

Methoden zur Prozessüberwachung und Modellierung
von Herstellungsprozessen rekombinanter Proteine
mit *Escherichia coli*

Von der Naturwissenschaftlichen Fakultät
der Gottfried Wilhelm Leibniz Universität
Hannover

zur Erlangung des Grades

Doktor der Naturwissenschaften
- Dr. rer. nat. -

genehmigte Dissertation

von

M.Sc. Dipl.-Ing. (FH) Christian Kaiser
Geboren am 03. Januar 1975 in Hannover

2010

Referent: Prof. Dr. T. Scheper
Korreferent: Prof. Dr. B. Hitzmann
Tag der Promotion: 20.01.2010

Hiermit versichere ich an Eides statt, die vorliegende Dissertation selbständig und nur unter Verwendung der angegebenen Hilfsmittel angefertigt zu haben. Ich versichere ferner, dass die Dissertation nicht schon als Diplomarbeit oder ähnliche Prüfungsarbeit verwendet worden ist.

Christian Kaiser

Hamburg, März 2010

Veröffentlichungen

Teile dieser Arbeit wurden bereits in folgenden Beiträgen vorab veröffentlicht:

Publikationen

- [1] Kaiser, C; Carvel, J.P., Luttmann, R.: Development of a sensitive and compact in-situ biomass measurement system for controlling and monitoring microbial fermentations using radio-frequency impedance, *BioProcess*, 01.2007
- [2] Kaiser, C., Pototzki, T., Ellert, A., Luttmann, R.: Applications of PAT-Process Analytical Technology in Recombinant Protein Processes with *Escherichia coli*, *Eng. Life Sci.*, **8** (2), p. 132–138, 2008
- [3] Luttmann, R.: Monitoring und Kontrolle von Bioreaktionsprozessen zur Herstellung heterologer Proteine mit *Escherichia coli*, Abschlussbericht für die Max-Buchner-Forschungstiftung Kennziffer 2627, 2007. – URL http://www.dechema.de/Forschung_und_Forschungsforderung/Max_Buchner__Forschungstiftung-p-190/Abgeschlossene_Projekte-p-907/2007-p-123070.html. Abruf: 25. März 2009
- [4] Luttmann, R., Cordes, C., Gollmer, K.-U.: Monitoring und Control von Bioreaktionsprozessen zur Herstellung heterologer Proteine mit *Escherichia coli*, Abschlussbericht: Programm zur Förderung angewandter Forschung an Fachhochschulen im Verbund mit der Wirtschaft (FH³); Förderzeitraum 09/04 – 12/07, bmbf-FKZ: 17 10B 04

Tagungsbeiträge

- [5] Kaiser, C; Bauch, T; Rohde A; Peuker T; Luttmann, R.: Applications of PAT-Process Analytical Technology in Recombinant Protein Processes with *Escherichia coli*, Vortrag ESBES 6, Salzburg, 2006
- [6] Kaiser, C.; Eicke, M.; Bauch, T.; Ellert, A.; Luttmann, R.: Process Analytical Technology in Recombinant *E. coli* Processes, Poster auf der European Bioperspectives 2007, Köln
- [7] Kaiser, C., Peuker, T., Bauch, T., Ellert, A., Luttmann, R.: PAT-Process Analytical Technology in Cultivation Processes with Recombinant *Escherichia coli*, Vortrag und Proceedings der CAB 10-Computer Applications in Biotechnology, Cancun, Mexiko, 2007 Pons, M. N. and v. d. Impe, J.; Eds., Elsevier, 2007

gratiarum actio

Die vorliegende Arbeit wurde in der Zeit vom März 2005 bis April 2009 im Forschungs- und Transferzentrum Bioprocess- und Analysetechnik der Hochschule für Angewandte Wissenschaften (HAW) Hamburg unter der Leitung von Prof. Dr.-Ing. Reiner Luttmann angefertigt.

Ihm gehört mein besonderer Dank, da diese Arbeit durch sein persönliches Engagement und seine Unterstützung erst ermöglicht wurde. Seine fachliche Unterstützung war mir beim Anfertigen dieser Arbeit eine große Hilfe. Sein hoher Anspruch und Perfektionismus an sich sowie an seine Mitarbeiter waren dabei nicht immer leicht hinzunehmen, haben dem Inhalt und der Form dieser Arbeit aber bestimmt sehr gut getan. Sie haben mich dazu bewegt die Darstellung meiner Methoden und Ergebnisse immer wieder zu hinterfragen und zu optimieren.

Herrn Prof. Dr. Thomas Scheper und Herrn Prof. Dr. Bernd Hitzmann vom Institut für Technische Chemie der Universität Hannover danke ich für die Übernahme der Referate.

Jörn Hülsmann, Florian Mariacher und Lena Pototzki danke ich für den Einsatz im Rahmen Ihrer Diplomarbeiten und die von Ihnen erzielten Ergebnisse. Herrn Prof. Dr.-Ing. Klaus-Uwe Gollmer danke ich für die kostenlose Bereitstellung seiner MATLAB[®]-Toolbox BIOSS sowie die anregenden und hilfreichen Diskussionen des Bioreaktionsmodells.

Finanziell wurde diese Arbeit durch das Promotionsförderprogramm der HAW Hamburg unterstützt. Ich möchte mich insbesondere bei Herrn Ralf Behrens für die gute Betreuung und die Durchsetzung der Verlängerung meiner finanziellen Unterstützung bedanken. Außerdem wurde diese Arbeit durch das bmbf (FKZ: 17 10B 04) sowie durch die Max-Buchner Forschungsstiftung (Kennziffer: 2627) gefördert.

Für das gute Arbeitsklima und den schönen Lebensabschnitt möchte ich insbesondere meinen Kollegen Andree, Hape, Uli, Ali, Matze, Torben, Fritze und allen die ich vergessen habe danken.

Allen Beteiligten unseres MoCoBi-Projektes in Magdeburg und Trier spreche ich hiermit meinen Dank für die nicht immer leichte aber schöne und lehrreiche Zeit aus.

Zum Abschluss möchte ich mich insbesondere bei meiner Patchwork-Familie bedanken. Bei meiner Freundin Petra für das große Verständnis für meine Arbeit und die Abstriche die wir dafür bei unserer kleinen Familie machen mussten. Bei meinen Eltern und Jutta für die großartige Unterstützung, die mir immer zu Gute gekommen ist und meinen Geschwistern für die schönen unbeschwerten Zeiten zusammen. Der letzte Dank gilt unseren Kindern Mila und Len, die mich jetzt nach dem Warum fragen, mich stets erheitern und mir soviel Freude und Spass bereiten.

*„Viel öfter kommt, was unverhofft,
als was man hofft“*

Plautus

Kurzfassung

Christian Kaiser

„Methoden zur Prozessüberwachung und Modellierung von Herstellungsprozessen rekombinanter Proteine mit *Escherichia coli*“

Im Rahmen dieser Arbeit wurden unterschiedliche Methoden der Beobachtung, Überwachung und Simulation von Herstellungsprozessen rekombinanter Proteine mit *Escherichia coli* erprobt. Diese basieren auf kommerziell erhältlichen Online- und Atline-Messsystemen für relevante Substrate und Produkte sowie analytischen und datengetriebenen mathematischen Methoden zur Auswertung und Weiterverarbeitung der Signale.

Nach der erfolgreichen Integration der verschiedenen Messsysteme in die Vorortautomatisierung des verwendeten Bioreaktorsystems mussten die Messsignale einer geeigneten Datenvorverarbeitung unterzogen werden. Dabei wurden Methoden zur Ausreissererkennung, diverse Datenfilter und unterschiedliche Interpolationsverfahren eingesetzt und ihre Anwendbarkeit auf die unterschiedlichen Signale beurteilt.

Für die Online-Überwachung des Hauptsubstrates Glukose, der Zelldichte sowie des eingesetzten Reporterproteins GFP auf Basis der 2D-Fluoreszenzspektroskopie wurden unterschiedliche datengetriebene Verfahren aus dem Fachgebiet der Chemometrik eingesetzt. Es wurden dabei sowohl univariate als auch multivariate Methoden zur linearen Modellkalibrierung angewendet und die Hauptkomponentenanalyse sowie die unabhängige Komponentenanalyse als Datenreduktionsmethoden gewählt. Um auch nichtlineare Zusammenhänge zwischen den biologischen Größen und den aufgenommenen Fluoreszenzsignalen berücksichtigen zu können, kamen künstliche neuronale Netze zum Einsatz.

Die besten Approximationen mit einem Fehler von unter 10% konnten mit Hilfe von multivariaten, linearen Modellen unter Verwendung der Original-Fluoreszenzspektren erzielt werden. Etwas größere Modellfehler ergaben sich bei der Verwendung von Hauptkomponenten und unabhängigen Komponenten als Eingangsgrößen der Modelle. Damit konnten hingegen deutlich kürzere Kalibrierzeiten verzeichnet und der Rauschanteil der Trajektorien deutlich verringert werden.

Auch die Anwendung von künstlichen neuronalen Netzen lieferte mit Vorhersagefehlern von 10% gute Ergebnisse, die durch die Vergrößerung des Eingangsdatenraumes noch reduziert werden könnten.

Im letzten Teil der Arbeit wurde ein Modell zur mathematischen Beschreibung des dynamischen Bioreaktionsprozesses entwickelt. Dabei wurde neben der Modellierung des Zellwachstums und der Glukoseaufnahme die Beschreibung der Acetatbildung unter sauerstoffunlimitierten und sauerstofflimitierten Bedingungen implementiert. Die Darstellung der Zielproteinbildung wurde mit Hilfe eines wachstumsabhängigen Anteils sowie über den zellinternen Induktorgehalt realisiert. Dafür wurde der aktive Transport des verwendeten chemischen Induktors IPTG in die Zelle modelliert. Zur mathematischen Beschreibung der Zielproteinakkumulation in der Medienphase wurde zusätzlich eine scherkraftabhängige Lyserate eingeführt. Dabei ergaben sich gute Übereinstimmungen zwischen den Messwerten am Bioreaktionsprozess und den Ausgaben des erstellten Modells.

Schlüsselwörter: *Escherichia coli*, Online-Prozessmonitoring, Bioprozesssimulation, univariate Modelle, multivariate Modelle, Hauptkomponentenanalyse, unabhängige Komponentenanalyse, künstliche neuronale Netze, GFP

Abstract

Christian Kaiser

„Process Monitoring and Process Modelling for Production Processes of Recombinant Proteins with *Escherichia coli*“

Within the scope of this work different methods for process monitoring and process simulation were developed and applied for production processes of recombinant proteins with *Escherichia coli*. The monitoring is based on commercially available on-line and at-line measurement systems for the detection of relevant substrates and products, followed by the application of data driven methods as well as analytical methods for processing of the measuring signals.

After a successful integration of the analysers into the automation of the bioreactor, different methods for data pre-processing were applied and evaluated. These included outlier detection, different filter functions as well as varying interpolation approaches.

Chemometric data driven methods were implemented to realise an online observation of the main carbon source glucose, the cell density and the reporter protein GFP based on 2D-fluorescence spectroscopy. Univariate and multivariate calibration were applied and additionally principal component analysis as well as independent component analysis were used for data mining purposes. Artificial neural networks were applied beside to take non-linearities between biological values and fluorescence measures into account.

Multivariate linear approaches, using the original fluorescence data for model calibration, indicated the best approximation of the process data and led to a root mean square error below 10 %. Slightly larger errors were found using principal components and independent components as model inputs, however the time demand for calibration procedures was significant smaller and the trajectories showed a much better signal to noise ratio.

Errors for artificial neural networks were approximately 10 % and gave suitable results. Error reduction should be feasible using larger input data sets.

Within the last section of this work a dynamic mathematical bioprocess model was developed. Besides the characterisation of cell growth and glucose consumption, acetate formation was taken into account for either sufficient oxygen supply or oxygen limited conditions. Description of target protein formation included a growth dependent term as well as the cell internal inducer content. Therefore the active transport of the used chemical inducer IPTG from the media phase into the cytoplasm was modelled. For a description of target protein accumulation in the media phase an additional shear force dependent lysis rate was established. The measurements were in good accordance with the outputs of the bioprocess model.

Keywords: *Escherichia coli*, on-line process monitoring, bioprocess simulation, univariate modelling, multivariate modelling, principal component analysis, independent component analysis, artificial neural nets, GFP

Nomenklatur – Messtechnik und Bioreaktionsmodell

| | | |
|-------------------------|---|-------------------------------------|
| a | := Parameter zur Anpassung der Trübungsmessung | $[\text{gl}^{-1}]$ |
| b | := Parameter zur Anpassung der Trübungsmessung | $[\text{AU}^{-1}]$ |
| C_{JtotK} | := molare Gesamtkonzentration der Komponente J im Teilsystem K | $[\text{mol l}^{-1}]$ |
| c_{IK} | := Massenkonzentration der Komponente I im Teilsystem K | $[\text{gl}^{-1}]$ |
| f_{bas} | := Parameter der basalen GFP-Produktion | $[-]$ |
| f_{c} | := kritische Frequenz der Kapazitätsmessung | $[\text{Hz}]$ |
| F_{K} | := Volumenstrom in oder aus dem Teilsystem K | $[\text{lh}^{-1}]$ |
| F_{Krel} | := relative Zufütterrate aus dem Teilsystem K | $[\%]$ |
| F_{nl} | := Begasungsrate der Komponente I unter Normbedingungen | $[\text{lh}^{-1}]$ |
| $G_{\text{I/X}}$ | := zellinterner molarer Gehalt der Komponente I | $[\text{mol g}^{-1}]$ |
| $g_{\text{I/X}}$ | := zellinterner Gehalt der Komponente I | $[\text{gg}^{-1}]$ |
| H_{I} | := Henrykoeffizient der Komponente I | $[\text{Nm kg}^{-1}]$ |
| IER | := volumetrische Produktionsrate einer Komponente I aller Zellen | $[\text{gl}^{-1} \text{h}^{-1}]$ |
| ITR | := volumetrische Transferrate einer Komponente I (von G in L) | $[\text{gl}^{-1} \text{h}^{-1}]$ |
| IUR | := volumetrische Aufnahme rate einer Komponente I aller Zellen | $[\text{gl}^{-1} \text{h}^{-1}]$ |
| K_{Iji} | := molare Inhibierungskonstante des Substrates j durch Substrat i | $[\text{mol l}^{-1}]$ |
| K_{I} | := molare Limitierungskonstante der Komponente I | $[\text{mol l}^{-1}]$ |
| k_{Iji} | := Inhibierungskonstante des Substrates j durch Substrat i | $[\text{gl}^{-1}]$ |
| k_{I} | := Limitierungskonstante der Komponente I | $[\text{gl}^{-1}]$ |
| k_{La} | := volumetrischer O_2 -Transferkoeffizient | $[\text{h}^{-1}]$ |
| M_{I} | := Molmasse der Komponente I | $[\text{g mol}^{-1}]$ |
| m_{IK} | := Masse der Komponente I im Teilsystem K | $[\text{g}]$ |
| N_{St} | := Rührerdrehzahl | $[\text{min}^{-1}]$ |
| p_{K} | := Gesamtdruck im Teilsystem K | $[\text{bar}]$ |
| pH | := pH-Wert | $[-]$ |
| p_{O_2} | := relativer Sauerstoffpartialdruck in der Flüssigphase | $[\%]$ |
| Q_{I} | := volumetrische Ver- oder Entsorgungsrate der Komponente I | $[\text{gl}^{-1} \text{h}^{-1}]$ |
| $Q_{\text{J/X}}$ | := molare zellspezifische Reaktionsrate der Komponente J | $[\text{mol g}^{-1} \text{h}^{-1}]$ |
| $q_{\text{I/X}}$ | := zellspezifische Reaktionsrate der Komponente I | $[\text{gg}^{-1} \text{h}^{-1}]$ |
| RQ | := molarer Respirationsquotient | $[\text{mol mol}^{-1}]$ |
| S_{cap} | := Messsignal der Kapazitätsmessung in der Flüssigphase | $[\text{pF cm}^{-1}]$ |
| S_{cap^∞} | := Hintergrundkapazität | $[\text{pF cm}^{-1}]$ |
| $S_{\text{ex/em}}$ | := Fluoreszenzmessung einer Wellenlängenkombination | $[\text{RFU}]$ |
| S_{turb} | := optisches Signal der Trübungsmessung in der Flüssigphase | $[\text{AU}]$ |
| t | := Zeit | $[\text{h}]$ |
| V_{K} | := Volumen des Teilsystems K | $[\text{l}]$ |
| V_{nM} | := Molvolumen unter Gasnormbedingungen | $[\text{l mol}^{-1}]$ |
| v_{vm} | := Gasvolumenstrom pro Flüssigphasenvolumen und Minute | $[\text{l l}^{-1} \text{min}^{-1}]$ |
| x_{IG} | := Stoffmengenanteil der Komponente I in der Gasphase | $[-]$ |
| $w_{\text{C/I}}$ | := Massenanteil von Kohlenstoff in der Komponente I | $[\text{gg}^{-1}]$ |
| $Y_{\text{X/O}}$ | := Ausbeutekoeffizient Zellmasse/Sauerstoff | $[\text{gg}^{-1}]$ |
| $Y_{\text{I/J}}$ | := stöchiometrisches Verhältnis von Komponente I zu J | $[\text{gg}^{-1}]$ |
| $Y_{\text{X/I}}$ | := Ausbeutekoeffizient Zellmasse/Komponente I | $[\text{gg}^{-1}]$ |

Griechische Formelzeichen

| | | |
|----------------|---|---------------------------------|
| $\alpha_{Z/X}$ | := Biofeucht-/Biotrockenmasseverhältnis | [-] |
| ρ_K | := Dichte im Teilsystem K | [g ^l ⁻¹] |
| λ | := Wellenlänge | [nm] |
| ϑ_L | := Temperatur in der Flüssigphase | [°C] |
| μ | := zellspezifische Wachstumsrate | [h ⁻¹] |

Indices

| | |
|--------------------|---|
| 0 | := Anfangsbedingung (Zeitpunkt null) |
| A | := Analysensystem |
| Ac | := Säure (<i>acid</i>) |
| AF | := Antischaum (<i>antifoam</i>) |
| AIR | := Luft |
| Al | := Lauge (<i>alkali</i>) |
| act | := aktiv (<i>active</i>) |
| at | := atline |
| C, CO ₂ | := Kohlenstoff (<i>carbon</i>) / Kohlenstoffdioxid (<i>carbon dioxide</i>) |
| CDW | := Biotrockenmasse (<i>cell dry weight</i>) |
| cal | := Kalibrierbedingungen |
| cap | := Kapazität |
| em | := Emission |
| est | := abgeschätzt (<i>estimate</i>) |
| ex | := Exitation |
| FDA | := Fließdiffusionsanalyse |
| FIA | := Fließinjektionsanalyse |
| G | := Gasphase |
| GSE | := gassensitive Elektrode |
| gr | := Wachstumsanteil (<i>growth</i>) |
| H | := Ernte (<i>harvest</i>) |
| HPLC | := Hochleistungsflüssigkeits-Chromatographie (<i>high performance liquid chromatography</i>) |
| Ind | := Induktor (IPTG) |
| IPTG | := Isopropyl- β -thiogalactopyranosid |
| i, j, k | := Laufvariable für ein Ereignis o.ä. |
| in | := Eingang (<i>input</i>) |
| L | := Reaktionsraum Flüssigphase (<i>liquid</i>) |
| lys | := lysiert |
| M | := Medienphase |
| m | := Maintenanceanteil |
| max | := Maximalwert |
| min | := Minimalwert |
| n | := Gasnormbedingungen |
| P1 | := Produkt 1 (Zielprotein GFP) |
| P3 | := Produkt 3 (Acetat) |
| Ptot | := Gesamtprotein |

| | | |
|------|----|--|
| R1 | := | Reservoir 1 (Glukose) |
| R2 | := | Reservoir 2 (Di-Ammoniumhydrogenphosphat-Lsg.) |
| S | := | Probenahme (<i>sample</i>) |
| S1 | := | Substrat 1 (Glukose) |
| S3 | := | Substrat 3 (Acetat) |
| T1 | := | Titration 1 (Säure) |
| T2 | := | Titration 2 (Lauge) |
| tot | := | Gesamtmenge dissoziierfähiger Größen |
| turb | := | Trübung (<i>turbidity</i>) |
| V | := | Verdampfungsverlust bei Begasung (<i>volatility</i>) |
| w | := | Sollwert |
| X | := | Biotrockenmasse |
| Z | := | Biofeuchtmasse |

Nomenklatur – Datenvorverarbeitung und Modellkalibrierung

| | | | |
|-----------------------------|----|--|----------------|
| B | := | Koeffizientenmatrix der inversen Kalibrierung | |
| b_I | := | Koeffizienten der inversen Kalibrierung für die Komponente I | |
| C | := | Kovarianzmatrix | |
| c_{IK} | := | Gelöstkonzentration der Komponente I im Teilsystem k | [$g l^{-1}$] |
| cov | := | Kovarianz | |
| E_x | := | Matrix der Modellfehler der klassischen Kalibrierung | |
| e_{Ix} | := | Modellfehler der klassischen Kalibrierung für die Komponente I | |
| E_y | := | Matrix der Modellfehler der inversen Kalibrierung | |
| e_{Iy} | := | Modellfehler der inversen Kalibrierung für die Komponente I | |
| $F(x)$ | := | Verteilungssummenfunktion der Größe x | |
| m | := | Parameter der Filterbreite | |
| n | := | Anzahl Messungen | |
| NORM | := | Summe der Koeffizienten des Savitzky-Golay Filters | |
| R | := | Korrelationsmatrix | |
| r_{jk} | := | Pearson'scher Korrelationskoeffizient zwischen x_j und x_k | |
| r_{min} | := | untere Grenze für den Pearson'schen Korrelationskoeffizienten | |
| RMSE | := | Root Mean Square Error (Wurzel des quadratischen Fehlers) | |
| RMSEC | := | Root Mean Square Error of Calibration (Kalibrierungsfehler) | |
| RMSEP | := | Root Mean Square Error of Prediction (Vorhersagefehler) | |
| S | := | Koeffizientenmatrix der klassischen Kalibrierung | |
| s_i | := | Koeffizienten der klassischen Kalibrierung | |
| T_i | := | Prüfgröße des Grubbs Tests | |
| $T_{n;1-\alpha}$ | := | kritischer Wert des Grubbs Tests | |
| $T_{\frac{\alpha}{2n};n-2}$ | := | kritischer Wert der Student t-Verteilung mit Freiheitsgrad n-2 | |
| X | := | Messmatrix des BioView [®] -Messgerätes | [RFU] |
| x_{ik} | := | Messwert der i-ten Zeile und der k-ten Spalte | [RFU] |
| x_j, x_k | := | Messvektor der j-ten bzw. k-ten Spalte | [RFU] |
| \bar{x}_k | := | Mittelwert der k-ten Spalte | [RFU] |

Griechische Formelzeichen

| | | |
|------------|----|--------------------|
| Σ | := | Summe |
| σ | := | Standardabweichung |
| σ^2 | := | Varianz |

Indices und Exponenten

| | | |
|------------|----|------------------------------------|
| * | := | vorhergesagte Größe |
| cen | := | zentrierte Daten |
| i, j, k, m | := | Laufvariable für ein Ereignis o.ä. |
| sim | := | simulierte Daten |
| skal | := | bereichsskalierte Daten |
| stn | := | standardisierte Daten |
| T | := | transponierte Matrix |

Nomenklatur – Hauptkomponentenanalyse

| | | |
|-------|----|--|
| D | := | quadrierte Diagonalmatrix S der Hauptkomponentenanalyse |
| P | := | Ladungsmatrix der Hauptkomponentenanalyse |
| p_i | := | Ladungsvektor der i-ten Spalte der Hauptkomponentenanalyse |
| S | := | reelle Diagonalmatrix der Singulärwertzerlegung |
| T | := | Factorscorematrix der Hauptkomponentenanalyse |
| t_i | := | Scorevektor der i-ten Spalte der Hauptkomponentenanalyse |
| U | := | unitäre Matrix der Singulärwertzerlegung |
| V | := | adjungierte Matrix der Singulärwertzerlegung |

Griechische Formelzeichen

| | | |
|--------------|----|-------------------------------|
| λ_i | := | Eigenwert i |
| σ_e^2 | := | erklärte (kumulative) Varianz |

Nomenklatur – Unabhängige Komponentenanalyse

| | | |
|----------|----|--|
| A | := | Mischmatrix der unabhängigen Komponentenanalyse |
| a_1 | := | Parameter der Kontrastfunktion |
| a_i | := | Element der Zufallsvariable Y |
| a_{ij} | := | Koeffizient der Mischmatrix A |
| c | := | Konstante des Fast-Fixed-Point Algorithmus |
| cosh | := | Cosinushyperbolicus |
| E | := | Erwartungswert |
| G | := | Kontrastfunktion zur Approximation der Negentropie |
| g | := | erste zeitliche Ableitung der Kontrastfunktion G |
| H | := | Entropie |
| I | := | mutual Information |
| I_q | := | Cluster-Qualitätsindex |
| J | := | negative Entropie (Negentropie) |

| | | |
|--------------------|----|--|
| J_G | := | Gütefunktional des Fast-Fixed-Point Algorithmus |
| kurt | := | Kurtosis |
| log | := | dekadischer Logarithmus |
| P | := | Verteilungssummenfunktion |
| p | := | Wahrscheinlichkeitsdichtefunktion |
| s_i | := | Quellsignale der Mischung |
| tanh | := | Tangenshyperbolicus |
| u | := | Argument der Kontrastfunktion |
| W | := | Inverse der Mischmatrix |
| w_i | := | Vektor der i -ten Zeile der inversen Mischmatrix |
| x | := | Zufallsvariable |
| x_j | := | Stichprobe der Zufallsvariable x |
| Y | := | diskrete Zufallsvariable |
| y | := | kontinuierliche Verteilungsfunktion |
| y_{GAUSS} | := | Gaussfunktion |

Griechische Formelzeichen

| | | |
|-----|----|---|
| v | := | Standardnormalverteilte Zufallsvariable |
|-----|----|---|

Nomenklatur – Künstliche Neuronale Netze

| | | |
|------------------|----|--|
| E_q | := | Fehlervektor einer Schicht q eines neuronalen Netzes |
| e_{qr} | := | Fehler eines Neurons r in der Schicht q |
| f_{act} | := | Aktivierungsfunktion |
| MSE | := | quadratischer Fehler (<i>Mean Square Error</i>) |
| net_r | := | Netzeingabe (Propagierungsfunktion) eines Neurons r |
| $O^{[q]}$ | := | Eingabevektor der Schicht q eines Neuronalen Netzes |
| o_{pq} | := | Eingabewert eines Neurons in der Schicht q aus dem vorgelagerten Neuron p |
| t_{qr} | := | tatsächlicher Ausgabewert des Neurons r in der Schicht q |
| $W^{[q]}$ | := | Gewichtsvektor der Schicht q eines neuronalen Netzes |
| w_{pq} | := | Gewicht eines Neurons p in der Schicht q |
| X | := | Eingabevektor eines neuronalen Netzes |
| x_k | := | Eingabewert eines Neurons |
| Y | := | Ausgabevektor eines neuronalen Netzes |
| y_k | := | Ausgabewert eines Neurons |

Griechische Formelzeichen

| | | |
|------------|----|--|
| δ_r | := | lokaler Fehler eines Neurons r |
| Θ_r | := | Schwellenwert (Bias) eines Neurons r |

Inhaltsverzeichnis

| | | |
|----------|--|-----------|
| 1 | Einleitung | 3 |
| 1.1 | Hintergrund und Motivation | 3 |
| 1.2 | Zielsetzung der Arbeit | 4 |
| 2 | Mikro- und molekularbiologische Grundlagen | 6 |
| 2.1 | Verwendete <i>Escherichia coli</i> -Stämme und Plasmide | 6 |
| 2.1.1 | <i>Escherichia coli</i> JM105: pGFP | 6 |
| 2.1.2 | <i>Escherichia coli</i> BL21(DE3): pRSETB | 7 |
| 2.1.3 | <i>Escherichia coli</i> DH5 α : pTrcHisB | 8 |
| 2.2 | Das Modellprotein GFP | 9 |
| 2.3 | Stoffwechsellvorgänge von <i>Escherichia coli</i> | 11 |
| 3 | Monitoring von Bioreaktionsprozessen | 15 |
| 3.1 | Prozessüberwachung in der Biotechnologie | 15 |
| 3.2 | Standard Messtechnik | 18 |
| 3.3 | Erweiterte Inline-Messtechnik | 19 |
| 3.3.1 | Möglichkeiten der Inline-Instrumentierung | 19 |
| 3.3.2 | Trübungsmessung | 21 |
| 3.3.3 | Kapazitäts- und Leitfähigkeitsmessung | 22 |
| 3.3.4 | 2D-Fluoreszenzspektroskopie | 24 |
| 3.4 | Eingesetzte Atline-Messtechnik | 28 |
| 3.4.1 | Mögliche Atline-Applikationen | 28 |
| 3.4.2 | Glukoseanalytik mit der ProcessTrace | 32 |
| 3.4.3 | Ammoniumanalytik mit der TAS2000 | 33 |
| 3.4.4 | Atline-Messung mittels HPLC | 34 |
| 4 | Automatisierungs- und Datenverarbeitungstechnik | 35 |
| 4.1 | Gewählter Automatisierungsaufbau | 35 |
| 4.2 | Anbindung der erweiterten Analysetechnik | 36 |
| 4.3 | Das SCADA-System MFCS/win | 37 |
| 4.4 | Automatisierung des Bioreaktionsprozesses | 39 |
| 5 | Chemometrische Methoden und Modelle | 42 |
| 5.1 | Motivation | 43 |
| 5.2 | Erweiterte Datenverarbeitung über Einbindung von MATLAB [®] | 44 |
| 5.3 | Datenvorverarbeitung | 45 |
| 5.3.1 | Dublettenentfernung | 45 |
| 5.3.2 | Ausreissererkennung | 47 |
| 5.3.3 | Datenfilterung | 50 |

| | | |
|----------|---|------------|
| 5.3.4 | Dateninterpolation | 51 |
| 5.4 | Kalibrierung linearer univariater Modelle | 53 |
| 5.4.1 | Datenzentrierung und Skalierung | 54 |
| 5.4.2 | Methodik der linearen Modellkalibrierung | 55 |
| 5.4.3 | Ergebnisse der univariaten Modellkalibrierung | 59 |
| 5.5 | Kalibrierung linearer multivariater Modelle | 67 |
| 5.5.1 | Hauptkomponentenanalyse | 69 |
| 5.5.2 | Unabhängige Komponentenanalyse | 77 |
| 5.5.3 | Ergebnisse der multivariaten Modellkalibrierung | 85 |
| 5.6 | Nichtlineare Modellierung mit künstlichen neuronalen Netzen | 90 |
| 5.6.1 | Aufbau neuronaler Netze | 91 |
| 5.6.2 | Training neuronaler Netze | 94 |
| 5.6.3 | Lernalgorithmen | 96 |
| 5.6.4 | Ergebnisse der nichtlinearen Modelle mit ANNs | 99 |
| 6 | Modellierung des dynamischen Bioreaktionsprozesses | 103 |
| 6.1 | Das Dreiphasenmodell | 106 |
| 6.2 | Die allgemeinen Bilanzgleichungen | 108 |
| 6.3 | Verwendete Reaktionskinetische Ansätze | 111 |
| 6.3.1 | Mathematische Beschreibung des Overflow-Metabolismus | 111 |
| 6.3.2 | Segregierter Modellansatz für die Zielproteinbildung | 120 |
| 7 | Zusammenfassung | 132 |
| 8 | Anhang | 134 |
| 8.1 | Verwendete Medien und Reagenzien | 134 |
| 8.1.1 | Vorkultur- und Batchmedium | 134 |
| 8.1.2 | Feedmedium | 135 |
| 8.1.3 | Reagenzien für die Ammoniumbestimmung mit der TAS2000 | 135 |
| 8.2 | Ergänzende Offline-Messtechnik | 136 |
| 8.2.1 | Optische Dichte Messung | 136 |
| 8.2.2 | Biotrockenmassebestimmung | 136 |
| 8.2.3 | Zielproteinanalytik | 136 |
| 8.2.4 | Ammoniumbestimmung | 137 |
| 8.2.5 | Acetat- und Glukosemessung über enzymatische Testkits | 137 |
| 8.2.6 | Keimzahlbestimmung | 138 |
| 8.3 | Modellparameter der Chemometrischen Modelle | 138 |
| 8.3.1 | Univariate Modelle mit Originalvariablen | 138 |
| 8.3.2 | Multivariate Modelle mit Originalvariablen | 139 |
| | Literaturverzeichnis | 140 |

1 Einleitung

Seitdem es 1972 der Arbeitsgruppe um den späteren Nobelpreisträger Paul Berg gelang Restriktionsenzyme zur Herstellung rekombinanter DNA zu verwenden [Jackson *et al.* (1972)] und 1973 das erste Mal ein gentechnisch veränderter *Escherichia coli* Stamm cloniert wurde [Cohen *et al.* (1973)], hat sich die Biotechnologie rasant weiterentwickelt. Das Potential der Rekombination wurde durch die Industrie schnell erkannt und bereits 1982 mit dem von Genentech und Eli Lilly produziertem Insulin das erste rekombinante Protein durch die amerikanische *Food and Drug Administration*¹ für den Markt zugelassen.

Heute sind in Europa über 130 Medikamente kommerziell erhältlich die von Wachstumsfaktoren über Antikörper und Enzyme bis zu Gerinnungsfaktoren reichen. Die meisten werden dabei unter Verwendung von Bakterien, Hefen, Hamster- oder humanen Zelllinien hergestellt [Dingermann and Zündorf (2007)].

Trotz der globalen Marktturbulenzen im Jahr 2008 konnten die börsennotierten Biotech-Unternehmen ihre Umsätze um 12% auf 89.7 Mrd. US\$ steigern. Der Reinverlust sank um 53% auf 1.4 Mrd. US\$ und der Gesamtsektor der Vereinigten Staaten von Amerika konnte erstmals einen Reingewinn von 0.4 Mrd. US\$ verbuchen [Ernst & Young (2009)]. Trotz eines deutlichen Einbruchs bei den Kreditaufnahmen gilt die Biotechnologie nach wie vor als Wachstumsmarkt und als die Schlüsseltechnologie des 21ten Jahrhunderts.

Sie hat sich dabei zu einem interdisziplinären Betätigungsfeld unterschiedlichster Fachrichtungen entwickelt. Neben diversen Disziplinen der Biologie finden insbesondere in der Produktion rekombinanter Produkte klassische ingenieurwissenschaftliche Fächer wie die Verfahrenstechnik und die Mess- und Regelungstechnik ihre Anwendung. Durch das immer detailliertere Wissen über die Prozesse und die damit verbundene immer größer werdende Datenflut sowie einen höheren Automatisierungsgrad wird die Nachfrage nach Fachkräften mit Informatikhintergrund in der Bioverfahrenstechnik immer größer.

1.1 Hintergrund und Motivation

Der finanzielle, zeitliche, apparative und regulatorische Aufwand, den die herstellende Industrie für die Zulassung neuer Pharmazeutika betreiben muss, ist enorm hoch. So dauert der Prozess vom Wirkstoff bis hin zum zugelassenen Medikament ungefähr 10 bis 12 Jahre und ist mit einem Kostenaufwand von ca. 800 Mio. US\$ verbunden [DiMasi *et al.* (2003)]. Dabei unterliegt die Entwicklung eines geeigneten Herstellungsprozesses ebenso wie die anschließende Wirkstoffproduktion nach erfolgreicher Zulassung strengen GMP²- und GLP³-Richtlinien.

¹FDA – *Food and Drug Administration*

²GMP – *Good Manufacturing Practice*

³GLP – *Good Laboratory Practice*

Der regulatorische Aufwand, den die pharmazeutische Industrie für den Zulassungsprozess neuer Arzneimittel betreiben muss, ist über die Jahre kontinuierlich angestiegen. Trotz der hohen empirischen Standards die dabei durch die FDA gesetzt wurden, gab es Hinweise, dass bei der regelmäßigen Überprüfung der laufenden Produktionen immer häufiger Warnbriefe, Rückrufe und sogar Produktionsunterbrechungen nötig waren [Dünnebier and Tups (2007)].

Die Intention der FDA die Effizienz der Pharmaproduktion bei gleichzeitiger Sicherstellung der Produktqualität zu realisieren, veranlasste die FDA zur Veröffentlichung des Entwurfs *PAT⁴ – a Framework for Innovative Pharmaceutical Manufacturing and Quality Assurance* [FDA (2004)]. Dieser stellt einen Paradigmenwechsel für die pharmazeutische Produktion dar und fordert ein wissensbasiertes Prozessverständnis auf Grundlage von

- modernen Prozessmessgeräten und Analysensystemen,
- multivariaten Datenerfassungs- und Auswertewerkzeugen,
- Überwachungs- und Regelungssystemen zum fortlaufenden Prozessmonitoring und zur Einhaltung von optimalen Prozessbedingungen sowie
- Methoden zur laufenden Weiterentwicklung der eingesetzten Systeme und des Wissensmanagements.

Viele der zur Umsetzung der PAT-Initiative einsetzbaren Verfahren sind dabei bereits aus anderen Fachgebieten wie z.B. der chemischen Produktion bekannt. Sie reichen von empirischen Methoden wie die Risikoanalyse und Ursachenforschung über statistische Versuchsplanung zur Identifikation der die einzelnen Prozessschritte signifikant beeinflussenden Parameter (KPIs⁵) bis zum Einsatz geeigneter Prozessführungsstrategien auf Basis dieser KPIs.

Ebenso beinhalten sie geeignete Online-Messtechniken sowie mechanistische und datengetriebene Modelle, die sowohl zur Prozessoptimierung als auch zur Prozessüberwachung eingesetzt werden können. Eine nähere Beschreibung der hier aufgeführten Methoden sowie Anwendungsbeispiele aus der pharmazeutischen Praxis finden sich bei Dünnebier and Tups (2007).

1.2 Zielsetzung der Arbeit

Die Zielsetzung dieser Arbeit war es, einen Beitrag zur PAT-Initiative zu leisten. Dabei sollten Wege aufgezeigt werden, wie die unterschiedlichen von der FDA genannten Ideen in der Forschung und Entwicklung des *Upstream-Processings* mit rekombinanten *Escherichia coli* eingesetzt werden können.

Im ersten Schritt kann dafür der Automatisierungsgrad der zu untersuchenden Fed-Batch Kultivierungen erhöht werden, um somit die Variationsbreite zwischen den Prozessen zu verringern und die Reproduzierbarkeit der Experimente zu erhöhen.

Um eine detailliertere Prozessüberwachung realisieren zu können, sollten zusätzlich zur Standardinstrumentierung weitere kommerziell erhältliche Analysensysteme an den Bioreaktor angebunden werden, um Schlüsselgrößen des Herstellungsprozesses echtzeitnah beobachten zu können.

⁴PAT – *Process Analytical Technology*

⁵KPI – *key performance indicators*

Für die Verarbeitung der gewonnenen Daten mussten dafür geeignete Methoden zur Datenaufbereitung erprobt und etabliert werden. Die anschließende Extraktion der relevanten Informationen aus den Prozessdaten sollte unter Einsatz verschiedener Verfahren zur Informationsextraktion und Datenkompression erzielt werden. Diese sollten mit Hilfe der Entwicklungsumgebung MATLAB[®] auf die verschiedenen Messsignale angewendet und ihre Güte beurteilt werden.

Für die Online-Überwachung der Zielproteinbildung während des laufenden Prozesses war in dieser Arbeit das grün fluoreszierende Reporter-Protein GFP als rekombinantes Produkt vorgegeben. Daher bot es sich an unter Einsatz der 2D-Fluoreszenzspektroskopie datengetriebene multivariate Modelle zu entwickeln, die neben der Zielproteinexpression auch weitere Schlüsselgrößen des Prozesses abbilden können.

Zur Beschreibung von Nichtlinearitäten zwischen den Prozessvariablen und den entsprechenden Messsignalen sind lineare Modelle nicht geeignet. Eine adäquate Methode um diese zu beschreiben stellen dabei Künstliche Neuronale Netze dar. Diese bieten zusätzlich den Vorteil, dass sie mit jedem zusätzlichem Kultivierungsdatensatz weiter trainiert und die Vorhersagegenauigkeiten verbessert werden können.

Die Prozesssynthese und Optimierung wird oft auf der Basis von mechanistischen Ansätzen durchgeführt. Diese basieren, im Gegensatz zu den datengetriebenen Verfahren, auf den grundlegenden Konzepten der Energie- und Massenerhaltung, den Phasengleichgewichten sowie den postulierten Reaktions- und Wachstumskinetiken. Um diese beschreiben zu können bedarf es eines dynamischen Prozessmodells. Zur Erstellung des Modells sollte eine systemtheoretische Analyse des Bioreaktionsprozesses durchgeführt werden, um darauf aufbauend wichtige Prozessgrößen mathematisch beschreiben und simulieren zu können.

2 Mikro- und molekularbiologische Grundlagen

Für die Herstellung rekombinanter Produkte stehen verschiedene Expressionssysteme zur Verfügung. Neben prokaryotischen Host-Systemen wie z.B. *Escherichia coli* sowie unterschiedliche Arten der *Bacillus*-Gattung werden außerdem Eukaryonten wie z.B. Hefen (*Pichia pastoris*, *Hansenula polymorpha*, etc.) und Pilze (*Aspergillus*), tierische (z.B. CHO¹) und humane aber auch pflanzliche Zellen eingesetzt.

Das fakultativ anaerobe, gramnegative, stäbchenförmige Bakterium *Escherichia coli* ist dabei das am häufigsten für die Expression rekombinanter Proteine eingesetzte System [Anderson and Krummen (2002)]. Dies liegt zum Einen daran, dass es hervorragend charakterisiert ist und eine große Anzahl an Expressionsvektoren zur Verfügung stehen [Swartz (2001)]. Zum Anderen macht das sehr schnelle Wachstum und die Möglichkeit hoher Zelldichten sowie die geringen Kosten für die benötigten Fermentationsmedien die Prozesse aus wirtschaftlichen Gründen sehr attraktiv.

Die in dieser Arbeit eingesetzten *E. coli* Stämme, Vektoren und rekombinanten Proteine werden im Folgenden vorgestellt. Bei dem rekombinanten Produkt handelt es sich dabei immer um das grün fluoreszierende Protein GFP² bzw. GFP-Derivate. Alle Systeme sind durch den chemischen Induktor Isopropyl- β -D-thiogalaktopyranosid (IPTG) induzierbar und weisen eine Ampicillinresistenz auf.

In weiteren Teilen dieses Kapitels werden dann Produkteigenschaften und kurz die grundlegenden Stoffwechselfgänge von *Escherichia coli* erläutert.

Die Kultivierungsbedingungen und der automatisierte Prozessablauf sind in Abschnitt 4.4 aufgezeigt. Die Zusammensetzungen der verwendeten Minimalmedien befindet sich im Anhang.

2.1 Verwendete *Escherichia coli*-Stämme und Plasmide

2.1.1 *Escherichia coli* JM105: pGFP

Zunächst wurde ein apathogener, gentechnisch veränderter, *E. coli* JM105-Stamm eingesetzt, der das Plasmid pGFP der Fa. CLONTECH, Palo Alto, CA, USA, trägt [Chalfie *et al.* (1994)]. Es handelt sich dabei um ein Thiamin auxotrophes K12 Derivat, das auf Minimalmedium wächst. Durch ein Klonierungsexperiment trägt der Stamm ein intaktes lacZ-Gen. Die längerfristige Stammhaltung erfolgte in Glycerol-Kryokulturen bei -80 °C.

¹CHO – Chinese Hamster Ovary

²GFP – Green Fluorescent Protein

Bei dem in Bild 2.1a gezeigten Plasmid handelt es sich um ein Hybrid aus dem Klonierungsvektor pUC19 und dem Gen für das grün fluoreszierende Protein. Laut Hersteller liegt das 3.344 kbp³ große Plasmid mit 300-500 Kopien pro Zelle vor und ist damit ein High Copy-Plasmid.

Bei der Expression wird ein Fusionsprotein aus einigen wenigen Aminosäuren der β -Galactosidase und GFP von den Zellen gebildet. Das Plasmid enthält den Replikationsursprung ori und die Ampicillinresistenz von pUC19, wobei der Genabschnitt des GFP unter der Kontrolle des lac-Promotors von pUC19 steht [Friehs (1999)].

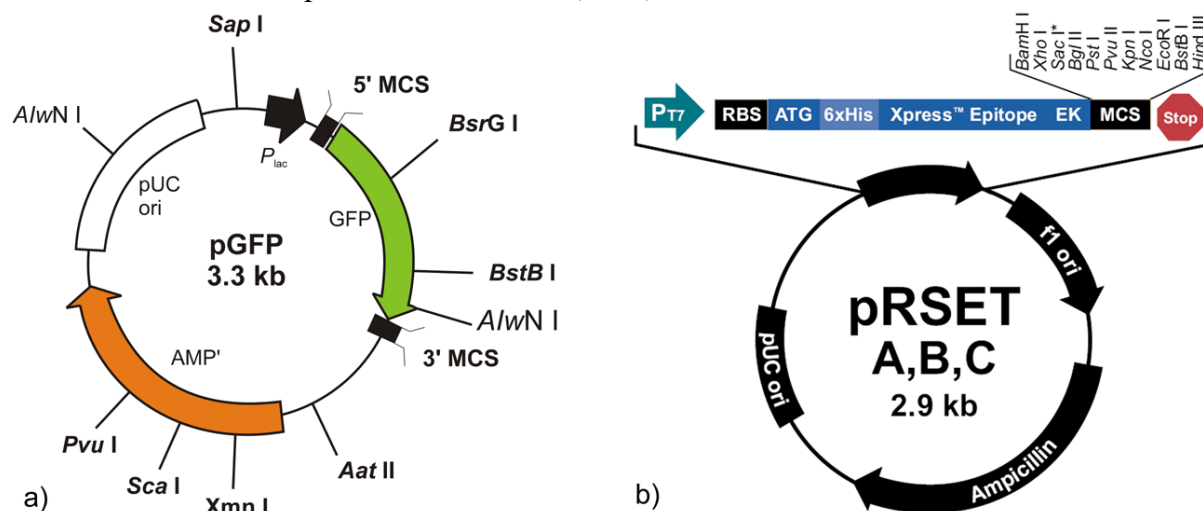


Bild 2.1: Die eingesetzten Expressionsplasmide a) pGFP und b) pRSETB [Invitrogen (2006a)]

2.1.2 *Escherichia coli* BL21(DE3): pRSETB

Bei dem in Bild 2.1b dargestellten Expressionsplasmid pRSETB der Fa. Invitrogen, Carlsbad, CA, USA, handelt es sich um ein 2,9 kbp großes Plasmid, welches als Low Copy-Plasmid in etwa 20 Kopien in der Zelle vorliegt [Invitrogen (2006a)]. Es besitzt als Marker eine Ampicillinresistenz sowie zur einfachen Proteinaufreinigung einen 6xHis-Tag.

Hinter dem His-Tag befindet sich zusätzlich ein von Invitrogen patentiertes Xpress[®]-Epitop, welches unter Verwendung eines entsprechenden Antikörpers z.B. zur Proteinquantifizierung mittels ELISA verwendet werden kann. Die folgende *Enterokinase Cleavage Site* (EK) ermöglicht die effektive Entfernung des Fusionstags. Unter Verwendung der *Multiple Cloning Site* (MCS) kann die gewünschte genetische Information für die Expression rekombinanter Produkte in den Vektor eingebracht werden [Schoepfer (1993)]. Im hier vorliegenden Fall war dieses das Gen des GFP-Wildtyps, welches nicht in Bild 2.1b dargestellt ist.

Durch das Vorhandensein eines T7-Promotors im Vektor vor dem klonierten Gen werden hohe Mengen an Messenger-RNA produziert. Im verwendeten Expressionstamm muss allerdings zur Initialisierung der Transkription das Gen der T7-Polymerase im Genom vorliegen, was beim eingesetzten *Escherichia coli* BL21 (DE3) Stamm gegeben ist [Johnston *et al.* (2000)].

Der Prozess der Induktion untergliedert sich dann in zwei Teilschritte. Durch Zugabe von IPTG wird das Gen der T7 RNA-Polymerase exprimiert. Die entstehende T7-Polymerase lagert

³kbp – kilo base pairs

sich im zweiten Schritt an den auf dem Plasmid befindlichen T7-Promotor an und startet die Ablesung des klonierten Genabschnittes. Die Bildung der mRNA des Zielproteins wird also nicht direkt durch die Zugabe von IPTG induziert, sondern indirekt [Studier *et al.* (1990)].

2.1.3 *Escherichia coli* DH5 α : pTrcHisB

In Bild 2.2 ist das eingesetzte Expressionsplasmid pTrcHisB der Fa. Invitrogen dargestellt. Wie pRSETB trägt es das Gen für eine Ampicillinresistenz. Es enthält den künstlichen starken trc-Promotor. Dabei handelt es sich um ein Hybrid aus der -35 Region des Tryptophan- und der -10 Region des lacUV5-Promotors. Die DNA-Polymerase von *Escherichia coli* erkennt diese Sequenz und bindet mit einer hohen Affinität, was hohe Expressionraten des Zielgens hervorruft [Invitrogen (2006b)].

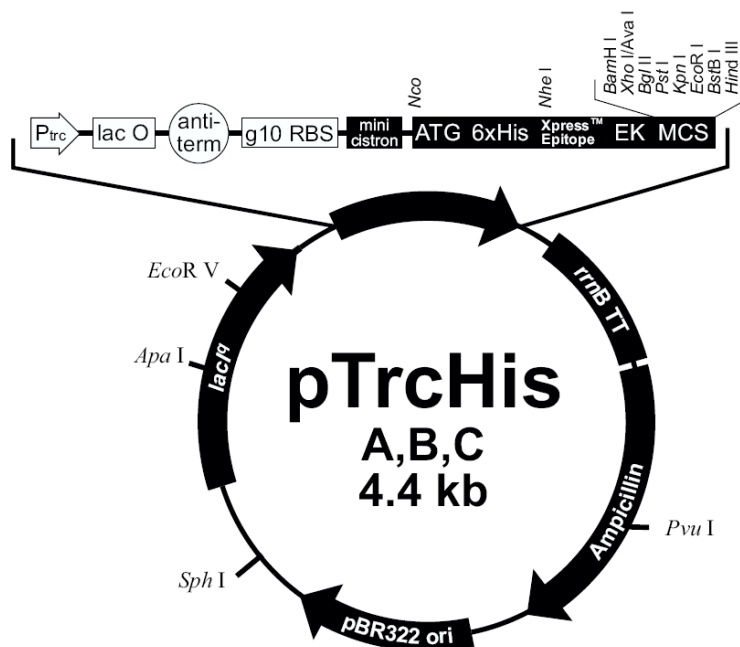


Bild 2.2: Das verwendete Expressionsplasmid pTrcHisB [Invitrogen (2006b)]

schweig) eingesetzt werden. Als rekombinantes Produkt wurde dabei das GFP-Derivat T-Sapphire gewählt, welches im Folgenden näher beschrieben wird.

Das T-Sapphire Gen wurde freundlicherweise von Herrn Dr. Oliver Griesbeck aus der Forschungsgruppe des Max-Planck Institutes für Neurobiologie in Martinsried, München, in den Vektor pRSETB (Fa. Invitrogen) kloniert und zur Verfügung gestellt. Die Integration in den verwendeten DH5 α -Stamm wurde dann von Frau Sabine Knappe (Arbeitsgruppe von Prof. Dr. Christiana Cordes am Fachbereich Chemie / Pharmatechnik der Hochschule Magdeburg-Stendal) vorgenommen und für diese Arbeit zur Verfügung gestellt.

Der Vektor enthält ebenfalls das zuvor beschriebene Xpress[®]-Epitop sowie eine *Enterokinase Cleavage Site* und eine *Multiple Cloning Site*. Das Plasmid trägt zusätzlich das lacI-Gen, dessen Genprodukt dafür sorgt, dass der Promotor ohne Induktion geschlossen bleibt [Invitrogen (2006b)]. Obwohl als solches beschrieben, wurde in dieser Arbeit eine gewisse Durchlässigkeit beobachtet, also eine Expression ohne Induktion. Der Promotor ist somit nicht vollständig geschlossen.

Da die in *E. coli* natürlich vorkommende DNA-Polymerase zur Genexpression verwendet wird, konnte zur Expression der Stamm DH5 α (6897, DSMZ Braunschweig)

2.2 Das Modellprotein GFP

Um ein Online-Monitoring der Zielproteinexpression während des Kultivierungsprozesse zu realisieren, wurde als Modellprotein das grün fluoreszierende Protein GFP verwendet. Eine Reihe von Autoren [Cha *et al.* (2000)][DeLisa *et al.* (1999)] konnten bereits zeigen, dass sich dieses unter Verwendung von Fluoreszenzsensoren zur quantitativen Beobachtung der intrazellulären Produktmessung im Bioreaktor eignet.

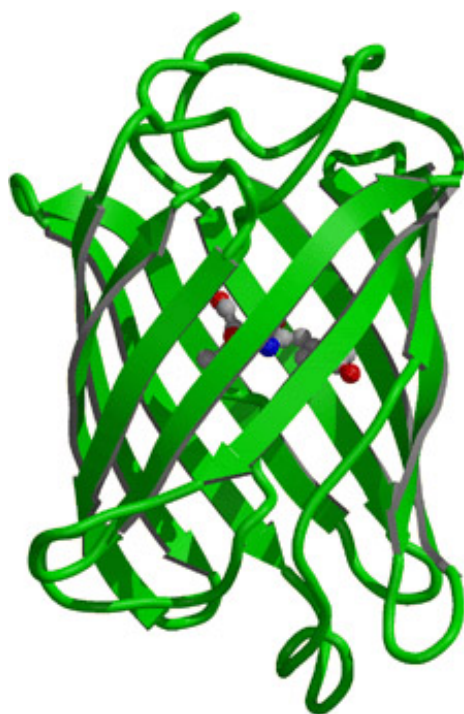


Bild 2.3: Räumliche Struktur des GFP-Moleküls [Wachter (2008)]

Das Wildtyp-GFP (wtGFP) wurde erstmals 1962 von Shimomura *et al.* (1962) bei dem Versuch entdeckt, die lumineszierende Komponente Aequorin aus der Tiefseequalle *Aequorea victoria* zu isolieren. Im Meeresorganismus wird das Chromophor des wtGFPs durch die Lichtemission der Chemielumineszenzreaktion des Photoproteins Aequorin bei 470 nm angeregt und emittiert bei 508 nm [Morise *et al.* (1974)].

Die DNA-Sequenz des aus 238 Aminosäuren bestehenden und 27 kDa schweren Proteins wurde 1992 von Prasher *et al.* (1992) veröffentlicht. Die in Bild 2.3 dargestellte einzigartige dreidimensionale β -Can Struktur besteht aus einer einzelnen Polypeptidkette, die einen ca. 40 Å langen Zylinder mit einem Durchmesser von 30 Å bildet. Das Innere des aus elf antiparallelen β -Faltblattstrukturen gebildeten Zylinders wird durch eine α -Helix durchzogen, in dessen Mitte sich das Chromophor befindet [Yang *et al.* (1996)]. Das Fluorophor wird dadurch vom umgebenden Lösungsmittel abgeschirmt und ist dadurch vor Fluoreszenzlöschung, die z.B. durch Sauerstoff hervorgerufen werden kann, geschützt, was die hohe Quantenausbeute des GFPs erklärt [Ormö *et al.* (1996)].

Weitere Vorzüge von GFP sind die hohe Stabilität über einen großen pH-Bereich (pH 5 - 12) und die erst bei hohen Temperaturen ($\vartheta > 65^{\circ}\text{C}$) einsetzende Denaturierung des Proteins [Yang *et al.* (1996)]. Zusätzlich ist es nahezu resistent gegenüber Proteasen und Detergenzien wie z.B. 1 % SDS und 8 M Harnstoff [Ward *et al.* (1982)].

Da für die Ausbildung des Chromophores keine spezifischen Enzyme oder Cofaktoren benötigt werden, sondern die Bildung autokatalytisch aus dem Polypeptid-Rückrad erfolgt, ist die rekombinante Expression des Proteins in verschiedenen Organismen möglich [Cubitt *et al.* (1995)]. Eine Fusion mit anderen Proteinen ist sowohl C- als auch N-terminal ohne Verlust der Fluoreszenz möglich und stellt einen weiteren Grund für den vielfältigen Einsatz von GFP dar [Chalfie (1995)][Chalfie *et al.* (1994)].

Wie in Bild 2.4 dargestellt, besitzt das Wildtyp-GFP zwei Anregungsmaxima bei 395 nm und bei 470 nm sowie ein Emissionsmaximum bei 509 nm. Die zwei Anregungsmaxima kommen dadurch zustande, dass das Fluorophor sowohl in der protonierten Hydroxyl-Form ($\lambda_{\text{ex}} = 395 \text{ nm}$) als auch in der deprotonierten Enolat-Form ($\lambda_{\text{ex}} = 470 \text{ nm}$) vorliegt. Dabei liegt das chemische Gleichgewicht auf der Seite der Hydroxyl-Form, was zu einer höheren Intensität

des ersten Anregungsmaximums führt. Da die Hydroxyl-Form im angeregten Zustand zunächst in die deprotonierte Form übergeht, bevor es Licht emittiert, geht die Lichtemission in beiden Anregungsfällen von der angeregten deprotonierten Form aus, wodurch GFP nur ein Emissionsmaximum aufweist. Der zyklische Prozess zwischen Photonenabsorption, Deprotonierung, Photoemission und Aufnahme eines Protons wird als „Förster-Zyklus“ bezeichnet und wurde erstmals 1948 beschrieben [Förster (1948)].

Da die notwendige Autooxidation zur Ausbildung der Fluoreszenzeigenschaften des Chromophors nahezu zwei Stunden dauert, die Fluoreszenzausbildung relativ schwach ist und die Proteinfaltung eine starke pH und Temperaturabhängigkeit zeigt [Tsien (1998)], ist das Wildtyp GFP nur bedingt als Fluoreszenzmarker im Prozess einsatzfähig. Es stehen heutzutage jedoch eine ganze Reihe von GFP-Varianten zur Verfügung, die sich nach Tsien (1998) in sieben unterschiedliche Klassen einteilen lassen.

Für das Online-Monitoring wurde eine Variante benötigt, die vor allem eine schnelle und stabile Chromophorausbildung sowie eine hohe Sensitivität aufweist. Ausserdem mussten aufgrund der Beschaffenheit des verwendeten BioView[®]-Spektroskopes die Anregungs- und Emissionsmaxima mindestens 40 nm auseinanderliegen um eine zuverlässige Messung zu gewährleisten. Deshalb wurde eine GFP-Mutante aus der dritten Klassifikation, das sogenannte T-Sapphire GFP, ausgewählt. Das eingesetzte T-Sapphire stellt eine Mutante des Sapphire GFPs dar. Sie ist von Zapata-Hommer and Griesbeck (2003) ausführlich beschrieben und zeichnet sich durch folgende Eigenschaften aus:

- Anregung $\lambda_{\text{ex}} = 399 \text{ nm}$
- Emission $\lambda_{\text{em}} = 511 \text{ nm}$
- verbesserte Faltung \Rightarrow höhere Leuchtkraft
- temperaturstabil
- stabil gegen pH-Änderungen in der Zelle
- schnelle Fluoreszenzausbildung (45 min.)
- Anwendung als Fusionsprotein möglich.

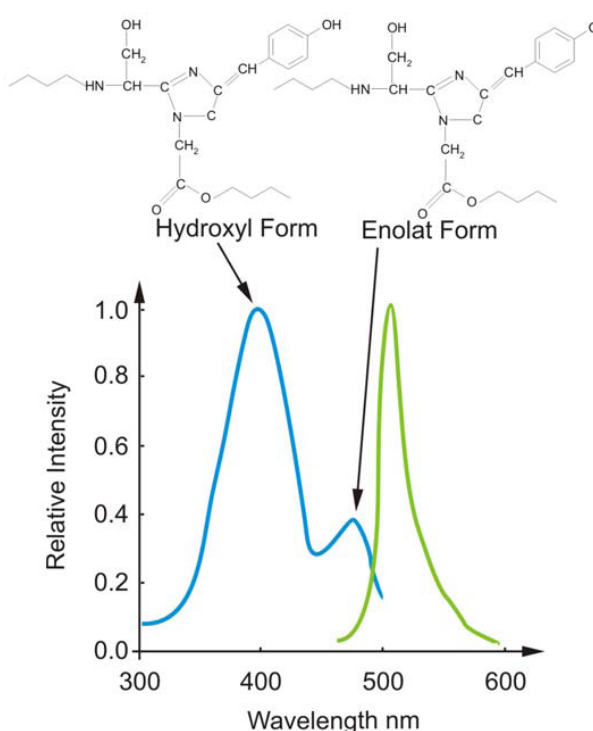


Bild 2.4: Anregungs- und Emissionsspektrum von wtGFP [Zoon (2008)]

2.3 Stoffwechselvorgänge von *Escherichia coli*

Die Umsetzung der eingesetzten Kohlenstoffquelle Glukose lässt sich im Wesentlichen in zwei Abschnitte unterteilen. Auf der einen Seite steht der Glukoseabbau, der als Katabolismus oder Energiestoffwechsel bezeichnet wird. Dieser beinhaltet unter aeroben Bedingungen die vollständigen Umsetzung des Substrats zu Kohlendioxid und Wasser unter Bildung der energie-reichen Verbindung Adenosintriphosphat (ATP) mit der dabei freigesetzten Energie [Madigen *et al.* (2000)].

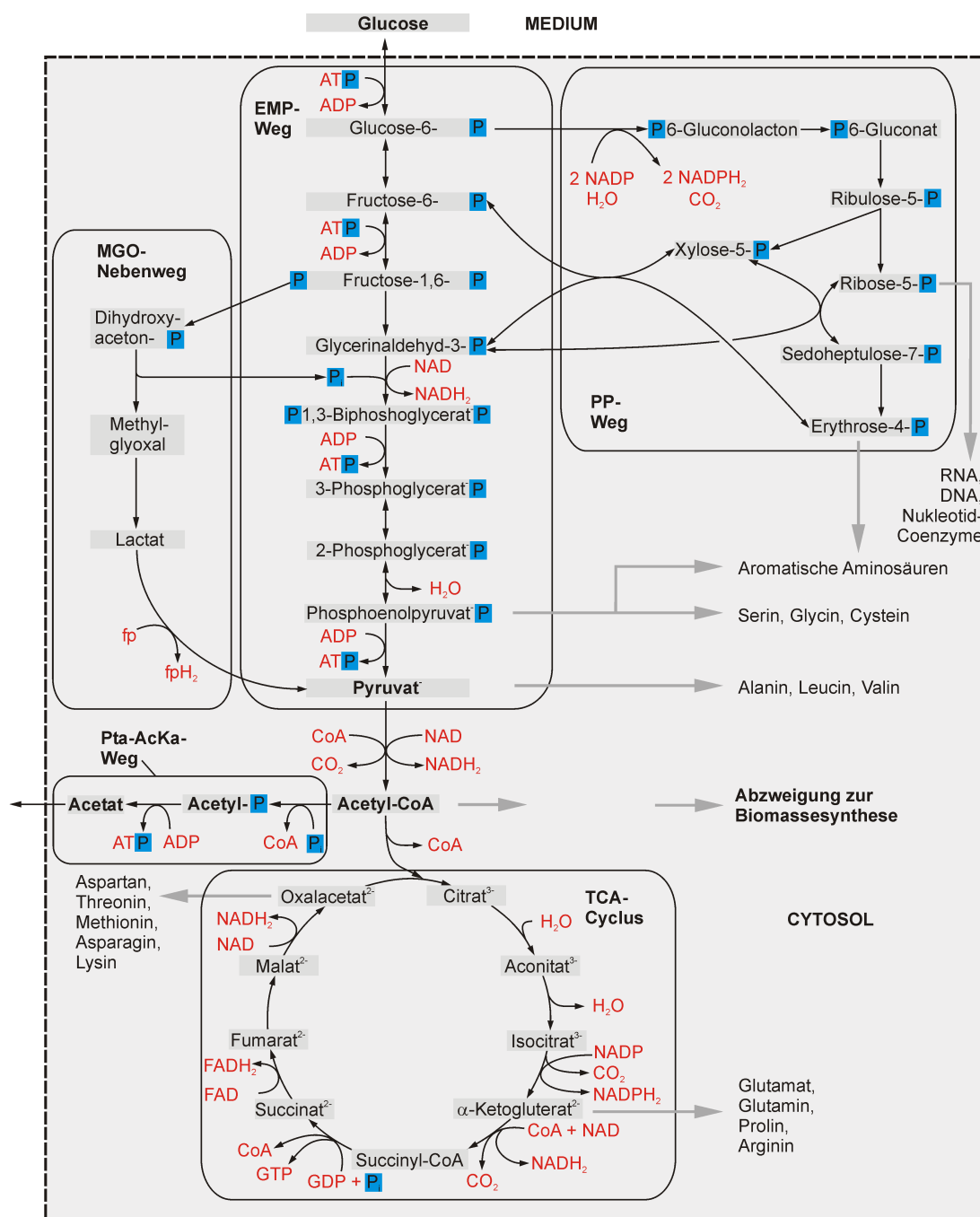


Bild 2.5: Stoffwechselwege des Glukoseabbaus in *E. coli* in Anlehnung an Madigen *et al.* (2000) und Kayser (1999)

Ein weiteres Ziel des Katabolismus ist es, das Substrat in kleinere Bruchstücke (Metabolite) zu zerlegen, die im Synthesestoffwechsel (Anabolismus) dazu genutzt werden, unter Energieverbrauch Biopolymere wie z.B. Proteine, RNA und DNA aufzubauen.

Die Glukose erfüllt also eine doppelte Funktion für die Zelle. Sie dient zum Einen als Ausgangspunkt für den Synthesestoffwechsel der benötigten Metabolite und zum Anderen als Energiequelle, um die für Zellwachstum und Proteinbildung erforderlichen chemischen Reaktionen aufrecht zu erhalten [Schlegel (1992)].

Der Glukosetransport in die Zelle, der an eine Phosphorylierung gekoppelt ist, erfolgt durch Gruppentranslokalisierung mit Hilfe eines Carrierproteins durch die Cytoplasmamembran. Durch den aktiven Transport kann das Substrat auch gegen ein Konzentrationsgefälle erfolgen. In der Zelle liegt die Glukose dann ausschließlich als Zuckerphosphat vor [van der Zee *et al.* (1996)].

Nach der Aufnahme der Glukose in die Zelle erfolgt der Abbau über Redoxreaktionen. Dafür stehen *Escherichia coli* unter aeroben Bedingungen im Wesentlichen vier unterschiedliche Stoffwechselwege zur Verfügung, die in Bild 2.5 dargestellt sind und im Folgenden kurz vorgestellt werden. Diese sind die Glycolyse, die nach Ihren Entdeckern auch Embden-Meyerhof-Parnas (EMP) Weg genannt wird, der Pentosephosphatweg (PP-Weg), der Methylglyoxal (MGO)-Nebenweg sowie der Tricarbonsäurezyklus (TCA). Die ersten drei Stoffwechselwege laufen dabei teilweise parallel ab [Kayser (1999)].

Der Großteil der Glukose wird in der Glycolyse bis zum Pyruvat umgewandelt. In den ersten energieverbrauchenden Schritten erfolgt eine Umwandlung durch Phosphorylierung und Isomerisierung zu Fruktose-1,6-bisphosphat. Der C₆-Körper wird anschließend in zwei C₃-Körper gespalten, die dann zu Pyruvat oxidiert werden. Dabei wird NAD reduziert und durch Substratkettenphosphorylierung ATP gewonnen [Schlegel (1992)]. Energetisch werden durch den Embden-Meyerhof-Parnas Weg für ein Glukosemolekül zwei ATP durch Substratkettenphosphorylierung gewonnen und zwei NAD reduziert.

Die Intermediate Pyruvat und seine Vorstufe Phosphoenolpyruvat des EMP-Weges werden auch als Metabolite für die Biosynthese unterschiedlicher Aminosäuren aus dem Stoffwechsel abgezogen.

Ungefähr 20 % der Glukose werden über den Pentosephosphatweg verstoffwechselt [Fuhrer *et al.* (2005)]. Das Enzym Glukose-6-phosphatdehydrogenase kontrolliert den Stofffluß durch den PP-Weg [Bonsignore and de Flora (1972)]. Es ist bei *Escherichia coli* NADP-spezifisch und reguliert die Generierung von NADPH, welches als Reduktionsäquivalent für biosynthetische Reaktionen von Bedeutung ist. Den wichtigsten Metabolit stellt dabei Ribose-5-phosphat dar, welcher als Monomer für die Synthese von RNA, DNA und Nucleotid-Coenzymen dient.

Es besteht eine reversible Verbindung zwischen dem PP- und dem EMP-Weg, so dass durch die Kombination der beiden Stoffwechselwege der Spiegel von NADPH, ATP sowie Ribose-5-phosphat und Pyruvat reguliert werden kann [Jeremy *et al.* (2007)].

Die Funktion des Methylglyoxal-Nebenwegs ist bisher nicht geklärt. Viele Mikroorganismen können über diesen Stoffwechselweg aus Fruktose-1,6-bisphosphat, über die Zwischenstufen Dihydroxyacetonphosphat, Methylglyoxal und Laktat, Pyruvat bilden. Da keine Substratkettenphosphorylierung stattfindet, ist der Energiebedarf dabei geringer als über den EMP-Weg. Methylglyoxal wirkt jedoch schon ab einer sehr geringen Konzentration als Zellgift und

reduziert das Wachstum durch die Inhibierung der DNA-Initiation und hemmt die Protein-Biosynthese [Inoue and Kimura (1995)][Fraval and McBrien (1980)][Fredberg *et al.* (1971)].

Mögliche Aufgaben des MGO-Nebenweges könnten das Entkoppeln von Anabolismus und Katabolismus sein oder aber der Wachstumsregulation dienen [Cooper (1984)]. Ebenso ist eine Funktion als Kontrollmechanismus des Embden-Meyerhof-Parnas Weg möglich [Hopper and Cooper (1971)].

Nachdem die Glukose durch die oben aufgeführten Stoffwechselwege bis zum Pyruvat oxidiert worden ist, wird dieses nach oxidativer Decarboxylierung zu Acetyl-CoA in den TCA-Zyklus⁴ eingeschleust. Zur Energiegewinnung kann dieses vollständig zu drei Molekülen CO₂, vier Molekülen NADH (Nicotinamid-Adenin-Dinukleotid) und einem Molekül FADH (Flavin-Adenin-Dinukleotid) oxidiert werden [Madigen *et al.* (2000)]. Ausserdem werden einige Intermediate des TCA als Monomere für den anabolen Stoffwechsels genutzt [Baldwin and Krebs (1981)].

Das fakultativ anaerobe Bakterium *E. coli* oxidiert die Glukose unter Abwesenheit von Sauerstoff nicht vollständig bis zum CO₂, sondern bildet Gärungsprodukte. Diese sind nach Böck and Sawers (1996) die Produkte der gemischten Säuregärung, Acetat, Succinat, Ethanol, Formiat, H₂ und CO₂. Die Gärungsprozesse dienen dabei der Regenerierung von NAD, die bei Abwesenheit von Sauerstoff nicht über die Atmungskette erfolgen kann [Clark (1989)].

Die Bildung von Acetat kann jedoch auch unter aeroben Bedingungen und einem Glukoeseüberschuss oder bei hohen Wachstumsraten unter substratlimitierten Voraussetzungen auftreten [Elmansi and Holms (1989)][Holms (1996)]. Hauptsächlich erfolgt die Acetatbildung über den in Bild 2.5 dargestellten Phosphotransacetylase-Acetatkinase (Pta-AcKa)-Weg. Dabei wird Acetyl-CoA in einer reversiblen Reaktion durch das Enzym Phosphotransacetylase zu Acetylphosphat und anschließend über Acetatkinase zu Acetat umgesetzt. Über diese Substratketten-Phosphorylierung wird ein Molekül ATP gebildet [Brown *et al.* (1977)]. Damit können die Zellen, obwohl die respiratorische Kapazität erschöpft ist, zusätzlich Energie gewinnen. Die Akkumulation des gebildeten Acetats inhibiert jedoch das Wachstum sowie die Proteinproduktion der Zellen [Luli and Strohl (1990)][Schimizu *et al.* (1988)].

Die Aktivität des Stoffwechselweges ist nach Wolfe (2005) abhängig vom pH-Wert, der Temperatur, der Medienzusammensetzung sowie der Sauerstoffverfügbarkeit. Die Phosphotransacetylase wird dabei durch Pyruvat aktiviert und durch NADH inhibiert [Suzuki (1969)].

Ein weiterer hier nicht dargestellter Weg der aeroben Acetatbildung ist nach Gennis and Reeves (1976) die Decarboxylierung von Pyruvat zu Acetat und CO₂. Diese wird durch Pyruvatoxidase katalysiert und dabei FAD zu FADH₂ reduziert.

Das von den Zellen gebildete Acetat kann bei Abwesenheit von Glukose oder unter glukoselimitierten Bedingungen von *Escherichia coli* als C-Quelle wieder verwertet werden. Da der Citrat-Zyklus nur ablaufen kann, wenn das Oxalacetat bei jedem Durchlauf des Zyklus regeneriert wird und damit keine Kohlenstoffverbindungen für Biosynthesereaktionen zur Verfügung ständen, müssen die Zellen zusätzlich auf einen weiteren Stoffwechselweg zurückgreifen.

⁴TCA-Zyklus – auch Krebszyklus, Zitronensäure- oder Zitratzyklus genannt

Der in Bild 2.6 dargestellte Glyoxylat-Zyklus stellt dabei eine anabolische Variante des Citrat-Zyklus dar. Er besteht im Wesentlichen aus den Reaktionen des TCA wobei zwei zusätzliche Enzyme benötigt werden. Zum Einen die Isocitratlyase, die Isocitrat in Succinat und Glyoxylat spaltet und zum Anderen Malatsynthase, die Glyoxylat und Acetyl-CoA zu Malat synthetisiert [Schlegel (1992)].

Durch die Spaltung von Isocitrat kann das gebildete Succinat oder ein daraus gebildetes Intermediat des Zitronensäure-Zyklus zur Biosynthese verwendet werden. Es kann zum Aufbau von C₄-Aminosäuren dienen, zu Oxalacetat oxidiert werden oder als dieses in Verbindung mit Phosphoenolpyruvat in Glukose umgewandelt werden. Durch die Verwendung des Glyoxylat-Zyklus beim Aceatabbau kann dadurch die zyklische Natur des TCA und damit der Energiebedarf der Zellen über den Katabolismus als auch der Aufbau von Biomasse und Biopolymeren aufrecht erhalten werden [Madigen *et al.* (2000)].

Nach der kurzen Übersicht über die mikro- und molekularbiologischen Eigenschaften des verwendeten *Escherichia coli* Host-Systems sowie des gewählten rekombinanten Produktes werden im nachfolgenden Kapitel die eingesetzten Messverfahren zur Quantifizierung unterschiedlicher Substrate und Produkte vorgestellt. Dabei wird zusätzlich zu den verwendeten Verfahren auch ein Überblick über die zur Zeit am Markt erhältliche Messtechnik gegeben.

