

A statistical and mechanistic, model-based analysis of spindle assembly checkpoint signalling

Von der Fakultät Konstruktions-, Produktions- und Fahrzeugtechnik
und dem Stuttgart Research Centre for Simulation Technology
der Universität Stuttgart zur Erlangung der Würde eines
Doktor-Ingenieurs (Dr.-Ing.) genehmigte Abhandlung

Vorgelegt von

Eva-Maria Geissen

aus Mainz

Hauptberichter: Prof. Dr. rer. nat. Nicole Radde

Mitberichter: Assistant Professor Dr. Silke Hauf

Tag der mündlichen Prüfung: 05.05.2017

Institut für Systemtheorie und Regelungstechnik

Universität Stuttgart

2017

Acknowledgements

I would like to thank everyone who supported me during the years leading to the completion of this thesis, including many people not mentioned below.

First and foremost, I am grateful to my advisor Prof. Dr. Nicole Radde for the opportunity to perform my doctoral research in her group, for her constant support and her patience and for giving me the freedom to work in my own way.

I would like to give my most sincere thanks to my former colleague, co-author and mentor Dr. Jan Hasenauer. Without his brilliant ideas and contagious enthusiasm this thesis would not be this thesis.

I want to thank Prof. Dr. Silke Hauf and Dr. Stephanie Heinrich for an exceptional supportive and successful collaboration. In addition, I want to thank Prof. Dr. Silke Hauf for being part of my doctoral examination committee.

I want to thank Prof. Dr. Peter Swain for hosting me for three month in his group at the Centre for Synthetic and Systems Biology (SynthSys) in Edinburgh, UK and the members of his group for welcoming me so warmly.

I want to thank all colleagues at the IST for creating such an agreeable working environment. In particular, I want to thank Dr. Christian Breindl, Wolfgang Halter, Dirke Imig, Dr. Andrei Kramer, Dr. Gerd Simon Schmidt, Beate Spinner and Dr. Patrick Weber for moral support, helpful discussion and all all the things that I will remember with joy. Special thanks go to Wolfgang Halter, Dr. Gerd Simon Schmidt and Dr. Zoltan Tuza for proofreading this thesis.

Last but by no means least, I want to thank my parents Roswitha and Rolf Geissen for their constant and unconditional support.

Eva-Maria Geissen
Heidelberg, Mai 2017

Contents

Acknowledgements	iii
Index of notation	ix
Abstract	xiii
Deutsche Kurzfassung	xv
1. Introduction	1
1.1. Research motivation	1
1.2. Research topic overview	2
1.3. Contribution of this thesis	4
1.4. Outline of this thesis	6
2. Background	7
2.1. <i>Schizosaccharomyces pombe</i> - a model organism	7
2.2. The spindle assembly checkpoint	7
2.2.1. Sensing and signal initiation at kinetochores	8
2.2.2. Relaying the wait anaphase signal	9
2.3. Single-cell data	9
2.3.1. Aspects of non-genetic cell-to-cell variability	10
2.3.2. Time-to-event data	10
2.3.3. Data censoring	11
2.4. Likelihood based parameter estimation and uncertainty analysis in a nutshell .	11
2.4.1. Frequentist approach	11
2.4.2. Bayesian approach	13
3. Experimental data	15
3.1. Yeast strains	15
3.2. Prometaphase length data	15
4. Statistical models for the quantification and analysis of cellular SAC phenotypes	19
4.1. Introduction	19
4.1.1. Censored data	20
4.1.2. Sources of censoring in biological data	21
4.1.3. Mixture models in the analysis of censored single-cell data	23
4.2. Problem formulation	25

4.3.	Multi-experiment mixture modelling of censored single-cell data (MEMO)	26
4.3.1.	Statistical models in the presence of censoring	26
4.3.2.	Formulation of model hypotheses	31
4.3.3.	Model parametrization	32
4.3.4.	Hypotheses testing via model selection	32
4.3.5.	Interpretation and further analysis	33
4.3.6.	Assessment of the impact of censoring on inference in SAC datasets	34
4.4.	Mixture model based analysis of variability in SAC single-cell microscopy data	34
4.4.1.	Inference of subpopulation structure from prometaphase data	38
4.4.2.	Quantification of the sensitivity of SAC functionality to perturbations	43
4.4.3.	Data based hypothesis testing on the mode of interaction of Mad2 and Mad3 in SAC signalling	44
4.5.	Summary and Discussion	48
5.	Mechanistic models of SAC signalling in heterogeneous populations	53
5.1.	Introduction	53
5.1.1.	Dynamic models of protein reaction networks	54
5.1.2.	Dynamics with uncertainty: random differential equations	56
5.1.3.	Propagation of uncertainty: Sigma-point based methods	57
5.1.4.	Ultrasensitivity in steady state input-output responses	59
5.1.5.	Model-based studies of SAC signalling	63
5.2.	Problem formulation	64
5.3.	A general framework for mechanistic models of populations with dual SAC phenotypes	65
5.3.1.	Formulation of a population model with phenotypic output	65
5.3.2.	Model simulation	67
5.3.3.	Model calibration from experimental data	69
5.3.4.	Assessment of model fit	70
5.4.	M1: A model for Slp1 inhibition via MCC formation	70
5.4.1.	M1: Model formulation	71
5.4.2.	M1: Model calibration	72
5.4.3.	M1: Assessment of model fit	73
5.4.4.	M1: Simulation based model analysis	75
5.4.5.	M1: Bayesian uncertainty analysis	77
5.4.6.	M1: Analysis of steady state input output response	79
5.5.	M2: A model for MCC conveyed APC inhibition	84
5.5.1.	M2: Model formulation	84
5.5.2.	M2: Model calibration	86
5.5.3.	M2: Assessment of model fit	88
5.5.4.	M2: Simulation based model analysis	88
5.5.5.	M2: Analysis of steady state input output response	91
5.6.	Summary and Discussion	94
6.	Conclusion	97
6.1.	Summary and conclusions	97

6.2. Outlook	98
Appendix	101
A. Derivation of the pdf and the cdf for the conditional random variable $X X \leq C$	101
B. Probability densities	102
C. Derivation of the gradient of the objective function	103
C.1. Gradient of the log-likelihood function	103
C.2. Gradient of the mixture distribution	105
C.3. Gradient of the probability densities	105
D. Approximation of $[Slp1]_{s_s}$ in Model M1 by Taylor series expansions	106
E. Derivation of parameter boundaries for parameter estimation in Chapter 5	108
Bibliography	109

Index of notation

Acronyms

Acronym	Description
AIC	Akaike information criterion
APC/C	anaphase promoting complex/cyclosome
Bub1	budding uninhibited by benzimidazole 1
Bub3	budding uninhibited by benzimidazole 3
BubR1	budding uninhibited by benzimidazole related 1
Cdc20	cell division cycle 20 homolog
cdf	cumulative probability function
CMad2	closed Mad2
CV	coefficient of variation
DD2	divided differences filter of second order
DNA	deoxyribonucleic acid
DRAM	delayed rejection adaptive Metropolis
FLAME	flow analysis with automated multivariate estimation
GFP	green fluorescence protein
inh	inhibitor
JCM	Joint Clustering and Matching
LOD	limit of detection
Mad1	mitotic arrest deficient 1
Mad2	mitotic arrest deficient 2
Mad3	mitotic arrest deficient 3
MAP	maximum a posteriori
MCC	mitotic checkpoint complex
MCMC	Markov chain Monte-Carlo
MEMO	multi experiment mixture modelling
MJP	Markov jump process
ML	maximum likelihood
MLE	maximum likelihood estimate
Mph1	Mps1p-like pombe homolog
Mps1	monopolar spindle 1
ODE	ordinary differential equation
ODE-MM	ODE constrained mixture modelling
OMad2	open Mad2
PCR	polymerase chain reaction

Acronyms - continued

Acronym	Description
PDE	partial differential equation
pdf	probability density function
PESTO	Parameter ESTimation TOolbox
Plo1	polo-related kinase 1
qPCR	quantitative polymerase chain reaction
RNA	ribonucleic acid
RODE	random ordinary differential equations
RRE	reaction rate equation
SAC	spindle assembly checkpoint
Slp1	sleepy homolog
SPB	spindle pole bodies
UT	unscented transform
WT	wild type

Notation

General notation: Throughout this thesis, we follow the common statistical notation and denote random variables with capital letters, such as X . To describe the corresponding probability (density) of the random variable, in the general case, lower-case letters are used. According to this $p(x)$ denotes the probability density of X . However, for the derivation of the statistical models we use the notation $f_X(x)$ to denote the probability density function of X and $F_X(x)$ for the corresponding cumulative distribution function.

Protein complexes are denoted by concatenating the involved proteins, separated by a colon.

Symbols

Symbol	Description
c_j	stoichiometric coefficient
\mathcal{D}	collection of all data
\mathcal{D}_i	data of experiment i
d_x	dimension of vector x
$\mathbb{E}[X]$	expected value of X
$J(x)$	Jacobian matrix
$\mathcal{J}(\theta)$	negative log-likelihood
$\mathcal{M}(\theta)$	model given parameters θ
n_x	dimension of vector x
n_H	Hill-coefficient
n_{eff}	effective Hill-coefficient
$\mathbb{P}(\mathcal{D})$	marginal probability to observe measurement data
$\mathbb{P}(\mathcal{D} \theta)$	likelihood of \mathcal{D} given θ

Symbols - continued

Symbol	Description
$\phi(x \varphi_s(u))$	probability density of a mixture component
$\varphi_s(u)$	subpopulation distribution parameters in general dependence on input u
$\pi(\theta)$	prior probability of θ
$\pi(\theta \mathcal{D})$	posterior probability of θ given \mathcal{D}
$\text{PL}(\theta_i)$	profile likelihood
\mathbb{R}	real numbers
\mathbb{R}_+	non-negative real numbers
R_j	j th chemical reaction
S_{local}	local sensitivity
S_j	j th chemical species
\mathcal{S}	MCMC sample
θ	vector of model parameters
$\tilde{\theta}$	meta-parameter in a mixture model
θ^{MAP}	maximum a posteriori estimate
θ^{ML}	maximum likelihood estimate
\mathbf{V}_X	covariance matrix of X
v_j	reaction rate of j th reaction
$\tilde{v}(x, \theta)$	rate law
$w_s(u)$	relative subpopulation size
\mathcal{X}_i	i th sigma point
\mathcal{Y}_i	transformed i th sigma point
y_i^j	uncensored observation
\bar{y}_i^j	right censored observation
$\underline{\bar{y}}_i^j$	interval censored observation

Abstract

The mechanisms that ascertain whether a phase of the cell cycle has been successfully completed and the conditions to proceed to the next phase are fulfilled are called checkpoints. One of them is the spindle assembly checkpoint (SAC), which clears for completion of cell division only if the conditions for a proper partitioning of the genetic material are fulfilled. Despite complete knowledge of its function for decades, the underlying mechanism on the molecular level is still not completely elucidated.

We have data at hand that show how persistent the SAC is in individual yeast cells, when the amounts of its signalling components are altered. Since these manipulations are done on the genetic level, the efficacy is the same for each cell of a strain. Therefore, one would expect the SAC to show a homogeneous response in such a clonal population of cells. However, the data reveal that SAC persistence, measured as duration of cell cycle arrest in prometaphase, is highly variable between cells of the same strain.

In this thesis we use statistical modelling to quantify the observed cell-to-cell variability and analyse subpopulation structures in clonal populations of yeast cells. The sophisticated statistical analysis is complemented by mechanistic modelling of the molecular mechanism of the SAC on the population level.

The statistical analysis of the data is hampered by the fact that the data are censored, i.e. that prometaphase length as the variable of interest is not completely observable in many cells. To account for this in the analysis and to exploit the information which is only accessible by simultaneously analysing the data from multiple stains, we propose a general framework for multi-experiment mixture modelling, named MEMO. Employing this framework, we show that reduction of the amount of individual SAC proteins results in a split of the clonal population of cells into subpopulations with opposing SAC phenotypes. While one subpopulation retains a completely functional SAC, a second subpopulation with an impaired SAC emerges and increases. We quantify the sensitivity of this effect as a function of type and amount of the manipulated protein. Such a quantification allows for the prediction of the subpopulation structure of yet unobserved protein manipulations.

The striking observation of phenotypically different subpopulations in a population of genetically identical cells is underscored by the fact that noise in the protein abundances is small. We complement the statistical analysis of the data with mechanistic models of the molecular mechanism of SAC signalling. By exploiting the information contained in the population split, we identify ultrasensitivity and potential bistability to be a property of the dynamical system that forms the SAC. This implies high sensitivity with respect to noise in the abundance of signalling and targeted proteins. Furthermore, we assess the contribution of different SAC components to the observed cell-to-cell variability.

While the statistical modelling framework proposed in this thesis can help to prevent misinterpretation of data in the presence of censoring, also in other single-cell data settings, our findings on the properties of the SAC signalling system provide novel insights into this intri-

cate molecular mechanism.

Deutsche Kurzfassung

Motivation

Die fundamentale Einheit aller bekannten lebenden Organismen ist die Zelle. Sowohl einzellige Lebewesen wie die Hefe, als auch das komplexe Gefüge des menschlichen Körpers, entstehen aus einer einzelnen Zelle durch wiederholte Teilung in einem Prozess der als Zellzyklus bezeichnet wird. Wie Rudolf Virchow schon 1855 feststellte: „omnis cellula e cellula“ – jede Zelle stammt von einer Zelle. Während jeder Zellteilung entstehen zwei Tochterzellen aus einer Mutterzelle. Für den Fortbestand des Organismus ist es von größter Wichtigkeit, dass das vor der Teilung verdoppelte genetische Material korrekt auf die beiden Tochterzellen verteilt wird. Jede Zelle muss die korrekte Anzahl und den korrekten Satz an Chromosomen bekommen. Um dies sicherzustellen, haben eukaryotische Organismen einen Überwachungsmechanismus entwickelt, der das Eintreten der Bedingungen überwacht, die nötig sind um eine korrekte Aufteilung zu gewährleisten, und alle weiteren Schritte im Zellzyklus blockiert, bis diese Bedingungen erfüllt sind.

Diese Aufgabe nimmt der Spindel-Assemblierungs-Checkpoint (SAC) wahr. Der SAC ist ein intrazellulärer biochemischer Signalmechanismus, der aus Interaktionen zwischen Proteinen aufgebaut ist. Fehlfunktionen in diesem Mechanismus können zu einer inkorrekten Verteilung der Chromosomen führen und damit zu einer abnormalen Anzahl von Chromosomen, ein Zustand der als Aneuploidie bezeichnet wird. Aneuploidie wird mit der Entstehung von Tumoren in Verbindung gebracht. Andererseits ist ein funktionsfähiger SAC ein potentieller Angriffspunkt für die Krebstherapie, da seine Aktivierung die Zellteilung verhindern kann. Einsichten in die Funktionsweise des SAC und seine Potentiale und Schwachpunkte sind daher erstrebenswert und könnten dabei helfen potentielle Angriffspunkte für eine gezielte Beeinflussung des SAC zu identifizieren. Trotz umfassendem Wissen über die molekularen Komponenten des SAC, sind jedoch das Zusammenspiel dieser Komponenten beim Aufbau und der Lösung des SAC induzierten Zellzyklusarrests, und vor allem die quantitativen Eigenschaften der Dynamik dieser Prozesse, noch nicht vollständig verstanden.

Die Systembiologie kombiniert Methoden der mathematischen Modellierung und computergesützte Verfahren mit experimentellen Daten, um ein konzeptionelles und quantitatives Verständnis biologischer Systeme zu erlangen. Dies ermöglicht die Vorhersage und genaue Simulation von komplexem biologischen Verhalten und dessen Dynamik. Daher eignen sich systembiologische Ansätze wenn die Dynamik einer Größe oder andere Eigenschaften nicht experimentell messbar sind. Zusätzlich können Modelle die Analyse vorhandener biologischer Daten unterstützen und die so gewonnenen Erkenntnisse können die Grundlage für mechanistische Modelle des zugrunde liegenden molekularen Prozesses bilden. Um den SAC Mechanismus in seiner Gesamtheit zu verstehen, ist es erfolgsversprechend traditionelle Forschungsmethoden der Biologie mit modellbasierten Ansätzen zu kombinieren, um mit verein-

ten Kräften den molekularen Mechanismus des SAC zu ergründen.

Durch eine Forschungskollaboration mit dem Labor von Frau Dr. Silke Hauf, damals am Friedrich Miescher Laboratorium der Max Planck Gesellschaft in Tübingen, stehen uns vielversprechende Daten zur SAC Funktionalität in Spaltheffen zur Verfügung. Die Spaltheffe *S. pombe* ist ein gängiger Modellorganismus für humane Zellen, da die am SAC beteiligten Strukturen dieser Hefe denen von humanen Zellen sehr ähnlich sind. Im Vergleich zu humanen Zellen sind Hefen jedoch leichter experimentell manipulierbar. In einer sorgfältig durchdachten Serie von Experimenten untersuchte Frau Dr. Stephanie Heinrich, damals in der Gruppe von Frau Dr. Hauf, auf der Ebene einzelner Zellen die Robustheit des SAC gegenüber Abweichungen in den Mengen der am Signal beteiligten Proteine. Diese Einzelzelldaten zeigen eine erhebliche Menge an nichtgenetischer Variabilität der SAC Funktionalität, obwohl experimentell ungewöhnlich geringe Schwankungen in den Mengen der am Signal beteiligten Proteine nachgewiesen werden konnten. Aufgrund der begrenzten Dauer und Frequenz der Messungen unterliegen diese Daten unterschiedlichen Arten von Zensierung, was eine Quantifizierung der Variabilität erheblich erschwert. Für eine zuverlässige Interpretation der Daten ist eine verlässliche Analyse jedoch unabdingbar. Bei diesem Problem haben sich modellbasierte statistische Verfahren als sehr nützlich erwiesen.

Im Folgenden werden wir Eigenschaften des SAC Mechanismus anhand statistischer und mechanistischer Modelle und den experimentellen Daten von Frau Dr. Heinrich erforschen. Für die Quantifizierung und Analyse der nichtgenetischen Variabilität werden wir ein statistisches Framework für die gleichzeitige Modellierung und Analyse zensierter Einzelzelldaten aus verschiedenen experimentellen Bedingungen vorstellen. Dieses Framework werden wir einsetzen, um die Subpopulationsstrukturen in den Daten und die Sensitivität des SAC gegenüber Abweichungen in den Mengen seiner Proteinkomponenten zu quantifizieren. Die Ergebnisse dieser statistischen Analyse werden dann eingesetzt werden, um anhand von mechanistischen Modellen Einblicke in die Eigenschaften und den Mechanismus des SAC zu ermöglichen und die Ursachen der beobachteten Variabilität zu ergründen.

Einführung in das Thema

Im Folgenden geben wir eine kurze Einführung in die drei zentralen Bestandteile dieser Arbeit, in der wir uns mit der Erforschung des (i) SAC mit Hilfe von (ii) statistischen und (iii) mechanistischen Modellen beschäftigen.

Der Spindel-Assemblierungs-Checkpoint

Der SAC ist ein Kontrollmechanismus der die korrekte Verteilung der Chromosomen bei der Zellteilung fördert. Damit spielt er eine wichtige Rolle bei der Erhaltung der Intaktheit des Genoms. Das Funktionsprinzip des SAC ist simpel: Chromosomen, die noch nicht korrekt an die mitotische Spindel angeheftet wurden, aktivieren den SAC der wiederum die Maschinerie des Zellzyklus blockiert, um ein Fortschreiten der Teilung zu verhindern. Durch die Aktivität des SAC wird die Zelle in der Prometaphase der Mitose, der Teilungsphase des Zellzyklus, angehalten. Die essentiell am Signalfluss beteiligten Proteine sind Mad1, Mad2, Bub1, Bub3 und BubR1 (Mad3 in *S. pombe*) zusammen mit der Proteinkinase Mps1 (Mph1 in *S. pombe*).

Die Verbindung zum Zellzyklus wird über das Protein Cdc20 (Slp1 in *S. pombe*) hergestellt. Der Eingriffspunkt des SAC im Zellzykluskontrollsystem ist der APC/C, ein Proteinkomplex der in seiner durch Cdc20 aktivierten Form die Teilung der Schwesterchromatiden und den Übergang zur Anaphase ermöglicht.

Das ultimative Ziel des SAC ist es eine vorzeitige Aktivierung des APC/C zu verhindern. Ein Großteil seiner Proteinkomponenten konzentriert sich dazu an unangehefteten Chromatiden und generiert dort zusammen ein Signal, das ins Zytoplasma diffundiert und den APC/C inhibiert. Dies geschieht zum einen durch die Sequestrierung seines Aktivators Cdc20 und zum anderen durch eine direkte Inhibition des APC/C. Der genaue Aufbau und die Dynamik dieses zellzyklusinhibierenden Signals sind jedoch auch ein Vierteljahrhundert nach der Identifizierung der molekularen Komponenten des SAC noch immer nicht vollständig aufgeklärt. In den vergangenen zehn Jahren wurden deshalb vermehrt auch modellbasierte Methoden eingesetzt, um die klassische biologische Forschung zu unterstützen. Die Bedeutung dieser Zusammenarbeit wurde auch von biologischer Seite betont. Seitdem wurden verschiedene Modelle zur Beschreibung der Vorgänge entwickelt, eine systemische Sichtweise des SAC konnte jedoch bisher nicht etabliert werden.

Modellbasierte statistische Analyse von heterogenen Einzelzelldaten

Einzelzelldaten sind immer häufiger verfügbar und zeigen, dass einzelne Zellen unterschiedlich auf denselben Stimulus oder dieselbe Behandlung reagieren. Damit führen Einzelzelldaten immer zu einer Verteilung in den gemessenen Werten. Diese Heterogenität, oder Zell-Zell-Variabilität, enthält wertvolle zusätzliche Informationen über den zugrundeliegenden molekularen Prozess. Sie kann sowohl quantitativ als auch qualitativ sein, wobei qualitativ in diesem Zusammenhang im Sinne von Subpopulationen mit qualitativ unterschiedlichen Eigenschaften zu verstehen ist. Das Vorhandensein von solchen Subpopulationen kann zu multimodalen und schiefen Verteilungen der Daten führen. Um solche Verteilungen richtig zu interpretieren und das Potential von Einzelzelldaten voll auszuschöpfen, müssen sie statistisch quantifiziert und analysiert werden.

Für diesen Zweck eignet sich das von Ronald A. Fisher (1890-1962), einem britischen Biologen und Statistiker, entwickelte Prinzip der modellbasierten statistischen Deduktion. Fishers Ansatz basiert auf einem statistischen Modell, das die Vorstellung von der Generierung der Daten beschreibt. Damit bildet das statistische Modell eine Verbindung zwischen den Daten und der Population aus der diese Beobachtungen stammen. Dies ermöglicht die Analyse der nicht beobachtbaren vollständigen Population anhand der durch die Daten gebildeten Stichprobe. Dem statistischen Modell liegt die Auffassung der gemessenen Werte als Realisierungen einer Zufallsvariable zugrunde. Diese Zufallsvariable muss über eine parametrische Verteilung spezifiziert werden.

Auf Grundlage dieses Prinzips wird in dieser Arbeit ein Modellierungsframework entwickelt, das in der Lage ist die Variabilität in zensierten Einzelzelldaten zu quantifizieren und zu analysieren, und dabei alle Datensätze gleichzeitig zu berücksichtigen. Unsere Daten bestehen aus den Prometaphaselängen der Zellen, die ein Maß für die Dauer des SAC induzierten Arrests in den Zellen sind. Eine Herausforderung bei der Analyse dieser Einzelzelldaten ist Zensierung. Zensierung bedeutet, dass aufgrund von Eigenschaften der Messmethode Beobachtungen teilweise nicht exakt bestimmbar sind, sondern lediglich Angaben über Grenzen

gemacht werden können. Wir zeigen, wie der probabilistische Hintergrund modellbasierter Verfahren hilft dieses Problem zu lösen, und dass die Anwendung des von uns entwickelten Frameworks Erkenntnisse ermöglicht die über eine reine Quantifizierung der Variabilität hinausreichen. Die mit Hilfe der statistischen Modelle erlangten Erkenntnisse bilden im zweiten Teil der Arbeit die Datenbasis für mechanistische Modelle der molekularen Vorgänge im Signalmechanismus des SAC.

Mechanistische Modellierung intrazellulärer Signalübertragung

Signalübertragungspfade erlauben es Zellen Signale aus der Umwelt sowie dem Zellinneren, zu übermitteln und zu prozessieren. Dabei detektieren Sensoren die Signale, die dann weitergeleitet und ausgewertet werden, um schließlich Aktuatoren zu aktivieren. Deren Aktivität führt zur Änderung des Zustands der Zelle und ihrer biochemischen Zusammensetzung. In Zellen werden Signale vor allem über die Menge und Aktivitäten von Proteinen und Proteinkomplexen codiert. Diese Größen werden durch verschiedene Prozesse, wie Synthese und Abbau, die Bildung von Proteinkomplexen und Phosphorylierung moduliert. Komplexe Zusammenhänge mit Rückkopplungen und Mitkopplungen sind dabei schwer zu interpretieren und nachzuvollziehen. Um das komplexe Verhalten von Signalpfaden mit ihren zahlreichen Komponenten und Interaktionen zu verstehen, können Methoden der mathematischen Modellierung eingesetzt werden. Modelle ermöglichen eine systemische Sicht der komplexen Vorgänge und die Systemtheorie dynamischer Systeme bietet wertvolle Analysemethoden. Die etablierten Modelle reichen von kleineren Modulen wiederholt auftretender Motive bis zu großen Netzwerken die vom Sensor bis zum Aktuator alle Schritte beinhalten. Dabei existieren unterschiedliche Arten von Modellen: Mechanistische Modelle basieren auf chemischer Reaktionskinetik, während phänomenologische Modelle die Interaktion zwischen den Signalmolekülen vorwiegend qualitativ beschreiben. Beide Modelltypen können sowohl auf stochastischen als auch deterministischen Regeln basieren.

Der SAC ist ein solcher intrazellulärer Signalübertragungsmechanismus. Die beteiligten Proteine bilden miteinander zahlreiche Komplexe und die aktivierenden und inhibierenden Interaktionen sind vielfältig. Wir werden mechanistische Modelle des SAC, basierend auf deterministischen Reaktionskinetiken, nutzen und diese um Variabilität in den Modellkomponenten erweitern, um die Vorgänge in heterogenen Populationen abzubilden.

Forschungsbeiträge und Gliederung der Arbeit

In dieser Arbeit wird der SAC mit zwei komplementären modellbasierten Ansätzen untersucht. Zum einen nutzen wir statistische Modelle, um die vorhandenen experimentellen Daten zu analysieren, und Subpopulationsstrukturen zu identifizieren und zu quantifizieren. Dies erlaubt uns die Sensitivität des SAC bezüglich der Mengen an signalübertragenden Proteinen zu quantifizieren und Aussagen darüber zu machen, welche Rolle einzelne Proteine bei diesem Verhalten spielen. Für die beschriebenen Analysen entwickeln wir ein Rahmenkonzept für die statistische Modellierung zensierter Einzelzelldaten, das in der Lage ist, gleichzeitig die Daten unterschiedlicher experimenteller Bedingungen zu berücksichtigen. Wir bezeichnen dieses Konzept und seine Implementierung als MATLAB Toolbox als MEMO (Multi-Experiment

mixture MOdelling). MEMO ist auf unterschiedliche Arten von Einzelzelldaten anwendbar und trägt dazu bei, die Fehlinterpretation zensierter Daten zu verhindern.

Zum anderen entwickeln wir mechanistische Modelle des biochemischen Signals des SAC in heterogenen Populationen, um Einblicke in die Systemeigenschaften des SAC zu erhalten. Mit Hilfe dieser Modelle ist es uns möglich, Hinweise auf die Ursachen der beobachteten Variabilität im Phänotyp und das Ausmaß der Variabilität auf molekularer Ebene zu generieren. Der Inhalt der Arbeit ist wie folgt gegliedert:

Kapitel 1 - Einleitung In diesem Kapitel umreißen wir die zentralen Bestandteile dieser Arbeit. Nach einer kurzen Einführung in die Funktion und Geschichte des SAC gehen wir auf den Einsatz statistischer Modelle in der Quantifizierung von Einzelzelldaten ein. Wir erörtern ihre speziellen Eigenschaften die eine Nutzung in diesem Zusammenhang ermöglichen. Des Weiteren geben wir eine kurze Einführung in die intrazelluläre Signalübertragung und die Modellierung von intrazellulären Signalübertragungsmechanismen.

Kapitel 2 - Grundlagen In diesem Kapitel beleuchten wir die biologischen und methodischen Grundlagen dieser Arbeit. Wir stellen die Spaltheefe *Schizosaccharomyces pombe* als Modellorganismus vor und fassen zusammen, was aktuell über den SAC auf molekularer Ebene bekannt ist. Außerdem erörtern wir verschiedene Aspekte im Zusammenhang mit Einzelzelldaten wie Zell-Zell-Variabilität, die Besonderheiten von Ereigniszeitdaten und Zensierung. Als Grundlage für die Kapitel 4 und 5 führen wir außerdem die Grundlagen Likelihood-basierter Parameterschätzverfahren ein.

Kapitel 3 - Experimentelle Daten In diesem Kapitel stellen wir die experimentellen Daten vor auf denen diese Arbeit basiert. Sie wurden uns von Frau Dr. Stephanie Heinrich zur Verfügung gestellt und sind im Rahmen ihrer Doktorarbeit im Labor von Silke Hauf am Friedrich Miescher Laboratorium der Max Planck Gesellschaft in Tübingen entstanden. Es handelt sich dabei um Einzelzelldaten aus Stämmen der Spaltheefe *S. pombe*, in denen die Expression von Schlüsselproteinen im SAC Signalpfad auf Promotorebene manipuliert wurde. Für einzelne Zellen wurde unter SAC aktivierenden Bedingungen die Zeitspanne gemessen für die der SAC die Teilung der Zelle verhinderte. Da Beobachtungszeitraum und Beobachtungsintervall zeitlich begrenzt sind tritt bei diesen Daten Intervall- und Rechtszensierung auf. Dies erschwert die Analyse und macht den Einsatz statistischer Modelle erforderlich.

Kapitel 4 - Statistische Modelle für die Quantifizierung und Analyse zellulärer SAC Phänotypen In Kapitel 4 entwickeln wir ein allgemeines Framework für die Quantifizierung und Analyse von zensierten Einzelzelldaten. Das Framework basiert auf Mixturmodellen, statistischen Modellen die durch die gewichtete Summation parametrischer Verteilungen multimodale Verteilung in Daten beschreiben können. Außerdem leiten wir die Verteilungen her, die sich durch den Einfluss von Datenzensierung ergeben um damit die Daten statistisch korrekt zu beschreiben. Um die Parameter der Verteilungen anhand der Daten zu schätzen, verwenden wir die Maximum-Likelihood-Methode, ein Schätzverfahren das durch Maximierung der gemeinsamen Wahrscheinlichkeit aller Datenpunkte unter einem Modell die plausibelsten Parameter bestimmt. Diese Methode erlaubt die gleichzeitige Berücksichtigung der

Information aus mehreren experimentellen Bedingungen durch eine gemeinsame Likelihood. Die probabilistischen Eigenschaften von Mixturmodellen erlauben außerdem den Vergleich verschiedener Modellhypothesen, und damit die Detektion von Subpopulationen in den Daten.

Im zweiten Teil des Kapitels quantifizieren und analysieren wir die Variabilität in den experimentellen Daten zur SAC Funktionalität. Wir zeigen, dass in Stämmen mit reduzierter Proteinmenge bimodale Verteilungen in den Einzelzelldaten auftreten, die durch die Präsenz von zwei Subpopulationen mit qualitativ unterschiedlichen Eigenschaften in Bezug auf die SAC Funktionalität verursacht werden. Unsere Analyse belegt, dass dabei in einer Subpopulation die Eigenschaften des Wildtyps erhalten sind. Diese Subpopulation verfügt also über einen voll funktionsfähigen SAC, wohingegen der SAC in der zweiten Population defekt ist. Des Weiteren quantifizieren wir die Sensitivität der SAC Funktionalität als Funktion der relativen Menge der an der Signalübertragung beteiligten Proteine. Die Ergebnisse der Quantifizierung dieser Sensitivität werden außerdem dazu verwendet, Hypothesen zur qualitativen Interaktion der Proteine in der Signalübertragung zu testen. Dazu werden Modelle, die diese unterschiedlichen Hypothesen beschreiben, anhand ihrer Vorhersagekraft bezüglich zusätzlicher experimenteller Daten bewertet.

Kapitel 5 - Mechanistische Modelle der SAC Signalübertragung in heterogenen Populationen In diesem Kapitel ergründen wir die Ursachen des Auftretens von Subpopulationen in den Populationen der genetisch identischen Hefezellen. Dazu stellen wir zuerst einen allgemeinen Modellierungsansatz vor, der darauf zugeschnitten ist Zell-Zell-Variabilität in einem deterministischen Modell des Signalmechanismus abzubilden und phänotypische Daten in das Modell zu integrieren. Dazu modellieren wir deterministische Variabilität innerhalb der Population als unimodale log-normale Verteilungen in verschiedenen Modellgrößen. Um im Modell zwischen den beiden unterschiedlichen Phänotypen des SAC zu unterscheiden, führen wir einen Schwellwert in der Konzentration der Modellgröße ein, die als Indikator für die Stärke des SAC Signals interpretiert werden kann.

Im zweiten Teil des Kapitels untersuchen wir zwei Modelle, die den molekularen Mechanismus des SAC in unterschiedlichem Umfang abbilden. Während das erste Modell sich auf die Beschreibung der Inhibition von Slp1/Cdc20 durch Sequestrierung im MCC, einem Multiproteinkomplex, beschränkt, erweitert das zweite Modell diesen Mechanismus um die inhibitorische Aktivität des SAC auf den Zellzyklus. Beide Modelle werden anhand der phänotypischen Subpopulationsanteile aus acht unterschiedlichen Hefestämmen kalibriert und zeigen in einer simulationsbasierten Analyse die Aufteilung in die beiden Phänotypen. Wir analysieren das stationäre Verhalten beider Modelle und zeigen, dass Ultrasensitivität bzw. Bistabilität der SAC Funktionalität bezüglich der Syntheserate in Verbindung mit minimaler deterministischer Variabilität der Zellen die Ursachen des gezeigten Modellverhaltens sind. Außerdem präsentieren wir modellgenerierte Hinweise darauf, dass dabei die Schwankung in der Synthese von Slp1 der bestimmende Einflussfaktor ist.

Kapitel 6 - Fazit In Kapitel 6 werden die Resultate der Arbeit zusammengefasst und offene Probleme und Ideen diskutiert.

1. Introduction

1.1. Research motivation

The fundamental unit of every known living organism is the cell. Unicellular organisms, such as yeast, as well as the complex entity of the human body, originate from a single cell by cell division in a process called the cell cycle. As Rudolf Virchow noted in 1855: “*omnis cellula e cellula*” - all cells come from cells. In each cell division, two daughter cells emerge from a mother cell. For the persistence of the organism it is of great importance that the previously duplicated genetic material is divided correctly between the daughter cells. Every cell has to receive the correct amount and the correct set of chromosomes. To ensure this, eukaryotic organisms have developed a surveillance mechanism that monitors the conditions needed to ensure a proper segregation of the chromosomes and blocks further actions until these conditions are fulfilled.

This task is executed by the spindle assembly checkpoint (SAC), an intracellular biochemical signalling mechanism constituted by interactions between proteins. Malfunctions in this mechanism can result in an incorrect distribution of chromosomes causing aneuploidy, the presence of an abnormal number of chromosomes in a cell. Aneuploidy is associated with tumorigenesis. On the other hand, a functional SAC is a potential therapeutic target in cancer treatment, since its activation can stop cells from dividing. Therefore, it is important to elucidate and quantify its strengths and weaknesses to reveal potential targets and strategies for interventions. Unfortunately, despite extensive knowledge on the molecular components of the SAC signalling machinery, the integrated action of the components in establishing and relieving the SAC mediated cell cycle arrest, and especially the quantitative dynamics of these processes, are not entirely understood. Signal generation is known to be sensitive, in terms of detecting small stimuli, and highly dynamic, but a quantitative systems view of the SAC is still lacking.

Systems biology combines mathematical modelling and computational methods with experimental data to develop a conceptual as well as quantitative understanding of biological systems, permitting prediction and accurate simulation of complex biological behaviours and their dynamics. Therefore, systems biology promises useful if direct measurements of dynamics, or other properties of interest, are not possible. In addition, models aid in the analysis of the data that is available and analysed data can form the basis for models of the underlying biochemical mechanism. To understand SAC signalling on the systems level, traditional biological research needs to be combined with model-based research to tackle the secrets of SAC signalling with joined forces.

A research collaboration with the lab of Dr. Silke Hauf, at that time located at the Friedrich Miescher Laboratory of the Max Planck Society in Tübingen, provided us with excellent data from fission yeast. Fission yeast is a common model organism for human cells. Its structures

involved in SAC signalling resemble the equivalent human ones quite well, while the manipulation of yeast cells is much easier to accomplish. In an elaborate series of experiments, Dr. Stephanie Heinrich experimentally probed the robustness of SAC functionality towards changes in some of the key signalling proteins. These single-cell data show a substantial amount of non-genetic variability in SAC functionality, although the noise in the amount of SAC proteins was shown to be unusually low in these cells. Due to limitations in the measurement process, these data on SAC functionality are subject to censoring. Censoring makes the quantification of single-cell population data non-trivial. However, for a thorough interpretation of the data they have to be carefully analysed. Fortunately, model-based statistical approaches promise useful in this situation.

In the following we investigate properties of the SAC signalling mechanism, by employing statistical and mechanistic models, together with the experimental data of our collaboration partners. For this, we derive a general framework for the simultaneous analysis of censored single-cell data from multiple experimental conditions. We use this framework to quantitatively characterise subpopulation structures in the data and the sensitivity of the SAC with respect to changes in the abundance of two of its key protein components. Furthermore, two mechanistic models that describe the molecular mechanism of SAC signalling to a different extent are employed to assess the sources of the observed sensitivity and predict sources of noise that challenge the SAC.

1.2. Research topic overview

This section introduces the three components constituting this thesis, in which we investigate the (i) intracellular signalling pathway known as spindle assembly checkpoint with the help of (ii) statistical models of its functionality and (iii) mechanistic models of the molecular interactions constituting its mechanism.

The Spindle Assembly Checkpoint

The SAC is a surveillance mechanism promoting proper chromosome segregation in dividing cells. In this respect, it has a crucial role in maintaining genomic integrity. The underlying principle of the SAC is quite simple. Chromosomes that are not yet prepared for division alert the SAC, that in turn blocks the cell cycle machinery, stopping cells in mitosis, the division phase of the cell cycle. However, a quarter of a century after the molecular identification of its key components, SAC signalling is still not fully understood and the SAC is still under extensive investigation (reviewed in Musacchio (2015) and London & Biggins (2014)).

The action of the SAC, back then still unknown, was first encountered in the early 1980s when cells were observed to arrest in mitosis upon treatment with chemicals that negatively interfere with the mitotic spindle, the protein structure that pulls chromosomes apart upon division (Umesono *et al.*, 1983; Zieve *et al.*, 1980). Therefore, the unknown mechanism was termed spindle assembly checkpoint. Today, it is sometimes simply called mitotic checkpoint. The core components of the SAC were identified in two independent seminal studies in 1991 (Hoyt *et al.*, 1991; Li & Murray, 1991). Thereafter, the target of the SAC within the cell cycle control system was identified as protein complex APC/C (Sudakin *et al.*, 1995). At

about the same time, elegant experiments linked the sensory component of SAC signalling to the kinetochores, the sites where chromosomes get attached to the mitotic spindle (Li *et al.*, 1995; Rieder *et al.*, 1995). Thereafter, the interface between the SAC and the cell cycle was found to be the protein Cdc20 (Slp1 in fission yeast) (Hwang *et al.*, 1998; Kim *et al.*, 1998). Now provided with all the components, the mechanism of signal transduction between the kinetochores and the cell cycle machinery remained under investigation and is still not completely understood (see Section 2.2 for a more detailed review on SAC signalling).

During the last decade, classical biological research on the SAC has been complemented by model-based approaches and the importance of joining forces has been stressed (Ciliberto & Shah, 2009; Joglekar, 2016). Mathematical models have since been used to model different aspects of the establishment of a SAC mediated cell cycle arrest as well as of the dynamics of cell cycle progression after SAC silencing (see Section 5.1.5 for a review on existing models). Although modelling revealed interesting aspects of SAC signalling, a systems view of the SAC is still lacking.

We have single-cell data at hand which consist of observations on the duration of the SAC-induced arrest of the cell cycle. This time span is a measure for the functionality of the SAC in the individual cells. We use these data to investigate properties of the SAC signalling pathway by employing two different types of models. An introduction to these model types is provided in the subsequent sections.

Model-based statistical analysis of heterogeneous single-cell data

Single-cell data become more and more abundant and reveal that individual cells react differently to the same stimulus or perturbation. Therefore, single-cell measurements will always lead to a distribution of values. This heterogeneity, manifested as cell-to-cell variability, contains valuable additional information about the underlying biological mechanisms (as detailed in Section 2.3.1). Heterogeneity can have a quantitative or qualitative character, with qualitative meaning several subpopulations with qualitatively different properties within one population. The presence of subpopulations can cause multimodal or skewed distributions. To interpret such distributions correctly and exploit the full potential of single-cell data, they have to be statistically analysed.

The model-based statistical inference framework developed by Ronald A. Fisher, a British biologist and statistician (Fisher, 1922), is an appropriate basis for the quantification and analysis of single-cell data. The approach builds on a statistical model that describes the idea of the generation process of the variable under consideration. With that, the statistical model provides a link between the observed units in the sample and the infinite population the sample was drawn from, enabling causal or analytic inferences to pertain to the whole population. The key point of a statistical model is the consideration of the observed sample as realizations of a random variable. This random variable has to be specified by a parametric distributional assumption (Sterba (2009) and references therein).

We build on this framework to quantify and analyse cell-to-cell variability in the functionality of the SAC under different signal perturbing conditions. In doing so, we face the challenge that the variable of interest is not always completely observable in our data, which is known as data censoring. However, by exploring the potential of model-based inference methods, sophisticated approaches that can deal with censoring and go beyond sole quantification of

variability can be derived, as it is shown in Chapter 4.

The results gained via statistical modelling form the data basis for models of the molecular signalling mechanism of the SAC, i.e. models that describe how the SAC arrest is generated on the molecular level, presented in Chapter 5. The following section provides a brief introduction to this type of models.

Mechanistic modelling of cellular signalling pathways

Signalling pathways enable cells to transmit and process signals originating from the environment or within the cell. Sensors, such as membrane receptors, sense signals that are then transmitted and integrated to activate actuators, such as transcription factors, to change the state and biochemical composition of the cell. Signalling is closely related to decision making, the processing of the information in the signal to determine required actions (Perkins & Swain, 2009). In cells, signals are encoded by the number and activity of mostly proteins and complexes of proteins. Dynamics are generated by synthesis and degradation, complex formation and modification through binding to other proteins or small molecules and phosphorylation, the covalent binding of phosphate residues. Complex behaviours including feedback and feedforward signalling are hard to grasp by intuition. In order to assess the complex behaviour of signalling pathways with their various components and interactions, researchers have adopted computational modelling techniques. Models enable a systems view of complex behaviours and the theory of dynamical systems provides a resource of powerful analysis methods (Wolkenhauer, 2014). Models exist for different scales from single reoccurring building blocks (Tyson *et al.*, 2003) to whole signal transduction networks all the way down from sensors to actuators (reviewed in Klipp & Liebermeister (2006)). Models of signalling pathways are of various types. While mechanistic models are based on chemical reaction kinetics and provide quantitative dynamics (Tyson & Novak, 2015), phenomenological models describe the interaction between the signalling species in a qualitative manner (Angeli *et al.*, 2004). Both model types can be of stochastic (Wilkinson, 2009) or deterministic nature (Raue *et al.*, 2013b).

The SAC is such a cellular signalling pathway. The involved proteins can form various complexes and activating and inhibiting interactions between them are numerous. We use mechanistic models, based on deterministic kinetics, combined with parametrised variability in the model species and model parameters, to model SAC signalling in heterogeneous populations in Chapter 5. These models are calibrated to results gained through statistical model-based analysis as presented in Chapter 4.

1.3. Contribution of this thesis

In this thesis, the system properties and the mechanism of SAC signalling are assessed in a twofold approach. We use statistical models to analyse censored single-cell data of SAC functionality and calibrate mechanistic models with phenotypic SAC data.

In the first part of this thesis, we present the statistical model-based

- quantification of cell-to-cell variability in the functionality of the SAC in several strains of *S. pombe* with manipulated protein abundances.
- detection and identification of subpopulations with different cellular SAC phenotypes in isogenic yeast populations under SAC signalling perturbing conditions.
- quantitative characterization of the sensitivity of SAC signalling with respect to changes in the abundance in its key signalling proteins Mad2 and Mad3.
- assessment of the mode of interaction between Mad2 and Mad3 in influencing SAC functionality.

We show that

- the yeast strains with altered abundances of Mad2 or Mad3 display up to two different phenotypes of SAC functionality within populations of isogenic cells, one of these phenotypes corresponding to the wild type (WT).
- SAC functionality exhibits a highly ultrasensitive behaviour with respect to the total amount of Mad2 available for signalling and an almost hyperbolic behaviour with respect to the amount of Mad3.
- Mad2 and Mad3 act cooperatively in promoting SAC functionality.

Since all these findings require the analysis of censored single-cell data from various experimental conditions, we propose

- a statistical modelling framework for the analysis of single-cell data from multiple experiments in the presence of censoring. The framework applies mixture models and is implemented as the MATLAB toolbox MEMO (Multi-Experiment mixture MOdelling). MEMO is also applicable to other types of censored and complete 1D data.

In the second part of this thesis, we present

- an assessment of the sources of non-genetic cell-to-cell variability in the cellular phenotype, based on mechanistic models of SAC signalling on the population level, calibrated to phenotypic data.

We show that

- ultrasensitivity is also a property of the dynamical systems our SAC models create.
- ultrasensitivity in the input output response of our model is caused by bistability in the strength of SAC signalling, arising from an ultrasensitive mechanism of stoichiometric inhibition embedded into several double negative feedback loops.
- this ultrasensitivity in combination with small differences in the amount of SAC signalling proteins can explain the observed population split. We predict that these differences between cells are more pronounced in the SAC target Slp1, than in the signalling components.

1.4. Outline of this thesis

As an introduction to the topic of this thesis, Chapter 2 surveys its biological and methodological background. The model organism *Schizosaccharomyces pombe* is introduced and the present knowledge of the molecular mechanism of SAC signalling is reviewed. Moreover, we elaborate on several aspects in the context of single-cell data. Since likelihood-based parameter estimation approaches are used for inference in this thesis, they are introduced in this chapter.

Chapter 3 presents the experimental data, on which this thesis project is founded. These data were provided by Dr. Stephanie Heinrich, a former Phd student in the group of Dr. Silke Hauf, at that time located at the Friedrich Miescher Laboratory in Tübingen.

