TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

Harri Pölönen

Quantification of Biomedical Data with Stochastic Parametric Models and Numerical Optimization

Julkaisu 915 • Publication 915

Tampere 2010

Tampereen teknillinen yliopisto. Julkaisu 915
Tampere University of Technology. Publication 915

Harri Pölönen

Quantification of Biomedical Data with Stochastic Parametric Models and Numerical Optimization

Thesis for the degree of Doctor of Science in Technology to be presented with due permission for public examination and criticism in Tietotalo Building, Auditorium TB219, at Tampere University of Technology, on the 8th of October 2010, at 12 noon.

Tampereen teknillinen yliopisto - Tampere University of Technology
Tampere 2010

ISBN 978-952-15-2442-4 (printed)
ISBN 978-952-15-2504-9 (PDF)
ISSN 1459-2045

Abstract

Accurate and robust quantification of measurement data is a key factor in biomedical research. However, the quantification is complicated by random noise, limited resolution and indirect nature of measurements.

In this study we quantified biomedical data by modeling the target, the acquisition process and the noise contamination. By combining these three components we built the model for the acquired data. The indirect acquisition was modeled as a forward projection from the target to the data. The random noise was handled by using stochastic model and treating the data as a single realization from the model. The data model was determined through adjustable parameters and the most likely parameters in terms of the acquired noisy data were searched for.

The search for best parameters in our stochastic parametric models led to mathematically inconvenient and challenging optimization problems. The most common challenges in our applications were multiple local optima, very large number of parameters, unknown gradient function and lack of reliable initialization. In order to solve these issues, we developed customized numerical optimization techniques by modifying standard algorithms and combining different types of optimization. The optimization techniques were implemented to distributed computing environment which enabled us to solve problems with very large number of parameters and amount of data.

By modeling static fluorescence microscopy images we achieved results which we could not obtain with the conventional methods and which in part helped to reveal significant differences between treatment groups. By modeling dynamic fluorescence microscopy data we managed to compensate cell movement and inhomogeneous fluorescence distribution well during the quantification. The results with simulated data imply that our method is very robust and accurate. With positron emission tomography (PET) data we were able to solve the huge parameter optimization problem which allowed us to quantify regional parameter heterogeneity with a novel approach. Overall, we believe that the stochastic parametric modeling is a very accurate and robust method to quantify biomedical data.

Preface

The work presented in this thesis has been carried out in the Department of Signal Processing at Tampere University of Technology during the years 2006-2010.

First of all, I would like to thank my supervisor Professor Ulla Ruotsalainen for her trust in me and never-ending optimism. In times when I saw only insuperable difficulties, she saw various promising possibilities. Her attitude has inspired me greatly and made this work possible.

I would like to thank the reviewers of the thesis, Ph.D. Michal Kozubek and Ph.D. Ferrante Neri, for making a great job in examining the thesis thoroughly and giving valuable feedback. I would like thank also Ph.D. Jussi Tohka for helping me greatly especially in the beginning of my work and career. All the other colleagues in the M^2oBSI research group deserve warm thanks for creating an inspiring and open-minded work environment. I would also like to express my gratitude to Professor Elina Ikonen and M.Sc Maurice Jansen from University of Helsinki for the fruitful collaboration and making biology understandable for me. I would also like to thank Rainer Wehkamp and Teppo Tammisto from Techila Technologies Ltd for guiding me in the fascinating world of distributed computing.

This thesis was financially supported by Academy of Finland (under the grant 213462, Finnish Centre of Excellence Program (2006 - 2011)), Doctoral Programme in Information Science and Engineering (TISE) and Jenny and Antti Wihuri Foundation. I am thankful for their support.

Finally, I would like to express my gratitude to my parents for letting me become me. Above all, this work would have never been completed without the constant support and care from the love of my life, Viivi.

Tampere, September 2010 *Harri Pölönen*

Supervisor:	Professor Ulla Ruotsalainen Department of Signal Processing Tampere University of Technology
Pre-examiners:	Michal Kozubek, Ph.D. Centre for Biomedical Image Analysis Masaryk University
	Ferrante Neri, Ph.D. Department of Mathematical Information Technology University of Jyväskylä
Opponent:	Professor Pekka Hänninen Laboratory of Biophysics, Institute of Biomedicine University of Turku

Contents

Abstract **i**

Preface **iii**

List of Publications **vii**

Abbreviations and Symbols **ix**

1 Introduction **1**

2 Biomedical imaging **3**
2.1 Data distortion 3
2.2 Noise contamination 5

3 Parametric modeling **9**
3.1 Definition and motivation 9
3.2 Model design 10
3.3 Model evaluation 14

4 Numerical Parameter Estimation **17**
4.1 Parameter space 17
4.2 Distributed computing environment 21
4.3 Developed methods 22

5 Selected Applications **29**
5.1 Fluorescence microscopy 29
5.1.1 Caveolin-1 protein distribution 31
5.1.2 Fluorescence recovery after photobleaching 34
5.2 Positron Emission Tomography 41

6 Results **43**
6.1 Caveolin-1 protein distribution 43
6.2 Fluorescence recovery after photobleaching 44
6.3 Positron Emission Tomography 48

7 Discussion **53**

Bibliography **57**

List of Publications

This Thesis is mainly based on the following publications. In the text, these publications are referred to as Publication [P*], where * is an integer.

P1 Harri Pölönen, Maurice Jansen, Elina Ikonen and Ulla Ruotsalainen: *Automatic FRAP Analysis with Inhomogeneous Fluorescence Distribution and Movement Compensation.* Accepted in Advances in Computational Biology (in series: Advances in Experimental Medicine and Biology, Vol. 680), Springer, 2010

P2 Harri Pölönen, Jussi Tohka and Ulla Ruotsalainen: *Automatic intensity quantification of fluorescence targets from microscope images with maximum likelihood estimation.* Proceedings of European Signal Processing Conference (EUSIPCO 2009), p. 1072-1076, Glasgow, Scotland, 2009

P3 Harri Pölönen, Jussi Tohka and Ulla Ruotsalainen: *Automatic Quantification of Fluorescence from Clustered Targets in Microscope Images.* Proceedings of 16th Scandinavian Conference on Image Analysis (SCIA 2009), Lecture Notes in Computer Science, Vol. 5575, p. 667-675, Oslo, Norway, 2009.

P4 Harri Pölönen, Jari Niemi and Ulla Ruotsalainen: *Error Corrected Estimation of Regional Kinetic Parameter Histograms Directly from PET Projections.* Submitted to Physics in Medicine and Biology.

P5 Maurice Jansen, Vilja M. Pietiäinen, Harri Pölönen, Laura Rasilainen, Mirkka Koivusalo, Ulla Ruotsalainen, Eija Jokitalo, and Elina Ikonen: *Cholesterol substitution increases the structural heterogeneity of caveolae.* Journal of Biological Chemistry, Vol. 283, Issue 21, p. 14610-14618, 2008.

The author's contribution

In Publications P1, P2 and P3 the author designed the parametric models according to provided biological and technical information, derived the functions and equations, performed simulations, quantified the data and mainly wrote the manuscripts. In P4 the author was involved in designing and performing the simulations and was responsible for writing the manuscript. In P4 Jari Niemi had equal contribution with the author of this thesis. In P5 the author was responsible for processing and quantifying the microscope image data.

Abbreviations and Symbols

BB	Barzilai-Borwein (optimization method)
DE	Differential evolution
GA	Genetic algorithm
FRAP	Fluorescence recovery after photobleaching
GFP	Green fluorescent protein
ML	Maximum likelihood
NFL	No free lunch (theorem)
p	Probability function
PET	Positron emission tomography
Poiss	Poisson distribution
PSF	Point spread function
SNR	Signal-to-noise ratio
t	Temporal coordinate
θ	Parameter vector
x,y,z	Spatial coordinates

Chapter 1

Introduction

Scientific research in the field of biology and medicine strives constantly to obtain detailed information from increasingly smaller targets and understanding of more complex behavior. Due to the chaotic and unpredictable nature of biology and physical limitations of the measurement devices, random noise is inevitable in the obtained measurements. In addition, the analysis of the measurements is complicated by the indirect measurement techniques, because complex issues, such as the functionality of human brain, cannot be measured directly without damaging the target.

The first means to analyze biomedical data was manual inspection. The manual analysis together with biomedical expertise has its place even in modern days e.g. in region-of-interest segmentation [1]. However, it rarely is suitable for quantitative data analysis. The main drawback is the variation between the conclusions of different experts as well as within the conclusions of a single expert [2, 3]. In addition, the capacity to process large amounts of data is very limited and many target properties, such as dynamic processes, are practically impossible to quantify manually.

The conventional approach to biomedical data quantification is based on data processing. The idea is to remove the noise from the data and to reverse the indirect measurement process so that the target could then be analyzed easily from the corrected data. However, the noise is irreversible,

random and exists in part on the same frequency interval as the useful signal. The quantification results are affected by the chosen (imperfect) noise processing method. Often the indirect measurement technique creates an ill-posed inversion problem which requires additional data regularization. Generally neither noise removal nor measurement inversion can be done without affecting the actual quantitative information from the target and the quality of results is thus decreased.

Various approaches to analyze biomedical data with parametric modeling have been presented [4, 5]. In parametric modeling, the prior knowledge about the target and the acquisition system is used to construct a model for the acquired data. The model is defined through parameters and the best parameter in terms of the data are searched for. However, it has been common to include simplifications to the models such as linear behavior assumption or data regularization in order to make the resulting parameter optimization faster and easier . We believe the data quantification should not be done in terms of mathematical convenience but in the terms of best quantification accuracy and robustness.

In this study we use stochastic parametric modeling together with numerical optimization to quantify fluorescence microscopy [6] and positron emission tomography (PET) data [7]. We perform the modeling without any noise filtering, data regularization or model modification in order to achieve optimal quantification accuracy. Our approach leads easily to challenging parameter optimization problems with multiple local optima and very large number of parameters. To solve the optimization problems we develop stochastic optimization algorithms and use a distributed computing environment. The stochastic algorithms provide the needed robustness for the optimization while the distributed computing environment makes it possible to solve the problem in a feasible time. We aim to achieve very accurate quantification through detailed models and accurate numerical optimization.

Chapter 2

Biomedical imaging

In this study a biomedical image denotes a measurement device output from a biological target. We limit ourselves roughly to image data and leave e.g. audio-based measurements out of the scope of this thesis. The term *medical* in the topic refers that we study dynamic properties and functionality of the biological targets rather than merely their anatomical structures.

2.1 Data distortion

A biomedical image can be seen as a projection from the target to the data, because each data element (image pixel) represents a combination of signal from several target locations. The nature of the measurement projection is application specific and may be caused e.g. by the physical limitations of the measurement device or indirect measurement technique. For example, in a fluorescence microscope image the distortion is caused, among other things, by the wave nature of the light and results as a blurred presentation of the target [8]. Thus fluorescence point source in the target is projected over to several pixels in the image. In PET the distortion is caused by the indirect measurement process and is much stronger to a human eye because the data does not look like the target at all [9]. See Fig.2.1 for simulated examples of data formation.

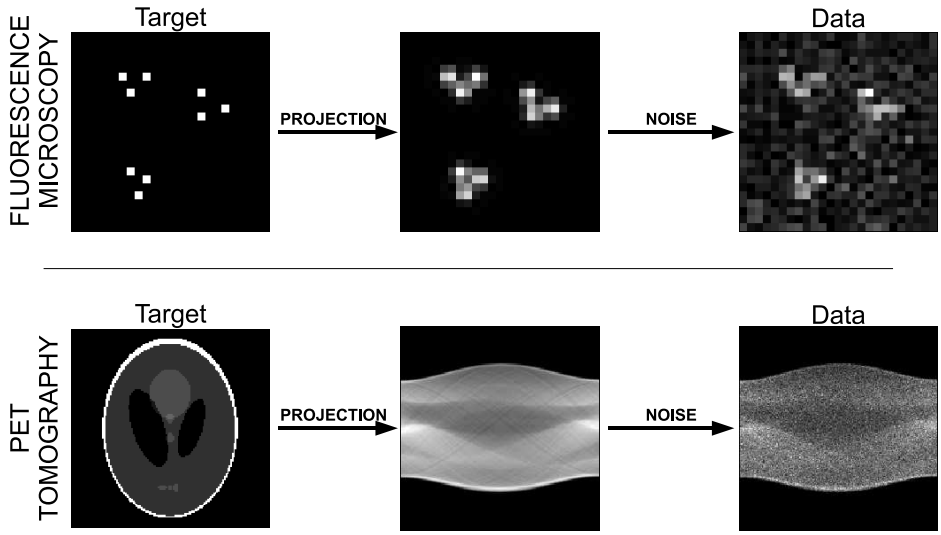

Figure 2.1: *Simulated examples of data formation through projection and noise contamination. In fluorescence microscopy the projection is caused by light diffraction and results as image blurring. The noise is mostly due to uncertainties in signal detection. In PET the projection is caused by the tomography measurement technique and the noise by the stochastic nature of radioactive tracer.*

The projections in the previous applications are different also by the signal processing or modeling point of view; in microscope image the smoothing is modeled through convolution [10] and in PET image through matrix multiplication [11]. Of course, these are merely models and in reality the data has been formed through stochastic physical phenomena like light diffraction in microscope lenses instead mathematical operations. We present the data formation in this study on a quite general level and in reality there are several additional application-specific details that affect the data and should be included in the model.

The fundamental question in terms of data quantification is whether the projection is uniquely and stably invertible. In many biomedical applications the inverse problems are ill-posed and thus a small change in

the data, such as noise contamination, may cause large and unpredictable changes in the quantification results if performed through inverse projection [12]. Inversion problems constitute a field of research in its own and they are not discussed in detail in this study. The bottom line is that we believe that the methods based on inverse projection are not the optimal way to analyze the data.

