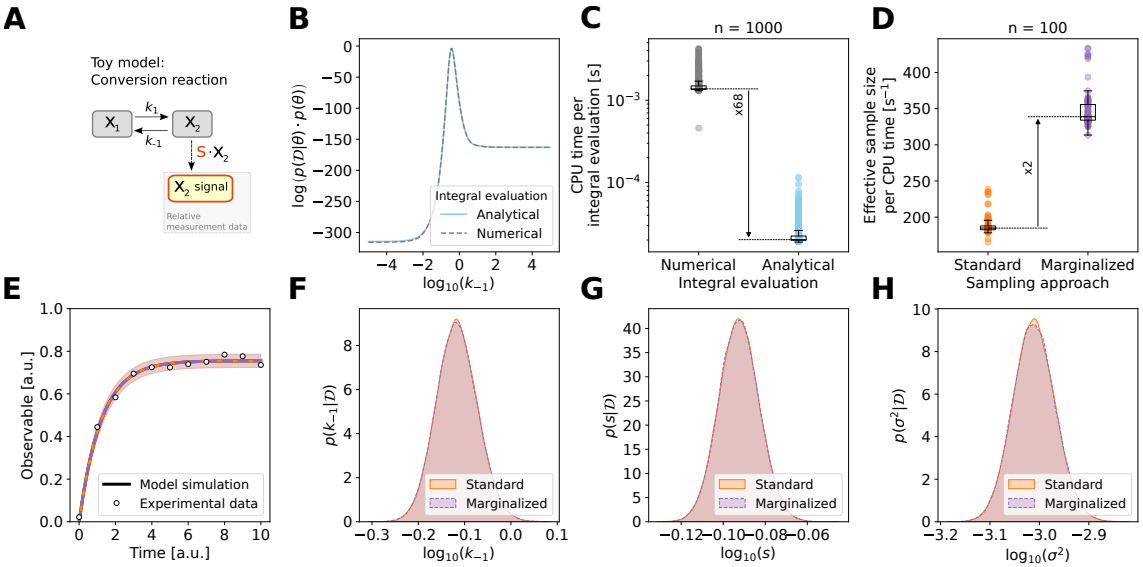

Figure 4.2.: Evaluation of the standard and marginalized sampling approach for the toy model. (A) Model representation. (B) Comparison of analytical vs numerical integration. (C) Time comparison analytical vs numerical integration. (D) Effective sample size per unit of time for 100 independent runs. (E) Model fit of the best sample found during sampling from the standard (orange) and marginalized (purple) approach. (F–H) Parameter marginal posterior distributions computed using a kernel density estimate for (F) the model parameter k_{-1}, and the recovered marginalized (G) scaling constant s and (H) noise variance σ^2.

size per unit of time increased by a factor of two, compared to the standard approach (Figure 4.2D). This indicates that the marginalized approach improved the mixing of the chains and, hence, provided a more efficient exploration of the parameter space. The model fit for the best sample found (i.e. maximizing the posterior) coincided for both approaches (Figure 4.2E) as did the marginal distributions for the model parameter k_{-1} (Figure 4.2F), and the recovered marginalized scaling constant s and noise variance σ^2 (Figure 4.2G–H).

4.3. Evaluation of the proposed method for three signaling models

In this section, we further study the performance of the proposed method (Figure 4.1B–C) on a collection of three mathematical models of different signaling processes with published experimental data. We found that the proposed method was consistently more efficient than the standard method for all tested models.

Table 4.1.: Key numbers and features of the considered models. The number of unknown model parameters n_θ, unknown observable parameters n_ξ and unknown noise parameters n_σ, which are effectively sampled, are reported.

Model ID	n_θ	n_ξ	n_σ	Feature	Reference
Toy model	1	1	1	Minimal example	–
M1	13	3	–	Signaling pathway I	(Fujita et al., 2010)
M2	6	3	3	Signaling pathway II	(Boehm et al., 2014)
M3	3	1	1	Bimodal posterior	(Leonhardt et al., 2014)

4.3.1. Description of the signaling models

For the evaluation of the marginalized approach, we employed three published mechanistic models based on ODEs with experimental data (Table 4.1). These models are briefly described in the following.

M1: Model of EGF-dependent AKT pathway

The model from Fujita et al. (2010) describes an EGF-dependent AKT pathway with 16 unknown parameters in total, of which 13 are model parameters and 3 are scaling constants (Table 4.1, M1). The available experimental data are a total of 144 data points under 6 different experimental conditions for 3 observables. For each data point, the corresponding variance of the measurement noise is provided, therefore it does not need to be estimated. The complete dataset is depicted in Annex Figure C.2.

M2: Model of STAT5 dimerization

The model from Boehm et al. (2014) describes a STAT5 dimerization pathway with 9 unknown parameters in total, of which 6 are model parameters and 3 are noise parameters. To this model, we have added 3 scaling constants (Table 4.1, M2), one per observable, for the sake of testing the proposed method. The available experimental data are a total of 48 data points for 3 observables. The complete dataset is depicted in Annex Figure C.4.

Models M1 and M2 were taken from the `PEtab` benchmark collection which is based on (Hass et al., 2019) and available at `https://github.com/Benchmarking-Initiative/Benchmark-Models-PEtab`.

M3: Model of mRNA transfection 

The model from Leonhardt et al. (2014) describes an mRNA transfection process in which cells are transfected with GFP-mRNA. The corresponding ODEs are

$$\begin{aligned} \dot{x}_1 &= -\delta \cdot x_1 \qquad x_1(t_0) = m_0 \\ \dot{x}_2 &= k_{TL} \cdot x_1 - \beta \cdot x_2 \qquad x_2(t_0) = 0 \end{aligned}$$

where t_0 denotes the onset time, k_{TL} the translation rate of mRNA to protein, m_0 the initial number of effectively translated mRNA molecules, and δ and β the degradation rates of mRNA and protein, respectively. The state vector $x = (x_1, x_2)^T$ consists of the concentrations of mRNA and protein, respectively. The observable is the protein concentration, which is given by

$$y(t) = s \cdot x_2(t) = s \cdot k_{TL} \cdot m_0 \cdot \frac{e^{-\beta(t-t_0)} - e^{-\delta(t-t_0)}}{\delta - \beta}.$$

The posterior distribution of this model is bimodal as the exchange of the degradation rates of mRNA δ and protein β results in the same dynamics for the observable. We considered the product $\hat{s} = s \cdot k_{TL} \cdot m_0$ as a single parameter and treated $\hat{s}$ as a scaling constant, therefore marginalized together with the noise parameter. Consequently, in our setting this model consists of 5 unknown parameters in total, of which we have 3 model parameters $\theta = (t_0, \beta, \delta)$, one scaling constant $\hat{s}$, and one noise parameter (Table 4.1, M3).

4.3.2. Comparison of standard and marginalization-based approach

To further test the performance of the proposed marginalized approach, we considered a collection of already published models M1–M3 of increasing size and complexity with experimental data (Table 4.1). For the model M1 the variance of the measurement noise is not estimated from the data, i.e. it is provided in the measurements, thus only the scaling constants needed to be estimated on top of the model parameters. Hence, we derived the corresponding marginal likelihood function integrating only s, which we provide in Annex D.1.

When using the marginalized approach for the models considered here, we could substantially reduce the dimensionality of the problem to up to half of the original size (Figure 4.3A). For all models, we performed 50 independent MCMC sampling runs using the parallel tempering algorithm with 10 temperatures (Łacki & Miasojedow (2015), Section 2.3.3). All the runs were initialized at the 10 best local optima found for during multi-start

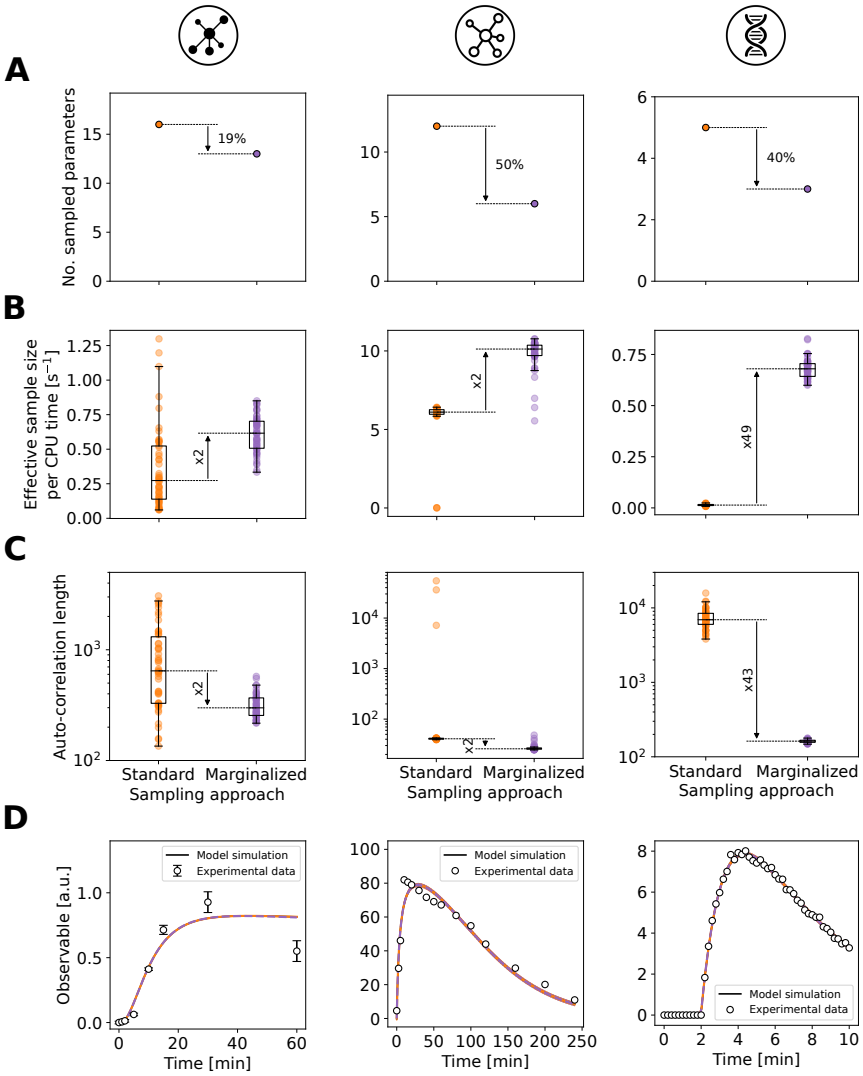

Figure 4.3.: Evaluation of the standard and marginalized sampling approach for the benchmark models. Models M1–M3 are shown from left to right. Results for the standard approach are shown in orange and marginalized in purple. (A) Number of sampled parameters. (B) Effective sample size per unit of time. (C) Auto-correlation length. (D) Model fit of the best sample found during sampling. A subset of the experimental data is shown for M1 and M2, their complete datasets are depicted in Annex Figures C.2 and C.4.

optimization (Raue et al., 2013b) for the standard and marginalized approach. While the standard approach worked, we found for all models that the marginalized approach showed a consistent superior performance in terms of effective sample size per unit of time (Figure 4.3B) and reduction in the auto-correlation length (Figure 4.3C). This confirmed our previous observation that the marginalized approach enabled a more efficient exploration of

the parameter space. Additionally, the model fits for the best sample found were identical for both approaches (Figure 4.3D) as well as the parameter marginal distributions (Annex Figures C.1, C.3 and C.5).

4.4. The marginalization-based approach improves transition rates between posterior modes

A very interesting result was the finding that the marginalized approach was for M3 nearly 50 times more efficient than the standard approach. Since M3 possesses a bimodal posterior (Table 4.1), we hypothesized that the huge improvement in the performance could be explained by an increased transition probability between the different posterior modes provided by the marginalized approach. These transitions occur when a single chain transitions to the neighboring mode. To confirm our hypothesis that the marginalized sampling approach improved transition rates between posterior modes within one single chain, we performed several independent MCMC runs using the single-chain adaptive Metropolis algorithm (Haario et al., 2001) and considering a wide range of temperature values for the posterior (4.1) and the marginalized posterior (4.2).

For the standard approach, the tempered versions of the posterior distribution are

$$p(\theta, s, \sigma^2|\mathcal{D})_T \propto \left(p(\mathcal{D}|\theta, s, \sigma^2)p(\theta)p(s, \sigma^2)\right)^{1/T}.$$

Accordingly, for the marginalized approach, the tempered versions of the marginalized posterior distribution are

$$p(\theta|\mathcal{D})_T \propto \left(p(\mathcal{D}|\theta)p(\theta)\right)^{1/T}.$$

We found that for the same temperature value, e.g. $T = 16$, the standard approach only explored one mode (Figure 4.4A–B) while in the marginalized approach transitions happened more frequently, being able to fully explore both modes (Figure 4.4D–E). Furthermore, we found that in the marginalized approach the transitions occurred already for lower temperatures than in the standard approach (Figure 4.4C). Our results demonstrate that by using the marginalized approach, the exploration of the parameter space was more efficient, and especially beneficial for more complex posteriors. Aiming to explain this behavior, we analyzed and compared the minimal energy path (Henkelman & Jónsson, 2000) for both approaches. However, our results indicated that the minimal energy path was practically identical between the two methods (Figure 4.5), which suggested a relation to the geometry of the paths.

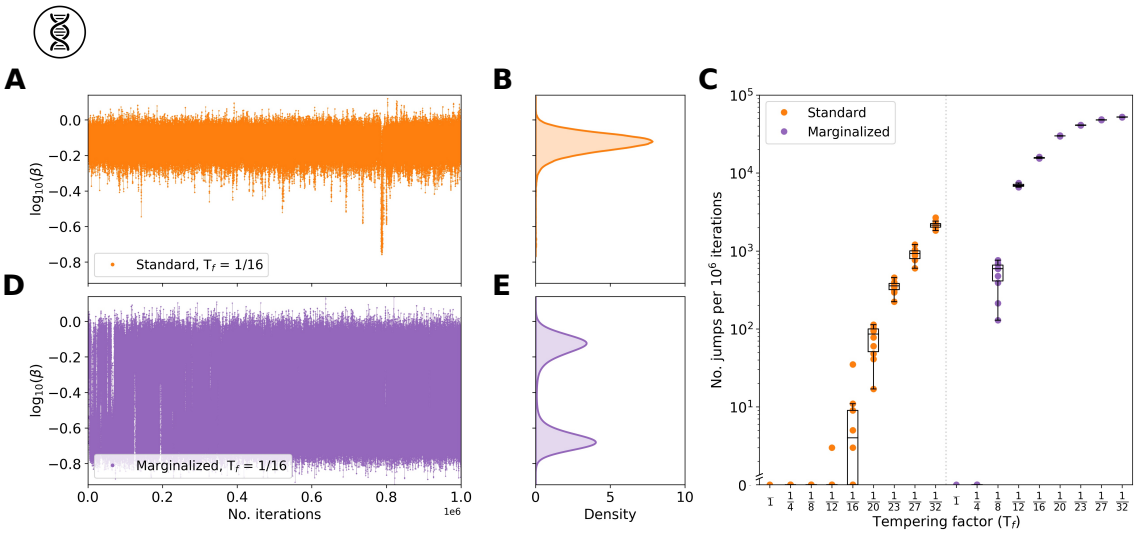

Figure 4.4.: Quantification of the transitions between the posterior modes for different tempering factors (T_f) for model M3. We refer to tempering factor T_f as the inverse of temperature T, therefore $T_f := 1/T$. (A) Parameter trace and (B) marginal distribution for the model parameter β obtained with the standard approach for $T_f = 1/16$. (C) Number of transitions per 10^6 iterations for a range of T_f values for the standard (orange) and marginalized approach (purple). (D) Parameter trace and (E) marginal distribution for the model parameter β obtained with the marginalized approach for $T_f = 1/16$. The depicted marginal distributions (B,E) were computed using a kernel density estimate.