In Chapter 4 MEMO, a statistical modelling framework for the quantification and analysis of censored single-cell data from multiple experimental conditions, is presented. This framework is employed to quantify and analyse cell-to-cell variability in SAC prometaphase single-cell data, quantify the sensitivity of SAC functionality towards changes in the abundance of signalling proteins Mad2 and Mad3 and examine the mode of interaction of these two proteins in determining the SAC phenotype.

The statistical analysis of the prometaphase length data is complemented by mechanistic models of SAC signalling in Chapter 5. SAC signalling is modelled on the population level by explicitly considering cell-to-cell variability in reaction rates and protein amounts. Two models describing the mechanism of SAC signalling to a different extent are derived, calibrated to phenotypic data and analysed.

Chapter 6 provides a summary of the main results and discusses conclusions. The thesis is concluded by an outlook.

2. Background

This chapter provides the background information on the key aspects relevant throughout this thesis. Since the experimental data stem from experiments with the yeast *Schizosaccharomyces pombe*, this organism is introduced in Section 2.1. Section 2.2 gives a survey of the biological system under study, the spindle assembly checkpoint. In Section 2.3 different aspects in connection with single-cell data are discussed. An insight into the aspects of model calibration based on data which are applied to statistical models in Chapter 4 and mechanistic models in Chapter 5 is given in Section 2.4. These are the likelihood based estimation of model parameters from data and the quantification of uncertainty associated with these parameters. This section addresses the aspects independent of the class of the calibrated model while specific issues for the statistical and the dynamical models respectively are exemplified in the corresponding chapters.

2.1. *Schizosaccharomyces pombe* - a model organism

Schizosaccharomyces pombe, also called fission yeast, is a unicellular rod shaped fungus. *S. pombe* grows through elongation on both ends and divides by medial fission into two daughter cells of equal size. It is a non-pathogenic, fast growing eukaryotic organism that resembles many features of multi cellular organisms. Its genome, consisting of three chromosomes, is relatively easy to manipulate (Forsburg & Rhind, 2006). Since its first isolation from African millet beer in 1890 and its first description by the German scientist P. Lindner in 1893 (Lindner, 1893), *S. pombe* has become an important model organism for mammalian in general and human cells in particular (Forsburg, 2005; Hoffman *et al.*, 2015). The structure of the fission yeast kinetochore resembles the one of higher organisms quite well (Pidoux & Allshire, 2004), which makes fission yeast suitable to study the cell cycle. However, in contrast to human cells, the expression of Cdc20 is cell cycle dependent and restricted to mitosis (Buck *et al.*, 2004; Yamada *et al.*, 2000).

2.2. The spindle assembly checkpoint

The biological system investigated here is the spindle assembly checkpoint. The spindle assembly checkpoint is a signalling mechanism that synchronizes cell division with conditions that considerably promote the correct distribution of chromosomes to the emerging daughter cells (Musacchio, 2015). In mammals it is essential for viability and its function and components are conserved in eukaryotes (Vleugel *et al.*, 2012). It prevents the cellular genomes from acquiring abnormal chromosome numbers, a condition termed aneuploidy. Aneuploidy has detrimental consequences on the cellular level (Torres *et al.*, 2007) as well as on the or-

ganism level. It may contribute to the development of cancer (Kops *et al.*, 2005; Santaguida & Amon, 2015). Furthermore, it is the leading cause of miscarriage and still birth in humans (Hunt & Hassold, 2008).

Operating in mitosis, the SAC is also known as mitotic or metaphase checkpoint. Mitosis is the phase in the cell cycle in which the genetic information is distributed in preparation for the subsequent division into two daughter cells. At this point, chromosomes have been duplicated beforehand. Therefore, they consist of two identical sister chromatids linked by the protein complex cohesin. The concept of chromatid distribution is as simple as it is ingenious: each of the two chromatids gets attached to its own part of the mitotic spindle emerging from the opposite parts of the cell. This condition is called biorientation of the chromosome. Biorientation ensures that each sister chromatid, after loosing cohesin, will be separated into different daughter cells when the cell divides. The contribution of the SAC is to delay the loss of sister chromatid cohesion until every chromosome achieved biorientation. The execution of this task is entrusted to a set of proteins and protein kinases, interacting to generate a signalling cascade of protein-protein interactions and phosphorylation that prevents chromatid separation, and therefore cell division, until optimal conditions are fulfilled.

The key players of this signal are the Mad1 and Mad2 (mitotic arrest deficient) (Li & Murray, 1991), Bub1 and Bub3 (budding uninhibited by benzimidazoles) (Hoyt *et al.*, 1991), BubR1 (Bub1-related 1, Mad3 in *S. pombe*) proteins together with the protein kinase Mps1 (monopolar spindle 1 (Weiss & Winey, 1996), Mph1 in *S. pombe* (Mps1p-like pombe homolog) (He *et al.*, 1998)). These proteins are conserved in all eukaryotic cells and essential for SAC signalling. The absence of either of these proteins leads to a dysfunctional SAC and cell division in the presence of unattached chromosomes. The target of the SAC is the APC/C (anaphase-promoting complex/cyclosome) (Primorac & Musacchio, 2013), a protein complex that, when activated, promotes the separation of the sister chromatids and anaphase onset. The APC/C needs to bind Cdc20 (cSlp1 in *S. pombe*), its coactivator, to become active (Yu, 2007). To block the exclusive binding of Cdc20 and APC/C is the ultimate purpose of SAC signalling.

2.2.1. Sensing and signal initiation at kinetochores

Sensors detect the status of a quantity and generate and relay a corresponding output signal. This exactly is the task of kinetochores in SAC signalling. The kinetochore is a protein assembly situated on the chromatid DNA which mediates the attachment to the spindle microtubules of the mitotic spindle (Cleveland *et al.*, 2003). Kinetochores sense the attachment status and activate a progression-inhibiting signal as long as attachment is not or not correctly achieved. However, the nature of sensing has not yet been completely elucidated. Two different models exist, one of which is of biochemical the other one of rather mechanical nature. The biochemical point of view attributes sensing and signal initiation to spindle proteins and SAC proteins competing for binding at the microtubule binding sites of the kinetochores (Hiruma *et al.*, 2015; Ji *et al.*, 2015). The mechanical model attributes sensing and signal initiation to a mechanical switch constituted of proteins that are close as long as the kinetochore is not attached and become separated in space upon attachment (Aravamudhan *et al.*, 2015). This switch is a “normally-closed” switch: separation after attachment leads to ceasing of signal transduction. Without an activating signal from the kinetochores the SAC is shut off. A possible resolution

to the differing notions is a combination of both principles that acts as the origin of the SAC signal (Joglekar & Aravamudhan, 2016).

However, it is known that missing or incorrect attachment leads to the recruitment of proteins to the kinetochore. With this equipment, kinetochores act as a “catalytic scaffold” in initiating the diffusible “wait anaphase” signal.

2.2.2. Relaying the wait anaphase signal

The wait anaphase signal relays the information on the kinetochores attachment status to the cell cycle control system represented by APC/C. This ubiquitin ligase tags proteins for degradation via the proteasome by linking them to the small protein ubiquitin. One of its targets is securin, a protein that protects chromatid cohesion (Thornton & Toczyski, 2003). By inhibiting the APC/C the SAC therefore protects securin and inhibits the separation of the chromatids.

The effector of the SAC signal is the MCC (mitotic checkpoint complex). The MCC consists of Mad2, BubR1(Mad3), Cdc20 and in many organism Bub3 (Chao *et al.*, 2012). The MCC on the one hand sequesters free Cdc20 in a complex in which it is not able to activate the APC/C and on the other hand blocks the APC/C from activation via free Cdc20. The steps that lead to the assembly of the MCC are only partially understood. According to the template model Mad1 bound CMad2 at unattached kinetochores serves as a template for a hetero dimer of Cdc20 bound CMad2 (De Antoni *et al.*, 2005). For that purpose OMad2 dimerizes with Mad1 bound C-Mad2 thereby adopting an activated conformation that enables the binding to Cdc20 and the concomitant transition to CMad2. CMad2 bound Cdc20 is not able to activate the APC/C. Mad2 is required to recruit the remaining components to the MCC. Therefore, it is likely, that they bind to the hetero dimer Cdc20:CMad2 to complete the MCC. The MCC is constantly turned over during persistent SAC activation. Its dissociation is attributed to at least two different mechanisms. First, the APC/C dependent degradation of Cdc20 bound to the APC via the MCC. This dissociation is paralleled by the release of Mad2. The second mechanism is assumed to focus an MCC not bound to the APC/C, possibly by mediating the conversion of CMad2 to OMad2 and without the destruction of Cdc20.

2.3. Single-cell data

In contrast to population average data gained by lysate-based methods, such as immunoblots, PCR or microarrays, single-cell data disclose that heterogeneity is ubiquitous even in isogenic populations. Single-cell data not only prevent from spurious conclusions caused by the loss of information due to averaging, but can also provide valuable additional information (Altschuler & Wu, 2010; Huang, 2009). The following sections elaborate on various aspects in the context of single-cell data. In Section 2.3.1 the term cell-to-cell variability is clarified and its relevance is outlined. Section 2.3.2 deals with a special type of single-cell data known as time-to-event data. Section 2.3.3 introduces data censoring.

2.3.1. Aspects of non-genetic cell-to-cell variability

Non-genetic cell-to-cell variability, also referred to as population heterogeneity, specifies the variability that can be observed in single-cell data from isogenic populations of cells experiencing the same culturing conditions and treatment. This variability is a result of deterministic as well as stochastic processes. In particular the stochastic aspect of cell-to-cell variability is termed cellular noise. Cellular noise is in turn further divided into intrinsic and extrinsic noise. While intrinsic noise refers to variation caused by the inherent stochastic and discrete nature of intracellular biochemical reactions, extrinsic noise refers to upstream sources of variation, like differences in cell cycle state. Extrinsic noise can reflect stochastic as well as deterministic influences. Deterministic extrinsic noise is also sometimes called regulated cell-to-cell variability as opposed to cellular noise (Snijder & Pelkmans, 2011). Sources of deterministic extrinsic noise may be differences due to unequal cell division (Huh & Paulsson, 2011).

Cell-to-cell variability is omnipresent in biological systems (Balázsi *et al.*, 2011). Clonal populations can show quantitative differences in gene expression and qualitatively distinct cellular phenotypes and subpopulations (Balaban *et al.*, 2004; Eldar & Elowitz, 2010). The magnitude and nature of variability within a population can differ significantly depending on the system under consideration (Pelkmans, 2012). The observed variability can serve as a potential source of information for the inference of the regulatory mechanisms (Li & You, 2013; Pelkmans, 2012; Rinott *et al.*, 2011) and the interpretation of effector screenings (Dey *et al.*, 2014). In contrast, averaging might even lead to spurious conclusions (Altschuler & Wu, 2010; Loo *et al.*, 2009).

Most studies on cell-to-cell variability are related to gene expression (Elowitz *et al.*, 2002; Swain *et al.*, 2002). But cell-to-cell variability can also arise without substantial contribution of transcriptional and translational processes, but through signal transduction (Colman-Lerner *et al.*, 2005; Jeschke *et al.*, 2013).

2.3.2. Time-to-event data

In time-to-event data the quantity of interest is the time that elapses between a start point and the observation of the event of interest. Time-to-event data are often called survival data no matter which event is observed (Altman & Bland, 1998). Common questions posed to these data concern the chance to survive upon a certain time or whether a certain treatment has an impact on this chance. Although being phenotypic data, time-to-event data can also give insights into the process leading to the event (Yurkovsky & Nachman, 2013). The distribution of event times can tell something about whether a single process is responsible or the slowest of a group of parallel processes determines the event, or a sequence of processes (Pedraza & Paulsson, 2007; Yurkovsky & Nachman, 2013). A bimodal distribution points to the existence of different subgroups (Nachman *et al.*, 2007). Statistical tests on the relation between the timing of different events can reveal whether the underlying processes are independent or not (Huang *et al.*, 2010) and whether they take place sequentially or run in parallel (Duffy *et al.*, 2012). A characteristic inevitably linked with time-to-event data is censoring.

2.3.3. Data censoring

Data are not always complete. They can be subject to censoring and truncation, which prevent the collection of observations from the whole data sample space. In censored data for some observations only partial information is available, while in truncated data the data itself is incomplete. In the case of censoring a value is recorded for every observation, but the exact value is not always known (Klein & Moeschberger, 2003). In contrast, truncation excludes observations which do not fall into a particular range. This is in most cases not done on purpose but caused by the study design. While we focus on censored data in this thesis, truncation is also an important aspect in the statistical analysis of data and has to be considered in the analysis when present (Dai & Wang, 2016).

Most experimental devices provide censored data due to limited resolution or experimental constraints. There are mainly three types of censoring. In case only an upper bound is determinable in a measurement this is called left censoring and the corresponding data is called left censored data. In this case the real value is known to be somewhere below the measured value. In other cases the measurement gives only a lower bound for the real value, e.g. if the measurement technique allows only for measurements up to a certain limit. Then the real value of the observation is known to be above this limit without knowing how much. This type of censoring is called right censoring. Sometimes the real value can only be narrowed down to lie between a lower and an upper bound. This is called interval censoring, since we only know that the real value is in the interval between the lower and upper bound. Data censoring hampers the statistical analysis of single-cell data since the quantity of interest cannot be fully observed in all cells which biases the statistics, if not properly accounted for. In Section 4.1.1 the different types of censored data are discussed in more detail.

2.4. Likelihood based parameter estimation and uncertainty analysis in a nutshell

This section introduces the principles of likelihood based parameter estimation as they are used throughout this thesis to estimate unknown parameters from data. This introduction is by no attempt considered complete, but rather restricted to the aspects applied for inference in Chapter 5 and Chapter 4, which present the results of this thesis. The notation is kept general since the same methods are applied to estimate the parameters of statistical models (Chapter 4) as well as mechanistic models (Chapter 5).

We consider the model $\mathcal{M}(\theta)$ which is a function of the unknown parameters θ . Furthermore, we consider a dataset $\mathcal{D} = \{\mathcal{D}_i\}_{i=1}^I$ consisting of data from I different experiments. The likelihood is the conditional probability of observing the data given the model and its parameters $\mathbb{P}(\mathcal{D}|\theta)$.

2.4.1. Frequentist approach

Frequentist parameter estimation assesses the quality of the model using the likelihood of the data given the model parameters, $\mathbb{P}(\mathcal{D}|\theta)$ (Scholz, 2004). Assuming that measurements are independent and identically distributed (i.i.d.), the likelihood function for multiple experimental

datasets is given as the product of the likelihood functions for the individual datasets,

$$\mathbb{P}(\mathcal{D}|\theta) = \prod_{i=1}^I \mathbb{P}(\mathcal{D}_i|\theta). \quad (2.1)$$

This likelihood function encodes the information about the optimal parameter values and parameter uncertainties present in the experimental data.

Parameter estimation: The maximum likelihood (ML) estimate

A maximum likelihood estimate (MLE) θ^{ML} is a parameter vector for which the likelihood takes its maximum value in an open region $\Omega \in \mathbb{R}^{n_\theta}$, hence $\forall \theta \in \Omega : \mathbb{P}(\mathcal{D}|\theta) \leq \mathbb{P}(\mathcal{D}|\theta^{\text{ML}})$ (Scholz, 2004). Accordingly, θ^{ML} is a solution to the optimization problem

$$\theta^{\text{ML}} = \arg \max_{\theta \in \Omega} \mathbb{P}(\mathcal{D}|\theta). \quad (2.2)$$

Often, the numerics of the optimization problem (2.2) and the efficiency of the implementation are improved by using the negative logarithm of the likelihood as objective function,

$$\mathcal{J}(\theta) = -\log \mathbb{P}(\mathcal{D}|\theta) = -\sum_{i=1}^I \log \mathbb{P}(\mathcal{D}_i|\theta). \quad (2.3)$$

The log transformation transforms the products in (2.1) into sums in (2.3) while it is monotone and therefore conserves the extrema and the shape of the level sets. Given that

$$\arg \max_x(x) = \arg \min_x(-x),$$

the reformulation yields the minimization problem

$$\theta^{\text{ML}} = \arg \min_{\theta \in \Omega} \mathcal{J}(\theta). \quad (2.4)$$

For strictly positive parameters θ it may also be advantageous to use a log-transformation of the parameters, $\xi = \log(\theta)$, and estimate the parameters in the logarithmic space. In this case, $\xi \in \Omega' = \{\xi \in \mathbb{R}^{n_\theta} | \exp(\xi) \in \Omega\}$ is estimated instead of $\theta \in \Omega$. The transformation into logarithmic space often improves efficiency of optimizers since all elements of ξ are of the same order of magnitude.

In our settings, the minimization problem (2.4) is nonlinear and non-concave (as the maximization problem (2.2) is non-convex). To solve (2.4), sophisticated optimization schemes are required. Commonly used global optimization methods are multi-start local optimization (Raue *et al.*, 2013b), evolutionary and genetic algorithms (Bäck, 1996), particle swarm optimizers (Yang, 2010), simulated annealing (Kirkpatrick *et al.*, 1983) and hybrid optimizers (Balsa-Canto *et al.*, 2008; Vaz & Vicente, 2007). For details we refer to available comprehensive surveys of local and global optimization procedures (Banga, 2008; Moles *et al.*, 2003; Raue *et al.*, 2013b; Weise, 2009). In this thesis we use multi-start local optimization, an approach which has been shown to be efficient for a wide class of problems (Raue *et al.*, 2013b).

Multi-start local optimization

Given the intricacy of the objective functions that result from the models we consider, local search algorithms are likely to get trapped in a local optimum. To circumvent this shortcoming multiple starts with different starting points are a promising approach. Starting points for individual local optimization runs can for example be generated by Latin hypercube sampling (Owen, 1992).

Confidence intervals

As the measurement data are limited, the parameters can often not be determined uniquely and the corresponding estimation problem is ill-posed (Hadamard, 1902). Confidence intervals can be computed via local sensitivity-based methods, e.g., the Wald approximation (Meeker & Escobar, 1995) or the Fisher information matrix (FIM) (Murphy & van der Vaart, 2000). Alternatively, bootstrapping (Joshi *et al.*, 2006), profile likelihoods (Murphy & van der Vaart, 2000; Raue *et al.*, 2009) and Markov chain Monte-Carlo methods (Girolami & Calderhead, 2011; Wilkinson, 2007) can be used. Nowadays, Bayesian methods and profile likelihoods (Murphy & van der Vaart, 2000; Raue *et al.*, 2013a) become more and more popular as they yield very reliable results (Raue *et al.*, 2013a).

Frequentist uncertainty analysis often employs profile likelihood methods. Profile likelihoods allow for a global uncertainty analysis of individual parameters by means of repeated optimization (Raue *et al.*, 2009). Therefore, for each estimated parameter θ_i a profile likelihood is computed by repeatedly fixing the parameter θ_i to values of a set and optimizing over all remaining parameters,

$$\text{PL}(\theta_i) = \max_{\theta_{j \neq i}} \mathbb{P}(\mathcal{D}|\theta), \theta \in \Omega.$$

The profile likelihood $\text{PL}(\theta_i)$ is the maximum likelihood value for a given value θ_i . A particular value of θ_i can be rejected, if the profile likelihood $\text{PL}(\theta_i)$ is low compared to the likelihood $\mathbb{P}(\mathcal{D}|\theta^{\text{ML}})$ at the globally optimal parameter point θ^{ML} . Cut-off values for the likelihood ratio for a particular significance level can be derived from the χ^2 distribution (Meeker & Escobar, 1995).

2.4.2. Bayesian approach

In the Bayesian approach of parameter estimation and uncertainty analysis, the available (prior) information on the parameters and the information gathered by collecting data, the likelihood, are merged. This is accomplished by the Bayes' theorem,

$$\pi(\theta|\mathcal{D}) = \frac{\mathbb{P}(\mathcal{D}|\theta)\pi(\theta)}{\mathbb{P}(\mathcal{D})} \propto \mathbb{P}(\mathcal{D}|\theta)\pi(\theta) \quad (2.5)$$

with

$$\mathbb{P}(\mathcal{D}) = \int_{\Omega} \mathbb{P}(\mathcal{D}|\theta)\pi(\theta)d\theta.$$

The posterior distribution $\pi(\theta|\mathcal{D})$ of the parameters θ given the data \mathcal{D} is determined by the likelihood $\mathbb{P}(\mathcal{D}|\theta)$, the prior distribution $\pi(\theta)$ and the marginal probability $\mathbb{P}(\mathcal{D})$. The marginal

probability is independent of the parameters. Therefore, the posterior probability is proportional to the product of likelihood and prior.

Parameter estimation: The maximum a posteriori (MAP) estimate

A maximum a posteriori estimate θ^{MAP} is a parameter vector for which the posterior probability takes its maximal value, hence $\forall \theta \in \Omega : \pi(\theta|\mathcal{D}) \leq \pi(\theta^{\text{MAP}}|\mathcal{D})$. Thus, MAP estimates provide the best agreement of model and data taking the prior knowledge into account. As the posterior probability is proportional to the product of likelihood and prior probability, θ^{MAP} is a solution to the optimization problem

$$\theta^{\text{MAP}} = \arg \max_{\theta \in \Omega} \mathbb{P}(\mathcal{D}|\theta)\pi(\theta).$$

This optimization problem can be reformulated similar to the corresponding ML problem (2.2).

Uncertainty analysis: Credibility intervals

In Bayesian statistics, the a posteriori uncertainty of model parameters depends on the information content of the data – encoded in the likelihood – and the prior information. Bayesian credibility intervals for the parameters (Chen & Shao, 1999) can, for instance, be computed using Laplace approximations at the MAP estimate, profile posteriors (Hug *et al.*, 2013) and Markov chain Monte-Carlo (MCMC) methods (Girolami & Calderhead, 2011; Hug *et al.*, 2013; Wilkinson, 2007).

MCMC methods are particularly useful if the posterior distribution $\pi(\theta|\mathcal{D})$ cannot be assessed analytically, but can be evaluated for every θ by computing $\mathbb{P}(\mathcal{D}|\theta)\pi(\theta)$. Then MCMC algorithms can be employed to generate a chain of parameters, $\theta^1, \theta^2, \dots, \theta^{n_S}$, by exploring $\pi(\theta|\mathcal{D})$. After convergence of the chain, the set $\mathcal{S} = \{\theta^j\}_{j=1}^{n_S}$ provides a representative sample from the posterior distribution. This sample \mathcal{S} reveals parameter uncertainties as well as correlations of parameters. The most common choice for the $100(1 - \alpha)\%$ credibility interval for a parameter θ_j is the $100(1 - \alpha)$ -th percentiles of the sample \mathcal{S} (DiCiccio & Efron, 1996).

3. Experimental data

This chapter reviews the data used for model-based inference in this thesis. All experiments were performed in the former group of Dr. Silke Hauf at the Friedrich Miescher Laboratory of the Max Planck Society in Tübingen. Dr. Stephanie Heinrich created all the strains and performed all the fluorescence microscopy experiments. We published the data together with a computational analysis in Heinrich *et al.* (2013).

3.1. Yeast strains

One aim of Stephanie Heinrichs PhD project was to investigate the robustness of SAC signalling. Since SAC signalling involves a series of protein-protein interactions, changes in the abundance of these proteins should affect the functionality. To systematically probe SAC activity following changes in protein abundance, strains of *S. pombe* differing in the amounts of several SAC proteins were created. The abundances of Mad1, Mad2 and Mad3 were modified by promoter modifications. The promoter of a gene is a regulatory element that determines the abundance of the gene product, the corresponding protein. The modifications resulted in a number of strains that will in the following be referred to by their relative protein abundance in relation to the wild type strain (Table 3.1). The proteins were expressed as green fluorescent protein (GFP) fusions from the modified promoter at the endogenous locus. The protein abundances relative to wild type were analysed by immunoblotting (original blots are available in Supplementary Figure 4 of Heinrich *et al.* (2013)). To create strains completely lacking a certain protein, the respective gene was deleted. In another set of strains the abundance of Slp1 was increased by inserting a second gene copy under its endogenous regulatory sequences. Promoter modifications, gene deletion and gene insertion resulted in a variety of strains with changed protein abundances in comparison to the wild type. The ones used for model-based inference in this thesis are identified in Table 3.1 together with the chapter in which they appear within this thesis.

3.2. Prometaphase length data

Dr. Stephanie Heinrich used the collection of yeast strains to study the dependency of SAC functionality on SAC protein abundance by the help of fluorescence microscopy. Therefore, SAC signalling was activated and the prometaphase length, which is a measure of SAC functionality, was determined for every cell (Figure 3.1). To be able to reliably activate SAC signalling, every strain expresses a temperature sensitive tubulin mutant. If these cells are grown at low temperatures, microtubule formation is prevented. The lack of microtubules leads to activation of SAC signalling at all kinetochores. Consequently, cells stop in mitotic

3. Experimental data

strain	% Mad1	% Mad2	% Mad3	% Slp1	Chapter
0% Mad2	100	0	100	100	4
10% Mad2	100	10	100	100	4
20% Mad2	100	20	100	100	4
40% Mad2	100	40	100	100	4
65% Mad2 P50	100	65	100	100	4
65% Mad2 & 30% Mad3	100	65	30	100	4
65% Mad2 & 60% Mad3	100	65	60	100	4
65% Mad2 & 120% Mad3	100	65	120	100	4
65% Mad2 P188	100	65	100	100	4
80% Mad2	100	80	100	100	4
WT Mad2	100	100	100	100	4
200% Mad2	200	100	100	100	4
0% Mad3	100	100	0	100	4
30% Mad3	100	100	30	100	4
60% Mad3	100	100	60	100	4
WT Mad3	100	100	100	100	4
120% Mad3	100	100	120	100	4
30% Mad1	30	100	100	100	5
2×Slp1	100	100	100	200	5
2×Slp1 & 30% Mad1	30	100	100	200	5
2×Slp1 & 65% Mad2	100	65	100	200	5
2×Slp1 & 30% Mad3	100	100	30	200	5

Table 3.1.: List of *S. pombe* strains created by Dr. Stephanie Heinrich and used in this thesis for model-based inference. Strains are named by their protein amounts relative to the wild type strain. All strains are listed together with the chapter of appearance for inference

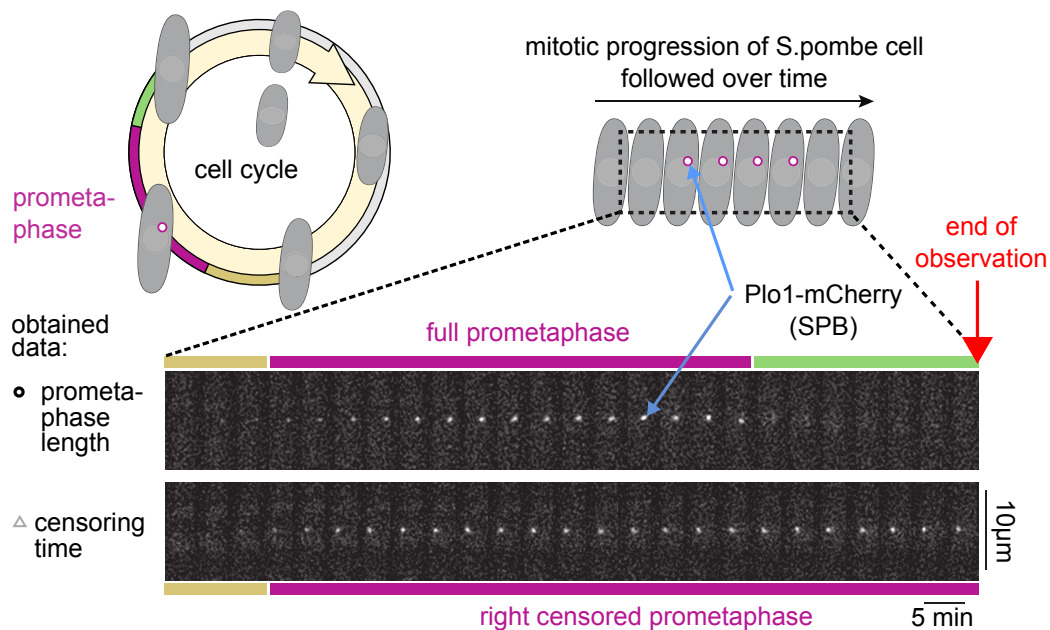


Figure 3.1.: Fluorescence microscopy live-cell imaging of *S. pombe* cells to assess the prometaphase lengths of individual cells. Localization of Plo1-mCherry to spindle pole bodies (SPB) was used to determine the prometaphase length. Cells were recorded every five minutes. A sustained Plo1 signal at the end of the observation lead to the recording of a censoring time. Figure adapted from Geissen *et al.* (2016).

prometaphase. To be able to determine how long the SAC stops a single cell, a fluorescence marker was used. The strains express the protein kinase Plo1 in a fluorescence tagged version (Plo1-mCherry). During prometaphase, Plo1 locates at the spindle pole bodies (SPB), the origins of the mitotic spindle. Therefore its localisation serves as a marker for prometaphase and its length. To assess SAC functionality cells were shifted to the restrictive temperature and imaged for the localisation of Plo1 every 5 minutes for at most 17 hours. This leads to right censoring, since not for all imaged cells the disappearance of Plo1 from the SPBs can be recorded. For cells in which Plo1 was observed to disappear the duration of Plo1 localisation to the SPBs is recorded as prometaphase length of that cell. If Plo1 is still present at the SPBs when the recording ends, the duration of Plo1 localisation up to this point is recorded as a censoring time. Since the cells enter mitosis asynchronously the censoring times of individual cells are distributed. The determined values are furthermore subject to interval censoring, since the disappearance of Plo1 is checked only every 5 minutes. The raw data are shown in Figure 4.5 and Figure 4.7 in Chapter 4.

4. Statistical models for the quantification and analysis of cellular SAC phenotypes

This chapter deals with the model-based quantification and analysis of cell-to-cell variability of SAC signalling in populations of genetically identical cells. The statistical analysis enables the quantification of subpopulation structures within the populations and the quantification of the sensitivity of SAC functionality with respect to the amount of key signalling proteins. For the model-based quantification and analysis of cell-to-cell variability we propose a modelling framework for the statistical analysis of censored single-cell data. We apply this framework to analyse SAC prometaphase length data. This chapter is based on the paper “MEMO: Multi-experiment mixture modelling of censored data” published in *Bioinformatics*, Oxford Journals (Geissen *et al.*, 2016).

MEMO is a framework for the quantification and statistical analysis based on finite mixture models and maximum likelihood estimation. The development of MEMO resulted from the demand for an approach that is able to quantify the heterogeneity in the censored SAC prometaphase length data. In Section 4.1 we elaborate on different types of censored data and sources of censoring in biological data. Moreover, we introduce mixture models as an approach for the modelling of censored single-cell data. The problems addressed in this chapter are formulated in Section 4.2. After presenting the statistical analysis approach in Section 4.3, we outline the pitfalls inherent in the analysis of censored data by analysing two exemplary SAC datasets with known subpopulation structures. SAC prometaphase length data is analysed in Section 4.4 taking into account that due to the measurement technique all datasets are interval censored and most of them are right censored. The chapter closes with a short summary and a discussion of the results in Section 4.5.

4.1. Introduction

Since the quantification of cell-to-cell variability in our SAC prometaphase data is hampered by the fact that the data is censored, this section elaborates on the properties of censored data and a certain type of statistical model that is able to capture these properties. Different types of censored data and the conditions that lead to their occurrence are illustrated in Section 4.1.1. Examples for the sources of censoring in a biological context are given in Section 4.1.2. In Section 4.1.3 the statistical modelling concept of finite mixture models is introduced, which is the underlying concept of all models derived in this chapter.

Throughout the present chapter we distinguish between observations and data. With observation we mean the information gathered for one individual cell or more generally speaking,

one unit. Data are the set of all observations in one study or experiment.

4.1.1. Censored data

Censored data arise if the variable of interest is not fully observable in all observations constituting a dataset. The conditions leading to the lack of full observability and the corresponding type of censored data are reviewed in this section. We give a formal introduction to the different types of censored data by briefly illustrating the process of data generation.

Right censored data

In right censored data an observation is only exact if it does not exceed a certain value, the right censoring value (Millard *et al.* (2012), Chapter 11). Beyond this value the observation only provides the information that the exact value is greater than the observed censoring value. Denote the exact value by X and the right censoring value by C . Let the observation T be

$$T = \min\{X, C\}$$

and δ^r be an indicator:

$$\delta^r = \begin{cases} 1 & \text{if } T = X, \\ 0 & \text{if } T = C. \end{cases}$$

The data involving right censoring is denoted as $\{(T_i, \delta_i^r) : i = 1, \dots, n\}$.

Depending on the setup, right censoring is classified into Type I, Type II and random censoring. In **Type I censoring** the censoring value is fixed and/or predetermined. In this context fixed means equal for every observation. This is for example the case if a measurement device has a limit up to which it can measure, like for example a scale or a thermometer. Type I censoring includes three possible subtypes: fixed, progressive, and generalized. Scales and thermometers produce fixed Type I censoring. Progressive Type I censoring arises if for example not only one scale is used but several, each having its own limit. The censoring value per scale is fixed and the specific scale used for a measured unit determines the censoring value. In generalized Type I censoring the censoring value is not fixed but predetermined. Generalized Type I censoring is for example observed if again only one scale is used for the measurement, but we imagine to weight cows that have to be lead on the scale each by their own cowboy of known weight. Subtracting the weight of the cowboy from the weight limit of the scale determines the censoring value for every weighted cow.

In **Type II censoring** the censoring value is not known in advance, but depends on the observations and a predefined procedure of collection. The number of uncensored observations is fixed in advance. The censoring value is given by the largest value observed within this observations. A common example are studies that end when a certain amount of observations have been recorded. Those studies often investigate time spans, and the time at which the acquired number of observations is recorded therefore equals the censoring value. An example would be a study in which the time is recorded that it takes rats to die after being poisoned all at the same time. The study might be predetermined to end with the death of the n th rat. Then this time would be the common censoring value for the observations of all rats still alive at this point in time, of course put to sleep immediately after the n th rats death.

In contrast, in **random censoring** the censoring value does neither have a fixed value for all units nor is it predetermined. In this case the censoring value is a random variable and censoring can be interpreted as a stochastic process competing with the process of interest for realization. For an illustrating example suppose that all rats were poisoned at the same time but this time the responsible PhD student wanted to clock off at 4:30pm and therefore put the survivors to sleep one after the other in random order starting from 4:00pm. Then the censoring value would be the time span between poisoning and euthanasia, which is different and random for every rat.

Left censored data

In left censoring the exact value of an observation is only recorded if it is above a certain limit given by the left censoring value (Millard *et al.* (2012), Chapter 11). For observations below the censoring value the censoring value is recorded and provides an upper bound for the exact value. Left censoring for example occurs if a measurement device has a detection limit and cannot measure values below the detection limit. Denote the exact value by X and the left-censoring value by C_l . Let the observation T be

$$T = \max\{X, C_l\}$$

and δ^l be an indicator:

$$\delta^l = \begin{cases} 1 & \text{if } T = X, \\ 0 & \text{if } T = C_l. \end{cases}$$

Data involving right censoring is denoted as $\{(T_i, \delta_i^r) : i = 1, \dots, n\}$. In general, the different censoring types described for right censoring (Type I, Type II, random) also exist for left censoring. Examples thereof are however rarely relevant in practice and therefore omitted here.

Interval censored data

An observation is interval censored if it is reported as being within a specified interval (Millard *et al.* (2012), Chapter 11). Denote the exact value by X . For interval censored observations an upper $T_u > X$ and a lower bound $T_l < X$ can be recorded. The data is denoted as $\{(T_{l,i}, T_{u,i}) : i = 1, \dots, n\}$. In interval censored data, every observation is censored. Interval censoring is produced by measurement devices with limited resolution where the precision of the device determines the interval. In other cases feasibility reasons determine the interval. Examples are clinical studies where patients can only be observed once a month and an observed change in their health status therefore can only be attributed to the time interval since the last appointment. Note that left censoring located between zero and the detection limit is in fact a special case of interval censoring.

4.1.2. Sources of censoring in biological data

Biological data are subject to multiple sources of censoring. In the following, we discuss some of these sources for common measurement techniques.

Time-lapse microscopy

Time-lapse microscopy is commonly used to monitor cellular dynamics (Coutu & Schroeder, 2013). When cellular events like division, differentiation or death are monitored, the elapsed time until these events occur in the individual cells is of interest for quantification. In time-lapse microscopy left and right censoring is caused by the limited time horizon of the experiment. An event of interest might occur before the observation starts, which leads to a left censored observation. On the other hand the event of interest might not occur before the end of observation, in which case the observation is subject to right censoring. Due to technical constraints and also with the purpose to limit cell stress, the image acquisition rate is limited in video microscopy. This leads to interval censoring of the observations. An event of interest newly observed in one frame must have occurred in the interval between this frame and the last one taken.

Quantitative fluorescence microscopy

Quantitative fluorescence microscopy aims to quantify cellular components by the emitted fluorescence (Waters, 2009). Left and right censoring are caused by background fluorescence and fluorescence saturation, respectively (Visscher *et al.*, 1994). Interval censoring is in this context introduced by limited sensitivity and the digital representation of the signal, for which the analogue signal has to be binned into a limited number of so called channels.

Flow cytometry

In flow cytometry data left and right censoring are caused by background fluorescence and fluorescence saturation, respectively (Pyne *et al.*, 2009; Visscher *et al.*, 1994). Interval censoring is due to limited sensitivity, automated binning and quantization of the analogue signal for digital processing. Flow cytometry data is also frequently subject to truncation, due to gating, the subjective process of selecting the observations of interest for further analysis.

Mass cytometry

In mass cytometry data, saturation of the detector causes right censoring (Nanita, 2013). Since ion counts are discrete these data are also interval censored. A second source of interval censoring in this context is the digital representation of the signal.

Single-cell qPCR

Single-cell qPCR aims to measure the number of RNA molecules in a sample. This is done by the amplification of the sample RNA by repeated PCR cycles until the detection limit is reached. The number of amplification cycles is limited and the maximum number of cycles sets the lower limit of detection (LOD), the minimum number of RNA molecules that can be detected. If the detection limit is not reached after the maximum number of cycles the only information is that the number of RNA molecules was below the detection limit. Therefore, this observation is left censored, while the number of amplification cycles is right censored (Boyer *et al.*, 2013; Buettner *et al.*, 2014; McCall *et al.*, 2014). Interval censoring is caused

in this context by the fact that the actual observations are the number of amplification cycles needed to reach the RNA detection limit.

4.1.3. Mixture models in the analysis of censored single-cell data

The statistical analysis of censored data requires specific methods with the most famous being introduced by Kaplan & Meier (1958). The Kaplan-Meier estimator is mainly used to estimate the survival function for survival data but it can deal with different time-to-event data in principle. Since the Kaplan-Meier estimator is a non-parametric estimator, no assumption on the type of distribution of the data has to be made. Problems arise when the largest data point is right censored. In this case the integral of the survival function is not defined and hence e.g. the mean survival time cannot be calculated.

In contrast, model-based approaches use parametric probability distributions to describe the data. Parametric distributions supply a closed description of the heterogeneity in the cell population. Based on parametric distributions, Duffy and colleagues introduced the idea of censoring caused by competing processes to the analysis of microscopy data. In the case of competing process, not a measurement process, but the competing processes themselves censor the observations of each other. They use parametric distributions to show how censoring through competing processes alters the shape of the observed distribution of events from the real distribution. This can even lead to correlations between the realized quantities of processes although the actual processes are not correlated (Duffy *et al.*, 2012).

If the standard parametric distributions are not suitable to describe the data, i.e. if data suggest multimodality, mixture distributions provide the needed flexibility (Everitt, 1996). Statistical models that make use of mixture distributions are known as finite mixture models (McLachlan & Peel, 2000). In the following we introduce finite mixture models and survey existing approaches for the model-based analysis of (censored) distributed data in a biological context.

Finite mixture models

In 1894, Karl Pearson fitted a mixture of two normal distribution to the ratio of forehead to body lengths of a thousand crabs sampled from the Bay of Naples. He found evidence, that the crab population consisted of two evolutionary diverging subpopulations (Pearson, 1894). Since then, finite mixture models have found entry in various fields of science such as biology, economics and sociology.

Finite mixture models are statistical models based on the assumption that distributed observations can be adequately approximated by a parametric probability density. In fact, finite mixture models assume a generative model: each observation is generated by a random phenomenon represented by a probability density. This probability density can itself be a mixture of a finite number of probability densities termed mixture components. Therefore, mixture models are appropriate to model multimodal distributions of observations. Hence, mixture models provide an elegant approach for the closed mathematical quantification of cell-to-cell variability beyond the calculation of sample moments. Moreover, the nature of mixture models provides the possibility to assess the subpopulation structure of a cell population.

Subpopulations are groups of cells that have quantitatively different properties. Mixing of observations from both groups can result in bimodal distributions. The use of a mixture model for the quantification of such data characterizes the subpopulations. The term subpopulation is used throughout this thesis to refer to the components of a mixture.

Mathematically a mixture model \mathcal{M}_{mix} takes the form

$$p(x|\theta, u) = \sum_{s=1}^S w_s(u) \phi(x|\varphi_s(u)) \quad (4.1)$$
$$\text{with } \sum_{s=1}^S w_s(u) = 1.$$

This model describes the density p of the quantity of interest X in a heterogeneous population consisting of subpopulations $s = 1, \dots, S$ with relative subpopulation sizes $w_s(u)$ and subpopulation distribution parameters $\varphi_s(u)$ for experimental condition u . The parameter vector θ comprises all parameters of the mixture model. In addition to $\varphi_s(u)$ and $w_s(u)$, this might also be metaparameters, as detailed later in Section 4.3.2. The properties of the individual subpopulations are described by the probability densities $\phi(x|\varphi_s(u))$. The relative subpopulation sizes sum to one. Subpopulation sizes and mixture parameters can depend on the experimental conditions u .

In this thesis, we utilize mixture models for the comprehensive quantification of cell-to-cell variability beyond the calculation of sample moments. The probabilistic character of mixture models render them a perfect approach to handle censored data, as it will be described in Section 4.3.

Single-cell data analysis approaches based on mixture models

Mixture modelling of single-cell data is receiving increasing attention due to a rising number of single-cell technologies (Buettner *et al.*, 2015; Crane *et al.*, 2014; Hoppe *et al.*, 2014). Besides application in DNA content analysis (Wang & Huang, 2007), most software packages were developed for flow cytometry data (Lo *et al.*, 2008; Pyne *et al.*, 2009, 2014) but can also be used for other data types. While most mixture modelling approaches do not account for censoring (Johnsson *et al.*, 2016; Pyne *et al.*, 2009, 2014) others consider selected types of censoring (Lee & Scott, 2012; McLachlan & Jones, 1988). Unfortunately, the latter do not provide a comprehensive, easily accessible framework. Therefore, such models are infrequently applied in a biological context, which entails certain risks. In the presence of mutually exclusive (competing) biological events, for example, disregarding right censoring can result in an incorrect interpretation of experimental data such as correlations between actually uncorrelated data (Duffy *et al.*, 2012). Hence, there is a need for simple-to-use computational methods to analyse censored population data.

Besides censoring, another challenge for computational analysis methods of single-cell data is the integration of data from multiple experimental conditions (e.g. different strengths of stimuli or multiple sampling times after an intervention on the biological system at hand) or multiple technical and biological replicates. Established approaches use a two step procedure for this purpose. First, individual samples are described independently with finite mixture

models. Thereafter, matching-based methods are applied to link the different samples (Pyne *et al.*, 2009), e.g. to decide upon the appearance of identical subpopulations. These methods rely on similarity between distributions under different conditions. In the case of large changes in the corresponding distribution between experimental conditions, matching methods are not able to map the populations. To address this shortcoming, a Joint Clustering and Matching (JCM) approach (Pyne *et al.*, 2014) has been introduced. JCM allows for a more rigorous matching across samples and the consideration of inter-sample variability. For this, a template model is fitted to the pooled samples and the individual samples are modelled as instances of this template by adding random effect terms to the template parameters. This approach is well-suited for analysing the size of different distinct subpopulations in different samples. However, JCM does not facilitate an automatic matching of subpopulations across different experimental conditions, and - like all other methods - does currently not incorporate hypothesis testing methods.

4.2. Problem formulation

The prometaphase length data provide the information on how long the SAC is able to maintain a cell cycle arrest depending on the amount of signalling proteins. This time span is a measure of SAC functionality. The data were derived from strains with altered protein abundances of either one or simultaneously two SAC signalling proteins. These SAC prometaphase length data exhibit significant cell-to-cell variability. Data of some strains appear to have a bimodal distribution. For many cells, especially for those with longer prometaphase lengths, the end of prometaphase was not recorded, which causes right censored data. Furthermore, the data is interval censored due to an inter observation time of five minutes. Cell-to-cell variability can be a precious source of information (see Section 2.3.1).

Problem 4.1. (Formulation of a modelling framework) *Given the observed cell-to-cell variability and the problem of data censoring in the SAC data, develop a modelling framework that is tailored to quantify the variability in censored single-cell data from multiple experimental conditions and to detect subpopulations where present.*

For an interpretation of the effects of the experimental interventions on SAC signalling, a rigorous statistical quantification and analysis of the data is crucial. We are especially interested in the characteristics of cell-to-cell variability in the SAC prometaphase lengths datasets in terms of numbers and in terms of subpopulations that suggest qualitative differences in SAC signalling.

Problem 4.2. (Characterization of cell-to-cell variability in SAC data) *Given the datasets on prometaphase lengths, quantify cell-to-cell variability in the SAC datasets regarding quantitative cell-to-cell variability and qualitative differences as indicated by subpopulations.*

The effect of an altered protein amount depends on the particular protein. Since the degree of alteration was determined, we are able to quantify the sensitivity the SAC exhibits towards each of the assessed proteins.

Problem 4.3. (Characterization of SAC sensitivity to perturbations) *Given the information on the corresponding perturbation in protein amount for every strain in addition to the observed prometaphase data, quantify the sensitivity of SAC signalling with respect to different SAC proteins via the functional dependency of SAC functionality on the relative amount of protein.*

In some strains not only one protein was altered but two of them simultaneously. What information can such a double perturbation provide on the mode of interaction of these proteins?

Problem 4.4. (Characterization of SAC signalling mechanism) *Given data from strains with double perturbations, characterize the mode of interaction of these proteins on SAC functionality.*

4.3. Multi-experiment mixture modelling of censored single-cell data (MEMO)

This section presents a statistical modelling framework for the analysis of censored single-cell data, termed Multi-Experiment mixture MOdelling (MEMO). Therewith we address Problem 4.1. The modelling framework is based on the application of finite mixture models to assess quantitative and qualitative variability in the data while considering multiple datasets simultaneously.

In Section 4.3.1, we derive the statistical models for different types of censoring. The subsequent sections describe the individual steps of the modelling work flow, hypothesis formulation (Section 4.3.2), model parametrization (Section 4.3.3), model selection (Section 4.3.4), and interpretation and further analysis (Section 4.3.5) and the methods and approaches applied in these steps. For a comprehensive overview, the work flow is illustrated in Figure 4.1. In Section 4.3.6 we apply our framework to two prometaphase datasets of known subpopulation structure, to illustrate the effects of censoring on inference. An implementation of the presented framework is available as the MATLAB toolbox MEMO (Multi-experiment mixture MOdelling). MEMO is available on github ([MEMO-toolbox.github.io/MEMO/](https://github.com/MEMO-toolbox/MEMO/)).

