

Ruprecht-Karls-Universität Heidelberg
Fakultät für Biowissenschaften
Bachelor Molekulare Biotechnologie

Inferring cell dynamics in multi-cellular systems with incomplete data

Bachelorarbeit

**Von Benjamin Moritz Frey
Aus Bad Soden am Taunus**

Juli 2021

Eigenständigkeitserklärung

Die vorliegende Bachelorarbeit wurde im BioQuant Heidelberg in der Abteilung für Modelling Infection and Immunity in der Zeit von 17.05.2021 bis 26.07.2021 angefertigt.

Gutachter der Arbeit: Dr. Frederik Graw
Modelling Infection and Immunity
BioQuant Heidelberg

Ich erkläre hiermit ehrenwörtlich, dass:

1. ich die vorliegende Bachelorarbeit selbständig unter Anleitung verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe;
2. die Übernahme wörtlicher Zitate aus der Literatur/Internet sowie die Verwendung der Gedanken anderer Autoren an den entsprechenden Stellen innerhalb der Arbeit gekennzeichnet wurde;
3. ich meine Bachelorarbeit bei keiner anderen Prüfung vorgelegt habe.

Ich bin mir bewusst, dass eine falsche Erklärung rechtliche Folgen haben wird.

Ort, Datum

Frankfurt, den 26.07.21

Unterschrift



Acknowledgments

First and foremost, I would like to thank both Dr. Frederik Graw and Nils Bundgaard for the support and helpfulness during the Bachelor thesis. Whenever I had questions or problems, I was welcomed with great support, educational answers and solutions. Despite the difficulties presented by the current pandemic, I felt welcome in the group.

Without Dr. Frederik Graw, I would not have had the opportunity to perform my thesis in such an interesting field. He introduced me to the fascinating field of systems biology, which I am grateful for. In addition to that, he always helped defining new steps and was open to new ideas. I truly admire his ability to motivate, even when progress slowed down at times. What I also appreciate is his effort to provide good mentoring even though the thesis had to be written remotely.

Another person deserving special thanks is Nils Bundgaard. He was always available and motivated to answer my questions, no matter how trivial they were. It was especially exciting to see, how he always took time to find solutions together and not just provide an “easy” solution. He was extremely knowledgeable and regularly proposed new ideas and steps to advance the project. Thanks to him, a lot of frustration was prevented, and the time could be spent on more productive tasks. He also helped me a lot finding my way into the project in a short amount of time and provided me with material that tremendously simplified the work on the project.

I would also like to thank the IT support of the bwCluster and bqCluster for helping with technical problems without whom the generation of data would have been impossible. They always seemed motivated to help and find solutions and provided good insights that lead to the solutions of even the most difficult problems.

Finally, I would like to thank my family and all the people in my life who supported me during the thesis and helped to motivate me.

Abstract

Parameter inference is an important part in generating multicellular models. Parametrization of such models can be performed with the help of Approximate Bayesian Computation algorithms. An established workflow involves the package FitMultiCell which combines Cellular Potts Models with an Approximate Bayesian Computation approach. For the analysis and parametrization of models describing cell motility behaviour, live-cell microscopy data is often used. Analysis of corresponding tracking data revealed that not all cell tracks are able to be tracked over the entire tracking period. It has been hypothesized that this data loss negatively influences the parameter inference within the FitMultiCell pipeline. In order to correct for this occurrence, this thesis presents a method to synthetically eradicate the problem of cell loss and to ensure a corrected parameter inference. This method is referred to as *subsampling*. The validation as core part of the thesis was performed step-by-step. Initially a proof of concept was generated with increasingly complex synthetic models. In addition to that, further factors such as the subsampling depth were examined and have shown to positively influence the parameter inference. Finally, the relationship between subsampling and the contact behaviour of cells was analysed. The obtained results imply a direct proportionality between cell dropout and cell contact dropout and present an extension to the existing FitMultiCell analysis.

Zusammenfassung

Parameterinferenz ist ein wesentlicher Bestandteil der Modellierung multizellulärer Systeme. Die Parametrisierung entsprechender Modelle kann unter anderem mithilfe der Approximate Bayesian Computation durchgeführt werden. Ein dazu maßgeschneidertes Paket ist das FitMultiCell Paket. Im Falle der Analyse und Parametrisierung von Zellmotilitätsmodellen werden häufig Daten aus der Echtzeit-Mikroskopie verwendet. Bei der Analyse einiger dieser Daten wurde beobachtet, dass nicht alle Zellen über den gesamten Zeitraum verfolgt werden können. Um den Datenverlust zu korrigieren, wird in dieser Arbeit eine Methode vorgestellt, die synthetisch den Verlust beachtet und eine korrigierte Parameterinferenz gewährleistet. Die Methode wurde als *subsampling* betitelt. Die Validierung dieser Methode wird der Kernteil dieser Arbeit sein und erfolgte schrittweise. Zunächst wurde über die Implementierung verschiedener vollständig synthetischer Modelle mit steigender Komplexität ein „proof of concept“ erzeugt. Im Anschluss wurden weitere Bestandteile der Methode wie die subsampling Tiefe untersucht. Hierbei konnte gezeigt werden, dass eine Verbesserung der Parameterinferenz durch höhere subsampling Tiefe erreicht werden kann. Weiterhin wurde der Zusammenhang zwischen subsampling und dem Kontaktverhalten simulierter Zellen untersucht. Die hierbei generierten Ergebnisse implizieren eine direkte Proportionalität zwischen dem Zellverlust und dem Zellkontaktverlust.

Contents

Eigenständigkeitserklärung	I
Acknowledgments	II
Abstract	III
Zusammenfassung	IV
List of Figures	VII
List of Tables	VIII
List of Supplementary Figures	IX
List of Supplementary Tables	X
Abbreviations	XI
Glossary	XII
1 Introduction	1
1.1 Cell motility and interaction	1
1.2 Cell dropout.....	3
1.3 Cellular Potts modelling.....	4
1.4 FitMultiCell	5
1.5 Aims	7
2 Methods	9
2.1 Software	9
2.2 Cell motility simulation and analysis.....	10
2.2.1 Cellular Potts modelling in Morpheus	10
2.2.2 Quantifying cell motility	11
2.3 Parameter fitting with Approximate Bayesian Computation	13

2.3.1 pyABC	13
2.3.2 Models and FitMultiCell settings	15
2.4 Subsampling and fit design	17
2.5 Contact analysis.....	19
2.6 Data analysis and comparison	20
3 Results	21
3.1 Cell dropout disrupts ground truth summary statistics.....	21
3.1.1 Motility statistics from incomplete cell tracks show deviations	21
3.2 Increasingly complex models allow validation.....	23
3.2.1 Distance composition varies when ignoring cell loss	28
3.2.2 Performance analysis reveals discrepancies in fitting scenarios	29
3.2.3 Examination of motility data shows deviances between Runs....	30
3.2.4 Credibility intervals imply effectiveness of subsampling	32
3.3 Multiple factors determine effect of subsampling depth	36
3.3.1 Altered fitting settings reveal effect of subsampling depth	38
3.4 Subsampled contact ratio matches cell dropout ratio.....	41
4 Discussion	46
4.1 Summary statistics show deviation when ignoring cell loss	47
4.2 Subsampling positively influences parameter inference.....	49
4.3 Effect of subsampling depth depends on particle number and model design	51
4.4 Contact analysis implies relationship between dropout and subsampling	54
4.5 Conclusion and Outlook	55
References	57
Supplementary	61

List of Figures

1	Schematic depiction of the INSPECT-3D workflow	2
2	Cellular Potts Model	5
3	FitMultiCell workflow and Schematic overview of the pABC algorithm	7
4	Visualization of the subsampling method.....	8
5	Comparison of complete and incomplete ground truth data	22
6	Dropout of Mock condition in dependency of the time	22
7	Comparison of different subsampling Run specifications	27
8	Distributions of fractional distances of each motility parameter	29
9	Credibility intervals for all three runs for Model 2.....	34
10	Credibility intervals of Model 3, Run 2 and 3	35
11	Aggregated distances in dependency of the subsampling depth.....	37
12	Comparison of the motility parameters for the subsampling depths, Model 3.a..	39
13	Comparison of the total contact count for different subsampling depths	43
14	Comparison of total contact count ratios for different subsampling strategies	44
15	Comparison of the cumulative count ratios with the dropout curves.....	45

List of Tables

1	Software	9
2	Values for the parameters decay time, persistence strength and haptotaxis.....	17
3	Prior distributions of the fitted parameters	17
4	Settings of the different fitting runs within the pyABC pipeline.....	19
5	Analysed subsampling depths in Model 3 and Model 3.a	19
6	Dropout type and rates used for the contact analysis	20
7	Ranges of acceptance rates for different models	30
8	Distances of the best particle from the last generation for each depth	41

List of Supplementary Figures

S1	Screenshot of the utilized R Shiny application	61
S2	Posterior parameter distributions of Model 3, Run 1	61
S3	Summary statistics of the fitting runs for Model 1.....	62
S4	Summary statistics of the fitting runs for Model 2.....	63
S5	Summary statistics of the fitting runs for Model 3.....	64
S6	Credibility intervals for all three runs for Model 1	65
S7	Credibility intervals for all three runs for Model 3	65
S8	Distribution of the ratio haptotaxis:PS for Model 3	66
S9	Credibility intervals different subsampling depths for Model 3.a.....	66
S10	Comparison MSD curves Model 3.a different subsampling depths	67
S11	Comparison of the ratio haptotaxis:persistence strength Model 3.a	67
S12	Comparison summary statistics for different depths for Model 3	68
S13	Credibility intervals for subsampling depths for Model 3	69
S14	Comparison of different dropouts in an exponential and linear scenario	70

List of Supplementary Tables

S1	Pseudo code for the sequential Monte Carlo ABC algorithm	71
S2	ε -lists used for the different models	71

Abbreviations

2D	2-dimensional
3D	3-dimensional
ABC	Approximate Bayesian Computation
CCD	Charge-coupled device
CD-4	Cluster of differentiation 4
CI	Credibility interval
CPM	Cellular Potts Model
DT	Decay-time
EGFP	Enhanced green fluorescent protein
ETI	Equal-tailed interval
GT	Ground truth (data)
HIV	Human immunodeficiency virus
HPDI	Highest posterior density interval
MCS	Monte Carlo steps
MSD	Mean squared displacement [μm^2]
n	Subsampling depth
ODE	Ordinary differential equations
PS	Persistence strength
SARS-CoV	Severe acute respiratory syndrome coronavirus
sd	Standard deviation
SMC	Sequential Monte Carlo
T cell	T-lymphocyte

Glossary

Complete	Set of ground truth data based on tracking data containing all tracked cells at all time points.
(Cell) Dropout	Can be understood as a vector that includes fractions of remaining cell tracks at each time point. Monotonically decreasing.
Dropout type	Type of function that is used to describe the cell dropout.
Dropout rate	Fraction of cell tracks left after a period of time. Last entry of the corresponding dropout (vector).
(FitMultiCell) Fit	Operation requiring a set of ground truth data and parameters which undergoes the FitMultiCell pipeline and outputs a database with parameter sets.
Ground truth data	Set of summary statistics that is used for a SMC ABC fit with FitMultiCell (here: Motility statistics of cell tracks).
Incomplete/ Subsampled	Set of ground truth data based on tracking data containing different tracked cell counts at different time points. Monotonically decreasing.
Run	Specific type of fit requiring a pre-defined CPM, set of ground truth data (subsampled or complete) and analysis settings (subsampling or standard).
Subsampling	Method, that artificially samples out cell tracks after a simulation. Follows a certain dropout rate and type.
Subsampling depth	Number of iterations, subsampling is applied on a set of cell tracks. Selected via the minimal aggregated distance between summary statistics.
Summary statistics	Set of data that summarizes and describes sample data (here: Motility statistics of cell tracks).

1 Introduction

1.1 Cell motility and interaction

Cells are found in tissues of any living organism on planet earth. Cell theory was first described in 1839 by Theodor Schwann (Schwann, 1838) and has since been steadily improved.

The motility behaviour of cells poses an especially interesting subcategory. As explained by Jouanneau and Thiery (2002), cell motility can be understood as the capability of cells to translocate on solid surfaces. This translocation is induced by the cytoskeleton, more specifically by F-actin. Cells advance via the formation of so-called pseudopods that are induced by actin polymerisation. Direction of the pseudopod protrusion defines the direction of the cell (Jouanneau and Thiery, 2002). This type of protrusion resembles the motility characteristics found in amoeba as observed in T cells (amongst others) (Dupré, 2015). Research of cell migration and motility can give insight into a variety of biological processes. Especially the influence on pathogenesis has been extensively studied. Commonly examined aspects are molecular changes and mutations of corresponding genes (Franz et al., 2002). As an example, Tumour metastasis is enabled by cell motility and favoured by factors such as CD155 (Sloan et al., 2004) or claudins (Webb et al., 2013).

However, molecular changes are not the only factors adjuvant to the development of disease when studying cell motility. The influence of cell motility on the infection behaviour of viral diseases represents another field of interest. As seen most prominently in the past 1.5 years, virus outbreaks have the potential to disrupt entire continents and cause severe damage to humans and the economy. As of July 2021, more than 180 million people have been infected with SARS-CoV-2 causing almost 4 million deaths (WHO, 2021). On a tissue level, such virus spread occurs via virus-transport, multiplication within the infected cell, their ability to escape the immune system and cell-to-cell interaction. Their detection and spread is dependent on the

motility of local cells and T cells migrating to the site of infection (Bocharov et al., 2016; Jackson et al., 2021). Another infectious disease that has been proven to be transmitted via cell-to-cell contact is HIV-1 (Agosto et al., 2015).

Imle et al. were able to demonstrate that cell motility and underlying collagen networks can influence the infection rate of HIV-1 on CD4 T cells (Imle et al., 2019). A workflow to study the infection behaviour was implemented in the form of INSPECT-3D. In order to examine infection rates, 3D *ex vivo* cell cultures were produced that simulate the infection dynamics of HIV-1 within tissue-like conditions. The cell culture consisted of primary human CD4 T lymphocytes. The T cells were observed via live-cell microscopy and analysed quantitatively to parametrize a Cellular Potts Model (Figure 1).

The motility behaviour of CD4 T cells is of special interest because of its ability to influence the spread of disease within local neighbourhoods. T cells possess a wide spectrum of motility strategies (Krummel et al., 2016). Those can be categorized into subdiffusion (Brownian type), superdiffusion (Lévy Type) and ballistic (highly directional) as reviewed by Krummel et al., 2016. The cell's motility behaviours depend on multiple factors such as their microenvironment, activation status or local chemokine concentrations (Dupré, 2015; Krummel et al., 2016). T cells perform a stop-and-go movement during which the cells scan and search their environment (Krummel et al., 2016). It has been shown that migration and motility of T cells is partly governed by their motion induced cell shape change and contact guidance with prevalent collagen fibres (Wolf et al., 2003). Additionally, Imle et al. (2019) were able to provide evidence that collagen networks influence the infection behaviour of HIV-1 in 3D cell cultures with present CD4 cells.

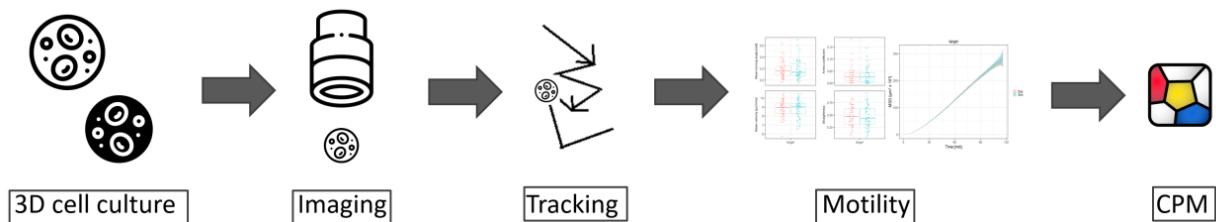


Figure 1 Schematic depiction of the INSPECT-3D workflow. The 3D *ex vivo* cultures are imaged via live-cell microscopy and tracked. Tracking data leads to motility data which is used to create a Cellular Potts Model (CPM).

In addition to cell motility, cell-to-cell interaction behaviour is another point of interest when observing pathogenesis. HIV infections become more efficient when directly contacting uninfected cells via so-called virological synapses (Agosto et al., 2015). But analysis of cell-to-cell interactions is not only important for the study of pathogenic processes but also a multitude of other biological pathways. As reviewed by Armingol et al. (2021), direct interactions are responsible for a plethora of biological processes such as homeostasis, cell differentiation, immune responses and many more (Armingol et al., 2021). Consequently, analysis of cell-to-cell interactions presents an important part when studying models of multicellular systems. As seen on the example of INSPECT-3D, such models are generated with the help of live-tracked cells requiring live-cell microscopy.

1.2 Cell dropout

Live-cell fluorescence imaging is based on the illumination of fluorescently labelled proteins in the target cell. Commonly used proteins are EGFP, mCherry or Luciferase. Fluorescence is generated when the target protein is hit with light at its specific excitation wavelength. Photons are absorbed and emitted at a different wavelength, which is then captured by a charge-coupled device (CCD) camera (Ettinger and Wittmann, 2014). Given the intricacy of live-cell fluorescence microscopy, tracking multicellular systems can present challenges. Based on the maintenance of the cellular environment, the choice of fluorescence marker and other potential disruptors, generated tracking data might be erroneous (Beltman et al., 2009; Jensen, 2013).

Recently it has been shown, that cell track loss can be observed when tracking multiple cells over a period of time (Harmel, 2020). This track loss partially resulted from the inability of the microscope to reliably track cells that left the 2D focussing plane (Harmel, 2020). The obtained tracking data was generated with CD4+ T cells and showed different cell counts at different time points (Harmel, 2020). As shown on the example of INSPECT-3D, data from live-cell microscopy can be applied to parametrize mathematical models such as Cellular Potts Models (Imle et al., 2019). A reliable quantification of cell dynamics and motility plays a crucial role in understanding and

predicting the progress of infectious pathogenic processes. In order to achieve this, mathematical models have been shown reliable (Daun and Clermont, 2007).

It will be explained later, how observed cell dropout in the experimental data can lead to disrupted solutions of the motility as reviewed by Beltman et al. (2009). A method to correct for this dropout will be the core part of this thesis.

1.3 Cellular Potts modelling

As stated, quantification via mathematical models can help to understand complex cellular systems on a molecular level with the help of computational power (Fischer, 2008). Recent progressions in computer science and hardware improvements have allowed for the creation of increasingly complex models with dedicated software such as Morpheus (Starruß et al., 2014) or PhysiCell (Ghaffarizadeh et al., 2018). These models can, given a user input, simulate a plethora of complex biological signal ways and systems. In the context of cell simulations, a well-established method presents the Cellular Potts Model (CPM). This model is based on the Potts model, a statistical model describing complex systems while considering nearest neighbour interactions (Ashkin and Teller, 1943; Beaudin, 2007). As described by Graner and Glazier in 1992, the CPM simulates cell behaviour based on the lattice class. Cells are represented by lattice sites that share the same ID. Unoccupied sites are referred to as medium (Graner and Glazier, 1992). The cell behaviour in the CPM is determined by an energy function H (also called the *Hamiltonian*). This energy is obtained by the sum of interfacial energies and energies that emerge because of the cell's deviation from its resting position (Hirashima et al., 2017). Cell dynamics emerge based on stochasticity caused by the minimization of the free energy H (Hirashima et al., 2017; Ouchi et al., 2003). In Figure 2, a simplification of the cell protrusion in a CPM can be seen. The concept of CPM has been proven effective in a multitude of application scenarios such as cyst formation (Cerruti et al., 2013), cell sorting (Graner and Glazier, 1992) or tumour biology (Szabó and Merks, 2013). Ever since it was proposed by Graner and Glazier, the CPM has been steadily upgraded. Additional parameters such as chemotaxis, leading the protrusion of cells along a chemical gradient, have been implemented (Savill and Hogeweg, 1997). CPMs have been applied in the context of

immunology for the study of wound healing (Scianna, 2015) and recently for the analysis of the infection behaviour of HIV-1 in CD4 cells (Imle et al., 2019). The development of cellular Potts models has been simplified and made accessible for broader audiences. In particular the software Morpheus (Starruß et al., 2014) presents a straight-forward way of generating CPMs with multiple parameter options to choose from.

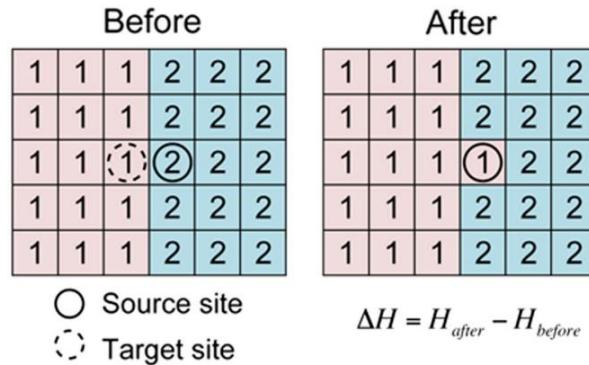


Figure 2 Cellular Potts Model. Cell protrusion is induced by change in free energy H . Protrusion occurs one step at a time. Taken and adapted from Hirashima et al. (2017).

1.4 FitMultiCell

As explained on the example of CPMs, parametrization of simulations of biological systems is an important part in generating a model that best fits a given set of data points (Gábor and Banga, 2015). Especially for CPMs, parametrization poses a difficult task caused by the stochastic nature of the model as well as potentially large parameter spaces (Jagiella et al., 2017). An established method to achieve parametrization in CPMs is the concept of Approximate Bayesian Computation (ABC). The advantage of ABC over conventional Bayesian computation methods is the redundancy of an explicit likelihood function (Boelts, 2018). In ABC, rather than calculating likelihood functions, simulated summary statistics are generated and compared to a prior set of summary statistics via a pre-defined metric (Beaumont et al., 2002). An example for a set of summary statistics is found with the motility data from a certain cell population. Within the ABC algorithm, each set of summary statistics is generated with the use of a parameter set (Beaumont, 2010). If a set threshold ϵ is undercut, the summary statistics along with its set of parameters will be accepted as a *particle*. Particles that do not reach ϵ will be rejected, hence it is also called rejection-based ABC. When a set

number of particles is reached, a new *generation* starts. The algorithm continues until convergence occurs and is terminated if specified termination criteria are reached (Beaumont, 2010). As an improvement to the established ABC algorithm, the sequential Monte Carlo (SMC) ABC was proposed in 2007 by Sisson et al. In this approach, successive reduction of the tolerance value ϵ via fitting a density kernel as well as weighted resampling of the already drawn points were implemented (Sisson et al., 2007; Beaumont et al., 2009). A more precise depiction of the algorithm will be presented in chapter 2.3 (a pseudo-code can be found in Table S1). The applications of ABC algorithms in biological applications have been used on a plethora of studies covering multiple biological fields such as pathogen spread or evolution in humans (Csilléry et al., 2010; Neuenschwander et al., 2008; Shriner et al., 2006).

Simulation of a CPM and subsequent parameter inference using an SMC ABC algorithm was proposed by Starruß et al. in the form of the package FitMultiCell (Starruß et al., 2019). In detail, it combines the simulation of a CPM in Morpheus with the subsequent analysis using the package pyABC (Klinger et al., 2018). pyABC presents the implementation of a highly parallelized pABC SMC algorithm on computing clusters (Jagiella et al., 2017). The pABC SMC algorithm distributes computational tasks between a master and slaves. The master initializes parameters and stores the results, the slaves compute the simulations with a given parameter set and calculate the distances of the summary statistics which are then stored (Jagiella et al., 2017, Figure 3; right).

