

Analyse und Optimierung eines  
industriellen Biotransformationsprozesses  
zur Herstellung von Tryptophan

Vom Fachbereich Chemie  
der Universität Hannover

zur Erlangung des Grades

Doktor der Naturwissenschaften

- Dr. rer. nat. -

genehmigte Dissertation

von

Dörte Solle

Geboren am 09.11.1968 in Lehrte

Hannover, im Juli 2003

Referent: PD. Dr. B. Hitzmann

Korreferent: Prof. Dr. T. Scheper

Tag der Promotion: Freitag, den 09.05.03

Die vorliegende Arbeit wurde in der Zeit von April 1999 bis April 2003 am Institut für Technische Chemie der Universität Hannover unter der Leitung von Herrn PD Dr. Bernd Hitzmann angefertigt. Für die interessante Themenstellung, die gebotene Arbeitsmöglichkeit und die freundliche Unterstützung möchte ich mich herzlich bei PD Dr. Bernd Hitzmann bedanken. Das mir entgegengebrachte Vertrauen hat diese Arbeit erst ermöglicht. Die anregenden Diskussionen mit Ihm hatten einen maßgeblichen Anteil am Erfolg der Arbeit.

Herrn Prof. Dr. Thomas Scheper danke ich für die Übernahme des Korreferates.

Der Deutschen Bundesstiftung Umwelt danke ich für die finanzielle Förderung im Rahmen des Förderschwerpunkts *Integrierte Biotechnologie Verbund Sensorik*.

Mein Dank gilt der AMINO GmbH, Frellstedt, für die Möglichkeit am industriellen Prozess zu forschen. Insbesondere Frau Dr. Birgit Kläßen und Herrn Dr. Robert Faurie danke ich für ihre Unterstützung. Für die Koordination der Messungen im Werk und den konstruktiven Erfahrungsaustausch danke ich auf diesem Wege Herrn Seja und seinen Mitarbeitern in Werk III.

Für den Aufbau der HPLC-Analytik, die Analyse der zahlreichen Proben und auch für ihre Hilfsbereitschaft danke ich Frau Martina Weiss.

Den Mitarbeitern des Institutes für Technische Chemie in Hannover danke ich für die angenehme Atmosphäre, die anregenden Diskussionen und die gute Zusammenarbeit. Mein besonderer Dank geht an den Arbeitskreis Hitzmann für die Unterstützung und das hervorragende Arbeitsklima auch in schwierigen Zeiten.

Carsten Protsch möchte ich für sein Engagement für unser Projekt und seinen uneingeschränkten Optimismus danken.

Mein Dank für die angenehme Zeit auch neben der Forschung geht an Daniela Böhl, Dirk Hüll, Gerald Ehlerding, Susan Baxmann, Christian Harbeck, Kirsten Hantelmann und dem KK.

Für die liebevolle Unterstützung danke ich meiner Familie und Klaus Kapels. Ohne Euch wäre es nie zu dieser Arbeit gekommen.

## **Analyse und Optimierung eines industriellen Biotransformationsprozesses zur Herstellung von Tryptophan**

L-Tryptophan wird bei der AMINO GmbH durch die Biotransformation von L-Serin und Indol mit dem Enzym Tryptophansynthase aus *Escherichia coli* Zellen hergestellt. Um Tryptophan auf dem Weltmarkt konkurrenzfähig anbieten und modernen Umweltstandards gerecht werden zu können, müssen die Substrate möglichst vollständig umgesetzt werden. Für eine optimale Umsetzung ist eine Prozessführung, abhängig vom aktuellen Prozesszustand, unerlässlich.

Zur Beurteilung des Prozesszustandes müssen die Konzentrationen der Substrate Indol und Serin und des Produktes Tryptophan online gemessen und ausgewertet werden. Zur Messung der Prozessvariablen wird die 2D-Fluoreszenzspektroskopie eingesetzt. Die Auswertung der Fluoreszenzmessungen erfolgt durch multivariate Methoden. Aus den Fluoreszenzspektren können durch Anwendung chemometrischer Modelle die Prozessvariablen online und intime mit einem mittleren Fehler von etwa 5 % für Indol und Tryptophan und einem Fehler von etwa 10 % für Serin vorhergesagt werden.

Zur Simulation des Prozesses wird ein Bioprozessmodell aufgestellt und die kinetischen Parameter des Modells identifiziert. Basierend auf der Sensitivitätsanalyse konnte die Anzahl der Modellparameter halbiert werden ohne dass das Modell signifikant an Leistungsstärke verliert. Das Prozessmodell wird zur Filterung der Onlinemesswerte aus der Fluoreszenzmessung eingesetzt, um rauschfreie Vorhersagen zu erhalten; hierdurch wird die Online-schätzung der Prozessvariablen auf mittlere 5 % für alle drei Variablen verbessert.

Unter Berücksichtigung der Prozessführung wird mit dem Bioprozessmodell die optimale Führung des Prozesses festgelegt. Eine Halbierung der Prozesszeit ist bei optimaler Prozessregelung möglich, was zu einer Energieeinsparung führt. Außerdem kann die Produktausbeute theoretisch bis zu 22 % gesteigert werden. Basierend auf der optimalen Prozessführungsstrategie und unter Berücksichtigung der online Analyse des aktuellen Prozesszustandes kann der Biotransformationsprozess Ressourcen schonend geregelt werden und eine Erhöhung der Raum-Zeit-Ausbeute erreicht werden.

Schlagworte: 2D-Fluoreszenzmessung, Multivariate Auswertung, Bioprozessmodell, Parameteridentifikation, Onlinevorhersage, Optimierung, Biotransformation, Tryptophan

## **Analysis and Optimisation of an Industrial Biotransformation Process for Tryptophan Production**

At the AMINO GmbH L-Tryptophan is produced by biotransformation of L-Serine and Indol using the *Escherichia coli* enzym Tryptophansynthase. To sell Tryptophan competitive on the world market and to justice modern ecology standards, the substrates have to used up totally. Therefore, monitoring and control of the process is obligate.

To determine the process conditions concentrations of both substrates and the product has to be measured, analysed and interpreted. Therefore the 2D-Fluorescence spectroscopy is used together with multivariate analyses to predict the process variables on line and in time based on the measurement. The root mean square error of prediction for Indol and Tryptophan is about 5 % and for Serin it is about 10 %.

A bioprocess model is established to simulate the process variables of the industrial process and the kinetic parameters are identified. Based on the sensitivity analysis the number of model parameters are reduced without losing significant simulation power. The model is used to filter the online measurement for low-noise predictions and to improve the prediction error about less than 5 % for all variables.

Considering the process conditions the bioprocess model is used to decide the optimal control. The result of the optimal process control is 50 % shorter process time for a complete biotransformation leading to saving energy. Besides the product yield can be improved theoretically up to 22 %. Based on the optimal process control strategy and considering the analysed process state the biotransformation can be controlled for saving of resources and improving yield.

Key words: 2D-Fluorescence spectroscopy, Multivariate Analysis, Bioprocess modelling, Parameter identification, Online prediction, Optimization, Biotransformation, Tryptophan

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b> .....	<b>1</b>
<b>2</b>	<b>Tryptophan und Tryptophanproduktion</b> .....	<b>2</b>
2.1	Tryptophan .....	2
2.2	Tryptophansynthese.....	3
2.3	Biotechnische Tryptophanproduktion .....	4
2.3.1	Herstellung des Serineluates.....	4
2.3.2	Herstellung der <i>Escherichia coli</i> B10 Zellen.....	5
2.3.3	Durchführung der Biotransformation .....	6
2.4	Das Enzym Tryptophansynthase .....	7
<b>3</b>	<b>Messungen am industriellen Prozess</b> .....	<b>9</b>
3.1	Messmethoden .....	9
3.1.1	2D-Fluoreszenzmessung.....	9
3.1.2	Referenzmessung.....	11
3.2	Messaufbau am industriellen Produktionsprozess.....	12
3.3	Messungen bei den Biotransformationen .....	13
3.3.1	Die beobachteten Tryptophanproduktionsprozesse.....	13
3.3.2	Onlinemesswerte .....	13
3.3.3	Offlinemesswerte.....	14
3.3.4	Fluoreszenzmessungen .....	15
<b>4</b>	<b>Mathematische Modelle</b> .....	<b>17</b>
4.1	Chemometrische Modelle.....	17
4.1.1	Hauptkomponentenanalyse.....	19
4.1.2	Regression nach der Methode der kleinsten Fehlerquadrate .....	22
4.1.3	Multidimensionale Regression .....	22
4.1.4	Vorhersage.....	23
4.1.5	Modellvalidierung .....	24
4.1.6	Sensitivitätsanalyse.....	25
4.1.7	Modellerstellung und -anwendung .....	26
4.2	Bioprozessmodell .....	27
4.2.1	Reaktionskinetik .....	28
4.2.2	Parameteridentifikation .....	29
4.2.3	Sensitivitätsanalyse.....	31
4.2.4	Optimierung.....	33
4.3	Onlineanwendung.....	33
4.3.1	Onlinevorhersage.....	34
4.3.2	Filterung der Messwerte .....	34

<b>5</b>	<b>Ergebnisse .....</b>	<b>37</b>
5.1	Das theoretische Prozessmodell .....	37
5.1.1	Parameteridentifikation .....	42
5.1.2	Parametersensitivität.....	47
5.1.3	Parameterreduktion.....	50
5.1.4	Modellvalidierung .....	55
5.2	Chemometrik .....	58
5.2.1	Hauptkomponentenanalyse eines Prozesses.....	58
5.2.2	Hauptkomponentenanalyse mehrerer Prozesse .....	63
5.2.3	Erstellung der chemometrischen Modelle .....	65
5.2.4	Anwendung der chemometrischen Modelle .....	70
5.3	Onlineanwendung der Modelle .....	73
5.3.1	Die Onlinesimulation.....	73
5.3.2	Die Onlinevorhersage .....	75
5.3.3	Filterung der Onlinevorhersagen .....	78
5.4	Optimierung.....	80
<b>6</b>	<b>Zusammenfassung und Ausblick.....</b>	<b>83</b>
<b>7</b>	<b>Literaturverzeichnis .....</b>	<b>86</b>
<b>8</b>	<b>Anhang .....</b>	<b>92</b>
8.1	Verzeichnis der Abkürzungen .....	92
8.2	Algorithmen.....	93
8.2.1	NIPALS-Algorithmus.....	93
8.2.2	PLS1-Algorithmus .....	94
8.3	Modelle in Simulinkdarstellung .....	95
8.4	Parameter des Kalman-Filters .....	99

# 1 Einleitung

Bei der AMINO GmbH, Frellstedt, wird aus Zuckerrübenmelasse u.a. Serin gewonnen, das unter Zugabe von Indol und dem Enzym Tryptophansynthase zur essenziellen Aminosäure Tryptophan umgesetzt wird. Die biotechnische Herstellung der pharmakologisch wirksamen Aminosäure L-Tryptophan stellt eine ökonomische und ökologische Alternative zur chemischen Synthese dar. Um das Produkt der Biotransformation auf dem Weltmarkt konkurrenzfähig anbieten zu können und modernen Umweltstandards bei der Produktion gerecht zu werden, müssen die Substrate möglichst vollständig in das Produkt umgesetzt werden. Für eine vollständige Umsetzung der Substrate ist eine Prozessführung in Abhängigkeit vom aktuellen Prozesszustand unerlässlich.

Biologische Vorgänge sind meist sehr komplex und daher nur begrenzt beobachtbar, außerdem findet die Reaktion in einer komplexen Matrix statt und unterliegt variierenden Prozessbedingungen. Zur Beurteilung des Prozesszustandes müssen die Konzentrationen der Substrate Indol und Serin sowie des Produktes Tryptophan online gemessen und verarbeitet werden. Da Indol und Tryptophan fluoreszieren, wird zur Messung der Prozessvariablen die 2D-Fluoreszenzspektroskopie eingesetzt. Zur Auswertung der Fluoreszenzmessungen werden multivariate Methoden genutzt, um die relevanten Informationen aus der großen Menge online ermittelter Daten zu extrahieren.

Zur Simulation des Prozesses wird ein Prozessmodell erstellt und die kinetischen Parameter der Reaktion werden identifiziert. Durch Variation der Prozessbedingungen wird mit dem Modell die optimale Führungsstrategie des Prozesses identifiziert. Basierend auf der optimalen Prozessführungsstrategie und unter Berücksichtigung des online ermittelten Prozesszustandes kann der Prozess Ressourcen schonend geführt und eine Erhöhung der Raum-Zeit-Ausbeute erreicht werden.

Ziel dieser Arbeit ist es, Monitoring und Control der biotechnischen Tryptophanproduktion zu verbessern. Zunächst soll hierfür eine schnelle und industrietaugliche Messung der Prozessvariablen realisiert werden; danach soll die Onlineauswertung der Daten ermöglicht sowie die Prozessführungsstrategie festgelegt werden. Dadurch kann die Optimierung zur Einsparung der Ressourcen und Erhöhung der Raum-Zeit-Ausbeute durchgeführt werden, um die Aminosäure L-Tryptophan unter ökologischen und ökonomischen Gesichtspunkten optimal zu produzieren.

## 2 Tryptophan und Tryptophanproduktion

Das Produkt der Biotransformation Tryptophan sowie seine Bedeutung und Wirkung wird in diesem Kapitel beschrieben. Verschiedene Synthesewege für Tryptophan werden vorgestellt, insbesondere die Biotransformation, ihre Ausgangsstoffe und der benutzte Biokatalysator. Für die enzymatische Bildung von Tryptophan durch den Biokatalysator werden die Bedingungen der Umsetzungen und ihre Grenzen aufgezeigt.

### 2.1 Tryptophan

L-Tryptophan ist eine essentielle, aromatische Aminosäure. In Abbildung 2.1 ist das Tryptophanmolekül als Valenzstrichformel und räumlich dargestellt. Tryptophan besteht aus einem Indolring und einer Alaninkette. Das optisch aktive Zentrum in der Alaninkette ist in Abbildung 2.1 gekennzeichnet.

Wie viele metabolisch aktive Substanzen ist das D-Isomer nicht biochemisch wirksam. Da Tryptophan nicht vom menschlichen Körper synthetisiert werden kann, muss es durch die Nahrung aufgenommen werden. In verschiedenen Obstsorten, wie Bananen und Erdbeeren, kommt Tryptophan vor, ebenso wie in tierischem Eiweiß, besonders in Milch.

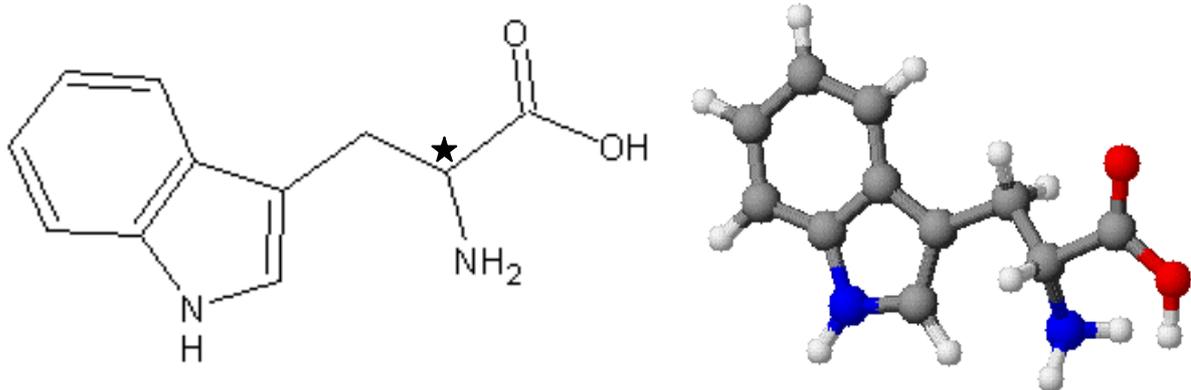


Abbildung 2.1: Tryptophan (links: 2D-Darstellung; rechts: 3D-Darstellung)

Tryptophan wird zu 98 % im menschlichen Organismus über Chinolinsäure zu Nicotinsäureamid und zu 2 % zu Serotonin metabolisiert; weshalb Nicotinsäure-Mangelercheinungen nur bei tryptophanarmer Ernährung oder eines gestörten Tryptophanstoffwechsels auftreten kann. Tryptophan wird sowohl als Nahrungsmittel- und Futtermittelzusatz verwendet, als auch in der Medizin als Sedativum und Antidepressivum eingesetzt. Darüber hinaus wird es als Rohstoff in Infusionslösungen eingesetzt und in der Mikrobiologie als Medienbestandteil benötigt.

Die pharmakologische Wirkung von L-Tryptophan beruht hauptsächlich auf der Anhebung des Serotoninspiegels im Zentralnervensystem. Serotonin ist ein Neurotransmitter und eine

Vorstufe des Epiphysenhormons Melatonin. Nach der Transmission im synaptischen Spalt wird das Serotonin wieder aufgenommen und anschließend abgebaut. Ein hoher Serotoninspiegel wirkt schlafinduzierend, ein niedriger hingegen führt zu Antriebslosigkeit, Depressionen und Ängsten. Da Serotonin im Gegensatz zum Tryptophan nicht die Blut-Hirn-Schranke passieren kann, führt entweder eine erhöhte Zufuhr von Tryptophan oder die Einnahme von selektiven Serotoninwiederaufnahmehemmern zu einem gleichmäßig hohen Serotoninspiegel im Gehirn.

Der vielfältige Einsatz von L-Tryptophan wurde 1989 abrupt unterbrochen, da Fälle des Eosinophilia-Myalgia-Syndrom (EMS) mit der Einnahme von L-Tryptophan in Zusammenhang gebracht wurden. EMS, eine in Einzelfällen tödlich verlaufende Krankheit, wird durch eine erhöhte Zahl eosinophiler Zellen charakterisiert und bewirkt Muskelschmerzen sowie eine Vielzahl unspezifischer Symptome. Nicht der Wirkstoff L-Tryptophan selbst, sondern eine beim Downstreamprocessing erfolgte Verunreinigung bei der Herstellung von L-Tryptophan durch Ethylen-bis-tryptophan war hierfür verantwortlich. Durch verbesserte Produktionsverfahren und stärkere Qualitätskontrollen ist L-Tryptophan seit 1995 wieder zugelassen, da das Produkt den hohen Anforderungen an ein Pharmazeutikum gerecht wird [FAEDER 1970].

## 2.2 Tryptophansynthese

Es sind eine Reihe von Verfahren zur Herstellung von L-Tryptophan bekannt. Die chemische Synthese geht von Indolringsystemen als Edukt aus und baut anschließend die Alaninkette auf. Eine weitere Möglichkeit besteht darin, den Indolring während der Tryptophansynthese aufzubauen, wie es bei der Warner-Moe-Synthese (Degussa 1978) Anwendung gefunden hat. Die Synthese eines durch Hydrolase spaltbaren Tryptophanderivates ist ebenso möglich. Die chemischen Synthesen machen eine Vielzahl von Tryptophanderivaten zugänglich, die zumeist als racemische Gemische aus D- und L-Tryptophan vorliegen [MÜLLER 1999].

Biotechnisch wird Tryptophan durch Fermentation von *Bacillus subtilis*, *Bacillus amyloliquefaciens*, *Escherichia coli* oder *Corynebacterium glutamicum* hergestellt, da Mikroorganismen im Gegensatz zu Wirbeltieren die Tryptophanbiosynthese durchführen können [MÜLLER 1999]. Für fermentative Herstellungen von Tryptophan müssen die Mikroorganismen so modifiziert werden, dass eine Überproduktion an Tryptophan angeregt wird. Außerdem darf keine Produkthemmung der am Tryptophananabolismus beteiligten Enzyme vorliegen. Die Tryptophan- und Substratkatabolismusaktivität sollte ebenfalls stark reduziert sein.

Statt der chemischen Synthese oder der fermentativen Herstellung von Tryptophan ist auch die Biosynthese oder Biotransformation aus Indolderivaten und Serin möglich. Der Vorteil der Biotransformation gegenüber der Fermentation oder der chemischen Synthese ist die

Selektivität der Enantiomere, die hohe Substratselektivität, die hohe Produktkonzentration und Reinheit des Rohproduktes. Die Qualität des Rohproduktes ist hierbei im wesentlichen durch die Qualität der eingesetzten Edukte bestimmt.

### 2.3 Biotechnische Tryptophanproduktion

L-Serin und Indol werden als Edukte der Biotransformation eingesetzt. Durch Kondensation der Edukte mittels Tryptophansynthase und dem Co-Enzym Pyridoxalphosphat (PLP) wird L-Tryptophan enzymatisch synthetisiert. Das Reaktionsschema ist in Abbildung 2.2 dargestellt.

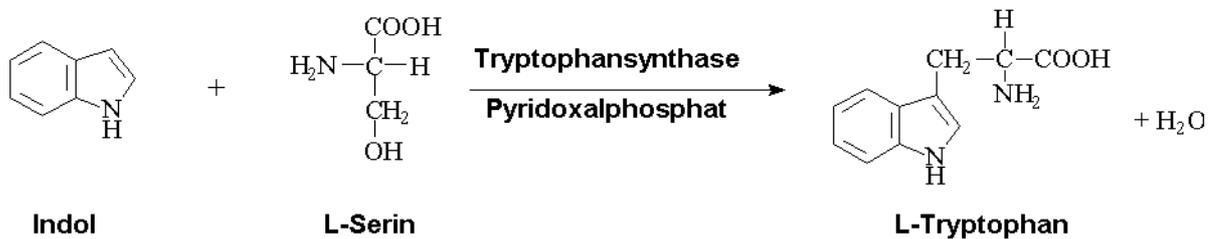


Abbildung 2.2: Die Reaktion von Indol und Serin zu Tryptophan

Bei der AMINO GmbH in Frellstedt werden aus Zuckerrübenmelasse durch chromatographische Aufreinigung verschiedene Eluate mit angereicherten Aminosäuren gewonnen. Eine dieser Aminosäuren ist Serin, das nach weiterer Aufarbeitung zur Tryptophanproduktion eingesetzt wird. Dem L-Serineluat wird Pyridoxalphosphat und Indol zugesetzt und die Biotransformation durch Zugabe von Tryptophansynthase in *Escherichia coli* B10 Zellen gestartet.

#### 2.3.1 Herstellung des Serineluates

Die bei der Zuckerherstellung anfallende Melasse wird durch Ionenausschlusschromatographie in Fraktionen aufgeteilt. Während des Chromatographieprozesses werden die Salze der Melasse von den übrigen Bestandteilen getrennt und eluieren in der Schlempefraktion (Salzfraktion). Die übrigen Bestandteile (Saccharose, Betain und Aminosäuren) werden nach ihrer Größe und Ladung aufgetrennt. Mischfraktionen werden gesammelt und der Säule erneut zur Trennung zugeführt, dadurch werden Fraktionen mit hoher Produktreinheit gewonnen. Eine optimale Bestimmung der Schnittgrenzen kann über Onlineanalysen durch Biosensoren oder optische Methoden erfolgen, wodurch hohe Produktreinheiten gewährleistet werden können ([HARBECK 2001A], [HARBECK 2001B], [IRTEL 2000], [ULBER 2000]).

Abbildung 2.3 zeigt das Fließdiagramm des gesamten Prozesses von der Melasseaufarbeitung zu D,L-Serin und Serineluat mit der nachfolgenden Biotransformation und der Aufreinigung zum Endprodukt.

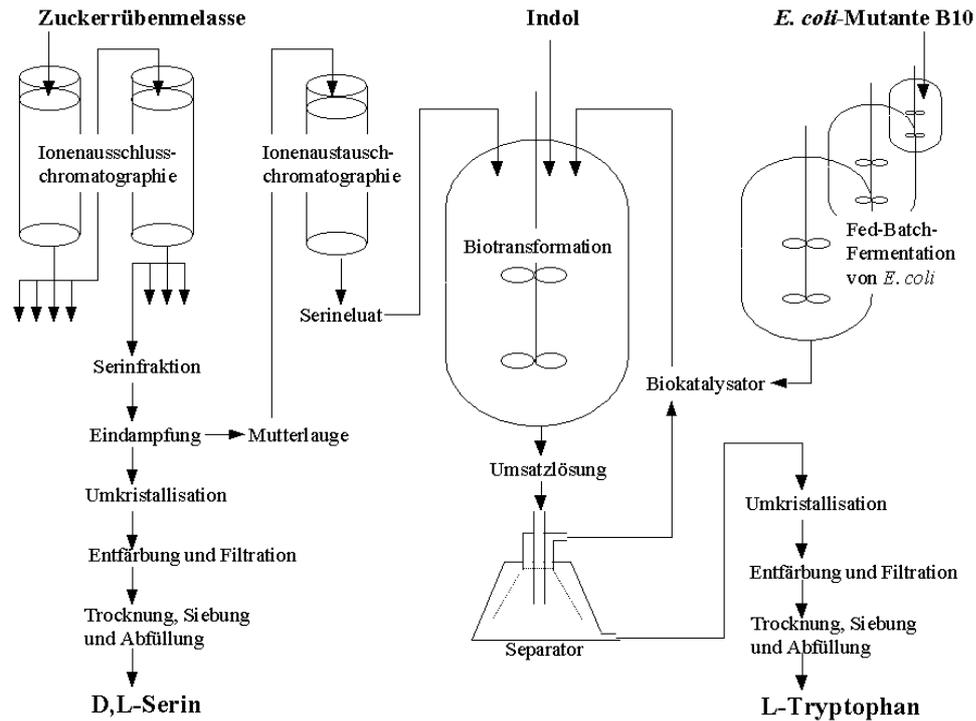


Abbildung 2.3: Fließdiagramm der Melasseaufarbeitung zur Tryptophanproduktion

Die Serinfraktion besteht zum größten Teil aus L-Serin und zu einem geringen Anteil aus D/L-Serin. Das Racemat wird durch ein Fällverfahren quantitativ abgetrennt und nach weiteren Aufreinigungen als Pharmawirkstoff (D,L-Serin) verkauft. Die verbleibende verdünnte Serinfraktion (Mutterlauge) wird durch Ionenaustauschchromatographie von noch verbliebenen Zuckern und anderen Aminosäuren abgetrennt. Dieses Serineluat wird für die Biotransformation eingesetzt.

Das Serineluat hat eine Serinkonzentration von 0,8 - 1,3 %, je nach Melasse und Aufarbeitungsgüte. Im Eluat liegt Serin zu über 95 % in L-Form vor, daneben enthält das Eluat noch geringe Konzentrationen verschiedener Aminosäuren und anderer Begleitstoffe. Das Eluat weist eine starke Braunfärbung auf, die abhängig von Rohstoff und Aufarbeitung variiert.

### 2.3.2 Herstellung der *Escherichia coli* B10 Zellen

Die Biotransformation kann mit freiem Enzym oder ganzen Zellen durchgeführt werden. Bei der Reaktion mit freien Enzymen findet keine Diffusions- und Transferlimitierung durch die Zellwand statt und es gibt keine unerwünschten Nebenreaktionen durch andere Enzyme. Bei der Verwendung ganzer Zellen arbeiten verschiedene Enzyme und Co-Faktoren in natürlicher Weise zusammen und erhöhen dadurch die Stabilität der Enzyme [MATEUS 1999]. Allerdings werden die katalytisch aktiven Enzyme durch Proteasen denaturiert. Die Diffusionslimitierung durch die Zellwand bei der Verwendung der ganzen Zelle wird durch den Einsatz von Indol reduziert, da Indol die Zellwand permeabilisiert [VORLOP 1983]. Unerwünschte Neben-

reaktionen können durch die Verwendung von Bakterienstämmen reduziert werden, die die entsprechenden Enzyme nicht exprimieren [KAWATA 1990]. Darüber hinaus ist für ganze Zellen die Möglichkeit der Immobilisierung und einer damit verbundenen Optimierung der Katalysatorausnutzung gegeben ([KLEIN 1984], [WAGNER 1982]).

Für die Biotransformation werden bei der AMINO GmbH ganze, Tryptophan auxotrophe, *Escherichia coli* Zellen vom Stamm B10 eingesetzt. Durch natürliche Mutation des Repressors wird das Enzym Tryptophansynthase überexprimiert und darüber hinaus Tryptophanase nicht exprimiert ([BANG 1981], [BANG 1983]). Durch diese Mutationen ist der Abbau des Produktes erheblich reduziert und die Umsetzung erfolgt mit einer hohen Ausbeute ([FUCHS 1986], [SCOGMAN 1984], [TRIBE 1979]). Der Mikroorganismus wird, wie in Abbildung 2.3 dargestellt, schrittweise mittels Fed-Batch-Fermentation (30 L-, 300 L- und 3000 L-Fermenter) in einer sterilen Fermenterkaskade angezchtet.

Nach der Umsetzung werden die Zellen von der Suspension separiert und erneut einem Prozess zugeführt (Abbildung 2.3). Durch den Einsatz ganzer Zellen ist die Separation von der Umsatzlösung und damit eine Wiederbenutzung des Katalysators möglich. Durch die Kombination von frischen Zellen mit Zellen, die bereits bei vorangegangenen Umsetzungen eingesetzt und abgetrennt wurden, wird versucht die Enzymaktivität für alle Prozesse auf gleichbleibendem Niveau zu halten.

### 2.3.3 Durchführung der Biotransformation

Die Biotransformation, wie in der Literatur beschrieben ([SOLLE 2001], [ULBER 2001B]), wird in einem 10.000 L Reaktor im Batch-Verfahren durchgeführt. Es werden etwa 8.000 L Serin-eluat im Reaktor vorgelegt, mit 150 g PLP versetzt, auf 40 °C (optimale Reaktionstemperatur) aufgeheizt und auf einen pH-Wert von 8,3 eingestellt. 10 - 15 kg Indol werden hinzugegeben und in der Reaktionsmasse gelöst. Das Indol wird durch Erwärmung auf 90 °C verflüssigt, durch eine mit Wasserdampf geheizten Leitung gepumpt und über einen Sprühkopf fein verteilt in die Reaktionsmasse gegeben. Durch die so erzeugte große Oberfläche wird ein schnelles Lösen, des sonst in wässriger Lösung schwerlöslichen Substrates erreicht.

Zum Start des Prozesses werden insgesamt etwa 750 L bereits eingesetzte und frische Zellen zugegeben. Eine online-HPLC bestimmt im 15 minütigen Abstand die Konzentrationen von Indol und Tryptophan. Abhängig von dieser Onlinemessung wird Indol nachdosiert. Aufgrund der Indolinhibierung und der problematischen Aufarbeitung des Indols nach der Umsetzung werden hohe Indolkonzentrationen während des Prozesses vermieden. Nach etwa 8 h Reaktionszeit ist der Großteil des Serins zu Tryptophan umgesetzt.

Nach der Reaktion wird der Reaktorinhalt in einen Vorratsbehälter umgepumpt und die Zellen in einem Separator von der Reaktionsmasse getrennt (Abbildung 2.3). Die separierte

Zellmasse wird gesammelt, bei etwa 8 °C gelagert und in weiteren Umsetzungen eingesetzt. Die nun zellfreie Tryptophanlösung wird durch Filtrierung mit Aktivkohle aufgereinigt und das Tryptophan über mehrere Stufen hochrein auskristallisiert. Durch Umkristallisationen kann das Produkt in Pharmaqualität hergestellt und verkauft werden.

## 2.4 Das Enzym Tryptophansynthase

Das Enzym Tryptophansynthase EC 4.2.1.20 (auch Tryptophan Synthetase [HATANAKA 1962], [FAEDER 1971], [VORLOP 1983], [YANOFSKY 1960]) katalysiert die Reaktion von Indol und L-Serin zu L-Tryptophan unter Pyridoxal-5-phosphat Sättigung. Darüber hinaus werden weitere Reaktionen, die in Tabelle 2.1 zusammengefasst sind, von dem Enzym katalysiert. Die Reaktionen finden an einem heterotetrameren Komplex  $\alpha_2\beta_2$  statt, der aus zwei  $\alpha$ -Untereinheiten und einem  $\beta$ -Untereinheiten-Dimer gebildet wird, wovon je eine  $\beta$ -Untereinheit ein Molekül Pyridoxal-5-phosphat bindet [KIRSCHNER 1975].

Tabelle 2.1: Reaktionen, die vom Enzym Tryptophansynthase katalysiert werden [Yanofsky 1960]

	Reaktion	Katalysiert von
1	IndGly $\longrightarrow$ Ind + Gly	$\alpha$ ; $\alpha_2\beta_2$
2	Ind + Ser $\xrightarrow{PLP}$ Trp + H <sub>2</sub> O	$\beta_2$ ; $\alpha_2\beta_2$
3	IndGly + Ser $\xrightarrow{PLP}$ Trp + Gly + H <sub>2</sub> O	$\alpha_2\beta_2$
4	Ser $\xrightarrow{PLP}$ Pyr + Amm	$\beta_2$
5	MEth + Ser $\longrightarrow$ SCys + H <sub>2</sub> O	$\beta_2$ ; $\alpha_2\beta_2$

IndGly	Indol-3-glycerol-phosphat	Ind	Indol
Gly	D-Glyceraldehyd-3-phosphat	Ser	L-Serin
Trp	L-Tryptophan	Pyr	Pyruvat
PLP	Pyridoxalphosphat	MEth	$\beta$ -Mercaptoethanol
SCys	S-hydroxyethyl-L-cystein	Amm	Ammonium

Die  $\alpha$ -Untereinheit, an die das Indol bindet, katalysiert die Reaktion 1 (Tabelle 2.1), an der als einziges kein L-Serin beteiligt ist. Der  $\alpha_2\beta_2$ -Enzymkomplex ist für diese Reaktion etwa 100-fach aktiver. Die  $\beta$ -Untereinheiten sind die aktiven Zentren für die Tryptophanproduktion (Reaktionen 2 + 3 (Tabelle 2.1) [HEILMANN 1978], [DREWE 1985], [FAEDER 1970]), abhängig von der ionischen Umgebung ist der  $\alpha_2\beta_2$ -Komplex jedoch etwa 30 mal aktiver, als die  $\beta$ -Einheit allein [YANOFSKY 1960].

Die  $\beta$ -Untereinheiten zeigt neben der Tryptophankatalyse auch eine Serindesaminierung (Reaktion 4 (Tabelle 2.1)) [CRAWFORD 1964]. Darüber hinaus katalysiert diese Untereinheit nach Bindung von Serin und Pyridoxal-5-phosphat weitere Reaktionen, an denen beide Substrate beteiligt sind (z.B. Reaktion 5 (Tabelle 2.1)) [LANE 1983B]. Die Desaminierung und

die Tryptophanbildung wird durch Komplexbildung von Indol mit der  $\alpha$ -Untereinheit inhibiert, daher zeigt der  $\alpha_2\beta_2$ -Enzymkomplex nur eine geringe Desaminierung und eine Inhibierung der Tryptophansynthese schon bei geringen Indolkonzentrationen. Hohe Ammoniumkonzentrationen verstärken die Aktivität der Desaminierung, aber auch die Aktivität der  $\beta$ -Untereinheit für die Tryptophanbildung ([YANOFSKY 1960], [LEJA 1995]).

Die Reaktion 3 (Tabelle 2.1) wird ausschließlich von dem vollständigen Enzymkomplex katalysiert. Obwohl diese Reaktion die Summe der Reaktionen 1 und 2 (Tabelle 2.1) zu sein scheint, konnte dieses nicht bestätigt werden, da in einem solchen System kein freies Indol als Zwischenprodukt nachgewiesen werden konnte. Das Zwischenprodukt Indol wird nicht freigesetzt sondern gelangt durch einen mit Lösungsmittel gefüllten Tunnel zum aktiven Zentrum der  $\beta$ -Untereinheit ([LANE 1991], [VOET 1992]).

Indol ist nicht nur ein Substrat für die Reaktion, sondern wirkt auch inhibierend auf den Enzymkomplex ab etwa 0,03 mM. Heilmann [HEILMANN 1978] beschreibt diese Inhibierung als keine kompetitive Hemmung, da das inhibierende Indol nicht in Konkurrenz zum Substrat L-Serin steht, außerdem schließt er eine Substratinhibierung aus. Lane [LANE 1983A] bestätigt, dass Indol an die  $\alpha$ -Untereinheit bindet und nicht an die  $\beta$ -Untereinheit, an die das L-Serin gebunden wird. Damit kann der Mechanismus einer kompetitiven Hemmung ausgeschlossen werden, es wird stattdessen eine Inhibierung durch Wechselwirkungen zwischen den Untereinheiten angenommen. Für die  $\beta$ -Untereinheit allein findet Heilmann [HEILMANN 1978] keine Indolinhibierung für Indolkonzentrationen bis 1 mM. Aufgrund dieser Unterschiede zwischen der  $\beta$ -Untereinheit und dem gesamten Enzymkomplex liegt im industriellen Prozess eventuell ein Teil des Enzyms als  $\beta$ -Untereinheit vor. Eine Trennung der Untereinheiten ist durch die ionische Umgebung, Temperatur oder pH-Wert-Änderungen möglich. Hierfür spricht, dass die Zellen relativ hohe Indolkonzentrationen ohne Inhibierung des Enzyms vertragen können, dagegen spricht, dass bis jetzt keine Desaminierung des L-Serins dieser Zellen beschrieben wurde.

### 3 Messungen am industriellen Prozess

Am industriellen Produktionsprozess zur Herstellung von Tryptophan wurden verschiedene Messungen durchgeführt. Die genutzten Messmethoden werden vorgestellt und beschrieben. Außerdem wird der Aufbau der Onlinemessung und die spätere Auswertung am Prozess vorgestellt. Des Weiteren werden die Messungen an den beobachteten Prozessen beschrieben und ihre Bedeutung für die Optimierung herausgestellt.

#### 3.1 Messmethoden

An der industriellen Biotransformation zur Herstellung von Tryptophan werden zur Beobachtung der Biotransformation ständig Online-HPLC-Analysen durchgeführt. Zusätzlich wurden bei verschiedenen Prozessen Offlineproben genommen und darüber hinaus Online-fluoreszenzmessungen durchgeführt. Die verschiedenen Messmethoden und ihre Ergebnisse werden beispielhaft vorgestellt.

##### 3.1.1 2D-Fluoreszenzmessung

Für die Beobachtung und Regelung von biotechnischen Prozessen ist die Messung der biologischen Größen notwendig. Ein wichtiges Feld der bioanalytischen Systeme stellen nicht-invasive optische Systeme dar. Informationen über den Bioprozess, wie zellinterne Größen und Produktbildung, können mit Fluoreszenzsensoren gewonnen werden, die in einem breiten Spektralbereich für Exzitation und Emission arbeiten. Eine Vielzahl von Fluorophoren (z.B. Aminosäuren und Vitamine) können hierbei gleichzeitig detektiert werden [ULBER 2001A]. Die Fluoreszenz ist von den Konzentrationen der Fluorophore abhängig, sie wird jedoch auch durch Inner-Filter-Effekte, Quenching, Kaskadeneffekte und die Dichte bzw. Viskosität beeinflusst [PERKAMPUS 1993]. Die 2D-Fluoreszenzmessung ist eine Onlinemessmethode, die intime und nicht invasiv, insitu und sogar *in vivo* Analysen ermöglicht, da hierbei Substanzen in der Reaktionsmasse und Vorgänge in der Zelle selbst in Echtzeit beobachtet und detektiert werden können ([LINDEMANN 1998], [BANG 1981], [LI 1991]). Durch die hohe Variabilität der 2D-Fluoreszenzspektroskopie ist diese Messmethode besonders geeignet, um die komplexen Vorgänge bei biotechnischen Prozessen zu beobachten und gegebenenfalls frühzeitig regelnd einzugreifen, wie in der Literatur mehrfach beschrieben ([BÖHL 2001A], [BÖHL 2003], [HITZMANN 1998B], [MAROSE 1998], [MUKHERJEE 1999]). Das 2D-Fluoreszenzspektrometer BioView (DELTA Light and Optics, Dänemark) ist ein speziell für industrielle Applikationen entwickeltes Messgerät. Die elektronischen und optischen Komponenten befinden sich in einem Spritzwasser geschützten Stahlgehäuse. Über einen

Lichtleiter wird der BioView durch einen Standardstutzen an das Messsystem gekoppelt, so dass eine direkt Messung der Fluorophore in dem Reaktionsgemisch möglich ist. Das Prinzip der Aufnahme eines 2D-Fluoreszenzspektrums ist in Abbildung 3.1 dargestellt [BÖHL 2001B].

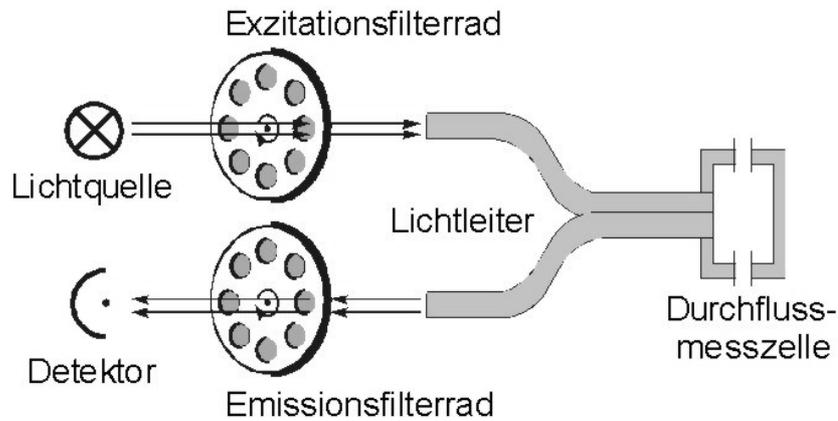


Abbildung 3.1: Messprinzip der 2D-Fluoreszenz

Licht einer Wellenlänge wird durch ein Filter im Filterrad ausgewählt, über den Lichtleiter eingestrahlt und die resultierende Fluoreszenz zu dem Detektor geleitet. Durch die Rotation der Filterräder, die eine Bandbreite von 20 nm haben, kann das gesamte Spektrum (270 – 550 nm Exzitation und 310 – 590 nm Emission, 120 Wellenlängenkombinationen) innerhalb einer Minute aufgezeichnet werden. Die 15\*15 Wellenlängenkombinationen umfassende Matrix ist oberhalb der Diagonale nicht mit Messpunkten besetzt, da die resultierende Fluoreszenz nur Licht geringerer Energie aufweist als das eingestrahlte Licht. In den Fluoreszenzspektren sind die biogenen Fluorophore bei charakteristischen Wellenlängenkombinationen erkennbar. In Abbildung 3.2 ist beispielhaft ein 2D-Fluoreszenzspektrum der Reaktionsmasse dargestellt.