4.5. Application example: A mechanistic model of gastric cancer signaling

The models we have considered until this point to test the marginalized approach ranged from small to medium size. To further evaluate the power and capabilities of the proposed method, we selected a substantially larger model: a pathway model used to describe the dynamics of cetuximab responder and non-responder gastric cancer cell lines, which we introduced in Chapter 3.

In this section, we focused on the individual cell line model for the MKN1 cell line, as we found that optimization for this model converged very reliably (Figure 3.3). The available experimental data for the MKN1 cell line are a total of 303 data points under 106 different experimental conditions for 31 observables. For each data point, the corresponding variance of the measurement noise is provided, therefore it does not need to be estimated (Section 3.2.2). Accordingly, we used the derivation presented in Annex D.1, which we also used for model M1. For details on the model and experimental data, we refer to Chapter 3.

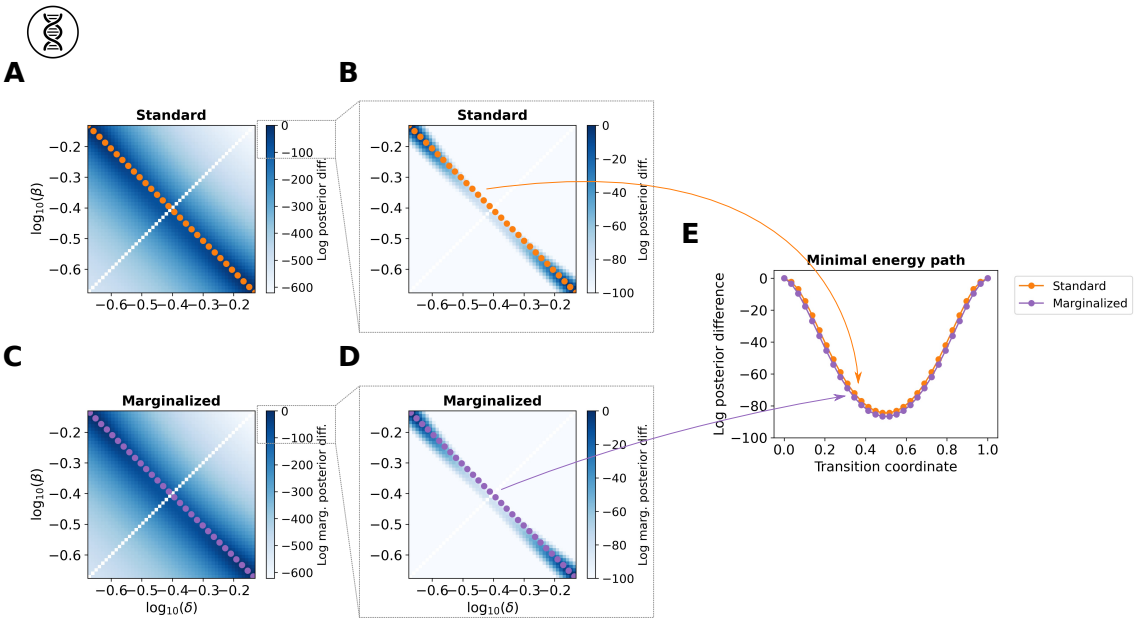

Figure 4.5.: Comparison of the minimal energy path for model M3. Landscape of the optimized (A,B) posterior and (C,D) marginalized posterior for different fixed values of the model parameters β and δ. The difference with respect to the maximal posterior value is depicted. (E) Transition coordinates for the minimal energy path. The minimal energy paths were constructed using the Python package `mep` available at `https://github.com/chc273/mep`.

This model contains 57 unknown parameters in total, of which 26 are model parameters and 31 are scaling constants. This implies that, by using our marginalized approach, we could reduce by more than half the dimensionality of the problem (Figure 4.6A). We found a perfect agreement (Pearson correlation, $r = 1$) between the analytical and numerical evaluation of the integral (Figure 4.6B) which confirmed the correctness of our analytical integration. For sampling, we employed two different algorithms, namely adaptive Metropolis (Haario et al., 2001) and parallel tempering (Łacki & Miasojedow, 2015), and performed 10 independent runs for each. As reported for M1–M3, the model fit for the best sample found was indistinguishable for both approaches (Figure 4.6C–D). However, we found differences in the marginal distributions of the model parameters, e.g., the marginalized approach found broader parameter distributions than the standard approach (Figure 4.6E–F). Interestingly, if we focused on the results obtained only by the standard approach for one of the sampled parameters, a huge variability is shown in the shape of its marginal distribution across the different runs and algorithms employed (Figure 4.6E). On the contrary, for the same parameter in the results only by the marginalized approach the shape of its marginal distribution was overlapping across all realizations (Figure 4.6F) for

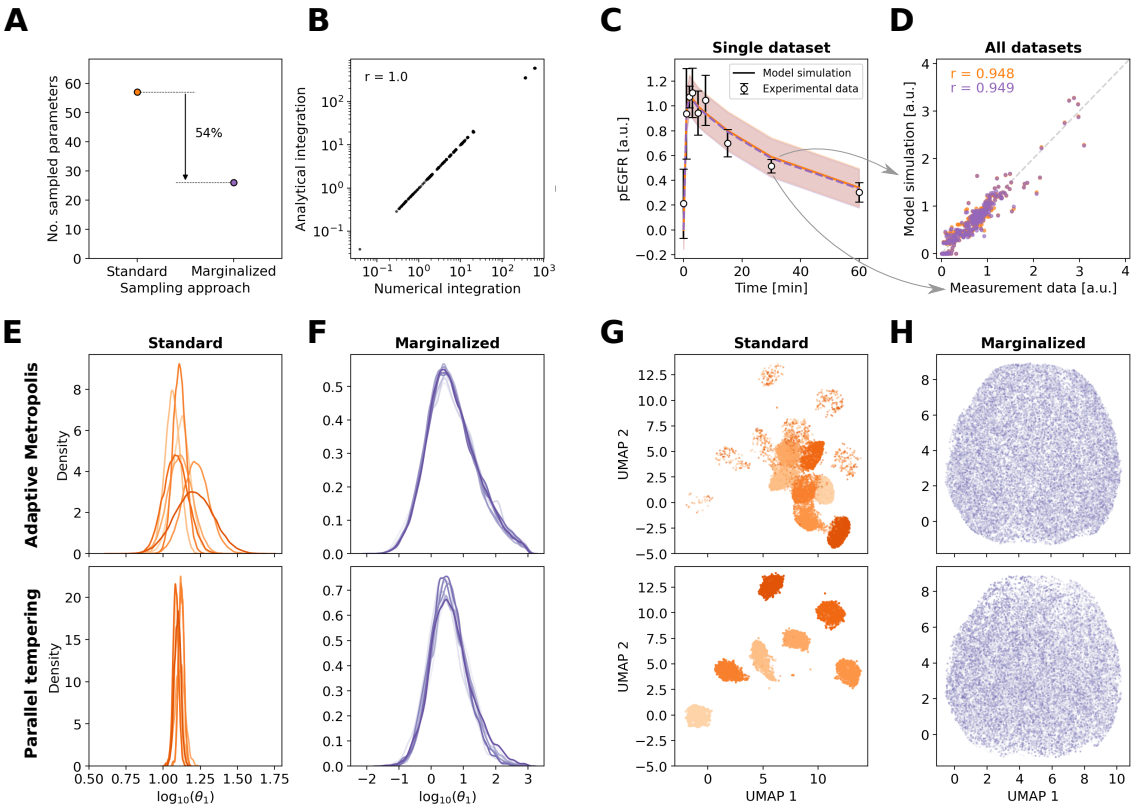

Figure 4.6.: Convergence of the marginalized approach for the gastric cancer model. (A) Number of sampled parameters. (B) Scatter plot for the agreement of analytical and numerical integration. Model fit of the best sample found during sampling for, (C) a subset of the experimental data and (D) the complete dataset in form of a scatter plot, the standard (orange) and marginalized approach (purple). Parameter marginal posterior distribution obtained using the (E) standard and (F) marginalized approach computed using a kernel density estimate for model parameter θ_1. Dimensionality reduction for all samples from all runs for the (G) standard and (H) marginalized approach using the UMAP representation. Different shades correspond to individual runs. (E–H) Results from adaptive Metropolis (top) and parallel tempering (bottom) are shown. The UMAPs were constructed using the Python package `umap` (McInnes et al., 2018).

all runs. To verify that the behavior observed for the parameter shown in Figure 4.6E–F is present in all other parameters, we analyzed the overall agreement of all parameter samples across all available runs for the standard and marginalized approach using the uniform manifold approximation and projection (UMAP) representation. We found that while in the standard approach all the individual runs clustered independently (Figure 4.6G), in the marginalized approach they were indistinguishable (Figure 4.6H). This revealed that: (i) in the marginalized approach all the individual runs actually converged closer to the posterior since all of them found the same distribution, and (ii) the standard approach failed for both

algorithms considered here. Our results suggest that the marginalized sampling approach is an excellent candidate for Bayesian parameter inference in large models with relative measurements.

In summary, the application of our marginalized approach to Bayesian inference for models with relative measurement data shows consistently that our approach yields the same marginal distributions for the parameters as the standard approach, while being highly more efficient in exploring the parameter space and enabling Bayesian inference of larger models, which was not possible before with the standard approach.

4.6. Summary and discussion

Bayesian parameter estimation for models with relative data requires, on top of the assumption of a statistical noise model, observable parameters, such as scaling or offset constants (Raue et al., 2013b). This leads to an increased dimensionality of the estimation problem, which in cases can imply an unfeasible assessment of the posterior distribution. To circumvent this, we introduced a framework in which we marginalize out these additional dimensions for additive Gaussian noise.

We evaluated the performance of our marginalized approach and compared it to the standard approach for four published models, with differences in their complexity. We showed a higher efficiency of the marginalized approach in terms of: (i) increased effective sample size per unit of time, and (ii) increased transition probabilities between posterior modes. We found our approach to be not only more efficient in exploring the parameter space, but also enabling the assessment of the posterior distribution for larger models, specifically the gastric cancer model presented in Chapter 3, for which the standard approach failed.

In this thesis, we mainly focused on the assessment of parameter uncertainties. However, this approach is also applicable for the evaluation of prediction uncertainties (Raue et al., 2013a). In (Villaverde et al., 2019) we found that the performance of standard MCMC sampling for prediction uncertainty quantification could suffer from convergence problems, especially for large models. This could be circumvented by using the marginalized approach, which was not available before.

The approach presented here is not limited to relative measurement data, but also applicable to absolute measurements. Since for these, the noise parameters would still have to be inferred. We provide the derivation in Annex D.2. Also, our approach can be used for combinations of relative and absolute data, and it is applicable to different measurement

process functions, e.g. offset constants (and their combination with scaling constants), and noise models, e.g. Laplace-distributed noise. In (Fedders, 2021), a Bachelor thesis that I supervised, we provided the marginalization-based approach for offset parameters under additive Gaussian noise.

The choice of conjugate priors for the marginalized parameters eased the analytical derivation of the marginal posterior. Yet, this implies in our case that observable and noise parameters are not independent. Mostly, this is not a problem since both parameters are related to the measurement process. However, in some cases, it may be that it is known that they are independent, therefore other prior distribution assumptions must be considered. It should be noted that the concept of marginalization is not restricted to integrals that are analytically solvable, but also numerical integration schemes can be considered. However, this would in part increase the required computation time (as observed in Figure 4.2C), but very likely the improved mixing properties would be maintained.

Our proposed approach is not bounded to only models based on ODEs, but in practise it can also be used with other types of models, e.g. based on partial or stochastic differential equations, as just the model simulation outputs and experimental data are needed. Furthermore, this method is widely applicable to mathematical models from different research fields, such as engineering, physics and ecology.

There exist MCMC algorithms that use the gradient information of the posterior, such as Hamilton Monte Carlo sampling (Graham & Storkey, 2017; Hoffman & Gelman, 2014) or the Metropolis-adjusted Langevin algorithm (Girolami & Calderhead, 2011). To apply our method to these samplers, derivatives of the marginal likelihood function need to be implemented.

In summary, the proposed approach provides a powerful tool for Bayesian parameter inference for mathematical models with observation-related parameters by reducing the dimension of the parameter space. It substantially benefits the efficiency of sampling-based approaches, and renders the generation of representative posterior samples for large models possible.

Chapter 5

Summary and conclusion

Mathematical models can be used to gain mechanistic insights into biological systems. Frequently, a biological question triggers a cyclic process which involves: (i) data collection and model development, (ii) parameter estimation, and (iii) uncertainty analysis. All these steps are interconnected and give rise to an iterative process with feedback loops between the different stages. Each step possesses different challenges depending on the biological questions and data types considered. In this thesis, we covered all these stages and addressed the research questions posed in Section 1.3. To address these challenges, we first introduced the necessary mathematical background and notation in Chapter 2. Based on this, the contributions of this thesis to tackle the open problems were presented in Chapters 3 and 4.

In Chapter 3, we helped addressing the lack of mathematical models in the context of gastric cancer research and we provided a basis for the understanding of the heterogeneity in patient response to cetuximab treatment. For this, we built a collection of mechanistic models based on ODEs describing the EGFR, the AKT and the ERK signaling pathways in a cetuximab responder and non-responder cell line. This model collection considered different potential sources of variability between the responder and non-responder cell line beyond their expression and mutation patterns. We found the model including cell line specific differences in only the EGFR turnover dynamics to be the most appropriate based on the AIC, which was supported by previous experimental studies. Parameter uncertainty analysis on the optimal candidate model revealed that less than 50% of the parameters were practically identifiable. In spite of this, the optimal model was able to reliably predict the response to multiple validation experiments. Furthermore, we employed the optimal model to predict response and resistance factors, in particular EGFR and MMET abundance which were experimentally confirmed by knockdown experiments. All these successful validations indicated that our proposed mechanistic model could be used to guide biomarker discovery in cetuximab treatment. Moreover, the model was implemented employing the SBML and `PEtab` standard formats, making it easily reusable to other researchers.

In Chapter 4, we tackled the limited applicability of sampling-based methods for Bayesian parameter inference to large models, in particular when these are calibrated on relative measurement data. In this specific case, observable and noise parameters need to be considered, hence increasing the dimensionality of the estimation problem. To this end, we proposed a marginalization-based approach that integrates over observation-related parameters. Under the assumption of additive Gaussian noise, we obtained the marginal likelihood – which only depends on the model parameters – by analytically integrating over the observable and noise parameters. Our proposed approach yielded a lower-dimensional posterior that can be explored more easily using sampling-based methods. Then, we derived the distributions to sample and retrieve the marginalized parameters. We tested the performance of our method in a collection of published models. Firstly, we found that the marginalization-based approach provided the same posterior distribution as sampling in the full parameter space. Secondly, we obtained a higher exploration efficiency in the parameter space in terms of increased effective sample size per unit of time and improved transition rates between posterior modes. Lastly, we showed how our method enabled the Bayesian quantification of parameter uncertainty in a large model, specifically the gastric cancer model presented in Chapter 3, which failed for standard methods. All these findings suggested that the marginalization-based approach is an excellent candidate for Bayesian parameter inference in models with observation-related parameters.