4.3.1. Statistical models in the presence of censoring

In this section, we derive the statistical models needed to model data which are subject to censoring. The statistical model of a dataset is the parametric probability density that describes the distribution from which the observations in this dataset are sampled. Censoring transforms this data generating density into an observable density. While we are interested in the data generating density, the experimental data are realizations from the observed density and first and foremost contain information on this density. However, probability theory allows for the derivation of the observed density as a function of the generating densities by taking into account which type of censoring the data is subject to. Therefore, observed density as function of the generating densities is our statistical model for the variable of interest. In the following, we derive the observed densities in case of interval censoring, right censoring and the combination of interval and right censoring in the i th experiment within a set of experiments.

We consider random censoring in a general setup where cells can undergo two mutually exclusive randomly distributed events of interest, such as cell death and cell division. However, one of the events may as well be a censoring event due to the end of the observation time. Then we have only one event of interest which is mutually exclusive with a censoring event. We denote Event I in the i th experiment with the random variable X_i and Event II in the i th experiment with random variable C_i . The i th experiment is described by the input variable u_i . X_i and C_i are independent random variables with probability densities $f_{X_i}(x_i|\theta, u_i)$ and $f_{C_i}(c_i|\theta, u_i)$, respectively. The cumulative distributions of X_i and C_i are denoted by $F_{X_i}(x_i|\theta, u_i)$ and $F_{C_i}(c_i|\theta, u_i)$, respectively. The densities $f_{X_i}(x_i|\theta, u_i)$ and $f_{C_i}(c_i|\theta, u_i)$ are in general assumed to be given by a mixture model as defined in Equation (4.1). Censoring transforms these densities into the observed densities. Therefore, we use additional random variables associated with the observed densities. They are introduced where needed in the following sections. The observed densities depend on the type of censoring and the respective models and corresponding likelihood functions will be derived in the following.

The statistical model in the absence of censoring

For completeness we start with the model for data in the absence of censoring. If only one event is possible or the supports of the event generating densities do not overlap, no censoring occurs, all observations are exact and the data are uncensored or complete. Under these circumstances the generating densities and observed densities are identical.

Consider the case in which X_i is the only event to occur and our measurement process does not cause censoring. We denote the random variable representing the observations with Y_i . For i.i.d. observations, uncensored data $\mathcal{D}_i = \{y_i^j\}_{j=1 \dots n_{y,i}}$ of $n_{y,i}$ observations are direct samples from the data generating density and the probability density of Y_i is

$$f_{Y_i}(y_i|\theta, u_i) = f_{X_i}(x_i|\theta, u_i).$$

Here the data provide information about the full data generating probability density, enabling reliable reconstruction for sufficiently large sample numbers $n_{y,i}$. This does not ensure that the parameters θ are identifiable. For mixture models, for instance, the problem of symmetry is well-known (Stephens, 2000).

In the absence of censoring, the likelihood function for data \mathcal{D}_i is given by

$$\mathbb{P}(\mathcal{D}_i|\theta) = \prod_{j=1}^{n_{y,i}} f_{Y_i}(y_i^j|\theta, u_i).$$

The statistical model accounting for interval censoring

For interval censoring we denote the random variable representing the observed censored quantity in the i th experiment with \bar{Y}_i . An interval censored observation \bar{y}_i provides the information that the corresponding exact value x_i lies in the interval $(\bar{y}_i - \Delta x, \bar{y}_i]$. The interval length is denoted by Δx . Accordingly, for experimental condition u_i the dataset consists of

realizations from

$$\begin{aligned} f_{\bar{Y}_i}(\bar{y}_i|\theta, u_i) &= \int_{\bar{y}_i - \Delta x}^{\bar{y}_i} f_{X_i}(x_i|\theta, u_i) dx_i \\ &= F_{X_i}(\bar{y}_i|\theta, u_i) - F_{X_i}(\bar{y}_i - \Delta x|\theta, u_i) \end{aligned}$$

with cumulative distribution

$$F_{X_i}(x_i|\theta, u_i) := \int_{-\infty}^{x_i} f_{X_i}(x'_i|\theta, u_i) dx'_i.$$

Interval censored data $\mathcal{D}_i = \{\bar{y}_i^l\}_{l=1, \dots, n_{\bar{y}_i}}$ provide information about the probability mass between two observation points. The precise shape of the probability density between observation points cannot be reconstructed but is merely restricted by the chosen distribution type. In the presence of interval censoring, the likelihood function for data \mathcal{D}_i is

$$\begin{aligned} \mathbb{P}(\mathcal{D}_i|\theta) &= \prod_{l=1}^{n_{\bar{y}_i}} f_{\bar{Y}_i}(\bar{y}_i^l|\theta, u_i) \\ &= \prod_{l=1}^{n_{\bar{y}_i}} \left(F_{X_i}(\bar{y}_i^l|\theta, u_i) - F_{X_i}(\bar{y}_i^l - \Delta x|\theta, u_i) \right). \end{aligned}$$

Here we assume that the length of all intervals is identical. This can easily be generalized.

The statistical model accounting for right censoring

For the derivation of the model, we consider two competing processes, one generating actual observations of the process of interest and the second generating observations such as the end of recording. Mutual exclusiveness in the context of right censoring has the effect that only the event occurring first can be detected and recorded as described in Section 4.1.1. In the presence of random right censoring due to a competing process, observations of the quantity of interest $\{y_i^j\}$ and observations of censoring $\{\bar{y}_i^k\}$ are recorded. These are realizations of the conditional random variables $Y_i := X_i|X_i \leq C_i$ and $\bar{Y}_i := C_i|C_i \leq X_i$, respectively. In the following we derive the densities of Y_i and \bar{Y}_i from the densities of X_i and C_i .

The densities of observed uncensored and right censoring observations for experimental condition u_i are the probability densities

$$\begin{aligned} f_{Y_i}(y_i|\theta, u_i) &= f_{X_i|X_i \leq C_i}(x_i|\theta, u_i) \\ &= \frac{f_{X_i, X_i \leq C_i}(x_i|\theta, u_i)}{P(X_i \leq C_i|\theta, u_i)}, \\ f_{\bar{Y}_i}(\bar{y}_i|\theta, u_i) &= f_{C_i|C_i \leq X_i}(c_i|\theta, u_i) \\ &= \frac{f_{C_i, C_i \leq X_i}(c_i|\theta, u_i)}{P(C_i \leq X_i|\theta, u_i)}, \end{aligned}$$

with joint distributions (derivation provided in Appendix A)

$$\begin{aligned} f_{X_i, X_i \leq C_i}(x_i|\theta, u_i) &= f_{X_i}(x_i|\theta, u_i)(1 - F_{C_i}(x_i|\theta, u_i)), \\ f_{C_i, C_i \leq X_i}(x_i|\theta, u_i) &= f_{C_i}(c_i|\theta, u_i)(1 - F_{X_i}(c_i|\theta, u_i)), \end{aligned}$$

and marginal probabilities for observing a valid or a censoring observation

$$\begin{aligned} P(X_i \leq C_i|\theta, u_i) &= \int_{-\infty}^{\infty} f_{X_i}(x_i|\theta, u_i)(1 - F_{C_i}(x_i|\theta, u_i))dx_i, \\ P(C_i \leq X_i|\theta, u_i) &= \int_{-\infty}^{\infty} f_{C_i}(c_i|\theta, u_i)(1 - F_{X_i}(c_i|\theta, u_i))dc_i. \end{aligned}$$

As analytical solutions of $P(X_i \leq C_i|\theta, u_i)$ and $P(C_i \leq X_i|\theta, u_i)$ are often not available, numerical integration might be necessary (Cook, 2008).

The density of C_i can have different shapes. In the case of random censoring, meaning that $f_{C_i}(c_i|\theta, u_i)$ is a smooth distribution, the likelihood function for data

$$\mathcal{D}_i = \left\{ \{y_i^j\}_{j=1, \dots, n_{y,i}}, \{\bar{y}_i^k\}_{k=1, \dots, n_{\bar{y},i}} \right\}$$

is proportional to

$$\begin{aligned} \mathbb{P}(\mathcal{D}_i|\theta) &\propto \left(\prod_{j=1}^{n_{y,i}} f_{Y_i}(y_i^j|\theta, u_i) \right) \left(\prod_{k=1}^{n_{\bar{y},i}} f_{\bar{Y}_i}(\bar{y}_i^k|\theta, u_i) \right) \\ &\propto \left(\prod_{j=1}^{n_{y,i}} f_{X_i}(y_i^j|\theta, u_i)(1 - F_{C_i}(y_i^j|\theta, u_i)) \right) \left(\prod_{k=1}^{n_{\bar{y},i}} f_{C_i}(\bar{y}_i^k|\theta, u_i)(1 - F_{X_i}(\bar{y}_i^k|\theta, u_i)) \right). \end{aligned}$$

In case of fixed Type I censoring at a single value \tilde{c}_i such that $\{\bar{y}_i^k\}_{k=1, \dots, n_{\bar{y},i}} = \tilde{c}_i \forall k$, which corresponds to a probability density which is a Dirac delta, $f_{\tilde{c}_i}(c_i|\theta, u_i) = \delta(c_i - \tilde{c}_i)$, the likelihood function simplifies to

$$\mathbb{P}(\mathcal{D}_i|\theta) \propto \left(\prod_{j=1}^{n_{y,i}} f_{X_i}(y_i^j|\theta, u_i) \right) \left(\prod_{k=1}^{n_{\bar{y},i}} (1 - F_{X_i}(\bar{y}_i^k|\theta, u_i)) \right).$$

This formulation exploits the tail probabilities $1 - F_{X_i}(\bar{y}_i^k|\theta, u_i)$ to capture the censoring.

Note that this likelihood function can also be used to avoid explicit modelling of the censoring process as a probability density. While this still allows for inference, a visual comparison of model and data requires an estimate of the censoring density (Geissen *et al.*, 2016), since f_{Y_i} and $f_{\bar{Y}_i}$ have to be evaluated for this purpose. Furthermore, both, $f_{X_i}(x_i|\theta, u_i)$ and $f_{C_i}(c_i|\theta, u_i)$, are needed to resample data for a goodness-of-fit analysis based on bootstrapping of the likelihood distribution of the objective function.

The statistical model accounting for interval and right censoring

In the presence of interval and right censoring, interval censored observations $\{\bar{y}_i^j\}$ and right censored observations $\{\bar{y}_i^k\}$ are recorded in experimental condition i . These observations are

realizations of the random variables \bar{Y}_i and \bar{Y}_i , respectively. To derive \bar{Y}_i and \bar{Y}_i and their respective densities from X_i and C_i we need to make an intermediate step and create the random variables X_i^+ and C_i^+ first. X_i^+ and C_i^+ are derived from X_i and C_i by discretisation. Loosely speaking, realizations of X_i and C_i are binned according to the censoring interval Δx . Binning here equals a round up of x_i and c_i to the next multiple of Δx . This yields the smallest multiple of Δx , x_i^+ , which is larger than x_i , and correspondingly c_i^+ . Without loss of generality we assume that measured time points are multiples of Δx , such that $\forall i, j \exists k', k''$ such that $\bar{y}_i^j = k' \Delta x$ and $\bar{y}_i^k = k'' \Delta x$. The densities of the conditional random variables $\bar{Y}_i := X_i^+ | X_i^+ \leq C_i^+$ and $\bar{Y}_i := C_i^+ | C_i^+ \leq X_i^+$ for experimental condition u_i are then derived as

$$\begin{aligned} f_{\bar{Y}_i}(\bar{y}_i^j | \theta, u_i) &= f_{X_i^+ | X_i^+ \leq C_i^+}(x_i^+ | \theta, u_i) \\ &= \frac{f_{X_i^+, X_i^+ \leq C_i^+}(x_i^+ | \theta, u_i)}{P(X_i^+ \leq C_i^+ | \theta, u_i)}, \\ f_{\bar{Y}_i}(\bar{y}_i^k | \theta, u_i) &= f_{C_i^+ | C_i^+ \leq X_i^+}(c_i^+ | \theta, u_i) \\ &= \frac{f_{C_i^+, C_i^+ \leq X_i^+}(c_i^+ | \theta, u_i)}{P(C_i^+ \leq X_i^+ | \theta, u_i)}, \end{aligned}$$

with joint distributions

$$\begin{aligned} f_{X_i^+, X_i^+ \leq C_i^+}(x_i^+ | \theta, u_i) &= (F_{X_i}(x_i^+ | \theta, u_i) - F_{X_i}(x_i^+ - \Delta x | \theta, u_i))(1 - F_{C_i}(x_i^+ | \theta, u_i)), \\ f_{C_i^+, C_i^+ \leq X_i^+}(c_i^+ | \theta, u_i) &= (F_{C_i}(c_i^+ | \theta, u_i) - F_{C_i}(c_i^+ - \Delta x | \theta, u_i))(1 - F_{X_i}(c_i^+ | \theta, u_i)), \end{aligned}$$

and marginal probabilities for observing uncensored or censored data,

$$\begin{aligned} P(X_i^+ \leq C_i^+ | \theta, u_i) &= \sum_{k' \in \mathbb{Z}} (F_{X_i}(k' \Delta x | \theta, u_i) - F_{X_i}((k' - 1) \Delta x | \theta, u_i))(1 - F_{C_i}(k' \Delta x | \theta, u_i)), \\ P(C_i^+ \leq X_i^+ | \theta, u_i) &= \sum_{k'' \in \mathbb{Z}} (F_{C_i}(k'' \Delta x | \theta, u_i) - F_{C_i}((k'' - 1) \Delta x | \theta, u_i))(1 - F_{X_i}(k'' \Delta x | \theta, u_i)). \end{aligned}$$

The cumulative distributions of X_i and C_i are denoted by $F_{X_i}(x_i | \theta, u_i)$ and $F_{C_i}(c_i | \theta, u_i)$, respectively.

In the case of random censoring the likelihood function for data

$$\mathcal{D}_i = \left\{ \left\{ \bar{y}_i^l \right\}_{l=1, \dots, n_{\bar{y}_i}}, \left\{ \bar{y}_i^k \right\}_{k=1, \dots, n_{\bar{y}_i}} \right\}$$

is proportional to

$$\begin{aligned} \mathbb{P}(\mathcal{D}_i | \theta) &\propto \left(\prod_{l=1}^{n_{\bar{y}_i}} f_{\bar{Y}_i}(\bar{y}_i^l | \theta, u_i) \right) \left(\prod_{k=1}^{n_{\bar{y}_i}} f_{\bar{Y}_i}(\bar{y}_i^k | \theta, u_i) \right) \\ &\propto \left(\prod_{l=1}^{n_{\bar{y}_i}} (F_{X_i}(\bar{y}_i^l | \theta, u_i) - F_{X_i}(\bar{y}_i^l - \Delta x | \theta, u_i))(1 - F_{C_i}(\bar{y}_i^l | \theta, u_i)) \right) \\ &\quad \left(\prod_{k=1}^{n_{\bar{y}_i}} (F_{C_i}(\bar{y}_i^k | \theta, u_i) - F_{C_i}(\bar{y}_i^k - \Delta x | \theta, u_i))(1 - F_{X_i}(\bar{y}_i^k | \theta, u_i)) \right). \end{aligned}$$

As before, for fixed Type I censoring at a value \tilde{c}_i , the likelihood function simplifies to

$$\mathbb{P}(\mathcal{D}_i|\theta) \propto \left(\prod_{l=1}^{n_{\tilde{y}_i}} f_{X_i}(\tilde{y}_i^l|\theta, u_i) \right) \left(\prod_{k=1}^{n_{\tilde{y}_i}} (1 - F_{X_i}(\tilde{y}_i^k|\theta, u_i)) \right).$$

4.3.2. Formulation of model hypotheses

The probably most intuitive way to formulate different hypotheses in a mixture modelling framework is the choice of the distribution type to model the data. We considered normal, log-normal, gamma and Johnson SU distributions (see Appendix B), but in general every parametric distribution can be employed. Since the framework proposed here allows for the simultaneous study of multiple experimental datasets, more general hypotheses can be assessed, such as for example the subpopulation structure, i.e. the number and identity of subpopulations. While the formulation of hypotheses on the number of subpopulations is straightforwardly encoded in the number of mixture components, hypotheses on the identity of subpopulations across several experimental conditions needs knowledge or at least intuition on the data generating process. The identity of subpopulations across different experimental conditions comprises two variants. Either the subpopulation conserves its properties or its properties evolve depending on the experimental condition. Hypotheses regarding condition-dependent population structures can be encoded in the functional dependency of the subpopulation sizes $w_s(u, \tilde{\theta})$ and distribution parameters $\varphi_s(u, \tilde{\theta})$ on the experimental condition u and meta-parameters $\tilde{\theta}$. These functional dependencies might be derived from the knowledge of the biological mechanism underlying the data or in a more phenomenological way by using semi-mechanistic models. Both variants are illustrated in the following sections.

Mechanistic model hypotheses

Appropriate functions $w_s(u, \tilde{\theta})$ and $\varphi_s(u, \tilde{\theta})$ can often be derived from the properties of the biological system, e.g., the underlying signalling pathway. This has been demonstrated for uncensored data (Hasenauer *et al.*, 2014a) and can be generalized to the case of censored data. Possible models include reaction rate equations (Hasenauer *et al.*, 2014a), linear noise approximations (Elf & Ehrenbarg, 2003; van Kampen, 2007), effective mesoscopic rate equations (Grima, 2010; Ramaswamy *et al.*, 2012) or moment equations (Engblom, 2006; Lee *et al.*, 2009). These ODE models allow for a mechanistic description of the single-cell and population dynamics, in particular the explicit consideration of intrinsic and/or extrinsic noise (Swain *et al.*, 2002). An example for the use of mechanistic models in the framework presented here can be found in Geissen *et al.* (2016).

Semi-mechanistic model hypotheses

For many processes no mechanistic models are available. In this case, general functions can be used to model $w_s(u, \tilde{\theta})$ and $\varphi_s(u, \tilde{\theta})$. These expressions can be specific for a certain mixture component and depend on the experimental condition. In case of a normal distribution with mixture parameters $\varphi(u) = (\mu(u), \sigma(u))$, exemplary hypotheses regarding the condition-dependence of the mean are for example:

- The mean is constant for all u : $\mu(u, \tilde{\theta}) = b$ with $\tilde{\theta} = b$.
- The mean changes linearly with u : $\mu(u, \tilde{\theta}) = m \cdot u + b$ with $\tilde{\theta} = (m, b)^T$.
- The mean increases with u in a Hill-type manner: $\mu(u, \tilde{\theta}) = \frac{\mu_{\max} u^n}{K^n + u^n}$ with $\tilde{\theta} = (\mu_{\max}, K, n)^T$.
- The mean decreases with u in a Hill-type manner: $\mu(u, \tilde{\theta}) = \frac{\mu_{\max} K^n}{K^n + u^n}$ with $\tilde{\theta} = (\mu_{\max}, K, n)^T$.

Similar or more general dependencies might be assumed for any function $w_s(u, \tilde{\theta})$ and $\varphi_s(u, \tilde{\theta})$. In the presence of multiple inputs also products of univariate functions and/or multi-variate functions might be used. An example for the use of semi-mechanistic models is provided below in Section 4.4.2.

4.3.3. Model parametrization

The unknown parameters of the mixture model parameters and possible meta-parameters have to be inferred from the available data. For that purpose the likelihood functions derived in Section 4.3.1 for individual experimental conditions i are used to construct the overall likelihood function. The joint likelihood function for multiple experimental datasets is given as the product of the likelihood functions for the individual datasets,

$$\mathbb{P}(\mathcal{D}|\theta) = \prod_{i=1}^I \mathbb{P}(\mathcal{D}_i|\theta).$$

To determine the maximum likelihood estimate of the parameters the approaches introduced in Section 2.4.1 can be used. MEMO makes use of the multistart optimization implementations within PESTO (Parameter Estimation TOOLbox, unpublished toolbox of Jan Hasenauer). In short, Latin hypercube sampling is used to generate starting points for the MATLAB routine `fmincon`. As default the interior point method is used with default setting (see MATLAB documentation for details). Since an analytical solution of the gradient of the objective function can be derived (provided in Appendix C), we provide `fmincon` with the analytical gradient of the objective function and the constraints. This facilitates convergence and reduces the computational cost as no finite difference approximation of the gradient is necessary.

4.3.4. Hypotheses testing via model selection

For model selection we used the Bayesian information criterion (BIC) for the results presented in this Chapter. MEMO furthermore implements the Likelihood ratio test and the Akaike information criterion (AIC).

Bayesian information criterion

The BIC is a probabilistic model selection criterion, derived using Bayesian arguments, which accounts for fit and model complexity. The BIC is defined as

$$\text{BIC}_k = -2 \log(\mathbb{P}(\mathcal{D}|\theta_k^{\text{ML}})) + n_{\theta,k} \log n_{\mathcal{D}},$$

where k is used to denote a particular model, θ_k^{ML} is the maximum likelihood estimate of the parameter vector, $n_{\theta,k}$ denotes the number of parameters in model k , and $n_{\mathcal{D}}$ denotes the number of observations in the dataset. The weight of the model complexity depends on the sample size $n_{\mathcal{D}}$. Models with lower BICs are preferable. The best model $\mathcal{M}_{\text{mix},k^*}$ (according to BIC) possesses the minimal BIC, $\text{BIC}_{\min} = \text{BIC}_{k^*}$ with $k^* = \arg \min_k \text{BIC}_k$. In practice, given the best model $\mathcal{M}_{\text{mix},k^*}$, the evidence against model $\mathcal{M}_{\text{mix},k}$ is considered to be

- *not worth more than a bare mention* for $0 < \text{BIC}_k - \text{BIC}_{\min} < 2$,
- *positive* for $2 < \text{BIC}_k - \text{BIC}_{\min} < 6$,
- *strong* for $6 < \text{BIC}_k - \text{BIC}_{\min} < 10$, and
- *very strong* for $10 < \text{BIC}_k - \text{BIC}_{\min}$.

For details we refer to Kass & Raftery (1995). In this thesis we considered a difference in the BIC values of 6 to be substantial.

Automated unravelling of subpopulation structures

The comparison of many alternative hypotheses can become cumbersome as it requires the definition of many models. In a setup with I experimental conditions, if for example linked by a common censoring process, the number of models to decide on one or two subpopulations in each condition is 2^I . We propose an automated backward model selection approach for the automated unravelling of the subpopulation structure. For single- and multi-experiment single-cell data, the implemented backward model selection iteratively reduces the model by eliminating subpopulations which are not supported by the experimental data. For this, all possible models which possess in the experimental conditions one subpopulation less than the current model are generated, compiled and fitted. Among the fitted models the best one is selected using the specified selection criteria along with a cut-off threshold. This process is repeated until no further improvement of the model is achieved. In comparison to a brute-force approach not all model alternatives are analysed, which decreases the computational complexity significantly. Furthermore, the automatic generation reduces the work load for the user and enables the analysis of thousands of model alternatives.

4.3.5. Interpretation and further analysis

Parameter estimation and model selection provide a set of mixture models capturing the data. MEMO provides different routines to visualize these models as well as a model-data comparison. Beyond the sole analysis of available data, multi-experiment mixture models enable predictions. For example, mixture models with relative subpopulation sizes $w_s(u)$ and subpopulation mixture parameters $\varphi_s(u)$ that functionally depend on the experimental condition can

be used to predict the outcomes of future experiments with different experimental conditions. Therefore, the mixture properties just have to be evaluated for the respective experimental setting u . MEMO also allows for the assessment of the prediction uncertainty, e.g. by using parameter samples obtained via MCMC sampling. This enables a thorough analysis of the predictive power of models and model validation.

The mixture model and its parameters can furthermore be used for subsequent analysis of the data generating process. The calculated subpopulation sizes $w_s(u)$ and subpopulation mixture parameters $\varphi_s(u)$ can, for instance, be used to unravel key properties of the underlying signalling pathway such as cooperative interactions between proteins. The processed data can even be used to inform mechanistic models. This was demonstrated in several recent publications (Duffy *et al.*, 2012; Heinrich *et al.*, 2013; Song *et al.*, 2010). For the use in subsequent analysis, MEMO provides estimates and confidence or credibility intervals for all essential model quantities.

4.3.6. Assessment of the impact of censoring on inference in SAC datasets

For an illustration of the effects of censoring on the inference of the subpopulation structure we compared the performance of MEMO with mixture modelling that disregards censoring. We chose two strains with known subpopulation structure. One strain has a completely dysfunctional SAC. This is reflected in the data set by tight unimodal short prometaphase lengths (Figure 4.2). The other strain has a functional SAC and therefore exhibits much longer prometaphase lengths, including a large portion of right censored observations (Figure 4.3). Using the BIC as model selection criterion to determine the number of subpopulations, mixture modelling ignoring censoring suggests a model with a mixture of two log-normal subpopulations for both data sets (Figure 4.2A and Figure 4.3A), while MEMO is able to identify the biologically plausible result of single log-normal populations (Figure 4.2B and Figure 4.3B).

4.4. Mixture model based quantification and analysis of variability in SAC single-cell microscopy data

Under normal conditions, the SAC precludes cell cycle progression as long as unattached chromosomes are under risk to get incorrectly segregated. To gain insights into the robustness of this mechanism, it was challenged by the modulation of critical proteins and the impact was assessed by measuring the time span the SAC is able to arrest the cell cycle when triggered (Heinrich *et al.*, 2013). Since the obtained data is subject to massive right censoring due to a limited observation time, the quantification, even in terms of sample moments, is non-trivial. Furthermore, a first look at these SAC single-cell prometaphase length data (Figure 4.5, A and C) raises two questions. The first question concerns the subpopulation structure of certain strains. Do they consist of subpopulations with different cellular phenotypes? The second question is related to the identity of the subpopulations. In particular, are there subpopulations that occur across the different strains, i.e. are not dependent on the experimental treatment? The first question is answered in Section 4.4.1. We used model selection to de-

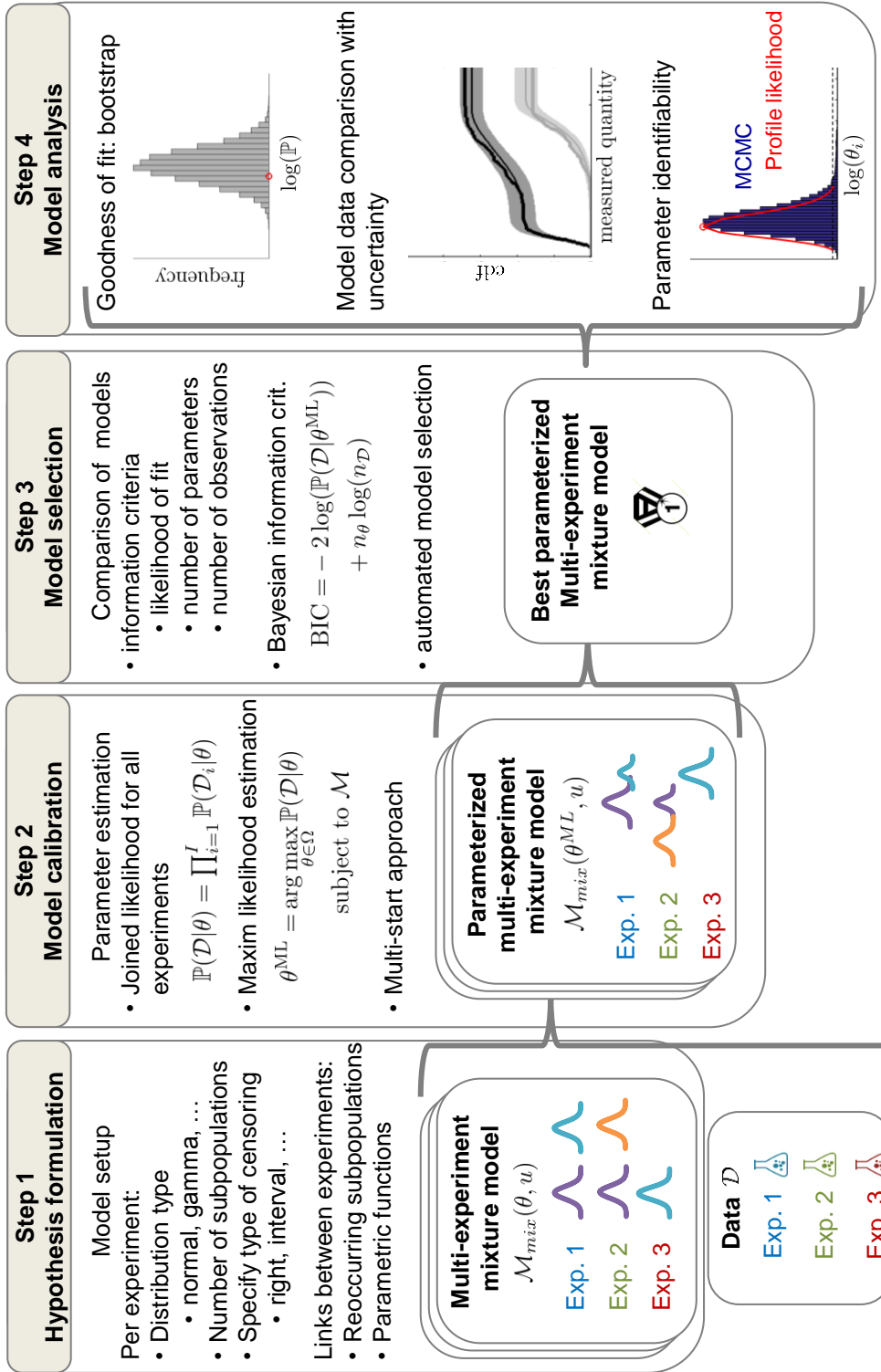


Figure 4.1.: Illustration of the proposed multi-experiment mixture modelling framework MEMO. The four main steps in the workflow are shown together with the respective outcome of each step.

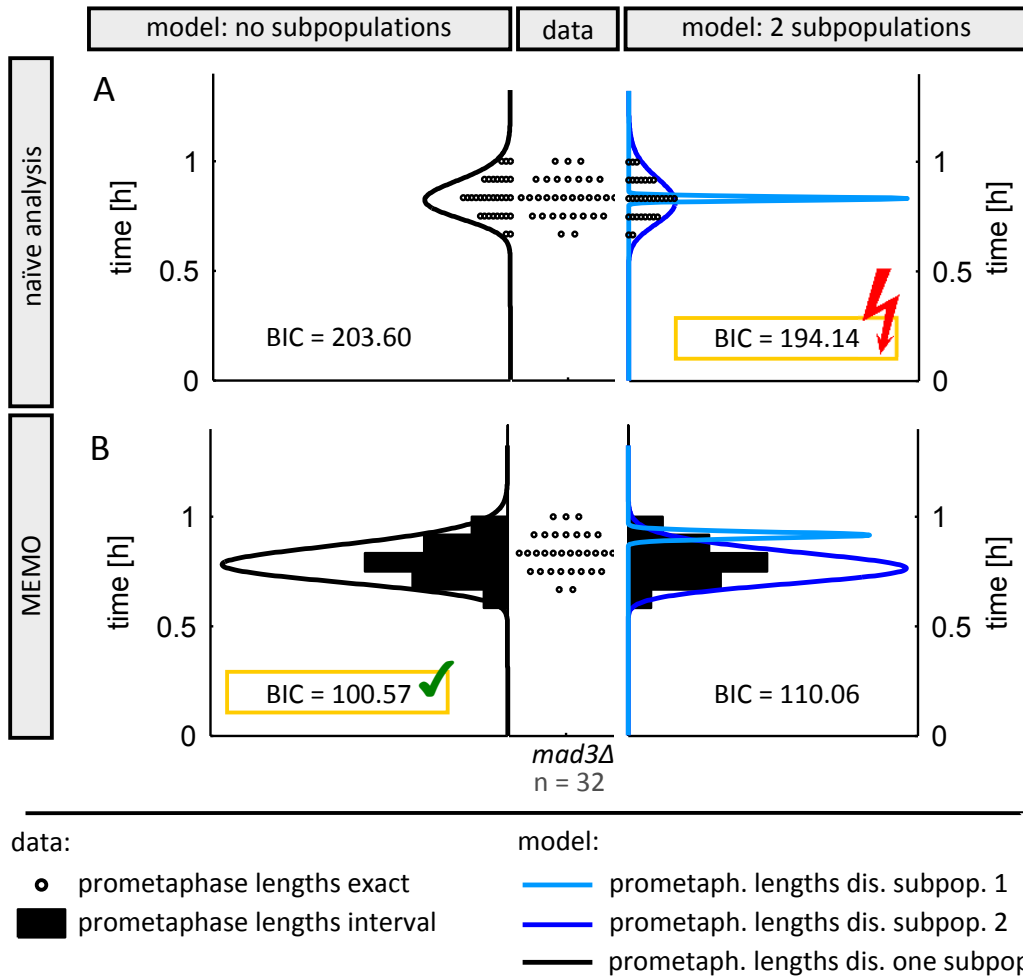


Figure 4.2.: Interval censoring has to be considered for accurate reconstruction of SAC functionality from fluorescence live-cell microscopy imaging. Circles and black bars indicate cells in which the entire prometaphase was recorded (interval censored data). For the *mad3Δ* strain (0% Mad3, dysfunctional SAC) unimodal prometaphase lengths are observed. We used the BIC to decide upon the number of subpopulations for different settings. (A) A naïve analysis, disregarding interval censoring, selects a statistical model with two subpopulations, while (B) MEMO selects a model with a single population. Figure adapted from Geissen *et al.* (2016).

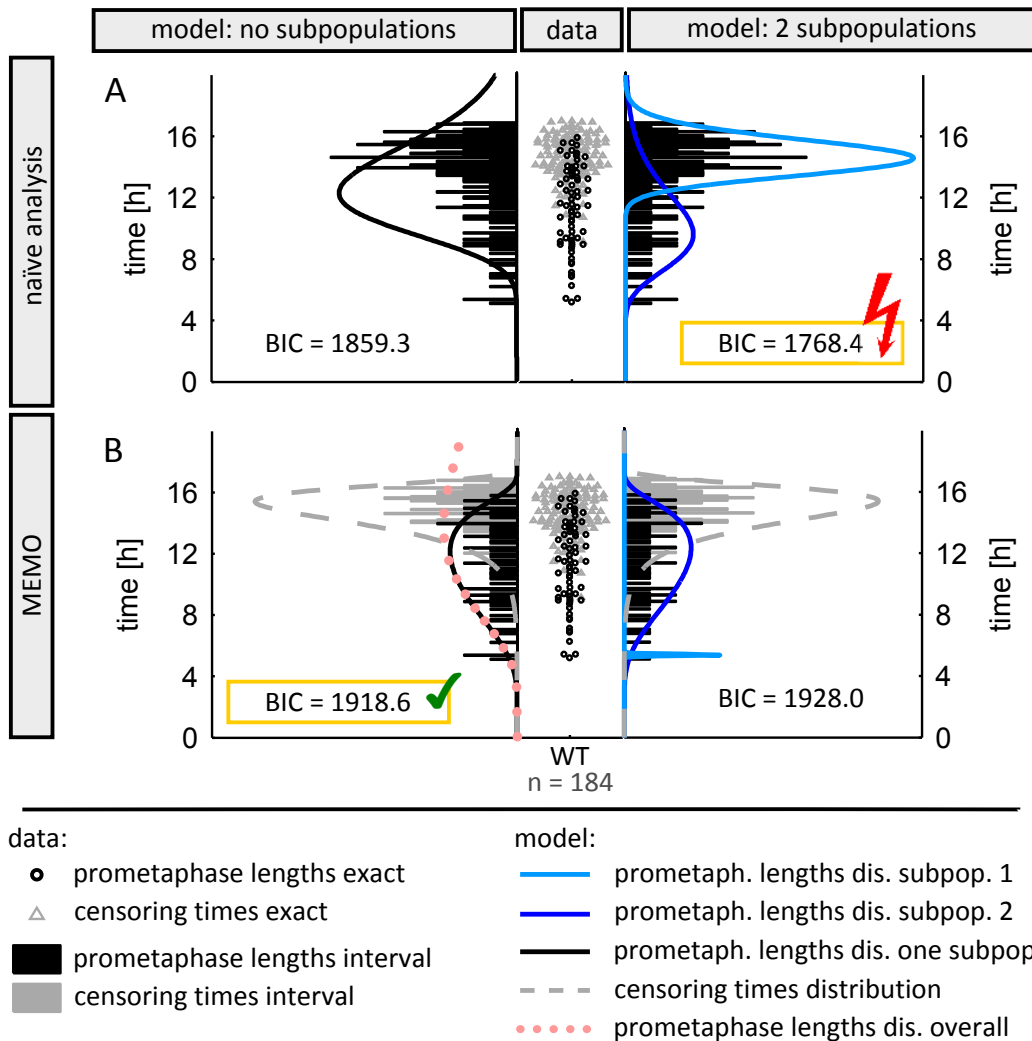


Figure 4.3.: Right censoring has to be considered for accurate reconstruction of SAC functionality from fluorescence live-cell microscopy imaging. Circles and black bars indicate cells in which the entire prometaphase was recorded (interval censored data). Triangles and gray bars indicate cells that were still in prometaphase when recording stopped (right censored data). For the wild type (WT) strain a large portion of right censored data are observed. (A) The commonly used approach to set prometaphase length of censored data to censoring time selects a statistical model with two subpopulations, while (B) MEMO selects a model with a single subpopulation and Johnson SU distributed censoring times. The pink dotted line depicts the reconstructed overall distribution of prometaphase lengths (i.e. the distribution that would be observed if the observation time was infinite) obtained using MEMO. Figure adapted from Geissen *et al.* (2016).

termine the number of subpopulations for every strain and quantified the full distribution of the variability in these subpopulations. Section 4.4.1 also answers the second question by identifying wild type like subpopulations across several non-wild type strains.

While the two questions above merely refer to sole data analysis and interpretation, the data may also contain information on the mechanism of the signalling pathway. What do the data tell us about the sensitivity of SAC functionality with respect to alterations in the respective proteins? And moreover, can we already learn something about the role of the single proteins without modelling the details and dynamics of the signalling mechanism? The sensitivity of SAC signalling regarding changes in the abundance of key proteins is quantified in Section 4.4.2. The latter question is addressed in Section 4.4.3, where we disclose whether Mad2 and Mad3 independently influence SAC functionality or whether they act cooperatively.

We denote the interval censored observations of prometaphase lengths as prometaphase lengths and the right censored observations with censoring times. Since cells enter mitosis asynchronously, while the end of observation is fixed to 17 hours after the start of the observation, the censoring times are distributed.

4.4.1. Inference of subpopulation structure from prometaphase data

In this section we address Problem 4.2 by analysing the quantitative and qualitative cell-to-cell variability in prometaphase lengths under SAC activating conditions. We use the framework introduced in the previous section to assess the subpopulation structure in strains differing in the expressed amounts of proteins (Mad2 or Mad3), which are essential for SAC functionality, in a hypothesis-driven approach. The subpopulation structure is given by the number, the properties and the identity of the distributions of prometaphase lengths in each strain.

Inference of the number of subpopulations

We use weighted mixtures of different distribution types and different numbers of mixture components to assess the qualitative and quantitative properties of the variability in prometaphase lengths in the different strains. Experimental data for all strains in which Mad2 or Mad3 abundance is altered are shown in Figure 4.5A and B, respectively. The recorded prometaphase lengths indicate the presence of cellular subpopulations with functional and dysfunctional SAC for certain strains. Cells with a functional SAC, e.g. wild type cells, have a minimum prometaphase time of at least 5 hours, while cells with dysfunctional SAC have shorter prometaphase lengths (Heinrich *et al.*, 2013). For strains with 65% and 80% Mad2 expression and strains with 30% and 60% Mad3 expression, subpopulations with either property seem to be present. The same parametrization of the distribution of right censoring times is used for all strains, as censoring is statistically identical. The parameters of the distributions of prometaphase lengths are estimated along with the subpopulation sizes, i.e. the weights $w_s(u)$ of the distributions, and the parameters of the censoring time distribution. For the sake of simplicity we analyse the datasets of strains with altered abundance of Mad2 and Mad3 separately, although it would be possible to analyse all datasets simultaneously in our framework.

The most complex statistical model, i.e. a model with two mixture components for every strain except the wild type, possesses many parameters. In order to find the minimal descrip-

tion of the data, MEMO is used to perform backward model selection (Figure 4.4A, setup 2). The successive simplification of the initial model resulted in a 33% reduction of the number of parameters. The most plausible model identifies two subpopulations in both 65% Mad2 strains and the 80% Mad2 strain, and only one subpopulation in all other strains (Figure 4.4A, setup 2). The analysis of the Mad3 strains identified two subpopulations in the 30% Mad3 and the 60% Mad3 strains (Figure 4.4B, setup 2).

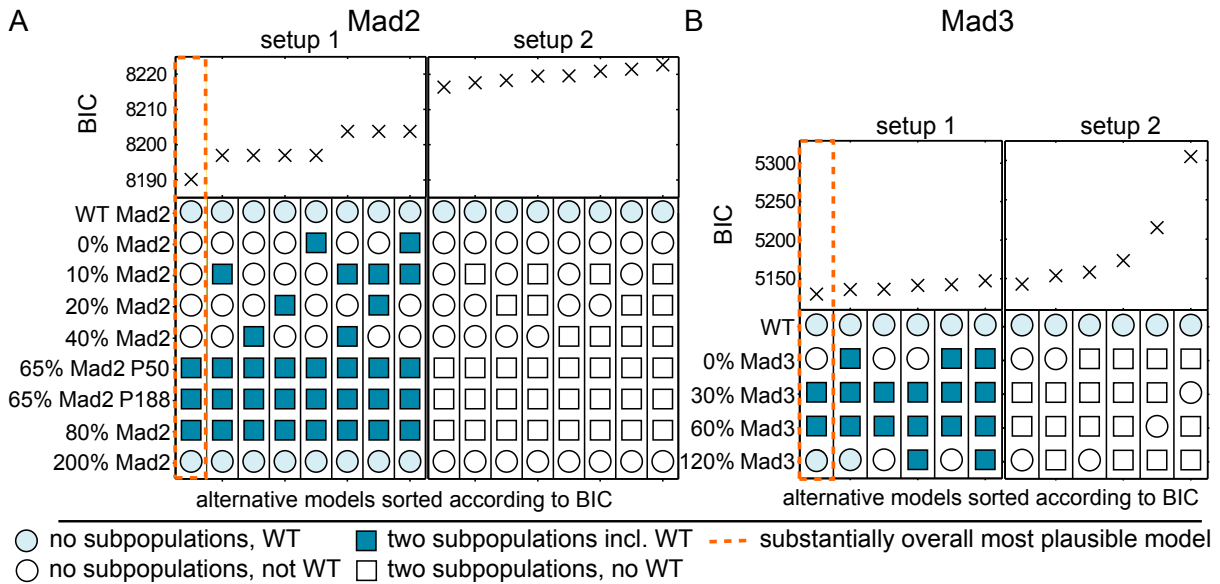


Figure 4.4.: We considered two initial models for the prometaphase length distributions: setup 1 – weighted mixture of wild type and strain specific distribution; and setup 2 – weighted mixture of two strain-specific distributions. Starting from these initial models, backward model selection was performed. In each step all possible individual simplifications were performed and the best model was selected. For each setup, the structures of the eight most plausible models are recorded, ranked according to their BIC. According to the most plausible model, the 200% Mad2 strain is indistinguishable from the wild type. 0% Mad2, 10% Mad2, 20% Mad2, and 40% Mad2 strains consist of unimodal distributions significantly different from wild type. Both 65% Mad2 strains and the 80% Mad2 strain consist of two subpopulations. Figure adapted from Geissen *et al.* (2016).

Inference of subpopulation identities

As the subpopulations with functional SAC seemed to possess similar parameters, we considered in the next step a weighted mixture of the wild type prometaphase length distribution and a strain-specific distribution, and again performed backward model selection (Figure 4.4A and B, setup 1). By comparing the BIC values for setups 1 and 2, we confirmed that the subpopulations with functional SAC have the statistical properties of the wild type. Furthermore, the existence of two subpopulations could be statistically substantiated for the three suspected strains in Mad2 and two strains in Mad3 data (Figure 4.4A and B, setup 1, leftmost columns). These results were confirmed for different distribution assumptions. The estimated fraction

4. Statistical models for the quantification and analysis of cellular SAC phenotypes

Dataset	WT subpopulation			strain specific subpopulation		
	w	μ	σ	w	μ	σ
WT	1.00	6.85	0.43			
Delta Mad2				1.00	3.77	0.14
10% Mad2				1.00	3.81	0.15
20% Mad2				1.00	3.97	0.15
40% Mad2				1.00	4.09	0.16
65% Mad2 P50	0.44	6.85	0.43	0.56	4.51	0.34
65% Mad2 P188	0.80	6.85	0.43	0.20	4.90	0.32
80% Mad2	0.82	6.85	0.43	0.18	4.74	0.35
200% Mad2	1.00	6.85	0.43			

Table 4.1.: MLE of parameter estimation of the most plausible model for Mad2 data resulting from automated hypothesis testing assuming one WT and one strain specific subpopulation per strain.

Dataset	WT subpopulation			strain specific subpopulation		
	w	μ	σ	w	μ	σ
WT	1.00	6.79	0.38			
Delta Mad3				1.00	3.86	0.11
30% Mad3	0.63	6.79	0.38	0.37	4.77	0.30
60% Mad3	0.90	6.79	0.38	0.10	4.83	0.22
120% Mad3	1.00	6.79	0.38			

Table 4.2.: MLE of parameter estimation of the most plausible model for Mad3 data resulting from automated hypothesis testing assuming one WT and one strain specific subpopulation per strain.

and distribution parameters of the log-normally distributed subpopulations are shown in Table 4.1 and Table 4.2. We assessed the identifiability of these parameters by calculating the profile likelihoods for the parameters and also by MCMC sampling and found that all parameters are identifiable. The selected models (leftmost column in setup 1 of Figure 4.4A and B) quantitatively agree with the observed experimental data (Figure 4.5).

