HELMHOLTZ ASSOCIATION
Alliance on Systems Biology
HelmholtzZentrum münchen
German Research Center for Environmental Health

TECHNISCHE UNIVERSITÄT MÜNCHEN

Parameter estimation and uncertainty quantification for image based systems biology

Sabrina Hroß

March 2016

TECHNISCHE UNIVERSITÄT MÜNCHEN
Lehrstuhl M12 (Mathematische Modellierung biologischer Systeme)

Parameter estimation and uncertainty quantification for image based systems biology

Sabrina Hroß

Vollständiger Abdruck der von der Fakultät Mathematik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:

Univ.-Prof. Dr. O. Junge

Prüfer der Dissertation:

1. Univ.-Prof. Dr. Dr. F. J. Theis
2. Univ.-Prof. Dr. B. Kaltenbacher (Alpen Adria Universität Klagenfurt/Österreich)
3. Univ.-Prof. Dr. Chr. Kuttler

Die Dissertation wurde am 01.04.2016 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 23.09.2016 angenommen.

Für meinen Vater.

Acknowledgements

I'm not a person of big words so bear with me while I'm trying to thank all those people who helped to get this thesis going.

First of all I want to thank my supervisor Fabian Theis for the chance to write my thesis in his group (and institute), for the scientific freedom to chose my own topics to pursue and the many opportunities he gave me to travel and meet interesting scientists. And thank you for all the non-scientific mountain related discussions ... it is really sad that we never managed to go on a skitour together.

Second I want to thank my mentor Jan Hasenauer for his patience with failing project ideas, broken MATLAB code and crappy writing skills. When you came along I finally found the topic of my work and no one knows if I would have ever reached this point without your support and ideas!

Thank you also to Barbara Kaltenbacher and Christina Kuttler for the many helpful comments with regard to the theoretical parts of this thesis, for the good cooperation in our other projects and the discussions at the Helena thesis committee meetings. I also want to thank all my experimental cooperation partners Michael Sixt and Nilima Prakash who trusted me with the analysis of their data.

This thesis would not be the same without the best roomie ever: Sabine! Thank you for all the times we had to laugh so loud somebody came to our office door to check what happened. Thank you for all the Sabrine moments, the fun events, the coffee breaks and the distractions when writing this thesis made me unbearably grumpy. I really hope our time as roomies is only the beginning of an everlasting friendship!

During the last five years there were a lot of ups and downs but it was always fun because I had great colleagues to celebrate the ups with me and to help forget the downs. Thanks to you all at the ICB and the data driven computational modelling group. And if no-one was there to cheer me up there was chocolate. Thank you Snack Bär :-)
Last I want to thank my family, without their constant encouragement I would have never managed to write this thesis. Thank you Svenja for being a great little sister. Thank you mom for always supporting me to the fullest no matter how hard your own life was. And thank you dad for always believing in me I am unbelievably sad that you will never read this lines.

Thank you Sven and Elisabeth for being my sunshine whenever there is rain!

Contents

Index of notation

Acronyms

Acronym	Description
AIC	Akaike information criterion
BIC	Bayesian information criterion
CCL21	chemokine (C-C motif) ligand 21
DC	dendritic cell
FRAP	fluorescent recovery after photobleaching
KKT	Karush-Kuhn-Tucker optimality conditions
LV	lymphoid vessel
MLE	maximum likelihood estimation
ODE	ordinary differential equation
PDE	partial differential equation
Pom1p	(phosphorylated) polarity protein kinase

Notation

General notation: Throughout this thesis we follow the common notation and denote the partial derivative with respect to for example x of a variable u by u_x.

Mathematics and statistics

Symbol	Description
$u^{(d)}$	d-th component of a vector
$\delta_{i,j}$	Dirac delta function
$\exp(x)$	exponential function
$\log(x)$	natural logarithm
$\mathcal{P}_{e_{\theta_i}}$	orthogonal projection in direction of θ_i
$\int_\Omega \dots dx$	integral over Ω
$\int_{\partial\Omega} \dots ds$	line integral over boundary of Ω
$\mathcal{N}(\mu, \sigma^2)$	normal distribution
$\mathcal{LN}(x\|\mu, \sigma^2)$	log-normal distribution
$\chi^2(\alpha, n_\theta)$	Chi-squared distribution
$\mathbb{N}$	natural numbers
$\mathbb{R}$	real numbers
$\mathbb{R}_+$	non-negative real numbers
$\mathbb{Z}$	integers
$H^1(\Omega)$	Sobolev space $W^{1,2}(\Omega)$
$(H^1(\Omega))^*$	Dual space of $H^1(\Omega)$
$\partial u/\partial \nu_C$	Konormal derivative with respect to the differential operator C

List of Figures

Abstract

In image based systems biology imaging techniques are combined with quantitative models. These are able to capture biological processes presenting multiple scales of complexity ranging from single cells to entire populations. Of special interest is the estimation of the underlying process parameters from the given imaging data and the quantification of the uncertainty in the obtained parameter values. Due to the spatial and temporal dimension of images or videos the models considered in this context are often semi-linear partial differential equations (PDE). While parameter estimation and uncertainty quantification for inverse problems with PDEs has been extensively studied in the last decades, e.g. in the context of physics or engineering, it fails to meet the requirements for image based systems biology. Especially partial observations, high noise levels and parameter indeterminacy are not yet discussed. We have developed a general set up for parameter estimation, uncertainty quantification and model selection in this context as well as a new efficient method to facilitate uncertainty analysis for inverse problems with semi-linear PDEs based on profile likelihoods.

We introduce a non-standard, statistically motivated objective function for the parameter estimation problem with semi-linear PDE constraints and partial observations from images. In contrast to the standard least-squares objective function the introduced likelihood function naturally incorporates both a statistical treatment of the image preprocessing as well as the parameter estimation. Subsequently, we propose a new approach to calculate profile likelihoods, which circumvents the time consuming repeated solution of the inverse problem and trivially allows for an adaptive profile calculation. We consider two applications: gradient formation during cell size control of fission yeast *Schizosaccharomyces pombe* and gradient formation for dendritic cell guidance in mammalians. For both applications we develop the complete scope of model based hypothesis testing for image data including the formulation of models based on competing biological hypothesis, tailored parameter estimation and uncertainty quantification, repeated model adjustments and finally model selection.

For both applications our analysis yielded new insights into open biological questions and motivated new experiments to validate the computational results. From a methodological point of view, however, the key result is the speed up of the profile likelihood calculation achieved with the introduced method. Without this gain on computation speed the uncertainty analysis for both applications would have been very time consuming or infeasible. Furthermore it is robust in the presence of partial observations, non-identifiabilties and high noise levels. This enables the application of the approach to the computationally demanding parameter estimation problems arising in spatio-temporal biological processes, e.g. blood-flow or tumour growth models.

Deutsche Zusammenfassung

Die bildbasierte System Biologie kombiniert bildgebende Verfahren mit quantitativen Modellen. Dies ermöglicht die Untersuchung biologischer Prozesse von der einzelnen Zelle bis hin zu ganzen Populationen. Von besonderem Interesse ist in diesem Fall die Schätzung der Parameter des zugrundeliegenden Prozesses auf Basis der Bilddaten und die Quantifizierung der Unsicherheiten in den ermittelten Parameterwerten. Auf Grund der räumlichen und zeitlichen Auflösung von Bildern oder Filmen sind die in diesem Zusammenhang betrachteten Modelle häufig semi-lineare partielle Differentialgleichungen (PDE). Obwohl in den letzten Jahrzehnten Parameterschätzung und Unsicherheitsanalyse für inverse Probleme mit PDEs ausführlich untersucht wurden, z.B. im Zusammenhang mit physikalischen oder technischen Fragestellungen, so sind sie doch nicht in der Lage die Voraussetzungen, die die bildbasierte System Biologie fordert, zu erfüllen. Insbesondere unvollständige Beobachtungen, hohe Messfehler und Parameterunbestimmtheit sind bisher nicht untersucht worden. Wir entwickelten ein allgemeines Set-up für die Parameterschätzung, Unsicherheitsanalyse und Modellselektion für diese Anforderungen, sowie eine neue effiziente Methode, welche die Unsicherheitsanalyse mit Hilfe von Likelihoodprofilen für inverse Probleme mit semi-linearen PDEs ermöglicht.

Wir führen eine spezielle, statistisch motivierte Kostenfunktion für die Parameterschätzung mit semi-linearen PDE Nebenbedingungen und teilweisen Beobachtungen ein. Im Gegensatz zur Standardmethode der kleinsten Fehlerquadrate beinhaltet die neu eingeführte Likelihoodfunktion auf natürliche Weise eine statistische Behandlung der Bildbearbeitung sowie der Parameterschätzung. In Folge dessen schlagen wir eine neue Vorgehensweise zur Berechnung von Likelihoodprofilen vor, welche die wiederholte Lösung des inversen Problems umgeht und eine adaptive Profilberechnung ermöglicht. Wir betrachten zwei Anwendungsbeispiele: Konzentrationsverläufe in der Zellgrößenkontrolle der Spalthefe *Schizosaccharomyces pombe* und Konzentrationsverläufe für die Zellwanderung dendritischer Zellen in Säugetieren. Für beide Anwendungen durchlaufen wir den gesamten Umfang des modellbasierten Hypothesentestes für Bilddaten, d.h. die mathematische Modellierung basierend auf den konkurrierenden biologischen Hypothesen, Parameterschätzung und Unsicherheitsanalyse sowie Modellanpassung und abschließende Modellselektion.

Unsere Analyse ergab für beide Anwendungen neue Erkenntnisse zu offenen biologischen Fragen und motivierte neue Experimente um die theoretischen Resultate zu validieren. Vom methodischen Standpunkt aus ist das Hauptresultat die Beschleunigung der Likelihoodprofilberechnung durch eine neu entwickelte Methode. Ohne diesen Zugewinn an Rechengeschwindigkeit wäre die Unsicherheitsanalyse für beide Anwendungen sehr zeitaufwändig oder gar unmöglich gewesen. Des weiteren, ist die Berechnung auch im Falle von unvollständiger Beobachtungen zuverlässig, Nicht-

Identifizierbarkeiten und hohen Messfehlern. Dies ermöglicht die Anwendung der Methode auf viele rechenaufwendige Parameterschätzungsprobleme für biologische Prozess, z.B. Blutfluß- oder Tumorwachstumsmodelle.

1. Introduction

1.1. Research motivation

Imaging is excellent to capture biological systems, which present multiple scales of complexity ranging from single molecules to entire cell populations. Image based systems biology combines imaging with quantitative models to gain new insights into the big picture of the underlying biological processes beyond molecular analysis (Sbalzarini, 2013; Verveer & Bastiaens, 2008; Megason & Fraser, 2007). The three essential steps of a systematic model based analysis of biological image data are acquisition and preprocessing of image data, deduction of mechanistic models and the computational simulation and calibration of the models to a given data set, including the parameter optimization and uncertainty quantification for the kinetic parameters of the model (Medyukhina *et al.*, 2015). To balance model detail and overfitting, models are developed in an iterative circle of model deduction, model calibration and model refinement and maybe the realization of additional validation experiments. This iterative process is referred to as the "Systems Biology Loop" (Kitano, 2002).

While a quantitative analysis of imaging data, e.g. cell sizes or shapes, is already widely used in biological research the key difference in image based systems biology is the model based analysis of the images. The quantitative information serves as the basis for the mathematical models and thus can yield insights and predictions on parameters and components, which are not directly measurable (see Medyukhina *et al.* (2015) for a detailed review). The key to successful predictions, however, is the estimation of the kinetic parameters of the mathematical model from the image data, which is composed of optimization and uncertainty quantification. Parameter optimization solves the inverse problem of finding the parameters which describe the given experimental data best by maximization of an objective function, which gives a measure for the discrepancy between model and data, e.g. the likelihood function. Uncertainty quantification for inverse problems is concerned with two key problems: the uncertainty in the obtained parameter values and the propagation of parameter uncertainty through the model. To address the first problem, information about the distribution of the obtained parameters around the optimal point is needed, as a wide distribution represents a high uncertainty in the optimal parameter values and and a narrow distribution indicates a well defined set of parameters. Common methods to obtain the parameter distribution are asymptotic approximations, profile likelihoods, bootstrapping or sampling methods. In the biological context it is known that the asymptotic and profiling methods perform as good as sampling with less computational effort this holds in the context of ordinary differential equations (ODE) (Hug & Raue, 2013) and for pressure wave propagation models, which are PDEs (Kenz *et al.*, 2013). Besides the general challenges for parameter estimation

from biological data optimization and uncertainty quantification both face problems specific to imaging data.

A major challenge for parameter optimization from images is the presence of other biological processes, which can not be disabled experimentally but are of no interest for the biological question at hand. These processes are called structured noise in the following as they are treated as noise during the estimation process but they posses structure in contrast to the unstructured measurement noise due to the underlying biological process, which generates them. Such processes have to be considered either in the model or during the parameter optimization process as they can bias or completely change the estimated kinetic parameters (Medyukhina *et al.*, 2015). The inclusion of structured noise as biological processes in the model leads to overly complicated models, which are not tailored to the hypothesis that should be addressed. In contrast preprocessing of the data in terms of filtering or extraction of features to remove the structured noise yields a reduction in number or even the dimension of the measurement data, which can lead to unintuitive uncertainties in the parameter estimates or a big amount of unused data (Sbalzarini, 2013).

Besides problems with the interpretation of parameter uncertainties also the quantification of parameter uncertainties itself poses a major challenge as the models considered in image based systems biology are prominently reaction-diffusion equations, i.e. semi-linear parabolic PDEs, and the state of the art methods for parameter optimization include regularization techniques (Hinze *et al.*, 2009). Regularization techniques are employed to resolve the ill-posedness of the inverse problems, due to non-identifiable parameters, partial observations and high noise levels, which occur in the biological context. For the regularized inverse problem existence and uniqueness of optimizers can be shown and efficient numerical schemes to solve the inverse problem exist (Engl *et al.*, 2009). However, the regularization enforces an artificial identifiability of the parameters, which is often a bad assumption in the biological context (Gutenkunst *et al.*, 2007).

1.2. Research topic

After the general introduction of parameter optimization and uncertainty quantification in the context of image based systems biology this section reviews the topic of imaging in biology and spatio-temporal models in more detail to motivate the research questions posed in this thesis.

1.2.1. Imaging in biology

A central aspect of biological research is the observation of biological systems ranging for example from the observation of mammalian populations to single cells of the same species. While the first is easy to study with the unaided eye the second can not be observed without a microscope. Since the microscope has been developed in the 1590s, it has opened the way to visually study small entities like bacteria or blood cells. And it has become one of the most essential tools for cellular biology. In the last decades, the development of high resolution microscopy in

combination with biochemical staining methods has transformed microscopy from a qualitative to a highly quantitative method (Antony *et al.*, 2013; Megason & Fraser, 2007; Yuste, 2005). At the heart of this development is the confocal fluorescence microscopy. It is able to detect small concentrations of single molecules with a good signal-to-background ratio in fixed or alive specimen (Yuste, 2005). In this work we concentrate on two confocal imaging techniques: live single cell imaging and stained tissue imaging.

Live single cell imaging. The key development, which facilitated the successful imaging of living cells in systems biology was the discovery of the green fluorescent protein (GFP). The GFP was discovered during the studies of bio luminescence in the jellyfish *Aequorea victoria* (Chalfie *et al.*, 1994). It can be introduced in all species which can be genetical engineered, e.g. yeast, C.elegans or mice. Once expressed it can be stimulated by UV or blue light. Nowadays there is a wide range of variants of GFP and newly developed fluorophores that allow to work with a rainbow of colors from green to red (Wiedenmann *et al.*, 2009; Kremers *et al.*, 2011). Together with confocal or two photon laser microscopy fluorescent, labelling allows the study of concentrations of proteins, mRNA or small molecules in single cells (Sung & McNally, 2011).

A very useful technique in combination with fluorescent labels and confocal microscopy is fluorescence recovery after photobleaching (FRAP). It can be used to measure diffusion rates of labelled proteins in living cells on time scales from seconds to minutes. For FRAP a region of interest in the cell is bleached by a high intensity light pulse. The region is then imaged over time and it becomes visible that the region, which lacked fluorescence in the beginning, starts to become brighter as fluorescent molecules diffuse from outside into the region. Thus the rate of brightness recovery after the bleaching is a direct measure of how fast molecules diffuse in this region of the cell (Reits & Neefjes, 2001; Sprague & McNally, 2005). For complex diffusion processes like the Pom1p gradient formation discussed in Chapter 5, which includes immobilization, degradation or clustering the rate of recovery is no-longer a direct measure for the diffusion rate.

How can diffusion parameters be determined from FRAP measurements in the presence of immobilization, degradation or clustering? (see Problem 5.1)

Immunostainings. The second imaging technique is also based on fluorescent proteins but no genetical engineering is needed. Immunostainings were introduced in the 1940s and this method is based on the combination of fluorescent molecules with antibodies applicable to fixated tissue slices (Coons *et al.*, 1941). While today there is a wide range of procedures how to stain tissue, the general idea remains the same: Tissue slices are fixated and the cell membranes are permeabilized in such a way that the antibodies labelled with a fluorescent molecule which are added in the next step can diffuse into the tissue and bind the proteins of interest. Finally the fluorescence is stimulated and imaged with fluorescence microscopy methods. The obtained images are analyzed with basic image processing techniques to obtain biological insight and model based analysis is rare. This is due to the fact that immunostainings are

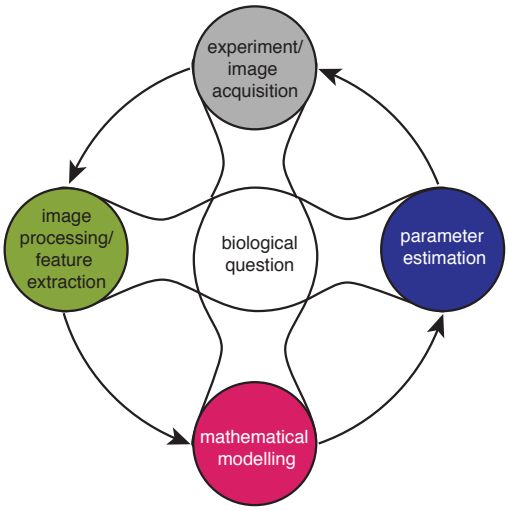


Figure 1.1: ***Image based systems biology loop.*** *The schematic shows the image based systems biology loop consisting of four steps: the experiment with the image acquisition, the image preprocessing combined with an extraction of features, the model building and the parameter estimation. Each step feed into the next one and the parameter estimation feed back to experiment via prediction and experimental design based on the calibrated model.*

often of poor quality and give only semi-quantitative information or the structured noise originating from biological process which can not be disabled experimentally poses major problems for parameter estimation. For complex biological hypothesis testing as discussed for the CCL21 gradient formation in Chapter 6 basic image processing techniques are insufficient and a model based analysis is needed.

How can parameter optimization and uncertainty quantification be performed on tissue scale immunostainings, which are corrupted by structured noise, to facilitated biological hypothesis testing? (see Problem 6.1 and 6.2)

1.2.2. Spatio-temporal models in systems biology

The first ground breaking spatio-temporal models in biology were the pattern formation models by Turing (1952) and Gierer & Meinhardt (1972). The models implicitly use an image based model calibration as their behaviour and hence the model parameters are chosen such that the model output is comparable to patterns observed in nature, e.g. sea shell patterns. These models are based on the combination of basic chemical reactions with diffusion and more than half a century later the basic idea remains the same but models consider for example complex regions or growing domains (Menshykau & Iber, 2013; Uzkudun *et al.*, 2015). First introduced in the context of morphogenesis, spatio-temporal models are present in nearly all biological fields ranging from physical models of cardiovascular flow over tumor growth to the simulation of single cells (Klann & Koeppl, 2012). The image based modeling process consist of four steps: the experiment and image acquisition, the image preprocesing and feature extraction, the model building and the parameter estimation (Medyukhina *et al.*, 2015; Sbalzarini, 2013) (see Figure 1.1). We concentrated on the questions and challenges arising from the fourth point. We considered the parameter optimization and the uncertainty analysis for spatio-temporal models in systems biology.

Parameter optimization. As mentioned before there is a considerable number of spatio-temporal deterministic models in the context of computational biology already available and the development progresses. Those models become especially

useful for biological applications if they can give insights beyond directly observable aspects of a biological process or make predictions for future experiments. For such questions models have to give qualitative or quantitative information about the process, therefore the knowledge of the parameter values used is crucial. In many cases these parameters can not be measured directly but have to be determined from measurement data by parameter estimation procedures (Lillacci & Khammash, 2010; Engl *et al.*, 2009). In computational biology two parameter estimation paradigms are prevalent: Frequentist and Bayesian approaches (Villaverde & Banga, 2014). The key difference of both approaches lies in the interpretation of a parameter (Stark, 2012). For a Bayesian a fixed parameter value does not exist, but all information about the parameters is given as a distribution and all probabilities quantify a degree of belief including prior knowledge of the statistician. For a frequentist a fixed, unknown parameter value exists and its probability is the frequency of its occurrence in infinitely many samples. This difference leads to the major difference in the parameter estimation as Bayesian methods often approximate the parameter distribution by sampling methods, e.g. MCMC, while frequentist methods consider one point estimated obtained by optimization methods, e.g. maximum likelihood estimation. There have been promising developments of Bayesian approaches to facilitate parameter estimation for spatio-temporal models in the last years (Bui-Thanh *et al.*, 2011; Sternfels & Earls, 2013) but we will focus on the Frequentist approach. The theory and algorithms recently developed in the field of PDE constrained optimization (Hinze *et al.*, 2009; Tröltzsch, 2009; Garvie & Trenchea, 2014) can be applied/adapted to many image based parameter estimation problems with the Frequentist approach, which results in fast, efficient and reliable optimization results (see Chapter 2). Those methods, however, have been developed for inverse problems in the context of physics or engineering applications and rely on regularization techniques to resolve the ill-posedness of the inverse problem due to measurement noise or parameter non-identifiabilities. Furthermore, only for the regularized inverse problem existence and uniqueness of optimizers can be proven mathematically (Engl *et al.*, 2009). Unfortunately the assumptions made for the regularization are unsuitable for the high amount of measurement noise and the indeterminacy of parameters in the context of biological applications (Gutenkunst *et al.*, 2007).)

How to define parameter optimization for semi-linear models from imaging data? (see Problem 3.1)

Uncertainty analysis. A great challenge in biological applications is the difficulty or impossibility to determine unique parameter values from measurement data (Jacquez & Perry, 1990; Gutenkunst *et al.*, 2007; Fernández Slezak *et al.*, 2010; Erguler & Stumpf, 2011). Furthermore, based on those parameters sensible decisions like drug dosages or medical treatments are made. These demand for a precise quantification of the uncertainty in the obtained parameters and the resulting predictions of the model. Regarding the lack of identifiability we can distinguish between two types: structural and practical identifiability (see Chapter 2 for a detailed comparison). Structural identifiability is a property of the dynamic model and the observations, which is independent of the measurements. It has to be verified by analytical methods. Hence, assessing structural identifiability for non-linear dynamical

systems is very hard and the methods are computationally expensive (Chis *et al.*, 2011). In contrast to the theoretical property of structural identifiability, practical identifiability takes into account the finite number of measurements, the noise level and the model errors. Practical identifiability analysis is an a posteriori analysis, which addresses the question if the parameters can uniquely be estimated with the the available experimental data (Raue *et al.*, 2009). This closely relates to the question of parameter uncertainty and both can be addressed in terms of confidence intervals. For spatio-temporal models the sensitivities or the Hessian matrix are used to study identifiability and uncertainty in the parameters (Ashyraliyev *et al.*, 2008; Banks *et al.*, 2010). For ordinary differential equations, however, it has been shown that those approaches fail to describe the uncertainty and obtain valid confidence intervals. For ordinary differential equations the profile likelihood was proven to be a good choice for uncertainty quantification and practical identifiability analysis (Murphy *et al.*, 2000; Raue *et al.*, 2009; Kreutz *et al.*, 2013). For discretized PDEs those methods can be applied, however, they can become infeasible for computationally demanding models as expected in most PDE applications.

How to can parameter uncertainties efficiently quantified for semi-linear models in image based systems biology? (see Problem 3.2 as well as Problem 4.1 and 4.2)

1.3. Contribution of this thesis

To address the aforementioned open problems, I developed:

- parameter optimization methods with semi-linear PDE constraints, partial observations and problem tailored likelihood functions to enable the parameter optimization for image based systems biology.
- an efficient profile likelihood calculation for optimization problems with semi-linear PDE constraints to address the problem of infeasible profile calculation for computationally demanding PDE models.
- tailored parameter estimation and model selection methods for single cell images and FRAP measurements to determine kinetic parameters for signaling gradients in fission yeast.
- tailored parameter estimation and model selection methods for immunostainings to determine the kinetic parameters for signaling gradients; a special focus is on the treatment of structured noise.

Apart from the methodological contributions the following biological insights were gained:

- there is strong evidence in the CCL21 immunostaining image data for a lymphoide vessel specificity of the heparan sulfate concentration and therefore the CCL21 gradient stabilization mechanism, this suggest a distinct functionality of different lymphoid vessel branches for dendritic cell guidance.

- in the existing single cell data for Pom1p there is empirically no evidence for clustering over auto phosphorylation or trans phosphorylation for gradient stabilization against input fluctuations.

This is no cumulative thesis but some of these contributions were in part already published in peer-reviewed journals or are currently submitted to peer-reviewed journals. Some parts and results will thus correspond or are identical with the following publications:

- **S. Hock**[1], J. Hasenauer, and F.J. Theis. Modeling of 2D diffusion processes based on microscopy data: parameter estimation and practical identifiability analysis. BMC Bioinformatics, 14 Suppl 1:S7, January 2013.

- R. Boiger, J. Hasenauer, **S. Hross**, and B. Kaltenbacher. Integration based profile likelihood calculation for PDE constrained parameter estimation problems. Inverse Problems *(accepted)*

- **S. Hross**[2], A. Fiedler[2] F.J. Theis and J. Hasenauer. Quantitative comparison of competing PDE models for Pom1p dynamics in fission yeast. FOSBE, 2016.

In addition to the publications, which directly contributed to this thesis, the author also participated in other research projects in the context of image based systems biology and spatio-temporal modelling. The results of those papers are related to the topic of this work but focus on complementary aspects. The following publications have been published:

- **S. Hock**[1],[2], Y.K. Ng[2], J. Hasenauer, D. Wittmann, D. Lutter, D. Trümbach, W. Wurst, N. Prakash, and F.J. Theis. Sharpening of expression domains induced by transcription and microRNA regulation within a spatio-temporal model of mid-hindbrain boundary formation. BMC Systems Biology, 7:48, 2013.

- D.M. Wittmann, **S. Hock**[1] and F.J. Theis, Truth-content and phase transitions of random boolean networks with generic logics. SIAM Journal on Applied Dynamical Systems, 12, 315351, 2013.

- F. Fröhlich, **S. Hross**, F.J. Theis and J. Hasenauer. Radial basis function approximations of Bayesian parameter posterior densities for uncertainty analysis. Lecture Notes Computer Science. 8859, 73-85, 2014.

- J. Hasenauer, N. Jagiella, **S. Hross** and F.J. Theis. Data-driven modelling of biological multi-scale processes. Journal Coupled Systems and Multiscale Dynamics, 2015.

- **S. Hross** and J. Hasenauer. Analysis of CFSE time-series data using division-, age- and label-structure population models. Bioinformatics, 2321-9, 2016.

[1]This paper was published under the authors maiden name
[2]These authors contributed equally to this paper

1.4. Outline of this thesis

This thesis starts by a short introduction to semi-linear PDEs in **Chapter** 2. The general form of the models, as well as weak solutions and existence results are summarized. In addition, this chapter contains a general overview of parameter optimization, uncertainty quantification methods and a summary of PDE constrained optimization results needed in this work.

In **Chapter** 3, the PDE constrained optimization problem for parameter optimization for semi-linear PDEs with partial observations from image data is introduced. We consider a special pixel based observation operator as well as the likelihood functions for the integrated noise model approach. Following the definitions, we derive sufficient and necessary optimality conditions for the problem and compare them to the Hessian of the optimization problem without PDE constraints. In the remainder of the chapter optimization based profile likelihood calculation for semi-linear PDEs is introduced to facilitate uncertainty quantification.

In **Chapter** 4 we address the inefficiency of the optimization based profile likelihood calculation with a simulation based profile likelihood calculation. We extend the existing methodology to problems with semi-linear PDE constraints and show the equivalence of full and reduced formulation. For both approaches the Hessian matrix of the objective function is needed, which is often unknown. Hence, we proceed by considering different approximation schemes for the Hessian matrix of the optimization problem. The approximations are subsequently analysed with respect to an application example.

In **Chapter** 5, we apply the introduced methods to study the Pom1 gradient formation, which is the key regulator of cell size control in fission yeast *Schizosaccharomyces pombe*. We discuss models for the competing hypothesis regarding gradient stabilization against noise, namely cluster formation, auto phosphorylation and trans phosphorylation of Pom1. The considered measurements are single cell live images and FRAP experiments and we perform a parameter optimization and uncertainty analysis, which in turn enables the model selection.

In **Chapter** 6, CCL21 gradient formation for dendritic cell guidance, the main application of this thesis, is discussed. We introduce models for three existing biological hypothesis regarding the formation of the CCL21 gradient. We perform a tailored parameter estimation utilizing the previously developed methods to compare estimation based on features to estimation on full images. Furthermore, we introduce and apply an integrated noise model and compare it to the existing filtering methods on simulated and real data. The chapter is concluded by a model selection, which suggests that lymphoid vessel branches have different functionality with regard to the dendritic cell guidance.

Chapter 7 concludes this work by giving a summary of the key results and a discussion in the context of existing methods. It is followed by an outline of potential extensions, applications and open problems.

2. Background

In this chapter, we state the model problems to be considered and discuss some of their theoretical aspects as well as the resulting parameter estimation problems and the basic theory of uncertainty quantification.

In Section 2.1 we state the general formulation of the models and give the application example considered in this work. Subsequently, we discuss parameter optimization and uncertainty quantification based on the reduced form of the optimization problem. We proceed by giving a short introduction to the parameter estimation problem in terms of a PDE constrained optimization and briefly discuss some theoretical aspects.

2.1. Semi-linear PDE models

We consider models, which combine diffusion processes with biochemical reaction kinetics. The reactions are formulated in terms of mass-action kinetics (MA)(Horn & Jackson, 1972) and give rise to semi-linear PDEs. In the following we distinguish between stationary models, i.e. the derivative with respect to time is zero, and dynamical models, which respectively results in either semi-linear elliptic or parabolic PDEs.

Throughout this work, Ω denotes a bounded domain in $\mathbb{R}^n$, with a $C^{0,1}$- boundary $\partial\Omega$ (for the definition see Tröltzsch (2009, page 21)). Additionally, we denote by $I := (0, T)$ a bounded time interval with $0 < T < \infty$. The state of the model is given by u, as we often have systems of PDEs we denote by $u^{(1)}, \dots, u^{(d)}$ the individual components of u, where d is the dimension of the system.

We consider models with the general form

$$\begin{aligned} u_t + C(u; \varphi) &= f(\varphi) \text{ in } \Omega \times I \\ \frac{\partial u}{\partial \nu_C} &= 0 \text{ in } \partial\Omega \times I \\ u &= u_0(\varphi) \text{ in } \Omega \times \{t = 0\}. \end{aligned} \tag{2.1}$$

Here and in the following, $\varphi \in \mathbb{R}_+^{n_\varphi}$ is a vector of kinetic parameters, the right hand side is given by a parameter dependent function $f(\varphi) \in L^2(\Omega)$. Furthermore, u_t denotes the partial derivative of u with respect to time and u_0 is the initial condition. We always assume that C is a semi-linear, elliptic, second order differential operator

with the form

$$C(u;\varphi) := \underbrace{-\sum_{i,j=1}^{n} a_{ij}(x;\varphi)\frac{\partial^2}{\partial x_i \partial x_j}u + \sum_{i=1}^{n} b_i(x;\varphi)\frac{\partial}{\partial x_i}u + c(x;\varphi)u}_{:=C_0(\varphi)u} + d(u;\varphi) \quad (2.2)$$

where $a_{ij}(x;\varphi)$, $b_i(x;\varphi)$, $c(x;\varphi)$ are scalar, linear functions of φ and $d(u;\varphi)$ is linear in φ. And $\partial u/\partial \nu_C$ denotes the conormal derivative with respect to C, i.e. $\nu^T C \nabla u = 0$. If φ is fixed we write $a_{ij}(x)$, $b_i(x)$, $c(x)$ and $d(u)$ for the sake of brevity. Furthermore, $A = a_{ij}{}_{i,j}$ is a symmetric, positive definite matrix with smallest eigenvalue greater or equal $\xi > 0$ (see Evans (1996)). C is a semi-linear operator. So we denote the linear part by $C_0(\varphi)$, which includes all derivatives, and the non-linear part by $d(u;\varphi)$. A classical physical example for C_0 is a diffusion process with substance degradation, i.e. $a_{ij} = \delta_{ij}$, $b_i = 0$, $c = \gamma$ and $d = 0$ ($\delta_{i,j}$ is the Dirac delta function, and γ is the degradation rate). Even though other boundary condition of Dirichlet or Robin type could be handled by the developed methods we will focus on homogeneous von Neumann conditions for the sake of brevity.

Often biological systems are in a steady state or quasi steady state. In these cases we assume that the derivative of u with respect to time is equal to zero. The resulting model is an elliptic equation with the same semi-linear, elliptic, second order differential operator as introduced above and the same boundary conditions. Of course no initial condition has to be given in this case. We consider stationary models of the form

$$\begin{aligned} C(u;\varphi) &= f(\varphi) \text{ in } \Omega \\ \frac{\partial u}{\partial \nu_C} &= 0 \text{ on } \partial\Omega. \end{aligned} \quad (2.3)$$

Stationary models

In the following we want to give a suitable framework for the description of stationary semi-linear PDEs. In this work we consider for stationary models $u \in H^1(\Omega)$, $f(\varphi) \in (H^1(\Omega))^*$ and it follows that $C : H^1(\Omega) \to (H^1(\Omega))^*$ for fixed φ.
Before we can define the weak solution of (2.3), we have to consider the non-linearity $d(u;\varphi)$. As mentioned before, the non-linearity in the equations is given by mass-action kinetics of second or higher order or Michaelis-Menten kinetics. We can conclude that for any meaningful kinetic parameter φ the mass in the reaction is conserved, i.e. if u is bounded almost everywhere for $x \in \Omega$, then

$$d(u;\varphi) \leq M \text{ and } d(0;\varphi) = 0 \quad (2.4)$$

a.e. in $x \in \Omega$ with upper bound $M > 0$ (Horn & Jackson, 1972). Furthermore, we demand that $d(u;\varphi)$ is continuous and monotonically increasing or decreasing in u. For some test function $v \in H^1(\Omega)$ we denote the bilinear form corresponding to the differential operator with

$$B(u,v) := \int_\Omega -\sum_{i,j=1}^{n} a_{ij}(x;\varphi)\frac{\partial^2}{\partial x_i \partial x_j}uvdx + \int_\Omega \sum_{i=1}^{n} b_i(x;\varphi)\frac{\partial}{\partial x_k}uvdx + \int_\Omega c(x;\varphi)uvdx$$

and define the right-hand side

$$F(u,v) := \int_\Omega f(\varphi)v \; dx.$$

Definition 2.1.1. *For the differential operator given in (2.2), which satisfies (2.4), the function $u \in H^1(\Omega)$ is called a weak solution of (2.3), if it satisfies the variational formulation:*

$$B(u,v) + \int_\Omega d(u;\varphi)vdx = F(u,v)$$

for all test functions $v \in (H^1(\Omega))^*$.

The existence and uniqueness strongly depend on the non-linearity $d(u;\varphi)$. With the previously made assumptions we can state the following existence theorem.

Theorem 2.1.1. *Given (2.4) and that $d(u;\varphi)$ is continuous and monotonically increasing or decreasing and $c(x,\varphi) \geq \delta > 0$ then (2.3) exhibits for all $f \in L^2(\Omega)$ exactly one weak solution $u \in H^1(\Omega)$.*

The existence and uniqueness of a weak solution u can be shown using the properties of monotone operators (the detailed proof can be found in Evans (1996) or Tröltzsch (2009)).

Dynamical models

For the parabolic case we need to account for the time-dependence of the state variable u and therefore introduce the space

$$W(I) = \{u | u \in L^2(I; H^1(\Omega)) \text{ and } u_t \in L^2(I; (H^1(\Omega))^*)\}.$$

Throughout this work we always assume that $u \in W(I)$ and $f \in (H^1(\Omega))^*$ for dynamical models of the form (2.1). It follows that $C : W(I) \to H^{-1}$ with the same properties given above for (2.2) and fixed φ. Again we have to consider the non-linearity in more detail before we address the weak solution of the problem. Of course (2.4) also holds for the time dependent case and all higher order MA kinetics are locally Lipschitz continuous in the state u hence we find that

$$\begin{aligned} |d(0;\varphi)| &\leq K \\ |d(\hat{u};\varphi) - d(\tilde{u};\varphi)| &\leq D(M)\,|\hat{u} - \tilde{u}|\,. \end{aligned} \tag{2.5}$$

for $\hat{u},\ \tilde{u} \in H^1(\Omega)$, $|\hat{u}| < M$, $|\tilde{u}| < M$, $K, M > 0$, $D(M) > 0$ and $x \in \Omega$.

Definition 2.1.2. *For the differential operator given in (2.2), which satisfies (2.4) and (2.5), the function $u \in W(I) \cap L^\infty(I \times \Omega)$ is called a weak solution of (2.1) if it satisfies the variational formulation:*

$$-\int_I \int_\Omega uv_t dxdt + \int_I \int_\Omega e(u,v) + d(u;\varphi)vdxdt = \int_I \int_\Omega f(\varphi)vdxdt + \int_\Omega u_0 v(\cdot,0)dx$$

for all test functions $v \in W(I)$, which satisfy $v(x,T) = 0$.

The existence and uniqueness of a weak solution u can be shown analogously to the ideas for the elliptic case.

Theorem 2.1.2. *Given (2.4), (2.5) and that $d(u;\varphi)$ is continuous and monotonically increasing or decreasing in u then there exists one weak solution $u \in W(I) \cap L^\infty(I \times \Omega)$ of 2.1 for all $f \in (H^1(\Omega))^*$, $u_0 \in L^2(\Omega)$.*

For a detailed proof we refer to the literature, e.g. (Tröltzsch, 2009, page 212). The strong assumptions on $d(u;\varphi)$ have to be made to show existence and uniqueness of solutions as well as existence of optimizers for the optimization problems considered later on. For models with non-linearities, which do not satisfy the requirements, existence and uniqueness have to be proven application dependent. This has been done for example for a Gierer-Meinhardt model in Garvie & Trenchea (2014). In this work, however, we only consider non-linearities, which satisfy (2.4) and (2.5).

Application example

We consider the basic process of production and degradation of a diffusive, biological substance in one dimension. Molecules of the substance, denoted by u, are produced in a region centred around the middle of the considered interval with a rate J. It is assumed that the production is highest in the middle of the region and decreases smoothly to the boundary, which is modelled by a Gaussian. Subsequently, the molecules diffuse with a rate D. The molecules can exist as monomers or dimers and the transition is described by the term αu^2 and u is the sum of both (the detailed deduction of the model equations is given in Section 5.2 or in Hersch *et al.* (2015)).

Example 2.1. *For the kinetic parameter vector $\varphi = (D, \alpha, J)$, $\Omega = [-7\mu m, 7\mu m]$ and $I = (0s, 100s)$, we consider the system*

$$\begin{aligned} u_t - D\triangle u + \alpha u^2 &= \frac{J}{2\pi\rho}\exp\left(-\frac{x^2}{\rho^2}\right) \ \textit{in } \Omega \times I \\ \frac{\partial u}{\partial \nu_C} &= 0 \ \textit{on } \partial\Omega \times I \\ u(0) &= u_0(\varphi) \ \textit{in } \Omega \times \{t = 0\}. \end{aligned}$$

In this example the components of the linear, elliptic differential operator C_0 are $a_{11} = D\delta_{11},\ b_1 = 0,\ c = 0$ and the non-linearity is $d(u;\varphi) = \alpha u^2$. The right side of the equation is given by f. This example is motivated by the application in Chapter 5 and will be used to evaluate the introduced uncertainty quantification methods in Chapter 4.

2.2. Parameter estimation

In this section we introduce the concepts of parameter optimization, uncertainty quantification and model selection for a given set of measured data. The key assumption is that the model is known as a function depending on the kinetic parameters either given analytically or numerically, i.e. we consider the reduced problem. To distinguish the reduced formulation from the formulation with PDE constraints,

we denote objective function, likelihood and profile likelihood for the reduced problem with a small letter and for the PDE constrained problem with a capital letter. Of special interest is the notion of parameter uncertainty and identifiability based on the calculation of confidence intervals.

2.2.1. Observation, measurement noise and likelihood function

As biological systems are complex and can react strongly to disturbances, the existing measurement techniques are limited and can often only capture scaled, partial observations of the system. Therefore we introduce the observation operator G as an operator which maps the solution of the model u onto the observation $y \in \mathbb{R}$.

$$G(s)u(t, x; \varphi) = y \tag{2.6}$$

with $\varphi \in \mathbb{R}_+^{n_\varphi}$ unknown parameter vector and $s \in \mathbb{R}_+^{n_s}$ is the vector of (possibly unknown) parameters of the observation operator. For example the scaled observation of the j-th model component for the measurement k would be given as

$$G(s)u(t, x; \varphi) = su_j(t_k, x_k; \varphi)$$

with $x_k \in \Omega$ and $t_k \in I$.

As a general setup we assume that we have a number of different observables denoted by the index $l = 1, \dots, N$ and for each observation a number of measurements at given time points and spatial locations (t_k, x_k) with $k = 1, \dots, M$,

$$G^{l,k}(s)u = y^{l,k}.$$

In this work we are concerned with the derivation of optimality conditions and uncertainty quantification methods which employ the Hessian of the objective function. To ensure the existence of the Hessian we always assume that the observation operators fulfill the following assumptions

Assumption 2.2.1 (Observation Operator)**.** *The operator $G^{l,k} : \mathbb{R}^{n_s} \times W(I) \cap L^\infty(I \times \Omega) \to \mathbb{R}$ for $l = 1, \dots, N$ and $k = 1, \dots, M$ is twice differentiable with respect to u and s.*

For sake of brevity we only consider the parabolic case as the elliptic problem is automatically included.