We discussed ideas for possible extensions of all contributions of this thesis in the corresponding Sections 3.6 and 4.6. These ideas comprised the extension of the marginalization-based approach to additional observable formulations and noise models, or the derivation of first-order derivatives of the marginal posterior. A further possible extension would be the gradient calculation via adjoint sensitivity analysis (Fröhlich et al., 2019; Schmiester et al., 2019). These extensions would improve even further the scalability of the proposed approach to large-scale models, such as the pan-cancer model introduced by Fröhlich et al. (2018). In Chapter 4, we only applied our approach to the mechanistic model for the cetuximab responder cell line, but we did not consider the integrated model for the cetuximab responder and non-responder as optimization was more challenging. With the method extensions previously mentioned, a global and exhaustive analysis of parameter and prediction uncertainties using MCMC sampling could be performed for the integrated model. This would provide additional mechanistic insights into the differences between both cell lines. We could consider further extensions of the integrated model to include spatial information, such as partial differential equations or more advanced agent-based models, or to include additional immunotherapy treatments targeting other members of the HER family, such as trastuzumab an anti-HER2 antibody. For the latter, part of the cell line experimental data that we published in (Ebert et al., 2020) and patient data that

we presented in (Haffner et al., 2021) could be employed for model calibration (on top of additional data sources). Moreover, the combination of trastuzumab with pembrolizumab, which is an anti-PD-1 antibody, has been shown to significantly improve treatment response (Janjigian et al., 2021).

To conclude this thesis, we developed (i) an integrated mathematical model to describe cetuximab response in multiple gastric cancer cell lines and (ii) an efficient method for Bayesian parameter inference based on the marginalization of observation-related parameters. For the first main contribution (i), we demonstrated the suitability and potential of the integrated model under an exhaustive array of analyses. For the second main contribution (ii), we showed that the proposed method could be used on real-application examples to substantially ease the assessment of parameter uncertainties and enable the integration of larger datasets with larger models. This demonstrates that the contributions developed in this thesis pave the way towards the aim of systems biology facilitating a deeper, holistic and mechanistic understanding of biological systems.

Appendix

A. PubMed search query terms

To generate Figure 1.1, we employed the following search query terms:

- For **cancer** – without distinction of types, labelled as *All cancer* in Figure 1.1:

  ("computational model" OR "mechanistic model" OR "mathematical model") AND "cancer".

- For **breast cancer**:

  ("computational model" OR "mechanistic model" OR "mathematical model") AND "breast cancer".

- For **lung cancer**:

  ("computational model" OR "mechanistic model" OR "mathematical model") AND "lung cancer".

- For **gastric cancer**:

  ("computational model" OR "mechanistic model" OR "mathematical model") AND ("stomach cancer" OR "gastric cancer").

B. Complete model-data fit of the considered gastric cancer models

Individual cell line model

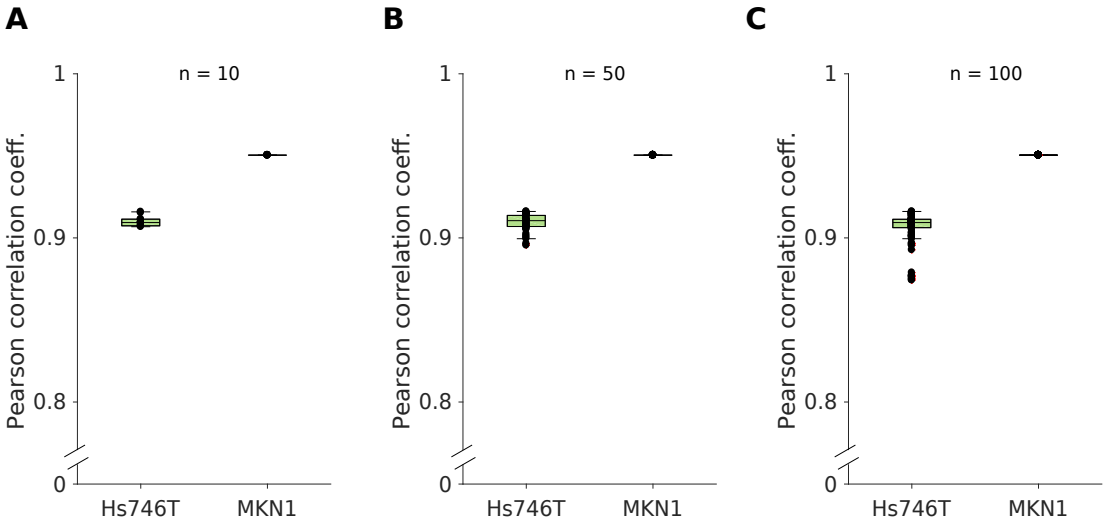

Figure B.1.: Overview on model simulation and data correlation for multiple parameter sets on the individual cell line models. Boxplots for the overall agreement of experimental data and model simulation for, (A) the best 10 parameter sets, (B) the best 50 parameter sets, and (C) the best 100 parameter sets. The individual model fits for Hs746T and MKN1 cells are shown. Adapted from Figure S7 of the thesis author's publication (Raimúndez et al., 2020).

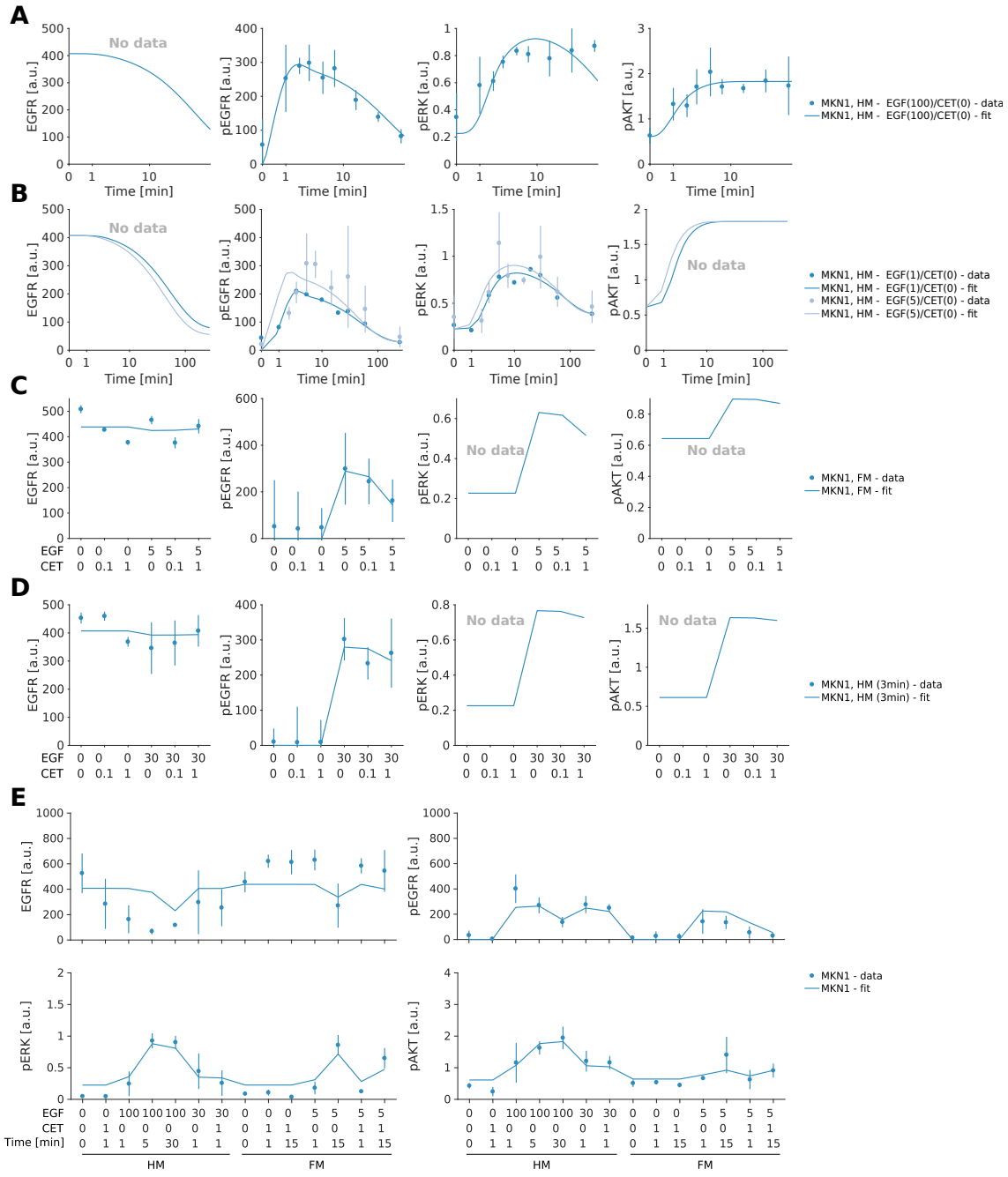

Figure B.2.: Model-data comparison for the MKN1 cell line, for datasets not depicted in Figure 3.3. (A-B) Time response to different EGF concentrations in starvation culture media (HM). (C) Dose response to EGF and cetuximab stimulation at 3 min in rich culture media (FM). (D) Dose response to EGF and cetuximab stimulation at 3 min in starvation culture media (HM). (E) Dose response to EGF and cetuximab stimulation at 0, 1, 15 and 30 min in full (FM) and starvation culture media (HM). (C-E) Specific EGF and cetuximab concentrations are shown along the X axis. Adapted from Figure S3 of the thesis author's publication (Raimúndez et al., 2020).

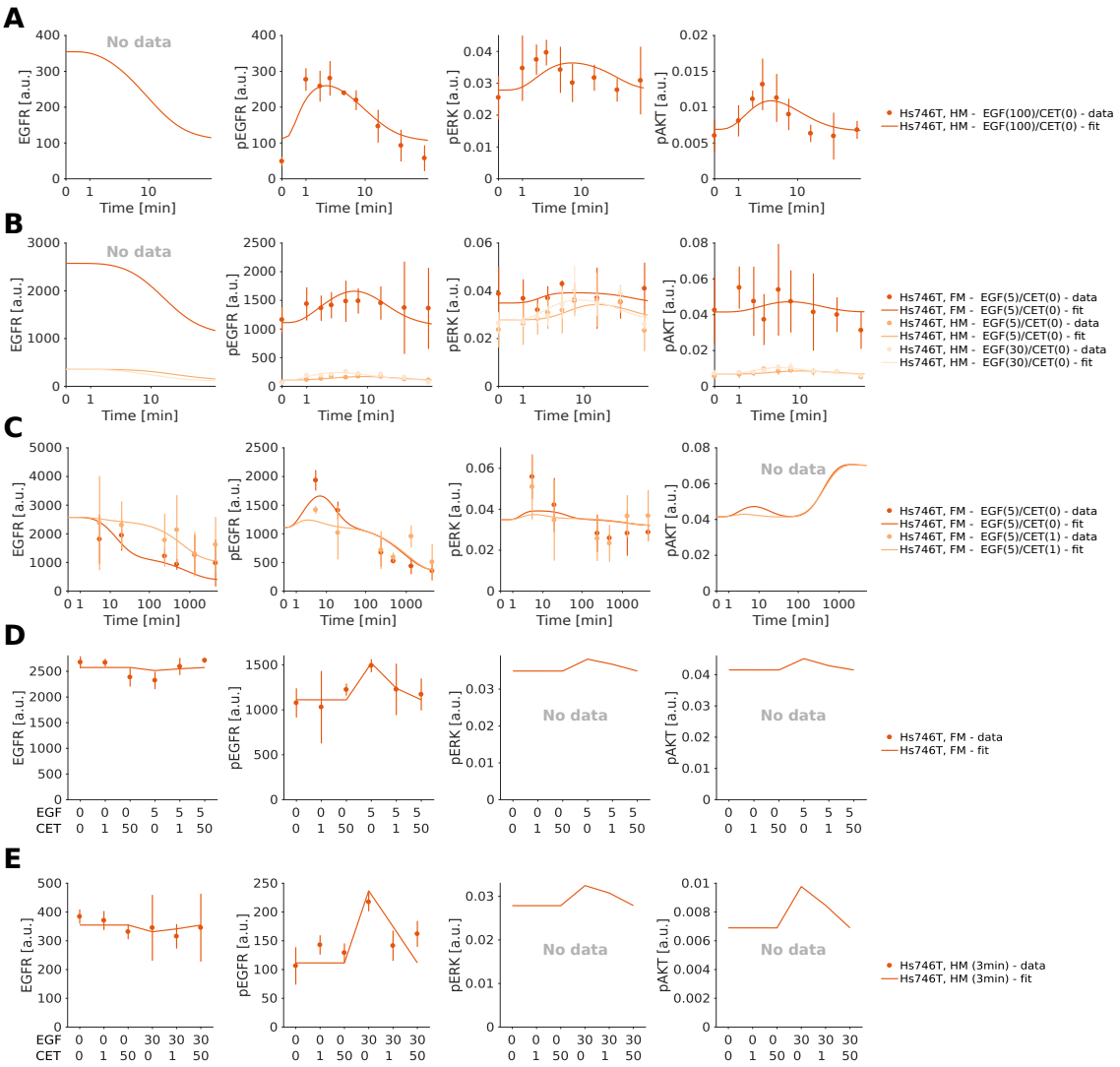

Figure B.3.: Model-data comparison for the Hs746T cell line, for datasets not depicted in Figure 3.4. (A) Time response to EGF stimulation in starvation culture media (HM). (B) Time response to EGF stimulation in full (FM) and starvation culture media (HM). (C) Time response to EGF and cetuximab stimulation in rich culture media (FM). (D) Dose response to EGF and cetuximab stimulation at 3 min in rich culture media (FM). (E) Dose response to EGF and cetuximab stimulation at 3 min in starvation culture media (HM). (D-E) Specific EGF and cetuximab concentrations are shown along the X axis. Adapted from Figure S4 of the thesis author's publication (Raimúndez et al., 2020).

Combined cell line model

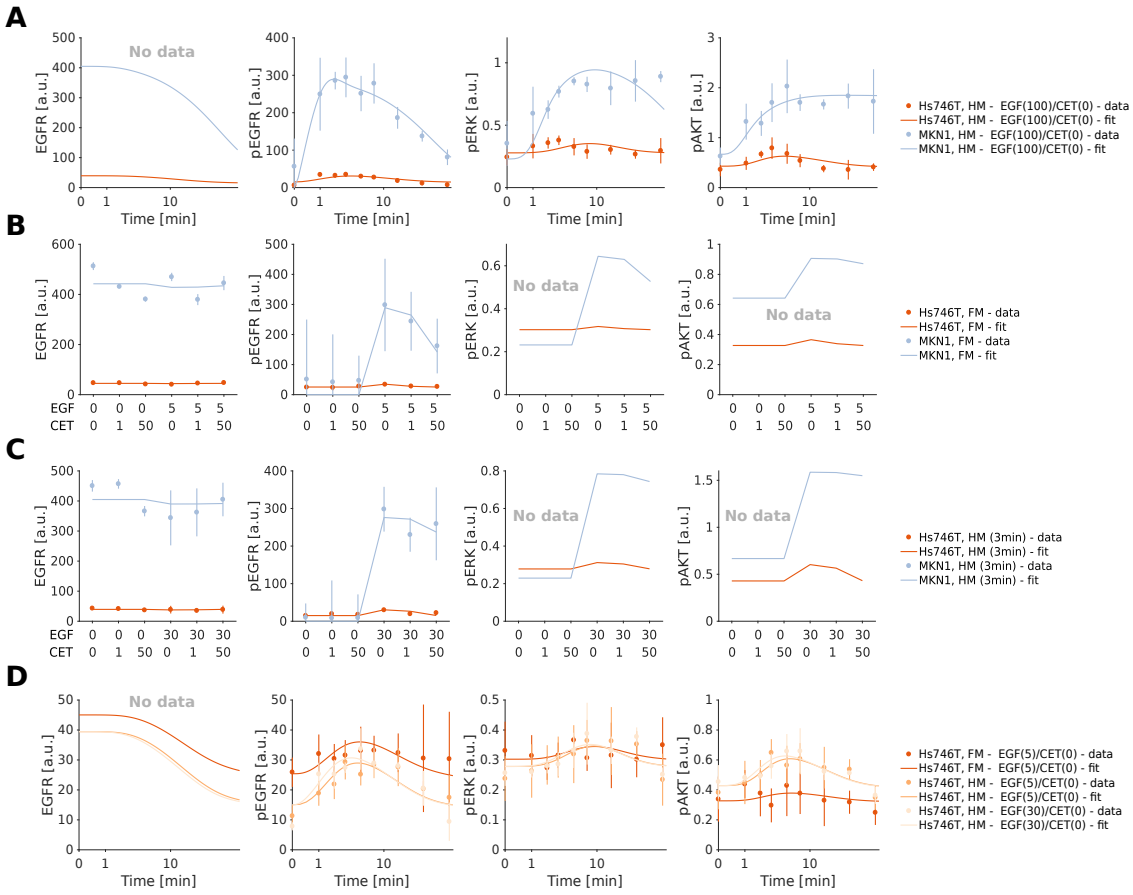

Figure B.4.: Model-data comparison for the combined fitting of MKN1 and Hs746T cell lines, for datasets not depicted in Figures 3.5 and 3.6. Model fits for the best model (M5). (A) Time response to EGF stimulation in starvation culture media (HM). (B) Dose response to EGF and cetuximab stimulation at 3 min in rich culture media (FM). (C) Dose response to EGF and cetuximab stimulation at 3 min in starvation culture media (HM). (D) Time response to EGF stimulation of Hs746T cells in full (FM) and starvation culture media (HM). (A-C) Experimental data for both cell lines. (B-C) Specific EGF and cetuximab concentrations are shown along the X axis. Adapted from Figure S5 of the thesis author's publication (Raimúndez et al., 2020).