4. Statistical models for the quantification and analysis of cellular SAC phenotypes

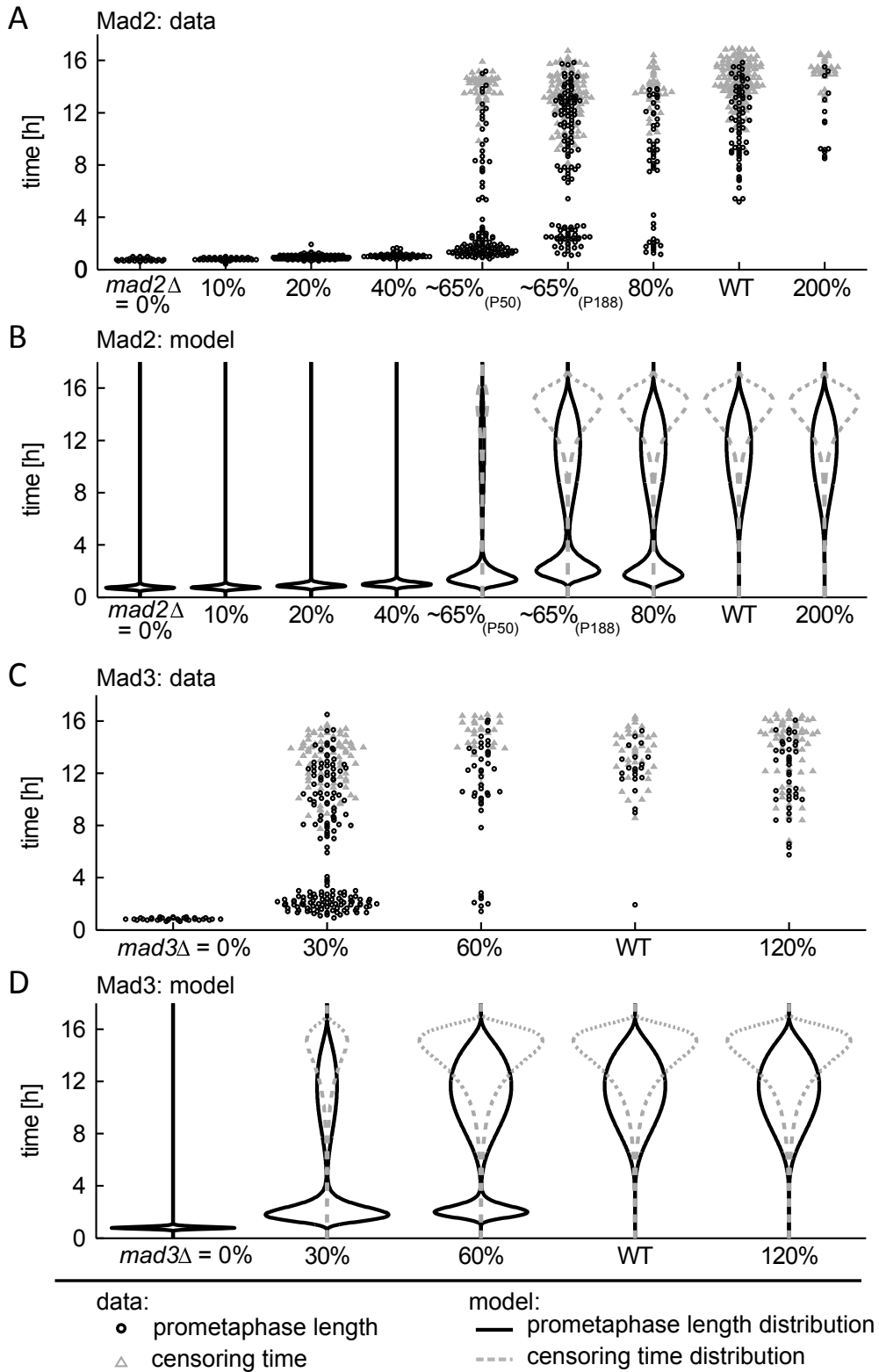


Figure 4.5 (previous page): Analysis of subpopulation structure of SAC functionality in different strains using MEMO. (A) Measured prometaphase length distributions for *S. pombe* strains with different Mad2 abundances. Circles indicate cells in which the entire prometaphase was recorded (prometaphase lengths, interval censored). Triangles indicate cells that were still in prometaphase when recording stopped (censoring times). Since cells enter prometaphase asynchronously, the times at which data are censored are distributed. (B) Model fit of distributions for prometaphase lengths (black lines) and censoring times (dashed gray lines) for the overall most plausible model selected by MEMO. To mimic the bee swarm plots in subfigure A, probability densities are vertically mirrored. Figure adapted from Geissen *et al.* (2016).

4.4.2. Quantification of the sensitivity of SAC functionality to perturbations

In the preceding section the information on the different lengths of prometaphases was used to characterize the the pattern of SAC phenotypes in the various strains. We did not consider the quantitative information on the cause of the observed pattern, which is the altered protein abundance in the different strains. This had the advantage that the analysis was independent of the error-prone data for the protein concentrations of the different strains. However, the functional dependency of the phenotypic pattern on the perturbation can provide additional constraints and insights.

The results from the preceding section already display three qualitative regimes of SAC functionality. A regime in which all cells have a functional SAC and a regime in which all cells have a dysfunctional SAC are separated by a regime where subpopulations of both phenotypes are present in the data. Furthermore, with increasing relative protein amount the fraction of cells with functional SAC increases (Table 4.1 and Table 4.2). In this section we address Problem 4.3 by quantifying the functional dependency of the output (fraction of cells with functional SAC) on the input (relative protein concentration) to gain information on these regimes and the sensitivity of the SAC with respect to perturbations in Mad2 and Mad3.

To investigate the functional dependency of the fraction of cells with functional SAC on the relative amount of protein we reanalysed the complete dataset of Mad2 and Mad3 data, taking this time the measured relative protein abundances \tilde{u}_k into account. These protein abundances were determined by immunoblotting by comparing the bands of mutant strains with different amounts of wild type extracts. While these measurements were done very carefully, visual immunoblot quantifications are error-prone. We chose to account for this by modelling a measurement error ϵ . Errors in the quantification of Mad2 and Mad3 were modelled using additive normally distributed measurement noise around the measured value with a standard deviation of 0.2. Therefore, we model the relative protein abundance as

$$u_k = \tilde{u}_k + \epsilon_k$$

with

$$\epsilon_k \sim \mathcal{N}(0, 0.2)$$

where \tilde{u}_k is the measured relative abundance in strain k and ϵ_k is the unknown measurement

error which has to be estimated together with the other unknown parameters from the prometaphase lengths data.

The functional dependency of the fraction of cells with functional SAC w is modelled as a Hill-type function. Hill-type functions are widely used to model nonlinear input-output dependencies such as dose-response curves. A Hill-function has two parameters: the parameter K can be interpreted as the threshold and n determines the steepness of the sigmoid shaped Hill-function. Therefore, Hill functions are perfectly suited to model switch like behaviours but in the special case of $n = 1$ allow also for more graded responses. Nevertheless, one has to keep in mind that the Hill-function does not have a mechanistic interpretation (Weiss, 1997).

We started out by modelling the fraction of cells with functional SAC w in single perturbation experiments with

$$w_{\text{Mad2}} = f_2(u_{\text{Mad2}}) = (1 + K_2^{n_2}) \frac{(u_{\text{Mad2}})^{n_2}}{(u_{\text{Mad2}})^{n_2} + K_2^{n_2}}, \quad (4.2)$$

for all strains with altered Mad2 abundance and

$$w_{\text{Mad3}} = f_3(u_{\text{Mad3}}) = (1 + K_3^{n_3}) \frac{(u_{\text{Mad3}})^{n_3}}{(u_{\text{Mad3}})^{n_3} + K_3^{n_3}}, \quad (4.3)$$

for data sets of strains with altered amount of Mad3. Here n_2 and K_2 (n_3 and K_3) are the parameters of the Hill-function in the model for Mad2 and Mad3, respectively and u_{Mad2} (u_{Mad3}) is the amount of protein relative to the wild type. Note that the Hill-functions have been scaled to yield 1 for $u = 1$.

We estimated all unknown parameters from the prometaphase data sets of all strains listed in Table 4.3. The estimated distribution parameters and the fractions ω , calculated from Equations (4.2) and (4.3), are very similar to the ones estimated in Section 4.4.2 where ω was not constrained (Table 4.3). This indicates that the parametrization of w was reasonable and did not cause artefacts. The estimated Hill coefficient is much higher for Mad2 ($n_2 = 12.3$) than for Mad3 ($n_3 = 2.1$). This indicates high sensitivity of the SAC functionality to changes in the Mad2 abundance, while the response to changes in the Mad3 abundance is more graded (Figure 4.6). The threshold parameter for reduction of Mad2 abundance was estimated with $K_2 = 0.65$, which indicates 65 % relative Mad2 abundance to be the threshold between the regimes of functional and dysfunctional SAC. A value around 0.65 was anticipated for Mad2 considering the differences observed in two strains with this relative protein abundance. The threshold for Mad3 reduction was estimated with $K_3 = 0.25$. Moreover, the model predicts a bimodal SAC phenotype for almost the entire range of Mad3 reduction, while this regime is more narrow for Mad2 (Figure 4.6).

4.4.3. Data based hypothesis testing on the mode of interaction of Mad2 and Mad3 in SAC signalling

In the previous section we quantified the sensitivity of SAC functionality with respect to the protein abundances of Mad2 and Mad3, respectively (Figure 4.6). For both proteins, a reduction from the normal amount results in a loss of SAC functionality, with Mad2 having a more pronounced effect. Moreover, for both proteins, increased protein amounts did not have

Dataset	WT subpop.			cond. subpop.			input dependency					
	w	μ	σ	w	μ	σ	\tilde{u}_k	ϵ_k	K_2	K_3	n_2	n_3
65% Mad2 _{P50}	0.44	6.82	0.41	0.56	4.51	0.34	0.65	-0.01	0.65		12.3	
65% Mad2 _{P188}	0.77	6.82	0.41	0.23	4.90	0.32	0.65	0.06	0.65		12.3	
80% Mad2	0.87	6.82	0.41	0.13	4.74	0.35	0.80	-0.05	0.65		12.3	
200% Mad2	1.00	6.82	0.41				2.00	0.00	0.65		12.3	
Delta Mad2	0.00	6.82	0.41	1.00	3.77	0.15	0.00	0.00				
10% Mad2	0.00	6.82	0.41	1.00	3.81	0.15	0.10	0.00	0.65		12.3	
20% Mad2	0.00	6.82	0.41	1.00	3.97	0.15	0.20	0.00	0.65		12.3	
40% Mad2	0.00	6.82	0.41	1.00	4.09	0.16	0.40	-0.00	0.65		12.3	
Delta Mad3	0.00	6.82	0.41	1.00	3.86	0.11	0.00	0.00				
30% Mad3	0.63	6.82	0.41	0.37	4.77	0.30	0.30	0.00		0.25		2.1
60% Mad3	0.91	6.82	0.41	0.09	4.83	0.22	0.60	-0.00		0.25		2.1
120% Mad3	1.00	6.82	0.41				1.20	0.00		0.25		2.1
WT	1.00	6.82	0.41				1.00	0.00				

Table 4.3.: MLE of parameter estimation for Hill type dependency of the fraction of cells with functional SAC (w) on the relative amount of protein (input u).

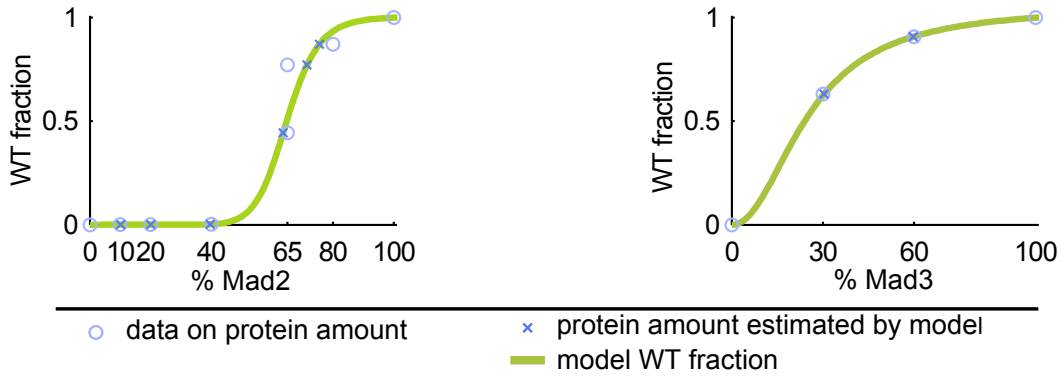


Figure 4.6.: Data driven sensitivity analysis of SAC functionality. A multi-experiment mixture model using a Hill-type description for the fraction of cells with functional SAC in Mad2 and Mad3 data (Equations (4.2) and (4.3)) was fitted to the data. Circles indicate the data for the measured relative protein abundance in the different strains, crosses the protein abundances that result from the estimation of the measurement error ϵ_k . Markers were plotted with the y-value being the model derived fraction of wild type cells (cells with functional SAC).

a detrimental effect on SAC functionality. In this section we address Problem 4.4 and answer the question which mode of action a simultaneous change in both protein amounts exhibits regarding its influence on SAC functionality. As before, we measure SAC functionality in terms of the fraction of the subpopulation of cells that conserves the wild type's properties, i.e. have a functional SAC.

The study of the combined effect of both proteins can give insights on how the proteins qualitatively interact in SAC signalling. One possibility is an independent effect of the amount of both proteins. Independence should be understood here in analogy to sequential action of the effects. In this case the combined effect can be modelled by multiplication of the observed fractions in the respective perturbation. Loosely speaking, a certain relative amount of one protein causes a certain fraction of cells with functional SAC and the relative amount of the second protein reduces this remaining fraction by the factor that is specific for this amount in this protein. To clarify this, imagine perturbations that individually cause strains with a fraction of 50 % cells with functional SAC. Independent effects would lead to a strain with 25 % SAC functionality if both perturbations were combined in one strain. An alternative to this mode of action is a cooperative action of both perturbations on the fraction of cells with functional SAC. This means that the two perturbations cooperate in the impact on the fraction of cells with a functional SAC. This would lead to less than 25 % functionality in our previous example.

Based on the findings on the functional dependencies of the fraction of cells with functional SAC w on the relative amount of protein reported in Section 4.4.2, we assessed these two competing hypotheses by comparing two models encoding these dependencies of w to experimental data (see below). The first hypothesis corresponds to a model of the fraction of cells with a functional SAC ω described by the product of the two Hill-type functions 4.2 and 4.3 derived in Section 4.4.2 (Model A).

Model A:

$$\begin{aligned} w_{\text{Mad2\&Mad3,A}} &= f_A(u_{\text{Mad2}}, u_{\text{Mad3}}) = w_{\text{Mad2}}(u_{\text{Mad2}}) \cdot w_{\text{Mad3}}(u_{\text{Mad3}}) \\ &= (1 + K_2^{n_2}) \frac{(u_{\text{Mad2}})^{n_2}}{(u_{\text{Mad2}})^{n_2} + K_2^{n_2}} (1 + K_3^{n_3}) \frac{(u_{\text{Mad3}})^{n_3}}{(u_{\text{Mad3}})^{n_3} + K_3^{n_3}} \end{aligned}$$

To take possible cooperative effects into account, in the second model the threshold parameters K_{ij} of these functions were described to be inversely proportional to the amount of the other protein as given by Equations 4.4.3 and 4.4.3. Note that both models share the same parameters, except of an additional parameter a in Model B. For inputs $u = 1$ (wild type protein amount) Model B equals Model A, while for lower inputs the K parameters are increased in comparison to Model A.

4.4. Mixture model based analysis of variability in SAC single-cell microscopy data

Dataset	WT subpopulation			strain specific subpopulation		
	w	μ	σ	w	μ	σ
65%Mad2 & 30%Mad3	0.22	6.82	0.41	0.78	4.45	0.32
65%Mad2 & 60%Mad3	0.66	6.82	0.41	0.34	4.91	0.24
65%Mad2 & 120%Mad3	0.86	6.82	0.41	0.14	4.67	0.23

Table 4.4.: MLE of parameter estimation for mixture model-based analysis of prometaphase length data from the indicated strains (double perturbation in Mad2 and Mad3). Data are given in Figure 4.7. The indicated μ and σ are parameters of log-normal densities.

Model B:

$$\begin{aligned}
 w_{\text{Mad2\&Mad3,B}} &= f_B(u_{\text{Mad2}}, u_{\text{Mad3}}) \\
 &= (1 + K_{23}^{n_2}) \frac{(u_{\text{Mad2}})^{n_2}}{(u_{\text{Mad2}})^{n_2} + K_{23}^{n_2}} (1 + K_{32}^{n_3}) \frac{(u_{\text{Mad3}_{\text{rel}}})^{n_3}}{(u_{\text{Mad3}})^{n_3} + K_{32}^{n_3}}
 \end{aligned}$$

with

$$K_{23} = K_2 \frac{a}{(a-1) + u_{\text{Mad3}}}$$

and

$$K_{32} = K_3 \frac{a}{(a-1) + u_{\text{Mad2}}}.$$

To assess how good each model predicts the outcome in double perturbation experiments, we analysed the data of three available datasets (65% Mad2 & 120% Mad3, 65% Mad2 & 60% Mad3, and 65% Mad2 & 30% Mad3, data in Figure 4.7) with the same approach as described in 4.4.1 to gain the wild type fraction of cells in these strains (Table 4.4). To obtain the corresponding model predictions, both models were parametrized with the parameters estimated from single perturbation data in Section 4.4.2. The additional parameter a in Model B was determined independently, to maximize the likelihood of the wild type fractions of the double perturbations, given the MLE for all other parameters derived in Section 4.4.2. It was found to be $a = 7.55$. To assess which of the models is more plausible under the given wild type fractions of the double perturbations, we calculated the BIC with respect to these data for both models. This was done by computing the likelihood of the data, given the respective model, by taking into account that Model B has one parameter more than Model A (Figure 4.8). While Model A has a BIC of 10.18, Model B has an BIC of -3.95 for the fit of these data and is therefore the substantially better model.

The results indicate that a cooperative effect of the relative amount of Mad2 and Mad3 on the fraction of wild type-like cells is more likely than an independent effect. This is consistent with Mad2 and Mad3 acting in the same complex to inhibit Cdc20/Slp1 (Heinrich *et al.*, 2013), thereby inhibiting cell cycle progression. It is consistent with the reasoning that Mad2 is not the limiting factor for SAC functionality in the strains with reduced amount of Mad2, since the simultaneous reduction of Mad3 further reduces SAC functionality (Heinrich, 2013).

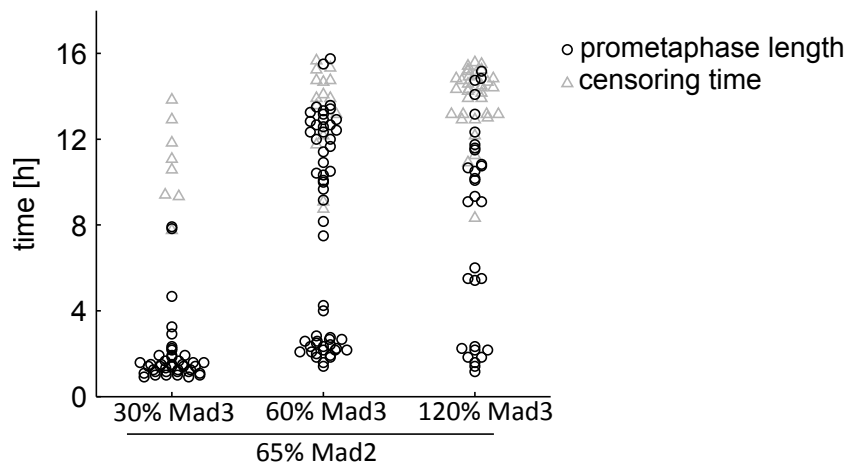


Figure 4.7.: Data for three strains with altered protein amount and Mad3 in a 65% Mad2 P188 background.

4.5. Summary and Discussion

In this chapter, we have assessed the information contained in the cell-to-cell variability of prometaphase length data from several yeast strains by model-based statistical inference. Furthermore, we have presented a general modelling framework that is tailored for such analysis.

In Section 4.3 we have addressed Problem 4.1 by introducing MEMO, a Multi-Experiment mixture MOdelling framework which is able to analyse samples from different experimental conditions simultaneously, can account for censoring and compares competing model hypotheses. MEMO employs finite mixture models and maximum-likelihood inference to determine the subpopulation structure and properties of heterogeneous cell populations. The multi-experiment feature is realised by a joined likelihood for all observations of all experiments. This feature enables the identification and linking of subpopulations across experiments. MEMO exploits the fact that censoring alters the distribution that would be observed without censoring, the data generating distribution, into the distribution that can be observed in the data. Since one is interested in the data generating distribution, one has to account for this effect when analysing the data and derive a model for the observed distribution from the data generating distribution. We have derived these statistical models for different types of censoring and illustrated the formulation of model hypotheses. Furthermore, we have introduced how the unknown parameters are estimated and how model selection is used to select between alternative hypotheses. For two data sets of known subpopulation structure we have compared MEMO with mixture model-based analysis that ignores data censoring to demonstrate the impact of censoring on inference and the resulting pitfalls. The ignorance of censoring tends to overestimate the number of subpopulations, detecting structures that are artefacts of censoring.

The developed modelling framework MEMO allows for its application to a broad range of data types, e.g. single-cell time-lapse and single-cell protein level snapshot data as demonstrated in Geissen *et al.* (2016). The approach can also be used in medical studies, where patients are not observed continuously or may drop out from the study. The current implementation of MEMO supports analytical functions to link experimental conditions. These

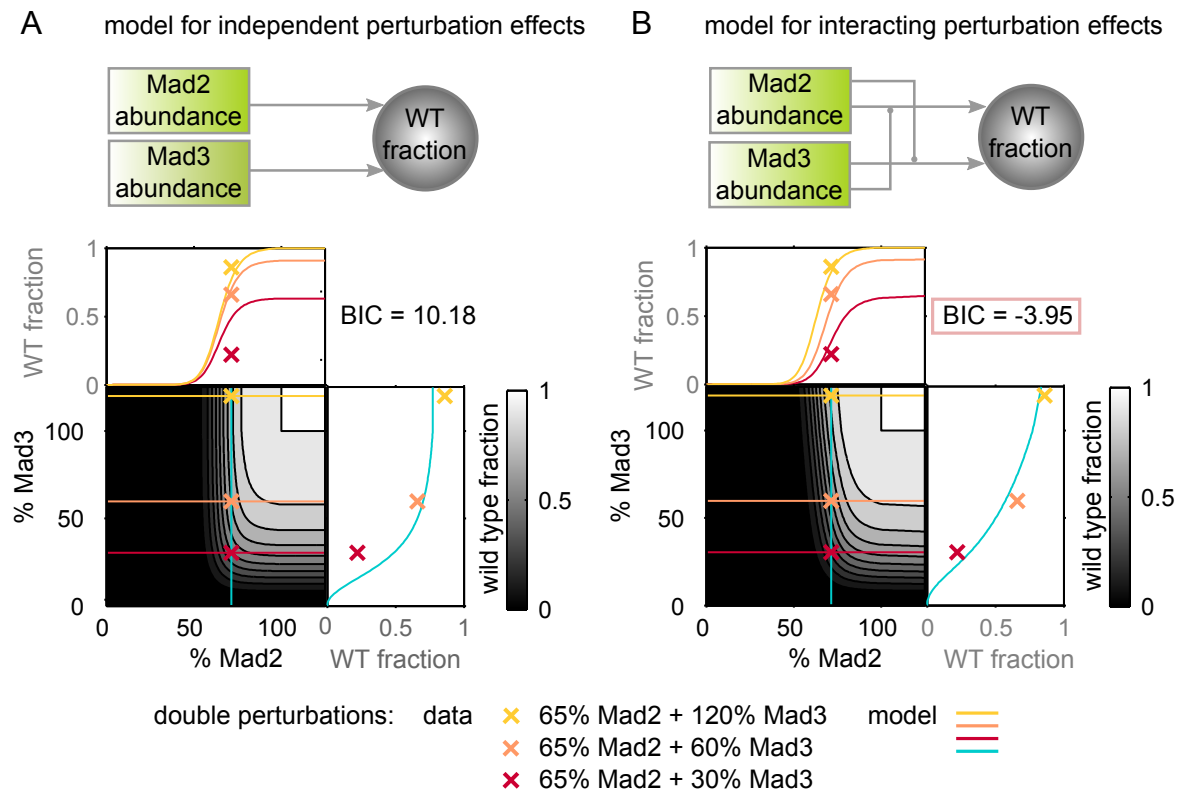


Figure 4.8.: Data-driven hypotheses testing of SAC perturbation response using MEMO. Comparison of model agreement for independent (A) and cooperative (B) perturbation effects of Mad2 and Mad3 with the measured fraction of wild type-like cells in strains with double perturbations. Crosses indicate data from three double perturbation strains. Curves show the wild type fraction in cut planes through the Mad2-Mad3 plane (as indicated by the same colour) as computed from the respective model. The model in (A) reflects independent effects of Mad2 and Mad3 perturbations by just multiplying the models for the individual Mad2 and Mad3 perturbations from Figure 4.6. The model in (B) considers cooperative effects of perturbations in Mad2 and Mad3 by modelling the threshold parameters K in each of the two Hill-type functions to be inversely proportional to the amount of the other protein. As indicated by the lower BIC of the model in B, a cooperative influence of both proteins on the wild type fraction is more likely. Figure adapted from Geissen *et al.* (2016).

functions encode hypotheses and can be derived from measurement data or mechanistic models, such as for example ordinary differential equation (ODE) constrained mixture models (ODE-MM) as described in Hasenauer *et al.* (2014a). ODE-MMs use mechanistic models of single cell behaviour and subpopulation structure to integrate data collected under different experimental conditions (Hasenauer *et al.*, 2014b; Thomas *et al.*, 2014), and could be used to reconstruct differences between subpopulations. MEMO provides an extension to ODE-MM as censored data can be studied and knowledge about the signalling pathway is not required. This renders MEMO more flexible and easier to use for explorative data analysis. Table 4.5 provides a comparison of the features of MEMO, the modelling packages FLAME (Pyne *et al.* (2009)), JCM (Pyne *et al.* (2014)), BayesFlow (Johnsson *et al.* (2016)), and the algorithms presented in Lee & Scott (2012).

In Section 4.4.1 we have addressed Problem 4.2. We have characterized the subpopulation structures of the different yeast strains regarding SAC functionality. The statistical modelling of variability in the prometaphase datasets of several strains has revealed that these strains contain cells with two different cellular phenotypes regarding the functionality of the SAC. One subpopulation of cells keeps the phenotype of wild type cells, despite the altered amount of protein. A second subpopulation displays a phenotype of impaired SAC functionality. Note that this effect is clearly not an artefact of inhomogeneous efficacy of the experimental manipulation, since the treatment is on the genetic level. Each cell has the altered promoter and therefore expresses the respective protein in a different amount as the wild type. Since all cells in a strain are genetically identical, there has to exist non-genetic variability to an extent that suffices to result in these quantitative differences in the cellular phenotype. The results obtained in this chapter can be used for subsequent mechanistic modelling, as we demonstrate in Chapter 5.

Cell-to-cell variability in SAC signalling has been reported before but not quantified beyond sample statistics. Moreover, potential subpopulations with different phenotypes are present in the data of other studies that perturbed SAC signalling (Morrow *et al.*, 2005; Saurin *et al.*, 2011; Thoma *et al.*, 2009). However, the interpretation of these data is difficult because it cannot be ruled out that in these studies the population split is an artefact of the experimental treatment, such as incomplete RNAi knock down. While Morrow *et al.* (2005) attribute the population split they observed to an incomplete knock-down in RNAi treated cells, other publications show bimodal patterns in SAC arrest but do not comment on the potential subpopulations at all (Saurin *et al.*, 2011; Thoma *et al.*, 2009).

In Section 4.4.2 we have addressed Problem 4.3 and quantified the sensitivity of SAC functionality with respect to Mad2 and Mad3. We gained the fraction of the subpopulation with wild type-like SAC phenotype as a function of the relative protein concentration. The dependency on Mad2 is well-approximated by a Hill-function with a Hill-coefficient of 12.3. Therefore the sensitivity towards changes in Mad2 is highly ultrasensitive. Sensitivity with respect to changes in Mad3 has a Hill-coefficient of 2.1 and is therefore more graded. The model predicts a population split for a wide range of relative Mad3 amounts.

In Section 4.4.3 we have addressed Problem 4.4 and assessed the mode of interaction of Mad2 and Mad3 in promoting a functional SAC. We have presented findings that indicate that Mad2 and Mad3 act cooperatively. A simultaneous reduction of both proteins has a more pronounced effect as it would be predicted for independent effects.

Properties		FLAME	JCM	Lee & Scott	BayesFlow	MEMO
Censoring	left	-	-	✓	-	✓
	right	-	-	✓	-	✓
	interval	-	-	-	-	✓
	distributed	-	-	-	-	✓
	fitted	-	-	-	-	✓
Truncation	left	-	-	✓	-	-
	right	-	-	✓	-	-
	fitted	-	-	-	-	-
Distributions	normal	-	-	✓	✓ ⁴	✓
	log-normal	-	-	-	-	✓
	gamma	-	-	-	-	✓
	skew-normal	✓	-	-	-	-
	t	✓	✓	-	-	-
	skew-t	✓	✓	-	-	-
	Johnson SU	-	-	-	-	✓
Data dimension	uni-variate	- ¹	✓	✓	✓	✓
	multi-variate	✓	✓	~ ²	✓	-
Multi-sample fitting		✓	✓	-	✓	✓
Multi-experiment fitting		-	- ³	-	-	✓
Simultaneous analysis of all data		-	-	-	-	✓
Automated model selection		-	-	-	-	✓

Table 4.5.: Comparison of the features of MEMO and other packages. ¹The methods used in FLAME allow for the analysis of univariate data, the implementation does however yield an error with the version available on January 20, 2016. ²The methods used by Lee & Scott (2012) allows for the analysis of multi-variate data, the comments in the code state however that it is only correctly implemented for uni-directional sampling in each coordinate. In the README it is furthermore stated, that the current implementation considers that the truncation is only on the first coordinate. ³JCM exploit prior knowledge of the subpopulation structure to perform the inter-condition matching, In general this will not be available. Furthermore, JCM does not allow for a description of the underlying mechanisms and hypothesis testing. ⁴ Skewed and/or heavy tailed distributions are handled by merging of Gaussian components into super components.

5. Mechanistic models of SAC signalling in heterogeneous populations

The statistical analysis of the prometaphase length data in Chapter 4 substantiated the presence of subpopulations with different cellular SAC phenotypes within the clonal populations of yeast cells. The present chapter addresses the origins of this split as concealed in the SAC signalling mechanism. Despite substantial knowledge of the proteins involved in SAC signalling, the interactions between these proteins, and especially the properties of the dynamics of these interactions are still mostly elusive. We use mechanistic modelling to gain insights into the properties of the SAC signalling mechanism. Thereby, our models focus on the molecular processes of SAC signal transduction described in Section 2.2.2. This chapter is based on the paper “Determinants of robustness in spindle assembly checkpoint signalling”, which we published in *Nature Cell Biology* (Heinrich *et al.*, 2013).

Section 5.1 provides the background information for this chapter concerning the chosen modelling approach and the aspects arising by considering heterogeneity. Section 5.2 delivers the formulation of the problems addressed in this chapter. In Section 5.3 we present a general modelling framework tailored to deal with our phenotypic population data. In Sections 5.4 and 5.5 we set up and analyse two models which describe the SAC signal transduction mechanism with different scopes. The first model, Model M1, describes the sequestration of Slp1 into the MCC. The second model, Model M2, extends this model by the activation of the APC/C through binding of Slp1 and the prevention of APC/C activation by binding to the MCC. Both models are analysed and the analysis provides answers to the sources behind the subpopulation split. Section 5.6 summarizes and discusses the insights into SAC signalling obtained in the present chapter. The findings are put into context and open questions are presented.

5.1. Introduction

This section introduces the concepts and methods that were utilized to gain the results presented in this chapter. We introduce reaction rate equations (RREs), which translate protein reaction networks into ODE models in Section 5.1.1. The integration of cell-to-cell variability introduces a source of randomness to these deterministic models, which is elaborated in Section 5.1.2. The resulting type of model, known as random differential equations (Soong, 1973), requires a special approach for the model simulation. The approach chosen here, the unscented transform, is outlined in Section 5.1.3.

The differing SAC phenotypes we discovered in Chapter 4 imply that SAC signalling has

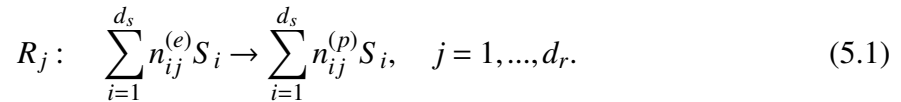
to be qualitatively different in the two subpopulations of cells although these cells show only small variability in the investigated signalling components. Big differences in an output caused by small differences in the input hint to ultrasensitivity in the input-output response of a system. An introduction to the concept of ultrasensitivity and its sources in steady state input-output responses of dynamical systems is given in Section 5.1.4.

5.1.1. Dynamic models of protein reaction networks

Systems biology employs mathematical and computational methods to study the dynamics of signalling pathways in cells. For this, a mathematical representation of the pathway under investigation, a model, is needed. One approach to model biochemical pathways are systems of ordinary differential equations (ODEs) called reaction rate equations (RREs). How these models are gained from the mechanism of signalling is introduced in this section.

Dynamical modelling in systems biology adopts nomenclature and principles from chemistry. Accordingly, models of cellular reaction networks consist of two essential components, species (S_1, S_2, \dots, S_{d_s}) and reactions (R_1, R_2, \dots, R_{d_r}). A chemical species is defined as an ensemble of chemically identical molecular entities, while chemical reaction refers to a process that results in the interconversion of chemical species (McNaught *et al.*, 1997). These definitions can be transferred to biochemistry where species may be e.g. proteins, RNA molecules or lipids and reactions describe e.g. dimerisation of species, their degradation or their synthesis.

In the first place, chemical reactions are defined by their educts (e) entering the reaction and the products (p) resulting from the reaction. They can be written as



Here, $n_{ij}^{(e)} \in \mathbb{N}_0$ and $n_{ij}^{(p)} \in \mathbb{N}_0$ are the stoichiometric amounts of species i as educt and product in reaction j , respectively. The stoichiometric amount is the number of molecules of a particular reactant i (educt or product) taking part in reaction j (Sauro, 2011). In contrast, the stoichiometric coefficient refers to the relative amount of substance that is consumed and/or produced in a reaction. The vector $c_j \in \mathbb{Z}^{d_s}$ contains the stoichiometric coefficients for every species i with regard to reaction j and is calculated by

$$c_j = \left(n_{ij}^{(p)} - n_{ij}^{(e)} \right)_{i=1, \dots, d_s}, \quad j = 1, \dots, d_r. \quad (5.2)$$

To build a dynamic model, in addition to this qualitative description of reactions in terms of products and educts, the kinetics of such a reaction needs to be quantified. A kinetic describes how fast and according to which rules the educts are converted into the products. Under the assumption that the abundance of the species is sufficiently large and the system is well stirred, i.e. spatially homogeneous, the species abundances can be treated as time dependent continuous quantities $x_i(t)$. In this case the velocity of reaction j is given by the reaction rate

$$v_j = \frac{dx_i}{dt} \frac{1}{c_{ij}}, \quad i = 1, \dots, d_s. \quad (5.3)$$

Reaction type	Rate law
$\emptyset \rightarrow \text{product}$	$\tilde{v}(x, \theta) = \theta$
$S_i \rightarrow \text{product}$	$\tilde{v}(x, \theta) = \theta x_i$
$S_{i_1} + S_{i_2} \rightarrow \text{product}$	$\tilde{v}(x, \theta) = \theta x_{i_1} x_{i_2}$
$2S_i \rightarrow \text{product}$	$\tilde{v}(x, \theta) = \theta x_i^2$
$\sum_{i=1}^{d_s} n_{ij}^{(e)} S_i \rightarrow \text{product}$	$\tilde{v}(x, \theta) = \theta \prod_{i=1}^{d_s} x_i^{n_{ij}^{(e)}}$

Table 5.1.: Deterministic rate laws as derived from mass action kinetics (Klipp *et al.*, 2005).

Equation 5.3 reveals that the reaction rate of an isolated reaction can in principle be determined by measuring the rate of change in either of the involved species and knowledge of its stoichiometric coefficient.

In general biochemical reactions in cells are neither separable nor measurable. Furthermore, reaction rates depend on the amount of the reacting species. To be able to model a reaction for a wide range of conditions, a quantitative description of this dependency is needed. Since, in the context of reaction networks, it is not feasible to model the dynamics of reactions from first principles of physics, so called macroscopic rate laws have to be used as an approximation. They are referred to as macroscopic since they do not represent mechanistic aspects of the interactions between molecules. The most common rate law is the so called kinetic law of mass action. It dates back to the early studies of chemical kinetics (reviewed in Voit *et al.* (2015)) and states that the reaction rate is proportional to the concentrations of the involved species raised to the power given by their respective stoichiometric amount. The proportionality constant is called the reaction rate constant. The kinetic law of mass action is rather a model than an actual law, although in chemistry it holds at least for elementary reactions in a volume. The rate laws resulting from different reaction types are given in Table 5.1.

Given the state vector of time dependent species concentrations $x(t) \in \mathbb{R}_+^{d_s}$ the RRE is given by an ordinary differential equation for each species

$$\dot{x} = g(x, \theta), \quad (5.4)$$

in which $g : \mathbb{R}_+^{d_s} \times \mathbb{R}_+^{d_\theta} \rightarrow \mathbb{R}_+^{d_s}$ is the vector field of the RRE and $\theta \in \mathbb{R}_+^{d_\theta}$ the vector of reaction rate constants. The vector field is determined by the reaction rates, as given by the rate laws in Table 5.1, and the stoichiometric coefficients (5.2), yielding

$$\dot{x} = \sum_{j=1}^{d_r} c_j \tilde{v}_j(x, \theta). \quad (5.5)$$

RREs are purely deterministic and do neither cover biological variability nor noise. Non-deterministic, stochastic alternatives to RREs are discrete state continuous time Markov jump processes (MJP) (Gillespie, 1977) and the Chemical Langevin equation (Gillespie, 2000). While the former considers the discrete nature of molecules and reactions, the latter is a system of continuous state stochastic differential equations, approximating MJPs. In contrast, extrinsic cell-to-cell variability can be incorporated by introducing random elements into the ODEs as will be described in the next section.

5.1.2. Dynamics with uncertainty: random differential equations

Even in a clonal population of cells, individual cells are likely to differ slightly. This might for example be differences in the amount of certain proteins or the available resources for protein synthesis and degradation. We have referred to this variability as non-genetic cell-to-cell variability in Section 2.3.1. More specifically, this corresponds to the deterministic sources of cell-to-cell variability called extrinsic variability. Considering extrinsic variability in dynamic models means a transition from the ODEs obtained in the previous section to random differential equations.

Random ordinary differential equations (RODEs) are defined in Banks *et al.* (2014) as ordinary differential equations involving random elements. In general, these random elements can be the initial conditions, the parameters (coefficients) and the inhomogeneous part of the RODE. Therefore, a RODE describes the propagation of uncertainty through a continuous time dynamical system as given by Equation (5.4) in the general form

$$\dot{X} = g(t, X), \quad X(0) = X_0,$$

where X is the state vector of random variables and $X_0 \in \mathbb{R}^{d_s}$ is a random vector (Banks *et al.*, 2014).

RODEs are used in different fields to model heterogeneity in deterministic systems. In this context extrinsic variability is considered as aleatory uncertainty. Aleatory uncertainty is caused by natural variations within an ensemble of individuals which represent individual realizations of the same dynamical system, e.g. particles, devices or cells. Aleatory uncertainty is opposed to epistemic uncertainty, which is uncertainty caused by incompleteness of knowledge, e.g. measurement errors. While epistemic uncertainty can be reduced, aleatory uncertainty is an intrinsic property of a system and therefore irreducible (Banks *et al.*, 2014).

We are interested in systems with variability in initial conditions and parameters. The corresponding RODE reads

$$\dot{x} = g(t, x; \Theta), \quad x(0) = X_0, \quad (5.6)$$

where $x \in \mathbb{R}^{d_s}$ is the state vector, $g : \mathbb{R}_+^{d_s} \times \mathbb{R}_+^{d_\theta} \rightarrow \mathbb{R}_+^{d_s}$ is a vector function of t and x , $\Theta \in \mathbb{R}^\Theta$ is the random vector of parameters, and $X_0 \in \mathbb{R}^{d_s}$ is the random vector of initial conditions.

Systems of nonlinear ODEs in general lack an analytical solution. Even if an analytical solution for the corresponding ODE of a RODE can be found, the resulting function of random variables is tedious to solve. Therefore, computational methods are employed. The common methods include Monte Carlo Methods, probabilistic collocation methods, also called deterministic equivalent modelling method (Webster *et al.*, 1996) and sigma-point based methods (Nørgaard *et al.*, 2000). All three methods solve individual deterministic realizations of the RODE, but differ in the approach to choose these realizations. Monte Carlo methods sample from the distribution of random elements while probabilistic collocation methods are based on quadrature rules. Sigma-point based methods are detailed in Section 5.1.3.

A different approach is to rewrite Equation (5.6) as a first order partial differential equation (PDE) and apply methods from this field to obtain a solution. Therefore, the time evolution of the joint probability density function of $(x(t; X_0, \Theta), \Theta)^T$ has to be derived. As shown in Banks

et al. (2014), one can define the augmented state vector $\tilde{x} = (x, \theta)^T$. Then Equation (5.6) can be rewritten as

$$\dot{\tilde{x}} = \tilde{g}(t, \tilde{x}), \quad \tilde{x}(0) = (X_0, \Theta)^T. \quad (5.7)$$

Here $\tilde{g}(t, \tilde{x}) = \begin{bmatrix} g(t, x; \theta) \\ 0_{d_\theta} \end{bmatrix}$, where 0_{d_θ} is a column vector of dimension d_θ with all vector elements being zeros. Equation (5.7) is often referred to as a crypto-deterministic formulation (Whittaker, 1937). Under some technical assumptions on the existence of the solution $x(t; X_0, \Theta)$ for Equation (5.6) (Ladde & Lakshmikantham (1980), Chapter 3), the joint probability density function $\tilde{\phi}_{X, \Theta}$ of $x(t; X_0, \Theta)$ and random vector Θ fulfils

$$\frac{\partial}{\partial t} \tilde{\phi}_{X, \Theta}(t, x, \theta) + \sum_{k=1}^n \frac{\partial}{\partial x_k} (g_k(t, x; \theta) \tilde{\phi}_{X, \Theta}(t, x, \theta)) = 0 \quad (5.8)$$

with initial condition $\tilde{\phi}_{X, \Theta}(0, x, \theta) = \tilde{\phi}_{X, \Theta}^0(x, \theta)$, and $\tilde{\phi}_{X, \Theta}^0$ is the joint probability density function of X_0 and Θ (Banks *et al.*, 2014). Given the joint probability density function $\tilde{\phi}_{X, \Theta}$, we can obtain the probability density function for $x(t; X_0, \Theta)$ by marginalization over θ

$$p(t, x) = \int_{\Omega_\theta} \tilde{\phi}_{X, \Theta}(t, x, \theta) d\theta,$$

where Ω_θ denotes the set of all possible values for θ .

Equation (5.8) is often referred to as the Dostupov-Pugachev equation. The Dostupov-Pugachev equation for RODEs with random initial conditions and random parameters corresponds to the Liouville equation for RODEs with random initial conditions only (Kozin, 1961), and the Fokker-Planck equation for stochastic differential equations with all diffusion coefficients being zero (Risken, 1996). The derivation of the Dostupov-Pugachev equation converts the problem of solving Equation (5.6) to one of solving an initial value problem involving a first-order partial differential equation. Its solution can be assessed via standard methods for the numerical solution of PDEs (Ames, 2014). An alternative is the use of the method of characteristics (Weiße *et al.*, 2010).

5.1.3. Propagation of uncertainty: Sigma-point based methods

Solving random differential equations, as introduced in the previous section, is a challenging problem. In this section we elaborate on methods that approximate these solutions called sigma-point based methods. Sigma-point based methods are an approach for the efficient, derivative free approximation of the propagation of random variables through nonlinear functions. The term comprises several approaches which are quite similar in their procedure (Särkkä (2013), S.92), although they come from different assumptions. Their classification as sigma-point methods is based on the finding that these methods can be interpreted as special cases of statistical linear regression (Lefebvre *et al.*, 2002; van der Merwe & Wan, 2003). Statistical linear regression is a method to linearise a nonlinear function of a random variable by taking into account the probabilistic spread of the random variable (van der Merwe & Wan, 2003).

Within the family of sigma-point methods, we focus on the unscented transform (UT) (Julier & Uhlmann, 1996) and the divided differences filter of second order (DD2, also called finite differences filter) (Ito & Xiong, 2000; Nørgaard *et al.*, 2000). While the UT is based on the reasoning that it should be easier to approximate a density than a nonlinear function, the DD2 is based on Stirling's interpolation formula for the interpolation of functions (Nørgaard *et al.*, 2000). Interestingly, the algorithms resulting from both approaches are almost the same. The basic idea is to approximate the transformation of the probability density by approximating the transformation of its moments. Therefore the distribution is represented by deterministically chosen points on its support, so called sigma-points, which have the same sample mean and covariance as the original distribution. These points are then transformed by the nonlinear function individually and the moments of the resulting distribution, also called posterior distribution, are derived by calculating of the sample moments from the weighted transformed points.

In detail, a random variable X (dimension n) is propagated through a nonlinear function, $y = g(x)$. The random variable X has mean $\mathbb{E}[X]$ and covariance V_X . To calculate the statistics of y , a matrix \mathcal{X} of $2n + 1$ sigma vectors \mathcal{X}_i is formed, according to

$$\begin{aligned} \mathcal{X}_0 &= \mathbb{E}[X], \\ \mathcal{X}_i &= \mathbb{E}[X] + \left(\sqrt{\frac{n}{1-\omega_0} V_X} \right)_i, & i = 1, \dots, n \\ \mathcal{X}_i &= \mathbb{E}[X] - \left(\sqrt{\frac{n}{1-\omega_0} V_X} \right)_{i-n}, & i = n+1, \dots, 2n \\ \omega_0 &= 1 - \frac{n}{h^2}, \end{aligned}$$

where $(\sqrt{V_X})_i$ denotes the i th row of a matrix square root such that $\sqrt{V_X} \sqrt{V_X}^T = V_X$. Note that we use here a generalized version of the notation in Nørgaard *et al.* (2000). The sigma points are propagated through the nonlinear function to yield the transformed sigma points

$$\mathcal{Y}_i = g(\mathcal{X}_i) \quad i = 0, \dots, 2n.$$

The mean of Y is approximated using a weighted sample mean of the transformed sigma points

$$\begin{aligned} \mathbb{E}[Y] &\approx \sum_{i=0}^{2n} \omega_i \mathcal{Y}_i, \\ \omega_i &= \frac{1-\omega_0}{2n}, & i = 1, \dots, 2n. \end{aligned}$$

For $X \sim \mathcal{N}(\mu, \sigma)$ the optimal choice is $h^2 = 3$ (Julier *et al.*, 2000; Nørgaard *et al.*, 2000).

For the UT the covariance of y is calculated via the sample variance with

$$\begin{aligned} \mathbf{P}_y^{UT} &\approx \sum_{i=0}^{2n} \omega_i^c (\mathcal{Y}_i - \bar{y})(\mathcal{Y}_i - \bar{y})^T, \\ \omega_0^c &= \omega_0 + 1 - \frac{h^2}{n} + \beta, \\ \omega_i^c &= \omega_i, & i = 1, \dots, 2n. \end{aligned}$$

where β is a parameter that incorporates prior knowledge on the density of x and $\beta = 2$ is optimal for Gaussian densities (Julier *et al.*, 2000). In contrast, the covariance according to the DD2 is calculated as

$$\begin{aligned} \mathbf{P}_y^{DD2} \approx & \frac{1}{4h^2} \sum_{i=1}^n [(\mathcal{Y}_i - \mathcal{Y}_{i+n})(\mathcal{Y}_i - \mathcal{Y}_{i+n})^T] \\ & + \frac{h^2 - 1}{4h^4} \sum_{i=1}^n [(\mathcal{Y}_i + \mathcal{Y}_{i+n} - 2\bar{y})(\mathcal{Y}_i + \mathcal{Y}_{i+n} - 2\bar{y})^T]. \end{aligned}$$

It has been shown that the DD2 has a marginally higher theoretical accuracy (Nørgaard *et al.*, 2000), but for practical application it has been shown that both perform equally well (van der Merwe & Wan, 2003). The accuracy of the approximation can be improved by decomposing the original density into a weighted mixture of Gaussian densities (van der Merwe, 2004). The mixture components themselves are then represented by sigma points and transformed individually. The transformed component distributions are used to reconstruct the transformation of the original distribution.

