

Technische Universität München

Fachgebiet für Systembiotechnologie

**Analyse der Qualität von Messdaten
in der Systembiologie**

Jana Tillack

Vollständiger Abdruck der von der Fakultät für Maschinenwesen
der Technischen Universität München zur Erlangung des akademischen Grades
eines
Doktor-Ingenieurs
genehmigten Dissertation.

Vorsitzender: Prof. Dr.-Ing. Dirk Weuster-Botz

Prüfer der Dissertation: 1. Prof. Dr.-Ing. Andreas Kremling
2. Prof. Dr. rer. nat. Wolfgang Wiechert, RWTH Aachen

Die Dissertation wurde am 10.08.2017 bei der Technischen Universität München
eingereicht und durch die Fakultät für Maschinenwesen am 20.12.2017 angenommen.

Zusammenfassung

In systembiologischen Projekten werden experimentelle Daten erhoben, interpretiert und für weitere Analysen und Studien als Eingangsgrößen eingesetzt. Dabei rückt das Stichwort Datenqualität immer weiter in den Fokus, denn nur eine hohe Datenqualität kann valide Erkenntnisse gewährleisten. Weiterhin spielt in interdisziplinären Projekten in der Systembiologie auch die Dokumentation des kompletten Workflows zur Erzeugung der Roh- bzw. sogar Nutzdaten eine wesentliche Rolle.

Die vorliegende Dissertation stellt einen allgemeinen Leitfaden, sowie verschiedene Modellansätze und Analysemethoden vor, um die Qualität erhobener Daten zu ermitteln und zu analysieren. Die beschriebenen Methoden können einerseits zur Verarbeitung von Roh- zu Nutzdaten eingesetzt werden, bieten aber andererseits auch die Möglichkeit, den kompletten Prozess von der Probenahme bis zu den Rohdaten zu analysieren und zu optimieren. Weiterhin erlauben sie über die Definition von Qualitätskriterien auch eine Kontrolle und Filterung der Rohdaten und stellen eine konsistente Dokumentation aller experimentellen und theoretischen Prozesse dar, die zur Datenerzeugung herangezogen wurden.

Der vorgestellte Leitfaden wird auf vier unterschiedliche Fragestellungen angewendet. Die Bereitstellung extrazellulärer Raten und ihrer Standardabweichungen für Stoffflussanalysen wird über ein biotechnologisches Prozessmodell gekoppelt an verschiedene statistische Methoden realisiert. Das Backscatter Signal aus einem Mikrofermentationssystem wird auf Basis eines algebraischen Gleichungssystems in Verbindung mit einer Monte-Carlo Simulation verarbeitet und Wachstumsraten werden automatisiert abgeschätzt. Ebenfalls ein algebraisches Gleichungssystem beschreibt die komplette Prozessierung von Metabolom-Proben von der Probenahme bis zur Analytik und erlaubt damit eine Quantifizierung der intrazellulären Konzentrationen. Sensitivitätsanalysen stellen die Basis dafür dar, Optimierungspotential im experimentellen Ablauf zu identifizieren. Zur Umrechnung von Fluoreszenzsignalen in intrazelluläre Konzentrationen kommt ein dynamisches, biochemisches Modell aus Differentialgleichungen zum Einsatz.

Abstract

During systems biology projects experimental data is collected, interpreted, and applied for further analyses and studies. Thereby, data quality plays a more and more important role. Only high data quality ensures valid scientific insights. Furthermore, documentation of experimental workflows generating and processing data plays an important role in interdisciplinary projects.

This dissertation presents a general guideline and various modeling and analyzing techniques to estimate and analyse experimental data quality. The introduced methods can be applied to process raw data to used data. But they also provide the opportunity to analyse and optimize the complete experimental workflow. In addition, the methods can be used for data checking or filtering and ensure a consistent documentation of all experimental and theoretical processes involved in the data generation workflow.

The suggested guideline is adapted to four types of systems biological data answering different biological questions. Extracellular rates and their standard deviations are prepared for metabolic flux analyses by applying a biotechnological process model and statistical methods. The backscatter signal taken from a microbioreactor is processed based on an algebraic equation system combined with Monte Carlo simulations and growth rates are estimated automatically. Also, an algebraic equation system describes the complete processing of metabolome samples from sampling to quantification. The model provides quantification of intracellular metabolites and sensitivity analyses identify potential to optimize the experimental procedure. Finally, a dynamic, biochemical model consisting of differential equations is applied for calibration of fluorescence intracellular concentrations.

Teile der vorliegenden Dissertation wurden bereits als Beiträge in Zeitschriften veröffentlicht oder auf Konferenzen vorgestellt.

Artikel in Zeitschriften

KÄSS, F., A. PRASAD, J. TILLACK, M. MOCH, H. GIESE, J. BÜCHS, W. WIECHERT UND M. OLDIGES: *Rapid assessment of oxygen transfer impact for Corynebacterium glutamicum*. Bioprocess and biosystems engineering, 2014.

TILLACK, J., N. PACZIA, K. NÖH, W. WIECHERT UND S. NOACK: *Error Propagation Analysis for Quantitative Intracellular Metabolomics*. Metabolites 2(4):1012-1030, 2012.

Artikel in Konferenzbänden

HANNEMANN-TAMÁS, R., J. TILLACK, M. SCHMITZ, M. FÖRSTER, J. WYES, K. NÖH, E. VON LIERES, U. NAUMANN, W. WIECHERT UND W. MARQUARDT: *First- and second order parameter sensitivities of a metabolically and isotopically non-stationary biochemical network model*. In: *Proceedings of the 9th International Modelica Conference*. Hrsg. von M. ZIMMER UND D. OTTER. Munich, 2012: 641-648.

TILLACK, J., P. DROSTE, N. HACKBARTH, W. WIECHERT UND K. NÖH: *Visually-Assisted Modeling of Kinetic Metabolic Networks - From Omix to Modelica and Back*. In: *Proceedings of the 7th International Conference on Mathematical Modeling*. Vienna, Austria: Argesim, 2012.

TILLACK, J., N. PACZIA, S. LEWEKE, M. OLDIGES, W. WIECHERT, K. NÖH UND S. NOACK: *Modellierung und detaillierte Fehleranalyse der Metabolom-Datenprozessierung*. In: *Chemie Ingenieur Technik*, 82(9):1556-1557, 2010.

Vorträge auf Konferenzen

TILLACK, J., P. DROSTE, N. HACKBARTH, W. WIECHERT UND K. NÖH: *Visually-Assisted Modeling of Kinetic Metabolic Networks - From Omix to Modelica and Back*. MATHMOD 2012, Vienna, Austria, 2012.

Poster auf Konferenzen

TILLACK, J., N. HACKBARTH, P. DROSTE, W. WIECHERT UND K. NÖH: *A Modeling and Simulation Workflow for the Evaluation of Kinetic Network Models with OMIX and the MODELICA Library DYNAMO*, 1st European Congress of Applied Biotechnology, Berlin, September 25-29, 2011.

TILLACK, J., N. PACZIA, S. LEWEKE, M. OLDIGES, W. WIECHERT, K. NÖH UND S. NOACK: *Modeling and Detailed Error Analysis of Metabolome Data Processing*, 28. DE-HEMA-Jahrestagung der Biotechnologen, Aachen, September 21-23, 2010.

TILLACK, J., N. PACZIA, S. LEWEKE, W. WIECHERT, S. NOACK UND K. NÖH: *Sample Protocols by Numbers: Acquiring, Modeling and Exposing Errors for Metabolome Data Analysis*, ASIM-Workshop on Trends in Computational Science and Engineering: Foundations of Modeling and Simulation, Jülich, March 3-5, 2010.

Danksagung

Mein ganz besonderer Dank gilt Frau Dr. Katharina Nöh für die hervorragende Betreuung, viele spannende Diskussionen und ihr stetiges Interesse und Engagement in den letzten Jahren.

Herrn Prof. Dr. Andreas Kremling danke ich für das in mich gesetzte Vertrauen, das stete Interesse am Fortschritt dieser Arbeit und die freundliche Unterstützung.

Ich danke Herrn Prof. Dr. Wolfgang Wiechert für viele Denkanstöße und die Möglichkeit, diese Arbeit am Forschungszentrum Jülich anfertigen zu können.

Für die finanzielle Unterstützung im Rahmen des FlexFit Projektes danke ich dem Bundesministerium für Bildung und Forschung.

Für die gute Zusammenarbeit und viele überfachliche Diskussionen zur Bildung eines gemeinsamen Wortschatzes möchte ich mich bei Friedrich Käß, Nicole Paczia, Nuriye Mustafi, Stephan Noack und Julia Frunzke bedanken.

Allen Kollegen der Arbeitsgruppe Modellierung und Simulation gilt mein Dank für die tolle Atmosphäre und Hilfsbereitschaft.

Aus ganzem Herzen danke ich meiner Familie für ihre Liebe und Unterstützung.

Inhaltsverzeichnis

Zusammenfassung	III
Abkürzungen	XIII
1 Einleitung	1
1.1 Motivation	1
1.2 Zielsetzung und Aufbau der Arbeit	4
2 Modellierungs- und Analysemethoden in der Systembiologie	7
2.1 Einleitung	7
2.2 Modellierung biotechnologischer Prozesse	9
2.3 Modellierung biochemischer Prozesse	9
2.4 Sensitivitätsanalyse	10
2.5 Finite-Differenzen Approximation	11
2.6 Linearisierte Statistik	12
2.7 Fehlerfortpflanzung nach Gauß	13
2.8 Monte-Carlo Simulation	14
2.9 Parameterschätzung	16
3 Modellierung der Proben- und Datenprozessierung	19
3.1 Einleitung und Motivation	19
3.2 Leitfaden	21
4 Schätzung extrazellulärer Raten unter Nutzung von Prozessmodellen	25
4.1 Einleitung	25
4.2 Messung extrazellulärer Konzentrationen	27
4.3 Schätzung extrazellulärer Raten	27
4.4 Fehlerabschätzung mittels linearisierter Statistik	30
4.5 Fehlerabschätzung mittels Monte-Carlo Bootstrapping	34
4.6 Zusammenfassung und Diskussion	36
5 Quantitative Auswertung von Wachstumskurven aus dem BioLector®	39
5.1 Einleitung	39
5.2 Modellierung der Prozessierung vom Backscatter Signal	42
5.3 Modellcharakterisierung	49
5.4 Zusammenfassung und Diskussion	53

6	Modellierung und Analyse der Prozessierung von Metabolom-Daten	55
6.1	Einleitung	55
6.2	Prozessierung quantitativer Metabolom-Daten	57
6.3	Modellierung der Metabolom-Daten Generierung	61
6.4	Modellcharakterisierung	67
6.5	Optimierung des experimentellen Workflows	74
6.6	Zusammenfassung und Diskussion	75
7	Sensormodellierung	77
7.1	Einleitung	77
7.2	Aufbau und Analyse des Sensormodells	79
7.3	Anwendung des Sensormodells als Werkzeug zur Kalibrierung	92
7.4	Zusammenfassung und Diskussion	94
8	Gesamtzusammenfassung und Ausblick	95
	Abbildungsverzeichnis	xv
	Tabellenverzeichnis	xvii
	Literaturverzeichnis	xix
A	Extrazelluläre Raten	xxvii
A.1	Prozessmodell	xxvii
A.2	Experimentelle Daten	xxviii
B	BioLector®-Daten	xxxii
B.1	Daten für die Kalibrierfunktion	xxxii
B.2	Lineare Kalibrierfunktionen	xxxii
B.3	Dokumentation des MATLAB-Werkzeuges für BioLector®-Daten	xxxiv
C	Metabolom-Daten	xlvi
C.1	Rohdaten	xlvi
C.2	Datenauswertung mit unterschiedlichen Modellvarianten	lv
C.3	Sensitivitätsanalyse	lv
D	Sensormodell	lix
D.1	Experimentelle Daten	lix
D.2	Simulation und Analysen	lix

Abkürzungen

Abkürzung	Bezeichnung
AD	Automatische Differentiation
BTM	Biotrockenmasse
<i>C. glutamicum</i>	<i>Corynebacterium glutamicum</i>
COV	Covarianzmatrix
DAE	Differential-algebraische Gleichung
FD	Finite-Differenzen
FIM	Fisher-Information Matrix
<i>G. oxydans</i>	<i>Gluconobacter oxydans</i>
IBG	Institut für Bio- und Geowissenschaften am Forschungszentrum Jülich
<i>in silico</i>	Experimente im Computer
<i>in vitro</i>	Experimente unter künstlichen Bedingungen
<i>in vivo</i>	Experimente im lebenden System
MC	Monte-Carlo (Simulation)

Abkürzungen und Symbole, welche nur in einem Kapitel relevant sind, sind in diesem in einem separaten Symbolverzeichnis erläutert.

Kapitel 1

Einleitung

Der schlimmste aller Fehler ist, sich keines solchen bewusst zu sein.

Thomas Carlyle

1.1 Motivation

1.1.1 Systembiologie

Die Systembiologie ist ein interdisziplinäres Fachgebiet, welches sich durch ein Zusammenspiel von natur-, ingenieurwissenschaftlichen, mathematischen und informationstechnischen Herangehensweisen und Methoden auszeichnet (KITANO, 2002; WIECHERT, 2004; KLIPP et al., 2009; KREMLING, 2012). Das Ziel der Systembiologie ist es, die dynamischen Prozesse in biologischen Organismen zu entschlüsseln, um die Organismen in ihrer Gesamtheit zu verstehen und Vorhersagen über ihr Verhalten treffen zu können. Dabei findet häufig ein iterativer Zyklus aus Laborexperimenten, Modellbildung und Analysen Anwendung. Einen möglichen Überblick über die Schritte, die Teil eines systembiologischen Projektes sind, gibt Abbildung 1.1.

Modelle bilden einen wichtigen Baustein der Systembiologie, welcher auch in der vorliegenden Arbeit eine bedeutende Rolle einnimmt. Es kommen zwei unterschiedliche Modell-Typen zum Einsatz. Der erste Modell-Typ stellt eine Beschreibung der Prozessierung von Proben und Daten dar und erlaubt damit eine Generierung valider Messdaten und ihrer Fehler. Als zweites werden Stoffwechsel-Modelle eingesetzt, um das Verhalten des biologischen Systems bei definierten Randbedingungen vorherzusagen.

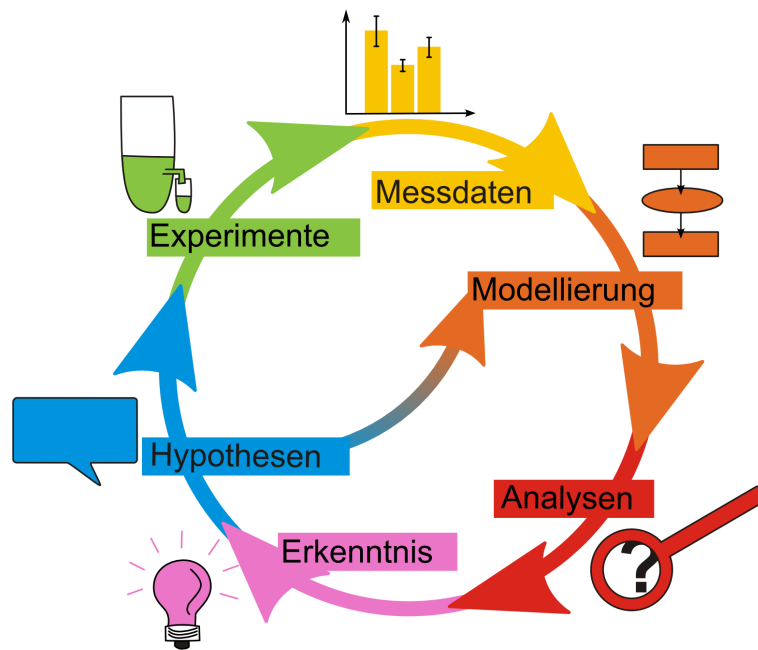


Abbildung 1.1: Systembiologischer Zyklus: Ausgehend von Wissen können Hypothesen erzeugt werden. Um diese Hypothesen bestätigen oder widerlegen zu können, wird eine Kette von Experimenten und Analytik sowie Modellierung und theoretischen Analysen angestoßen. Die daraus resultierenden Erkenntnisse können zum Erzeugen neuer Hypothesen herangezogen werden und der Zyklus beginnt erneut.

1.1.2 Multi-Omics Daten

„Omics“-Technologien spielen eine zentrale Rolle in der Systembiologie. In ihrer Gesamtheit aus Transkriptomik, Proteomik, Metabolomik und Fluxomik ermöglichen sie eine detaillierte Analyse komplexer, regulatorischer Mechanismen, welche die Basis aller zellulären Prozesse bilden (BÜSCHER et al., 2012). Allerdings hängen die Interpretation und Nutzbarkeit von multi-omics Daten stark von der Qualität der Daten ab. Dabei sollte neben dem Anstreben einer hohen Datenqualität vor allem daran gearbeitet werden, die Qualität von Daten möglichst genau quantifizieren zu können.

1.1.3 Datenqualität

Die Genauigkeit einer Messung kann durch die beiden Terme „Richtigkeit“ und „Präzision“ charakterisiert werden (Abbildung 1.2). Während die Richtigkeit die systematische Abweichung eines erhobenen Messwertes vom tatsächlichen Wert ist, ist die Präzision direkt an die Reproduzierbarkeit einer Messung gekoppelt. Sie definiert die Größe der Streuung mehrerer äquivalenter Messungen um den Erwartungswert.

In der Statistik werden die Werte für Richtigkeit und Präzision als Verzerrung und Varianz beschrieben.

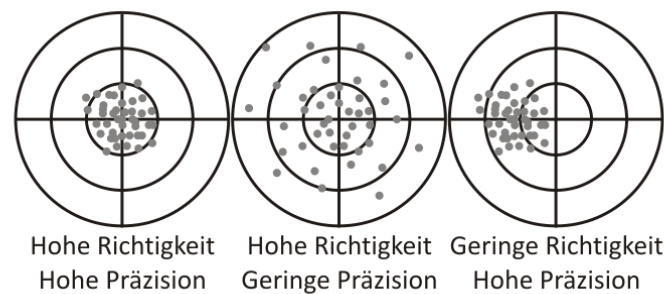


Abbildung 1.2: Schematische Darstellung der Terme Richtigkeit und Präzision einer Messung.

Aus wissenschaftlicher Sicht ist die Richtigkeit einer Messung wichtiger als die Präzision, da nur eine hohe Richtigkeit ein möglichst exaktes Abbild der Realität gewährleistet. Diese Exaktheit wiederum ist notwendig um nutzbare Hypothesen zu erstellen und zu testen.

Im Bereich der Biotechnologie und ganz besonders bei der Analyse komplexer biologischer Proben aus spezifischen Kultivierungs-Experimenten ist sowohl die Bestimmung der Richtigkeit einer Messung als auch das Erzielen einer hohen Richtigkeit eine schwierige Aufgabe.

Der Grund dafür liegt in vielen technischen (z.B. Bioreaktor) und biologischen (z.B. Zellwachstum) Faktoren, welche nur zum Teil zugänglich sind.

Eine Möglichkeit sich trotz allem der Qualität biologischer Daten zu nähern ist das Erheben von Replikaten der Probe (biologische Replikate) und Messung (technische Replikate). Dies ist jedoch nicht immer und vor allem nicht für jeden Prozessschritt möglich. Oft müssten dafür eine sehr große Menge teilweise komplexer experimenteller Prozesse durchgeführt werden und manchmal kann der interessierende Prozesszustand kein zweites Mal „angefahren“ werden.

Eine Alternative zu experimentellen Replikaten bieten modellgestützte Datenauswertungen mit integrierter Fehlerfortpflanzungen. Dafür werden die Schritte der Probengenerierung möglichst detailliert mit Hilfe eines Modells abgebildet. Dieses Modell berechnet dann aus verfügbaren experimentellen Daten und ihren Fehlern, im Weiteren als Rohdaten bezeichnet, die gewünschten Messwerte inklusive ihrer Standardabweichungen, im Weiteren als Nutzdaten bezeichnet.

Statistische Verfahren und Modelle können die Generierung valider Nutzdaten z.B. zu Interpretations- und Analysezwecken nachvollziehbar und somit reproduzierbar gewährleisten.

1.2 Zielsetzung und Aufbau der Arbeit

Die vorliegende Dissertation ist in der Arbeitsgruppe Modellierung und Simulation des Forschungszentrums Jülich am Institut für Bio- und Geowissenschaften entstanden. In dieser Dissertation wird eine Methode vorgestellt und angewendet, welche es ermöglicht die Qualität von Nutzdaten in der Systembiologie zu ermitteln und zu optimieren. Durch eine lückenlose Modellierung aller Schritte von der Probenahme über die Rohdaten bis zu den Nutzdaten, wird ein Modell bereit gestellt, welches zur validen Datenprozessierung und -interpretation sowie zur Identifizierung und Quantifizierung von Fehlerquellen und Optimierungspotential in der Prozessierungskette eingesetzt werden kann.

Nach einer Einleitung in die in der Arbeit eingesetzten theoretischen Methoden in Kapitel 2, wird in Kapitel 3 ein Leitfaden zur generellen Vorgehensweise bei der Datenverarbeitung und -analyse vorgestellt. Dieser enthält alle Schritte von der Analyse der Prozesskette, über die Modellierung bis zur Optimierung der experimentellen Abläufe und hat einen speziellen Fokus auf der Ermittlung valider Standardabweichungen. Dieser Leitfaden finden daraufhin Anwendung bei verschiedenen Datentypen und Fragestellungen der Systembiologie.

Kapitel 4 beschäftigt sich mit der Bereitstellung von extrazellulären Raten für Anwendungen der Stoffflussanalyse. Die Auswertung von Wachstumskurven aus dem BioLector®, einem Mikrofermentationssystem, steht im Mittelpunkt des Kapitel 5. In Kapitel 6 wird die Prozessierung von Metabolomdaten modelliert und analysiert. Abschließend wird in Kapitel 7 ein Stoffwechsel-Modell in Form eines biochemischen Differentialgleichungssystems eingesetzt, um aus einem Fluoreszenzsignal Rückschlüssen auf die dazugehörigen Metabolit-Konzentration ziehen zu können.

Kapitel 2

Modellierungs- und Analysemethoden in der Systembiologie



In der vorliegenden Dissertation werden etablierte systembiologische Methoden der Modellbildung und Modellanalyse eingesetzt. Diese Methoden werden eingeführt und kurz erläutert. Die einzelnen Abschnitte stellen eine Zusammenfassung der jeweiligen Methode dar und liefern die zum Verständnis der nachfolgenden Kapitel notwendigen Vokabeln.

2.1 Einleitung

Modelle nehmen in der vorliegenden Arbeit eine zentrale Rolle ein. Methoden der Modellbildung und -analyse, welche in der Systembiologie etabliert sind, werden in diesem Kapitel kurz erläutert und im folgenden Verlauf der Arbeit bei verschiedenen Fragestellungen eingesetzt.

Die Anwendung der verschiedenen Methoden folgt in der Regel aufeinander. Basierend auf Wissen und Hypothesen wird ein mathematisches Modell formuliert und mit experimentellen Daten kalibriert, welches anschließend mit mathematischen Methoden analysiert wird. Daraufhin können mit experimentellen Daten die Modellparameter geschätzt werden. Das angepasste Modell wird dann wiederum bei der Planung neuer Experimente eingesetzt. Ein möglicher Ablauf der nachfolgend vorgestellten Methoden und die dazugehörigen einführenden Abschnitte sind in Tabelle 2.1 aufgeführt.

Tabelle 2.1: Übersicht über die etablierten Vorgehensweisen und Methoden in der Systembiologie, welche dieser Arbeit zugrunde gelegt werden.

Schritt/Methode	Erläuterung/Abschnitt
Analyse der Fragestellung	auf Basis der biologischen Frage werden die zur Beantwortung der Frage wesentlichen Prozesse identifiziert
Modellbildung	Formulierung der Prozesse in einem Modell als vereinfachte Abbildung der Wirklichkeit Modellierung biotechnologischer Prozesse (2.2) Modellierung biochemischer Prozesse (2.3)
Modell-Kalibrierung	reale Daten werden eingesetzt, um sinnvolle Bereiche für die Modellparameter zu identifizieren
Modell-Simulation	Ausführen des Modells zur Umwandlung der Eingangs- in Ausgangsgrößen
Modell-Charakterisierung	Untersuchung des Modell-Verhaltens und Rückschließen auf Prozess- bzw. Systemeigenschaften
Sensitivitätsanalyse	Finite-Differenzen Approximation (2.4, 2.5) Monte-Carlo Bootstrapping (2.8.1)
Statistische Analyse	Linearisierte Statistik (2.6) Fehlerfortpflanzung nach Gauß (2.7) Fehlerfortpflanzung mittels Monte-Carlo Simulation (2.8.2)
Parameterschätzung	Schätzung der Modellparameter auf Basis experimenteller Daten Parameterschätzung (2.9)
Modell-Validierung	Überprüfen des Modellverhaltens für unabhängige Daten
Experimentelles Design	Planen neuer, informationsreicher Experimente siehe 'Sensitivitätsanalyse' und 'Statistische Analyse'

2.2 Modellierung biotechnologischer Prozesse

Bioprozesse sind Prozesse in Bioreaktoren, bei welchen Mikroorganismen oder Zellkulturen unter definierten Bedingungen bezüglich Wachstum oder Produktion kultiviert werden. Je nach Anwendung kann zwischen verschiedenen Arten von Bioreaktoren und Prozessführungen unterschieden werden (WILLIAMS, 2002). Überlicherweise findet der Rührreaktor Anwendung und gängige Prozess-Typen sind Batch, Fed-Batch und kontinuierlicher Modus.

Biotechnologische Prozessmodelle führen zum Einen zum verbesserten Verständnis der ablaufenden Prozesse und erlauben zum Anderen eine Überwachung und Steuerung der Reaktoren (BAILEY, 1998). Aus diesem Grund ist diese Art der Modellierung eine gängige Methode im Bereich der biotechnologischen Prozessentwicklung und -optimierung.

Ein typisches Bioprozessmodell besteht aus einem Satz nichtlinearer, kontinuierlicher differential-algebraischer Gleichungen (DAE), welche den Zustand im Reaktor mit Hilfe von Kinetiken beschreiben (ROELS, 1983). Es kann zwischen strukturierten und unstrukturierten Modellen unterschieden werden. Den unstrukturierten Modellen liegen die Vereinfachungen zu Grunde, dass die Zelle als homogene Masse angenommen wird, welche gleichmäßig wächst und sich teilt (PRÄVE et al., 1994). Im einfachsten Fall beschreibt ein Bioprozessmodell die Umsetzung von Substraten zu Produkten und Biomasse. Als Systemgrößen werden in der Regel extrazelluläre Konzentrationen eingesetzt, deren zeitliche Änderungen durch Raten (Substratverbrauchs-, Produktbildungs- und Wachstumsraten) beschrieben werden.

In Kapitel 4 der vorliegenden Dissertation wird ein Prozessmodell an statistische Methoden gekoppelt, um extrazelluläre Raten und ihre Standardabweichungen zu ermitteln.

2.3 Modellierung biochemischer Prozesse

Die mathematische Modellierung spielt seit Beginn des zwanzigsten Jahrhunderts eine immer wichtiger werdende Rolle in der Biochemie. Die ersten mathematisch beschriebenen, biochemischen Vorgänge waren enzymatische Umsetzungen von Substrat(en) zu Produkt(en) (HENRI, 1902; MICHAELIS und MENTEN, 1913). Heutzutage werden mathematische Modelle genutzt, um den zellulären Stoffwechsel mit Hilfe von Kinetiken abzubilden und zu untersuchen (RIZZI et al., 1997; VASEGHI et al., 1999; CHASSAGNOLE et al., 2002; VISSER et al., 2004; ALTINTAS et al., 2006; MAGNUS et al., 2006). Die Kinetiken können dabei nahezu beliebig komplex werden, häufig wird jedoch auf verfügbare, vereinfachte Ansätze wie

Powerlaw (SAVAGEAU und VOIT, 1982), Linlog (WESTERHOFF und VAN DAM, 1987) oder Convenience (LIEBERMEISTER und KLIPP, 2006) zurückgegriffen.

Alternativ zur dynamischen Modellierung werden genomweite Netzwerke stöchiometrie-basiert modelliert (FEIST et al., 2007; KJELDSSEN und NIELSEN, 2009). Stöchiometrische Netzwerke minimieren den Effekt der nicht im Modell enthaltenen Umgebung, decken jedoch keine regulatorischen Effekte, z.B. Inhibierungen, ab. Außerdem gelten sie nur unter metabolischer Stationarität und sind grundsätzlich unterbestimmt.

Für die Modellierung biochemischer Prozesse gibt es einige generelle Modellierungssprachen (SBML, MODELICA) und -programme (MATLAB, DYMOLA).

In Kapitel 7 dieser Dissertation findet ein dynamisches Modell zur Analyse von Fluoreszenz-Daten Anwendung.

2.4 Sensitivitätsanalyse

Die Sensitivitätsanalyse ist eine Analyseform für komplexe Systeme mit Wechselwirkungen zwischen den Modellparametern. Sie untersucht den Einfluss jedes Parameters auf das Systemverhalten und die Zustandsgrößen. Ein Parameter gilt als sensitiv in Bezug auf eine Zustandsgröße, wenn seine Änderung eine signifikante Veränderung der Zustandsgröße bewirkt. Eine Definition der Sensitivitätsanalyse kann SALTELLI et al. (2000) entnommen werden.

Die Frage, wie stark sich die Änderung eines Parameters auf die Zustandsgrößen des Systems auswirkt, spielt u.A. im Metabolic Engineering eine wesentliche Rolle. Wenn man Kenntnis darüber besitzt, welche Parameter zu einer Erhöhung einer erwünschten Zustandsgröße führen, kann das System gezielt verändert werden. Anwendung findet die Sensitivitätsanalyse auch beim Design neuer Experimente, da *in silico* abgeschätzt werden kann, welche Messungen zu welchen Messzeitpunkten für die spätere Parameterschätzung den höchsten Informationsgehalt haben.

Die Methoden der Sensitivitätsanalyse können in zwei Gruppen eingeteilt werden. So kann zwischen lokalen und globalen Sensitivitätsmethoden unterschieden werden. Der Unterschied von lokalen zu globalen Sensitivitätsfunktionen ist, dass jeweils nur der untersuchte Parameter variiert wird, alle anderen werden konstant gehalten. Eine lokale Sensitivitätsanalyse ist die Finite-Differenzen Approximation, welche in Abschnitt 2.5 vorgestellt wird. Eine globale Methode ist die Monte-Carlo Simulation. Sie wird in Abschnitt 2.8 erläutert.

Die Frage, wie sich die Änderung eines Parameters auf die Zustandsgrößen des Systems auswirkt, spielt in vielen Bereichen dieser Arbeit eine Rolle.

2.5 Finite-Differenzen Approximation

Die Finite-Differenzen (FD) Approximation ist eine Methode der lokalen Sensitivitätsanalyse (vgl. Abschnitt 2.4). Die lokalen Sensitivitäten sind als partielle Ableitungen jeder Zustandsgröße c_i nach jedem Parameter α_j definiert.

$$S_{i,j}(t) = \frac{\partial c_i(t, \alpha)}{\partial \alpha_j} \quad \forall i = 1 \dots m, \forall j = 1 \dots p \quad (2.1)$$

Die Sensitivitätsmatrix (**S**) erlaubt Aussagen darüber zu treffen, ob und wie stark die Änderung eines Parameters die Änderung einer Zustandsgröße bedingt. So bedeuten z.B. negative Werte in der Matrix ($S_{i,j} < 0$), dass die Erhöhung des Parameters j den Wert der Zustandsgröße i verringert.

Bei der FD Approximation wird die Ableitung durch einen Differenzenquotienten approximiert. Dieser beschreibt das Verhältnis der Veränderung einer Größe zu der Veränderung einer zweiten Größe. Dieses Vorgehen zur Berechnung von Ableitungen ist unter der Annahme möglich, dass die Änderung des Parameters $\Delta\alpha$ ausreichend klein ist und der Parameter sich im Intervall $[\alpha + \Delta\alpha, \alpha - \Delta\alpha]$ linear verhält.

Nutzt man den zentralen Differenzenquotienten für die FD Approximation, verändert man den ausgewählten Parameter j sowohl um $\Delta\alpha_j$ nach oben als auch nach unten. Alle anderen Parameter $\alpha_{k=1\dots p \neq j}$ bleiben jeweils unverändert. Die Sensitivität kann annähernd beschrieben werden durch (SALTELLI et al., 2000):

$$S_{i,j}(t) \approx \frac{c_i(t, \alpha_j + \Delta\alpha_j, \alpha_{k=1\dots p \neq j}) - c_i(t, \alpha_j - \Delta\alpha_j, \alpha_{k=1\dots p \neq j})}{2 \cdot \Delta\alpha_j} \quad (2.2)$$

Vorteile der FD Approximation sind eine einfache Implementierung und ein geringer Rechenaufwand. Die Ergebnisse hängen jedoch von der gewählten Schrittweite $\Delta\alpha_j$ ab. Ist diese zu klein, kann es zu Rundungs- bzw. Konditionierungsfehlern kommen. Wird $\Delta\alpha_j$ dagegen zu groß gewählt, kann die Annahme der Linearität bei nichtlinearen Modellen oft nicht mehr erfüllt werden. Die abgeschätzten Sensitivitäten sind dann fehlerhaft.

Eine Alternative zur FD Approximation ist die Automatische Differentiation (AD, GRIEWANK und WALTHER, 2008).

2.6 Linearisierte Statistik

Mithilfe einer Sensitivitätsmatrix \mathbf{S} (z.B. aus Abschnitt 2.5) können Standardabweichungen für alle Modellparameter abgeschätzt werden. Man erhält damit ein Maß der Genauigkeit der Parameterschätzung. Auf die dafür notwendigen Berechnungsschritte wird im Folgenden eingegangen:

1 Berechnung der Fisher-Informationsmatrix (FIM)

$$\mathbf{FIM}(t) = \mathbf{S}(t)^T \cdot \mathbf{W}(t) \cdot \mathbf{S}(t) \quad (2.3)$$

Die Fisher-Informationsmatrix ist eine Größe aus der Statistik, mit welcher Parameterschätzungen bewertet werden können.

Die Gewichtsmatrix \mathbf{W} berechnet sich durch die Beziehung $\mathbf{W} = \frac{1}{\mathbf{var}}$, wobei \mathbf{var} eine Diagonalmatrix ist, welche die Varianzen der gemessenen Zustandsgrößen enthält. Die Gewichtsmatrix dient der Normierung von Messungen unterschiedlicher Größenordnungen.

2 Aufsummierung der FIM über die Zeit

$$\mathbf{F}_{sum} = \sum_i \mathbf{FIM}(t_i) \quad (2.4)$$

Bei dynamischen Modellen ergibt sich für jeden Zeitpunkt eine Fisher-Informationsmatrix. Um einen Wert für den Informationsgehalt über die komplette Zeit und damit im nächsten Schritt eine Parametervarianz pro Parameter zu erhalten, werden diese über die Zeit aufsummiert.

3 Berechnung der Parameter-Kovarianzmatrix

$$\mathbf{Cov} = \mathit{inv}(\mathbf{F}_{sum}) \quad (2.5)$$

Die Kovarianzmatrix enthält auf der Hauptdiagonalen die Varianzen aller betrachteten Parameter. Alle anderen Einträge der Matrix beinhalten Kovarianzen und stellen damit ein Merkmal für die Beziehung zweier Parameter dar. Die Matrix berechnet sich aus der Inversen der aufsummierten Fisher-Informationsmatrix. Dieser Schritt ist nur dann möglich, wenn es sich bei der Fisher-Informationsmatrix um eine reguläre Matrix handelt.

Vor allem bei Netzwerken mit komplexen Kinetiken oder vielen Modellparametern kann es vorkommen, dass Korrelationen zwischen den Parametern dazu führen, dass die Parameter strukturell nicht identifizierbar sind und damit ihre Standardabweichungen bei linearisierten Methoden nicht aus der Sensitivitätsmatrix berechnet werden können. Es treten Singularitäten auf, die verhindern, dass die Inverse der zeitlich aufsummierten Fisher-Informationsmatrix berechnet werden kann. Dieses Problem kann behoben werden, indem eine bestimmte Anzahl der unbekannt, korrelierten Parameter festgesetzt oder eine Modellreduktion durchgeführt wird.

4 Berechnung des Vektors der Standardabweichungen (σ) aus den Diagonalelementen der Kovarianzmatrix

$$\sigma = \sqrt{\text{diag}(\mathbf{Cov})} \quad (2.6)$$

Die Standardabweichung ist ein Maß für die Streuung eines Parameters um seinen Mittelwert. Umso geringer diese Streuung ist, desto besser kann der jeweilige Parameter bestimmt werden.

2.7 Fehlerfortpflanzung nach Gauß

Werden mathematische Formeln eingesetzt um Rohdaten in Nutzdaten umzurechnen, müssen ebenfalls die Messgenauigkeiten der Rohdaten in Standardabweichungen der Nutzdaten überführt werden. Dabei kommen Regeln der Fehlerfortpflanzung zum Einsatz.

Liegt eine mathematische Funktion mit mehreren unabhängigen, fehlerbehafteten Variablen x_i vor und weichen die Variablen um einen Wert Δx_i vom „richtigen“ Wert ab, kann die Standardabweichung des gesuchten Nutzdatums Δy mittels Taylorreihe abgeschätzt werden. Sind die Standardabweichungen der Variablen genügend klein, kann nach dem Glied erster Ordnung abgebrochen werden.

$$y = f(x_1, x_2, \dots, x_n) \quad (2.7)$$

$$(\Delta y)^2 \approx \left(\frac{\partial y}{\partial x_1} \cdot \Delta x_1 \right)^2 + \left(\frac{\partial y}{\partial x_2} \cdot \Delta x_2 \right)^2 + \dots + \left(\frac{\partial y}{\partial x_n} \cdot \Delta x_n \right)^2 \quad (2.8)$$

Diese Fehlerfortpflanzung ist nur anwendbar, wenn sich die Funktion bei Änderung der Messgrößen x_i im Rahmen ihrer Messgenauigkeiten Δx_i hinreichend linear verhält. Aus diesem Grund spricht man hierbei auch von linearer Fehlerfortpflanzung.

Für Funktionen mit nichtlinearem Verhalten kann z.B. die Monte-Carlo Simulation eingesetzt werden (vgl. Abschnitt 2.8.2).

2.8 Monte-Carlo Simulation

Die Monte-Carlo (MC) Simulation ist eine weitere Möglichkeit, Standardabweichungen von Parametern zu schätzen oder die Fortpflanzung von Standardabweichungen durch Modelle zu untersuchen. Die Idee der Monte-Carlo Simulation basiert auf dem Gesetz der großen Zahlen.

Unter dem Begriff Monte-Carlo Simulation sind verschiedene Anwendung bekannt, daher wird im Folgenden explizit auf die in der Arbeit genutzten Vorgehensweisen des parametrischen Bootstrapps eingegangen.

2.8.1 Bootstrap-Methode

Das MC Bootstrapping ist, im Gegensatz zur FD Approximation, eine Methode der globalen Sensitivitätsanalyse. Das bedeutet, dass alle Parameter parallel variiert werden und ihr Einfluss auf die Zustandsgrößen analysiert wird (vgl. Abschnitt 2.4).

Ziel des MC Bootstrapping ist es, durch wiederholte Optimierungen eine Verteilung jedes zu schätzenden Parameters zu erhalten und aus dieser die Varianz des Parameters abzuschätzen. Je größer die Zahl der Optimierungen, umso genauer die Verteilung, aber auch umso höher der Rechenaufwand. Die optimale Anzahl an MC Simulationen muss für jede Fragestellung separat ermittelt werden.

Folgende Schritte werden für eine vorgegebene Anzahl an Iterationen durchgeführt:

1. Die gemessenen oder simulierten Konzentrationsverläufe werden um eine zufallsbehaftete Standardabweichung verrauscht. So entstehen im Rahmen der Messgenauigkeit bei jeder Iteration andere Konzentrationsverläufe aus denen synthetische Datensätze generiert werden.
2. Zusätzlich werden die Startwerte der zu optimierenden Parameter vor jeder Optimierung (vgl. Abschnitt 2.9) um 20-50% verrauscht, um die Optimierung von unterschiedlichen Punkten aus zu starten, damit ein breiteres Spektrum des Optimierungsgebietes abzudecken und im besten Fall das globale Optimum zu finden.

3. Optimierung: Minimierung der Fehlerquadratsumme für jeden Startwert.
4. Nach der Optimierung wird der Parametersatz mit der kleinsten Fehlerquadratsumme ermittelt und gespeichert.

Nach dem Abschluss der Optimierung, liegt aus jeder Iteration für jeden Parameter ein optimierter Wert vor. Die große Menge an Daten kann in einem Histogramm dargestellt werden.

Datenauswertung

Bei der Auswertung der durch die MC Simulation erhaltenen Daten wird eine Verteilung ermittelt, welche die Daten bestmöglich beschreibt. In der Regel findet dabei eine Normalverteilung ($N(\mu, \sigma^2)$) Anwendung, mit welcher Mittelwert (μ) und Standardabweichung (σ) für jeden Parameter abgeschätzt werden.

Vorher werden die Daten des MC Bootstrapping wie folgt vorbereitet:

Ausreißertest Als Ausreißer werden nach JOSHI et al., 2006 diejenigen Werte bezeichnet, die weit entfernt vom Mittelwert liegen und dadurch einen großen Einfluss auf die Statistik haben. Für die Identifizierung der Ausreißern werden die Daten der Größe nach sortiert und in 4 Abschnitte geteilt, jeder dieser Abschnitte enthält die gleiche Anzahl an Daten. Die Grenze zwischen zwei Vierteln wird als Quartil (Q) bezeichnet. Alle Werte die außerhalb des Bereiches $[Q_1 - 1,5 \cdot (Q_3 - Q_1) \quad Q_3 + 1,5 \cdot (Q_3 - Q_1)]$ liegen, werden aus dem Datenvektor entfernt.

Vertrauensintervall Das Vertrauensintervall ist ein statistisches Maß und gibt den Bereich an, in dem sich ein zu schätzender Parameter bei unendlicher Wiederholung der Schätzung mit einer definierten Wahrscheinlichkeit z.B. 95% befindet. Für die Auswertung der Ergebnisse von Bootstrapping Analysen, werden die Daten außerhalb des 95%-Vertrauensbereiches aus dem Datenvektor entfernt.

In dieser Arbeit wird das MC Bootstrapping zur Schätzung extrazellulärer Raten in Kapitel 4 eingesetzt.

2.8.2 Fehlerfortpflanzung mittels Monte-Carlo Simulation

Wie im Abschnitt 2.7 eingeführt, können Modelle genutzt werden, um Rohdaten in Nutzdaten umzuwandeln und gleichzeitig eine Genauigkeit der Nutzdaten zu ermitteln. Für nichtlineare Modelle stößt die Fehlerfortpflanzung nach Gauß schnell an ihre Grenzen. Eine Alter-

native, welche ohne die Annahme der Linearität auskommt, ist die MC Simulation. Hierbei wird eine ausreichend große Anzahl an Nutzdaten, jeweils basierend auf im Rahmen ihrer Messgenauigkeit verrauschten Rohdaten, erzeugt und aus der Nutzdatenverteilung Mittelwert und Standardabweichung abgeschätzt.

In der vorliegenden Arbeit wird diese MC Simulation in Kapitel 6 zur Analyse von Metabolom-Daten eingesetzt.

2.9 Parameterschätzung

Ziel einer Parameterschätzung ist es, die Werte der Parameter eines Modells zu ermitteln, mit denen eine Simulation des Modells den durch die experimentellen Daten vorgegebenen Verlauf am besten wiedergeben. Eine Parameterschätzung ist ein Optimierungsproblem.

Die Optimierung ist ein Teilgebiet der Mathematik, welche vielfach und in unterschiedliche Komplexität Anwendung in der Systembiologie findet (BANGA, 2008) und sich damit beschäftigt, optimale Parameter eines Systems zu finden. Parameter sind dann optimal, wenn eine Zielfunktion mit ihnen ihr Minimum erreicht. Gleichungs- und Ungleichungsnebenbedingungen begrenzen das Suchgebiet und schränken damit die Werte ein, welche die zu optimierenden Parameter annehmen dürfen.

Es gibt eine ganze Reihe von Optimierungsverfahren, die je nach zu optimierendem System eingesetzt werden und unterschiedliche Methoden nutzen (PAPAGEORGIU, 1996). In erster Unterscheidung kann man alle Verfahren in zwei Gruppen einteilen, die globalen und die lokalen Schätzverfahren.

Lokale Verfahren suchen in einer durch die Nebenbedingungen festgelegten Umgebung um den Startpunkt. Die meisten numerischen Lösungsverfahren verfolgen die Liniensuch- (Line Search-) Strategie (z.B. die gradientenbasierte Optimierung), alternativ gibt es Trust-Region-Verfahren. Lokale Optimierungsverfahren finden in der Regel ein Optimum, welches in der Nähe des Startpunktes liegt. Dies muss jedoch nicht das globale Optimum sein.

Genetische/evolutionäre Optimierungsverfahren sind globale Verfahren und an biologische Vorbilder angelehnt. Neue Suchpunkte werden durch die Anwendung stochastischer Operationen, z.B. Selektion, Mutation und Rekombination, festgelegt. Diese Gruppe von Algorithmen eignet sich insbesondere für solche Optimierungsaufgaben, bei denen die lokalen

Verfahren auf Grund hoher Komplexität des Systems oder der Existenz zahlreicher Optima versagen. Der Nachteil ist ein hoher Rechenaufwand, da nicht gezielt in einer Richtung, sondern zufällig in alle Richtungen gesucht wird.

Genetische/Evolutionäre Verfahren können auch mit klassischen lokalen Optimierungsmethoden kombiniert werden. Man spricht dabei von hybriden Verfahren.

Eingesetzt wird die Parameterschätzung mittels Optimierung in Kapitel 4 bei einem Bioprozessmodell und in Kapitel 7 bei einem biochemischen Modell.

Kapitel 3

Modellierung der Proben- und Datenprozessierung

Die Prozessierung biologischer Proben und Daten erfolgt in der Regel nach Laborprotokollen, die eine Vielzahl unterschiedlicher Schritte umfassen. Der Fehler des Gesamtprozesses setzt sich aus denjenigen der einzelnen Schritte zusammen. Eine lückenlose Modellierung bietet die Möglichkeit, präzise und damit reproduzierbare Nutzdaten zu erzeugen. Außerdem kann ein solches Modell eingesetzt werden, um Fehlerquellen erst zu identifizieren und dann zu quantifizieren. Dieses Wissen über Fehlerquellen kann letztendlich zur Optimierung der kompletten Prozessierungskette eingesetzt werden. Dieses Kapitel gibt einen allgemeinen Überblick und stellt einen Leitfaden für die generelle Vorgehensweise auf, welcher in den nachfolgenden Kapiteln Anwendung findet.

3.1 Einleitung und Motivation

Daten bilden einen wesentlichen Baustein in allen Wissenschaften und eine hohe Datenqualität ist die Grundlage fundierter wissenschaftlicher Erkenntnisse. In der vorliegenden Dissertation steht die Qualität systembiologischer Daten im Fokus. Durch den Einsatz und die Kombination verschiedener etablierter Modellierungs- und Analyse-Methoden wird die Generierung diverser systembiologischer Daten analysiert und hinsichtlich Präzision und Reproduzierbarkeit optimiert. Der modellbasierte Ansatz, welcher dabei Anwendung findet, bietet neben einer detaillierten Datenauswertung und Prozessbeschreiben das Potential, die modellierten Vorgänge mit mathematischen Methoden zu analysieren. Aufbauend auf diesen *in silico* Analysen können z.B. die experimentellen Prozesse optimiert und die Datenqualität verbessert werden.

Die Analyse der Qualität systembiologischer Daten ist von großem Interesse.