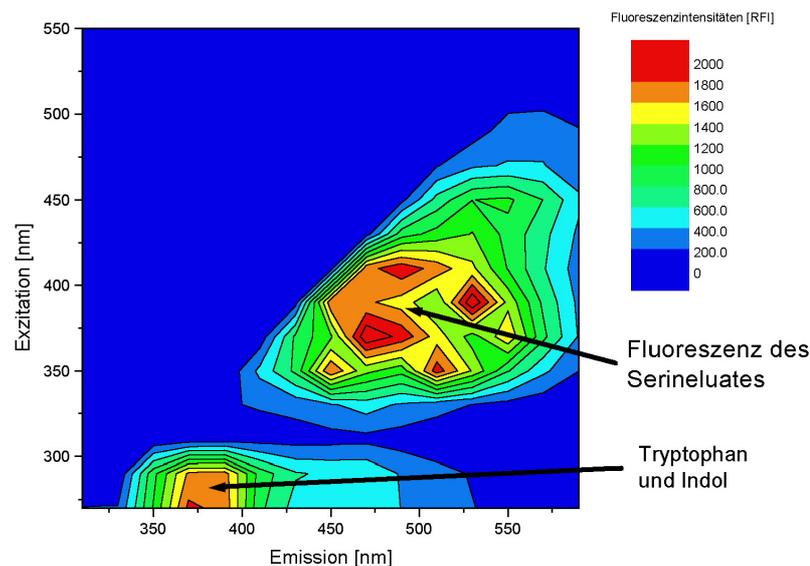


Abbildung 3.2: 2D-Fluoreszenzspektrum der Reaktionsmasse

Indol und Tryptophan fluoreszieren bei  $290 \text{ nm}_{\text{Ex}} / 370 \text{ nm}_{\text{Em}}$  daher wird durch die Tryptophanproduktion im Laufe des Prozesses die Änderung der Indolkonzentration überdeckt und eine direkte Beobachtung der Konzentrationsänderungen erschwert. Neben Tryptophan und Indol liegen in dem Serineluat noch weitere fluoreszierende Stoffe vor. Diese Fluoreszenz kann keinem bestimmten Inhaltsstoff zugeordnet werden. Sie wird jedoch durch die Zugabe von Indol, PLP und den Zellen, bzw. der Komplexbildung zwischen diesen verändert [PROTSCH 2003]. Die 2D-Fluoreszenzspektroskopie eignet sich zur Beobachtung der Biotransformation, da sowohl das Substrat Indol als auch des Produkt Tryptophan fluoreszieren.

### 3.1.2 Referenzmessung

An dem Biotransformationsreaktor ist ein Bypass angeschlossen, in dem das Reaktionsgemisch zirkulär gepumpt wird. Die am Bypass über die Cross-Flow-Filtration (Abbildung 3.3) angeschlossene Online-HPLC-Messung detektiert im Abstand von 15 Minuten die Konzentrationen von Indol und Tryptophan. Die Messergebnisse werden vom Bedienpersonal genutzt, um den Prozess zu regeln und das Substrat Indol nachzudosieren.

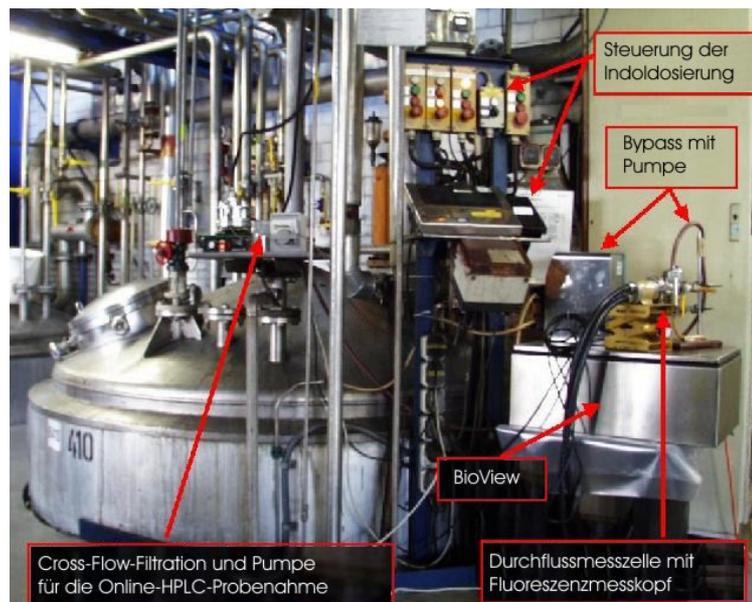


Abbildung 3.3: Reaktorkopf mit Indoldosierung und Fluoreszenzmessung

Neben der Probenahme für die Online-HPLC durch eine Cross-Flow-Filtration ist in Abbildung 3.3 noch die manuelle Indoldosierungseinheit zu erkennen, mit der die Zugabemenge eingestellt und die Zugabe gestartet werden kann. Von der AMINO GmbH werden die Serinkonzentration des Eluates vor der Umsetzung und die Indol-, Serin- und Tryptophankonzentration nach der Umsetzung mittels HPLC analysiert. Indol und Tryptophan werden über eine Indol-HPLC erfasst, Serin über eine Aminosäure-HPLC. Hierfür werden die Zellen von der Lösung abgetrennt und die Lösung durch Verdünnung für die Messung vorbereitet. Um genauere Informationen über die Vorgänge während des

Prozesses zu gewinnen wurden darüber hinaus Offlineproben direkt aus der Durchflusszelle für die Fluoreszenzmessung entnommen. Um die Indolabbaukinetiken zwischen den einzelnen Indolzugaben bestimmen zu können beträgt die Probenahmezykluszeit zu Beginn des Prozesses 5 Minuten. Nach 2 Stunden der Umsetzungen wurde die Probenahmezykluszeit sukzessiv auf 20 Minuten reduziert, da typischerweise nach dieser Zeit keine größeren Veränderungen mehr beobachtet werden können. Die Proben für die Indol-HPLC wurden mit Methanol versetzt, um die Reaktion sofort zu stoppen. Die Offlineproben für die Aminosäure-HPLC wurden sofort in flüssigem Stickstoff eingefroren. Nach dem Auftauen werden die Proben erst verdünnt und dann die Zellen von der Lösung abgetrennt. Die Proben wurden nach weiterer Aufarbeitung am Institut für Technische Chemie in Hannover analysiert [PROTSCH 2003].

### 3.2 Messaufbau am industriellen Produktionsprozess

In dem Bypass wurde zusätzlich eine Durchflusszelle (Werkstatt des Instituts für Technische Chemie, Hannover) zur 2D-Fluoreszenzmessung installiert und mit dem Flüssigkeitslichtleiter des BioView verbunden. Über einen weiteren Lichtleiter ist der BioView mit einem PC verbunden, der das Messgerät steuert und die Messungen aufzeichnet. Über die serielle Schnittstelle wurde ein Computer zur Auswertung der Messung angeschlossen. Basierend auf der Auswertung der Messung kann die Zugabe von Indol manuell dosiert werden. Eine Skizze des Aufbaus ist auf Abbildung 3.4 zu sehen.

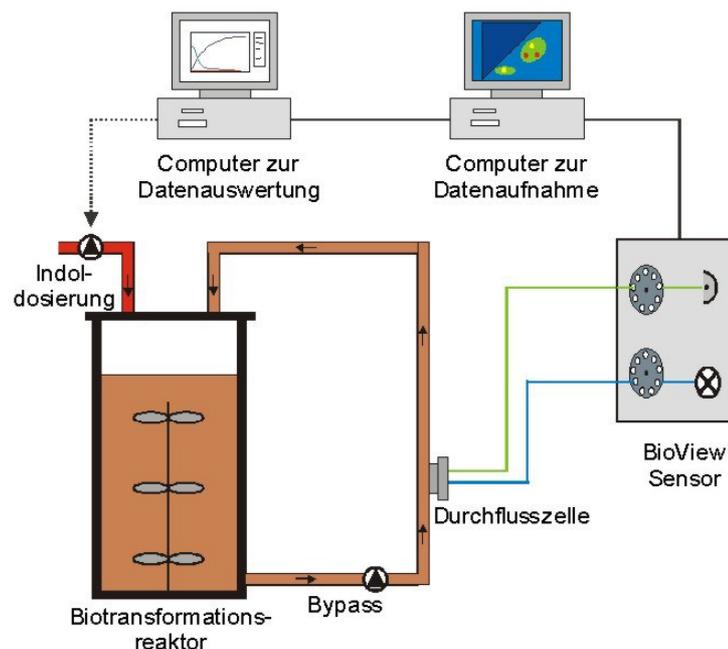


Abbildung 3.4: Aufbau der 2D-Fluoreszenzmessung am industriellen Tryptophanproduktionsreaktor

### 3.3 Messungen bei den Biotransformationen

#### 3.3.1 Die beobachteten Tryptophanproduktionsprozesse

Für die Systemanalyse des Biotransformationsprozesses stehen zehn Prozesse zur Verfügung, die mit Probenahme und Fluoreszenzmessungen begleitet wurden. Im November 2001 wurden sechs Prozesse (U1 – U6) mit Offlinemessungen und Fluoreszenzmessungen beobachtet. Zur Verifizierung der Modelle wurden im Mai 2002 vier weitere Produktionsprozesse (U7 – U10) mit Probenahme und Fluoreszenzmessung begleitet.

Die Prozesse U1 – U2, U4 – U6 und U10 zeigen Umsetzungen, die basierend auf der Online-HPLC-Messung vom Bedienpersonal geführt wurden. Der Prozess U3 wurde basierend auf Onlinevorhersagen aus den Fluoreszenzspektren und ersten Optimierungskriterien gefahren. Die Prozesse U7 – U9 wurden mit geringen Serinkonzentrationen durchgeführt, um den Serinverlust zu minimieren. Die Prozesse U7 und U8 wurden mit einem Teil des Serineluates gestartet und das restliche Serineluat erst während des Prozesses zugegeben wurde. Der Prozess U9 wurde mit Wasser angesetzt und dreifach konzentriertes eingedampftes Serineluat im Fed-Batch-Verfahren nachdosiert, um möglichst gleichbleibende geringe Serinkonzentrationen aufrecht zu erhalten.

#### 3.3.2 Onlinemesswerte

Über die Online-HPLC-Messung werden die Konzentrationen von Indol und Tryptophan während der Biotransformation in einem Abstand von 15 Minuten beobachtet. Da die Probenahme aufgrund der Probenahmestrecke mit einer Zeitverzögerung von etwa 5 Minuten stattfindet und die Auswertung der Probe 10 Minuten dauert, gibt die Online-HPLC den Zustand der Umsetzung mit etwa 15 Minuten Verzögerung an. Daher ist die derzeit am Prozess installierte Messmethode nur bedingt für eine optimale Prozessregelung geeignet. Der Vergleich der Online-HPLC-Messwerte mit den Offlinemesswerten in Abbildung 3.5 zeigt die Unterschiede der beide Messmethoden für die Konzentrationen der Prozessvariablen Indol und Tryptophan.

Bis zur ersten Stunde des Prozesses stimmt die Tryptophankonzentration der beiden Messungen gut überein, danach steigt die Tryptophankonzentration der Offlinemessung stärker an, als die der Onlinemessung, so dass nach 4 h Prozesszeit die Messwerte der beiden Messmethoden um 0,2 Gew.-% abweichen.

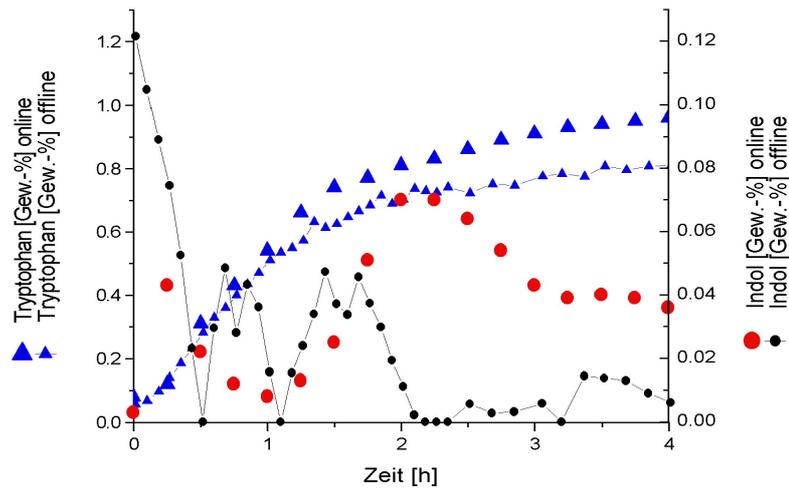


Abbildung 3.5: Vergleich der Offline- und Onlinemesswerte des Prozesses U1

Nach zwei Stunden liegt die online gemessene Indolkonzentration deutlich über der offline gemessenen Konzentration, wobei die Onlinemessung eine gewisse Trägheit bei der Detektion von schnellen Änderungen in der Konzentration aufweist. Die Messung ist mit einem relativ hohen Fehler für die absoluten Konzentrationen behaftet. Für die gefundene Tryptophankonzentration und unter Berücksichtigung der zugegebenen Menge Indol werden zu hohe Restindolkonzentrationen durch die Online-HPLC-Messung bestimmt. Somit ist entweder die Tryptophan- oder die Indolanalytik mit einem Fehler behaftet. Daher werden die Onlinemesswerte nicht für die genauere Analyse des Prozesses genutzt.

### 3.3.3 Offlinemesswerte

Von den Biotransformationen, die mit Offlineprobenahme begleitet wurden, liegen pro Prozess etwa 40 Messwerte der Prozessvariablen vor. Ein Prozess ist beispielhaft durch die Konzentrationsverläufe der Offlinemesswerte in Abbildung 3.6 dargestellt.

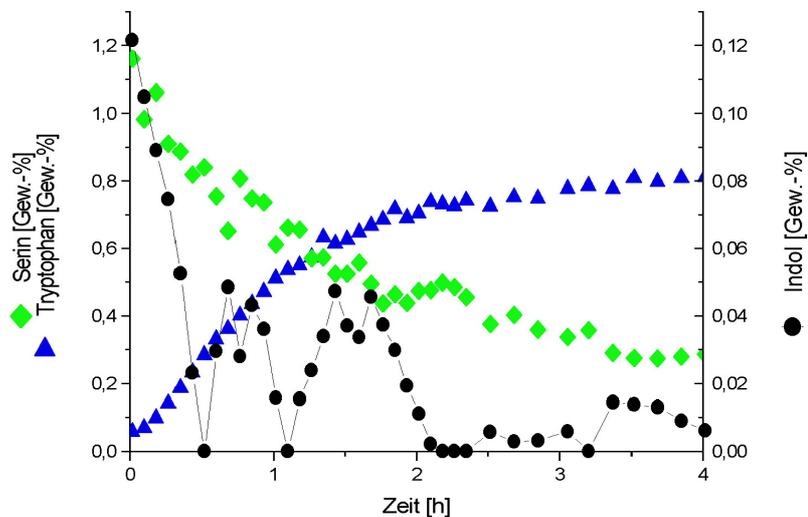


Abbildung 3.6: Verlauf der Offlinemesswerte von Serin, Tryptophan und Indol während des Prozesses U1

Deutlich zu erkennen ist die Abnahme des Substrates Serin und die Zunahme des Produktes Tryptophan. Um den Verlauf der Indolkonzentration besser abschätzen zu können, wurden die Messpunkte durch Linien verbunden, die jedoch nicht die Konzentrationen repräsentieren. Anhand der Indolmesswerte sind die Indolzugaben bei 0,52 h, 0,78 h, 1,13 h, 1,3 h, 1,35 h, 1,58 h, 2,57 h, 2,82 h, 3,07 h, 3,34 h und 3,59 h zu erkennen. Im Vergleich zu den Online-messwerten werden hier die schnellen Änderungen der Indolkonzentration detektiert und die Bilanz der Indol- und Tryptophankonzentration abhängig von den Indolzugaben stimmt.

### 3.3.4 Fluoreszenzmessungen

Für die zehn Prozesse, die mit Probenahme und Offlineanalytik beobachtet wurden, liegen ebenso Aufnahme von 2D-Fluoreszenzspektren vor. Für die Aufnahme eines Spektrums benötigt der BioView knapp eine Minute, so daß pro Prozess circa 500 Spektren für 8 h Umsetzungszeit zur Verfügung stehen. Pro Spektrum wurden 120 Messpunkte aufgezeichnet, dies entspricht einer 15\*15 Matrix von Exzitations- und Emissionswellenlängenkombinationen im Bereich 270 – 550 nm Exzitation und 310 – 590 nm Emission.

Für die Darstellung der Fluoreszenzverläufe in Abbildung 3.7 wurden signifikante Wellenlängenkombinationen ausgewählt. Dargestellt sind die Fluoreszenzverläufe im Bereich der Tryptophanfluoreszenz (Abbildung 3.2) bei 290 nm<sub>EX</sub> / 370 nm<sub>EM</sub> und aus dem Bereich der Fluoreszenz des Serineluates (Abbildung 3.2) bei 370 nm<sub>EX</sub> / 450 nm<sub>EM</sub>.

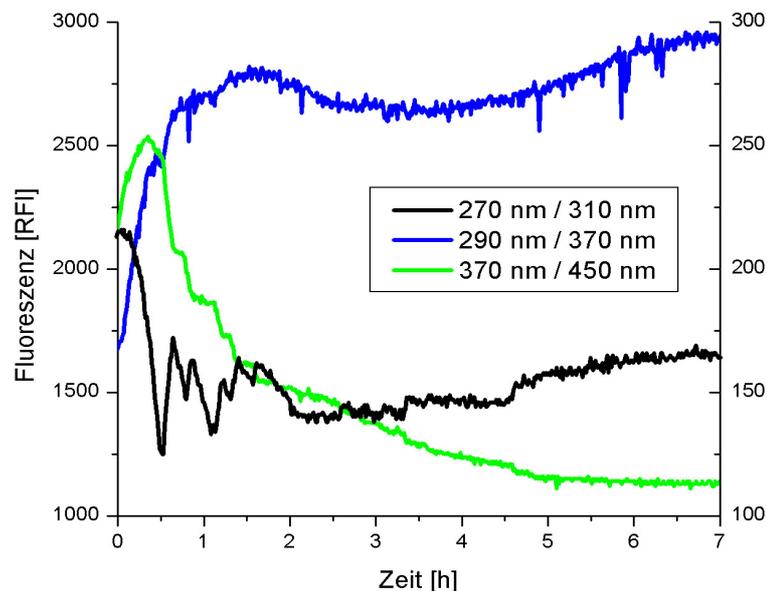


Abbildung 3.7: Verläufe ausgewählter Fluoreszenzwellenlängen während des Prozesses U1

Außerdem ist in Abbildung 3.7 der Intensitätsverlauf der Wellenlänge 270 nm<sub>EX</sub> / 310 nm<sub>EM</sub> dargestellt. Obwohl diese Wellenlänge nicht im Zentrum der Indolfluoreszenz liegt, zeigt sie den Verlauf der Indolkonzentration offenbar sehr gut. In der Abbildung ist der Anstieg der

Tryptophanfluoreszenz ( $290 \text{ nm}_{\text{Ex}} / 370 \text{ nm}_{\text{Em}}$ ) deutlich zu erkennen. Dieser ist so stark, dass die Indolfluoreszenz überdeckt wird und sich eine überlagerte Fluoreszenz für beide Inhaltsstoffe ergibt. Eine chemometrische Auswertung der Fluoreszenzspektren ist unerlässlich, da die Konzentrationen der einzelnen Prozessvariablen nicht direkt mit einer oder mehreren Wellenlängenkombinationen korreliert werden können [HITZMANN 1998A]. Die Abnahme der Fluoreszenz des Serineluates ( $370 \text{ nm}_{\text{Ex}} / 450 \text{ nm}_{\text{Em}}$ ) könnte nach weiterer Auswertung dazu geeignet sein die Konzentration von Serin vorherzusagen. Dies ist in sofern beachtenswert, da Serin selber nicht fluoresziert und eine Vorhersage von Serin aus der 2D-Fluoreszenz im Vorfeld nicht zu erwarten war.

## 4 Mathematische Modelle

In dieser Arbeit werden verschiedene mathematische Modelle zur Berechnung und Auswertung vorgestellt. Ein Bereich der mathematischen Modellierung erfolgt über die chemometrische Modellerstellung zur Vorhersage der Konzentrationen der Prozessvariablen aus den 2D-Fluoreszenzspektren (chemometrische Modelle). Zum anderen wird für die Simulation des Prozesses und die Optimierung ein mathematisches Prozessmodell genutzt, mit dem abhängig von den Umsetzungsbedingungen die Konzentrationen der Prozessvariablen berechnet werden können (Prozessmodelle).

### 4.1 Chemometrische Modelle

Die Chemometrik ist eine chemische Teildisziplin, die mathematische und statistische Methoden nutzt, um chemische Verfahren und Experimente optimal zu planen, durchzuführen und auszuwerten. So können aus den experimentellen Messdaten ein Maximum an chemisch relevanten, problembezogenen Informationen gewonnen werden. Ziele der Chemometrik sind Molekulares Design, Syntheseplanung, Optimierung und statistische Versuchsplanung in Bezug auf die Vorbereitung chemischer Experimente. Was die neue Qualität der Chemometrik gegenüber den herkömmlichen Anwendungen von Mathematik und Statistik in der Chemie ausmacht ist die Vieldimensionalität, hohe Variabilität und Korrelation der Daten, mit denen sie sich beschäftigt, die Komplexität der zu lösenden Probleme, und demzufolge der multidimensionale Charakter der mathematischen und statistischen Techniken [DANZER 2001]. Die von der Chemometrik genutzten Methoden sind in Tabelle 4.1 zusammengestellt.

Tabelle 4.1: Teilgebiete und Problemkreise der Chemometrik [DANZER 2001]

Stochastik	Datenanalyse	Modellierung	Systemtheorie
Statistik Fehlertheorie	Faktoren- und Hauptkomponentenanalyse	Korrelationsanalyse	Informationstheorie
Versuchsplanung	Clusteranalyse	Multivariate Regression	Signaltheorie und -verarbeitung
Optimierung	Mustererkennung	Softmodellierung (PLS)	Zuverlässigkeitstheorie
Prozesskontrolle	Klassifikation	Kalibration	Operations Research
Probenahmetheorie	Datenvorbehandlung		

In der Chemie ist man im Rahmen von Kalibrationsmethoden an den Zusammenhängen zwischen Variablen interessiert. Ziel der Regressionsanalysen ist es ein Modell für das Verhältnis zwischen den unabhängigen Variablen  $x$  (Messwerten) und den abhängigen Variablen  $y$  (Prozessvariablen) zu finden. Unter der Voraussetzung, dass unabhängige Messungen vorliegen

und die Messwerte normalverteilt sind, wird eine lineare Funktion nach der Methode der kleinsten Fehlerquadrate nach Gleichung (4.1) an die Messwerte angepasst.

$$y = b \cdot x + e \quad (4.1)$$

Wobei  $y$  die Konzentration der Prozessvariablen,  $x$  den Messwert,  $b$  den Regressionskoeffizient und  $e$  den Fehler repräsentiert. Werden mehrere Messgrößen genutzt um die Prozessvariable zu bestimmen, so erweitert sich die lineare Regression zur multilinearen Regression (MLR) in den Gleichungen (4.2) für  $m$  verschiedene Messgrößen.

$$y = \sum_{j=1}^m b_j \cdot x_j + e \quad (4.2)$$

Die Gleichung beschreibt die multilinearen Abhängigkeiten zwischen Messwert und Prozessvariable für nur eine Probe. Für die Erweiterung auf  $n$  Proben wird  $y_i$  mit  $i = 1 \dots n$  in Gleichung (4.3) als Spaltenvektor geschrieben,  $b$  bleibt ein Spaltenvektor und ändert seine Dimension nicht, ebenso wie  $y$  wird  $e$  zu einem Spaltenvektor mit  $n$  Komponenten und  $x_j$  wird von einem Zeilenvektor auf die Matrix  $X$  erweitert.

$$\vec{y} = X \cdot \vec{b} + \vec{e} \quad (4.3)$$

Die Zeilen der Datenmatrix  $X$  bilden Merkmalsvektoren der Dimension  $m$  einer Probe, die Spalten Messreihen an  $n$  Proben. Für die Modellierung des Zusammenhanges zwischen mehreren Zielgrößen  $y$  und mehreren Einflussgrößen  $x$  wird das Regressionsproblem multivariat und es liegen sowohl Matrizen von abhängigen Größen  $Y$  als auch von unabhängigen Größen  $X$  vor. Für eine Lösung nach der Methode der kleinsten Fehlerquadrate (Least-Square-Lösung) ergibt sich der Parametervektor  $b$  nach Gleichung (4.4).

$$\vec{b} = (X'X)^{-1} X' \vec{y} \quad (4.4)$$

In Gleichung (4.4) wird das Hauptproblem der MLR deutlich, denn die inverse Matrix zu  $X'X$  existiert zumeist nicht aufgrund einer Kollinearität zwischen  $x$ -Variablen. In realen Datensätzen sind nur in Ausnahmefällen alle Variablen voneinander unabhängig, meist korrelieren die Variablen mehr oder weniger stark untereinander. Die reale Dimension der Datenmatrix ist in diesem Fall geringer als die Anzahl der  $x$ -Variablen (Spalten).

Um dennoch einen Zusammenhang zwischen den unabhängigen und abhängigen Variablen berechnen zu können, wird durch eine Hauptkomponentenanalyse im Koordinatensystem der  $x$ -Variablen ein neues Koordinatensystem geringerer Dimension aufgespannt, welches die Daten mit geringem Informationsverlust beschreibt. Bevor die Hauptkomponentenanalyse zur Datenreduktion durchgeführt werden kann, werden die Daten vorverarbeitet, da sie nur selten in der Form vorliegen, die für die direkte Auswertung geeignet ist.

Als Preprocessing der Daten wird häufig die Zentrierung und die Standardisierung durchgeführt. Für die Zentrierung nach Gleichung (4.6) wird von jeder Variablen der Mittelwert nach Gleichung (4.5) subtrahiert, so dass die transformierten Variablen den Mittelwert Null besitzen. Die y-Variablen werden nur zentriert, wenn die Hauptkomponentenanalyse auch für diese Variablen durchgeführt werden soll, bzw. mehrere y-Variablen vorliegen und gleichzeitig aus den x-Variablen bestimmt werden sollen.

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij} \quad (4.5)$$

$$x_{ij}^{cent} = x_{ij} - \bar{x}_j \quad (4.6)$$

Wobei der Index j die x-Variablen, der Index i die Proben und n die Anzahl der Proben bedeuten. Eine Skalierung der Daten ist nicht immer sinnvoll und sollte je nach Form der Daten festgelegt werden. In vielen Fällen ist es günstig, die Variablen so zu skalieren, dass die resultierende Standardabweichung eins beträgt. Diese Transformation wird häufig auch als Autoskalierung bezeichnet [DANZER 2001]. Zur Autoskalierung der Daten wird die Varianz  $Dev(x_j)$  einer Messreihe nach Gleichung (4.7) berechnet und jeder Messwert dieser Messreihe durch diese Varianz dividiert, wie in Gleichung (4.8) beschrieben .

$$Dev(x_j) = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{n-1} \quad (4.7)$$

$$x_{ij}^{skal} = \frac{x_{ij}}{\sqrt{Dev(x_j)}} \quad (4.8)$$

Da alle standardisierten Variablen unabhängig von ihrem Informationsgehalt gleich gewichtet werden, kann diese Transformation jedoch auch unerwünschte Auswirkungen haben. Eine problemspezifische Wichtung einzelner Variablen oder Variablengruppen wird erreicht, indem die entsprechenden Variablen mit einem individuellen Wichtungsfaktoren multipliziert werden. Generell führt geeignetes Preprocessing der Variablen dazu, dass problemspezifisches Wissen im Datenmaterial Berücksichtigung findet und dadurch dessen Komplexität vermindert wird [DANZER 2001].

#### 4.1.1 Hauptkomponentenanalyse

Nach der Datenvorverarbeitung kann die Hauptkomponentenanalyse der Daten durchgeführt werden. Ziel der Hauptkomponentenanalyse (PCA, Principal Component Analysis) ist es, die Variablen der Datenmatrix durch eine geringere Zahl von untereinander unabhängiger Hauptkomponenten (PC) so darzustellen, dass die ursprüngliche Information der Variablen weitge-

hend erhalten bleibt. Voraussetzung hierfür ist, dass die Variablen zu einem gewissen Grad redundant sind, dass also Zusammenhänge zwischen den Variablen der Datenmatrix bestehen. Die Hauptkomponentenanalyse ist eine Methode, bei der die Matrix  $X$  des Ranges  $m$  als Summe von Matrizen  $M$  des Ranges 1 beschrieben wird, wie in Gleichung (4.9) dargestellt. Mit dem Rang einer Matrix wird die Anzahl der Spaltenvektoren bezeichnet, die linear unabhängig sind, und deren Anzahl die der linear unabhängigen Zeilenvektoren nicht überschreitet. Der Rang beziffert also die wirklich zu Grunde liegende Dimension einer Matrix.

$$X = M_1 + M_2 + M_3 + \dots + M_m \quad (4.9)$$

Diese Rang 1 Matrizen können auch als Produkt von zwei Vektoren, einem Scorevektor  $\vec{t}$  und einem Loadingvektor  $\vec{p}$  geschrieben werden. In Gleichung (4.11) wurden die Loadingvektoren  $\vec{p}_i'$  aus Gleichung (4.10) zeilenweise und die Scorevektoren  $\vec{t}_i$  spaltenweise zu Matrizen zusammengefasst.

$$X = \vec{t}_1 \vec{p}_1' + \vec{t}_2 \vec{p}_2' + \vec{t}_3 \vec{p}_3' + \dots + \vec{t}_m \vec{p}_m' \quad (4.10)$$

$$X = TP' \quad (4.11)$$

Eine ausführliche Beschreibung der Hauptkomponentenanalyse ist bei Vandeginste zu finden [VANDEGINSTE 1998]. Im Datenraum der Variablen, in dem die einzelnen Proben durch Punkte darstellbar sind, ist eine Hauptkomponente die Gerade, die die Daten mit der kleinsten Fehlersumme über alle Dimensionen wiedergibt. In Abbildung 4.1 ist dies für drei Variablen dargestellt, wobei unzentrierte Daten zur Darstellung genutzt wurden.

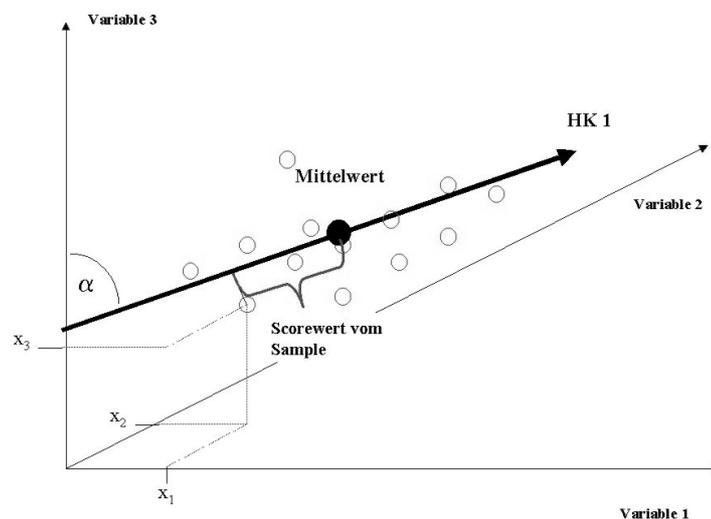


Abbildung 4.1: Eine Hauptkomponente im 3-dimensionalen Raum der Datenmatrix  $X$

Ein Loadingvektor enthält die Cosinuswerte der Winkel ( $\alpha$ ) zwischen der Hauptkomponente und den Achsen im ursprünglichen Raum. Ein Loadingvektor hat somit genau so viele Komponenten wie es  $x$ -Variablen gibt. Außerdem wird die Rotation des neuen Koordinatensystems im ursprünglichen Raum hierdurch festgelegt, so stellt ein Loading den Zusammen-

hang der Hauptkomponente zu den ursprünglichen Variablen her. Ein Scorevektor enthält die Abstände zwischen Koordinatenursprung, bzw. Hauptkomponentenmittelpunkt und den Projektionspunkten einer jeden Probe auf die Hauptkomponente. Ein Scorevektor hat somit genau so viele Komponenten wie Proben im Datensatz vorhanden sind und stellt die Hauptkomponente, bzw. die neue unabhängige Variable dar.

Die Hauptkomponenten werden sukzessiv berechnet. Nach Berechnung der ersten PC wird die Matrix  $\vec{t}_1 \vec{p}_1'$  von der ursprünglichen Datenmatrix subtrahiert, um eine neue Matrix zu erhalten, aus der die zweite PC berechnet wird. Die Berechnung der Hauptkomponenten erfolgt durch den NIPALS-Algorithmus [VANDEGINSTE 1998] (siehe Anhang 8.2.1). Der NIPALS-Algorithmus legt die Hauptkomponenten durch Iteration so fest, dass jeweils die größte noch in den Daten vorhandene Varianz mit der Hauptkomponente beschrieben wird und die Hauptkomponenten orthogonal zueinander stehen. Durch die Hauptkomponentenanalyse wird somit eine sukzessive varianzmaximierende orthogonale Rotationstransformation<sup>1</sup> der Originaldaten in das neue Koordinatensystem der Hauptkomponenten durchgeführt.

Da mit jeder Hauptkomponente die größte noch vorhandene Varianz der Daten beschrieben und damit die relevanten Informationen extrahiert werden, bleibt nach wenigen Iterationsschritten nur noch Informationen über das Messrauschen übrig. Zur Beurteilung der relevanten Anzahl von Hauptkomponenten und damit der Dimension oder Komplexität der Daten, können verschiedene Methoden benutzt werden. Ein Kriterium ist die Eigenwertanalyse der Hauptkomponenten. Ebenso kann durch die Beobachtung der beschriebenen Gesamtvarianz die Anzahl der zu benutzenden PCs festgelegt werden. Die Festlegung der zu benutzenden Anzahl von Hauptkomponenten sollte immer unter Kenntnis und Berücksichtigung der Originaldaten, deren Struktur und Komplexität getroffen werden.

Für die anschließende Hauptkomponentenregression (PCR, Principal Component Regression) stehen nun linear unabhängige x-Variablen, repräsentiert durch die Scorematrix, zur Verfügung. Durch die Reduktion der Datenmenge auf die relevanten Informationen kann nun eine multilineare Regression mit den Hauptkomponenten durchgeführt werden. Durch die Orthogonalität der Variablen in der Scorematrix ist die Least-Square Bestimmung des Parametervektors (Regressionskoeffizienten) nach Gleichung (4.4) ohne Schwierigkeiten möglich.

---

<sup>1</sup> Im Allgemeinen wird von der orthogonalen Rotationstransformation [Danzer 2001] gesprochen, wobei nicht die Transformation selbst orthogonal ist, sondern das Ergebnis der Transformation.

### 4.1.2 Regression nach der Methode der kleinsten Fehlerquadrate

Die Partial Least Square Regression (PLS) berücksichtigt im Gegensatz zu einem PCA-Modell, dass es Abhängigkeiten innerhalb der x-Variablen und zwischen den x-Variablen und y-Variablen gibt. Der Unterschied besteht darin, dass bei der PLS nicht nur die Varianz der Datenmatrix X sondern auch die Varianz der y-Variablen berücksichtigt wird [FRANK 1993]. Es gehen also sowohl die x-Variablen als auch die y-Variablen in die Berechnung mit ein. Hierfür wird die Datenmatrix nicht wie bei der PCA auf einen Scorevektor projiziert, sondern über einen Loadinggewichtsvektor mit Hilfe der Regression der kleinsten Fehlerquadrate auf den y-Vektor projiziert [MARTENS 1989] (siehe Anhang 8.2.2). Die Vorteile der PLS-Regression gegenüber einer PCR sind unter anderem, dass durch den Austausch der Informationen während der Kalkulation der Hauptkomponenten nach der PLS keine Regression mehr nötig ist, dass auch Korrelationen bei sehr komplexen Daten hergestellt werden können und dass die Kalibration unter Berücksichtigung der Korrelation zwischen den Datensätzen robuster ist, als wenn nur die Varianz der Datenmatrix X berücksichtigt wird [BROWN 1992].

Es gibt zwei Algorithmen zur Berechnung einer PLS-Regression, den PLS-1 und den PLS-2 Algorithmus. Bei der PLS-1 wird für jede y-Variable ein eigenes Modell berechnet, bei der PLS-2 hingegen wird ein Modell für alle y-Variablen gleichzeitig berechnet. Da bei der PLS-2 Regression analog zur PCR für alle y-Variablen die gleichen Scorevektoren benutzt werden, sind diese nicht für die Berechnung jeder y-Variable optimiert. Dies ist ein Nachteil besonders für komplexe Datensätze und bringt einen Verlust in der Vorhersagequalität mit sich. Die PLS-1 Regression liefert genauere Vorhersagen, da die Modelle für die Vorhersage jeder einzelnen Variable optimiert sind. Ein Nachteil der PLS-1 Methode ist jedoch der erheblich größere Zeitaufwand zur Berechnung der einzelnen Modelle. Der größte Vorteil der PLS-1 Regression zeigt sich bei der Analyse eines Datensystems mit sehr unterschiedlichen Verläufen der y-Variablen und komplexen Abhängigkeiten zwischen den x- und y-Variablen.

### 4.1.3 Multidimensionale Regression

Zweidimensionale Messmethoden wie die 2D-Fluoreszenzspektroskopie liefern für jede Messung eine Datenmatrix. Für mehrere Proben erhält man damit eine Hypermatrix mit 3 Indices. Wie in Abbildung 4.2 dargestellt, kann man sich diese als Datenquader vorstellen.

Bilineare Verfahren wie PCR und PLS können auch auf multidimensionale Daten angewendet werden, hierfür müssen die multidimensionalen Daten "entfaltet" werden. Zur Entfaltung eines 2D-Spektrums werden die einzelnen Intensitäten einer Messung einfach hintereinander aufgelistet. Bei den bilinearen Verfahren werden diese Messwerte dann als linear unabhängige Daten angenommen. Daher sind die Auswertemethoden selbst multivariat, aber nicht multidimensional. Sie ziehen keinen Vorteil aus der Multidimensionalität der Daten. Durch

die Entfaltung der multidimensionalen Daten werden die Modelle sehr komplex, da die Anzahl an Parameter enorm hoch ist; sie sind schwierig zu interpretieren, da die ursprüngliche Struktur der Daten und die Information der Multidimensionalität verloren geht.

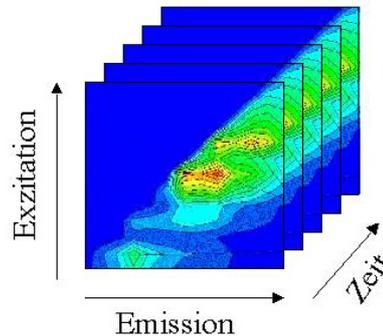


Abbildung 4.2: Darstellung einer Hypermatrix für eine zweidimensionale Messmethode

Es gibt verschiedene multidimensionale Regressionsverfahren, wie PARAFAC oder Tucker3 Modelle, die analog zur PCR die Regression nach Reduktion der Daten durch Projektion auf die Scores durchführen; in Analogie zur PLS gibt es die N-way PLS als multidimensionale Regression. Die N-way-PLS Modelle (im Folgenden auch N-way-Modelle) werden zwar nicht über die Regression der kleinsten Fehlerquadrate bestimmt, aber in Anlehnung an den Grundgedanken der PLS wird hier die Kovarianz zwischen den abhängigen und unabhängigen Variablen betrachtet und somit die Korrelation zwischen diesen Datensätzen berücksichtigt. Weiterführende Angaben zum den Algorithmen der unterschiedlichen multidimensionalen Regressionsmethoden sind bei Bro [BRO 1998] zu finden.

#### 4.1.4 Vorhersage

Als Vorhersage wird die Anwendung der Modelle auf entsprechende x-Variablen zur Schätzung der y-Variablen bezeichnet. Zur Berechnung der Vorhersage werden die Daten ebenso vorverarbeitet, wie die Daten zur Modellerstellung. Dies bedeutet, dass von der Datenmatrix  $X$  der Mittelwert subtrahiert wird, der bei der Modellerstellung benutzt wurde. Bei einem PCR Modell wird nach dem Preprocessing die Datenmatrix  $X$  mit den Loadings des Modells multipliziert, die als hinreichend für die Beschreibung der Daten identifiziert wurden. Die Berechnung der Scorematrix  $T$  für die Vorhersage aus der Datenmatrix  $X$  mit der Loadingmatrix  $P$  ist durch Gleichung (4.12) gegeben.

$$T = X \cdot P \quad (4.12)$$

Sowohl die Loadingmatrix  $P$  als auch die Scorematrix  $T$  enthalten nicht alle Hauptkomponenten, sondern nur so viele, wie bei der Modellerstellung als hinreichend festgelegt wurden.