Let us present an example of an ill-posed inverse problem with a positive 5×5 matrix S defined as

$$S = \begin{pmatrix} 54 & 56 & 46 & 61 & 31 \\ 48 & 50 & 55 & 58 & 27 \\ 20 & 40 & 45 & 34 & 33 \\ 43 & 21 & 41 & 11 & 0 \\ 8 & 17 & 10 & 45 & 11 \end{pmatrix}.$$

By introducing a very small additive noise from uniform distribution $U(0, 0.5)$ we get a noisy matrix S_n, which in this case is defined as

$$S_n \approx \begin{pmatrix} 54.30 & 56.00 & 46.11 & 61.08 & 31.24 \\ 48.08 & 50.25 & 55.05 & 58.32 & 27.35 \\ 20.17 & 40.03 & 45.27 & 34.38 & 33.16 \\ 43.27 & 21.05 & 41.18 & 11.27 & 0.10 \\ 8.03 & 17.46 & 10.26 & 45.28 & 11.07 \end{pmatrix}.$$

Inverse matrices S^{-1} and S_n^{-1} exist and can easily be calculated, but they are totally different as can be seen in Fig.2.2. The numerical values of S^{-1} are on interval $\begin{bmatrix} -6290.76 & 5130.39 \end{bmatrix}$ while the values of S_n^{-1} are on interval $\begin{bmatrix} -258039.56 & 1.11 \end{bmatrix}$. Although the added noise was very small in comparison to the matrix values, it had drastic quantitative effects on the matrix inverse. The inversion problem with matrix S_n was ill-posed, which can be seen from the relatively low determinant value $det(S_n) \approx 0.127$.

2.2 Noise contamination

The measurement devices are not perfect in a sense that the exact amount of approaching signal cannot be detected flawlessly. Instead, due to device

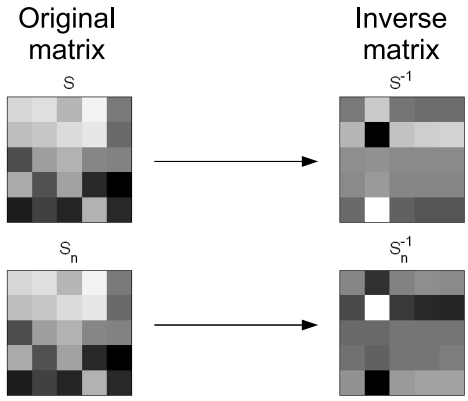

Figure 2.2: *On the left: matrix S without noise and with very small additive noise (S_n). On the right: Inversions of matrices S and S_n. The addition of a small error in S_n results as large differences in the inverted matrices. The problem is therefore ill-posed.*

imperfectness the measurement is in part stochastic by nature and generally two slightly different outputs would be obtained even from two equal targets. In addition to device imperfection, the detected signal itself can be stochastic by nature due to the variability in the used radioactive tracer, fluorescent label or the target itself [13]. All this stochastic undesired variation in the data is here defined as *noise*. For example, the exact wavelength emitted by a fluorescent label is not fixed but varies randomly around the theoretical wavelength. In the microscope signals of different wavelengths are detected differently and may even be blocked by the excitation filter.

Because the noise is stochastic by nature, it is impossible to know how much noise has been realized in a single pixel of the obtained data. Therefore the noise cannot be removed completely from the data e.g. by filtering as conventionally has been done. Inevitably any noise removal operation modifies the useful information in the data and thereby the quantification cannot be performed with optimal accuracy. Because the actual amount of noise in a single data element is not known, the quantitative effects of the noise removal are difficult to determine. However, statistically the noise re-

alizations in the data, or in a region of interest, obey typically some known probability distribution.

Chapter 3

Parametric modeling

3.1 Definition and motivation

In this study the quantification of biomedical data is performed by replicating the data formation. The data model is constructed from separate models for the target, the projection and the noise. These models are defined through parameters so that the data model is determined by the parameter values. The idea is to compress the essential information of the target into a finite, and preferably small, number of parameters which have a biophysically feasible meaning. The noise is always random in each data element and cannot be modeled as such, but we use stochastic noise model to take care of the statistical noise. Note that here we aim to model the data completely and thus we do not need to perform any pre-processing to the data. The overview of our approach can be seen in Fig.3.1.

There are two obvious advantages in the stochastic modeling approach. First, the projection is performed in the same direction (forward) as in the real acquisition and the inversion problem needs not to be solved. Secondly, the noise distribution can be taken into account and used in the data quantification instead of trying to remove the noise from the data. Thus optimal quantification accuracy can be achieved because the information in the data is not lost due to pre-processing.

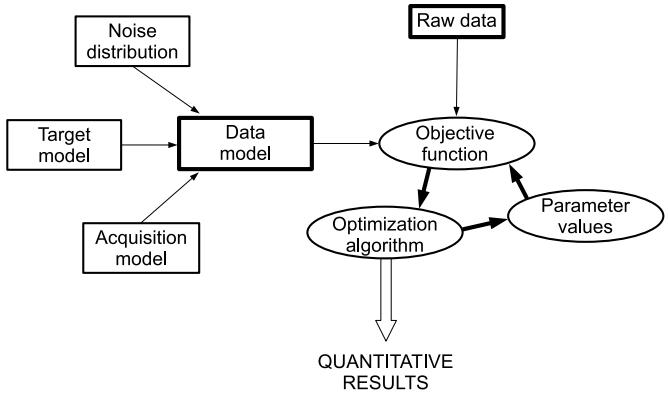

Figure 3.1: *A schematic overview of parametric modeling. The data model is built from the target model and the acquisition device model. Noise is included as a stochastic model with known probability distribution. The data model is defined through parameter values. The objective function value between the data model and the acquired data is maximized with the optimization algorithm to get the best parameter values. The optimal parameter values give the quantitative results.*

3.2 Model design

General principles

The parametric model must be designed from the biophysical point of view so that it corresponds to the reality as well as possible and the parameters have a feasible biophysical meaning. Naturally, biophysical expertise is required in the model design stage as well as prior knowledge about the anatomy and behavior of the target. In our selected applications, both the acquisition system[7, 8] and the target structure[14, 15] are known quite in detail and therefore parametric modeling is a feasible approach. We have worked in collaboration with biologists, medical doctors and physicists (among others) to create realistic models.

There have been studies where the quantification is performed through simplified parametric modeling. For example, in the analysis of fluorescence

dynamics from microscopy images spatial average values have been used [16]. However, it has been shown that ignoring the spatial variation causes significant error [17]. Similarly in PET, there are methods that *could* be used to quantify every data pixel, but they have been implemented to give only regional values [18]. It is obvious that with spatial averaging some of the information in the data is lost or modified. In some cases the model is turned into a linear problem, but the nature rarely follows linear scheme [19, 20]. The model simplifications are included probably due to difficulties with the resulting parameter optimization problem. We believe that the parameter optimization must serve the model design and the quantification needs, not vice versa.

Level of detail

Too detailed models pose a theoretical problem, because the balance between the amount of data and the number of quantified parameters gets worse. If there are less data points (containing unique information) than there are parameters, the problem is mathematically underdetermined and it may not be possible to define a unique solution for the problem.

The increase in the amount of detail does not always make the model more useful. For example, free diffusion of fluorescence in a cell is essentially based on the random motion of the individual fluorescent proteins [21, 22]. However, there is no sense in modeling every protein individually if the microscope pixel size is hundred times larger than the protein. The resulting optimization problem would have practically infinite number of different solutions and the quantification would be impossible. On the other hand, if too general model is used, the obtained parameters may not describe the target behavior appropriately.

In real biomedical measurement the amount of information does not necessarily increase by acquiring the data with higher (temporal or spatial) resolution and using correspondingly higher resolution model. The problem is that the total amount of detected signal is typically limited and independent of the used resolution. Then the relative variation in pixel values increases with the resolution due to Poisson distributed noise. This

can be seen if we define signal-to-noise ratio (SNR) as a proportion between the signal mean (μ) and the noise deviation (σ). When the signal is Poisson distributed with some Poisson parameter α, the signal-to-noise ratio is

$$\mathrm{SNR} = \frac{\mu\left(\mathrm{Poiss}(\alpha)\right)}{\sigma\left(\mathrm{Poiss}(\alpha)\right)} = \frac{\alpha}{\sqrt{\alpha}} = \sqrt{\alpha}. \tag{3.1}$$

If we now increase the resolution and thus decrease the amount of signal per pixel, the SNR decreases accordingly and the quantification results may not be improved. This is the case with both fluorescence microscopy and PET, where the increase in the (temporal or spatial) resolution may not lead to significantly better results.

We choose the level of detail in the model with the following criteria. First, the model must be realistic and sufficiently detailed in terms of biology and physiology, so that the *essential* components in the data formation are included. The parameters must have a biophysically feasible meaning and so that the data could be quantified through parameter optimization. On the other hand, the parameter optimization problem must theoretically have a unique solution. Therefore, the model cannot have more parameters than the information in the data allows.

Objective function

The data quantification with parametric modeling is performed by searching the best parameters in terms of the data. The definition of the "best parameters" is dependent on the chosen measure of goodness referred i.e. objective function. The parameters that are optimal in terms of one objective function are generally not optimal in terms of another.

Many simple and commonly used objective functions are based on the 2-norm (or euclidean distance) between the data and the model (e.g. [23, 24]). For example, the mean squared error (MSE) is defined self-evidently as

$$\mathrm{MSE}\left(\theta|D\right) = \min_{\theta} \sum_{i} \left(D_i - M_i(\theta)\right)^2, \tag{3.2}$$

where D denotes the data, θ the model parameters, $M(\theta)$ the data model (defined through parameters θ) and subindex i refers to a specific pixel

in the image. MSE is analogous to least squares fitting (LSQ), which has been widely used in various applications such as regression analysis and analysis of variance (ANOVA) techniques in fitting a known parametric function or model to a group of observations [25]. It is rarely feasible to try to completely minimize MSE distance between model and the data, because the data consists not only of some useful information but of the noise as well. Because the realized noise is always unknown it cannot be modeled with a deterministic criterion. In [26, 27, 28] the MSE criterion was compared to probabilistic approach and the accuracy of MSE was found less accurate. Also we have noted in our applications that 2-norm based objective functions are not optimal.

We measure the parameter goodness through probabilistic point of view. Although the noise is always random and unknown in a single pixel, each data pixel is a realization from some stochastic distribution. The noise distribution is obviously additional information and should be taken into use instead of just omitting or trying to remove it. For example, in applications based on radioactive decomposition the observed quantities are sums of individual independent instances, which is the very definition of the Poisson distribution and can thus be modeled.

A likelihood value gives the answer to a question *With what probability the data is originated from the given stochastic model?* Formally the likelihood can be written as:

$$L(\theta) = \min_{\theta} \prod_{i} p\left(D_i | M_i(\theta)\right), \tag{3.3}$$

where p denotes the probability or likelihood function and other notations are equal to those in Eq.3.2. Note that the probability distribution function is actually defined by the parameters θ. The question can then be further modified to *Which parameter values have the best probability to produce the data?* The answer to this question is searched for in data quantification with stochastic modeling.

Each data pixel is assigned a unique probability density/mass function from which the realized pixel value is assumed to be originated. The density function can be dependent on the model parameters, pixel location and the

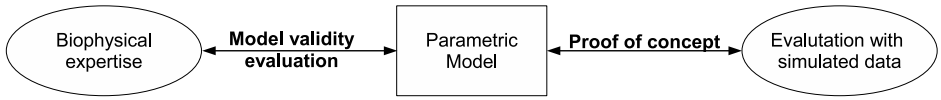

Figure 3.2: *The model design is evaluated with biophysical expertise together with prior knowledge of the target and the acquisition system. Simulated data is used to evaluate the internal integrity of the model and the performance of the parameter optimization algorithm.*

model pixel value. Thereby, each image pixel may have a different density function in comparison to other pixels, and their joint probability defines the likelihood function value.

Because the variation in the biomedical data comes often from several sources, it may be useful to construct a customized distribution into the model. For example, normal distribution is useful in many applications but negative values do not occur in many real acquisition devices. Thus the full normal distribution may not be adequate. However, a likelihood function can easily be modified to cover only the positive values and give zero likelihood for negative values.

3.3 Model evaluation

We build the parametric models from separate components whose validity has already been evaluated. For example, the PET model is based on widely accepted kinetic model of the radioactive tracer behavior in the tissue [29]. Also the discrete radon transform is widely accepted as a PET scanning model [11] and the free diffusion in a cell is modeled with the hundred years old Brownian motion theory [22]. Thus the parametric modeling approach does not necessarily require new theory to be derived but the current knowledge can be used in a novel way.

The final model evaluation should be done with real acquired data from a living target with known biophysical properties. However, the true properties of a living target are always unknown. In some cases it is possible to construct artificial targets, such as a physical PET phantom having known

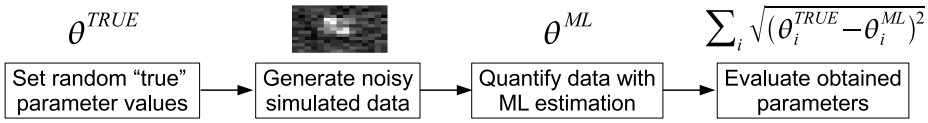

Figure 3.3: *Procedure to test the numerical parameter optimization with simulated data. The simulated data is created according to the target model and acquisition model with pre-defined parameters* θ^{TRUE}*. Noise is randomly drawn from the noise model. The simulated data is quantified with maximum likelihood estimation in order to obtain estimated* θ^{ML}*. The distance, such as 2-norm, between* θ^{TRUE} *and* θ^{ML} *tells how well the parameters can be estimated.*

radioactivity concentration, but they lack the natural variation of living targets. Comparing modeling results to a conventional method's results merely tells if there is a difference between them but not which results are the correct ones.

We suggest that the evaluation of the model should be done in two separate phases: the proof-of-concept and the biophysical authenticity (see Fig.3.2). The latter is omitted here as it should be done by biophysical experts based on the current biological knowledge of the target and physical knowledge of the acquisition system. In the proof-of-concept it needs to be shown that the model itself is robust, unambiguous and overall usable in every aspect. This can be tested by setting some predefined values for the model parameters, creating simulated data with the model and quantifying the data (see Fig. 3.3). It is not at all self-evident that the parameters can be determined correctly from the simulated data, even though the simulated data was created through the very same model as the parameters are estimated. The reason for this is the included random noise, which inevitably destroys a portion of the information in the data. With simulations it can be shown that the noisy data can be handled robustly and it can be estimated how reliable results can be obtained with a certain noise level. Another issue which can be tested with simulations is the parameter optimization. For example, the robustness of the optimization algorithm can be tested by selecting multiple different algorithm initializations for a

single simulated data set and monitoring whether the same likelihood value can be achieved from all initializations.