The implementation of CPM models that mimic observed cell behaviour from live-cell microscopy data uses a constant cell number at each simulated time step (Starruß et al., 2014). Analysis of the simulated data within the FitMultiCell framework calculates summary motility statistics based on those simulations with the complete cell track data. As explained, SMC ABC algorithms compare the calculated summary statistics to a prior set of summary statistics (Figure 3; left). So far it is not clear how well these methods perform when presented with incomplete prior data. Especially the influence of incomplete prior data on the parameter inference of cell motility dynamics is yet to be investigated.

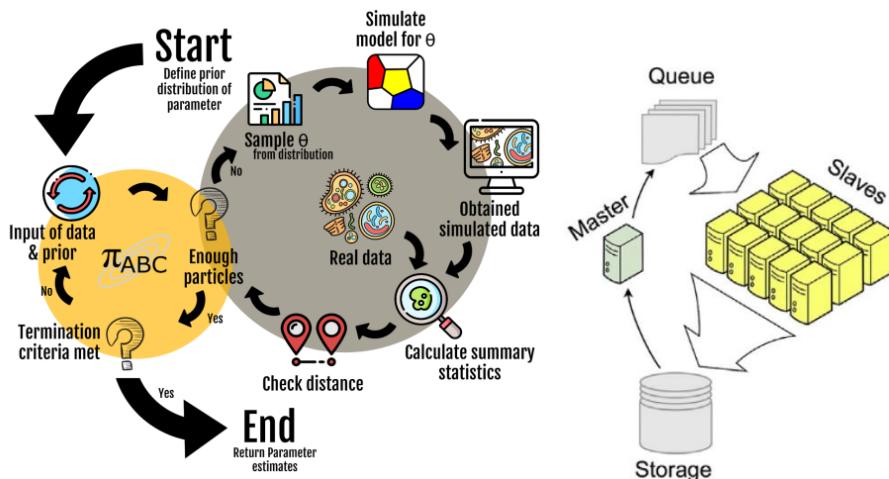


Figure 3 Left: FitMultiCell workflow, taken from fitmulticell.gitlab.io, **Right:** Schematic overview of the pABC algorithm parallelized for high performance clusters. Taken and adapted from Jagiella et al. (2017).

1.5 Aims

The intention of this thesis is to infer cell dynamic parameters in multi-cellular systems when presented with incomplete prior data. In the context of the FitMulticell framework, a comparison of incomplete prior data to complete simulated data is hypothesized to pose a problem on the parametrization. In the work from Harmel (2020) a method that addressed and corrected this problem was defined, the subsampling (Figure 4).

Since subsampling is not established within the FitMultiCell framework, its validation will be the core part of this thesis. The proposed working plan is to generate synthetic data for different models in order to analyse the potential effects of the subsampling. The models will be designed to simulate the behaviour of one or two cell types under different biological conditions. Analysis of the cell's motility under consideration of a cell dropout will be performed as part of the validation. Finally, the effects of different subsampling depths on the parameter inference as described by Harmel (2020) will be studied as well as the influence of subsampling on the contact behaviour of different cell populations. Subsampling will be intended to be used as an extension of the existing FitMultiCell framework when confronted with incomplete prior data.

1 Introduction

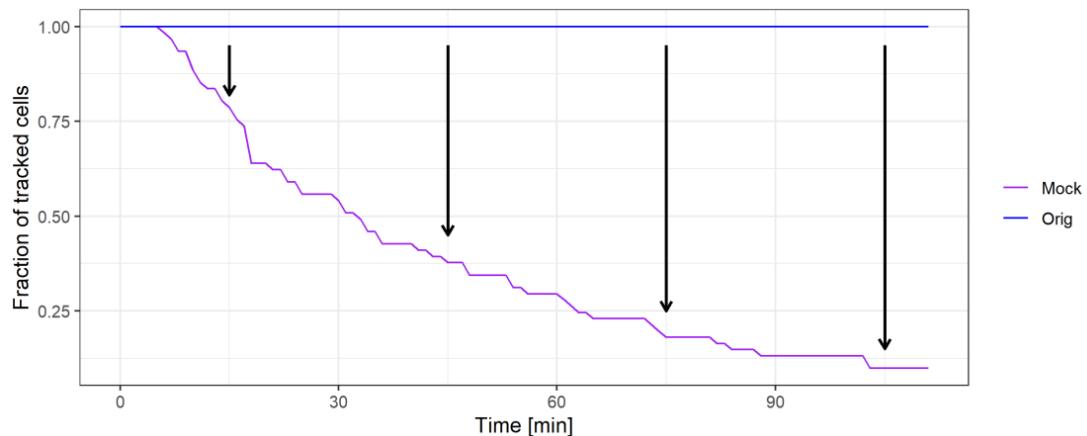


Figure 4 Visualization of the subsampling method. At each sampling step, a fraction of the cells is artificially sorted out (indicated by black arrows). The purple line corresponds to the dropout, that can be observed after subsampling was applied (in this case adapted from Harmel (2020), hence labelled with Mock). The blue line indicates the cell tracks without dropout (here: original tracks, will be referred to as complete later on).

2 Methods

2.1 Software

In order to generate data and perform the simulations, a multitude of software was necessary. The main tools and packages are listed in Table 1. In addition to the calculations on a local machine, the high-performance cluster of the BioQuant in Heidelberg as well as the bwForCluster MLS&WISO were utilized. The FitMultiCell pipeline and pyABC analysis were executed on the clusters. The statistical analysis was performed locally in R with the help of different packages. The most important packages were MotilityLab, dplyr, bayestestR and ggplot.

Table 1 Software along with the versions. Version differences between the clusters are mentioned in brackets. The versions in the brackets refer to the MLS&WISO cluster.

Software	Version
RStudio	1.4.1106
R	4.0.5 (4.1.0)
bayestestR	0.9.0
Dplyr	2.1.0
Tidyverse	1.3.0
MotilityLab	0.2-4
Python	3.7.6 (3.9.5)
pyABC	0.10.14 (0.10.15)
FitMultiCell	0.0.3
Morpheus	2.2.0

2.2 Cell motility simulation and analysis

2.2.1 Cellular Potts modelling in Morpheus

In order to analyse and fit the simulated data, this data must be produced first. This was achieved with the help of Morpheus, a modelling environment for the creation of multi-cellular systems (Starruß et al., 2014). All simulated models were Cellular Potts Models as described in section 1.3. For the model design, a step-by-step approach was chosen.

The following settings applied to all models (in some models, an additional cell type was simulated additionally). A space with the dimensions of $250 \times 250 \times 250$ voxels was created, with 1 voxel corresponding to $1 \mu\text{m}$ in real space. A total of 64 cells per cell type were simulated and tracked over a time period of 7 200 MCS with recording steps of 60 MCS, resulting in 121 individual time-dependant x- and y- and z-coordinates per cell type. For the analysis, the track data was outputted in a Morpheus logger file, where 1 MCS corresponded to 1 second. The initial 240 MCS were considered within the analysis to account for the adjustment of cell size and motion by the CPM. The surface asperity was set to 0.7 and the volume to $180 \mu\text{m}^3$ with a strength of 1 for both parameters. The CPM neighbourhood was set to 2.

The CPM further required different parameters in order to calculate the Hamiltonian (free energy) H (section 1.3 for reference). As explained, cell protrusion in a CPM occurs when the free energy H is minimized. H is given as

$$H = E_{surface} + E_{volume} + E_{PS} + E_{Hapto} \quad (1)$$

(Hirashima et al., 2017). The parameter definitions can be found in the documentation window in Morpheus (Starruß et al., 2014). The surface constraint $E_{surface}$ penalizes deviations from the targeted surface area where λ_S is the strength of constraint, $s_{\sigma,t}$ the surface area at time t and s_{target} the targeted surface area of cell σ .

2 Methods

$$E_{surface} = \sum_{\sigma} \lambda_S (s_{\sigma,t} - s_{target})^2 \quad (2)$$

The volume constraint E_{volume} analogly penalizes deviations from the targeted volume V_{target} with λ_V being a Langrangian multiplier and $v_{\sigma,t}$ being the volume at time t of cell σ (Graner and Glazier, 1992).

$$E_{volume} = \sum_{\sigma} \lambda_V (v_{\sigma,t} - V_{target})^2 \quad (3)$$

Another parameter that was simulated is the persistent motion E_{PS} which stands for the tendency of the cells to maintain their direction of movement. s_{PS} stands for the persistence strength and α is the angle between two consecutive movements. An additional requirement of the persistent motion not found in the equation is the decay time DT . It represents the memory of direction of the cell movement.

$$E_{PS} = -s_{PS} \cdot \cos(\alpha) \quad (4)$$

A factor that favours the updates of the cells in the direction of a certain attractant is the haptotaxis E_{Hapto} . An underlying collagen network represents this attractant in the models. Its strength is determined by the strength of the haptotaxis (Starruß et al., 2014). Protrusion of a cell or a copy attempt is undertaken when ΔH falls below a certain threshold. Here T is the temperature. This threshold is calculated via the Boltzmann probability (Chen et al., 2007):

$$P(\Delta H) = \begin{cases} 1, & \text{if } \Delta H \leq 0 \\ e^{-\Delta H/T}, & \text{else} \end{cases} \quad (5)$$

2.2.2 Quantifying cell motility

After a simulation is completed, a tracking dataset is generated by Morpheus in which time-dependant x- and y- and z-coordinates of each cells' centre-of-mass are logged. The cell tracks can be represented as time-dependant vectors $\vec{p}_c(t) = (x(t), y(t), z(t))$ where c stands for the individual cell and t for the time (holds for all following

explanations). The motility statistics were calculated with the help of the R package *MotilityLab* (Textor et al., 2016).

Velocity

The velocity v_c is calculated by dividing the track length by the total duration and taking the average of all speed values between two points. In *MotilityLab* the function *speed()* is used.

$$v_c = \frac{1}{n-1} \sum_{i=0}^{n-1} \frac{\sqrt{(\vec{p}_c(t_{i+1}) - \vec{p}_c(t_i))^2}}{t_{i+1} - t_i} \quad (6)$$

Mean turning angle

To analyse the cell's ability to change direction, the mean turning angle is calculated. This is achieved by averaging the overall angle over all adjacent segments ϕ of a cell track. The resulting mean turning angle is α_c . The corresponding function is *meanTurningAngle()*.

$$\alpha_c = \frac{1}{n-1} \sum_{i=0}^{n-1} \phi((\vec{p}_c(t_i), (\vec{p}_c(t_{i+1}))) \quad (7)$$

Arrest coefficient

Moving forward, turning and moving backwards are not the only possible motility patterns cells can undergo. It is also possible for the cells to stay in one position or oscillate/ wiggle for a period of time. The arrest coefficient a_c characterizes whether a cell is in motion or not. It is defined as a time-fraction in which the cells velocity is slower than a given threshold. The used velocity threshold v_c was set to 2 $\mu\text{m}/\text{min}$. This threshold is commonly used for simulated T-cell motility (Beltman et al., 2009). The total number of cells tracks is represented by n .

$$a_c = \frac{|\{v_c < 2\}|}{n} \quad (8)$$

Straightness

The straightness s_c explains to what extent the cells move in a straight line. It is calculated by dividing the individual displacement by the total track length. The straightness is thus a number $s_c \in [0,1]$. It is calculated via the *straightness()* function.

$$s_c = \frac{\sqrt{(\vec{p}_c(t_n) - (\vec{p}_c(t_0))^2}}{\sum_{i=0}^{n-1} \sqrt{(\vec{p}_c(t_{i+1}) - (\vec{p}_c(t_i))^2}} \quad (9)$$

Mean square displacement

In addition to the four motility statistics and to analyse the type of motion the cells undergo; the mean square displacement MSD was calculated. It calculates iteratively the squared displacement of the cells over strictly monotonically increasing time steps τ . The function *squareDisplacement()* is used.

$$s(\tau) = \frac{1}{n-\tau} \sum_{i=0}^{n-\tau} ((\vec{p}_c(t_{i+\tau}) - (\vec{p}_c(t_i))^2 \quad (10)$$

The mean squared displacement is obtained by averaging the square displacement over all time steps and cells c .

$$MSD = \frac{1}{c} \sum_{i=1}^c \frac{1}{n-1} \sum_{i=2}^{n-1} s_c(\tau_i) \quad (11)$$

2.3 Parameter fitting with Approximate Bayesian Computation

2.3.1 pyABC

Analysis and parametrization were performed with help of FitMultiCell, a pipeline which orchestrates the multi-cellular system modelling in Morpheus and the analysis with

pyABC (Figure 3). As explained, pyABC is an approximate Bayesian Computation algorithm using a sequential Monte Carlo approach (ABC-SMC) (Klinger et al., 2018).

The principle of a rejection-based ABC algorithm is to fit a set of model parameters θ_s using an underlying prior set of parameters θ and data D (also called ground truth data or GT). The parameters θ_s are sampled randomly from a uniform distribution with heuristic boundaries (Bundgaard, 2020). For each step, a dataset T is simulated under a predefined model M and corresponding summary statistics are calculated (each simulation along with its calculated summary statistic and θ_s is called a particle). M is the utilized CPM with the previously explained motility parameters (more detailed explanation in section 2.4). The summary statistics of D and M are then used to calculate a pre-defined cost function d which is compared to a threshold ε . This threshold can either be set manually or be obtained from a previous generation. Particles which fulfil $d < \varepsilon$ are accepted and kept until a pre-defined number of particles is reached. The parameter sets θ_s of the particles are weighted by their distance d (small distance \rightarrow high weight). A new generation begins with an updated distribution that takes the weights and parameter sets into consideration. The distribution is subsequently smoothed via a kernel operation (Jagiella et al., 2017). The algorithm stops when a pre-defined minimum ε is reached.

In order to evaluate the individual particles, the cost function d had to be defined. The values for d were calculated using the summary statistics of the prior (D) and simulated data (M). The means and standard deviations of both motility datasets (summary statistics) were used to calculate d_{MSD} and d_{Mot} respectively. They are denoted by μ_g, μ_s, σ_g and σ_s where g stands for *given* and s for *simulated*. The resulting distance metric for the MSD is defined as

$$d_{MSD} = \frac{1}{n} \sum_{i=0}^n \left(\frac{\mu_g(t_i) - \mu_s(t_i)}{\sigma_g(t_i)} \right)^2 \quad (12)$$

where the squared distance of each time interval t between the given and simulated data for each cell c is calculated and divided by the standard deviation of the given data (σ_g).

d_{Mot} was calculated by dividing the squared distance of the two datasets over time and dividing by the standard deviation of the given data per cell per motility parameter. To account for a difference in dispersion between the data, the coefficient of variation was added.

$$d_{Mot} = \frac{(\mu_{g,c} - \mu_{s,c})^2}{\sigma_{g,c}^2} + \left(\frac{\sigma_{g,c}}{\mu_{g,c}} - \frac{\sigma_{s,c}}{\mu_{s,c}} \right)^2 \quad (13)$$

The total distance was calculated by adding up the individual distances and weighing the distance for the MSD. This was done in order to correct for the different dimensionality of the individual motility parameters.

$$d_{Total} = \sum_{i \in Mot} d_i + \sum_{j \in MSD} 0.01 \cdot d_j \quad (14)$$

A pseudo-code of the SMC-ABC algorithm can be found in Table S1. In an additional step, the influence of the individual distances on the total distance was examined. This was achieved by calculating the distances per particle and dividing them by d_{Total} .

2.3.2 Models and FitMultiCell settings

In order to validate the subsampling method, different multi-cellular models with increasing complexity were created. Different parameters were set in the models and used for the parameter inference. The parameters were set based on a desired range of motility parameters.

- **Model 1** included one cell type and two parameters: Persistence strength and decay time.

- **Model 2** included one cell type and three parameters: Persistence strength, decay time and haptotaxis.
- **Model 3** included two cell types and three parameters (per cell type): Persistence strength, decay time and haptotaxis. The second cell type was defined to possess different motility properties compared to the first cell type with the intention to mimic a faster cell type.
- **Model 3.a** was an updated version of Model 3 that was used to investigate the subsampling depth. The difference to Model 3 was that the values for the persistence strength were sampled from a normal distribution. This resulted in a wider range of motility characteristics.

As briefly explained in section 2.2.1, the parameters persistence strength and decay time play a role in the persistent motion. The persistence strength indicates how strong the persistence motion of a given cell is while the decay time quantifies the memory of direction. The haptotaxis on the other hand is a measurement of directional movement in response to substrate bound molecules. It was introduced to set the intensity of the underlying collagen network. All parameters influenced the motility behaviour of the cells. The descriptions of the parameters were defined in Morpheus (Starruß et al., 2014). A detailed motivation of the different model designs is presented in section 3.2. A summary of the set parameters can be found in Table 2. The prior distributions of the parameters fitted with the FMC pipeline are shown in Table 3.

For the fitting with pyABC, an ε -list with a minimum ε of 1.5 (3 for the two-cell model) was generated by averaging the ε -values from previously executed FitMultiCell fits. The lists were implemented in order to ensure comparability between the individual fits. Since the fits were performed on different clusters with different parameter sets and ground truth data, it was necessary to be able to compare the results of the fits to ultimately be able to evaluate potential effects of the subsampling. The minimum acceptance rate was set to 0.01. The ε -lists can be found in Table S2. The particles per generation were set to 1 000 (50 for the analysis of the subsampling depth). The prior parameters were sampled from a uniform (PS, haptotaxis) and a discrete (DT) distribution (Table 3).

2 Methods

Table 2 Values for the parameters **decay time**, **persistence strength** and **haptotaxis** are set using a range of desired output parameters. In squared brackets are the values for a second cell type, if implemented.

Parameter	Model 1	Model 2	Model 3	Model 3.a
Decay time	15	15	15 [12]	15 [12]
Persistence strength	35	43	43 [57]	rand_norm(43 [57], 20)
Haptotaxis	/	-1 000	-1 000 [-500]	-1 000 [-500]

Table 3 Prior distributions of the fitted parameters. The type of distribution as well as the boundaries are shown. If a second cell type was simulated, the same prior distributions were used.

Parameter	Prior distribution
Decay time	Discrete [0, 60]
Persistence strength	Uniform [0, 100]
Haptotaxis	Uniform [-2 500, 0]

2.4 Subsampling and fit design

In order to account for the discussed problem of cell dropout, a subsampling strategy was implemented. The incentive was to positively influence the parameter inference within the FitMultiCell pipeline when presented with incomplete tracking data. Subsampling works by randomly dropping a fraction of cell tracks x at each timepoint t of the cell track file outputted by a Morpheus simulation. The dropout follows a pre-defined strategy. For all fitted models, the strategy was adapted from the dropout of the Mock condition as seen in Harmel (2020, Figure 6). A dropout vector along with a simulated tracking dataset and a time vector are required to generate the subsampled tracking data.

The validation of the subsampling was performed via a step-by-step approach. The first step was the implementation of the models introduced in section 2.3. The next step was the definition of distinct fitting scenarios. This was necessary in order to be able to draw conclusions on whether subsampling influenced the parameter inference

2 Methods

and to fathom how it performed in comparison to a fit without subsampling. The fits were defined as follows and will be motivated in detail in section 3.2. For each implemented model, three fits were performed. The following rules applied to the fits.

- 1) The three fits were named *Run 1*, *Run 2* and *Run 3*.
- 2) The settings of each *Run* differed in two variables: the *ground truth data (GT)* and the *analysis*. The applied settings are depicted in Table 4. The variables for each run can be partitioned as follows:
 - a. For each *Run* the *ground truth* was calculated from a set of simulated cell tracks. If the tracks were subsampled before calculation of the ground truth summary statistics, it will be described as *subsampled*. If they were not subsampled, it will be depicted as *complete*.
 - b. In each *Run* the *analysis* can either include subsampling or not. If subsampling was applied, it will be depicted as *subsampling*. If not, it will be depicted as *standard*.
 - c. For the presentation of the results the term *Run 1/2/3* with the according model will be used.

The calculated ground truth summary statistics of the different *Runs* consisted of the motility statistics of the simulated cell tracks (see section 2.2.2). Each model had a distinct set of *complete* (*Run 1*) and *subsampled* (*Run 2 and 3*) ground truth data. Motility statistics and summary statistics will be used synonymously and refer to the same set of calculated statistics.

After the examination of the models, the effects of the subsampling depth were studied. A depth of n means that subsampling was applied iteratively n -times and for each iteration corresponding summary statistics (motility statistics) were calculated and compared to the prior data. The distances of the subsampled datasets were then normalized by their maximal value and aggregated. The lowest aggregated distance sample was chosen, and further evaluated within the FitMultiCell pipeline. The analysed subsampling depths are listed in Table 5. For the comparison of the models in section 3.2 all models in *Run 3* were fitted with a depth of $n = 50$.

2 Methods

Table 4 Settings of the different fitting Runs within the pyABC pipeline. The field “ground truth” stands for the prior data, the field “analysis” refers to whether subsampling was used in the analysis or not.

Setting	Run 1	Run 2	Run 3
Ground truth	Complete	Subsampled	Subsampled
Analysis	Standard	Standard	Subsampling
Depth	x	x	50

Table 5 Analysed subsampling depths in Model 3 and Model 3.a.

Model	Depth
Model 3	[1, 5, 10, 25, 50]
Model 3.a	[1, 6, 25]

2.5 Contact analysis

In addition to the analysis of the subsampling depth on the parameter inference and summary statistics, the effect of subsampling on the contact behaviour was examined. Analog to the outputted tracking files for the cell coordinates, logger files with cell-to-cell contacts were generated by Morpheus. The logger files included the time points as well as the corresponding cell ids and the contact duration of each contact. For the contact analysis, Model 3 was selected.

First, the unique interactions slow-slow, slow-fast and fast-fast were calculated for different subsampling depths and compared to complete tracking data. As a next step, the ratios of the total cell contacts were compared to the dropout observed in the subsampling. This was achieved by dividing the total cell-to-cell contacts per interaction type to the original contact number. Calculation of the dropout ratio of the cell tracks was performed as follows:

$$(I) \text{ } Ratio_{Dropout} = \frac{\int_{t=Start}^{End} f(t)dt}{\int_{t=Start}^{End} c(t)dt} \quad (16)$$

$$(II) \text{ } c(t) = n_{initial}$$

$$(II) \text{ } f(t) = \text{Dropout rate}$$

Here t stands for the time steps in minutes, $n_{initial}$ for the number of simulated cells, $f(t)$ for the dropout rate and $c(t)$ for the original cell number without dropout.

In addition to the contact ratios, the cumulative contact ratios were compared. For this, first the cumulative unique contacts were calculated for a specific subsampling depth. Those were then divided by the respective original contacts in order to generate cumulative contact ratios. For a better comparison to the cell dropout curves, the ratios cumulative (dropout/total tracks) were calculated. Different scenarios were selected (Table 6). The scenarios were partitioned into dropout type, and dropout rate. A graphical representation be found in Figure S14.

Table 6 Dropout type and rates used for the contact analysis. The dropout rates refer to the fraction of data that is lost after the simulation period of 2 h. The dropout type indicates what function was used to perform the dropout. An overview of the dropout rates and the corresponding total data losses can be found in Figure S14.