We assume that measurements considered during the parameter estimation process are a combination of the observation with a measurement noise term and we denote them by $\overline{y}^{l,k}$. In this work we consider two types of noise, namely additive, normally distributed noise and multiplicative, log-normally distributed noise. For the first case we obtain

$$\overline{y}^{l,k} = y^{l,k} + \epsilon_{l,k} \text{ with } \epsilon_{l,k} \sim \mathcal{N}(0, \sigma_{l,k}^2)$$

and hence $\overline{y}^{l,k} \sim \mathcal{N}(y^{l,k}, \sigma_{l,k}^2)$. This is the noise type most often considered in the literature. Furthermore, for known variances $\sigma_{j,k}^2$ it yields the well known weighted

least-squares objective function. In the case of log-normally distributed measurement noise we have the form

$$\overline{y}^{l,k} = y^{l,k} \cdot \epsilon_{l,k} \text{ with } \epsilon_{l,k} \sim \mathcal{LN}(0, \sigma_{l,k}^2)$$

and hence $\overline{y}^{l,k} \sim \mathcal{LN}(y^{l,k}, \sigma_{l,k}^2)$. The methodology developed in Chapter 3 for uncertainty quantification for PDE constraint inverse problems can also be applied to other noise types as long as a likelihood function exists.
For the case of additive, normally distributed measurement noise we obtain the corresponding likelihood function for $\theta = (\sigma, s, \varphi)$ as

$$l(\theta) = \prod_{l=1}^{N} \prod_{k=1}^{M} \frac{1}{\sqrt{2\pi}\sigma_{l,k}} \exp\left(-\frac{(\overline{y}^{l,k} - y^{l,k})^2}{2\sigma_{l,k}^2} \right). \tag{2.7}$$

And for the case of multiplicative, log-normally distributed noise and for measurements strictly greater zero we obtain

$$l(\theta) = \prod_{l=1}^{N} \prod_{k=1}^{M} \frac{1}{\sqrt{2\pi}\sigma_{l,k}\overline{y}^{l,k}} \exp\left(-\frac{(\log(\overline{y}_{l,k}) - \log(y^{l,k}))^2}{2\sigma_{l,k}^2} \right). \tag{2.8}$$

Based on the likelihood function the maximum likelihood estimator (MLE) of the unknown parameter $\theta \imath n \Theta \subset \mathbb{R}^{n_\sigma + n_s + n_\varphi}$ is defined as the parameter value $\hat{\theta}$, which maximizes the likelihood function, i.e.

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmax}}\, l(\theta). \tag{2.9}$$

Due to better numerical accuracy the maximum likelihood estimation is considered in terms of the negative logarithm of the likelihood

$$j(\theta) = -\log(l(\theta))$$

and the MLE is given as

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmin}}\, j(\theta).$$

For the above discussed additive, normally distributed noise for known $\sigma_{l,k}$ we obtain

$$\begin{aligned} j(\theta) &= \frac{1}{2} \sum_{l=1}^{N} \sum_{k=1}^{M} \log(2\pi\sigma_{l,k}^2) + \frac{(\overline{y}^{l,k} - y^{l,k})^2}{\sigma_{l,k}^2} \\ &= const + \sum_{l=1}^{N} \sum_{k=1}^{M} \frac{(\overline{y}^{l,k} - y^{l,k})^2}{\sigma_{l,k}^2}. \end{aligned}$$

This is the weighted least squares objective function plus an additive constant. In most applications $\sigma_{l,k}$ is unknown and has to be estimated based on equation (2.7) or (2.8).

2.2.2. Identifiability analysis

A great challenge in biological applications is the difficulty or impossibility to determine unique parameter values from measurement data (Gutenkunst *et al.*, 2007). We can distinguish between two types: structural and practical identifiability. Structural identifiability is a property of dynamic model and observation, which is independent of the measurements. It has to be analyzed by analytical methods and assessing structural identifiability for non-linear dynamical systems is hard and the methods are computationally expansive (Chis *et al.*, 2011). In contrast to the theoretical property of structural identifiability, practical identifiability takes into account the finite number of measurements, the noise level and model errors. Practical identifiability analysis is an a posteriori analysis, which addresses the question when the parameters can be uniquely estimated with the available experimental data. This closely relates to the question of parameter uncertainty and both can be addressed in terms of confidence intervals.

Definition 2.2.1. *We call a parameter θ_i practically non-identifiable if the likelihood based confidence interval extends infinitely or beyond the considered parameter region in at least one direction.*

2.2.3. Confidence intervals

To quantify parameter identifiability in terms of confidence intervals we are going to introduce confidence intervals or confidence regions. In this work we introduce and consider likelihood based confidence intervals, which are good approximations for parameter uncertainty in the case of a small number of data points (Venzon & Moolgavkar, 1988; Meeker & Escobar, 1995; Raue *et al.*, 2009).

Definition 2.2.2 (Confidence region and interval)**.**
For $\hat{\theta} \in \Theta$ which maximizes (2.9) we define the likelihood based confidence region for the confidence level α as

$$CI_\alpha(\theta) = \left\{ \theta \in \Theta \,\middle|\, \frac{l(\theta)}{l(\hat{\theta})} \geq \exp\left(-\frac{\delta_\alpha}{2}\right) \right\}$$

with $\delta_\alpha = \chi^2(\alpha, n_\theta)$. The confidence interval for a single parameter component θ_i is the projection of CI onto the parameter direction e_{θ_i}

$$CI_\alpha(\theta_i) = \mathcal{P}_{e_{\theta_i}} CI_\alpha(\theta).$$

The critical part in determining the bounds of confidence regions is the calculation of the root of $l(\theta)/l(\hat{\theta}) - \exp\left(-\frac{\delta_\alpha}{2}\right)$. Furthermore, the confidence region is not always connected and there might exist multiple roots of the equation. In the following we describe variants to approximate the root and discuss their strengths.

The most commonly used method to calculate the quotient is based on the second order approximation of the negative, logarithm of the likelihood function around the optimum, i.e.

$$j(\hat{\theta} - \triangle\theta) = j(\hat{\theta}) + \frac{\partial j}{\partial \theta}(\hat{\theta})(\triangle\theta) + \frac{\partial^2 j}{\partial^2 \theta}(\hat{\theta})(\triangle\theta)^2.$$

In the definition of the CI we consider $log(l(\theta)/l(\hat{\theta})) = j(\hat{\theta}) - j(\theta)$ hence the confidence region for θ_i are approximated by

$$CI_\alpha(\theta_i) = \left\{ \hat{\theta} + \triangle\theta \in \Theta \left| -\frac{\partial^2 j}{\partial^2 \theta}(\hat{\theta})(\triangle\theta)^2 \leq \frac{\delta_\alpha}{2} \right. \right\}. \tag{2.10}$$

with $\triangle\theta_j = 0$ for $j \neq i$ and $\triangle\theta_j \neq 0$ for $i = j$. If the noise level is given, the Hessian of the likelihood function is the Fisher Information matrix. This method is often used in applications. However, these confidence intervals tend to underestimate the uncertainty in the parameters (Banks *et al.*, 2010). Furthermore, the method is sufficient to hint at structurally non-identifiable parameters (zero eigenvalues of the Hessian) but is not sufficient to determine practically non-identifiable parameters. A better suited method to determine exact confidence regions and practically non-identifiability are likelihood profiles (Murphy *et al.*, 2000; Raue *et al.*, 2009).

The likelihood profile for a parameter θ_i, $i = 1, \ldots, n_\theta$ of θ is defined as

$$pl_c(\theta_i) = \max_{\theta \in \Theta} l(\theta) \text{ subject to } \theta_i = c. \tag{2.11}$$

Based on the likelihood profile (the upper and lower bounds of) the confidence region can be calculated as

$$CI_\alpha(\theta_i) = \left\{ c \left| \frac{pl_c(\theta_i)}{l(\hat{\theta})} \geq \exp\left(-\frac{\delta_\alpha}{2}\right) \right. \right\}.$$

Another important property of $pl_c(\theta_i)$ is that it gives the precise shape of the likelihood function in direction θ_i, if we calculate it for a continuous curve c.

2.2.4. Model selection

A key question in biological modeling and parameter estimation is the decision for or against a biological hypothesis based on the given experimental data. To this end models are introduced based on the hypothesis that the parameters are estimated and the best model is selected. Following Burnham & Anderson (2002) we focus on the Akaike information criterion and the Bayesian information criterion for model selection as they are superior to standard hypothesis testing.

Akaike information criterion (AIC) The AIC relies on the maximum likelihood estimate $\hat{\theta}$ of each model:

$$AIC = -2\log(l(\hat{\theta})) + 2n_\theta$$

with n_θ number of parameters in the model. As the AIC is only based on the MLE and the number of parameters, it is applicable to a wide range of models. It can be used if the true model is not in the set of tested models and also non-nested models can be considered. During the model selection it has to be kept in mind that the AIC tends to choose larger models in comparison to other selection criteria (Burnham & Anderson, 2002).

Bayesian information criterion (BIC) The BIC as well as the AIC are calculated with the maximum likelihood estimate $\hat{\theta}$ of each model:

$$BIC = -2\log(l(\hat{\theta})) + n_\theta \log(M)$$

with n_θ number of parameters in the model and M is the number of data points. Like the AIC, the BIC can be used to compare a wide range of models, which do not have to be nested.

Based on AIC and BIC the best model to select is the model with the smallest BIC or AIC value. If the difference between AIC/BIC values for two competing models is below 10, there is essentially no empirical support for one model over the other.

2.3. PDE constrained optimization

In this section we introduce the notation used for the PDE constrained optimizations considered in this work. In contrast to the previous section we consider the model as a constraint and not as a given function. The optimization is then performed with regard to the parameter vector $\theta = (\sigma, s, \varphi)$ and the function u, which satisfies the model constraints. To distinguish the objective and likelihood function for the constrained case from the reduced formulation, we denote them by capital letters. Furthermore, we review some basic principles of optimization with semi-linear PDE constraints like the Lagrangian formulation as well as reduced problems, sensitivities and adjoints. We sketch the general ideas here and refer the interested reader to the books of Tröltzsch (2009) and Hinze *et al.* (2009) for the detailed theory and proofs.

For the existence of an optimal solution of the optimization problem with for non-linear PDE constraint we follow Hinze *et al.* (2009, page 55) and consider the optimization problem in a general form

$$\begin{aligned} &\min_{\theta\in\Theta, u\in U} J(\theta, u) \\ &\text{subject to } e(\theta, u) = 0 \end{aligned} \tag{2.12}$$

where $J : \Theta \times H^1(\Omega) \to \mathbb{R}$ is the negative logarithm of the likelihood function, e.g. (2.7) or (2.8) and $e : \mathbb{R}^{n_\theta} \times V \to V^*$ in the elliptic and $e : \mathbb{R}^{n_\theta} \times W(I) \to L^2(I, V^*)$ in the parabolic case. The model is given as the constraint and for a more general consideration the right hand side, the boundary conditions and the initial conditions are subsumed in one equality constraint. With the models introduced in the previous sections we can write $e(\theta, u)$ as

$$e(\theta, u) = C(u; \varphi) - f(\varphi) - \frac{\partial u}{\partial \nu_C}$$

or

$$e(\theta, u) = u_t + C(u; \varphi) - f(\varphi) - \frac{\partial u}{\partial \nu_C} + (u - u_0).$$

Here $C(u; \varphi)$ is the semi-linear elliptic operator encoding for example the diffusion and all mass action kinetics higher then zeroth order and $f(\varphi)$ encodes all zeroth order mass action kinetics or model independent inputs of the model. Three key assumptions have to be made to guarantee the existence of optimal solutions:

- $e(\theta, u) = 0$ has a bounded solution, i.e. there exists a solution u, which satisfies the PDE as well as boundary conditions and initial conditions of the model for a parameter θ and this solution has an upper bound almost everywhere.
- $(\theta, u) \mapsto e(\theta, u)$ is continuous under weak convergence.
- $J(\theta, u)$ is sequentially weakly lower semi-continuous, i.e. every sequence $(\theta_n, u_n)_n$ of parameter and solution pairs, which converges weakly against an optimal pair (θ, u) we have $J(\theta, u) \leq \liminf_{n\to\infty} J(\theta_n, u_n)$.

Then we can state the following theorem as in Hinze *et al.* (2009, page 55)

Theorem 2.3.1. *If the assumptions above hold Problem (2.12) has an optimal solution* $(\hat{\theta}, \hat{u})$.

Hence at least one optimum exists, however, its uniqueness is not guaranteed by the assumptions. Also local optima are possible.

2.3.1. Lagrangian formulation

To find the optimality conditions for optimization problems with possibly complex semi-linear PDE constraints, the Karush-Kuhn-Tucker (KKT) theory for optimality conditions has to be applied. This becomes difficult for semi-linear PDEs as operators, functionals and matching spaces have to be chosen in a way that differentiability is guaranteed. In this work we do not focus on the deduction of optimality conditions therefore we introduce the formal Lagrange principle. We considered the differential operators formally and assume that the state and all derivatives are L^2 functions with L^2 inner products. We will sketch the main results for the considered systems here and all proofs, detailed derivations and the discussion of more general function spaces can be found in Tröltzsch (2009) and Hinze *et al.* (2009). For the sake of brevity we subsumed noise, observation and kinetic parameters and write only θ.

We define the Lagrangian function of (2.12) with constraints given by (2.3), as

$$\mathcal{L}(\theta, u, p) = J(\theta, u) + \int_\Omega (C_0(\theta)u + d(u;\theta) - f(\theta))p dx + \int_{\partial\Omega} \frac{\partial u}{\partial \nu_C} p ds.$$

Following Theorem 2.3.1 an optimal point $(\hat{\theta}, \hat{u})$ exists and for this point the derivative of $\mathcal{L}$ with respect to θ and u has to vanish. It follows that a function $p \in H^1(\Omega) \cap C(\overline{\Omega})$ exists that fulfills

$$\nabla_\theta \mathcal{L}(\hat{\theta}, \hat{u}, p)(\theta - \hat{\theta}) = 0 \ \forall \theta \in \Theta \tag{2.13}$$

$$\nabla_u \mathcal{L}(\hat{\theta}, \hat{u}, p)u = 0 \ \forall u \in H^1(\Omega). \tag{2.14}$$

If we consider (2.14) we obtain

$$\nabla_u \mathcal{L}(\hat{\theta}, \hat{u}, p)u = J_u(\hat{\theta}, \hat{u})u + \int_\Omega (C_0(\hat{\theta})p + d_u(\hat{u};\hat{\theta})p)u dx + \int_{\partial\Omega} \frac{\partial p}{\partial \nu_{C^*}} u ds = 0$$

for all $u \in H^1(\Omega)$. This is the variational formulation of the weak solution p of

$$\begin{aligned} C_0^*(\hat{\theta})p + d_u(\hat{u};\hat{\theta})p &= -J_u(\hat{\theta},\hat{u}) \text{ in } \Omega \\ \frac{\partial p}{\partial \nu_{C^*}} &= 0 \text{ on } \partial\Omega, \end{aligned} \tag{2.15}$$

which we will call adjoint equation of the elliptic problem in the following.
Following the same steps as for the elliptic case we can write the Lagrangian function of (2.12) with constraint (2.1), as

$$\begin{aligned} \mathcal{L}(\theta,u,p) =& J(\theta,u) + \int_I \int_\Omega (u_t + C_0(\theta)u + d(u;\theta) - f(\theta))p dx dt \\ &+ \int_I \int_{\partial\Omega} \frac{\partial u}{\partial \nu_C} p ds dt + \int_\Omega (u_0(\theta) - u_0(\theta))p(0,x)dx. \end{aligned}$$

And we obtain the adjoint equation for the parabolic problem as

$$\begin{aligned} -p_t + C_0^*(\hat{\theta})p + d_u(\hat{u};\hat{\theta})p &= -J_u(\hat{\theta},\hat{u}) \text{ in } \Omega \times I \\ \frac{\partial p}{\partial \nu_{C^*}} &= 0 \text{ on } \partial\Omega \times I \\ p &= 0 \text{ on } \Omega \times \{t = T\}. \end{aligned} \tag{2.16}$$

Due to the non-linearity it is not possible to show existence of unique optimal parameter sets $\hat{\theta}$ in general. For the uncertainty quantification and identifiability analysis developed in this work, however, we require only existence of local optimizers, which can be shown for the considered models and strong requirements for $d(u;\varphi)$ (Hinze *et al.*, 2009; Tröltzsch, 2009). Furthermore, we often expect that for the model and data considered no unique optimizer can be found, i.e. we have non-identifiable problems.

2.3.2. Reduced problem

Often the reduced form of the problem (2.12) is considered as it is given for example by analytical or numerical solutions of the PDE. To constructed the reduced problem we assume that for each $\theta \in \Theta$ there exists a corresponding unique solution $u(\theta)$. By inserting $u(\theta)$ in $J(\theta,u)$ we obtain the reduced objective function, denoted by a small letter in contrast to the non-reduced objective function,

$$j(\theta) := J(\theta, u(\theta)). \tag{2.17}$$

For the solution of the optimization the derivative of $j(\theta)$ has to be computed. There are two methods to do this: the forward and the adjoint sensitivity approach. We introduce both in the following.

The forward sensitivity approach is based on the directional derivatives of u with respect to θ. For a given direction θ_i the chain rule of differentiation yields

$$\frac{d}{d\theta_i} j(\theta) = J_u(\theta, u(\theta)) \frac{du}{d\theta_i} + J_{\theta_i}(\theta, u(\theta)).$$

In this equation the forward sensitivities occur as $\frac{du}{d\theta_i} := u_{\theta_i}$ and for the considered model types the forward sensitivities can be calculated by solving the linearized state equations corresponding to (2.3)

$$\begin{aligned} C_u(u,\theta)u_{\theta_i} + C_{\theta_i}(u,\theta) &= f_{\theta_i}(\theta) \text{ in } \Omega \\ \frac{\partial u_{\theta_i}}{\partial \nu_C} &= 0 \text{ on } \partial\Omega. \end{aligned}$$

And for model (2.1) we obtain

$$\begin{aligned} u_{\theta_i,t} + C_u(u,\theta)u_{\theta_i} + C_{\theta_i}(u,\theta) &= f_{\theta_i}(\theta_i) \text{ in } \Omega \times I \\ \frac{\partial u_{\theta_i}}{\partial \nu_C} &= 0 \text{ on } \partial\Omega \times I \\ u_{\theta_i} &= \frac{du_0}{d\theta_i} \text{ in } \Omega \times \{t = 0\}. \end{aligned}$$

So to obtain the derivative of $j(\theta)$ we have to solve as many linear PDEs as we have parameters in the system. For models with many parameters as they often occur in biological applications the computational effort grows linearly with the number of parameters.

The adjoint sensitivity approach is based on the solution p of adjoint equation (2.15) or (2.16) introduced previously. Furthermore, the adjoint operator of the PDE (2.3) or (2.1) has to be calculated, which we denote by $e(\theta, u)^*$. Then we can derive that

$$j'(\theta) = e_\theta(\theta, u(\theta))^* p(\theta) + J_\theta(\theta, u(\theta)).$$

So if the adjoint equation can be differentiated with respect to the parameters for the given model, the calculation of the derivative by the adjoint sensitivity approach only involves the solution of one additional PDE.

To summarize this chapter, we began by introducing the two model types we consider in this work: elliptic, steady state models and parabolic, dynamical models. Both models are semi-linear and the differential operator can be decomposed in a linear differential operator and a non-linear part, which is given by biologically motivated reaction terms for example mass-action or Michaelis-Menten kinetics. The non-linearity satisfies strict assumptions regarding boundedness and monotonicity and therefore the models have unique weak solutions. Furthermore, these characteristics enable the existence of optimizers if the models are the constraint of an optimization problem as discussed above. Such optimization problems arise if we consider parameter estimation problems, which we also discussed in this chapter. A special focus was on the general introduction of likelihood functions and likelihood based uncertainty analysis, which will be extended to the special case of semi-linear PDE models and image measurements in the following.

3. Image based parameter estimation and uncertainty quantification

Todays experimental methods yield quantitative measurements of molecules in-vivo can be obtained by a combination of antibody stainings and high resolution microscopy. This yields two or three dimensional spatially resolved images of the underlying biological process (Huang *et al.*, 2010; Schumann *et al.*, 2010; Weber *et al.*, 2013). Despite this high resolution data, the usage of quantitative spatio-temporal models is limited. In recent years the field of PDE-constrained optimization emerged, providing the theory and methods to estimate parameters of PDEs (Hinze *et al.*, 2009). Nevertheless, specific problems occurring in biological problems, like partial observations, sparse measurements and high noise levels, have yet to be addressed as it has been done for ODE parameter estimation problems. In particular, appropriate uncertainty quantification methods for the efficient and reliable analysis of practical identifiability are not available (Raue *et al.*, 2009). In this chapter we want to introduce the general setting of image based parameter optimization and uncertainty quantification for semi-linear PDEs.

After introducing the general setup for images based parameter estimation and the problem formulation in Section 3.1, we proceed by formulating the second order approximation of the parameter uncertainty in Section 3.2. A special focus lies on the deduction of the Hessian matrix for the full PDE constrained problem and the comparison to the Hessian matrix of the reduced problem this is also part of the paper by Boiger *et al.* (2016). This Section is followed by the introduction of a profile likelihood analysis based on the reduced formulation of the problem in Section 3.3. This part of the chapter is already published in the paper of Hock *et al.* (2013).

3.1. Introduction and problem statement

The key question we want to address in this chapter is the parameter estimation for reaction-diffusion models based on microscopy images with a special focus on the development of uncertainty quantification methods for such estimation processes. Each parameter estimation process starts with the biological measurements. The application examples in this thesis consider measurements of signaling molecules, which consist of microscopy images or stacks of microscopy images obtained by fluorescent stainings or labels for the certain protein of interest. These can either be images of single cells as in the Pom1P application example in Chapter 5 or tissue scale images as discussed in the CCL21 application example in Chapter 6.

For both types of images the standard procedure to estimate parameters is an sequential process consisting of an image preprocessing phase to remove e.g. struc-

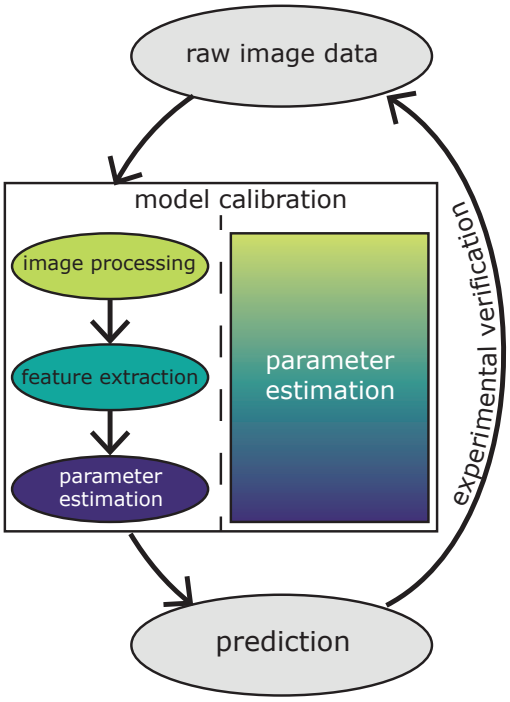


Figure 3.1: ***Model calibration from microscopy images.*** *The schematic shows the modeling process with imaging data, model calibration, prediction and experimental verification. In the following we will distinguish between two types of model calibration processes: The standard, which reduces the data by image processing and feature selection is shown on the left side. The new method, which gradually passes through the steps by incorporating the whole data in the estimation process is shown on the right side.*

tured or unstructured noise, an extraction of features like size or intensity and the following parameter estimation for a mathematical model (Sbalzarini, 2013). A major challenge for parameter estimation from tissue scale microscopy images is the presence of structured noise, which can strongly bias the estimation process. In the sequential approach the noise is removed by filtering methods. We propose a new integrated approach which treats it as an additional noise term. The difference between both processes in schematically depicted in Figure 3.1.

As introduced in Section 2.1 we consider semi-linear partial differential equations of the general form:

$$\begin{aligned} u_t + C_0(\varphi)u + d(u;\varphi) &= f(\varphi) \text{ in } \Omega \times I \\ \frac{\partial u}{\partial \nu_C} &= 0 \text{ in } \partial\Omega \times I \\ u &= u_0(\varphi) \text{ in } \Omega \times \{t = 0\}. \end{aligned} \tag{3.1}$$

$C_0(\varphi)$ is the linear, elliptic partial differential operator, $d(u;\varphi)$ is the non-linearity and $f(\varphi)$ is the right side as defined in Section 2.1. In the following we will always assume that the state of the model only depends on the model parameters φ, the observation only depends on the observation parameters s and the likelihood parameters are subsumed under the identifier σ. All three can be part of the estimation process with $\theta = (\sigma, s, \varphi) \in \Theta = \mathbb{R}^{n_\theta}$, $n_\theta = n_\sigma + n_s + n_\varphi$.

3.1.1. Observation operator for microscopy images

In Section 2.2 we introduced observation operators for parameter estimation in general. Here we want to introduce an observation operator tailored to the use with

two dimensional microscopy images as they arise for example from tissue slice immunostainings. The operator can either be used for the full images in combination with the integrated estimation approach or it can be applied to the images after filtering for the sequential approach.
The observation operator and the corresponding likelihood function were introduced in the context of 2D microscopy images and are used in the application in Chapter 6 (Hock *et al.*, 2013). A time-series of images was taken at time points $t_k \in I$ for $k = 1, \ldots, M$, which are not necessarily equally spaced. For each time point the number of pixels is denoted with n_k and the pixel area with $\mathcal{A}_{p_l}$ for $l = 1, \ldots, n_k$. We consider a partial observation of the full state $u(t, x; \varphi)$, given by the possibly parameter dependent function h. The l-th observation operator $G^{l,k}$ maps the function u for a given time point to the observation $y^{l,k} \in \mathbb{R}$.

Definition 3.1.1 (pixel based observation operator)**.** *The l-th component of the pixel based observation operator for the time point t_k with $G^{l,k} : \mathbb{R}^{n_s} \times W(I) \cap L^\infty(I \times \Omega) \to \mathbb{R}$, is defined as*

$$G^{l,k}(s, u) = b + \int_{\mathcal{A}_{p_l}} h(s, u(t_k, x; \varphi))dx, \tag{3.2}$$

with $k = 1, \ldots, M$, $t_k \in I$, $l = 1, \ldots, n_k$ and $b \in \mathbb{R}_+$. h is a continuous function of $n_s + d$ variables, which is linear with respect to both arguments.

Here b denotes a constant off-set due to background luminescence.

3.1.2. Integrated noise model

For both approaches (sequential and integrated) we always consider a likelihood functions to estimate the kinetic parameters of the underlying biological process based on reaction-diffusion models (see Section 2.2). The actual form of the likelihood depends on the type of measurement noise.

The sequential parameter estimation approach often considers additive normal noise or multiplicative log-normal noise, which yields the likelihood functions described in Section 2.2:

$$L_1(\theta, y) = \prod_{l,k=1}^{N,M} \frac{1}{\sqrt{2\pi}\sigma_{l,k}} \exp\left(-\frac{(\overline{y}_{l,k} - y^{l,k})^2}{2\sigma_{l,k}^2}\right)$$

and

$$L_2(\theta, y) = \prod_{l,k=1}^{N,M} \frac{1}{\sqrt{2\pi}\sigma_{l,k}\overline{y}^{l,k}} \exp\left(-\frac{(\log(\overline{y}^{l,k}) - \log(y^{l,k})^2}{2\sigma_{l,k}^2}\right).$$

Parameter estimation with those functions assumes that the measured quantity at every pixel of an image is generated by the underlying model and the assumed measurement noise. Therefore, it requires a pre-processing of the images by filtering or the extraction of a specific feature to remove all information from the images, which is assumed to be not considered by the model or is Gaussian white noise. Otherwise the estimated parameters can be biased. On the other hand, the removal of data, i.e. filtering, or the restriction to certain elements of the data, i.e. feature

extraction, can change the perceived uncertainty in the estimates. To circumvent both, bias and changes in the uncertainty, we introduce the integrated approach in the following.
In the case of the integrated noise model, we assume that the data is generated by a mixture of at least two components: a model dependent one and an independent one. Both processes are combined as a weighted sum in the likelihood function with weight w_1 for the model dependent part and weight w_2 for the structured noise. Both weights add up to one and we require that in the case of no noise $w_2 = 0$ and the likelihood becomes the standard likelihood L_1 or L_2. This leaves room for a lot of model dependent combinations. However, we only want to show the two likelihoods considered in Chapter 6.3. First we assume that the structured noise is normally distributed. We obtain

$$\begin{aligned} L_3(\theta, y) = \prod_{l,k=1}^{N,M} w_1 \frac{1}{\sqrt{2\pi}\sigma_{l,k}\overline{y}^{l,k}} \exp\left(-\frac{(\log(\overline{y}^{l,k}) - \log(y^{l,k})^2}{2\sigma_{l,k}^2}\right) \\ + w_2 \frac{1}{\sqrt{2\pi}\sigma_{2,l,k}} \exp\left(-\frac{(\overline{y}^{l,k} - \mu_2)^2}{2\sigma_{2,l,k}^2}\right). \end{aligned} \tag{3.3}$$

Second, the structured noise is log-normally distributed

$$\begin{aligned} L_4(\theta, y) = \prod_{l,k=1}^{N,M} w_1 \frac{1}{\sqrt{2\pi}\sigma_{l,k}\overline{y}^{l,k}} \exp\left(-\frac{(\log(\overline{y}^{l,k}) - \log(y^{l,k})^2}{2\sigma_{l,k}^2}\right) \\ + w_2 \frac{1}{\sqrt{2\pi}\sigma_{2,l,k}\overline{y}^{l,k}} \exp\left(-\frac{(\log(\overline{y}^{l,k}) - \mu_2)^2}{2\sigma_{2,l,k}^2}\right), \end{aligned} \tag{3.4}$$

in this case we obtain the parameter vector $\sigma = (\{\sigma_{l,k}\}_{l,k}, \{\sigma_{2,l,k}\}_{l,k}, w_1, w_2)$.

These are very specialized likelihood functions for parameter estimation with 2D microscopy images and the theory we develop in the following holds for this type of likelihood as well as for all likelihood functions, which fulfill the following assumptions:

- the parameter dependencies are as given above for $\theta = (\sigma, s, \varphi)$.
- the observation operator $G^{l,k}$ is twice differentiable with respect to u and s for $l = 1, \ldots, N$, $k = 1, \ldots, M$.
- the likelihood function is twice differentiable with respect to $y^{l,k}$ and σ.
- the negative logarithm of the likelihood function denoted by J can be written component wise such that

$$J(\sigma, y) = \sum_{l,k=1}^{N,M} J^{l,k}(\sigma, y^{l,k}).$$

3.1.3. Problem statement

The general parameter estimation problem considered is given by the following constrained optimization problem

$$\min_{s\in\mathbb{R}^{n_s},\sigma\in\mathbb{R}^{n_\sigma},\varphi\in\mathbb{R}^{n_\varphi},u\in W(I)\cap L^\infty} J(\sigma,y)$$
$$\text{s.t.}\begin{cases} G(s)u = y \\ u_t + C(u;\varphi) = f(\varphi) \text{ in } \Omega\times I \\ \dfrac{\partial u}{\partial \nu_C} = 0 \text{ in } \partial\Omega\times I \\ u = u_0(\varphi) \text{ in } \Omega\times\{t=0\} \end{cases} \tag{3.5}$$

The equality constraint is to be understood in the weak sense and J is the negative logarithm of the likelihood functions as described in Section 2.2. For optimization problems of the above form with ODE instead of PDE constraints the parameter estimation and identifiability analysis has been thoroughly discussed before, e.g. Raue *et al.* (2009). A special focus of this work is on the observation operator, which has not yet been considered in the context of optimization with semi-linear PDE constraints, and the problem specific likelihood functions for the integrated noise model.

Problem 3.1 (Uncertainty analysis for semi-linear PDE models)**.** *Given the semi-linear, spatio-temporal model, the observation operator and the problem dependent likelihood function derive the corresponding Hessian matrix for second order parameter uncertainty approximation.*

In Section 2.2 the profile likelihood calculation was introduced as a uncertainty quantification method superior to the second order approximation by the Hessian matrix. This was done for the reduced problem and the ODE case but a discussion in the PDE context is still missing.

Problem 3.2 (Profile likelihood for semi-linear PDE models)**.** *Given the parameter estimation problem (3.5) for imaging data develop a profile likelihood base uncertainty and practical identifiability analysis.*

3.2. Parameter estimation

Before we address uncertainty quantification for the parameter estimation problem we deduce the sufficient and necessary optimality conditions for the optimization problem with observation (3.5). We consider only the parabolic case in detail based on the formal Lagrange principle as the elliptic case can be considered to be a special case of the parabolic one. Based on the necessary optimality conditions we conclude in which cases the observation operator might induce non-identifiability in our models.

3.2.1. Optimality conditions

For simplicity of calculation we assume that we can directly calculate the equality constraint with the observation operator and consider the objective function

$$J(\sigma, G(s)u) = \sum_{l,k=1}^{N,M} J^{l,k}(\sigma, G^{l,k}(s)u)$$

and $\theta = (\sigma, s, \varphi)$. The Lagrange function corresponding to the constrained minimization problem (3.5) reads then

$$\begin{aligned}\mathcal{L}(\theta, u, p) = &\sum_{l,k=1}^{N,M} J^{l,k}(\sigma, G^{l,k}(s)u) \\ &+ \int_I \int_\Omega (u_t + C_0(\varphi)u + d(u,\varphi) - f(\varphi))p dx dt \\ &+ \int_I \int_{\partial\Omega} \partial_{\nu_C} u p ds dt + \int_\Omega (u_0 - u_0(\varphi))p(0)dx.\end{aligned} \tag{3.6}$$

If we compare the obtained Lagrangian to the Lagrangian without observation operator introduced in Section 2.3.1 only the first term is different.

Sufficient optimality condition Let $(\hat{\theta}, \hat{u})$ be a minimizer and also a KKT point of (3.5), i.e. the MLE. We have the corresponding Lagrange multiplier p and $\hat{\theta} = (\hat{\sigma}, \hat{s}, \hat{\varphi})$. This implies that the gradient of the objective function and the Lagrangian function vanishes at this point and we have

$$\nabla_{\sigma,s,\varphi,u,p}\mathcal{L}(\hat{\sigma}, \hat{s}, \hat{\varphi}, \hat{u}, p) = 0 \tag{3.7}$$

where (skipping the arguments $(\sigma, s, \varphi, u, p)$ of $\mathcal{L}$ for the moment)

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \sigma_i} &= \sum_{l,k=1}^{M,N} J^{l,k}_{\sigma_i}(\sigma, G^{l,k}(s)u)) \text{ for } i = 1, \dots, n_\sigma \\ \frac{\partial \mathcal{L}}{\partial s_i} &= \sum_{l,k=1}^{M,N} J^{l,k}_{y}(\sigma, G^{l,k}(s)u)) G^{l,k}_{s_i}(s)u \text{ for } i = 1, \dots, n_s \\ \frac{\partial \mathcal{L}}{\partial \varphi_i} &= \int_I \int_\Omega (C_{0,\varphi_i}(\varphi)u + d_{\varphi_i}(u,\varphi) - f_{\varphi_i}(\varphi))p dx dt \text{ for } i = 1, \dots, n_\varphi \\ \frac{\partial \mathcal{L}}{\partial u}[du] &= \sum_{l,k=1}^{M,N} J^{l,k}_{y}(\sigma, G^{l,k}(s)u)) G^{l,k}(s)[du] \\ &\quad + \int_{I,\Omega} (-p_t + C^*_{0,u}(\varphi)p + d_u(u;\varphi)p)[du]dxdt + \int_\Omega p(T)[du]dx \\ \frac{\partial \mathcal{L}}{\partial p}[dp] &= \int_{I,\Omega} (u_t + C_{0,u}(\varphi)u + d(u;\varphi))[dp]dxdt + \int_\Omega (u(0) - u_0(\varphi))[dp](0)dx.\end{aligned} \tag{3.8}$$

Here $[du]$ denotes the direction of the derivative. We see that the observation operator becomes part of the derivatives with respect to the observation parameter s

and also with respect to the state. But this does not happen in the equations for the kinetic parameters and the model equation. We can especially observe that the observation operator becomes part of the adjoint state equation at the MLE (see Section 2.3 for the deduction):

$$\begin{aligned} -p_t + C_0(\hat{\varphi})^* p + d_u(\hat{u}, \hat{\varphi}) p &= -\sum_{l,k=1}^{N,M} J_y(\hat{\sigma}, G^{l,k}(\hat{s})\hat{u}) G^{l,k}(\hat{s}) \text{ in } \Omega \times I \\ \frac{\partial p}{\partial \nu_{C^*}} &= 0 \text{ on } \partial\Omega \times I \\ p &= 0 \text{ on } \Omega \times \{t = T\}. \end{aligned} \tag{3.9}$$

For the pixel based observation operator (3.2) the term $G^{l,k}(\hat{s})$ equals the scaled image size or zero if the component is not considered. Hence it can be considered to be a weighting of the components of J_y.