For extensive simulations it is necessary to have sufficient calculation power. Fortunately distributed calculation environments suit well to simulations because each simulation can be run separately on a computer of its own. Thereby as many simulations can be performed simultaneously as there are computers in the grid. The distributed calculation environments are described more in Section 4.2.

Chapter 4

Numerical Parameter Estimation

The parameter optimization phase must not be overlooked in the quantification because the quality of the results is crucially dependent on the success of the parameter optimization. The purpose of this section is to present typical issues that we have experienced with the parameter estimation of stochastic models. We also show the approaches that we have developed to solve the problems. Overall, we want to point out that the nature of the problem has to be analyzed properly and optimization method chosen accordingly.

4.1 Parameter space

Structure

The parameter space denotes the allowed parameter values that is searched for the likelihood function maximum. The structure of the parameter space is dependent on the model design and may vary largely between different models and applications. For example, if there are discrete parameters, such as the number of target components, the parameter space is not continuous

or differentiable in that dimension. This would set additional requirements for the optimization algorithm. On the other hand, in some cases the objective function is known to be convex which helps the optimization greatly. For a convex function, a local optimum has to be actually the global optimum.

The parameter space in stochastic models may be constrained based on the physiologically realistic range of values, but generally we aim to build unconstrained optimization problems. The motivation for this is, that we want to allow also unexpected results which in many cases may be the most valuable. However, in order to keep the model unambiguous we some times have to limit the possible range of parameter values. For example, in publications P3 and P2 we had to limit the intensity of spots to positive values in order to avoid unfeasible target objects with negative number of proteins and emitting "negative light". These objects would have made the model unambiguous. On the other hand, in publication P4 it was crucial to let the parameters obtain also negative and physically absurd values in order to obtain the pure theoretical maximum likelihood estimate distribution.

Due to the nonlinear nature of physical phenomena and the corresponding model equations, the stochastic parametric models create typically nonlinear optimization problems. For example, the behavior of kinetic parameters in publication P4 are determined through differential equations [29] and in publication P1 the solution to diffusion equation (heat kernel) contains an exponential term [21, 22]. This is unfortunate, because linear problems are well-studied and the tools to solve such optimization problems would be readily provided [30]. However, the requirement that the data should be a linear combination of the target or its properties is unrealistic due to typically nonlinear distortion and inter-dependencies between model components. There are attempts to convert the parameter estimation problem into a linear problem such as graphical analysis of PET data [31, 32]. Inevitably something is lost in the translation, although data processing may be sped up drastically.

For the maximum likelihood estimation to be feasible, the global maximum for the likelihood should be finite and unambiguous. The validity

of the requirements has to be checked in the model evaluation phase. In practice, the unambiguity requirement fails most easily with parametric models if the model is not designed carefully. If the model is unambiguous, the solution may look good but the quantification may not correspond to the reality at all. The finiteness requirement may fail e.g. if the number of target objects is estimated as an parameter. Then the likelihood may get better with every added object, because they may adapt to the realized errors in the pixels. Then the likelihood optimum is approached slowly by adding more target objects but may not be (theoretically) achieved with finite number of target objects.

Gradient

The gradient or derivative is a valuable aid in optimization. It tells the direction where the change in the likelihood function is locally largest. Unfortunately, the gradient does not point to the direction of global optimum but it can surely be used to speed up the optimization. Because the gradient tells a good direction to proceed in terms of all the estimated parameters and improvement in all the parameter values may be achieved with a single step. The drawback is that because the gradient always shows the *local* descent direction, it easily leads the optimization to a local optimum. Another problem is the choice of good step length to the direction of (negative) gradient, which is discussed later on in Section 4.3.

Often the gradient is not known e.g. due to complicated form of the likelihood function. Even if the analytic expression for the gradient could not be solved in practice, it does not mean that the likelihood function would not be theoretically differentiable. The gradient can be estimated numerically through finite differences approximation [33], but unfortunately this is slow to compute with large number of parameters. It takes twice as many function evaluations as there are parameters in each central (two-sided) finite difference estimation. Additionally, Hessian matrix can be utilized but the size of the full Hessian matrix is often huge and the optimization very slow. There are some modified algorithms, such as quasi-Newton methods [33], to estimate the finite difference and Hessian in order to make it

faster and less memory consuming. However, in practice we have noted that gradient-based optimization methods are useful only if the analytic form for the gradient is available. Especially with large-scale problems it is faster to use some non-gradient method than spend time in numerical gradient approximation which, after all, gives only a local descent direction.

Without analytic or estimated gradient the likelihood value, i.e. the joint probability of pixels in all time points, is the only information of the goodness of the specific parameters and the optimization problem must be solved with that information. The optimization problems can be solved without the gradient or its estimate, although the optimization problem is then a bit more challenging.

Scale

Capacity of measurement devices increases constantly and there are finer and finer structures that can be quantified with the most advanced devices [34]. Similarly grown biomedical knowledge allows, and requires, more detailed models and therefore more parameters to estimate. The problem from the optimization point of view is not the actual number of parameters, but the number of possible combinations between them. If we limit ourselves into thinking that there are two possible search directions (increase or decrease) for a single parameter, there are 2^n possible different search directions in a n-dimensional parameter space. Thereby if the number of parameters increases linearly, the number of possible search direction and the complexity of the optimization problem increases exponentially. Because the parameters usually have continuous values, in practice the amount of possible search directions is endless.

In theory, if the region of feasible parameter values is closed, the global optimum can be achieved by performing a local optimization from a sufficient number of initializations. However, even a relatively small number of parameters can lead to huge number of initializations. In the application of publication P3 we could typically have e.g. 16 parameters to estimate and if we take only 10 different initial values for each parameter, we end up with total number of 10^{16} different initializations for parameters. There-

fore we have aimed to optimization methods that are able to find the global optimum from all feasible initializations.

According to popular, although not scientifically justifiable or precisely accurate, Moore's law, the computation power approximately doubles every two years [35], i.e. is proportional to $\sqrt{2}^y$, where y denotes the number of years passed. For comparison, the number of search directions was previously noted to grow exponentially with a larger base number. This approximation is merely heuristic, but it suggests that the optimization problems with huge amount of parameters will not be solved simply by relying on the development of computers and increased computational power. The temporal and spatial resolution limits of the acquisition devices has hardly been reached yet, so it can be expected that there will be problems with the storage space and computation power in the future.

4.2 Distributed computing environment

Distributed computing environment, or a *grid* of computers, is a group of connected computing devices that can be used to perform a large computational task [36]. The reason why distributed environments are considered in this thesis is that they are more affordable than a super computer but provide massive amount of computation power. The grid may be constructed without extra infrastructure based on the idle processor time of desktop computers, such as classroom computers, and there are free of charge grid management systems available. Thus they are a realistic choice in universities and research centers.

The significant difference in comparison to a super computer is that the computers in a grid cannot run the same core processes or use the same data simultaneously due to slow communication between the computers. For example, if we have a strictly deterministic iterative optimization algorithm, which in each step needs the results from the previous step, it gets no instant advantage of the grid. Often in stochastic modeling neither the likelihood function nor the data is directly separable so that they could be quantified simultaneously in separate parts. Thus it is not so straightforward to solve

our parameter optimization problems in a computer grid. Therefore, we have modified and combined standard optimization algorithms in order to take advantage of the massive computational capacity of the distributed environment.

4.3 Developed methods

Randomized line search

If the gradient of the likelihood function is known, it obviously should be used to speed up the optimization. However, there still exists a question of how long step should be taken to the direction of the negative gradient. There are numerous theoretically good or even theoretically optimal ways (under certain conditions) to determine the step size [33]. In publication P4 we used a dynamic step length given by the Barzilai-Borwein (BB) formula [37, 38]. The optimization was quite slow due to large number of parameters but we managed to decrease the time cost after various modifications and program code tuning to a feasible level and could obtain the desired results. However, for more extensive simulations and for larger image resolution we had to develop a faster method and named it as *randomized line search*.

Let us assume that the step length given by the BB formula is not (locally) optimal on the one-dimensional line along the gradient. If we then make a *small* random mutation to the step length, there is theoretically a 50% chance that the new step length provides a better likelihood value. If we further make similar random mutations in n consecutive iterations of the algorithm, there is a 0.5^n probability that the step length has been improved in every iteration. This probabilistic property is the basis of our method. By using a grid of m computers, there is $1-(1-0.5^n)^m$ chance that at least one of the computers has improved the step length in every iteration. In practice we have noticed that the optimization can be sped up drastically this way by using random mutations to the BB step length and a "sufficient" number of computers. In exemplary Fig.4.1 there are 20 search paths with randomly chosen step lengths toward the negative gradient at each iteration. As can be seen, by good luck some of the search paths end

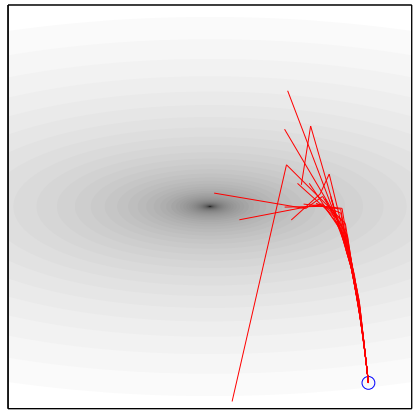

Figure 4.1: *Randomized line search on a two-dimensional surface. The red lines denote 20 search paths from the same initial location marked by a blue circle. The optimum is the dark spot in the middle of the image. Each search path used unique randomly chosen step lengths in the direction of their local negative gradient. Due to "good luck" in step length selection, some of the search paths arrive significantly closer to the optimum than the others.*

up to better locations than the others, which is the key idea of our method.

The above routine is further improved with re-initialization as follows. Each computer in the grid starts from the same initialization and optimize iteratively with unique random mutations to the BB step length. Some of the computers proceed to a better likelihood value than others due to "good luck" in step length choice (note that every computer has also its own unique gradient values after the first step). However, over time the luck settles and all the computers have had approximately equal number of lucky and unlucky step length mutations. To avoid this, within every few iterations we take the current best parameters among all the grid computers and use it as a new initialization for all the other grid computers. This way all the computers continue the optimization from the "most lucky" case.

Note that the globally best search path is not necessarily found by taking locally best step lengths. It is possible that a computer in a grid that has chosen many locally "bad" step lengths finally ends up to the best likelihood value after n iterations. This logic is implemented also in BB method, which doesn't aim to improve the likelihood value at every iteration but tries to find a good search path instead.

When the stochastic model is not directly separable and the selected parameter optimization method is strictly iterative, it is not straightforward to get any advantage of the distributed computing environment. However, this way we have used stochastic logic and a computer grid to improve a deterministic optimization method. The randomized step size method is compared to the unmodified BB method and to a method with locally optimal steps in the Chapter 6. The method can obviously be implemented with other step length selection methods than Barzilai-Borwein.

Iterative multiresolution

Another approach to speed up the optimization is to iteratively increase the resolution of the model during the optimization. In the beginning of the optimization we used a low resolution model with a small amount of parameters, but increased the resolution and the number of parameters as the optimization proceeded. The low resolution model obtained from the first iterations served as an initialization to the consecutive iterations with higher resolution. From the signal processing view, only the low frequency information in the data was used in the beginning of the optimization and the full information was taken into account at the end.

The pixels in lower resolution models are regional averages of the higher resolution model pixels. Thus the method works best when there is only small pixel-to-pixel variation in the data. Therefore the iterative multiresolution optimization did not work well in our publication P4, because we defined the virtual tissue to have significant heterogeneity i.e. spatial variation. The iterative multiresolution method has been used in some PET data quantification approaches [39, 40, 41]. However, the iterative resolution approach could be used also in temporal space and this might be a

more robust way because dynamic actions in the target appear more or less similar in consecutive time frames. In addition, by using iterative time resolution the spatial heterogeneity of the target can be preserved. Naturally it depends on the application which kind of iterative method can be used. The flexibility of the parametric models would allow to implement the iterative resolution logic in various ways of parameter estimation. For example, we have tried an iterative system matrix in PET parameter estimation so that at first only the largest system matrix elements are considered in the likelihood function and small values are set to zero. Our tests so far imply that the iterative system matrix may slightly speed up the optimization. We still need to test how useful this approach will be with large-scale problems, though.

In dynamic fluorescence microscopy data quantification the iterative multiresolution method has not been used in previous studies. We implemented iterative multiresolution approach in publication P1 and it was a key factor in achieving good accuracy. With our stochastic model the optimization was quite slow with full resolution, because the gradient was unknown. This is a good example of a case where advanced optimization methods allow more complex models and thus provide better quantification.

Evolutionary algorithms

Evolutionary optimization algorithms replicate natures way of improving species (or an isolated population) through gene mixing [42, 43]. In the evolutionary algorithms there is a population of parameter vectors and new parameter vectors are created by mixing the population members with each other. The objective function values of the parameter vectors are then used to decide which parameter vectors (population members) are killed and which survive. This cycle is repeated until the termination criterion is satisfied, for example until the diversity in the population is lost.

The two evolutionary algorithms that we have found most useful are genetic algorithm (GA) and differential evolution (DE). In GA the population members are typically mixed pairwise with a crossover operator and a whole generation of new parameter vectors is created at once [44]. This

new generation then replaces the previous population and the generation cycle is continued until termination criteria are met. Naturally, there are various different genetic algorithms with unique rules for mixing process, parameter mutation, generation survival and parent selection [45]. We used GA in the estimation of caveolae intensity distributions in publication P5. Our genetic algorithm was a modified version of the one presented in [46].