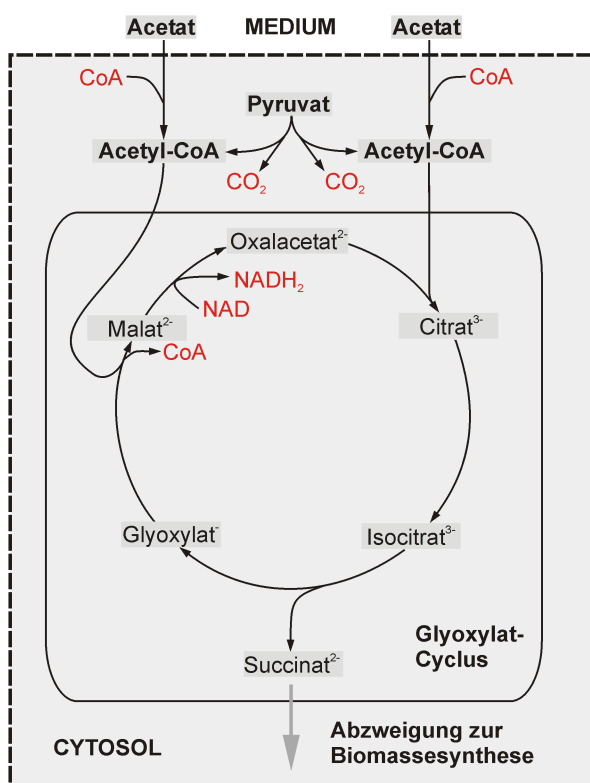


Bild 2.6: Der Glyoxylat-Zyklus nach Madigen *et al.* (2000)

3 Monitoring von Bioreaktionsprozessen

Unter einem Monitoring von Bioreaktionsprozessen versteht man die messtechnische Erfassung von chemischen, physikalischen und biologischen Daten sowie deren Aufbereitung und anschließende Analyse. Die Zielsetzungen sind dabei eine Prozessüberwachung und Parameteridentifikation und, aufbauend auf den gewonnenen Erkenntnissen, die Entwicklung von geeigneten Prozessführungsstrategien zur Ausbeutesteigerung und Prozesskostenminimierung. Der Weg von den gewonnenen Rohdaten zum gesicherten Wissen ist in Bild 3.1 aufgezeigt

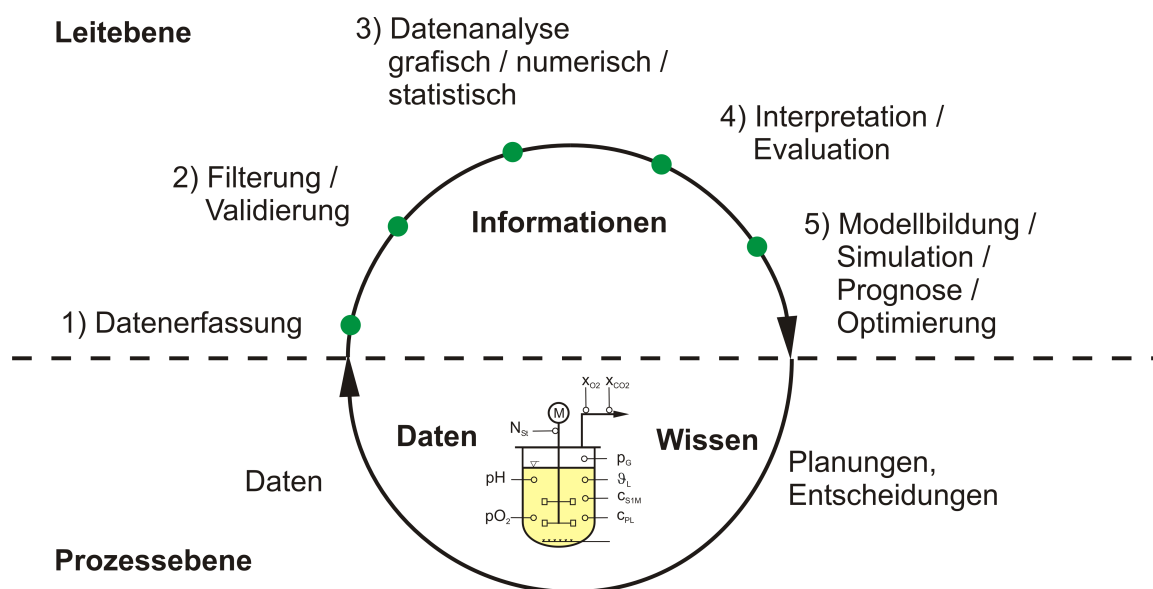


Bild 3.1: Der Weg von den Rohdaten zum gesicherten Wissen [Peuker (2006)]

Dieses Kapitel beschäftigt sich im Folgenden mit den messtechnischen Möglichkeiten an Kultivierungsprozessen und den in dieser Arbeit eingesetzten Verfahren. Auf die Messdatenerfassung und -aufbereitung sowie erweiterte Methoden der Prozessüberwachung und Simulation wird dann in den folgenden Kapiteln näher eingegangen.

3.1 Prozessüberwachung in der Biotechnologie

Die Anwendung von physikalischen und chemisch(-biologisch)en Verfahren zur Überwachung des die Zellen umgebenden Mediums sowie des physiologischen Zellzustandes während des Prozesses ist in der Forschung in den letzten Jahren kontinuierlich angestiegen. Dieses ist auch auf die Fortschritte der rekombinanten Techniken und des damit verbundenen Gewinnpotentials neuartiger Pharmazeutika zurückzuführen. Dabei eignen sich die meisten Messverfahren für die Identifikation von Prozessparametern in näherungsweise ideal durchmischten Laborreaktoren. Für den Einsatz am Produktionsprozess ist meist die notwendige Robustheit nicht gegeben, die Aussagekraft der Messungen aufgrund der Inhomogenität der Kulturbrühe zu gering

und/oder der Kostenaufwand zu hoch. In der produzierenden pharmazeutischen Industrie stellen die strengen cGMP-Richtlinien einen weiteren Hinderungsgrund für den Einsatz erweiterter Messtechniken an Kultivierungsprozessen dar.

Daher beschränkt sich die Messtechnik an Bioprozessen in den allermeisten Fällen auch heute noch auf die bewährten Messgrößen wie Temperatur, Masse, Druck, Leitfähigkeit, Gas- und Flüssigkeitsströme, Schaumhöhe, Rührergeschwindigkeit und Leistungseintrag zur physikalischen Charakterisierung sowie pH- und pO_2 -Wert, Redoxpotential und die Molenbrüche von Sauerstoff und Kohlendioxid im Abgas für die chemische Charakterisierung [Känsäkoski *et al.* (2006)].

Die zur Produktion aktiver Biopharmazeutika (API¹) eingesetzten biologischen Wirtssysteme werden jedoch wesentlich von den physikalischen und chemischen Parametern des sie umgebenden Mediums beeinflusst. Um die Qualitätsschwankungen der Produkte so gering wie möglich zu halten, müssen die Einflussfaktoren bekannt sein. Dieses bedingt ein größeres Verständnis der Wechselwirkung der Wirtssysteme mit ihrer Umgebung und damit ein besseres Prozessmonitoring, das die Grundlage für eine optimale Führung des Prozesses mit einer gleichbleibend hohen Produktqualität darstellt.

Aus diesen Gründen hat die US-Amerikanische *Food and Drug Administration* im Jahr 2004 die PAT-Initiative „PAT – A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance“ [FDA (2004)] auf den Weg gebracht. Durch die Vorgaben der Initiative, die mehr einen Leitfaden als eine strikte Richtlinie darstellt, werden die Hersteller der produzierenden pharmazeutischen Industrie unter Anderem dazu angehalten, durch den Einsatz von geeigneten Messsystemen die Schwankungen in der Produktqualität zu minimieren.

Grundsätzlich lässt sich die Art der Messungen zur prozessbegleitenden Analyse in Inline-, Online-, Atline- und Offline-Verfahren unterteilen, wobei auf Letztere an dieser Stelle nicht näher eingegangen werden soll. Die Definitionen in Anlehnung an die FDA [FDA (2004)] sind in Bild 3.2 gegeben.

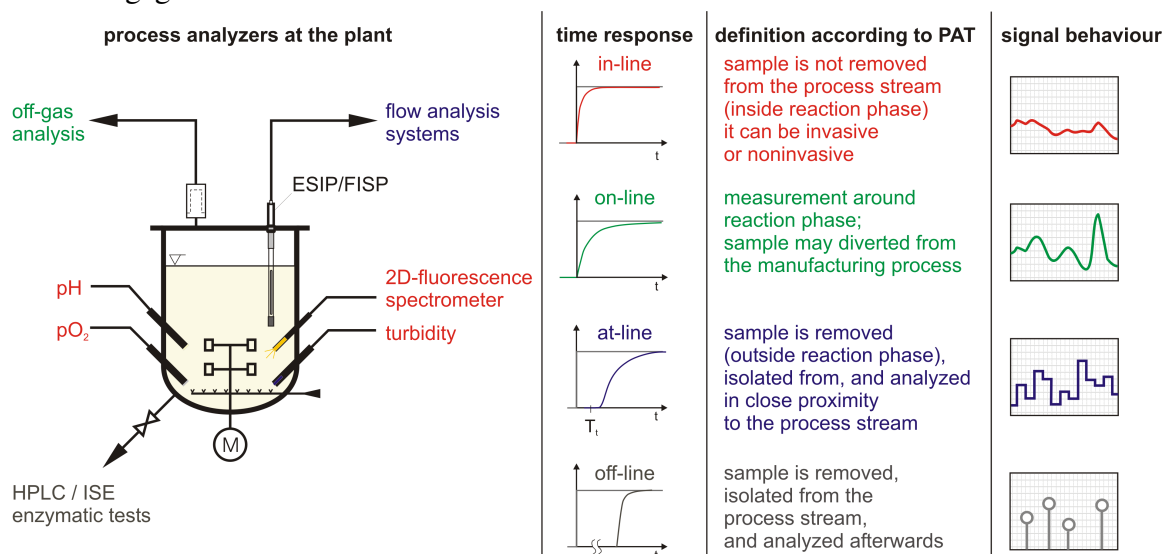


Bild 3.2: Definition der Art von Messverhalten [Peuker (2006)]

¹API – Active Pharmaceutical Ingredients

Für die konstruktive Beschaffenheit der Sensoren ist es vor allem massgeblich, ob sie innerhalb des sterilen Bereiches (*insitu*) eingesetzt werden, also direkt im Bioreaktor oder in einem Bypass mit Rückfluss in den Reaktor, oder aber außerhalb der Sterilbarrieren (*exsitu*). *Insitu*-Sensoren müssen sterilisierbar sein, wohingegen *Exsitu*-Sensoren die ausreichende Versorgung mit zellbehaftetem oder zellfreiem Medium und damit die Installation eines geeigneten *Insitu*-Probenahmesystems erfordern.

Die zellbehaftete automatisierte Probenahme zur Bestimmung von zellinternen Komponenten stellt dabei das größte Problem dar. Sie muss ohne Kontaminationsrisiko erfolgen und die Aktivität der Zellen muss durch die Zugabe geeigneter Reagenzien gestoppt werden, um eine Verfälschung der Messergebnisse zu verhindern. Dieses ist reproduzierbar nur durch automatisierte Probenahmesysteme zu realisieren [Hiller *et al.* (2007)][Buchholz *et al.* (2002)].

Für die Versorgung der Analysengeräte mit zellfreiem Medium stehen unterschiedliche kommerziell verfügbare Inline-Filtrationssysteme zur Verfügung, bei denen die Filtrationsmembran als Sterilbarriere dient [van de Merbel *et al.* (1996)].

Das für diese Arbeit gewählte messtechnische Setup, die eingesetzte Zufütterungsperipherie und die zugehörigen Waagen sind in Bild 3.3 schematisch dargestellt. Im weiteren Verlauf dieses Kapitels werden die einzelnen Messsysteme vorgestellt. Dabei sind die einzelnen Abschnitte nach der Art des Messverhaltens gegliedert. Zu Beginn eines jeden Abschnittes wird dabei ein kurzer Überblick über die verfügbaren Messsysteme gegeben.

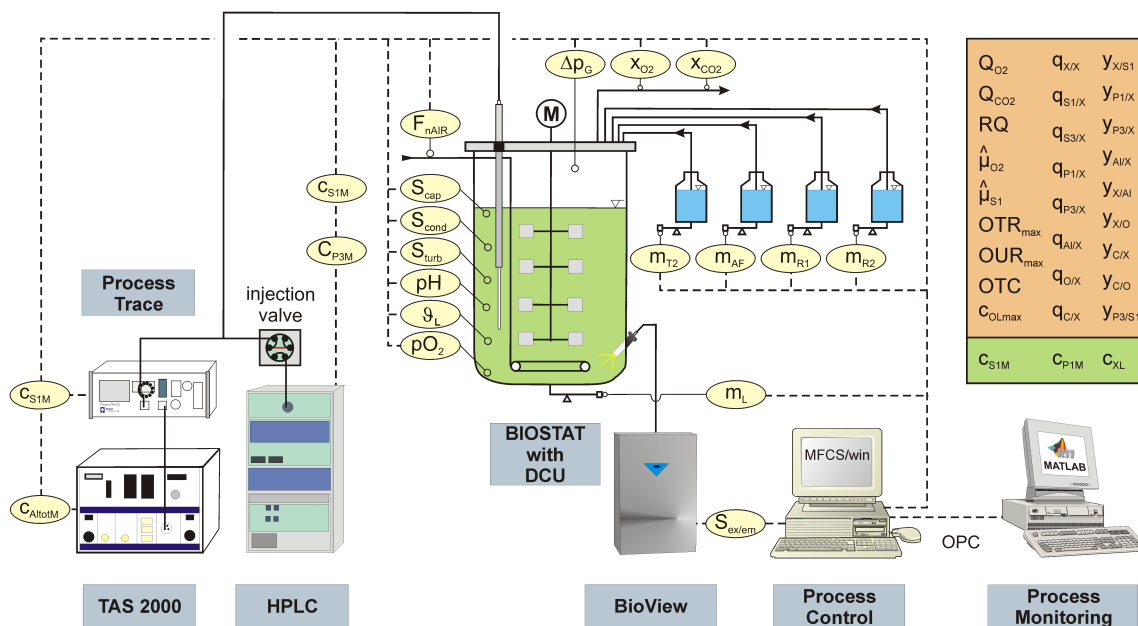


Bild 3.3: Schematische Darstellung des gewählten Anlagensetups

3.2 Standard Messtechnik

Zur Kultivierung der Mikroorganismen wurde ein *insitu* sterilisierbarer Laborbioreaktor der Fa. Sartorius Stedim Biotech GmbH, Melsungen, mit 10 l Arbeitsvolumen eingesetzt. Der „open-frame“ Rührkesselreaktor BIostat® C mit Obenantrieb besitzt ein h/d-Verhältnis von 3:1. Der in Bild 3.4 gezeigte Bioreaktor wird über eine Micro-DCU² gesteuert und ist werksseitig mit der zugehörigen Standardmesstechnik für den relativen Gelöstsauerstoffpartialdruck pO_2 , pH-Wert, Temperatur sowie Füllstand und Schaumalarm ausgestattet. Die Micro-DCU beinhaltet Standard-Regelkreise für die Temperatur, den pH-Wert und die Schaumbekämpfung sowie eine pO_2 -Drehzahlregelung.

Zusätzlich zur Luftbegasungseinrichtung des Bioreaktors ($F_{nAIRmax} = 30 \text{ lmin}^{-1}$) ist der Fermenter mit einer externen Gasmischstation ausgestattet, mit der eine Sauerstoffbegasung bis maximal 5 lmin^{-1} zu realisieren ist. Der Fermenterüberdruck wird über eine Drucksonde, die hinter dem Abgaskühler installiert ist, überwacht.

Für die Bestimmung von Sauerstoff und Kohlendioxid im Abgasweg steht eine Abgasanalytik Multor 610 der Fa. Sick|Maihak, Hamburg zur Verfügung. Die paramagnetische O_2 -Messung erfolgt über ein OXOR-Modul, die CO_2 -Messung mittels nichtdispersiver Infrarot-Messung³ über eine UNOR-Einheit.

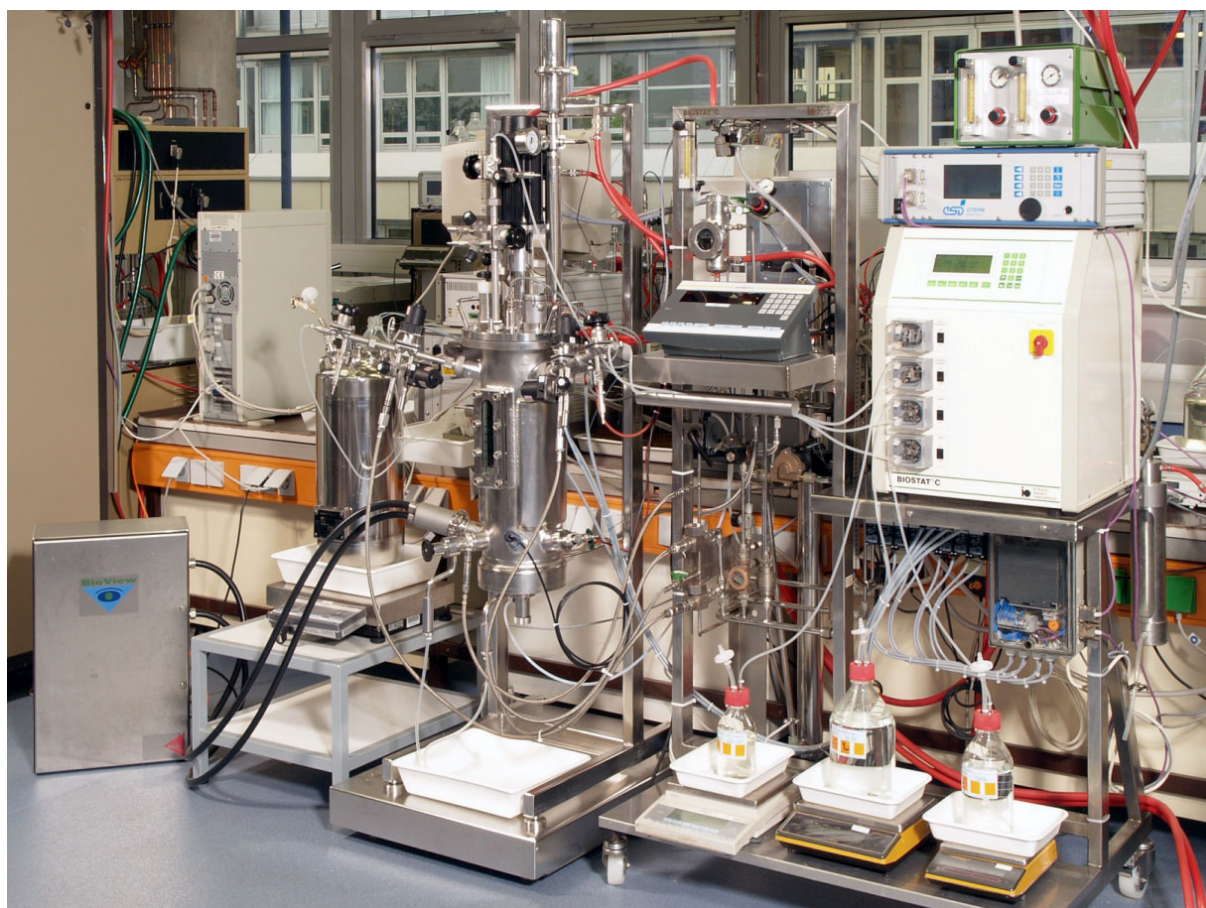


Bild 3.4: Der eingesetzte Bioreaktor BIostat® C

²DCU – Digital Control Unit

³NichtDispersive InfraRot-Messung – NDIR-Messung

3.3 Erweiterte Inline-Messtechnik

3.3.1 Möglichkeiten der Inline-Instrumentierung

Inline-Messungen finden direkt in der flüssigen Reaktionsphase statt. Sie liefern meist stetige Signale und zeichnen sich durch geringe Zeitverzögerungen aus [FDA (2004)]. Neben den klassischen Inline-Messsonden für pH, Temperatur und pO₂ existieren eine Reihe weiterer Systeme, die im Folgenden kurz vorgestellt werden.

Optische Methoden bieten sich für den Inline-Einsatz in Bioprozessen an. Die einfachste und am weitesten verbreitete Applikation stellen die Trübungssonden zur Abschätzung der Zelldichte dar. Für eine genauere Charakterisierung der Zellen, die auch die Größenverteilung und eine Beurteilung der Zellaggregation für höhere Zellen einschließt, können *In situ*-Mikroskope eingesetzt werden [Joeris *et al.* (2002)][Lindner *et al.* (2007)].

Großes Potential besitzen die Verfahren der optischen Spektroskopie. Die Nahinfrarot-Spektroskopie⁴ ist eine der am weitesten verbreiteten Methoden zur Bioprozessüberwachung. Viele biologisch interessante Komponenten wie aliphatische und aromatische Kohlenwasserstoffe, Amine und Hydroxylgruppen absorbieren im NIR-Bereich bei Wellenlängen zwischen 2.0 und 2.5 μm [Marose *et al.* (1999)]. Für die Kalibration der Systeme müssen im allgemeinen chemometrische Methoden eingesetzt werden [Känsäkoski *et al.* (2006)].

Arnold *et al.* (2003a) haben die NIRS erfolgreich zum Monitoring von Glukose, Laktat und Ammonium in Fed-Batch Prozessen mit CHO-K1-Zellen eingesetzt. Sie stellten aber auch heraus, dass die sterilisierbaren Glasfasersonden zur Transmissionsmessung eine schlechte Auflösung oberhalb von 2.1 μm aufweisen [Arnold *et al.* (2002)]. Außerdem haben Gasblasen, Temperaturunterschiede, Fouling und Vibrationen durch das Rührwerk negative Auswirkungen auf die Messmethode [Arnold *et al.* (2003b)][Arnold *et al.* (2002)]. Aus diesen Gründen stellen fast alle NIR-Applikationen Atline-Anwendungen dar. Typische Messgrößen sind Biomasse, Zucker, organische Säuren, Ammonium und Alkohole [Crowley *et al.* (2005)][Navratil *et al.* (2005)][Macaloney *et al.* (1997)].

Für den Inline-Einsatz scheint die *Fourier-Transformed* Infrarot Spektroskopie (FTIR) im mittleren Infrarot-Bereich⁵ besser geeignet zu sein. ASI Applied Systems (Mettler-Toledo) haben eine Sonde (DiComp) entwickelt, die auf dem ATR⁶-Prinzip basiert, sterilisierbar ist und durch einen 25 mm Sicherheitsstutzen in den Bioreaktor eingebracht werden kann. Mehrere Arbeitsgruppen konnten zeigen, dass sich diese Technik für die Messung von Glukose, Laktat, Acetat und Ethanol in Biosuspensionen eignet [Doak and Phillips (1999)][Pollard *et al.* (2002)][Rhiel *et al.* (2002)][Kornmann *et al.* (2003)][Kornmann *et al.* (2004)].

Weitere spektroskopische Methoden stellen die Raman-Spektroskopie sowie die Fotoakustische Spektroskopie dar. Cannizzarro *et al.* (2003) konnte die Raman-Spektroskopie erfolgreich zum Monitoring von Glukose und zellintern gebildeten Karotenoiden mit Hefezellen einsetzen. Die fotoakustische Spektroskopie eignet sich für die Abgasanalytik, wie z.B. im Fermentation Monitor 1313 der Fa. Innova, Dänemark.

⁴NIRS – *Near InfraRed Spectroscopy*

⁵MIR – *Mid InfraRed*

⁶ATR – *Attenuated Total Reflection*

Bei den in dieser Arbeit eingesetzten Inline-Verfahren handelt es sich um eine Trübungsmessung, eine Kapazitäts- und Leitfähigkeitsmessung sowie um ein Messsystem für die 2D-Fluoreszenzspektroskopie, auf deren Anwendung in Abschnitt 3.3.4 näher eingegangen wird.

Trübungs- und Kapazitätsmessungen stellen produktberührende Methoden zur Online-Abschätzung der Biomasse dar. Diese ist eine Schlüsselgröße in der Prozessbeobachtung und bildet die Grundlage für Online-Berechnungen zellspezifischer und volumetrischer Reaktionsraten. Unter Verwendung geeigneter mathematischer Modelle ist aber auch über die nicht produktberührende Fluoreszenzspektroskopie eine Online-Zelldichteschätzung zu realisieren. Diese Methoden werden in Kapitel 5 ausführlich diskutiert.

Exemplarisch ist in Bild 3.5 die Online-Abschätzung der Zelldichte über die Trübungsmessung ($c_{XL_{turb}}$), die Kapazitätsmessung ($c_{XL_{cap}}$) und die Abgasbilanzierung ($c_{XL_{est}}$) während einer Kultivierung dargestellt. Es ist offensichtlich, dass die Messergebnisse der einzelnen Systeme, insbesondere für höhere Zelldichten, voneinander abweichen. Dieses ist auf die zu Grunde liegenden Messprinzipien unterschiedlicher Zelleigenschaften sowie die mathematischen Anpassungen für die jeweiligen Prozessphasen zurückzuführen, die im Folgenden näher erläutert werden.

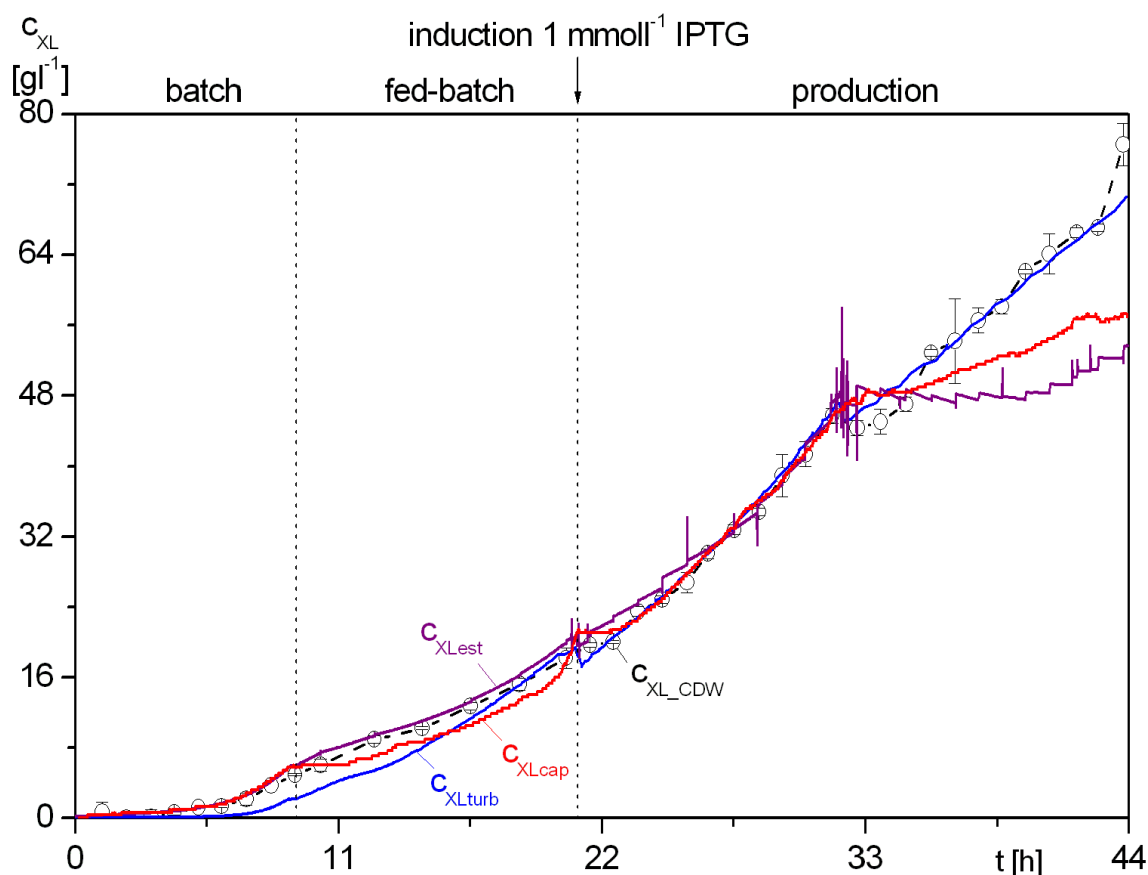


Bild 3.5: Ergebnisse verschiedener Methoden zur Zelldichtebestimmung

3.3.2 Trübungsmessung

Trübungsmessungen sind die am häufigsten eingesetzten Inline-Messsysteme zur Online-Abschätzung der unbekanntes Zelldichte in der Kulturbrühe. Das Messprinzip basiert auf der Abschwächung des Lichts beim Durchgang durch das Prozessmedium. Diese wird massgeblich von der Menge der in der Flüssigphase dispergierten Teilchen beeinflusst. Dabei gibt es keine Selektion zwischen aktiver und toter Biomasse oder evtl. auftretenden Salzfällungen.

Einen weiteren wesentlichen Einfluß auf das Trübungssignal haben die Rührergeschwindigkeit und die Begasungsrate. Die Anzahl und Größe der die Flüssigphase durchströmenden Gasblasen beeinflussen die Lichtstreuung erheblich und stellen damit eine wesentliche Störgröße des Messsystems dar.



Bild 3.6: Die eingesetzte Trübungsmesstechnik [optek-Danulat (2005)]

Für die Trübungsmessungen wurde eine Sonde der Fa. optek-DANULAT, Essen, Typ ASD25-BT-N-5 mit einer optischen Pfadlänge von 5 mm verwendet. Der Messwertempfänger ist ein Einstrahl-Absorptionsphotometer, in welchem das Lichtspektrum im nahen Infrarotbereich zur Messung verwendet wird (840 - 910 nm). Über ein Glasfaserkabel wird das mittels einer Wolf-ramlampe erzeugte Licht zum Messaufnehmer geleitet. Die Sonde besitzt einen 25 mm Portanschluss und ist CIP und SIP geeignet.

Zugehörig zum Messaufnehmer wird der Messumformer Control 4000 der Fa. optek eingesetzt. Über dieses Modul ist eine Anzeige, Speicherung und Nachbearbeitung der Messwerte möglich. Die implementierte Software A2 ist vielseitig konfigurierbar. Als Grundeinstellung wurden softwareseitig lediglich die Ausgänge, der Nullpunkt im Medium und die Parameter des implementierten Tiefpassfilters eingestellt. Control 4000 verfügt über einen 0/4-20 mA-Ausgang zur Bereitstellung der Rohdaten. Die Datenaufzeichnung des analogen Signals $S_{\text{turb}} \in [0 - 4 \text{ AU}]$ erfolgte unter Verwendung des Prozessleitsystems MFCS/win.

Die Zelldichte c_{XLturb} ,

$$c_{\text{XLturb}}(t) = a \cdot (e^{b \cdot S_{\text{turb}}(t)} - 1), \quad (3.1)$$

mit den aus mehreren *E. coli* Kultivierungen ermittelten Anpassungsparametern,

$$a = 0.0378 \text{ g l}^{-1} \text{ und } b = 3.05 \text{ AU}^{-1},$$

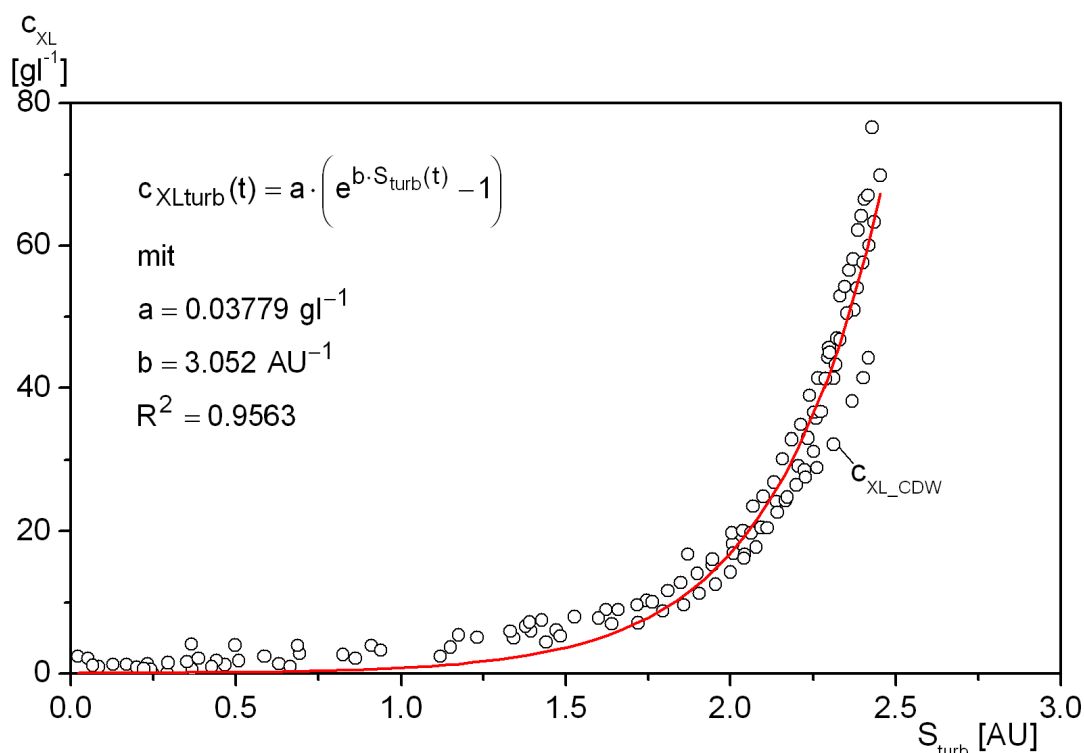


Bild 3.7: Kennlinie der Trübungssonde ASD25-BT-N-5 in *E. coli* Kultivierungen

kann unter Verwendung eines exponentiellen Ansatzes approximiert werden.

Die stammspezifischen Parameter a und b wurden postexperimentell bestimmt, wie in Bild 3.7 gezeigt. Durch die Implementierung von Gl. (3.1) in das Prozessleitsystem ist damit eine Online-Abschätzung der unbekanntenen Zelldichte über die Trübungsmessung möglich.

3.3.3 Kapazitäts- und Leitfähigkeitsmessung

Das Prinzip der Kapazitätsmessung in Biosuspensionen beruht auf den isolierenden Eigenschaften intakter Zellmembranen. Da sowohl die Flüssigkeit innerhalb der Zellen als auch das zellumgebende Medium aus ionischen Lösungen bestehen, kommt es beim Anlegen eines elektrischen Feldes zu einer Polarisation der Zellen. Der Grad der Polarisation im Wechselstromfeld ist dabei von der angelegten Frequenz abhängig. Grundsätzlich gibt es unterschiedliche Bereiche, in denen sich frequenzabhängige Polarisationen ergeben. Im Bereich von 0.1 - 100 MHz wird diese als β -Dispersion bezeichnet.

Der Verlauf der β -Dispersion in Abhängigkeit unterschiedlicher Zelldichten und Zellgrößen ist in Bild 3.8 dargestellt. Bei hohen Frequenzen haben die Ladungsträger nicht ausreichend Zeit sich im elektrischen Feld auszurichten. Es bildet sich ein Plateau aus, so dass nur die sogenannte Hintergrundkapazität $S_{cap\infty}$ detektiert werden kann. Mit abnehmender Frequenz nimmt die Polarisation der Zellen zu und das Messsignal wird stärker, bis ein weiteres Polarisationsplateau erreicht wird. Die Frequenz, bei der sich die Hälfte der Polarisation einstellt, wird als kritische Frequenz f_c bezeichnet. Um das Signal-Rausch Verhältnis zu verbessern wird meistens die Hintergrundkapazität von den Messsignalen geringerer Frequenz subtrahiert. Eine detaillierte Beschreibung der Prinzipien der dielektrischen Spektroskopie findet sich bei Davey *et al.* (1993).

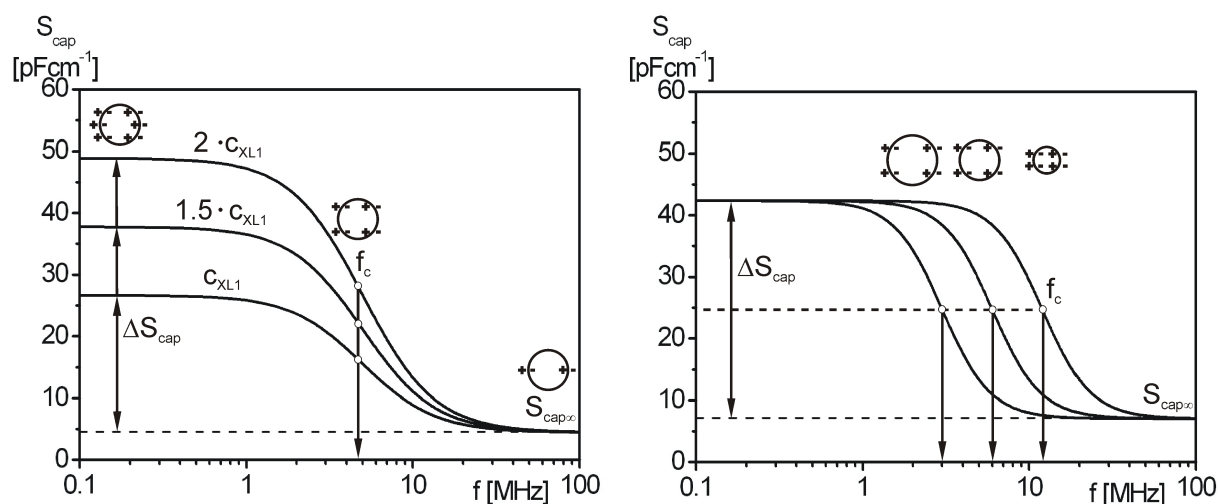


Bild 3.8: Prinzipieller Verlauf der β -Dispersion in Abhängigkeit der Messfrequenz
links: Zelldichteabhängigkeit rechts: Zellgrößenabhängigkeit

Es wurden zwei verschiedene Messsysteme der Fa. AberInstruments Ltd., Aberystwyth, GB, im Verlauf dieser Arbeit eingesetzt. Dabei handelte es sich um den Biomass Monitor 200 (BM200) und den BM220, der in Bild 3.9 dargestellt ist. Die Messungen erfolgten bei 1.21 MHz und die Detektion der Hintergrundkapazität bei 15.65 MHz.

Als Sonden wurden zwei unterschiedliche Ringelektroden mit der gleichen Charakteristik eingesetzt. Die eine war für den Einsatz in einem 19 mm Deckelstutzen vorgesehen, die andere für den Einbau in den unteren Stutzenkranz des Bioreaktors (25 mm Sicherheitsstutzen).

Der BM220 ermöglicht die Aufnahme der Kapazität bei 25 Frequenzen im Bereich von 0.1 - 19.49 MHz (logarithmische Unterteilung), wodurch nicht nur die Detektion des von intakten Zellmembranen eingeschlossenen Biovolumens möglich wird, sondern auch auf die Größenverteilung der Zellen zurückgeschlossen werden kann, wie Cannizzaro *et al.* (2003) und Hauttmann and Müller (2001) für höhere Zellen zeigen konnten. Neben der Kapazität werden außerdem die Leitfähigkeit sowie die Temperatur in der Kulturbrühe über den Biomassmonitor gemessen.

Der Vorteil der Kapazitätsmessung gegenüber der Trübungsmessung liegt darin, dass nur Zellen mit intakter Zellmembran (lebende Zellen) detektiert werden, wohingegen bei der Trübungsmessung auch Zelldebris, Salzfällungen sowie Rührerdrehzahl und Begasungsrate den Messwert wesentlich beeinflussen. Da mittels der Inline-Impedanzmessung jedoch das einge-



Bild 3.9: Der Biomass Monitor BM220 zur Inline-Impedanzspektroskopie

schlossene Biovolumen detektiert wird, muss bei der Interpretation des Messsignals auch die Größenänderung der Zellen in Betracht gezogen werden. Metabolische Umstellungen, wie z.B. bei Wachstum auf Acetat, sowie Änderungen der zellspezifischen Wachstumsrate während der Fed-Batch Phase haben einen wesentlichen Einfluß auf die Zellgröße von *E. coli* [Kubitschek (1969)][Cullum and Vicente (1969)] und beeinflussen damit das Messsignal entscheidend.

3.3.4 2D-Fluoreszenzspektroskopie

Als Fluoreszenz wird die ungerichtete Emission von Licht bezeichnet, die im Unterschied zur Phosphoreszenz nur so lange auftritt, wie die fluoreszierenden Moleküle (Fluorophore) durch Lichtabsorption angeregt werden [Wedler (2004)]. Da für die Fluoreszenzfähigkeit von Molekülen freie π -Elektronen verantwortlich sind, finden sich Fluorophore in Verbindungen mit freien Elektronenpaaren, wie z.B. gesättigten Aromaten, Stickstoff- und Sauerstoffverbindungen.

Durch die Absorption elektromagnetischer Strahlung im UV/Vis-Bereich wird ein freies Elektron in einen angeregten Zustand $S1'$ überführt. Im Folgenden geht das Elektron durch innere Umwandlung, die auf Vibrations- und Rotationsvorgängen beruht, in den ersten angeregten

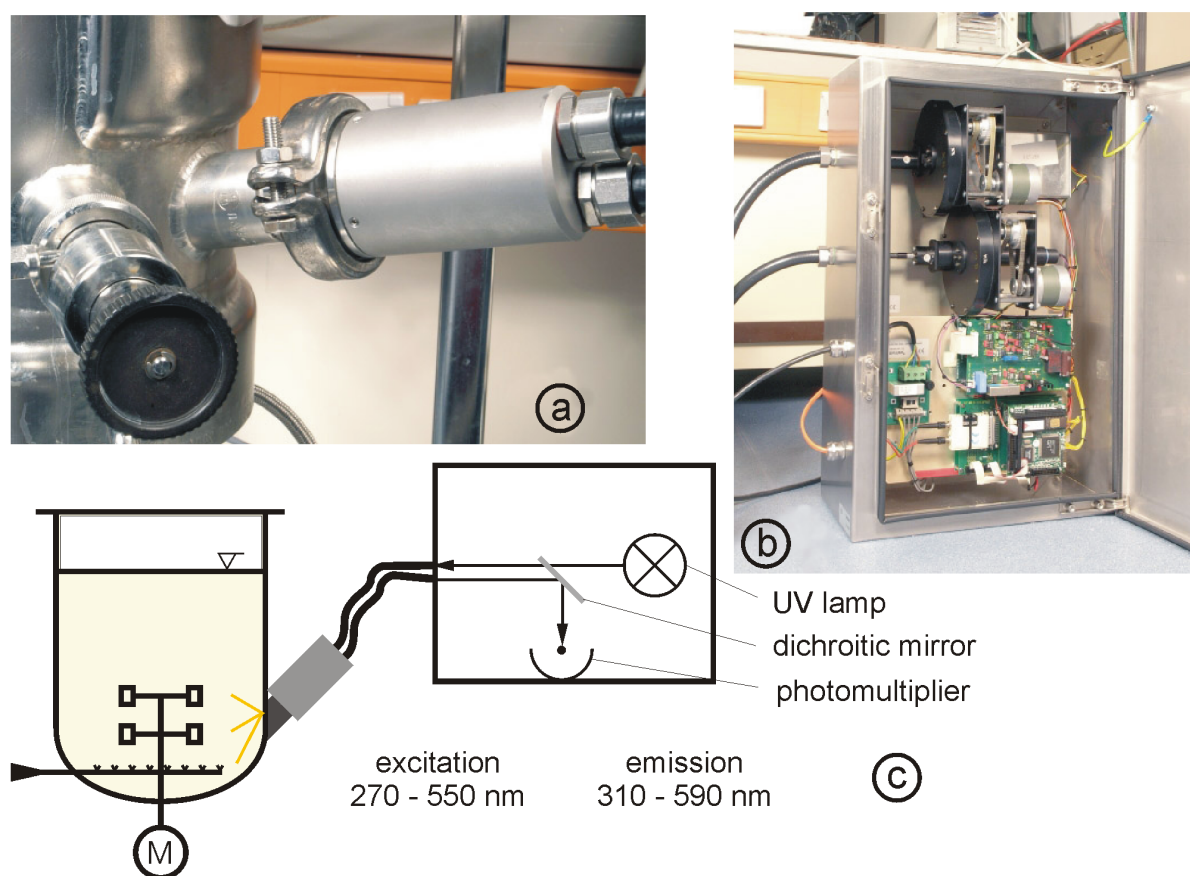


Bild 3.10: Einsatz des BioView®-Messsystems am BIostat® C

- Montage des Sensorkopfes im 25 mm Sicherheitsstutzen
- Elektronik und Mechanik des BioView®-Spektrometers
- Schematische Darstellung der Reaktorankopplung und des Messgerätes

Singlettzustand S1 über. Dieser Übergang ist strahlungslos. Bei einem spontanen Übergang des Elektrons in den Grundzustand S0 wird die Energie in Form eines Lichtquantums emittiert und kann als Fluoreszenz gemessen werden. Der gesamte Vorgang dauert zwischen 1 - 10 ns. Die Energiedifferenz zwischen den Zuständen S1' und S1 wird als „Stokes-Shift“ bezeichnet und ist dafür verantwortlich, dass das emittierte Licht grundsätzlich langwelliger und damit energieärmer als das Anregungslicht ist [Wedler (2004)].

Fluoreszenzmessungen in Biosuspensionen werden schon seit den 70iger Jahren zur Korrelation der Fluoreszenz des Nicotinamid-Adenin-Dinucleotid(phosphat)s (NAD(P)H) mit der Biomasse eingesetzt [Zabriskie and Humphrey (1978)]. Verschiedene Autoren konnten in den folgenden Jahrzehnten jedoch zeigen, dass es durchaus Fluoreszenzsignale gibt, die ausserhalb der NAD(P)H-Fluoreszenz liegen und bessere Korrelationen mit der Biomasse aufzeigen [Scheper and Schügerl (1986)][Li *et al.* (1991)]. Dies führte zu der Entwicklung der 2D-Fluoreszenzspektroskopie, bei der sich über einen weiten Bereich unterschiedlicher Anregungs- und Emissionswellenlängen ganze Fluoreszenzspektren aufnehmen lassen.

Mit Hilfe der 2D-Fluoreszenzspektroskopie konnte für unterschiedliche Organismen gezeigt werden, dass es Bereiche hoher Korrelation der Fluoreszenzspektren mit Prozessvariablen wie Biomasse, Substrat- und Produktkonzentrationen gibt [Tartakovsky *et al.* (1996)][Hilmer (1996)][Lindemann (1998)][Marose *et al.* (1998)].

Die Interpretation der Fluoreszenzspektren wird häufig durch die komplexen Wechselwirkungen und Überlagerungen der einzelnen Fluorophore erheblich erschwert. Der Einsatz chemometrischer Methoden zur Kalibrierung des Messsystems stellt dabei einen effektiven Weg dar ein robustes und zuverlässiges Monitoring auf der Basis der 2D-Fluoreszenzspektroskopie zu realisieren [Stärk (2002)][Solle (2003)][Surribas *et al.* (2006)][Lantz *et al.* (2006)].

Bei dem in dieser Arbeit eingesetzten Messsystem handelt es sich um das 2D-Fluoreszenzspektrometer BioView[®] der Fa. DELTA Light & Optics, Lyngby, Dänemark. Das in Bild 3.10 dargestellte 2D-Fluoreszenzspektrometer besitzt einen Anregungsbereich von 270 - 550 nm und einen Emissionsbereich von 310 - 590 nm, jeweils mit einer Bandbreite von 20 nm. Das Anregungslicht wird von einer gepulsten Xenon-Blitzlichtlampe erzeugt. Die entsprechenden Exzitations- und Emissionswellenlängenbereiche werden durch zwei unabhängige Filterräder mit verschiedenen Filtern realisiert. Zusätzlich befinden sich in den Filterrädern noch Neutraldichte-Filter (ND-Filter) mit einer Transmission oberhalb von 450 nm zur Exzitation (ExND) und Messung des wellenlängenabhängigen Streulichts (NDEm).

Die sich ergebenden und in Bild 3.11 exemplarisch dargestellten 2D-Fluoreszenzspektren bestehen aus insgesamt 150 Filterkombinationen und benötigt eine Analysezeit von ca. einer Minute. Der Sensorkopf des eingesetzten BioView[®] Fluoreszenzmonitors wird über einen optischen Adapter mit Quarzglasfenster nicht produktberührend in den Bioreaktor eingebracht. Dieser Adapter ist in einen 25 mm Sicherheitsstutzen des unteren Stutzenkranzes des Bioreaktors montiert (Bild 3.10a).

Die Messungen des Fluoreszenzlichtes erfolgen nach dem „Open-end“-Prinzip im Winkel von 180° zum Anregungslicht. Durch Anregung der Oberfläche vor dem Sensorkopf wirken sich Veränderungen in der Eindringtiefe des Anregungslichts und der „Inner-Filter“-Effekt weniger stark aus [Stärk (2002)]. Der eigentliche Sensor mit der Lichtquelle, Photomultiplier, Spiegel und Filterrädern für die Exzitation und Emission sowie die zugehörige Mechanik und Elektronik sind in einer Spritzwasser geschützten Edelstahlbox untergebracht (Bild 3.10b).

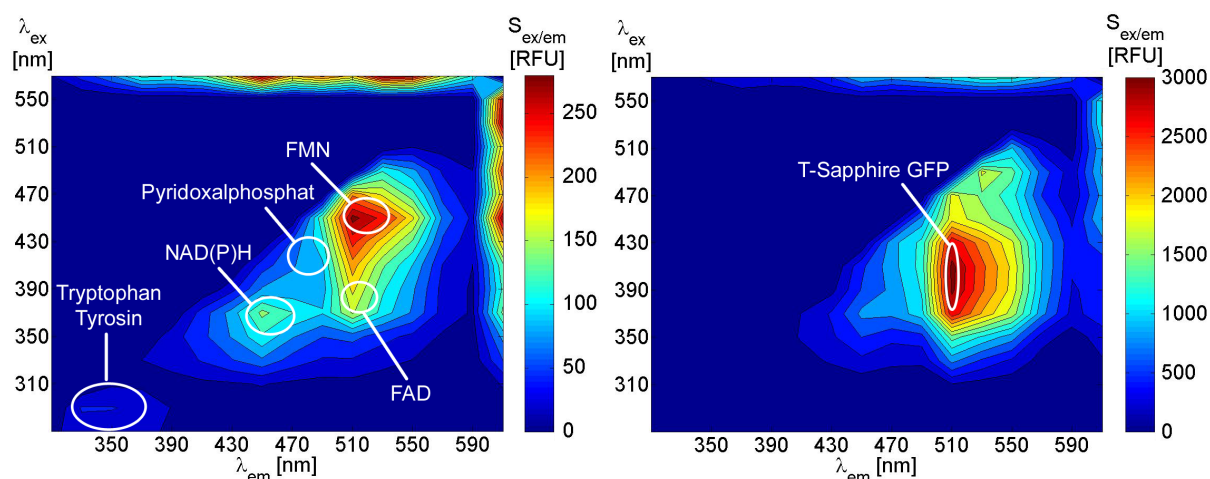


Bild 3.11: Fluoreszenzspekten während einer *E. coli* Kultivierung
links: Batch-Phase rechts: Produktionsphase

Zu Beginn einer Batch-Phase ist die Intensität der Fluoreszenz im Bereich der Aminosäure Tryptophan sowie des Elektronen transportierenden Coenzym Nicotinsäureamid-Adenin-Dinucleotid Hydroniumionens (NADH) und des Elektronencarriers und Co-Faktors Flavin-Adenin-Dinucleotid (FAD) am ausgeprägtesten. In der Produktionsphase des Fed-Batches werden vor allem die Signale von FAD und FMN (Flavine Mono-Nucleotide) durch die hohe Intensität der Fluoreszenz der im gleichen Frequenzbereich liegenden Signale des T-Sapphire GFP überlagert.

Eine Übersicht über die zu erwartenden Fluorophore und deren Anregungs- und Emissionsmaxima bietet Tabelle 3.1. Die zeitlichen Verläufe während einer Kultivierung sind in Bild 3.12 dargestellt.

Die Fluoreszenzmessung kann von diversen Größen beeinflusst werden, die in Tabelle 3.2 aufgeführt sind. Die biologischen Einflussfaktoren wie Zellkonzentration und metabolischer Zustand stellen dabei Größen dar, die über die Fluoreszenzmessung detektiert werden und die Grundlage eines Prozessmonitorings bilden sollen. Sie basieren auf der zeitlichen Konzentrationsänderung zellinterner biogener Fluorophore wie Proteine, Vitamine und Coenzyme. Die

Tabelle 3.1: Anregungs- und Emissionsmaxima biogener Fluorophore sowie von T-Sapphire GFP [Wolfbeis *et al.* (1985)][Zapata-Hommer and Griesbeck (2003)][Stärk (2002)]

| # | Fluorophore | ex/em [nm] |
|---|---------------------|------------|
| 1 | NADH, NADPH | 340/460 |
| 2 | Pyridoxin | 324/390 |
| 3 | Pyridoxalphosphat | 410/470 |
| 4 | Tryptophan, Tyrosin | 287/348 |
| 5 | FAD | 370/520 |
| 6 | FMN | 450/520 |
| 7 | T-Sapphire GFP | 399/511 |

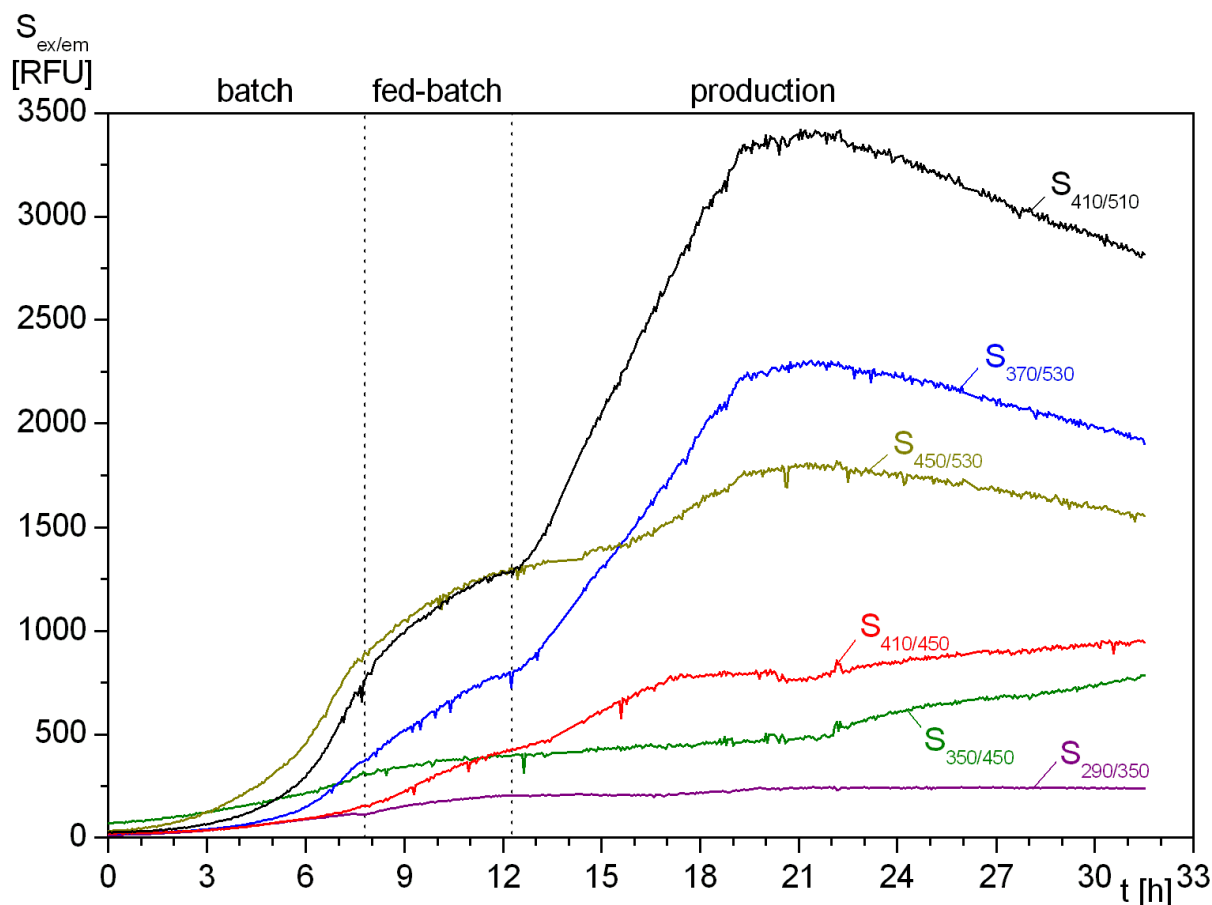


Bild 3.12: Zeitlicher Verlauf ausgewählter Wellenlängenkombinationen

Messungen können aber auch durch zellmorphologische Änderungen beeinflusst werden, welche die Eindringtiefe des Lichtes in die Zelle sowie die Absorptions- und Reflektionseigenschaften der Mikroorganismen verändern können und damit Einfluss auf die Fluoreszenzantwort nehmen.

Die Fluoreszenzspektren verschiedener Fluorophore (z.B. Pyridoxin) sind stark pH-Wert abhängig. Durch Protonierung wird die Ladungsverteilung im Molekül verändert, was sowohl die Intensität als auch die Lage der Anregungs- und Emissionsmaxima verändern kann [Zander (1981)][Li and Humphrey (1992)]. Durch unterschiedliche Polaritäten verschiedener Lösungsmittel kann die Polarisierung von Fluorophoren beeinflusst werden und zu spektralen Verschiebungen führen. Außerdem können protonische Lösungsmittel durch die Ausbildung von Wasserstoffbrückenbindungen die Fluoreszenzeigenschaften modifizieren [Schwedt (1989)].

Temperaturerhöhungen verursachen im Allgemeinen eine Abnahme der Fluoreszenzintensitäten, da die Wahrscheinlichkeit desaktivierender Stöße durch die höhere Molekülbewegung zunimmt. Einen gegenteiligen Effekt hat die Zunahme der Viskosität und der Dichte des Fluorophor umgebenden Mediums.

Da optische Messungen wesentlich von der Trübung des Mediums beeinflusst werden, haben auch die Rührerdrehzahl sowie die Begasungsrate einen merklichen Einfluss auf die Fluoreszenzmessung [Li and Humphrey (1992)]. Durch eine Trübungszunahme wird die Eindringtiefe des Anregungslichtes verringert. Die mittlere freie Weglänge des Streulichtes wird kleiner,

Tabelle 3.2: Einflussgrößen der Fluoreszenzmessung [Stärk (2002)]

| Art der Beeinflussung | Einflussgrößen |
|-----------------------|---|
| biologisch | Zellkonzentration Metabolischer Zustand Morphologischer Zustand |
| chemisch | pH-Wert Lösungsmittel |
| physikalisch | Temperatur Viskosität, Dichte des Mediums Begasung, Rührerdrehzahl Trübung Fluoreszenzlöschung (Quenching) „Inner-Filter“-Effekt Kaskadeneffekt |

so dass es oberhalb eines Sättigungsbereiches, trotz zunehmender Trübung, zu einer Abnahme der Streulichtintensität kommt.

Dynamische Fluoreszenzlöschung wird durch die Kollision der Fluorophore mit unterschiedlichen Molekülen (z.B. Sauerstoff) hervorgerufen. Dabei kommt es zu einer Energieübertragung von dem Fluorophor auf das entsprechende Molekül. Bei der statischen Fluoreszenzlöschung kommt es zur Ausbildung nicht fluoreszierender Komplexe aus Fluorophor und verschiedenen Ionen wie z.B. Iodid, Nitrat, Cr^{3+} oder Cu^{2+} .

Die bei hoher Konzentration der Fluorophore auftretende Konzentrations-Fluoreszenzlöschung wird als „Inner-Filter“-Effekt bezeichnet. Bedingt durch die starke Absorption des Anregungslichtes von den Fluorophoren wird die Eindringtiefe verringert und Fluorophormoleküle in größerem Abstand des Sensorkopfes nicht mehr angeregt. Dieses bedingt eine Abnahme der Fluoreszenzantwort bei gleichzeitiger Zunahme der Fluorophorkonzentration, insbesondere bei Küvettenmessungen [Li and Humphrey (1992)]. Der Kaskadeneffekt kann bei Fluorophorgemischen auftreten. Dabei kommt es zu einer Anregung eines Fluorophors durch die Lichtemission eines weiteren Fluorophors. Die Emissionen beider Fluorophore werden dabei entscheidend voneinander beeinflusst.

3.4 Eingesetzte Atline-Messtechnik

3.4.1 Mögliche Atline-Applikationen

Atline-Messsysteme zeichnen sich dadurch aus, dass die zu analysierende Probe steril, mit oder ohne Zellen, aus dem Bioreaktor entnommen wird und zeitnah einem angeschlossenen Analysengerät zugeführt wird (Bild 3.2). Sowohl der Transport der Probe zum Messgerät als auch der sich anschließende photometrische oder biochemische Messvorgang sind dabei totzeitbehaftet. Es ergeben sich retardierte, zeitdiskrete Messwerte. Diese werden in der sich anschließenden Datenverarbeitung bis zum nächsten Messergebnis festgehalten (*Sample and Hold*).

Ein Großteil der zur Bioprozessüberwachung während der Kultivierung zur Verfügung stehenden Methoden stellen die Atline-Verfahren dar. Durch die Verknüpfung unterschiedlicher

Verfahren ergibt sich eine fast beliebige Anzahl von Methoden zur Analyse relevanter Komponenten. Häufig handelt es sich bei diesen Verfahren um sehr spezifische Lösungsansätze, die wartungsintensiv sind und zumeist nicht die nötige Robustheit für den Einsatz in einem industriellen Umfeld besitzen.

Gängige, verhältnismäßig robuste Verfahren sind Analysen von niedermolekularen zell-externen Medienbestandteilen wie Zucker, Karbonsäuren und Aminosäuren mittels HPLC [Peucker (2006)][van de Merbel *et al.* (1992)][van de Merbel *et al.* (1995)]. In einigen Fällen können auch zell-externe rekombinante Proteine unter Verwendung von FPLC⁷-Techniken atline detektiert werden [Cornelissen (2004)]. Für die Analyse flüchtiger Substanzen wie Alkohole können Gaschromatographen (GC) eingesetzt werden, wobei auf Grund der besseren Automatisierbarkeit hierfür zumeist Massenspektrometer (MS) zum Einsatz kommen [Wittmann (2007)][Schügerl (2001)]. Diese wurden in den letzten Jahren auch vermehrt zur Analyse von Biopolymeren eingesetzt. Aufgrund der relativ aufwendigen Probenvorbereitung werden Massenspektrometer, außer für die Abgasanalytik, jedoch (noch) nicht zum atline Einsatz an Bioprozessen eingesetzt.

Die Fließinjektionsanalytik stellt eine sehr vielseitige und flexible Analysentechnik dar. Nahezu alle Komponenten, für die spezifische und sensitive Assays bestehen, können mit dieser Technik atline detektiert werden, da prinzipiell alle Detektoren die sich für den Durchflussbetrieb eignen, eingesetzt werden können. Interessant ist insbesondere der Einsatz von FIA-Systemen zum Nachweis von Biopolymeren. Durch die Automatisierung der sonst sehr pipetierintensiven Assays kann nicht nur eine höhere Genauigkeit erzielt werden [Klockewitz (2000)], sondern auch Zielproteinexpressionen atline überwacht werden [Reinecke and Scheper (1997)][Scheper *et al.* (1993)]. Einen guten Überblick über den Einsatz von FIA-Techniken an Bioprozessen bietet Schügerl (2001).

In der zweiten Generation der Injektionsanalytik wird durch den Einsatz von Spritzenpumpen (SIA – *Sequential Injection Analysis*) anstelle von peristaltischen Pumpen die Zuverlässigkeit und Standzeit der Systeme deutlich verbessert, wobei sich die Analysenzeiten durch die langsameren Spritzenpumpen vergrößern [Hansen (2004)]. Die anwendbaren Methoden bleiben aber die gleichen. Die dritte Generation der Injektionsanalytik ist die *Lab-On-Valve* Technologie, die eine Miniaturisierung der SIA-Technik darstellt [Wang and Hansen (2003)].

Verschiedene kommerzielle Anbieter vertreiben Multikomponentenanalysatoren, welche die Standardmessverfahren für mehrere relevante Komponenten in einem Messgerät zusammenfassen. Anbieter sind z.B. Nova Biomedical, Waltham, USA, mit ihrer BioProfile[®]-Palette und YSI Incorporated, Yellow Springs, USA, mit den YSI-Analysatoren. Analyten sind dabei z.B. Zucker, Alkohole, Ammonium und Phosphor aber auch Spurenelemente wie Kalium und Natrium. Teilweise gibt es auch Systeme mit bildverarbeitenden Methoden zur morphologischen Charakterisierung sowie zur Vitalitätsbestimmung der Zellen.

Eine weitere Methode zur Identifikation von Zellcharakteristika wie Zellgröße und Zellform, aber auch für die Beurteilung von Zellkomponenten wie DNA-Gehalt, Gesamtprotein- und Lipidgehalt stellt die Durchflußcytometrie dar. Atline-Applikationen sind jedoch eher selten [Sitton and Scrienc (2008)].

Abschließend bietet Tabelle 3.3 eine Übersicht über mögliche Verfahren zur Überwachung von Bioprozessen.

⁷FPLC – *Fast Liquid Chromatography*

Tabelle 3.3: Mögliche Messtechniken zur (quasi) Echtzeit-Überwachung von Bioprozessen

| Messmethode | Analyt | Literatur |
|-----------------------------|--|--|
| Dielektrische Spektroskopie | aktives Biovolumen, Zellgrößenverteilung | [Cannizzaro <i>et al.</i> (2003)], [Hauttimann and Müller (2001)] |
| Mikroskopie & Bildanalyse | Zellmorphologie | [Joeris <i>et al.</i> (2002)], [Lindner <i>et al.</i> (2007)] |
| Durchflußcytometrie | Zellmorphologie, Zellparameter wie Membranzustand, Vitalität, DNA/RNA-Gehalt, Gesamtprotein etc. | [Sitton and Scrienc (2008)], [Clementschitsch <i>et al.</i> (2005)] |
| 2D-Fluoreszenzspektroskopie | Zelldichte, Zucker, Ethanol, Aminosäuren, Stoffwechsellstand | [Stärk (2002)], [Solle (2003)], [Kolleck (2006)], [Marose <i>et al.</i> (1998)], [Surribas <i>et al.</i> (2006)], [Lantz <i>et al.</i> (2006)] |
| IR-Spektroskopie | Biomasse, Zucker, Karbonsäuren, Aminosäuren, Ammonium, Alkohole | [Crowley <i>et al.</i> (2005)], [Kormann <i>et al.</i> (2004)], [Kondepati and Heise (2008)], [Navratil <i>et al.</i> (2005)] |
| Raman-Spektroskopie | Glukose, Karotinoide | [Cannizzaro <i>et al.</i> (2003)] |
| Massenspektroskopie | Abgaskomponenten | [Wittmann (2007)], [Schügerl (2001)] |
| HPLC | Zucker, Karbonsäuren, Aminosäuren | [Peuker (2006)], [van de Merbel <i>et al.</i> (1992)], [van de Merbel <i>et al.</i> (1995)] |
| FPLC | zellexterne rekombinante Proteine | [Cornelissen (2004)] |
| Fließinjektionsanalytik | Zucker, Karbonsäuren, Formaldehyd, IgG's, Harnstoff, Penicilin V | [Horstkotte <i>et al.</i> (2006)] [Schügerl (2001)], [Reinecke and Scheper (1997)], [Scheper <i>et al.</i> (1993)] |

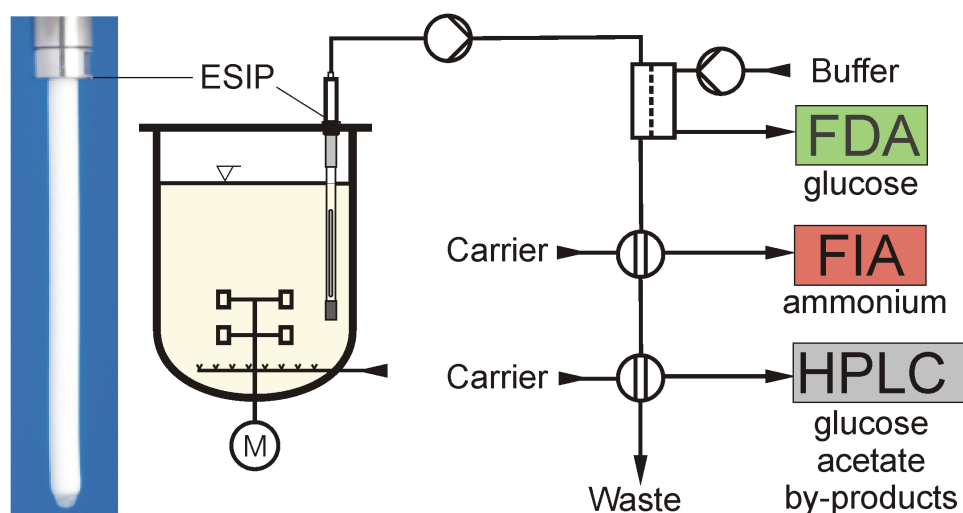


Bild 3.13: Die eingesetzte Atline-Messtechnik mit der verwendeten ESIP-Filtrationssonde

Die in dieser Arbeit eingesetzten Atline-Messsysteme sowie die serielle Anordnung ihrer Detektoren ist in Bild 3.13 schematisch dargestellt. Die zellfreie Probenahme wurde mit einer *insitu* sterilisierbaren Filtrationssonde (ESIP) der Fa. Trace Analytics, Braunschweig, realisiert. Der Filtrationsfinger besteht aus Polypropylen, hat eine definierte Porenweite von $0.2 \mu\text{m}$ und wird vor Inbetriebnahme mit Alkohol hydrophilisiert. Es lässt sich mit dieser Sonde ein maximaler Analysenstrom $F_{A\text{max}}$ von 2 mlmin^{-1} realisieren.

Die Fließdiffusionsanalytik (FDA) zur Glukosebestimmung und die Fließinjektionsanalytik (FIA) für die Ammonium-Stickstoff-Quantifizierung wurden während aller Prozesse eingesetzt, das System für die Hochleistungsflüssigkeitschromatographie (HPLC) wurde nur in

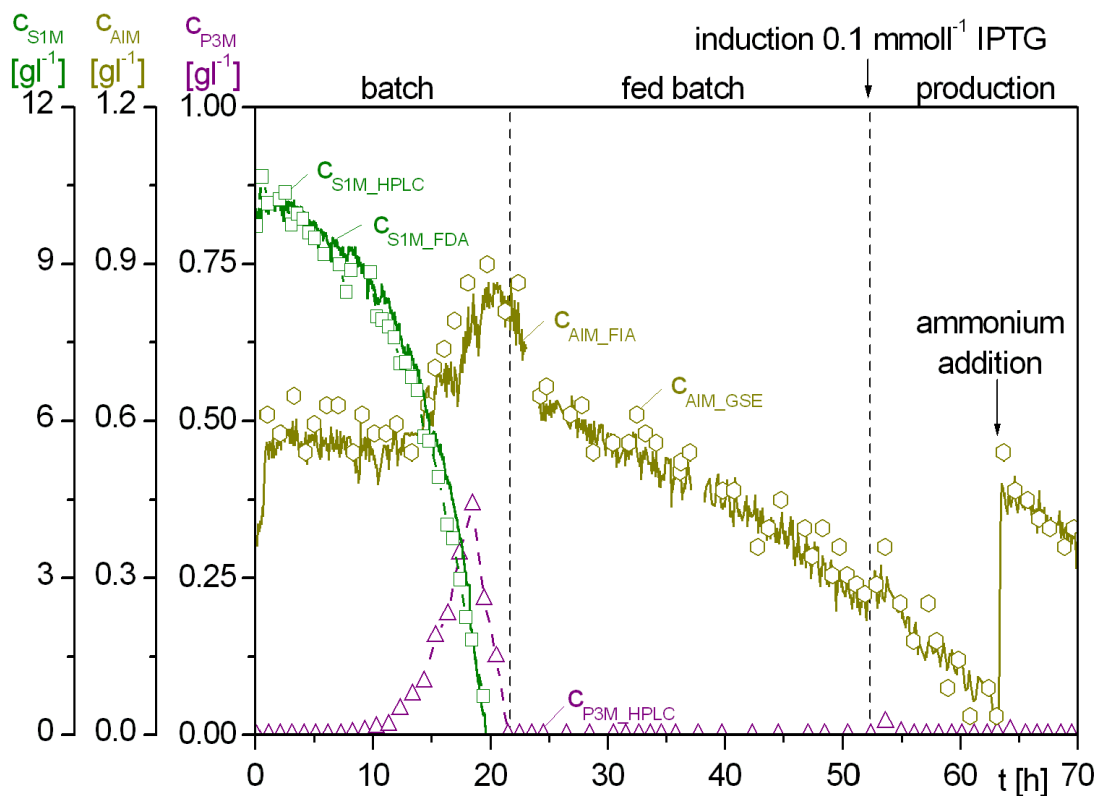


Bild 3.14: Messergebnisse der eingesetzten Atline-Systeme

einigen Fällen zur Validierung der Atline- und Offline-Methoden verwendet.

Exemplarisch sind Ergebnisse der eingesetzten Atline-Messungen in Bild 3.14 dargestellt. Es ist offensichtlich, dass die Ergebnisse der Atline-Glukosemessung mittels Fließdiffusionsanalytik (c_{SIM_FDA}) eine gute Übereinstimmung mit den Ergebnissen der Offline-HPLC-Messung (c_{SIM_HPLC}) zeigen.

Für die Validierung der Atline-Ammoniumanalytik (c_{AIM_FIA}) wurden Offline-Messungen mit einer gassensitiven Elektode durchgeführt (c_{AIM_GSE}). Auch hier zeigt sich eine gute Übereinstimmung der Messergebnisse über den gesamten Kultivierungsverlauf. Zur Vermeidung einer N-Limitierung wurde der Kultur gegen Ende des Prozesses sterile Ammoniumdihydrogenphosphat-Lösung zugesetzt. Die Erhöhung der Ammoniumkonzentration um 0.4 g l^{-1} entspricht dabei der zugegebenen Menge und bestätigt damit die Richtigkeit der Messung.

Die Konzentration des Stoffwechselnebenproduktes Acetat (c_{P3M_HPLC}) wurde mit der Atline-HPLC bestimmt. Die Messung wurde in weiteren Experimenten mit Hilfe von kommerziell erhältlichen enzymatischen Test validiert (Daten nicht dargestellt).

Im Folgenden wird auf die einzelnen Atline-Messsysteme näher eingegangen und die ihnen zugrunde liegenden Prinzipien kurz erläutert.

3.4.2 Glukoseanalytik mit der ProcessTrace

Für die Glukosebestimmung während der Batch-Phase sowie zum Monitoring einer Glukoseakkumulation während der Fed-Batch Phase wurde das in Bild 3.15 dargestellte Fließdiffusionsanalysengerät ProcessTrace der Fa. Trace Analytics eingesetzt.

Der Donor (Sample mit Analyt) strömt kontinuierlich über eine Diffusionsmembran und lädt dabei einen Akzeptor (Carrier) aus enzymstabilisierendem Puffer auf. Dieser Strom liefert

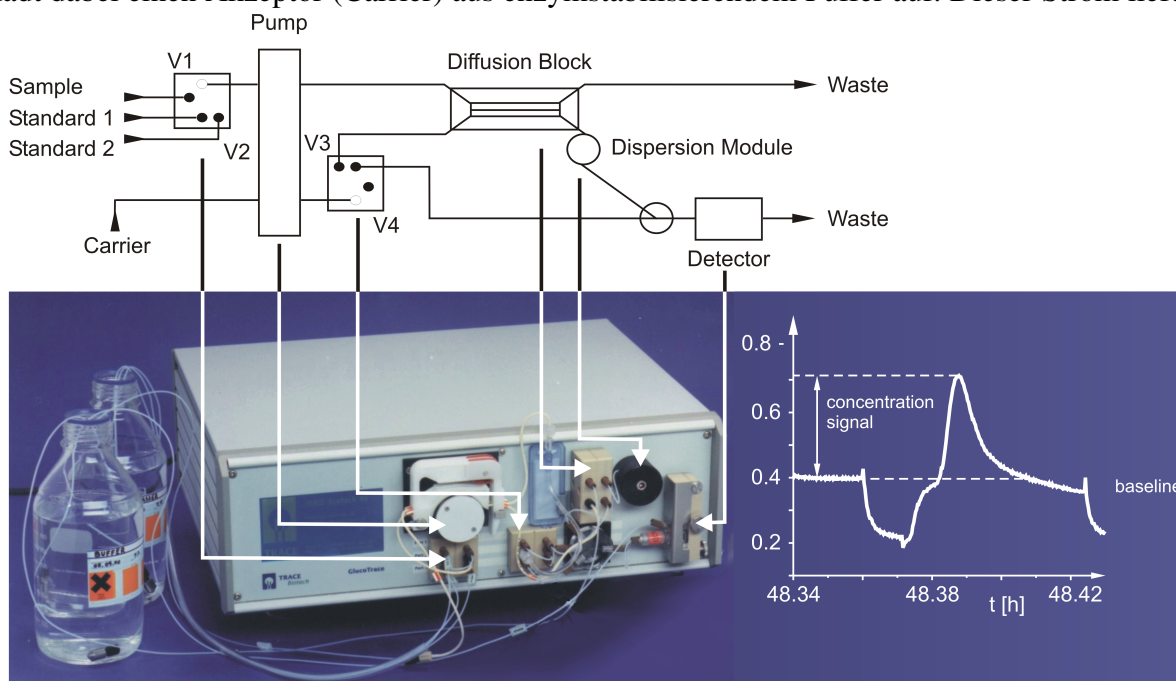


Bild 3.15: Das eingesetzte Fließdiffusionsanalysengerät ProcessTrace

im Detektor eine Basislinie. Im „stop-flow“-Verfahren wird dann ein definiertes Volumen des Akzeptors in Abhängigkeit der Analytkonzentration angereichert. Während dieser Zeit fließt unbelasteter Carrier zum Detektor. Nach Wiederaufnahme des Flusses erreicht das angereicherte Segment den Detektor und liefert dadurch den entsprechenden Konzentrationswert des Analyten. Die Glukoseanalyse erfolgt in der Messzelle über die spezifische Oxidation des Analyten mittels Glukoseoxidase. Das bei der Reaktion gebildete Wasserstoffperoxid wird anschließend an einer Pt-Elektrode reduziert, wobei pro Glukosemolekül zwei Elektronen gebildet und amperometrisch detektiert werden.

Die Glukosemessung findet im Bereich von $0.5 - 40 \text{ g l}^{-1}$ statt, die Standardabweichung vom oberen Messwert wird vom Hersteller mit 2 % angegeben.

3.4.3 Ammoniumanalytik mit der TAS2000

Für die Überwachung der Stickstoffquelle Ammonium während der durchgeführten Kultivierungen wurde ein Fließinjektionsanalysegerät TAS2000 der Fa. Jüke Systemtechnik, Altenberge, eingesetzt. Das modular aufgebaute System entspricht der modifizierten DIN 38406 T23 und ist in Bild 3.16 dargestellt. Die DIN-Norm wurde dahingehend angepasst, dass eine sequentielle anstatt einer kontinuierlichen Probeninjektion verwendet wurde. Außerdem musste der Analyt im Messsystem verdünnt werden, um eine zuverlässige Analyse im relevanten Messbereich von $0.2 - 2.0 \text{ g l}^{-1}$ ($0.012 - 0.12 \text{ mol l}^{-1}$) realisieren zu können.