Hessian matrix For the deduction of the Hessian matrix we consider a Taylor expansion around the MLE. We take any other $(\tilde{\sigma}, \tilde{s}, \tilde{\varphi}, \tilde{u})$ that satisfies the constraint (3.5). For the difference of the cost function values we have:

$$J(\tilde{s}, \tilde{\sigma}, \tilde{\varphi}, \tilde{u}) - J(\hat{\sigma}, \hat{s}, \hat{\varphi}, \hat{u}) = \mathcal{L}(\tilde{s}, \tilde{\sigma}, \tilde{\varphi}, \tilde{u}, p) - \mathcal{L}(\hat{\sigma}, \hat{s}, \hat{\varphi}, \hat{u}, p)$$

Due to equation (3.7) the first order Taylor term vanishes if we expand the difference in the Lagrangian functions. The second order term of the Taylor expansion contains the Hessian matrix we are interested in. We use the notation $s^\mu := \hat{s} + \mu(\tilde{s} - \hat{s})$, $\sigma^\mu := \hat{\sigma} + \mu(\tilde{\sigma} - \hat{\sigma})$, $u^\mu := \hat{u} + \mu(\tilde{u} - \hat{u})$, $\varphi^\mu := \hat{\varphi} + \mu(\tilde{\varphi} - \hat{\varphi})$ for the direction in the Taylor expansion and can write formally:

$$\mathcal{L}(\tilde{\cdot}, p) - \mathcal{L}(\hat{\cdot}, p) = \int_0^1 \begin{pmatrix} \tilde{s} - \hat{s} \\ \tilde{\sigma} - \hat{\sigma} \\ \tilde{\varphi} - \hat{\varphi} \\ \tilde{u} - \hat{u} \end{pmatrix}^T \nabla^2 \mathcal{L}_{s,\sigma,\varphi,u}(s^\mu, \sigma^\mu, \varphi^\mu, u^\mu, p) \begin{pmatrix} \tilde{s} - \hat{s} \\ \tilde{\sigma} - \hat{\sigma} \\ \tilde{\varphi} - \hat{\varphi} \\ \tilde{u} - \hat{u} \end{pmatrix} d\mu$$

To obtain the Hessian matrix we calculate the second order derivative of $\mathcal{L}$ and obtain the following integral

$$\begin{aligned} &\int_0^1 \nabla^2 \mathcal{L}_{s,\sigma,\varphi,u}(s^\mu, \sigma^\mu, \varphi^\mu, u^\mu, p)[(\tilde{s} - \hat{s}, \tilde{\sigma} - \hat{\sigma}, \tilde{\varphi} - \hat{\varphi}, \tilde{u} - \hat{u})^2] d\mu \\ &= \int_0^1 \Big(\sum_{i,j=1}^{n_s} (\tilde{s}^i - \hat{s}^i)(\tilde{s}^j - \hat{s}^j) \sum_{l,k=1}^{N,M} J_{yy}^{l,k}(\sigma^\mu, G^{l,k}(s^\mu)u^\mu) G_{s_i}^{l,k}(s^\mu) u^\mu G_{s_j}^{l,k}(s^\mu) u^\mu \\ &+ \sum_{i,j=1}^{n_s} (\tilde{s}^i - \hat{s}^i)(\tilde{s}^j - \hat{s}^j) \sum_{l,k=1}^{N,M} J_y^{l,k}(\sigma^\mu, G^{l,k}(s^\mu)u^\mu) G_{s_i s_j}^{l,k}(s^\mu) u^\mu \end{aligned}$$

$$
\begin{aligned}
&+ 2\sum_{i=1}^{n_s}\sum_{j=1}^{n_\sigma}(\tilde{s}^i-\hat{s}^i)(\tilde{\sigma}^j-\hat{\sigma}^j)\sum_{l,k=1}^{N,M} J^{l,k}_{\sigma_j y}(\sigma^\mu, G^{l,k}(s^\mu)u^\mu)G^{l,k}_{s_i}(s^\mu)u^\mu \\
&+ 2\sum_{i=1}^{n_s}(\tilde{s}^i-\hat{s}^i)\sum_{l,k=1}^{N,M} J^{l,k}_{yy}(\sigma^\mu, G^{l,k}(s^\mu)u^\mu)(G^{l,k}_{s_i}(s^\mu)u^\mu)G^{l,k}(s^\mu)(\tilde{u}-\hat{u}) \\
&+ 2\sum_{i=1}^{n_s}(\tilde{s}^i-\hat{s}^i)\sum_{l,k=1}^{N,M} J^{l,k}_{y}(\sigma^\mu, G^{l,k}(s^\mu)u^\mu)G^{l,k}_{s_i}(s^\mu)(\tilde{u}-\hat{u}) \\
&+ \sum_{i,j=1}^{n_\sigma}(\tilde{\sigma}^i-\hat{\sigma}^i)(\tilde{\sigma}^j-\hat{\sigma}^j)\sum_{l,k=1}^{N,M} J^{l,k}_{\sigma_i\sigma_j}(\sigma^\mu, G^{l,k}(s^\mu)u^\mu) \\
&+ 2\sum_{i=1}^{n_\sigma}(\tilde{\sigma}^i-\hat{\sigma}^i)\sum_{l,k=1}^{N,M} J^{l,k}_{\sigma_i y}(\sigma^\mu, G^{l,k}(s^\mu)u^\mu)G^{l,k}(s^\mu)(\tilde{u}-\hat{u}) \\
&+ \sum_{i,j=1}^{n_\varphi}(\tilde{\varphi}^i-\hat{\varphi}^i)(\tilde{\varphi}^j-\hat{\varphi}^j)\int_0^T\int_\Omega (C_{0,\varphi_i\varphi_j}(\varphi^\mu)u^\mu + d_{\varphi_i\varphi_j}(u^\mu;\varphi^\mu) - f_{\varphi_i\varphi_j}(\varphi^\mu))p dx dt \\
&+ 2\sum_{i=1}^{n_\phi}(\tilde{\varphi}^i-\hat{\varphi}^i)\int_0^T\int_\Omega (C_{0,\varphi_i}(\varphi^\mu)(\tilde{u}-\hat{u}) + d_{\varphi_i u}(u^\mu;\varphi^\mu)(\tilde{u}-\hat{u}))p dx dt \\
&+ \sum_{l,k=1}^{N,M} J^{l,k}_{yy}(\sigma, G^{l,k}(s^\mu)(u^\mu))(G^{l,k}(s^\mu)(\tilde{u}-\hat{u}))^2 \\
&+ \int_0^T\int_\Omega d_{uu}(u^\mu;\varphi^\mu)(\tilde{u}-\hat{u})^2 p dx dt\Big) d\mu
\end{aligned}
\tag{3.10}
$$

under the assumption that the differentiation order of d with respect to u and φ is interchangeable. This representation is based on the Hessian matrix for the problem with semi-linear PDE constraints, which also contains elements that consider the direction of the PDE constraint, i.e. $\tilde{u}-\hat{u}$. For the uncertainty approximation with respect to θ as introduced in Section 2.2 the Hessian matrix representation is needed only with respect to the parameter direction, i.e.

$$
J(\tilde{s},\tilde{\sigma},\tilde{\varphi},\tilde{u}) - J(\hat{s},\hat{\sigma},\hat{\varphi},\hat{u}) = \begin{pmatrix} \tilde{s}-\hat{s} \\ \tilde{\sigma}-\hat{\sigma} \\ \tilde{\varphi}-\hat{\varphi} \end{pmatrix}^T H \begin{pmatrix} \tilde{s}-\hat{s} \\ \tilde{\sigma}-\hat{\sigma} \\ \tilde{\varphi}-\hat{\varphi} \end{pmatrix}.
$$

After reordering the terms in (3.10) with respect to the directions and formulation of the summation as an vector-matrix product we obtain $H = \int_0^1 H(\mu)\,d\mu$ of the form

$$
H(\mu) = \begin{pmatrix} H^{ss}(\mu) & H^{s\sigma}(\mu) & H^{s\varphi}(\mu) \\ H^{s\sigma}(\mu) & H^{\sigma\sigma}(\mu) & H^{\sigma\varphi}(\mu) \\ H^{s\varphi}(\mu) & H^{\sigma\varphi}(\mu) & H_1^{\varphi\varphi}(\mu) + H_2^{\varphi\varphi}(\mu) + H_3^{\varphi\varphi}(\mu) + H_4^{\varphi\varphi}(\mu) \end{pmatrix}. \tag{3.11}
$$

We can directly give the parts that are independent of $\tilde{u} - \hat{u}$ as

$$
\begin{aligned}
H^{ss}(\mu) &= \left\{ \sum_{l,k=1}^{N,M} J^{l,k}_{yy}(\sigma^\mu, G^{l,k}(s^\mu)u^\mu)(G^{l,k}_{s_i}(s^\mu)u^\mu, G^{l,k}_{s_j}(s^\mu)u^\mu) + J^{l,k}_{y}(\sigma^\mu, G^{l,k}(s^\mu)u^\mu)G^{l,k}_{s_i s_j}(s^\mu)u^\mu \right\}_{i,j} \\
H^{s\sigma}(\mu) &= \left\{ \sum_{l,k=1}^{N,M} J^{l,k}_{\sigma_j y}(\sigma^\mu, G^{l,k}(s^\mu)u^\mu)G^{l,k}_{s_i}(s^\mu)u^\mu \right\}_{i,j} \\
H^{\sigma\sigma}(\mu) &= \left\{ \sum_{l,k=1}^{N,M} J^{l,k}_{\sigma_i \sigma_j}(\sigma^\mu, G^{l,k}(s^\mu)u^\mu) \right\}_{i,j} \\
H_1^{\varphi\varphi}(\mu) &= \left\{ \int_0^T \int_\Omega (C_{0,\varphi_i\varphi_j}(\varphi^\mu)u^\mu + d_{\varphi_i\varphi_j}(u^\mu;\varphi^\mu) - f_{\varphi_i\varphi_j}(\varphi^\mu))p dx dt \right\}_{i,j}
\end{aligned}
\tag{3.12}
$$

To obtain the other parts of H we replace the difference $u^\tau = \tilde{u} - \hat{u}$. As a first step we consider the PDEs satisfied by $(\tilde{\varphi}, \tilde{u})$ and $(\hat{\varphi}, \hat{u})$ to find the PDE for u^τ. They are admissible and full fill the PDE constraint, i.e.

$$
\begin{aligned}
&\tilde{u}_t + C_0(\tilde{\varphi})\tilde{u} + d(\tilde{u};\tilde{\varphi}) - f(\tilde{\varphi}) = 0,\ \tilde{u}(0) = u_0(\tilde{\varphi}) \\
&\hat{u}_t + \underbrace{C_0(\hat{\varphi})\hat{u} + d(\hat{u};\hat{\varphi}) - f(\hat{\varphi})}_{=C(\hat{\varphi},\hat{u})} = 0,\ \hat{u}(0) = u_0(\hat{\varphi}).
\end{aligned}
$$

By subtraction we have

$$
u^\tau_t + C(\tilde{\varphi},\tilde{u}) - C(\hat{\varphi},\hat{u}) = 0,\ u^\tau(0) = u_0(\tilde{\varphi}) - u_0(\hat{u}).
$$

Hence for the difference $(\varphi^{i,\tau}, u^\tau) = (\tilde{\varphi}^i - \hat{\varphi}^i, \tilde{u} - \hat{u})$ with the first order Taylor expansion of the differential operator $C(\tilde{\varphi},\tilde{u}) - C(\hat{\varphi},\hat{u}) = \int_0^1 \nabla_{\varphi_i,u} C(\varphi^\mu, u^\mu)(\varphi^\tau, u^\tau) d\mu$ we have

$$
\begin{aligned}
u^\tau_t + \int_0^1 C_0(\varphi^\mu) + d_u(\varphi^\mu, u^\mu) d\mu\, u^\tau &= -\int_0^1 C_{0,\varphi_i}(\varphi^\mu)u^\mu + d_{\varphi_i}(\varphi^\mu, u^\mu) - f_{\varphi_i}(\varphi^\mu) d\mu\, \varphi^{i,\tau} \\
u^\tau(0) &= u_0(\tilde{\varphi}) - u_0(\hat{\varphi}).
\end{aligned}
$$

Note that this is a linear PDE in u^τ and we define the corresponding solution operator as

$$
\begin{aligned}
&S^{\varphi,u,\psi,v} : L^2(0,T;(H^1(\Omega))^*(\Omega)) \to W(0,T), \\
&w = S^{\varphi,u,\psi,v} g \text{ solves } \begin{cases} w_t + \int_0^1 C_0(\varphi + \mu\psi) + d_u(\varphi + \mu\psi, u + \mu v) d\mu\, w = g \\ w(0) = 0 \\ \frac{\partial w}{\partial \nu_C} = 0 \end{cases}
\end{aligned}
\tag{3.13}
$$

so that we can write

$$
u^\tau = -S^{\hat{\varphi},\hat{u},\varphi^\tau,u^\tau} \underbrace{\int_0^1 C_{0,\varphi_i}(\varphi^\mu)u^\mu + d_{\varphi_i}(\varphi^\mu, u^\mu) - f_{\varphi_i}(\varphi^\mu) d\mu}_{:=v_i}\ \varphi^\tau.
$$

This gives us the means to replace u^τ in (3.10) by the solution operator, which linearly depends on $\varphi^\tau = \tilde{\varphi} - \hat{\varphi}$. So for the mixed terms in H we obtain

$$\begin{aligned} H^{s\varphi}(\mu) = -\Bigg\{ &\sum_{l,k=1}^{N,M} J_{yy}^{l,k}(\sigma^\mu, G^{l,k}(s^\mu)u^\mu)(G_{s_i}^{l,k}(s^\mu)u^\mu)G^{l,k}(s^\mu)S^{\hat{\varphi},\hat{u},\varphi^\tau,u^\tau}v_j \\ &+ \sum_{l,k=1}^{N,M} J_{y}^{l,k}(\sigma^\mu, G^{l,k}(s^\mu)u^\mu)G_{s_i}^{l,k}(s^\mu)S^{\hat{\varphi},\hat{u},\varphi^\tau,u^\tau}v_j \Bigg\}_{i,j} \\ H^{\sigma\varphi}(\mu) = -\Bigg\{ &\sum_{l,k=1}^{N,M} J_{\sigma_i y}^{l,k}(\sigma^\mu, G^{l,k}(s^\mu)u^\mu)G^{l,k}(s^\mu)S^{\hat{\varphi},\hat{u},\varphi^\tau,u^\tau}v_j \Bigg\}_{i,j} . \end{aligned} \tag{3.14}$$

Finally for the parts with the second order derivatives with respect to the kinetic parameters φ we have

$$\begin{aligned} H_2^{\varphi\varphi}(\mu) &= \Bigg\{ \sum_{l,k=1}^{N,M} J_{yy}^{l,k}(\sigma, G^{l,k}(s^\mu)(u^\mu))(G^{l,k}(s^\mu)S^{\hat{\varphi},\hat{u},\varphi^\tau,u^\tau}v_i)(G^{l,k}(s^\mu)S^{\hat{\varphi},\hat{u},\varphi^\tau,u^\tau}v_j) \Bigg\}_{i,j} \\ H_3^{\varphi\varphi}(\mu) &= \left\{ \int_0^T \int_\Omega (C_{0,\varphi_i}(\varphi^\mu) + d_{\varphi_i u}(u^\mu;\varphi^\mu))S^{\hat{\varphi},\hat{u},\varphi^\tau,u^\tau}v_j p dx dt \right\}_{i,j} \\ H_4^{\varphi\varphi}(\mu) &= \left\{ \int_0^T \int_\Omega d_{uu}(u^\mu;\varphi^\mu)(S^{\hat{\varphi},\hat{u},\varphi^\tau,u^\tau}v_i)(S^{\hat{\varphi},\hat{u},\varphi^\tau,u^\tau}v_j) p dx dt \right\}_{i,j} . \end{aligned}$$

Necessary optimality condition For the deduced Hessian matrix H we investigate the necessary optimality condition (see Tröltzsch (2009, page 220)). If the objective function components are strictly convex $J_{yy}^{l,k} \geq \gamma^{l,k} > 0$ then $H_2(\mu)$ is positive semi-definite

$$x^T H_2(\mu) x \geq \sum_{l,k=1}^{M,N} \gamma^{l,k} \Big(G^{l,k} \sum_{i=1}^{n_\varphi} u_i x_i \Big)^2$$

with null space

$$\mathfrak{N}(H_2) = \bigcap_{l,k=1}^{M,N} \mathfrak{N}\Big(\Big\{ G^{l,k} u^i \Big\}_{i=1,\dots,n_\varphi} \Big)$$

where

$$u^i = \int_0^1 S^{\hat{\varphi},\hat{u},\varphi^\tau,u^\tau}(C_{0,\varphi_i}(\varphi^\mu)u^\mu + d_{\varphi_i}(\varphi^\mu,u^\mu) - f_{\varphi_i}(\varphi^\mu)) d\mu \,.$$

It follows that the identifiability of our problem, i.e. the existence of unique solutions, directly depends on the observation operator. The terms involving H_1, H_3, H_4 will be small if the residuals $J_y^{l,k}(G^{l,k}u)G^{l,k}$ are small, since they contain the adjoint state p, which solves a PDE with the sum of these residuals as right hand side

$$p_t - C_u(\hat{\varphi},\hat{u})^* p = \sum_{l,k=1}^{N,M} J_y^{l,k}(G^{l,k}(\hat{s})\hat{u})G^{l,k}(\hat{s}) \,, \quad p(T) = 0 \tag{3.15}$$

cf. the next to last component in (3.8). If the exact data are attainable for an optimal choice of the parameters (in the sense that the residuals vanish for exact data), then these terms will be of the order of the measurements noise. In case of first order mass action kinetics, C is linear with respect to φ and hence H_1 vanishes.

3.2.2. Integrated noise model method

We considered the objective function given by the integrated noise model method (3.3) and (3.4). For this Likelihood functions the objective function is given as

$$J(\sigma, s, \varphi, u) = \sum_{l,k=1}^{M,N} J_1^{l,k}(\sigma, G^{l,k}(s)u) + J_2^{l,k}(\sigma)$$

where $J_1 = -\log(L_1)$ and $J_2 = -\log(L_2)$. It is clear that for both components of the objective function of the integrated noise model method denoted by $H_{int.}(\mu)$ changes only with respect to σ compared to the Hessian (3.11). Hence we have

$$H_{int.}^{mn}(\mu) = H^{m,n}(\mu)$$

for $m, n \in \{s, \varphi\}$ and $H^{m,n}(\mu)$ are the submatrices calculated above with $J = w_1 J_1$. Furthermore as J_2 is independent of s and φ this holds also for $n = \sigma$ and $m \in \{s, \varphi\}$. Hence we have to consider only the second derivative of J with respect to σ. We have

$$H_{int.}^{\sigma\sigma} = \left\{ \sum_{l,k=1}^{M,N} w_1 J_{1\sigma_i\sigma_j}^{l,k} + w_2 J_{2\sigma_i\sigma_j}^{l,k} \right\}_{i,j}$$

and all the considerations made in the previous sections also hold for the integrated noise model approach. Up to this point we derived the Hessian matrix for the given spatio-temporal model, the observation operator and the imaging specific integrated noise model likelihood function as stated in Problem 3.1. The last point left to address is the comparison between reduced and full formulation of the optimization problem.

3.2.3. Full versus reduced formulation

In the following we compare the formulation (3.11) of the Hessian with the formulation we obtained when we considered the reduced optimization problem as it is commonly done for ODEs (see Section 2.3.2). By eliminating the state constraint we obtain the reduced formulation of (3.5) as

$$\min_{\theta=(\sigma,s,\varphi)\in\Theta} j(\theta) \qquad \text{with } j(\theta) = J(\sigma, G(s)\xi(\varphi)) \tag{3.16}$$

using the parameter-to-state map

$$\xi : \mathbb{R}^{n_\varphi} \to W(I),$$
$$u = \xi(\varphi) \text{ solves } \begin{cases} u_t + C_0(\varphi)u + d(u;\varphi) = f(\varphi) \\ u(0) = u_0(\varphi) \\ \frac{\partial u}{\partial \nu_C} = 0. \end{cases}$$

With stationarity at the MLE $\nabla_\theta j(\hat{\sigma}, \hat{s}, \hat{\varphi}) = 0$ we obtain, for any $\tilde{\theta} = (\tilde{\sigma}, \tilde{s}, \tilde{\varphi}) \in \Theta$,

$$j(\tilde{\sigma}, \tilde{s}, \tilde{\varphi}) - j(\hat{\sigma}, \hat{s}, \hat{\varphi}) = \begin{pmatrix} \tilde{s} - \hat{s} \\ \tilde{\sigma} - \hat{\sigma} \\ \tilde{\varphi} - \hat{\varphi} \end{pmatrix}^T \int_0^1 \nabla_\theta^2 j(s^\mu, \sigma^\mu, \varphi^\mu)\, d\mu \begin{pmatrix} \tilde{s} - \hat{s} \\ \tilde{\sigma} - \hat{\sigma} \\ \tilde{\varphi} - \hat{\varphi} \end{pmatrix} \tag{3.17}$$

where obviously

$$\nabla_\theta^2 j^{m,n}(s^\mu, \sigma^\mu, \varphi^\mu) = H^{m,n}(\mu)$$

for $m, n \in \{s, \sigma\}$ and $H^{m,n}(\mu)$ as in (3.12) if we set $u^\mu = \xi(\varphi^\mu)$. For linear ξ this is exact for non-linear ξ there is an arbitrary approximation error due to the fact that $u^\mu = \xi(\hat{\varphi}) + \mu(\xi(\tilde{\varphi}) - \xi(\hat{\varphi}))$ in (3.12).

Next we consider the derivatives of j which contain mixed terms with φ as in (3.14). In both cases the derivative of u^μ with respect to φ is needed. We use the fact that the parameter-to-state map satisfies the PDE constraint

$$\begin{cases} \xi(\varphi)_t + C_0(\varphi)\xi(\varphi) + d(\xi(\varphi); \varphi) = f(\varphi) \\ \xi(\varphi)(0) = u_0(\varphi). \end{cases}$$

Hence the first derivative with respect to φ^i can be given in terms of an solution operator for this PDE. We obtain

$$\begin{aligned} &S^{\varphi,u,0,0} : L^2(0, T; (H^1(\Omega))^*) \to W(0, T), \\ &v = S^{\varphi,u,0,0} g \text{ solves } \begin{cases} v_t + (C_0(\varphi) + d_u(u; \varphi))\, v = g \\ v(0) = u_0(\varphi) \end{cases} \end{aligned}$$

so we can write

$$\xi_{\varphi_i}(\varphi) = -S^{\varphi,u,0,0} \underbrace{\left(C_{0\varphi_i}(\varphi)\xi(\varphi) + d_{\varphi_i}(\xi(\varphi); \varphi) - f_{\varphi_i}(\varphi)\right)}_{:= w_i}.$$

This gives us the means to derive

$$\begin{aligned} H_{int.}^{s\varphi} &= \left\{ \sum_{l,k=1}^{N,M} J_{yy}^{l,k}(\sigma^\mu, G^{l,k}(s^\mu)\xi(\varphi^\mu))(G_{s_i}^{l,k}(s^\mu)\xi(\varphi^\mu))G^{l,k}(s^\mu)S^{\hat{\varphi},\hat{u},0,0} w_j \right\}_{i,j} \\ H_{int.}^{\sigma\varphi} &= \left\{ \sum_{l,k=1}^{N,M} J_{\sigma_i y}^{l,k}(\sigma^\mu, G^{l,k}(s^\mu)\xi(\varphi^\mu))G^{l,k}(s^\mu)S^{\hat{\varphi},\hat{u},0,0} w_j \right\}_{i,j}. \end{aligned}$$

These terms are equal to (3.14) except for the definition of u^μ as mentioned before. For the derivative with respect to φ we also need the second order derivative of $\xi(\varphi)$ with respect to φ. With the same reasoning as for the first derivative we can write

$$\begin{aligned} \xi_{\varphi_i\varphi_j}(\varphi) = -S^{\varphi,u,0,0} \Big(& C_{0\varphi_i\varphi_j}(\varphi)\xi(\varphi) + d_{\varphi_i\varphi_j}(\xi(\varphi); \varphi) - f_{\varphi_i\varphi_j} \\ & + (C_{0\varphi_i}(\varphi) + d_{\varphi_i u}(\varphi, \xi(\varphi)))\xi_{\varphi_j}(\varphi) \\ & + (C_{0\varphi_j}(\varphi) + d_{\varphi_i u}(\varphi, \xi(\varphi)))\xi_{\varphi_i}(\varphi) \\ & + d_{uu}(\varphi, \xi(\varphi))\xi(\varphi)_{\varphi_i}\xi(\varphi)_{\varphi_j} \Big) \\ =: & -S^{\varphi,u,0,0} w_{i,j}(\varphi). \end{aligned}$$

This yields

$$\begin{aligned} j_{\varphi_i\varphi_j}(s^\mu,\sigma^\mu,\varphi^\mu) = & \sum_{l,k=1}^{N,M} J_{yy}^{l,k}(\sigma^\mu, G^{l,k}(s^\mu)\xi(\varphi^\mu))(G^{l,k}(s^\mu)S^{\varphi^\mu,u^\mu,0,0}w_i)(G^{l,k}(s^\mu)S^{\varphi^\mu,u^\mu,0,0}w_j) \\ & + \sum_{l,k=1}^{M,N} J_{y}^{l,k}(\sigma^\mu, G^{l,k}(s^\mu)\xi(\varphi^\mu))G^{l,k}(s^\mu)\xi_{\varphi_i\varphi_j}(\varphi^\mu) \\ = & \sum_{l,k=1}^{N,M} J_{yy}^{l,k}(\sigma^\mu, G^{l,k}(s^\mu)\xi(\varphi^\mu))(G^{l,k}(s^\mu)S^{\varphi^\mu,u^\mu,0,0}w_i)(G^{l,k}(s^\mu)S^{\varphi^\mu,u^\mu,0,0}w_j) \\ & - \int_0^T w_{i,j}(\varphi^\mu)p^\mu dt\,. \end{aligned}$$

where, similarly to the next to last equation in (3.8), we have defined p^μ as the weak solution to the dual problem

$$\begin{cases} p_t^\mu - C_u(\varphi^\mu,u^\mu)^* p^\mu = \sum_{l,k=1}^{N,M} J_y^{l,k}(\sigma, G^{l,k}(s^\mu)\xi(\varphi^\mu))G^{l,k}(s^\mu) \\ p^\mu(T) = 0. \end{cases}$$

So up to the distinction in the definition of u^μ as discussed above, the two difference representations coincide. In other words, the Hessian at the solution (corresponding to formally setting $\mu, \varphi^\tau, u^\tau = 0$ instead of integration over the interval $(0,1)$) of course has to be the same if reduction is done after derivation from the PDE constrained problem and if the reduced formulation is differentiated.

We derived the Hessian matrix based difference representation for the semi-linear model including the observation operator and the imaging specific likelihood function to address Problem 3.1. Subsequently we compared the obtained matrix to the representation obtained by the reduced problem formulation and found that in case of non-linearity in the parameter-to-state map they do not coincide. They differ with respect to the definition of u^μ. In case of the constrained optimization $u^\mu = \xi(\hat{\varphi}) + \mu(\xi(\tilde{\varphi}) - \xi(\hat{\varphi}))$, i.e it is given as a direction in the function space. For the reduced formulation $u^\mu = \xi(\hat{\varphi} + \mu(\tilde{\varphi} - \hat{\varphi}))$, i.e. it is given as a direction in the parameter space.

3.3. Profile Likelihood

The previously derived Hessian matrix can be used for confidence interval calculation as in (2.10). This local approximation, however, is not reliable for non-linear problems or global uncertainty bounds. The PL as introduced in Section 2.2.3 is a tool to quantify the uncertainty of the MLE and to determine global uncertainty bounds. We give the formulation in terms of the reduced problem (3.16) here (Hock *et al.*, 2013; Lockley *et al.*, 2015). The MLE is calculated for a one-dimensional subspace of the parameter space. In our case we calculate the profile likelihood for the unknown parameters of interest, i.e. θ_i. For parameter θ_i, $\mathrm{pl}(\theta_i)$ is computed by the re-optimization of all parameters $\theta_j \neq \theta_i$ along the profile of parameter θ_i (Murphy

et al., 2000):

$$\mathrm{pl}(\theta_i) = \max_{\theta_{j\neq i}} l(\theta) = \exp\left(-\min_{\theta_{j\neq i}} j(\theta)\right). \tag{3.18}$$

The minimization must full fill the same constraints as in (3.16). This can be repeated for all parameters θ_i, $i = 1, \ldots, n_\theta$, and allows the evaluation of the likelihood ratio $R(\theta_i) = \mathrm{pl}_c(\theta_i)/l(\hat{\theta})$ for the individual parameters. Based on the likelihood ratio $R(\theta_i)$ we determine globally valid confidence intervals for the parameter θ_i,

$$CI_\alpha(\hat{\theta}_i) = \left\{\theta_i \,\middle|\, R(\theta_i) \geq \exp\left(-\frac{\delta_\alpha}{2}\right)\right\},$$

with confidence level α and the corresponding likelihood ratio threshold $\delta_\alpha = \chi^2(\alpha, 1)$ (Murphy *et al.*, 2000). In the following we want to analyse the confidence intervals obtained with the Hessian approximation and the PL for the previously introduced Example 2.1 with simulated data.

3.3.1. Mathematical model

We considered the Example 2.1 for $\varphi = (D, \alpha, J)^T$, $\Omega = [-7, 7][\mu\mathrm{m}]$ and $I = (0, 100)[\mathrm{s}]$ with the model equation

$$\begin{cases} u_t - D\triangle u + \alpha u^2 = \frac{J}{2\pi\rho}\exp\left(-\frac{x^2}{\rho^2}\right) & \text{in } \Omega \times I \\ \frac{\partial u}{\partial \nu_C} = 0 & \text{on } \partial\Omega \times I \\ u(0) = u_0 & \text{in } \Omega \times \{t = 0\}. \end{cases} \tag{3.19}$$

The steady state of the model

$$u^\infty(x) = A\frac{x_0^2}{(x + x_0)^2}$$

with

$$A = \sqrt[3]{\frac{3}{2\alpha D}\left(\frac{J}{2\pi\rho}\exp\left(-\frac{x^2}{\rho^2}\right)\right)^{2/3}} \text{ and } x_0 = \sqrt{\frac{6D}{\alpha A}}.$$

Artificial data To obtain data close to the real application situation for which the model was developed, we produced simulated data similar to the data given in Section 5.1.2. We considered three data sets: a steady state profile, absolute molecule numbers and a FRAP measurement. Therefore, we calculated the steady state for the parameter values given in Figure 3.2A), added additive normal distributed noise with $\varepsilon_1 \sim \mathcal{N}(0, 0.1)$ and scaled the profile to one. Additionally, we obtained the absolute molecule number by integration of the profile. For the FRAP measurement we used the steady state value as initial condition but set all values in the region $Q = [-2.7\mu\mathrm{m}, 2.7\mu\mathrm{m}]$ to zero. Consequently, we simulated the model until $T = 30$ and calculated at $t = \{1, 3, 5, 8, 10, 12, 18, 20, 30\}[\mathrm{s}]$ the overall intensity of the profile scaled by the same constant as the steady state profile. The artificial data is depicted in Figure 3.2B)-C).

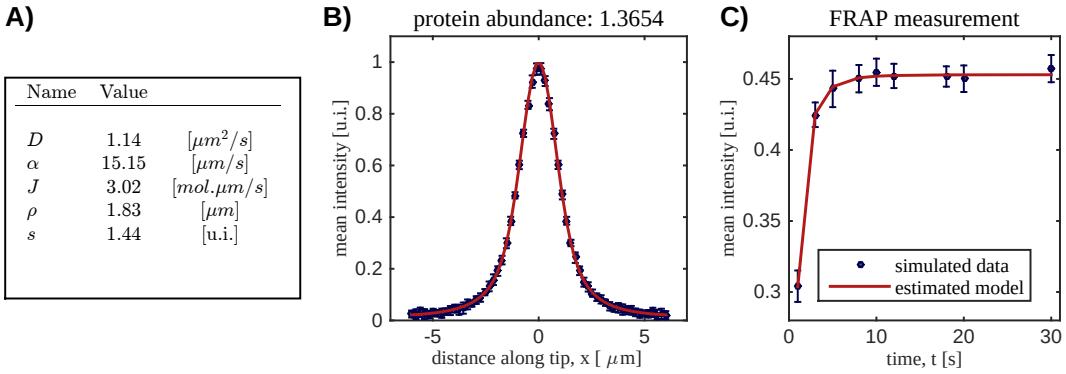


Figure 3.2.: ***artificial data for example 2.1*** *A) Parameters used to perform the data simulation. B) Mean intensity profile. The steady state of the model was calculated and additive normal noise $\varepsilon \sim \mathcal{N}(0, 0.1)$ was added and the profile scaled to one. From those measurements the mean and standard deviation was calculated. Furthermore the mean protein abundance and standard deviation was calculated based on those replica. C) FRAP measurements after full tip bleaching. Again the model was simulated and to each replica additive normal noise $\varepsilon \sim \mathcal{N}(0, 0.1)$ was added and the profile scaled. The red line is the output of the model after parameter estimation.*

Before we performed profile likelihood calculation on this data we had to establish the parameter estimation process. Therefore we introduced three different observation operators g, one for each measurement. First we considered the mean intensity profiles. The profiles are the steady state concentrations scaled to one. Hence the first observation operator is given as

$$G_1(u^\infty) = \frac{u^\infty(x;\varphi)}{\max\limits_x u^\infty(x;\varphi)}$$

here u^∞ denotes the stationary limit of the model. Measurements were taken at 60 equally spaced spatial points $x_k \in [-7, 7][\mu m]$ and $k = 1, \ldots, 60$. We assumed that the measurement error is normally distributed with standard deviation obtained from the measurements. Hence the objective function is given as

$$J_1(\theta) = \frac{1}{2}\sum_{k=1}^{60} \log(2\pi\sigma_{1,k}^2) + \frac{(\overline{y}_{1,k} - G_{1,k})^2}{\sigma_{1,k}^2}.$$

For the total protein abundance in steady state we consider the integral over the tip region and obtain the observation operator.

$$G_2(u^\infty) = \int_{-7}^{7} u^\infty(x;\varphi)dx.$$

Again we assumed that the measurement error is normally distributed with standard deviation obtained from the measurements. Hence the objective function is given as

$$J_2(\theta) = \frac{1}{2}\log(2\pi\sigma_2^2) + \frac{(\overline{y}_2 - G_2)^2}{\sigma_2^2}.$$

Finally for the FRAP measurements we had to consider a scaled version of the dynamic time dependent models. Hence we obtained

$$G_3(p, p^\infty; s) = \int_Q su(x, t; \varphi)dx.$$

Measurements with $t_k \in [0, 30][\text{s}]$ with $k = 1, \ldots, 9$ were taken. Again we assumed that the measurement error is normally distributed with standard deviation obtained from the measurements. Hence for the data points t_k the objective function is given as

$$J_3(\theta) = \frac{1}{2}\sum_{k=1}^{10} \log(2\pi\sigma_{3,k}^2) + \frac{(\overline{y}_{3,k} - G_{3,k})^2}{\sigma_{3,k}^2}.$$

For the estimation process we consider different combinations of the three measurements with the objective function

$$J(\theta, y) = \sum_{i=1}^{M} J_i(\theta).$$

The results of the parameter estimation performed with this objective function are shown in Figure 3.2.

Implementation The model was implemented in MATLAB with a finite difference scheme using a step length of $h = 0.01\mu$m in the spatial variable and then the resulting ODE was solved with the Sundials Toolbox. 100 multi-starts with a latin hypercube sampling of the parameter space were performed with the PESTO Toolbox, developed in the group of data-driven dynamical modelling, to obtain the MLE and the corresponding parameter $\hat{\theta}$. The first and second order sensitivity equations were also implemented in MATLAB using the toolbox AMICI (*http://amici-developer.github.io/AMICI/*). Based on the second order sensitivity equations the Hessian matrix was calculated. The optimization based profile calculation was performed with the PESTO Toolbox.

3.3.2. Uncertainty analysis

For the uncertainty analysis of all five parameters we used three different data sets: the full data set (M=3), the steady state data set (M=2) and the steady state profile (M=1). This combination yielded identifiable parameters, practical non-identifiable parameters and structural non-identifiable parameters (see Table 3.1).

We found that for the identifiable parameters the Hessian Matrix approximated the confidence intervals well and yielded a similar interval as the PL. For the practical non-identifiable parameters the Hessian was a good approximation close to the optimum, however, it overestimated the size of the confidence interval. For the structural non-identifiable parameters the Hessian approximation gave the same result as the profile likelihood in the set parameter bounds. As a representative illustration we considered the PL and the resulting confidence intervals for the parameter α shown in Figure 3.3.

Table 3.1.: ***Parameter identifiability for Example 2.1*** *Comparison of the identifiability of the five parameters for the different data sets. The identifiability of a parameter was analyzed based on the PLs calculated with the optimization based approach..*

Parameter	M=3	M=2	M=1
D	identifiable	prac. non-identifiable	prac. non-identifiable
α	identifiable	prac. non-identifiable	struc. non-identifiable
J	identifiable	prac. non-identifiable	struc. non-identifiable
ρ	identifiable	identifiable	identifiable
s	identifiable	struc. non-identifiable	struc. non-identifiable

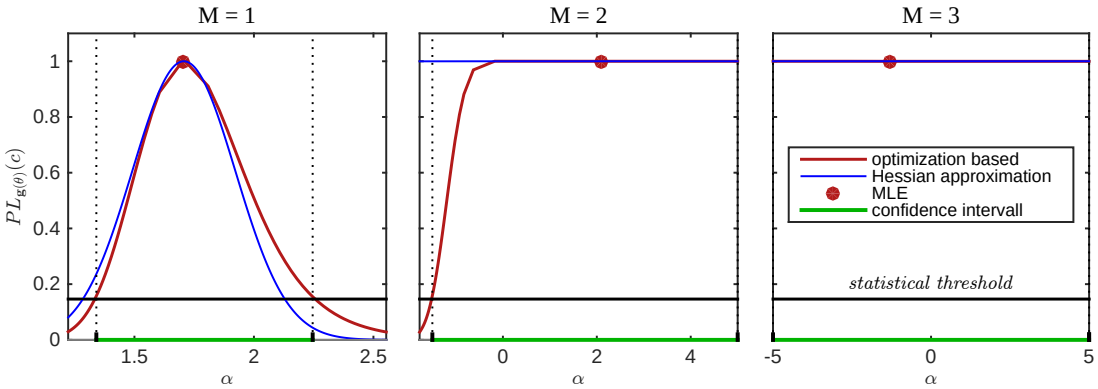


Figure 3.3.: ***Parameter identifiability analysis for Example 2.1 for parameter*** α *PL obtained by the optimization based approach for parameter* α

The introduced optimization based profile likelihood calculation addresses Problem 3.2 as it facilitates the rigorous definition of uncertainty bounds compared to local approximative methods like the approximation with the Hessian matrix. Unfortunately the calculation procedure includes a full solution of the optimization problem for a fixed number of grid points in each parameter direction. For computationally demanding problems with PDEs this method is rather insufficient. In the next chapter we introduce a more efficient profile calculation procedure based on the previously deduced Hessian formulation (3.11) and compare the uncertainty quantification for the parameters based on the Hessian approximation, the profile likelihood for the reduced system and the newly introduced efficient profile calculation methods.

4. Efficient profile likelihood calculation

In the last chapter we introduced the existing methods to calculate PLs based on the completely reduced form of the likelihood for semi-linear PDEs. It needs the repeated solutions of the optimization problem, which is very time consuming or even infeasible for large PDE systems. To circumvent this repeated optimization Chen & Jennrich (2002) introduced simulation based PL calculation. The method proved to be efficient and accurate in the case of known model solution and corresponding Hessian. In this chapter we want to extend the method to the PDE constraint case to develop a more efficient method to calculate PLs for parameter estimation problems with semi-linear PDE constraints.

First we give the key idea behind simulation based PL calculation and the problems we address in this chapter. We introduce the formulation of the method for PDE constraints in the full and reduced form. Following the theoretical considerations we discuss approximations for the PL calculation in case of unknown Hessian matrix. We conclude this chapter by an evaluation of the PL calculation performance an application examples.

Preliminary results of the evaluation were discussed in the master thesis by Frank 2013, which was co-supervised by the author of this thesis. The results of this chapter have been published recently in the paper of Boiger *et al.* (2016), which was co-authored by the author of this thesis.

4.1. Introduction and problem statement

The PL has been introduced as a constrained maximization of the likelihood function with an scalar equality constraint on the parameter of interest. In Chen & Jennrich (1996, 2002) the PL calculation is considered to be a constrained maximization of the logarithm of the likelihood. For comparison with the previously introduced method we consider it in terms of a minimization of the negative logarithm of the likelihood. The model solution is assumed to be given, i.e. the reduced parameter estimation formulation is considered. For $j = -\log(l(\theta))$ we obtain

$$\begin{aligned} \theta_c = &\underset{\theta \in \Theta}{\operatorname{argmin}}\, j(\theta) \\ &\text{s.t. } g(\theta) = c, \end{aligned} \tag{4.1}$$

in which g is the profile function. To obtain $\mathrm{pl}_c(\theta_i)$ as introduced in Section 2.2 with θ_i the i-th component of θ then $g(\theta) = \theta_i$. In comparison to the standard formulation the consideration of g also allows the consideration of non-linear parameter

combinations. For example identifiability of products of parameter components can be considered, which is often useful for scaled observations.

The major drawback of the optimization based approach is that it only implicitly uses the knowledge about parameter dependencies given by the model and therefore in each step an optimization is needed to reach the PL parameter value. Chen & Jennrich (2002) introduced a method which approximates the curve in the parameter spaces needed for the profile calculation based on the local structure of the objective function given by the Hessian of the reduced objective function (see Figure 4.1). In the following the parameter curve for any c is denoted by θ_c and we introduce the Lagrangian multiplier λ_c such that, we obtain the optimal points of (4.1) as solutions of

$$\begin{aligned} \nabla_\theta j(\theta_c) + \lambda_c \nabla_\theta g(\theta_c) &= 0 \\ g(\theta_c) &= c. \end{aligned} \tag{4.2}$$

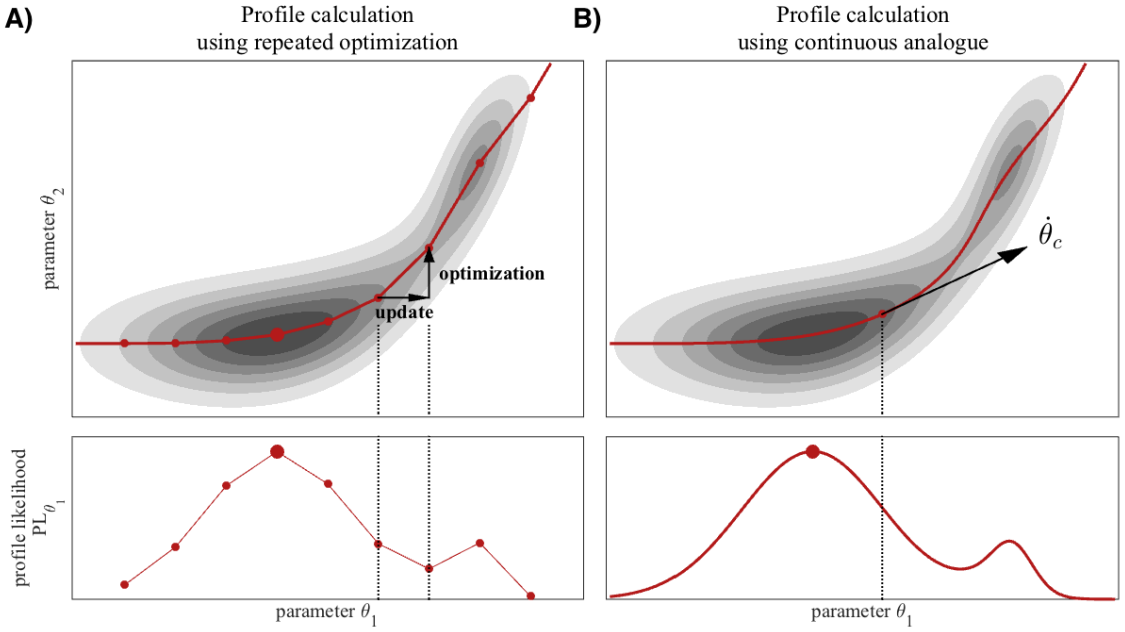


Figure 4.1.: ***Comparison of optimization and simulation based profile calculation*** *A) Optimization based PL calculation procedure. In one parameter direction a fixed step is taken and the the other parameters are optimized to obtain the point on the profile. B) Simulation based PL calculation method. The curve in parameter space is given by the direction of* $(\dot{\theta}, \dot{\lambda})^T$.

If we differentiate (4.2) with respect to c it has to satisfy the ordinary differential equation

$$\begin{pmatrix} \nabla_\theta^2 j(\theta_c) + \lambda_c \nabla_\theta^2 g(\theta_c) & \nabla_\theta g(\theta_c) \\ \nabla_\theta g^T(\theta) & 0 \end{pmatrix} \begin{pmatrix} \dot{\theta}_c \\ \dot{\lambda}_c \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{4.3}$$

where $\dot{\theta}_c$ and $\dot{\lambda}_c$ are the derivatives with respect to c. The solution of this ODE yields the parameter curve θ_c and in turn $\mathrm{pl}_c(g(\theta))$ for any c.

4.1.1. Problem statement

The simulation based PL calculation by Chen & Jennrich (2002) considers an optimization problem without model constraints. In the context of uncertainty quantification for parameter optimization problems as discussed in this work the model is always given as constraint. Hence the first problem to be addressed is the following

Problem 4.1 (Formulation of simulation based PL calculation)**.** *Formulate the simulation based PL calculation method for semi-linear PDE constraint problems.*

The simulation based PL calculation relies on the local structure of the objective function given by the Hessian matrix. In most applications the calculation of the exact Hessian matrix is infeasible and Chen & Jennrich (2002) introduced an approximative scheme. The second problem addressed approximation methods for the PDE constraint case.

Problem 4.2 (Hessian approximation for PL calculation)**.** *Introduce and evaluate efficient methods to approximate the Hessian for the approximative, simulation based PL calculation.*

These problems will be discussed and solved through the chapter and evaluated on the previously introduced application example.

4.2. Differential equation based profile calculation

Alternatively to applying the approach (4.3) directly to the reduced form of the maximization problem we want to consider the PDE constrained setting of Section 3.1.3 and derive a system of ODEs for the profiles as functions of c. Therefore we consider (4.1) for the PDE constrained problem and obtain θ_c as optimal point of

$$\begin{aligned} &\min_{s\in\mathbb{R}^{n_s},\sigma\in\mathbb{R}^{n_\sigma},\varphi\in\mathbb{R}^{n_\varphi},u\in W(I)\cap L^\infty} J(\sigma, G(s)u) \\ &\text{s.t.} \quad \begin{cases} u_t + C(u;\varphi) = f(\varphi) & \text{in } \Omega\times I \\ \frac{\partial u}{\partial \nu} = 0 & \text{on } \partial\Omega\times I \\ u(0) = u_0(\varphi) & \text{on } \Omega\times\{t=0\} \\ g(\theta) = c. \end{cases} \end{aligned} \tag{4.4}$$

with $\theta = (\sigma, s, \varphi)$. We will see in the following that the distinction between model parameters φ, observable parameters s and noise parameters σ leads to a simplified notation for the introduced Hessian matrix. The Lagrange function $\tilde{\mathcal{L}}$ for the resulting constrained minimization problem can be written in terms of the Lagrangian $\mathcal{L}$ of the parameter optimization problem introduced in Section 3.2, i.e.

$$\tilde{\mathcal{L}}(\theta, u, p, \lambda_c) = \mathcal{L}(\theta, u, p) + \lambda_c(g(\theta) - c) \tag{4.5}$$

Hence the first order optimality conditions read as

$$\begin{aligned} \nabla_\theta \mathcal{L}(\theta_c, u, p) + \lambda_c \nabla_\theta g(\theta_c) &= 0 \\ \nabla_u \mathcal{L}(\theta_c, u, p) &= 0 \\ \nabla_p \mathcal{L}(\theta_c, u, p) &= 0 \\ g(\theta_c) &= c. \end{aligned}$$

Differentiating with respect to c we obtain the following system for the evolution of $(\theta_c, u, p, \lambda_c)$

$$B \begin{pmatrix} \dot{\theta}_c \\ \dot{u} \\ \dot{p} \\ \dot{\lambda}_c \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

with

$$B = \begin{pmatrix} \nabla_\theta^2 \mathcal{L}(\theta_c, u, p) + \lambda_c \nabla_\theta^2 g(\theta_c) & \nabla_\theta \nabla_u \mathcal{L}(\theta_c, u, p) & \nabla_\theta \nabla_p \mathcal{L}(\theta_c, u, p) & \nabla_\theta g(\theta_c) \\ \nabla_u \nabla_\theta \mathcal{L}(\theta_c, u, p) & \nabla_u^2 \mathcal{L}(\theta_c, u, p) & \nabla_u \nabla_p \mathcal{L}(\theta_c, u, p) & 0 \\ \nabla_p \nabla_\theta \mathcal{L}(\theta_c, u, p) & \nabla_p \nabla_u \mathcal{L}(\theta_c, u, p) & 0 & 0 \\ \nabla_\theta g^T(\theta_c) & 0 & 0 & 0 \end{pmatrix}.$$

For singular B this is a differential algebraic equation of order one or higher. If B is regular we obtain an ordinary differential equation by inversion of B. Note that for linear parameter functions g the Hessian of g equals to zero and the upper three times three part of the matrix correspond to the Hessian deduced in Section 3.2.1. This formulation includes the derivative of the PDE solution u and the adjoint state p with respect to c. Both are not needed for the calculation of PL with respect to the parameter and we proceeded to consider B only in terms of θ. This can be done by consideration of H as defined in (3.11), i.e.

$$\begin{pmatrix} H + \lambda_c \nabla_\theta^2 g(\theta_c) & \nabla_\theta g(\theta_c) \\ \nabla_\theta g(\theta)^T & 0 \end{pmatrix} \begin{pmatrix} \dot{\theta}_c \\ \dot{\lambda}_c \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{4.6}$$

The solution of the differential equation yields the parameter curve θ_c and therefore the profile likelihood $\mathrm{PL}_c(g(\theta))$.
Following the results from Section 3.2.3 we found that the formulation in terms of the reduced and the full problem coincide up to the definition of u^μ. In case of non-linearity we saw a difference between both approaches. Hence we conclude that the PL calculation based on the reduced formulation (4.3) is in the non-linear case only an approximation to the method. Therefore, the approximation formulation for simulation based PL calculation has to be considered.

