

Fortschritt-Berichte VDI

VDI

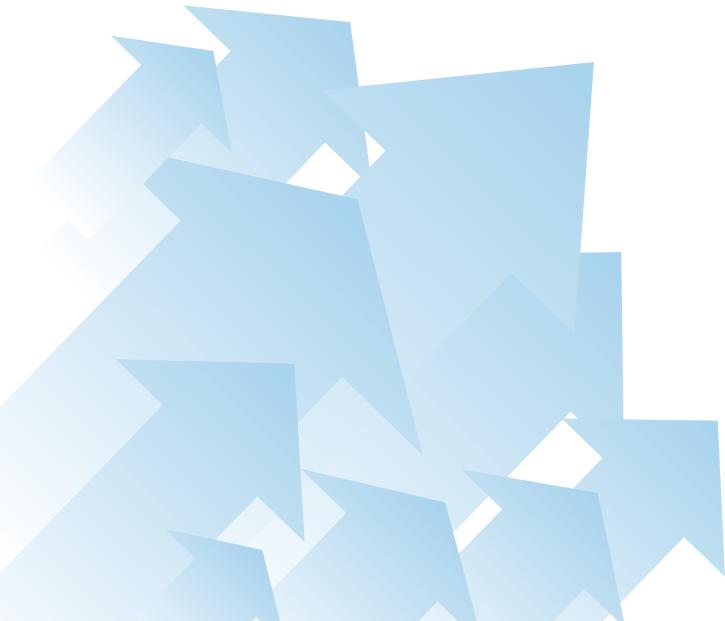
Reihe 17

Biotechnik/
Medizintechnik

M.Sc. Dipl.-Ing. (FH) Jan-Patrick Voß,
Hamburg

Nr. 293

Anwendung spektroskopischer Messverfahren und Multivariater Datenanalyse zur Bewertung und Beobachtung von Bioprozessen



Anwendung spektroskopischer Messverfahren und Multivariater Datenanalyse zur Bewertung und Beobachtung von Bioprozessen

Von der Naturwissenschaftlichen Fakultät der
Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des Grades
Doktor der Naturwissenschaften (Dr. rer. nat.)

genehmigte Dissertation
von
Jan-Patrick Voß, M.Sc. Dipl.-Ing. (FH)
geboren am 26.04.1984 in Hamburg

2017

Referent: Prof. Dr. rer. nat. Thomas Scheper
Koreferent: Prof. Dr. rer. nat. Bernd Hitzmann
Tag der Promotion: 31. Juli 2017

Fortschritt-Berichte VDI

Reihe 17

Biotechnik/
Medizintechnik

M.Sc. Dipl.-Ing. (FH) Jan-Patrick Voß,
Hamburg

Nr. 293

Anwendung spektroskopischer
Messverfahren und
Multivariater Datenanalyse
zur Bewertung und
Beobachtung von Bioprozessen

VDI verlag

Voß, Jan-Patrick

Anwendung spektroskopischer Messverfahren und Multivariater Datenanalyse zur Bewertung und Beobachtung von Bioprozessen

Fortschr.-Ber. VDI Reihe 17 Nr. 293. Düsseldorf: VDI Verlag 2017.

178 Seiten, 111 Bilder, 36 Tabellen.

ISBN 978-3-18-329317-9, ISSN 0178-9600,

€ 67,00/VDI-Mitgliederpreis € 60,30.

Für die Dokumentation: Bioprozessmonitoring – Spektroskopie – Multivariate Datenanalyse – Chemometrie – Process Analytical Technology – *Pichia pastoris* – Malariaimpfstoff

Die vorliegende Arbeit befasst sich mit dem Bioprozessmonitoring unter Verwendung spektroskopischer Messverfahren und multivariater Datenanalyse nach den Grundsätzen von PAT – *Process Analytical Technology*. Mit NIR-Spektroskopie und dem Verfahren *Soft Independent Modelling of Class Analogy* (SIMCA) wurde eine Qualitätsbewertung von Hefeextrakten realisiert. Im Vordergrund stand jedoch die Quantifizierung nicht direkt messbarer Größen aus NIR-, Raman- und 2D-Fluoreszenzspektren in pharmazeutischen Produktionsprozessen mit *Pichia pastoris*. Eine entsprechende Online-Bestimmung mit der Methode *Partial Least Squares Regression* (PLSR) kam weiterführend zur Regelung der Glycerolkonzentration zum Einsatz. Darüber hinaus wurde die Verwendung nichtspektraler Online-Daten zur Prozessbeobachtung erprobt. Dabei gelang mit Hilfe des nichtlinearen Verfahrens *Support Vector Regression* (SVR) unter anderem die Bestimmung zellspezifischer Reaktionsraten.

Bibliographische Information der Deutschen Bibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliographische Daten sind im Internet unter <http://dnb.ddb.de> abrufbar.

Bibliographic information published by the Deutsche Bibliothek

(German National Library)

The Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliographie (German National Bibliography); detailed bibliographic data is available via Internet at <http://dnb.ddb.de>.

© VDI Verlag GmbH · Düsseldorf 2017

Alle Rechte, auch das des auszugsweisen Nachdruckes, der auszugsweisen oder vollständigen Wiedergabe (Fotokopie, Mikrokopie), der Speicherung in Datenverarbeitungsanlagen, im Internet und das der Übersetzung, vorbehalten.

Als Manuskript gedruckt. Printed in Germany.

ISSN 0178-9600

ISBN 978-3-18-329317-9

Referent: Prof. Dr. rer. nat. Thomas Scheper
Koreferent: Prof. Dr. rer. nat. Bernd Hitzmann
Tag der Promotion: 31. Juli 2017

Vorwort

Die vorliegende Arbeit entstand in der Zeit von Januar 2013 bis März 2017 an der Hochschule für Angewandte Wissenschaften (HAW) Hamburg im Forschungs- und Transferzentrum für Bioprozess- und Analysentechnik unter der Leitung von Prof. Dr.-Ing. Reiner Luttmann.

Ihm gilt mein herzlichster Dank für das Vertrauen, dass er in mich gesetzt hat. Für seinen ansteckenden Enthusiasmus und die wertvollen Hilfestellungen bei fachlichen Fragen sowie dafür so viel von ihm und durch die Arbeit in seinem Labor gelernt zu haben.

Herrn Prof. Dr. Thomas Scheper von der Leibniz Universität Hannover sowie Herrn Prof. Dr. Bernd Hitzmann von der Universität Hohenheim danke ich für die Übernahme der Referate.

Für finanzielle Unterstützung möchte ich mich bei der Leitung der Fakultät Life Sciences der HAW Hamburg bedanken. Außerdem wurde diese Arbeit durch das Bundesministerium für Bildung und Forschung gefördert (Förderkennzeichen 1756X09 und 031567).

Den Firmen Polytec und Kaiser Optical Systems danke ich für die Leihgabe von Spektrometer-systemen sowie für technische Unterstützung. MKS data analytics solutions (ehemals Umetrics) gilt mein Dank für die Bereitstellung benötigter Softwarelizenzen.

Den Mitarbeitern des Labors BPA danke ich für die schöne Zeit und das freundschaftliche Arbeitsklima. Frau Prof. Dr. Gesine Cornelissen stand mir jeder Zeit mit konstruktivem Rat zur Seite. Die Laboringenieure Hans-Peter Bertelsen und Ulrich Scheffler haben durch ihre kollegiale und einsatzfreudige Unterstützung maßgeblich zum Erfolg der Arbeit beigetragen.

Besonderer Dank gilt Nina Mittelheuser, Vignesh Rajamanickam und Roman Lemke für den Einsatz im Rahmen ihrer Masterarbeiten und für die dabei erzielten Ergebnisse. Ohne Euch wäre diese Arbeit nicht möglich gewesen.

Allen weiteren lieben Kollegen, vor allem Roman, Sanja, Christian, Jessica, David, Sven, Jens, Kristof und Sarah danke ich auch dafür, dass der Spaß an der Arbeit nie verloren ging.

Zum Abschluss möchte ich meiner Familie und vor allem meiner Freundin Janet für das Verständnis, die mentale Unterstützung sowie jederzeit offene Ohren danken.

*I've done the math enough
to know the dangers of our second-guessing*

Tool

Inhaltsverzeichnis

1. Einleitung und Zielsetzung	1
1.1 PAT – Ein Werkzeug moderner pharmazeutischer Produktion	1
1.2 Zielsetzung dieser Arbeit.....	2
2. Mikro- und molekularbiologische Grundlagen	4
2.1 Das potentielle Malariavakzin D1M1H als Zielprodukt.....	4
2.2 Das eingesetzte Expressionssystem	5
2.2.1 Die methylotrophe Hefe <i>Pichia pastoris</i>	5
2.2.2 Transformation des Organismus.....	6
3. Eingesetzte Bioprosesstechnik	7
3.1 Kultivierung von <i>Pichia pastoris</i>	7
3.1.1 Der klassische Herstellungsprozess rekombinanter Proteine	7
3.1.2 Intensivierung der Produktion durch zyklische Prozessführung	8
3.2 Die verwendete Bioreaktoreinheit	9
3.3 Erweiterte MSR- und Automatisierungstechnik	10
3.3.1 Automatisierungsaufgaben	10
3.3.2 Der substratlimitierte Glycerol-Fed-Batch	11
3.3.3 Online-Estimierung der Zelldichte	12
3.3.4 Inline-Messung und Regelung der Methanolkonzentration	13
3.3.5 Atline-Quantifizierung des Zielpunktes	13
3.3.6 Realisierung der zyklischen Fahrweise.....	14
3.3.7 Abgasanalyse und Gasbilanzen.....	14
3.4 Prozessbegleitende Offline-Analysentechnik.....	15
3.4.1 Probenahme	15
3.4.2 Zelldichtebestimmung	15
3.4.3 Glycerol- und Methanolanalytik.....	16
3.4.4 Ammoniummessung.....	16
3.4.5 Gesamtproteinbestimmung	16
3.4.6 Enzymatischer Alkoholoxidasesnachweis	16
3.4.7 Berechnung von Konzentrationen in der Flüssigphase	17

4. Spektroskopische Messverfahren als erweiterte PAT-Werkzeuge	18
4.1 Eine Übersicht über PAT-Analysatoren.....	18
4.2 Eingesetzte Spektroskopieverfahren.....	19
4.2.1 Nahinfrarotspektroskopie.....	19
4.2.2 Raman-Spektroskopie	22
4.2.3 2D-Fluoreszenzspektroskopie	24
4.3 Extraktion relevanter Informationen aus Spektren	26
5. Einführung in die Multivariate Datenanalyse (MVDA)	27
5.1 Zielsetzung bei der Anwendung der MVDA	27
5.2 Datenvorbereitung	28
5.2.1 Struktur und Modifizierung multivariater Datensätze.....	28
5.2.2 Zentrierung und Skalierung	28
5.2.3 Datenvorverarbeitung für Spektren	29
5.3 Die Hauptkomponentenanalyse (PCA)	31
5.3.1 Dimensionsreduktion durch Hauptkomponenten	31
5.3.2 Das mathematische Modell der PCA.....	32
5.3.3 Berechnung der Hauptkomponenten.....	33
5.4 Partial Least Squares Regression (PLSR).....	34
5.4.1 Multivariate Kalibrierung mittels PLSR	34
5.4.2 Das mathematische Modell der PLSR.....	35
5.4.3 Berechnung der PLS-Komponenten.....	37
5.5 Erstellung multivariater Modelle	39
5.5.1 Generelle Anforderungen an das Datenmaterial.....	39
5.5.2 Ausreißerdetektion für multivariate Daten	39
5.5.3 Variablenelektion	42
5.5.4 Validierung multivariater Modelle	43
5.6 Multivariate Klassifizierung	45
5.6.1 Allgemeine Informationen	45
5.6.2 Soft Independent Modelling of Class Analogy (SIMCA).....	45
5.6.3 Validierung von Klassifikatoren	47

5.7	Support Vector Machines (SVM).....	48
5.7.1	SVM als multivariates Klassifizierungsverfahren	48
5.7.2	Berechnung einer optimalen Trennebene.....	50
5.7.3	Kernel-Funktionen zur Abbildung nichtlinearer Beziehungen.....	52
5.7.4	Erweiterung zur Support Vector Regression (SVR).....	53
5.8	Eingesetzte MVDA-Software.....	55
6.	Qualitätsbewertung von Hefeextrakten mit NIR-Spektroskopie	56
6.1	Motivation und Zielsetzung	56
6.2	Stand der Wissenschaft	57
6.3	Der gewählte Messaufbau	57
6.4	Vorstellung des Probenpools	58
6.5	Vorverarbeitung der Spektraldaten	58
6.6	Explorative Datenanalyse und Probenselektion	58
6.7	Entwicklung eines SIMCA-Klassifizierungsmodells	61
6.8	Externe Validierung des Modells.....	64
7.	Offline-Prädiktion relevanter Variablen in Bioreaktionsprozessen	66
7.1	Eine Machbarkeitsstudie anhand von Offline-Analysen	66
7.2	Stand der Wissenschaft	66
7.3	Untersuchte Prozessgrößen und vorhandenes Probenmaterial.....	67
7.4	Entwicklung von PLSR-Modellen	68
7.4.1	Der komplexe Prozess der PLSR-Modellentwicklung.....	68
7.4.2	Eine exemplarische Darstellung bei der PLSR-Modellentwicklung	69
7.4.3	Schwierigkeiten bei Betrachtung der zellhaltigen Flüssigphase L	76
7.5	Prädiktion der Zelldichte.....	78
7.6	Prädiktion der Glycerolkonzentration	80
7.7	Prädiktion der Ammoniumkonzentration	85
7.8	Prädiktion der Gesamtproteininkonzentration.....	87
7.9	Prädiktion des zellinternen Alkoholoxidasegehaltes	90
8.	Online-Monitoring mit spektroskopischen Verfahren.....	93
8.1	Der untersuchte zweistufige Produktionsprozess	93
8.1.1	Die verwendete Bioreaktoranlage	93

8.1.2	Die parallel/sequentielle Prozessführung	94
8.1.3	Die erweiterte Prozess-EDV zur Anwendung der MVDA.....	96
8.2	Skizzierung des Versuchsaufbaus	98
8.3	Prädiktion der Zelldichte und der Glycerolkonzentration.....	99
8.3.1	Bereitstellung geeigneten Datenmaterials.....	99
8.3.2	Ergebnisdarstellung	100
8.4	Prädiktion des zellinternen Alkoholoxidasegehaltes	103
8.4.1	Bereitstellung geeigneter Kalibrierdaten	103
8.4.2	Ergebnisdarstellung	103
8.5	Prädiktion der Gesamtproteinkonzentration	104
8.6	Anwendung der nichtlinearen SVR	105
9.	Regelung der Glycerolkonzentration	110
9.1	Das Regelungskonzept	110
9.2	Theoretische Betrachtung des Regelungsproblems	111
9.2.1	Elemente des Regelkreises	111
9.2.2	Charakterisierung des Streckenverhaltens	111
9.2.3	Einführung des linearisierten Streckenmodells	114
9.2.4	Vernachlässigung der Dynamik des Messsystems	116
9.2.5	Regel- und Stellverhalten	117
9.3	Untersuchung der Dynamik des Regelungsproblems.....	118
9.3.1	Übertragungsfunktionen des Regelkreises	118
9.3.2	Eigenwerte des geschlossenen Regelkreises.....	119
9.3.3	Schwingungsverhalten des Regelkreises.....	121
9.3.4	Vorgabe des Regelkreisverhaltens.....	121
9.4	Technische Vorgaben der Glycerolregelung	122
9.4.1	Prozesstechnische Umsetzung	122
9.4.2	Bereitstellung der Regelgröße durch ein PLSR-Modell	123
9.4.3	Durchführung notwendiger Online-Berechnungen	124
9.4.4	Vernetzung beteiligter Softwaresysteme	124
9.5	Experimentelle Erprobung der Substratregelung	125
9.5.1	Regelung im aperiodischen Grenzfall.....	125
9.5.2	Regelung der Glycerolkonzentration im Schwingfall	127

10. Anwendung der MVDA auf nichtspektroskopische Daten	129
10.1 Prädiktion nicht direkt messbarer Prozessgrößen	129
10.2 Vorbereitung der Modellerstellung	130
10.2.1 Gewählte Zielgrößen	130
10.2.2 Berechnung zellspezifischer Reaktionsraten	131
10.2.3 Bereitstellung idealisierter Kalibrierdaten.....	132
10.2.4 Auswahl und Bereitstellung der Prädiktorvariablen	138
10.2.5 Erzeugung benötigter Datensätze und Datenvorverarbeitung.....	138
10.3 Ergebnisdarstellung.....	139
10.3.1 Prädiktion von Zustandsgrößen	139
10.3.2 Prädiktion zellspezifischer Reaktionsraten.....	143
11. Zusammenfassung	147
12. Anhang	149
12.1 Kulturmedien	149
12.2 Kultivierungsbedingungen.....	150
12.3 Reaktionskinetische Parameter	151
12.4 Offline-Messungen	151
13. Quellenverzeichnis	153
Veröffentlichungen des Autors	160

Nomenklatur

Allgemeine Abkürzungen

AOX	Alkoholoxidase
BMBF	Bundesministerium für Bildung und Forschung
BPRC	<i>Biomedical Primate Research Centre</i>
D1M1H	Potentielles Malariavakzin, Fusionsprotein aus <i>PfAMA1</i> und <i>PfMSP1</i>
DCU	Bioreaktor-Kontrolleinheit (<i>Digital Control Unit</i>)
EMA	<i>European Medicines Agency</i>
FDA	<i>Food and Drug Administration</i>
HPLC	<i>High Performance Liquid Chromatography</i>
ICH	<i>International Conference on Harmonization</i>
IMAC	<i>Immobilized Metal chelate Affinity Chromatography</i>
MVDA	Multivariate Datenanalyse
NIR, NIRS	Nahinfrarot, Nahinfrarotspektroskopie
OPC	<i>Open Platform Communications</i>
PAT	<i>Process Analytical Technology</i>
PCA	Hauptkomponentenanalyse (<i>Principle Component Analysis</i>)
<i>PfAMA1</i>	<i>Plasmodium falciparum</i> Apical Membrane Antigen 1
<i>PfMSP1</i>	<i>Plasmodium falciparum</i> Merozoite Surface Protein 1
PLSR	<i>Partial Least Squares Regression</i>
POD	Peroxidase
QbD	<i>Quality by Design</i>
RI	Bréchungsindex (<i>refractive index</i>)
RTR	<i>Real-Time Release</i>
SCADA	<i>Supervisory Control and Data Acquisition</i>
SIMCA	<i>Soft Independent Modelling of Class Analogy</i>
SNV	Standard Normal Variate
SVM, SVR	<i>Support Vector Machines, Support Vector Regression</i>
UV/Vis	Ultraviolet/Visuell, Wellenlängenbereich des elektromagnetischen Spektrums
WHO	<i>World Health Organization</i>

Prozessgrößen

A_K	UV/Vis- oder NIR-Absorptionsmessung im Teilsystem K	[AU]
a_{P_1}	Eigenwert der Produktbildungsräte	[h^{-1}]
C_{IK}	Stoffmengenkonzentration von Komponente I in Teilsystem K	[mol $^{-1}$]
C_{P2K}	AOX-Aktivität in Teilsystem K	[UI $^{-1}$]
c_{IK}	Massenkonzentration von Komponente I in Teilsystem K	[g $^{-1}$]
D_A	Gesamtverdünnungsfaktor eines Aufschlussansatzes	[\cdot]
d	Schichtdicke	[cm]
E_{ABTS}	Extinktionsmessung des Chromophors ABTS	[\cdot]

e_{S1L}	Regeldifferenz der Glycerolkonzentration in der Flüssigphase	[gl ⁻¹]
E_L	Trübungsmessung in der Flüssigphase	[AU]
F_K	Volumenstrom in/aus Teilsystem K	[l h ⁻¹]
F_{nl}	Begasungsrate der Komponente I unter Normbedingungen	[l h ⁻¹]
$g_{P2/X}$	zellspezifische AOX-Aktivität	[Ug ⁻¹]
I_K	Raman-Intensitätsmessung in Teilsystem K	[lU]
J_I	Gütekriterium bei Anpassung einer Komponente I	[div.]
IUR	volumetrische Aufnahmerate der Komponente I	[gl ⁻¹ h ⁻¹]
K_J	Verstärkungsfaktor eines Teilsystems J	[div.]
$k_{I,a}$	volumetrischer Sauerstofftransferkoeffizient	[h ⁻¹]
k_{S1}	Monod-Limitierungskonstante für Glycerol	[gl ⁻¹]
M_I	Molare Masse von Komponente I	[gmol ⁻¹]
m_{IK}	Masse von Komponente I in Teilsystem K	[g]
m_K	Masse von Teilsystem K	[g]
N_{St}	Rührerdrehzahl	[min ⁻¹]
p_G	Überdruck in der Gasphase	[bar]
pH	pH-Wert	[\cdot]
pO_2	relativer Gelöstsauerstoffpartialdruck	[$\%$]
Q_I	volumetrischer Massenstrom von Komponente I	[gl ⁻¹ h ⁻¹]
q_{IX}	zellspezifische Reaktionsrate der Komponente I	[h ⁻¹]
R	Korrelationskoeffizient	[\cdot]
RQ	Respirationsquotient	[molmol ⁻¹]
S_K	Fluoreszenzmessung in Teilsystem K	[RFU]
s	Eigenwert	[h ⁻¹]
t	Zeit	[h]
T_J	Zeitkonstante eines Teilsystems J	[h]
V_K	Volumen des Teilsystems K	[l]
V_{nM}	Molares Normvolumen	[l mol ⁻¹]
x_{IG}	Stoffmengenanteil von Komponente I in der Gasphase	[\cdot]
y_{IJ}	Ausbeutekoeffizient der Komponente I aus Komponente J	[gg ⁻¹]
$\alpha_{Z/X}$	Massenverhältnis von feuchten zu trockenen Zellen	[gg ⁻¹]
ε_{ABTS}	molarer Extinktionskoeffizient von ABTS	[l μ mol ⁻¹ cm ⁻¹]
ϑ	Dämpfungsgrad	[\cdot]
ϑ_L	Temperatur in der Flüssigphase	[°C]
λ	Wellenlänge	[nm]
μ	Zellwachstumsrate	[h ⁻¹]
ν	Wellenzahl	[cm ⁻¹]
ρ_K, ρ_Z	Dichte des Teilsystems K, Dichte feuchter Zellen	[gl ⁻¹]
σ	Abklingkonstante	[h ⁻¹]
ω	Kreisfrequenz	[h ⁻¹]

Prozessrelevante Indizes**Komponenten**

Ac	Säure (<i>acid</i>)	P1	D1M1H (Zielprodukt)
AF	Antischaum (<i>anti foam</i>)	P2	Alkoholoxidase (AOX)
AIR	Luft	Ptot	Gesamtprotein
AI	Base (<i>alkali</i>)	S1	Glycerol (Substrat 1)
C, CO ₂	Kohlenstoffdioxid	S2	Methanol (Substrat 2)
N ₂	Stickstoff	X	Biotrockenmasse
O, O ₂	Sauerstoff	Z	Biofeuchtmasse

Teilsysteme

A	Analyse, Aufschlussansatz	R1	Glycerolvorlage
B	Puffersystem	R2	Methanolvorlage
G	Gasphase	R3	Medienvorlage
H	Ernte (<i>harvest</i>)	S	Probe (<i>sample</i>)
L	Flüssigphase (Kulturbrühe)	T1	Säurevorlage
M	Medienphase (Überstand)	T2	Basevorlage

Laufindizes

i	Substrate, Zeitpunkte	p	Reaktoren
j	Zeitpunkte	u	Zeitpunkte
k	Batches/Zyklen, Zeitpunkte		

Zustände, Orte

0	Anfangsbedingung	n	Normbedingungen
ap	aperiodisch	nir	Nahinfrarot
at	atline	off	offline
cdw	Biotrockenmasse	on	online
em	Emission	op	Arbeitspunkt (<i>operating point</i>)
est	estimiert	P	proportional, Periodendauer
ex	Anregung (<i>exitation</i>)	pls	vorhergesagt mit PLSR
fia	Fließinjektionsanalyse	R	Regler
flu	2D-Fluoreszenz	ram	Raman
gr	Wachstumsanteil (<i>growth</i>)	rel	relativ
hplc	gemessen mit HPLC	S	Strecke
I	Integration	svr	vorhergesagt mit SVR
in	Eingang (<i>inlet</i>)	tot	total
m	Erhaltungsstoffwechsel	turb	Trübung (<i>turbidity</i>)
max	maximal	w	Sollwert
min	minimal	zi	Störgröße

Matrizen und Vektoren in der MVDA

B	(m × v)	PLS-Regressionskoeffizientenmatrix
b _j	(m × 1)	Spaltenvektor mit PLS-Regressionskoeffizienten
D	(n × m)	Datenmatrix (Spektren oder Prozessdaten)
d _j	(n × 1)	Spaltenvektor der Datenmatrix D
E	(n × m)	Residuenmatrix des X-Datenraums
F	(n × v)	Residuenmatrix des Y-Datenraums
G	(n × v)	Residuenmatrix des PLS-Regressionsansatzes
M	(n × v)	Messdatenmatrix (Zielgrößen bei der PLSR)
P	(m × r)	Loadingmatrix des X-Datenraums
p _I	(m × 1)	Loading-Spaltenvektor der Komponente I (X-Daten)
Q	(v × r)	Loadingmatrix des Y-Datenraums
q _I	(v × 1)	Loading-Spaltenvektor der Komponente I (Y-Daten)
T	(n × r)	Scorematrix des X-Datenraums
t _I	(n × 1)	Score-Spaltenvektor der Komponente I (X-Daten)
U	(n × r)	Scorematrix des Y-Datenraums
u _I	(n × 1)	Score-Spaltenvektor der Komponente I (Y-Daten)
w	(m × 1)	Wichtungs-Spaltenvektor der Entscheidungsfunktion (SVR)
W	(m × r)	Weightmatrix eines PLSR-Modells
w _I	(m × 1)	Weight-Spaltenvektor der Komponente I
X	(n × m)	modifizierte Datenmatrix (Spektren oder Prozessdaten)
x _j	(n × 1)	Spaltenvektor der Datenmatrix X
Y	(n × v)	modifizierte Messdatenmatrix (Zielgrößen bei der PLSR und SVR)
y _h	(n × 1)	Spaltenvektor der Messdatenmatrix Y

MVDA-relevante Bezeichnungen

b	Bias der Entscheidungsfunktion (SVR)
C	Güteparameter zur Wichtung der Fehler bei der SVR
c	Koeffizient der inneren Beziehung eines PLSR-Modells
DModX	Distanz zum Modell im X-Datenraum (Ausreißerdetektion)
E _C	Klassifizierungsfehler [%]
F	Wert einer F-Verteilung
g	Trennebene, Entscheidungsfunktion (SVR)
K	Kernel-Funktion (SVR)
L, L _d	Lagrange-Funktional (SVR)
m	Anzahl an Variablen/Spalten in der Datenmatrix X
n	Anzahl an Zeilen in der Datenmatrix X
n _K	(allgemein) Anzahl an Beobachtungen in der Gruppe K
r	Anzahl an Komponenten/Spalten in der Scorematrix T
R ² _{cv}	Güte der Vorhersage der Kreuzvalidierung (<i>cross validation cv</i>)

R_P^2	Güte der Vorhersage der externen Validierung (<i>prediction P</i>)
R_X^2	Güte der Anpassung/erklärter Anteil der Varianz (Datenraum X)
R_Y^2	Güte der Anpassung/erklärter Anteil der Varianz (Datenraum Y)
RMSEcv	Vorhersagefehler bei der Kreuzvalidierung [div.]
RMSEP	Vorhersagefehler bei der externen Validierung [div.]
r_{Tal}	Radius einer Hotelling T^2 -Ellipse in Richtung der Komponenten I
S	Spanne zwischen den Eingangsdaten (SVR)
s_{di}	Zeilenstandardabweichung bezogen auf Datenmatrix D
s_{dj}	Spaltenstandardabweichung bezogen auf Datenmatrix D
Se	Sensitivität eines Klassifikators
Sp	Spezifität eines Klassifikators
SSY	Maß für den Anteil erklärter Varianz des Y-Datenraums
T^2	Hotelling T^2 -Wert (Ausreißerdetektion)
v	Anzahl an Variablen/Spalten in der Datenmatrix Y
VIP	Variable Importance on Projection (Variablenelektion)
α	Irrtumswahrscheinlichkeit einer statistischen Hypothese
α_i, β_i	Lagrange-Multiplikatoren der Beobachtung i (SVR)
β_K	Wahrscheinlichkeit der Zugehörigkeit zu einer Klasse K
γ	Parameter des Gauß'schen RBF-Kernels (SVR)
ε	Fehlertoleranz-Parameter bei der SVR

MVDA-relevante Indizes**Laufindizes**

i	Beobachtungen, Objekte	I	Haupt-/PLS-Komponenten
j	X-Variablen	z	Beobachtungen (SVR)
h	Y-Variablen		

Weitere Indizes und Zustände

-	Arithmetischer Mittelwert	opt	optimal
^	Modellschätzwert	p	positiv (Klassenzugehörigkeit)
abs	absolut	nor	normiert
av	Mittelwert (<i>average</i>)	n	negativ (Klassenzugehörigkeit)
crit	kritischer Wert	mc	mittenzentriert
CS	Kalibrierdatensatz	f	falsch (Klassifizierung)
cv	Kreuzvalidierung	de1	erste Ableitung
P	Vorhersage (<i>prediction</i>)	t	wahr (Klassifizierung)
PS	Vorhersagedatensatz	sv	Stützvektor (<i>support vector</i>)
rel	relativ	tot	total
snv	SNV-gefiltert	uv	autoskaliert (<i>unit variance</i>)
te	temporär	VS	Validierdatensatz

Kurzfassung

Jan-Patrick Voß

**Anwendung spektroskopischer Messverfahren
und Multivariater Datenanalyse zur Bewertung
und Beobachtung von Bioprozessen**

Die *Process Analytical Technology* (PAT) Initiative empfiehlt den Einsatz fortschrittlicher Analysensysteme in der pharmazeutischen Produktion. Die vorliegende Arbeit beschreibt wichtige Schritte bei der Etablierung von spektroskopischen Messverfahren in Bioprozessen.

Ein erstes Ziel war die Entwicklung eines Qualitätsbewertungsverfahrens für Hefeextrakte, basierend auf NIR-Spektroskopie und multivariater Klassifizierung. Die gewählte Methode *Soft Independent Modelling of Class Analogy* (SIMCA) führte mit einem Klassifizierungsfehler von 1,5 % zu einem sehr guten Ergebnis.

Ein überwiegender Teil dieser Arbeit widmete sich der Beobachtung zyklischer Kultivierungsprozesse der methylotrophen Hefe *Pichia pastoris* zur Herstellung des potentiellen Malariavakzins D1M1H. Hierbei wurde die Quantifizierung von Zelldichte und AOX-Gehalt sowie Glycerol-, Ammonium- und Produktkonzentration mit *Partial Least Squares Regression* (PLSR), basierend auf NIR-, Raman- und 2D-Fluoreszenzspektren zunächst offline erprobt.

Im Anschluss erfolgte eine Übertragung der erarbeiteten Methoden auf den Online-Betrieb. Dabei kam neben der PLSR auch das nichtlineare Verfahren *Support Vector Regression* (SVR) zum Einsatz. Dieses verbesserte unter anderem die Bestimmung der Glycerolkonzentration mit Raman-Spektroskopie und erreichte einen Vorhersagefehler von ca. 3 %.

Zur Regelung von Glycerol wurden Raman-Spektren und PLSR erfolgreich eingesetzt und damit die technische Relevanz der multivariaten Kalibrierung über eine reine Prozessbeobachtung hinaus demonstriert. Die Realisierung erfolgte mit einer Störgrößenaufschaltung und einer ergänzenden adaptiven Regelung der Abweichungen vom Sollwert.

Bei der technischen Umsetzung kamen mit SIMATIC SIPAT, Umetrics SIMCA und MATLAB® eine Reihe industrierelevanter Softwarepakete innerhalb einer komplexen Prozess-EDV in einem zweistufigen Produktionsprozess zum Einsatz.

Den Abschluss dieser Arbeit bildet die Anwendung multivariater Kalibrierverfahren auf nicht-spektroskopische Online-Prozessgrößen. Hierbei wurden die obigen biotechnologischen Variablen über 15 klassische Online-Messgrößen eines Bioreaktorprozesses ermittelt. Eine Bereitstellung geeigneter Trainingsdaten erfolgte durch Simulation und Parameteridentifikation von Modellen mit Massenbilanzen zur Glättung von Offline-Analysen.

Die daraus resultierende Bestimmung der Zielproduktkonzentration erreichte mit der nicht-linearen SVR einen Vorhersagefehler von 3,2 %. Die Quantifizierung zellspezifischer Reaktionsraten für Zellmasse, Glycerol, Methanol, Sauerstoff und Produkt war auf diesem Wege ebenfalls erfolgreich. Mit einem Fehler von 2,7 % war die Prädiktion der Aufnahmerate von Methanol am besten.

Schlüsselworte: Bioprozessmonitoring; Spektroskopie; Multivariate Datenanalyse; Chometrie; Process Analytical Technology; *Pichia pastoris*; Malariaimpfstoff

Abstract

Jan-Patrick Voß

**Application of spectroscopic measurement methods
and multivariate data analysis for evaluation
and monitoring of bioprocesses**

The *Process Analytical Technology* (PAT) initiative proposes the application of advanced analysis systems in pharmaceutical production. This work describes important steps in establishing spectroscopic measurement methods in bioprocesses.

A first objective was the development of a quality evaluation system for yeast extracts based on NIR-spectroscopy and multivariate classification. The selected method *Soft Independent Modelling of Class Analogy* (SIMCA) gave excellent results by obtaining a classification error of 1.5 %.

Most of the work was dedicated to the monitoring of cyclic cultivation processes of the methyotropic yeast *Pichia pastoris* for production of the potential malaria vaccine D1M1H. First, quantification of cell density, AOX content as well as glycerol, ammonia and product concentration with *Partial Least Squares Regression* (PLSR) based on NIR-, Raman- and 2D-fluorescence spectra was tested off-line.

A transfer of developed methods to the on-line operation took place subsequently. Besides PLSR, the non-linear method *Support Vector Regression* (SVR) was also used on-line and enabled the improvement of glycerol determination with Raman-Spectroscopy. A prediction error of approx. 3 % was achieved.

For glycerol control, Raman spectra and PLSR were successfully applied, thus demonstrating the technical relevance of multivariate calibration beyond mere process monitoring. The implementation was carried out with a feed-forward control and an additional adaptive compensation of the deviations from the setpoint.

A complex data processing system, involving software with industrial relevance like SIMATIC SIPAT, Umetrics SIMCA and MATLAB[®], was used for the technical implementation in a two-stage production process.

In conclusion of this work, multivariate calibration methods were applied to non-spectroscopic on-line process data. Here, the biotechnological variables mentioned above were determined using 15 classical on-line measurements of a bioreactor process. Suitable training data was provided by smoothing off-line measurements via simulation of models with mass balances and parameter identification.

The resulting determination of target protein concentration with the non-linear SVR obtained an error of 3.2 %. In addition, quantification of cell specific reaction rates for cell mass, glycerol, methanol, oxygen and product was also successful in this context. With an error of 2.7 %, prediction of methanol uptake performed the best.

Keywords: bioprocess monitoring; spectroscopy; multivariate data analysis; chemometrics; Process Analytical Technology; *Pichia pastoris*; malaria vaccine

1. Einleitung und Zielsetzung

1.1 PAT – Ein Werkzeug moderner pharmazeutischer Produktion

Die Herstellung pharmazeutischer Produkte unterliegt strengen Vorschriften durch Zulassungsbehörden wie der *U.S. Food and Drug Administration* (FDA) und der *European Medicines Agency* (EMA). Dabei stehen die Gewährleistung einer hohen reproduzierbaren Qualität und Effizienz der Arzneimittel und damit die Sicherheit des Patienten im Vordergrund.

Auch kleinere Änderungen bestehender Produktionsprozesse bedürfen, auch bei technischen Fortschritten, daher einer Genehmigung durch die FDA. Dem steht die Industrie aufgrund des damit verbundenen hohen finanziellen Aufwandes kritisch gegenüber. In der Vergangenheit hat dieser Umstand die Implementierung neuer Technologien und wissenschaftlicher Erkenntnisse in die Herstellung von Pharmaka nahezu unterbunden [RATHORE & WINKLE, 2009].

In den Jahren 2002 und 2003 kündigte die FDA den Start zweier Initiativen an, um Innovationen in den Unternehmen zu fördern und regulatorische Hürden dabei zu mindern [BONDI & DRENNEN, 2011]. Die finale Version der Initiative „Pharmaceutical CGMPs for the 21st Century - A Risk-Based Approach“ aus dem Jahre 2004 beschreibt einen systematischen Entwicklungsansatz der Herstellung von Pharmaka bei dem die Qualität des Produktes bereits bei dessen Herstellung sichergestellt werden soll [FDA, 2004b].

Eine fundamentale Annahme dieses Quality by Design (QbD) genannten Konzeptes liegt darin, dass die Produktqualität von der Prozessvariabilität abhängt und nur dann gesichert werden kann, wenn kritische Quellen dieser Variabilität identifiziert, verstanden und kontrolliert werden können [JAIN, 2014; YU et al., 2014].

Aus diesem Grund veröffentlichte die FDA zur Unterstützung der QbD-Initiative den Industrieleitfaden „PAT - A Framework for Innovative Pharmaceutical Development, Manufacturing and Quality Assurance“, der die Einbindung moderner Analyse- und Kontrollverfahren in die Prozessentwicklung behandelt [FDA, 2004a].

Beide Konzepte sind bereits international anerkannt und adaptiert, was durch eine Reihe von nachfolgenden Richtlinien der *International Conference on Harmonization* (ICH) dokumentiert wurde [ICH, 2005, 2008, 2009]. Die FDA hat damit einen starken Paradigmenwechsel in der pharmazeutischen Produktion angestoßen.

Die *Process Analytical Technology* (PAT) Richtlinie beruht auf prozessbegleitenden, möglichst zeitnahen Messungen kritischer Qualitäts- und Leistungskenngrößen von Rohmaterialien, Zwischen- und Endprodukten sowie des ablaufenden Prozesses selbst. Sie dient damit der wichtigen Erzeugung von Produkt- und Prozessverständnis [JUNKER & WANG, 2006; GLASSEY et al., 2011].

Das Ziel von PAT ist letztlich eine Regelung von Prozesskenngrößen, um damit die finale Produktqualität ggf. dynamisch korrigieren zu können. Das Resultat soll der *Real-Time Release* (RTR) sein, bei dem die Qualität des Produktes nicht am Ende der Herstellung durch Laboranalysen nachgewiesen werden muss, sondern bereits durch den Prozessverlauf sichergestellt ist [STREEFLAND et al., 2013; RATHORE, 2014].

PAT stellt daher auch eine Sammlung an Werkzeugen (*PAT Tools*) zur Realisierung des QbD-Konzeptes dar [GLASSEY et al., 2011]. Wie in Bild 1.1 angedeutet, werden insbesondere moderne Prozessanalysatoren sowie multivariate Analyseverfahren zur Beherrschung großer Datenmengen empfohlen und gefördert [FDA, 2004a].



Bild 1.1: Werkzeuge und Ziele der PAT-Richtlinie [VALERO, 2013]

In der chemisch-pharmazeutischen Industrie sind bereits einige einfache Prozesse, wie z.B. Vermahlungen oder Tablettierungen unter PAT-Aspekten zugelassen. Aus dem biopharmazeutischen Sektor sind dagegen nur sehr wenige Anwendungen bekannt [STREEFLAND et al., 2013; RATHORE, 2014].

Als Ursache gilt weithin die hohe Komplexität biotechnologischer Kultivierungsprozesse, von denen die Produktqualität maßgeblich beeinflusst wird. Die Begasung und Durchmischung aeroben Prozesse, komplexe Kulturmedien, die Zellen selbst sowie nichtlineare Verläufe und Abhängigkeiten der Prozesse erschweren eine robuste Messung relevanter Variablen. Eine hohe Stamm- und Prozessspezifität sowie Inhomogenitäten im großen Maßstab stehen einer industriellen Umsetzung zusätzlich im Wege [MERCIER et al., 2014; SIMON et al., 2015].

1.2 Zielsetzung dieser Arbeit

Das Ziel dieser Arbeit ist, einen Beitrag zur PAT-Initiative zu leisten. Dabei stand die Anwendung spektroskopischer Messverfahren und der Multivariaten Datenanalyse (MVDA) als PAT-Werkzeuge im Fokus.

Eine erste Aufgabenstellung bestand in der Entwicklung eines Qualitätsbewertungssystems für verschiedene Hefeextrakte, basierend auf Nahinfrarotspektroskopie (NIRS). Hefeextrakte kommen als Medienbestandteil in Kultivierungsprozessen zum Einsatz und haben durch ihren hohen Nährstoffgehalt einen starken Einfluss auf das Wachstum der Zellen sowie die Produktivität des Prozesses [KEIL & TILKINS, 2013; MOSSER et al., 2013].

Da jedoch eine Abhängigkeit von der Produktcharge besteht, ist eine Qualitätsprüfung des Rohmaterials notwendig. Durch den Einsatz der NIRS und der MVDA sollte daher unter PAT-Aspekten ein schnelles Verfahren entwickelt werden, das keine aufwendigen Laboranalysen erfordert.

Am Forschungs- und Transferzentrum für Bioprozess- und Analysentechnik der HAW Hamburg wurden unter der Leitung von Prof. Dr.-Ing. R. Luttmann in den letzten Jahren zyklische Prozesse zur Herstellung des potentiellen Malariavakzins D1M1H mit der methyloptrophen Hefe *Pichia pastoris* aufgebaut [MARTENS, 2014; FRICKE, 2014; LÖGERING, 2015] und unter QbD- sowie PAT-Aspekten weiterentwickelt und untersucht [FRICKE, 2014; BORCHERT, 2015].

Diese Arbeit sollte an die vorangegangen Forschungsergebnisse unter PAT-Gesichtspunkten anknüpfen. Daher beschäftigt sich ein überwiegender Teil mit der Online-Beobachtung der zyklischen Produktionsprozesse. Hierbei sollten nicht direkt online bestimmbare Variablen durch spektroskopische Messungen und mit Hilfe multivariater Kalibrierverfahren abgeschätzt werden.

In einem ersten Schritt dieses Arbeitspaketes erfolgte anhand von spektralen Offline-Messungen mit den Methoden Nahinfrarot, Raman sowie 2D-Fluoreszenz eine Machbarkeitsstudie zur Quantifizierung ausgewählter Zielgrößen. Hierbei handelt es sich um die Zeldichte (c_{XL}) sowie um die Konzentrationen in der Flüssigphase von Glycerol (c_{SIL}), Ammonium (C_{AtoL}) und sekretiertem Gesamtprotein (c_{ProL}). Weiterhin sollte auch die Bestimmung des zellinternen Gehaltes $g_{P2/X}$ eines Stoffwechselenzyms der Hefe getestet werden. Als multivariates Regressionsverfahren kommt dabei *Partial Least Squares* (PLSR) zum Einsatz.

Durch die Offline-Analysen werden Störquellen, wie beispielsweise die Begasung in Verbindung mit einer Durchmischung des Bioreaktors, ausgeschlossen. Zusätzlich ist eine Untersuchung des Einflusses von Zellen auf die multivariaten Kalibriermodelle möglich. Die Untersuchungen sollten daher optimale Einstellungen, beispielsweise hinsichtlich der Vorverarbeitung der Spektren oder des gewählten Spektralbereiches, hervorbringen.

In einem zweiten Schritt waren erfolgreiche Methoden und gewonnene Erkenntnisse auf den Online-Betrieb zu übertragen. Zur Kompensation erwarteter Störeinflüsse sollte hier auch die nichtlineare Regressionsmethode *Support Vector Regression* (SVR) erprobt und mit der PLSR verglichen werden.

Als weiterer Schritt zur Etablierung einer PAT-Anwendung wurde die Regelung der Glycerolkonzentration c_{SIL} mit spektroskopischer Messung und PLSR zum Ziel gesetzt. Hierbei sollten eine Störgrößenaufschaltung sowie eine Adaption der Reglerparameter zum Einsatz kommen, um das stark nichtlineare Prozessverhalten zu kompensieren.

Im letzten Teil dieser Arbeit war die Anwendung der multivariaten Regressionsverfahren PLSR und SVR auf nichtspektroskopische Prozessdaten auszuweiten. Ziel dabei war, ebenfalls eine Vorhersage nicht direkt messbarer Variablen zur Erweiterung des Prozessmonitorings nach PAT-Gesichtspunkten ohne komplexe spektrale Messungen zu erreichen.

Neben Zustandsgrößen wie der Zeldichte, Substrat- oder Produktkonzentrationen sollte hierbei auch die Bestimmung zellspezifischer Reaktionsraten, wie beispielsweise die Zielproduktbildung $q_{P1/X}$, näher untersucht werden. Hierfür standen eine Reihe von Inline- und Online-Messgrößen, wie beispielsweise Waagensignale oder die Ergebnisse einer Abgasbilanzierung aus hochinstrumentierten Bioreaktoren zur Verfügung.

2. Mikro- und molekularbiologische Grundlagen

2.1 Das potentielle Malariavakzin D1M1H als Zielprodukt

Malaria ist eine der gefährlichsten Infektionskrankheiten des Menschen. Trotz eines Rückgangs von Infektionen und Todesfällen erkrankten im Jahr 2015 schätzungsweise 214 Millionen Menschen. Für 438 Tausend Patienten verlief die Krankheit tödlich [WHO, 2015].

An der Entwicklung möglicher Impfstoffe wird seit Jahrzehnten intensiv geforscht. Dennoch ist bisher kein Vakzin kommerziell erhältlich. Mit Mosquirix™, einem Produkt der Fa. Glaxo-Smith-Kline, Brentford, UK, erhielt der erste Malariaimpfstoff im Juli 2015 eine Zulassung der Europäischen Arzneimittel-Agentur (EMA) [EMA, 2015].

Der Erreger *Plasmodium* nutzt eine Reihe von Mechanismen, um dem menschlichen Immunsystem zu entgehen. Hieraus erwachsen Schwierigkeiten bei der Entwicklung eines umfassend wirksamen Impfstoffes. Insbesondere weisen mögliche Antigene eine hohe Diversität zwischen den Stämmen einer Spezies auf. Eine Vakzinierung mit entsprechendem Antigen ist daher nicht gegen die gesamte Spezies, sondern nur stammspezifisch wirksam [KENNEDY et al., 2002; HEALER et al., 2004].

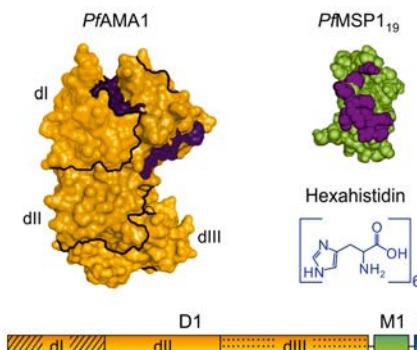


Bild 2.1: Komponenten und vereinfachte Struktur von D1M1H [MARTENS, 2014]

Bei dem in Bild 2.1 gezeigten artifiziellen Konstrukt D1M1H handelt es sich um ein potentielles Malariavakzin, das in dieser Arbeit als Zielprotein dient. Es hat eine Molekulmasse von 63 kDa und besteht aus den drei Komponenten D1, M1 und H wobei letztere für einen Hexahistidinschwanz (His-Tag) steht. Dieser vereinfacht die Aufreinigung und Quantifizierung des Produktes mit der Metallchelat-Affinitätschromatographie (IMAC).

Die Komponenten D1 und M1 sind von den Oberflächenproteinen *Apical Membrane Antigen 1* (*PfAMA1*) und *Merozoite Surface Protein 1* (*PfMSP1*) abgeleitet. Sie entstammen *Plasmodium falciparum*, dem Erreger der besonders schweren Malaria tropica. Es handelt es sich um zwei aussichtsreiche Kandidaten möglicher Impfstoffe, die jedoch den beschriebenen Polymorphismus aufweisen [NIKODEM & DAVIDSON, 2000; KOCKEN et al., 2002; MITCHELL et al., 2004].

Am *Biomedical Primate Research Centre* (BPRC), Rijswijk, Niederlande, wurde deshalb ein sogenannter *Diversity-Covering* (DiCo) Ansatz entwickelt, der für das *PfAMA1* drei Varianten (D1, D2, D3) hervorbrachte. Diese decken den natürlichen Polymorphismus des Oberflächenproteins zu 97 % ab [REMARQUE et al., 2008].

Durch Kombination der DiCo-Proteine mit zwei Versionen (M1, M2) eines 19 kDa schweren Fragmentes von *P/MSP1* wurden darüber hinaus artifizielle Fusionsproteine geschaffen, die einen höheren Immunisierungserfolg erzielen als die Verabreichung einzelner Varianten der Antigene [FABER et al., 2007, 2013].

Zugunsten einer Produktivitätssteigerung mit dem gewählten Expressionssystem *Pichia pastoris* wurde das sich ergebende Gesamtkonstrukt D1M1D2M2D3 auf die beiden Untereinheiten D1M1 und D2M2D3 aufgeteilt. Der Nachteil hierbei besteht in der Notwendigkeit, ein geeignetes Vakzingemisch zu entwickeln [MARTENS, 2014].

Eine detaillierte Beschreibung der Entwicklung und Eigenschaften der obigen Malariavakzinkandidaten befindet sich in [MARTENS, 2014].

2.2 Das eingesetzte Expressionssystem

2.2.1 Die methylotrophe Hefe *Pichia pastoris*

Die zu der Abteilung *Ascomycota* gehörende Hefe *Pichia pastoris* ist ein einzelliger Mikroorganismus und eines der wichtigsten biotechnologischen Wirtssysteme zur Herstellung artfremder Proteine.

Seit 1993 kann *Pichia pastoris* in akademischen Forschungs- und Entwicklungsarbeiten lizenziert frei eingesetzt werden und gehört dadurch zu den am besten charakterisierten Expressionssystemen. Weiterhin zeichnet sich die Hefe durch eine sehr gute Handhabbarkeit hinsichtlich gentechnischer Arbeiten sowie ihrer Kultivierung aus [HIGGINS & CREGG, 1998].

Weitere Vorteile von *Pichia pastoris* bei der Produktion rekombinanter Proteine bestehen zum einen in der Fähigkeit zur Sekretion, wodurch sich die Aufreinigung des Zielproduktes vereinfacht. Zum anderen beherrscht die eukaryotische Hefe viele posttranskriptionale Modifikationen und ist dadurch in der Lage, die meisten Proteine in ihrer biologisch aktiven Form herzustellen [CREGG et al., 2000].

Es handelt sich um einen fakultativ methylotrophen Organismus. Dies bedeutet, dass Methanol als einzige Kohlenstoff- und Energiequelle genutzt werden kann. Für eine Verwertung von Methanol ist das Enzym Alkoholoxidase (AOX) hauptverantwortlich. Die Expression von AOX wird durch Methanol selbst induziert. Andere Substrate wie zum Beispiel Glucose oder Glycerol hemmen dagegen den Methanolstoffwechsel und werden bevorzugt metabolisiert [ELLIS et al., 1985].

Die Hefe besitzt zwei Gene, die für zwei Varianten der Alkoholoxidase (AOX1, AOX2) kodieren. Bei Wachstum auf Methanol kann AOX bis zu 30 % des gelösten zellinternen Proteins ausmachen. Die Ursache ist eine geringe Affinität des Enzyms zum Edukt Sauerstoff, was mit einer verstärkten Expression kompensiert wird. Verantwortlich hierfür ist der sehr starke Promotor P_{AOX1} , der die Transkription des AOX1 kontrolliert. Letzteres ist somit für den Großteil des Methanolstoffwechsels verantwortlich [KOUTZ et al., 1989].

Unterstellt man der Kontrolle dieses Promotors das Gen eines heterologen Proteins kann *Pichia pastoris* für dessen biotechnologische Herstellung genutzt werden. Man erhält dann ein Expressionssystem, das sich besonders durch hohe Produktausbeuten auszeichnet. Durch den Induktionsmechanismus mit Methanol ist die Proteinexpression außerdem steuerbar, was eine flexible Prozessführung ermöglicht [TSCHOPP et al., 1987; ROMANOS, 1995].

2.2.2 Transformation des Organismus

Die gentechnische Modifizierung von *Pichia pastoris* erfolgte mit dem kommerziellen Vektor-System Invitrogen™ pPICZαA der Fa. Thermo Fisher Scientific, Waltham, USA.

Nach der Klonierung des Zielgens in das ringförmige Plasmid, dessen Amplifikation in *Escherichia coli* erfolgt, sowie einer nachfolgenden Linearisierung des Vektors, ist eine Einschleusung des genetischen Materials in *Pichia pastoris* durch Elektroporation vorzunehmen. Bild 2.2 zeigt schematisch den Ablauf der Transformation.

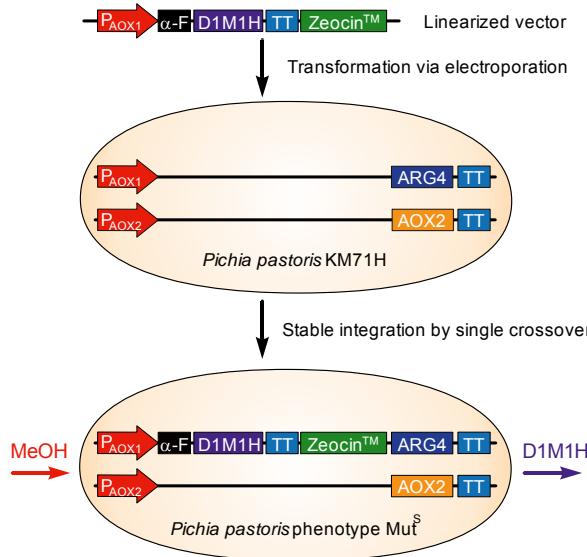


Bild 2.2: Transformation von *Pichia pastoris* KM71H zur Erzeugung eines Produktionsstammes mit dem Phänotyp Mut^S [MARTENS, 2014]

Das Plasmid enthält einen Großteil der Sequenz des AOX1-Promotors (P_{AOX1}) in deren Bereich es zu einer Anlagerung an die homologe Wirts-DNA kommt. Durch ein sogenanntes Crossover wird die Expressionscassette stabil in das Genom von *Pichia pastoris* integriert [TSCHOPP et al., 1987; HIGGINS & CREGG, 1998].

Neben der Gensequenz des Zielproduktes D1M1H wird außerdem ein sekretorisches Signal (α -F) für dessen Ausschleusung sowie eine AOX1-Transkriptionsterminationssequenz (TT) und ein Zeocin™-Resistenzgen für die Isolation erfolgreicher Transformanden in das Erbgut der Hefe eingebracht.

In der vorliegenden Arbeit kam der Stamm *Pichia pastoris* KM71H zum Einsatz. Bei diesem ist das AOX1-Gen durch eine Fremdsequenz (ARG4) ersetzt worden. Ein solcher Organismus kann Methanol nur noch mit Hilfe des AOX2 verwerten. Dies äußert sich in einer geringeren Wachstumsrate auf diesem Substrat. Der entsprechende Phänotyp der Hefe wird daher *methanol utilization slow* (Mut^S) genannt [HIGGINS & CREGG, 1998].

3. Eingesetzte Bioprozesstechnik

3.1 Kultivierung von *Pichia pastoris*

3.1.1 Der klassische Herstellungsprozess rekombinanter Proteine

Der in Bild 3.1 gezeigte klassische Prozessverlauf zur Herstellung rekombinanter Proteine mit *Pichia pastoris* besteht aus drei Phasen.

In einem ersten Abschnitt wachsen die Zellen (c_{XL}) unlimitiert auf dem Primärsubstrat Glycerol (c_{S1M}). Dieses wird zu Beginn in einer bestimmten Menge vorgelegt und bis zum Ende der Batchphase vollständig verbraucht. In dem darauffolgenden Fed-Batch findet substratlimiteretes Wachstum statt. Durch ein exponentielles Profil der Glycerolzufütterrate F_{R1} wird die Zellteilungsrate μ in dieser Phase limitiert, aber konstant gehalten.

Die beiden ersten Prozessphasen werden somit zur Erzeugung von Zellmasse genutzt.

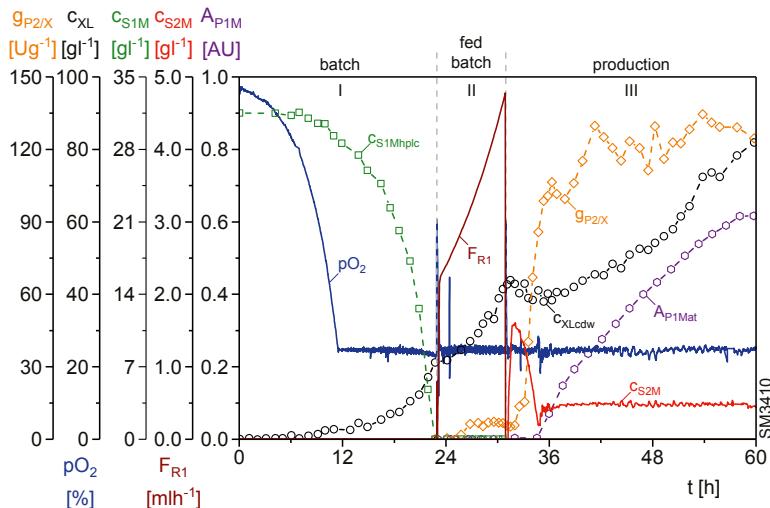


Bild 3.1: Beispiel eines klassischen Produktionsprozesses mit *Pichia pastoris*

Durch eine Dereprimierung des AOX-Promotors unter glycerollimitierten Bedingungen kann eine Adaption der Zellen an die Verwertung von Methanol beschleunigt werden [HIGGINS & CREGG, 1998]. Deswegen dient der Glycerol-Fed-Batch auch der Vorbereitung der Induktion in der nachfolgenden Produktionsphase.

Im diesem Prozessabschnitt wird die Expression und Sekretion des Zielproduktes (A_{P1M}) durch Zufütterung des Sekundärsubstrates Methanol induziert. Die Regelung der Methanolkonzentration c_{S2M} vermeidet eine Substratlimiterierung in der Produktionsphase.