- Viele Protokolle werden eingesetzt, weil sie sich im Laboralltag bewährt haben, nicht aber, weil sie zur Generierung spezifischer Daten optimal sind.
- Die Reproduzierbarkeit von Forschungsergebnissen rückt immer mehr in das Bewusstsein der Wissenschaftler. SOPs (Standard Operating Procedures) im Labor sind die Grundlage für reproduzierbare Rohdaten. Standards in der Datenprozessierung fehlen häufig.
- In anderen Naturwissenschaften, vor allem in der Physik (TAYLOR, 1997), spielt die Betrachtung von Messfehlern bereits eine wesentlich größere Rolle.
- Für jede Technik gibt es spezifische Metadaten (Bioreaktor, Medium, Säule, etc.).
- Nicht alle Zwischenschritte können durch Messungen verfolgt und/oder unabhängig voneinander betrachtet werden.

Zur detaillierten Analyse des Workflows der Nutzdatenerzeugung ist daher eine formale Beschreibung der Wirklichkeit - ein mathematisches Modell - notwendig. Solch ein Modell beschreibt jeden einzelnen Schritte der Proben- und Datenprozessierung. Es erlaubt die Betrachtung des Datenflusses und, durch die Kopplung mit statistischen Methoden, eine lückenlose Quantifizierung der Datengenauigkeit. Das Wissen über die Genauigkeit von Daten ist essentiell und spielt vor Allem dann eine Rolle, wenn Signifikanz nachgewiesen werden soll oder wenn die Nutzdaten wiederum Eingangsparameter mathematischer Analysen sind, z.B. zur Parameterschätzung eingesetzt werden.

Neben der Nachvollziehbarkeit und der Beschreibung hat ein Modell außerdem den Vorteil, dass es über Simulationen Verbesserungspotential aufzeigen kann. Eine Anpassung des experimentellen Workflows auf Basis der durch ein Modell erzeugten Vorschläge, kann zu einer verbesserten Datenqualität führen.

Die Modellierung der Proben- und Datenprozessierung führt somit zu Antworten auf die folgenden essentiellen Fragen:

- Wie genau sind die Nutzdaten, die mit dem betrachteten Protokoll erzeugt wurden?
- Wie wirken sich jeder Schritt der Prozessierung und jede Messung auf die Datenqualität aus?
- Wie kann die Datenqualität erhöht werden?
- Kann eine zusätzliche oder eine alternative Messung zu genaueren Nutzdaten führen?

Im Anschluss ist der ideale Ablauf zur Optimierung der Nutzdatenerzeugung in Form eines Leitfadens dargestellt, welcher in den nachfolgenden Kapiteln für verschiedenen Datentypen Anwendung findet.

3.2 Leitfaden

3.2.1 Analyse des Workflows

Der erste Schritt auf dem Weg zur Rohdaten-Prozessierung unter Nutzung eines Modells beinhaltet die genaue Analyse des experimentellen Workflows. Zu Beginn dieser Betrachtungen ist zu klären, wofür welche Nutzdaten benötigt werden. Die Grundlage zur Klärung von Zielsetzung und Datenlage bilden immer die biologischen Fragen, welche unter Zuhilfenahme der Nutzdaten beantwortet werden können. Für die bessere Anschaulichkeit der nachfolgenden Betrachtungen wird die biologisch relevante Frage im Mittelpunkt stehen, mit welcher Aufnahmerate ein bestimmtes Substrat von einem Organismus aufgenommen werden kann.

Häufig sind die gewünschten Daten, wie z.B. die Substrataufnahmerate, nicht direkt messbar. Für diesen Fall schließt sich die Frage an, welche Rohdaten (1) messbar sind und (2) in einem direkt Zusammenhang zu den benötigten Nutzdaten stehen. Die Substrataufnahmerate kann mit Hilfe einer zeitlichen Reihe von Substratkonzentrationen mathematisch bestimmt werden. Aber auch Konzentrationen können nur indirekt gemessen werden. So ist der Messwert, den ein Massenspektrometer ausgibt, eine Peakfläche, welche über eine lineare Regression in eine Konzentration umgerechnet wird.

Sobald diese grundlegenden Fragen analysiert und beantwortet werden konnten, rückt der experimentelle Workflow, die Probenprozessierung, in den Fokus. Dieser Workflow beschreibt, wie experimentelle Proben beginnend mit der Probenahme prozessiert werden und endet mit den Rohdaten. Die detaillierte Analyse des Workflows umfasst dabei auch die Analytik und die experimentellen Methoden. Der Workflow sollte dabei so detailliert wie möglich aufgelöst werden, um später ein präzises Modell aufstellen zu können.

Zusätzlich zur Probenprozessierung soll auch die Prozessierung der Daten in das Modell mit einfließen. Die Datenprozessierung umfasst alle Schritte, welche notwendig sind, um aus den Rohdaten Nutzdaten zu erzeugen, z.B. die Daten zur Erzeugung der Parameter für

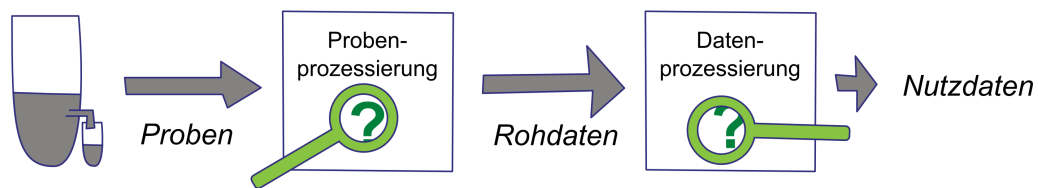


Abbildung 3.1: Analyse der einzelnen Schritte im Workflow zur Nutzdatengenerierung.

die lineare Regression. Weiterhin ist an dieser Stelle auch zu klären, ob für eine vollständige Datenprozessierung zusätzliche Messungen erhoben werden müssen, z.B. die Daten der Regressionsgeraden.

Alle diese Informationen beschreiben den Prozess der Nutzdatengenerierung (Abbildung 3.1) und fließen direkt in den nächsten Schritt - die Modellbildung - ein. Aus diesem Grund ist es besonders wichtig, auf die folgenden Aspekte zu achten:

- Welche zusätzlichen Daten (neben den primär betrachteten Rohdaten) werden benötigt? Welche davon sind messbar?
- Können für alle benötigten Daten Standardabweichung - zufällige Fehler - quantifiziert werden?
- An welchen Stellen im Workflow können systematische Effekte auftreten, welche den Wert der Rohdaten verfälschen? Gibt es eine Möglichkeit, diese Effekte zu quantifizieren?

Am Ende dieses Prozesses steht eine detaillierte, verbale Beschreibung aller Schritte von der Probenahme zu den Nutzdaten. Diese Beschreibung bildet die Grundlage der Modellierung.

3.2.2 Modellierung der Zustände und Fehler

Die im vorangegangenen Abschnitt erarbeitete verbale Beschreibung der Nutzdatenerzeugung dient als Grundlage des Modells für (1) die Korrektur der Rohdaten, (2) deren Umwandlung in Nutzdaten und (3) die Bestimmung der Genauigkeit der Nutzdaten. Der Workflow der Proben- und Datenprozessierung wird in mathematische Formeln übersetzt (Abb. 3.2).

Die Bestimmung der Genauigkeiten der Nutzdaten wird durch die Einbettung des Modells in eine Routine zur Berechnung von Standardabweichungen realisiert (Abbildung 3.3). Dadurch ist es möglich, die Fortpflanzung der Standardabweichungen aller eingehenden Roh-

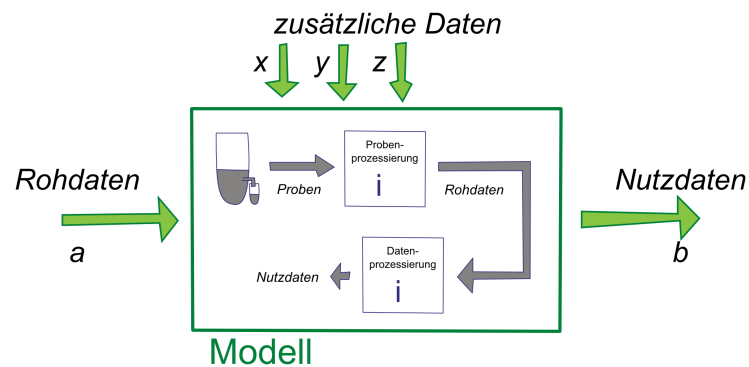


Abbildung 3.2: Das Modell - eine formale Beschreibung des Workflows zur Nutzdatengenerierung.

daten auf die Standardabweichung der Nutzdaten zu untersuchen. Abhängig von den Charakteristika des Modells werden dabei verschiedene Methoden der Sensitivitätsanalyse und Fehlerfortpflanzung eingesetzt (vgl. Abschnitte 2.4 - 2.8).

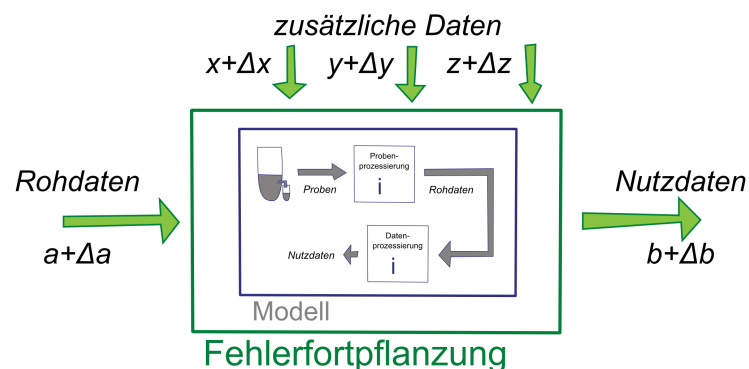


Abbildung 3.3: Einbetten des Modells in eine Fehlerfortpflanzungsroutine.

3.2.3 Modellvalidierung und -charakterisierung

Bevor das erstellte Modell verlässlich zur Verarbeitung von Rohdaten und zur Optimierung des experimentellen Workflows eingesetzt werden kann, ist eine Charakterisierung und eine Validierung des Modells notwendig. Dabei wird das Modell getestet, um sein Verhalten zu analysieren. Bildet das Modell nicht alle, für die biologische Fragestellung relevante, Gegebenheiten realitätsgetreu ab, muss es überarbeitet werden.

Weiterhin ist von Interesse, ob das Modell eventuellen Limitierungen unterliegt. Solche Limitierungen sollten identifiziert werden, um eine reproduzierbare Datenverarbeitung zu gewährleisten. Eine mögliche Limitierung des zu Beginn eingeführten Beispiels ist der Linearitätsbereich der linearen Regression für die Umrechnung der Peakflächen in Konzentrationen.

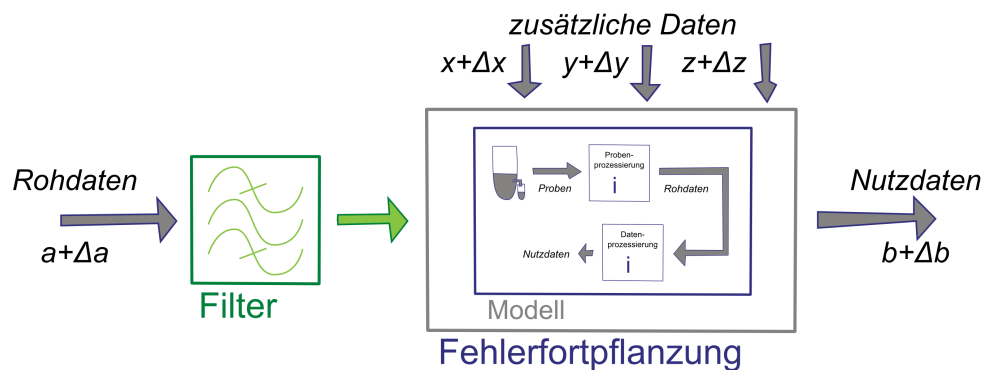


Abbildung 3.4: Vorgeschalteter Rohdatenfilter zur Gewährleistung reproduzierbarer Nutzdaten.

Das Ergebnis der Analysen legt fest, zwischen welchen Grenzwerten die Rohdaten liegen müssen, damit diese mit dem Modell verarbeitet werden können (Abb. 3.4). Liegen Daten z.B. außerhalb des Linearitätsbereiches der linearen Regression, würde dies einen generellen Ausschluss von der Verarbeitung bedeuten.

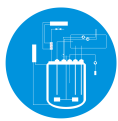
Eine weitere Anwendung für die Rohdatenfilterung liegt in der Qualitätssicherung. Theoretische Vorabanalysen mit dem vorliegenden Modell können die Rohdaten hinsichtlich ihrer Werte und/oder Genauigkeiten einschränken, um eine definierte Qualität der Nutzdaten zu gewährleisten.

3.2.4 Optimierung des experimentellen Workflows

Nach einer detaillierten Charakterisierung kann das erstellte Modell neben der Erzeugung reproduzierbarer Nutzdaten dafür eingesetzt werden um den experimentellen Workflow der Probenprozessierung zu analysieren und zu optimieren. Dafür können der Workflow oder die benötigten Rohdaten und/oder ihre Standardabweichungen künstlich (*in silico*) variiert und die Auswirkung dieser Veränderung auf die Nutzdaten analysiert werden. Vorteil dieser *in silico* Analyse ist vor allem eine zeitliche Ersparnis. Bevor eine Veränderung des Workflows mit hohem zeitlichen Aufwand direkt im Labor getestet wird, kann der Einfluss der Veränderung theoretisch abgeschätzt werden.

Kapitel 4

Schätzung extrazellulärer Raten unter Nutzung von Prozessmodellen



Extrazelluläre Raten, Stoffflüsse über die Zellmembran, stellen eine wichtige Eingangsgröße im Bereich der Flussanalyse dar. Neben dem Absolutwert dieser Größe hat auch die realistische Abschätzung ihrer Genauigkeit einen großen Einfluss auf das Ergebnis statistischer Analysen. Um dieses Wissen bereitstellen zu können, wird in diesem Kapitel ein Prozessmodell an mathematische Methoden der Schätzung von Parametern und deren Standardabweichungen gekoppelt. Die Funktionsweise dieser Kombination wird am Beispiel von Daten des Organismus *Gluconobacter oxydans* beschrieben. Die experimentellen Daten wurden von Tanja Hanke (AG Mikrobielle Physiologie, IBG-1) erhoben.

4.1 Einleitung

4.1.1 Motivation

Eine extrazelluläre Rate beschreibt den Fluss eines Stoffes, welchen ein Organismus mit seiner Umgebung austauscht, z.B. Substrat oder Produkt. Für die Analyse von Stoffflüssen in einem Organismus (ORTH et al., 2010), den intrazelluläre Raten, stellen extrazelluläre Raten eine notwendige Eingangsgröße dar. Neben den Raten selbst spielt auch ihre Qualität eine große Rolle, da sich diese direkt auf die Qualität der Schätzungen auswirkt.

Standardmäßig werden Regressionsanalysen zur Abschätzung der extrazellulären Raten eingesetzt (GLACKEN et al., 1988; GOUDAR, 2012) und in wenigen Ausnahmen werden diese sogar an eine detaillierte Fehlerberechnung gekoppelt (MURPHY und YOUNG, 2013). Vor allem für Fragestellungen, bei denen nur sehr wenige Informationen über das zugrundeliegende biologische System vorliegen, ist eine Ratenschätzung via Regressionsanalyse sinnvoll.

Die Methode stößt allerdings sehr schnell an ihre Grenzen, wenn mehrere Substrate gleichzeitig aufgenommen und verstoffwechselt werden oder biphasisches Wachstum vorliegt.

Eine Alternative zur Regressionsanalyse stellt ein biotechnologisches Prozessmodell dar. Ein solches Modell wird im Folgenden eingesetzt und an verschiedene statistische Methoden gekoppelt, um valide, extrazelluläre Raten und ihre Standardabweichungen abzuschätzen.

4.1.2 Anwendung *Gluconobacter oxydans*

Parallel zum methodischen Teil dieses Kapitels, werden die präsentierten Methoden direkt zur realistischen Schätzung der extrazellulären Raten von *Gluconobacter oxydans* 621H¹ (*G. oxydans*) für die ¹³C-Stoffflussanalyse (SAUER, 2006; WIECHERT, 2001) angewendet. Ziel dieser Stoffflussanalyse war es, den Glukose Katabolismus über einen partiell zyklischen Pentose-Phosphat Weg zu untersuchen. Detaillierte Hintergründe und die Ergebnisse sind in HANKE et al., 2013 veröffentlicht.

4.1.3 Symbolverzeichnis

Tabelle 4.1: Verzeichnis aller Symbole dieses Kapitels.

Symbol	Bezeichnung	Einheit
x	Biomasse	
GLC	Glukose	
GLK	Gluconat	
IIKGA	2-Ketogluconat	
VKGA	5-Ketogluconat	
μ	spezifische Wachstumsrate	[h ⁻¹]
π	spezifische Produktionsrate	[h ⁻¹]
r	spezifische Rate	[mmol g _{BTM} ⁻¹ h ⁻¹]
Y	Ausbeute	[-]
K	Sättigungskonstante	[g L ⁻¹]
K _i	Inhibierungskonstante	[g L ⁻¹]
μ_{max}, π_{max}	maximale Raten	[h ⁻¹]
δ	Standardabweichung	

¹*Gluconobacter oxydans* DSM 2343 (ATCC 621H) (PRUST et al., 2005)

4.2 Messung extrazellulärer Konzentrationen

Extrazelluläre Raten sind nicht direkt messbar, können jedoch aus der zeitlich aufgelösten Anreicherung bzw. Verringerung von extrazellulären Konzentrationen ermittelt werden.

Die experimentelle Bestimmung extrazellulärer Konzentrationen ist ein Standardverfahren in der Biotechnologie. Die Konzentrationsdaten für die betrachtete Anwendung *G. oxydans* wurden mittels HPLC (Gluconat, 5-Ketogluconat, 2-Ketogluconat) und enzymatischem Test (Glukose) erhoben. Nähere experimentelle Informationen können der Veröffentlichung (HANKKE et al., 2013) entnommen werden. Die experimentellen Daten befinden sich im Anhang A.2.

4.3 Schätzung extrazellulärer Raten

Die Bestimmung extrazellulärer Raten auf Basis extrazellulärer Konzentrationen und mit Hilfe von Prozessmodellen ist eine gängige Methode, welche in Abbildung 4.1 dargestellt ist. Im Folgenden werden die wesentlichen Schritte dieses Ablaufes vorgestellt.

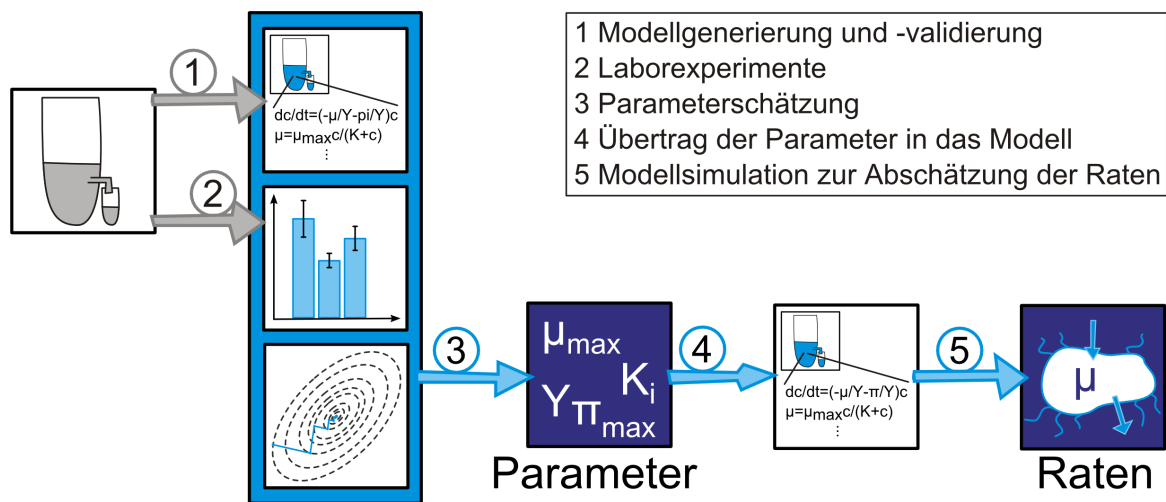


Abbildung 4.1: Schätzung extrazellulärer Raten auf Basis eines Prozessmodells: Nachdem Experimente durchgeführt und ein Prozessmodell aufgestellt wurden, kann durch den Einsatz einer Optimierungsmethode eine Parameterschätzung durchgeführt werden. Die ermittelten Parameter werden in das Modell übertragen und die extrazellulären Raten durch eine Modellsimulation abgeschätzt.

4.3.1 Aufstellen eines Prozessmodells

Ein biotechnologisches Prozessmodell (vgl. Abschnitt 2.2) beschreibt den Zusammenhang zwischen den Konzentrationen und den Raten durch ein Differential-algebraisches Gleichungssystem. Die extrazellulären Raten werden durch Kinetiken beschrieben und typische Modellparameter sind dabei die kinetischen Parameter maximale Rate, Ausbeute, Sättigungs-, Aktivierungs- und Inhibierungskonstante.

Die Grundlage zur realistischen Schätzung der extrazellulären Raten von *G. oxydans* bildet ein Modell, welches den Batch-Prozess im Bioreaktor beschreibt. Dafür wurde ein unstrukturierter Modellansatz gewählt. Das Modell wurde durch Dr. Stephan Noack bereitgestellt.

Das Modell beschreibt die zeitliche Änderung von fünf Zustandsgrößen durch sechs extrazelluläre Raten, welche über Monod-Kinetiken definiert sind, und enthält 21 kinetische Parameter. Das Modell stellt eine Abstraktion eines kleinen biologischen Systems dar, in welchem Glukose zu Gluconat, 2-Ketogluconat und 5-Ketogluconat umgesetzt wird. Biomasse kann aus Glukose und Gluconat gebildet werden, wobei allerdings das Wachstum auf Gluconat durch die Anwesenheit von Glukose inhibiert wird.

Eine schematische Darstellung des modellierten Systems ist in Abbildung 4.2 zu finden und die Modellgleichungen können dem Anhang A.1 entnommen werden.

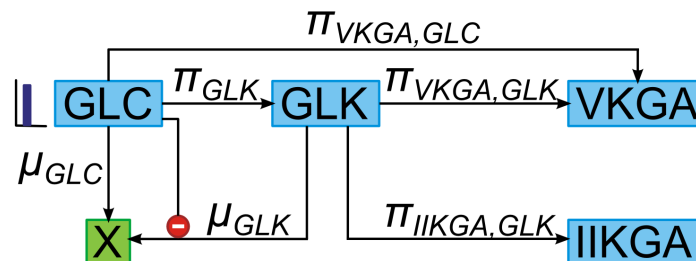


Abbildung 4.2: Netzwerkdarstellung des Systems für die Prozessmodellierung: Zu Beginn der Simulation wird die Konzentration von Glukose (GLC) auf 80 g L^{-1} gesetzt und die dynamische Verstoffwechslung durch die Wachstums- (μ) und Produktbildungsraten (π) beobachtet.

4.3.2 Schätzung der kinetischen Parameter

Auf Basis des Prozessmodells und unter Einsatz der experimentellen Daten und einer Optimierungsmethode können Werte für die kinetischen Parameter ermittelt werden, mit welchen

das Modell die gemessenen Metabolitkonzentrationen am besten beschreibt. Eine Gegenüberstellung der gemessenen und simulierten Konzentrationen ist in Abbildung 4.3 oben zu finden.

Tabelle 4.2 enthält die Ergebnisse der einfachen Parameterschätzung. Fünf kinetische Parameter, deren optimierte Werte bei multiplen Optimierungen der unteren Schranke sehr nah kamen, wurden für weitere Analysen auf Null gesetzt und somit für die folgenden Fehlerabschätzungen nicht weiter berücksichtigt. Dadurch werden die Raten μ_{GLK} (Umsetzung von Gluconat zu Bioamse) und $\pi_{VKGA,GLK}$ (Umsetzung von Gluconat zu 5-Ketogluconat) Null. Die Zahl der Kinetiken im Modell verringert sich auf vier und die Zahl der Parameter auf 16.

Tabelle 4.2: Mittels Parameteranpassung geschätzte kinetische Parameter inklusive Startwert für die Optimierung und den oberen und unteren Schranken. Die markierten Parameter werden in den folgenden Analysen auf Null gesetzt.

Parameter	Startwert	untere Schranke	obere Schranke	Ergebnis
$\mu_{max,GLC}$	3.30E-01	1.00E-08	1.00E+03	6.97E-01
$\mu_{max,GLK}$	1.00E-02	1.00E-08	1.00E+03	1.00E-08
$\pi_{max,GLK}$	5.00E+00	1.00E-08	1.00E+03	3.46E+01
$\pi_{max,IKGA,GLK}$	5.00E-01	1.00E-08	1.00E+03	3.34E+00
$\pi_{max,VKGA,GLK}$	1.00E-01	1.00E-08	1.00E+03	1.00E-08
$\pi_{max,VKGA,GLC}$	1.00E-01	1.00E-08	1.00E+03	3.98E+01
$K_{GLC,X}$	1.00E-02	1.00E-08	1.00E+03	4.68E+01
$K_{GLC,GLK}$	1.00E-02	1.00E-08	1.00E+03	6.15E+01
$K_{GLK,X}$	1.00E-02	1.00E-08	1.00E+03	1.61E-08
$K_{IKGA,GLK}$	1.00E-02	1.00E-08	1.00E+03	1.96E+01
$K_{VKGA,GLK}$	1.00E-02	1.00E-08	1.00E+03	1.00E-08
$K_{VKGA,GLC}$	1.00E-02	1.00E-08	1.00E+03	1.00E+03
$Y_{x,GLC}$	6.00E-01	1.00E-08	6.00E-01	6.00E-01
$Y_{x,GLK}$	6.00E-01	1.00E-08	6.00E-01	9.71E-05
$Y_{GLK,GLC}$	9.00E-01	1.00E-08	1.09E+00	1.00E+00
$Y_{IKGA,GLK}$	1.00E+00	1.00E-08	9.89E-01	9.89E-01
$Y_{VKGA,GLK}$	1.00E+00	1.00E-08	9.89E-01	1.32E-04
$Y_{VKGA,GLC}$	1.00E+00	1.00E-08	1.08E+00	9.48E-01
$k_{i,GLC}$	1.00E+00	1.00E-08	1.00E+03	2.32E-08
$c_{x,0}$	5.00E-02	1.00E-08	1.00E+03	4.19E-02
$c_{GLC,0}$	8.07E+01	7.50E+01	1.00E+03	8.08E+01

4.3.3 Simulation der extrazellulären Raten

Mit den ermittelten kinetischen Parametern führt eine Modellsimulation zu einer Abschätzung der extrazellulären Raten (Abbildung 4.3 unten und Tabelle 4.3). Für die späteren

Stoffflussanalysen wurden die beiden Szenarien zu den Zeitpunkten 8h (Phase 1) und 14.25h (Phase 2) ausgewählt, an welchen die Raten bestimmt werden sollten. In der ersten Phase findet hauptsächlich eine Umsetzung von Glukose zu Gluconat statt, während die zweite Phase von der Umsetzung von Gluconat zu 2-Ketogluconat bestimmt wird.

Tabelle 4.3: Durch Modellsimulation erzeugte extrazelluläre Raten (r in $[\text{mmol g}_{BTM}^{-1} \text{h}^{-1}]$) und μ in $[\text{h}^{-1}]$) zu zwei ausgewählten Zeitpunkten nach klassischer Parameteranpassung.

Rate	t_1	t_2
r_{GLC}	-7.22E+01	-3.57E-01
r_{GLK}	4.64E+01	-1.14E+01
r_{IIKGA}	1.15E+01	1.17E+01
r_{VKGA}	5.97E+00	2.10E-02
μ_{GLC}	2.72E-01	1.52E-03
μ_{GLK}	7.75E-18	2.27E-15
r_{CO_2}/CER	4.50E+00	7.14E+00

Neben ihrem Wert ist auch die Qualität der Raten von großem Interesse, da diese die Ergebnisse der Stoffflussanalysen beeinflusst und vor allem bei der Bestimmung der Anpassungsgüte des Modells und dem experimentellen Design eine wesentliche Rolle spielt. Die mathematische Schätzung der Standardabweichungen der extrazellulären Raten wird im Folgenden auf zwei Wegen realisiert, welche final miteinander verglichen werden.

4.4 Fehlerabschätzung mittels linearisierter Statistik

Die Standardabweichungen der extrazellulären Raten werden auf Basis des Prozessmodells mit zwei unterschiedlichen Ansätzen abgeschätzt. Der erste Ansatz ist ein gemischter Ansatz aus linearisierter Statistik und Monte-Carlo Simulation. Eine Kombination beider Methoden bietet eine schnelle, lokale Sensitivitätsanalyse der Parameter und berücksichtigt, durch die MC Simulation, die Nichtlinearitäten der zeitlichen Verläufe der extrazellulären Raten.

Bevor die einzelnen Schritte (vgl. Abb. 4.4) im Detail am Beispiel erläutert werden, seien sie im Folgenden kurz zusammengefasst und dabei auf die methodischen Kapitel verwiesen:

- 1. Abschätzung der Parameter-Standardabweichungen:** Berechnung der Sensitivitätsmatrix mittels Finite-Differenzen Approximation (vgl. Abschnitte 2.4 und 2.5). Gefolgt von linearisierter Statistik mittels Fisher-Informations-Matrix und Covarianzmatrix (vgl. Abschnitt 2.6).

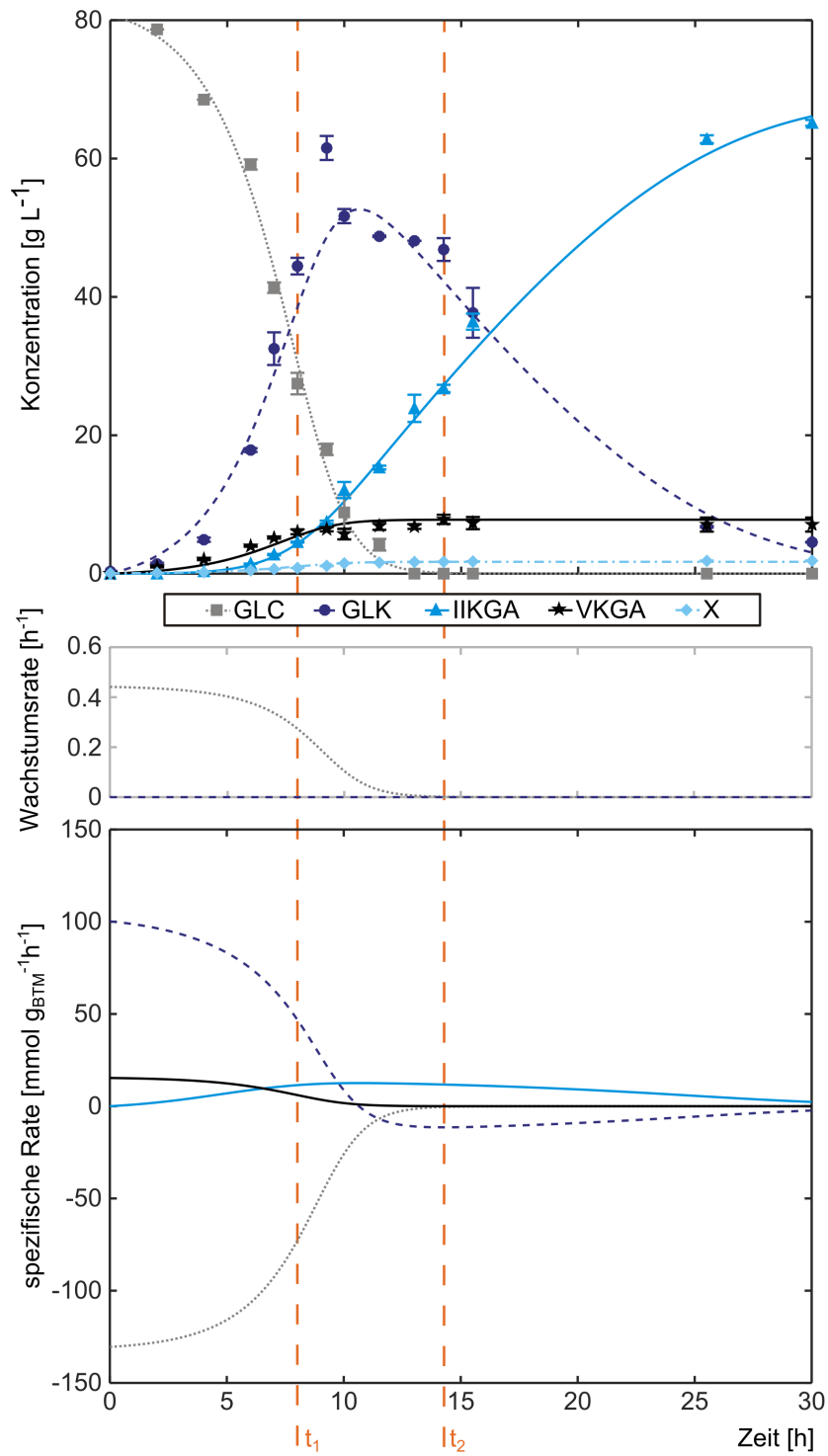


Abbildung 4.3: Zeitliche Verläufe der Konzentrationen simuliert (Linien) und gemessen (Symbole) und der durch Modellsimulation erzeugten Wachstums-, Verbrauchs- und Bildungsraten.

2. Bestimmung der Standardabweichung der extrazellulären Raten: Monte-Carlo Simulation mittels Verrauschen der kinetischen Parameter und multiplen Simulationen (vgl. Abschnitt 2.8).

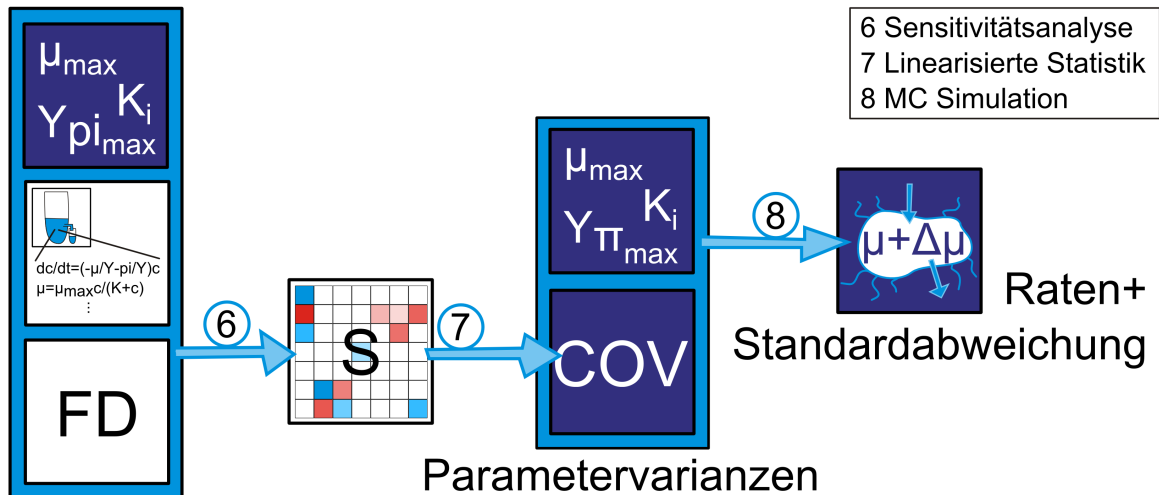


Abbildung 4.4: Ablauf der Bestimmung von extrazellulären Rate und ihrer Standardabweichungen auf Basis linearisierter Statistik: Die in Abbildung 4.1 ermittelten kinetischen Parameter, das Prozessmodell und die Methoden der Finiten-Differenzen Approximation werden genutzt um eine Sensitivitätsmatrix zu erstellen. Diese Sensitivitätsmatrix wiederum bildet die Grundlage der linearisierten Statistik, welche via Fisher-Informations Matrix zur Covarianzmatrix führt. Daraufhin werden die Parameter im Rahmen ihrer soeben ermittelten Varianzen verrauscht und mittels MC Simulation extrazelluläre Raten und ihre Standardabweichungen abgeschätzt.

4.4.1 Parameter-Standardabweichungen

Die zu ermittelnden Standardabweichungen der kinetischen Parameter hängen direkt von den experimentell bestimmten Standardabweichungen der Konzentrationsmessungen ab. Die lokale Sensitivitätsanalyse mittels FD untersucht den Einfluss jedes Parameters auf jede Messgröße. Die nachfolgende linearisierte Statistik kann dann Auskunft darüber geben, welcher Parameter durch die vorhandenen Messdaten wie genau bestimmt werden kann.

Ein bekanntes Problem bei der Bestimmung von Standardabweichungen kinetischer Parameter, welches auch bei der Anwendung *G. oxydans* auftrat, ist das der strukturellen Identifizierbarkeit (vgl. Abschnitt 2.6). Eine Berechnung der Covarianzmatrix, welche u.A. die Parametervarianzen enthält, ist nicht möglich.

Eine Möglichkeit, dieses Problem zu umgehen ist eine Minimierung der Korrelationen zwischen den betrachteten Parametern. Dafür werden einzelne Parameter konstant gehalten und

somit bei der Sensitivitätsanalyse nicht betrachtet. Der komplette Workflow aus Sensitivitätsanalyse und linearisierter Statistik wird mit dieser Einschränkung wiederholt. Die Ergebnisse für vier verschiedene Kombinationen konstant gehaltener Parameter sind in Tabelle 4.4 dargestellt.

Tabelle 4.4: Mittels linearisierter Statistik bestimmte Standardabweichungen der kinetischen Parameter: Die grau markierten Parameter, wurden für die jeweilige Berechnung der Varianzen (A-D) festgehalten, um Korrelationen zwischen Parametern zu minimieren und damit strukturelle Identifizierbarkeit zu erreichen.

Parameter	δ_A	δ_B	δ_C	δ_D	δ_{final}
$\mu_{max,GLC}$	0.25%	0.24%	0.22%	0.085%	0.085%
$\pi_{max,GLK}$	2.77%	2.77%	2.47%	0.66%	0.66%
$\pi_{max,IIKGA,GLK}$	0.88%	0.88%	0.30%	0.19%	0.19%
$\pi_{max,VKGA,GLC}$	60.21%	0.35%	60.20%	fest	60.20%
$K_{GLC,X}$	0.55%	0.54%	0.48%	0.19%	0.19%
$K_{GLC,GLK}$	4.05%	4.04%	3.45%	0.95%	0.95%
$K_{IIKGA,GLK}$	2.00%	2.00%	1.35%	0.47%	0.47%
$K_{VKGA,GLC}$	64.59%	fest	64.58%	fest	64.58%
$Y_{x,GLC}$	73.39%	73.29%	61.79%	fest	61.79%
$Y_{x,GLK}$	2308.02%	2307.96%	fest	fest	2308.02%
$Y_{GLK,GLC}$	1.39%	0.86%	1.39%	0.68%	0.68%
$Y_{IIKGA,GLK}$	1.36%	1.36%	0.70%	0.12%	0.12%
$Y_{VKGA,GLK}$	73275.83%	73273.81%	fest	fest	73275.83%
$Y_{VKGA,GLC}$	21.09%	20.14%	18.28%	5.89%	5.89%
$c_{x,0}$	0.081%	0.079%	0.072%	0.039%	0.039%
$c_{GLC,0}$	0.0012%	0.0012%	0.0012%	0.0012%	0.0012%

Die Ergebnisse zeigen, dass von den 16 zu bestimmenden Parametern 11 mit einer Standardabweichung von weniger als 6% bestimmt werden können. Drei Parameter können mit einer Standardabweichung von etwa 60% nur sehr ungenügend vorhergesagt werden. Diese drei Parameter ($\pi_{max,VKGA,GLC}$, $K_{VKGA,GLC}$ und $Y_{x,GLC}$) stehen alle direkt im Zusammenhang mit der Umsetzung von Glukose zu 5-Ketogluconat bzw. zu Biomasse.

Zwei der Ausbeute-Koeffizienten ($Y_{x,GLK}$, $Y_{VKGA,GLK}$) können mit einer ermittelten Standardabweichung von über 1000% nicht mit den vorliegenden Messdaten bestimmt werden. Der Grund dafür liegt bei den in Tabelle 4.2 auf Null gesetzten Parametern. Dies führt dazu, dass die Raten μ_{GLK} und $\pi_{VKGA,GLK}$ ebenfalls Null sind und die Ausbeute-Koeffizienten dementsprechend jeden beliebigen Wert annehmen können, ohne dass es einen Einfluss auf das System hat. Beim Szenario C sind genau diese beiden Parameter konstant gehalten. Es liefert also die am meisten realistische Abschätzung der anderen Parameter-Standardabweichungen.

4.4.2 Standardabweichung der extrazellulären Raten

Zur Schätzung der extrazellulären Raten und ihrer Standardabweichungen wurden auf Basis der geschätzten kinetischen Parameter (Tab. 4.2, rechte Spalte) und der finalen Standardabweichungen (Tab. 4.4, rechte Spalte) Monte-Carlo Simulationen durchgeführt. Dafür wurden für 1000 Simulationen die Parameter im Rahmen der für sie ermittelten Genauigkeit verrauscht und die Raten simuliert. Aus den daraus erhaltenen Ratenverteilungen, wurden, eine Normalverteilung annehmend, jeweils Mittelwerte und Standardabweichungen berechnet (Tab. 4.5).

Tabelle 4.5: Geschätzte extrazelluläre Raten (r in $[\text{mmol g}_{BTM}^{-1} \text{h}^{-1}]$ und μ in $[\text{h}^{-1}]$) und ihre Standardabweichungen bestimmt mit dem gemischten Ansatz aus linearisierter Statistik und MC Simulation.

Rate	t_1		t_2	
	Wert	δ	Wert	δ
r_{GLC}	-6.90E+01	29.65%	-5.27E-01	108.31%
r_{GLK}	4.93E+01	37.73%	-1.17E+01	10.07%
r_{IIKGA}	1.16E+01	9.87%	1.20E+01	6.45%
r_{VKGA}	3.70E-01	22.89%	1.25E-02	131.17%
μ_{GLC}	2.68E-01	30.93%	3.54E-03	138.60%
μ_{GLK}	2.12E-06	676.14%	1.14E-05	258.19%
r_{CO_2}/CER	3.98E+00	19.14%	7.44E+00	47.80%

Die extrazellulären Raten können mit dem gemischten Ansatz aus linearisierter Statistik und MC Simulation mit einer Standardabweichung von 10-40% zum ersten Zeitpunkt und mit einer Standardabweichung von 7-140% zum zweiten Zeitpunkt bestimmt werden. Eine Ausnahme bildet die Wachstumsrate μ_{GLK} für Wachstum auf Gluconat, die in beiden Fällen eine größere Standardabweichung aufweist. Die geringste Standardabweichung kann zu beiden Zeitpunkten bei der Bildungsrate von 2-Ketogluconat beobachtet werden.

4.5 Fehlerabschätzung mittels Monte-Carlo Bootstrapping

Dem gemischten Ansatz liegt die Annahme zu Grunde, dass sich die Parameter-Sensitivitäten über die lokale, lineare Methode der Finiten-Differenzen Approximation abschätzen lassen.

Ist diese Annahme falsch, kann dies zu falsch zu kleinen Parameterstandardabweichungen führen. Eine alternative, vollständig nichtlineare Abschätzung der Raten-Standardabweichungen liefert das Monte-Carlo Bootstrapping (vgl. Abschnitt 2.8, Abbildung 4.5). Die Nachteile des MC Bootstrapping sind ein hoher Zeit- und Rechenbedarf.

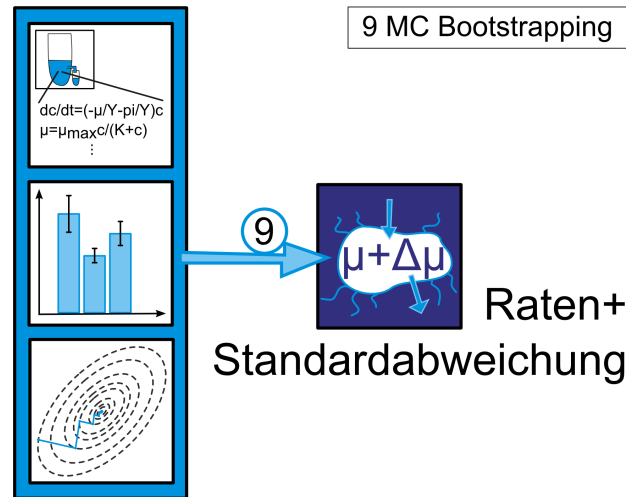


Abbildung 4.5: Ablauf der Bestimmung von extrazellulären Raten und ihrer Standardabweichungen auf Basis von Monte-Carlo Bootstrapping: Verrauschte experimentellen Daten werden im Rahmen des Monte-Carlo Bootstrapping zur wiederholten Parameterschätzung gefolgt von der Modellsimulation zur Erzeugung der extrazellulären Raten genutzt. Dafür sind neben den Daten das Prozessmodell und eine Optimierungsmethode notwendig.

Bei der Anwendung dieser Methode werden die experimentellen Daten im Rahmen ihrer Messgenauigkeit und dem angenommenen Messmodell folgend verrauscht und die extrazellulären Raten und ihre Standardabweichungen mit Hilfe des MC Bootstrapping und der darin integrierten Optimierung ermittelt. Die für die Anwendung *G. oxydans* geschätzten extrazellulären Raten sind in Tabelle 4.6 zu finden.

Tabelle 4.6: Durch MC Bootstrapping geschätzte extrazelluläre Raten (r in $[\text{mmol g}_{BTM}^{-1} \text{h}^{-1}]$) und μ in $[\text{h}^{-1}]$) und ihre Standardabweichungen.

Rate	t_1		t_2	
	Wert	δ	Wert	δ
r_{GLC}	-1.04E+02	70.59%	-2.69E-01	323.47%
r_{GLK}	7.93E+01	96.36%	-1.79E+01	85.33%
r_{IIKGA}	1.48E+01	69.43%	1.46E+01	75.33%
r_{VKGA}	8.28E+00	80.82%	1.33E-02	333.19%
μ_{GLC}	3.34E-01	62.85%	6.48E-03	466.08%
μ_{GLK}	6.24E-10	483.23%	8.62E-09	123.99%
$r_{CO_2/CER}$	6.66E+00	68.80%	8.34E+00	58.41%

Das Verrauschen der Messdaten vor der Optimierung erzeugt eine sehr große Streuung der geschätzten Parameter und damit auch der Raten. Daraus resultieren große Standardabweichungen von bis zu 500%. Dies bedeutet im Umkehrschluss, dass die Daten nicht oder nur sehr eingeschränkt für ^{13}C -Stoffflussanalysen eingesetzt werden können, da sich die hohen Messvarianzen auf die Varianzen der zu schätzenden Flüsse übertragen.

4.6 Zusammenfassung und Diskussion

In diesem Kapitel wurden verschiedene Ansätze zur Bestimmung von extrazellulären Raten und ihren Fehlern auf Basis extrazellulärer Konzentrationsmessungen vorgestellt. In den Tabellen 4.7 und 4.8 sind die Ergebnisse der drei Ansätze gegenübergestellt. Aufgeführt sind jeweils die geschätzten extrazellulären Raten zu den definierten Zeitpunkten in Phase 1 und 2 der Kultivierung und, wenn vorhanden, deren Standardabweichungen. Diese Ergebnisse zeigen deutliche Unterschiede sowohl in den Mittelwerten als auch in den Standardabweichungen.

Tabelle 4.7: Vergleich der geschätzten extrazellulären Raten (r in $[\text{mmol g}_{BTM}^{-1} \text{h}^{-1}]$) und μ in $[\text{h}^{-1}]$) in der ersten Phase (nach 8 Stunden). Gegenübergestellt sind die Raten, welche (1) via Simulation nach Parameterschätzung, (2) durch den Einsatz des kombinierten Ansatzes aus linearisierter Statistik und MC Simulation und (3) mittels MC Bootstrapping ermittelt wurden.