Another evolutionary algorithm that we have used is differential evolution [47, 48]. The fundamental difference in comparison to GA is that the population members are replaced one by one instead of updating the whole generation at once. Another difference is that in DE the new parameter vectors ("population's children") are mixed up from three vectors as

$$\theta_4 = \theta_1 + k(\theta_2 - \theta_3). \tag{4.1}$$

where typically $0.5 < k < 1.0$. Again, there are additional details, such as different crossover operators, that can be implemented in DE algorithms in several ways [49]. However, in practice we have based our algorithms on the standard DE/rand/1/bin formula because we could not obtain significant improvement in the optimization robustness or in the time cost with modified DE structures. We have used the dithered DE, though, in which the scale factor k in Eq. (4.1) is chosen randomly for every parameter vector creation [48]. Due to various different modifications of GA and DE algorithms, the separation between them can sometimes be a bit unclear. Thereby we haven't stuck to the formal definition of either algorithm but have modified our algorithms based on the needs and performance in our application. See Fig. 1 of Publication P3 for pseudo-code of our simplified DE algorithm. In biomedical quantification DE has been used in neural networks optimization but not widely in modeling approaches [50, 51].

We have noticed that evolutionary algorithms possess excellent capability to solve problems with multiple local optima and they are very robust with noisy data due to stochastic components such as crossover operator and random parent selection. However, the convergence is often quite slow and it cannot be guaranteed that the global optimum is reached in a single run of the algorithm. To overcome these issues, we have developed

hybrid algorithms that include both deterministic and stochastic components. The structure of the hybrid algorithm is described more in detail in Publication P1. The combination of Nelder-Mead and DE algorithm can be seen as a modification of memetic approach [52]. However, we have added also a random search step which has proven to be surprisingly useful in our simulations. We have compared the hybrid algorithm to several other algorithms, including pure genetic algorithm and a line search with finite difference gradient approximations, and found out that the hybrid worked most robustly.

The choice of algorithm settings, such as population size or mutation probability, is crucial in the performance of the optimization algorithm and with incorrect settings the algorithm may not be robust or may be useless due to slowness of the convergence. The most optimal algorithm settings can also be dependent e.g. on the data noise level and thus vary from data to another. We have developed a novel approach to determine good settings for the optimization algorithm by using simulated data. We quantified the simulated data with evolutionary algorithm having randomly chosen settings. The quantification was repeated with several different algorithm settings and the convergence monitored. The settings that produce the desired results most robustly and rapidly were searched for. Thus, we used a stochastic method to determine the settings of a stochastic algorithm, which was then used to quantify the data.

Chapter 5

Selected Applications

The scale of the selected applications is quite wide, ranging from cell imaging to whole body scanning. However, there are several analogous issues both in data formation and model building.

5.1 Fluorescence microscopy

Fluorescence microscope is used to study various properties of living targets such as drug response of a human cell. The imaging is based on the phenomena of fluorescence. The target or its components are labeled with a fluorescent protein (fluorophore) such as green fluorecence protein (GFP) and exposed to a light of specific wavelength [53, 54]. The fluorescent label gets excited due to the light energy and after a very small time interval it emits the light back. Due to the electron relaxation from the excitation state to the ground state, a portion of the vibration energy is lost and the emitted light has a higher wavelength than the excitation light. The phenomena is known as Stoke's shift. These two light sources can then be separated or blocked by their wavelength with a filter such as a stained glass or a dichroic mirror. Various different fluorescence microscope techniques have been developed to match specific imaging purposes [55]. In this study we concentrate on live cell imaging, although fixed-cell imaging, such as

Fluorescence In Situ Hybridization (FISH), is used widely..

The raw pixel size of a fluorescent microscope image is at best around 100nm which way more than e.g. size of a single protein or many other structures of interest. In addition, both the noise and the distortion effects exist and complicate the quantification. The acquired image is therefore not a direct one-to-one presentation of the target. The *noise* in the image is actually from several sources including illumination noise, shot noise and stray light. It is commonly assumed that the shot noise, originated from the uncertainty in the signal detectors, is dominant in the image. The shot noise is assumed to follow Poisson distribution due to unreliability of independent detections. Thereby the intensity values in the image can be assumed to be theoretically from Poisson distribution with the noiseless intensity as the Poisson parameter. Due to imperfect detection of the signal, the distribution in pixels is not exactly Poisson but a scaled (or shifted) Poisson distribution with unequal mean and variance. In fluorescence microscope images non-stochastic noise may also occur, such as fixed-pattern noise due to non-uniformity of the detectors [56]. However, this type of noise can be estimated from a separate measurement quite easily and thus we concentrate on the stochastic noise in this study.

The distortion in the microscope image is caused by the wave nature of the light and diffraction in the microscope lenses and other components, which results as a smoothed image. The function that models the smoothing, called point spread function (PSF), may be sometimes difficult to determine experimentally but theoretical formulas have been developed for it [57, 58]. For example, the lateral PSF in an ideal case where the resolution is limited by the lens diffraction, is

$$PSF(r) = \left(2J_1\left(\frac{2\pi Ar}{\lambda}\right)\right)^2 \frac{1}{r^2}, \tag{5.1}$$

where r is the (euclidean) distance from the center, A is the numerical aperture of the lens, J_1 is the Bessel function of first kind and λ is the wavelength of the used fluorescence. The PSF gives the proportional amount of signal that statistically diffracts to a distance r from the image centroid $r = 0$ of

a point source. Although the PSF is fixed with each distance r, in practice the diffraction occurs randomly.

The sharpness of the image is defined by the shape of the point spread function and from the signal processing point of view the image is formed through convolution between the (unknown) true target and the PSF. From the above formula (5.1) it can be seen that the shape of the PSF is dependent both on the lens and the fluorescence wavelength. Unfortunately these two issues cannot be improved much. Imaging with a shorter wavelength would give less smoothing, but UV light kills living targets or at least disturbs their behavior. Living cells need also a specific composition of immersion fluid. Thereby the image smoothing cannot be avoided just by improving the experiment settings because the amount of smoothing, i.e. the shape of PSF, is limited by the laws of physics.

The term "resolution" is used in several contexts with microscope images. Optical resolution is used to denote the lowest distance of two point sources at which they still can be detected individually (resolved) from the image. This limit is defined by the Rayleigh criterion. However, this limitation does not concern modeling approach because the point sources and their properties can be estimated quite accurately although they cannot be detected individually from the image. In this study, the *resolution* or *image resolution* denotes always the amount of pixels in the image.

5.1.1 Caveolin-1 protein distribution

Caveolae are invaginations on plasma membrane and they are assumed to play an important role in e.g. endocytosis and cell lipid transport [59, 60, 61]. This in turn may be an important factor e.g. in cardio-vascular diseases [62]. Caveolae consist of varying amount of caveolin-1 protein. The differences in the proteins amounts between caveolae structures was our point of interest.

The caveolin-1 protein can be labeled with green fluorescent protein (GFP) and then observed with a fluorescence microscope (see Fig. 5.1). The quantification is not trivial because the proteins are just few nanometers in size and thus beyond the microscope resolution to be detected in-

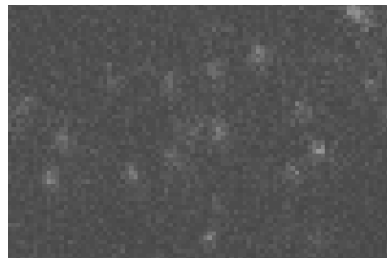

Figure 5.1: *A fluorescence microscope image of GFP stained caveolae. The bright spots present caveolae structures on a cell membrane. The noise in the image is clearly visible and makes the quantification more difficult.*

dividually. The diameter of the caveolae, however, is approximately 50-100nm and they can be seen as spots in the microscope image [63]. The spots' brightness is approximately, though not exactly, linearly related to the amount of fluorescent proteins in the caveolae. Thereby the amount of proteins can be quantified by the intensity of the corresponding spot in the image.

The conventional approach to microscopy data quantification is based on data processing and image restoration [64]. In the conventional approach the noise in the image is decreased e.g. by low-pass filtering but more complex methods have been proposed [65, 66]. However, the noise reduction cannot be perfect in practice with any method, because the noise in the image is random. The filtering performs inevitably some additional smoothing and destroys a portion of the useful information in the image. Especially the quantitative effects of noise reduction are a problem, although the visual quality of the image may be improved.

The projection, i.e. the smoothing according to the PSF, have been attempted to invert through deconvolution. The discrete two-dimensional deconvolution is an ill-posed problem[67] and typically cannot be solved exactly (except in some special cases[68]). Generally the original target cannot be reconstructed from noiseless data perfectly. The strong noise contamination makes the problem even more challenging. The processed image is then conventionally quantified by fitting a two-dimensional Gaus-

sian distribution to each spot, or using more heuristic methods [69, 63].

Our approach to this problem is to model the target and the acquisition parametrically and search for the best model parameters in terms of the data. The model image was built by first adding a constant value to the whole image representing both background and environmental signal as well as cell autofluorescence. In this case it was not of interest to know where this constant background signal was originated from. We investigated only small regions of the cell at once and allowed a different constant background value for each region and therefore the constant background assumption was realistic. We modeled the fluorescence images from plasma membrane with caveolae in two different ways.

In publication P3 the caveolae were modeled as two-dimensional Gaussian intensity distributions and the aim was to evaluate the quantitative errors caused by image filtering and local optimization algorithms. We used mean squared objective function which is analogous to maximum likelihood estimation with statistically independent image pixels. Although the pixel values are not statistically independent in fluorescence microscopy images due to the convolution with PSF, the results were sufficiently good. The function to be minimized was

$$f(I|\theta) = \frac{1}{nm} \sum_{x=1}^{n} \sum_{y=1}^{m} \left(I(x,y) - C(x,y|\theta) - \beta\right)^2, \tag{5.2}$$

where θ are the parameters to be estimated, I is the observed image with resolution $n \times m$, C is the image model and β is the background level. The details of the model can be seen in P3.

In P2 we further developed the method by replacing Gaussian component with theoretical point spread function. In the PSF-based model we assumed Poisson-type noise and used maximum likelihood estimation. We assumed that a pixel (x, y) of the noisy image N is distributed as

$$N(x,y) \sim Poiss\left(\rho\sqrt{C(x,y)}\right) - \sqrt{C(x,y)}\left(\rho - \sqrt{C(x,y)}\right), \tag{5.3}$$

where ρ controls the noise variance. The objective function was then defined

as the joint probability function of pixel-wise Poisson probabilities

$$f(I|\theta) = \prod_x^n \prod_y^m \frac{\lambda(x,y|\theta)^{\Omega(x,y|\theta)} e^{-\lambda(x,y|\theta)}}{\Omega(x,y|\theta)!} \tag{5.4}$$

where

$$\lambda(x,y|\theta) = \rho\sqrt{C(x,y|\theta)} \tag{5.5}$$

$$\Omega(x,y|\theta) = I(x,y) + \sqrt{C(x,y|\theta)}(\rho - \sqrt{C(x,y|\theta)}). \tag{5.6}$$

The motivation behind the above definition is that theoretically the noise in the microscope image is assumed to follow Poisson distribution with variance equal to the square root of the signal, i.e. noiseless pixel value [8]. However, perfect signal detection cannot be achieved and in practice the noise variance is not exactly equal to the signal square root. In our model, the variance of $N(x,y)$ in Eq. (5.3) is equal to $\rho\sqrt{C(x,y)}$ and thus the noise variance can be controlled with the estimated parameter ρ. The expectation of $N(x,y)$ is

$$\rho\sqrt{C(x,y)} - \rho\sqrt{C(x,y)} + C(x,y) = C(x,y)$$

and it is thus not affected by the noise variance parameter.

5.1.2 Fluorescence recovery after photobleaching

In fluorescence recovery after photobleaching (FRAP) experiment the fluorescent molecules are exposed to a high power laser in a (small) region of the cell. The process is called bleaching and it causes the fluorescent proteins lose their ability to fluoresce due to chemical damage. However, the total fluorescence in the bleach area recovers as the unbleached fluorescent proteins from the surrounding regions move to the bleached region and simultaneously the bleached proteins move out of the region. The recovery is monitored and the shape of the recovery curve imply e.g. the nature of protein movement (e.g. active transport or Brownian motion). The FRAP experiment can be used to study various cell structure or behavior issues such as bacterial cytoskeleton and protein binding [70, 71].

The first methods for FRAP quantification were presented in separate studies by Edidin et al and by Axelrod et al both in 1976 [72, 73]. They derived closed-form solutions for the intensity recovery in the bleached region over time for several experimental cases. The solutions are formally dependent on the half-time of the recovery which is the time point when 50 percent of the full intensity in the region has recovered. The diffusion rate can then be determined by calculating the average intensity of the bleached region from each image and by fitting them to the formal solution. The general outline of conventional FRAP quantification can be seen in Fig. 5.2. The analytical solutions contain additional parameters, such as bleach profile, which need to be pre-defined correctly in order to obtain a correct value for diffusion coefficient. Various other equations for FRAP quantification have been proposed for specific cases along increased biological knowledge of cell dynamics [74, 75, 76]. There is no doubt about the mathematical or physical correctness of these equations, but robustness with true noisy data is questionable. All these methods share the common problem of fitting noisy observations to a fixed one-dimensional theoretical parametric curve. Additional factors, such as heterogeneous fluorescence distribution and sample movement are conventionally treated separately through image processing which in turn affects the curve fitting results.