Die maximale Wachstumsrate auf Methanol erreichen die Zellen jedoch erst nach vollständiger Umstellung des Metabolismus. Dieser ist über den zellinternen AOX-Gehalt $g_{P2/X}$ beobachtbar.

Eine Auflistung typischer Sollwerte geregelter Größen sowie die Zusammensetzung des verwendeten Kulturmediums befinden sich im Anhang.

3.1.2 Intensivierung der Produktion durch zyklische Prozessführung

Die maximal mögliche Dauer einer Produktionsphase und damit die Produkttausbeute sind bei der oben beschriebenen Fahrweise aus verfahrenstechnischen Gründen oder aufgrund einer möglichen Produktinstabilität begrenzt [MARTENS, 2014].

Als Folge davon müssten kurze Prozesse durchgeführt werden, die jedoch einen hohen Vorbereitungsaufwand mit sich bringen, der wiederum hohe Kosten verursacht. Deshalb wurden zyklische Prozessführungsstrategien entwickelt, die ausführlich in [MARTENS, 2014] beschrieben werden.

Diese repetitive Fahrweise sieht eine mehrfache Wiederholung der vorgestellten Prozessphasen ohne Neustart der Kultivierung vor. Dies wird dadurch realisiert, dass nach einer unvollständigen Ernte ein Rest der Kulturbrühe im Bioreaktor verbleibt und mit frischem Medium verdünnt als Inokulum eines weiteren Zyklus dient.

Bei der in dieser Arbeit verwendeten Kultivierungsstrategie erfolgt zusätzlich vor Beginn einer jeden Produktionsphase (III) eine Verdünnung der Kulturbrühe auf eine einheitliche Zelldichte. Dies ist in Bild 3.2 gezeigt.

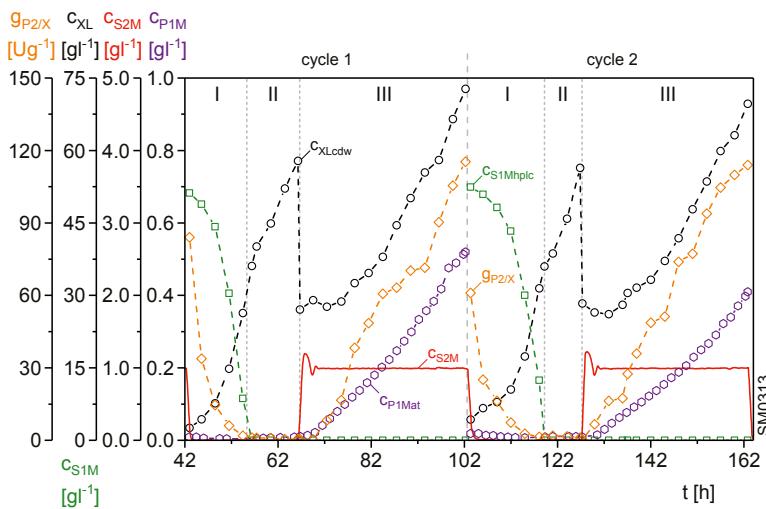


Bild 3.2: Zyklische Prozessführung zur Herstellung von D1M1H

Durch die identischen Startbedingungen bei einer zyklischen Produktion eignet sich diese Fahrweise auch für eine Prozessoptimierung mit statistischer Versuchsplanung (DoE) [MARTENS, 2014]. Eine vollständige Deaktivierung des Methanolstoffwechsels in den Glycerol-Batch- und Fed-Batchphasen bewirkt zudem eine Regeneration der Zellen, was eine langfristige Aufrechterhaltung einer derartigen Prozessführung ermöglicht [FRICKE, 2014].

3.2 Die verwendete Bioreaktoreinheit

Für einen Großteil der experimentellen Arbeiten wurde ein hochinstrumentierter Bioreaktor vom Typ BIOSTAT® C10 der Fa. B. Braun Biotech International, Melsungen, eingesetzt. Das Arbeitsvolumen des Edelstahl-Rührkesselreaktors beträgt 10 l. Bild 3.3 zeigt den eingesetzten Bioreaktor mit seiner erweiterten peripheren Ausstattung.

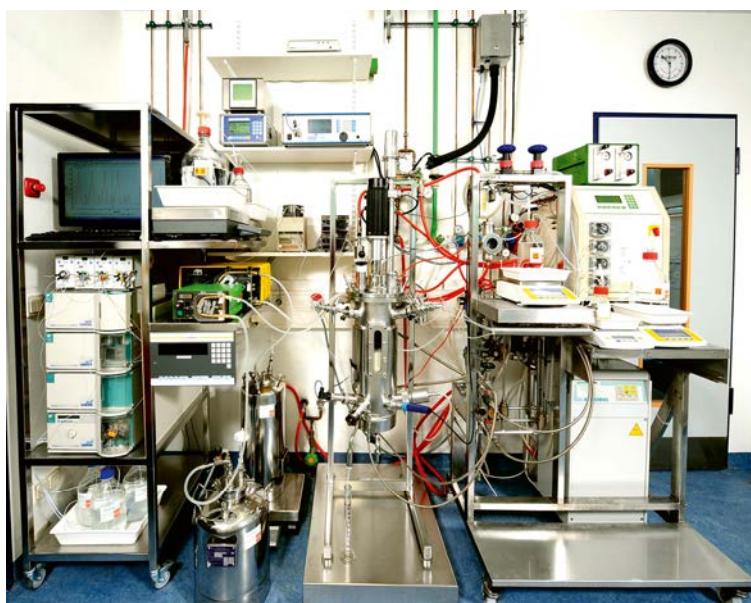


Bild 3.3: Hochinstrumentierter Forschungsbioreaktor BIOSTAT® C10

Eine Regelung der Standardmessgrößen Temperatur ϑ_L , pH-Wert, Begasungsraten F_{nl} , Rütteldrehzahl N_{St} , Druck p_G und Gelöstsauerstoffgehalt pO_2 erfolgt über eine digitale Kontrolleinheit (Mikro-DCU).

Die SCADA-Software *Multi Fermenter Control System for Windows* (MFCS/win) der Fa. Sartorius Stedim Biotech, Guxhagen, ermöglicht über eine zugeschnittene Prozess-EDV eine Datenerfassung und Datenverarbeitung sowie eine übergeordnete Steuerung des Bioreaktorsystems. Sie erlaubt darüber hinaus die Einbindung weiterer peripherer Geräte.

Beispielsweise verfügt der Bioreaktor zur vollständigen Bilanzierung des Prozesses über eine Reihe von Waagen. Tabelle 3.1 enthält eine Aufstellung der beteiligten Prozessgrößen und der vorhandenen peripheren Ausstattung der Anlage.

Auf die Funktionsweise der erweiterten Messtechnik und auf die notwendige Prozessautomatisierung wird im nächsten Abschnitt eingegangen.

Tabelle 3.1: Ausstattung des BIOSTAT® C10

Prozessgröße	Symbol	Hersteller/Typ
Methanolkonzentration	c_{S2M}	Biotechnologie Kempe Alcoline® Silikonflachmembransonde
rekombinantes Zielprodukt (Atline-HPLC)	c_{P1Mat}	VWR-Hitachi LaChrom Elite®, Säule: GE Healthcare HisTrap™ Fast Flow
Schaum im Reaktorkopfraum	D_F	B. Braun Biotech kapazitive Sonde
Trübung der Kulturbrühe	E_L	optek-Danulat ASD25-BT-N-5
zellfreie Probenahme	F_A	TRACE Filtrationssonde, Gilson Minipuls 3
Ernterate	F_H	Watson-Marlow 604U, max. 165 min ⁻¹
Begasungsrate Luft	F_{nAIR}	Bronkhorst High Tech MFC, 0 - 20 lmin ⁻¹
Begasungsrate N ₂	F_{nN2}	Bronkhorst High Tech MFC, 0 - 20 lmin ⁻¹
Begasungsrate O ₂	F_{nO2}	Bronkhorst High Tech MFC, 0 - 20 lmin ⁻¹
Glycerolzufütterung	F_{R1}	Watson-Marlow 101U/R, max. 32 min ⁻¹
Methanolzufütterung	F_{R2}	Watson-Marlow 101U/R, max. 32 min ⁻¹
Refreshmedienzugabe	F_{R3}	Watson-Marlow 503U, max. 220 min ⁻¹
Masse Erntetank	m_H	Sartorius LP32000 S, max. 32 kg
Masse Flüssigphase	m_L	Sartorius F300 S, max. 303 kg
Masse Glycerolreservoir	m_{R1}	Sartorius BA 2100, max. 2.1 kg
Masse Methanolreservoir	m_{R2}	Sartorius LP16000 S, max. 16 kg
Masse Säurereservoir	m_{T1}	Sartorius BA 2100, max. 2.1 kg
Masse Basereservoir	m_{T2}	Sartorius QA7DCE-S, max. 7.5 kg
Rührerdrehzahl	N_{St}	3 × Sechsstahl-Scheibenrührer, Antrieb: Heidolph, 1.240 W, max. 1.500 min ⁻¹
Druck im Reaktorkopfraum	p_G	Phillips P13, 0 - 2,0 bar
pH-Wert	pH	Hamilton EasyFerm PLUS K8 160, pH 0 - 14
relativer Gelöstsauerstoffpegel	pO_2	Mettler-Toledo PN 52200103
Abgas-Stoffmengenanteil CO ₂	x_{CO2}	BlueSens BCP-CO2, 0 - 25 %
Abgas-Stoffmengenanteil O ₂	x_{O2}	BlueSens BCP-O2, 0 - 50 %
Flüssigphasentemperatur	ϑ_L	B. Braun Biotech Pt100
Doppelmanteltemperatur	ϑ_{DU}	B. Braun Biotech Pt100

3.3 Erweiterte MSR- und Automatisierungstechnik

3.3.1 Automatisierungsaufgaben

Durch Orientierung an der in Abschnitt 3.1.2 vorgestellten zyklischen Kultivierungsstrategie lassen sich drei wesentliche Arbeitsschritte identifizieren, die nicht allein mit der Standardinstrumentierung zu bewältigen sind:

- Initialisierung und Steuerung der substratlimitierten Glycerolzufütterung F_{R1}
- Regelung der Methanolkonzentration c_{S2M} in der Produktionsphase
- Durchführung von Harvest- und Refreshoperationen F_H bzw. F_R

Daher ist der Bioreaktor mit einer, in Bild 3.4 beispielhaft gezeigten, erweiterten mess- und verfahrenstechnischen Ausstattung versehen, die eine Durchführung der notwendigen Operationen erlaubt.

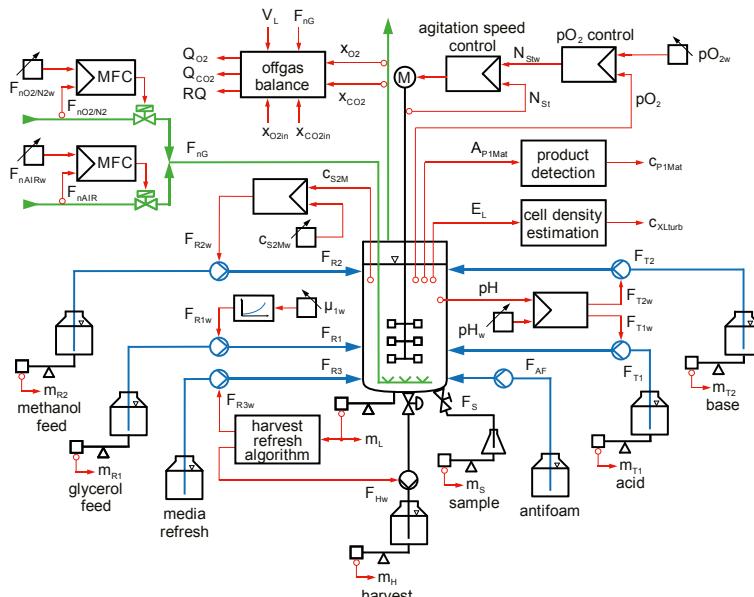


Bild 3.4: MSR Struktur des BIOSTAT® C10

In den folgenden Abschnitten werden die genannten Automatisierungsaufgaben sowie die über eine Standardinstrumentierung hinausgehende Messtechnik kurz vorgestellt. Detaillierte Beschreibungen der einzelnen Methoden finden sich in einer Reihe vorangegangener Arbeiten [CORNELISSEN, 2004; PEUKER, 2006; KAISER, 2010; MARTENS, 2014; FRICKE, 2014; LÖGERING, 2015; ELLERT, 2015; BORCHERT, 2015].

3.3.2 Der substratlimitierte Glycerol-Fed-Batch

Für die automatische Aktivierung einer Glycerol-Fed-Batchphase ist zunächst eine Bestimmung des Startzeitpunktes t_j notwendig. Hierbei ermöglicht eine Überwachung des Gelöstsauergehaltes pO_2 die Erkennung des Batchendes.

Ein plötzlicher Anstieg des geregelten pO_2 offenbart eine starke Reduktion des Sauerstoffbedarfs, die von dem vollständigen Verbrauch des Substrates am Ende einer Batchphase herührt. Ab diesem Zeitpunkt sollen substratlimitierte Bedingungen durch Steuerung der zellspezifischen Wachstumsrate μ_{1wj} auf Glycerol (Substrat S1) eingestellt werden.

Die Realisierung erfolgt mit einer exponentiellen Nachführung des Sollwertes F_{R1w} der Glycerolzufütterung,

$$F_{R1w}(t) = F_{R1j} \cdot e^{\mu_{1wj} \cdot (t - t_j)}, \quad (3.1)$$

der mit seinem Startwert F_{R1j} ,

$$F_{R1j} = F_{R1}(t_j) = \frac{(\mu_{1wj} + y_{X/S1gr} \cdot q_{S1/Xm})}{y_{X/S1gr} \cdot c_{S1R1}} \cdot c_{XLj} \cdot V_{Lj} \quad (3.2)$$

mit

t_j	:= Startzeitpunkt einer exponentiellen Zufütterungsphase	[h]
F_{R1j}	:= Zufütterrate aus Reservoir R1 zum Startzeitpunkt t_j	[l/h ⁻¹]
μ_{1wj}	:= Sollwert der geforderten Wachstumsrate bei t_j	[h ⁻¹]
c_{XLj}	:= Zelldichte bei t_j	[g/l ⁻¹]
V_{Lj}	:= Volumen der Flüssigphase bei t_j	[l]
$q_{S1/Xm}$:= zellspezifische Glycerolaufnahmerate für den Erhaltungsstoffwechsel	[h ⁻¹]
$y_{X/S1gr}$:= Glycerol-Ausbeutekoeffizient für den Zellzuwachs	[-]
c_{S1R1}	:= Glycerolkonzentration im Reservoir R1	[g/l ⁻¹],

online berechenbar ist. Die dafür benötigten reaktionskinetischen Parameter $q_{S1/Xm}$ und $y_{X/S1gr}$ wurden der Arbeit von Martens [MARTENS, 2014] entnommen und sind im Anhang gegeben.

Das Reaktionsvolumen V_{Lj} lässt sich durch Wägung der Flüssigphase zum Zeitpunkt t_j online bestimmen. Die Abschätzung der Zelldichte c_{XLj} erfolgt ebenfalls online und wird mit der im nächsten Abschnitt vorgestellten Trübungsmessung realisiert.

Für die Detektion des Batchendes und die Umsetzung der nachfolgenden μ -Steuerung kam ein in Visual Basic programmiertes Zusatzprogramm (MFCS-Tool) zum Einsatz. Dieses ist in [ELLERT, 2015] näher beschrieben.

3.3.3 Online-Estimierung der Zelldichte

Mit einer Sonde der Fa. optek-Danulat, Essen, wird in der Flüssigphase des Bioreaktors eine Trübungsmessung durchgeführt. Diese beruht maßgeblich auf einer Abschwächung von Nah-infrarotstrahlung durch dispergierte Zellen.

Aus dem Trübungssignal E_L wird die Zelldichte c_{XLturb} ,

$$c_{XLturb}(t) = k_1 \cdot (e^{k_2 \cdot E_L(t)} - 1) \quad (3.3)$$

mit

E_L	:= Trübung der Flüssigphase	[AU]
k_1	:= Zelldichteparameter	[g/l ⁻¹]
k_2	:= Trübungsparameter	[AU ⁻¹],

mit Hilfe eines Exponentialansatzes approximiert.

Die Parameter k_1 und k_2 der nichtlinearen Kennlinie sind postexperimentell durch eine Anpassung von c_{XLturb} an den Verlauf der offline ermittelten Zelldichte c_{XLodw} zu bestimmen.

Dies erfolgt durch Minimierung der Fehlerquadrate mit dem Simplex-Algorithmus von Nelder und Mead [NELDER & MEAD, 1965] mit der Software MATLAB®, The Mathworks, Natick, USA. Ein beispielhaftes Ergebnis ist in Bild 3.5 gezeigt.

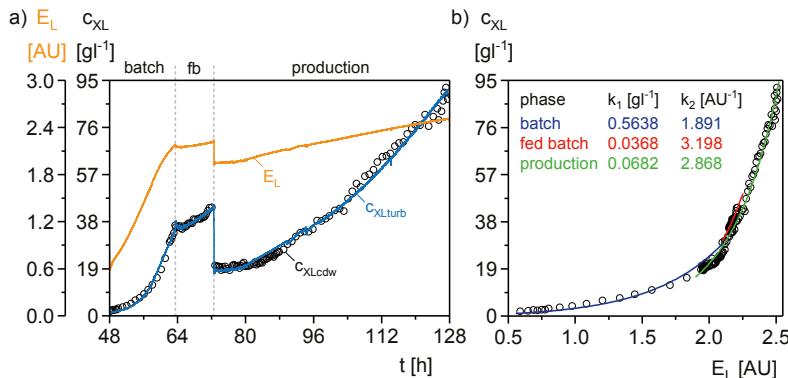


Bild 3.5: Bestimmung der Zelldichte $c_{XL,turb}$ durch Trübungsmessung
a) Kultivierungsverlauf, b) Kennlinien der Trübungsanpassung

Durch eine Abhängigkeit des Trübungssignals E_L von der Rührerdrehzahl N_{St} sowie aufgrund einer Morphologieänderung der Zellen bei Wechsel des Substrates empfiehlt es sich, die Bestimmung eines Parametersatzes für jede Prozessphase getrennt vorzunehmen.

Durch Implementierung von Gl. (3.3) und Hinterlegung der Parameter k_1 und k_2 in MFCS/win steht damit eine Online-Bestimmung der Zelldichte zur Verfügung.

3.3.4 Inline-Messung und Regelung der Methanolkonzentration

Für die Messung der Methanolkonzentration c_{S2M} kam eine Flachmembransonde der Fa. Biotechnologie Kempe, Kleinmachnow, zum Einsatz. In [ELLERT, 2015] findet sich eine detaillierte Beschreibung der zugrunde liegenden Messmethode.

Methanol diffundiert über eine Silikonmembran aus der Flüssigphase des Bioreaktors in den Innenraum der Sonde, wo es mit einem Trägergasstrom zu einem Gassensor (TGS2620) der Fa. FIGARO Engineering, Osaka, Japan, transportiert wird.

Das Messprinzip beruht auf einer Veränderung der elektrischen Leitfähigkeit des Sensormaterials durch reduzierende Verbindungen. Daher lässt sich der messbare elektrische Widerstand für die Bestimmung der Methanolkonzentration heranziehen.

Die Aufrechterhaltung einer konstanten Methanolkonzentration c_{S2M} in der Produktionsphase wird unter Verwendung eines PI-Regelalgorithmus in MFCS/win umgesetzt. Das SCADA-System unterstützt hierfür die Einrichtung digitaler Regler.

3.3.5 Atline-Quantifizierung des Zielproduktes

In einer Induktionsphase ist vor allem die prozessnahe Beobachtung der Produktbildung von großem Interesse. Zu diesem Zweck kam eine Atline-HPLC zum Einsatz, deren Messaufbau in Bild 3.6 gezeigt ist.

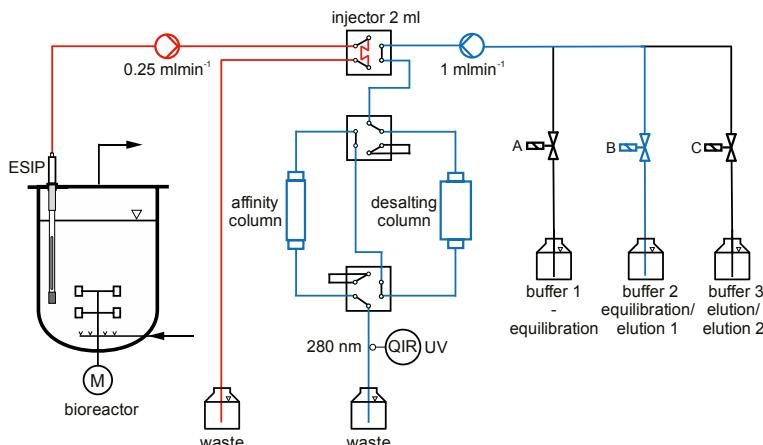


Bild 3.6: Fließschema der Atline-HPLC zur Detektion des Zielproduktes, [MARTENS, 2014]

Über eine Filtrationssonde (ESIP) der Fa. TRACE Analytics, Braunschweig, wird das Chromatographiesystem (VWR-Hitachi, Radnor, USA) mit einem zellfreien Probenstrom aus dem Bioreaktor versorgt. Nach der Injektion der Probe erfolgt zunächst eine Entsalzung und Umpuffierung mit einer Größenausschlussäule (HiTrap™ Desalting, 5 ml, GE Healthcare, Chalfont St. Giles, UK). Danach bindet das Zielprodukt mit seinem His-Tag an eine Affinitätsäule (Nickel-IMAC, HisTrap™ Crude FF, 1 ml, GE Healthcare).

Eine zweistufige Elution erlaubt die Quantifizierung sowohl des Zielproteins als auch auftretender Spaltprodukte. Die Detektion erfolgt photometrisch im UV-Bereich bei einer Wellenlänge von 280 nm.

Die Entwicklung und Funktionsweise der Atline-Chromatographiemethode wird ausführlich in [MARTENS, 2014] dargestellt.

3.3.6 Realisierung der zyklischen Fahrweise

Für eine Umsetzung der Ernte- und Refresh-Operationen ist der Bioreaktor mit einem Medien- und einem Erntetank ausgestattet. Letzterer, sowie der Bioreaktor selbst, befinden sich auf Waagen (Sartorius, Göttingen).

Durch Überwachung der Waagensignale sowie eine Ansteuerung zweier Pumpen (Watson-Marlow, Wilmington, USA) wird eine automatische Durchführung der Prozeduren erreicht. Die Programmierung der dafür notwendigen Algorithmen erfolgt als Rezept nach dem ANSI/ISA S88 Standard in MFCS/win.

3.3.7 Abgasanalyse und Gasbilanzen

In der Abgasstrecke der Anlage sind Gasanalysatoren der Fa. BlueSens gas sensor, Herten, für Sauerstoff und Kohlenstoffdioxid installiert. Das Messprinzip des Sauerstoffsensors basiert auf einer elektrochemischen Sauerstoffpumpzelle. Die Messung des Kohlenstoffdioxids erfolgt dagegen über die Absorption von Infrarotstrahlung bei zwei Wellenlängen.

Mit Kenntnis der Begasungsrate F_{nG} , des Reaktionsvolumens V_L sowie der Stoffmengenanteile x_{OG} und x_{CG} im Ein- und Austritt der Begasung lässt sich eine Bilanzierung der beteiligten Gas-Komponenten O_2 und CO_2 durchführen.

Die Bestimmung der volumetrischen Sauerstoffeintragsrate Q_{O_2} ,

$$Q_{O_2}(t) = Q_{O_2\max}(t) \cdot \frac{x_{OGin}(t) \cdot [1 - x_{CG}(t)] - x_{OG}(t) \cdot [1 - x_{CGin}(t)]}{1 - x_{OG}(t) - x_{CG}(t)}, \quad (3.4)$$

und der Kohlenstoffdioxidaustragsrate Q_{CO_2} ,

$$Q_{CO_2}(t) = Q_{CO_2\max}(t) \cdot \frac{x_{CG}(t) \cdot [1 - x_{OGin}(t)] - x_{CGin}(t) \cdot [1 - x_{OG}(t)]}{1 - x_{OG}(t) - x_{CG}(t)}, \quad (3.5)$$

mit ihren jeweiligen Maximalwerten Q_{imax} ,

$$Q_{imax}(t) = \frac{F_{nG}(t) \cdot M_I}{V_{nm} \cdot V_L(t)} \quad (3.6)$$

mit

Q_{imax}	:= theoretisch maximale volumetrische Eintragsrate, $I = O_2, CO_2$	$[gl^{-1}h^{-1}]$
M_I	:= molare Masse, $I = O_2, CO_2$	$(32\text{ gmol}^{-1}, 44\text{ gmol}^{-1})$
x_{IGin}	:= Stoffmengenanteil in der Begasung, $I = O (O_2), C (CO_2)$	$[-]$
x_{IG}	:= Stoffmengenanteil im Abgas, $I = O (O_2), C (CO_2)$	$[-]$
F_{nG}	:= Gesamtbegasungsrate unter Normbedingungen	$[l h^{-1}]$
V_{nm}	:= molares Normvolumen	$(22,41\text{ lmol}^{-1})$
V_L	:= Volumen der Flüssigphase	$[l]$

sowie die Berechnung des molaren Respirationskoeffizienten RQ ,

$$RQ(t) = \frac{M_{O_2} \cdot Q_{CO_2}(t)}{Q_{O_2}(t) \cdot M_{CO_2}}, \quad (3.7)$$

kann in MFCS/win online erfolgen, wodurch die Gasbilanzen eine Beobachtung des Zellstoffwechsels und eine Beurteilung der Zellaktivität erlauben.

3.4 Prozessbegleitende Offline-Analysentechnik

3.4.1 Probenahme

Eine Beprobung der Prozesse erfolgte zumeist in unregelmäßigen Abständen gemäß des Probenbedarfs des jeweiligen experimentellen Vorhabens. Die entnommene Kulturbrühe wurde aliquotiert, zentrifugiert und bis zur späteren Durchführung der Offline-Analytik in Kulturrührstand und Zellpellet getrennt bei -20°C gelagert.

3.4.2 Zelldichtebestimmung

Für die Bestimmung der Zelldichte c_{XLcdw} über die Biotrockenmasse wurden Zellpellets von 1 ml Aliquots der Kulturbrühe bei 80°C für 48 h getrocknet und anschließend gewogen. Durch Differenzbildung mit den Leermassen der Probengefäße erhält man die Biotrockenmasse, aus der sich die Zelldichte c_{XLcdw} leicht berechnen lässt.

3.4.3 Glycerol- und Methanolanalytik

Die Quantifizierung der Substratkonzentrationen c_{SiMHplc} erfolgte über Ionenausschlusschromatographie mit der Säule Rezex™ RHM-Monosaccharide der Fa. Phenomenex®, Torrance, USA.

Zur Durchführung kam ein HPLC-System der Serie LaChrom Elite® der Fa. VWR-Hitachi, Radnor, USA, zum Einsatz. Dieses verfügt über ein Interface, ein Pumpenmodul (L-7100), einen Autosampler (L-7250) und einen Säulenofen (L-7360). Zur Detektion beider Analyten diente ein Brechungsindexdetektor (RI-Detektor, L-7490). Die Vermessung sowie Auswertung der Proben erfolgte mit der Software D-7000 HSM.

Vor Injektion der Kulturüberstände wurden diese mit einem 0,22 µm Spritzenfilter von etwaigen Partikeln befreit. Das Injektionsvolumen betrug 40 µl. Als mobile Phase diente 10 mM Schwefelsäure, die mit einer Flussrate von 0,6 ml min⁻¹ befördert wurde. Die Säule erfordert eine Temperierung auf 60 °C.

Die Retentionszeit des Glycerols beträgt 14,1 min, wohingegen Methanol nach 18,7 min detektiert werden kann. Für die Ermittlung der Konzentrationen wurde mit den jeweiligen Peakflächen des RI-Signals für beide Substrate eine Zehnpunktkalibrierung durchgeführt.

3.4.4 Ammoniummessung

Das Fließinjektionsanalysensystem TAS 2000 der Fa. Jüke Systemtechnik, Altenberge, wurde für die Bestimmung der Ammoniumstoffmengenkonzentration C_{AltotMfa} nach der Berthelot-Methode eingesetzt. Die Messung basiert auf der Reaktion von Ammonium mit Hypochlorit und Salicylsäure zu Indophenolblau, das bei einer Wellenlänge von 605 nm photometrisch detektiert werden kann [KAISET, 2010].

3.4.5 Gesamtproteinbestimmung

Zur Quantifizierung der Gesamtproteininkonzentration c_{ProtMoff} im Medium kam das QuickStart™ Protein Assay Kit der Fa. Bio-Rad Laboratories, Hercules, USA, zum Einsatz. Dieses beruht auf der Methode von Bradford [BRADFORD, 1976]. Es wurde der Standardprozedur für Mikrotiterplatten gemäß der Anleitung des Herstellers gefolgt.

3.4.6 Enzymatischer Alkoholoxidasanachweis

Zur Bestimmung der zellinternen Alkoholoxidaseaktivität $g_{\text{P2/Xoff}}$ wurde zunächst ein mechanischer Zellaufschluss mit Glasperlen zur Freisetzung der Alkoholoxidase (AOX) durchgeführt. Darauf folgte ein gekoppeltes Enzymassay nach [JANSSEN & RUELIUS, 1968] unter Verwendung des nichttoxischen ABTS™ (2,2'-Azino-bis(3-Ethylbenzothiazolin-6-Sulfonsäure) als alternativer Chromophor [CHILDS & BARDSLEY, 1975].

Das Assay beruht auf einer Umsetzung von Methanol zu Wasserstoffperoxid durch AOX,



und der Oxidation von ABTS™ durch H_2O_2 ,



mit Hilfe des Enzyms Peroxidase (POD).

Der oxidierte Zustand ABTSTM_{ox} ist über das Lambert-Beer'sche Gesetz bei einer Wellenlänge von 405 nm photometrisch quantifizierbar.

Eine Einheit U der AOX-Aktivität ist definiert zu 1 µmol umgesetzten ABTSTM pro Minute. Die AOX-Aktivität C_{P2Ai} im Aufschlussansatz A,

$$C_{P2Ai} = \frac{D_{Ai}}{\varepsilon_{ABTS} \cdot d} \cdot \dot{E}_{ABTSi} \quad (3.8)$$

mit

i	:= Probenindex	[·]
C _{P2Ai}	:= volumetrische AOX-Aktivität im Aufschlussansatz der Probe i	[Ug ⁻¹]
D _{AI}	:= Gesamtverdünnung des Aufschlussansatzes von Probe i	[·]
dot{E} _{ABTSi}	:= zeitliche Extinktionsänderung (ABTS TM _{ox}) der Probe i	[min ⁻¹]
ε _{ABTS}	:= molarer Extinktionskoeffizient (ABTS TM _{ox}) ([KEESEY, 1987] 36,8 lµmol ⁻¹ cm ⁻¹)	
d	:= Schichtdicke des Reaktionsansatzes	[cm],

ist demnach von der linearen Extinktionsänderung \dot{E} abhängig, deren Bestimmung mit Hilfe von zehn Extinktionsmessungen im Abstand von jeweils 30 Sekunden erfolgt.

Die zeitlich zugeordnete zellspezifische AOX-Aktivität g_{P2X}:

$$g_{P2X}(t_i) = \frac{C_{P2Ai}}{c_{Xai}} \quad (3.9)$$

mit

g _{P2X}	:= zellspezifische AOX-Aktivität zum Zeitpunkt t _i	[Ug ⁻¹]
c _{Xai}	:= eingestellte Zeldichte im Aufschlussansatz der Probe i	[gl ⁻¹],

ergibt sich aus der im Aufschlussansatz enthaltenen Zeldichte c_{Xai} und wird in dieser Arbeit als Maß für den zellinternen AOX-Gehalt verwendet.

3.4.7 Berechnung von Konzentrationen in der Flüssigphase

Eine Offline-Bestimmung von Konzentrationen im Bioreaktor erfolgt meist im Kulturüberstand (Medienphase M). Soll die Konzentration auf die Kulturbreue (Flüssigphase L) bezogen werden, ist zu beachten, dass sich diese aus dem Medium M und der Biofeuchtphase Z zusammensetzt. Dies ist bei hohen Zeldichten nicht zu vernachlässigen.

Die Konzentration c_{IL} einer Komponente I im Reaktionsraum (Flüssigphase L),

$$c_{IL}(t) = \left(1 - \frac{\alpha_{Z/X}}{\rho_Z} \cdot c_{XL}(t)\right) \cdot c_{IM}(t) \quad (3.10)$$

mit

c _{IK}	:= Konzentration der Komponente I in Phase K	[gl ⁻¹]
α _{Z/X}	:= Massenverhältnis von feuchten zu trockenen Zellen	(3,7 gg ⁻¹)
ρ _Z	:= Dichte der feuchten Zellen	(1.150 gl ⁻¹),

lässt sich aus der entsprechenden Konzentration c_{IM} im Medium und der zugeordneten Zeldichte c_{XL} berechnen.

4. Spektroskopische Messverfahren als erweiterte PAT-Werkzeuge

4.1 Eine Übersicht über PAT-Analysatoren

Die PAT-Initiative der FDA empfiehlt den Einsatz von Inline-, Online- und Atline-Analysatoren mit dem Ziel, das Verständnis und die Kontrolle über den betrachteten Bioprozess zu verbessern. Der Begriff Analytoren umfasst dabei Systeme, die physikalische, biochemische oder mathematische Analysen hervorbringen [FDA, 2004a].

Insbesondere steht die Qualität des Zielproduktes im Fokus von PAT. Daher sollen vor allem kritische Qualitäts- und Leistungskenngrößen von Rohmaterialien, Zwischen- und Endprodukten sowie der ablaufenden Prozesse selbst analytisch erfasst werden [FDA, 2004a].

Offline-Analysen wie Bioassays sind bei der Überwachung von Bioprozessen hinsichtlich des Zielproduktes heutzutage häufig noch alternativlos. Werden diese Analysen jedoch zeit- und prozessnah durchgeführt und dienen der Erweiterung des Prozessverständnisses, so sind solche Methoden durchaus auch den PAT-Werkzeugen zuzuordnen [STREEFLAND et al., 2013].

Wie in Bild 4.1 gezeigt, umfassen PAT-Analysatoren von Bioprozessen demnach Offline- und Atline-Verfahren, modellgestützte mathematische Methoden sowie fortschrittliche Inline-Messsysteme. Hierzu gehört zum Beispiel eine Reihe von optischen Spektroskopieverfahren.

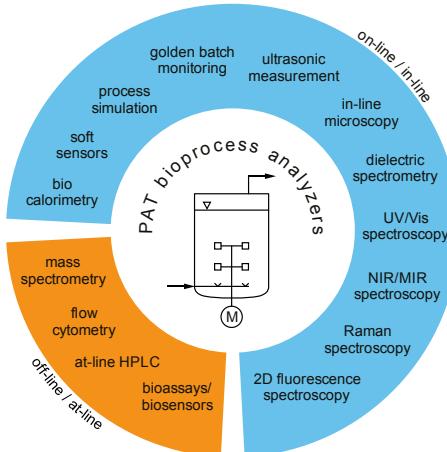


Bild 4.1: Eine graphische Übersicht über mögliche PAT-Analysatoren

Diese sind in der Bioprozesstechnik von besonderem Interesse. Sie haben den Vorteil, schnell, zerstörungsfrei und nichtinvasiv zu sein. Weiterhin können mit spektroskopischen Verfahren eine Vielzahl von physikalischen und chemischen Größen simultan gemessen werden [SKIBSTED & ENGELSEN, 2010].

Technische Fortschritte, beispielsweise im Bereich der Lichtleitertechnik, haben zu der Entwicklung von robusten und kostengünstigen Spektrometersystemen geführt, die flexibel in Herstellungsprozessen eingesetzt werden können [LOURENÇO et al., 2012]. Insbesondere die Nahinfrarotspektroskopie wurde bereits vielfach erfolgreich auch in Bioprozessen erprobt [HOEHSE et al., 2015].

Aufgrund des komplexen Charakters und der Vielzahl verschiedener Bioprozesse mit unterschiedlichen zu beobachtenden Qualitäts- und Leistungskenngrößen besteht allerdings noch ein großer Forschungsbedarf. Diese Arbeit untersucht daher die Eignung verschiedener spektroskopischer Verfahren hinsichtlich ihrer Anwendung an Bioprozessen.

Dabei wurden die drei Methoden Nahinfrarot, Raman und 2D-Fluoreszenz eingesetzt und miteinander verglichen. Diese Messverfahren werden in den folgenden Abschnitten näher beschrieben.

4.2 Eingesetzte Spektroskopieverfahren

4.2.1 Nahinfrarotspektroskopie

Der infrarote Spektralbereich wird im elektromagnetischen Spektrum vom sichtbaren Licht und den Mikrowellen begrenzt. Er ist aufgeteilt in das nahe Infrarot (NIR, ca. 760 bis 3.000 nm), das mittlere Infrarot (MIR, ca. 3.000 nm bis 30 µm) und das ferne Infrarot (FIR, ca. 30 µm bis 1.000 µm) [LOTTSPEICH & ENGELS, 2012].

Die IR-Spektroskopie basiert auf der Absorption von Infrarotstrahlung durch Materie. Dabei werden Moleküle zu Schwingungen angeregt. Die dafür notwendige Energie bzw. die Wellenlänge der anregenden Strahlung ist abhängig von den jeweils schwingenden Massen und den entsprechenden Bindungsstärken. Deshalb sind IR-Absorptionsbanden charakteristisch für bestimmte funktionelle Gruppen und diese Methode somit für Analysezwecke geeignet.

Voraussetzung für die Absorption von Infrarotstrahlung ist eine Änderung des Dipolmoments des Gesamt moleküls im Verlauf einer Schwingung. Aufgrund dessen können allein heteronukleare Moleküle detektiert werden. Homonukleare Verbindungen, die kein Dipolmoment besitzen, absorbieren keine Infrarotstrahlung. In Bild 4.2 sind mögliche Schwingungen in Wasser- und Kohlenstoffdioxidmolekülen gezeigt.

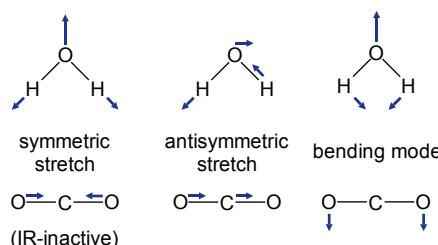


Bild 4.2: Mögliche Schwingungsmodi von H₂O- und CO₂-Molekülen

Aufgrund des Dipolcharakters können neben den sogenannten Fundamentalschwingungen auch Oberton- und Kombinationsschwingungen auftreten, für deren Anregung jedoch höhere Energiemengen notwendig sind. Weiterhin sind die Anregungswahrscheinlichkeiten geringer, sodass die Banden der Obertöne weniger stark ausgeprägt sind. In Bild 4.3 sind Molekülschwingungen anhand des Potentialfeldes eines anharmonischen Oszillators veranschaulicht.

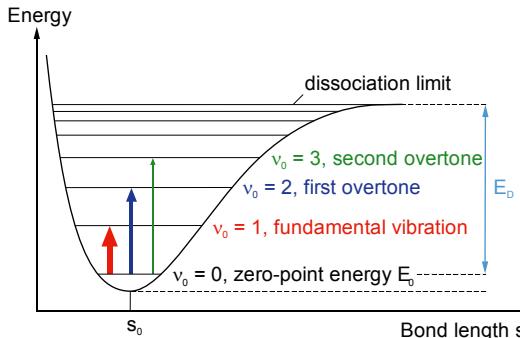


Bild 4.3: Schwingungsniveaus im Potentialfeld eines anharmonischen Oszillators

Die Nullpunktenergie E_0 ist ungleich dem Energieminimum, da Moleküle selbst bei 0 K noch Schwingungen um ihre Ruhelage ausführen. Die Dissoziationsenergie E_D führt zu einem Zerfall der Verbindung in ihre Radikale [SKRABAL, 2009]. Zwischen der Nullpunktenergie und der Dissoziationsgrenze können die Verbindungen bestimmte Schwingungsniveaus einnehmen. Die Höhe der Pfeile zeigt die zur Anregung notwendige Energie an. Die Breite der Pfeile steht für die Stärke der Absorptionsbanden im Spektrum.

Die höchsten Dipolmomente und damit die höchsten Intensitäten weisen Moleküle mit Wasserstoffatomen auf. Diese Verbindungen absorbieren Strahlung im MIR-Bereich. Daher sind die Banden ihrer Oberton- und Kombinationsschwingungen häufig im NIR-Bereich zu finden. Das NIR-Spektrum wird demnach von CH-, OH-, SH- und NH-Gruppen dominiert und eignet sich folglich gut zur Analyse von organischen Substanzen [SIESLER, 2002].

Der Vorteil der NIR-Spektroskopie gegenüber der MIR-Spektroskopie besteht in den verhältnismäßig schwachen Absorptionsbanden. Daher kann auf eine aufwendige Proben vorbereitung verzichtet werden [SIESLER, 2002].

Die Nachteile der NIR-Spektroskopie basieren auf der Vielzahl an Übergangsmöglichkeiten, die zu einer hohen Redundanz an Informationen in einem Spektrum führen. Folglich kommt es zur Überlagerung der ohnehin schon breit ausgeprägten Absorptionsbanden, was eine Auswertung des Spektrums erschwert [MARTIN, 1992].

Wasser ist zudem ein starker Absorber, dessen erster Oberton in etwa bei 1.450 nm auftritt und häufig die Signale anderer Analyten überlagert. Dieser Umstand erschwert die Analyse biotechnologischer Prozesse. Hinzu kommt eine hohe Temperaturabhängigkeit der NIR-Spektroskopie, was möglichst äquivalente Versuchsbedingungen erfordert.

In dieser Arbeit wurde das Prozessspektrometersystem PSS 1720 der Fa. Polytec, Waldbronn, eingesetzt. Dieses enthält eine Wolfram-Halogenlampe als Lichtquelle sowie eine Detektoreinheit, die wiederum aus einem Polychromator (Prisma) und aus einen Diodenzeilendetektor besteht. Durch die Abwesenheit beweglicher Teile in der Detektoreinheit wird im industriellen Produktionsumfeld eine hohe Robustheit erzielt. Die gleichzeitige Abbildung aller Wellenlängen auf dem Detektor ermöglicht außerdem sehr kurze Messzeiten.

Über SMA-Lichtleiteranschlüsse können unterschiedliche Sondentypen mit dem Gerät verbunden werden. Bild 4.4 zeigt das Spektrometersystem sowie die beiden verwendeten Sondentypen.



Bild 4.4: Prozessspektrometersystem der Fa. Polytec, a) Grundgerät PSS 1720, b) Transmissionssonde PSS-H-CT1, c) Reflexionssonde PSS-H-CR1

Die Messung erfolgt abhängig von der Probenbeschaffenheit, entweder in Transmission oder diffuser Reflexion. Bei der Transmissionsmessung wird die Strahlung durch eine definierte Schichtdicke der Probe zum Detektor geleitet. Dies erlaubt bei der Auswertung die Heranziehung des Lambert-Beer'schen Gesetzes. Die Transmissionsmessung ist nur für flüssige und transparente Proben geeignet.

Für intransparente, pulverförmige oder pastöse Proben sowie Suspensionen wird in diffuser Reflexion gemessen. Hierbei wird Licht in die Probe eingestrahlt und die von den Probenpartikeln reflektierte Strahlung detektiert. Aus dieser werden die Absorptionseigenschaften der Probe extrahiert. In Bild 4.5 sind die beiden Messverfahren illustriert.

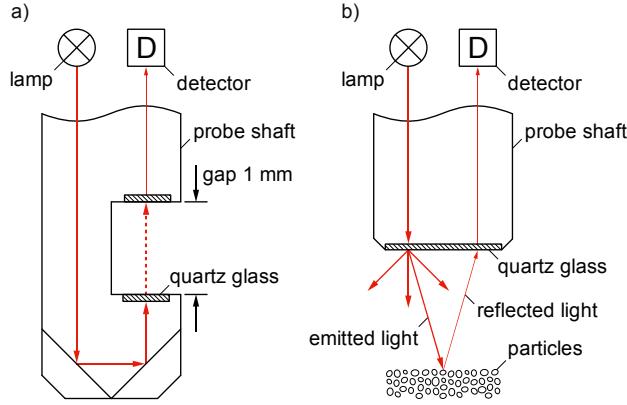


Bild 4.5: Messverfahren bei der NIR-Spektroskopie, a) Transmissionsmessung, b) Messung in diffuser Reflexion

Ein Nachteil der Messung in diffuser Reflexion besteht in der Abhängigkeit der Reflexions-eigenschaften von der Partikelgröße sowie von der Anregungswellenlänge des eingestrahlten Lichts. Eine direkte chemische Interpretation der Messergebnisse wird dadurch erschwert.

Die Aufzeichnung der Spektren erfolgte in einem Wellenlängenbereich von 850 nm bis 1.650 nm. Es wurde eine Schrittweite von 2 nm gewählt, wodurch die verwendeten NIR-Spektren aus 401 Einzelmessungen bestehen.

4.2.2 Raman-Spektroskopie

Wie bei der NIR-Spektroskopie kommt es bei der Raman-Spektroskopie zur Anregung von Molekülschwingungen. Allerdings beruht die Raman-Spektroskopie nicht auf der Absorption sondern vielmehr auf der Streuung von Strahlung.

In dem an einem Molekül gestreuten Licht können neben den Linien der Erregerstrahlung auch Begleitlinien mit geringeren Intensitäten beobachtet werden. Diese Linien weisen eine Verschiebung der Wellenlänge gegenüber der Erregerstrahlung auf und werden als Raman-Linien bezeichnet [SKRABAL, 2009]. In Bild 4.6 ist das Auftreten des Raman-Effektes illustriert.

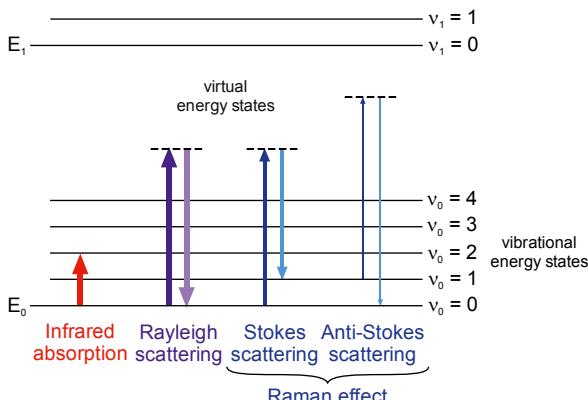


Bild 4.6: Erklärungsmodell zum Auftreten des Raman-Effektes

Das gängige Modell zur Erklärung des Raman-Effektes involviert virtuelle Energiezustände, in die ein Molekül durch Bestrahlung mit monochromatischem Licht gehoben werden kann. Diese instabilen Energieniveaus liegen zwischen dem Elektronengrundzustand E_0 eines Moleküls mit den möglichen Schwingungsniveaus v_0 und einem angeregten Elektronenzustand E_1 .

Beim Verlassen eines dieser Zustände emittiert das angeregte Molekül ein Photon. Wird dabei wieder der Grundzustand E_0 erreicht, weist die emittierte Strahlung die gleiche Energie auf wie die Erregerstrahlung. In diesem Fall spricht man von Rayleigh-Streuung. Diese enthält keine chemische Information über die vorliegende Substanz.

Fällt das Molekül jedoch auf einen Schwingungszustand oberhalb des Grundzustandes zurück, besitzt das emittierte Photon eine niedrigere Energie als die Erregerstrahlung. Die resultierende Streuung wird als Stokes-Streuung bezeichnet. Ebenso kann sich ein Molekül bereits

in einem angeregten Schwingungszustand befinden und nach weiterer Anregung wieder seinen Grundzustand einnehmen. In diesem Fall, der Anti-Stokes-Streuung hat das emittierte Photon eine höhere Energie als die Erregerstrahlung. Stokes- und Anti-Stokes-Streuung bilden zusammen den Raman-Effekt [VANDENABEELE, 2013].

Da sich die auftretenden Energiedifferenzen bestimmten Schwingungsniveaus zuordnen lassen, enthält der Raman-Effekt chemische Information über das angeregte Molekül. Eine Voraussetzung für das Auftreten von Molekülschwingungen in der Raman-Spektroskopie ist eine periodische Polarisierbarkeit der Moleküle.

Die Polarisierbarkeit beschreibt das Maß, in dem sich Elektronen innerhalb eines Moleküls durch Anlegen eines elektrischen Feldes verschieben lassen. Dieses wird hier durch die elektromagnetische Erregerstrahlung erzeugt. Eine Verschiebung der Elektronen in einem Molekül führt zu einer Deformation der Elektronenbahnen, was sich wiederum auf die Atomabstände und Bindungswinkel auswirkt und damit Molekülschwingungen induziert.

Insbesondere homonukleare Gruppen wie C-C, C=C, oder S-S weisen eine hohe Polarisierbarkeit und damit intensive Raman-Banden auf. Damit bildet die Raman-Spektroskopie den komplementären Teil zur IR-Spektroskopie, bei der ausschließlich heteronukleare Gruppen schwingen. Durch die Wahl der Erregerwellenlänge ist eine Ausrichtung auf die Untersuchung bestimmter funktioneller Gruppen möglich [PERKAMPUS, 1993].

In Raman-Spektren wird die Intensität der Raman-Banden gegen den sogenannten Raman-Shift des Streulichtes in Wellenzahlen ν [cm^{-1}] aufgetragen. Der Raman-Shift ist die Wellenzahldifferenz zur Erregerlinie. In der Regel wird sich auf die Auswertung der Stokes'schen Banden beschränkt, da die Anti-Stokes'schen Banden wesentlich schwächer ausgeprägt sind. Stokes'sche Banden weisen niedrigere Wellenzahlen als die Erregerlinie auf. Deshalb befinden sie sich im Raman-Spektrum links von dieser. Der Raman-Shift wird folglich in absteigender Reihenfolge dargestellt [BÖCKER, 1997].

Der Vorteil der Raman-Spektroskopie insbesondere im Vergleich mit der NIR-Spektroskopie besteht darin, dass die Spektren ein definiteres Bandenmuster aufweisen und somit chemisch spezifischere Analysen erlauben. Weiterhin ist Wasser nicht Raman-aktiv, was eine Auswertung von Spektren aus Bioprozessen vereinfacht.

Die Nachteile der Raman-Spektroskopie erwachsen aus den schwachen Banden des Raman-Effektes. So weist die Raman-Spektroskopie eine vergleichsweise niedrige Sensitivität auf. Weiterhin werden Erregerstrahlungen mit hoher Intensität benötigt, um überhaupt Raman-Streuung nachweisen zu können. Diese Strahlungen werden mit Lasern erzeugt, für die der Aspekt der Arbeitssicherheit zu berücksichtigen ist. Darüber hinaus kann die intensive Laserstrahlung insbesondere organische Proben zur Fluoreszenz anregen, was die Auswertung der Spektren erheblich stört.

In dieser Arbeit kamen die Geräte RamanRxn1™ und RamanRxn2™ der Fa. Kaiser Optical Systems, Ann Arbor, USA, zum Einsatz. Die Geräte unterscheiden sich hauptsächlich dahingehend, dass an ersterem nur eine Sonde zurzeit betrieben werden kann, wohingegen es sich bei dem in Bild 4.7 gezeigten RamanRxn2™ um ein Mehrkanalspektrometer handelt, an dem über einen sogenannten Multiplexer bis zu vier Sonden gleichzeitig betrieben werden können.



Bild 4.7: RamanRxn2™ Mehrkanalspektrometer der Fa. Kaiser Optical Systems

In beiden verwendeten Spektrometersystemen kamen NIR-Diodenlaser der Laserklasse 3B mit einer Anregungswellenlänge λ_{ex} von 785 nm zum Einsatz. Die maximale Laserleistung beträgt 400 mW. Die Messungen wurden mit über Lichtleiter verbundenen bIO-LAB Tauchsonden (Kaiser) durchgeführt.

Die Optik der Sonde fokussiert das parallele Laserlicht ca. 7 mm vor der Sondenspitze. Ähnlich zur NIR-Messtechnik der diffusen Reflexion wird die erzeugte Streustrahlung von derselben Optik aufgenommen und zum Detektor geleitet. Vor dem Detektor wird die intensive Rayleigh-Streustrahlung herausgefiltert.

Die Aufzeichnung der Spektren erfolgte mit einem Charge-Coupled-Device (CCD) Detektor im Wellenzahlbereich von 200 cm^{-1} bis 3.200 cm^{-1} . Bei einem gewählten Inkrement von 1 cm^{-1} ergeben sich daraus 3.001 Einzelmessungen für ein Raman-Spektrum.

4.2.3 2D-Fluoreszenzspektroskopie

Bei der Absorption von Strahlung können Moleküle in einen angeregten Elektronenzustand E_1 gehoben werden. Beim Übergang des Moleküls in seinen Ursprungszustand E_0 kommt es zu einer Abgabe der zuvor aufgenommenen Energie. Erfolgt dies in Form ungerichteter elektromagnetischer Strahlung, wird dieser in Bild 4.8 illustrierte Vorgang Fluoreszenz genannt. In der Fluoreszenzspektroskopie wird diese Strahlung, anstelle des Intensitätsverlustes der Erregerstrahlung (Absorption), messtechnisch erfasst [LAKOWICZ, 2006].

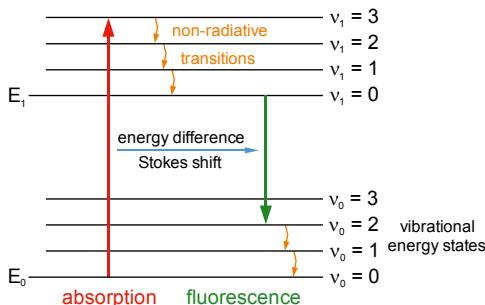


Bild 4.8: Jablonski Diagramm zur Verdeutlichung der Fluoreszenz

Durch die Absorption von Strahlung im UV/Vis-Bereich werden Moleküle üblicherweise auf einen Schwingungszustand ($v_1 > 0$) des angeregten Elektronenzustandes E_1 angehoben. In sehr kurzer Zeit (10^{-12} s) findet anschließend über intramolekulare Vibrations- und Rotationsvorgänge eine strahlungslose Deaktivierung statt. Dabei verliert das Molekül Schwingungsenergie bis zum Erreichen des niedrigsten Niveaus ($v_1 = 0$) von E_1 [LAKOWICZ, 2006].

Die eigentliche Fluoreszenzemission erfolgt dann in der Form von meist im visuellen Bereich liegender Strahlung. Auch im Elektronengrundzustand E_0 nimmt das Molekül zunächst ein höheres Schwingungsniveau ($v_0 > 0$) ein. Eine weitere Entspannung auf den Grundzustand ($v_0 = 0$) erfolgt wieder in sehr kurzer Zeit. Durch die strahlungsfreien inneren Übergänge ist Fluoreszenzstrahlung üblicherweise energieärmer und dadurch langwelliger als die Erregerstrahlung. Diese Energiedifferenz wird auch als Stokes-Shift bezeichnet [LAKOWICZ, 2006].

Typische Fluorophore sind aromatische Moleküle mit konjugierten Ringstrukturen. Andere organische Substanzen fluoreszieren in der Regel nicht. Im Bereich der Bioanalytik erlaubt die Fluoreszenzspektroskopie einen Nachweis von Proteinen, bei denen eine Fluoreszenzemission insbesondere durch die aromatischen Reste der Aminosäuren Tryptophan, Tyrosin und Phenylalanin hervorgerufen wird. Weiterhin sind auch die Co-Enzyme NADH und FAD fluoreszierend [LAKOWICZ, 2006].

Eine weit verbreitete und die in dieser Arbeit angewandte Methode ist die 2D-Fluoreszenzspektroskopie. Hierbei werden Emissionsspektren für verschiedene Anregungswellenlängen aufgenommen. Dieses Vorgehen hat den Vorteil, dass mehr Analyten gleichzeitig detektiert werden können. Ein weiterer Vorteil der Fluoreszenzspektroskopie ist eine im Allgemeinen hohe Sensitivität.

Als Nachteil kann dagegen aufgeführt werden, dass viele Substanzen nicht fluoreszieren. Des Weiteren kann es innerhalb einer komplexen Messmatrix zu Wechselwirkungen kommen, die eine Interpretation und Auswertung der Fluoreszenzspektren erschweren. Ein Beispiel dafür sind sogenannte Quenching-Effekte, bei denen funktionelle Gruppen in der Nähe eines Fluorophors die emittierte Strahlung absorbieren.

In dieser Arbeit wurde das in Bild 4.9 dargestellte 2D-Fluoreszenzspektrometer BioView® der Fa. DELTA Light & Optics, Lyngby, Dänemark, verwendet.

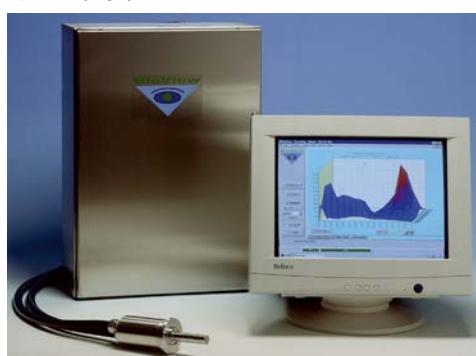


Bild 4.9: 2D-Fluoreszenzspektrometer von DELTA Light & Optics

Als Lichtquelle des verwendeten Spektrometers dient eine Xenon-Blitzlichtlampe. Die entsprechenden Kombinationen aus Anregungs- und Emissionswellenlängen werden durch zwei unabhängige Filterräder realisiert. Als Detektor kommt ein Photomultiplier zum Einsatz. Lichtquelle, Detektor, Filterräder sowie die zugehörige Mechanik und Elektronik sind in einem spritzwassergeschützten Edelstahlgehäuse untergebracht.

Die Fluoreszenzanregung erfolgte im Wellenlängenbereich λ_{ex} von 270 nm bis 550 nm mit einer Schrittweite von 20 nm. Die Fluoreszenzemission wurde mit gleichem Inkrement im Bereich λ_{em} von 310 nm bis 590 nm aufgezeichnet, wobei aufgrund des Stokes-Shift nur solche Emissionswellenlängen λ_{em} berücksichtigt werden, die größer sind als die jeweilige Anregungswellenlänge. Daraus resultieren 150 Emissionswerte pro aufgenommenem 2D-Fluoreszenzspektrum.