We used the DD2 version of the sigma-point methods for the simulation of our mechanistic SAC models as described in Section 5.3.2.

5.1.4. Ultrasensitivity in steady state input-output responses

In the preceding sections we addressed dynamical aspects of signalling models. But, dynamical systems have interesting features beyond dynamics. For the analysis of a dynamic system the determination and characterization of its steady state properties is of special interest. The steady state \bar{x} of a dynamical system, described by a system of ODEs, given a set of parameters θ and an input u , is defined by

$$\dot{x} = g(\bar{x}, \theta, u) = 0. \quad (5.9)$$

A steady state is a condition in which the system states do not change over time any more, but achieve a constant value depending on the input u and the parameters θ . Accordingly, the system output $\bar{y} = h(\bar{x})$ is constant in steady state. Plotting the steady state output \bar{y} over the corresponding input u yields the steady state input-output response curve of the system. This curve visualizes characteristic features of the system, e.g. the sensitivity of the output with respect to changes in the input (Thomaseth *et al.*, 2017).

This section gives an introduction to ultrasensitivity, a property of the steady-state input output responses of biochemical systems. In this context sensitivity refers to the degree of which fold changes in some control parameter, the input, are amplified to fold changes in the output. We elaborate on the definition and quantification of ultrasensitivity and particular mechanisms that result in ultrasensitive steady state input-output responses. We focus here roughly on the extent needed within the scope of this thesis. For a more comprehensive overview we refer to a recent review series on ultrasensitivity (Ferrell & Ha, 2014a,b,c). The following paragraph attempts to clarify the concept of ultrasensitivity and introduces existing measures for its quantification.

Definition and quantification of ultrasensitivity

The term ultrasensitivity originates from a paper of Albert Goldbeter and Daniel E. Koshland in 1981 in which the authors use it “to describe an output response that is more sensitive to change in stimulus than the hyperbolic (Michaelis-Menten) equation”(Goldbeter & Koshland, 1981). Today ultrasensitivity is commonly defined as “a property of input-output relationships that makes them switch like in character” (Ferrell & Ha, 2014a). This means that a small fold change in the input produces a larger fold change in the output, resembling switching from an “Off” state to an “On” state without intermediates. In terms of signal transduction this corresponds to an amplification of the signal change by the system. The definitions of ultrasensitivity given above, despite being illustrative, are not applicable for a rigorous quantitative assessment of the sensitivity exhibited by the response of a system. In the following several global and local measures of sensitivity are given to define and discriminate ultrasensitive responses. The applicability of the different approaches depends on the response curve under study.

For response curves with a sigmoidal (S-shaped) shape it is a common approach to determine the degree of sensitivity by approximating the response with a Hill equation

$$\text{output} = \frac{\text{input}^{n_H}}{K^{n_H} + \text{input}^{n_H}}. \quad (5.10)$$

The resulting Hill coefficient n_H is determined by the steepness of the S shape and serves as a global measure of sensitivity. Thereby $n_H > 1$ indicates an ultrasensitive response. Responses with $n_H = 1$ are called hyperbolic and responses with $n_H < 1$ are called subsensitive (Koshland *et al.*, 1982). For $n_H = 1$ the Hill equation equals the Michaelis-Menten equation. For higher values of n_H the shape more and more equals an ideal switch. With that the Hill equation connects to both of the informal definitions of ultrasensitivity cited above. The second parameter K equals the input value for which the output reaches half of its maximum value. The Hill equation is an approximation of the response of a cooperative enzyme ($n_H > 1$), where the Hill coefficient describes the degree of cooperativity. It is named after Archibald Hill who introduced it in an attempt to describe the binding curve of oxygen to hemoglobin in 1910 (Hill, 1910). Ultrasensitivity in this system arises from positive cooperative binding. Hemoglobin has multiple binding sites for oxygen molecules whereby the binding of an oxygen molecule promotes the binding of the following molecule. Unfortunately, the applicability of the Hill equation as a measure of sensitivity is limited to sigmoidal response curves, and therefore solely to responses with a maximum value.

A second global measure of ultrasensitivity is the $EC_{90} : EC_{10}$ ratio (Koshland *et al.*, 1966), also called response coefficient. The EC_{10} is the input leading to 10% of the maximum output and EC_{90} the input yielding 90% of the maximum output. Therefore, it describes the fold change in the input needed to achieve a 9 fold change in the output. For a hyperbolic response this value is 81. In the context of ultrasensitivity it was introduced by Goldbeter and Koshland as a general global measure for ultrasensitivity independent of the Hill equation (Goldbeter & Koshland, 1981). The authors defined an ultrasensitive response as to have an $EC_{90} : EC_{10}$ ratio smaller than 81 (Goldbeter & Koshland, 1981), meaning that less than an 81 fold input change is needed to generate a 9 fold increase in the output. However, this global measure of sensitivity is again not applicable to all ultrasensitive responses, because it

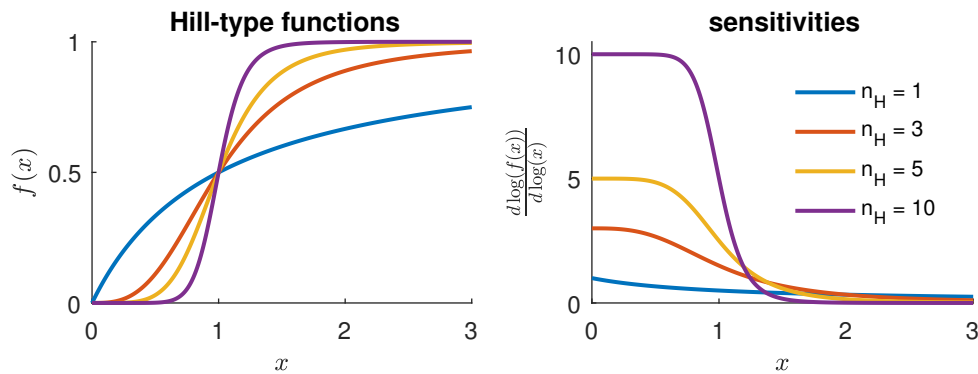


Figure 5.1.: Hill-type functions and corresponding local sensitivities with $K = 1$ and different values of the Hill coefficient n_H .

requires the output to be bounded, i.e. to saturate for high input values.

Another approach to globally quantify the sensitivity of a response curve is the effective Hill exponent. It combines the Hill equation with the $EC_{90} : EC_{10}$ ratio. The effective Hill exponent is defined as the Hill coefficient n_{eff} of the Hill equation that shares the $EC_{90} : EC_{10}$ ratio with the questioned response curve. The effective Hill coefficient is related to the $EC_{90} : EC_{10}$ ratio by

$$n_{eff} = \frac{\log_{10}(81)}{\log_{10}\left(\frac{EC_{90}}{EC_{10}}\right)}$$

(Taketa & Pogell, 1965). As for the Hill equation, ultrasensitivity is given for $n_{eff} > 1$. Again, the measure is only applicable if the output converges to a maximum value.

The requirement for a bounded output or a certain shape of the response can be circumvented if the sensitivity of the response curve is assessed locally. As mentioned above, sensitivity relates a change in the input to the corresponding change in the output. Therefore, a local sensitivity measure can be derived by assuming an infinitesimal change in the input (Higgins, 1965). This results in the ratio called local sensitivity

$$S_{local} = \lim_{\Delta input \rightarrow 0} \frac{\frac{\Delta output}{output}}{\frac{\Delta input}{input}} = \frac{doutput}{dinput} \frac{input}{output} = \frac{d \log(output)}{d \log(input)}. \quad (5.11)$$

Sometimes (e.g. in metabolic control analysis) Equation (5.11) is also defined as response coefficient, which can lead to confusion with the $EC_{90} : EC_{10}$ ratio. S_{local} equals the slope of a log-log plot of the output versus the input, and therefore it also reflects the polynomial order of the response curve. Higher values indicate higher local sensitivity. Ultrasensitivity is determined by $S_{local} > 1$. Although S_{local} specifies the sensitivity at every point along the response curve, it does not necessarily provide the information which range a signal change should span to yield a signal amplification that results in a physiologically meaningful effect. This is illustrated best when considering the sensitivity of response curves of Hill functions with different n_H (Figure 5.1). The local sensitivity equals the Hill-coefficient of a Hill-equation for inputs approaching zero, whereas it is substantially lower at the half maximum response where differences in the input lead to clearly distinguishable outputs.

Signalling mechanisms generating ultrasensitive responses

There are several motifs and mechanisms that produce an ultrasensitive input output response without employing a Hill equation in its kinetics. The probably most widely known mechanism besides positive cooperativity is called **zero-order ultrasensitivity** (Goldbeter & Koshland, 1981). Goldbeter and Koshland found that if two opposing reactions (a forward and a backward reaction) operate in saturation of Michaelis-Menten kinetics, a regime called zero-order, a small change in one of the enzymes activity leads to an amplified change in its product. This is based on the demand that in steady state the two fluxes balance: the increased enzyme activity increases the flux towards its product which can due to the saturation only be balanced by shifting the steady state in the direction of this product, to increase the substrate for the backward reaction.

Coherent feed-forward regulation motifs (Mangan & Alon, 2003) yield ultrasensitive responses from reactions described with mass action kinetics, that is without a role of zero-order effects. The characteristic of these motifs is that the input positively feeds into the output more than once which amplifies the effect of changes in the input. Such an input might for example increase the abundance and the activity of the output species, or increase its abundance and at the same time inhibit its degradation. Therefore, this mechanism is also known as *multi-step signalling* (Zhang *et al.*, 2013) or *multi-step ultrasensitivity* (Ferrell & Ha, 2014b).

Positive feedback, and equivalently double negative feedback, can produce ultrasensitive input output responses (Ferrell & Ha, 2014b). Positive feedback is revealed by a circuit in the interaction graph of the system (Radde, 2010). The interaction graph of a system as given by a RRE is a signed directed graph with vertices given by the model species and signed edges defined by the Jacobian matrix of the system (Domijan & Pécou, 2012). The Jacobian matrix is given by

$$J(x) = \begin{bmatrix} \frac{\partial g}{\partial x_1} & \dots & \frac{\partial g}{\partial x_{d_s}} \end{bmatrix}.$$

The signs of the individual components $J_{ij} = \frac{\partial g_i}{\partial x_j}$ indicate positive and negative influence of the species j on the species i , which are indicated by an edge from species j to species i in the interaction graph. A circuit of the graph is a sequence of distinct vertices that start and ends in the same vertex. The sign of the circuit is given by the product of the signs of the edges along that sequence. A circuit with a positive sign indicates positive feedback in the system. Positive feedback amplifies changes in the input and this leads to an ultrasensitive response. Under certain conditions positive feedback can cause bistability in a system (Ferrell & Ha, 2014c). A bistable system has three steady states for a range of inputs: two stable steady states are separated by an unstable steady state. In a biological system only positive and stable steady states can be realized. To which of them the system output will converge depends on its initial condition. An increase in the input that spans the critical point will result in ultrasensitivity since the system jumps from one stable steady state to the next. Whether a system exhibits bistability depends on the system design but also on the parameters. A given system can be bistable for one set of parameter values but fail to do so with a slightly different set.

Molecular titration, also called **inhibitor ultrasensitivity**, is a motif that generates ultrasensitivity without the response curve being sigmoidal (Buchler & Louis, 2008). The term

molecular titration refers to the analogy to the method of chemical titration. Chemical titration measures the unknown concentration of a substance (the analyte) by continuously adding a sequestering substance (the titrant) until the amount of titrant equals the amount of analyte. Below the equivalence point, free titrant is buffered into a complex with the analyte. The equivalence point is marked by the sudden detection of free titrant as soon as the amount of titrant exceeds the amount of analyte and cannot be buffered any longer. The biochemical analogue is a heterodimerization reaction in which an inhibitory species sequesters a target by binding to form a heterodimeric complex. The response curve of this system can for example be the steady state amount of active, unbound target in dependence of the total amount of target. As long as the inhibitor is in excess, the target will be almost completely sequestered and the amount of free target is low. As soon as the total amount of target protein exceeds the amount of inhibitor, active target can accumulate. At the boundary between buffering and sudden accumulation ultrasensitivity occurs.

5.1.5. Model-based studies of SAC signalling

The cell biological research into SAC signalling has been complemented by model-based approaches in the last decade. Mathematical models are used to model the establishment of a SAC arrest as well as the dynamics of cell cycle progression after SAC silencing. The obtained models can be classified into biophysical models and molecular models (Ciliberto & Shah, 2009). While the biophysical models aim to provide conceptual insights into the mechanism of signalling, the molecular models try to reconcile signalling on the mechanistic level. In the following sections we review relevant model-based research on the SAC.

Biophysical SAC models

Doncic *et al.* (2005) provided the cornerstone of quantitative SAC modelling with a study that investigates different generic molecular reaction-diffusion schemes, considering biophysical limitations as given in mitosis of budding yeast. They found that, considering realistic diffusion rates, direct inhibition of Cdc20 at unattached kinetochores cannot account for strong inhibition and fast inactivation kinetics, while the production of a diffusible stoichiometric inhibitor at the kinetochore, which is constitutively disassembled, complies with the spatial and temporal constraints. In a further model-based study Doncic and colleagues showed that inhibition through sequestration is more robust regarding changes in the Cdc20 concentration than inhibition by degradation (Doncic *et al.*, 2006).

Building on the work of Doncic and colleagues, Sear & Howard (2006) introduced measurements of numbers and dynamics of SAC proteins at kinetochores and confirmed that direct inhibition at kinetochores is not feasible. Moreover, they showed that, given their data, a diffusible inhibitor is not sufficient for cell cycle inhibition. They proposed a model in which kinetochore-derived inhibitors can produce further inhibitors in the cytoplasm, which in turn do not have the potential to produce further inhibitors. This model was later extended by Mistry *et al.* (2008), to include upstream kinase activity as well as the dynamics and influence of kinetochore-microtubule interactions.

Another study considering the spatial and temporal constraints is provided by Dao Duc & Holcman (2012). They use a Markov chain to model APC/C activation by Cdc20. In

their model free Cdc20 can be degraded through binding to the MCC and is produced by dissociation from a preformed complex. The model parameters are constrained by temporal and spatial constraints. They find that the production rate and the degradation rate need to be balanced to reliably prevent APC/C activation and allow for its fast activation after the SAC is silenced.

Molecular models

Molecular models consider the known SAC proteins, their interactions and the kinetics of these interactions to simulate the pathway of signalling. Simonetta and colleagues use in vitro measurements of single reactions in SAC signalling and known protein abundances to model a simplified signalling (Simonetta *et al.*, 2009). They could determine to which extent catalysis accelerates inhibitor formation at unattached kinetochores, but found that their model is not able to account for the dynamics of SAC activation.

In a reverse engineering approach using the quantitative phenotype of gene deletion mutants Doncic and colleagues attempted to deduce interactions between SAC proteins (Doncic *et al.*, 2009). They predict that the cooperation between Cdc20 sequestering and degradation creates an optimized inhibition. They screened thirty million possible networks and concluded that Mad2 and MCC sequester Cdc20, while a complex formed by Bub3 and Mad3 and the MCC are responsible for its degradation.

In a whole series of publications Ibrahim and colleagues have successively extended a model of SAC signalling. Starting with the generation of the Cdc20:Mad2 complex at kinetochores (Ibrahim *et al.*, 2008b), via the assembly of the MCC (Ibrahim *et al.*, 2009), up to the inhibition of the APC/C by the MCC (Ibrahim *et al.*, 2008a) the authors devised detailed molecular models of SAC signalling. Since the authors introduced mathematical *ad hoc* formalism to model speculative roles of the kinetochores, the interpretability of the results of these studies in biological terms is rather limited. Albeit in an artificial framework, this work may provide a study in parameters that are able to recapitulate dynamics of SAC signalling.

5.2. Problem formulation

In order to gain insights into SAC signalling, the information available on the involved proteins and the reactions need to be translated into a mathematical model of the signalling mechanism.

Problem 5.1. (Formulation of a mechanistic model of SAC signalling) *Given the published protein interactions in SAC signalling as described in Section 2.2.2, formulate the corresponding RREs.*

Our experimental data consist of fractions of subpopulations, identified in the prometaphase lengths of single *S. pombe* cells. A mechanistic model of SAC signalling, formulated as RREs, describes SAC signalling on the level of signal transmitting proteins. Therefore, the experimental data does not correspond to one of the model species. Furthermore RREs are deterministic and as such restricted to the representation of an average cell. To model signalling on the population level, we need to introduce cell-to-cell variability into the RREs.

Problem 5.2. (Formulation of a framework to model phenotypic variability in SAC signalling) *Given the phenotypic data and the observed cell-to-cell-variability, formulate a modelling framework that maps SAC signalling to the SAC phenotype and includes cell-to-cell variability.*

Chapter 4 substantiated a bimodal phenotype of SAC functionality within the isogenic population of yeast cells. Signalling has to be qualitatively different in the two subpopulations. Differences in signalling can be caused by differences in the amount of signalling molecules or differences in reactions generating the signal. The latter is quite unlikely in an isogenic population. Bimodality of total protein concentrations is observed due to burst like protein expression. However, we could show that the key signalling components do not have bimodal distributions and that the noise in the abundance of these proteins is small (Heinrich *et al.*, 2013).

Problem 5.3. (Determination of the sources of phenotypic bimodality) *Given that the signalling molecules do not have a bimodal distribution, determine other possible sources of the observed phenotypic bimodality.*

5.3. A general framework for mechanistic models of populations with dual SAC phenotypes

In this section we address Problem 5.2. A general framework to utilize binomial phenotype data and cell-to-cell variability in a deterministic model setup is derived. This includes model simulation, calibration with data and assessment of the model fit.

5.3.1. Formulation of a population model with phenotypic output

In this section the general formulation of a population model, tailored to describe the dual phenotype SAC data, is presented. We will start with deriving a general model of the dynamics of a single cell in the population and then make the transition to the model of a heterogeneous population.

The protein reactions which constitute SAC signalling are determined by the genome and therefore are the same in every single cell. We assume that stochastic effects (intrinsic noise) are negligible in SAC signalling. Therefore, a deterministic model of SAC signalling can be derived from the known protein reactions (Section 2.2.2) as a RRE (Section 5.1.1) based ODE model in the state space representation

$$\mathcal{M}(\theta) : \begin{cases} \dot{x} = g(x, \theta), & x(0) = x_0(\theta) \\ y = h(x, \theta) \end{cases} . \quad (5.12)$$

The model \mathcal{M} is defined by the dynamics \dot{x} of the state vector of time dependent species concentrations x with initial condition x_0 and the model output y . The parameter vector is denoted by $\theta \in \mathbb{R}^{d_\theta}$. The model output is mapped from the state space by the output mapping h . For our purposes, the model output needs to be the SAC phenotype derived from the models'

states. SAC signalling suppresses the accumulation of proteins and complexes that promote the transition to anaphase. Therefore, a model of SAC signalling has to include such a species and this species can be regarded as a measure for the strength of signalling. To discriminate between the two SAC phenotypes we assume this anaphase activating species to reach a threshold. This is a common simplification in SAC modelling (Doncic *et al.*, 2005, 2006; Mistry *et al.*, 2008), and is based on the assumption that very small amounts of the anaphase promoting species are insufficient to initiate anaphase, because the system would otherwise not be robust. In our model this translates into two possible values of the output mapping

$$h(x, \theta) = \begin{cases} 0 & \text{if } \exists t \text{ such that } x_i(t) > \theta_j \\ 1 & \text{if } \forall t \ x_i(t) \leq \theta_j \end{cases}, \quad (5.13)$$

where x_i with $i \in \{1, \dots, d_s\}$ is the model state representing the anaphase promoting model species and θ_j with $j \in \{1, \dots, d_s\}$ is the parameter representing its threshold. Simulated trajectories crossing the threshold indicate parametrizations of the model that result in a dysfunctional SAC phenotype B ($y = 0$), while trajectories staying below the threshold indicate the functional SAC phenotype A ($y = 1$). The threshold is unknown and has to be estimated from data together with all other unknown parameters.

A cell population is an ensemble of individual cells, which are likely to differ slightly due to non genetic cell-to-cell variability. From the data we know that single cells differ quantitatively (prometaphase lengths) and qualitatively (SAC phenotype). We included extrinsic cell-to-cell variability to model the behaviour of the whole cell population. We chose the reaction rate constants and initial conditions that correspond to the manipulated quantities in the yeast strains to be variable within the population, instead of a single value. Therefore the parameters θ are split into $\Theta = [\theta_1, \Theta_2]^T$, where θ_1 is the vector of scalar parameters, whereas Θ_2 contains the parameters which are represented as a random variable. The initial conditions are as well split in the scalar and random parts $X_0 = [x_{10}, X_{20}]^T$. With this, the model states become random variables and the model becomes a system of random ordinary differential equations (RODEs), i.e. ordinary differential equations with random initial conditions and random model parameters (Banks *et al.*, 2014).

The model of the heterogeneous population is denoted by

$$\tilde{\mathcal{M}}(\theta) : \begin{cases} \dot{X} = g(X, \Theta), & \tilde{X}(0) = X_0 \\ \tilde{y} = \tilde{h}(X, \Theta) \end{cases}, \quad (5.14)$$

consisting of the system of random ordinary differential equations \dot{X} and the population output mapping \tilde{h} . In such a model, every model species is associated with a probability density $p_i(t, x)$ which evolves over time. The output mapping \tilde{h} maps the relative fraction of cells with functional SAC from the state vector by calculating the probability mass below the threshold

$$\tilde{h}(X, \Theta) = \max \left(\int_0^{\theta_j} p(t, x_i) dx_i \right).$$

The fraction of cells with non-functional SAC can be easily derived from $1 - \tilde{y}$.

To describe the variability in the initial conditions and parameters, we employ the probability density of the log-normal distribution. This is a common assumption for the distribution

5.3. A general framework for mechanistic models of populations with dual SAC phenotypes

of cell-to-cell variability in biological systems (Furusawa *et al.*, 2005). Therefore, the k th distributed initial condition is distributed according to

$$X_{20_k} \sim \log\mathcal{N}(\mu_k, \sigma_k),$$

and the l th distributed parameter is distributed according to

$$\Theta_{2_l} \sim \log\mathcal{N}(\mu_l, \sigma_l).$$

Hence, every dimension of variability is represented by two unknown parameters, the scale and location parameters σ and μ of the corresponding log-normal distribution. The variability is assumed to be uncorrelated, such that it can be modelled by univariate densities.

5.3.2. Model simulation

Given the model specified by (5.14), the simulation of the dynamics of the cell population corresponds to the propagation of the initial densities of the states in time by a nonlinear dynamic system with uncertain parameters. The most intuitive approach for this task is the Monte Carlo approach (Helton & Davis, 2003). It resembles the population in a natural way by simulating it as an ensemble of individual cells. Therefore, different values of reaction rate constants and initial conditions are drawn from the corresponding probability densities and the model is simulated with these values to generate a sample from the state densities. By summing up the model outputs of the individual simulations as given by (5.13) and relating the sum to the total number of simulations, one would obtain a Monte Carlo estimate of the probabilities to observe the functional SAC phenotype. This probability corresponds to the fraction of cells observed in this subpopulation. However, the number of simulated cells required to achieve a high precision is large. Furthermore, when this simulation approach is used for parameter estimation, the resulting objective function exhibits stochastic fluctuations. This renders the application of efficient gradient-based methods impractical and the optimization of the objective function computationally intractable.

To circumvent these problems, we employed the sigma-point based approach described in Section 5.1.3 as an efficient approximation for the transformation of the probability densities by the numerical solution of the nonlinear ODE system. To improve the approximation, this approach was combined with the decomposition of the log-normal densities into mixtures of Gaussian densities. The mixture components can be transformed individually, with three sigma-points each, and the overall distribution of the states can be reconstructed as the weighted mixture of the components posterior densities at every time point during the simulation (Figure 5.2).

Since every log-normal density can easily be derived via a transformation of the standard normal (Gaussian) density, the decomposition has to be derived only once for the standard normal distribution. For an efficient implementation, the decomposition was conducted in an intelligent way. We determined the optimal decomposition of a standard normal density into 24 Gaussian mixture components, which all share the same standard deviation. The means of the mixture component densities were constrained to be positioned equidistantly, spaced in a way that the spacing equals the shared, but a priori unknown, standard deviation multiplied by $\sqrt{3}$. The chosen spacing between the means of the mixture components reduces the number

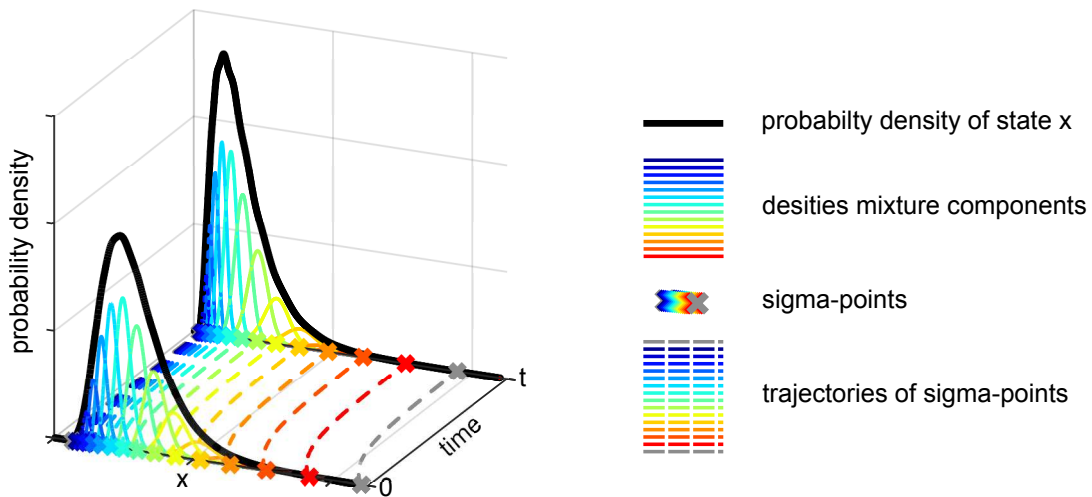


Figure 5.2.: Illustration of the propagation of a probability density. In this example, the density at time 0 is decomposed into 14 mixture components. The resulting 16 sigma-points (see text for an explanation of this number) are simulated. At time t the transformed sigma-points are used to calculate the 14 transformed mixture components. The overall distribution at time t is computed as the weighted sum of the transformed mixture components.

of numeric model evaluations, since the means of the two neighbours of every component constitute the two sigma-points needed in addition to the components own mean. In the one dimensional case this results in the reduction to 26 forward simulations for the approximation of 24 Gaussian distributions instead of 72 ($24 \cdot 3$) simulations if the component means are not constrained in such a way. In the case of two variable dimensions - as in the models presented in Sections 5.4 and 5.5 - the number of forward simulations is reduced from 5481 ($72 \cdot 72$) to 576 ($24 \cdot 24$). The optimal mixture model was determined by the estimation of the mixture weights and the common standard deviation of all mixture components. The estimation was performed by the minimization of the sum of squared errors between the mixture of Gaussians and the standard normal distribution, given the constraints on the location of the means. This approach to approximate the state densities has the advantage that no integral has to be computed for the mapping of the output from the states. By employing normal distributions the probability mass below the threshold can be easily computed via the evaluation of the cumulative distribution function (cdf) of the normal distribution for the threshold value.

For the numerical solution of the ODEs the Sundial (Hindmarsh *et al.*, 2005) solver CVODES Serban & Hindmarsh (2005) was used. CVODES is a stiff and nonstiff ODE initial value problem solver that provides the derivatives $\frac{dx}{d\theta}$, also called sensitivities, by solving the sensitivity equations $\frac{d}{dt} \frac{dx}{d\theta}$ in addition to the ODE system. For this the solver was provided with the Jacobian of the ODE system. The sensitivities are needed to provide the optimization algorithm with the gradient of the objective function.

5.3.3. Model calibration from experimental data

The unknown model parameters - reaction rate constants, the threshold value and the distribution parameters of the log-normal cell-to-cell variability - have to be estimated from the experimental data. We used time lapse microscopy prometaphase lengths data of *S. pombe* strains with changed abundances of the SAC proteins Mad1-3 in combination with wild type and doubled synthesis rates of Slp1. In this setup a strain is equivalent to an experimental condition and the terms are used interchangeably in the following.

We consider only the phenotypic information in the data, namely whether the SAC is functional (cellular phenotype A) or dysfunctional (cellular phenotype B). Therefore, the dataset of a single experimental condition i is denoted by $\mathcal{D}_i = \{n_{A,i}, n_{B,i}\}$ where n_A and n_B are the number of cells reported for phenotype A and B, respectively. These numbers were derived by a data analysis as described in Chapter 4. MEMO was used to estimate the fraction of cells in each phenotypic subpopulation for every experimental condition. By taking the number of total cells in an experiment condition into account, the number of cells in each of the subpopulations can be calculated. The overall dataset is given by $\mathcal{D} = \{\mathcal{D}_i\}_{i=1\dots d_e}$.

Since there are only two possible outcomes, the likelihood $\mathbb{P}(\mathcal{D}_i|\tilde{\theta})$ of the data given the parameters is given by the binomial distribution

$$\mathbb{P}(\mathcal{D}_i|\tilde{\theta}) = p(n_{A,i}, n_{B,i}|\tilde{\theta}) = \frac{(n_{A,i} + n_{B,i})!}{n_{A,i}!n_{B,i}!} q_{A,i}^{n_{A,i}}(\tilde{\theta}) q_{B,i}^{n_{B,i}}(\tilde{\theta}) \quad (5.15)$$

for each individual experiment e . Here $n_{A,i}$ and $n_{B,i}$ are the number of cells reported for phenotype A and B, respectively in dataset \mathcal{D}_i and $q_{A,i} = \tilde{y}$ and $q_{B,i} = 1 - \tilde{y}$ are the probabilities to observe a cell of phenotype A and B, respectively, as predicted by a model as defined by Equation 5.14 for a given parametrization. The vector of unknown parameters $\tilde{\theta}$ consists of reaction rate constants, the threshold and the scale and location parameters σ and μ of the log-normal densities describing variability in the initial conditions and reaction rate constants. Since the individual experimental conditions are statistically independent, the overall likelihood is obtained by multiplying the likelihoods of the individual experiments

$$\mathbb{P}(\mathcal{D}|\theta) = \prod_{i=1}^I \mathbb{P}(\mathcal{D}_i|\theta). \quad (5.16)$$

An independent optimization of the likelihoods of individual experimental conditions is not possible, as the different experiments share the kinetic parameters, the threshold and the scale and location parameters of the log-normal cell-to-cell variability that do not depend on the experimental condition.

As described in Section 2.4.1, the parameters were estimated in the logarithmic space by minimizing the negative logarithm of 5.16. Since the assumed threshold is artificially introduced for the purpose of modelling, we wanted to ensure robustness of the model predictions with respect to the estimated threshold. To achieve this, in every optimization step the objective function was not only evaluated for the current threshold, but also for thresholds smaller and larger by a factor of 1.221. This factor results from adding 0.2 and -0.2 to the threshold in the logarithmic space. The log-likelihoods obtained for these three thresholds were added and the sum was divided by 3. The resulting values can be interpreted as average likelihood

	WT		30% Mad1		65% Mad2		30% Mad3	
% Slp1	100	200	100	200	100	200	100	200
$q_{B,i}$	0	0.26	0.03	0.65	0.56	0.84	0.37	0.94
$n_{B,i}$	0	30	5	46	92	60	87	45
$n_i = n_{B,i} + n_{A,i}$	451	117	157	71	164	71	234	48

Table 5.2.: Experimental data used for model calibration.

function of the interval $[1/1.221, 1.221] \times \text{threshold}$. By using this average in the optimization, the parameter estimation yielded parameter combinations for which the model output is not sensitive with respect to the threshold.

We used the data of eight strains to calibrate the models. These are the wild type, 30% Mad1, 65% Mad2 and 30% Mad3 strains with both normal (100 %) and doubled (200 %) rate of Slp1 synthesis (Table 5.2).

5.3.4. Assessment of model fit

Parameter estimation, as described in the previous section, yields a MLE of the parameter vector. To assess how well the model, parametrized with the MLE, describes the data, we made use of the 98% confidence interval, as given by the two-sided 98% confidence level of the probability density derived under the hypothesis that the MLE parametrized model is correct. Since there are two subpopulations whose probabilities of occurrence sum to 1, this probability mass is given by a binomial distribution

$$P(x = k) = \binom{n}{k} q_B^k (1 - q_B)^{n-k}. \quad (5.17)$$

The two parameters of this distribution are given by the probability q_B to observe a cell of population B, as given by the fraction obtained from the MLE parametrized model, and the total number of cells n in the experiment. If the experimentally determined fraction fell within the confidence interval, we would conclude that there is significant evidence that the model could be correct. If the experimental observation was not included in the interval, we would have to reject the model's correctness on the 1% level of significance.

5.4. M1: A model for Slp1 inhibition via MCC formation

The statistical analysis of the experimental data detects a split into two populations when Mad2 or Mad3 are decreased or Slp1 is increased relative to wild type conditions. This hints to a mechanism involving MCC formation to be the source of the population split. In the following sections a model of sequestration of Slp1 into the MCC is presented and analysed in order to address Problem 5.1 and Problem 5.3. This includes the model's calibration with experimental data and the assessment of the uncertainty associated with this calibration. Due to the size of the model and its analytical steady states, a rigorous uncertainty analysis with Bayesian methods was feasible. While we focus on the aspect of APC:Slp1 inhibition by

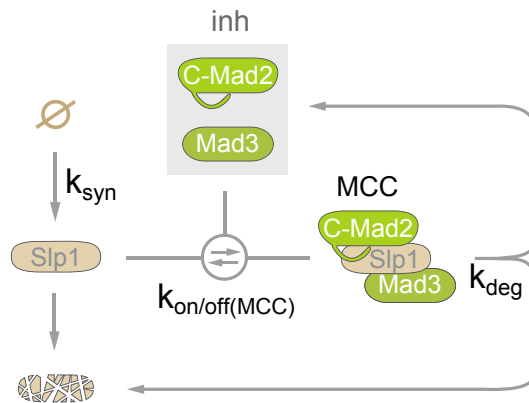


Figure 5.3.: Schematic representation of species and reactions of Model M1. (Figure adapted from Heinrich *et al.* (2013))

the indirect mechanism of Slp1 sequestration in this section, we will extend the model to the direct inhibition of APC:Slp1 in Section 5.5.

5.4.1. M1: Model formulation

We formulated a core model of MCC formation based on the following information:

- (i) Slp1 is synthesized in mitosis (Yamada *et al.*, 2000).
- (ii) Slp1 is an unstable protein with a half-life in the range of 15 min (Heinrich *et al.*, 2013; Sczaniecka *et al.*, 2008).
- (iii) Accumulation of Slp1 is not drastically different in cells with or without an active checkpoint (Heinrich *et al.*, 2013). We therefore assume that the degradation rates of Slp1 and of Slp1 as part of the MCC are similar.
- (iv) Slp1 reaches approx. 20 nM (Heinrich *et al.*, 2013).
- (v) Maximal Slp1 concentration is reached in about 120 min after start of mitosis at 16 °C (the temperature at which we assessed checkpoint activity) (Heinrich *et al.*, 2013).
- (vi) Mad2 and Mad3 bind Slp1 as stoichiometric inhibitors (Burton & Solomon, 2007; Chao *et al.*, 2012; Fang, 2002).
- (vii) Mad2 and Mad3 are stable proteins (Heinrich *et al.*, 2013; Sczaniecka *et al.*, 2008). Hence, synthesis and degradation can be neglected.

To reduce the complexity of the model, we merged all proteins that are known to inhibit Slp1 in one species termed inhibitor. The inhibitor is analogous to Mad2 and Mad3 that is competent to bind Slp1 ('active' Mad2/Mad3). Binding of the inhibitor to Slp1 leads to the formation of the MCC (Figure 5.3). Using mass action kinetics, the model is described by the

following ODEs:

$$\frac{d[\text{Slp1}]}{dt} = k_{syn} - k_{deg}[\text{Slp1}] + k_{off(MCC)}[\text{MCC}] - k_{on(MCC)}[\text{Slp1}][\text{inh}] \quad (5.18)$$

$$\frac{d[\text{inh}]}{dt} = k_{deg}[\text{MCC}] + k_{off(MCC)}[\text{MCC}] - k_{on(MCC)}[\text{Slp1}][\text{inh}] \quad (5.19)$$

$$\frac{d[\text{MCC}]}{dt} = k_{on(MCC)}[\text{Slp1}][\text{inh}] - k_{off(MCC)}[\text{MCC}] - k_{deg}[\text{MCC}] \quad (5.20)$$

with initial conditions

$$\begin{aligned} [\text{Slp1}](0) &= 0 \\ [\text{inh}](0) &= [\text{inh}]_T \\ [\text{MCC}](0) &= 0 \end{aligned}$$

in which $[\text{Slp1}]$ denotes the concentration of Slp1, $[\text{inh}]$ denotes the concentration of the inhibitor, $[\text{MCC}]$ denotes the concentration of the Slp1:inhibitor complex; k_{syn} is the synthesis rate and k_{deg} the degradation rate of Slp1 and $[\text{inh}]_T$ denotes the total amount of the inhibitor; $k_{on(MCC)}$ and $k_{off(MCC)}$ are binding and dissociation rates of Slp1 and the inhibitor. The degradation rate of Slp1 within the MCC is assumed to be equal to the degradation rate of free Slp1, k_{deg} (see (iii)). The model fulfils the conservation relation

$$[\text{inh}]_T = [\text{inh}] + [\text{MCC}].$$

The model state considered in the output mapping is the concentration of free Slp1, $[\text{Slp1}]$. The threshold in $[\text{Slp1}]$ should be low, because even low levels of mammalian Cdc20 efficiently promote anaphase (Malureanu *et al.*, 2010; Wolthuis *et al.*, 2008). We assume the synthesis rate of Slp1 k_{syn} as well as $[\text{inh}]_T$ in the different strains to exhibit cell-to-cell variability. They are assumed to be log-normally distributed with parameters μ_{syn} and σ_{syn} and $\mu_{[\text{inh}]_T}$ and $\sigma_{[\text{inh}]_T}$, respectively.

5.4.2. M1: Model calibration

To estimate the unknown parameters from data we used the number of cells with active and inactive SAC under checkpoint-activating conditions from eight different strains (Table 5.2). We employed maximum likelihood estimation to determine the optimal model parameters. We estimated the kinetic parameters $k_{on(MCC)}$, $k_{off(MCC)}$, k_{deg} and the distribution parameters μ_{syn} , σ_{syn} , $\mu_{[\text{inh}]_T, WT}$, $\sigma_{[\text{inh}]_T, WT}$, $\mu_{[\text{inh}]_T, 30\%Mad1}$, $\sigma_{[\text{inh}]_T, 30\%Mad1}$, $\mu_{[\text{inh}]_T, 65\%Mad2}$, $\sigma_{[\text{inh}]_T, 65\%Mad2}$, $\mu_{[\text{inh}]_T, 30\%Mad3}$, $\sigma_{[\text{inh}]_T, 30\%Mad3}$ and the Slp1 threshold. This yields in total 14 parameters that we constrained to ranges resulting from the following constraints:

- CV of k_{syn} between 0.05 and 0.5.
- lower bound of mean of k_{syn} : 0.17 mol/ min; calculated from 20 nM Slp1 after 120 min (see (iv) and (v)), assuming no degradation.
- upper bound of mean of k_{syn} : 1.98 mol/ min calculated from 20 nM Slp1 after 120 min, assuming the upper bound for the degradation rate.

	σ_{syn}	μ_{syn}	$\sigma_{[inh]_T}$	$\mu_{[inh]_T}$	k_{deg}
min	0.05	-1.793	0.05	-0.1116	0.01733 min ⁻¹
max	0.4724	0.5717	0.47238	3.9108	0.09902 min ⁻¹

Table 5.3.: Parameter boundaries for estimation.

- degradation rates should result in a Slp1 half-life between 7 and 40 min (see (ii)).
- $1 \times 10^{-5} \text{ nM}^{-1} \text{ min}^{-1} < k_{on} < 1 \times 10^5 \text{ nM}^{-1} \text{ min}^{-1}$.
- $1 \times 10^{-5} \text{ nM} < K_d = k_{off}(MCC)/k_{on}(MCC) < 1 \times 10^5 \text{ nM}$.
- $1 \text{ nM} < E([inh]_T) = < 50 \text{ nM}$ (Heinrich *et al.*, 2013) with corresponding CV between 0.05 and 0.5.
- Slp1 threshold for anaphase onset between 0.1 and 20 nM.

The dissociation constant K_d and the CV and mean of the total inhibitor concentration $[inh]_T$ are lumped parameters of several biological parameters that are not included in this simple model. Hence, these model parameters do not have an exact biological equivalent. The parameter boundaries for the estimation were derived from these constraints (Table 5.3 see Appendix E for a derivation).

Multi-start optimization using the sigma-point method yielded maximum likelihood estimates for the kinetic and distribution parameters and the Slp1 threshold (Tables 5.4 and 5.5). Based on the MLE the means of the total inhibitor concentrations and of the synthesis rate k_{syn} can be calculated together with the respective CVs (Table 5.6). The MLE tells us, that model and data fit best for a scenario in which there is very little noise in the abundance of the inhibitor under wild type conditions (CV \approx 0.05). The same holds for the 30% Mad1 strain (CV \approx 0.06). In contrast, according to the MLE, the inhibitor abundance has to exhibit substantially more noise in the 30% Mad3 strain (CV \approx 0.22) and to be even more noisy in the 65% Mad2 strain (CV \approx 0.5) to allow for a split of the population into the observed fractions. The variability in the rate of Slp1 synthesis is moderately high (CV \approx 0.31).

5.4.3. M1: Assessment of model fit

To assess how well the model, parametrized with the MLE, describes the data, we utilized binomial confidence intervals (Figure 5.4). We find that for the nominal threshold of the MLE (x-axis value = 1) the experimentally observed fractions are inside the 98 % confidence intervals for all strains. By varying the threshold, we find that the model fit is not sensitive to the choice of the threshold. We conclude that M1 can describe the main characteristics of the process while satisfying our requirement to be robust with respect to the threshold for anaphase activation.

5. Mechanistic models of SAC signalling in heterogeneous populations

k_{deg}	$K_d = k_{off}(MCC)/k_{on}(MCC)$	$k_{on}(MCC)$	threshold
0.0353 min^{-1}	$6.6156 \times 10^{-5} \text{ nM}$	$392.1326 \text{ nM}^{-1} \text{ min}^{-1}$	0.1034 nM

Table 5.4.: MLE of kinetic parameters and threshold resulting from multi-start optimization.

$[\text{inh}]_T$	wild type	30% Mad1	65% Mad2	30% Mad3	k_{syn}
μ	3.8734	3.5366	2.9306	3.0791	0.3096
σ	0.0507	0.0572	0.4719	0.2167	-0.4077

Table 5.5.: MLE of distribution parameters of inhibitor concentrations and synthesis rate resulting from multi-start optimization.

	wild type	30% Mad1	65% Mad2	30% Mad3	k_{syn}
	$[\text{inh}]_T$	$[\text{inh}]_T$	$[\text{inh}]_T$	$[\text{inh}]_T$	
mean	48.1675 nM	34.4063 nM	20.9464 nM	22.2551 nM	0.6978 / min
CV	0.0507	0.0573	0.4995	0.2192	0.3172

Table 5.6.: Means and CVs of lognormally distributed entities calculated from MLE.

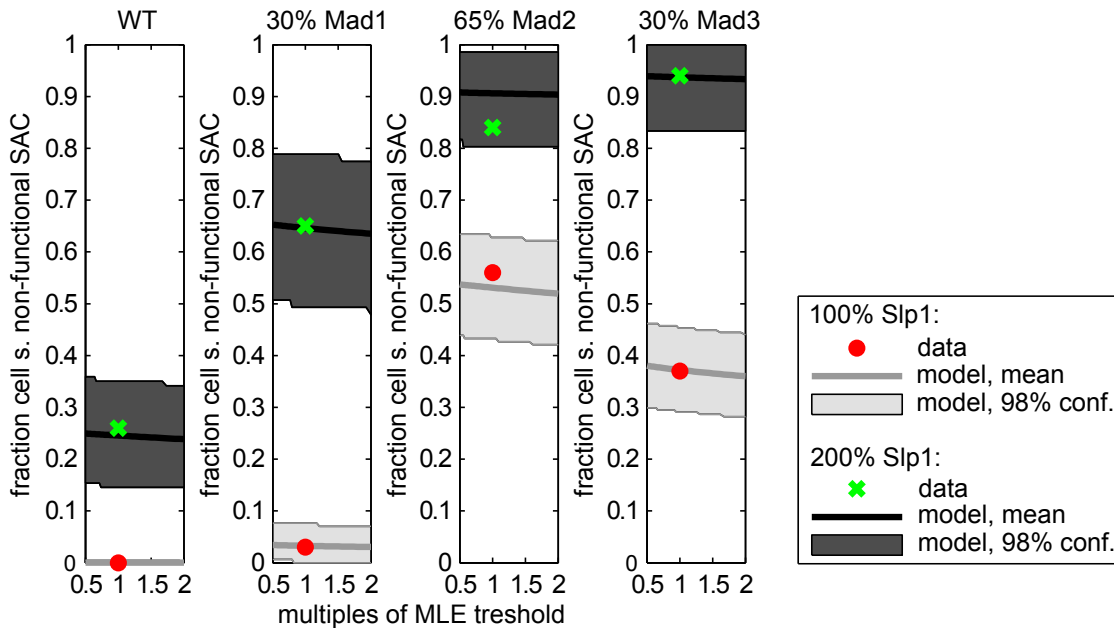


Figure 5.4.: Assessment of model fit for Model M1 exemplary for fraction of cells with non-functional SAC (subpopulation B). Binomial 98% confidence intervals (light resp. dark grey area) of MLE model fit over a range of threshold values around the best fit for each experimental condition. Bold lines indicate the resulting fraction of cells in population B when assuming the corresponding threshold. Green crosses and red circles indicate the experimentally observed fractions of cells with non-functional SAC in 200% and 100% Slp1 backgrounds, respectively.

5.4.4. M1: Simulation based model analysis

A representative sample of single-cell trajectories for model M1 gives an impression how the model in cooperation with cell-to-cell variability leads to a population split (Figure 5.5). The trajectory plots reveal that the Slp1 response is highly heterogeneous within the simulated populations. Depending on the strain, many cells keep very small values of [Slp1] (indicating a functional SAC, population A), while others reach high [Slp1] levels above the threshold (indicating a non-functional SAC, population B).

To analyse how inhibitor concentration and Slp1 synthesis rate differ between populations A and B, we evaluated the parameter distributions corresponding to the individual populations. The values of the two parameters were sorted in two groups depending on whether the simulated [Slp1] trajectories exceeded or remained below the threshold in the simulation. For each group the frequency distributions of the inhibitor concentrations and the Slp1 synthesis rates were computed (Figure 5.6). We find that the distributions of inhibitor concentrations differ only slightly between population A and B for most strains. The Slp1 synthesis rate allows for a better discrimination in this simulation study but this rate cannot be measured experimentally.