Die so erhaltenen Scores  $T$  werden mit den Regressionskoeffizienten  $b$  multipliziert. Die Multiplikation nach Gleichung (4.13) ergibt dann die Vorhersage für die Prozessvariable  $\hat{y}$ .

$$\vec{\hat{y}} = T \cdot \vec{b} \quad (4.13)$$

Für die Vorhersage der  $y$ -Variablen mit einem PLS Modell ist die direkte Berechnung der Vorhersage über die Regressionskoeffizienten, die nach der Modellerstellung zugänglich sind (siehe Anhang 8.2.2) möglich. Hierbei wird die Vorhersage mit den Regressionskoeffizienten  $\vec{b}$  und dem Achsenabschnitt  $b_0$  nach Gleichung (4.14) berechnet.

$$\vec{y} = b_0 + X' \vec{b} \quad (4.14)$$

Die Vorhersage der  $y$ -Variablen über ein multidimensionales Regressionsmodell erfolgt entsprechend der Vorhersage über ein PLS Modell.

#### 4.1.5 Modellvalidierung

An die Ergebnisse der Modellierung sind zwei Anforderungen zu stellen: Erstens müssen Modelle anhand der untersuchten Daten gültig und nachvollziehbar sein (interne Validität); zweitens sollen die erhaltenen Ergebnisse auch auf zukünftige Messungen übertragbar und verallgemeinerbar sein (externe Validität). Die externe Validität kann nur getestet werden, indem das erhaltene Modell mit neuen, unbekanntem Daten überprüft wird. Da die Daten in der Regel nur eine mehr oder weniger repräsentative Stichprobe der Grundgesamtheit darstellen, hat eine verbesserte interne Validität nicht automatisch eine höhere externe Validität zur Folge. Um eine Übermodellierung zu vermeiden, sollte ein Modell mit möglichst geringer Zahl von Parametern auskommen, also eine dem Problem angepasste Komplexität aufweisen. Zur Schätzung der internen Validität werden z.B. Testdatensätze oder die Kreuzvalidierung herangezogen. Für die Bestimmung der externen Validität der Modelle können aus den Vorhersagen und Referenzwerten verschiedene Gütemaße berechnet werden, z.B. die Summe der Fehlerquadrate zwischen den vorhergesagten und beobachteten Werten PRESS (Predictive Residual Sum of Squares) in Gleichung (4.15) bzw. RMSEP (Root Mean Square Error of Prediction) in Gleichung (4.16).

$$PRESS = \sum_{i=1}^n (\hat{c}_i - c_i)^2 \quad (4.15)$$

$$RMSEP = \sqrt{\frac{PRESS}{n}} \quad (4.16)$$

Hierbei sind  $c_i$  die Referenzkonzentrationen und  $\hat{c}_i$  die vorhergesagten Konzentrationen für  $n$  verschiedene Proben. Die Angabe eines PRESS hat die quadratische Dimension der Messung, die eines RMSEP die Dimension der Messung der Prozessvariable. Eine Vergleichbarkeit der

Vorhersagefehler für verschiedene Prozessvariablen ist wünschenswert und wird durch Bildung des Quotienten zwischen der Fehlergröße und der maximal vorkommenden Konzentration erreicht. Dieser relative Fehler ist folglich in Prozent anzugeben.

#### 4.1.6 Sensitivitätsanalyse

Die Sensitivitätsanalyse identifiziert die Spektrenbereiche, die für die Vorhersage einer Prozessvariable hohe Relevanz besitzen. Dies kann dadurch geschehen, dass für eine Probe die x-Variablen nacheinander entweder um einen absoluten oder einen relativen Betrag variiert werden. Der Abstand der variierten Vorhersage und der unveränderten Vorhersage ergibt das Maß der Relevanz dieser Wellenlänge für die Vorhersage. Zur besseren Vergleichbarkeit werden die Abstände in Relation zum maximalen Abstand gesetzt und es ergeben sich Sensitivitätsspektren mit Werten zwischen null, keine Sensitivität und eins, maximale Sensitivität. Bei PLS Modellen kann die Sensitivität der x-Variablen  $Sens_j$  direkt durch die Regressionskoeffizienten  $\vec{b}$  ermittelt werden, die zur Erzeugung eines Sensitivitätsspektrums nur noch in Relation zueinander gestellt werden müssen, dies geschieht entsprechend Gleichung (4.17).

$$Sens_j = \frac{|\vec{b}_j|}{\max \vec{b}} \quad (4.17)$$

Die Ergebnisse der Sensitivitätsanalyse werden als Sensitivitätsspektrum dargestellt, da hierdurch ein anschaulicher Vergleich mit den Messdaten ermöglicht wird. Anhand der Sensitivitätsspektren werden die relevanten Spektrenbereiche identifiziert und können eventuell auf bestimmte Komponenten wie Proteine, Flavine oder einzelne Fluorophore zurückgeführt werden. Hierdurch lässt sich die biologische Bedeutung der Modelle interpretieren. Darüber hinaus ist es möglich aufgrund der Sensitivitätsspektren die Messungen auf relevanten Wellenlängenkombinationen zu beschränken und hierdurch die Zeit für die Aufnahme eines Spektrums deutlich zu verringern. Hierfür wird ein Relevanzniveau von z.B. 40 % festgelegt und nur die Wellenlängenkombinationen zur Modellbildung berücksichtigt, die eine höhere Sensitivität aufweisen. Die Vorhersagen aus einem reduzierten Spektrum sollten nicht deutlich von den Vorhersagen aus einem vollständigen Spektrum abweichen. Die etwas geringere Vorhersagegüte kann durch Filterung der Vorhersagen, die jetzt in deutlich größerer Zahl vorliegen, ausgeglichen werden. Ebenso ist auch eine Verbesserung der Vorhersage möglich, da eventuell vorhandenes Messrauschen nicht mehr für die Vorhersage berücksichtigt wird.

#### 4.1.7 Modellerstellung und -anwendung

Für die Erstellung von chemometrischen Modellen wurde der UNSCRAMBLER 7.6 SR-1 (Camo, Norwegen) benutzt, da mit diesem Softwaretool eine schnelle und vielfältige Berechnung der Modelle möglich ist. Darüber hinaus bieten die Darstellungen der Ergebnisse die notwendige Variabilität zur Interpretation und Beurteilung der Modellgüte, und der Umgang mit den enormen Datenmengen ist übersichtlich und beherrschbar.

Zur Analyse der Struktur und Komplexität der Messdaten wird eine Hauptkomponentenanalyse (PCA) durchgeführt. Die Ergebnisse der PCA geben Aufschluss über die Variationen innerhalb eines Prozesses und zwischen den Prozessen. Basierend hierauf kann durch Bestimmung der Distanzen zwischen den einzelnen Prozessen gegebenenfalls eine Klassifizierung der Prozesse durchgeführt werden. Eine Klassifizierung der Prozesse sollte durchgeführt werden, wenn die Unterschiede in der Fluoreszenz zwischen den einzelnen Prozessen größer sind als die Unterschiede innerhalb der einzelnen Prozesse.

Die Bestimmung der Anzahl der Hauptkomponenten erfolgt über die Analyse der beschriebenen Gesamtvarianz und des minimalen RMSEP der Validierung. Die interne Validität der Modelle wird durch die Kreuzvalidation [OTTO 1997] bestimmt. Die externe Validität erfolgt über die Berechnung des RMSEP.

Aufgrund der hohen Komplexität der Datenmatrix und den sehr unterschiedlichen Verläufen der Prozessvariablen werden für Vorhersagen der Prozessvariablen PLS1 und N-way-Modelle berechnet. Für die Erstellung von N-way-Modellen wurde MATLAB 6.5.0.180913a (The MathWorks, Inc., Natick, USA) und die N-way-Toolbox von Rasmus Bro (N-way Toolbox 1.04 for MATLAB; Anderssen and Bro, 2000; <http://www.models.kvl.dk/source>) benutzt.

Zur Onlineanwendung der Modelle werden die aus den Modellen benötigten Parameter, wie Mittelwerte und Regressionskoeffizienten, exportiert und in ein selbst erstelltes Vorhersageprogramm (programmiert mit BORLAND DELPHI 5, Inprise Corp., USA) eingelesen.

Für eine weitere Methode der Modellerstellung werden nicht die Offlinemesswerte als Y-Variablen herangezogen sondern die durch Simulationen gegebenen Konzentrationsverläufe. Hierfür müssen die kinetischen Modellparameter bekannt sein. Doch je nach Messmethode und Prozess können diese eventuell auch direkt aus den Onlinemesswerten identifiziert werden. Durch diese simultane Parameteridentifikation und chemometrische Modellerstellung, wie sie in der Literatur beschrieben ist ([GEISLER 2002], [SOLLE 2003]), werden die zum Teil sehr aufwändigen Offlinemessungen nur noch geringfügig benötigt.

## 4.2 Bioprozessmodell

Die universelle mathematische Methode für die Prozessmodellierung sind Differenzialgleichungssysteme. Die komplizierten Wechselwirkungen zwischen der chemischen Reaktion und den gleichzeitig ablaufenden Transportvorgängen lassen sich hierdurch mathematisch erfassen. Dazu wird ein sogenannter Bilanzraum definiert, in dem die zeitliche Änderung bestimmter Zustandsgrößen beschrieben wird. Die Änderungen beruhen auf der Differenz zwischen den in den Bilanzraum eintretenden und austretenden Strömen und der Erzeugung oder dem Verbrauch im Inneren des betrachteten Raumes. Die allgemeine Stoffbilanz der Komponente  $i$  ist durch Gleichung (4.18) gegeben. Aufgrund der isothermen Reaktionsführung kann eine Berücksichtigung der Wärmebilanz vernachlässigt werden.

$$\frac{\partial c_i}{\partial t} = -\text{div}(uc_i) + \text{div}(D_i \text{grad}(c_i)) + \sum_j v_{ij} r_j \quad (4.18)$$

Mit  $\frac{\partial c_i}{\partial t}$  der Akkumulation,  $-\text{div}(uc_i)$  der erzwungenen Konvektion,  $\text{div}(D_i \text{grad}(c_i))$  der effektiven Diffusion (Dispersion) und  $\sum_j v_{ij} r_j$  der Reaktion. Für Rührkesselreaktoren wird ideale Durchmischung angenommen, so dass innerhalb des Reaktors keine Konzentrationsgradienten auftreten. Ein Zustrom wird ebenfalls sofort vollständig vermischt und im Abfluss liegen die Konzentrationen wie im Reaktor vor. Dadurch werden die Konzentrationen der Zustandsgrößen  $c_i$  durch die allgemeine Gleichung (4.19) bei konstantem Volumen beschrieben.

$$V \frac{dc_i}{dt} = \dot{V}_0 c_{i0} - \dot{V}_a c_i + V \sum_j v_{ij} r_j \quad (4.19)$$

Die Zu- und Abflüsse  $\dot{V}_0$  und  $\dot{V}_a$  werden durch die Reaktorauslegung, bzw. die Reaktionsführung bestimmt, ebenso die Konzentration der Komponente  $i$  im Zufluss  $c_{i0}$ . Der Reaktionsterm  $V \sum_j v_{ij} r_j$  ist durch die Kinetik der ablaufenden Reaktionen gegeben.

Die Tryptophanproduktion wird im Fed-Batch-Verfahren durchgeführt, daher ist der Abfluss  $\dot{V}_a=0$ . Durch das Kondensat und die Zellzugabe wird das Volumen während der Reaktion vergrößert, so dass neben den zu betrachtenden drei Konzentration für Indol, Serin und Tryptophan auch die zeitliche Änderung des Volumens beschrieben werden muss. Es ergibt sich für das Prozessmodell ein Differenzialgleichungssystem mit vier Gleichungen, das die in Gleichung (4.20) – (4.23) beschriebene Grundstruktur aufweist.

$$\frac{dI}{dt} = \frac{\dot{m}_I}{V} - \frac{\dot{V}_{BFM} + \dot{V}_{Kond}}{V} I + v_i \quad (4.20)$$

$$\frac{dS}{dt} = -\frac{\dot{V}_{BFM} + \dot{V}_{Kond}}{V} S + v_S \quad (4.21)$$

$$\frac{dT}{dt} = -\frac{\dot{V}_{BFM} + \dot{V}_{Kond}}{V} T + v_T \quad (4.22)$$

$$\frac{dV}{dt} = \dot{V}_{BFM} + \dot{V}_{kond} \quad (4.23)$$

Mit I der Indolkonzentration, S der Serinkonzentration und T der Tryptophankonzentration,  $\dot{V}_{BFM}$  und  $\dot{V}_{kond}$  dem Volumenstrom der Biofeuchtmasse und des Kondensats und  $\dot{m}_I$  dem Massenstrom von Indol während der Indolzugabe.

#### 4.2.1 Reaktionskinetik

Für eine enzymatische Reaktion ist die Michaelis-Menten-Kinetik, entsprechend Gleichung (4.24), die Grundlage.

$$v = \mu_{\max} \frac{S}{K_m + S} \quad (4.24)$$

Mit  $v$  der Reaktionsgeschwindigkeit, der maximalen Reaktionsgeschwindigkeit  $\mu_{\max}$ , der Substratkonzentration  $S$  und der Michaelis-Menten-Konstanten  $K_m$ . Die Michaelis-Menten-Kinetik bewirkt, dass für sehr kleine Substratkonzentrationen eine Reaktion 1. Ordnung vorliegt, bei Substratkonzentrationen deutlich über der Michaelis-Menten-Konstante  $S \gg K_m$  liegt die Reaktionsgeschwindigkeit konstant bei  $\mu_{\max}$ .

Die Enzymkinetik der Reaktion von L-Serin und Indol zu L-Tryptophan wird bei PLP-Sättigung zu einer Zweisubstrat-Reaktion. Heilmann [HEILMANN 1978] konnte keine Entscheidung zwischen einem random- oder einem ordered-Mechanismus treffen. Basierend hierauf wurde von Vorlop [VORLOP 1983] die Reaktionskinetik in Gleichung (4.25) für diese Reaktion angegeben.

$$v = \mu_{\max} \frac{S \cdot I}{S(I + K_m^{I1}) + K_m^S \left( I + K_m^{I2} + \frac{I^2}{K_i} \right)} \quad (4.25)$$

Hierbei ergibt sich die Reaktionsgeschwindigkeit  $v$  aus der maximalen Reaktionsgeschwindigkeit  $\mu_{\max}$ , den Konzentrationen der Substrate Serin  $S$  und Indol  $I$ , der Inhibierungskonstante  $K_i$  und den Michaelis-Menten Konstanten  $K_m^{I1}$  und  $K_m^{I2}$  für Indol und  $K_m^S$  für Serin. Vorlop geht von einem Random-Mechanismus aus und erweitert die Kinetik um eine kompetitive Indolhemmung. Obwohl es sich nicht um eine kompetitive Hemmung handelt,

erzielt Vorlop mit dieser mathematischen Beschreibung der Reaktion gute Ergebnisse für immobilisierte Zellen. In seinen Versuchen mit freien Zellen gelangte er zu einer abweichenden Hemm-kinetik, für die er jedoch kein Modell angibt [VORLOP 1983].

Lane [LANE 1983A] stellt eine sequenzielle Bindung von L-Serin an das Enzym fest, die von der Indolbindung gefolgt wird. Es kann also von einem Ordered-Mechanismus für die Kinetik der Tryptophanbildung an Tryptophansynthase ausgegangen werden. Da Indol und Serin sich gegenseitig in der Bindung an das Enzym nicht beeinflussen, ist die Geschwindigkeitsgleichung für einen Ordered-Mechanismus in Gleichung (4.26) gegeben [BISSWANGER 2000].

$$v = \mu_{\max} \frac{S_1}{K_m^1 + S_1} \cdot \frac{S_2}{K_m^2 + S_2} \quad (4.26)$$

So dass die Reaktionsterme  $v$  dieses Prozessmodells folgendermaßen aussehen:

$$v_I = -\mu_{\max} \frac{S}{K_m^S + S} \cdot \frac{I}{K_m^I + I} \quad (4.27)$$

$$v_S = -\mu_{\max} \frac{S}{K_m^S + S} \cdot \frac{I}{K_m^I + I} \quad (4.28)$$

$$v_T = \mu_{\max} \frac{S}{K_m^S + S} \cdot \frac{I}{K_m^I + I} \quad (4.29)$$

Aufgrund der Struktur der Reaktionsterme, sind die Differenzialgleichungssysteme nicht analytisch lösbar. Eine numerische Lösung des Gleichungssystems (4.27) – (4.29) ist durch die Runge-Kutta-Iteration möglich. Hierbei werden die Zustandsgrößen für möglichst kleine Zeitintervalle sukzessiv berechnet. Die so erhaltenen Trajektorien stellen die Simulation des Prozesses über die Zeit dar.

#### 4.2.2 Parameteridentifikation

Die numerische Integration der Differenzialgleichungen ist jedoch nur möglich, sofern die Startbedingungen und alle Konstanten bekannt sind. Bei komplexen biologischen Reaktionen sind insbesondere die kinetischen Konstanten bzw. Parameter nicht immer bekannt, oder sie variieren in Abhängigkeit von den Umsetzungsbedingungen. Somit ist die Bestimmung der kinetischen Parameter für eine zutreffende Simulation eines biotechnischen Prozesses unerlässlich.

Zur Parameteridentifikation wird die Simulation für eine bestimmte Parameterkombination mit den Messwerten des Prozesses verglichen und der quadratische Fehler berechnet. Für die optimale Parameterkombination ist dieser Fehler minimal. Die zu optimierende Zielgröße (Fehlerfunktion) ist daher von mehreren Einflussgrößen (Parametern) abhängig. Durch die

Wechselwirkung zwischen den Einflussgrößen führt das Variieren eines Parameters bei konstant halten der anderen Parameter meist nicht zum Optimum.

Man unterscheidet lokale und globale Optima. Im Gegensatz zur univariaten Optimierung kann bei multivariaten Optimierungsaufgaben nur selten garantiert werden, dass wirklich das globale Optimum gefunden wurde. In der Regel finden Optimierungsmethoden nur das nächste Optimum, welches nur bei geeigneten Startwerten das globale Optimum ist. Der Erfolg der verschiedenen Optimierungsmethoden hängt damit außer vom Datenmaterial auch vom Vorwissen des Anwenders über den Bereich des Optimums ab. Eine optimale Optimierungsmethode, die sich für alle Problemstellungen eignet, existiert nicht [DANZER 2001].

Bei der sequenziellen Simplex-Optimierung wird die Wirkungsfläche lokal durch einen Simplex beschrieben. Ein Simplex ist eine geometrische Figur, die für  $m$  Parameter  $m+1$  Eckpunkte besitzt. Zunächst wird an einem Punkt des Optimierungsraumes ein Startsimplex ermittelt. Im nächsten Schritt wird bei der einfachsten Version der Simplex-Optimierung ein neuer Simplex erhalten, indem der Punkt mit der schlechtesten Lösung am Schwerpunkt der anderen Punkte gespiegelt wird. Nach einigen Optimierungsschritten wird ein Optimum erreicht, welches sich durch ein Oszillieren des Simplexes bemerkbar macht.

Die richtige Wahl der Schrittweite spielt bei dieser einfachen Variante der Simplexoptimierung eine große Rolle, da bei zu großer Schrittweite das Optimum entweder ungenau oder gar nicht gefunden wird. Bei kleiner Schrittweite werden dagegen sehr viele Iterationsschritte benötigt und vermehrt lokale Minima gefunden. Eine erweiterte Version des Algorithmus (Nelder-Mead-Suche) verwendet aus diesem Grund eine variable Größe der Schrittweite. Die Optimierung sollte möglichst mehrfach mit verschiedenen Startwerten durchgeführt werden, um sicherzustellen, dass nicht nur ein lokales Optimum gefunden wurde (nähere Angaben zum Simplex-Verfahren sind bei Danzer [DANZER 2001] und Otto [OTTO 1997] beschrieben).

Genetische Algorithmen gehören zur Klasse der globalen Optimierungsalgorithmen. Konventionelle diskrete Optimierungsmethoden versagen häufig bei komplizierten Zielfunktionen mit einer großen Zahl von lokalen Optima und Aufgaben mit vielen Parametern. Genetische Algorithmen finden jedoch in diesen hochdimensionalen Suchräumen oft gut Näherungslösungen.

Die Genetischen Algorithmen wurden in Analogie zur Evolution in der Natur entwickelt. Die Anpassung von Lebewesen an die Umwelt kann als Optimierungsprozess aufgefasst werden, der durch Rekombination und Mutation des Erbgutes und anschließender Selektion der Individuen nach dem Darwinschen Prinzip des "survival of the fittest" gesteuert wird.

Zu Beginn wird eine Population von möglichen Lösungen, den Individuen, zufällig erzeugt, wobei ein Individuum aus einem Vektor besteht, der Werte für die anzupassenden Parameter enthält. Die Individuen werden durch eine problemspezifische Fitnessfunktion bewertet und

erhalten Reproduktionschancen proportional zur ermittelten Güte der Lösung. Durch den Prozess aus Selektion, Rekombination und Übernahme der besten Individuen in die nächste Generation werden unter bestimmten Voraussetzungen nach einer hinreichenden Anzahl von Generationen optimale oder annähernd optimale Lösungen erhalten [MASSART 1997].

Für das hier aufgestellt Optimierungsproblem der Parameteranpassung wurde eine Kombination verschiedener Optimierungsalgorithmen gewählt. Zuerst wird durch einen genetischen Algorithmus der Bereich des globalen Optimums ermittelt und daran anschließend durch ein Simplex-Verfahren das Optimum näher bestimmt. Der Vorteil dieser Reihenfolge liegt darin, dass die Startparameter für den Simplex durch das Ergebnis des genetischen Algorithmus gegeben sind und dass der genetische Algorithmus zuvor im Raum der biologisch sinnvollen Parametergrenzen nach einem Optimum gesucht hat. Um die zuvor gewählten Grenzen für die Parameter nicht durch die Anpassung mit dem Simplexverfahren zu verlassen, wurde bei den hier beschriebenen Anpassungen der Nelder-Mead-Algorithmus erweitert. Die Erweiterung erfolgte durch die Berücksichtigung der festgelegten Parametergrenzen bei jeder Veränderung der Parameter durch die Subroutinen innerhalb des Algorithmus.

Zur Parameteranpassung werden die Prozessbedingungen, wie Startkonzentrationen, Biomassezugaben, Feed-Profil für Indol und Startvolumen für die Simulation initialisiert. Das Differenzialgleichungssystem wurde für die Simulationen in SIMULINK 4.0 (The MathWorks, Inc., Natick, USA) eingegeben. Für die Fehlerberechnung der Simulation werden die Offlinemesswerte der Prozessvariablen Indol, Serin und Tryptophan eingelesen. Als Optimierungskriterium werden die Summen der quadratischen Abweichungen zwischen Simulation und Messwert der einzelnen Prozessvariablen berechnet. Die Fehlerquadratsummen der Prozessvariablen werden addiert, wobei die Fehlersumme für Indol mit dem Faktor 10 gegenüber den anderen gewichtet wird, um die unterschiedliche Größenordnung der Prozessvariablen auszugleichen und eine möglichst zutreffende Simulation für die Indolkonzentration zu erhalten. Die Fehlerfunktion ist durch Gleichung (4.30) gegeben.

$$E = 10 \cdot \sum_{i=1}^n \left( I_i^{sim} - I_i^{mess} \right)^2 + \sum_{i=1}^n \left( S_i^{sim} - S_i^{mess} \right)^2 + \sum_{i=1}^n \left( T_i^{sim} - T_i^{mess} \right)^2 \quad (4.30)$$

### 4.2.3 Sensitivitätsanalyse

Nach der Parameteridentifikation sollten die Parameter auf Sensitivität überprüft werden. Die Sensitivität der Parameter für die Optimierungsaufgabe ist von den Messwerten abhängig, da ein Parameter, der sich für kleine Konzentrationen im Modell auswirkt, nicht durch Messwerte im hohen Konzentrationsbereich bemerkbar machen kann. Die Parametersensitivität kann durch verschiedene Methoden bestimmt werden. Die Darstellung der univariaten Fehler-

verläufe liefert einen ersten Hinweis. Durch Variation eines Parameters bei konstant halten der anderen Parameter kann der Einfluss dieses Parameters auf den Fehler nach Gleichung (4.30) dargestellt werden. Die unterschiedlichen Steigungen der Fehlerfunktionen in der Nähe des Optimums und die Breite der Optima zeigen deutlich, wie stark sich die Veränderungen eines Parameters auf den Fehler auswirken.

Die numerische Bestimmung der Parametersensitivität nach Hearne [Hearne 1985] und die Berechnung der Fisherinformationsmatrix wie sie bei Munack [Munack 1989] und Schneider ([SCHNEIDER 1999], [Schneider 1995]) beschrieben wird, gibt genauere Aufschlüsse über die Anpassbarkeit der Modellparameter mit den vorliegenden Messwerten. Die Bestimmung der Parametersensitivität nach Hearne [HEARNE 1985] erfolgt über den Vergleich der Nominaltrajektorie bei der optimalen Parameterkombination (Gleichung (4.31)) mit den Trajektorien die bei der Variation der einzelnen Parameter entstehen (Gleichung (4.32)).

$$\frac{dx_i}{dt} = f_i(x, p) \quad (4.31)$$

$$x_i(p + \Delta p, t) = x_i(p, t) + \Delta x_i(p, \Delta p, t) \quad (4.32)$$

Mit  $x_i$  der Trajektorie bei der optimalen Parameterkombination  $p$ , bzw. der variierten Parameterkombination  $p + \Delta p$  und der Differenz der Trajektorien  $\Delta x_i$ . Die relative prozentuale Differenz zwischen jeder veränderten Trajektorie und der Nominaltrajektorie wird für jede Parametervariation nach Gleichung (4.33) berechnet.

$$N_{ij}(t_i) = 100 \frac{\Delta x_i(t_i)}{x_i(t_i)} \quad (4.33)$$

Das Maß für die Differenz zwischen der Nominaltrajektorie und der veränderten Trajektorie kann definiert werden als:

$$\int_{t_0}^{t_n} \left( N_{ij}(t_i)^T N_{ij}(t_i) \right) dt \quad (4.34)$$

Die Integration der zeitabhängigen Matrix  $M$  als Resultat aus der Multiplikation von  $N$  mit ihrer Transponierten, erfolgt über die Simpson Integrationsformel nach Gleichung (4.35).

$$\int_{t_0}^{t_n} M(t) dt \approx \frac{\Delta t}{3} \left[ M_1 + M_n + 2 \sum_{k=3}^{n-2} M_k + 4 \sum_{k=2}^{n-1} M_k \right] = A \quad (4.35)$$

Der zum höchsten Eigenwert korrespondierende Eigenvektor der resultierenden zeitunabhängigen Matrix  $A$  legt die Ordnung der Parameter in Abhängigkeit ihrer Sensitivität fest. Der höchste Wert des Eigenvektors zeigt den sensitivsten Parameter. Werden die Werte des

Eigenvektors in Relation zum höchsten Wert gestellt, so erhält man die prozentuale Sensitivität jedes einzelnen Parameters für die Simulation.

Bei der Bestimmung der Parametersensitivität über die Fisherinformationsmatrix wird das Differenzialgleichungssystem durch die partiellen Ableitungen nach den Zustandsvariablen und nach den Parametern betrachtet. Das selbe gilt für die Messgleichungen. Hierüber ist das Differenzialgleichungssystem der Zustandssensitivitäten zugänglich, welches durch analytische Integration lösbar ist. Unter Berücksichtigung der Messfehlerkovarianzmatrix und den Zeitpunkten der Messung kann die Fisherinformationsmatrix berechnet werden, die über die Eigenwertanalyse eine Aussage über die Parameterbestimmbarkeit mit den Messwerten ermöglicht [MUNACK 1989].

Ziel der Sensitivitätsanalyse ist es, dass das Modell um die Parameter reduziert werden kann, die nur eine sehr geringe Sensitivität aufweisen. Diese Parameterreduktion ist bei sehr komplexen Modellen wünschenswert, da durch die hohe Komplexität ein größerer Rechenaufwand nötig ist ohne eine signifikant bessere Simulation zu erzeugen. Die Parameterreduktion sollte jedoch immer unter Berücksichtigung der biologischen Bedeutung der Parameter und Zielsetzung der Modellierung erfolgen.

#### **4.2.4 Optimierung**

Basierend auf dem Prozessmodell kann der Prozess zur Tryptophanproduktion abhängig von der Prozessführung simuliert werden. Durch Veränderung der Umsetzungsbedingungen ist es möglich die optimalen Prozessbedingungen abhängig von den Optimierungskriterien zu identifizieren. Sofern die Optimierungskriterien mathematisch bestimmbar sind und die veränderlichen Prozessbedingungen frei gegeben werden, ist eine mathematische Optimierung durch die oben beschriebenen Optimierungsalgorithmen erstrebenswert.

Bei dem hier beschriebenen Prozess ist die Indolkonzentration, bzw. die Indolzugabe die Stellgröße der Optimierung. Eine hohe Produktkonzentration und eine geringe Indolkonzentration zum Ende des Prozesses sind die Optimierungskriterien.

### **4.3 Onlineanwendung**

Die beschriebenen Verfahren zur Identifikation des Prozesszustandes sind nur dann für eine ökonomischere und ökologischere Produktion sinnvoll, wenn der Prozesszustand online ermittelt werden kann und die Ergebnisse zur optimaleren Prozesssteuerung genutzt werden. Die Onlinevorhersage aus den 2D-Fluoreszenzspektren kann ebenso genutzt werden wie die Simulation unter Berücksichtigung der Prozessführung. Beide Methoden vereint werden darüber hinaus eine Vorhersage für die möglichst optimale weitere Prozessregelung ergeben.

### 4.3.1 Onlinevorhersage

Zur Onlinevorhersage der Prozessvariablen aus den 2D-Fluoreszenzspektren wird ein Computer an den BioView-Messrechner angeschlossen. Über die serielle Verbindung dieser Computer werden die Spektren zur weiteren Verarbeitung übertragen. Ein Programm liest die serielle Schnittstelle aus und berechnet die Vorhersage der Prozessvariablen nach Gleichung (4.18), basierend auf den chemometrischen Modellparametern. Die Vorhersage wird sowohl graphisch als auch numerisch ausgegeben, so dass sowohl der zeitliche Verlauf der Konzentrationen als auch die genauen aktuellen Konzentrationen für den Nutzer zugänglich sind. Ein Screenshot laufenden Programms während eines Prozesses ist in Abbildung 4.3 dargestellt.

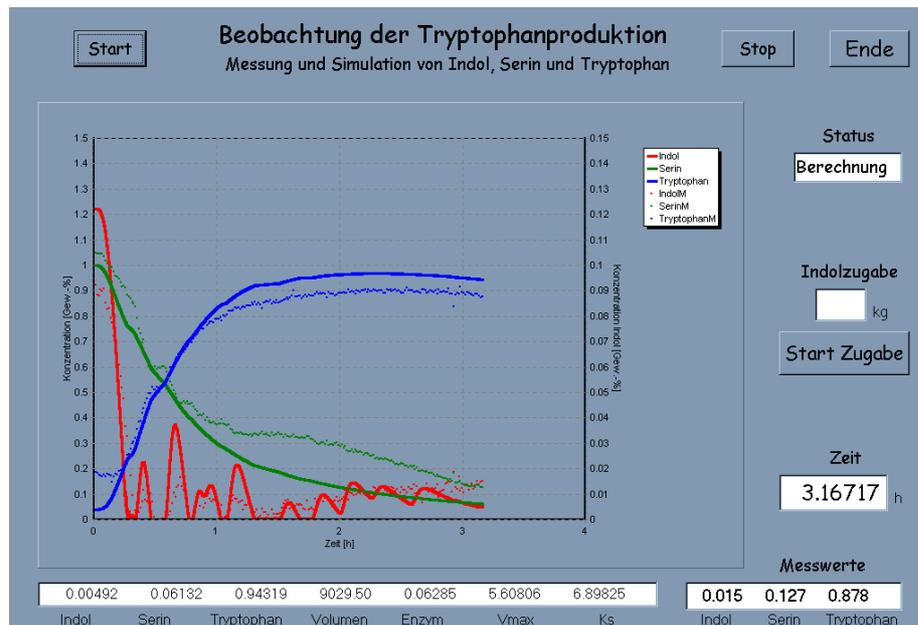


Abbildung 4.3: Screenshot des Programms zur Onlinevorhersage der Prozessvariablen

Die einzelnen Punkte zeigen die diskreten Vorhersagen der Prozessvariablen während der Umsetzung. Im unteren Teil des Fensters werden die aktuellen Messwerte der Konzentrationen numerisch ausgegeben. Die Linien stellen die gefilterten Verläufe der Konzentrationen dar. Die Filterung wird im nächsten Abschnitt beschrieben. Die aktuellen Schätzungen der Prozessvariablen werden unter der Graphik angezeigt. Außerdem werden die aktuelle Prozesslaufzeit und der Status des Programms (Berechnung der Schätzung oder Filterung) angegeben. Ein Eingabefeld für die Indolzugabemengen und Steuerelemente für den Beginn der Zugabe, den Start der Prozesszustandsschätzung und das Beenden ergänzen die graphische Oberfläche.

### 4.3.2 Filterung der Messwerte

Zur Beobachtung eines biotechnischen Prozesses werden meistens die relevanten Zustandsvariablen direkt oder indirekt gemessen. Viele Bioprozessvariablen sind nur über Offlinemessungen mit einer starken Zeitverzögerung detektierbar. Durch moderne Onlineanalytoren

können einige Prozessvariablen zumindest indirekt am Prozess beobachtet werden. Messwerte eines biologischen Systems sind aufgrund der komplexen Matrix jedoch häufig mit stärkeren Messfehlern, bzw. Messrauschen behaftet. Insbesondere bei Onlinemessungen ist es möglich die Messwerte zu filtern, um das Messrauschen zu kompensieren. Insbesondere für die Regelung von Prozessen sind zuverlässige Onlinemesswerte unabdingbar, da basierend auf der Messung der Prozess geregelt wird und ein verrauschtes Messsignal nicht für eine Regelung geeignet ist.

Für Prozessvariablen, die sich im Verhältnis zum Messintervall nur geringfügig ändern, können Mittelwertfilter eingesetzt werden, wie der gleitende Durchschnitt oder der Savitzky-Golay-Filter [OTTO 1997]. Ändert sich die Prozessvariable jedoch schnell, so sind diese "nachtragenden" Filter ungeeignet, da sie sich nur allmählich an den aktuellen Zustand annähern. Für solche Prozesse ist der Kalman-Filter geeignet, der über ein Prozessmodell, den aktuellen Zustand des Prozesses schätzt und bei Eintreffen eines Messwertes diesen filtert und die Zustandsschätzung aktualisiert [HITZMANN 2000]. Darüber hinaus werden die weiteren Zustandsschätzungen basierend auf dem gefilterten Messwert berechnet bis ein neuer Messwert eintrifft [GEE 1996].

Ist das für den Kalman-Filter genutzte Prozessmodell nicht linear und kontinuierlich und die Messungen zeitdiskret so wird der Filter als kontinuierlich-diskreter erweiterter Kalman-Filter bezeichnet. Das nichtlineare Prozessmodell geht über die Modellgleichung (4.36) in die kontinuierliche Zustandsschätzung ein.

$$\frac{dx(t)}{dt} = f(x(t), u(t), t) + z(t) \quad (4.36)$$

Mit  $x(t)$  dem Zustandsvektor der Prozessvariablen, der Funktion  $f(x, u, t)$  zur Berechnung der Prozessvariablen abhängig von den Zustandsvariablen, der Zeit und  $u(t)$  dem Zugangsvektor der Prozessbedingungen und  $z(t)$  dem Prozessrauschen. Die Messungen werden durch die Mess- oder Ausgangsgleichung (4.37) berücksichtigt.

$$y(t_i) = h(x(t_i)) + s(t_i) \quad (4.37)$$

Mit  $y(t_i)$  dem Messwert zu den diskreten Zeitpunkten  $t_i$ ,  $h(x(t_i))$  dem wahren Wert der Zustandsvariablen und  $s(t_i)$  dem Messrauschen. Das Prozess- und Messrauschen wird als weißes Rauschen angenommen, so dass der Mittelwert gleich null, die Varianz jedoch von null verschieden ist. Die Stärke des Rauschens wird durch die Prozessrauschleistungsmatrix  $Q$  in Gleichung (4.38) und die Messrauschkovarianzmatrix  $R$  in Gleichung (4.39) beschrieben.

$$Q \cdot \delta(\tau) = E\{z(t) \cdot z^T(t + \tau)\} \quad (4.38)$$

$$R = E\{s(t_i) \cdot s^T(t_i)\} \quad (4.39)$$

Wobei E jeweils der Erwartungswert der Größe ist. Die Messrauschkovarianzmatrix ist die Diagonalmatrix der Varianzen des Messrauschens. Das Wissen über die Güte der Messung ist in R enthalten und findet so Berücksichtigung bei der Filterung. Die Prozessrauschleistungsmatrix ist entsprechend ein Maß für die Modellgüte. Trotz der Berücksichtigung des Rauschens wird der wahre Prozesszustand vom geschätzten abweichen, diese Abweichung wird in der Schätzfehlerkovarianzmatrix P durch Gleichung (4.40) berücksichtigt.

$$P_i(t) = E\{(x_i(t) - \hat{x}_i(t))^2\} \quad (4.40)$$

Wobei  $x_i(t)$  den wahren Wert und  $\hat{x}_i(t)$  den Schätzwert des Zustandsvektor bedeuten. Die Schätzfehlerkovarianzmatrix wird kontinuierlich nach Gleichung (4.41) berechnet, solange kein Messwert eintrifft.

$$\frac{dP(t)}{dt} = F(t) \cdot P(t) + P(t) \cdot F^T(t) + Q \quad (4.41)$$

Wobei F(t) die Matrix der partiellen Ableitungen der Modellgleichungen nach den Zustandsvariablen, also die Jakobimatrix von  $f(x(t), u(t), t)$  ist. Sobald ein Messwert eintrifft wird die Schätzfehlerkovarianz nach Gleichung (4.42) neu berechnet.

$$P_{neu}(t_i) = [I - K(t_i) \cdot H(t_i)] \cdot P(t_i) \cdot [I - K(t_i) \cdot H(t_i)]^T + K(t_i) \cdot R \cdot K^T(t_i) \quad (4.42)$$

Wobei I die Einheitsmatrix, K(t<sub>i</sub>) die Kalman-Verstärkungsmatrix und H(t<sub>i</sub>) die Jacobi-Matrix von  $h(x(t_i))$  ist. Die Kalman-Verstärkungsmatrix berechnet sich zuvor durch Gleichung (4.43) ebenso aus H(t<sub>i</sub>) und P(t<sub>i</sub>).

$$K(t_i) = P(t_i) \cdot H^T(t_i) \cdot [H(t_i) \cdot P(t_i) \cdot H^T(t_i) + R]^{-1} \quad (4.43)$$

Die Kalman-Verstärkungsmatrix wird neben der Neuberechnung der Schätzfehlerkovarianzmatrix bei Eintreffen eines Messwertes auch für die neue Zustandsschätzung durch Gleichung (4.44) benötigt.

$$\hat{x}_{neu}(t_i) = \hat{x}(t_i) + K(t_i) \cdot [y(t_i) - h(\hat{x}(t_i), t_i)] \quad (4.44)$$

Die Kalman-Verstärkungsmatrix wird von der Messrauschkovarianzmatrix und Prozessrauschleistungsmatrix in der Art beeinflusst, dass sich bei kleinem Messrauschen die Zustandsschätzung am Messwert orientiert und bei kleinem Prozessrauschen eher am Prozessmodell.

## 5 Ergebnisse

Die Offlinemesswerte und Fluoreszenzspektren der in Kapitel 3.3.3 beschriebenen zehn Prozesse (U1 – U10) werden zur Auswertung herangezogen. Es wird die Entwicklung des Prozessmodells zur Simulationen der am Industrieprozess zur Tryptophanproduktion gemessenen Offlinemesswerten beschrieben. Durch die Anpassung der kinetischen Modellparameter an die Offlinemesswerte wird das Prozessmodell determiniert und verschiedene Simulationen des Prozesses werden vorgestellt. Dieses Modell beschreibt den Prozess unter unterschiedlichen Umsetzungsbedingungen und wird zur Optimierung des Prozesses genutzt. Aus den gemessenen Fluoreszenzspektren und Offlinemesswerten werden chemometrische Modelle zur Vorhersage der Prozessvariablen erstellt und die Modelle und Ihre Entwicklung vorgestellt. Außerdem sollen prozessrelevante Informationen aus den Fluoreszenzspektren ermittelt werden. Die chemometrischen Modelle werden zur Onlinevorhersage der Prozessvariablen während einer Umsetzung eingesetzt. Darüber hinaus können die Onlinevorhersagen durch den Kalman-Filter mit dem Prozessmodell verknüpft werden und noch präzisere Schätzwerte für die Prozessvariablen Indol, Serin und Tryptophan erzielt werden. Basierend auf den Onlineschätzungen der Prozessvariablen und unter Berücksichtigung der idealen Prozessbedingungen kann der Prozess optimaler geführt werden. Die Modellerstellung, die Anwendung der Modelle und die Optimierung werden in diesem Kapitel vorgestellt.