4.3 Extraktion relevanter Informationen aus Spektren

Wie bereits erwähnt, können Spektren eine Vielzahl an chemischen und physikalischen Informationen der vermessenen Probe enthalten. Allerdings liegen diese häufig redundant vor und können sich überlagern. Darüber hinaus ist der Informationsgehalt im Verhältnis zur Datenmenge eher gering.

Für eine Auswertung der Spektren sind demnach Methoden der Multivariaten Datenanalyse (MVDA) unumgänglich und werden von der FDA als eigenständiges Werkzeug für die Umsetzung von PAT angesehen [FDA, 2004a; SMALL, 2006; SKIBSTED & ENGELSEN, 2010; MERCIER et al., 2014].

Für die Anwendung auf spektrale Daten werden vor allem dimensionsreduzierende Verfahren wie die Hauptkomponentenanalyse (PCA) verwendet. Dies wird insbesondere auf die Einfachheit und Effektivität dieser Verfahren zurückgeführt [RATHORE et al., 2011].

Weiterhin erlauben diese Methoden, anhand vielfältiger statistischer und grafischer Werkzeuge, eine einfache Interpretation der Daten und tragen somit zur Erweiterung des Prozessverständnisses bei [LOURENÇO et al., 2012]. Eine Einführung in die Multivariate Datenanalyse ist im folgenden Kapitel 5 gegeben.

5. Einführung in die Multivariate Datenanalyse (MVDA)

5.1 Zielsetzung bei der Anwendung der MVDA

Das allgemeine Ziel der Multivariaten Datenanalyse (MVDA) ist die Verarbeitung großer Datenmengen. Dabei soll eine Interpretation der Daten vereinfacht und relevante Informationen extrahiert werden. Eine grundlegende Technik der MVDA ist die sogenannte Hauptkomponentenanalyse (PCA), die in Abschnitt 5.3 näher erläutert wird.

Ein erstes Ziel dieser Arbeit war die Qualitätsbewertung von Hefeextrakten mittels NIR-Spektroskopie. Grundlage derartiger Methoden sind multivariate Klassifizierungsverfahren, die Variationen von Spektren derart herausarbeiten, dass eine Unterscheidung verschiedener Klassen ermöglicht wird. Das Konzept ist in Bild 5.1 illustriert.

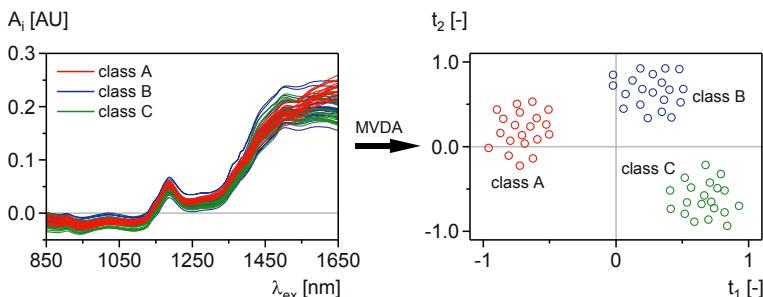


Bild 5.1: Klassifizierung von Hefeextrakten mit MVDA basierend auf NIR-Spektren

Basierend auf dem Resultat der Klassifizierung kann eine Qualitätsbewertung erfolgen. Weit verbreitet ist das auf der PCA beruhende Verfahren *Soft Independent Modelling of Class Analogy* (SIMCA), das in Abschnitt 5.6.2 vorgestellt wird.

Der überwiegende Teil dieser Arbeit widmet sich der Bestimmung von Bioprozessgrößen aus Spektren. Das Ziel dabei ist die Online-Beobachtung schwer messbarer Variablen zur Erweiterung des Bioprozessmonitors. Bild 5.2 zeigt diesen Ansatz.

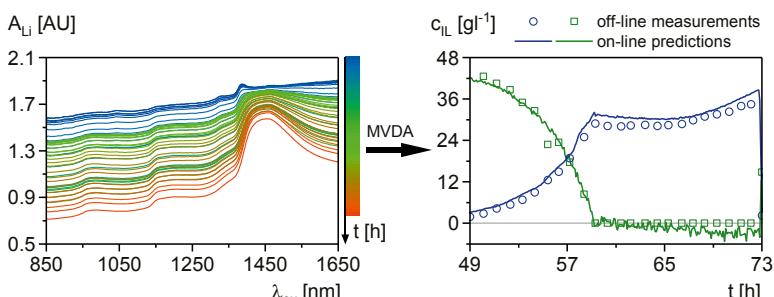


Bild 5.2: Prädiktion von Bioprozessgrößen mit MVDA aus Spektren

Die Abschätzung von Prozessvariablen wird in der MVDA als Vorhersage bezeichnet. Hierfür sind multivariate Kalibrierverfahren erforderlich, die eine Beziehung zwischen den chemischen Informationen in den Spektren und den Referenzmessungen der entsprechenden Prozessgrößen herstellen. Die dabei eingesetzte Methode *Partial Least Squares Regression* (PLSR) ist eine Erweiterung der PCA und kann lineare Zusammenhänge zwischen Spektren und Analyten modellieren. Die PLSR wird in Abschnitt 5.4 erläutert.

Neben der weit verbreiteten PLSR wurde in dieser Arbeit auch die Methode *Support Vector Regression* (SVR) untersucht. Die SVR ist in der Lage, auch nichtlineare Beziehungen abzubilden. Diese Methode ist in Abschnitt 5.7 zu finden.

5.2 Datenvorbereitung

5.2.1 Struktur und Modifizierung multivariater Datensätze

Vor Anwendung der MVDA werden die gemessenen Daten in einer Matrix D ($n \times m$) abgelegt. Diese besteht aus $i = 1, 2, \dots, n$ Zeilen (Beobachtungen) und $j = 1, 2, \dots, m$ Spalten (Variablen). Ein Datensatz von beispielsweise 50 NIR-Spektren hat folglich $n = 50$ Zeilen und $m = 401$ Spalten.

Die Modifizierung der Rohdaten mit einer geeigneten Vorverarbeitungsmethode ist häufig zweckmäßig und kann ausschlaggebend für den Erfolg oder Misserfolg der MVDA sein. So können beispielsweise Störungen aus den Daten entfernt werden, die eine multivariate Modellbildung erschweren. Auch die Anwendung mehrerer Vorverarbeitungsmethoden auf denselben Datensatz ist eine übliche Vorgehensweise.

Durch eine Modifizierung werden die Daten in eine Matrix X ($n \times m$) überführt. Diese hat die gleichen Dimensionen wie die nicht modifizierte Datenmatrix D und stellt den Eingangsdatensatz für die eigentliche Anwendung der MVDA dar.

Einige übliche und in dieser Arbeit angewandte Vorbehandlungen werden in den folgenden beiden Abschnitten vorgestellt.

5.2.2 Zentrierung und Skalierung

Eine der einfachsten Vorverarbeitungsmethoden ist die spaltenweise Zentrierung oder Mittenzentrierung (*mean centering mc*).

Zur Berechnung eines mittenzentrierten Messwertes x_{ijmc} ,

$$x_{ijmc} = d_{ij} - \bar{d}_j, \quad (5.1)$$

wird von dem Messwert d_{ij} der jeweilige Spaltenmittelwert \bar{d}_j ,

$$\bar{d}_j = \frac{1}{n} \cdot \sum_{i=1}^n d_{ij} \quad (5.2)$$

mit

$n :=$ Anzahl der Messwerte in Spalte j,

subtrahiert.

Die Mittenzentrierung kann zu einer Reduktion der Modellkomplexität beitragen und verbessert insbesondere die Interpretierbarkeit multivariater Modelle, die auf der PCA beruhen [KESSLER, 2008]. Sie wird zumeist für spektrale Daten eingesetzt.

Wenn die Variablen der Datenmatrix D aus unterschiedlichen physikalischen Abläufen stammen, weisen diese differierende numerische Größen und damit auch Varianzen auf. Solche führen zu einer ungewollten und ungleichen Wichtung der Variablen bei der Bildung eines multivariaten Modells.

Aus diesem Grund handelt es sich bei der sogenannten Autoskalierung um eine weit verbreitete Vorbehandlungsmethode. Sie bewirkt eine Standardisierung der Daten auf eine einheitliche Varianz von 1 (*unit variance* uv).

Zur Berechnung eines autoskalierten Messwertes x_{ijuv} ,

$$x_{ijuv} = \frac{d_{ij} - \bar{d}_j}{s_{dj}}, \quad (5.3)$$

wird der mittenzentrierte Messwert durch die Spaltenstandardabweichung s_{dj} ,

$$s_{dj} = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^n (d_{ij} - \bar{d}_j)^2}, \quad (5.4)$$

dividiert.

Im Gegensatz zur Mittenzentrierung wird die Autoskalierung nicht für Spektren verwendet, da dort die hohe Varianz einer Variablen der gesuchten chemischen Information entspricht. Eine Autoskalierung irrelevanter Variablen mit niedriger Varianz ist demnach nicht zweckmäßig und führt lediglich zu einer Verstärkung des Rauschens. Spezielle Vorverarbeitungsmethoden für spektrale Daten werden im nächsten Abschnitt behandelt.

5.2.3 Datenvorverarbeitung für Spektren

Spektren können Störungen enthalten, die durch bestimmte mathematische Datenfilter entfernt beziehungsweise korrigiert werden können.

Eine häufig verwendete Methode für die Behandlung spektraler Daten ist das *Standard Normal Variate* (SNV) Filter.

Ein SNV-gefilterter Messwert x_{ijsnv} ,

$$x_{ijsnv} = \frac{d_{ij} - \bar{d}_i}{s_{di}}, \quad (5.5)$$

wird durch Subtraktion des Zeilenmittelwertes \bar{d}_i ,

$$\bar{d}_i = \frac{1}{m} \cdot \sum_{j=1}^m d_{ij} \quad (5.6)$$

mit

$m :=$ Anzahl der Messwerte in einer Zeile i,

vom Messwert d_{ij} und anschließender Division durch die Zeilenstandardabweichung s_{di} ,

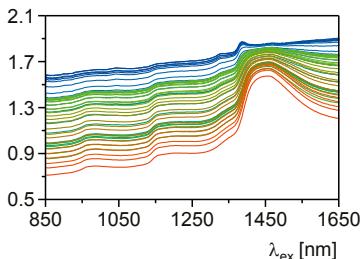
$$s_{di} = \sqrt{\frac{1}{m-1} \cdot \sum_{j=1}^m (d_{ij} - \bar{d}_i)^2}, \quad (5.7)$$

berechnet.

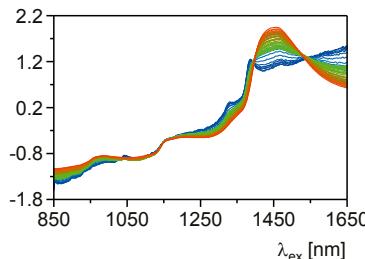
Das SNV-Filter ist eine Zeilenoperation und nicht mit der Autoskalierung zu verwechseln, die basierend auf dem Spaltenvektor d_j berechnet wird.

Die SNV-Filterung führt zu einer Standardisierung der Spektren und korrigiert Basislinienverschiebungen sowie wellenlängenabhängige Streueffekte in Spektren partikelbehafteter Proben [KESSLER, 2008].

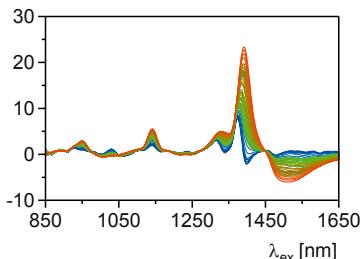
a) $A_{Li} = d_i [\text{AU}]$



b) $A_{Lisnv} = x_{isnv} [-]$



c) $A_{Lie1} = x_{ide1} [\text{mAU nm}^{-1}]$



d) $A_{Lie2} = x_{ide2} [\text{mAU nm}^{-2}]$

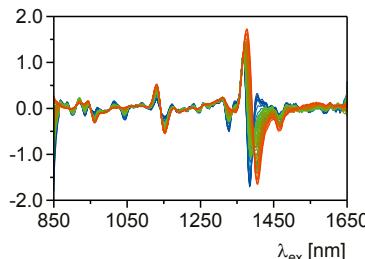


Bild 5.3: Vorverarbeitung von NIR-Spektren, a) unverändert, b) SNV-gefiltert, c) Modifizierung durch erste bzw. d) durch zweite Ableitung

Bild 5.3a zeigt NIR-Spektren der Flüssigphase eines Bioreaktionsprozesses. In Bild 5.3b sind dagegen SNV-gefilterte Spektren dargestellt. Eine weitere häufig anzutreffende Vorverarbeitungsmethode ist die Ableitung von Spektren. Dies ist in Bild 5.3c und Bild 5.3d ebenfalls am Beispiel der NIR-Spektroskopie gezeigt.

Die Ableitung stellt eine sehr effektive Methode zur Basislinienkorrektur dar und verbessert außerdem die spektrale Auflösung. Ein Nachteil ist die schlechtere chemische Interpretierbarkeit, da sich das Erscheinungsbild der Spektren stark verändert [NAES, 2004].

Ein durch Bildung der ersten Ableitung modifiziertes Element x_{ijde1} ,

$$x_{ijde1} = \frac{\partial d_{ij}}{\partial \lambda_{exj}} \approx \frac{d_{ij+1} - d_{ij-1}}{\lambda_{exj+1} - \lambda_{exj-1}}, \quad (5.8)$$

entspricht z.B. bei der NIR-Spektroskopie der Absorptionssteigung bei der Anregungswellenlänge λ_{exj} .

Die numerische Differenzierung von Spektren wird in der Regel mit dem Verfahren von Savitzky und Golay durchgeführt. Dieses beruht auf der Anpassung eines fortlaufenden Datenfensters an ein Polynom nach der Methode der kleinsten Fehlerquadrate. Durch Differenzierung des Polynoms stehen dann auch die Ableitungen des betrachteten Signals zur Verfügung [SAVITZKY & GOLAY, 1964]. Die Größe des Datenfensters bestimmt dabei das Maß der Signallägung, die mit dieser Methode automatisch erzielt wird.

5.3 Die Hauptkomponentenanalyse (PCA)

5.3.1 Dimensionsreduktion durch Hauptkomponenten

Die Hauptkomponentenanalyse (*Principle Component Analysis PCA*) ist eine zentrale Methode der MVDA. Viele weitere Verfahren basieren auf der PCA. Daher wird deren Funktionsweise an dieser Stelle näher erläutert.

Das Ziel der PCA besteht darin, eine Reduktion der Dimension der Ursprungsdaten zu erreichen, um eine Analyse und Interpretation dieser Daten zu vereinfachen. Dazu werden im Prinzip Variablen des betrachteten multivariaten Datensatzes mit ähnlichem Informationsgehalt in sogenannte Hauptkomponenten zusammengefasst. Diese werden daher auch latente Variable genannt.

Zur Verdeutlichung wird im Folgenden ein Datensatz mit einer Variablenanzahl m von 3 betrachtet. Die Variablen x_1 , x_2 und x_3 sind die Spalten der modifizierten Datenmatrix X und bilden ein Koordinatensystem mit der Dimension m. Jede Beobachtung (Zeile) des Datensatzes kann innerhalb dieses in Bild 5.4a gezeigten dreidimensionalen Datenraums als ein Punkt dargestellt werden.

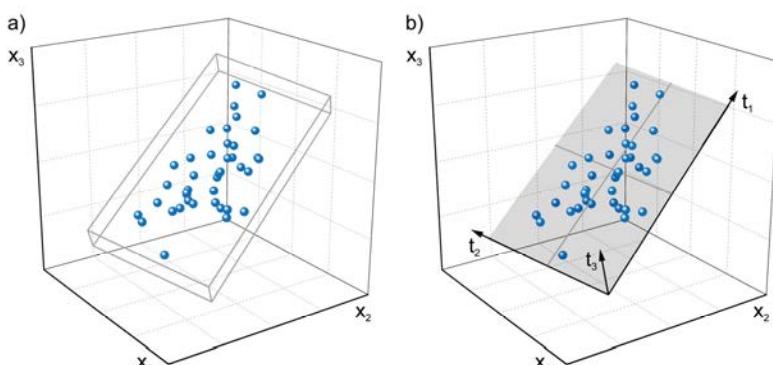


Bild 5.4: a) Anordnung von Beobachtungen im dreidimensionalen Datenraum X,
b) Beschreibung der Daten durch die Bildung von Hauptkomponenten

Die Beobachtungen weisen dort eine bestimmte räumliche Anordnung auf, die durch einen Quader angedeutet ist. Werden die Daten entlang der größten Varianz innerhalb dieser Anordnung durch eine Gerade beschrieben, so nennt man diese Gerade die erste Hauptkomponente t_1 . Die zweite Hauptkomponente t_2 erhält man durch eine weitere Gerade, die orthogonal zur ersten verläuft und erneut ein Maximum der Varianz der Daten in dieser Richtung beinhaltet.

Ist die Restvarianz der Daten vernachlässigbar klein, kann man auf die Berechnung der dritten Hauptkomponente t_3 verzichten, sodass die gewünschte Dimensionsreduktion erreicht wird. Bild 5.4b deutet die Beschreibung der Daten durch die beiden Hauptkomponenten t_1 und t_2 an, die eine Fläche im dreidimensionalen Raum bilden.

5.3.2 Das mathematische Modell der PCA

In der allgemeinen Form des PCA-Modells wird die Datenmatrix X,

$$X = T \cdot P^T + E \quad (5.9)$$

mit

$X := (n \times m)$ modifizierte Datenmatrix

$n :=$ Anzahl an Beobachtungen bzw. Proben (Zeilen) in X

$m :=$ Anzahl an Variablen (Spalten) in X

$T := (n \times r)$ Scorematrix

$r :=$ Anzahl gebildeter Hauptkomponenten (Spalten in T)

$P^T := (r \times m)$ transponierte Loadingmatrix

$E := (n \times m)$ Residuenmatrix,

in eine Scorematrix T, eine Loadingmatrix P und eine Residuenmatrix E zerlegt.

Die Residuenmatrix E hat die gleichen Dimensionen wie die Datenmatrix X und enthält den Teil der Varianz in den Daten aus X, der nicht durch die gebildeten Hauptkomponenten beschrieben wird.

Die Scorematrix T besteht aus n Zeilen und $l = 1, 2, \dots, r$ Spalten. Jeder berechneten Hauptkomponente wird damit ein Wert (Score) für jede Beobachtung zugewiesen. Dadurch werden die Beobachtungen in einem neuen Koordinatensystem abgebildet. In Analogie zu den Spaltenvektoren (Variablen) x_j wird der Begriff Hauptkomponente (latente Variable) häufig für die Spaltenvektoren t_i verwendet.

Die Loadingmatrix P besteht dagegen aus m Zeilen und r Spalten. Hier ist jeder Hauptkomponente ein Wert (Loading) für die Variablen der Datenmatrix X zugewiesen. Die Spaltenvektoren p_i sind die Richtungsvektoren der Hauptkomponenten im Originaldatenraum X. Die Loadings beschreiben somit die Orientierung der Hauptkomponenten in diesem Datenraum.

Bild 5.5 illustriert die Bedeutung der Scores und Loadings. Es ist eine Anzahl n von 3 Beobachtungen sowie eine Hauptkomponente t_1 in einem Raum gezeichnet, der aus einer Anzahl m von 2 Variablen x_j besteht. Durch eine Untersuchung der geometrischen Struktur erkennt man, dass ein Loading p_{j1} dem Cosinus des Winkels zwischen der gewählten Variablen x_j und der Hauptkomponenten t_1 entspricht.

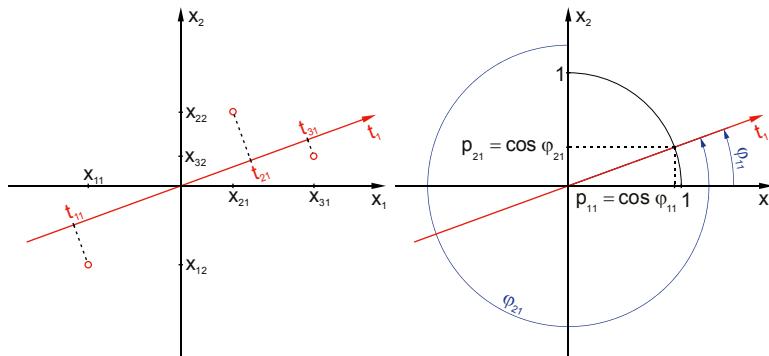


Bild 5.5: Verdeutlichung der Scores und Loadings bei der PCA

5.3.3 Berechnung der Hauptkomponenten

Die am häufigsten verwendete Methode zur Lösung der PCA ist der *Nonlinear Iterative Partial Least Squares* (NIPALS) Algorithmus. Es handelt sich dabei um ein Näherungsverfahren zur Berechnung der *Scores* und *Loadings* [WOLD, 1966]. Zur Berechnung einer Hauptkomponente t_i beginnt das Verfahren mit einer beliebigen Lösung, die iterativ verbessert wird, bis sie eine definierte Fehlerschwelle unterschreitet. Der NIPALS-Algorithmus umfasst dabei eine Reihe von Arbeitsschritten, die im Folgenden vorgestellt werden sollen.

- 1) Ausgangspunkt ist die Zuweisung des temporären Scorevektors t_{ite} ,

$$t_{ite} = x_j, \quad (5.10)$$

mit dem Spaltenvektor x_j der Datenmatrix X mit der höchsten Varianz s_{xj}^2 ,

$$s_{xj}^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2. \quad (5.11)$$

- 2) Zu diesem Scorevektor wird der zugehörige temporäre Loadingvektor p_{ite} ,

$$p_{ite} = \frac{X^T \cdot t_{ite}}{t_{ite}^T \cdot t_{ite}}, \quad (5.12)$$

durch Projektion der Datenmatrix X auf den Unterraum t_{ite} berechnet.

Der gesuchte Richtungsvektor der Hauptkomponente p_i ,

$$p_i = \frac{p_{ite}}{\|p_{ite}\|} = \frac{p_{ite}}{\sqrt{p_{ite}^T \cdot p_{ite}}}, \quad (5.13)$$

ergibt sich durch Normierung des temporären Loadingvektors p_{ite} auf die Länge 1.

- 3) Den zugehörigen korrigierten Scorevektor t_i ,

$$t_i = X \cdot p_i, \quad (5.14)$$

erhält man wiederum durch Projektion der Matrix X auf den Unterraum p_i .

- 4) Durch Bildung des Gütekriteriums J_i ,

$$J_i = \|t_{ite} - t_i\| = \sqrt{\sum_{l=1}^n (t_{ite} - t_{il})^2}, \quad (5.15)$$

wird im vierten Schritt der temporäre mit dem korrigierten Scorevektor über die Quadratsumme der Differenzen verglichen.

- 5) Bei Überschreiten eines vordefinierten Schweltenwertes J_{crit} wird der temporäre Scorevektor t_{ite} ,

$$t_{ite} = t_i, \quad (5.16)$$

mit dem unter 3) berechneten Scorevektor t_i belegt und es beginnt eine erneute Iteration des Algorithmus bei Schritt 2.

Andernfalls ist die Berechnung der Hauptkomponente t_i abgeschlossen und es wird mit Schritt 6 fortgefahrene.

- 6) Für die Berechnung einer weiteren Hauptkomponente mit $i = i + 1$ wird zunächst die neue Residuenmatrix E ,

$$E = X - t_i \cdot p_i^T, \quad (5.17)$$

bestimmt, indem die durch die Hauptkomponente t_i beschriebene Information aus der Datenmatrix X entfernt wird.

- 7) In einem letzten Schritt wird die neue Datenmatrix X ,

$$X = E, \quad (5.18)$$

für den Neustart des Algorithmus bei Schritt 1 aus der unter 6) bestimmten Residuenmatrix E festgelegt.

Die Schritte 1 bis 7 werden entweder so oft wiederholt, bis alle möglichen Hauptkomponenten berechnet wurden oder bis ein bestimmter Anteil der Gesamtvarianz durch das PCA-Modell erklärt ist. Die maximale Anzahl berechenbarer Hauptkomponenten entspricht der kleineren Anzahl an Beobachtungen n oder Variablen m .

Die Schritte 2 und 3 zur Berechnung der Scores und Loadings des PCA-Modells stellen Lösungen nach der Methode der kleinsten Fehlerquadrate (*Least Squares*) dar, von denen der Name des NIPALS-Algorithmus herrührt.

5.4 Partial Least Squares Regression (PLSR)

5.4.1 Multivariate Kalibrierung mittels PLSR

Das Ziel der multivariaten Kalibrierung ist die Abschätzung, in der MVDA auch Vorhersage genannt, schwer messbarer Variablen anhand einer Reihe von Größen, die messtechnisch einfacher oder schneller zu erfassen sind. In dieser Arbeit werden dafür hauptsächlich Spektren zur Prädiktion von Bioprozessgrößen verwendet.

Für die Vorhersage wird ein mathematisches Modell benötigt, das den Zusammenhang zwischen den Spektren und den Zielgrößen beschreibt. Ein Modell wird mit Hilfe geeigneter

Kalibriertdaten (Spektren und Referenzmessungen) erzeugt. Hierfür kann z.B. das Verfahren *Partial Least Squares Regression* (PLSR) zum Einsatz kommen.

Auch bei der PLSR befinden sich die Spektren in der modifizierten Datenmatrix X mit den Dimensionen $(n \times m)$. Die Referenzwerte werden dagegen zunächst in einer Messdatenmatrix M zusammengefasst. Diese besteht auch aus n Zeilen, hat jedoch $h = 1, 2, \dots, v$ Spalten. Vor Verwendung dieser Daten werden sie ebenfalls einer Vorbehandlung unterzogen. In der Regel wird sich hier auf die in Abschnitt 5.2.2 beschriebene Autoskalierung beschränkt. Die modifizierten Messdaten befinden sich dann in Matrix Y .

Prinzipiell ist die PLSR dazu in der Lage mehrere Ziel- oder Y -Variablen gleichzeitig zu verarbeiten ($v \geq 2$). Dieser Ansatz wird PLS2 genannt. Meistens ist die Bildung separater PLSR-Modelle für einzelne Y -Variablen ($v = 1$) vorteilhaft, da dadurch eine individuellere Modellbildung möglich ist [KESSLER, 2008].

Aus diesem Grund wurde in dieser Arbeit ausschließlich mit den sogenannten PLS1-Modellen gearbeitet. Die folgenden theoretischen Betrachtungen der PLSR behandeln jedoch den allgemeinen Fall der PLS2, in dem die PLS1 als Sonderfall enthalten ist.

5.4.2 Das mathematische Modell der PLSR

Bei der PLSR wird ein multivariater Ansatz verfolgt, mit dem die Zielgrößen in der Matrix Y ,

$$Y = X \cdot B + G \quad (5.19)$$

mit

$Y := (n \times v)$ autoskalierte Messdatenmatrix

$X := (n \times m)$ modifizierte Datenmatrix

$B := (m \times v)$ PLSR-Koeffizientenmatrix

$G := (n \times v)$ Residuenmatrix des Regressionsansatzes,

aus den Variablen in X abgeschätzt werden sollen.

Das erste Ziel bei der Bildung eines PLS-Regressionsmodells ist die Bestimmung der darin enthaltenen Regressionskoeffizientenmatrix B . Bild 5.6 zeigt die dabei beteiligten Matrizen.

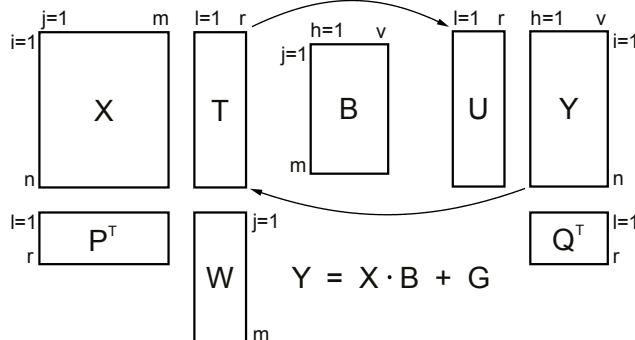


Bild 5.6: Korrelation der Datenmatrizen X und Y durch eine PLS-Regression

Wie bereits erwähnt, basiert die PLSR auf der Hauptkomponentenanalyse. Im Prinzip wird für die Datenmatrix X,

$$X = T \cdot P^T + E \quad (5.20)$$

mit

X := (n × m) modifizierte Datenmatrix

T := (n × r) Scorematrix von X

P^T := (r × m) transponierte Loadingmatrix von X

E := (n × m) Residuenmatrix von X,

sowie die Messdatenmatrix Y,

$$Y = U \cdot Q^T + F \quad (5.21)$$

mit

Y := (n × v) autoskalierte Messdatenmatrix

U := (n × r) Scorematrix von Y

Q^T := (r × v) transponierte Loadingmatrix von Y

F := (n × v) Residuenmatrix von Y,

gleichzeitig eine PCA durchgeführt.

Für die Herleitung der Koeffizientenmatrix B muss es dabei zu einem Informationsaustausch kommen, um beide Scorematrizen T und U aufeinander abzubilden. Dieser Austausch ist in Bild 5.6 durch Pfeile angedeutet.

Die bei der PLSR berechneten latenten Variablen werden somit von beiden Datenräumen beeinflusst. Dadurch unterscheiden sich die Scores T und die Loadings P von denen einer reinen PCA des X-Datenraumes. Aus diesem Grund spricht man im Falle der PLSR auch nicht von Hauptkomponenten sondern von PLS-Komponenten. Auch der Begriff Faktoren ist dabei gebräuchlich.

Ein weiteres wichtiges Element des PLSR-Modells ist die Weightmatrix W. Diese hat die gleichen Dimensionen wie die Loadingmatrix P und wird auch als gewichtete Loadingmatrix von X bezeichnet. Sie stellt die Verbindung zwischen den beiden beteiligten Datenräumen her [KESSLER, 2008].

Die Berechnung der PLS-Komponenten und damit der Scores T und U sowie der zugehörigen Loadings P und Q und den Weights W wird im nächsten Abschnitt ausführlich beschrieben.

Es lässt sich zeigen, dass die gesuchte Regressionskoeffizientenmatrix B,

$$B = W \cdot (P^T \cdot W)^{-1} \cdot Q^T \quad (5.22)$$

mit

B := (m × v) PLSR-Koeffizientenmatrix

W := (m × r) Weightmatrix der PLSR

P^T := (r × m) transponierte Loadingmatrix von X

Q^T := (r × v) transponierte Loadingmatrix von Y,

aus den Weight- und Loadingmatrizen W, P und Q berechnet werden kann [WOLD et al., 1993].

5.4.3 Berechnung der PLS-Komponenten

Die Berechnung der PLS-Komponenten wird mit einer modifizierten Variante des NIPALS-Algorithmus durchgeführt [WOLD, 1975]. Dabei kommen einige lokale Modelle für X,

$$X = T \cdot W^T + E' \quad (5.23)$$

und

$$X = U \cdot W^T + E'', \quad (5.24)$$

sowie für Y,

$$Y = T \cdot Q^T + F', \quad (5.25)$$

zur Anwendung, mit denen die entsprechenden Datenräume ebenfalls beschrieben werden können.

Wie bei der Berechnung der Hauptkomponenten eines PCA-Modells umfasst der NIPALS-Algorithmus zur Bestimmung einer PLS-Komponente einige Schritte, die ggf. iterativ wiederholt werden, bis eine Konvergenz erreicht wird.

- Der Algorithmus beginnt mit der Zuweisung des temporären Y-Scorevektors u_{ite} ,

$$u_{ite} = y_h, \quad (5.26)$$

aus der Spalte h der Messdatenmatrix Y mit der höchsten Varianz.

- Über das lokale Modell in Gl. 5.24 und unter Vernachlässigung der Residuenmatrix E'' wird daraus der temporäre Weightvektor w_{ite} ,

$$w_{ite} = \frac{X^T \cdot u_{ite}}{u_{ite}^T \cdot u_{ite}}, \quad (5.27)$$

über einen Least-Squares-Ansatz berechnet.

Dieser wird durch Normierung auf die Länge 1 in den gesuchten Weightvektor w_i ,

$$w_i = \frac{w_{ite}}{\|w_{ite}\|} = \frac{w_{ite}}{\sqrt{w_{ite}^T \cdot w_{ite}}}, \quad (5.28)$$

transformiert.

- Über das lokales Modell in Gl. 5.23 und unter Vernachlässigung der Residuen E' lässt sich daraus der Scorevektor t_i ,

$$t_i = X \cdot w_i, \quad (5.29)$$

durch Projektion der Matrix X auf den Unterraum w_i berechnen.

- Der Scorevektor t_i erlaubt einerseits eine Berechnung des X-Loadingvektors p_i ,

$$p_i = \frac{X^T \cdot t_i}{t_i^T \cdot t_i}, \quad (5.30)$$

gemäß der allgemeinen Form des PCA-Modells (Gl. 5.20) und andererseits eine Berechnung des Y-*Loadingvektors* q_i ,

$$q_i = \frac{Y^T \cdot t_i}{t_i^T \cdot t_i}, \quad (5.31)$$

über das dritte lokale Modell (Gl. 5.25) der PLSR unter Vernachlässigung von F' .

- 5) In einem fünften Schritt wird der korrigierte Y-Scorevektor u_i ,

$$u_i = \frac{Y \cdot q_i}{q_i^T \cdot q_i}, \quad (5.32)$$

durch Projektion der Matrix Y auf den Unterraum q_i bestimmt.

- 6) Analog zum NIPALS Algorithmus der PCA wird wieder ein Gütfunktional J_i ,

$$J_i = \|u_{ite} - u_i\| = \sqrt{\sum_{i=1}^n (u_{ite} - u_i)^2}, \quad (5.33)$$

berechnet, welches als Maß für die Konvergenz des Verfahrens herangezogen wird.

- 7) Bei Überschreiten eines Schwellenwertes J_{crit} wird dem temporären Scorevektor u_{ite} ,

$$u_{ite} = u_i, \quad (5.34)$$

der zuvor berechneten Scorevektor u_i zugewiesen und bei Schritt 2 eine weitere Iteration des Algorithmus begonnen.

Bei Unterschreiten des Schwellenwertes ist die Berechnung der aktuellen PLS-Komponente abgeschlossen und es folgt Schritt 8.

- 8) Für die Berechnung einer weiteren PLS-Komponente mit $i = i + 1$ erfolgt zunächst eine Neubestimmung der Residuenmatrizen E,

$$E = X - t_i \cdot p_i^T, \quad (5.35)$$

und F,

$$F = Y - u_i \cdot q_i^T, \quad (5.36)$$

indem die enthaltene Information in den Scores und Loadings t_i und p_i bzw. u_i und q_i der aktuellen PLS-Komponente aus den Datenmatrizen X und Y entfernt wird.

- 9) In einem letzten Schritt werden dann die Datenmatrizen X,

$$X = E, \quad (5.37)$$

und Y,

$$Y = F, \quad (5.38)$$

für den Neustart des NIPALS-Algorithmus der PLSR bei Schritt 1 mit den Residuenmatrizen E und F besetzt.

Auch bei der PLSR kann der Algorithmus nach der Bestimmung der benötigten Anzahl an Komponenten gestoppt werden. Die Anzahl erforderlicher Faktoren ist vom Vorhersagefehler abhängig und wird über diverse Methoden der Modellvalidierung (siehe Abschnitt 5.5.3) festgelegt [KESSLER, 2008]. Nach Beendigung des NIPALS-Algorithmus erfolgt eine Berechnung der Regressionskoeffizientenmatrix B gemäß Gl. (5.22).

5.5 Erstellung multivariater Modelle

5.5.1 Generelle Anforderungen an das Datenmaterial

Eine große Rolle bei der Bildung multivariater Modelle spielt das verwendete Datenmaterial. Dieses sollte repräsentative Beobachtungen enthalten, die den untersuchten Prozess möglichst vollständig beschreiben. Der Informationsgehalt der Beobachtungen ist somit höher zu bewerten, als die Menge an Datenpunkten [KESSLER, 2008].

Im Falle eines PLSR-Modells zur Quantifizierung einer Substratkonzentration sollte beispielsweise der gesamte relevante Konzentrationsbereich in den zur Modellkalibrierung verwendeten Daten enthalten sein. Dabei ist weiterhin ein Maximum an Varianz weiterer möglicher Einflussgrößen auf die Prädiktorvariablen (Spektren) abzubilden. Anzustreben ist außerdem eine uniforme Verteilung der Konzentrationswerte, um eine Wichtung des Modells auf bestimmte Zustände zu vermeiden [KESSLER, 2008].

Reale Datensätze erfüllen häufig nicht alle Anforderungen, da sie beispielsweise prozessbegleitend erzeugt werden müssen. Dieser Umstand ist bei der Modellerstellung zu beachten und das erstellte Modell kritisch zu bewerten.

5.5.2 Ausreißerdetektion für multivariate Daten

Aufgrund des Ansatzes der kleinsten Fehlerquadrate bei PCA und PLSR können Ausreißer einen hohen Einfluss auf das resultierende multivariate Modell haben [WOLD et al., 1987]. Aus diesem Grund sollten die Daten gegebenenfalls von Ausreißern bereinigt werden.

Eine Beobachtung, die sich stark von anderen unterscheidet, kann mit Hilfe des Hotelling T^2 -Tests identifiziert werden. Die dafür notwendige Hotelling T^2 -Statistik stellt eine Generalisierung der Student t-Statistik für multivariate Daten dar und überprüft Beobachtungen hinsichtlich einer Normalverteilung [HOTELLING, 1951].

Der Hotelling T^2 -Wert einer Beobachtung i,

$$T_i^2 = \sum_{l=1}^r \frac{(t_{il} - \bar{t}_l)^2}{s_{il}^2} \quad (5.39)$$

mit

t_{il} := Score der Komponente l für die Beobachtung i

\bar{t}_l := Mittelwert des Scorevektors t_l

s_{il}^2 := Varianz der Scores von Komponente l,

beschreibt den normierten Abstand der Beobachtung zum Schwerpunkt des Modells für alle berechneten Hauptkomponenten.

Bei der Verwendung mittenzentrierter oder autoskalierter Daten ist der Score-Mittelwert \bar{t}_l jeder Komponente l gleich 0, wodurch sich der Schwerpunkt des Modells im Ursprung befindet.

Bei einer Beobachtung mit einem Hotelling T^2 -Wert oberhalb des kritischen Wertes $T_{crit\alpha}^2$,

$$T_{crit\alpha}^2 = \frac{r \cdot (n - 1)}{n - r} \cdot F_{(a,r,n-r)} \quad (5.40)$$

mit

α := gewählte Irrtumswahrscheinlichkeit

r := Anzahl berechneter Hauptkomponenten

n := Anzahl an Beobachtungen

$F_{(a,r,n-r)}$:= Kritischer Wert einer F-Verteilung mit einer Irrtumswahrscheinlichkeit α sowie r und $n - r$ Freiheitsgraden,

handelt es sich mit einer gewählten Irrtumswahrscheinlichkeit α um einen Ausreißer. Ein typischer Wert für α ist 5 %, woraus das übliche 95 % Vertrauensintervall resultiert.

Der benötigte kritische F-Wert lässt sich aus der kumulativen Verteilungsfunktion einer F-Verteilung in Abhängigkeit der Irrtumswahrscheinlichkeit α sowie den r und $n - r$ Freiheitsgraden berechnen oder aus Tabellenwerken ablesen (z.B. [Ross, 2006]).

Als grafische Werkzeuge zur Identifizierung von Ausreißern mit Hilfe der Hotelling T^2 -Statistik kommen häufig sogenannte *Score Scatter Plots* zum Einsatz. In diesen sind die Scores einer Komponente gegen die Scores einer anderen Komponente aufgetragen und erlauben damit eine einfache grafische Analyse des multivariaten Modells.

Beobachtungen mit ungewöhnlichen Eigenschaften erscheinen in den Randbereichen der sich ergebenden Punktwolke. Durch Einzeichnung einer Hotelling T^2 -Ellipse für die betrachteten Hauptkomponenten, lassen sich die Daten auf Ausreißer überprüfen.

Bei einer gewählten Irrtumswahrscheinlichkeit α von 5 % handelt es sich bei Proben außerhalb der Ellipse demnach mit einer Wahrscheinlichkeit von 95 % um Ausreißer. Bild 5.7a zeigt einen *Score Scatter Plot* mit einer Hotelling T^2 -Ellipse für die Scorevektoren t_1 und t_2 .

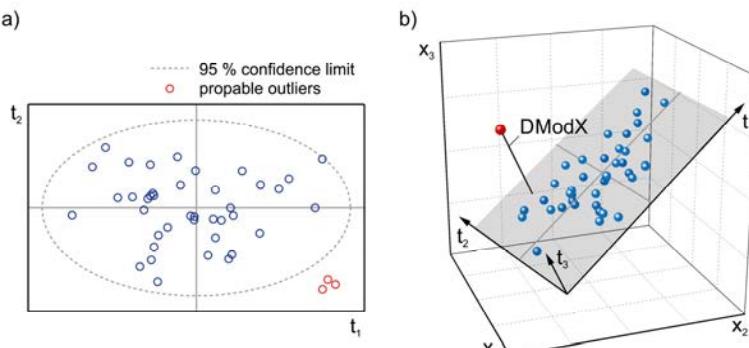


Bild 5.7: a) Score Scatter Plot mit Hotelling T^2 -Ellipse als 95 % Vertrauengrenze
b) DModX als orthogonale Distanz einer Probe zur Modellebene

Die Radien der Hotelling T^2 -Ellipse r_{Tot} ,

$$r_{\text{Tot}} = \sqrt{\frac{2 \cdot (n - 1)}{n - 2} \cdot F_{(\alpha, 2, n-2)}} \cdot s_{\text{tl}} \quad (5.41)$$

mit

$F_{(\alpha, 2, n-2)}$:= Kritischer Wert einer F-Verteilung mit einer Irrtumswahrscheinlichkeit α sowie 2 und $n - 2$ Freiheitsgraden,

ergeben sich aus der Wurzel des kritischen Wertes T^2_{critx} gemäß Gl. (5.40) für $r = 2$ und aus den Standardabweichungen s_{tl} der beiden betrachteten Hauptkomponenten (z.B. $I = 1, 2$).

Eine weitere Möglichkeit zur Identifikation von Ausreißern bietet die DModX (*Distance to Model in X*) genannte Kenngröße. Der DModX ist ein Maß für den orthogonalen Abstand einer Beobachtung zu der Modellebene (Hyperebene) im Originaldatenraum X (siehe Bild 5.7b). Er zeigt somit Abweichungen einer Beobachtung auf, die durch das Modell nicht beschrieben werden.

Der DModX basiert auf der in der Literatur häufig anzutreffenden Q-Statistik und wird von der Software Umetrics SIMCA anstelle der Q-Residuen verwendet [ERIKSSON et al., 2001].

Die absolute Distanz DModX_{absi} einer Beobachtung i,

$$\text{DModX}_{\text{absi}} = \sqrt{\frac{1}{m - r} \cdot \sum_{j=1}^m e_{ij}^2} \quad (5.42)$$

mit

m := Anzahl an Variablen in der Datenmatrix X

r := Anzahl an berechneten Hauptkomponenten

e_{ij} := Element der Residuenmatrix E bei der PCA,

entspricht der residualen Standardabweichung einer Beobachtung i für alle Variablen.

Mit Einführung der normierten Distanz DModX_{nor} einer Beobachtung i,

$$\text{DModX}_{\text{nor}} = \frac{\text{DModX}_{\text{absi}}}{\text{DModX}_{\text{av}}} = \text{DModX}_i, \quad (5.43)$$

wird eine Vergleichbarkeit der Distanzen für unterschiedliche Modelle erreicht. Im Weiteren wird diese normierte Distanz zum Modell allgemein mit DModX bezeichnet.

Die für eine Normierung benötigte mittlere Distanz DModX_{av},

$$\text{DModX}_{\text{av}} = \sqrt{\frac{1}{(n - r - r_0) \cdot (m - r)} \cdot \sum_{i=1}^n \sum_{j=1}^m e_{ij}^2} \quad (5.44)$$

mit

n := Anzahl an Beobachtungen in der Datenmatrix X

r_0 = 1 bei zentrierten Modellen, sonst 0 [ERIKSSON et al., 2001],

stellt die residuale Standardabweichung aller n Beobachtungen dar.

In Analogie zum Hotelling T^2 -Wert kann mit Hilfe einer F-Statistik ein kritischer Wert DmodX_{critx} für eine gewählte Irrtumswahrscheinlichkeit α berechnet werden [ERIKSSON et al., 2001], oberhalb dessen eine Beobachtung als Ausreißer gilt.

5.5.3 Variablenselektion

Für multivariate Modelle sollten nur solche X-Variablen herangezogen werden, die relevante Informationen enthalten. Irrelevante Prädiktoren erhöhen lediglich die Modellkomplexität und können messtechnisches Rauschen mit in das Modell einbringen. Beides verschlechtert zu meist die Leistung des Modells [KESSLER, 2008].

Eine Möglichkeit der Identifikation irrelevanter Variablen für PLSR-Modelle besteht in der Be trachtung der Regressionskoeffizienten im Spaltenvektor b_j . Diese können zu diesem Zweck gegen den Variablenindex j aufgetragen werden.

Regressionskoeffizienten mit einem Wert von oder nahe bei 0 deuten darauf hin, dass die zugehörige Variable keine Rolle bei der Quantifizierung der entsprechenden Zielgröße spielt. Ein Ausschluss der Variable sollte die Leistung des Modells also nicht direkt beeinflussen aber die Modellkomplexität senken.

Vorsicht ist bei diesem Verfahren bei der Verwendung nichtskalierter X-Variablen geboten, da hier die Höhe des Wertes des Regressionskoeffizienten auch von der Höhe der Messwerte der entsprechenden Variable abhängt. Dies ist meist bei der Verwendung von Spektren der Fall, da diese nicht autoskaliert werden.

Deshalb wurde von Wold et al. eine *Variable Influence on Projection* (VIP) genannte Kenngrö ße entwickelt, die den Beitrag einer X-Variablen an der durch das Modell erklärten Y-Varianz beschreibt [WOLD et al., 1993; FARRÉS et al., 2015]. In dieser Arbeit wird eine durch [CHONG & JUN, 2005] modifizierte Form dieser Kenngrö ße genutzt.

Zur Berechnung der Kenngrö ße VIP einer X-Variable j ,

$$\text{VIP}_j = \sqrt{\frac{m \cdot \sum_{l=1}^r (\text{SSY}_l \cdot w_{jl\text{nor}}^2)}{\text{SSY}_{\text{tot}}}} \quad (5.45)$$

mit

m := Anzahl an X-Variablen

r := Anzahl berechneter PLS-Komponenten,

werden die durch die PLS-Komponenten l erklärten Anteile der Y-Varianz SSY_l ,

$$\text{SSY}_l = c_l^2 \cdot t_l^T \cdot t_l \quad (5.46)$$

mit

t_l := $(n \times 1)$ X-Scorevektor der PLS-Komponente l

c_l := Koeffizient der inneren Beziehung des PLSR-Modells für Komponente l ,

zu den entsprechenden normierten PLS-Weights $w_{jl\text{nor}}$ dieser Variable,

$$w_{jl\text{nor}} = \frac{w_{jl}}{\sqrt{w_l^T w_l}} \quad (5.47)$$

mit

w_{jl} := Weight der Komponente l für die Variable j

w_l := $(m \times 1)$ Weight-Spaltenvektor der Komponente l ,

in Beziehung gesetzt und dann auf ein Maß des insgesamt durch das Modell erklärten Anteils der Y-Varianz SSY_{tot} ,

$$SSY_{tot} = \sum_{l=1}^r SSY_l, \quad (5.48)$$

bezogen.

Der in die Berechnung des Anteils erklärter Varianz SSY_l mit eingehende Koeffizient der inneren Beziehung c_l ,

$$c_l = \frac{u_l^T \cdot t_l}{t_l^T \cdot t_l}, \quad (5.49)$$

mit

$u_l := (n \times 1)$ Y-Scorevektor der PLS-Komponente l,

kann aus den X- und Y-Scorevektoren des PLSR-Modells berechnet werden.

5.5.4 Validierung multivariater Modelle

Die Validierung ist einer der wichtigsten Arbeitsschritte bei der Erstellung multivariater Modelle, da sie die Entscheidungsgrundlage für die Anzahl zu berechnender Komponenten liefert.

Der durch das Modell erklärte Anteil der Varianz R_x^2 ,

$$R_x^2 = 1 - \frac{1}{n \cdot m \cdot s_x^2} \cdot \sum_{i=1}^n \sum_{j=1}^m e_{ij}^2 \quad (5.50)$$

mit

$e_{ij} :=$ Element der Residuenmatrix E,

an der Gesamtvarianz s_x^2 der Datenmatrix X,

$$s_x^2 = \frac{1}{n \cdot m} \cdot \sum_{i=1}^n \sum_{j=1}^m (x_{ij} - \bar{x})^2 \quad (5.51)$$

mit

$x_{ij} :=$ Element der Datenmatrix X,

ist ein Maß dafür, wie gut ein Modell die Eingangsdaten beschreibt und wird insbesondere für die Bewertung von PCA-Modellen verwendet. Der dabei berücksichtigte Mittelwert \bar{x} ,

$$\bar{x} = \frac{1}{n \cdot m} \cdot \sum_{i=1}^n \sum_{j=1}^m x_{ij}, \quad (5.52)$$

ist der Gesamtmittelwert der Datenmatrix X.

Bei der Untersuchung von PLSR-Modellen wird zusätzlich der erklärte Anteil der Varianz R_y^2 ,

$$R_y^2 = 1 - \frac{1}{n \cdot v \cdot s_y^2} \cdot \sum_{i=1}^n \sum_{h=1}^v f_{ih}^2, \quad (5.53)$$

an der Gesamtvarianz der Matrix Y betrachtet, der sich in Analogie zu R_x^2 aus der Residuenmatrix F mit den Dimensionen $(n \times v)$ berechnet.

Die Größen R_X^2 und R_Y^2 werden auch als Güte der Anpassung bezeichnet. Für PLSR-Modelle ist jedoch die Güte der Vorhersage von größerer Bedeutung.

Das liegt darin begründet, dass durch die Berechnung weiterer Komponenten auch die erklärte Varianz ansteigt, bis hin zu einem Maximalwert von 1. Dabei wird jedoch auch zufällige Varianz in X oder Y beschrieben, bei der es sich zum Beispiel um messtechnisches Rauschen handeln kann. In einem solchen Fall spricht man von Überanpassung (*Overfit*) des Modells.

Ein *Overfit* wirkt sich meist negativ auf die Leistung des Modells bei der Vorhersage unbekannter Proben aus. Daher sollte eine Validierung von PLSR-Modellen mit Proben durchgeführt werden, die nicht an der eigentlichen Kalibrierung des Modells beteiligt sind.

Mehr noch sollte ein unabhängiger Validierdatensatz (*validation set VS*) verwendet werden, der aus einer separaten Messreihe stammt. Dies wird auch als externe Validierung eines Modells der Zielgröße y_h bezeichnet.

Die Güte der Vorhersage (*prediction P*) R_{Ph}^2 einer Variable h ,

$$R_{Ph}^2 = 1 - \frac{\sum_{i=1}^{n_{VS}} (y_{ViH} - \hat{y}_{ViH})^2}{\sum_{i=1}^{n_{VS}} (y_{ViH} - \bar{y}_{ViH})^2} \quad (5.54)$$

n_{VS} := Anzahl an Beobachtungen im Validierdatensatz VS

y_{ViH} := Messwert der Zielgröße y_h für das Objekt i in VS

\hat{y}_{ViH} := Modellschätzwert der Zielgröße y_h für das Objekt i in VS

\bar{y}_{ViH} := Mittelwert der Zielgröße y_h in VS,

entspricht dem Bestimmtheitsmaß R^2 der linearen Anpassung einer Auftragung von Modellschätzwerten gegen ihre Referenzmessungen.

Weiterhin erfolgt eine Bewertung von PLSR-Modellen anhand des mittleren Vorhersagefehlers (*root mean square error of prediction*) $RMSEP_h$,

$$RMSEP_h = \sqrt{\frac{1}{n_{VS}} \cdot \sum_{i=1}^{n_{VS}} (y_{ViH} - \hat{y}_{ViH})^2}, \quad (5.55)$$

der als wichtigste Kenngröße bei der Validierung des Modells zur Prädiktion der Zielgröße y_h betrachtet werden kann.

Zum Zwecke der Vergleichbarkeit von Modellen unterschiedlicher Zielgrößen wird in dieser Arbeit außerdem ein auf den jeweiligen Kalibrierbereich (*calibration set CS*) normierter relativer Fehler $RMSEP_{hrel}$,

$$RMSEP_{hrel} = \frac{RMSEP_h}{y_{CSHmax} - y_{CSHmin}} \cdot 100\% \quad (5.56)$$

mit

y_{CSHmax} := maximaler Wert der Zielgröße y_h im Kalibrierdatensatz CS

y_{CSHmin} := minimaler Wert von y_h in CS,

zur externen Validierung von PLSR-Modellen eingesetzt.

Steht kein geeigneter Validierdatensatz zur Verfügung, wird eine interne Validierungsmethode verwendet. Bei diesem Verfahren, auch Kreuzvalidierung (*cross validation cv*) genannt, wird jede Probe des Kalibrierdatensatzes genau einmal von der Modellbildung ausgeschlossen und dann anhand des so erzeugten Untermodells vorhergesagt.

Im Falle der Kreuzvalidierung werden die vorgestellten Kenngrößen zur PLSR-Modellvalidierung mit R_{cv}^2 beziehungsweise RMSEcv bezeichnet.

5.6 Multivariate Klassifizierung

5.6.1 Allgemeine Informationen

Multivariate Klassifizierungsverfahren haben ihren Ursprung in den 1930er Jahren in der klassischen Statistik [FISHER, 1936]. Viele verschiedene Methoden entstanden parallel in den 1970er Jahren vor allem in den aufstrebenden Fachgebieten der Informatik (maschinelles Lernen) oder der chemischen Analytik (Chemometrie) [BRERETON, 2015]. Mögliche Anwendungen reichen von Gesichtserkennungssystemen, Überwachung der Qualität von Lebensmitteln bis hin zur medizinischen Diagnosefindung.

In der Literatur werden die Begriffe Mustererkennung (*pattern recognition*) oder Diskriminanzanalyse häufig synonym verwendet.

Im folgenden Abschnitt wird die insbesondere im chemometrischen Umfeld weit verbreitete Methode SIMCA vorgestellt.

5.6.2 Soft Independent Modelling of Class Analogy (SIMCA)

Die multivariate Klassifizierungsmethode *Soft Independent Modelling of Class Analogy* (SIMCA) wurde von Svante Wold auf Basis der PCA entwickelt [WOLD, 1976; WOLD & SJÖSTRÖM, 1977]. Es handelt sich um ein sogenanntes überwachtes Verfahren, bei dem für jede betrachtete Klasse ein unabhängiges PCA-Modell erstellt wird.

Unter der Voraussetzung, dass die verwendeten Daten in der Lage sind, Unterschiede zwischen den Klassen und Gemeinsamkeiten innerhalb einer Klasse zu beschreiben, eignet sich die SIMCA-Methode sehr gut zur Abschätzung der Klassenzugehörigkeit unbekannter Objekte (Proben) [ERIKSSON et al., 2001].

Entscheidend für die Einteilung einer unbekannten Probe i aus einem Vorhersagedatensatz (*prediction set*) PS zu einer Klasse K ist die Distanz der Probe i zum PCA-Modell dieser Klasse.

Als Maß hierfür wird die erweiterte Distanz DModX+,

$$\text{DModX+}_{\text{KPSi}} = \sqrt{\text{DModX}_{\text{KPSi}}^2 + \text{DModT}_{\text{KPSi}}^2}, \quad (5.57)$$

herangezogen, die sich als die geometrische Länge aus der orthogonalen Distanz DModX_{KPSi} der Probe zur Modellebene und einer hier als DModT_{KPSi} bezeichneten Distanz der Probe zur Vertrauengrenze in der Modellebene berechnet [ERIKSSON et al., 2001].

Basierend auf DModX+, dessen Konzept in Bild 5.8 veranschaulicht wird, lässt sich für jede unbekannte Probe i eine Wahrscheinlichkeit β_{KPSi} für die Zugehörigkeit zu der betrachteten Klasse K berechnen [ERIKSSON et al., 2001]. Für die Klasseneinteilung wird zweckmäßig eine Grenzwahrscheinlichkeit β_{crit} von 0,5 verwendet.

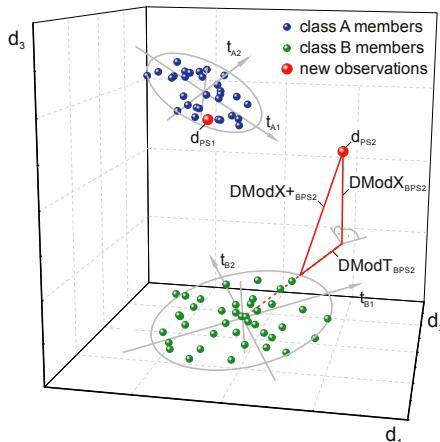


Bild 5.8: Verdeutlichung der Klasseneinteilung anhand des DModX+

In Bild 5.8 sind zwei PCA-Modelle für die Klassen A (blau) und B (grün) gezeigt. Die Modelle bestehen jeweils aus zwei Hauptkomponenten und sind zum Zwecke der Veranschaulichung im Originaldatenraum D mit $m = 3$ Variablen dargestellt.

Die Modelle basieren auf den Daten ihrer jeweiligen Kalibrierdatensätze, wohingegen die beiden neuen Proben d_{PSi} (rot) unbekannt sind. Die Probe d_{PS1} befindet sich in der Modelldomäne der Klasse A und würde somit in diese eingeteilt werden. Die Probe d_{PS2} liegt hingegen weit außerhalb beider Klassenmodelle. Daher ist sie keiner der beiden Klassen zuzuordnen.

Das Konzept der orthogonalen Distanz zur Modellebene DModX wird vor allem zur Detektion von Ausreißern verwendet (siehe Abschnitt 5.5.2) und basiert auf der Betrachtung des modifizierten Datenraums X.

Die orthogonale Distanz $DModX_{KPSi}$ einer unbekannten Probe i zum PCA-Modell K,

$$DModX_{KPSi} = \frac{1}{DModX_{Kav}} \cdot \sqrt{\frac{1}{m - r_K} \cdot \sum_{j=1}^m (x_{PSij} - \hat{x}_{KPSij})^2} \quad (5.58)$$

mit

$DModX_{Kav}$:= mittlere Distanz zum PCA-Modell K (siehe Gl. (5.44))

r_K := Anzahl an Hauptkomponenten im Modell der Klasse K

\hat{x}_{KPSij} := X-Schätzwert der Probe i und Variable j gemäß Klasse K,

lässt sich aus den zur Probe i zugehörigen Messwerten x_{PSij} und aus den zugeordneten Schätzwerten \hat{x}_{KPSij} berechnen.