4.3. Approximation properties

For the Hessian matrix, which considers the PDE constraints, the solution of the differential equation (4.6) yields an exact likelihood profile. If the Hessian is considered for the reduced problem formulation, we find that this no longer holds for non-linear parameter-to-state maps. Furthermore, the calculation of the exact Hessian matrix is often computationally very demanding or infeasible. For the cases without Hessian matrix an approximation with a positive semi-definite matrix W can be used (Chen & Jennrich, 2002). The approximative differential equation is given as

$$\begin{pmatrix} -W(\theta_c) + \lambda_c \nabla_\theta^2 g(\theta_c) & \nabla_\theta g(\theta_c) \\ \nabla_\theta g(\theta)^T & 0 \end{pmatrix} \begin{pmatrix} \dot{\theta}_c \\ \dot{\lambda}_c \end{pmatrix} = \begin{pmatrix} -\gamma \nabla_\theta l(\theta) \\ 1 \end{pmatrix}. \tag{4.7}$$

Here $\gamma \nabla_\theta l(\theta)$ with $\gamma > 0$ is a correction term, which for increasing γ, directs the solution back to the true solution if a derivation occurs. For positive, semi-definite

W it can be shown that the approximation converges to the PL, if γ is chosen big enough (Chen & Jennrich, 2002).

4.3.1. Hessian approximation schemes

Then we address the question how to proceed when the Hessian is approximated. The approximation is needed if the calculation of the Hessian is infeasible as it needs the second order sensitivity equations of the model. Chen & Jennrich (2002) proposed the identity matrix as a good choice for regression models. In the context of PDEs, however, we found that this approximation is not suitable. Our studies showed that it completely fails in the presence of non-identifiable parameters (see Table 4.1). For the identity matrix the correction term has to be chosen very large, which results in a computationally demanding, stiff ODE system. In the following we consider four approximation schemes:

- the identity matrix
- the exact Hessian for s and σ and the Identity matrix for φ
- the exact Hessian for s and σ and $H_2^{\varphi\varphi}$ for φ
- the Hessian H and $H_2^{\varphi\varphi}$ for φ.

In the following we discuss the corresponding matrices W in detail.

For the approximation with the identity matrix we obtained

$$W_1(\theta) = \begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix}.$$

For the other approximations we recalled $H = \int_0^1 H(\mu)\, d\mu$,

$$H(\mu) = \begin{pmatrix} H^{ss}(\mu) & H^{s\sigma}(\mu) & H^{s\varphi}(\mu) \\ H^{s\sigma}(\mu) & H^{\sigma\sigma}(\mu) & H^{\sigma\varphi}(\mu) \\ H^{s\varphi}(\mu) & H^{\sigma\varphi}(\mu) & H_1^{\varphi\varphi}(\mu) + H_2^{\varphi\varphi}(\mu) + H_3^{\varphi\varphi}(\mu) + H_4^{\varphi\varphi}(\mu) \end{pmatrix}$$

with sub-matrices given as in Section 3.2.1. It becomes clear that $H^{ss}, H^{s\sigma}, H^{\sigma\sigma}$ can be calculated independently of the PDE. Especially, if the objective function and the observation operator full fill the assumptions made in the beginning of Section 3.1. Hence, the second approximation we propose is

$$W_2(\theta) = \begin{pmatrix} H^{ss}(\mu) & H^{s\sigma}(\mu) & & & \\ H^{\sigma s}(\mu) & H^{\sigma\sigma}(\mu) & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix}. \tag{4.8}$$

We expect this approximation to perform well for non-identifiable noise or scaling parameters. For strongly non-linear kinetic parameters it has the same problems as

the identity matrix.

To improve the approximation results for the kinetic parameters we considered the submatrix $H_2^{\varphi\varphi}$ as introduced in Section 3.2.1. This matrix contains the second derivative of the objective function multiplied by the product of the first-order derivatives with respect to φ. This matrix is often called Fisher information matrix (see Section 2.2.3). To calculate $H_2^{\varphi\varphi}$ we calculated v_i by forward sensitivity equations and set $S^{\varphi,u,0,0}(C_{0,\varphi}(\varphi)u - d_\varphi(\varphi,u) + f_\varphi(\varphi))$ as done in Section 3.2.3. We obtained the two approximations

$$W_3(\theta) = \begin{pmatrix} H^{ss}(\mu) & H^{s\sigma}(\mu) & 0\dots 0 \\ H^{\sigma s}(\mu) & H^{\sigma\sigma}(\mu) & 0\dots 0 \\ 0 & 0 & \\ \vdots & \vdots & H_1^{\varphi\varphi}(\mu) \\ 0 & 0 & \end{pmatrix} \tag{4.9}$$

and

$$W_4(\theta) = \begin{pmatrix} H^{ss}(\mu) & H^{s\sigma}(\mu) & H^{s\varphi}(\mu) \\ H^{\sigma s}(\mu) & H^{\sigma\sigma}(\mu) & H^{\sigma\varphi}(\mu) \\ H^{s\varphi}(\mu) & H^{\sigma\varphi}(\mu) & H_1^{\varphi\varphi}(\mu) \end{pmatrix}. \tag{4.10}$$

Following the considerations in Section 3.2.1 all matrices proposed are positive semi definite and can be used as approximations according to Chen & Jennrich (2002).

4.4. Performance evaluation

In the following we evaluated the performance of the introduced simulation based PL calculation. We considered three evaluation criteria:

- efficiency with respect to the number of PDE solutions
- accuracy compared to the optimization based profile calculation
- correctness of decision regarding identifiability.

All three points were evaluated for the exemplary parameter estimation problem introduced in Section 2.1 for simulated data introduced in Section 3.3. The performance is evaluated on all possibly occurring identifiability problems. The application on real data is discussed in Chapter 5 and Chapter 6.

4.4.1. Correctness of identifiability analysis

As the first evaluation criteria we considered the precision of the decision regarding the identifiability. This is the minimal requirement every PL calculation should fulfill as it is the key tool for practical identifiability analysis (Raue *et al.*, 2009). We used the standard optimization based process to calculate PLs for all five parameters on the full data set (M=3), the steady state data set (M=2) and only the steady state profile (M=1). This combination yielded identifiable parameters, practical non-identifiable parameters and structural non-identifiable parameters (see Table 4.1).

We subsequently performed the profile calculation with the approximation W_1-W_4 on the same data sets and compared the obtained results regarding the identifiability of the parameters.
We found that for the identifiable case ($M = 3$) all methods gave the right results, even though the profiles obtained by W_1 and W_2 were considerably narrower than those obtained by the benchmark calculation. In case of practical and structural non-identifiability, W_3 and W_4 still came to the right decisions while W_1 and W_2 completely failed. This also explains why Chen & Jennrich (2002) proposed the identity matrix as a good approximation as they only considered identifiable parameters. In the presence of possible non-identifiability, however, it is a bad choice. Also the variation of the adaption parameter γ between values of $1e-3$ up to $1e5$ did not yield other results for W_1 and W_2.

4.4.2. Accuracy of the method

As the next point we analyzed the accuracy of the obtained profile. A key aspect of uncertainty quantification is the calculation of the confidence interval bounds, which can be obtained from the profile likelihood as explained in the Background Section 2.2. Therefore, a good approximation of the profile likelihood should yield the same confidence bounds like the optimization based approach. To compare the accuracy of the calculated profiles we defined 20 points within the confidence region of each parameter. We obtained the confidence interval by the optimization based approach. For 20 points we calculated the square distance of the calculated PL to the benchmark PL. In case that the points where not inside the approximated profiles, the value was set to zero at this point (maximal distance possible). The results are displayed in Figure 4.2.

We found that for W_3 and W_4 the method performed well and always yielded errors below 1 for all measurement situations. In contrast W_1 and W_2 completely failed the test. The profiles did not resemble the shape of the Benchmark. This could be slightly tuned by choosing high values of γ but even then the profiles were much to narrow.

4.4.3. Efficiency of the method

The last aspect we analyzed was the performance with respect to the calculation efficiency. As we will apply the method to problems with computationally demanding objective function, we assumed that the objective function evaluation was the bottle neck in this setup. Therefore, we evaluated the method with respect to the number of objective function evaluations, which is proportional to the CPU time. The results are shown in Figure 4.3.
We found that for almost all parameters the number of objective function values needed lay several orders of magnitudes below the number of function evaluations needed for the optimization based method. For W_1 and W_2, however, this result is meaningless as the obtained profiles do not resemble the profiles used as benchmark.

Table 4.1.: ***Parameter identifiability for Example 2.1*** *Comparison of the results regarding identifiability of the model parameters for the different data set with respect to the approximation schemes. For identifiable parameters all approximations give the right decision for non-identifiability, however, W_1 and W_2 failed.*

M=3

	Benchmark	W_1	W_2	W_3	W_4
D	identifiable	✓	✓	✓	✓
α	identifiable	✓	✓	✓	✓
J	identifiable	✓	✓	✓	✓
ρ	identifiable	✓	✓	✓	✓
s	identifiable	✓	✓	✓	✓

M=2

	Benchmark	W_1	W_2	W_3	W_4
D	practical non-identifiable	×	×	✓	✓
α	practical non-identifiable	×	×	✓	✓
J	practical non-identifiable	×	×	✓	✓
ρ	identifiable	✓	✓	✓	✓
s	structural non-identifiable	×	×	✓	✓

M=1

	Benchmark	W_1	W_2	W_3	W_4
D	practical non-identifiable	×	×	✓	✓
α	structural non-identifiable	×	×	✓	✓
J	structural non-identifiable	×	×	✓	✓
ρ	identifiable	✓	✓	✓	✓
s	structural non-identifiable	×	×	✓	✓

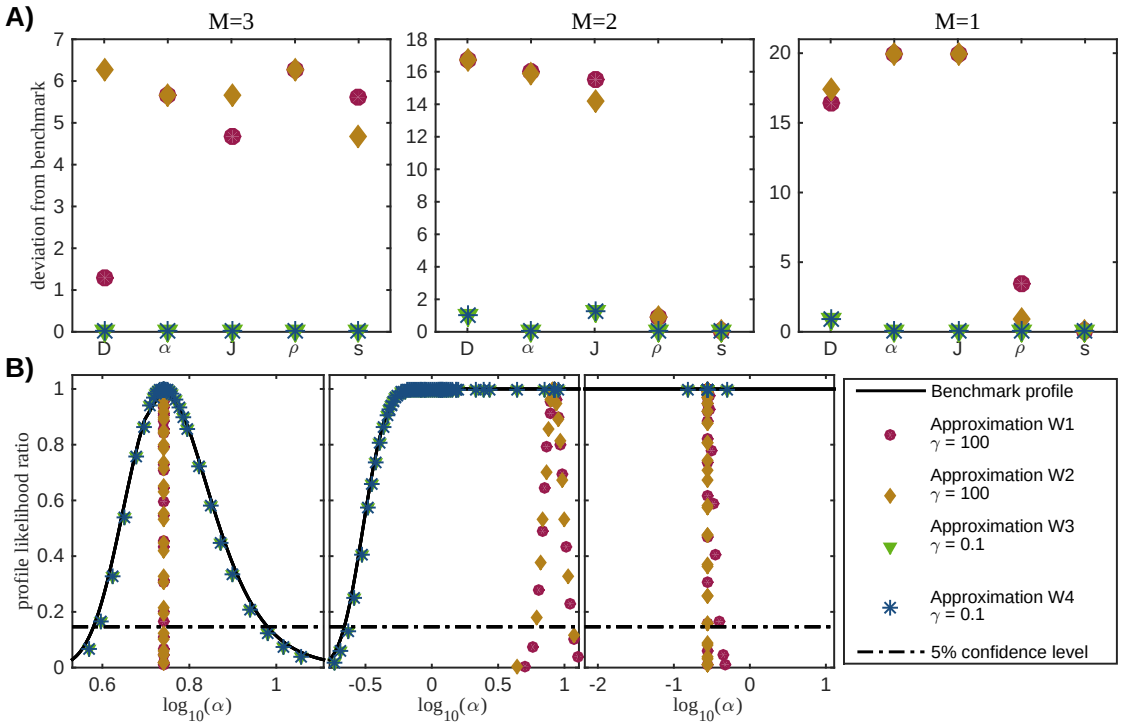
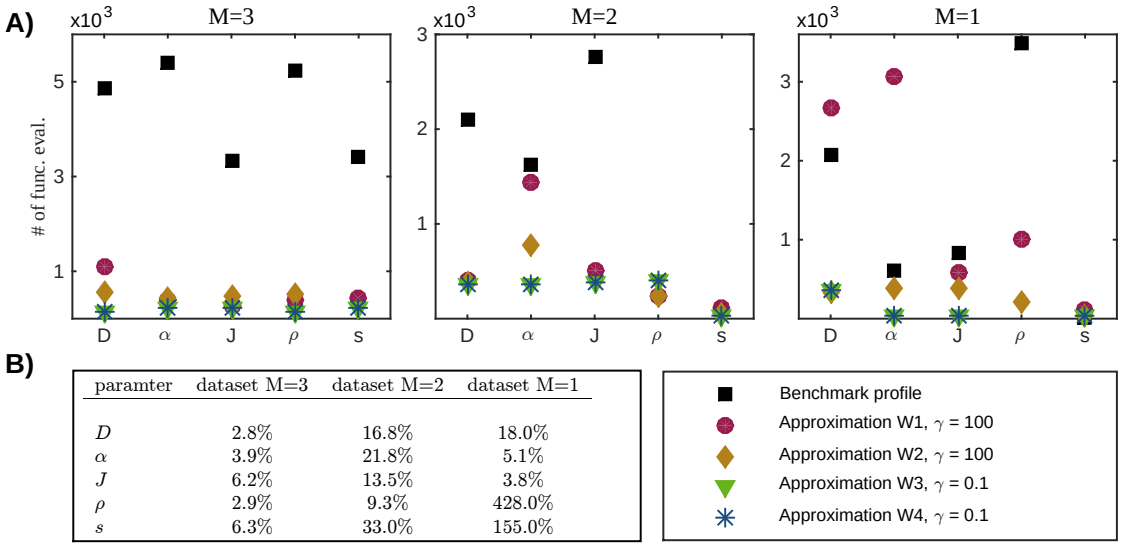


Figure 4.2.: ***Accuracy evaluation for the new method*** *A) Error with respect to the standard optimization based approach for measurement situation M=3, M=2 and M=1. B) Profiles for the parameter α obtained for the three measurement combinations. For $M = 3$ the parameter is identifiable and with the reduction of measurements the parameter becomes practical non-identifiable and the structural non-identifiable. Approximation W_3 and W_4 nicely reproduced the profiles.*

4.5. Conclusion

In the frequentist setting of parameter estimation profile likelihoods are the most reliable method for practical identifiability analysis and uncertainty quantification. Unfortunately the computation time for the standard optimization based profile calculation renders the application to semi-linear, computation heavy PDE models infeasible. We addressed this problem by a combination of the Hessian deduced in Chapter 3 with the ODE formulation of the profile calculation problem, which yielded an simulation based PL calculation method which can be widely applied to inverse problems with semi-linear PDEs. To reduce computation time we developed and evaluated approximations of the original method, which do not need the calculation of the exact Hessian. This enabled the profile likelihood analysis of estimation problems with computational expensive models. For the considered application example with a relatively low computation time the optimization based profile calculation took several hours in contrast to minutes used by the new method.

For the simulation based PL calculation we considered three key aspects: the ODE formulation of the constrained optimization problem, the approximation schemes and the evaluation on an application example. We found that with the Hessian



paramter	dataset M=3	dataset M=2	dataset M=1
D	2.8%	16.8%	18.0%
α	3.9%	21.8%	5.1%
J	6.2%	13.5%	3.8%
ρ	2.9%	9.3%	428.0%
s	6.3%	33.0%	155.0%

Figure 4.3.: ***Efficiency evaluation for the new method*** *A) Number of function evaluations for the standard approach, the new method and the considered approximations with data set M= 3, data set M = 2 and data set M = 1. We observed that in nearly all situations the approximations need less objective function evaluations than the standard approach. B) Table with percentage of function evaluations used by the approximation W4 compared to the optimization based method. We observed that in nearly all cases less than 20% of the evaluation needed for the optimization based approach are required.*

matrix introduced in Chapter 3 the constrained optimization problem considered during the profile calculation could be formulated as an ODE, especially in the case of identifiable parameters. For most applications, however, to calculate Hessian is computationally demanding and not all parameters might be identifiable. In this case an approximation has to be considered and we introduced several possible approximations of the Hessian matrix based on the Fisher Information matrix. Finally the evaluation of the method on an application motivated 1D semi-linear PDE parameter estimation problem showed its impressive improvement of calculation speed in comparison to the standard optimization based approach. Furthermore, it became apparent that the best approximation choice is the Fisher Information matrix regarding profile accuracy and correctness of decisions regarding the identifiability of parameters.

Based on the numerical results of this chapter we can conclude that the newly developed efficient profile calculation method outperforms the standard optimization based approach by regarding the number of function evaluations. Despite this increase in calculation speed the accuracy and the correctness of decisions regarding parameter identifiability is comparable to that of the standard approach. From an application point of view we can conclude that the full Hessian is not needed to obtain reliable results but can be approximated with the often easy to obtain Fisher Information matrix. The Fisher information matrix showed as good a performance as an approximation to the Hessian.

Recapitulating the research questions posed in Section 4.1.1 the results obtained in this chapter answer the questions. Based on the efficient profile calculation for ODEs we introduced the method for semi-linear PDEs. Furthermore, we proposed Hessian matrix approximations and evaluated them on an application motivated example, which showed an impressive improvement of calculation speed compared to the standard approach.

From a numerical point of view the efficient implementation of the Hessian matrix remains open. Especially, for more complex models this might be computationally very demanding and hamper the profile calculation. Another open question is the formulation of the problem as ODE is the matrix B is not invertible, i.e. in case of non-identifiable parameters. In this case the formulation (4.6) becomes an DAE instead of an ODE and the DAE has to be solved numerically. In this study we used the approximation formulation (4.7) and the pseudo inverse in case of non-identifiable parameters, which worked acceptable well.

In addition to the improvements of computation time and the consideration of the DAE equations future studies should be addressed to find an optimal γ value. Up to this point γ is chosen by hand but a more structural choice should be possible. For large values the ODE becomes increasingly stiff resulting in increased computation time but small values result in inaccurate profiles. The optimal choice should be a trade off between speed and accuracy. Furthermore, an adaptive update scheme for γ could be possible depending on the accuracy of the Hessian matrix approximation. Parameter estimation and uncertainty quantification for semi-linear PDE models, e.g in the field of image based systems biology, strongly relies on efficient methods for semi-linear, computational demanding PDE problems. The introduced efficient profile calculation can address this demand. In the following chapters we will apply the method during parameter estimation and model selection for two biological problems.

5. Single cell images: growth length regulation in fission yeast

In the previous sections we developed novel methods for image based parameter optimization and uncertainty analysis. In this chapter we apply these methods to a series of models for the formation of chemical gradients, which occur in a number of biological processes. A convenient method to measure these gradients is the attachment of a fluorescent label to the gradient forming protein. This protein is then expressed in each cell with the attached label and the cells can be imaged. From those single cell images information about the gradient can be extracted and it can be used to estimate parameters of the gradient formation process. We considered the Pom1p gradient formation as an example of such an image based estimation process. We especially considered different models describing competing biological hypothesis regarding the Pom1 gradient formation. We estimated the model parameters based on the experimental data, assess the parameter uncertainties and finally perform model selection to gain insight in the gradient formation process.

We begin by introducing the biological background and the measurements obtained by (Saunders *et al.*, 2012). In Section 5.2 we introduce four Pom1 gradient formation models considered in this work. In Section 5.3 we perform the parameter estimation for the models and the given data. And in Section 5.4 we finally selected the best model based on the given data set. All results in this chapter were published in the paper of Hross *et al.* (2016).

5.1. Introduction

In the following we will give the biological background for the Pom1 gradient formation, the data considered and then state the research problem of this chapter.

5.1.1. Biological Background

From bacteria to mammalian cells we observe a remarkably constant characteristic cell size, which is attuned to the tissue structure and function and strongly influences organs and organ sizes. To achieve such tight control of the size, the process of cell growth and division has to be controlled in a precise manner. For the model organism fission yeast *Schizosaccharomyces pombe* this is realized by Pom1 (Almeida & Tyers, 2009; Moseley *et al.*, 2009; Martin & Berthelot-Grosjean, 2009). Fission yeast is a rod shaped eucaryotic cell, which divides along the middle axis to form two symmetrical daughter cells. It has been discovered that the polarity factor Pom1 forms a gradient in the cell membranes. Its concentration is strongest at the tips and then dilutes towards the cell middle. As long as the cell is still small Pom1

inhibits the cell division mechanism positioned at the cell middle. When the cell grows, however, the concentration of Pom1 at the cell middle decreases and the cell division mechanism becomes activated. See Figure 5.2B) for a schematic of the process.

5.1.2. Measurement data

For parameter estimation and model selection purposes we use the single cell image data published by Saunders *et al.* (2012). Four types of measurement data are considered here and discussed in the following.

Mean intensity curves For the mean intensity curves 98 cells were imaged and for each cell a cortical mask, i.e. a line along the cell membrane, was defined. Along this mask the Pom1 intensities were measured and as each cell has two independent tips, 196 intensity profiles were collected. Based on the 196 samples the membrane was divided into 60 equally sized regions and for each the mean intensity profile was calculated as well as the standard deviation in the intensity. The mean profile normalized by the mean maximal intensity and the corresponding standard deviation are shown in Figure 5.3B).

FRAP measurements The second set of measurements is based on Fluorescence Recovery after Photobleaching (FRAP). One cell tip was photobleach in a region corresponding to nearly the full tip, while the other tip was used as control. 13 cells were photobleached in this manner and imaged over 600s. Based on those images the mean recovery curve of the intensity and the corresponding standard deviation in the bleaching region was calculated (see Figure 5.3D).

In addition only the half of the tip of a cell was bleached to measure diffusion from one side of the tip into the bleached region. Therefore, 14 cells were bleached and imaged over 60s. From those images the half tip recovery curve and the corresponding standard deviation in the bleaching region scaled by the maximum Pom1 intensity of the tip was calculated (see Figure 5.3D). Both FRAP measurements are used to estimate the parameters of all Pom1 gradient formation models.

Total protein abundance The last set of measurements is the total Pom1 protein abundance in each cell. Therefore the fluorescence intensities of rlc1 and spn4, two other proteins of fission yeast with well known protein abundances, were imaged with the same exposure. Based on the intensities of those reference proteins the total amount of 5000 ± 1900 Pom1 molecules in the cells was estimated (see supplement of (Saunders *et al.*, 2012) for details). For the estimation process we assume that the amount of protein is split equally between the tips.

For the detailed experimental methods and measurement techniques we refer the reader to the original work of Saunders *et al.* (2012).

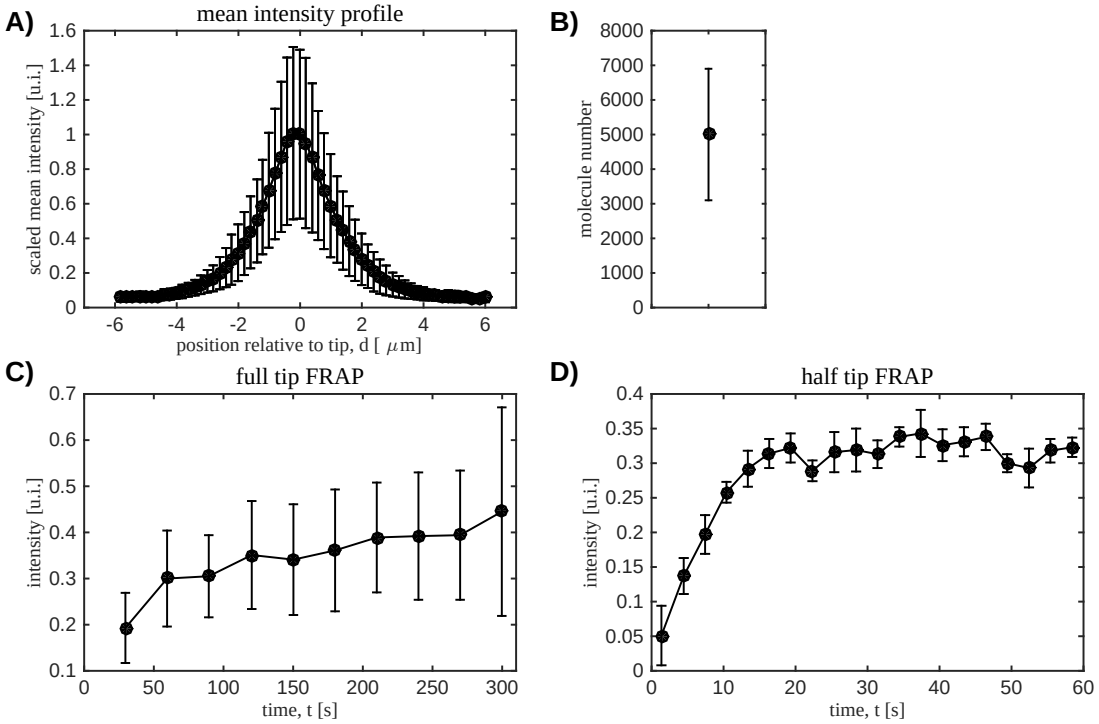


Figure 5.1.: ***Pom1 measurement data.*** *A) Mean intensity curve obtained from 196 steady state measurements. B) Total protein abundance in one cell. The integration over one tip equals to half the amount of total proteins. C),D) Full tip and Half Tip FRAP measurements The errorbars indicate the measured standard deviation.*

5.1.3. Problem statement

While the general idea of the gradient formation could be confirmed by knock-out experiments the detailed formation, especially the stabilization of the gradient against noise and fluctuations remains unclear. Two mechanisms were proposed for the formation of the Pom1 gradient and its stabilization against fluctuations: cluster formation (Saunders *et al.*, 2012; Saunders, 2015) and autophosphorylation Hachet *et al.* (2011, 2012); Hersch *et al.* (2015) (all models are schematically depicted in Figure 5.2). While both models roughly describe the available experimental data, a quantitative comparison is missing.

Problem 5.1 (Modeling of CCL21 gradient formation)**.** *Given the two biological hypothesis and the measurement data estimate the kinetic parameters and perform model selection to choose the most plausible hypothesis based on the given data for Pom1 gradient formation.*

5.2. Modeling of Pom1 gradient formation

In this section we introduce four different models for Pom1 gradient formation:

- the minimal description by a source-diffusion degradation model
- the non-linear cluster formation model (Saunders *et al.*, 2012)
- the auto phosphorylation model (Hachet *et al.*, 2012)
- the multiple site phosphorylation model (Hersch *et al.*, 2015)

The basic gradient formation process and all models are schematically depicted in Figure 5.2.
Before we discuss each of the Pom1 gradient formation models in detail, we want to address the general cell geometry used. The cells have a rod shape, hence we considered a cylindrical body with half spheres as tips. Those cells are symmetrical with respect to the horizontal and vertical axis and we modeled the process only along one line on the surface with $d \in \Omega = [-L, L]$ and $L = 7\mu m$ (see Figure 5.2A)). It has to be kept in mind that this simplification completely neglects delusion effects due to the curvature of the cell tips. In contrast to Saunders *et al.* (2012); Hersch *et al.* (2015) we did not project the process onto a line but we discretized the system on the surface of the cell and assumed that there is a constant diffusion rate on the cell surface. This enables us to simulate the whole tip region in one step, in contrast to previous work where the singularity due to the projection had to be treated by artificial boundary conditions. We consider for all models only the boundary at the division axis of the cells, where we assume no-flux conditions, i.e. for the Pom1 concentration given and denoted by u we obtain

$$\frac{\partial}{\partial \nu} u = 0 \text{ for } d = \{-L; L\} \tag{5.1}$$

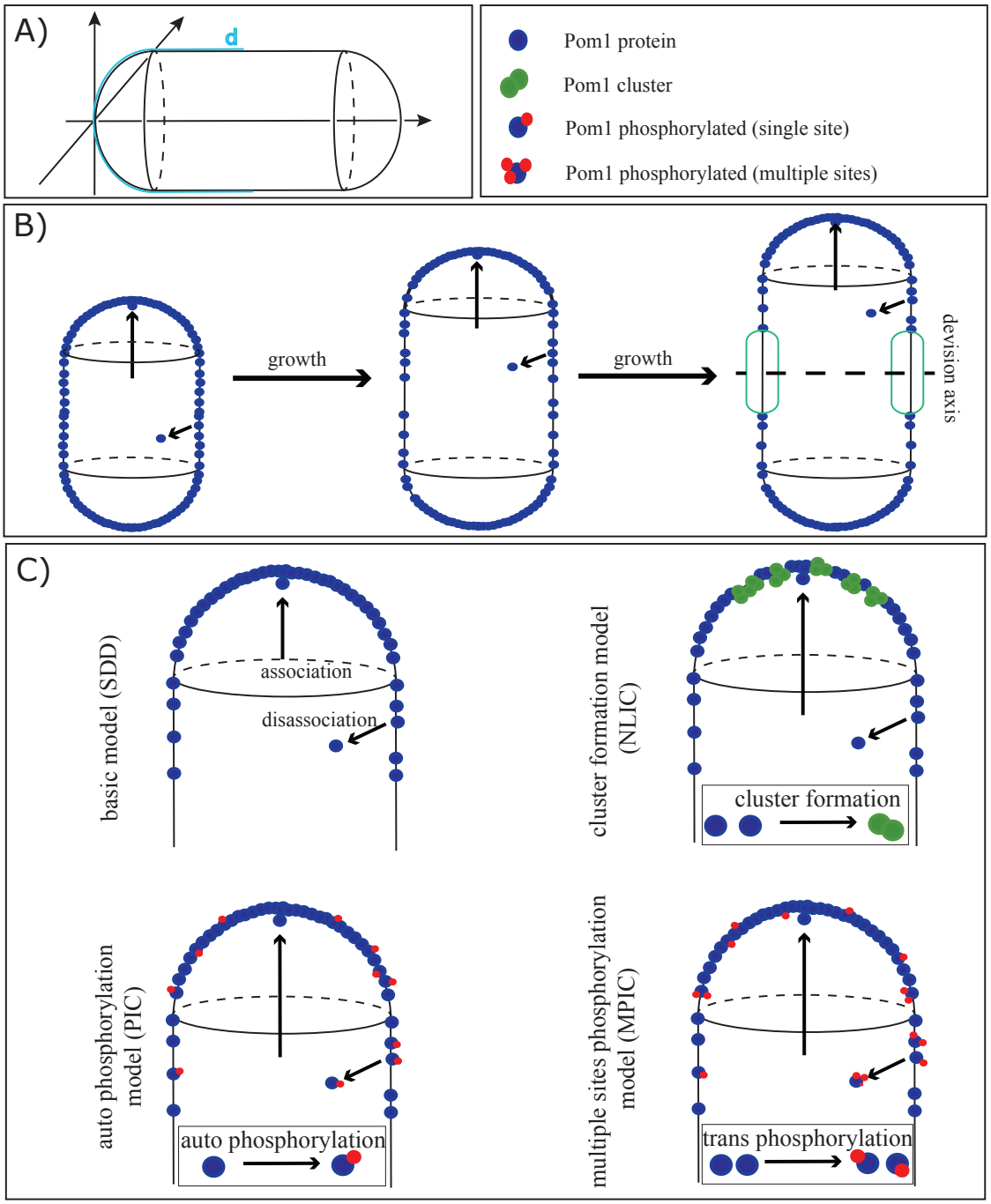


Figure 5.2.: *Schematic of the Pom1 model.* *A) Schematic of the fission yeast geometry. Along the blue line the models are considered. B) Cell size control in fission yeast. The gradient from the tips is constant and if the cell grows the concentration at the cell middle reduce until a region with low concentration arises. This is where division will take place. C) Schematic of the four Pom1 gradient formation models considered. All share the same process of association and disassociation but they differ regarding cluster formation or phosphorylation processes.*

and regarding the initial conditions we assume $u(0,d) = 0$ for $d \in \Omega$ and for the FRAP experiments we assume that

$$\begin{cases} u(0,d) = 0 & \text{for } d \in Q \\ u(0,d) = u^{\infty}(d) & \text{otherwise.} \end{cases} \tag{5.2}$$

With $Q \subset \Omega$ we denoted the photo bleached region and u^{∞} is the steady state of the model.

5.2.1. Source-diffusion-degradation model (SDD)

The first model is introduced in Saunders *et al.* (2012) and considered a simple source-diffusion-degradation model. Pom1 associates to the cell membrane in a circular region at the tip of the cell (see Figure 5.2). The molecules bound per μm per second are given by the rate constant J and the association region is considered as a Gaussian shape with width ρ. Once associated to the membrane Pom1, denoted by u, it diffuses along the membrane with a diffusion constant D and disassociates with a rate μ. The resulting PDE model for $t \in I = [0,T]$ and parameter $\theta = (D, \mu, J, \rho)^T$ with units $([\mu m/s^2], [\mu m/s], [mol\mu m/s], [\mu m])$ is given by

$$\frac{\partial}{\partial t}u = D\frac{\partial^2}{\partial d^2}u - \mu u + \frac{J}{\sqrt{2\pi}\rho}e^{-d^2/2\rho^2} \text{for } d \in \Omega. \tag{5.3}$$

For the boundaries we assumed that there is no Pom1 flow at the division axis of the cells and we obtained (5.1).

5.2.2. Non-linear interacting cluster formation model (NLIC)

We considered the cluster formation model proposed in Saunders *et al.* (2012), which includes a non-linear interaction between slow and fast moving Pom1 components. This model assumes that the Pom1 gradient is stabilized against fluctuations in the input rate by cluster formation. Two Pom1 states are possible: fast moving single molecules denoted by u and slow moving clusters denoted by u_c. Slow clusters are assumed to form mainly at the tips and therefore generate a kind of traffic jam which buffers the variance in the Pom1 input at the tips and guarantees a more or less stable gradient away from the tips. In this model a fraction of the Pom1 molecules, denoted by ε, associated to the membrane at the tips are single molecules and the other fraction are clusters, denoted by $1-\varepsilon$. Only single molecules can disassociate from the membrane as in the SDD model. The cluster formation process considered is a simple complex formation process: when a single molecule meets a cluster it can become part of the cluster with rate β and a cluster can fragment into the single molecule state with rate α. Hence for $t \in [0,T]$ and parameter $\theta = (D, D_c, \mu, \alpha, \beta, J, \rho, \varepsilon)^T$ the model equations are given as

$$\begin{cases} \frac{\partial}{\partial t}u = D\frac{\partial^2}{\partial d^2}u + \alpha u_c - \beta u u_c - \mu u + \varepsilon\frac{J}{\sqrt{2\pi}\rho}e^{-d^2/2\rho^2} & \text{for } d \in \Omega \\ \frac{\partial}{\partial t}u_c = D_c\frac{\partial^2}{\partial d^2}u_c - \alpha u_c + \beta u u_c + (1-\varepsilon)\frac{J}{\sqrt{2\pi}\rho}e^{-d^2/2\rho^2} & \text{for } d \in \Omega. \end{cases} \tag{5.4}$$

For both species we assume that there is no Pom1 flow at the division axis of the cells and we obtain for both the boundary condition (5.1). To ensure that the diffusion constant of the clusters is truly slower than the single molecule, the diffusion

coefficient of the Pom1 clusters is parametrized as fraction $\xi_c \leq 1$ of the diffusion coefficient of free Pom1, $D_c = \xi_c D$.

A key weakness of this model is that there is no mass conservation of Pom1 molecules in the membrane. Clusters of molecule size $n > 2$ are transformed to a single molecule upon fragmentation and all other molecules are simply lost. Furthermore, all cluster sizes are subsumed in u_c.

5.2.3. Autophosphorylation model with one phosphorylation site (PIC)

The model, which considers autophosphorylation of Pom1 as buffering mechanism against fluctuations in the membrane association of Pom1, was introduced by Hachet *et al.* (2011). Like the NLIC model a fastdiffusing component with diffusion rate D and a slow diffusing component with diffusion rate D_p are considered. However, they represent not clusters and single molecules but phosphorylated (u_p) and unphosphorylated (u) forms of Pom1. In this model Pom1 is associated to the membrane in an unphosphorylated form. Once on the surface it can phosphorylate with rate α and then it can disassociate from the membrane. It is assumed that the phosphorylated and unphosphorylated form have different diffusion rates. For $t \in [0, T]$ and parameter $\theta = (D, D_p, \mu, \alpha, J, \rho)^T$ the model equations are given as

$$\begin{cases} \frac{\partial}{\partial t} u = D \frac{\partial^2}{\partial d^2} u - \alpha u + \frac{J}{\sqrt{2\pi}\rho} e^{-d^2/2\rho^2} & \text{for } d \in \Omega \\ \frac{\partial}{\partial t} u_p = D_p \frac{\partial^2}{\partial d^2} u_p + \alpha u_p - \mu u_p & \text{for } d \in \Omega. \end{cases} \tag{5.5}$$

For the boundaries we assume that there is no Pom1 flow at the division axis of the cells for both species and we obtain (5.1) for each. As for the NLIC model, the diffusion coefficient of the phosphorylated Pom1 is parametrized as fraction $\xi_p \leq 1$ of the diffusion coefficient of unphosphorylated Pom1, $D_p = \xi_p D$.

5.2.4. Phosphorylation model with multiple phosphorylation site (MPIC)

The second Pom1 model, which considers phosphorylation of Pom1 as buffering mechanism against fluctuations in the membrane association of Pom1 was introduced by Hersch *et al.* (2015). Here, phosphorylation is also considered, but not with only one but with multiple phosphorylation sites (up to 6 possible sites have been found in Pom1). Furthermore, the phosphorylation is no auto-phosphorylation process but happens by trans-phosphorylation, i.e. two Pom1 molecules meet and phosphorylate each other with rate α. The full model would consider eight Pom1 species. Hersch *et al.* (2015) reduced the model and showed that a one dimensional model with concentration depended disassociation is a good approximation of the full model. For $t \in [0, T]$ and parameter $\theta = (D, \alpha, J, \rho)^T$ the model equations are given as

$$\frac{\partial}{\partial t} u = D \frac{\partial^2}{\partial d^2} u - \alpha u^\gamma + \frac{J}{\sqrt{2\pi}\rho} e^{-d^2/2\rho^2} \text{for } d \in \Omega \tag{5.6}$$

with γ chosen set to 2. For the boundaries we assume that there is no Pom1 flow at the division axis of the cells and we obtain again (5.1). For this model the steady state can be analytically calculated as

$$u^{\infty}(d) = A\frac{d_0^2}{(d+d_0)^2}$$

with

$$A = \sqrt[3]{\frac{3}{2\alpha D}\left(\frac{J}{2\pi\rho}\exp\left(-\frac{d^2}{2\rho^2}\right)\right)^{2/3}} \text{ and } d_0 = \sqrt{\frac{6D}{\alpha A}}.$$

5.3. Parameter estimation for the Pom1 single cell image data

In the following section we performed the parameter estimation on the data obtained by Saunders *et al.* (2012) using the models described in the previous sections. The measurements are a combination of steady state data and dynamic FRAP measurements as described in 5.1.2 and for each measurement we considered a different observation operator g. In the following we describe the observation operator for each measurement in detail. The models consider different components and we always calculated with respect to the total amount of Pom1 u_{tot} with

$$u_{\text{tot}} = \begin{cases} u + s_c u_c & \text{for models (5.4)} \\ u + u_p & \text{for models (5.5)} \\ u & \text{for models (5.3), (5.6).} \end{cases} \tag{5.7}$$

The fluorescence intensity of a Pom1 cluster is a multiple of the single molecule intensity. Determined from the data the range of the intensity is between one and the maximal cluster size of 200 (Saunders, 2015), i.e. $1 < s_c < 200$.