Rate	(1) Simuliert	(2) kombinierter Ansatz		(3) Monte-Carlo	
	Wert	Wert	δ	Wert	δ
r_{GLC}	-7.22E+01	-6.90E+01	29.65%	-1.04E+02	70.59%
r_{GLK}	4.64E+01	4.93E+01	37.73%	7.93E+01	96.36%
r_{IIKGA}	1.15E+01	1.16E+01	9.87%	1.48E+01	69.43%
r_{VKGA}	5.97E+00	3.70E-01	22.89%	8.28E+00	80.82%
μ_{GLC}	2.72E-01	2.68E-01	30.93%	3.34E-01	62.85%
μ_{GLK}	7.75E-18	2.12E-06	676.14%	6.24E-10	483.23%
r_{CO_2}/CER	4.50E+00	3.98E+00	19.14%	6.66E+00	68.80%

Die extrazellulären Raten, welche mittels MC Bootstrapping ermittelt wurden, sind im Durchschnitt 1.5x so groß wie die mit dem kombinierten Ansatz geschätzten Raten. Die Standardabweichungen sind im Durchschnitt sogar 3x (t_1) bis 5x (t_2) so groß. Eine Ausnahme bildet hierbei die GLK-gekoppelte Wachstumsrate, welche mittels MC Bootstrapping mit einer geringeren Standardabweichung bestimmt werden konnte.

Abschließend kann festgestellt werden, dass die mittels MC Bootstrapping geschätzten Raten eine zu große Standardabweichung aufweisen, um diese Ergebnisse für Stoffflussanaly-

Tabelle 4.8: Vergleich der geschätzten extrazellulären Raten (r in $[\text{mmol g}_{BTM}^{-1} \text{h}^{-1}]$ und μ in $[\text{h}^{-1}]$) in der zweiten Phase (nach 14.25 Stunden). Gegenübergestellt sind die Raten, welche (1) via Simulation nach Parameterschätzung, (2) durch den Einsatz des kombinierten Ansatzes aus linearisierter Statistik und MC Simulation und (3) mittels MC Bootstrapping ermittelt wurden.

Rate	Simuliert	lineare Statistik		Monte-Carlo	
	Wert	Wert	δ	Wert	δ
r_{GLC}	-3.57E-01	-5.27E-01	108.31%	-2.69E-01	323.47%
r_{GLK}	-1.14E+01	-1.17E+01	10.07%	-1.79E+01	85.33%
r_{IIKGA}	1.17E+01	1.20E+01	6.45%	1.46E+01	75.33%
r_{VKGA}	2.10E-02	1.25E-02	131.17%	1.33E-02	333.19%
μ_{GLC}	1.52E-03	3.54E-03	138.60%	6.48E-03	466.08%
μ_{GLK}	2.27E-15	1.14E-05	258.19%	8.62E-09	123.99%
r_{CO_2}/CER	7.14E+00	7.44E+00	47.80%	8.34E+00	58.41%

sen einsetzen zu können. Dadurch, dass bei diesem Ansatz die Messdaten verrauscht und auf deren Basis Parameterschätzung und Modellsimulationen zur Ermittlung der extrazellulären Raten durchgeführt werden, kommt es zu einer großen Streuung der Raten, welche in sehr großen Standardabweichungen münden.

Der alternative, kombinierte Ansatz beinhaltet die Annahme, dass sich die Parameter-Sensitivitäten in unmittelbarer Umgebung zum betrachteten Parameterwert linear verhalten, so dass die FD Approximation angewendet werden kann. Diese Annahme ist für das vorliegende System vertretbar und somit wurden in HANKE et al., 2013 die Ergebnisse des kombinierten Ansatzes als Eingangsgrößen für eine ^{13}C -Stoffflussanalyse eingesetzt.

Sowohl der kombinierte Ansatz als auch die reine MC Simulation liefert Abschätzungen der Standardabweichungen der extrazellulären Raten. Eine Möglichkeit, die Ergebnisse detailliert zu bewerten, stellt die Analyse der Rohdaten dar. Diese Analyse kann zeigen, welche zugrundeliegenden Annahmen eventuell verletzt sind und welche der Methoden zu bevorzugen ist.

Kapitel 5

Quantitative Auswertung von Wachstumskurven aus dem BioLector®



Dieses Kapitel beschäftigt sich mit der Gewinnung quantitativer Zeitverläufe der Biomasse und der automatischen Schätzung von Wachstumsraten aus Experimenten in dem Mikrofermentationssystem BioLector®. Bedingt durch das eingesetzte Messprinzip der Rückstreuung ist eine Kalibrierung notwendig. Diese Kalibrierung wird etabliert, im Rahmen der Fehlerfortpflanzung untersucht und in einen Workflow zur Datenauswertung vom Rohsignal aus dem BioLector® bis hin zur Wachstumsrate eingebettet. Das Projekt wurde in Kooperation mit Friedrich Käss (AG Bioprozesse und Bioanalytik, IBG-1) durchgeführt.

5.1 Einleitung

5.1.1 Das Mikrofermentationssystem BioLector®

Der BioLector® ist ein von m2p-labs¹ vertriebenes Mikrofermentationssystem (FUNKE et al., 2009). Es erlaubt Experimente in Volumina von 100 bis 2000 μL im Parallelansatz (48 oder 96 Batch-Fermentationen) mit Echtzeit-Überwachung weniger Prozessparameter und gewährleistet dadurch die Durchführung vieler Experimente in kurzer Zeit, sogenannte Hochdurchsatz-Experimente. Neben der Biomasse können im BioLector® die Prozessparameter pH-Wert, Sauerstoffkonzentration und Fluoreszenz erfasst werden, diese werden jedoch in diesem Kapitel nicht weiter betrachtet.

Experimente im BioLector® werden in Mikrotiterplatten durchgeführt (vgl. Abbildung 5.1). Diverse Parameter der Fermentation können separat in jeder der einzelnen Kavitäten oder für alle Kavitäten einer Platte eingestellt werden. Auf Grund dieser Eigenschaft ist ein Einsatz-

¹m2p-labs GmbH, Aachen, Deutschland

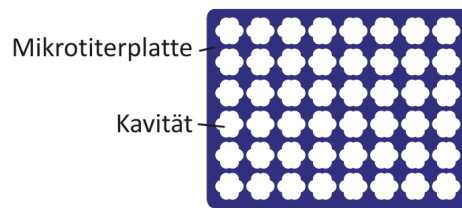


Abbildung 5.1: Mikrotiterplatte mit 48 Kavitäten. Auf Grund der Blütenform der Kavitäten werden diese Platten Flowerplates genannt. Durch die Schikanen können erhöhte Sauerstofftransferaten erzielt werden.

gebiet für den BioLector® der Vergleich eines Stammes bei unterschiedlichen Prozessparametern (gelöst Sauerstoff, vorgelegte Menge an Substrat, uvm.). Die häufigste Grundlage für diesen Vergleich ist die Wachstumsrate. Diese ergibt sich aus der Änderung der Menge an Zellen über der Zeit, welche wiederum aus dem Onlinesignal der Biomasse abgeleitet wird.

Im BioLector® erfolgt eine optische Messung der Biomasse über die Methode der Rückstreuung (SAMORSKI et al., 2005). Die Methode ist in Abbildung 5.2 dargestellt. Ein Lichtsignal wird von unten in eine Kavität gegeben. Das Licht wird an den Zellen in der Kavität gebrochen und reflektiert. Das Rückstreulicht, im Folgenden als Backscatter bezeichnet, wird gemessen. Das Backscatter Signal steht also in einem direkten Verhältnis zur Anzahl an Zellen in der Kavität.

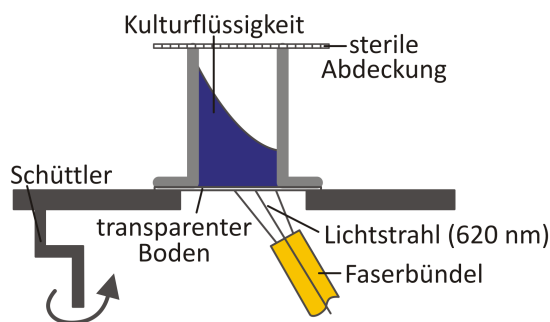


Abbildung 5.2: Messprinzip der Methode der Rückstreuung zur Bestimmung der Biomasse im BioLector® nach KENSY et al. (2009a).

5.1.2 Motivation

Zur Bestimmung von Wachstumsraten und zellspezifischen Substratverbrauchs- und Produktbildungsraten ist ein quantitatives Signal der Biomasse zwingend notwendig. Das oben erläuterte Funktionsprinzip der Rückstreuung birgt jedoch Herausforderungen bei der Aus-

wertung der Rohdaten aus dem BioLector®. Das Backscatter Signal hängt neben der Biomasse zusätzlich z.B. vom Füllvolumen in einer Kavität, dem eingesetzten Substrat oder dem Organismus ab. Außerdem beeinflussen optische Grenzen den Zusammenhang zwischen Backscatter Signal und Biomasse.

Bisher ist es allerdings üblich, die Zusammenhänge zwischen Backscatter Signal und Biomasse sowie zwischen Backscatter Signal und Füllstand als linear anzunehmen (KENSY et al., 2009a; KENSY et al., 2009b). Diese Annahme muss auf Basis der komplexen Zusammenhänge in der Kavität überprüft werden (vgl. Abbildung 5.3), um fehlerhaften Daten vorzubeugen.

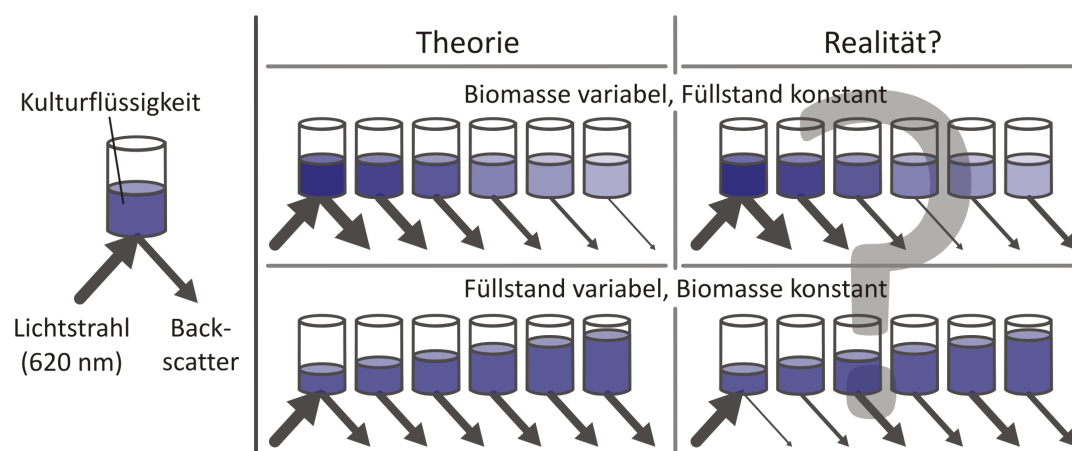


Abbildung 5.3: Herausforderung bei der Biomasse-Quantifizierung im BioLector®: Die Zusammenhänge zwischen Backscatter Signal und Biomasse sowie zwischen Backscatter Signal und Füllstand werden als linear bzw. konstant angenommen (vgl. Spalte „Theorie“). Wie verhält sich das Backscatter Signal in Abhängigkeit von Füllstand und Biomasse wirklich? Eine möglicher, nichtlinearer Zusammenhang ist in der Spalte „Realität“ dargestellt.

Im Folgenden wird die in Kapitel 3 eingeführte Methode der modellbasierten Rohdatenprozessierung angewendet. Die Methode führt zur Etablierung eines Werkzeugs, welches eine automatische Prozessierung der Biomasse-Daten aus dem BioLector® inklusive Fehlerfortpflanzung gewährleistet und somit u.A. die Nutzdaten in Form von Biotrockenmassen und Wachstumsraten und ihre Varianzen direkt für anschließende Analysen bereitstellt. Der Kern dieses Werkzeugs ist eine mit experimentellen Daten (KÄSS, 2015) erstellte Kalibrationsfunktion, welche das Backscatter Signal unter Berücksichtigung des Parameters „Füllstand“ in ein Biomassesignal überführt. Das Werkzeug wird außerdem genutzt, um das Backscatter Signal und damit auch die Einflüsse von Füllstand und Biomasse auf das Signal detailliert zu analysieren. Weiterhin findet eine Untersuchung der Varianzen von Roh- und Nutzsignal statt.

5.1.3 Symbolverzeichnis

Tabelle 5.1: Verzeichnis aller Symbole dieses Kapitels.

Symbol	Bezeichnung	Einheit
BS_{roh}	Backscatter Signal der Rohdaten	[a.u.]
BS_{blind}	Backscatter Signal der Blindprobe	[a.u.]
BS_{s1}	um Blindwert korrigierte Backscatter Daten	[a.u.]
BS_{s2}	gemitteltes Backscatter Signal aller Replikate	[a.u.]
N	Anzahl der Replikate pro Prozessbedingung	[-]
V	Füllvolumen einer Kavität	[μL]
KF	Kalibrierfunktion	
BTM	Biotrockenmasse	[g L^{-1}]
μ	Wachstumsrate	[h^{-1}]
μ_{max}	Maximale Wachstumsrate	[h^{-1}]

5.2 Modellierung der Prozessierung vom Backscatter Signal

Abbildung 5.4 gibt eine Übersicht über die Schritte der Datenprozessierung. Jeder dieser Schritte wird nachfolgend beschrieben und in mathematische Formeln übersetzt.

5.2.1 Datengrundlage

Bei Experimenten im BioLector® wird das Backscatter Signal, nach Zeitpunkt und Kavität aufgelöst, automatisiert abgespeichert. Das daraus resultierende Datenformat bildet die Grundlage der nachfolgenden Rohdaten-Prozessierung. Vor dem ersten Schritt der Datenprozessierung steht das Extrahieren notwendiger Daten. Eine Analyse der Probenprozessierung ist bei dieser Anwendung nicht notwendig, da die eingesetzten Rohdaten direkt durch Onlinemessungen erzeugt wurden.

Grundlage der nachfolgenden Betrachtungen sind Experimente, welche von Friedrich Käß durchgeführt wurden (KÄSS et al., 2014; KÄSS, 2015). Bei den Experimenten wurde der

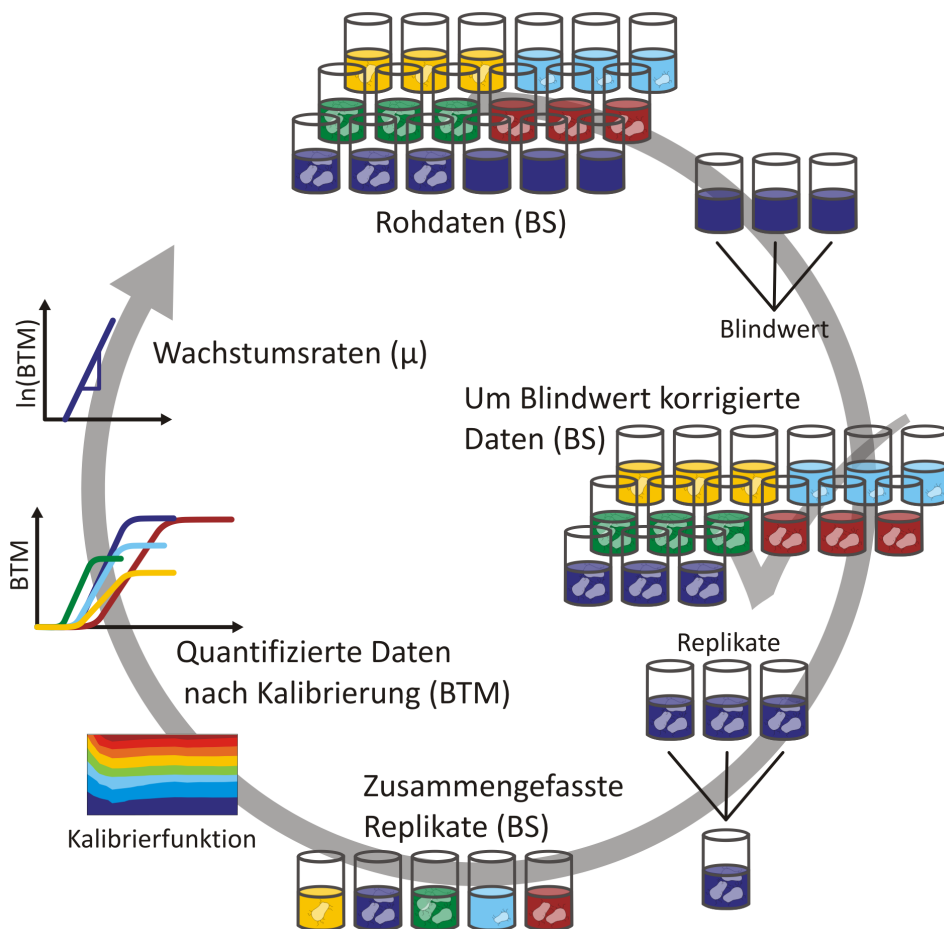
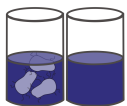


Abbildung 5.4: Workflow der Auswertung des Backscatter Signals aus dem BioLector®: Einlesen der Rohdaten verschiedener Szenarien (Farbe), Korrektur um den Wert der Blindprobe, Verarbeitung der Replikate jedes Szenarios, Kalibrierung und Quantifizierung der Biotrockenmassen, Wachstumsratenberechnung.

Stamm *Corynebacterium glutamicum* ATCC13032 (Wildtyp) und definiertes CgXII Minimalmedium (KEILHAUER et al., 1993) mit unterschiedlichen Glukose-Konzentrationen eingesetzt.

5.2.2 Blindwertkorrektur

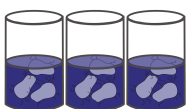


In einem ersten Bearbeitungsschritt werden die Backscatter Rohdaten um ihren Blindwert korrigiert. Der Blindwert einer Kavität ist das Backscatter Signal einer vergleichbaren Kavität, welche zwar Kulturflüssigkeit und Medium, jedoch keine Biomasse enthält. Diese Kavität wird Blindprobe genannt. Die Partikel im Medium beeinflussen den Wert der Rückstreuung. Das Backscatter Signal der Blindprobe (BS_{blind}) wird genutzt, um den Einfluss der Störgrößen zu bestimmen und alle weiteren Rohdaten (BS_{roh}) um diesen Wert zu korrigieren:

$$BS_{s1} = BS_{roh} - BS_{blind} \quad (5.1)$$

Wurde der Blindwert in mehreren Kavitäten gleichzeitig, d.h. als Replikat, bestimmt, fließt der Mittelwert aller Blindwerte in diese Berechnung ein.

5.2.3 Zusammenfassung der Replikate



Die um den Blindwert korrigierten Backscatter Daten (BS_{s1}) werden im nächsten Schritt in Bezug auf die Replikate verarbeitet. Das bedeutet, dass Kavitäten, welche unter exakt den selben Bedingungen und mit identischen Prozessparametern kultiviert wurden, zu einem Messwert zusammengefasst werden. Dies geschieht durch die Berechnung des Mittelwertes:

$$BS_{s2} = \frac{\sum_{i=1}^N BS_{s1,i}}{N} \quad (5.2)$$

N stellt dabei die Anzahl der Replikate dar, welche zu einem Wert (BS_{s2}) zusammengefasst werden.

Im nächsten Schritt werden die vorprozessierten Daten kalibriert.

5.2.4 Kalibrierung des Backscatter Signals

Um das optische Backscatter Signal aus dem BioLector® quantifizieren zu können, ist eine Kalibrierung notwendig. Dafür muss zuerst der Zusammenhang zwischen den Größen Füllstand, Biomasse und Backscatter Signal aufgelöst werden. Um dies zu erreichen wurde das Backscatter Signal in verschiedenen Experimenten für definierte Kombinationen aus Füllstand (300 bis 1500 μL) und Biomassekonzentration (0 bis 18 g L^{-1}) aufgenommen. Abbildung 5.5 enthält die Messdaten (weiße Kreise), welche ausführlich in Anhang B.1 zu finden sind.

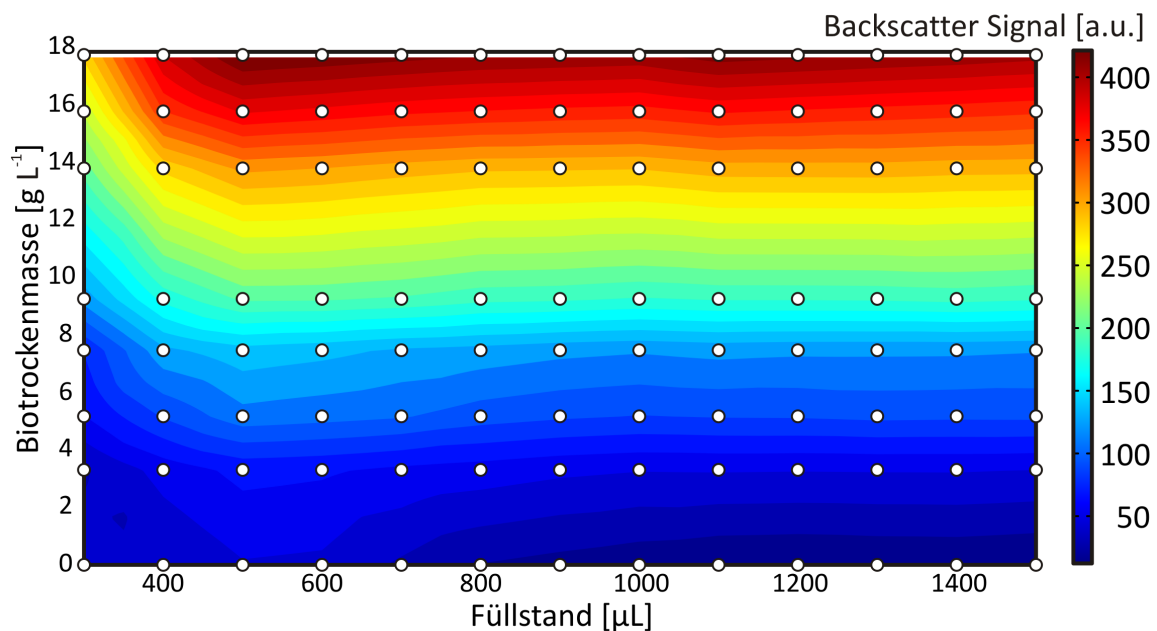


Abbildung 5.5: Kalibrierfunktion zur Quantifizierung des optischen Backscatter Signals aus dem BioLector®: Die weißen Kreise stehen für Messdaten aus Experimenten, welche die Füllstände von 300-1500 μL und eine Biomassekonzentration von 0 bis 18 g L^{-1} abdecken. Zwischen diesen vorliegenden Messungen wurde die Funktion des Backscatter Signals stückweise quadratisch interpoliert.

Die experimentellen Daten stellen die Basis der Schätzung der Parameter einer Kalibrierfunktion dar. Die Parameter der quadratischen Funktion der Form $BS = a_1 \cdot BTM^2 + a_2 \cdot BTM + a_3$ wurden mit Hilfe der MATLAB-Funktion 'polyfit' eingebettet in eine Monte-Carlo Simulation ermittelt. Die Funktionsparameter für verschiedene Füllvolumina und die dazugehörigen Standardabweichungen sind in Tabelle 5.2 aufgeführt.

Die Funktionsparameter (a_1 - a_3) können genutzt werden, um Backscatter Signale in Biotrockenmassen umzurechnen. Dabei kommt Gleichung 5.3 zum Einsatz. Außerdem ergeben

Tabelle 5.2: Parameter der quadratischen Kalibrierfunktionen: Die experimentellen Daten aus Tabelle B.1 wurden zur Erstellung einer quadratischen Kalibrierfunktion ($BS = a_1 \cdot BTM^2 + a_2 \cdot BTM + a_3$) eingesetzt. Die Kalibrierfunktion kann zur Berechnung von Biotrockenmasse (BTM) aus den mit dem BioLector® erhobenen Backscatter Signalen (BS) eingesetzt werden.

V [μ L]	a_1	Δa_1	a_2	Δa_2	a_3	Δa_3
300	0.6414	0.1230	3.1875	2.0414	32.8562	6.6411
400	0.6584	0.1127	9.1074	1.6568	31.5548	3.9803
500	0.7442	0.1353	9.0039	2.0753	45.1352	6.6860
600	0.7368	0.1339	8.9718	1.9344	42.9908	4.9666
700	0.6755	0.1392	10.2983	1.9652	31.8464	4.5905
800	0.6375	0.1397	11.2432	1.9492	21.6489	4.4273
900	0.6348	0.1444	11.4765	2.0101	16.7963	4.3336
1000	0.6257	0.1413	11.7312	1.9217	12.9647	3.8706
1100	0.6274	0.1441	12.3962	1.9331	10.3078	3.6912
1200	0.6026	0.1396	12.7155	1.8824	9.2394	3.4725
1300	0.5762	0.1291	12.9243	1.7398	9.9806	2.9035
1400	0.5539	0.1277	13.1240	1.7497	10.0169	3.2644
1500	0.5320	0.1155	13.3854	1.5695	8.2537	2.5101

sich die dazugehörigen Standardabweichungen der Biotrockenmassen mittels Gauß'scher Fehlerfortpflanzung (Gleichung 5.4).

$$BTM = \frac{\sqrt{BS + \frac{a_2^2}{a_1} - a_3 - \frac{a_2}{\sqrt{a_1}}}}{\sqrt{a_1}} \quad (5.3)$$

$$\Delta BTM \approx \sqrt{\left(\frac{\partial BTM}{\partial a_1} \cdot \Delta a_1\right)^2 + \left(\frac{\partial BTM}{\partial a_2} \cdot \Delta a_2\right)^2 + \left(\frac{\partial BTM}{\partial a_3} \cdot \Delta a_3\right)^2} \quad (5.4)$$

Der ermittelte Zusammenhang zwischen Füllstand, Biomasse und Backscatter wird im Folgenden als Kalibrierfunktion bezeichnet und bildet die Grundlage für die Kalibrierung aller nachfolgenden experimentellen Daten.

$$BTM = KF(BS_{s2}, V, a_1, a_2, a_3) \quad (5.5)$$

Liegen nun Backscatter Werte aus einem BioLector®-Experiment vor, können diese direkt unter Einsatz der Messung des Füllstandes (V) und der Kalibrierfunktion (KF) inklusive ihrer Funktionsparameter (a_1 , a_2 und a_3) in Biotrockenmasse (BTM) umgerechnet werden.

Die Kalibrierung ist für Daten aus Experimenten mit *C. glutamicum* und bei Einsatz des CgXII-Mediums oder eines ähnlichen Mediums ohne Weiteres anwendbar. Für andere Orga-

nismen oder veränderte Medien, bei denen die Änderung einen Einfluss auf die Trübung des Mediums hat, sollte eine neue Kalibrierung durchgeführt werden, da dies Auswirkungen auf das Backscatter Signal haben kann.

5.2.5 Bestimmung der Wachstumsrate

Eine Wachstumsrate beschreibt die durchschnittliche relative Zunahme der Zellzahl pro Zeiteinheit. In der Regel wird exponentielles Wachstum der Zellen angenommen, wenn diese ohne jegliche Limitierung wachsen. Die ideale Wachstumskurve einer Population in einem Batchansatz, wie er in jeder Kavität einer Mikrotiterplatte theoretisch abläuft, ist in Abbildung 5.6 dargestellt und gliedert sich in vier Phasen. Dabei ist vor allem die exponentielle Phase für die nachfolgenden Untersuchungen relevant. Gerade für die automatische Bestimmung der Wachstumsrate aus einer solchen Kurve sind jedoch auch die anderen Phasen nicht zu vernachlässigen.

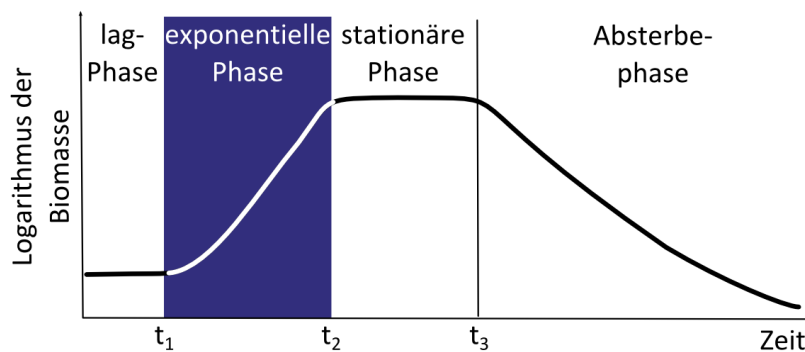


Abbildung 5.6: Ideale Wachstumskurve einer Population im Batchansatz und die vier Phasen, in welche sich der Verlauf einteilen lässt.

Die Wachstumsrate wird bestimmt, indem die Biotrockenmasse logarithmiert und daraufhin stückweise durch lineare Funktionen beschrieben wird (Abb. 5.7 D). Der Anstieg dieser Funktionen (Abb. 5.7 E) stellt dann die für das jeweilige Stück ermittelte Wachstumsrate dar.

$$\mu_n = \frac{\ln(BTM(t_n)) - \ln(BTM(t_{n-1}))}{t_n - t_{n-1}} \quad (5.6)$$

$$\mu_{max} = \max(\mu_n) \quad (5.7)$$

Die Herausforderung in der (semi-)automatischen Bestimmung der Wachstumsrate liegt darin, die Zeitpunkte, welche die einzelnen Phasen voneinander abgrenzen (vgl. t_1 und t_2 in

Abbildung 5.6), möglichst genau zu bestimmen. Des Weiteren sollten der Berechnung jeder einzelnen Wachstumsrate ausreichend viele Datenpunkte zugrunde gelegt werden.

Erschwert wird die automatische Bestimmung der Wachstumsrate, wenn die Wachstumskurve nicht ideal verläuft. Ein möglicher Grund dafür ist der parallele Einsatz unterschiedlicher Substrate, wenn dieser zu biphasischem Wachstum führt (vgl. Kapitel 4).

5.2.6 Überblick der Prozessierungsschritte

Werden alle beschriebenen Prozessierungsschritte auf die Rohdaten angewendet, so verändern sich das Backscatter Signal wie in Abbildung 5.7 schematisch für ein Beispiel aufgetragen. Aus dem Backscatter Signal können so die Biotrockenmasse und die maximale Wachstumsrate jeweils mit ihren Standardabweichungen ermittelt werden.

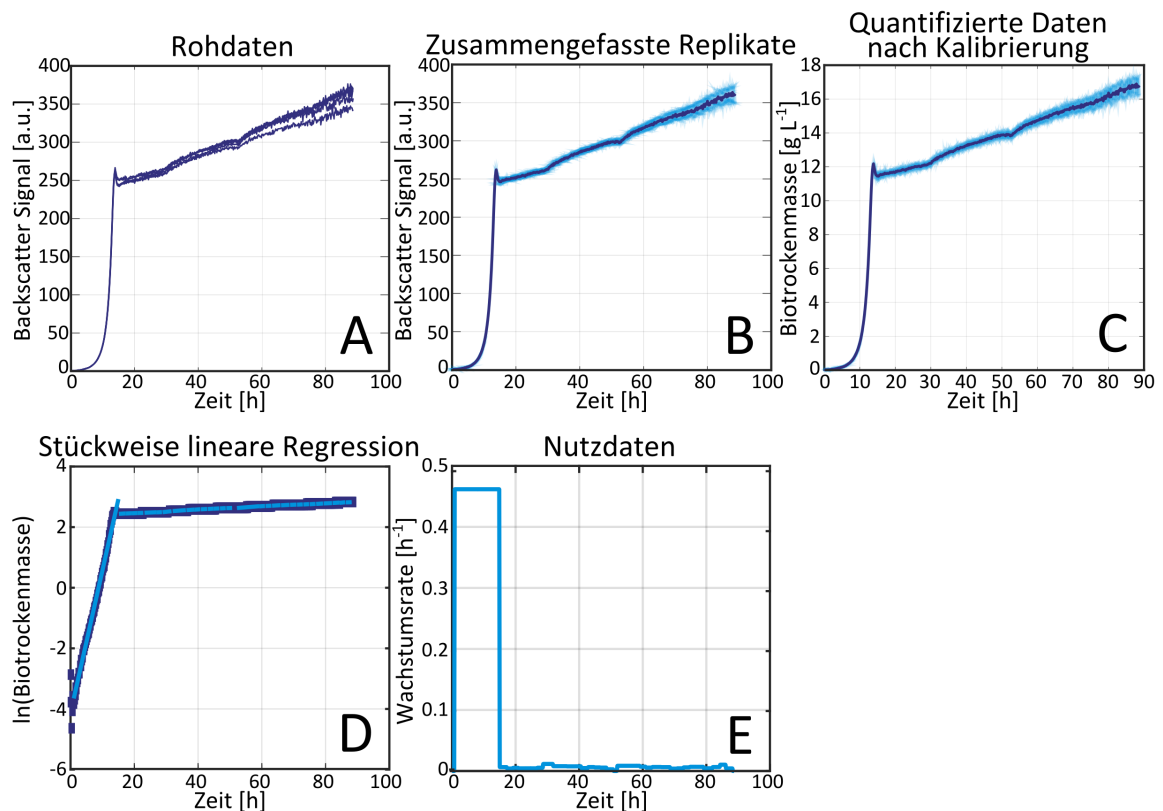


Abbildung 5.7: Übersicht der Prozessierungsschritte vom Backscatter Signal (A, Rohdatum) zur Biotrockenmasse (C, Nutzdatum) und zur maximalen Wachstumsrate (E, Nutzdatum).

5.2.7 MATLAB-Werkzeug

Die im Vorfeld detailliert beschriebene, modellbasierte Methode zur Auswertung von Backscatter Signalen aus dem BioLector® wurde in einem MATLAB-Werkzeug umgesetzt. Das Werkzeug bietet die folgenden Einsatz- und Einstellmöglichkeiten:

1. Festlegen der Anzahl unterschiedlicher Datensätze.
2. Zuordnung der Kavitäten zu den einzelnen Datensätzen.
3. Identifizierung der Kavitäten, welche zur Blindwertkorrektur herangezogen werden oder Eingabe spezifischer Blindwerte.
4. Auswahl der Füllvolumina der Kavitäten eines Datensatz.
5. Festlegung der durchzuführenden Bearbeitungsschritte:
 - Ausschließlich Prozessierung des Backscatter Signals und der Replikate.
 - Einsatz der vordefinierten, quadratischen Kalibrierfunktion, einer vordefinierten linearen Kalibrierfunktion (vgl. Anhang B.2) oder einer eigenen linearen Kalibrierung durch manuelle Eingabe der Parameter und Standardabweichungen für Anstieg und Achsenabschnitt.
 - Abschätzung der Wachstumsraten nach Definition einer minimalen Anzahl an Datenpunkten pro Regression (Standardwert: 10).
 - Visualisierung.

Die Benutzeroberfläche des Werkzeugs ist in Abbildung 5.8 zu sehen und die komplette Dokumentation kann dem Anhang B.3 entnommen werden.

5.3 Modellcharakterisierung

5.3.1 Betrachtung der Einflussparameter Biomasse und Füllstand

Bisher ist die gängige Annahme bei der Bestimmung der Biotrockenmasse aus dem Backscatter Signal von BioLector®-Experimenten die Linearität. Auf Basis der Kalibrationsexperimente soll diese Annahme nun im Detail überprüft werden.

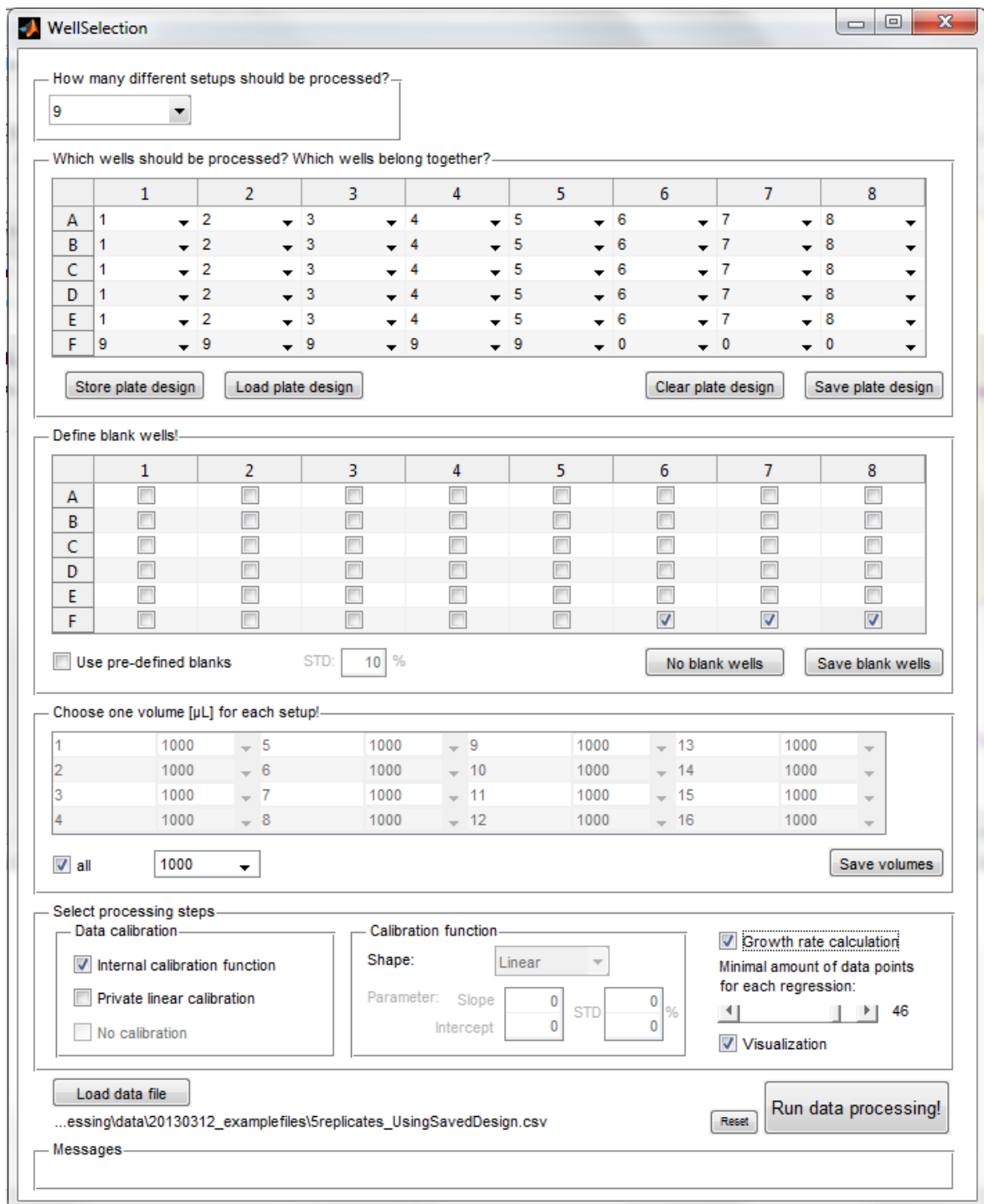


Abbildung 5.8: Benutzeroberfläche der entwickelten MATLAB-Anwendung zur Auswertung von BioLector®-Daten.

Abbildung 5.9 zeigt den Zusammenhang zwischen Biotrockenmasse und Backscatter Signal für verschiedene Füllstände. Die Annahme der Linearität, angedeutet durch die gestrichelte Linie, kann für Füllstände größer als 300 μL und Biotrockenmassen größer als 8 g L^{-1} , was einem Backscatter Signal von mindestens 200 a.u. entspricht, bestätigt werden. Der Zusammenhang zwischen Backscatter Signal und Biotrockenmassen weicht jedoch vor allem bei geringen Füllständen und geringen Biotrockenmassen von der angenommenen Linearität ab.

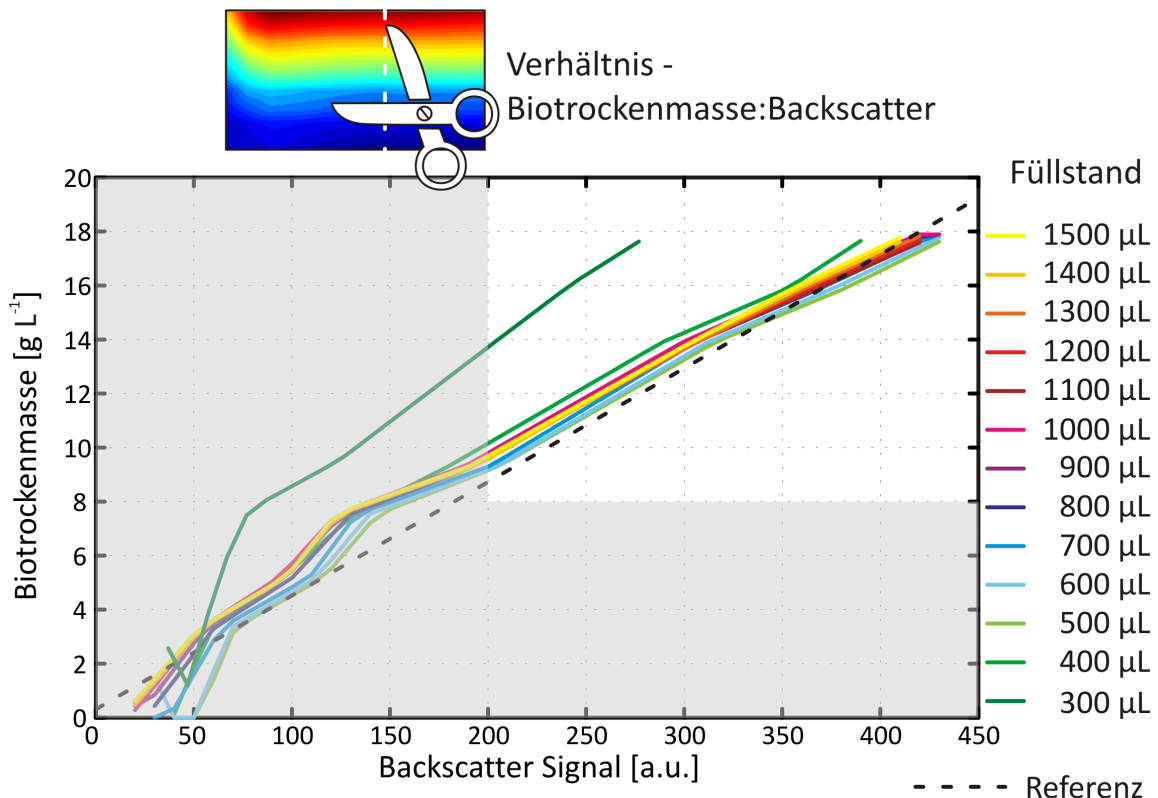


Abbildung 5.9: Zusammenhang zwischen Biotrockenmasse und Backscatter Signal für verschiedene Füllstände. Der Vergleich mit der Annahme eines linearen Zusammenhanges (gestrichelte Linie) führt zum grauen Bereich, in welchem diese Annahme nicht gerechtfertigt ist.

In Abbildung 5.10 ist der Einfluss des Füllstandes auf das Backscatter Signal für verschiedenen Biotrockenmassen aufgelöst. Es wurde angenommen, dass die Füllvolumina keinen Einfluss auf den Zusammenhang zwischen Backscatter Signal und Biotrockenmasse haben. Um diese Annahme zu bestätigen, müsste die Abbildung ausschließlich zur x-Achse parallele Geraden (vgl. gestrichelte Linie für eine Biotrockenmasse von 10 g L^{-1}) enthalten. Die Annahme kann jedoch nur für einen eingeschränkten Bereich mit Backscatter Signal zwischen 75 und 350 a.u. und Füllstand größer $500 \mu\text{L}$ bestätigt werden.

Zusammenfassend kann festgestellt werden, dass geringe Volumina oder eine geringe Bio-

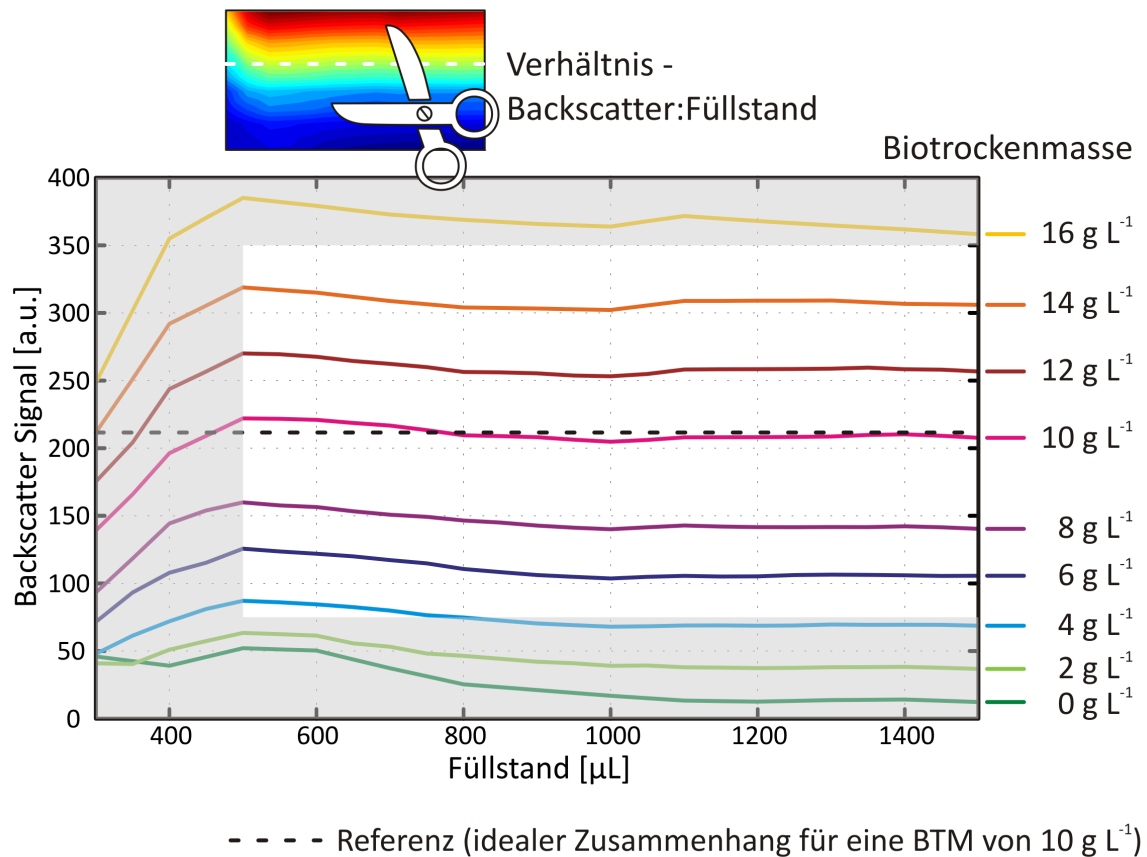


Abbildung 5.10: Einfluss des Füllstandes auf das Backscatter Signal für verschiedene konstante Biotrockenmassen. Der Vergleich mit der Annahme, dass der Füllstand keinen Einfluss auf das Backscatter Signal hat (gestrichelte Linie) führt zum grauen Bereich, in welchem diese Annahme nicht gerechtfertigt ist.

trockenmasse dazu führen, dass sich das Backscatter Signal nicht linear zur tatsächlich vorhandenen Biotrockenmasse verhält. Der Einsatz einer nichtlinearen Kalibrierfunktion ist zur validen Datenauswertung unumgänglich. Der Grund für die Nichtlinearitäten liegt vermutlich im Messprinzip der Rückstreuung.

Dass sehr hohe Biotrockenmassen auf Grund eines Sättigungseffektes zu einem nichtlinearen Zusammenhang zwischen Biotrockenmasse und Backscatter Signal führen können, wurde bereits in KENSY et al. (2009b) beschrieben, konnte jedoch in den unseren Analysen zugrundeliegenden Experimenten nicht beobachtet werden, da sich die gemessenen Backscatter Signale weit unterhalb der kritischen Werte befanden.

5.3.2 Untersuchung der Messgenauigkeiten

Die experimentellen Daten der Kalibrierexperimente wurden außerdem für die Analyse der Messgenauigkeit der Backscatter Signale herangezogen. Die Ergebnisse dieser Untersuchungen fließen ebenfalls in die Bewertung des Backscatter Signales hinsichtlich Nutzbarkeit für Dateninterpretation und weitere Analysen ein.

Abbildung 5.11 zeigt die ermittelten relativen Standardabweichungen zugehörig zu den Backscatter Signalen aus Abbildung 5.5. Hier ist zu sehen, dass hohe relative Standardabweichungen von über 10% vor allem bei geringen Konzentrationen der Biotrockenmasse von unter 2 g L^{-1} auftreten.