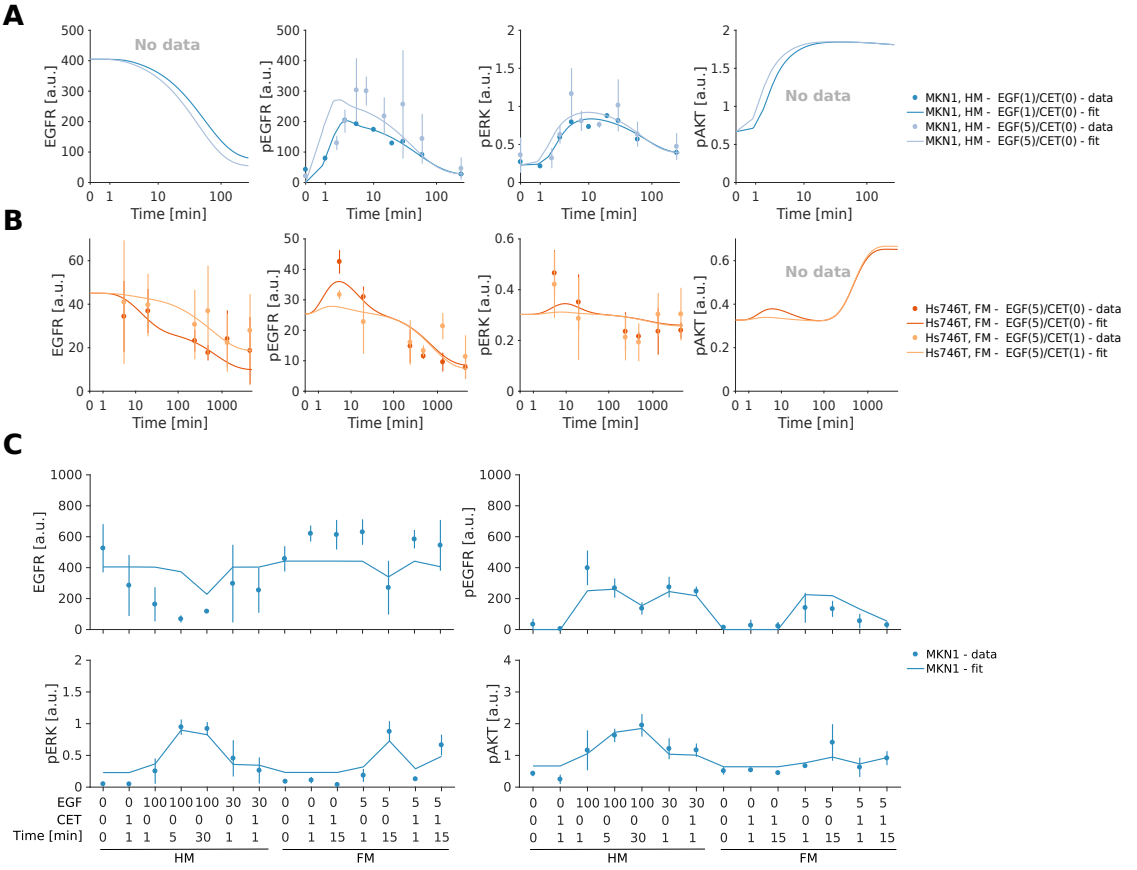

Figure B.5.: Model-data comparison for the combined fitting of MKN1 and Hs746T cell lines, for datasets not depicted in Figures 3.5 and 3.6. Model fits for the best model (M5). (A) Time response to EGF and cetuximab stimulation of MKN1 cells in starvation culture media (HM). (B) Time response to EGF and cetuximab stimulation of Hs746T cells in rich culture media (FM). (C) Dose response to EGF and cetuximab stimulation at 0, 1, 15 and 30 min of MKN1 cells in rich (FM) and starvation culture media (HM). Specific EGF and cetuximab concentrations, time points and culture media, are shown along the X axis. Adapted from Figure S6 of the thesis author's publication (Raimúndez et al., 2020).

Model prediction of pAKT insensitivity to cetuximab treatment in MKN1 cells

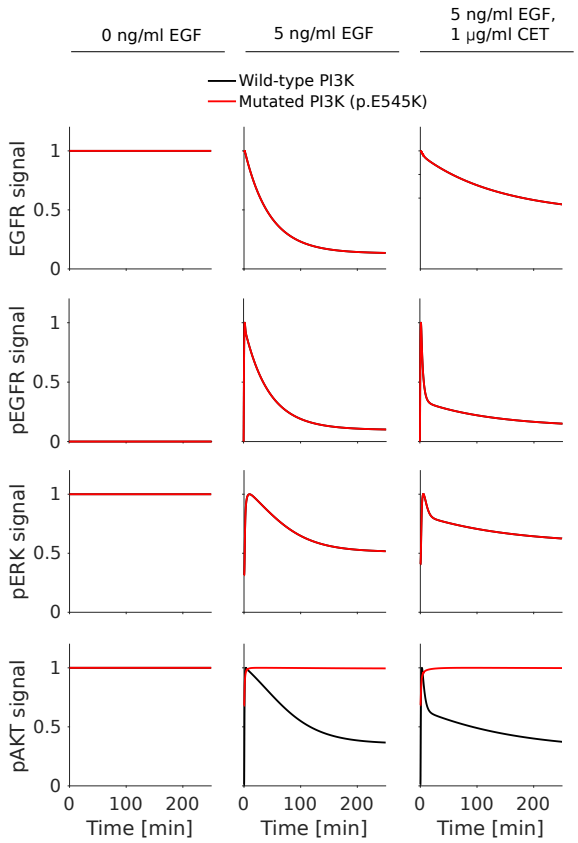

Figure B.6.: Expression of wild-type PI3K recovers pAKT levels sensitivity to cetuximab treatment in MKN1 cell line. Model prediction of time response of wild-type PI3K (black line) shows a reduction in AKT activity compared to PI3K p.E545K (red line), which remains insensitive. The signal is normalized with respect to the maximum activity level for each observed component. Adapted from Figure S2 of the thesis author's publication (Raimúndez et al., 2020).

C. Parameter distributions and complete model-data fit of models M1–M3

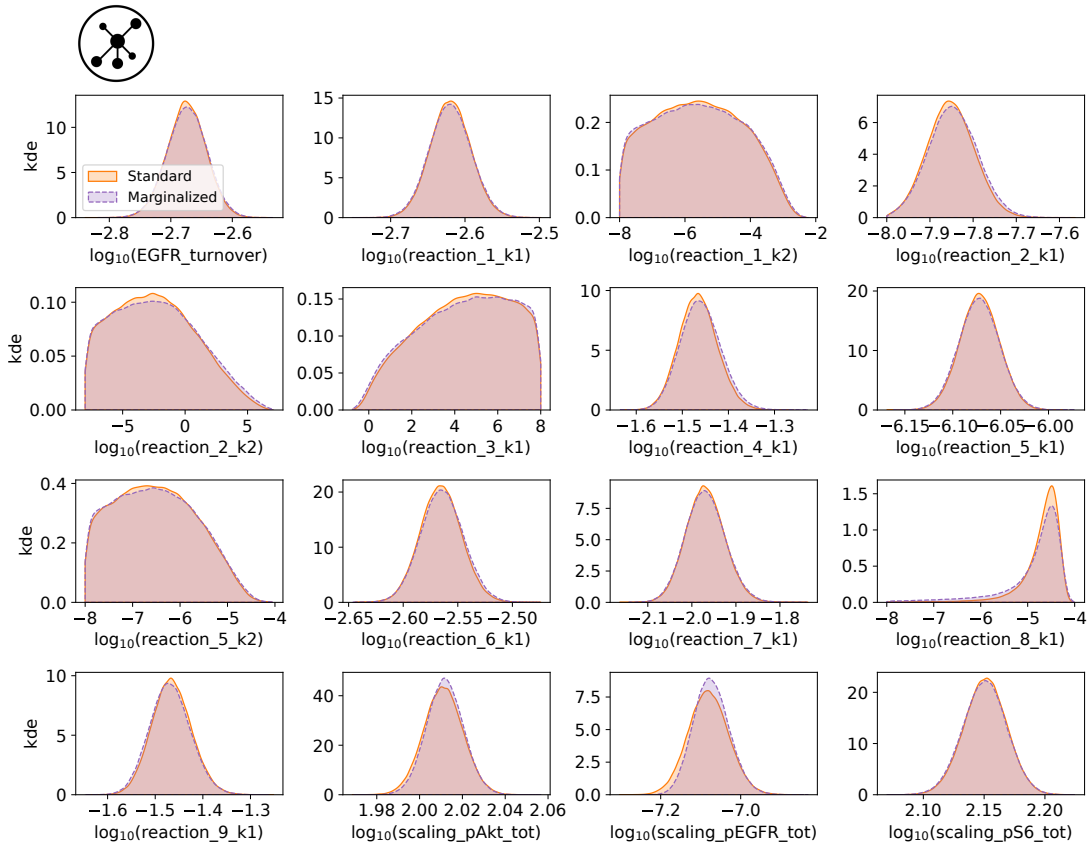

Figure C.1.: Parameter marginal posterior distributions computed using a kernel density estimate for model M1. The marginalized parameters correspond to those denoted with *scaling_**.

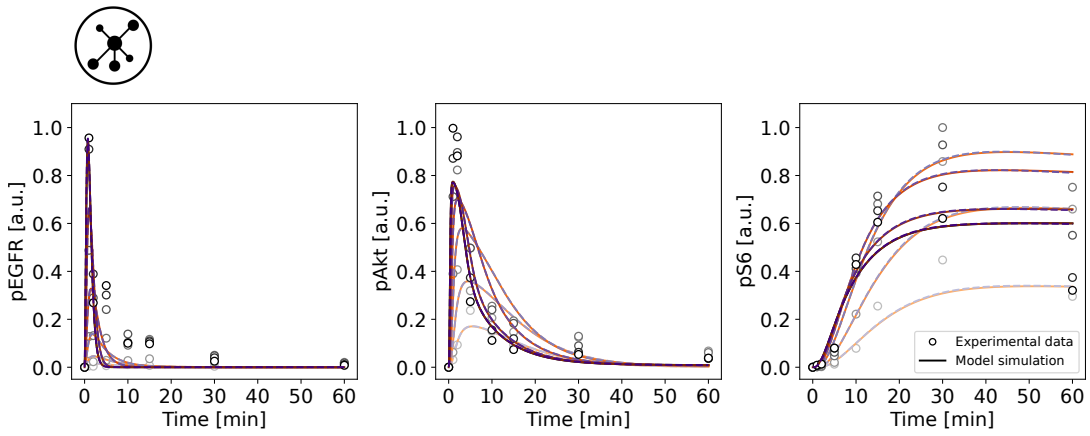

Figure C.2.: Complete dataset and model fit for model M1. Model simulation of the best sample found for the standard approach is depicted in orange and for the marginalized approach in purple. Different shades indicate different experimental conditions.

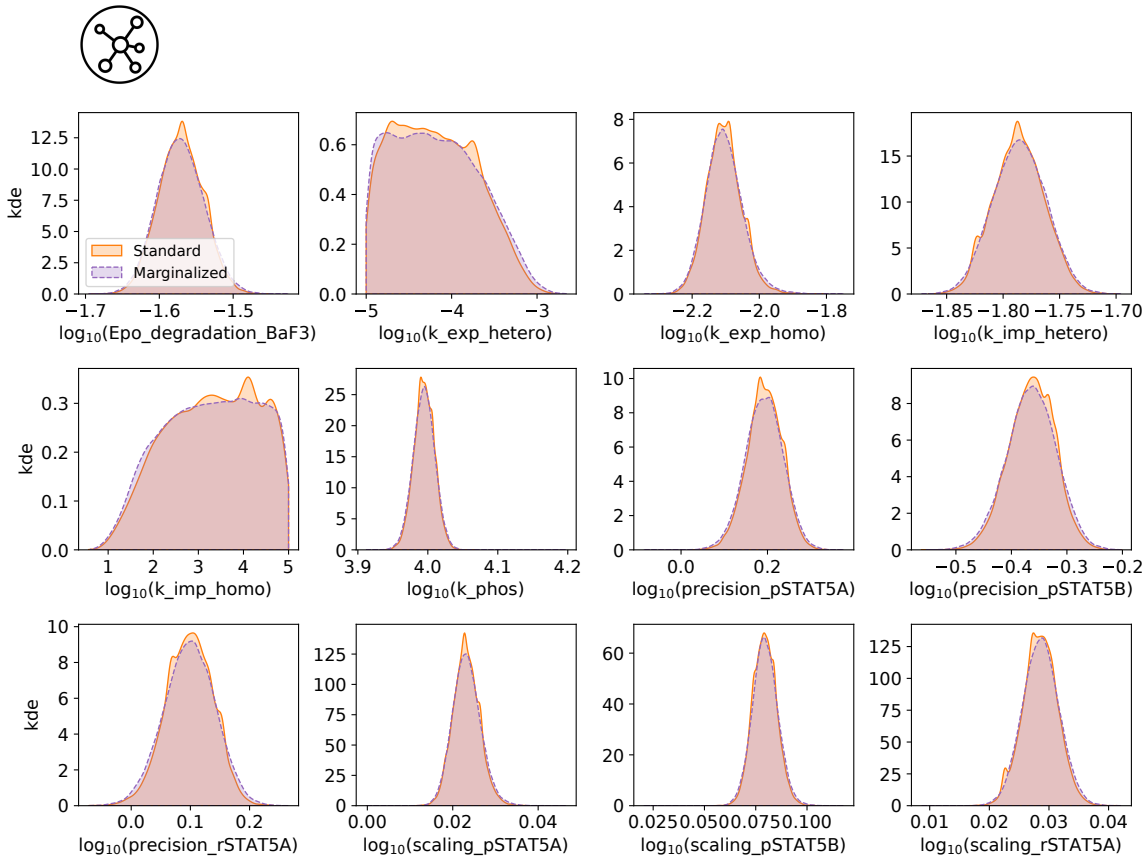

Figure C.3.: Parameter marginal posterior distributions computed using a kernel density estimate for model M2. The marginalized parameters correspond to those denoted with *scaling_** and *precision_**.