Mean intensity curves First we consider the mean intensity profiles. The profiles are a combination of all Pom1 species scaled to one. The first observation operator is given as

$$g_1(u_{\text{tot}}^{\infty}; s) = \frac{u_{\text{tot}}^{\infty}(d;\varphi)}{\max\limits_{d} u_{\text{tot}}^{\infty}(d;\varphi)}$$

here u_{tot}^{∞} denotes the stationary limit of the model and $s = s_c$ denotes the scaling. Measurements are taken at 60 equally spaced spatial points with $d_k \in \Omega$ and $k = 1, \ldots, 60$ and for each point the mean of the 196 profiles and the standard deviation are given. Following the law of large numbers, we assumed that the error of the measured mean $\overline{y}_{1,k}$ is normally distributed. The standard deviation of the distribution is set to the empirically determined standard error of mean, $\sigma_{1,k}$. Hence, the objective function is given as

$$J_1(\theta, g_1) = \frac{1}{2}\sum_{k=1}^{60}\log(2\pi\sigma_{1,k}^2) + \frac{(\overline{y}_{1,k} - y_{1,k})^2}{\sigma_{1,k}^2}.$$

FRAP measurements For both FRAP measurements we have to consider the dynamic models. At each time point the overall intensity in the predefined bleaching region $d \in Q$ is calculated and normalized by the overall intensity in that region before bleaching.

$$g_{2/3}(u_{\text{tot}}, u_{\text{tot}}^{\infty}; s) = \frac{\int_Q u_{\text{tot}}(d, t; \varphi) dd}{\int_Q u_{\text{tot}}^{\infty}(d; \varphi) dd}$$

as before u_{tot}^{∞} denotes the stationary limit of the model. Preliminary estimation results showed that scaling by the total abundance gives bad fitting results for the measurement data. The scaling for those relative measurements seems to be unknown and we introduced an additional unknown scaling factor and adapted the observation operator to

$$g_{2/3}(u_{\text{tot}}, u_{\text{tot}}^{\infty}; s) = s_{2/3} \int_Q u_{\text{tot}}(d, t; \varphi) dd.$$

For the full tip FRAP 10 measurements at time points $t_k \in [0, 300]$[s] with $k = 1, \ldots, 10$ were taken. Again we assumed that the measurement error is additive, normally distributed with standard deviation obtained from the measurements. Hence, for the data points t_k the objective function is given as

$$J_2(\theta, g_2) = \frac{1}{2} \sum_{k=1}^{10} \log(2\pi\sigma_{2,k}^2) + \frac{(\overline{y}_{2,k} - y_{2,k})^2}{\sigma_{2,k}^2}.$$

For the half tip FRAP 20 measurements at time points $t_k \in [0, 60]$[s] with $k = 1, \ldots, 20$ were taken. Again we assumed that the measurement error is additive normally distributed with standard deviation obtained from the measurements. Hence, for the data points t_k the last objective function is given as

$$J_3(\theta, y_3) = \frac{1}{2} \sum_{k=1}^{20} \log(2\pi\sigma_{3,k}^2) + \frac{(\overline{y}_{3,k} - y_{3,k})^2}{\sigma_{3,k}^2}.$$

Total protein abundance For the total protein abundance in steady state we consider the integral over the tip region and obtain the observation operator.

$$g_4(u_{\text{tot}}^{\infty}; s) = \int_{-L}^{L} u_{\text{tot}}^{\infty}(d; \varphi) dd.$$

We assume that the measurement error is additive, normally distributed with standard deviation obtained from the measurements. Hence, the objective function is given as

$$J_4(\theta, y_4) = \frac{1}{2} \log(2\pi\sigma_4^2) + \frac{(\overline{y}_4 - y_4)^2}{\sigma_4^2}.$$

For the estimation we combined all four measurements. Therefore the i-th component of the overall observation operator $G(u_{\text{tot}}, u_{\text{tot}}^{\infty}; s)$ is given by $g_i(u_{\text{tot}}, u_{\text{tot}}^{\infty}; s)$ and

$$y_{i,k} = g_i(u_{\text{tot}}(d_k, t_k; \varphi), u_{\text{tot}}^{\infty}(d_k, t_k; \varphi); s)$$

and

$$J(\theta, y) = \sum_{i=1}^{4} J_i(\theta, y_i).$$

The general estimation problem is then given as

$$\begin{aligned} &\min_{s\in\mathbb{R}^{n_s},\varphi\in\mathbb{R}^{n_\varphi},u_{\text{tot}}\in W(I)\cap L^\infty} J(\theta, y) \\ &\text{s.t.} \begin{cases} G(u_{\text{tot}}, u_{\text{tot}}^\infty; s) = y \\ u_{\text{tot}} \text{ as in (5.7)} \\ u, u_{c/p} \text{ solution of (5.3), (5.4), (5.5) or (5.6)} \\ u^\infty, u_{c/p}^\infty \text{ stationary limit of } u, u_{c/p}. \end{cases} \end{aligned} \tag{5.8}$$

Implementation All models were discretized with a finite-differences scheme and implemented in MATLAB and compiled and simulated with the AMICI toolbox (*amici-developer.github.io/*), which uses the SUNDIALS solver suite (Hindmarsh *et al.*, 2005). As the numerical simulation using AMICI was computationally efficient and robust, we also used it to compute the steady state of the models. If (1) the numerical simulation failed or (2) the numerical simulation did not yield a steady state for the unperturbed system (steady state condition: $\partial u \partial t < 10^{-6}$ for $t = 2.5 \cdot 10^5$), we set $J_i(\theta)$ to infinity. We performed multi-start local optimization using the Parameter EStimation TOolbox (PESTO) (Hross & Hasenauer, 2016). The starting points for the local optimizations are drawn from a latin hypercube spanning seven orders of magnitude for most parameters. For each of the models we performed the estimation by multi-start local optimization with PESTO, which uses the MATLAB optimization function *fmincon*. To speed up the estimation process and improve accuracy exact gradients were calculated based on the sensitivity equations of each model. Finally 100 local optimizations were performed with starting points chosen by a latin hyper cube sampling. For models (5.3) and (5.6) based on the good convergence of the multi-start global optimization we could conclude that we had found the global optimum for both models (see Figure 5.3 A). For model (5.5) and (5.4) the optimization did not converge as well due to indeterminacy of the parameters. We increased the number of multi-starts to 400 as more then 15% of the multi-starts converged to the same optimal value we assumed that this was also a global optimum.

Regarding the absolute molecule number models (5.3), (5.5) and (5.6) fitted it reasonably well. Unfortunately, the model (5.4) does not ensure mass conservation which might be considered as problematic. Recently, a new cluster formation model was introduced by Saunders (2015), which should be analyzed in future work. Regarding the other measurements we observed that all models produced a good fit for the mean intensity profiles. Also for full and half tip FRAP, the model fit is mostly within the uncertainty of the measurement. For the full tip FRAP it appears however that the model dynamics are slower than suggested by the data.
Following the estimation process we performed a short identifiability analysis with the previously introduced simulation based PL calculation method. This step is

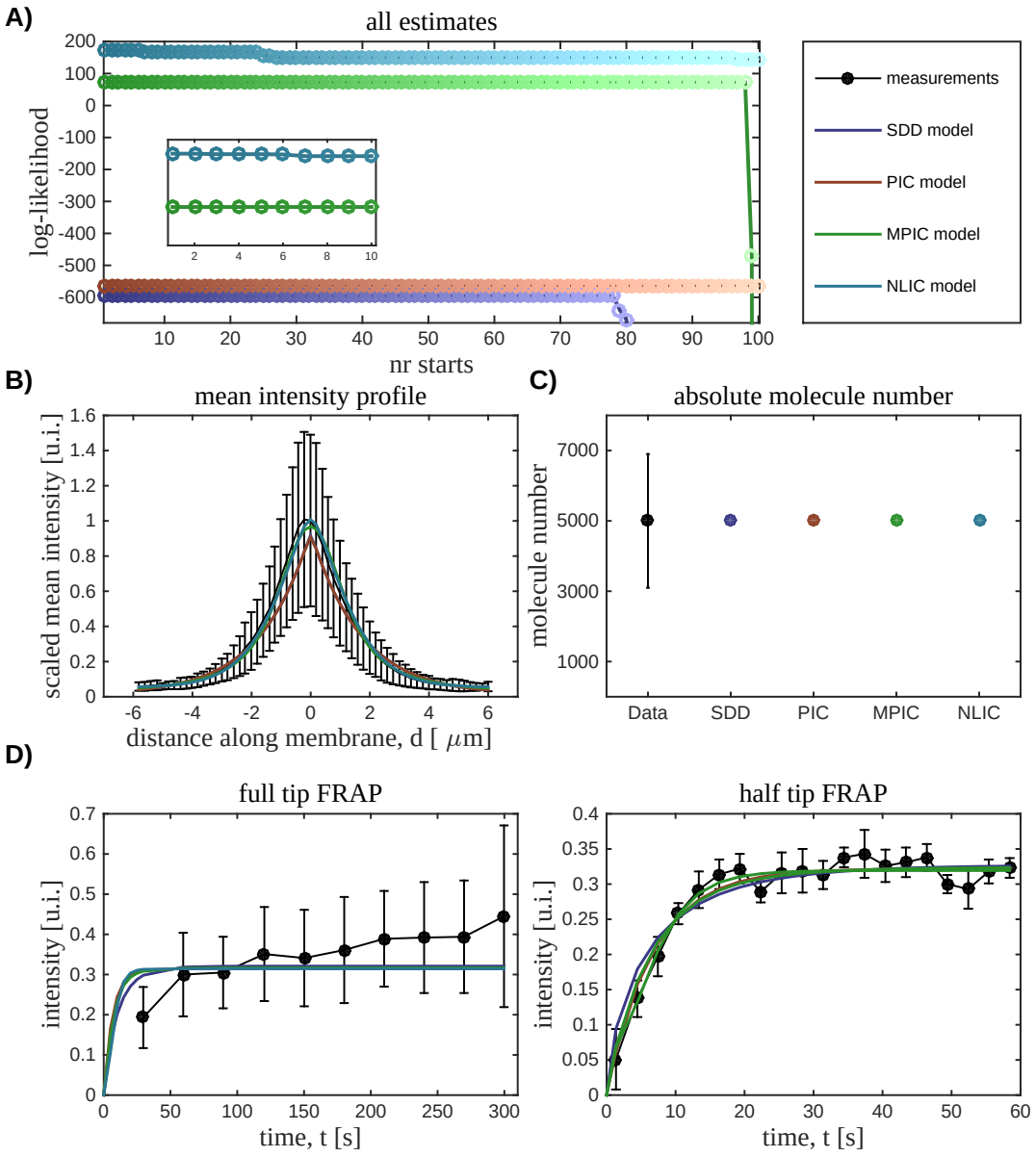


Figure 5.3.: ***Estimation results Pom1 models*** *A) Comparison of the maximal log-likelihood results from 100 optimization starts. We observe that the estimation of all models converged to the same optimum in more than 70% of the runs and the value obtained by model MPIC is maximal. B) Mean intensity profile measurement overlay with simulation results from all four models. We observe that all models are able to give a reasonable description of the mean intensity profile. C) Total protein abundance measurements and data. All models replicated the total number of proteins except the NLIC model, which can not give a molecule number due to problems with the molecule preservation in the model. D) Half and full tip FRAP measurements. The half tip measurements are nicely fitted by all models but the dynamics of all models seem to be slower than suggested by the full tip measurements.*

Table 5.1.: ***Model selection for the Pom1 measurement data.*** *AIC and BIC values calculated for each model. The model with the lowest AIC/BIC value is selected. A difference between two BIC/AIC values smaller than 10 is considered to be indecisive.*

	AIC	ΔAIC	decision	BIC	ΔBIC	decision
SDD	1206.3	1529.7	rejected	1223.9	1517.2	rejected
NLIC	**-323.4**	0	optimal	**-293.2**	0	optimal
PIC	1142.7	1466.1	rejected	1165.3	1458.6	rejected
MPIC	-131.6	191.7	rejected	-114.1	179.2	rejected

often emitted between estimation and model selection. The model selection criteria (AIC/BIC) introduced in Section 2.2.4 only yield reliable results for identifiable parameters otherwise the assumptions for the criteria are not full filled. We found that for models (5.3) and (5.6) all parameters are practical identifiable given the measurements. For models (5.5) and (5.4) the optimization already hinted at indeterminacies. We found that for model (5.5) both diffusion rates D, D_p and the disassociation rate μ are practical non-identifiable. The parameter D is practical non-identifiable in decreasing direction and the parameters D_p and μ are practical non-identifiable in increasing direction.

5.4. Model selection

In this section we determined, which model explains the measurement data best based on the parameter estimation results from the previous section. The results of the multi-starts depicted in Figure 5.3 A already showed that the SDD model is outperformed by the others. The likelihood values obtained by the three other models are so close that the selection is not trivial and we used the AIC and BIC model selection criteria as introduced in Section 2.2. The results are shown in Table 5.1.

We found that with AIC and BIC the SDD model is clearly rejected in favour of the more complex models, which include a buffering effect against noisy inputs. With regard to the AIC/BIC the NLIC model is selected.

To summarize the model selection results we found that all models with a buffering effect against noise in the Pom1 influx perform better than the basic model and the NLIC model performs best.

5.5. Conclusion

Single cell images combined with fluorescent protein fusions and photobleaching are a commonly used tool for many gradient formation processes. For the Pom1 gradient formation process there is a wealth of single cell data already published. For four existing models with competing hypothesis for the gradient formation we performed a rigorous model fitting, uncertainty analysis and model selection. While all models were able to explain the data we found that all models considering a buffering of the Pom1 gradient against noise in the influx of Pom1 outperformed the basic model.

The model based hypothesis testing for the Pom1 gradient formation process from single cell images considered three key aspects: parameter estimation, uncertainty quantification and hypothesis based model selection. We found that all four models could explain the data within the measurement errors. Despite this we found that the dynamics of all models seems to be slower than the full-tip data suggested. While all models were able to explain the data qualitatively, we found that the NLIC model, incorporating clustering as buffering mechanism, outperformed the other models. Unfortunately, the NLIC model does not ensure mass conservation which might be considered as problematic. Recently, a new cluster formation model was introduced by Saunders (2015), which should be analysed in future work.

Recapitulating the the research questions posed in Section 5.1.3 we could answer the question which hypothesis for Pom1 gradient stabilization explains the data best. For the three chosen models representing the two different hypothesis of Pom1 gradient formation the model selection showed that a buffering by cluster formation explains the data best.

The performed parameter optimization and the uncertainty quantification gave additional insights beyond the BIC/AIC base model selection. During the parameter optimization we found that none of the models could explain both FRAP measurements simultaneously. The discrepancy between full and half tip measurements points to either a measurement error or the need for new hypothesis and models for the gradient formation process, which can describe the two-scale characteristic of the data.

On the other hand, the identifiability analysis showed not all parameters for the models NLIC and PIC are identifiably. This yielded convergence problems for the optimization as well as the PL calculation. Furthermore the non-identifiability violates the assumptions of the AIC/BIC model selection criteria and might influence the model selection. To dissolve this either new measurements, for example Tea4 measurements (responsible for Pom1 membrane association) or single measured profiles in contrast to the considered mean profiles could be considered.

Apart from Problem 5.1 the Pom1 gradient formation process was a good real life application example for the efficient profile likelihood calculation method introduced in Chapter 4. The method showed an impressive improvement of profile calculation efficiency for the models SDD and MPIC. This increase in calculation speed enabled

the PL calculation for the computationally more demanding models PIC and NLIC.

Single cell imaging in combination with fluorescent proteins is a convenient tool to study not only temporal but also spatial organization of proteins in cells. This additional spatial dimension allows for more insight into the organization and structural functioning of cells. The related models, however, can be computationally demanding and need efficient analytical tools to extract and test biological hypothesis. We used the Pom1 gradient formation process as an application example for the developed efficient uncertainty quantification methods and we hope that the method will be applied to many gradient formation processes in the future.

6. Tissue scale images: dendritic cell guidance in adaptive immune response

In the previous section, we consider a one-dimensional gradient formation model in a single cell. In the following we consider a tissue scale gradient formation process to underline the scalability of the image based parameter estimation approach we introduced. The measurements considered are immunostainings of tissue slices, which show the formation of a gradient in the whole tissue. For this type of data a special treatment of structured noise, for example given by inhomogeneity of the tissue, has to be addressed during the estimation process and we successfully applied the integrated noise model developed in Section 3.1. Based on the previously introduced methods and a step by step modeling process we addressed the biological question if heperan sulfate concentrations differ between tissue and lymphoid vessels (LV).

We start by introducing the model for the basic source-diffusion process of gradient formation similar to the Pom1p problem in one-dimensional and two-dimensional with an homogeneous heperan sulfate concentration in Section 6.2.1. In Section 6.2.2 we introduce the model with different heparan sulfate concentrations in tissue and LV and in Section 6.2.3 we concentrate on the model with different heparan sulfate concentrations in each LV. The models introduced in these sections are the basis for the parameter estimation in Section 6.3 and the model selection in Section 6.4.

6.1. Introduction and problem statement

In the following we will introduce the biological background for the gradient formation of the signaling molecule CCL21, the data considered and then state the research problem of this chapter.

6.1.1. Biological background

In contrast to the previous chapter, which considered gradient formation for cell size control, this chapter concentrates on haptic movement of cells. Based on the greek word "haptesthai" meaning "to touch", *haptic* movement describes a cell crawling along a surface based on adhesion forces. In general two types of cell movement can be considered: random motion or a directed movement. We considered haptic movement of cells along a signal gradient in tissue, referred to as *haptotaxis*, where "taxis" is the greek word for "order". In contrast to chemotaxis, where the signal molecules diffuse freely, the gradient of signal molecules for haptotaxis is bound to

the tissue surface that the cell crawls along. So while a chemotactic signal can be compared to a smell in the air, haptotactic signals are like emergency lights on the floor of an airplane.
To stay with this analogy, this chapter focuses on the question how the emergency lights are installed. A suitable model system to study haptotaxis in-vivo is dendritic cell guidance towards LV. Dendritic cells (DC) are key players in the adaptive immune response of mammalians. They collect antigen material throughout the body and transport it to the lymph nodes where they present it to T-cells to trigger an adaptive immune response. It has recently been shown that DCs perform haptic movement in an ordered fashion in and outside the LV based on a chemoattractant, a small signal molecule, called CCL21 (Schumann *et al.*, 2010). The interesting ability of CCL21 is that it can exist in two states: as a freely diffusing signal or bound to the tissue surface by heperan sulfate (later only called sugar). The existing theory is that CCL21 is produced and secreted by the LV. The diffusing CCL21 is consequently bound to the surrounding tissue surface and forms a steep gradient. This gradient in turn can be sensed by the DC moving in the tissue and guides them towards the nearest LV. This guidance can be impressively shown by visual tracking of DCs added onto a tissue slice. In a short time all cells have moved inside of the LV in this tissue (see Weber *et al.* (2013) for the movies). While all the key ideas of the gradient formation could be verified by experiments, it remains an open question how the gradient is formed in-vivo detail.

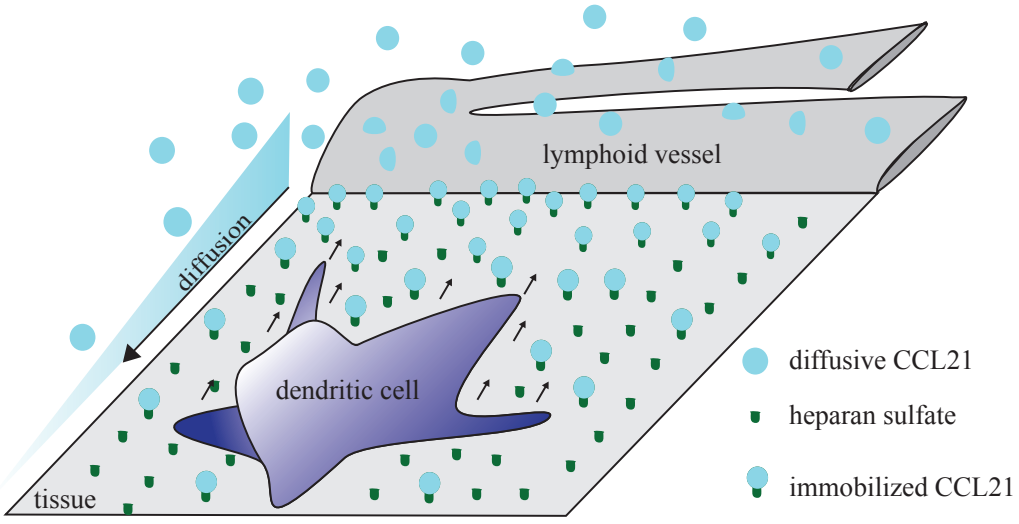


Figure 6.1.: ***Schematic of the haptotaxis model.*** *The soluble chemokine CCL21 (denoted by blue balls) is secreted at the lymphatic vessels (LV). It diffuses freely in the tissue. The heparan sulfate, later referred to as sugar, is equally distributed in the tissue (denoted by the green cups). The CCL21 is immobilized by complex formation with the heparan sulfate. The mature dendritic cells (DC) follow the immobilized CCL21 gradient towards the LV.*

6.1.2. Measurement data

To study the formation of CCL21 gradients, we used confocal microscopy images of immunostainings of mouse ear sheets (Weber *et al.*, 2013). To obtain this data a

thin sheet of a mouse ear was treated with a fluorescent antibody, which binds itself to CCL21 and can be detected under the microscope. After treatment with the antibody the sheet was washed to ensure that all soluble CCL21 and left over antibody had been removed and only the immobilized CCL21 was measured. Microscopy images were taken for 9 different ear sheets (labeled with $k = 1, 2, 3, 4, 12, 13, 14, 15, 16$ later on). For each slice an additional staining with a different fluorescent color was applied to detect the surface of the LV. From those images the LV masks were obtained by manual image segmentation. For the detailed description of the experimental methods, the laboratory animals and the data collection we refer the reader to the original work (Weber *et al.*, 2013). A representative combination of image and mask data is shown in Figure 6.2 and the full data set is shown in the Appedix A.

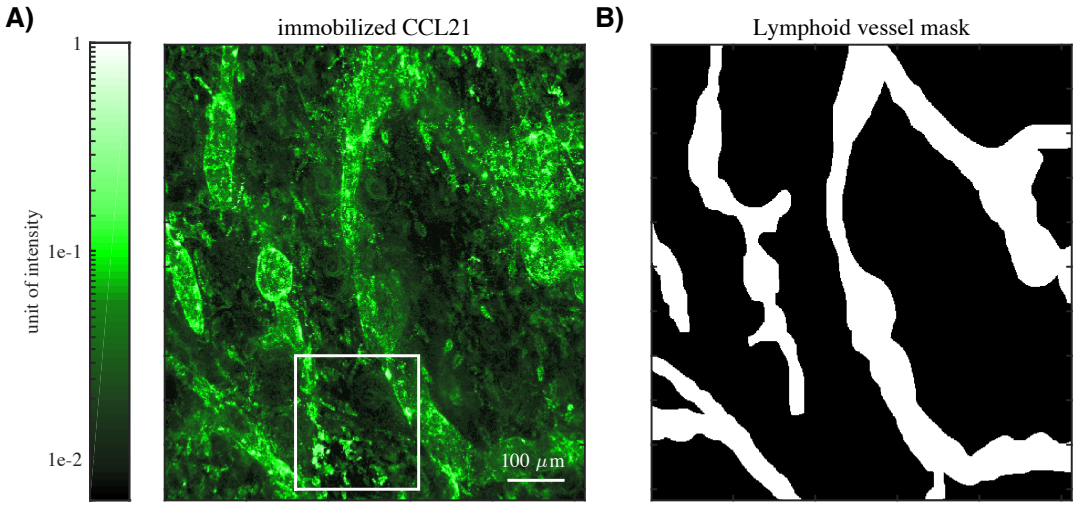


Figure 6.2.: ***CCL21 immunostainings and LV masks.*** *A) representative image of CCL21 immunostaining, later on called measurement CCL21-1, with green color coding for better visibility of the signal. The white box markes a region with especially many bright spots discussed in the image processing section. B) Corresponding LV masks obtain by an additional staining with an LV surface marker and manual segmentation of the lymphoid region (Weber* et al.*, 2013).*

Data analysis

For this analysis we defined a region of interest Ω equal to 400x400 pixels to exclude boundary effects. And we denoted with Q the mapping, which represents the LV mask (see Figure 6.2). To quantify the CCL21 intensities according to their distance from the LV we introduced a distance map as in Weber *et al.* (2013).

Definition 6.1.1 (Distance map)**.** *Let $\Omega \subset \mathbb{R}^2$ be the image and $Q : \Omega \to \mathbb{R}$, $x \mapsto \{0, 1\}$ the LV mapping such that*

$$\Upsilon = \{x \in \Omega | Q(x) \equiv 1\}$$

is the LV region. The distance map is given as

$$dm(x) = \min_{y \in \Upsilon} |x - y| \,, \ \forall x \in \Omega. \tag{6.1}$$

In the following sections we will use this mapping for image analysis and parameter estimation. Based on the distance map a mean intensity for each individual CCL21 staining image is calculated depending on the distance from the LV. For a given set of distances $r_0, \ldots, r_n$ the mean intensity for all pixels with $r_{i-1} < dm(x) < r_i$ is calculated. In Figure 6.3 the mean intensity for each image and the mean overall measurements are shown. Based on this extracted mean intensity, the parameters of the models introduced in Section 6.2 will be estimated in Section 6.3.2.

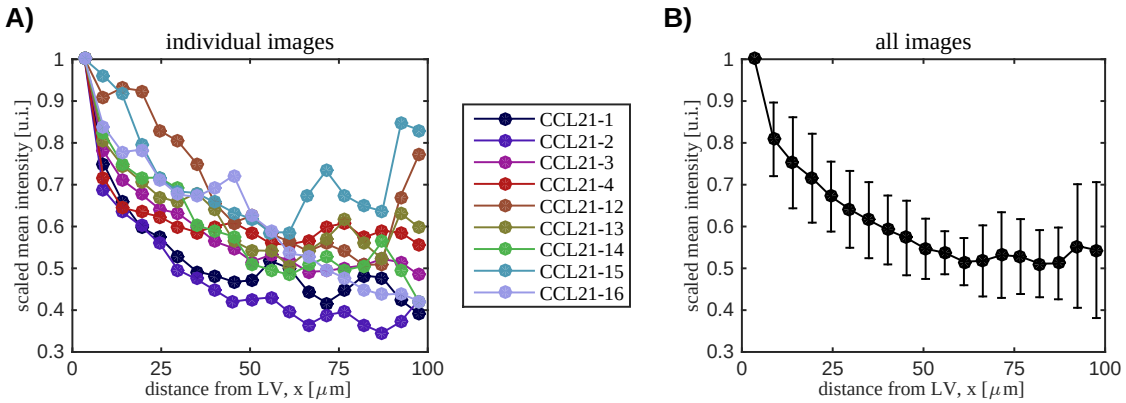


Figure 6.3.: ***Distance measure.*** *A) Distance measure scaled by the maximal value was calculated for each image individually. The used grid was* $r_j = 1+2*j$ *for* $j = 0, \ldots, 20$. *It is easily visible that the gradient steepness of the distance measure strongly varies between the images. B) Average CCL21 intensity across images for different distances from the lymphoid vessels.*

6.1.3. Image processing

The distance dependent fluorescent intensity provides a first quantification of the data. To analyze the data in detail, we performed a visual inspection. We found high intensity spots in regions away from the lymphoid vessels. We assumed they originate from other cell types migrating towards the lymphoid vessels or inhomogeneities in the tissue, which react differently to the staining than the other tissue. These spots, however, are not of interest for the process considered later on and can be seen as structured noise produced by the measurement procedure. In the following we discuss how to filter these bright spots to obtain an adjusted image.

Maximally stable extremal regions filtering

For the detection of the bright spots in the staining images we applied a maximally stable extremal region (MSER) filtering introduced by Matas *et al.* (2004) and the implementation introduced by Nistér & Stewénius (2008). It is based on a water shedding mechanism, where the image is considered as a landscape with heights given by the image gray scale value. This landscape is flooded by gradually raising the water level until the whole image is immersed. The algorithm keeps track of the connected water regions for each water level. MSER is a method to detect high intensity regions for example in fluorescence and brightfield images and has already been successfully applied to detect cells (Buggenthin *et al.*, 2013). We used the

fast C implementation of the algorithm used in Buggenthin *et al.* (2013). To find the bright spots in our microscopy images we adjusted the water level to a value needed to find the regions corresponding to the structured noise. We also tuned the maximal and minimal size of a region assumed to be part of the structured noise.

After identifying high fluorescent cells we adapted the obtained regions to prevent effects at the boundary of the cells. This was facilitated by dilation the obtained region by 1px using the MATLAB routine `imdilate` of the image processing toolbox. Additionally, we neglected all regions inside the LV regions, as we assumed that the structured noise is not found inside the LVs. Figure 6.4A) shows a representative filtering mask as well as the adjusted image for the measurement CCL21-1.

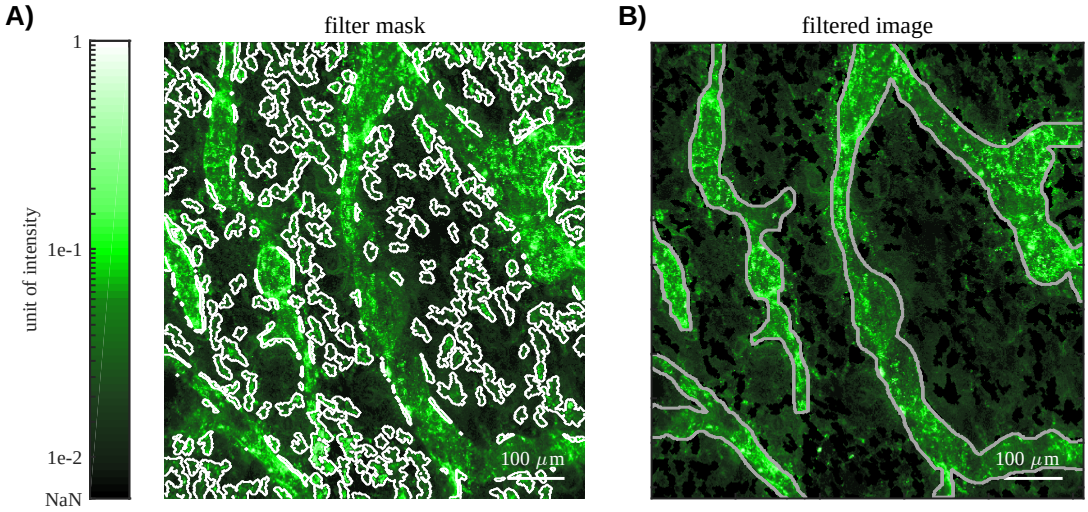


Figure 6.4.: ***Filtered images.*** *A) Immunostaining image CCL21-1 with the outline of the region obtained by the MSER filter. For the filtering process all points inside the LV region were neglected. We used a minimal region size of 50 pixels, a maximal region size of 200 pixels, a threshold level of 1 for the MSER algorithm and a dilation of 1 pixel for the obtained mask. B) Adjusted image with all points indicated by the filter removed. The application of the filter resulted in the removal of 16% of the pixels.*

Distance measure

Analogously to the distance map for the raw images a mean intensity for each individual filtered CCL21 image is calculated depending on the distance from the LV. For the same region of interest and set of distances $r_0, \ldots, r_n$ as in Figure 6.3. The mean intensity for all pixels in the filtered images with $r_{i-1} < dm(x) < r_i$ is calculated. The results are shown in Figure 6.5.

We can observe that the distance measure no longer varies as strongly as for the raw data shown in Figure 6.3. The obtained curves are smooth as the spikes which were introduced by the bright spots are removed from the data. Furthermore, the steepness of the distance measure of all images is now comparable between the nine

individual measurements. Finally, we see that the mean distance measure of the combined images shows a smooth gradient, which has a considerably larger range than the gradient obtained by the raw images. It is important to note how strongly the distance measure in influenced by the outliers (bright spots) in the data, which indicates that it is no stable measure to obtain an estimate for the diffusion length of the chemokine CCL21.

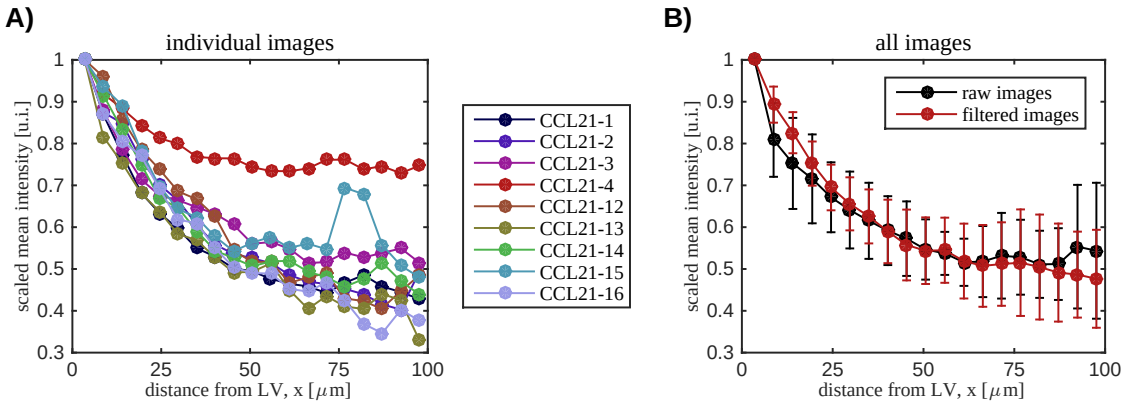


Figure 6.5.: ***Distance measure for the filtered regions.*** *Distance measure scaled by the maximal value was calculated for each filtered image individually. The used grid was $r_j = 1 + 2 * j$ for $j = 0, \ldots, 20$. It is visible that the gradient steepness is similar between the images except for CCL21-4, which has high background values. B) Average CCL21 intensity across images for different distances from the lymphoid vessels for the raw images and the filtered images.*

Filtered region analysis

In the following sections we will introduce models for the CCL21 gradient formation process and estimate the model parameters from the image data. To asses the efficiency of the proposed estimation procedure we used simulated data based on the true LV structure and some realistic kinetic parameters. However, the treatment of the structured noise in the data is of special interest and we introduced such artifacts in the simulated data. To obtain simulated data as close as possible to the real data we analyzed the size and shape distribution of the bright spots in the filtered real images. For the size of the bright spots we fitted a normal distribution to the area size of the convex hulls for each segmented region. To exclude clumped cells we restricted the region size to maximally 250px^2. We obtained a nearly normally distributed spot size with mean $\mu = 149\text{px}^2$ and variance $\sigma = 47\text{px}^2$. Furthermore we consider the elliptic shape of the bright spots in terms of the length of the major and minor axis and fitted a two-dimensional normal distribution with mean minor axis $\mu_a = 9.5\text{px}$ and major axis $\mu_b = 19.3\text{px}$ and covariance matrix $\Sigma_{a,b} = \begin{pmatrix} 19.2 & -4.2 \\ -4.2 & 7.5 \end{pmatrix}$. The fitting was done with the MATLAB routines `normfit` and `gmdistribution`. The distributions of the data and our assumed distributions are displayed in Figure 6.6.

6.1.4. Problem statement

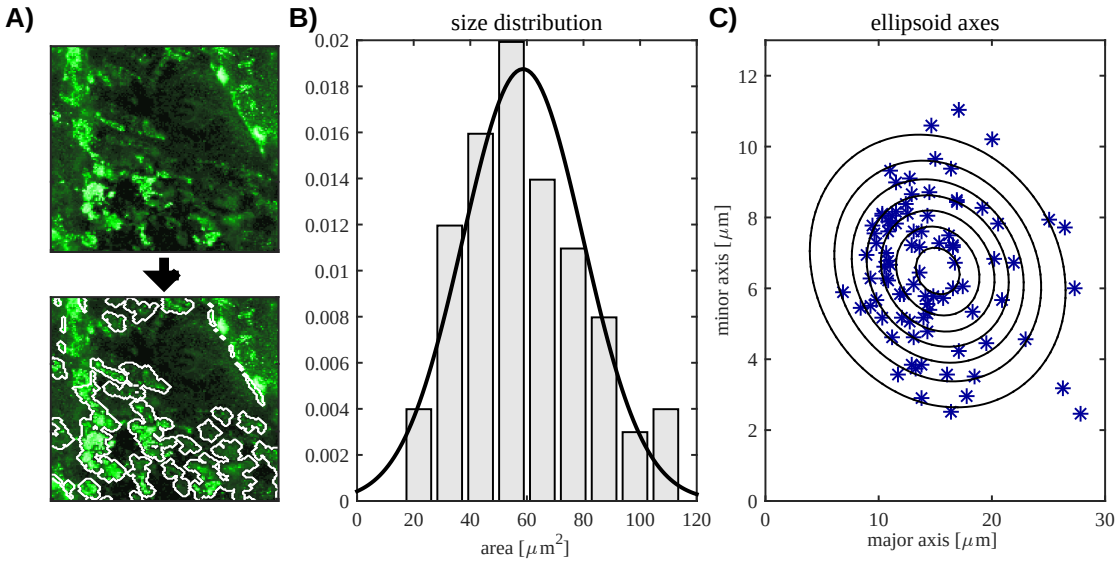


Figure 6.6.: ***Region properties.*** *A) Zoom in of region with many spots for measurement CCL21-1. The lower panel displays the filtering masked obtained for this image. B) Bar plot for the area size distribution obtained from each segmented region and line plot of the normal distribution fitted to the data. C) Scatter plot of minor versus major axis of the surrounding ellipsoid calculated for each segmented region and contour of the normal distribution fitted to the data.*

By in-vitro and in-vivo experiments the key features of CCL21 gradient formation could be confirmed and with immunostainings it is possible to quantitatively measure the gradient in tissue. The exact gradient formation process, however, remains unclear. Especially the influence and importance of the spatial organization of heparan sulfate for the gradient formation remains uncertain. While the measurement data is available a quantitative, model based analysis of the CCL21 gradient formation process is missing.

Problem 6.1 (Modeling of CCL21 gradient formation)**.** *Given the biological knowledge about the CCL21 gradient formation process and the measurement data develop models for the different heparan sulfate conditions, estimate the kinetic parameters and perform model selection to choose the most plausible description of CCL21 gradient formation for the given data.*

The tissue scale images considered for this model based hypothesis testing need a tailored parameter estimation method. In particular the structured noise found by the preliminary data analysis in Section 6.1.2 has to be considered in detail.

Problem 6.2 (Parameter estimation of tissue scale images)**.** *Given the immunostaining images develop noise models and corresponding likelihood functions to choose the best parameter estimation method for this particular tissue scale data.*

Both problems considered are closely related and will be solved through the chapter to arrive at a decision to improve the description of the process of CCL21 gradient

formation.

6.2. Modeling of CCL21 gradient formation

The deduction of the CCL21 gradient formation models was an iterative process to step by step define the spatial sugar distribution. We started with a basic description of the process and refined the models with regard to the spatial structure of the sugar with respect to the parameter estimation results from the model. The model development process is schematically depicted in Figure 6.7.

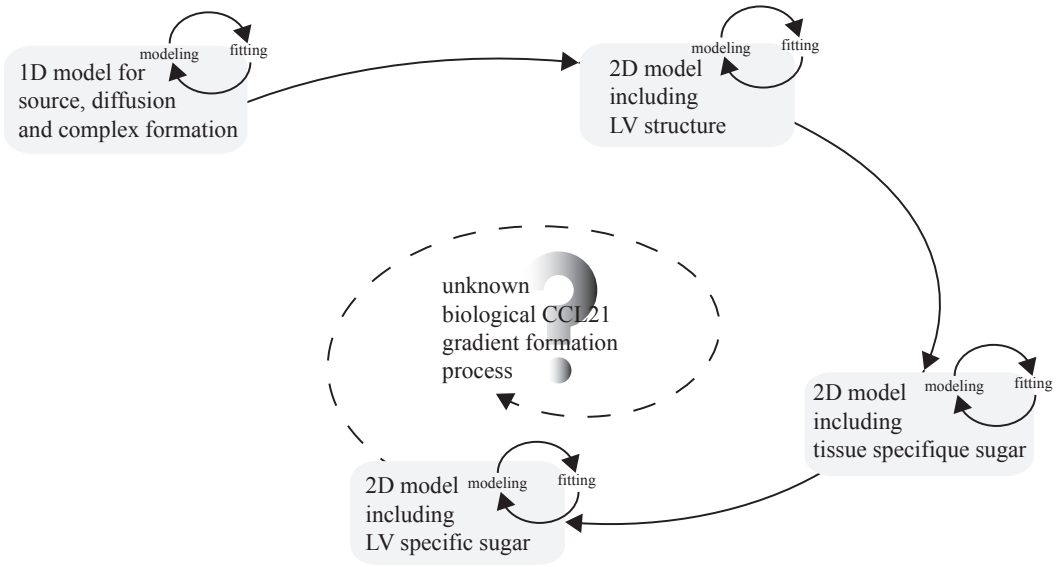


Figure 6.7.: ***Modeling process for the CCL21 gradient formation.*** *We started with a simplistic one-dimensional discription of the gradient formation process and iteratively increased the model complexity homing in on the true underlying process. Each step in the modeling process included a model calibration to the data and an analysis of the models capacity to explain the data (depicted by the gray boxes).*

6.2.1. Model of basic source-diffusion-complex formation process

The initial model for CCL21 gradient formation considers the signaling protein CCL21, denoted by u which is produced constantly at a spatially distributed source Q, the lymphatic vessels in the tissue, with a rate α. As CCL21 is a small, soluble protein it is assumed to diffuse freely in the tissue, with a diffusion constant D, furthermore it is degraded with a degradation rate γ. The CCL21 gradient is immobilized through a reversible complex formation with the tissue bound sugar, denoted by s. The immobilized CCL21 protein is denoted by c, which is the gradient sensed by the dendritic cells. Hence the chemical reactions considered in the model are:

$$\emptyset \xrightarrow{\alpha \cdot Q} u,\ u \xrightarrow{\gamma} \emptyset \text{ and } s + u \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} c. \tag{6.2}$$

One-dimensional model

In one dimension there is no spatially distributed source but the source is positioned at distance zero and can be interpreted as an influx. We studied the problem for three state variables u,s and c. Each variable is a function of the distance from the lymphoid vessel position $x \in \Omega = [0, 110][\mu\text{m}]$, time $t \in [0, T][\text{s}]$ and a set of unknown parameters $\theta = (D, \alpha, k_1, k_{-1}, \gamma)$. For a detailed description and deduction of such models see Wartlick *et al.* (2009). The model is given by

$$\begin{cases} \frac{\partial}{\partial t}u - D\frac{\partial^2}{\partial x^2}u = -k_1 usu + k_{-1}c - \gamma u & \text{for } x \in \Omega, t \in [0, T] \\ \frac{\partial}{\partial t}s = -k_1 us + k_{-1}c & \text{for } x \in \Omega, t \in [0, T] \\ \frac{\partial}{\partial t}c = k_1 us - k_{-1}c & \text{for } x \in \Omega, t \in [0, T] \end{cases}$$

with initial conditions

$$u(0, x) = c(0, x) = 0 \text{ and } s(0, x) = S_0 \ \forall x \in \Omega$$

and boundary conditions

$$\frac{\partial}{\partial \nu_C}u(t, x)\Big|_{x=110\mu\text{m}} = 0 \text{ and } \frac{\partial}{\partial \nu_C}u(t, x)\Big|_{x=0\mu\text{m}} = \alpha \ \forall t \in [0, T].$$

Steady state All measurements given in Section 6.1.2 are assumed to be steady state measurements and hence we have to consider the steady state for the model for further analysis, which is given as

$$\begin{cases} 0 = \frac{D}{\gamma}\frac{\partial^2}{\partial x^2}u(x) - u(x) & \text{for } x \in \Omega \\ \frac{\partial}{\partial \nu_C}u(x) = 0 & \text{for } x = 110\mu\text{m} \\ \frac{\partial}{\partial \nu_C}u(x) = \alpha & \text{for } x = 0\mu\text{m}. \end{cases} \tag{6.3}$$

and for s and c we obtain

$$s(x) = \frac{S_0}{\frac{1}{K_D}u(x) + 1} \text{ and } c(x) = \frac{S_0 p(x)}{u(x) + K_D} \text{ for } x \in \Omega. \tag{6.4}$$

Here $K_D = k_{-1}/k_1$ is the disassociation constant. For the stationary limit we can solve the equation analytically (Wartlick *et al.*, 2009) and obtain for $\sqrt{D/\gamma} \ll 110\mu\text{m}$, which is a reasonable assumption given the data shown in Figure 6.3,

$$u(x) = \frac{\alpha}{\sqrt{D/\gamma}}e^{-x/\sqrt{D/\gamma}}.$$

If we consider the model in terms of the rescaled variable $\widehat{u} = u/K_D$ it becomes clear that the maximal possible parameter set which can be identified, independent of measurement noise and observations, based on steady state data is $\theta = (D/\gamma, \alpha/\gamma K_D, S_0)$.