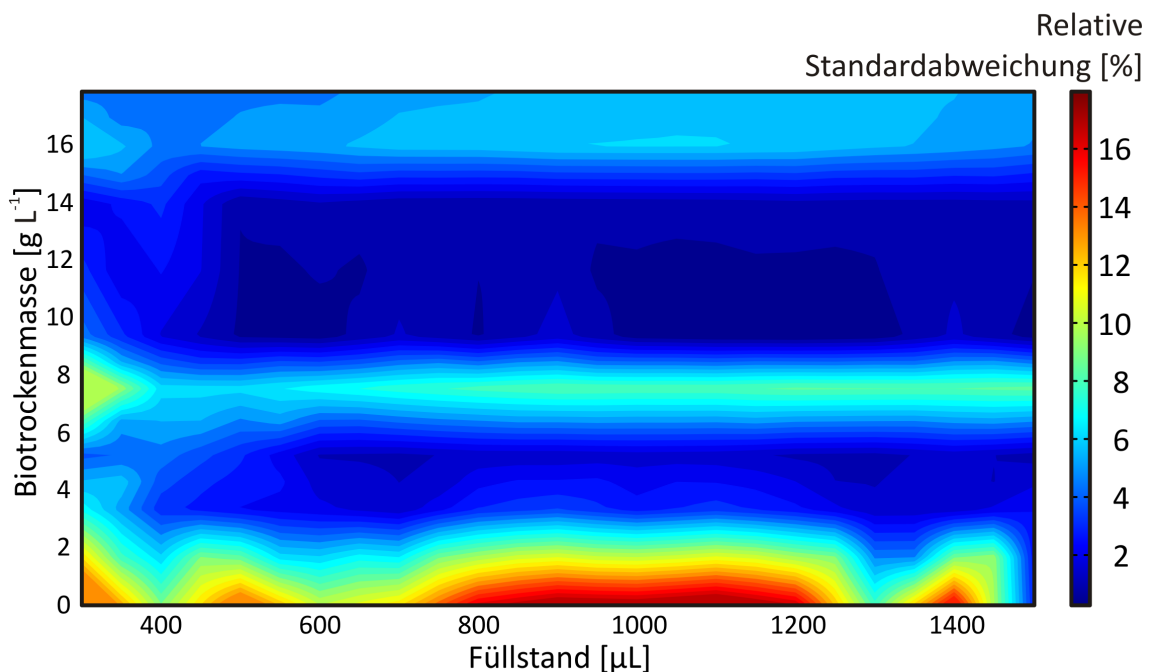


Abbildung 5.11: Relative Standardabweichungen der Backscatter Signale in Abhängigkeit von Füllstand und Biomasse zugehörig zu den Backscatter Signalen aus Abbildung 5.5.

Die detaillierte Ansicht aus Abbildung 5.11 wird in Abbildung 5.12 abstrahiert dargestellt, um die Bereiche signifikant erhöhter Standardabweichungen besser identifizieren zu können.

5.4 Zusammenfassung und Diskussion

Es wurde eine modellbasierte Methode entwickelt, welche zur automatischen, Auswertung von Backscatter Signalen aus dem BioLector® eingesetzt werden kann. Diese Methode wur-

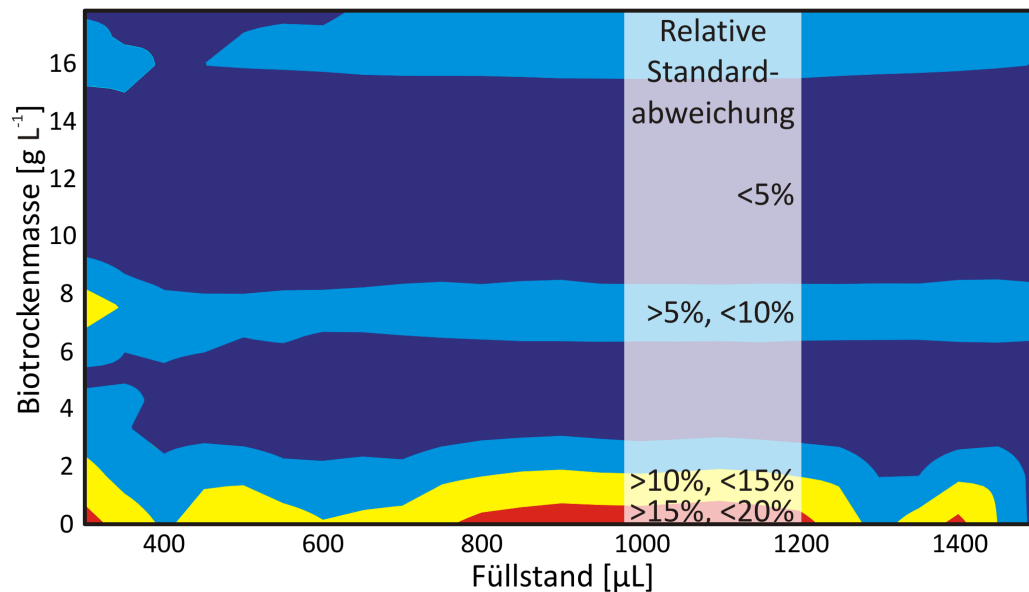


Abbildung 5.12: Abstrahierte Darstellung der relativen Standardabweichungen der Backscatter Signale aus Abbildung 5.11: Zur besseren Übersicht wurden die relativen Standardabweichungen in Bereiche von <5%, <10%, <15% und >15% eingeteilt und farblich getrennt.

de in einem MATLAB-Werkzeug umgesetzt. Das Werkzeug stellt valide Verläufe der Biotrockenmasse und Wachstumraten mit Standardabweichungen zur Verfügung und fand bereits in verschiedenen Projekten Anwendung (VOGES und NOACK, 2012; KÄSS et al., 2014).

Auf Basis experimenteller Daten von *C. glutamicum* wurde eine Kalibrierfunktion erstellt und in das Werkzeug integriert. Die Funktion kann für Experimente mit *C. glutamicum* und Medien ähnlich dem CgXII-Medium direkt eingesetzt werden. Für andere Organismen oder veränderte Medien müssen einmalig neue experimentelle Daten zur Erstellung einer Kalibrierfunktion erhoben werden. Diese kann dann einfach im Werkzeug etabliert werden. Danach kann die präsentierte Methode auch zur Charakterisierung dieser Organismen eingesetzt werden.

Wird das Werkzeug mit der aktuellen Kalibrierfunktion zur Auswertung von Backscatter Signalen eingesetzt, kann für den Großteil der Daten eine Standardabweichung von unter 10% erwartet werden.

Es konnte gezeigt werden, dass der Einsatz einer Kalibrierfunktion für das analysierte System vor allem im Bereich geringer Biotrockenmassen und Füllstände notwendig ist, da sich in diesem Bereich Backscatter Signal und Biotrockenmasse nicht linear verhalten.

Kapitel 6

Modellierung und Analyse der Prozessierung von Metabolom-Daten



Die Quantifizierung intrazellulärer Metabolite ist ein zentrales Thema der Systembiologie, denn eine Quantifizierung bietet die einzige Möglichkeit, Daten aus verschiedenen Protokollen miteinander zu vergleichen. Im Mittelpunkt dieses Kapitels steht die Einführung einer Methode, welche die modellbasierte Quantifizierung von Metabolomdaten ermöglicht. Diese Methode zeichnet sich durch eine Modellierung aller Schritte der Probenprozessierung und eine konsistente Betrachtung aller Messfehler aus. Am Beispiel eines Metabolom-Datensatzes von *Corynebacterium glutamicum* erfolgt eine detaillierte Untersuchung des experimentellen Ablaufes. Dadurch können sowohl kritische Schritte in Bezug auf Datenqualität als auch Verbesserungspotentiale aufgedeckt werden. Das Projekt wurde in Kooperation mit Nicole Paczia (AG Bioprozesse und Bioanalytik, IBG-1) durchgeführt.

6.1 Einleitung

6.1.1 Motivation

Bei der Untersuchung des Metaboloms eines Organismus weisen die zu analysierenden Metabolite große Unterschiede in ihren chemischen Eigenschaften auf und ihre Konzentrationen variieren von Pikomol bis Millimol (OLDIGES et al., 2007). Auf Grund dieser Komplexität der Proben birgt die Metabolomanalyse eine ganze Reihe von Herausforderungen und es sind verschiedene Protokolle zur Quantifizierung der Metabolitkonzentrationen verfügbar (VILLAS-BÔAS et al., 2005), welche versuchen sich diesen Herausforderungen zu stellen.

Um verlässliche Daten zu erhalten ist das mehrfache Wiederholen eines Experimentes mit all seinen Probenprozessierungsschritten, das Generieren sogenannter biologischer Replika-te, notwendig. Dieses Vorgehen ist oftmals zu aufwendig oder aus biologischen Gründen nicht möglich. Folglich werden oft nur ausgewählte Teile eines Experimentes wiederholt (technische Replikate), um daraus eine grobe Abschätzung der Genauigkeit der erhobenen Messdaten abzuleiten.

Präzise Metabolomdaten sind die Grundlage vieler Analysemethoden und bilden damit die Basis fundierter Erkenntnisse und Hypothesen. Die Genauigkeit der Daten hängt von vielen biologischen und biotechnologischen Faktoren u.A. während der Probenprozessierung ab (NOACK und WIECHERT, 2014). Eine detaillierte Beschreibung der Probenprozessierung und der dabei auftretenden Fehlerquellen gewährleistet eine Prozessierung der Rohdaten, welche (1) nach aktuellem Wissensstand präzise Nutzdaten mit (2) validen Standardabweichungen bereitstellt. Darüber hinaus bietet nur eine solche Beschreibung die Möglichkeit, die Probenprozessierung systematisch mit mathematischen Methoden zu analysieren und zu optimieren.

In diesem Kapitel wird eine Methode vorgestellt, welche die Berechnung der Fehlerfortpflanzung für die Prozessierung der Metabolomdaten erlaubt. Zentrales Element der Methode ist ein Modell des kompletten Ablaufes der Probenprozessierung zur Quantifizierung von intrazellulären Metaboliten. In dieser Arbeit lag der Fokus auf einem Protokoll, welches am Institut für Bio- und Geowissenschaften (IBG-1) des Forschungszentrum Jülich eingesetzt wird. Für dieses Protokoll wurden alle aktuell bekannten und experimentell verfügbaren Einflussfaktoren, welche die Genauigkeit der intrazellulären Metabolitkonzentration beeinflussen, bestimmt (PACZIA, 2012). Diese Informationen flossen direkt in die Modellierungsarbeit.

6.1.2 Symbolverzeichnis

Tabelle 6.1: Verzeichnis aller Symbole dieses Kapitels.

Symbol	Bezeichnung	Einheit
c_{CDW}	Biotrockenmasse	[g L ⁻¹]
ν_{CDW}	Biotrockenmassespezifisches Biovolumen	[μ L mg ⁻¹]
ν_{brs}	Volumenspezifisches Biovolumen	[μ L mL ⁻¹]
V_{brs}	Volumen der Probe aus dem Bioreaktor	[μ L]
V_{cyt}	Zytosolisches Volumen	[μ L]

Tabelle 6.2: Verzeichnis aller Symbole dieses Kapitels (Fortsetzung).

Symbol	Bezeichnung	Einheit
V_{exc}	Volumen der Extraktionsflüssigkeit	[μL]
V'_{exc}	Vollständiges Extraktionsvolumen	[μL]
V_{res}	Restvolumen nach Zellseparation	[μL]
V_{que}	Volumen der Quenchingflüssigkeit	[μL]
c_{std}	Metabolitkonzentration des Standards	[μM]
c_{exc}	Metabolitkonzentration im Extrakt	[μM]
c_{cyt}	Zytosolische/intrazelluläre Metabolitkonzentration	[μM]
c_{qsn}	Metabolitkonzentration im Quenchingüberstand	[μM]
c_{cub}	Metabolitkonzentration im Kulturüberstand	[μM]
c_{lea}	Leakage Konzentration	[μM]
$u_{[...]}$	Peakflächen Quotient der jeweiligen Probenotypen	[-]
$\eta_{12C, [...]}$	^{12}C Peakfläche der jeweiligen Probenotypen	[counts]
$\eta_{13C, [...]}$	^{13}C Peakfläche der jeweiligen Probenotypen	[counts]

6.2 Prozessierung quantitativer Metabolom-Daten

Abbildung 6.1 gibt eine Übersicht über ein Protokoll der Probenprozessierung für die intrazelluläre Metabolomanalyse. Im Folgenden wird jeder der sieben Schritte des Protokolls kurz beschrieben. Dabei wird auf mögliche Quellen von Unsicherheiten während der Probenprozessierung eingegangen. Diese Beschreibung des experimentellen Ablaufes stellt die wichtigste Grundlage für die nachfolgende Modellierung dar. Für detaillierte Informationen zum Protokoll sei auf die Dissertation von Nicole Paczia (PACZIA, 2012) verwiesen.

6.2.1 Bestimmung der Biomasse

Die Bestimmung der Biomasse liefert das Referenz-Volumen für die intrazelluläre Metabolitkonzentration und hat damit einen großen Einfluss auf den Absolutwert dieser Konzentration.

Mit Blick auf die Genauigkeit der Messgröße zur Bestimmung der Biomasse ist es wichtig zu beachten, dass das für die intrazelluläre Metabolomanalyse relevante Volumen von der Summe der Zellen mit intakter Zellmembran, dem sogenannten Biovolumen, bestimmt wird (HARRIS et al., 1987). Im Gegensatz dazu wird das Biovolumen in vielen Studien jedoch

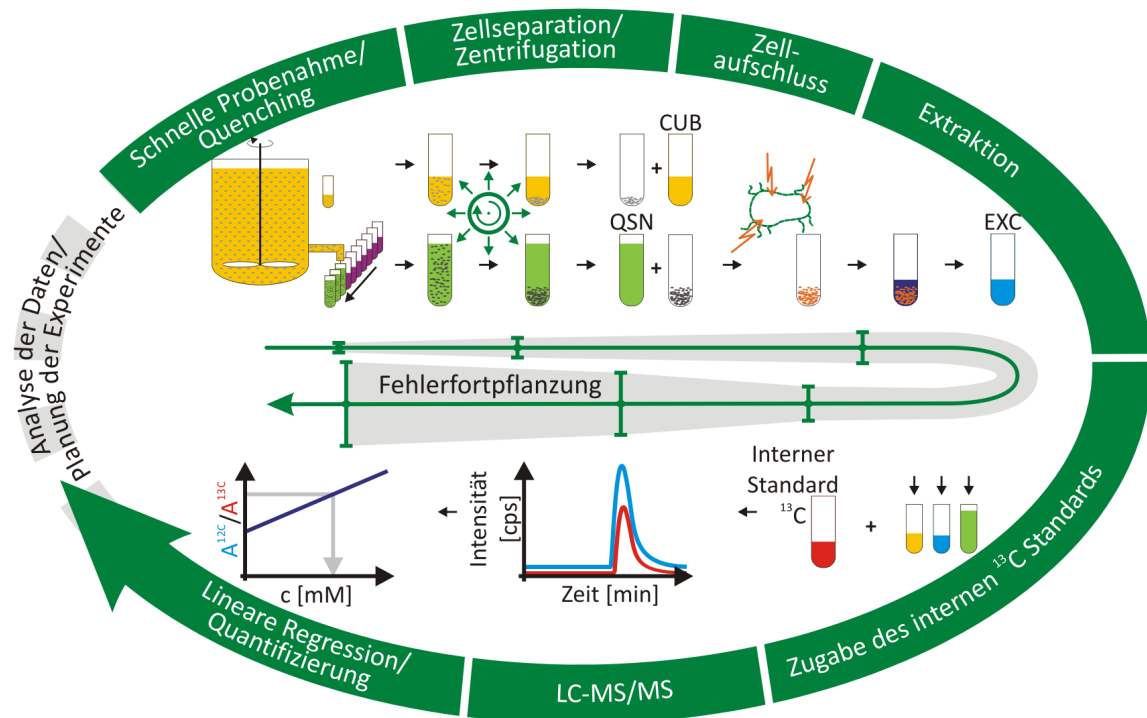


Abbildung 6.1: Experimenteller Ablauf der Datenprozessierung zur Gewinnung von intrazellulären Metabolitkonzentrationen am IBG-1 des Forschungszentrum Jülich.

nur durch Messen der Biotrockenmasse und Zuhilfenahme eines Literaturwertes für das spezifische intrazelluläre Biovolumen abgeschätzt (BUZIOLE et al., 2002). Dies führt dazu, dass als Messgröße die Summe der Biomassen intakter und nicht intakter Zellen aufgenommen wird. Außerdem liegt der Wert des spezifischen Volumens einer Zelle mit einer verhältnismäßig großen Varianz vor, da das Biovolumen über den Zyklus der Zellteilung variiert und zusätzlich von Prozessparameter, wie z.B. dem osmotischen Druck, beeinflusst wird.

6.2.2 Probenahme

Die Probenahme ist der Schritt, welcher die Prozessierung der Proben einleitet. Neben der für alle Schritte geltenden Voraussetzung der Kontaminationsfreiheit der Probe ist hierbei vor allem auf ein exaktes Einhalten des definierten Probevolumens und ein schnelles Inaktivieren des Stoffwechsels der Zellen (siehe Quenching) zu achten.

Neben manueller Probenahme (BOLTEN et al., 2007) wurden in den letzten Jahrzehnten eine große Zahl automatischer Probenahme-Einrichtungen entwickelt (SCHÄDEL und FRANCOLARA, 2009). Die automatischen Systeme basieren entweder auf aktiven Pumpen oder auf

Regelventilen, welche Über- oder Unterdruck als treibende Kraft nutzen. Das Probenvolumen unterliegt einer systematischen Abweichung vom Sollwert, welche in Abhängigkeit der gewählten Probennahme-Methode variieren und vor dem eigentlichen Experiment bestimmt werden kann.

6.2.3 Quenching

Eine große Herausforderung bei der Quantifizierung von intrazellulären Metaboliten ist das Quenching. Diese Methode wird genutzt, um den Stoffwechsel des untersuchten Organismus direkt nach der Probenahme abzustoppen und damit den *in vivo* Zustand bei der Kultivierung möglichst genau abzubilden. Dieses Ziel wird durch eine geringe Stabilität einzelner Metabolite und schnelle enzymatische Umsetzungen zwischen den Metaboliten erschwert. In den vergangenen Jahren wurden viele unterschiedliche Quenchingmethoden etabliert (VAN GULIK et al., 2012). Für die Anwendung bei Bakterien hat sich die Methode des Quenchings mit kaltem Methanol als Standardvorgehen durchgesetzt. Dabei wird die Probe direkt in ein definiertes Volumen der Quenchinglösung, welche unter anderem 50-100% Methanol enthält und mit einer Temperatur von -20 bis -80°C vorliegt, überführt (TAYMAZ-NIKEREL et al., 2009; WINDER et al., 2008; FAIJES et al., 2007).

In mehren Studien konnte belegt werden, dass während des Quenchings signifikante Mengen intrazellulärer Metabolite „verloren gehen“ können (MEINERT et al., 2013; WELLERDIEK et al., 2009; LINK et al., 2008). Mögliche Erklärungen für die, für das sogenannte Leakage ursächliche, veränderte Integrität der Zellmembran liefern der Temperatursprung beim Quenching, der osmotische Druck, der Einfluss organischer Lösungsmittel oder eine Kombination dieser Faktoren (LINK et al., 2008). Unabhängig von der Ursache kann der Leakageeffekt den Wert jeder einzelnen intrazellulären Konzentration nach unten verfälschen.

Das Ziel der Prozessierung quantitativer Metabolom-Daten sollte demnach sein, das Leakage der Metabolite zu quantifizieren und, falls notwendig, die intrazellulären Konzentrationen um das Leakage zu korrigieren. Dafür werden zusätzlich zur Probe des Extraktes (vgl. Abbildung 6.1, EXC) eine Probe des Quenchingüberstandes nach Zellseparation (vgl. Abbildung 6.1, QSN), zur Bestimmung der Metabolite außerhalb der Zellen nach dem Quenching, und eine zellfreie Probe der Kulturlösung direkt aus dem Reaktor (vgl. Abbildung 6.1, CUB), zur Bestimmung der Metabolite außerhalb der Zelle vor dem Quenching, untersucht (PACZIA et al., 2012).

Mit dieser vorgestellten Erweiterungen sind drei fehlerbehaftete Konzentrationsmessungen für jeden Metaboliten notwendig um die finale intrazelluläre Konzentration berechnen zu können.

6.2.4 Zellseparation

Die Separation der gequenchten Zellen vom extrazellulären Medium wird im optimalen Fall vollständig und unter minimalem Eintrag physikalischer Energie durchgeführt. Die Zeit bis zur Separation sollte kurz und die Separation selbst möglichst schnell sein, um die Verweilzeit der Zellen in der Quenchinglösung und damit den Effekt des Leakage zu minimieren (CANELAS et al., 2008).

Die Zellen werden via Zentrifugation (KONING und VAN DAM, 1992) oder Filtration (SÁEZ und LAGUNAS, 1976) vom extrazellulären Medium getrennt. Für beide Methoden ist eine möglichst kurze Separationszeit der beste Kompromiss um die oben erwähnten Voraussetzungen zu erfüllen (BOLTEN et al., 2007), dabei können jedoch Verluste nicht ausgeschlossen werden. Unabhängig von der ausgewählten Methode bleibt extrazelluläres Volumen zwischen den Zellen, an den Gefäßwänden haftend oder in den Poren des Filters (Totvolumen) zurück. Dieses Restvolumen verfälscht die Probe entweder durch eine Erhöhung oder eine Verdünnung der Menge an intrazellulärem Metabolit.

6.2.5 Zellaufschluss und Extraktion

Auf die Zellseparation folgt die Extraktion der intrazellulären Metabolite von den flüssigen Bestandteilen des Cytoplasmas. Für diesen Schritt werden Methoden eingesetzt, die den Zellaufschluss und die direkt folgende Metabolitextraktion miteinander verbinden. Diese Methoden gewährleisten einen kompletten Zellaufschluss sowie eine vollständige Extraktion aller Metabolite und verhindern den chemischen Ab- oder Umbau der Metabolite. Die Wahl der eingesetzten Lösungen ist abhängig von den nachfolgenden analytischen Techniken und der Kompatibilität der Lösungen mit diesen Techniken.

Neben diesen grundsätzlichen Kriterien, beeinflusst der zu untersuchende Organismus und die Eigenschaften seiner Zellwand die Wahl der Zellaufschlussmethode. Die verschiedenen etablierten Methoden nutzen physikalische (Temperatur, Ultraschall, Mahlung, uvm.), chemische (Chloroform, uvm.) oder biologische (enzymatischer Aufschluss, uvm.) Techniken (MAHARJAN und FERENCI, 2003). Für den Einsatz bei Prokaryoten sind eine ganze Reihe

von Extraktionsflüssigkeiten bekannt (WINDER et al., 2008; FAIJES et al., 2007; SHIN et al., 2010). Unabhängig von der gewählten Extraktionsflüssigkeit, kommt es jedoch oft nur zu einem unvollständigen Mischen des Biovolumens mit der Extraktionsflüssigkeit, was unter Anderem von der Effektivität des Zellaufschlusses abhängt und je nach Metabolit variieren kann (MAHARJAN und FERENCI, 2003).

6.2.6 Analytik

Der letzte Schritt der Probenprozessierung ist die Quantifizierung der intrazellulären Metabolite mit analytischen Methoden, welche eine hohe Sensitivität und Selektivität gewährleisten, einen breiten linearen Bereich und ein umfangreiches analytisches Spektrum bieten sowie eine Quantifizierung über verfügbare Standards erlauben.

Im Bereich Metabolomanalyse wird seit einigen Jahren die LC-ESI-MS/MS Technik favorisiert (BAIDOO et al., 2012), da sie die oben erwähnten Kriterien gut erfüllt. Trotz der vielen positiven Eigenschaften dieser analytischen Methode hat sie jedoch auch Nachteile. Es ist allgemein bekannt, dass in Abhängigkeit der Komplexität und der chemischen Eigenschaften (pH-Wert, Polarität und Salzkonzentration) der biologischen Matrix Ionensuppression ein Problem vor allem bei ESI-MS basierten Methoden darstellt (JESSOME und VOLMER, 2006). Aus diesem Grund ist die aus der Analytik resultierende Peakintensität oder -fläche nicht direkt proportional zur Menge des zu quantifizierenden Metabolites.

Durch den Einsatz eines internen ^{13}C -Standards, der zu quantifizierenden Metabolite mit ausschließlich ^{13}C -Kohlenstoffatomen, kann der Effekt der biologischen Matrix korrigiert werden (WU et al., 2005).

6.3 Modellierung der Metabolom-Daten Generierung

Eine Anwendung des im vorangegangenen Kapitel beschriebenen experimentellen Protokolls liefert Messungen der Peakflächen jedes einzelnen Metaboliten für die drei Prozessproben Extrakt, Quenchingüberstand und Kulturüberstand. Im Folgenden wird ein Modell für die Umrechnung der Peakflächen in intrazelluläre Konzentrationen vorgestellt. Für die folgende Beschreibung der einzelnen notwendigen Schritte wird der experimentelle Workflow rückwärts beschritten. Nach einer einführenden, allgemeinen Betrachtung des Workflows und der Abbildung des aktuellen Standes, wird das Modell nach und nach verfeinert,

indem auf die unterschiedlichen, bereits eingeführten Quellen für Unsicherheiten eingegangen wird.

6.3.1 Bestimmung der Extraktkonzentration

In einem ersten Schritt muss die, für einen Metaboliten gemessene, Peakfläche der Extraktprobe in die dazugehörige Konzentration bezogen auf das eingesetzte Extraktvolumen umgerechnet werden. Beginnend mit einem klassischen Kalibrationsansatz werden dafür definierte Konzentrationen, sogenannte Standards $c_{i,std}$, vermessen und ein lineares Modell für die resultierenden Standard-Peakflächen $\eta_{i,std}$ formuliert:

$$\eta_{i,std} = c_{i,std} \cdot m + n \quad (6.1)$$

Die Modellparameter Anstieg m und Achsenabschnitt n werden z.B. durch die Anwendung einer gewichteten linearen Regression bestimmt. Die gewichtete lineare Regression ist eine Methode der kleinsten Quadrate, welche die Varianz der Standardmessungen mit einbezieht (GANS, 1992). Nach Umstellen des Modells in Gleichung 6.1 kann aus der gemessenen Peakfläche η_{exc} des Metaboliten aus der Extraktprobe die Extraktkonzentration c_{exc} bestimmt werden:

$$c_{exc} = \frac{\eta_{exc} - n}{m} \quad (6.2)$$

6.3.2 Bestimmung der intrazellulären Konzentration

Die in Gleichung 6.2 bestimmte Extraktkonzentration muss jedoch weiter umgerechnet werden, um die Konzentration zu erhalten, welche sich auf das cytosolische Volumen der extrahierten Zellen bezieht. Dafür muss die Verdünnung während der Extraktion berücksichtigt werden. Vorab wird das cytosolische Volumen V_{cyt} aus den Messungen der Biotrockenmasse c_{CDW} , dem *a priori* definierten Volumen der Probe aus dem Reaktor V_{brs} und einem Literaturwert für das Biotrockenmasse spezifische Biovolumen ν_{CDW} berechnet:

$$V_{cyt} = c_{CDW} \cdot V_{brs} \cdot \nu_{CDW} \quad (6.3)$$

Anschließend wird das Extraktionsvolumen V'_{exc} als Summe aus cytosolischem Volumen V_{cyt} und *a priori* definiertem Volumen der Extraktionsflüssigkeit V_{exc} bestimmt:

$$V'_{exc} = V_{cyt} + V_{exc} \quad (6.4)$$

Die cytosolische Metabolitkonzentration c_{cyt} wird schließlich wie folgt bestimmt:

$$c_{cyt} = c_{exc} \cdot \frac{V'_{exc}}{V_{cyt}} = \frac{\eta_{exc} - n}{m} \cdot \left(1 + \frac{V_{exc}}{c_{CDW} \cdot V_{brs} \cdot \nu_{CDW}} \right) \quad (6.5)$$

An diesem Punkt liegt ein Wert für die intrazelluläre Metabolitkonzentration vor. Auf Grund der Unsicherheiten wie z.B. Matrixeffekte oder Leakage, welche durch die aktuellen Modellgleichungen nicht abgedeckt werden, wird diese Metabolitkonzentration nicht die wahre *in vivo* Konzentration abbilden.

6.3.3 Minimierung systematischer Fehler

Jede Prozessgröße eines experimentellen Workflows kann mit systematischen Fehlern behaftet sein, welche für alle Durchläufe eines Experimentes gleich oder variabel sind. Ist die Abweichung des Messwertes vom wahren Wert für verschiedenen experimentelle Durchläufe konstant, kann sie im Vorherein in einem separaten Experiment durch technische Replikate bestimmt (Tabelle 6.3) und für die Korrektur des Messwertes in allen weiteren Auswertungen genutzt werden. Ein typischer Messwert, der diese Art Unsicherheit aufweist, ist das Probenvolumen V_{brs} (vgl. Gleichung 6.3).

Im Gegensatz dazu können systematische Fehler auftreten, welche für jeden experimentellen Durchlauf erneut bestimmt werden müssen, da sie mit den Prozessbedingungen variieren. In diesem Fall sind zusätzliche, artfremde Messungen oder der Einsatz von Standards notwendig, um den Fehler zu korrigieren. Ein typischer Messwert, bei dem variable systematische Fehler auftreten, ist die finale intrazelluläre Konzentration c_{cyt} .

Um die finale intrazelluläre Konzentration um die bekannten, variablen, systematischen Fehler zu korrigieren, wird das Grundmodell (vgl. Gleichung 6.5) im Folgenden Stück für Stück verfeinert.

6.3.4 Matrixeffekt

Die Methode der Isotopen-Verdünnungs-Massenspektrometrie (IDMS) (WU et al., 2005) wird eingesetzt, um systematische, analytische Abweichungen, welche durch sogenannte Matrixeffekte (z.B. Ionensuppression) ausgelöst werden, zu korrigieren. Dabei wird eine vorbereitete Mischung, der interne Standard, zu jeder Metabolom-Probe gegeben, welche die vollständig ^{13}C -markierten Analoga aller Targetmetabolite enthält. Die Mischung aus Probe und internem Standard wird dann via LC-MS/MS vermessen und analysiert. Dies liefert für jedes Targetmolekül zwei Peakflächen, $\nu_{12\text{C}}$ des ^{12}C -Analyten aus der Metabolom-Probe und $\nu_{13\text{C}}$ des ^{13}C -Analogon aus dem internen Standard. Der Quotient beider Peakflächen gewährleistet eine Korrektur der Matrixeffekte, da beide Proben bei der Messung denselben Effekten ausgesetzt waren.

$$v = \frac{\eta_{12\text{C}}}{\eta_{13\text{C}}} \quad (6.6)$$

Dieses Vorgehen unterliegt der Annahme, dass die Matrixeffekte für ^{12}C -Analyt und ^{13}C -Analogon identisch sind. Außerdem liefert der Einsatz von IDMS nur dann sinnvolle Ergebnisse, wenn ^{12}C -Analyt und ^{13}C -Analogon in gleicher Weise ionisiert wurden und damit der gleichen Ionensuppression unterliegen. Diese Annahme ist im jeweiligen linearen Messbereich des Detektors gegeben, da die Ionenquelle in diesem Bereich ungesättigt ist (LIANG et al., 2003).

Unter diesen Annahmen und Beachtung des in Gleichung 6.6 eingeführten Quotienten können die Gleichungen 6.1 und 6.2 umformuliert werden:

$$\begin{aligned} v_{i,std} &= c_{i,std} \cdot m + n \\ c_{exc} &= \frac{v_{exc} - n}{m} \end{aligned} \quad (6.7)$$

6.3.5 Messung des Biovolumens

Das Biovolumen der Biomasse in der Quenching-Probe dient als Referenzvolumen für die Menge intrazellulärer Metabolite. Daher ist eine direkte und Experiment spezifische Messung des Biovolumens zwingend notwendig. Die Methode des Coulter-Counters liefert ein Histogramm, welches der Anzahl der gezählten Partikel das von ihnen verdrängte Volumen gegenüberstellt. Der größere Teil des verdrängten Volumens ist auf das Cytoplasma zurückzuführen, so dass mit dieser Methode Zelltrümmer und nicht-intakte Zellen von intakten

Zellen unterschieden werden können (SONNLEITNER et al., 1992). Wird die Methode des Coulter-Counters an Stelle der klassischen Bestimmung der Biotrockenmasse (vgl. Gleichung 6.3) genutzt um das volumenspezifische Biovolumen ν_{brs} zu ermitteln, kann das cytosolische Volumen genauer bestimmt werden:

$$V_{cyt} = V_{brs} \cdot \nu_{brs} \quad (6.8)$$

6.3.6 Unvollständige Zellseparation

Die Separation der Zellen von der sie umgebenden Flüssigkeit, dem Quenchingüberstand, mittels Zentrifugation oder Filtration ist nie vollständig. Als Folge dieser unvollständigen Zellseparation verbleiben Rückstände des Quenchingüberstandes in der Biomasse-Probe. Dieser Effekt führt zu einer Erhöhung des Extraktionsvolumens (ISHIZAKI et al., 1995). Mit Berücksichtigung dieses systematischen Effektes kann Gleichung 6.4 angepasst werden:

$$V'_{exc} = V_{cyt} + V_{exc} + V_{res} \quad (6.9)$$

In dieser Gleichung charakterisiert der Term V_{res} das nach der Zellseparation zurückgebliebene Volumen des Quenchingüberstandes. Durch die Erhöhung des Extraktionsvolumens führt das Restvolumen zu einer Verdünnung der Extraktkonzentration.

Außerdem müssen die im Quenchingüberstand enthaltenen Metabolite beachtet werden. Ihre Vernachlässigung führt dazu, dass mit Gleichung 6.7 eine Extraktkonzentration bestimmt wird, welche falsch zu hoch ist. Der Grund dafür liegt in den Metaboliten des Quenchingüberstandes, welche sich im Restvolumen befinden.

Um diese systematische Abweichung korrigieren zu können ist die Bestimmung der Metabolitkonzentrationen im Quenchingüberstand c_{qsn} notwendig. Dies geschieht mit Gleichung 6.7. Bevor die Metabolitkonzentration im Quenchingüberstand zur Korrektur der Extraktkonzentration eingesetzt werden kann, muss diese auf das Extraktionsvolumen bezogen werden. Dies geschieht über den Verdünnungsterm, welcher einen Quotienten zweier Volumina darstellt:

$$c_{exc} = \frac{v_{exc} - n}{m} - c_{qsn} \cdot \frac{V_{res}}{V'_{exc}} \quad (6.10)$$

Die so bestimmte Extraktkonzentration kann noch immer von der tatsächlichen Konzentration abweichen. Eine mögliche Ursache dafür stellt die spezifische Extraktionseffizienz dar,

welche vom Grad der Zerstörung der Zellen und von den Eigenschaften des Extraktionsmittels abhängt. Eine Korrektur der Extraktkonzentration um den Faktor einer unvollständigen Extraktion kann mithilfe eines passenden internen Standards erfolgen. Ein solcher Standard sollte in definierter Menge intrazellulär vorliegen und seine Messung wäre ein direktes Maß für die Extraktionseffizienz. Da ein solcher Standard im Moment nicht verfügbar ist, kann der Fehler, welcher durch eine unvollständige Extraktion hervorgerufen wird, nicht bestimmt werden. Aus diesem Grund, wird er in dem hier vorliegenden Modell nicht weiter betrachtet und von einer Extraktionseffizienz von 100% ausgegangen.

6.3.7 Leakage

Der letzte Schritt der Verfeinerung des Modells beschäftigt sich mit der Korrektur von potentiell Metabolitverlust durch Leakage. Grundlage dafür ist die Bestimmung der Metabolitkonzentration im Kulturüberstand c_{cub} durch den Einsatz von Gleichung 6.7. Zusammen mit den bereits bestimmten Metabolitkonzentrationen für Extrakt und Quenchingüberstand ergibt sich somit ein komplettes Bild der Metabolitverteilungen vor und nach dem Quenching (vgl. Abbildung 6.2).

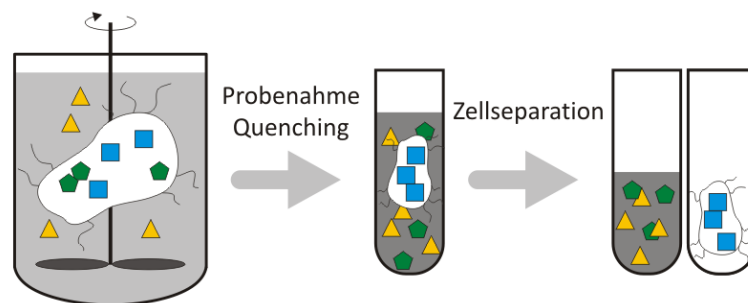


Abbildung 6.2: Darstellung der Leakage-Korrektur: Gesucht ist die intrazelluläre Metabolitkonzentration *in vivo* (Quadrat+Fünfeck). Während des Quenchings gelangen Metabolite aus der Zelle (Fünfeck), vermischen sich dort mit den Metaboliten des Kulturüberstandes (Dreieck) und bilden den sogenannten Quenchingüberstand (Fünfeck+Dreieck). Durch Messung der drei Proben Kulturüberstand, Quenchingüberstand und Extrakt (Quadrat) kann die *in vivo* Konzentration abgeschätzt werden.

Die Konzentration der durch Leakage „verloren gegangenen“ intrazellulären Metabolite ergibt sich unter Berücksichtigung der Probenverdünnungen:

$$c_{lea} = c_{qsn} \cdot \frac{V_{brs} + V_{que} - V_{cyt}}{V_{cyt}} - c_{cub} \cdot \frac{V_{brs}}{V_{cyt}} \quad (6.11)$$

Diese Gleichung zur Berechnung von Leakage unterliegt der Annahme, dass Leakage den Verlust von Metaboliten über die Membran, z.B. mittels Diffusion, darstellt und nicht durch Zellschädigung hervorgerufen wird. Wird der Metabolitverlust durch Zellschädigungen hervorgerufen, muss neben den Metaboliten auch frei gesetztes cytosolisches Volumen berücksichtigt werden.

Werden alle eingeführten Anpassungen im Modell berücksichtigt, verändert sich Gleichung 6.5 zu:

$$\begin{aligned}
 c_{cyt} &= c_{exc} \cdot \frac{V'_{exc}}{V_{cyt}} + c_{lea} \\
 &= \frac{1}{m \cdot V_{brs} \cdot \nu_{brs}} \cdot \begin{pmatrix} v_{exc} \cdot (V_{brs} \cdot \nu_{brs} + V_{exc} + V_{res}) \\ + v_{qsn} \cdot (V_{brs} + V_{que} - V_{brs} \cdot \nu_{brs}) \\ - v_{cub} \cdot V_{brs} - n \cdot (V_{exc} + V_{que} + V_{res}) \end{pmatrix} \quad (6.12)
 \end{aligned}$$

Damit liegt eine detaillierte Beschreibung der Probenprozessierung vor, welche die Korrektur bekannter systematischer Abweichungen enthält.

6.4 Modellcharakterisierung

6.4.1 Datengrundlage

Das eingeführte Modell (Gleichung 6.12) zur Auswertung von Metabolom-Daten wird im Folgenden auf Basis eines Beispieldatensatzes charakterisiert. Die Daten stammt aus der exponentiellen Wachstumsphase von Kultivierung des Organismus *Corynebacterium glutamicum*. Bei dem Experiment wurden der lysinproduzierende Stamm *C. glutamicum* DM1800 (VAN OYEN et al., 2012) und definiertes Glukose-Medium CGXII eingesetzt (KEILHAUER et al., 1993). Der Datensatz umfasst Messungen der drei Proben (Extrakt, Kulturüberstand und Quenchingüberstand) für insgesamt 44 Metabolite des Zentralstoffwechsels und ist im Anhang C.1 aufgeführt.

6.4.2 Überprüfen des Linearitätsbereiches

Bevor die experimentellen Daten für die Auswertung heran gezogen werden, werden sie im Hinblick auf die Einhaltung des durch die Standardmessung bestimmten Linearitätsbe-

reiches untersucht und gefiltert (Abb. 6.3). Für den Datensatz führte die Überprüfung des Linearitätsbereiches zu folgenden Ergebnissen:

- A** Die Daten von 17 Metaboliten können ohne Einschränkungen ausgewertet werden.
- B** Für fünf Metabolite wird für mindestens eine der Proben die obere Grenze des Linearitätsbereiches überschritten. Eine Verdünnung der entsprechenden Proben ist notwendig.
- C** Die Hälfte der Daten des Datensatzes erfüllen nicht die Anforderungen und eine Auswertung würde ggf. zu falschen Aussagen führen. Einige der Daten können mit zusätzlichen Annahme, z.B. der, dass für den jeweils betrachteten Metaboliten kein Leakage beobachtet wurde, verarbeitet werden. Allerdings ist diese Möglichkeit für jeden Metaboliten separat zu überprüfen.

Auf Basis der 17 Metabolite, deren experimentelle Daten für die nachfolgende Prozessierung geeignet sind, wird im Folgenden der Effekt der im Abschnitt 6.3 eingeführten Korrekturschritte diskutiert.

6.4.3 Einfluss der Korrekturen systematischer Fehler

Um den Einfluss der Korrektur jedes bekannte, systematischen Effektes zu untersuchen, wird ausgehend vom Referenzmodell ohne Korrektur dieser Effekte (vgl. Gleichung 6.5) ein Korrekturschritt nach dem nächsten integriert (vgl. Gleichung 6.6-6.12) und jeweils die Metabolitkonzentration ermittelt.

Die Ergebnisse dieser Studie sind in Abbildung 6.4 dargestellt und können wie folgt zusammen gefasst werden:

- Bei 15 der 17 Metabolite führt der Einsatz eines internen Standards (IDMS) zu einer Erhöhung der intrazellulären Metabolitkonzentration. Die Erhöhung beträgt im Durchschnitt etwa 80%.
- Die Messung des spezifischen Biovolumens (ν_{brs}) führt, wie erwartet, zu einem kleineren cytosolischen Volumen und resultiert damit in einer 1.5-fach höheren intrazellulären Metabolitkonzentration. Außerdem führt diese, im Vergleich zur klassischen Biotrockenmassebestimmung, genauere Methode zu einer Senkung der relativen Standardabweichungen auf durchschnittlich 66%.

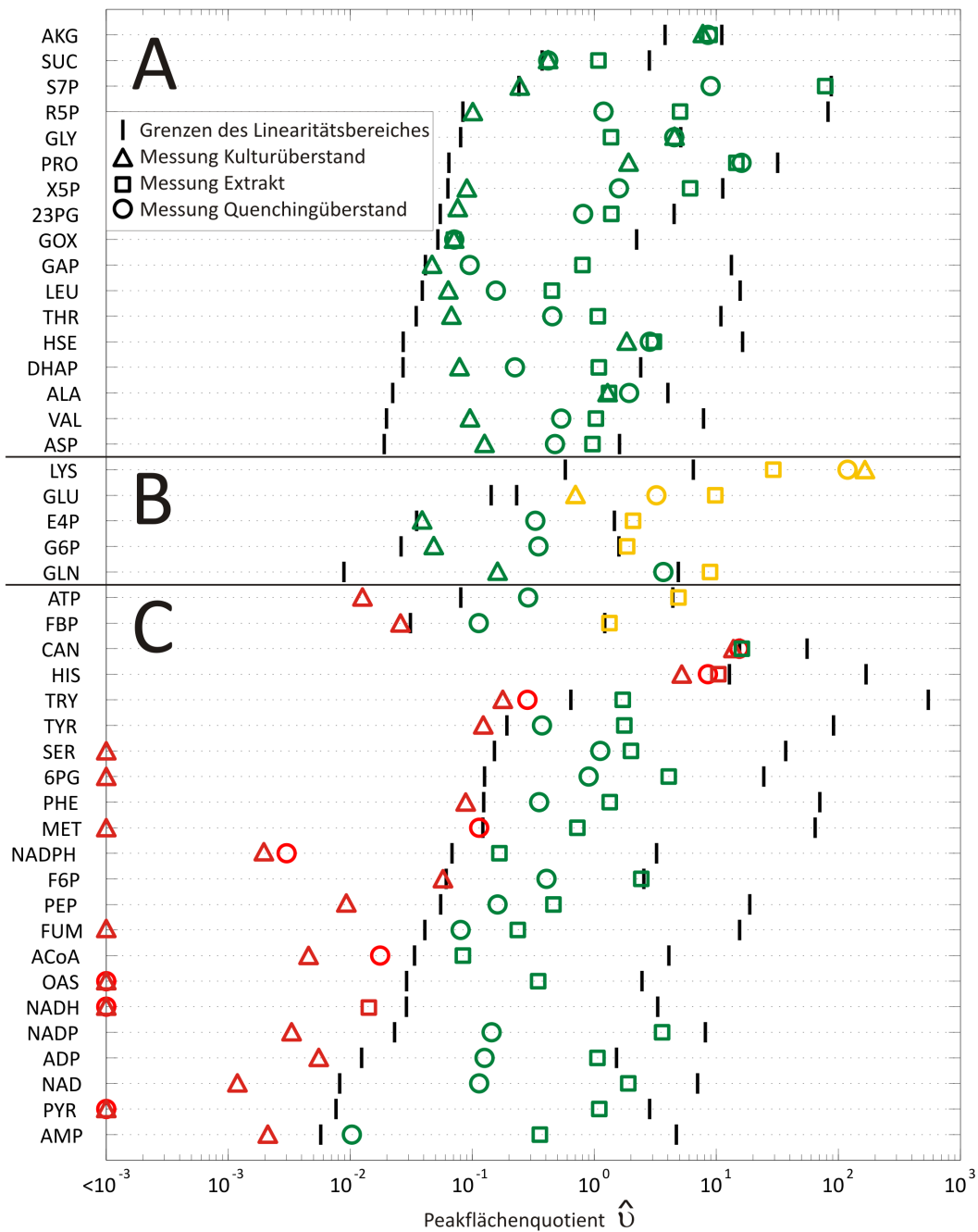


Abbildung 6.3: Überprüfen der Peakflächenquotienten auf Einhaltung des durch die Standardmessung bestimmten Linearitätsbereiches. Die vertikalen schwarzen Linien kennzeichnen die unteren und oberen Grenzen des Linearitätsbereiches. Die drei Symbole markieren den Peakflächenquotienten für Extrakt (Quadrat), Quenchingüberstand (Kreis) und Kulturüberstand (Dreieck). Die Farben der Symbole geben an, ob sich der Peakflächenquotienten unterhalb (rot), oberhalb (gelb) oder innerhalb (grün) des Linearitätsbereiches befindet. Entsprechend dieses Farb-codes wurden alle Metabolite in drei Kategorien eingeteilt: (A) Datensätze, welche ohne Einschränkung prozessiert werden können; (B) Proben, die nach einer Verdünnung erneut vermessen werden müssen; (C) Fälle, bei denen mindestens einer der Peakflächenquotienten unterhalb des Linearitätsbereiches liegt, daher kann der Metabolit nicht mit dem vorgestellten Modell quantifiziert werden.

- Für nahezu alle Metabolite führt die Berücksichtigung eines Restvolumens nach der Zellseparation (V_{res}) nur zu kleinen Änderungen der intrazellulären Metabolitkonzentration. Dies kann damit erklärt werden, dass die unvollständige Zellseparation mit (1) Metabolitverdünnung und (2) zusätzlichen Metaboliten aus dem Quenchingüberstand zwei sich aufhebende Effekte bedingt (vgl. Abschnitt 6.3.6). Der Effekt dieser Korrekturschritte auf die Standardabweichung ist mit einer Erhöhung um durchschnittlich 4% auch sehr gering.
- Die Korrektur des Leakage hat den größten Einfluss auf die intrazelluläre Metabolitkonzentration. Dieser Schritt führt im Durchschnitt zu einer Verfünffachung der Konzentration. Für sechs der Metabolite erhöht sich die relative Standardabweichung der Konzentration durch diesen Schritt erheblich auf 200-1000% des Wertes ohne Leakage-Korrektur. Für die anderen Metabolite bleibt der Fehler durch diesen Korrekturschritt nahezu unverändert.