Diese anhand des PCA-Modells der Klasse K bestimmten Schätzwerte \hat{x}_{KPSij} ,

$$\hat{x}_{KPSij} = t_{KPSi} \cdot p_{Kj}^T \quad (5.59)$$

mit

p_{Kj}^T := ($r \times 1$) Spaltenvektor der transponierten Loadingmatrix P_K des Modells K,

ergeben sich aus den Scores t_{KPSi} der unbekannten Probe i und mit den *Loadings* des PCA-Modells K für die betrachtete Variable j.

Der Score-Zeilenvektor t_{KPSi} der unbekannten Probe im Modell K,

$$t_{KPSi} = x_{PSi} \cdot P_K \quad (5.60)$$

mit

$t_{KPSi} := (1 \times r)$ Zeilenvektor der Scores einer unbekannten Probe i im Modell K

$x_{PSi} := (1 \times m)$ Zeilenvektor der Eingangsdaten (gemessen) einer unbekannten Probe i

$P_K := (m \times r)$ Loadingmatrix des PCA-Modells der betrachteten Klasse K,

ist wiederum abhängig von den Messwerten x_{PSi} der unbekannten Probe i und berechnet sich aus der Loadingmatrix P_K der Klasse K.

5.6.3 Validierung von Klassifikatoren

Die Bewertung eines Klassifikators erfolgt auf Basis von Konfusionsmatrizen. Diese Technik stammt ursprünglich aus der Signalentdeckungstheorie [PETERSON et al., 1954]. Für jede betrachtete Klasse ist eine solche Matrix, auch Kontingenztabelle genannt, zu erstellen. Bild 5.9 zeigt den Aufbau einer Konfusionsmatrix.

		class member	
		positive	negative
predicted member	yes	n_{tp} true positive	n_{fn} false negative
	no	n_{fp} false positive	n_{tn} true negative

Bild 5.9: Konfusionsmatrix zur Bewertung eines Klassifikators

Eine Konfusionsmatrix fasst die Ergebnisse der Einteilung einer Gesamtanzahl von n Objekten in die jeweilig betrachtete Klasse zusammen. Da es sich um einen Validierungsschritt handelt, ist die Klassenzugehörigkeit aller n Objekte bekannt.

Vorausgesetzt die Gesamtanzahl n,

$$n = n_p + n_n \quad (5.61)$$

mit

n_p := Anzahl klassenzugehöriger (positiver) Objekte

n_n := Anzahl klassenfremder (negativer) Objekte,

setzt sich aus klassenzugehörigen (n_p) und klassenfremden (n_n) Proben zusammen, ergeben sich für die Validierung eines Klassifikators vier mögliche Resultate.

Bei korrekter Zuordnung einer klassenzugehörigen (positiven) Probe zur betrachteten Klasse, wird das Ergebnis als *true positive* (tp) bezeichnet. Schlägt die Einteilung durch den Klassifi-

kator dagegen fehl, lautet das Ergebnis *false positive* (fp). Analog dazu treten für klassenfremde (negative) Proben mit *true negative* (tn) und *false negative* (fn) zwei weitere mögliche Ergebnisse auf.

Die auf diese Möglichkeiten entfallenen Objektanzahlen werden in der Konfusionsmatrix notiert und für die Ableitung einer Reihe von Kenngrößen genutzt.

Die Sensitivität Se eines Klassifikators,

$$Se = \frac{n_{tp}}{n_{tp} + n_{fp}} = \frac{n_{tp}}{n_p} \quad (5.62)$$

mit

n_{tp} := Anzahl richtig eingeteilter positiver Objekte

n_{fp} := Anzahl falsch vorhergesagter positiver Objekte,

ist definiert zu dem Anteil korrekt vorhergesagter klassenzugehöriger Objekte.

Die Spezifität Sp eines Klassifikators,

$$Sp = \frac{n_{tn}}{n_{tn} + n_{fn}} = \frac{n_{tn}}{n_n} \quad (5.63)$$

mit

n_{tn} := Anzahl richtig eingeteilter negativer Objekte

n_{fn} := Anzahl falsch vorhergesagter negativer Objekte,

beschreibt dagegen wie gut klassenfremde Objekte abgelehnt werden.

Die Gesamtleistung eines Klassifikators kann in einem Fehler E_C der Klassifizierung,

$$E_C = \frac{n_{fp} + n_{fn}}{n} \cdot 100\%, \quad (5.64)$$

als Gesamtanteil falsch vorhergesagter Objekte zusammengefasst werden.

Die Nutzung von Kontingenztabellen setzt diskrete oder binäre Klassifikatoren voraus, die eine ja/nein-Entscheidung über die Klassenzugehörigkeit liefern. Die meisten multivariaten Klassifizierungsverfahren erzeugen jedoch kontinuierliche Größen wie zum Beispiel eine Wahrscheinlichkeit β der Klassenzugehörigkeit. Dies macht eine Definition geeigneter Grenzwerte notwendig.

Andersherum betrachtet ergibt sich daraus jedoch auch die Möglichkeit einer Optimierung der Klassifikatoren hinsichtlich der Sensitivität Se oder Spezifität Sp, je nachdem ob ein eher konservatives oder liberales Verhalten des Klassifikators erwünscht ist [FAWCETT, 2004].

5.7 Support Vector Machines (SVM)

5.7.1 SVM als multivariates Klassifizierungsverfahren

Die Methodik der *Support Vector Machines* wurde von Vapnik und Chervonenkis im Rahmen der statistischen Lerntheorie beschrieben [VAPNIK & CHERVONENKIS, 1974]. Der Methode gelang in den 1990er Jahren der Durchbruch, nachdem einige Erweiterungen den Algorithmus beispielsweise zur nichtlinearen multivariaten Regression befähigten [VAPNIK, 2000].

Ursprünglich handelt es sich jedoch um ein multivariates Klassifizierungsverfahren, dessen Funktionsweise im Folgenden erläutert werden soll.

Ausgegangen wird zunächst von zwei sich nicht überlappenden, linear separierbaren Klassen A und B im m -dimensionalen Raum der Eingangsdaten X . Gesucht wird eine optimale Trennebene g_{opt} . Bild 5.10a illustriert dies anhand von $m = 2$ X-Variablen.

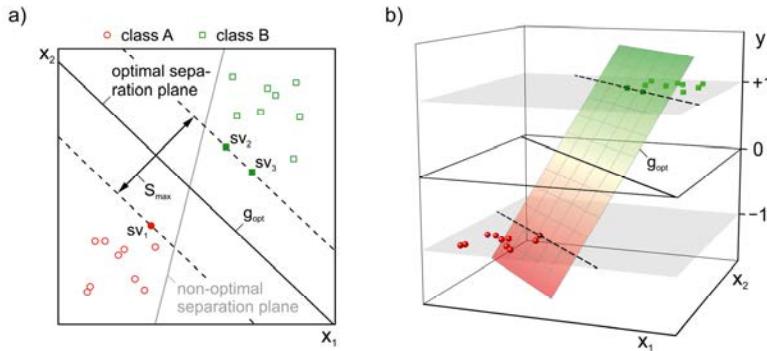


Bild 5.10: a) Nicht überlappende, linear trennbare Klassen für $m = 2$ Dimensionen
b) Übertragung des Problems in den $m + 1$ dimensionalen Raum

Entsprechend seiner Klassenzugehörigkeit wird jeder Beobachtung x_i der Eingangsdaten einer von zwei möglichen Werten $y_i \in \{-1, +1\}$ zugewiesen. Diese Information der Klassentrennung überträgt das Klassifizierungsproblem in einen $m + 1$ dimensionalen Raum, welcher in Bild 5.10b dargestellt ist.

Für die Einteilung eines Objektes i lässt sich eine lineare Entscheidungsfunktion g ,

$$g(x_i) = \hat{y}_i = x_i \cdot w + b = x_{i1} \cdot w_1 + x_{i2} \cdot w_2 + \dots + x_{im} \cdot w_m + b \quad (5.65)$$

mit

\hat{y}_i := Schätzwert der Klassenzugehörigkeit

x_i := $(1 \times m)$ Daten-Zeilenvektor des Objektes i

w := $(m \times 1)$ Wichtungs-Spaltenvektor

b := Bias der Entscheidungsfunktion,

definieren, welche eine (Hyper-) Ebene innerhalb dieses Raums beschreibt.

Die Klasse einer unbekannten Beobachtung x_{PSi} ,

$$\text{class}(x_{PSi}) = \begin{cases} A, & \hat{y}_{PSi} < 0 \\ B, & \hat{y}_{PSi} > 0 \end{cases}, \quad (5.66)$$

kann anhand des Vorzeichens des Schätzwertes \hat{y}_{PSi} bestimmt werden.

Der unbekannte Wichtungsvektor w und der Bias b der Entscheidungsfunktion sind über das Modelltraining mit den Daten x_i und y_i zu ermitteln.

5.7.2 Berechnung einer optimalen Trennebene

Es existiert eine unendliche Anzahl möglicher Trennebenen. Dies ist in Bild 5.10a angedeutet. Gesucht wird die optimale Trennebene g_{opt} , welche eine maximale Spanne S_{max} zwischen den Eingangsdaten aufweist.

Die Spanne wird auf jeder Seite durch mindestens ein Objekt der entsprechenden Klasse begrenzt. Diese Beobachtungen der Trainingsdaten werden Stützvektoren genannt und sind für die Methode namensgebend.

Für einen Stützvektor x_{svi} soll damit sein Schätzwert \hat{y}_{svi} ,

$$\hat{y}_{\text{svi}} = x_{\text{svi}} \cdot w + b \stackrel{!}{=} y_{\text{svi}} \in \{-1, +1\}, \quad (5.67)$$

dem gegebenen Wert seiner Klassenzugehörigkeit y_{svi} entsprechen.

Mit dieser Bedingung lässt sich zeigen, dass die zu maximierende Spanne S ,

$$S = \frac{2}{\|w\|} = \frac{2}{\sqrt{w^T \cdot w}}, \quad (5.68)$$

von der Länge des Wichtungsvektors w abhängt.

Aus numerischer Zweckmäßigkeit wird daraus das zu minimierende Gütekriterium J_g ,

$$J_g = \frac{1}{2} \cdot \|w\|^2 = \frac{1}{2} \cdot w^T \cdot w = \frac{1}{2} \cdot \sum_{j=1}^m w_j^2, \quad (5.69)$$

abgeleitet [KECMAN, 2005].

Aufgrund der in Gl. (5.67) beschriebenen Voraussetzung muss die gesuchte optimale Trennebene g_{opt} außerdem die Nebenbedingung,

$$y_i \cdot (x_i \cdot w + b) \geq 1, \quad (5.70)$$

für die Trainingsdaten x_i und y_i erfüllen.

Dies stellt ein klassisches quadratisches Optimierungsproblem dar, welches über den Sattelpunkt eines Lagrange-Funktionalen L ,

$$L(w, b, \alpha) = \frac{1}{2} \cdot w^T \cdot w - \sum_{i=1}^n \alpha_i \cdot [y_i \cdot (x_i \cdot w + b) - 1] \quad (5.71)$$

mit

n := Anzahl Beobachtungen im Datensatz X

α_i := Lagrange-Multiplikator der Beobachtung i,

gelöst werden kann [KECMAN, 2005].

Dabei werden zunächst die unbekannten α_i durch Maximierung eines dualen Lagrange-Funktionalen L_d ,

$$L_d(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \cdot \sum_{i=1}^n \sum_{z=1}^n \alpha_i \cdot \alpha_z \cdot y_i \cdot y_z \cdot x_i \cdot x_z^T \quad (5.72)$$

mit

$z :=$ weiterer Zählindex der Objekte, $z = 1, 2, \dots, n$,

und mit den Nebenbedingungen,

$$\sum_{i=1}^n \alpha_i \cdot y_i = 0 \quad (5.73)$$

und

$$\alpha_i \geq 0, \quad (5.74)$$

aus den Trainingsdaten identifiziert.

Bei allen Beobachtungen mit einem $\alpha_i > 0$ handelt es sich um Stützvektoren. Alle übrigen Objekte haben demnach ein α_i von 0.

Die gesuchten Parameter der optimalen Trennebene, der Wichtungsvektor w ,

$$w = \sum_{i=1}^n \alpha_i \cdot y_i \cdot x_i^T, \quad (5.75)$$

und der Bias b ,

$$b = \frac{1}{n_{sv}} \cdot \sum_{i=1}^{n_{sv}} (y_{svi} - x_{svi} \cdot w) \quad (5.76)$$

mit

n_{sv} := Anzahl der Stützvektoren,

lassen sich dann berechnen.

Anhand Gl. (5.72) ist zu erkennen, dass bei der Berechnung der optimalen Trennebene die Eingangsdaten X in Form von Skalarprodukten eingehen.

Damit hat die Dimensionalität (Anzahl der Variablen m) von X keinen starken Einfluss auf den benötigten Rechenaufwand. Dieser ist dagegen maßgeblich von der Anzahl n an Beobachtungen im Kalibrierdatensatz abhängig, da ebenso viele Lagrange-Multiplikatoren α_i identifiziert werden müssen. *Support Vector Machines* sind demnach besonders für hochdimensionale Daten mit verhältnismäßig wenigen Objekten geeignet.

Für die Identifizierung der α_i kommen spezielle Algorithmen zur Lösung des quadratischen Optimierungsproblems zum Einsatz. Häufig verwendet wird der *Sequential Minimal Optimization* (SMO) Algorithmus [PLATT, 1999].

5.7.3 Kernel-Funktionen zur Abbildung nichtlinearer Beziehungen

In Bild 5.11a sind zwei Klassen A und B mit einer nichtlinearen Klassengrenze in einem Raum zweier X-Variablen gezeigt. In einem solchen Fall ist die direkte Anwendung eines linearen Ansatzes für die Trennebene ungeeignet.

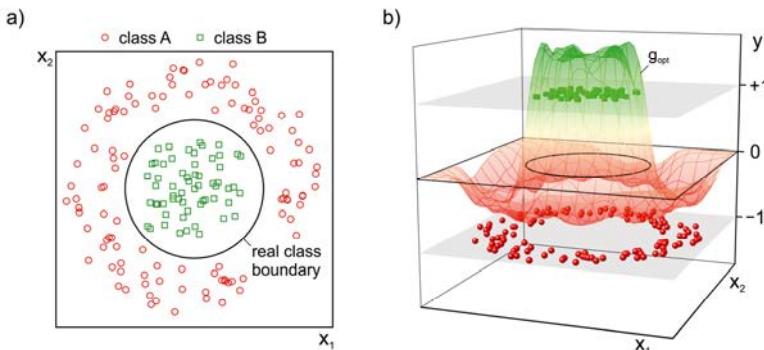


Bild 5.11: a) Zwei Klassen mit nichtlinearer Klassengrenze für $m = 2$ Dimensionen
b) Entscheidungsfunktion der Klassifizierung als nichtlineare Oberfläche

Durch mathematische Transformationen der Eingangsdaten X lässt sich dieser ($n \times m$) dimensionale Datenraum in einem Merkmalsraum F ($n \times M$) abbilden. Die Anzahl M der Variablen (Merkmale) dieses Raums kann dabei beliebig hoch gewählt werden.

Die Idee ist, dass sich durch die Wahl einer geeigneten Abbildungsvorschrift Φ ein linearer Ansatz der Entscheidungsfunktion g_F ,

$$g_F(X) = \Phi(X) \cdot w_F + b_F, \quad (5.77)$$

auf den Raum F anwenden lässt, wodurch der zur Klassentrennung benötigte nichtlineare Verlauf von g_F erzeugt wird. Dies ist in Bild 5.11b illustriert.

Das für die Lösung dieses Problems zu maximierende duale Lagrange-Funktional L_d ,

$$L_d(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \cdot \sum_{i=1}^n \sum_{z=1}^n \alpha_i \cdot \alpha_z \cdot y_i \cdot y_z \cdot \Phi(x_i) \cdot \Phi^T(x_z), \quad (5.78)$$

beinhaltet analog zu Gl. (5.72) die Skalarprodukte der in F abgebildeten Objekte $\Phi(x_i)$.

Bei sehr hoher Dimension M des Merkmalsraumes benötigt die Berechnung dieser Skalarprodukte (sowie die Ausführung der Abbildung Φ selbst) einen sehr hohen Rechenaufwand.

Dieser wird durch die Verwendung sogenannter *Kernel-Funktionen* K ,

$$K(x_i, x_z) = \Phi(x_i) \cdot \Phi^T(x_z), \quad (5.79)$$

umgangen, welche ausschließlich mit den Eingangsdaten in X operieren.

Dadurch wird sogar die Verwendung, unendlichdimensionaler Eigenschaftsräume ermöglicht, die beispielsweise durch die Gauß'sche radiale Basisfunktion (RBF),

$$K(x_i, x_z) = e^{-\frac{1}{\gamma} \cdot \|x_i - x_z\|^2} = e^{-\frac{1}{\gamma} \cdot (x_i - x_z) \cdot (x_i - x_z)^T} \quad (5.80)$$

mit

γ := Parameter des Gauß'schen RBF-Kernels,

erzeugt werden.

Bei dieser muss der Parameter γ über eine Kreuzvalidierung ermittelt werden. Für stark verrauschte Daten erzeugt ein kleiner Wert für γ eine komplexe Trennebene und kann damit zu einem *Overfit* führen [KECMAN, 2005].

Es existieren verschiedene *Kernel*-Funktionen. Einige sind in Tabelle 5.1 aufgelistet. In dieser Arbeit wurde jedoch ausschließlich mit dem Gauß'schen RBF-*Kernel* gearbeitet.

Tabelle 5.1: Häufig verwendete *Kernel*-Funktionen [KECMAN, 2005]

Kernel-Funktion	Beschreibung
$K(x_i, x_z) = x_i \cdot x_z^T$	Linearer Kernel
$K(x_i, x_z) = (x_i \cdot x_z^T + 1)^d$	Polynom vom Grad d
$K(x_i, x_z) = \tanh(x_i \cdot x_z^T + c)$	Mehrlagiges Perzeptron
$K(x_i, x_z) = (\ x_i - x_z\ ^2 + \beta)^{-\frac{1}{2}}$	Inverse multiquadratische Funktion

Bei der Verwendung nichtlinearer *Kernel* ist eine explizite Darstellung des Wichtungsvektors w nicht mehr möglich. Stattdessen kann die Entscheidungsfunktion g ,

$$g(x) = \hat{y} = \sum_{i=1}^n \alpha_i \cdot y_i \cdot K(x_i, x), \quad (5.81)$$

direkt mit den Trainingsdaten und der verwendeten *Kernel*-Funktion notiert werden.

Abhängig von der Art des verwendeten *Kernels* kann der Bias b entweder vernachlässigt oder implizit in der *Kernel*-Funktion berücksichtigt werden [KECMAN, 2005].

5.7.4 Erweiterung zur Support Vector Regression (SVR)

Das prinzipielle Vorgehen bei der Entwicklung eines Regressionsverfahrens mittels *Support Vector Machines* entspricht dem bei der Entwicklung eines Klassifikators.

Ausgehend von einer linearen Beziehung, wird eine Regressionsfunktion f ,

$$f(X) = \hat{y} = X \cdot w + b, \quad (5.82)$$

allgemein als linearer Zusammenhang im multivariaten Raum X definiert.

Jede Beobachtung i der Trainingsdaten besitzt anstelle diskreter Werte der Klassenzugehörigkeit jedoch einen kontinuierlichen, autoskalierten Messwert y_i .

Für die Berechnung der Regressionsfunktion soll über das Gütekriterium J_f ,

$$J_f = \frac{1}{2} \cdot w^T \cdot w + C \cdot \sum_{i=1}^n E(x_i, y_i) \quad (5.83)$$

mit

C := Güteparameter zur Wichtung der Fehler, $C > 0$,

auch die Summe der Kalibrierfehler minimiert werden.

Der Fehler E einer Beobachtung i ,

$$E(x_i, y_i) = \begin{cases} |y_i - f(x_i)| - \varepsilon, & |y_i - f(x_i)| < \varepsilon \\ 0, & \text{otherwise} \end{cases} \quad (5.84)$$

mit

ε := akzeptierte Fehlertoleranz,

beschreibt die absolute Abweichung von der Regressionsfunktion unter Berücksichtigung einer Fehlertoleranz ε .

Diese muss zusammen mit dem Parameter C zur Wichtung der Kalibrierfehler vom Anwender gewählt oder in einer Kreuzvalidierung ermittelt werden. Kleine Werte von ε sowie große Werte von C erzeugen ein komplexes Modell und können somit zu einem *Overfit* führen.

Das im Falle der Regression zu lösende quadratische Optimierungsproblem führt zur einem dualen Lagrange-Funktional L_d ,

$$\begin{aligned} L_d(\alpha_i, \beta_i) = & -\frac{1}{2} \cdot \sum_{i=1}^n \sum_{z=1}^n (\alpha_i - \beta_i) \cdot (\alpha_z - \beta_z) \cdot x_i \cdot x_z^T \\ & - \sum_{i=1}^n (\varepsilon - y_i) \cdot \alpha_i - \sum_{i=1}^n (\varepsilon + y_i) \cdot \beta_i, \end{aligned} \quad (5.85)$$

welches mit den Nebenbedingungen,

$$\sum_{i=1}^n (\alpha_i - \beta_i) = 0 \quad (5.86)$$

und

$$C \geq \alpha_i, \beta_i \geq 0, \quad (5.87)$$

durch Identifikation der Lagrange-Multiplikatoren α_i und β_i zu maximieren ist.

Im Vergleich mit einem SVM-Klassifikator müssen bei der SVR demnach doppelt so viele Parameter bestimmt werden.

Da die Trainingsdaten x_i auch hier als Skalarprodukt in L_d eingehen (siehe Gl. (5.85)), können auch bei der Regression *Kernel*-Funktionen zur Modellierung nichtlinearer Beziehungen ohne weitere Maßnahmen eingesetzt werden.

Beobachtungen mit einem α_i oder $\beta_i > 0$ sind Stützvektoren. Für alle anderen Objekte sind beide Lagrange-Multiplikatoren gleich 0.

Der Wichtungsvektor w der Regressionsfunktion,

$$w = \sum_{i=1}^n (\alpha_i - \beta_i) \cdot x_i^T, \quad (5.88)$$

wird deshalb nur von den Stützvektoren beeinflusst.

Diese liegen entweder oberhalb ($\alpha_i > 0$) oder unterhalb ($\beta_i > 0$) des Toleranzbandes der Regressionsfunktion, welches eine Weite von 2ϵ besitzt. Bild 5.12 zeigt Stützvektoren und ein ϵ -Toleranzband bei der Modellierung eines nichtlinearen Verlaufes mit SVR für eine X-Variable.

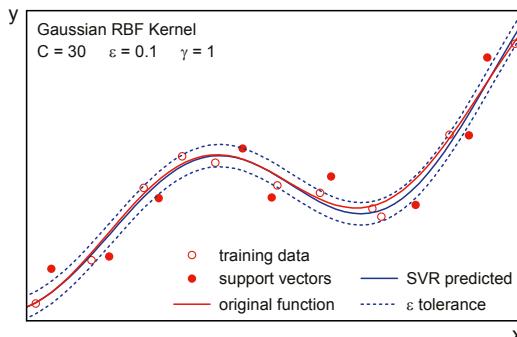


Bild 5.12: Mit SVR modellierter, nichtlinearer Verlauf ($m = 1$)

Stützvektoren, dessen Lagrange-Multiplikatoren einen Wert $< C$ aufweisen, werden freie Stützvektoren genannt. Alle anderen heißen gebundene Stützvektoren.

Die Berechnung des Bias b der Regressionsfunktion,

$$b = \frac{1}{n_{fsv}} \sum_{i=1}^{n_{fsv}} \begin{cases} y_i - x_i \cdot w - \epsilon, & C > \alpha_i > 0 \\ y_i - x_i \cdot w + \epsilon, & C > \beta_i > 0 \end{cases} \quad (5.89)$$

mit

n_{fsv} := Anzahl freier Stützvektoren,

erfolgt durch Mittelung über diese freien Stützvektoren [KECMAN, 2005].

5.8 Eingesetzte MVDA-Software

In dieser Arbeit wurden zwei Softwarepakete für die Erstellung multivariater Modelle genutzt. Die Software SIMCA 13.0.3 der Fa. MKS data analytics solutions (ehemals Umetrics), Umeå, Schweden, wurde für die Berechnung von PLS-Regressionsmodellen und SIMCA-Klassifikationsmodellen verwendet.

Für die Berechnung von Modellen, basierend auf *Support Vector Machines* (SVR), kam die PLS_Toolbox 8.1.1 von Eigenvector Research, Manson, USA, zum Einsatz. Hierbei handelt es sich um eine Zusatzsoftware für MATLAB®, The Mathworks, Natick, USA, das in der Version 8.5.0/R2015a zur Verfügung stand. SVR-Modelle werden hierin mit einer Implementierung der LIBSVM-Bibliothek berechnet [CHANG & LIN, 2011].

6. Qualitätsbewertung von Hefeextrakten mit NIR-Spektroskopie

6.1 Motivation und Zielsetzung

Hefeextrakt wird zumeist aus dem Lysat von Zellen der Spezies *Saccharomyces cerevisiae* gewonnen. Durch die Wahl und Steuerung des Aufschlussverfahrens (z.B. Autolyse, Thermolyse) können unterschiedliche Produktklassen gewonnen werden, die sich in ihrer Zusammensetzung unterscheiden. Nach Abtrennung unlöslicher Zellbestandteile wird das Extrakt in flüssiger oder in sprühgetrockneter Pulverform weiterverarbeitet.

Ein Einsatzgebiet von Hefeextrakt ist die biotechnologische Produktion, in der es als Bestandteil von Kulturmedien zum Einsatz kommt. Neben verschiedenen Kohlenstoffquellen enthält es unter anderem Aminosäuren und Spurenelemente, womit ein breites Spektrum des Nährstoffbedarfes verschiedener Mikroorganismen abgedeckt wird.

Hefeextrakt kann damit das Wachstum aber auch die Produktivität verschiedener biotechnologischer Hostsysteme positiv beeinflussen [KEIL & TILKINS, 2013; MOSSER et al., 2013]. Eine Maximierung mikrobiellen Wachstums konnte mit Variation des Extraktes oder durch die Verwendung einer optimalen Mischung verschiedener Klassen erreicht werden [WILMES, 2012].

Der Vertrieb solcher stamm- oder produktspezifisch optimierter Mischungen könnte ein lohnendes Geschäftsfeld kleinerer Hefeextraktproduzenten sein, die sich auf Kunden aus dem biotechnologischen Umfeld fokussieren.

Ein Hindernis bei der Herstellung derartig spezialisierter Produkte ist die natürliche Chargenschwankung, die bei der Produktion der Grundklassen beobachtet wird. Hieraus kann chargenabhängig unterschiedliches mikrobielles Wachstum resultieren [WILMES, 2012].

Das Ziel dieser Arbeit war die Entwicklung eines Bewertungsverfahrens für drei Hefeextraktklassen. Als Qualitätsmerkmal wurde dabei die Übereinstimmung einer Charge mit den mittleren Eigenschaften ihrer entsprechenden Klasse herangezogen. Der Ausschluss abweichender Chargen bei der Herstellung spezialisierter Mischungen würde deren Qualität gewährleisten.

Die Durchführung umfangreicher Analysen, hinsichtlich relevanter Komponenten der Hefeextrakte, stellt eine sehr zeit- und kostenintensive Aufgabe dar. Darüber hinaus sind viele Inhaltsstoffe unbekannt oder sind aufgrund des komplexen Charakters der Hefeextrakte nur schwer isolier- und quantifizierbar.

Es konnte gezeigt werden, dass Nahinfrarotspektren von Hefeextrakten mit dem Kultivierungserfolg von Mikroorganismen in Beziehung gesetzt werden können [KASPROW et al., 1998]. Daher wurde auf die Bestimmung spezifischer Inhaltsstoffe verzichtet und als Analysemethode einzig die Nahinfrarotspektroskopie (NIRS) gewählt. Diese hat den Vorteil schnell und zerstörungsfrei zu arbeiten und kaum Probenvorbereitung zu benötigen [BLANCO & VILLARROYA, 2002].

Für die Entwicklung der Qualitätsbewertung basierend auf NIRS kam in dieser Arbeit das multivariate Klassifizierungsverfahren *Soft Independent Modelling of Class Analogy* (SIMCA) zum Einsatz (siehe Abschnitt 5.6.2).

6.2 Stand der Wissenschaft

Multivariate Klassifizierungsverfahren wurden zusammen mit der NIRS bereits erfolgreich eingesetzt. Insbesondere eine Qualitätsbewertung von Nahrungsmitteln wurde auf diese Art bereits mehrfach realisiert [BERRUETA et al., 2007; HUANG et al., 2008; WOODCOCK et al., 2008].

Hier kann auf einen breiten Erfahrungsschatz zurückgegriffen werden, der sich in einer hohen Anzahl an Veröffentlichungen niederschlägt. Die Methodik wurde zum Beispiel zur Authentizitätsprüfung von Weinen [ARVANITOYANNIS et al., 1999], zur Qualitätsbewertung von Weizenmehlen [COCCHE et al., 2005], oder der schnellen Detektion von Aflatoxin B1 in Getreiden [FERNÁNDEZ-IBÁÑEZ et al., 2009] eingesetzt.

Die Untersuchung von Rohmaterialien im pharmazeutischen Umfeld mit NIR-Spektroskopie stellt ein klassisches Beispiel einer PAT-Anwendung dar [SKIBSTED & ENGELSEN, 2010]. Dies wurde jedoch auch schon vor der Veröffentlichung der PAT-Initiative erfolgreich praktiziert. Gemperline et al. nutzen bereits 1989 NIR-Reflexionsspektren mit der Methode SIMCA, um die Eignung verschiedener Rohmaterialien bei der Herstellung von Tabletten zu bewerten [GEMPERLINE et al., 1989].

Für die Zeit nach Erscheinen der PAT-Initiative können jedoch vergleichsweise wenig Literaturstellen aus dem biopharmazeutischen Sektor gefunden werden. Kirdar et al. nutzten NIRS und PCA, um den Erfolg von Zellkulturen auf die Chargenschwankungen der pulverförmigen Zellkulturmedien zurückzuführen [KIRDAR et al., 2009]. Märk et al. konnten die Qualität eines Zwischenproduktes bei der Antibiotikaproduktion mittels NIRS überwachen [MÄRK et al., 2010].

6.3 Der gewählte Messaufbau

Die pulverförmigen Hefeextraktproben wurden in einem speziellen Gefäß vorgelegt und auf dem Kontaktmesskopf PSS-H-B01 der Fa. Polytec, Waldbronn, wie in Bild 6.1 gezeigt, positioniert. Die Messung erfolgte mit der Methode der diffusen Reflexion.



Bild 6.1: Messaufbau mit dem Reflexionskontaktmesskopf PSS-H-B01

Das Probengefäß wurde individuell von der Fa. Quarzglas Komponenten und Service QCS, Maintal, angefertigt. Es verfügt über einen plangeschliffenen Boden aus Quarzglas und ist damit für Anwendungen mit Nahinfrarotspektroskopie geeignet.

Für eine Messung wurden etwa 10 g Hefeextrakt pulver benötigt. Die Vermessung der Proben erfolgte in Dreifachbestimmung. Zwischen diesen Messungen wurde das Gefäß grob gesäuert und die Probe neu vorgelegt. Zwischen der Vermessung unterschiedlicher Proben erfolgte eine gründliche Reinigung.

Die Messungen fanden bei den Standardmesseinstellungen des Spektrometers statt. Die Integrationszeit betrug 15 ms und es wurden 100 Akkumulationen aufgenommen. Eine Dunkelkorrektur und der Weißabgleich erfolgten automatisch über Schließmechanismen des Detektors und des Kontaktmesskopfes.

6.4 Vorstellung des Probenpools

Das Qualitätsbewertungsverfahren sollte für die drei Hefeextraktklassen A, B und C eines Herstellers entwickelt werden. Klasse A ist von einem hohen Gehalt freier Aminosäuren gekennzeichnet, während die Klassen B und C einen hohen Peptidgehalt aufweisen. Klasse B weist zudem einen relativ hohen Gehalt an Nukleotiden auf.

Es standen 93 Hefeextraktproben zur Verfügung. Wie in Tabelle 6.1 dargestellt, teilen sich diese zu je 32 auf die Klassen A und C auf. Auf die Klasse B entfallen 26 Proben. Zusätzlich wurden drei externe Proben vermessen, die bei der Modellvalidierung als Negativbeispiele verwendet werden sollten. Bei diesen Proben handelt es sich um kommerziell erhältliche Hefeextrakte anderer Hersteller.

Tabelle 6.1: Probenpool für die Entwicklung von NIR- und MVDA-Bewertungsverfahren

Klasse	Probenanzahl	Einzelmessungen
A	32	96
B	26	78
C	32	96
extern	3	9
Gesamt	93	279

6.5 Vorverarbeitung der Spektraldaten

Als Methode der Datenvorverarbeitung wurde das *Standard Normal Variate* (SNV) Filter gewählt. Außerdem erfolgte eine Mittenzentrierung der NIR-Spektren.

Das SNV-Filter ist deshalb für die Spektren der pulverförmigen Hefeextrakte geeignet, da insbesondere die bei Messungen in diffuser Reflexion an partikulärem Probematerial auftretenden Streueffekte von dieser Methode kompensiert werden [KESSLER, 2008]. Die Überlegenheit des SNV-Filters gegenüber anderen Methoden in ähnlichen Anwendungen der NIR-Spektroskopie wurde bereits gezeigt [CANDOLFI et al., 1999a; CHEN et al., 2006].

6.6 Explorative Datenanalyse und Probenselektion

Zum Zwecke der Analyse der Hefeextraktspektren wurde zunächst ein auf allen Proben basierendes PCA-Modell erstellt. Dieses Modell besteht aus zwei Hauptkomponenten, die gemeinsam eine Varianz der Spektraldaten R_x^2 von 92,1 % beschreiben.

In dem für diese beiden Hauptkomponenten in Bild 6.2 gezeigten Score Scatter Plot ist bereits eine gute Trennung der Klasse A (rot) von den anderen beiden Klassen zu erkennen. Die Klassen B (blau) und C (grün) weisen eine gewisse Überlappung auf. Die als graue Dreiecke gekennzeichneten externen Proben zeigen zum Teil Ähnlichkeiten mit den Klassen B und C, liegen aber eher abseits der Hauptgruppen.

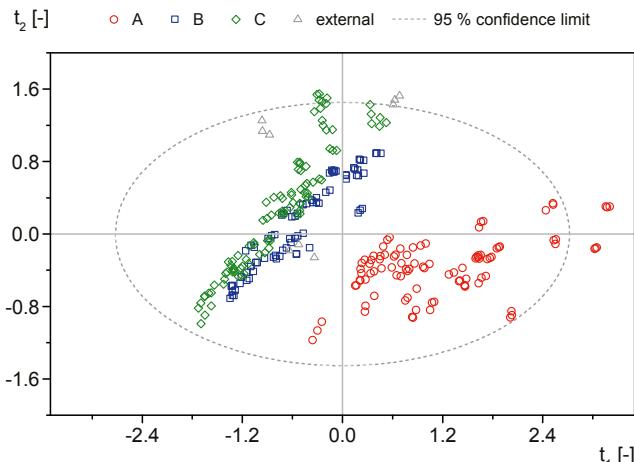


Bild 6.2: Score Scatter Plot des PCA-Modells aller Hefeextraktproben

Kleinere Gruppen von jeweils drei Symbolen entsprechen den Dreifachmessungen der Proben. Einige ungewöhnliche Objekte, die relativ weit von der Hauptgruppe entfernt positioniert sind, können insbesondere der Klasse A zugeordnet werden.

Ein multivariates Klassifizierungsmodell für eine Qualitätsbewertung sollte auf Proben basieren, die den mittleren Eigenschaften der Klasse entsprechen. Aus diesem Grund wurden in einem nächsten Schritt PCA-Modelle der einzelnen Klassen erzeugt und diese auf Ausreißer untersucht.

Für die PCA-Klassenmodelle wurden jeweils so viele Hauptkomponenten berechnet, um mindestens 95 % der Varianz in den Spektraldaten zu beschreiben. Diese Grenze wurde frei gewählt. Wies eine Probe einen Hotelling T^2 -Wert außerhalb des 95 % Konfidenzintervalls auf, so wurde diese Probe als Ausreißer erachtet und folglich von der Modellbildung ausgeschlossen.

Bild 6.3a zeigt den Score Scatter Plot und Bild 6.3b den Hotelling T^2 Plot eines PCA-Modells mit zwei Hauptkomponenten für alle 32 Proben der Klasse A.

Es ist eine relativ kompakte Gruppe in der Mitte des Score Scatter Plots zu erkennen, von denen sich einige ungewöhnliche Proben abheben. Diese sind durch ihre Probennummer als potentielle Ausreißer kenntlich gemacht.

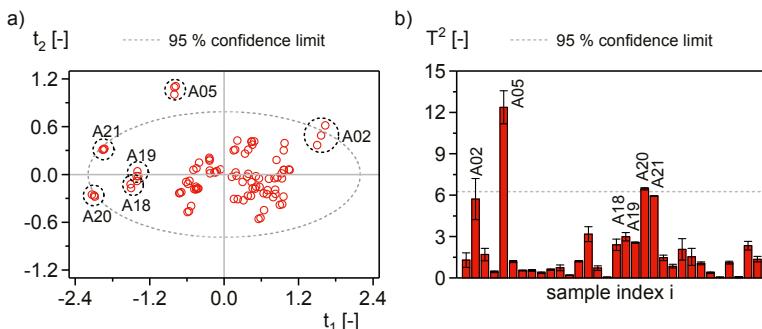


Bild 6.3: PCA-Modell der Klasse A zur Ausreißerdetektion, $n = 96$ (32), $r = 2$,
a) Score Scatter Plot, b) Hotelling T^2 Plot

Allerdings konnte nur die Probe A05 als eindeutiger Ausreißer identifiziert werden, da die anderen markierten Proben innerhalb oder auf der Vertrauensgrenze liegen. Folglich wurde zunächst nur diese Probe von der Modellbildung ausgeschlossen und ein neues PCA-Modell gebildet.

Basierend auf diesem aktualisierten Modell konnten zwei weitere Proben als Ausreißer identifiziert werden. Dieses Vorgehen wurde mit dem weiter reduzierten Datensatz so oft wiederholt, bis keine Probe mehr als Ausreißer zu erkennen war und eine homogene Verteilung der Proben im Score Scatter Plot erreicht wurde. Das endgültige Resultat mit 26 von 32 Proben ist in Bild 6.4 gezeigt.

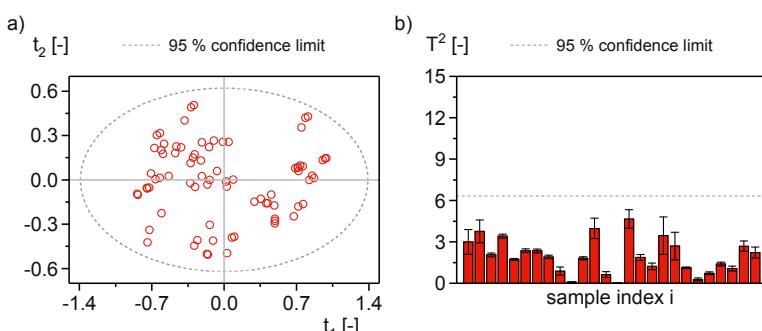


Bild 6.4: PCA-Modell der Klasse A ohne Ausreißer, $n = 78$ (26), $r = 2$,
a) Score Scatter Plot, b) Hotelling T^2 Plot

Alle sechs in Bild 6.3 als potentielle Ausreißer hervorgehobenen Proben wurden auch tatsächlich als Ausreißer erkannt und von der Bildung des multivariaten Klassifizierungsmodells für Klasse A ausgeschlossen.

Das gleiche Vorgehen für die Klassen B und C führte zu der Erkennung von zwei bzw. fünf Ausreißern.

6.7 Entwicklung eines SIMCA-Klassifizierungsmodells

Ein SIMCA-Klassifizierungsmodell besteht aus je einem PCA-Modell für jede der betrachteten Klassen und bestimmten Entscheidungsregeln für die Klassifizierung. Diese Entscheidungsregeln können im Voraus festgelegt werden. Die Konfidenzintervalle für den Hotelling T^2 -Wert und des DmodX+ wurden zu 95 % gewählt. Die Grenzwahrscheinlichkeit β_{crit} für die Einteilung in eine Klasse betrug 0,5.

Die jeweilige Anzahl erforderlicher Hauptkomponenten der PCA-Klassenmodelle muss dagegen gesondert ermittelt werden. Die Schwierigkeit dieser Aufgabe ist bekannt, da keine generellen Regeln gegeben sind [CANDOLFI et al., 1999b]. Häufig kommen hierbei Kreuzvalidierungsverfahren, angewandt auf die einzelnen PCA-Modelle, zum Einsatz.

Dieses Vorgehen optimiert eine generelle Beschreibung der Klassen-Spektraldaten durch die einzelnen PCA-Modelle, nicht aber die Güte eines vollständigen SIMCA-Klassifikators. In dieser Arbeit wurde deshalb ein Ansatz verfolgt, bei dem das gesamte SIMCA-Modell einer Kreuzvalidierung unterzogen wird. Dabei stellt der Klassifizierungsfehler E_C (siehe Abschnitt 5.6.3) das entscheidende Kriterium dar.

Es wurden vier Kreuzvalidierungs durchgänge durchgeführt und sichergestellt, dass jede Probe nur einmal als Teil eines Validierdatensatzes zum Einsatz kam sowie keine Trennung der Dreifachmessungen erfolgte. Zuvor als Ausreißer deklarierte Proben (siehe Abschnitt 6.6) waren ebenfalls genau einmal Teil einer der sich ergebenen vier Validierdatensätze. Die drei externen Proben wurden bei diesem Vorgehen nicht berücksichtigt.

Diese Kreuzvalidierungsprozedur wurde für acht SIMCA-Klassifizierungsmodelle durchgeführt, in denen die drei PCA-Klassenmodelle jeweils die gleiche Anzahl an Hauptkomponenten beinhalten ($r = 1 \dots 8$).

Die Klassifizierungsergebnisse jedes dieser acht Modelle lassen sich in je einer Konfusionsmatrix für die drei Klassen zusammenfassen. Diese sind am Beispiel des SIMCA-Modells mit jeweils $r = 1$ in Tabelle 6.2 gezeigt.

Tabelle 6.2: Konfusionsmatrizen der drei Klassen eines SIMCA-Modells mit $r = 1$ für alle PCA-Klassenmodelle, Auswertung aller Einzelmessungen

	Klasse A		Klasse B		Klasse C	
positiv eingeteilt	pos	neg	pos	neg	pos	neg
ja	78	0	66	30	75	10
nein	0	192	6	168	6	179
$\Sigma (n = 270)$	78	192	72	198	81	189

Den Konfusionsmatrizen lässt sich entnehmen, dass für die Klasse A bereits ein perfekter Klassifikator mit nur einer Hauptkomponente erzeugt werden konnte. Alle Proben (78 Messungen) der Klasse A wurden korrekt in diese eingeteilt. Darüber hinaus wurde keine der klassenfremden Proben (192 Messungen) fälschlicher Weise der Klasse A zugewiesen. Im Gegensatz dazu weisen die Klassen B und C für dieses SIMCA-Modell noch einige Fehlklassifizierungen auf.

Beispielsweise wurden je zwei Proben (sechs Messungen) nicht ihrer jeweiligen Klasse zugewiesen. Für Klasse B handelt es sich hierbei um die Proben B02 und B06 (siehe Bild 6.5).

Weiterhin wurden der Klasse B fälschlicher Weise zehn negative Proben (30 Messungen) zugewiesen. Bild 6.5 kann entnommen werden, dass diese Proben ausschließlich Vertreter der Klasse C sind. Hierzu sind die entsprechenden Proben rechts oben im Bild anhand ihrer Probennummer hervorgehoben.

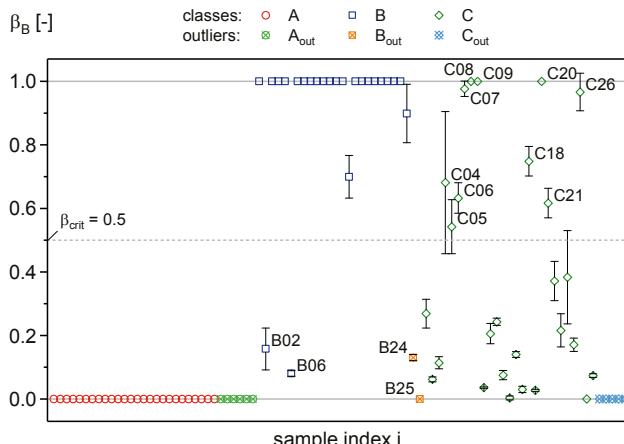


Bild 6.5: Wahrscheinlichkeit β_B der Zugehörigkeit zur Klasse B aller Proben für ein SIMCA-Modell mit $r = 1$ für alle PCA-Klassenmodelle

In obiger Abbildung sind die Wahrscheinlichkeiten β_B der Zugehörigkeit zur Klasse B für alle Proben dargestellt. Bei Überschreiten der gewählten Grenzwahrscheinlichkeit β_{crit} von 0,5 werden die Proben in Klasse B eingeteilt. Viele Proben der Klasse C (grün) befinden sich oberhalb dieser Grenze. Die beiden Ausreißer der Klasse B, B24 und B25 (orange), wurden dagegen korrekt als nicht klassenzugehörig erkannt.

Analog dazu stammen die zehn falsch-positiven Zuweisungen der Klasse C (siehe Tabelle 6.2) ausschließlich aus Klasse B (nicht gezeigt). Dies lässt darauf schließen, dass eine Hauptkomponente nicht zur Trennung der Klassen B und C ausreicht, da diese starke Ähnlichkeiten aufweisen. Dieser Eindruck entstand bereits bei Betrachtung des Score Scatter Plots des Gesamt-PCA-Modells in Bild 6.2.

Ein übersichtlicherer Vergleich der getesteten SIMCA-Klassifikatoren mit unterschiedlicher Komponentenanzahl kann anhand der in Abschnitt 5.6.3 vorgestellten Charakteristiken erfolgen. Hierzu sind in Bild 6.6 die Selektivität Se und die Spezifität Sp für die drei Klassen in Abhängigkeit der Anzahl r verwendeter Hauptkomponenten gezeigt.

In Bild 6.6a ist zu erkennen, dass die Selektivität Se für alle Klassen mit steigender Komponentenanzahl abnimmt. Dies bedeutet, dass ein sinkender Anteil an klassenzugehörigen Proben korrekt in die entsprechende Klasse eingeteilt wird. Dieses Verhalten deutet auf einen Overfit der PCA-Klassenmodelle bei hohen Komponentenanzahlen hin.

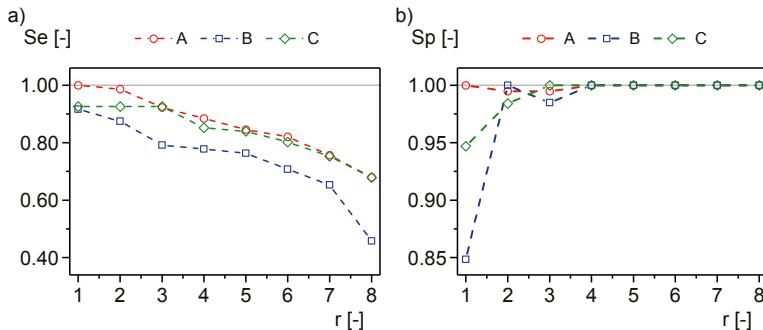


Bild 6.6: Bewertung der Klassifizierung anhand a) der Selektivität Se und b) der Spezifität Sp der drei Klassen abhängig von der Anzahl r berechneter Hauptkomponenten

Die Spezifität Sp , gezeigt in Bild 6.6b, steigt in der Tendenz für die Klassen B und C mit der Komponentenanzahl r an und liegt für alle Klassen ab vier Hauptkomponenten auf ihrem Maximalwert 1. Für viele berechnete Hauptkomponenten ist es also sehr unwahrscheinlich, Proben in die falsche Klasse einzuteilen.

Vor dem Hintergrund der angestrebten Entwicklung eines Qualitätsbewertungsverfahrens sollten, um falsch-positive Ergebnisse zu vermeiden, Klassifikatoren mit hoher Spezifität gewählt werden. Auf eine maximale Selektivität kann dagegen verzichtet werden, da bei der vorliegenden Fragestellung und unter Vernachlässigung ökonomischer Erwägungen, falsch-negative Einteilungen eher toleriert werden können.

Die Ergebnisse lassen sich für die Festlegung der optimalen Anzahl an Hauptkomponenten anhand des Klassifizierungsfehlers E_C zusammenfassen, der in Bild 6.7 in Abhängigkeit der Anzahl berechneter Hauptkomponenten gezeigt ist.

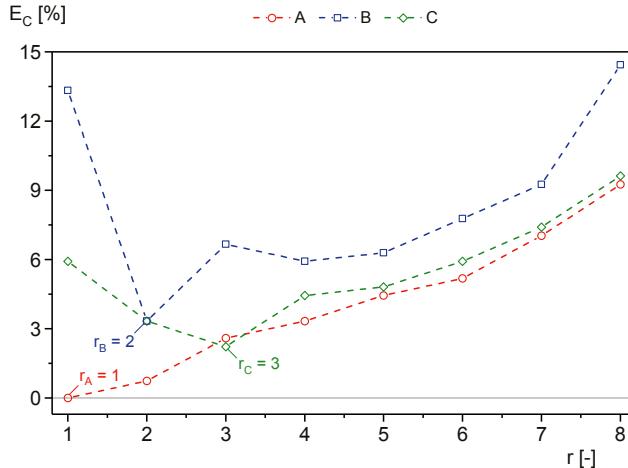


Bild 6.7: Klassifizierungsfehler E_C der drei Klassen in Abhängigkeit von r

Gemäß der durchgeführten Kreuzvalidierungsprozedur haben die Klassen A (1), B (2) und C (3) eine unterschiedliche Hauptkomponentenanzahl r für minimale Klassifizierungsfehler E_c des Gesamtmodells. Folglich wurden diese Werte gewählt, damit der SIMCA-Klassifikator optimale Ergebnisse erzielen kann. Tabelle 6.3 zeigt die finalen Konfusionsmatrizen des gewählten SIMCA-Modells.

Tabelle 6.3: Konfusionsmatrizen der drei Klassen des finalen SIMCA-Modells mit $r_A = 1$, $r_B = 2$, $r_C = 3$, Auswertung aller Einzelmessungen

	Klasse A		Klasse B		Klasse C	
	pos	neg	pos	neg	pos	neg
positiv eingeteilt						
ja	78	0	63	0	75	0
nein	0	192	9	198	6	189
$\Sigma (n = 270)$	78	192	72	198	81	189

Für die Klasse A bleibt der perfekte Klassifikator erhalten. Von Klasse B wurden dagegen drei Proben (neun Messungen) und von Klasse C zwei Proben (sechs Messungen) als nicht klas-senzugehörig erkannt. Wie bereits erwähnt ist dieses Verhalten jedoch akzeptabel. Erfreulich ist, dass für die Klassen B und C, trotz der beobachteten Gemeinsamkeiten, keine falsch-positiven Einteilungen mehr auftreten.

6.8 Externe Validierung des Modells

Zur abschließenden Bewertung des im vorherigen Abschnitt entwickelten SIMCA-Klassifikators soll eine externe Validierung durchgeführt werden. Da für die Arbeiten kein weiteres als das bereits beschriebene Probenmaterial zur Verfügung stand, wurden Kalibrier- und Validier-datensätze aus den vorhandenen Daten mit Hilfe des Kennard-Stone-Algorithmus erzeugt.

Dieser nutzt die euklidische Länge von Vektoren, um uniform verteilte Objekte mit den größten Abständen in einem Datenraum nacheinander für den Kalibrierdatensatz einer multivariaten Modellbildung auszuwählen. Der Algorithmus arbeitet sequentiell und wird zweckmäßig nach einer vorher festgelegten Anzahl an Objekten gestoppt [KENNARD & STONE, 1969].

Die Aufteilung der Proben mit dem Kennard-Stone-Algorithmus erfolgte separat für jede Klas-se mit den vorverarbeiteten sowie von Ausreißern bereinigten Daten. Ca. 60 % der Proben wurden dabei für den Kalibrierdatensatz ausgewählt und Dreifachmessungen nicht getrennt.

Die übrigen Proben sind dann positive Exemplare des Validierdatensatzes. Als Negativbe-i-spiele enthält dieser zusätzlich die in Abschnitt 6.6 detektierten Ausreißer der entsprechenden Klasse sowie alle Proben der anderen Klassen und die drei Extrakte exzerner Hersteller.

Das Resultat des Kennard-Stone-Algorithmus ist in Bild 6.8 anhand eines *Score Scatter Plots* eines PCA-Modells der Klasse A gezeigt. In roter Farbe sind die algorithmisch bestimmten Proben des Kalibrierdatensatzes dargestellt. Diese spannen den Datenraum gleichmäßig auf. Positive Vertreter des Validierdatensatzes (grau) liegen verteilt dazwischen.

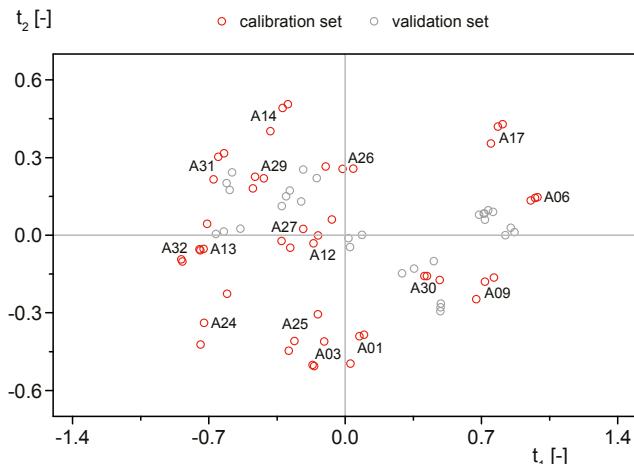


Bild 6.8: Verteilung von Proben der Klasse A auf Kalibrier- und Validierdatensätze durch den Kennard-Stone-Algorithmus

Basierend auf dem oben vorgestellten Kalibrierdatensatz wurde ein neues SIMCA-Klassifizierungsmodell gemäß der im vorherigen Abschnitt ermittelten optimalen Einstellungen erzeugt und mit dem Validierdatensatz überprüft. Die Ergebnisse sind in Tabelle 6.4 zusammengefasst.

Tabelle 6.4: Kenngrößen des finalen SIMCA-Klassifikators für die externe Validierung

Klasse	r	Se [-]	Sp [-]	E _C [%]	n _{CS}	n _{VS} (positiv)	n _{VS} (negativ)
A	1	1	0,99	0,74	48 (16)	30 (10)	105 (35)
B	2	1	0,98	1,48	45 (15)	27 (9)	108 (36)
C	3	1	1	0	51 (17)	30 (10)	105 (35)
Ø		1	0,99	0,74	-	-	-

Für alle Klassen wurde eine Selektivität Se von 1 erzielt. Dieses sehr gute Ergebnis ist auch auf die Auswahl repräsentativer Datensätze durch den Kennard-Stone-Algorithmus zurückzuführen, da dieser Extrapolationen bei der Validierung verhindert.

Auch für die Spezifität Sp aller Klassen wurden hohe Werte erreicht. Lediglich eine Einzelmessung eines Ausreißers der Klasse A konnte nicht korrekt eingeteilt werden. Der Klasse B wurden zwei Einzelmessungen einer Probe der Klasse C fälschlicherweise zugewiesen.

Hierin besteht ein Unterschied zu der Kreuzvalidierung, bei der die Spezifität maximiert wurde. Die externe Validierung zeigt auf, dass dies nicht dem realen Verhalten des Klassifikators entspricht. Eine umfassende Validierung empfiehlt sich daher in jedem Fall.

Abschließend sei hervorgehoben, dass der Klassifizierungsfehler E_C im Durchschnitt aller Klassen bei unter 1 % liegt. Dies stellt ein sehr gutes Ergebnis dar.

7. Offline-Prädiktion relevanter Variablen in Bioreaktionsprozessen

7.1 Eine Machbarkeitsstudie anhand von Offline-Analysen

Dieser Abschnitt stellt die Ergebnisse einer Machbarkeitsstudie für die Quantifizierung relevanter Bioprozessgrößen mit spektroskopischen Methoden und multivariater Datenanalyse an einem zyklischen Produktionsprozess mit *Pichia pastoris* dar.

Das Ziel der Untersuchungen war ein Vergleich der drei verwendeten Spektroskopiearten NIR, Raman und 2D-Fluoreszenz sowie eine generelle Bewertung der Eignung dieser Methoden für die Quantifizierung ausgewählter Zielgrößen.

Darüber hinaus sollten die Arbeiten Strategien zur Bildung von PLSR-Modellen hervorbringen. Dies beinhaltet unter anderem die Wahl geeigneter Vorverarbeitungsmethoden sowie die Bestimmung relevanter Spektralbereiche. Weiterhin war der Einfluss von Zellen in der vermessenen Matrix auf das Quantifizierungsergebnis zu untersuchen.

Die spektralen Messungen fanden offline, das heißt nach Ablauf des Prozesses, statt. Durch den Verzicht auf eine Inline-Messung wurden damit zunächst mögliche Störquellen, wie zum Beispiel die Begasung in Verbindung mit einer starken Durchmischung der Kulturbrühe, ausgeschlossen. Der Inline-Betrieb wird in einem späteren Teil der Arbeit näher untersucht.

7.2 Stand der Wissenschaft

Die Beobachtung von Kultivierungsprozessen mit spektroskopischen Messverfahren und multivariater Datenanalyse gilt heutzutage als typisches Beispiel für eine PAT-Anwendung. Ähnlich wie bei der Klassifizierung von Rohmaterialien wurde dies jedoch auch schon vor der Veröffentlichung der PAT-Initiative betrieben.