Two-dimensional model

The two-dimensional model also includes three state variables: u, s and c. And again each variable is a function of the spatial location $x \in \Omega = [0, 604]^2[\mu\text{m}^2]$, time $t \in [0, T][\text{s}]$ and a set of unknown parameters $\theta = (D, \alpha, k_1, k_{-1}, \gamma)$. The model considered is:

$$\begin{cases} \frac{\partial}{\partial t}u - D\triangle u = \alpha Q(x) - k_1 us + k_{-1}c - \gamma u & \text{for } x \in \Omega,\ t \in [0, T] \\ \frac{\partial}{\partial t}s = -k_1 us + k_{-1}c & \text{for } x \in \Omega,\ t \in [0, T] \\ \frac{\partial}{\partial t}c = k_1 us - k_{-1}c & \text{for } x \in \Omega,\ t \in [0, T] \end{cases}$$

$\triangle$ denotes the Laplace operator in two dimensions. The initial conditions considered are

$$\forall x \in \Omega : u(0, x) = c(0, x) = 0 \text{ and } s(0, x) = S_0\ \forall x \in \Omega.$$

Regarding the boundary conditions no-flux boundary conditions were chosen. In case of a "cut-out" region of tissue, this means there is no flux out of the tissue. The boundary conditions are given as

$$\frac{\partial}{\partial \nu_C}u(t, x)\Big|_{x\in\partial\Omega} = 0\ \forall t \in [0, T].$$

Steady state The resulting model is:

$$\begin{cases} 0 = D\triangle u + \alpha Q(x) - \gamma u & \text{for } x \in \Omega \\ \frac{\partial}{\partial \nu_C}u(t, x)\Big|_{x\in\partial\Omega} = 0 & \text{for } x \in \partial\Omega. \end{cases} \tag{6.5}$$

and for s and c we obtain

$$s(x) = \frac{S_0}{\frac{1}{K_D}u(x) + 1} \text{ and } c(x) = \frac{S_0 p(x)}{u(x) + K_D} \text{ for } x \in \Omega \tag{6.6}$$

here $K_D = k_{-1}/k_1$ is the disassociation constant like in the one-dimensional case. Again we consider the model in terms of the rescaled variable $\widehat{u} = u/K_D$ and see that the the maximal possible parameter set which can be identified does not vary compared to the one-dimensional case. Independent of measurement noise and observations, based on steady state data we can identify $\theta = (D/\gamma, \alpha/\gamma K_D, S_0)$.

6.2.2. Model for tissue specific sugar concentrations

We extended the basic model by considering two different sugar concentrations for the lymphoid vessel region, denoted by S_L, and the surrounding tissue, denoted by S_0. Hence we have the same reaction equations (6.2) as for the basic model but a varying initial value for s. As this adaptation is only visible in the two-dimensional set-up we do not consider the one-dimensional model here. Furthermore we only consider the steady state model: For the three state variables u, s and c, each a function of the spatial location $x \in \Omega = [0, 604]^2[\mu\text{m}^2]$ and a set of unknown parameters $\theta = (D, \alpha, k_1, k_{-1}, \gamma, S_0, S_L)$, the considered model is given as

$$\begin{cases} 0 = D\triangle u + \alpha Q(x) - \gamma u & \text{for } x \in \Omega \\ \frac{\partial}{\partial \nu_C}u\Big|_{x\in\partial\Omega} = 0 & \text{for } x \in \partial\Omega. \end{cases} \tag{6.7}$$

and for s and c we obtain the changed equations

$$s(x) = \frac{S_L Q(x) + S_0(1 - Q(x))}{\frac{1}{K_D}u(x) + 1} \text{ and } c(x) = \frac{(S_L Q(x) + S_0(1 - Q(x)))u(x)}{u(x) + K_D} \quad (6.8)$$

for $x \in \Omega$. Here we also consider the rescaled variable $\widehat{u} = u/K_D$ and see that the set of structural identifiable parameters is $\theta = (D/\gamma, \alpha/\gamma K_D, S_0, S_L)$.

6.2.3. Model with lymphoid vessel specific sugar concentration

Based on new experimental evidence which suggests that a variation in the individual lymphoid vessel branches might influence the dendritic cell migration (Kilarski *et al.*, 2013), we adapted the model considering different sugar concentration for each lymphoid vessel branch and the surrounding tissue. As before we only consider the two-dimensional case and the steady state model. Before we can write down the model here we define

$$Q(x) = \sum_{k}^{n_k} q_k(x) \text{ with } q_k(x) : \Omega \to \{0, 1\}$$

where n_k is the number of LV branches and q_k denotes the individual LV branches which do not overlap. Then we could write the model for $x \in \Omega = [0, 604]^2[\mu\text{m}^2]$ and a set of unknown parameters $\theta = (D, \alpha, k_1, k_{-1}, \gamma, S_0, S_{L,1}, \ldots, S_{L,n_k})$ and the three state variables u, s and c as,

$$\begin{cases} 0 = D\triangle u + \alpha Q(x) - \gamma u & \text{for } x \in \Omega \\ \frac{\partial}{\partial \nu_C} u\Big|_{x \in \partial\Omega} = 0 & \text{for } x \in \partial\Omega. \end{cases} \quad (6.9)$$

and for s and c we obtain the changed equations

$$s(x) = \frac{\sum_k S_{L,k} q_k(x) + S_0(1 - Q(x))}{\frac{1}{K_D}u(x) + 1} \text{ and } c(x) = \frac{(\sum_k S_{L,k} q_k(x) + S_0(1 - Q(x)))u(x)}{u(x) + K_D} \quad (6.10)$$

for $x \in \Omega$. If we consider the rescaled variable $\widehat{u} = u/K_D$ for the model above we see that the maximal possible parameter set which can be identified is $\theta = (D/\gamma, \alpha/(\gamma K_D), S_0, S_{L,1}, \ldots, S_{L,n_k})$. Hence the sugar concentration can always be identified for steady state data. In the following chapters we will perform parameter estimation for those three model and in Section 6.4 we will determine the best fit model for the considered data.

6.3. Parameter estimation for CCL21 tissue-scale microscopy images

In the following we perform the parameter estimation for the basic source-diffusion-complex-formation process described in the previous section. As mentioned in the beginning of this chapter the measurements are microscopy images of the immunostainings for immobilized CCL21 (Weber *et al.*, 2013) and we modeled the observation as

$$y = b + c(x), \quad (6.11)$$

where b is a background intensity produced by autofluorescence of the tissue. All parameter estimation problems considered in the following are for the steady state of each model, as we assumed that the measurements are taken close to an equilibrium state of the system.
As mentioned in the previous parts of this work a major challenge in image based parameter estimation is the presence of biological processes, which are of no interest for the biological question at hand but can not be disabled experimentally. They introduce a structured noise in the images, which has to be considered in the parameter estimation process as it might bias the obtained estimates. For the CCL21 data this is a prominent problem as we observe bright spots in the data (see Figure 6.2). Possible origins of the spots could be migrating cells or inhomogeneities, which react differently to the immunostaining but are not considered crucial for the process. In this chapter we compared the parameter estimation on an extracted features of the microscopy data to an integrated approach which treats the structured noise as an additional noise term in the likelihood function as introduced in Section 3.1. We considered two features: the one-dimensional distance measure introduced in Section 6.1.2 and the two-dimensional images obtained by filtering the data as explained in Section 6.1.3. Based on this comparison we choose the best performing method to proceed with the model selection in Section 6.4.
In the following we always verified the validity of the estimation on simulated data and then applied the method to the real data to obtain trustworthy parameter estimates. In Section 6.3.1 we introduce the simulated data we considered for the estimation process. To obtain data as close as possible to the real data we used the image properties deduced in Section 6.1.3. Section 6.3.2 concentrates on the estimation of the kinetic parameters for the extracted distance measure and in Section 6.3.3 we used the two-dimensional images for the parameter estimation.

6.3.1. Simulated data

To assess the ability of the proposed procedure to recover the true parameter value we applied the method to simulated data, generated in a way that it is as close as possible to the real data as introduced in Section 6.1.3 (see also Figure 6.3).

A crucial aspect to obtain data close to the real data is to find a set of reasonable parameter values. To asses the similarity between the generated data and the real data we used the distance measure introduce before. By hand we chose a set of parameters, which gave a distance measure for the simulated image, which is close to the distance measure of the filtered CCL21-1 image with the filtering parameters as used in Figure 6.5. The resulting parameters are shown in Figure 6.8 A) and the distance measures are compared in Figure 6.8 B).
Next we added structured noise by introducing bright spots. We generated those bright spots in the simulated data by randomly choosing uniformly distributed spatial points outside the LV region and added an ellipsoid. The ellipsoid size was chosen by drawing random samples from the minor and major axis distribution obtained from the images. As the surrounding ellipsoid overestimates the size of the extracted regions we scaled the axis lengths such that we obtained a size distribution close to the one obtained from the data (see Figure 6.6). As we assessed the per-

formance of the estimation for different strength of structured noise, we generated data with varying numbers of bright spots. We simulated images with 20, 40, 80, 160, 320 and 640 ellipsoids and for each spot number we created 30 independent samples.
Finally we added multiplicative log-normally distributed noise ε to each image with $\varepsilon \sim \mathcal{LN}(1, \sigma^2)$. A reasonable parameter value for the variance σ^2 was also chosen together with the other parameters as $\sigma = 0.38$. One exemplary image with 320 bright spots is shown in Figure 6.8 C). In the following we are going to study the bias of the estimated parameters due to the side process of firoblast migration towards the LV visible in the images based on this simulated data and choose the best performing method for bias correction to apply it to the real data.

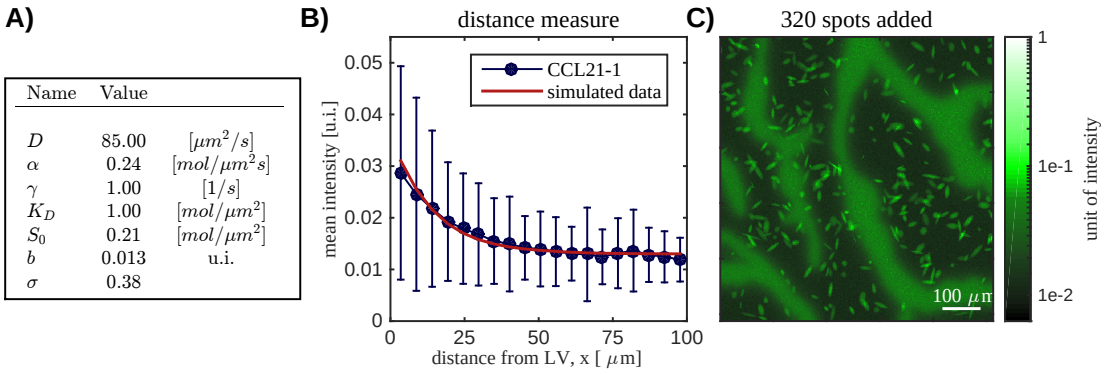


Figure 6.8.: ***Simulated data for the basic source-diffusion-complex formation process.*** *A) Parameter values assumed for the generation of the simulated data. B) Comparison of distance measure obtained from the filtered CCL21-1 image and the one given by the simulated model and the corresponding observable $y(x)$. C) Representative image of the simulated data for 320 added ellipsoids and multiplicative log-normally distributed noise generated with the parameters given in A) and the model for the basic source-diffusion-complex-formation process.*

6.3.2. Distance measure

For the parameter estimation we first considered the distance measure extracted from each image as data. For the estimation of the kinetic parameters based on the distance measure we applied two approaches: first we estimated the parameters using the one-dimensional model introduced in Section 6.2.1 and second we calculated the distance measure based on the simulation of the corresponding two-dimensional model introduced in the same section. For both methods we used the simulated data to verify the validity and then applied it to the distance measures obtained from the real data.

One-dimensional model

For the parameter estimation of the one-dimensional model from the distance measure we considered the system state u,s and c in the domain $\Omega = [0, 110]\mu m$. As the scaling of the images and the corresponding distance measures varies strongly we

introduced an observation operator for each measurement, which includes a scaling parameter $s = (s_1, \ldots, s_N)$ where N is the number of images and an additive offset b, which is assumed to be due to autofluorescence of the tissue. The l-th observation operator maps the state u onto the spatial points given by the distance measure grid $r_k = 1 + 2k$ for $k = 0, \ldots, 20$. The observation operator is given by

$$y_{l,k} = G_{l,k}(s, b) \begin{pmatrix} u \\ s \\ c \end{pmatrix} = s_l \cdot (b + c(x_{r_k})) \quad \text{for } l = 1, \ldots, N.$$

From the previous analysis of the one-dimensional CCL21 model we knew that the parameter combinations D/γ, $\alpha/(\gamma K_D)$ and S_0 are structural identifiable if we have full observations. We considered the parameter estimation the parameter set $\theta = (D/\gamma, \alpha/(\gamma K_D), S_0, b, s_1, \ldots, s_N)$. For the partial, scaled observation introduced above only the product $s_l S_0$ occurs in the equation as well as the product $s_l b$ hence only those are structural identifiable.
For the distance measure we can not deduce the noise model from the measurement technique. However, within each bin we know the mean and standard deviation of the data shown by the errorbars in Figure 6.8. We assume that this error is additive and normally distributed with known variance, i.e

$$\overline{y}_{l,k} = y_{l,k} + \varepsilon_{l,k} \text{ with } \varepsilon \sim \mathcal{N}(0, \sigma_{l,k}).$$

Following Section 2.2 we consider as objective function for the parameter estimation the reduced negative log-likelihood function

$$j(\theta) = \frac{1}{2} \sum_{l=1}^{N} \sum_{k=1}^{20} \log(2\pi\sigma_{l,k}^2) + \left(\frac{\overline{y}_{l,k} - y_{l,k}}{\sigma_{l,k}} \right)^2. \tag{6.12}$$

As the parameters are positive, the parameter estimation was performed in terms of the logarithm of the parameter θ. To further increase the performance of the optimization and the profile likelihood calculation we implemented the sensitivity equations for the one-dimensional model and used them to calculate the gradient and the Fisher information matrix of the objective function.

Simulated data As the simulated data is on the same scale for each image we can fix the scaling variable in the observation function to $s_l = 1$ for $l = 1, \ldots, N$ and we then should be able to estimate the underlying parameters at least for the spot and noise free data set. We performed the estimation on a single data set as scaling plays no role for the simulated data. We performed 100 local optimizations with varying starting points and between 80-95% of the starts converged to the same optimal likelihood value. During the analysis of the estimation we found three key results:

1. the distance measure is sensitive to the introduced bright spots in the data (see Figure 6.3 A)

2. the parameters $\alpha/(\gamma K_D)$ and S_0 are practically non-identifiable independent of the spot number and their product $S_0\alpha/(\gamma K_D)$ is only identifiable for a confidence level bigger then 7%. The true value is not included in the confidence interval.

3. the estimated parameter values for D/γ and b are well identifiable for low numbers of spots in the data.

To address the first problem it is sufficient to consider the distance map calculated from the filtered images. However, it showed that the distance measure without preprocessing of the image data is no reliable measure to estimate kinetic parameters or to justify a biological hypothesis. Following this observation we performed the parameter estimation task on the distance map obtained from the filtered CCL21 images.

The second result was obtained by calculation of the PL for the estimated parameters. The parameter $\alpha/(\gamma K_D)$ showed PLs, which did not decrease below the confidence level threshold for decreasing values of $\alpha/(\gamma K_D)$ and the likelihood profiles did not cross the threshold for s_0 increasing values. This hinted at the identifiability of the combination of both and we found that $S_0 \cdot \alpha/(\gamma K_D)$ showed a well defined optimum. The PL, however, never falls below the 5% confidence level threshold (see Figure 6.9) leading to infinitely extending confidence intervals. Furthermore, we see that the MLE varies by nearly an order of magnitude from the true value, which is not included in the obtained confidence intervals. As expected, the one-dimensional model is not able to explain the underlying two-dimensional process with respect to these parameters.
The third result becomes obvious if we consider the obtained parameter values shown in Table 1. For both D/γ and b the MLE is close to the true value for all numbers of introduced spots. Furthermore, both parameters show identifiable PLs, however, with increasing amount of introduced spots the filtering is no longer as efficient and also the amount of data left for the estimation decreases. This increases the variance at each spot number and finally leads to practical non-identifiability of the background fluorescence b. So the ability to estimate parameters from the distance measure strongly depends on the noise level of the measurements and the performance of the filtering.

CCL21 data For the real CCL21 data we performed the parameter estimation for the one-dimensional model on all 9 measurements simultaneously (see Figure 6.10). Following the results of the estimation process with the simulated data we filtered each image before calculating the distance measure. As the estimated parameter value for $S_0\alpha/(\gamma K_D)$ was expected to differ from the true underlying parameter we concentrated on the estimation results for D/γ and b. Furthermore, all images were on a strongly varying intensity level and we also estimated the scaling parameters s_j for $j = 1, 2, 3, 4, 12, 13, 14, 15, 16$ (see Section 6.1.2 for a discussion of the data properties). For the simulated data we found that the parameters $S_0\alpha/(\gamma K_D)$ and b become practically non-identifiable if the noise level increases and the key question we wanted to answer was if the noise level in the true images allows for identifiability of the parameters based on the distance measure.
We found that the combination of the nine filtered CCL21 images is sufficient for practical identifiability of D/γ with respect to a 5% confidence level. The confidence interval, however, spans several orders of magnitude. Hence we obtained a good guess for further estimation processes but we had no confidence that the value represents the true physical diffusion constant.

Table 6.1.: *Simulated data: estimated parameters for the one-dimensional source-diffusion-complex formation model* *Parameter estimation performed on the filtered simulated data. Each table displays the estimated parameters for the best fit out of 100 for the number of introduced spots. Note the difference between the true and estimated value of $S_0\alpha/\gamma K_D$.*

no spots

	$\log_{10}(\theta_i)$	$\log_{10}(\hat{\theta}_i)$	$\mathrm{CI}_{0.05}(\theta_i)$
D/γ	1.93	1.93	$[-0.15; 3.08]$
$S_0\alpha/\gamma K_D$	-1.30	-0.66	$[-1.02; \infty]$
b	-1.89	-1.88	$[-2.01; -1.83]$

 # spots: 20

	$\log_{10}(\theta_i)$	$\log_{10}(\hat{\theta}_i)$	$\mathrm{CI}_{0.05}(\theta_i)$
D/γ	1.93	1.94	$[-0.16; 3.17]$
$S_0\alpha/\gamma K_D$	-1.30	-0.66	$[-1.02; \infty]$
b	-1.89	-1.88	$[-2.03; -1.82]$

 # spots: 40

	$\log_{10}(\theta_i)$	$\log_{10}(\hat{\theta}_i)$	$\mathrm{CI}_{0.05}(\theta_i)$
D/γ	1.93	1.92	$[-0.19; 3.13]$
$S_0\alpha/\gamma K_D$	-1.30	-0.65	$[-1.03; \infty]$
b	-1.89	-1.88	$[-2.03; -1.83]$

 # spots: 80

	$\log_{10}(\theta_i)$	$\log_{10}(\hat{\theta}_i)$	$\mathrm{CI}_{0.05}(\theta_i)$
D/γ	1.93	1.94	$[-0.17; 3.14]$
$S_0\alpha/\gamma K_D$	-1.30	-0.66	$[-1.02; \infty]$
b	-1.89	-1.88	$[-2.03; -1.83]$

 # spots: 160

	$\log_{10}(\theta_i)$	$\log_{10}(\hat{\theta}_i)$	$\mathrm{CI}_{0.05}(\theta_i)$
D/γ	1.93	1.94	$[-0.25; 3.27]$
$S_0\alpha/\gamma K_D$	-1.30	-0.63	$[-1.04; \infty]$
b	-1.89	-1.88	$[-2.07; -1.83]$

 # spots: 320

	$\log_{10}(\theta_i)$	$\log_{10}(\hat{\theta}_i)$	$\mathrm{CI}_{0.05}(\theta_i)$
D/γ	1.93	1.92	$[-0.46; 3.40]$
$S_0\alpha/\gamma K_D$	-1.30	-0.59	$[-1.06; \infty]$
b	-1.89	-1.88	$[-2.12; -1.82]$

 # spots: 640

	$\log_{10}(\theta_i)$	$\log_{10}(\hat{\theta}_i)$	$\mathrm{CI}_{0.05}(\theta_i)$
D/γ	1.93	1.85	$[-2.14; 4.34]$
$S_0\alpha/\gamma K_D$	-1.30	-0.57	$[-1.32; \infty]$
b	-1.89	-1.85	$[-\infty; -1.76]$

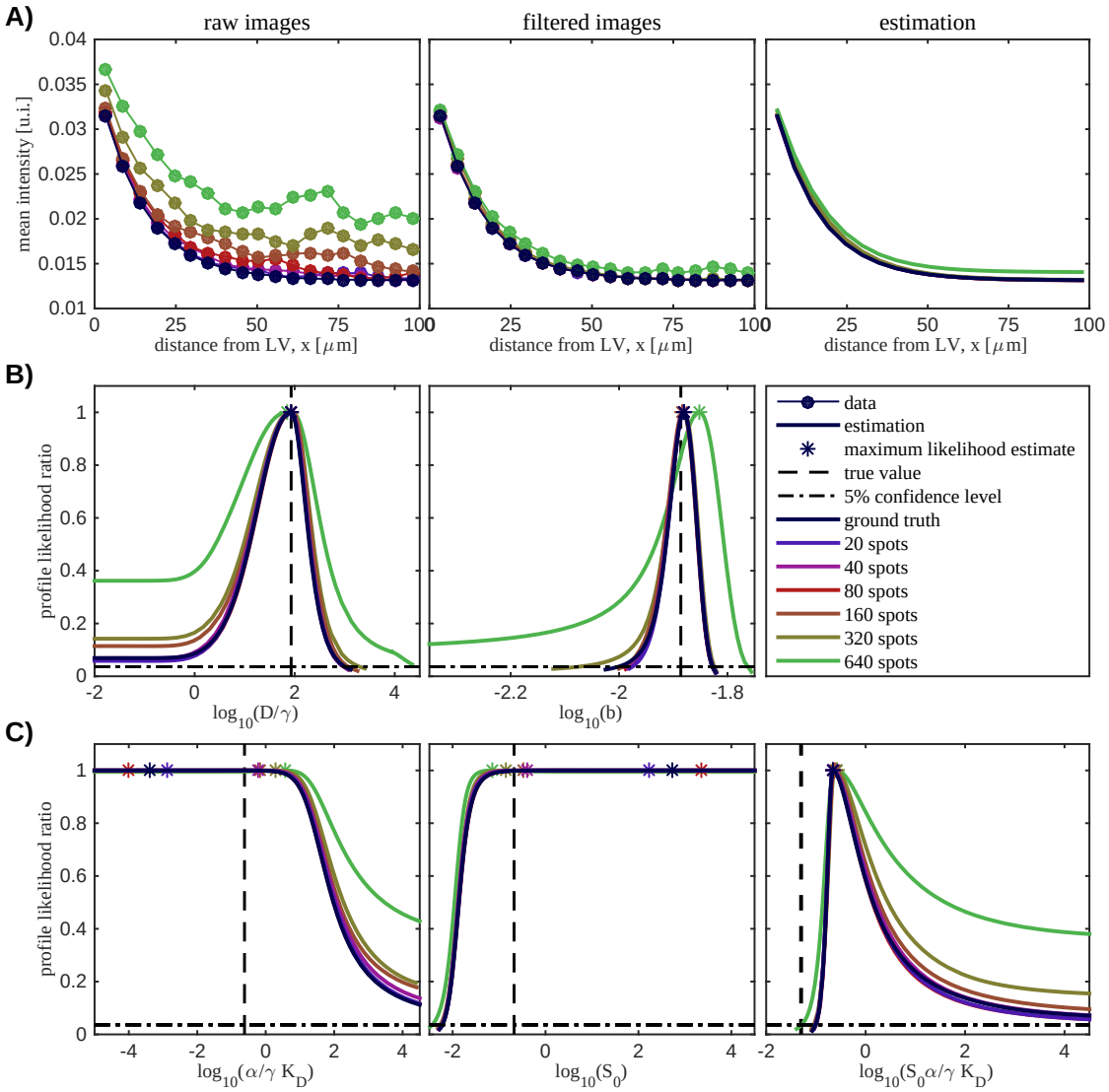


Figure 6.9.: ***Estimation results for the one-dimensional source-diffusion-complex formation model on simulated data.*** *A) Distance measure estimated from raw simulated images, from filtered images and the estimated distance measure. It becomes clearly visible how the increasing number of artifacts changes the slope of the curve for the unfiltered data while the filtered curves are nearly identical. Furthermore we see that the estimated curves match the data nicely and are similar for all spot numbers. B) Likelihood ration of the parameters D/γ and b. For both parameters the true value is nicely estimated, however, D/γ is practical non-identifiable for a 5% confidence level for all spot numbers. C) Likelihood ratio of the parameters $\alpha/(\gamma K_D)$ and S_0. We observed that both are practical non-identifiable but show a correlation. The third panel shows the likelihood ratio calculated for the product $S_0\alpha/(\gamma K_D)$, which is practical identifiable.*

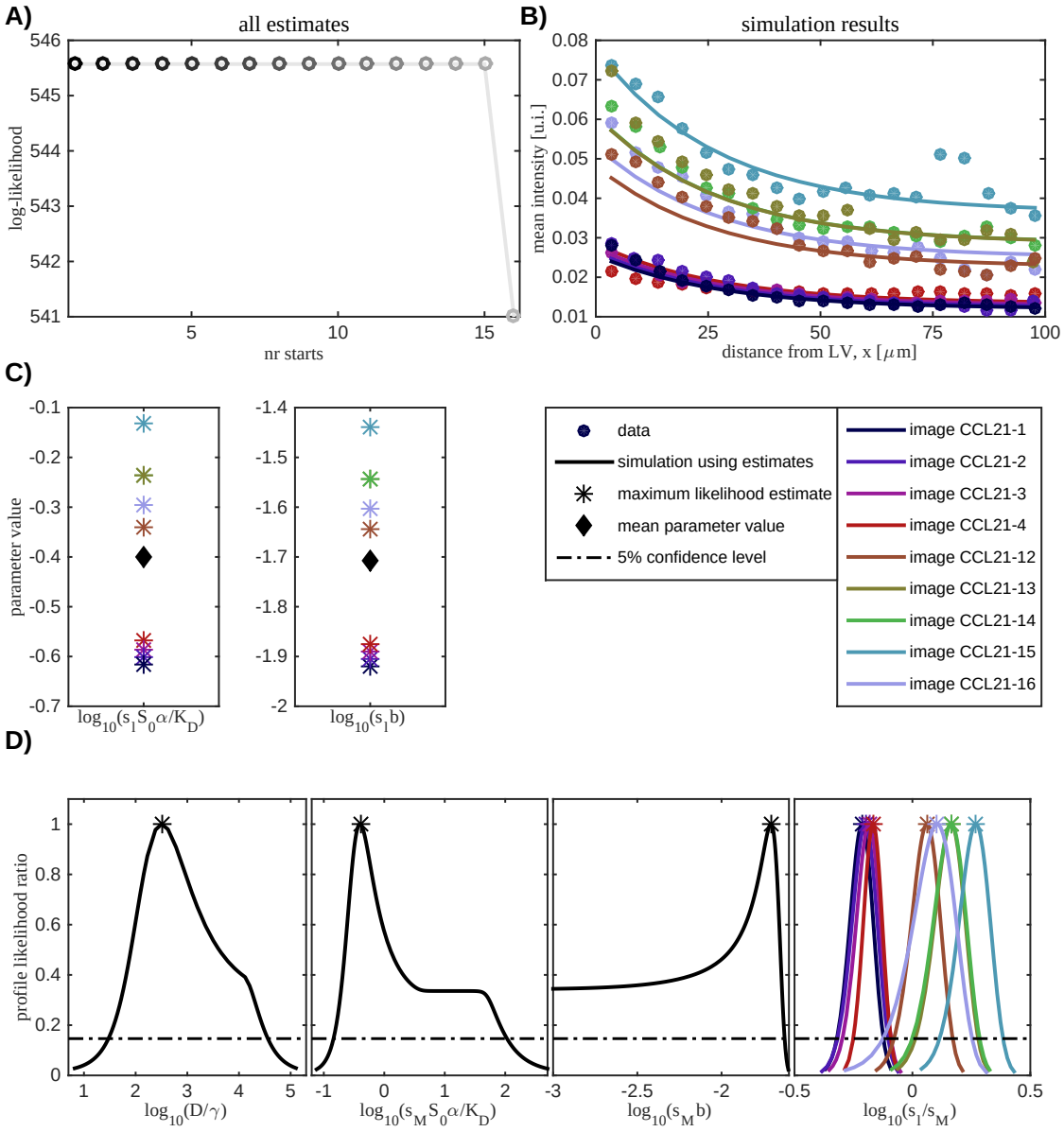


Figure 6.10.: ***Simultaneous estimation of the one-dimensional source-diffusion-complex formation model for the full CCL21 data set.*** *A) Out of 16 performed optimization starts 15 converged to the same log-likelihood value. Hence we assumed that the likelihood for the simultaneous consideration of the nine CCL21 images has a global optimum, which is reach by more than 90% of the optimization runs. B) The distance measure for each image as well as the distance measures obtained by the simulation with the maximum likelihood parameters. C) Scaled values for $S_0\alpha/(\gamma K_D)$ and b obtained from experiments 1-4 and experiments 12-16 cluster. Furthermore, they are either lower or higher than the mean value obtained from all experiments. This dependency is lost without the scaling factors and the mean parameter value yields a bad fit for all images. D) Likelihood profiles for the parameters D/γ, $s_{mean}S_0\alpha/(\gamma K_D)$, $s_M b$ and the sclaing factors s_l/s_M. s_M is the geometric mean of the scaling factors.*

Based on the nine images b is structural non-identifiable, however, the scaled background value $s_l b$ is structural identifiable and yields a background value for each individual image. As this scaling is a measurement specific constant we introduced the mean scaling for all images to obtain a mean background value. For the mean calculation we used the geometric mean as it is the intuitive choice if we consider the logarithm of the parameters. We observed that scaled value of $S_0\alpha/(\gamma K_D)$ is well defined for all images and all scaling factors in relation to s_{mean} are practical identifiable. In contrast, the scaled value for the background is practically non-identifiable. The estimated parameter gives us an initial guess for the parameter value but are not well determined and we have to proceed to a more complicated model to obtain reliable estimates of the true underlying parameter. The obtained values, however, are used in the following to obtain a good range for parameter guesses.

Two-dimensional model

To address the problem that the one-dimensional model is not sufficient to estimate all the kinetic parameters from the distance measure we calculated the distance measure for the two-dimensional source-diffusion-complex formation model introduced in Section 6.2.1. We considered the square domain $\Omega = [0, 604]^2[\mu\text{m}]$ and the observation operator for this case not only includes the background and the scaling but also the distance measure calculation applied to the state component c, which we denoted by $\text{DM}(c)$. We considered the observation

$$y_{l,k} = G_{l,k}(s, b) \begin{pmatrix} u \\ s \\ c \end{pmatrix} = s_l \cdot (b + \text{DM}(c(x))).$$

From the previous analysis we know that the structural identifiable parameters for full observations are D/γ, $\alpha/(\gamma K_D)$ and S_0. It is not directly clear how the observation operator including the distance measure influences the identifiability of the parameters. But again we can conclude that we can only estimate scaled values for S_0 and b. We consider the estimation process in terms of four kinetic parameters and the scaling parameters, i.e. $\theta = (D/\gamma, \alpha/(\gamma K_D), S_0, b, s_1, \ldots, s_N)$.
Regarding the noise model we make the same assumptions as for the one-dimensional model and again use the objective function given by (6.12).

Simulated data First we considered the simulated data to asses the quality of the estimated parameters. As the simulated data includes no scaling of the individual images, we fixed the scaling parameter $s_l = 1$ for $l = 1, \ldots, N$ to one. Learning from the previous analysis we considered the distance measure obtained from the filtered images for the estimation. We performed 25 multi-start local optimization runs and again obtained a good convergence of the optimization and assumed that we found a global optimal likelihood value.
Considering the estimation results we found that comparable with the one-dimensional model only the product of S_0 and $\alpha/(\gamma K_D)$ is identifiable with a broad PL. In contrast to the one-dimensional case the true value is contained in the confidence interval and the obtained MLE is close to it (see Figure 6.11). Furthermore, we observed that with a growing number of introduced spots the PL becomes broad so that the

Table 6.2.: ***Simulated data: estimated parameters for the two-dimensional source-diffusion-complex formation model*** *Parameter estimation performed on the filtered simulated data. Each table displays the estimated parameters for the best fit out of 25 for the number of introduced spots.*

no spots

	$\log_{10}(\theta_i)$	$\log_{10}(\hat{\theta}_i)$	$\mathrm{CI}_{0.05}(\theta_i)$
D/γ	-3.26	-3.34	$[-4.34; -1.77]$
$S_0\alpha/\gamma K_D$	-2.88	-1.24	$[-1.68; 2.15]$
b	-1.87	-1.86	$[-4.34; -1.78]$

 # spots: 20

	$\log_{10}(\theta_i)$	$\log_{10}(\hat{\theta}_i)$	$\mathrm{CI}_{0.05}(\theta_i)$
D/γ	-3.26	-3.34	$[-4.34; -1.77]$
$S_0\alpha/\gamma K_D$	-2.88	-1.24	$[-1.68; 2.15]$
b	-1.87	-1.86	$[-4.34; -1.78]$

 # spots: 40

	$\log_{10}(\theta_i)$	$\log_{10}(\hat{\theta}_i)$	$\mathrm{CI}_{0.05}(\theta_i)$
D/γ	-3.26	-3.51	$[-4.34; -2.51]$
$S_0\alpha/\gamma K_D$	-2.88	-1.16	$[-1.63; 1.23]$
b	-1.87	-1.86	$[-2.01; -1.81]$

 # spots: 80

	$\log_{10}(\theta_i)$	$\log_{10}(\hat{\theta}_i)$	$\mathrm{CI}_{0.05}(\theta_i)$
D/γ	-3.26	-3.49	$[-4.34; -2.50]$
$S_0\alpha/\gamma K_D$	-2.88	-1.18	$[-1.63; 1.19]$
b	-1.87	-1.86	$[-2.01; -1.81]$

 # spots: 160

	$\log_{10}(\theta_i)$	$\log_{10}(\hat{\theta}_i)$	$\mathrm{CI}_{0.05}(\theta_i)$
D/γ	-3.26	-3.51	$[-4.34; -2.43]$
$S_0\alpha/\gamma K_D$	-2.88	-1.15	$[-1.65; 1.35]$
b	-1.87	-1.86	$[-2.04; -1.81]$

 # spots: 320

	$\log_{10}(\theta_i)$	$\log_{10}(\hat{\theta}_i)$	$\mathrm{CI}_{0.05}(\theta_i)$
D/γ	-3.26	-3.48	$[-4.34; -2.38]$
$S_0\alpha/\gamma K_D$	-2.88	-1.14	$[-1.64; 1.50]$
b	-1.87	-1.86	$[-\infty; -1.81]$

 # spots: 640

	$\log_{10}(\theta_i)$	$\log_{10}(\hat{\theta}_i)$	$\mathrm{CV}_{0.05}$
D/γ	-3.26	-3.34	$[-4.34; -1.77]$
$S_0\alpha/\gamma K_D$	-2.88	-1.24	$[-1.68; 2.15]$
b	-1.87	-1.86	$[-\infty; -1.78]$

PL no longer decreases below the confidence level threshold in the parameter region considered. For increasing spot numbers the parameter b became practically non-identifiable as in the one-dimensional case.
Considering the estimated parameter values shown in Table 6.2 we saw that the MLE is always close to the true value for each parameter. This is promising as it shows that a simplistic feature as the considered distance measure is sufficient to obtain an estimate of the parameters if the structure of the LV is known and considered in the model. So the analysis of the distance measure is a good initial analysis for the process. If we consider the confidence intervals for D/γ we see that they vary over a range of several orders of magnitude, showing that even though the parameter value can be estimated it is not a precise estimate and needs further considerations. The obtained values can be used as rough initial guesses to perform a more sophisticated method as introduced in the following sections.

CCL21 data For the real CCL21 data we again performed the parameter estimation for the two-dimensional model on all nine measurements simultaneously and we considered the distance measure obtained from the filtered images. The results are shown in Figure 6.12. Following the results from the simulated data we expected to be able to estimate D/γ and b as well as $S_0\alpha/(\gamma K_D)$. As all images are on a strongly varying intensity level and we also estimated the scaling parameters s_l for $l = 1, 2, 3, 4, 12, 13, 14, 15, 16$.
We found that the combination of the nine filtered CCL21 images is sufficient to estimate the parameter D/γ. Furthermore, its confidence interval corresponding to the 5% confidence level is one order of magnitude smaller than in the one-dimensional case. Thus we obtained a better guess for further estimation processes but we still have no confidence that the MLE represents the true physical diffusion constant. Based on the nine images we could only identify scaled values for b and $S_0\alpha/K_D$ and again we used the geometric mean denoted by s_{mean} as mean scaling. Based on the mean scaling we obtained for b the schematically same practical non-identifiable profile as in the one-dimensional case. For $S_0\alpha/(\gamma K_D)$, however, we obtained a defined profile, which shows that it is identifiable with a reasonably tight confidence interval.
Finally if we compared the one-dimensional and the two-dimensional model for the estimation from all nine measurements. We found that the two-dimensional model yields a slightly higher likelihood value. If we quantified the evidence for the two-dimensional model in contrast to the one-dimensional model in terms of the BIC we observed that the difference between both models is barely mentionable. Thus we concluded that a model selection on both models with the nine measurement images would not suggest the use of the more complicated and computationally expensive two-dimensional model. Our additional analysis with simulated data, however, showed that the parameter values for $S_0\alpha/(\gamma K_D)$ can only be estimated with the two-dimensional model. The combined analysis showed that the two-dimensional model is the right choice even if the measurements are only a one-dimensional feature.