5. Mechanistic models of SAC signalling in heterogeneous populations

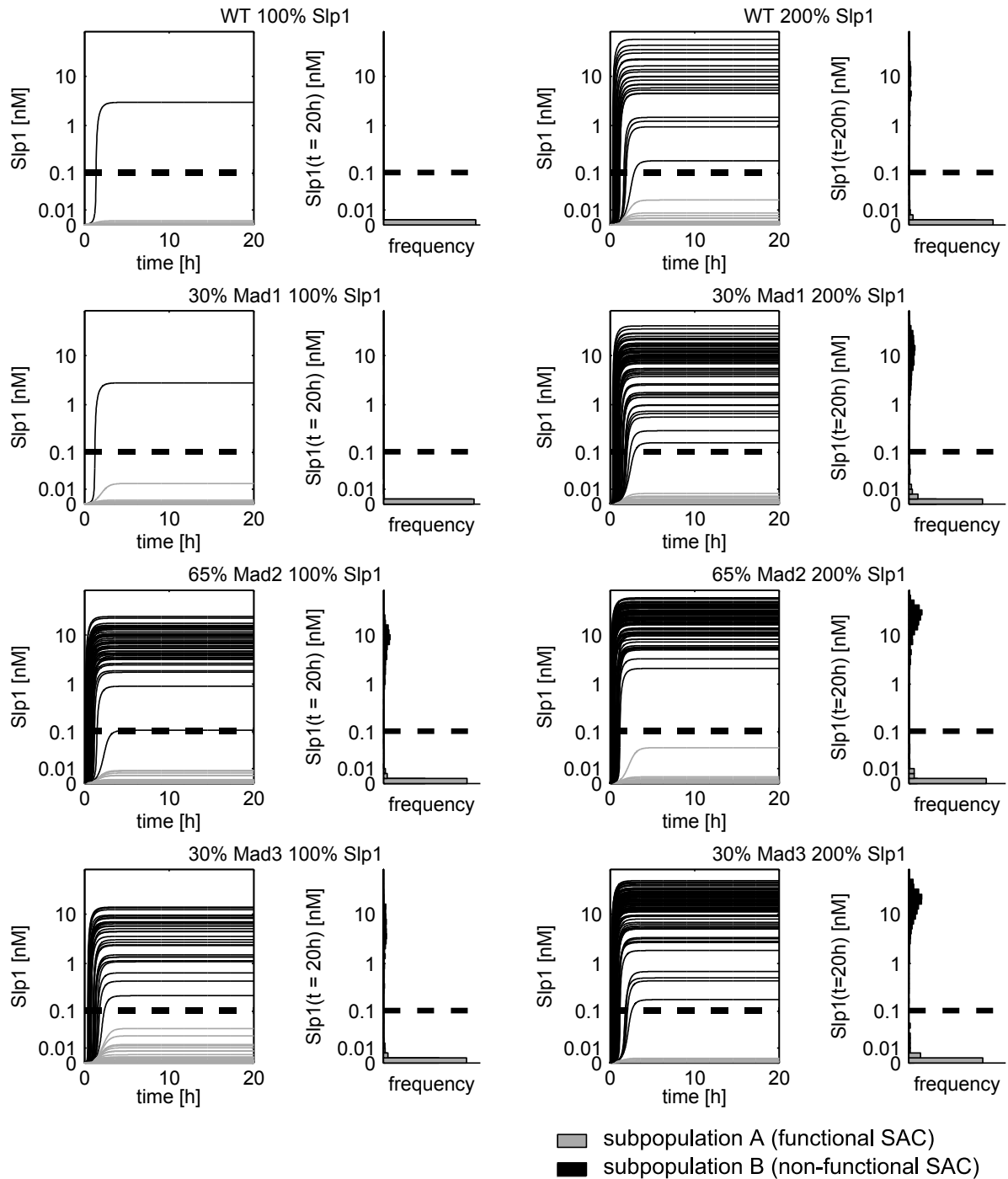


Figure 5.5 (previous page): A sample of 100 cells was simulated by sampling inhibitor concentration and Slp1 synthesis rate for each cell from the estimated distribution for each strain. For every strain the time-dependent concentration of Slp1 in individual cells is shown. Trajectories printed in black cross the threshold (dashed line) and therefore represent cells with non-functional SAC (population B). Trajectories staying below the threshold are plotted in grey. They represent cells with functional SAC (population A). Note that we use a nonlinear y-axis, which is roughly linear for $[Slp1]$ below 0.01 and becomes progressively logarithmic ($Y = \log([Slp1]+0.01)$) above 0.01. The scale is related to the logicle scale used for the visualization of flow cytometry data. The frequency distribution of the Slp1 concentration at the end of the simulation is plotted on the right of every trajectory plot.

5.4.5. M1: Bayesian uncertainty analysis

By calibrating Model M1 with experimental data we could show that this model can reproduce the data while showing robustness with respect to the threshold that determines the phenotype. However, the question remained whether these conclusions drawn from the best fit are a general feature of the model given the data, or whether the data allowed to fit the model comparably well with qualitatively different features. This question can be answered by an uncertainty analysis. One useful approach especially if one is interested in multi dimensional relations are sampling based Bayesian methods as introduced in Section 2.4.2.

Sampling based methods allow to assess the multi-variate posterior distribution of the parameters in a Bayesian framework. However, sampling demands for extensive model simulation. Even by using a sigma-point based approximation, the simulation of the time-dependent system is computationally intensive. Therefore, a rigorous uncertainty analysis of model M1 was impracticable. However, the estimation results above suggested that the system almost reached its steady state after 20 hours. We therefore decided to focus on the steady state of M1 to perform an uncertainty analysis. The advantage of this steady state is its existing analytical solution (see below). Without the need to simulate the model, it is feasible to assess the features population by an individual based approach. For this purpose the analytical solution of the steady state of single cells were calculated from random samples of the distribution of inhibitor and Slp1 synthesis rates. The probabilities p_A and p_B then were efficiently computed by calculating the ratio of cells below and above the threshold. The efficient computation of p_A and p_B enables the fast evaluation of the likelihood function and thus a rigorous sampling based uncertainty analysis.

To study the uncertainty of the parameters of M1, we employed a Bayesian approach with a flat prior constraint to the parameter set specified above (Table 5.3). To explore the parameter set we employed adaptive Markov chain Monte Carlo (adaptive MCMC) sampling. Using the MATLAB Toolbox DRAM (<http://helios.fmi.fi/~lainema/mcmc/>) we generated a converged MCMC sample of the posterior distribution and evaluated its statistics. The maximum a posteriori parameter estimate of the steady state version of M1 found by MCMC sampling (Tables 5.7 and 5.8) is for many parameters surprisingly close to the estimate for the dynamic version of M1 (Tables 5.4 and 5.5). The fits of the observed data are almost indistinguishable (Figure 5.7). This substantiated our simplification and we analysed the parameter uncertain-

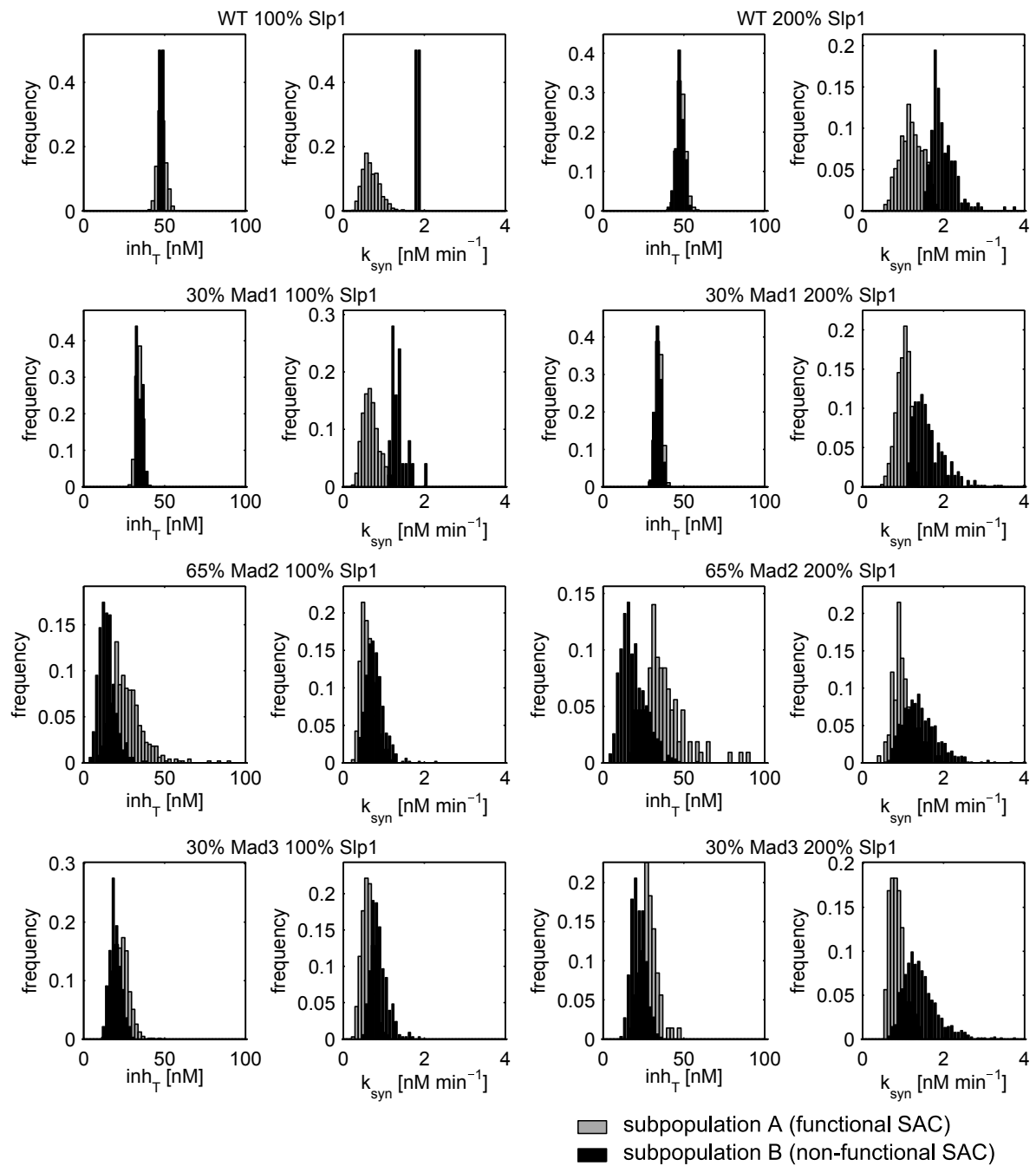


Figure 5.6.: Histograms of the frequency distributions of values of the synthesis rate and the total inhibitor concentration in populations A (grey) and B (black) for all strains as assessed by simulation of Model M1. For better visibility, the histograms of both subpopulations are normalized. Therefore, distributions can appear bimodal although all samples come from unimodal distributions.

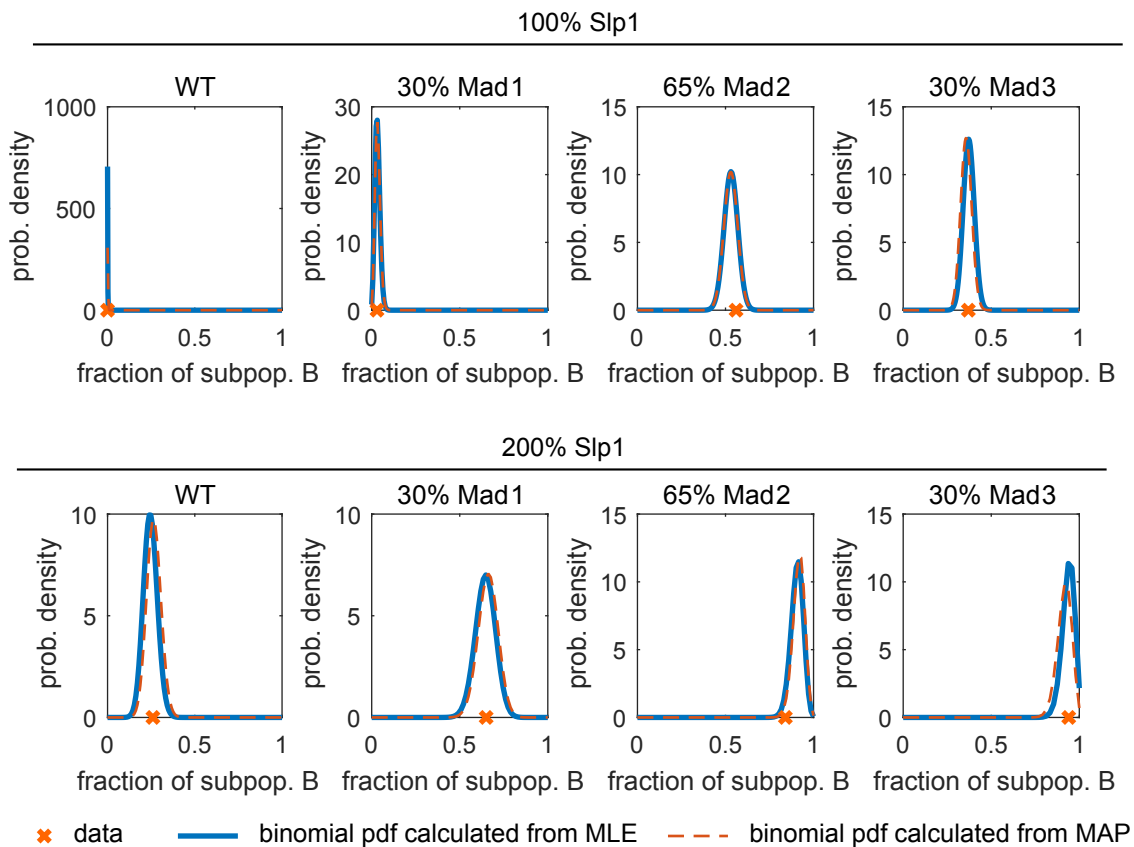


Figure 5.7.: Comparison of the model fits for the dynamic version of Model M1 (MLE, estimated using multi-start optimization) and the steady state version of Model M1 calibrated by MCMC sampling (here the MAP estimate of the MCMC sample is depicted) with the data. The binomial pdfs are calculated from Equation (5.17).

ties based upon the MCMC sample (Figure 5.8). The key finding of this analysis is that the variability in the inhibitor concentration in WT cells has to be small to explain the observed fraction of population A and B. Furthermore, the variability in the Slp1 synthesis rate is high compared to the variability of inhibitor concentrations in WT cells.

5.4.6. M1: Analysis of steady state input output response

The analysis of the kinetic version of M1 revealed that the split of the population is insensitive with respect to the threshold (Figure 5.4) and that the steady states of the Slp1 trajectories appear to lie in two qualitatively different regimes (Figure 5.5). This is a clear hint to ultrasensitivity of the steady state input output response, where small differences in the input are capable of steering the trajectories into qualitatively different steady state regimes. To verify this conjecture and to understand the underlying mechanism, we analysed the steady state properties of M1.

The sequestration of Slp1 into the MCC by Slp1-inhibition competent Mad2/Mad3 in Model M1 is similar to a model of a mechanism called *in vivo* molecular titration, which has been shown to exhibit ultrasensitivity (Buchler & Louis, 2008). Molecular titration is one

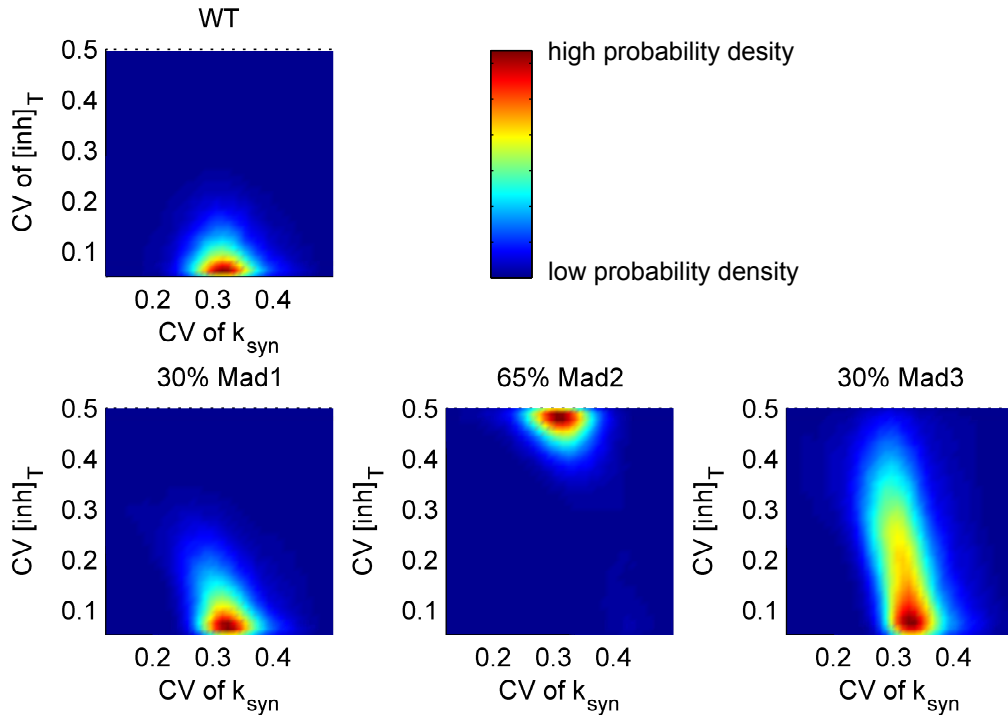


Figure 5.8.: Two dimensional marginal probability densities for the noise (CV) of Slp1 synthesis rate (k_{syn}) and total inhibitor concentration $[inh]_T$ as calculated from a converged MCMC sample of the model parameters.

$\sigma_{k_{syn}}$	$\mu_{k_{syn}}$	k_{deg}	$K_d = k_{off}(MCC)/k_{on}(MCC)$	threshold
0.2908	0.3175	0.0716 / min	0.1066 nM	1.1788 nM

Table 5.7.: MAP estimate of kinetic parameters and threshold determined by MCMC sampling of the steady state version of M1.

$[inh]_T$	wild type	30% Mad1	65% Mad2	30% Mad3
μ	3.8945	3.5694	2.9338	3.1228
σ	0.0733	0.1058	0.4710	0.2882

Table 5.8.: MAP estimate of distribution parameters of total inhibitor concentrations determined by MCMC sampling of the steady state version of M1.

of the ultrasensitivity inducing motifs introduced on Page 62. *In vivo* refers to the additional consideration of the synthesis and degradation of all model species as expected for reactions in living cells (Buchler & Louis, 2008). Model M1 differs from the Buchler model in that the inhibitor (the titrant) is stable (no degradation) but catalyses the degradation of Slp1 (the analyte) via the MCC (the heterodimeric complex). It has been shown that such "catalytic degradation" reactions cut off ultrasensitivity in case that the free titrant itself undergoes degradation as well (Buchler & Louis, 2008).

To determine whether Model M1 can indeed exhibit ultrasensitivity, we performed an analysis of the steady state of M1, the steady state fluxes and the point where the system changes its buffering behaviour (the equivalence point) as a function of the synthesis rate of Slp1 (k_{syn}). In steady state, the fluxes of Slp1 synthesis and degradation are balanced,

$$k_{syn} = k_{deg}[\text{Slp1}] + k_{deg}[\text{MCC}].$$

The total inhibitor abundance $[\text{inh}]_T$ is subject to mass conservation such that

$$[\text{inh}]_T = \text{const.} = [\text{inh}] + [\text{MCC}].$$

Employing these two properties, we can determine an analytical expression for the steady state by setting Equations (5.19) - (5.20) to zero. By keeping only the physically meaningful (positive) solutions of the emerging quadratic equations, the steady state of M1 is given by

$$k_{deg}[\text{Slp1}]_{ss} = \frac{k_{syn} - k_{deg}[\text{inh}]_T - k_{deg}K_D}{2} + \sqrt{\left(\frac{k_{syn} - k_{deg}[\text{inh}]_T - k_{deg}K_D}{2}\right)^2 + k_{syn}k_{deg}K_D}, \quad (5.21)$$

$$k_{deg}[\text{inh}]_{ss} = \frac{-k_{syn} + k_{deg}[\text{inh}]_T - k_{deg}K_D}{2} + \sqrt{\left(\frac{-k_{syn} + k_{deg}[\text{inh}]_T - k_{deg}K_D}{2}\right)^2 + k_{deg}K_D[\text{inh}]_T}, \quad (5.22)$$

$$k_{deg}[\text{MCC}]_{ss} = \frac{k_{syn} + k_{deg}[\text{inh}]_T + k_{deg}K_D}{2} + \sqrt{\left(\frac{k_{syn} + k_{deg}[\text{inh}]_T + k_{deg}K_D}{2}\right)^2 - k_{syn}k_{deg}[\text{inh}]_T}. \quad (5.23)$$

With

$$K_D = \frac{k_{off}(\text{MCC}) + k_{deg}}{k_{on}(\text{MCC})} = \frac{[\text{inh}]_{ss}[\text{Slp1}]_{ss}}{[\text{MCC}]_{ss}}$$

we denote the so called *in vivo* dimer dissociation constant. It is given by the ratio of educts and products in steady state (in chemistry termed equilibrium) *in vivo*, i.e. under the consideration of synthesis and/or degradation of the involved species.

Equation (5.21) relates the steady state amount of free Slp1 to the total amount of inhibitor $[\text{inh}]_T$, the synthesis rate of Slp1 k_{syn} and the remaining kinetic parameters of the model. The evaluation of Equation (5.21) for the MAP parameter estimate over a range of values for k_{syn} points to ultrasensitive behaviour for a range of k_{syn} values as determined from a log log plot (Figure 5.9). The evaluation for different values of $[\text{inh}]_T$ reveals that increasing values of $[\text{inh}]_T$ shift the response curve and the ultrasensitive region to higher values of k_{syn} . Hence $[\text{inh}]_T$ sets a threshold in k_{syn} .

Given the analytical solution of the Slp1 steady state (Equation (5.21)), we set out to analyse its features with respect to k_{syn} and $[\text{inh}]_T$. By employing the local sensitivity measure as defined in Section 5.1.4, the sensitivity of the steady state input output response of M1 reads

$$S = \frac{d\log([\text{Slp1}]_{ss})}{d\log(k_{syn})} = \frac{2k_{syn} \left(\frac{k_{syn} + K_D k_{deg} - k_{deg} [\text{inh}]_T}{2\sqrt{(K_D k_{deg} - k_{syn} + k_{deg} [\text{inh}]_T)^2 + 4K_D k_{deg} k_{syn}}} + \frac{1}{2} \right)}{k_{syn} - K_D k_{deg} + \sqrt{(K_D k_{deg} - k_{syn} + k_{deg} [\text{inh}]_T)^2 + 4K_D k_{deg} k_{syn}} - k_{deg} [\text{inh}]_T} \quad (5.24)$$

The maximum of S can be determined by calculating the derivative of Equation (5.24) with respect to k_{syn} and setting it to zero. This yields

$$S_{max} = \frac{1}{2} + \frac{\sqrt{\frac{[\text{inh}]_T}{K_D} + 1}}{2},$$

as the maximum sensitivity of $[\text{Slp1}]_{ss}$ with respect to changes in k_{syn} . The ratio $\frac{[\text{inh}]_T}{K_D}$ is called the *in vivo* stoichiometric binding parameter (Buchler & Louis, 2008). The value of this ratio is an indicator whether a binding reaction operates under stoichiometric binding conditions. Stoichiometric binding conditions are defined as conditions in which the chemical equilibrium favours complex formation over the free educts (Correia & Detrich III, 2009). Therefore, stoichiometric binding conditions are indicated by $\frac{[\text{inh}]_T}{K_D} > 1$ or $[\text{inh}]_T > K_D$. Hence, the maximum sensitivity of M1 will always be greater than 1 under stoichiometric binding conditions and M1 accordingly exhibits ultrasensitive behaviour.

The synthesis rate at which the sensitivity takes its maximum value is

$$k_{syn} = k_{deg}([\text{inh}]_T + K_D).$$

This is the synthesis rate up to which the degradation and binding are able to roughly balance synthesis. It represents the analogy to the equivalence point in titration theory. To reveal the mechanism behind the ultrasensitive behaviour, we characterized the steady state of Slp1 for values of k_{syn} below and above the equivalence point by performing a first order Taylor series expansion in k_{syn} and $k_{deg}([\text{inh}]_T + K_D)$ similar to Buchler & Louis (2008) (cf. Appendix D). For $k_{syn} \ll k_{deg}([\text{inh}]_T + K_D)$ the Taylor series yields

$$[\text{Slp1}]_{ss} \approx \frac{\frac{k_{syn}}{k_{deg}}}{\frac{[\text{inh}]_T}{K_D} + 1} \quad (5.25)$$

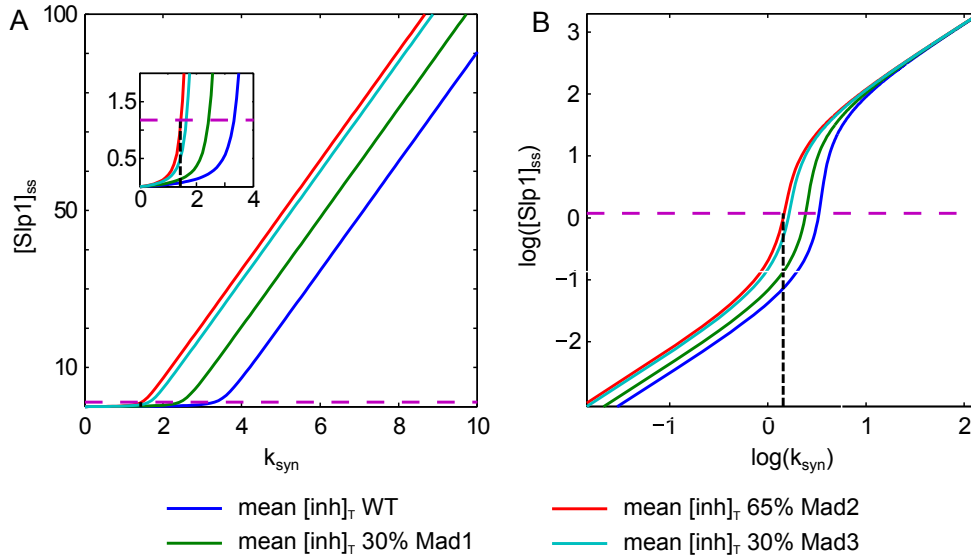


Figure 5.9.: Steady state of [Slp1] as a function of k_{syn} if M1 is parametrized with the MAP estimate for total inhibitor concentration $[inh]_T$ estimated in Mad strains. (A) shows a linear representation with the inset zooming in the area around the origin. (B) shows the double logarithmic representation to indicate the ultrasensitive range. The dotted black line marks the estimated mean value of k_{syn} . The dashed line indicates the Slp1 threshold that determines SAC functionality.

Equation (5.25) reveals that for small values of k_{syn} the steady state abundance of Slp1 is linear in the ratio of the synthesis and degradation rate constants (Figure 5.9). This ratio equals the steady state concentration that would result if the synthesis of Slp1 was only balanced by its degradation, i.e. without consideration of complex formation. This maximum amount is scaled down by a factor determined by the stoichiometric binding parameter. Therefore, the stoichiometric binding parameter regulates the strength of buffering in this regime. High values lead to strong buffering of free Slp1. For $k_{syn} \gg k_{deg}([inh]_T + K_D)$ the Taylor series expansion in k_{syn} yields

$$[Slp1]_{ss} \approx \frac{k_{syn}}{k_{deg}} - ([inh]_T + K_D). \quad (5.26)$$

In this regime (above the equivalence point) the amount of free Slp1 in steady state is again proportional to the maximum amount possible given the ratio of synthesis and degradation. However, this amount is reduced by the subtraction of the total inhibitor concentration plus the *in vivo* dissociation constant. Therefore, free Slp1 grows sub-linearly with the synthesis rate while the inhibitor loses more and more influence with growing synthesis rate.

In summary, the steady state response of Model M1 consists of three regimes. For low values of k_{syn} the accumulation of free [Slp1] is buffered by degradation and complex formation, with the buffering strength depending on the *in vivo* dimer dissociation constant K_D . For sufficiently high values of K_D almost no free Slp1 accumulates. For high values of k_{syn} the buffering capacity is exceeded by the synthesis of Slp1 and free Slp1 accumulates. Around the equivalence point, situated at the transition between these two regimes, ultrasensitivity arises.

The ultrasensitivity is caused by the transition from strong buffering with almost no Slp1 to a strong increase of free Slp1 after the inhibitor loses the influence on the steady state. Hence, Model M1 exhibits ultrasensitivity with respect to the rate of Slp1 synthesis, despite catalysed degradation of Slp1.

5.5. M2: A model for MCC conveyed APC inhibition

In this section Model M2, an extended version of Model M1, is introduced to address Problem 5.1 and Problem 5.3 by considering more biological knowledge. While Model M1 comprises only the SAC signalling reactions that lead to the prevention of the accumulation of free Slp1, M2 also captures the steps that link the SAC to the cell cycle. This is the direct inhibition of APC/C's activation via complex formation with the MCC, in addition to its inhibition by sequestration of Slp1. The APC/C, in turn, causes the disintegration of the MCC by mediating the degradation of Slp1 from APC/C bound MCC. Such double negative feedbacks are known to be sources of ultrasensitivity (Ferrell & Ha, 2014c). This chapter deals with the setup, parametrization and analysis of Model M2. Since the model extension causes the loss of an analytical solution for the steady state, a sampling based assessment of the parameter uncertainty is no longer feasible. Furthermore, the steady state input output response is analysed by employing a simulation based approach.

5.5.1. M2: Model formulation

We extended model M1 described in Chapter 5.4 by reactions based on the following additional information:

- (viii) APC/C is a stable complex (Schwanhäusser *et al.*, 2011). Hence, synthesis and degradation can be neglected.
- (ix) APC/C is inhibited by binding to the MCC (Fang, 2002; Herzog *et al.*, 2009; Sudakin *et al.*, 2001).
- (x) APC/C is activated by Slp1 (Fang *et al.*, 1998; Kramer *et al.*, 1998).
- (xi) Slp1 is degraded as part of the MCC when bound to the APC/C (Foster & Morgan, 2012; Ge *et al.*, 2009; Ma & Poon, 2011; Mansfeld *et al.*, 2011; Nilsson *et al.*, 2008; Pan & Chen, 2004; Reddy *et al.*, 2007; Uzunova *et al.*, 2012).
- (xii) In analogy to M1, we assume APC/C-independent degradation of free Slp1, although the Slp1 ortholog Cdc20 is degraded in an APC/C-dependent manner (Foe *et al.*, 2011).

The additional assumptions lead to a model with six species (Figure 5.10). Using mass

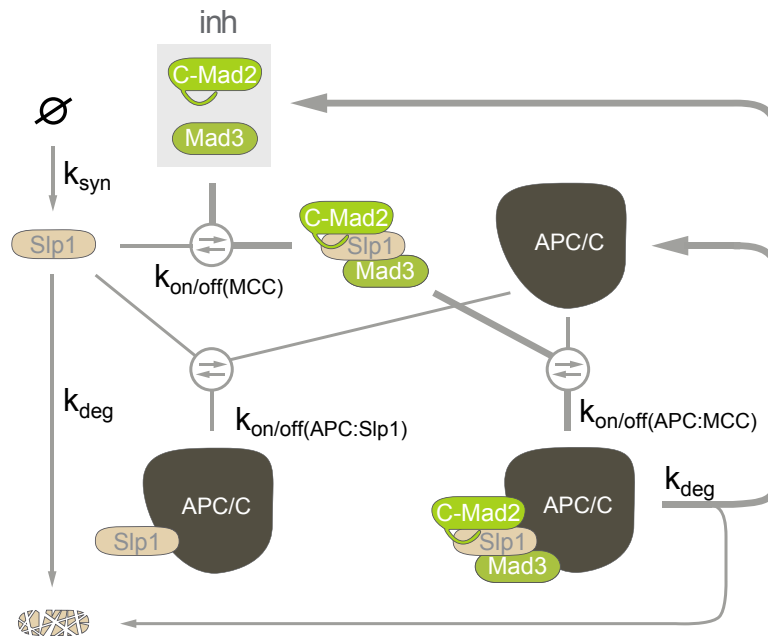


Figure 5.10.: Schematic representation of species and reactions of Model M2. (Figure adapted from Heinrich *et al.* (2013))

action kinetics, the model is described by the following ODEs:

$$\frac{d[\text{APC}]}{dt} = -v_1 - v_2 + v_5 \quad (5.27)$$

$$\frac{d[\text{APC} : \text{MCC}]}{dt} = v_2 - v_5 \quad (5.28)$$

$$\frac{d[\text{APC} : \text{Slp1}]}{dt} = v_1 \quad (5.29)$$

$$\frac{d[\text{Slp1}]}{dt} = -v_1 - v_3 + v_4 - v_6 \quad (5.30)$$

$$\frac{d[\text{inh}]}{dt} = v_5 - v_3 \quad (5.31)$$

$$\frac{d[\text{MCC}]}{dt} = -v_2 + v_3 \quad (5.32)$$

with

$$v_1 = k_{on(\text{APC:Slp1})}[\text{APC}][\text{Slp1}] - k_{off(\text{APC:Slp1})}[\text{APC} : \text{Slp1}]$$

$$v_2 = k_{on(\text{APC:MCC})}[\text{APC}][\text{MCC}] - k_{off(\text{APC:MCC})}[\text{APC} : \text{MCC}]$$

$$v_3 = k_{on(\text{MCC})}[\text{inh}][\text{Slp1}] - k_{off(\text{APC:MCC})}[\text{MCC}]$$

$$v_4 = k_{syn}$$

$$v_5 = k_{deg}[\text{APC} : \text{MCC}]$$

$$v_6 = k_{deg}[\text{Slp1}]$$

and initial conditions

$$\begin{aligned}
 [\text{APC}](0) &= [\text{APC}]_T \\
 [\text{APC} : \text{MCC}](0) &= 0 \\
 [\text{APC} : \text{Slp1}](0) &= 0 \\
 [\text{Slp1}](0) &= 0 \\
 [\text{inh}](0) &= [\text{inh}]_T \\
 [\text{MCC}](0) &= 0
 \end{aligned}$$

in which $[\text{APC}:\text{MCC}]$ denotes the concentration of the APC/C:MCC complex in which APC/C is inhibited, $[\text{APC}:\text{Slp1}]$ denotes the concentration of the active APC/C:Slp1 complex, $[\text{MCC}]$ denotes the concentration of the Slp1:inhibitor complex, $[\text{Slp1}]$ denotes the concentration of free Slp1, $[\text{APC}]$ denotes the concentration of APC/C, and $[\text{inh}]$ denotes the concentration of the inhibitor. As before, the inhibitor is analogous to Mad2/Mad3 that is competent to bind Slp1 ('active' Mad2/Mad3). The model parameters are the Slp1 synthesis and degradation rates, k_{syn} and k_{deg} , and the binding and dissociation rates of different complexes, $k_{on(X)}$ and $k_{off(X)}$. The model fulfils the conservation relations

$$\begin{aligned}
 [\text{APC}]_T &= [\text{APC}] + [\text{APC} : \text{Slp1}] + [\text{APC} : \text{MCC}] \\
 [\text{inh}]_T &= [\text{inh}] + [\text{MCC}] + [\text{APC} : \text{MCC}]
 \end{aligned}$$

in which $[\text{APC}]_T$ denotes the total concentration of APC/C and $[\text{inh}]_T$ denotes the total concentration of the inhibitor. The model state considered in the output mapping is the concentration of APC/C:Slp1, $[\text{APC}:\text{Slp1}]$. We assume the synthesis rate of Slp1 k_{syn} as well as $[\text{inh}]_T$ in the different strains, to exhibit cell-to-cell variability. They are assumed to be log-normally distributed with parameters μ_{syn} and σ_{syn} and $\mu_{[\text{inh}]_T}$ and $\sigma_{[\text{inh}]_T}$, respectively.

5.5.2. M2: Model calibration

In order distinguish between qualitatively different outcomes (functional SAC vs. dysfunctional SAC) for different cells, we assume that APC/C:Slp1 needs to reach a threshold for anaphase to occur. The synthesis rate of Slp1 (k_{syn}) as well as $[\text{inh}]_T$ in the different strains are assumed to be log-normally distributed with parameters μ and σ . This yields in total 19 parameters that we constrained, as before, to the following ranges:

- CV of k_{syn} between 0.05 and 0.5.
- lower bound of mean of k_{syn} : 0.17 mol/ min; calculated from 20 nM Slp1 after 120 min (see (iv) and (v)), assuming no degradation.
- upper bound of mean of k_{syn} : 1.98 mol/ min calculated from 20 nM Slp1 after 120 min, assuming the upper bound for the degradation rate.
- degradation rates should result in a Slp1 half-life between 7 and 40 min (see (ii)).
- $1 \times 10^{-5} \text{ nM}^{-1} \text{ min}^{-1} < k_{on} < 1 \times 10^5 \text{ nM}^{-1} \text{ min}^{-1}$.

k_{deg}	$[APC]_T$	threshold in $[APC : Slp1]$
0.0616 min^{-1}	19.2690 nM	1.0147 nM

Table 5.9.: MLE of total APC, Slp1 degradation rate and threshold of Model M2 resulting from multi-start optimization.

$k_{on}(MCC)$	$k_{on}(APC:MCC)$	$k_{on}(APC:Slp1)$
$1.4225 \times 10^3 \text{ min}^{-1} \text{ nM}^{-1}$	$1.9263 \text{ min}^{-1} \text{ nM}^{-1}$	$4.0955 \text{ min}^{-1} \text{ nM}^{-1}$
$K_d(MCC)$	$K_d(APC:MCC)$	$K_d(APC:Slp1)$
$22.7780 \times 10^{-4} \text{ nM}$	2.5582 nM	0.2560 nM

Table 5.10.: MLE of complex binding and dissociation rate constants resulting from multi-start optimization.

- $1 \times 10^{-5} \text{ nM} < K_d = k_{off}/k_{on} < 1 \times 10^5 \text{ nM}$.
- $1 \text{ nM} < E([inh]_T) < 50 \text{ nM}$ (Heinrich *et al.*, 2013) with corresponding CV between 0.05 and 0.5.
- APC/C:Slp1 threshold for anaphase onset between 0.1 and 20 nM.

The dissociation constant K_d and the CV and mean of the total inhibitor concentration $[inh]_T$ are lumped parameters of several biological parameters that are not included in this simplified model. Hence, these model parameters do not have an exact biological equivalent. The parameter boundaries for estimation were derived from these constraints as before (Table 5.3, see Appendix E for a derivation).

Multi-start optimization using the sigma-point method yielded maximum likelihood estimates for the total amount of APC/C, the threshold in APC/C:Slp1 and the degradation rate (Table 5.9) and for the complex binding and dissociation rates and distribution parameters (Tables 5.10 and 5.11). Based on the MLE the means of the total inhibitor concentrations and of the synthesis rate k_{syn} can be calculated together with the respective CVs (Table 5.12). As already observed for Model M1, model and data fit best for a scenario in which there is very little noise in the abundance of the inhibitor under wild type conditions of inhibitor concentration ($CV \approx 0.09$). The same holds for the 30% Mad1 strain ($CV \approx 0.08$). In contrast, according to the MLE, the inhibitor abundance has to exhibit substantially more noise in the 30% Mad3 strain ($CV \approx 0.21$) and to be even more noisy in the 65% Mad2 strain ($CV \approx$

$[inh]_T$	wild type	30% Mad1	65% Mad2	30% Mad3	k_{syn}
μ	3.2600	2.6577	1.9059	2.1000	-1.0654
σ	0.0922	0.0845	0.3854	0.2087	0.3446

Table 5.11.: MLE of distribution parameters of inhibitor concentrations and synthesis rate resulting from multi-start optimization.

	wild type [inh] _T	30% Mad1 [inh] _T	65% Mad2 [inh] _T	30% Mad3 [inh] _T	k_{syn}
mean	26.1606 nM	14.3145 nM	7.2440 nM	8.3459 nM	0.3684 min ⁻¹
CV	0.0924	0.0847	0.4002	0.2110	0.3551

Table 5.12.: Means and CVs of log-normally distributed entities calculated from MLE.

0.4) to allow for a split of the population into the observed fractions. The variability in the rate of Slp1 synthesis is moderately high (CV \approx 0.36). The results are very similar to the results obtained with Model M1, not only in the qualitative outcome, but also in the absolute numbers.

5.5.3. M2: Assessment of model fit

For the maximum likelihood estimate we assessed the fit of M2 (Figure 5.11). As before, we utilized binomial confidence intervals. As for M1, we found that M2 can describe the main characteristics of the process. For the nominal threshold of the MLE (x-axis value = 1) the experimentally observed fractions are inside the 98 % confidence intervals for all strains. By varying the threshold, we found that the model fit is not sensitive to the choice of the threshold. Therefore also our requirement to be robust with respect to the threshold for anaphase activation is satisfied.

5.5.4. M2: Simulation based model analysis

While the fits of the experimentally observed fractions of population A and B are similar for M1 and M2, the dynamics of the underlying pathways are quite different (Figure 5.12). While M1 shows a long tail towards high concentrations of the anaphase activation species ([Slp1]), which is a result of the almost linear dependency of the steady state on k_{syn} and [inh]_T above the equivalence point, M2 shows a bimodal distribution in the concentration of the active species ([APC:Slp1]). Individual cells either have [APC:Slp1] close to zero or at high concentrations.

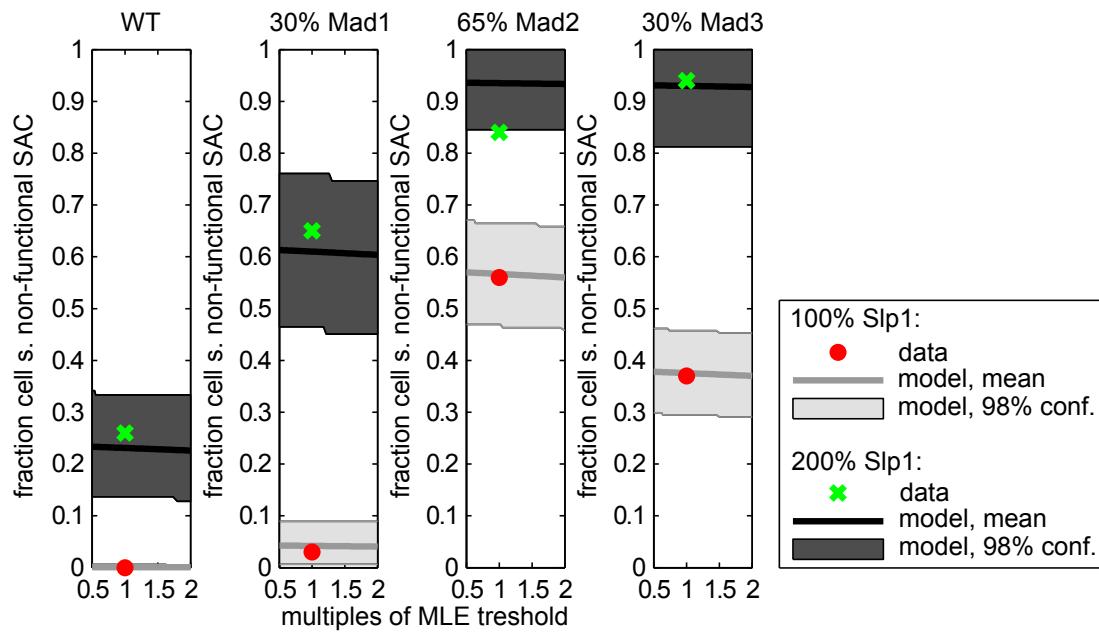


Figure 5.11.: Assessment of model fit for Model M2 exemplarily for fraction of cells with non-functional SAC (subpopulation B). Binomial 98 % confidence intervals (light resp. dark grey area) of MLE model fit over a range of threshold values around the best fit for each experimental condition. Bold lines indicate the resulting fraction of cells in population B when assuming the corresponding threshold. Green crosses and red circles indicate the experimentally observed fractions of cells with non-functional SAC in 200% and 100% Slp1 backgrounds, respectively.

5. Mechanistic models of SAC signalling in heterogeneous populations

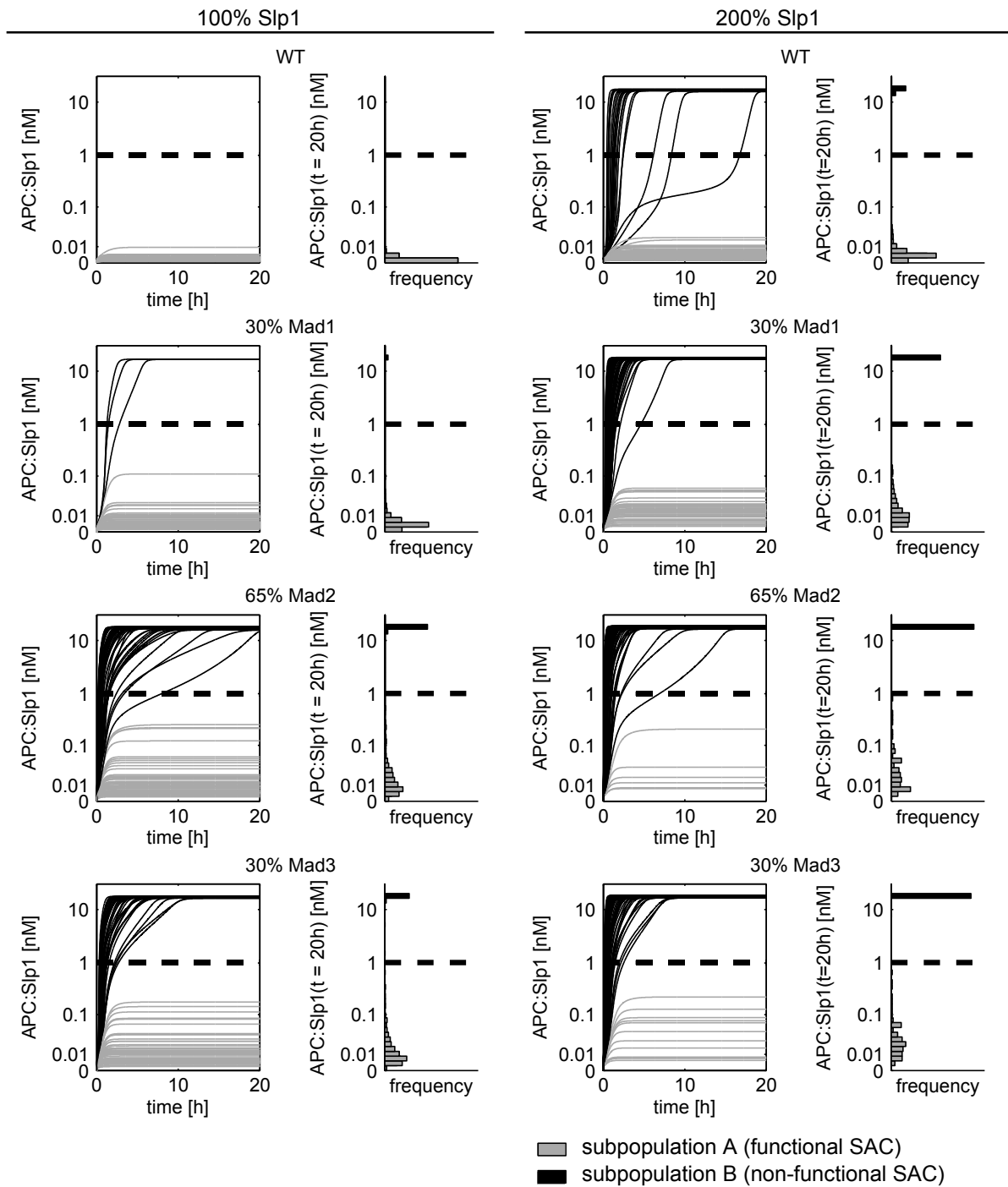


Figure 5.12 (previous page): A sample of 100 cells was simulated by sampling inhibitor concentration and Slp1 synthesis rate for each cell from the estimated distribution for each strain. For every strain the time-dependent concentration of APC:Slp1 in individual cells is shown. Trajectories printed in black cross the threshold (dashed line) and therefore represent cells with non-functional SAC in population B. Trajectories staying below the threshold are plotted in grey. They represent cells with functional SAC (population A). Note that we use a nonlinear y-axis, which is roughly linear for $[APC:Slp1] < 0.01$ and becomes progressively logarithmic ($y = \log([APC : Slp1] + 0.01)$). The scale is related to the logicle scale used for the visualization of flow cytometry data. The frequency distribution of the APC:Slp1 concentration at the end of the simulation is plotted to the right of every trajectory plot.