Die Injektion eines Probesegementes von $20 \mu\text{l}$ in den kontinuierlich fließenden Reagenzienstrom R1 (Trinatriumcitrat) erfolgt im zeitlichen Abstand von ca. 5 min (Bild 3.16). In der ersten Reaktionsschleife erfolgt die Verdünnung mit dem Carrier Wasser. Bei Anwesenheit von Nitroprussid und Salicylat (R2) bildet der Ammoniumstickstoff im alkalischen Milieu mit Hypochlorid (R3) Chloramin. Die Reaktion erfolgt in einem auf 45°C temperierten Wasserbad, um das chemische Gleichgewicht konstant zu halten. Der blau-grüne Indophenolfarbstoff wird bei 605 nm im Durchflußphotometer detektiert und die Ammoniumkonzentration über die Integration der Peakfläche softwareseitig berechnet.

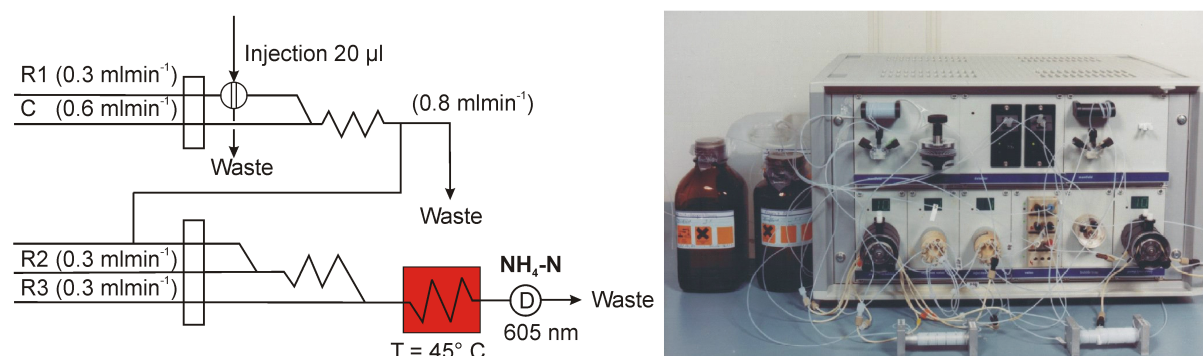


Bild 3.16: Das eingesetzte Fließinjektionsanalysegerät TAS2000

3.4.4 Atline-Messung mittels HPLC

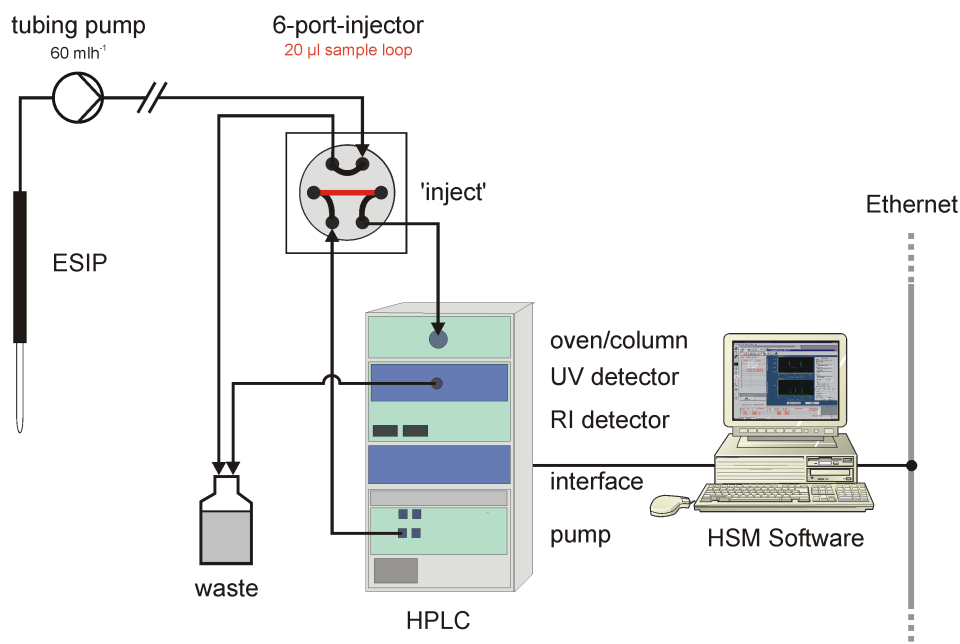


Bild 3.17: Aufbau eines Atline-Messverfahrens mit einer HPLC

Zur Atline-Quantifizierung von Acetat, Glukose, Laktat und Pyruvat wurde der in Bild 3.17 dargestellte Messaufbau mit einer HPLC der Fa. VWR International / Hitachi, Darmstadt vorgenommen.

Die Versorgung mit zellfreiem Medium erfolgt, wie bei allen eingesetzten Atline-Systemen, unter Verwendung einer ESIP Filtrationssonde sowie einer Peristaltikpumpe. Die Probenaufgabe erfolgt über ein Injektionsventil, dessen 20 µl große Probenschleife kontinuierlich gefüllt wird. Zu Beginn einer jeden Messung wird durch Schalten des Injektors das Probesegment auf die Säule aufgegeben.

Die Analyse der Probe erfolgt nach dem Prinzip der hydrophoben Wechselwirkungen (HIC⁸). Es wurde dafür eine Aminex HPX-87H Säule der Fa. Biorad, München, bei einer Ofentemperatur von 50 °C verwendet. Als Laufmittel wurde 25 mM H₂SO₄ mit einer Flußrate von 0.6 mlmin⁻¹ gewählt. Die Bestimmung von Acetat und Glukose erfolgte über einen RI⁹-Detektor während Laktat und Pyruvat über einen DA¹⁰-Detektor gemessen wurde. Eine Analyse dauerte dabei insgesamt 30 min.

⁸HIC – *Hydrophobic Interaction Chromatography*

⁹RI – *Refraction Index*

¹⁰DA – *Diode Array*

4 Automatisierungs- und Datenverarbeitungstechnik

4.1 Gewählter Automatisierungsaufbau

Die Hauptziele der Datenaufnahme sowie der Automatisierung sind die Prozessbeobachtung unter Online-Einbindung mathematischer Methoden sowie die Sicherstellung reproduzierbarer Kultivierungen. Dieses erfordert eine zentrale Datenverfügbarkeit für die Elemente der Prozesssteuerung sowie eine Datenverarbeitung mit MATLAB[®]. Die Automatisierungsinfrastruktur, welche die Datenübertragung zwischen dem Leitrechner und den einzelnen Komponenten der Prozessmess- und Steuerungstechnik ermöglicht, ist in Bild 4.1 dargestellt. Das zentrale Element stellt das PC-gestützte Prozessleitsystem MFCS¹/win dar.

Dieses SCADA²-System des Bioreaktorherstellers Sartorius Stedim Biotech wird für die Datenaufnahme und Automatisierung eingesetzt. Bei der Anbindung der einzelnen Teile der Prozesstechnik an den Leitrechner werden drei unterschiedliche Datenübertragungskonzepte

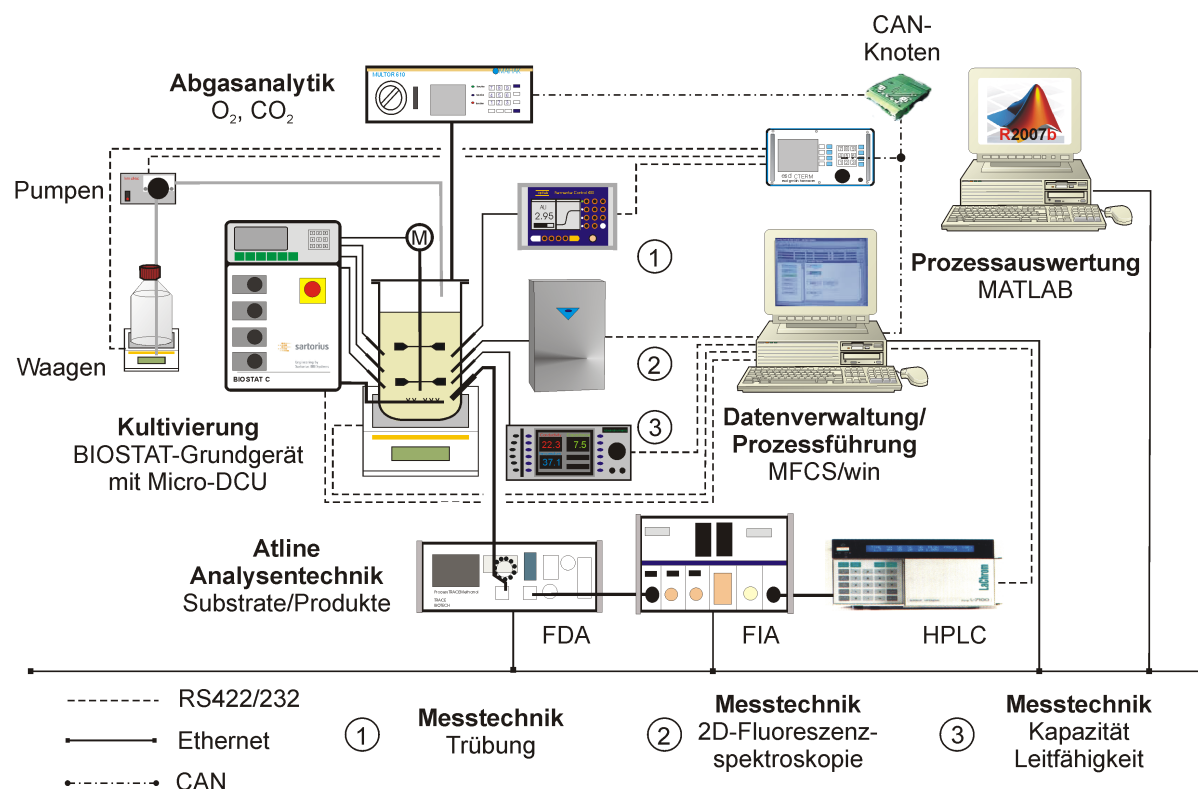


Bild 4.1: Der gewählte Automatisierungsaufbau

¹MFCS – Multi Fermenter Control System

²SCADA – Supervisory Control and Data Acquisition

verwendet. Die Auswahl der entsprechenden Schnittstelle hing dabei von den durch den jeweiligen Hersteller vorgegebenen Möglichkeiten, den Schnittstellenkapazitäten des Prozessrechners sowie den Ansprüchen an die Kommunikation ab.

Die Einbindung der Vorortautomatisierung des verwendeten Bioreaktors sowie der diversen Waagen und Analysengeräte erfolgt seriell. Zur Integration der Atline-HPLC, der TAS2000 und des MATLAB®-PCs wird das Ethernet verwendet. Alle weiteren Analysensysteme sowie die für einen Fed-Batch benötigte Zufütterungsperipherie sind mit Hilfe von I/O-Rechnern (esd – electronic systems design, Hannover) in den Leitrechner integriert. Dafür wird das Feldbussystem CAN – *Controller Area Network* eingesetzt, in das der Leitrechner mit einer PC-CAN Karte und einem MFCS/win-OPC Treiber integriert ist.

4.2 Anbindung der erweiterten Analysetechnik

Die zentrale Erfassung aller Messdaten sowie der Ergebnisse der Online-Berechnungen erfolgt in der Access® basierten Datenbank von MFCS/win. Für die Kommunikation der einzelnen Prozesskomponenten mit der Datenbank sind dabei diverse Treiber erforderlich. Die Verknüpfung der einzelnen Systeme ist in Bild 4.2 aufgezeigt.

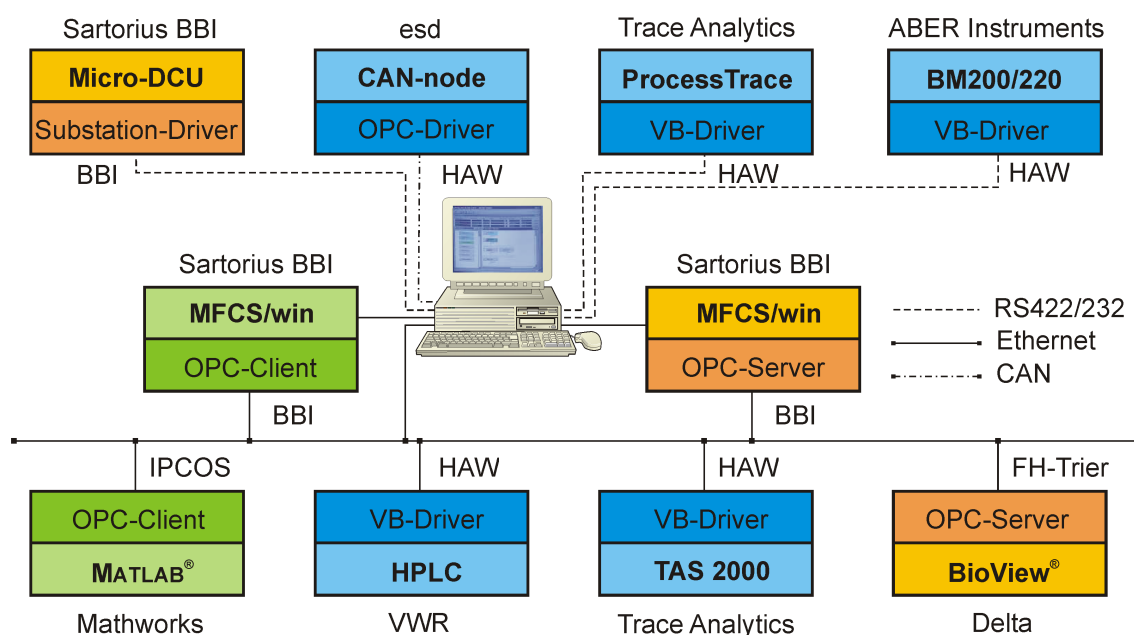


Bild 4.2: Anbindung der verwendeten Komponenten an MFCS/win

Für die seriell angebotenen Analysensysteme wurden an der HAW unter VisualBasic® Treiber entwickelt, über die das jeweilige RS232-Protokoll ausgelesen wird und die Daten anschließend auf die entsprechende MFCS-Variable geschrieben werden. Der Datenaustausch der über TCP/IP³ angebotenen HPLC sowie der TAS2000 wurde unter Verwendung von ODCB⁴ realisiert. Dabei werden die Messdaten auf eine USER-Datenbank, die sich ebenfalls auf dem Leitrechner befindet, geschrieben, die von MFCS/win zyklisch abgefragt wird.

Der Datenaustausch zwischen MATLAB® sowie dem eingesetzten CAN-Netz und MFCS erfolgte unter Verwendung des OPC-Standards. Diese herstellerübergreifende Schnittstelle er-

³TCP/IP – Transmission Control Protocol/Internet Protocol

⁴ODCB – *Open DataBase Connectivity*

möglicht die Datenübertragung mittels TCP/IP unter Verwendung einer Server/Client Technologie. Die Daten werden dabei vom OPC-Server von MFCS zur Verfügung gestellt und können so vom MATLAB[®]-Client entgegengenommen und ausgelesen werden. Dabei ist es dem MATLAB[®]-Client auch möglich, Daten auf den MFCS-Server zurückzuschreiben, die dann zyklisch vom Prozessleitsystem in die Datenbank aufgenommen werden.

Die Messdaten des BioView[®] stehen als XML-Daten zur Verfügung. Diese werden über einen vom Institut für Softwaretechnik des Umweltcampus Birkenfeld (Fachhochschule Trier) entwickelten OPC-Server zur Verfügung gestellt und zyklisch von MATLAB[®] abgefragt. Der Datentransfer von MATLAB[®] an MFCS erfolgt dann über die oben beschriebene OPC-Verbindung.

4.3 Das SCADA-System MFCS/win

MFCS/win ist ein modular aufgebautes, batchorientiertes Server/Client Leitrechner System. Es wurde in Version 2.1 für die Prozesssteuerung, Datenaufnahme und Prozessvisualisierung sowie für die Dokumentation der durchgeführten Experimente eingesetzt. In Bild 4.3 sind die einzelnen Module des Systems schematisch dargestellt.

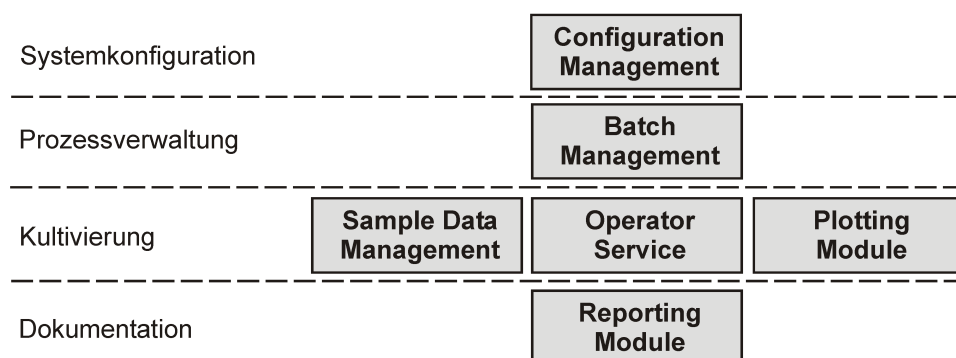


Bild 4.3: Module des SCADA-Systems MFCS/win

MFCS/win setzt sich aus sechs Modulen zusammen. Diese sind für die vier Hauptaufgaben, Konfiguration des Systems, Verwaltung der Kultivierungen, Durchführung der Kultivierungen sowie die Dokumentation der Prozesse und Systemeinstellungen zuständig.

Die Konfiguration des Systems erfolgt über das *Configuration Management*. Hier werden neben den grundsätzlichen Systemeinstellungen die benötigten Variablen und Oberflächen für die Kultivierungen definiert. Es können verschiedenartige Variablen angelegt und für diese Alar-me und Nachrichten vorgegeben werden. Daneben besteht die Möglichkeit einfache Berechnungen zu programmieren und online durchzuführen. Neben den Grundrechenarten können auch logische Operatoren verwendet werden. Für die Variablen lassen sich grafische Oberflächen konfigurieren, sogenannte *Trend-Displays* und *Variable-Displays*, die dann innerhalb des *Operator-Service* aufgerufen und zur Prozessvisualisierung genutzt werden können.

Für die Steuerung und Regelung der Prozesse können, zusätzlich zu den Standard-Regelkreisen der Frontend-Station Micro-DCU, über PID-Module neue Regler definiert und zur Prozessautomation verwendet werden. Die grafische Darstellung und der Zugriff auf die konfigurierten Regler im *Operator-Service*, erfolgt über *Group-Displays*, die zuvor im *Configuration Management* angelegt werden müssen.

Die Prozesse können mit Hilfe von Rezepten, die dem ISA⁵ S88-Standard genügen, geführt werden. Die Visualisierung erfolgt dabei über benutzerspezifische Fließschemata, den sogenannten Rezepten. Ein Rezept oder auch *Batch* besteht dabei immer aus den Blöcken INIT, FERMENTATION und TERMINATION. Der INIT Block dient der Initialisierung der benötigten Regler, Variablen und Berechnungen, der TERMINATION Block der Sicherstellung stabiler Zustände und dem abschließenden Speichern der Variablen für die Prozessdokumentation. Die einzelnen Blöcke sind dabei wieder in Phasen unterteilt.

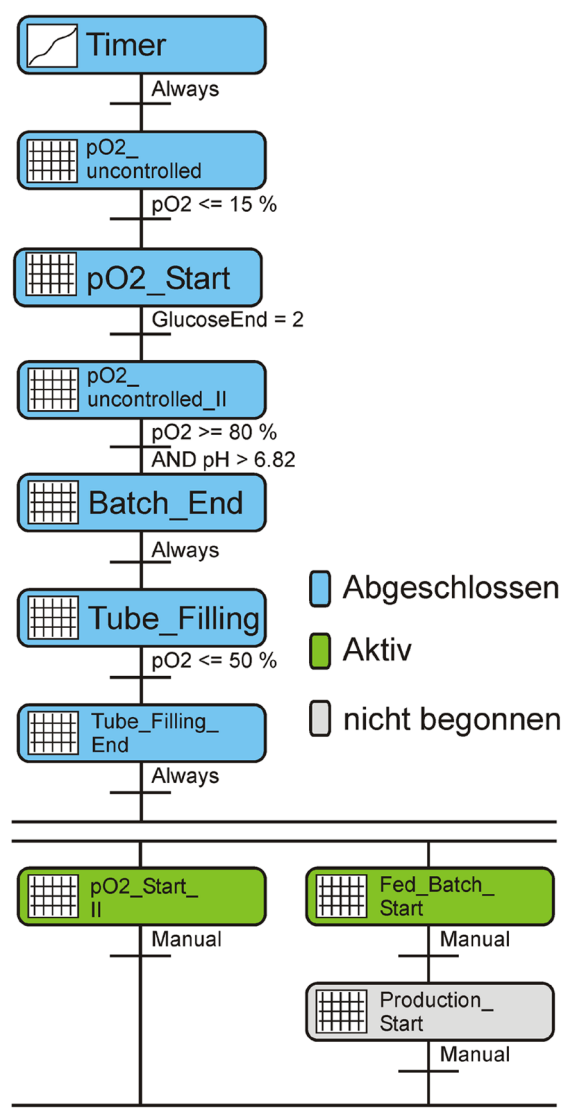


Bild 4.4: MFCs Rezept für die Prozessautomatisierung

und Nachrichten werden hier visualisiert und dabei akustisch unterlegt. Der *Operator Service* stellt damit quasi die Leitwarte des Prozesses dar.

Für die Einbindung der Ergebnisse von prozessbegleitenden Offline-Messungen steht das *Sample Data Management* zur Verfügung. Hier erfolgt die Eingabe der Messwerte (z.B. der OD-Messung) über eine grafische Oberfläche. Mit entsprechendem Zeitstempel werden die Er-

Die Automatisierungsaufgaben für den Prozessablauf befinden sich im Block Fermentation. Bild 4.4 zeigt die Standardautomatisierung der durchgeführten Prozesse. Dafür werden einzelne Prozessphasen definiert, die entweder über spezifische Prozessbedingungen oder manuell aktiviert und beendet werden. Innerhalb der Prozessphasen ist es möglich, Variablen zu initiieren, Berechnungen zu starten und zu beenden sowie Reglermodi und Sollwerte zu verändern. Es können mit Hilfe der Rezeptsteuerung bis zu 20 Automatisierungspfade parallel abgearbeitet werden. Die Programmierung von Schleifen ist jedoch nicht möglich, da ein bereits durchlaufener Prozessblock nicht erneut aktiviert werden kann.

Die Verwaltung der Kultivierungen erfolgt über das *Batch Management*. Hier können Prozesse, unter Verwendung der oben beschriebenen Rezepte, in der Datenbank von MFCs angelegt werden. Nach Prozessende können hier die durchgeführten Batches eingesehen, abgeschlossen oder gelöscht werden.

Für die Durchführung von Kultivierungen stehen drei Module zur Verfügung. Im Mittelpunkt steht dabei der *Operator Service*. Er beinhaltet die grafische Oberfläche für die Prozessführung und Prozessüberwachung. Hier werden die im *Batch Management* angelegten Prozesse initialisiert. Auf die vorkonfigurierten Regler und Variablen sowie die gewählte Prozessautomatisierung kann von hier zugegriffen und der Prozess damit geführt werden. Alle auftretenden Alarme

⁵ISA – The Instrumentation, Systems and Automation Society's Execution Board

gebnisse direkt in die MFCS-Datenbank geschrieben und stehen damit digital zur Verfügung.

Zur anspruchsvollen Visualisierung mehrerer Prozessvariablen kann das *Plotting Modul* verwendet werden. Es können alle Daten eines Prozesses aus der MFCS-Datenbank über der Zeit dargestellt werden. Außerdem können einzelne Grafiken vordefiniert und als Vorlagen abgespeichert werden.

Über das *Reporting Modul* können die Messdaten der durchgeführten Prozesse exportiert werden. Zudem können vollständige Berichte erstellt werden. Die Aufzeichnung von Alarmen und Nachrichten sowie die Erfassung von Online- und Offline-Messdaten werden dabei unter MFCS gemäß cGMP-Bestimmungen durchgeführt. Durch die Erweiterung um das optionale 21CFR11-Modul werden Benutzerwechsel und andere Prozesseingriffe voll dokumentiert (*Audit-Trail*) und erfüllen damit die Richtlinien von Part 11⁶.

4.4 Automatisierung des Bioreaktionsprozesses

Zur Sicherstellung reproduzierbarer Kultivierungsergebnisse wurde der Bioreaktionsprozess weitestgehend unter Verwendung eines S88-Rezeptes automatisiert. Ein typischer Prozessverlauf sowie die wichtigsten Aufgaben der Automatisierung sind in Bild 4.5 dargestellt. Im Wesentlichen lässt sich der Prozess in drei Phasen unterteilen. Die unlimitierte Batch-Phase, die substratlimitierte Fed-Batch Phase und die Produktionsphase, während der ebenfalls substratlimitiert Glukose zugefüttert wird.

Zu Beginn eines jeden Prozesses erfolgte die Initialisierung aller Variablen und Berechnungen sowie der verwendeten Regler über das gewählte MFCS-Rezept. Die Anfangsglukosekonzentration

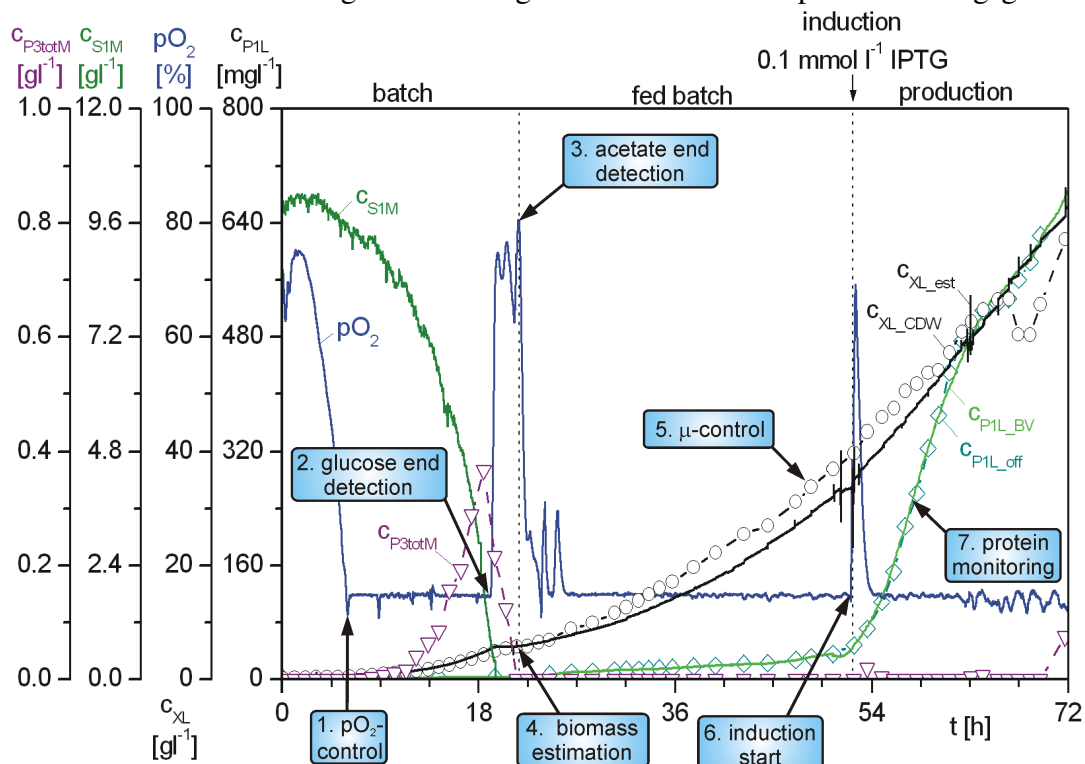


Bild 4.5: Biologische Größen eines typischen Prozessverlaufes

⁶Part 11 – *Electronic Records, Electronic Signatures – Scope and Application* – Richtlinie der FDA

zentration c_{S1M0} betrug dabei 10 gl^{-1} . Der pH-Wert wurde während des gesamten Prozess auf 6.8 mit einem Totband von 0.05 pH-Einheiten geregelt. Dafür wurden 25%ige Ammoniumhydroxidlösung und 3 molare Phosphorsäure vorgelegt, die jeweils als Stickstoff- bzw. Phosphor-Quelle für die Mikroorganismen dienen. Der Sollwert für die Flüssigphasentemperatur betrug 37°C . Bei einem Anfangsvolumen V_{L0} von 10 l erfolgte eine Luftbegasung mit 0.5 vvm. Ein ausreichender Sauerstofftransfer wurde zu Kultivierungsbeginn durch eine konstante Rührerdrehzahl des eingesetzten Scheibenrührers von 400 rpm gewährleistet. Alle Kultivierungen wurden auf M9 Minimalmedium [Riesenberg *et al.* (1991)] durchgeführt, welches gemäß DeLisa *et al.* (1999) angesetzt wurde.

Die Inokulation des Bioreaktors erfolgte mit einem Liter Vorkultur, wodurch sich Anfangszellichten von $0.15 - 0.3 \text{ gl}^{-1}$ ergaben. Durch die Inkubierung des Fermenterinhaltens wird die INIT-Phase der Prozessautomatisierung abgeschlossen. Im Anschluß werden die einzelnen Phasen der Prozessautomation des Blocks FERMENTATION, der in Bild 4.4 dargestellt ist, schrittweise durchlaufen.

Die erste Automatisierungsaufgabe während der Batch-Phase stellt die Initialisierung der $p\text{O}_2$ -Drehzahlregelung dar. Dies erfolgt, sobald der relative Sauerstoffpartialdruck in der Flüssigphase den Sollwert von 15% unterschreitet. Der Verbrauch der vorgelegten Glukose (c_{S1M}) wird durch ein starkes Ansteigen des $p\text{O}_2$ -Wertes und eine dadurch bedingte Reduktion der Rührerdrehzahl N_{St} indiziert. Beide Prozessgrößen werden für die Detektion des vollständigen Glukoseverbrauches verwendet. Um eine zuverlässige Batchend-Erkennung auch bei oszillierender $p\text{O}_2$ -Drehzahlregelung zu gewährleisten, muss, innerhalb von 4 Zyklen a 5 Sekunden, der $p\text{O}_2$ -Wert mindestens 23% überschreiten und die Rührerdrehzahl um mehr als 40 rpm sinken. In diesem Fall wird die $p\text{O}_2$ -Drehzahlregelung in MAN genommen, um die sich anschließende Polyauxie besser beobachten zu können.

Für die Mittelung der Messwerte sowie die Erkennung der oben aufgeführten Betriebszustände wird ein an der HAW entwickeltes Visual Basic[®] Tool, das sogenannte MFCS-Tool welches in Bild 4.6 dargestellt ist, eingesetzt.

Nach dem vollständigen Glukoseverbrauch beginnen die Zellen das zuvor gebildete Stoffwechselprodukt Acetat (c_{P3totM}) zu verstoffwechseln, wodurch sich der pH-Wert in den basischen Bereich verschiebt. Durch die wesentlich geringere Wachstumsgeschwindigkeit auf Acetat stellt sich ein Sauerstoffpartialdruck ein, der deutlich über 15% liegt. Nach dem vollständigen Verbrauch des Acetats steigt der $p\text{O}_2$ -Wert erneut an. Überschreitet dieser einen fest vorgegebenen Wert (hier 80 %) wird er zusammen mit dem pH-Wert, der durch den Abbau von Acetat höher als 6.8 liegt, als Bedingung für den Übergang zur Fed-Batch Phase verwendet.

Es wurden sowohl Prozesse gefahren, bei denen das gesamte Acetat vor Initialisierung der Fed-Batch Phase verstoffwechselt wurde (wie in Bild 4.5 dargestellt), als auch Prozesse, bei denen nur der Beginn des Acetatabbaus abgewartet wurde und das Acetat dann am Anfang der Fed-Batch Phase parallel zur limitiert zugegebenen Glukose verbraucht wurde.

Zu Beginn der Fed-Batch Phase wird die Begasungsrate F_{nAIR} auf 12 lmin^{-1} erhöht und die $p\text{O}_2$ -Drehzahlregelung wieder in den Betriebszustand AUTO gesetzt ($p\text{O}_{2w} = 15\%$) sowie die exponentielle Glukosezufütterung gestartet.

Die zellspezifische Wachstumsrate wird auf einen fest vorgegebenen Sollwert $\mu_w < \mu_{max}$ gesteuert und der Prozess dadurch substratlimitiert gefahren.

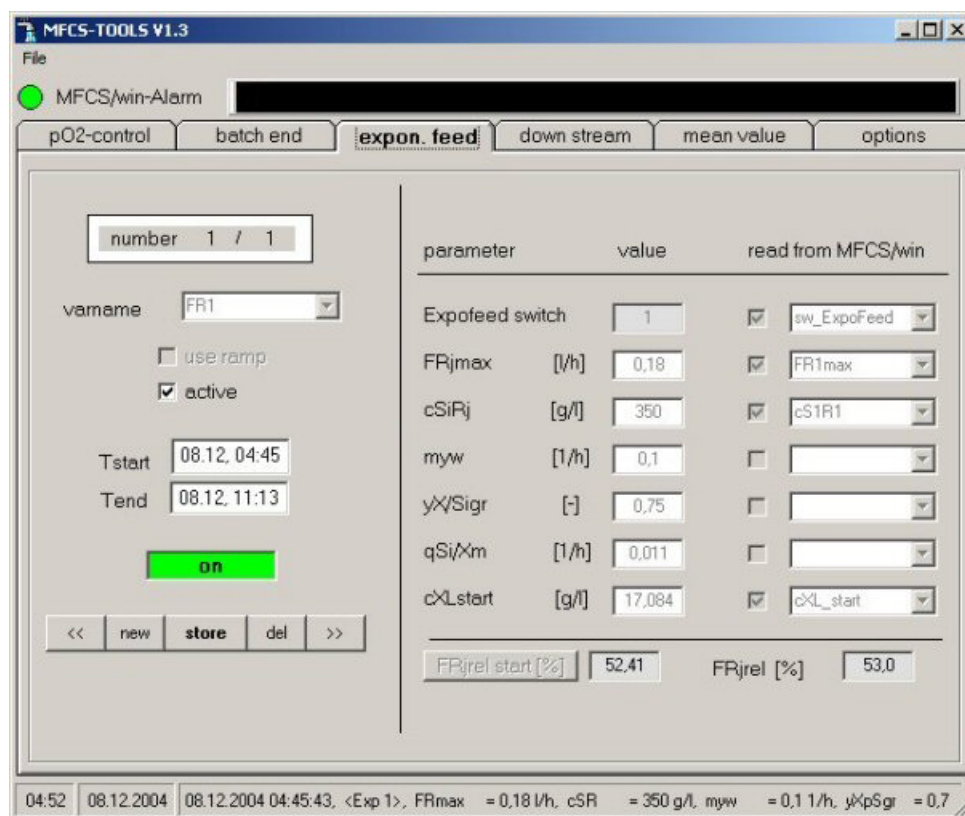


Bild 4.6: Grafische Oberflächen des MFCS-Tools

Der dafür vorzugebende Sollwert der relativen Zufütterrate F_{R1relw} ,

$$F_{R1relw}(t) = \frac{\mu_w(t_k) \cdot V_L(t_k) \cdot c_{XL}(t_k)}{y_{X/S1w}(t_k) \cdot c_{S1R1}} \cdot \frac{100\%}{F_{R1max}} \cdot e^{\mu_w(t_k) \cdot (t-t_k)}, \quad (4.1)$$

wird mit Hilfe des in Bild 4.6 gezeigten MFCS-Tools in die Prozessautomatisierung integriert.

Neben der Eingabe von festen Parametern, wie die Glukosekonzentration im Zufütterreservoir c_{S1R1} und die maximale Zufütterrate F_{R1max} , wird auch die Zelldichte c_{XL} und das Zellvolumen V_L zum Zeitpunkt t_k beim Start der Fed-Batch Phase oder bei Wechsel des μ -Sollwertes aktualisiert. Das Volumen sowie die Zelldichte stehen dabei als Online-Messgrößen zur Verfügung.

Der wachstumsgesteuerte Ausbeutekoeffizient auf Glukose $y_{X/S1w}$,

$$y_{X/S1w}(t_k) = \frac{\mu_w(t_k) \cdot y_{X/S1gr}}{\mu_w(t_k) + y_{X/S1gr} \cdot q_{S1/Xm}}, \quad (4.2)$$

wird dabei in Abhängigkeit von der vorgegebenen Wachstumsrate μ_w über die postexperimentell bestimmten Reaktionsparameter, Zuwachsausbeutekoeffizient auf Glukose $y_{X/S1gr}$ sowie Maintenanceanteil der zellspezifischen Glukoseaufnahme $q_{S1/Xm}$, berechnet. Die zugrunde liegenden reaktionskinetischen Ansätze sind in Kapitel 6 ausführlich dargestellt, die verwendeten Gleichungen sind ausführlich bei Peuker (2006) und Cornelissen (2004) beschrieben.

Sobald die gewünschte Induktionszelldichte erreicht ist, wird die Expression des rekombinanten Zielproteins durch die Zugabe des chemischen Induktors IPTG eingeleitet. Dabei wurde in den meisten Fällen mit einer Konzentration von 0.1 mmol l^{-1} gearbeitet.

5 Chemometrische Methoden und Modelle

Die Auswertung und Modellierung biotechnologischer Prozesse erfordert eine systematische Analyse und Bewertung der erzielten Messergebnisse. Die hierfür zur Verfügung stehenden mathematischen und statistischen Methoden werden in der chemischen Teildisziplin der Chemometrik zusammengefasst.

Die Chemometrik oder auch Chemometrie ist ein relativ junges Wissenschaftsgebiet, das sich Ende der Sechziger-, Anfang der Siebzigerjahre herausbildete und bald ein eigenständiges Arbeitsfeld etablierte [Danzer *et al.* (2001)]. Die Gegenstandsbezeichnung der *International Chemometrics Society* geht auf eine Definition von B.R. Kowalski [Kowalski (1978)] zurück und lautet:

„Die Chemometrik ist eine chemische Teildisziplin, die mathematische und statistische Methoden nutzt, um chemische Verfahren und Experimente optimal zu planen, durchzuführen und auszuwerten und so ein Maximum an chemisch relevanten, problembezogenen Informationen aus den experimentellen Messdaten zu gewinnen.“

Viele der Verfahren, die in Tabelle 5.1 aufgeführt sind, werden in den verschiedenen Teilgebieten der Chemie schon lange eingesetzt und waren in der Mathematik bekannt. Die Weiterentwicklung der Computertechnik hat die praktische Anwendbarkeit dieser Methoden auf umfangreiche und komplexe Daten ermöglicht und hat damit der Chemometrik zum Durchbruch verholfen. Die 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 machen die neue Qualität der Chemometrik aus [Danzer *et al.* (2001)].

Tabelle 5.1: Übersicht über die Methoden der Chemometrik [Danzer *et al.* (2001)]

| Stochastik | Datenanalyse | Modellierung | Systemtheorie |
|----------------------------|---|----------------------------|-------------------------------------|
| Statistik Fehlertheorie | Faktorenanalyse Hauptkomponenten- analyse | Korrelationsanalyse | Informationstheorie |
| Versuchsplanung | Clusteranalyse | Multivariate Regression | Signaltheorie Signalverarbeitung |
| Optimierung | Mustererkennung | Softmodellierung (PLS) | Zuverlässigkeitstheorie |
| Prozesskontrolle | Klassifikation | Kalibration | Operations Research |
| Probenahmethorie | Datenvorbehandlung | | |

5.1 Motivation

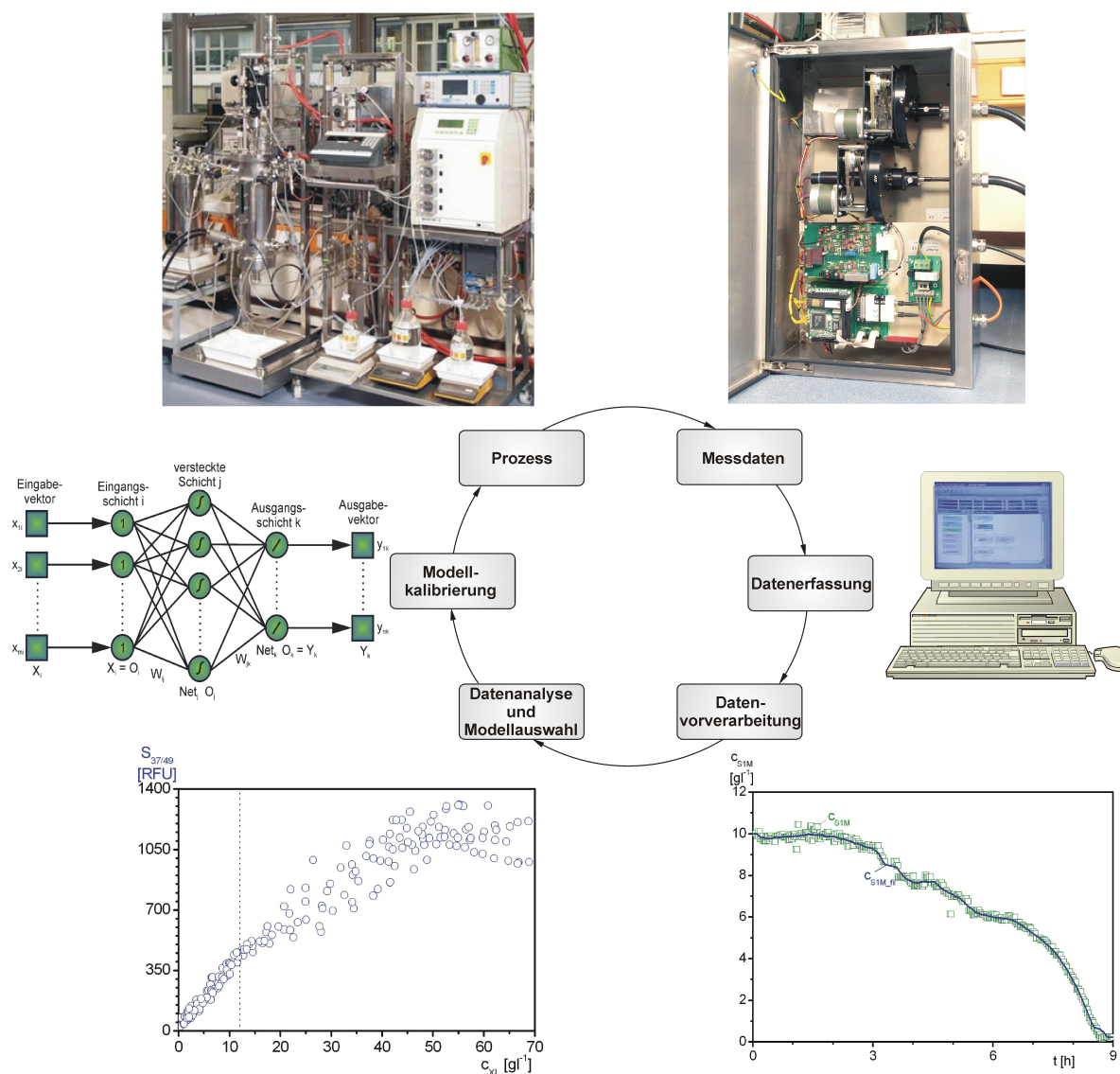


Bild 5.1: Vom Prozess zum Modell

Eines der Ziele dieser Arbeit war die Prozessüberwachung auf Basis der 2D-Fluoreszenzspektroskopie. Bild 5.1 zeigt dabei den Weg auf, der vom Prozess bis zur Kalibrierung eines geeigneten datengetriebenen Modells durchlaufen werden muss.

Die ersten drei Schritte, der Prozess und die darin verfolgte Prozessstrategie, die verwendeten Messsysteme und die dabei eingesetzten Techniken der Datenerfassung, sind bereits in den vorherigen Kapiteln erläutert worden. Die Datenvorverarbeitung, Datenanalyse und Modellauswahl sowie die Modellkalibrierung sind Gegenstand dieses Kapitels.

Zunächst müssen die aufgezeichneten Messdaten einer geeigneten Vorverarbeitung unterzogen werden, die Methoden der Ausreissererkennung, Datenfilterung und Dateninterpolation beinhaltet.

Anschließend kann eine Datenanalyse durchgeführt werden. Hierfür werden klassische Methoden wie die Korrelationsanalyse eingesetzt. Diese werden sowohl auf die Originalspektren

als auch auf latente Variable, die mit Hilfe von Datenreduktionsmethoden wie der Hauptkomponenten¹- und der unabhängigen Komponentenanalyse² ermittelt werden, angewandt.

Auf Basis der aus der Datenanalyse gewonnenen Erkenntnisse können dann geeignete lineare und nichtlineare Modelle ausgewählt werden, die sich im letzten Schritt mit Hilfe der zur Verfügung stehenden Messdaten kalibrieren lassen. Anschließend können die Modelle zur Prozessüberwachung eingesetzt werden.

Alle hier aufgeführten Methoden gehören dabei in das Fachgebiet der Chemometrik. Im Verlauf dieses Kapitels soll nun an ausgewählten Beispielen der Weg von der Datenvorverarbeitung bis zum einsatzfähigen Modell sowie die verwendeten Methoden näher erläutert werden.

5.2 Erweiterte Datenverarbeitung über Einbindung von MATLAB[®]

Für die Auswertung der mit Hilfe der in Kapitel 3 beschriebenen Methoden gewonnenen Messdaten wurde in dieser Arbeit die wissenschaftliche Software MATLAB[®] in der Version 7.5 Release 2007b eingesetzt. MATLAB[®] ist ein mathematisch orientiertes Programmpaket und bietet eine interaktive Umgebung an, um rechenintensive Aufgaben schneller als mit herkömmlichen Programmiersprachen wie Fortran, C oder C++ lösen zu können.

Neben den mathematischen Grundfunktionen gibt es eine Reihe von Toolboxen, die Sammlungen von Funktionen für bestimmte Applikationen, wie z.B. die Bioinformatik, statistische Verfahren oder auch künstliche neuronale Netze enthalten. Durch eine Reihe von grafischen Benutzeroberflächen wird dem Anwender der Einstieg in die Funktionen der Toolboxen erleichtert. Die für diese Arbeit verwendeten Toolboxen sind in Tabelle 5.2 aufgeführt.

Sämtliche mathematischen Auswertungen, chemometrische Methoden und Modelle sowie die Simulationen, die in den weiteren Kapiteln zu finden sind, wurden unter Verwendung von MATLAB[®] durchgeführt.

Tabelle 5.2: Übersicht der verwendeten Toolboxen

| Toolbox | Version | Eigenschaft |
|--|---------|--|
| Genetic Algorithm and Direct Search TB | V2.2 | Erweiterung der Optimization TB um genetische Algorithmen und weitere Funktionen zur Minimalsuche |
| Neural Networks TB | V5.1 | Verschiedene Typen neuronaler Netze und entsprechende Trainingsalgorithmen |
| Optimization TB | V3.1.2 | Optimierungsalgorithmen, Minimierung, nichtlineare Gleichungslöser, PLS |
| Statistics TB | V6.1 | Statistische Tests, Deskriptive Statistik, Verteilungsfunktionen, Regressions- und Varianzanalyse, multivariate Statistik, Klassifikation, Design of Experiments |
| Symbolic Math TB | V3.2.2 | Symbolische Mathematik |

¹PCA – *Principal Component Analysis*

²ICA – *Independent Component Analysis*

Für die dynamische Prozesssimulation wurde zusätzlich die von Prof. Dr.-Ing. Klaus-Uwe Gollmer, Umweltcampus Birkenfeld, FH Trier entwickelte Toolbox BLOSS³ verwendet. Diese bietet den Vorteil, dass die Modellgleichungen in symbolischer Form programmiert werden können. Die Umsetzung in die matrixbasierte Schreibweise erfolgt dann über den Precompiler von BLOSS, der auf die Symbolic Toolbox und den darin enthaltenen Maple-Kernel zurückgreift.

Für die Einbindung mathematischer Methoden zur Online-Prozessüberwachung wurde die Software, wie in Abschnitt 4.2 dargestellt, in das verwendete SCADA-System MFCS/win integriert.

5.3 Datenvorverarbeitung

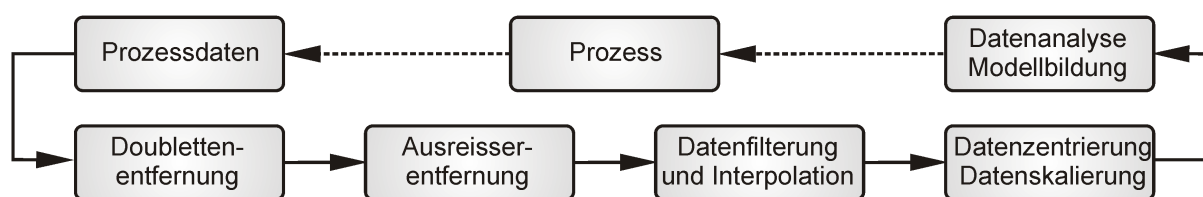


Bild 5.2: Die einzelnen Schritte der Datenvorverarbeitung

Die einzelnen Schritte der Datenvorverarbeitung sowie die Reihenfolge in der sie durchlaufen werden, sind in Bild 5.2 aufgezeigt. Dabei ist es nicht notwendig, immer alle Verfahren auf die jeweiligen Messdaten anzuwenden. Begründungen für den Einsatz der einzelnen Verfahren finden sich in den nachfolgenden Unterkapiteln.

Die Doubletten- und Ausreisserentfernung sowie die Datenfilterung wird im Folgenden zunächst an einer exemplarischen Atline-Glukosemessung während der Batch-Phase eines Prozesses erläutert. Die verwendeten Interpolationsmethoden werden anhand von Offline-Messdaten vorgestellt. Die Datenzentrierung und Datenskalierung, die am Beispiel der Fluoreszenzspektroskopie dargestellt wird, erfolgt aus didaktischen Gründen zu Beginn von Abschnitt 5.4.

5.3.1 Doublettenentfernung

In Bild 5.3 sind die Messdaten einer Glukoseanalytik für die Batch-Phase eines Prozesses gezeigt. Die Messung erfolgte, wie in Kapitel 3 erläutert, atline mit einer Fließdiffusionsanalyse.

Bedingt durch das eingesetzte SCADA-System MFCS/win wird, jedesmal wenn ein neuer Messwert zur Verfügung steht, der vorhergehende Wert sowie der aktuelle Messwert in die Datenbank geschrieben. Dieses führt dazu, dass jeder Messwert zweimal vorliegt (Doublette).

Um eine daraus resultierende falsche Gewichtung der Messwerte bei der Anwendung mathematischer und statistischer Methoden zu vermeiden, muss der redundante (zum späteren Zeitpunkt gespeicherte) Messwert vor Verwendung der Daten entfernt werden. Dieses kann über ein einfaches MATLAB[®]-Skript realisiert werden, in welchem der jeweilige Wert mit dem

³BLOSS – *Biotechnological Simulation System*

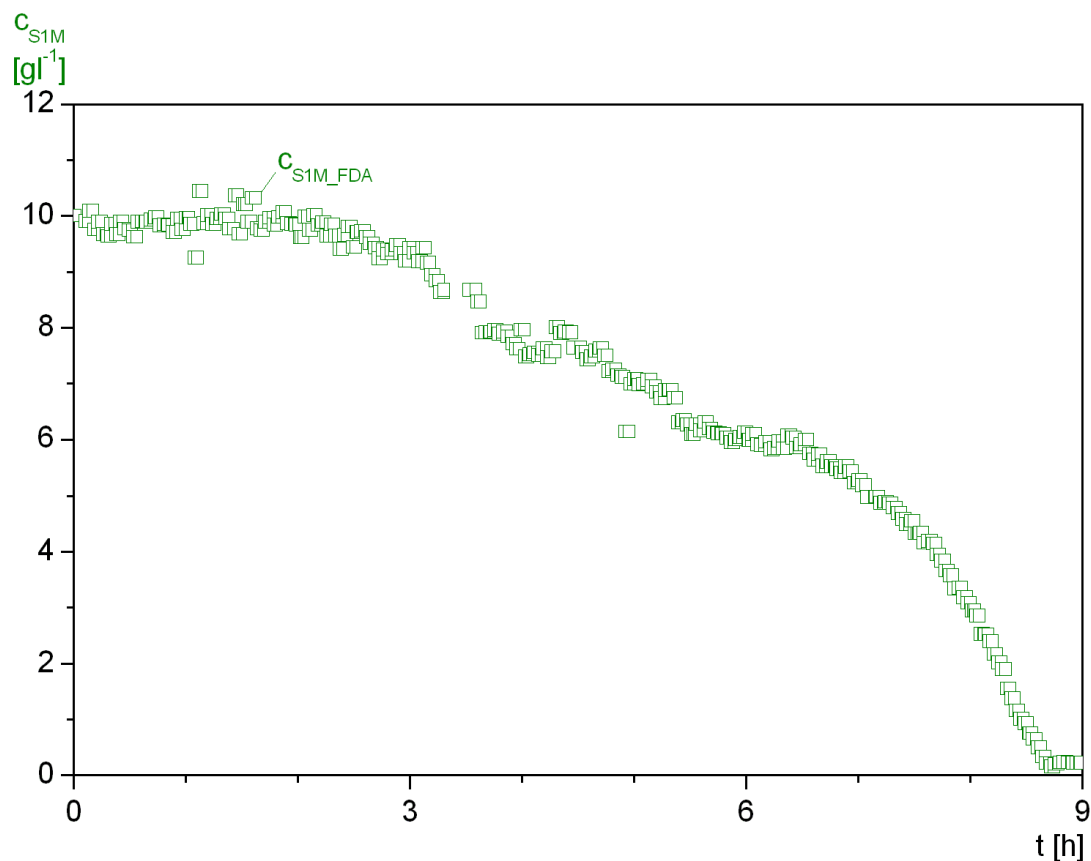


Bild 5.3: Messdaten einer Atline-Glukoseanalytik mit Doubletten

vorherigen verglichen wird. Haben beide Messwerte den gleichen Betrag, so wird der jeweils letzte aus dem Datensatz gelöscht.

Um zu verhindern, dass Werte aus dem Datensatz entfernt werden, die tatsächliche Messergebnisse darstellen und auf eine zweimalige Messung mit dem selben Ergebnis zurückzuführen sind, muss eine weitere Bedingung in das Skript aufgenommen werden. Tritt ein Messwert viermal auf, so sind nur der zweite und vierte Wert aus dem Datensatz zu löschen. Der erste und dritte Wert stellen dann die tatsächlichen Messergebnisse dar.

Die Zeitintervalle zwischen den Messdaten können hierfür als Indikator verwendet werden. Da MFCS den jeweils aktuellen Wert immer 15 Sekunden nach dem alten Messwert in die Datenbank schreibt, kann dieser darüber erkannt und aus dem Datensatz entfernt werden.

5.3.2 Ausreissererkennung

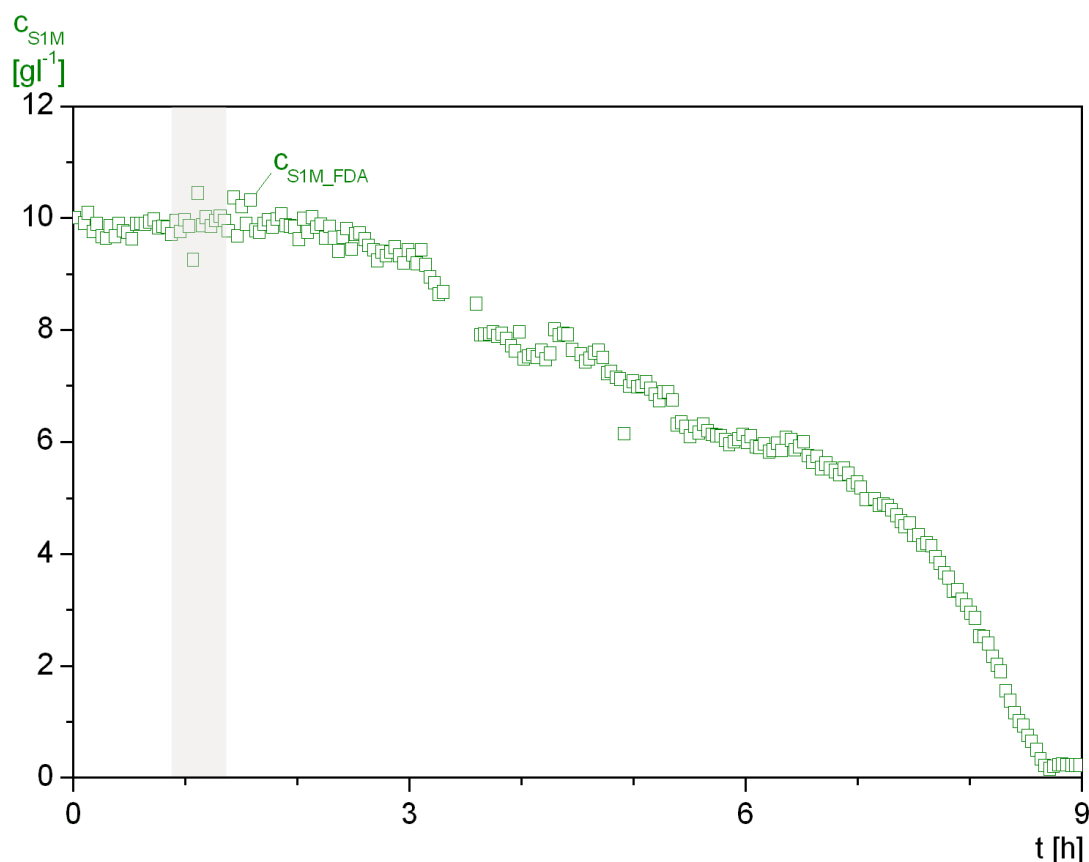


Bild 5.4: Messdaten der von Doubletten befreiten Atline-Glukoseanalytik mit Ausreißern

Im nächsten Schritt sollen Ausreißer aus dem Datensatz entfernt werden, die auf eine fehlerhafte Funktion des jeweiligen Messsystems zurückzuführen sind.

Die Ausreißerentfernung ist insbesondere für die Messdaten relevant, die für eine Kalibrierung der Modelle eingesetzt werden sollen. Da fehlerhafte Ergebnisse oft extreme Abweichungen aufweisen, haben sie einen negativen Einfluss auf die zur Modellkalibrierung verwendeten Regressionsmethoden und verfälschen somit die Vorhersagen der Modelle.

Die Ausreißer wurden auf Grundlage des Grubbs-Tests [Grubbs (1969)] detektiert und aus dem Datensatz entfernt. Der Test basiert auf der Annahme einer normalverteilten Grundgesamtheit für 3 - 150 Einzelmessungen [Otto (1997)]. Dies ist für die Messdaten zumeist nicht gegeben, kann bei einer ausreichend kleinen Intervalllänge jedoch näherungsweise angenommen und der Test somit angewendet werden.

Das in Bild 5.4 gewählte Zeitintervall ist mit 25 Minuten gewählt worden und grau hinterlegt. Trägt man die in diesem Intervall liegenden Messdaten im Wahrscheinlichkeitsnetz einer Normalverteilung auf, so ergibt sich bei normalverteilten Daten eine Gerade. Die Ordinate ist dafür so skaliert, dass die Verteilungssummenfunktion $F(c_{S1M})$ der Normalverteilung bei Auftragung linearisiert wird.

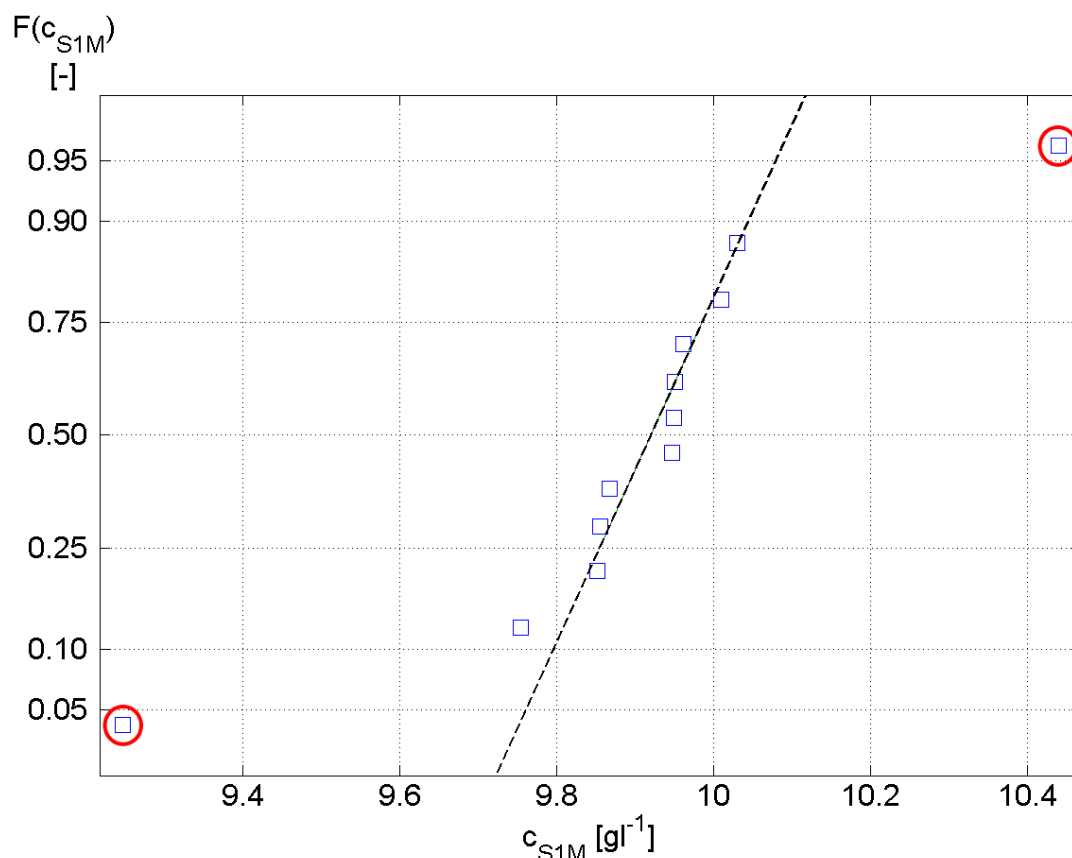


Bild 5.5: Messwerte des betrachteten Zeitintervalls im Wahrscheinlichkeitsnetz der Normalverteilung

Wie aus Bild 5.5 ersichtlich wird, ist diese Voraussetzung für die Messdaten im gewählten Zeitintervall näherungsweise gegeben. Die beiden eingekreisten extremen Werte, die nicht auf der Geraden liegen, müssen dabei im folgenden Algorithmus als Ausreißer detektiert werden.

Vor Durchführung der Überprüfung werden jeweils der Mittelwert \bar{x} ,

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}, \quad (5.1)$$

sowie die Standardabweichung σ ,

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}, \quad (5.2)$$

für die Messwerte im entsprechenden Zeitintervall bestimmt.

Zur Durchführung des Tests wird die Nullhypothese, dass ein Messwert x_i aus n Messungen kein Ausreißer ist, angewendet. Sie wird verworfen, wenn die Prüfgröße T_i ,

$$T_i = \frac{|x_i - \bar{x}|}{\sigma} < T_{n;1-\alpha}, \quad (5.3)$$

in Abhängigkeit des gewählten Signifikanzniveaus α und der Anzahl der Messungen die Bedingung nicht erfüllt. Es wird dabei der Abstand des jeweiligen Messwertes x_i vom Mittelwert im Verhältnis zur Standardabweichung beurteilt.

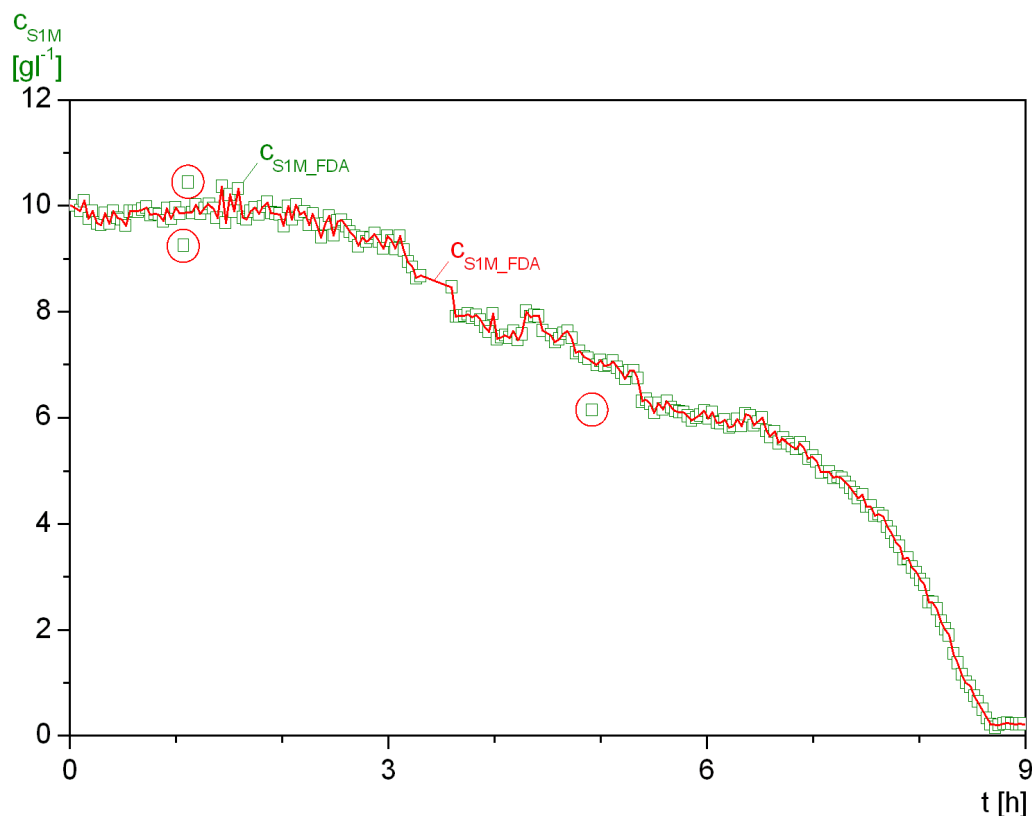


Bild 5.6: Messdaten der Atline-Glukoseanalytik mit identifizierten Ausreißern

Der kritische Wert $T_{n;1-\alpha}$ des Grubbs-Tests,

$$T_{n;1-\alpha} = \frac{n-1}{\sqrt{n}} \cdot \sqrt{\frac{t_{\alpha/2n,n-2}^2}{n-2+t_{\alpha/2n,n-2}^2}}, \quad (5.4)$$

lässt sich in Abhängigkeit des kritischen Wertes⁴ der Studentischen t-Verteilung mit einem Freiheitsgrad von $n-2$ berechnen [Hartung (1989)]. Die Quantile der t-Verteilung sind in MATLAB[®] hinterlegt, so dass auf diese zugegriffen werden kann.

Bild 5.6 zeigt die gekennzeichneten Ausreißer (eingekreiste Symbole) sowie die von den Ausreißern befreiten Daten (Linie). Es wurde für den Test ein Signifikanzniveau von 85 % verwendet, wodurch die drei im Plot markierten Ausreißer detektiert und aus dem Datensatz entfernt werden konnten.

⁴auch Quantile genannt

5.3.3 Datenfilterung

Datenfilter werden nach Otto (1997) in der Vorbehandlung von analytischen Signalen mit dem Ziel eingesetzt,

- das Signal Rausch-Verhältnis zu verbessern,
- die Ableitung des Signals zur Subtraktion des Untergrundes und zur Verbesserung der visuellen Auflösung zu realisieren,
- die Integration des Signals für quantitative Auswertungen zu ermöglichen.

Vor dem Hintergrund der hier vorliegenden Aufgabenstellung steht vor allem die Verbesserung des Signal Rausch-Verhältnisses im Vordergrund. Es stehen dafür eine Reihe von Filteralgorithmen zur Verfügung.

Die einfachste Möglichkeit bieten gleitende Mittelwerte, bei denen der gefilterte Wert x_k^* ,

$$x_k^* = \frac{1}{2m+1} \sum_{j=-m}^{j=m} x_{k+j}, \quad (5.5)$$

aus den Rohsignalwerten x_k innerhalb der Filterbreite $2m$ berechnet wird.

Ein weiterer sehr verbreiteter Algorithmus ist das Savitzky-Golay (SG) Filter. Bei dieser Glättung werden die Rohdaten unterschiedlich gewichtet, was einer Filterung mit einem Polynom zweiter oder dritter Ordnung entspricht (beide weisen die gleichen Filterkoeffizienten auf).

Nach Festlegung der Filterbreite kann der k -te Datenpunkt x_k^* ,

$$x_k^* = \frac{1}{\text{NORM}} \sum_{j=-m}^{j=m} a_j \cdot x_{k+j}, \quad (5.6)$$

berechnet werden. Die benötigten Filterkoeffizienten a_j können aus entsprechenden Tabellen entnommen werden, wobei NORM der Summe der Koeffizientenwerte entspricht und als Normalisierungsfaktor bezeichnet wird [Otto (1997)]. Um das SG-Filter anwenden zu können, müssen die Messdaten zuvor auf eine äquidistante Zeitachse interpoliert werden.

Ein weiterer Vorteil des Savitzky-Golay Filters besteht darin, dass mit Hilfe von tabellarisch dokumentierten Filterkoeffizienten die Berechnung der zeitlichen Ableitungen des Messsignals über Gl. (5.6) erfolgen kann.

Weitere Möglichkeiten der Datenfilterung stellen mathematische Transformationen, wie z.B. die Fourier- oder die Hadamard-Transformation, dar. Dabei werden die Daten nach der Transformation mit geeigneten Filterfunktionen multipliziert und anschließend wieder in die Zeitdomäne rücktransformiert. Des Weiteren stehen eine Reihe von Filtern zur Verfügung, die auf der Kombination von lokaler Regression und polynomischer Interpolation basieren. Einen tieferen Einblick in die Filtertheorie bieten Shumway and Stoffer (2006) und Chatfield (2003).

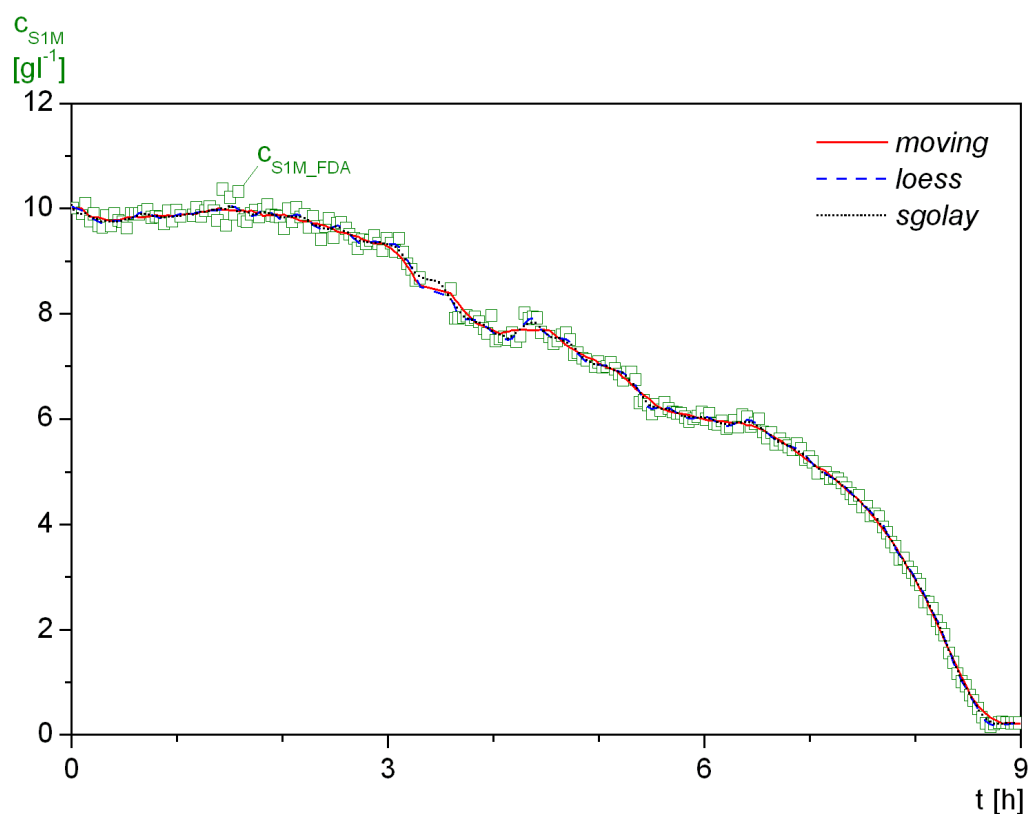


Bild 5.7: Unterschiedliche Methoden der Datenfilterung

Für die Datenfilterung wurden die unter MATLAB[®] verfügbaren Filterfunktionen eingesetzt. Exemplarisch sind die Ergebnisse für drei dieser Datenfilter in Bild 5.7 aufgezeigt. Für alle Filter wurde das gleiche Zeitintervall wie für die Ausreissererkennung gewählt.

Bei den dargestellten Ergebnissen wurde zum Einen ein gleitender Mittelwertfilter (*moving*), das Savitzky-Golay Filter (*sgolay*) sowie eine Kombination aus lokaler Regression und linearer Interpolation (*loess*) eingesetzt.

Es ist offensichtlich, dass alle drei Filter bei dem hier vorliegenden Messsignal sehr ähnliche Ergebnisse zeigen. Der gleitende Mittelwert zeigt dabei die stärkste Signalglättung. Im Fall der Filterung des Glukosesignals wurde deshalb im Folgenden diese Methode als Filter gewählt.

Die Auswahl des geeigneten Filters erfolgte für jede Signalquelle individuell. In den allermeisten Fällen der hier vorliegenden Signale reicht der gleitende Mittelwert jedoch aus.

5.3.4 Dateninterpolation

In vielen Fällen ist es sinnvoll bzw. notwendig, die erzielten Messdaten auf eine äquidistante Zeitachse zu interpolieren. Das im vorherigen Abschnitt vorgestellte Savitzky-Golay Filter erfordert z.B. äquidistante Zeitabstände. Des Weiteren kann die Präzision sowie die Geschwindigkeit von numerischen Differentialgleichungslösern durch den Vergleich von Simulationsergebnissen mit Messwerten, die auf einer sinnvoll gewählten Zeitachse liegen, entscheidend verbessert werden (Kapitel 6).

Zur Durchführung von Korrelationsanalysen (Abschnitt 5.4.2) aber auch zum Training von

Neuronalen Netzen (Abschnitt 5.6) müssen die eingesetzten Datensätze die gleiche Zeitachse aufweisen. Auch hierfür müssen Verfahren der Dateninterpolation eingesetzt werden.

Für die Messdateninterpolation wurden mehrere unter MATLAB[®] verfügbare Verfahren erprobt. Diese sind die lineare Interpolation (*linear*), die Interpolation über kubische Splines (*spline*) und das Verfahren der stückweisen Hermite-Interpolation (*pchip*⁵).

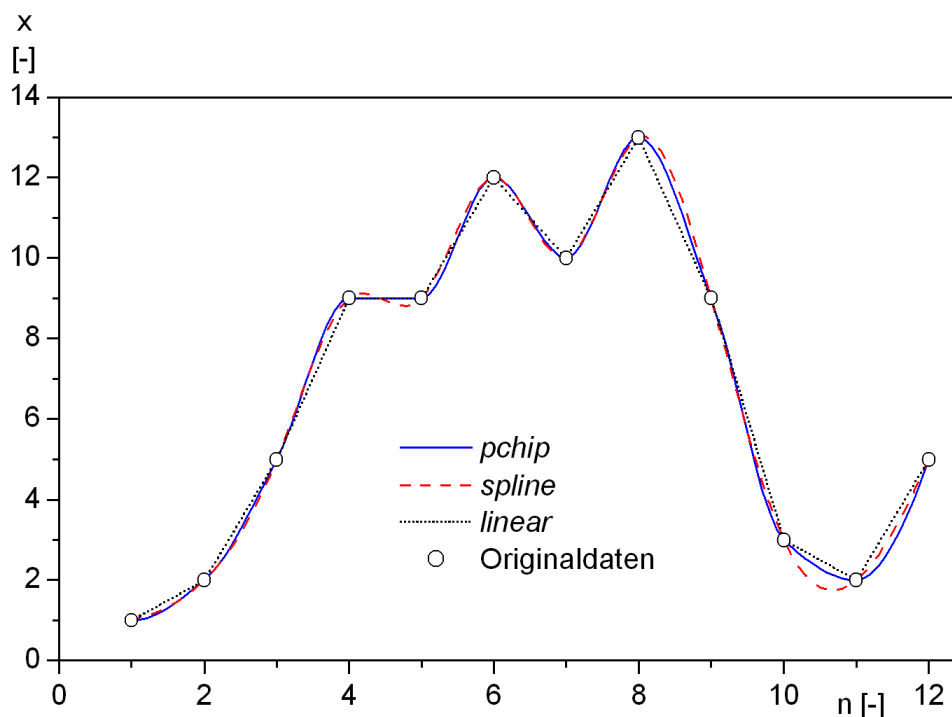


Bild 5.8: Vergleich unterschiedlicher Methoden der Dateninterpolation

Letztere Methode bietet gegenüber der Interpolation mit kubischen Splines den Vorteil, dass sie kein Überschwingen verursacht und weniger zur Oszillation neigt, wie auch aus Bild 5.8 ersichtlich wird. Eine ausführliche Beschreibung dieser Verfahren findet sich bei Hyman [Hyman (1983)] sowie Fritsch [Fritsch and Carlson (1980)].