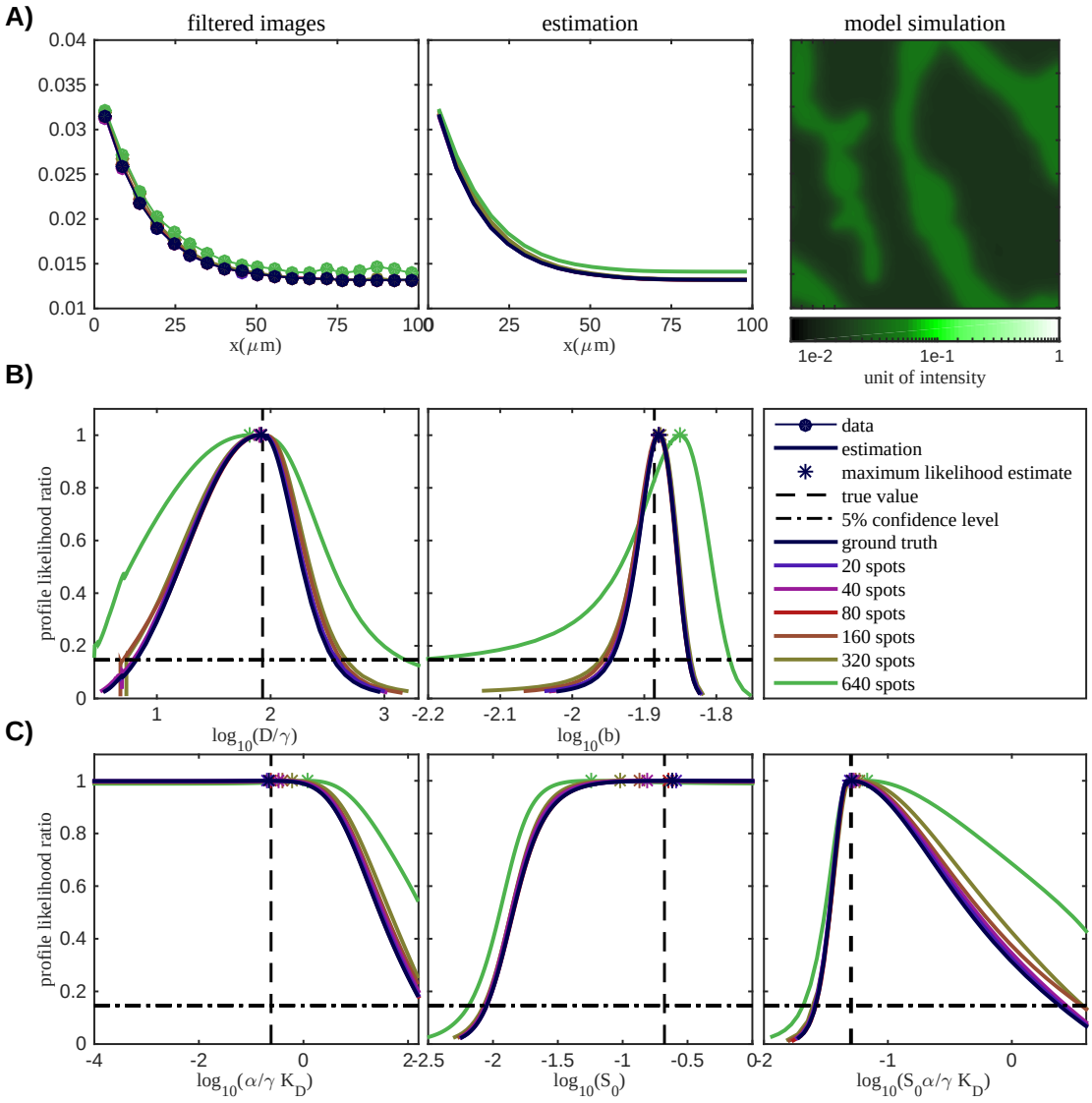


Figure 6.11.: ***Estimation results for the two-dimensional source-diffusion-complex formation model and the distance measure for simulated data.*** *A) Distance measure calculated from filtered images was used as data and the estimated distance measure obtained from the two-dimensional model simulation. We observed that the estimated curves match the data nicely and are similar for all spot numbers. The third panel shows one representative two-dimensional model simulation for the estimated parameters in case of 640 spots. B) Likelihood ration of the parameters D/γ and b. For both parameters the true value is nicely estimated. In comparison to the one-dimensional model the confidence intervals become smaller and both parameters are practical identifiable for all spot numbers. C) Likelihood ratio of the parameters $\alpha/(\gamma K_D)$ and S_0. We observed that both are practical non-identifiable but show a correlation. The third panel shows the likelihood ratio calculated for the product $S_0\alpha/(\gamma K_D)$, which is practical identifiable.*

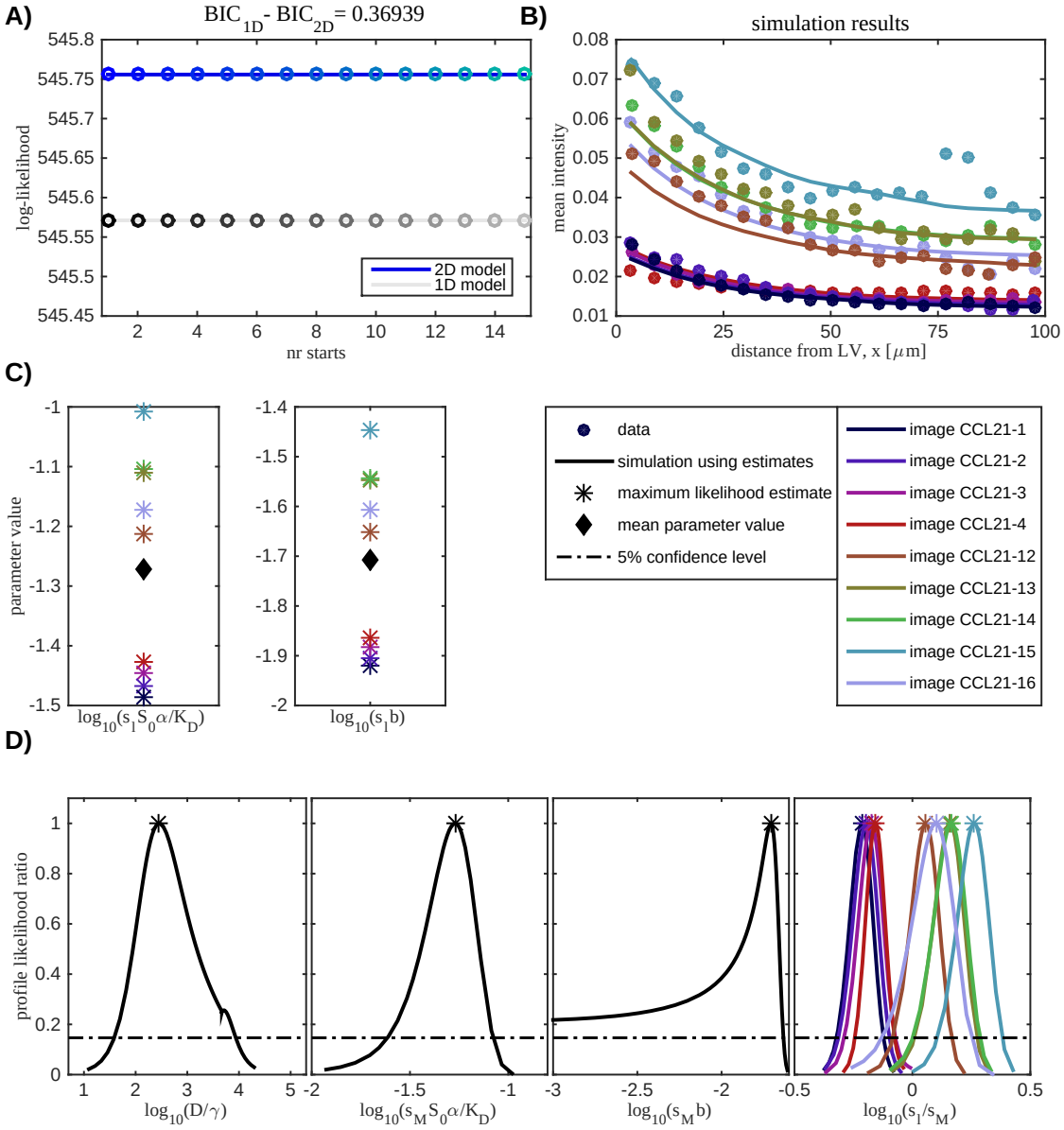


Figure 6.12.: ***Simultaneous estimation of the two-dimensional source-diffusion-complex formation model for the full CCL21 data set.*** *A) Comparison of the maximal log-likelihood results from 15 optimization starts. We observe that the two-dimensional model yields a slightly better likelihood value than the one-dimensional model. If we consider the difference in the BIC, however, the evidence for the two-dimensional model is "not worth more than a bare mention" B) The distance measure for each image as well as the distance measures obtained by the simulation with the maximum likelihood parameters. C) Scaled values for $S_0\alpha/(\gamma K_D)$ and b. We observed that experiments 1-4 and experiments 12-16 cluster. D) Likelihood profiles for the parameters D/γ, $s_M S_0\alpha/(\gamma K_D)$, $s_M b$ and the scaling factors s_l/s_M. s_M is the geometric mean of the scaling factors. $s_M b$ shows the same behavior as in the one-dimensional case.*

6.3.3. Full image

In the previous sections we estimated the kinetic parameters of the source-diffusion-complex formation model introduced in Section 6.2.1 from the distance measure calculated based on the data. This distance measure is a strong abstraction of the original two-dimensional image data, which is suitable to visualize the existence of a gradient. To overcome the deficiencies of the estimation results, namely the problems to recover the true underlying parameters, the wide confidence intervals and the missing spatial sugar distribution we addressed the question how to estimate the kinetic parameters considering the full image data. A special focus was on the treatment of the bright spots during the estimation process. We considered two approaches: first we estimated the parameters from a filtered image and second we considered a noise model, which includes the bright spots as additional noise term and considers the whole image as introduced in Section 3.1.2.

As we considered the two-dimensional image data we only concentrated on the two-dimensional model. Hence the state of the system is u,s and c in the square domain $\Omega = [0, 604]^2[\mu\text{m}^2]$. For the full images we consider an offset b and a scaling parameter $s = (s_1, \ldots, s_N)$ with N number of measurements. The l-th observation operator for the full image maps the state onto the integral over each of the pixels $\mathcal{A}_k$, $k = 1, \ldots, M$, of the image as introduced in Section 3.1.1.

$$y_{l,k} = G_{l,k}(s, b) \begin{pmatrix} u \\ s \\ c \end{pmatrix} = s_l b + \int_{\mathcal{A}_k} s_l c(x) dx \quad \text{for } l = 1, \ldots, N,\ k = 1, \ldots, M.$$

For high image resolutions in comparison to the discretization grid size the integral can be neglected and only the interpolation of the state onto the pixel grid is considered. Following the analysis in Section 6.2.1 we knew that the parameters D/γ, $\alpha/(\gamma K_D)$, S_0 and b can be identified for full noise free observations. With the introduced scaling we can only identify $s_l S_0$ and $s_l b$ for each experiment.

Concerning the measurement technique and the resulting errors on the images we deduced a noise model. We assumed that taking a microscopy picture is a counting process and the noise on the final image is multiplicative and log-normally distributed, i.e.

$$\overline{y}_{l,k} = y_{l,k} \cdot \varepsilon_{l,k} \text{ with } \varepsilon_{l,k} \sim \mathcal{LN}(0, \sigma_l^2) \quad \text{for } l = 1, \ldots, N,\ k = 1, \ldots, M.$$

It follows that $\overline{y}_{l,k} \sim \mathcal{LN}(y_{l,k}, \sigma^2)$ and the reduced, negative log-likelihood function is given by

$$j(\theta) = \frac{1}{2} \sum_{l=1}^{N} \sum_{k=1}^{M} \log(2\pi\sigma_{l,k}^2 \overline{y}_{l,k}) + \left(\frac{\log(\overline{y}_{l,k}) - \log(y_{l,k})}{\sigma_{l,k}} \right)^2. \tag{6.13}$$

As before, we performed the parameter estimation in terms of the logarithm of the parameters $\log(\theta) = \log((D/\gamma, \alpha/(\gamma K_D), S_0, b, s_1, \ldots, s_N, \sigma_1, \ldots, \sigma_M))$. Additionally we considered the sensitivity system of (6.5) to calculate the gradient and Fisher information matrix of the objective function.

Filtered data First we considered the parameter estimation on filtered data, which removed the bright spots. The filtering procedure introduced in Section 6.1.3 returned a mask of pixels, which are to be removed from the data. Consequently we considered a reduced image with $k = 1, \ldots, M_f$ pixels. For the estimation process we could directly use the objective function (6.13) without consideration of a different observation operator only a reduced number of pixels.

Integrated noise model Second we considered an adaptive noise model considering a structured noise given by the bright spots observed in the data. We assumed that our data is distributed as the mixture of two processes: a model dependent one and an independent one. As we had no knowledge about the structured noise or its distribution we could not directly deduce a noise model for the two processes, however, we required that in the case of no side processes the noise model for the standard process applied. Therefore we considered the distribution of the ration $\overline{y}_{l,k}/y_{l,k}$ to be the weighted sum of a log-normal distribution and an additional noise term, i.e.

$$\frac{\overline{y}_{l,k}}{y_{l,k}} \sim w_1 \mathcal{LN}(1, \sigma^2) + w_2 P_2.$$

We assumed that the noise term is also log-normally distributed with unknown mean and variance, $P_2 = \mathcal{N}(\mu_2, \sigma_2^2)$, i.e.

$$j(\theta) = -\log \prod_{l,k=1}^{N,M} \left(\frac{w_1 \exp\left\{ -\frac{(\log(\overline{y}_{l,k}) - \log(y_{l,k}))^2}{2\sigma_1^2} \right\}}{\sqrt{2\pi}\sigma_1 \overline{y}_{l,k}} + \frac{w_2 \exp\left\{ -\frac{(\log(\overline{y}_{l,k}) - \mu_2)^2}{2\sigma_2^2} \right\}}{\sqrt{2\pi}\sigma_2 \overline{y}_{l,k}} \right).$$

Performance evaluation on simulated data

First, we considered the performance of the methods for the simulated data. We assumed that $s_l = 1$ and $\sigma_l = \sigma$ for $l = 1, \ldots, N$. To define a set of matching parameter bounds and a good initial mesh for discretization we used the estimated parameters obtained from the fit to the distance measure of the two-dimensional model. We performed 25 local multi-start optimizations with the raw data, the filtered data and the adaptive noise model applied to the raw data for each of 30 replicates individually. For all methods we obtained a good convergence of the optimization to a single optimal objective function value. In the following we will compare the performance of all three approaches in terms of the parameter identifiability, the obtained distance measure and the bias of the estimated parameters due to the introduced bright spots.

Before we could draw conclusions concerning the bias of the estimates we had to analyze if the considered parameter is identifiable. To study the identifiability of the parameters we used two methods: the bootstrap method and the profile likelihood method. As we have generated the simulated data in a way such that each replicate is based on the same simulation with a different realization of the noise model, the estimates obtained for each of the 30 replicates correspond to the estimates needed to obtain the bootstrap distribution of the parameters. For one of the estimates

we calculated the profile likelihood for each parameter. We found that indeed all the considered parameters are identifiable for all three approaches even though the width of the profiles strongly varies depending on the number of introduced spots.

Next we compared the approaches with respect to the ability to describe the process. Therefore, we first considered the difference between the images and the simulation. We observed that by eye there is no clearly visible difference between the result obtained by the three approaches (see Figure 6.13A). Next we analyzed the residual error distribution and compared it to the noise model used in the estimation process. We found that for the unfiltered, raw images the obtained residual distribution is not in good accordance with the residuum distribution calculated from the images. Hence, the assumption of normal distributed noise is indeed not well suited for this problem. Furthermore we observed that the filtering as well as the integrated noise model approach yield residual distribution, which are well in accordance with the residuum obtained from the data. Last we calculated the distance measure for uncorrupted data as well as for the data with bright spots. We found that the distance measure obtained from the raw images describes the corrupted data quite well but is far away from the real underlying process. In comparison both filtering and integrated noise model approach perfectly reconstructed the distance measure for the uncorrupted data from the images with spots. Figure 6.13 shows this results for the images with 320 added spots.

Finally we considered the bias in the estimated parameters due to the spots in the images. It is clear that the parameters obtained from the raw images start to deviate from the true value with increasing spot number. We observed that the background parameter b and the noise parameter σ are strongly overestimated in the raw images and for more than 40 introduced spots the true value can no longer be found in the 95% CI. For the three other parameters the CIs become wide but the true parameter can still be found in the 95% CI. For all parameters the CI obtained for the filtered images are much smaller than for the raw images. As expected the deviation from the true values is much smaller than for the raw images. However, it still shows the same trend. For the parameters $\alpha/(\gamma K_D)$, S_0 and σ^2 the true value is no longer included in the 95% CI for spot numbers bigger than 160. Of course this could be addressed with a filtering tailored especially to those images with more than 160 spots, however, here we wanted to compared the methods without prior knowledge of the spot numbers in the images. With the adaptive noise model approach we obtain even tighter CI and we see that up to the maximal number of 640 spots we can recover the parameters D/γ, b and σ^2. Especially interesting is that the underestimation of the diffusion length D/γ observable for the raw and filtered images can not be found for the adaptive noise model. For $\alpha/(\gamma K_D)$ and S_0 we observe a deviation from the true value and for spot numbers larger than 320 spots the true value is no longer included in the 95% CI.

Performance evaluation on CCL21 data

Next we investigated the performance of the three methods on the real images. This was done in two steps: first, we considered a single measurement (CCL21-1)

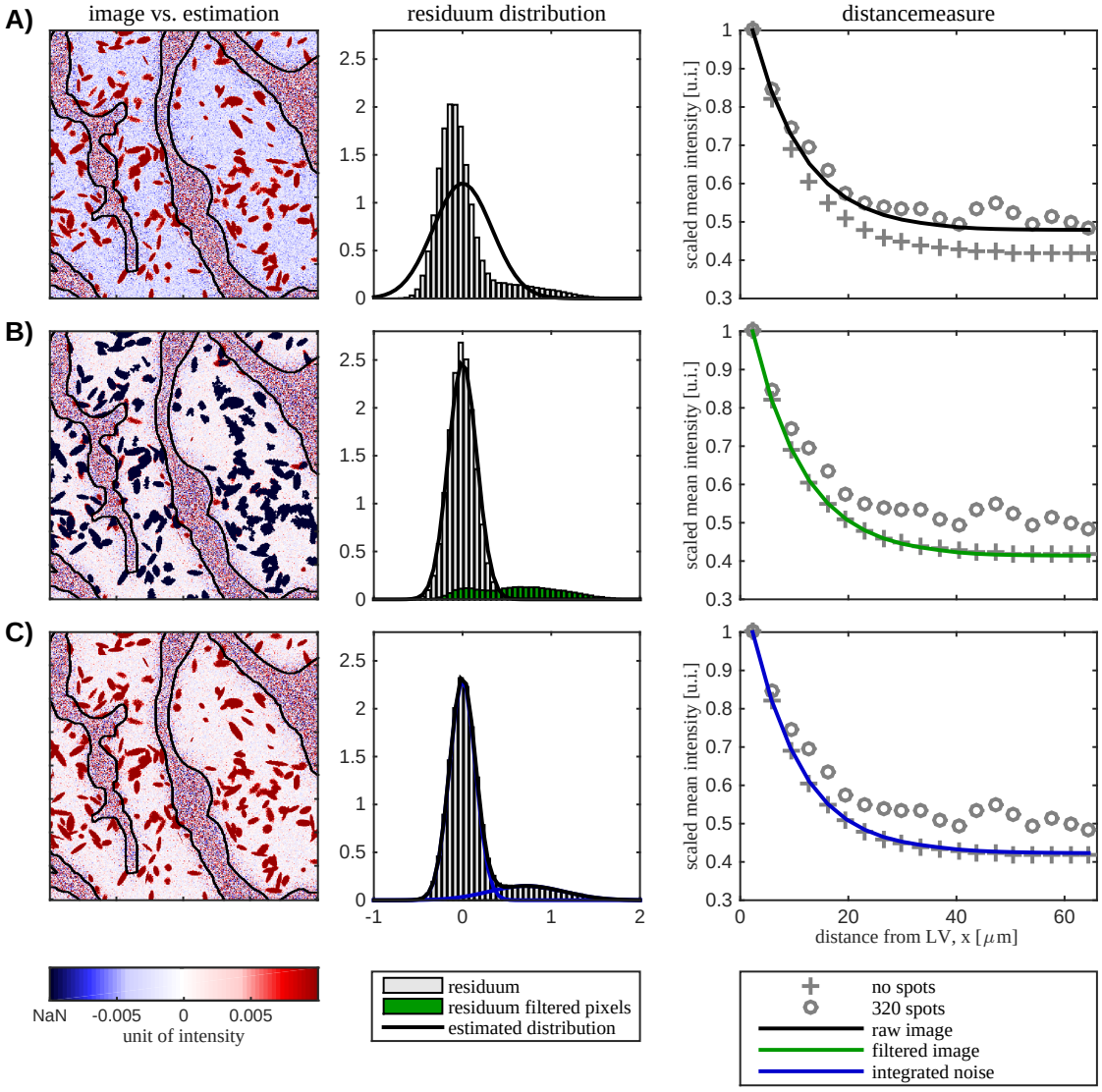


Figure 6.13.: ***Comparison of estimation results based on the simulated data with 320 spots.*** *A-C) First panel: difference between the raw image and the image obtained by the estimation. Second panel: residuum distribution and the estimated noise distribution. Third panel: distance measure obtained for the data with spots, the true value without spots as well as the one obtained from the estimation. A) Estimation performed on the raw, untreated images. We observe that the background is overestimated due to the introduced spots as visible in the first panel and confirmed by the distance measures shown in the last panel. Furthermore, panel two shows that the assumption of normal distributed noise is not valid for this estimation. B) Estimation performed on the filtered images. Dark spots in the first panels are filtered regions, which are not considered in the parameter estimation. The second panel shows that the assumed normal distribution and the residuum match well and the third panel shows that the obtained distance measure is a near perfect reconstruction of the true value. C) Estimation performed using the integrated model based method.*

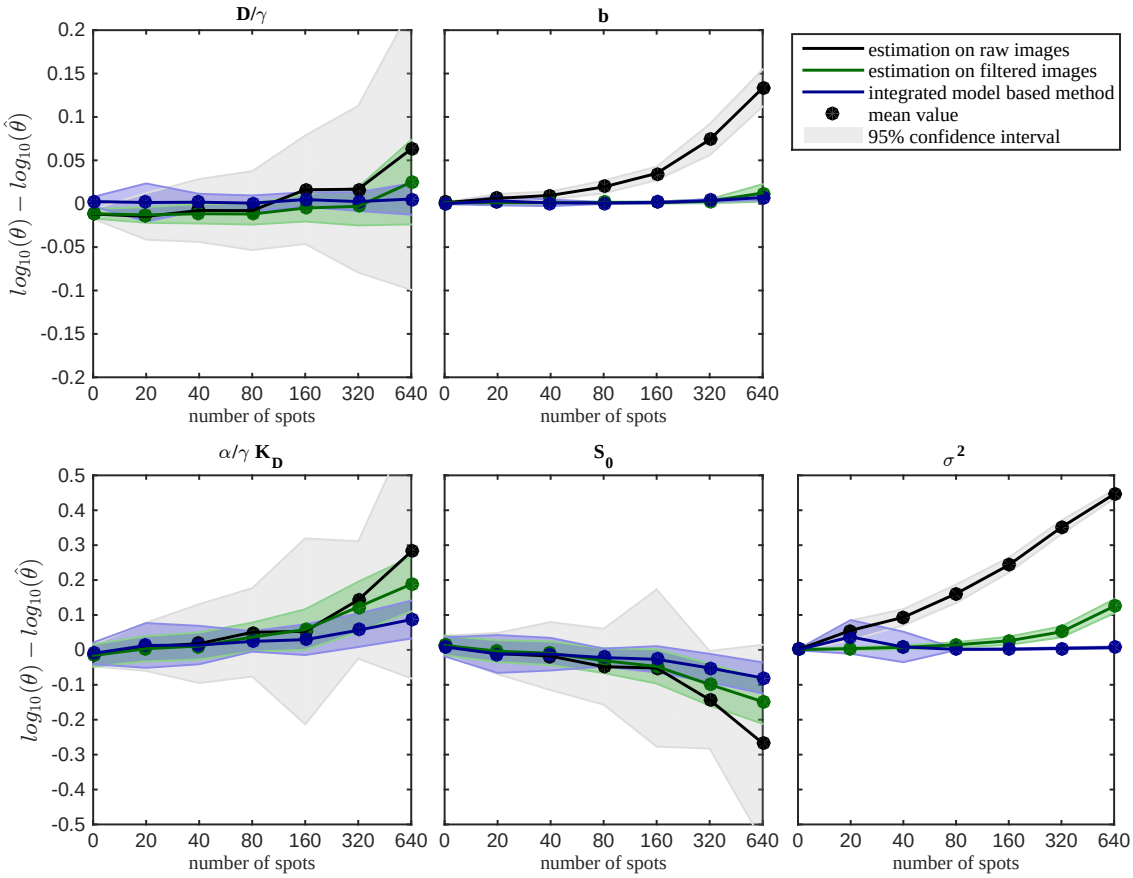


Figure 6.14.: ***Deviation of estimates from true value for simulated data.*** *Based on 30 samples for each spot number of the simulated data we estimated the parameters and calculated the deviation from the true value. Circles mark the mean deviation of the estimated values and the bands are the corresponding 95% confidence intervals calculated from the 30 samples. The integrated noise model outperforms the filtering with respect to all parameters if the number of spots is high.*

and compared the performance of the three methods visually and with respect to the distance measure. Second, we performed the parameter estimation on all nine images simultaneously and compared the methods regarding the likelihood values, parameter value and identifiability of the parameters.
Our analysis of the estimation based on the image CCL21-1 showed similar results as we obtained for the simulated data (see Figure 6.13 and Figure 6.15). For the real data filtering and adaptive noise model both cannot completely detect the spots. It is however sufficient to reduce the estimation bias for the background value (see Figure 6.15A). Naturally, a residuum distribution for the real images is not as smooth as for the simulated ones and the noise distribution is not as close a fit. However, we still see an improvement for filtering and adaptive noise model. As we have no noise free data to compare the distance measures, we calculated the distance measure for the filtered images as reference for the filtering approach. For the adaptive noise model we defined a weighted distance measure based on weighted data. For the distance map defined by (6.1) for a given set of distances $r_0, \ldots, r_n$ the pixels in the

i-th bin are denoted by $M_{r_i} = \{x \in \Omega | r_{i-1} < dm(x) < r_i\}$ and the weighted mean intensity is

$$\frac{\sum_{x \in M_{r_i}} \omega_k \overline{y}_{l,k}}{\sum_{x \in M_{r_i}} \omega_k}$$

with

$$\omega_k = \frac{\exp\left(-\frac{\log(\overline{y}_{l,k}) - \log y_{l,k})^2}{2\sigma_{l,k}}\right)}{\sqrt{2\pi}\sigma_{l,k}\overline{y}_{l,k}}.$$

We observed that both distance measures are well captured by the respective methods.

Next, we considered the performance of the parameter estimation based on the simultaneous consideration of all nine images. In this case we had 4 kinetic parameters and nine scaling parameters in all approaches. For raw images and filtered images we had additional nine noise parameters (in summary 22 parameters); in contrast, the integrated noise model considered on top of the kinetic and scaling parameters 32 noise parameters (in summary 45 parameters). In a pre-study this difference in the number of parameters resulted in convergence problems for the integrated noise approach. We addressed this problem by parameter optimization for the individual images to obtain an initial guess for the noise parameters. Subsequently we set bounds such that the noise parameters can vary within two magnitudes around the initial guess. This resulted in a good convergence and for all methods more than 75% of the optimization converged to the same optimal value with parameter values insight the bounds. We assumed that all methods found a global optimal value. Based on the AIC/BIC introduced in Section 2.2 there is a strong evidence that the integrated noise model is better suited for the CCL21 data then the estimation from raw images. The filtering approach and the integrated noise model, however, cannot be compared by AIC/BIC as they use different data sets. With respect to the obtained parameter values we found that they are similar and vary within less then one order of magnitude between the models. Additionally, the kinetic parameters are identifiable with all approaches considered.
Summarizing the findings we can say that parameter estimation based on the raw images in the presence of measurement artefacts like the introduced spots is not recommended as it yields unreliable parameter estimates. Both proposed methods, the filtering and the integrated noise model, performed considerably better in reconstructing the true parameter values. Especially, for the simulated data we showed that the integrated noise model also outperformed the filtering approach as the true parameters could be recovered for all numbers of introduced spots. However, it has to be kept in mind that the integrated noise model was designed especially for data with the spot type introduced in the simulated data. In general we found that without a filtering especially tailored to each image, the integrated noise model always outperformed the filtering for all spot numbers as well as on the real data. So choosing one method over the other is a trade-off between the know-how and time needed for the manual adaptation of the filtering method and the computational complexity of the noise parameter estimation. For the considered model the integrated noise model is the method of choice and we will apply it for the model selection in the following section. We saw, however, that the increased number of noise parameters

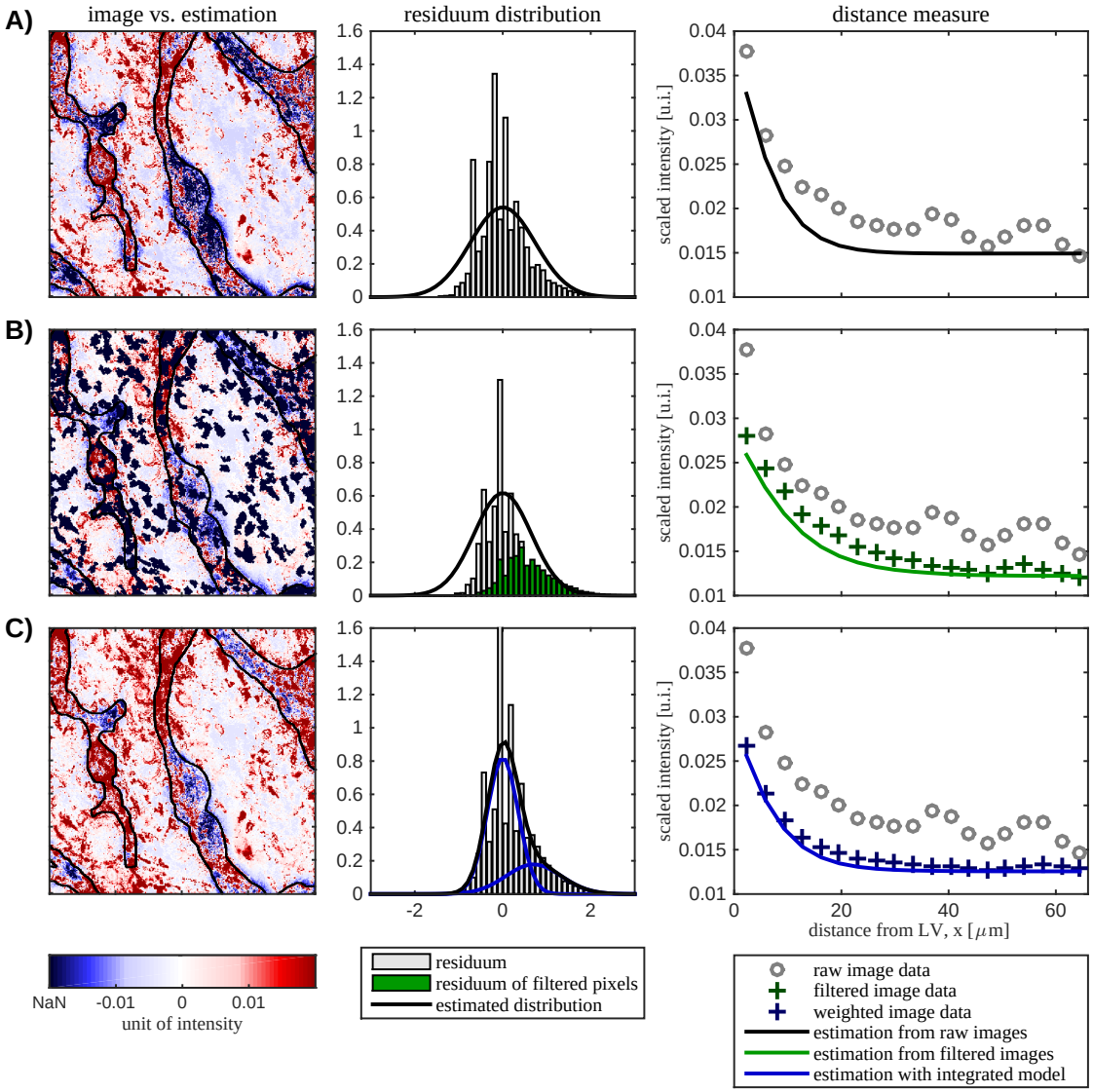


Figure 6.15.: ***Comparison of estimation results based on the measurement data CCL21-1.*** *A-C) First panel: difference between the data and the estimation. Second panel: residuum distribution and the estimated noise distribution. Third panel: distance measure obtained for data CCL21-1 and from the estimation. A) Estimation performed on the raw, untreated images. B) Estimation performed on the filtered images. Dark spots in the first panels are filtered regions, which are not considered in the parameter estimation. C) Estimation performed using the integrated noise model based method. In this case the background is not over estimated and we obtain the best match between estimated distance measure and the distance measure obtained by weighting the pixels of the image according to the estimated noise model.*

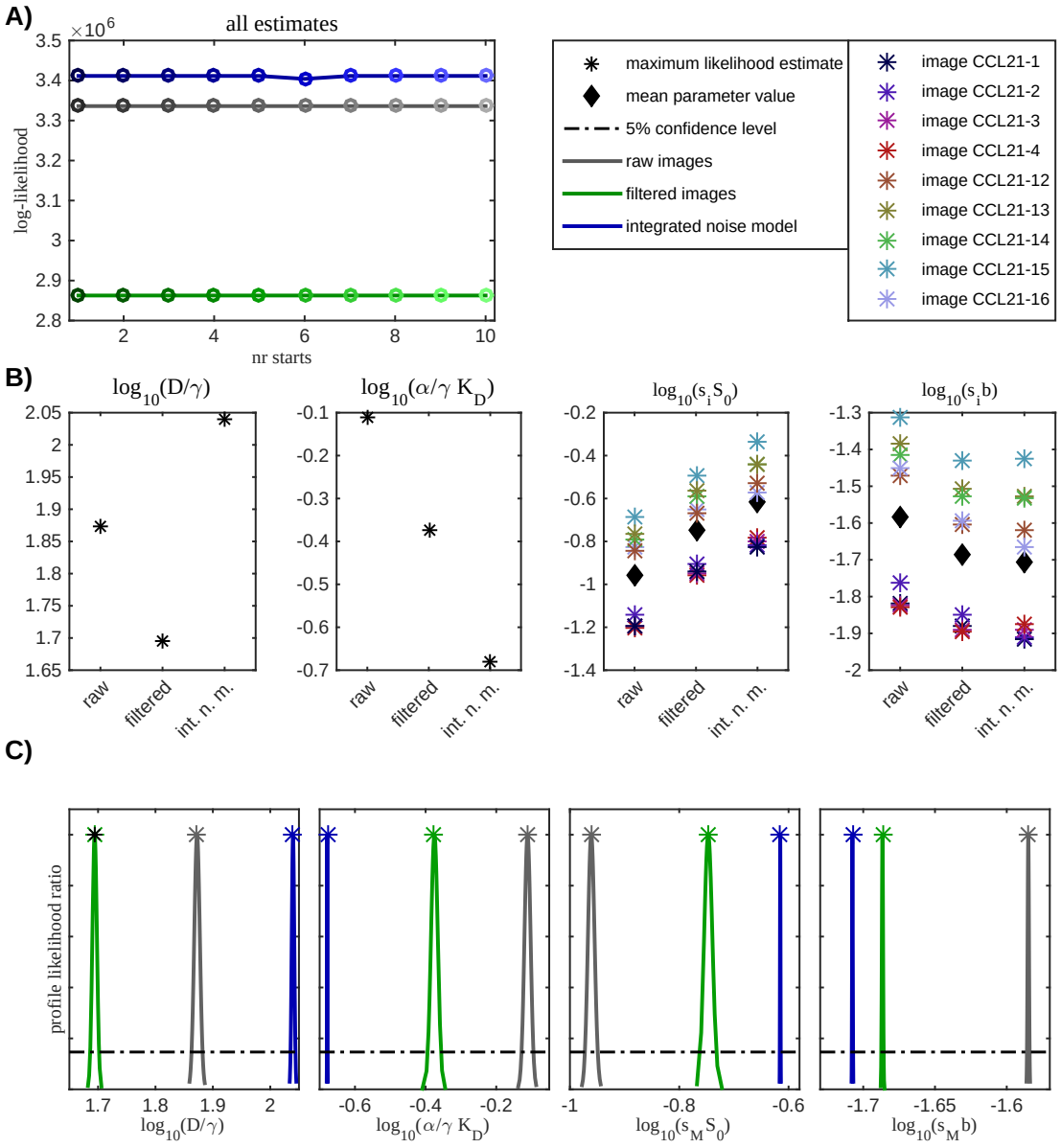


Figure 6.16.: ***Comparison of estimation results based on the simultaneous consideration of the nine CCL21 images*** *A) Comparison of the maximal log-likelihood results from 15 optimization starts. We observe that the estimation on raw and filtered images as well as the integrated noise model converged to a global optimum in all runs. B) Shows the estimated parameters D/γ and $\alpha/\gamma K_D$ and scaled values for S_0 and b. D) Likelihood profiles for the parameters D/γ, $\alpha/\gamma K_D$,$s_M S_0$, $s_M b$. s_M is the geometric mean of the scaling factors.*

of this approach scales with the number of considered measurements and can yield convergence problems for the optimization process. Possible adjustments to increase the convergence rate of the optimization for the integrated noise model are discussed in the Conclusion and Outlook section of this chapter.

6.4. Model selection

As mentioned in Section 6.2 we found that during the estimation process the estimated CCL21 concentration and the measured concentration considerably varied especially at the LV. This is a biologically interesting point as it might suggest a special marking of different LV regions with the chemokine signal (Kilarski *et al.*, 2013). Such differences in the bound CCL21 could be influenced by the concentration of sugar in those regions, which controls the amount of bound CCL21. In this section we want to methodically address this biological question by model selection considering three models:

- Model 1: no difference of sugar between LV and surrounding tissue (see Section 6.2.1).
- Model 2: different sugar concentration at the LV regions (see Section 6.2.2).
- Model 3: varying sugar concentration in each LV segment in the images (see Section 6.2.3).

We want to select the best model based on the Bayesian information criterion

$$BIC = -2J(\hat{\theta}) + \log(n)k$$

with n number of data points, k number of parameters and $\hat{\theta}$ is the MLE. For the model selection we proceeded in three steps: comparison of one single image by eye, comparison of all images by BIC and comparison of the simultaneous estimation of the nine images by BIC. For the estimation process we considered the filtering method as we saw in the previous section that the integrated noise model has convergence problems for high numbers of parameters. For the model selection we consider models with up to 33 kinetic parameters.

First, we considered the single image CCL21-3. In this image we observed a background region with low signal as well as LV regions with a high or a low intensity (see Figure 6.17A). We expected that Model 2 would be able to account for the change in the background and Model 3 would be able to account for all three phenomena. To compare models we also considered the distance measure of the estimates. This feature, however, is such a crude simplification of the data, that no difference between the models is visible by eye (see Figure 6.17B).

Second, we performed a parameter estimation for all three models on each of the nine images individually. Thus we could determine if the choice of the best model strongly would strongly vary between the images. Such a variance could be a biological dependence due to different regions in the mouse ear. For this comparison the difference of the parameter number between the models is maximally 5 and thus the

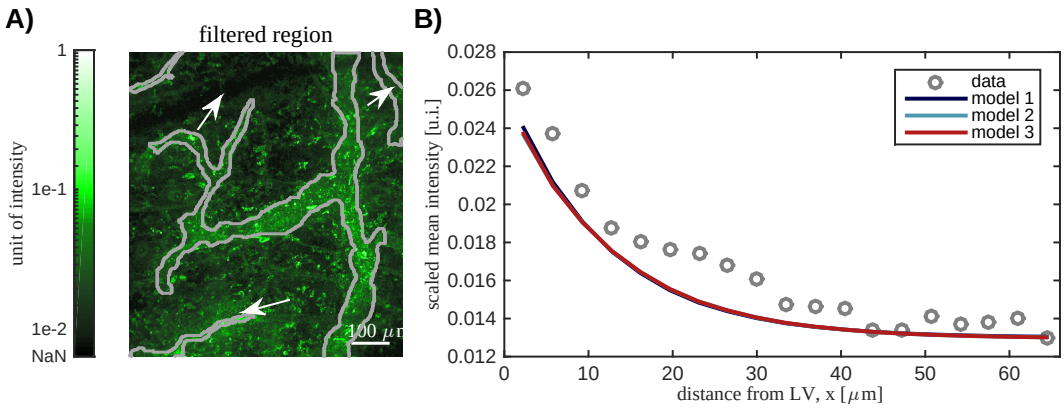


Figure 6.17.: ***Comparison of model performance with respect to image CCL21-3*** *A) Image CCL21-3 with overlay of LV mask. The LV regions with high and low intensity levels, as well as the low intensity background region we can observe in this image are marked with arrows. B) Distance measure calculated for the best fit parameters for the three models. By eye there is no visible difference in the distance measure. Hence it is inadequate to perform the model selection on this feature.*

penalization of the parameter number in the BIC is relatively small. We saw that except for image CCL21-16 there is always a strong evidence to choose Model 3 and not Model 1 and Model 2. For the image CCL21-16 the evidence is considerably lower but the choice remains the same (see Figure 6.18B). In image CCL21-16 we only observed one big LV and a small segment of another LV (see Figure 3), hence the increase in parameters of Model 3 is penalized strongly by the BIC in favor of Model 2.

Third, we performed the parameter estimation for all three models simultaneously on all nine images. The multi-start optimization of all three models nicely converged to a single optimal value and with respect to the reached maximum likelihood value Model 3 outperforms the other models (see Figure 6.18A). Here, the parameter number for Model 2 is only one higher than in Model 1. For Model 3, however, we considered 29 additional parameters in comparison to Model 1. Thus, the parameter penalization term of the BIC has a strong impact in this model selection process. The huge difference in the obtained likelihood value, however, counterweighted the parameter number and we see that there is a strong evidence for Model 3. Hence the selection for the single images also holds true in case of the simultaneous consideration of all nine images (see Figure 6.18B). Furthermore we found that the Models have a huge impact on the common kinetic parameters. For Model 2 and Model 3 the diffusion range D/γ and the secretion $\alpha/(\gamma K_D)$ are significantly increased, which suggests a stronger dynamic in the gradient formation. The sugar concentration in the tissue S_0, however, is significantly smaller in Model 2 and Model 3. The Background b value is nearly unchanged and also the spread of the estimated parameters over the different experiments is not influenced by the models. Finally we considered the additional parameters of Model 2 and Model 3 and we observed that for experiment 1-4, which have multiple LVs with strongly varying size, the individual sugar concentrations strongly spread. In contrast, experiments 12-16,

which showed LVs of nearly the same size also have similar values for the individual sugar concentrations.

In summary we found that the model for tissue specific sugar concentrations (Model 2) and the model for lymphoid vessel specific sugar concentration (Model 3) both describe the process of CCL21 gradient formation at the LV better than the basic model (Model 1). The difference in LV intensity strongly biased the estimated kinetic parameters obtained by Model 1 and this problem is addressed by Model 2 as well as by Model 3. Furthermore, we found that the best model choice is Model 3. This hints strongly at a functional marking of different LV regions with the chemokine signal. This prediction is supported by the findings of Kilarski *et al.* (2013).

6.5. Conclusion

Tissue scale images are a data source for studying spatial aspects of biological processes. Unfortunately, analysis by eye often only yields small quantities of the insight that can possibly be gained by a rigorous model based analysis. For example the CCL21 gradient is only visible to the expert eye in the original images and while a selective marking of LV regions with high CCL21 intensities can also be seen, the importance of those phenomena can not be assessed. In a step by step process we built up a model based analysis of the CCL21 images concluding in the answer that there is indeed a significant difference in the sugar concentrations, and therefore in the CCL21 signal, in the different LV vessels.

The model based analysis of the CCL21 images considered three key aspects: data representation, parameter estimation methodology and hypothesis based model selection. We found that the distance measure considered in Weber *et al.* (2013) is a too crude simplification of the CCL21 data, especially without filtering of the images. For the reliable estimation of the kinetic parameters the full two-dimensional images have to be considered. The parameter estimation process with the raw untreated images, however, is not accurate enough and either a filtering or the newly developed integrated noise model approach have to be applied. Both methods yield reliable estimation results. The assessment of the three proposed models representing the biological hypothesis of no difference in sugar concentration between LV and surrounding tissue, different sugar concentration in the LV region and different sugar concentration in each LV with the BIC was the third aspect. This yielded strong evidence for the hypothesis that there exists a different sugar concentration in each LV.