Dropout type	Dropout rates
Linear	[0.3, 0.7, 0.99]
Exponential	[0.3, 0.7, 0.99]

2.6 Data analysis and comparison

Comparisons of the data was performed with the help of the R packages *ggplot2*, *dplyr* and *shiny* (Figure S1). Credibility intervals were calculated and plotted for each parameter. Credibility intervals show, in analogy to the frequentists' confidence interval, a range of values for parameters in set intervals (Hespanhol et al., 2019). 25%, 50% and 95% credibility intervals were chosen to compare the runs. They were calculated with the R package *bayestatR* and will be mostly abbreviated with CI in the results and discussion.

3 Results

3.1 Cell dropout disrupts ground truth summary statistics

The aim of this thesis was to examine and validate the method subsampling. The main idea behind this method was to account for *incomplete* ground truth data that might impair the parameter inference within FitMultiCell fits. As explained, SMC ABC fits as found in the FitMultiCell pipeline require ground truth summary statistics in order to be able to perform parameter inference.

The validation of the subsampling method was structured into different parts. At first, the necessity of the subsampling was shown. Next, the effects of subsampling with different CPMs with increasing complexity were evaluated. Lastly, additional factors such as the subsampling depth and its effect on the contact behaviour were analysed. All results shown in sections 3.1 and 3.2 were generated using a synthetic version of the dropout observed by Harmel (2020) for the Mock condition (Figure 6).

3.1.1 Motility statistics from incomplete cell tracks show deviations

The problem that subsampling intends to correct stems from using an *incomplete* set of ground truth data in a FitMultiCell fit where the incompleteness during the analysis is not considered. *Incomplete* refers to a cell dropout in the tracking data that is used to calculate the ground truth motility statistics. This means that the cell tracks in the tracking file do not remain constant over time but decrease monotonically (Figure 6). The corresponding ground truth data is generated via calculation of the summary motility statistics of the simulated cell tracking file.

A comparison of an *incomplete* set of ground truth motility data to a *complete* set from the same tracking data revealed significant differences of all motility parameters. The position of the means as well as the width of the distributions showed significant deviations. In addition to that, the MSD showed a significant decrease for all time points (Figure 5).

3 Results

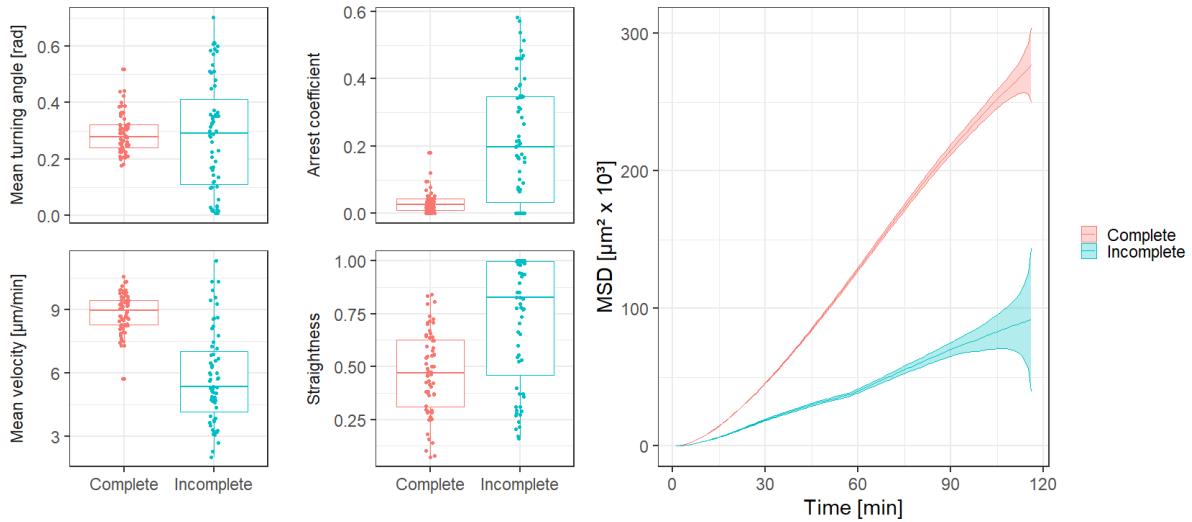


Figure 5 Comparison of complete and incomplete ground truth data. Summary statistics of the complete ground truth data (red) and incomplete ground truth data (blue) are shown. The boxplots show the motility parameters for each cell. MSD stands for mean square displacement. The ground truth data was chosen from an arbitrarily selected model with one cell type and two parameters that influence the motility behaviour.

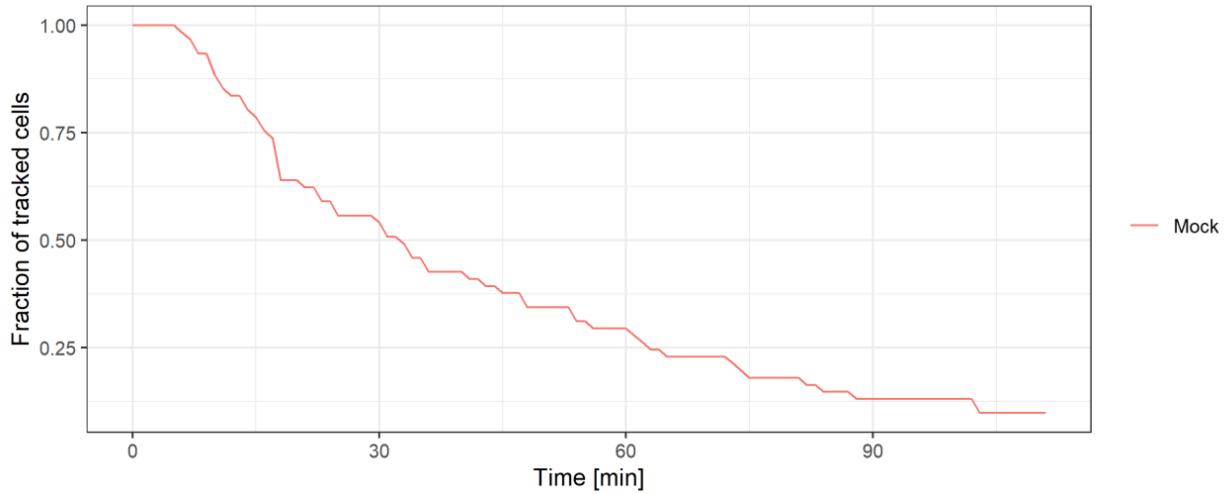


Figure 6 Dropout of Mock condition in dependency of the time. Adapted from the CD4+ T cell data generated by Prof. Dr. Oliver Fackler). This dropout was used for the analysis of the different models and subsampling depths.

As shown in Figures 5 and 6 *incomplete* cell track data led to significant differences in the motility statistics of the ground truth data when compared to the *complete* motility statistics. If used in a FitMultiCell fit, these differences were hypothesized to result in an impaired parameter inference. The next step was to figure out whether this impairment occurred and to validate a viable way to correct for the dropout, the subsampling method.

3.2 Increasingly complex models allow validation

In order to implement a method that corrected the impairment observed in the ground truth data (Figure 5) within the FitMultiCell analysis, subsampling was developed. A representation of how the method works can be found in Figure 1. At each timepoint, a pre-defined fraction of cells is dropped artificially so that it corresponds to the observed track loss used for the calculation of the ground truth data. The applied dropout can be found in Figure 6.

The validation of subsampling was implemented incrementally by applying subsampling on different Cellular Potts Models. The models were designed to be increasingly complex in terms of the parameter composition, biological conditions and scenarios. This design choice was implemented in order to fathom which parameters influencing the motility behaviour might be picked up by the subsampling and which not. With an increase in model complexity, the number of potential points of interaction and disruption for the cells would rise as well. Additional introduction of a dense collagen network created another layer of possible interactions in the system. It was hypothesized that increasingly complex models would also result in more factors missing when confronted with *incomplete* cell track data, which then had to be corrected by the subsampling. The term complexity of the models refers to the number of implemented parameters and cell types influencing the motility behaviour of the cells. Analog to the model design, the analysis of the models was performed incrementally, so first Model 1 was analysed, then Model 2 was analysed, and so on. The set parameter values are summarized in Table 2.

- **Model 1:** A model that simulated 64 cells of one cell type. The parameters persistence strength (PS) and decay time (DT) defined the motility of the cells. Motility characteristics of T cells were simulated which resulted in overall fast motility characteristics. Model 1 was implemented to evaluate how subsampling performs in a simple model with one cell type and without interactions with an environment.

Total number of parameters fitted: 2

- **Model 2:** A model that simulated 64 cells of one cell type. In addition to the same PS and DT as in Model 1, a third parameter was introduced; the haptotaxis. Along with the haptotaxis, a dense collagen network was implemented. The value of the haptotaxis indicates how strong the cells interact with the network. As explained in section 1.1, influence of collagen on the cell motility has been shown (Imle et al., 2019). The collagen network and haptotaxis were introduced to study how a collagen network influencing the cell motility could affect the parameter inference.

Total number of parameters fitted: 3

- **Model 3:** A model that simulated 64 cells for two cell types. One cell type was designed to possess overall slower motility properties, the other to have faster motility properties. The cells were accordingly labelled. For each cell type, the three previously explained parameters were defined with differing values between the cells. It was studied, how the interaction between the cell types as well as with the underlying collagen network influenced the parameter inference. When presented with a greater number of parameters that can influence the cell motility, a greater number of potential factors could be missing with incomplete cell tracks. The effect of this dropout was analysed as well.

Total number of parameters fitted: 6

- **Model 3.a:** An extension of Model 3, that differed in the values for the PS. The values for the PS were sampled from a normal distribution with a standard deviation of 20. Model 3.a was implemented solely for the study of the subsampling depth that will follow in chapter 3.3 and will thus be excluded from the analysis in this chapter. The motivation behind Model 3.a was to create cells with a wider range of motility parameters in order to increase the stochasticity of the model.

Total number of parameters fitted: 6

As described above, the model design was incremental. In addition to the model implementations, different fitting scenarios for each model were developed. This was done to be able to analyse the effects of subsampling on the parameter inference and to compare it to standard fitting settings. Each fit required the following components:

- i) A set of ground truth data and parameters.
- ii) A CPM that generates simulated data from which the summary statistics are calculated within the FitMultiCell pipeline.
- iii) Application of the subsampling strategy in the analysis or standard analysis.
- iv) General fitting settings (number of accepted particles and an ε -list).

For i), the ground truth data included either *subsampled* or *complete* motility statistics from the respective models. The prior parameter distributions of the models are shown in Table 3.

For ii), the described Models 1-3 were chosen.

For iii) *subsampling* could either be included in the analysis or not which was then referred to as *standard*.

For iv), in total 1 000 particles per generation as well as an ε -list (Table S2) were implemented. As seen in Table S2, the ε -lists differ between a one-cell (Model 1 and 2) and a two-cell (Model 3) model. The reason for this was that the number of parameters in the two-cell model was double the number of parameters in the one-cell model. This resulted in a two-fold increase in the number of distances to be considered. A correction of this was applied by doubling the ε -list.

Using the rules i)-iv), three so-called *Runs* were defined for each of the Models 1-3. The exact specifications of the *Runs* can be found in section 2.4 and an overview can be found in Figure 7.

It is of utmost importance for the following results to always consider that the described *Runs* 1-3 were performed for each Model and that each *Run* was an independent fit within FitMultiCell with a distinct set of ground truth data. Therefore, in total 9 FitMultiCell fits were performed for this part of the analysis.

The ground truth data for the fits was obtained by calculating summary motility statistics of simulated cell tracks using the respective Model. As seen in Figure 7, ground truth data could be either *complete* or *subsampled*. A *complete* set of ground truth data meant calculated summary motility statistics from a simulated set of cell tracks. A

subsampled set of ground truth data referred to the same set of simulated cell tracks, with the difference that they were *subsampled* before calculating summary motility statistics.

A categorization of the *Runs* was implemented in order to obtain conclusive results about the effectivity of the subsampling. Each *Run* had a distinct purpose:

- *Run 1* was performed in order to examine whether the applied Model and fitting settings were able to infer the set parameters in the first place and to serve as a point of reference for the following runs.
- *Run 2* was performed in order to fathom whether implementation of a *subsampled* set of ground truth data would impair the parameter inference and summary statistics when ignoring the *subsampling* in the analysis (*standard analysis*).
- *Run 3* was performed in order to evaluate whether implementation of *subsampling* in the analysis while having *subsampled* ground truth data would lead to an improved parameter inference in comparison to *Run 2* and whether the parameter inference was comparable to *Run 1*.

In summary, the design of the validation strategy for the subsampling was carefully planned in order to comply with a multitude of potential scenarios. Different biological conditions and influences on the cell motility were considered in the form of a dense collagen network and multiple cell types. These factors were hypothesized to show the effects of a cell track dropout on the motility behaviour as well as the parameter inference to different degrees. Subsequent implementation of a subsampling strategy would thus require accounting for different levels of information loss when correcting the dropout.

In order to precisely validate if subsampling truly corrected the cell dropout and improved the parameter inference, the explained fitting scenarios (*Runs*) were defined (Figure 7). For each Model, this set of fits was performed which would allow for a

comprehensive and quantitative evaluation of the effects of the subsampling method in comparison to settings where its implementation was considered.

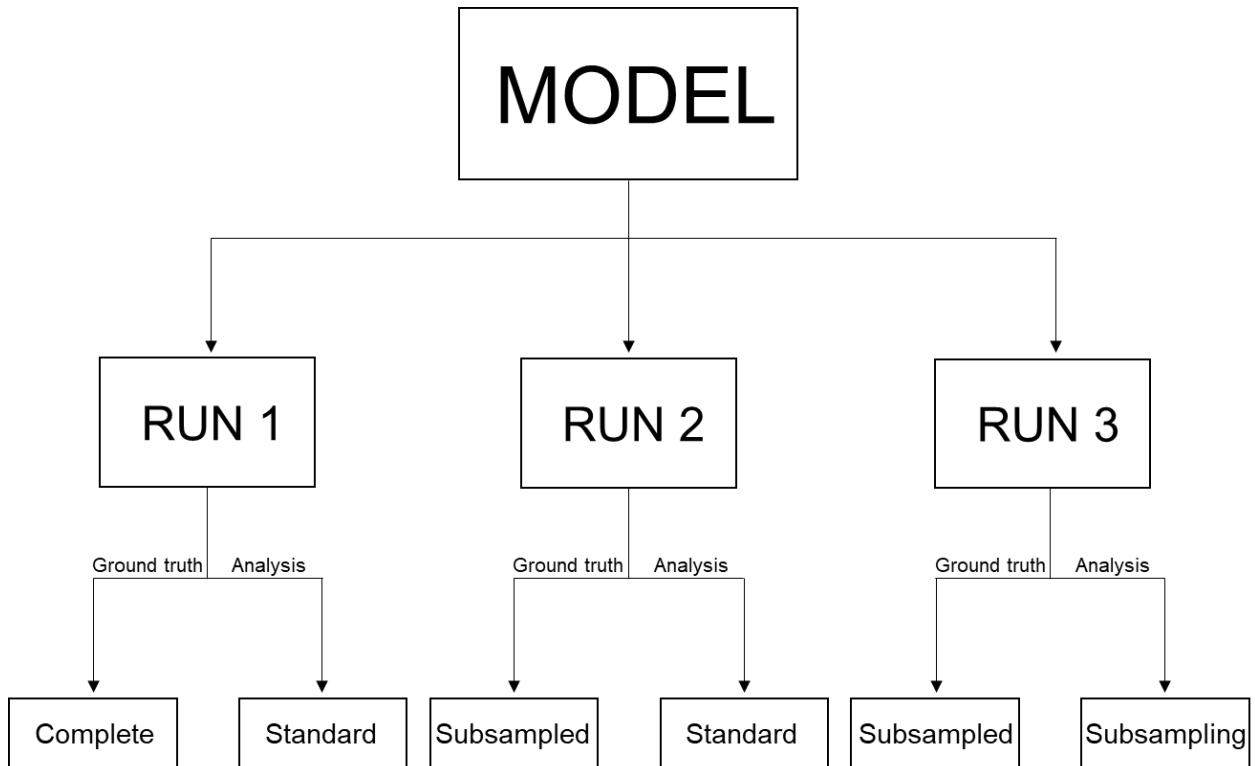


Figure 7 Comparison of different subsampling Run specifications. For each Model, three runs were applied. Each Run had a distinct set of ground truth data (complete or subsampled) and analysis settings (standard or subsampling). In total, 9 fits were performed, each of which can be categorized into Model \times Run \times : where $x \in [1,2,3]$.

The evaluation of the runs was performed by analysing different factors. First, the performance of the runs was examined by inspecting the acceptance rates of the particles for individual generations. Secondly, the summary statistics of the best particles of the final generations of each *Run* were analysed. Those were compared to the respective ground truth summary statistics. And finally, the parameter distributions and credibility intervals of the final generations of a *Run* were evaluated. Before analysing the different runs, a general inspection of the composition of the central cost function d (equation 14) was performed.

3.2.1 Distance composition varies when ignoring cell loss

A factor that plays a key role in the FitMultiCell framework is the distance metric (or cost function) d . As explained in section 2.3.1, it measures how well the motility statistics of a simulation match the prior summary statistics. It holds, that

$$d_{Total} = \sum_{i \in Mot} d_i + \sum_{j \in MSD} 0.01 \cdot d_j .$$

Analysis of the above distance composition was performed for each *Run* of the analysed models. The incentive was to evaluate whether the subsampling affected the composition of the distance metric d and if so whether the composition of the distance should be taken into consideration in the following evaluation of the fits. To compare the distance compositions, fractions of the distance of each motility parameter of the total distance $d_{Particle}$ were calculated. This was iterated over all particles of the final generation of each *Run*.

Comparison of the fractional composition of the total distances for Model 3 (representative for all models) showed no significant differences between *Run 1* and *Run 3*. In both runs, the MSD showed a higher mean compared to the other parameters for both cell types. In addition to that, all parameters except the straightness of the slow cell type *Run 3* showed a lower mean compared to the rest of the parameters. In *Run 2*, a significant increase of the mean of the straightness was observed. This increase was accompanied by a slight increase of the MSD mean as compared to the rest of the parameters in *Run 2*. Relative to the means of *Run 1* and *Run 3*, the mean of the MSD in *Run 2* stayed in a similar range (Figure 8).

In summary, the results showed no significant difference in the composition of the distances between *Run 1* and *Run 3*. A deviation of the distance composition in *Run 2* was expected. As a result, no further evaluation of the distance composition was required for the subsequent evaluations of the model fits. The evaluation of the fits started with analysis of the acceptances rates.

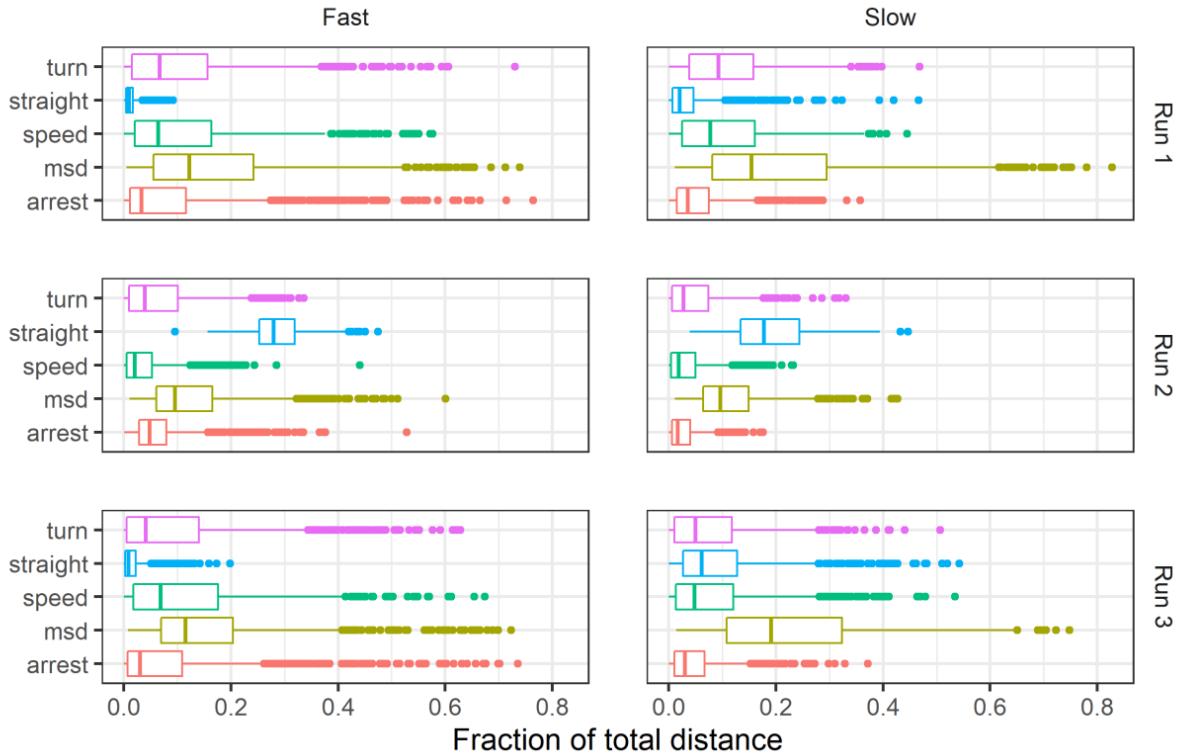


Figure 8 Distributions of fractional distances of each motility parameter. Fractions of the distances of each motility parameter of the total distance were calculated per particle. 1 000 particles were analysed for Model 3, generation 16 and shown as a boxplot. The total distances were in the ranges [0.4, 2.8] for Run 1, [1.6, 2.8] for Run 2 and [0.4, 2.8] for Run 3. The full parameter names are abbreviated. The parameters are highlighted with colours.

3.2.2 Performance analysis reveals discrepancies in fitting scenarios

Before analysing the summary statistics and credibility intervals of the models, evaluation of the fitting performance was undertaken. In addition to the later discussed summary statistics and parameter distributions, the performance of the fits can give a first impression of how well the fits performed in comparison to each other. The performance was measured via the acceptance rates of the generations of each *Run* and Model. A higher acceptance rate generally indicates better overall correspondence of the simulated particles with the ground truth data.

The first evaluated model was Model 1. The acceptance rates of *Run 1* and *Run 2* were in a range between 6% (generation 15) and 80% (generation 1) for *Run 2* and between 40% (generation 16) and 70% (generation 2) for *Run 1*. *Run 3* showed a lower

acceptance rate throughout with a peak acceptance rate of 50% and with the lowest acceptance rate of 1%. The reason for the low acceptance rates in Model 1 will be discussed in section 4.1.

For Model 2, *Run 1* and *2* had comparable acceptance rates ranging from 17% [30%, *Run 2*] to 90%. *Run 3* showed a reduced acceptance rate range of [10%, 50%].

Run 1 and *Run 2* of Model 3 showed again similar performance with acceptance rates from 5% to 60% in *Run 1* and 5% to 70% in *Run 2*. Evaluation of the acceptance rates of *Run 3* showed a slight decrease in peak acceptance rate with a range from 12% to 50%. See Table 7 for a summary of the acceptance rates.

Table 7 Ranges of acceptance rates for different models. Squared brackets contain lowest and highest acceptance rates of each run, without specifying the generation.

Run	Model 1	Model 2	Model 3
1	[45%, 70%]	[17%, 90%]	[5%, 70%]
2	[6%, 80%]	[30%, 90%]	[5%, 60%]
3	[1%, 50%]	[10%, 50%]	[12%, 50%]

The performance analysis by itself was not sufficient to rate the quality of the fits. An examination of the summary statistics was performed next to find out whether the fits were able to accurately fit the ground truth statistics.