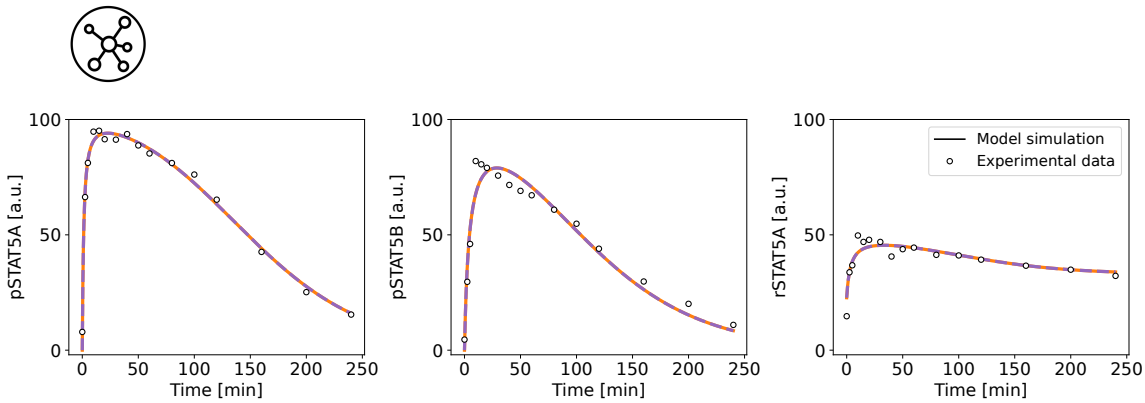

Figure C.4.: Complete dataset and model fit for model M2. Model simulation of the best sample found for the standard approach is depicted in orange and for the marginalized approach in purple.

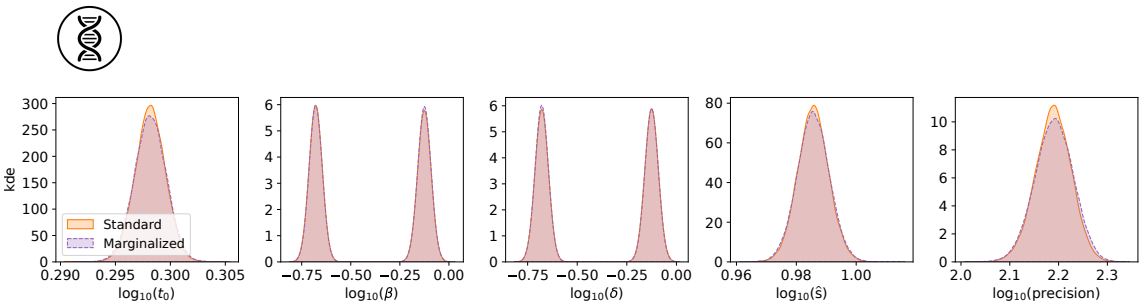

Figure C.5.: Parameter marginal posterior distributions computed using a kernel density estimate for model M3. The marginalized parameters correspond to $\hat{s}$ and *precision*.

D. Marginalization-based approach for additional observable structures under additive Gaussian noise

Here, we provide the derivation of the marginalization-based approach for additional observable structures as mentioned in Chapter 4 and we use the notation employed therein.

D.1. Relative measurement data with known variance of the measurement noise

In the case that the variance (or standard deviation) of the measurement noise is known (i.e. measured), we still need to infer the values for the scaling parameters s. Hence, we present here the corresponding marginalization assuming a Gaussian-distributed prior given by

$$p(s) = \left(\frac{\hat{\lambda}}{2\pi}\right)^{1/2} \exp\left(-\frac{\hat{\lambda}}{2}(s-\mu)^2\right)$$

where $\mu \in \mathbb{R}$ and $\hat{\lambda} \in \mathbb{R}_+$ are the known mean and precision of s respectively.

We start with

$$p(\mathcal{D}|\theta) = \left(\frac{\lambda}{2\pi}\right)^{n_t/2} \underbrace{\int_{-\infty}^{\infty} \exp\left(-\frac{\lambda}{2}\sum_{k=1}^{n_t}\left(\overline{y}_k - s\cdot h_k\right)^2\right) p(s)\, ds}_{(*)} \tag{D.1}$$

with

$$
\begin{aligned}
(*) &= \int_{-\infty}^{\infty} \exp\left(-\frac{\lambda}{2}\sum_{k=1}^{n_t}(\overline{y}_k - s\cdot h_k)^2\right)\left(\frac{\hat{\lambda}}{2\pi}\right)^{1/2}\exp\Big(-\frac{\hat{\lambda}}{2}(s-\mu)^2\Big)\,ds \\
&= \left(\frac{\hat{\lambda}}{2\pi}\right)^{1/2}\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2}\left(\lambda\sum_{k=1}^{n_t}h_k^2+\hat{\lambda}\right)s^2+\left(\lambda\sum_{k=1}^{n_t}\bar{y}_k h_k+\hat{\lambda}\mu\right)s-\frac{1}{2}\left(\lambda\sum_{k=1}^{n_t}\bar{y}_k^2+\hat{\lambda}\mu^2\right)\right)ds
\end{aligned}
\tag{D.2}
$$

and with constants

$$
a := \frac{1}{2}\left(\lambda\sum_{k=1}^{n_t}h_k^2+\hat{\lambda}\right),\qquad b := \left(\lambda\sum_{k=1}^{n_t}\bar{y}_k h_k+\hat{\lambda}\mu\right),\qquad c := -\frac{1}{2}\left(\lambda\sum_{k=1}^{n_t}\bar{y}_k^2+\hat{\lambda}\mu^2\right).
$$

We can solve $(*)$ using (4.9), therefore we obtain

$$
\begin{aligned}
(*) &= \left(\frac{\hat{\lambda}}{2\pi}\right)^{1/2}\sqrt{\frac{\pi}{\frac{1}{2}\left(\lambda\sum_{k=1}^{n_t}h_k^2+\hat{\lambda}\right)}}\exp\left(\frac{\left(\lambda\sum_{k=1}^{n_t}\bar{y}_k h_k+\hat{\lambda}\mu\right)^2}{4\frac{1}{2}\left(\lambda\sum_{k=1}^{n_t}h_k^2+\hat{\lambda}\right)}-\frac{1}{2}\left(\lambda\sum_{k=1}^{n_t}\bar{y}_k^2+\hat{\lambda}\mu^2\right)\right) \\
&= \left(\frac{\hat{\lambda}}{\lambda\sum_{k=1}^{n_t}h_k^2+\hat{\lambda}}\right)^{1/2}\exp\left(\frac{1}{2}\left(\frac{\left(\lambda\sum_{k=1}^{n_t}\bar{y}_k h_k+\hat{\lambda}\mu\right)^2}{\lambda\sum_{k=1}^{n_t}h_k^2+\hat{\lambda}}-\lambda\sum_{k=1}^{n_t}\bar{y}_k^2-\hat{\lambda}\mu^2\right)\right) \\
&= \left(\frac{\hat{\lambda}}{C}\right)^{1/2}\exp\left(\frac{1}{2}\left(\frac{\left(\lambda\sum_{k=1}^{n_t}\bar{y}_k h_k+\hat{\lambda}\mu\right)^2}{C}-\lambda\sum_{k=1}^{n_t}\bar{y}_k^2-\hat{\lambda}\mu^2\right)\right)
\end{aligned}
$$

with

$$
C := \lambda\sum_{k=1}^{n_t}h_k^2+\hat{\lambda}.
$$

Substitution of $(*)$ in (D.1) yields

$$
p(\mathcal{D}|\theta) = \left(\frac{\lambda}{2\pi}\right)^{n_t/2}\left(\frac{\hat{\lambda}}{C}\right)^{1/2}\exp\left(\frac{1}{2}\left(\frac{\left(\lambda\sum_{k=1}^{n_t}\bar{y}_k h_k+\hat{\lambda}\mu\right)^2}{C}-\lambda\sum_{k=1}^{n_t}\bar{y}_k^2-\hat{\lambda}\mu^2\right)\right).
$$

Similarly, in the case of having time-point specific variances, we can generalize to

$$
p(\mathcal{D}|\theta) = \left(\frac{1}{2\pi}\right)^{n_t/2}\left(\frac{\hat{\lambda}}{C}\right)^{1/2}\left(\prod_{k=1}^{n_t}\lambda_k\right)^{1/2}\exp\left(\frac{1}{2}\left(\frac{\left(\sum_{k=1}^{n_t}\lambda_k\bar{y}_k h_k+\hat{\lambda}\mu\right)^2}{C}-\sum_{k=1}^{n_t}\lambda_k\bar{y}_k^2-\hat{\lambda}\mu^2\right)\right)
$$

with $C := \sum_{k=1}^{n_t} \lambda_k h_k^2 + \hat{\lambda}$.

Retrieving the marginalized observable parameters

For s we can consider the integrand of (D.2). We receive

$$\begin{aligned} s &\sim \exp\left(-\frac{1}{2}\left(\sum_{k=1}^{n_t}\lambda_k h_k^2 + \hat{\lambda}\right)s^2 + \left(\sum_{k=1}^{n_t}\lambda_k \bar{y}_k h_k + \hat{\lambda}\mu\right)s\right) \\ &\sim \exp\left(-\frac{1}{2}\frac{\left(s - \frac{\left(\sum_{k=1}^{n_t}\lambda_k \bar{y}_k h_k + \hat{\lambda}\mu\right)}{\left(\sum_{k=1}^{n_t}\lambda_k h_k^2 + \hat{\lambda}\right)}\right)^2}{1/\left(\sum_{k=1}^{n_t}\lambda_k h_k^2 + \hat{\lambda}\right)}\right) \\ &\sim \mathcal{N}\left(\mu' = \frac{\left(\sum_{k=1}^{n_t}\lambda_k \bar{y}_k h_k + \hat{\lambda}\mu\right)}{\left(\sum_{k=1}^{n_t}\lambda_k h_k^2 + \hat{\lambda}\right)}, (\sigma')^2 = \left(\sum_{k=1}^{n_t}\lambda_k h_k^2 + \hat{\lambda}\right)^{-1}\right) \end{aligned}$$

The distribution is well-defined if $\mu' \in \mathbb{R}$ and $(\sigma')^2 \in \mathbb{R}_+$, which holds for both and is clear by definition.

D.2. Absolute measurement data with unknown variance of the measurement noise

Sometimes the experimental data available is absolute, therefore there is no need to include scaling parameters. However, noise parameters (σ^2) need to be inferred. We present here the marginalization of the likelihood for the case of absolute data, i.e. only integrating out σ^2, assuming a Gamma-distributed prior given by

$$p(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}\exp\left(-\beta\lambda\right), \quad \text{where } \lambda := 1/\sigma^2,$$

and with $\alpha, \beta \in \mathbb{R}_+$ denoting the hyper-parameters of the prior distribution, and $\Gamma(\cdot)$ the Gamma function.

We start with

$$\begin{aligned} p(\mathcal{D}|\theta) &= \int_0^\infty \left(\frac{\lambda}{2\pi}\right)^{n_t/2} \exp\left(-\frac{\lambda}{2}\sum_{k=1}^{n_t}\left(\overline{y}_k - h_k\right)^2\right)\frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}\exp\left(-\beta\lambda\right)\, d\lambda \\ &= \left(\frac{1}{2\pi}\right)^{n_t/2}\frac{\beta^\alpha}{\Gamma(\alpha)}\int_0^\infty \lambda^{(n_t/2)+\alpha-1}\exp\left(-\frac{\lambda}{2}\sum_{k=1}^{n_t}\left(\overline{y}_k - h_k\right)^2 - \beta\lambda\right)\, d\lambda \\ &= \left(\frac{1}{2\pi}\right)^{n_t/2}\frac{\beta^\alpha}{\Gamma(\alpha)}\int_0^\infty \lambda^{(n_t/2)+\alpha-1}\exp\left(-\lambda C\right)\, d\lambda \end{aligned} \tag{D.3}$$

with

$$C := \beta + \frac{1}{2}\sum_{k=1}^{n_t} (\overline{y}_k - h_k)^2 \,.$$

Using the transformation $\omega = \lambda C$, we get

$$\begin{aligned} p(\mathcal{D}|\theta) &= \left(\frac{1}{2\pi}\right)^{n_t/2} \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \left(\frac{\omega}{C}\right)^{(n_t/2)+\alpha-1} \exp(-\omega)\ \frac{1}{C} d\omega \\ &= \left(\frac{1}{2\pi}\right)^{n_t/2} \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{1}{C}\right)^{(n_t/2)+\alpha} \int_0^\infty \omega^{(n_t/2)+\alpha-1} \exp(-\omega)\ d\omega \\ &= \left(\frac{1}{2\pi}\right)^{n_t/2} \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{1}{C}\right)^{z} \int_0^\infty \omega^{z-1} \exp(-\omega)\ d\omega \end{aligned} \tag{D.4}$$

with $z := \alpha + n_t/2$.

As the integral in (D.4) follows the form of (4.12), we obtain

$$\begin{aligned} p(\mathcal{D}|\theta) &= \left(\frac{1}{2\pi}\right)^{n_t/2} \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{1}{C}\right)^{\alpha+n_t/2} \cdot \Gamma\left(\alpha + \frac{n_t}{2}\right) \\ &= \frac{(\beta/C)^\alpha}{\Gamma(\alpha)(2\pi C)^{n_t/2}} \cdot \Gamma\left(\alpha + \frac{n_t}{2}\right). \end{aligned} \tag{D.5}$$

After substituting C, we finally obtain the marginal likelihood (D.5) for absolute data with unknown variance of the measurement noise.

Retrieving the marginalized noise parameters

We know from the integrand of (D.3) that λ has – up to a constant – the distribution of a Gamma distribution, mainly

$$\lambda \sim \text{Gamma}(\alpha' = \alpha + \frac{n_t}{2}, \beta' = C).$$

The distribution is well-defined if $\alpha', \beta' \in \mathbb{R}_+$, which holds for both and is clear by definition.

Similarly, as $\sigma^2 = 1/\lambda$, we conclude that

$$\sigma^2 \sim \text{Inv-Gamma}(\alpha' = \alpha + \frac{n_t}{2}, \beta' = C).$$

Bibliography

Lorenz Adlung, Sandip Kar, Marie-Christine Wagner, Bin She, Sajib Chakraborty, Jie Bao, Susen Lattermann, Melanie Boerries, Hauke Busch, Patrick Wuchter, Anthony D Ho, Jens Timmer, Marcel Schilling, Thomas Höfer, & Ursula Klingmüller. Protein abundance of AKT and ERK pathway components governs cell type-specific regulation of proliferation. *Mol. Syst. Biol.*, 13(1): 904, 2017. ISSN 1744-4292. doi: 10.15252/msb.20167258.

H. Akaike. Information theory and an extension of the maximum likelihood principle. In *2nd International Symposium on Information Theory, Tsahkadsor, Armenian SSR*, volume 1, pages 267–281. Akademiai Kiado, 1973. doi: 10.1007/978-1-4612-1694-0_15.

Philipp M. Altrock, Lin L. Liu, & Franziska Michor. The mathematics of cancer: integrating quantitative models. *Nature Reviews Cancer*, 15(12):730–745, November 2015. doi: 10.1038/nrc4029.