More recently, spatial and stochastic estimation methods have been introduced [77, 78]. In Tannert's method, the fluorescence concentration within each pixel in the image during the experiment is modeled by simulating diffusion [79]. Then the diffusion coefficient which gives the closest match to the acquired data in terms of 2-norm is searched for. The results are reported to be improved in comparison to conventional methods. This is expectable because the whole image data is used in such a spatial model. In Jonasson's study a likelihood objective function for diffusion analysis was derived, which gives more correct and flexible handling of the image noise [80]. Also the effect of point spread function was considered, although not implemented. However, strong assumptions were made about the experiment such as isotropic diffusion and strictly Gaussian bleach profile. Later, the method has been updated to handle also general bleach profiles [81]. The likelihood function was optimized numerically with the help of partial

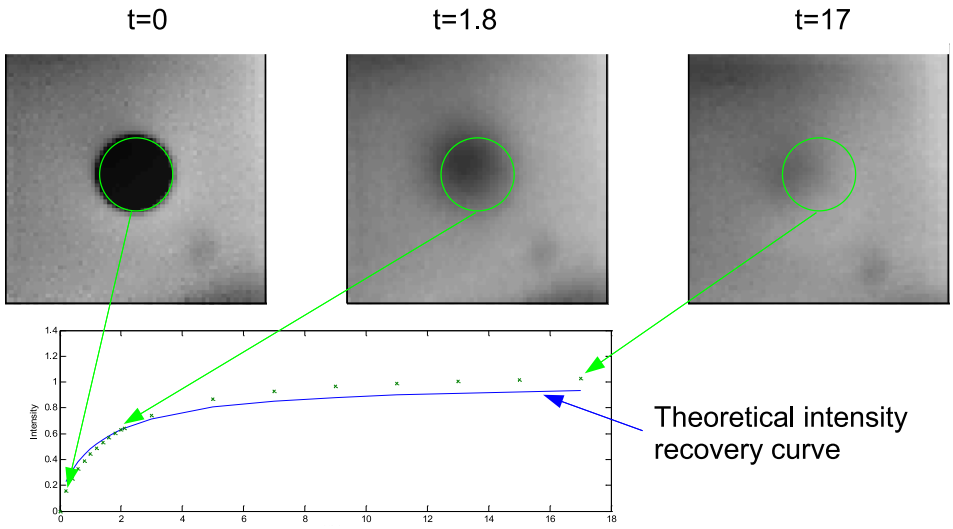

Figure 5.2: *Conventional FRAP quantification. The average intensity within the bleach region in each image of the FRAP time series is computed. Theoretical fluorescence recovery equation is fitted to the series of average intensities.*

derivatives. A good initialization was reported to be needed in order to achieve the maximum of the likelihood function due to several local optima in the parameter space. In practice it may be difficult to find a sufficiently good initialization and to evaluate the quality of an initialization.

In practice there are two severe issues that complicate the FRAP quantification and that may be difficult to avoid: the heterogeneity of the fluorescence distribution and the sample movement. Conventionally the theoretical diffusion equations have assumed that the fluorescent molecules are uniformly distributed in the equilibrium state, which is rarely the case in real cell. To tackle this problem, the data has been conventionally subtracted or divided by the average of pre-bleach images in order to convert the data into proportional fluorescence concentration form [82]. The approach is basically valid, if the exact inhomogeneous fluorescence distribution could be obtained from the pre-bleach images. However, noise will inevitably be

present in the pre-bleach image and the inhomogeneity estimate becomes erroneous.

The image movement can have several origins such as cell motility or mechanical drift. The accuracy of a fluorescence microscope is nanometers and it is difficult to achieve perfect stillness of a living target on that scale. If the microscope is even slightly unbalanced or the temperature changes on the other side of the sample, the target will move in the image. In conventional methods this movement must be separately estimated and pre-processed in order the theoretical diffusion solutions to be valid. This introduces an additional pre-processing step, which together with other possible processing steps like noise removal and deconvolution, takes the data even further from the original acquired images. The quantitative effects of the pre-processing steps to the are quite hard to determine.

In our parametric model, the sample movement and inhomogeneous fluorescence distribution are estimated simultaneously with the main point of interest i.e. the rate of fluorescence recovery. Overview of our FRAP model is shown in Fig. 5.3. The image movement is modeled as a dynamic linear dislocation of the whole image, i.e. the whole cell or the whole visible fraction of the cell. This kind of movement is mainly due to previously mentioned instrument or environmental issues. Cell motility, however, may be modeled more accurately with some non-linear or stochastic function. The movement is parametrized with two quantities: the velocity and the direction of movement. In two dimensions this can be defined with two variables: angular direction and movement speed. The movement has obviously quite a drastic effect on the match between the model output and the acquired data, and therefore correct values for movement parameters are often found easily right in the beginning of the optimization process.

Conventionally inhomogeneous distribution of fluorescent molecules in equilibrium state is estimated from the pre-bleach images. We modeled the inhomogeneity for each image pixel parametrically. We used the whole time series to estimate the inhomogeneity in interaction with the other model parameters. The inhomogeneity matrix defines the natural fluorescence capacity of the pixels, i.e. the relative amount of fluorescence protein in equilibrium state. It may not be necessary to model the heterogeneity of all

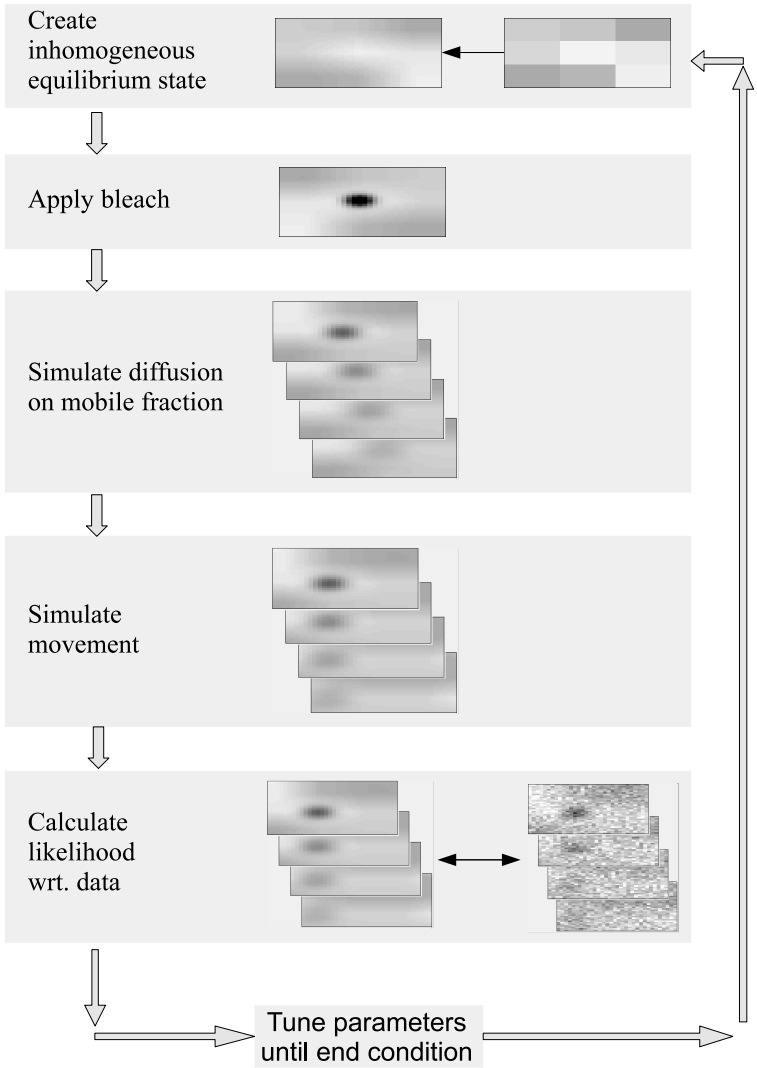

Figure 5.3: *The workflow of the developed FRAP quantification method. The model for dynamic FRAP data is built by creating heterogeneous initial distribution, applying bleach to the distribution, simulating fluorescence diffusion and finally virtually moving the cell in the image. All the steps are defined through the parameter values. The most likely parameters in terms of the acquired FRAP data are searched for.*

the pixels individually, though, but a lower resolution model can be used and interpolated to match the image resolution. Other components in the model are immobile fraction, bleach laser profile and noise variance.

The pixel-wise likelihood function was formulated as

$$p(x,t|\theta) = \frac{\left(\rho\sqrt{M(x,t|\theta)}\right)^{N(x,t)+\sqrt{M(x,t|\theta)}\left(\rho-\sqrt{M(x,t|\theta)}\right)} e^{-\left(\rho\sqrt{M(x,t|\theta)}\right)}}{\left(N(x,t)+\sqrt{M(x,t|\theta)}(\rho-\sqrt{M(x,t|\theta)})\right)!} \tag{5.7}$$

where x denotes the pixel, t denotes time, M is the data model constructed from parameters θ, N is the measured FRAP data and ρ controls the noise variance. The data model M for the measured FRAP data was constructed from several components as

$$M(\theta) = m\Big(\gamma B(H|\alpha,\beta,\Sigma)+\big(1-\gamma\big)\big(H+f(B(H|\alpha,\beta,\Sigma)-H|D)\big)\Big|\tau,r_l,r_z,\rho\Big), \tag{5.8}$$

where m is the function for image motion, f is the function for diffusion, B is the function for bleaching, H defines the cell inhomogeneity and θ is a vector consisting of all the estimated parameters.

To emphasize the fact that parametric models easily lead to quite complicated optimization problems, we show here how the actual image model M in the Eq. 5.8 is constructed. The outmost of the nested functions in the likelihood function (5.7) controls the image movement and is defined as

$$m(X|\tau,r_l,r_z) = \left\{ \int_{\tilde{x}-0.5}^{\tilde{x}+0.5} X(z)\ d^n z \;\middle|\; x \in X \right\}, \tag{5.9}$$

where $\tilde{x} = (x_1 + tr_l\cos\tau, x_2 + tr_l\sin\tau, x_3 + r_z)$. The movement is controlled with the three estimable parameters τ, r_l and r_z. The bleach profile Γ in bleaching function B is defined as

$$\Gamma(x|\alpha,\beta,\Sigma) = 1 - \min\left\{\beta, \frac{\alpha}{(2\pi)^{n/2}|\Sigma|^{1/2}}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)\right\}, \tag{5.10}$$

where α, β and Σ are estimable parameters and μ is the (known) bleach location. The bleaching is performed in practice as a multiplication with the bleach profile matrix Γ and the fluorescence concentration matrix. This model can be used to model both uniform or Gaussian bleach profile as well as incomplete or complete bleaching. See Fig. 2.1 in P1 for examples of bleaching profiles.

The actual point of interest, the diffusion, is modelled as a convolution

$$f(x,t) = \int_{\mathbb{R}^n} \Phi(x - z|D,t) F(x,0)\ d^n z, \tag{5.11}$$

where Φ is the discrete heat kernel defined as

$$\Phi(x|D,t) = \frac{1}{(4\pi Dt)^{n/2}} e^{-\frac{|x|^2}{4Dt}}. \tag{5.12}$$

The heterogeneity or cell topology H is defined as a matrix with a fixed resolution that is then upscaled to match to the microscope image resolution.

It can be easily seen from the above equation that the likelihood function would have been very difficult to differentiate. Actually, we have not even expressed the likelihood function (5.7) as a closed-form expression of the parameters but only of its components. The full analytical expression for the likelihood function probably would have been useless in practice. For the numerical modeling and optimization it was sufficient and a lot easier to build the components numerically and the combine them. For example, the integral equation (5.11), which controls the sample movement, simplifies in practice to a weighted sum of pixel values.

The FRAP application demonstrates the versatility of parametric models. The sample movement and inhomogeneity were implemented in the model and therefore the data "corrections" could be done simultaneously with the data quantification. In conventional approach the corrections are made prior to quantification and the possible errors made in this step will results in erroneous parameters of interest. Thus, in parametric modeling there is constant interaction between the data "correction" and quantification so that both of them can be performed with optimal accuracy.

5.2 Positron Emission Tomography

Positron emission tomography (PET) is an imaging technique which can be used to monitor functional processes in the body. In PET a radioactive tracer is injected into the blood flow and its concentrations in the tissue are monitored by the PET scanner. From the dynamic tissue concentrations, the tissue behavior can be deducted based on theoretical models such as kinetic model. The PET scanner does not produce a direct image of the tracer concentrations, but a collection of scans called sinogram. Each sinogram bin value is a sum (a line integral) over the tracer concentrations in the line between a scanner detector pair. In addition, the detected bin values are noisy due to stochastic behavior of the radioactive tracer.

The scanning can be described in PET by the Radon transform, which can be performed in practice through a matrix multiplication. The Radon transform is in theoretic continuous case uniquely invertible (proved already in 1917 by Johann Radon himself [83, 84]) but in true life the transform is discrete and the inversion problem is ill-posed [85]. Quantification of kinetic parameters, or even time-activities, from noisy PET sinogram data poses a Poisson inverse problem [86] which is difficult to be solved robustly. There have been studies where the kinetic parameters are computed directly from sinograms with stochastic parametric modeling [87, 39, 88]. However, all the previous methods have been applied with some kind of regularization or likelihood penalization [89]. The motivation for the data regularization or likelihood penalization is to avoid the maximum likelihood estimation error. However, we needed to obtain the kinetic parameters with the full estimation error in order to use the ML estimation theory.

We created in the publication P4 an unique method to determine the regional distribution of kinetic parameters directly from PET sinograms based on the maximum likelihood estimation. The determination of kinetic parameters through large-scale non-penalized likelihood estimation without any regularization was a new approach in itself, but we also used the maximum likelihood theory to correct the error in the obtained estimate histograms. Here we discuss mostly the resulting large-scale optimization problem and omit the statistical error correction.

The likelihood for kinetic parameter model was defined as

$$p(M|\theta) = \frac{1}{(2\pi)^{m/2}\sqrt{\det(\Sigma)}} \exp\left((M - S(\theta))^{\mathrm{T}} \Sigma^{-1} (M - S(\theta))\right), \quad (5.13)$$

where M is the data, Σ is a diagonal matrix constructed from S and $S(\theta)$ is the model sinogram created from the parameters θ. The details of the model structure can be found in the Publication P4.

Although we had an analytic formula for the gradient of the likelihood, the optimization problem was still quite challenging due to its large scale and time consuming calculations. With standard 128×128 image resolution we had about 15000 parameters to estimate without constraints. The parameters were allowed to obtain also negative values to get the statistical information, although physiologically the negative values make no sense. We needed to obtain the maximum likelihood estimates very accurately in order to be able to apply the ML estimation theory. With the 128×128 image resolution, a single calculation of likelihood/gradient function value took initially about 2.5 seconds on a standard desktop computer. Therefore, we had to pay a lot of attention to the programming of the estimation code and the selection of optimization method.

Chapter 6

Results

6.1 Caveolin-1 protein distribution

Our first approach to caveolin-1 quantification was simple and heuristic by nature [90]. The method managed to outperform the compared another heuristic method by Pelkmans mostly due to capability to quantify overlapping spots in the image. However, the method had severe weakness such as strong dependence on the initialization of the algorithm. The data was also pre-processed with deconvolution and filtering. Because we were not satisfied with the quantification results of the method, we developed the stochastic modeling approach.