To analyse whether inhibitor concentration or Slp1 synthesis rate in populations A and B could be distinguished experimentally, we evaluated the parameter distributions corresponding to the individual populations. For $[APC:Slp1]$ trajectories that exceeded or that remained below the threshold, we collected the inhibitor concentrations and the Slp1 synthesis rates and computed the corresponding frequency distributions (Figure 5.13). We find that the distributions of inhibitor concentrations differ only slightly between population A and B for most strains. The Slp1 synthesis rate allows for a better discrimination according to our model, but this rate unfortunately cannot directly be measured experimentally.

5.5.5. M2: Analysis of steady state input output response

The trajectories and frequency plots in Figure 5.12 hint to qualitatively different steady states of the model state $[APC:Slp1]$ that determine the identity of the individual cells. To understand the cause of the bimodality and the resulting robustness to threshold alterations, we analysed the steady state input output response of M2. Due to the lack of an analytical solution for the steady state, an analytical assessment of the steady state properties of M2 is not possible. To test for multi-stability of the model output $[APC : Slp1]$, we performed a simulation based bifurcation analysis with respect to the rate of Slp1 synthesis, k_{syn} (Figure 5.14). Simulations were computed using the maximum likelihood parameters and the estimated mean inhibitor concentration in WT cells. The analysis revealed that (1) For low k_{syn} , Model M2 possesses a globally asymptotically stable steady state with low $[APC:Slp1]$ concentrations below the threshold, corresponding to a functional SAC. (2) For high k_{syn} , there exists a globally asymptotically stable steady state with high $[APC:Slp1]$ above the threshold, corresponding to a dysfunctional checkpoint. (3) For intermediate values of k_{syn} , M2 possesses three steady states of which two are stable and correspond to a functional and a dysfunctional SAC, respectively. These stable steady states are separated by an unstable steady state. This bistability allows for a threshold behaviour with respect to k_{syn} , meaning that below a certain rate of Slp1 synthesis, the critical synthesis rate, the SAC is functional, while above the threshold the SAC is deficient. In the vicinity of the critical synthesis rate ultrasensitivity arises. Since the initial condition for $[APC : Slp1]$ is zero, only the lower steady state is physiologically relevant in the range of k_{syn} in which two stable steady states exist.

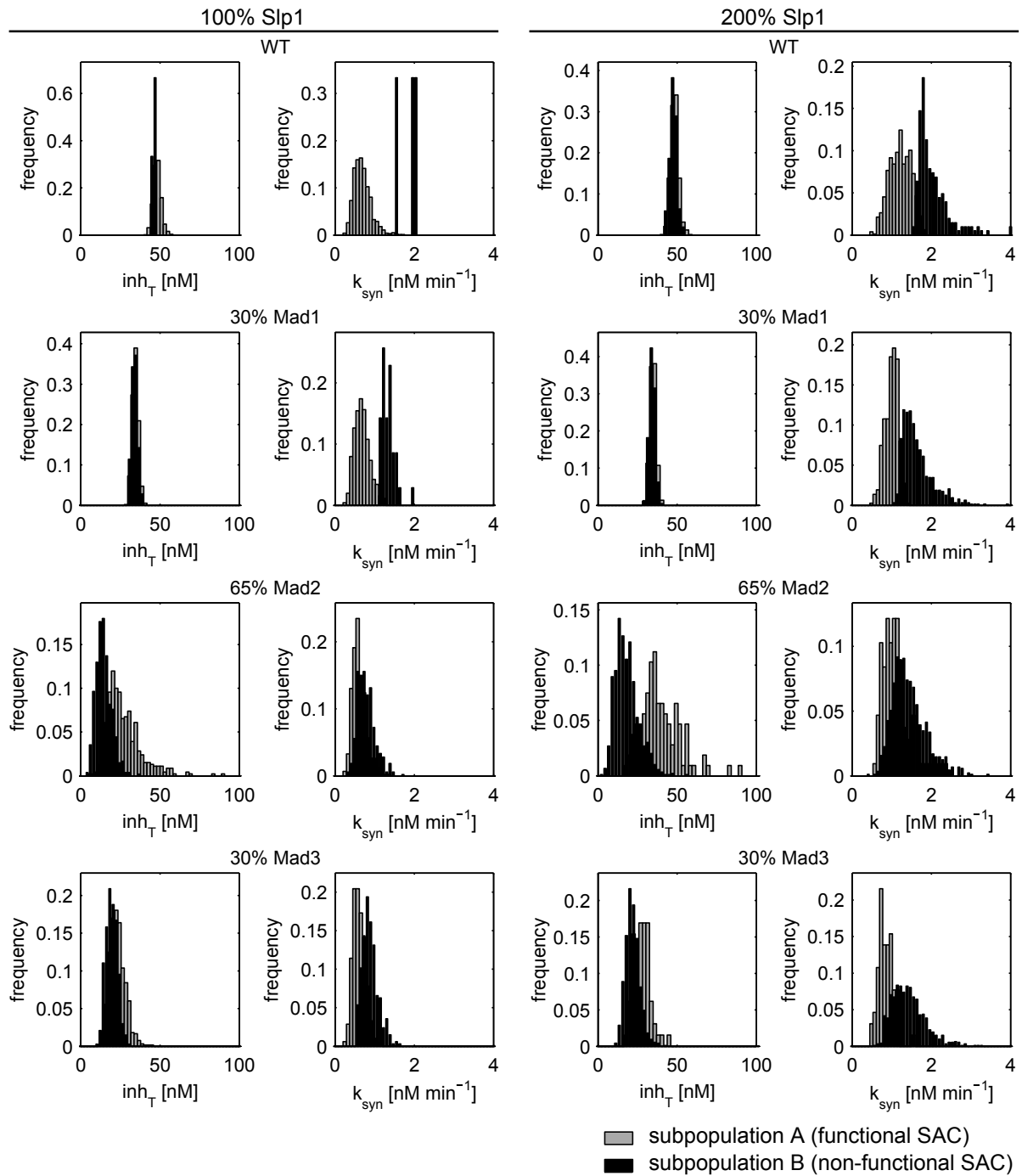


Figure 5.13.: Histograms of the frequency distributions of values of the synthesis rate and the total inhibitor concentration in populations A (grey) and B (black) for all strains as assessed by simulation of Model M2. For better visibility, the histograms of both subpopulations are normalized. Therefore, distributions can appear bimodal although all samples come from unimodal distributions.

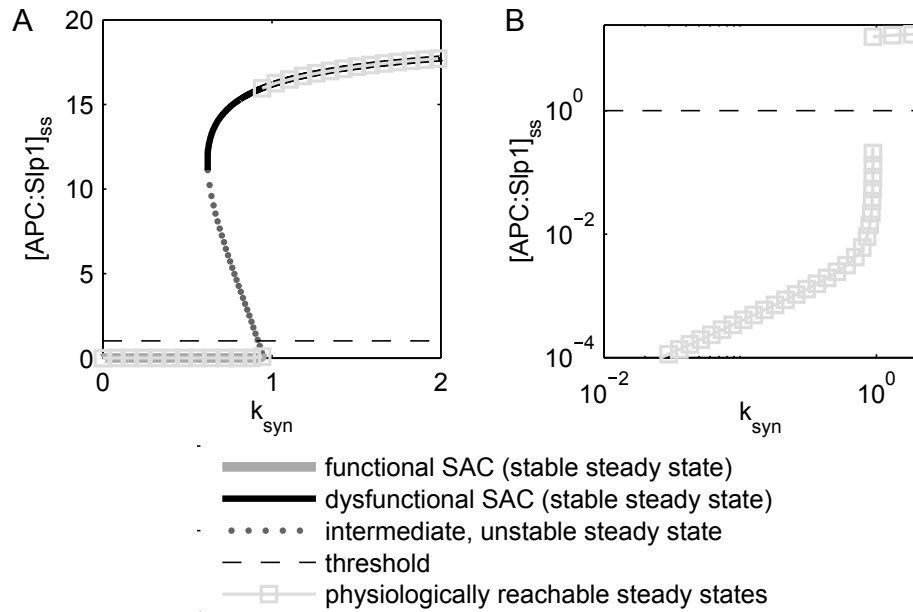


Figure 5.14.: Results of simulation based steady state analysis of Model M2. (A) The steady states of the MLE parametrized model are plotted over a range of values for k_{syn} . Considering the MLE, two saddle node bifurcations lead to the existence of two stable steady states separated by an unstable steady state over a range of Slp1 synthesis rate (k_{syn}) values. Steady states are represented in different colours and line styles according to their stability properties. Since the initial condition of APC:Slp1 is zero, only the lower steady state in the bistable range of k_{syn} can be reached physiologically. At the critical synthesis rate the steady state solution jumps from the lower branch of stable steady states to the high branch, which causes a high sensitivity at this point. (B) The log-log plot of solely the physiologically possible steady states shows the ultrasensitive region of model M2 (slope > 1).

For the maximum likelihood estimates of the parameters and the mean wild type inhibitor concentration, the critical Slp1 synthesis rate is 0.943 nM/min. Furthermore, the bistability and the switch-like change allow for robustness with respect to the threshold as lower and upper steady state are separated.

A necessary condition for bistability is positive feedback (Gouzé, 1998; Thomas, 1981). Although such a feedback might not be obvious from the model structure of Model M2, it is revealed through the corresponding interaction graph (cf. Page 62). Therefore, the systems matrix of signs of the Jacobian matrix of the ODE system (Equations (5.28) to (5.32)) has to be evaluated. The matrix of signs of the Jacobian matrix of Model M2 is given by

$$\text{sgn}(J(x)) = \begin{pmatrix} - & + & + & - & 0 & - \\ + & - & 0 & 0 & 0 & + \\ + & 0 & - & + & 0 & 0 \\ - & 0 & + & - & - & + \\ 0 & + & 0 & - & - & + \\ - & + & 0 & + & + & - \end{pmatrix}. \quad (5.33)$$

The rows of matrix (5.33) indicate the dynamics of every species with the order given by the

model Equations (5.28) to (5.32), while the columns correspond to the species in the same order. Every matrix entry indicates the influence of the species given by the column on the species given by the row. A minus indicates a negative interaction, a plus a positive interaction and zero indicates no interaction between the model states. Positive feedback is indicated by symmetry of the signs with respect to the main diagonal of the matrix. As revealed by (5.33), Model M2 has several positive and double negative feedbacks. As an example, the double negative feedback between the model species [APC] and [MCC] is indicated in green in (5.33) and the double negative feedback between [APC] and [Slp1] is indicated in red. Besides positive feedback, a further requirement for bistability is an ultrasensitive mechanism in the system (reviewed in Ferrell (2002)). Here, this is given by the ultrasensitive mechanism shown for Model M1, which is embedded in M2. Model M2 adds a second step of Slp1 sequestration. Slp1 is now buffered in [MCC] and [APC : MCC]. As soon as the whole inhibitor is involved in buffering, exceeding Slp1 binds to free APC and Slp1 is predominantly buffered in the MCC. The observed ultrasensitivity of Model M2 stems from bistability of the model output with respect to k_{syn} .

5.6. Summary and Discussion

In this chapter, we have devised and analysed two models of the mechanism of protein mediated SAC signalling to investigate the sources of the observed bimodality in SAC functionality and gain insights into the signalling dynamics. Model M1 in Section 5.4 models one aspect of SAC mediated cell cycle inhibition, the inhibition of Slp1 due to sequestration into the mitotic checkpoint complex (MCC). Model M2 in Section 5.5 includes Model 1 and extends it by the activation and inhibition of the APC/C by Slp1 and the MCC, respectively.

As substantiated in Chapter 4, the bimodality in prometaphase length can be interpreted as a split of the isogenic cell population into two subpopulations with different cellular SAC phenotypes: functional and dysfunctional. In the set up of our mechanistic models, phenotype affiliation is determined via a threshold in the model state that represents the molecular species that is an indicator of SAC functionality (Slp1 in Model M1, APC/C:Slp1 in Model M2). By using the available data on subpopulation fractions from eight different *S. pombe* strains to calibrate the parameters, we could show, that both models can reproduce these data. Furthermore, we have confirmed that the model outcomes are robust with respect to the assumed threshold. Hence, we conclude that the population split in the models is not an artefact of our threshold assumption. Therefore, the sources causing the split in the models are also possible sources of the population split *in vivo*.

The analysis of the models revealed ultrasensitivity of the output of both models with respect to the Slp1 synthesis rate, in combination with cell-to-cell variability, to be the source of the population split. Ultrasensitivity in the steady state input output response (small fold changes in the input of a system cause large fold changes in its output) is known to have the ability to transform unimodally distributed cell-to-cell variability in the input into a bimodal distribution in the output (Dobrzyński *et al.*, 2014; Ochab-Marcinek & Tabaka, 2010). Differences between cells are selectively amplified in certain ranges of the input. In our models, unimodal cell-to-cell variability in the Slp1 synthesis rate and in the amount of Slp1 binding competent Mad2/Mad3 is transformed into a bimodal distribution in the phenotype.

The ultrasensitivity of Model M1 is caused by a mechanism related to molecular titration: the total amount of inhibitor and the rate of Slp1 synthesis determine a threshold, separating a regime of buffering from a regime of inhibitor saturation causing excess of free Slp1. Input variability spanning this threshold is transformed into bimodal distributions. By demonstrating ultrasensitivity for Model M1, we could show that *in vivo* molecular titration models can exhibit ultrasensitive behaviour despite of catalytic degradation of the target if the inhibitor is conserved, i.e. does not undergo synthesis or degradation. This refutes earlier conclusions on the effect of catalytic degradation on ultrasensitivity in molecular titration models (Buchler & Louis, 2008).

Model M2 exhibits ultrasensitivity due to bistability in the steady states of the model. There exists a range in the Slp1 synthesis rate for which the model output has two stable steady states: one with a low amount of activated APC and one with a substantially higher amount. Each steady state defines one phenotype. The threshold is situated in between the two steady states. With increasing synthesis rate a critical value is reached for which the system switches from the lower to the upper steady state. This switch corresponds to an extreme sensitivity to fold changes around this synthesis rate. Bistability can be induced by positive or double negative feedback in a system that additionally contains an ultrasensitive mechanism (reviewed in Ferrell (2002)). Model M2 has several of these feedbacks and contains Model M1 as an ultrasensitivity generating mechanism. Its bistable properties sharpen the ultrasensitive behaviour observed in Model M1 and induce the characteristic sigmoid shape of the response curve.

The second aspect in the generation of bimodality in our models, besides ultrasensitivity, is cell-to-cell variability. We have included two sources of cell-to-cell variability in our models: the total amount of the inhibitor and the synthesis rate of Slp1. The MLEs of both models and a rigorous uncertainty analysis of Model M1 using Bayesian methods, indicated higher noise in the Slp1 synthesis rate than in the amount of inhibition competent SAC proteins (Figure 5.8). The low noise in inhibitory SAC proteins is in agreement with the experimental data, showing that the noise in the expression of SAC inhibitory proteins is unusually low (Heinrich *et al.*, 2013). Due to a lack of data on Slp1 variability, the model estimate on Slp1 synthesis variability cannot be evaluated. Given the strong increase in transcription, and potentially translation for Slp1 at the start of mitosis (Heinrich *et al.*, 2013), it is conceivable that this may cause a larger variability.

Simulation studies of both models show that the distribution of inhibitor concentrations differ only slightly between the two phenotypes. This is in agreement with the observation that there is no significant difference between cells from the different phenotypes in the examined strains (Heinrich *et al.*, 2013). The distributions of Slp1 synthesis rates allow for a better discrimination of the phenotypes. This model prediction could be substantiated if it was possible to determine the distribution of Slp1 synthesis rates on the single cell level.

The SAC mechanism has the ability to amplify small differences between cells, the cell-to-cell variability, into opposing phenotypes: a fully functional SAC or a dysfunctional SAC. The ability to create bimodal distributions is usually understood as a strategy for cell populations to survive in fluctuating environments (Fraser & Kaern, 2009; Kussell & Leibler, 2005). In case of SAC signalling, however, variability in the phenotype is actually not appreciated. Given the high sensitivity, to guarantee reliable SAC signalling the level of proteins has to be kept in a tight window. The question remains why such a crucial mechanism is so sensitive to changes in its components abundances. One reason could lie in the dynamics of SAC silencing, which

have to be fast (Ciliberto & Shah, 2009).

Bimodal distributions in protein concentrations are frequently observed in isogenic cell populations. The causes of the bimodality are attributed to two effects: intrinsic stochastic effects and deterministic, noise induced effects. Stochastic effects are commonly observed in gene expression (reviewed in Kaern *et al.* (2005)). Stochasticity arises from slow transitions between promoter states. The resulting burst like protein expression can lead to populations that contain a mixture of expressing and non-expressing cells. Deterministic, noise induced effects are linked to the properties of regulatory and signalling dynamical systems (Kim & Sauro, 2012). Bistability enables switching between two alternative cellular states of gene expression or signalling, inducing bimodality on the population level. But also systems which are not bistable but exhibit an ultrasensitive input output response can amplify the variability in the input signal into a bimodal distribution (Birtwistle *et al.* (2012); Dobrzyński *et al.* (2014), our own results). Interestingly, Ochab-Marcinek & Tabaka (2010) could analytically show for a model of gene expression regulation that bimodality can also arise through non-linear but not ultrasensitive response curves if the noise distribution in the input has certain properties. While gene expression leads to bimodality in the total amount of a protein, solely signalling induced effects cause bimodality, e.g. in the active form of a protein or in the distribution of signal encoding complexes.

We are striving to use the whole distribution as characterized by our statistical models for model calibration. Since we are not aware of an approach to do so, the sole consideration of the quantiles of the event distribution instead of the whole distribution could be a first step in this direction. In this approach the likelihood function could be constructed equivalent to Equation (5.15), but evaluated for several time points given by the quantiles of the prometaphase length distribution. For a particular experimental condition, the times that correspond to the quantiles of the prometaphase length distribution would have to be calculated from our statistical models. Furthermore, the number of cells corresponding to each quantile would have to be determined from the respective quantile and the total number of cells in the subpopulation. These numbers are equivalent to $n_{B,e}$ in Equation (5.15). Then the distribution of the species of interest could be determined at the quantile times by applying the sigma-point based approach. From these distributions the probability mass above the threshold value could be evaluated at the quantile time points. This values correspond to $q_{B,e}$ in Equation (5.15). Given $n_{B,e}$ $q_{B,e}$ at each quantile time the likelihood can be evaluated at each quantile time. The likelihood for the particular experiment is then given by the product of the likelihoods at each quantile time. However, this approach has not been tested so far and its value remains to be evaluated.

6. Conclusion

6.1. Summary and conclusions

In this thesis, the spindle assembly checkpoint (SAC) has been investigated through two model-based approaches. First, we have employed statistical models to quantify and analyse cell-to-cell variability in time-to-event single-cell data on SAC functionality. In addition, we have devised mechanistic models of the SAC signalling molecular pathway to address the system properties of the pathway as a dynamical system.

First, provided with censored experimental data of the duration of the SAC induced cell cycle arrest in single cells, we have developed a flexible framework for the model-based statistical analysis of censored and uncensored single-cell data. The framework is based on mixture models and enables the simultaneous, automated analysis of univariate data acquired under multiple experimental conditions. Therefore, we refer to it as Multi-experiment mixture Modelling or MEMO Geissen *et al.* (2016). MEMO employs maximum-likelihood inference and allows for testing of competing hypotheses. Our framework can be applied to a variety of different single-cell data types as demonstrated for NFG-induced Erk phosphorylation data in Geissen *et al.* (2016). MEMO can help to avoid the misinterpretation of censored data and, since it is available as an easy to use Matlab toolbox, makes statistical analysis readily available for non-statisticians.

We have employed this framework to quantify the variability in the experimental data and to extract the information about the SAC contained in these data. In this way, we have classified phenotypes of SAC functionality by establishing a quantitative linkage between the time the SAC arrests a cell in prometaphase and its phenotype. Furthermore, we have established a quantitative relation between the relative amounts of the key SAC signalling proteins Mad2 and Mad3 and the probability to establish a stable SAC arrest. We have shown that the outcome is highly sensitive to changes in Mad2. With decreasing amount of Mad2, the probability to establish a stable SAC arrest decreases gradually until the protein amount reaches a threshold and the probability drops rapidly to zero. In contrast, cells have a non-zero probability to establish a stable SAC arrest for an almost complete range of the amount of Mad3 between zero and wild type. This indicates that the amounts of Mad2 and Mad3 have different predictive power in predicting the outcome for a single cell, at least in cells of *S. pombe*. Predicting the outcome of SAC functionality based on protein amounts can be interesting to inform clinicians whether the SAC is functional in a certain type of cancer and can be targeted for treatment.

In the second part of this thesis, we have used the phenotypic data on SAC functionality to calibrate mechanistic models of SAC signalling. We have found that the sensitivity we observe in the statistical analysis of the prometaphase length data can be explained with the properties of the dynamical system defined by our models. The dynamical system resulting from the

calibration of our model of APC/C inhibition (Model M2) shows ultrasensitivity in the input output response that is caused by bistability. This bistability is caused by the ultrasensitive motif of MCC formation (Model M1) embedded into multiple positive and double negative feedbacks. Note that we used only mass action kinetics to model the protein interactions. Therefore, the exhibited ultrasensitivity and bistability is an actual property of the dynamical system and not artificially introduced by using ultrasensitive kinetics.

Bistable switches are a common motif in cell cycle signalling. They are mainly found at transitions between cell cycle phases (Novak *et al.*, 2001; Tyson & Novak, 2015). However, in the context of SAC signalling the meaning of such a clear distinction between two opposing states is not obvious. Nevertheless, it would be beneficial to assess the model predicted bistability experimentally. However, the classical experimental approach to assess bistability, which is to start in different initial conditions and determine the resulting steady state, is not feasible *in vivo*. *In vitro* experiments might prove useful, but since the kinetics of the studied system involve synthesis and degradation, the exact setup remains elusive.

In conclusion, our framework for the analysis of censored data can help to prevent misinterpretation of data and the information we gained on SAC signalling reveals previously unknown system properties of the dynamic system and might guide further experiments.

6.2. Outlook

MEMO is currently restricted to the analysis of censored and uncensored univariate data. An extension of MEMO to truncated and multivariate data is possible, the latter poses, however, several challenges. Among others, the evaluation of the likelihood for censored data requires the calculation of multivariate integrals (McLachlan & Jones, 1988), which is already computationally intensive for the bivariate case (Cadez *et al.*, 2002). Potential solutions might be provided by sparse grids (see Burkardt (2014) and references therein). As the analysis of multivariate data is currently not possible with MEMO, to resolve this issue one needs to combine it with preprocessing and dimension reduction approaches (Angerer *et al.*, 2015). Moreover, depending on the experimental setup, prior to analysis, the data may have to be corrected for experimental biases that mask the biological population structure (Buettner *et al.*, 2015). To facilitate the biological interpretation of the results, a hierarchical view on cell populations should be incorporated in MEMO (Usoskin *et al.*, 2015).

Regarding the mechanistic modelling of SAC signalling, it would be interesting to use the information on the prometaphase length distribution of a subpopulation, as gained by statistical modelling, instead of the subpopulation fractions for inference. For this, only the distributions of the subpopulations with dysfunctional SAC can be employed, since only for a dysfunctional SAC the event generating the data (end of prometaphase) is an event which can be captured by the models of SAC signalling we consider. The end of prometaphase in the subpopulation with functional SAC cannot directly be captured by a model of SAC signalling, since in this case other mechanisms such as mitotic slippage come into play (Brito & Rieder, 2006; Gascoigne & Taylor, 2009; Huang *et al.*, 2010). The consideration of distributional information of prometaphase length for model calibration would capture more information about the dynamics of the individual signalling components. It could also lead to more accurate estimates of cell-to-cell variability and therefore to a better understanding of the crucial

factors determining the outcome of signalling.

However, the modelling of event triggered measurement data with ODEs, in particular on the population level, is non-trivial. On the one hand, parameter estimation is challenging since for models with discrete events the analytical gradient of the objective function, which is needed for efficient gradient based optimization, cannot be derived by the available modelling toolboxes (Fröhlich *et al.*, 2016). However, a recent progress in this context is the implementation of sensitivity equations for differential equation models with events in a MATLAB interface for the solver CVODES (Fröhlich *et al.*, 2016). On the other hand, the consideration of events in populations models provides certain challenges in the construction of the objective function. In our present setup, the sigma-point based approach makes the time dependent distributions of the model species available. However, to compare the model to the data an output mapping is needed which maps the distribution of one of the model species to the distribution of events over time. There exists a theory for such threshold crossing problems in the context of stochastic differential equations, i.e. stochastic processes including diffusion, which is applied for problems known as first hitting time or first passage time processes (Iyer-Biswas & Zilman, 2015; Redner, 2001; Valov, 2009). However, whether and how this theory can be adapted for the application to RODEs is beyond the scope of this thesis.

As a first attempt, the sole consideration of the quantiles of the event distribution instead of the whole distribution could be promising for parameter estimation. As described in Section 5.6, the objective function can be constructed similarly to our present one. Whether this approach proves useful remains to be investigated.

Appendix

A. Derivation of the pdf and the cdf for the conditional random variable $X|X \leq C$

Assuming that C and X are independent random variables with pdfs f_X and f_C and cdfs F_X and F_C , we derive in the following the cdf and the pdf of the conditional random variable $X|X \leq C$, which results from the X and C being competing processes such that a realization x of X can only be observed if it is smaller than a realization c of C and vice versa. The conditional random variable $X|X \leq C$ has a cdf given by

$$F_{X|X \leq C}(x) = \frac{P(\{X \leq x\} \cap \{X \leq C\})}{P(X \leq C)} = P(X \leq x | X \leq C)$$

where

$$\begin{aligned} P(\{X \leq x\} \cap \{X \leq C\}) &= P(X \leq x, X \leq C) \\ &= \int_{-\infty}^x \int_x^{\infty} f_X(x) f_C(c) dc dx \\ &= \int_{-\infty}^x f_X(x) (1 - F_C(x)) dx \end{aligned}$$

is the joint distribution, and the marginal distribution is given by

$$\begin{aligned} P(X \leq C) &= \iint_{x \leq c} f_X(x) f_C(c) dc dx = \\ &= \int_{-\infty}^{\infty} \int_x^{\infty} f_X(x) f_C(c) dc dx \\ &= \int_{-\infty}^{\infty} f_X(x) (1 - F_C(x)) dx. \end{aligned}$$

Therefore, the cdf reads

$$F_{X|X \leq C}(x) = \frac{\int_{-\infty}^x f_X(x) (1 - F_C(x)) dx}{\int_{-\infty}^{\infty} f_X(x) (1 - F_C(x)) dx}$$

and the pdf reads

$$f_{X|X \leq C}(x) = \frac{f_X(x) (1 - F_C(x))}{\int_{-\infty}^{\infty} f_X(x) (1 - F_C(x)) dx}.$$

The cdf and pdf of the conditional random variable $C|C \leq X$ can be derived accordingly .

B. Probability densities

The interpretation of the distribution parameters $\varphi_s(u)$ depends on the distribution assumption. In the current version of MEMO four distributions for the subpopulations are supported:

- **Normal distribution:** With distribution parameters $\varphi = (\mu, \sigma)$ the probability density function (pdf) and cumulative distribution function (cdf) are

$$\phi(x|\varphi) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right),$$

and

$$\Phi(x|\varphi) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x \exp\left(-\frac{1}{2}\left(\frac{\tau-\mu}{\sigma}\right)^2\right) d\tau = \Lambda\left(\frac{x-\mu}{\sigma}\right).$$

With $\Lambda(y)$ we denote the cdf of the standard normal distribution $\mathcal{N}(0, 1)$ throughout this documentation.

- **Log-normal distribution:** The pdf and cdf are

$$\phi(x|\varphi) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma x} \exp\left(-\frac{1}{2}\left(\frac{\log(x)-\mu}{\sigma}\right)^2\right) & \text{for } x > 0 \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\Phi(x|\varphi) = \frac{1}{\sqrt{2\pi}\sigma} \int_0^x \frac{1}{\tau} \exp\left(-\frac{1}{2}\left(\frac{\log(\tau)-\mu}{\sigma}\right)^2\right) d\tau = \Lambda\left(\frac{\log(x)-\mu}{\sigma}\right),$$

with distribution parameters $\varphi = (\mu, \sigma)$.

- **Johnson-SU distribution:** The pdf and cdf are

$$\phi(x|\varphi) = \frac{\sigma}{\lambda \sqrt{2\pi} \sqrt{z^2 + 1}} \exp\left(-\frac{1}{2}\left(\gamma + \sigma \sinh^{-1} z\right)^2\right) \quad \text{with } z = \frac{x-\xi}{\lambda} \text{ and } y = \gamma + \sigma \sinh^{-1} z$$

and

$$\Phi(x|\varphi) = \int_0^x \phi(\tau|\varphi) d\tau = \Lambda(y),$$

with distribution parameters $\varphi = (\gamma, \sigma, \lambda, \xi)$.

- **Gamma distribution:** The pdf and cdf are

$$\phi(x|\varphi) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x),$$

$$\Phi(x|\varphi) = \frac{\gamma(\alpha, \beta x)}{\Gamma(\alpha)} = \frac{1}{\Gamma(\alpha)} \int_0^{\beta x} \tau^{\alpha-1} \exp(-\tau) d\tau,$$

with distribution parameters $\varphi = (\alpha, \beta)$. $\Gamma(\alpha)$ denotes the gamma function evaluated at α and $\gamma(\alpha, \beta x)$ the lower incomplete gamma function.

C. Derivation of the gradient of the objective function

To facilitate the convergence and reduce the computational cost of the optimization, MEMO provides the optimization routine with the gradient of the objective function with respect to the parameter vector θ . In the following subsections we provide all gradients needed for the case of uncensored data (y_i^j), interval censored data (\bar{y}_i^j) and right censored data (\underline{y}_i^j), assuming a model for the censoring. For maximum generality the distribution parameters φ are assumed to be functions of θ .

C.1. Gradient of the log-likelihood function

Since the objective function is the logarithm of the likelihood, the gradient of this log-likelihood function is provided.

For the case of uncensored data the gradient reads

$$\frac{d \log(\mathbb{P}(\mathcal{D}_i|\theta))}{d\theta} \propto \frac{d}{d\theta} \log \left(\prod_{j=1}^{n_{y,i}} f_{X_i}(y_i^j|\theta, u_i) \right) = \sum_{j=1}^{n_{y,i}} \frac{d}{d\theta} \log (f_{X_i}(y_i^j|\theta, u_i)) = \sum_{j=1}^{n_{y,i}} \frac{1}{A} \frac{df_{X_i}(y_i^j|\theta, u_i)}{d\theta},$$

with $A = f_{X_i}(y_i^j|\theta, u_i)$.

For the case of interval censored data the gradient reads

$$\begin{aligned} \frac{d \log(\mathbb{P}(\mathcal{D}_i|\theta))}{d\theta} &\propto \frac{d}{d\theta} \log \prod_{l=1}^{n_{\bar{y},i}} f_{\bar{Y}_i}(\bar{y}_i^l|\theta, u_i) \\ &= \sum_{l=1}^{n_{\bar{y},i}} \frac{d}{d\theta} \log (f_{\bar{Y}_i}(\bar{y}_i^l|\theta, u_i)) \\ &= \sum_{l=1}^{n_{\bar{y},i}} \frac{1}{B} \frac{d}{d\theta} (F_{X_i}(\bar{y}_i^l|\theta, u_i) - F_{X_i}(\underline{y}_i^l - \Delta x|\theta, u_i)) \end{aligned}$$

with $B = F_{X_i}(\bar{y}_i^l|\theta, u_i) - F_{X_i}(\underline{y}_i^l - \Delta x|\theta, u_i)$.

For the case of right censored data the gradient reads

$$\begin{aligned}
\frac{d \log(\mathbb{P}(\mathcal{D}_i|\theta))}{d\theta} &\propto \frac{d}{d\theta} \log \left(\prod_{j=1}^{n_{y,i}} f_{X_i}(y_i^j|\theta, u_i)(1 - F_{C_i}(y_i^j|\theta, u_i)) \right) \left(\prod_{k=1}^{n_{\bar{y},i}} f_{C_i}(\bar{y}_i^k|\theta, u_i)(1 - F_{X_i}(\bar{y}_i^k|\theta, u_i)) \right) \\
&= \sum_{j=1}^{n_{y,i}} \frac{d}{d\theta} \log(f_{X_i}(y_i^j|\theta, u_i)(1 - F_{C_i}(y_i^j|\theta, u_i))) + \sum_{k=1}^{n_{\bar{y},i}} \frac{d}{d\theta} \log(f_{C_i}(\bar{y}_i^k|\theta, u_i)(1 - F_{X_i}(\bar{y}_i^k|\theta, u_i))) \\
&= \sum_{j=1}^{n_{y,i}} \frac{1}{C} \frac{d}{d\theta} (f_{X_i}(y_i^j|\theta, u_i)(1 - F_{C_i}(y_i^j|\theta, u_i))) + \sum_{k=1}^{n_{\bar{y},i}} \frac{1}{D} \frac{d}{d\theta} (f_{C_i}(\bar{y}_i^k|\theta, u_i)(1 - F_{X_i}(\bar{y}_i^k|\theta, u_i))) \\
&= \sum_{j=1}^{n_{y,i}} \frac{1}{C} \left(\frac{df_{X_i}(y_i^j|\theta, u_i)}{d\theta} (1 - F_{C_i}(y_i^j|\theta, u_i)) - f_{X_i}(y_i^j|\theta, u_i) \frac{dF_{C_i}(y_i^j|\theta, u_i)}{d\theta} \right) \\
&\quad + \sum_{k=1}^{n_{\bar{y},i}} \frac{1}{D} \left(\frac{df_{C_i}(\bar{y}_i^k|\theta, u_i)}{d\theta} (1 - F_{X_i}(\bar{y}_i^k|\theta, u_i)) - f_{C_i}(\bar{y}_i^k|\theta, u_i) \frac{dF_{X_i}(\bar{y}_i^k|\theta, u_i)}{d\theta} \right)
\end{aligned}$$

with $C = f_{X_i}(y_i^j|\theta, u_i)(1 - F_{C_i}(y_i^j|\theta, u_i))$ and $D = f_{C_i}(\bar{y}_i^k|\theta, u_i)(1 - F_{X_i}(\bar{y}_i^k|\theta, u_i))$.

For the case of interval and right censored data the gradient reads

$$\begin{aligned}
\frac{d \log(\mathbb{P}(\mathcal{D}_i|\theta))}{d\theta} &= \frac{d}{d\theta} \log \left(\prod_{l=1}^{n_{\bar{y},i}} f_{X_i}(\bar{y}_i^l, \theta, u_i) \prod_{k=1}^{n_{\bar{y},i}} f_{C_i}(\bar{y}_i^k, \theta, u_i) \right) \\
&= \sum_{l=1}^{n_{\bar{y},i}} \frac{d}{d\theta} \left(\log \left((F_{X_i}(\bar{y}_i^l|\theta, u_i) - F_{X_i}(\bar{y}_i^l - \Delta x|\theta, u_i)) (1 - F_{C_i}(\bar{y}_i^l|\theta, u_i)) \right) \right) \\
&\quad + \sum_{k=1}^{n_{\bar{y},i}} \frac{d}{d\theta} \left(\log \left((F_{C_i}(\bar{y}_i^k|\theta, u_i) - F_{C_i}(\bar{y}_i^k - \Delta x|\theta, u_i)) (1 - F_{X_i}(\bar{y}_i^k|\theta, u_i)) \right) \right) \\
&= \sum_{l=1}^{n_{\bar{y},i}} \frac{1}{G} \left(\int_{\bar{y}_i^l - \Delta x}^{\bar{y}_i^l} \frac{df_{X_i}(x|\theta, u_i)}{d\theta} (1 - F_{C_i}(x|\theta, u_i)) - f_{X_i}(x|\theta, u_i) \frac{dF_{C_i}(x|\theta, u_i)}{d\theta} dx \right) \\
&\quad + \sum_{k=1}^{n_{\bar{y},i}} \frac{1}{H} \left(\int_{\bar{y}_i^k - \Delta x}^{\bar{y}_i^k} \frac{df_{C_i}(x|\theta, u_i)}{d\theta} (1 - F_{X_i}(x|\theta, u_i)) - f_{C_i}(x|\theta, u_i) \frac{dF_{X_i}(x|\theta, u_i)}{d\theta} dx \right)
\end{aligned}$$

with $G = \int_{\bar{y}_i^l - \Delta x}^{\bar{y}_i^l} f_{X_i}(x|\theta, u_i)(1 - F_{C_i}(x|\theta, u_i))dx$ and $H = \int_{\bar{y}_i^k - \Delta x}^{\bar{y}_i^k} f_{C_i}(x|\theta, u_i)(1 - F_{X_i}(x|\theta, u_i))dx$.

In case f_{X_i} or f_{C_i} are given by a mixture distribution $p(x|\theta)$, these equations include gradients of mixture distributions, which are derived in the following section. The gradients of non-mixture probability densities can be found in Section C.3.

C.2. Gradient of the mixture distribution

The gradient of the mixture distribution

$$p(x|\theta) = \sum_{s=1}^S w_s(\theta)\phi(x|\theta),$$

with respect to the parameter vector θ is

$$\frac{dp(x|\theta)}{d\theta} = \sum_{s=1}^S \frac{dw_s(\theta)}{d\theta} \phi(x|\theta) + w_s(\theta) \frac{d\phi(x|\theta)}{d\theta}.$$

The gradient of the cumulative mixture distribution

$$P(x|\theta) = \sum_{s=1}^S w_s(\theta)\Phi(x|\theta),$$

with respect to the parameter vector θ is

$$\frac{dP(x|\theta)}{d\theta} = \sum_{s=1}^S \frac{dw_s(\theta)}{d\theta} \Phi(x|\theta) + w_s(\theta) \frac{d\Phi(x|\theta)}{d\theta}.$$

C.3. Gradient of the probability densities

In the following the gradients of the different types of probability densities ϕ (see B) and cumulative probability functions (cdf) Φ with respect to the parameter vector θ are provided for different distribution types.

- **Normal distribution:**

With

$$y = \frac{x - \mu}{\sigma}$$

the gradients of the pdf and cdf are

$$\begin{aligned} \frac{d\phi(x|\theta)}{d\theta} &= \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left(-\frac{1}{2}y^2\right) \left(y \frac{\partial\mu}{\partial\theta} + (y^2 - 1) \frac{\partial\sigma}{\partial\theta} \right), \\ \frac{d\Phi(x|\theta)}{d\theta}(x) &= -\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}y^2\right) \left(\frac{d\mu}{d\theta} + y \frac{d\sigma}{d\theta} \right). \end{aligned}$$

- **Log-normal distribution:**

With

$$y = \frac{\log(x) - \mu}{\sigma}$$

the gradients of the pdf and cdf are

$$\begin{aligned} \frac{d\phi(x|\theta)}{d\theta} &= \frac{1}{\sqrt{2\pi}\sigma^2 x} \exp\left(-\frac{1}{2}y^2\right) \left(y \frac{\partial\mu}{\partial\theta} + (y^2 - 1) \frac{\partial\sigma}{\partial\theta} \right), \\ \frac{d\Phi(x|\theta)}{d\theta} &= -\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}y^2\right) \left(\frac{d\mu}{d\theta} + y \frac{d\sigma}{d\theta} \right). \end{aligned}$$

• **Johnson-SU distribution:**

With

$$z = \frac{x - \bar{y}}{\lambda} \quad \text{and} \quad y = \gamma + \sigma \sinh^{-1} z$$

the gradients of the pdf and cdf are

$$\begin{aligned} \frac{d\phi(x|\theta)}{d\theta} &= \frac{\sigma z^2 + \sigma^2 z y \sqrt{z^2 + 1} - \sigma(z^2 + 1)}{\lambda^2(z^2 + 1)^{\frac{3}{2}}} \frac{\partial \lambda}{\partial \theta} + \frac{1 - \sigma y \sinh^{-1} z}{\lambda \sqrt{z^2 + 1}} \frac{\partial \sigma}{\partial \theta} \\ &\quad + \frac{\sigma z + \sigma^2 y \sqrt{z^2 + 1}}{\lambda^2(z^2 + 1)^{\frac{3}{2}}} \frac{\partial \bar{y}}{\partial \theta} - \frac{\sigma y}{\lambda \sqrt{z^2 + 1}} \frac{\partial \gamma}{\partial \theta}, \\ \frac{d\Phi(x|\theta)}{d\theta} &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y^2\right) \left(-\frac{\sigma}{\lambda \sqrt{1+z^2}} \left(z \frac{\partial \lambda}{\partial \theta} + \frac{\partial \bar{y}}{\partial \theta} \right) + \frac{\partial \gamma}{\partial \theta} + \sinh^{-1} z \frac{\partial \sigma}{\partial \theta} \right). \end{aligned}$$

• **Gamma distribution:** The gradients of the pdf and cdf are

$$\begin{aligned} \frac{d\phi(x|\theta)}{d\theta} &= \phi(x|\theta) \left(\log(\beta) - \underbrace{\frac{\Gamma'(\alpha)}{\Gamma(\alpha)}}_{=\Psi(\alpha)} + \log(x) \right) \frac{\partial \alpha}{\partial \theta} + \phi(x|\theta) \left(\frac{\alpha}{\beta} - x \right) \frac{\partial \beta}{\partial \theta}, \\ \frac{d\Phi(x|\theta)}{d\theta} &= \left(\frac{\partial \gamma(\alpha, \beta x)}{\partial \alpha} - \Phi(x|\theta) \Psi(\alpha) \right) \frac{\partial \alpha}{\partial \theta} + \frac{(\beta x)^{\alpha-1} \exp(-\beta x)}{\Gamma(\alpha)} \frac{\partial \beta}{\partial \theta}, \end{aligned}$$

in which $\Gamma(\alpha)$ denotes the gamma function evaluated at α ,

$$\Gamma'(\alpha) = \frac{d\Gamma(\alpha)}{d\alpha} = \int_0^\infty x^{\alpha-1} \log(x) \exp(-x) dx \quad \text{and} \quad \Psi(\alpha) \text{ denotes the digamma function.}$$

D. Approximation of $[\text{Slp1}]_{ss}$ in Model M1 by Taylor series expansions

The steady state of $[\text{Slp1}]$ in Model M1 is given by

$$\begin{aligned} [\text{Slp1}]_{ss} &= \frac{1}{k_{deg}} \frac{k_{syn} - k_{deg}[\text{inh}]_T - k_{deg}K_D}{2} \\ &\quad + \frac{1}{k_{deg}} \sqrt{\left(\frac{k_{syn} - k_{deg}[\text{inh}]_T - k_{deg}K_D}{2} \right)^2 + k_{syn}k_{deg}K_D}. \end{aligned}$$

We are interested in the behaviour for $k_{syn} \ll k_{deg}([\text{inh}]_T + K_D)$ and $k_{syn} \gg k_{deg}([\text{inh}]_T + K_D)$. Algebraic manipulations lead to

$$\frac{[\text{Slp1}]_{ss}}{K_D} = \frac{\frac{k_{syn}}{k_{deg}K_D} - \left(\frac{[\text{inh}]_T}{K_D} + 1 \right) + \sqrt{\left(\frac{k_{syn}}{k_{deg}K_D} - \left(\frac{[\text{inh}]_T}{K_D} + 1 \right) \right)^2 + 4 \frac{k_{syn}}{k_{deg}K_D}}}{2}.$$

To simplify we set $x := \frac{k_{syn}}{k_{deg}K_D}$ and $y := (\frac{[\text{inh}]_T}{K_D} + 1)$, which yields

$$f(x,y) = \frac{x - y + \sqrt{(x-y)^2 + 4x}}{2}$$

and the regimes of interest become $x \ll y$ and $x \gg y$. To approximate $f(x,y)$ for $x \ll y$ we perform a Taylor series expansion of $x \mapsto f(x,y_0)$ around $x_0 = 0$ for $y_0 > 0$:

$$\begin{aligned} T_{f(0,y_0)}(x) &= f(0,y_0) + \frac{\partial f}{\partial x}(0,y_0)x + O(x^2) \\ &\stackrel{y_0 > 0}{=} \frac{1}{y_0}x + O(x^2), \end{aligned}$$

with

$$\frac{\partial f}{\partial x} = \frac{1}{2} \left(1 + \frac{2(x-y) + 4}{2\sqrt{(x-y)^2 + 4x}} \right).$$

Resubstitution of x and y leads to

$$[\text{Slp1}]_{ss} \approx \frac{\frac{k_{syn}}{k_{deg}}}{\frac{[\text{inh}]_T}{K_D} + 1}$$

for $k_{syn} \ll k_{deg}([\text{inh}]_T + K_D)$.

To approximate the steady state for $k_{syn} \gg k_{deg}([\text{inh}]_T + K_D)$ we perform the Taylor series expansion of $y \mapsto f(x_0,y)$ around y_0

$$T_{f(x_0,y_0)}(y) = \frac{x_0 - y_0 + \sqrt{(x_0 - y_0)^2 + 4x_0}}{2} + \frac{1}{2} \left(\frac{y_0 - x_0}{\sqrt{(x_0 - y_0)^2 + 4x_0}} - 1 \right) y + O(y^2)$$

For $x_0 \gg y_0$

$$\lim_{x_0 \rightarrow \infty} \frac{x_0 - y_0 + \sqrt{(x_0 - y_0)^2 + 4x_0}}{2x_0} = 1$$

and

$$\lim_{x_0 \rightarrow \infty} \frac{1}{2} \left(\frac{y_0 - x_0}{\sqrt{(x_0 - y_0)^2 + 4x_0}} - 1 \right) = -1.$$

Therefore for $x_0 \gg y_0$

$$T_{f(x_0,y_0)}(y) \approx x_0 - y.$$

Resubstitution of x and y leads to

$$[\text{Slp1}]_{ss} \approx \frac{k_{syn}}{k_{deg}} - ([\text{inh}]_T + K_D)$$

for $k_{syn} \gg k_{deg}([\text{inh}]_T + K_D)$.