3.2.3 Examination of motility data shows deviances between Runs

After evaluation of the performance of the individual models it was examined how good the fits were able to approximate the applied ground truth data. Each *Model* and each *Run* were taken into consideration.

Model 1

Analysis of the fitted summary statistics revealed that *Run 1* achieved an accurate reproduction of the ground truth parameters. Looking at the boxplots, a correspondence of the means and widths of all four motility parameters could be observed (Figure S3; top). The same applied to the MSD where no apparent discrepancy between the fitted data and the ground truth was observable.

This tendency changed in *Run 2*, where the means and widths of the boxplots differed for all motility parameters. The discrepancy became more apparent for the MSD, where the overall MSD from the subsampled ground truth data was significantly lower than the simulated data (Figure S3; middle).

Run 3 was able to recreate the motility parameters with good approximation, especially regarding the shape of the MSD curve (Figure S3; bottom).

Model 2

In Model 2 the complexity was increased with the introduction of an underlying collagen network and the haptotaxis. Summary statistics of the runs showed similar trends as observed with Model 1. An overall correspondence between the motility parameters *Run 1* and *Run 3* were observed while more significant discrepancies in *Run 2* could be shown (Figure S4).

The most remarkable difference was found in the MSD and straightness. The distinct shape of the MSD curve observed in the ground truth data was not matched by the fitted data in *Run 2* (Figure S4; middle). After application of the subsampling in the analysis in *Run 3*, the shape was accurately fitted (Figure S4; bottom).

Model 3

In Model 3 a second, slower cell type was introduced. Analysis of the summary statistics showed the same tendencies described in Model 2 (correspondence of the parameters in *Run 1* and *Run 3*, deviations of MSD and straightness in *Run 2*). These trends applied to both cell types. More significant discrepancies were observed in the MSD of the slow cell type (Figure S5).

In summary, the results of the analysis of the summary statistics showed, that a *standard* analysis in *Run 2* with *subsampled* ground truth data led to inaccuracies when comparing the ground truth data to the simulated motility statistics. Application of *subsampling* in the analysis in *Run 3* led to an improvement across all models and was comparable to the accuracy of *Run 1* (Figure S3, S4, S5). Overall, the discrepancies between the fitted motility parameters and the ground truth motility parameters were more significant in Model 1 than they were in Models 2 and 3. In order to find out whether the differences for the motility statistics between the runs also applied to the parameter inference, it was analysed next.

3.2.4 Credibility intervals imply effectiveness of subsampling

The principle of inferring parameters with FitMultiCell is to use a prior set of parameters that influence the motility behaviour of the cells and a prior set of summary statistics (ground truth data). The ABC algorithm uses this data to try to fit the simulated data. The main goal of a fit is therefore to infer the correct parameters used in the original model. Analysis of the generated parameter distributions and credibility intervals was therefore a crucial part in the evaluation and validation of the effectiveness of subsampling.

Model 1

In *Run 1*, both true parameter values were found within the 25% credibility interval of the parameter distributions (Figure S6).

For *Run 2* and *Run 3*, none of the two parameters were found in the fit. The true parameter for the persistence strength was significantly outside the 95% CI. For the decay time, a similar observation was made, however it was not as dominant as for the persistence strength (Figure S6). A discussion of the inaccuracy of *Run 2* and *3* will be presented in chapter 4.

Model 2

The CIs for *Run 1* showed that the true parameter of the decay time was found within the middle of the 25% CI. The CIs of both the haptotaxis and persistence strength showed a positive (haptotaxis) and negative (PS) shift from the respective true values (Figure 9).

Run 2 was not able to infer all true parameter values. While the decay time was approximated within the 25% CI, the most significant deviation from the true parameter value was found with the haptotaxis (Figure 9). A remarkable observation was that the same shift of the CIs of the haptotaxis and PS as seen in *Run 1* was found for *Run 2* as well as *Run 3*.

Overall *Run 3* was able to most accurately approximate the true parameter values. The true values for decay time and persistence strength were both found within the 25% CIs and the haptotaxis just outside the 50% CI (Figure 9). Especially in comparison to *Run 2* a significant improvement of the parameter inference was observed.

In order to examine the explained shift of the haptotaxis and persistence strength, a ratio plot of the two parameters was generated (for Model 3, Figure S8).

3 Results

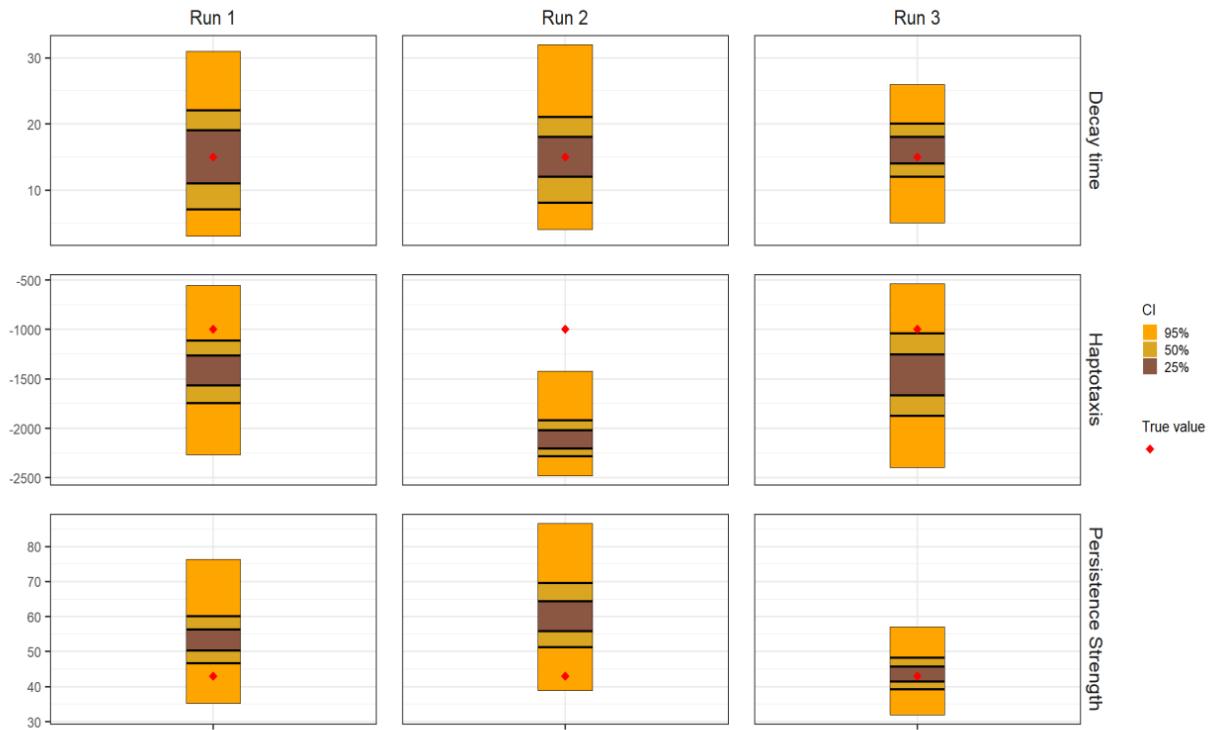


Figure 9 Credibility intervals for all three runs for Model 2. The parameters decay time, haptotaxis and persistence strength are shown. The red rhomb indicates the true parameter values. The effect of subsampling on the parameter inference can be observed when comparing the credibility intervals for Run 2 and Run 3. The credibility intervals (CI) were calculated from the respective final generations of each run (generation 15 for Run 3). The CIs 25%, 50% and 95% are shown and highlighted with different colour.

Model 3

In Model 3 the total number of parameters was increased from three to six by implementing a second cell type. Here, the motility behaviours of both cell types were considered.

The true value of the decay time was found within the 25% CI in *Run 1* for both cell types, whereas the haptotaxis and persistence strength showed the same shifts as observed in Model 2 (Figure S7). A ratio plot for both cell types showed that the true value ratios

$$\frac{\text{Haptotaxis}}{\text{Persistence strength}} = \begin{cases} -43.26, & \text{if cell type == "slow"} \\ -8.77, & \text{if cell type == "fast"} \end{cases}$$

3 Results

were found in *Run 1*. Minor deviations from the true ratios were observed in *Run 2* and *3* (Figure S8). A dependency between the parameters was assumed and supported by examination of the 2D density plots of the two parameters. The plot showed a tendency of a negative correlation (Model 3, *Run 1*: Figure S2).

Analysis of the CIs of *Run 2* revealed that only one of the six fitted parameters (decay time of the fast cell type) was found within the 25% CI. The persistence strength of the slow cell type was not found in any of the CIs. The remaining four parameters were found in the respective 95% CIs (Figure 10; left).

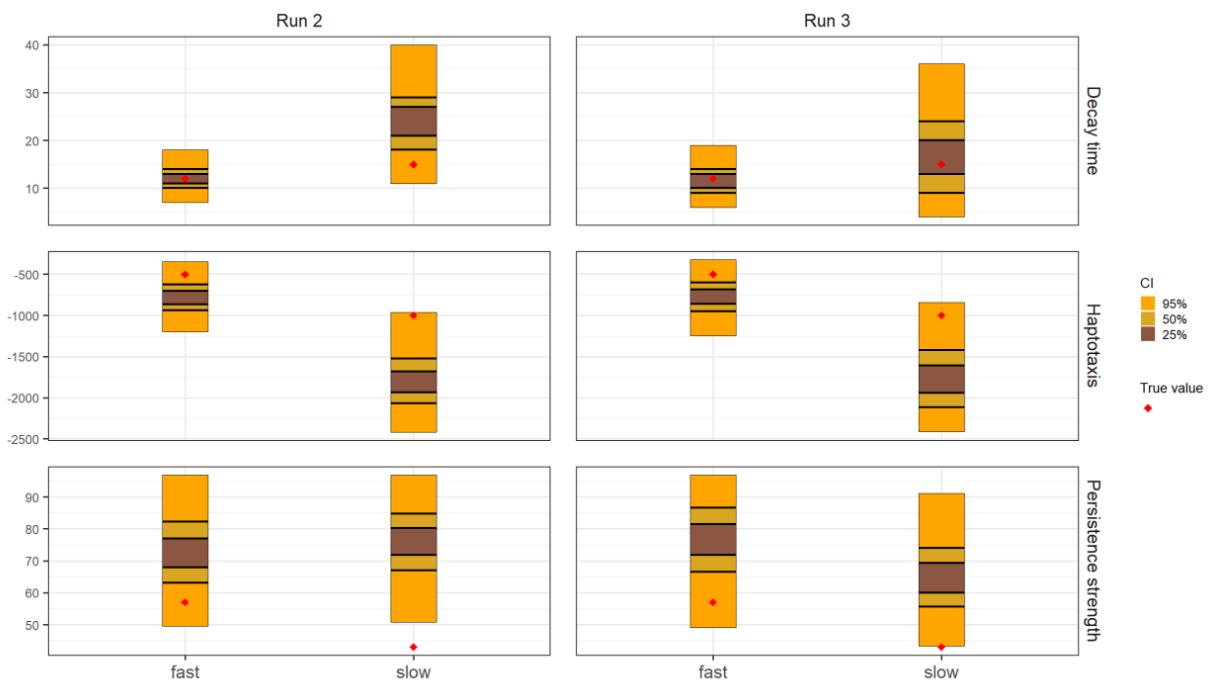


Figure 10 Credibility intervals of Model 3, Run 2 and 3. **Left:** Subsampled data is used as ground truth data with standard analysis. **Right:** Subsampled data is used as ground truth data with subsampling in the analysis. The parameters decay time, haptotaxis and persistence strength were fitted for each cell type. The true parameter values are indicated in red. CI stands for credibility interval (25%, 50% and 95% shown and highlighted with colours).

In *Run 3* an improved inference of the parameters of the slow cell type was found (Figure 10; right). The decay time was found within the 25% CI, as compared to the 95% CI in *Run 2*. Again, implementation of subsampling resulted in an improved parameter inference when compared to *Run 2*. The difference in accuracy was more prominent in the slow cell type than the fast cell type.

Direct comparisons of the credibility intervals for the different runs of the different models showed that implementation of subsampling led to an improved parameter inference when presented with *subsampled* ground truth data. In summary, the summary statistics as well as the posterior parameter distributions of *Run 2* performed significantly worse than *Run 3* across all models (exemption Model 1). With a proof of the effectiveness of the subsampling method on the parameter inference and motility statistics, deeper analysis of the effects of the subsampling depth on the parameter inference was performed next. A change in subsampling depth was hypothesized to further improve the effects of subsampling on the parametrization.

3.3 Multiple factors determine effect of subsampling depth

With the previous results providing evidence for the effectiveness of the subsampling, the next step was to determine how toggling the subsampling depth influenced the accuracy of the fits. Obtaining results for the subsampling depth could give insight into optimal fitting settings in future applications. The subsampling depth, as defined in section 2.4 describes how often subsampling is applied on a tracking dataset. This is achieved by subsampling a set of tracking data n times and then calculating the motility statistics and selecting the sample with the lowest aggregated distance.

In theory, a higher subsampling depth could correspond to a better approximation of the motility data and to a lower aggregated distance. In analogy to Harmel (2020), a plot comparing different subsampling depths to the total aggregated distances was generated. Subsampling depths from $n = 2$ to $n = 100$ in steps of 1 were used to calculate the aggregated distances with the ground truth data from Model 2 (here, Model 2 was selected representative for the observations. Analysis of the subsampling depth of the other models showed identical results). For each depth, a total of 100 samples were generated. For each subsampling depth, the best sample, the unsampled aggregated distances and the sampled distances were plotted. The results can be found in Figure 11.

With increasing subsampling depth, the aggregated distances decreased monotonically in all three scenarios. Comparing the different types of sampling, the

3 Results

best sample performed with the lowest aggregated distance across all subsampling depths. The plot also revealed that beginning at a subsampling depth of $n \approx 15$, no significant decrease in the aggregated distance was observable. Increasing the subsampling depth even further over $n \approx 25$, a nearly constant aggregated distance for all scenarios was observed.

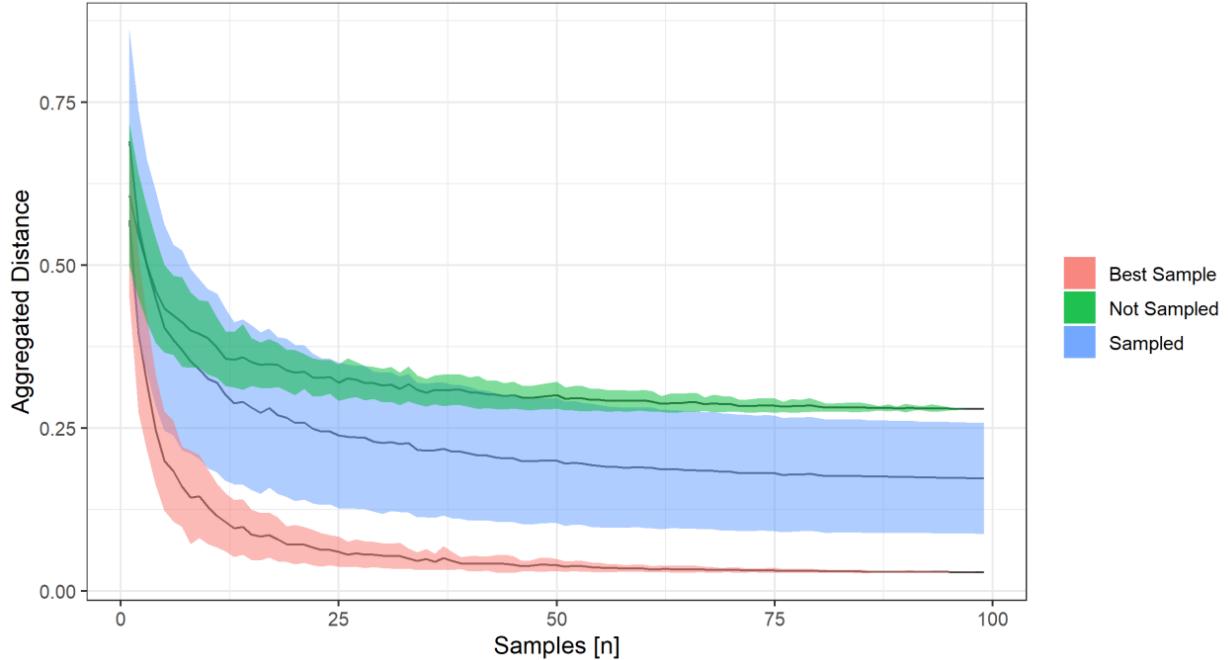


Figure 11 Aggregated distances in dependency of the subsampling depth. Three scenarios: Best Sample (red), Not Sampled (green) and sampled (blue) are shown. For each scenario, every subsampling depth was calculated 100 times and the corresponding aggregated distances generated. The aggregated distances were calculated by normalizing the absolute distances by their respective maximum, individual ratios of the distances were calculated and divided by the total distance. n stands for the subsampling depth. This plot shows results from the analysis performed on Model 2 and is representative for all models.

These findings led to the assumption that a higher subsampling depth could improve the performance of the pyABC fits, especially with regards to the motility data. In order to verify or falsify this assumption, a set of fits was performed. There were different aspects considered when comparing different subsampling depths:

- (1) Ensure comparable simulation conditions between all runs.
- (2) Select multiple depths with sufficient variation.

For (1) an ε -list was generated based on the previous subsampling runs. The list can be found in Table S2. Additionally, the acceptance rates as well as the distribution of the distances of the respective final generations were analysed to ensure homogeneity between the runs. For (2), observation made in Figure 11 were used to select different subsampling depths.

All runs used for the analysis of the subsampling depth were fitted with Model 3.a. The subsampling depth function was programmed to always select the best sample as described in Figure 11.

3.3.1 Altered fitting settings reveal effect of subsampling depth

The following subsampling depths were compared:

$$n = [1, 6, 25]$$

As seen in the Figure 11, the most significant changes in aggregated distance were observed in the range from $n = 1$ to $n = 15$. Because of those results, two subsampling depths ($n = 1$, $n = 6$) in that range were selected. In addition to that, no significant reduction in aggregated distance for $n > 25$ was visible (see Figure 11). As a result, a maximum subsampling depth of $n = 25$ was chosen.

For the evaluation of the different subsampling depths, the same criteria as before were utilized i.e., the summary statistics of the best particles of the respective last generations and the parameter distributions (credibility intervals) of the last generations. For the fitting itself, the number of accepted particles per generation was reduced from 1 000 to 50. In addition to that, a modified version of Model 3 was introduced, labelled Model 3.a. In Model 3.a, the parameter values for the persistence strength were sampled from a normal distribution with a standard deviation of 20. The motivation behind this change was to generate a broader range of motility parameters in the subsampling depth (Figure 12, purple boxplots slow cell type). By that, a higher stochasticity in the ground truth data could be ensured which increased the difficulty of the fits.

3 Results

The analysis did confirm significant differences between the individual motility parameters in the fits of the different depths. In Figure 12 the width as well as the means of the individual motility parameters differed significantly between the different subsampling depths. The deviations were most apparent for the fast cell type. Here the lowest subsampling depth of $n = 1$ underestimated the ground truth motility distributions as well as the mean for the turning angle and arrest coefficient, while it overestimated the velocity and straightness (Figure 12; left). These observations were less prominent with the slower cell type that generally showed a wider distribution in the ground truth data. The summary statistics for the subsampling depth of $n = 6$ showed an improvement in terms of the mean values of the simulated data in comparison to the ground truth data for the fast cell type. The subsampling depth of $n = 25$ managed to accurately fit the simulated summary statistics to the ground truth data. Especially the motility parameter distribution of the fast cell type showed no significant differences compared to the ground truth data. The MSD plot showed no significant differences between the subsampling depths (Figure S10).

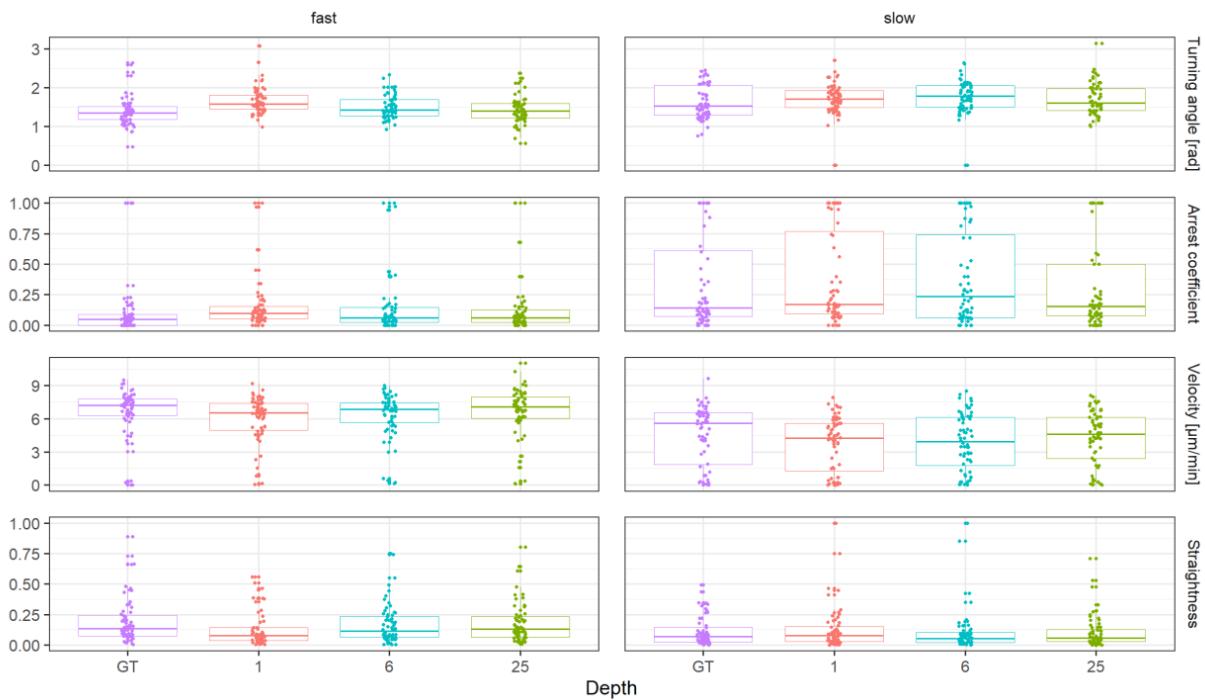


Figure 12 Comparison of the motility parameter (excluding MSD) for the subsampling depths $n = 1, 6$ and 25 for Model 3.a. **Left:** Summary statistics of the fast cell type. **Right:** Summary statistics of the slow cell type. All statistics were calculated with Model 3.a for the final generation of each Run. GT stands for the ground truth data and is highlighted in purple. Every colour stands for a distinct fitting Run with a different subsampling depth.

3 Results

Another factor that defines the accuracy of a fit is the posterior parameter distribution. It can provide a measurement of how well the parameter inference of a fit has performed. The CIs 25%, 50% and 95% were selected and plotted in a grid for all fitted parameters and depths.