Insbesondere die NIR- und die 2D-Fluoreszenzspektroskopie kamen dabei zum Einsatz. Es konnten verschiedene Größen in Prozessen mit *Escherichia coli* [HALL et al., 1996], *Saccharomyces cerevisiae* [LINDEMANN et al., 1998], Insektenzellen [RILEY et al., 1997] sowie Streptomyzeten [ARNOLD et al., 2000] mithilfe multivariater Verfahren beobachtet werden.

Nach 2004 stieg die Anzahl an Veröffentlichungen in diesem Themenfeld stark an [POMERANTSEV & RODIONOVA, 2012].

In Produktionsprozessen rekombinanter Proteine mit *Pichia pastoris* konnte die Zelldichte, sowie Substrat- und Produktkonzentrationen aus NIR- bzw. 2D-Fluoresenzspektren vorhergesagt werden [CROWLEY et al., 2005; SURRIBAS et al., 2006].

Auch Säugetierzellkulturen zur Produktion monoklonaler Antikörper wurden mit NIR- und 2D-Fluoreszenzspektroskopie überwacht [HENRIQUES et al., 2009; JOSE et al., 2011]. Für die Beobachtung multipler Prozessgrößen in CHO-Zell-Prozessen wurde insbesondere die Raman-Spektroskopie eingesetzt [ABU-ABSI et al., 2011; BERRY et al., 2015; WHELAN et al., 2012]. Craven et al. nutzten PLSR-Modelle in einem solchen Versuchsaufbau zu Regelung der Glukosekonzentration [CRAVEN et al., 2014].

Paul et al. untersuchten die Eignung der Raman-Spektroskopie in Verbindung mit der PLSR zur Bestimmung verschiedener Größen eines Prozesses mit *Pichia pastoris* anhand artifizieller Proben [PAUL et al., 2016]. Darüber hinaus existieren nur wenige Publikationen, die sich mit der Anwendung der Raman-Spektroskopie an Bakterien- oder Hefeprozessen befassen.

Ein Review aus dem Jahre 2012 gibt einen guten Überblick über Veröffentlichungen hinsichtlich der Beobachtung von Bioprozessen mit den drei in dieser Arbeit untersuchten spektroskopischen Verfahren [LOURENÇO et al., 2012]. Eine Zusammenfassung auch aktueller Veröffentlichungen hinsichtlich der Verwendung der NIR-Spektroskopie in Kultivierungsprozessen findet sich in [HOEHSE et al., 2015].

Die am häufigsten verwendete chemometrische Methode zur Prädiktion von Bioprozessgrößen ist die PLSR [POMERANTSEV & RODIONOVA, 2012]. Aber auch Methoden wie *Multiple Lineare Regression* (MLR) [SMALL, 2006], *Principle Component Regression* (PCR) [RHEE & KANG, 2007] oder nichtlineare Methoden wie Künstliche Neuronale Netze (ANN) [LUCHNER et al., 2015] wurden bereits erfolgreich eingesetzt.

7.3 Untersuchte Prozessgrößen und vorhandenes Probenmaterial

Gegenstand der ersten Untersuchungen war hauptsächlich eine Kultivierung des Wildtypstamms DSM 70382 von *Pichia pastoris*, die gemäß der in Abschnitt 3.1.2 vorgestellten zyklischen Kultivierungsstrategie durchgeführt wurde. Die starke Beprobung eines dieser Zyklen erzeugte das notwendige Datenmaterial für eine multivariate Kalibrierung. Die Kultivierung mit der internen Kennzeichnung NM3213 ist in Bild 7.1 gezeigt.

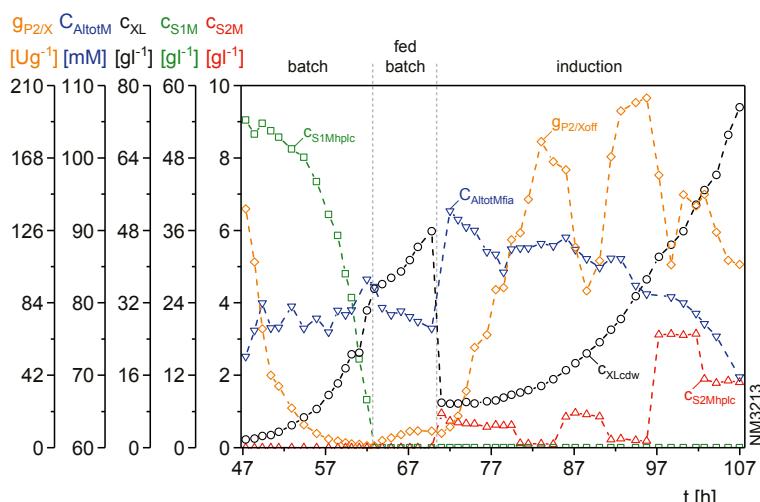


Bild 7.1: Verlauf von Offline-Analysedaten in der Kultivierung NM3213

Dargestellt sind die Offline-Analysen der Zelldichte c_{XLcdw} (schwarz), der Glycerolkonzentration $c_{S1Mhplic}$ (grün), der Methanolkonzentration $c_{S2Mhplic}$ (rot), der Ammoniumkonzentration $C_{AlloMfia}$ (blau) sowie der zellspezifischen Alkoholoxidaseaktivität $g_{P2/Xoff}$ (orange). Diese Größen waren mit spektroskopischen Methoden zu quantifizieren.

Insgesamt ergaben sich aus dem untersuchten Zyklus 50 Proben, die jeweils in Dreifachbestimmung mit den drei getesteten spektroskopischen Methoden vermessen wurden.

Die Wahl eines Wildtypstammes lag darin begründet, dass Zellmaterial zu Analysezwecken für ein anderes Projekt an ein kooperierendes Labor ohne S1-Akkreditierung geliefert wurde. Aus diesem Grund war kein rekombinantes Produkt in dieser Kultivierung vorhanden.

Die Quantifizierung des sekretierten Gesamtproteins c_{PlotM} wurde dagegen anhand der in Bild 7.2 gezeigten Kultivierung SM0313 durchgeführt. Für diese kam der rekombinante Stamm *Pichia pastoris* KM71H D1M1H zum Einsatz. In fünf Zyklen wurden insgesamt 120 Proben genommen, von denen 65 spektral vermessen werden konnten.

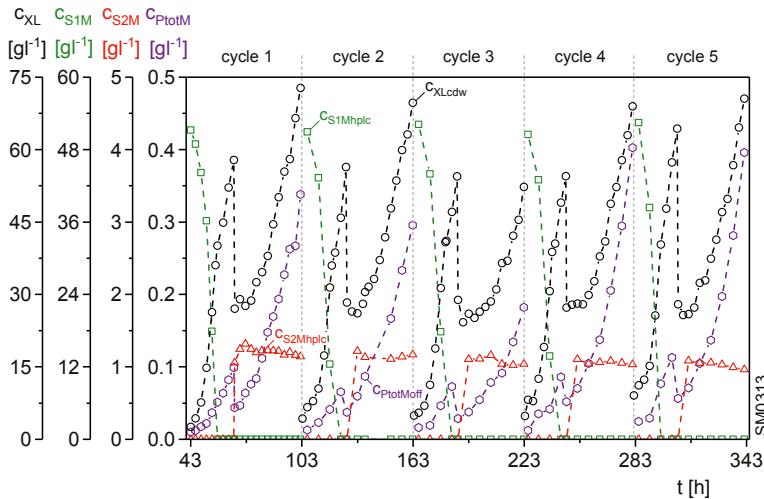


Bild 7.2: Verlauf der Kultivierung SM0313 in fünf Zyklen

Darüber hinaus kam dieser Prozess auch für eine externe Validierung des Kalibriermodells der Glycerolkonzentration c_{S1M} in der Medienphase (erstellt mit NM3213) zum Einsatz, da es sich bei dieser separaten Kultivierung um ein unabhängiges Testset handelt.

7.4 Entwicklung von PLSR-Modellen

7.4.1 Der komplexe Prozess der PLSR-Modellentwicklung

Wichtige Faktoren für die Vorhersagekraft von PLSR-Modellen sind die An- bzw. Abwesenheit von Ausreißern, die gewählte Datenvorverarbeitung, der verwendete Spektralbereich sowie vor allem die Anzahl der berechneten PLS-Komponenten.

Die entsprechenden Einstellungen sind vom Benutzer zu ermitteln. Die Schwierigkeit in der Erstellung bzw. Optimierung eines PLSR-Modells besteht aber darin, dass die genannten Faktoren voneinander abhängen. So kann die Wahl des Spektralbereiches beispielweise beeinflussen, ob eine bestimmte Beobachtung als Ausreißer identifiziert wird oder nicht.

Die Entwicklung und Optimierung von PLSR-Modellen zur Quantifizierung bestimmter Größen mit Hilfe spektraler Daten ist demnach ein komplexer und iterativer Prozess und lässt sich nur schwer nach einem gleichen Schema durchführen, da der Ablauf dieses Prozesses von der jeweiligen Problemstellung abhängt.

Durch geeignete Annahmen kann die Modellbildung beschleunigt werden. Das setzt jedoch eine gute Kenntnis des vorliegenden Datenmaterials und des zugrundeliegenden Bioprozesses voraus.

Der folgende Abschnitt soll anhand des Beispiels der Vorhersage der Glycerolkonzentration c_{S1M} im Kulturüberstand mit Hilfe der NIR-Spektroskopie das in dieser Arbeit prinzipielle Vorgehen bei der Entwicklung von PLSR-Modellen verdeutlichen.

7.4.2 Eine exemplarische Darstellung bei der PLSR-Modellentwicklung

Dafür stehen prinzipiell die 50 Proben der Kultivierung NM3213 zur Verfügung. Ein Großteil davon stammt jedoch aus der Produktionsphase des Prozesses und weist somit eine Glycerolkonzentration von 0 g l^{-1} auf.

Um eine Wichtung des zu erzeugenden PLSR-Modells auf diesen Zustand zu vermeiden, wurdet sich auf die 21 Proben aus der Batchphase und der Glycerol-Fed-Batchphase beschränkt. Bild 7.3 zeigt die NIR-Spektren vom Überstand dieser Kultivierung.

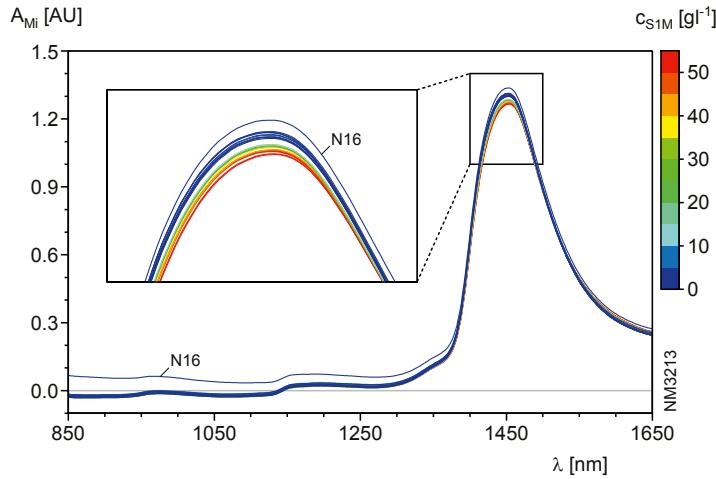


Bild 7.3: NIR-Transmissionsspektren des Überstandes gefärbt nach c_{S1M}

Eine Korrelation zwischen den Absorptionswerten A_{Mi} der Spektren und der Glycerolkonzentration c_{S1M} ist insbesondere im hervorgehobenen Bereich gut zu erkennen. Es fällt weiterhin auf, dass das Spektrum der Probe N16 von den anderen abweicht.

Vor der Erstellung eines PLSR-Modells sollten die Spektraldaten daher auf Ausreißer untersucht werden. Dies wird anhand eines PCA-Modells mithilfe der in Abschnitt 5.5.2 vorgestellten Kenngrößen durchgeführt. Um mögliche Unterschiede zwischen den Spektren durch Basislinienverschiebungen zu eliminieren, kamen die nach der Wellenlänge abgeleiteten (und mittenzentrierten) Spektren zum Einsatz.

PCA-Modelle zur Ausreißerdetektion basieren in dieser Arbeit auf so vielen Hauptkomponenten, dass mindestens 95 % der Varianz in den Spektraldaten (R_X^2) beschrieben wird. Dies ist

zumeist bereits mit zwei Hauptkomponenten der Fall, wodurch sich der *Score Scatter Plot* der ersten beiden Hauptkomponenten (t_1 , t_2) gut zur Ausreißerdetektion eignet.

Im Zweifelsfall (wie im folgenden Beispiel) und bei mehreren Hauptkomponenten wurden zusätzlich die DModX- und Hotelling- T^2 -Werte des Gesamtmodells betrachtet.

Bild 7.4a zeigt den *Score Scatter Plot* eines PCA-Modells. Alle Einzelmessungen der Probe N1 haben hohe Hotelling- T^2 -Werte und liegen außerhalb des 95 % Konfidenzintervalls. Es handelt sich mit einer Wahrscheinlichkeit von 95 % also um einen Ausreißer. Daher wurde diese Probe von der weiteren Modellbildung ausgeschlossen.

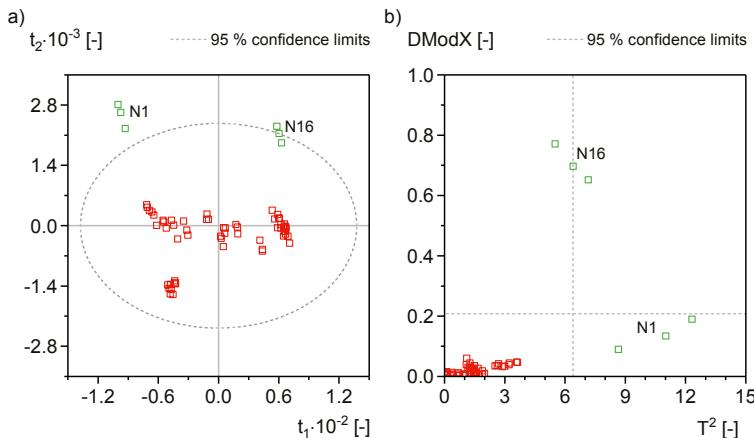


Bild 7.4: a) *Score Scatter Plot* eines PCA-Modells mit $r = 2$, $R_X^2 = 0,982$
b) Auftragung des DModX gegen Hotelling- T^2

Die Probe N16 liegt auf der Grenze des Hotelling- T^2 -Konfidenzintervalls und kann folglich bei ausschließlicher Betrachtung des *Score Scatter Plots* nicht eindeutig als Ausreißer identifiziert werden. Eine Betrachtung von Bild 7.4b zeigt jedoch, dass die Probe N16 zusätzlich einen sehr hohen DModX aufweist. Aus diesem Grund wurde sie ebenfalls als Ausreißer erachtet und von der Modellbildung ausgeschlossen.

Nach der Entfernung von Ausreißern wird mit den verbliebenen Daten ein weiteres PCA-Modell erstellt, um sie, wie oben beschrieben, erneut auf Ausreißer zu untersuchen. Dieses Verfahren ist notwendig, da sich durch das veränderte Datenmaterial ein neues Modell ergibt.

Im vorliegenden Beispiel konnten keine weiteren Ausreißer identifiziert werden. Daher kamen die verbliebenen 19 Proben nun zur Berechnung des finalen PLSR-Modells zum Einsatz. Dabei wurde zunächst die gleiche Vorverarbeitungsmethode (Ableitung und Mittenzentrierung) wie bei der PCA genutzt.

Zur Festlegung der Anzahl r an PLS-Komponenten ist im Idealfall der Vorhersagefehler RMSEP der externen Validierung in Abhängigkeit von r zu betrachten. Steht kein unabhängiger Testdatensatz zur Verfügung ist man auf den Fehler RMSEcv der Kreuzvalidierung angewiesen. Dieser Weg soll hier demonstriert werden.

In Bild 7.5 weist der Fehler RMSEcv ein Minimum bei fünf PLS-Komponenten auf. Daher wurde zunächst ein solches Modell gebildet. Allerdings sollte dieses einer weiteren Prüfung unterzogen werden, um einen möglichen *Overfit* aufzudecken.

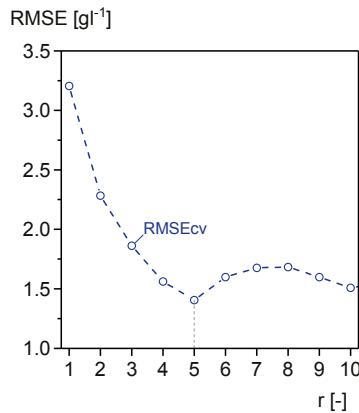


Bild 7.5: Vorhersagefehler RMSEcv in Abhängigkeit der PLS-Komponentenanzahl r

Im vorliegenden Fall weisen z.B. die in Bild 7.6 dargestellten PLS-Regressionskoeffizienten b_j einen stark schwankenden Verlauf in Abhängigkeit des Variablenindex j auf. Dieses Verhalten deutet auf einen *Overfit* hin.

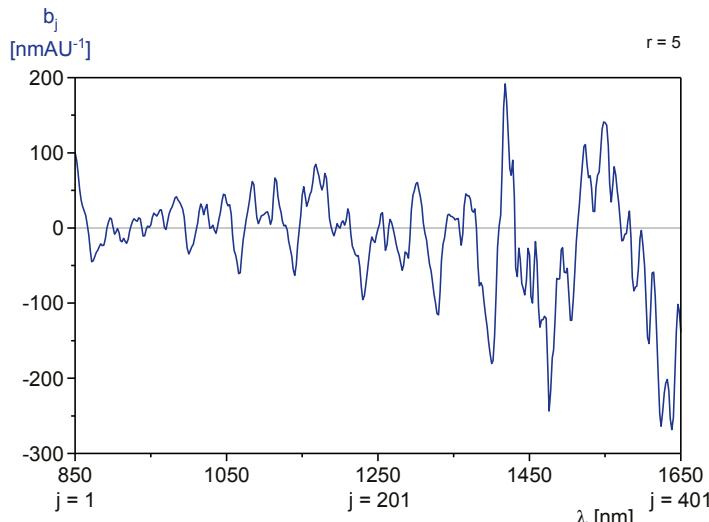


Bild 7.6: PLS-Regressionskoeffizienten b_j in Abhängigkeit des Variablenindex j
($\hat{=}$ Wellenlänge λ_j)

In NIR-Spektren sind Absorptionswerte benachbarter Wellenlängen meist stark korreliert und weisen einen gleichmäßigen Verlauf auf. Demnach sollten auch entsprechende PLS-Regressionskoeffizienten aufeinander folgender Variablen ein ähnliches Verhalten zeigen. Bei stark schwankenden Koeffizienten wird daher davon ausgegangen, dass das PLSR-Modell mess-technisches Rauschen beschreibt, welches den ungleichmäßigen Verlauf verursacht.

Nach Betrachtung der Regressionskoeffizienten der PLSR-Modelle mit einer Faktorenanzahl r von 2 bis 4, wurde sich für das Modell mit 3 PLS-Komponenten entschieden.

Die Regressionskoeffizienten b_j dieses so überarbeiteten Modells sind in Bild 7.7 zu sehen. Für den vorderen Bereich des NIR-Spektrums weisen die b_j nun viel kleinere Werte und einen gleichmäßigeren Verlauf auf. Die macht das Vorliegen eines *Overfits* weniger wahrscheinlich.

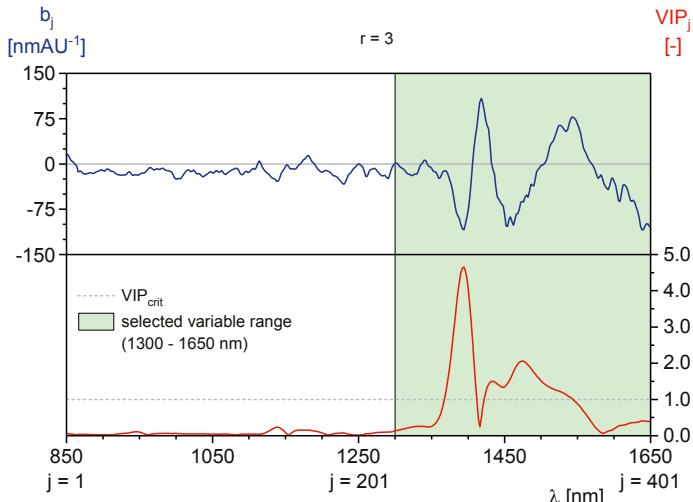


Bild 7.7: PLS-Regressionskoeffizienten b_j und *Variable Importance in Projection* VIP_j zur Variablenelektion aufgetragen gegen den Index j

Im hinteren Bereich zeigen sich dagegen zusammenhängende Bereiche höherer Werte. Diese haben den größten Einfluss auf eine Vorhersage mit diesem Modell, was eine Möglichkeit zur Variablenelektion aufzeigt. Bei zusätzlicher Betrachtung des VIP_j (Gl. (5.45)), ebenfalls in Bild 7.7 gezeigt, bestätigt sich die geringe Bedeutung des vorderen Spektralbereiches für die Vorhersage der Glycerolkonzentration c_{S1M} .

Durch eine Auswahl des Spektralbereiches von 1.300 nm bis 1.650 nm konnte die Variablenanzahl m von 401 auf 176 reduziert werden. Der Vorhersagefehler RMSEcv eines entsprechenden Modells verbesserte sich bei unverändertem r (3) dabei von 1,84 gl^{-1} auf 1,59 gl^{-1} .

Nach der Variablenelektion ist wiederum eine Ausreißerprüfung vorzunehmen, da sich das Datenmaterial erneut verändert hat. Diese verlief hier allerdings negativ.

Ist die Entwicklung des Modells über die vorgestellten Schritte abgeschlossen, sollten noch weitere grafische Werkzeuge zurate gezogen werden, die eine Beurteilung des Modells ermöglichen oder auf Fehlerquellen hinweisen können. Eine solche Grafik ist in Bild 7.8 gezeigt.

Hierin sind die durch Kreuzvalidierung vorhergesagten Glycerolkonzentrationen c_{S1Mnir} gegen ihre Referenzwerte $c_{S1Mhplc}$ aufgetragen.

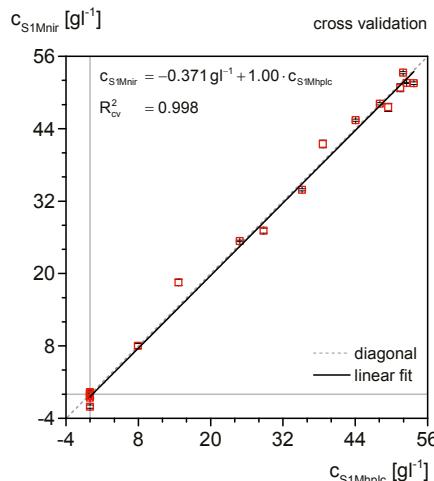


Bild 7.8: Vorhersage c_{S1Mnir} aufgetragen gegen die Referenzwerte $c_{S1Mhplc}$

Im Idealfall liegen die Punkte auf einer Geraden durch den Ursprung mit der Steigung eins. Bei einem nichtlinearen Verlauf können ein Wechsel der Vorverarbeitungsmethode oder die Berechnung weiterer Faktoren Abhilfe schaffen. Ist dies nicht der Fall, sollte die Eignung der verwendeten Methode (z.B. PLSR) angezweifelt werden.

Bild 7.9 zeigt die Vorhersagen und deren Offline-Referenzwerte im Kultivierungsverlauf.

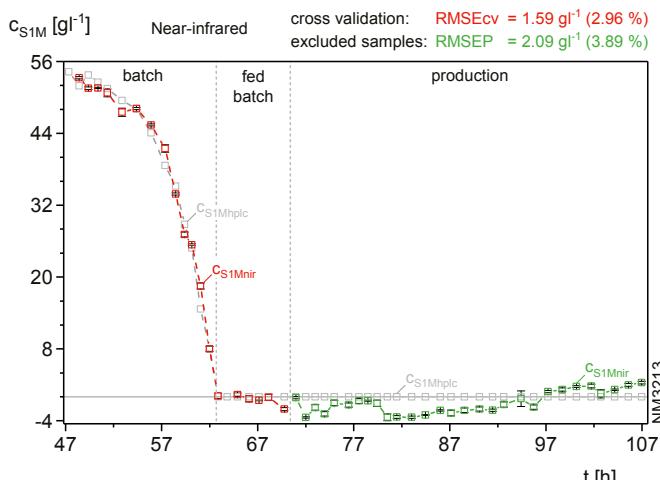


Bild 7.9: Vorhersage c_{S1Mnir} im Überstand im Rahmen einer Kreuzvalidierung

Die Glycerolkonzentration c_{S1M} in der Batch- und der Fed-Batchphase kann gut vorhergesagt werden (rot). Etwas stärkere Abweichungen sind bei der Vorhersage der Produktionsphase (grün) zu erkennen. Entsprechende Proben waren nicht an der Modellbildung beteiligt.

Ein Schwachpunkt des vorliegenden Modells liegt in der Verteilung der Konzentrationswerte. So ist anhand von Bild 7.8 und Bild 7.9 zu erkennen, dass einige Proben mit einer Glycerolkonzentration von 0 gl^{-1} Teil des Kalibrierdatensatzes waren. Weiterhin sind hohe Konzentrationen von über 40 gl^{-1} ebenfalls relativ häufig. Kleinere Werte sind dagegen nur selten vertreten. Die Verteilung der Daten weist also nicht die geforderte Uniformität auf. Dies lässt sich jedoch mit der äquidistanten Probenahme im untersuchten Wachstumsprozess begründen.

Eine externe Modellvalidierung mit einem unabhängigen Validierdatensatz ist hier mit den Proben der in Bild 7.10 gezeigten Kultivierung SM0313 zusätzlich möglich. Stünde ein solcher Datensatz nicht zur Verfügung, wäre die Modellbildung an dieser Stelle für die gewählte Art der Datenvorverarbeitung abgeschlossen.

Aber auch hier weisen viele Proben eine Glycerolkonzentration c_{S1M} von 0 gl^{-1} auf. In Bild 7.10 sind die Vorhersagen der meisten dieser Proben in grün dargestellt. Sie wurden jedoch nicht zur externen Validierung herangezogen.

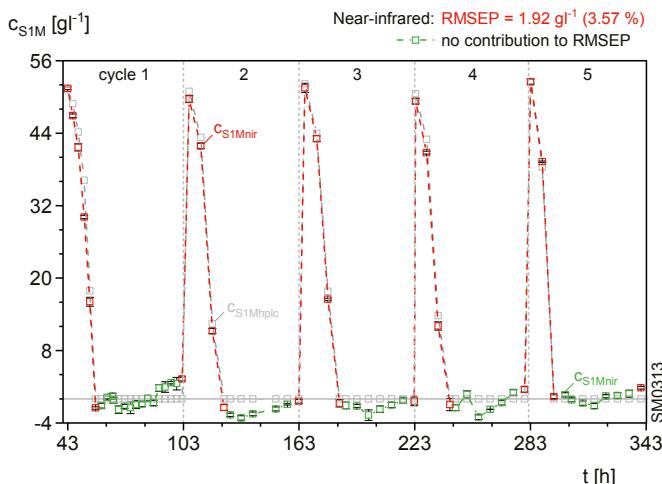


Bild 7.10: Vorhersage c_{S1Mnr} im Überstand für ein unabhängiges Testset

Der Fehler RMSEP der externen Validierung wurde nur mit den in rot dargestellten Vorhersagen berechnet und dabei zu $1,92 \text{ gl}^{-1}$ ermittelt. Die Vorhersagegüte des Modells ist damit für Proben des unabhängigen Testsets etwas schlechter als für Proben des Kalibrierdatensatzes ($\text{RMSE}_{\text{cv}} = 1.59 \text{ gl}^{-1}$). Dieses Verhalten ist aufgrund der Unabhängigkeit der beiden Datensätze zu erwarten.

Abschließend soll das gebildete Modell noch einmal im Hinblick auf die Anzahl r zu berechnender PLS-Komponenten untersucht werden. Dazu ist in Bild 7.11 neben dem Vorhersagefehler RMSE_{cv} der Kreuvalidierung auch der Vorhersagefehler RMSEP in Abhängigkeit von r aufgetragen.

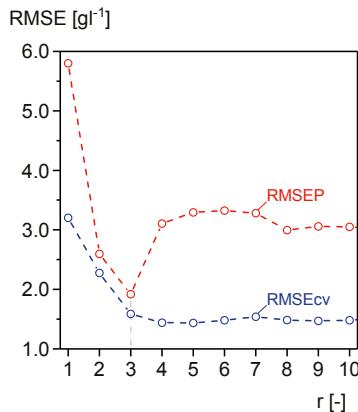


Bild 7.11: Fehler RMSEcv und RMSEP in Abhängigkeit der PLS-Komponentenanzahl r

Dieser weist ein Minimum bei drei PLS-Komponenten auf. Dies entspricht der zuvor gewählten Komponentenanzahl. Für Modelle mit mehr als drei PLS-Komponenten liegt, aufgrund des höheren Fehlers, wahrscheinlich ein *Overfit* vor.

Bei dem vorgestellten Vorgehen der PLSR-Modellentwicklung wurde die Vorverarbeitungsmethode zu Beginn festgelegt. Zur Erzeugung eines optimierten Modells, sollten aber auch andere Verfahren getestet werden. Tabelle 7.1 fasst die Kenngrößen des vorgestellten Modells (de1 + mc) zusammen und stellt außerdem einen Vergleich mit der Verwendung des SNV-Filters (snv + mc) und der alleinigen Verwendung der Mittenzentrierung (mc) an.

Tabelle 7.1: Vergleich unterschiedlicher Vorverarbeitungsmethoden bei der Vorhersage von c_{S1M} im Überstand mit der NIR-Spektroskopie

	mc	de1 + mc	snv + mc
Spektralbereich	komplett	1300 - 1650 nm	1300 - 1650 nm
$n_{CS}^1)$	[\cdot]	57 (19)	57 (19)
r	[\cdot]	4	3
R_x^2	[\cdot]	0,999	0,999
R_Y^2	[\cdot]	0,991	0,997
R_{cv}^2	[\cdot]	0,982	0,995
RMSEcv	[$g l^{-1}$]	2,92	1,59
RMSEcv _{rel}	[%]	5,43	2,96
$n_{VS}^1)$	[\cdot]	78 (26)	78 (26)
R_p^2	[\cdot]	0,982	0,991
RMSEP	[$g l^{-1}$]	5,25	1,92
RMSEP _{rel}	[%]	9,77	3,57

¹⁾ Die Klammern geben die Anzahl unabhängiger Proben an

Eine Spektrenvorverarbeitung, die nur auf Mittenzentrierung (mc) beruht, brachte das schlechteste Ergebnis hervor. Die Modellentwicklung führte dabei nicht zu einer Eingrenzung des Spektralbereiches. Die Verwendung eines SNV-Filters (snv + mc) lieferte beinahe identische Ergebnisse zur ersten Ableitung (de1 + mc). Im vorliegenden Fall fiel die Wahl weiterhin auf diese Methode, da mit ihr das beste Ergebnis bei der externen Validierung erzielt wurde.

7.4.3 Schwierigkeiten bei Betrachtung der zellhaltigen Flüssigphase L

Zum Zweck der Bewertung des Störeinflusses von Zellen bei der PLSR-Modellentwicklung kamen Spektren von der zellbehafteten Kulturbrühe zum Einsatz. Als Zielgröße diente dabei die Glycerolkonzentration c_{S1L} in der Flüssigphase.

Wie für das Modell für c_{S1M} wurden zunächst nur die Proben N1 bis N21 aus den Batch- und Fed-Batchphasen genutzt, um eine möglichst uniform verteilte Zielgröße zu gewährleisten. Es wurde der gleiche Spektralbereich (1.300 - 1.650 nm) gewählt und die gleiche Vorverarbeitung (de1 + mc) darauf angewandt. Ausreißer konnten hier nicht entdeckt werden.

Mit diesen Daten konnte ein Modell mit nur einer PLS-Komponente gebildet werden, das einen mittleren Vorhersagefehler der Kreuzvalidierung RMSEcv von 2,76 g l^{-1} erreichte.

Die Vorhersage der Proben N1 bis N21 (rot), gezeigt in Bild 7.12, weist eine gute Anpassung an die Referenzwerte in der Batchphase auf. Dies verleiht zunächst zu der Schlussfolgerung, dass die Glycerolkonzentration c_{S1L} in der zellhaltigen Flüssigphase quantifiziert werden kann.

Der höhere Vorhersagefehler RMSEcv dieses Modells im Vergleich mit dem Modell der Medienphase M (1,59 g l^{-1}) lässt sich sowohl durch die Präsenz störender Zellen in der Matrix, als auch durch den notwendigen Wechsel des Messverfahrens von Transmission auf diffuse Reflexion begründen.

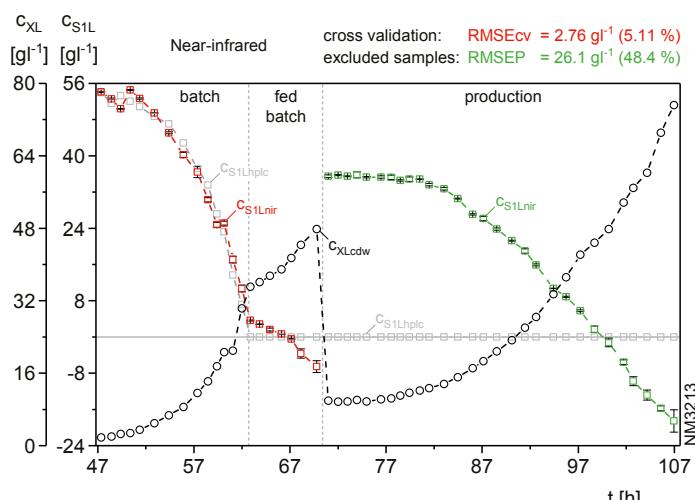


Bild 7.12: Versuch der Vorhersage von c_{S1L} in der Flüssigphase mit der NIR-Spektroskopie im Vergleich mit der Zelldichte c_{XL}

Tatsächlich gibt dieses Ergebnis jedoch eine völlig falsche Einschätzung der Modellgüte ab. Dies wird anhand der Vorhersagen in der Glycerol-Fed-Batchphase und insbesondere in der Produktionsphase anhand der Prädiktion (grün) von nicht an der Modellbildung beteiligten Proben deutlich.

Insgesamt weist der Verlauf von c_{S1Nir} eine horizontale Spiegelsymmetrie mit der ebenfalls in Bild 7.12 gezeigten Zelldichte c_{XL} auf. Dies zeigt eindrucksvoll, dass die Zelldichte in dem erzeugten Glycerolmodell offensichtlich die dominierende Rolle spielt.

Ursache hierfür ist, dass Zellen einerseits ein starkes Signal im NIR-Spektrum hervorrufen (siehe Abschnitt 7.5) und andererseits, dass Zelldichte und Glycerolkonzentration in einer Batchphase proportional zueinander verlaufen und damit hoch korreliert sind. Für alle 21 an der Modellbildung beteiligten Proben aus der Batchphase und der Glycerol-Fed-Batchphase wurde ein Korrelationskoeffizient R von -0,977 ermittelt.

Diese Korrelation lässt sich für eine indirekte Quantifizierung von c_{S1M} bewusst heranziehen. Ein entsprechendes Modell ist jedoch nur für die Batchphase eines immer gleich laufenden (z.B. zyklischen) Prozesses gültig, da das PLSR-Modell nicht auf tatsächlichen Signalen des Glycerols im Spektrum beruht.

Aus diesem Grund wurden in einem weiteren Modell alle 50 Proben der Kultivierung NM3213 verwendet, um die Korrelation R von c_{XL} und c_{S1L} auf einen Wert von -0,529 zu reduzieren und das Modell möglichst auf einem Signal des Glycerols basieren zu lassen. Diese Maßnahme geht allerdings weiter zulasten der Uniformität der Zielgrößenverteilung, da hierbei die Anzahl an Proben mit c_{S1L} von 0 g l^{-1} stark ansteigt.

Die Vorhersage mit diesem Modell ist in Bild 7.13 gezeigt und wird mit einem Fehler RMSEcv von 7,20 g l^{-1} bewertet. Es wird deutlich, dass auch dieses Modell nicht von praktischem Nutzen ist. Allerdings wird auf diese Weise eine zu optimistische Einschätzung der Modellgüte bei hauptsächlicher Verwendung korrierender Proben aus der Batchphase vermieden.

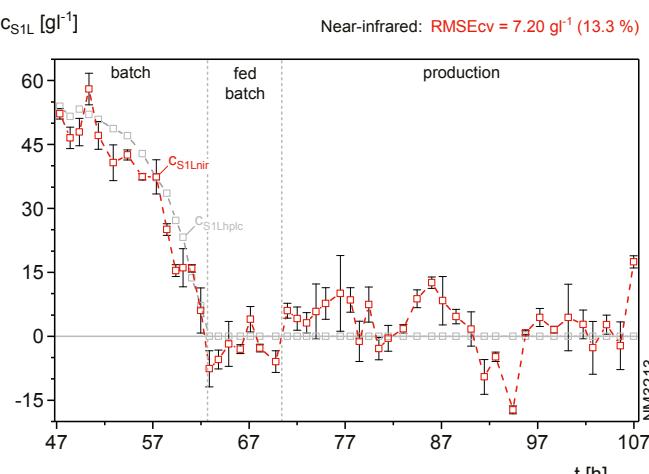


Bild 7.13: Vorhersage von c_{S1L} in der Flüssigphase mit der NIR-Spektroskopie

7.5 Prädiktion der Zelldichte

Optische Verfahren sind etablierte Methoden zur Bestimmung der Zelldichte. So wird beispielsweise die Trübungsmessung konventionell am Bioprozess eingesetzt (siehe Abschnitt 3.3.3). Eine Quantifizierung der Zelldichte mit den getesteten spektroskopischen Verfahren sollte daher gut möglich sein.

In Bild 7.14 sind die nach der Zelldichte c_{XL} gefärbten NIR-Spektren der Kulturbrühe gezeigt. Ein Zusammenhang zwischen c_{XL} und den Verläufen der Spektren ist deutlich zu erkennen.

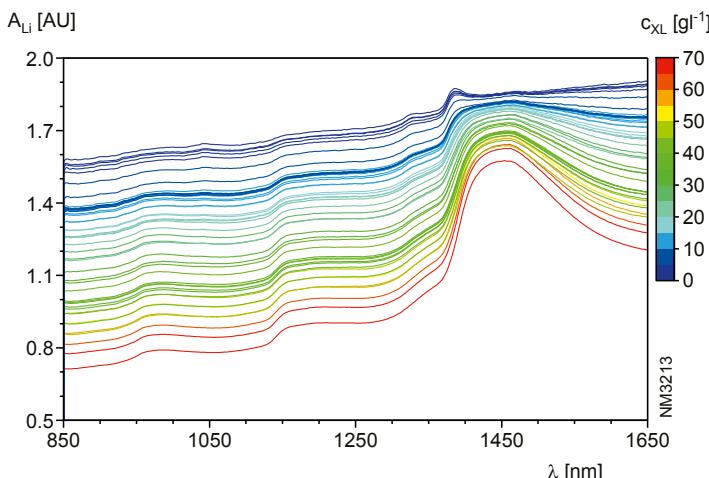


Bild 7.14: NIR-Reflexionsspektren im Prozessverlauf gefärbt nach der Zelldichte c_{XL}

Die Abnahme der Absorptionswerte mit steigender Zelldichte lässt sich durch die Verwendung des Messprinzips der diffusen Reflexion erklären, die hier zum Einsatz kam. Zellfreies Kulturmödium reflektiert nicht und weist hier somit die höchsten (vermeintlichen) Absorptionswerte auf.

Die Ergebnisse der Modellbildung sind in Bild 7.15 dargestellt. Für alle drei spektroskopischen Verfahren kann der Verlauf der Zelldichte sehr gut wiedergegeben werden.

Bei der Bildung eines PLSR-Modells zur Bestimmung der Zelldichte mit der NIR-Spektroskopie benötigt ein optimales Modell nur zwei PLS-Komponenten. Der Vorhersagefehler der Kreuzvalidierung $\text{RMSE}_{\text{cv,rel}}$ wurde zu 4,37 % ermittelt.

Ein ähnliches Ergebnis, mit einem Fehler $\text{RMSE}_{\text{cv,rel}}$ von 4,40 %, konnte mit der Raman-Spektroskopie und mit drei PLS-Komponenten erzielt werden. Die 2D-Fluoreszenzspektroskopie brachte das beste Modell hervor, das mit fünf PLS-Komponenten jedoch auch das komplexeste ist. Es hat einen Vorhersagefehler $\text{RMSE}_{\text{cv,rel}}$ von 3,16 %.

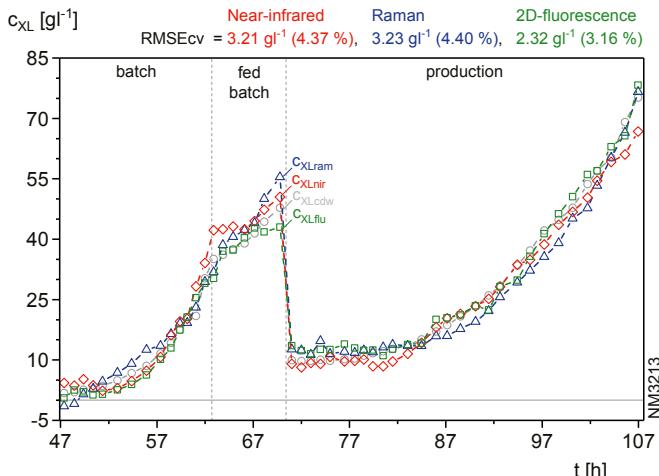


Bild 7.15: Vorhersage der Zelldichte c_{XL} mit verschiedenen spektroskopischen Verfahren

Alle Einstellungen und Kenngrößen der gebildeten Modelle sind in Tabelle 7.2 aufgelistet.

Tabelle 7.2: Vergleich der Vorhersage von c_{XL} mit spektroskopischen Verfahren

	Nahinfrarot	Raman	2D-Fluoreszenz
Spektralbereich	komplett	komplett	komplett
Vorverarbeitung	snv + mc	de1 + mc	mc
$n_{cs}^1)$	[\cdot]	150 (50)	150 (50)
r	[\cdot]	2	3
R_x^2	[\cdot]	0,994	0,972
R_y^2	[\cdot]	0,972	0,975
R_{cv}^2	[\cdot]	0,974	0,971
RMSEcv	[gl^{-1}]	3,21	3,23
RMSEcv _{rel}	[%]	4,37	4,40
			3,16

¹⁾ Die Klammern geben die Anzahl unabhängiger Proben an

Bei Betrachtung von Bild 7.16a kann man erkennen, dass mit der NIR-Spektroskopie hohe Konzentrationen schlechter als kleinere vorhergesagt werden. Eine mögliche Ursache hierfür könnten nichtoptimale Messeinstellungen sein, die zu einer Sättigung des NIR-Detektors bei höheren Zelldichten führen.

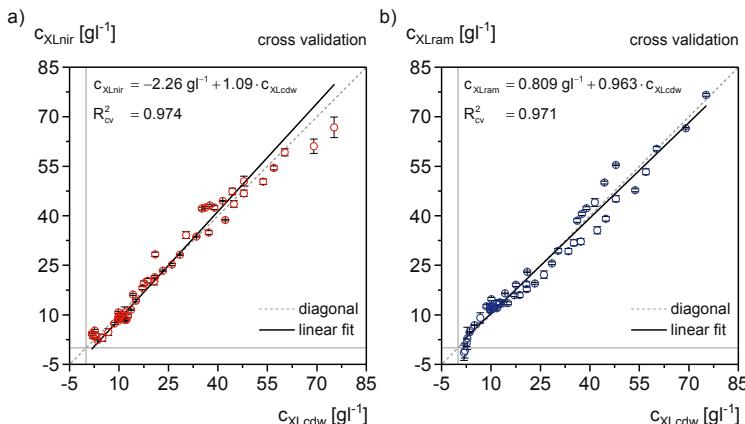


Bild 7.16: Vorhersagen der Zelldichte c_{XL} aufgetragen gegen die Referenzwerte $c_{XL\text{cdw}}$
a) NIR-Spektroskopie b) Raman-Spektroskopie

Für die Raman-Spektroskopie ist in Bild 7.16b zu sehen, dass besonders im Bereich kleiner Zelldichten kein linearer Zusammenhang zwischen der Vorhersage $c_{XL\text{ram}}$ und den Referenzwerten $c_{XL\text{cdw}}$ vorliegt. Dies konnte auch durch eine Änderung der Modelleinstellungen nicht behoben werden.

Eine Bestimmung der Zelldichte ist mit allen drei getesteten spektroskopischen Verfahren wie erwartet möglich. Die Modellgüte ist vergleichbar, wobei jede Spektroskopieart ihre Vor- und Nachteile hat. So weist z.B. das gebildete NIR-Modell Schwächen bei hohen und das Raman-Modell Schwächen bei niedrigen Zelldichten auf. Das 2D-Fluoreszenzmodell liefert das beste Ergebnis bei gleichzeitig höchster Modellkomplexität.

7.6 Prädiktion der Glycerolkonzentration

Die Möglichkeit zur Verwendung der NIR-Spektroskopie zur Quantifizierung der Glycerolkonzentration wurde in Abschnitt 7.4 ausführlich dargelegt. Der folgende Abschnitt stellt die Ergebnisse der anderen getesteten spektroskopischen Verfahren, insbesondere der Raman-Spektroskopie, dar.

Bei Betrachtung der Raman-Spektren in Bild 7.17 ist ein deutlicher Zusammenhang zwischen der Höhe des Signals im Spektrum und der Glycerolkonzentration c_{S1M} gegeben. Insbesondere der Doppelpeak in der Nähe der Wellenzahl ν von 2.900 cm^{-1} weist eine hohe Korrelation mit c_{S1M} auf. Dieses Signal wird wahrscheinlich durch Strettschwingungen der C-C-Bindungen innerhalb des Glycerolmoleküls verursacht [VANDENABEELE, 2013].

Eine Untersuchung der PLS-Regressionskoeffizienten b_j und der *Variable Importance in Projection* VIP_j anhand von Bild 7.18 führte zu einer Einschränkung des Raman-Spektrums auf den Bereich zwischen 2.750 cm^{-1} und 3.050 cm^{-1} . Durch diese Maßnahme konnte die X-Variablenanzahl m von 3.001 auf 301 reduziert werden.

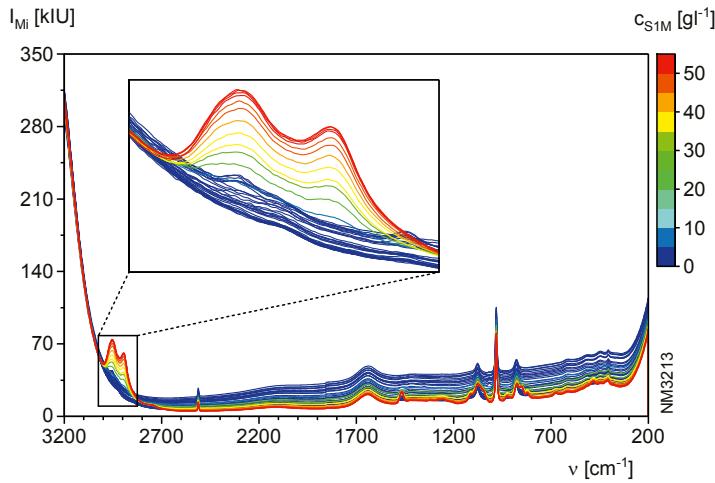


Bild 7.17: Raman-Spektren im Prozessverlauf gefärbt nach der Glycerolkonzentration c_{S1M}

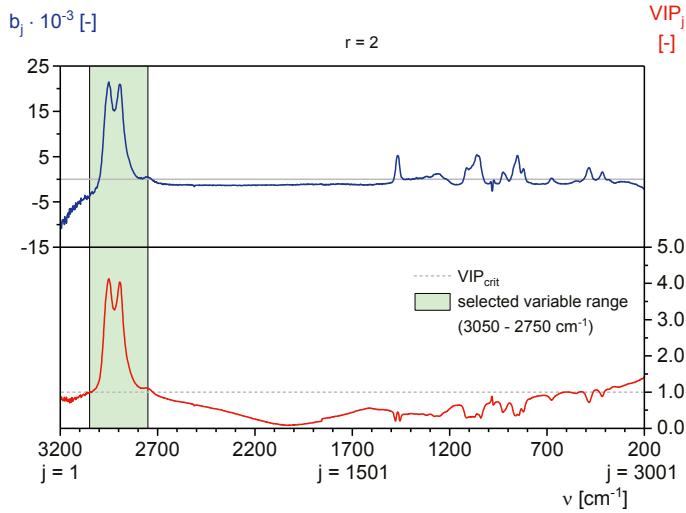


Bild 7.18: Variablenelektion bei der Modellerstellung zur Vorhersage von c_{S1M}

Nach der Variablenelektion wurde mit nur einer PLS-Komponente ein optimales Modell erstellt. Der Vorhersagefehler der Kreuzvalidierung $RMSE_{cv, rel}$ beträgt 2,87 %. Bild 7.19 zeigt die gute Anpassung an die Referenzdaten anhand des Verlaufs der vorhergesagten Glycerolkonzentration $c_{S1M,ram}$ (blau). Auch nicht an der Modellbildung beteiligte Proben mit Konzentrationen von 0 $g l^{-1}$ (orange) werden gut vorhergesagt.

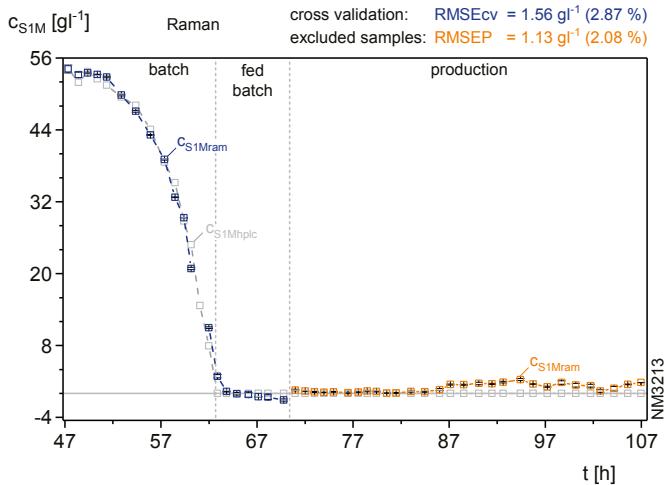


Bild 7.19: Vorhersage c_{S1Mram} im Überstand im Rahmen einer Kreuzvalidierung

Die externe Validierung des Modells mit den Proben der Kultivierung SM0313, dargestellt in Bild 7.20, bestätigt das gute Ergebnis der Modellierung. Der Vorhersagefehler $RMSEP_{rel}$ ist, wie erwartet, mit 3,70 % höher als der Vorhersagefehler der Kreuzvalidierung.

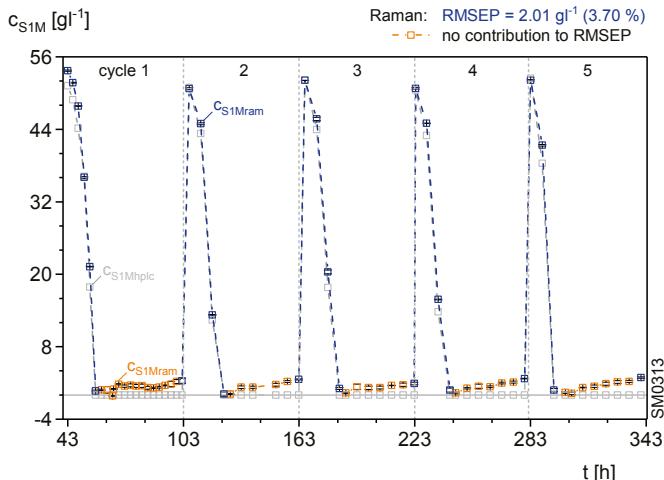


Bild 7.20: Vorhersage c_{S1Mram} im Überstand für ein unabhängiges Testset

Die Referenzwerte der in orange dargestellten Vorhersagen weisen Konzentrationen von 0 g l^{-1} auf und waren nicht teil der Validierung, um eine Wichtung des $RMSEP$ auf diesen Zustand zu vermeiden.

Zur Vorhersage der Glycerolkonzentration c_{S1L} in der zellhaltigen Kulturbrühe mit Raman-Spektroskopie wurde ein Modell mit allen 50 Proben der Kultivierung NM3213 entwickelt (vgl. Abschnitt 7.4.3). Dies benötigt im Vergleich zum Modell von c_{S1Lram} im Überstand, anstelle von einer, eine Anzahl von drei PLS-Komponenten. Bild 7.21 zeigt die Vorhersage von c_{S1Lram} .

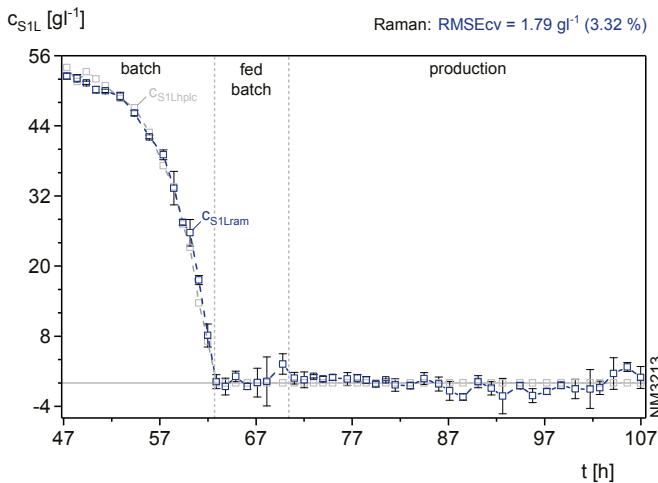


Bild 7.21: Vorhersage c_{S1Lram} in der Kulturbrühe im Rahmen einer Kreuzvalidierung

Der Vorhersagefehler $\text{RMSE}_{\text{cv, rel}}$ der Kreuzvalidierung beträgt 3,32 % und ist damit, wie erwartet, etwas höher als für das Modell ohne störende Zellen in der Matrix. Konzentrationen im Bereich von 0 gl^{-1} werden ebenfalls zufriedenstellend vorhergesagt.

Bei Betrachtung von Bild 7.22 erkennt man, dass mit Raman-Spektroskopie im Gegensatz zur Zelldichte (siehe Bild 7.16b) zwischen Vorhersage und Referenzwerten der Glycerolkonzentration c_{S1L} ein linearer Zusammenhang besteht.

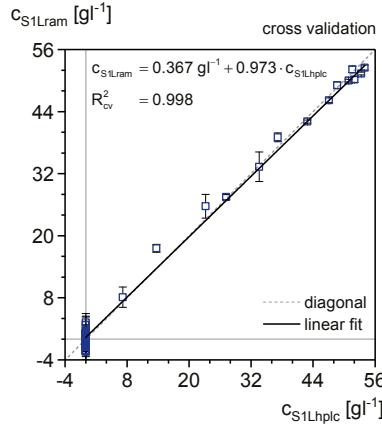


Bild 7.22: Vorhersage c_{S1Lram} aufgetragen gegen die Referenzwerte $c_{S1L\text{hplc}}$

Die Quantifizierung der Glycerolkonzentrationen c_{S1M} und c_{S1L} mit den drei eingesetzten spektroskopischen Verfahren sind anhand üblicher Kenngrößen in Tabelle 7.3 für den Kulturüberstand und in Tabelle 7.4 für die Kulturbrühe zusammengefasst.

Tabelle 7.3: Vergleich spektroskopischer Verfahren zur Vorhersage von c_{S1M} im Überstand

	Nahinfrarot	Raman	2D-Fluoreszenz
Spektralbereich	1300 - 1650 nm	3050 - 2750 cm ⁻¹	komplett
Vorverarbeitung	de1 + mc	snv + mc	mc
$n_{CS}^1)$	[\cdot]	57 (19)	60 (20)
r	[\cdot]	3	1
R_x^2	[\cdot]	0,999	0,991
R_Y^2	[\cdot]	0,997	0,996
R_{cv}^2	[\cdot]	0,995	0,995
RMSEcv	[gl^{-1}]	1,59	1,56
RMSEcv _{rel}	[%]	2,96	2,87
$n_{VS}^1)$	[\cdot]	78 (26)	78 (26)
R_P^2	[\cdot]	0,991	0,997
RMSEP	[gl^{-1}]	1,92	2,01
RMSEP _{rel}	[%]	3,57	3,70
			103

¹⁾ Die Klammern geben die Anzahl unabhängiger Proben an

Tabelle 7.4: Vorhersage von c_{S1L} in der Kulturbrühe mit NIR- und Raman-Spektroskopie

	Nahinfrarot	Raman
Spektralbereich	1300 - 1650 nm	3050 - 2750 cm ⁻¹
Vorverarbeitung	de1 + mc	snv + mc
$n_{CS}^1)$	[\cdot]	150 (50)
r	[\cdot]	4
R_x^2	[\cdot]	0,999
R_Y^2	[\cdot]	0,908
R_{cv}^2	[\cdot]	0,859
RMSEcv	[gl^{-1}]	7,20
RMSEcv _{rel}	[%]	13,3
		3,32

¹⁾ Die Klammern geben die Anzahl unabhängiger Proben an

Eine erfolgreiche Nutzung der 2D-Fluoreszenzspektroskopie zur Quantifizierung der Glycerolkonzentration c_{S1M} wurde nicht erwartet, da Glycerol nicht fluoresziert. Zu Testzwecken wurde dennoch versucht ein valides Modell zu erstellen. Dies schlug im Rahmen einer externen Validierung jedoch völlig fehl.

Der Vorhersagefehler RMSEP wird im Überstand zu $56,0 \text{ gl}^{-1}$ berechnet und ist damit höher als die maximale Glycerolkonzentration im Kalibrierdatensatz ($54,3 \text{ gl}^{-1}$). Auf eine weitere Modellbildung für Messungen in der Flüssigphase mit 2D-Fluoreszenz wurde daher verzichtet.

Während Vorhersagen der Glycerolkonzentration über PLSR-Modelle aus NIR- und Raman-Spektralen des Überstandes gut möglich sind, ist die 2D-Fluoresenzspektroskopie damit für diese Aufgabe ungeeignet.

Zellen in der vermessenen Matrix stören die spektralen Messungen unterschiedlich stark. Während eine Messung in der zellhaltigen Kulturbrühe mit der NIR-Spektroskopie nicht sinnvoll erscheint, eignet sich die Raman-Spektroskopie jedoch sehr gut, da entsprechende Modelle nur geringfügig höhere Fehler aufweisen als solche für den Kulturüberstand.