Zusammenfassend kann mit dieser Studie gezeigt werden, dass ohne die Korrekturen der bekannten systematischen Fehler alle betrachteten Konzentrationen mit einem durchschnittlichen Faktor von 26 falsch zu klein bestimmt werden. Beim Vergleich der ermittelten Standardabweichungen von Referenzmodell und korrigiertem Modell sind zwei Fälle zu unterscheiden. Für sechs der 17 Metabolite erhöht sich die relative Standardabweichung auf 130-1000%. Für die anderen 11 Metabolite sinkt die relative Standardabweichung auf 60-80%. Diese großen Unterschiede sind im Wesentlichen auf den Schritt der Leakage-Korrektur zurückzuführen.

Der vorangeschaltete Linearitätstest filtert Peakflächenquotienten unterhalb des Linearitätsbereiches heraus. Dadurch werden automatisch Metabolite eliminiert, deren Daten keine oder nur sehr geringe Konzentrationen im Kultur- sowie Quenchingüberstand aufweisen. Dies bedeutet, dass Metabolite, welche für Kultur- und Quenchingüberstand Peakflächen nahe Null und für die Extrakt-Probe einen signifikant höheren und damit für die Regression geeigneten Wert aufweisen, separat betrachtet werden müssen. Metabolite, die diese Charakteristik aufweisen, benötigen eventuell keine Leakage-Korrektur und somit kann direkt die ermittelte Extraktkonzentration (Gleichung 6.7) als finale Konzentration angesehen werden. Metabolite des betrachteten Datensatzes, auf die dies zutrifft, sind unter anderem Pyruvat und NADPH (vgl. Abbildung 6.3 Abschnitt C.). Dieses noch offene Problem ist schwierig automatisiert zu lösen, da es häufig Einzelfallunterscheidungen und damit manuelles Eingreifen bedarf.

Das Modell (Gleichung 6.12) gewährleistet die Bestimmung der intrazellulären Metabolit-

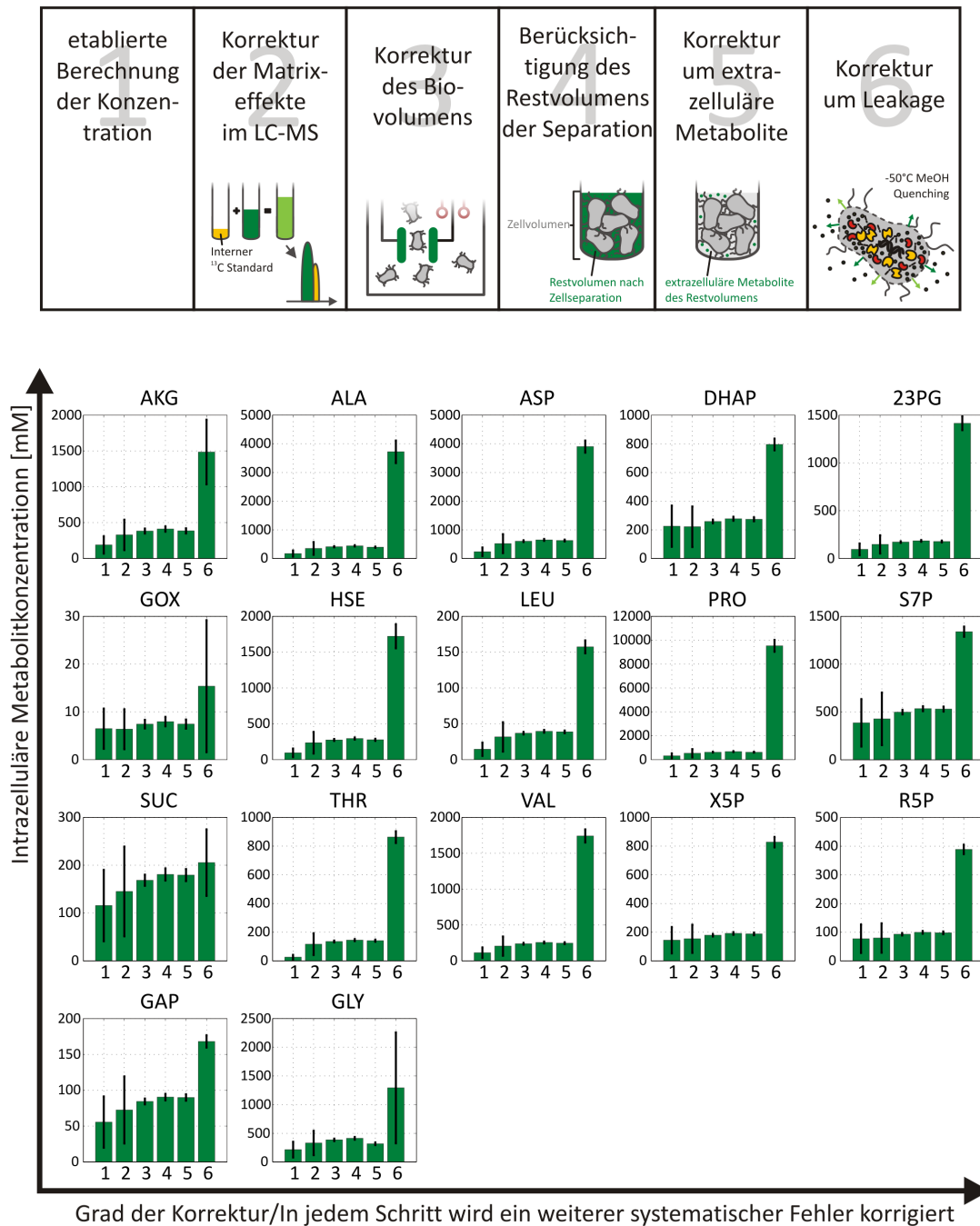


Abbildung 6.4: Veränderung der intrazellulären Metabolitkonzentration bei kumulativer Korrektur der systematischen Fehler. (1) Referenzmodell; (2) Matrixeffekt; (3) Messung des Biovolumens; (4) Metabolitverdünnung durch unvollständige Zellseparation; (5) zusätzliche Metabolite durch unvollständige Zellseparation; (6) Leakage. Der Fehlerbalken symbolisiert jeweils die 2σ -Umgebung der ermittelten Genauigkeit der Konzentrationen um den Mittelwert. Alle Werte können der Tabelle C.12 entnommen werden.

konzentration, welche den *in vivo* Zustand des Metaboloms abbildet. Bei aktuellem Wissensstand ist es die beste Möglichkeit der Verarbeitung von Daten, welche aus dem spezifischen, vorgestellten Protokoll entstanden sind.

6.4.4 Fortpflanzung zufälliger Fehler

Das beschriebene Modell kann genutzt werden um intrazelluläre Metabolitkonzentrationen zu berechnen. Der nun folgenden Abschnitt beschäftigt sich damit, zusätzlich die Genauigkeit dieser Konzentrationen zu bestimmen. Dafür werden zuerst die zufälligen Fehler aller Messgrößen, ihre Varianzen, benötigt. Abbildung 6.5 zeigt eine schematische Darstellung des Modells inklusive aller fehlerbehafteten Messgrößen.

Zur Berechnung der Varianz der intrazellulären Metabolitkonzentration ist außerdem eine Methode der Fehlerfortpflanzung nötig. Auf Grund der Nichtlinearität des Modells wird die Monte-Carlo Simulation (EFRON, 1979), welche in Abschnitt 2.8 beschrieben wurde, als nichtlineare Methode der Fehlerfortpflanzung eingesetzt.

Die Funktionsweise der Monte-Carlo Simulation kann verglichen werden mit einer *in silico* Generierung vieler biologischer Replikate. Für die Nutzung der Methode wird angenommen, dass alle fehlerbehafteten Messgrößen normalverteilt und sowohl ihr Mittelwert als auch ihre Varianz bekannt sind. Während der Monte-Carlo Simulation werden die Mittelwerte aller Variablen zufällig im Rahmen ihrer Varianzen verrauscht und das Modell mit jedem dieser Sätze an Variablen simuliert. Das Ergebnis der Monte-Carlo Simulation ist eine Zufallsverteilung der intrazellulären Metabolitkonzentration, welche zur Bestimmung von Mittelwert und Varianz herangezogen werden kann.

Die Varianzen aller Variablen wurden in Vorab-Experimenten bestimmt (vgl. PACZIA, 2012) und sind in Tabelle 6.3 aufgelistet. Die experimentell bestimmten Standardabweichungen liegen im Bereich von 0.78% für das Quenchingvolumen (V_{que}) bis 19.05% für das Restvolumen nach der Zellseparation (V_{res}).

Unter Berücksichtigung aller Fehler wurden die intrazellulären Metabolitkonzentrationen erneut bestimmt. Tabelle 6.4 enthält die modell-basiert ermittelten intrazellulären Metabolitkonzentrationen des Referenzmodells im Vergleich zu den modell-basiert ermittelten Konzentrationen des verfeinerten Modells inklusive aller Korrekturen und der Fehlerfortpflanzung. Die Unterschiede beider Ergebnisse sind leicht ersichtlich und verdeutlichen damit die Wichtigkeit der detaillierten Datenprozessierung während der Metabolomanalyse.

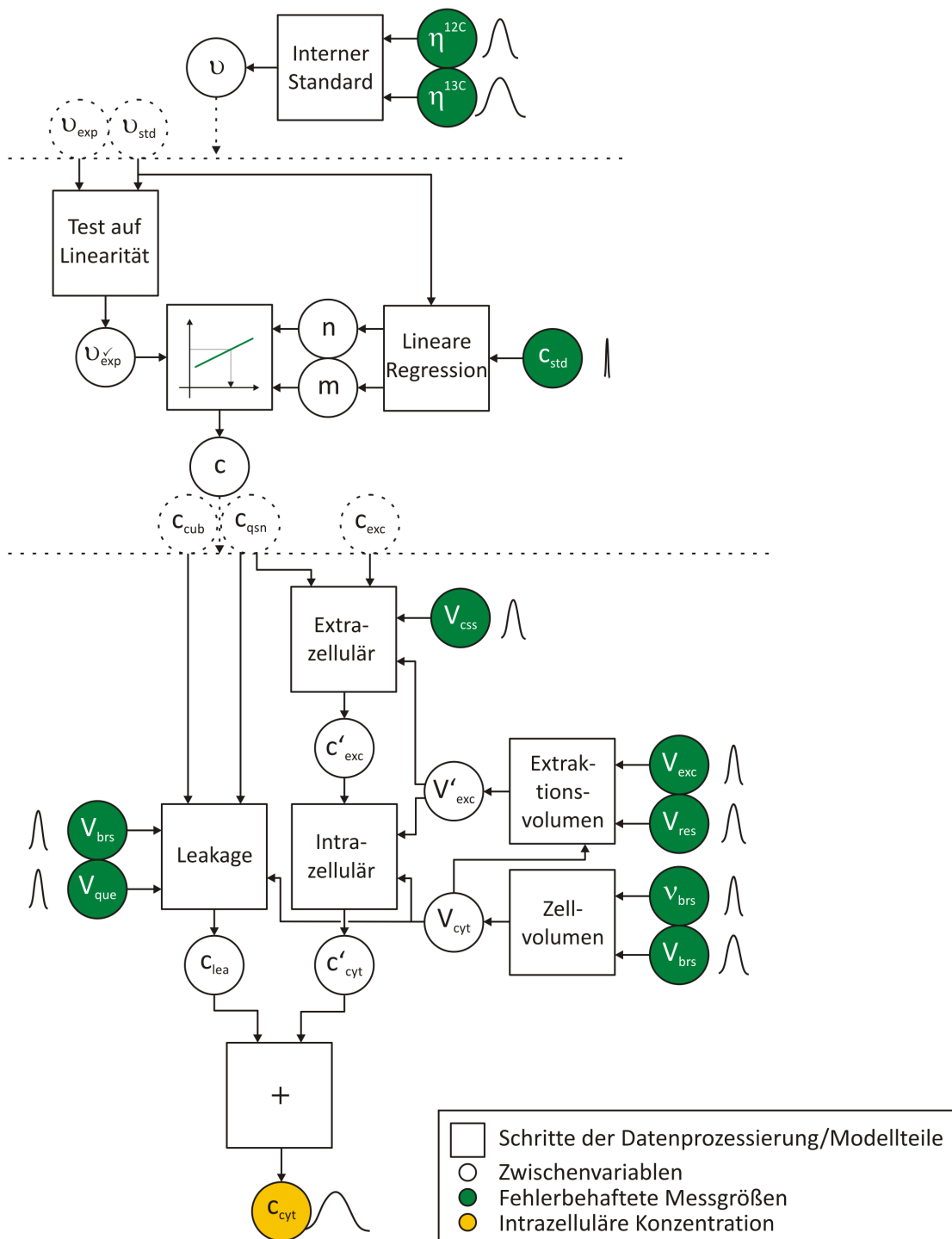


Abbildung 6.5: Schematische Darstellung des Modells zur Auswertung von Metabolom-Daten inklusive Fortpflanzung der Varianzen aller Messgrößen (grün) auf die Varianz der intrazellulären Konzentration (gelb). Die einzelnen Quadrate enthalten Teile des Modells. Die gestrichelten Linien separieren verschiedene Modellteile bzw. mehrere Modellgrößen, welche nach denselben Vorschriften verarbeitet werden.

Tabelle 6.3: Messgrößen des Protokolls der Metabolom Probenprozessierung und ihre Fehler. Systematischer (Tendenz) und zufälliger (Varianz) Fehler jeder Messgröße wurden unabhängig voneinander experimentell bestimmt (PACZIA, 2012).

Messgröße	Einheit	Sollwert	Messwert	Tendenz	Standardabweichung
c_{CDW}	[g L ⁻¹]	-	3.95	-	0.17
ν_{cdw}	[μ L mg ⁻¹]	-	1.93 ^a	-	0.96
ν_{brs}	[μ L ml ⁻¹]	-	5.205	n.b.	0.10
V_{brs}	[μ L]	5000	4782	218	38.26
V_{exc}	[μ L]	1350	1372.275	22.275	17.43
V_{res}	[μ L]	-	100.61	n.b.	19.17
V_{que}	[μ L]	15000	15472.5	472.5	120.69
$\eta_{12C, std_0.25}$	[counts]	-	4.29E+06	n.b.	1.67E+05
η_{12C, std_1}	[counts]	-	6.94E+06	n.b.	4.30E+05
η_{12C, std_5}	[counts]	-	1.64E+07	n.b.	7.25E+05
$\eta_{13C, std_0.25}$	[counts]	-	6.63E+06	n.b.	6.48E+04
η_{13C, std_1}	[counts]	-	6.24E+06	n.b.	1.29E+05
η_{13C, std_5}	[counts]	-	6.18E+06	n.b.	9.94E+04

^aaus RÖNSCH et al. (2003), n.b. nicht bestimmbar

6.5 Optimierung des experimentellen Workflows

Nachdem das Modell auf Basis eines Beispieldatensatzes charakterisiert wurde, soll dieses zur Analyse des Optimierungspotentials des experimentellen Workflows eingesetzt werden.

6.5.1 Sensitivitätsanalyse

Sensitivitätsanalysen (vgl. Abschnitt 2.4) erlauben eine *in silico* Analyse des Einflusses der Messgenauigkeit aller Eingangsgrößen auf die Varianz der intrazellulären Konzentration. Bei den Analysen wurde jede Varianz um 50% ihres experimentell bestimmten Wertes reduziert und damit die Möglichkeit simuliert, den jeweiligen Messwert doppelt so genau zu messen wie bisher.

Die Ergebnisse der Sensitivitätsanalyse sind in Abbildung 6.6 exemplarisch für drei Metabolite aus verschiedenen Teilen des Zentralstoffwechsels dargestellt und für alle Metabolite im Anhang in den Tabelle C.14 und C.13 zu finden. Die ¹²C und ¹³C Peakflächen von Standard und Quenchingüberstand gefolgt von der Messung des Biovolumens wurden als einflussreichste Messgrößen identifiziert. Eine Erhöhung der Genauigkeit der intrazellulären Metabolitkonzentration ist mit einer verbesserten massenspektrometrischen Messung und einer genaueren Biovolumenbestimmung möglich. Den größten Einfluss haben die Genauigkeit

Tabelle 6.4: Intrazelluläre Metabolikonzentrationen bestimmt durch die Anwendung der zwei Modellansätze der Datenprozessierung.

Metabolit	Intrazelluläre Metabolikonzentration [μM]	
	Referenzmodell (Gl. 6.5)	Verfeinertes Modell (Gl. 6.12)
Glycerinaldehyd-3-Phosphat	25.14	167.80 \pm 8.56
Dihydroxyacetonphosphat	102.14	793.93 \pm 40.78
2/3-Phosphoglycerat	43.55	1412.06 \pm 72.16
Ribose-5-Phosphat	34.94	387.93 \pm 16.38
Xylose-5-Phosphat	64.84	825.92 \pm 36.64
Sedoheptulose-7-Phosphat	174.85	1337.71 \pm 53.62
α -Ketoglutarat	85.41	1473.90 \pm 448.60
Succinat	52.15	204.78 \pm 68.83
Glyoxylat	2.94	15.34 \pm 13.77
Glycin	97.13	1283.37 \pm 964.32
Alanin	78.01	3714.27 \pm 392.85
Valin	51.03	1737.33 \pm 91.52
Aspartat	106.70	3894.39 \pm 211.47
Homoserin	43.00	1718.66 \pm 165.63
Threonin	11.43	860.61 \pm 41.17
Leucin	6.55	157.01 \pm 9.03
Prolin	145.86	9509.24 \pm 495.28

der ^{12}C Peakfläche des Standards, sowie die Genauigkeiten der ^{12}C und ^{13}C Peakflächen des Quenchingüberstandes.

6.6 Zusammenfassung und Diskussion

Die detaillierte Analyse eines Protokolls zur Prozessierung von Metabolom-Daten liefert Einblicke in die experimentellen Abläufe der Metabolomanalyse und kann zur Bewertung und Identifikation von Optimierungspotential eingesetzt werden. Das Protokoll der Prozessierung wurde in MATLAB modelliert und kann, durch seinen modularen Aufbau, leicht an neue Protokolle angepasst werden. Das vorgestellte Modell ermöglicht eine quantitative Auswertung der Rohdaten inklusive Fehlerbetrachtung.

Das Modell stellt weiterhin die Grundlage dafür dar, die Genauigkeit der Metabolom-Daten bezogen auf die eingesetzten Rohdaten zu analysieren. Dadurch können kritische Schritte im Prozessierungs-Protokoll identifiziert und Optimierungspotentiale aufgedeckt werden.

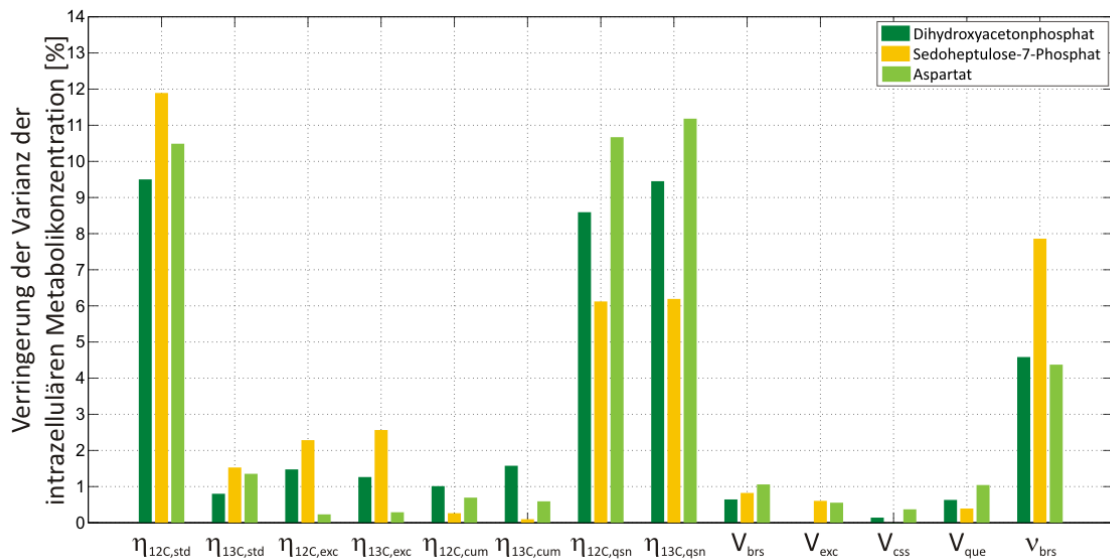


Abbildung 6.6: Sensitivität der Varianz der intrazellulären Metabolitkonzentration in Bezug auf die Varianzen aller Messgrößen. Die Verringerung jeder einzelnen Varianz auf 50% des experimentell bestimmten Wertes führt zu einer Verminderung der Varianz der Metabolitkonzentration um den dargestellten Wert.

Für den vorliegenden Beispieldatensatz von *C. glutamicum* konnten die Peakflächen von Standard und Quenchingüberstand sowie das volumenspezifische Biovolumen als einflussreiche Messwerte identifiziert werden. Eine Verbesserung der Genauigkeit dieser Messwerte würde zu einer deutlich geringeren Standardabweichung der intrazellulären Konzentration führen.

Kapitel 7

Sensormodellierung



Optische Biosensoren bieten eine schnelle Möglichkeit, Organismen auf die Produktion eines Targetmoleküls zu überprüfen und damit optimale Stämme eines Organismus für die biotechnologische Produktion zu identifizieren. Um das Onlinesignal der Fluoreszenz in die es induzierende Konzentration umzurechnen, ist eine Kalibrierung notwendig. In diesem Kapitel wird ein mechanistisches Modell vorgestellt, welches den angesprochenen Zusammenhang zwischen Konzentration und Fluoreszenz abbildet. Das Modell wird an experimentelle Daten angepasst und genutzt, um einerseits eine Kalibrierfunktion zu erstellen und andererseits neues Wissen über die stattfindenden Mechanismen zu erlangen. Die in diesem Projekt eingesetzten experimentellen Daten stammen von Nuriye Mustafi (AG Populationsheterogenität und Signaltransduktion, IBG-1).

7.1 Einleitung

7.1.1 Motivation

Die Optimierung industriell eingesetzter Mikroorganismen ist ein wesentlicher Schwerpunkt in der Biotechnologie. Das Ziel ist es, die Mikroorganismen so zu verändern, dass sie die maximal mögliche Menge eines industriell relevanten Targetmoleküls produzieren. Die Stammentwicklung erfolgt häufig in einem sehr zeitintensiven Ansatz aus zufallsbasierter Mutagenese, Screening, Selektion und rationalem Design der Pathways (BAILEY, 1991).

Der Einsatz genetisch kodierter optischer Biosensoren, welche die Abundanz eines Targetmoleküls über ein Fluoreszenzsignal anzeigen, in Kombination mit dem Messverfahren der Durchflusszytometrie bietet die Möglichkeit eines Hochdurchsatz FACS (fluorescence-activated cell sorting) Screenings von Stämmen auf Einzelzellebene. Über zufällige Mutagenese erzeugte Stämme eines Organismus werden zuerst auf Fluoreszenz überprüft und

nach der Intensität ihrer Fluoreszenz sortiert. Zur Verifikation potentieller Produzenten, werden die Stämme mit der stärksten Fluoreszenz anschließend im BioLector® kultiviert. Ein solcher Workflow (vgl. Abbildung 7.1), welcher eine schnelle Identifizierung potentieller 'Superproduzenten' ermöglicht, wurde am IBG-1 etabliert (BINDER et al., 2012).

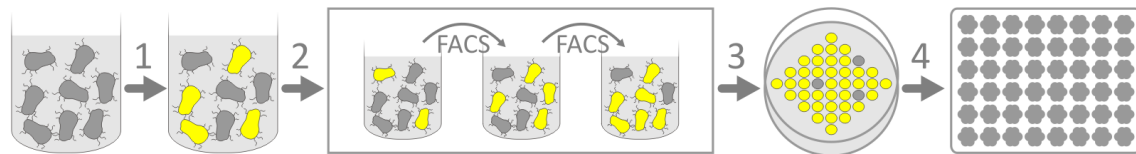


Abbildung 7.1: Biosensor-basiertes Hochdurchsatz FACS Screening bestehend aus den Schritten (1) Mutagenese, (2) Screening via FACS, (3) Sortierung auf Platten und (4) Kultivierung in Mikrotiterplatten im BioLector® (Abbildung nach MUSTAFI et al., 2012).

Um das Fluoreszenzsignal dieses Workflows interpretieren und für weitere Analysen einsetzen zu können, muss der Zusammenhang zwischen Fluoreszenz und der die Fluoreszenz induzierenden Konzentration des Targetmoleküls bekannt sein. Ein kinetisches Modell bildet die einzige Möglichkeit diesen Zusammenhang zu beschreiben.

Auf Basis verfügbarer biochemischer Informationen in Form von Literatur und Nutzdaten aus Kalibrierexperimenten im BioLector® wird ein kinetisches Modell aufgestellt, welches den Zusammenhang zwischen der Fluoreszenz und der Konzentration beschreibt. Es wird untersucht, ob dieses Modell für zukünftige Kalibrierungen und die Planung neuer Experimente genutzt werden kann.

7.1.2 Methioninproduktion in *Corynebacterium glutamicum*

Corynebacterium glutamicum ist ein Bakterium mit großer Bedeutung für die Biotechnologie (KINOSHITA et al., 1957). Es ist sehr gut untersucht und wird für die Produktion vieler biotechnologisch relevanter Substanzen (u.A. L-Glutamat und L-Lysin) eingesetzt.

Methionin ist eine essentielle proteinogene, schwefelhaltige Aminosäure, welche überwiegend Anwendung in Medizin und Tierernährung findet. Die Entwicklung eines Stammes zur Produktion von Methionin ist daher von großem Interesse (IKEDA, 2003).

Die experimentelle Kette bestehend aus Biosensor-Entwicklung, chemisch induzierter Mutagenese und Hochdurchsatz Screening wird eingesetzt, um veränderte Stämme des Bakteriums *C. glutamicum* zu erzeugen und diejenigen Stämme zu identifizieren, welche in der

Lage sind eine maximal mögliche Menge an Methionin zu produzieren (MUSTAFAI et al., 2012).

Der optische Biosensor, welcher auf erhöhte intrazelluläre Methioninkonzentrationen reagiert, wurde als Plasmid auf Basis des Regulators Lrp (leucine-responsive protein) konstruiert und an *eyfp* gekoppelt. Dadurch wird die Expression von *eyfp* und damit die Produktion des gelb fluoreszierenden Proteins eYFP (enhanced yellow fluorescent protein) indirekt über die Konzentration von Methionin reguliert.

7.1.3 Symbolverzeichnis

Tabelle 7.1: Verzeichnis aller Symbole dieses Kapitels.

Symbol	Bezeichnung
$DP_{[ex]}$	Dipeptid (intra-/extrazellulär)
$AA_{[ex]}$	Aminosäure/Substrat (intra-/extrazellulär)
$SENS_{[a]}$	(aktiver) Sensor
$PROM_{[a]}$	(aktiver) Promotor
$PROT_{[f]}$	(fluoreszierendes) Protein
EX	Exporter
upt_{DP}	Aufnahmerate des Dipeptids
exc_S	Exkretionsrate des Substrates
bind	Sensor-Substrat-Bindungsrate
act	Aktivierung des Promotors
pro	Produktionsrate des Proteins
del	Reifung des Proteins
$prod_{EX}$	Produktionsrate des Exporters
$degr_{EX}$	Abbaurrate des Exporters

7.2 Aufbau und Analyse des Sensormodells

7.2.1 Biochemische Modellierung

Der Zusammenhang zwischen der intrazellulären Konzentration der Aminosäure Methionin und der Fluoreszenz des Proteins eYFP wird mit einem biochemischen Modell abgebildet.

Tabelle 7.2: Modell-Parameter, welche in diesem Kapitel von Bedeutung sind.

Symbol	Bezeichnung	Einheit
μ	Wachstumsrate	$[\text{min}^{-1}]$
D	Diffusionskonstante	[-]
$vmax_{uptDP}$	maximale Aufnahme rate des Dipeptides	$[\text{mM min}^{-1}]$
km_{uptDP}	Michaelis-Menten Konstante der Aufnahme rate	[mM]
$vmax_{excS}$	Maximale Exkretionsrate	$[\text{mM min}^{-1}]$
km_{excS}	Michaelis-Menten Konstante der Exkretionsrate	[mM]
$kd_{bind, fwd}$	Konstante der Sensor-Substrat-Bindung	[mM]
$kd_{bind, bwd}$	Konstante der Sensor-Substrat-Spaltung	[mM]
$kd_{act, fwd}$	Konstante der Promotoraktivierung	$[\text{mM}^{-1}\text{min}^{-1}]$
$kd_{act, bwd}$	Konstante der Promotordeaktivierung	$[\text{min}^{-1}]$
k_{pro}	Geschwindigkeitskonstante der Proteinproduktion	$[\text{min}^{-1}]$
d_{del}	Konstante der Proteinreifung	$[\text{min}^{-1}]$
ka_{prodEX}	Geschwindigkeitskonstante der Exporterproduktion	$[\text{min}^{-1}]$
k_{degrEX}	Geschwindigkeitskonstante des Exporterabbaus	$[\text{min}^{-1}]$
$time_{shift}$	Reifung	[min]
$DP_{ex,0}$	Startkonzentration Dipeptid (extrazellulär)	[mM]
$SENS_0$	Startkonzentration Sensor	[mM]
$PROM_0$	Startkonzentration Promotor	[mM]
DP_{in}	Zugabe Dipeptid	[mM]
n_{DP}	Anteil Methionin-enthaltendes Dipeptid am insgesamt zugegebenem Dipeptid	[-]
$n_{DP,S}$	Anzahl der Methionin-Moleküle im Dipeptid	[-]

Das betrachtete System ist in Abbildung 7.2 dargestellt. Die darin enthaltenen biochemischen Prozesse werden rein deterministisch beschrieben, auch wenn einzelne Schritte, wie z.B. die Gen-Regulation stochastisch ablaufen. Dieser Ansatz ist für eine Population von Zellen, wie sie im Folgenden betrachtet wird, gerechtfertigt. In den nächsten Abschnitten wird das betrachtete biochemische System schrittweise eingeführt und durch Kinetiken beschrieben.

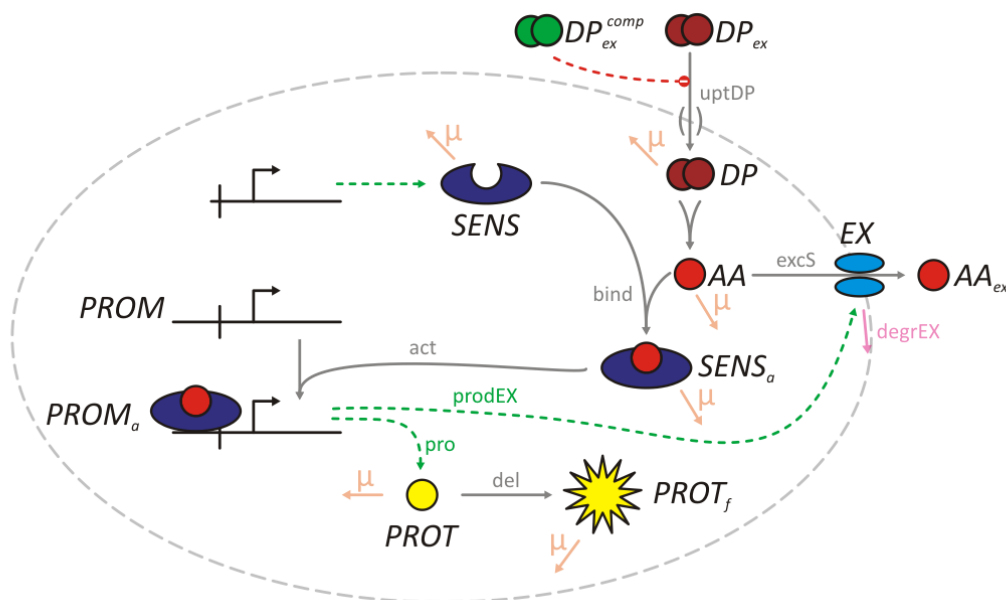


Abbildung 7.2: Darstellung der im Sensormodell integrierten biochemischen Prozesse zur Beschreibung des Zusammenhangs zwischen einer intrazellulären Aminosäurekonzentration (AA) und einem Fluoreszenzsignal ($PROT_f$).

Dipeptidaufnahme und -hydrolyse

Um für Test- und Kalibrierzwecke eine möglichst hohe intrazelluläre Konzentration des Targetmoleküls Methionin zu erreichen, werden Dipeptide als Substrate eingesetzt. Dieses Vorgehen wird in VRLJIC et al., 1996 beschrieben. Die Dipeptide werden zum Medium hinzugegeben, durch die Zellen aufgenommen und direkt hydrolysiert. Aufnahme und Hydrolyse des Dipeptides werden im Modell zusammen gefasst, da beide Mechanismen biochemisch nicht auseinander gehalten werden können. Die zusammengefassten Reaktionen werden durch eine Michaelis-Menten-Kinetik beschrieben. Parametergrenzen für spätere Analysen konnten TRÖTSCHEL et al., 2005 entnommen werden.

Die Zugabe des Methionin-enthaltenden Dipeptids erfolgt immer in Kombination mit einem Dipeptid, welches kein Methionin enthält. Beide Dipeptide nutzen denselben Transportme-

chanismus, so dass die Konkurrenz des zweiten Dipeptides beim Aufnahmefluss über eine kompetitive Substrathemmung berücksichtigt werden muss. Das Verhältnis beider Dipeptide ist in verschiedenen Experimenten variabel und kann mit dem Parameter n_{DP} eingestellt werden.

Als vereinfachende Annahme werden die Parameter zur Aufnahme des Dipeptides, km_{uptDP} und $vmax_{uptDP}$, mit denen der Aufnahme des konkurrierenden Dipeptides, km_{uptDP2} und $vmax_{uptDP2}$, gleichgesetzt.

$$upt_{DP} = \frac{vmax_{uptDP} \cdot DP_{ex}}{km_{uptDP} \cdot \left(1 + \frac{DP_2}{km_{uptDP2}}\right) + DP_{ex}} \quad (7.1)$$

$$upt_{DP2} = \frac{vmax_{uptDP2} \cdot DP_2}{km_{uptDP2} \cdot \left(1 + \frac{DP_{ex}}{km_{uptDP}}\right) + DP_2} \quad (7.2)$$

$$\frac{dDP_{ex}}{dt} = n_{DP} \cdot DP_{in} - upt_{DP} \quad DP_{ex}(0) = DP_{ex,0} \quad (7.3)$$

$$\frac{dDP_2}{dt} = (1 - n_{DP}) \cdot DP_{in} - upt_{DP2} \quad DP_2(0) = DP_{2,0} \quad (7.4)$$

Exkretion

Die Exkretion von Methionin erfolgt über das Transportsystem brnFE. BrnFE ist ein bekannter Exporter für verzweigtkettige Aminosäuren (KENNERKNECHT et al., 2002; TRÖTSCHEL et al., 2005).

In LANGE et al., 2011 wurde gezeigt, dass der eingesetzte Biosensor die Expression von brnFE in Anwesenheit von Methionin erhöht. Es wird vermutet, dass ein aktivierter Promotor neben der Proteinproduktion auch zur Produktion des Exporters führt. Aus diesem Grund wurde ein Exporter eingeführt, welcher in Abhängigkeit des aktiven Promotors produziert und mit einer konstanten Rate abgebaut wird.

Wie für die Aufnahme rate, konnten auch für die Exkretion Parametergrenzen für später Analysen TRÖTSCHEL et al., 2005 entnommen werden.

$$prod_{EX} = ka_{prodEX} \cdot PROM_a \quad (7.5)$$

$$degr_{EX} = k_{degrEX} \cdot EX \quad (7.6)$$

$$\frac{dEX}{dt} = prod_{EX} - degr_{EX} \quad EX(0) = EX_0 \quad (7.7)$$

Die Exkretionsrate von Methionin ergibt sich dann wie folgt:

$$exc_{AA} = \frac{vmax_{excAA} \cdot EX \cdot AA}{km_{excAA} + AA} \quad (7.8)$$

$$\frac{dAA_{ex}}{dt} = \frac{exc_{AA}}{D} \quad AA_{ex}(0) = AA_{ex,0} \quad (7.9)$$

Sensor-Aktivierung

Das Binden von Methionin an den Sensor aktiviert den Sensor. Beschrieben wird diese Bindung durch eine reversible Reaktion. Die Reaktionsrate wird durch die Konzentration an vorliegendem Sensor nach oben begrenzt.

$$bind_{fwd} = \frac{SENS \cdot AA}{kd_{bind,fwd} + AA} \quad (7.10)$$

$$bind_{bwd} = \frac{SENS_a}{kd_{bind,bwd} + SENS_a} \quad (7.11)$$

$$bind = bind_{fwd} - bind_{bwd} \quad (7.12)$$

Die Gesamt-Sensorkonzentration, als Summe aus freiem, inaktivem Sensor ($SENS$), freiem, aktivem Sensor ($SENS_a$) und gebundenem Sensor ($PROM_a$), wird als konstant angenommen. Dies ist unter der Annahme erfüllt, dass sich die Neubildung des Sensors und die „Verdünnung“ durch Wachstum gegenseitig aufheben.

Die Bindung von Methionin an den Sensor ist der letzte Prozess, an welchem Methionin direkt beteiligt ist. Die zeitliche Änderung der intrazellulären Methioninkonzentration ergibt sich unter Betrachtung der bisher beschriebenen Prozesse nach folgender Differentialgleichung:

$$\frac{dAA}{dt} = n_{DP,AA} \cdot upt_{DP} \cdot D - exc_{AA} - bind - \mu \cdot AA \quad AA(0) = AA_0 \quad (7.13)$$

Der letzte Term, $\mu \cdot AA$, beschreibt die durch Wachstum bedingte Verdünnung, welcher bei allen intrazellulären Komponenten berücksichtigt wurde.

Die Konzentration des inaktiven Sensors verringert sich durch die Bindung von Methionin, was die Aktivierung des Sensors zur Folge hat. Der zweite Term definiert die Rate der Sen-

sorneubildung und gewährleistet die Annahme der konstanten Gesamt-Sensorkonzentration.

$$\frac{dSENS}{dt} = -bind + \mu \cdot SENS_a \quad SENS(0) = SENS_0 \quad (7.14)$$

Promotor-Aktivierung

Durch den aktiven Sensor wird der Promotor Lrp aktiviert. Es wird von einer konstanten Menge Promotor ausgegangen. Die Rate der Promotor-Aktivierung ist abhängig von der vorliegenden Konzentration an aktivem Sensor und inaktivem Promotor. Leider sind für diesen Prozessschritt keine Werte aus der Literatur bekannt.

$$act_{fwd} = kd_{act,fwd} \cdot SENS_a \cdot PROM \quad (7.15)$$

$$act_{bwd} = kd_{act,bwd} \cdot PROM_a \quad (7.16)$$

$$act = act_{fwd} - act_{bwd} \quad (7.17)$$

$$\frac{dPROM}{dt} = -act \quad PROM(0) = PROM_0 \quad (7.18)$$

$$\frac{dPROM_a}{dt} = act \quad PROM_a(0) = PROM_{a,0} \quad (7.19)$$

Die Konzentration des aktive Sensors erhöht sich durch die Bindung von Methionin an den Sensor und verringert sich durch die Bindung des Sensors an den Promotor, sowie durch die durch das Wachstum bedingte Verdünnung.

$$\frac{dSENS_a}{dt} = bind - act - \mu \cdot SENS_a \quad SENS_a(0) = SENS_{a,0} \quad (7.20)$$

eYFP Produktion und Fluoreszenz

Durch den Promotor wird die Produktion des fluoreszierenden Proteins eYFP reguliert. Am Ende der Synthese des Proteins steht ein Reifungsprozess, welcher verschiedene Umwandlungsschritte zusammenfasst. Umgesetzt wird dieser verzögernde Aspekt der Reifung durch eine zeitliche Verschiebung der Konzentrationskurve nach der Simulation um den variablen

Parameter $time_{shift}$. Die Reifezeit wird mit 30-50 Minuten geschätzt.

$$pro = k_{pro} \cdot PROM_a \quad (7.21)$$

$$del = d_{del} \cdot PROT \quad (7.22)$$

$$\frac{dPROT}{dt} = pro - del - \mu \cdot PROT \quad PROT(0) = PROT_0 \quad (7.23)$$

$$\frac{dPROT_f}{dt} = del - \mu \cdot PROT_f \quad PROT_f(0) = PROT_{f,0} \quad (7.24)$$

7.2.2 Modellstatistik

Das in Abbildung 7.2 dargestellte und die Gleichungen 7.2 bis 7.24 enthaltende Modell beschreibt den Zusammenhang zwischen der intrazellulären Konzentration von Methionin und einem Fluoreszenzsignal, welches durch diese Konzentration induziert wird.

Das Modell ist ein Differential-algebraisches Gleichungssystem, welches elf Differentialgleichungen und elf algebraische Gleichungen in Form von Kinetiken umfasst und 21 Parameter enthält.

7.2.3 Datengrundlage

Den nachfolgenden Analysen liegen experimentelle Daten aus dem Mikrofermentationssystem BioLector® zu Grunde. Für den Test des Biosensors wurden Experimente mit *C. glutamicum* unter Zugabe verschiedener Dipeptid und Dipeptidmischungen (jeweils gesamt 3 mM) durchgeführt. Dabei wurden das Backscatter Signal und die Fluoreszenz als Onlinesignal aufgenommen (MUSTAFAI et al., 2012). Zusätzlich wurden intrazelluläre Methioninkonzentrationen via Silikonölzentrifugation und HPLC ermittelt.

Im weiteren Verlauf der Arbeit spielen zwei Datensätze eine zentrale Rolle. Der erste Datensatz wurde unter Einsatz von 100% des Dipeptides Met-Met erzeugt und besteht aus Messungen der Methionin-Konzentration, der Fluoreszenz und der Biomasse. Eine Übersicht über die integrierten experimentelle Daten gibt Tabelle D.1. Die Rohdaten des Backscatter Signals wurden nach der in Kapitel 5 eingeführten Methode prozessiert und eine maximale Wachstumsrate von $0.474 \pm 0.022 \text{ h}^{-1}$ wurde abgeschätzt. Dieser Datensatz wird zur Modellkalibrierung eingesetzt.

Der zweite Datensatz stellt den Zusammenhang zwischen der jeweils maximalen intrazellulären Konzentration und der jeweils maximalen Fluoreszenz aus Experimenten unter Einsatz verschiedener Dipeptid-Mischungen dar. Die Daten können der Tabelle D.2 entnommen werden und werden zum Testen des Modells zur Datenkalibration eingesetzt.

7.2.4 Kalibration des biochemischen Modells

Auf Basis der experimentellen Daten des ersten Datensatzes wurden mit dem biochemischen Modell unter verschiedenen Annahmen Parameter-Optimierungsstudien durchgeführt. Das Modell sollte damit an die experimentellen Daten angepasst werden, um es später für die Kalibrierung von Fluoreszenz-Daten nutzen zu können.

Eine Übersicht über alle Optimierungen kann der Tabelle D.3 entnommen werden. Im Weiteren wird hier nur auf die Ergebnisse der elften Optimierung eingegangen, da diese alle Erkenntnisse der vorangegangenen Optimierungen vereint und somit die, für das vorliegende Problem, beste Optimierung liefert.

In Optimierung 11 wurde eine multiple Parameteranpassung durchgeführt. Die Startparameter wurden im Bereich von maximal 20% um den gesetzten Startwert zufallsbasiert variiert und dann jeweils basierend auf der gewichteten Summe kleinster Quadrate optimiert. Mittelwert und Standardabweichung aller mit den zehn verschiedenen optimierten Parametersätzen erzeugten Simulationsergebnisse sind in Abbildung 7.3 dargestellt. Die Simulation der intrazellulären Konzentration zeigt vor allem in der dynamischen Phase in den ersten zwei Stunden eine starke Variation. Die Simulation der Fluoreszenz zeigt nach der Optimierung eine gute Übereinstimmung mit den Messdaten.

Um einen besseren Eindruck von den mittels Optimierung erzeugten Parametersätzen zu erhalten, sind diese für drei ausgewählte Optimierungen in Tabelle 7.3 aufgeführt. Der Effekt der unterschiedlichen Parameter auf die simulierten Verläufe der intrazellulären Konzentration und der Fluoreszenz ist in Abbildung 7.4 zu beobachten.

Für den späteren Einsatz des Modells zu Kalibrierungszwecken spielen die maximale intrazelluläre Methionin-Konzentration und die maximale Fluoreszenz eine wesentliche Rolle. Die Simulationen zeigen alle drei nahezu identische maximale Fluoreszenzwerte, weisen aber große Unterschiede in den maximalen Konzentrationen auf.

Erreicht wird die größte maximale Konzentration in 11-8 durch hohe Werte in den kinetischen Parametern bei der Aufnahme und Hydrolyse des Dipeptides (*uptDP*). Durch eine

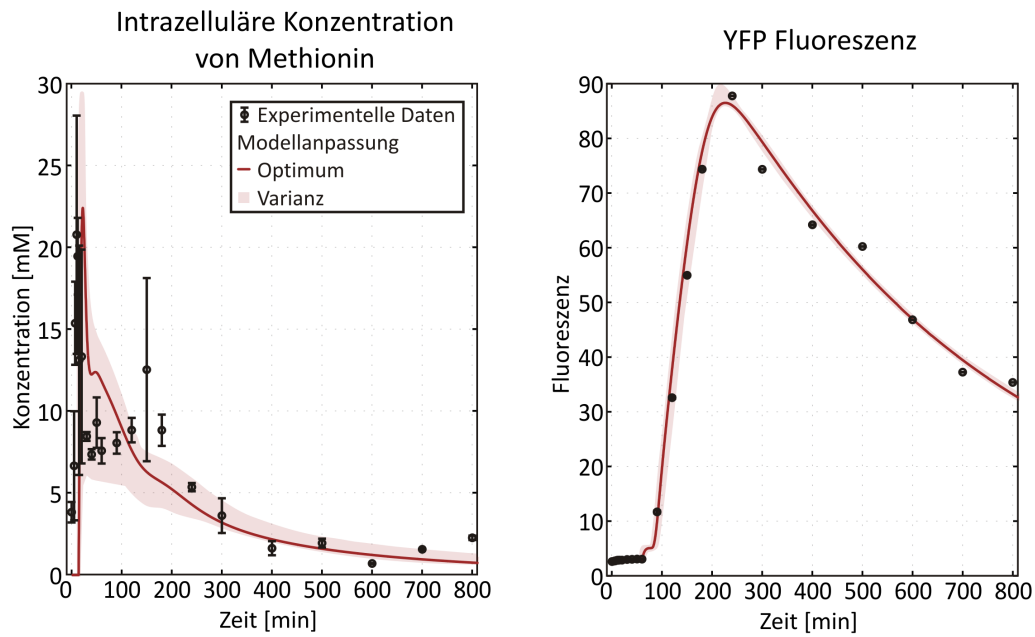


Abbildung 7.3: Simulationsergebnisse nach multipler Parameteranpassung (Optimierung 11): Dargestellt sind jeweils die experimentellen Daten und Mittelwert und Standardabweichung der dazugehörigen Simulationen für die intrazelluläre Methionin-Konzentration (links) und die Fluoreszenz (rechts).