When comparing the CIs of the different subsampling depths, it was observed that the true parameter value for the decay time was found most closely in the highest subsampling depth $n = 25$. Here, the true values were found just outside the 25% CIs. An in-depth discussion of the appearance of the decay time intervals will follow in chapter 5. For the lower subsampling depths $n = 1$ and $n = 6$, the true parameter values for the decay time were found best for the slow cell types within the respective 95% and 50% CIs. Analysis of the remaining parameters showed that again, the highest subsampling depth was able to most accurately find the true parameter values with the true parameters of the fast cell type on the borders of the 25% CI and for the slow cell type within the 95% CIs (Figure S9). The true values for the persistence strength of the lowest depth were found within similar borders as compared to the subsampling depth of $n = 25$, however the haptotaxis was significantly more off the true values (Figure S9). The same applied for the subsampling depth $n = 6$. In analogy to the previously discussed models, a plot depicting the ratios

$$\frac{\text{Haptotaxis}}{\text{Persistence strength}}$$

was created which showed convergence to the true parameter ratio with increasing subsampling depth (Figure S11).

Another criterion that can be used to compare the efficacy of two runs is a metric, that gives an indication of how good a fit approximated the prior ground truth data. A commonly used metric in Bayesian statistics is the Kullback-Leibler divergence (Jiang et al., 2018) where the log ratio between the probability distributions of the posterior and prior data is calculated. In the case of pyABC fitting, the discussed cost function d (section 2.3.1) was utilized to compare the runs. It must be considered, however, that comparability between the runs needs to be ensured in order to adequately compare

the distances. As discussed, this was ensured with the implementation of an ϵ -list. For the comparison, the distance of the best particle of the final generation of each depth was used as well as the total number of evaluated particles across all generations. As seen in Table 8 for a depth of $n = 1$, a minimal distance of 1.93 was obtained. The total number of evaluated particles for all generations was 4 501. For a subsampling depth of $n = 6$, the best distance decreased to 1.37 with 5 160 total evaluated particles. The subsampling depth $n = 25$ resulted in a best particle with $d_{best} = 0.5$ with 5 037 total evaluated particles.

Table 8 Distances of the best and worst particle from the last generation for each depth. Distances were obtained from the dataset created during the pyABC fit. The number of evaluated particles was calculated by adding up the evaluated particles from each generation. The average per generation was additionally calculated. All depths were analysed using Model 3.a.

Depth	d_{best}	$n_{evaluated}$ particles	$\bar{n}_{evaluated}$ particles
1	1.93	4 501	281.3
6	1.37	5 160	322.5
25	0.50	5 037	315.0

3.4 Subsampled contact ratio matches cell dropout ratio

The evaluation of the quality of the FitMultiCell fits was performed based on the performance, motility statistics and parameter distributions of the runs. The Bayesian comparison of the runs was performed with calculated ground truth data using different motility statistics. Deviations in the motility analysis have been shown when a cell dropout was present (Figure 5). The dropout of cells does not only influence the resulting motility statistics but might also influence cell-to-cell interactions. In order to find out, how it could influence the cell-to-cell interactions quantitatively, an according analysis was performed. For this, Model 3 was used to simulate cell interactions. Individual contacts between the cells were tracked for different simulations. It was differentiated between slow-slow, slow-fast and fast-fast interaction types.

3 Results

First, a constant dropout was used, and the cell contacts were compared for different subsampling depths. Here, the dropout from the previous analysis was selected (Figure 6). For each depth, 5 samples were generated and plotted with their respective mean counts and standard deviations. A subsampling depth of 0 corresponded to the tracked cell contacts of the complete dataset. As seen in Figure 13 for all types of contacts, the contact number remained approximately constant with increasing subsampling depth. The total number of interactions was lowest for the slow-slow interactions and highest for the slow-fast interaction.

Based on these observations, it was hypothesized that a constant ratio between the number of dropped out contacts and the total contacts might exist. In order to investigate this, different dropout rates and types were examined. The dropout types linear and exponential were compared along with dropout rates of 30%, 70% and 99%. An overview of the dropout types and rates is given in Figure S14. Analysis of the resulting contact ratios was performed with a constant subsampling depth of $n = 20$ and 20 samples per dropout.

The analysis revealed that for both exponential and linear dropout rates, an overall correspondence of the mean ratios between the interaction type and the dropout ratio was observable. This especially hold true for a dropout rate of 30%, where the means of both dropout types as well as the contact types were found nearly identical. For the dropout rate 70% and 99%, more apparent deviations were observed. Generally, the ratios of the contact types tended to undershoot the value of the dropout ratio. Within the contact types, the slow-slow interaction seemed to show the most prominent scattering with the means mostly undershooting the mean of the dropout ratio (Figure 14).

In order to evaluate whether the observed contact dropout ratios corresponded to the dropout curves (Figure S14), the cumulative (contact) count ratios were analysed (Figure 15). A subsampling depth of $n = 10$ was selected and 10 samples per dropout type and rate were simulated. The cumulative contacts were calculated for each interaction type and divided by the total contacts per interaction type, resulting in the

3 Results

cumulative (contact) count ratios. In order to ensure comparability, the cumulative dropouts of the cell tracks (Figure S14) were calculated and corrected by the total cell tracks to generate the cumulative dropout ratios. The mean of the 10 simulations was plotted alongside the corrected cumulative dropout rate of the cell tracks (shown in red; it will be referred to as *red line*).

While the ratios of the contacts showed little deviation when compared to the ratio of the cell track dropout (Figure 14), the cumulative count ratios showed more significant deviations from the red line (Figure 15). Across all dropout types and rates, a negative deviation from the red line could be observed for the slow-fast interaction contact ratios. This deviation was most prominent for the 30% and 70% dropout rate. The cumulative count ratios of the remaining interaction types showed less significant deviations. The slow-slow interaction type tended to undershoot the red line for the first 60 min of the simulation and overshoot it for the remaining 60 min. The overall shapes of the red lines however were matched in close approximation by the cumulative count ratios. This became more apparent for the higher dropout rates (especially 99%) (Figure 15). Another remarkable result was observed across all dropout types and rates (except the 99% drop out of the exponential dropout type) with the slow-slow cumulative ratio. A sharp increase in the cumulative count ratio beginning on the 1-hour mark could be observed. This observation will be discussed in section 4.4.

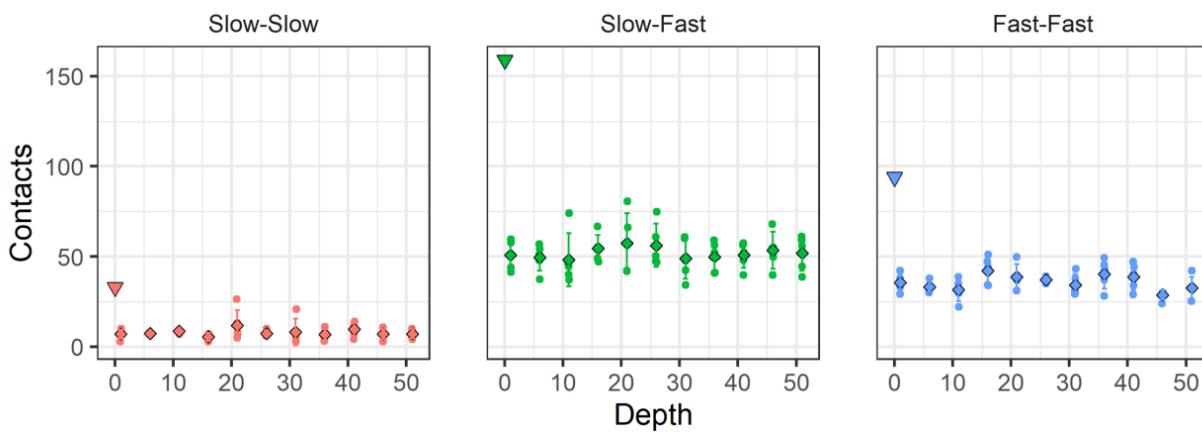


Figure 13 Comparison of the total contact count different subsampling depths. The contact counts were divided into the interaction types slow-slow, slow-fast and fast-fast. The inverted triangles depict the number of contacts of the complete dataset. Subsampling depths from 1 to 51 in steps of 5 are shown. For each depth, 5 samples were generated and the mean and sd calculated. The mean for each depth is highlighted via a rhomb.

3 Results

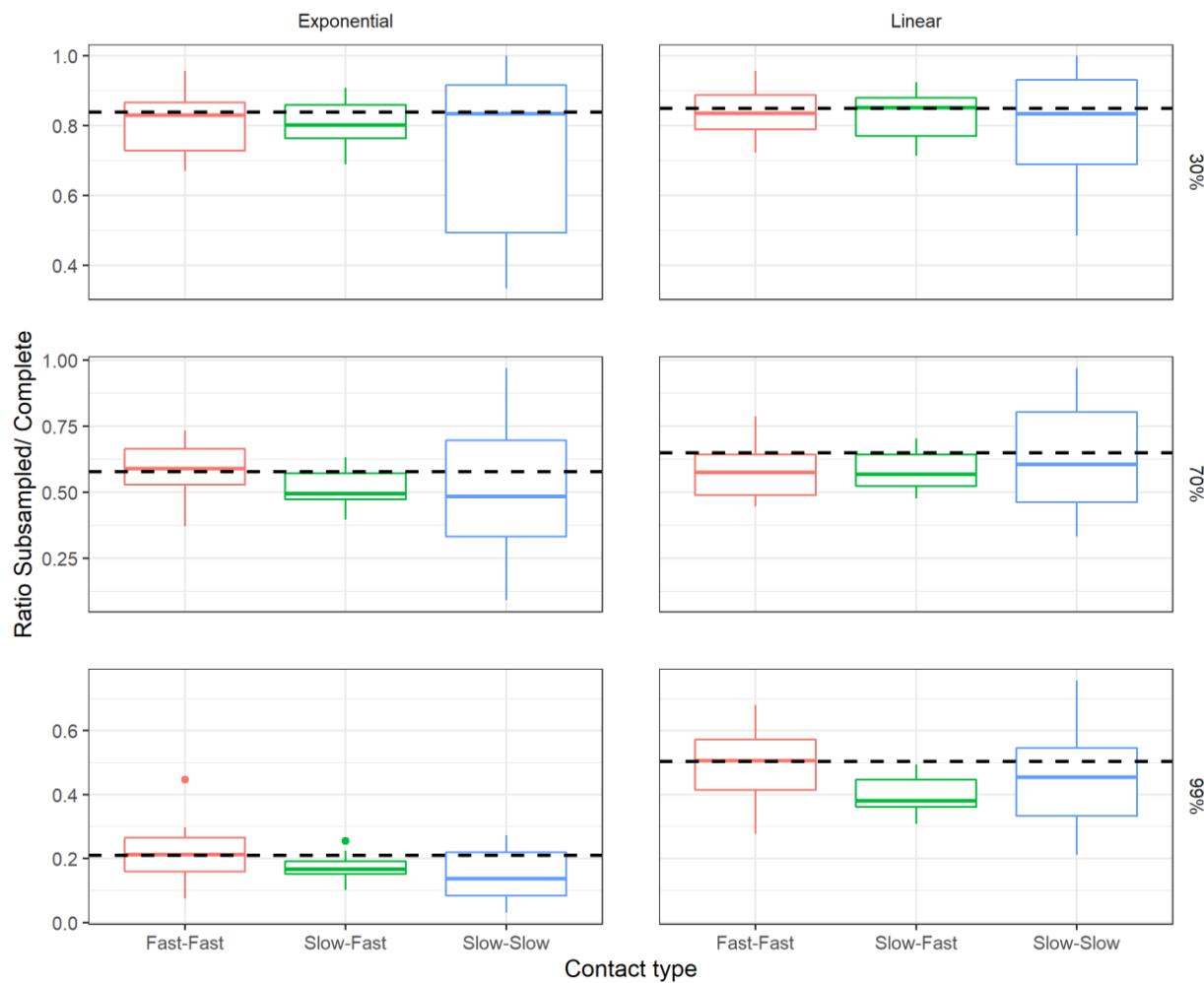


Figure 14 Comparison of total contact count ratios for different subsampling strategies. On the left, exponential dropouts were simulated with a subsampling depth of $n = 10$. 10 samples were generated per dropout. On the right, a linear dropout was simulated under the same conditions. For both strategies, dropout rates of 30%, 70% and 99% were examined. The dashed black lines indicate the dropout ratio as observed in generated tracking data (Figure S14; bottom). The different interaction types are highlighted in different colours.

3 Results

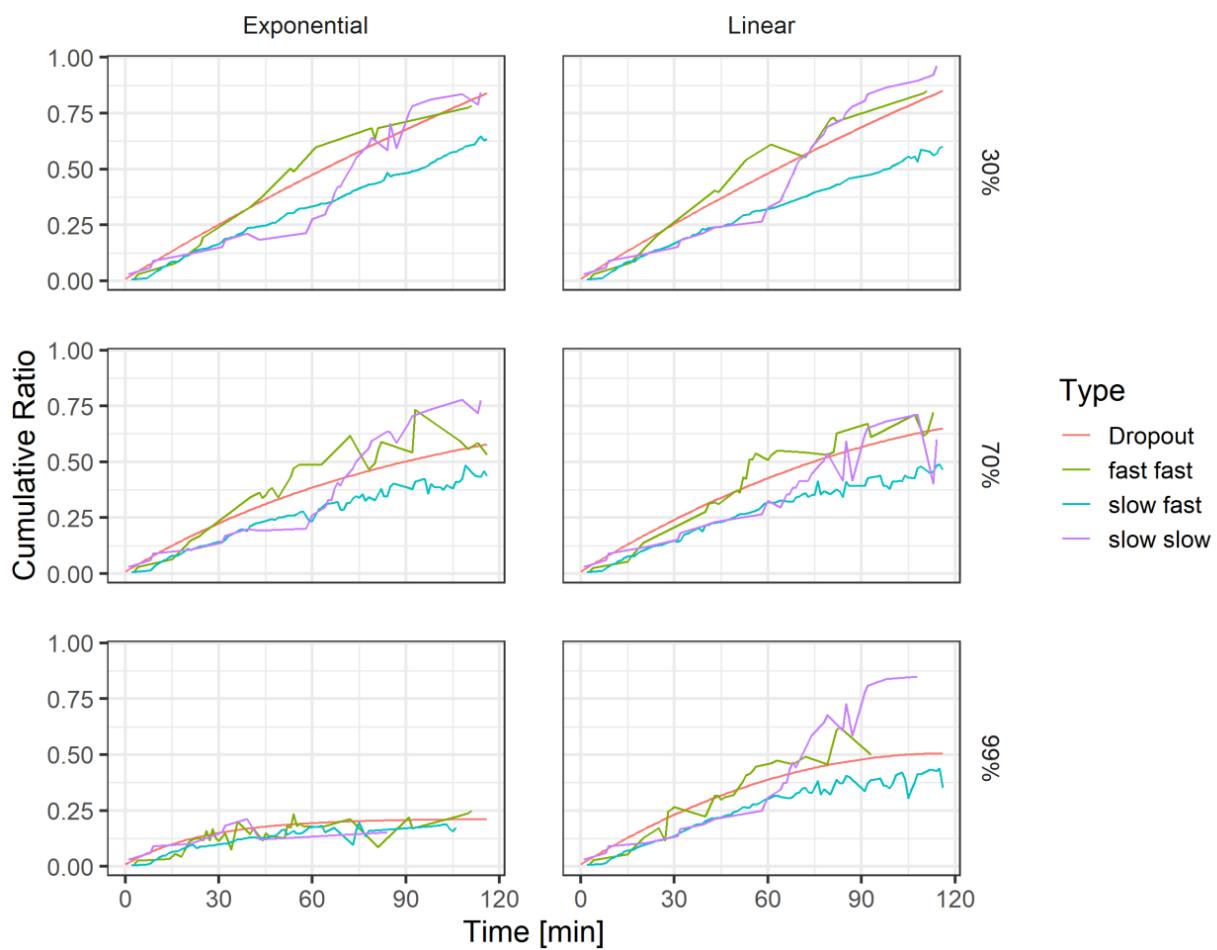


Figure 15 Comparison of the cumulative count ratios with the dropout curves. Dropout types linear (right) and exponential (left) as well as the dropout rates 30%, 70% and 99% are shown. The red line indicates the cumulative dropouts of the cell tracks (Figure S14). The coloured lines show the means of the cumulative contact ratios for the different interaction types for 10 repetitions. The y-axis shows cumulative ratios of the counts of the interaction types in comparison to the cumulative ratio of the dropout of the cell tracks.

4 Discussion

The aim of this thesis was to examine and validate the method subsampling in the context of parameter inference within the FitMultiCell framework. This method was motivated by observations made in live-cell microscopy tracking data. Here, in contrast to an assumed constant number of cells over a tracking period, a cell dropout over time has been observed (Harmel, 2020). In combination with the fitting of a CPM within the FitMultiCell workflow (Starruß et al., 2019), this dropout was hypothesized to impair the parameter inference. As a solution, the subsampling method was proposed by Harmel (2020) and was validated in this thesis.

The validation of the method was structured into different parts. In a first step, it was examined whether the implementation of subsampling in the FitMultiCell analysis had any effect on the performance of the fits. This was achieved by implementing different models that were increasingly complex. For each model, different fitting settings were compared in order to study out the effects of the subsampling and to see whether different biological conditions (collagen network, additional cell type) would influence the parameter inference when applying subsampling.

In a previous work from Harmel, he showed that with an increasing subsampling depth, a decrease in the aggregated distance would be induced. It was theorized that this also applied to the accuracy of the fits with FitMultiCell. Subsequently, different subsampling depths were examined and fitted.

Finally, the effects of subsampling on the cell-to-cell contact behaviour were examined. A two-cell model was implemented, and the analysis was performed for different subsampling depths and dropout strategies.

Since subsampling can be understood as a new method and extension of the FitMultiCell pipeline, not many previous works will be referenced in the discussion.

4.1 Summary statistics show deviation when ignoring cell loss

Before evaluating the parameter inference, the obtained results for the summary statistics will be discussed. As explained, three runs were defined that differed in the selection of their respective ground truth data and analysis setting. In *Run 2* significant deviations from the ground truth data were observed across all models (Figures S3-S5). Those observations complied with the previous assumptions that *incomplete* ground truth data could lead to inaccuracies in the fitting of the summary statistics. The motility parameters straightness and MSD showed the most significant deviations from the *complete* ground truth data across the models. When taking into consideration how the summary statistics were calculated (Textor et al., 2016), an explanation can be given.

Starting with the MSD, it was apparent that the distinct shape of the MSD as observed in the ground truth data for Model 2 and 3 was not matched by either of the models in *Run 2* (Figures S3, S4, S5; middle). The simulated cell tracks within the FitMultiCell pipeline were analysed without consideration of the cell dropout in the ground truth data. Consequently, the MSD was calculated by normalized summation of the squared displacement of a full cell count at each given time point, resulting in the observed shapes of the curve. In contrast, *Run 3* was able to accurately simulate the correct shape of the MSD curve for all models (Figures S3, S4, S5; bottom). Implementation of subsampling led to a correction of the MSD for all models which showed its positive effect on the fitting of the summary statistics.

The second parameter that showed significant deviation compared to the ground truth data was the straightness. The exact reasoning behind this deviation will be explained in the next section. Again, addition of the subsampling showed an improvement and led to a better fit in terms of the motility parameter (Figures S3, S4, S5).

4 Discussion

These results were in accordance with the previous results from Harmel (2020). In his fits, he was able to account for the cell dropout in the experimental data with the help of subsampling and generated overall matching summary statistics (Harmel, 2020).

Despite those general observations, a rudimental difference has been observed between Model 1 and Models 2/3. As explained, Models 2 and 3 considered an implemented dense collagen network that influenced the motility of the cells. When comparing the ground truth motility statistics (and fitted motility statistics) of Model 1 without consideration of a collagen network to the ground statistics (and fitted motility statistics) of Models 2 and 3 where a collagen networks was considered, a significant difference was observed. Whereas the *complete* and *subsampled* ground truth statistics in Model 1 showed significant differences between all motility parameters (Figure 5), Models 2 and 3 showed significantly less variation (see Figures S3, S4, S5). Additionally, the distinct shape of the MSD curve when implementing subsampled ground truth data was only observed in Models 2 and 3, where a dense collagen network was applied.

The same observations applied to the analysed fitted motility statistics. The motility parameters of the *subsampled* ground truth motility statistics mostly differed from the *complete* motility statistics regarding the MSD and straightness in Models 2/3. Even though the introduction of a collagen network in Model 2 as well as a second cell type in Model 3 introduced a multitude of additional complexity and interaction possibilities, it did not lead to a significant change of all motility statistics parameters. The effects of collagen networks on the motility statistics of *subsampled* and *complete* ground truth data should therefore be strongly considered when fitting within the FitMultiCell pipeline.

For future analysis it should be studied, how the implementation of a looser collagen network might influence the difference between *subsampled* and *complete* ground truth data (in fits within the FitMultiCell pipeline).

4.2 Subsampling positively influences parameter inference

As shown on Models 1, 2 and 3, the introduction of subsampling resulted in an improved parameter inference. For Models 2 and 3, a significant improvement of the relative positioning of the true parameter values within the CIs was observed (Figures 9, 10, S6, S7). The results seemed to confirm the earlier obtained results from the motility statistics where subsampling positively influenced the accuracy of the fits. It must be added that Model 2; *Run 3* included only 15 generations compared to the 16 defined generations in the fit. Yet, this did not influence the accuracy of the parameter inference as seen in Figure 9.

However, these results must be interpreted with caution. The observed results differed for the CIs of Model 1 *Run 2* and *3*, where both *Runs* failed to find the true parameter values opposed to the accurate results from the remaining Models. It would have been expected that Model 1 with an overall lower model complexity would have been able to better approximate the true parameter values. The true parameter values of the persistence strength for both parameters were found significantly outside the 95% credibility interval (Figure S6). Since the CIs of *Run 1* showed good results, the problem might have resulted from the subsampled ground truth data. However, it is also possible, that the poor accuracy of the fit resulted from a local minimum that was reached within the multi-dimensional parameter space. The problem of high-dimensionality in ABC algorithms has been described before (Csilléry et al., 2010; Templeton, 2009). The results observed in the analysis of the performance of the fits also seemed to strengthen this assumption. Here, significantly lower acceptance rates were observed for *Run 2* and *Run 3* (Table 7). Despite these observations for Model 1, there was uniform agreement between the results of the other Models that supported the effectiveness of the subsampling.

Another factor that initially did not correspond with the assumptions, were the observations made with the CIs of the haptotaxis and persistence strength. As seen in Figures 9 and S7 a tendency of the haptotaxis to underestimate the true parameter value as well as of the persistence strength to overestimate the true parameter value

4 Discussion

was observed. Further analysis showed that a dependency between the two parameter values was present. A constant ratio between the parameters has been observed across the Models 2, 3 and 3.a and was found to be approximated by all *Runs*, including *Run 2*.

In addition to the parameter distributions and summary statistics, analysis on the composition of the distances within the pyABC fits was performed. The results showed no significant differences between the distance composition of *Run 1* and *3* (Figure 8; top and bottom). This was appreciable since subsampling was implemented in order to correct the cell dropout while remaining accurate in terms of fitting and comparable to a fit without any cell dropout. *Run 2* showed a deviation in the distance composition with the straightness contributing the most to the distance for both cell types (Figure 8; middle). This corresponded to the previous observation that showed the straightness to be the motility parameter with the most significant deviations. Examination of the formulas for the individual motility parameters as described by Textor et al. (2016), showed that the straightness is obtained by dividing the adjusted track length of a cell over the entire simulation time by its individual sub lengths over the time. If a track loss is not accounted for in form of subsampling, deviation of the straightness would become more apparent resulting in a higher distance and thus in a higher relative part of the aggregated distance.