7.7 Prädiktion der Ammoniumkonzentration

Vor der Modellbildung zur Quantifizierung der Ammoniumkonzentration wurden vier Proben hinsichtlich ihrer Offline-Referenzmessungen als Ausreißer identifiziert.

Für die in Bild 7.23 gezeigte Vorhersage der Ammoniumkonzentration C_{AltotM} mit der NIR-Spektroskopie wurde im Überstand ein relativer Fehler $\text{RMSE}_{\text{cv,rel}}$ von 11,5 % erreicht.

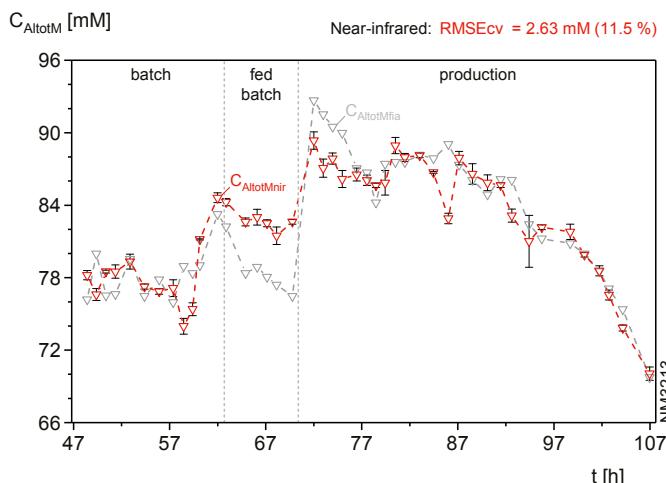


Bild 7.23: Vorhersage von C_{AltotM} im Überstand mit der NIR-Spektroskopie

Trotz dieses hohen Fehlers kann der tendenzielle Verlauf von C_{AltotM} relativ gut vorhergesagt werden. Vor dem Hintergrund der stark fehlerbehafteten Referenzwerte und in Ermangelung einer Online-Messung könnte ein derartiges Modell dennoch einen Mehrwert für die Prozessbeobachtung liefern.

Die in Bild 7.24 dargestellte Prädiktion der Ammoniumkonzentration C_{AltotL} in der Flüssigphase bildet mit einem Fehler von 7,24 % unerwarteter Weise die Referenzdaten besser ab. Dies äußert sich auch in einem glatteren Verlauf im Vergleich mit der Vorhersage von C_{AltotM} in der Medienphase.

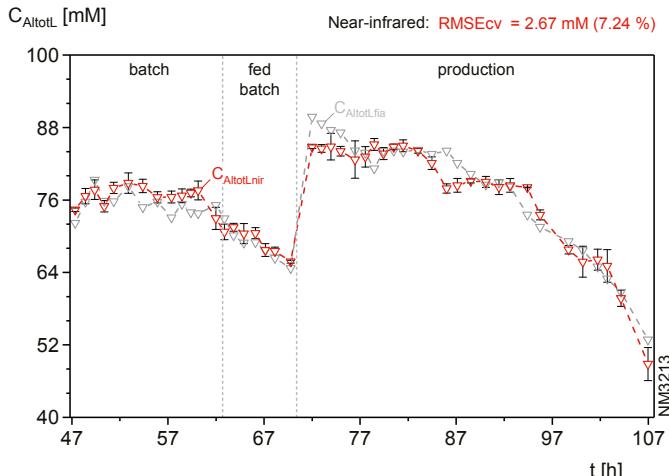


Bild 7.24: Vorhersage von C_{AltotL} in der Kulturbrühe mit der NIR-Spektroskopie

Der Grund für diese Verbesserung liegt wahrscheinlich an einem starken Einfluss der Zellen auf die Spektren und der hohen Korrelation von C_{AltotL} mit der Zelldichte c_{XL} ($R = -0,806$, vgl. Abschnitt 7.4.3). Die Vorhersage von C_{AltotL} in der Flüssigphase beruht demnach wahrscheinlich auf einer indirekten Quantifizierung und ist folglich kritisch zu betrachten.

Das gleiche Verhalten mit weiter verminderten Fehlern konnte bei der Untersuchung der Raman-Spektroskopie beobachtet werden. Für die 2D-Fluoreszenzspektroskopie ergeben sich die schlechtesten Ergebnisse bei gleichzeitig höchster Modellkomplexität.

Die Kenngrößen aller gebildeten Modelle zur Vorhersage der Ammoniumkonzentration sind in Tabelle 7.5 (C_{AltotM}) und Tabelle 7.6 (C_{AltotL}) aufgeführt.

Tabelle 7.5: Vergleich spektroskopischer Vorhersagen von C_{AltotM} im Kulturüberstand

	Nahinfrarot	Raman	2D-Fluoreszenz
Spektralbereich	1300 - 1650 nm	komplett	komplett
Vorverarbeitung	de1 + mc	mc	mc
$n_{\text{CS}}^1)$	[-]	132 (44)	138 (46)
r	[-]	3	2
R_x^2	[-]	0,999	0,981
R_y^2	[-]	0,778	0,825
R_{cv}^2	[-]	0,754	0,784
RMSEcv	[mM]	2,63	2,51
RMSEcv _{rel}	[%]	11,5	11,0
			13,8

¹⁾ Die Klammern geben die Anzahl unabhängiger Proben an

Tabelle 7.6: Vergleich spektroskopischer Vorhersagen von $C_{\text{AltoI}_{\text{L}}}$ in der Flüssigphase

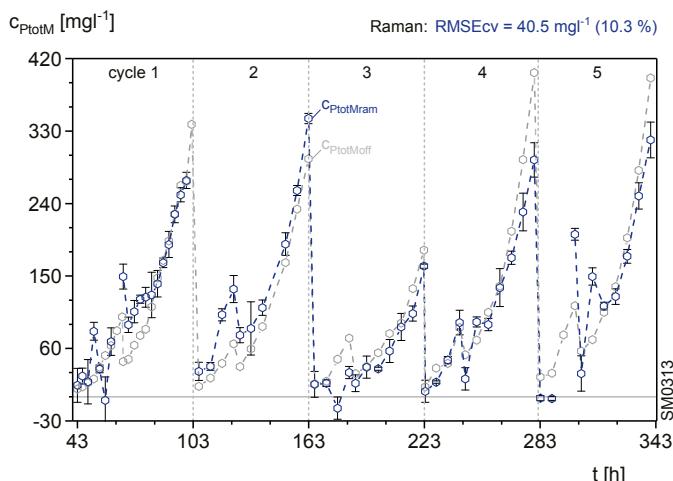
	Nahinfrarot	Raman	2D-Fluoreszenz
Spektralbereich	1300 - 1650 nm	komplett	komplett
Vorverarbeitung	de1 + mc	mc	mc
$n_{\text{CS}}^1)$	[\cdot]	138 (46)	138 (46)
r	[\cdot]	3	4
R_x^2	[\cdot]	0,999	0,999
R_y^2	[\cdot]	0,913	0,947
R_{cv}^2	[\cdot]	0,887	0,930
RMSEcv	[mM]	2,67	2,11
RMSEcv _{rel}	[%]	7,24	5,72
			9,00

¹⁾ Die Klammern geben die Anzahl unabhängiger Proben an

7.8 Prädiktion der Gesamtproteinkonzentration

Modelle zur Prädiktion der Gesamtproteinkonzentration c_{PlotM} in der Medienphase basieren auf maximal 62 Proben der Kultivierung SM0313. Eine externe Validierung dieser Modelle oder eine Untersuchung der Flüssigphase war leider nicht möglich.

In Bild 7.25 sind Offline-Analysen von c_{PlotM} und die Vorhersagen mit der Raman-Spektroskopie gezeigt.

Bild 7.25: Vorhersage von c_{PlotM} im Überstand mit der Raman-Spektroskopie

Eine Quantifizierung von c_{PlotM} ist trotz des vergleichsweise hohen Wertes des relativen Vorhersagefehlers RMSEcv_{rel} von 10,3 % durchaus möglich. Dies gilt auch für das etwas schlechtere Ergebnis mit der NIR-Spektroskopie (nicht gezeigt).

Es ist allerdings wahrscheinlich, dass Korrelationen mit anderen Substanzen in der Medienphase zu einer indirekten Vorhersage führen, da Proteine im relevanten Konzentrationsbereich kein eigenes verwertbares Signal im Raman-Spektrum liefern [PAUL et al., 2016].

Das beste Resultat für die Quantifizierung von c_{PlotM} mit spektroskopischen Verfahren liefert in dieser Arbeit die 2D-Fluoreszenzspektroskopie mit einem relativen Vorhersagefehler RMSEcv_(re) von 6,69 %. Ursache hierfür ist ein in Bild 7.26 deutlich zu erkennendes Signal der im Medium gelösten Proteine im 2D-Fluoreszenzspektrum.

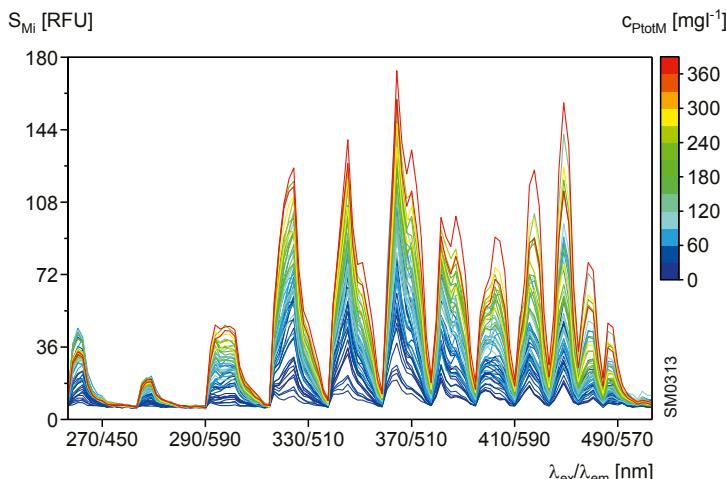


Bild 7.26: 2D-Fluoreszenzspektren gefärbt nach der Gesamtproteinkonzentration c_{PlotM}

Bild 7.27 zeigt die Vorhersage von c_{PlotM} mit 2D-Fluoreszenzspektroskopie aufgetragen gegen ihre Referenzmessungen. Es ist ein linearer Zusammenhang zu erkennen.

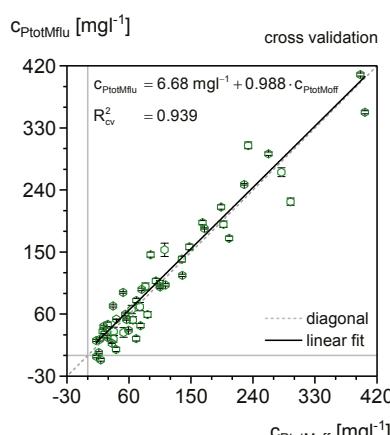


Bild 7.27: Vorhersage c_{PlotMflu} aufgetragen gegen die Referenzwerte c_{PlotMoff}

In Bild 7.28 ist dagegen die Vorhersage c_{PlotMfu} im Kultivierungsverlauf dargestellt.

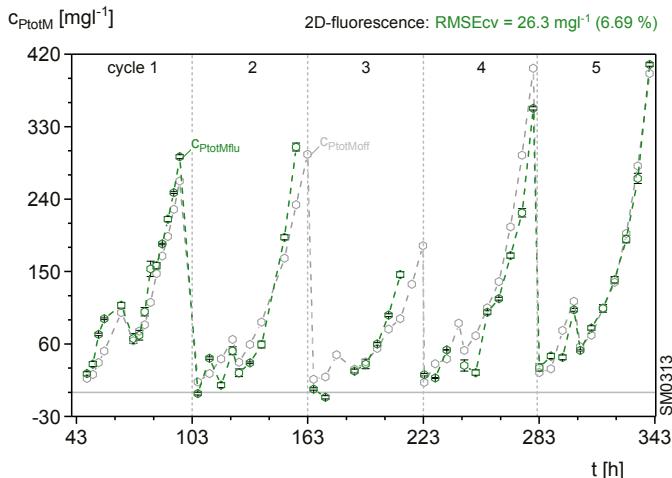


Bild 7.28: Vorhersage von c_{PlotM} im Überstand mit 2D-Fluoreszenzspektroskopie

Auffällig sind geringere Schwankungen der Mehrfachmessungen und kleinere Fehler bei hohen Konzentrationswerten im Vergleich zur Vorhersage mit Raman-Spektroskopie (Bild 7.25). Eine Schwäche der Trainingsdaten ist ein Mangel an hohen Gesamtproteinkonzentrationen. Dies ist durch die Prozessdynamik bei konstanter Probenahmefrequenz begründet.

Ein Vergleich der Modelleinstellungen und Vorhersageergebnisse der drei untersuchten spektroskopischen Verfahren ist in Tabelle 7.7 gegeben.

Tabelle 7.7: Vergleich spektroskopischer Verfahren zur Vorhersage von c_{PlotM} im Überstand

	Nahinfrarot	Raman	2D-Fluoreszenz
Spektralbereich	1300 - 1650 nm	1800 - 200 cm ⁻¹	komplett
Vorverarbeitung	de1 + mc	mc	mc
$n_{\text{CS}}^1)$	[\cdot]	186 (62)	183 (61)
r	[\cdot]	4	4
R_X^2	[\cdot]	0,999	0,999
R_Y^2	[\cdot]	0,778	0,858
R_{cv}^2	[\cdot]	0,811	0,887
RMSEcv	[mg l^{-1}]	49,0	40,5
RMSEcv _{rel}	[%]	12,5	10,3
			26,3
			6,69

¹⁾ Die Klammern geben die Anzahl unabhängiger Proben an

7.9 Prädiktion des zellinternen Alkoholoxidasegehaltes

Die zellspezifische AOX-Aktivität $g_{P2/X}$ gibt Aufschluss über den Induktionszustand der Zellen von *Pichia pastoris*. Eine Online-Quantifizierung von $g_{P2/X}$ ist daher für die Prozessentwicklung von großem Interesse. Aus diesem Grund wurden PLSR-Modelle mit 2D-Fluoreszenzspektroskopie erstellt. Auf eine Darstellung von NIR- und Raman-Spektroskopie wird verzichtet, da diese Messverfahren keine validen Modelle hervorbrachten.

Zunächst sollte geklärt werden, ob eine Modellbildung prinzipiell möglich ist und 2D-Fluoreszenzspektren von Zellsuspensionen Signale zur Quantifizierung von $g_{P2/X}$ enthalten.

Dazu wurden Zellpellets der Proben mit Phosphatpuffer auf eine einheitliche Zelldichte c_{XB} von $2,5 \text{ g l}^{-1}$ in der Pufferphase B eingestellt und offline vermessen. Da es sich bei dem AOX-Gehalt $g_{P2/X}$ um eine zellspezifische Größe handelt, bleibt diese dabei konstant. Eine Beeinflussung der Modellbildung durch unterschiedliche Zelldichten wird auf diese Weise vermieden.

Bild 7.29 zeigt die Prädiktion von $g_{P2/X}$ für die derart angepassten (*adjusted*) Proben.

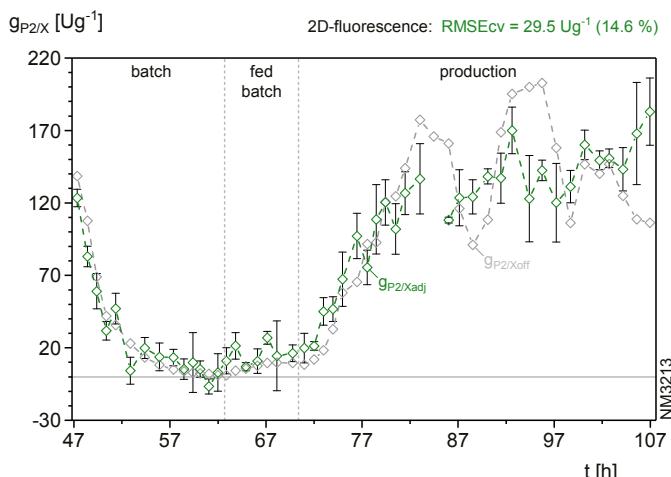


Bild 7.29: Vorhersage von $g_{P2/X}$ der angepassten Proben (adj) im Puffersystem B

Das Modell erreicht einen relativ hohen Vorhersagefehler $\text{RMSE}_{\text{cv},\text{rel}}$ von 14,6 %. Dennoch kann der prinzipielle Verlauf des AOX-Gehaltes $g_{P2/X}$ gut wiedergegeben werden. Indizien für ein *Overfit* gibt es nicht. Daraus kann geschlussfolgert werden, dass eine Offline-Quantifizierung des zellinternen Enzymgehaltes $g_{P2/X}$ prinzipiell möglich ist.

Im nächsten Schritt sollte ein PLSR-Modell zur Prädiktion von $g_{P2/X}$ direkt aus den originären Messungen in der Flüssigphase L erstellt werden.

Zu diesem Zweck wurde eine virtuelle AOX-Aktivität C_{P2Loft} in der Flüssigphase,

$$C_{P2Loft}(t) = c_{XLcdw}(t) \cdot g_{P2/Xoff}(t), \quad (7.1)$$

definiert, die aus den Offline-Daten der zellspezifischen AOX-Aktivität $g_{P2/Xoff}$ sowie der entsprechenden Zelldichte c_{XLcdw} berechnet werden kann.

Diese theoretische Größe wurde der nachfolgenden Modellbildung unterzogen. Ein Nachteil dabei ist, dass C_{P2L} und die Zelldichte c_{XL} mit einem Koeffizienten R von 0,733 relativ stark korrelieren (vgl. Abschnitt 7.4.3). Die direkte Vorhersage von C_{P2L} ist in Bild 7.30 gezeigt.

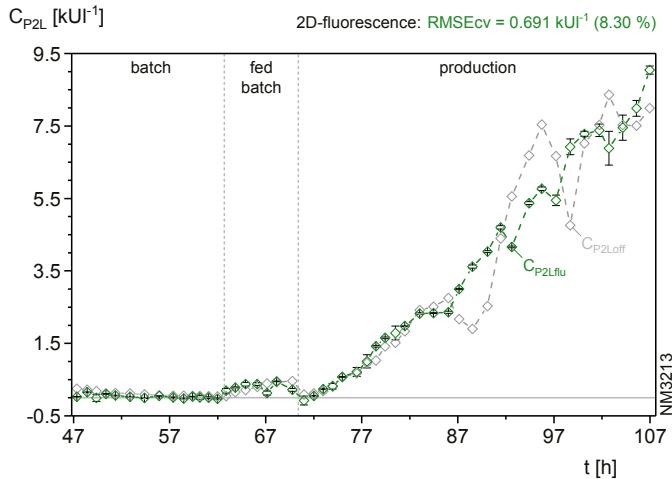


Bild 7.30: Vorhersage C_{P2Lflu} der virtuellen AOX-Aktivität in der Flüssigphase L

Die in Bild 7.31 gezeigte Vorhersage $g_{P2/Xflu}$ der zellspezifischen AOX-Aktivität,

$$g_{P2/Xflu}(t) = \frac{C_{P2Lflu}(t)}{c_{XLflu}(t)}, \quad (7.2)$$

kann dann aus den Vorhersagen C_{P2Lflu} und c_{XLflu} (siehe Abschnitt 7.5) leicht berechnet werden.

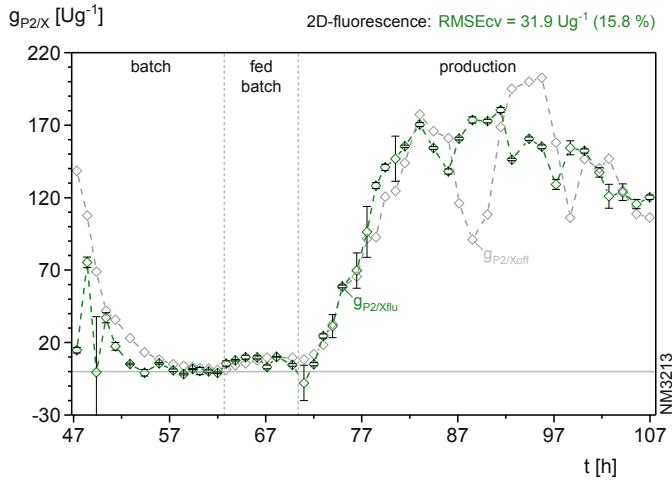


Bild 7.31: Vorhersage von $g_{P2/X}$ aus Fluoreszenzmessungen in der Kulturbrühe L

Der AOX-Gehalt wird zu Beginn des Prozesses nur schlecht vorhergesagt. Auch die Abweichungen der Mehrfachmessungen sind dort sehr hoch. Dies ist vermutlich darauf zurückzuführen, dass die Spektren durch niedrige Zelldichten in dieser Prozessphase nur geringe Signale zur Quantifizierung von $g_{P2/X}$ enthalten.

Der weitere Verlauf des Prozesses zeigt eine bessere Übereinstimmung mit den gemessenen Werten. Die starken Schwankungen bei ca. 90 h können nicht wiedergegeben werden. Es ist nicht auszuschließen, dass es sich dabei um Messfehler der Referenzwerte handelt.

Insgesamt kann ein relativer Vorhersagefehler $RMSEcv_{rel}$ von 15,8 % ermittelt werden. Dieser Wert liegt nur geringfügig oberhalb des Fehlers bei der Vorhersage von $g_{P2/X}$ aus den angepassten Proben (Bild 7.29). Im Vergleich dazu fallen außerdem die Abweichungen der Mehrfachmessungen weniger stark aus.

Besonders hervorzuheben ist, dass ein leichter Anstieg des AOX-Gehaltes in der Glycerol-Fed-Batchphase richtig vorhergesagt wird. Dieser kommt durch eine Dereprimierung des AOX-Promotors bei einer Glycerollimitierung zustande. Eine Vorhersage dieses Verhaltens lässt auf eine gewisse Sensitivität dieses Bestimmungsverfahrens schließen.

In Tabelle 7.8 sind die Ergebnisse der Vorhersage von $g_{P2/X}$ noch einmal zusammengefasst.

Tabelle 7.8: Vergleich verschiedener Ansätze zur Vorhersage der AOX-Aktivität $g_{P2/X}$

Messmatrix	Puffer B	Flüssigphase L	berechnet aus L
Größe	$g_{P2/Xadj}$	C_{P2Lflu}	$g_{P2/Xflu}$
Spektralbereich	komplett	komplett	-
Vorverarbeitung	mc	mc	-
$n_{cs}^1)$	[\cdot]	147 (49)	150 (50)
r	[\cdot]	3	4
R_x^2	[\cdot]	0,826	0,999
R_y^2	[\cdot]	0,883	0,955
R_{cv}^2	[\cdot]	0,796	0,938
$RMSEcv$	[div.]	$29,5 \text{ Ug}^{-1}$	$0,691 \text{ kU l}^{-1}$
$RMSEcv_{rel}$	[%]	14,6	8,30
			15,8

¹⁾ Die Klammern geben die Anzahl unabhängiger Proben an

8. Online-Monitoring mit spektroskopischen Verfahren

8.1 Der untersuchte zweistufige Produktionsprozess

8.1.1 Die verwendete Bioreaktoranlage

Im Cluster *BIOKATALYSE 2021* wurde im Rahmen eines BMBF-Forschungsprojektes

„Entwicklung vollautomatischer integrierter Bioprozesse in einer industrielkompatiblen Pilotanlage zur Herstellung rekombinanter Enzyme“

die in Bild 8.1 gezeigte Anlage zur Herstellung der rekombinanten Lipase CALB mit *Pichia pastoris* aufgebaut. In weiteren Forschungsprojekten wurde diese auch zur Herstellung des Malariavakzins D1M1H eingesetzt.



Bild 8.1: Downscale-Pilotanlage zur Herstellung rekombinanter Proteine

Es handelt sich dabei um eine hochinstrumentierte Forschungsanlage zur Entwicklung und Optimierung industrieller Produktionsprozesse im *Downscale*. Das Gesamtkonzept ist auf die Durchführung integrierter Bioprozesse ausgerichtet. Daher sind Komponenten für *Upstream*- sowie für *Downstream*-Operationen vorhanden.

Hierbei ist beispielsweise eine Zellabtrennung durch Separation vorgesehen (rechts im Bild). Die Aufreinigung rekombinanter Proteine war jedoch nicht Teil dieser Arbeit. Deshalb wird auf die *Downstream*-Operationen nicht näher eingegangen. Sie werden jedoch im Zuge der Entwicklung und Erprobung der Gesamtanlage ausführlich in [LÖGERING, 2015] und [BORCHERT, 2015] beschrieben.

Im *Upstream*-Bereich besteht die Anlage aus zwei Biorektoren, die über eine Rohrleitung fest miteinander verbunden sind. Hierüber wird der Transfer zur Durchführung einer zweistufigen Prozessführung mittels Überdruck im ersten Kessel realisiert. Bild 8.2 zeigt die Kopplung der beiden Edelstahlreaktoren über ein Ernteventil.

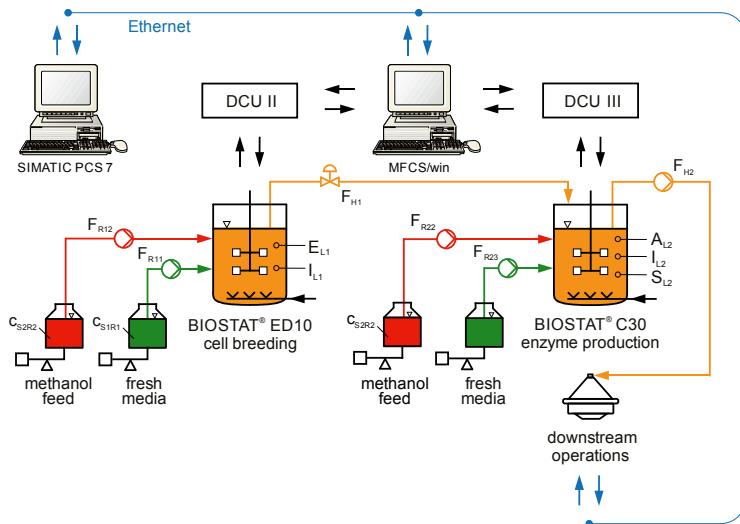


Bild 8.2: Schematische Darstellung des Anlagenaufbaus

Ein BIOSTAT® ED10 der Fa. B. Braun Biotech International, Melsungen, dient mit einem Arbeitsvolumen von 10 l als Zellanzuchtbioreaktor. Für die Produktion kommt ein BIOSTAT® C30 der Fa. Sartorius Stedim Biotech, Guxhagen, der bis zu 30 l beinhalten kann, zum Einsatz.

Die Ausstattung beider Reaktorsysteme entspricht der in Abschnitt 3.2 vorgestellten Ausstattung des BIOSTAT® C10. Eine detaillierte Auflistung der verwendeten peripheren Gerätschaften befindet sich in [BORCHERT, 2015].

Die Grundautomatisierung der Bioreaktoren erfolgt über digitale Kontrolleinheiten (DCU II, DCU III) und MFCS/win. In dem SCADA-System wurden außerdem Rezepte nach dem ANSI/ISA Standard S88 zur Automatisierung der zyklischen Teilprozesse der zweistufigen Produktion hinterlegt.

Für eine übergeordnete Automatisierung und insbesondere die Einbindung der Komponenten für *Downstream*-Operationen in den Gesamtprozess wurde weiterhin das Prozessleitsystem SIMATIC PCS 7, Siemens, München, implementiert [LÖGERING, 2015].

8.1.2 Die parallel/sequentielle Prozessführung

Die parallel/sequentielle Prozessführung basiert auf der in Abschnitt 3.1.2 vorgestellten zyklischen Kultivierungsstrategie. Im Gegensatz dazu wurde hier jedoch auf die Durchführung eines substratlimitierten Glycerol-Fed-Batches verzichtet, da eine einfache industrierelevante Prozessführung das Ziel der Entwicklungsarbeiten war [LÖGERING, 2015].

Zur Erläuterung der zweistufigen Fahrweise ist in Bild 8.3 ein vollständiger Kulturzyklus einer längeren Produktionskampagne zur Herstellung des Malariavakzins D1M1H dargestellt. Er besteht aus einer sequentiellen Zellanzucht mit Vorinduktion und einer Proteinexpression, die voneinander entkoppelt in zwei Bioreaktoren ablaufen [BORCHERT, 2015].

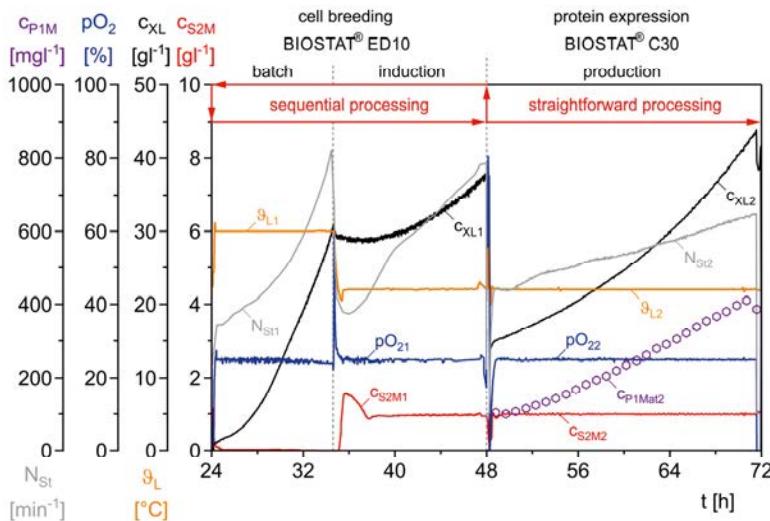


Bild 8.3: Vollständiger Produktionszyklus der zweistufigen Prozessführung

Der Zellanzuchtprozess findet im Bioreaktor BIORSTAT® ED10 statt. Hier wachsen die Zellen zunächst unlimitiert auf dem Primärsubstrat Glycerol an. Nach Abschluss der etwa elfstündigen Batchphase erfolgt durch eine Methanolzufütterung die Umstellung des Zellmetabolismus auf den Methanolstoffwechsel. Diese auch als Vorinduktion bezeichnete Prozessphase ist an dem anfänglich fehlenden Zellwachstum zu erkennen.

Nach Abschluss eines solchen Zellanzuchzyklus wird ca. 80 % der Kulturbrühe vom BIORSTAT® ED10 auf den Produktionsreaktor BIORSTAT® C30 transferiert. Hier wird zunächst die gewünschte Startzelldichte c_{XL20} mit frischem Basismedium ohne Glycerol und ohne Methanol eingestellt. Die folgende Produktionsphase dient ausschließlich der Expression des Zielproteins D1M1H und verläuft mit bereits vollständig induzierten Zellen in einem methanolgeregelten Fed-Batchprozess.

Die verbliebenen 20 % des Zellmaterials im Anzuchtreaktor werden mit frischem glycerolhaltigen Medium auf die Startzelldichte c_{XL10} verdünnt und als Inokulum des nächsten Zellanzuchzyklus verwendet. Dies entspricht dem bekannten Vorgehen bei einer sequentiellen Fahrweise. Die aktuelle Zellanzucht und der Produktionsprozess laufen nun parallel in beiden Reaktoren ab. Dies ist in Bild 8.4 illustriert.

Vor jedem Zyklus wird der Produktionsbioreaktor vollständig entleert und die so geerntete Kulturbrühe dem Downstream zugeführt. Damit wird der Beginn eines weiteren Produktionslaufes mit frischen Zellen aus dem Anzuchtreaktor vorbereitet. Durch die Wahl einer Zyklusdauer von 24 Stunden, inklusive Zelltransfer- und Refresh-Operationen, wiederholen sich die parallelen Zyklen dann täglich. Eine detaillierte Beschreibung der zweistufigen Prozessführung findet sich in [LÖGERING, 2015].

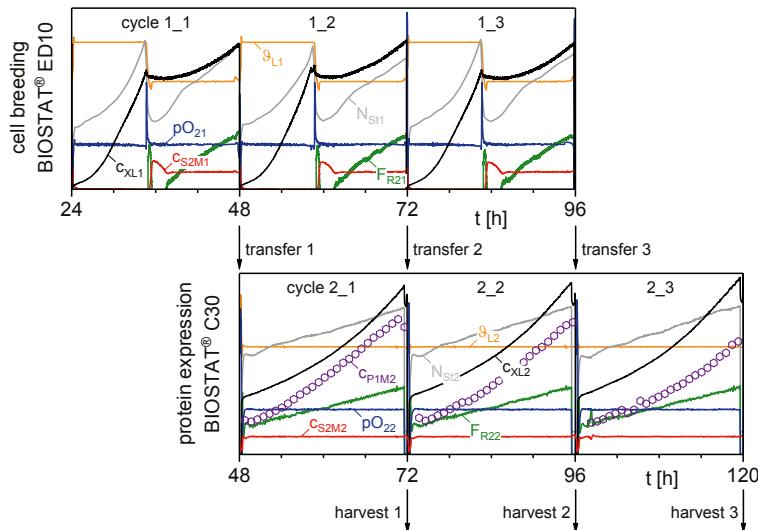


Bild 8.4: Zweistufiger, paralleler Produktionsprozess [LÖGERING, 2015]

Die Kultivierungsstrategie zeichnet sich durch eine hohe Reproduzierbarkeit aus, die eine Produktion mit gleichebleibender Qualität über einen Zeitraum von mehreren Wochen erlaubt [BORCHERT, 2015]. Es wird außerdem eine hohe Flexibilität erreicht, da durch die Entkopplung der Teilprozesse und mit unveränderter Zellanzucht, stamm- oder produktsspezifische Produktionsbedingungen im zweiten Reaktor erprobt werden können.

Weiterhin eignet sich die Prozessführung zur Optimierung der Proteinexpression, da durch die sequentielle Zellanzucht zu Beginn eines jeden Produktionszyklus gleiche Ausgangsbedingungen vorliegen [LÖGERING et al., 2011].

8.1.3 Die erweiterte Prozess-EDV zur Anwendung der MVDA

Die Online-Anwendung der MVDA, beispielsweise zur Vorhersage von Substratkonzentrationen aus Spektren, stellt einige Anforderungen an die Prozess-EDV. So müssen benötigte Eingangsdaten multivariater Modelle in einem geeigneten Format bereitgestellt und eine Rückführung erzeugter Ergebnisse an die Prozessleittechnik sichergestellt werden.

Zu diesem Zweck wurde in vorangegangenen Arbeiten eine komplexe Datenverarbeitung aufgebaut, deren Struktur für den gesamten integrierten Bioprozess in Bild 8.5 gezeigt ist. Dabei wurde die bereits vorhandene Prozess-EDV sowohl um benötigte Softwarepakete als auch um neue Computerhardware ergänzt [BORCHERT, 2015].

Die beteiligten Geräte stammen von verschiedenen Herstellern, weisen ein unterschiedliches Alter auf und wurden zu unterschiedlichen Zeitpunkten in das System integriert. Der Aufbau des Netzwerkes ist daher historisch gewachsen und erfordert die Nutzung unterschiedlicher Anschlusstechnologien, wie z.B. Ethernet, ProfiBus und die seriellen Schnittstellen RS232 und RS422.

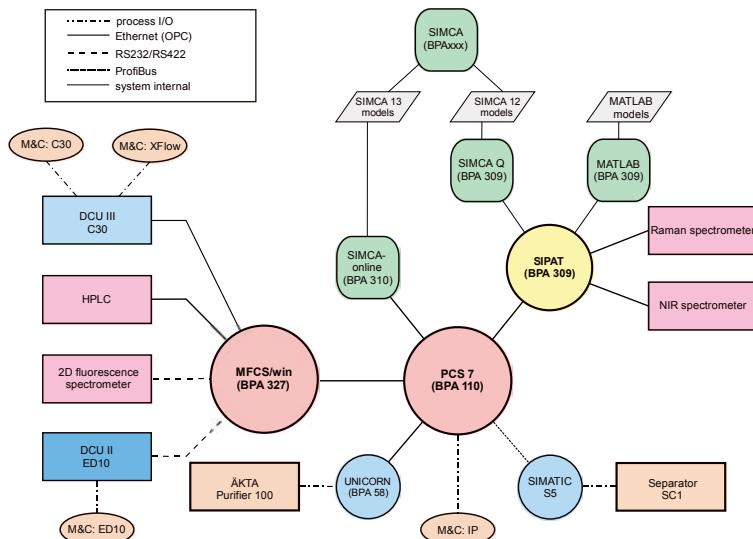


Bild 8.5: Vernetzung der erweiterten Prozess-EDV [BORCHERT, 2015]

Auf der linken Seite der Abbildung befinden sich die Mess- und Stelleinheiten (M&C) der beiden Bioreaktoren C30 und ED10, die neben der Filtrationseinheit XFlow an ihre jeweilige digitale Kontrolleinheit DCU II bzw. DCU III angeschlossen sind.

Zusammen mit einer Atline-HPLC und einem 2D-Fluoreszenzspektrometer sind die beiden DCU mit dem Rechner BPA 327 verbunden, der hauptsächlich für den Betrieb des SCADA-Systems MFCS/win zur Datenerfassung und einer Automatisierung der beiden Teilprozesse des *Upstreams* verantwortlich ist.

Im Zentrum des Netzwerks befindet sich auf dem Rechner BPA 110 das übergeordnete Prozessleitsystem SIMATIC PCS 7, welches den *Upstream*-Prozess mit den *Downstream*-einheiten Separator und Äcta Purifier (unten im Bild) zusammenführt und damit eine Integration (IP) des Bioprozesses erlaubt. Weiterhin wird über das Labornetzwerk (Ethernet) mit Hilfe von PCS 7 eine Kommunikation verschiedener beteiligter Rechnersysteme mit dem OPC-Protokoll ermöglicht.

Eine detaillierte Darstellung der bis zu diesem Punkt beschriebenen Prozess-EDV befindet sich in [LÖGERING, 2015].

Entscheidend für die Einbindung der MVDA ist die speziell für PAT-Anwendungen entwickelte Software SIMATIC SIPAT, Siemens, München, die auf dem Prozessrechner BPA 309 installiert wurde. SIPAT sammelt Daten unterschiedlicher Quellen und legt diese in einem MVDA-gerechten Format in einer Datenbank ab. Unter Verwendung individueller Treiber und Schnittstellen ist dabei auch die Verarbeitung spektraler Daten (z.B. Raman, NIR) unterschiedlicher Gerätehersteller möglich, die oft verschiedene Strategien bei der Datenverarbeitung verfolgen.

Weiterhin ist SIPAT dazu in der Lage, modellgestützte Berechnungen durch externe Software-systeme auszulösen. Unterstützt werden beispielsweise die MVDA-Software SIMCA, MKS

data analytical solutions, Umeå, Schweden, sowie die numerische Entwicklungsumgebung MATLAB®, The Mathworks, Natick, USA.

Die dafür benötigten Modelldateien werden auf Arbeitsplatzrechnern (BPA xxx) erstellt und in SIPAT eingebettet, wo über sogenannte Methoden eine Steuerung und Automatisierung der Berechnungen umgesetzt wird. Durch eine Verteilung über OPC-Anbindungen werden damit erzeugte Werte im Labornetzwerk zur Verfügung gestellt und können durch die Prozessleittechnik beispielsweise für *Monitoring*- und *Control*-zwecke genutzt werden.

Auf einem weiteren Rechner (BPA 310) ist die Software SIMCA-online installiert, die eine Regelung von Prozessqualitätsattributen durch *Model Predictive Multivariate Control* (MPMC) erlaubt. Dies war nicht Teil dieser Arbeit und ist in [LUTTMANN et al., 2015] beschrieben.

In [BORCHERT, 2015] befindet sich darüber hinaus eine detaillierte Beschreibung von SIPAT sowie eine Dokumentation von dessen Einbindung in die Prozess-EDV inklusive der Verwendung von SIMCA und MATLAB®.

8.2 Skizzierung des Versuchsaufbaus

Das Online-Monitoring mit spektroskopischen Messverfahren wurde in der oben beschriebenen zweistufigen Anlage erprobt. Es standen je ein NIR- und 2D-Fluoreszenzspektrometer mit Einzelsonden sowie ein Raman-Gerät mit zwei Sonden zur Verfügung (siehe Abschnitt 4.2.2). Bild 8.6 zeigt den Versuchsaufbau aus Sicht des Produktionsbioreaktors der zweistufigen Anlage.

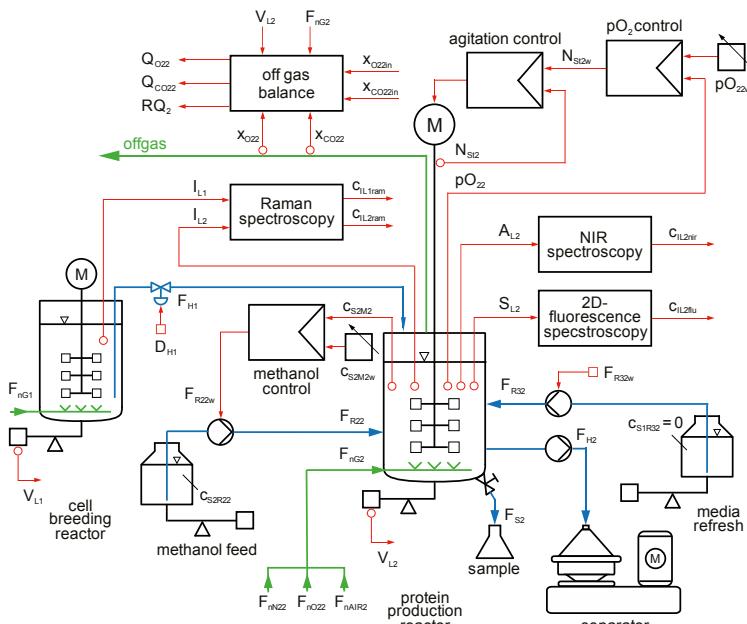


Bild 8.6: Spektroskopische Verfahren im zweistufigen *Upstream*-Prozess

Die erste Raman-Sonde kam im Zellanzuchtreaktor zum Einsatz. Die zweite Sonde sowie die NIR- und 2D-Fluoreszenzsonden wurden dagegen im Produktionsreaktor verbaut. Im Gegensatz zu den Offline-Analysen der Machbarkeitsstudie (siehe Kapitel 7) wurden die Spektrometer bei diesen Arbeiten inline betrieben. Eine Mehrfachmessung einzelner Proben entfällt dadurch.

Die Messeinstellungen der verwendeten Spektrometersysteme ergaben sich aus Erfahrungswerten der Machbarkeitsstudie sowie weiterer Vorversuche. Weiterhin richten sich die Messzeiten nach der angestrebten Datenaufzeichnungsfrequenz, die für die PAT-Software SIPAT mit drei Minuten festgelegt wurde.

Für das Raman-Gerät hat sich hinsichtlich der Sättigung des Detektors eine optimale Integrationszeit von 35 Sekunden herausgestellt. Durch die Wahl von fünf Akkumulationen und unter Berücksichtigung eines gewissen Zeitüberhangs für die Datenverarbeitung wurde das dreiminütige Messintervall nicht überschritten. Bei parallelem Betrieb beider Sonden verdoppelt sich dieses jedoch auf etwa sechs Minuten.

Für das NIR-Spektrometer wurde eine Integrationszeit von 15 Millisekunden gewählt. Die Durchführung von 10.000 Akkumulationen stellt ein sehr gutes Signal-zu-Rausch-Verhältnis her und führt zu einem Messintervall von 2,5 Minuten, was ebenfalls Raum für die nötige Datenverarbeitung lässt.

Eine Messung mit dem 2D-Fluoreszenzspektrometer nimmt in etwa 1,3 Minuten in Anspruch. Mit zwei Akkumulationen zwecks Signalglättung wurde auch hier das angestrebte Messintervall von drei Minuten nicht überschritten.

8.3 Prädiktion der Zelldichte und der Glycerolkonzentration

8.3.1 Bereitstellung geeigneten Datenmaterials

Der Zellanzuchtprozess RL0415 wurde für die Modellentwicklung zur Vorhersage der Zelldichte c_{XL1} und der Glycerolkonzentration c_{S1L1} verwendet. Auf Grund der Tatsache, dass der Zellanzuchtreaktor nur über einen Anschluss einer Spektralmessung verfügt, kam für die Vorhersage der beiden Zielgrößen ausschließlich die Raman-Spektroskopie zum Einsatz.

Zur Erzeugung geeigneter Referenzmessungen wurden die fünf in Bild 8.7 gezeigten Zyklen intensiv beprobt und auf einen Kalibrierdatensatz CS, bestehend aus den Zyklen 1_0, 1_4 und 1_6 sowie einen Validierdatensatz VS aufgeteilt, der die Zyklen 1_2 und 1_3 beinhaltet.

Bei den Zyklen 1_2 und 1_6 handelt es sich um normale Zellanzuchtyziken, wie sie gemäß der in Abschnitt 8.1.2 vorgestellten Fahrweise vorgesehen sind. Zyklus 1_0 folgt dagegen direkt dem Animpfen des Bioreaktors (*Startup-Zyklus*) und unterscheidet sich daher von einem normalen Zyklus in der anfänglichen Glycerolkonzentration c_{S1L10} , der Startzelldichte c_{XL10} sowie in der Dauer der durchlaufenen Prozessphasen.

Einen besonderen Ablauf zeigen die Zyklen 1_3 und 1_4. Mit dem Ziel die Korrelation der beiden untersuchten Variablen zu reduzieren (siehe Abschnitt 7.4.3) wurden hier Glycerol-Spikes durchgeführt. Das heißt, dass dem laufenden Prozess eine definierte Menge Glycerol zugegeben wurde, was zu einem starken Anstieg (*Spike*) der Glycerolkonzentration c_{S1L1} führte.

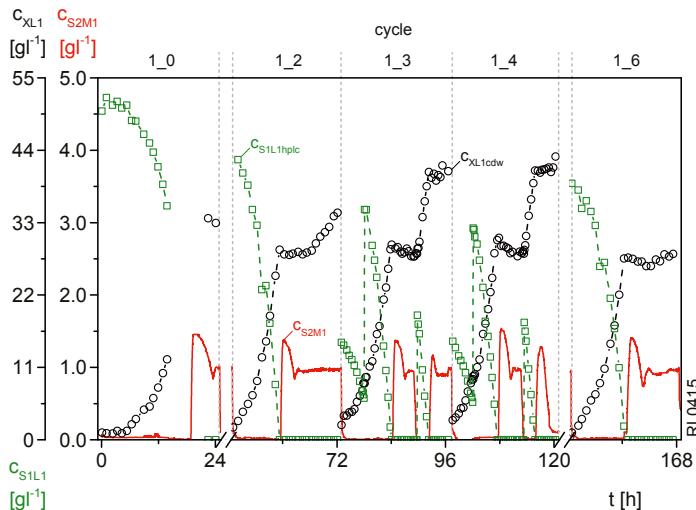


Bild 8.7: Verlauf von fünf Zyklen der Kultivierung RL0415

Der erste *Spike* wurde während der Batchphase ca. fünf Stunden nach Beginn des jeweiligen Zyklus durchgeführt. Die insgesamt eingesetzte Menge an Glycerol entspricht der eines normalen Zyklus. Damit wurde die Dauer der Batchphase nicht verändert.

Der zweite *Spike* fand ca. 17 h nach Zyklusbeginn während der Induktionsphase statt. Hierzu wurde zunächst die Methanolregelung deaktiviert und der Spike erst nach vollständiger Metabolisierung des verbliebenen Methanols durchgeführt. Nach Verbrauch des zugegebenen Glycerols erfolgte eine Reaktivierung der Methanolzufütterung.

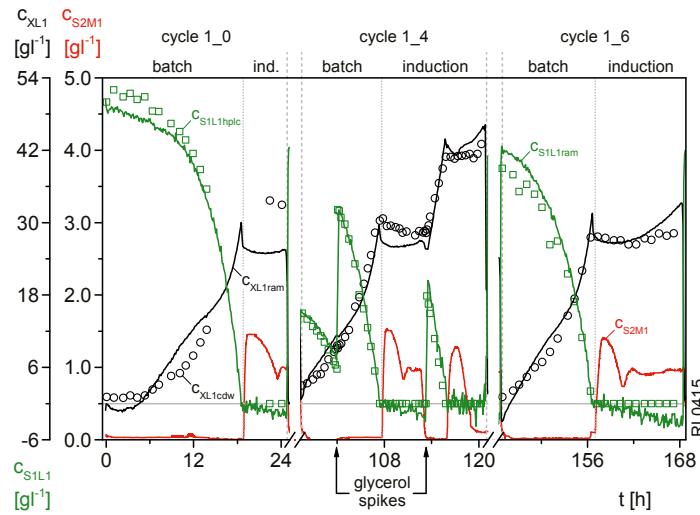
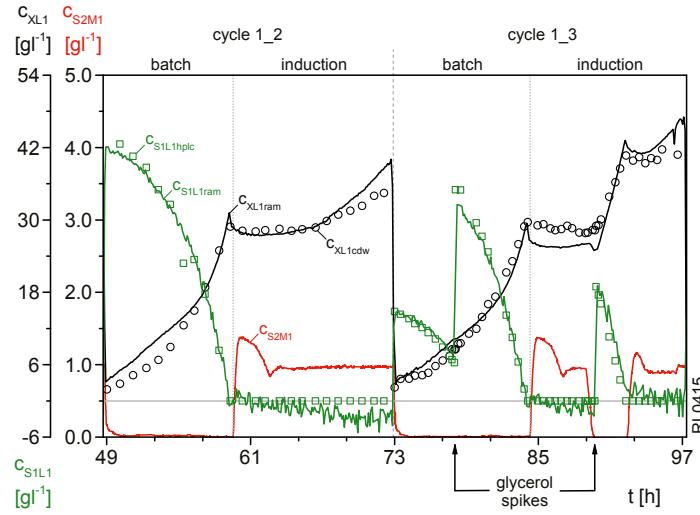
8.3.2 Ergebnisdarstellung

Für das Online-Monitoring von Zelldichte c_{XL1} und Glycerolkonzentration c_{S1L1} mittels Raman-Spektroskopie wurden jeweils die in der Machbarkeitsstudie ermittelten Einstellungen der PLSR-Modelle verwendet. Diese konnten als optimal für den Online-Betrieb bestätigt werden. Lediglich die Anzahl berechneter PLS-Komponenten unterscheidet sich jeweils für beide betrachteten Zielgrößen.

Die absoluten Vorhersagefehler für das Online-Modell von c_{XL1} sind niedriger als die des Offline-Modells. Aufgrund eines kleineren Kalibrierbereichs der Online-Daten fallen die relativen Fehler jedoch höher aus.

Beim Online-Modell von c_{S1L1} sind auch die absoluten Vorhersagefehler höher. Dies war zu erwarten, da bei spektroskopischer Inline-Messung mit Signalstörungen zu rechnen ist. Diese beeinträchtigen die Vorhersagekraft multivariater Modelle.

Bild 8.8 zeigt die Vorhersagen im Verlauf der drei Zyklen des Kalibrierdatensatzes, während in Bild 8.9 das Ergebnis der externen Validierung dargestellt ist. Der prinzipielle Verlauf beider Prozessgrößen kann wiedergegeben werden. Hohe Glycerolkonzentrationen und Werte von 0 gl^{-1} werden allerdings nicht gut vorhergesagt.

Bild 8.8: Prädiktion von c_{XL1} und c_{S1L1} für den Kalibrierdatensatz CSBild 8.9: Prädiktion von c_{XL1} und c_{S1L1} für den Validierdatensatz VS

Schlechter ist jedoch die Vorhersage der Zelldichte zu bewerten, dessen Verlauf nicht der wahren Prozessdynamik (z.B. exponentielle Zunahme) entspricht.

Dies ist auch anhand von Bild 8.10a zu erkennen. Die vorhergesagte Zelldichte $c_{XL1fram}$ aufgetragen gegen die Referenzwerte c_{XLcdw} weist insbesondere bei niedrigen Werten einen nicht-linearen Verlauf auf. Dies wurde schon in der Machbarkeitsstudie (Abschnitt 7.5) beobachtet.

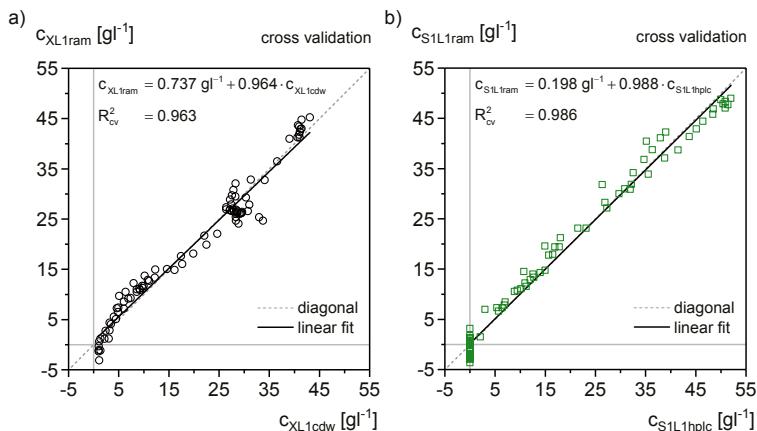


Bild 8.10: Mit PLSR vorhergesagte Zielgrößen aufgetragen gegen ihre Referenzwerte

a) Zelldichte c_{XL1ram} b) Glycerolkonzentration $c_{S1L1ram}$

Bild 8.10b verdeutlicht dagegen, dass die Raman-Spektroskopie in Verbindung mit PLSR besser für die Vorhersage der Glycerolkonzentration c_{S1L1} geeignet ist. So lässt sich hier für den gesamten betrachteten Konzentrationsbereich ein linearer Zusammenhang herstellen.

In Tabelle 8.1 sind abschließend die Einstellungen sowie die Kenngrößen der erzeugten Online-PLSR-Modelle den entsprechenden Werten der Offline-Modelle gegenübergestellt. Positiv hervorzuheben ist, dass für beide Zielgrößen ein besseres Ergebnis bei der externen Validierung erzielt werden konnte. Dies macht einen *Overfit* eher unwahrscheinlich.

Tabelle 8.1: Offline- und Online-Vorhersage von c_{XL1} und c_{S1L1} mit Raman-Spektroskopie

		Zelldichte c_{XL1}		Glycerolkonzentration c_{S1L1}	
		offline	online	offline	online
$n_{cs}^1)$	[·]	150 (50)	90 (90)	150 (50)	90 (90)
r	[·]	3	2	2	3
R_x^2	[·]	0,972	0,976	0,951	0,972
R_y^2	[·]	0,975	0,964	0,992	0,987
R_{cv}^2	[·]	0,971	0,963	0,998	0,986
RMSEcv	[$g l^{-1}$]	3,23	2,54	1,79	2,07
RMSEcv _{rel}	[%]	4,40	6,03	3,32	3,98
$n_{vs}^1)$	[·]	-	74 (74)	-	69 (69)
R_p^2	[·]	-	0,970	-	0,978
RMSEP	[$g l^{-1}$]	-	2,11	-	1,85
RMSEP _{rel}	[%]	-	5,01	-	3,56

¹⁾ Die Klammern geben die Anzahl unabhängiger Proben an

8.4 Prädiktion des zellinternen Alkoholoxidasegehaltes

8.4.1 Bereitstellung geeigneter Kalibrierdaten

In Abschnitt 7.9 wurde die Möglichkeit aufgezeigt, den zellinternen AOX-Gehalt $g_{P2/X}$ mit 2D-Fluoreszenzspektroskopie offline zu quantifizieren. An dieser Stelle soll eine Übertragung des dort entwickelten Modells auf den Online-Betrieb erprobt werden.

Dazu wurde in den in Bild 8.11 gezeigten Zyklen 2_2 und 2_3 der Kultivierung XX0415 im Produktionsbioreaktor jeweils zu Beginn einer Phase unlimitierten Wachstums auf Glycerol eingerichtet, um eine Abnahme des zellinternen AOX-Gehaltes zu bewirken. Damit sollte die für eine PLSR-Modellerstellung benötigte Varianz in den Daten erzeugt werden.

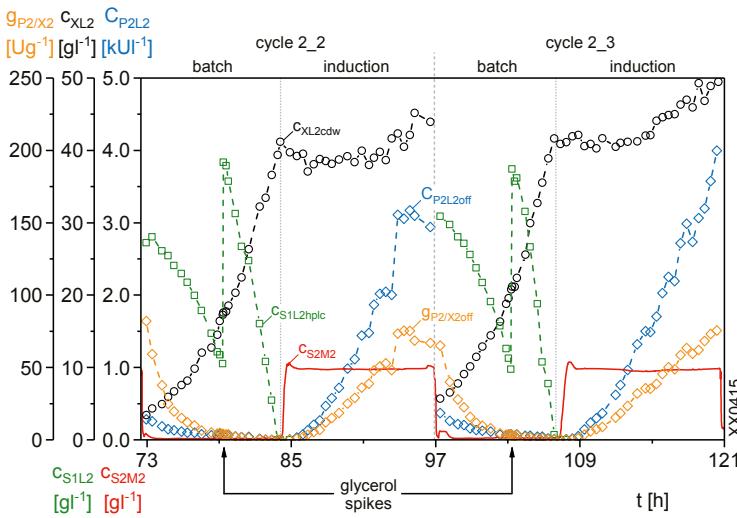


Bild 8.11: Prozesszyklen zur Bildung von PLSR-Modellen für die Quantifizierung von $g_{P2/X}$

Wie in Abschnitt 7.9 beschrieben, kam auch hier als Zielgröße die virtuelle AOX-Aktivität C_{P2L2} in der Flüssigphase zum Einsatz, die aus dem zellinternen AOX-Gehalt $g_{P2/X}^{off}$ und der Zelldichte c_{XL2cdw} berechnet wurde.

8.4.2 Ergebnisdarstellung

Für die Vorhersage der AOX-Aktivität C_{P2L2} in der Flüssigphase konnte kein valides Online-Modell erstellt werden.

Für den relativen Vorhersagefehler $RMSEcv_{rel}$ wurde mit 6,62 % zwar ein niedrigerer Wert als für das Offline-Modell (8,30 %, siehe Abschnitt 7.9) erzielt, jedoch zeigt die in Bild 8.12 dargestellte Vorhersage einen unerwarteten Verlauf, der eine schlechte Anpassung an die experimentellen Daten darstellt.