J. Bachmann, A. Raue, M. Schilling, M. E. Böhm, C. Kreutz, D. Kaschek, H. Busch, N. Gretz, W. D. Lehmann, J. Timmer, & U. Klingmüller. Division of labor by dual feedback regulators controls JAK2/STAT5 signaling over broad ligand range. *Mol. Syst. Biol.*, 7(1):516, July 2011.

T. Bäck. *Evolutionary algorithms in theory and practice: evolution strategies, evolutionary programming, genetic algorithms*. Oxford University Press, New York and Oxford, 1996.

B. Ballnus, S. Hug, K. Hatz, L. Görlitz, J. Hasenauer, & F. J. Theis. Comprehensive benchmarking of Markov chain Monte Carlo methods for dynamical systems. *BMC Syst Biol*, 11(63):63, 2017. doi: 10.1186/s12918-017-0433-1.

Benjamin Ballnus, Steffen Schaper, Fabian J Theis, & Jan Hasenauer. Bayesian parameter estimation for biochemical reaction networks using region-based adaptive parallel tempering. *Bioinformatics*, 34(13):i494–i501, July 2018. doi: 10.1093/bioinformatics/bty229.

Yung-Jue Bang, Eric Van Cutsem, Andrea Feyereislova, Hyun C Chung, Lin Shen, Akira Sawaki, Florian Lordick, Atsushi Ohtsu, Yasushi Omuro, Taroh Satoh, Giuseppe Aprile, Evgeny Kulikov, Julie Hill, Michaela Lehle, Josef Rüschoff, Yoon-Koo Kang, & ToGA Trial Investigators. Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): a phase 3,

open-label, randomised controlled trial. *Lancet*, 376(9742):687–97, Aug 2010. doi: 10.1016/S0140-6736(10)61121-X.

J. Barretina, G. Caponigro, N. Stransky, K. Venkatesan, A. A. Margolin, S. Kim, C. J. Wilson, J. Lehár, G. V. Kryukov, D. Sonkin, A. Reddy, M. Liu, L. Murray, M. F. Berger, J. E. Monahan, P. Morais, J. Meltzer, A. Korejwa, J. Jané-Valbuena, F. A. Mapa, J. Thibault, E. Bric-Furlong, P. Raman, A. Shipway, I. H. Engels, J. Cheng, G. K. Yu, J. Yu, P. Aspesi, Jr, M. de Silva, K. Jagtap, M. D. Jones, L. Wang, C. Hatton, E. Palescandolo, S. Gupta, S. Mahan, C. Sougnez, R. C. Onofrio, T. Liefeld, L. MacConaill, W. Winckler, M. Reich, N. Li, J. P. Mesirov, S. B. Gabriel, G. Getz, K. Ardlie, V. Chan, V. E. Myer, B. L. Weber, J. Porter, M. Warmuth, P. Finan, J. L. Harris, M. Meyerson, T. R. Golub, M. P. Morrissey, W. R. Sellers, R. Schlegel, & L. A. Garraway. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, 483(7391):603–607, Mar. 2012. doi: 10.1038/nature11003.

Sophie Bekisz & Liesbet Geris. Cancer modeling: From mechanistic to data-driven approaches, and from fundamental insights to clinical applications. *Journal of Computational Science*, 46:101198, October 2020. doi: 10.1016/j.jocs.2020.101198.

Richard E. Bellman. *Adaptive Control Processes*. Princeton University Press, December 1961. doi: 10.1515/9781400874668.

David M. Blei, Alp Kucukelbir, & Jon D. McAuliffe. Variational Inference: A Review for Statisticians. *Journal of the American Statistical Association*, 112(518): 859–877, April 2017. doi: 10.1080/01621459.2017.1285773.

Martin E Boehm, Lorenz Adlung, Marcel Schilling, Susanne Roth, Ursula Klingmueller, & Wolf D Lehmann. Identification of isoform-specific dynamics in phosphorylation-dependent STAT5 dimerization by quantitative mass spectrometry and mathematical modeling. *Journal of Proteome Research*, 13(12):5685–5694, 2014.

S. Boyd & L. Vandenberghe. *Convex Optimisation*. Cambridge University Press, UK, 2004.

Freddie Bray, Mathieu Laversanne, Elisabete Weiderpass, & Isabelle Soerjomataram. The ever-increasing importance of cancer as a leading cause of premature death worldwide. *Cancer*, 127(16):3029–3030, June 2021. doi: 10.1002/cncr.33587.

P.J. Brockwell & R.A. Davis. *Introduction to Time Series and Forecasting*. Springer Texts in Statistics. Springer International Publishing, 2016. ISBN 9783319298542.

S. Brooks, A. Gelman, G. Jones, & X.L. Meng. *Handbook of Markov Chain Monte Carlo*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. CRC Press, 2011. ISBN 9781420079425.

S. P. Brooks & G. O. Roberts. Assessing convergence of Markov chain Monte Carlo algorithms. *Stat. Comp.*, 8(4):319–335, 1998.

Stephen P. Brooks & Andrew Gelman. General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*, 7(4):434–455, 1998. doi: 10.1080/10618600.1998.10474787.

K. P. Burnham & D. R. Anderson. *Model selection and multimodel inference: A practical information-theoretic approach*. Springer, New York, NY, 2nd edition, 2002. ISBN 978-0-387-95364-9. doi: 10.1007/b97636.

Ben Calderhead. *Differential geometric MCMC methods and applications*. PhD thesis, University of Glasgow, 2011.

O.-T. Chis, J. R. Banga, & E. Balsa-Canto. Structural identifiability of systems biology models: A critical comparison of methods. *PLoS ONE*, 6(11):e27755, Nov. 2011. doi: 10.1371/journal.pone.0027755.

C. Cobelli & J. J. DiStefano. Parameter and structural identifiability concepts and ambiguities: a critical review and analysis. *Am. J. Physiol. Regul. Integr. Comp. Physiol.*, 239(1):7–24, Aug. 1980.

E. A. Coddington & N. Levinson. *Theory of ordinary differential equations*. McGraw-Hill, New York, 1955.

Joel E Cohen. Mathematics is biology's next microscope, only better biology is mathematics' next physics, only better. *PLoS Biology*, 2(12):e439, December 2004. doi: 10.1371/journal.pbio.0020439.

T. F. Coleman & Y. Li. An interior trust region approach for nonlinear minimization subject to bounds. *SIAM J. Optim.*, 6:418–445, 1996.

Lorenzo Contento, Noemi Castelletti, Elba Raimúndez, Ronan Le Gleut, Yannik Schälte, Paul Stapor, Ludwig Christian Hinske, Michael Hölscher, Andreas Wieser, Katja Radon, Christiane Fuchs, & Jan Hasenauer and. Integrative modelling of reported case numbers and seroprevalence reveals time-dependent test efficiency and infection rates. October 2021. doi: 10.1101/2021.10.01.21263052.

David Croft, Gavin O'Kelly, Guanming Wu, Robin Haw, Marc Gillespie, Lisa Matthews, Michael Caudy, Phani Garapati, Gopal Gopinath, Bijay Jassal, Steven Jupe, Irina Kalatskaya, Shahana Mahajan, Bruce May, Nelson Ndegwa, Esther Schmidt, Veronica Shamovsky, Christina Yung, Ewan Birney, Henning Hermjakob, Peter D'Eustachio, & Lincoln Stein.

Reactome: A database of reactions, pathways and biological processes. *Nucleic Acids Res.*, 39(Database issue):D691–7, Jan 2011. doi: 10.1093/nar/gkq1018.

Michael De La Maza & Deniz Yuret. Dynamic hill climbing. *AI expert*, 9:26–26, 1994.

Andrea Degasperi, Dirk Fey, & Boris N Kholodenko. Performance of objective functions and optimisation procedures for parameter estimation in system biology models. *npj Syst Biol Appl*, 3(1):20, 2017. doi: 10.1038/s41540-017-0023-2.

JT den Dunnen, R Dalgleish, DR Maglott, RK Hart, MS Greenblatt, J McGowan-Jordan, AF Roux, T Smith, SE Antonarakis, & PE Taschner. HGVS Recommendations for the Description of Sequence Variants: 2016 Update. *Hum Mutat*, 2016.

Karolin Ebert, Gwen Zwingenberger, Elena Barbaria, Simone Keller, Corinna Heck, Rouven Arnold, Vanessa Hollerieth, Julian Mattes, Robert Geffers, Elba Raimúndez, Jan Hasenauer, & Birgit Luber. Determining the effects of trastuzumab, cetuximab and afatinib by phosphoprotein, gene expression and phenotypic analysis in gastric cancer cell lines. *BMC Cancer*, 20(1), October 2020. doi: 10.1186/s12885-020-07540-7.

J. A. Egea, M. Rodriguez-Fernandez, J. R. Banga, & R. Marti. Scatter search for chemical and bio-process optimization. *J. Global Optim.*, 37(3):481–503, Oct. 2007. doi: 10.1007/s10898-006-9075-3.

M. Fedders. Posterior marginalisation for offset and noise parameters in observation functions of mechanistic models. Bachelor thesis, University of Bonn, 2021.

D. Fey, M. Halasz, D. Dreidax, S. P. Kennedy, J. F. Hastings, N. Rauch, A. Garcia Munoz, R. Pilkington, M. Fischer, F. Westermann, W. Kolch, B. N. Kholodenko, & D. R. Croucher. Signaling pathway models as biomarkers: Patient-specific simulations of JNK activity predict the survival of neuroblastoma patients. *Sci. Signal.*, 8(408), Dec. 2015. doi: 10.1126/scisignal.aab0990.

Ronald Aylmer Fisher. On the mathematical foundations of theoretical statistics. *Philos. Trans. R. Soc. London, Ser. A*, 222:309–368, 1922. doi: 10.1098/rsta.1922.0009.

Roger Fletcher & Michael JD Powell. A rapidly convergent descent method for minimization. *Comp J*, 6(2):163–168, 1963. doi: 10.1093/comjnl/6.2.163.

A Flobak, A Baudot, E Remy, L Thommesen, D Thieffry, M Kuiper, & A Laegreid. Discovery of drug synergies in gastric cancer cells predicted by logical modeling. *PLoS Comput Biol*, 2015.

Fabian Fröhlich, Thomas Kessler, Daniel Weindl, Alexey Shadrin, Leonard Schmiester, Hendrik Hache, Artur Muradyan, Moritz Schütte, Ji-Hyun Lim,

Matthias Heinig, Fabian J. Theis, Hans Lehrach, Christoph Wierling, Bodo Lange, & Jan Hasenauer. Efficient parameter estimation enables the prediction of drug response using a mechanistic pan-cancer pathway model. *Cell Syst.*, 7(6):567–579.e6, Dec. 2018. ISSN 2405-4712. doi: https://doi.org/10.1016/j.cels.2018.10.013.

Fabian Fröhlich, Carolin Loos, & Jan Hasenauer. Scalable inference of ordinary differential equation models of biochemical processes. In G. Sanguinetti & V. A. Huynh-Thu, editors, *Gene Regulatory Networks: Methods and Protocols*, volume 1883 of *Methods in Molecular Biology*, chapter 16, pages 385–422. Humana Press, 1 edition, 2019.

Kazuhiro A Fujita, Yu Toyoshima, Shinsuke Uda, Yu-ichi Ozaki, Hiroyuki Kubota, & Shinya Kuroda. Decoupling of receptor and downstream signals in the Akt pathway by its low-pass filter characteristics. *Sci Signal*, 3(132):ra56, Jul 2010. doi: 10.1126/scisignal.2000810.

A. Gelman, G. O. Roberts, & W. R. Gilks. Efficient metropolis jumping rules. In J. M. Bernardo, J. O. Berger, A. P. Dawid, & A. F. M. Smith, editors, *Bayesian Statistics*, pages 599–608. Oxford University Press, Oxford, 1996.

A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, & D.B. Rubin. *Bayesian Data Analysis, Third Edition.* Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis, 2013. ISBN 9781439840955.

Andrew Gelman & Donald B. Rubin. Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7(4):457 – 472, 1992. doi: 10.1214/ss/1177011136.

J. Geweke. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In J. M. Bernardo, A. F. M. Smith, A. P. Dawid, & J. O. Berger, editors, *Bayesian Statistics*, volume 4, pages 169–193. University Press Oxford, 1992.

Charles J. Geyer. Practical Markov Chain Monte Carlo. *Statistical Science*, 7(4):473 – 483, 1992. doi: 10.1214/ss/1177011137.

Giulia Giordano, Franco Blanchini, Raffaele Bruno, Patrizio Colaneri, Alessandro Di Filippo, Angela Di Matteo, & Marta Colaneri. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nat Med*, 26(6):855–860, June 2020. ISSN 1546-170X. doi: 10.1038/s41591-020-0883-7.

M. Girolami & B. Calderhead. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. R. Statist. Soc. B*, 73(2):123–214, Mar. 2011. doi: 10.1111/j.1467-9868.2010.00765.x.

Donald Goldfarb. A family of variable-metric methods derived by variational means. *Math Comp*, 24(109):23–26, 1970. ISSN 00255718, 10886842. doi: 10.1090/s0025-5718-1970-0258249-6.

Matthew M Graham & Amos J Storkey. Continuously tempered hamiltonian monte carlo. *arXiv preprint arXiv:1704.03338*, 2017.

R. N. Gutenkunst, J. J. Waterfall, F. P. Casey, K. S. Brown, C. R. Myers, & J. P. Sethna. Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput. Biol.*, 3(10):1871–1878, Oct. 2007.

H. Haario, E. Saksman, & J. Tamminen. An adaptive Metropolis algorithm. *Bernoulli*, 7(2):223–242, 2001.

H. Haario, M. Laine, A. Mira, & E. Saksman. DRAM: Efficient adaptive MCMC. *Stat. Comp.*, 16(4):339–354, 2006. doi: 10.1007/s11222-006-9438-0.

Ivonne Haffner, Katrin Schierle, Elba Raimúndez, Birgitta Geier, Dieter Maier, Jan Hasenauer, Birgit Luber, Axel Walch, Katharina Kolbe, Jorge Riera Knorrenschild, Albrecht Kretzschmar, Beate Rau, Ludwig Fischer von Weikersthal, Miriam Ahlborn, Gabriele Siegler, Stefan Fuxius, Thomas Decker, Christian Wittekind, & Florian Lordick. HER2 expression, test deviations, and their impact on survival in metastatic gastric cancer: Results from the prospective multicenter VARIANZ study. *Journal of Clinical Oncology*, 39(13): 1468–1478, May 2021. doi: 10.1200/jco.20.02761.

M Hara, H Nakanishi, K Tsujimura, M Matsui, Y Yatabe, T Manabe, & M Tatematsu. Interleukin-2 potentiation of cetuximab antitumor activity for epidermal growth factor receptor-overexpressing gastric cancer xenografts through antibody-dependent cellular cytotoxicity. *Cancer Science*, 2008.

Kazuto Harada, Yoshifumi Baba, Hironobu Shigaki, Takatsugu Ishimoto, Keisuke Miyake, Keisuke Kosumi, Ryuma Tokunaga, Daisuke Izumi, Mayuko Ohuchi, Kenichi Nakamura, Yuki Kiyozumi, Junji Kurashige, Masaaki Iwatsuki, Yuji Miyamoto, Yasuo Sakamoto, Naoya Yoshida, Masayuki Watanabe, & Hideo Baba. Prognostic and clinical impact of PIK3CA mutation in gastric cancer: pyrosequencing technology and literature review. *BMC Cancer*, 16:400, 07 2016. doi: 10.1186/s12885-016-2422-y.