E. Derivation of parameter boundaries for parameter estimation in Chapter 5

Boundaries for k_{deg}

The boundaries for k_{deg} are calculated by assuming a minimum and maximum half-life of Slp1 of 7 min and 40 min, respectively, which leads to

$$k_{deg,max} = \frac{\log(2)}{7 \text{ min}} = 0.099 \text{ min}^{-1},$$

$$k_{deg,min} = \frac{\log(2)}{40 \text{ min}} = 0.0173 \text{ min}^{-1}.$$

Boundaries for k_{syn}

The boundaries for k_{syn} can be assessed through the observations that Slp1 accumulation is similar in cells with and without an active SAC and 20 nM Slp1 accumulate in 120 min. For k_{syn} upper and lower bounds for the mean are calculated because also the measurement of 20 nM Slp1 accumulating in 120 min corresponds to the mean of a population. The lower bound for the mean of k_{syn} is calculated from the extreme case of no degradation ($k_{deg} = 0$) of [Slp1]. Then 20 nM of Slp1 would be synthesized in 120 min yielding

$$k_{syn,mean,min} = \frac{20 \text{ nM}}{120 \text{ min}} = 0.17 \text{ nM min}^{-1}$$

Without SAC activity, no MCC is formed and the dynamics of Slp1 in our model reduce to

$$\frac{d[\text{Slp1}]}{dt} = k_{syn} - k_{deg}[\text{Slp1}]. \quad (\text{E.1})$$

In this model of Slp1 accumulation all [Slp1] is subject to degradation. Assuming a maximum degradation rate yields an upper bound for the mean of k_{syn} . This ODE (E.1) can be solved to

$$[\text{Slp1}](t) = \frac{k_{syn}}{k_{deg}}(1 - \exp(-t k_{deg})). \quad (\text{E.2})$$

Transforming (E.2) and inserting the maximum degradation rate yields

$$k_{syn,mean,max} = \frac{20 \text{ nM } k_{deg,mean,max}}{1 - \exp(-120 \text{ min } k_{deg,mean,max})}.$$

Since we assume a log-normal density for k_{syn} the bounds for μ and σ are derived from the bounds for the mean. The mean of a log-normal distribution is given by

$$mean = \exp\left(\mu + \frac{\sigma^2}{2}\right), \quad (\text{E.3})$$

and the standard deviation by

$$std = \sqrt{(\exp(\sigma^2) - 1) \exp(2\mu + \sigma^2)}$$

The CV is given by

$$CV = \frac{std}{mean} = \sqrt{\exp(\sigma^2) - 1}. \quad (E.4)$$

Therefore the bounds for σ_{syn} can be calculated considering a CV between 0.05 and 0.5 which gives

$$\begin{aligned} \sigma_{syn,min} &= \sqrt{\log(0.05^2 + 1)} = 0.4724, \\ \sigma_{syn,max} &= \sqrt{\log(0.5^2 + 1)} = 0.05. \end{aligned}$$

From (E.3) then the bounds for μ_{syn} can be calculated which gives

$$\begin{aligned} \mu_{syn,min} &= \log(k_{syn,mean,min}) - \frac{\sigma_{syn,min}^2}{2} = -1.7930, \\ \mu_{syn,max} &= \log(k_{syn,mean,max}) - \frac{\sigma_{syn,max}^2}{2} = 0.5717. \end{aligned}$$

Boundaries for $[inh]_T$

To calculate the bounds for the total concentration of the effective inhibitor we use the information that its CV is between 0.05 and 0.5 and the mean between 1 nM and 50 nM and a log-normal distribution. From (E.4) we can again compute toe bounds for $\sigma_{[inh]_T}$, which yields

$$\begin{aligned} \sigma_{[inh]_T,max} &= \sqrt{\log(0.5^2 + 1)} = 0.05, \\ \sigma_{[inh]_T,min} &= \sqrt{\log(0.05^2 + 1)} = 0.4724. \end{aligned}$$

And from the mean inhibitor between 1 nM and 50 nM from this and (E.3) the bounds for μ_{inh} become

$$\begin{aligned} \mu_{inh,max} &= \log(50) - \frac{\sigma_{[inh]_T,min}^2}{2} = 3.9108, \\ \mu_{inh,min} &= \log(1) - \frac{\sigma_{[inh]_T,max}^2}{2} = -0.1116. \end{aligned}$$

Bibliography

- Altman, D. G. & Bland, J. M. (1998). Time to event (survival) data. *BMJ*, *317*(7156), 468–469.
- Altschuler, S. J. & Wu, L. F. (2010). Cellular heterogeneity: do differences make a difference? *Cell*, *141*(4), 559–563.
- Ames, W. F. (2014). *Numerical methods for partial differential equations*. Academic press.
- Angeli, D., Ferrell, J. E., & Sontag, E. D. (2004). Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feedback systems. *Proc Natl Acad Sci U S A*, *101*(7), 1822–1827.
- Angerer, P., Haghverdi, L., Büttner, M., Theis, F. J., Marr, C., & Buettner, F. (2015). destiny: diffusion maps for large-scale single-cell data in R. *Bioinformatics*.
- Aravamudhan, P., Goldfarb, A. A., & Joglekar, A. P. (2015). The kinetochore encodes a mechanical switch to disrupt spindle assembly checkpoint signalling. *Nat Cell Biol*, *17*(7), 868–879.
- Bäck, T. (1996). *Evolutionary algorithms in theory and practice: evolution strategies, evolutionary programming, genetic algorithms*. New York and Oxford: Oxford University Press.
- Balaban, N. Q., Merrin, J., Chait, R., Kowalik, L., & Leibler, S. (2004). Bacterial persistence as a phenotypic switch. *Science*, *305*(5690), 1622–1625.
- Balázsi, G., van Oudenaarden, A., & Collins, J. J. (2011). Cellular decision making and biological noise: from microbes to mammals. *Cell*, *144*(6), 910–925.
- Balsa-Canto, E., Peifer, M., Banga, J. R., Timmer, J., & Fleck, C. (2008). Hybrid optimization method with general switching strategy for parameter estimation. *BMC Syst. Biol*, *2*(26).
- Banga, J. R. (2008). Optimization in computational systems biology. *BMC Syst. Biol*, *2*(47).
- Banks, H., Hu, S., & Thompson, W. C. (Eds.) (2014). *Modeling and inverse problems in the presence of uncertainty*. Monographs and textbooks in pure and applied mathematics. Boca Raton: Chapman & Hall/CRC.
- Birtwistle, M. R., Rauch, J., Kiyatkin, A., Aksamitiene, E., Dobrzyski, M., Hoek, J. B., Kolch, W., Ogunnaike, B. A., & Kholodenko, B. N. (2012). Emergence of bimodal cell population responses from the interplay between analog single-cell signaling and protein expression noise. *BMC Syst Biol*, *6*, 109.
- Boyer, T. C., Hanson, T., & Singer, R. S. (2013). Estimation of low quantity genes: a hierarchical model for analyzing censored quantitative real-time pcr data. *PLoS One*, *8*(5), e64900.
- Brito, D. A. & Rieder, C. L. (2006). Mitotic checkpoint slippage in humans occurs via cyclin b destruction in the presence of an active checkpoint. *Curr Biol*, *16*, 1194–1200.

- Buchler, N. E. & Louis, M. (2008). Molecular titration and ultrasensitivity in regulatory networks. *J Mol Biol*, 384(5), 1106–1119.
- Buck, V., Ng, S. S., Ruiz-Garcia, A. B., Papadopoulou, K., Bhatti, S., Samuel, J. M., Anderson, M., Millar, J. B. A., & McInerney, C. J. (2004). Fkh2p and Sep1p regulate mitotic gene transcription in fission yeast. *J Cell Sci*, 117(Pt 23), 5623–5632.
- Buettner, F., Moignard, V., Göttgens, B., & Theis, F. J. (2014). Probabilistic PCA of censored data: accounting for uncertainties in the visualization of high-throughput single-cell qPCR data. *Bioinformatics*, 30(13), 1867–1875.
- Buettner, F., Natarajan, K. N., Casale, F. P., Proserpio, V., Scialdone, A., Theis, F. J., Teichmann, S. A., Marioni, J. C., & Stegle, O. (2015). Computational analysis of cell-to-cell heterogeneity in single-cell rna-sequencing data reveals hidden subpopulations of cells. *Nat Biotechnol*, 33(2), 155–160.
- Burkardt, J. (2014). The truncated normal distribution. Tech. rep., Florida State University (FSU), Florida, USA.
- Burton, J. L. & Solomon, M. J. (2007). Mad3p, a pseudosubstrate inhibitor of APCCdc20 in the spindle assembly checkpoint. *Genes Dev*, 21(6), 655–667.
- Cadez, I. V., Smyth, P., McLachlan, G. J., & McLaren, C. E. (2002). Maximum likelihood estimation of mixture densities for binned and truncated multivariate data. *Machine Learning*, 47(1), 7–34.
- Chao, W. C. H., Kulkarni, K., Zhang, Z., Kong, E. H., & Barford, D. (2012). Structure of the mitotic checkpoint complex. *Nature*, 484(7393), 208–213.
- Chen, M.-H. & Shao, Q.-M. (1999). Monte Carlo estimation of Bayesian credible and HPD intervals. *J Comput Graphical Statist*, 8(1), 69–92.
- Ciliberto, A. & Shah, J. V. (2009). A quantitative systems view of the spindle assembly checkpoint. *The EMBO journal*, 28, 2162–2173.
- Cleveland, D. W., Mao, Y., & Sullivan, K. F. (2003). Centromeres and kinetochores: from epigenetics to mitotic checkpoint signaling. *Cell*, 112(4), 407–421.
- Colman-Lerner, A., Gordon, A., Serra, E., Chin, T., Resnekov, O., Endy, D., Pesce, C. G., & Brent, R. (2005). Regulated cell-to-cell variation in a cell-fate decision system. *Nature*, 437(7059), 699–706.
- Cook, J. D. (2008). Numerical computation of stochastic inequality probabilities. Tech. rep., UT MD Anderson Cancer Center Department of Biostatistics.
- Correia, J. J. & Detrich III, H. W. (2009). *Biophysical Tools for Biologists: In Vivo Techniques*, vol. 89. Academic Press.
- Coutu, D. L. & Schroeder, T. (2013). Probing cellular processes by long-term live imaging—historic problems and current solutions. *J Cell Sci*, 126, 3805–3815.
- Crane, M. M., Clark, I. B. N., Bakker, E., Smith, S., & Swain, P. S. (2014). A microfluidic system for studying ageing and dynamic single-cell responses in budding yeast. *PLoS One*, 9(6), e100042.
- Dai, H. & Wang, H. (2016). *Analysis for Time-to-Event Data under Censoring and Truncation*. Elsevier Science & Technology.

- Dao Duc, K. & Holcman, D. (2012). Using default constraints of the spindle assembly checkpoint to estimate the associated chemical rates. *BMC Biophys*, 5, 1.
- De Antoni, A., Pearson, C. G., Cimini, D., Canman, J. C., Sala, V., Nezi, L., Mapelli, M., Sironi, L., Faretta, M., Salmon, E. D., & Musacchio, A. (2005). The Mad1/Mad2 complex as a template for Mad2 activation in the spindle assembly checkpoint. *Curr Biol*, 15(3), 214–225.
- Dey, G., Gupta, G. D., Ramalingam, B., Sathe, M., Mayor, S., & Thattai, M. (2014). Exploiting cell-to-cell variability to detect cellular perturbations. *PLoS one*, 9, e90540.
- DiCiccio, T. J. & Efron, B. (1996). Bootstrap confidence intervals. *Statist Sci*, 11(3), 189–228.
- Dobrzyński, M., Nguyen, L. K., Birtwistle, M. R., von Kriegsheim, A., Blanco Fernandez, A., Cheong, A., Kolch, W., & Kholodenko, B. N. (2014). Nonlinear signalling networks and cell-to-cell variability transform external signals into broadly distributed or bimodal responses. *J R Soc Interface*, 11(98), 20140383.
- Domijan, M. & Pécou, E. (2012). The interaction graph structure of mass-action reaction networks. *J Math Biol*, 65(2), 375.
- Doncic, A., Ben-Jacob, E., & Barkai, N. (2005). Evaluating putative mechanisms of the mitotic spindle checkpoint. *Proc Natl Acad Sci U S A*, 102(18), 6332–6337.
- Doncic, A., Ben-Jacob, E., & Barkai, N. (2006). Noise resistance in the spindle assembly checkpoint. *Mol Syst Biol*, 2, 2006.0027.
- Doncic, A., Ben-Jacob, E., Einav, S., & Barkai, N. (2009). Reverse engineering of the spindle assembly checkpoint. *PLoS one*, 4, e6495.
- Duffy, K. R., Wellard, C. J., Markham, J. F., Zhou, J. H. S., Holmberg, R., Hawkins, E. D., Hasbold, J., Dowling, M. R., & Hodgkin, P. D. (2012). Activation-induced B cell fates are selected by intracellular stochastic competition. *Science*, 335(6066), 338–341.
- Eldar, A. & Elowitz, M. B. (2010). Functional roles for noise in genetic circuits. *Nature*, 467(9), 1–7.
- Elf, J. & Ehrenborg, M. (2003). Fast evaluation of fluctuations in biochemical networks with the linear noise approximation. *Genome Res.*, 13, 2475–2484.
- Elowitz, M. B., Levine, A. J., Siggia, E. D., & Swain, P. S. (2002). Stochastic gene expression in a single cell. *Science*, 297(5584), 1183–1186.
- Engblom, S. (2006). Computing the moments of high dimensional solutions of the master equation. *Appl Math Comp*, 180, 498–515.
- Everitt, B. S. (1996). An introduction to finite mixture distributions. *Stat Methods Med Res*, 5(2), 107–127.
- Fang, G. (2002). Checkpoint protein BubR1 acts synergistically with Mad2 to inhibit anaphase-promoting complex. *Mol Biol Cell*, 13(3), 755–766.
- Fang, G., Yu, H., & Kirschner, M. W. (1998). Direct binding of CDC20 protein family members activates the anaphase-promoting complex in mitosis and G1. *Mol Cell*, 2(2), 163–171.
- Ferrell, J. E. (2002). Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Curr Opin Cell Biol*, 14, 140–148.

- Ferrell, J. E., Jr & Ha, S. H. (2014a). Ultrasensitivity part i: Michaelian responses and zero-order ultrasensitivity. *Trends Biochem. Sci.*, 39(10), 496–503.
- Ferrell, J. E., Jr & Ha, S. H. (2014b). Ultrasensitivity part ii: multisite phosphorylation, stoichiometric inhibitors, and positive feedback. *Trends Biochem Sci*, 39(11), 556–569.
- Ferrell, J. E., Jr & Ha, S. H. (2014c). Ultrasensitivity part iii: cascades, bistable switches, and oscillators. *Trends Biochem Sci*, 39(12), 612–618.
- Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 222, 309–368.
- Foe, I. T., Foster, S. A., Cheung, S. K., DeLuca, S. Z., Morgan, D. O., & Toczyski, D. P. (2011). Ubiquitination of Cdc20 by the APC occurs through an intramolecular mechanism. *Curr Biol*, 21(22), 1870–1877.
- Forsburg, S. L. (2005). The yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*: models for cell biology research. *Gravit Space Biol Bull*, 18(2), 3–9.
- Forsburg, S. L. & Rhind, N. (2006). Basic methods for fission yeast. *Yeast*, 23(3), 173–183.
- Foster, S. A. & Morgan, D. O. (2012). The APC/C subunit Mnd2/Apc15 promotes Cdc20 autoubiquitination and spindle assembly checkpoint inactivation. *Mol Cell*, 47(6), 921–932.
- Fraser, D. & Kaern, M. (2009). A chance at survival: gene expression noise and phenotypic diversification strategies. *Mol Microbiol*, 71, 1333–1340.
- Fröhlich, F., Theis, F. J., Rädler, J. O., & Hasenauer, J. (2016). Parameter estimation for dynamical systems with discrete events and logical operations. *Bioinformatics*.
- Furusawa, C., Suzuki, T., Kashiwagi, A., Yomo, T., & Kaneko, K. (2005). Ubiquity of log-normal distributions in intra-cellular reaction dynamics. *Biophysics*, 1, 25–32.
- Gascoigne, K. E. & Taylor, S. S. (2009). How do anti-mitotic drugs kill cancer cells? *J Cell Sci*, 122, 2579–2585.
- Ge, S., Skaar, J. R., & Pagano, M. (2009). APC/C- and Mad2-mediated degradation of Cdc20 during spindle checkpoint activation. *Cell Cycle*, 8(1), 167–171.
- Geissen, E.-M., Hasenauer, J., Heinrich, S., Hauf, S., Theis, F. J., & Radde, N. E. (2016). MEMO: multi-experiment mixture model analysis of censored data. *Bioinformatics*, 32, 2464–2472.
- Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *J Phys Chem*, 81(25), 2340–2361.
- Gillespie, D. T. (2000). The chemical Langevin equation. *J Chem Phys*, 113(1), 297–306.
- Girolami, M. & Calderhead, B. (2011). Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J R Statist Soc B*, 73(2), 123–214.
- Goldbeter, A. & Koshland, D., Jr (1981). An amplified sensitivity arising from covalent modification in biological systems. *Proc Natl Acad Sci U S A*, 78(11), 6840–6844.
- Gouzé, J.-L. (1998). Positive and negative circuits in dynamical systems. *J Biol Syst*, 06(01), 11–15.
- Grima, R. (2010). An effective rate equation approach to reaction kinetics in small volumes:

- Theory and application to biochemical reactions in nonequilibrium steady-state conditions. *J Chem Phys*, 133(035101).
- Hadamard, J. (1902). Sur les problèmes aux dérivées partielles et leur signification physique. In *Princeton University Bulletin*. 49–52.
- Hasenauer, J., Hasenauer, C., Hucho, T., & Theis, F. J. (2014a). ODE constrained mixture modelling: A method for unraveling subpopulation structures and dynamics. *PLoS Comput Biol*, 10(7), e1003686.
- Hasenauer, J., Wolf, V., Kazeroonian, A., & Theis, F. J. (2014b). Method of conditional moments (MCM) for the chemical master equation. *J Math Biol*, 69(3), 687–735.
- He, X., Jones, M. H., Winey, M., & Sazer, S. (1998). Mph1, a member of the Mps1-like family of dual specificity protein kinases, is required for the spindle checkpoint in *S. pombe*. *J Cell Sci*, 111(12), 1635–1647.
- Heinrich, S. (2013). *Qualitative and quantitative analysis of spindle assembly checkpoint signaling in Schizosaccharomyces pombe*. Ph.D. thesis, Eberhard Karls Universität Tübingen.
- Heinrich, S., Geissen, E., Kamenz, J., Trautmann, S., Widmer, C., Drewe, P., Knop, M., Radde, N., Hasenauer, J., & Hauf, S. (2013). Determinants for robustness in spindle assembly checkpoint signalling. *Nat Cell Biol*, 15(11), 1328–1339.
- Helton, J. C. & Davis, F. J. (2003). Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliability Engineering & System Safety*, 81(1), 23–69.
- Herzog, F., Primorac, I., Dube, P., Lenart, P., Sander, B., Mechtler, K., Stark, H., & Peters, J.-M. (2009). Structure of the anaphase-promoting complex/cyclosome interacting with a mitotic checkpoint complex. *Science*, 323(5920), 1477–1481.
- Higgins, J. (1965). Dynamics and control in cellular reactions. *Control of energy metabolism*, 13–46.
- Hill, A. V. (1910). The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *J Physiol (Lond)*, 40, 4–7.
- Hindmarsh, A. C., Brown, P. N., Grant, K. E., Lee, S. L., Serban, R., Shumaker, D. E., & Woodward, C. S. (2005). SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers. *ACM Trans Math Softw*, 31(3), 363–396.
- Hiruma, Y., Sacristan, C., Pachis, S. T., Adamopoulos, A., Kuijt, T., Ubbink, M., von Castel-mur, E., Perrakis, A., & Kops, G. J. P. L. (2015). Cell division cycle. competition between mps1 and microtubules at kinetochores regulates spindle checkpoint signaling. *Science*, 348(6240), 1264–1267.
- Hoffman, C. S., Wood, V., & Fantes, P. A. (2015). An ancient yeast for young geneticists: a primer on the schizosaccharomyces pombe model system. *Genetics*, 201(2), 403–423.
- Hoppe, P. S., Coutu, D. L., & Schroeder, T. (2014). Single-cell technologies sharpen up mammalian stem cell research. *Nat Cell Biol*, 16(10), 919–927.
- Hoyt, M. A., Totis, L., & Roberts, B. T. (1991). *S. cerevisiae* genes required for cell cycle arrest in response to loss of microtubule function. *Cell*, 66(3), 507–517.
- Huang, H.-C., Mitchison, T. J., & Shi, J. (2010). Stochastic competition between mechanis-

- tically independent slippage and death pathways determines cell fate during mitotic arrest. *PLoS One*, 5(12), e15724.
- Huang, S. (2009). Non-genetic heterogeneity of cells in development: more than just noise. *Development*, 136, 3853–3862.
- Hug, S., Raue, A., Hasenauer, J., Bachmann, J., Klingmüller, U., Timmer, J., & Theis, F. J. (2013). High-dimensional Bayesian parameter estimation: Case study for a model of JAK2/STAT5 signaling. *Math Biosci*, 246(2), 293–304.
- Huh, D. & Paulsson, J. (2011). Non-genetic heterogeneity from stochastic partitioning at cell division. *Nat Gen*, 43(2), 95–102.
- Hunt, P. A. & Hassold, T. J. (2008). Human female meiosis: what makes a good egg go bad? *Trends Genet*, 24(2), 86–93.
- Hwang, L. H., Lau, L. F., Smith, D. L., Mistrot, C. A., Hardwick, K. G., Hwang, E. S., Amon, A., & Murray, A. W. (1998). Budding yeast Cdc20: a target of the spindle checkpoint. *Science*, 279, 1041–1044.
- Ibrahim, B., Diekmann, S., Schmitt, E., & Dittrich, P. (2008a). In-silico modeling of the mitotic spindle assembly checkpoint. *PLoS One*, 3, e1555.
- Ibrahim, B., Dittrich, P., Diekmann, S., & Schmitt, E. (2008b). Mad2 binding is not sufficient for complete Cdc20 sequestering in mitotic transition control (an in silico study). *Biophys Chem*, 134(1), 93–100.
- Ibrahim, B., Schmitt, E., Dittrich, P., & Diekmann, S. (2009). In silico study of kinetochore control, amplification, and inhibition effects in mcc assembly. *Bio Systems*, 95, 35–50.
- Ito, K. & Xiong, K. (2000). Gaussian filters for nonlinear filtering problems. *IEEE Trans Autom Contro*, 45(5), 910–927.
- Iyer-Biswas, S. & Zilman, A. (2015). First passage processes in cellular biology. *arXiv preprint arXiv:1503.00291*.
- Jeschke, M., Baumgärtner, S., & Legewie, S. (2013). Determinants of cell-to-cell variability in protein kinase signaling. *PLoS Comput Biol*, 9(12), e1003357.
- Ji, Z., Gao, H., & Yu, H. (2015). Kinetochore attachment sensed by competitive Mps1 and microtubule binding to Ndc80c. *Science*, 348(6240), 1260–1264.
- Joglekar, A. P. (2016). A cell biological perspective on past, present and future investigations of the spindle assembly checkpoint. *Biology*, 5.
- Joglekar, A. P. & Aravamudhan, P. (2016). How the kinetochore switches off the spindle assembly checkpoint. *Cell Cycle*, 15(1), 7–8.
- Johnsson, K., Wallin, J., & Fontes, M. (2016). Bayesflow: latent modeling of flow cytometry cell populations. *BMC Bioinf*, 17(1), 25.
- Joshi, M., Seidel-Morgenstern, A., & Kremling, A. (2006). Exploiting the bootstrap method for quantifying parameter confidence intervals in dynamical systems. *Metabolic Eng*, 8, 447–455.
- Julier, S., Uhlmann, J., & Durrant-Whyte, H. F. (2000). A new method for the nonlinear transformation of means and covariances in filters and estimators. *IEEE Trans Autom Control*, 45(3), 477–482.

- Julier, S. & Uhlmann, J. K. (1996). A general method for approximating nonlinear transformations of probability distributions. Technical report, Robotics Research Group, Department of Engineering Science, University of Oxford.
- Kaern, M., Elston, T. C., Blake, W. J., & Collins, J. J. (2005). Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet*, 6, 451–464.
- Kaplan, E. L. & Meier, P. (1958). Nonparametric estimation from incomplete observations. *J Amer Statist Assn*, 53(282), 457–481.
- Kass, R. E. & Raftery, A. E. (1995). Bayes factors. *J Am Stat Assoc*, 90(430), 773–795.
- Kim, K. H. & Sauro, H. M. (2012). In search of noise-induced bimodality. *BMC Biol*, 10, 89.
- Kim, S. H., Lin, D. P., Matsumoto, S., Kitazono, A., & Matsumoto, T. (1998). Fission yeast slp1: an effector of the mad2-dependent spindle checkpoint. *Science*, 279(5353), 1045–1047.
- Kirkpatrick, S., Gelatt Jr, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, 220(4598), 671–680.
- Klein, J. P. & Moeschberger, M. L. (2003). *Survival Analysis Techniques for Censored and Truncated Data*. Springer, second ed.
- Klipp, E., Herwig, R., Kowald, A., Wierling, C., & Lehrach, H. (2005). *Systems biology in practice*. Wiley-VCH, Weinheim.
- Klipp, E. & Liebermeister, W. (2006). Mathematical modeling of intracellular signaling pathways. *BMC Neurosci*, 7 Suppl 1, S10.
- Kops, G. J., Weaver, B. A., & Cleveland, D. W. (2005). On the road to cancer: aneuploidy and the mitotic checkpoint. *Nat Rev Cancer*, 5(10), 773–785.
- Koshland, D., Jr, Goldbeter, A., & Stock, J. (1982). Amplification and adaptation in regulatory and sensory systems. *Science*, 217(4556), 220–225.
- Koshland, D., Jr, Némethy, G., & Filmer, D. (1966). Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry*, 5(1), 365–385.
- Kozin, F. (1961). On the probability densities of the output of some random systems. *ASME J Appl Mech*, 28(2), 161–164.
- Kramer, E. R., Gieffers, C., Hölzl, G., Hengstschläger, M., & Peters, J. M. (1998). Activation of the human anaphase-promoting complex by proteins of the CDC20/Fizzy family. *Curr Biol*, 8(22), 1207–1210.
- Kussell, E. & Leibler, S. (2005). Phenotypic diversity, population growth, and information in fluctuating environments. *Science*, 309, 2075–2078.
- Ladde, G. S. & Lakshmikantham, V. (1980). *Random differential inequalities*. Elsevier.
- Lee, C. H., Kim, K. H., & Kim, P. (2009). A moment closure method for stochastic reaction networks. *J Chem Phys*, 130(13), 134107.
- Lee, G. & Scott, C. (2012). Em algorithms for multivariate gaussian mixture models with truncated and censored data. *Comput Stat Data Anal*, 56(9), 2816–2829.
- Lefebvre, T., Bruyninckx, H., & Schuller, J. D. (2002). Comment on "a new method for the nonlinear transformation of means and covariances in filters and estimators" [with authors' reply]. *IEEE Trans Autom Control*, 47(8), 1406–1409.

- Li, B. & You, L. (2013). Predictive power of cell-to-cell variability. *Quant Biol*, 1(2), 131–139.
- Li, R. & Murray, A. W. (1991). Feedback control of mitosis in budding yeast. *Cell*, 66(3), 519–531.
- Li, X., Nicklas, R. B., *et al.* (1995). Mitotic forces control a cell-cycle checkpoint. *Nature*, 373(6515), 630–632.
- Lindner, P. (1893). Schizosaccharomyces pombe n. sp. neuer Gärungserreger. *Wochenschrift für Brauerei*, 10, 1298–1300.
- Lo, K., Brinkman, R. R., & Gottardo, R. (2008). Automated gating of flow cytometry data via robust model-based clustering. *Cytometry A*, 73, 321–332.
- London, N. & Biggins, S. (2014). Signalling dynamics in the spindle checkpoint response. *Nat Rev Mol Cell Biol*, 15(11), 736–747.
- Loo, L.-H., Lin, H.-J., Singh, D. K., Lyons, K. M., Altschuler, S. J., & Wu, L. F. (2009). Heterogeneity in the physiological states and pharmacological responses of differentiating 3T3-L1 preadipocytes. *J Cell Biol*, 187(3), 375–384.
- Ma, H. T. & Poon, R. Y. C. (2011). Orderly inactivation of the key checkpoint protein mitotic arrest deficient 2 (MAD2) during mitotic progression. *J Biol Chem*, 286(15), 13052–13059.
- Malureanu, L., Jeganathan, K. B., Jin, F., Baker, D. J., van Ree, J. H., Gullon, O., Chen, Z., Henley, J. R., & van Deursen, J. M. (2010). Cdc20 hypomorphic mice fail to counteract de novo synthesis of cyclin B1 in mitosis. *J Cell Biol*, 191(2), 313–329.
- Mangan, S. & Alon, U. (2003). Structure and function of the feed-forward loop network motif. *Proc Natl Acad Sci U S A*, 100, 11980–11985.
- Mansfeld, J., Collin, P., Collins, M. O., Choudhary, J. S., & Pines, J. (2011). APC15 drives the turnover of MCC-CDC20 to make the spindle assembly checkpoint responsive to kinetochore attachment. *Nat Cell Biol*, 13(10), 1234–1243.
- McCall, M. N., McMurray, H. R., Land, H., & Almudevar, A. (2014). On non-detects in qPCR data. *Bioinformatics*, 30(16), 2310–2316.
- McLachlan, G. & Peel, D. (2000). *Finite mixture models*. Wiley Series in Probability and Statistics: Applied Probability and Statistics. Wiley-Interscience, New York.
- McLachlan, G. J. & Jones, P. N. (1988). Fitting mixture models to grouped and truncated data via the EM algorithm. *Biometrics*, 44(2), 571–578.
- McNaught, A. D., Wilkinson, A., *et al.* (1997). *Compendium of chemical terminology. IUPAC recommendations*. Oxford (United Kingdom) Blackwell.
- Meeker, W. Q. & Escobar, L. A. (1995). Teaching about approximate confidence regions based on maximum likelihood estimation. *Am. Stat.*, 49(1), 48–53.
- Millard, S. P., Neerchal, N. K., & Dixon, P. (2012). *Environmental Statistics with R*. CRC.
- Mistry, H. B., MacCallum, D. E., Jackson, R. C., Chaplain, M. A. J., & Davidson, F. A. (2008). Modeling the temporal evolution of the spindle assembly checkpoint and role of Aurora B kinase. *Proc Natl Acad Sci U S A*, 105(51), 20215–20220.
- Moles, C. G., Mendes, P., & Banga, J. R. (2003). Parameter estimation in biochemical pathways: A comparison of global optimization methods. *Genome Res*, 13, 2467–2474.

- Morrow, C. J., Tighe, A., Johnson, V. L., Scott, M. I. F., Ditchfield, C., & Taylor, S. S. (2005). Bub1 and aurora B cooperate to maintain BubR1-mediated inhibition of APC/CCdc20. *J Cell Sci*, *118*, 3639–3652.
- Murphy, S. A. & van der Vaart, A. W. (2000). On profile likelihood. *J Am Stat Assoc*, *95*(450), 449–485.
- Musacchio, A. (2015). The molecular biology of spindle assembly checkpoint signaling dynamics. *Curr Biol*, *25*(20), R1002–R1018.
- Nachman, I., Regev, A., & Ramanathan, S. (2007). Dissecting timing variability in yeast meiosis. *Cell*, *131*(3), 544–556.
- Nanita, S. C. (2013). Quantitative mass spectrometry independence from matrix effects and detector saturation achieved by flow injection analysis with real-time infinite dilution. *Anal Chem*, *85*(24), 11866–11875.
- Nilsson, J., Yekezare, M., Minshull, J., & Pines, J. (2008). The APC/C maintains the spindle assembly checkpoint by targeting Cdc20 for destruction. *Nat Cell Biol*, *10*(12), 1411–1420.
- Nørgaard, M., Poulsen, N. K., & Ravn, O. (2000). New developments in state estimation for nonlinear systems. *Automatica*, *36*(11), 1627 – 1638.
- Novak, B., Pataki, Z., Ciliberto, A., & Tyson, J. J. (2001). Mathematical model of the cell division cycle of fission yeast. *Chaos*, *11*(1), 277–286.
- Ochab-Marcinek, A. & Tabaka, M. (2010). Bimodal gene expression in noncooperative regulatory systems. *Proc Natl Acad Sci U S A*, *107*, 22096–22101.
- Owen, A. B. (1992). A central limit theorem for Latin hypercube sampling. *Journal of the Royal Statistical Society. Series B (Methodological)*, 541–551.
- Pan, J. & Chen, R.-H. (2004). Spindle checkpoint regulates γ Cdc20p stability in *Saccharomyces cerevisiae*. *Genes Dev*.
- Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Philos Trans R Soc Lond. A*, 71–110.
- Pedraza, J. M. & Paulsson, J. (2007). Random timing in signaling cascades. *Mol Syst Biol*, *3*, 81.
- Pelkmans, L. (2012). Using cell-to-cell variability—A new era in molecular biology. *Science*, *336*(6080), 425—426.
- Perkins, T. J. & Swain, P. S. (2009). Strategies for cellular decision-making. *Mol Syst Biol*, *5*, 326.
- Pidoux, A. L. & Allshire, R. C. (2004). Kinetochores and heterochromatin domains of the fission yeast centromere. *Chromosome Res*, *12*(6), 521–534.
- Primorac, I. & Musacchio, A. (2013). Panta rhei: the APC/C at steady state. *J Cell Biol*, *201*(2), 177–189.
- Pyne, S., Hu, X., Wang, K., Rossin, E., Lin, T., Maier, L., Baecher-Allan, C., McLachlan, G., Tamayo, P., Hafler, D., De Jager, P., & Mesirov, J. (2009). Automated high-dimensional flow cytometric data analysis. *Proc Natl Acad Sci U S A*, *106*(21), 8519–8124.
- Pyne, S., Lee, S. X., Wang, K., Irish, J., Tamayo, P., Nazaire, M.-D., Duong, T., Ng, S.-K., Hafler, D., Levy, R., Nolan, G. P., Mesirov, J., & McLachlan, G. J. (2014). Joint modeling

- and registration of cell populations in cohorts of high-dimensional flow cytometric data. *PLoS One*, 9(7), e100334.
- Radde, N. (2010). Fixed point characterization of biological networks with complex graph topology. *Bioinformatics*, 26(22), 2874–2880.
- Ramaswamy, R., González-Segredo, N., Sbalzarini, I., & Grima, R. (2012). Discreteness-induced concentration inversion in mesoscopic chemical systems. *Nat Comm*, 3(779).
- Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., & Timmer, J. (2009). Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(25), 1923–1929.
- Raue, A., Kreutz, C., Theis, F. J., & Timmer, J. (2013a). Joining forces of Bayesian and frequentist methodology: A study for inference in the presence of non-identifiability. *Phil Trans Royal Soc A*, 371(1984).
- Raue, A., Schilling, M., Bachmann, J., Matteson, A., Schelke, M., Kaschek, D., Hug, S., Kreutz, C., Harms, B. D., Theis, F. J., Klingmüller, U., & Timmer, J. (2013b). Lessons learned from quantitative dynamical modeling in systems biology. *PLoS ONE*, 8(9), e74335.
- Reddy, S. K., Rape, M., Margansky, W. A., & Kirschner, M. W. (2007). Ubiquitination by the anaphase-promoting complex drives spindle checkpoint inactivation. *Nature*, 446(7138), 921–925.
- Redner, S. (2001). *A guide to first-passage processes*. Cambridge University Press, Cambridge.
- Rieder, C. L., Cole, R. W., Khodjakov, A., & Sluder, G. (1995). The checkpoint delaying anaphase in response to chromosome monoorientation is mediated by an inhibitory signal produced by unattached kinetochores. *J Cell Biol*, 130(4), 941–948.
- Rinott, R., Jaimovich, A., & Friedman, N. (2011). Exploring transcription regulation through cell-to-cell variability. *Proc Natl Acad Sci U S A*, 108(15), 6329–6334.
- Risken, H. (1996). *The Fokker-Planck equation: Methods of solution and applications*. Springer Berlin / Heidelberg, 2nd ed.
- Santaguida, S. & Amon, A. (2015). Short-and long-term effects of chromosome mis-segregation and aneuploidy. *Nat Rev Mol Cell Biol*, 16(8), 473–485.
- Särkkä, S. (2013). *Bayesian filtering and smoothing*, vol. 3. Cambridge University Press.
- Saurin, A. T., van der Waal, M. S., Medema, R. H., Lens, S. M. A., & Kops, G. J. P. L. (2011). Aurora B potentiates Mps1 activation to ensure rapid checkpoint establishment at the onset of mitosis. *Nat Commun*, 2, 316.
- Sauro, H. M. (2011). *Enzyme kinetics for Systems Biology*. Future Skill Software.
- Scholz, F. W. (2004). Maximum likelihood estimation. *Encyclopedia of Statistical Sciences*, John Wiley & Sons, Inc., chap. Maximum Likelihood Estimation.
- Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., & Selbach, M. (2011). Global quantification of mammalian gene expression control. *Nature*, 473(7347), 337–342.
- Sczaniecka, M., Feoktistova, A., May, K. M., Chen, J.-S., Blyth, J., Gould, K. L., & Hardwick,

- K. G. (2008). The spindle checkpoint functions of Mad3 and Mad2 depend on a Mad3 KEN box-mediated interaction with Cdc20-anaphase-promoting complex (APC/C). *J Biol Chem*, 283(34), 23039–23047.
- Sear, R. P. & Howard, M. (2006). Modeling dual pathways for the metazoan spindle assembly checkpoint. *Proc Natl Acad Sci U S A*, 103(45), 16758–16763.
- Serban, R. & Hindmarsh, A. C. (2005). CVODES: the sensitivity-enabled ODE solver in SUNDIALS. In *ASME 2005 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. American Society of Mechanical Engineers, 257–269.
- Simonetta, M., Manzoni, R., Mosca, R., Mapelli, M., Massimiliano, L., Vink, M., Novak, B., Musacchio, A., & Ciliberto, A. (2009). The influence of catalysis on Mad2 activation dynamics. *PLoS Biol*, 7(1), e10.
- Snijder, B. & Pelkmans, L. (2011). Origins of regulated cell-to-cell variability. *Nat Rev Mol Cell Biol*, 12(2), 119–125.
- Song, C., Phenix, H., Abedi, V., Scott, M., Ingalls, B. P., Kærn, M., & Perkins, T. J. (2010). Estimating the stochastic bifurcation structure of cellular networks. *PLoS Comput Biol*, 6(3), e1000699.
- Soong, T. T. (1973). *Random differential equations in science and engineering*. Elsevier.
- Stephens, M. (2000). Dealing with label switching in mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(4), 795–809.
- Sterba, S. K. (2009). Alternative model-based and design-based frameworks for inference from samples to populations: From polarization to integration. *Multivariate Behav Res*, 44, 711–740.
- Sudakin, V., Chan, G. K., & Yen, T. J. (2001). Checkpoint inhibition of the APC/C in HeLa cells is mediated by a complex of BUBR1, BUB3, CDC20, and MAD2. *J Cell Biol*, 154(5), 925–936.
- Sudakin, V., Ganoth, D., Dahan, A., Heller, H., Hershko, J., Luca, F. C., Ruderman, J. V., & Hershko, A. (1995). The cyclosome, a large complex containing cyclin-selective ubiquitin ligase activity, targets cyclins for destruction at the end of mitosis. *Mol Biol Cell*, 6, 185–197.
- Swain, P. S., Elowitz, M. B., & Siggia, E. D. (2002). Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc Natl Acad Sci U S A*, 99(20), 12795–12800.
- Taketa, K. & Pogell, B. M. (1965). Allosteric inhibition of rat liver fructose 1,6-diphosphatase by adenosine 5'-monophosphate. *J Biol Chem*, 240, 651–662.
- Thoma, C. R., Toso, A., Gutbrodt, K. L., Reggi, S. P., Frew, I. J., Schraml, P., Hergovich, A., Moch, H., Meraldi, P., & Krek, W. (2009). VHL loss causes spindle misorientation and chromosome instability. *Nat Cell Biol*, 11, 994–1001.
- Thomas, P., Popovic, N., & Grima, R. (2014). Phenotypic switching in gene regulatory networks. *Proc Natl Acad Sci U S A*, 111(19), 6994–6999.
- Thomas, R. (1981). On the relation between the logical structure of systems and their ability to generate multiple steady states or sustained oscillations. *Numerical Methods in the Study*

- of Critical Phenomena*, Berlin, Heidelberg: Springer Berlin Heidelberg. 180–193.
- Thomaseth, C., Kuritz, K., Allgwer, F., & Radde, N. (2017). The circuit-breaking algorithm for monotone systems. *Math Biosci*, 284, 80–91.
- Thornton, B. R. & Toczyski, D. P. (2003). Securin and B-cyclin/CDK are the only essential targets of the APC. *Nat Cell Biol*, 5(12), 1090–1094.
- Torres, E. M., Sokolsky, T., Tucker, C. M., Chan, L. Y., Boselli, M., Dunham, M. J., & Amon, A. (2007). Effects of aneuploidy on cellular physiology and cell division in haploid yeast. *Science*, 317(5840), 916–924.
- Tyson, J. J., Chen, K. C., & Novak, B. (2003). Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr Opin Cell Biol*, 15, 221–231.
- Tyson, J. J. & Novak, B. (2015). Models in biology: lessons from modeling regulation of the eukaryotic cell cycle. *BMC Biol*, 13, 46.
- Umesono, K., Toda, T., Hayashi, S., & Yanagida, M. (1983). Two cell division cycle genes NDA2 and NDA3 of the fission yeast *Schizosaccharomyces pombe* control microtubular organization and sensitivity to anti-mitotic benzimidazole compounds. *J Mol Biol*, 168(2), 271–284.
- Usoskin, D., Furlan, A., Islam, S., Abdo, H., Lönnerberg, P., Lou, D., Hjerling-Leffler, J., Haeggström, J., Kharchenko, O., Kharchenko, P. V., Linnarsson, S., & Ernfors, P. (2015). Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. *Nat Neurosci*, 18(1), 145–153.
- Uzunova, K., Dye, B. T., Schutz, H., Ladurner, R., Petzold, G., Toyoda, Y., Jarvis, M. A., Brown, N. G., Poser, I., Novatchkova, M., Mechtler, K., Hyman, A. A., Stark, H., Schulman, B. A., & Peters, J.-M. (2012). APC15 mediates CDC20 autoubiquitylation by APC/C(MCC) and disassembly of the mitotic checkpoint complex. *Nat Struct Mol Biol*, 19(11), 1116–1123.
- Valov, A. V. (2009). *First passage times: Integral equations, randomization and analytical approximations*. Ph.D. thesis, University of Toronto.
- van der Merwe, R. (2004). *Sigma-point Kalman filters for probabilistic inference in dynamic state-space models*. Ph.d. thesis, Oregon Health & Science University.
- van der Merwe, R. & Wan, E. (2003). Sigma-point kalman filters for probabilistic inference in dynamic state-space models. In *Proceedings of the Workshop on Advances in Machine Learning, Montreal, Canada (2003)*.
- van Kampen, N. G. (2007). *Stochastic processes in physics and chemistry*. Amsterdam: North-Holland, 3rd ed.
- Vaz, A. & Vicente, L. (2007). A particle swarm pattern search method for bound constrained global optimization. *J. Global Optim.*, 39(2), 197–219.
- Visscher, K., Brakenhoff, G., & Visser, T. (1994). Fluorescence saturation in confocal microscopy. *J Microsc*, 175(2), 162–165.
- Vleugel, M., Hoogendoorn, E., Snel, B., & Kops, G. J. P. L. (2012). Evolution and function of the mitotic checkpoint. *Dev Cell*, 23(2), 239–250.
- Voit, E. O., Martens, H. A., & Omholt, S. W. (2015). 150 years of the mass action law. *PLoS*

- Comput Biol*, 11(1), e1004012.
- Wang, H. & Huang, S. (2007). Mixture-model classification in DNA content analysis. *Cytometry A*, 71(9), 716–723.
- Waters, J. C. (2009). Accuracy and precision in quantitative fluorescence microscopy. *J Cell Biol*, 185, 1135–1148.
- Webster, M. D., Tatang, M. A., & McRae, G. J. (1996). Application of the probabilistic collocation method for an uncertainty analysis of a simple ocean model. Tech. rep., MIT.
- Weise, T. (2009). Global Optimization Algorithms: Theory and Application. ebook, Nature Inspired Computation and Applications Laboratory (NICAL), University of Science and Technology, China.
- Weiss, E. & Winey, M. (1996). The *Saccharomyces cerevisiae* spindle pole body duplication gene MPS1 is part of a mitotic checkpoint. *J Cell Biol*, 132(1-2), 111–123.
- Weiss, J. N. (1997). The Hill equation revisited: uses and misuses. *The FASEB Journal*, 11(11), 835–841.
- Weiß, A. Y., Middleton, R. H., & Huisinga, W. (2010). Quantifying uncertainty, variability and likelihood for ordinary differential equation models. *BMC Syst. Biol.*, 4(144).
- Whittaker, E. (1937). *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies*. Cambridge University Press.
- Wilkinson, D. J. (2007). Bayesian methods in bioinformatics and computational systems biology. *Briefings in Bioinf.*, 8(2), 109–116.
- Wilkinson, D. J. (2009). Stochastic modelling for quantitative description of heterogeneous biological systems. *Nat. Rev. Genet.*, 10(2), 122–133.
- Wolkenhauer, O. (2014). Why model? *Front Physiol*, 5, 21.
- Wolthuis, R., Clay-Farrace, L., van Zon, W., Yekezare, M., Koop, L., Ogink, J., Medema, R., & Pines, J. (2008). Cdc20 and Cks direct the spindle checkpoint-independent destruction of cyclin A. *Mol Cell*, 30(3), 290–302.
- Yamada, H. Y., Matsumoto, S., & Matsumoto, T. (2000). High dosage expression of a zinc finger protein, Grt1, suppresses a mutant of fission yeast slp1(+), a homolog of CDC20/p55CDC/Fizzy. *J Cell Sci*, 113 (Pt 22), 3989–3999.
- Yang, X. (2010). *Nature-inspired metaheuristic algorithms*. Luniver Press, 2nd ed.
- Yu, H. (2007). Cdc20: a WD40 activator for a cell cycle degradation machine. *Mol Cell*, 27(1), 3–16.
- Yurkovsky, E. & Nachman, I. (2013). Event timing at the single-cell level. *Brief Funct Genomics*, 12(2), 90–98.
- Zhang, Q., Bhattacharya, S., & Andersen, M. E. (2013). Ultrasensitive response motifs: basic amplifiers in molecular signalling networks. *Open Biol*, 3(4), 130031.
- Zieve, G. W., Turnbull, D., Mullins, J. M., & McIntosh, J. R. (1980). Production of large numbers of mitotic mammalian cells by use of the reversible microtubule inhibitor nocodazole. nocodazole accumulated mitotic cells. *Exp Cell Res*, 126, 397–405.