### 5.1 Das theoretische Prozessmodell

Die Grundlagen eines Prozessmodells sind in Kapitel 4.2 dargestellt. Für ein hinreichend genaues Prozessmodell muss eine beschreibende Kinetik gefunden werden. In der Literatur [VORLOP 1983] und [MATEUS 1996] wird eine Tryptophanbildungskinetik (Gleichung (4.25)) beschrieben. Diese Kinetik gilt für Umsetzungen mit immobilisierten Zellen und wurde nicht für den industriellen Prozess entwickelt. Darüber hinaus wird sie durch die Erkenntnisse von Lane [LANE 1983A] über den Ordered-Mechanismus aufgehoben (siehe Kapitel 4.2.1). In der Kinetik von Vorlop [VORLOP 1983] wird außerdem eine Indolinhibierung berücksichtigt, die im industriellen Prozess aufgrund zu geringer Indolkonzentrationen nicht beobachtet werden konnte und daher auch nicht während des industriellen Prozesses beschrieben werden kann. Deshalb wird diese Kinetik nicht in dem Prozessmodell für die Simulation des Umsetzungsprozesses mit freien Zellen im Industriemaßstab genutzt. Anstatt der Kinetik aus der Literatur wird für die Entwicklung des Prozessmodells die Ordered-Kinetik aus Kapitel 4.2.1 (Gleichungen (4.27) – (4.29)) als Grundlage genommen. Diese allgemeine Kinetik für Enzymreaktionen mit zwei Substraten wird in das allgemeine

Prozessmodell aus Kapitel 4.2 (Gleichungen (4.20) – (4.23)) eingefügt. Das resultierende Differenzialgleichungssystem für die Umsetzung der Substrate Indol und Serin zum Produkt Tryptophan ist durch die Gleichungen (5.1) – (5.4) gegeben.

$$\frac{dI}{dt} = \frac{\dot{m}_I}{V} - \frac{\dot{V}_{BFM} + \dot{V}_{Kond}}{V} I - \mu_{\max} \frac{S}{K_m^S + S} \cdot \frac{I}{K_m^I + I} \quad (5.1)$$

$$\frac{dS}{dt} = -\frac{\dot{V}_{BFM} + \dot{V}_{Kond}}{V} S - \mu_{\max} \frac{S}{K_m^S + S} \cdot \frac{I}{K_m^I + I} \quad (5.2)$$

$$\frac{dT}{dt} = -\frac{\dot{V}_{BFM} + \dot{V}_{Kond}}{V} T + \mu_{\max} \frac{S}{K_m^S + S} \cdot \frac{I}{K_m^I + I} \quad (5.3)$$

$$\frac{dV}{dt} = \dot{V}_{BFM} + \dot{V}_{kond} \quad (5.4)$$

Mit  $I$  der Indolkonzentration,  $S$  der Serinkonzentration und  $T$  der Tryptophankonzentration,  $\dot{V}_{BFM}$  und  $\dot{V}_{kond}$  dem Volumenstrom der Biofeuchtmasse und des Kondensats und  $\dot{m}_I$  dem Massenstrom von Indol während der Indolzugabe. Die maximale Reaktionsrate  $\mu_{\max}$  ist ein Parameter, in den sowohl die Enzymmenge, als auch die Aktivität und andere enzyspezifische Größen eingehen. Daher wird die maximale Reaktionsrate durch die Zellzugabe zu Beginn des Prozesses beeinflusst und ist nicht während des gesamten Prozesses konstant. Um die zeitliche Veränderung der Zellmasse und damit die Enzymmenge in das Modell aufzunehmen, wird das Differenzialgleichungssystem um eine Gleichung für die Zellmasse erweitert. Die eingeführte Größe  $X$  gibt den relativen Teil an Biofeuchtmasse im System an und ergibt sich durch Multiplikation mit einer Konstante  $k$  zu der maximalen Reaktionsrate  $\mu_{\max}$ . Durch die Gleichung (5.5) wird die Zunahme der relativen Zellmasse im System  $X$  durch die Zellzugabe und deren Abnahme durch die Verdünnung beschrieben. Die Zellzugabe wird durch den Volumenstrom  $\dot{V}_{BFM}$  und die relative Zellmassenkonzentration  $X_0$  in der Zugabe beschrieben. Gleichung (5.6) beschreibt den Zusammenhang zwischen der relativen Zellmassen und der maximalen Reaktionsrate  $\mu_{\max}$ .

$$\frac{dX}{dt} = \frac{\dot{V}_{BFM}}{V} X_0 - \frac{\dot{V}_{BFM} + \dot{V}_{Kond}}{V} X \quad (5.5)$$

$$\mu_{\max} = k \cdot X \quad (5.6)$$

Bei der Beobachtung der industriellen Prozesse trat ein Serinverlust auf, ein Teil des eingesetzten L-Serins wird weder zu Tryptophan umgesetzt, noch verbleibt es in der Reaktionsmasse. Eine Racemisierung zu D-Serin kann ausgeschlossen werden, da bei der Analytik zwi-

schen den Enantiomeren nicht unterschieden wird und daher D-Serin von der Analytik erfasst wird. Der Substratverbrauch wird auf eine Nebenreaktion des Serins zurückgeführt. Eine Desaminierung des Serins zu Pyruvat und Ammoniak um könnte hierfür verantwortlich sein. Zum einen wird eine desaminierende Nebenaktivität des Enzyms Tryptophansynthase selbst beschrieben, zum anderen können weitere desaminierende Enzyme eventuell noch in der Zelle aktiv sein. Da noch eine erhebliche Zellaktivität nachgewiesen werden konnte [KLABEN 2003], kann eine solche Verstoffwechselung des Serins als Kohlenstoff- oder Stickstoffquelle der Zellen nicht ausgeschlossen werden.

Um den Verbrauch an Serin im Modell zu berücksichtigen und eine hinreichende Beschreibung der Serinmesswerte zu ermöglichen, muss für die Nebenreaktion eine beschreibende Kinetik gefunden werden. Hierfür wird der Serinverlust zu jedem Messzeitpunkt aus dem eingesetzten Serin, dem noch verbleibenden Serin und den zu Tryptophan umgesetzten Serin berechnet. Da die berechneten Werte für den Serinverlust den Messwertschwankungen unterliegen, wurde der Verlauf des Serinverlustes durch die Funktion in Gleichung (5.7) angepasst.

$$SV(t) = A(1 - e^{-Bt}) \quad (5.7)$$

Wobei  $SV(t)$  der Verlauf des Serinverlustes bedeutet. Durch die Anpassung wurde der Parameter  $A$  mit 0,55 Gew.-% und der Parameter  $B$  mit  $0,98 \text{ h}^{-1}$  bestimmt. Zur Identifikation der Kinetik für die Nebenreaktion wird die 1. Ableitung des Serinverlustes berechnet, um die Änderung des Serinverlustes über die Zeit zu betrachten. Aus der resultierenden Funktion (Gleichung (5.8)) für den Serinverlust kann die Kinetik der Nebenreaktion bestimmt werden.

$$SV'(t) = A \cdot B \cdot e^{-Bt} \quad (5.8)$$

Die in Abbildung 5.1 dargestellte Serinkonzentration ist durch die Offlinemesswerte des Prozesses U5 gegeben. Die aus den Messwerten berechneten Serinverluste, sowie die Anpassung des Verlustes und die 1. Ableitung zur Identifikation der Kinetik der Nebenreaktion sind ebenfalls in der Darstellung abgebildet.

Mit der Abnahme der Serinkonzentration steigt der Serinverlust bis auf über 0,5 Gew.-% bei 4 h Prozesszeit an. Dieser Verlust ist nicht allein durch die Volumenzunahme und den damit verbundenen Verdünnungseffekt zurückzuführen, da dieser nur etwa 10 % ausmacht. Die Ableitung der Funktion des Serinverlustes zeigt zu Beginn des Prozesses einen Verlust von etwa 0,5 Gew.-%/h und sinkt binnen vier Stunden Prozesslaufzeit auf null ab. Auffällig ist die hohe Korrelation der zeitlichen Änderung des Serinverlustes mit dem Verlauf der Serinkonzentration selbst, so dass angenommen werden kann, dass der Serinverlust von der Serinkonzentration direkt abhängig ist. Daher wird für die Nebenreaktion näherungsweise eine Kinetik erster Ordnung angenommen.

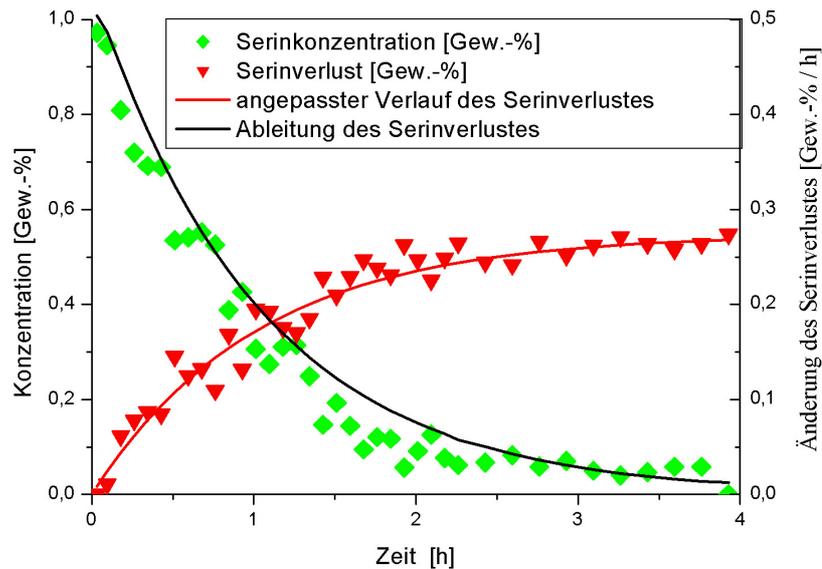


Abbildung 5.1: Die Serinkonzentration, der Serinverlust und die Änderung des Serinverlustes am Beispiel des Prozesses U5

Ist der Serinverlust von der Serinkonzentration abhängig, so sollten Prozesse mit hohen Serinkonzentrationen auch einen hohen Verlust aufweisen. Für die meisten Prozesse ist nur die Serinstart- und Serinendkonzentration bekannt. Aber gerade zu Beginn des Prozesses bei hohen Serinkonzentrationen tritt offenbar der höchste Verlust auf. Daher sollte der Serinverlust in Abhängigkeit von der Serinkonzentration des Eluates eine Korrelation aufweisen. In Abbildung 5.2 sind die Ausbeuten an Tryptophan in Bezug auf den Serineinsatz der Prozesse gegenüber den entsprechenden Serinstartkonzentrationen von allen Prozessen eines Jahres aufgetragen.

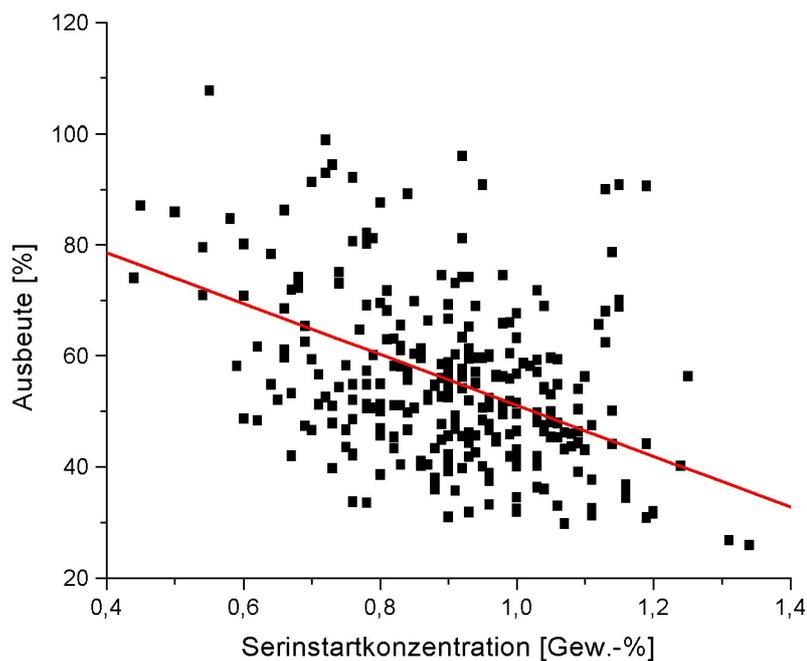


Abbildung 5.2: Korrelation der Ausbeuten zur Serinstartkonzentration aller Prozesse des Jahres 2001

Die Gegenüberstellung von Serinstartkonzentration und Ausbeute weist eine starke Streuung auf. Ausbeuten von über 100 % können auf Messfehler zurückgeführt werden, des weiteren wirken sich unterschiedliche Prozessführung und Unterschiede im Serineluat auf die Streuungen aus. Trotzdem lässt sich eine deutliche Tendenz erkennen, die durch die rot eingezeichnete Gerade wiedergegeben wird. Somit wird der direkte Zusammenhang zwischen dem Serinverlust und der Serinkonzentration über die Prozesse des Jahres 2001 bestätigt.

Für die Nebenreaktion des Serins muss auch die zeitlichen Änderung der Biomasse berücksichtigt werden, da der Biokatalysator für die Nebenreaktion ebenso durch die Zugabe der Zellmasse aufgebaut wird, wie bei der Tryptophanproduktion. Unter Berücksichtigung der Nebenreaktion und der zeitlichen Änderung der Biomasse ergibt sich für das Prozessmodell das durch die Gleichungen (5.9) – (5.13) dargestellte Differenzialgleichungssystem.

$$\frac{dI}{dt} = \frac{\dot{m}_I}{V} - \frac{\dot{V}_{BFM} + \dot{V}_{Kond}}{V} I - k \cdot X \frac{S}{K_m^S + S} \cdot \frac{I}{K_m^I + I} \quad (5.9)$$

$$\frac{dS}{dt} = -k_S \cdot X \cdot S - \frac{\dot{V}_{BFM} + \dot{V}_{Kond}}{V} S - k \cdot X \frac{S}{K_m^S + S} \cdot \frac{I}{K_m^I + I} \quad (5.10)$$

$$\frac{dT}{dt} = -\frac{\dot{V}_{BFM} + \dot{V}_{Kond}}{V} T + k \cdot X \frac{S}{K_m^S + S} \cdot \frac{I}{K_m^I + I} \quad (5.11)$$

$$\frac{dV}{dt} = \dot{V}_{BFM} + \dot{V}_{kond} \quad (5.12)$$

$$\frac{dX}{dt} = \frac{\dot{V}_{BFM}}{V} X_0 - \frac{\dot{V}_{BFM} + \dot{V}_{Kond}}{V} X \quad (5.13)$$

Der Volumenstrom des Kondensats beträgt während der gesamten Umsetzung etwa 33 L/h. Der Volumenstrom der Biomassezugabe beträgt während der Zugabezeit etwa 2800 L/h und ist im weiteren Verlauf des Prozesses null. Da sich die gekühlte Zellmasse jedoch erst in der Reaktionsmasse verteilen und temperieren muss, verzögert sich die Aktivität der Zellmasse gegenüber der Zellzugabe. Für die Simulation wird daher nicht die Sprungfunktion für die Zellzugabe angenommen sondern ein Zugabeprofil, welches sich durch eine Rührkesselkaskade ergibt. Aus Vorversuchen wurde die Zellzugabefunktion durch Anpassung der Zellzugabe an den Verlauf der Fluoreszenzintensitäten einer charakteristischen Wellenlängenkombination während der Zellzugabe bestimmt. Hierdurch wurde für die Zellzugabe eine Reaktorzahl von vier und eine Verweilzeit von 0,038 h bei einem mittleren Fehler von 4,34 % bestimmt. Die Funktion für die Indolzugabe wird durch die Zeitpunkte der Indolzugabe und die dazugehörigen Indolmengen bestimmt, wobei von einer Reinheit des Indol von 95 %

ausgegangen werden kann. Der Massenstrom der Indolzugabe beträgt etwa 100 kg/h. Auch hier muss eine Verzögerung durch die Verteilung und Lösung des Indol berücksichtigt werden. Daher wurde auch hier die Zugabe als Profil einer Rührkesselkaskade angenommen und die Funktion an den Verlauf der Fluoreszenzintensitäten einer charakteristischen Wellenlängenkombination angepasst. Mit einem mittleren Fehler von 2,16 % wurde für die Indolzugabe eine Kesselzahl von vier bei einer Verweilzeit von 0,018 h bestimmt.

Das Startvolumen  $V(0)$  wird durch die Menge an Serineluat festgelegt und steht vor Prozessbeginn fest. Die Konzentration von Indol zu Beginn des Prozesses  $I(0)$  kann aus der Menge an vorgelegtem Indol und dem Startvolumen bestimmt werden. Die Serinstartkonzentration  $S(0)$  ist aus der späteren Offlineanalyse bekannt. Die Produktkonzentration zu Beginn des Prozesses ist null, wobei das in der Biofeuchtmasse der gebrauchten Zellen vorhandene Tryptophan mit etwa 0,5 Gew.-% angenähert werden kann und in der Bilanz und der Simulation berücksichtigt werden sollte. Die Menge an Biomasse liegt vor der Biomassezugabe bei null.

Das Modell enthält somit noch vier zu bestimmende Parameter:  $k$  als Geschwindigkeitskonstante für die Reaktion der Tryptophanbildung,  $k_S$  als Geschwindigkeitskonstante für die Nebenreaktion 1. Ordnung und  $K_m^I$  sowie  $K_m^S$  als Michaelis-Menten-Konstanten der beiden Substrate Serin und Indol. Für die Prozesssimulation müssen diese Parameter bestimmt werden, hierfür wird aus dem Differenzialgleichungssystem ein SIMULINK<sup>®</sup> Modell erstellt. Die SIMULINK<sup>®</sup> Darstellung des Prozessmodells ist im Anhang 8.3 zu finden.

### 5.1.1 Parameteridentifikation

Zur Parameteranpassung des Prozessmodells werden Simulationen des Modells durchgeführt und das Ergebnis der Simulation mit den Offlinemesswerten verglichen. Die Abweichung der Simulation von den Messwerten nach Gleichung (4.30) gilt als Maß für die Güte der Parameterkombination, die für die Simulation benutzt wurde. Durch Variation der Parameter wird mit Hilfe von Optimierungsalgorithmen die Parameterkombination identifiziert, die die Messwerte am besten beschreibt. Die Parameter sollten jedoch nicht beliebige Werte annehmen, sondern nur im Rahmen ihrer biologischen Bedeutung liegen.

Zur Ermittlung der biologisch sinnvollen Grenzen für die Michaelis-Menten-Konstanten während der Anpassung der kinetischen Parameter stehen verschiedene Veröffentlichungen zur Verfügung. In Tabelle 5.1 sind die Literaturwerte für die Michaelis-Menten-Konstanten dieser Reaktion neben den entsprechenden Quellen angegeben.

Die erwähnten Werte der kinetischen Konstanten für das freie Enzym und die ganze Zelle sind sehr unterschiedlich. Dies ist darauf zurückzuführen, dass die Substratverteilungen am

Enzym innerhalb der Zellen und beim freien Enzym unterschiedlich sind. Diese Differenz nimmt mit dem Alter der Zellen und der Permeabilisierung der Zellwand ab. Die Abhängigkeit der Konstanten vom Zellalter wird auch für die Geschwindigkeit der Reaktion beschrieben (Zellaktivität 160 - 230 mg Trp/h/gBFM) [VORLOP 1983].

Tabelle 5.1: Literaturwerte für die Michaelis-Menten-Konstanten [mmol/L]

Enzymform	$K_m^I$	$K_m^S$	Literatur
freies Enzym (Bindung an das Enzym)	0,0037 <sup>1)</sup>	0,077	[Heilmann 1978]
freies Enzym (Bindung an den Komplex)	0,014	0,313 <sup>1)</sup>	[Heilmann 1978]
freies Enzym	0,165	24,6	[Hatanaka 1962]
freies Enzym	0,17	1,5	[Yanofsky 1960]
freies Enzym	0,013	0,43	[Lane 1991]
Ganzzelle	3,6		[Vorlop 1983]

<sup>1)</sup> Diese Konstanten wurden unter der Annahme eines Random-Mechanismus angegeben.

Es liegen nicht ausreichend Werte für Reaktionen mit ganzen Zellen vor, die eine sinnvolle Wahl der Grenzen erlauben. Außerdem sind die Konstanten unter Laborbedingungen mit reinen Substratlösungen identifiziert worden. Es kann nicht davon ausgegangen werden, dass die Bedingungen im Industriemaßstab, in einer komplexen Matrix und bei unterschiedlichem Zellalter vergleichbar sind. Daher müssen nicht nur die Grenzen für die Geschwindigkeitskonstanten  $k_s$  und  $k$  festgelegt werden, sondern auch die Grenzen für die Michaelis-Menten-Konstanten.

Als Grenzen für die Michaelis-Menten-Konstanten werden  $10^{-6}$  mol/L und das 10-fache der jeweiligen maximal Konzentration der Substrate angenommen. Als untere Grenze ist 0 mol/L nicht zulässig, da unzulässige arithmetische Operationen zustande kommen, wenn während der Simulationen Substratkonzentrationen von 0 mol/L erreicht werden. Das 10-fache der maximalen Konzentration der Substrate sollte als obere Grenze ausreichen, da in der Nähe dieser Grenze die Michaelis-Menten-Konstanten kinetisch nicht mehr wirksam sind. In diesem Fall kann das Modell dann um den entsprechenden Parameter reduziert werden, da in diesem Fall eine Reaktion 1. Ordnung bezüglich des Substrates vorliegen würde. Die Konstanten für die Enzymaktivität der Tryptophanproduktion und der Nebenreaktion sollte zwischen 1 und 100 liegen, wie Voruntersuchungen ergeben haben. Die gewählten Grenzen für die Parameter sind in Tabelle 5.2 zusammengefasst.

Tabelle 5.2: Grenzen der Parameter für die Parameteridentifikation

Parameter	untere Grenze	obere Grenze
k [mol/L/h]	0	100
$k_S$ [h <sup>-1</sup> ]	0	100
$K_{m_I}$ [mol/L]	$1 \cdot 10^{-6}$	0,1
$K_{m_S}$ [mol/L]	$1 \cdot 10^{-6}$	1

Die Parameteridentifikation wurde für die sechs Prozesse durchgeführt, die im November 2001 mit Offlinemessungen begleitet wurden. Es wurde für jeden einzelnen Prozess die optimale Parameterkombination ermittelt. Um eine allgemeingültige Parameterkombination zu erhalten wurden für die Prozesse U1 – U5 eine gemeinsame Parameteridentifikation durchgeführt. Die Ergebnisse der Anpassung für die einzelnen Parameter sind in Tabelle 5.3 zusammenfassend dargestellt. Die angegebene Fehlerquadratsumme wurde jeweils nach Gleichung (4.30) berechnet, so dass diese Größe von der Anzahl der Messwerte, die für die Anpassung benutzt wurden abhängt.

Tabelle 5.3: Ergebnisse der Parameteranpassungen von sechs verschiedenen Prozessen

	U1	U2	U3	U4	U5	U6	U1 – U5
k [mol/L/h]	9,0	13,9	31,5	5,5	17,2	35,3	6,2
$k_S$ [h <sup>-1</sup> ]	1,9	6,3	5,9	5,2	7,7	6,3	4,4
$K_m^I$ [mol/L]	0,0029	0,0012	0,0056	0,0008	0,00015	0,0019	0,0046
$K_m^S$ [mol/L]	1	1	1	0,50	0,69	1	0,18
Fehlerquadratsumme [mol <sup>2</sup> /L <sup>2</sup> ]	0,20	0,21	0,29	0,15	0,18	0,16	2,82

Die Unterschiede bei den Anpassungen fallen besonders bei den Werten für die kinetischen Parameter der maximalen Reaktionsgeschwindigkeiten der Haupt- und Nebenreaktion k und  $k_S$  auf. Die Geschwindigkeitskonstante k weist eine starke Streuung auf, die nur durch eine unterschiedliche Konzentration oder Aktivität des Enzyms in der Biofeuchtmasse zu erklären ist. Diese Unterschiede können durch die Separation der Zellen von der Reaktionsmasse der vorangegangenen Umsetzung entstehen oder auch durch die Anzahl der Prozesse, die die Zellen durchlaufen haben. Für das Zellalter wurde eine statistische Verteilung des Umsatzes für die einzelnen Zellklassen berechnet, die in Abbildung 5.3 dargestellt ist.

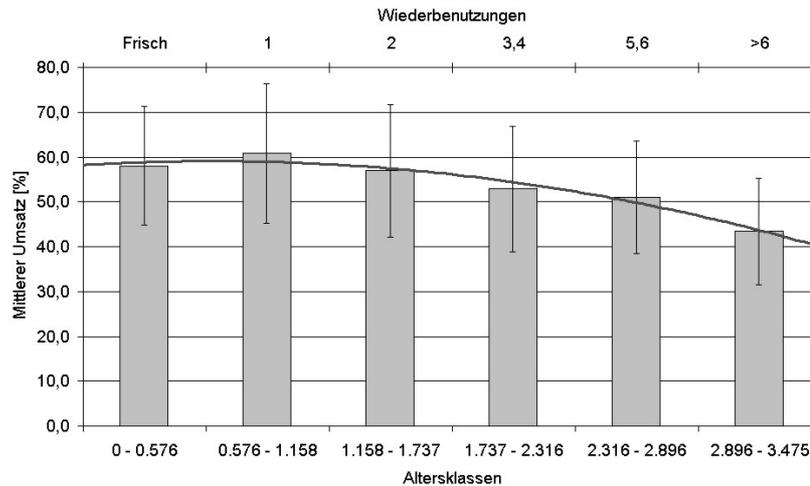


Abbildung 5.3: Verteilung des mittleren Umsatzes auf verschiedene Zellklassen

In der Darstellung ist zu erkennen, dass der mittlere Umsatz für einmal eingesetzte Zellen größer ist, als für erstmalig benutzte. Ab einer mittleren Wiederbenutzung der Zellen von mehr als eins nimmt die erreichte Ausbeute wieder ab.

Die unterschiedliche Konzentration an Enzym lässt sich zu Beginn des Prozesses nicht bestimmen. Die Anpassung über die ersten fünf Prozesse ergibt einen Wert für diesen Parameter, der nicht im mittleren Bereich der Einzelanpassungen liegt und daher für die meisten Prozesse zu niedrig ist. Die Unterschiede der Werte des Parameters  $k_S$  zeigt ebenso große Unterschiede, die nicht mit den Schwankungen für  $k$  korreliert werden können. Hier scheint die Anpassung über fünf Prozesse einen guten mittleren Wert für den Parameter zu ergeben.

Die Anpassung der Michaelis-Menten-Konstante für Indol weist zum Teil sehr kleine Werte auf, wobei die Anpassung über fünf Prozesse im oberen Bereich der Einzelanpassungen liegt. Die während eines Prozesses auftretenden Indolkonzentrationen liegen zwischen 0 mol/L und 0,01 mol/L bei einer geschätzten mittleren Indolkonzentration von 0,003 mol/L. Damit ist die Michaelis-Menten-Konstante für Indol nur zu wenigen Zeitpunkten während einer Umsetzung kinetisch wirksam und damit durch die Messwerte auch nur schwer identifizierbar. Die entsprechende Konstante für Serin hingegen liegt für die meisten Prozesse an der gewählten Obergrenze. Für die gemeinsame Anpassung über mehrere Prozesse liegt der Wert dieses Parameters jedoch deutlich unterhalb der Werte der Einzelanpassungen. Diese deutlichen Unterschiede zwischen Einzelanpassungen und der gemeinsamen Anpassung lassen auf eine geringe Sensitivität dieser Parameter schließen, da diese Parameter offenbar die Simulation nicht besonders stark beeinflussen.

Die Unterschiede in den Fehlern der Anpassung deuten auf eine unterschiedlich starke Streuung der Messwerte oder eine unterschiedlich gute Beschreibung der Messwerte hin. Der Fehler für die gemeinsame Anpassung liegt deutlich über der Summe der Fehler aus den

Einzelanpassungen, da die Anpassung über fünf Prozesse jeden einzelnen Prozess natürlich nicht optimal wiedergibt. Als Beispiel des Ergebnisses einer Parameteranpassung wird die Simulation mit der, für diesen Prozess bestimmten Parameterwerte in Abbildung 5.4 mit den Offlinemesswerten dargestellt.

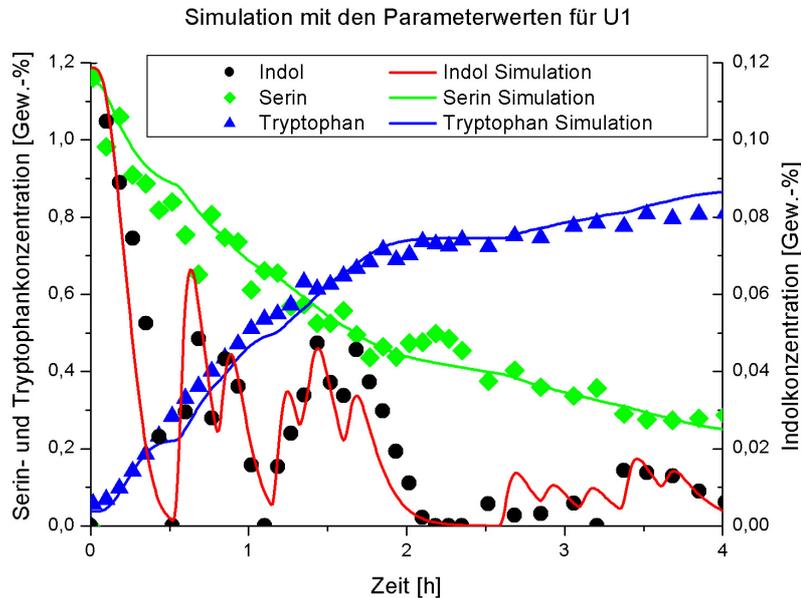


Abbildung 5.4: Die Simulation des Prozesses U1 mit der Parameterkombination dieses Prozesses und den Offlinemesswerten der Prozessvariablen

Die Simulation beschreibt die Offlinemesswerte von Serin gut. Die Messwerte von Tryptophan liegen zu Beginn des Prozesses meist oberhalb der Simulation. Das Prozessmodell kann bei vollständigem Indolverbrauch jedoch keine höheren Tryptophankonzentrationen beschreiben. Die Indolmesswerte, insbesondere die Anstiege durch die Indolzugaben und die nachfolgende Abreaktion des Indols, werden von der Simulation gut wiedergegeben.

Die Parameteranpassung über fünf Prozesse hat jedoch deutlich andere Parameterwerte ergeben als bei der Einzelparameteranpassung. Dass die Simulationen trotz der unterschiedlichen Parameterwerte auch in diesem Fall der charakteristische Verlauf der Messwerte wiedergibt, zeigt die Abbildung 5.5.

Die Simulation von Serin beschreibt die Offlinemesswerte nicht so gut wie noch zuvor. Der über fünf Prozesse identifizierte Parameter für die Nebenreaktion ist größer, weshalb die Simulation der Serinmesswerte nicht mehr so gut gelingt. Die Simulation von Tryptophan verläuft zwar anders, beschreibt die Messwerte jedoch nicht signifikant schlechter. An der Simulation von Indol ist die größere Reaktionsgeschwindigkeit der Hauptreaktion zu erkennen. Trotz der unterschiedlichen Parameter geben beide Simulationen den Verlauf der Messwerte wieder. Die Fehlerquadratsumme der Simulation steigt von  $0,20 \text{ mol}^2/\text{L}^2$  für die Parameterwerte dieses Prozesses auf  $1,30 \text{ mol}^2/\text{L}^2$  für die Parameter aus mehreren Prozessen.

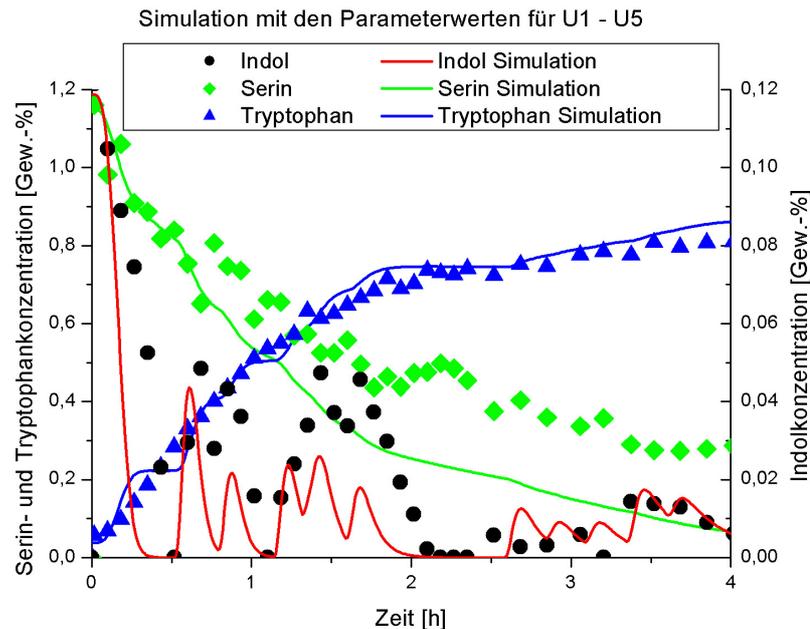


Abbildung 5.5: Die Simulationen des Prozesses U1 mit der Parameterkombination aus fünf Prozessen

Die über fünf Prozesse identifizierte Parameterkombination erweist sich auch bei großen Unterschieden zu der individuell optimalen Parameterkombination als geeignet den Prozess zu beschreiben. Die Simulation der Indolkonzentration und damit die Anpassung des Parameters  $k$  sollte für eine Onlineanwendung verbessert werden oder für den einzelnen Prozess genauer abzuschätzen sein. Insbesondere, da eine Optimierung oder Regelung des Prozesses nur basierend auf der Kenntnis der Indolkonzentration möglich ist. Eine Abschätzung der unterschiedlichen Enzymaktivität und damit eine Verbesserung der Simulation durch genauere Bestimmung des Parameters  $k$  ist wünschenswert und kann durch die Auswertung der Fluoreszenzspektren zugänglich gemacht werden, wie in Kapitel 5.2.3 dargestellt.

### 5.1.2 Parametersensitivität

Durch den Vergleich der beiden Simulationen sind die Wirkungsweisen der beiden Geschwindigkeitsparameter deutlich zu erkennen. Eine Variation dieser Parameter wirkt sich deutlich auf die Simulation aus. Für die beiden Michaelis-Menten-Konstanten sind entsprechende Effekte trotz der zum Teil großen Unterschiede zwischen den beiden dargestellten Parameterwerten, nicht sofort ersichtlich.

Die Relevanz der Parameter für die Simulation kann mit Hilfe der Sensitivitätsuntersuchung berechnet werden. Parameter mit einer geringen Sensitivität können aus dem Modell entfernt werden, ohne dass der Fehler der Simulation signifikant ansteigt. Eine Vereinfachung des Modells ist wünschenswert, um den Rechenaufwand während der Simulationen und Onlineanwendungen zu reduzieren und unnötige Fehlerquellen zu beseitigen.

Die Variation eines Parameters bei konstanten übrigen Parametern zeigt die in Abbildung 5.6 dargestellten univariaten Parametersensitivitäten. Hierbei wird für jeden Parameter abhängig von der prozentualen Abweichung von dem optimalen Parameterwert der Fehler zwischen Simulation und Messwerten nach Gleichung (4.30) berechnet und dargestellt.

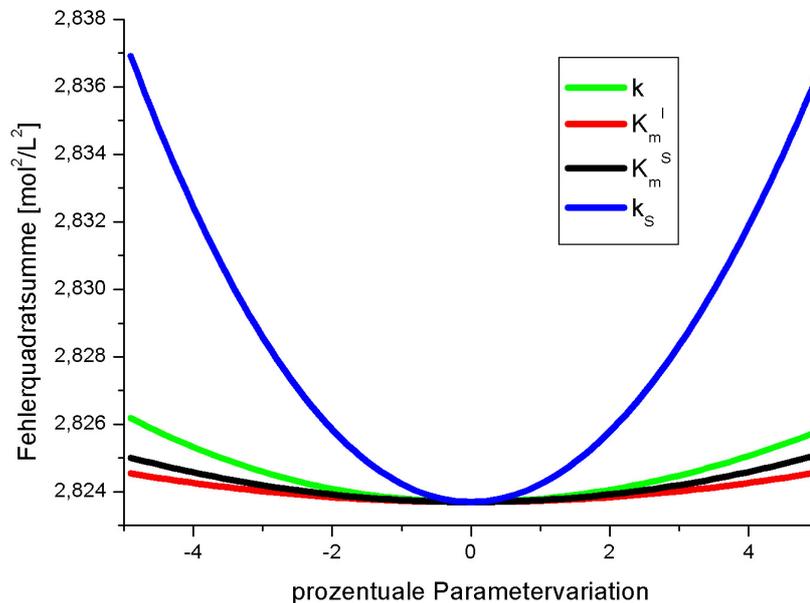


Abbildung 5.6: Univariate Parametersensitivitäten für die Parameteranpassung über U1 – U5

Anhand der univariaten Parametersensitivitäten ist zu erkennen, dass die kinetischen Parameter für die Enzymaktivitäten  $k$  und  $k_S$  die Simulation stärker beeinflussen und somit besser bestimmbar sind. Wobei der Parameter  $k_S$  für die Nebenreaktion in seiner Sensitivität deutlich über dem Parameter  $k$  für die Tryptophanbildung liegt. Die univariaten Parametersensitivitäten zeigen außerdem, dass die Michaelis-Menten-Konstante für Indol  $K_m^I$  schlechter bestimmt werden kann als die entsprechende Konstante für Serin  $K_m^S$ .

Bei der numerischen Sensitivitätsanalyse nach Hearne [HEARNE 1985] ergibt sich eine Sensitivität von 100 % für den Parameter  $k_S$ , gefolgt von  $k$  mit 5,2 % Sensitivität,  $K_m^S$  mit 4,3 % und  $K_m^I$  weist die geringste Sensitivität von 3,7 % auf. Entsprechend den univariaten Sensitivitäten ergibt sich für die numerisch bestimmten Sensitivitäten die Reihenfolge, dass  $k_S$  am besten bestimmbar ist, gefolgt von  $k$ ,  $K_m^S$  und  $K_m^I$ . Erstaunlich ist jedoch die unerwartet geringe Sensitivität von 5,2 % für  $k$ , dies ist aus der univariaten Darstellung und dem Vergleich der Simulation für unterschiedliche Parameterkombinationen nicht sofort ersichtlich. Auch die geringe Sensitivität von  $K_m^I$  ist in sofern erstaunlich, da dieser Parameter bei den Parameteranpassungen im Vergleich zum Parameter  $K_m^S$  zumeist kinetisch wirksame Werte aufweist.

Für eine Analyse der Ergebnisse der Sensitivitätsuntersuchungen eignet sich die Betrachtung der semirelativen Zustandssensitivitätsfunktionen. Hier ist zu entnehmen, zu welchen Zeitpunkten des Prozesses, bzw. bei welchen Konzentrationen der Prozessvariablen, die Para-

meter bestimmbar sind und wie sich die einzelnen Parameter auf die Simulation der Prozessvariablen auswirken. Die Zustandssensitivitätsfunktionen für Indol entsprechen in etwa denen des Tryptophans in Abbildung 5.7, weshalb auf diese Darstellung verzichtet wurde.

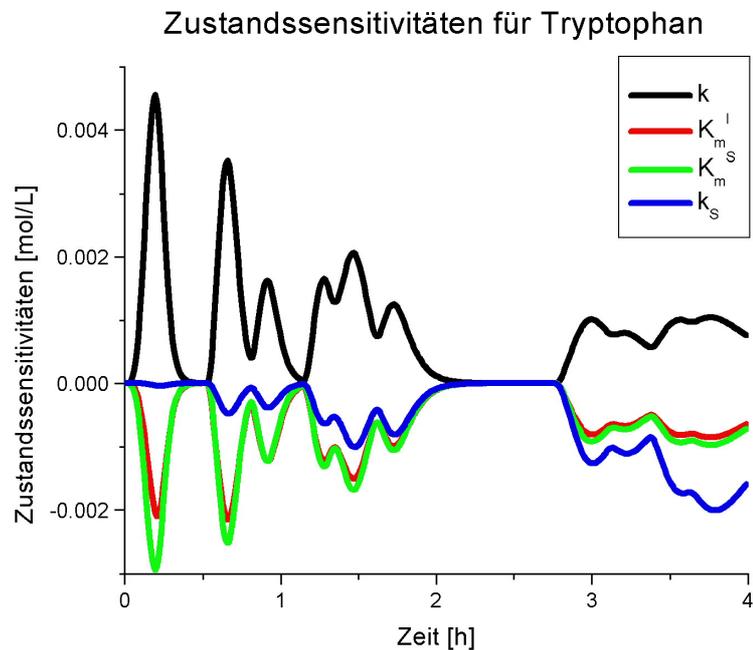


Abbildung 5.7 Zustandssensitivitätsfunktionen der vier Parameter für Tryptophan bei dem Prozess U1

Die Zustandssensitivitätsfunktion für  $k_S$  zeigt, dass dieser Parameter sich kaum auf die Simulation von Indol und Tryptophan auswirkt, da die Funktion zumeist deutlich geringerer Werte aufweist. Nach etwa 3 h Prozesszeit erhöht sich die Sensitivität jedoch enorm, da z.B. durch eine größere Serinnebenreaktionskonstante kein Serin mehr für die Hauptreaktion zur Verfügung gestellt wird und die Trajektorien einen anderen Verlauf nehmen. Für die beiden Michaelis-Menten-Konstanten ergibt sich ein sehr ähnliches Bild während des Prozesses. Die Beträge der Zustandssensitivitätsfunktionen zeigen dabei die Reihenfolge der Bestimmbarkeit der Parameter, welche für  $K_m^S$  größer als für  $K_m^I$  ist. Eigentlich sollte die Michaelis-Menten-Konstanten  $K_m^I$  für kleine Indolkonzentrationen wirksam und damit identifizierbar sein, was aber anhand der Zustandssensitivitäten nicht zu erkennen ist. Für den Parameter  $k$  ergeben sich geringfügig größere absolute Werte der Zustandssensitivitätsfunktion, die aber im Vergleich zu den Zustandssensitivitätsfunktionen der Michaelis-Menten-Konstanten einen sehr ähnlichen Verlauf haben. Zu erkennen ist, dass diese drei Parameter bei hohen Indolkonzentrationen, bzw. für hohe Reaktionsgeschwindigkeiten direkt nach den Indolzugaben am besten bestimmbar sind. Die Zustandssensitivitätsfunktionen der Parameter für die Simulation von Serin sind in Abbildung 5.8 dargestellt.