H. Hass, C. Kreutz, J. Timmer, & D. Kaschek. Fast integration-based prediction bands for ordinary differential equation models. *Bioinformatics*, 32(8):1204–1210, 2016. doi: 0.1093/bioinformatics/btv743.

Helge Hass, Kristina Masson, Sibylle Wohlgemuth, Violette Paragas, John E. Allen, Mark Sevecka, Emily Pace, Jens Timmer, Joerg Stelling, Gavin MacBeath, Birgit Schoeberl, & Andreas Raue. Predicting ligand-dependent tumors from multi-dimensional signaling features. *npj Syst Biol Appl*, 3(1):27, 2017. doi: 10.1038/s41540-017-0030-3.

Helge Hass, Carolin Loos, Elba Raimúndez-Álvarez, Jens Timmer, Jan Hasenauer, & Clemens Kreutz. Benchmark problems for dynamic modeling of intracellular processes. *Bioinformatics*, 35(17):3073–3082, 01 2019. ISSN 1367-4803. doi: 10.1093/bioinformatics/btz020.

Trevor Hastie, Robert Tibshirani, & Jerome H. Friedman. *The Elements of Statistical Learning: data mining, inference, and prediction.* Springer Series in Statistics. Springer-Verlag, New York, 2 edition, 2005.

W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 51(1):97–109, April 1970.

S Heindl, E Eggenstein, S Keller, J Kneissl, G Keller, K Mutze, S Rauser, G Gasteiger, I Drexler, A Hapfelmeier, H Höfler, & B Luber. Relevance of MET activation and genetic alterations of KRAS and E-cadherin for cetuximab sensitivity of gastric cancer cell lines. *J Cancer Res Clin Oncol*, 2012.

Graeme Henkelman & Hannes Jónsson. Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. 113(22): 9978–9985, December 2000. doi: 10.1063/1.1323224.

L. A. Herzenberg, J. Tung, W. A. Moore, L. A. Herzenberg, & D. R. Parks. Interpreting flow cytometry data: A guide for the perplexed. *Nat. Immunol.*, 7(7): 681–685, July 2006. doi: 10.1038/ni0706-681.

Keegan Hines, Thomas Middendorf, & Richard Aldrich. Determination of parameter identifiability in nonlinear biophysical models: A Bayesian approach. *The Journal of general physiology*, 143, 02 2014. doi: 10.1085/jgp.201311116.

Matthew D Hoffman & Andrew Gelman. The No-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(1):1593–1623, 2014.

S. Hross & J. Hasenauer. Analysis of CFSE time-series data using division-, age- and label-structured population models. *Bioinformatics*, 32(15):2321–2329, Aug. 2016. doi: 10.1093/bioinformatics/btw131.

Gorjan Hrustanovic, Bianca J. Lee, & Trever G. Bivona. Mechanisms of resistance to EGFR targeted therapies. *Cancer Biol. Ther.*, 14(4):304–314, Apr. 2013.

M. Hucka, A. Finney, H. M. Sauro, H. Bolouri, J. C. Doyle, H. Kitano, A. P. Arkin, B. J. Bornstein, D. Bray, A. Cornish-Bowden, A. A. Cuellar, S. Dronov, E. D. Gilles, M. Ginkel, V. Gor, I. I. Goryanin, W. J. Hedley, T. C. Hodgman, J.-H. Hofmeyr, P. J. Hunter, N. S. Juty, J. L. Kasberger, A. Kremling, U. Kummer, N. Le Novère, L. M. Loew, D. Lucio, P. Mendes, E. Minch, E. D. Mjolsness, Y. Nakayama, M. R. Nelson, P. F. Nielsen, T. Sakurada,

J. C. Schaff, B. E. Shapiro, T. S. Shimizu, H. D. Spence, J. Stelling, K. Takahashi, M. Tomita, J. Wagner, & J. Wang. The systems biology markup language (SBML): A medium for representation and exchange of biochemical network models. *Bioinformatics*, 19(4):524–531, 2003. doi: 10.1093/bioinformatics/btg015.

S. Hug, A. Raue, J. Hasenauer, J. Bachmann, U. Klingmüller, J. Timmer, & F. J. Theis. High-dimensional Bayesian parameter estimation: Case study for a model of JAK2/STAT5 signaling. *Math. Biosci.*, 246(2):293–304, Nov. 2013. doi: 10.1016/j.mbs.2013.04.002.

S. Hug, M. Schwarzfischer, J. Hasenauer, C. Marr, & F. J. Theis. An adaptive scheduling scheme for calculating Bayes factors with thermodynamic integration using Simpson's rule. *Stat. Comput.*, 26(3):663–677, May 2016. doi: 10.1007/s11222-015-9550-0.

Trey Ideker, Timothy Galitski, & Leroy Hood. A new approach to decoding life: systems biology. *Annu. Rev. Genomics Hum. Genet.*, 2(1):343–372, 2001.

B. Ingalls. *Mathematical modelling in systems biology: An introduction*. MIT Press, 2013.

Yelena Y. Janjigian, Akihito Kawazoe, Patricio Yañez, Ning Li, Sara Lonardi, Oleksii Kolesnik, Olga Barajas, Yuxian Bai, Lin Shen, Yong Tang, Lucjan S. Wyrwicz, Jianming Xu, Kohei Shitara, Shukui Qin, Eric Van Cutsem, Josep Tabernero, Lie Li, Sukrut Shah, Pooja Bhagia, & Hyun Cheol Chung. The KEYNOTE-811 trial of dual PD-1 and HER2 blockade in HER2-positive gastric cancer. *Nature*, 600(7890):727–730, December 2021. doi: 10.1038/s41586-021-04161-3.

Claire Jean-Quartier, Fleur Jeanquartier, Igor Jurisica, & Andreas Holzinger. In silico cancer research towards 3r. *BMC Cancer*, 18(1), April 2018. doi: 10.1186/s12885-018-4302-0.

H. Jeffreys. *Theory of Probability*. Oxford University Press, Oxford, 3rd edition, 1961.

M. Kanehisa, S. Goto, M. Furumichi, M. Tanabe, & M. Hirakawa. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.*, 38(Database issue):D355–D360, Jan. 2010. doi: 10.1093/nar/gkp896.

S. Kang, A. G. Bader, & P. K. Vogt PK. Phosphatidylinositol 3-kinase mutations identified in human cancer are oncogenic. *Proc. Natl. Acad. Sci. USA*, 102(3):802–807, Jan. 2005.

Eva-Maria Kapfer, Paul Stapor, & Jan Hasenauer. Challenges in the calibration of large-scale ordinary differential equation models. *IFAC-PapersOnLine*, 52(26):58–64, Dec. 2019.

R. E. Kass & A. E. Raftery. Bayes factors. *J Am Stat Assoc*, 90(430):773–795, 1995. doi: 10.2307/2291091.

Simone Keller, Julia Kneissl, Verena Grabher-Meier, Stefan Heindl, Jan Hasenauer, Dieter Maier, Julian Mattes, Peter Winter, & Birgit 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, 12 2017. doi: 10.1186/s12885-017-3822-3.

William Ogilvy Kermack, A. G. McKendrick, & Gilbert Thomas Walker. A Contribution to the Mathematical Theory of Epidemics. *P Roy Soc A-Math Phy*, 115 (772):700–721, 1927. doi: 10.1098/rspa.1927.0118.

B. N. Kholodenko. Untangling the signalling wires. *Nat. Cell Biol.*, 9(3):247–249, Mar. 2007. doi: 10.1038/ncb0307-247.

GP Kim & A Grothey. Targeting colorectal cancer with human anti-EGFR monoclonocal antibodies: focus on panitumumab. *Biologics*, 2008.

J. Kim & B. Schoeberl. Beyond static biomarkers – The dynamic response potential of signaling networks as an alternate biomarker? *Sci. Signal.*, 8(408):fs21, Dec. 2015. doi: 10.1126/scisignal.aad4989.

Paul Kirk, Thomas Thorne, & Michael PH Stumpf. Model selection in systems and synthetic biology. *Current Opinion in Biotechnology*, 24(4):767–774, August 2013. doi: 10.1016/j.copbio.2013.03.012.

S. Kirkpatrick, C. D. Gelatt Jr, & M. P. M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, May 1983. doi: 10.1126/science.220.4598.671.

H. Kitano. Systems biology: A brief overview. *Science*, 295(5560):1662–1664, Mar. 2002.

P Klein, D Mattoon, MA Lemmon, & J Schlessinger. A structure-based model for ligand binding and dimerization of EGF receptors. *Proc Natl Acad Sci USA*, 2004.

E. Klipp, R. Herwig, A. Kowald, C. Wierling, & H. Lehrach. *Systems biology in practice*. Wiley-VCH, Weinheim, 2005. ISBN 978-3-527-31078-4.

J. Kneissl, S. Keller, T. Lorber, S. Heindl, G. Keller, I. Drexler, A. Hapfelmeier, H. Höfler, & B. Luber. Association of amphiregulin with the cetuximab sensitivity of gastric cancer cell lines. *Int. J. Oncol.*, 41(2):733–744, Aug 2012. doi: 10.3892/ijo.2012.1479.

C. Kreutz, M. M. Bartolome Rodriguez, T. Maiwald, M. Seidl, H. E. Blum, L. Mohr, & J. Timmer. An error model for protein quantification. *Bioinformatics*, 23 (20):2747–2753, Sept. 2007. doi: 10.1093/bioinformatics/btm397.

Clemens Kühn, Christoph Wierling, Alexander Kühn, Edda Klipp, Georgia Panopoulou, Hans Lehrach, & Albert J Poustka. Monte Carlo analysis of an ODE model of the sea urchin endomesoderm network. *BMC Syst Biol*, 3(1):1, 2009.

I Kuperstein, E Bonnet, H-A Nguyen, D Cohen, E Viara, L Grieco, S Fourquet, L Calzone, C Russo, M Kondratova, M Dutreix, E Barillot, & A Zinovyev. Atlas of cancer signalling network: A systems biology resource for integrative analysis of cancer data with Google maps. *Oncogenesis*, 4:e160, 07 2015.

Mateusz Krzysztof Łacki & Błażej Miasojedow. State-dependent swap strategies and automatic reduction of number of temperatures in adaptive parallel tempering algorithm. *Statistics and Computing*, 26(5):951–964, June 2015. doi: 10.1007/s11222-015-9579-0.

David Lake, Sonia A. L. Corrêa, & Jürgen Müller. Negative feedback regulation of the ERK1/2 MAPK pathway. 73(23):4397–4413, June 2016. doi: 10.1007/s00018-016-2297-8.

Nicolas Lartillot & Hervé Philippe. Computing Bayes factors using thermodynamic integration. *Systematic biology*, 55(2):195–207, 2006.

C. Leonhardt, G. Schwake, T. R Stögbauer, S. Rappl, J. T. Kuhr, T. S. Ligon, & J. O. Rädler. Single-cell mRNA transfection studies: Delivery, kinetics and statistics by numbers. *Nanomedicine: Nanotechnology, Biology, and Medicine*, 10(4):679–688, May 2014. doi: 10.1016/j.nano.2013.11.008.

David A. Levin, Yuval Peres, & Elizabeth L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, 2006.

Chen Li, Marco Donizelli, Nicolas Rodriguez, Harish Dharuri, Lukas Endler, Vijayalakshmi Chelliah, Lu Li, Enuo He, Arnaud Henry, Melanie I. Stefan, Jacky L. Snoep, Michael Hucka, Nicolas Le Novère, & Camille Laibe. BioModels database: An enhanced, curated and annotated resource for published quantitative kinetic models. *BMC Syst Biol*, 4:92, 2010.

Shiqing Li, Karl R Schmitz, Philip D Jeffrey, Jed J W Wiltzius, Paul Kussie, & Kathryn M Ferguson. Structural basis for inhibition of the epidermal growth factor receptor by cetuximab. *Cancer Cell*, 7(4):301–11, Apr 2005. doi: 10.1016/j.ccr.2005.03.003.

William Link & Mitchell Eaton. On thinning of chains in MCMC. *Methods in Ecology and Evolution*, 3:112 – 115, 06 2011. doi: 10.1111/j.2041-210X.2011.00131.x.

Alison C. Lloyd. The regulation of cell size. *Cell*, 154(6):1194–1205, 2013.

Carolin Loos, Sabrina Krause, & Jan Hasenauer. Hierarchical optimization for the efficient parametrization of ODE models. *Bioinf.*, 34(24):4266–4273, July 2018. doi: 10.1093/bioinformatics/bty514.

Florian Lordick, Yoon-Koo Kang, Hyun-Cheol Chung, Pamela Salman, Sang Cheul Oh, György Bodoky, Galina Kurteva, Constantin Volovat,

Vladimir M Moiseyenko, Vera Gorbunova, Joon Oh Park, Akira Sawaki, Ilhan Celik, Heiko Götte, Helena Melezínková, Markus Moehler, & Arbeitsgemeinschaft Internistische Onkologie and EXPAND Investigators. Capecitabine and cisplatin with or without cetuximab for patients with previously untreated advanced gastric cancer (EXPAND): a randomised, open-label phase 3 trial. *Lancet Oncol*, 14(6):490–9, May 2013. doi: 10.1016/S1470-2045(13)70102-5.

Florian Lordick, William Allum, Fátima Carneiro, Emmanuel Mitry, Josep Tabernero, Patrick Tan, Eric Van Cutsem, Cornelis van de Velde, & Andrés Cervantes. Unmet needs and challenges in gastric cancer: the way forward. *Cancer Treat Rev*, 40(6):692–700, Jul 2014. doi: 10.1016/j.ctrv.2014.03.002.

Jennifer L Macdonald-Obermann & Linda J Pike. Different epidermal growth factor (EGF) receptor ligands show distinct kinetics and biased or partial agonism for homodimer and heterodimer formation. *J Biol Chem*, 289(38):26178–26188, Sep 2014. doi: 10.1074/jbc.M114.586826.

C. Maier, C. Loos, & J. Hasenauer. Robust parameter estimation for dynamical systems from outlier-corrupted data. *Bioinformatics*, 33(5):718–725, Mar. 2017. doi: 10.1093/bioinformatics/btw703.

Leland McInnes, John Healy, Nathaniel Saul, & Lukas Grossberger. UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software*, 3 (29):861, 2018. doi: 10.21105/joss.00861.

W. Q. Meeker & L. A. Escobar. Teaching about approximate confidence regions based on maximum likelihood estimation. *Am Stat*, 49(1):48–53, 1995. doi: 10.1080/00031305.1995.10476112.

Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, & Edward Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.

Blazej Miasojedow, Eric Moulines, & Matti Vihola. An adaptive parallel tempering algorithm. *J. Comput. Graph. Stat.*, 22(3):649–664, 2013.

J. Mockus. *Bayesian Approach to Global Optimization: Theory and Applications*. Mathematics and its Applications. Springer Netherlands, 2011. ISBN 9789401068987.

Radford M Neal. Sampling from multimodal distributions using tempered transitions. *Statistics and computing*, 6(4):353–366, 1996.

John A Nelder & Roger Mead. A simplex method for function minimization. *Comput. J.*, 7(4):308–313, 1965. doi: 10.1093/comjnl/7.4.308.

Jorge Nocedal. Updating quasi-newton matrices with limited storage. *Mathematics of computation*, 35(151):773–782, 1980. doi: 10.2307/2006193.

Jorge Nocedal & Stephen Wright. *Numerical Optimization*. Springer Science & Business Media, 2006. doi: 10.1007/b98874.

K. Oda, Y. Matsuoka, A. Funahashi, & H. Kitano. A comprehensive pathway map of epidermal growth factor receptor signaling. *Mol. Syst. Biol.*, 1(2005.0010), May 2005. doi: 10.1038/msb4100014.