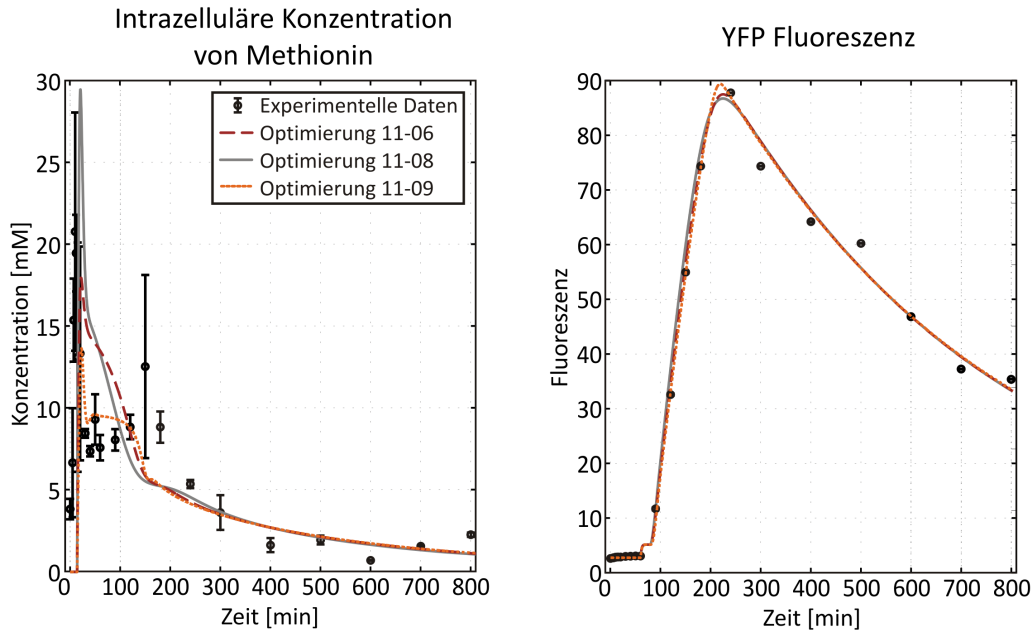


Abbildung 7.4: Ausgewählte Simulationsergebnisse nach der Optimierung: Zeitliche Verläufe der intrazellulären Methionin-Konzentration (links) und der Fluoreszenz (rechts) nach Optimierung der kinetischen Parameter (Läufe 11-6, 11-8 und 11-9). Zum Vergleich sind zusätzlich die experimentellen Daten dargestellt.

Tabelle 7.3: Ausgewählte Optimierungsergebnisse: Die resultierenden Parameter aus den Läufen 6, 8 und 9 der multiplen Optimierung 11 (siehe Tabelle D.4 für eine Übersicht über alle Ergebnisse der Optimierung 11). Grau hinterlegte Parameter werden durch experimentelle Gegebenheiten definiert und während der Optimierung auf ihren festen Werten gehalten und nicht angepasst. Fett hervorgehobene Werte wurden durch Literatur (Randbedingungen) bzw. Messungen (Startwerte) unterstützt.

Parameter	Randbedingungen		Start	11-6	11-8	11-9
$DP_{ex,0}$	fest auf 0, Dipeptid-Zugabe durch DP_{in}					
$SENS_0$	0.01	1000	10.9711	9.1321	16.0794	10.3257
$PROM_0$	1	100	100.0000	4.7984	8.1696	20.0551
DP_{in}	fest auf 30, 1/10 min zugegeben entspricht das 3 mM					
n_{DP}	fest auf 1, 100% Zugabe vom Met-Met					
$n_{DP,S}$	fest auf 2, da zwei Moleküle Met aus Dipeptid Met-Met					
μ	0.001	0.6	0.0074	0.0018	0.0018	0.0018
D	10	10000	461.5249	70.5143	116.4746	47.8458
$vmax_{uptDP}$	0.0001	1	0.0182	0.0345	0.0458	0.0264
km_{uptDP}	0.002	20	2.2922	0.4823	0.9590	0.1371
$vmax_{excS}$	0.0001	10	0.0027	0.0071	0.0057	0.0034
km_{excS}	2	20	8.5056	16.4055	18.0738	20.0000
$kd_{bind,fwd}$	2	20000	49.0934	43.3175	79.4041	57.1191
$kd_{bind,bwd}$	0.0001	100	0.0020	0.0007	0.0007	0.0007
$kd_{act,fwd}$	0.0001	100	1.0000	0.5174	0.5889	0.5344
$kd_{act,bwd}$	0.0001	100	0.8004	0.4810	0.5339	1.3249
k_{pro}	0.0001	10	0.0109	0.1160	0.0612	0.1219
d_{del}	0.0001	100	0.3790	0.5525	0.4619	0.4617
ka_{prodEX}	0.0001	10000	289.3430	822.2022	285.0046	667.6544
k_{degrEX}	0.0001	10000	0.5882	2.1751	0.6499	0.9244
$time_{shift}$	30	60	40	60	60	60

im Vergleich zu den anderen Optimierungen hohe Sensorkonzentration ($SENS_0$) und eine hohe Konstante bei der Methionin-Sensor-Bindung ($bind$), wird das intrazelluläre Methionin ebenfalls sehr schnell gebunden. Die Konzentration sinkt nach dem Anstieg ebenfalls schnell wieder ab. Auf Grund einer sehr geringen Proteinbildungsrate, wird der Effekt der hohen Methionin-Konzentration nicht auf die Fluoreszenz übertragen, so dass diese auch gut angepasst werden konnte.

Die Kalibration des biochemischen Modells liefert sinnvolle Modell-Parameter. Unter Einsatz dieser Parameter fällt die Modellvorhersage bereits sehr gut mit den experimentellen Beobachtungen zusammen.

7.2.5 Charakterisierung der Modelleigenschaften

Sensitivitätsanalysen (vgl. Abschnitt 2.4) geben eine Hinweis darauf, welche Parameter mit den eingesetzten Messwerten bestimmbar sind und liefern außerdem mit Hilfe der linearisierten Statistik (vgl. Abschnitt 2.6) zu jedem geschätzten Parameter eine Standardabweichung. Außerdem können sie zur Planung neuer Experimente eingesetzt werden.

Für die Analyse und Charakterisierung des für die Anwendung der Fluoreszenzkalibrierung neuen Modellansatzes liefern lokale, lineare Methoden eine weniger rechenintensive Alternative zu den globalen, nichtlinearen Verfahren.

Abbildung 7.5 zeigt eine Sensitivitätsmatrix als Ergebnis einer Sensitivitätsanalyse auf Basis einer Finiten-Differenzen Approximation für die Optimierung 11-9. Eine vergleichbare Matrix für die Optimierung 11-6 ist im Anhang D.2 zu finden. Je stärker farbig ein Feld, umso größer der Einfluss des Parameters auf den Messwert. Nimmt man Messungen der intrazellulären Konzentration (S) und der Fluoreszenz ($PROT_f$) an, so sind ohne weitere Einschränkungen nur 4 Parameter (dunkelgrün hervorgehoben) bestimmbar. Auch die anderen potentiellen Messgrößen weisen, wie die Matrix zeigt, keine oder nur eine sehr geringe Sensitivität gegenüber den anderen Parametern auf. Zusätzliche Messgrößen würden keinen großen Informationsgewinn bedeuten. Der Informationsgewinn potentieller, zusätzlicher Messzeitpunkte kann dieser Matrix nicht entnommen werden. Eine Antwort auf diese Frage können Sensitivitätsanalysen mit synthetisch erzeugten Daten geben.

Durch Korrelationen zwischen den Parametern tritt auch bei dieser Anwendung das Problem der strukturellen Identifizierbarkeit auf. Bei der Abschätzung der Parameter-Standardabweichungen sind nur dann nutzbare Ergebnisse zu erwarten, wenn einzelne Parameter aus den

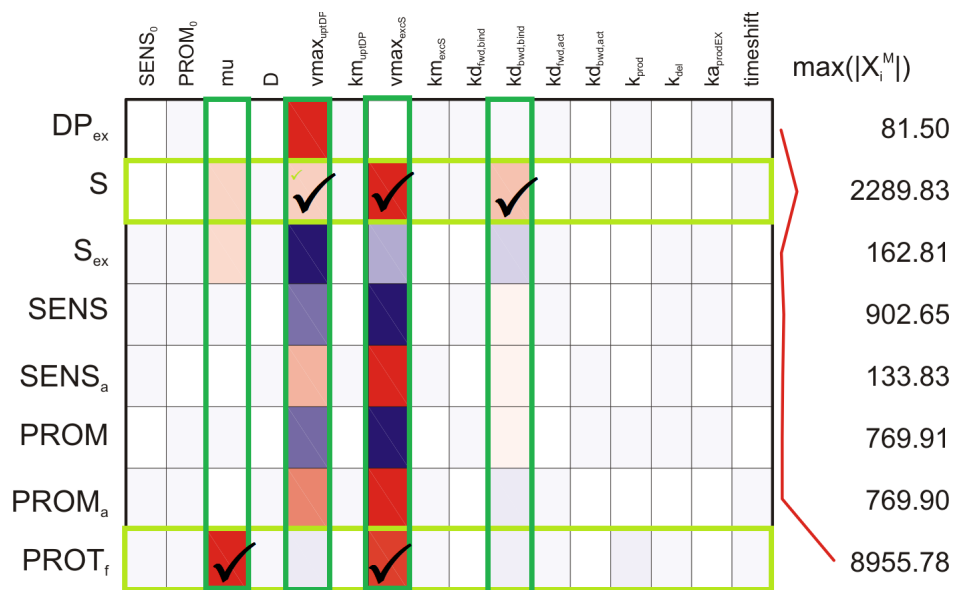


Abbildung 7.5: Sensitivitätsmatrix 11-9: Eine Erhöhung des Parameters führt zu einer Verringerung der (potentiellen) Messung, wenn die Zelle rot eingefärbt ist und zu einer Erhöhung der (potentiellen) Messung, wenn die Zelle blau eingefärbt ist. Je kräftiger die Farbe, umso größer ist der Einfluss. Die Messgrößen der intrazellulären Konzentration (S) und der Fluoreszenz (PROT_f) sind sensitiv gegenüber den grün hervorgehobenen Parametern. Die maximale Sensitivität einer jeden Zeile (rechts) dient der Normierung der Sensitivitätswerte und definiert damit die Farbskala.

Tabelle 7.4: Parameterstandardabweichungen zu ausgewählten Optimierungen: Via Sensitivitätsanalyse und linearisierter Statistik wurden Standardabweichungen zu den durch die Optimierung ermittelten Parametern abgeschätzt. Auf Grund von Parameter-Korrelationen mussten bei der Statistik jeweils zwei Parameter konstant gehalten werden, um eine Bestimmbarkeit der anderen Parameter zu erreichen.

Parameter	11-6			11-9		
	Wert	Standardabweichung		Wert	Standardabweichung	
$SENS_0$	9.1321	0.1567	1.72%	10.3257	0.2033	1.97%
$PROM_0$	4.7984	0.0896	1.87%	20.0551	0.4258	2.12%
μ	0.0018	0.0001	6.34%	0.0018	0.0001	6.32%
D	70.5143	4.0798	5.79%	47.8458	2.2374	4.68%
$vmax_{uptDP}$	0.0345	0.0012	3.60%	0.0264	0.0006	2.34%
km_{uptDP}	0.4823	0.0419	8.68%	0.1371	0.0251	18.32%
$vmax_{excS}$	0.0071		fest	0.0034	0.0011	31.86%
km_{excS}	16.4055	8.3440	50.86%	20.0000		fest
$kd_{bind, fwd}$	43.3175	2.0491	4.73%	57.1191	2.6540	4.65%
$kd_{bind, bwd}$	0.0007	0.0002	31.73%	0.0007	0.0002	24.69%
$kd_{act, fwd}$	0.5174	0.1915	37.02%	0.5344		fest
$kd_{act, bwd}$	0.4810	0.1860	38.67%	1.3249	0.0323	2.44%
k_{pro}	0.1160	0.0059	5.09%	0.1219	0.0068	5.57%
d_{del}	0.5525	0.1373	24.85%	0.4617	0.0995	21.55%
ka_{prodEX}	822.2022		fest	667.6544		fest
k_{degrEX}	2.1751	0.6027	27.71%	0.9244	0.2990	32.34%

Analysen herausgenommen werden, indem sie fest auf ihrem ermittelten Wert gehalten oder in direkten Bezug zu einem anderen Parameter gesetzt werden.

Die ermittelten Parameter-Standardabweichungen sind in Tabelle 7.4 zusammengefasst. Werden jeweils zwei Parameter während der Optimierung konstant gehalten und nicht bei der Optimierung berücksichtigt, führt dies zu guten Schätzungen der anderen Parameter. Jeweils die Hälfte der noch verbleibenden 16 Parameter kann mit einer Standardabweichung von unter 10% bestimmt werden.

7.3 Anwendung des Sensormodells als Werkzeug zur Kalibrierung

Das angepasste biochemische Modell soll genutzt werden, um Fluoreszenz-Daten in Konzentrations-Daten umzurechnen. Dafür werden in diesem Abschnitt weitere Simulationsergebnisse mit neuen Messdaten (D.2) verglichen.

Das Modell wurde für 3 mM des Dipeptids Met-Met validiert. Um es als Werkzeug zur Kalibrierung einsetzen zu können, muss es jedoch auch für andere Dipeptide und Mengen nachvollziehbare und vor allem richtige Ergebnisse liefern.

Für eine Kalibrierfunktion werden aus den dynamischen Verläufen für verschiedene Dipeptid-Mengen maximale spezifische Fluoreszenzen und maximale intrazelluläre Konzentrationen gegeneinander aufgetragen. In Abbildung 7.6 sind zwei gemessene (rot) und drei simulierte (blau) Kalibrationen gegenübergestellt. Die drei simulierten Kalibrationen wurden mit den Modell-Parametern aus den Optimierungen 11-6, 11-8 und 11-9 erzeugt.

Im Bereich niedriger Mengen Dipeptid ist ein großer Unterschied zwischen den gemessenen und simulierten Verläufen erkennbar. Das Modell liefert auch bei sehr geringer eingesetzter Menge an Dipeptid eine maximale intrazelluläre Konzentration von mehr als 5 mM, während die Messwerte in diesem Bereich nahezu 0 mM erreichen. Wird dagegen Dipeptid im Bereich von 50% Met-Ala bis hin zu 100% Met-Met eingesetzt, gibt die Simulation 11-8 die experimentellen Verläufe sehr gut wieder.

Eine mögliche Erklärung für die Unterschiede im Bereich geringer Konzentrationen, sind unterschiedliche Parametersätze für Szenarien mit einer sehr hohen und Szenarien mit einer sehr geringen intrazellulären Konzentration. Die intrazelluläre Konzentration dient, durch die Bindung an den Biosensor, als Regulator für einige im Modell betrachtete Reaktionen, u.A.

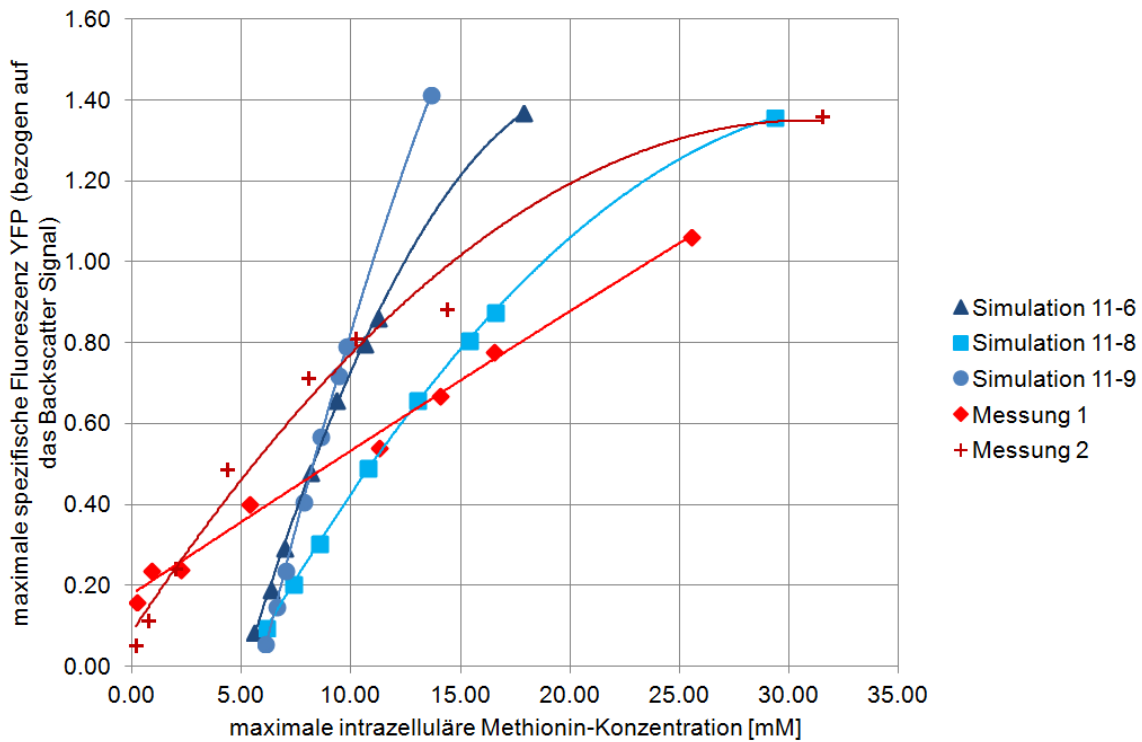


Abbildung 7.6: Kalibrierung: Gegenüberstellung der Messdaten (D.2) in rot und Simulationen in blau. Die Simulationen wurden für die drei Parameterkombinationen der Läufe 6, 8 und 9 der Optimierung 11 durchgeführt. Alle Kalibrierkurven enthalten Mess- bzw. Simulationenwerte für folgende Dipeptid-Anteile: 100% Met-Met, 100% Ala-Met, 90% Ala-Met, 70% Ala-Met, 50% Ala-Met, 30% Ala-Met, 20% Ala-Met, 10% Ala-Met. Die Kurven zwischen den Messdaten beschreiben eine quadratische Funktion auf Basis der jeweiligen Messpunkte.

für den Exporter. Weiterhin ist es möglich, dass biochemische Schwellwerte für Enzyme, welche nicht im Modell abgebildet werden, das Systemverhalten beeinflussen.

7.4 Zusammenfassung und Diskussion

Um Fluoreszenzdaten eines Biosensors kalibrieren und damit in Konzentrationen umrechnen zu können wurde ein biochemisches Modell vorgestellt. Das Modell beschreibt den Zusammenhang zwischen einem vorgelegten Dipeptid, der intrazellulären Konzentration und einem Fluoreszenzsignal auf Basis von Kinetiken.

Das Modell wurde für das reale System in *C. glutamicum* mit einem Lrp-Sensor für die Aminosäure Methionin aus Literaturdaten und in Zusammenarbeit mit der AG Frunzke (IBG-1) aufgebaut. Nach mehreren Validierungs-Durchgängen des Modells mit experimentellen Daten, konnten Parameterkombinationen identifiziert werden, mit welchen das Modell die experimentellen Verläufe gut wiedergeben kann.

Beim Testen des Modells als Kalibrierfunktion wurden Defizite im Bereich geringer Fluoreszenzen ersichtlich. Um eine präzisere Kalibrierung durchführen zu können, sollte das Modell mit dynamischen Messdaten, welche diesen Bereich abdecken, erneut angepasst werden.

Das entwickelte Modell stellt einen Startpunkt für die modellbasierte Kalibrierung von Fluoreszenzdaten dar, bedarf jedoch noch weiterer Optimierung. Die nächsten Schritte der Modellanpassungen setzen darauf auf, dass das Modell in der Lage sein muss, verschiedene Szenarien parallel abbilden zu können. Die Kalibrierung des Modells mit nur einem dieser Szenarien führt zu mehreren optimalen Parametersätzen. Eine Schärfung der Ergebnisse lässt sich eventuell mit einer Ensemble Modellierung mit Datensätzen aus verschiedenen Szenarien erreichen. Auch können Änderungen der Modellstruktur, z.B. durch die Integration stochastischer Prozesse und Schwellwerte, die Qualität der Modellvorhersagen verbessern.

Kapitel 8

Gesamtzusammenfassung und Ausblick

Die Analyse der Qualität von Daten ist ein wichtiges Feld in der Systembiologie, da Daten am Schnittpunkt zwischen experimentellen und theoretischen Betrachtungen liegen und damit einen großen Einfluss auf beide Bereiche haben. Die Interdisziplinarität der Systembiologie ist der Grund dafür, dass Daten zwischen Kollegen unterschiedlicher Fachrichtungen weitergegeben werden, was eine valide und gut dokumentierte Datenprozessierung und -auswertung mit integrierter Fehlerbetrachtung unabdingbar macht.

Im Mittelpunkt dieser Dissertation steht die Methode der modellbasierten Aufbereitung von Rohdaten zu Nutzdaten. Ein spezieller Fokus wurde dabei auf die Fehlerfortpflanzung gelegt. Das Modell als ganzheitliche Beschreibung der biologischen bzw. biotechnologischen Vorgänge bietet neben der Datenprozessierung viele Möglichkeiten. Die beschriebenen Vorgänge können mit statistischen Methoden analysiert und Hypothesen falsifiziert und generiert werden. Ein valides Modell kann für das Design neuer Experimente und zur Optimierung der experimentellen Prozesse eingesetzt werden.

Zum Einstieg in die Methode wurde ein Workflow beschrieben, welcher die Herangehensweise aufzeigt. Anschließend wurde der Workflow auf unterschiedliche Fragestellungen und Datentypen der Systembiologie angewendet

Extrazelluläre Raten stellen eine wichtige Eingangsgröße bei Stoffflussanalysen dar. In Kapitel 4 werden diese mit einem biotechnologischen Prozessmodell abgeschätzt und über lineare sowie nichtlineare Statistik mit Standardabweichungen versehen.

Wachstumsraten aus Daten aus Mikrofermentationssystemen automatisiert abzuschätzen ist Inhalt des Kapitel 5. Dabei findet ein algebraisches Gleichungssystem zur Beschreibung der Datenauswertung Anwendung, in welchem eine Kalibrierfunktion integriert ist. Das komplette Modellsystem gibt Hinweise über die Genauigkeit von Nutzdaten und deckt Limitierungen der Anwendung auf.

Die Quantifizierung von Konzentrationen steht im Mittelpunkt des Kapitel 6. Dafür wurde ein algebraisches Modell der Probenprozessierung erstellt und in eine komplette nichtlineare Fehlerfortpflanzung mittels MC-Simulationen gebettet. Das Modell kann zur validen Datenverarbeitung eingesetzt werden und liefert durch eine Sensitivitätsanalyse Stellschrauben zur Optimierung des Prozesses.

In Kapitel 7 wird ein DAE-System zur Kalibrierung von mittels Biosensoren erzeugten Fluoreszenzdaten vorgestellt. Nach der Optimierung der kinetischen Parameter wurden Defizite in der Vorhersagefähigkeit des Modells festgestellt. Aus diesem Grund ist das Modell nur bedingt für Kalibrierungen einsetzbar und sollte weiter angepasst werden.

Es kamen verschiedene Modellansätze und Analysemethoden zum Einsatz, welche durch ihre Modularität schnell auf andere Datentypen angepasst werden können. Somit stellt diese Dissertation eine Grundlage dafür dar, Datenqualität weiter in den Fokus der Systembiologie zu rücken und Prozesse der Datengenerierung und -aufbereitung modellbasiert zu analysieren und zu optimieren.

Ein weiteres Einsatzgebiet für diese Art der Modelle liegt beim Stichwort Hochdurchsatz-Experimente in der Kopplung der Modelle an Onlinemessungen. Damit könnten diese direkt als Qualitäts-Indikator im Labor Anwendung finden.

Abbildungsverzeichnis

1.1	Systembiologischer Zyklus	2
1.2	Schematische Darstellung der Terme Richtigkeit und Präzision einer Messung.	3
3.1	Analyse der einzelnen Schritte im Workflow zur Nutzdatengenerierung.	22
3.2	Das Modell - eine formale Beschreibung des Workflows zur Nutzdatengenerierung.	23
3.3	Einbetten des Modells in eine Fehlerfortpflanzungsroutine.	23
3.4	Vorgeschalteter Rohdatenfilter zur Gewährleistung reproduzierbarer Nutzdaten.	24
4.1	Schätzung extrazellulärer Raten auf Basis eines Prozessmodells.	27
4.2	Netzwerkdarstellung des Systems für die Prozessmodellierung.	28
4.3	Zeitliche Verläufe von Experiment und Simulation.	31
4.4	Ablauf der Bestimmung von extrazellulären Rate und ihren Standardabweichungen auf Basis linearisierter Statistik.	32
4.5	Ablauf der Bestimmung von extrazellulären Raten und ihrer Standardabweichungen auf Basis von Monte-Carlo Bootstrapping.	35
5.1	Mikrotiterplatte mit 48 Kavitäten.	40
5.2	Messprinzip der Methode der Rückstreuung zur Bestimmung der Biomasse im BioLector®	40
5.3	Herausforderung bei der Biomasse-Quantifizierung im BioLector®.	41
5.4	Workflow der Auswertung des Backscatter Signals aus dem BioLector®.	43
5.5	Kalibrierfunktion zur Quantifizierung des optischen Backscatter Signals aus dem BioLector®	45
5.6	Ideale Wachstumskurve einer Population im Batchansatz.	47
5.7	Übersicht der Prozessierungsschritte von den Rohdaten zu den Nutzdaten.	48
5.8	Benutzeroberfläche der entwickelten MATLAB-Anwendung zur Auswertung von BioLector®-Daten.	50

5.9	Zusammenhang zwischen Biotrockenmasse und Backscatter Signal für verschiedene Füllstände.	51
5.10	Einfluss des Füllstandes auf das Backscatter Signal für verschiedene konstante Biotrockenmassen.	52
5.11	Relative Standardabweichungen der Backscatter Signale.	53
5.12	Abstrahierte Darstellung der relative Standardabweichungen der Backscatter Signale.	54
6.1	Experimenteller Ablauf der Datenprozessierung zur Gewinnung von intrazellulären Metabolitkonzentrationen am IBG-1 des Forschungszentrum Jülich.	58
6.2	Darstellung der Leakage-Korrektur.	66
6.3	Überprüfen der Peakflächenquotienten auf Einhaltung des durch die Standardmessung bestimmten Linearitätsbereiches.	69
6.4	Veränderung der intrazellulären Metabolitkonzentration bei kumulativer Korrektur der systematischen Fehler.	71
6.5	Schematische Darstellung des Modells zur Auswertung von Metabolom-Daten inklusive Fehlerfortpflanzung.	73
6.6	Sensitivität der Varianz der intrazellulären Metabolitkonzentration in Bezug auf die Varianzen aller Messgrößen.	76
7.1	Biosensor-basiertes Hochdurchsatz FACS Screening.	78
7.2	Darstellung der im Sensormodell integrierten biochemischen Prozesse.	81
7.3	Simulationsergebnisse nach multipler Parameteranpassung.	87
7.4	Ausgewählte Simulationsergebnisse nach der Optimierung.	87
7.5	Sensitivitätsmatrix 11-9.	90
7.6	Kalibrierung: Vergleich Messdaten und Simulationen.	93
D.1	Ergebnisse der Optimierung 11.	lxi
D.2	Sensitivitätsmatrix 11-6.	lxiv

Tabellenverzeichnis

2.1	Übersicht über die etablierten Vorgehensweisen und Methoden in der Systembiologie.	8
4.1	Symbolverzeichnis zum Abschnitt Analyse von Metabolomdaten.	26
4.2	Mittels Parameteranpassung geschätzte kinetische Parameter.	29
4.3	Durch Modellsimulation erzeugte extrazelluläre Raten zu zwei ausgewählten Zeitpunkten nach klassischer Parameteranpassung.	30
4.4	Mittels linearisierter Statistik bestimmte Standardabweichungen der kinetischen Parameter.	33
4.5	Geschätzte extrazelluläre Raten und ihre Standardabweichungen bestimmt mit dem gemischten Ansatz aus linearisierter Statistik und MC Simulation.	34
4.6	Durch MC Bootstrapping geschätzte extrazelluläre Raten und ihre Standardabweichungen.	35
4.7	Vergleich der geschätzten extrazellulären Raten zum ersten Zeitpunkt.	36
4.8	Vergleich der geschätzten extrazellulären Raten zum zweiten Zeitpunkt.	37
5.1	Symbolverzeichnis zum Abschnitt Analyse von BioLector®-Daten.	42
5.2	Parameter der quadratischen Kalibrierfunktionen.	46
6.1	Symbolverzeichnis zum Abschnitt Analyse von Metabolomdaten.	56
6.2	Symbolverzeichnis zum Abschnitt Analyse von Metabolomdaten (Fortsetzung).	57
6.3	Messgrößen des Protokolls der Metabolom Probenprozessierung und ihre Fehler.	74
6.4	Intrazelluläre Metabolikonzentrationen bestimmt durch die Anwendung der zwei Modellansätze der Datenprozessierung.	75
7.1	Symbolverzeichnis zum Abschnitt Sensormodellierung.	79
7.2	Symbolverzeichnis zum Abschnitt Sensormodellierung: Modell-Parameter.	80
7.3	Ausgewählte Optimierungsergebnisse.	88

7.4	Parameterstandardabweichungen zu ausgewählten Optimierungen.	91
A.1	Zeitaufgelöste extrazelluläre Konzentrationen.	xxix
A.2	Zeitaufgelöste extrazelluläre Konzentrationen (Fortsetzung).	xxix
B.1	Experimentelle Daten zur Erstellung der Kalibrierfunktion.	xxxi
B.2	Standardabweichung der experimentellen Daten zur Erstellung der Kalibrierfunktion.	xxxii
B.3	Parameter der linearen Kalibrierfunktion.	xxxiii
C.1	Metabolom-Datensatz: Standard-Messungen (1/5).	xlvi
C.2	Metabolom-Datensatz: Standard-Messungen (2/5).	xlvi
C.3	Metabolom-Datensatz: Standard-Messungen (3/5).	xlix
C.4	Metabolom-Datensatz: Standard-Messungen (4/5).	l
C.5	Metabolom-Datensatz: Standard-Messungen (5/5).	li
C.6	Metabolom-Datensatz: Messung Biovolumen	li
C.7	Metabolom-Datensatz: Peakflächen (1/5).	li
C.8	Metabolom-Datensatz: Peakflächen (2/5).	lii
C.9	Metabolom-Datensatz: Peakflächen (3/5).	liii
C.10	Metabolom-Datensatz: Peakflächen (4/5).	liv
C.11	Metabolom-Datensatz: Peakflächen (5/5).	lv
C.12	Modell-basierte Analyse der Veränderung der intrazellulären Metabolitkonzentration bei kumulativer Korrektur der systematischen Fehler.	lvi
C.13	Ergebnisse der Sensitivitätsanalyse (Volumina).	lvii
C.14	Ergebnisse der Sensitivitätsanalyse (Peakflächenquotienten).	lviii
D.1	Übersicht über integrierte experimentelle Daten: Datensatz 1.	lx
D.2	Übersicht über integrierte experimentelle Daten: Datensatz 2.	lxi
D.3	Übersicht über alle Optimierungen.	lxii
D.4	Ergebnisse der Optimierung 11.	lxiii

Literaturverzeichnis

- ALTINTAS, M. M., C. K. EDDY, M. ZHANG, J. D. MCMILLAN und D. S. KOMPALA: *Kinetic modeling to optimize pentose fermentation in Zymomonas mobilis*. *Biotechnology and Bioengineering* 94(2):273–295, 2006
- BAIDOO, E. E. K., P. I. BENKE und J. D. KEASLING: *Mass spectrometry-based microbial metabolomics*. *Methods in Molecular Biology* 881:215–78, 2012
- BAILEY, J. E.: *Toward a science of metabolic engineering*. *Science* 252:1668–1675, 1991
- BAILEY, J. E.: *Mathematical modeling and analysis in biochemical engineering: past accomplishments and future opportunities*. *Biotechnology Progress* 14(1):8–20, 1998
- BANGA, J. R.: *Optimization in computational systems biology*. *BMC Systems Biology* 2:47, 2008
- BINDER, S. et al.: *A high-throughput approach to identify genomic variants of bacterial metabolite producers at the single-cell level*. *Genome Biology* 13(5):R40, 2012
- BOLTEN, C. J., P. KIEFER, F. LETISSE, J.-C. PORTAIS und C. WITTMANN: *Sampling for metabolome analysis of microorganisms*. *Analytical Chemistry* 79(10):3843–3849, 2007
- BÜSCHER, J. M. et al.: *Global network reorganization during dynamic adaptations of Bacillus subtilis metabolism*. *Science* 335(6072):1099–1103, 2012
- BUZIOL, S. et al.: *New bioreactor-coupled rapid stopped-flow sampling technique for measurements of metabolite dynamics on a subsecond time scale*. *Biotechnology and Bioengineering* 80(6):632–6, 2002
- CANELAS, A. B. et al.: *Leakage-free rapid quenching technique for yeast metabolomics*. *Metabolomics* 4(3):226–239, 2008
- CHASSAGNOLE, C., N. NOISOMMIT-RIZZI, J. W. SCHMID, K. MAUCH und M. REUSS: *Dynamic modeling of the central carbon metabolism of Escherichia coli*. *Biotechnology and Bioengineering* 79(1):53–73, 2002

EFRON, B.: *Bootstrap Methods: Another Look at the Jackknife*. *Annals of Statistics* 7(1):1–26, 1979

FAIJES, M., A. E. MARS und E. J. SMID: *Comparison of quenching and extraction methodologies for metabolome analysis of Lactobacillus plantarum*. *Microbial Cell Factories* 6:27, 2007

FEIST, A. M. et al.: *A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information*. *Molecular Systems Biology* 3:121, 2007

FUNKE, M., S. DIEDERICHS, F. KENSY, C. MÜLLER und J. BÜCHS: *The baffled microtiter plate: increased oxygen transfer and improved online monitoring in small scale fermentations*. *Biotechnology and Bioengineering* 103(6):1118–28, 2009

GANS, P.: *Data Fitting in the Chemical Sciences*. Chichester, UK: John Wiley & Sons, 1992: 270.

GLACKEN, M. W., E. ADEMA und A. J. SINSKE: *Mathematical descriptions of hybridoma culture kinetics: I. Initial metabolic rates*. *Biotechnology and Bioengineering* 32(4):491–506, 1988

GOUDAR, C. T.: *Computer programs for modeling mammalian cell batch and fed-batch cultures using logistic equations*. *Cytotechnology* 64(4):465–475, 2012

GRIEWANK, A. und A. WALTHER: *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*. 2. Aufl. SIAM, 2008: 105. ISBN: 978-0-89871-659-7.

HANKE, T. et al.: *Combined fluxomics and transcriptomics analysis of glucose catabolism via a partially cyclic pentose phosphate pathway in Gluconobacter oxydans 621H*. *Applied and Environmental Microbiology* 79(7):2336–48, 2013

HARRIS, C. M. et al.: *Dielectric permittivity of microbial suspensions at radio frequencies: a novel method for the real-time estimation of microbial biomass*. *Enzyme and Microbial Technology* 9(3):181–186, 1987

HENRI, V.: *Theorie generale de l'action de quelques diastases*. *Comptes rendues l'Academie des sciences* 135(1902):916–919, 1902

IKEDA, M.: *Amino acid production processes*. *Advances in Biochemical Engineering/Biotechnology* 79:1–35, 2003

ISHIZAKI, A., K. YAMAMOTO und Y. FURUTA: *A new method for the accurate and rapid determination of the concentrations of intracellular metabolites in cells during fermentation*. Biotechnology Techniques 9(6):409–412, 1995

JESSOME, L. L. und D. A. VOLMER: *Ion Suppression: A Major Concern in Mass Spectrometry*. LCGC 24(5), 2006

JOSHI, M, A. SEIDEL-MORGENSTERN und A. KREMLING: *Exploiting the bootstrap method for quantifying parameter confidence intervals in dynamical systems*. Metabolic Engineering 8(5):447–455, 2006

KÄSS, F.: *Scale-Down of oxygen supply in bioprocess development with Corynebacterium glutamicum*. Dissertation. Technischen Universität Berlin, 2015.

KÄSS, F. et al.: *Rapid assessment of oxygen transfer impact for Corynebacterium glutamicum*. Bioprocess and Biosystems Engineering 37(12):2567–2577, 2014

KEILHAUER, C, L. EGGELING und H. SAHM: *Isoleucine synthesis in Corynebacterium glutamicum: molecular analysis of the ilvB-ilvN-ilvC operon*. Journal of Bacteriology 175(17):5595–5603, 1993

KENNERKNECHT, N. et al.: *Export of L-Isoleucine from Corynebacterium glutamicum: a Two-Gene-Encoded Member of a New Translocator Family*. Journal of Bacteriology 184(14):3947–3956, 2002

KENSY, F., C. ENGELBRECHT und J. BÜCHS: *Scale-up from microtiter plate to laboratory fermenter: evaluation by online monitoring techniques of growth and protein expression in Escherichia coli and Hansenula polymorpha fermentations*. Microbial Cell Factories 8:68, 2009

KENSY, F., E. ZANG, C. FAULHAMMER, R.-K. TAN und J. BÜCHS: *Validation of a high-throughput fermentation system based on online monitoring of biomass and fluorescence in continuously shaken microtiter plates*. Microbial Cell Factories 8:31, 2009

KINOSHITA, S., S. UDAKA und M. SHIMONO: *Studies on the amino acid fermentation: I. Production of L-glutamic acid by various microorganisms*. The Journal of General and Applied Microbiology 3:193–205, 1957

KITANO, H.: *Systems biology: a brief overview*. Science 295(5560):1662–1664, 2002

KJELDSSEN, K. R. und J. NIELSEN: *In silico genome-scale reconstruction and validation of the Corynebacterium glutamicum metabolic network*. Biotechnology and Bioengineering 102(2):583–597, 2009

KLIPP, E. et al.: *Systems Biology: A Textbook*. 2009: 569. ISBN: 3527318747. DOI: 10.3797/scipharm. arXiv:arXiv:1011.1669v3.

KONING, W. de und K. VAN DAM: *A method for the determination of changes of glycolytic metabolites in yeast on a subsecond time scale using extraction at neutral pH*. Analytical Biochemistry 204(1):118–123, 1992

KREMLING, A.: *Kompendium Systembiologie - Mathematische Modellierung und Modellanalyse*. Vieweg + Teubner, 2012.

LANGE, C. et al.: *Lrp of Corynebacterium glutamicum controls expression of the brnFE operon encoding the export system for l-methionine and branched-chain amino acids*. Journal of Biotechnology, 2011

LIANG, H. R., R. L. FOLTZ, M MENG und P BENNETT: *Ionization enhancement in atmospheric pressure chemical ionization and suppression in electrospray ionization between target drugs and stable-isotope-labeled internal standards in quantitative liquid chromatography/tandem mass spectrometry*. Rapid Communications in Mass Spectrometry 17(24):2815–21, 2003

LIEBERMEISTER, W. und E. KLIPP: *Bringing metabolic networks to life: convenience rate law and thermodynamic constraints*. Theoretical Biology and Medical Modelling 3:41, 2006

LINK, H., B. ANSELMANT und D. WEUSTER-BOTZ: *Leakage of adenylates during cold methanol/glycerol quenching of Escherichia coli*. Metabolomics 4(3):240–247, 2008

MAGNUS, J. B., D. HOLLWEDEL, M. OLDIGES und R. TAKORS: *Monitoring and modeling of the reaction dynamics in the valine/leucine synthesis pathway in Corynebacterium glutamicum*. Biotechnology Progress 22(4):1071–1083, 2006

MAHARJAN, R. P. und T. FERENCI: *Global metabolite analysis: the influence of extraction methodology on metabolome profiles of Escherichia coli*. Analytical Biochemistry 313(1):145–54, 2003

MEINERT, S. et al.: *Quantitative quenching evaluation and direct intracellular metabolite analysis in Penicillium chrysogenum*. Analytical Biochemistry 438(1):47–52, 2013

- MICHAELIS, L. und M. L. MENTEN: *Die Kinetik der Invertin-Wirkung*. Biochemische Zeitschrift 49:333–369, 1913
- MURPHY, T. A. und J. D. YOUNG: *ETA: Robust software for determination of cell specific rates from extracellular time courses*. Biotechnology and Bioengineering 100(6), 2013
- MUSTAFAI, N., A. GRÜNBERGER, D. KOHLHEYER, M. BOTT und J. FRUNZKE: *The development and application of a single-cell biosensor for the detection of l-methionine and branched-chain amino acids*. Metabolic Engineering:1–9, 2012
- NOACK, S. und W. WIECHERT: *Quantitative metabolomics: a phantom?* Trends in Biotechnology 32(5):238–244, 2014
- OLDIGES, M. et al.: *Metabolomics: current state and evolving methodologies and tools*. Applied Microbiology and Biotechnology 76(3):495–511, 2007
- ORTH, J. D., I. THIELE und B. Ø. PALSSON: *What is flux balance analysis?* Nature Biotechnology 28(3):245–248, 2010
- PACZIA, N.: *Metabolomanalyse von Mikroorganismen - Optimierung und Konzepte zur Automatisierung*. Dissertation. Universität Bielefeld, 2012.
- PACZIA, N. et al.: *Extensive exometabolome analysis reveals extended overflow metabolism in various microorganisms*. Microbial Cell Factories 11(1):122, 2012
- PAPAGEORGIU, M: *Optimierung: Statische, dynamische, stochastische Verfahren für die Anwendung*. 2. Aufl. München: R. Oldenbourg Verlag GmbH, 1996.
- PRÄVE, P., U. FAUST, W. SITTIG und D. A. SUKATSCH: *Handbuch der Biotechnologie*. Oldenbourg Industrieverlag, 1994: 1045.
- PRUST, C. et al.: *Complete genome sequence of the acetic acid bacterium Gluconobacter oxydans*. Nature Biotechnology 23(2):195–200, 2005
- RIZZI, M., M. BALTES, U. THEOBALD und M. REUSS: *In vivo analysis of metabolic dynamics in Saccharomyces cerevisiae: II. Mathematical model*. Biotechnology and Bioengineering 55(4):592–608, 1997
- ROELS, J. A.: *Energetics and Kinetics in Biotechnology*. Elsevier Biomedical Press, 1983: 330. ISBN: 9780444804426.

RÖNSCH, H., R. KRÄMER und S. MORBACH: *Impact of osmotic stress on volume regulation, cytoplasmic solute composition and lysine production in Corynebacterium glutamicum MH20-22B*. Journal of Biotechnology 104(1-3):87–97, 2003

SÁEZ, M. J. und R. LAGUNAS: *Determination of intermediary metabolites in yeast - critical-examination of effect of sampling conditions and recommendations for obtaining true levels*. Molecular and Cellular Biochemistry 13(2):73–78, 1976

SALTELLI, A, K CHAN und E. M. SCOTT: *Sensitivity Analysis*. John Wiley & Sons, 2000.

SAMORSKI, M, G. MÜLLER-NEUEN und J. BÜCHS: *Quasi-Continuous Combined Scattered Light and Fluorescence Measurements: A Novel Measurement Technique for Shaken Microtiter Plates*. Biotechnology and Bioengineering 92(1):61–68, 2005

SAUER, U.: *Metabolic networks in motion: 13C-based flux analysis*. Molecular Systems Biology 2:62, 2006

SAVAGEAU, M. A. und E. O. VOIT: *Power-law approach to modeling biological systems I. Theory*. Journal of Fermentation Technology 60:221–228, 1982

SCHÄDEL, F. und E. FRANCO-LARA: *Rapid sampling devices for metabolic engineering applications*. Applied Microbiology and Biotechnology 83(2):199–208, 2009

SHIN, M. H., D. Y. LEE, K.-H. LIU, O. FIEHN und K. H. KIM: *Evaluation of sampling and extraction methodologies for the global metabolic profiling of Saccharophagus degradans*. Analytical Chemistry 82(15):6660–6, 2010

SONNLEITNER, B., G. LOCHER und A. FIECHTER: *Biomass determination*. Journal of Biotechnology 25(1-2):5–22, 1992

TAYLOR, J. R.: *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*. University Science Books, 1997: 327. ISBN: 9780935702750.

TAYMAZ-NIKEREL, H. et al.: *Development and application of a differential method for reliable metabolome analysis in Escherichia coli*. Analytical Biochemistry 386(1):9–19, 2009

TRÖTSCHEL, C., D. DEUTENBERG, B. BATHE, A. BURKOVSKI und R. KRÄMER: *Characterization of Methionine Export in Corynebacterium glutamicum*. Journal of Bacteriology 187(11):3786–3794, 2005

VAN GULIK, W. M. et al.: *Fast Sampling of the Cellular Metabolome*. In: *Microbial Systems Biology: Methods and Protocols*. Hrsg. von A. NAVID. Bd. 881. Methods in Molecular Biology. Totowa, NJ: Humana Press, 2012. Kap. 10:279–306. ISBN: 978-1-61779-826-9. DOI: 10.1007/978-1-61779-827-6.

VAN OUYEN, J., S. NOACK, M. BOTT, A. RETH und L. EGGELING: *Improved L-lysine production with Corynebacterium glutamicum and systemic insight into citrate synthase flux and activity*. Biotechnology and Bioengineering:1–28, 2012

VASEGHI, S., A. BAUMEISTER, M. RIZZI und M. REUSS: *In Vivo Dynamics of the Pentose Phosphate Pathway in Saccharomyces cerevisiae*. Metabolic Engineering 1:128–140, 1999

VILLAS-BÔAS, S. G., J. HØJER-PEDERSEN, M. AKESSON, J. SMEDSGAARD und J. NIELSEN: *Global metabolite analysis of yeast: evaluation of sample preparation methods*. Yeast 22(14):1155–69, 2005

VISSER, D., J. W. SCHMID, K. MAUCH, M. REUSS und J. J. HEIJNEN: *Optimal re-design of primary metabolism in Escherichia coli using linlog kinetics*. Metabolic Engineering 6(4):378–390, 2004

VOGES, R. und S. NOACK: *Quantification of proteome dynamics in Corynebacterium glutamicum by 15N-labeling and selected reaction monitoring*. Journal of Proteomics 75(9):2660–2669, 2012

VRLJIC, M., H. SAHM und L. EGGELING: *A new type of transporter with a new type of cellular function: L-lysine export from Corynebacterium glutamicum*. Molecular Microbiology 22:815–826, 1996

WELLERDIEK, M. G., D. WINTERHOFF, W. REULE, J. BRANDNER und M. OLDIGES: *Metabolic quenching of Corynebacterium glutamicum: efficiency of methods and impact of cold shock*. Bioprocess and Biosystems Engineering 32(5):581–592, 2009

WESTERHOFF, H. V. und K. VAN DAM: *Thermodynamics and control of biological free-energy transduction*. Amsterdam: Elsevier, 1987.

WIECHERT, W.: *13C metabolic flux analysis*. Metabolic Engineering 3(3):195–206, 2001

WIECHERT, W.: *Systembiologie: eine interdisziplinäre Herausforderung*. Schöningh, Paderborn, 2004.

WILLIAMS, J. a.: *Keys to bioreactor selections*. Chemical Engineering Progress 98(3):34–41, 2002

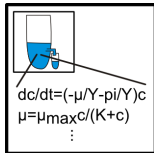
WINDER, C. L. et al.: *Global metabolic profiling of Escherichia coli cultures: an evaluation of methods for quenching and extraction of intracellular metabolites*. Analytical Chemistry 80(8):2939–48, 2008

WU, L. et al.: *Quantitative analysis of the microbial metabolome by isotope dilution mass spectrometry using uniformly ¹³C-labeled cell extracts as internal standards*. Analytical Biochemistry 336(2):164–71, 2005

Anhang A

Extrazelluläre Raten

A.1 Prozessmodell



Differentialgleichungen:

$$\frac{dc_x}{dt} = (\mu_{GLC} + \mu_{GLCN}) \cdot c_x,$$
$$c_x(0) = c_{x,0}$$

$$\frac{dc_{GLC}}{dt} = \left(-\frac{\mu_{GLC}}{Y_{x,GLC}} - \frac{\pi_{GLK}}{Y_{GLK,GLC}} - \frac{\pi_{VKGA,GLC}}{Y_{VKGA,GLC}} \right) \cdot c_x,$$
$$c_{GLC}(0) = c_{GLC,0}$$

$$\frac{dc_{GLK}}{dt} = \left(-\frac{\mu_{GLK}}{Y_{x,GLK}} + \pi_{GLK} - \frac{\pi_{IIKGA,GLK}}{Y_{IIKGA,GLK}} - \frac{\pi_{VKGA,GLK}}{Y_{VKGA,GLK}} \right) \cdot c_x,$$
$$c_{GLK}(0) = c_{GLK,0}$$

$$\frac{dc_{IIKGA}}{dt} = \pi_{IIKGA,GLK} \cdot c_x,$$
$$c_{IIKGA}(0) = c_{IIKGA,0}$$

$$\frac{dc_{VKGA}}{dt} = (\pi_{VKGA,GLK} + \pi_{VKGA,GLC}) \cdot c_x,$$
$$c_{VKGA}(0) = c_{VKGA,0}$$

Monod-Kinetiken:

$$\mu_{GLC} = \mu_{max,GLC} \cdot \frac{c_{GLC}}{K_{GLC,X} + c_{GLC}}$$

$$\mu_{GLK} = \mu_{max,GLK} \cdot \frac{c_{GLK}}{K_{GLK,X} + c_{GLK}} \cdot \frac{k^i_{GLC}}{k^i_{GLC} + c_{GLC}}$$

$$\pi_{GLK} = \pi_{max,GLK} \cdot \frac{c_{GLC}}{K_{GLC,GLK} + c_{GLC}}$$

$$\pi_{IIKGA,GLK} = \pi_{max,IIKGA,GLK} \cdot \frac{c_{GLK}}{K_{IIKGA,GLK} + c_{GLK}}$$

$$\pi_{VKGA,GLK} = \pi_{max,VKGA,GLK} \cdot \frac{c_{GLK}}{K_{VKGA,GLK} + c_{GLK}}$$

$$\pi_{VKGA,GLC} = \pi_{max,VKGA,GLC} \cdot \frac{c_{GLC}}{K_{VKGA,GLC} + c_{GLC}}$$

Bildungsrate von Kohlendioxid:

$$r_{CO_2} = \frac{\dot{V}_g}{V_r \cdot 27.47} \left(Y_{CO_2,out} - Y_{CO_2,in} \frac{1 - Y_{O_2,in} - Y_{CO_2,in}}{1 - Y_{O_2,out} - Y_{CO_2,out}} \right)$$

mit \dot{V}_g als vordefinierte Begasungsrate. $Y_{O_2,in}$, $Y_{CO_2,in}$ und $Y_{O_2,out}$, $Y_{CO_2,out}$ definieren die Anteile von O_2 und CO_2 in Zu- und Abgas.