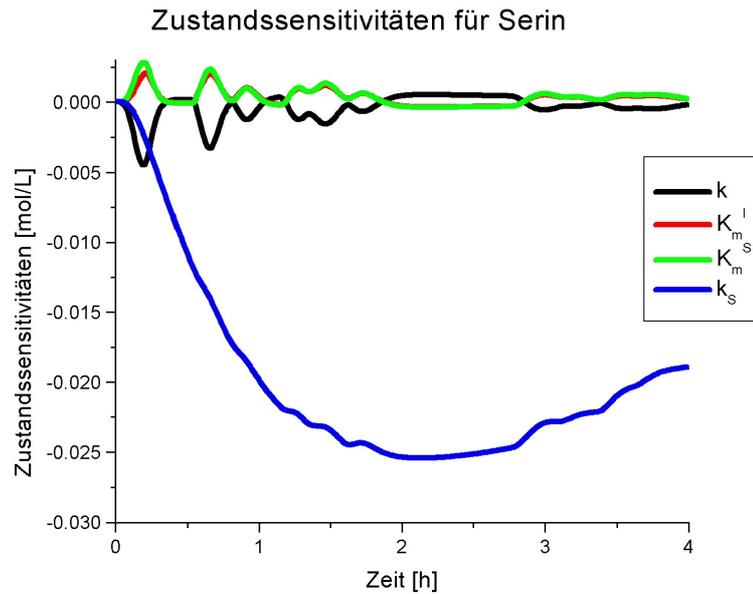


Abbildung 5.8 Zustandssensitivitätsfunktionen der vier Parameter für Serin bei dem Prozess U1

Für die Michaelis-Menten-Konstanten und den Parameter  $k$  ergeben sich auch hier sehr ähnliche Zustandssensitivitätsfunktionen, welche unter Vorzeichenumkehr auch denen für Tryptophan und Indol entsprechen. Die Zustandssensitivitäten zeigen dabei die schon zuvor beobachtete Reihenfolge der Bestimmbarkeit:  $k$  besser als  $K_m^S$  besser als  $K_m^I$ . Die Zustandssensitivitätsfunktion für  $k_S$  zeigt, dass dieser Parameter sich sehr stark auf die Simulation von Serin auswirkt. Der Einfluss von  $k_S$  auf die Simulation von Serin ist darüber hinaus sehr viel größer als die der anderen Parameter bei allen anderen Simulationen, so dass hierdurch die Auswertung über alle Prozessvariablen dahingehend beeinflusst wird, dass dieser Parameter am sensitivsten ist. Die Sensitivitätsanalyse hat ergeben, dass der Parameter  $k_S$  am stärksten die Simulation beeinflusst und daher am besten bestimmbar ist. Die beiden Michaelis-Menten-Konstanten hingegen weisen eine sehr geringe Sensitivität auf die Simulationen auf.

### 5.1.3 Parameterreduktion

Die Ergebnisse der Parameteranpassung und der Sensitivitätsanalyse lassen darauf schließen, dass die beiden Michaelis-Menten-Konstanten und die Geschwindigkeitskonstante der Produktbildung kaum Auswirkungen auf die Simulationen besitzen. Die Parameteranpassung hat ergeben, dass  $K_m^S$  zumeist an der gewählten Obergrenze liegt und die Reaktion somit keiner typischen Enzymkinetik sondern einer Reaktion 1. Ordnung unterliegt. Da die während einer Umsetzung auftretenden Serinkonzentrationen deutlich unterhalb von  $K_m^S$  liegen, ergibt sich eine Kinetik 1. Ordnung bezüglich Serin. Der Parameter  $K_m^I$  weist eine noch geringere Sensitivität auf als  $K_m^S$ . Daher kann das Modell auch um diesen Parameter reduziert werden

ohne die Ergebnisse der Simulationen allzu stark zu beeinflussen. Da  $K_m^I$  zumeist an der Untergrenze angepasst wurde ergibt sich eine Kinetik 0. Ordnung bezüglich Indol. Somit ist die Reaktionsgeschwindigkeit nicht mehr von der Indolkonzentration abhängig. Bei einem vollständigen Indolverbrauch kann jedoch keine Tryptophanbildung mehr stattfinden und die Reaktionsrate der Hauptreaktion muss auf null gesetzt werden. Die Geschwindigkeitskonstante  $k$  weist mit unter 10 % eine sehr geringe Sensitivität auf. Dieser Modellparameter muss jedoch im Modell enthalten bleiben, da eine Umrechnung von der eingesetzten Biomasse auf die Enzymaktivität erfolgen muss. Trotz geringer Sensitivität verbleibt dieser Parameter im Modell. Das reduzierte Modell ist durch die Gleichungen (5.14) – (5.19) gegeben.

$$RV = \begin{cases} I > 0 & k \cdot X \cdot S \\ I = 0 & 0 \end{cases} \quad (5.14)$$

$$\frac{dI}{dt} = \frac{\dot{m}_I}{V} - \frac{\dot{V}_{BFM} + \dot{V}_{Kond}}{V} I - RV \quad (5.15)$$

$$\frac{dS}{dt} = -k_S \cdot X \cdot S - \frac{\dot{V}_{BFM} + \dot{V}_{Kond}}{V} S - RV \quad (5.16)$$

$$\frac{dT}{dt} = -\frac{\dot{V}_{BFM} + \dot{V}_{Kond}}{V} T + RV \quad (5.17)$$

$$\frac{dV}{dt} = \dot{V}_{BFM} + \dot{V}_{kond} \quad (5.18)$$

$$\frac{dX}{dt} = \frac{\dot{V}_{BFM}}{V} X_0 - \frac{\dot{V}_{BFM} + \dot{V}_{Kond}}{V} X \quad (5.19)$$

Wobei  $RV$  die Reaktionsrate der Tryptophanbildung bedeutet. Als Modellparameter verbleiben die beiden kinetischen Konstanten für die Enzymaktivität  $k$  und  $k_S$ . Damit ist das Vorhandensein von Indol, bzw. die Indolzugabe und die Geschwindigkeit der Tryptophanproduktion, abhängig von der Serinkonzentration, für den Prozess bestimmend. Die Tryptophanbildung wird als eine Kinetik 1. Ordnung bezüglich Serin beschrieben, genauso wie die Nebenreaktion. Die Limitierung der Reaktion durch Indol wird jedoch weiterhin berücksichtigt. Weitere kinetische Einflüsse, wie Inhibierung durch Indol, treten in den Hintergrund und können bei den gemessenen Konzentrationen während der beobachteten Prozesse nicht identifiziert werden. Die Einheit der Konstante  $k$  ändert sich entsprechend einer Kinetik 1. Ordnung zu  $[h^{-1}]$ . Entsprechend der oben beschriebenen Parameteridentifikation werden für das reduzierte, zweiparametrische Modell die verbliebenen Modellparameter neu bestimmt. Die Ergebnisse der Parameteranpassung sind in Tabelle 5.4 zusammengefasst.

Tabelle 5.4: Identifizierte Parameter für das zweiparametrische Modell der Prozesse U1 – U6

	U1	U2	U3	U4	U5	U6	U1 – U5
$k$ [ $\text{h}^{-1}$ ]	4,0	8,5	7,7	7,6	13,0	12,7	7,7
$k_s$ [ $\text{h}^{-1}$ ]	1,9	6,7	6,3	5,0	7,3	6,9	4,3
Fehlerquadratsumme [ $\text{mol}^2/\text{L}^2$ ]	0,22	0,20	0,38	0,16	0,24	0,18	3,26

Die Anpassungen für den Parameter  $k_s$  zeigen kaum Unterschiede zu denen des vierparametrischen Modells. Die Anpassungen des Parameters  $k$  hingegen haben sich stark verändert, so haben sich nicht nur die Werte selbst, sondern auch das Verhältnis der einzelnen Anpassungen untereinander geändert. Hierbei fällt auf, dass die Parameter für die maximalen Geschwindigkeiten bei dem zweiparametrischen Modell nun eine gewisse Korrelation aufweisen: Umsetzungen mit hohen maximalen Reaktionsraten zeigen auch eine hohe Nebenreaktionsrate, wie in Abbildung 5.9 vergleichend dargestellt.

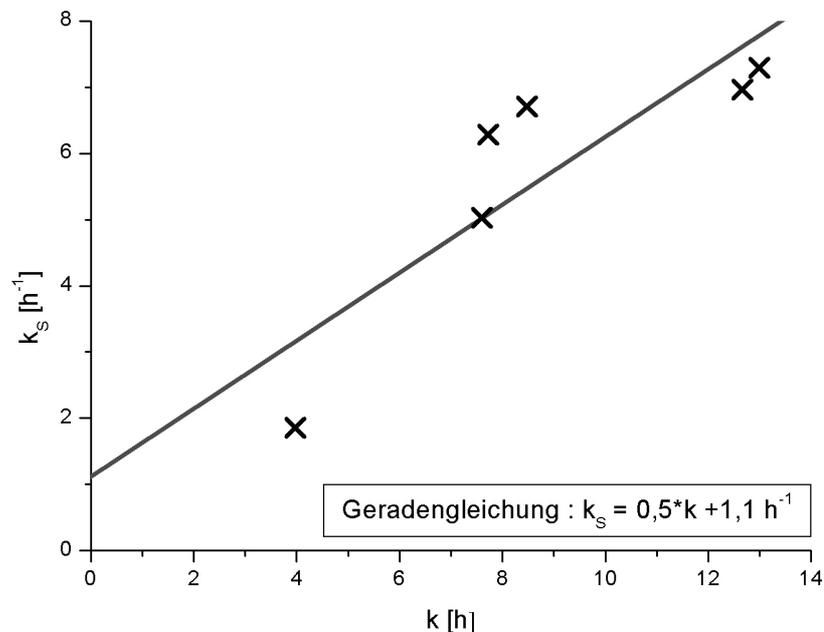


Abbildung 5.9: Korrelation der beiden Parameter für die maximalen Reaktionsgeschwindigkeiten

Aus den dargestellten Werten kann durch lineare Regression der Zusammenhang der beiden Parameter ermittelt werden. Die Geradengleichung weist eine Steigung von 0,5 auf, damit ist die Nebenreaktion etwa halb so groß, wie die Hauptreaktion. Der Achsenabschnitt weist eventuell auf einen weiteren Serinverlust hin, der für jede Umsetzung konstant bleibt und nicht von der Enzymmenge abhängt.

Die Korrelation zeigt, dass beide maximalen Geschwindigkeiten von der bis jetzt nicht zu bestimmenden Enzymmenge in der zugegebenen Biofeuchtmasse abhängen. Dies würde bedeuten, dass eine Erhöhung der Enzymmenge für jede einzelne Umsetzung zwar eine schnellere Tryptophanproduktion, aber auch einen größeren Serinverlust nach sich ziehen würde.

Eine Abschätzung der Aktivität der Enzyme zu Beginn des Prozesses würde die Qualität von Onlinesimulationen erheblich verbessern, da hiermit beide kinetischen Parameter bestimmt werden könnten. Da die Größe der Parameter von den jeweils eingesetzten Zellen abhängt und daher nicht für alle Prozesse gleich ist, sollten die allgemeingültigen Parameterwerte nur benutzt werden, solange, die individuellen Parameterwerte nicht bekannt sind.

Die Anpassungsfehler des zweiparametrischen Modells unterscheiden sich nicht gravierend von denen des vierparametrischen Modells, so dass die Reduktion keine signifikante Verschlechterung der Simulation bedeutet. Ein Vergleich der Simulationen mit der über fünf Prozesse identifizierten Parameterkombination zeigt, wie unterschiedlich die Ergebnisse der beiden vorgestellten Modelle sind. Die Simulationen des vierparametrischen Modells in Abbildung 5.5 und des zweiparametrischen Modells in Abbildung 5.10 zeigen die Unterschiede der Modelle anhand des Prozesses U1 für die Parameterwerte aus den Anpassungen über U1 – U5 auf.

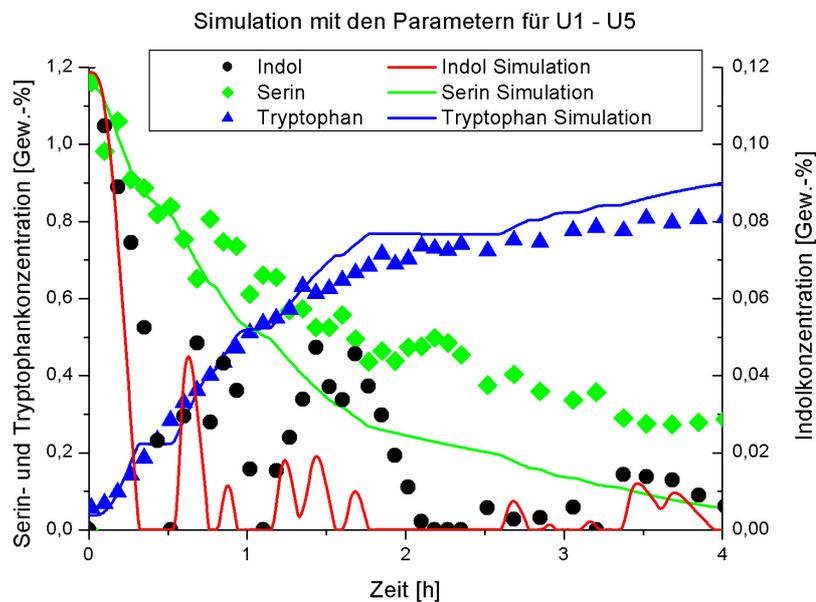


Abbildung 5.10 Simulationen des Prozesses U1 mit dem zweiparametrischen Modell

Die Simulationen von Serin unterscheiden sich kaum, da für beide Modelle die allgemeine Geschwindigkeitskonstante für die Nebenreaktion zu gering sind. An der Simulation für Indol ist im Vergleich der beiden Simulationen zu erkennen, wie die Reaktion 1. Ordnung bezüglich Serin für die Tryptophanbildung sich im Laufe der Umsetzung mit Abnahme der Serinkonzentration auswirkt, da im Fall des zweiparametrischen Modells die Reaktionsgeschwindigkeit mit der Serinkonzentration abnimmt. Die Simulation von Tryptophan zeigt, dass durch die Reduzierung der Parameter die Indolkonzentration nicht mehr auf die Reaktionsgeschwindigkeit wirkt, es sei denn es ist kein Indol mehr in der Reaktionsmasse vorhanden.

Im Vergleich zeigen die Darstellungen außerdem, dass die Parameterreduktion zwar zu einer anderen aber nicht viel schlechteren Simulation führt. Die quadratische Fehlersumme der

Simulationen des Prozesses U1 für die allgemeine Parameterkombination zu den Offline-messwerten hat sich von  $1,30 \text{ mol}^2/\text{L}^2$  beim vierparametrischen Modell auf  $1,48 \text{ mol}^2/\text{L}^2$  beim zweiparametrischen Modell erhöht. Der Test auf Linearität nach Mandel [FUNK 1992] ergibt einen Prüfwert von 4,55 für die beiden verschiedenen Modelle. Aus der F-Verteilung zeigt sich mit einer Irrtumswahrscheinlichkeit von 1 %, dass das vierparametrische Modell die Messwerte nicht signifikant besser beschreibt als das zweiparametrische. Daher sollte das Modell mit weniger Parameter für die Simulation genutzt werden.

Wegen der Veränderungen der Parameterwerte bei der Parameteranpassung, bedingt durch den Einsatz des zweiparametrischen Modells, wird auch für dieses Modell die Sensitivität der Parameter untersucht. In Abbildung 5.11 sind die Parametersensitivitäten für die verbliebenen Parameter dargestellt. Da nur noch zwei Parameter angepasst werden müssen, kann hier die bivariate Fehlerfläche dargestellt werden.

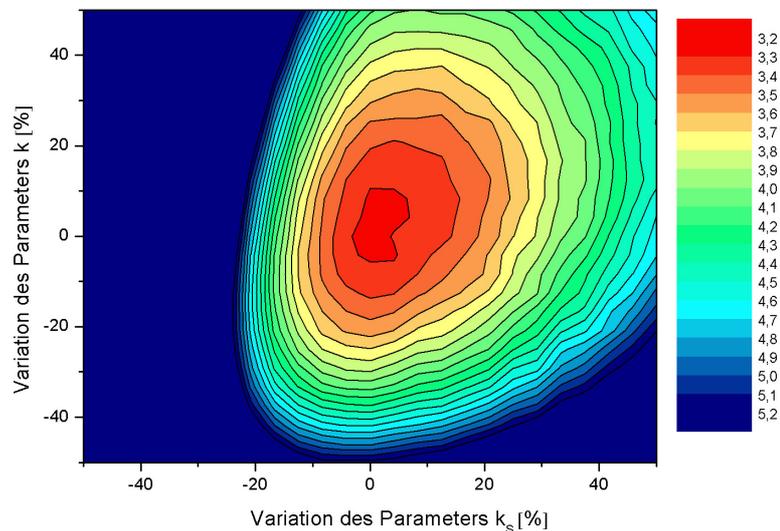


Abbildung 5.11: Bivariate Parametersensitivitäten für das zweiparametrische Modell

Das Optimum ist für die Variation des Parameters  $k$  breiter als für den Parameter  $k_s$ , der somit besser identifizierbar ist. Die bivariaten Parametersensitivitäten zeigen Unterschiede zu den entsprechenden Verläufen bei dem vierparametrischen Modell, die unterschiedliche Breite des Optimums ist nicht mehr so stark ausgeprägt. Dies deutet auf eine größere Sensitivität des Parameters  $k$  hin, als sie noch beim vierparameterigen Modell zu finden war.

Die numerische Berechnung der Sensitivitäten ergibt wiederum eine Sensitivität von 100 % für  $k_s$ . Die Sensitivität für  $k$  ist jedoch auf 43,5 % gestiegen, was die Beobachtung aus der bivariaten Parametersensitivität bestätigt und der beobachteten Relevanz dieses Parameteres für die Simulation entspricht. Somit ist eine weitere Parameterreduktion nicht sinnvoll. Die Zustandssensitivitätsfunktionen des zweiparametrischen Modells zur Analyse der Veränderungen der Sensitivitäten lassen sich nicht berechnen, da die Differenzialgleichung für Indol

nicht mehr stetig ist und daher die partiellen Ableitungen nach der Zustandsvariable Indol nicht mehr gebildet werden können.

Die Sensitivitätsanalyse hat gezeigt, dass beide noch vorhandene Parameter gut bestimmbar sind. Die beobachtete Korrelation der Parameter ist biologisch sinnvoll und sollte für die weiteren Untersuchungen berücksichtigt werden. Das beschriebene Modell und die identifizierten Parameter sollten in der Lage sein den Umsetzungsprozess hinreichend gut zu beschreiben.

#### 5.1.4 Modellvalidierung

Die identifizierten Parameterwerte sollten für die meisten Prozesse eine geeignete Simulation ergeben. Trotz der zum Teil großen Unterschiede zwischen den Umsetzungen haben sich die Parameterwerte aus fünf Prozessen als hinreichend gut erwiesen, um für weitere Umsetzungen oder auch Onlinesimulationen, unter Berücksichtigung der entsprechenden Prozessbedingungen, ein gutes Ergebnis zu erzielen.

Ein Prozess, der nicht für die Parameteridentifikation genutzt wurde, ist die Umsetzung U6 für die die individuell besten Parameter schon in Tabelle 5.4 dargestellt wurden. Die resultierende Simulation ist in Abbildung 5.12 mit den ermittelten Offlinemesswerten dargestellt.

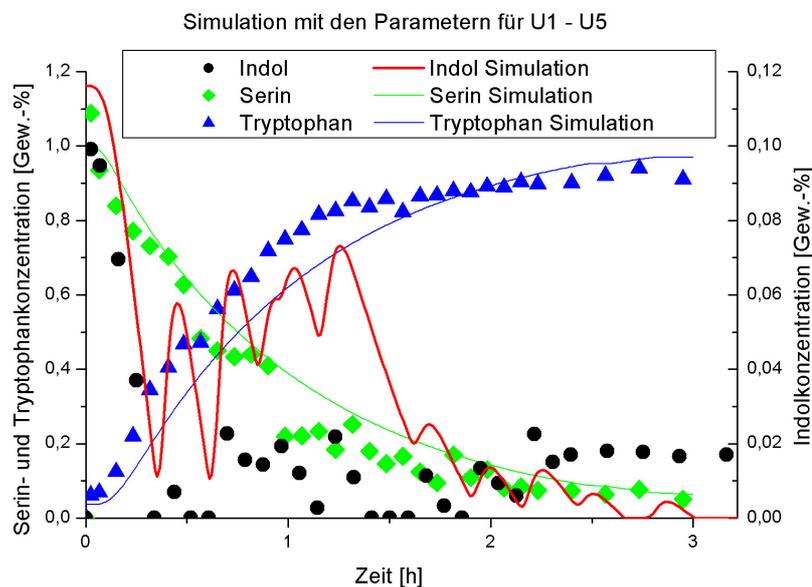


Abbildung 5.12: Simulation des Prozesses U6 mit der identifizierten Parameterkombination mit den ermittelten Offlinemesswerten

Die Simulation des Prozesses U6 zeigt mäßige Übereinstimmungen mit den Offlinemesswerten. Die Reaktionsgeschwindigkeit für die Hauptreaktion scheint anhand der Indol- und Tryptophansimulation etwas zu niedrig, was durch die Ergebnisse der Einzelparameteranpassung in Tabelle 5.4 bestätigt wird. Die Simulation von Serin ist gut gelungen, was auf geringere Unterschiede in der Aktivität der Nebenreaktion schließen lässt. Die Simulation

dieses Prozesses, die nicht für die Parameteridentifikation benutzt wurde zeigt, dass das Modell und die identifizierten Parameterwerten unter Berücksichtigung der Umsetzungsbedingungen eine Aussagekraft zur Beschreibung des Prozesses aufweist.

Die Prozesse U7 – U9 wurden unter veränderten Bedingungen durchgeführt. Bei diesen Umsetzungen wurde versucht durch geringe Serinkonzentrationen die Nebenreaktion nicht zu begünstigen, daher wurde nur ein Teil des Serins vor Prozessstart vorgelegt, und das restliche Serin während des Prozesses hinzugegeben. Wenn das theoretische Prozessmodell die Kinetik der Reaktion korrekt beschreibt, sollten auch diese Umsetzungen hinreichend gut simuliert werden können. Abbildung 5.13 zeigt beispielhaft den Prozess U7, die sich durch eine Serinzugabe nach zwei und vier Stunden Prozesslaufzeit von der anderen Umsetzungen unterscheiden.

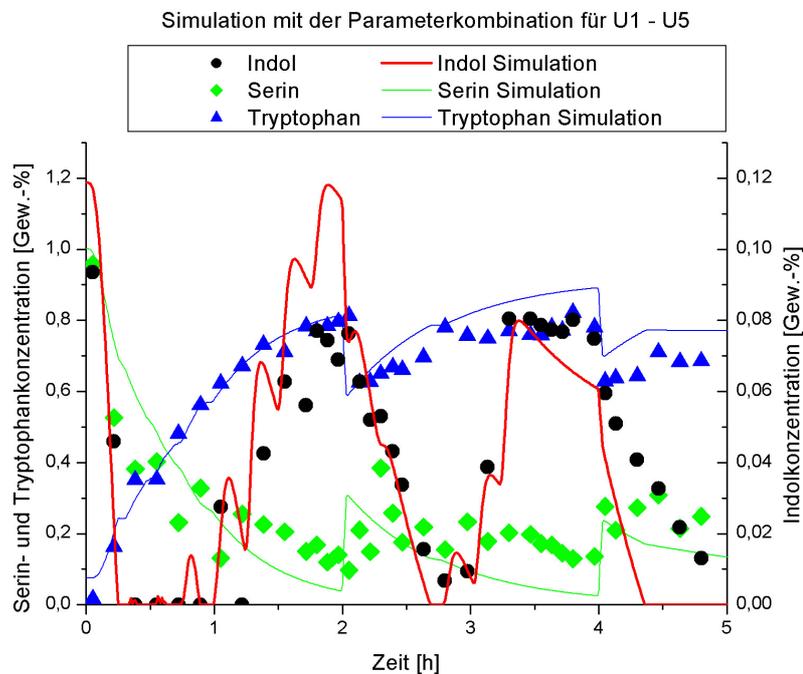


Abbildung 5.13: Simulation des Prozesses U7 mit dem zweiparametrischen Prozessmodell

Die Simulation zeigt, dass die Messwerte der Indolkonzentration recht gut beschrieben werden. Kurz vor den Serinzugaben erhöht sich jeweils die Indolkonzentration, da aufgrund der geringen Serinkonzentration die Reaktionsgeschwindigkeit erheblich herabgesetzt ist. Dieser Effekt wird vom Modell korrekt beschrieben. Abgesehen von der recht starken Streuung der Serinmesswerte, werden auch diese Werte von der Simulation gut beschrieben. Die Simulation zeigt jedoch meist eine zu geringe Serinkonzentration an, was auf einen real niedrigeren Wert für  $k_S$  bei dieser Umsetzung hindeutet. Die Simulation der Tryptophankonzentration liegt nach der ersten Serinzugabe über den Messwerten. Da die Indolmesswerte gut wiedergegeben werden und von einer 100 %-igen Ausbeute in Bezug auf Indol ausgegangen werden kann, kann es sich hier nur um Messfehler oder Volumeneffekte handeln, z.B.

durch einen höheren Volumenstrom des Kondensats. Entsprechende Volumeneffekte sind bei diesen Umsetzungen nicht auszuschließen, da nur abgeschätzt werden konnte, wieviel Serineluat jeweils zugegeben wurde. Wurde zu diesem Zeitpunkt mehr Serineluat zugegeben als für die Simulation angenommen, so wäre der Verdünnungseffekt durch die Zugabe höher und die folgenden Tryptophankonzentrationen niedriger.

Für die Prozesse U7 – U10 aus dem Messzeitraum April 2002 gibt es auch jeweils optimale Parameterwerte, für die die Simulation die Messwerte am besten beschreibt. In Tabelle 5.5 sind die identifizierten Parameter aus den Prozessen U1 – U5 mit der optimalen Parameterwerte für die Prozesse U7 – U10 verglichen.

Tabelle 5.5: Vergleich der allgemeinen und der individuellen Parameterkombination für U7 bis U10

	U1 – U5	U7	U8	U9	U10	U1 – U10
$k [h^{-1}]$	7,7	5,2	4,3	11,0	5,0	7,8
$k_S [h^{-1}]$	4,3	3,5	3,5	11,0	1,9	4,5
Fehlerquadratsumme [ $mol^2/L^2$ ]	3,26	0,68	0,85	0,16	0,58	10,9

Wie in Abbildung 5.13 zu sehen, ist der Wert des Parameters  $k_S$  aus der Anpassung über mehrere Prozesse für die Umsetzung U7 zu groß, der Unterschied für den Parameter  $k$  fällt bei der Simulation nicht so stark auf. Der Fehler bei den Einzelanpassungen liegt zumeist deutlich über den bisher beobachteten Fehlersummen. Die nicht exakt bestimmbar Volumina für die Zugaben des Serineluats und die damit verbundenen Anpassungsfehler sind ein Grund hierfür.

Obwohl die Parameter nur über fünf Prozesse identifiziert wurden, zeigt sich, dass die resultierenden Parameter geeignet sind, den Prozess allgemein zu beschreiben. Eine Parameteranpassung über alle zur Verfügung stehenden Prozesse ergibt geringfügig andere Parameterwerte, wie in Tabelle 5.5 vergleichend dargestellt. Durch Berücksichtigung der Umsetzungen U6 – U10 wird eine weitere Vielfalt an Prozessen (auch unter anderen Umsetzungsbedingungen) für die Bestimmung der Parameter herangezogen.

Die zuvor beobachtete Korrelation der Parameter  $k$  und  $k_S$  wird bei diesen Prozessen nicht überall wieder gefunden. Dies gilt insbesondere für die Prozesse U7 – U9 und könnte an den veränderten Umsetzungsbedingungen liegen. Eigentlich sollte durch die Reaktionsführung mit einer geringeren Serinkonzentration die Nebenreaktion benachteiligt werden, da bis zur Durchführung dieser Umsetzungen die Reaktionsgeschwindigkeit für die Tryptophanbildungsreaktion in Abhängigkeit von der Indol- und Serinkonzentration angenommen wurde. Offensichtlich wird die Nebenreaktion jedoch deutlich begünstigt. Dies lässt auf eine von der Serinkonzentration unabhängige Nebenreaktion schließen. Dies wurde bei der Korrelation der

beiden Parameter in Abbildung 5.9 durch den Achsenabschnitt auch beobachtet, widerspricht jedoch der Beobachtung bei der Identifikation des Serinverlustes in Abbildung 5.1. Zur eindeutigen Klärung sollte die Nebenreaktion und ihre Kinetik näher untersucht werden.

Es konnte gezeigt werden, dass das erstellte Modell geeignet ist, die Biotransformation zur Tryptophanproduktion zu beschreiben; dies gilt auch für Prozesse, die nicht für die Modellerstellung benutzt wurden. Die wahren Parameterwerte werden sich jedoch nur durch weitere gezielte Versuche zur Bestimmung der Tryptophanbildungskinetik und der Enzymaktivität im industriellen Umsetzungsprozess identifizieren lassen. Da die wahren Parameterwerte nur angenähert werden können, erscheint eine Anpassung der Parameter an die aktuellen Reaktionsgeschwindigkeiten während einer Onlineanwendung sinnvoll.

## 5.2 Chemometrik

Zur Beobachtung der Prozessvariablen wurden während der zehn Prozesse U1 – U10 Fluoreszenzspektren aufgenommen. Zur chemometrischen Auswertung werden die online aufgenommenen Fluoreszenzspektren und die offline ermittelten Referenzwerte benutzt. Zur Beschreibung der Spektren und Identifizierung der prozessrelevanten Informationen in den Fluoreszenzspektren wird eine Hauptkomponentenanalyse durchgeführt. Außerdem wird die Erstellung und Anwendung von Modellen zur Vorhersage der Prozessvariablen und anderer Größen aus den Fluoreszenzspektren vorgestellt.

### 5.2.1 Hauptkomponentenanalyse eines Prozesses

Die während eines Prozesses aufgenommenen Fluoreszenzspektren ändern sich im Laufe der Umsetzung. Informationen über die Konzentrationsänderungen von Fluorophoren sind hieraus zu erhalten. Da sowohl Indol als auch Tryptophan fluoreszieren, sollten die Konzentrationsverläufe dieser Prozessvariablen in den Fluoreszenzspektren wieder zu finden sein.

Grafische Darstellungen über alle gemessenen 120 Wellenlängenkombinationen sind kaum zu interpretieren, daher wird zuerst eine Hauptkomponentenanalyse zur Datenreduktion der Fluoreszenzmesswerte eines Prozesses durchgeführt. Die resultierenden Hauptkomponenten, repräsentiert durch die Scores und Loadings, geben die größten, in den Fluoreszenzspektren enthaltenen Veränderungen wieder. Die einzelnen Scores stellen den Verlauf kollinearer Fluoreszenzintensitäten über die Prozesslaufzeit dar. Die Loadings zeigen den Einfluss der einzelnen Wellenlängen für die Berechnung der Hauptkomponenten. Die Darstellung der ersten beiden Scores über die Zeit eines Prozesses in Abbildung 5.14 gibt Aufschluss über die relevantesten Veränderungen der Fluoreszenzintensitäten eines Prozesses.

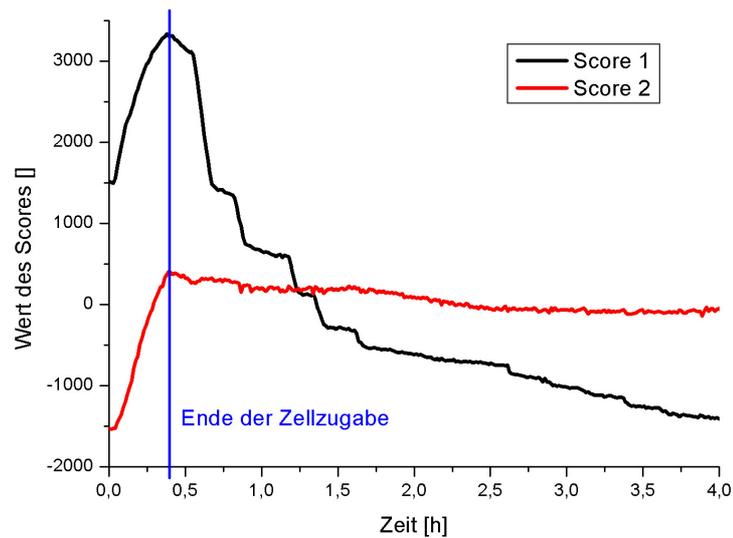


Abbildung 5.14: Darstellung der Scores der ersten beiden Hauptkomponenten der Spektren des Prozesses U1

Mit der ersten Hauptkomponente wird die größte Varianz der Fluoreszenz während eines Prozesses beschrieben. Die größten Änderungen der Fluoreszenzintensitäten werden zu Beginn des Prozesses durch einen Anstieg und nach Ende der Zellzugabe durch eine sukzessive Abnahme im Spektrum hervorgerufen. Die zweite Hauptkomponente beschreibt ausschließlich einen Anstieg zu Beginn der Umsetzung.

Durch welche Wellenlängenkombinationen die Veränderungen in der ersten Hauptkomponente hervorgerufen werden, ist in dem Loadingspektren der Hauptkomponente in Abbildung 5.15 zu erkennen. Ein Loadingspektrum stellt die Werte eines Loadings in Anlehnung an ein 2D-Spektrum dar, so dass anhand dieser Darstellung leicht Vergleiche mit einem Fluoreszenzspektrum selbst hergestellt werden können.

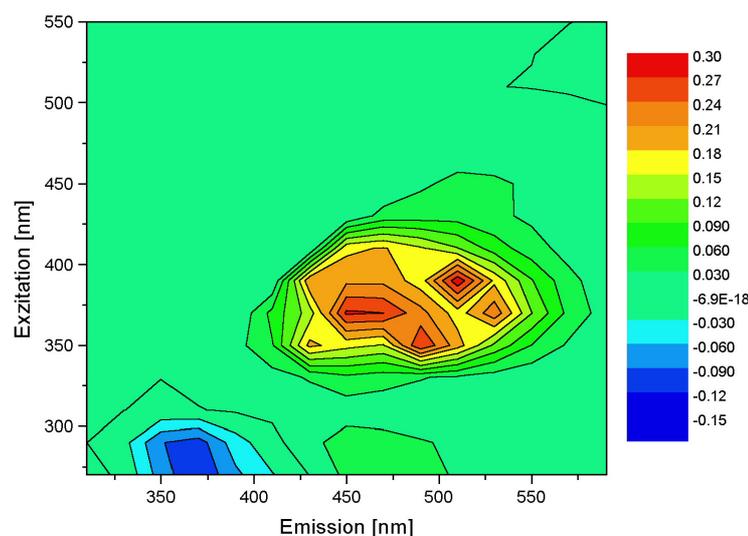


Abbildung 5.15: Loadingspektrum der ersten Hauptkomponente des Prozesses U1

Aus dem Loadingspektrum ist zu entnehmen, dass die erste Hauptkomponente hauptsächlich von den Fluoreszenzänderung bei  $370 - 410 \text{ nm}_{\text{Ex}} / 430 - 530 \text{ nm}_{\text{Em}}$  (Fluoreszenz des Serin-

eluates) und der Tryptophanfluoreszenz im Bereich  $270 - 290 \text{ nm}_{\text{Ex}} / 370 \text{ nm}_{\text{Em}}$  beeinflusst wird. Die Fluoreszenz im Bereich des Serineluates steigt durch die Zellzugabe stark an und nimmt im Laufe der Umsetzung ab. In diesem Bereich muss es ein Fluorophor geben, welches durch die Zellzugabe aufgebaut wird und mit fortschreitender Tryptophanproduktion abgebaut wird. Der Verlauf der ersten Hauptkomponente resultiert aus den Änderungen in diesem Bereich des Spektrums.

Durch welche Wellenlängenkombinationen die Veränderungen in der zweiten Hauptkomponente hervorgerufen werden ist in dem Loadingspektren der Hauptkomponente in Abbildung 5.16 zu erkennen.

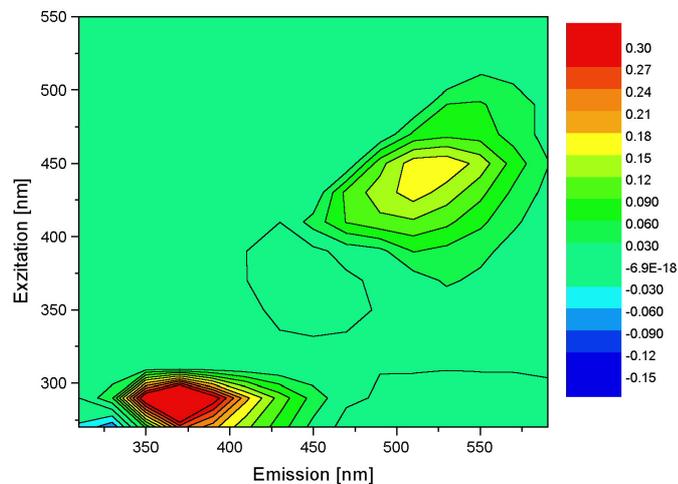


Abbildung 5.16: Loadingspektren der zweiten Hauptkomponente

Die zweite Hauptkomponente wird hauptsächlich von der Tryptophanfluoreszenz im Bereich  $270 - 290 \text{ nm}_{\text{Ex}} / 350 - 390 \text{ nm}_{\text{Em}}$  und untergeordnet von dem zweiten Bereich der Serineluatfluoreszenz im Bereich  $430 - 450 \text{ nm}_{\text{Ex}} / 510 - 530 \text{ nm}_{\text{Em}}$  beeinflusst. Somit beschreibt die zweite Hauptkomponente hauptsächlich die durch die Zellzugabe ausgelöste Fluoreszenzänderungen im Tryptophanbereich ergänzt durch die Fluoreszenz im Bereich des Serineluates bei  $430 \text{ nm}_{\text{Ex}} / 510 \text{ nm}_{\text{Em}}$ . Die Tryptophanfluoreszenz wird durch die Zellzugabe und die hierdurch eingebrachten Proteine beeinflusst, die nachfolgende Produktion von Tryptophan wirkt sich jedoch offenbar nur noch untergeordnet auf die Intensitäten dieser Fluoreszenzwellenlänge aus.

Die Abbildung 5.17 zeigt, dass auch der dritte Score einer Hauptkomponentenanalyse hauptsächlich durch die Fluoreszenzen des Serineluates und des Tryptophanbereichs bestimmt wird und keine neuen Informationen enthält.

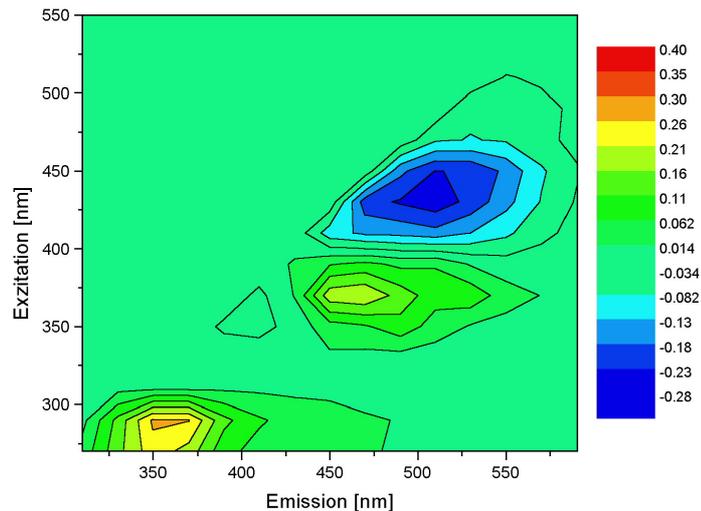


Abbildung 5.17: Loadingspektrum der dritten Hauptkomponente

Bei der Hauptkomponentenanalyse der Spektren eines Prozesses wurde beobachtet, dass keiner der Scores eine Ähnlichkeit zum Verlauf der Indolkonzentration aufweist. Obwohl die Wellenlängenkombinationen  $270 \text{ nm}_{\text{Ex}} / 310 \text{ nm}_{\text{Em}}$  und  $270 \text{ nm}_{\text{Ex}} / 330 \text{ nm}_{\text{Em}}$  eine hohe Korrelation zu dem Verlauf dieser Prozessvariable aufweisen. Die relativ geringe Fluoreszenz dieser Wellenlängen geht im allgemeinen Rauschen der stark fluoreszierenden Wellenlängen mit geringem Informationsgehalt unter. Daher wurde eine Hauptkomponentenanalyse durchgeführt, bei der die Wellenlängenkombination  $270 \text{ nm}_{\text{Ex}} / 310 \text{ nm}_{\text{Em}}$  eine individuelle Wichtung von 20 und die Wellenlängenkombination  $270 \text{ nm}_{\text{Ex}} / 330 \text{ nm}_{\text{Em}}$  eine individuelle Wichtung von zwei erhält, um die relativ geringen Fluoreszenzintensitäten auszugleichen. In Abbildung 5.18 sind die Scores der dritten Hauptkomponente ohne Wichtung der Fluoreszenzwellenlängen und der zweiten Hauptkomponente mit Wichtung der Wellenlängenkombinationen mit dem Verlauf der Fluoreszenzintensitäten der ersten gewichteten Wellenlängenkombination und der Indolkonzentration dargestellt.

Durch die Wichtung der ersten beiden Fluoreszenzwellenlängen enthält der Score der zweiten Hauptkomponente andere prozessrelevante Informationen und lässt sich mit dem Verlauf der Indolkonzentration vergleichen. So beschreibt die zweite Hauptkomponente die Abnahme der Indolkonzentration während der ersten halben Stunde deutlich besser als die Fluoreszenzwellenlänge selbst. Nach gut einer Stunde Prozesslaufzeit zeigen die Fluoreszenz und der Score jedoch einen nicht zu erklärenden Anstieg, der sich nicht in der Indolkonzentration wiederfindet. Die Fluoreszenz der Wellenlänge  $270 \text{ nm}_{\text{Ex}} / 310 \text{ nm}_{\text{Em}}$  verläuft auch nach der zweiten Stunde ähnlich zu der beobachteten Indolkonzentration, der Score hingegen nimmt nicht mehr ab sondern verbleibt fast konstant bei etwa null.