The method based on mixtures of Gaussian distributions presented in Publication P3 worked well, is quite robust and easily modifiable to tackle a specific problem. In Publication P2 the method was further modified by replacing the Gaussian mixture with a mixture of the theoretical point spread functions. We compared these two methods with simulated data and the latter gave better results when the simulated target consisted of point sources. However, the PSF-based approach sets an additional requirement for the experiment because the target objects (here caveolae) have to be strictly smaller than the pixel size of the acquired image. If the objects are larger than the pixel size, and essentially not point sources, the resulting

spots in the image are not shaped as a PSF anymore.

In these methods the notable issue was that the numerical optimization was not so simple as one might first think. In P3 we showed that the chosen optimization algorithm affects to the quantification results especially with noisy data and in the case of several overlapping spots. Because our modeling approach always uses raw noisy data, the choice of optimization method needed to be done carefully. In this case, the number of parameters was actually pretty low (about 10-20 parameters). This implies that even relatively simple stochastic parametric models together with noisy data have a tendency to cause complicated optimization problems and therefore require careful consideration.

The stochastic parametric modeling was used successfully in Publication P5 to obtain biologically valuable results of caveolin-1 protein distributions under varying environments. Previously we tried to analyze the same data both with the Pelkmans method and with our heuristic method [63, 90]. However, there was too much variation in the results within the test cases and it was not possible to detect significant differences between test groups. In other words, the stochastic modeling approach was required to obtain biologically significant quantitative results.

6.2 Fluorescence recovery after photobleaching

We created simulated FRAP data to analyze the optimization problem. The parameter optimization problem in the FRAP model had multiple local optima and due to various interdependent model components, it was very difficult to find a good initialization. We tried some local optimization algorithms, such as Nelder-Mead method [91], and they failed to achieve the global optimum. To solve the problem, we developed a hybrid evolutionary algorithm based on differential evolution, random search and simplex algorithm. We chose to use a hybrid algorithm instead of a pure evolutionary algorithm because we had noticed that the pure evolutionary algorithms were too slow in this case. We chose DE/rand/1/bin variant of the differential evolution because our previous tests had shown that it was faster

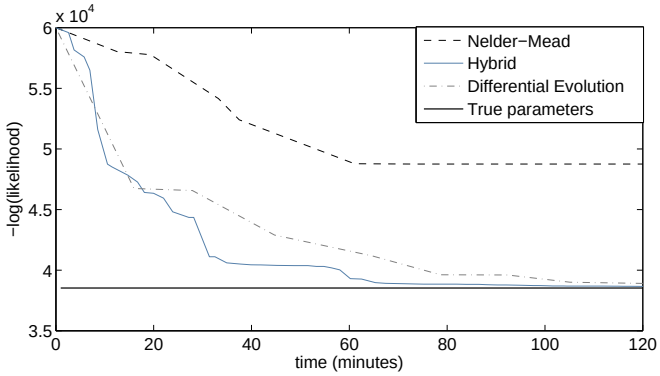

Figure 6.1: *Development of the negative log-likelihood value in parameter optimization with different algorithms and a simulated FRAP data. The constant True parameters line denotes the likelihood of the defined parameter values of the simulation, which is very close to the global likelihood maximum. The stochastic algorithms (Hybrid and Differential Evolution) manage to get close to the global optimum but Nelder-Mead converges to a local optima near the initialization. The lines appear piecewise linear because the optimization algorithms saved intermediate results quite rarely (2-20 minutes interval).*

that e.g. DE/rand/1/exp variant. The hybrid algorithm works basically so that the deterministic part (simplex) converges rapidly towards a local optimum, while the stochastic parts (random search, DE) help finding the global optimum. A comparison of optimization results with different algorithms can be seen in Fig. 6.1.

The resolution of the cell inhomogeneity model (matrix H in the likelihood function 5.7) has to be fixed in order to use our model. The problem is that if the resolution is too low, the inhomogeneity may not be modeled well enough and the quantification results may therefore be erroneous. On the other hand, if the resolution is high, there may be quite high number of parameters to be estimated. In Fig. 6.2 can be seen the estimated inhomogeneities with different resolutions from a real FRAP data. It is obvious that the inhomogeneity is significant in this data and the homogeneous, i.e. 1×1 resolution, is not enough to describe the target. In the highest resolu-

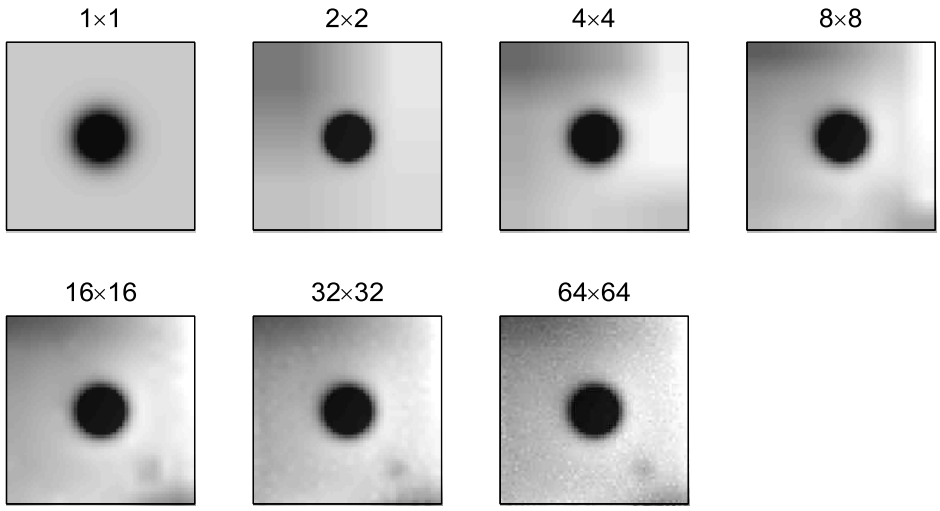

Figure 6.2: *The estimated inhomogeneous equilibrium state from a real FRAP data with various model resolutions. The images are interpolated from low resolution model to match the actual data resolution. The* (1×1) *image corresponds to a homogeneous model. There can be seen some undesired pixel-to-pixel variation in the full resolution model* (64×64).

tions there can be seen some pixel-to-pixel variation in the inhomogeneity matrix which indicates over-fitting. Fortunately, the over-fitting with high resolution inhomogeneity models does not seem to disturb the quantification because the obtained likelihood values and the quantified parameters are quite equal when using 8×8 or higher resolution. See Fig. 6.3 for development of the likelihood function value in terms of the inhomogeneity model resolution. The quantified parameter values are presented in Table 6.1 and it can be seen that there are no large changes in the parameter values when the resolution increases higher than 8×8. Especially, the rate of diffusion D is obtained equally well from 8×8 model and 64×64 model. However, if the inhomogeneity is completely ignored and a homogeneous fluorescence distribution assumed (1×1 model resolution) , the obtained parameter D value is largely erroneous.

Table 6.1: *Quantified parameter values for a real FRAP experiment with various inhomogeneity matrix resolutions.*

Resolution	D	β	ρ	r_l	τ	Σ	α
1×1	3.16	0.92	25.72	0.94	5.80	0.51	486.11
2×2	1.48	0.85	10.62	0.22	6.19	0.22	652.43
4×4	1.65	0.90	6.80	0.23	6.30	0.34	525.12
8×8	1.77	0.88	5.50	0.23	6.76	0.32	546.96
16×16	1.77	0.88	5.25	0.23	6.75	0.32	575.58
32×32	1.77	0.88	5.25	0.23	6.75	0.32	552.35
64×64	1.77	0.88	5.72	0.23	6.75	0.32	575.58

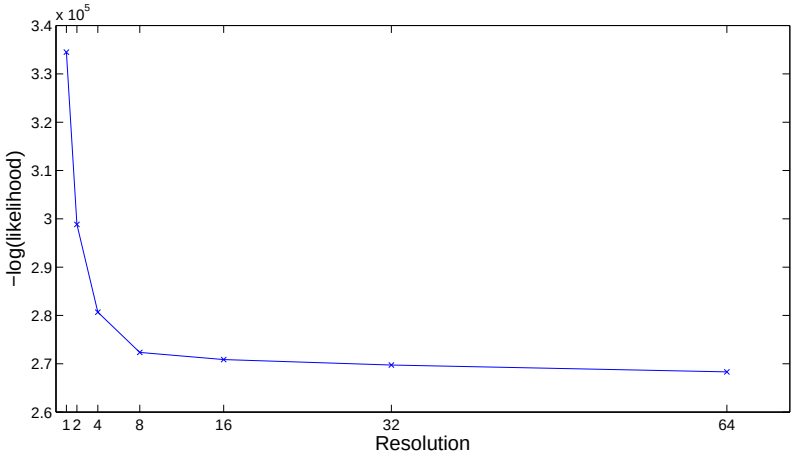

Figure 6.3: *The obtained maximum likelihood value from a real FRAP data with various model resolutions. The horizontal axis denotes the square root of the number of pixels in the FRAP model. The likelihood improves only slightly when resolution increases higher than* 8×8*. The values correspond to the images in Fig. 6.2. The likelihood with homogeneous (*1×1 *resolution) model is remarkably worse than with the heterogeneous models.*

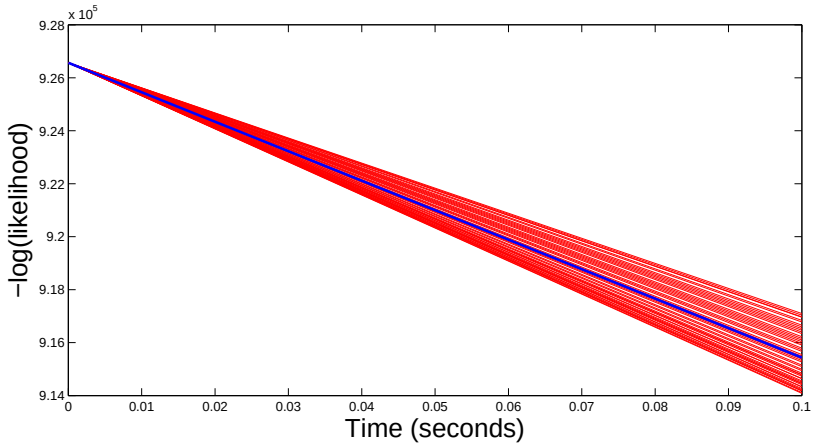

Figure 6.4: *Development of the negative log-likelihood in one iteration with 50 different step lengths from the same initialization to the direction of the negative gradient in simulated PET data. Blue line corresponds to the standard Barzilai-Borwein step length and the red lines corresponds to the randomized step lengths. Approximately 50 percent of the randomized step lengths give better likelihood values than the standard BB step length.*

6.3 Positron Emission Tomography

Publication P4 introduces an unique methodology to obtain the regional kinetic parameter histograms and to correct the estimation error in them. The kinetic parameters were estimated for each voxel, which resulted in a large-scale optimization problem with typically tens of thousands of parameters. In order to utilize maximum likelihood theory in the error correction, the parameters had to be estimated without any data regularization or likelihood function penalization. We managed to solve this huge optimization problem and obtained results that we could not have been achieved with any conventional method.

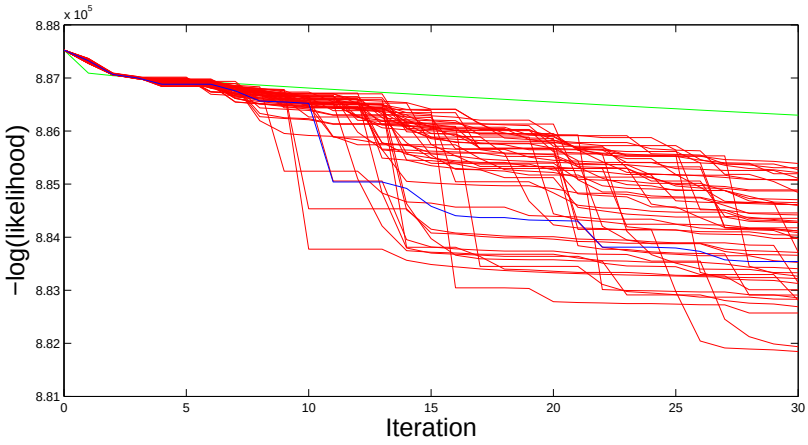

Figure 6.5: *Development of negative log-likelihood in 30 iterations and 50 different runs of randomized line search algorithm from the same initialization in simulated PET data. Blue line denotes the standard Barzilai-Borwein method. Green line denotes the steepest descent method, where locally optimal step length is used. Due to good luck in step length selection, some of the algorithm runs achieve remarkably better likelihood after 30 iterations than the unmodified BB method or the steepest descent method.*

Parameter optimization

Due to the large number of parameters and relatively slow calculation of likelihood/gradient function, a novel approach was needed to take advantage of the distributed calculation environment in the numerical parameter optimization. Each kinetic parameter contributes to several sinogram elements through Radon transform. Reversibly, (almost) every sinogram element is dependent from several kinetic parameters. Therefore the problem is not directly separable into subprocesses that could be solved simultaneously so that the final result could be obtained by merging the subprocesses.

We applied the randomized line search presented in Section 4.3. The behavior of the likelihood function in one iteration with 50 randomly chosen step lengths in the direction of negative gradient is shown in Fig.6.4.