A.2 Experimentelle Daten

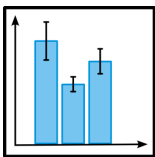


Tabelle A.1: Zeitaufgelöste extrazelluläre Konzentrationen und ihre Standardabweichungen in $[\text{g L}^{-1}]$.

Zeit [h]	GLC		X		GLK	
	val	std	val	std	val	std
0.00	80.6960	0.0000	0.0500	0.0016	0.3100	0.1114
2.00	78.6786	0.0000	0.1100	0.0160	1.3900	0.1875
4.00	68.5209	0.0000	0.2100	0.0049	4.9000	0.3076
6.00	59.1386	0.7401	0.4700	0.0333	17.8600	0.2765
7.00	41.3375	0.7592	0.6600	0.0171	32.5200	2.3611
8.00	27.4675	1.5538	0.8200	0.0000	44.4500	1.2021
9.25	17.9599	0.8193	1.1200	0.0369	61.5300	1.7368
10.00	8.8403	2.5004	1.4900	0.0077	51.6800	1.0229
11.50	4.1780	0.8441	1.6000	0.0190	48.7600	0.1119
13.00	0.0000	0.0000	1.6800	0.0190	48.1100	0.0000
14.25	0.0000	0.0000	1.7100	0.0190	46.8500	1.6487
15.50	0.0000	0.0000	1.7500	0.0380	37.7000	3.6030
25.50	0.0000	0.0000	1.8400	0.0570	6.7700	0.1277
30.00	0.0000	0.0000	1.8600	0.0321	4.5400	0.0469

Tabelle A.2: Zeitaufgelöste extrazelluläre Konzentrationen und ihre Standardabweichungen in $[\text{g L}^{-1}]$ (Fortsetzung).

Zeit [h]	IIKGA		VKGA	
	val	std	val	std
0.00	0.0000	0.0000	0.0000	0.0000
2.00	0.0000	0.0000	0.9600	0.0667
4.00	0.3400	0.0128	2.0700	0.1295
6.00	1.4400	0.0136	3.9700	0.0812
7.00	2.7500	0.0937	5.1600	0.1039
8.00	4.5500	0.0707	6.0500	0.3536
9.25	7.4500	0.2121	6.4200	0.3075
10.00	12.0800	1.1701	5.7200	0.7695
11.50	15.3300	0.2740	6.9000	0.5739
13.00	23.8900	1.9689	6.8100	0.3471
14.25	26.7600	0.5356	7.8400	0.6653
15.50	36.4200	1.1617	7.3000	0.8799
25.50	62.8000	0.5657	7.0900	1.0154
30.00	65.1500	0.4266	7.0900	1.0154

Anhang B

BioLector®-Daten

B.1 Daten für die Kalibrierfunktion

Der Ermittlung der Kalibrierfunktion liegen sieben BioLector®-Experimente zu Grunde. Für jede Kombination von Füllvolumen und Biotrockenmasse wurde das Backscatter Signal in drei Replikaten bestimmt (Experimente von Friedrich Käß; KÄSS, 2015).

Tabelle B.1: Experimentelle Daten zur Erstellung der Kalibrierfunktion: Die Backscatter Signale (in a.u.) sind in Abhängigkeit von Füllvolumen (V in μL) und definierter Biotrockenmasse (BTM in g L^{-1}) dargestellt.

V \ BTM	0	3.33	5.21	7.53	9.33	13.91	17.89	15.91
300	45.95	37.81	66.47	81.41	126.95	209.92	289.72	246.48
400	39.16	58.76	95.64	131.63	180.43	289.09	395.00	353.12
500	52.10	70.99	116.39	143.69	205.98	315.92	437.51	382.56
600	50.37	68.84	113.01	139.22	205.34	312.13	433.20	376.65
700	37.26	63.85	109.17	132.90	201.42	305.92	427.94	370.16
800	25.41	60.58	100.75	129.79	193.87	301.12	422.73	366.32
900	21.15	56.22	96.37	125.26	192.32	300.43	420.69	363.19
1000	16.92	53.92	93.67	122.95	188.58	299.42	418.30	361.30
1100	13.38	54.35	95.19	125.77	191.23	306.08	429.15	368.92
1200	12.57	53.93	95.48	124.12	191.29	306.43	425.56	365.23
1300	13.72	54.23	97.44	124.00	191.81	306.71	421.66	361.99
1400	14.16	54.35	96.77	123.85	194.19	304.27	418.30	359.14
1500	12.30	53.08	97.02	122.40	191.20	303.67	413.67	355.65

Tabelle B.2: Standardabweichung der experimentellen Daten zur Erstellung der Kalibrierfunktion: Die Standardabweichungen der Backscatter Signale (in a.u.) sind in Abhängigkeit von Füllvolumen (V in μL) und definierter Biotrockenmasse (BTM in g L^{-1}) dargestellt.

$V \backslash \Delta\text{BTM}$	0	3.33	5.21	7.53	9.33	13.91	17.89	15.91
300	7.77	2.70	2.78	10.42	6.02	4.62	13.44	16.20
400	3.61	2.02	4.52	8.47	3.58	9.73	19.85	16.94
500	7.78	1.84	3.88	9.18	1.54	2.79	21.15	19.95
600	5.21	1.50	1.58	9.94	1.09	4.21	20.79	20.56
700	4.45	1.01	1.42	10.28	3.05	3.00	22.93	21.78
800	4.20	1.97	1.64	10.66	1.54	2.83	23.19	21.67
900	3.81	2.14	1.57	10.69	3.25	2.96	24.28	22.43
1000	3.00	1.57	1.66	10.38	0.98	3.09	24.65	22.42
1100	2.48	1.91	1.62	10.64	0.62	3.30	25.85	22.86
1200	2.12	1.50	1.27	10.83	0.48	3.69	24.96	22.24
1300	1.13	0.86	1.22	10.74	0.88	3.45	23.83	20.66
1400	2.33	0.89	1.64	10.69	3.03	3.47	23.48	19.48
1500	0.40	1.38	1.19	10.79	1.11	3.71	21.63	17.61

B.2 Lineare Kalibrierfunktionen

Zur Erstellung einer linearen Funktion nach dem Schema $BS = m \cdot \text{BTM} + n$, mit m als Anstieg und n als Achsenabschnitt, wird folgende Berechnungsvorschrift zur gewichteten linearen Regression verwendet:

1. Berechnung der Varianzmatrix

$$\text{VAR}_{BS} = \Delta BS^2 \quad (\text{B.1})$$

2. Berechnung der Systemmatrix in der Vandermonde-Form

$$\mathbf{M}_{SYS} = \begin{pmatrix} 1 \\ 1 \quad \text{BTM} \\ \vdots \end{pmatrix} \quad (\text{B.2})$$

3. Berechnung der Kovarianzmatrix der Regressionsparameter

$$\text{COV}_{\alpha} = \begin{pmatrix} \text{COV}_n & \text{COV}_{n,m} \\ \text{COV}_{m,n} & \text{COV}_m \end{pmatrix} \quad (\text{B.3})$$

$$\text{COV}_{\alpha} = (\mathbf{M}_{SYS}^T \cdot \text{VAR}_{BS}^{-1} \cdot \mathbf{M}_{SYS})^{-1} \quad (\text{B.4})$$

4. Berechnung der Regressionsparameter

$$\text{PARA}_\alpha = \begin{pmatrix} n \\ m \end{pmatrix} \tag{B.5}$$

$$\text{PARA}_\alpha = \text{COV}_\alpha \cdot \mathbf{M}_{SYS}^T \cdot \text{VAR}_{BS}^{-1} \cdot \text{BS} \tag{B.6}$$

Die finalen Parameter können zur Umrechnung von Backscatter Signalen in Biotrockenmasse genutzt werden:

$$BTM = \frac{BS - n}{m} \tag{B.7}$$

Und die Standardabweichung ergibt sich mittels Gauß'scher Fehlerfortpflanzung:

$$\Delta BTM = \sqrt{\frac{-1}{m} \cdot \Delta n^2 - \frac{BS - n}{m^2} \cdot \Delta m^2 + \frac{1}{m} \cdot \Delta BS^2} \tag{B.8}$$

Mit der beschriebenen Methoden wurden die in Tabelle B.3 dargestellten Funktionsparameter und deren Standardabweichungen ermittelt.

Tabelle B.3: Parameter der linearen Kalibrierfunktion: Die experimentellen Daten aus Tabelle B.1 wurden zur Erstellung einer linearen ($BS = m \cdot BTM + n$) Kalibrierfunktion genutzt. Die Kalibrierfunktion kann zur Berechnung von Biotrockenmasse (BTM) aus den mit dem BioLector® erhobenen Backscatter Signalen (BS) eingesetzt werden.

V [L]	m	Δm	n	Δn
300	-4.2195	2.9519	14.8625	0.4084
400	8.9936	2.3651	18.1127	0.4169
500	-1.6314	2.3223	22.4641	0.2722
600	0.0623	1.8406	21.9543	0.2445
700	-6.8543	1.4489	22.1635	0.2547
800	-8.1331	1.9396	21.7092	0.2494
900	-11.8283	1.9551	21.8182	0.2776
1000	-12.2086	1.6443	21.5136	0.2086
1100	-9.0569	1.6375	21.4650	0.1909
1200	-9.6880	1.3623	21.4994	0.1598
1300	-2.7911	0.8416	20.4760	0.1350
1400	-13.2587	1.2547	21.6759	0.2490
1500	10.4313	0.3902	18.7957	0.1069

B.3 Dokumentation des MATLAB-Werkzeuges für BioLector®-Daten

Die Dokumentation befindet sich in der nachfolgend angehängten PDF.

Getting started

Performing BIOLECTOR cultivations produces a huge amount of data. *BioLectorDataProcessing* is a MATLAB tool supporting the evaluation of biomass measurements. Data evaluation consists of different steps. The raw backscatter signals will be calibrated to bio dry weight and logarithmized. The logarithmic curve will be used for growth rate calculations; therefore it will be split in linear sections. The tool is able to handle replicates and includes error propagation.

Starting BIOLECTOR data evaluation the zipped file *BioLectorDataProcessing.zip* has to be unpacked.

Folder structure

- **BioLectorDataProcessing** contains
 - The main MATLAB file *WellSelection.m*.
 - **data** contains the BIOLECTOR files (*.csv) organized in subfolders for different projects organized by yourself; for files which already have been processed the data is stored in a mat-file in this subfolder. Also the result files (*_processed_time-stamp.csv) will be stored in here.
 - **docu** contains figures of the calibration function and this document for documentation.
 - **mat** contains mat-files including variables MATLAB needs for data processing.
 - **mfiles** contains all MATLAB files (*.m, *.fig) needed for data processing.
- Please keep this folder structure! The folders **mfiles** and **mat** are required for data evaluation.

Before data processing with MATLAB

- Create a project subfolder in the data folder. Choose an explicit and traceable folder title.
- Copy the BIOLECTOR file containing the data you want to process into this subfolder.

MATLAB

- Start MATLAB and change the **current folder** to *BioLectorDataProcessing* (cp. Figure 1)
- Start the **MATLAB file** *WellSelection.m* by double click on the m-file in 'current folder'; press the green **Run-button** in the opened editor window or F5 on your keyboard.
- A new window (Figure 2) will appear...next steps in detail in section **Setup**.

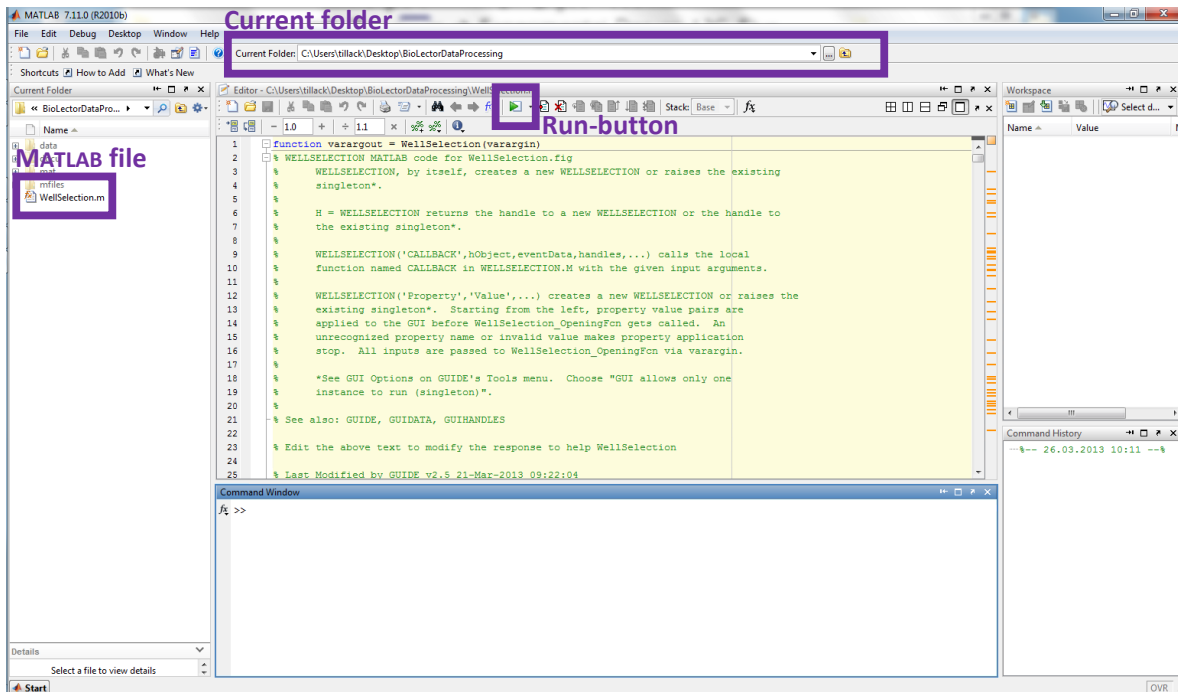


Figure 1: Main MATLAB window.

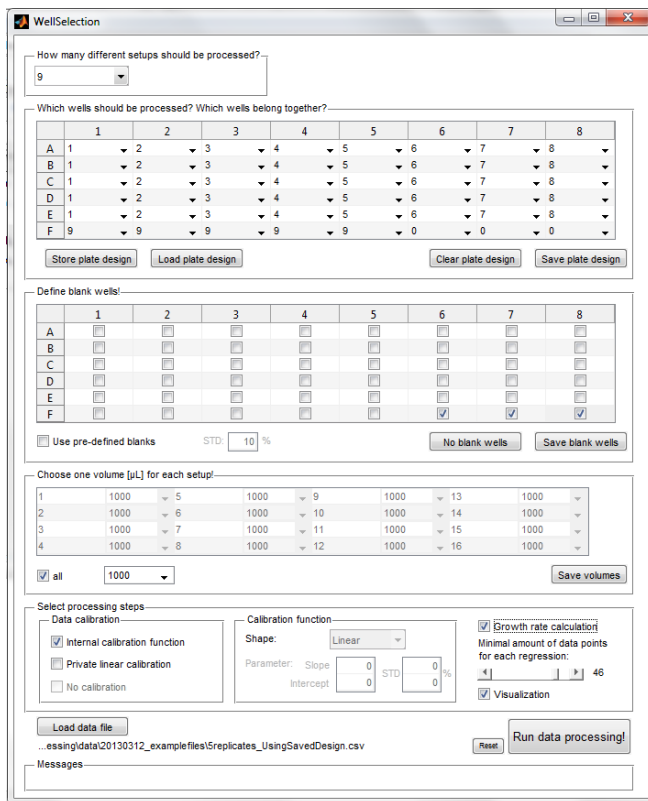



Figure 2: Graphic user interface of the tool to define all process parameters.

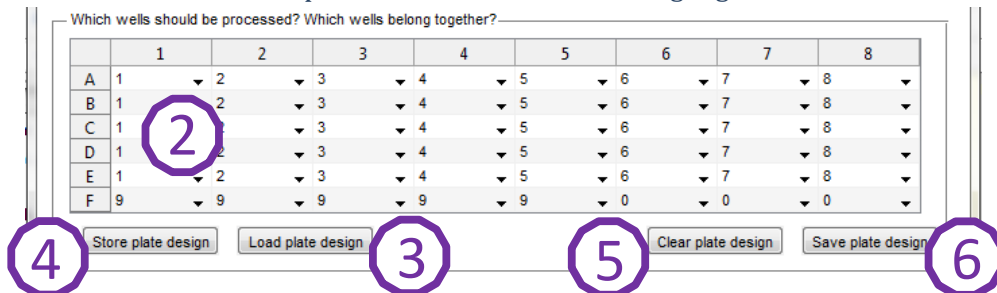
Setup

How many different setups should be processed?

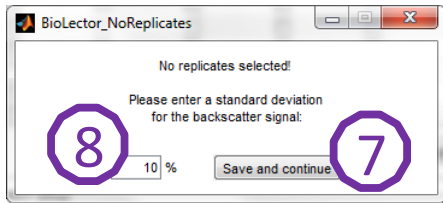


- Choose the amount of different setups (1) you want to compare within this run. e.g. within 48 wells 16 different setups, each with 3 replica.
- One setup should contain at least one well and each setup must contain the same number of wells/replicates!

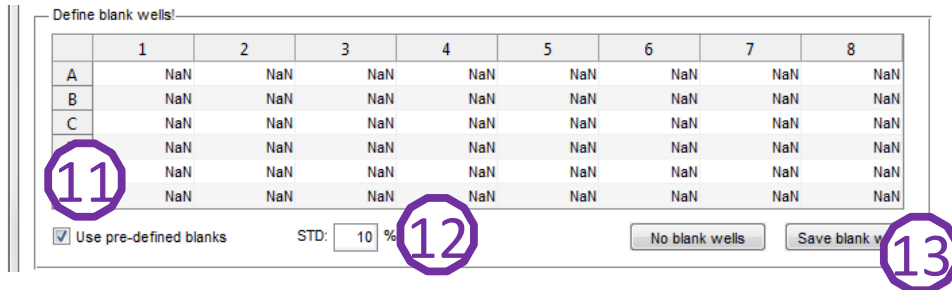
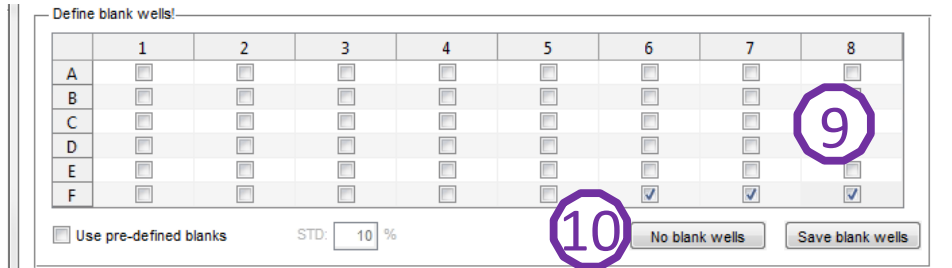
Which wells should be processed? Which wells belong together?



- Define the plate design including the information, which wells should be processed and which wells belong together (2) as they are replicates. All wells which are part of one setup (received the same number), will be processed together afterwards. All wells which are not selected (selected as 0), will not be processed.
- Please note: Allocate the setup numbers sequentially starting with setup 1!
- Definition can happen:
 - By manual selection
 - Loading a pre-defined plate design (3)
(example: ...\\data\\20130312_examplefiles\\PlateDesign_5replicates.mat)
- As soon as a plate design should be used a second time, storing (4) and loading might be reasonable.
- If you have chosen 16, 24 or 48 different setups, a plate design will be defined automatically. It can easily be removed with button (5).
- Finally, the plate design has to be saved (6) for this run and the next section can be filled.
- If no replicates were chosen as all wells are independent cultivations, a new window opens up to define a fixed standard error for the processing of the backscatter signal. It requests to enter a standard deviation for the backscatter signal (8). The default standard deviation is set to 10% and will be used for analysis if the window is closed, not using button (7).

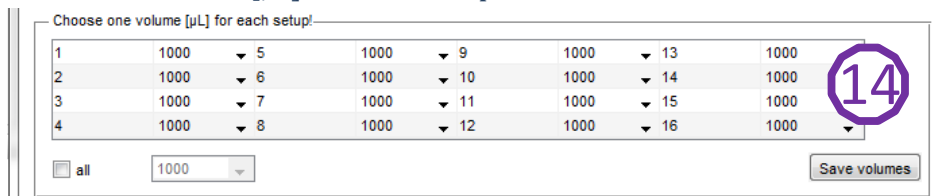


Define blank wells!



- This section should be used to define blanks if available.
- If no blanks are available, choose button (10) and pass on to the next section.
- If the evaluated plate includes blank wells, they can be chosen by marking the relevant wells (9, upper figure). Blank wells will not be evaluated and represent no additional setup.
- Pre-defined blank values (e.g. backscatter value of medium without cells from a previous experiment) can manually be added to each well after selecting this option (11, lower figure).
- For pre-defined blank values a standard deviation has to be added (12, default 10%).
- Finally, the information about blank values has to be saved (13) for this run and the next section can be filled.

Choose one volume [µL] for each setup!



Choose one volume [μL] for each setup!

1	1000	5	1000	9	1000	13	1000
2	1000	6	1000	10	1000	14	1000
3	1000	7	1000	11	1000	15	1000
	1000	8	1000	12	1000	16	1000

all 1000 Save volume

- In this section a filling volume should be assigned to the different setups, as biomass calibration used by this tool is depending on the applied cultivation volume.
- This can be done separately for each setup (14, upper figure).
- Alternatively one volume can be defined for the complete plate after selecting this option (15, lower figure).
- For a run with more than 16 setups there is no possibility to define one separate volume for each setup. If an independent evaluation is desired, please divide the analysis in different runs.
- Save the volumes (16) and go on with the next section.

Select processing steps

Select processing steps

Data calibration

Internal calibration function Private linear calibration No calibration

Calibration function

Shape: Linear

Parameter: Slope 0 STD 0% Intercept 0

Growth rate calculation

Minimal amount of data points for each regression: 46

Visualization

- Select the steps of data processing for this run within this section.
- Transfer the backscatter signal to bio dry weight via calibration (17):
 - ...with the internal linear or non-linear function. The parameter of the internal calibration functions can be seen in the appendix.
 - ...using own parameters of a linear calibration function.
 - ...skip data calibration.

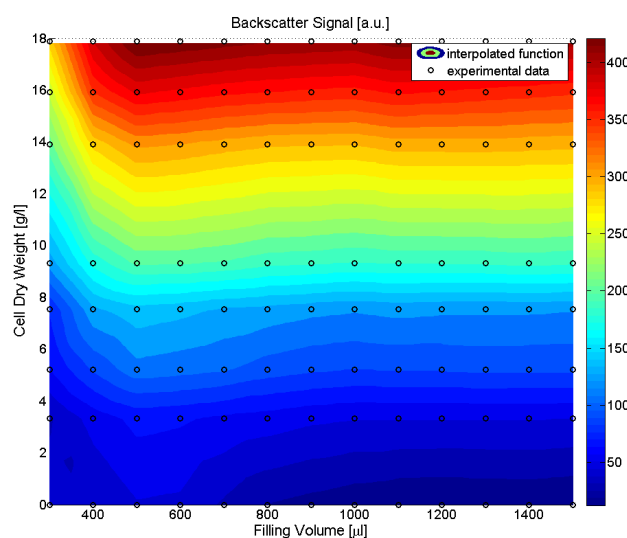


Figure 3: Internal calibration function. Experiments have been done for different combinations of bio dry weight and filling volume (black circles). The rest of the backscatter surface has been filled with interpolation.

- Define the calibration function, if needed.
 - The shape of the internal calibration function can be linear or non-linear (18). For data processing with blank values, only the linear shape can be chosen. This is due to the fact that intercept will be already corrected with the blank values.
 - The parameters intercept and slope have to be defined (19) after selecting the private calibration function. Here, parameters estimated in separate analyses should be added including their standard deviations.
- Estimating the growth rate (20); for this step data calibration will be required. This option is not available for 48 setups due to excel limitations (maximum number of columns). If an independent evaluation is desired, please divide the analysis in different runs.
 - For growth rate estimation a parameter called “minimal amount of data points for each regression” might be adapted. The effect of this parameter to the growth rate estimation can be seen in Figure 4 for an example. Usually, an optimal parameter depends on the total amount of data points, but also the shape of the growth curve might affect the choice. Best practice will be to test with the default parameter, 10 data points, and adapt (if necessary) after taking a look at figure 2 of your MATLAB results.
- Visualizing the different steps of data processing and the growth rate estimation procedure (21). This option is not available for more than 16 setups due to clarity reasons.

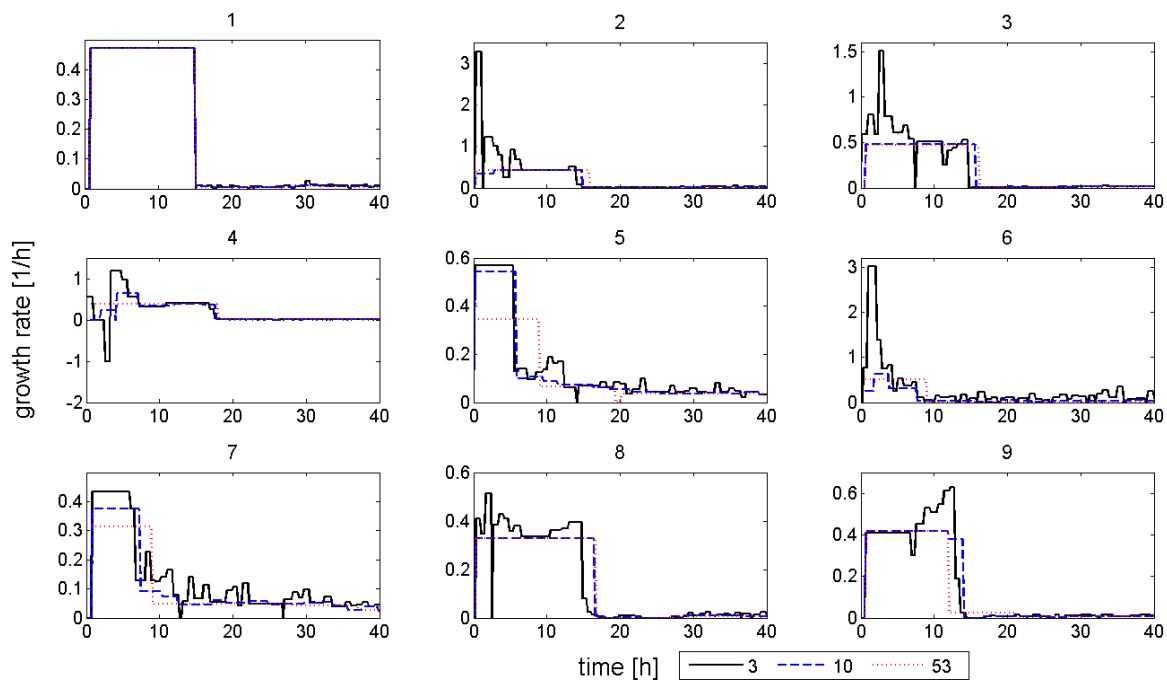


Figure 4: Effect of different parameters for the “minimal amount of data points for linear regression” to the growth rates. The parameter has been set to 3 (minimal possible value, black and solid line), 10 (known as optimal value, blue and dashed line) and 53 (maximal possible value, red and dotted line). The main effect can be seen for the parameter 3; the curve of estimated growth rates is in contrast to the other curves much less constant. Sometimes, the small parameter value leads to effects of partially negative (setup 4) or very high (setup 2 and 6) growth rates. The growth rate estimated with the parameter values 10 and 53 are comparable for that example.

Load BIOLECTOR file and start data processing



- Before starting data processing, the BIOLECTOR file including the experimental raw data has to be loaded (22).
- Now, data processing can be started (23). The results will be stored automatically in the project subfolder as *BioLectorFileName_processed_time-stamp.csv*. The content depends on the chosen processing steps. Also, depending on the chosen processing steps, MATLAB figures will appear.
- To continue with another run and adapted setup parameters, resetting (24) of this mask is often but not basically required. Confusing errors can often be solved resetting the mask.

Messages



- Messages at the end of the window should help to find and correct errors (Error: ...), display some unusual assumptions during processing (Warning: ...) and signalize the successful completion of data processing (Information: ...).

Example

This section describes an example which should run and can therefore be used as a test case. After starting the MATLAB file the following *WellSelection.m* steps should be made:

1. Choose 9 setups
2. Load the plate design from `.../data/20130312_examplefiles/PlateDesign_5replicates.mat` and save the plate design
3. Select F6, F7 and F8 as blank wells and continue with save.
4. Set the filling volume of all setups to 1000µL and save the volumes.
5. Select the internal calibration function.
6. Select growth rate calculation and adjust 50 data points for each regression.
7. Load the BioLECTOR file `.../data/20130312_examplefiles/ 5replicates_UsingSavedDesign.csv`
8. Run data processing!

This example data processing results in a CSV-file and three MATLAB figures.

CSV-file: `BioLectorFileName_processed_time-stamp.csv`

1	wells/replic1- (1 9 17 25 33)				2- (2 10 18 26 34)				3- (3 11 19 27 35)			
2	time [min]	back.scatter [a.u.]	bio.dry.weight [g/L]	growth.rate [1/h]	time [min]	back.scatter [a.u.]	bio.dry.weight [g/L]	growth.rate [1/h]	time [min]	back.scatter [a.u.]	bio.dry.v	
3	value	std(abs)	value	std(abs)	value	std(abs)	value	std(abs)	value	std(abs)	value	
4	1.13244	1.24533333	0.66518419	0.05788575	0.03092427	0	0	1.13796	0.90133333	1.07910611	0.0418959	
5	10.98876	0.49333333	0.36993243	0.02293119	0.01719669	0	0	10.995	0.45533333	0.8638692	0.02216486	
6	21.0078	0.21333333	0.41695323	0.00991619	0.01938111	0	0	21.015	-0.02666667	0.97334475	-0.00123952	
7	31.01628	0.60333333	0.26201145	0.02804422	0.01218189	0	0	31.02252	0.29733333	0.91023623	0.01382069	
8	41.01384	0.36466667	0.34687173	0.01695048	0.01612418	0	0	41.02068	0.18266667	0.84233604	0.00849074	
9	50.98968	0.528	0.29115288	0.02454257	0.01353555	0.46988009	0.0025613	50.9958	0.192	0.92003804	0.00892457	
10	61.00284	0.46333333	0.38294908	0.02153672	0.01780152	0.46988009	0.0025613	61.00944	0.15133333	0.92395982	0.0070343	
11	71.01864	0.518	0.28560462	0.02407775	0.01327757	0.46988009	0.0025613	71.02452	0.178	0.89697826	0.00827382	
12	80.99712	0.53266667	0.36156604	0.02475948	0.01680808	0.46988009	0.0025613	81.00276	0.29466667	0.8191581	0.01369674	
13	91.00076	0.636	0.32837479	0.02956264	0.01526625	0.46988009	0.0025613	91.00692	0.078	0.92569974	0.00362561	
14	101.01972	0.78133333	0.33274615	0.03631804	0.01547076	0.46988009	0.0025613	101.02572	0.14133333	0.89745195	0.00656947	
15	111.00612	0.748	0.30792856	0.03476864	0.01431715	0.46988009	0.0025613	111.01248	0.14	0.81816869	0.0065075	
16	121.0008	0.98533333	0.4070258	0.04580004	0.01892464	0.46988009	0.0025613	121.00692	0.11133333	0.81542627	0.00517501	
17	131.01852	0.97266667	0.27281862	0.04521162	0.01268877	0.46988009	0.0025613	131.02524	0.19066667	0.81359081	0.00886259	
18	140.99832	1.046	0.32959066	0.04862031	0.01532733	0.46988009	0.0025613	141.00588	0.346	0.84246245	0.01608282	
19	151.01832	1.31533333	0.33566352	0.0611395	0.01561362	0.46988009	0.0025613	151.02408	0.25733333	0.90875189	0.0119614	
20	161.00412	1.26	0.21965883	0.05856749	0.01022599	0.46988009	0.0025613	161.01	0.21	0.92341757	0.00976125	
21	171.0066	1.272	0.25756767	0.05912527	0.01198411	0.46988009	0.0025613	171.01212	0.394	0.798312951	0.01831396	
22	180.99972	1.60733333	0.31532523	0.07471228	0.01467488	0.46988009	0.0025613	181.00344	0.43533333	0.92396951	0.02023522	
23	191.0124	1.63866667	0.28968949	0.07616872	0.01348563	0.46988009	0.0025613	191.0184	0.35666667	0.98267492	0.01657863	
24	201.02364	1.93533333	0.2118254	0.08995842	0.00988466	0.46988009	0.0025613	201.02928	0.39133333	0.86243993	0.01819001	
25	210.9924	1.962	0.23815961	0.09119795	0.01110544	0.46988009	0.0025613	210.99924	0.528	0.88759225	0.02454257	
26	221.00076	2.21533333	0.32866396	0.10297342	0.0153096	0.46988009	0.0025613	221.00784	0.53933333	0.88796401	0.02506936	
27	231.0256	2.41866667	0.2046216	0.11242479	0.00957353	0.46988009	0.0025613	231.01536	0.59866667	0.89570076	0.0278273	
28	241.02516	2.75266667	0.27446311	0.12794982	0.01281782	0.46988009	0.0025613	241.03092	0.60266667	0.88449421	0.02801323	
29	251.00364	2.78866667	0.18833481	0.12962318	0.00884399	0.46988009	0.0025613	251.01084	0.73466667	0.89048863	0.03414887	
30	261.01212	2.892	0.19460216	0.13442633	0.00913897	0.46988009	0.0025613	261.01728	0.704	0.94245955	0.03272342	
31	271.0254	3.19666667	0.13019217	0.14858789	0.00622078	0.46988009	0.0025613	271.03152	0.70866667	0.91442332	0.03294034	
32	281.03364	3.51333333	0.27586228	0.16330723	0.01292009	0.46988009	0.0025613	281.0388	0.74733333	0.93098335	0.03473765	
33	291.01832	3.006	0.2029781	0.16761457	0.00958048	0.46988009	0.0025613	291.02112	0.788	0.84455314	0.03662792	
34	301.0236	3.80866667	0.17986106	0.17694199	0.00853459	0.46988009	0.0025613	301.03116	0.73066667	0.88791338	0.03396295	

Figure 5: The CSV-file contains all information gained during data processing. In case of the example including all available steps the table comprises time [min], mean backscatter signal [a.u.], bio dry weight [g/L] and (partially linear) growth rate [1/h] (e.g. column B, C, E, G, respectively) for each setup (e.g. setup 1 consisting of wells 1,9,17,25,33). Standard deviations are also included in this file (e.g. column D, F and H).

A setup consists of at least one well, the well numbers are stored, too. The numbers can be related to the position in the well as follows:

	1	2	3	4	5	6	7	8
A	1	2	3	4	5	6	7	8
B	9	10	11	12	13	14	15	16
C	17	18	19	20	21	22	23	24
D	25	26	27	28	29	30	31	32
E	33	34	35	36	37	38	39	40
F	41	42	43	44	45	46	47	48

MATLAB Figure: Visualization of the processed data

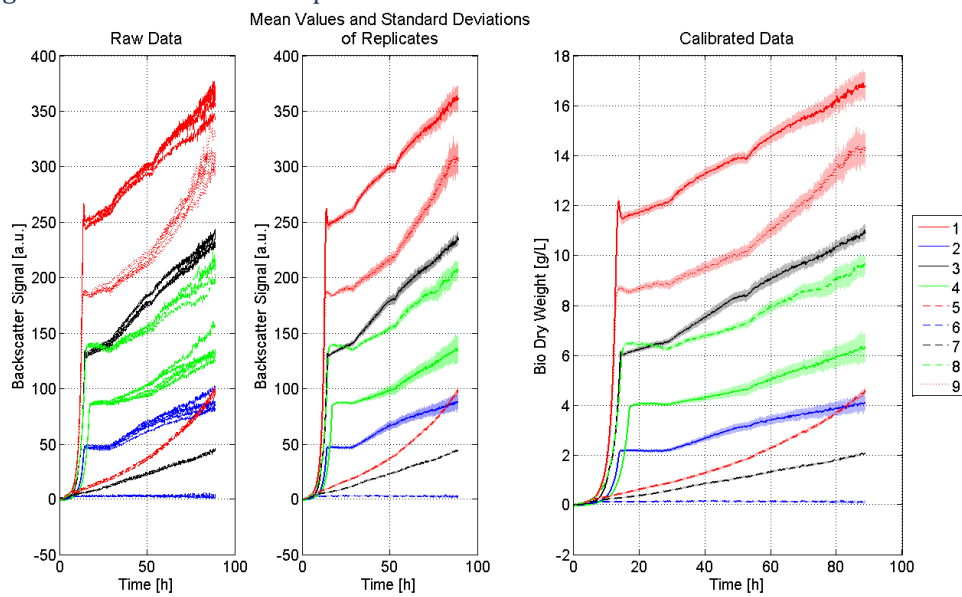


Figure 6: Data after each processing step; left: Raw data without any processing, backscatter curves belonging to one setup have the same line style; mid: Backscatter signal after mean value calculation, the standard deviations are shown as error tubes around the each curve; right: Bio dry weight after data calibration. The legend allocates the curves to the setups.

MATLAB figure: Estimating the growth rates with piecewise linear regression

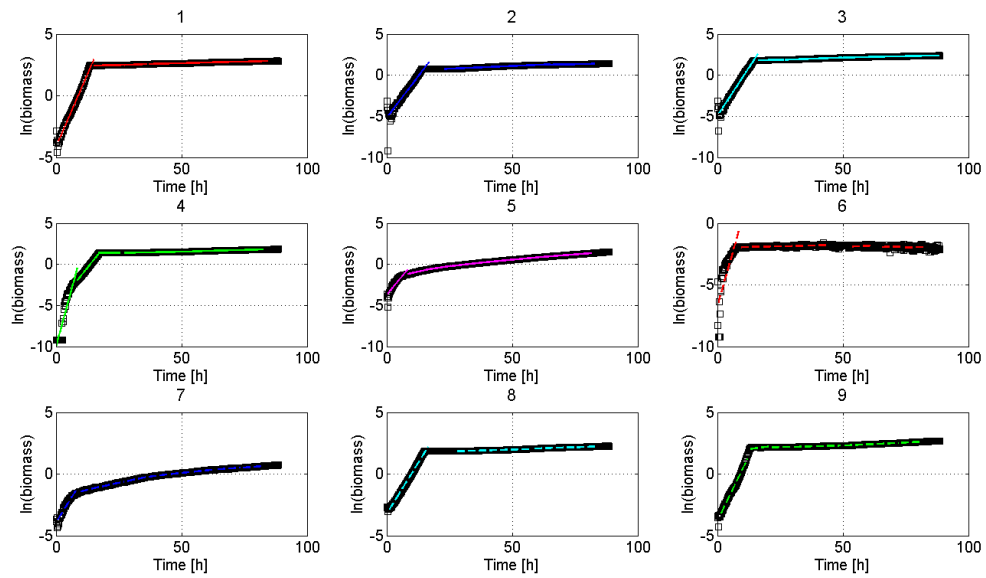


Figure 7: Linear regression of the logarithmic bio dry weight curves for each setup (number above the diagrams). Large deviations of the colored line from the logarithmic bio dry weight curves in black should not occur. If they occur, an adaption of the “minimal amount of data points for linear regression” (20) might be helpful.

MATLAB figure: Growth rates

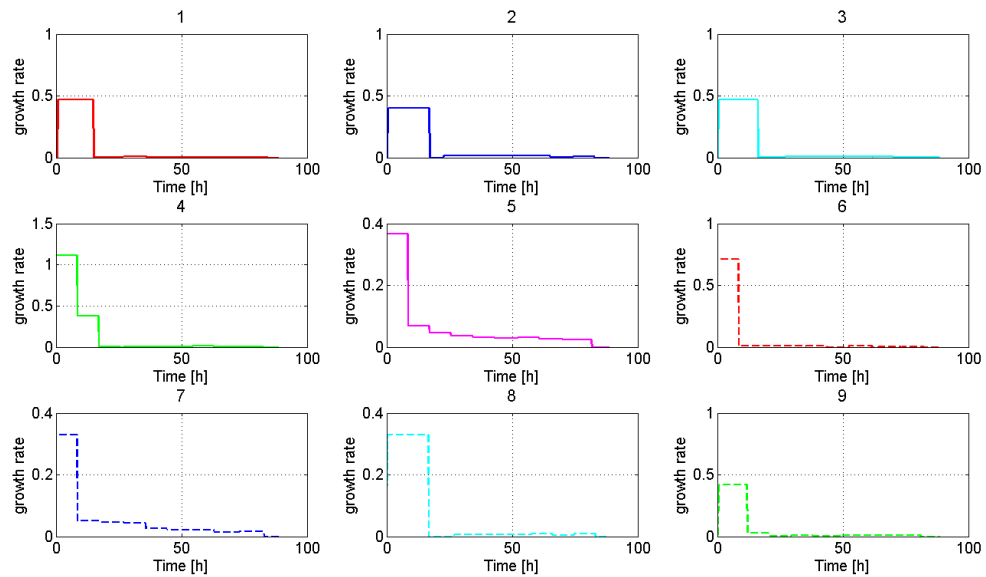


Figure 8: Estimated growth rates, which are stored in the CSV-file, too. The maximal growth rate can directly be taken from this figure.

Appendix

The Table 1 and Table 2 include the parameters of the non-linear and linear calibration function for different well filling volumes. Both functions can also be found in the file *CalibrationFunctionParameter.xlsx*.

Table 1: Parameters and standard deviations of the internal non-linear calibration function. The function to calculate the bio dry weight is: $bdw = (\sqrt{\text{backscatter} + a2^2/4/a1 - a3} - a2/2/\sqrt{a1})/\sqrt{a1}$.

Volume [μL]	a1		a2		a3	
	Value	Std	Value	Std	Value	Std
300	0.64	0.12	3.19	2.04	32.86	6.64
400	0.66	0.11	9.11	1.66	31.55	3.98
500	0.74	0.14	9.00	2.08	45.14	6.69
600	0.74	0.13	8.97	1.93	42.99	4.97
700	0.68	0.14	10.30	1.97	31.85	4.59
800	0.64	0.14	11.24	1.95	21.65	4.43
900	0.63	0.14	11.48	2.01	16.80	4.33
1000	0.63	0.14	11.73	1.92	12.96	3.87
1100	0.63	0.14	12.40	1.93	10.31	3.69
1200	0.60	0.14	12.72	1.88	9.24	3.47
1300	0.58	0.13	12.92	1.74	9.98	2.90
1400	0.55	0.13	13.12	1.75	10.02	3.26
1500	0.53	0.12	13.39	1.57	8.25	2.51

Table 2: Parameters and standard deviations of the internal linear calibration function. The function to calculate the bio dry weight is: $bdw = (\text{backscatter} - a1)/a2$. The resulting linear curve has the more intuitive parameters intercept and slope ($bdw = \text{slope} * \text{backscatter} + \text{intercept}$).

Volume [μL]	a1		a2		Intercept Value	Slope Value
	Value	Std	Value	Std		
300	-4.22	2.95	14.86	0.41	0.28	0.07
400	8.99	2.37	18.11	0.42	-0.50	0.06
500	-1.63	2.32	22.46	0.27	0.07	0.04
600	0.06	1.84	21.95	0.24	-0.00	0.05
700	-6.85	1.45	22.16	0.25	0.31	0.05
800	-8.13	1.94	21.71	0.25	0.37	0.05
900	-11.83	1.96	21.82	0.28	0.54	0.05
1000	-12.21	1.64	21.51	0.21	0.57	0.05
1100	-9.06	1.64	21.47	0.19	0.42	0.05
1200	-9.69	1.36	21.50	0.16	0.45	0.05
1300	-2.79	0.84	20.48	0.13	0.14	0.05
1400	-13.26	1.25	21.68	0.25	0.61	0.05
1500	10.43	0.39	18.80	0.11	-0.55	0.05

Anhang C

Metabolom-Daten

C.1 Rohdaten

Der Datensatz zur Analyse des modell-basierten Ansatzes und des experimentellen Ablaufes zur Metabolom-Datenanalyse enthält Rohdaten für 44 Metabolite bestehend aus Standard-Messungen, Peakflächen und dem mittels Coulter-Counter bestimmten Biovolumen.