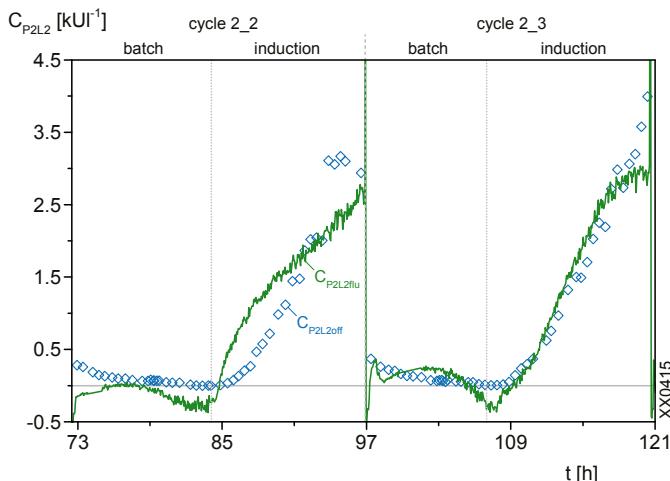


Bild 8.12: Online-Prädiktion von C_{P2L2} in der Flüssigphase mit 2D-Fluoreszenz

Bis auf die Anzahl verwendeter PLS-Komponenten stimmen die Modelleinstellungen für die Offline- und Online-Anwendungen überein. Eine Änderung der Einstellungen des Online-Modells führt stets zu einer weiteren Verschlechterung des Ergebnisses.

Daraus muss der Schluss gezogen werden, dass eine Online-Quantifizierung des zellinternen AOX-Gehaltes nicht in dieser Form erfolgen kann. Als mögliche Ursachen können Störungen im Online-Betrieb (Begasung, Durchmischung) oder auch ein *Overfit* im Offline-Modell mit indirekter Quantifizierung genannt werden.

Auf eine Berechnung des AOX-Gehaltes $g_{P2/X2flu}$ aus der Vorhersage der virtuellen AOX-Aktivität $C_{P2L2flu}$ in der Flüssigphase wurde daher verzichtet, da weitere Fehler bei der Vorhersage eines Online-Signals der Zelldichte eine Verschlechterung des Ergebnisses erwarten lässt.

8.5 Prädiktion der Gesamtproteinkonzentration

Eine Online-Prädiktion der Gesamtproteinkonzentration c_{PtotL2} war leider auch nicht erfolgreich.

Mit der Durchführung von Spikes mit bovinem Serumalbumin (BSA), analog zum Vorgehen bei der Online-Vorhersage der Glycerolkonzentration c_{S1L1} (Abschnitt 8.3), wurde ein Datensatz erzeugt, der durch eine minimierte Korrelation von c_{PtotL} und c_{XL} möglichst valide Modelle hervorbringen sollte.

Keine der drei untersuchten spektroskopischen Verfahren war in der Lage, die damit erzeugten stufenförmigen Verläufe von c_{PtotL} abzubilden. Vielmehr wurde anhand der Vorhersagen deutlich, dass die Modelle trotz der oben beschriebenen Bemühungen von Signalen der Zellen in den Spektren der Kulturbrühe dominiert werden.

Eine Anwendung der Methode zur Prädiktion von sekretierten Proteinen erscheint daher nicht sinnvoll.

8.6 Anwendung der nichtlinearen SVR

Für die Untersuchung der *Support Vector Regression* (SVR) zur Vorhersage der Zelldichte c_{XL1} und Glycerolkonzentration c_{SIL1} kamen dieselben Datensätze (CS, VS) wie zur Bildung von PLSR-Modellen zum Einsatz (siehe Abschnitt 8.3).

Als spektrale Vorverarbeitungsmethode für beide Variablen diente das SNV-Filter. Auf eine zusätzliche Mittenzentrierung wurde verzichtet, da bei der SVR die Vorteile in Bezug auf die Modellinterpretierbarkeit entfallen. Die Prädiktion von c_{XL1} erfolgte mit den vollständigen Raman-Spektren, wohingegen sich für c_{SIL1} der Bereich $2.750 - 3.050 \text{ cm}^{-1}$ als sinnvoll erwies.

Der Ablauf der Modellbildung ist in den wesentlichen Schritten für beide Variablen gleich. Da-her wird an dieser Stelle nur die Modelloptimierung für die Zelldichte c_{XL1} dargestellt.

Zur Ermittlung möglicher Vorhersagefehler sowie sinnvoller Wertebereiche der Parameter γ , C und ε wurden einige Voruntersuchungen durchgeführt. Dabei wurde der Parameter ε zu- nächst auf einen beliebig gewählten Wert von 0,02 eingestellt und für die anderen beiden Parameter eine Rastersuche durchgeführt.

Hierfür erfolgte eine logarithmische Variation der Parameter γ und C in sehr großen Bereichen sowie eine Modellbildung für jede Kombination mit anschließender externer Validierung. Dadurch konnte der RMSEP in Abhängigkeit beider Parameter in einem *Contourplot* (siehe Bild 8.13) dargestellt werden. In diesem kann man erkennen, dass drei lokale Minima des Fehlers RMSEP in dem durch die zwei Modellparameter aufgespannten Suchraum existieren. Diese sind als Kreuze gekennzeichnet.

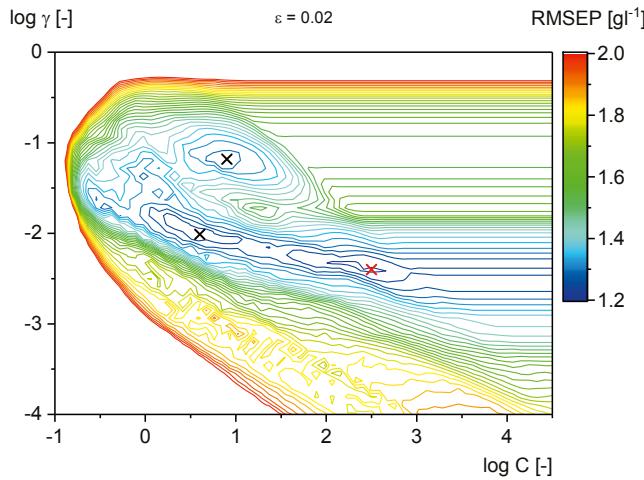


Bild 8.13: Contourplot des Vorhersagefehlers RMSEP für die Vorhersage der Zelldichte c_{XL1} mit SVR bei logarithmischer Variation der Parameter γ und C

Für die weitere Modellbildung wurde der Parametersatz mit dem niedrigsten Vorhersagefehler ausgewählt. Dieser liegt in Bild 8.13 auf der rechten Seite bei einem $\log \gamma$ von -2,4 und einem $\log C$ von 2,5.

In einem weiteren Schritt erfolgte mit diesen Werten und bei logarithmischer Variation des dritten Modellparameters ε eine in Bild 8.14 dargestellte Untersuchung der Vorhersagefehler RMSEcv und RMSEP sowie der Anzahl n_{sv} genutzter Stützvektoren.

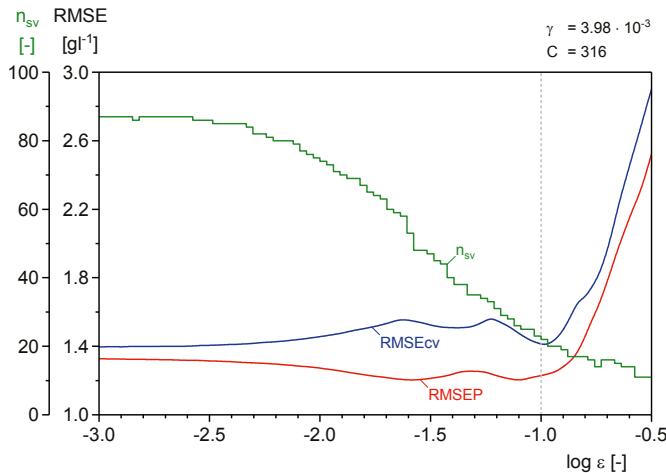


Bild 8.14: Vorhersagefehler RMSEcv und RMSEP sowie Anzahl n_{sv} benötigter Stützvektoren in Abhängigkeit des Parameters ε bei der Vorhersage von c_{XL1} mit SVR

Es ist zu erkennen, dass für ein $\log \varepsilon$ von -1 sehr niedrige Werte der Vorhersagefehler vorliegen und wenig Stützvektoren benötigt werden. Deren Anzahl sollte so gering wie möglich gehalten werden. Erhöht man ε über 0,1 so steigen die Vorhersagefehler drastisch an, während bei abfallendem ε die Anzahl der nötigen Stützvektoren n_{sv} in die Höhe gehen.

Daher wurde für den letzten Schritt ein ε von 0,1 gemeinsam mit den zuvor festgelegten Werten für γ und C als Startpunkt für eine finale Modelloptimierung mit dem Simplex-Verfahren von Nelder und Mead gewählt [NELDER & MEAD, 1965].

Das dafür verwendete Gütekriterium J_{svr} ,

$$J_{svr}(\gamma, C, \varepsilon) = \frac{RMSEP(\gamma, C, \varepsilon)}{a_{svr}} + \frac{n_{sv}(\gamma, C, \varepsilon)}{n_{svmax}} \quad (8.1)$$

mit

a_{svr} := Wichtungsfaktor bei der SVR-Modelloptimierung
 n_{svmax} := maximale Anzahl an Supportvektoren,

sorgt für eine simultane Minimierung des Vorhersagefehlers RMSEP sowie der Anzahl n_{sv} benötigter Stützvektoren, deren maximale Anzahl n_{svmax} von 90 gleich der Anzahl an Proben im Kalibrierdatensatz n_{CS} ist. Der Wichtungsfaktor a_{svr} wurde dagegen mit einem Wert von 3 frei gewählt.

Die Modelloptimierung führte zu den in Tabelle 8.2 gezeigten Parametersätzen für die SVR-Modelle der Zelldichte c_{XL1} und der Glycerolkonzentration c_{S1L1} .

Tabelle 8.2: Ermittelte Parameter der SVR-Modelle zur Vorhersage von c_{XL1} und c_{S1L1}

Parameter	c_{XL1}	c_{S1L1}
γ	[-]	$3,27 \cdot 10^{-3}$
C	[-]	407
ε	[-]	$8,42 \cdot 10^{-2}$
n_{sv}	[-]	24
		35

Die Vorhersagen beider Größen für den Kalibrierdatensatz sind in Bild 8.15 im Vergleich mit den Ergebnissen der PLSR (grau) gezeigt. In beiden Fällen konnte eine deutliche Verbesserung erzielt werden.

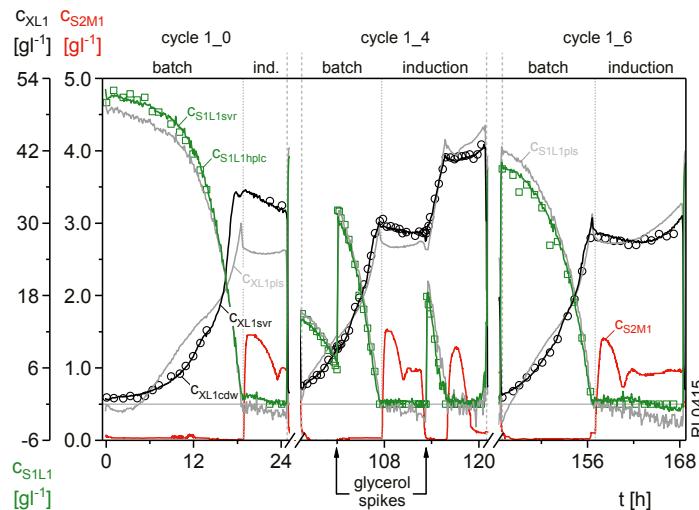


Bild 8.15: Vorhersage beider Zielgrößen mit der SVR für den Kalibrierdatensatz

Die Zelldichte c_{XL1svr} spiegelt mit ihrem exponentiellen Anstieg in den Batchphasen sowie sowie mit dem Abfall während des Umbaus vom Glycerol- auf den Methanolstoffwechsel das Wachstumsverhalten hervorragend wieder. Der exponentielle Abfall der Glycerolkonzentration c_{S1L1} in den Batchphasen sowie Glycerol-Spikes werden ebenfalls gut dargestellt. Auch die Phasen ohne Glycerol sind relativ gut getroffen.

Bei der Betrachtung der externen Validierung, gezeigt in Bild 8.16, ist die Verbesserung weniger ausgeprägt. Die Zelldichte c_{XL1} wird nur in Zyklus 1_3 sehr gut vorhergesagt. In Zyklus 1_2 liegt die Prädiktion allerdings oberhalb der Offline-Messwerte c_{XL1cdw} . Dennoch reduziert sich der Fehler RMSEP von $2,11 \text{ gl}^{-1}$ (PLSR) auf $1,23 \text{ gl}^{-1}$ (SVR).

Das wahre Prozessverhalten wird ebenfalls viel besser mit SVR dargestellt. Dies ist auch in Bild 8.17a zu erkennen, in der die Prädiktion c_{XL1svr} nahezu mit ihren Referenzwerten c_{XL1cdw} übereinstimmt.

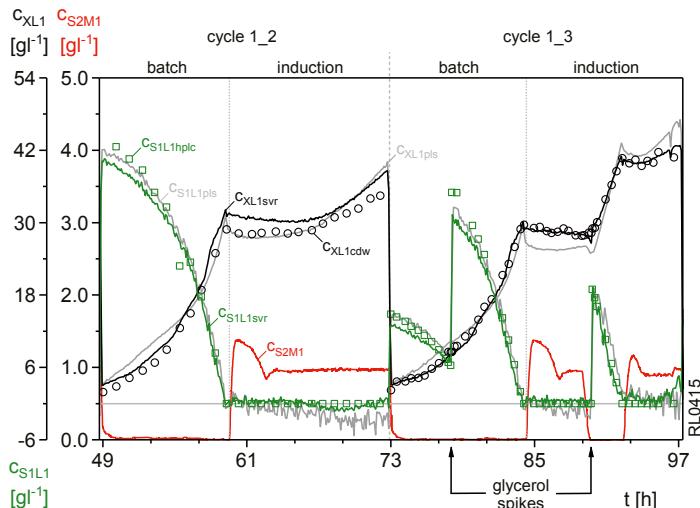


Bild 8.16: Vorhersage beider Zielgrößen mit der SVR für den Validierdatensatz

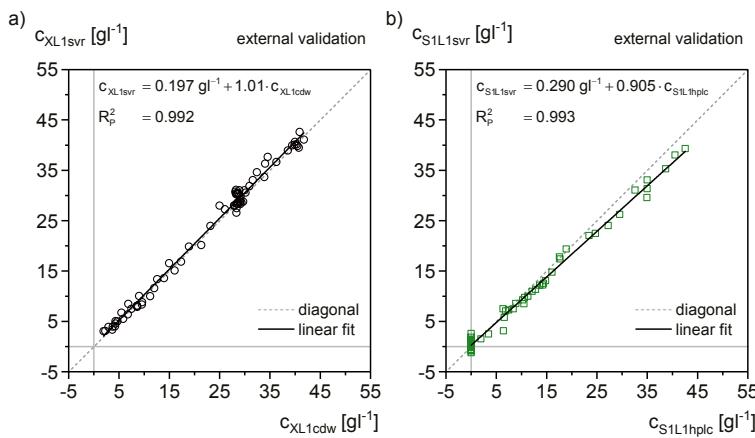


Bild 8.17: Vorhersagen beider Zielgrößen mit der SVR aufgetragen gegen ihre Referenzwerte a) Zelldichte c_{XL1svr} b) Glycerolkonzentration $c_{S1L1svr}$

Im Vergleich zu der entsprechenden Abbildung des PLSR-Modells (Abschnitt 8.3.2, Bild 8.10a) ist ein deutlicher linearer Zusammenhang zwischen der Vorhersage c_{XL1svr} und den Referenzwerten zu erkennen.

Die Darstellung des prinzipiellen Verhaltens der Glycerolkonzentration c_{S1L1} wird ebenfalls verbessert. Der Vorhersagefehler RMSEP verringerte sich von $1,85 \text{ gl}^{-1}$ (PLSR) auf $1,58 \text{ gl}^{-1}$ (SVR). Jedoch wird dies hauptsächlich auf die genauere Abbildung glycerolfreier Phasen zurückgeführt.

Hohe Werte von c_{S1L1} werden dagegen zu niedrig vorhergesagt. Dies ist auch anhand von Bild 8.17b zu erkennen, in der zwar ein linearer Zusammenhang zwischen der Vorhersage $c_{S1L1svr}$ mit den Referenzwerten besteht, die Gerade jedoch eine Steigung von < 1 aufweist.

Ein Vergleich der SVR mit der PLSR für die Online-Prädiktion von c_{XL1} und c_{S1L1} mit Raman-Spektroskopie ist abschließend anhand üblicher Kenngrößen in Tabelle 8.3 gegeben.

Tabelle 8.3: Vergleich der SVR und der PLSR bei Vorhersage von c_{XL1} und c_{S1L1}

		Zelldichte c_{XL1}		Glycerolkonzentration c_{S1L1}	
		PLSR	SVR	PLSR	SVR
n_{CS}	[\cdot]	90	90	90	90
R^2_{cv}	[\cdot]	0,963	0,992	0,986	0,996
RMSEcv	[gl^{-1}]	2,54	1,23	2,07	1,22
RMSEcv _{rel}	[$\%$]	6,03	2,92	3,98	2,35
n_{VS}	[\cdot]	74	74	69	69
R^2_p	[\cdot]	0,970	0,992	0,978	0,993
RMSEP	[gl^{-1}]	2,11	1,23	1,85	1,58
RMSEP _{rel}	[$\%$]	5,01	2,92	3,56	3,04

9. Regelung der Glycerolkonzentration

9.1 Das Regelungskonzept

In diesem Kapitel soll die Regelung der Glycerolkonzentration c_{S1L} im Bioreaktor als mögliche PAT-Anwendung näher untersucht werden.

Diese fand im Zellanzuchtbioreaktor BIOSTAT® ED10 statt, da entsprechende multivariate Modelle für diesen Teil der zweistufigen Anlage entwickelt wurden (siehe Abschnitt 8.3). Da der Produktionsbioreaktor nicht Gegenstand dieser Untersuchungen war, wird auf die Kennzeichnung der Reaktornummer ($p = 1$) im Index betrachteter Größen verzichtet.

Bild 9.1 zeigt den verwendeten Bioreaktor mit den wichtigsten für die Glycerolregelung benötigten Gerätschaften.

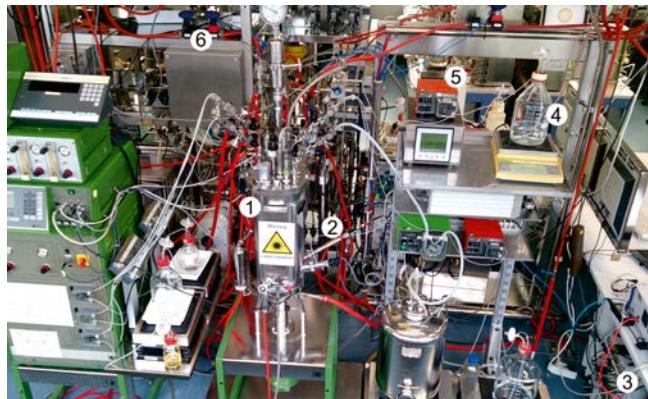


Bild 9.1: Zellanzuchtbioreaktor BIOSTAT® ED10, 1) Bioreaktor, 2) Raman-Sonde, 3) Raman-Spektrometer (Rückseite), 4) Glycerolreservoir, 5) Substratpumpe, 6) Abgas-Analysesystem

Als Regelgröße wird die aus Raman-Spektrien I_L und durch ein PLSR-Modell vorhergesagte Glycerolkonzentration c_{S1Lram} herangezogen. Bild 9.2 zeigt eine schematische Darstellung des Regelungskonzeptes.

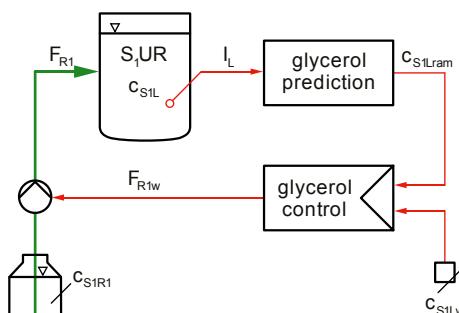


Bild 9.2: Schematische Darstellung des Konzeptes zur Regelung von c_{S1L}

Glycerol wird im Bioreaktor von den Hefezellen mit der volumetrischen Aufnahmerate S_1UR verbraucht. Ein Vergleich von c_{S1Lram} mit dem Sollwert c_{S1Lw} wird über einen Regelalgorithmus in das Stellsignal F_{R1w} einer Substratpumpe überführt. Diese fördert Glycerollösung mit der Konzentration c_{S1R1} aus einem Reservoir R1 in den Bioreaktor, um die Regelung von c_{S1L} zu bewirken.

9.2 Theoretische Betrachtung des Regelungsproblems

9.2.1 Elemente des Regelkreises

In Bild 9.3 ist ein Blockschaltbild mit den vier Teilsystemen des Regelkreises gezeigt.

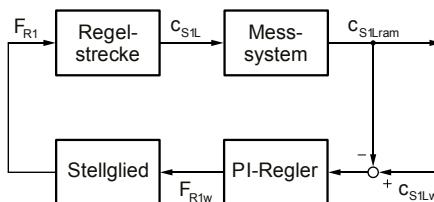


Bild 9.3: Regelungstechnisches Ersatzschaltbild zur Glycerolregelung

Der Bioreaktor stellt die Regelstrecke dar, die systemtheoretisch betrachtet die Zufütterrate F_{R1} in die Glycerolkonzentration c_{S1L} in der Flüssigphase transformiert. Bei dem Messsystem handelt es sich um das Raman-Spektrometer mit nachfolgender Datenverarbeitung, das alle drei Minuten einen Messwert c_{S1Lram} der Glycerolkonzentration zur Verfügung stellt. Für den Regler wurde ein PI-Algorithmus gewählt.

9.2.2 Charakterisierung des Streckenverhaltens

Als erster Schritt bei den theoretischen Untersuchungen des Regelkreises erfolgt eine mathematische Beschreibung der Regelstrecke. Diese beruht auf der Massenbilanz des Substrates Glycerol (S_1) in der Flüssigphase L,

$$\begin{aligned}\dot{m}_{S1L}(t) &= F_{R1}(t) \cdot c_{S1R1} - q_{S1/X}(t) \cdot c_{XL}(t) \\ &= \dot{V}_L(t) \cdot c_{S1L}(t) + V_L(t) \cdot \dot{c}_{S1L}(t)\end{aligned}\quad (9.1)$$

mit

V_L	:= Volumen der Flüssigphase L	[l]
c_{S1L}	:= Glycerolkonzentration in L	[gl ⁻¹]
F_{R1}	:= Glycerolzufütterrate aus Reservoir R1	[l h ⁻¹]
c_{S1R1}	:= Glycerolkonzentration in R1	[gl ⁻¹]
$q_{S1/X}$:= zellspezifische Glycerolaufnahmerate	[h ⁻¹]
c_{XL}	:= Zeldichte (cdw) in L	[gl ⁻¹].

Mit Annahme einer idealisierten Fed-Batch-Kultivierung,

$$\dot{V}_L(t) = F_{R1}(t), \quad (9.2)$$

und Einführung der volumetrischen Glycerolaufnahmerate S_1UR ,

$$S_1UR(t) = q_{S1/X}(t) \cdot c_{XL}(t), \quad (9.3)$$

erhält man die Beschreibung einer nichtlinearen Streckendynamik,

$$\dot{c}_{S1L}(t) = \frac{F_{R1}(t)}{V_L(t)} \cdot [c_{S1R1} - c_{S1L}(t)] - S_1UR(t), \quad (9.4)$$

in der c_{S1L} die Regelgröße, F_{R1} die Steuergröße sowie V_L eine messbare und S_1UR eine nicht direkt messbare Störgröße darstellen.

Im Weiteren soll S_1UR in Gl. (9.4) durch die indirekt messbare volumetrische Sauerstoffaufnahmerate OUR ersetzen,

$$OUR(t) = q_{1O/X}(t) \cdot c_{XL}(t) \quad (9.5)$$

mit

$OUR :=$ volumetrische O_2 -Aufnahmerate $[g^{-1}h^{-1}]$

$q_{1O/X} :=$ zellspezifische O_2 -Aufnahme mit Substrat S1 $[h^{-1}]$,

ersetzt werden.

Diese wird über das quasistationäre volumetrische O_2 -Fließgleichgewicht,

$$Q_{O2}(t) = OTR(t) = OUR(t) \quad (9.6)$$

mit

$Q_{O2} :=$ O_2 -Versorgungsrate aus der Gasphase $[g^{-1}h^{-1}]$

$OTR :=$ O_2 -Transferrate von der Gas- in die Flüssigphase $[g^{-1}h^{-1}]$,

ermittelt.

Die volumetrische O_2 -Versorgungsrate Q_{O2} wird aus den sechs messbaren Prozessgrößen F_{nG} , V_L , x_{OGin} , x_{CGin} , x_{OG} und x_{CG} (siehe Abschnitt 3.3.7) bilanziert und ersetzt die unbekannte O_2 -Aufnahmerate OUR.

Zur Beschreibung von S_1UR muss das in Bild 9.4 gezeigte Reaktionsschema herangezogen werden.

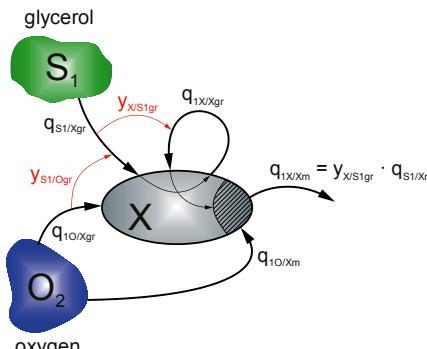


Bild 9.4: Reaktionsschema mit zeitinvarianten Parametern

Im Allgemeinen werden die zellinternen Reaktionen für Zellwachstum sowie Substrat- und Sauerstoffaufnahme,

$$q_{1X/X}(t) = y_{X/S1}(t) \cdot q_{S1/X}(t) = y_{1X/O}(t) \cdot q_{1O/X}(t) \quad (9.7)$$

mit

$$q_{1X/X} := \text{zellspezifische Zellreaktionsrate auf Glycerol} \quad [h^{-1}]$$

$$y_{X/S1} := \text{Zell/Substrat-Ausbeute bei Wachstum auf Glycerol} \quad [-]$$

$$y_{1X/O} := \text{Zell/Sauerstoff-Ausbeute bei Wachstum auf Glycerol} \quad [-],$$

durch Ausbeutekoeffizienten verknüpft, deren Zeitvarianz durch die Aufnahmeraten des Erhaltungsstoffwechsels (*maintenance*) $q_{S1/Xm}$ für Substrat und $q_{1O/Xm}$ für Sauerstoff verursacht werden. In Bild 9.4 ist dieser Einfluss auf das Zellzuwachsverhalten berücksichtigt.

Die zellspezifische Glycerolaufnahmerate $q_{S1/X}$,

$$q_{S1/X}(t) = q_{S1/Xgr}(t), \quad (9.8)$$

dient zunächst vollständig dem Zuwachs an Zellmasse,

$$q_{1X/Xgr}(t) = y_{X/S1gr} \cdot q_{S1/X}(t), \quad (9.9)$$

der durch den zeitinvarianten Zellzuwachs-Ausbeutekoeffizienten $y_{X/S1gr}$ charakterisiert wird.

Ein Teil des umgesetzten Substrates S1 geht allerdings durch die Zellverlustrate $q_{1X/Xm}$,

$$q_{1X/Xm} = y_{X/S1gr} \cdot q_{S1/Xm}, \quad (9.10)$$

wieder verloren, die zeitinvariante und proportional zur Substratmaintenancerate $q_{S1/Xm}$ angenommen wird.

Die beobachtbare zellspezifische Zellreaktionsrate $q_{1X/X}$,

$$q_{1X/X}(t) = q_{1X/Xgr}(t) - q_{1X/Xm} = y_{X/S1gr} \cdot (q_{S1/X}(t) - q_{S1/Xm}), \quad (9.11)$$

beinhaltet somit den Zuwachs durch die Substrataufnahmerate $q_{S1/X}$ und den Verlust durch die Substratmaintenancerate $q_{S1/Xm}$.

Die zellspezifische O_2 -Aufnahmerate $q_{1O/X}$ auf Glycerol,

$$q_{1O/X}(t) = q_{1O/Xgr}(t) + q_{1O/Xm} = y_{1X/Ogr} \cdot q_{1X/Xgr}(t) + q_{1O/Xm}, \quad (9.12)$$

beinhaltet hingegen den Sauerstoffbedarf bei Aufnahme von S1, als auch den der Energiegewinnung für den Erhaltungsstoffwechsel.

Überführt man dieses Reaktionsverhalten in das volumetrische Reaktionsgleichgewicht,

$$y_{X/S1}(t) \cdot S_1UR(t) = y_{1X/O}(t) \cdot OUR(t), \quad (9.13)$$

so erhält man nach einigen Umformungen die Beschreibung der Störgröße S_1UR ,

$$S_1UR(t) = \frac{y_{1X/Ogr}}{y_{X/S1gr}} \cdot \frac{q_{1X/X}(t) + y_{X/S1gr} \cdot q_{S1/Xm}}{q_{1X/X}(t) + y_{X/S1gr} \cdot q_{S1/Xm} + y_{1X/Ogr} \cdot q_{1O/Xm}} \cdot Q_{O2}(t), \quad (9.14)$$

durch die O_2 -Eintragsrate Q_{O2} .

Da die Glycerolregelung durch die Wahl eines Sollwertes c_{S1Lw} von 10 g l⁻¹ eine unlimitierte und hohe Wachstumsrate $q_{1X/X}$ gewährleistet, ist der Einfluss der Sauerstoffmaintenancerate $q_{1O/Xm}$ vernachlässigbar.

Ermittelt man aus Vorversuchen die beiden Wachstums-Ausbeutekoeffizienten, so ist aus dem stöchiometrischen Glycerol/Sauerstoff-Quotienten $y_{S1/Ogr}$,

$$y_{S1/Ogr} = \frac{y_{1X/Ogr}}{y_{X/S1gr}}, \quad (9.15)$$

die unbekannte Störgröße S_1UR ,

$$S_1UR(t) = y_{S1/Ogr} \cdot Q_{O2}(t), \quad (9.16)$$

nunmehr direkt berechenbar und bei der Auslegung des Regelkreises zu berücksichtigen.

Die für eine Berechnung des Parameters $y_{S1/Ogr}$ benötigten Ausbeutekoeffizienten $y_{X/S1gr}$ und $y_{1X/Ogr}$ wurden der Arbeit von Martens entnommen und sind in Tabelle 9.1 aufgeführt.

Tabelle 9.1: Globale Reaktionsparameter zur Berechnung der Glycerolaufnahmerate S_1UR aus der Sauerstoffeintragsrate Q_{O2} , aus [MARTENS, 2014]

$y_{X/S1gr}$	$y_{1X/Ogr}$	$y_{S1/Ogr}$
0,887	2,47	2,78

Abschließend sei die finale Streckendynamik,

$$\begin{aligned} \dot{c}_{S1L}(t) &= \frac{F_{R1}(t)}{V_L(t)} \cdot [c_{S1R1} - c_{S1L}(t)] - y_{S1/Ogr} \cdot Q_{O2}(t) \\ &= f(c_{S1L}, F_{R1}, V_L, Q_{O2}, t), \end{aligned} \quad (9.17)$$

nochmals in Abhängigkeit der vier Streckenvariablen notiert.

9.2.3 Einführung des linearisierten Streckenmodells

Das Streckenverhalten weist ein hohes nichtlineares Verhalten auf. Zum einen sind die Steuergröße F_{R1} und die Regelgröße c_{S1L} multiplikativ verknüpft und zum anderen steigen die Störgrößen V_L und Q_{O2} , beide bedingt durch ansteigenden Substrat- und Sauerstoffbedarf, exponentiell an.

Um dieses Verhalten zu beherrschen und damit die Substratkonzentration konstant zu halten, bedürfen die erforderlichen Reglerparameter einer Nachführung.

Die klassische Regelungstheorie beruht allerdings auf linearem Systemverhalten um einen festen Arbeitspunkt des Regelkreises mit zeitinvarianten Prozessparametern.

Dies wird hier erreicht, indem die Streckendifferentialgleichung um einen gleitenden Arbeitspunkt linearisiert wird und die Strecken- sowie die Reglerparameter in zeitdiskreten Abständen nachgeführt werden.

Der Arbeitspunkt eines Regelkreises wird so definiert, dass zu einem Zeitpunkt t_{op} die Regelgröße c_{S1L} ,

$$c_{S1L}(t_{op}) = c_{S1Lw}(t_{op}) =: c_{S1Lop}, \quad (9.18)$$

ihrem aktuellen Sollwert entspricht und die zeitliche Ableitung \dot{c}_{S1L} ,

$$\dot{c}_{S1L}(t_{op}) = f_{op} = \frac{F_{R1op}}{V_{Lop}} \cdot [C_{S1R1} - c_{S1Lop}] - y_{S1/Ogr} \cdot Q_{O2op} = 0, \quad (9.19)$$

verschwindet.

Bei einem gleitenden Arbeitspunkt sind zu diskreten Zeitpunkten t_k die aktuellen Werte,

$$V_{Lop} = V_L(t_k) =: V_{Lk}, \quad (9.20)$$

$$Q_{O2op} = Q_{O2}(t_k) =: Q_{O2k}, \quad (9.21)$$

$$c_{S1Lop} = c_{S1L}(t_k) =: c_{S1Lwk}, \quad (9.22)$$

aus der Sollwertvorgabe und den Messwerten von V_L und Q_{O2} bekannt.

Hieraus ist die erforderliche Zufütterrate F_{R1op} ,

$$F_{R1op} = F_{R1k} = \frac{y_{S1/Ogr} \cdot V_{Lk} \cdot Q_{O2k}}{C_{S1R1} - c_{S1Lwk}}, \quad (9.23)$$

online berechenbar.

Mit der Definition des Systems der Abweichungen,

$$\Delta c_{S1L}(t) = c_{S1L}(t) - c_{S1Lwk}, \quad (9.24)$$

$$\Delta \dot{c}_{S1L}(t) = \dot{c}_{S1L}(t), \quad (9.25)$$

$$\Delta F_{R1}(t) = F_{R1}(t) - F_{R1k}, \quad (9.26)$$

$$\Delta V_L(t) = V_L(t) - V_{Lk}, \quad (9.27)$$

$$\Delta Q_{O2}(t) = Q_{O2}(t) - Q_{O2k}, \quad (9.28)$$

lässt sich nun durch eine Taylor-Reihenentwicklung mit Vernachlässigung der Ableitungen höherer Ordnung,

$$\begin{aligned} \Delta \dot{c}_{S1L}(t) &= \underbrace{f_{op}}_{0 \text{ gl } h^{-1}} + \left. \frac{\partial f}{\partial c_{S1L}} \right|_{op} \cdot \Delta c_{S1L}(t) + \left. \frac{\partial f}{\partial F_{R1}} \right|_{op} \cdot \Delta F_{R1}(t) \\ &\quad + \left. \frac{\partial f}{\partial V_L} \right|_{op} \cdot \Delta V_L(t) + \left. \frac{\partial f}{\partial Q_{O2}} \right|_{op} \cdot \Delta Q_{O2}(t), \end{aligned} \quad (9.29)$$

eine lineare Streckendifferentialgleichung herleiten.

Damit lässt sich die regelungstechnische Normalform der Regelstrecke,

$$T_{Sk} \cdot \Delta \dot{c}_{S1L}(t) + \Delta c_{S1L}(t) = K_{Sk} \cdot (\Delta F_{R1}(t) + K_{z1k} \cdot \Delta V_L(t) + K_{z2k} \cdot \Delta Q_{O2}(t)), \quad (9.30)$$

notieren.

Es handelt sich um ein Verzögerungszeitsystem 1. Ordnung mit drei proportionalen Eingängen.

Die Streckenzeitzkonstante T_{Sk} ,

$$T_{Sk} = \frac{V_{Lk}}{F_{R1k}}, \quad (9.31)$$

und die Streckenstellverstärkung K_{Sk} ,

$$K_{Sk} = \frac{c_{S1R1} - c_{S1Lwk}}{F_{R1k}}, \quad (9.32)$$

bestimmen jeweils bei t_k die Auslegung der Reglerparameter, während die Volumenstörverstärkung K_{z1k} ,

$$K_{z1k} = -\frac{F_{R1k}}{V_{Lk}}, \quad (9.33)$$

und die Störverstärkung der O_2 -Aufnahme K_{z2k} ,

$$K_{z2k} = -\frac{y_{S1/Ogr} \cdot V_{Lk}}{c_{S1R1} - c_{S1Lwk}}, \quad (9.34)$$

das Störübertragungsverhalten des Regelkreises beeinflussen.

9.2.4 Vernachlässigung der Dynamik des Messsystems

Im Falle der Regelungsversuche nimmt die Erzeugung eines neuen Messwertes mit dem Raman-Spektrometer eine Zeit von drei Minuten in Anspruch. Der vorhergehende Messwert wird während dieser Zeit beibehalten. Diese Funktionsweise wird als *Sample-And-Hold* (SAH) bezeichnet.

Vergleicht man die Haltezeit T_H von 0,05 h mit der minimalen Streckenzeitzkonstante T_{Smin} ,

$$T_{Smin} = \frac{V_{Lmin}}{F_{R1max}} = \frac{8l}{0.2l h^{-1}} = 40h, \quad (9.35)$$

wird deutlich, dass man die Dynamik des Messverfahrens bei der Auslegung des Regelkreises vernachlässigen kann.

9.2.5 Regel- und Stellverhalten

Für die Substratregelung wird ein PI-Regler eingesetzt. Bild 9.5 verdeutlicht das Regel- und Stellverhalten.

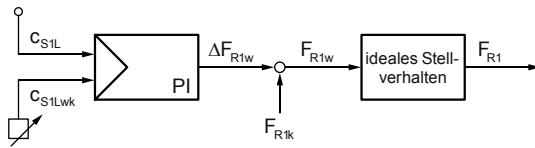


Bild 9.5: Zusammenhang zwischen Reglerausgang und Prozesssteuergröße

Der Sollwert Δc_{S1Lw} im System der Abweichungen,

$$\Delta c_{S1Lw}(t) = c_{S1Lw}(t) - c_{S1L}(t) = 0, \quad (9.36)$$

ist im nachgeführten Regelkreis immer null.

Die Regeldifferenz e_{S1L} ,

$$e_{S1L}(t) = c_{S1Lw}(t) - c_{S1L}(t) = -\Delta c_{S1L}(t), \quad (9.37)$$

entspricht somit der negativen Regelgröße $-\Delta c_{S1L}$.

Der Reglerausgang, die Stellgröße ΔF_{R1w} ,

$$\Delta F_{R1w}(t) = -K_{Rk} \cdot \left(\Delta c_{S1L}(t) + \frac{1}{T_{Ik}} \cdot \int_{t_k}^t \Delta c_{S1L}(\tau) d\tau \right), \quad (9.38)$$

wird daher bei zyklischer Aktualisierung des Arbeitspunktes F_{R1k} durch Rücksetzung des I-Anteils bei t_k zunächst nur vom P-Anteil bestimmt.

Der Sollwert der nachfolgenden Zufütterrate F_{R1w} ,

$$F_{R1w}(t) = F_{R1k} + \Delta F_{R1w}(t), \quad (9.39)$$

wird durch Addition der Stellgröße ΔF_{R1w} und des aktuellen Arbeitspunktes F_{R1k} eingestellt.

Die Dynamik der Prozesssteuergröße, die Zufütterrate F_{R1} ,

$$F_{R1}(t) = F_{R1w}(t), \quad (9.40)$$

wird durch Annahme eines idealisierten Stellgliedes (Zufütterpumpe) vernachlässigt.

Allerdings ist in der Praxis die Zufütterrate mit

$$F_{R1}(t) \in [0, F_{R1max}],$$

durch die Pumpe beschränkt.

Dadurch wird der Stellbereich des Reglers,

$$\Delta F_{R1w}(t) \in [-F_{R1k}, F_{R1max} - F_{R1k}] = [\Delta F_{R1min}, \Delta F_{R1max}],$$

eingegrenzt.

9.3 Untersuchung der Dynamik des Regelungsproblems

9.3.1 Übertragungsfunktionen des Regelkreises

Zur Bewertung und Vorgabe des dynamischen Verhaltens des Regelkreises werden die energiefreien Teilsysteme in den Bildbereich Laplace-transformiert. Das dynamische Verhalten des Regelkreises ist aus Bild 9.6 zu entnehmen.

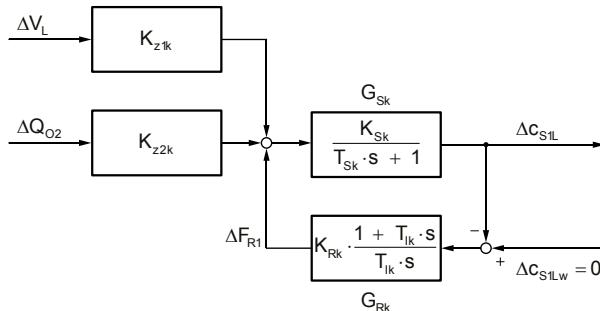


Bild 9.6: Blockschaltbild der Substratregelung im Bildbereich

Die Streckenübertragungsfunktion G_{Sk} ,

$$G_{Sk}(s) = \frac{K_{Sk}}{T_{Sk} \cdot s + 1} = \frac{\Delta c_{S1L}(s)}{\Delta F_{R1}(s)} \quad (9.41)$$

mit

$$\begin{aligned} K_{Sk} &:= \text{bei } t_k \text{ ermittelte Verstärkung der linearisierten Regelstrecke} & [\text{ghl}^{-2}] \\ T_{Sk} &:= \text{bei } t_k \text{ ermittelte Verzögerungszeitkonstante der Regelstrecke} & [\text{h}], \end{aligned}$$

überträgt die Steuergröße $\Delta F_{R1}(s)$ auf die Regelgröße $\Delta c_{S1L}(s)$.

Die Regelstrecke besitzt mit

$$p_{2k} = -\frac{1}{T_{Sk}}, \quad (9.42)$$

einen zeitvarianten Eigenwert (Pol).

Der PI-Regler mit der Reglerübertragungsfunktion G_{Rk} ,

$$G_{Rk}(s) = K_{Rk} \cdot \frac{1 + T_{Ik} \cdot s}{T_{Ik} \cdot s} \quad (9.43)$$

mit

$$\begin{aligned} K_{Rk} &:= \text{bei } t_k \text{ nachzuführende Reglerverstärkung} & [\text{l}^2 \text{g}^{-1} \text{h}^{-1}] \\ T_{Ik} &:= \text{bei } t_k \text{ nachzuführende Integrationszeitkonstante} & [\text{h}], \end{aligned}$$

besitzt mit

$$p_1 = 0, \quad (9.44)$$

einen zeitinvarianten Pol und mit

$$z_{ik} = -\frac{1}{T_{ik}}, \quad (9.45)$$

die nachzuführende Nullstelle des Regelkreises.

Mit der komplexen Übertragungsfunktion des offenen Regelkreises G_{0k} ,

$$G_{0k}(s) = \frac{K_{0k} \cdot (1 + T_{ik} \cdot s)}{T_{ik} \cdot s \cdot (1 + T_{sk} \cdot s)}, \quad (9.46)$$

mit der zugehörigen Kreisverstärkung K_{0k} ,

$$K_{0k} = K_{sk} \cdot K_{rk}, \quad (9.47)$$

ist die Dynamik des Regelkreises bereits festgelegt.

Die Führungsübertragungsfunktion des geschlossenen Regelkreises G_{wk} ,

$$G_{wk}(s) = \frac{G_{0k}(s)}{1 + G_{0k}(s)} = \frac{K_{0k} \cdot (1 + T_{ik} \cdot s)}{T_{ik} \cdot T_{sk} \cdot s^2 + (1 + K_{0k}) \cdot T_{ik} \cdot s + K_{0k}}, \quad (9.48)$$

beschreibt ein PD-T₂-Verhalten mit der Proportionalverstärkung K_{pw} ,

$$K_{pw} = 1, \quad (9.49)$$

und der Differentialverstärkung K_{dwk} ,

$$K_{dwk} = T_{ik}. \quad (9.50)$$

Die komplexe Störübertragungsfunktion G_{zik} der jeweiligen Störgröße z_i ,

$$G_{zik}(s) = \frac{K_{zik} \cdot K_{sk} \cdot T_{ik} \cdot s}{\mathcal{A}_k(s)} \quad (9.51)$$

mit

$$K_{zik} := \text{bei } t_k \text{ ermittelte Störverstärkung des Störeinflusses } z_i,$$

besitzt den gleichen Nenner wie G_{wk} und beschreibt ein D-T₂-Verhalten.

Eine sprungförmige Störung erzeugt somit nur eine gegen Null gehende transiente Störantwort, während bei einer sprungförmigen Führungsänderung der Istwert Δc_{s1L} in dem neuen Sollwert Δc_{s1Lw} enden wird.

9.3.2 Eigenwerte des geschlossenen Regelkreises

Der Nenner \mathcal{A}_k ,

$$\mathcal{A}_k(s) = T_{ik} \cdot T_{sk} \cdot s^2 + (1 + K_{0k}) \cdot T_{ik} \cdot s + K_{0k}, \quad (9.52)$$

stellt das charakteristische Polynom des Regelkreises dar und bestimmt bei Vorgabe von T_{sk} und K_{sk} sowie durch die Wahl von T_{ik} und K_{rk} das Regelverhalten.

Dieses lässt sich anhand der beiden Lösungen, den Eigenwerten $s_{1,2k}$ des geschlossenen Regelkreises,

$$s_{1,2k} = -\frac{1 + K_{0k}}{2 \cdot T_{Sk}} \pm \sqrt{\frac{(1 + K_{0k})^2}{4 \cdot T_{Sk}^2} - \frac{K_{0k}}{T_{Ik} \cdot T_{Sk}}}, \quad (9.53)$$

als T_2 -Verhalten bestimmen und ist bei positiven Parametern global stabil.

Sind beide Eigenwerte negativ und reell, so ist der Regelkreis zudem nicht schwingungsfähig. Ist die Diskriminante in Gl. (9.53) negativ, so erhält man eine konjugiert komplexe Lösung mit negativem Realteil. In diesem Fall schwingt sich der Regelkreis stabil ein.

Bei Kenntnis von K_{Sk} und T_{Sk} sowie Vorgabe von T_{Ik} hängen die Lösungen $s_{1,2k}$ von der Wahl von K_{Rk} ab. Dieses Verhalten, mit K_{0k} parametriert, ist aus den beiden Wurzelortskurven in Bild 9.7 gut abzulesen.

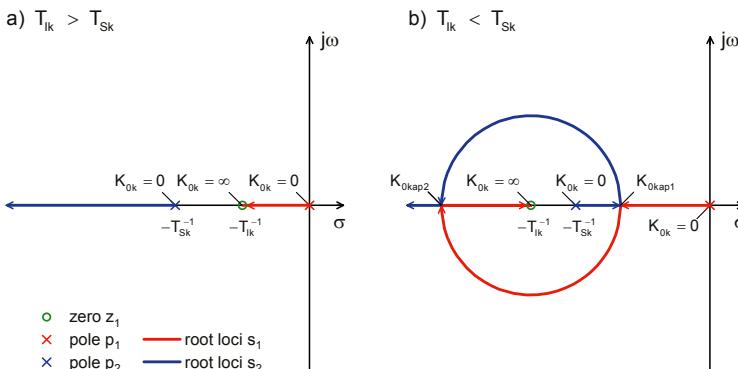


Bild 9.7: Wurzelortskurven in Abhängigkeit der Wahl von T_{Ik}

In Bild 9.7a wurde $T_{Ik} > T_{Sk}$ gewählt. Mit wachsendem K_{0k} bleiben die Eigenwerte (Pole) in der linken s-Halbebene. Der erste Pol wandert bei $K_{0k} = \infty$ in die Nullstelle z_1 und der zweite Pol gegen minus unendlich. Der Regelkreis schwingt in keinem Fall.

Wird $T_{Ik} < T_{Sk}$ gewählt, so wechselt, wie in Bild 9.7b gezeigt, das Regelkreisverhalten. Für $K_{0k} \in [0, K_{0kap1}]$ und $K_{0k} \geq K_{0kap2}$ verhält er sich aperiodisch, während er für $K_{0k} \in]K_{0kap1}, K_{0kap2}[$ schwingt.

Mit der Bedingung $T_{Ik} < T_{Sk}$ können die Kreisverstärkungen $K_{0kap1,2}$ für den aperiodischen Grenzfall,

$$K_{0kap1,2} = \frac{2 \cdot T_{Sk}}{T_{Ik}} \cdot \left(1 \pm \sqrt{1 - \frac{T_{Ik}}{T_{Sk}}} \right) - 1, \quad (9.54)$$

aus der dann verschwindenden Diskriminante in Gl. (9.53) berechnet werden.

9.3.3 Schwingungsverhalten des Regelkreises

Die klassische Notation des charakteristischen Polynoms für ein schwingendes T_2 -Verhalten und der Vergleich mit \mathcal{A}_k^c ,

$$s^2 + 2 \cdot \vartheta_k \cdot \omega_{0k} \cdot s + \omega_{0k}^2 = s^2 + \frac{1 + K_{0k}}{T_{Sk}} \cdot s + \frac{K_{0k}}{T_{Ik} \cdot T_{Sk}} = 0, \quad (9.55)$$

beinhaltet die Kreisfrequenz ω_{0k} des ungedämpften Systems,

$$\omega_{0k} = \sqrt{\frac{K_{0k}}{T_{Ik} \cdot T_{Sk}}}, \quad (9.56)$$

und den Dämpfungsgrad ϑ_k ,

$$\vartheta_k = \frac{1 + K_{0k}}{2 \cdot \sqrt{K_{0k}}} \cdot \sqrt{\frac{T_{Ik}}{T_{Sk}}}, \quad (9.57)$$

der im Bereich $K_{0k} \in]K_{0kap1}, K_{0kap2}[$ unter 1 liegt.

Die Lösungen von Gl. (9.55), die Eigenwerte $s_{1,2k}$ des geschlossenen Regelkreises,

$$s_{1,2k} = \sigma_{Rk} \pm j\omega_{Rk} = -\vartheta_k \cdot \omega_{0k} \pm j\sqrt{1 - \vartheta_k^2} \cdot \omega_{0k}, \quad (9.58)$$

beschreiben mit der Abklingkonstanten σ_{Rk} ,

$$\sigma_{Rk} = -\frac{1 + K_{0k}}{2 \cdot T_{Sk}}, \quad (9.59)$$

und der Eigenkreisfrequenz ω_{Rk} des gedämpften Systems,

$$\omega_{Rk} = \sqrt{\frac{K_{0k}}{T_{Ik} \cdot T_{Sk}} - \frac{(1 + K_{0k})^2}{4 \cdot T_{Sk}^2}}, \quad (9.60)$$

das Verhalten eines schwingenden Regelkreises.

9.3.4 Vorgabe des Regelkreisverhaltens

Für die Berechnung der Reglerparameter eines schwingungsfähigen Systems wird zweckmäßiger Weise einer der Reglerparameter, K_R oder T_I , sowie der Dämpfungsgrad ϑ festgelegt. Mit $\vartheta \geq 1$ erhält man ein nicht schwingendes System und mit $\vartheta < 1$ ein schwingendes System.

Von besonderem Interesse sind solche Reglereinstellungen, bei denen gerade kein Schwingungsverhalten mehr auftritt. Dieser aperiodische Grenzfall wird bei einem Dämpfungsgrad von $\vartheta = 1$ erreicht.

Mit der Online-Ermittlung der beiden Streckenparameter T_{Sk} (Gl. (9.31)) und K_{Sk} (Gl. (9.32)) zu festen Zeitpunkten t_k ist eine Auslegung des Verhaltens der Substratregelung mit den vorgestellten theoretischen Grundlagen leicht möglich.

Ein Ziel könnte ein nichtschwingendes Verhalten (reelle negative Eigenwerte) mit schnellem transienten Abklingen (hohe negative Realteile) sein. Die Lösung wäre eine Einstellung von T_{Ik} geringfügig über T_{Sk} mit einem hohen K_{Rk} (siehe Bild 9.7a).

Dies war in der Praxis allerding nicht möglich, da die Regler der Anlage eine Eingabe von T_i oberhalb von $T_{Smin} = 40$ h nicht zuließen.

Daher war ein gegenüber T_{Sk} kleines T_{Ik} (hoher I-Anteil des Reglers) zu wählen, um die Regelung in der Nähe des zweiten aperiodischen Grenzfallen K_{0kap2} (siehe Bild 9.7b) zu halten.

Hierfür boten sich zwei Möglichkeiten an. In beiden Strategien wurde die Reglerverstärkung K_R fixiert und damit die Verstärkung des offenen Regelkreises K_{0k} ,

$$K_{0k} = K_R \cdot K_{Sk}, \quad (9.61)$$

proportional zur Streckenverstärkung K_{Sk} eingestellt.

Ein nichtschwingendes Verhalten wurde dann mit Vorgabe des aperiodischen Grenzfalls ($\theta = 1$) durch Anpassung von T_i erzwungen.

Die dafür erforderliche Integrationszeitkonstante T_{Ik} ,

$$T_{Ik} = T_{Ikap} = \frac{4 \cdot T_{Sk} \cdot K_R \cdot K_{Sk}}{(1 + K_R \cdot K_{Sk})^2}, \quad (9.62)$$

wird dann zusammen mit F_{Rik} jeweils bei t_k verändert.

Will man hingegen bei festem K_R Schwingungen mit einer definierten Periodendauer T_P ,

$$T_P = \frac{2 \cdot \pi}{\omega_{Rk}}, \quad (9.63)$$

zulassen, so ist T_{Ik} ,

$$T_{Ik} = T_{Ikosc} = \frac{4 \cdot K_R \cdot K_{Sk} \cdot T_{Sk} \cdot T_P^2}{(4 \cdot \pi \cdot T_{Sk})^2 + (1 + K_R \cdot K_{Sk})^2 \cdot T_P^2}, \quad (9.64)$$

mit einer anderen Vorschrift nachzuführen. Das Abklingverhalten σ_{Rk} wird sich dann entsprechend Gl. (9.59) verhalten.

Diese beiden Methoden wurden experimentell erprobt. Die erforderlichen Maßnahmen zur Realisierung werden im Weiteren erläutert.

Andere Adoptionsmöglichkeiten, z.B. Vorgabe von σ_R oder der Überschwingweite über den Sollwert sind aus Gl. (9.60) oder der Literatur zu entnehmen.

9.4 Technische Vorgaben der Glycerolregelung

9.4.1 Prozesstechnische Umsetzung

In den folgenden Abschnitten soll auf die technische Umsetzung der Glycerolregelung, insbesondere im Hinblick auf die verwendete Prozess-EDV, näher eingegangen werden. Die in Bild 9.8 gezeigte erweiterte schematische Darstellung des Regelungskonzeptes zeigt die für eine Glycerolregelung notwendigen Operationen.

Dabei kann insbesondere die Erzeugung der Regelgröße durch Messung von Raman-Spektren und der Anwendung eines PLSR-Modells als ein erster Schritt hervorgehoben werden. Das Vorgehen hierbei wird in Abschnitt 9.4.2 beschrieben.

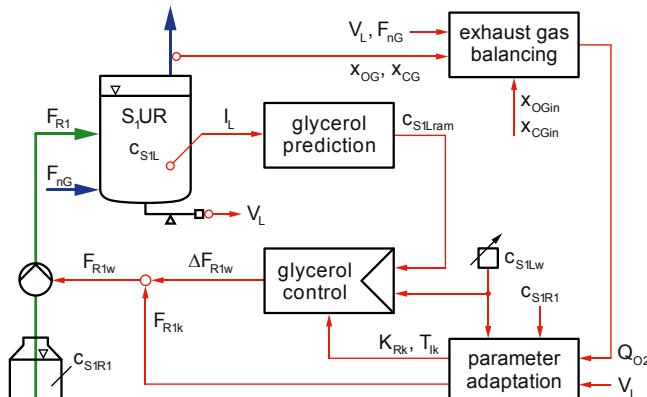


Bild 9.8: Schematische Darstellung der erweiterten Ausstattung zur Regelung der Glycerolkonzentration

Eine weitere wichtige Aufgabe ist die Nachführung des Arbeitspunktes bzw. die Adaption der Reglerparameter und die dafür benötigte Berechnung der Sauerstoffeintragsrate Q_{O_2} durch eine Abgasbilanzierung. Darauf wird in Abschnitt 9.4.3 näher eingegangen.

In Abschnitt 9.4.4 findet sich abschließend eine Beschreibung der Interaktion beteiligter Softwaresysteme sowie eine detaillierte Auflistung der verwendeten Programme.

9.4.2 Bereitstellung der Regelgröße durch ein PLSR-Modell

Das verwendete Online-PLSR-Modell unterscheidet sich in seinem Aufbau und der Vorhersagegüte von dem in Abschnitt 8.3.2 dargestellten finalen Modell. Dies liegt darin begründet, dass aufgrund einer auslaufenden Leihfrist des Raman-Spektrometers, die Modellentwicklungsarbeiten zum Zeitpunkt der Regelungsexperimente noch nicht abgeschlossen waren.

Der hauptsächliche Unterschied der Online- und Offline-Modelle besteht in der Variablenauswahl. Bei der Bildung des Online-Modells wurden die kompletten Raman-Spektren verwendet und nicht eingegrenzt. Dies führte zu einem Vorhersagefehler RMSEP von $3,91 \text{ gl}^{-1}$ im Vergleich zu $1,85 \text{ gl}^{-1}$ des finalen Offline-Modells.

Die Entwicklung des verwendeten PLSR-Modells wurde mit der Software SIMCA durchgeführt. Für die Anwendung des PLSR-Modells zur Online-Prädiktion der Glycerolkonzentration c_{S1Lram} kam das Programm SIMCA-Q zum Einsatz. Hierbei handelt es sich um ein reines Berechnungsmodul, das über keine grafische Oberfläche verfügt.

Die Aufnahme der Raman-Spektren erfolgte mit der Spektrometer-Software iC Raman. Zur Übertragung der Spektren an SIMCA-Q war die Zusatzsoftware Kaiser Data Link notwendig über die außerdem eine Einbindung der SIMCA-Modelldatei erfolgte.

Eine weitere wichtige Funktion von Kaiser Data Link bestand darin, die von SIMCA-Q berechnete Regelgröße c_{S1Lram} im Labornetzwerk über das OPC-Protokoll für weitere beteiligte EDV-Systeme zur Verfügung zu stellen.