To address Problem 6.2, a parameter estimation process was designed to estimate the parameters from the two-dimensional images. Furthermore it was tested for its reliability, efficiency and applicability for the CCL21 images. We found that the integrated noise model introduced in Chapter 3 is an adequate choice for image based parameter estimation. For the CCL21 application example it outperformed the filtering approach without any necessary adaption to the images.
From a biological point of view we addressed the question of gradient formation raised in Problem 6.1 by model development, parameter estimation and model selection for the CCL21 data. We found that the CCL21 gradient formation can

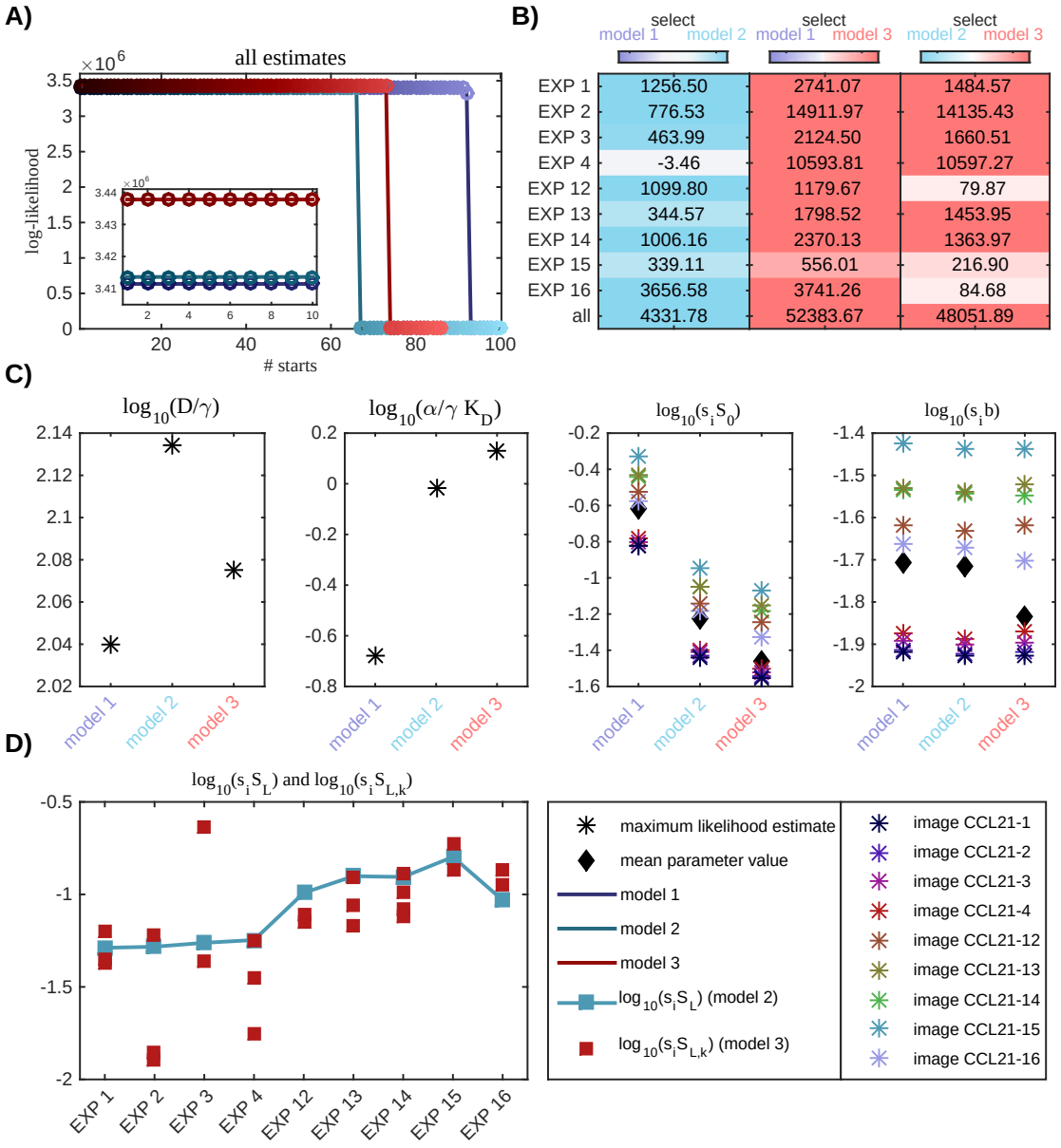


Figure 6.18.: ***Model selection on all nine CCL21 images*** *A)Comparison of the maximal log-likelihood results from 15 optimization starts. We observe that the estimation on Model 1, 2 and 3 converged to a global optimum in all runs and the value obtained by Model 3 is maximal. B) BIC of the estimates for each individual image and for the simultaneous consideration of all nine images. The color coding indicates which model is favored. C) Comparison of the shared optimal parameters of the three models. D) Comparison of the additional parameters in Model 2 and Model 3.*

indeed be described by source-diffusion-complex formation process, however, there is a strong evidence that the complex formation has a distinctive local patterning due to different concentrations of the sugar (heparan sulfates).

In a subsequent step, the simultaneous inference using multiple imaging data with the integrated noise model could be addressed. The computational bottleneck is the repeated choice of kinetic parameters, which yield numerical, instable simulations during the update steps for the noise parameters. We propose the design of an optimization procedure, which treats kinetic parameters and noise parameters separately. Thus the repeated, time consuming simulation of the CCL21 model can be done with a high accuracy, while the fast but slow converging optimization of the noise parameters can be performed more often.

In addition to the possible numerical improvements for the integrated noise model approach there are a number of possible biological studies, which could be proposed based on the findings of this chapter. First, of course a experimental verification of the importance of the lymphoid vessel sugar concentration for the CCL21 gradient formation is of great interest. From a theoretical point of view the reduction or the complete deletion of haparan sulfate production in the LV should strongly affect the gradient formation. A systematic reduction of the sugar production could yield a dilution series based on which the estimation of the absolute sugar concentration might become possible. Furthermore, the behavior of dendritic cells applied to a mouse slice with such a reduced sugar concentration in the LV vessels could prove the functionality of the different concentrations in different LV regions. While a study suggesting this functionality has already been done by Kilarski *et al.* (2013), the data obtained proved insufficient to perform a model based analysis to quantitatively analyze the validity of this hypothesis.
With the development of new imaging technologies biological measurements move away from single data point observations towards more spatially organized data. This additional spatial dimension yields significantly more insight into the biological processes but comes with a high demand for accurate, adaptive and efficient analytical tools to extract and test biological hypothesis. We demonstrated with the CCL21 gradient formation that the methods proposed in this thesis can contribute to the understanding of complicated spatio-temporal processes.

7. Conclusion and Outlook

The key elements to exploit imaging techniques with biological models to gain insights beyond visual analysis are parameter estimation and model based hypothesis testing, unfortunately both fields are still in the fledgling stages when it comes to spatio-temporal models as encountered in image based systems biology. Despite the developments for inverse problems with semi-linear PDEs, key elements like partial observations, parameter identifiability, non-standard noise models or uncertainty quantification are still open questions. In this thesis we novel tools to address image based parameter estimation and uncertainty quantification for semi-linear PDEs as they are encountered in biological applications, culminating in the development of an efficient profile likelihood calculation for semi-linear PDE models. This enabled the rigorous model based analysis of different hypothesis of Pom1p and CCL21 gradient formation yielding new biological insights and motivating new experiments to further elucidate those biological processes.

Summarizing our results we began by introducing imaging specific structures like partial or pixel based observation operators. The newly introduced non-standard parameter estimation approach, which considers an integrated noise model instead of the step wise process of feature extraction and parameter estimation considered in the field was of special interest. Based on this consideration the parameter estimation problem in a very generic set-up was established as an optimization problem with semi-linear PDE constraints for which optimality conditions and the Hessian matrix were deduced. The Hessian matrix is only a second order approximation to parameter uncertainty and often fails in the presence of parameter indeterminacies. Therefore, the profile likelihood was introduced. However, standard approaches to calculate the profile likelihood have been shown to be infeasible for computationally demanding problems like semi-linear PDEs (Hock *et al.*, 2013). To tackle the problem of profile likelihood calculation for semi-linear PDE models, an efficient ODE based formulation of the profile likelihood calculation was developed following the method for ODEs by Chen & Jennrich (2002). The combination of the the ODE formulation with the previously deduced Hessian matrix or efficient approximations based on the Fisher Information matrix, in case of missing or computationally infeasible Hessian information, gave an enhanced profile likelihood calculation. For the application example on which we evaluated the method a decrease of up to 90% in the number of function evaluations yielded a tremendous reduction in computation time while the accuracy of the profiles was still close to the standard optimization based calculation. This impressive speed up enabled the profile likelihood based uncertainty analysis to be used for the real life applications of the Pom1p and the CCL21 gradient formation. For both applications we also observed a strong reduction in the number of function evaluations up to 70%. Besides the efficiency of the newly introduced profile calculation method we also used the CCL21 application

example as proof of concept for the introduced integrated noise model approach. We compared the standard filtering approach and the integrated noise model approach on simulated and real data and found a good performance of the new method.

Regarding the discussed parameter estimation and uncertainty quantification methods for image based systems biology we can give two key conclusions. First, based on the numerical results for the application example, the Pom1p gradient formation models and the CCL21 model based analysis, we demonstrated that the newly developed efficient profile calculation method outperforms the standard optimization based approach by several orders of magnitude regarding the number of function evaluations. This increase in calculation speed does not reduce the accuracy as the obtained profiles and the decisions regarding parameter identifiability are comparable to that of the standard approach thus enabling the profile calculation for computationally demanding models. From an application point of view we can conclude that as well as for the application example and the real life applications the full Hessian is not needed to obtain reliable results but can be approximated with the often easier to obtain Fisher Information matrix. Furthermore, the use of the Fisher information matrix stabilized the method against approximation errors or near zero eigenvalues in the Hessian matrix, especially for non-identifiable parameters. Second, the application of the integrated noise model approach to the simulated and the real data of the tissue scale application example outperformed the standard filtering approach. Hence we can conclude that without loss of estimation precision the new method enables a comprehensive analysis of uncertainties for which no data is discarded and no specialists knowledge is needed to tune it (compared to filtering methods).

Recapitulating the research questions posed in Section 1.1 the results obtained in this thesis answer the posed questions. We introduced the general parameter estimation problem and based on the efficient profile calculation for ODEs we introduced a method for semi-linear PDEs. Furthermore, we proposed Hessian matrix approximations and evaluated them on an application motivated example, which showed an impressive improvement of calculation speed compared to the standard approach. To reinforce the theoretical findings we used the developed methods to perform a rigorous model based analysis for single cell image data and tissue scale image data to elucidate the underlying biological processes.

In the application chapters we proposed new biological experiments and further studies to verify the predictions made by the models. Beyond that the key point for improvements and further studies is the optimal choice for the adaption parameter γ in (4.6). In this work we chose γ by hand, i.e. the visual comparison of benchmark and approximated profile and to our knowledge there is no work yet proposing how to chose that parameter efficiently. Especially, the accuracy of the profiles strongly depends on this parameter and a more structural choice should be possible. For large values the ODE becomes increasingly stiff resulting in increased computation time. Furthermore, too narrow profiles occur if γ is very large but small values result in inaccurate profiles so the optimal choice should be a trade off between speed and accuracy. We propose the development of an adaptive update scheme

for γ depending on the accuracy of the Hessian matrix approximation. Another point, which should be addressed in further studies are the convergence problems of the developed integrated noise model. Here, the estimation of the noise parameters and the kinetic parameters should be separated as the noise model has a rather big overhead of parameters. These scale up the optimization process, however, they do not need repeated simulation of the model in each step.

Beyond the calculation of profile likelihoods, the proposed scheme can be employed to calculate profile posteriors. This enables a Bayesian uncertainty analysis which might be more efficient than parameter sampling (Hug & Raue, 2013). In the last years, however, there has been a rapid development of Bayesian techniques intended for PDE models (Bui-Thanh *et al.*, 2011; Bui-Thanh & Girolami, 2014). A comparison of the obtained uncertainty measures, i.e. marginals, with the profile likelihoods described has already been done for ODEs (Raue *et al.*, 2013) and would be interesting for PDEs as well. Another recent development is the use of model reduction methods to perform Bayesian uncertainty analysis for large scale models (Sternfels & Earls, 2013). Model reduction could also be applied to the optimization process for the maximum likelihood calculation and an especially nice aspect of model reduction is the possibility to give exact error bounds, which are missing for the here introduced method.

Parameter estimation and uncertainty quantification for image based systems biology is still in its fledgling state especially as it strongly relies on efficient methods for semi-linear, computational demanding PDE problems, which are currently not available. This hampers the development of quantitative spatio-temporal models from image data compared to the very often considered ordinary differential equation models. The spatial scale, however, offers a wealth of information far beyond what time series data can supply as spatial structures, gradients and tissue specifications have great influence on the biological processes. In this work we took the first steps towards introducing efficient methods, which will allow the uncertainty analysis of semi-linear PDE problems with similar to the state of the art for ODE problems. Furthermore, we introduced a comprehensive integrated noise model approach, which is especially useful for analysis of tissue scale image data. The combination of both should enable the application of a rigorous uncertainty quantification to a large number of existing biological spatio-temporal models and facilitate the analysis of existing data to gain new insights.

Appendix

A. Additional Data and Estimation results

Table 1.: ***CCL21 data: estimated parameters for the 1D source-diffusion-complex formation model*** *Parameter estimation performed simultaneously on the nine filtered CCL21 staining images. The estimated parameters are obtained by the best fit out of 100 initial points sampled with a pace filling design.*

	$\log_{10}(\theta_i)$	$\log_{10}(\hat{\theta}_i)$	$\mathrm{CI}_{0.05}(\theta_i)$
D/γ	-1.39	2.50	$[0.82; 5.10]$
$s_M(S_0\alpha\gamma K_D)$	11.21	10.99	$[-1.12; 2.70]$
$s_M b$	10.95	9.68	$[-8.83; -1.60]$
s_1/s_M	-12.81	-11.60	$[-0.37; -0.09]$
s_2/s_M	-12.81	-11.59	$[-0.38; -0.06]$
s_3/s_M	-12.81	-11.57	$[-0.35; -0.05]$
s_4/s_M	-12.81	-11.56	$[-0.29; -0.07]$
s_{12}/s_M	-12.81	-11.33	$[-0.15; 0.21]$
s_{13}/s_M	-12.81	-11.23	$[-0.06; 0.32]$
s_{14}/s_M	-12.81	-11.23	$[-0.09; 0.33]$
s_{15}/s_M	-12.81	-11.12	$[0.04; 0.43]$
s_{16}/s_M	-12.81	-11.29	$[-0.28; 0.31]$

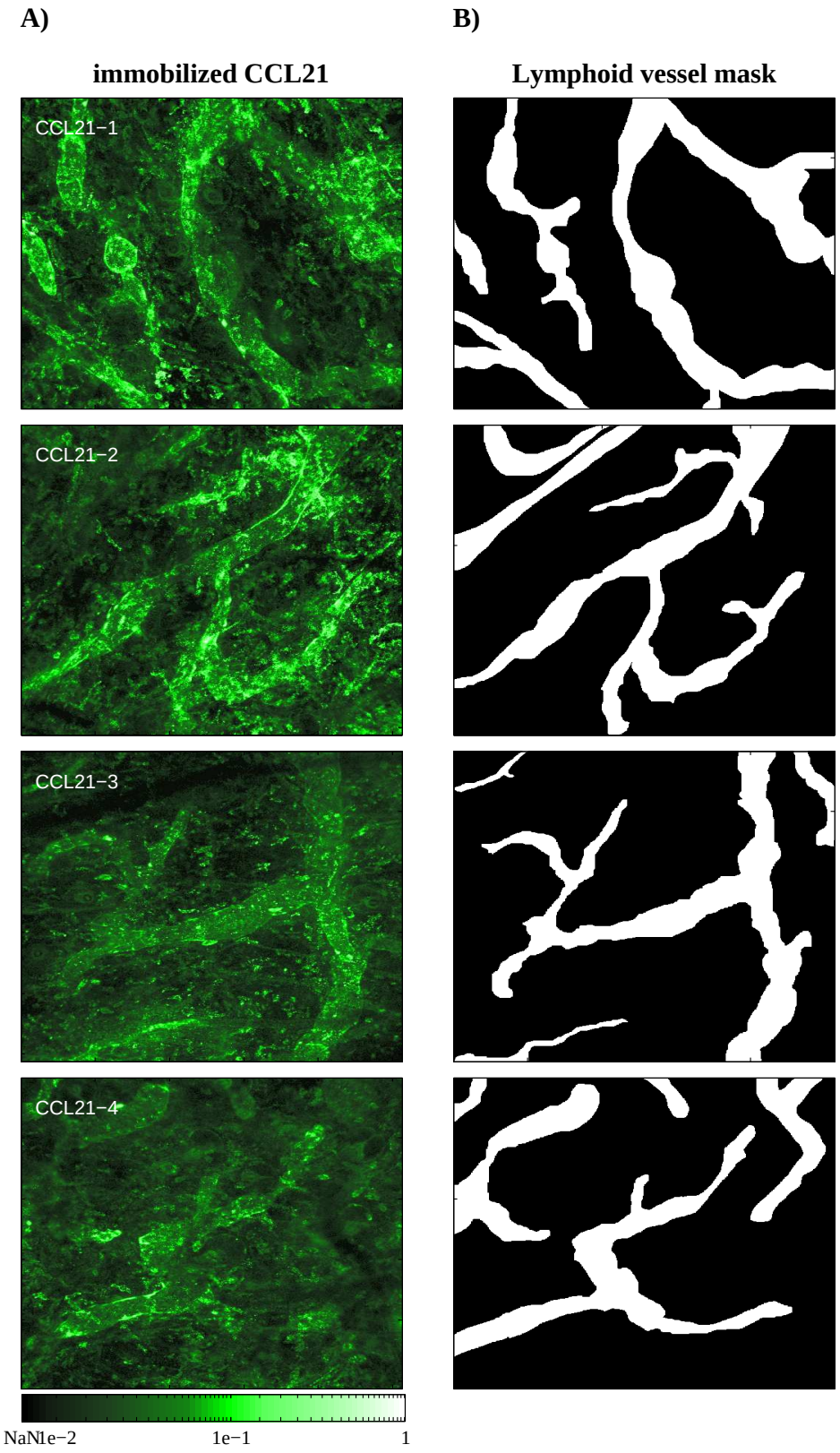


Figure 1.: *CCL21 immonostainings and LV masks experiments 1-4*

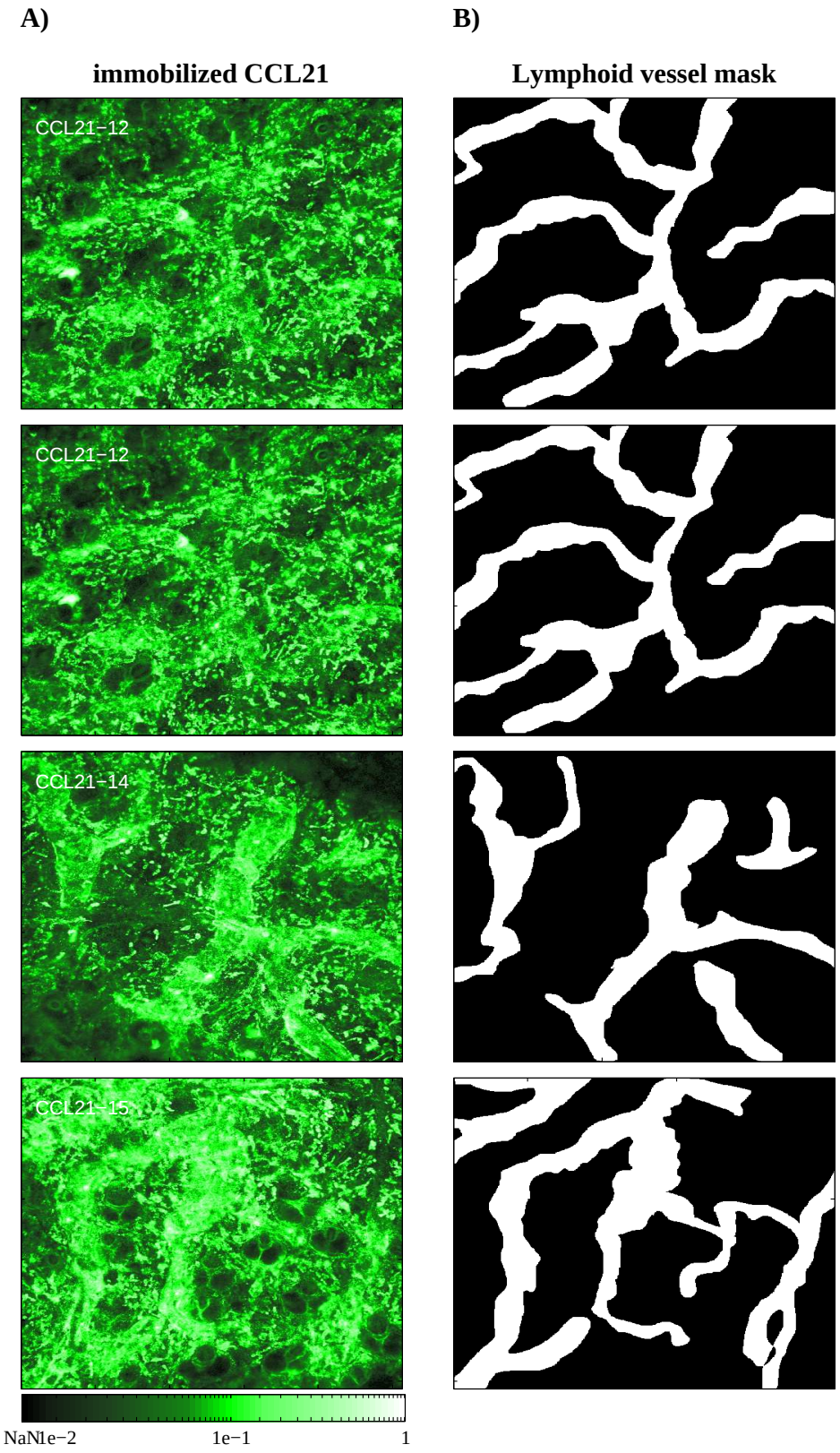


Figure 2.: *CCL21 immonostainings and LV masks experiments 12-14*

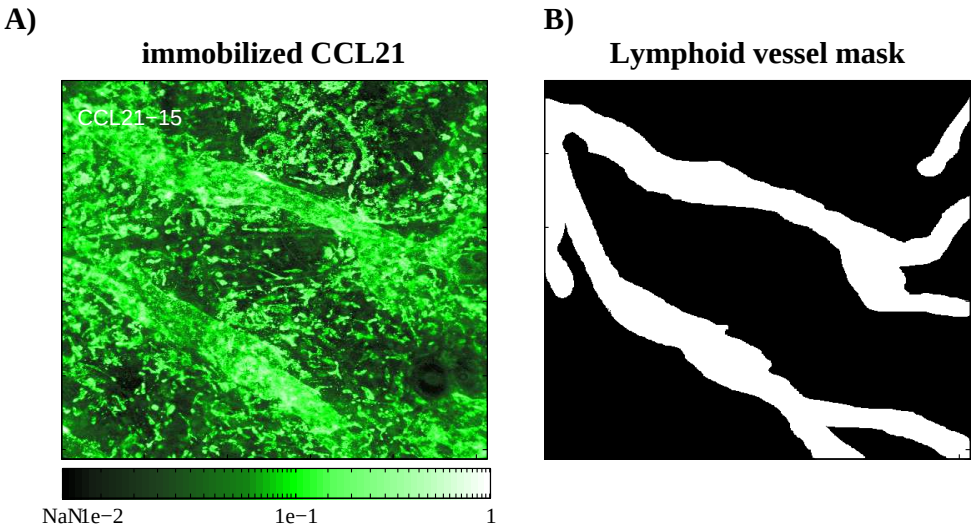


Figure 3.: *CCL21 immonostainings and LV masks experiments 16*

Table 2.: ***CCL21 data: estimated parameters of 2D model based on distance measure*** *Parameter estimation performed simultaneously on the nine filtered CCL21 staining images. The estimated parameters are obtained by the best fit out of 25 initial points sampled with a space filling design. As we saw that only the scaled parameters are identifiable for this estimation we displayed only the scaled value for each image. θ_i is the initial guess and $\hat{\theta}_i$ is the MLE.*

	$\log_{10}(\theta_i)$	$\log_{10}(\hat{\theta}_i)$	$\mathrm{CI}_{0.05}(\theta_i)$
D/γ	-1.39	2.46	$[1.10; 4.29]$
$s_M(S_0\alpha\gamma K_D)$	-1.08	-0.80	$[-2.78; -1.13]$
$s_M b$	-1.34	-1.24	$[-11.71; -1.71]$
s_1/s_M	-0.53	-0.69	$[-0.21; 0.18]$
s_2/s_M	-0.53	-0.67	$[-0.20; 0.09]$
s_3/s_M	-0.53	-0.65	$[-0.34; 0.15]$
s_4/s_M	-0.53	-0.63	$[-0.20; 0.11]$
s_{12}/s_M	-0.53	-0.42	$[0.02; 0.46]$
s_{13}/s_M	-0.53	-0.31	$[0.13; 0.54]$
s_{14}/s_M	-0.53	-0.31	$[0.16; 0.56]$
s_{15}/s_M	-0.53	-0.21	$[0.15; 0.70]$
s_{16}/s_M	-0.53	-0.37	$[0.00; 0.59]$

Table 3.: ***Simulated data: estimated parameters for the 2D source-diffusion-complex formation model based on raw images*** *Parameter estimation performed on the raw simulated data. Each table displays the estimated parameters for the best fit out of 25 for the number of introduced spots.*

no spots

	$\log_{10}(\theta_i)$	$\log_{10}(\hat{\theta}_i)$	$\mathrm{CI}_{0.05}(\theta_i)$
D/γ	1.93	1.92	$[1.91; 1.93]$
$\alpha/\gamma K_D$	-0.62	-0.65	$[-0.69; -0.62]$
S_0	-0.68	-0.65	$[-0.68; -0.62]$
b	-1.89	-1.88	$[-1.89; -1.88]$

 # spots: 20

	$\log_{10}(\theta_i)$	$\log_{10}(\hat{\theta}_i)$	$\mathrm{CI}_{0.05}(\theta_i)$
D/γ	1.93	1.90	$[1.89; 1.91]$
$\alpha/\gamma K_D$	-0.62	-0.64	$[-0.68; -0.60]$
S_0	-0.68	-0.67	$[-0.70; -0.63]$
b	-1.89	-1.88	$[-1.88; -1.88]$

 # spots: 40

	$\log_{10}(\theta_i)$	$\log_{10}(\hat{\theta}_i)$	$\mathrm{CI}_{0.05}(\theta_i)$
D/γ	1.93	1.96	$[1.95; 1.97]$
$\alpha/\gamma K_D$	-0.62	-0.70	$[-0.76; -0.66]$
S_0	-0.68	-0.60	$[-0.65; -0.56]$
b	-1.89	-1.88	$[-1.88; -1.88]$

 # spots: 80

	$\log_{10}(\theta_i)$	$\log_{10}(\hat{\theta}_i)$	$\mathrm{CI}_{0.05}(\theta_i)$
D/γ	1.93	1.95	$[1.93; 1.96]$
$\alpha/\gamma K_D$	-0.62	-0.68	$[-0.74; -0.62]$
S_0	-0.68	-0.63	$[-0.68; -0.58]$
b	-1.89	-1.86	$[-1.87; -1.86]$

 # spots: 160

	$\log_{10}(\theta_i)$	$\log_{10}(\hat{\theta}_i)$	$\mathrm{CI}_{0.05}(\theta_i)$
D/γ	1.93	1.99	$[1.98; 2.00]$
$\alpha/\gamma K_D$	-0.62	-0.84	$[-0.95; -0.75]$
S_0	-0.68	-0.49	$[-0.57; -0.39]$
b	-1.89	-1.85	$[-1.85; -1.85]$

 # spots: 320

	$\log_{10}(\theta_i)$	$\log_{10}(\hat{\theta}_i)$	$\mathrm{CI}_{0.05}(\theta_i)$
D/γ	1.93	1.91	$[1.89; 1.93]$
$\alpha/\gamma K_D$	-0.62	-0.41	$[-0.48; -0.34]$
S_0	-0.68	-0.88	$[-0.93; -0.82]$
b	-1.89	-1.80	$[-1.80; -1.80]$

 # spots: 640

	$\log_{10}(\theta_i)$	$\log_{10}(\hat{\theta}_i)$	$\mathrm{CI}_{0.05}(\theta_i)$
D/γ	1.93	1.95	$[1.93; 1.98]$
$\alpha/\gamma K_D$	-0.62	-0.32	$[-0.41; -0.24]$
S_0	-0.68	-0.97	$[-1.03; -0.91]$
b	-1.89	-1.75	$[-1.75; -1.74]$

Table 4.: ***Simulated data: estimated parameters for the 2D source-diffusion-complex formation model based on filtered images*** *Parameter estimation performed on the filter simulated data. Each table displays the estimated parameters for the best fit out of 25 for the number of introduced spots.*

no spots

	$\log_{10}(\theta_i)$	$\log_{10}(\hat{\theta}_i)$	$\mathrm{CI}_{0.05}(\theta_i)$
D/γ	1.93	1.92	$[1.91; 1.93]$
$\alpha/\gamma K_D$	-0.62	-0.65	$[-0.70; -0.62]$
S_0	-0.68	-0.65	$[-0.68; -0.61]$
b	-1.89	-1.88	$[-1.89; -1.88]$

 # spots: 20

	$\log_{10}(\theta_i)$	$\log_{10}(\hat{\theta}_i)$	$\mathrm{CI}_{0.05}(\theta_i)$
D/γ	1.93	1.92	$[1.91; 1.93]$
$\alpha/\gamma K_D$	-0.62	-0.64	$[-0.68; -0.60]$
S_0	-0.68	-0.66	$[-0.69; -0.63]$
b	-1.89	-1.88	$[-1.89; -1.88]$

 # spots: 40

	$\log_{10}(\theta_i)$	$\log_{10}(\hat{\theta}_i)$	$\mathrm{CI}_{0.05}(\theta_i)$
D/γ	1.93	1.92	$[1.91; 1.93]$
$\alpha/\gamma K_D$	-0.62	-0.64	$[-0.68; -0.60]$
S_0	-0.68	-0.66	$[-0.69; -0.63]$
b	-1.89	-1.89	$[-1.89; -1.88]$

 # spots: 80

	$\log_{10}(\theta_i)$	$\log_{10}(\hat{\theta}_i)$	$\mathrm{CI}_{0.05}(\theta_i)$
D/γ	1.93	1.92	$[1.91; 1.93]$
$\alpha/\gamma K_D$	-0.62	-0.61	$[-0.65; -0.57]$
S_0	-0.68	-0.69	$[-0.72; -0.66]$
b	-1.89	-1.88	$[-1.89; -1.88]$

 # spots: 160

	$\log_{10}(\theta_i)$	$\log_{10}(\hat{\theta}_i)$	$\mathrm{CI}_{0.05}(\theta_i)$
D/γ	1.93	1.93	$[1.93; 1.94]$
$\alpha/\gamma K_D$	-0.62	-0.58	$[-0.63; -0.55]$
S_0	-0.68	-0.71	$[-0.74; -0.68]$
b	-1.89	-1.89	$[-1.89; -1.88]$

 # spots: 320

	$\log_{10}(\theta_i)$	$\log_{10}(\hat{\theta}_i)$	$\mathrm{CI}_{0.05}(\theta_i)$
D/γ	1.93	1.92	$[1.91; 1.93]$
$\alpha/\gamma K_D$	-0.62	-0.48	$[-0.51; -0.44]$
S_0	-0.68	-0.79	$[-0.82; -0.76]$
b	-1.89	-1.88	$[-1.88; -1.88]$

 # spots: 640

	$\log_{10}(\theta_i)$	$\log_{10}(\hat{\theta}_i)$	$\mathrm{CI}_{0.05}(\theta_i)$
D/γ	1.93	1.91	$[1.90; 1.92]$
$\alpha/\gamma K_D$	-0.62	-0.47	$[-0.51; -0.43]$
S_0	-0.68	-0.81	$[-0.84; -0.77]$
b	-1.89	-1.86	$[-1.87; -1.86]$

Bibliography

Almeida, R. & Tyers, M. (2009). Cell size control: governed by a spatial gradient. *Developmental Cell*, **17**, 3–4.

Antony, P.M.A., Trefois, C., Stojanovic, A., Baumuratov, A.S. & Kozak, K. (2013). Light microscopy applications in systems biology: opportunities and challenges. *Cell communication and signaling : CCS*, **11**, 24.

Ashyraliyev, M., Jaeger, J. & Blom, J.G. (2008). Parameter estimation and determinability analysis applied to Drosophila gap gene circuits. *BMC systems biology*, **2**, 83.

Banks, H.T., Holm, K. & Robbins, D. (2010). Standard Error Computations for Uncertainty Quantification in Inverse Problems: Asymptotic Theory vs. Bootstrapping. *Mathematical and Computer Modelling*, **52**, 1610–1625.

Boiger, R., Hasenauer, J., Hross, S. & Kaltenbacher, B. (2016). Integration based profile likelihood calculation for PDE constrained parameter estimation problems. *Inverse Problems*.

Buggenthin, F., Marr, C., Schwarzfischer, M., Hoppe, P.S., Hilsenbeck, O., Schroeder, T. & Theis, F.J. (2013). An automatic method for robust and fast cell detection in bright field images from high-throughput microscopy. *BMC bioinformatics*, **14**, 297.

Bui-Thanh, T. & Girolami, M. (2014). Solving large-scale PDE-constrained Bayesian inverse problems with Riemann manifold Hamiltonian Monte Carlo. *Inverse Problems*, **30**, 114014.

Bui-Thanh, T., Ghattas, O. & Higdon, D. (2011). Adaptive Hessian-based nonstationary gaussian process response surface method for probability density approximation with application to bayesian solution of large-scale inverse problems. *SIAM Journal on Scientific Computing*, **34**, 2837–2871.

Burnham, K. & Anderson, D. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach (2nd ed)*, vol. 172. Springer-Verlag New York, 2nd edn.

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W. & Prasher, D.C. (1994). Green fluorescent protein as a marker for gene expression. *Science (New York, N.Y.)*, **263**, 802–805.

Chen, J.s. & Jennrich, R.I. (1996). The Signed Root Deviance Profile and Confidence Intervals in Maximum Likelihood Analysis. *Journal of the American Statistical Association*, **91**, 993.

Chen, J.S. & Jennrich, R.I. (2002). Simple Accurate Approximation of Likelihood Profiles. *Journal of Computational and Graphical Statistics*, **11**, 714–732.

Chis, O.T., Banga, J.R. & Balsa-Canto, E. (2011). Structural identifiability of systems biology models: a critical comparison of methods. *PloS One*, **6**, e27755.

Coons, A.H., Creech, H.J. & Jones, N.R. (1941). Immunological properties of an antibody containing a fluorescent group. *Proceedings of the Society ...*, **47**, 200–202.

Engl, H.W., Flamm, C., Kügler, P., Lu, J., Müller, S. & Schuster, P. (2009). Inverse problems in systems biology. *Inverse Problems*, **25**, 123014.

Erguler, K. & Stumpf, M.P.H. (2011). Practical limits for reverse engineering of dynamical systems: a statistical analysis of sensitivity and parameter inferability in systems biology models. *Molecular bioSystems*, **7**, 1593–602.

Evans, L. (1996). Partial Differential Equations.

Fernández Slezak, D., Suárez, C., Cecchi, G.a., Marshall, G. & Stolovitzky, G. (2010). When the optimal is not the best: parameter estimation in complex biological models. *PloS One*, **5**, e13283.

Garvie, M. & Trenchea, C. (2014). Identification of space-time distributed parameters in the Gierer-meinhardt reaction-diffusion system. *SIAM Journal on Applied Mathematics*, **74**, 147–166.

Gierer, a. & Meinhardt, H. (1972). A theory of biological pattern formation. *Kybernetik*, **12**, 30–9.

Gutenkunst, R.N., Waterfall, J.J., Casey, F.P., Brown, K.S., Myers, C.R. & Sethna, J.P. (2007). Universally sloppy parameter sensitivities in systems biology models. *PLoS Computational Biology*, **3**, 1871–78.

Hachet, O., Berthelot-Grosjean, M., Kokkoris, K., Vincenzetti, V., Moosbrugger, J. & Martin, S.G. (2011). A phosphorylation cycle shapes gradients of the DYRK family kinase Pom1 at the plasma membrane. *Cell*, **145**, 1116–28.

Hachet, O., Bendezú, F.O. & Martin, S.G. (2012). Fission yeast: in shape to divide. *Current Opinion in Cell Biology*, **24**, 858–64.

Hersch, M., Hachet, O., Dalessi, S., Ullal, P., Bhatia, P., Bergmann, S. & Martin, S.G. (2015). Pom 1 gradient buffering through intermolecular auto-phosphorylation. *Molecular systems biology*, **7**.

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

Hinze, M., Pinnau, R., Ulbrich, M. & Ulbrich, S. (2009). *Optimization with PDE Constraints*, vol. 23 of *Mathematical Modelling: Theory and Applications*. Springer Netherlands, Dordrecht.

Hock, S., Hasenauer, J. & Theis, F.J. (2013). Modeling of 2D diffusion processes based on microscopy data: parameter estimation and practical identifiability analysis. *BMC bioinformatics*, **14 Suppl 1**, S7.

Horn, F. & Jackson, R. (1972). General mass action kinetics. *Archive for Rational Mechanics and Analysis*, **47**.

Hross, S. & Hasenauer, J. (2016). Analysis of CFSE time-series data using division-, age- and label-structure population models. *Bioinformatics*, **32**, 2321–9.

Hross, S., Fiedler, A., Theis, j.F. & Hasenauer, J. (2016). Quantitative comparison of competing PDE models for Pom1p dynamics in fission yeast. *FOSBE*.

Huang, B., Babcock, H. & Zhuang, X. (2010). Breaking the diffraction barrier: Super-resolution imaging of cells. *Cell*, **143**, 1047–1058.

Hug, S. & Raue, A. (2013). High-dimensional Bayesian parameter estimation : Case study for a model of JAk2/STAT5 signaling. *Mathematical Biosciences*, **246**.

Jacquez, J.a. & Perry, T. (1990). Parameter estimation: local identifiability of parameters. *The American journal of physiology*, **258**, E727–36.

Kenz, Z., Banks, H. & Smith, R. (2013). Comparison of Frequentist and Bayesian Confidence Analysis Methods on a visoelastic stenosis model. *SIAM Journal on Uncertainty Quantification*, **1**, 348–369.

Kilarski, W.W., Güç, E., Teo, J.C.M., Oliver, S.R., Lund, A.W. & Swartz, M.a. (2013). Intravital immunofluorescence for visualizing the microcirculatory and immune microenvironments in the mouse ear dermis. *PloS one*, **8**, e57135.

Kitano, H. (2002). Computational systems biology. *Nature*, **420**, 206–210.

Klann, M. & Koeppl, H. (2012). Spatial simulations in systems biology: from molecules to cells. *International journal of molecular sciences*, **13**, 7798–827.

Kremers, G.J., Gilbert, S.G., Cranfill, P.J., Davidson, M.W. & Piston, D.W. (2011). Fluorescent proteins at a glance. *Journal of cell science*, **124**, 157–160.

Kreutz, C., Raue, A., Kaschek, D. & Timmer, J. (2013). Profile likelihood in systems biology. *The FEBS journal*, **280**, 2564–71.

Lillacci, G. & Khammash, M. (2010). Parameter estimation and model selection in computational biology. *PLoS Computational Biology*, **6**.

Lockley, R., Ladds, G. & Bretschneider, T. (2015). Image based validation of dynamical models for cell reorientation. *Cytometry Part A*, **87**, 471–480.

Martin, S.G. & Berthelot-Grosjean, M. (2009). Polar gradients of the DYRK-family kinase Pom1 couple cell length with the cell cycle. *Nature*, **459**, 852–856.

Matas, J., Chum, O., Urban, M. & Pajdla, T. (2004). Robust wide-baseline stereo from maximally stable extremal regions. In *Image and Vision Computing*, vol. 22, 761–767.

Medyukhina, A., Timme, S., Mokhtari, Z. & Figge, M.T. (2015). Image-based systems biology of infection. *Cytometry Part A*, **87**, 462–470.

Meeker, W.Q. & Escobar, L.A. (1995). On Maximum Likelihood Estimation. *The American Statistician*, **49**, 48–53.

Megason, S.G. & Fraser, S.E. (2007). Imaging in Systems Biology. *Cell*, **130**, 784–795.

Menshykau, D. & Iber, D. (2013). Kidney branching morphogenesis under the control of a ligand-receptor-based Turing mechanism. *Physical biology*, **10**, 046003.

Moseley, J.B., Mayeux, A., Paoletti, A. & Nurse, P. (2009). A spatial gradient coordinates cell size and mitotic entry in fission yeast. *Nature*, **459**, 857–860.

Murphy, S.a.S., Vaart, A.V.D.A.V.D. & van der Vaart, A. (2000). On profile likelihood. *Journal of the American Statistical Association*, **95**, 449–465.

Nistér, D. & Stewénius, H. (2008). Linear time maximally stable extremal regions. *08 Proceedings of the 10th ECCV (Marseille, France)*, **5303**, 183–196.

Raue, a., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U. & Timmer, J. (2009). Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics (Oxford, England)*, **25**, 1923–9.

Raue, A., Kreutz, C., Theis, F.J. & Timmer, J. (2013). Joining forces of Bayesian and frequentist methodology: a study for inference in the presence of non-identifiability. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, **371**, 20110544.

Reits, E.a. & Neefjes, J.J. (2001). From fixed to FRAP: measuring protein mobility and activity in living cells. *Nature cell biology*, **3**, E145–E147.

Saunders, T.E. (2015). Aggregation-fragmentation model of robust concentration gradient formation. *Physical Review E*, **91**, 022704.

Saunders, T.E., Pan, K.Z., Angel, A., Guan, Y., Shah, J.V., Howard, M. & Chang, F. (2012). Noise reduction in the intracellular pom1p gradient by a dynamic clustering mechanism. *Developmental cell*, **22**, 558–72.

Sbalzarini, I.F. (2013). Modeling and simulation of biological systems from image data. *BioEssays : news and reviews in molecular, cellular and developmental biology*, **35**, 482–90.

Schumann, K., Lämmermann, T., Bruckner, M., Legler, D.F., Polleux, J., Spatz, J.P., Schuler, G., Förster, R., Lutz, M.B., Sorokin, L. & Sixt, M. (2010). Immobilized chemokine fields and soluble chemokine gradients cooperatively shape migration patterns of dendritic cells. *Immunity*, **32**, 703–13.

Sprague, B.L. & McNally, J.G. (2005). FRAP analysis of binding: proper and fitting. *Trends in cell biology*, **15**, 84–91.

Stark, P. (2012). Constraints versus Priors. *Journal on Uncertainty Quantification*, **3**, 586–598.

Sternfels, R. & Earls, C.J. (2013). Reduced-order model tracking and interpolation to solve PDE-based Bayesian inverse problems. *Inverse Problems*, **29**, 075014.

Sung, M.H. & McNally, J.G. (2011). Live cell imaging and systems biology. *Wiley interdisciplinary reviews. Systems biology and medicine*, **3**, 167–182.

Tröltzsch, F. (2009). *Optimale Steuerung partieller Differentialgleichungen*. Vieweg+Teubner, Wiesbaden.

Turing, A.M. (1952). The chemical basis of morphogenisis. *Philosophical Transactions of the Royal Societs of London B*, **237**, 37–72.

Uzkudun, M., Marcon, L. & Sharpe, J. (2015). Data-driven modelling of a gene regulatory network for cell fate decisions in the growing limb bud. *Molecular systems biology*, **7**, 1–15.

Venzon, D. & Moolgavkar, S. (1988). A Method for Computing Profile-Likelihood-Based Confidence Intervals. *Applied Statistics*, **37**, 87–94.

Verveer, P.J. & Bastiaens, P.I.H. (2008). Quantitative microscopy and systems biology: Seeing the whole picture. *Histochemistry and Cell Biology*, **130**, 833–843.

Villaverde, A.F. & Banga, J.R. (2014). Reverse engineering and identification in systems biology: strategies, perspectives and challenges. *Journal of the Royal Society Interface*, **11**, 20130505.

Wartlick, O., Kicheva, A. & González-Gaitán, M. (2009). Morphogen gradient formation. *Cold Spring Harbor perspectives in biology*, **1**, a001255.

Weber, M., Hauschild, R., Schwarz, J., Moussion, C., de Vries, I., Legler, D.F., Luther, S.A., Bollenbach, T. & Sixt, M. (2013). Interstitial Dendritic Cell Guidance by Haptotactic Chemokine Gradients. *Science*, **339**, 328–332.

Wiedenmann, J., Oswald, F. & Nienhaus, G.U. (2009). Fluorescent proteins for live cell imaging: Opportunities, limitations, and challenges. *IUBMB Life*, **61**, 1029–1042.

Yuste, R. (2005). Fluorescence microscopy today. *Nature Methods*, **2**, 902–904.