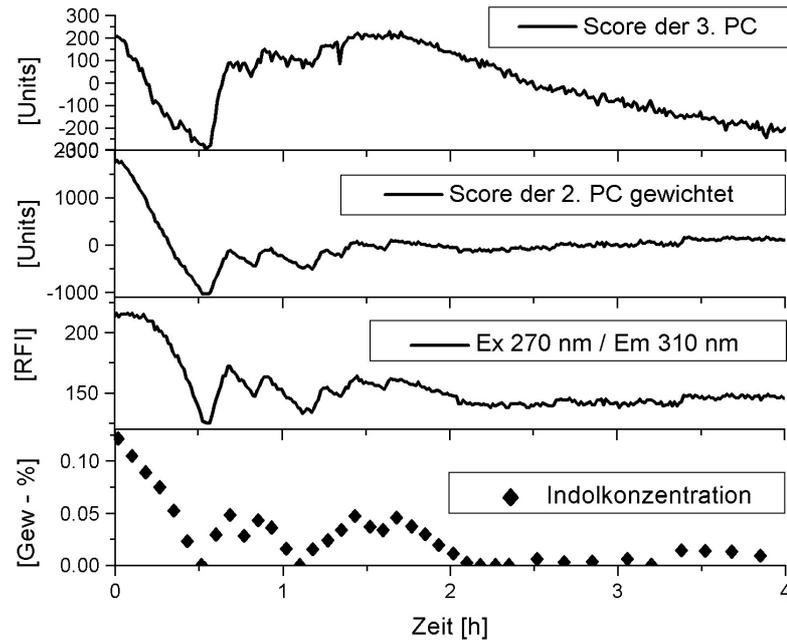


Abbildung 5.18: Vergleich der Scores mit und ohne Wichtung einzelner Wellenlängen

Durch die Wichtung einzelner Wellenlängen verändert sich nicht erst die dritte Hauptkomponente, sondern schon die zweite und enthält damit andere, bisher noch nicht beschriebene Informationen. In Abbildung 5.19 ist das Loadingspektrum der zweiten Hauptkomponente mit Wichtung einzelner Wellenlängen dargestellt. Die erste Hauptkomponente repräsentiert wie zuvor die starken Änderungen der Fluoreszenz des Serineluates.

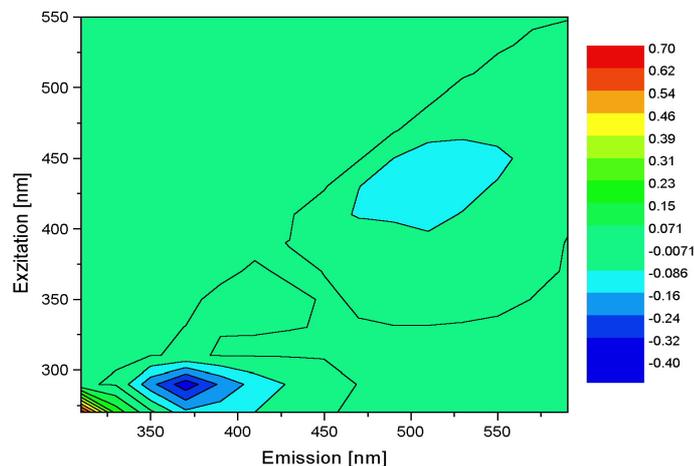


Abbildung 5.19: Loading der 2. Hauptkomponente mit Wichtung

Der Verlauf der zweiten Hauptkomponente wird nun durch die Wellenlängenkombination  $270 \text{ nm}_{\text{Ex}} / 310 \text{ nm}_{\text{Em}}$  und die Tryptophanfluoreszenz bestimmt. Der Verlauf der Fluoreszenzintensitäten dieser Wellenlänge korreliert zum Verlauf der Indolkonzentration. Im Vergleich der beiden Hauptkomponentenanalysen werden durch die Wichtung der Wellenlängenkombinationen unterschiedliche prozessrelevante Informationen beschrieben. Diese Datenvorverarbeitung ist sinnvoll, da hierdurch Prozesswissen in die Analyse einfließt und das Ergebnis

erheblich verbessert wird. Die Hauptkomponentenanalyse der Daten eines Prozesses hat gezeigt, dass die prozessrelevanten Informationen im Fluoreszenzspektrum enthalten sind und damit die Erstellung von Modellen für die Prozessvariablen möglich sein sollte.

### 5.2.2 Hauptkomponentenanalyse mehrerer Prozesse

Die Fluoreszenzmessungen der Prozesse finden innerhalb der komplexen Matrix des Serineluates statt. Je nach Ausgangsmaterial und Aufarbeitung variiert die Reaktionsmasse in seiner Farbe und Zusammensetzung. Insbesondere die Farbigeit des Eluates wirkt sich auf die Fluoreszenzmessung aus. Zur Untersuchung der Unterschiede in den einzelnen Fluoreszenzverläufen der Prozesse werden die Ergebnisse der Hauptkomponentenanalyse über die Spektren mehrerer Prozesse vorgestellt. Die Scores der Hauptkomponenten beschreiben die Varianzen in den Spektren, so dass in dem Scoreplot in Abbildung 5.20 dargestellt werden kann, wie unterschiedlich die einzelnen Prozesse verlaufen und ob die Unterschiede innerhalb eines Prozesses größer sind als zwischen den Prozessen.

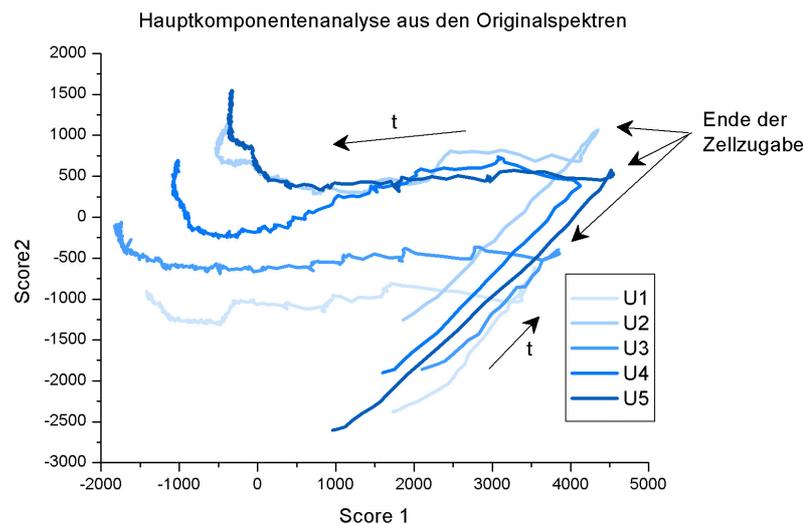


Abbildung 5.20: Darstellung mehrerer Prozesse durch den Scoreplot der Hauptkomponentenanalyse

In dem Scoreplot wurden für die einzelnen Prozesse unterschiedliche Farben ausgewählt. Die Unterschiede, bzw. Varianzen innerhalb eines Prozesses sind erheblich größer, als die Varianzen zwischen den Prozessen. Die einzelnen Prozesse sind im Scoreplot durch eine Parallelverschiebung in Richtung der zweiten Hauptkomponente angeordnet. Die Startpunkte aller Prozesse liegen bei etwa 1000 – 2000 auf der 1. PC und -1000 – -3000 auf der 2. PC. Es sind hier die Unterschiede im Serineluat zu erkennen. Im Allgemeinen zeigen die Verläufe zu Beginn einen starken Anstieg der ersten beiden Hauptkomponenten. Danach sinkt der Score der erste Hauptkomponente sukzessiv ab, wobei der Score der zweiten Hauptkomponente zumeist auf dem selben Niveau verbleibt. Auffallend bei dieser Hauptkomponentenanalyse über die Spektren mehrerer Prozesse ist, dass die Informationen der gewichteten Wellen-

längenkombinationen nicht mehr so stark die zweite Hauptkomponente beeinflussen, wie noch bei der Hauptkomponentenanalyse über die Spektren eines Prozesses. Die Unterschiede zwischen den einzelnen Prozessen überlagern die Varianzen der gewichteten Wellenlängen, so dass diese Informationen erst in späteren Hauptkomponenten beschrieben wird.

Durch die Hauptkomponentenanalyse aus Differenzspektren werden die Unterschiede zu Beginn der Prozesse entfernt, da das Spektrum vor Umsetzung von allen anderen Spektren subtrahiert wird und somit nur die Änderungen der Fluoreszenzen im Spektrum auftreten.

Die Unterschiede zwischen den Prozessen, die durch die Zellzugabe entstehen, können durch die Berechnung aus Differenzspektren nicht ausgeglichen werden. Trotz des einheitlichen Startpunktes der Umsetzungen unterscheiden sie sich nach der Zellzugabe deutlich. Es ist nicht möglich den Startpunkt der Berechnung auf den Zeitpunkt des Endes der Zellzugabe zu legen und damit die Unterschiede, die durch die Zellzugabe eingebracht werden auszugleichen. Der Zeitpunkt des Endes der Zellzugabe ist am industriellen Produktionsprozess nicht eindeutig bestimmbar, außerdem findet nach der Zellzugabe die Tryptophanproduktion mit hoher Geschwindigkeit statt; die hierdurch entstehenden starken Veränderungen im Spektrum machen eine Bestimmung des Nullspektrums ohne große Fehler unmöglich.

Für die spätere Modellbildung aus mehreren Prozessen zur Vorhersage der Prozessvariablen müssen also für gleiche Konzentrationen der Prozessvariablen sehr große Unterschiede zwischen den Umsetzungen berücksichtigt werden. Somit erschwert die Zellzugabe die Modellbildung höchstwahrscheinlich sehr und es müssen erheblich mehr Hauptkomponenten zur Vorhersage der Prozessvariablen berücksichtigt werden.

Mit Hilfe der Loadings der Hauptkomponenten lassen sich die Scores weiterer Prozesse berechnen und mit den in Abbildung 5.20 dargestellten Umsetzungen vergleichen. In Abbildung 5.21 ist der Prozess U6 durch Transformation der Fluoreszenzspektren mit Hilfe der ersten beiden Loadings der Hauptkomponentenanalyse über die Prozesse U1 – U5 in dem Scoreplot der anderen Umsetzungen dargestellt.

Die neu hinzugekommene Umsetzung reiht sich zwischen den Verläufen der Scores der anderen Prozesse ein. Obwohl die Spektren dieser Umsetzung nicht für die Hauptkomponentenanalyse berücksichtigt wurden, zeigen sich durch die Projektion keine gravierenden Unterschiede im Scoreplot, so dass davon ausgegangen werden kann, dass die Prozesse im Rahmen ihrer Variabilität vergleichbar sind. Es sollte also eine Modellbildung und Übertragung auf die dargestellte Umsetzung möglich sein.

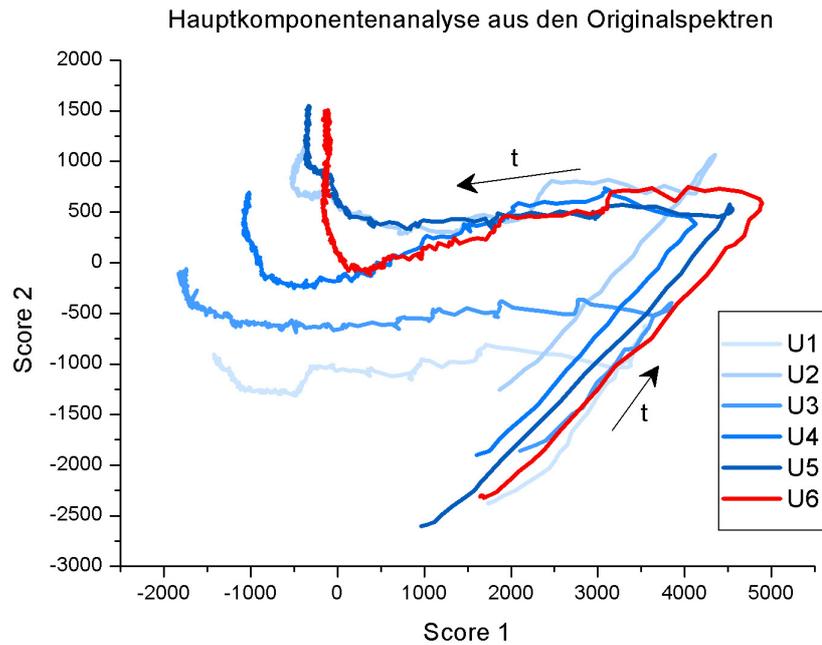


Abbildung 5.21: Vergleichende Darstellung der Scores unter Berücksichtigung eines weiteren Prozesses

Die Betrachtung der Hauptkomponentenanalyse mehrerer Prozesse hat gezeigt, dass die Umsetzungen durchaus Unterschiede aufweisen, aber dennoch vergleichbar sind. Werden Differenzspektren zur Berechnung der Modelle benutzt, so steigt die Vergleichbarkeit der Modelle an, die größten Unterschiede werden jedoch durch die Zellzugabe erzeugt, die nicht aus den Spektren eliminiert werden kann.

### 5.2.3 Erstellung der chemometrischen Modelle

Für die Erstellung chemometrischer Modelle werden die einzelnen Intensitäten der 120 Fluoreszenzwellenlängenkombinationen als unabhängige Variablen verwendet. Die Offline-messwerte der drei Prozessvariablen werden als abhängigen Größen benutzt. Da pro Prozess etwa 40 Offlineproben genommen wurden, werden auch nur die dazugehörigen Spektren für die Modellbildung genutzt.

Aus den Daten der Prozesse U1 – U5 werden für alle drei Prozessvariablen je ein PLS Modell erstellt. Für die interne Validierung der Modelle wird die Kreuzvalidierung benutzt. Hierbei werden jeweils die Daten eines Prozesses nicht für die Modellbildung herangezogen sondern für die Validierung benutzt. Hierdurch wird die Übertragbarkeit der Modelle auf andere Prozesse bei der Validierung berücksichtigt. In Tabelle 5.6 werden die erstellten Modelle für die Prozessvariablen in Abhängigkeit von der Hauptkomponentenanzahl anhand der RMSEPs und der beschriebenen Gesamtvarianz verglichen.

Tabelle 5.6: Vergleich der chemometrischen Modelle anhand der Anzahl an Hauptkomponenten

			1. PC	2. PC	3. PC	4. PC	5. PC
Indol	Kalibration	RMSEP [Gew-%]	0,0205	0,0139	0,0089	0,0086	0,0077
		Varianz [%]	65	88	97	99	99
	Validierung	RMSEP [Gew-%]	0,0214	0,0157	0,0105	<b>0,0091</b>	0,0097
		Varianz [%]	59	83	94	97	98
Serin	Kalibration	RMSEP [Gew-%]	0,141	0,077	0,072	0,066	0,63
		Varianz [%]	73	96	97	99	99
	Validierung	RMSEP [Gew-%]	0,171	<b>0,088</b>	0,102	0,109	0,105
		Varianz [%]	66	94	95	97	98
Tryptophan	Kalibration	RMSEP [Gew-%]	0,097	0,075	0,047	0,041	0,039
		Varianz [%]	75	96	97	99	100
	Validierung	RMSEP [Gew-%]	0,113	0,101	0,073	<b>0,055</b>	0,080
		Varianz [%]	68	94	94	98	99

In Tabelle 5.6 ist zu erkennen, dass bei der Validierung die durch die Hauptkomponenten beschriebene Varianz der Fluoreszenzspektren meist deutlich geringer ist als bei der Kalibrierung. Dies ist auf die schon beschriebenen Unterschiede zwischen den Umsetzungen zurückzuführen, da die Prozesse, die nicht für die Kalibrierung benutzt wurde, jeweils nicht so gut beschrieben werden kann. Für alle drei Modelle werden fast 100 % der Varianz der Spektren spätestens mit der fünften Hauptkomponente beschrieben. Mit weiteren Hauptkomponenten kann keine relevante Information mehr zur Verfügung gestellt werden. Die Festlegung der zu benutzenden Hauptkomponentenanzahl der Modelle erfolgt durch den minimalen RMSEP der Validierung. Für das Modell zur Vorhersage von Indol und Tryptophan sollten vier Hauptkomponenten benutzt werden und für Serin zwei.

Welche Bereiche der Fluoreszenzspektren für die Beschreibung der Konzentrationen sensitiv sind, zeigen die Regressionskoeffizienten der einzelnen Fluoreszenzwellenlängen an. Durch Normierung der Beträge der Regressionskoeffizienten werden in Analogie zu den Loadingspektren der Hauptkomponentenanalyse Sensitivitätsspektren für die PLS-Modelle erstellt. Das Sensitivitätsspektrum des Modells mit vier Hauptkomponenten zur Vorhersage von Indol aus den Fluoreszenzspektren ist in Abbildung 5.22 dargestellt.

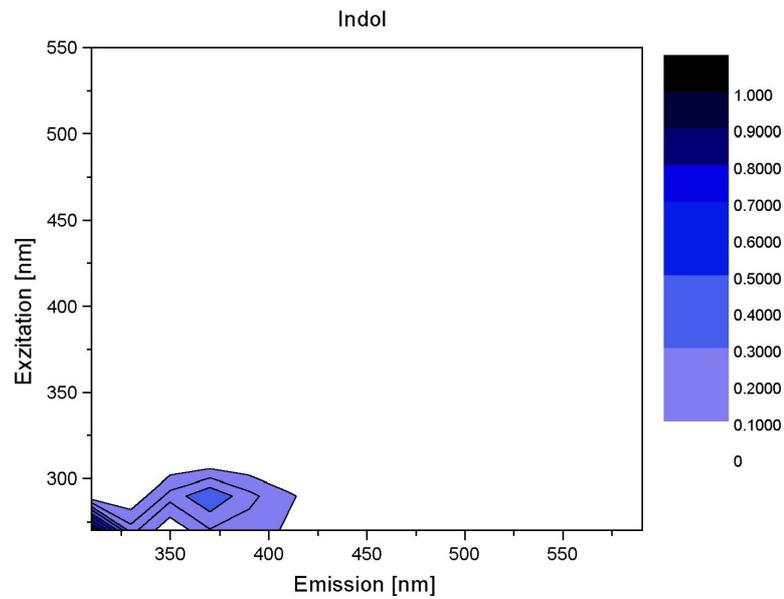


Abbildung 5.22: Sensitivitätsspektrum des Indolmodells

Für die Vorhersage von Indol ist hauptsächlich die erste gewichtete Wellenlänge ( $270 \text{ nm}_{\text{Ex}}$  /  $310 \text{ nm}_{\text{Em}}$ ) verantwortlich und mit untergeordneter Sensitivität der Bereich der Tryptophanfluoreszenz ( $290 \text{ nm}_{\text{Ex}}$  /  $370 \text{ nm}_{\text{Em}}$ ). Alle anderen Bereiche des Spektrum spielen für die Vorhersage von Indol keine Rolle. Damit wird die schon zuvor beobachtete Korrelation zwischen der Wellenlänge  $270 \text{ nm}_{\text{Ex}}$  /  $310 \text{ nm}_{\text{Em}}$  und der Indolkonzentration auch bei der Modellbildung wieder gefunden. Eine lineare Regression ausschließlich auf diese Wellenlänge ist jedoch nicht möglich, da die Intensität der Tryptophanfluoreszenz zusätzlich berücksichtigt werden muss. Das Sensitivitätsspektrum des Modells zur Vorhersage von Serin mit zwei Hauptkomponenten ist in Abbildung 5.23 dargestellt.

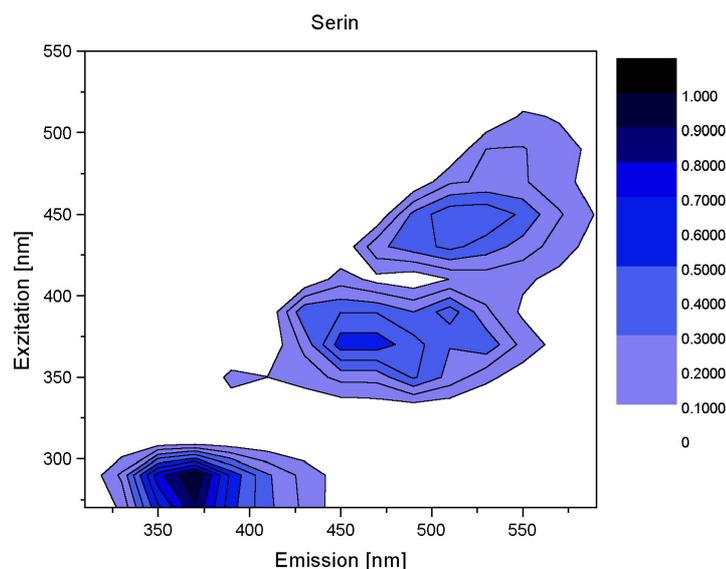


Abbildung 5.23: Sensitivitätsspektrum des Serinmodells

Für das Modell von Serin ist die Tryptophanfluoreszenz und die Fluoreszenz im Bereich des Serineluates am signifikantesten. Die hohe Korrelation der Fluoreszenz des Serineluates mit dem Verlauf der Serinkonzentration wurde schon während der Hauptkomponentenanalyse beobachtet. Obwohl Serin selbst nicht fluoresziert, wird mit dem Eluat offenbar ein Fluorophor eingebracht, welches sich gut mit der Serinkonzentration korrelieren lässt. Die hohe Sensitivität der Tryptophanfluoreszenz ist auf die Korrelation zwischen Tryptophan-zunahme und Serinabnahme zurückzuführen. In Abbildung 5.24 ist das Sensitivitätsspektrum des Tryptophanmodells mit vier Hauptkomponenten dargestellt.

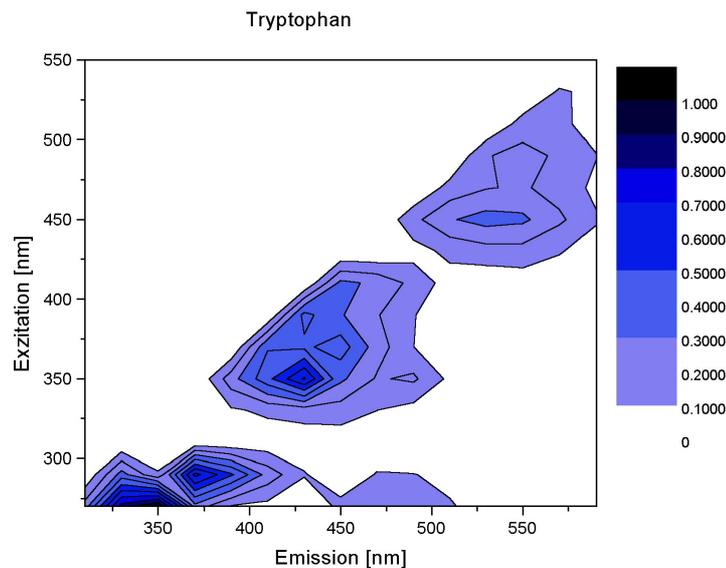


Abbildung 5.24: Sensitivitätsspektrum des Tryptophanmodells

Das Sensitivitätsmaximum des Tryptophanmodells liegt dicht neben dem für Indol und nicht bei der eigentlichen Tryptophanfluoreszenz bei  $290 \text{ nm}_{\text{Ex}} / 370 \text{ nm}_{\text{Em}}$ . Dass die höchste Sensitivität nicht auf der Tryptophanfluoreszenz selbst liegt ist dadurch zu erklären, dass die Fluoreszenz dieser Wellenlänge mit jeder Indolzugabe sofort stark ansteigt. Die Tryptophankonzentration steigt jedoch erst mit dem Verbrauch von Indol an, somit müssen andere Bereiche des Spektrums für die Vorhersage genutzt werden. Für die Vorhersage der Tryptophankonzentration werden außerdem weitere Bereiche des Spektrums genutzt als bei dem Modell für Indol. Dies unterstreicht die komplexen Zusammenhänge für die Vorhersage von Tryptophan aus den Fluoreszenzspektren.

Anhand der Sensitivitätsspektren ist zu erkennen, dass für alle drei Prozessvariablen unterschiedliche Bereiche des Spektrums für die Vorhersage berücksichtigt werden müssen. Für alle drei Modelle sind die als sensitiv identifizierten Spektrenbereiche sinnvoll.

Neben den Prozessvariablen Indol, Tryptophan und Serin sind noch Informationen über die Biomasse, bzw. die Proteinmenge im Spektrum enthalten, wie bei der Hauptkomponentenanalyse zu erkennen war. Die Änderung der Fluoreszenzintensität der Wellenlänge  $350 \text{ nm}_{\text{Ex}}$

/ 470 nm<sub>Em</sub> durch die Zellzugabe ist bei den Prozessen verschieden. Diese Unterschiede können nicht auf die Biomasse zurückgeführt werden, da die Menge an zugegebener Biofeuchtmasse für die Prozesse fast identisch ist. Somit muss die in der Biofeuchtmasse enthaltene Protein- oder Enzymmenge hierfür verantwortlich sein. Durch vorherige Separation der Zellen, kann die Konzentration der Biomasse in der Biofeuchtmasse unterschiedlich sein. Es gibt jedoch keine Offlinemesswerte der Proteinmenge, die zur Modellbildung genutzt werden könnten. Für die Geschwindigkeiten, die abhängig sind von der Enzymmenge, wurden jedoch ebenfalls unterschiedliche Werte bei den einzelnen Umsetzungen gefunden, wie in Kapitel 5.1.3 durch die Parameteranpassung beschrieben. Da es sich bei den Geschwindigkeitsparametern um Größen handelt, die während des Prozesses konstant bleiben, kann keine Regression mit den sich ändernden Intensitäten der Fluoreszenzspektren durchgeführt werden. Es muss aus den Spektren für jede Umsetzung ein aussagekräftiger Wert gefunden werden, der das Maß der zugegebenen Enzymmenge repräsentiert. Da die Fluoreszenz der Wellenlänge 350 nm<sub>Ex</sub> / 470 nm<sub>Em</sub> mit der Biomassezugabe ansteigt und im Anschluss daran wieder abfällt, sollte die maximale Änderung der Fluoreszenzintensität zu der zugegebenen Enzymmenge in Relation stehen. Dieser Wert ist in Abbildung 5.25 gegen die Geschwindigkeitskonstanten  $k$  der Prozesse U1 – U5 dargestellt. Die Abbildung zeigt eine hohe Korrelation zwischen der maximalen Differenz der Fluoreszenzintensität und dem Wert der Geschwindigkeitskonstanten  $k$ . Damit sind die verschiedenen Parameterwerte bei der Anpassung in Kapitel 5.1.3 nicht zufällig sondern zeigen prozessspezifische Information und lassen sich durch die Fluoreszenzmessungen ebenso wieder finden.

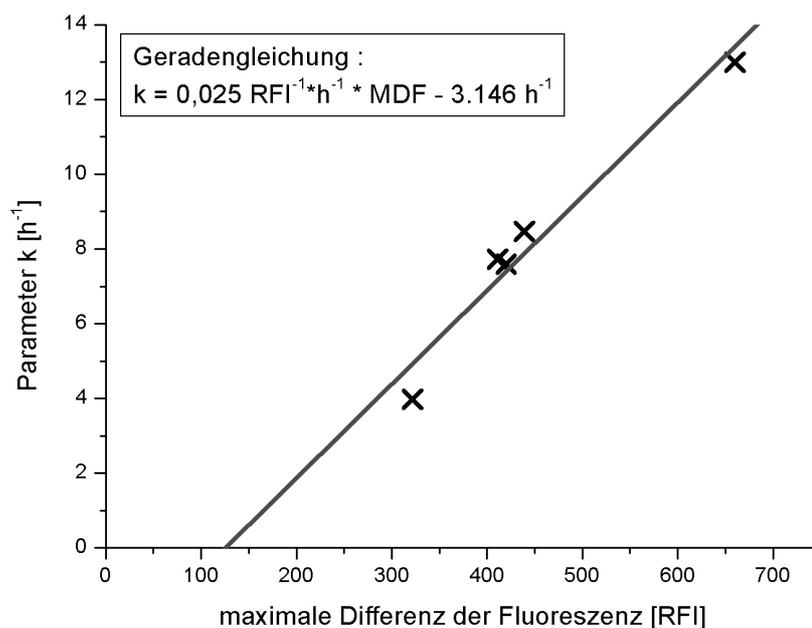


Abbildung 5.25: Korrelation der maximalen Differenz der Fluoreszenzintensität 350 nm<sub>Ex</sub> / 470 nm<sub>Em</sub> zum Geschwindigkeitsparameter  $k$  der Hauptreaktion der Prozesse U1 – U5

Durch lineare Regression ist aus der maximalen Differenz der Fluoreszenzintensität der Wellenlänge 350 nm<sub>Ex</sub> / 470 nm<sub>Em</sub> der Parameter  $k$  abschätzbar. Die aus den dargestellten Werten der Prozesse U1 – U5 bestimmte Geradengleichung ist im Diagramm angegeben.

Es konnte gezeigt werden, dass prozessrelevante Informationen in den Spektren enthalten sind und durch die PLS-Regressionen mit den Konzentrationsverläufen der Prozessvariablen korreliert werden können. Die erstellten chemometrischen Modelle beziehen sich auf unterschiedliche Spektralbereiche für die drei Prozessvariablen und geben die beobachteten Korrelationen gut wieder. Die Modellerstellung und die Präzision der Vorhersagen ist stark von der Unterschiedlichkeit der Umsetzungen abhängig. Insbesondere die Zellzugabe hat einen großen Einfluss auf die Fluoreszenzänderungen, wodurch die Modelle zur Vorhersage der Prozessvariablen an Komplexität zunehmen. Die Fluoreszenzänderungen durch die Zellzugabe weisen jedoch eine hohe Korrelation zu den identifizierten Geschwindigkeitskonstanten der Prozesse auf, so dass die relative Enzymaktivität über die Fluoreszenzmessung zugänglich ist.

#### 5.2.4 Anwendung der chemometrischen Modelle

Für die externe Validierung der Modelle werden die Messdaten des Prozesses U6 herangezogen und die Prozessvariablen Indol, Serin und Tryptophan vorhergesagt. Die Vorhersage aus allen Spektren des Prozesses U6 sind den Offlinemesswerten der Prozessvariablen in Abbildung 5.26 gegenübergestellt.

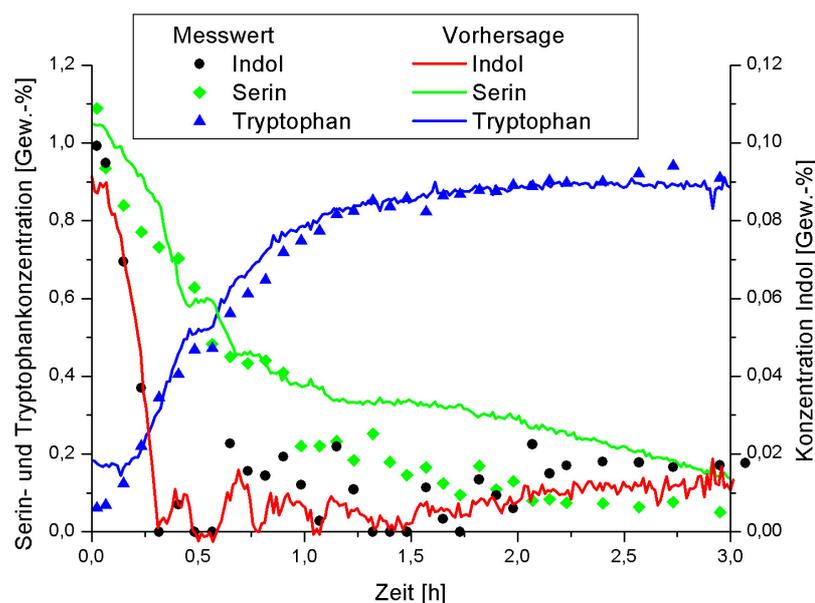


Abbildung 5.26: Offlinemesswerte des Prozesses U6 und Vorhersagen der Prozessvariablen mit PLS Modellen aus den Prozessdaten von U1 – U5

Die Vorhersagen des Prozesses U6 beschreiben den Verlauf der Offlinemesswerte gut. Insbesondere die Vorhersage von Tryptophan gelingt sehr gut. Die Vorhersage für Indol zeigt die

Änderungen der Konzentration durch die Indolzugaben gut, allerdings ist diese Vorhersage stark verrauscht und die Qualität nimmt mit der Prozesszeit deutlich ab. Die Vorhersage für Serin gelingt nicht besonders gut. Der starke Abfall der Serinkonzentration bei etwa 1 h wird von der Vorhersage nicht beschrieben. Die Serinnebenreaktion und damit der Serinverlust kann aus den Fluoreszenzspektren offenbar nicht hinreichend beschrieben werden. Diese unterschiedlichen Qualitäten der Vorhersagen werden durch die in Tabelle 5.7 angegebenen Fehler bestätigt. Die Vorhersagegüte der Modelle wird durch den RMSEP zu den Offline-messwerten der Prozessvariablen und den relativen Fehler in Bezug auf den Maximalwert angegeben.

Tabelle 5.7: Fehler für die Vorhersage der Prozessvariablen aus den Fluoreszenzspektren

	absoluter Fehler	relativer Fehler
Indol	0,0067 Gew.-%	6,74 %
Serin	0,135 Gew.-%	12,4 %
Tryptophan	0,043 Gew.-%	4,53 %

Nur die Vorhersage von Tryptophan liegt im Bereich von 5 % Genauigkeit. Dieser geringe Fehler bestätigt die Beobachtung einer guten Vorhersage von Tryptophan aus Abbildung 5.26. Wobei der Fehler der ersten drei Messungen überdurchschnittlich groß ist. Während der Zellzugabe, bzw. der ersten 20 Prozessminuten, sollte der Vorhersage von Tryptophan also nicht viel Gewicht gegeben werden. Die geringe Genauigkeit der Serinvorhersage ist für eine Onlineanwendung tolerierbar, da die Konzentration von Serin nicht direkt für die Optimierung des Prozesses benötigt wird. Allerdings wäre eine Detektion des Serinverbrauchs und damit der vollständige Umsatz des Substrates von Vorteil. Die Genauigkeit der Vorhersage von Indol ist wohl für eine direkte Regelung basierend auf den Vorhersagen nicht geeignet. Eine Verbesserung der Indolvorhersage ist unbedingt notwendig für eine optimale Führung des Prozesses. Durch Verwendung von N-way-PLS-Modellen lassen sich die Vorhersagefehler reduzieren. Die Fehler sind entsprechend in Tabelle 5.8 dargestellt.

Tabelle 5.8: Fehler für die Vorhersage der Prozessvariablen aus den Fluoreszenzspektren

	absoluter Fehler	relativer Fehler
Indol	0,0056 Gew.-%	5,64 %
Serin	0,147 Gew.-%	13,5 %
Tryptophan	0,042 Gew.-%	4,50 %

Die Vorhersagefehler für Indol und Tryptophan sind gesunken, jedoch auch hier reicht die Genauigkeit der Indolvorhersage nicht für eine Regelung aus. Der Fehler für die Vorhersage

von Serin ist leider angestiegen. In der Abbildung 5.27 sind die Vorhersagen der N-way-PLS mit den Offlinemesswerten abgebildet und geben Aufschluss über die Veränderungen zu den PLS-Modellen in Abbildung 5.26.

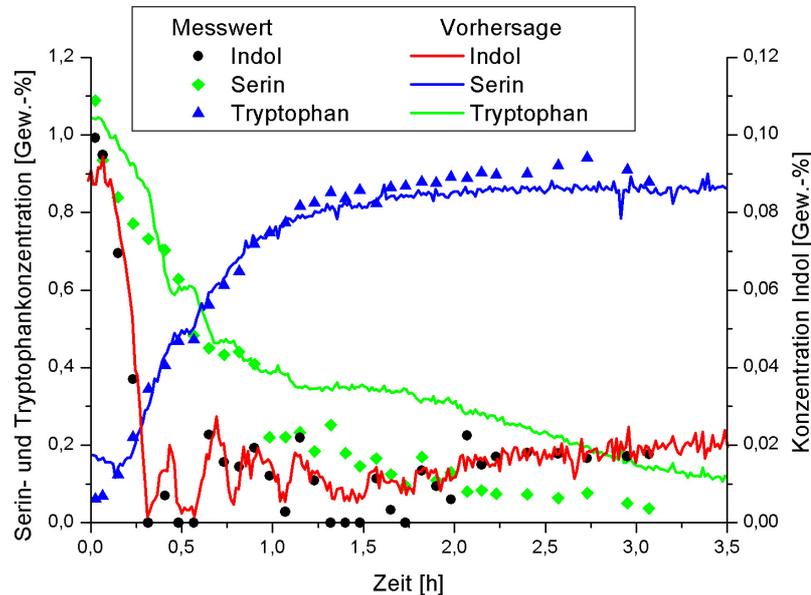


Abbildung 5.27: Offlinemesswerte des Prozesses U6 und Vorhersagen der Prozessvariablen mit N-way-Modellen aus den Prozessdaten von U1 – U5

Die Änderungen der Indolkonzentrationen durch die Indolzugaben werden von diesem Modell gut beschrieben, was die Reduzierung des Fehlers für diese Vorhersage erklärt. Leider wird ein vollständiger Verbrauch von Indol während der ersten 1,5 h nicht korrekt wieder gegeben. Die Vorhersage von Tryptophan zeigt einen ähnlich guten Verlauf wie die Vorhersage des PLS-Modells, ist jedoch im allgemeinen geringer, was hier die Reduzierung des Fehlers erklärt. Im Gegensatz dazu liegt die Vorhersage von Serin etwas oberhalb von der Vorhersage aus dem PLS-Modell. Auch hier ist die Veränderung des Fehlers auf die Verschiebung der Vorhersage zurückzuführen, die jedoch eine Verschlechterung mit sich bringt. Aus den Fluoreszenzspektren des Prozesses U6 kann die zugegebene Enzymmenge über die Bestimmung der maximalen Differenz der Fluoreszenzintensität der Wellenlänge  $350 \text{ nm}_{\text{Ex}} / 470 \text{ nm}_{\text{Em}}$  ermittelt werden. Aus dieser Berechnung kann auf den Wert der Geschwindigkeitskonstanten über die Ergebnisse der linearen Regression geschlossen werden. Bei Bestimmung der maximalen Fluoreszenzdifferenz der Wellenlänge  $350 \text{ nm}_{\text{Ex}} / 470 \text{ nm}_{\text{Em}}$  hat sich ein Wert von 590 RFI ergeben, dies entspricht einem  $k$  von  $11,6 \text{ h}^{-1}$ . Nach der beobachteten Korrelation der beiden Geschwindigkeitsparameter zueinander ergibt sich für die Nebenreaktion eine Geschwindigkeitskonstante von  $6,9 \text{ h}^{-1}$ . Durch die Parameteranpassung haben sich optimale Werte von  $12,7 \text{ h}^{-1}$  für  $k$  und  $6,9 \text{ h}^{-1}$  für  $k_s$  ergeben. Die Parameterwerte, die aus den Prozessen U1 – U5 berechnet wurden, liegen bei  $7,66 \text{ h}^{-1}$  und  $4,4 \text{ h}^{-1}$ . Damit liegen die aus den Fluoreszenzspektren berechneten Parameter dicht bei den individuell für diesen Prozess

besten Parameterwerten und unterscheiden sich deutlich von den Parameterwerten, die über mehrere Prozesse bestimmt wurden. Durch die Abschätzung der Parameter aus den Fluoreszenzspektren kann der Prozess deutlich besser simuliert werden, als mit den Parametern, die über verschiedene Prozesse identifiziert wurden.

Es konnte gezeigt werden, dass die chemometrischen Modelle die Prozessvariablen von anderen Prozessen mit unterschiedlicher Qualität beschreiben. Die Vorhersagen für die Prozessvariablen Indol und Tryptophan gelingen gut. Die Vorhersage für Serin ist suboptimal. Da Serin selbst nicht fluoresziert war eine Vorhersage dieser Variablen eigentlich nicht zu erwarten. Außerdem konnte gezeigt werden, dass die Parameter der Geschwindigkeitskonstanten aus den Fluoreszenzspektren abgeleitet werden können.

### **5.3 Onlineanwendung der Modelle**

Die vorgestellten theoretischen und chemometrischen Modelle zur Simulation der Prozesse und Vorhersage der Prozessvariablen aus den Fluoreszenzspektren wurden am industriellen Prozess online eingesetzt. Die Ergebnisse dieser Onlineanwendungen werden in diesem Kapitel vorgestellt. Die Onlineanwendung der Modelle fand im April 2002 an dem Prozess U10 statt, somit wurde für die folgenden Darstellungen ein Prozess aus einem anderen Umsetzungszeitraum benutzt. Dies gibt auch Aufschluß über die Langzeitstabilität der Modelle. Im Vergleich zu den Umsetzungen U7 – U9 wurde dieser Prozess vergleichbar zu den Umsetzungen U1 – U6 durchgeführt und repräsentiert eine typische Biotransformation.

#### **5.3.1 Die Onlinesimulation**

Für die Onlinesimulation wird das Bioprozessmodell eingesetzt. Unter Berücksichtigung der aktuellen Umsetzungsbedingungen, wie Abschätzung der zugegebenen Biofeuchtmasse, Abschätzung der Serinstartkonzentration aus vorangegangenen Umsetzungen und der jeweiligen Indolzugaben kann eine Simulation in Echtzeit parallel zum Prozess durchgeführt werden. In Abbildung 5.28 ist das Ergebnis der Simulation mit der allgemeinen Parameterkombination mit den ermittelten Offlinemesswerten dargestellt.