Zur Identifikation des jeweils geeigneten Interpolationsverfahrens wurde eine Kreuzvalidierung [Brereton (2003)] für die vorhandenen Datensätze durchgeführt. Dafür wurde aus den zu interpolierenden Datensätzen, bei einer ungeraden Anzahl an Messpunkten, jeder zweite Wert aus dem Datensatz entfernt. Anschließend wurden die Daten mit dem entsprechenden Verfahren interpoliert.

Die Gütebewertung der Verfahren erfolgte über die Wurzel des mittleren quadratischen Fehlers, den RMSE⁶,

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (c_i^* - c_i)^2}{n-1}}, \quad (5.7)$$

wobei c_i den tatsächlichen und c_i^* den interpolierten Konzentrationswert darstellen.

⁵pchip – piecewise cubic hermite interpolating polynomial

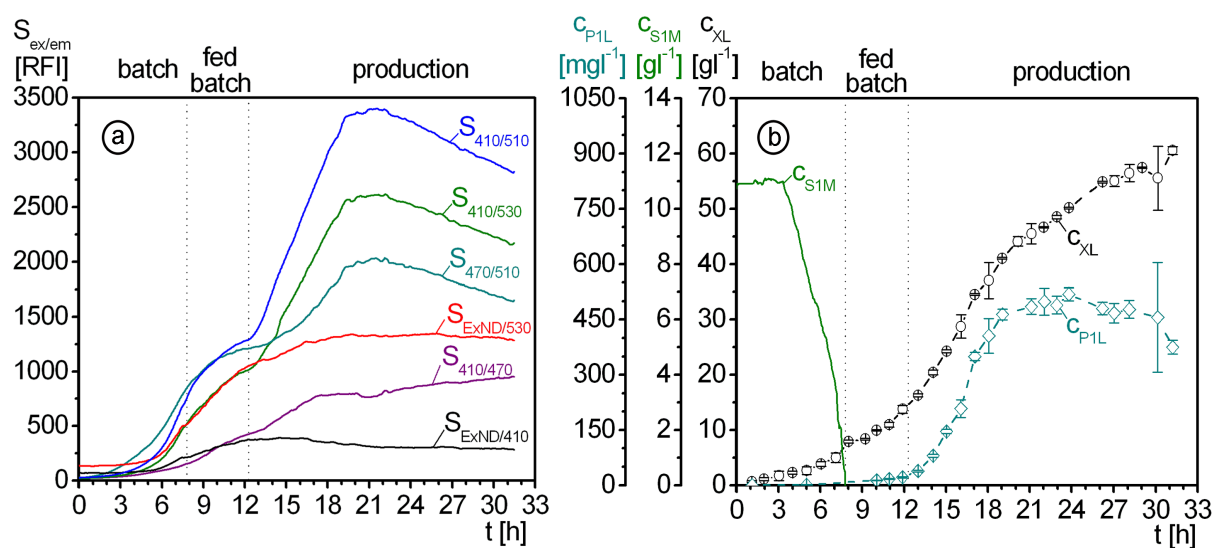
⁶RMSE – Root Mean Square Error

Tabelle 5.3: Mittels Kreuzvalidierung bestimmte Fehler von Interpolationsverfahren für unterschiedliche Messgrößen

| Messgröße | Fehlerart | Interpolationsmethode | | |
|-----------|-----------------------------|-----------------------|---------------|--------------|
| | | <i>linear</i> | <i>spline</i> | <i>pchip</i> |
| c_{XL} | RMSE [gl^{-1}] | 0.7507 | 0.7540 | 0.7259 |
| | RMSE [%] | 1.453 | 1.458 | 1.406 |
| c_{SIM} | RMSE [gl^{-1}] | 0.2883 | 0.2928 | 0.2898 |
| | RMSE [%] | 4.198 | 4.242 | 4.213 |
| c_{PIL} | RMSE [mg l^{-1}] | 20.00 | 21.84 | 19.17 |
| | RMSE [%] | 3.875 | 3.947 | 3.465 |

Die Fehler der einzelnen Verfahren sind für die Messgrößen Zelldichte c_{XL} , Glukosekonzentration c_{SIM} und die Zielproteinkonzentration c_{PIL} in Tabelle 5.3 aufgeführt. Der prozentuale Fehler ergibt sich, indem der RMSE durch den Wertebereich des Datensatzes geteilt wird. Dadurch ist die Vergleichbarkeit der einzelnen Ergebnisse untereinander gegeben.

5.4 Kalibrierung linearer univariater Modelle

**Bild 5.9:** Darstellung einiger BioView®-Signale (a) und der Zielgrößen des Prozesses (b)

Nachdem die Messdaten entsprechend den vorherigen Abschnitten aufbereitet worden sind, ist es die Aufgabe der Datenanalyse, mathematische Zusammenhänge zwischen den 2D-Fluoreszenzspektren und den restlichen Messdaten aufzudecken. Exemplarisch sind einige Wellenlängenkombinationen der BioView®-Signale sowie die Zielgrößen des Prozesses in Bild 5.9 dargestellt.

Zunächst sollen dafür univariate Modelle zur Prozessbeobachtung vorgestellt werden. Dabei gilt es, lineare Beziehungen zwischen den Prozessschlüsselgrößen Zelldichte c_{XL} , der Gesamtzielproteinkonzentration c_{PIL} für die gesamte Prozessdauer sowie der Glukosekonzentration c_{SIM} während der Batch-Phase und einzelnen Wellenlängenpaaren der BioView®-Signale aufzufinden.

Die Messignale der 2D-Fluoreszenzmessungen stellen mathematisch die Originalmatrix X ,

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{pmatrix} \begin{array}{l} \text{Spektrum Messung 1} \\ \text{Spektrum Messung 2} \\ \vdots \\ \text{Spektrum Messung } m, \end{array} \quad \downarrow t \quad (5.8)$$

dar. Sie ist eine $m * n$ Matrix mit den 150 Wellenlängenkombinationen in den n Spalten und den einzelnen Messungen über der Zeit in den m Zeilen. Die Breite der Matrix ist damit durch das Messgerät und die Länge der Matrix durch die Prozesszeit und die Messintervalle festgelegt.

Die einzelnen Elemente der Matrix x_{ik} sind durch ihre Indizes ik somit eindeutig dem i -ten Messzeitpunkt sowie der k -ten Wellenlängenkombination zugeordnet. Der zeitliche Ablauf der Messung eines Wellenlängenpaares wird daher durch den Spaltenvektor x_k dargestellt.

Bevor im Folgenden das prinzipielle Vorgehen und die Ergebnisse der univariaten Modelle unter Verwendung der BioView[®]-Messungen vorgestellt werden, soll noch kurz die Datenzentrierung und Skalierung erläutert werden, die ein wichtiges Werkzeug der Datenvorverarbeitung mehrdimensionaler Daten darstellt.

5.4.1 Datenzentrierung und Skalierung

Die Datenzentrierung führt zur Eliminierung des Ordinatenabschnittes in den prozessierten Daten. Da sich dadurch die benötigten Matrixoperationen für uni- und multivariate Kalibrierung sowie bei Datenreduktionsmethoden wie z.B. der Hauptkomponentenanalyse vereinfachen, ist sie ein häufig verwendetes Verfahren.

Die zentrierten Messwerte x_{ik}^{cen} ,

$$x_{ik}^{cen} = x_{ik} - \bar{x}_k, \quad (5.9)$$

ergeben sich durch Subtraktion des jeweiligen Spaltenmittelwertes \bar{x}_k der k -ten Spalte vom Variablenwert x_{ik} in Zeile i und in Spalte k [Otto (1997)]. Die Messwerte werden dabei so verschoben, dass ihr Schwerpunkt im Ursprung des Koordinatensystems liegt.

Falls die Eingangsdaten einen starken Hintergrund aufweisen (z.B. Eigenfluoreszenz des Mediums), ist dieser vor der Anwendung der Datenzentrierung zu entfernen, da es sonst zu einer Gewichtung des Hintergrundes in den zentrierten Daten kommt [Brereton (2003)].

Die Datenskalierung wird eingesetzt, um zentrierte Messwerte mit stark verschiedenen Wertebereichen nicht unterschiedlich zu gewichten. Dabei sind mehrere Verfahren in der Literatur beschrieben [Brereton (2003)][Otto (1997)][Danzer *et al.* (2001)]. Sie werden insbesondere beim Training neuronaler Netze, aber auch bei der Parameteridentifikation dynamischer Systeme eingesetzt.

Die Datenskalierung kann aber auch dazu führen, dass z.B. Wellenlängenkombinationen, die nur Rauschen enthalten, mit in die Gewichtung der anschließend verwendeten multivariaten Methoden einfließen. Um dieses zu verhindern, muss eine geeignete Vorauswahl der einzelnen Messgrößen schon vor der Datenskalierung vorgenommen werden [Brereton (2003)][Danzer *et al.* (2001)].

Es wurden in dieser Arbeit zwei Verfahren zur Datenskalierung verwendet. Bei der Standardisierung⁷ werden die zentrierten Daten x_{ik}^{cen} durch die Standardabweichung der k-ten Spalte σ_k dividiert,

$$x_{ik}^{std} = \frac{x_{ik}^{cen}}{\sigma_k}. \quad (5.10)$$

Die so skalierten Messwerte zeichnen sich dann durch einen Mittelwert von 0 aus, besitzen eine Varianz von 1 und die Länge der Vektoren beträgt $\sqrt{n-1}$.

Eine einfachere Methode, bei denen die Daten nur auf einen Wertebereich von -1 bis +1, aber nicht hinsichtlich ihrer Varianz skaliert werden, ist die Bereichsskalierung. Unter Verwendung des größten, x_k^{max} , sowie des kleinsten Wertes x_k^{min} der jeweiligen Spalte k, können dann die bereichsskalierten Messpunkte x_{ik}^{skal} ,

$$x_{ik}^{skal} = 2 \cdot \frac{x_{ik} - \bar{x}_k}{x_k^{max} - x_k^{min}}, \quad (5.11)$$

berechnet werden.

5.4.2 Methodik der linearen Modellkalibrierung

Das prinzipielle Vorgehen bei der Kalibrierung und der Beurteilung der erstellten linearen Modelle ist in Bild 5.10 aufgezeigt. Die elf zur Verfügung stehenden Datensätze wurden dafür entsprechend der vorangegangenen Abschnitte aufbereitet und anschließend in acht Trainingsdatensätze für die Parameteridentifikation und drei Datensätze für die Modellvalidierung unterteilt. Ein Datensatz besteht dabei immer aus den BioView[®]-Signalen (Eingangsdaten) sowie den zugehörigen Prozessvariablen (Zieldaten) einer Kultivierung.

Zum Auffinden linearer Zusammenhänge zwischen den einzelnen Wellenlängenkombinationen und den zu modellierenden Prozessgrößen wurde zunächst eine Korrelationsanalyse durchgeführt.

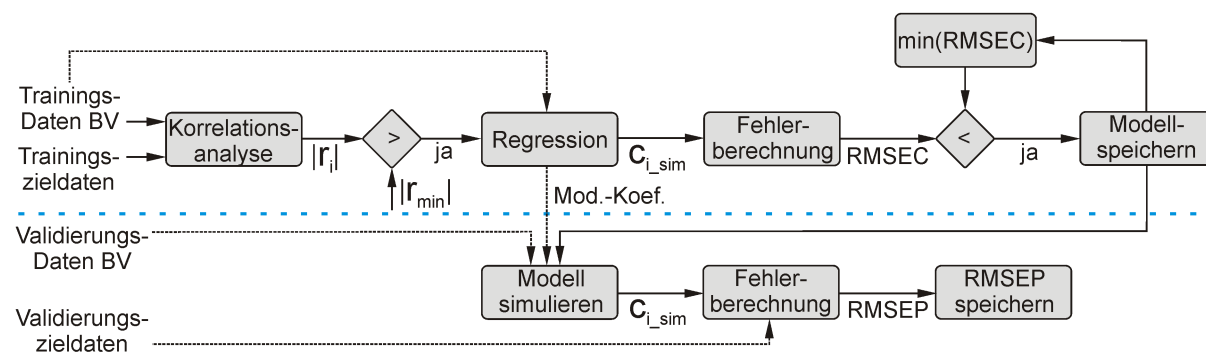


Bild 5.10: Prinzipielles Vorgehen bei der Kalibrierung univariater Modelle

⁷auch Autoskalierung genannt

Um die Korrelationsanalyse anwenden zu können muss zunächst die Varianz-Kovarianzmatrix⁸ berechnet werden. Die Kovarianz $\text{cov}(x_j, x_k)$ zwischen zwei gleich großen Vektoren x_j und x_k ,

$$\text{cov}(x_j, x_k) = \frac{1}{n-1} \sum_{i=1}^m (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k) \quad j, k = 1 \dots n; j \neq k, \quad (5.12)$$

ist ein Maß für die Streuung zweier Variablenpaare zueinander. Sie stellt das zweite zentrale Moment der Statistik für zweidimensionale Daten dar und ist damit das Pendant zur Varianz eindimensionaler Daten, die als Quadrat der Standardabweichung aus Gl. (5.2) definiert ist.

Im n -dimensionalen Raum ergibt sich die Kovarianzmatrix C ,

$$C = \begin{pmatrix} \sigma^2(x_1) & \text{cov}(x_1, x_2) & \dots & \text{cov}(x_1, x_n) \\ \text{cov}(x_2, x_1) & \sigma^2(x_2) & \dots & \text{cov}(x_2, x_n) \\ \vdots & \dots & \ddots & \vdots \\ \text{cov}(x_n, x_1) & \text{cov}(x_n, x_2) & \dots & \sigma^2(x_n) \end{pmatrix}, \quad (5.13)$$

aus den Kovarianzen der Variablenpaare sowie den Varianzen der Einzelvariablen als symmetrische $n \times n$ Matrix. Die Hauptdiagonale der Kovarianzmatrix ist dabei mit der jeweiligen Varianz des entsprechenden Datenvektors besetzt.

Bei der Verwendung von zentrierten Daten vereinfacht sich Gl (5.13) entsprechend zu

$$C = \frac{X^T \cdot X}{m-1}. \quad (5.14)$$

Da die Kovarianz von der Metrik der Daten abhängig ist, kann sie nur bei Variablen mit gleichen (skalierten) oder ähnlichen Wertebereichen zur Beurteilung herangezogen werden. Eine von diesen Parametern unabhängige Größe stellt der (Personsche) Korrelationskoeffizient bzw. die Korrelationsmatrix dar.

Der Korrelationskoeffizient r_{jk} ,

$$r(x_j, x_k) = \frac{\text{cov}(x_j, x_k)}{\sigma(x_j)\sigma(x_k)} := r_{jk}, \quad (5.15)$$

berechnet sich aus dem Quotienten der entsprechenden Kovarianz und dem Produkt der zugehörigen Standardabweichungen. Er ist damit ein dimensionsloses Maß und liegt immer im Wertebereich von -1 bis +1. Werte nahe +1 weisen dabei auf einen proportionalen Zusammenhang mit positiver Steigung und Werte nahe -1 auf einen proportionalen Zusammenhang mit negativer Steigung und geringer Streuung hin.

Bei der Auftragung eines Datenpaares gegeneinander, wie in Bild 5.11 dargestellt, ergibt sich bei hoher Korrelation nahezu eine Gerade. Bei Werten, die nahe Null liegen, besteht kein linearer Zusammenhang. Es ergibt sich bei der Auftragung eine strukturlose Punktwolke.

⁸auch Dispersions- oder kurz Kovarianzmatrix genannt

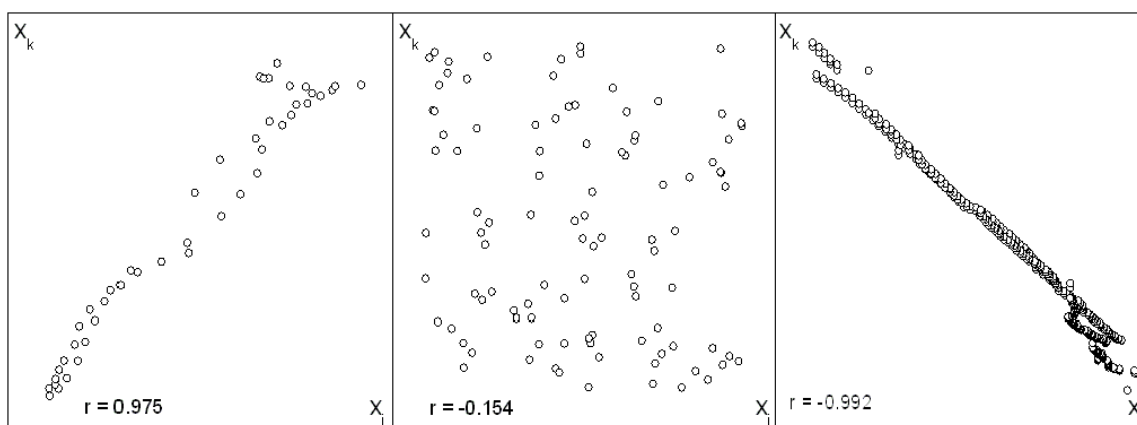


Bild 5.11: Positive, keine sowie negative Korrelation eines Variablenpaares x_j, x_k

Bei der Verwendung autoskalierter Daten (Abschnitt 5.4.1, Gl. (5.10)) zur Berechnung der Kovarianzmatrix entspricht diese der Korrelationsmatrix R ,

$$R = \begin{pmatrix} 1 & r_{12} & \dots & r_{1n} \\ r_{21} & 1 & \dots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \dots & 1 \end{pmatrix}. \quad (5.16)$$

Bei der Interpretation des Korrelationskoeffizienten ist zu beachten, dass er keine Aussage über einen kausalen Zusammenhang der betrachteten Variablenpaare macht, sondern lediglich die Wahrscheinlichkeit eines linearen Zusammenhanges beschreibt. Bei der Beurteilung der erzielten Ergebnisse muss daher stets das Wissen um die Herkunft der betrachteten Datensätze mit einfließen, um Fehlinterpretationen zu vermeiden.

Sind die Korrelationskoeffizienten r der einzelnen BioView[®]-Signale mit den entsprechenden Prozessgrößen des Trainingsdatensatzes bestimmt, werden diese im Anschluß mit einem Grenzwert $|r_{\min}|$ verglichen. Im nächsten Schritt des in Bild 5.10 aufgezeigten Vorgehens bei der Modellkalibrierung werden dann über eine lineare Regression die Parameter des entsprechenden Modells identifiziert. Für diesen Vorgang, der auch als Modellkalibrierung bezeichnet wird, stehen die in Bild 5.12 dargestellten Verfahren zur Verfügung.

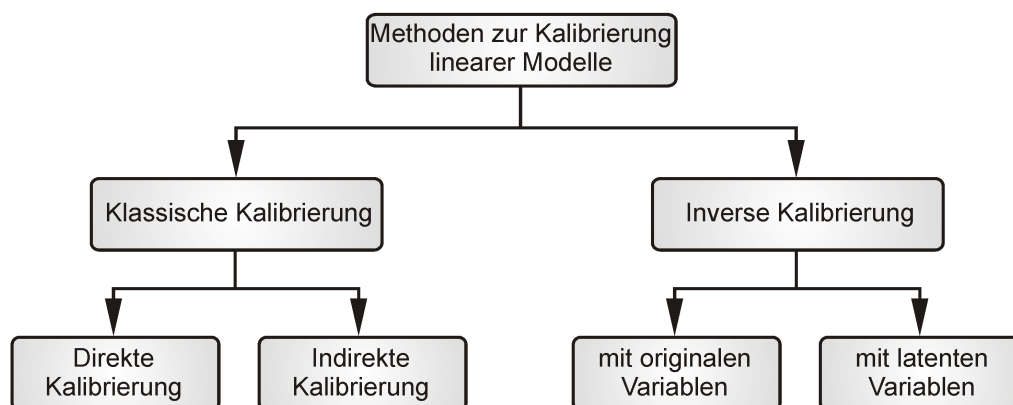


Bild 5.12: Methoden der linearen Kalibrierung [Danzer *et al.* (2004)]

Grundsätzlich lassen sich die Methoden in klassische und inverse Kalibrierung unterteilen. Die Auswahl des geeigneten Verfahrens richtet sich dabei nach der Fehlerbehaftung der Messgrößen. Der Unterschied beider Verfahren liegt dabei im funktionalen Zusammenhang von der abhängigen Größe, hier das BioView[®]-Signal eines Wellenlängenpaares, und der sie verursachenden Konzentration.

Bei chemischen Messungen, bei denen die Konzentration der Komponente z.B. über Standards mit sehr hoher Genauigkeit vorgegeben ist und der Fehler der Messung größtenteils durch das Messgerät bestimmt wird, kommt die klassische Kalibrierung zum Einsatz. Die abhängige Größe x_k ,

$$x_k = C_i \cdot s + e_x \quad (5.17)$$

stellt dabei in diesem Fall den Messvektor einer Wellenlängenkombination k des BioView[®]-Messgerätes über die Zeit dar. Die Matrix C_i ist in der ersten Spalte mit 1 besetzt, um den Ordinatenabschnitt zu berücksichtigen, und enthält in der zweiten Spalte die zu modellierenden Konzentrationswerte. Diese sind abhängig von den Elementen des Koeffizientenvektors s , der den Ordinatenabschnitt sowie die Steigung der proportionalen Verknüpfung enthalten. Der Vektor e_x enthält die Modellfehler zum jeweiligen Messzeitpunkt.

Für die Berechnung des Koeffizientenvektors s ,

$$s = (C_i^T \cdot C_i)^{-1} \cdot C_i^T \cdot x_k \quad (5.18)$$

muss damit die Pseudoinverse der Konzentrationsmatrix gebildet werden.

Für die multivariate Kalibrierung mit mehreren Eingangs- und/oder Ausgangsgrößen kann die Koeffizientenmatrix S ,

$$S = (C^T \cdot C)^{-1} \cdot C^T \cdot X \quad (5.19)$$

in Anlehnung an Gl. (5.18) allgemein formuliert werden.

Die indirekte Kalibrierung wird eingesetzt, wenn keine klassische Kalibration des Messsystems mit Standards möglich ist. Dieses ist im Falle des 2D-Sensors gegeben, da die komplexe Messmatrix aus zellinternen Fluorophoren nicht quantitativ nachgebildet werden kann. Außerdem sind die Konzentrationen der Fluorophore nicht bekannt.

Da der Fehler des BioView[®]-Systems im Vergleich zu den am Prozess eingesetzten Offline- und Atline-Analysen als geringer eingeschätzt werden kann, wird die zu modellierende Konzentration c_i ,

$$c_i = X_k \cdot b + e_y \quad (5.20)$$

als Funktion des Wellenlängenpaares x_k beschrieben.

Die unbekannte Koeffizientenmatrix B ,

$$B = (X^T \cdot X)^{-1} \cdot X^T \cdot C \quad (5.21)$$

wird dann für den allgemeinen Fall der multivariaten Kalibrierung über die pseudoinverse Matrix der BioView[®]-Spektren berechnet, worin der gravierende Unterschied zur klassischen Kalibrierung in Gl. (5.19) liegt.

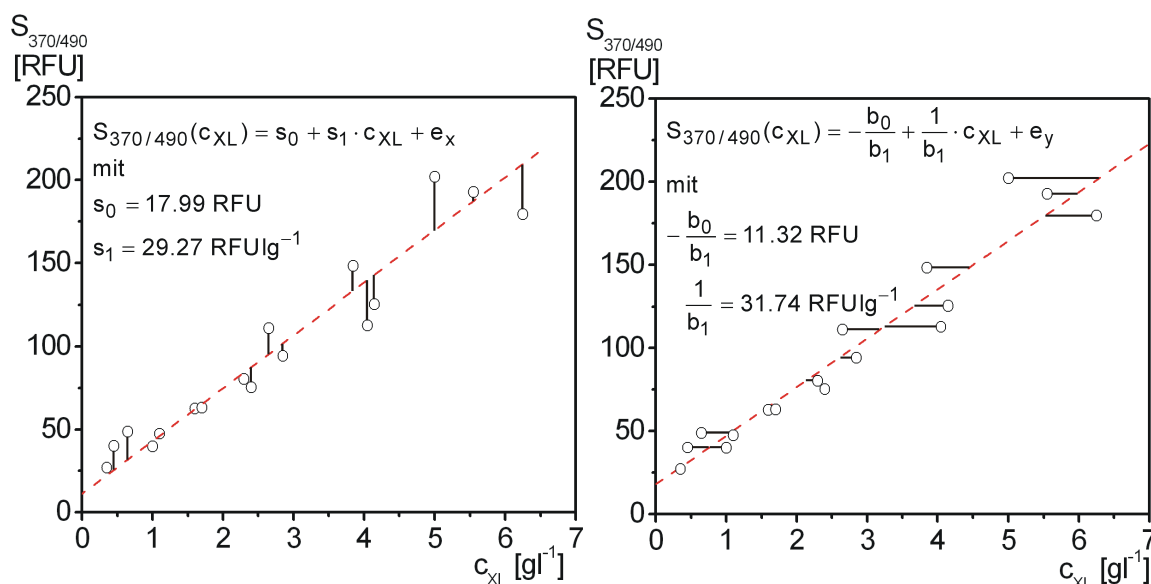


Bild 5.13: Regressionsgeraden für die klassische (links) und die inverse Kalibrierung (rechts)

Den in den Gln. (5.18) bzw. (5.19) und (5.21) vorgestellten Ansätzen zur Kalibrierung der Modelle liegen unterschiedliche Ansätze zur Auffindung der Koeffizientenmatrix S bzw. B bezüglich der Fehlerverteilung zugrunde. Wie in Bild 5.13 dargestellt, ergeben sich dadurch unterschiedliche Werte für die Koeffizienten der Modelle.

Nach der Korrelationsanalyse und der Bestimmung des Regressionskoeffizientenvektors b erfolgt gemäß Bild 5.10 eine Beurteilung der Modellgüte über die Wurzel der mittleren Fehlerquadrate (Gl. (5.7)). Zur Unterscheidung der Operationen wird der Fehler der Kalibrierung als RMSEC^9 und der Fehler des Validierungsdatensatzes als RMSEP^{10} bezeichnet.

Ist der Kalibrierungsfehler kleiner als der zuletzt gespeicherte, werden die aktuellen Modellkoeffizienten, die zugehörige Wellenlängenkombination sowie der Fehler gespeichert. Anschließend wird die entsprechende Prozessgröße unter Verwendung der zugehörigen Validierungsdaten berechnet und der RMSEP bestimmt und gespeichert. Letzterer stellt ein Maß für die Vorhersagegenauigkeit der erstellten Modelle dar.

5.4.3 Ergebnisse der univariaten Modellkalibrierung

Exemplarisch ist in Bild 5.14 das Messsignal einer Wellenlängenkombination des BioView[®]-Sensors ($S_{370/490}$) über die Zelldichte aufgetragen. Offensichtlich sind die Daten im Bereich bis 12 g l^{-1} hoch korreliert. Für größere Zelldichten nimmt die Streuung deutlich zu. Dies zeigt sich auch für die meisten anderen Wellenlängenkombinationen (Daten nicht dargestellt).

Einer der Gründe hierfür ist die Induktion, die bei den meisten Kultivierungen bei einer Zelldichte von 12 g l^{-1} durchgeführt wurde. Die anschließend stark ansteigende Fluoreszenz des T-Sapphire-GFP überlagert und beeinflusst dabei die angrenzenden Fluoreszenzbereiche.

Bild 5.14 macht deutlich, dass eine lineare Modellierung der Zelldichte im unteren Bereich möglich ist. Für höhere Zelldichten zeigen sich jedoch bei allen Wellenlängenkombinationen,

⁹RMSEC – Root Mean Square Error of Calibration

¹⁰RMSEP – Root Mean Square Error of Prediction

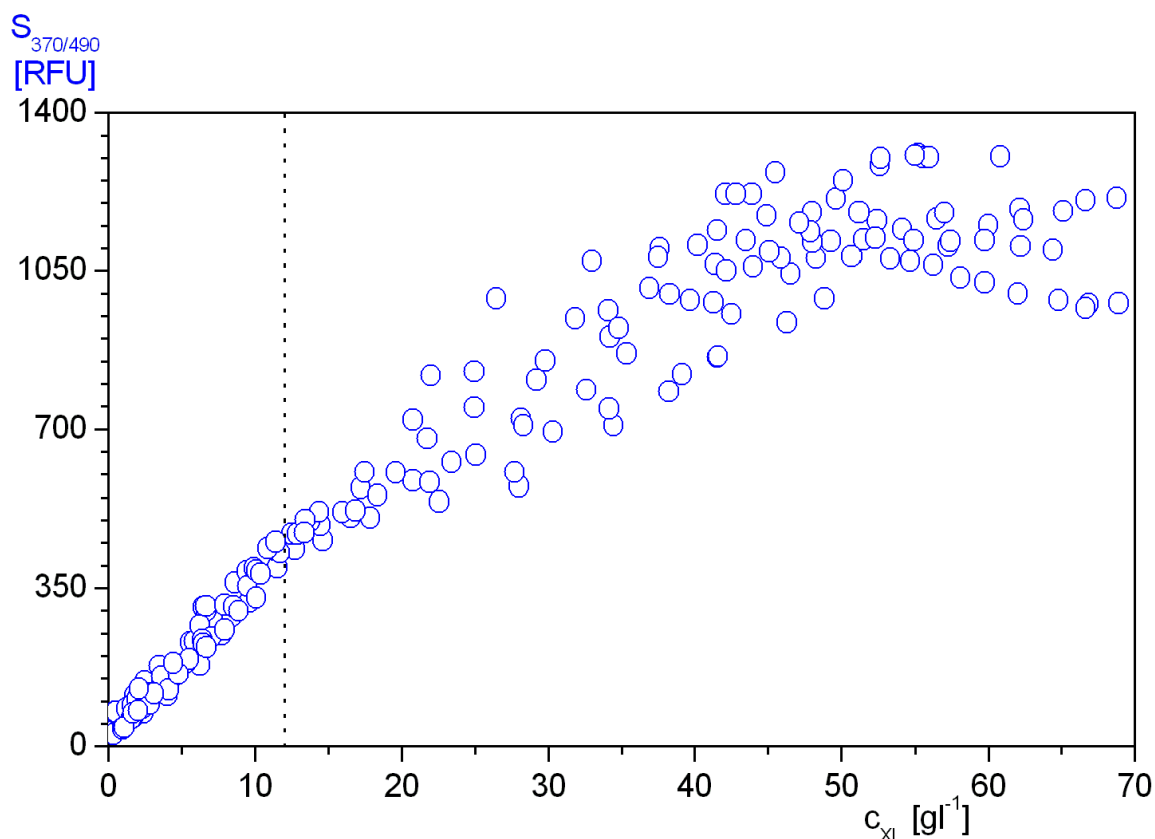


Bild 5.14: Verlauf des Fluoreszenzsignals $S_{370/490}$ (FAD-Bereich) über die Zelldichte c_{XL}

neben der wesentlich höheren Streuung der Messwerte, auch Sättigungseffekte, die mit linearen Modellen nicht zu lösen sind. Da diese Effekte teilweise auch schon während der Fed-Batch Phase vor der Induktion auftreten, werden die Spektren im Folgenden abschnittsweise für die einzelnen Prozessphasen Batch, Fed-Batch und Produktion analysiert.

Während der unlimitierten Batch-Phase zeigen nahezu alle Fluoreszenzsignale einen exponentiellen Verlauf. Wie verschiedene Autoren zeigen konnten, korreliert die Konzentration der Fluorophore NAD(P)H sowie des Pyridoxalphosphates während des unlimitierten Batches mit der Zelldichte sowie der Glukosekonzentration [Marose (1999)][Clementschtisch *et al.* (2005)].

Die Korrelationskoeffizienten der einzelnen BioView[®]-Signale mit der Zelldichte und der Glukosekonzentration für die Batch-Phase sind in Bild 5.15 dargestellt. Die relevanten Fluoreszenzbereiche sind dabei nahezu identisch.

Da Glukose keinerlei fluoreszierende Eigenschaften aufweist, kann die Bestimmung nur indirekt aus den an die Stoffwechselwege von *Escherichia coli* gekoppelten Fluorophoren erfolgen. Hohe Korrelationen ergeben sich in mehreren Bereichen. Der Zelldichteverlauf und die Glukosekonzentration werden dabei am Besten durch die Fluoreszenz im Bereich des Pyridoxalphosphats ($S_{410/470}$) wiedergegeben. Es ergeben sich Korrelationskoeffizienten für die Zelldichte von 0.96 bis 0.97 und 0.94 bis 0.95 für die Glukosekonzentration.

Auffällig ist auch die gute Korrelation der Glukosekonzentration und der Zelldichte mit der Fluoreszenz im Zielproteinbereich. Da der Gehalt des rekombinanten Zielproteins in den Zellen vor der Induktion annähernd konstant ist, korreliert die Fluoreszenz im GFP-Bereich ($S_{390/510}$) mit den Zielvariablen c_{XL} und c_{S1M} .

Da sich im Verlauf der Fed-Batch Phase, insbesondere wenn die Induktion bei höheren Zelldichten erfolgte, Sättigungsphänomene in den Spektren zeigen, wurde für die Anpassung der Zelldichte ein exponentieller Ansatz gewählt. Durch Logarithmieren werden die Zieldaten linearisiert und es kann weiterhin die Korrelationsanalyse sowie die lineare Regression angewendet werden.

Auch während der Fed-Batch-Phase weisen die Bereiche des Pyridoxalphosphates sowie die Streulichtmessungen die höchste Korrelation mit der Zelldichte auf. Die in Bild 5.16 dargestellte Anpassung führte dabei zu den besten Ergebnissen und erfolgte über die Streulichtmessung ExND/em590.

Problematisch stellen sich die Anpassungen während der Produktionsphase dar. Neben der starken Fluoreszenz des Zielproteins, die auch angrenzende Fluoreszenzbereiche beeinflusst, kommen die in Abschnitt 3.3.4 beschriebenen Effekte wie z.B. der Inner-Filter Effekt und der Kaskadeneffekt immer mehr zum Tragen. Dadurch kommt es zu einer immer größer werdenden Streuung der Messdaten der unterschiedlichen Kultivierungen und die Präzision der Vorhersagen nimmt ab.

Im Fall des Inner-Filter Effektes kann es auch zu einer Abnahme der Fluoreszenzintensität durch eine zu hohe Konzentration an Fluorophoren kommen. Dieser Effekt macht die Erstellung von verlässlichen Modellen nahezu unmöglich. Für eine logarithmische Anpassung ergeben sich lediglich im Streulichtbereich noch Korrelationskoeffizienten von 0.9. Die beste Anpassung ergibt sich, wie in Tabelle 5.4 aufgeführt, über die gleiche Streulichtmessung wie im Fed-Batch.

Am gravierendsten wirken sich die oben genannten Effekte auf die Anpassung der GFP-Konzentration über die Fluoreszenzmessung aus. Die Sättigungsphänomene für höhere GFP-Konzentrationen in Fed-Batch Prozessen von *E. coli* Kulturen wurden auch von Reischer *et al.* (2004) beschrieben.

Die in Bild 5.17 dargestellten Korrelationskoeffizienten der Fluoreszenzmessung mit der Zielproteinkonzentration während der Produktionsphase machen deutlich, dass die höchsten Werte an der Flanke des Emissionsmaximums bei Wellenlängen von 550 - 590 nm zu finden sind. Der höchste Korrelationskoeffizient ergibt sich dabei mit 0.9 für die Wellenlängenkombi-

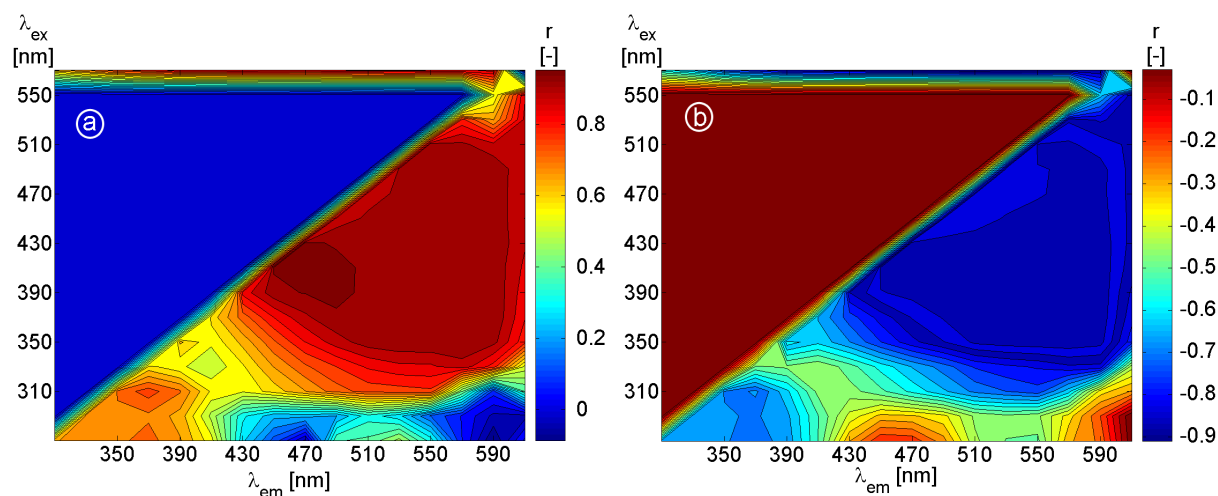


Bild 5.15: Korrelationskoeffizienten für die Zelldichte (a) sowie die Glukosekonzentration (b) während der Batch-Phase

nation 390/590 nm, wobei ein linearer Zusammenhang zwischen GFP-Konzentration und Fluoreszenzmessung verwendet wurde.

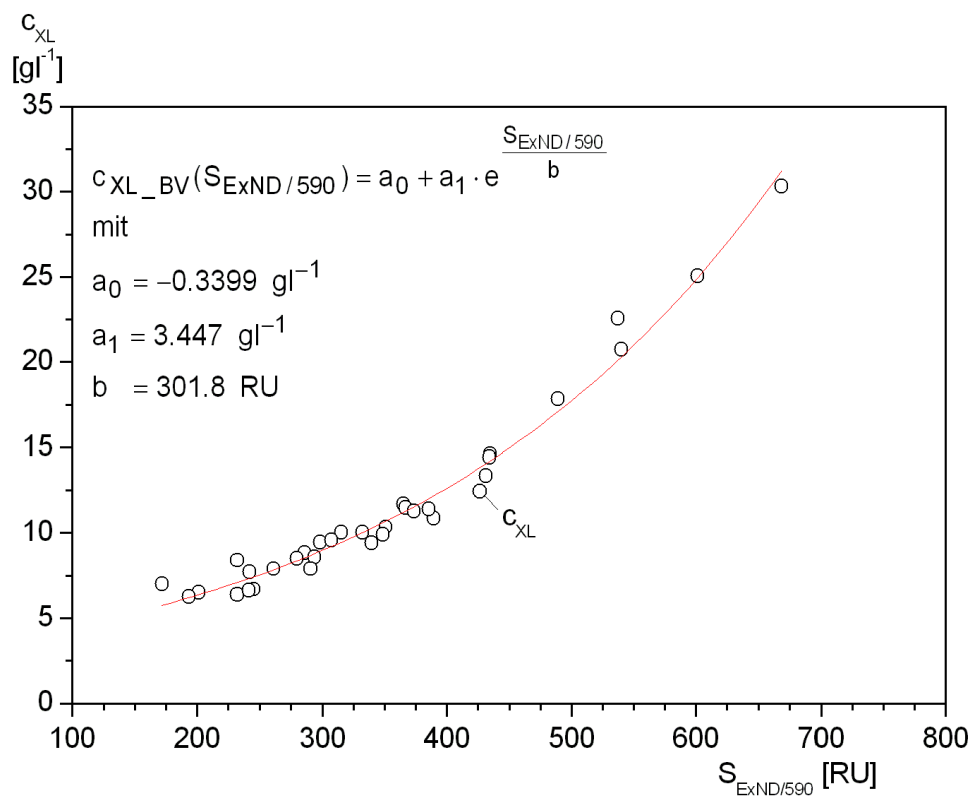


Bild 5.16: Korrelation der Zelldichte mit dem Neutraldichtefilter $S_{ExND/590}$

In Bild 5.18 sind die über die Korrelationsanalyse abgeleiteten univariaten Approximationen im Vergleich mit Offline- und Atline-Messungen für einen Validierungsdatensatz dargestellt. Der Verlauf der rekonstruierten Glukosekonzentration während der Batch-Phase liegt durchgehend unter den atline bestimmten Werten. Insbesondere während der ersten vier Stunden spiegelt die modellierte Größe jedoch einen wesentlich plausibleren Verlauf wieder als die Atline-Messung. Der Zeitpunkt des vollständigen Glukoseverbrauches, welcher auch für die Prozessautomatisierung relevant ist, wird durch die Approximation sehr gut wiedergegeben.

Die zeitlichen Verläufe der Zelldichte und der Zielproteinkonzentration geben den Trend der offline bestimmten Größen über alle Prozessphasen sehr gut wieder. Lediglich am Ende der Produktionsphase weicht die rekonstruierte Zelldichte von der Biotrockenmasse gravierend ab. Ob dieses nur durch Sättigungs- und Innerfiltereffekte oder aber auch durch eine Abnahme der zellinternen Fluorophore bedingt ist, konnte im Rahmen dieser Arbeit nicht geklärt werden.

Auch für den zweiten Validierungsdatensatz, der in Bild 5.19 dargestellt ist, zeigen die simulierten Größen eine zufriedenstellende Übereinstimmung mit den Messdaten. Der Zeitpunkt des vollständigen Glukoseverbrauches wird in diesem Fall jedoch eine halbe Stunde zu früh durch das verwendete Modell angezeigt. Die Zelldichte wird, wie auch beim Ersten Validierungsdatensatz, während der Batch und der Fed-Batch Phase sehr gut wiedergegeben. Lediglich während der Produktionsphase ist eine deutlich Abweichung von den Offline-Messdaten zu erkennen. Der Verlauf der Zielproteinkonzentration wird gut wiedergegeben, jedoch kommt es beim Übergang zwischen Fed-Batch und Produktionsphase, durch die Verwendung unterschiedlicher Modellparameter, zu einer Abnahme der modellierten Proteinkonzentration.

Die Fehler der Modelle, die verwendeten Wellenlängenkombinationen sowie die Art der

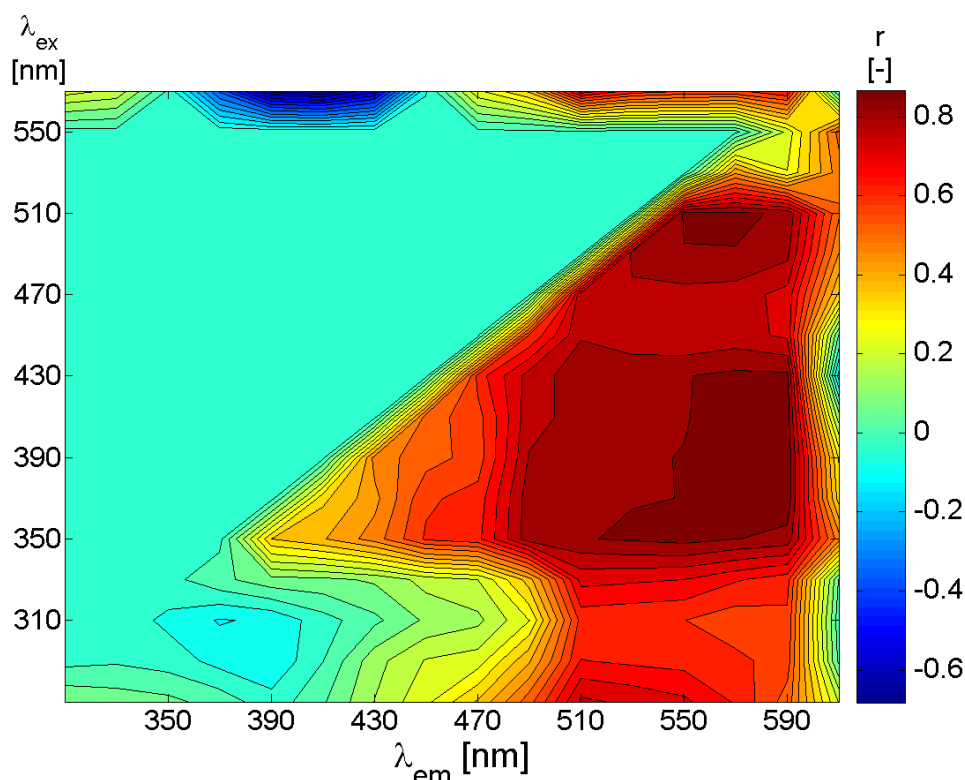


Bild 5.17: Korrelationskoeffizienten für die Zielproteinkonzentration während der Produktionsphase

Anpassung sind in Tabelle 5.4 aufgeführt. Die entsprechenden Modellkoeffizienten befinden sich im Anhang. Bemerkenswert ist die Tatsache, dass in mehreren Fällen der Validierungsfehler geringer ist als der Trainingsfehler. Durch die hohe Variabilität in den Trainingsdatensätzen, von denen drei weitere exemplarisch in den Bild 5.20 bis 5.22 dargestellt sind, kommt es zu großen Streuungen hinsichtlich Zellvitalität, Prozessführung und Induktionszeitpunkt in den Trainingsdatensätzen, die zu relativ hohen Fehlern führen. Die geringeren Validierungsfehler, die sich aus den drei verwendeten Validierungsdatensätzen ergeben, sprechen daher für eine gute Generalisierungsfähigkeit der gefundenen Ansätze.

Tabelle 5.4: Fehler der erstellten univariaten Modelle

| Zielgröße | Phase | Anpassung | Eingangsgröße | RMSEC | | RMSEP | |
|------------------|---------------------|--------------|-----------------------|-----------------------|------|-----------------------|------|
| | | | | [g l ⁻¹] | [%] | [g l ⁻¹] | [%] |
| c _{XL} | Batch | linear | S _{410/490} | 0.56 | 8.35 | 0.44 | 6.09 |
| | Fed-Batch | exponentiell | S _{ExND/590} | 0.70 | 3.13 | 0.74 | 12.7 |
| | Produktion | exponentiell | S _{ExND/590} | 6.85 | 11.7 | 4.60 | 10.8 |
| c _{S1M} | Batch | linear | S _{430/470} | 0.83 | 7.15 | 1.26 | 11.4 |
| c _{P1L} | Batch/ Fed-Batch | exponentiell | S _{390/590} | 2.42·10 ⁻³ | 7.32 | 1.21·10 ⁻³ | 4.03 |
| | Produktion | linear | S _{390/570} | 63.9·10 ⁻³ | 9.24 | 54.6·10 ⁻³ | 9.21 |

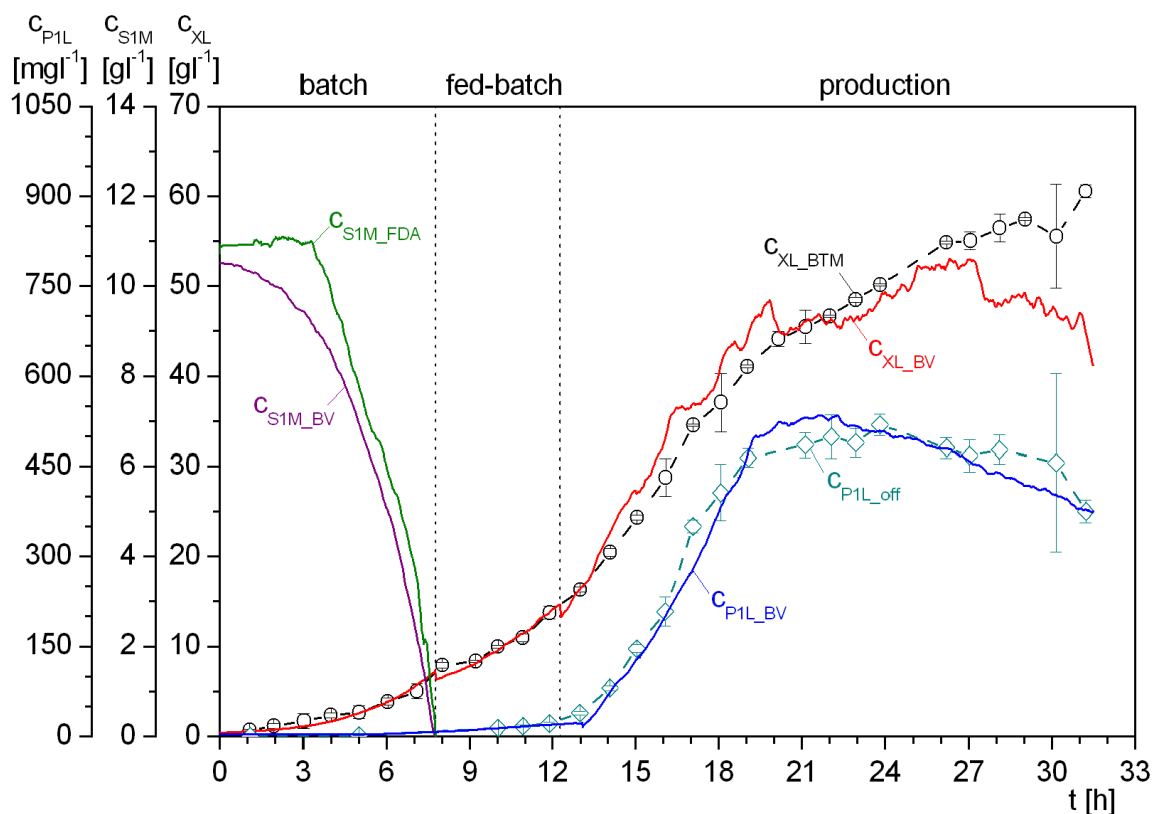


Bild 5.18: Ergebnis der erstellten univariaten Modelle für den Validierungsdatensatz CK0707

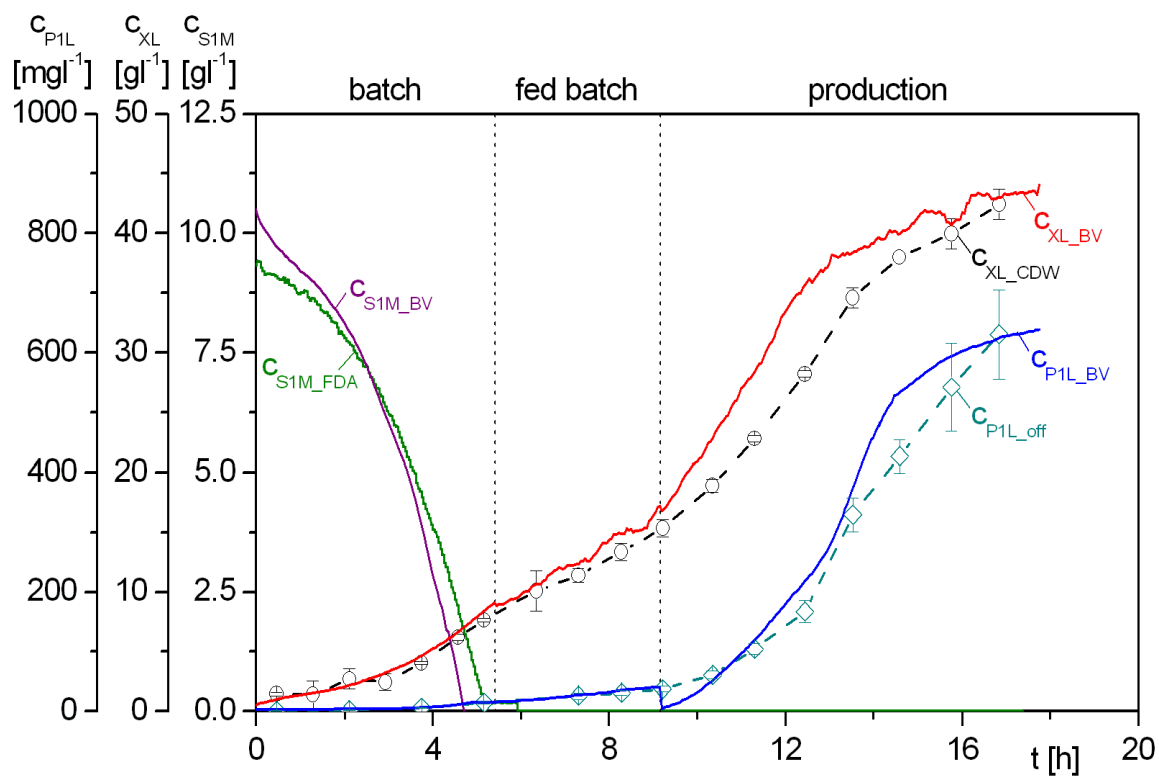


Bild 5.19: Ergebnis der erstellten univariaten Modelle für den Validierungsdatensatz CK2208

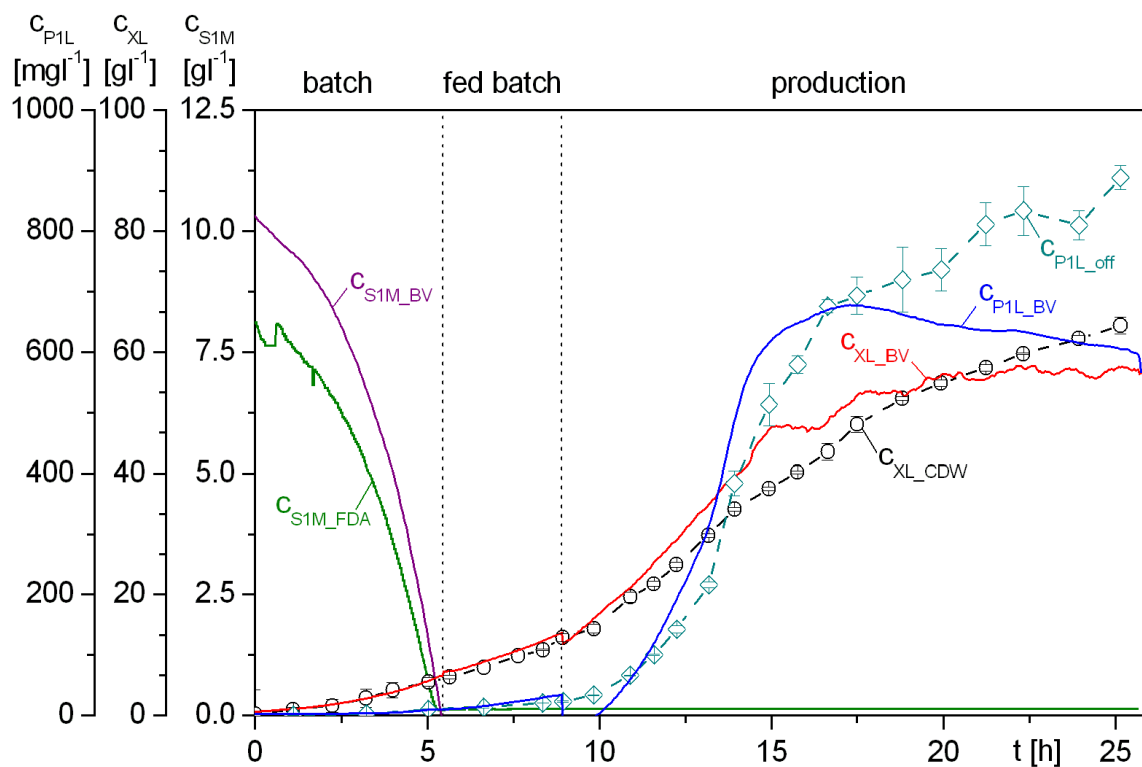


Bild 5.20: Ergebnis der erstellten univariaten Modelle für den Trainingsdatensatz CK4407

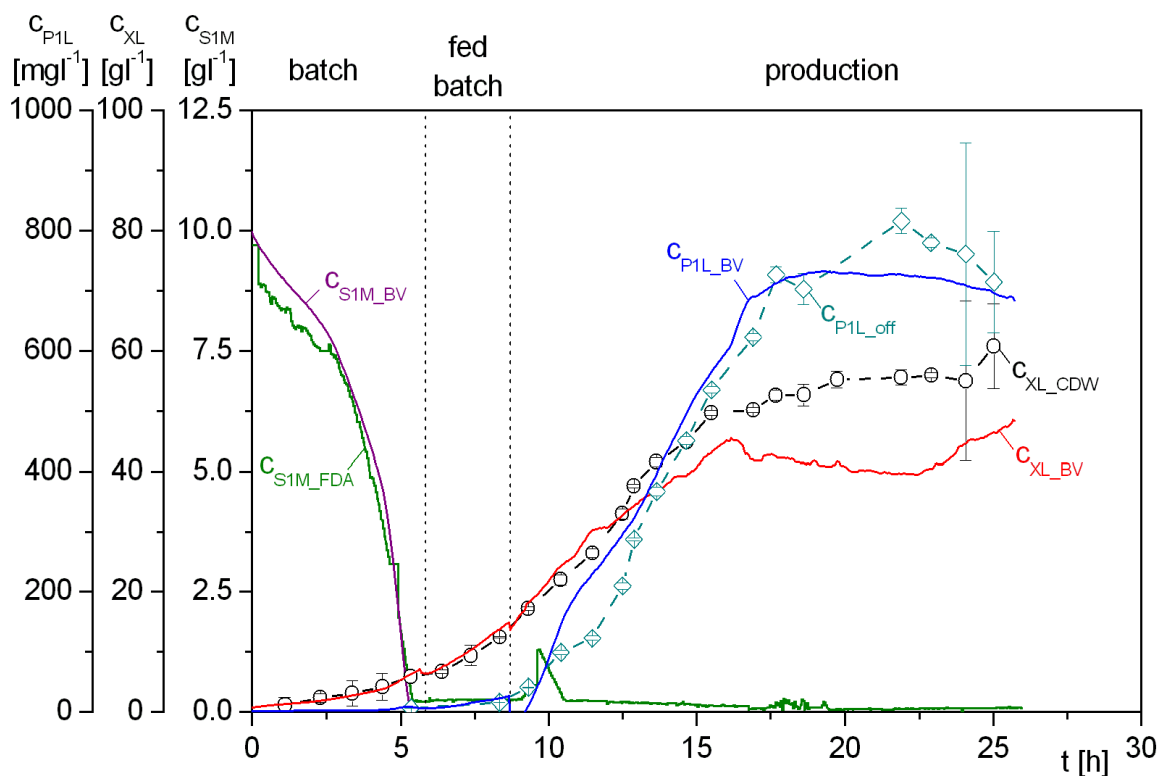


Bild 5.21: Ergebnis der erstellten univariaten Modelle für den Trainingsdatensatz CK4406

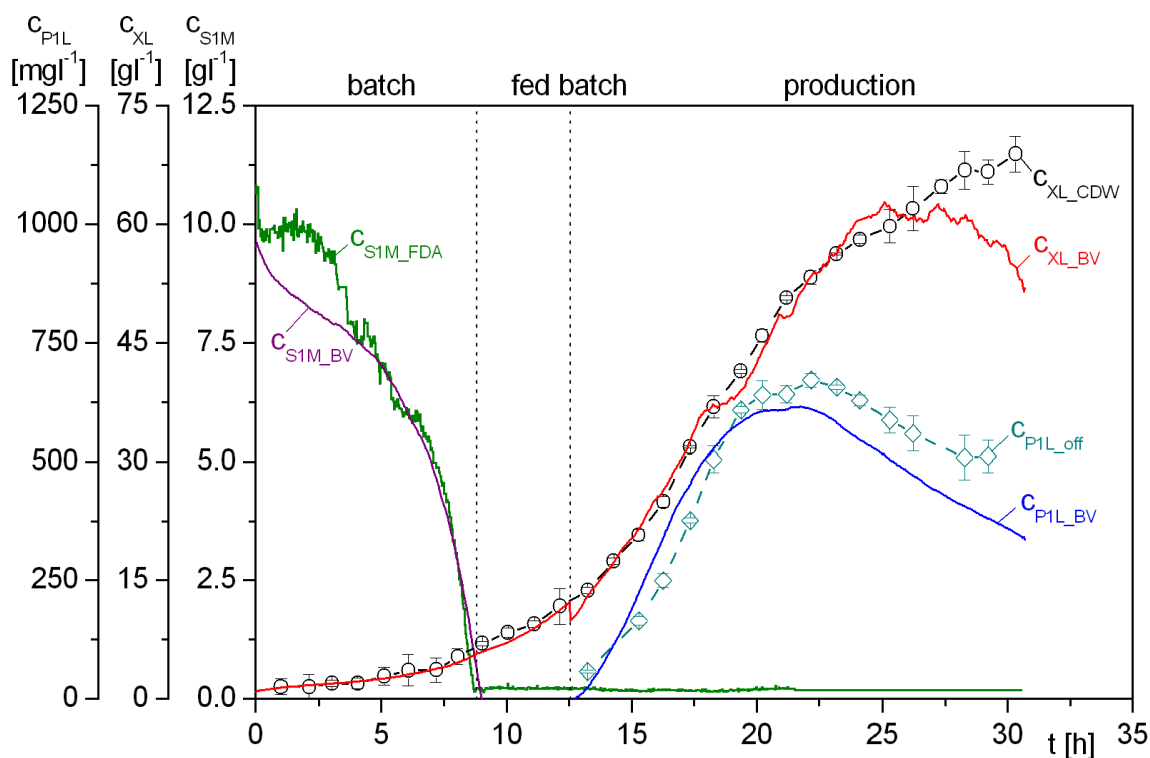


Bild 5.22: Ergebnis der erstellten univariaten Modelle für den Trainingsdatensatz CK0607

Mit den relativ einfachen Ansätzen konnten bereits gute Anpassungen gefunden werden. Über multivariate Modelle, unter Verwendung mehrerer Wellenlängenkombinationen zur Approximation der Prozessgrößen, kann die Vorhersagegenauigkeit und Zuverlässigkeit im Allgemeinen noch verbessert werden.

Die große Anzahl an Wellenlängenkombinationen erschwert dabei die Auswahl. Außerdem wird durch die linearen Abhängigkeiten der Messdaten die Präzision der Matrixinversion, die gemäß Gl. (5.21) zur Kalibrierung der Modelle notwendig ist, verringert. Im Folgenden werden deshalb zusätzlich zur Modellkalibrierung mit den Original-BioView[®]-Daten auch geeignete Methoden zur Datenreduktion vorgestellt, auf die Messdaten angewendet und zur Vorhersage der zu bestimmenden Prozessgrößen eingesetzt.

5.5 Kalibrierung linearer multivariater Modelle

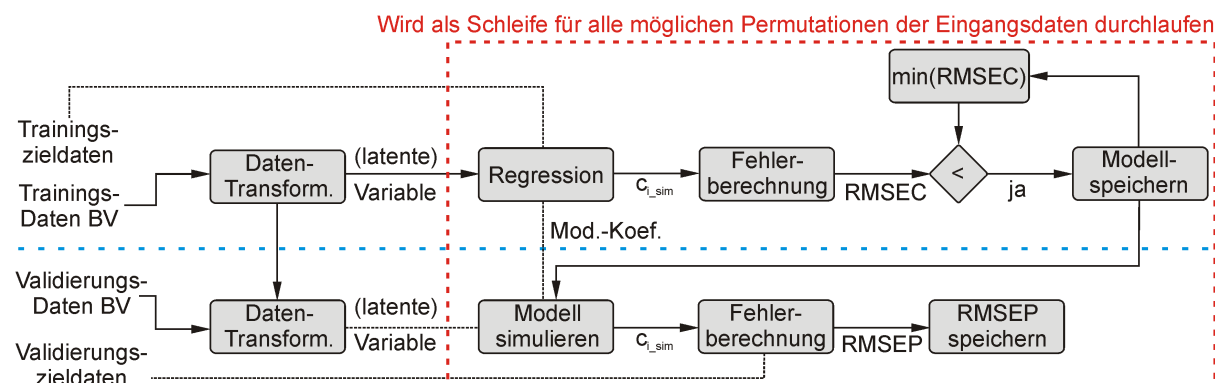


Bild 5.23: Prinzipielles Vorgehen bei der Kalibrierung multivariater Modelle

In Bild 5.23 ist das in dieser Arbeit verwendete prinzipielle Vorgehen bei der Kalibrierung multivariater Modelle dargestellt. Das Verfahren unterscheidet sich dabei von dem in Abschnitt 5.4.2 vorgestellten Ablauf in zwei wesentlichen Punkten.

Anstelle der Korrelationsanalyse, die bei der univariaten Modellkalibrierung der Regression vorgeschaltet ist, tritt in diesem Fall eine Transformation der BioView[®]-Signale zur Datenreduktion der 150 Fluoreszenzspektren.

Insgesamt wurden drei unterschiedliche Verfahren angewendet, um aus den Eingangsdaten der BioView[®]-Messungen die Prozessgrößen Zelldichte, Glukose- sowie Zielproteinkonzentration zu rekonstruieren.

Im ersten Fall wurde die Kalibrierung der Modelle mit den tatsächlichen Messsignalen durchgeführt. In diesem Fall entfiel die Transformation, die der Iterationsschleife vorgeschaltet ist. Es wurden maximal drei Wellenlängenkombinationen zur Kalibrierung verwendet, um die Konditionierung der zu invertierenden Messmatrix im Rahmen zu halten. Eine höhere Anzahl an Wellenlängenkombinationen führte zu einem starken Rauschanteil in den modellierten Prozessgrößen.

Im zweiten Fall wurden die BioView[®]-Daten zunächst einer Hauptkomponentenanalyse unterzogen, die in Abschnitt 5.5.1 näher erläutert wird. Diese legt auf der Basis einer Varianzanalyse ein neues orthogonales Koordinatensystem durch die spektralen Daten des BioView[®]-Messgerätes. Durch diese Transformation kann mit einigen wenigen Vektoren ein Großteil der in den Messdaten enthaltenen Informationen wiedergegeben werden. Es konnten dadurch 99,9 % der in den 150 zur Verfügung stehenden BioView[®]-Spektren enthaltenen Information durch vier Hauptkomponenten beschrieben und dadurch wesentlich kürzere Kalibrierzeiten erreicht werden.

Im dritten Fall wurde eine *Independent Component Analysis* zur Merkmalsextraktion eingesetzt. Diese wird häufig in der Bildbearbeitung zur Reduktion von komplexen Daten verwendet und ist in Abschnitt 5.5.2 ausführlich beschrieben. Die ICA basiert, wie auch die PCA, auf der linearen Transformation der Daten in ein neues Koordinatensystem. Dabei wird jedoch nicht nur die Varianzanalyse als Grundlage für die Berechnung des neuen Koordinatensystems angewendet sondern auch Statistiken höherer Ordnungen, mit dem Ziel die Daten statistisch so unabhängig wie möglich zu machen.

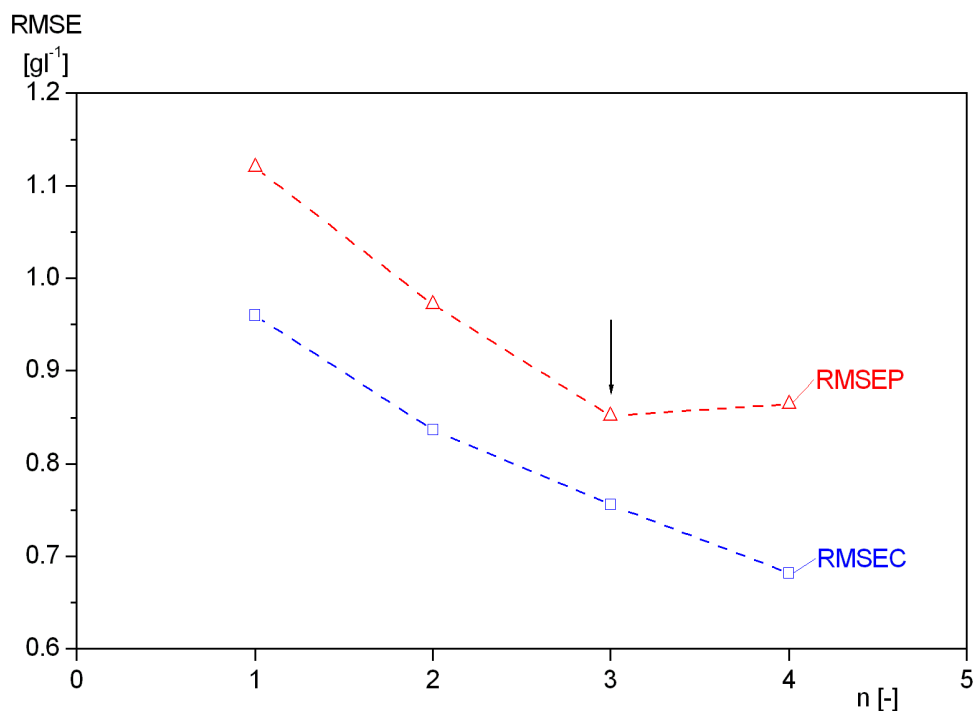


Bild 5.24: Verlauf von RMSEC und RMSEP bei der multivariaten Kalibrierung als Funktion der gewählten Anzahl von Eingangsvektoren

Der zweite wesentliche Unterschied bei der Kalibrierung multivariater Modelle ist die Einführung einer Iterationschleife (gestrichelte Linie in Bild 5.23). Diese wird für alle möglichen Permutationen der Eingangsdaten durchlaufen, um die optimale Kombination an Eingangsvektoren aufzufinden. Dabei werden zunächst zwei Eingangsvektoren zur Modellkalibrierung verwendet und alle möglichen Kombinationen der zu Verfügung stehenden Eingangsvektoren erzeugt.

Im Anschluß wird der Modellfehler des erzeugten Modells für die Trainingsdatensätze berechnet (Gl. (5.7)). Ist dieser kleiner als der zuletzt gespeicherte, werden die aktuellen Modellparameter abgespeichert und der Fehler des Validierungsdatensatzes berechnet.

Ist dieses Prozedere für alle möglichen Kombinationen der zur Verfügung stehenden Eingangsvektoren abgeschlossen, wird die Anzahl der Eingangsvektoren um eins erhöht. Anschließend wird die Iterationsschleife erneut für alle möglichen Permutationen durchlaufen und das oben beschriebene Verfahren wiederum angewendet.

Da der Fehler der Kalibrierung (RMSEC) durch die Vergrößerung der Anzahl an Eingangsvektoren (Eingangsdatenraum) kleiner wird muss ein weiteres Kriterium verwendet werden, welches ein *overfitting* der zu erstellenden Modelle verhindert.

Zur Bestimmung der Verallgemeinerungsfähigkeit wird deshalb der Fehler der Modellvalidierung (RMSEP) herangezogen. Wie in Bild 5.24 dargestellt, werden zunächst sowohl der RMSEC als auch der RMSEP kleiner, je mehr Komponenten zur Modellierung verwendet werden. Sobald der Wert des RMSEP jedoch beginnt wieder größer zu werden, wird die Kalibrierung abgebrochen und das Modell mit dem kleinsten Fehler für die Validierungsdaten ausgewählt.

5.5.1 Hauptkomponentenanalyse

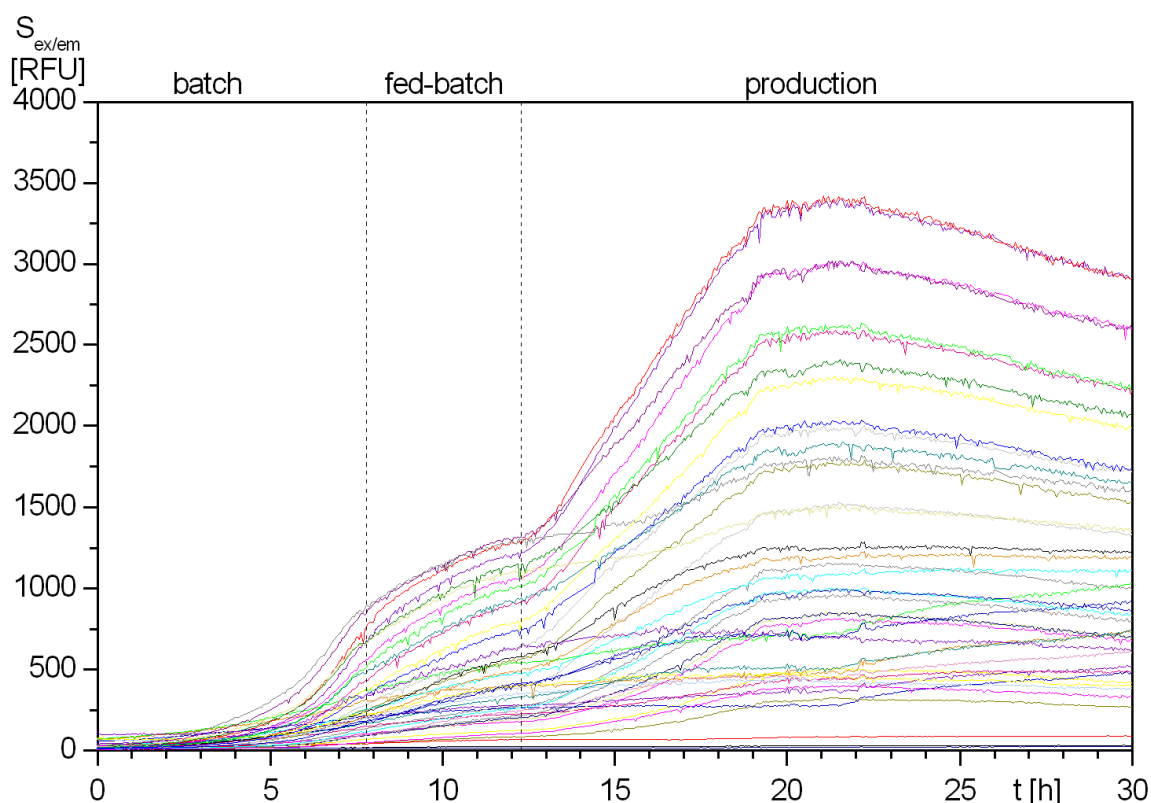


Bild 5.25: Mehrere Wellenlängenpaare des BioView[®]-Spektrums bei einer Kultivierung

Die Hauptkomponentenanalyse gehört zu den faktoriellen Methoden des unüberwachten Lernens. Sie stellt eine der am weit verbreitetsten Methoden zur Reduktion multivariater Daten dar, die ein hohes Maß an Kolinearitäten aufweisen. Die Originaldaten werden dabei in ein neues, relevanteres Koordinatensystem transformiert, mit dem Ziel untereinander unkorrelierte Variablen (Hauptkomponenten¹¹) zu erhalten. Durch die Beschränkung auf die ersten PCs wird eine Datenreduktion erreicht, die tatsächliche Information extrahiert und das Signal-Rausch Verhältnis verbessert [Danzer *et al.* (2001)]. Zusätzlich wird die für die Modellkalibrierung benötigte Invertierbarkeit $X^T X$ gewährleistet.

Voraussetzung für die erfolgreiche Anwendung der PCA ist ein gewisses Maß an Redundanz in den Ausgangsdaten. Bei Betrachtung der in Bild 5.25 dargestellten Fluoreszenzdaten des eingesetzten BioView[®]-Spektrometers eines Kultivierungsverlaufes wird deutlich, dass diese Voraussetzung eindeutig erfüllt ist.

Zur Veranschaulichung der PCA werden oft grafische Darstellungen wie in Bild 5.26 genutzt. Der Einfachheit halber sind hier nur zweidimensionale Daten dargestellt. Trägt man diese Daten gegeneinander auf, so entsprechen die Achsen x_1 und x_2 des Koordinatensystems den Wertebereichen der beiden Variablen.

¹¹auch Faktoren, Scores, PCs oder allgemein latente Variable genannt

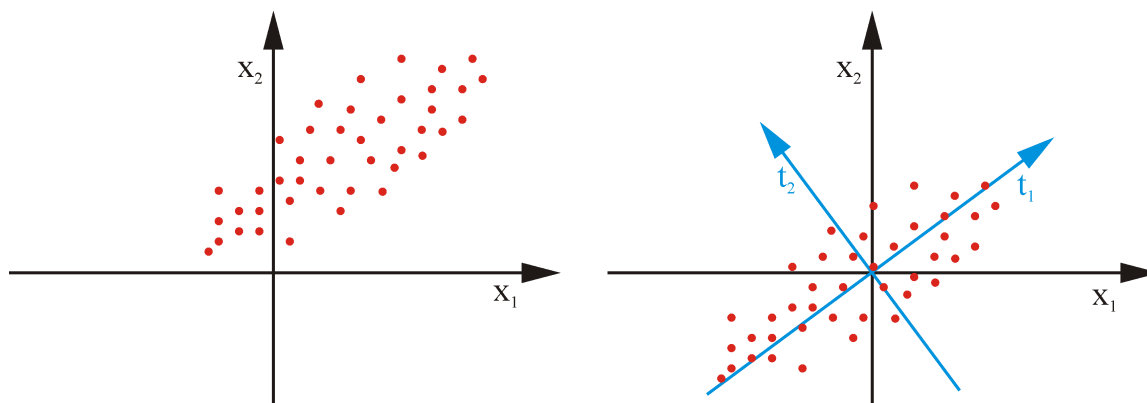


Bild 5.26: Geometrische Interpretation der Hauptkomponentenanalyse

Die Originaldaten werden gemäß Abschnitt 5.4.1 über eine Zentrierung in den Ursprung des Koordinatensystems verschoben. Anschließend wird das Koordinatensystem so gedreht, dass die erste Hauptachse t_1 in Richtung der größten Streuung der Daten zeigt, d.h. dass die erste Hauptachse bzw. Hauptkomponente den größten Wertebereich aufweist. Dieses entspricht dem gleichen Prinzip, das auch bei der linearen Regression zugrunde liegt.