Finally, another more rudimentary factor might be influential to the outcome of the credibility intervals. As described in (Kendall and O'Hagan, 1994) there are multiple ways to calculate credibility intervals based on a posterior parameter distribution. In order to ensure comparability and because it fit the posterior distributions the best, the equal-tailed interval (ETI) was selected for the analysis in this thesis. Selection of other methods to calculate the credibility intervals such as the highest posterior density interval (HPDI) might reveal minimally different posterior parameter distributions. Yet, comparability between the individual runs can only be ensured when utilizing the same approach across all fits, so both methods might be workable. Since the comparison of different credibility interval calculation methods was not a crucial part of the validation

of subsampling, it was not further pursued. If necessary, additional analysis could be performed regarding this topic.

In conclusion, the results showed that, in general, the implementation of a subsampling strategy succeeded in improving the parameter inference as well as the fitting of summary statistics in a FitMultiCell fit when using incomplete tracking data. For different models designed to cover multiple biological conditions and challenges, a significant improvement in the parameter inference as well as the accuracy of the motility statistics could be observed. Nevertheless, attention should be paid to the differences in ground truth data when implementing dense collagen networks into the models as described in section 4.1.

4.3 Effect of subsampling depth depends on particle number and model design

The presented results for the subsampling depth (Figure 12, S9, S10) seemed to confirm previous results from Harmel (2020). In Figure 11 an increase in subsampling depth resulted in a decrease in aggregated distance, which could be shown for all studied models. This was most prominent for the best sample as defined by Harmel (2020). Based on these observations, the subsampling depth was studied.

For the examination, Model 3.a was introduced. In contrast to the previous models, Model 3.a did not use constant prior parameter values but sampled values for the persistence strengths from a normal distribution. This change was implemented intentionally in order to increase the stochasticity in the ground truth data (see Figure 12; GT statistics). The obtained results showed that an increase in subsampling depth in fact improved the accuracy of the fits both in terms of the summary statistics and the parameter inference (Figures 12, S9, S10, S11).

As an additional way to compare the depths, the distances of the best particles were compared in Table 8. With the distances describing the accuracy of the shown motility statistics, the previously stated results were verified. An increase in subsampling depth

led to a decrease in the total distance of the best particles. Additional evaluation of the total number of simulated particles per depth showed that no drastic increase in the number of evaluated particles was observed. The most significant increase occurred between $n = 1$ and $n = 6$ (Table 8), however no increase in the number of evaluated particles was observed for the highest subsampling depth. The same applied to the average number of evaluated particles per generation. These findings indicated no significant effects of the subsampling depth on the total number of evaluated particles.

The evaluation of the parameter distribution showed surprising results. The CIs from the comparison of the subsampling depth showed an apparent absence of the 25% and 50% CIs for the decay time. This can be explained by the fact that the decay time was sampled from a discrete prior distribution. The low number of 50 particles per generation in combination with a discrete prior distribution led to a low range of simulated parameter values for the decay time. The respective 25% and 50% CIs consisted of only one or two distinct values resulting in the observed CIs. These observations were made based on the parameter distributions and summary tables.

However, given the small number of accepted particles, caution must be taken when interpreting the effects of these results. In a previous set of fits that compared the subsampling depths $n = [1, 6, 10, 25, 50]$ for Model 3 with 1 000 accepted particles per generation, no significant differences in the subsampling depth for both the summary statistics and the CIs were observed (Figures S12, S13). These findings were consistent across all five subsampling depths. In the figures, three subsampling depths were selected for reference (Figures S12, S13). According to Figure 11, an increase in the accuracy of the fits via a decrease of the aggregated distance would have been expected, at least for the summary statistics. Yet, these findings might be explained by the fact that because of the large number of accepted particles per generation, a large sample size of particles was evaluated and thus only the best of these particles were selected. In contrast, the lower number of particles in Model 3 forced the algorithm to rely more on the additional sampling performed by the subsampling.

As an example, for a subsampling depth of $n = 1$ and $n = 25$:

1. Fitting with Model 3 and a total number of accepted particles of 1 000:
 - $n = 1$: A total of $1 \cdot 1\ 000 = 1\ 000$ particles are evaluated.
 - $n = 25$: A total of $25 \cdot 1\ 000 = 25\ 000$ particles are evaluated.
2. Fitting with Model 3.a and a total number of accepted particles of 50:
 - $n = 1$: A total of $1 \cdot 50 = 50$ particles are evaluated.
 - $n = 25$: A total of $25 \cdot 50 = 1\ 250$ particles are evaluated.

Now consider that every 10th evaluated particle is accepted by the algorithm. This would result in the following constellations:

1. For scenario 1 with a subsampling depth of $n = 1$; 100 particles would be accepted on average. For a depth of $n = 25$; 2 500 particles would be accepted per generation.
2. For scenario 2, with a depth of $n = 1$; on average 5 particles would be accepted whereas with a depth of $n = 25$; on average 125 particles per generation would be accepted.

This shows how the number of accepted particles per generation can have a great influence on the effect of the subsampling depth. Statistically speaking, it is much more likely to draw a good particle out of a set of 2 500 particles than it is out of a set of 5 particles. Consequently, a fit with a lower number of accepted particles will be more likely to contain particles with worse motility parameters than a fit with a higher number of accepted particles.

Taking the computational expanse of a fit into consideration, a lower number of acceptance particles in combination with a higher subsampling depth could probably be the most efficient way to design a pipeline when applying subsampling. The reason for this is that subsampling takes significantly less time than running an entire simulation. This is especially true when implementing models that take a longer time for the simulation. Here an increase of the subsampling depth while decreasing the number of accepted particles might result in a more efficient and quicker overall fit without compromising accuracy.

Because these findings are the first of their kind, they need to be viewed with caution and might not be representative of a broader application. Multiple factors need to be considered when thinking about the efficiency and accuracy of a subsampled FitMultiCell fit such as the model design, the subsampling depth, number of accepted particles per generation and many more.

4.4 Contact analysis implies relationship between dropout and subsampling

Analysis of the contact behaviour of subsampled data showed that a relationship between the dropout ratio as observed for the cell tracks and the contact count ratio of the subsampled data to the original contact counts seemed to be present (Figure 14). While the subsampling depth did not influence the unique contact counts, a change of the dropout strategy with a constant subsampling depth seemed to greatly influence the counts as well the contact count ratios (Figures 13-15). As presented in section 3.4, the contact count ratios for the different dropout strategies seemed to directly correlate with the respective dropout ratios of the cell tracks. These findings were consistent across both dropout types (linear, exponential) as well as all dropout rates (30%, 70%, 99%) with minor deviations (Figure 14).

Subsequent analysis of the cumulative (contact) count ratios seemed to confirm these findings. While the slow-fast curve consistently undershot the red line, the remaining interaction types generally approximated the red line accurately (Figure 15). However, further data collection and analysis would be needed to determine the exact relationship between the dropouts. The observed volatile course of the cumulative contact ratio curves of the different interaction types resulted from the calculation of the mean over 10 repetitions per interaction type. The individual curves per repetition would show a monotonic increase over time as seen with the cumulative dropout ratios (Figure 15).

Another surprising observation was the consistent increase of the slow-slow cumulative (contact) count ratio after 60 min (Figure 15; slow-slow interactions). This

increase was consistent across all dropout scenarios and could be explained upon closer observation of the contact logger used for the analysis. The contact logger included the time points and cell ids of each unique contact. Several contactless periods were observed between the slow cells until the 60 min mark. After 60 min a sharp increase in cell contacts was observable. This resulted in a sharp increase of the contact ratios for the slow-slow interaction (Figure 15; slow-slow interactions). The only exception to this tendency was found with the exponential dropout type at a dropout rate of 99%. As seen in Figure S14, an exponential dropout of 99% led to a sharp drop of cells in the first 60 min. After 60 min, less than 10% of the cells were left. Because of this, the missing sudden increase in slow-slow interaction as observed in the other plots could be explained. Since the contact count for the slow-slow interactions only started to increase significantly after the 60 min mark, the main proportion of it was filtered out when applying the exponential 99% dropout.

4.5 Conclusion and Outlook

The results proposed in this thesis were able to confirm the effectiveness of subsampling within pyABC fitting. It was shown that through the application of subsampling in the analysis of the FitMultiCell workflow, an improvement in the parameter inference could be obtained. Furthermore, it was shown that ignoring the cell dropout led to a decrease in accuracy of the pyABC fits for both the parameter inference and motility statistics.

The results suggested that FitMultiCell fitting is strongly dependant on the set of cell tracking data used for the calculation of the ground truth statistics. Cell tracking data is the starting point for the motility calculations and a central part of the fitting pipeline. It should therefore be taken into consideration before application of a fit in the FitMultiCell pipeline. It was shown that incomplete tracking data causes significant deviations of the corresponding ground truth motility parameters (Figure 5) as well as of the fitted motility statistics and parameters (Figures S3-S5, 9, S6, S7).

Considerable progress has also been made in the analysis of the effect of the subsampling depth. Depending on the fitting settings such as the number of accepted

4 Discussion

particles per generation, a significant improvement in accuracy when using a higher subsampling depth has been observed. However, it was also shown that these findings have their limitations when increasing the number of particles or decreasing the stochasticity of the model. These results require more data in order to draw viable conclusions.

Additionally, the effect of subsampling on the contact behaviour of different cells has been studied. Here, the obtained data showed conclusive results with regards to the relationship between the dropout of the cells and the corresponding dropout of the cell-to-cell interactions. The findings suggested a direct proportionality between the dropout ratios which was further supported by analysis of the cumulative contact ratios.

In summary, I was able to provide a first validation of subsampling as an extension of the established FitMultiCell analysis. When confronted with incomplete cell track data, one should consider the application of a subsampling method to obtain the best possible parameter inference. If further proven effective, subsampling could be applied to the analysis of any experimental cell tracking dataset that uses the FitMultiCell framework and contains incomplete tracking data. In addition to that, the specialized application of subsampling on the FitMultiCell framework could be generalized to the parametrization of models relying on incomplete cell tracking data with arbitrary ABC algorithms.

References

- Agosto, L.M., Uchil, P.D., and Mothes, W. (2015). HIV cell-to-cell transmission: effects on pathogenesis and antiretroviral therapy. *Trends Microbiol.* 23, 289–295.
- Armingol, E., Officer, A., Harismendy, O., and Lewis, N.E. (2021). Deciphering cell–cell interactions and communication from gene expression. *Nat. Rev. Genet.* 22, 71–88.
- Ashkin, J., and Teller, E. (1943). Statistics of Two-Dimensional Lattices with Four Components. *Phys. Rev.* 64, 178–184.
- Beaudin, L. (2007). A Review of the Potts Model A Review of the Potts Model : 8.
- Beaumont, M.A. (2010). Approximate Bayesian Computation in Evolution and Ecology. *Annu. Rev. Ecol. Evol. Syst.* 41, 379–406.
- Beaumont, M.A., Zhang, W., and Balding, D.J. (2002). Approximate Bayesian computation in population genetics. *Genetics* 162, 2025–2035.
- Beaumont, M.A., Cornuet, J.M., Marin, J.M., and Robert, C.P. (2009). Adaptive approximate Bayesian computation. *Biometrika* 96, 983–990.
- Beltman, J.B., Marée, A.F.M., and de Boer, R.J. (2009). Analysing immune cell migration. *Nat. Rev. Immunol.* 9, 789–798.
- Bocharov, G., Meyerhans, A., Bessonov, N., Trofimchuk, S., and Volpert, V. (2016). Spatiotemporal Dynamics of Virus Infection Spreading in Tissues. *PLoS One* 11, e0168576.
- Boelts, J. (2018). Model Comparison in Approximate Bayesian Computation.
- Bundgaard, N. (2020). Analyzing Cellular Dynamics In HIV-1 Spread Within A Complex Environment.
- Cerruti, B., Puliafito, A., Shewan, A.M., Yu, W., Combes, A.N., Little, M.H., Chianale, F., Primo, L., Serini, G., Mostov, K.E., et al. (2013). Polarity, cell division, and out-of-equilibrium dynamics control the growth of epithelial structures. *J. Cell Biol.* 203, 359–372.
- Chen, N., Glazier, J.A., Izquierdo, J.A., and Alber, M.S. (2007). A parallel implementation of the Cellular Potts Model for simulation of cell-based morphogenesis. *Comput. Phys. Commun.* 176, 670–681.
- Csilléry, K., Blum, M.G.B., Gaggiotti, O.E., and François, O. (2010). Approximate Bayesian Computation (ABC) in practice. *Trends Ecol. Evol.* 25, 410–418.

References

- Daun, S., and Clermont, G. (2007). In silico modeling in infectious disease. *Drug Discov. Today Dis. Model.* 4, 117–122.
- Dupré, L. (2015). T Lymphocyte Migration : An Action Movie Starring the Actin and Associated Actors. 6.
- Ettinger, A., and Wittmann, T. (2014). Chapter 5 - Fluorescence live cell imaging. In *Quantitative Imaging in Cell Biology*, J.C. Waters, and T.B.T.-M. in C.B. Wittman, eds. (Academic Press), pp. 77–94.
- Fischer, H.P. (2008). Mathematical modeling of complex biological systems: from parts lists to understanding systems behavior. *Alcohol Res. Health* 31, 49–59.
- Franz, C.M., Jones, G.E., and Ridley, A.J. (2002). Cell Migration in Development and Disease. *Dev. Cell* 2, 153–158.
- Gábor, A., and Banga, J.R. (2015). Robust and efficient parameter estimation in dynamic models of biological systems. *BMC Syst. Biol.* 9, 74.
- Ghaffarizadeh, A., Heiland, R., Friedman, S.H., Mumenthaler, S.M., and Macklin, P. (2018). PhysiCell: An open source physics-based cell simulator for 3-D multicellular systems. *PLOS Comput. Biol.* 14, e1005991.
- Graner, F., and Glazier, J.A. (1992). Simulation of biological cell sorting using a two-dimensional extended Potts model. *Phys. Rev. Lett.* 69, 2013–2016.
- Harmel, C. (2020). Determining the Effect of HIV- Δ Nef Mutation on Viral Spread in Complex Environments.
- Hespanhol, L., Vallio, C.S., Costa, L.M., and Saragiotto, B.T. (2019). Understanding and interpreting confidence and credible intervals around effect estimates. *Brazilian J. Phys. Ther.* 23, 290–301.
- Hirashima, T., Rens, E.G., and Merks, R.M.H. (2017). Cellular Potts modeling of complex multicellular behaviors in tissue morphogenesis. *Dev. Growth Differ.* 59, 329–339.
- Imle, A., Kumberger, P., Schnellbächer, N.D., Fehr, J., Carrillo-Bustamante, P., Ales, J., Schmidt, P., Ritter, C., Godinez, W.J., Müller, B., et al. (2019). Experimental and computational analyses reveal that environmental restrictions shape HIV-1 spread in 3D cultures. *Nat. Commun.* 10.
- Jackson, L., Rodel, H., Hwa, S.-H., Cele, S., Ganga, Y., Tegally, H., Bernstein, M., Giandhari, J., Team, C.-K., Gosnell, B.I., et al. (2021). SARS-CoV-2 cell-to-cell spread occurs rapidly and is insensitive to antibody neutralization. *BioRxiv* 2021.06.01.446516.

References

- Jagiella, N., Rickert, D., Theis, F.J., and Hasenauer, J. (2017). Parallelization and High-Performance Computing Enables Automated Statistical Inference of Multi-scale Models. *Cell Syst.* **4**, 194–206.e9.
- Jensen, E.C. (2013). Overview of Live-Cell Imaging: Requirements and Methods Used. *Anat. Rec.* **296**, 1–8.
- Jiang, B., Wu, T.Y., and Wong, W.H. (2018). Approximate Bayesian computation with Kullback-Leibler divergence as data discrepancy. *Int. Conf. Artif. Intell. Stat. AISTATS 2018* **84**, 1711–1721.
- Jouanneau, J., and Thiery, J.P. (2002). Tumor Cell Motility and Invasion. *J.R.B.T.-E. of C.* (Second E. Bertino, ed. (New York: Academic Press), pp. 467–473.
- Kendall, M.G., and O'Hagan, A. (1994). Kendall's advanced theory of statistics Vol. 2B, Vol. 2B, (London [etc.]: Arnold).
- Klinger, E., Rickert, D., and Hasenauer, J. (2018). pyABC: distributed, likelihood-free inference. *Bioinformatics* **34**, 3591–3593.
- Krummel, M.F., Bartumeus, F., and Gérard, A. (2016). T cell migration, search strategies and mechanisms. *Nat. Rev. Immunol.* **16**, 193–201.
- Neuenschwander, S., Largiadèr, C.R., Ray, N., Currat, M., Vonlanthen, P., and Excoffier, L. (2008). Colonization history of the Swiss Rhine basin by the bullhead (*Cottus gobio*): inference under a Bayesian spatially explicit framework. *Mol. Ecol.* **17**, 757–772.
- Ouchi, N., Glazier, J., Rieu, J.-P., Upadhyaya, A., and Sawada, Y. (2003). Improving the realism of the cellular Potts model in simulations of biological cells. *Phys. A Stat. Mech. Its Appl.* **329**, 451–458.
- Savill, N.J., and Hogeweg, P. (1997). Modelling Morphogenesis: From Single Cells to Crawling Slugs. *J. Theor. Biol.* **184**, 229–235.
- Schwann, T. (1838). Ueber die Analogie in der Structur und dem Wachsthum der Thiere und Pflanzen. Neue Not Geb Nat Heil.
- Scianna, M. (2015). An extended Cellular Potts Model analyzing a wound healing assay. *Comput. Biol. Med.* **62**, 33–54.
- Shriner, D., Liu, Y., Nickle, D.C., and Mullins, J.I. (2006). EVOLUTION OF INTRAHOST HIV - 1 GENETIC DIVERSITY DURING CHRONIC INFECTION. *Evolution (N. Y.)* **60**, 1165–1176.
- Sisson, S.A., Fan, Y., and Tanaka, M.M. (2007). Sequential Monte Carlo without likelihoods. *Proc. Natl. Acad. Sci.* **104**, 1760 LP – 1765.

References

- Sloan, K.E., Eustace, B.K., Stewart, J.K., Zehetmeier, C., Torella, C., Simeone, M., Roy, J.E., Unger, C., Louis, D.N., Ilag, L.L., et al. (2004). CD155/PVR plays a key role in cell motility during tumor cell invasion and migration. *BMC Cancer* *4*, 73.
- Starruß, J., de Back, W., Brusch, L., and Deutsch, A. (2014). Morpheus: a user-friendly modeling environment for multiscale and multicellular systems biology. *Bioinformatics* *30*, 1331–1332.
- Starruß, J., Schälte, Y., Hasenauer, J., Graw, F., de Back, W., Bundgaard, N., Brusch, L., and Alamoudi, E. (2019). Fitmulticell.
- Szabó, A., and Merks, R.M.H. (2013). Cellular potts modeling of tumor growth, tumor invasion, and tumor evolution. *Front. Oncol.* *3*, 87.
- Templeton, A.R. (2009). Statistical hypothesis testing in intraspecific phylogeography: nested clade phylogeographical analysis vs. approximate Bayesian computation. *Mol. Ecol.* *18*, 319–331.
- Textor, A.J., Dannenberg, K., Berry, J., Burger, G., and Textor, M.J. (2016). Package ‘MotilityLab’.
- Toni, T., Welch, D., Strelkowa, N., Ipsen, A., and Stumpf, M.P.H. (2009). Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J. R. Soc. Interface* *6*, 187–202.
- Webb, P.G., Spillman, M.A., and Baumgartner, H.K. (2013). Claudins play a role in normal and tumor cell motility. *BMC Cell Biol.* *14*, 19.
- WHO (2021). COVID-19 Weekly Epidemiological Update 35. World Heal. Organ. 1–3.
- Wolf, K., Müller, R., Borgmann, S., Bröcker, E.-B., and Friedl, P. (2003). Amoeboid shape change and contact guidance: T-lymphocyte crawling through fibrillar collagen is independent of matrix remodeling by MMPs and other proteases. *Blood* *102*, 3262–3269.

Supplementary

The supplementary figures are structured into summary statistics, parameter distributions, subsampling depth, contact analysis and tables.

Subsampling analysis

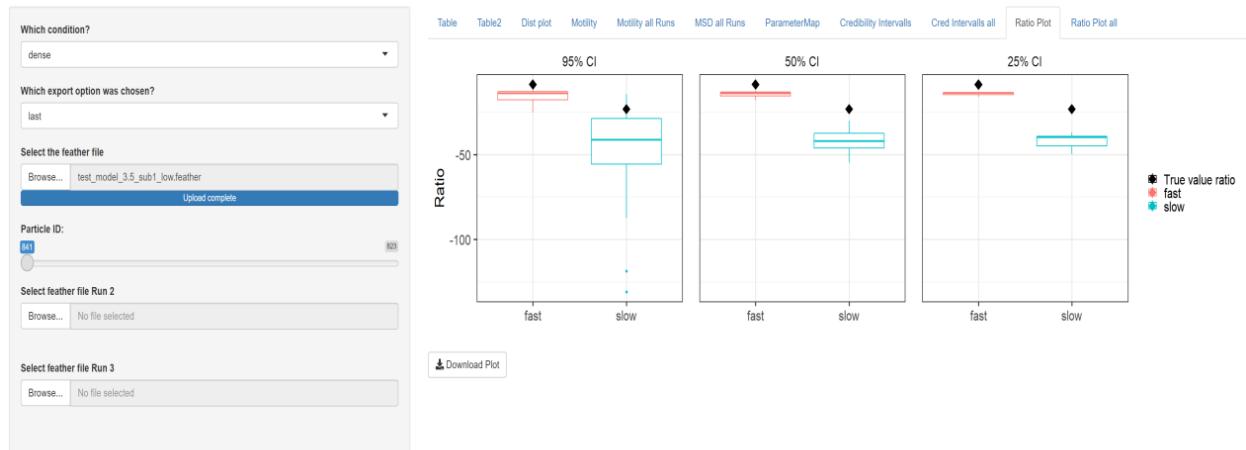


Figure S1 Screenshot of the utilized R Shiny application.