Die Simulation des Prozesses U10 beschreibt den Verlauf der Messwerte gut. Die Anfangsreaktionsgeschwindigkeit scheint zu hoch zu sein, wie an der simulierten Indolkonzentration zu erkennen ist; im weiteren Verlauf des Prozesses ist dieser Effekt nicht mehr zu erkennen. Die Simulation von Tryptophan liegt deutlich über den Offlinemesswerten. Die Bilanz der Reaktion wird in der Simulation berücksichtigt und die Indolkonzentration wird während dieser Prozesszeit korrekt wiedergegeben. Somit kann die Differenz zwischen Messwert und Simulation nur durch Volumeneffekte, bzw. eine größere Verdünnung erklärt

werden. Nach Ende der Umsetzung lag das Gesamtvolumen etwa 10 % über dem durch die Simulation berechneten Volumen. Die Simulation von Serin zeigt, dass die Geschwindigkeitskonstante für die Nebenreaktion als zu groß für diese Umsetzung angenommen wurde. Die Quadratische Fehlersumme der Simulation liegt bei  $1,88 \text{ mol}^2/\text{L}^2$ .

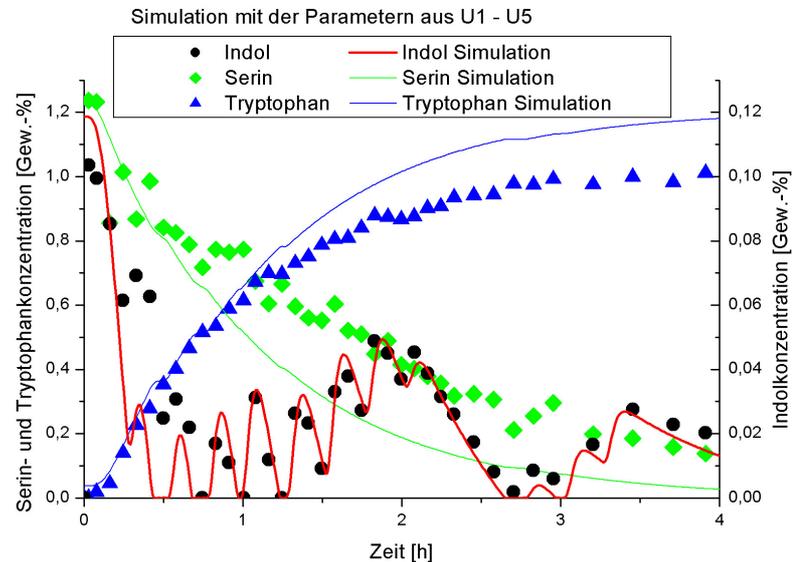


Abbildung 5.28: Onlinesimulation des Prozesses U10 sowie die ermittelten Offlinemesswerte

Durch Abschätzung der kinetischen Parameter aus den Fluoreszenzspektren haben sich Parameterwerte von  $k = 4,8 \text{ h}^{-1}$  und  $k_s = 3,5 \text{ h}^{-1}$  ergeben. Bei der Simulation mit diesen Parameterwerten ergibt sich ein Simulationsfehler von  $1,42 \text{ mol}^2/\text{L}^2$  und damit eine deutliche Verbesserung. Die Abschätzung der Parameter aus den Fluoreszenzspektren ist erst nach der Zellzugabe also etwa 20 Minuten nach Prozessstart möglich. Darauf aufbauend kann die weitere Simulation des Prozesses in Echtzeit durchgeführt werden. Die aus der Parameteranpassung bestimmten ideale Parameterwerte liegt bei  $k = 5,0 \text{ h}^{-1}$  und  $k_s = 1,9 \text{ h}^{-1}$  und einem Simulationsfehler von  $0,58 \text{ mol}^2/\text{L}^2$ . Diese nochmals erhebliche Reduzierung des Fehlers wird insbesondere durch den deutlich geringeren Wert für die Nebenreaktionsgeschwindigkeit und einer damit verbundenen besseren Vorhersage der Serinkonzentration hervorgerufen.

Damit hat sich gezeigt, dass das Prozessmodell auch nach längerer Zeit noch sehr leistungsfähig ist und seine Gültigkeit nicht verliert. Die Parameter sind immer noch recht gut aus den Fluoreszenzspektren abschätzbar und erzeugen eine Simulation des Prozesses von höherer Qualität als eine Simulation mit den über mehrere Prozesse bestimmten Parameterwerten. Insbesondere die Simulation der Indolkonzentration mit den abgeschätzten Parametern kann als Grundlage einer Regelung eingesetzt werden.

### 5.3.2 Die Onlinevorhersage

Während des Prozesses U10 wurden online und intime Fluoreszenzspektren aufgenommen. Die Fluoreszenzspektren wurden online mit Hilfe der chemometrischen Modelle verarbeitet, so dass eine Vorhersage der Prozessvariablen Indol, Serin und Tryptophan intime am industriellen Biotransformationsprozess erfolgte. In Abbildung 5.29 sind die Vorhersagen der Modelle mit den ermittelten Offlinemesswerten dargestellt.

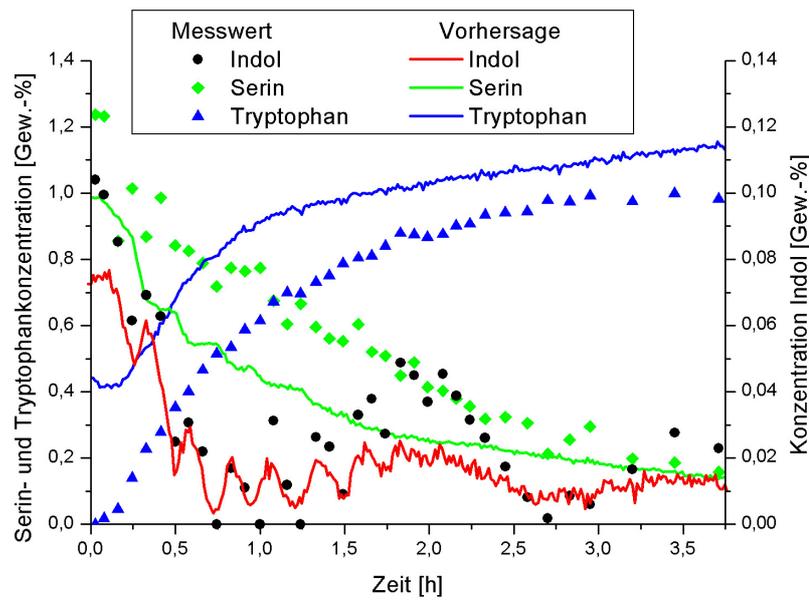


Abbildung 5.29: Offlinemesswerte des Prozesses U10 und Vorhersagen der Prozessvariablen mit PLS Modellen aus den Prozessdaten von U1 – U5

Die Vorhersage des Prozesses U10 gelingt nicht besonders gut. Bei der Vorhersage von Tryptophan zeigt sich eine deutliche Verschiebung um etwa 0,3 Gew.-%, so dass die Vorhersage von Tryptophan während des Prozesses etwa 33 % zu hoch ausfällt. Die Indolkonzentration zeigt den Verlauf, bedingt durch die Indolzugabe recht gut, jedoch stimmen die vorhergesagten Werte nicht überein. Zu Beginn des Prozesses liegen die Vorhersagen unterhalb der Messwerte und der gemessene Indolanstieg bei 2 h wird von den Vorhersagen nicht wiedergegeben. Die Serinkonzentrationen der Vorhersage liegen während des gesamten Prozesses unter den ermittelten Offlinemesswerten.

Ein Grund für die Verschiebungen der Vorhersagen könnte der Zeitunterschied sein. Während der Zeit zwischen Modellerstellung und Anwendung hat sich eventuell das Serineluat derart verändert, dass die Modelle keine gültigen Vorhersagen mehr liefern. Eine Verbesserung der Vorhersagen sollte dann durch die Benutzung von Differenzspektren für die Vorhersage erreicht werden, da die Unterschiede, die mit dem Eluat eingebracht werden, eliminiert werden. Die sehr gravierenden Fehler der Vorhersage aus den Originalspektren werden in Tabelle 5.9 den Fehler der Vorhersage aus Differenzspektren gegenüber gestellt.

Tabelle 5.9: Fehler für die Vorhersage der Prozessvariablen der Prozesse U10 mit PLS Modellen aus den Prozessdaten von U1 – U5

Prozessvariable	Originalspektren		Differenzspektren	
	absoluter Fehler	relativer Fehler	absoluter Fehler	relativer Fehler
Indol	0,0162 Gew.-%	15,62 %	0,0176 Gew.-%	16,96 %
Serin	0,2208 Gew.-%	17,85 %	0,1959 Gew.-%	15,84 %
Tryptophan	0,2604 Gew.-%	26,63 %	0,1050 Gew.-%	10,75 %

Die Vorhersagen der drei Prozessvariablen aus den Spektren sind, entsprechend dem Verlauf der Vorhersagen in Abbildung 5.29, sehr groß. Der Vergleich zu der Vorhersage aus Differenzspektren zeigt eine deutliche Verbesserung. Durch die Vorhersage aus Differenzspektren wird die Verschiebung, die durch die Veränderung des Serineluates hervorgerufen wird, offenbar kompensiert. Auch für die Vorhersage aus Differenzspektren sind die Fehler für Tryptophan jedoch noch zu hoch. Für Indol wird die Vorhersage aus Differenzspektren schlechter, hier wurde in Abbildung 5.29 auch keine erhebliche Verschiebung gegenüber den Messwerten beobachtet. Die Vorhersage von Serin aus Differenzspektren gelingt nur geringfügig besser. Während der externen Validierung in Kapitel 5.2.4 konnte gezeigt werden, dass die N-way-Modelle geringere Vorhersagefehler liefern als die PLS-Modelle. Daher werden die Ergebnisse auch für diese Modelle in Tabelle 5.10 für die Vorhersagen aus den Spektren und den Differenzspektren vorgestellt.

Tabelle 5.10: Fehler für die Vorhersage der Prozessvariablen des Prozesses U10 mit N-way-Modellen aus den Prozessdaten von U1 – U5

Prozessvariable	Originalspektren		Differenzspektren	
	absoluter Fehler	relativer Fehler	absoluter Fehler	relativer Fehler
Indol	0,0136 Gew.-%	13,1 %	0,0178 Gew.-%	17,2 %
Serin	0,2025 Gew.-%	16,4 %	0,0858 Gew.-%	6,9 %
Tryptophan	0,2431 Gew.-%	24,3 %	0,1167 Gew.-%	11,7 %

Die Fehler für alle drei Prozessvariablen der Vorhersagen mit N-way-Modellen aus den Fluoreszenzspektren sind auch für diese Umsetzung geringer als die Vorhersagen mit PLS-Modellen. Die Vorhersagefehler sind jedoch auch für diese Modelle viel zu hoch. Für Tryptophan ergibt sich auch hier eine deutliche Verbesserung beim Einsatz von Differenzspektren, doch der Fehler liegt immer noch zu hoch. Für Serin bedeutet die Vorhersage aus Differenzspektren mit N-way-Modellen eine enorme Verbesserung, wobei der relative Vorhersagefehler erstmalig unter 10 % liegt. Auch bei der Anwendung von N-way-Modellen auf die Differenzspektren vergrößert sich der Fehler für die Vorhersage von Indol gegenüber

der Berechnung aus den Originalspektren. Da sich die Vorhersagen zwischen der Benutzung der verschiedenen Modelle nicht erheblich unterscheiden wird die Anwendung von N-way-Modellen auf Differenzspektren in Abbildung 5.30 vorgestellt. Die Abbildung zeigt die Veränderungen gegenüber den Vorhersagen aus den Originalspektren mit PLS-Modellen in Abbildung 5.29.

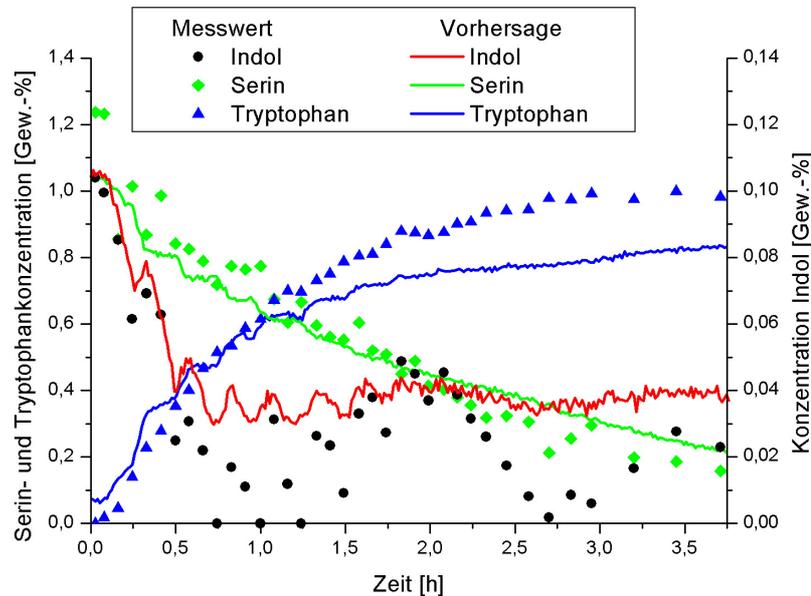


Abbildung 5.30: Offlinemesswerte des Prozesses U10 und Vorhersagen der Prozessvariablen mit N-way-Modellen für Differenzspektren aus den Prozessdaten von U1 – U5

Der Verlauf der Vorhersagen zeigt, wie erwartet, für alle drei Prozessvariablen eine Verschiebung gegenüber der Vorhersage aus den Originalspektren. Die Werte der Vorhersagen für Indol und Serin liegen höher, was für Serin den geringeren Fehler und für Indol den höheren Fehler erklärt. Die Vorhersagen von Tryptophan sind zu niedrigeren Werten verschoben, was hier die Verbesserung der Vorhersage bedingt.

Für Serin wird eine gute Vorhersage für den Einsatz der N-way-Modelle und Differenzspektren erreicht. Es ist jedoch auch an der Serinvorhersage zu erkennen, dass die Vorhersagefehler nicht nur durch eine Parallelverschiebung hervorgerufen werden. Die Vorhersagen sind auch in ihrer Änderung zu gering, so beginnt die Vorhersage von Tryptophan zutreffend, erreicht jedoch ab 1,5 h nicht mehr die richtigen Werte und liegt bei 3,5 h etwa 0,02 Gew.-% unterhalb der Messwerte. Die Serinvorhersage beginnt mit zu niedrigen Werten und endet mit zu hohen Werten. Die Vorhersage von Indol beginnt bei der zutreffenden Konzentration von 1,05 Gew.-%, erreicht jedoch nie eine Konzentration von unter 0,3 Gew.-%.

Die Stauchung der Vorhersagen kann nicht auf die unterschiedliche Fluoreszenz des Serineluates zurück geführt werden, da diese Unterschiede durch die Bildung von Differenzspektren aus der Vorhersage entfernt werden. Die Abweichungen der Vorhersagen könnten nur auf unterschiedliches Quenchen der Fluoreszenzintensitäten während des Prozesses

zurückgeführt werden. Falls die Änderung des Serineluates für die schlechten Vorhersagen die Ursache ist, so müsste für eine korrekte Vorhersage ein gleichbleibendes Serineluat zur Verfügung gestellt werden. Dies ist jedoch prozesstechnisch nicht realisierbar, da es sich um einen natürlichen Rohstoff handelt. Eine gleichbleibende Qualität der Melasse kann nicht garantiert werden. Daher müssten möglichst viele unterschiedliche Serineluate für die Modellbildung genutzt werden, was langwierig und mit einem enorm hohen Messaufwand verbunden ist. Auch eine immer wieder erneute Modellerstellung für den aktuellen Zustand des Serineluates ist aufgrund des hohen Messaufwandes nicht möglich.

Filtereffekte sind nicht nur auf das Serineluat selbst zurückzuführen, sondern können auch vom Aufbau der Messung bedingt werden. Die Fluoreszenzmessung wird in einer Messzelle durchgeführt, die in dem Bypass installiert ist. Nach jeder Umsetzung wird die Bypasspumpe ausgestellt und die Reaktionsmasse verbleibt mehrere Stunden bis zur nächsten Umsetzung im Bypass und in der Messzelle. Insbesondere durch das restliche, noch vorhandene Indol werden Verfärbungen hervorgerufen, die sich auf der Glasscheibe absetzen können und die Messung beeinflussen. Aber auch die Reaktionsmasse selbst altert während dieser Zeit und es fallen Bestandteile aus der Lösung aus. Da die Messzelle vor der Beobachtung des Prozesses U10, im Gegensatz zu den anderen dargestellten Umsetzungen, mehrere Wochen im Bypass installiert war, können derartige Effekte ebenso wie die Änderung des Serineluates für die schlechten Vorhersagen verantwortlich sein. Falls der Messaufbau die Messung in der beschriebenen Weise beeinflusst, sollte eine regelmäßige Reinigung der Messzelle Abhilfe schaffen. Eine direkte Messung der Fluoreszenz im Reaktor würde hier keine Änderung bringen, da auch an der Reaktorwand Ablagerungen auftreten können.

### **5.3.3 Filterung der Onlinevorhersagen**

Für eine hinreichend gute Vorhersage der Prozessvariablen aus den Fluoreszenzspektren, wie es bei dem Prozess U6 in Kapitel 5.2.4 vorgestellt wurde, kann durch die Filterung der Daten das Messrauschen reduziert werden. Da ein leistungsstarkes Bioprozessmodell für den Prozess zur Verfügung steht, kann eine Filterung der Vorhersagen durch den Kalman-Filter implementiert werden. Bei dieser Methode werden die Messungen durch Berücksichtigung von Prozesswissen aus dem Bioprozessmodell gefiltert. Die Vorhersagen der Prozessvariablen des Prozesses U6 sind in Abbildung 5.31 mit Simulationswerten dargestellt. Die Filterung der Vorhersagen mit Hilfe des Prozessmodells durch den Kalman-Filter sind ebenso, wie die Offlinemesswerte des Prozesses im Diagramm aufgeführt.

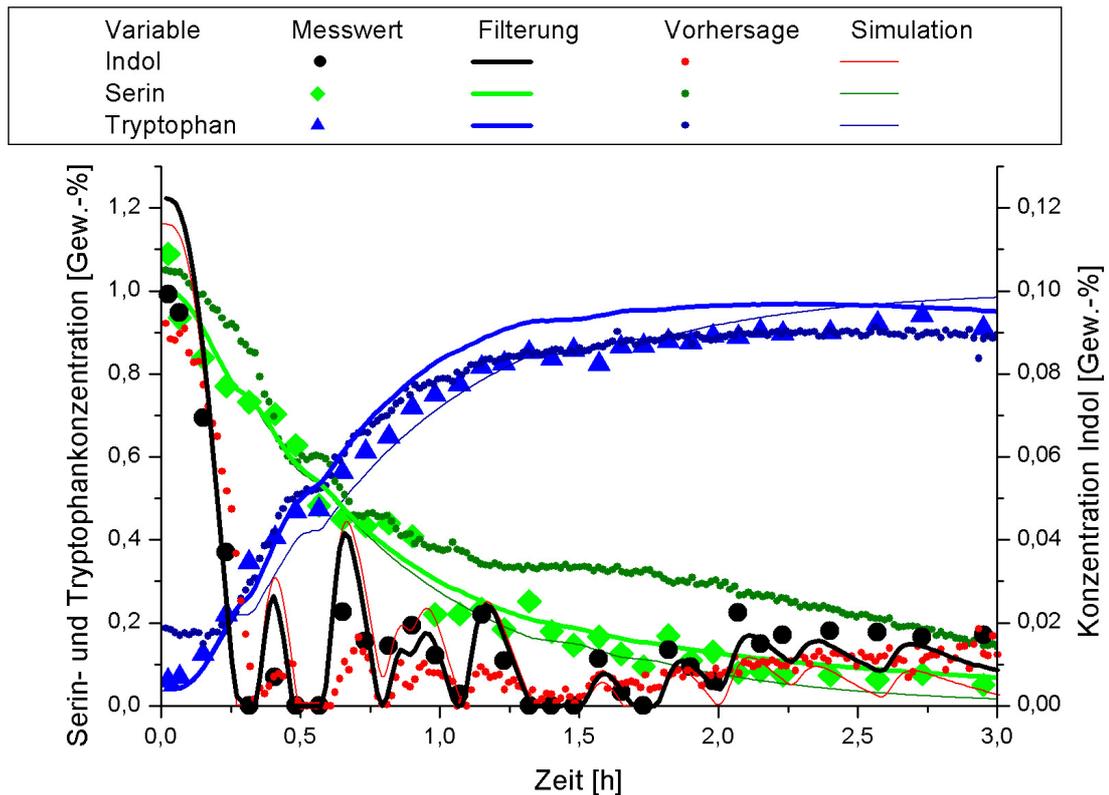


Abbildung 5.31: Vorhersagen, Simulation, Filterung und Offlinemesswerte der Prozessvariablen

Die Abbildung zeigt, dass durch die Filterung der Vorhersagen der geschätzte Verlauf der Prozessvariablen ruhiger ist und kein Messrauschen mehr zeigt. Je nach Wahl der Parameter des Kalmanfilters (siehe Anhang 8.4) verlaufen die Schätzungen dichter an der Simulation oder dichter an der Vorhersage. So wurde für die Serinschätzung ein großes Messwertrauschen angenommen, da die Vorhersage von Serin mit einem großen Fehler behaftet ist. Hierdurch wird die Schätzung von Serin gegenüber der Vorhersage deutlich verbessert. Für die Indolschätzung wurde ebenfalls mehr Gewicht auf die Simulation gelegt, um die Bilanz der Reaktion nicht durch die Filterung zu verfälschen. Da für die Vorhersage von Tryptophan ein geringer Fehler bestimmt wurde, wurde für die Schätzung dieser Variable mehr Gewicht auf die Messung gelegt. Die Fehler der Schätzung nach der Filterung der Vorhersagen sind in Tabelle 5.11 mit den Fehlern der Vorhersagen dargestellt.

Tabelle 5.11: Vergleich der Fehler von Vorhersagen und Schätzungen

	Vorhersagefehler		Fehler nach der Filterung	
	Absolut [Gew.-%]	Relativ [%]	Absolut [Gew.-%]	Relativ [%]
Indol	0,006	6,1	0,003	2,7
Serin	0,134	12,3	0,062	5,7
Tryptophan	0,047	5,0	0,078	8,1

Anhand der Fehler ist zu erkennen, dass durch Filterung der Vorhersagen zum Teil deutliche Verbesserungen für die Schätzung der Prozessvariablen erreicht werden. Der geringe Fehler der Tryptophanvorhersage wird durch die Filterung geringfügig erhöht. Dieser Fehler für die Schätzung der Tryptophankonzentration wird jedoch zu Gunsten der verbesserten Vorhersage der Serin- und Indolkonzentration in Kauf genommen. Für die Schätzung der Serinkonzentration ist die Filterung von enormen Vorteil, ebenso wie für die Schätzung von Indol.

Durch die Filterung der Vorhersagen sind die Schätzungen der drei Prozessvariablen für eine Regelung des Prozesses geeignet. Die hervorragende Schätzung von Indol, welche Grundlage einer Regelung ist, kann für eine optimale Führung durch das Bedienpersonal oder den Aufbau einer Automatisierung des Prozesses herangezogen werden.

## 5.4 Optimierung

Der Biotransformationsprozess zur Tryptophanherstellung bei der AMINO GmbH soll im Rahmen dieser Arbeit unter ökologischen und ökonomischen Gesichtspunkten optimiert werden. Hierfür soll die Produktion von Tryptophan unter optimaler Ressourcenausnutzung gesteigert werden. Neben der Ausbeute an Tryptophan in Bezug auf den Einsatz des Substrates Serin ist ebenso die Prozessdauer und damit einhergehend die Einsparung von Energie das Ziel der Optimierung. Die Ausbeute an Tryptophan bezogen auf Serin liegt derzeit nicht bei 100 %, das andere Substrat Indol wird unter den gegebenen Umsetzungsbedingungen jedoch nahezu zu 100 % zu Tryptophan umgesetzt.

Serin könnte ebenso zu 100 % umgesetzt werden, wenn keine Serinnebenreaktion stattfinden würde. Je nach Beschaffenheit der eingesetzten Biomasse wurden unterschiedliche Nebenreaktionsaktivitäten beobachtet. Eine gezielte Untersuchung der Aktivität wäre also wünschenswert. Durch gezielte Mutationen am Mikroorganismus könnte die Nebenreaktionsaktivität ebenfalls unterdrückt werden, diese Vorgehensweise ist jedoch nicht Bestandteil dieser Arbeit und würde eine gezielte Untersuchung der Nebenreaktion voraussetzen.

Im Rahmen dieser Arbeit soll unter den derzeitigen Umsetzungsbedingungen eine Optimierung der Tryptophanproduktion durch Änderung der Prozessführung erreicht werden. Mit Hilfe des Bioprozessmodells kann der Umsetzungsprozess simuliert werden und damit der Verlust des Substrates identifiziert werden. Außerdem kann eine Prozessführung erarbeitet werden, die den Substratverlust minimiert. Durch die vorgestellte Onlinemessung des Prozesses ist es möglich, den aktuellen Prozesszustand durch die Messung und Auswertung zu bestimmen und hierauf aufbauend die Optimierung am Prozess durchzuführen.

Das Bioprozessmodell zeigt eine gleichzeitige Haupt- und Nebenreaktion in Abhängigkeit von der Serinkonzentration. Somit kann die Nebenreaktion nicht vollständig unterdrückt

werden. Die Nebenreaktion findet bei den hier beobachteten Prozessen meist mit etwas mehr als der halben Geschwindigkeit der Hauptreaktion statt, so dass bei Vorhandensein von Serin ein entsprechender Teil des Serins verloren geht. Die Tryptophanbildungsreaktion ist nicht von der Indolkonzentration abhängig, so dass durch Variation dieser Konzentration nicht die Haupt- oder Nebenreaktion begünstigt werden kann. Falls kein Indol in der Reaktionsmasse mehr vorhanden ist, findet allerdings ausschließlich die Nebenreaktion statt, da das zweite Substrat für die Tryptophanbildung fehlt. In folgedessen sind die Prozesszustände während einer Umsetzung zu vermeiden, bei denen kein Indol im System vorhanden ist. Bei ständiger Indolversorgung und ausreichender Enzymaktivität sollte eine erhöhte Produktausbeute bezüglich des Substrates Serin erreichbar sein.

Derzeit sind die Zeitpunkte eines vollständigen Indolverbrauchs am Prozess nicht identifizierbar. Durch das hier vorgestellte Mess- und Auswertesystem, sind diese Zustände jedoch unmittelbar detektierbar und können durch eine weitere Indolzugabe eliminiert werden.

Durch Eliminierung der Prozesszustände ohne Indol hätten die Serinverluste während der Prozesse U1 – U10 verringert werden können. Die minimierten Serinverluste sind in Tabelle 5.12 den am Prozess beobachteten Serinverlusten gegenüber gestellt.

Tabelle 5.12: Vergleich der Serinverluste mit und ohne Optimierung der Indolzugabe

	Parameter		Messwerte		Optimierung	
	k [h <sup>-1</sup> ]	k <sub>S</sub> [h <sup>-1</sup> ]	Verlust <sup>2</sup> [%]	Ausbeute <sup>2</sup> [%]	Verlust <sup>2</sup> [%]	Ausbeute <sup>2</sup> [%]
U1	4,0	1,9	-15,1	90,7	-16,2	123,3
U2	8,5	6,7	3,5	104,4	-0,4	118,8
U3	7,7	6,3	4,2	95,5	0,2	113,5
U4	7,6	5,0	-1,2	105,5	-4,2	124,5
U5	13,0	7,3	-2,6	111,3	-7,0	134,4
U6	12,7	6,9	-0,9	112,5	-7,9	135,6
U7	5,2	3,5	6,4	96,8	-10,7	106,6
U8	4,3	3,5	12,0	77,5	-3,8	107,1
U9 <sup>3</sup>	11,0	11,0	10,1	65,3	20,0	60,4
U10	5,0	1,9	-6,2	105,7	-17,8	135,2
Mittelwert			0,00	100,00	-7,53	122,11

<sup>2</sup> Es werden relative Verluste und Ausbeuten bezogen auf die mittleren Ausbeuten und Verluste angegeben. Die Werte der Optimierung geben die Änderungen gegenüber der mittleren Ausbeute und dem mittleren Verlust an.

<sup>3</sup> Da es sich bei dem Prozess U9 aufgrund der Prozessführung offenbar um einen nicht vergleichbaren Prozess handelt, wird dieser nicht für die Mittelwertberechnung berücksichtigt.

Die Höhe der theoretisch möglichen Ausbeute ist von den Werten für die beiden Reaktionsraten abhängig. Die Verluste liegen bei hohen Reaktionsgeschwindigkeiten (U5 - U6) im Bereich der Mittelwerte und die Ausbeuten oberhalb der Mittelwerte. Wie an den Werten von U9 zu erkennen ist, spielt jedoch auch das Verhältnis zwischen  $k$  und  $k_S$  eine Rolle, da hier für beide derselbe Werte identifiziert wurde, ist die Ausbeute an Tryptophan sehr gering. Bei dem Prozess U10 hingegen ist die Reaktionsgeschwindigkeit nicht besonders hoch, dafür die Nebenreaktionsgeschwindigkeiten signifikant niedriger im Verhältnis zur Hauptreaktion. Insbesondere für diese Umsetzungen sind bei geeigneter Prozessregelung Verbesserungen zu erwarten.

Die Verluste über neun Prozesse zeigen, dass durch eine optimale Regelung des Prozesses der mittlere Verlust um 7,53 % gesenkt werden kann. Die Ausbeuten können um 22 % verbessert werden. Diese Steigerung der Ausbeute ist bei optimaler Regelung zudem in der Hälfte der Prozesszeit zu erreichen, so dass die Raum-Zeit-Ausbeute von derzeit etwa 0,1 Gew.-%<sub>TRP</sub>/h auf 0,25 Gew.-%<sub>TRP</sub>/h gesteigert werden kann.

Durch die dargestellten Messungen, Auswertungen und Analysen der Umsetzungsprozesse ist mit dem vorgestellten Messsystem eine Identifikation des Prozesszustandes möglich. Basierend auf diesem aktuellen Prozesszustand ist weiterhin eine Regelung der Indolzugabe und damit eine optimale Führung des Prozesses möglich. Allein hierdurch kann die Raum-Zeit-Ausbeute signifikant gesteigert werden. Durch die Prozessbeobachtung hat sich jedoch gezeigt, dass 4 h Prozesszeit für eine enzymatische Reaktion des Substrates zum Produkt bei hinreichender Indolversorgung völlig ausreichend sind. Somit könnte die Prozesszeit und damit der Energieverbrauch zum Antrieb des Rührers, Beheizten des Kessels und der Indolleitung eingespart werden, ebenso könnten doppelt so viele Umsetzungen pro Zeit durchgeführt werden. Der weiterhin vorhandene Serinverlust ist auf eine Zellaktivität zurückzuführen, die nicht durch die Regelung des Prozesses unterbunden werden kann. Diese Nebenaktivität der Zellen könnte durch den Einsatz von Inhibitoren oder eines genetisch veränderten Stammes unterdrückt werden. Wobei ein derartiger Eingriff in den industriellen Prozess unerwünschte Nebeneffekte haben könnte.

## 6 Zusammenfassung und Ausblick

L-Tryptophan wird bei der AMINO GmbH durch die Biotransformation von L-Serin und Indol mit dem Enzym Tryptophansynthase aus *Escherichia coli* Zellen hergestellt. Derzeit wird der Prozess durch manuelle Regelung der Indolzugabe abhängig von den Ergebnissen einer Online-HPLC-Analytik geführt. Ziel dieser Arbeit war es, Monitoring und Control der biotechnischen Tryptophanproduktion zu verbessern.

Dafür wurden zehn verschiedene Prozesse aus zwei unterschiedlichen Umsetzungszeiträumen mit 2D-Fluoreszenzspektroskopie und durch Offlinemessungen begleitet. In den entnommenen Proben wurden die Konzentrationen der beiden Substrate Indol und L-Serin und des Produktes L-Tryptophan bestimmt.

Die Fluoreszenzmessungen sollten mit Hilfe chemometrischer Modelle ausgewertet und somit der aktuelle Prozesszustand online ermittelt werden. Die Hauptkomponentenanalyse der Fluoreszenzmesswerte zeigte, dass relevante Informationen über die Konzentrationen der Prozessvariablen in den Spektren vorhanden sind. Somit wurden aus den 2D-Fluoreszenzspektren von fünf Prozessen mit den dazugehörigen Offlinemesswerten für die Konzentrationen von Indol, L-Serin und L-Tryptophan PLS-Modelle zur Vorhersage dieser prozessrelevanten Größen berechnet. Die Modelle wurden aus den Daten mehrerer Prozesse erstellt, um möglichst viel Variationen im Modell zu berücksichtigen und die externe Vorhersagegüte für möglichst viele verschiedene Prozesse zu verbessern. Die Vorhersage eines neuen Prozesses zeigt für Indol und Tryptophan Vorhersagefehler von etwa 5 %, nur für Serin zeigten sich Fehler von mehr als 10 %. Für einen weiteren Prozess wurden große Schwierigkeiten bei der Vorhersage der Prozessvariablen beobachtet. Durch Verwendung von Differenzspektren zur Modellerstellung und Vorhersage konnten die Unterschiede in den Fluoreszenzspektren eliminiert werden, die durch die Melasse einer anderen Kampagne in das Serineluat eingebracht wurden. Trotzdem waren nur suboptimale Vorhersagen möglich, da die Änderungen der Fluoreszenzintensitäten bei Konzentrations-änderung der Fluorophore nicht mehr vergleichbar waren. Hierfür können fluoreszenz-spezifische Effekte, wie z.B. der Inner-Filter-Effekt oder der Kaskaden-Effekt verantwortlich sein.

Zur theoretischen Beschreibung der Offlinemesswerte sollte ein Bioprozessmodell erstellt und anhand von Simulationen die Prozessführungsstrategie festgelegt werden. Zur Erstellung eines geeigneten Bioprozessmodells mussten die Kinetiken der stattfindenden Reaktionen hinreichend gut mathematisch beschrieben werden. Für den am industriellen Prozess beobachteten Serinverlust wurde eine Kinetik 1. Ordnung bezüglich Serin bestimmt. Somit konnte der Serinverlust beschrieben und als Nebenreaktion identifiziert werden, vermutlich ist eine Serin desaminierende Restaktivität der Zellen ist hierfür verantwortlich. Eine genaue

Untersuchung dieser Nebenreaktion, ihrer Kinetik und der begünstigenden Bedingungen für die Nebenreaktion ist insbesondere für eine Optimierung der Produktausbeute wünschenswert.

Zur theoretischen Beschreibung der Messwerte aller drei Prozessvariablen musste außerdem eine Kinetik der Reaktion von Indol und Serin zu Tryptophan für das Bioprozessmodell gefunden werden. Für die Hauptreaktion wurde je eine Michaelis-Menten-Kinetik für die Substrate angenommen. Durch Variation der kinetischen Parameter und Verwendung verschiedener Optimierungsalgorithmen wurden die optimalen Parameterwerte für fünf Prozesse identifiziert. Eine Sensitivitätsanalyse der Parameter hat gezeigt, dass beide Michaelis-Menten-Konstanten sowohl für Serin als auch für Indol nicht relevant für das Ergebnis der Simulation sind. Durch Reduktion des Modells um diese beiden Parameter wurde ein neues Modell erstellt, wobei sich für Serin eine Kinetik 1. Ordnung und für Indol eine Kinetik 0. Ordnung ergeben hat.

Die Ergebnisse einer Parameteranpassungen der einzelnen Prozesse ergaben, dass die Aktivität der Enzyme für die Haupt- und die Nebenreaktionen korrelieren. Obwohl bei den Prozessen sowohl frische als auch von einer Umsatzlösung separierte Zellen gemischt eingesetzt wurden, traten enorme Unterschiede in der Enzymmenge oder aber Enzymaktivität auf. Die Bestimmung dieser Aktivität war im Rahmen dieser Arbeit nicht möglich, da weitere Messwerte nicht zur Verfügung standen.

Zur Verbesserung der Simulation wurden die Werte der kinetischen Parameter aus den Änderungen der Fluoreszenzintensität während der Zellzugabe abgeschätzt. Es konnte gezeigt werden, dass die zugegebene Enzymmenge, bzw. -aktivität für jeden Prozess individuell aus den Fluoreszenzspektren bestimmt werden kann. Diese Abschätzung ist erst nach Beendigung der Zellzugabe möglich. Danach kann jedoch mit den abgeschätzten Parameterwerten eine für die Regelung des Prozesses hinreichende Onlinesimulation des Prozesses ermöglicht werden. Die Anwendung des Modells auf weitere Prozesse hat gezeigt, dass zutreffende Simulationen der Konzentrationsverläufe der Prozessvariablen Indol, Serin und Tryptophan möglich sind. Auch Prozesse, die unter veränderten Bedingungen geführt wurden, wie z.B. erneute Serinzugaben und Verwendung hochkonzentrierter Serinlösung in Wasser, konnten mit dem erstellten Bioprozessmodell beschrieben werden. Das Bioprozessmodell ist darüber hinaus auch für die Simulation der Biotransformation gültig, wenn Serineluat aus Melasse anderer Kampagnen für den Prozess eingesetzt wird.

Die bei der Parameteridentifikation der fünf Prozesse beobachtete Korrelation der beiden Parameter untereinander, also das Verhältnis zwischen Haupt- und Nebenreaktion, konnte für andere Prozesse nicht durchgehend gefunden werden. Das Verhältnis variiert abhängig von dem Serineluat und den eingesetzten Zellen selbst. Durch die hier vorgestellten Studien am

industriellen Prozess, war eine genaue Analyse oder Identifikation der Bedingungen für diese Unterschiede im Rahmen dieser Arbeit nicht möglich. Dies sollte aber genauer untersucht werden, da das Verhältnis zwischen Haupt- und Nebenreaktion für eine optimale Produktausbeute ausschlaggebend ist. Die Kinetik der Nebenreaktion und die Aktivität der Zellen für die Desaminierung des Serins müssen für eine effektive Optimierung der Biotransformation genauer untersucht werden. Eventuell sollte auch die Möglichkeit des Einsatzes eines genetisch veränderten Bakterienstammes zur vollständigen Beseitigung der Nebenreaktion in Betracht gezogen werden.

Durch Kombination der Simulation und Onlineanalyse des Prozesszustandes sollte eine optimale Führung des Prozesses zur Einsparung von Ressourcen und eine Erhöhung der Raum-Zeit-Ausbeute realisiert werden. Für die Verknüpfung von Bioprozessmodell und chemometrischen Modell zur Vorhersage der Prozessvariablen wurde ein Kalman-Filter eingesetzt. Anhand der Zustandsschätzung des Kalman-Filters konnte erfolgreich gezeigt werden, dass das Messrauschen reduziert wurde und kontinuierliche Schätzungen der Prozessvariablen online möglich waren. Durch die Onlineschätzung der Prozessvariablen am industriellen Prozess kann das Substrat L-Serin durch weitere Indolzugaben vollständig umgesetzt werden. Mit Hilfe des Bioprozessmodells konnte die optimale Prozessführung durch Regelung der Indolzugabe identifiziert werden. Hierdurch kann die Umsetzungszeit von acht auf vier Stunden reduziert werden.

Basierend auf der Zustandsschätzung des Kalman-Filters oder auf der Onlinesimulation ist somit eine verbesserte Prozessregelung möglich, die zu Einsparungen an Energiekosten und zur besseren Ressourcenausnutzung führt. Eine Steigerung der Ausbeute ist aufgrund der Nebenreaktion tatsächlich nur um 22 % möglich, da die eingesetzten Zellen die maximale Ausbeute bzw. den minimalen Verlust bestimmen. Daher sollten insbesondere die Ursachen für die Nebenreaktion Ziel weiterer Untersuchungen sein. Durch eine optimale Regelung des Prozesses kann die Ausbeute dennoch erheblich gesteigert und gleichzeitig die Prozesszeit wesentlich verkürzt werden, was zu einer signifikanten Verbesserung der Raum-Zeit-Ausbeute führt.