Tabelle C.1: Metabolom-Datensatz: Standard-Messungen (1/5). Grau hinterlegte Messungen sind von der Regression ausgeschlossen.

c_{std} μM	$\eta_{12C, std}$ counts	$\eta_{13C, std}$ counts	$\eta_{12C, std}$ counts	$\eta_{13C, std}$ counts	$\eta_{12C, std}$ counts	$\eta_{13C, std}$ counts
	ACoA		ADP		AKG	
25	10207312.41	718178.166	9181451.846	2716261.716	18220971.89	1207635.218
10	4283797.161	591921.605	3997017.804	2622924.302	9317897.08	839758.581
5	2040734.01	500383.075	2022552.46	2735550.755	5797894.324	770136.735
2.5	995063.777	472070.712	1133187.431	2824120.79	4174861.974	735864.944
1	406327.954	449846.157	458495.898	3020151.75	3161870.745	707390.863
0.5	209524.743	430777.108	252786.433	3011701.654	2761814.41	679704.465
0.25	108044.942	427476.881	181456.764	2980269.919	2645739.153	677988.24
0.1	48143.038	439323.472	89057.254	2936303.653	2532068.795	671938.24
0.05	25585.574	421712.788	51916.72	3118574.179	2509567.406	668442.653
0.025	15799.518	470881.416	37055.724	2976567.33	2416724.967	636386.859
	ALA		AMP		ASP	
25	1295133.292	322855.773	12758196.46	2716261.716	1776293.51	1105711.544
10	772325.607	426551.517	4950057.546	2622924.302	671522.692	840479.848
5	421281.24	460989.443	2479597.38	2735550.755	403129.906	985712.074
2.5	214723.547	472253.307	1204300.164	2824120.79	226507.664	971831.647
1	102822.793	405518.412	462053.253	3020151.75	85879.693	666760.854
0.5	61575.722	487554.347	218192.208	3011701.654	58436.64	853819.462
0.25	23121.067	413748.409	129120.802	2980269.919	31077.145	743999.89
0.1	19436.363	385710.147	50601.691	2936303.653	17159.769	744470.541
0.05	6662.75	313053.895	28189.961	3118574.179	13194.069	593876.719
0.025	6893.277	309314.437	17060.423	2976567.33	10051.808	529976.347

Tabelle C.2: Metabolom-Datensatz: Standard-Messungen (2/5). Grau hinterlegte Messungen sind von der Regression ausgeschlossen.

c_{std} μM	$\eta_{12C, std}$ counts	$\eta_{13C, std}$ counts	$\eta_{12C, std}$ counts	$\eta_{13C, std}$ counts	$\eta_{12C, std}$ counts	$\eta_{13C, std}$ counts
	ATP		CAN		DHAP	
25	11964541.5	2716261.716	176172273.9	3172724.426	20356224.45	3713569.784
10	4422802.373	2622924.302	117768879.8	3702546.377	7041964.871	2936103.932
5	2124812.879	2735550.755	86891016.44	3722041.192	3558249.611	2895169.267
2.5	859500.886	2824120.79	69468151.62	3541443.506	1774583.825	2965010.538
1	243101.632	3020151.75	61610111.02	3587255.697	725278.756	3039879.521
0.5	142388.924	3011701.654	55367080.76	3406865.014	396473.96	3008063.53
0.25	86535.21	2980269.919	54159560.26	3389921.215	267073.491	3076649.266
0.1	52244.256	2936303.653	53840009.11	3344369.696	131500.856	3143532.55
0.05	41209.777	3118574.179	50786533.6	3181627.367	109458.604	3169208.554
0.025	34094.673	2976567.33	50099082.19	3233510.322	80648.254	2977071.07
	E4P		F6P		FBP	
25	248951.292	89156.593	10256521.24	774774.3	2424789.296	1990604.663
10	254166.806	173844.734	6415660.389	1110567.876	853744.746	1695246.743
5	128671.327	180758.76	3371288.617	1322331.478	475196.521	1678065.019
2.5	73741.559	178695.092	1667895.598	1361958.421	246957.26	1704161.492
1	30504.696	181305.872	609609.309	1450914.001	132460.83	1689369.457
0.5	19864.95	174143.986	321242.866	1542979.304	85724.354	1906198.993
0.25	13701.064	178173.738	177706.637	1472879.258	70411.65	1724249.927
0.1	9660.554	175712.429	99644.097	1631375.206	61350.18	1804564.547
0.05	9003.257	174825.823	74400.839	1642407.311	55765.84	1794841.103
0.025	6783.446	193789.748	56718.757	1461038.191	55396.405	1781559.167
	FUM		G6P		GAP	
25	1639681.64	105773.059	5885172.015	1981664.957	11237676.71	845367.825
10	677953.602	112612.992	6558311.896	4135639.853	2256063.716	419692.292
5	330698.015	109705.784	3723524.768	4349436.797	1049164.737	392741.571
2.5	153944.337	107939	1895266.591	4348201.422	507992.827	388212.667
1	62503.503	109924.65	823749.478	4454161.628	231513.979	436902.072
0.5	29777.072	101476.118	506732.383	4458856.405	116388.036	427109.548
0.25	17736.853	95671.828	324521.732	4245878.091	69201.634	428039.457
0.1	8794.47	100326.247	189937.083	4492814.475	29544.976	398303.391
0.05	4207.522	102918.252	136953.497	4458899.388	19061.659	390015.721
0.025	0	100061.028	121707.724	4665991.644	19475.375	472068.673
	GLN		GLU		GLY	
25	8874681.879	1814547.358	4311814.476	13772259.77	19660.804	3851.209
10	6566808.457	3273157.935	3086940.902	13358409.46	7937.436	4851.541
5	3775553.483	3921313.203	2684421.867	14058463.03	5822.355	5236.306
2.5	2225935.584	4391615.302	2483985.561	15307131.58	2941.768	7019.207
1	729676.259	3359861.983	1722279.731	11666261.5	1425.01	3653.609
0.5	458282.563	4645383.05	1826847.88	12950864.04	466.578	5904.087
0.25	188216.89	3473868.931	1587781.931	10713913.94	514.359	5639.335
0.1	84802.9	3316045.63	1472711.407	10557778.45	569.054	4524.942
0.05	32011.803	2498207.544	1065089.072	7805579.771	318.521	4305.08
0.025	19956.28	2246242.294	1022042.657	7142532.869	366.644	4567.898

Tabelle C.3: Metabolom-Datensatz: Standard-Messungen (3/5). Grau hinterlegte Messungen sind von der Regression ausgeschlossen.

c_{std} μM	$\eta_{12C, std}$ counts	$\eta_{13C, std}$ counts	$\eta_{12C, std}$ counts	$\eta_{13C, std}$ counts	$\eta_{12C, std}$ counts	$\eta_{13C, std}$ counts
	GOX		HIS		HSE	
25	665021.736	105773.059	16441591.4	97232	2915121.845	177639.79
10	250503.912	112612.992	7294109.576	87217.306	1946480.955	308239.154
5	124636.237	109705.784	4107880.892	87000.422	1213625.874	381578.915
2.5	59979.783	107939	2169063.514	80281.287	723287.346	387240.573
1	24941.75	109924.65	1357587.303	81048.128	213975.772	338277.24
0.5	14367.11	101476.118	1118960.844	87259.286	148682.387	554713.432
0.25	9824.215	95671.828	585081.376	58396.595	56763.365	368516.885
0.1	7009.947	100326.247	415584.216	50535.317	16989.548	259925.445
0.05	5379.992	102918.252	228688.728	27913.889	10317.43	215777.494
0.025	3598.62	100061.028	235182.339	31542.765	5301.593	195190.3
	LEU		LYS		MET	
25	23830730.17	1514589.252	2266521.076	349220.761	5485824.654	84918.587
10	15177734.52	2166796.827	1024529.254	329991.39	4146486.848	136482.325
5	8411805.053	2549775.171	525561.086	297730.493	2557860.988	196471.507
2.5	4344906.686	2593169.495	315059.687	272161.494	1521290.996	221626.12
1	1581567.801	2222351.052	220368.447	258450.423	480710.959	179953.027
0.5	944234.251	2626928.304	170286.909	277078.131	291466.622	220306.201
0.25	374773.36	2015030.569	135159.036	233327.308	128036.226	195958.543
0.1	156759.258	1912564.884	67624.737	179726.43	42418.292	167828.298
0.05	79420.817	1501482.982	43032.109	128366.199	16861.136	123318.289
0.025	55350.896	1419522.298	44204.675	119671.188	14305.009	116918.51
	NAD		NADH		NADP	
25	6630417.139	943526.877	3121202.903	943526.877	5230738.926	288334.661
10	2523597.494	859150.067	1192863.035	859150.067	2075264.608	255174.945
5	1333854.399	927509.736	632837.57	927509.736	1072155.086	271422.156
2.5	678523.87	974255.931	284889.975	974255.931	515294.348	266585.277
1	282843.504	980181.223	80329.433	980181.223	204823.896	354305.631
0.5	151192.22	1019141.293	29455.047	1019141.293	102661.363	276871.737
0.25	89554.55	1005907.099	-1776.729	1005907.099	60347.224	278876.252
0.1	31142.486	1048301.389	5911.459	1048301.389	22161.955	288547.045
0.05	16612.783	1026019.433	3479.576	1026019.433	11873.029	275486.946
0.025	8288.403	1010761.939	2319.597	1010761.939	6952.602	300547.869
	NADPH		OAS		PEP	
25	4305730.706	288334.661	627016.304	74343.268	15689064.59	834274.823
10	1857869.615	255174.945	186796.819	76045.844	5780406.157	761184.666
5	878846.688	271422.156	96811.015	89727.088	2806551.357	724484.616
2.5	423343.11	266585.277	58222.269	91315.921	1353775.775	717503.971
1	159029.474	354305.631	12508.016	72027.04	538459.962	690472.342
0.5	75736.319	276871.737	8102.606	90484.298	265714.707	657213.695
0.25	45680.933	278876.252	2267.906	78147.225	159926.742	667681.215
0.1	19683.161	288547.045	1813.076	76547.881	73382.203	649729.738
0.05	11037.576	275486.946	1365.556	63135.421	50963.656	643159.524
0.025	7125.132	300547.869	1645.735	52721.519	36434.261	661485.757

Tabelle C.4: Metabolom-Datensatz: Standard-Messungen (4/5). Grau hinterlegte Messungen sind von der Regression ausgeschlossen.

c_{std} μM	$\eta_{12C, std}$ counts	$\eta_{13C, std}$ counts	$\eta_{12C, std}$ counts	$\eta_{13C, std}$ counts	$\eta_{12C, std}$ counts	$\eta_{13C, std}$ counts
	PHE		PRO		PYR	
25	14725197.37	208956.58	6710613.421	210924.919	164142.711	57893.68
10	9532427.515	339178.42	5340885.17	393246.348	61243.309	52198.719
5	5101423.04	380637.948	3150951.537	482378.809	31291.402	51957.298
2.5	2493144.901	353608.844	1785323.034	536419.005	15551.168	50168.291
1	898140.552	253899.668	655361.259	482617.771	6256.731	47790.553
0.5	444469.147	266185.201	346227.905	498874.409	3360.834	41406.543
0.25	189378.494	217677.223	170088.899	462008.484	1018.47	41273.211
0.1	86260.82	262541.852	75557.391	445210.285	460.939	42509.199
0.05	29748.821	150878.23	40774.393	348840.681	1291.414	45964.098
0.025	20623.529	166884.963	20021.991	311047.585	306.024	39998.305
	R5P		S7P		SER	
25	11579768.51	140479.693	10091011.91	67379.827	2759545.917	74343.268
10	5694102.711	171262.587	6959371.8	79462.315	1160380.825	76045.844
5	2921572.737	194391.232	3568498.877	84816.847	641074.32	89727.088
2.5	1486139.778	179898.815	1817411.255	83287.002	337095.851	91315.921
1	567790.212	200111.101	728891.729	88640.52	145828.749	72027.04
0.5	276457.542	198657.393	388008.029	90151.42	75229.981	90484.298
0.25	153255.037	201804.556	219103.696	77371.519	35251.952	78147.225
0.1	54784.081	194966.969	77328.679	86189.142	22093.899	76547.881
0.05	31018.178	196855.541	35933.807	83006.81	18497.214	63135.421
0.025	16388.431	195593.578	20304.37	84418.29	8033.887	52721.519
	SUC		THR		TRY	
25	6157656.066	1014603.291	3929932.125	361695.161	12779285	23365.774
10	2969864.708	1051027.675	2175854.339	451762.928	8061067.32	36824.086
5	1668031.189	1019069.809	1311218.035	587350.378	4280589.437	43373.673
2.5	1000113.066	1089400.904	742641.99	671298.875	2307605.131	44065.979
1	601671.687	1034436.399	264576.422	539892.378	683641.378	30961.645
0.5	461860.738	1018033.628	145632.992	574366.693	369959.596	38606.216
0.25	402576.026	1001006.288	65456.217	543224.178	156061.98	29054.457
0.1	355159.307	1016850.829	29445.205	527273.606	53869.125	14619.33
0.05	356447.122	986837.657	17717.77	397214.714	18858.598	19677.916
0.025	333558.388	892977.608	13188.215	380076.47	11268.491	17552.232
	TYR		VAL		X5P	
25	5408109.118	59204.839	255859.367	32576.677	14460010.71	343212.86
10	2995750.427	77313.111	156652.179	47594.229	5628253.72	285338.405
5	1873359.095	94486.538	84765.567	50769.632	2902955.563	255845.403
2.5	1061545.873	101377.17	38195.23	54287.764	1440966.817	286205.024
1	374725.076	81922.328	16222.967	45966.672	572215.536	291149.591
0.5	226248.109	101855.526	7918.791	46998.123	272830.421	297014.034
0.25	91237.859	77635.93	3479.64	38423.305	158059.55	285259.862
0.1	37294.093	86781.341	620.931	37893.523	54407.063	288033.989
0.05	12673.981	58289.877	1047.859	26721.558	24729.301	276190.568
0.025	10741.29	56048.349	556.334	28117.305	17456.198	276039.553

Tabelle C.5: Metabolom-Datensatz: Standard-Messungen (5/5). Grau hinterlegte Messungen sind von der Regression ausgeschlossen.

c_{std} μM	$\eta_{12C, std}$ counts	$\eta_{13C, std}$ counts	$\eta_{12C, std}$ counts	$\eta_{13C, std}$ counts
	6PG		23PG	
25	7222444.307	134205.137	1598294.068	169326.12
10	2751011.574	112442.4	700115.833	155336.165
5	1368179.458	111397.358	355081.169	158448.708
2.5	666898.699	113840.196	186824.909	164581.879
1	267887.016	115936.583	75838.18	171710.511
0.5	138747.267	115139.664	42403.279	167247.334
0.25	75483.546	106892.007	31342.994	166026.741
0.1	33018.403	106963.606	14706.408	171820.298
0.05	19053.187	103395.565	12598.291	171120.105
0.025	13397.275	106160.172	9896.71	181325.605

Tabelle C.6: Metabolom-Datensatz: Messung Biovolumen
Biovolumen

ν_{brs} μL
5.13
5.28

Tabelle C.7: Metabolom-Datensatz: Peakflächen von den Proben Extrakt, Kulturüberstand und Quenchingüberstand (1/5).

Extrakt		Kulturüberstand		Quenchingüberstand	
$\eta_{12C, exc}$	$\eta_{13C, exc}$	$\eta_{12C, cub}$	$\eta_{13C, cub}$	$\eta_{12C, qsn}$	$\eta_{13C, qsn}$
ACoA					
41228.031	482245.109	2042.248	399815.447	7408.901	388174.844
42095.489	500542.885	1794.085	391721.867	6658.099	399826.897
39475.562	484582.485	1538.081	385647.639	7102.657	417797.827
ADP					
2667749.505	2484893.027	15151.133	2682240.75	348951.016	2758247.559
2620353.14	2480269.251	16351.095	2766455.47	359336.52	2841284.788
2891612.355	2733926.195	12830.174	2560835.337	347177.139	2771013.416
AKG					
4961764.244	563234.395	5259214.735	680117.037	5214099.06	612445.248
5163617.957	586205.644	5302284.676	685406.285	5695402.97	665861.509
5233284.312	593067.483	5292514.656	679989.748	5682821.046	664607.782
ALA					
271291.147	201221.172	224031.53	173920.425	324624.803	174290.111
225508.779	177021.357	162909.873	127785.501	250247.972	127595.097
311301.077	234858.742	172489.651	136896.988	362100.586	182703.869
AMP					
925096.959	2552555.438	5520.539	2682240.75	30348.29	2758247.559
948077.858	2480269.251	5769.117	2766455.47	27237.639	2841284.788
904124.641	2733926.195	6005.138	2760835.337	28838.741	2771013.416

Tabelle C.8: Metabolom-Datensatz: Peakflächen von den Proben Extrakt, Kulturüberstand und Quenchingüberstand (2/5).

Extrakt		Kulturüberstand		Quenchingüberstand	
$\eta_{12C,exc}$	$\eta_{13C,exc}$	$\eta_{12C,cub}$	$\eta_{13C,cub}$	$\eta_{12C,qsn}$	$\eta_{13C,qsn}$
ASP					
386297.564	405867.406	42250.155	324752.174	204974.49	436085.886
392672.268	405211.095	34412.237	290850.375	192458.236	394596.992
405467.96	415502.561	39072.492	300437.45	192002.158	408042.811
ATP					
12255817.43	2484893.027	39795.656	2682240.75	856161.491	2758247.559
12181271.85	2480269.251	32184.133	2766455.47	801512.601	2841284.788
12809229.18	2630529.711	29358.538	2560835.337	661744.828	2450121.145
CAN					
54945051.1	3384457.844	60918131.74	4390448.24	60923490.85	3978644.971
55540938.87	3425435.504	59846869.81	4364908.451	61718577	4027934.616
56778874.38	3511253.51	60655103.95	4373025.181	59885246.38	3862577.441
DHAP					
3415190.388	3085490.738	239229.308	3001468.183	687006.695	3079644.148
3233456.276	2984941.875	245538.611	3142285.546	688874.262	3037509.319
3230815.95	3026351.376	236106.76	3026748.316	692595.554	3119729.103
E4P					
347636.958	164577.721	5062.209	133193.324	37156.873	110588.895
318437.183	154494.797	4947.612	122670.731	43483.343	134681.049
322947.499	155558.609	4143.509	108849.383	35167.427	107533.402
F6P					
3469317.135	1411937.445	76106.287	1327284.701	514924.79	1250958.272
3340666.229	1363620.474	84337.585	1455092.264	505428.113	1268813.355
3264179.207	1363306.386	71693.271	1254768.632	561954.299	1374158.892
FBP					
1842109.711	1390298.433	20452.508	917079.334	111549.148	976144.114
1718362.155	1271074.367	28798.113	1007521.423	112097.977	1011804.177
1630530.863	1223718.483	25463.7	960494.642	120114.47	1045265.432
FUM					
23817.341	98447.7	0	99757.353	7769.794	95809.539
24190.676	103232.201	0	97164.366	7777.078	98365.011
24107.189	103475.151	0	97058.973	7733.929	95630.595
G6P					
7541913.886	4073256.22	120703.936	2508843.342	739454.705	2128714.359
7539576.008	4006066.952	125158.666	2603608.606	893723.102	2561565.811
7564784.109	4119309.678	115714.349	2360420.58	849413.478	2418173.139
GAP					
257594.088	323137.26	10863.503	218476.847	21641.158	229275.491
286979.787	358340.709	9233.518	207988.287	19000.884	203939.671
249541.316	311268.382	9720.932	209531.654	24914.566	253833.825
GLN					
7677219.9	857126.96	245946.562	1537415.809	3998476.995	1105538.583
6395658.185	717533.623	205995.808	1274402.223	3872790.448	1052792.06
8373338.123	956369.309	219049.548	1358744.974	4476520.178	1197633.203
GLU					
27731045.36	2832408.672	3142602.423	4517663.67	12393735.36	3807196.283
25566927.39	2595111.637	2568984.147	3693492.07	12801505.54	3961587.249
34525967.44	3522747.726	2636680.251	3690019.158	14402216.05	4483330.033

Tabelle C.9: Metabolom-Datensatz: Peakflächen von den Proben Extrakt, Kulturüberstand und Quenchingüberstand (3/5).

Extrakt		Kulturüberstand		Quenchingüberstand	
$\eta_{12C,exc}$	$\eta_{13C,exc}$	$\eta_{12C,cub}$	$\eta_{13C,cub}$	$\eta_{12C,qsn}$	$\eta_{13C,qsn}$
GLY					
4107.392	2986.971	13712.314	3121.846	10916.399	2567.116
3976.174	2903.181	9395.939	2044.36	11554.651	2573.464
5196.969	3808.857	7865.957	1695.504	13822.482	2861.231
GOX					
7136.321	98447.7	7163.802	99757.353	6904.65	95809.539
7433.479	103232.201	6872.029	97164.366	6949.913	98365.011
7364.378	103475.151	6732.903	97058.973	6784.107	95630.595
HIS					
262619.476	25319.451	38007.291	5025.363	25872.609	3038.397
249954.174	24179.666	2615.886	485.867	19784.662	2314.802
357781.15	34395.414	34871.531	12998.416	42812.956	5015.703
HSE					
368012.843	118375.391	219442.055	121275.064	223980.362	78336.723
310344.689	101652.132	167010.969	90592.66	276167.445	96438.812
399025.52	129541.514	193606.285	104168.067	294735.65	104160.536
LEU					
437353.532	960520.079	32284.47	542358.213	81727.055	541733.656
384554.305	830547.857	28809.62	425359.197	75858.901	470756.205
474596.832	1104915.157	27619.077	433313.533	95207.819	606375.012
LYS					
4012476.893	136266.7	1935954.314	11336.913	712136.776	5991.332
3352137.693	119030.696	1147249.979	6830.45	506296.273	4272.025
6050254.231	199885.391	3161653.976	20452.443	2349311.361	19702.236
MET					
77309.463	104416.902	0	87799.554	9553.578	91411.619
62992.502	87905.638	0	57781.426	9157.077	77120.14
88406.877	122178.019	0	67212.542	10693.282	91639.65
NAD					
1754992.315	915025.669	1228.701	994902.649	112404.967	985860.224
1754571.002	922262.889	1121.291	1025941.593	117248.022	1046346.798
1735443.267	921107.303	1265.9	1001179.2	112019.513	967769.028
NADH					
14273.015	915025.669	0	994902.649	0	985860.224
10977.476	887595.776	0	1025941.593	0	1046346.798
13443.217	921107.303	0	1001179.2	0	967769.028
NADP					
709001.461	192164.623	461.19	155406.735	22412.033	160021.021
674220.607	189873.651	395.435	153131.31	20299.531	136449.336
722276.808	204163.72	539.603	123136.913	22501.939	156194.555
NADPH					
32401.441	192164.623	266.935	155406.735	535.671	160021.021
31486.843	189873.651	71.613	153131.31	173.543	136449.336
33927.032	204163.72	454.205	123136.913	686.967	156194.555
OAS					
8377.585	23664.737	0	31528.054	0	37174.138
8758.771	23030.092	0	27619.422	0	27913.024
9116.56	29856.471	0	30060.402	0	32302.244

Tabelle C.10: Metabolom-Datensatz: Peakflächen von den Proben Extrakt, Kulturüberstand und Quenchingüberstand (4/5).

Extrakt		Kulturüberstand		Quenchingüberstand	
$\eta_{12C,exc}$	$\eta_{13C,exc}$	$\eta_{12C,cub}$	$\eta_{13C,cub}$	$\eta_{12C,qsn}$	$\eta_{13C,qsn}$
PEP					
317792.563	692433.141	2499.112	146021.224	29347.689	162835.894
330128.379	714511.424	207.425	151385.967	27640.355	181766.752
333040.074	714609.64	1395.777	148559.81	25598.55	170808.33
PHE					
187306.932	137881.707	10610.907	112043.626	34846.252	106424.629
164241.779	124420.263	9137.492	103395.37	39122.808	108724.609
199764.061	148796.697	8645.067	104833.339	46267.882	124398.903
PRO					
3167142.883	212572.321	415040.281	221444.922	2475297.512	165661.045
2689419.6	187343.195	287552.911	149720.821	2848930.673	170956.835
3380247.758	233158.521	372537.565	192882.591	3149549.073	188425.202
PYR					
42257.325	38056.908	0	33780.224	0	30294.729
43186.28	39266.067	0	20602.065	0	33792.311
43029.879	38951.601	0	30130.035	0	34091.877
R5P					
836697.546	167373.24	18633.537	155195.774	185693.091	158502.609
910992.812	176896.462	14548.99	175507.39	188280.542	156344.513
897987.12	179357.883	14833.044	147197.071	197896.093	166141.662
S7P					
6004731.079	77263.223	17827.988	70359.495	633349.6	72918.914
5990920.195	76724.498	16646.816	73073.568	634472.258	67737.843
5770665.688	73914.045	16868.005	66674.98	648513.021	73096.754
SER					
47671.568	23664.737	0	31528.054	39268.622	35297.316
46356.36	23030.092	0	27619.422	31654.486	27913.024
58938.705	29856.471	0	30060.402	36028.1	32302.244
SUC					
974273.864	908962.203	357106.172	857001.717	358114.484	859058.782
960730.937	884768.759	363287.584	869753.213	381231.119	905860.924
968092.478	904458.987	366367.961	890025.133	375488.986	893584.526
THR					
129509.176	121767.594	17590.674	252089.21	91148.087	202512.131
99239.344	91746.475	11724.558	183777.31	94672.859	207465.155
139758.742	130532.994	12790.103	185597.16	100422.507	221501.973
TRY					
18101.107	11069.689	1315.495	7233.825	2265.575	9164.955
14805.828	8692.767	1168.749	6227.087	2236.198	7664.522
17065.27	9487.907	819.914	4956.457	3309.288	10653.318
TYR					
45515.671	25761.32	3041.96	26855.205	5230.606	13310.883
31488.649	18348.623	2622.139	21584.035	5396.749	16413.116
38864.564	21705.672	2963.85	21912.101	8376.288	21084.439
VAL					
27591.201	26801.379	1391.588	14344.906	5241.044	10547.497
23714.701	22769.825	1170.887	11471.994	6542.871	11585.407
27830.859	26893.737	1247.663	14153.121	7817.54	14316.663

Tabelle C.11: Metabolom-Datensatz: Peakflächen von den Proben Extrakt, Kulturüberstand und Quenchingüberstand (5/5).

Extrakt		Kulturüberstand		Quenchingüberstand	
$\eta_{12C,exc}$	$\eta_{13C,exc}$	$\eta_{12C,cub}$	$\eta_{13C,cub}$	$\eta_{12C,qsn}$	$\eta_{13C,qsn}$
X5P					
1582012.369	258679.415	25073.685	260521.458	421746.156	265629.798
1705627.733	279858.595	27053.429	279385.519	436993.231	274100.062
1672269.523	276480.948	20825.449	266413.725	454769.899	285499.403
6PG					
157067.238	38114.175	0	6093.467	5514.714	6166.413
203305.725	49466.614	0	5658.586	5122.111	5757.726
193084.015	48504.096	0	6940.42	5784.337	6325.379
23PG					
149485.727	108149.988	7180.495	49508.96	43279.772	53713.643
147469.812	109102.244	7361.054	46389.394	44435.594	54971.505
148883.693	106053.12	5444.51	33737.617	45735.375	56069.791

C.2 Datenauswertung mit unterschiedlichen Modellvarianten

C.3 Sensitivitätsanalyse

Die Ergebnisse der Sensitivitätsanalysen zeigen die Einflüsse der Qualität der Eingangsdaten auf die Qualität der intrazellulären Konzentration.

Tabelle C.12: Modell-basierte Analyse der Veränderung der intrazellulären Metabolitkonzentration und ihrer Standardabweichung bei kumulativer Korrektur der systematischen Fehler. (1) Referenzmodell; (2) Matrixeffekt; (3) Messung des Biovolumens; (4) Metabolitverdünnung durch unvollständige Zellseparation; (5) zusätzliche Metabolite durch unvollständige Zellseparation; (6) Leakage.

1 Referenzmodell			2 Matrixeffekte			3 Biovolumen-Messung		
μ	σ_{abs}	σ_{rel}	μ	σ_{abs}	σ_{rel}	μ	σ_{abs}	σ_{rel}
85.4074	4.1272	4.83%	148.1336	8.1528	5.50%	381.1634	17.2991	4.54%
78.0104	2.9079	3.73%	158.8417	6.324	3.98%	408.5946	10.1115	2.47%
106.7045	4.0305	3.78%	233.968	9.4787	4.05%	601.9316	15.4585	2.57%
102.1379	3.8424	3.76%	100.2476	4.0098	4.00%	257.9257	6.4496	2.50%
25.13535	0.93677	3.73%	32.74225	1.29809	3.96%	84.24006	2.05483	2.44%
97.132	3.8108	3.92%	149.236	6.3007	4.22%	383.8349	11.0018	2.87%
2.93897	0.16402	5.58%	2.88028	0.19995	6.94%	7.41063	0.45847	6.19%
42.998	1.5935	3.71%	106.8163	4.2914	4.02%	274.7834	6.8687	2.50%
6.55077	0.2416	3.69%	14.30997	0.56939	3.98%	36.82148	0.90883	2.47%
145.8647	5.3933	3.70%	242.7879	9.7853	4.03%	624.5719	16.0311	2.57%
34.935	1.2843	3.68%	35.9901	1.4068	3.91%	92.5839	2.1792	2.35%
174.8529	6.4873	3.71%	193.4687	7.5978	3.93%	497.7324	11.8831	2.39%
52.1531	2.2027	4.22%	65.2831	3.0392	4.66%	167.9343	5.8149	3.46%
11.42668	0.42556	3.72%	52.3006	2.08451	3.99%	134.55609	3.3216	2.47%
51.0286	1.906	3.74%	92.4663	3.7609	4.07%	237.8647	6.1987	2.61%
64.8404	2.4242	3.74%	69.1839	2.7447	3.97%	178.006	4.3626	2.45%
43.5525	1.6493	3.79%	66.8609	2.6881	4.02%	172.0027	4.3263	2.52%
4 Metabolitverdünnung			5 zusätzliche Metabolite			6 Leakage		
μ	σ_{abs}	σ_{rel}	μ	σ_{abs}	σ_{rel}	μ	σ_{abs}	σ_{rel}
408.6109	18.6585	4.57%	382.6357	17.9355	4.69%	1473.901	224.2989	15.22%
438.0142	11.0727	2.53%	394.5703	10.4589	2.65%	3714.2677	196.4251	5.29%
645.2866	16.9081	2.62%	624.3007	16.3898	2.63%	3894.3944	105.7336	2.72%
276.5012	7.0716	2.56%	272.9654	6.9772	2.56%	793.9277	20.3914	2.57%
90.30455	2.25235	2.49%	89.75603	2.23884	2.49%	167.79665	4.27937	2.55%
411.4789	11.9878	2.91%	316.5229	11.9039	3.76%	1283.3697	482.1593	37.57%
7.94402	0.49316	6.21%	7.42346	0.47855	6.45%	15.33628	6.88457	44.89%
294.5636	7.5115	2.55%	276.267	7.15	2.59%	1718.6617	82.8142	4.82%
39.47309	0.99517	2.52%	38.63746	0.9756	2.53%	157.00649	4.51255	2.87%
669.5456	17.5466	2.62%	619.8611	16.5634	2.67%	9509.24	247.6383	2.60%
99.2503	2.3938	2.41%	97.6854	2.3541	2.41%	387.9343	8.1915	2.11%
533.5805	13.031	2.44%	529.4609	12.9214	2.44%	1337.7111	26.8083	2.00%
180.0235	6.3072	3.50%	178.7729	6.2805	3.51%	204.779	34.4158	16.81%
144.24852	3.64117	2.52%	140.26649	3.54412	2.53%	860.60557	20.58712	2.39%
254.9875	6.7785	2.66%	246.1854	6.5562	2.66%	1737.3305	45.7581	2.63%
190.8283	4.7863	2.51%	187.4871	4.6983	2.51%	825.9173	18.32	2.22%
184.3893	4.7312	2.57%	177.2961	4.5717	2.58%	1412.0635	36.0796	2.56%

Tabelle C.13: Ergebnisse der Sensitivitätsanalyse (Volumina): Die Verringerung jeder einzelnen Varianz auf 50% des experimentell bestimmten Wertes führt zu folgender Verminderung der Varianz der Metabolitkonzentration in %.

	AKG	ALA	ASP	DHAP	GAP	GLY
V_{brs}	-0.39309	0.44585	1.0594	0.64477	0.71146	0.33689
V_{exc}	-0.58158	-0.66664	0.55534	-0.098808	-0.33946	-0.0090281
V_{res}	-0.46729	-0.027388	0.37124	0.14149	0.15958	-0.0097398
V_{que}	-0.63626	0.21996	1.0424	0.63203	-0.038445	0.70473
V_{brs}	-0.21685	0.6369	4.3741	4.5841	4.1235	-0.42252
	GOX	HSE	LEU	PRO	R5P	S7P
V_{brs}	0.44451	0.31215	0.99272	0.27316	1.1672	0.82465
V_{exc}	-0.064678	-0.1712	0.6171	-0.0020473	-0.10103	0.60419
V_{res}	0.43849	-0.2923	0.57014	0.15316	-0.12908	-0.066092
V_{que}	0.19411	0.16795	0.19793	0.2526	0.55837	0.3935
V_{brs}	0.15853	1.1437	3.8729	4.3221	6.8023	7.8599
	SUC	THR	VAL	X5P	23PG	
V_{brs}	0.23001	0.61529	0.73732	1.8718	1.1195	
V_{exc}	0.12	0.53798	-0.069526	0.26583	0.53302	
V_{res}	0.15226	0.16984	0.32501	0.48991	-0.4128	
V_{que}	0.1587	0.31312	0.34417	0.40792	0.35753	
V_{brs}	0.16831	5.1986	4.8214	7.0589	4.8812	

Tabelle C.14: Ergebnisse der Sensitivitätsanalyse (Peakflächenquotienten): Die Verringerung jeder einzelnen Varianz auf 50% des experimentell bestimmten Wertes führt zu folgender Verminderung der Varianz der Metabolitkonzentration in %.

	AKG	ALA	ASP	DHAP	GAP	GLY
$\nu_{12C, std}$	1.9681	2.1643	10.487	9.5015	7.6444	-0.062125
$\nu_{13C, std}$	-0.054558	0.25364	1.3533	0.8013	0.43234	-0.084186
$\nu_{12C, exc}$	-0.57563	0.17305	0.23169	1.4749	2.5374	-0.12859
$\nu_{13C, exc}$	-0.10594	-0.062407	0.29068	1.2631	2.6444	0.00054068
$\nu_{12C, cub}$	7.2369	4.5888	0.6945	1.0115	1.4462	8.9663
$\nu_{13C, cub}$	7.714	4.541	0.59115	1.5755	1.0847	9.2708
$\nu_{12C, qsn}$	10.115	12.424	10.67	8.5934	7.1774	10.02
$\nu_{13C, qsn}$	10.412	12.876	11.18	9.4497	7.6888	10.146
	GOX	HSE	LEU	PRO	R5P	S7P
$\nu_{12C, std}$	0.50018	2.7613	7.0766	11.614	9.319	11.892
$\nu_{13C, std}$	0.1183	-0.3753	0.90724	0.78172	1.4416	1.5279
$\nu_{12C, exc}$	0.55607	0.01932	0.77128	-0.37761	0.61512	2.2857
$\nu_{13C, exc}$	0.28555	-0.58394	0.60735	-0.30807	1.0402	2.5653
$\nu_{12C, cum}$	9.4538	4.3692	2.2105	-0.030776	0.15984	0.26449
$\nu_{13C, cum}$	9.4028	4.3735	1.887	-0.19052	0.41608	0.096628
$\nu_{12C, qsn}$	10.417	12.25	11.89	9.5529	9.0578	6.1219
$\nu_{13C, qsn}$	10.573	12.524	12.262	10.519	9.5037	6.1922
	SUC	THR	VAL	X5P	23PG	
$\nu_{12C, std}$	0.60579	10.282	12.973	13.13	10.029	
$\nu_{13C, std}$	0.71116	1.168	1.4322	1.5268	0.84216	
$\nu_{12C, exc}$	0.53093	0.3291	0.21984	1.1598	0.17462	
$\nu_{13C, exc}$	0.67845	0.6483	0.32196	1.2577	0.67437	
$\nu_{12C, cum}$	8.8303	0.37075	0.12371	0.50009	0.20519	
$\nu_{13C, cum}$	9.3915	0.12446	0.055984	0.5285	0.35117	
$\nu_{12C, qsn}$	10.171	10.562	9.4759	8.5086	11.181	
$\nu_{13C, qsn}$	10.006	10.431	9.4768	8.9639	11.402	

Anhang D

Sensormodell

D.1 Experimentelle Daten

D.1.1 Datensatz 1: komplett

Der erste Datensatz besteht aus Messungen der Methionin-Konzentration, der Fluoreszenz und der Biomasse. Die Rohdaten wurden vorverarbeitet, um einheitliche Zeitpunkte, valide Messfehler und eine maximale Wachstumsrate (unter Einsatz des in Kapitel 5 entwickelten Werkzeugs) zu erhalten.

D.1.2 Datensatz 2: Kalibrierdaten

Dieser Datensatz stellt den Zusammenhang zwischen intrazellulären Konzentrationen und Fluoreszenzmessungen für verschiedene Dipeptid-Anteile dar.

D.2 Simulation und Analysen

D.2.1 Parameteranpassung

Einen Überblick über die Optimierungen mit dem finalen Modell 7.2 und dem Datensatz D.1 gibt Tabelle D.3. Die angepassten Parameter der Optimierung 11 sind in Tabelle D.4 zusammengefasst und die damit erzeugten Simulationen sind in Abbildung D.1 dargestellt.

Tabelle D.1: Übersicht über integrierte experimentelle Daten: Datensatz 1. Aus den Werten der Biomasse wurde eine maximale Wachstumsrate von $0.474 \pm 0.022 \text{ h}^{-1}$ ermittelt.

Zeit [min]	Konzentration [mM]		Fluoreszenz		Biomasse [g L^{-1}]	
	Wert	δ	Wert	δ	Wert	δ
0	3.829150434	0.627053386	1.09475	0.020182658	1.113693548	0.321079278
5	6.662349509	3.328729414	1.138388384	0.047166261	1.182089643	0.302314327
7	15.36285146	2.534459648	1.160001174	0.059778687	1.231941023	0.289950039
10	20.76288378	7.273700501	1.192420358	0.077840068	1.306718095	0.273172476
12	19.4510793	2.344291926	1.200891272	0.080218262	1.355650033	0.263227444
15	13.10112069	6.997055412	1.204490552	0.077374962	1.291117583	0.281613857
20	13.33062273	6.523779006	1.210489352	0.072673704	1.183563501	0.316713425
30	8.453269763	0.269412946	1.261609372	0.086224211	1.106222573	0.311726331
40	7.360168379	0.318257929	1.273900752	0.065489355	1.103500719	0.317937311
50	9.29746885	1.540774786	1.297926748	0.073149415	1.128974824	0.301894839
60	7.577301102	0.775199989	1.286848914	0.040458992	1.110890533	0.30325144
90	8.057574157	0.654513622	5.128107171	0.113684582	1.208607351	0.282441071
120	8.838436959	0.74540649	14.42294998	0.086066946	1.454617286	0.241794713
150	12.53027972	5.588975879	24.38083415	0.072087421	1.695323847	0.216529284
180	8.83187327	0.952399564	33.01677681	0.055222807	2.015990174	0.192685902
240	5.36160297	0.251578574	38.98135086	0.004329868	2.989599678	0.152431691
300	3.621708295	1.061871838	33.01002806	0.014000432	4.666728873	0.126751866
400	1.632581244	0.428917639	28.4920567	0.031017903	11.0440645	0.098592103
500	1.943360217	0.271353684	26.72680453	0.011792065	18.32477427	0.092627622
600	0.698686258	0.025910879	20.76488064	0.008399417	19.98570298	0.092319854
700	1.566185871	0.001753622	16.50050802	0.003677461	19.42064575	0.092355253
800	2.266779006	0.132721945	15.66779145	0.00867568	19.03043219	0.092374488

Tabelle D.2: Übersicht über integrierte experimentelle Daten: Datensatz 2, Maximale intrazelluläre Methionin-Konzentrationen [mM] und die dazugehörigen maximalen Fluoreszenzen [spezifische Fluoreszenz].

	(1) - 27.02.2011		(2) - 01.03.2011	
	Konzentration	Fluoreszenz	Konzentration	Fluoreszenz
10% Ala-Met	0.204746932	0.16	0.167826362	0.054
20% Ala-Met	0.825495965	0.237	0.722928038	0.114
30% Ala-Met	2.204367036	0.24	1.952659282	0.243
50% Ala-Met	5.301378557	0.402	4.351729622	0.488
70% Ala-Met	11.26270526	0.542	8.051732271	0.716
90% Ala-Met	14.03275919	0.669	10.19141119	0.813
100% Ala-Met	16.50817978	0.779	14.36354306	0.886
3mM Met-Met	25.52051235	1.062	31.51386075	1.362

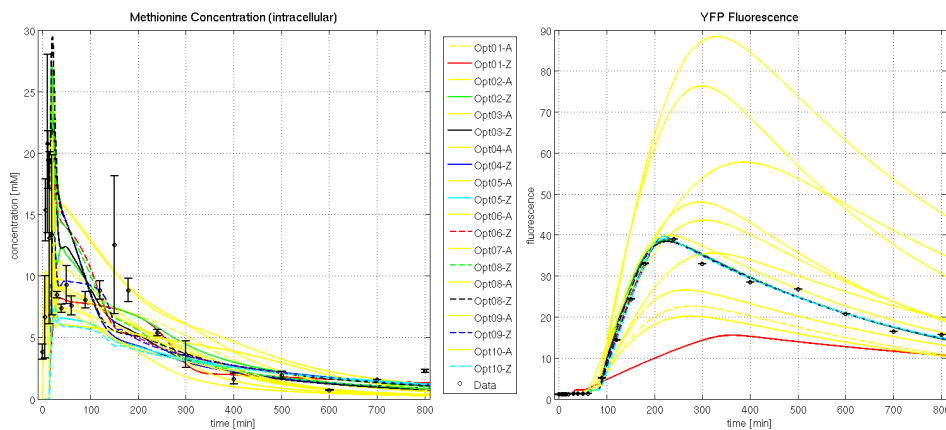


Abbildung D.1: Ergebnisse der Optimierung 11: Dargestellt sind jeweils die simulierten Verläufe der intrazellulären Methionin-Konzentration und des Fluoreszenz-Signals erzeugt mit verrauschten Parametern vor der Optimierung (gelb) und mit den optimierten Parametern (vgl. Tabelle D.4).

Tabelle D.3: Übersicht über alle Optimierungen.

- Opt1 Optimierung der Modellparameter auf Basis der neuen experimentellen Daten der intrazellulären Methionin-Konzentration und des Fluoreszenz-Signals (D.1). Gute Anpassung der Fluoreszenz, keine Anpassung Konzentration. Grund: Das Dipeptid wird so gut wie gar nicht aufgenommen.
- Opt2 Ausschließlich Einsatz der intrazellulären Konzentration als Messwert. Schlechte Anpassung der Dynamik des Konzentrationsverlaufes, im stationären Bereich passt die Anpassung. Keine Übereinstimmung der des Fluoreszenz-Signales mit dem Messwert. Der Parameter $PROM_0$ stößt an den oberen Grenzwert.
- Opt3 Erhöhung des oberen Grenzwertes für $PROM_0$ auf 100. $PROM_0$ wird bei der Optimierung auf einen Wert angepasst, der unter dem alten Grenzwert wert? Veränderte Parameter, allerdings ähnliche Anpassung der Messwert-Verläufe wie bei Opt2.
- Opt4 Festsetzen aller Parameter, die im Zusammenhang mit Dipeptid-Aufnahme und Substrat-Exkretion stehen. Oszillierender Konzentrationsverlauf im dynamischen Abschnitt bis 300 min.
- Opt5 Anpassung der Standardabweichungen aller Messungen auf 20%. Beste bisherige Anpassung der experimentellen Verläufe.
- Opt6 Zurücksetzen des oberen Grenzwertes für $PROM_0$ auf 10 - Kurve des Fluoreszenz-Signales ist flacher.
- Opt7 Erste multiple Optimierung mit 10 Läufen: Verrauschen der Startwerte aller Parameter zufallsbasiert um maximal 10%. Lauf 4 zeigt starke Abweichungen.
- Opt8 Weitere multiple Optimierung mit 50%igem Rauschen und 10 Läufen.
- Opt9 Bisher wurde das Fluoreszenz-Signal auf den simulierten Verlauf von $PROT$ angepasst. Nun Fitting mit $PROT_f$, welches ein zeitlich verschobenen Verlauf zu $PROT$ aufweist. Multiple Optimierung mit 20%igem Rauschen und 10 Läufen. Gute Anpassungen, beste in Lauf 8. Seltsames Verhalten der simulierten Konzentration (kurzer Abfall, dann wieder Anstieg) bei 30 min.
- Opt10 Zurücksetzen der Standardabweichungen der Messdaten auf ihre ursprünglichen, experimentell bestimmten Werte. Änderung der Startwerte auf die angepassten Parameter von Opt9-8. Multiple Optimierungen mit 20%igem Rauschen führen zu mit Opt9 vergleichbaren Ergebnissen. Abweichungen in den angepassten Konzentrationsverläufen.
- Opt11 Multiple Optimierungen mit 20%igem Rauschen und den ursprünglichen Startwerten führen zu den finalen Optimierungsergebnissen.

Tabelle D.4: Ergebnisse der Optimierung 11: Erweiterung der Tabelle 7.3

Parameter	11-1	11-2	11-3	11-4	11-5	11-6	11-7	11-8	11-9	11-10
$DP_{ex,0}$										
$SENS_0$	10.1121	10.8802	11.4179	7.7814	9.2680	9.1321	11.8964	16.0794	10.3257	18.0878
$PRO M_0$	4.5962	6.4327	8.3005	5.6077	4.7649	4.7984	6.5161	8.1696	20.0551	3.6373
DP_{in}										
n_{DP}										
$n_{DP,S}$										
μ										
D										
$vmax_{uptDP}$	0.0010	0.0019	0.0019	0.0019	0.0018	0.0018	0.0019	0.0018	0.0018	0.0018
kn_{uptDP}	50.9568	106.4442	66.6331	57.7093	45.5942	70.5143	87.7815	116.4746	47.8458	19.3932
$vmax_{excS}$	0.0139	0.0517	0.0465	0.0507	0.0295	0.0345	0.0411	0.0458	0.0264	0.0282
km_{excS}	0.2458	1.3438	1.0345	1.1374	0.3010	0.4823	0.7721	0.9590	0.1371	0.1855
$kd_{bind, f,ud}$	0.0060	0.0041	0.0035	0.0093	0.0126	0.0071	0.0046	0.0057	0.0034	0.0074
$kd_{bind, bud}$	20.0000	20.0000	20.0000	20.0000	20.0000	16.4055	20.0000	18.0738	20.0000	20.0000
$kd_{act, f,ud}$	18.1839	70.2503	61.7129	30.7217	37.2896	43.3175	68.5411	79.4041	57.1191	78.0967
$kd_{act, bud}$	0.0001	0.0006	0.0010	0.0019	0.0001	0.0007	0.0016	0.0007	0.0007	0.0002
k_{pro}	0.7182	0.9566	0.8242	0.5861	0.8083	0.5174	1.2432	0.5889	0.5344	0.8135
d_{del}	7.7662	0.3647	0.2843	0.9417	0.6711	0.4810	0.7722	0.5339	1.3249	0.2757
ka_{prodEX}	0.0392	0.1208	0.0827	0.1376	0.1865	0.1160	0.0871	0.0612	0.1219	0.1371
k_{degrEX}	0.5790	0.1731	0.2506	1.0043	0.5531	0.5525	0.4199	0.4619	0.4617	0.4735
$time_{shift}$	135.7789	890.5793	304.1234	248.5479	603.2475	822.2022	710.8384	285.0046	667.6544	239.2809
	0.2711	0.7592	0.5114	0.7817	1.5812	2.1751	1.1986	0.6499	0.9244	1.1281
	30.0000	60.0000	60.0000	60.0000	60.0000	60.0000	60.0000	60.0000	60.0000	60.0000

D.2.2 Sensitivitätsanalyse

Abbildung D.2 zeigt die Sensitivitätsmatrix zur Optimierung 11-6.

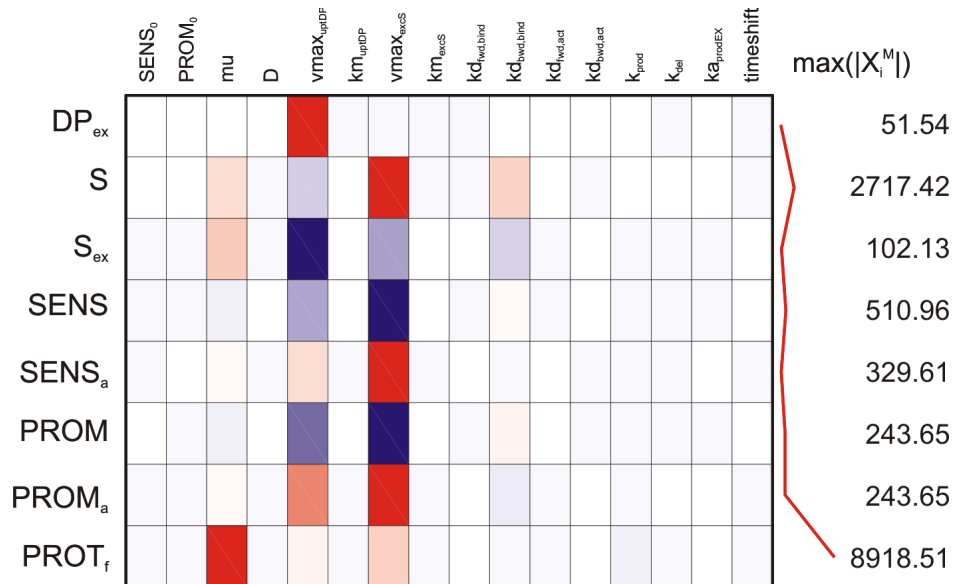


Abbildung D.2: Sensitivitätsmatrix 11-6: Eine Erhöhung des Parameters führt zu einer Verringerung der (potentiellen) Messung, wenn die Zelle rot eingefärbt ist und zu einer Erhöhung der (potentiellen) Messung, wenn die Zelle blau eingefärbt ist. Je kräftiger die Farbe, umso größer ist der Einfluss.