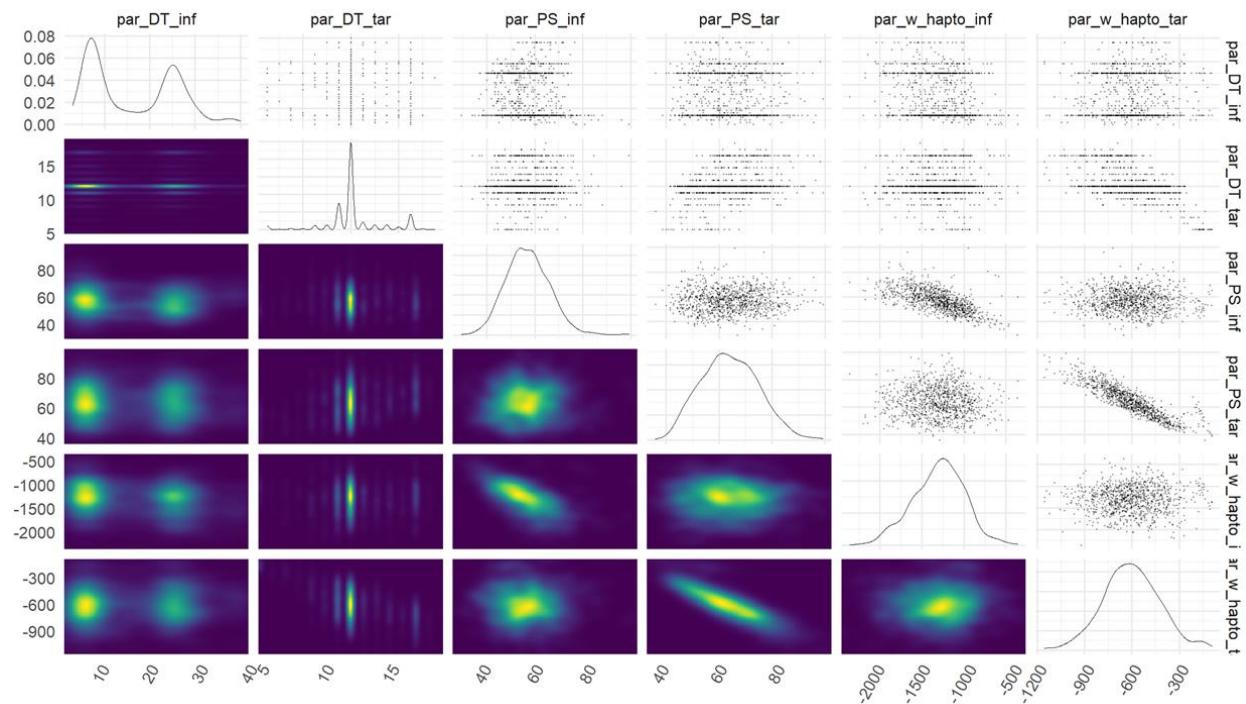


Figure S2 Posterior parameter distributions of Model 3, Run 1. The plots show the posterior parameter distributions of all fitted parameter in Model 3, Run 1, final generation as well as 2D density plots of each combination. A negative correlation between the parameters haptotaxis and persistence strength can be observed for both cell types.

Model 1 summary statistics

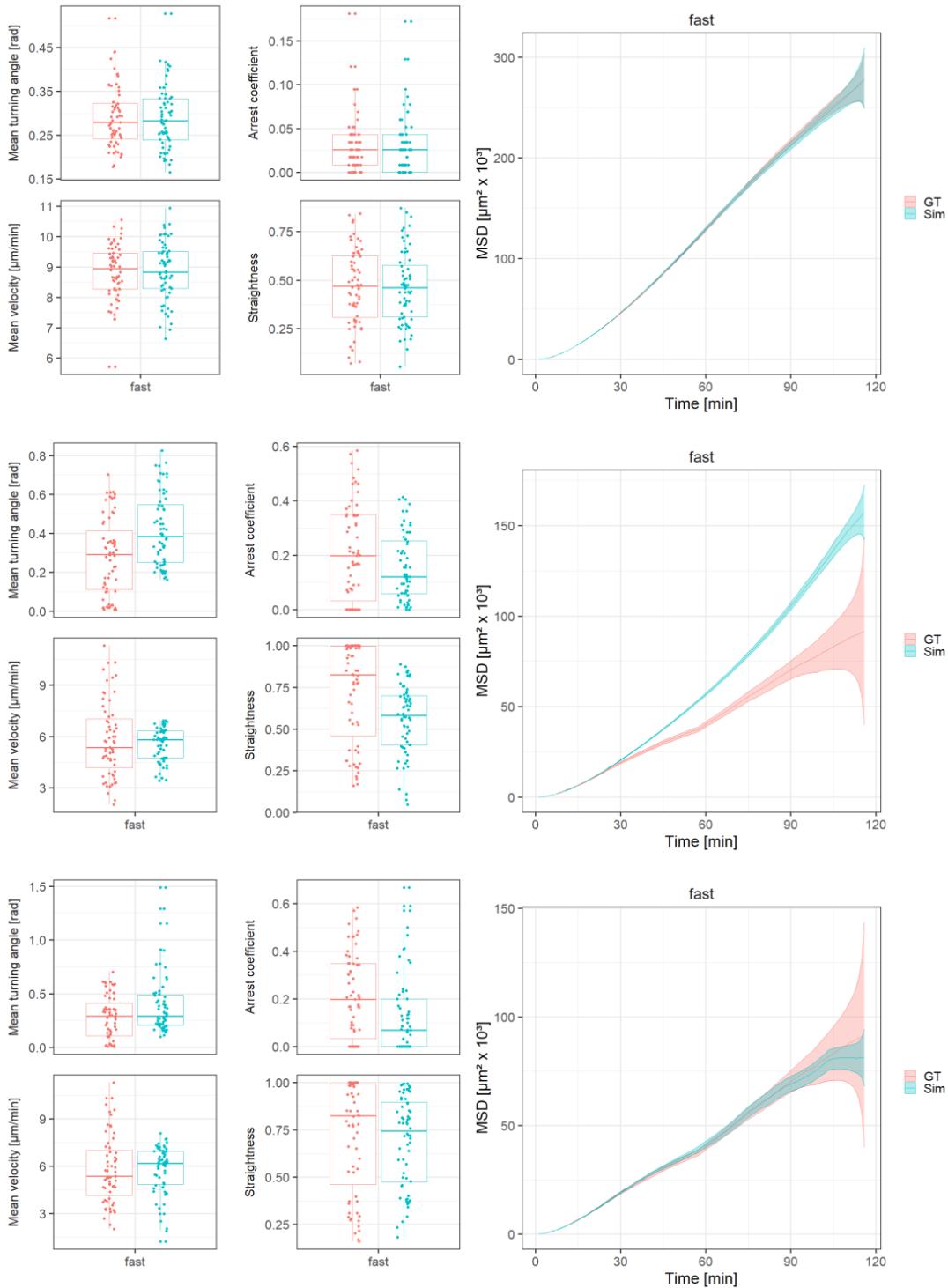


Figure S3 Summary statistics of the fitting runs for Model 1. The best particle of the last generation of each run is shown. Red indicates the ground truth motility data and blue the simulated motility data. **Top:** Run 1 with complete ground truth data and standard analysis. **Middle:** Run 2 with subsampled ground truth data and subsampling in the analysis. Run 2 was not able to accurately fit the motility parameters. **Bottom:** Run 3 with subsampled ground truth data and subsampling in the analysis. MSD stands for mean square displacement.

Model 2 Summary statistics

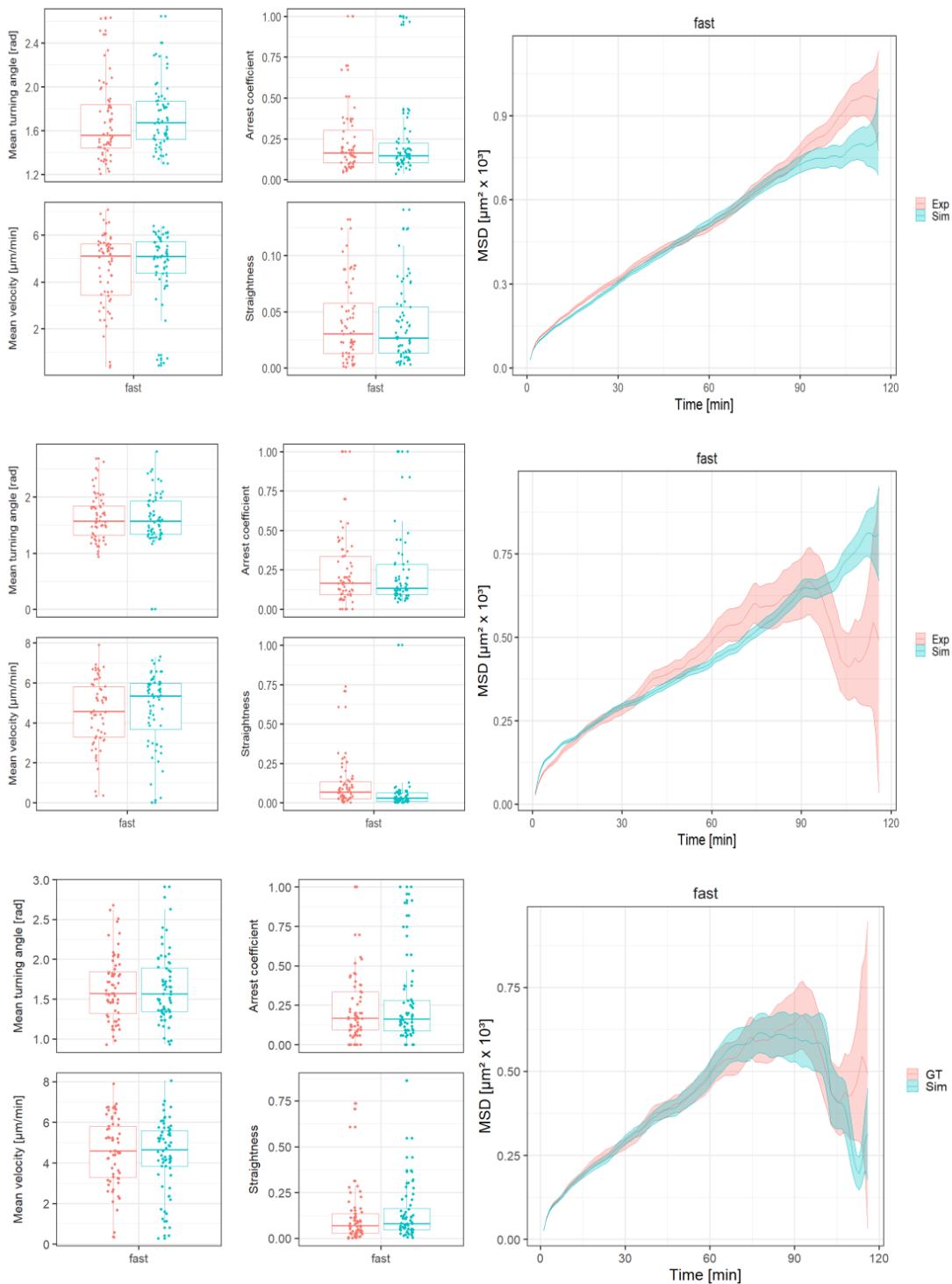


Figure S4 Summary statistics of the fitting runs for Model 2. The best particle of the last generation (generation 15 for Run 3) of each run is shown. Red indicates the ground truth motility data and blue the simulated motility data. **Top:** Run 1 with complete ground truth data and standard analysis. **Middle:** Run 2 with subsampled ground truth data and subsampling in the analysis. **Bottom:** Run 3 with subsampled ground truth data and subsampling in the analysis. Significant deviations were observed when comparing Run 2 to Run 1 and Run 3. Especially the shape of the MSD curve shows deviations. MSD stands for mean square displacement.

Model 3 Summary statistics

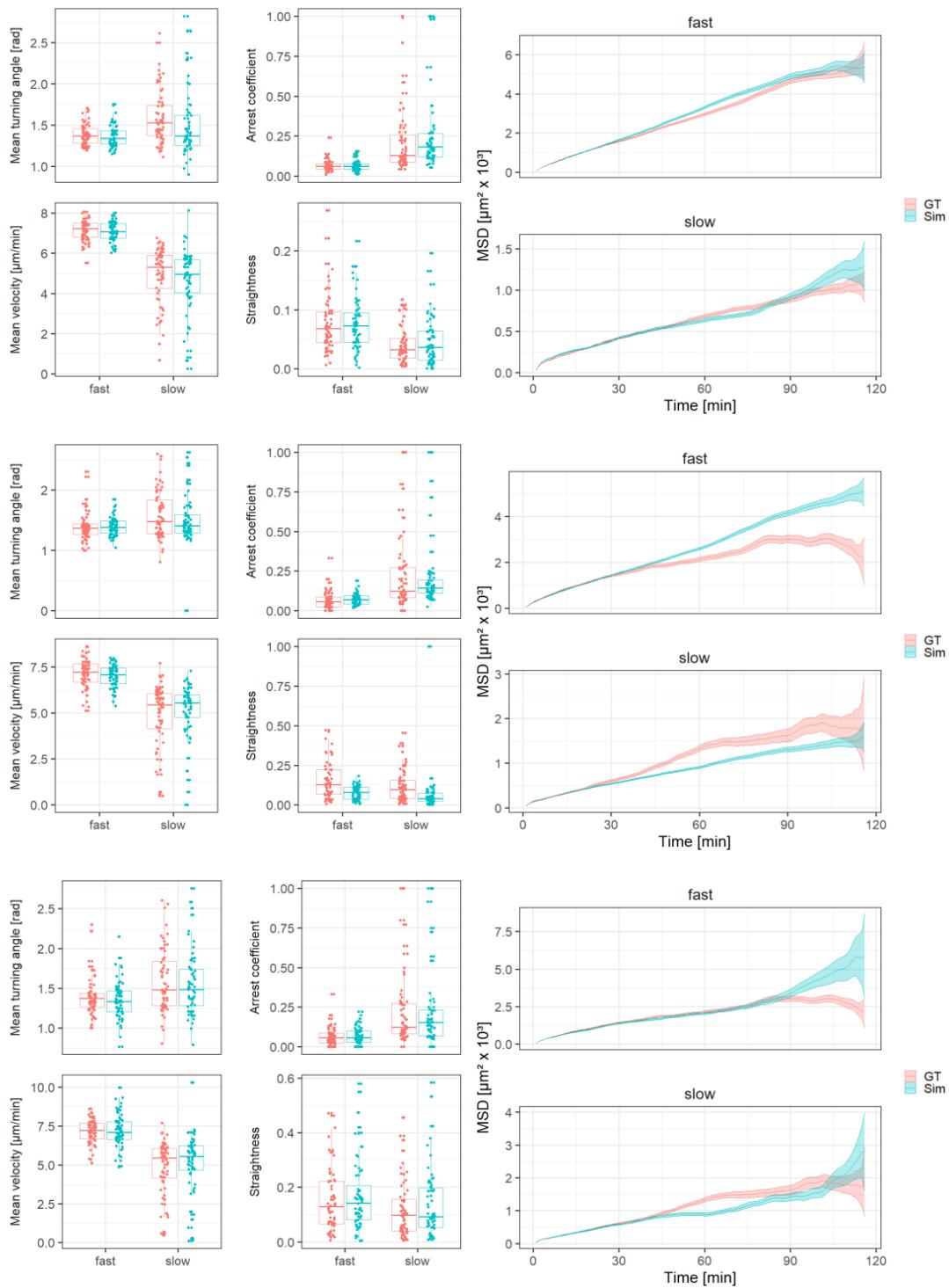


Figure S5 Summary statistics of the fitting runs for Model 3. The best particle of the last generation of each run is shown. Red indicates the ground truth motility data and blue the simulated motility data. **Top:** Run 1 with complete ground truth data and standard analysis. **Middle:** Run 2 with subsampled ground truth data and subsampling in the analysis. **Bottom:** Run 3 with subsampled ground truth data and subsampling in the analysis. The most significant deviations were observed for Run 2. MSD stands for mean square displacement.

Parameter distributions (credibility intervals)

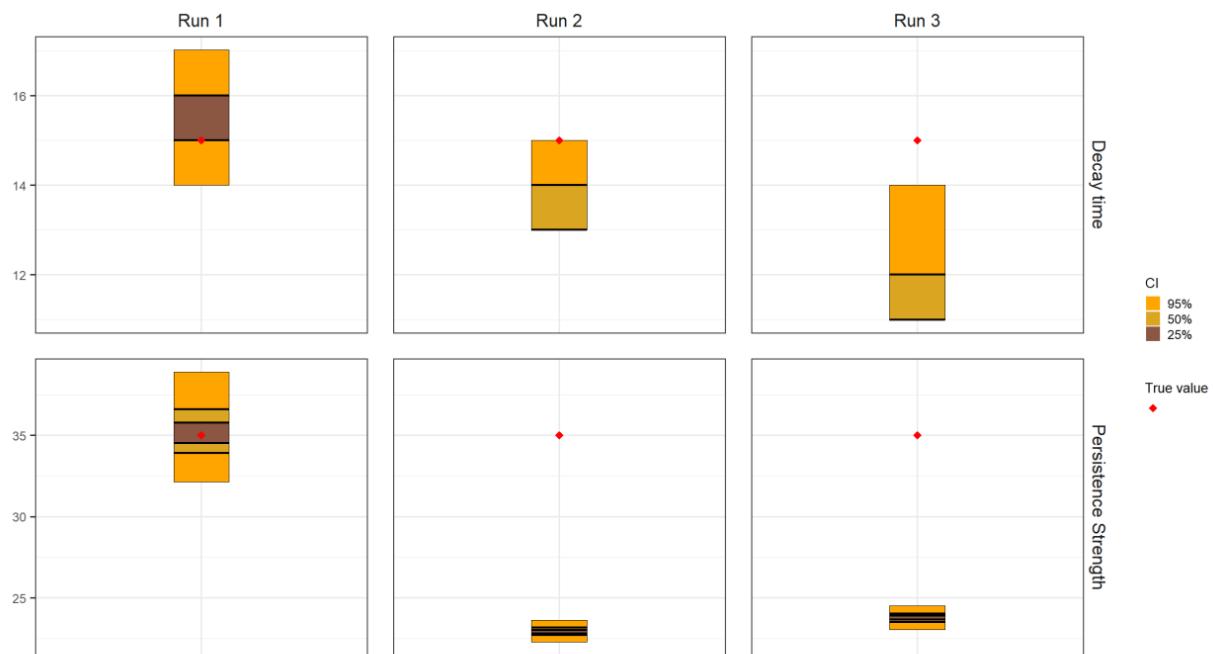


Figure S6 Credibility intervals for all three runs for Model 1. The parameters decay time and persistence strength are shown. The red rhomb indicates the true parameter values. The credibility intervals (CI) refer to the final generation of each run. The CIs 25%, 50% and 95% are shown and highlighted in different colours.

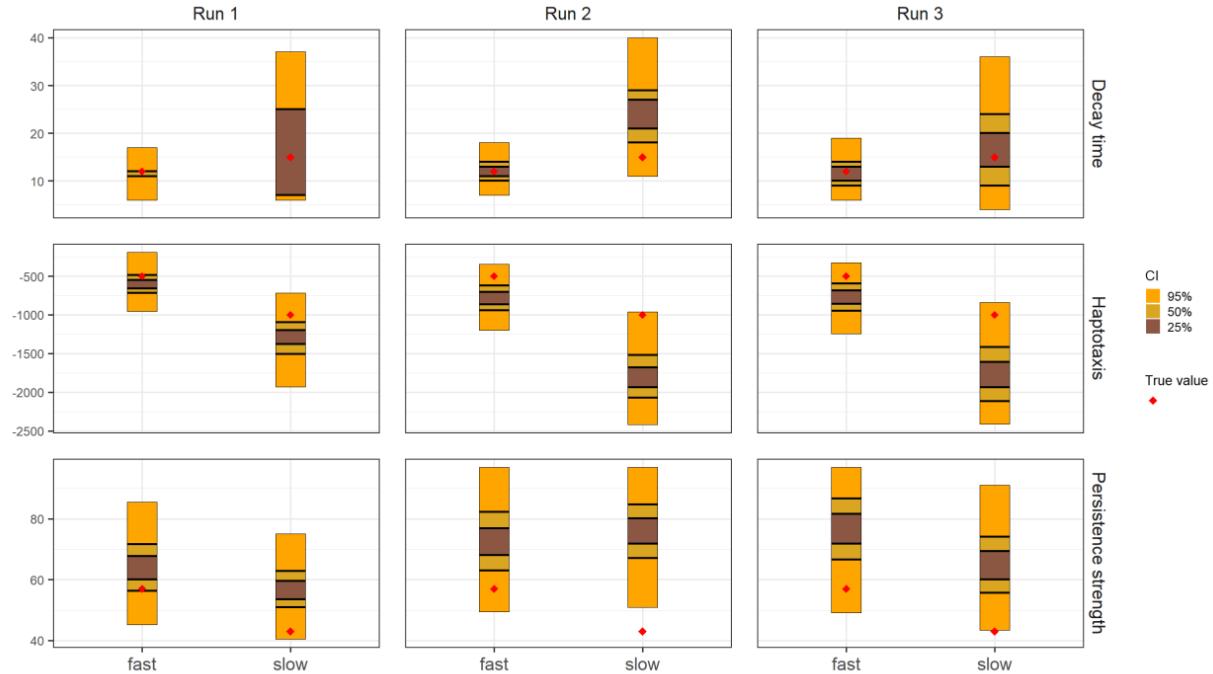


Figure S7 Credibility intervals for all three runs for Model 3. The parameters decay time, haptotaxis and persistence strength are shown for both cell types. The red rhomb indicates the true parameter values. The credibility intervals (CI) were calculated from the respective final generations of the individual runs. The CIs 25%, 50% and 95% are shown and highlighted in different colours.

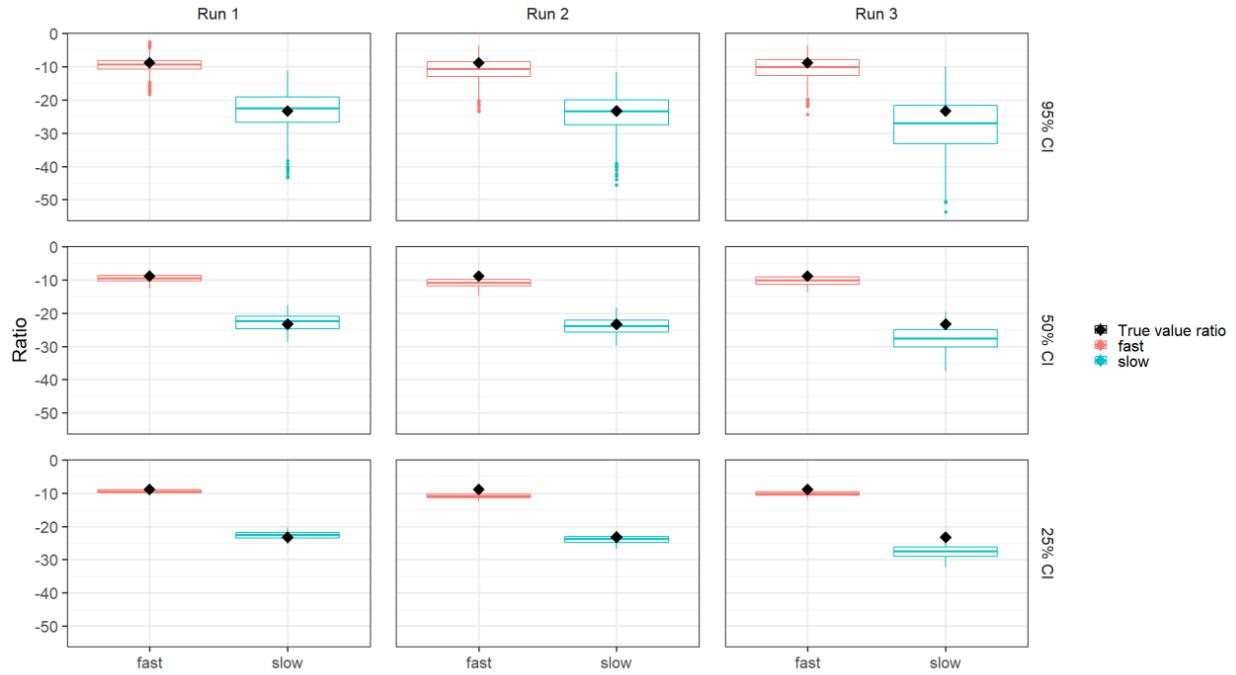


Figure S8 Distribution of the ratio haptotaxis:PS for Model 3. Comparison of the ratios for all three Runs. The ratios were calculated from the parameter values of the particles of the final generation. The credibility intervals 25%, 50% and 95% are depicted. True value ratios are indicated by the black rhomb. CI stands for credibility interval. The fast cell type is shown in red and the slow cell type in blue.