## 7 Literaturverzeichnis

- [BANG 1981] W.-G. Bang: Untersuchungen zur L-Tryptophan-Bildung mit freien und immobilisierten *Escherichia coli* W 3100-Zellen, Dissertation, TU Braunschweig, 1981
- [BANG 1983] W.-G. Bang, S. Lang, H. Sahm, F. Wagner: Production of L-Tryptophan by *Escherichia coli* Cells, *Biotechnol. Bioengin.*, Vol. 25, 999-1011 (1983)
- [BISSWANGER 2000] H. Bisswanger: *Enzymkinetik - Theorie und Methoden*, Wiley-VCH Verlag Weinheim, 2000
- [BÖHL 2001A] D. Böhl, D. Solle, H.-J. Toussaint, C. Lindemann, B. Hitzmann, T. Scheper: Anwendung der Fluoreszenzspektroskopie zur Beobachtung von Bierbrauprozessen, *GIT-Laborzeitschrift*, 5, 2001
- [BÖHL 2001B] D. Böhl, D. Solle, H.-J. Toussaint, C. Lindemann, B. Hitzmann, T. Scheper: Use of Fluorescence Spectroscopy for the Monitoring of Beer Brewing Processes, *GIT-Laboratory Journal*, 5(4), 2001
- [BÖHL 2003] D. Böhl, D. Solle, B. Hitzmann, T. Scheper: Fluorescence monitoring during cultivation of *Claviceps purpurea* and chemometric modelling for bioprocess characterisations, unveröffentlicht
- [BRO 1998] R. Bro: *Multi-way analysis in the food industry - models, algorithms and applications*; doctoral thesis, Rasmus Bro, 1998
- [BROWN 1992] S. D. Brown, R. S. Bear, T. B. Blank; *Chemometrics*, *Anal. Chem.*, Vol. 64, 22R-49R, 1992
- [CRAWFORD 1964] I. P. Crawford, J. Ito: Serine Deamination by B protein of *Escherichia coli* Tryptophan Synthase, *Biochemistry*, Vol. 51 (1964)
- [DANZER 2001] K. Danzer, H. Hobert, C. Fischbacher, K.-U. Jagemann: *Chemometrik - Grundlagen und Anwendungen*, Springer (2001)
- [DREWE 1985] W. F. Drewe, M. F. Dunn: Detection and Identification of Intermediates in the Reaction of L-Serine with *Escherichia coli* Tryptophan Synthase via Rapid-Scanning Ultraviolet-Visible Spectroscopy, *Biochemistry*, 24, 3977-3987, 1985
- [FAEDER 1970] Feader, E. J., Hammes, G. G. (1970): Kinetic Studies of Tryptophan Synthetase. Interaction of Substrates with b subunit, *Biochemistry*, Vol. 9, No. 21, 4043-4049
- [FAEDER 1971] E. J. Feader, G. G. Hammes (1971): Kinetic Studies of Tryptophan Synthetase. Interaction of L-Serine, Indole, and Tryptophan with the Native Enzyme, *Biochemistry*, Vol. 10, No. 6, 1041-1045, 1971

- [FAURIE 1999] R. Faurie, G. Fries: From Sugar beet molasses to Lyphan<sup>®</sup> - Integrated Quality Management from the Raw Material to the Drug, *Advances in Experimental Medicine and Biology*, 467 (1999) 443-452
- [FUCHS 1986] C. Fuchs: Untersuchungen zur L-Tryptophan-Bildung mit *Escherichia coli* B10, Diplomarbeit, Universität Braunschweig
- [FUNK 1992] W. Funk, V. Dammann, G. Donnevert: Qualitätssicherung in der Analytischen Chemie, VCH Weinheim, 1992
- [FRANK 1993] I. E. Frank, J. H. Friedman: A statistical view of some chemometrics Regression Tools, *Technometrics* Vol 35 No 2, 109-135 (1993), American statistical assoziation and the american society for quality control
- [GEE 1996] D. A. Gee, W. F. Ramirez: On-Line State Estimation and parameter identification for batch fermentation, *Biotechnol. Prog.* Vol. 12, 132-140, (1996)
- [GEISLER 2002] D. Geissler, D. Solle, E. Stärk, T. Scheper, H. Märkl, B. Hitzmann: Berechnung chemometrischer Modelle basierend auf theoretischen Modellen und 2D-Fluoreszenzspektren, *Chemie Ingenieur Technik*, 74(8), 2002
- [HARBECK 2001A] C. Harbeck, B. Hitzmann, C. Lindemann, P. Sosnitzer, R. Faurie, T. Scheper: Verwendung der 2D-Fluoreszenzspektroskopie zur On-line-Überwachung der chromatographischen Melasseentzuckerung, *Chemie Ingenieur Technik* (73), Wiley-VCH-Verlag, Weinheim, 2001
- [HARBECK 2001B] C. Harbeck, C. Lindemann, P. Sosnitzer, R. Faurie, T. Scheper: Fluorescence Spectroscopy for Online Monitoring of Chromatographic Molasses Desugarization, *G.I.T. Laboratory Journal* &/2001
- [HATANAKA 1962] M. Hatanaka, E. A. White, K. Horibata, I. P. Crawford: A Study of the Catalytic Properties of *Escherichia coli* Tryptophan Synthetase, a Two-Component Enzyme, *Archives of Biochemistry and Biophysics*, 97, 596-606, (1962)
- [HEARNE 1985] J. W. Hearne: Sensitivity analysis of parameter combinations, *Appl. Math. Modelling*, 1985, Vol. 9, Butterworth & CO, 1985
- [HEILMANN 1978] H.-D. Heilmann: On the mechanism of action of *Escherichia coli* tryptophan synthase, *Biochimica et Biophysica Acta* 522 (1978) 614-624, Elsevier/North-Holland Biomedical Press
- [HITZMANN 1998A] B. Hitzmann, T. Pekeler, C. Lindemann, S. Marose, T. Scheper: Chemometric Models for The On-Line estimation of Bioprocess variables from 2-D-Fluorescence spectra, *Computer Applications in Biotechnology – CAB7*, Osaka (1998)

- [HITZMANN 1998B] B. Hitzmann, S. Marose, C. Lindemann, T. Scheper: New Sensor Application in european Industries for detailed Bioprocess monitoring, Computer Applications in Biotechnology – CAB7, Osaka (1998)
- [HITZMANN 2000] B. Hitzmann, O. Broxtermann, Y.-L. Cha, O. Sobieh, E. Stärk, T. Scheper: The control of glucose concentration during yeast fed-batch cultivations using a fast measurement complemented by an extended Kalman filter, Bioprocess Engineering, 23, 337-341, 2000
- [IRTEL 2000] F. Irtel, P. Sosnitza, M. Busse, R. Faurie, L. Fischer, R. Ulber, T. Scheper: Langzeiteinsatz von Biosensorsystemen zur Überwachung industrieller Chromatographieprozesse, Chemie Ingenieur Technik (5), 2000
- [KAWATA 1990] Y. Kawata, N. Tsujimoto, S. Tani, T. Mizobata, M. Tokushige: Role of Tryptophan 248 in The Active Site of Tryptophanase From *Escherichia Coli*, Biochem. And Biophys. Res. Comm., Vol. 173, No. 2 (1990)
- [KIRSCHNER 1975] K. Kirschner, R. L. Wiskocil, M. Foehn, L. Rezeau: Tryptophan Synthase from *Escherichia coli* – Improved Purification Procedure for alpha-Subunit and Binding Studies with Substrate Analogues, Eur. J. Biochem., Vol. 60, No 2, 513-523, 1975
- [KLABEN 2003] Mündliche Mitteilung, 2003
- [KLEIN 1984] J. Klein, J. Stock, K.-D. Vorlop, F. Wagner, S. Lang: Optimierung der L-Tryptophan-Produktion mit immobilisierten Zellen von *E. coli* B10, Forum Mikrobiologie 7 (1984) Sonderheft Biotechnologie
- [LANE 1983A] A. N. Lane, K. Kirschner: The Catalytic Mechanism of Tryptophan Synthase from *Escherichia coli* - Kinetics of the Reaction of Indole with the Enzyme-L-Serine Complex, Eur. J. Biochem. 129, 571-582 (1983)
- [LANE 1983B] A. N. Lane, K. Kirschner: The Mechanism of Binding of L-Serin to Tryptophan Synthase from *Escherichia coli*, Eur. J. Biochem. 129, 561-570 (1983)
- [LANE 1991] A. N. Lane, K. Kirschner: Mechanism of the Physiological Reaction Catalysed by Tryptophan Synthase from *Escherichia coli*, Biochemistry 30 (1991), 479-484
- [LEJA 1995] C. A. Leja, E. U. Woehl, M. F. Dunn: Allosteric Linkages between  $\beta$ -Site Covalent Transformations and  $\alpha$ -Site Activation and Deactivation in the Tryptophan Synthase Bienenzyme Complex, Biochemistry, 34, 6552-6561, 1995

- [LI 1991] J.-K. Li, A. E. Humphrey: Use of fluorometry for monitoring and control of a bioreactor, *Biotechnol. Bioeng.*, 37, 1043 - 1049
- [LINDEMANN 1998] C. Lindemann: Anwendung der 2D-Fluoreszenzspektroskopie zur Bioprozessbeobachtung, *Fortschritt-Berichte VDI Reihe 17*, Nr. 171, Düsseldorf: VDU Verlag 1998
- [MAROSE 1998] S. Marose, C. Lindemann, T. Scheper: Two-Dimensional Fluorescence Spectroscopy: A New Tool for On-Line Bioprocess Monitoring, *Biotechnol. Prog.*, 14, 63 – 74, 1998
- [MARTENS 1989] H. Martens, T. Naes: *Multivariate Calibration*, John Wiley & Sons, Chichester
- [MASSART 1997] D. L. Massart, B. G. M. Vandeginste, L. M. C. Buydens, S. De Jong, P. J. Lewi, J. Smeyers-Verbeke: *Handbook of Chemometrics and Qualimetrics – Part A*, Elsevier Science, Amsterdam
- [MATEUS 1996] D. M. R. Mateus, S.S. Alves, M.M.R. da Fonseca: Model for the production of L-tryptophan from L-serine and indole by immobilized cells in a three-phase liquid-impelled loop reactor, *Bioprocess Engineering* 14 (1996) 151-158, Springer-Verlag
- [MATEUS 1999] D. M. R. Mateus, S.S. Alves, M.M.R. da Fonseca: Diffusion in Cell-Free and cell Immobilising k-Carrageenan Gel Beads with and without Chemical Reaction, *Biotechnology Bioengineering*, Vol. 63, No 5, 1999
- [MÜLLER 1999] B. Müller: *Synthese, Analytik und Bildungsbedingungen von Kontaminationen in biotechnologisch hergestelltem Tryptophan*, Dissertation, Universität Hamburg, 1999
- [MUKHERJEE 1999] J. Mukherjee, C. Lindemann, T. Scheper: Fluorescence monitoring during cultivation of *Enterobacter aerogenes* at different oxygen levels, *Appl. Microbiol. Biotechnol.* 52, 484-494, 1999
- [MUNACK 1989] A. Munack: Optimal Feeding Strategy for Identification of Monod-Type Models by Fed-Batch Experiments, *Computer Applications in Fermentation Technology*, 1989, S. 195 - 204
- [OTTO 1997] M. Otto: *Chemometrie – Statistik und Computereinsatz in der Analytik*, VCH Weinheim
- [PERKAMPUS 1993] H.-H. Perkampus: *Lexikon Spektroskopie*, VCH Weinheim, 1993
- [PROTSCH 2003] Mündliche Mitteilung, 2003
- [SCHNEIDER 1995] R. Schneider, A. Munack: Improvements in the On-Line Parameter Identification of Bioprocesses, *Computer Applications in Biotechnology*, 1995,

- [SCHNEIDER 1999] R. Schneider: Untersuchung eines adaptiven prädikativen Regelungsverfahrens zur Optimierung von bioverfahrenstechnischen Prozessen, Fortschritt Berichte VDI Reihe 8 Nr. 755, Düsseldorf: VDI Verlag 1999
- [SCOGMAN 1984] G. S. Scogman, J.-E. Sjöström: Factors Affecting the Biosynthesis of L-Tryptophan by Genetically Modified Strains of *Escherichia coli*, *Journal of General Microbiology*, 130, 3091-3100, 1984
- [SOLLE 2001] D. Solle, C. Protsch, R. Ulber, K. Seja, B. Willke, R. Faurie, B. Hitzmann, T. Scheper: Optimierung der industriellen Tryptophanproduktion, *Biospektrum Sonderausgabe Sensorik*, Hrsg.: R. Erb und S. Heiden, Spektrum Akademischer Verlag, Heidelberg (2001)
- [SOLLE 2003] D. Solle, D. Geissler, E. Stärk, T. Scheper, B. Hitzmann: Chemometric Modelling based on 2D-Fluorescence Spectra without a Calibration Measurement, *Bioinformatics* (19), No. 2 (2003), 173-177, Oxford University Press
- [TRIBE 1979] D. E. Tribe, J. Pittard: Hyperproduction of Tryptophan by *Escherichia coli*: genetic Manipulation of the Pathway Leading to Tryptophan Formation, *Applied and Environmental Microbiol.*, Vol. 38, No 2, 181-190, 1979
- [ULBER 2000] R. Ulber, R. Faurie, P. Sosnitsa, L. Fischer, E. Stärk, C. Harbeck, T. Scheper: Monitoring and control of industrial downstream processing of sugar beet molasses, *Journal of Chromatography A*, 882 (2000), 329-334, Elsevier,
- [ULBER 2001A] R. Ulber, B. Hitzmann, T. Scheper: Innovative Bioprozessanalytik - Neue Wege zum besseren Verständnis biotechnologischer Prozesse, *Chemie Ingenieur Technik* (73), 19-26. 2001, Wiley-VCH Weinheim
- [ULBER 2001B] R. Ulber, C. Protsch, D. Solle, B. Hitzmann, B. Willke, R. Faurie, T. Scheper: Optimierung der industriellen Tryptophanproduktion durch den Einsatz bioanalytischer Systeme, *CIT*, Vol. 73 (2001)
- [VANDEGINSTE 1998] B. G. M Vandeginste, D. L. Massart, L. M. C. Buydens, S. De Jong, P. J. Lewi, J. Smeyers-Verbeke: *Handbook of Chemometrics and Qualimetrics – Part B*, Elsevier Science, Amsterdam
- [VOET 1992] D. Voet, J. G. Voet: *Biochemie*, Wiley & Sons Weinheim
- [VORLOP 1983] K.-D. Vorlop: Entwicklung von Verfahren zur Polymerfixierung von Mikroorganismen und Anwendung der Biokatalysatoren zur Spaltung von Penicillin G und Synthese von L-Tryptophan, Dissertation, TU Braunschweig, 1983
- [WAGNER 1982] F. Wagner, S. Lang, W.-G. Bang, K. D. Vorlop, J. Klein: Production of L-Tryptophan with immobilized Cells, *Enzyme Engineering*, Vol 6, 1982

- [YANOFSKY 1960] C. Yanofsky: Tryptophan Synthetase System, Bacteriol Rev. 24 (2), 1960

## 8 Anhang

### 8.1 Verzeichnis der Abkürzungen

Abkürzung	Bedeutung	Abkürzung	Bedeutung
A	Koeffizient	PCR	Principal Component Regression
Amm	Ammonium	PLP	Pyridoxal-5-Phosphat
b	Regressionskoeffizient	PLS	Partial Least Square Regression
B	Koeffizient	PRESS	Predictive Residual Sum of Squares
°C	Celsius	Q	Prozessrauschleistungsmatrix
$\hat{c}$	Vorhergesagte Konzentration	R	Messrauschkovarianzmatrix
D	Diffusion	RMSEP	Root Mean Square Error of Prediction
Dev	Varianz	RV	Reaktionsgeschwindigkeit
div	Divergenz	s	Messrauschen
e	Fehler	S	Serinkonzentration
E	Erwartungswert	S	Substratkonzentration
EC	Enzyme Classification	SCys	S-hydroxyethyl-L-cystein
EMS	Eosinophilia-Myalgia-Syndrom	Sens	Sensitivitätswert einer Wellenlängenkombination
f	Prozessvariablengleichung	Ser	Serin
F	Jakobimatrix von f	SV	Serinverlust
Gew.-%	Gewichtsprozent	t	Zeit
Gly	D-Glyceraldehyd-3-phosphat	$\vec{t}$	Scorevektor
grad	Gradient	T	Scorematrix
h	Messwertgleichung	T	Tryptophankonzentration
H	Jakobimatrix von h	Trp	Tryptophan
HPLC	High Performance Liquid Chromatography	u	Prozessbedingungen
i	Laufvariable für die Messung	u	Strömungsgeschwindigkeit
I	Indolkonzentration	v	Stöchiometrische Zahl
Ind	Indol	V	Volumen
IndGly	Indol-3-glycerol-phosphat	$\dot{V}_{BFM}$	Volumenstrom der Biomassezugabe
j	Laufvariable für die Messgröße	$\dot{V}_{Kond}$	Volumenstrom des Kondensats
K	Kalmanverstärkungsmatrix	x	unabhängige Variable
$K_m$	Michaelis-Menten-Konstante	X	Matrix der unabhängigen Variablen (Messwerte)
L	Liter	X	Relative Biomasse
m	Anzahl der Messgrößen	$X^{-1}$	Inverse der Matrix X
MEth	Mercaptoethanol	$\bar{x}$	Mittelwert der Variablen x
$\dot{m}_i$	Massenstrom der Indolzugabe	$x^{cent}$	zentrierte Variable x
mM	millimolar	$x^{skal}$	skalierte Variable x
MLR	Multi-Lineare-Regression	X'	Transponierte der Matrix X
$nm_{Em}$	Emissionswellenlänge	$X_0$	Relative Biomasse in der Biomassezugabe
$nm_{Ex}$	Exzitationswellenlänge	y	abhängige Variable
$\mu_{max}$	maximale Reaktionsgeschwindigkeit	Y	Matrix der abhängigen Variablen (Konzentrationen)
n	Anzahl der Messungen	z	Prozessrauschen
NIPALS	Nonlinear Iterative Partial Least Square		
$\vec{p}$	Loadingvektor		
P	Loadingmatrix		
P	Schätzfehlerkovarianzmatrix		
PC	Principal Component		
PCA	Principal Component Analysis		

## 8.2 Algorithmen

### 8.2.1 NIPALS-Algorithmus

Der Algorithmus berechnet jede Komponente schrittweise. Die Berechnung der Komponente erfolgt iterativ über wiederholte Regressionen der Datenmatrix  $X$  auf den Scorevektor  $t$  für eine Verbesserung des Loadingvektors  $p$  und der Datenmatrix  $X$  auf diesen Loadingvektor  $p$  für eine Verbesserung des Scorevectors  $t$ .

Auswahl eines Startvektors  $t$  aus den Spalten der Datenmatrix  $X$

Verbesserte Schätzung des Loadingvektors  $p$  für diese Hauptkomponente durch Projektion der Matrix  $X$  auf  $t$ :

$$p'=(t't)^{-1}t'X$$

Scalierung von  $p$  auf die Länge 1:

$$p=p(p'p)^{-0.5}$$

Verbesserte Schätzung des Scorevectors  $t$  für diese Hauptkomponente durch Projektion der Matrix  $X$  auf  $p$ :

$$t=Xp(p'p)^{-1}$$

Berechnung des Eigenwertes  $\tau$ :

$$\tau=t't$$

Wenn die Differenz der Eigenwerte  $\tau$  aus dieser und der vorangegangenen Iteration kleiner ist als das Konvergenzkriterium (z.B. 0,0001), ist die Hauptkomponente berechnet, falls dies nicht zutrifft werden die Schritte 2-6 wiederholt.

Berechnung der Matrix für die nächste Hauptkomponente:

$$X_{\text{neu}}=X-tp'$$

### 8.2.2 PLS1-Algorithmus

Der Algorithmus berechnet jede Komponente schrittweise. Die Berechnung der Komponente erfolgt über die Regressionen der kleinsten Fehlerquadrate bei der Bestimmung des Loadinggewichtsvektor  $w$ , des Scorevektors  $t$  und des Loadingvektors  $q$ .

Schätzung des Loadinggewichtsvektors  $w$  durch Regression der kleinsten Fehlerquadrate des lokalen Modells:

$$X=yw'+E$$

Die Lösung der Regression lautet:

$$w=cX'y$$

wobei  $c$  der Skalierungsfaktor ist, der den Loadinggewichtsvektor auf die Länge 1 skaliert durch:

$$c=(y'XX'y)^{-0.5}$$

Schätzung des Scorevektors  $t$  durch Regression der kleinsten Fehlerquadrate des lokalen Modells:

$$X=tw'+E$$

Schätzung des Loadingvektors  $q$  durch Regression der kleinsten Fehlerquadrate des lokalen Modells:

$$y=tq+f$$

Die Lösung der Regression lautet:

$$q=y't/t't$$

Berechnung der neuen Datenmatrix  $X$  und des  $y$ -Vektors durch Subtraktion der berechneten Hauptkomponente:

$$X_{\text{neu}}=X-tp'$$

$$y_{\text{neu}}=y-tq$$

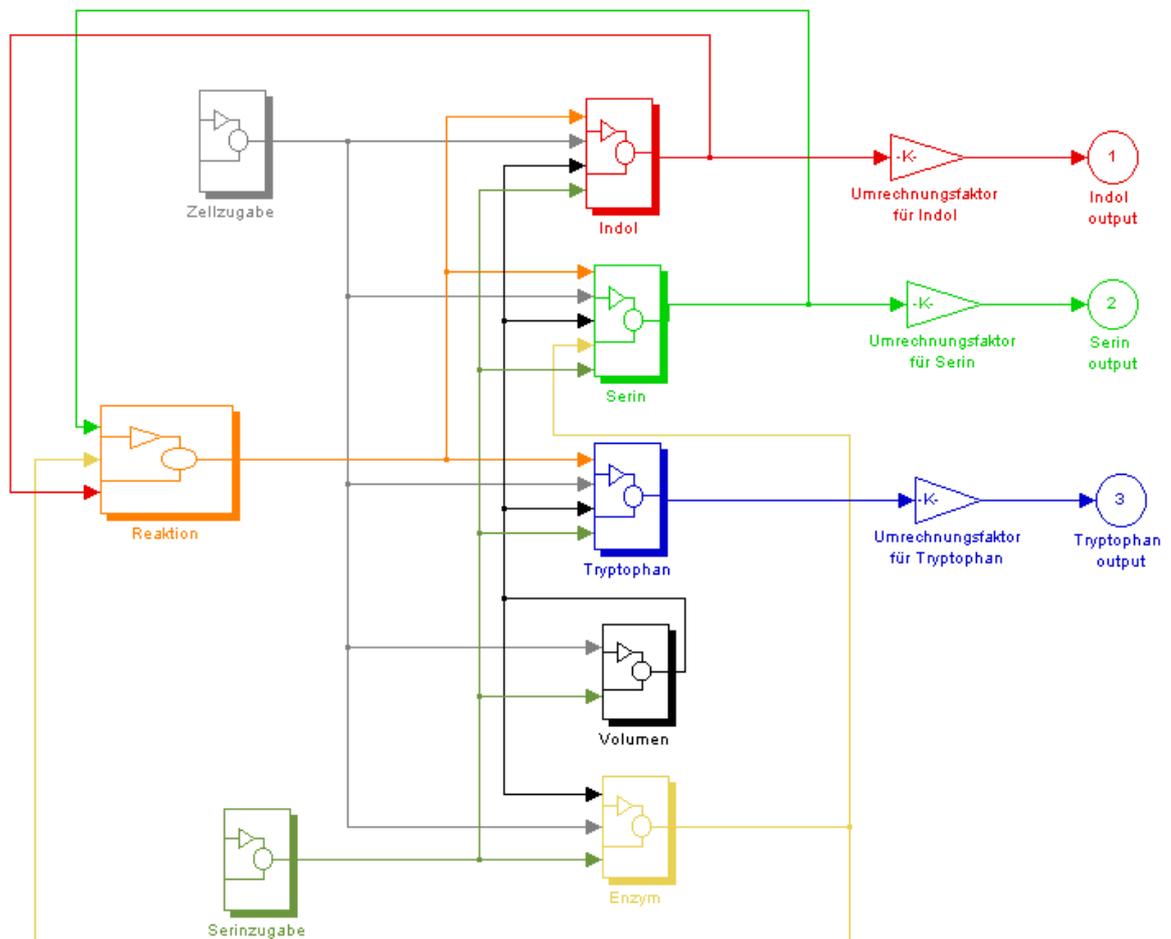
Berechnung der Regressionkoeffizienten für die Vorhersage:

$$b=W(P'W)^{-1}q$$

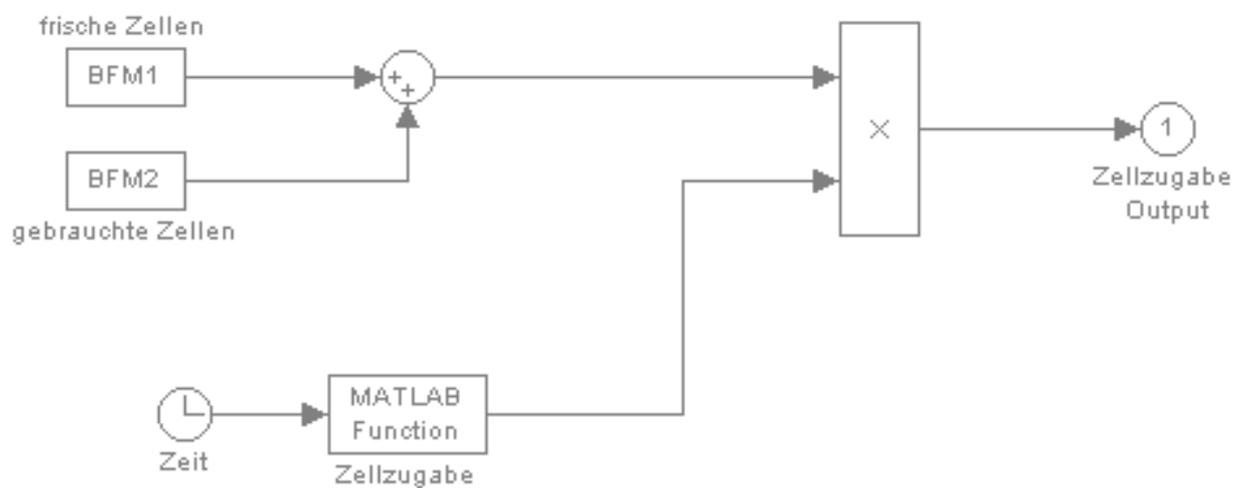
$$b_0=\bar{y}-\bar{x}'b$$

### 8.3 Modelle in Simulinkdarstellung

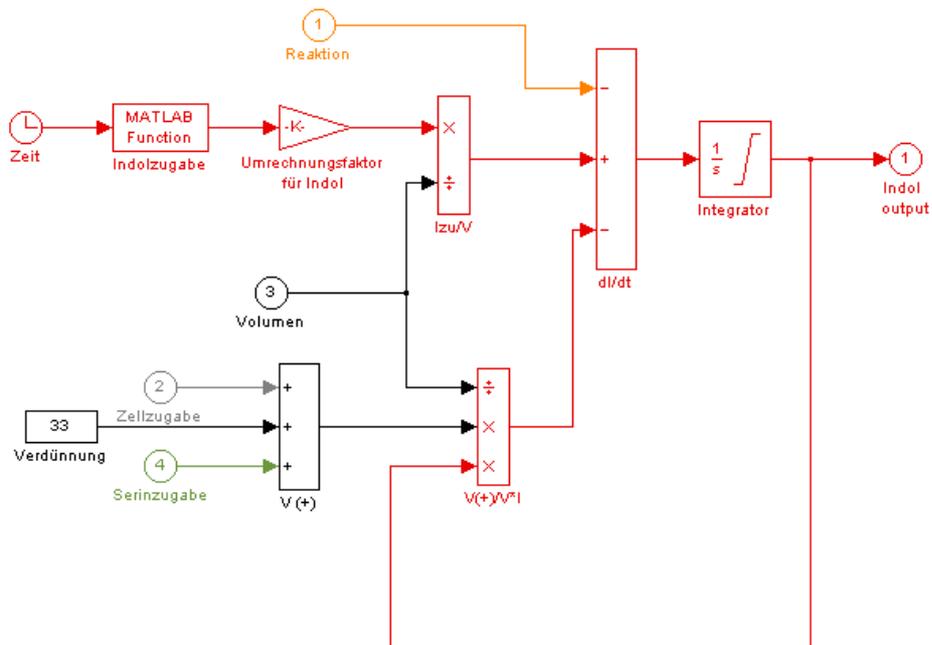
Die Grundstruktur



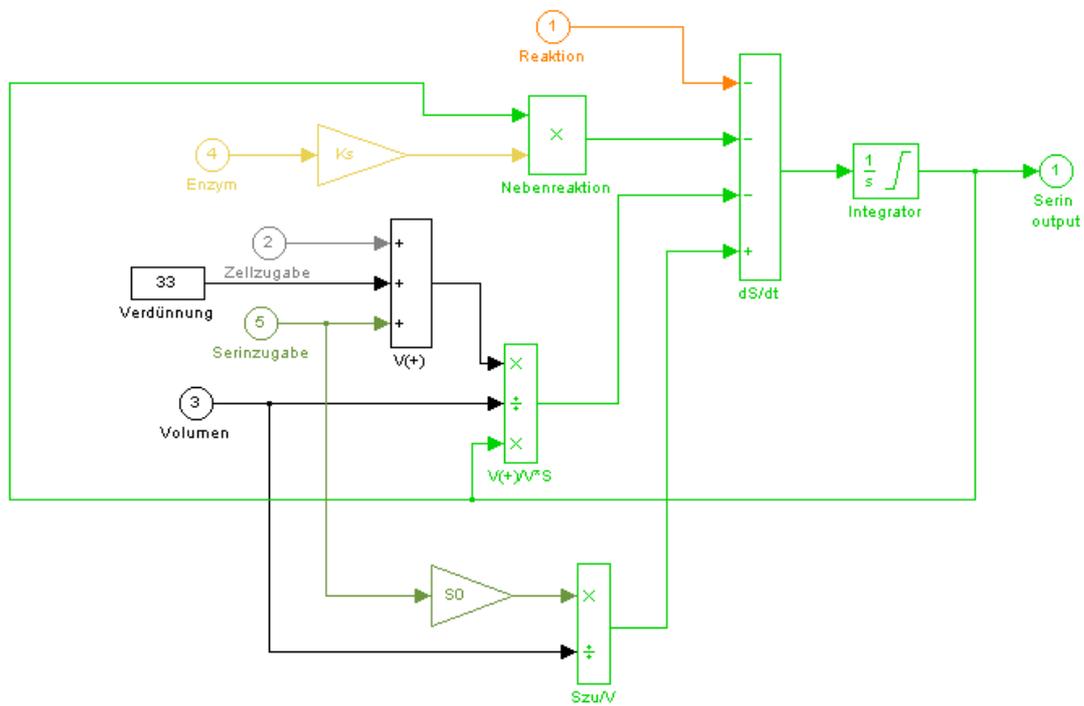
Die Zellzugabe



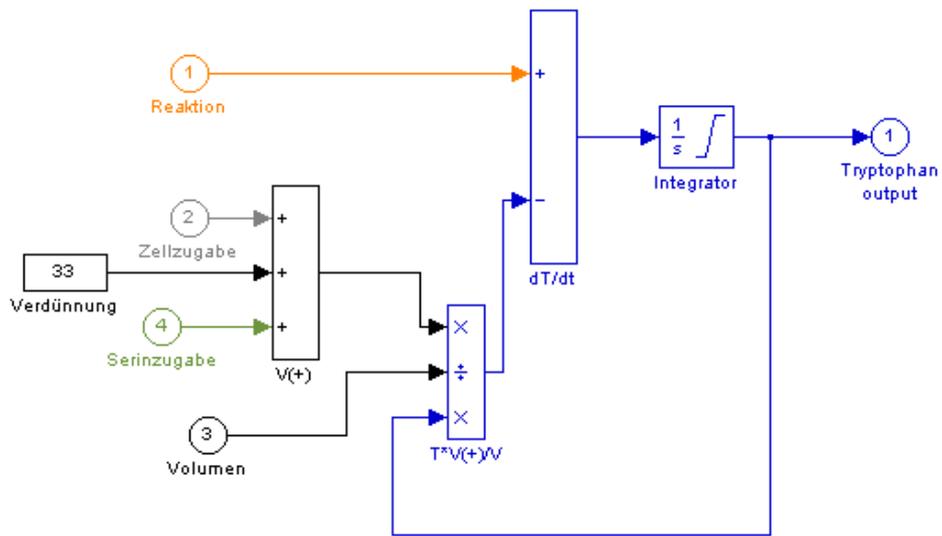
## Das Indolsubmodell



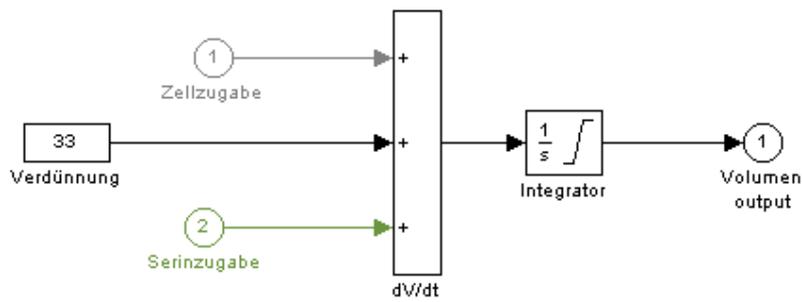
## Das Serin submodell



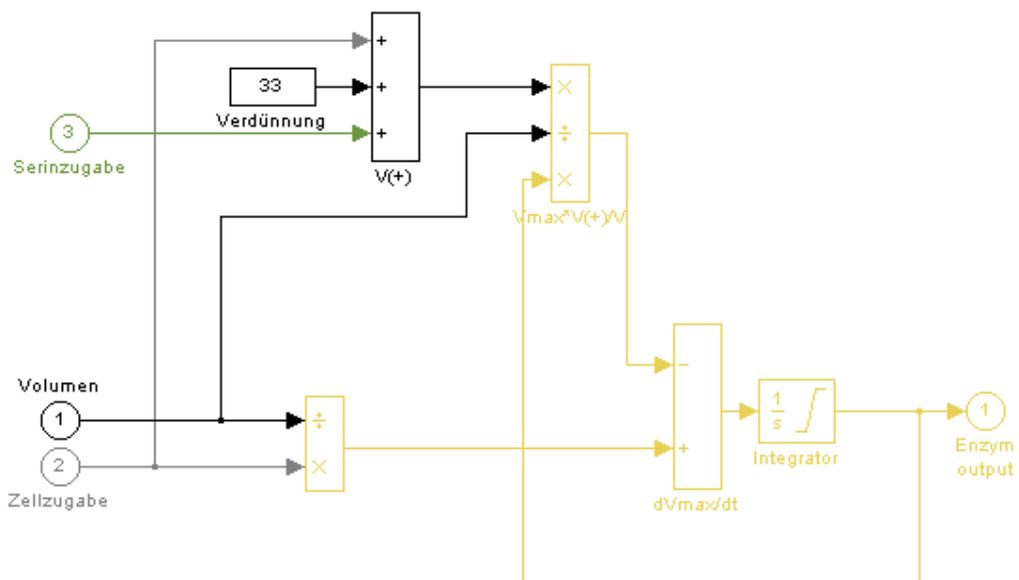
### Das Tryptophansubmodell



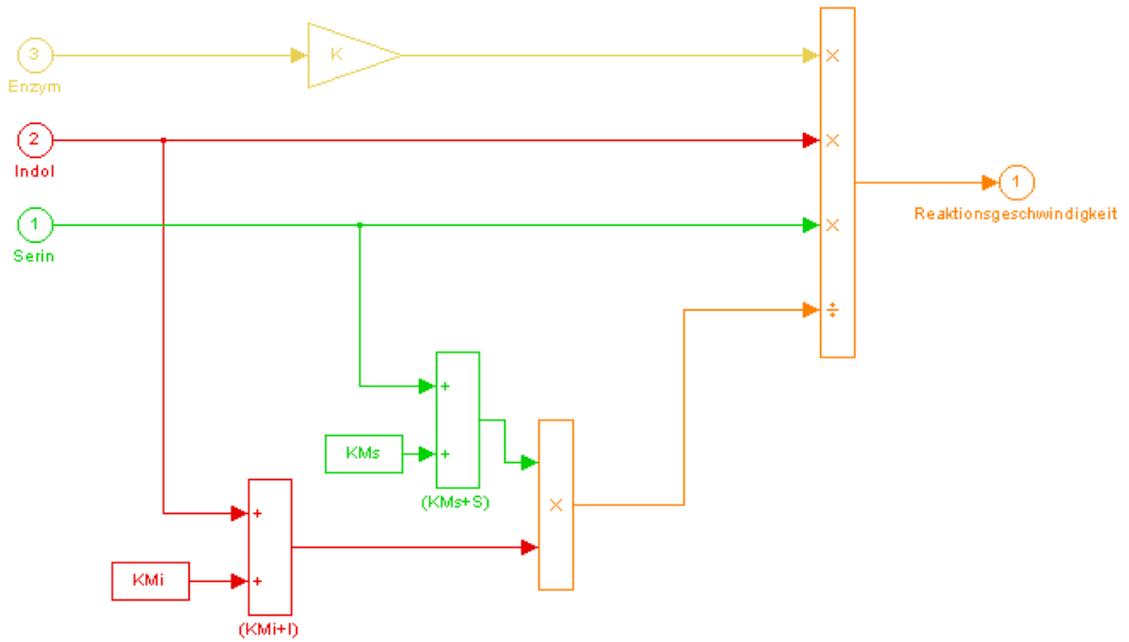
### Das Volumensubmodell



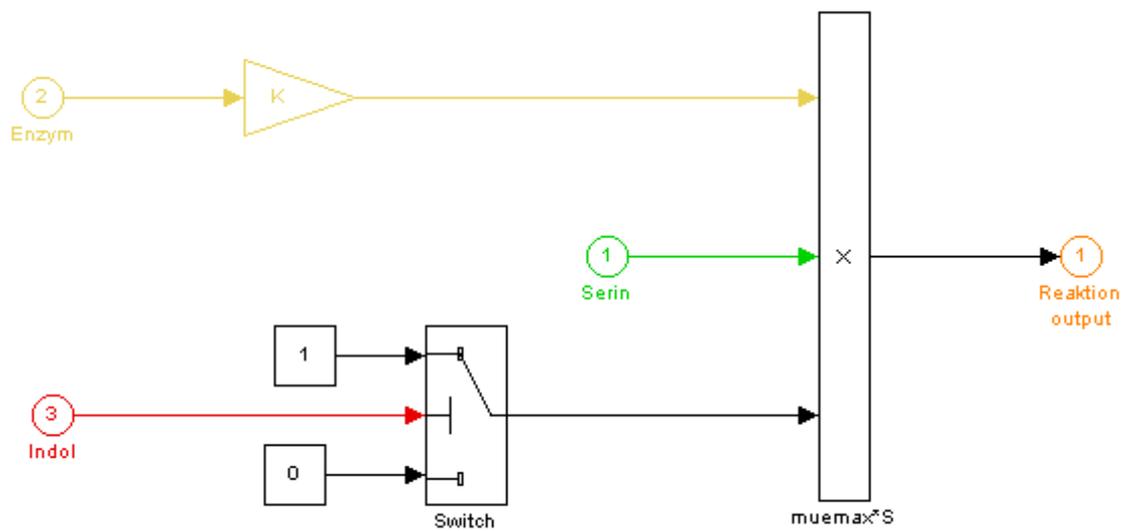
### Das Enzymsubmodell



Die Reaktionsgeschwindigkeit des vierparametrischen Modells



Die Reaktionsgeschwindigkeit des zweiparametrischen Modells



## 8.4 Parameter des Kalman-Filters

Prozessrauschleistungsmatrix

0,0001	0,0	0,0	0,0	0,0	0,0	0,0
0,0	0,0001	0,0	0,0	0,0	0,0	0,0
0,0	0,0	0,0001	0,0	0,0	0,0	0,0
0,0	0,0	0,0	0,1	0,0	0,0	0,0
0,0	0,0	0,0	0,0	0,001	0,0	0,0
0,0	0,0	0,0	0,0	0,0	100	0,0
0,0	0,0	0,0	0,0	0,0	0,0	0,1

Messrauschkovarianzmatrix

0,01	0,0	0,0
0,0	0,05	0,0
0,0	0,0	0,005

Startwerte der Schätzfehlerkovarianzmatrix

0,001	0,0	0,0	0,0	0,0	0,0	0,0
0,0	0,001	0,0	0,0	0,0	0,0	0,0
0,0	0,0	0,001	0,0	0,0	0,0	0,0
0,0	0,0	0,0	100	0,0	0,0	0,0
0,0	0,0	0,0	0,0	0,01	0,0	0,0
0,0	0,0	0,0	0,0	0,0	0,1	0,0
0,0	0,0	0,0	0,0	0,0	0,0	0,1

## Wissenschaftlicher Werdegang

Dörte Solle

geboren am 09.11.1968 in Lehrte  
ledig

### Schulbildung

1975 – 1979 Grund- und Hauptschule IV in Burgdorf  
1979 – 1981 Orientierungsstufe in Burgdorf  
1981 – 1986 Gymnasiums Burgdorf  
1986 – 1990 Goetheschule Hannover  
Abschluß: Abitur mit der Note 1,4  
Prüfungsfächer: Mathematik, Chemie, Philosophie, Musik

### Studium

10/1990 –09/1992 Universität Hannover, Chemie (Diplom)  
10/1992 –09/ 1998 Universität Hannover, Chemie / Biologie (Lehramt an Gymnasien)  
Abschluß: 1. Staatsexamen für höheres Lehramt, Note 2,6  
04/1999 –05/ 2003 Universität Hannover, Chemie (Promotion)  
Abschluß: Promotion zum Doktor der Naturwissenschaften, Note 1  
seit 04/2002 Fernuniversität Hagen, Bachelor of Science Mathematik / Informatik  
(Gasthörerin)

## Publikationen

- D. Solle, D. Geissler, E. Stärk, T. Scheper, B. Hitzmann: Chemometric Modelling based on 2D-Fluorescence Spectra without a Calibration Measurement, *Bioinformatics* (19), No. 2, 173-177, Oxford University Press 2003
- D. Solle, C. Protsch, R. Ulber, K. Seja, B. Willke, R. Faurie, B. Hitzmann, T. Scheper: Optimierung der industriellen Tryptophanproduktion, *Biospektrum Sonderausgabe Sensorik*, Hrsg.: R. Erb und S. Heiden, Spektrum Akademischer Verlag, Heidelberg 2001
- D. Geissler, D. Solle, E. Stärk, T. Scheper, H. Märkl, B. Hitzmann: Berechnung chemo-metrischer Modelle basierend auf theoretischen Modellen und 2D-Fluoreszenzspektren, *Chemie Ingenieur Technik*, 74(8), 2002
- R. Ulber, C. Protsch, D. Solle, B. Hitzmann, B. Willke, R. Faurie, T. Scheper: Optimierung der industriellen Tryptophanproduktion durch den Einsatz bioanalytischer Systeme, *CIT*, Vol. 73, 2001
- D. Böhl, D. Solle, H.-J. Toussaint, C. Lindemann, B. Hitzmann, T. Scheper: Use of Fluorescence Spectroscopy for the Monitoring of Beer Brewing Processes, *GIT-Laboratory Journal*, 5(4), 2001
- D. Böhl, D. Solle, H.-J. Toussaint, C. Lindemann, B. Hitzmann, T. Scheper: Anwendung der Fluoreszenzspektroskopie zur Beobachtung von Bierbrauprozessen, *GIT-Laborzeit-schrift*, 5, 2001