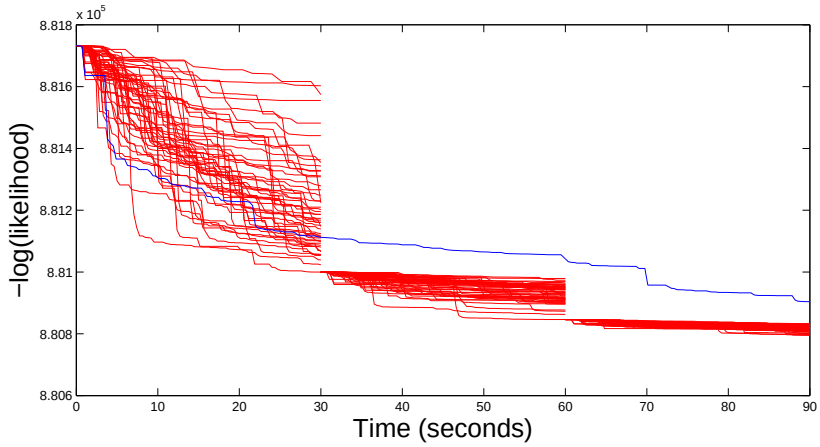

Figure 6.6: *Development of negative log-likelihood in 50 different parallel runs of randomized line search algorithm in simulated PET data. Blue line denotes the standard Barzilai-Borwein method. Green line denotes.the steepest descent method, where locally optimal step length is search along the gradient. The randomized line search algorithm is here re-initialized twice, after 30 seconds and after 60 seconds. The runs are re-initialized with the best parameter vector among all the runs and thus continue from the same point. In the long run this method outperforms the standard BB method remarkably.*

The random step-lengths were small mutations to the step-length proposed by the BB algorithm. Each optimization run was performed on a computer of its own in a distributed computing environment. Thus, it took approximately the same amount of time to calculate the 50 different runs as single run. As can be seen, approximately half of the random step-lengths produced a better likelihood than the unmodified BB method, which was expected because the mutations were small enough.

The optimization was then continued on these 50 computers from the unique parameter values they achieved in the first iteration of the algorithm. Each computer had now its own gradient because the runs on different computers ended in different places in the parameters space in the first iteration. Randomized steps were then taken in the direction of the

local negative gradients. The evolution of the negative logarithms of the likelihood values in the computers is presented in Fig.6.5. It can be seen that there were some few computers that have had "lucky" step-length guesses several times and had a substantially better likelihood than the standard BB method. However, after 30 iterations the luck already started to settle and there were only 12 computers that had better likelihood value than the BB. Therefore, it was time to re-initialize the runs.

In Fig.6.6 is shown the effect of re-initialization. It can be seen that in the long run this method gives significantly better results than the standard BB method. This way, we have used stochastic logic to improve a deterministic and strictly iterative optimization method. Of course, it is case dependent how often the re-initialization of the runs can be done and how much the optimization can be sped up with this approach. There are issues such as network delay and the number of grid computers that affect the success of the randomized line search method. In this case we used network file system which enabled fast communication between the grid computers.

In linear programming, the optimal step size e.g. the step which maximized the likelihood along the local gradient line, is not theoretically the best choice in the long run. We have noted this same behavior in our nonlinear applications and it can be seen also in Fig. 6.5. The "locally optimal" step size means here, that we used the step length that produces the largest change in the likelihood function. In the first iteration the optimal step size produces obviously the best result but after several iterations the other methods have better development. In a closer inspection, we noticed oscillatory behavior in the parameter values. In reality, the search for the optimal step size for every iteration would take a lot of time and the method would be quite useless. In the Fig.6.5 we compared only the number of iterations, but the optimal line search method used actually 30 times more computation time.

Chapter 7

Discussion

We have shown in this study that stochastic parametric modeling with powerful numerical optimization is a very accurate and robust method to quantify biomedical data. We also showed that the approach can be applied to real research problems.

We were able to determine the intensity distributions of caveolae structures from the real fluorescence microscope images very accurately with developed model. The developed method helped to achieve biologically significant difference between treatment groups in Publication P5 and the results have already received some interest [92, 93]. There are still some issues that could be included in the acquisition model and might further improve the quantification results. For example, we used quite a simplified point spread function in our experiments. An experimentally determined PSF could be used instead of a theoretical one which would allow e.g. position-dependent distortion modeling. It also would be interesting to correct the estimation error in the intensity histograms in a similar manner as with PET data in Publication P4. The error-correction might give more accurate estimates for absolute caveolin-1 protein concentrations in the caveolae. The differences between the treatment groups in our Publication P5 would probably stay unchanged because the estimation error would be quite invariant between the data sets. A more accurate caveolin-1 pro-

tein quantification might be useful e.g. in cancer research [94, 95].

Results with our FRAP model showed that the modeling approach was very accurate and could adapt well to the cell movement and heterogeneity. The results showed that both the movement and non-homogeneous fluorescence distribution affect the quantification results significantly and the error increases if these issues are ignored. There has been interest to take the inhomogeneity into account in FRAP quantification also in some other studies [96, 97]. However, the previous modeling solutions contain model simplification such as symmetric diffusion assumption or require separate estimation of the heterogeneity [77, 79]. Our approach is the most versatile so far because it requires no data pre-processing, estimates the heterogeneous fluorescence distribution automatically and compensates the cell movement. We quantified real microscopy data from a non-living target and the results imply that the method is able to handle real microscope data. Because of the excellent performance of the model so far, we believe the method could reveal some new information of the cell dynamics if applied to real experimental data. However, the model should still be tested with living cells in order to estimate whether the target behavior is modeled accurately enough. Also the three-dimensional nature of the true measurement should be taken into account. The optimization problem in FRAP model had many local optima and contained at worst thousands of parameter. Our hybrid algorithm was able to solve it robustly, accurately and with feasible time cost.

The idea of PET quantification with parametric modeling was presented already in 1985 [98]. However, no method with implementation has been published to estimate the kinetic parameters directly from the projections for every voxel without any regularization. Neither the correction of statistical maximum likelihood estimation error has been presented in any previous PET study. Thus we believe our PET data quantification method is unique and that our method is currently the most accurate way to determine the tissue heterogeneity from PET data. However, extensive comparative evaluation of our method and other possible methods should still be performed. Our acquisition model needs still to be improved in order to be used with real PET data. Fortunately many acquisition issues, such

as attenuation and scatter correction, can be directly implemented into the system matrix of the PET model [99, 100]. We were able to solve the huge optimization problem with tens of thousands of parameters through the Barzilai-Borwein gradient method. Our randomized line search method together with distributed computing environment speeds up the optimization significantly and will help us to apply the model with higher resolution and more complex models (three compartments with three or four kinetic parameters). The accurate solution of the large-scale optimization made it possible to determine the pure voxel-wise kinetic parameters including the ML estimation error. This was a crucial component in the success of our method and the results could not have been obtained with approximate optimization methods.

The optimization problems with stochastic models tend to be challenging because the model components take care of very different operations such as movement correction or tracer kinetics modeling. Some of the parameters affect the likelihood value through theoretical equations containing e.g. exponential terms while other parameters can have a very simple effect on the likelihood function. Thus the parameter space can be discrete in one dimension, continuous in another and have very large variation in the scale of the different dimensions. Therefore, the numerical solution of the problem requires not only large computational power but also suitable methods and proper analysis of the problem. Furthermore, this relates to the No Free Lunch (NFL) Theorem [101]. The NFL states that the average performance of any two optimization algorithms is equal over all possible problems. Therefore, we have not suggested any specific algorithm to be used in biomedical data quantification. Instead, we have presented purpose-built algorithms for each presented application and emphasize that the algorithms are problem-specific. The performance of our developed algorithms may not be the fastest over all possible methods but, most importantly, they manage to give out the desired quantitative results with a feasible time cost.

The nature of living biological targets limits the resolution and the quality of the data. For example, the fluorescence imaging causes photodamage to cells and therefore the number of images cannot be increased infinitely

without inflicting cell behavior [55]. Similarly, the quality of PET projections is dependent on the amount of injected radioactive tracer. Naturally, the image quality cannot be preferred to safety of the patient [102, 103]. Therefore, both temporal and spatial sampling in measurements from living targets is typically limited and the quality of the data may not be improved infinitely merely through improvement in the acquisition devices.

In order to correctly design the model, prior knowledge about the target and the acquisition system is required. If the model is too strictly defined based on prior knowledge, anything that is not modeled cannot be found from the target. This contradiction should be kept in mind when designing the model and not to set too narrow limits for the allowed values of parameters. The same contradiction has to be considered also with the conventional methods, though.

Stochastic parametric models do not offer a fast and easy way to data quantification because the better quantification accuracy comes with a price. The approach contains several stages where expertise is needed, such as target modeling, acquisition modeling, parameter optimization and implementation of the whole model into a computer program. Therefore scientific collaboration over several fields is required as well as sufficient computational resources in order to take full advantage of the approach. The optimization problems with our method are usually very time consuming and cannot be solved instantly by any algorithm. Thus stochastic parametric modeling is currently more suitable to research purposes than e.g. for clinical use. However, the stochastic modeling has got wider attention only quite recently and there is still much room for improvement in the methods and their implementation.

In the stochastic parametric modeling approach the acquired data is not modified by pre-processing and all the available information, such as noise distribution, is used to determine the model parameters. Therefore we believe this is the optimal way to quantify noisy biomedical measurements.

Bibliography

[1] T. W. Chow, S. Takeshita, K. Honjo, C. E. Pataky, P. L. St Jacques, M. L. Kusano, C. B. Caldwell, J. Ramirez, S. Black, and N. P. Verhoeff. Comparison of manual and semi-automated delineation of regions of interest for radioligand PET imaging analysis. *BMC nuclear medicine*, 7:2, 2007.

[2] S. L. Breen, J. Publicover, S. De Silva, G. Pond, K. Brock, B. O'Sullivan, B. Cummings, L. Dawson, A. Keller, J. Kim, J. Ringash, E. Yu, A. Hendler, and J. Waldron. Intraobserver and interobserver variability in GTV delineation on FDG-PET-CT images of head and neck cancers. *International journal of radiation oncology, biology, physics*, 68(3):763–770, 2007.

[3] T. W. Nattkemper, T. Twellmann, H. Ritter, and W. Schubert. Human vs machine: evaluation of fluorescence micrographs. *Computers in biology and medicine*, 33(1):31–43, 2003.

[4] J. Kay. Statistical models for PET and SPECT data. *Statistical methods in medical research*, 3(1):5–21, 1994.

[5] P. Sarder and A. Nehorai. Estimating locations of quantum-dot-encoded microparticles from ultra-high density 3-D microarrays. *IEEE transactions on nanobioscience*, 7(4):284–287, 2008.

[6] F. W. D. Rost. *Fluorescence microscopy, Volume I.* Cambridge University Press, 1992.

[7] M. M. Ter-Pogossian, M. E. Phelps, E. J. Hoffman, and N. A. Mullani. A positron-emission transaxial tomograph for nuclear imaging (PET). *Radiology*, 114(1):8998, 1975.

[8] R. D. Goldmana and D. L. Spector. *Live cell imaging: a laboratory manual*. CSHL Press, 2005.

[9] R. E. Carson, M-E. Daube-Witherspoon, and P. Herscovitch. *Quantitative functional brain imaging with positron emission tomography*. Elsevier, 1998.

[10] Michiel Müller. *Introduction to confocal fluorescence microscopy*. SPIE Press, 2006.

[11] G. Beylkin. Discrete radon transform. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 35(2):162 – 172, feb 1987.

[12] M. M. Lavrentev, S. M. Zerkal, and O. E. Trofimov. *Computer modelling in tomography and ill-posed problems*. VSP, 2001.

[13] Victor W Pike. PET radiotracers: crossing the blood-brain barrier and surviving metabolism. *Trends Pharmacol Sci*, 30(8):431–440, Aug 2009.

[14] B. Alberts. *Molecular biology of the cell: Reference edition*. Garland Science, 2008.

[15] H. Damasio. *Human brain anatomy in computerized images*. Oxford University Press US, 2005.

[16] E. A. Reits and J. J. Neefjes. From fixed to FRAP: measuring protein mobility and activity in living cells. *Nat Cell Biol*, 3(6):E145–E147, Jun 2001.

[17] B. L. Sprague, F. Muller, R. L. Pego, P. M. Bungay, D. A. Stavreva, and J. G. McNally. Analysis of binding at a single spatially localized cluster of binding sites by fluorescence recovery after photobleaching. *Biophysical journal*, 91(4):1169–1171, 2006.

[18] I. S. Yetik and J. Qi. Direct estimation of kinetic parameters from the sinogram with an unknown blood function. In *Biomedical Imaging: Nano to Macro, 2006. 3rd IEEE International Symposium on*, pages 295 –298, april 2006.

[19] M. Ichise, H. Toyama, R. B. Innis, and R. E. Carson. Strategies to improve neuroreceptor parameter estimation by linear regression analysis. *J Cereb Blood Flow Metab*, 22(10):1271–1281, Oct 2002.

[20] P. G. Coxson, R. H. Huesman, and L. Borland. Consequences of using a simplified kinetic model for dynamic pet data. *J Nucl Med*, 38(4):660–667, Apr 1997.

[21] A. Einstein. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik*, 322, 1905.

[22] A. Einstein, A. Beck, and P. Havas. *The Collected Papers of Albert Einstein: The Swiss years, writings, 1900-1909*. Princeton University Press, 1989.

[23] J. Zhou, J.-L. Coatrieux, and L. Luo. Noniterative sequential weighted least squares algorithm for positron emission tomography reconstruction. *Comput Med Imaging Graph*, 32(8):710–719, Dec 2008.

[24] J. Tellinghuisen. The least-squares analysis of data from binding and enzyme kinetics studies weights, bias, and confidence intervals in usual and unusual situations. *Methods Enzymol*, 467:499–529, 2009.

[25] M. H. DeGroot. *Probability and Statistics*. Addison-Wesley, 1980.

[26] P. F. Price. A comparison of the least-squares and maximum-likelihood estimators for counts of radiation quanta which follow a Poisson distribution. *Acta Crystallographica Section A*, 35(1):57–60, Jan 1979.

[27] F. O'Sullivan. A study of least squares and maximum likelihood for image reconstruction in positron emission tomography. *The Annals of Statistics*, 23(4):1267–1300, 1995.

[28] A. V. Abraham, S. Ram, J. Chao, E. S. Ward, and R. J. Ober. Quantitative study of single molecule location estimation techniques. *Opt Express*, 17(26):23352–23373, Dec 2009.

[29] H. Watabe, Y. Ikoma, Y. Kimura, M. Naganawa, and M. Shidahara. PET kinetic analysis–compartmental model. *Annals of nuclear medicine*, 20(9):583–588, 2006.

[30] S. I. Gass. *Linear programming: methods and applications*. Courier Dover Publications, 5th edition, 2003.

[31] J. Yguerabide, J. A. Schmidt, and E. E. Yguerabide. Lateral mobility in membranes as detected by fluorescence recovery after photobleaching. *Biophysical journal*, 40(1):69–75, 1982.