Die Orthogonalität wird dadurch gewährleistet, dass die zweite Hauptachse t_2 senkrecht auf der ersten Hauptachse steht. Sie wird so angeordnet, dass sie den verbleibenden maximalen Varianzanteil der Daten erklärt. Nach diesem Prinzip werden (bei höherdimensionalen Daten) alle weiteren Hauptkomponenten konstruiert.

Die zentrierte $m \times n$ BioView[®]-Messmatrix X aus Gl. (5.8),

$$X = T \cdot P^T = t_1 \cdot p_1^T + t_2 \cdot p_2^T + t_3 \cdot p_3^T + \dots + t_n \cdot p_n^T, \quad (5.22)$$

wird dabei durch das Produkt aus der $m \times n$ Faktorscorematrix T und der transponierten $n \times n$ Ladungsmatrix P , die sich aus den Ladungen¹² der einzelnen Hauptkomponenten zusammensetzt, dargestellt. Die Spaltenlänge m entspricht der Anzahl der Messungen bzw. Objekte und die Zeilenlänge n der Anzahl an Wellenlängenkombinationen oder allgemein der Quantität der Merkmale.

Die am häufigsten verwendeten Algorithmen zur Berechnung der transponierten Ladungsmatrix P^T , die auch für die lineare Regression eingesetzt werden, sind NIPALS¹³ [Wold *et al.* (1987)] oder die Singulärwertzerlegung¹⁴ [Golub and Loan (1983)].

Die in dieser Arbeit eingesetzte Singulärwertzerlegung stellt ein ähnliches Verfahren wie die Eigenwertzerlegung dar, sie kann aber auch auf nicht quadratische Matrizen angewendet werden kann [Press *et al.* (1986)]. Dabei wird die Matrix X ,

$$X = U \cdot S \cdot V^T, \quad (5.23)$$

in drei kleinere Matrizen zerlegt, wobei S eine Diagonalmatrix darstellt.

¹²auch Loadings genannt

¹³NIPALS – *Nonlinear Iterative Partial Least Square*

¹⁴SVD – *Single Value Decomposition*

Zur Berechnung der drei unbekanntenen Matrizen U , S und V wird zunächst eine Gauß'sche Transformation von Gl. (5.23),

$$X^T \cdot X = (U \cdot S \cdot V^T)^T U \cdot S \cdot V^T = V \cdot S^T \cdot U^T \cdot U \cdot S \cdot V^T, \quad (5.24)$$

durchgeführt,

Da U eine orthogonale Matrix ist, welche die auf eine Länge von eins normierten Faktoren aus Gl. (5.22) enthält, ergibt das Produkt $U^T U$ die Einheitsmatrix und kann im Folgenden weggelassen werden. Die Diagonalmatrix S kann dann über die Matrix D ,

$$D = S^T \cdot S = S^2, \quad (5.25)$$

beschrieben werden. Damit vereinfacht sich Gl. (5.24) zu

$$X^T \cdot X = V \cdot D \cdot V^T, \quad (5.26)$$

wobei die rechte Seite formal die Eigenwertzerlegung der symmetrischen Matrix $X^T X$ darstellt. Die Eigenwerte der Messmatrix X werden damit über die Diagonalmatrix S aus Gl. (5.23) dargestellt, wobei der Betrag der Eigenwerte ein direktes Maß für den Varianzanteil eines Faktors darstellt [Danzer *et al.* (2001)]. Die Eigenvektoren der Matrix X , die der Ladungsmatrix aus Gl. (5.22) entspricht, bilden die Matrix V .

Nach der Berechnung der Ladungsmatrix kann durch einfache Umformung von Gl. (5.22) die Faktormatrix T ,

$$T = X \cdot P, \quad (5.27)$$

berechnet werden.

Die orthogonale Rotationstransformation und damit die Hauptkomponentenanalyse lässt sich demnach auf ein Eigenwertproblem zurückführen. Die Gesamtvarianz der Ausgangsdaten bleibt durch die Rotationstransformation erhalten, jedoch werden die Anteile auf den Achsen neu verteilt. Die Scorematrix enthält dabei die Abstände der jeweiligen Messpunkte zwischen dem Projektionspunkt und dem Hauptkomponentenmittelpunkt. Die Cosinuswerte der Winkel zwischen der jeweiligen Hauptkomponente und den ursprünglichen Achsen im Raum beinhaltet die Loadingmatrix.

Durch die lineare Transformation wird noch keine Reduktion des Datenraumes erreicht. Da die Hauptkomponenten aber nach dem Kriterium maximaler Varianz bestimmt werden, können die höheren PCs vernachlässigt werden.

Es können verschiedene heuristische und statistische Kriterien zur Entscheidung über die Anzahl signifikanter Hauptkomponenten verwendet werden. Die häufigsten Kriterien sind nach Otto [Otto (1997)]

- der Anteil der erklärten Varianz,
- das Eigenwert-Eins-Kriterium (Kaiser-Kriterium),
- der Scree-Test und
- die Kreuzvalidierung (*Cross validation*).

Bei Auswahl von d Hauptkomponenten ergibt sich bei dem in dieser Arbeit angewendeten Verfahren der Anteil der erklärten (kumulativen) Varianz σ_e^2 ,

$$\sigma_e^2 = \frac{\sum_{i=1}^d \lambda_i}{\sum_{i=1}^p \lambda_i} \cdot 100\% , \quad (5.28)$$

aus dem Verhältnis der Summe der ersten d Eigenwerte zur Summe aller p Eigenwerte und wird üblicherweise in Prozent angegeben. Die Summe aller Quadrate einer Spalte der Score Matrix stellen dabei den entsprechenden Eigenwert dar. Durch Festlegen eines Grenzwertes für die erklärte Varianz werden dann alle Hauptkomponenten, die den restlichen Anteil der Gesamtvarianz oberhalb dieser Grenze beschreiben, verworfen.

Da der Rauschanteil eine gleichmäßige Verteilung über alle Hauptkomponenten aufweist, die analytische Information sich aber auf die ersten Faktoren konzentriert, kommt es zu einer Optimierung des Signal-Rausch Verhältnisses.

Weitergehende Beschreibungen der Hauptkomponentenanalyse finden sich in der Literatur bei Brereton (2003), Brereton (1992) und Mardia *et al.* (1979).

Für die Durchführung der PCA wurden die BioView[®]-Datensätze zentriert. Die PCA kann sowohl mit nicht skalierten als auch mit skalierten Daten durchgeführt werden (Abschnitt 5.4.1). Bei der Verwendung nicht skalierten Daten besteht die Gefahr, dass Komponenten, die eine geringe Fluoreszenz im Vergleich zu anderen Komponenten aufweisen, keine Berücksichtigung finden.

Bei der Verwendung skalierten Daten ist zu beachten, dass auch Bereiche, die nur Rauschen enthalten, in die Hauptkomponenten mit einfließen. Aus diesem Grund wurden neun Wellenlängenkombinationen aus den Messdaten entfernt, die hauptsächlich Messrauschen enthielten. Die verworfenen Wellenlängenkombinationen sind in Tabelle 5.5 aufgeführt.

Tabelle 5.5: Hauptsächlich Messrauschen enthaltende Wellenlängenkombination

| Messsignal $S_{ex/em}$ | | | | | | | | |
|------------------------|---------------|---------------|---------------|----------------|---------------|----------------|----------------|---------------|
| $S_{270/410}$ | $S_{270/430}$ | $S_{270/570}$ | $S_{270/590}$ | $S_{270/EmND}$ | $S_{290/590}$ | $S_{290/EmND}$ | $S_{310/EmND}$ | $S_{550/590}$ |

Bei Anwendung der PCA ergeben sich die in Tabelle 5.6 aufgeführten Werte für die erklärte Varianz nach Gl. (5.28). Die Hauptkomponentenanalyse wurde dabei sowohl für die gesamte Kultivierungsdauer, als auch für die einzelnen Prozessphasen Batch, Fed-Bach und Produktion durchgeführt. Bei den angegebenen Werten in Tabelle 5.6 handelt es sich um Mittelwerte der Datensätze aller verwendeten Trainingskultivierungen. Die PCA wurde jeweils auf zentrierte

und auf autoskalierte Daten (Abschnitt 5.4.1) angewendet. Es zeigt sich, dass sich die Gesamtvarianz bei der Verwendung autoskalierter Daten auf mehr PCs verteilt als bei der Verwendung nicht autoskalierter Daten.

Als Grenze für die zur Modellkalibrierung einzusetzenden PCs wurde für zentrierte Daten 99.9 %, für autoskalierte Daten 99.0 % der erklärten Varianz verwendet. Oberhalb dieser Grenzen wiesen die ermittelten PCs hauptsächlich Rauschen auf.

Tabelle 5.6: Erklärte Varianz der ersten Hauptkomponenten aller Trainingsdatensätze

| Prozessphase | Daten- vorverarbeitung | Erklärte Varianz σ_e^2 [%] | | | |
|--------------|---------------------------|-----------------------------------|--------|--------|--------|
| | | PC_1 | PC_2 | PC_3 | PC_4 |
| Gesamt | zentriert | 97.0 | 2.58 | 0.35 | 0.06 |
| | autoskaliert | 91.6 | 6.39 | 1.33 | 0.27 |
| Batch | zentriert | 99.3 | 0.57 | 0.11 | 0.02 |
| | autoskaliert | 91.9 | 3.84 | 1.34 | 0.72 |
| Fed-Batch | zentriert | 99.6 | 0.35 | 0.03 | 0.01 |
| | autoskaliert | 97.8 | 0.94 | 0.45 | 0.28 |
| Produktion | zentriert | 98.4 | 1.20 | 0.26 | 0.07 |
| | autoskaliert | 78.7 | 15.1 | 3.42 | 1.30 |

Zur Interpretation der PCA werden grafische Darstellungen herangezogen. Neben den zur Verfügung stehenden Methoden zur Bestimmung der Anzahl signifikanter Hauptkomponenten werden vor allem die Darstellung der Scores und Loadings verwendet.

Im Faktorladungsplot in Bild 5.27 werden die Anteile der Variable bzw. hier speziell die Wellenlängenkombination dargestellt, die auf die entsprechende Hauptkomponente „laden“. Variable, die nahe dem Ursprung liegen und damit einen geringen Wert aufweisen, haben keinen oder nur geringen Einfluss auf die entsprechende Hauptkomponente. Wellenlängenkombinationen, die ähnliche Beträge aufweisen, haben vergleichbare Anteile.

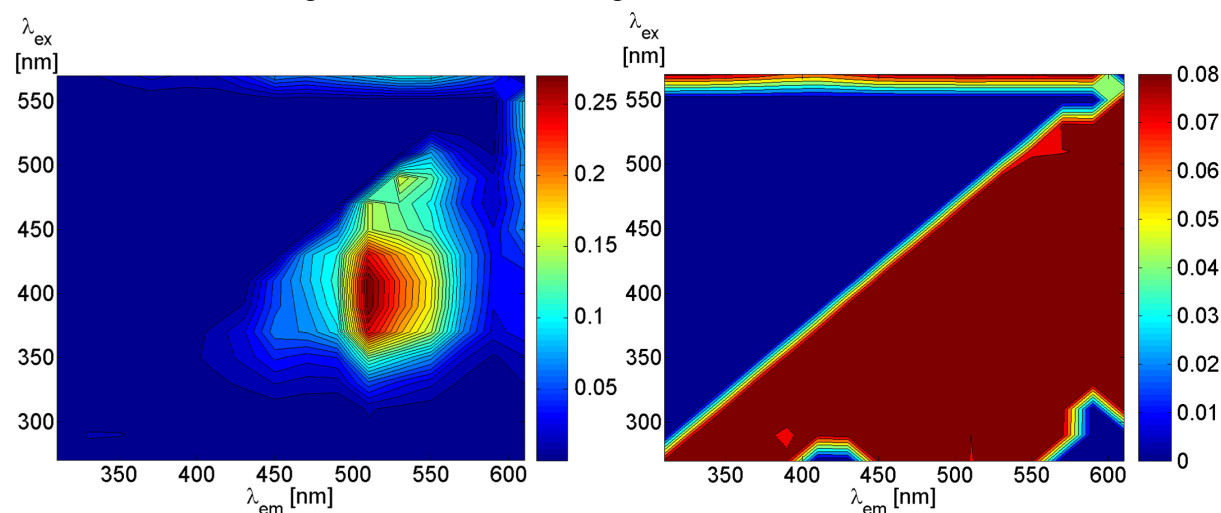


Bild 5.27: Loadings der ersten Hauptkomponente für den gesamten Kultivierungsverlauf
links: zentrierte Daten rechts: autoskalierte Daten

Die Darstellung gibt also Aufschluss darüber, von welchen Fluoreszenzbereichen die jeweilige Hauptkomponente massgeblich beeinflusst wird. Darüber können wieder Rückschlüsse auf die detektierten Fluorophore gezogen werden.

Beim Vergleich der Loadings der ersten Hauptkomponente beider oben aufgeführten Verfahren (Bild 5.27) wird deutlich, dass bei der Verwendung autoskalierter Daten keine spezifischen Fluoreszenzregionen zu erkennen sind. Die erste Hauptkomponente setzt sich damit aus dem gesamten Fluoreszenzspektrum zusammen.

Auch für die höheren Hauptkomponenten ergeben sich wesentlich unschärfere Fluoreszenzbereiche bei der Verwendung autoskalierter Daten. Dieses erschwert die Interpretation der gewonnenen PCs. Allerdings können im Falle autoskalierter Daten auch Fluoreszenzbereiche auf die höheren Hauptkomponenten laden, die nur eine geringe Fluoreszenzintensität aufweisen und im Falle nicht autoskalierter Daten keine Berücksichtigung finden.

Für die richtige Interpretation der BioView[®]-Daten können beide Verfahren von Bedeutung sein. Im Verlauf dieser Arbeit wurden deshalb beide Konzepte verfolgt und die jeweiligen Ergebnisse miteinander verglichen.

Üblich ist die Darstellung der PCs im sogenannten Biplot oder auch Scoreplot, wie in Bild 5.29 dargestellt. Dabei werden die signifikanten Hauptkomponenten gegeneinander aufgetragen. Gravierende Änderungen im Prozessverlauf werden dabei durch eine Richtungsänderung des Graphen sichtbar.

In der Zeitreihenanalyse ist außerdem noch die Auftragung der PCs über der Zeit von Bedeutung, wie in Bild 5.28 dargestellt. Diese erlaubt eine Beurteilung der einzelnen Hauptkomponenten im Prozessverlauf. In dem hier vorliegenden Anwendungsfall wird damit der prinzipielle zeitliche Verlauf der Fluoreszenzspektren wieder gegeben.

Bild 5.29 zeigt den Biplot der ersten beiden PCs der Kultivierung CK4407. In den Regionen, in welchen der Graph linear verläuft, sind die Komponenten klar getrennt. Die PCs verhalten sich proportional zueinander.

Während der Batch-Phase steigen beide PCs in gleichem Maße an. Der Graph verläuft daher linear. Die erste Änderung im Prozess ist der vollständige Verbrauch der vorgelegten Glukose am Ende der Batch-Phase ($t_1 = 5.4$ h). Dieser wird durch den Rückgang der zweiten Hauptkomponente indiziert. Während der substratlimitierten Fed-Batch Phase steigen beide PCs wieder proportional an. Lediglich die Reduktion der Zufütterraten am Anfang der Fed-Batch Phase wird in der grafischen Darstellung ersichtlich ($t_2 = 5.5$ h).

Zu Beginn der Produktionsphase ($t_3 = 8.9$ h) erreicht die zweite Hauptkomponente einen Scheitelpunkt und nimmt im Folgenden, bezogen auf die erste Hauptkomponente, überproportional ab. Dieses ändert sich erst durch die Reduktion der vorgegebenen Wachstumsrate ($t_4 = 14.0$ h). Am Ende der Produktionsphase kommt es zu einer starken Abnahme der Zielproteinproduktivität. Offensichtlich wird dieses sowohl durch den Rückgang der ersten Hauptkomponente als auch durch das erneute Ansteigen der zweiten Komponente ($t_5 = 17.6$ h).

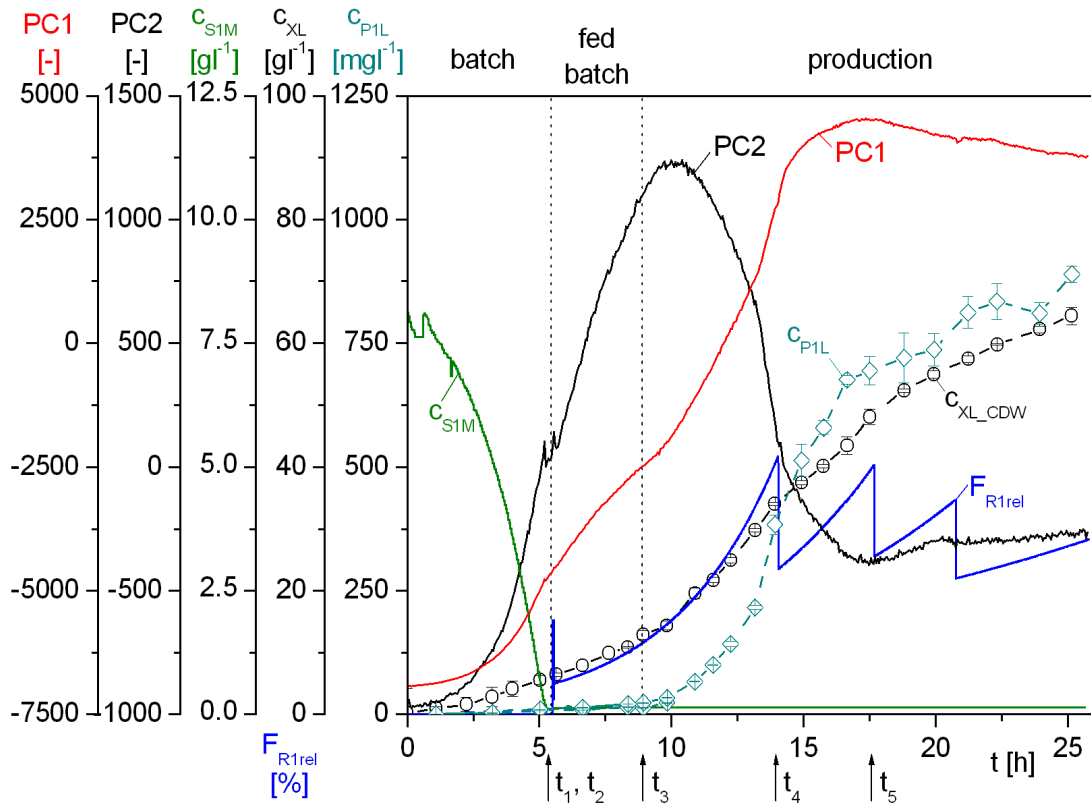


Bild 5.28: Zeitlicher Verlauf der ersten beiden Hauptkomponenten PC1 und PC2 sowie der Prozessgrößen Zelldichte c_{XL} , Glukosekonzentration c_{SIM} , Zielproteinkonzentration c_{P1L} und relative Glukosezufütterung F_{R1rel} während der Kultivierung CK4407

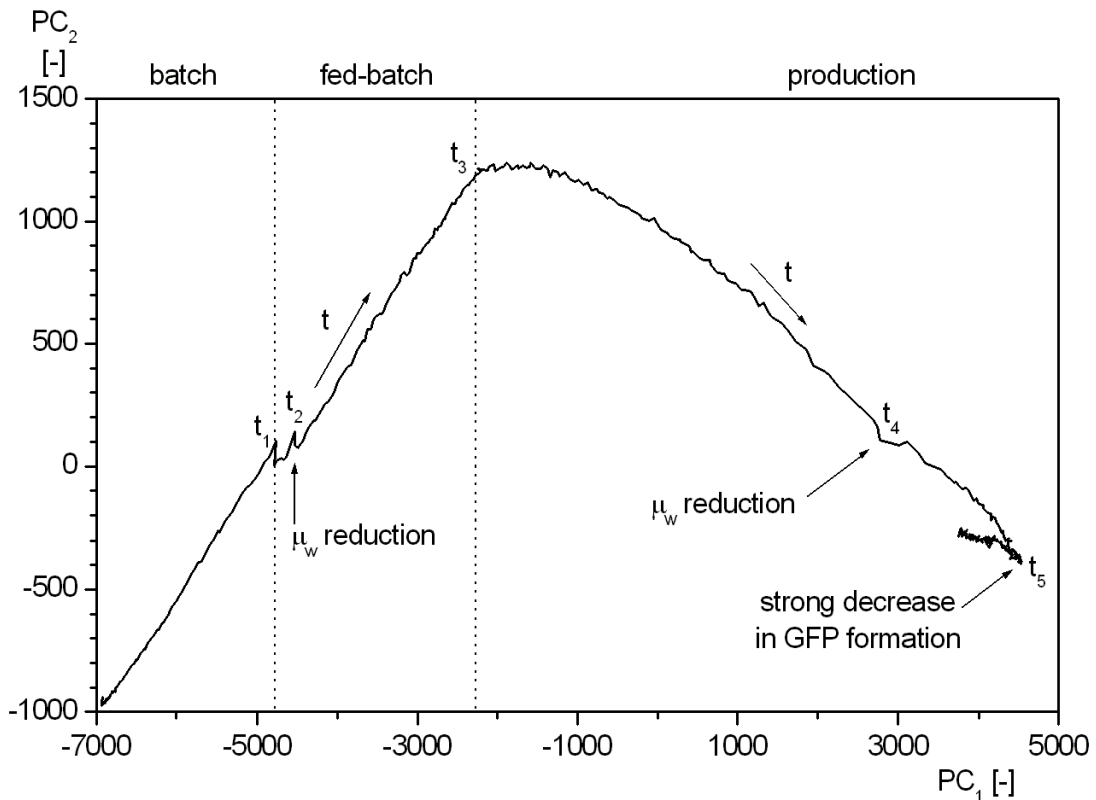


Bild 5.29: Biplot der ersten beiden Hauptkomponenten der Kultivierung CK4407

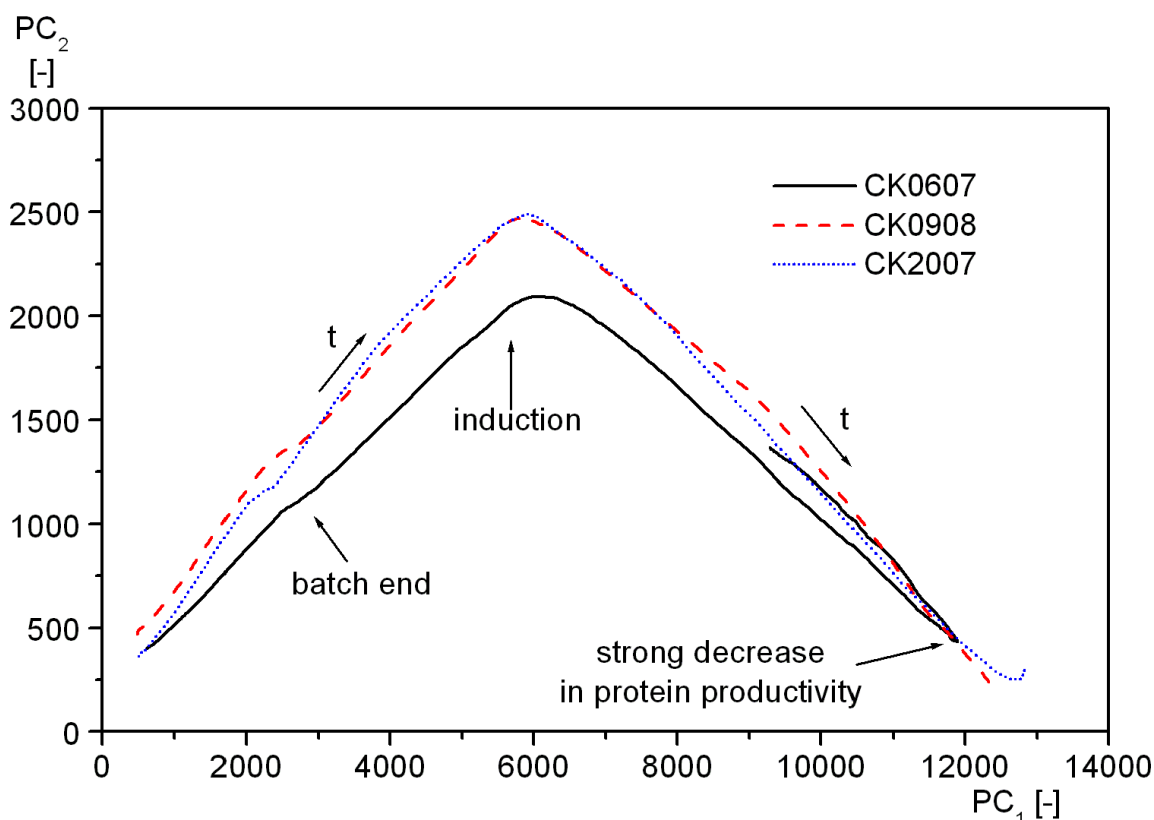


Bild 5.30: Biplot der ersten beiden PCs für drei Kultivierungen

Beim Vergleich der Biplots mehrerer Kultivierungen werden Unterschiede im Prozessverlauf eindeutig sichtbar. In Bild 5.30 sind exemplarisch drei Kultivierungen dargestellt, von denen zwei sehr ähnlich verlaufen sind, was sich in den entsprechenden Scoreplots widerspiegelt.

Auffällig ist, dass die Varianz der zweiten Hauptkomponente der Kultivierung CK0607 deutlich geringer ist, als die der anderen beiden Kultivierungen. Im Prozessverlauf schlägt sich dieses zunächst durch eine wesentlich längere Batch-Phase nieder. Diese dauerte 8,8 Stunden im Vergleich zu ca. 4,6 Stunden für die Kultivierungen CK0908 und CK2007. Da die Animpfdichte in allen drei Fällen ähnlich war, ist dieses wahrscheinlich auf Limitierungen im Verlauf der Vorkultur zurückzuführen. Die dadurch bedingte wesentlich geringere maximale Wachstumsrate im unlimitierten Batch wird durch die geringere Varianz der zweiten Hauptkomponente wiedergegeben.

Die Induktion bei einer geringeren Zelldichte von 13 g l^{-1} (CK0607), im Gegensatz zu ca. 20 g l^{-1} bei den beiden anderen Kultivierungen, wird durch eine geringere Varianz der zweiten Hauptkomponente deutlich sichtbar. Während bei den Kultivierungen CK2007 und CK0908 die Zielproteinexpression nahezu identisch verlief, brach die GFP-Bildung bei der Kultivierung CK0607 ein, was durch eine Richtungsänderung im Biplot ersichtlich wird. Im Folgenden kam die Expression nahezu vollständig zum Erliegen und die Zielproteinkonzentration nahm durch den Verdünnungseffekt der Zufütterung ab.

Die vergleichende Darstellung der Hauptkomponenten unterschiedlicher Kultivierungen im Biplot erlaubt eine Beurteilung der Vergleichbarkeit der einzelnen Kultivierungen. Darüber können Störungen im Prozessverlauf erkannt werden. Mögliche Anwendungen wären z.B. das frühzeitige Erkennen des Produktivitätsverlustes der Zellen und eine damit erzwungene Zellernte.

5.5.2 Unabhängige Komponentenanalyse

Ein weiteres Verfahren zur Merkmalsextraktion und Datenreduktion ist, neben der zuvor beschriebenen PCA, die unabhängige Komponentenanalyse. Sie ist ein statistisches Verfahren, mit dem Ziel Signale durch eine lineare Transformation in die ihnen zugrundeliegenden unabhängigen Komponenten zu zerlegen. Sie fordert, neben der linearen Unabhängigkeit, welche auch die Grundlage der Hauptkomponentenanalyse darstellt, dass die zu bestimmenden Komponenten möglichst statistisch unabhängig voneinander sind.

Dabei wird davon ausgegangen, dass sich die Messsignale aus der Superposition verschiedener unabhängiger Quellen zusammensetzen, wobei jedoch sowohl das Quellsignal als auch die Mischung unbekannt sind. Bei den hier vorliegenden BioView[®]-Daten entspricht diese Superposition zum Einen der Überlagerung der an verschiedenen Stoffwechselwegen beteiligten biogenen Fluorophore und zum Anderen der Beeinflussung verschiedener Fluoreszenzbereiche durch den Kaskadeneffekt.

Das am häufigsten verwendete Beispiel, um die sogenannte blinde Quellentrennung oder auch *Blind Source Separation*¹⁵ zu illustrieren, ist das sogenannte Cocktailparty-Problem. Reden in einem Raum zwei Menschen durcheinander und sind an mindestens zwei verschiedenen Orten im Raum Mikrophone zur Geräuschaufnahme installiert, so stellen die aufgenommenen Messsignale x die Linearkombination der statistisch unabhängigen Signale s der Sprecher dar. Die Aufgabe der ICA ist es, die unabhängigen Signale zu rekonstruieren.

Unter Annahme einer linearen Zusammensetzung ergibt sich für die beiden Messsignale x_1 und x_2 ,

$$\begin{aligned} x_1(t) &= a_{11} \cdot s_1(t) + a_{12} \cdot s_2(t) \\ x_2(t) &= a_{21} \cdot s_1(t) + a_{22} \cdot s_2(t) , \end{aligned} \quad (5.29)$$

jeweils eine Linearkombination von s_1 und s_2 .

Das Hauptproblem besteht darin, dass nicht nur die Quellsignale s_i sondern auch die Mischungsmatrix A unbekannt sind. Anstatt statistische Annahmen (Varianzmaximierung) über die Signale s_i zu treffen (wie bei der PCA) und darüber die Koeffizienten a_{ij} zu identifizieren, liegt der ICA die Annahme zugrunde, dass die Signale stochastisch voneinander unabhängig sind. Dabei reicht es auch, wenn diese Annahme näherungsweise erfüllt wird [Hyvärinen and Oja (2000)].

Allgemein formuliert seien nun $x_1 \dots x_n$,

$$x_k(t) = a_{k1} \cdot s_1(t) + a_{k2} \cdot s_2(t) + \dots + a_{kn} \cdot s_n(t) , \quad (5.30)$$

die gewichteten Summen von n unabhängigen Komponenten. Des Weiteren werden sowohl die Quellsignale s_i , die den separierten unabhängigen Fluoreszenzantworten des biologischen Systems entsprechen, als auch die Messsignale x_k , welche die einzelnen Wellenlängenpaare des BioView[®]-Messgerätes darstellen, nicht mehr als Funktionen der Zeit betrachtet, sondern als Zufallsvariablen. Damit stellen die Messungen x_k Stichproben der Zufallsvariablen dar. Durch eine Zentrierung der Variablen x und s besitzen beide den Erwartungswert Null.

¹⁵ *Blind Source Separation* – BSS

In Matrixschreibweise kann das ICA-Modell,

$$x = A \cdot s, \quad (5.31)$$

nach den gesuchten Quellsignalen s ,

$$s = W \cdot x, \quad (5.32)$$

umgestellt werden, wobei die Inverse der Mischmatrix A auf Grund der Uneindeutigkeit der ICA mit W bezeichnet wird. Da sowohl A als auch s unbekannt sind, ergeben sich zwei Freiheitsgrade. Daraus folgt, dass die Varianzen der ICs nicht bestimmbar sind, da jeder skalare Faktor in s in A ausgeglichen werden kann.

Die Varianz ist deshalb unerheblich für die ICA und wird in den meisten Verfahren auf $E\{\sigma_i^2\} = 1$ normiert. Bedingt durch die Freiheitsgrade ist das Vorzeichen der unabhängigen Komponenten nicht eindeutig bestimmbar. Die Reihenfolge der ICs ist beliebig, da sich jede Änderung der Reihenfolge in s in A wieder ausgleichen läßt [Hyvärinen and Oja (2000)].

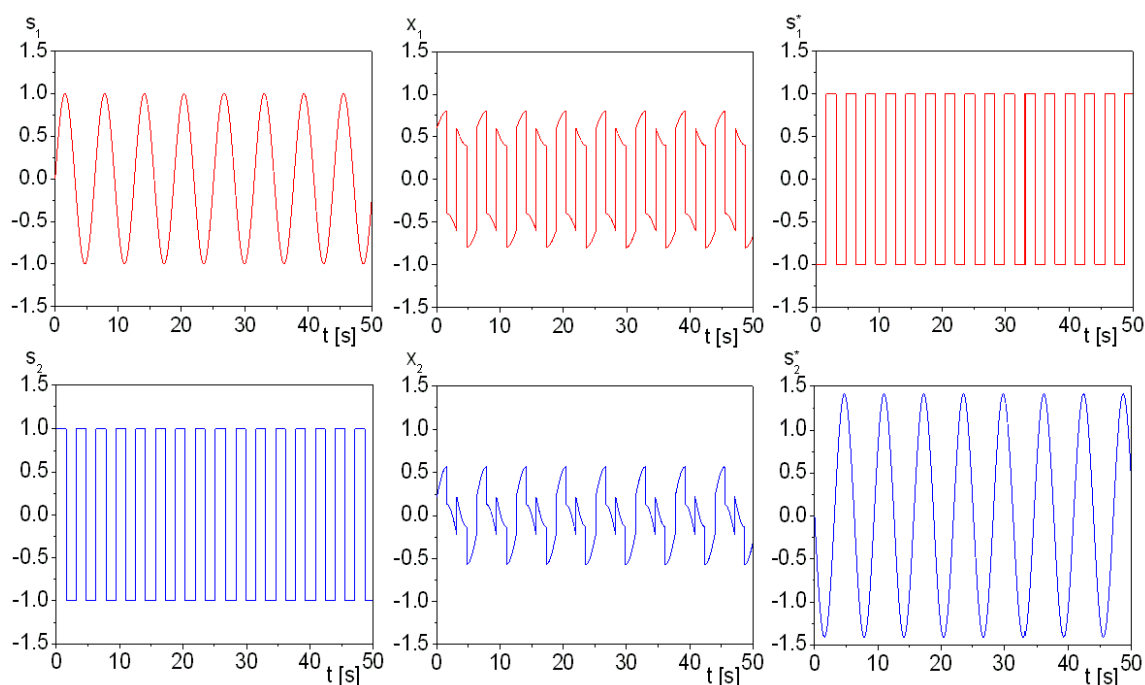


Bild 5.31: Anwendung der ICA auf die Mischung von zwei Quellsignalen
 links: Ausgangssignale mitte: Mischung der Ausgangssignale
 rechts: Approximation der Ausgangssignale unter Verwendung der ICA

In Bild 5.31 ist ein Beispiel für eine simple Anwendung der ICA gegeben. Auf der linken Seite sind die beiden Quellsignale gezeigt, die in der Mitte der Abbildung miteinander vermischt worden sind. Durch die ICA ist es möglich, die beiden Quellsignale zu rekonstruieren (rechts). Die nicht mehr zu extrahierenden Informationen über die Reihenfolge, die Amplitudenhöhe sowie das Vorzeichen gehen dabei jedoch verloren.

Sollen unabhängige Komponenten als Eingangsgrößen zur Modellkalibrierung verwendet werden, bedeutet dieses, dass die ICA einmal berechnet und dann für die restlichen Daten die selbe Entmischungsmatrix W verwendet werden muss, damit die Varianzen und die Reihenfolge der verwendeten Daten konsistent sind.

Zur Zerlegung der Messmatrix des BioView[®]-Fluoreszenzensors in stochastisch unabhängige Komponenten muss im Folgenden ein mathematisches Maß gefunden werden. Die stochastische Unabhängigkeit ist gegeben, wenn die Variable y_1 keine Information über y_2 enthält und umgekehrt. Mathematisch kann dies über die Wahrscheinlichkeitsdichtefunktion p definiert werden.

Lässt sich diese bei n Variablen vollständig faktorisieren,

$$p(y_1, y_2, \dots, y_n) = p_1(y_1) \cdot p_2(y_2) \cdot \dots \cdot p_n(y_n), \quad (5.33)$$

dann sind die einzelnen Komponenten voneinander unabhängig.

Illustrieren lässt dich dieses an zwei Würfeln. Die Wahrscheinlichkeit, dass der erste Würfel eine drei zeigt, ist $p(y_1 = 3) = 1/6$. Das Gleiche gilt für den zweiten Würfel, $p(y_2 = 3) = 1/6$. Die Wahrscheinlichkeit, dass beide Würfel eine drei zeigen ist dann vollständig faktorierbar, $p(y_1 = 3)p(y_2 = 3) = 1/36 = p(y_1 = 3, y_2 = 3)$ und die beiden Ereignisse sind voneinander statistisch unabhängig, was sich auch mit der intuitiven Vorstellung deckt [Schneegaß (2004)].

Eine schwächere Form der Unabhängigkeit ist die Unkorreliertheit, wie sie in der Hauptkomponentenanalyse (Abschnitt 5.5.1) zu Grunde gelegt ist. Es läßt sich aber zeigen, dass die PCA die in Gl. (5.33) gestellte Bedingung nicht erfüllt [Hyvärinen and Oja (2000)]. Dieses ist darauf zurückzuführen, dass dort nur Statistiken zweiter Ordnung (Kovarianz) verwendet werden, bei der ICA aber Statistiken aller Ordnungen Anwendung finden.

Da die stochastische Unabhängigkeit jedoch die Unkorreliertheit der Variablen impliziert, kann die Hauptkomponentenanalyse zur Datenvorverarbeitung für die ICA dienen und damit den Datenraum reduzieren sowie das Problem simplifizieren [Hyvärinen and Oja (2000)].

Die Basis der ICA ist der Grenzwertsatz der Statistik:

„Die Summe von vielen beliebig verteilten Zufallsgrößen ist näherungsweise normalverteilt.“

Dies bedeutet, dass die stochastischen Eigenschaften der Sensorsignale gaußähnlicher sind als die der zu identifizierenden unabhängigen Komponenten. Um eine geeignete Projektion durch W zu finden, kann die Gaußähnlichkeit verwendet werden.

Daraus ergibt sich die fundamentale Einschränkung der Independent Component Analysis, dass die unabhängigen Komponenten nicht gaußverteilt sein dürfen, da das Verfahren ansonsten nicht konvergieren kann.

Begründet liegt dies in der Symmetrie der Verteilung, die in Bild 5.32 für zweidimensionale gaußverteilte Daten dargestellt ist. Es lässt sich keine Koordinatenrichtung finden, welche die Verteilungsdichtefunktion dieser Daten ändert. Ein wesentlicher Unterschied zwischen den beiden Koordinatentransformationen der PCA und der ICA wird ebenfalls in Bild 5.32 ersichtlich. Bei der Independent Component Analysis werden im Gegensatz zur Hauptkomponentenanalyse auch nicht orthogonale Koordinatensysteme als Basis verwendet.

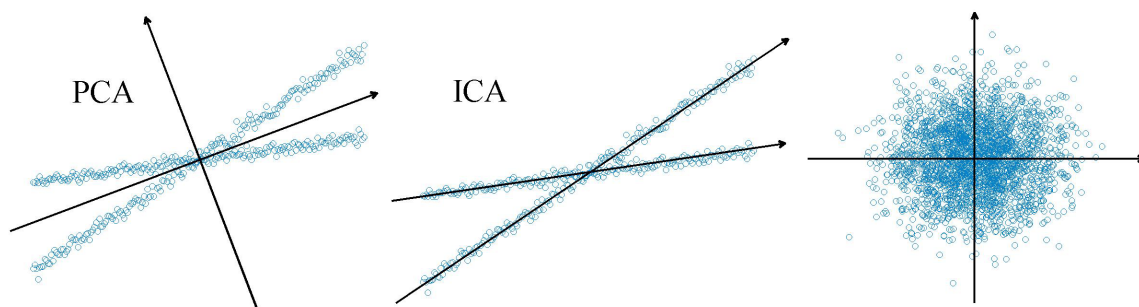


Bild 5.32: Grafische Darstellung der ICA und der PCA

links: PCA mitte: ICA rechts: Scoreplot der multivariaten Normalverteilung

Die üblichsten Verfahren zur Berechnung der Gaußähnlichkeit sind die Kurtosis und die Negentropie. Die Kurtosis ist der Parameter der deskriptiven Statistik, der die Wölbung einer statistischen Verteilung x beschreibt. Sie stellt die normierte Form des vierten zentralen Moments dar,

$$\text{kurt}(x) = E\{x^4\} - 3(E\{x^2\})^2, \quad (5.34)$$

und vereinfacht sich unter der Voraussetzung normierter Varianz zu,

$$\text{kurt}(x) = E\{x^4\} - 3, \quad (5.35)$$

wobei der Wert null einer Normalverteilung entspricht.

Für die meisten nicht gaußartigen Verteilungen ist die Kurtosis ungleich null. Bei positiven Werten wird die Verteilung als supergaußförmig, leptokurtisch oder auch spitzgipflig bezeichnet. Negative Werte indizieren subgaußförmige Verteilungen, die auch als platykurtisch oder flachgipflig bezeichnet werden.

Ein Nachteil der Kurtosis ist die Anfälligkeit gegenüber Ausreißern, insbesondere bei kleinen Datensätzen [Huber (1985)]. Eine detailliertere Beschreibung des Einsatzes der Kurtosis für die ICA findet sich bei Delfosse and Loubaton (1995).

Die Entropie H ist in der Informationstheorie ein Maß für den mittleren Informationsgehalt eines Zeichens. Ein Zeichen, welches in einer Quelle häufig vorkommt, besitzt weniger Informationsgehalt als ein Zeichen, dessen Aufenthaltswahrscheinlichkeit geringer ist.

Für eine diskrete Zufallsvariable Y ist die Entropie,

$$H(Y) = - \sum_i P(Y = a_i) \cdot \log P(Y = a_i), \quad (5.36)$$

definiert als die Summe aus dem Produkt aller möglichen Werte a_i von Y und dem zugehörigen dekadischen Logarithmus. In differentieller Schreibweise kann die Entropie auch als eine kontinuierliche Verteilungsfunktion,

$$H(y) = - \int f(y) \cdot \log(f(y)) \cdot dy, \quad (5.37)$$

dargestellt werden.

Da die Entropie lediglich eine Aussage im Vergleich mit der jeweiligen Gaußverteilung zulässt, wird diese normalisiert. Dazu wird die beliebige Zufallsvariable y ins Verhältnis zur Entropie der Gaußverteilung $H(y_{\text{gauss}})$, welche die gleiche Kovarianzmatrix wie y besitzt, gesetzt. Das führt auf die negative Entropie J ,

$$J(y) = H(y_{\text{gauss}}) - H(y) , \quad (5.38)$$

die kurz als Negentropie bezeichnet wird. Sie stellt ein Maß für die Abweichung einer Verteilung von der Gaußverteilung dar, ist immer positiv und wird nur zu null, wenn die Zufallsvariable gaußverteilt ist [Cover and Thomas (1991)][Papoulis (1991)].

Die Negentropie stellt ein sehr gutes und robustes Maß für die Nicht-Gaußartigkeit von Variablen dar. Ein Nachteil ist, dass die Wahrscheinlichkeitsdichtefunktion von y bekannt sein oder geschätzt werden muss, was sehr rechenintensiv ist [Hyvärinen and Oja (2000)]. Verschiedene Ansätze zur Approximation der Negentropie finden sich in der Literatur [R. and Sibson (1987)][Hyvärinen (1998)].

In dem in dieser Arbeit zur Approximation der Negentropie verwendeten Fast-Fixed-Point Algorithmus [Hyvärinen (1998)],

$$J(y_i) \approx c [E \{G(y_i)\} - E \{G(v)\}]^2 , \quad (5.39)$$

stellt c eine Konstante dar. G ist eine Kontrastfunktion, die der Approximation der Negentropie dient, auf die etwas später eingegangen wird, und v eine standardnormalverteilte¹⁶ Zufallsvariable.

Unter Berücksichtigung von Gl. (5.32) kann das zu optimierende Funktional,

$$J_G(w) = [E \{G(w^T \cdot x)\} - E \{G(v)\}]^2 , \quad (5.40)$$

unter der Bedingung $E \{(w^T \cdot x)^2\} = 1$, aufgestellt werden.

Bei Verwendung der *Mutual Information* I ,

$$I(y_1, y_2, \dots, y_n) = C - \sum_i J(y_i) , \quad (5.41)$$

die ein informationstheoretisches Maß für die Abhängigkeit von Zufallsvariablen darstellt, kann das Optimierungsproblem für alle ICs formuliert werden. Je größer der Wert für die *Mutual Information* wird, desto unabhängiger sind die Variablen. Je größer also die Negentropie,

$$J_G(w) = \max \sum_{i=1}^n J_G(w_i) , \quad (5.42)$$

wird, desto unabhängiger sind die entsprechenden Variablen.

Da Gl. (5.42) iterativ über geeignete Gradientenabstiegserfahren gelöst werden muss, werden differenzierbare Kontrastfunktionen zur Approximation der Negentropie verwendet.

¹⁶standardisierte Normalverteilung siehe Gl.(5.10)

Eine gute allgemeine Kontrastfunktion, die in dieser Arbeit ausschließlich eingesetzt wurde, stellt nach Hyvärinen and Oja (2000) der Logarithmus der Cosinus-Hyperbolikusfunktion dar,

$$G(u) = \frac{1}{a_1} \cdot \log \cosh(a_1 \cdot u) , \quad (5.43)$$

die in ihrer ersten Ableitung auf den Tangenshyperbolikus führt,

$$g(u) = \tanh(a_1 \cdot u) . \quad (5.44)$$

Detaillierte Beschreibungen möglicher Kontrastfunktionen sowie der Methoden der Gradientenabstiegsverfahren finden sich bei Hyvärinen and Oja (2000) und Hyvärinen (1998).

Eines der Hauptprobleme der ICA liegt darin begründet, dass die ICs stochastische Größen darstellen. Auf Grund der endlichen Größe der Stichprobe sind die Approximationen daher fehlerbehaftet. Außerdem gibt es keine Gewähr dafür, dass die in einem Durchlauf bestimmten Komponenten nicht auf einem lokalen Minima der verwendeten Kontrastfunktion beruhen und damit suboptimale Werte liefern [Himberg *et al.* (2004)][Himberg and Hyvärinen (2003)].

Auf Grund dessen hat das Neural Networks Research Center der Universität Helsinki eine MATLAB[®]-Toolbox entwickelt, welche die Zuverlässigkeit des verwendeten Fast-Fixed-Point Algorithmus sowie der zugrundeliegenden statistischen Verfahren wesentlich erhöht. Die sogenannten ICASSO-Funktionen sind frei im Internet erhältlich [Laboratory of Computer and Information Science: Adaptive Informatics Research Center (2005)] und wurden für diese Arbeit verwendet. Im Folgenden soll das Vorgehen zum Auffinden aussagekräftiger unabhängiger Komponenten näher erläutert werden.

Im ersten Schritt werden die Parameter des zugrundeliegenden Algorithmus festgelegt. Im Wesentlichen sind das

- die Kontrastfunktion aus Gl. (5.43),
- die parallele oder serielle Berechnung der ICs,
- die Verwendung fixer oder variabler Anfangswerte und/oder
- *Bootstrapping*¹⁷ der verwendeten BioView[®]-Signale,
- sowie die Anzahl der Berechnungen.

Die Wahl der parallelen bzw. der seriellen Berechnung der ICs hatte im konkreten Anwendungsfall keinen erkennbaren Einfluss auf die Ergebnisse der ICA. Eine nähere Beschreibung der beiden Verfahren findet sich bei Hyvärinen and Oja (2000). Um eine gute Verallgemeinerungsfähigkeit der ICs sicherzustellen, wurden variable Anfangswerte für die Lösung von Gl. (5.42) verwendet. Insgesamt wurden die ICs 500 mal berechnet.

Im zweiten Schritt wird die ICA unter den oben angegebenen Bedingungen wiederholt durchgeführt. Dabei werden die Ergebnisse geclustert. Als Ähnlichkeitsmaß werden die Absolutwerte der Korrelationskoeffizienten zwischen den einzelnen ICs verwendet.

¹⁷*Bootstrapping*–iterative Vermehrung der Datenmenge

Die Clustererstellung erfolgt mittels hierarchischer Clusteranalyse, wobei für die Aggregation der Cluster das *average-linkage* Kriterium angewendet wird. Eine detaillierte Beschreibung des angewendeten Verfahrens findet sich bei Himberg *et al.* (2004) sowie Himberg and Hyvärinen (2003). Eine gute Einführung in die Clusteranalyse, die wie auch die ICA und die PCA zu den chemometrischen Verfahren des unüberwachten Lernens gehören, gibt Otto (1997).

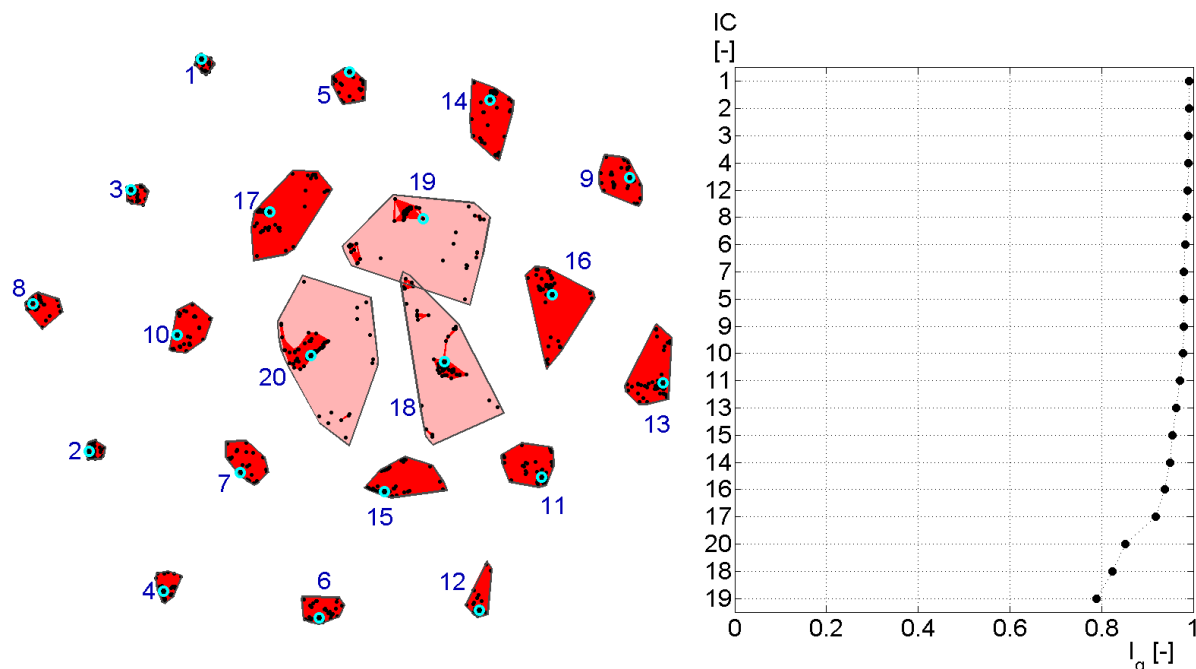


Bild 5.33: Clusteranalyse der ICA unter Verwendung der ICASSO-Funktionen

Im dritten Schritt erfolgt die Beurteilung der Güte der gebildeten Cluster. Dafür stehen grafische Darstellungen zur Verfügung, von denen zwei exemplarisch in Bild 5.33 wiedergegeben sind.

Nachdem die ICA für den dargestellten Fall 80 mal durchgeführt wurde, ergeben sich bei Annahme von 20 ICs die oben dargestellten Cluster. Dabei ist jedes Cluster entsprechend der zugehörigen Komponente nummeriert und der Schwerpunkt des Clusters mit einem offenen Kreis gekennzeichnet.

Je kleiner die Fläche ist, die ein Cluster in der 2D-Darstellung einnimmt, desto höher ist die Güte der identifizierten Komponente. Dieses spiegelt sich auch im einheitenlosen Cluster-Qualitätsindex $I_q \in [0, 1]$ wieder. Je höher der Wert des Qualitätsindex ist, desto größer ist die Güte des entsprechenden Clusters [Himberg and Hyvärinen (2003)].

Die am klarsten getrennten unabhängigen Komponenten ergaben sich, wenn zunächst eine PCA mit zentrierten Daten ohne Autoskalierung vorgenommen wurde. Dafür wurden sowohl die Trainings- als auch die Validierungsdaten der BioView[®]-Messungen mit der selben Ladungsmatrix wie im vorherigen Abschnitt transformiert. Anschließend wurde mit Hilfe der oben beschriebenen ICASSO-Funktionen die Entmischungsmatrix W aus Gl. (5.32) für alle Trainingsdatensätze bestimmt und die BioView[®]-Validierungsdaten mit dieser Matrix transformiert. Im Suchraum von 1 bis 20 unabhängigen Komponenten ergab sich für drei ICs die klarste Trennung.

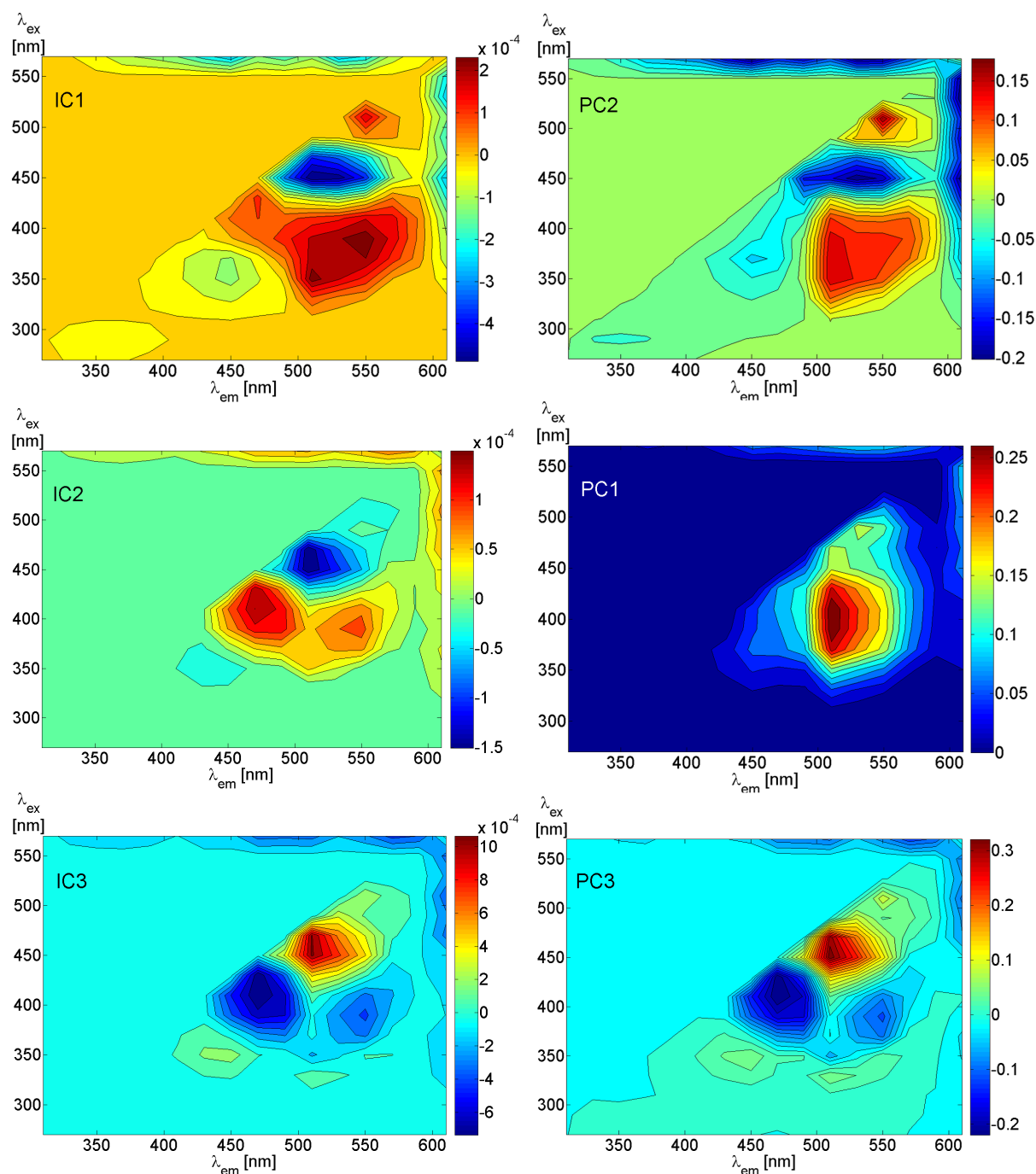


Bild 5.34: Koeffizientenmatrix und Loadings der ersten drei Komponenten der ICA und der PCA

Zur Interpretation der unabhängigen Komponenten können ebenfalls grafische Verfahren verwendet werden.

Die 2D-Darstellung der Zeilen der Koeffizientenmatrix W (Gl. (5.32)) ist vergleichbar mit der Darstellung der Loadingvektoren, wie in Bild 5.34 aufgezeigt. Sie gibt ebenfalls Aufschluß über die einzelnen Fluoreszenzbereiche, aus denen die jeweilige Komponente gebildet wird. Beim Vergleich der Loadings der PCA mit der Koeffizientenmatrix W fällt auf, dass sich insbesondere die Zusammensetzung der ersten beiden Komponenten in mehreren Bereichen unterscheidet.

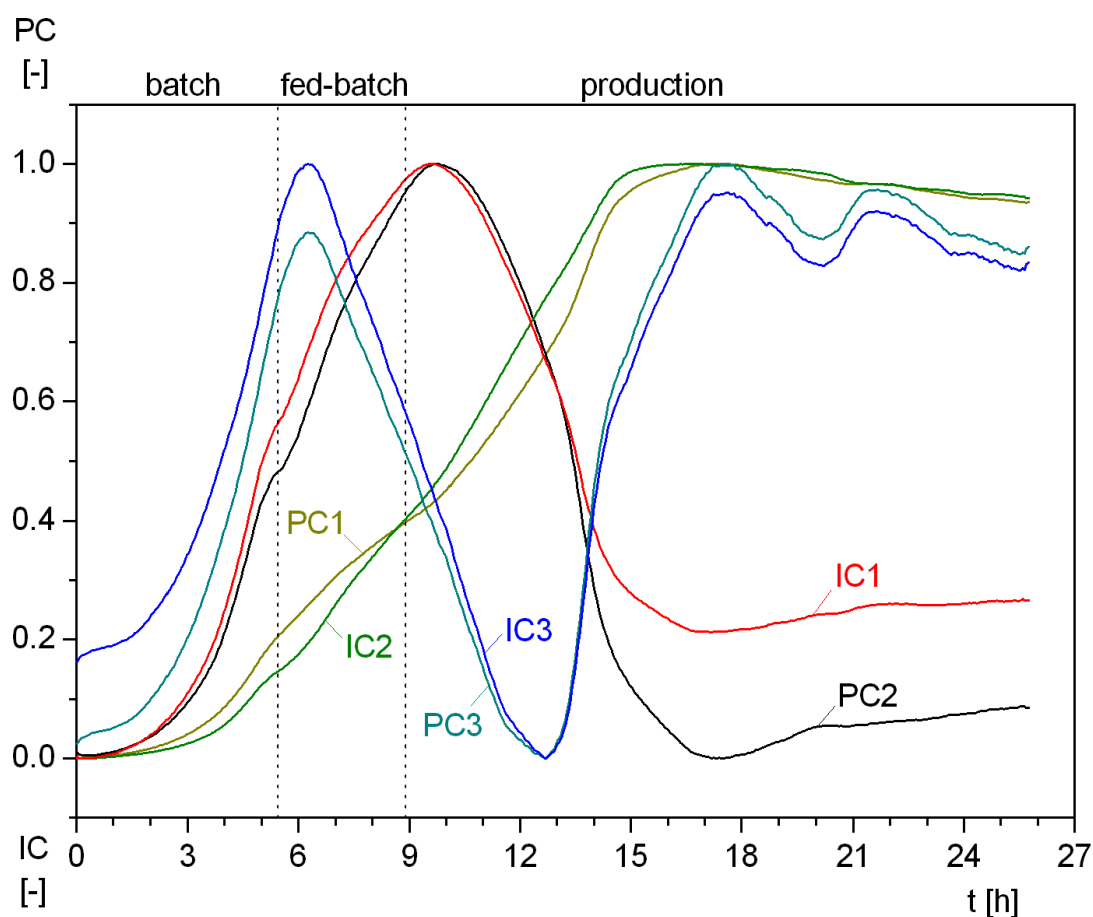


Bild 5.35: Zeitlicher Verlauf der ersten drei PCs und ICs einer Kultivierung

Die bereichsskalierten zeitlichen Verläufe der ersten drei Komponenten beider Verfahren sind in Bild 5.35 dargestellt. Es zeigt sich, dass die prinzipiellen Verläufe der bestimmten Komponenten ähnlich sind, was auch darauf zurückzuführen ist, dass der ICA eine PCA vorgeschaltet war. Dieses war jedoch notwendig, da sich ansonsten keine klar getrennten Komponenten ergaben. In wie weit sich die Unterschiede in den Komponenten auf die Modellbildung auswirken, ist im nachfolgenden Abschnitt aufgeführt.

5.5.3 Ergebnisse der multivariaten Modellkalibrierung

Für die oben beschriebenen drei Verfahren zur Kalibrierung der multivariaten Modelle sind die Anpassungen für die Zelldichte exemplarisch in Bild 5.36, für den Validierungsdatensatz CK0707 und die zugehörigen Fehler aller verwendeten Datensätze in Tabelle 5.7, dargestellt.

Die beste Anpassung an die Offline-Daten zeigten dabei die mit den Original-BioView[®]-Daten erstellten Modelle. Die Nachteile bei der Verwendung der Originaldaten liegen in einem höheren Rauschanteil sowie im wesentlich höheren Zeitaufwand zur Kalibrierung. Bei den hier vorliegenden 150 Wellenlängenkombinationen und drei Eingangsgrößen ergeben sich mehr als 500.000 Kombinationen, die angepasst werden müssen.

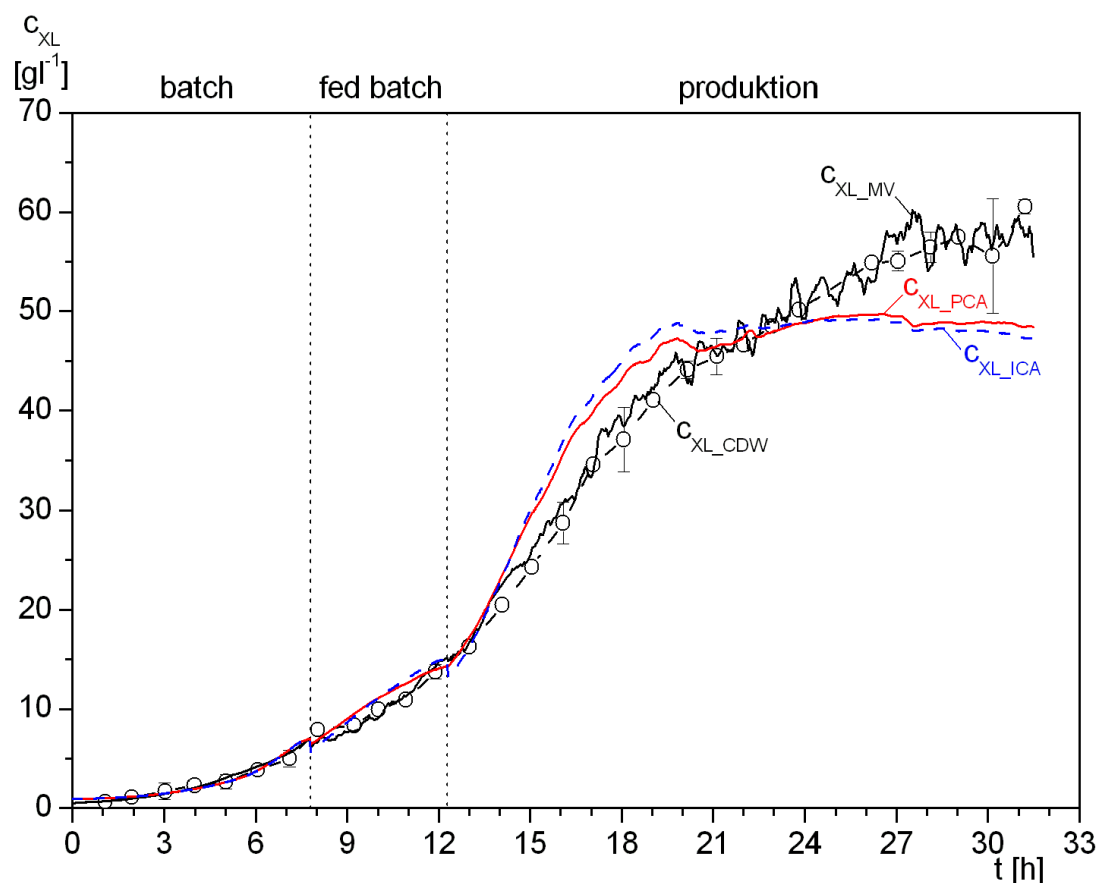


Bild 5.36: Zelldichteverläufe verschiedener multivariater Modelle

für den Validierungsdatensatz CK0707

Eingangsdaten: c_{XL_MV} := Originalvariablen c_{XL_PCA} := PCs c_{XL_ICA} := ICs

Tabelle 5.7: Fehler der besten multivariaten Modelle für die Zelldichte c_{XL}

| Daten | Anzahl Eingangs- größen | Phase | Anpassung | RMSEC | | RMSEP | |
|----------|-------------------------------|------------|--------------|----------------------|------|----------------------|------|
| | | | | [g l ⁻¹] | [%] | [g l ⁻¹] | [%] |
| BioView® | 2 | Batch | linear | 0.431 | 6.43 | 0.336 | 4.65 |
| | 3 | Fed-Batch | linear | 0.569 | 2.55 | 0.592 | 10.2 |
| | 2 | Produktion | exponentiell | 3.64 | 6.21 | 1.59 | 3.73 |
| PCs | 3 | Batch | linear | 0.563 | 8.40 | 0.398 | 5.51 |
| | 2 | Fed-Batch | linear | 1.57 | 7.03 | 0.750 | 12.9 |
| | 4 | Produktion | linear | 6.69 | 11.4 | 5.97 | 14.0 |
| ICs | 3 | Batch | linear | 0.506 | 7.54 | 0.304 | 4.21 |
| | 2 | Fed-Batch | linear | 1.45 | 6.46 | 1.43 | 24.5 |
| | 2 | Produktion | linear | 6.74 | 11.5 | 6.75 | 15.8 |

Die auf Basis der ICA und der PCA erstellten Modelle unterscheiden sich sowohl im Fehler als auch im Verlauf nicht wesentlich voneinander. Während der Produktionsphase zeigt sich jedoch eine wesentlich größere Abweichung von der offline bestimmten Biotrockenmasse als bei der Verwendung von Originaldaten. Die Verbesserung des Signal-Rausch Verhältnisses durch die verwendeten Datenreduktionsmethoden spiegelt sich im wesentlich ruhigerem Verlauf der modellierten Größen im Vergleich mit den Modellen der Originaldaten wieder.

Da die Modellierung mit Hilfe von autoskalierten Daten sowie mit abschnittsweise bestimmten PCs zu keiner Verbesserung der Modellfehler führte, wurden diese der Einfachheit halber nicht weiter verwendet.

Tabelle 5.8: Fehler der multivariaten Modelle auf Basis der BioView[®]-Originaldaten

| Daten | Anzahl Eingangs- größen | Phase | Anpassung | RMSEC | | RMSEP | |
|------------------|-------------------------------|---------------------|-----------|-----------------------|------|-----------------------|------|
| | | | | [gl ⁻¹] | [%] | [gl ⁻¹] | [%] |
| C _{S1M} | 2 | Batch | linear | 0.760 | 6.55 | 1.10 | 9.92 |
| C _{P1L} | 2 | Batch/ Fed-Batch | linear | 3.17·10 ⁻³ | 9.59 | 2.49·10 ⁻³ | 8.30 |
| | 2 | Produktion | linear | 45.6·10 ⁻³ | 6.60 | 40.5·10 ⁻³ | 6.91 |

In Abb 5.37 und 5.38 sind exemplarisch die Ergebnisse der multivariaten Modelle unter Verwendung von Originaldaten für zwei Validierungsdatensätze und in Bild 5.39 bis 5.41 für drei Trainingsdatensätze dargestellt. Die zugehörigen Fehler und die gewählten Anpassungen für die Bestimmung der Zelldichte, der Glukosekonzentration und der Zielproteinkonzentration sind in Tabelle 5.7 und 5.8 aufgeführt. Die verwendeten Koeffizienten der einzelnen Modelle sowie die zugehörigen Wellenlängenpaare können im Anhang eingesehen werden.

Für die Zelldichte sowie die Zielproteinkonzentration ergeben sich sehr präzise Verläufe. Die Modellfehler konnten im Vergleich mit den univariaten Modellen aus Abschnitt 5.4.3 für alle Messgrößen reduziert werden. Insbesondere die Zelldichteverläufe sowie die Zielproteinkonzentration während der Produktionsphase konnten unter Verwendung von mehreren Wellenlängenpaaren erheblich verbessert werden.

Lediglich für die Glukosekonzentration konnten sowohl der Validierungs- als auch der Trainingsfehler nicht gravierend verbessert werden. Auch die korrekte Wiedergabe des vollständigen Zuckerverbrauchs konnte nicht wesentlich optimiert werden.

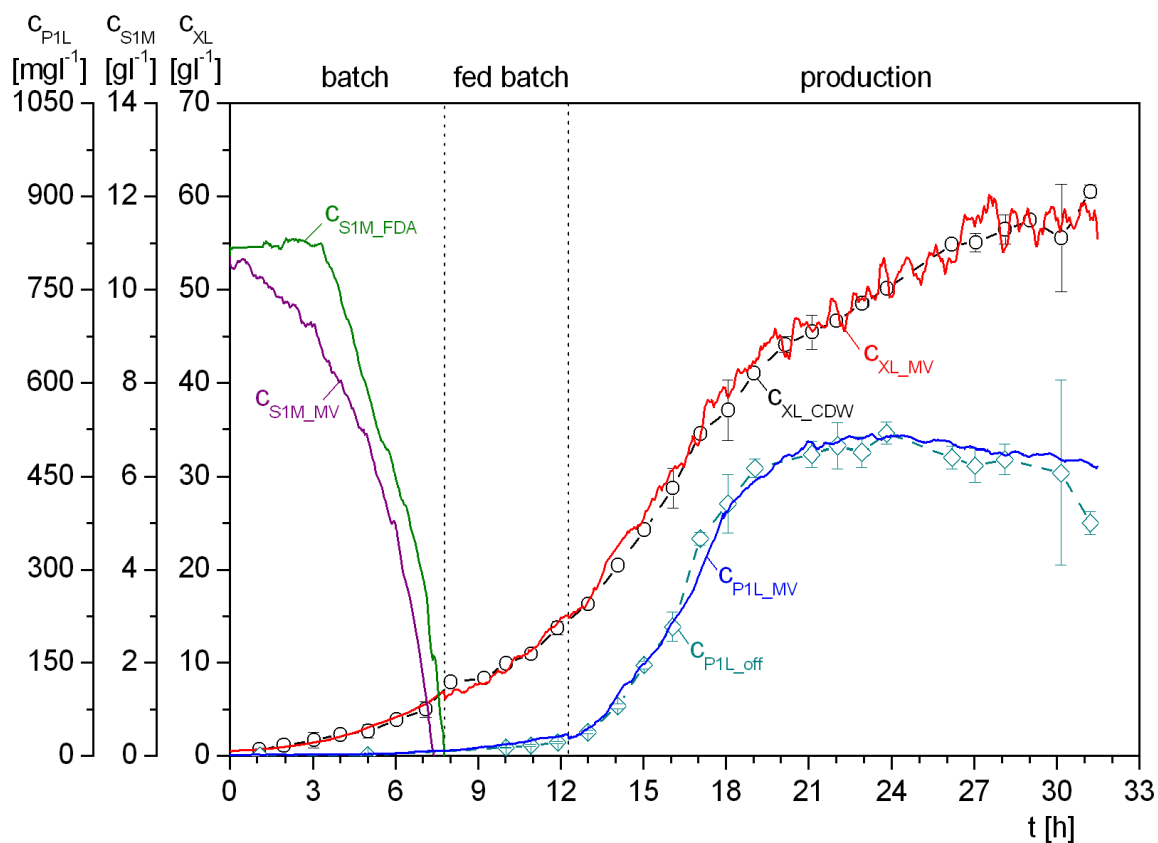


Bild 5.37: Multivariate Modelle auf Basis der BioView[®]-Originaldaten für den Validierungsdatensatz CK0707

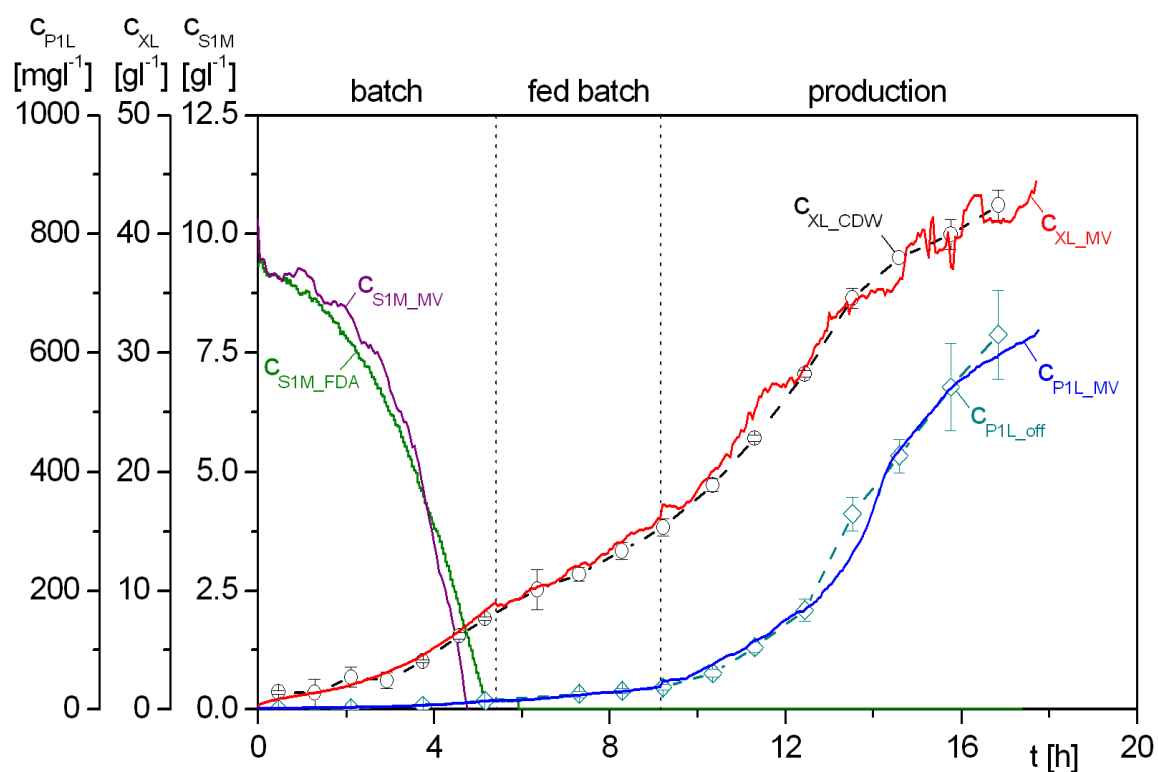


Bild 5.38: Multivariate Modelle auf Basis der BioView[®]-Originaldaten für den Validierungsdatensatz CK2208

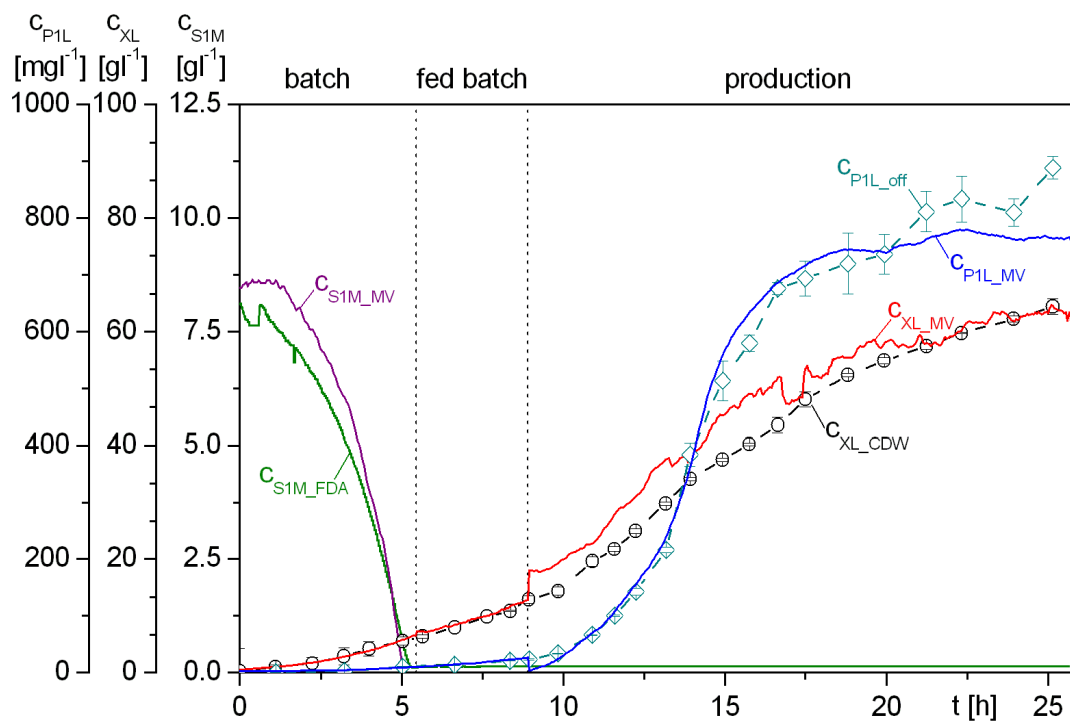


Bild 5.39: Multivariate Modelle auf Basis der BioView[®]-Originaldaten für den Trainingsdatensatz CK4407

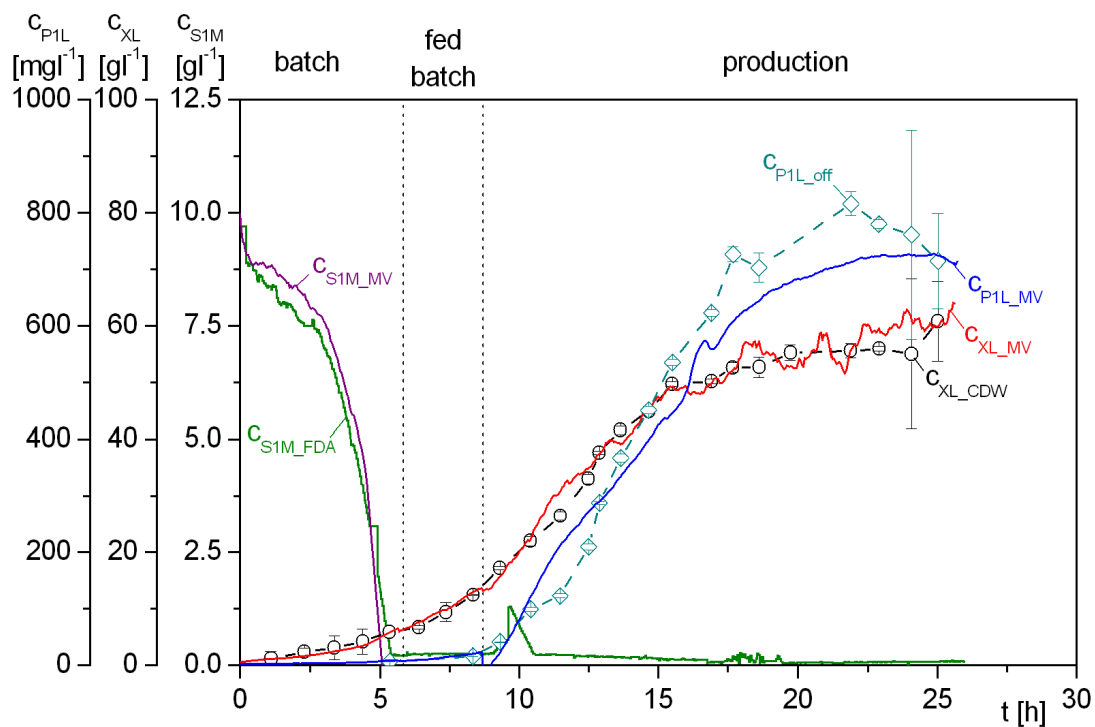


Bild 5.40: Multivariate Modelle auf Basis der BioView[®]-Originaldaten für den Trainingsdatensatz CK4406

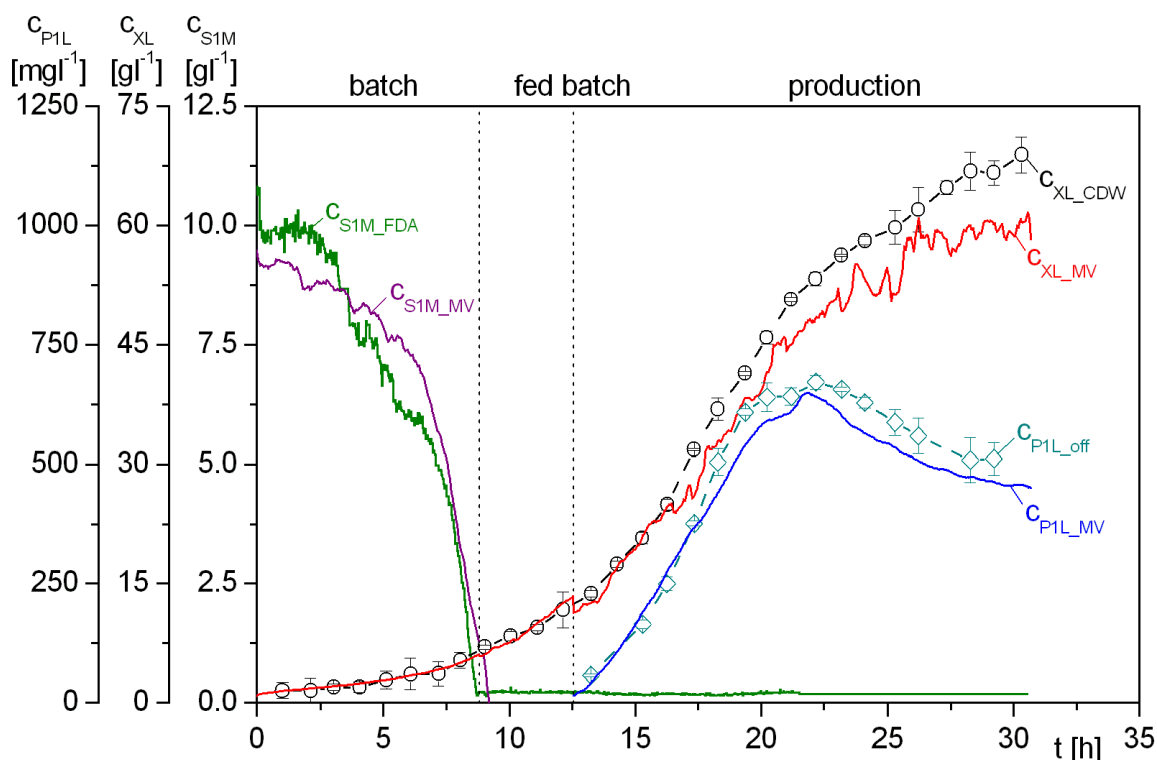


Bild 5.41: Multivariate Modelle auf Basis der BioView[®]-Originaldaten für den Trainingsdatensatz CK0607

5.6 Nichtlineare Modellierung mit künstlichen neuronalen Netzen

Als neuronale Netze werden informationsverarbeitende Systeme bezeichnet, die aus einer großen Anzahl einfach aufgebauter Verarbeitungseinheiten, den sogenannten Neuronen bestehen, die sich Signale über gerichtete Verbindungen zusenden [Zell (1997)]. Aufbau und Funktion sind dabei den biologischen Entsprechungen nachempfunden.