Subsampling depth

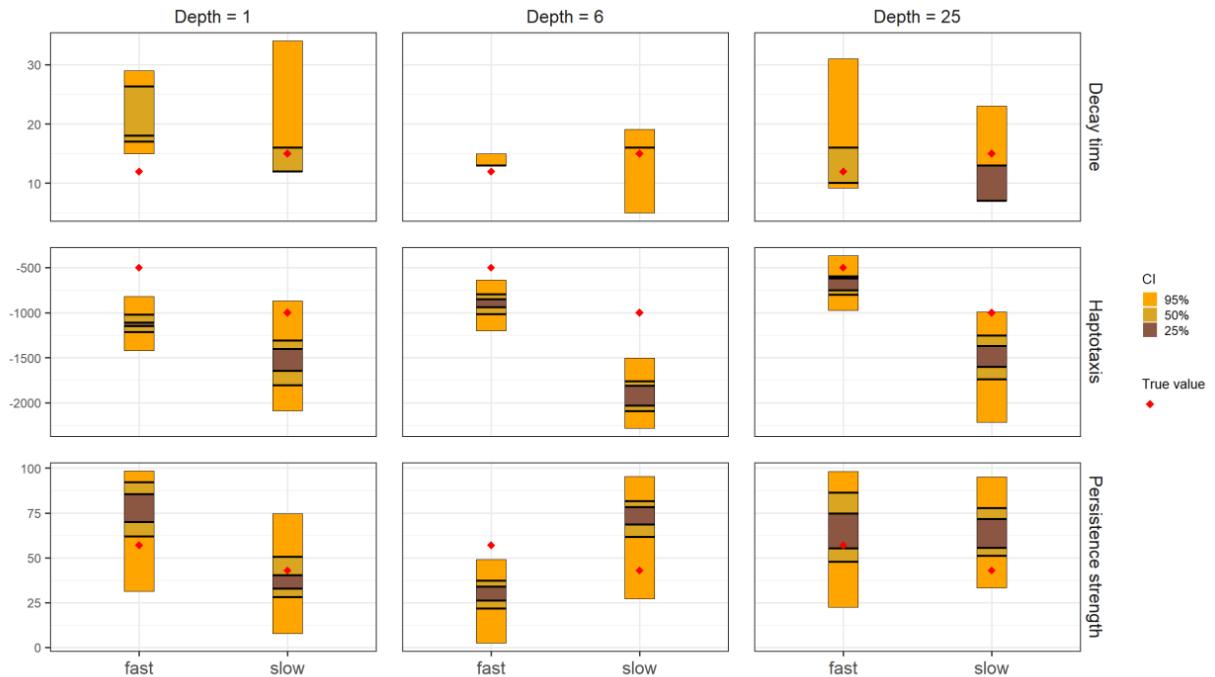


Figure S9 Credibility intervals for different subsampling depths for Model 3.a. The parameters DT, haptotaxis and PS and the depths $n = 1$, $n = 6$ and $n = 25$ are shown. The credibility intervals (CI) 25%, 50% and 95% are shown and the true parameter values are highlighted in red.

Supplementary

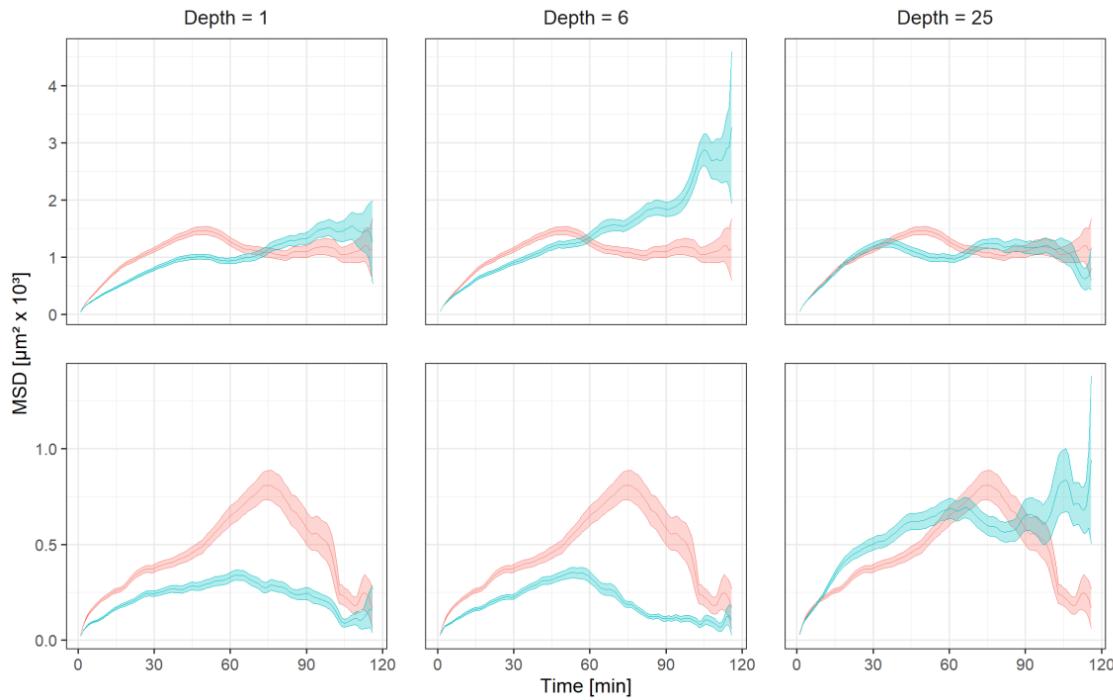


Figure S10 Comparison of the MSD curves for Model 3.a for different subsampling depths. The depths $n = 1$, $n = 6$ and $n = 25$ are depicted for the ground truth data and simulated data. An approximation of the true shape of the mean square displacement (MSD) curves can be observed with increasing subsampling depth. Red indicates the ground truth statistics and blue the simulated MSD statistics.

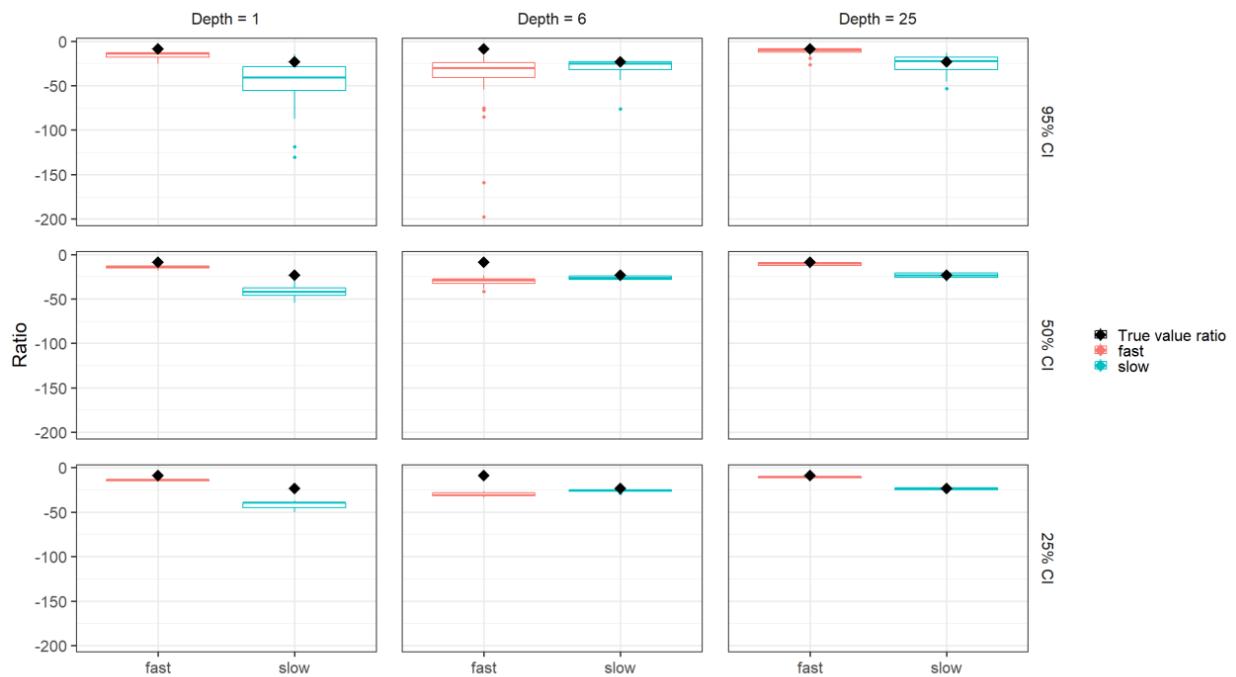


Figure S11 Comparison of the ratio haptotaxis:persistence strength for all credibility intervals for Model 3.a. The depths $n = 1$, $n = 6$ and $n = 25$ are depicted for the final generation. The black rhombs indicate the true parameter ratios. Blue shows the slow cell type and red the fast cell type.

Supplementary

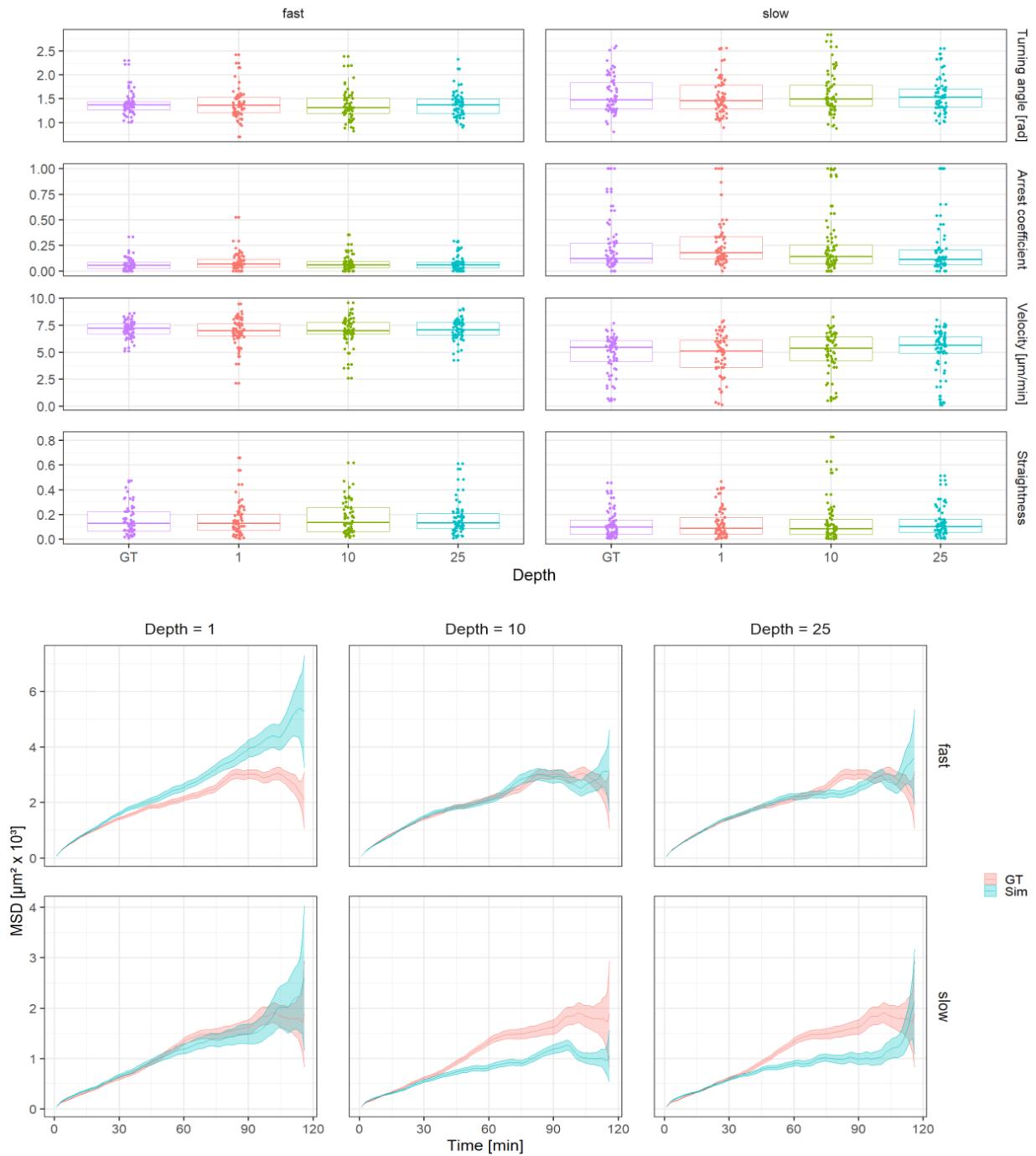


Figure S12 Comparison of the summary statistics for different depths for Model 3. The Summary statistics are shown for the last particle of the final generations of each fit. The shown depths were selected from the depths [1, 5, 10, 25, 50] and are representative for the overall observed characteristics in different colours. **Top:** Subsampling depths $n = 1$, $n = 10$ and $n = 25$ are shown, motility parameters turning angle, arrest coefficient, velocity and straightness are depicted for both cell types. Ground truth data (GT) is indicated in purple and simulated data in different colours for each depth. **Bottom:** Mean square displacement (MSD) of the best particles of the last generation for each depth. Subsampling depths $n = 1$, $n = 10$ and $n = 25$ are shown. The MSD curves are depicted with the ground truth curve shown in red and the simulated curve shown in blue.

Supplementary

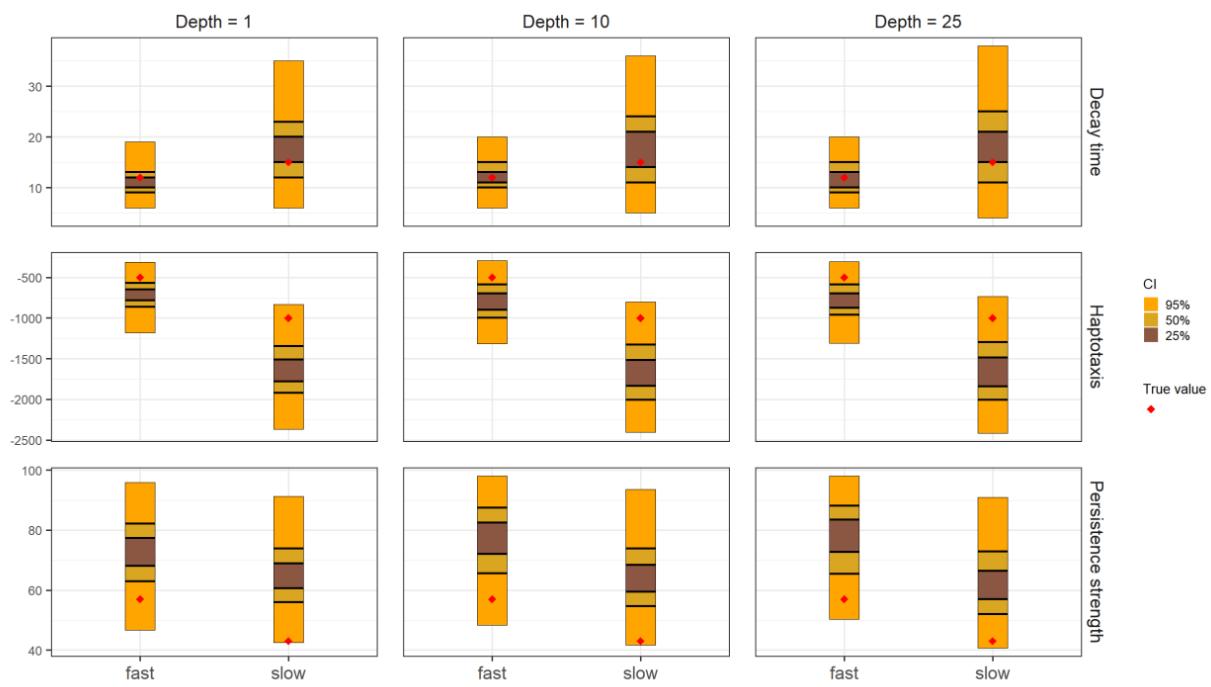


Figure S13 Credibility intervals for subsampling depths for Model 3. Subsampling depths $n = 1, 10, 25$ are depicted. The true parameter values are indicated in red. The credibility intervals (CI) 25%, 50% and 95% are shown and highlighted with different colours. The shown depths were selected from the depths [1, 5, 10, 25, 50] and are representative for the overall observed characteristics.

Contact analysis

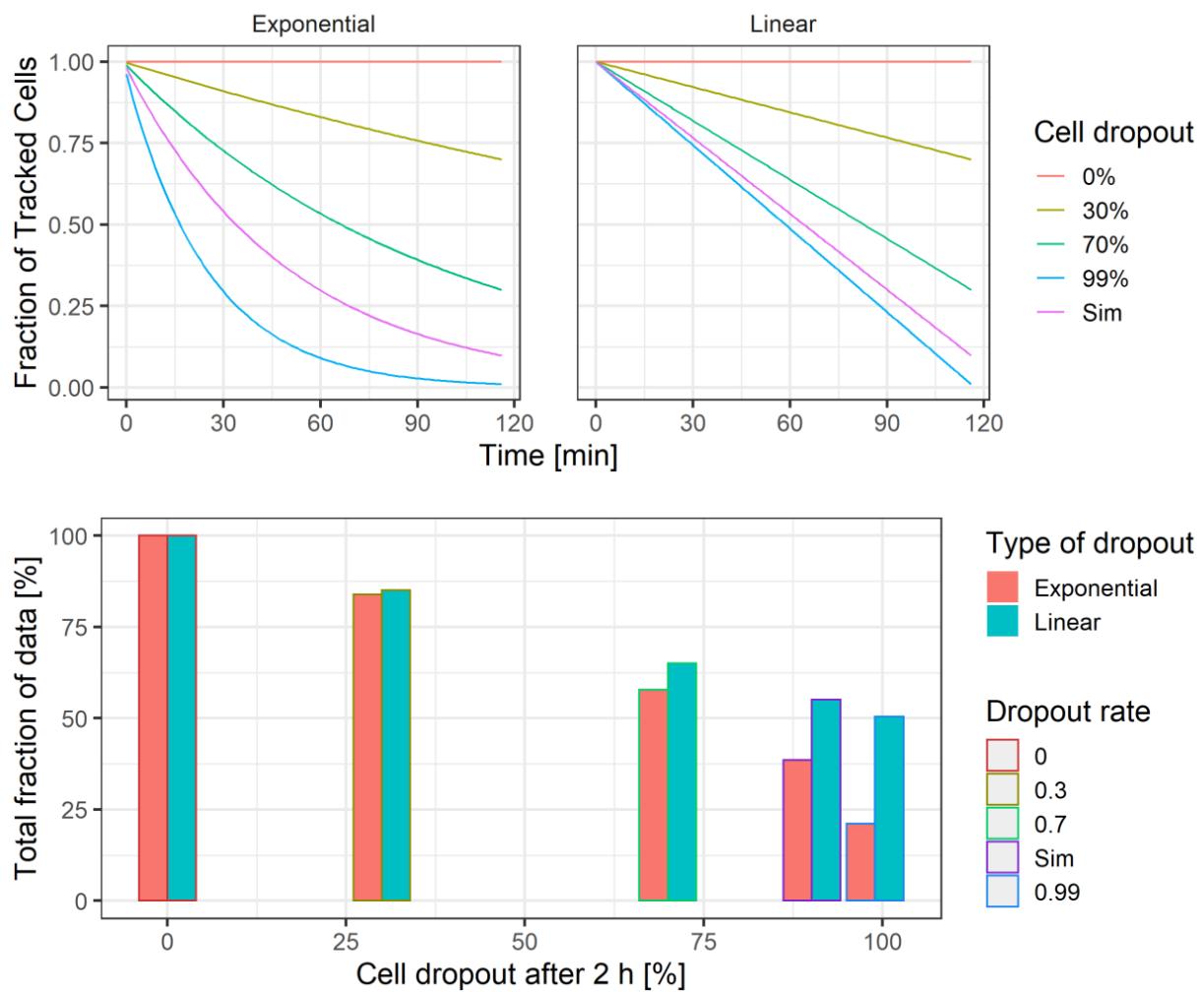


Figure S14 Comparison of different dropouts in an exponential and linear scenario. **Top:** The dropout rates 0%, 30%, 70%, 99% and the simulation dropout rate of 90.9% are shown and highlighted with different colours. An exponential dropout type as well as a linear dropout type were simulated. **Bottom:** Fractions of total dropout for different dropout rates and types in comparison. The fractions were obtained by dividing the area under the curves in the top plot by the respective area under the curve when no dropout is observed. The filling colour indicates the dropout type and the surrounding colour the dropout rate.

Tables

Table S1 Pseudo code for the sequential Monte Carlo Approximate Bayesian Computation algorithm. Taken from Toni et al. (2008). θ stands for the summary statistics (Toni et al., 2009).

```

S1 Initialize  $\epsilon_1, \dots, \epsilon_T$ .
      Set the population indicator  $t=0$ .
S2.0 Set the particle indicator  $i=1$ .
S2.1 If  $t=0$ , sample  $\theta^{**}$  independently from  $\pi(\theta)$ .
      Else, sample  $\theta^*$  from the previous population
           $\{\theta_{t-1}^{(i)}\}$  with weights  $w_{t-1}$  and perturb the
          particle to obtain  $\theta^{**} \sim K_t(\theta|\theta^*)$ , where  $K_t$  is a
          perturbation kernel.
      If  $\pi(\theta^{**})=0$ , return to S2.1.
      Simulate a candidate dataset  $x^* \sim f(x|\theta^{**})$ .
      If  $d(x^*, x_0) \geq \epsilon_t$ , return to S2.1.
S2.2 Set  $\theta_t^{(i)} = \theta^{**}$  and calculate the weight for
      particle  $\theta_t^{(i)}$ ,
      
$$w_t^{(i)} = \begin{cases} 1, & \text{if } t = 0, \\ \frac{\pi(\theta_t^{(i)})}{\sum_{j=1}^N w_{t-1}^{(j)} K_t(\theta_{t-1}^{(j)}, \theta_t^{(i)})}, & \text{if } t > 0. \end{cases}$$

      If  $i < N$ , set  $i = i + 1$ , go to S2.1.
S3 Normalize the weights.
      If  $t < T$ , set  $t = t + 1$ , go to S2.0.

```

Table S2 ϵ -lists used for the different models. For the 1 cell models (Model 1 and 2) the minimum was set to 1.5, for the 2-cell Model 3, to 3. An updated list based on the previous runs was generated for the comparison of the subsampling depth (Model 3 and 3.a).

Model	ϵ -list
Model 1, 2	[260, 212.4, 124.9, 67.9, 39.8, 26.5, 20.2, 15.7, 12.5, 9.7, 7.5, 5.9, 4.7, 3.6, 2.9, 1.9, 1.5]
Model 3	[520, 424, 250, 136, 80, 53, 40, 31, 25, 19, 15, 12, 9, 7, 6, 4, 3]
Model 3, 3.a (Subsampling depth)	[343.5, 123.9, 72, 49.5, 35.7, 27.6, 22, 16.3, 12.4, 9.7, 8, 6.5, 5.4, 4.6, 3.8, 3.3, 2.8]