[32] C. S. Patlak, R. G. Blasberg, and J. D. Fenstermacher. Graphical evaluation of blood-to-brain transfer constants from multiple-time uptake data. *J Cereb Blood Flow Metab*, 3(1):1–7, Mar 1983.

[33] J. Nocedal and S. J. Wright. *Numerical optimization*. Springer, 1999.

[34] G. H. Patterson. Fluorescence microscopy below the diffraction limit. *Seminars in cell & developmental biology*, 20(8):886–893, 2009.

[35] G. E. Moore. Cramming more components onto integrated circuits. *Electronics*, 38(8), 1965.

[36] H. Attiya and J. Welch. *Distributed computing: fundamentals, simulations and advanced topics*. Wiley-IEEE, 2nd edition, 2004.

[37] J. Barzilai and J. M. Borwein. Two-Point Step Size Gradient Methods. *IMA J Numer Anal*, 8(1):141–148, 1988.

[38] M. Raydan. *The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem*, volume 7, pages 26–33. 1997.

[39] M. E. Kamasak, C. A. Bouman, E. D. Morris, and K. Sauer. Direct reconstruction of kinetic parameter images from dynamic PET data. *IEEE transactions on medical imaging*, 24(5):636–640, 2005.

[40] P. Thevenaz, U. E. Ruttimann, and M. Unser. A pyramid approach to subpixel registration based on intensity. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 7(1):27–41, 1998.

[41] Y. De Witte, J. Vlassenbroeck, and L. Van Hoorebeke. A Multiresolution Approach to Iterative Reconstruction Algorithms in X-Ray Computed Tomography. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 2010.

[42] D. Ashlock. *Evolutionary computation for modeling and optimization.* Springer, 2006.

[43] K. A. De Jong. *Evolutionary computation : a unified approach.* Cambridge, Mass. : MIT Press, 2006.

[44] D. E. Goldberg. *Genetic algorithms in search, optimization, and machine learning.* Addison-Wesley, 1989.

[45] M. Gen and R. Cheng. *Genetic algorithms and engineering optimization.* Wiley-IEEE, 2000.

[46] J. Tohka, E. Krestyannikov, I. D. Dinov, A. M. Graham, D. W. Shattuck, U. Ruotsalainen, and A. W. Toga. Genetic algorithms for finite mixture model based voxel classification in neuroimaging. *IEEE transactions on medical imaging*, 26(5):696–711, 2007.

[47] K.V. Price. Differential evolution: a fast and simple numerical optimizer. In *Fuzzy Information Processing Society, 1996. NAFIPS. 1996*

Biennial Conference of the North American, pages 524 –527, 19-22 1996.

[48] K. V. Price, R. M. Storn, and J. A. Lampinen. *Differential evolution - A practical approach to global optimization*. Natural computing series. Springer-Verlag, 2007.

[49] F. Neri and V. Tirronen. Recent advances in differential evolution: a survey and experimental analysis. *Artificial Intelligence Review*, 33(1-2):61–106, 2010.

[50] V. Gupta, C. C. Chan, and P. T. Sian. A differential evolution approach to PET image de-noising. *Conf Proc IEEE Eng Med Biol Soc*, 2007:4173–4176, 2007.

[51] W. Van Geit, E. De Schutter, and P. Achard. Automated neuron model optimization techniques: a review. *Biological cybernetics*, 99(4-5):241–241, 2008.

[52] E. Mininno and F. Neri. A memetic differential evolution approach in noisy optimization. *Memetic Computing*, 2(2):111–135, June 2010.

[53] T. Stearns. Green fluorescent protein. The green revolution. *Curr Biol*, 5(3):262–264, Mar 1995.

[54] R. Y. Tsien. The green fluorescent protein. *Annual Review of Biochemistry*, 67:509–544, 1998.

[55] David J Stephens and Victoria J Allan. Light microscopy techniques for live cell imaging. *Science*, 300(5616):82–86, Apr 2003.

[56] J. B. Pawley. *Handbook of biological confocal microscopy*. Springer, 2006.

[57] D. Thomann, D. R. Rines, P. K. Sorger, and G. Danuser. Automatic fluorescent tag detection in 3D with super-resolution: application to the analysis of chromosome movement. *Journal of microscopy*, 208(Pt 1):49–64, 2002.

[58] S. F. Gibson and F. Lanni. Experimental test of an analytical model of aberration in an oil-immersion objective lens used in three-dimensional light microscopy. *Journal of the Optical Society of America. A, Optics and image science*, 9(1):154–156, 1992.

[59] R. G. Anderson. The caveolae membrane system. *Annual review of biochemistry*, 67:199–225, 1998.

[60] R. G. Parton and K. Simons. The multiple faces of caveolae. *Nature reviews. Molecular cell biology*, 8(3):185–194, 2007.

[61] X. Su and N. A. Abumrad. Cellular fatty acid uptake: a pathway under construction. *Trends in endocrinology and metabolism: TEM*, 20(2):72–77, 2009.

[62] A. Rahman and K. Sward. The role of caveolin-1 in cardiovascular regulation. *Acta Physiol (Oxf)*, 195(2):231–235, 2009.

[63] L. Pelkmans and M. Zerial. Kinase-regulated quantal assemblies and kiss-and-run recycling of caveolae. *Nature*, 436(7047):128–133, Jul 2005.

[64] P. J. Verveer, M. J. Gemkow, and T. M. Jovin. A comparison of image restoration approaches applied to three-dimensional confocal and wide-field fluorescence microscopy. *Journal of microscopy*, 193(1):50–61, 1999.

[65] S. Wolter, M. Schüttpelz, M. Tscherepanow, S. Van de Linde, M. Heilemann, and M. Sauer. Real-time computation of subdiffraction-resolution fluorescence images. *J Microsc*, 237(1):12–22, Jan 2010.

[66] J. Boulanger, C. Kervrann, P. Bouthemy, P. Elbau, J.-B. Sibarita, and J. Salamero. Patch-based nonlocal functional for denoising fluorescence microscopy image sequences. *IEEE Trans Med Imaging*, 29(2):442–454, Feb 2010.

[67] F. Colonna and G. R. Easley. The multichannel deconvolution problem: A discrete analysis. *Journal of Fourier Analysis and Applications*, 10, 2004.

[68] A. Karoui, A. Ammari, and B. Selmi. Exact solutions of some discrete deconvolution problems. *Journal of Fourier Analysis and Applications*, 13, 2007.

[69] C. M. Anderson, G. N. Georgiou, I. E. Morrison, G. V. Stevenson, and R. J. Cherry. Tracking of cell surface receptors by fluorescence digital imaging microscopy using a charge-coupled device camera. Low-density lipoprotein and influenza virus receptor mobility at 4 degrees C. *J Cell Sci*, 101 (Pt 2):415–425, Feb 1992.

[70] P. Vats, J. Yu, and L. Rothfield. The dynamic nature of the bacterial cytoskeleton. *Cell Mol Life Sci*, 66(20):3353–3362, Oct 2009.

[71] B. L. Sprague and J. G. McNally. FRAP analysis of binding: proper and fitting. *Trends Cell Biol*, 15(2):84–91, Feb 2005.

[72] D. Axelrod, D. E. Koppel, J. Schlessinger, E. Elson, and W. W. Webb. Mobility measurement by analysis of fluorescence photobleaching recovery kinetics. *Biophys J*, 16(9):1055–1069, Sep 1976.

[73] M. Edidin, Y. Zagyansky, and T. J. Lardner. Measurement of membrane protein lateral diffusion in single cells. *Science*, 191(4226):466–468, Feb 1976.

[74] D. M. Soumpasis. Theoretical analysis of fluorescence photobleaching recovery experiments. *Biophys J*, 41(1):95–97, Jan 1983.

[75] M. Kang and A. K. Kenworthy. A closed-form analytic expression for FRAP formula for the binding diffusion model. *Biophys J*, 95(2):L13–L15, Jul 2008.

[76] K. S. Zadeh and H. J. Montas. A class of exact solutions for biomacromolecule diffusion-reaction in live cells. *J Theor Biol*, Mar 2010.

[77] O. N. Irrechukwu and M. E. Levenston. Improved estimation of solute diffusivity through numerical analysis of FRAP experiments. *Cellular and Molecular Bioengineering*, 2:104 – 117, 2009.

[78] Y. H. Sniekers and C. C. van Donkelaar. Determining diffusion coefficients in inhomogeneous tissues using fluorescence recovery after photobleaching. *Biophys J*, 89(2):1302–1307, Aug 2005.

[79] A. Tannert, S. Tannert, S. Burgold, and M. Schaefer. Convolution-based one and two component FRAP analysis: theory and application. *Eur Biophys J*, 38(5):649–661, Jun 2009.

[80] J. K. Jonasson, N. Lorén, P. Olofsson, M. Nydén, and M. Rudemo. A pixel-based likelihood framework for analysis of fluorescence recovery after photobleaching data. *J Microsc*, 232(2):260–269, Nov 2008.

[81] J. K. Jonasson, J. Hagman, N. Lorén, D. Bernin, M. Nydén, and M. Rudemo. Pixel-based analysis of FRAP data with a general initial bleaching profile. *J Microsc*, 239(2):142–153, Aug 2010.

[82] E. D. Siggia, J. Lippincott-Schwartz, and S. Bekiranov. Diffusion in inhomogeneous media: theory and simulations applied to whole cell photobleach recovery. *Biophys J*, 79(4):1761–1770, Oct 2000.

[83] J. Radon. Über die bestimmung von funktionen durch ihre integralwerte längs gewisser mannigfaltigkeiten. *Ber. Schs. Akad. Wissenschaft. Leipzig Math. Phys. Kl.*, 69:262267, 1917.

[84] J. Radon. On the determination of functions from their integral values along certain manifolds. *Medical Imaging, IEEE Transactions on*, 5(4):170 –176, 1986.

[85] L. B. Klebanov, T. J. Kozubowski, and S. T. Rachev. *Ill-posed problems in probability and stability of random sums*. Nova Publishers, 2006.

[86] A. Antoniadis and J. Bigot. Poisson inverse problems. *The Annals of Statistics*, 34:2132–2158, 2006.

[87] I. S. Yetik and J. Qi. Direct estimation of kinetic parameters from the sinogram with an unknown blood function. In *Biomedical Imaging: Nano to Macro, 2006. 3rd IEEE International Symposium on*, pages 295 –298, april 2006.

[88] G. Wang and J. Qi. Analysis of penalized likelihood image reconstruction for dynamic PET quantification. *IEEE transactions on medical imaging*, 28(4):608–610, 2009.

[89] J. Verhaeghe, D. Van de Ville, I. Khalidov, Y. D'Asseler, I. Lemahieu, and M. Unser. Dynamic PET reconstruction using wavelet regularization with adapted basis functions. *IEEE transactions on medical imaging*, 27(7):943–949, 2008.

[90] H. Pölönen, M. Jansen, J. Tohka, E. Ikonen, and U. Ruotsalainen. Automatic method for caveolar structure detection and intensity distribution analysis from microscopy images. In *Proc. European Signal Processing Conference (EUSIPCO)*, pages 1107–1111, 2007.

[91] J. A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7(4):308–313, January 1965.

[92] D. Cunningham, K. Spychala, K. W. McLarren, L. A. Garza, C. F. Boerkoel, and G. E. Herman. Developmental expression pattern of the cholesterogenic enzyme NSDHL and negative selection of NSDHL-deficient cells in the heterozygous Bpa1H/+ mouse. *Molecular Genetics and Metabolism*, 98(4):356 – 366, 2009.

[93] J. Sanchez-Wandelmer, A. Davalos, E. Herrera, M. Giera, S. Cano, G. de la Pena, M. A. Lasuncion, and R. Busto. Inhibition of cholesterol biosynthesis disrupts lipid raft/caveolae and affects insulin receptor activation in 3T3-L1 preadipocytes. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1788(9):1731 – 1739, 2009.

[94] T. C. Thompson, S. A. Tahir, L. Li, M. Watanabe, K. Naruishi, G. Yang, D. Kadmon, C. J. Logothetis, P. Troncoso, C. Ren,

A. Goltsov, and S. Park. The role of caveolin-1 in prostate cancer: clinical implications. *Prostate Cancer Prostatic Dis*, 13(1):6–11, Mar 2010.

[95] P. Dias Pereira, C. C. Lopes, A. J F Matos, P. P. Cortez, F. Grtner, R. Medeiros, and C. Lopes. Caveolin-1 in diagnosis and prognosis of canine mammary tumours: Comparison of evaluation systems. *J Comp Pathol*, Feb 2010.

[96] J. Beaudouin, F. Mora-Bermudez, T. Klee, N. Daigle, and J. Ellenberg. Dissecting the contribution of diffusion and interactions to the mobility of nuclear proteins. *Biophys J*, 90(6):1878–1894, Mar 2006.

[97] M. A. Hallen and A. T. Layton. Expanding the scope of quantitative FRAP analysis. *J Theor Biol*, 262(2):295–305, Jan 2010.

[98] R. E. Carson and K. Lange. The EM parametric image reconstruction algorithm. *J. Am. Statist. Assoc.*, (202), 1985.

[99] A. J. Reader. The promise of new PET image reconstruction. *Phys Med*, 24(2):49–56, Jun 2008.

[100] P. J. Markiewicz, M. Tamal, P. J. Julyan, D. L. Hastings, and A. J. Reader. High accuracy multiple scatter modelling for 3d whole body pet. *Phys Med Biol*, 52(3):829–847, Feb 2007.

[101] D.H. Wolpert and W.G. Macready. No free lunch theorems for optimization. *Evolutionary Computation, IEEE Transactions on*, 1(1):67 –82, apr. 1997.

[102] M. A. Seltzer, S. A. Jahan, R. Sparks, D. B. Stout, N. Satyamurthy, M. Dahlbom, M. E. Phelps, and J. R. Barrio. Radiation dose estimates in humans for (11)C-acetate whole-body PET. *J Nucl Med*, 45(7):1233–1236, Jul 2004.

[103] The 2007 recommendations of the international commission on radiological protection. ICRP publication 103. *Ann ICRP*, 37(2-4):1–332, 2007.