Das erste Modell für Neuronen wurde 1943 von W.S. McCulloch und W. Pitts entworfen [Pitts and McCulloch (1947)]. Heutzutage werden neuronale Netze fachübergreifend für die unterschiedlichsten Applikationen eingesetzt und stellen ein effektives Werkzeug zur Modellierung komplexer Probleme dar [Liao and Wen (2007)]. Diese reichen von der Prozessüberwachung, -diagnose und -steuerung über den Einsatz in Autopiloten von Flugzeugen bis hin zur Prognose von Börsenkursen [Widrow *et al.* (1988)].

Auf Grund der Sättigungseffekte, die bei höheren Zelldichten in den Fluoreszenzspektren auftraten und die lineare Modellierung erschwerten, wurde hier auf ANNs¹⁸ zurückgegriffen und damit eine verbesserte nichtlineare Anpassung der 2D-Fluoreszenzmessungen an die relevanten Prozessgrößen zu erreichen. Wegen der großen Vernetztheit neuronaler Netze sowie des nichtlinearen Charakters, der im Verlauf dieses Abschnittes näher erläutert wird, gelten ANNs als Universalapproximatoren, mit denen nahezu jede Funktion angepasst werden kann.

Die Struktur künstlicher neuronaler Netze ist zwar an das biologische Vorbild angelehnt, stellt aber meistens eine starke Vereinfachung der biologischen Nervensysteme dar. Die Anzahl

¹⁸ANN – Artificial Neural Networks

der Neuronen sowie die Verbindungen in KNNs¹⁹ ist oftmals um ein Vielfaches geringer, als in ihren biologischen Entsprechungen.

Chemische Interaktionen zwischen räumlich benachbarten Neuronen, durch die sie als Gruppe gehemmt oder angeregt werden können, finden häufig keine Berücksichtigung. Die genauen zeitlichen und chemischen Abläufe in den Axonen und Synapsen werden ebenfalls nicht modelliert.

Da künstliche neuronale Netze immer problembezogen konstruiert werden, ist die genaue Nachbildung für die Lösung technischer und naturwissenschaftlicher Problemstellung auch meistens nicht notwendig [Otto (1997)].

Von dem biologischen Vorbild bleibt die hohe Parallelität der Neuronen erhalten. Diese verarbeiten die Aktivierung der Vorgängerneuronen und die Stärke der Verbindungen zu einer Ausgabe. Durch die Modifikation der Verbindungsgewichte wird das Netz während des Trainings an die vorgegebenen Trainingsdatensätze adaptiert, was der Plastizität der Synapsen im biologischen Vorbild entspricht.

5.6.1 Aufbau neuronaler Netze

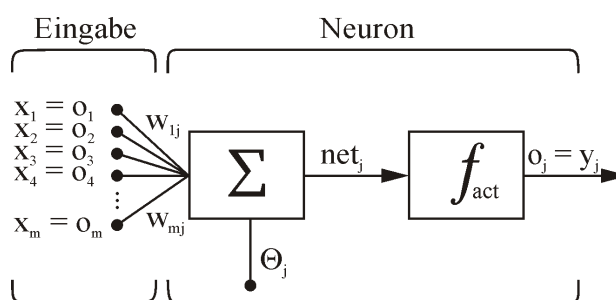


Bild 5.42: Aufbau eines einzelnen Neurons

Neuronale Netze lassen sich durch ihre Neuronen, die Gewichte w (*weights*), die Propagierungsfunktion net sowie die angewendete Lernregel vollständig charakterisieren [Danzer *et al.* (2001)].

Bild 5.42 zeigt das Modell eines einzelnen Neurons j . Es empfängt von m weiteren Neuronen die Eingabesignale, die den Vektor $X^{[0]}$ bzw. $O^{[0]}$ ergeben. Dabei enthalten diese die einzelnen Ausgabewerte der m vorgelagerten Neuronen $o_1 \dots o_m$ bzw. $x_1 \dots x_m$. Dieser wird mit den Gewichten w_{ij} des Neurons multipliziert und unter Berücksichtigung des Bias²⁰ Θ_j , durch den der Wertebereich der Ausgabe modifiziert wird, ergibt sich die eigentliche Netzeingabe net_j ,

$$net_j(t) = \sum_{i=1}^m (o_i(t) \cdot w_{ij}) + \Theta_j . \quad (5.45)$$

Bevor das aggregierte Signal das Neuron wieder verlässt, erfolgt eine Transformation mit Hilfe einer Aktivierungsfunktion f_{act} zum Ausgangssignal o_j bzw. y_j ,

$$y_j = o_j = f(net_j) . \quad (5.46)$$

¹⁹KNN – Künstliche Neuronale Netze

²⁰Bias – Schwellenwert

Wichtige Aktivierungsfunktionen sind in Tabelle 5.9 dargestellt. Wenn für das Netztraining (Abschnitt 5.6.2) Gradientenabstiegsverfahren verwendet werden, müssen die verwendeten Funktionen differenzierbar sein.

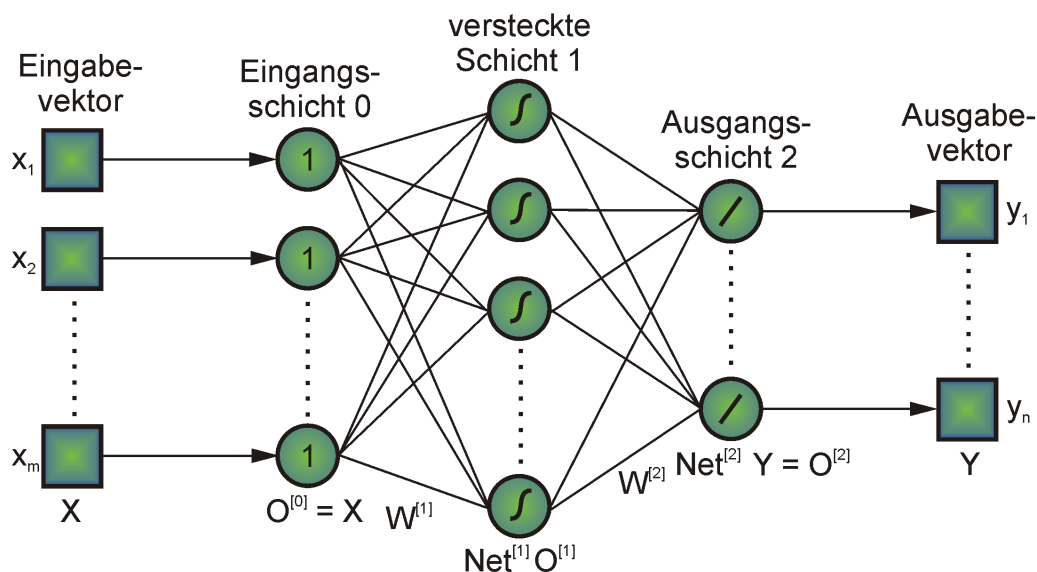


Bild 5.43: Schema eines ebenenweise vollständig verbundenen Feedforward-Netztes

Der Aufbau neuronaler Netze erfolgt in Schichten, die hier mit $0, 1, 2, \dots, s$ bezeichnet werden. Ein dreischichtiges Netz ist in Bild 5.43 schematisch dargestellt.

Die Anzahl der Neuronen in der Eingangsschicht entspricht der Menge der Elemente des Eingabevektors. Im konkreten Fall sind das die zum Netztraining verwendeten Hauptkomponenten bzw. unabhängigen Komponenten. Die Neuronen der Eingangsschicht leiten die Daten an jedes Neuron der nachfolgenden verdeckten Schicht weiter. Dabei wird in den meisten Fällen die Identität als Transferfunktion der Eingangsschicht verwendet, so dass die Eingangsdaten nicht modifiziert werden.

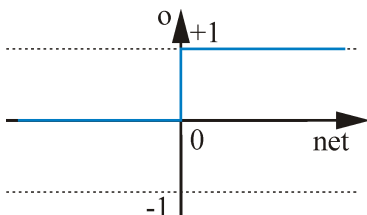
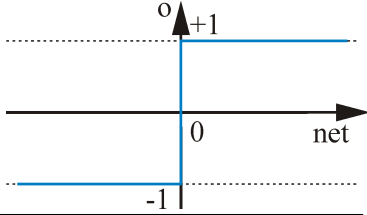
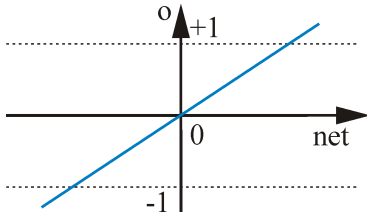
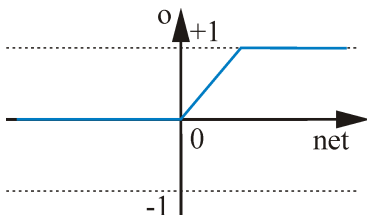
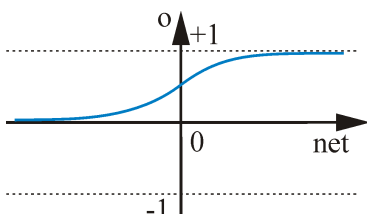
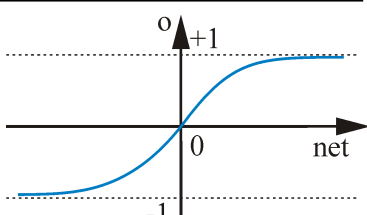
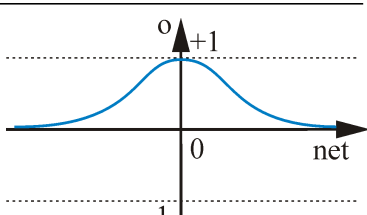
Die Verbindungsgewichte zwischen den einzelnen Schichten sind in den Matrizen $W^{[1]}$ und $W^{[2]}$ zusammengefasst. Sie enthalten die Verbindungsgewichte w_{ij} (Schicht 1) bzw. w_{jk} (Schicht 2) für jedes Neuron. Dabei entspricht die erste Laufvariable i bzw. j der Anzahl an vorgelagerten Neuronen und die zweite (j bzw. k) der Anzahl an Neuronen in der folgenden Schicht. Bei drei Neuronen in der Eingangsschicht und vier Neuronen in der verdeckten Schicht ergibt sich dabei eine 3×4 Matrix für $W^{[1]}$. Die entsprechenden Verbindungsgewichte w_{ij} geben die Stärke der jeweiligen Verknüpfung vor.

Die Aktivierungsfunktionen net sowie die Ausgänge o der einzelnen Neuronen einer Schicht werden in den zugehörigen Vektoren Net bzw. O zusammengefasst.

Die Neuronen der Ausgangsschicht haben die Aufgabe, die Ausgabe der verdeckten Schicht zusammenzufassen und an den Ausgabevektor weiterzuleiten. Die Anzahl der Neuronen in der Ausgangsschicht ist durch die Größe des Zielvektors festgelegt. In den allermeisten Fällen werden für die verdeckte Schicht sigmoidale Aktivierungsfunktionen und für die Ausgangsschicht lineare Transferfunktionen eingesetzt [Rojas (1996)], so auch in dieser Arbeit.

Theoretisch ist die Anzahl der Schichten unbegrenzt, jedoch sind für die meisten Lösungen

Tabelle 5.9: Aktivierungsfunktionen für neuronale Netze

| Übertragungsfunktion | 0. Ableitung | Funktionsverlauf |
|---------------------------------|---|---|
| Schwellenwertlogik (binär) | $o = \begin{cases} 1 & \text{für } \text{net} \geq 0 \\ 0 & \text{für } \text{net} < 0 \end{cases}$ |  |
| Schwellenwertlogik (bipolar) | $o = \begin{cases} 1 & \text{für } \text{net} \geq 0 \\ -1 & \text{für } \text{net} < 0 \end{cases}$ |  |
| lineare Transferfunktion | $o = c \cdot \text{net}$ |  |
| lineare Schwellenfunktion | $o = \begin{cases} 1 & \text{für } c \cdot \text{net} \geq 1 \\ 0 & \text{für } c \cdot \text{net} < 0 \\ c \cdot \text{net} & \text{für } 0 \leq c \cdot \text{net} < 1 \end{cases}$ |  |
| Sigmoidfunktion | $o = \frac{1}{1 + e^{c \cdot \text{net}}}$ |  |
| Tangenshyperbolicus | $o = \frac{e^{c \cdot \text{net}} - e^{-c \cdot \text{net}}}{e^{c \cdot \text{net}} + e^{-c \cdot \text{net}}}$ |  |
| Radial Bias Funktion | $o = e^{-\frac{\text{net}^2}{r^2}}$ |  |

von Klassifikations- und Kalibrationsaufgaben mehrschichtige Perceptrons²¹ mit einer verdeckten Schicht ausreichend [Danzer *et al.* (2001)].

Bei dem hier gezeigten Netz handelt es sich um ein ebenenweise vollständig verbundenes Netz. Jedes Neuron der vorhergehenden Schicht ist mit jedem Neuron der darauffolgenden Schicht verbunden. Da kein Pfad existiert, der von einem Neuron direkt oder über zwischengeschaltete Neuronen wieder zurück zu diesem Neuron führt, spricht man von einem rückkopplungsfreien²²-Netz.

Neben dem in Bild 5.43 dargestellten Netz wurden noch Netze mit sogenannten *shortcut connections* eingesetzt, welche Verbindungen zwischen Neuronen besitzen, die Ebenen überspringen. Da diese aber zu keiner wesentlichen Verbesserung der Ergebnisse führten, wurde im weiteren Verlauf dieser Arbeit darauf verzichtet.

Es gibt noch eine Reihe weiterer Netzarchitekturen, wie z.B. rekurrente Netze mit direkter Rückkopplung, die für die Prozessvorhersage eingesetzt werden können. Die Neuronen besitzen dabei eine Verbindung von ihrer Ausgabe zurück zur Eingabe, so dass sie ihre eigene Aktivierung verstärken oder abschwächen. Auf Grund der Anzahl der zur Verfügung stehenden Datensätze wurde jedoch vom Einsatz rekurrenter Netze abgesehen, da diese für eine zuverlässige Prozessvorhersage mehr Trainingsdatensätze benötigen.

5.6.2 Training neuronaler Netze

Damit ein neuronales Netz für eine gegebene Aufgabenstellung eingesetzt werden kann, muss es über eine fest vorgegebene Vorschrift modifiziert/trainiert werden. Grundsätzlich gibt es diverse Methoden um ein Netz anhand von bestehenden Trainingsdatensätzen an ein gegebenes Problem zu adaptieren [Rojas (1996)].

- Entwicklung neuer Verbindungen
- Löschen existierender Verbindungen
- Modifikation der Verbindungsstärke (Veränderung der Gewichte)
- Modifikation des Schwellenwertes
- Modifikation der Aktivierungs- bzw. Ausgabefunktion
- Entwicklung neuer Zellen
- Löschen bestehender Zellen

Die meisten Algorithmen zielen auf eine Modifikation der Verbindungsstärke sowie der Schwellenwerte ab. Dies bedeutet, dass zusätzliche Algorithmen implementiert werden müssen, um die optimale Netzwerkarchitektur sowie geeignete Aktivierungs- und Ausgabefunktionen aufzufinden.

Grundsätzlich wird zwischen unüberwachtem, überwachtem und bestärkendem Lernen unterschieden [Otto (1997)].

²¹MLP – *Multi Layer Perceptrons*

²²*feedforward*

Beim unüberwachten Lernen werden dem Netz lediglich die Eingangsdaten zur Verfügung gestellt, jedoch keine Zieldaten. Dieses Verfahren wird z.B. in der Bilderkennung eingesetzt. Durch die Klassifikation der Eingangsdaten lassen sich unbekannte Objekte automatisch erkennen. Ein typisches Beispiel stellen die selbsorganisierenden Karten²³ von Kohonen dar [Kohonen (1997)].

Stehen sowohl Eingangs- als auch Zieldaten zur Verfügung kann die Methode des überwachten Lernens eingesetzt werden. Anhand der Abweichung des Netzausgangs von den tatsächlichen Zieldaten wird das Netz über die angewendete Lernregel so lange modifiziert, bis ein vorgegebenes Abbruchkriterium erfüllt ist. Ein typisches überwachtetes Lernverfahren stellt der Backpropagation-Algorithmus dar, der in Abschnitt 5.6.3 näher erläutert wird.

Im Gegensatz zum überwachten Lernen wird beim bestärkendem Lernen dem Netz lediglich mitgeteilt, ob seine Ausgabe korrekt oder inkorrekt ist. Das Netz erfährt nicht den exakten Wert des Fehlers. Diese Methode stellt zwar ein biologisch plausibleres Verfahren dar, ist aber oft deutlich langsamer als das überwachte Lernen [Otto (1997)].

In dieser Arbeit wurde ausschließlich die Methode des überwachten Lernens eingesetzt. Um die Leistungsfähigkeit der neuronalen Netze während des Trainings beurteilen zu können, wurden die zur Verfügung stehenden Daten, die auch schon für die uni- und multivariaten Modelle verwendet wurden, zunächst in einen Trainingsdatensatz und einen Datensatz zur externen Validierung aufgeteilt.

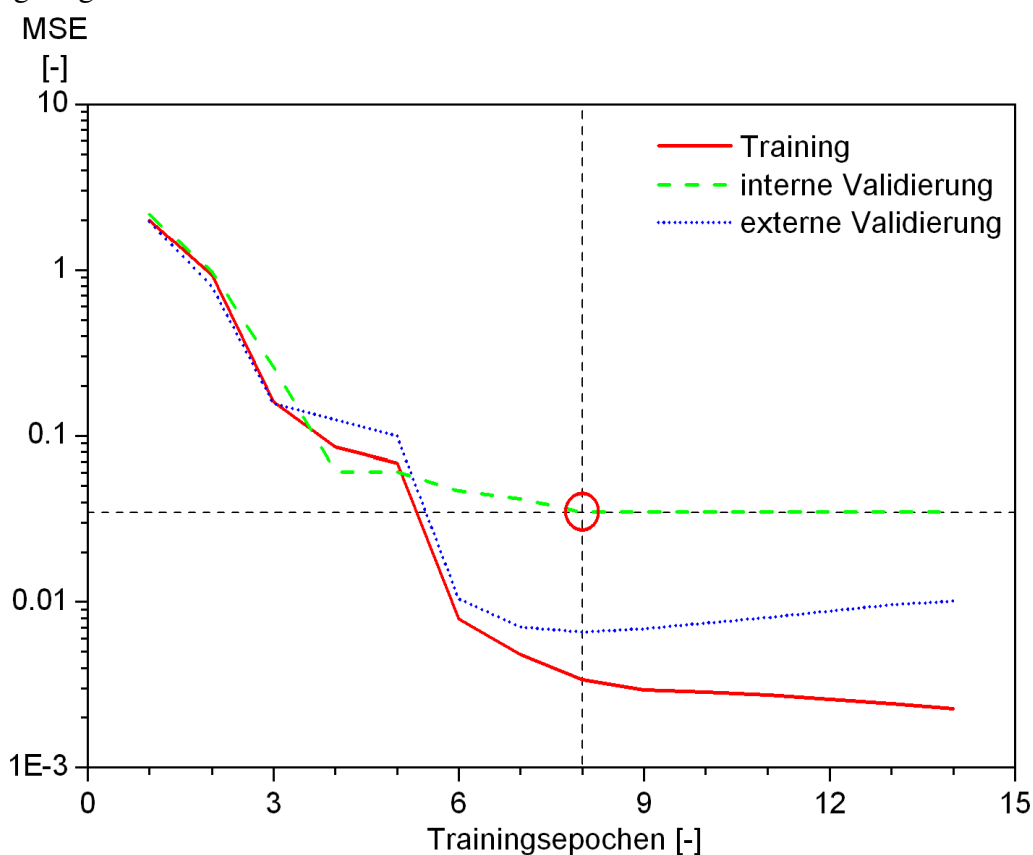


Bild 5.44: Verlauf des MSE während des Netztrainings für Trainings-, interne sowie externe Validierungsdaten

²³SOM – *Self Organizing Maps*

Die Trainingsdaten wurde noch einmal durch die MATLAB[®]-Funktion *dividevec* unterteilt, wobei die Eingangs- und Zieldaten nach dem Zufallsprinzip aufgeteilt werden. Für das Training der Netze werden 60% der Daten verwendet und über diese der Lernfehler bestimmt. Die restliche 40% des Trainingsdatenraumes werden für die interne Validierung des Netzes verwendet. Mittels dieses Datensatzes wird während des Netztrainings die Generalisierungsfähigkeit des Netzes überprüft.

Sobald der interne Validierungsfehler innerhalb der letzten sechs Iterationsschritte nicht mehr sinkt, wird das Netztraining abgebrochen und das erste Netz mit dem geringsten Validierungsfehler gespeichert. Die sich durch diese Aufteilung ergebenden Fehler eines Trainingsdurchlaufes sind exemplarisch in Bild 5.44 dargestellt. Die besten Netzparameter wurde dabei nach acht Trainingsepochen gefunden.

Über die externen Validierungsdaten kann dann die Approximations- und Generalisierungsfähigkeit des vollständig trainierten Netzes überprüft werden. Der externe Validierungsfehler kann als eigentliches Maß für die Güte des neuronalen Netzes angesehen werden.

Die Gütebeurteilung erfolgt in MATLAB[®] über den MSE²⁴,

$$\text{MSE} = \sum_{i=1}^n (c_i^* - c_i)^2, \quad (5.47)$$

der sich aus der Summe der quadratischen Abweichungen der tatsächlichen Konzentration c_i von der vorhergesagten Konzentration c_i^* ergibt.

5.6.3 Lernalgorithmen

Für die sinnvolle Änderung der Netzparameter während des Lernvorgangs werden geeignete Algorithmen benötigt. Die meisten Vorgaben gehen auf die 1949 von Hebb formulierte Hebb'sche Lernregel zurück:

„Wenn ein Axon der Zelle A [...] Zelle B erregt und wiederholt und dauerhaft zur Erzeugung von Aktionspotenzialen in Zelle B beiträgt, so resultiert dies in Wachstumsprozessen oder metabolischen Veränderungen in einer oder in beiden Zellen, die bewirken, dass die Effizienz von Zelle A in Bezug auf die Erzeugung eines Aktionspotenzials in B größer wird.“

Dieses konnte Hebb anhand von Veränderungen der synaptischen Übertragung zwischen Neuronen nachweisen [Hebb (1949)] und gilt damit als Entdecker der synaptischen Plastizität.

Mathematisch kann die Regel für die Aktivierung eines einzelnen Neurons durch die Eingabe w_{ij} ,

$$w_{ij}(t+1) = w_{ij}(t) + \eta \cdot x_i \cdot o_j \Leftrightarrow \Delta w_{ij} = \eta \cdot x_i \cdot o_j, \quad (5.48)$$

beschrieben werden, wobei sich die Änderung des Gewichtes Δw_{ij} aus dem Produkt der Lernrate η mit der Eingabe x_i und der Ausgabe o_j des Neurons berechnen läßt. Diese Lernregel kann bei einfachen Eingangs-Ausgangsneuronen eingesetzt werden, die binäre bzw. diskrete Übertragungsfunktionen beinhalten.

²⁴MSE – Mean Square Error

Beim Einsatz mehrschichtiger Netze mit nichtlinearen Aktivierungsfunktionen wird häufig der Backpropagation-Algorithmus angewendet, der auf der generalisierten Delta-Regel beruht. Dieser stellt eine Spezialisierung der Hebbischen-Regel dar [Rojas (1996)].

Der Gesamtfehler E eines Netzes in Abhängigkeit von der gewählten Gewichtsmatrix $W^{[2]}$,

$$E(W^{[2]}) = \sum_k e_k(w_{jk}) = \frac{1}{2} \sum_k (t_k - o_k)^2, \quad (5.49)$$

kann über die Summe der Fehlerquadrate der Ausgabeschichten k berechnet werden. Dabei stellt t_k den tatsächliche Wert und o_k die Ausgabe des Netzes dar.

Die Änderung der Gewichtsmatrix,

$$\nabla W^{[2]} = -\eta \cdot \nabla E(W^{[2]}), \quad (5.50)$$

ergibt sich aus dem Produkt des Gradienten des Netzfehlers und der Lernrate.

Die Änderung eines einzelnen Gewichtes w_{jk} ,

$$\Delta w_{jk} = -\eta \cdot \frac{\partial E(W^{[2]})}{\partial w_{jk}}, \quad (5.51)$$

kann aus der ersten partiellen Ableitung der Fehlerfunktion nach dem entsprechenden Gewicht berechnet werden.

Unter Verwendung der Kettenregel,

$$\frac{\partial E(W^{[2]})}{\partial w_{jk}} = \frac{\partial E(W^{[2]})}{\partial \text{net}_k} \cdot \frac{\partial \text{net}_k}{\partial w_{jk}}, \quad (5.52)$$

kann Gl. (5.51) so erweitert werden, dass sich für den ersten Faktor aus Gl. (5.52) der lokale Fehler δ_k ,

$$\delta_k = -\frac{\partial E(W_{jk})}{\partial o_k} \cdot \frac{\partial o_k}{\partial \text{net}_k}, \quad (5.53)$$

bei erneuter Anwendung der Kettenregel einführen lässt.

Die Änderung des Fehlers in Abhängigkeit des Netzausgangs,

$$-\frac{\partial E(W^{[2]})}{\partial o_k} = -\frac{\partial}{\partial o_k} \frac{1}{2} \sum_k (t_k - o_k)^2 = (t_k - o_k), \quad (5.54)$$

wird dann unter Verwendung von Gl. (5.49) formuliert.

Da die erste zeitliche Ableitung der Aktivierungsfunktion f_{act} ,

$$\frac{\partial o_k}{\partial \text{net}_k} = f'_{\text{act}}, \quad (5.55)$$

die Änderung des Netzausgangs in Abhängigkeit des Netzeingangs darstellt, vereinfacht sich Gl. (5.53) zu

$$\delta_k = (t_k - o_k) \cdot f'_{\text{act}}. \quad (5.56)$$

Der zweite Faktor aus Gl. (5.52), die Änderung des Netzeingangs in Abhängigkeit der Gewichtsänderung,

$$\frac{\partial \text{net}_k}{\partial w_{jk}} = \frac{\partial}{\partial w_{jk}} \sum_j ((o_j(t)w_{jk}) + \Theta_i) = o_j, \quad (5.57)$$

entspricht unter Berücksichtigung von Gl. (5.45) dem Ausgang des vorgelagerten Neurons.

Damit kann Gl. (5.51) zur verallgemeinerten Delta-Regel,

$$\Delta w_{jk} = \eta \cdot \delta_k \cdot o_j, \quad (5.58)$$

umgestellt werden. Unter Berücksichtigung von Gl. (5.56) für den lokalen Fehler ergibt sich dann die Änderung für die Gewichte der Ausgangsschicht,

$$\Delta w_{jk} = \eta \cdot (t_k - o_k) \cdot f'_{\text{act}} \cdot o_j. \quad (5.59)$$

Die Änderung der Gewichte für die verdeckte Schicht,

$$\Delta w_{ij} = \eta \cdot \sum_k (\delta_k \cdot w_{jk}) \cdot f'_{\text{act}} \cdot o_i, \quad (5.60)$$

ergibt sich bei gleichem Vorgehen wie für die Ausgangsschicht. Diese Herleitung des Backpropagation Algorithmus ist angelehnt an Pao (1989).

Wie alle Gradientenabstiegsverfahren beinhaltet der Backpropagation-Algorithmus eine Reihe von Problemen. Dieses sind das Verweilen in lokalen Minima, flache Plateaus, in denen der Gradient sehr klein wird oder das Verfahren stagnieren kann, direkte und indirekte Oszillation in steilen Schluchten und das Verlassen guter Minima.

Durch die Variation der Lernrate η kann auf diese Effekte Einfluss genommen werden. Oszillation und das Verlassen guter Minima können durch die Wahl einer kleinen Lernrate vermieden werden. Dadurch wird jedoch nicht nur die Trainingszeit erhöht, sondern auch das Verweilen in lokalen Minima und die Stagnation des Verfahrens wird bei kleinen Gradienten noch verstärkt [Universität Münster (2007)].

Da es sich bei dem Training neuronaler Netze um ein allgemeines Parameteroptimierungsproblem handelt, können eine Vielzahl von Optimierungsmethoden aus der numerischen Mathematik eingesetzt werden, die zu einer erheblichen Verbesserung hinsichtlich des Konvergenzgebietes, der Konvergenzgeschwindigkeit, der Stabilität und des Berechnungsaufwandes führen können. Da die Ergebnisse von Fall zu Fall sehr unterschiedlich sein können, lässt sich keine allgemeingültige Aussage darüber machen, welches Verfahren für den entsprechenden Anwendungsfall am geeignetsten ist. Es wurden verschiedene Algorithmen, die unter MATLAB[®] zur Verfügung stehen, angewendet und die Ergebnisse miteinander verglichen.

Die Verfahren sowie die zur Verfügung stehenden MATLAB[®]-Funktionen lassen sich unterteilen in

- Heuristische Verbesserung des Gradientenalgorithmus
 - Gradientenalgorithmus mit Momentum-Term
 - Verwendung einer adaptiven Lernrate: *traingda*, *traingdx*, *trainscg*
- Konjugierter Gradientenalgorithmus: *traincgf*, *traincgp*, *traincgb*, *trainscg*
- Newton und Quasi-Newton Algorithmen
 - Quickprop: *trainbfg*, *trainoff*
 - Levenberg-Marquardt Algorithmus: *trainlm*
- Pruning-Techniken

Von den zur Verfügung stehenden Verfahren wurden die Funktionen *trainlm*, *trainrp*, *trainscg*, *traincgb*, *traingdx* und *trainoss* vorab für das bestehende Problem auf ihre Anwendbarkeit für das gegebene Problem getestet. Der konjugierte Gradientenalgorithmus der MATLAB[®]-Funktion *trainscg* lieferte dabei die geringsten Fehler und wurde im Folgenden zum Training der neuronalen Netze verwendet. Eine nähere Beschreibung der Verfahren findet sich bei Demuth *et al.* (2008) und Rojas (1996).

5.6.4 Ergebnisse der nichtlinearen Modelle mit ANNs

Für das Training der in dieser Arbeit verwendeten Feed-Forward Netze wurden als Eingangsdaten die Scores der Hauptkomponentenanalyse bzw. die unabhängigen Komponenten der ICA eingesetzt. Es wurden 1 – 20 Zwischenschichtneuronen verwendet und für jede sich so ergebende Netzwerkarchitektur jeweils 100 Netze trainiert. Da die Anfangswerte für die Gewichte der Netze willkürlich vom Trainingsalgorithmus gewählt werden, die Fehleroberfläche des Suchraumes jedoch zerklüftet sein kann, wird so eine gewisse statistische Sicherheit gewährt, um gute Minima zu erreichen. Das jeweils beste Netz wird von MATLAB[®] gespeichert und kann anschließend zur Simulation der Prozesse verwendet werden

Die Prozessverläufe der besten auf Basis der ICA ermittelten Netze sind in Bild 5.45, die der PCA in Bild 5.46 dargestellt. Die Ergebnisse für einen weiteren Validierungsdatensatz sind in Bild 5.47 und Bild 5.48 aufgezeigt. Es ergaben sich bei der Verwendung von Hauptkomponenten ähnliche Fehler wie bei den mit unabhängigen Komponenten trainierten Netzen. Die Fehler für beide Methoden sowie die Größe des Eingangsdatenraumes und die Anzahl der verwendeten Zwischenschichtneuronen sind in Tabelle 5.10 und Tabelle 5.11 aufgeführt.

Um zufriedenstellende Ergebnisse hinsichtlich der Anpassung der Prozessdaten zu erhalten, mussten auch bei der Verwendung von ANNs je zwei Netze erstellt werden. Dabei erfolgte die Anpassung zum Einen für die Batch und die Fed-Batch Phase und zum Anderen für die Produktionsphase. Die sich dabei ergebenden Fehler entsprachen in etwa den der unter Abschnitt 5.5.3 erstellten multivariaten Modelle. Der Vorteil, den die neuronalen Netze jedoch gegenüber den multivariaten Modellen aufweisen, liegt darin, dass sie bei einer ausreichend großen Anzahl an Trainingsdatensätzen auch für unterschiedliche Kultivierungsstrategien zuverlässig eingesetzt werden können. Dass die Netze nach jeder Kultivierung mit den neuen Daten weitertrainiert werden können, stellt einen weiteren großen Vorteil dar.

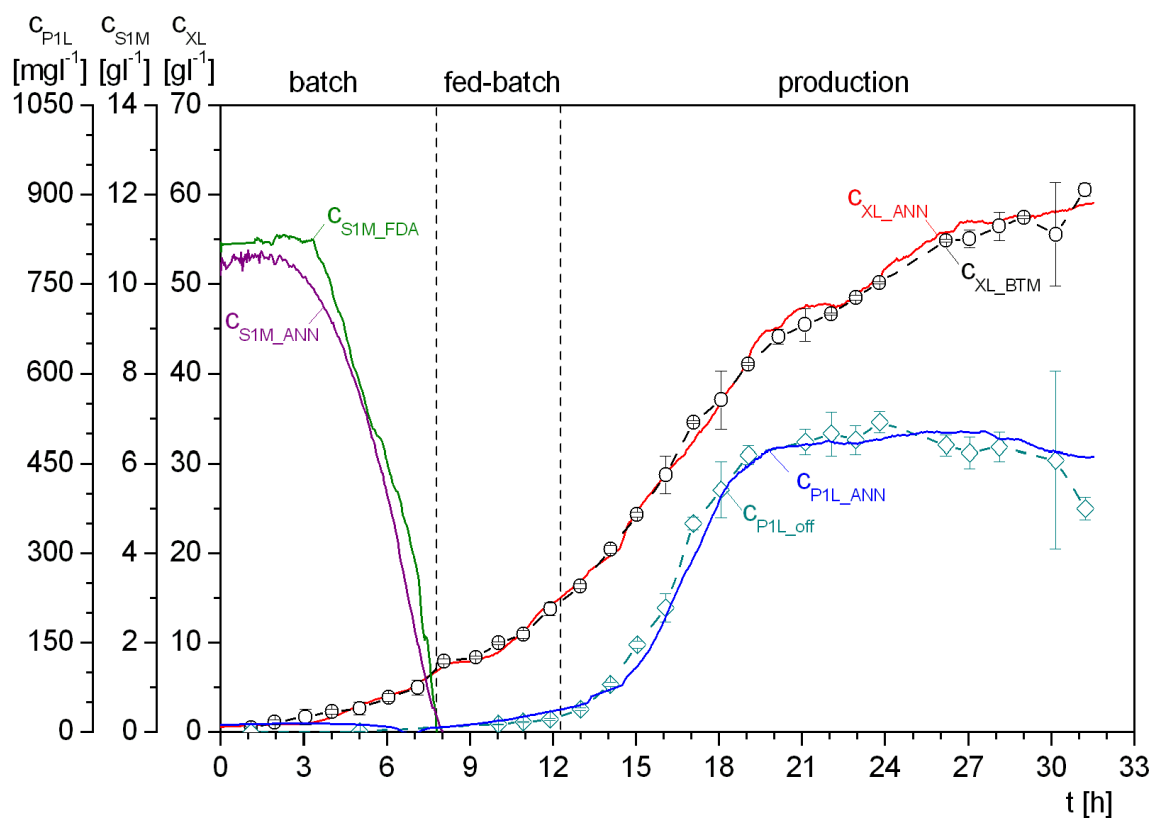


Bild 5.45: Verlauf der besten nichtlinearen Modelle auf Basis der ICA (CK0707)

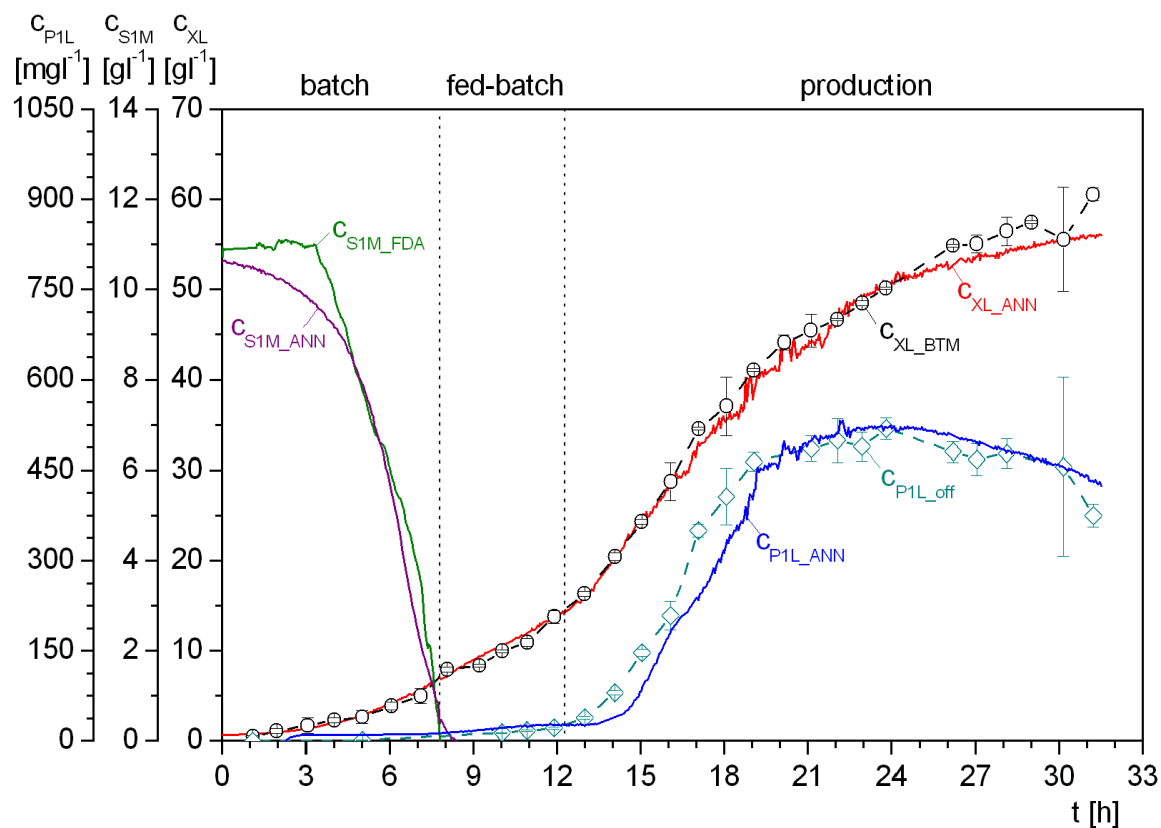


Bild 5.46: Verlauf der besten nichtlinearen Modelle auf Basis der PCA (CK0707)

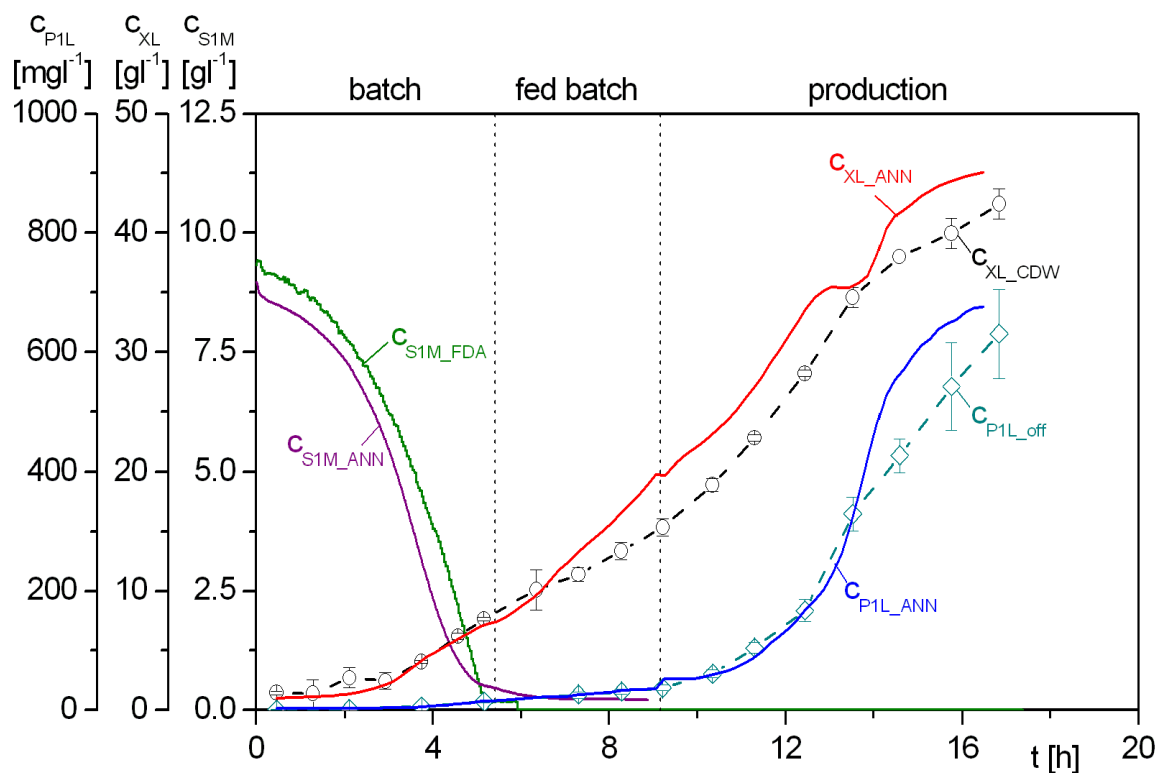


Bild 5.47: Verlauf der Ergebnisse der erstellten ANNs unter Verwendung von ICs für den Validierungsdatensatz CK2208

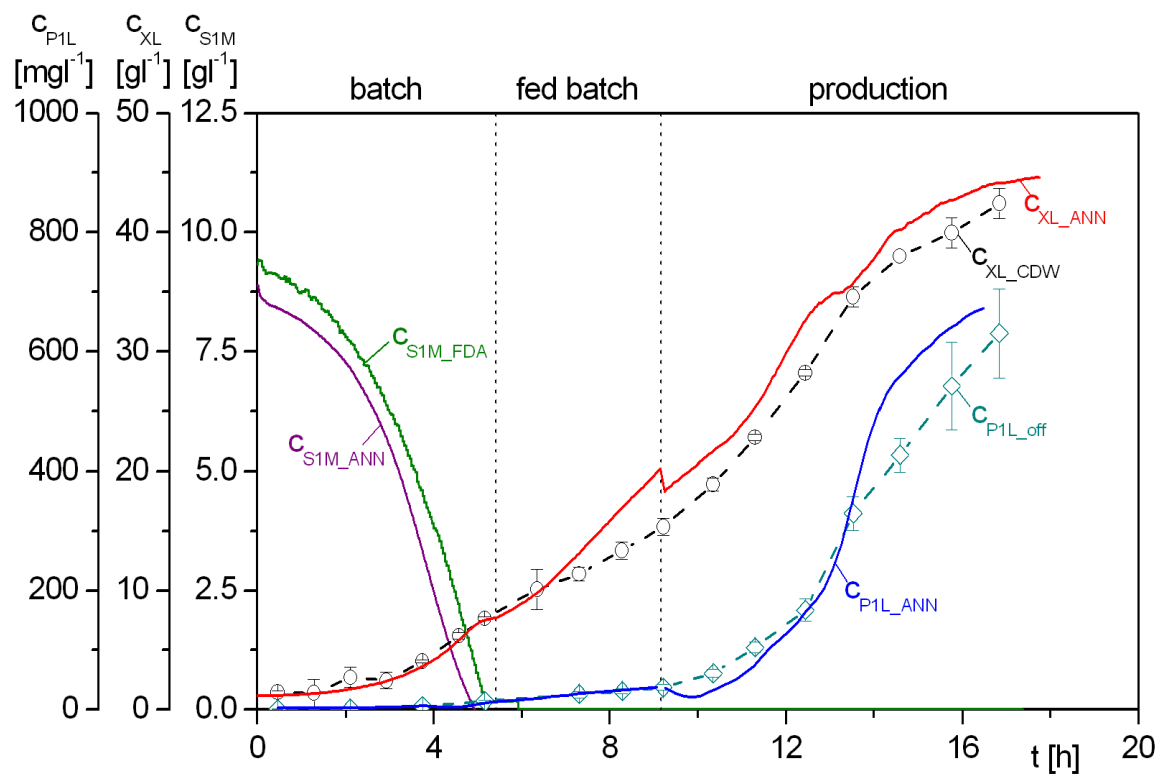


Bild 5.48: Verlauf der Ergebnisse der erstellten ANNs unter Verwendung von PCs für den Validierungsdatensatz CK2208

Tabelle 5.10: Fehler der besten ANNs auf Basis der ICA

| Daten | Anzahl Eingangs- größen | Anzahl ZS- Neuronen | Phase | RMSEC | | RMSEP | |
|------------------|-------------------------------|---------------------------|---------------------|-----------------------|------|-----------------------|------|
| | | | | [gl ⁻¹] | [%] | [gl ⁻¹] | [%] |
| c _{XL} | 3 | 3 | Batch/ Fed-Batch | 2.37 | 7.92 | 1.14 | 8.32 |
| | 3 | 8 | Produktion | 3.30 | 8.85 | 3.70 | 12.5 |
| c _{PIL} | 3 | 7 | Batch/ Fed-Batch | 7.14·10 ⁻³ | 7.83 | 9.21·10 ⁻³ | 9.21 |
| | 3 | 4 | Produktion | 85.9·10 ⁻³ | 9.91 | 50.5·10 ⁻³ | 8.50 |
| c _{SIM} | 3 | 1 | Batch | 0.814 | 8.85 | 1.06 | 9.54 |

Tabelle 5.11: Fehler der besten ANNs auf Basis der PCA

| Daten | Anzahl Eingangs- größen | Anzahl ZS- Neuronen | Phase | RMSEC | | RMSEP | |
|------------------|-------------------------------|---------------------------|---------------------|-----------------------|-------|-----------------------|------|
| | | | | [gl ⁻¹] | [%] | [gl ⁻¹] | [%] |
| c _{XL} | 3 | 4 | Batch/ Fed-Batch | 3.065 | 10.25 | 1.53 | 11.2 |
| | 3 | 12 | Produktion | 2.80 | 7.50 | 2.41 | 8.16 |
| c _{PIL} | 3 | 9 | Batch/ Fed-Batch | 6.18·10 ⁻³ | 6.78 | 2.33·10 ⁻³ | 9.38 |
| | 3 | 5 | Produktion | 87.8·10 ⁻³ | 10.1 | 56.6·10 ⁻³ | 9.60 |
| c _{SIM} | 3 | 1 | Batch | 0.894 | 7.76 | 0.855 | 8.59 |

6 Modellierung des dynamischen Bioreaktionsprozesses

Im Gegensatz zu den in Kapitel 5 vorgestellten datengetriebenen Modellen wird in diesem Kapitel ein dynamisches Modell zur Prozesssimulation verwendet. Es basiert auf den grundlegenden Konzepten der Energie- und Massenerhaltung, den Phasengleichgewichten sowie den postulierten Reaktions- und Wachstumskinetiken.

Die datengetriebenen Modelle stellen *Black-Box*-Verfahren dar, die hauptsächlich zur Prozessüberwachung dienen. Beim dynamischen Modell steht das Prinzip von Ursache und Wirkung im Vordergrund. Es wird dabei versucht die erzielten Messdaten über analytische Ansätze zu simulieren und damit ein besseres Verständnis der zugrunde liegenden Reaktionskinetiken und der Zielproteinexpression zu erreichen.

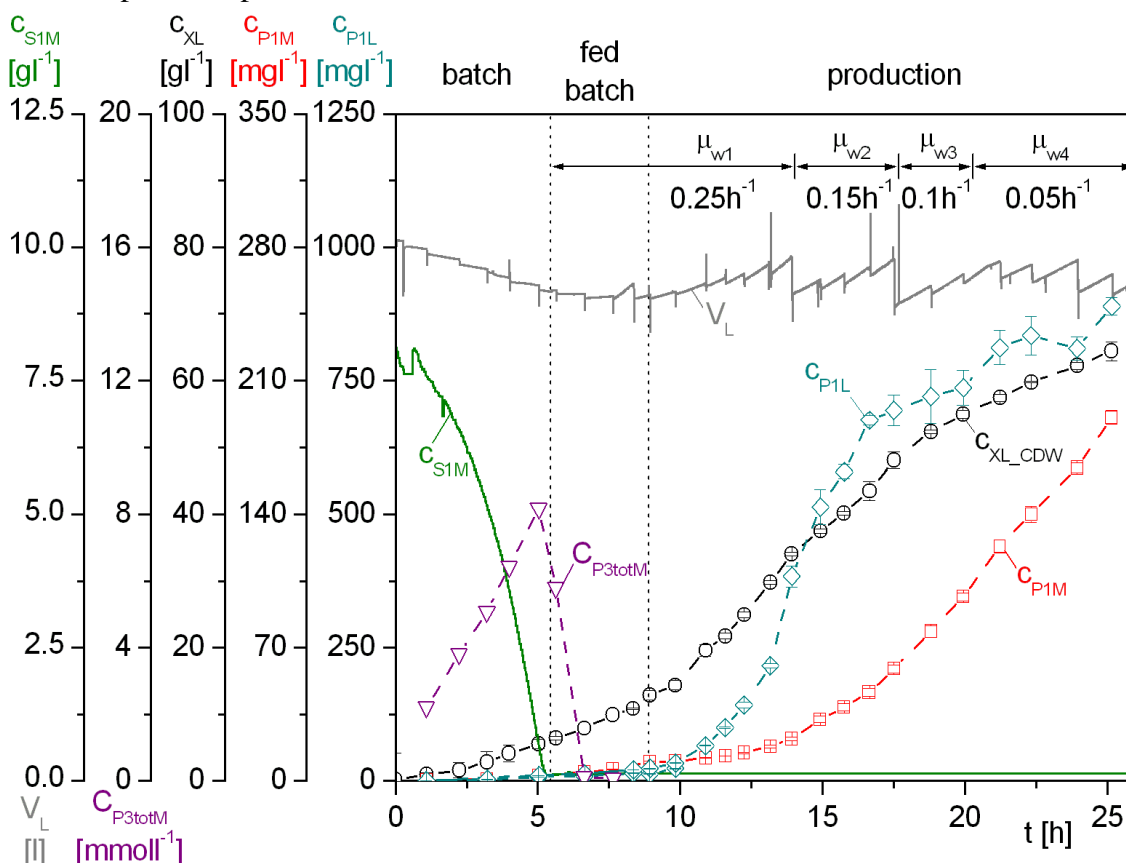


Bild 6.1: Exemplarische Darstellung des Reaktionsvolumens V_L , der Biotrockenmasse c_{XL} und der Zielproteinkonzentration c_{P1L} im Reaktionsvolumen sowie der Zielproteinkonzentration c_{P1M} , der Acetatkonzentration C_{P3totM} und der Glukosekonzentration c_{S1M} im Medienüberstand M (CK4407)

Exemplarisch ist in Bild 6.1 ein Kultivierungsverlauf dargestellt. Um neben der Zelldichte c_{XL} auch die relevanten Konzentrationen von Glukose c_{S1M} während der Batch-Phase sowie des Stoffwechselproduktes Acetat C_{P3totM} und der Gesamtzielproteinkonzentration c_{PIL} beschreiben zu können, muss zunächst das Volumen der Reaktionsphase L exakt simuliert werden. Dieses wird im Wesentlichen von der diskontinuierlichen Probenahme für die Offline-Analytik sowie dem kontinuierlichem zellfreien Analysenstrom für die Atline-Messsysteme und die einzelnen Zufüterraten beeinflusst.

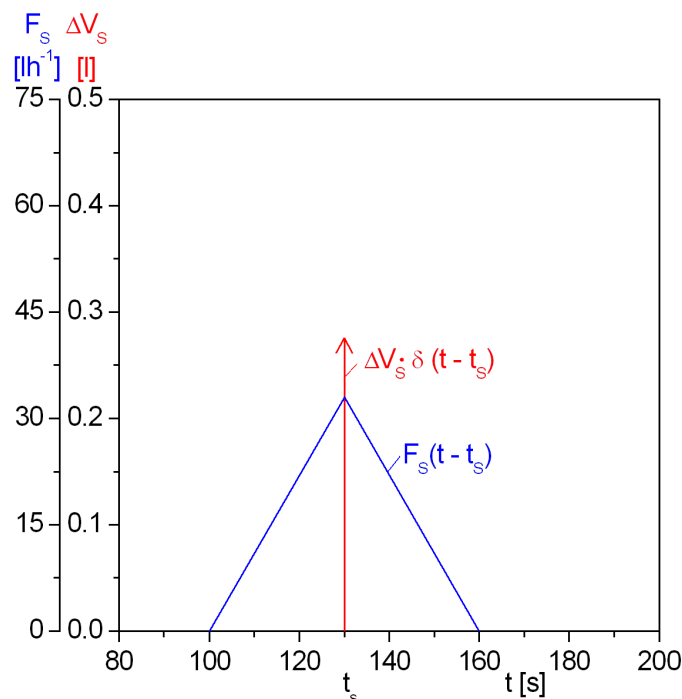


Bild 6.2: Exemplarische Darstellung des Probenahmevolumens ΔV_S und der fiktiven Probenahmerate F_S

nahmerate, wie in Bild 6.2 dargestellt, zu modellieren. Dieser Vorgang erstreckt sich über 60 Sekunden und geht als jeweilige Probenahmerate $F_S(t - t_s)$ in die Simulation ein.

Die Bildung von Acetat kann während der unlimitierten Batch-Phase beobachtet werden. Sie geht auf eine Überproduktion von Acetyl-CoA zurück, welches aufgrund der limitierten Kapazität des Citratzyklus nicht oxidativ bis zum Stoffwechselintermediat Acetat abgebaut wird [Wolfe (2005)]. Für die mathematische Beschreibung des sogenannten Crabtree-Effektes erfolgt die Einführung eines Overflow-Metabolismus in Anlehnung an Xu *et al.* (1999) sowie Kleman and Strohl (1994).

Relevant ist vor allem die Modellierung der Zielproteinbildung. Es wurde hierfür ein wachstumsabhängiger Ansatz nach Leveau and Lindow (2001) gewählt, der zum Einen eine basale Proteinexpression nichtinduzierter Zellen einschliesst und eine Abhängigkeit vom zellinternen Induktorgehalt sowie von der Wachstumsrate berücksichtigt. Die Beschreibung des aktiven IPTG-Transports in die Zelle erfolgte in Anlehnung an einen von Bentley *et al.* (1991) formulierten Mechanismus.

Das Hauptproblem bei der Volumensimulation stellt die diskontinuierliche Probenahme dar, die eine Unstetigkeitsstelle in der Volumensimulation verursacht. Das entnommene Probevolumen V_S muss daher in einen Volumenstrom F_S umgerechnet werden. Das Zeitintervall darf dabei nicht zu klein gewählt werden, da der Rechenaufwand des Differentialgleichungslösers dadurch sehr groß wird und damit die Simulationsdauer erheblich erhöht wird. Außerdem kann dieses zu Ungenauigkeiten in der Volumensimulation führen und dadurch gravierende Abweichungen der Simulationsergebnisse vom Realvolumen auftreten.

Es hat sich für den unter MATLAB[®] verwendeten Differentialgleichungslöser *ODE15s* als zweckmäßig erwiesen, die nahezu impulsförmige Probenahmerate des Volumens ΔV_S bei $t = t_s$ durch eine dreieckpulsförmige Probenahmerate

Tabelle 6.1: Zustands- und Steuergrößen der Simulation

| Steuergrößen | | Zustandsgrößen | | | |
|------------------|---------------------|--|---------------------|-----------------------|---|
| Formelzeichen | Einheit | Bezeichnung | Formelzeichen | Einheit | Bezeichnung |
| F_A | $[\text{lh}^{-1}]$ | zellfreier Analysenstrom | C_{IndM} | $[\text{mol l}^{-1}]$ | Induktorkonzentration in der Medienphase |
| F_{AF} | $[\text{lh}^{-1}]$ | Antischaummittelzugabe | $G_{\text{Ind/X}}$ | $[\text{mol g}^{-1}]$ | zellspezifischer Induktorgehalt |
| F_{Ind} | $[\text{lh}^{-1}]$ | Induktorzufütterung | c_{PIM} | $[\text{g l}^{-1}]$ | Zielproteinkonzentration in der Medienphase |
| F_{hG} | $[\text{lh}^{-1}]$ | Begasungsrate unter Normbedingungen | $g_{\text{P1/X}}$ | $[\text{gg}^{-1}]$ | zellspezifischer Zielproteingehalt |
| F_{R1} | $[\text{lh}^{-1}]$ | Zufütterung aus Reservoir 1 (Glukose) | C_{P3totM} | $[\text{mol l}^{-1}]$ | Acetatkonzentration in der Medienphase |
| F_{R2} | $[\text{lh}^{-1}]$ | Zufütterung aus Reservoir 2 (Ammonium) | c_{SIM} | $[\text{g l}^{-1}]$ | Glukosekonzentration in der Medienphase |
| ΔV_S | [l] | Volumen einer Probenahme | c_{XLact} | $[\text{g l}^{-1}]$ | Zelldichte der aktiven Zellen |
| F_{T1} | $[\text{lh}^{-1}]$ | Säuretitration (Phosphorsäure) | c_{XLlys} | $[\text{g l}^{-1}]$ | fiktive Zelldichte der lysierten Zellen |
| F_{T2} | $[\text{lh}^{-1}]$ | Laugentitration (Ammoniumhydroxidlösung) | V_L | [l] | Volumen der flüssigen Reaktionsphase |
| N_{St} | $[\text{min}^{-1}]$ | Rührerdrehzahl | V_Z | [l] | Volumen der Zellphase |
| $p\text{O}_2$ | [%] | relativer Gelöstsauerstoffpartialdruck | | | |

Um die Akkumulation des Zielproteins c_{P1M} im zellfreiem Medienvolumen V_M charakterisieren zu können, wurde eine von der Rührerdrehzahl abhängige Lyserate eingeführt.

Eine Übersicht über die verwendeten Steuergrößen sowie die simulierten Zustandsgrößen des Modells bietet Tabelle 6.1.

6.1 Das Dreiphasenmodell

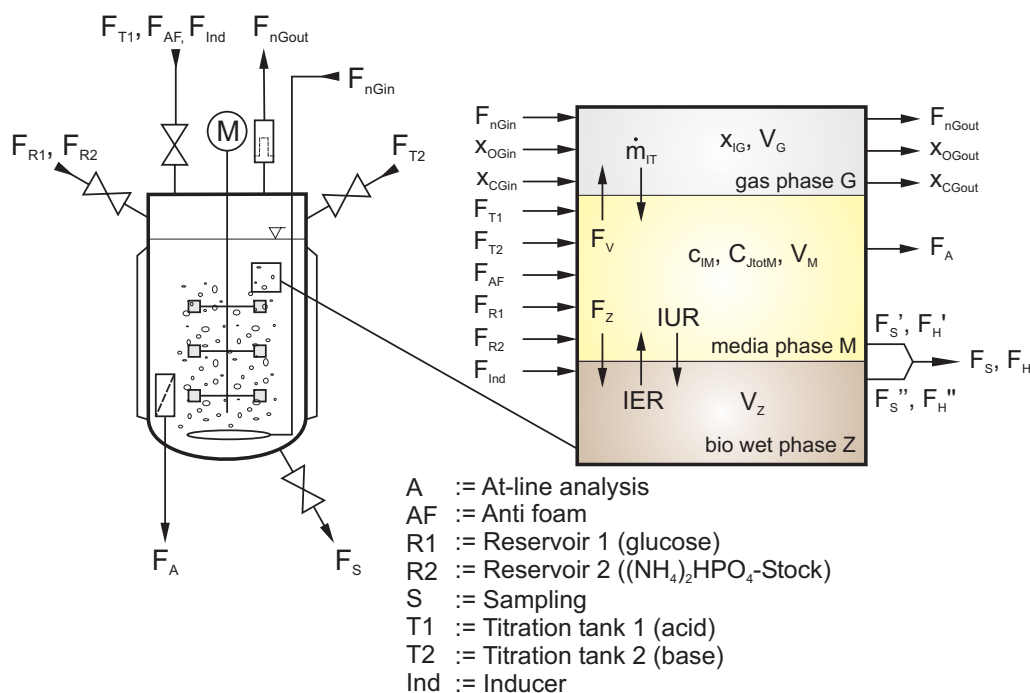


Bild 6.3: Makroskopisches Volumenelement eines Bioreaktors zur Bilanzierung eines Dreiphasenmodells [Peuker (2006)]

Für die Bilanzierung des Bioreaktorinhaltes wird üblicherweise eine Unterteilung in eine Gasphase G und eine Flüssigphase L vorgenommen. Da bei hohen Zelldichten die feuchte Biomasse einen nicht unbedeutenden Anteil der Flüssigphase einnimmt, die Bestimmung der messtechnisch erfassten Konzentrationen für Glukose, Acetat und Ammonium jedoch nur in der Medienphase (ohne Zellen) erfolgt, wird in dieser Arbeit eine Unterteilung in das zellfreie Medienvolumen V_M und das Zellvolumen V_Z gemäss Peuker (2006) vorgenommen.

Die simulierten Volumina und die Zelldichte (durchgezogene Linien) sowie die zugehörigen Prozessdaten (offene Symbole) sind exemplarisch in Bild 6.4 dargestellt. Dabei wird deutlich, dass die Biofeuchtphase am Ende der Kultivierung ca. 30% des gesamten Reaktionsvolumens einnimmt. Die korrekte Simulation der Volumina ist dabei von entscheidender Bedeutung, um die akkurate Berechnung der beteiligten Konzentrationen in den Bilanzgleichungen zu gewährleisten.

Dabei ergibt sich das gesamte flüssige Reaktionsvolumen V_L ,

$$V_L(t) = V_M(t) + V_Z(t) , \quad (6.1)$$

wie in Bild 6.3 dargestellt, als Summe aus dem Medienvolumen V_M und dem Zellvolumen V_Z .

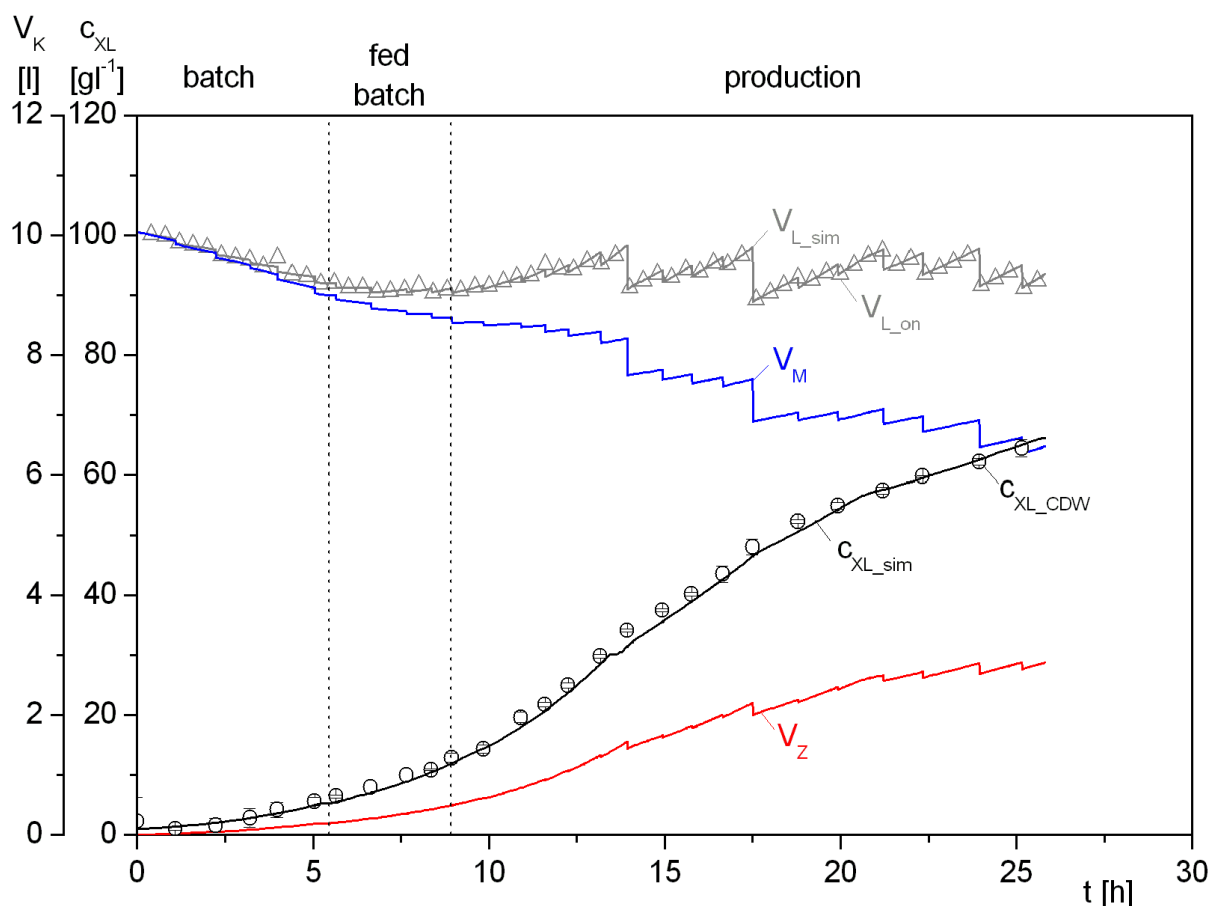


Bild 6.4: Verlauf des Reaktionsvolumens V_L , der Medienphase V_M und der Zellphase V_Z sowie der Zelldichte c_{XL} während des Kultivierungsprozesses CK4407
Symbole: Messgrößen Linien: simulierte Größen

Das Biofeuchtvolumen V_Z ,

$$V_Z(t) = \frac{\alpha_{Z/X}}{\rho_Z} \cdot m_{XL}(t), \quad (6.2)$$

kann über das als konstant angenommenem Verhältnis $\alpha_{Z/X}$ von Biofeucht- zu Biotrockenmasse und die ebenfalls als zeitinvariant postulierte Dichte der Biofeuchtphase ρ_Z direkt proportional aus der Zellmasse m_{XL} berechnet werden.

Die durch das Zellwachstum bedingte Volumenverlustrate F_Z des Medienvolumens V_M ,

$$F_Z(t) = q_{X/X}(t) \cdot V_Z(t) = \mu(t) \cdot V_Z(t) = \mu(t) \cdot \frac{\alpha_{Z/X}}{\rho_Z} \cdot m_{XL}(t), \quad (6.3)$$

ist dabei von der zellspezifischen Zellreaktionsrate $q_{X/X}$ („Wachstumsrate“ μ) und dem augenblicklichem Volumen V_Z der feuchten Zellen aus Gl. (6.2) abhängig.

Die im Folgenden zu bilanzierenden Stoffströme und Transferraten sind ebenfalls in Bild 6.3 dargestellt. Die Sauerstoffversorgung der Zellen sowie der Austrag des produzierten Kohlendioxids wird durch die Begasung sichergestellt. Über die Begasungsrate F_{nGin} und die bekannten Molenbrüche für Sauerstoff x_{OGin} und Kohlendioxid x_{CGin} am Eingang sowie die Analyse der gleichen Gaskomponenten (x_{OGout} und x_{CGout}) am Ausgang lässt sich die Gasphase, wie bei

Peuker (2006) ausführlich dargestellt, bilanzieren. Der Austrag von Wasser durch die Begasung bei schlechter Gasrückkühlung wird in der Verdunstungsrate F_V berücksichtigt, die im Weiteren mit 10 mlh^{-1} als konstant angenommen wird.

Der Stofftransport einer Komponente I von der Flüssig- in die Gasphase wird über die vorzeichenbehaftete Transferrate \dot{m}_{IT} dargestellt. Die entsprechende volumetrische Aufnahme- bzw. Produktionsrate IUR^1 bzw. IER^2 der Zellen beschreibt hingegen den auf V_L bezogenen Stofftransport der Medienkomponente in die Zelle bzw. von der Zelle in die Medienphase.

Die zugeführten Komponenten strömen ausschließlich in die Medienphase, aus der auch der zellfreie Analysenstrom F_A für die Atline-Analytik entnommen wird. Die zellbehaftete Probenahme F_S erfolgt aus dem gesamten flüssigen Reaktionsraum L und wird deshalb entsprechend der Volumenverhältnisse V_M bzw. V_Z zu V_L aufgeteilt.

6.2 Die allgemeinen Bilanzgleichungen

Die Grundlage der Modellierung stellen die dynamischen Bilanzgleichungen von ausgewählten Komponenten des Prozesses dar. Diese führen auf gewöhnliche Differentialgleichungen erster Ordnung, die mit Hilfe von numerischen Dgl.-Sollern unter MATLAB[®] gelöst werden können und ein vereinfachtes Abbild des Bioreaktionsprozesses darstellen.

Die Basis eines Bioreaktionsmodells bildet dabei immer die Bilanz der Zelldichte c_{XL} ,

$$\dot{c}_{XL}(t) = \left(-\frac{F_{in}(t) - F_A(t)}{V_L(t)} + q_{X/X}(t) \right) \cdot c_{XL}(t), \quad (6.4)$$

die auf den Reaktionsraum V_L bezogen ist.

Sie setzt sich aus Termen für die Verdünnung durch die Zufütterung und die Aufkonzentrierung durch den zellfreien Analysenstrom F_A sowie aus der Zellreaktion zusammen.

Dabei ergibt sich die gesamte Zufütterung F_{in} ,

$$F_{in}(t) = F_{R1}(t) + F_{R2}(t) + F_{T1}(t) + F_{T2}(t) + F_{AF}(t) + F_{Ind}(t) - F_V(t), \quad (6.5)$$

aus der Summe aller Zufüterraten und dem negativ zu berücksichtigenden Verdunstungsterm F_V .

Der abfließende Volumenstrom F_{out} ,

$$F_{out}(t) = F_A(t) + F_S(t) + F_H(t), \quad (6.6)$$

setzt sich aus der kontinuierlichen zellfreien Probenahme F_A für die Atline-Analytik sowie der diskontinuierlichen Probenahmerate F_S und der Ernterate F_H zusammen.