L. A. Ogilvie, C. Wierling, T. Kessler, H. Lehrach, & B. M. H. Lange. Predictive modeling of drug treatment in the area of personalized medicine. *Cancer Informatics*, 14(S4):95–103, 2015. doi: 10.4137/CIN.S19330.

Keisuke Okamura. Interdisciplinarity revisited: evidence for research impact and dynamism. *Palgrave Communications*, 5(1), November 2019. doi: 10.1057/s41599-019-0352-4.

D. B. Owen. A table of normal integrals. *Communications in Statistics - Simulation and Computation*, 9(4):389–419, January 1980. doi: 10.1080/03610918008812164.

Chi Hoon Park, Sung Yun Cho, Jae Du Ha, Heejung Jung, Hyung Rae Kim, Chong Ock Lee, In-Young Jang, Chong Hak Chae, Heung Kyoung Lee, & Sang Un Choi. Novel c-Met inhibitor suppresses the growth of c-Met-addicted gastric cancer cells. *BMC Cancer*, 2016.

David R Penas, Patricia González, José A Egea, Julio R Banga, & Ramón Doallo. Parallel metaheuristics in computational biology: An asynchronous cooperative enhanced scatter search method. *Procedia Comput. Sci.*, 51:630–639, 2015. doi: 10.1016/j.procs.2015.05.331.

Sara Pilotto, Anastasios Gkountakos, Luisa Carbognin, Aldo Scarpa, Giampaolo Tortora, & Emilio Bria. MET exon 14 juxtamembrane splicing mutations: clinical and therapeutical perspectives for cancer therapy. *Ann Transl Med*, 5(1):2, Jan 2017. doi: 10.21037/atm.2016.12.33.

George Poste. Bring on the biomarkers. *Nature*, 2011.

Elba Raimúndez, Simone Keller, Gwen Zwingenberger, Karolin Ebert, Sabine Hug, Fabian J. Theis, Dieter Maier, Birgit Luber, & Jan Hasenauer. Model-based analysis of response and resistance factors of cetuximab treatment in gastric cancer cell lines. *PLoS Comput. Biol.*, 16(3):e1007147, 2020.

Elba Raimúndez, Erika Dudkin, Jakob Vanhoefer, Emad Alamoudi, Simon Merkt, Lara Fuhrmann, Fan Bai, & Jan Hasenauer. COVID-19 outbreak in Wuhan demonstrates the limitations of publicly available case numbers for epidemiological modeling. *Epidemics*, 34:100439, 2021. ISSN 1755-4365. doi: https://doi.org/10.1016/j.epidem.2021.100439.

A. Raue, C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmüller, & J. Timmer. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(25):1923–1929, May 2009. doi: 10.1093/bioinformatics/btp358.

A. Raue, C. Kreutz, F. J. Theis, & J. Timmer. Joining forces of Bayesian and frequentist methodology: A study for inference in the presence of non-identifiability. *Philos T Roy Soc A*, 371(1984), 2013a. doi: 10.1098/rsta.2011.0544.

A. Raue, M. Schilling, J. Bachmann, A. Matteson, M. Schelke, D. Kaschek, S. Hug, C. Kreutz, B. D. Harms, F. J. Theis, U. Klingmüller, & J. Timmer. Lessons learned from quantitative dynamical modeling in systems biology. *PLoS ONE*, 8 (9):e74335, Sept. 2013b. doi: 10.1371/journal.pone.0074335.

A. Raue, B. Steiert, M. Schelker, C. Kreutz, T. Maiwald, H. Hass, J. Vanlier, C. Tönsing, L. Adlung, R. Engesser, W. Mader, T. Heinemann, J. Hasenauer, M. Schilling, T. Höfer, E. Klipp, F. J. Theis, U. Klingmüller, B. Schöberl, & J.Timmer. Data2Dynamics: a modeling environment tailored to parameter estimation in dynamical systems. *Bioinformatics*, 31(21):3558–3560, 2015. doi: 10.1093/bioinformatics/btv405.

J. Renart, J. Reiser, & G. R. Stark. Transfer of proteins from gels to diazobenzyloxymethyl-paper and detection with antisera: a method for studying antibody specificity and antigen structure. *Proc. Natl. Acad. Sci. USA*, 76(7):3116–3120, July 1979.

F. Rigat & A. Mira. Parallel hierarchical sampling: a general-purpose class of multiple-chains MCMC algorithms. *Comp. Stat. Data Anal.*, 56(6):1450–1467, June 2012. doi: 10.1016/j.csda.2011.11.020.

C. P. Robert & G. Casella. *Monte Carlo Statistical Methods*. Springer, 2004.

G. O. Roberts, A. Gelman, & W. R. Gilks. Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability*, 7(1):110–120, 1997.

K Roepstorff, MV Grandal, L Henriksen, SL Knudsen, M Lerdrup, L Grovdal, BM Willumsen, & B van Deurs. Differential effects of egfr ligands on endocytic sorting of the receptor. *Traffic*, 2009.

Thomas P. Runarsson & Xin Yao. Stochastic ranking for constrained evolutionary optimization. *IEEE T Evolut Comput*, 4(3):284–294, 2000. doi: 10.1109/4235.873238.

Mauricio D. Sacchi, Tadeusz J. Ulrych, & Colin J. Walker. Interpolation and extrapolation using a high-resolution discrete fourier transform. *IEEE Transaction on Signal Processing*, pages 31–38, 1998a.

M.D. Sacchi, T.J. Ulrych, & C.J. Walker. Interpolation and extrapolation using a high-resolution discrete fourier transform. *IEEE Transactions on Signal Processing*, 46 (1):31–38, 1998b. doi: 10.1109/78.651165.

Julio Saez-Rodriguez & Nils Blüthgen. Personalized signaling models for personalized treatments. *Molecular Systems Biology*, 16(1), January 2020. doi: 10.15252/msb.20199042.

Malcolm Sambridge. A parallel tempering algorithm for probabilistic sampling and multimodal optimization. *Geophys. J. Int.*, page ggt342, 2013.

M. J. Sanderson, I. Smith, I. Parker, & M. D. Bootman. Fluorescence Microscopy. 2014(10), October 2014. doi: 10.1101/pdb.top071795.

S. Sasagawa, Y.-i. Ozaki, K. Fujita, & S. Kuroda. Prediction and validation of the distinct dynamics of transient and sustained ERK activation. *Nat. Cell Biol.*, 7:365–373, Mar. 2005. doi: 10.1038/ncb1233.

Y. Schälte, P. Stapor, & J. Hasenauer. Evaluation of derivative-free optimizers for parameter estimation in systems biology. *IFAC-PapersOnLine*, 51(19):98–101, 2018. 7th Conference on Foundations of Systems Biology in Engineering FOSBE 2018.

Yannik Schälte, Fabian Fröhlich, Paul Stapor, Jakob Vanhoefer, Daniel Weindl, Paul Jonas Jost, Dantong Wang, Polina Lakrisenko, Elba Raimúndez, Dilan Pathirana, Leonard Schmiester, Philipp Städter, Lorenzo Contento, Simon Merkt, Erika Dudkin, Stephan Grein, & Jan Hasenauer. pyPESTO - Parameter EStimation TOolbox for Python, 2021. https://doi.org/10.5281/ZENODO.2553546.

Leonard Schmiester, Yannik Schälte, Fabian Fröhlich, Jan Hasenauer, & Daniel Weindl. Efficient parameterization of large-scale dynamic models based on relative measurements. *Bioinformatics*, 36(2):594–602, 07 2019. ISSN 1367-4803. doi: 10.1093/bioinformatics/btz581.

Leonard Schmiester, Yannik Schälte, Frank T. Bergmann, Tacio Camba, Erika Dudkin, Janine Egert, Fabian Fröhlich, Lara Fuhrmann, Adrian L. Hauber, Svenja Kemmer, Polina Lakrisenko, Carolin Loos, Simon Merkt, Wolfgang Müller, Dilan Pathirana, Elba Raimúndez, Lukas Refisch, Marcus Rosenblatt, Paul L. Stapor, Philipp Städter, Dantong Wang, Franz-Georg Wieland, Julio R. Banga, Jens Timmer, Alejandro F. Villaverde, Sven Sahle, Clemens Kreutz, Jan Hasenauer, & Daniel Weindl. PEtab—interoperable specification of parameter estimation problems in systems biology. *PLOS Computational Biology*, 17(1):1–10, 01 2021. doi: 10.1371/journal.pcbi.1008646.

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, & 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.

Gideon Schwarz. Estimating the dimension of a model. *Ann Stat*, 6(2):461–464, 1978. doi: 10.1214/aos/1176344136.

Bhuminder Singh, Graham Carpenter, & Robert J. Coffey. EGF receptor ligands: recent advances. 5:2270, September 2016. doi: 10.12688/f1000research.9025.1.

Kieran Smallbone & Pedro Mendes. Large-scale metabolic models: From reconstruction to differential equations. *Ind. Biotechnol.*, 9(4):179–184, 2013. doi: 10.1089/ind.2013.0003.

A. Sokal. *Monte Carlo Methods in Statistical Mechanics: Foundations and New Algorithms*, pages 131–192. Springer US, Boston, MA, 1997. ISBN 978-1-4899-0319-8. doi: 10.1007/978-1-4899-0319-8_6.

A Sorkin & JE Duex. Quantitative analysis of endocytosis and turnover of epidermal growth factor (EGF) and EGF receptor. *Curr Protoc Cell Biol*, 2010.

S. L. Spencer & P. K. Sorger. Measuring and modeling apoptosis in single cells. *Cell*, 144(6):926–939, Mar. 2011. doi: 10.1016/j.cell.2011.03.002.

Paul Stapor, Fabian Fröhlich, & Jan Hasenauer. Optimization and profile calculation of ODE models using second order adjoint sensitivity analysis. *Bioinformatics*, 34 (13):i151–i159, 2018.

Bernhard Steiert, Jens Timmer, & Clemens Kreutz. L1 regularization facilitates detection of cell type-specific parameters in dynamical systems. *Bioinformatics*, 32(17): i718–i726, 2016. doi: 10.1093/bioinformatics/btw461.

Hyuna Sung, Jacques Ferlay, Rebecca L. Siegel, Mathieu Laversanne, Isabelle Soerjomataram, Ahmedin Jemal, & Freddie Bray. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. 71(3):209–249, February 2021. doi: 10.3322/caac.21660.

I. Swameye, T. G. Müller, J. Timmer, O. Sandra, & U. Klingmüller. Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by databased modeling. *Proc. Natl. Acad. Sci. USA*, 100(3):1028–1033, Feb 2003.

A. Tarantola. *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM, 2005.

A. N. Taylor & T. D. Kitching. Analytic methods for cosmological likelihoods. *Monthly Notices of the Royal Astronomical Society*, 408(2):865–875, August 2010. doi: 10.1111/j.1365-2966.2010.17201.x.

Marco Viceconti, Adriano Henney, & Edwin Morley-Fletcher. In silico clinical trials: how computer simulation will transform the biomedical industry. *International Journal of Clinical Trials*, 3(2):37, May 2016. doi: 10.18203/2349-3259.ijct20161408.

Alejandro F Villaverde, Jose A Egea, & Julio R Banga. A cooperative strategy for parameter estimation in large scale systems biology models. *BMC systems biology*, 6 (1):75, 2012.

Alejandro F. Villaverde, Fabian Froehlich, Daniel Weindl, Jan Hasenauer, & Julio R Banga. Benchmarking optimization methods for parameter estimation in large kinetic models. *Bioinformatics*, page bty736, 2018.

Alejandro F. Villaverde, Elba Raimúndez-Álvarez, Jan Hasenauer, & Julio R. Banga. A comparison of methods for quantifying prediction uncertainty in systems biology. *IFAC-PapersOnLine*, 2019.

WD Vousden, Will M Farr, & Ilya Mandel. Dynamic temperature selection for parallel tempering in markov chain monte carlo simulations. *Mon. Not. R. Astron. Soc.*, 455(2):1919–1937, 2016.

A. Wächter & L. T. Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Math. Program.*, 106(1): 25–57, 2006. doi: 10.1007/s10107-004-0559-y.

P. Weber, J. Hasenauer, F. Allgöwer, & N. Radde. Parameter estimation and identifiability of biological networks using relative data. In S. Bittanti, A. Cenedese, & S. Zampieri, editors, *Proc. of the 18th IFAC World Congress*, volume 18, pages 11648–11653, Milano, Italy, Aug. 2011. doi: 10.3182/20110828-6-IT-1002.01007.

A. Wells. EGF receptor. *Int. J. Biochem. Cell Biol.*, 31(6):637–643, Jun. 1999.

Franz-Georg Wieland, Adrian L. Hauber, Marcus Rosenblatt, Christian Tönsing, & Jens Timmer. On structural and practical identifiability. *Current Opinion in Systems Biology*, 25:60–69, March 2021. doi: 10.1016/j.coisb.2021.03.005.

D. J. Wilkinson. Bayesian methods in bioinformatics and computational systems biology. *Briefings in Bioinf.*, 8(2):109–116, Mar 2007.

Kristy J Wilson, Jennifer L Gilmore, John Foley, Mark A Lemmon, & David J Riese, 2nd. Functional selectivity of EGF family peptide growth factors: implications for cancer. *Pharmacol Ther*, 122(1):1–8, Apr 2009. doi: 10.1016/j.pharmthera.2008.11.008.

Jian-Ming Xu, Yan Wang, You-Liang Wang, Yan Wang, Tao Liu, Ming Ni, Man-Sheng Li, Li Lin, Fei-Jiao Ge, Chun Gong, Jun-Yan Gu, Ru Jia, He-Fei

Wang, Yu-Ling Chen, Rong-Rui Liu, Chuan-Hua Zhao, Zhao-Li Tan, Yang Jin, Yun-Ping Zhu, Shuji Ogino, & Zhi-Rong Qian. PIK3CA Mutations Contribute to Acquired Cetuximab Resistance in Patients with Metastatic Colorectal Cancer. *Clin Cancer Res*, 23(16):4602–4616, Aug 2017. doi: 10.1158/1078-0432.CCR-16-2738.

W. Yang, J. Soares, P. Greninger, E. J. Edelman, H. Lightfoot, S. Forbes, N. Bindal, D. Beare, J. A. Smith, I. R. Thompson, S. Ramaswamy, P. A. Futreal, D. A. Haber, M. R. Stratton, C. Benes, U. McDermott, & M. J. Garnett. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucl. Acids Res.*, 41(Database issue):D955–D961, Jan. 2013. doi: 10.1093/nar/gks1111.

Lianhai Zhang, Jie Yang, Jie Cai, Xiaoming Song, Jianyun Deng, Xuesong Huang, Dawei Chen, Mengmeng Yang, Jean-Pierre Wery, Shuangxi Li, Aiwen Wu, Ziyu Li, Zhongwu Li, Yiqiang Liu, Yiyou Chen, Qixiang Li, & Jiafu Ji. A subset of gastric cancers with EGFR amplification and overexpression respond to cetuximab therapy. *Sci Rep*, 3:2992, Oct 2013. doi: 10.1038/srep02992.

Ben Zhao, Lu Wang, Hong Qiu, Mingsheng Zhang, Li Sun, Ping Peng, Qianqian Yu, & Xianglin Yuan. Mechanisms of resistance to anti-EGFR therapy in colorectal cancer. *Oncotarget*, 8(3):3980–4000, Jan 2017. doi: 10.18632/oncotarget.14012.

S Zhao & H Chen. Modeling the epidemic dynamics and control of COVID-19 outbreak in China. *Quant Biol*, 8:11–19, 2020. doi: 10.1007/s40484-020-0199-0.