Die zeitliche Änderung des flüssigen Reaktionsvolumens V_L ,

$$\dot{V}_L(t) = F_{in}(t) - F_{out}(t), \quad (6.7)$$

¹IUR – *Volumetric Uptake Rate* der Komponente I

²IER – *Volumetric Evolution Rate* der Komponente I

kann dann einfach aus der Differenz der gesamten Zufütterung aus Gl. (6.5) und dem abfließenden Volumenstrom aus Gl. (6.6) bestimmt werden. Hierbei werden gleiche Stoffeigenschaften im Zulauf und im Reaktor vorausgesetzt.

Um dem in Abschnitt 6.1 vorgestellten Dreiphasenmodell zu entsprechen, wird zunächst die Bilanzgleichung für das Volumen der feuchten Zellen V_Z ,

$$\dot{V}_Z(t) = F_Z(t) - (F_S(t) + F_H(t)) \cdot \frac{V_Z(t)}{V_L(t)}, \quad (6.8)$$

formuliert. Dabei vergrößert sich das Volumen durch das Zellwachstum gemäß Gl. (6.3) und wird durch die diskontinuierliche Probenahme sowie die Zellernte vermindert. Beide werden entsprechend des Volumenverhältnisses von feuchten Zellen zum Reaktionsvolumen berücksichtigt wird.

Die Volumenbilanz der Medienphase V_M ,

$$\dot{V}_M(t) = F_{in}(t) - F_A(t) - F_Z(t) - (F_S(t) + F_H(t)) \cdot \frac{V_M(t)}{V_L(t)}, \quad (6.9)$$

berücksichtigt alle Zufüterraten sowie die zellfreie Probenahme. Wie in Gl. (6.8) müssen hier ebenfalls das Zellwachstum, die Probenahme und die Zellernte berücksichtigt werden.

Das unbekannte Volumen der Medienphase V_M ,

$$V_M(t) = V_L(t) - V_Z(t) = V_L(t) \cdot \left(1 - \frac{\alpha_{Z/X}}{\rho_Z} \cdot c_{XL}(t) \right), \quad (6.10)$$

kann bei bekanntem Reaktionsvolumen und bekannter Zelldichte unter Verwendung von Gl. (6.1) und Gl. (6.2) einfach berechnet werden, so dass die Medienbilanz aus Gl. (6.9) nicht gelöst werden muss.

Nachdem die Volumenbilanzgleichungen für die drei Phasen vollständig formuliert wurden, können die Massenbilanzgleichungen aufgestellt werden.

Die allgemeine Massenbilanz einer nicht dissoziierenden und filtrierbaren Komponente I in der Medienphase M,

$$\begin{aligned} \dot{m}_{IM}(t) &= \dot{m}_{Iin}(t) - \dot{m}_{Iout}(t) + \dot{m}_{IT}(t) + \dot{m}_{IMreact}(t) \\ &= \sum (F_K(t) \cdot c_{IK}) - \left[\frac{F_S(t) + F_H(t)}{V_L(t)} + \frac{F_A(t)}{V_M(t)} \right] \cdot m_{IM}(t) + \dot{m}_{IT}(t) \pm q_{I/X}(t) \cdot m_{XL}(t), \end{aligned} \quad (6.11)$$

setzt sich aus den einzelnen Termen für den Zu- und Ablauf, die Transferrate von der Gas in die Flüssigphase \dot{m}_{IT} sowie die Reaktion über die Zellmasse zusammen. Da sich bei der Verwendung von Massenbilanzen durch die diskontinuierliche Probenahme Unstetigkeitsstellen ergeben, die den Zeitaufwand der Differentialgleichungslöser erheblich vergrößern, wird im Folgenden mit Konzentrationsbilanzen anstelle von Massenbilanzen gearbeitet.

Unter Berücksichtigung der Produktregel kann die Massenbilanz für die Komponente I in Phase M,

$$\dot{m}_{IM}(t) = \dot{V}_M(t) \cdot c_{IM}(t) + V_M(t) \cdot \dot{c}_{IM}(t), \quad (6.12)$$

auch über die Konzentration der Komponente c_{IM} im Volumen V_M der Medienphase sowie beider zeitlichen Ableitungen formuliert werden.

Damit kann Gl. (6.11) in die allgemeine Konzentrationsbilanz einer nicht dissoziierenden Komponente I in der Medienphase M,

$$\dot{c}_{IM}(t) = \frac{\rho_Z}{\rho_Z - \alpha_{Z/X} \cdot c_{XL}(t)} \cdot \left[\frac{\sum F_K(t) \cdot c_{IK} - F_{in}(t) \cdot c_{IM}(t)}{V_L(t)} + \frac{q_{X/X}(t) \cdot \alpha_{Z/X} \cdot c_{XL}(t)}{\rho_Z} \cdot c_{IM}(t) + ITR(t) \pm q_{I/X}(t) \cdot c_{XL}(t) \right], \quad (6.13)$$

überführt werden. Sie setzt sich aus Zunahme und Verdünnung durch die Zufütterung, aus der Aufkonzentrierung durch Wachstum, dem Ein- oder Austrag über die Gasphase und der Reaktion der jeweiligen Komponente zusammen.

Für eine dissoziierende Komponente C_{JtotM} ,

$$\dot{C}_{JtotM}(t) = \frac{\rho_Z}{\rho_Z - \alpha_{Z/X} \cdot c_{XL}(t)} \cdot \left[\frac{\sum F_K(t) \cdot C_{JtotK} - F_{in}(t) \cdot C_{JtotM}(t)}{V_L(t)} + \frac{q_{X/X}(t) \cdot \alpha_{Z/X} \cdot c_{XL}(t)}{\rho_Z} \cdot C_{JtotM}(t) + \frac{JTR(t)}{M_J} \pm Q_{J/X}(t) \cdot c_{XL}(t) \right], \quad (6.14)$$

werden molare Konzentrationen anstelle von Massenkonzentrationen bilanziert.

Werden zellinterne Komponenten betrachten, wie im Fall des eingesetzten Induktors IPTG, wird nicht die Konzentration sondern der zellspezifische Gehalt $g_{I/X}$ der jeweiligen Komponente I betrachtet. Dessen Änderung,

$$\dot{g}_{I/X}(t) = q_{I/X}(t) - q_{I/Xcon}(t) - q_{X/X}(t) \cdot g_{I/X}(t), \quad (6.15)$$

wird über die zellspezifische Reaktionsrate $q_{I/X}$ (Aufnahme – positiv, Abgabe – negativ) und die zellspezifische Konversionsrate $q_{I/Xcon}$ (Verbrauch – positiv, Bildung – negativ) beeinflusst. Darüber hinaus wird $g_{I/X}$ durch das Wachstum (Zellteilung) grundsätzlich verringert.

Da im hier verwendeten Modell lediglich der zellinterne Gehalt des von den Zellen nicht metabolisierten Induktors modelliert wird, kann Gl. (6.15),

$$\dot{g}_{I/X}(t) = q_{I/X}(t) - q_{X/X}(t) \cdot g_{I/X}(t), \quad (6.16)$$

vereinfacht ohne Konversionsterm notiert werden.

Um zwischen Massenkonzentrationen und molaren Konzentrationen innerhalb der Zelle zu unterscheiden, werden für letztere Grossbuchstaben verwendet, so dass sich in Analogie zu Gl. (6.16) die Änderung des zellspezifischen molaren Gehalts $G_{J/X}$ für eine nicht metabolisierte Komponente J,

$$\dot{G}_{J/X}(t) = \frac{1}{M_J} \cdot q_{J/X}(t) - q_{X/X}(t) \cdot G_{J/X}(t), \quad (6.17)$$

ergibt. Dabei muss zusätzlich die molare Masse M_J der Komponente J eingeführt werden, um weiterhin die zellspezifische (Massen-) Aufnahmerate $q_{J/X}$ nutzen zu können.

6.3 Verwendete Reaktionskinetische Ansätze

Die Basis für die Modellierung der biotischen Phase, die den Verbrauch der Substrate sowie die Bildung der Produkte beschreiben, stellen die Reaktionskinetiken dar. Dabei wird in dieser Arbeit zunächst ein unstrukturiertes Modell auf Basis von Monod-Kinetiken und dem Bottle-Neck Prinzip postuliert, welches an Xu *et al.* (1999) und Peuker (2006) angelehnt ist. Im Folgenden wird dieses weiter verfeinert und ein segregierter Ansatz für die Zelldichte und die Zielproteinbildung verwendet.

6.3.1 Mathematische Beschreibung des Overflow-Metabolismus

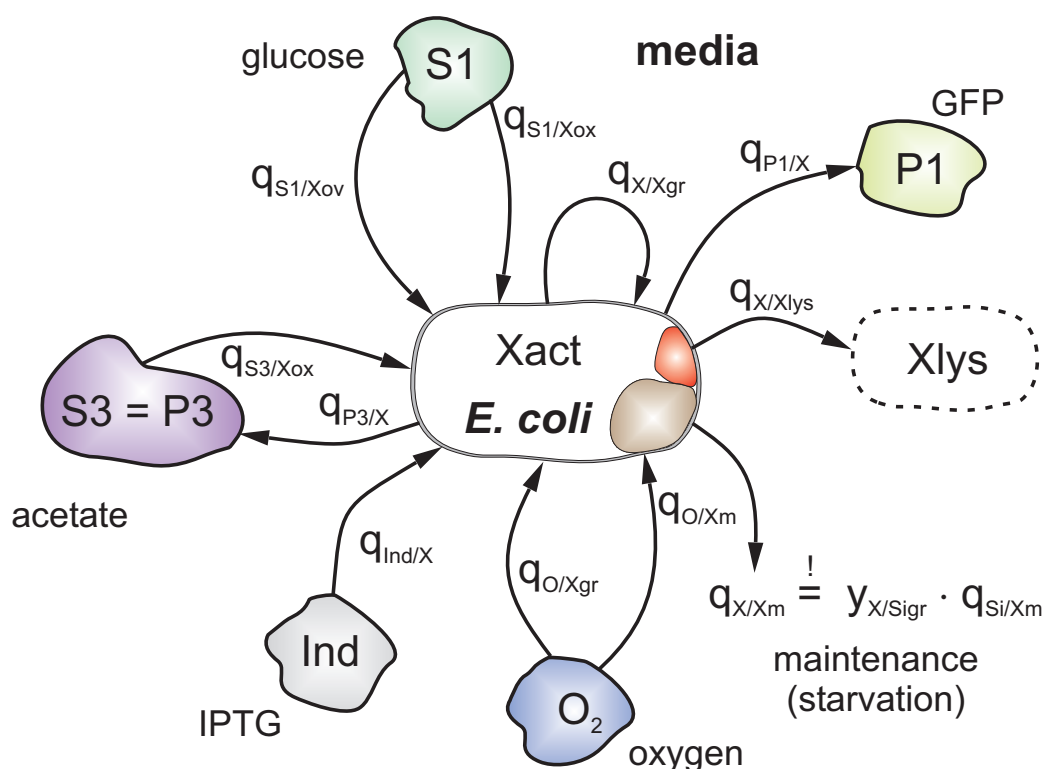


Bild 6.5: Schematische Darstellung des für die Simulation verwendeten Reaktionsschemas

Das für die Modellierung verwendete Reaktionsschema ist in Bild 6.5 dargestellt. Dabei metabolisiert *Escherichia coli* das Hauptsubstrat Glukose (S1) oxidativ ($q_{S1/Xox}$). Bei einem Überangebot von Glukose und limitierter oxidativer Kapazität der Zellen kann es jedoch, wie in Abschnitt 2.3 näher beschrieben, zur Ausbildung eines sogenannten Overflow-Metabolismus ($q_{S1/Xov}$) kommen. Dieses führt zur Bildung des Nebenproduktes Acetat ($q_{P3/X}$), welches bei glukoselimitierter Fahrweise und ausreichender Sauerstoffversorgung durch die Zellen wieder abgebaut werden kann ($q_{S3/Xox}$).

Sowohl für die oxidative Glukoseaufnahme als auch für die Acetataufnahme müssen die Zellen dabei ausreichend mit Sauerstoff ($q_{O/Xgr}$) versorgt werden. Neben der Beschreibung des Zellzuwachses ($q_{X/Xgr}$) wird auch der Substraterhaltungsstoffwechsel in Form einer fiktiven Sterberate $q_{X/Xm}$ berücksichtigt. Der Sauerstoffbedarf des Erhaltungsstoffwechsels wird dabei über die O₂-Maintenancerate ($q_{O/Xm}$) wiedergegeben.

Die GFP-Bildung wird über die zellspezifische Reaktionsrate $q_{p1/X}$ beschrieben, die sich aus einer basalen Expression vor der Induktion und einem durch den zellinternen Induktorgehalt sowie die zellspezifische Wachstumsrate gesteuerten Anteil zusammensetzt. Der aktive Transport des chemischen Induktors IPTG wird durch die zellspezifische Reaktionsrate $q_{Ind/X}$ dargestellt.

Um die Zielproteinakkumulation in der Medienphase beschreiben zu können, wird in Abschnitt 6.3.2 zusätzlich eine zellspezifische Lyserate $q_{X/Xlys}$ eingeführt.

Zur Beschreibung des Zellwachstums wird zunächst das Bottle-Neck der Wachstumskinetik formuliert und anschließend die zur Simulation benötigten Gleichungen aufgestellt.

Die beobachtbare zellspezifische Wachstumsrate $q_{X/X}$,

$$\mu(t) = q_{X/X}(t) = \text{Min} \{ \mu_{1max} + q_{X/Xm}(t), q_{X/XSgr}(t), q_{X/XOgr}(t) \} - q_{X/Xm}(t), \quad (6.18)$$

wird dabei durch die maximale Wachstumsrate auf Glukose μ_{1max} begrenzt und wird im limitierten Fall entweder durch die substratgesteuerte Wachstumsrate $q_{X/XSgr}$ oder die sauerstoffgesteuerte Zuwachsrate $q_{X/XOgr}$ vorgegeben.

Die Verknüpfung zwischen der nicht beobachtbaren Zuwachsrate $q_{X/Xgr}$ und der messbaren (Netto-) Wachstumsrate $q_{X/X}$ besteht in der ebenfalls nicht beobachtbaren fiktiven Sterberate $q_{X/Xm}$,

$$q_{X/Xm}(t) = q_{X/Xgr}(t) - q_{X/X}(t), \quad (6.19)$$

die den Substraterhaltungstoffwechsel in die Zellbilanz transformiert.

Für die Glukoseaufnahme wird eine Monod-Kinetik mit nicht-kompetitiver Inhibierung durch das Stoffwechselprodukt Acetat C_{S3totM} angenommen [Kleman and Strohl (1994)][Xu *et al.* (1999)].

Die potentielle Glukoseaufnahme rate $q_{S1/Xpot}$,

$$q_{S1/Xpot}(t) = q_{S1/Xmax}(t) \cdot \frac{c_{S1M}(t)}{k_{S1} + c_{S1M}(t)} \cdot \frac{K_{I13}}{K_{I13} + C_{S3totM}(t)}, \quad (6.20)$$

kann dann aus der maximal möglichen Glukoseaufnahme rate $q_{S1/Xmax}$ und dem Produkt des Monod-Terms sowie des Inhibierungsterms berechnet werden. Dabei stellen c_{S1M} die Glukosekonzentration, C_{S3totM} die Acetatkonzentration sowie k_{S1} und K_{I13} die entsprechende Limitierungs- bzw. Inhibierungskonstante dar.

Die in Gl. (6.20) enthaltene maximale Glukoseaufnahme rate $q_{S1/Xmax}$,

$$q_{S1/Xmax}(t) = \frac{\mu_{1max}}{Y_{X/S1gr}} + q_{S1/Xm}(t), \quad (6.21)$$

ergibt sich aus dem zeitinvarianten Quotienten von maximaler Wachstumsrate μ_{1max} auf dem Hauptsubstrat Glukose mit dem zugehörigen Zuwachsausbeutekoeffizienten $Y_{S1/Xgr}$ sowie der Addition einer Substratmaintenancerate $q_{S1/Xm}$.

Die Substratmaintenancerate für Glukose $q_{S1/Xm}$,

$$q_{S1/Xm}(t) = q_{S1/Xm0} + q_{S1/XmInd}(t) , \quad (6.22)$$

wird über einen zeitinvarianten Anteil $q_{S1/Xm0}$ sowie mit einem induktionsabhängigen Anteil $q_{S1/XmInd}$ postuliert.

Der induzierbare zeitvariante Anteil $q_{S1/XmInd}$,

$$q_{S1/XmInd}(t) = k_{S1m} \cdot g_{P1/X}(t) , \quad (6.23)$$

ist dabei über einen konstanten Faktor k_{S1m} an den zellinternen Zielproteingehalt $g_{P1/X}$ gekoppelt.

Die fiktive Sterberate $q_{X/Xm}$ aus Gl. (6.19),

$$q_{X/Xm}(t) = y_{X/S1gr} \cdot q_{S1/Xm}(t) , \quad (6.24)$$

ist dann unter Verwendung des Zuwachsausbeutekoeffizienten $y_{X/S1gr}$ von Glukose direkt aus der Substratmaintenancerate $q_{S1/Xm}$ berechenbar.

In Analogie zur Kinetik der Glukoseaufnahme wird für die Sauerstoffaufnahme ebenfalls eine Monod-Kinetik zugrundegelegt, die eine Inhibierung durch Acetat beinhaltet.

Damit kann die zellspezifische oxidative Kapazität $q_{O/Xcap}$,

$$q_{O/Xcap}(t) = q_{O/Xmax}(t) \cdot \frac{c_{OM}(t)}{k_O + c_{OM}(t)} \cdot \frac{K_{IO3}}{K_{IO3} + C_{P3totM}(t)} , \quad (6.25)$$

über das Produkt aus der maximalen zellspezifischen Sauerstoffaufnahmerate $q_{O/Xmax}$, den O_2 -Monod-Term, der die Gelöstsauerstoffkonzentration c_{OM} in der Medienphase sowie die Limitierungskonstante k_O für Sauerstoff beinhaltet, und den Inhibierungsterm aus Acetatkonzentration C_{P3totM} mit zugehöriger Inhibierungskonstante K_{IO3} formuliert werden.

Für die in dieser Arbeit durchgeführten Simulationen wurde die Gelöstsauerstoffkonzentration c_{OM} ,

$$c_{OM}(t) = \frac{c_{OM100}(t)}{100\%} \cdot pO_2(t) , \quad (6.26)$$

aus dem gemessenem relativen Sauerstoffpartialdruck pO_2 berechnet, der als Steuergröße in die Simulation einfließt.

Die Sättigungskonzentration des Gelöstsauerstoffes in der Medienphase c_{OM100} ,

$$c_{OM100} = \frac{p_{Gcal} \cdot x_{OGcal}}{H_{O2}} , \quad (6.27)$$

lässt sich dabei aus dem Molenbruch des Sauerstoffs x_{OGcal} und dem Reaktordruck p_{Gcal} während der pO_2 -Kalibrierung sowie dem Henrykoeffizienten H_{O2} berechnen. Da alle Kultivierungen bei konstanter Temperatur durchgeführt wurden, kann die Zeitabhängigkeit der Sauerstofflöslichkeit im Folgenden vernachlässigt werden.

Die in Gl. (6.25) benötigte maximale zellspezifische Sauerstoffaufnahme $q_{O/X_{\max}}$,

$$q_{O/X_{\max}}(t) = \frac{\mu_{1\max} + q_{X/X_m}(t)}{y_{X/O_{gr}}} + q_{O/X_m}, \quad (6.28)$$

kann in Anlehnung an Gl. (6.21) formuliert werden. Dabei muss zusätzlich der zellspezifische Sauerstoffbedarf q_{O/X_m} für die Umsetzung der Maintenancespeicherstoffe berücksichtigt werden.

Die maximale Reaktionsrate für Sauerstoff wird nur an die Glukoseaufnahme gekoppelt, um die Konsistenz der Gleichung bei der folgenden Formulierung des Overflow-Metabolismus nach Xu *et al.* (1999) zu gewährleisten. Um diesen zu modellieren, wird die zellspezifische Glukoseaufnahme $q_{S1/X}$,

$$q_{S1/X}(t) = q_{S1/X_{ox}}(t) + q_{S1/X_{ov}}(t), \quad (6.29)$$

in einen oxidativ zur Zellbildung verwertbaren Anteil $q_{S1/X_{ox}}$ sowie einen Overflow-Anteil $q_{S1/X_{ov}}$, der zur Acetatbildung führt, unterteilt.

Die oxidative zellspezifische Glukoseaufnahme $q_{S1/X_{ox}}$,

$$q_{S1/X_{ox}}(t) = \text{Min} \{ q_{S1/X_{pot}}(t), y_{S1/O} \cdot q_{O/X_{cap}}(t) \}, \quad (6.30)$$

wird aus dem Minimum der potentiellen Glukoseaufnahme aus Gl. (6.20) und dem Produkt eines zeitinvarianten stöchiometrischen Faktors $y_{S1/O}$ mit der zellspezifischen oxidativen Kapazität $q_{O/X_{cap}}$ aus Gl. (6.25) bestimmt.

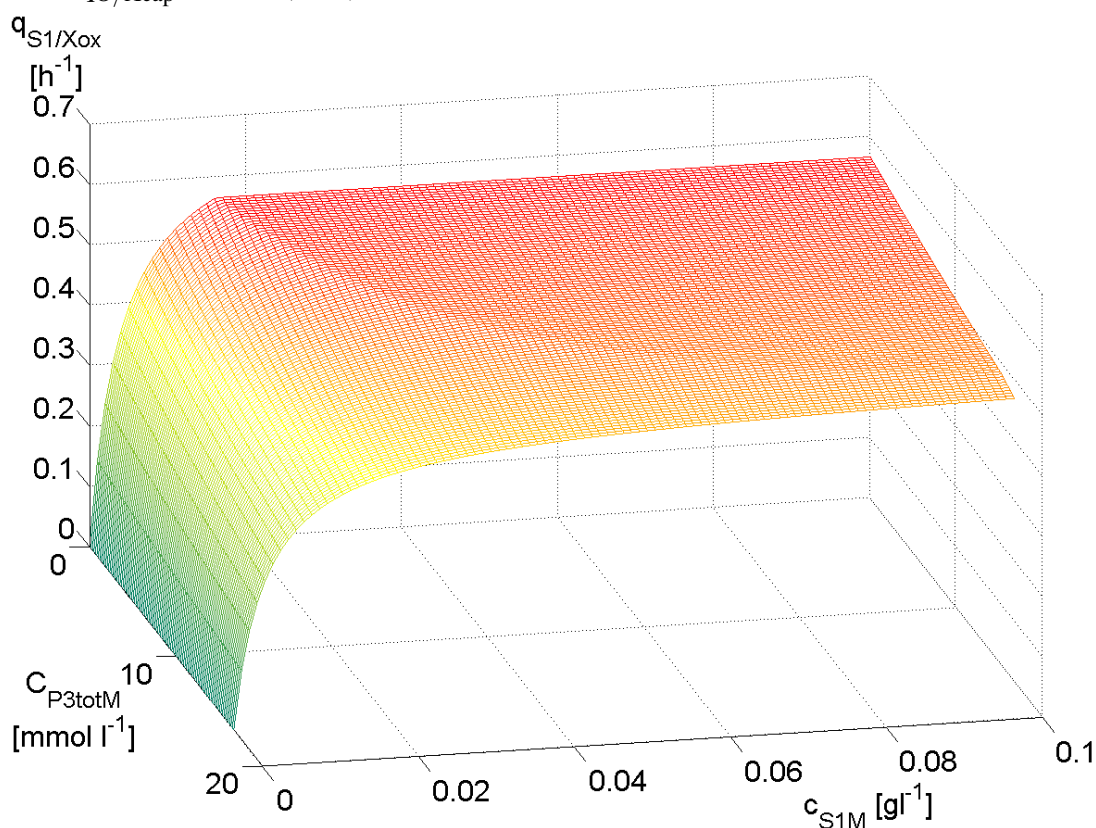


Bild 6.6: Verlauf der oxidativen Glukoseaufnahme $q_{S1/X_{ov}}$ in Abhängigkeit der Glukosekonzentration c_{S1M} sowie der Acetatkonzentration C_{P3totM}

Die sich daraus ergebende Abhängigkeit der oxidativen Kapazität von der limitierenden Glukose- sowie der inhibierenden Acetatkonzentration ist in Bild 6.6 dargestellt.

Überschreitet die potentielle Glukoseaufnahme die oxidative Kapazität der Zellen, kommt es zur Bildung von Acetat.

Die zellspezifische Glukoseaufnahme des Overflow-Metabolismus $q_{S1/Xov}$,

$$q_{S1/Xov}(t) = q_{S1/Xpot}(t) - q_{S1/Xox}(t), \quad (6.31)$$

kann dabei bei bekannter oxidativer Kapazität der Zellen über eine einfache Umstellung von Gl. (6.29) berechnet werden.

Unter Berücksichtigung des Acetatausbeutekoeffizienten auf Glukose $y_{P3/S1}$ steht die zellspezifische Acetatbildungsrate $q_{P3/X}$,

$$q_{P3/X}(t) = y_{P3/S1} \cdot q_{S1/Xov}(t), \quad (6.32)$$

damit zur Verfügung und ist direkt proportional zum Overflow-Anteil der Glukoseaufnahme. Die in Bild 6.7 dargestellte Acetatproduktionsrate tritt damit nur auf, wenn die oxidative Substrataufnahmekapazität der Zellen geringer ist als die potentielle Glukoseaufnahme [Xu *et al.* (1999)].

Die Acetatbildung tritt damit entweder bei einem Überangebot von Glukose unter nicht sauerstofflimitierten Bedingungen oder aber unter sauerstofflimitierten Bedingungen bei der Anwesenheit von Glukose auf [Wolfe (2005)].

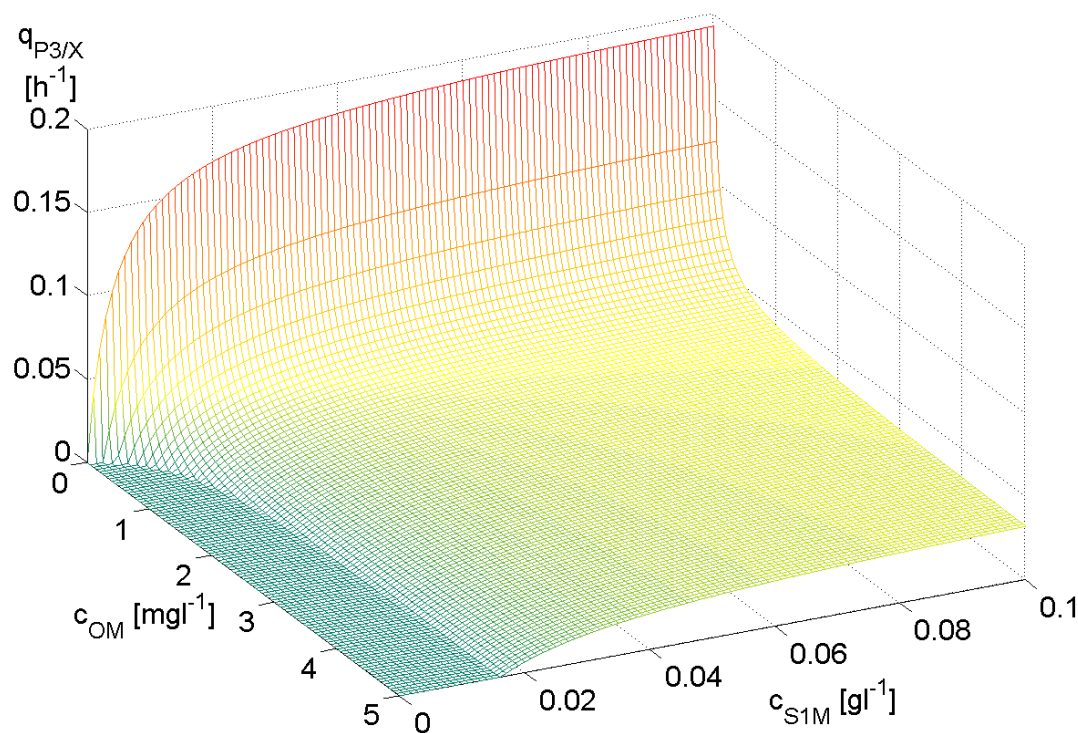


Bild 6.7: Verlauf der Acetatbildungsrate $q_{P3/X}$ in Abhängigkeit der Glukosekonzentration c_{S1M} sowie der Glöstsauerstoffkonzentration c_{OM}

Der stöchiometrische Koeffizient einer unvollständigen Glukoseverwertung bis zum Acetat $y_{P3/S1}$,

$$y_{P3/S1} = \frac{y_{X/S1gr} - y_{X/S1ov}}{y_{X/S3gr}}, \quad (6.33)$$

muss in der Summe der oxidativen Ausbeute für Glukose entsprechen. Deshalb setzt er sich aus der Differenz des Zuwachsausbeutekoeffizienten auf Glukose $y_{X/S1gr}$ und dem Ausbeutekoeffizienten des Overflow-Metabolismus $y_{X/S1ov}$ im Verhältnis zum Zuwachsausbeutekoeffizienten auf Acetat $y_{X/S3gr}$ zusammen.

Der Ausbeutekoeffizienten des Overflow-Metabolismus $y_{X/S1ov}$,

$$y_{X/S1ov} = \frac{3 \cdot w_{C/S3} \cdot y_{X/S1gr} - 2 \cdot w_{C/S1} \cdot y_{X/S3gr}}{3 \cdot w_{C/S3} - 2 \cdot w_{C/X} \cdot y_{X/S3gr}}, \quad (6.34)$$

kann dabei nach Schneider (1999) über die Zuwachsausbeutekoeffizienten $y_{X/Sigr}$ für die Biomasse X aus Glukose und aus Acetat mit Hilfe der Massenanteile $w_{C/I}$ von Kohlenstoff in Glukose (S1), in Acetat (S3 = P3) und in der Biomasse (X) berechnet werden.

Damit ist die substratgesteuerte Overflow-Zuwachsrate $q_{X/XSov}$,

$$q_{X/XSov}(t) = y_{X/S1gr} \cdot q_{S1/Xox}(t) + y_{X/S1ov} \cdot q_{S1/Xov}(t), \quad (6.35)$$

aus Gl. (6.18) für den Fall der Acetatproduktion bestimmbar. Sie setzt sich aus den über die entsprechenden Substratausbeutekoeffizienten transformierten zellspezifischen Glukoseaufnahmeraten des oxidativen und des Overflow-Metabolismus zusammen.

Bei glukoselimitierter Fahrweise und ausreichender oxidativer Kapazität der Zellen kann das gebildete Acetat von den Zellen wieder metabolisiert werden [Varma and Palsson (1994)].

Die zugrundeliegende potentielle Acetataufnahmerate $q_{S3/Xpot}$,

$$q_{S3/Xpot}(t) = q_{S3/Xmax}(t) \cdot \frac{C_{S3totM}(t)}{K_{S3} + C_{P3totM}(t)} \cdot \frac{k_{I31}}{k_{I31} + c_{S1M}(t)}, \quad (6.36)$$

wird analog zu Gl. (6.20) und Gl. (6.25) formuliert. Die Acetataufnahme wird dabei durch die vorliegende Glukose nicht-kompetitiv gehemmt. K_{S3} stellt dabei die Limitierungskonstante von Acetat und k_{I31} deren Inhibierungskonstante durch Glukose dar.

Die maximale Acetataufnahmerate $q_{S3/Xmax}$,

$$q_{S3/Xmax}(t) = \frac{\mu_{3max}}{y_{X/S3gr}} + q_{S3/Xm}(t), \quad (6.37)$$

kann in Anlehnung an Gl. (6.21) formuliert werden. Sie berechnet sich aus der maximalen Wachstumsrate μ_{3max} , dem Zuwachsausbeutekoeffizienten $y_{X/S3gr}$ und der Substratmaintenancerate $q_{S3/Xm}$ von Acetat.

Um die Konsistenz der Gleichungen zu wahren, muss letztere über

$$q_{S3/Xm}(t) = \frac{y_{X/S1gr}}{y_{X/S3gr}} \cdot q_{S1/Xm}(t) \quad (6.38)$$

aus der Substratmaintenancerate für Glukose und dem Quotienten der zugehörigen Zuwachsausbeutekoeffizienten ermittelt werden.

Damit ist die substratgesteuerte oxidative Zuwachsrate $q_{X/XSox}$,

$$q_{X/XSox}(t) = y_{X/S1gr} \cdot q_{S1/Xox}(t) + y_{X/S3gr} \cdot q_{S3/Xpot}(t) , \quad (6.39)$$

für den Fall, dass kein Overflow-Metabolismus vorliegt, bestimmbar. Sie ergibt sich aus der oxidativen Glukose- und der Acetataufnahmerate.

Die substratgesteuerte Zellzuwachsrate $q_{X/XSgr}$,

$$q_{X/XSgr}(t) = \text{Max} \{ q_{X/XSov}(t), q_{X/XSox}(t) \} , \quad (6.40)$$

ergibt sich aus dem Maximum von oxidativer Glukoseaufnahme und einem eventuellen Overflowanteil (Gl. (6.35)) oder der rein oxidativen Glukoseaufnahme und einer etwaigen Acetataufnahme gemäß Gl. (6.39). Eine simultane Acetatproduktion und Acetataufnahme ist ausgeschlossen.

Die letzte unbekannt Grösse des Bottle-Necks aus Gl. (6.18) stellt dann die sauerstoffgesteuerte Zellzuwachsrate $q_{X/XOgr}$,

$$q_{X/XOgr}(t) = y_{X/Ogr} \cdot (q_{O/Xcap}(t) - q_{O/Xm}) + y_{X/S1ov} \cdot q_{S1/Xov}(t) , \quad (6.41)$$

dar. Sie enthält zum einen das Produkt des Zuwachsausbeutekoeffizienten $y_{X/Ogr}$ für Sauerstoff mit der Differenz von oxidativer Kapazität der Zellen und dem Sauerstoffbedarf für die Umsetzung der Maintenancespeicherstoffe. Zum anderen wird der Wachstumsanteil des Overflow-Metabolismus berücksichtigt. Letzterer benötigt zwar keinen Sauerstoff, stellt aber einen Wachstumsanteil für den zu bestimmenden Bottle-Neck in Gl. (6.18) dar.

Nachdem der Bottle-Neck des Zellwachstums vollständig beschrieben ist, können abschließend die tatsächlichen zellspezifischen Reaktionsraten ermittelt werden, die für die Lösung der Dgln. aus Abschnitt 6.2 benötigt werden.

Die zellspezifische oxidative Glukoseaufnahmerate $q_{S1/Xox}$ wird unter Verwendung von Gl. (6.30) berechnet. Im Fall des Overflow-Metabolismus wird der Anteil der Wachstumsrate, der nicht durch die oxidative Glukoseaufnahme abgedeckt werden kann, über Gl. (6.31) bestimmt. Damit können die zellspezifische Glukoseaufnahmerate gemäß Gl. (6.29) sowie die Acetatbildungsrate nach Gl. (6.32) abschließend ermittelt werden.

Abschließend kann die beobachtbare Wachstumsrate $q_{X/X}$ in Gl. (6.18) aus der Zuwachsrate $q_{X/Xgr}$,

$$q_{X/Xgr}(t) = \text{Min} \{ q_{X/XSgr}(t), q_{X/XOgr}(t) \} , \quad (6.42)$$

berechnet werden. Diese ergibt sich aus dem Bottle-Neck der Substratbedingungen aus Gl. (6.40) sowie der Sauerstoffbedingungen aus Gl. (6.41).

Im Fall einer Acetataufnahme ergibt sich die zugehörige Reaktionsrate $q_{S3/X}$,

$$q_{S3/X}(t) = \text{Min} \left\{ \frac{q_{X/Xgr}(t) - y_{X/S1gr} \cdot q_{S1/Xox}(t)}{y_{X/S3gr}}, q_{S3/Xpot}(t) \right\}, \quad (6.43)$$

aus dem Minimum des nicht durch die oxidative Glukoseaufnahme abgedeckten Wachstumsanteils sowie der unter Gl. (6.36) bestimmten potentiellen Acetataufnahmerate.

Bild 6.8 zeigt exemplarisch die Acetatbildung während der Batch-Phase und die anschließende Wiederaufnahme zu Beginn der Fed-Batch Phase. Die offline bestimmte Acetatkonzentration C_{P3totM} (offene Symbole) stimmt dabei gut mit der simulierten Größe (durchgezogene Linie) überein. Der Zeitpunkt des vollständigen Acetatverbrauchs, der sich in dem ansteigendem Sauerstoffpartialdruck pO_2 bei $t = 6.2$ h widerspiegelt, wird durch das Modell gut wiedergegeben.

Zusätzlich sind sowohl die oxidative Glukoseaufnahmerate $q_{S1/Xox}$ als auch die zur Acetatbildung führende Glukoseaufnahmerate des Overflow-Metabolismus $q_{S1/Xov}$ dargestellt. Es kommt zur Acetatbildung, wenn die potentielle zellspezifische Glukoseaufnahmerate $q_{S1/Xpot}$ das Produkt aus dem stöchiometrischen Faktor $y_{S1/O}$ und der zellspezifischen Sauerstoffaufnahmekapazität $q_{O/Xcap}$ überschreitet.

Zu Beginn der Fed-Batch Phase beinhaltet die zellspezifische Zellreaktionsrate $q_{X/X}$ noch den Wachstumsanteil durch die Acetataufnahme. Nach dem vollständigem Acetatverbrauch

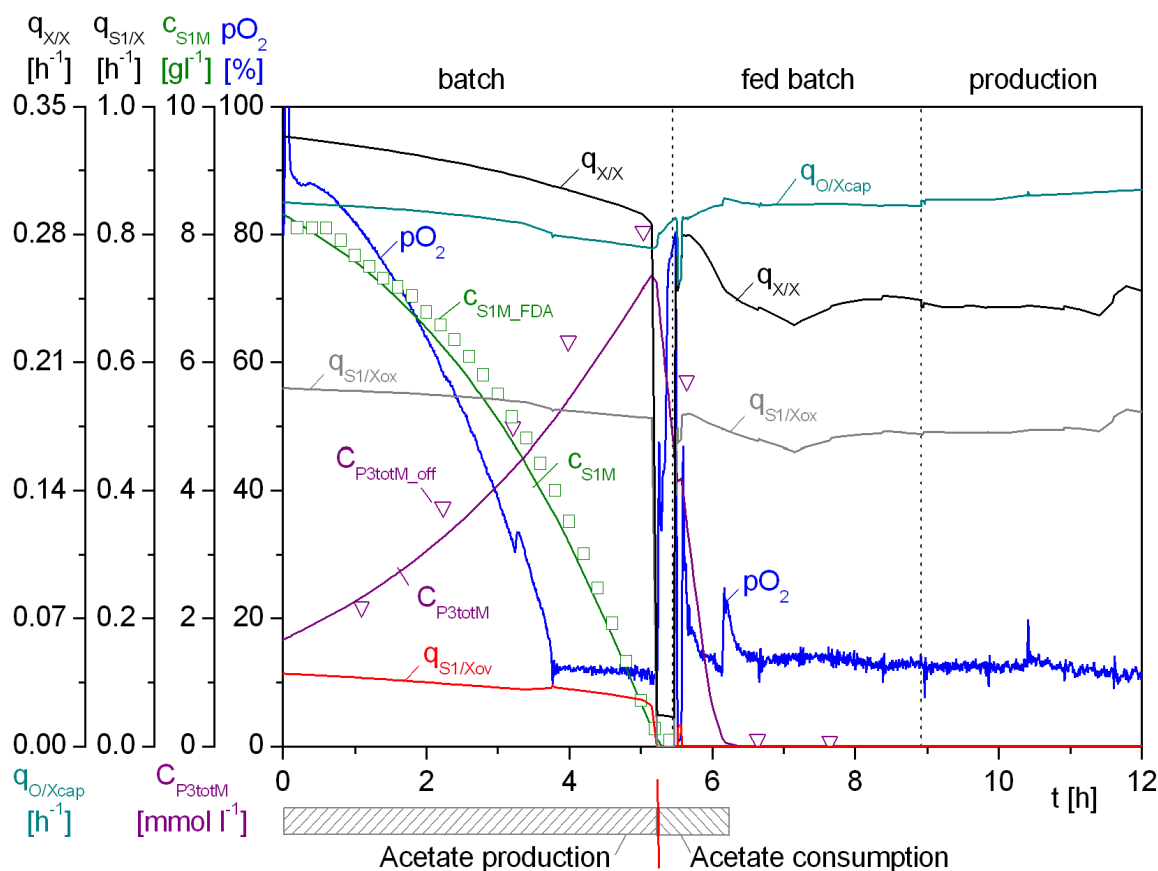


Bild 6.8: Acetatbildung und Acetataufnahme während der Batch und der Fed-Batch Phase der Kultivierung CK4407

wird das Zellwachstum lediglich durch die Glukoseaufnahme gesteuert, so dass beide dann Raten proportional zueinander verlaufen.

Eine detailliertere Darstellung des Übergangs zwischen Batch und Fed-Batch Phase zeigt Bild 6.9. Zusätzlich zu den in Bild 6.8 dargestellten Größen sind hier noch die zellspezifische Acetataufnahmerate $q_{S3/X}$ sowie die Drehzahl N_{St} aufgezeigt. Auf die Darstellung der Acetatbildungsrate $q_{P3/X}$ wurde aus Gründen der Übersichtlichkeit verzichtet. Sie ist gemäß Gl. (6.32) der zellspezifische Glukoseaufnahmerate $q_{S1/X_{ov}}$ des Overflow-Metabolismus direkt proportional.

Sobald die vorgelegte Glukose nahezu vollständig verbraucht ist, beginnen die Zellen Acetat zu metabolisieren. Durch den geringeren Sauerstoffbedarf steigt die Gelöstsauerstoffkonzentration an und die pO_2 -Drehzahlregelung wird durch die Prozessautomatisierung ausgeschaltet, um die sich anschließende Diauxie besser beobachten zu können.

Durch das weitere Ansteigen der Gelöstsauerstoffkonzentration wird bei einem pO_2 -Wert von mehr als 80% die Fed-Batch Phase initialisiert und die Glukosezufütterung sowie die pO_2 -Drehzahlregelung gestartet. Bedingt durch den höheren Sauerstoffbedarf sowie das nicht stossfreie Umschalten des Reglers kommt es zwischen $t = 5.55$ h und $t = 5.62$ h zu einer kurzfristigen Sauerstofflimitierung, die einen erneuten Overflow-Metabolismus verursacht. Sobald die ausreichende Sauerstoffversorgung der Zellen sichergestellt ist, werden Glukose und Acetat parallel abgebaut, bis letzteres bei $t = 6.3$ h vollständig metabolisiert ist

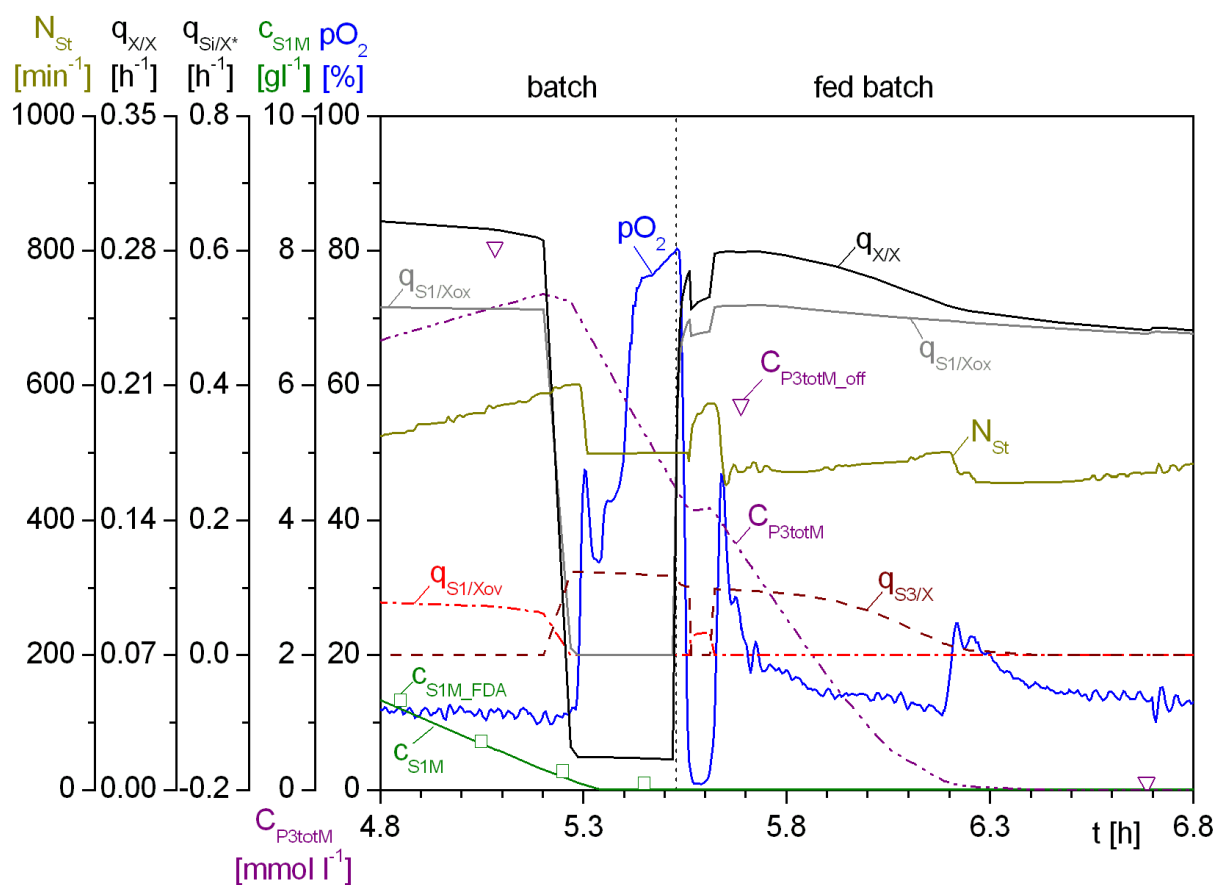


Bild 6.9: Diauxie zwischen den Prozessphasen Batch und Fed-Batch (CK4407)

6.3.2 Segregierter Modellansatz für die Zielproteinbildung

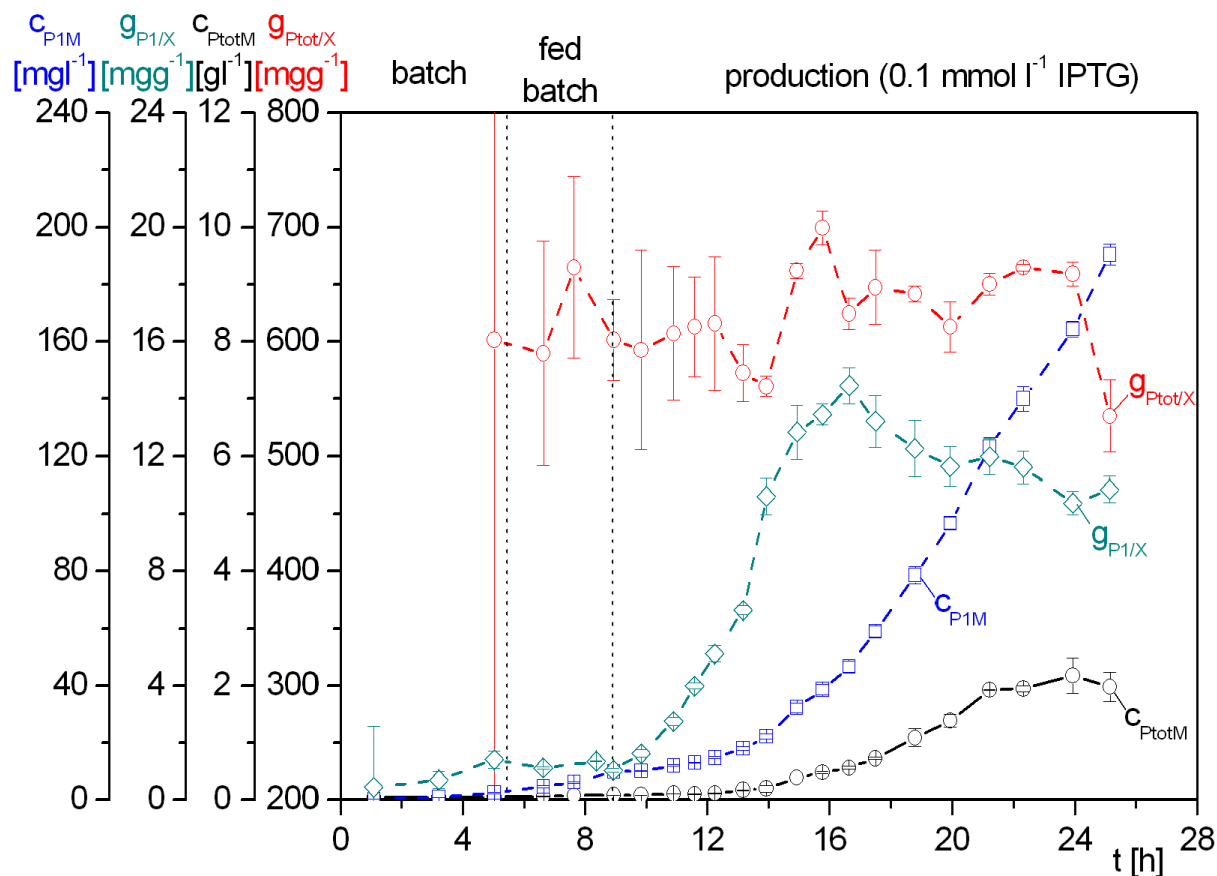


Bild 6.10: Exemplarischer Verlauf von Ziel- (P1) und Gesamtproteingehalt (Ptot) sowie deren entsprechende Konzentrationen c_{P1M} und c_{PtotM} im Medienüberstand (CK4407)

Da biologische Prozesse komplexe interne Regulationsmechanismen aufweisen, die oftmals nichtlineares Verhalten zeigen, sind unstrukturierte Modelle in ihrem Gültigkeitsbereich oft stark limitiert. Eine ausreichende Beschreibung der beobachteten Phänomene ist daher häufig mit diesen Modellen nicht möglich [Heine (2000)]. Auch in dem hier vorliegenden Fall konnten mit Hilfe eines unstrukturierten Modellansatzes keine zufriedenstellende Darstellung der Zielproteinexpression erreicht werden. Daher wurde ein segregiertes Modell aufgestellt. Da segregierte und strukturierte Ansätze die Mikroorganismen in unterschiedliche Gruppen unterteilen, können mit ihnen verschiedene physiologische Zellzustände zur gleichen Zeit berücksichtigt werden [Zheng *et al.* (2005)].

Auffällig in den in Bild 6.10 dargestellten Offline-Prozessdaten ist vor allem, dass ein nicht unerheblicher Teil des Ziel- sowie des Gesamtproteins im Medienüberstand nachgewiesen werden konnte. Die GFP-Konzentration c_{P1M} in der Medienphase nimmt dabei im Vergleich zur Gesamtproteinkonzentration c_{PtotM} im Medienüberstand im Verhältnis zu den entsprechenden zellinternen Gehältern $g_{P1/X}$ wesentlich stärker zu. Dies liegt in der hohen Stabilität des Zielproteins gegenüber den Umgebungsbedingungen wie z.B. Proteasen begründet, welche die zell-eigenen Proteine zersetzen [Yang *et al.* (1996)][Ward *et al.* (1982)].

Als Ursache für die Proteinakkumulation im Überstand wurde eine permanente Zelllyse postuliert. Die zellspezifische Lyserate $q_{X/Xlys}$,

$$q_{X/Xlys}(t) = q_{X/Xlys_max} \cdot \left(\frac{N_{St}(t) - N_{Stmin}}{N_{Stmax} - N_{Stmin}} \right)^3, \quad (6.44)$$

ist dabei von den Scherkräften im Reaktor abhängig, die sich proportional zum Leistungseintrag und damit zur dritten Potenz der Drehzahl verhalten [Levin (2005)]. Die Lyse setzt erst bei einer unteren Rührerdrehzahl N_{Stmin} ein und erreicht bei voller Drehzahl N_{Stmax} ihren Maximalwert $q_{X/Xlys_max}$.

Die Änderung der lysierten Zelldichte c_{XLlys} ,

$$\dot{c}_{XLlys}(t) = \left(-\frac{F_{in}(t) - F_A(t)}{V_L(t)} \right) \cdot c_{XLlys}(t) + q_{X/Xlys}(t) \cdot c_{XLact}(t), \quad (6.45)$$

wird damit von Konzentrations- und Verdünnungsvorgängen sowie dem Produkt aus Lyserate und der aktiven Zelldichte c_{XLact} beeinflusst.

Damit muss die Zelldichtebilanz aus Gl. (6.4),

$$\dot{c}_{XLact}(t) = \left(-\frac{F_{in}(t) - F_A(t)}{V_L(t)} + q_{X/X}(t) - q_{X/Xlys}(t) \right) \cdot c_{XLact}(t), \quad (6.46)$$

umgeschrieben werden, so dass die Lyserate Berücksichtigung findet.

Zur Berechnung der gravimetrisch bestimmbaren Biotrockenmasse c_{XL} ,

$$c_{XL}(t) = c_{XLact}(t) + 0.5 \cdot c_{XLlys}(t), \quad (6.47)$$

wird die aktive Zelldichte vollständig und die lysierte Zelldichte mit einer Gewichtung von 50 % herangezogen.

Die Trockensubstanz von Bakterien besteht hauptsächlich aus Polymeren und setzt sich ungefähr aus 50% Proteinen, 10–20% Zellwandbestandteilen, 10–20% RNA, 3–4% DNA und ca. 10% Lipiden zusammen [Schlegel (1992)]. Da von lysierten Zellen nur noch die unlösliche Fraktion der Zellbestandteile, nicht aber die löslichen Bestandteile des Zytoplasmas, in die Biomassebestimmung eingehen, wurde hier ein Korrekturfaktor von 50% eingeführt.

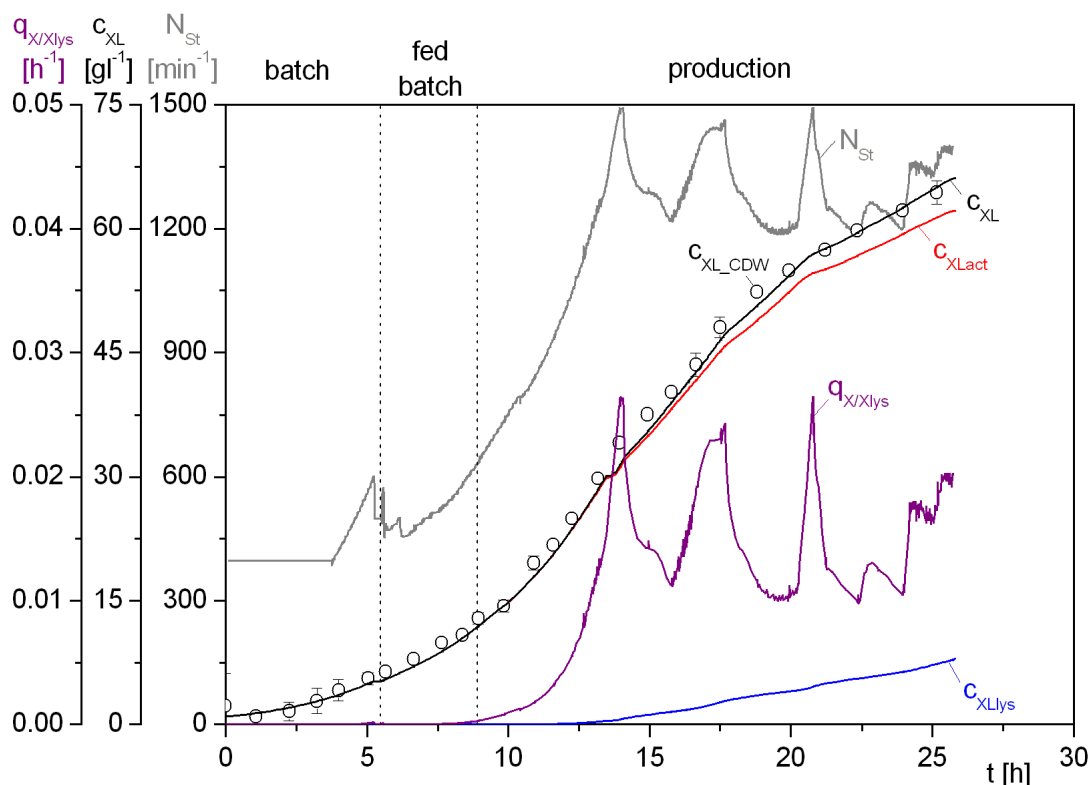


Bild 6.11: Verlauf der offline bestimmten Zelldichte c_{XL_CDW} sowie der simulierten Zelldichte c_{XL} mit dem aktiven Bestandteil c_{XLact} und dem lysierten Anteil c_{XLlys} sowie der zellspezifischen Lyserate $q_{X/Xlys}$ und der Rührerdrehzahl N_{St} (CK4407)

Die simulierten Daten für die aktive und lysierte Zelldichte, sowie die offline bestimmte Biotrockenmasse sind exemplarisch in Bild 6.11 dargestellt. Der Verlauf der Lyserate spiegelt dabei den prinzipiellen Verlauf der Rührerdrehzahl gemäß Gl. (6.44) wieder. Am Ende des Prozesses entspricht die lysierte Zelldichte ungefähr 8 gl^{-1} und damit etwas mehr als 10% der gesamten Zellen.

Für die Modellierung der Zielproteinexpression wird zunächst der aktive Transport des IPTGs aus der Medienphase in die Zelle in Anlehnung an Bentley *et al.* (1991) modelliert. Dafür wird die zellspezifische IPTG-Aufnahmerate $q_{Ind/X}$,

$$q_{Ind/X}(t) = y_{Ind/X} \cdot \frac{C_{IndM}(t)}{K_{Ind} + C_{IndM}(t)} \cdot q_{X/X}(t), \quad (6.48)$$

eingeführt, die von der extrazellulären IPTG-Konzentration im Medium C_{IndM} und der zellspezifischen Wachstumsrate $q_{X/X}$ sowie den Parametern $y_{Ind/X}$ und K_{Ind} abhängig ist und ebenfalls eine Sättigungskinetik darstellt.

Die Änderung des zellinternen molaren IPTG-Gehalts $G_{Ind/X}$,

$$\dot{G}_{Ind/X}(t) = \frac{1}{M_{Ind}} \cdot q_{Ind/X}(t) - q_{X/X}(t) \cdot G_{Ind/X}(t), \quad (6.49)$$

wird dabei von dem in Gl. (6.48) beschriebenen aktiven Transport des IPTGs in die Zelle sowie vom Verdünnungseffekt durch das Zellwachstum beeinflusst.

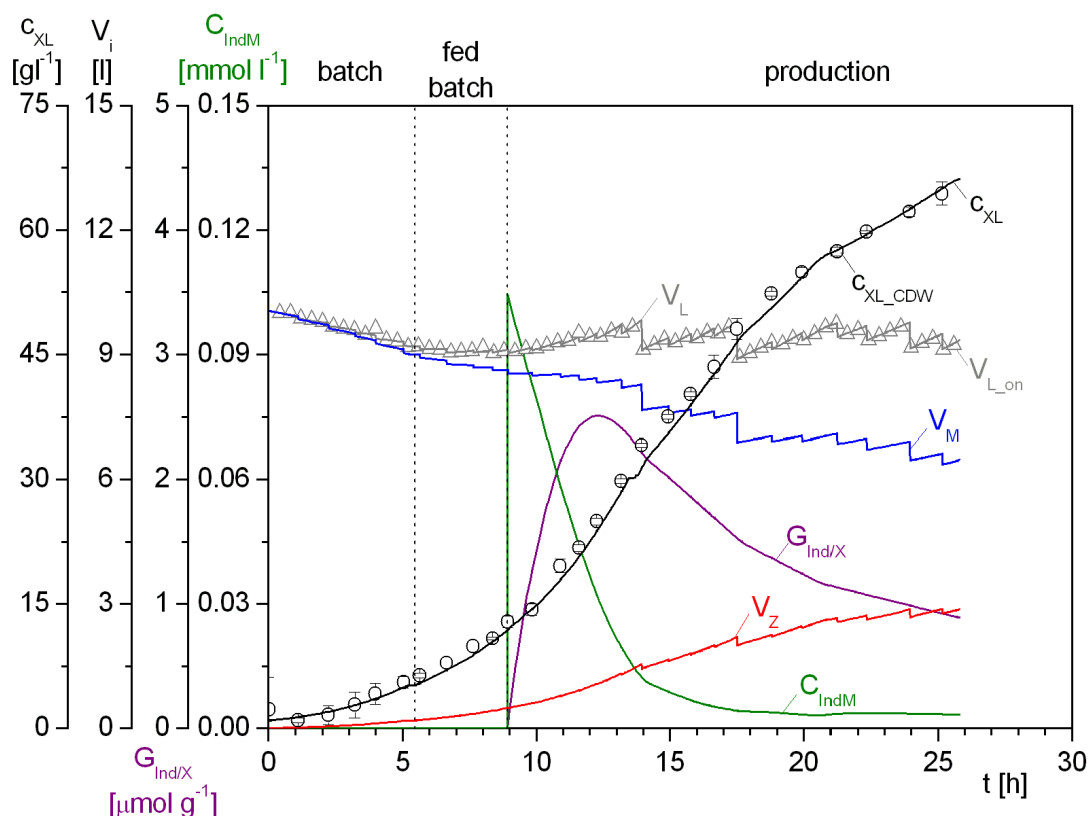


Bild 6.12: Simulierter Verlauf der Zelldichte c_{XL} , der Volumina des Dreiphasenmodells V_L , V_M und V_Z sowie der zellexternen Induktorkonzentration C_{IndM} und des zellinternen Induktorgehalts $G_{Ind/X}$ (CK4407)

Die Bilanz der in Gl. (6.48) benötigte Induktorkonzentration C_{IndM} in der Medienphase,

$$\begin{aligned} \dot{C}_{IndM}(t) = & \frac{F_{Ind}(t)}{V_M(t)} \cdot C_{IndR3} - \left(\frac{F_{in}(t)}{V_M(t)} - \frac{\alpha_{Z/X}}{\rho_Z} \cdot \frac{q_{X/X}(t)}{V_M(t)} \cdot V_L(t) \cdot c_{XLact}(t) \right) \cdot C_{IndM}(t) \\ & - \frac{1}{M_{Ind}} \cdot \frac{V_L(t)}{V_M(t)} \cdot q_{Ind/X}(t) \cdot c_{XLact}(t) + \frac{V_L(t)}{V_M(t)} \cdot q_{X/Xlys}(t) \cdot G_{Ind/X}(t) \cdot c_{XLact}(t) , \end{aligned} \quad (6.50)$$

ist eine Funktion des zugeführten Induktors, der Verdünnungs- sowie Konzentrationseffekte, des Induktortransports in die Zelle sowie der Freisetzung von IPTG durch die postulierte Zelllyse. Exemplarisch ist der Verlauf von zellexterner Induktorkonzentration und zellinternem Gehalt in Bild 6.12 dargestellt.

Nach der Beschreibung des zellinternen Induktorgehaltes $G_{Ind/X}$ kann dieser zur Formulierung einer Produktkinetik herangezogen werden. Zusätzlich dazu wird ein wachstumsabhängiger Ansatz gewählt, der eine basale GFP-Produktion bei nicht induzierten Zellen einschließt und an Leveau and Lindow (2001) angelehnt ist.

Die zellspezifische Produktbildungsrate $q_{P1/X}$,

$$q_{P1/X}(t) = q_{P1/Xmax} \cdot \left(\frac{q_{X/X}(t)}{\mu_{1max}} \right)^\psi \cdot \left[f_{bas} + (1 - f_{bas}) \cdot \left(\frac{G_{Ind/X}(t)}{K_{GInd} + G_{Ind/X}(t)} \right)^\zeta \right], \quad (6.51)$$

wird dabei von der zellspezifischen Wachstumsrate $q_{X/X}$ sowie dem zellinternen Induktorgehalt $G_{Ind/X}$ beeinflusst.

Der Parameter f_{bas} steuert dabei die basale Transkription vor der Induktion. Diese tritt dann auf, wenn kein oder nicht genügend aktiver Repressor durch die Zellen gebildet wird [Friebs (1999)]. Der zellinterne Zielproteingehalt zum Induktionszeitpunkt lag bei dem verwendeten DH5 α -Stamm über alle Kultivierungen bei $2.2 \text{ mg g}^{-1} \pm 0.28 \text{ mg g}^{-1}$. Das entspricht in etwa 15% des maximal erreichten Zielproteingehalts nach der Induktion.

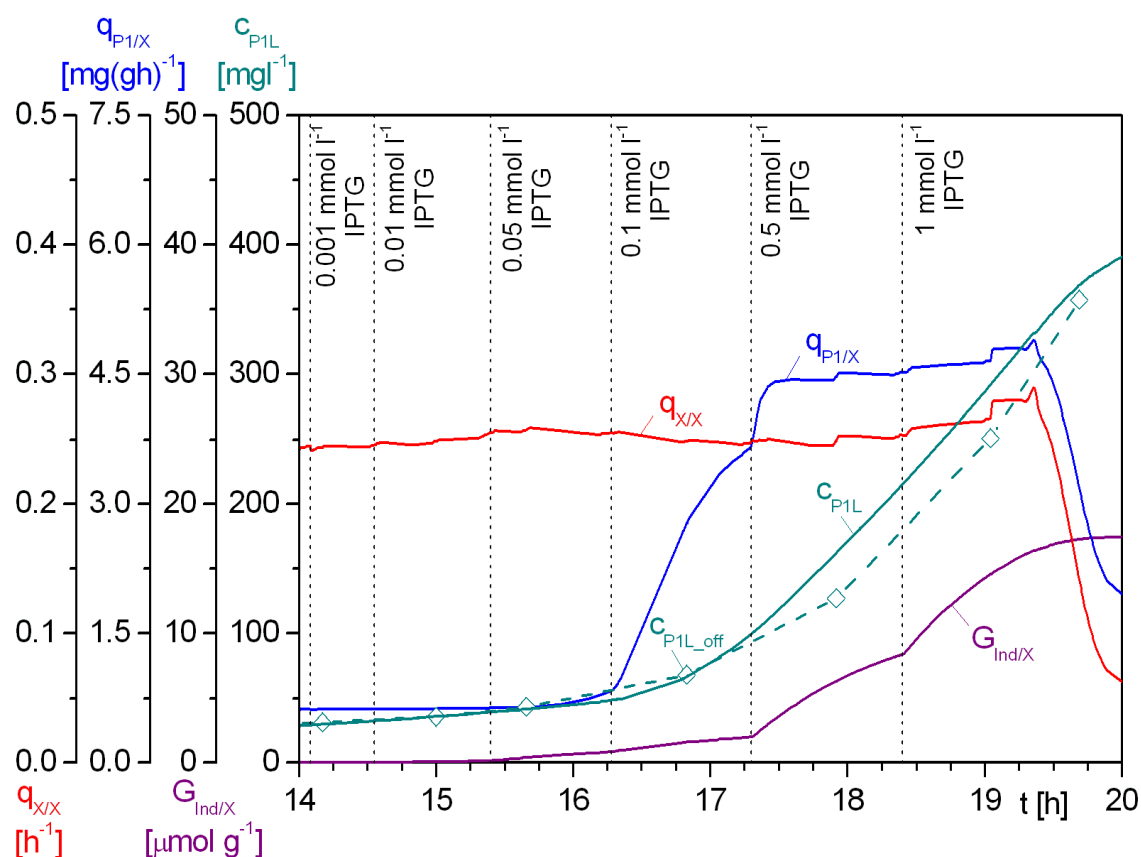


Bild 6.13: Verlauf von Wachstumsrate $q_{X/X}$, zellspezifischer Produktbildungsrate $q_{P1/X}$, zellinternem Induktorgehalt $G_{Ind/X}$ und der Zielproteinkonzentration c_{P1M} in der Medienphase in einer Produktionsphase mit stufenweiser Erhöhung der Induktorkonzentration (CK4907)

Die Abhängigkeit der Zielproteinexpression vom zellinternen Induktorgehalt $G_{Ind/X}$ konnte mit Hilfe von zwei Kultivierungen mit stufenweiser Erhöhung der Induktorkonzentration bestimmt werden. Wie in Bild 6.13 exemplarisch dargestellt, wurden dafür bei konstanter Wachstumsrate sechs unterschiedliche IPTG-Konzentrationen eingestellt. Dabei wird deutlich, dass die Zielproteinexpression erst ab einem zellinternen Gehalt von ca. $1 \mu\text{mol g}^{-1}$ die basale Expression deutlich überschreitet. Ab einem Gehalt von ca. $3 \mu\text{mol g}^{-1}$ kann dann keine gravie-

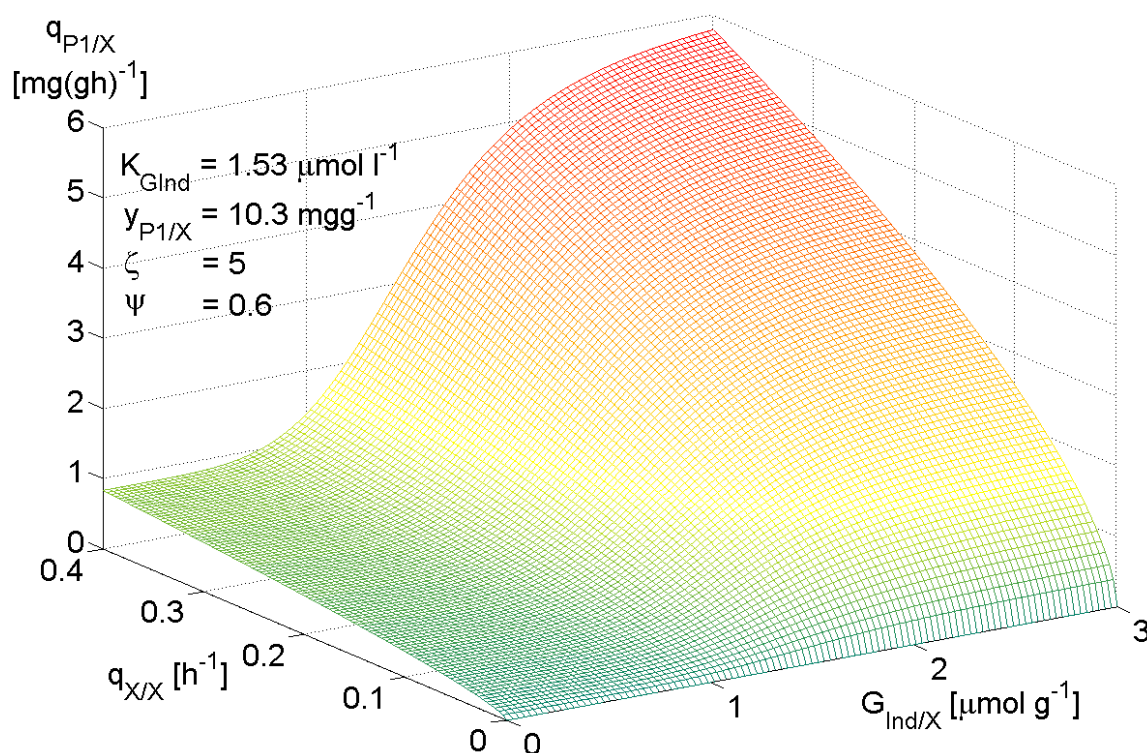


Bild 6.14: Die zellspezifische Produktbildungsrate $q_{P1/X}$ in Abhängigkeit von der Wachstumsrate $q_{X/X}$ und des zellinternen Induktorgehaltes $G_{Ind/X}$ nach Gl. (6.51)

rende Steigerung der Produktbildungsrate $q_{P1/X}$ durch eine Erhöhung des IPTG-Gehaltes mehr erreicht werden.

Die drei Parameter f_{bas} , K_{GInd} und ζ in Gl. (6.51) dienen dabei der Anpassung der Produktkinetik an den zellinternen Induktorgehalt. Die Zelltoxizität des Induktors IPTG bei hohen Konzentrationen wird durch diese Kinetik nicht berücksichtigt.

Um zusätzlich eine nichtlineare Abhängigkeit der Produktbildung von der zellspezifischen Wachstumsrate zu simulieren, wurde der Exponent ψ in Gl. (6.51) eingeführt. Die sich dadurch ergebende Kinetik in Abhängigkeit der zellspezifischen Wachstumsrate $q_{X/X}$ sowie des zellinternen IPTG-Gehalts $G_{Ind/X}$ ist in Bild 6.14 aufgezeigt.

Die Änderung des zellinternen Zielproteingehalts $g_{P1/X}$,

$$\dot{g}_{P1/X}(t) = q_{P1/X}(t) - q_{X/X}(t) \cdot g_{P1/X}(t), \quad (6.52)$$

ist durch die Differenz von zellspezifischer Produktbildungsrate $q_{P1/X}$ (Gl. (6.51)) und der Ausdünnung durch den Zellzuwachs gegeben.

Die GFP-Akkumulation im Medienüberstand \dot{c}_{P1M} ,

$$\dot{c}_{P1M}(t) = \frac{V_L(t)}{V_M(t)} \cdot q_{X/Xlys}(t) \cdot g_{P1/X}(t) \cdot c_{XLact}(t) - \frac{F_{in}(t)}{V_M(t)} \cdot c_{P1M}(t), \quad (6.53)$$

kann dann über die in Gl. (6.44) postulierte Zellyserate sowie die Verdünnung durch Zufütterung beschrieben werden.

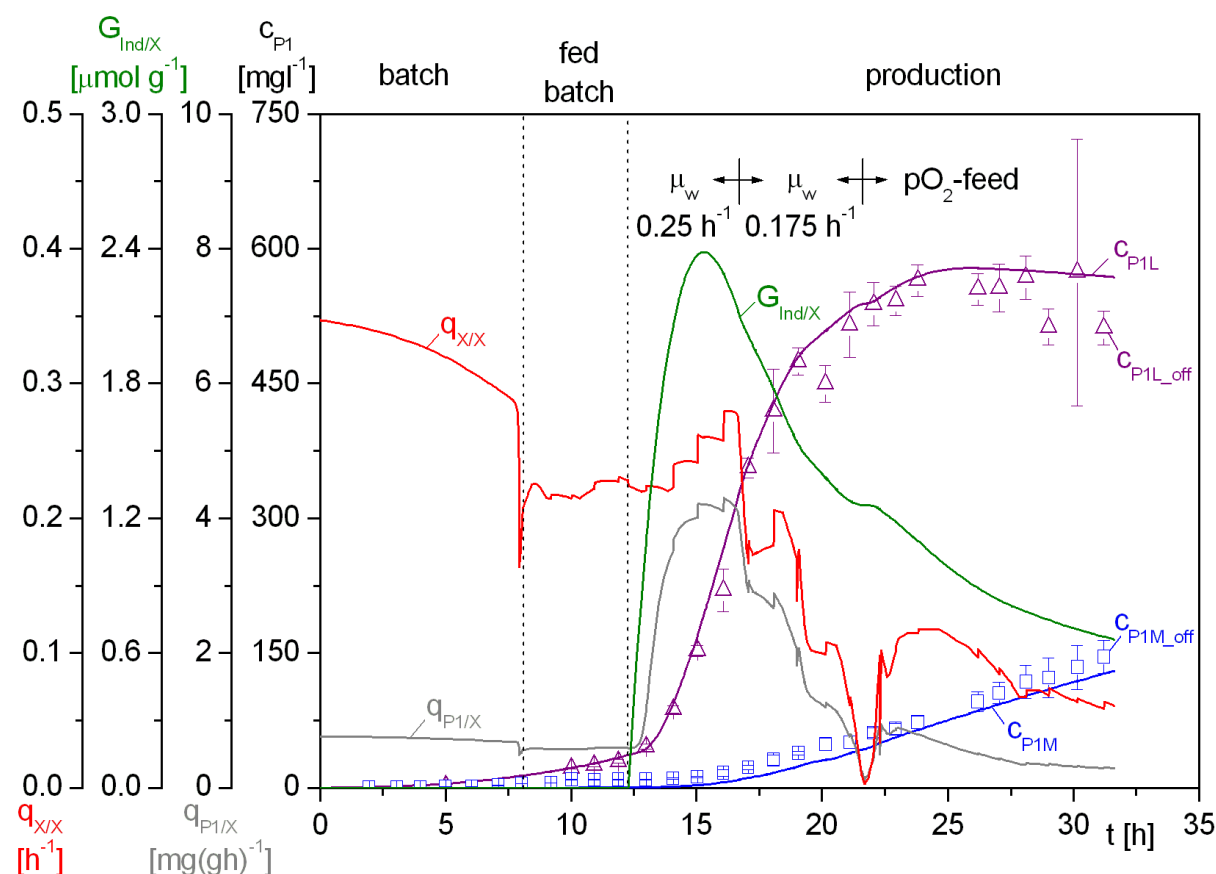


Bild 6.15: Exemplarischer Verlauf eines Produktionsprozesses mit der Zielproteinkonzentration in der Reaktionsphase c_{P1L} und in der Medienphase c_{P1M} , des zellinternen Induktorgehalts $G_{Ind/X}$, der zellspezifischen Wachstumsrate $q_{X/X}$ und der zellspezifischen Produktionsrate $q_{P1/X}$ (CK0707)

Die mittels Inline-Fluoreszenzmessung online beobachtbare Gesamtkonzentration des Zielproteins in der Reaktionsphase c_{P1L} ,

$$c_{P1L}(t) = g_{P1/X}(t) \cdot c_{XLact}(t) + \frac{V_M(t)}{V_L(t)} \cdot c_{P1M}(t), \quad (6.54)$$

wird somit durch den zellinternen Gehalt über $g_{P1/X}$ und c_{XLact} und das in der Medienphase akkumulierte Zielprotein c_{P1M} bestimmt.

Bild 6.15 zeigt den exemplarischen Verlauf eines typischen Expressionsprozesses sowie die entscheidenden Einflussfaktoren.

Bedingt durch die basale Proteinexpression kommt es schon vor der Induktion zur GFP-Bildung, die über die zellspezifische Reaktionsrate $q_{P1/X}$ gekennzeichnet ist. Diese wird auch durch die zellspezifische Wachstumsrate $q_{X/X}$ gesteuert (Gl. (6.51)).

Nach der Zugabe von 0.1 mmol l^{-1} IPTG bei $t = 12.4 \text{ h}$ baut sich der zellinterne Induktorgehalt $G_{Ind/X}$ auf und damit steigt die Produktionsleistung der Zellen gravierend an. Bis zur Reduktion der vorgegebenen zellspezifischen Wachstumsrate von 0.25 auf 0.175 h^{-1} bei $t = 17.6 \text{ h}$ exprimiert *Escherichia coli* das Zielprotein mit einer zellspezifischen Produktbildungsrate von 4 mg(gh)^{-1} .

Bedingt durch die geringere Wachstumsrate sowie den abfallenden Induktorgehalt sinkt die Produktionsleistung der Zellen im Folgenden deutlich ab und führt zu einer Stagnation der Gesamtzielproteinkonzentration c_{P1L} gegen Ende des Prozesses.

Sowohl die durch die Zellyse bedingte GFP-Konzentration im Medienüberstand c_{P1M} als auch die GFP-Konzentration in der gesamten Reaktionsphase L werden durch das zugrunde liegende Modell gut wiedergegeben.

In Bild 6.16 ist ein weiterer Kultivierungsverlauf dargestellt. Die Zielproteinkonzentrationen c_{P1L} wird auch hier während der Produktionsphase gut durch das Modell wiedergegeben. Die Simulationsdaten der Proteinexpression vor der Induktion weisen dabei in allen Simulationen die größten Fehler auf. Dies ist auf eine geringere Anzahl an Messdaten zur Parameteridentifikation für diese Phase, aber auch auf die größeren Fehler bei der Offline-Quantifizierung im unteren Konzentrationsbereich zurückzuführen.

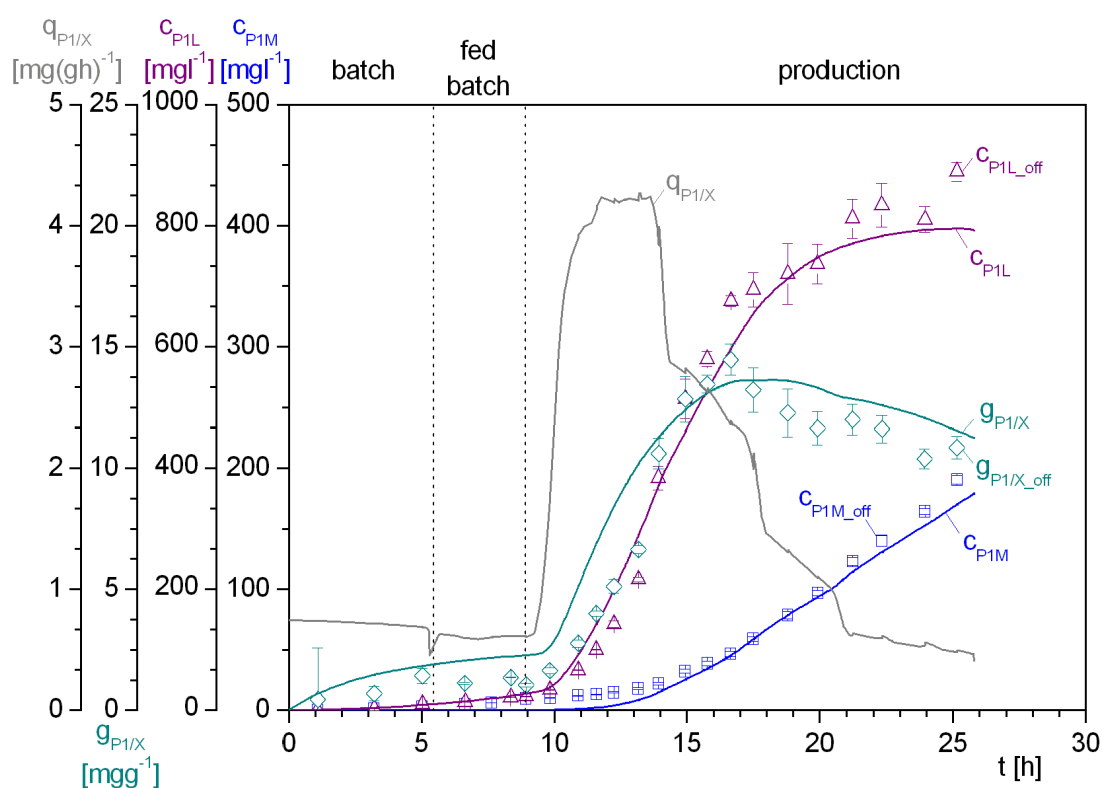


Bild 6.16: Verlauf der Offline-Messwerte sowie der simulierten Daten für die Zielprotein-konzentration im Medienüberstand c_{P1M} , die Gesamtzielproteinkonzentration im Reaktionsvolumen c_{P1L} sowie für die zellspezifische Produktbildungsrate $q_{P1/X}$ und des zellinternen Produktgehaltes $g_{P1/X}$ während der Kultivierung CK4407

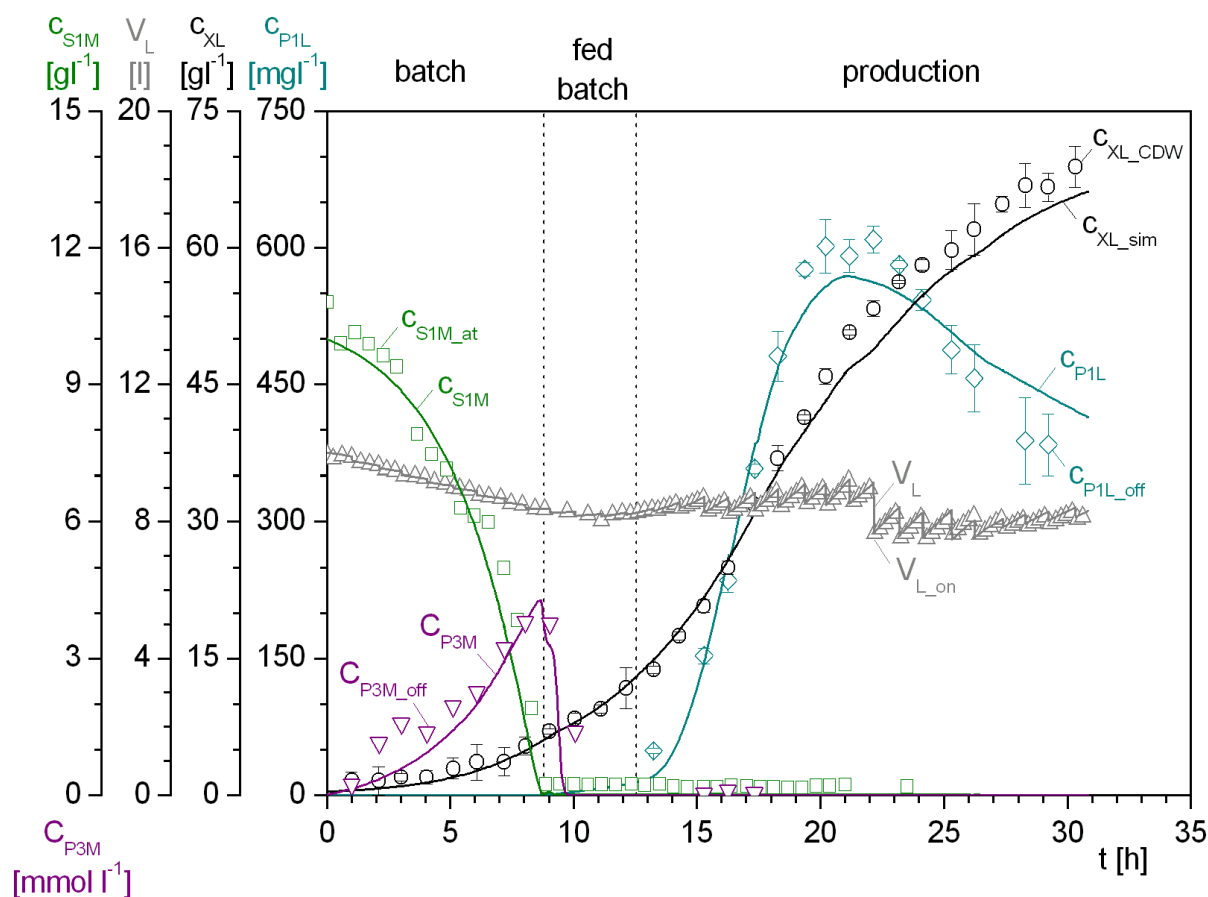


Bild 6.17: Exemplarischer Verlauf von Mess- und Simulationsdaten für die Zelldichte c_{XL} und die Zielproteinkonzentration c_{P1L} im Reaktionsvolumen V_L sowie für die Glukosekonzentration c_{S1M} und die Acetatkonzentration c_{P3totM} in der Medienphase M während der Kultivierung CK0607

Abschließend sind in Bild 6.17 wichtige Offline-, Atline- und Online-Messdaten, die durch das gewählte Modell beschrieben werden, dargestellt. Die Verläufe der Glukosekonzentration c_{S1M} und der Acetatkonzentration c_{P3totM} werden dabei sehr gut durch die Simulation wiedergegeben.

Die Zielproteinkonzentration c_{P1L} ist im Rahmen der Messgenauigkeit zufriedenstellend modelliert. Insbesondere für hohe Produkttiter ist der Fehler der Proteinquantifizierung aufgrund von vielen Verdünnungsschritten hoch und an eine akkurate Biotrockenmassebestimmung gekoppelt. Es ist daher nicht zu klären, ob die Abweichungen zwischen Mess- und Simulationsdaten hauptsächlich durch das Modell oder aber die Offline-Bestimmung der entsprechenden Komponenten verursacht wurde.

Grundsätzlich gibt das formulierte Modell die Kultivierungsergebnisse sehr gut wieder und kann damit zu einer Prozesssynthese und zu einer Optimierung von Fermentationsstrategien herangezogen werden.

Die mit den Simulationen identifizierten sowie aus der Literatur entnommenen Reaktionsparameter sind in den folgenden Tabellen 6.2, 6.3 und 6.4 zusammengefasst.

Tabelle 6.2: Übersicht der verwendeten Parameter der Simulation

| Parameter | Einheit | Bezeichnung | Wert | Herkunft | Lit.-wert | Literatur |
|----------------|------------------------|---|----------------------|------------------------|----------------------|------------------------------|
| $\alpha_{Z/X}$ | [-] | Verhältnis von Biofeucht- zu Biotrockenmasse | 4.67 | experimentell bestimmt | - | |
| ρ_Z | [g l ⁻¹] | Dichte der Biofeuchtphase | 1005 | experimentell bestimmt | - | |
| μ_{1max} | [h ⁻¹] | maximale Wachstumsrate auf Glukose (S1) | 0.326 | identifiziert | - | |
| μ_{3max} | [h ⁻¹] | maximale Wachstumsrate auf Acetat (S3) | 0.1 | Literatur | 0.1 | Peuker (2006) |
| ζ | [-] | Hill-Koeffizient der Zielprotein-kinetik | 5 | identifiziert | 2 | Leveau and Lindow (2001) |
| ψ | [-] | Koeffizient des Wachstumsterms der Zielprotein-kinetik | 0.6 | identifiziert | - | |
| f_{bas} | [-] | Parameter der basalen GFP-Produktion | 0.14 | identifiziert | - | |
| k_{S1m} | [h ⁻¹] | Parameter der induktionsabhängigen Maintenaneraterate | 0.0810 | identifiziert | - | |
| K_{GInd} | [mol l ⁻¹] | Limitierungskonstante der Zielprotein-kinetik | $1.53 \cdot 10^{-3}$ | identifiziert | $0.32 \cdot 10^{-3}$ | Leveau and Lindow (2001) |
| K_{Ind} | [mol l ⁻¹] | Limitierungskonstante des IPTG-Transports | $0.08 \cdot 10^{-3}$ | identifiziert | $0.04 \cdot 10^{-3}$ | Bentley <i>et al.</i> (1991) |
| K_{I13} | [mol l ⁻¹] | Inhibierungskonstante der Glukoseaufnahme durch die Acetatkonzentration | $37.7 \cdot 10^{-3}$ | identifiziert | - | |

Tabelle 6.3: Fortsetzung Übersicht der verwendeten Parameter der Simulation

| Parameter | Einheit | Bezeichnung | Wert | Herkunft | Lit.-wert | Literatur |
|------------------|-----------------|--|----------------------|-----------------|---------------------|----------------------------|
| k_{I31} | $[g\ l^{-1}]$ | Inhibierungskonstante der Acetataufnahme durch die Glukosekonzentration | $1.84 \cdot 10^{-3}$ | identifiziert | – | |
| K_{I03} | $[mol\ l^{-1}]$ | Inhibierungskonstante der Sauerstoffaufnahme durch die Acetatkonzentration | $88.7 \cdot 10^{-3}$ | identifiziert | – | |
| k_O | $[g\ l^{-1}]$ | Limitierungskonstante der Monod-Kinetik für die Sauerstoffaufnahme | $0.07 \cdot 10^{-3}$ | identifiziert | – | |
| k_{S1} | $[g\ l^{-1}]$ | Limitierungskonstante der Monod-Kinetik für die Glukoseaufnahme | $4.0 \cdot 10^{-3}$ | Literatur | $4.0 \cdot 10^{-3}$ | Cockshott and Bogle (1999) |
| K_{S3} | $[mol\ l^{-1}]$ | Limitierungskonstante der Monod-Kinetik für die Acetataufnahme | $0.81 \cdot 10^{-3}$ | identifiziert | – | |
| N_{Stmax} | $[min^{-1}]$ | maximale Rührerdrehzahl zur Berechnung der Lyserate | 1500 | Reaktor bedingt | – | |
| N_{Stmin} | $[min^{-1}]$ | minimale Rührerdrehzahl zur Berechnung der Lyserate | 390 | identifiziert | – | |
| q_X/X_{lysmax} | $[h^{-1}]$ | maximale Zellyserate | 0.03 | identifiziert | – | |
| q_O/X_m | $[h^{-1}]$ | Maintenanteanteil der zellspezifischen Sauerstoffaufnahme | 0.0442 | identifiziert | 0.02 | Peuker (2006) |
| q_{P1}/X_{max} | $[h^{-1}]$ | maximale zellspezifische Produktbildungsrate für das Zielprotein | $0.01 \cdot 10^{-3}$ | identifiziert | – | |

Tabelle 6.4: Fortsetzung Übersicht der verwendeten Parameter der Simulation

| Parameter | Einheit | Bezeichnung | Wert | Herkunft | Lit.-wert | Literatur |
|-----------------|------------|---|--------|---------------|-----------|------------------------------|
| q_{S1}/X_{m0} | $[h^{-1}]$ | induktionsunabhängiger Anteil der Glukose-Maintenancerate | 0.0147 | identifiziert | 0.04 | Xu <i>et al.</i> (1999) |
| $w_C/S1$ | [-] | Massenanteil von Kohlenstoff in Glukose | 0.4 | Stöchiometrie | 0.4 | Schneider (1999) |
| $w_C/S3$ | [-] | Massenanteil von Kohlenstoff im Acetat | 0.4 | Stöchiometrie | 0.4 | Schneider (1999) |
| w_C/X | [-] | Massenanteil von Kohlenstoff in den Zellen | 0.5 | Stöchiometrie | 0.5 | Schneider (1999) |
| y_{Ind}/X | [-] | Parameter des Induktortransportes | 4.2 | identifiziert | 3.3 | Bentley <i>et al.</i> (1991) |
| y_{S1}/O | [-] | Glukoseausbeutekoeffizient für Sauerstoff | 1.88 | Stöchiometrie | - | |
| y_X/O_{gr} | [-] | Zuwachsausbeutekoeffizient für Sauerstoff | 1.34 | identifiziert | 1.44 | Peuker (2006) |
| y_X/S_{lgr} | [-] | Zuwachsausbeutekoeffizient für Glukose | 0.536 | identifiziert | 0.52 | Peuker (2006) |
| $y_X/S3_{gr}$ | [-] | Zuwachsausbeutekoeffizient für Acetat | 0.2 | Literatur | 0.2 | Peuker (2006) |

7 Zusammenfassung

Die Zielsetzung dieser Arbeit war die Erprobung und Bewertung unterschiedlicher Verfahren zur Online-Überwachung von wichtigen Schlüsselgrößen in Kultivierungsprozessen mit *Escherichia coli*. In Anlehnung an die Empfehlungen der *Process Analytical Technology*-Initiative der US amerikanischen *Food and Drug Administration* wurden dafür unterschiedliche Prozessmessgeräte und Analysensysteme, multivariate Datenerfassungs- und Auswertewerkzeuge sowie die Entwicklung, Simulation und Parameteridentifikation mit dynamischen Prozessmodellen eingesetzt.

Zunächst wurden kommerziell erhältliche Analysensysteme in die Vorortautomatisierung des eingesetzten Bioreaktorsystems integriert. Anschließend erfolgte ein Vergleich und eine Beurteilung verschiedener Verfahren zur Online-Überwachung der Zelldichte sowie eine Validierung der verwendeten Atline-Systeme mittels Offline-Analysen.

Mit Hilfe der mathematischen Entwicklungsumgebung MATLAB[®] wurden die erzielten Messdaten einer geeigneten Datenvorverarbeitung unterzogen. Dieses umfasste eine Ausreisererkennung auf Basis des Grubbs-Tests sowie die Anwendung und Beurteilung verschiedener Datenfilter und Interpolationsverfahren.

Für die Online-Überwachung des Hauptsubstrates Glukose, der Zelldichte sowie des produzierten Reporterproteins GFP wurden, auf Basis des 2D-Fluoreszenzspektroskopes BioView[®], unterschiedliche datengetriebene Verfahren aus dem Fachgebiet der Chemometrik eingesetzt. Zur Aufdeckung linearer Korrelationen zwischen den interessierenden Prozessgrößen und den gewonnenen Fluoreszenzdaten erfolgte zunächst eine Korrelationsanalyse.

Auf Basis der gewonnenen Informationen aus der Korrelationsanalyse wurden dann lineare univariate sowie multivariate Modelle unter Verwendung der originalen, vorverarbeiteten Fluoreszenzspektren kalibriert und die Ergebnisse miteinander verglichen.

Im nächsten Schritt wurden die Hauptkomponentenanalyse sowie die unabhängige Komponentenanalyse auf die BioView[®]-Spektren angewendet, um zum einen den Eingangsdatenraum für die Modellkalibrierung zu verringern und zum anderen ein besseres Signal-Rausch Verhältnis zu erzielen.

Es wurde dabei gezeigt, dass die besten Approximationen mit einem Fehler von unter 10% mit Hilfe von multivariaten, linearen Modellen unter Verwendung der Original-Fluoreszenzspektren erzielt werden. Etwas größere Modellfehler ergaben sich bei der Verwendung von Hauptkomponenten und unabhängigen Komponenten als Eingangsgrößen. Diese zeigen jedoch einen wesentlich geringeren Rauschanteil in den Vorhersagen und zeichnen sich durch eine wesentlich kürzere Kalibrierzeit aus. Zwischen den Modellen auf Basis der Hauptkomponentenanalyse sowie der unabhängigen Komponentenanalyse konnten keine signifikanten Unterschiede festgestellt werden.

Für alle Prozessgrößen mussten die Kalibrierungen abschnittsweise für die Prozessphasen Batch, Fed-Batch und Produktion erfolgen, um verlässliche und aussagekräftige Vorhersagen zu erzielen.

Für die Erprobung künstlicher neuronaler Netze wurden dann die unabhängigen Komponenten sowie die Hauptkomponenten als Eingangsdaten gewählt, um den Zeitaufwand für das Netztraining in einem realistischen Rahmen zu halten. Die Netze lieferten mit Vorhersagefehlern von 10% gute Vorhersagen. Dabei wurde für jede Prozessgröße je ein Netz für die Abschnitte Batch/Fed-Batch sowie die Produktionsphase trainiert.

Die neuronalen Netze bieten den Vorteil, dass sie sich mit jedem weiteren zur Verfügung stehenden Datensatz weitertrainieren lassen. Sie bieten sich also insbesondere in einem industriellen Umfeld an, in dem viele Kultivierungen mit dem gleichen Produktionsstamm durchgeführt werden, da die Güte der Vorhersage mit der Vergrößerung des Eingangsdatenraumes weiter steigt.

Im letzten Teil der Arbeit wurde ein Modell zur mathematischen Beschreibung des dynamischen Bioreaktionsprozesses entwickelt. Dabei wurde neben der Modellierung des Zellwachstums und der Glukoseaufnahme die Acetatbildung unter sauerstoffunlimitierten und sauerstofflimitierten Bedingungen implementiert.

Die Beschreibung der Zielproteinbildung erfolgte mit Hilfe eines wachstumsabhängigen Anteils sowie über den zellinternen Induktorgehalt. Dafür musste der Transport des verwendeten chemischen Induktors IPTG in die Zelle modelliert werden. Zur mathematischen Beschreibung der Zielproteinakkumulation in der Medienphase wurde zusätzlich eine scherkraftabhängige Lyserate eingeführt.

Es ergaben sich gute Übereinstimmungen zwischen den Messwerten des Bioreaktionsprozesses und dem erstellten Modell. Dieses kann damit im Folgenden zur Prozesssynthese und zur Optimierung von Kultivierungsabläufen herangezogen werden.

8 Anhang

8.1 Verwendete Medien und Reagenzien

8.1.1 Vorkultur- und Batchmedium

Tabelle 8.1: Zusammensetzung des verwendeten M9 Minimal-Mediums für die Vorkultur und den Batch nach Riesenberg *et al.* (1991)

| Komponente | Konzentration [gl⁻¹] |
|--|--|
| Glukose | 10.0 |
| MgSO ₄ · 7 H ₂ O | 1.2 |
| K ₂ H ₂ PO ₄ | 13.3 |
| (NH ₄) ₂ HPO ₄ | 3.6 |
| Citronensäure | 1.7 |
| | [mg⁻¹] |
| Fe-(III)-Citrat | 100.0 |
| CoCl ₂ · 6H ₂ O | 2.5 |
| MnCl ₂ · 4H ₂ O | 15.0 |
| CuCl ₂ · 2H ₂ O | 1.5 |
| H ₃ BO ₃ | 3.0 |
| Na ₂ MoO ₄ · 2H ₂ O | 2.5 |
| Zn(CH ₃ COO) ₂ · 2H ₂ O | 13.0 |
| EDTA | 8.4 |
| Ampicillin | 100 |
| ThiaminHCl | 4.5 |

Glukose und MgSO₄ werden getrennt sterilisiert. Thiamin und Ampicillin aufgrund ihrer Temperaturempfindlichkeit sterilfiltriert.

Die Lösung der Salze wird vor der Sterilisation mit NaOH auf pH 6.3 eingestellt. Nach der Sterilisation erfolgt die Einstellung des pH-Wertes auf 6.8. Durch dieses Vorgehen werden Salzprecipitationen verhindert. Anschließend werden alle Lösungen vereint.

8.1.2 Feedmedium

Tabelle 8.2: Zusammensetzung des verwendeten Zufüttermediums nach Riesenberg *et al.* (1991)

| Komponente | Konzentration [gl⁻¹] |
|--|--|
| Glukose | 500 |
| MgSO ₄ · 7 H ₂ O | 20 |
| | [mg⁻¹] |
| Fe-(III)-Citrat | 40.0 |
| CoCl ₂ · 6H ₂ O | 4.0 |
| MnCl ₂ · 4H ₂ O | 15.0 |
| CuCl ₂ · 2H ₂ O | 2.5 |
| H ₃ BO ₃ | 5.0 |
| Na ₂ MoO ₄ · 2H ₂ O | 4.0 |
| Zn(CH ₃ COO) ₂ · 2H ₂ O | 16.0 |
| EDTA | 13.0 |

Glukose, MgSO₄ sowie die Salzlösungen werden getrennt sterilisiert, um Salzausfällungen zu vermeiden. Eine pH-Wert Korrektur des nicht gepufferten Mediums entfällt. Es ist auf eine ausreichende Durchmischung der Komponenten auf Grund der hohen Viskosität der Glukose-lösung zu achten.

8.1.3 Reagenzien für die Ammoniumbestimmung mit der TAS2000

Tabelle 8.3: Reagenz 1

| Komponente | Konzentration [gl⁻¹] |
|--|--|
| Na ₃ C ₆ H ₅ O · 2 H ₂ O | 40.0 |

Tabelle 8.4: Reagenz 2

| Komponente | Konzentration [gl⁻¹] |
|---|--|
| NaC ₇ H ₅ O ₃ | 34.0 |
| Na ₂ [Fe(CN) ₅ NO] · 2 H ₂ O | 0.40 |

Tabelle 8.5: Reagenz 3

| Komponente | Konzentration [gl^{-1}] |
|--|------------------------------------|
| $\text{NaC}_3\text{Cl}_2\text{N}_3\text{O}_3 \cdot 2 \text{H}_2\text{O}$ | 0.80 |
| NaOH | 10.0 |

8.2 Ergänzende Offline-Messtechnik

8.2.1 Optische Dichte Messung

Die Zellkonzentration wurde prozessbegleitend über eine Trübungsmessung bestimmt. Sie wurde mit einem UV/Vis-Photometer (Ultrospec 3000, Fa. Amersham Pharmacia Biotech, Buckinghamshire, UK) bei 600 nm gemessen, wobei die Proben so verdünnt wurden, dass die Extinktion zwischen 0.1 und 0.5 lag. Es wurden jeweils Doppelbestimmungen durchgeführt

8.2.2 Biotrockenmassebestimmung

Zur Ermittlung der Biotrockenmasse wurde 1 ml Bakteriensuspension in vorgetrockneten, abgewogenen Kunststoffcaps zentrifugiert (5 min bei 13000 Umin^{-1} , Centrifuge 5417C, Eppendorf AG, Hamburg) und die Pellets im Trockenschrank 24 h bei 80°C bis zur Gewichtskonstanz getrocknet. Es erfolgte eine Doppelbestimmung.

8.2.3 Zielproteinanalytik

Probenvorbereitung

Im ersten Schritt wurden die Zellpellets in PBS-Puffer (pH 7.0) resuspendiert und auf das Ausgangsvolumen von 1 ml eingestellt. Anschließend erfolgte eine Verdünnung auf eine Zelldichte von 3 gl^{-1} . Zur Quantifizierung der Anteile von löslichem und aggregiertem Zielprotein wurden die Zellen bei 50 W Leistungseintrag 3.5 min auf Eis sonifiziert (Labsonic U der Fa. Sartorius BBI Systems, Melsungen).

Die Trennung des löslichen vom unlöslichen Anteil erfolgte durch Zentrifugation bei 10000 g und 4°C für 30 min. Durch zwei Waschschrte mit PBS-Puffer wurde die vollständige Trennung der beiden Fraktionen gewährleistet.

Die Solubilisierung der im Pellet verbliebenen Inclusionbodyfraktion wurde unter Einwirkung von 8 M Harnstoff und 100 mM β -Mercaptoethanol in PBS Puffer für 1 h bei Raumtemperatur und leichtem Vortexen (1000 rpm) erreicht.

SDS-PAGE zur Proteinanalytik

Für die Zielproteinanalytik mittels SDS-Page wurden kommerzielle Criterion XT[®] 10-20% Tris-Tricine-Gradientengele der Fa. Bio-Rad Laboratories, München, sowie die zugehörigen Proben- und Laufpuffer, eingesetzt. Zur Größenbestimmung wurde ein Prestained Broad Band

Standard 7.1-209 kDa (Bio-Rad), zur Zielproteinquantifizierung ein GFP-Standard der Fa. Roche Diagnostics, Penzberg, mit einer klaren Bande bei 27 kDa, verwendet.

Die Färbung der Proteinbanden erfolgte unter Verwendung von Coomassie Brilliant Blue 250, die Digitalisierung der Gele mit einem MultiImager[®] und die Auswertung und Quantifizierung mit Hilfe der Software QuantityOne[®] (alle BioRad).

Offline-Fluoreszenzmessung von GFP

Zusätzlich zur Quantifizierung mittels SDS-Page wurden Fluoreszenzmessungen mit einem GENios Reader der Fa. Tecan Deutschland, Crailsheim durchgeführt. Die Steuerung und Programmierung des Multidetektionssystems für Fluoreszenz-, Absorptions- und Lumineszenzmessungen erfolgt über die zugehörige Software MAGELLAN[®]. Für die Messung mit einer Exzitation von 400 nm und einer Emission von 510 nm wurden schwarze 96 Well Mikrotiterplatten MaxiSorp[®] der Fa. Nunc, Wiesbaden eingesetzt.

Die Kalibrierung des Systems erfolgte über einen GFP-Standard der Fa. Roche Diagnostics, Penzberg, aus dem Lösungen mit definierten GFP-Konzentration im Bereich von 0 bis 10 mg l⁻¹ hergestellt wurden. In dem gewählten Messbereich zeigte die Quantifizierung mittels SDS-Page eine sehr gute Übereinstimmungen mit der Fluoreszenzmessung.

8.2.4 Ammoniumbestimmung

Zur Validierung der Atline-Analytik mit der TAS2000 wurde die Ammoniumkonzentration im Kulturüberstand mit einer gassensitiven Ammonium-Elektrode der Fa. Mettler Toledo, Gießen bestimmt. Dazu wurden 25 ml Probe 10 molare NaOH zugesetzt, um den Ammoniak aus der Lösung auszutreiben. Dieser passierte die gaspermeable Membran der Sonde und führte zu einer Spannung, die über ein pH-Meter Delta 345 (Mettler Toledo) detektiert wurde. Nach dem Nernst'schen Gesetz,

$$C_{\text{AltotM}} = C_{\text{Std}} \cdot 10^{\frac{U_{\text{Std}} - U_{\text{Probe}}}{b}}, \quad (8.1)$$

lassen sich mit Hilfe des Faktors b, der aus einer Kalibriergeraden mit bekannten Ammoniaklösungen (U_{Std}) ermittelt wurde und aus den gemessenen Spannungswerten U_{Probe} [V] die Konzentrationen C_{AltotM} [mol l⁻¹] errechnen.

8.2.5 Acetat- und Glukosemessung über enzymatische Testkits

Zur Bestimmung der unbekanntenen Acetatkonzentration sowie zur Kontrolle der durchgeführten Glukosemessungen mit der FDA und der HPLC wurden enzymatische UV-Tests (Nr. 0148 261 Acetat; NR. 0148 283 Glukose) der Fa. Roche Diagnostics, Mannheim, verwendet. Dabei wurden die enzymatischen Umsetzungen über die NADH Extinktion mit einem UV/Vis Photometer Ultrospec 3000 bei einer Wellenlänge von 340 nm detektiert.

8.2.6 Keimzahlbestimmung

Zur Bestimmung der Plasmidstabilität wurden während der Kultivierungen die Keimzahlen bestimmt. Alle 2-3 Stunden wurde eine Fermenterprobe entsprechend der aktuellen Zelldichte in PBS-Puffer (pH 7.0) verdünnt (10^6 - 10^{10}). Davon wurden 100 μ l auf LB-Medium enthaltende Petrieschalen ausplattiert, 24 Stunden bei 37°C inkubiert und im Anschluß ausgezählt. Es erfolgte jeweils eine Doppelbestimmung auf selektiven (Ampicillin) und nicht selektiven Platten.

8.3 Modellparameter der Chemometrischen Modelle

8.3.1 Univariate Modelle mit Originalvariablen

Tabelle 8.6: Parameter der besten univariaten Modelle mit Originalvariablen für die Zelldichte

| Variable | Prozessphase | Anpassung | Signal | Koef. | Wert |
|-----------------|--------------|--------------|-----------------------|----------------|-----------|
| c _{XL} | Batch | linear | - | a ₀ | -0.9306 |
| | | | S _{310/410} | a ₁ | 0.04660 |
| c _{XL} | Fed-batch | linear | - | a ₀ | 2.3384 |
| | | | S _{410/EmND} | a ₁ | -0.07730 |
| c _{XL} | Produktion | exponentiell | - | a ₀ | 4.298 |
| | | | S _{470/510} | a ₁ | -0.004192 |

Tabelle 8.7: Parameter der besten univariaten Modelle mit Originalvariablen für die Zielproteinkonzentration

| Variable | Prozessphase | Anpassung | Signal | Koef. | Wert |
|------------------|-----------------|-----------|----------------------|----------------|--------|
| c _{PIL} | Batch/Fed-batch | linear | - | a ₀ | -1.719 |
| | | | S _{390/570} | a ₁ | 0.5729 |
| c _{PIL} | Produktion | linear | - | a ₀ | -380.4 |
| | | | S _{270/510} | a ₁ | 9.001 |

Tabelle 8.8: Parameter des besten univariaten Modells mit Originalvariablen für die Glukosekonzentration

| Variable | Prozessphase | Anpassung | Signal | Koef. | Wert |
|------------------|--------------|-----------|----------------------|----------------|---------|
| c _{SIM} | Batch | linear | - | a ₀ | 19.33 |
| | | | S _{350/590} | a ₁ | -0.7923 |

8.3.2 Multivariate Modelle mit Originalvariablen

Tabelle 8.9: Parameter der besten multivariaten Modelle mit Originalvariablen für die Zelldichte

| Variable | Prozessphase | Anpassung | Signal | Koef. | Wert |
|-----------------|--------------|--------------|-----------------------|----------------|-----------|
| c _{XL} | Batch | linear | - | a ₀ | -0.9306 |
| | | | S _{310/410} | a ₁ | 0.04660 |
| | | | S _{410/470} | a ₂ | 0.04210 |
| c _{XL} | Fed-batch | linear | - | a ₀ | 2.3384 |
| | | | S _{410/EmND} | a ₁ | -0.07730 |
| | | | S _{ExND/330} | a ₂ | 0.1261 |
| | | | S _{ExND/590} | a ₂ | 0.07130 |
| c _{XL} | Produktion | exponentiell | - | a ₀ | 4.298 |
| | | | S _{470/510} | a ₁ | -0.004192 |
| | | | S _{490/530} | a ₂ | 0.004102 |

Tabelle 8.10: Parameter der besten multivariaten Modelle mit Originalvariablen für die Zielproteinkonzentration

| Variable | Prozessphase | Anpassung | Signal | Koef. | Wert |
|------------------|-----------------|-----------|----------------------|----------------|----------|
| c _{PIL} | Batch/Fed-batch | linear | - | a ₀ | -1.719 |
| | | | S _{390/570} | a ₁ | 0.5729 |
| | | | S _{410/530} | a ₂ | -0.08640 |
| c _{PIL} | Produktion | linear | - | a ₀ | -380.4 |
| | | | S _{270/510} | a ₁ | 9.001 |
| | | | S _{350/490} | a ₂ | 1.0613 |

Tabelle 8.11: Parameter des besten multivariaten Modells mit Originalvariablen für die Glukosekonzentration

| Variable | Prozessphase | Anpassung | Signal | Koef. | Wert |
|------------------|--------------|-----------|-----------------------|----------------|---------|
| c _{S1M} | Batch | linear | - | a ₀ | 19.33 |
| | | | S _{350/590} | a ₁ | -0.7923 |
| | | | S _{ExND/310} | a ₂ | -0.2512 |

Literaturverzeichnis

- [Anderson and Krummen 2002] ANDERSON, D.C. ; KRUMMEN, L.: Recombinant protein expression for therapeutic applications. In: *Curr. Opin. Biotechnol.* 13 (2002), S. 117–123
- [Arnold *et al.* 2003a] ARNOLD, S.A. ; CROWLEY, J. ; WOODS, N. ; HARVEY, L.M. ; MCNEIL, B.: In situ near infrared spectroscopy to monitor key analytes in mammalian cell cultivation. In: *Biotechnology and Bioengineering* 84 (2003), S. 13–19
- [Arnold *et al.* 2002] ARNOLD, S.A. ; GAESAKOO, R. ; HARVEY, L.M. ; MCNEIL, B.: Use of At-Line and *In-Situ* Near-Infrared Spectroscopy to Monitor Biomass in an Industrial Fed-Batch *Escherichia coli* Process. In: *Biotechnology and Bioengineering* 80 (2002), S. 405–413
- [Arnold *et al.* 2003b] ARNOLD, S.A. ; HARVEY, L.M. ; MCNEIL, B. ; HALL, J.W.: Employing near-infrared spectroscopic methods of analysis for fermentation monitoring and control: Part 2, implementation strategies. In: *BioPharm International* 16 (2003), Nr. 1, S. 47–49+70
- [Baldwin and Krebs 1981] BALDWIN, J.E. ; KREBS, H.: The Evolution of Metabolic Cycles. In: *Nature* 291 (1981), S. 381–382
- [Bentley *et al.* 1991] BENTLEY, W.E. ; DAVIS, R.H. ; KOMPALA, D.S.: Dynamics of Induced CAT Expression in *E. coli*. In: *Biotechnology and Bioengineering* 38 (1991), S. 749–760
- [Böck and Sawers 1996] BÖCK, A. ; SAWERS, G.: *Escherichia coli and Salmonella*. Washington D.C. : American Society for Microbiology Press, 1996
- [Bonsignore and de Flora 1972] BONSIGNORE, A. ; FLORA, A. de: Regulatory Properties of Glucose-6-phosphat Dehydrogenase. In: *Curr. Top. Cell Reg.* 6 (1972), S. 21
- [Brereton 1992] BRERETON, Richard G.: *Multivariate Pattern Recognition in Chemometrics, Illustrated by Case Studies*. Elsevier Science Ltd, 1992. – ISBN 0-444-89784-4
- [Brereton 2003] BRERETON, Richard G.: *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*. Chichester : Wiley, 2003. – ISBN 0-471-48978-6
- [Brown *et al.* 1977] BROWN, T.D. ; JONES-MORTIMER, M.C. ; KORNBERG, H.I.: The enzymatic interconservation of acetate and acetyl-coenzyme A in *Escherichia coli*. In: *J. Gen. Microbiol.* 102 (1977), S. 327–336
- [Buchholz *et al.* 2002] BUCHHOLZ, A. ; HURLEBAUS, J. ; WANDREY, C. ; TAKORS, R.: Metabolomics: quantification of intracellular metabolite dynamics. In: *Biomolecular Engineering* 19 (2002), S. 5–15
- [Cannizzaro *et al.* 2003] CANNIZZARO, C. ; GÜGERLI, R. ; MARISON, I. ; STOCKER, U. von: On-Line Biomass Monitoring of CHO Perfusion Culture with Scanning Dielectric Spectroscopy. In: *Biotechnology and Bioengineering* 84 (2003), S. 597–610

- [Cannizzarro *et al.* 2003] CANNIZZARRO, C. ; RHIEL, M. ; MARISON, I. ; STOCKAR, U. von: On-Line Monitoring of *Phaffia rhodozyma* Fed-Batch Process With *In Situ* Dispersive Raman Spectroscopy. In: *Biotechnology and Bioengineering* 113 (2003), S. 668–680
- [Cha *et al.* 2000] CHA, H.J. ; WU, C.-F. ; VALDES, J.J. ; RAO, G. ; BENTLEY, W.E.: Observation of green fluorescent protein as a fusion partner in genetically engineered *Escherichia coli*: Monitoring protein expression and solubility. In: *Biotechnology and Bioengineering* 67 (2000), S. 565–574
- [Chalfie 1995] CHALFIE, M.: Green fluorescent protein. In: *Photochem. Photobiol.* 62 (1995), S. 651–656
- [Chalfie *et al.* 1994] CHALFIE, M. ; TU, Y. ; EUSKIRCHEN, G. ; WARD, W. W. ; PRASHER, D. C.: A green fluorescent protein as a marker for gene expression. In: *Science* 263 (1994), S. 802–805
- [Chatfield 2003] CHATFIELD, C.: *The Analysis of Time Series: An Introduction*. CRC Press LLC, 2003. – ISBN 1-584-88317-0
- [Clark 1989] CLARK, D.P.: The fermentation pathways of *Escherichia coli*. In: *FEMS Microbiol. Rev.* 63 (1989), S. 223–234
- [Clementsich *et al.* 2005] CLEMENTSCHITSCH, F. ; KERN, J. ; PÖTSCHACHER, F. ; BAYER, K.: Sensor combination and chemometric modelling for improved process monitoring in recombinant *E. coli* fed-batch cultivations. In: *Journal of Biotechnology* 120 (2005), Nr. 2, S. 183–196
- [Cockshott and Bogle 1999] COCKSHOTT, A.R. ; BOGLE, I.D.L.: Modelling the effects of glucose feeding on a recombinant *E. coli* fermentation. In: *Bioproc. Engin.* 20 (1999), S. 83–90
- [Cohen *et al.* 1973] COHEN, S.N. ; CHANG, A.C.Y. ; BOYER, H.W. ; HELLING, R.B.: Construction of biologically Functional Bacterial Plasmids *In Vitro*. In: *Proceedings of the National Academy of Sciences of the United States of America* 70 (1973), Nr. 11, S. 3240–3244
- [Cooper 1984] COOPER, R.A.: Metabolism of Methylglyoxal in Microorganisms. In: *Ann. Rev. Microbiol.* 38 (1984), S. 49–68
- [Cornelissen 2004] CORNELISSEN, G.: *Integrierte Bioprozessentwicklung zur Herstellung pharmakologisch wirksamer Proteine mit Pichia pastoris*. Hannover, Universität Hannover, Dissertation, 2004
- [Cover and Thomas 1991] COVER, T.M. ; THOMAS, J.A.: *Elements of Information Theory*. Wiley, 1991. – ISBN 3-18-375508-4
- [Crowley *et al.* 2005] CROWLEY, J. ; ARNOLD, S.A. ; WOOD, N. ; HARVEY, L.M. ; MCNEIL, B.: Monitoring a high cell density recombinant *Pichia pastoris* fed-batch bioprocess using transmission and reflectance near infrared spectroscopy. In: *Enzyme and Microbial Technology* 36 (2005), Nr. 5–6, S. 621–628

- [Cubitt *et al.* 1995] CUBITT, A.B. ; HEIM, R. ; ADAMS, S.R. ; BOYD, A.E. ; GROSS, L.A. ; TSIEN, R.Y.: Understanding, improving and using green fluorescent proteins. In: *TIBS* 10 (1995), S. 448–455
- [Cullum and Vicente 1969] CULLUM, J. ; VICENTE, M.: Cell growth and length distribution in *Escherichia coli*. In: *Journal of Bacteriology* 134 (1969), S. 330–337
- [Danzer *et al.* 2001] DANZER, K. ; HOBERT, H. ; FISCHBACHER, C. ; JAGEMANN, K.-U.: *Chemometrik: Grundlagen und Anwendungen*. Berlin, Heidelberg, New York : Springer, 2001. – ISBN 3-540-41291-3
- [Danzer *et al.* 2004] DANZER, Klaus ; OTTO, Matthias ; LLOYD, A. C.: Guidelines for Calibration in Analytical chemistry Part2: Multispecies Calibration. In: *Pure Appl. Chem.* 76 (2004), Nr. 6, S. 1215–1225
- [Davey *et al.* 1993] DAVEY, C.L. ; DAVEY, H.M. ; KELL, D.B. ; TODD, R.W.: Introduction to the dielectric estimation of cellular biomass in real time, with special emphasis on measurements at high volume fractions. In: *Analytica Chimica Acta* 279 (1993), S. 155–161
- [Delfosse and Loubaton 1995] DELFOSSE, N. ; LOUBATON, P.: Adaptive blind source separation of independent sources: a deflation approach. In: *Signal Processing* 45 (1995), S. 59–83
- [DeLisa *et al.* 1999] DELISA, M.P. ; LI, J. ; RAO, G. ; WEIGAND, W.A. ; BENTLEY, W.E.: Monitoring GFPoperon fusion protein expression during high cell density cultivation of *Escherichia coli* using an on-line optical sensor. In: *Biotechnology and Bioengineering* 65 (1999), S. 54–64
- [Demuth *et al.* 2008] DEMUTH, H. ; BEALE, M. ; HAGAN, M. ; THE MATHWORKS (Hrsg.): *Neural Networks Toolbox User's Guide*. 2008. – URL <http://www.mathworks.de/access/helpdesk/help/helpdesk.html>. – Abruf: 12. April 2008
- [DiMasi *et al.* 2003] DIMASI, J.A. ; HANSEN, R.W. ; GRABOWSKI, H.G.: The proce of innovation: new estimates of drug development costs. In: *Journal of Health Economics* 22 (2003), Nr. 2, S. 151–185
- [Dingermann and Zündorf 2007] DINGERMAN, T. ; ZÜNDORF, I. ; JOHANN WOLFGANG GOETHE-UNIVERSITÄT FRANKFURT: INSTITUT FÜR PHARMAZEUTISCHE BIOLOGIE (Hrsg.): *Einführung in Neuronale Netze*. 2007. – URL <http://wwwmath.uni-muenster.de/lehre/material/wwwnscript/probleme1.html>. – Abruf: 25. Februar 2009
- [Doak and Phillips 1999] DOAK, D.L. ; PHILLIPS, J.A.: In Situ Monitoring of an *Escherichia coli* Fermentation Using a Diamond Composition ATR Probe and Mid-Infrared Spectroscopy. In: *Biotechnol. Prog.* 15 (1999), S. 529–539
- [Dünnebieer and Tups 2007] DÜNNEBIER, G. ; TUPS, H.: FDA PAT Initiative - Eine Anwendersicht zu technischen Möglichkeiten und aktueller industrieller Umsetzung. In: *Chemie Ingenieur Technik* 79 (2007), Nr. 12, S. 2019–2028

- [Elmansi and Holms 1989] ELMANSI, E.M.T. ; HOLMS, W.H.: Control of Carbon Flux to Acetate Excretion During Growth of *Escherichia coli* in Batch and Continuous Cultures. In: *Journal of General Microbiology* 135 (1989), S. 2875–2883
- [Ernst & Young 2009] ERNST & YOUNG: Beyond Borders: Ernst & Young Global Biotech Report. (2009)
- [FDA 2004] FDA: *Guidance for Industry: PAT – A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance*. 2004. – URL <http://www.fda.gov/cder/guidance/6419fn1.htm>. – Abruf: 25. Januar 2005
- [Förster 1948] FÖRSTER, T.: Zwischenmolekulare Energiewanderung und Fluoreszenz. In: *Ann. Physik* 6 (1948), Nr. 2, S. 55
- [Fraval and McBrien 1980] FRAVAL, H.N.A. ; MCBRIEN, D.C.H.: The Effect of Methylglyoxal on Cell Division and the Synthesis of Protein and DNA in Synchronous and Asynchronous Cultures of *Escherichia coli* B/r. In: *J. Gen. Microbiol.* 117 (1980), S. 127–134
- [Fredberg *et al.* 1971] FREDBERG, W.B. ; KISTLER, W.S ; LIN, E.C.C.: Lethal Synthesis of Methylglyoxal by *Escherichia coli* During Unregulated Glycerol Metabolism. In: *J. Bacteriol.* 108 (1971), S. 137–144
- [Friehs 1999] FRIEHS, K.H.: *Maßnahmen zur Verbesserung der Produktion von rekombinanten Proteinen und Plasmid-DNS*. Bielefeld, Universität Bielefeld, Habilitationsschrift, 1999
- [Fritsch and Carlson 1980] FRITSCH, F.N. ; CARLSON, R.E.: Monotone Piecewise Cubic Interpolation. In: *SIAM J. Numerical Analysis* 17 (1980), S. 238–246
- [Fuhrer *et al.* 2005] FUHRER, T. ; FISCHER, E. ; SAUER, U.: Experimental identification and quantification of glucose metabolism in seven bacterial species. In: *Journal of Bacteriology* 187 (2005), S. 1581–1590
- [Gennis and Reeves 1976] GENNIS, R.B. ; REEVES, H.C.: Pyruvate oxidase. In: MARTONOSI, A.N. (Hrsg.): *The Enzymes of Biological Membranes* Bd. 2. N.Y., USA : Plenum, New York, 1976, S. 493–504
- [Golub and Loan 1983] GOLUB, Gene. H. ; LOAN, Charles F. V.: *Matrix computations*. North Oxford Academic, 1983. – ISBN 0-946-53605-8
- [Grubbs 1969] GRUBBS, Frank: Procedures for Detecting Outlying Observations in Samples. In: *Technometrics* 11 (1969), Nr. 1, S. 1–21
- [Hansen 2004] HANSEN, E.H.: The impact of flow injection on modern chemical analysis: has it fulfill our expectations? And where are we going? In: *Talanta* 64 (2004), S. 1076–1083
- [Hartung 1989] HARTUNG, Joachim: *Statistik: Lehr- und Handbuch der angewandten Statistik*. München, Wien : Oldenbourg, 1989. – ISBN 3-486-57890-1
- [Hauttmann and Müller 2001] HAUTTMANN, S. ; MÜLLER, J.: *In situ* biomass characterization by impedance spectroscopy using a full bridge circuit. In: *Bioprocess Biosys. Eng.* 24 (2001), S. 137–141

- [Hebb 1949] HEBB, Donald. O.: *The organization of behavior; a neuropsychological theory*. New York : Wiley-Interscience, 1949
- [Heine 2000] HEINE, T.: *Modellgestützte Überwachung und Führung von Fed-Batch-Prozessen zur Antibiotikaproduktion*. Berlin, Technische Universität Berlin, Dissertation, 2000
- [Hiller *et al.* 2007] HILLER, J. ; FRANCO-LARA, E. ; PAPAIOANNOU, V. ; WEUSTER-BOTZ, D.: Fast sampling and quenching procedures for microbial metabolic profiling. In: *Biotechnol. Lett.* 29 (2007), S. 1161–1167
- [Hilmer 1996] HILMER, J.-M.: *In-situ-Bioprozesskontrolle am Beispiel der In-vivo-Biolumineszenz und der 2D-Fluoreszenzspektroskopie*. Hannover, Universität Hannover, Dissertation, 1996
- [Himberg and Hyvärinen 2003] HIMBERG, J. ; HYVÄRINEN, A.: ICASSO: Software for investigating the reliability of ICA estimates by clustering and visualization. In: *Processings of the IEEE 13th Workshop on Neural Networks for Signal Processing* (2003), S. 259–268
- [Himberg *et al.* 2004] HIMBERG, J. ; HYVÄRINEN, A. ; ESPOSITO, F.: Validating the independent components of neuroimaging time-series via clustering and visualisation. In: *NeuroImage* 22 (2004), Nr. 3, S. 1214–1222
- [Holms 1996] HOLMS, W.H.: Flux analysis and control of the central metabolic pathways in *Escherichia coli*. In: *FEMS Microbiol. Rev.* 19 (1996), S. 85–116
- [Hopper and Cooper 1971] HOPPER, D.J. ; COOPER, R.A.: The Regulation of *Escherichia coli* Methylglyoxal Synthase; A New Control Site in Glycolysis? In: *FEBS Letters* 13 (1971), S. 213–216
- [Horstkotte *et al.* 2006] HORSTKOTTE, B. ; WERNER, E. ; WIEDEMEIER, S. ; ELSHOLZ, O. ; CERDA, V. ; LUTTMANN, R.: At-line determination of formaldehyde in bioprocesses by sequential injection analysis. In: *Analytica Chimica Acta* 559 (2006), S. 248–256
- [Huber 1985] HUBER, P.: Projection pursuit. In: *The Annals of Statistics* 13 (1985), Nr. 2, S. 435–475
- [Hyman 1983] HYMAN, J.M.: Accurate Monotonicity Preserving Cubic Interpolation. In: *J. Sci. Stat. comput* 4 (1983), Nr. 4, S. 645–654
- [Hyvärinen 1998] HYVÄRINEN, A.: New approximations of differential entropy for independent component analysis and projection pursuit. In: *Advances in Neural Information Processing Systems* 10 (1998), S. 273–279. – MIT Press
- [Hyvärinen and Oja 2000] HYVÄRINEN, Aapo ; OJA, Erkki: Independent Component Analysis: Algorithms and Applications. In: *Neural Networks* 13 (2000), Nr. 4-5, S. 411–430
- [Inoue and Kimura 1995] INOUE, Y. ; KIMURA, A.: Methylglyoxal and Regulation of its Metabolism in Microorganisms. In: *Adv. Microbiol. Physiol.* 37 (1995), S. 177–227
- [Invitrogen 2006a] INVITROGEN: *pRSET A, B, and C: For high-level expression of recombinant proteins in E. coli*, 2006. – Catalog no. V351-20

- [Invitrogen 2006b] INVITROGEN: *pTrcHis A, B, and C: Vectors for Expression of Recombinant Proteins Containing N-Terminal 6xHis Tags in E. coli*, 2006. – Catalog no. V360-20
- [Jackson *et al.* 1972] JACKSON, D.A. ; SYMONS, R.H. ; BERG, P.: Biochemical Method for Inserting New Genetic Information into DNA of Simian Virus: Circular SV40 DNA Molecules Containing Lambda Phage Genes and the Galactose Operon of *Escherichia coli*. In: *Proceedings of the National Academy of Sciences of the United States of America* 69 (1972), Nr. 10, S. 2904–2909
- [Jeremy *et al.* 2007] JEREMY, J.M. ; TYMOCZKO, J.L. ; STRYER, L.: *Styer Biochemie*. Heidelberg : Spektrum Akademischer Verlag, 2007. – ISBN 978-3-8274-1800-5
- [Joeris *et al.* 2002] JOERIS, K. ; FRERICHS, J.G. ; KONSTANTINOV, K. ; SCHEPER, T.: *In-Situ* microscopy: Online process monitoring of mammalian cell cultures. In: *Cytotechnology* 38 (2002), S. 129–134
- [Johnston *et al.* 2000] JOHNSTON, K. ; CLEMENTS, A. ; VENKATARAMANI, R.N. ; TRIEVEL, R.C. ; MARMORSTEIN, R.: Coexpression of Proteins in Bacteria Using T7-Based Expression Plasmids: Expression of Heteromeric Cell-Cycle and Transcriptional Regulatory Complexes. In: *Protein Expression and Purification* 20 (2000), S. 435–443
- [Känsäkoski *et al.* 2006] KÄNSÄKOSKI, M. ; KURKINEN, M. ; WEYMARN, N. von ; NIEMELÄ, P. ; NEUBAUER, P. ; JUUSO, E. ; EERIKÄINEN, T. ; TURUNEN, S. ; SUHONEN, P. ; VTT BUISNESS FROM TECHNOLOGY (Hrsg.): *Process analytical technology (PAT) needs and applications in the bioprocess industry*. 2006. – URL <http://www.vtt.fi/inf/pdf/workingpapers/2006/W60.pdf>. – Abruf: 08. August 2008
- [Kayser 1999] KAYSER, A.N.: *Stoffflüsse in Escherichia coli TGI unter aeroben, glucoseli-mitierten Bedingungen*. Berlin, Technische Universität Braunschweig, Dissertation, 1999
- [Kleman and Strohl 1994] KLEMAN, G.L. ; STROHL, W.R.: Acetate Metabolism by *Escherichia coli* in High Cell Density Fermentation. In: *Appl. Environ. Microbiol.* 57 (1994), S. 3952–3958
- [Klockewitz 2000] KLOCKEWITZ, K.: *Entwicklung und Einsatz eines FIA-Immunanalytensystems zur schnellen On-line-Bestimmung von pharmazeutisch relevanten humanen Proteinen in Bioprozessen*. Hannover, Universität Hannover, Dissertation, 2000
- [Kohonen 1997] KOHONEN, T.: *Self-Organizing Maps*. Berlin : Springer-Verlag, 1997
- [Kollecker 2006] KOLLECKER, M.: *Analyse von 2D-Fluoreszenzspektren durch multivariate Auswerteverfahren zur Regelung von Bioprozessen*. Hannover, Universität Hannover, Dissertation, 2006
- [Kondepatti and Heise 2008] KONDEPATI, V.R. ; HEISE, H.M.: The Potential of Mid and Near-infrared Spectroscopy for Reliable Monitoring of Bioprocesses. In: *Current Trends in Biotechnology and Pharmacy* 2 (2008), Nr. 1, S. 117–132
- [Kornmann *et al.* 2003] KORNMANN, H. ; RHIEL, M. ; CANNIZZARRO, C. ; MARISON, I. ; STOCKAR, U. von: Methodology for Real-Time, Multianalyte Monitoring of Fermentations Using an *In-Situ* Mid-Infrared sensor. In: *Biotechnology and Bioengineering* 82 (2003), S. 702–709

- [Kornmann *et al.* 2004] KORNMANN, H. ; VALENTINOTTI, S. ; DUBOC, P. ; MARISON, I. ; STOCKAR, U. von: Methodology for Real-Time, Multianalyte Monitoring of Fermentations Using an *In-Situ* Mid-Infrared sensor. In: *Journal of Biotechnology* 113 (2004), S. 231–245
- [Kowalski 1978] KOWALSKI, B.R.: In: *Chem. Ind.* 22 (1978), S. 882. – London
- [Kubitschek 1969] KUBITSCHKEK, H.E.: Growth during the bacterial cell cycle. Analysis of cell size distributions. In: *Biophysics* 9 (1969), S. 792–809
- [Laboratory of Computer and Information Science: Adaptive Informatics Research Center 2005] LABORATORY OF COMPUTER AND INFORMATION SCIENCE: ADAPTIVE INFORMATICS RESEARCH CENTER: *Icasso: software for investigating reliability of ICA estimates by clustering and visualization*. 2005. – URL <http://www.cis.hut.fi/projects/ica/icasso/>. – Abruf: 25. Juni 2007
- [Lantz *et al.* 2006] LANTZ, A.E. ; ; JORGENSEN, P. ; POULSEN, E. ; LINDEMANN, C. ; OLSSON, L.: On-line cell mass monitoring of *Saccharomyces cerevisiae* cultivations by multi-wavelength fluorescence. In: *Journal of Biotechnology* 114 (2006), S. 199–208
- [Leveau and Lindow 2001] LEVEAU, J.H.J. ; LINDOW, S.E.: Predictive and Interpretive Simulation of Green Fluorescent Protein Expression in Reporter Bacteria. In: *Journal of Biotechnology* 183 (2001), S. 6752–6762
- [Levin 2005] LEVIN, M.: *Pharmaceutical Process Scale-Up*. East Hanover, New Jersey : CRC Press, 2005. – ISBN 1-574-44876-4
- [Li *et al.* 1991] LI, J.-K. ; ASALI, E.C. ; HUMPHREY, A.E.: Monitoring cell concentration and activity by multiple excitation fluorometry. In: *Biotechnol. Prog.* 7 (1991), S. 21–27
- [Li and Humphrey 1992] LI, J.-K. ; HUMPHREY, A.E.: Factors affecting culture fluorescence when monitoring bioreactors. In: *J. Ferm. Bioeng.* 74 (1992), Nr. 2, S. 101–111
- [Liao and Wen 2007] LIAO, S.-H. ; WEN, C.H.: Artificial neural networks classification and clustering methodologies and applications – literature analysis from 1995 to 2005. In: *Expert Systems with Applications* 32 (2007), S. 1–11
- [Lindemann 1998] LINDEMANN, C.: *Anwendungen der 2D-Fluoreszenzspektroskopie zur Bioprozessbeobachtung*. Hannover, Universität Hannover, Dissertation, 1998
- [Lindner *et al.* 2007] LINDNER, P. ; KRABICHLER, C. ; RUDOLPH, G. ; SCHEPER, T. ; HITZMANN, B.: Application of *In-Situ*-Microscopy and Digital Image Processing in Yeast Cultivations. In: *Proceedings of the 10th International IFAC Symposium on Computer Applications in Biotechnology*, IEEE Press, 2007, S. 245–249
- [Luli and Strohl 1990] LULI, G.W. ; STROHL, W.R.: Comparison of Growth, Acetate Production, and Acetate Inhibition of *Escherichia coli* Strains in Batch and Fed-Batch Fermentations. In: *Applied and Environmental Microbiology* 56 (1990), Nr. 4, S. 1004–1011
- [Macaloney *et al.* 1997] MACALONEY, G. ; HALL, J.W. ; NOCH fehlt: The utility and performance of near-infra red spectroscopy in simultaneous monitoring of multiple components in a high cell density recombinant *Escherichia coli* production process. In: *Bioprocess Engineering* 17 (1997), Nr. 3, S. 157

- [Madigen *et al.* 2000] MADIGEN, M. ; MARTINKO, J. ; PARKER, J.: *Brock Mikrobiologie*. Heidelberg : Spektrum Akademischer Verlag, 2000. – ISBN 3-8274-0566-1
- [Mardia *et al.* 1979] MARDIA, K.V. ; KENT, J.T. ; BIBBY, J.: *Multivariate Analysis*. London : Academic Press, 1979
- [Marose 1999] MAROSE, S.: *Bioprozessanalytik mit zweidimensionaler Fluoreszenzspektroskopie*. Hannover, Universität Hannover, Dissertation, 1999
- [Marose *et al.* 1998] MAROSE, S. ; LINDEMANN, C. ; SCHEPER, T.: Two-dimensional fluorescence spectroscopy: A new tool for on-line bioprocess monitoring. In: *Biotechnol. Prog.* 14 (1998), Nr. 1, S. 63–74
- [Marose *et al.* 1999] MAROSE, S. ; LINDEMANN, C. ; ULBER, R. ; SCHEPER, T.: Optical sensor systems for bioprocess monitoring. In: *Trends in Biotechnology* 17 (1999), S. 30–34
- [van de Merbel *et al.* 1992] MERBEL, N.C. van de ; KOOL, I.M. ; LINGEMAN, H. ; BRINKMAN, U.A.T. ; KOLHORN, A. ; DERIJKE, L.C.: On-line monitoring of fermentation processes by ultrafiltration and column liquid chromatography. In: *Chromatographia* 33 (1992), S. 525–532
- [van de Merbel *et al.* 1996] MERBEL, N.C. van de ; LINGEMAN, H. ; BRINKMAN, U.A.T.: Sampling and analytical strategies in on-line bioprocess monitoring and control. In: *Journal of Chromatography* 725 (1996), Nr. 1, S. 13–27
- [van de Merbel *et al.* 1995] MERBEL, N.C. van de ; ZUUR, P. ; FRIJLINK, M. ; HOLTHUIS, J.J.M. ; LINGEMAN, H. ; BRINKMAN, U.A.T.: Automated monitoring of amino acids during fermentation processes using on-line ultrafiltration and column liquid chromatography: application to fermentation medium improvement. In: *Analytica Chimica Acta* 303 (1995), Nr. 2-3, S. 175–185
- [Morise *et al.* 1974] MORISE, H. ; SHIMOMURA, O. ; JOHNSON, F.H. ; WINANT, J.: Intermolecular energy transfer in the bioluminescent system of *Aequorea*. In: *Biochemistry* 13 (1974), S. 2656–2662
- [Navratil *et al.* 2005] NAVRATIL, M. ; NORBERG, A. ; LEMBREN, L. ; MANDENIUS, C.-F.: On-line multi-analyzer monitoring of biomass, glucose and acetate for growth rate control of a *Vibrio cholerae* fed-batch cultivation. In: *Journal of Biotechnology* 115 (2005), S. 67–79
- [optek-Danulat 2005] OPTEK-DANULAT: *Firmenhomepage*. 2005. – URL <http://www.optek.com>. – Abruf: 13. Februar 2008
- [Ormö *et al.* 1996] ORMÖ, M. ; CUBITT, A.B. ; KALLIO, K. ; GROSS, L.A. ; TSIEN, R.Y. ; REMINGTON, S.J.: Crystal structure of the *Aequorea victoria* green fluorescent protein. In: *Science* 273 (1996), S. 1392–1395
- [Otto 1997] OTTO, Matthias: *Chemometrie*. Weinheim : VCH Verlagsgesellschaft mbH, 1997. – ISBN 3-527-28849-X
- [Pao 1989] PAO, Yoh-Han: *Adaptive pattern recognition and neural networks*. Boston, MA, USA : Addison-Wesley Longman Publishing Co., Inc., 1989. – ISBN 0-201-12584-6

- [Papoulis 1991] PAPOULIS, A.: *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, 1991. – ISBN 0-073-66011-6
- [Peuker 2006] PEUKER, T.: *Modellgestützte Verfahren zur Beobachtung und Steuerung bioverfahrenstechnischer Prozesse*. Hannover, Universität Hannover, Dissertation, 2006
- [Pitts and McCulloch 1947] PITTS, W. ; MCCULLOCH, S.: How we know universal perception of auditory and visual forms. In: *Bulletin of Mathematical Biology* 9 (1947), Nr. 3, S. 127–147
- [Pollard *et al.* 2002] POLLARD, D.J. ; BUCCINO, R. ; CONNORS, N.C. ; KIRSCHNER, T.F. ; OLEWINSKI, R.C. ; SAINI, K. ; SALMON, P.M.: Real-time analyte monitoring of a fungal fermentation, at pilot scale, using *in situ* mid-infrared spectroscopy. In: *Bioproc. Biosyst. Eng.* 24 (2002), S. 13–24
- [Prasher *et al.* 1962] PRASHER, D.C. ; ECKENRODE, V.K. ; WARD, W.W. ; PRENDERGAST, F.G. ; CORMIER, M.J.: Primary structure of the *Aequorea victoria* green fluorescent protein. In: *Gene* 59 (1962), S. 229–233
- [Press *et al.* 1986] PRESS, W. ; BRIAN, P. ; SAUL, S.A. ; TEUKOLSKY, A. ; VETTERLING, W.T.: *Numerical Recipes: The Art of Scientific Computing*. Cambridge : Cambridge University Press, 1986
- [R. and Sibson 1987] R., M. J. ; SIBSON: What is projection pursuit? In: *Journal of the Royal Statistical Society, Ser. A* 150 (1987), S. 1–36
- [Reinecke and Scheper 1997] REINECKE, M. ; SCHEPER, T.: Fast on-line flow injection analysis system for IgG monitoring in bioprocesses. In: *Journal of Biotechnology* 59 (1997), Nr. 1–2, S. 145–153
- [Reischer *et al.* 2004] REISCHER, H. ; STRIEDNER, I. Schotola G. ; PÖTSCHACHER, F. ; BAYER, K.: Sensor combination and chemometric modelling for improved process monitoring in recombinant *E. coli* fed-batch cultivations. In: *Journal of Biotechnology* 108 (2004), S. 115–125
- [Rhiel *et al.* 2002] RHIEL, M. ; DUCOMMUN, P. ; BOLZONELLA, I. ; STOCKAR, I. Marisonand U. von: *In Situ* Monitoring of Freely Suspended and Immobilized Cell Cultures Based on Mid- Infrared Spectroscopic Measurements. In: *Biotechnology and Bioengineering* 77 (2002), S. 174–185
- [Riesenberger *et al.* 1991] RIESENBERGER, D. ; SCHULZ, V. ; KNORRE, W.A. ; POHL, H.-D. ; KORZ, D. ; SANDERS, E.A. ; ROSS, A. ; DECKWER, W.-D.: High cell density cultivation of *Escherichia coli* at controlled specific growth rate. In: *Journal of Biotechnology* 20 (1991), S. 17–28
- [Rojas 1996] ROJAS, R.: *Neural Networks – A Systematic Introduction*. Springer, 1996. – ISBN 3-540-60505-3
- [Scheper *et al.* 1993] SCHEPER, T. ; BRANDES, W. ; MASCHKE, H. ; PLÖTZ, F. ; MÜLLER, C.: Two FIA-based biosensor systems studied for bioprocess monitoring. In: *Journal of Biotechnology* 31 (1993), S. 345–356

- [Scheper and Schügerl 1986] SCHEPER, T. ; SCHÜGERL, K.: Characterization of bioreactors by *in-situ*-fluorometry. In: *Journal of Biotechnology* 3 (1986), S. 221–229
- [Schimizu *et al.* 1988] SCHIMIZU, N. ; FUKUZONO, S. ; FUJIMORI, K. ; NISHIMURA, N. ; ODAWARA, Y.: Fed-batch cultures of recombinant *Escherichia coli* with inhibitory substance monitoring. In: *Journal of Fermentation Technology* 66 (1988), S. 187–191
- [Schlegel 1992] SCHLEGEL, H.G.: *Allgemeine Mikrobiologie*. Stuttgart, New York : Thieme Verlag, 1992. – ISBN 3-13-444607-3
- [Schneegaß 2004] SCHNEEGASS, D. ; UNIVERSITÄT LÜBECK (Hrsg.): *Lektion zur Principal Component Analysis und zur Independent Component Analysis: DaMiT-Projekt*. 2004. – URL <http://www.inb.uni-luebeck.de/~schneega/lektion.pdf>. – Abruf: 10. Juli 2008
- [Schneider 1999] SCHNEIDER, R.: *Untersuchung eines adaptiven prädiktiven Regelungsverfahrens zur Optimierung von bioverfahrenstechnischen Prozessen*. VDI-Verlag, 1999. – ISBN 3-18-375508-4
- [Schoepfer 1993] SCHOEPFER, Ralf: The pRSET family of T7 promoter expression vectors for *Escherichia coli*. In: *Gene* 124 (1993), S. 83–85
- [Schügerl 2001] SCHÜGERL, K.: Progress in monitoring, modeling and control of bioprocesses during the last 20 years. In: *Journal of Biotechnology* 85 (2001), S. 149–173
- [Schwedt 1989] SCHWEDT, G.: *Fluorimetrische Analyse. Methode und Anwendung*. Weinheim, Deerfield Beach, Basel : Wiley-VCH, 1989. – ISBN 3-527-25870-1
- [Shimomura *et al.* 1962] SHIMOMURA, O. ; JOHNSON, F.H. ; SAIGA, Y.: Extraction, purification, and properties of aequorin, a bioluminescent protein from the luminous hydromedusa, *Aequorea*. In: *J. Cell. Comp. Physiol.* 59 (1962), S. 223–239
- [Shumway and Stoffer 2006] SHUMWAY, R.H. ; STOFFER, D.S.: *Time Series Analysis and Its Applications*. Springer, 2006. – ISBN 9780387293172
- [Sitton and Scrienc 2008] SITTON, G. ; SRIENC, F.: Mammalian cell culture scale-up and fed-batch control using automated flow cytometry. In: *Journal of Biotechnology* 2 (2008), S. 174–180
- [Solle 2003] SOLLE, D.: *Analyse und Optimierung eines industriellen Biotransformationsprozesses zur Herstellung von Tryptophan*. Hannover, Universität Hannover, Dissertation, 2003
- [Stärk 2002] STÄRK, Egbert: *Bioprozessanalytik durch Online-Vorhersage von Bioprozessgrößen mittels 2D-Fluoreszenzspektroskopie und multivariater Auswerteverfahren*. Hannover, Universität Hannover, Dissertation, 2002
- [Studier *et al.* 1990] STUDIER, F.W. ; ROSENBERG, A.H. ; DUNN, J.J. ; DUBENDORF, J.W.: Use of T7 polymerase to direct expression of cloned genes. In: *Methods Enzymol.* 185 (1990), S. 60–89

- [Surribas *et al.* 2006] SURRIBAS, A. ; GEISSLER, D. ; GIERSER, A. ; SCHEPER, T. ; HITZMANN, B. ; MONTESINOS, J.L. ; VALERO, F.: State variables monitoring by *in situ* multi-wavelength fluorescence spectroscopy in heterologous protein production in *Pichia pastoris*. In: *Journal of Biotechnology* 124 (2006), S. 412–419
- [Suzuki 1969] SUZUKI, T.: Phosphotransacetylase of *Escherichia coli* B, activation by pyruvate and inhibition by NADH and certain nucleotides. In: *Biochim. Biophys. Acta* 191 (1969), S. 559–569
- [Swartz 2001] SWARTZ, J.R.: Advances in *Escherichia coli* production of therapeutic proteins. In: *Curr. Opin. Biotechnol.* 12 (2001), S. 195–201
- [Tartakovsky *et al.* 1996] TARTAKOVSKY, B. ; SHEINTUCH, M. ; HILMER, J.-M. ; SCHEPER, T.: Application of Scanning Fluorometry for Observation and Modelling of a Fermentation Process. In: *Biotechnol. Prog.* 12 (1996), S. 126–131
- [Tsien 1998] TSIEN, R.Y.: The green fluorescent protein. In: *Annu. Rev. Biochem* 67 (1998), S. 509–544
- [Universität Münster 2007] UNIVERSITÄT MÜNSTER ; NEURO FUZZY AG (Hrsg.): *Einführung in Neuronale Netze*. 2007. – URL <http://wwwmath.uni-muenster.de/lehre/material/wwwnscript/probleme1.html>. – Abruf: 25. Juli 2008
- [Varma and Palsson 1994] VARMA, A. ; PALSSON, B.O.: Stoichiometric Flux Balance Models Quantitatively Predict Growth and Metabolic By-Product Secretion in Wild-Type *Escherichia coli* W3110. In: *Appl. Environ. Microbiol.* 60 (1994), S. 3724–3731
- [Wachter 2008] WACHTER, R.M. ; ARIZONA STATE UNIVERSITY (Hrsg.): *Protein Maturation*. 2008. – URL http://rebekka-wachter.com/Protein_Maturation.html. – Abruf: 21. Mai 2009
- [Wang and Hansen 2003] WANG, J. ; HANSEN, E.H.: Sequential injection lab-on-valve: the third generation of flow injection analysis. In: *Trends in Analytical Chemistry* 22 (2003), Nr. 4, S. 225–231
- [Ward *et al.* 1982] WARD, W.W. ; PRENTICE, H.J. ; ROTH, A.F. ; CODY, C.W. ; REEVES, S.C.: Spectral perturbations of the *Aequorea* green-fluorescent protein. In: *Photochem. Photobiol.* 35 (1982), S. 803–808
- [Wedler 2004] WEDLER, G.: *Lehrbuch der physikalischen Chemie*. Wiley-VCH, 2004. – ISBN 3-527-31066-5
- [Widrow *et al.* 1988] WIDROW, B. (Hrsg.) ; MORROW (Hrsg.) ; GSCHWENDTNER (Hrsg.): *DARPA Neural Network Study*. Lexington, MA : Afcea International Press, 1988. – ISBN 0-916159-17-5
- [Wittmann 2007] WITTMANN, C.: Review: Fluxome analysis using GC-MS. In: *Microbial Cell Factories* 6 (2007)
- [Wold *et al.* 1987] WOLD, S. ; ESBENSEN, K. ; GELADI, P.: Principal component analysis. In: *Chemometrics and Intelligent Laboratory Systems* 2 (1987), S. 37–52

- [Wolfbeis *et al.* 1985] WOLFBEIS, O.S. ; POSCH, H.E. ; KRONEIS, H.W.: Fiber optical fluorosensor for determination of halothane and or oxygen. In: *Anal. Chem.* 57 (1985), S. 2556–2561
- [Wolfe 2005] WOLFE, Alan J.: The Acetate Switch. In: *Microbiology and Molecular Biology Reviews* 69 (2005), Nr. 1, S. 12–50
- [Xu *et al.* 1999] XU, B. ; JAHIC, M. ; ENFORS, S.-O.: Modeling of Overflow Metabolism in Batch and Fed-Batch Cultures of *Escherichia coli*. In: *Biotechnol. Prog.* 15 (1999), S. 81–90
- [Yang *et al.* 1996] YANG, F. ; MOSS, L.G. ; JR., G.N. P.: The molecular structure of green fluorescent protein. In: *Nature Biotechnol.* 14 (1996), S. 1246–1251
- [Zabriskie and Humphrey 1978] ZABRISKIE, D.W. ; HUMPHREY, A.E.: Estimation of Fermentation Biomass Concentration by Measuring Culture Fluorescence. In: *Appl. Environ. Microbiol.* 35 (1978), Nr. 2, S. 337–343
- [Zander 1981] ZANDER, M.: *Fluorimetrie*. Berlin, Heidelberg, New York : Springer-Verlag, 1981. – ISBN 3-540-10512-3
- [Zapata-Hommer and Griesbeck 2003] ZAPATA-HOMMER, O. ; GRIESBECK, O.: Efficiently folding and circularly permuted variants of the Sapphire mutant of GFP. In: *BMC Biotechnol.* 3 (2003), S. 5
- [van der Zee *et al.* 1996] ZEE, J.R. van der ; POSTMA, P.W. ; HELLINGWERF, K.J.: Quantitative Conversion of Glucose Into Glucose-6-phosphate by Intact *Escherichia coli* cells. In: *Biotechnol. Appl. Biochem.* 24 (1996), S. 225–230
- [Zell 1997] ZELL, A.: *Simulation neuronaler Netze*. München : R. Oldenbourg Verlag, 1997. – ISBN 3-486-24350-0
- [Zheng *et al.* 2005] ZHENG, Z.-Y. ; YAO, S.-J. ; LIN, D.-Q.: Using a kinetic model that considers cell segregation to optimize hEGF expression in fed-batch cultures of recombinant *E. coli*. In: *Bioprocess Biosyst. Eng.* 27 (2005), S. 143–152
- [Zoon 2008] ZOON, P.D. ; UNIVERSITEIT VAN AMSTERDAM (Hrsg.): *Single Molecule Spectroscopy by Confocal Microscopy*. 2008. – URL <http://staff.science.uva.nl/~200n/sms/SMS.html>. – Abruf: 13. April 2008