9.4.3 Durchführung notwendiger Online-Berechnungen

Die Nachführung des Arbeitspunktes der Glycerolregelung bzw. die Online-Berechnung der im Arbeitspunkt benötigten Glycerolzufüllrate F_{R1k} , und damit die Anpassung der Regelkreisparameter T_{Sk} , K_{Sk} und T_{Ik} , wurde mit einem MATLAB-Skript realisiert. Die dafür notwendige Abgasbilanzierung mit der Berechnung der Sauerstoffeintragsrate Q_{O2} wurde ebenfalls online über ein MATLAB-Skript durchgeführt.

Die Ausführung der Berechnungsskripte wurde durch die PAT-Software SIMATIC SIPAT organisiert. Dort erfolgten eine Einbindung der Skriptdateien sowie deren Versorgung mit den jeweils benötigten Eingangsdaten. Diese werden von SIPAT aus dem Netzwerk über das OPC-Protokoll eingelesen. Auch eine Verteilung der Berechnungen im Netzwerk fand über die OPC-Funktionalität statt.

SIPAT wurde so konfiguriert, dass alle drei Minuten der aktualisierte Wert der Glycerolzufüllrate F_{R1k} im Arbeitspunkt sowie der adaptierte Reglerparameter T_{Ik} für die Regelung mit MFCS/win zur Verfügung stand. Das Berechnungsintervall deckt sich dabei mit dem Messintervall des Raman-Spektrometers.

9.4.4 Vernetzung beteiligter Softwaresysteme

Eine wichtige Rolle bei der Glycerolregelung spielt das SCADA-System MFCS/win. Der verwendete PI-Regelalgorithmus wurde hier eingerichtet. Auch die Berechnung der Steuergröße F_{R1w} , aus der benötigten Zufüllrate im Arbeitspunkt F_{R1k} und dem Reglerausgang ΔF_{R1w} sowie die Ansteuerung der Glycerolpumpe fanden in MFCS/win statt.

Weiterhin erfolgte dort auch die Aufnahme der für die Online-Berechnungen benötigten Variablen, wie zum Beispiel die Größen der Abgasbilanzierung sowie deren Verteilung über das OPC-Protokoll im Netzwerk der Anlage.

Eine besondere Funktion im Zusammenspiel aller beteiligten Softwaresysteme nimmt dabei das Programm SIMATIC WinCC ein, das mit seiner umfangreichen OPC-Server-Funktionalität die Organisation der Datenströme im Labornetzwerk übernimmt [LÖGERING, 2015]. Bild 9.9 illustriert die Vernetzung der bei der Glycerolregelung beteiligten Computerprogramme.

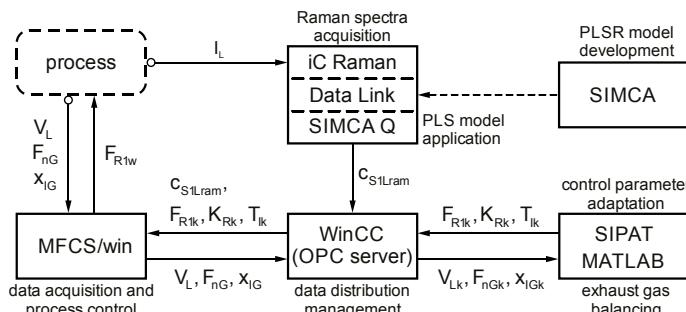


Bild 9.9: Vernetzung und Aufgaben der beteiligten Softwaresysteme

Abschließend sind die verwendeten Computerprogramme in Tabelle 9.2 nochmals detailliert aufgeführt.

Tabelle 9.2: Detaillierte Auflistung der beteiligten Softwaresysteme

Software	Version	Hersteller
Kaiser Data Link	1.1	Kaiser Optical Systems, Ann Arbor, USA
iC Raman	3.0	Kaiser Optical Systems, Ann Arbor, USA
MATLAB	8.3.0 (R2014a)	The Mathworks, Natick, USA
MFCS/win	3.0 (Level 32)	Sartorius Stedim Systems, Guxhagen
SIMATIC SIPAT	4.0.0.0	Siemens, München
SIMATIC WinCC	7.0 + SP2	Siemens, München
SIMCA	13.0.3	Data Analytical Solutions, Umeå, Schweden
SIMCA-Q	13.0.3	Data Analytical Solutions, Umeå, Schweden

9.5 Experimentelle Erprobung der Substratregelung

9.5.1 Regelung im aperiodischen Grenzfall

Im Folgenden werden zwei unterschiedliche Experimente zur Erprobung der Regelung von c_{S1L} diskutiert.

In Bild 9.10 ist der Verlauf der Störgrößen V_{Lk} und Q_{O2k} sowie die Arbeitspunkte T_{Sk} , K_{Sk} und F_{R1k} der linearisierten Strecke gezeigt.

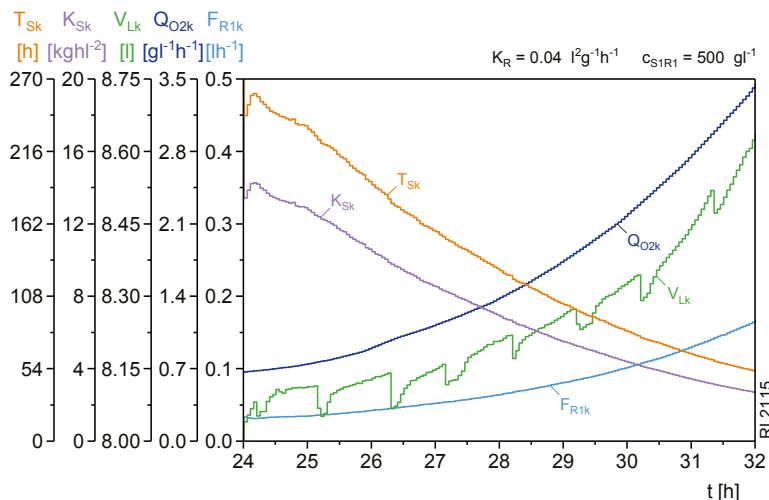


Bild 9.10: Verlauf der Parameter im Arbeitspunkt der linearisierten Strecke

Der Arbeitspunkt der Zufütterrate F_{R1k} (Gl. (9.23)) verläuft proportional zur Sauerstoffmassentransferrate \dot{m}_{OTk} ,

$$\dot{m}_{OTk} = V_{Lk} \cdot Q_{O2k}, \quad (9.65)$$

die unabhängig vom Volumen V_{Lk} ist.

Die gewählte Strategie des laufenden Arbeitspunktes F_{R1k} entspricht somit einem *Feed-Forward Control*, um das exponentielle Wachstum der Zellmasse m_{XL} im Reaktor zu erfüllen.

Beim ersten Experiment wurde K_R auf $0,04 \text{ l}^2\text{g}^{-1}\text{h}^{-1}$ fixiert und T_i auf dem Wert des aperiodischen Grenzfalles T_{ikap} (Gl. (9.62)) nachgeführt. In Bild 9.11 ist der Verlauf der Regelung dargestellt.

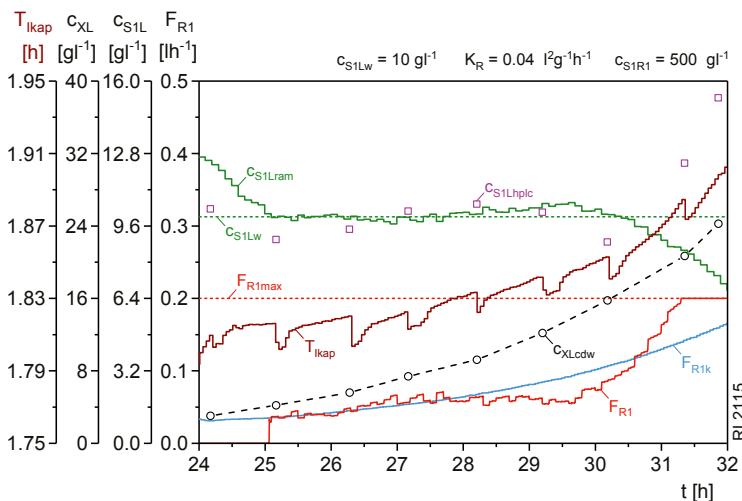


Bild 9.11: Verlauf der adaptiven aperiodischen Substratregelung

Bei $t = 25 \text{ h}$ erreicht die Regelgröße c_{S1Lram} den Sollwert c_{S1Lw} . Der Regler befindet sich somit im Arbeitspunkt und wird aktiviert. Damit sind die Steuergröße F_{R1} und die Zieltrajektorie F_{R1k} dort identisch.

Für die folgenden vier Stunden ist das erwartete Regelungsverhalten zu beobachten, dass durch deutliche, von der Regeldifferenz e_{S1L} abhängige, Steuergrößensprünge ΔF_{R1w} gekennzeichnet ist. Diese sind auf die relativ hohe Reglerverstärkung K_R zurückzuführen.

Die erforderliche aperiodische Integrationszeitkonstante T_{ikap} (Gl. (9.62)) ist viel kleiner als die im Verlauf des Prozesses abfallende Streckenzeitzkonstante T_{Sk} .

Da die Streckenverstärkung K_{Sk} und damit K_{ok} ebenfalls abfallen, bedarf es einer geringen positiven Korrektur von T_{ik} , um den aperiodischen Grenzfall bei K_{0kap2} beizubehalten.

Ab $t = 27,5 \text{ h}$ übersteigt die Regelgröße c_{S1Lram} ihren Sollwert c_{S1Lw} . Der Regler führt die Steuergröße F_{R1} zurück, sodass sie unterhalb der erforderlichen Zielgröße F_{R1k} liegt.

Die damit verbundene Abnahme von c_{S1L} wird zunächst allerdings nur im Verlauf des Offline-Referenzwertes $c_{S1L\text{hplc}}$ bei $t = 29,2$ h, deutlich.

Die spektrale Messung $c_{S1L\text{ram}}$ folgt diesem Verhalten hingegen nur langsam. Nach Unterschreiten des Sollwertes fährt der Regler F_{R1} kräftig hoch, bis die maximale Zufütterrate $F_{R1\text{max}}$ der Pumpe erreicht wird.

Die Offline-Messung zeigt das wahre Prozessverhalten. Der Prozesswert $c_{S1L\text{hplc}}$ steigt weit über den Sollwert an.

Die Vorhersage $c_{S1L\text{ram}}$ mit Raman-Spektroskopie hingegen versagt. Sie fällt immer weiter ab und führt somit zu einem instabilen Regelverhalten.

Die Fehlmessung verdeutlicht, dass sich das Online-PLSR-Modell an den Spektralanteilen der Zelldichte orientiert. Da letztere ab etwa $t = 30$ h stark ansteigt, führt dies zu einem Versagen der Prädiktion eines geregelten Verlaufes von c_{S1L} .

Dieses Verhalten könnte womöglich mit der Verwendung des finalen Modells (siehe Abschnitt 8.3.2) verhindert werden.

9.5.2 Regelung der Glycerolkonzentration im Schwingfall

Zur Erprobung der Vorgabe des Regelverhaltens wurde auch eine Regelung im abklingenden Schwingfall mit konstanter Periodendauer T_p gewählt. Auch hierbei wurde mit einem K_R von $0,012 \text{ l}^2 \text{g}^{-1} \text{h}^{-1}$ eine konstante Reglerverstärkung, allerdings viel kleiner als beim aperiodischen Grenzfall, eingestellt.

Die Integrationszeitkonstante T_{Ik} wurde dabei aus T_{Sk} , K_{Sk} , K_R und T_p mit T_{Ikosc} (Gl. (9.64)) in einer oszillierenden Regelung nachgeführt.

In Bild 9.12 ist der Verlauf der Arbeitspunkte der linearisierten Strecke im zweiten Experiment gezeigt. Die Störgrößen V_{Lk} und Q_{O2k} sind vergleichbar zur ersten Erprobung.

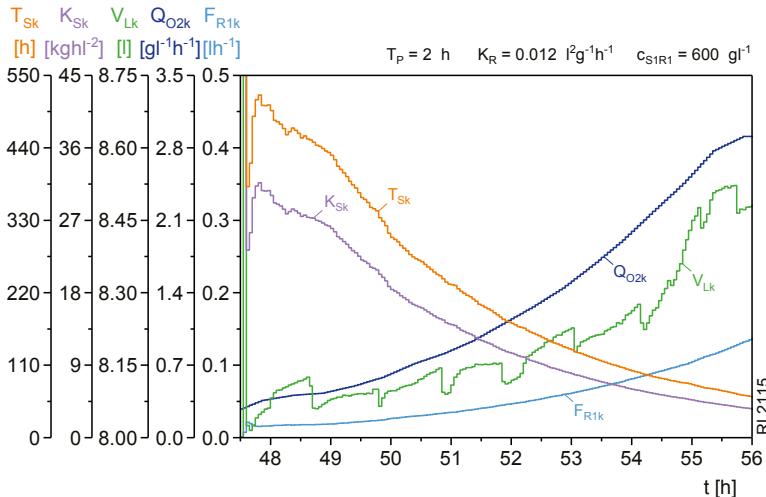


Bild 9.12: Verlauf der Parameter im Arbeitspunkt der linearisierten Strecke

Durch Wahl von c_{S1R1} zu 600 gl^{-1} verläuft F_{R1k} allerdings auf einem niedrigeren Niveau. Damit werden die Streckenparameter T_{Sk} und K_{Sk} nahezu verdoppelt.

In Bild 9.13 ist der Verlauf der Regelung mit einer festen Periodendauer T_p von 2 h gezeigt. Diese wurde bei $t = 47,8 \text{ h}$ aktiviert und zeigte zunächst das gewünschte Verhalten. Die erreichte Periodendauer ist mit zwei senkrechten Linien angedeutet.

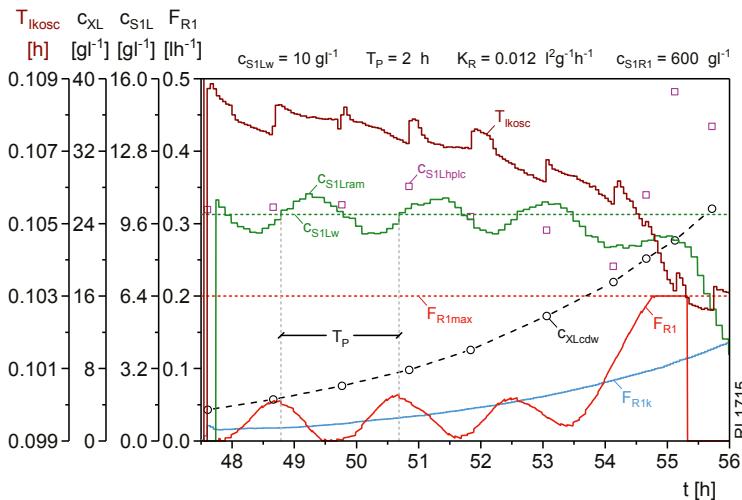


Bild 9.13: Verlauf der Regelung mit vorgegebener Oszillation

Im späteren Verlauf des Prozesses ist das gleiche Fehlverhalten wie bei der aperiodischen Regelung zu beobachten. Die spektral bestimmte Regelgröße c_{S1Lram} folgt nicht mehr den realen Prozesswerten ($c_{S1Lhplc}$) und führt zu einer Überfütterung. Daher wurde das Experiment bei $t = 55,3 \text{ h}$ beendet.

Abschließend ist zu bemerken, dass eine Substratregelung auf der Basis des hier verwendeten PLSR-Modells nicht zum Erfolg führte. Die Erprobung komplexerer Modellbildungen war aufgrund der bereits erwähnten auslaufenden Leihdauer des Raman-Gerätes leider nicht mehr möglich.

10. Anwendung der MVDA auf nichtspektroskopische Daten

10.1 Prädiktion nicht direkt messbarer Prozessgrößen

In diesem Abschnitt wird die Anwendung der MVDA auf nicht spektroskopische Prozessdaten beschrieben. Ziel war auch dabei die Online-Bestimmung von Prozessvariablen, wie zum Beispiel die Produktkonzentration. Weiterhin sollte auch die Ermittlung von zellspezifischen Reaktionsraten getestet werden.

Die Untersuchungen wurden unter Verwendung des aus Abschnitt 7.3 bekannten zyklischen Prozesses durchgeführt, der in Bild 10.1 anhand ausgewählter Größen dargestellt ist.

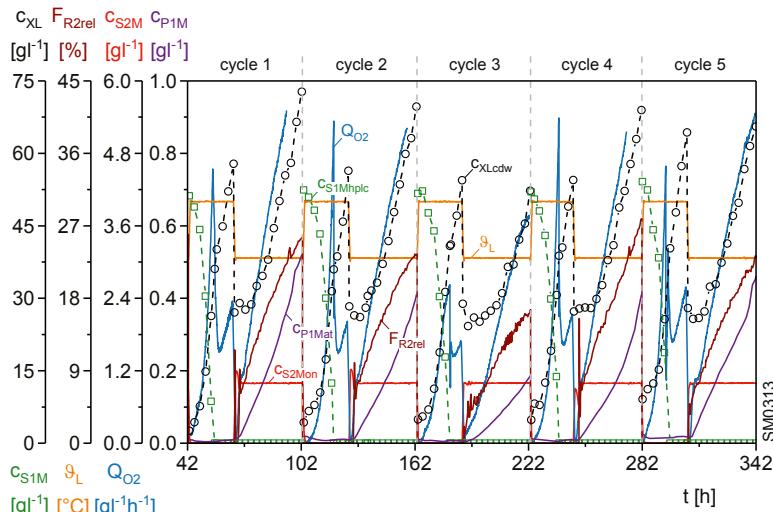


Bild 10.1: Ausgewählte Bioprozessgrößen im Verlauf der untersuchten Kultivierung

Als X-Variablen sollten hier anstelle von Spektren Inline- oder Online-Messgrößen des Bio-reaktionsprozesses verwendet werden. Beispiele sind die Trübungsmessung, Waagensignale oder die Ergebnisse einer Abgasbilanzierung. Details zu der Auswahl verwendeter Variablen sowie der Erzeugung der benötigten Datensätze für die MVDA befinden sich in Abschnitt 10.2.

Als Methoden der multivariaten Kalibrierung sollten hier die PLSR sowie die SVR eingesetzt und verglichen werden. Beide Methoden wurden bereits erfolgreich bei der Entwicklung derartiger *black box soft sensor* Modelle eingesetzt.

Le et al. nutzen PLSR- und SVR-Modelle zur Vorhersage von Produkt- und Laktatkonzentrationen am Ende eines Prozesses zur Herstellung von Antikörpern mit CHO Zellen anhand von 35 Offline- und Online-Prozessgrößen [LE et al., 2012].

Mehrere Quellen verglichen *black box soft sensor* SVR-Modelle mit anderen nichtlinearen Methoden, wie z.B. künstliche neuronale Netze (ANN), in Kultivierungsprozessen verschiedener Hostsysteme. Hierbei stellte sich eine Überlegenheit der SVR-Methodik heraus [DESAI et al., 2006; LI & YUAN, 2006; WANG et al., 2006].

Borchert nutzte in seiner Dissertation Online-Prozessgrößen zur Prädiktion der Zelldichte, der Glycerolkonzentration und der Produktkonzentration in einem zyklischen Produktionsprozess mit *Pichia pastoris*. Aus diesen Vorhersagen konnten dann zellspezifische Reaktionsraten online berechnet werden [BORCHERT, 2015].

Dieses Kapitel soll an die Arbeiten von Borchert anknüpfen. Hinsichtlich der Bestimmung zell-spezifischer Reaktionsraten wird jedoch ein anderer Ansatz verfolgt, der darin besteht, diese mittels multivariater Modelle direkt vorherzusagen. Hierbei war eine andere Bereitstellung geeigneter Kalibrierdaten zu finden.

10.2 Vorbereitung der Modellerstellung

10.2.1 Gewählte Zielgrößen

Primär sollten die geplanten Untersuchungen zunächst multivariate Modelle zur Vorhersage der Zustandsgrößen Zelldichte c_{XL} , Glycerolkonzentration c_{S1M} sowie der Zielproduktkonzentration c_{P1M} hervorbringen. Weiterhin sollte auch eine Prädiktion des zellinternen Alkoholoxidasegehaltes $g_{P2/X}$ mit dem Blackbox-Ansatz untersucht werden. Ein besonderes Interesse bestand dann in der Entwicklung multivariater Modelle für die direkte Vorhersage der zellspezifischen Reaktionsraten $q_{I/X}$. In Tabelle 10.1 sind die gewählten Zielgrößen aufgeführt.

Tabelle 10.1: Auflistung der gewählten Zielgrößen und Herkunft der Kalibrierdaten

h	Zielgröße y_h	Erzeugung geeigneter Kalibrierdaten
1	c_{XL} [gl^{-1}]	Anpassung von Simulationsverläufen durch
2	c_{S1M} [gl^{-1}]	Modellbildung, Simulation und Parameter-
3	c_{P1M} [gl^{-1}]	identifizierung
4	$g_{P2/X}$ [Ug^{-1}]	Interpolierte Werte eines enzymatischen Assays
5	$q_{X/X}$ [h^{-1}]	
6	$q_{O/X}$ [h^{-1}]	
7	$q_{S1/X}$ [h^{-1}]	Bilanzierung des Bioreaktionsprozesses unter Verwendung der simulierten Prozessgrößen
8	$q_{S2/X}$ [h^{-1}]	
9	$q_{P1/X}$ [h^{-1}]	

Ein wichtiger Aspekt ist die Bereitstellung geeigneter Kalibrierdaten. Für die untersuchten Zustandsgrößen standen vergleichsweise wenige Referenzmessungen zur Verfügung, die zusätzlich messtechnische Fehler beinhalteten. Dieser Umstand erschwert insbesondere die Erzeugung von Kalibrierdaten für die zellspezifischen Reaktionsraten $q_{I/X}$, da deren Berechnungen die zeitlichen Ableitungen der Zustandsgrößen des Prozesses benötigen.

Aus diesem Grund wurden für diese Untersuchungen differenzierbare Kalibrierdaten durch Simulation der Zielgrößen erzeugt. Darauf wird detailliert in Abschnitt 10.2.3 eingegangen. Im Folgenden werden zunächst die Rechenvorschriften der zellspezifischen Reaktionsraten $q_{I/X}$, die aus einer Bilanzierung des Bioreaktionssystems hervorgehen, hergeleitet.

10.2.2 Berechnung zellspezifischer Reaktionsraten

Die Darstellung der Zellteilungsrate $q_{X/X}$ in einem Fed-Batchprozess,

$$q_{X/X}(t) = \frac{F_{in}(t) - F_A}{V_L(t)} + \frac{\dot{c}_{XL}(t)}{c_{XL}(t)} \quad (10.1)$$

mit

F_{in} := Eingangsvolumenstrom $[l/h^{-1}]$

F_A := zeitinvariante zellfreier Atline-Analysenstrom $[l/h^{-1}]$

V_L := Volumen der Flüssigphase $[l]$

c_{XL} := Zelldichte in der Flüssigphase $[gl^{-1}]$,

lässt sich leicht aus der allgemeinen Bilanz für die Zellmasse m_{XL} herleiten.

Die Verdünnung der Zelldichte c_{XL} durch den Eingangsvolumenstrom F_{in} ,

$$F_{in}(t) = F_{R1}(t) + F_{R2}(t) + F_{T2}(t) \quad (10.2)$$

mit

F_{R1} := Glycerolzufütterrate $[l/h^{-1}]$

F_{R2} := Methanolzufütterate $[l/h^{-1}]$

F_{T2} := Volumenstrom der Titration mit Base $[l/h^{-1}]$,

wird darin als Summe der eingehenden Teilvolumenströme berücksichtigt.

Der zellfreie Atline-Analysenstrom F_A führt zu einer Konzentrierung der Zelldichte. Die Verdampfung der Medienphase mit gleichem Effekt wird dagegen vernachlässigt.

Eine Berechnung der zellspezifischen Sauerstoffaufnahmerate $q_{O_2/X}$,

$$q_{O_2/X}(t) = \frac{Q_{O_2}(t)}{c_{XL}(t)} \quad (10.3)$$

mit

Q_{O_2} := Sauerstoffeintragsrate $[gl^{-1}h^{-1}]$,

ist bei Annahme eines quasi-stationären Fließgleichgewichtes sehr einfach, da alle konvektiven Ströme vernachlässigbar und die benötigten Größen c_{XL} und Q_{O_2} online berechenbar sind.

Für weitere Berechnungen ist das Volumen der Medienphase V_M ,

$$V_M(t) = \left(1 - \frac{\alpha_{Z/X}}{\rho_Z} \cdot c_{XL}(t)\right) \cdot V_L(t) \quad (10.4)$$

mit

V_K := Volumen des Teilsystems K, $K = M, L$ $[l]$

$\alpha_{Z/X}$:= Verhältnis von Biofecht- zu Biotrockenmasse $[-]$

ρ_Z := Dichte der Biofeuchtphase $[gl^{-1}]$,

sowie die durch das Zellwachstum hervorgerufene Volumenverlustrate F_Z der Medienphase,

$$F_Z(t) = \frac{\alpha_{Z/X}}{\rho_Z} \cdot q_{X/X}(t) \cdot c_{XL}(t) \cdot V_L(t), \quad (10.5)$$

erforderlich.

Die zellspezifischen Substrataufnahmeraten $q_{Si/X}$ mit $i = 1$ für Glycerol und $i = 2$ für Methanol,

$$q_{Si/X}(t) = \frac{F_{Ri}(t) \cdot c_{SiRi}}{V_L(t) \cdot c_{XL}(t)} - \frac{F_{in}(t) - F_Z(t)}{V_L(t)} \cdot \frac{c_{SiM}(t)}{c_{XL}(t)} - \frac{V_M(t)}{V_L(t)} \cdot \frac{\dot{c}_{SiM}(t)}{c_{XL}(t)} \quad (10.6)$$

mit

c_{SiRi} := Konzentration des Substrates Si im entsprechenden Reservoir Ri $[g l^{-1}]$

c_{SiM} := Konzentration des Substrates Si in der Medienphase M $[g l^{-1}]$,

sowie die zellspezifische Zielproduktbildungsrat $q_{P1/X}$,

$$q_{P1/X}(t) = \frac{F_{in}(t) - F_Z(t)}{V_L(t)} \cdot \frac{c_{P1M}(t)}{c_{XL}(t)} + \frac{V_M(t)}{V_L(t)} \cdot \frac{\dot{c}_{P1M}(t)}{c_{XL}(t)}, \quad (10.7)$$

können aus der allgemeinen Bilanzgleichung einer Komponente I in der Medienphase M eines Dreiphasensystems hergeleitet werden [CORNELISSEN, 2004].

10.2.3 Bereitstellung idealisierter Kalibrierdaten

Aus dem vorherigen Abschnitt geht hervor, dass für die Berechnung der zellspezifischen Reaktionsraten die Ableitungen der Zustandsgrößen c_{XL} , c_{SiM} und c_{P1M} benötigt werden. Eine Differenzierung gemessener Werte ist numerisch jedoch sehr sensitiv und verstärkt etwaige Fehler gemessener Eingangsdaten [AHNERT & ABEL, 2007].

Aus diesem Grund wurden die Zustandsgrößen durch Simulation an die entsprechenden Offline- oder Atline-Messsignale angepasst. Dieses Vorgehen ermöglicht die Berechnung der Ableitungen ohne numerische Differenzierung der Originalmessdaten. In Bild 10.2 ist das Ergebnis der Anpassung der idealisierten Variablen $c_{ik\text{Kest}}$ grafisch dargestellt.

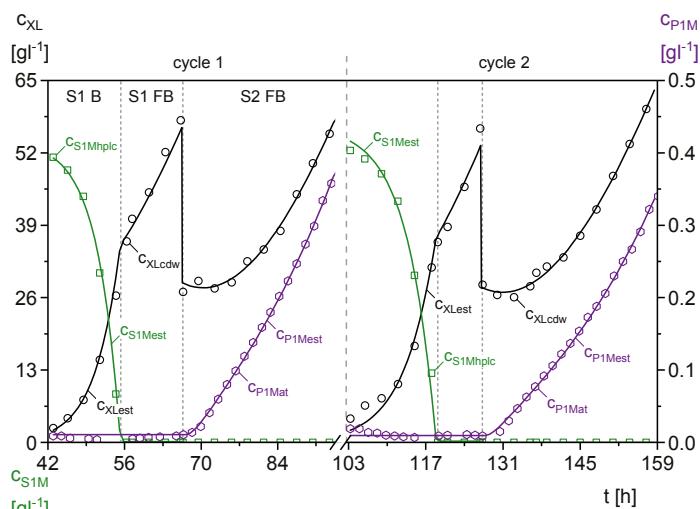


Bild 10.2: Anpassung der Zelldichte c_{XLest} , der Glycerolkonzentration c_{S1Mest} und der Produktkonzentration c_{P1Mest} an ihre Referenzmessungen in zwei Zyklen

Eine Simulation der Methanolkonzentration c_{S2M} war dabei nicht notwendig, da diese sehr gut geregelt wurde und somit aus c_{XL} und der Methanolzufütterrate F_{R2} die Berechnung eines wenig verrauchten Verlaufes der Methanolaufnahmerate $q_{S2/X}$ möglich war.

In einem ersten Schritt wurden die Zelldichte c_{XLest} und die Glycerolkonzentration c_{S1Mest} in den Glycerol-Batch- und Fed-Batchphasen über ein einfaches Modell des Bioreaktionsprozesses durch Identifikation von Reaktionsparametern und Anfangswerten angepasst.

Für die Berechnung der Zelldichte c_{XLest} in Zyklus k,

$$\dot{c}_{XLest}(t) = \left[q_{X/X}(t) - \frac{F_{in}(t) - F_A}{V_L(t)} \right] \cdot c_{XLest}(t) \quad (10.8)$$

mit

$$c_{XLest}(t_{0k}) = c_{XL0k}, \quad (10.8a)$$

wird eine reaktionskinetische Beschreibung der Zellteilungsrate $q_{X/X}$,

$$q_{X/X}(t) = \mu_{1maxk} \cdot \frac{c_{S1Mest}(t)}{c_{S1Mest}(t) + k_{S1}} \quad (10.9)$$

mit

$$\begin{aligned} \mu_{1maxk} &:= \text{maximale Zellwachstumsrate auf Glycerol in Zyklus k} & [h^{-1}] \\ k_{S1} &:= \text{Monod-Limitierungskonstante für Glycerol} & [gl^{-1}], \end{aligned}$$

benötigt, die in der Simulation einer einfachen Monod-Substratkinetik ohne Maintenacerate $q_{S1/Xm}$ folgen soll.

Die simulierte Glycerolaufnahmerate $q_{S1/X}$,

$$q_{S1/X}(t) = \frac{q_{X/X}(t)}{y_{X/S1k}} \quad (10.10)$$

mit

$$y_{X/S1k} := \text{Zellausbeutekoeffizient für Glycerol in Zyklus k} \quad [-],$$

verläuft unter Annahme eines zeitinvarianten Ausbeutekoeffizienten $y_{X/S1k}$ proportional zur Zellteilungsrate und ermöglicht über die Bilanz der Glycerolkonzentration,

$$\dot{c}_{S1Mest}(t) = \frac{F_{R1}(t)}{V_M(t)} \cdot c_{S1R1} - \frac{F_{in}(t) - F_Z(t)}{V_M(t)} \cdot c_{S1Mest}(t) - \frac{V_L(t)}{V_M(t)} \cdot q_{S1/X}(t) \cdot c_{XLest}(t) \quad (10.11)$$

mit

$$c_{S1Mest}(t_{0k}) = c_{S1M0k}, \quad (10.11a)$$

eine Berechnung der Glycerolkonzentration c_{S1Mest} in der Medienphase.

Die Bilanz des Flüssigvolumens V_L ,

$$\dot{V}_L(t) = F_{R1}(t) + F_{R2}(t) + F_{T2}(t) - F_A - F_S \quad (10.12)$$

mit

$$F_S := \text{zeitinvarianter gemittelter Probenahmeverolumenstrom} \quad [l h^{-1}]$$

und mit

$$V_L(t_{0k}) = V_{L0k}, \quad (10.12a)$$

setzt sich aus den ein- und austretenden Volumenströmen zusammen und wird zur Berechnung von V_M und F_z gemäß Gl. (10.4) und Gl. (10.5) benötigt.

Die Glycerolzufütterrate F_{R1} ,

$$F_{R1}(t) = \begin{cases} 0 & , t_{0k} \leq t < t_{jk} \\ F_{R1jk} \cdot e^{\mu_{1wk} \cdot (t - t_{jk})} & , t_{jk} \leq t < t_{0k+1} \end{cases} \quad (10.13)$$

mit

$$\begin{aligned} t_{0k} &:= \text{Startzeitpunkt des Zyklus } k & [h] \\ t_{jk} &:= \text{Beginn der Zufütterung in Zyklus } k & [h] \\ \mu_{1wk} &:= \text{fester Sollwert der Wachstumsrate in Zyklus } k & [h^{-1}] \end{aligned}$$

nimmt am Ende der jeweiligen Batchphase zum Zeitpunkt t_{jk} den Wert der initialen Glycerolzufütterrate F_{R1jk} ,

$$F_{R1jk} = F_{R1}(t_{jk}) = \frac{\mu_{1wk}}{y_{X/S1k} \cdot c_{S1R1}} \cdot c_{XL}(t_{jk}) \cdot V_L(t_{jk}), \quad (10.14)$$

an und wird im weiteren Verlauf bis t_{0k+1} mit einem exponentiellen Profil nachgeführt.

Das Ende der Batchphase ist durch den Verbrauch des Substrates charakterisiert. Die Zeit t_{jk} des Beginns der Glycerolzufütterung ist aus numerischen Gründen zu dem Zeitpunkt definiert, an dem die Glycerolkonzentration c_{S1Mest} ,

$$c_{S1Mest}(t_{jk}) < 0,1 \cdot k_{S1}, \quad (10.15)$$

ein Zehntel der angenommenen Limitierungskonstante k_{S1} unterschreitet.

Während die Methanolzufütterung F_{R2} in den Batch- und Fed-Batchphasen unberücksichtigt bleibt, wird der Basevolumenstrom F_{T2} ,

$$F_{T2}(t) = -\frac{\dot{m}_{T2}(t)}{\rho_{T2}} \quad (10.16)$$

mit

$$\begin{aligned} m_{T2} &:= \text{Masse des Basereservoirs} & [g] \\ \rho_{T2} &:= \text{Dichte des Basereservoirs} & [g \cdot l^{-1}] \end{aligned}$$

durch numerische Differentiation mit der Methode nach Savitzky und Golay [SAVITZKY & GOLAY, 1964] aus dem entsprechenden Online-Waagensignal m_{T2} approximiert. Dies ist hier möglich, da das Waagensignal ein hohes Signal-zu-Rauschverhältnis aufweist.

Die Anpassung der Simulationsverläufe der Zeldichte c_{XLest} und der Substratkonzentration c_{S1Mest} an die Offline-Messwerte erfolgte durch Minimierung des Gütefunktionalen J_{XSk} ,

$$J_{XSk} = \min_{\rho_k} \sum_{u=0}^{n_{Bk}-1} \left[\left(\frac{c_{XLest}(t_{uk}) - c_{XLcdw}(t_{uk})}{c_{XLmaxk}} \right)^2 + \left(\frac{c_{S1Mest}(t_{uk}) - c_{S1Mhpdc}(t_{uk})}{c_{S1Mmaxk}} \right)^2 \right] \quad (10.17)$$

mit

p_k	:= Vektor der unbekannten Parameter und Anfangswerte in Zyklus k	[div.]
n_{Bk}	:= Anzahl der Offline-Messwerte in der Glycerol-Batch- und Fed-Batchphase des betrachteten Zyklus k	[·]
t_{uk}	:= Offline-Messzeitpunkte innerhalb von Zyklus k	[h]
c_{XLcdw}	:= Offline-Messwert der Zelldichte	[gl ⁻¹]
$c_{S1Mhplc}$:= Offline-Messwert der Glycerolkonzentration	[gl ⁻¹]
c_{IKmaxk}	:= Maximalwerte der Offline-Messwerte zur Wichtung der Variablen	[gl ⁻¹],

für jeden der fünf betrachteten Kultivierungzyklen separat durch Identifizierung der unbekannten Modellparameter μ_{1maxk} , μ_{1wk} , $y_{X/S1k}$ sowie der Startwerte c_{XL0k} und c_{S1M0k} .

Als Optimierungsmethode wurde der Simplex-Algorithmus von Nelder und Mead gewählt [NELDER & MEAD, 1965]. In Tabelle 10.2 sind die identifizierten Modellparameter aller Zyklen aufgelistet. Tabelle 10.3 enthält dagegen die zyklusunabhängigen Parameter.

Tabelle 10.2: Modellparameter für die Erzeugung idealisierter Kalibrierdaten

Parameter	k = 1	2	3	4	5
μ_{1maxk} [h ⁻¹]	0,233	0,186	0,147	0,181	0,221
μ_{1wk} [h ⁻¹]	0,0533	0,0577	0,0614	0,0614	0,0769
$y_{X/S1k}$ [-]	0,678	0,662	0,699	0,711	0,631
c_{XL0k} [gl ⁻¹]	2,16	2,25	3,81	3,71	2,23
c_{S1M0k} [gl ⁻¹]	52,5	54,0	53,5	53,4	54,7

Tabelle 10.3: Zyklusunabhängige Modellparameter

Parameter	Wert	Parameter	Wert
c_{S1R1} [gl ⁻¹]	400	$\alpha_{Z/X}$ [-]	3,7
c_{S2R2} [gl ⁻¹]	790	ρ_{R2} [gl ⁻¹]	790
F_A [lh ⁻¹]	0,015	ρ_{T2} [gl ⁻¹]	950
F_S [lh ⁻¹]	0,010	ρ_Z [gl ⁻¹]	1050
k_{S1} [gl ⁻¹]	0,2		

Eine Simulation der Zelldichte während der Produktionsphase ist nicht zuverlässig möglich. Die Gründe hierfür liegen vor allem in der komplexen Natur der Induktionsmechanismen bei Umschaltung vom Glycerol- auf den Methanolstoffwechsel. Darüber hinaus ist Abnahme der Zellteilungsrate gegen Ende einiger Zyklen beobachtbar, jedoch nicht erklärbar.

Aus diesem Grund wurde in der Produktionsphase die Zelldichte c_{XLest} ,

$$c_{XLest}(t) = d_{0k} + d_{1k} \cdot (t - t_{0k}) + d_{2k} \cdot (t - t_{0k})^2 + d_{3k} \cdot (t - t_{0k})^3 \quad (10.18)$$

mit

$$t_{0k} := \text{Startzeitpunkt der Produktionsphase des Zyklus k} \quad [h],$$

für jeden Zyklus k mit einem Polynom dritten Grades approximiert.

Dieses Vorgehen erlaubt eine analytische Berechnung der Änderung von c_{XLest} ,

$$\dot{c}_{XLest}(t) = d_{1k} + 2 \cdot d_{2k} \cdot (t - t_{0k}) + 3 \cdot d_{3k} \cdot (t - t_{0k})^2. \quad (10.19)$$

Der Koeffizientenvektor d_k des Polynoms,

$$d_k = (X_k^\top \cdot X_k)^{-1} \cdot X_k^\top \cdot y_k, \quad (10.20)$$

lässt sich durch Lösung des überbestimmten Gleichungssystems,

$$\begin{bmatrix} c_{XLcdw}(t_{0k}) \\ c_{XLcdw}(t_{1k}) \\ \dots \\ c_{XLcdw}(t_{(n_{pk}-1)k}) \end{bmatrix}_{y_k} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & (t_{1k} - t_{0k}) & (t_{1k} - t_{0k})^2 & (t_{1k} - t_{0k})^3 \\ \dots & \dots & \dots & \dots \\ 1 & (t_{(n_{pk}-1)k} - t_{0k}) & (t_{(n_{pk}-1)k} - t_{0k})^2 & (t_{(n_{pk}-1)k} - t_{0k})^3 \end{bmatrix}}_{X_k} \cdot \begin{bmatrix} d_{0k} \\ d_{1k} \\ d_{2k} \\ d_{3k} \end{bmatrix}_{d_k}, \quad (10.21)$$

mit einem Ansatz der kleinsten Fehlerquadrate,

$$J_{Xk} = \min_{d_k} \sum_{u=0}^{n_{pk}-1} (c_{XLest}(t_{uk}) - c_{XLcdw}(t_{uk}))^2 \quad (10.22)$$

mit

d_k := Vektor der unbekannten Koeffizienten d_{jk} in Zyklus k [gl⁻¹h⁻¹]

n_{pk} := Anzahl der Offline-Messwerte in der Produktionsphase von Zyklus k [-],

für jeden Zyklus k leicht berechnen. Die Koeffizienten sind in Tabelle 10.4 aufgelistet.

Tabelle 10.4: Koeffizienten der Polynome zur Berechnung von c_{XLest} in der Produktionsphase

Parameter	k = 1	2	3	4	5
t_{0k} [h]	66,5	127,1	187,1	245,8	305,8
d_{0k} [gl ⁻¹]	28,7	27,9	26,2	27,7	26,5
d_{1k} [gl ⁻¹ h ⁻¹]	-0,495	-0,482	-0,318	-0,564	-0,325
d_{2k} [gl ⁻¹ h ⁻²]	$6,76 \cdot 10^{-2}$	$6,55 \cdot 10^{-2}$	$2,55 \cdot 10^{-2}$	$7,21 \cdot 10^{-2}$	$6,24 \cdot 10^{-2}$
d_{3k} [gl ⁻¹ h ⁻³]	$-4,37 \cdot 10^{-4}$	$-4,51 \cdot 10^{-4}$	$9,55 \cdot 10^{-5}$	$-6,05 \cdot 10^{-4}$	$-6,10 \cdot 10^{-4}$

Unter Verwendung der Zelldichte c_{XLest} lässt sich die Produktionskonzentration c_{P1Mest} im Medium,

$$\dot{c}_{P1Mest}(t) = -\frac{F_{in}(t) - F_Z(t)}{V_M(t)} \cdot c_{P1Mest}(t) + \frac{V_L(t)}{V_M(t)} \cdot q_{P1/X}(t) \cdot c_{XLest}(t) \quad (10.23)$$

mit

$$c_{P1Mest}(t_{0k}) = c_{P1M0k}, \quad (10.23a)$$

durch Simulation dieser Bilanzgleichung berechnen. Die zeitliche Ableitung ergibt sich dann direkt aus Gl. (10.23).

Die in der Produktionsphase relevante Methanolzufütterrate F_{R2} ,

$$F_{R2}(t) = -\frac{\dot{m}_{R2}(t)}{\rho_{R2}} \quad (10.24)$$

mit

$$\begin{aligned} m_{R2} &:= \text{Masse des Methanolreservoirs} & [g] \\ \rho_{R2} &:= \text{Dichte des Methanolreservoirs} & [g l^{-1}] \end{aligned}$$

wird analog zu F_{T2} (Gl. (10.16)) aus dem Signal der Methanolreservoirwaage m_{R2} ermittelt.

Die zellspezifische Produktbildungsrates $q_{P1/Xest}$

$$q_{P1/Xest}(t) = a_{P1k} \cdot (q_{P1/Xmaxk} - q_{P1/Xest}(t)) \quad (10.25)$$

mit

$$q_{P1/Xest}(t_{0k}) = 0 \text{ h}^{-1} \quad (10.25a)$$

und mit

$$\begin{aligned} a_{P1k} &:= \text{Eigenwert der Produktbildungsrates in Zyklus k} & [\text{h}^{-1}] \\ q_{P1/Xmaxk} &:= \text{maximale Produktbildungsrates in Zyklus k} & [\text{h}^{-1}] \end{aligned}$$

wird als P-T₁-Element simuliert und stellt damit eine vereinfachte Beschreibung des Induktionsvorganges dar.

Die Anpassung des Simulationsverlaufes der Produktkonzentration $c_{P1/Mest}$ in der Medienphase an die Messwerte der Atline-Analytik erfolgte durch Minimierung des Gütefunktionalen J_{P1k} ,

$$J_{P1k} = \min_{a_{P1k}, q_{P1/Xmaxk}, c_{P1M0k}} \sum_{i=0}^{n_{Ak}-1} (c_{P1Mest}(t_{ik}) - c_{P1Mat}(t_{ik}))^2 \quad (10.26)$$

mit

$$\begin{aligned} n_{Ak} &:= \text{Anzahl der Atline-Messwerte des Zielproduktes in Zyklus k} & [-] \\ t_{ik} &:= \text{Atline-Messzeitpunkte innerhalb von Zyklus k} & [\text{h}] \end{aligned}$$

für jeden der fünf betrachteten Kultivierungszyklen durch Identifikation der Parameter $q_{P1/Xmaxk}$, a_{P1k} sowie der Startwerte c_{P1M0k} . Diese sind in Tabelle 10.5 dargestellt.

Tabelle 10.5: Identifizierte Parameter zur Simulation des Zielproduktes

Parameter	$k = 1$	2	3	4	5
t_{0k} [h]	66,5	127,1	187,1	245,8	305,8
a_{P1k} [h^{-1}]	1,34	1,05	0,599	0,810	0,855
$q_{P1/Xmaxk}$ [h^{-1}]	0,405	0,309	0,186	0,270	0,351
c_{P1M0k} [mg l^{-1}]	12,6	9,66	5,43	8,51	17,0

Die Produktkonzentration $c_{P1/Mest}$ in den Glycerol-Batch- und Fed-Batchphasen wurde als konstant angenommen und dem identifizierten Startwert c_{P1M0k} in der Produktionsphase des entsprechenden Zyklus gleichgesetzt.

Geeignete Kalibrierdaten für die zellspezifische AOX-Aktivität $g_{P2/Xest}$ wurden durch lineare Interpolation aus den Offline-Messdaten (siehe Abschnitt 3.4.6) erzeugt.

10.2.4 Auswahl und Bereitstellung der Prädiktorvariablen

Als mögliche Prädiktorvariablen wurden zunächst alle online am Prozess erhobenen Größen in Betracht gezogen. Hierbei kann es sich beispielsweise um Inline-Messungen von Temperatur und pH-Wert als auch um Waagensignale der Substrat- und Titrationsvorlagen handeln. Auch online leicht berechenbare Größen wie die Ergebnisse der Abgasbilanzierung (z.B. O₂-Eintragsrate Q_{O₂}) wurden als X-Variablen berücksichtigt.

Größen, die während des gesamten Prozesses unverändert bleiben, enthalten keine relevanten Informationen zur Vorhersage der Zielgrößen und fallen deshalb als X-Variablen aus. Allerdings kommen diese Größen ggf. bei der Berechnung anderer Prädiktoren zum Einsatz. Beispiele sind die Begasungsrate F_{nG} und der Gelöstsauerstoffpegel pO₂.

Weiterhin blieben das Flüssigvolumen V_L und der Respirationskoeffizient RQ unberücksichtigt, da diese eine Verschlechterung der Vorhersagekraft der multivariaten Modelle verursachten. Dies konnte in Voruntersuchungen festgestellt werden.

Insgesamt kamen 15 Online-Prozessgrößen für die Vorhersage der vorgestellten Zielvariablen zum Einsatz. Diese sind in Tabelle 10.6 aufgeführt.

Tabelle 10.6: Auflistung der eingesetzten Online-Prädiktorvariablen

j	Variable	Beschreibung	Erzeugung der Werte
1	E _L	[AU]	Trübung
2	c _{S2M}	[gl ⁻¹]	Methanolkonzentration
3	θ _L	[°C]	Temperatur
4	pH	[·]	pH-Wert
5	Δm _{R1}	[g]	Masse Glycerolreservoir
6	Δm _{R2}	[g]	Masse Methanolreservoir
7	Δm _{T2}	[g]	Masse Basereservoir
8	F _{R1w}	[l/h ⁻¹]	Glycerolzufütterrate
9	F _{R2w}	[l/h ⁻¹]	Methanolzufütterrate
10	N _{Stw}	[min ⁻¹]	Rührerdrehzahl
11	x _{OG}	[·]	Sauerstoffgehalt
12	x _{CG}	[·]	Kohlenstoffdioxidgehalt
13	Q _{O₂}	[gl ⁻¹ h ⁻¹]	O ₂ -Eintragsrate
14	Q _{CO₂}	[gl ⁻¹ h ⁻¹]	CO ₂ -Austragsrate
15	k _L a	[h ⁻¹]	O ₂ -Transferkoeffizient

10.2.5 Erzeugung benötigter Datensätze und Datenvorverarbeitung

Durch die Generierung simulierter Y-Variablen und die Herkunft der X-Variablen aus Inline- und Online-Messungen oder Berechnungen stehen mehrere Tausend Beobachtungen für die fünf Zyklen des untersuchten Prozesses zur Verfügung.

Da der Rechenaufwand bei der SVR mit der Anzahl n an Beobachtungen stark ansteigt und dieses Regressionsverfahren mit der PLSR verglichen werden sollte, wurde die Beobachtungen auf eine Anzahl reduziert, die für eine SVR-Modellerstellung noch praktisch sinnvoll ist.

Durch lineare Interpolation wurden aus den Vektoren der estimaerten Zielgrößen sowie der gemessenen X-Variablen insgesamt 380 Zeitpunkte als Beobachtungen für die Modellbildung ausgewählt. Diese haben innerhalb der Zyklen jeweils einen Abstand von ca. 45 Minuten.

Die Prozessgrößen des X-Datenraumes wurden im Folgenden mit Hilfe einer PCA auf Ausreißer untersucht und zwei Datenpunkte von der Modellbildung ausgeschlossen. Zur Durchführung einer externen Validierung wurden die Zyklen 3, 4 und 5 als Kalibrierdatensatz gewählt. Die Zyklen 1 und 2 dienen folglich als Validierdatensatz.

Für beide getesteten Verfahren wurde als Datenvorverarbeitungsmethode des X- und des Y-Datenraums die Autoskalierung gewählt. Im Falle der PLSR wurden die X-Variablen zusätzlich mittenzentriert.

10.3 Ergebnisdarstellung

10.3.1 Prädiktion von Zustandsgrößen

Bild 10.3 zeigt die Vorhersagen der Zelldichte c_{XL} sowie der Glycerolkonzentration c_{S1M} mit der PLSR im Rahmen einer Kreuzvalidierung. In der Tendenz können die Größen vorhergesagt werden, jedoch werden zu Beginn der Zyklen, insbesondere in 3 und 4, die Zelldichte zu niedrig und die Glycerolkonzentration zu hoch vorhergesagt.

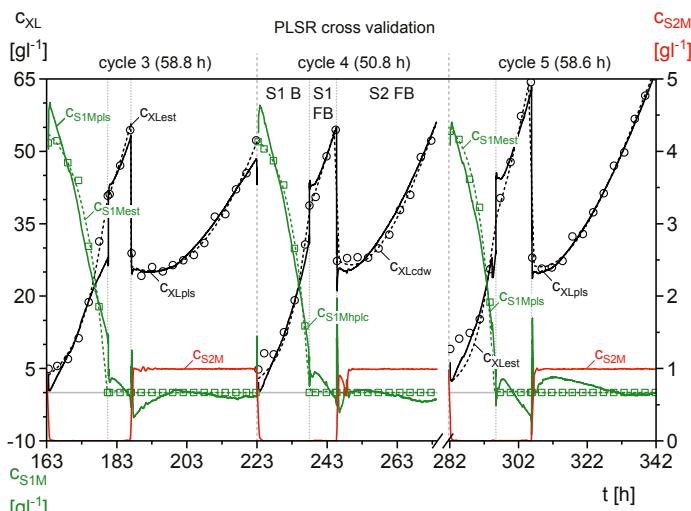


Bild 10.3: PLSR-Vorhersagen $c_{XL\text{Pls}}$ und $c_{S1M\text{Pls}}$ für den Kalibrierdatensatz

Die PLSR war somit nicht dazu in der Lage, die exponentiellen Verläufe dieser Zustandsgrößen mit den gewählten X-Variablen zufriedenstellend abzubilden, da die Vorhersagen eher lineare Verläufe zeigen. Zusätzlich wurden Konzentrationen $c_{S1\text{Mest}}$ von 0 gl^{-1} in den Glycerol- und den Methanol-Fed-Batchphasen nur schlecht abgeschätzt.

Ein besseres Ergebnis lässt sich mit der SVR erzielen. Dies ist in Bild 10.4 dargestellt. Die exponentiellen Verläufe beider Zielgrößen können durch die nichtlineare Methode in der Regel sehr gut wiedergegeben werden.

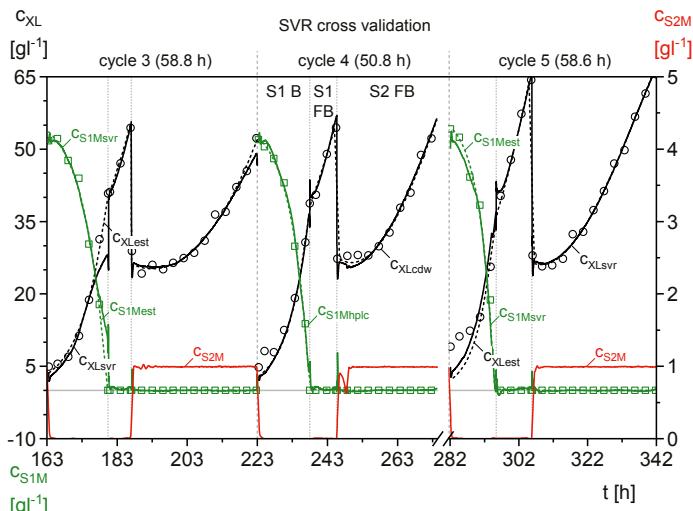


Bild 10.4: SVR-Vorhersagen c_{XLsvr} und c_{S1Msvr} für den Kalibriertatensatz

Bild 10.5 zeigt im Vergleich dazu die Vorhersagen mit der SVR für den Validierdatensatz. Hier weist c_{S1Msvr} zu Beginn beider Zyklen Schwächen auf. Weiterhin wird der Zelldichtezuwachs im Glycerol-Fed-Batch des Zyklus 1 als zu stark vorhergesagt.

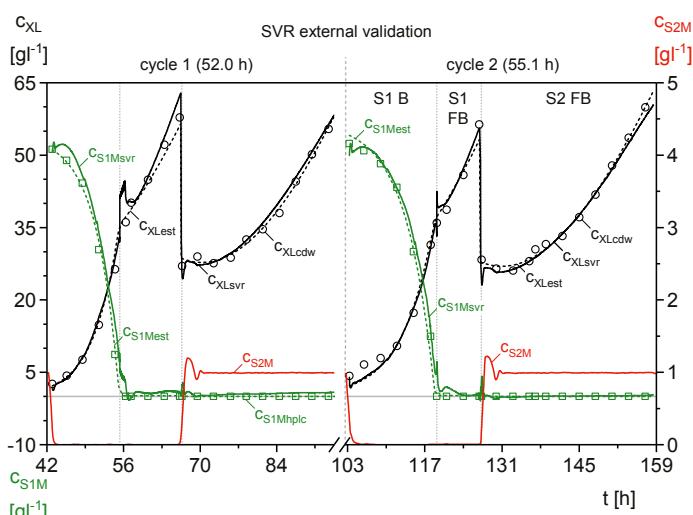


Bild 10.5: SVR-Vorhersagen c_{XLsvr} und c_{S1Msvr} für den Validierdatensatz

Die Unterschiede beider Regressionsmethoden werden bei Betrachtung der folgenden Abbildungen anhand der Prädiktion von c_{S1M} deutlich. Für die PLSR (Bild 10.6a) zeigt sich ein klar nichtlinearer Zusammenhang zwischen Vorhersage c_{S1Mpls} und den Referenzwerten c_{S1Mest} . Dieser ist im Fall der SVR (Bild 10.6b) wesentlich weniger stark ausgeprägt. Daher ist die SVR in diesem Fall besser für die Anwendung auf Online-Prozessdaten geeignet als die PLSR.

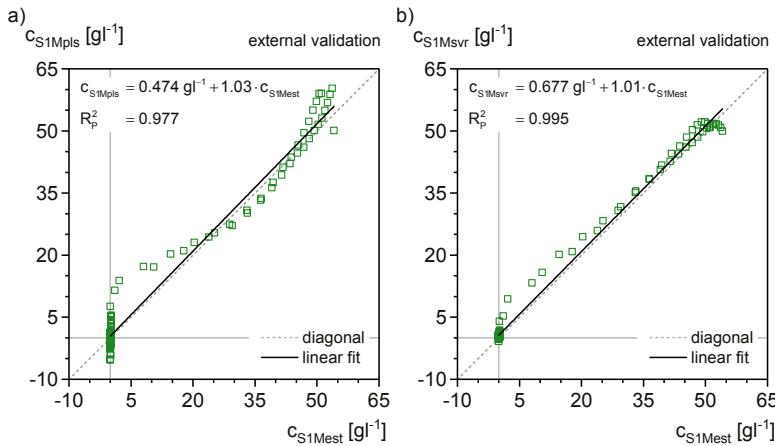


Bild 10.6: Vorhersagen von c_{S1M} aufgetragen gegen ihre Referenzwerte, a) PLSR, b) SVR

Auch für die Vorhersage der Zielproduktkonzentration c_{P1M} und des zellinternen AOX-Gehaltes $g_{P2/X}$ werden unterschiedliche Resultate erzielt. Bild 10.7 zeigt die besseren Vorhersagen mit der SVR, während auf eine grafische Darstellung der Ergebnisse mit der PLSR verzichtet wird.

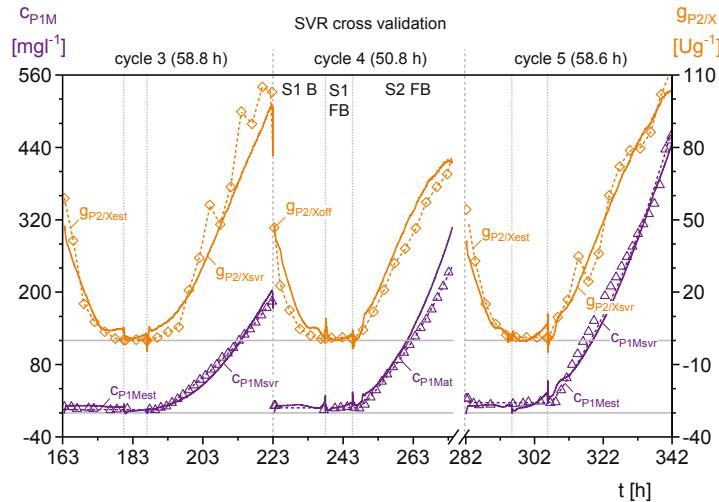


Bild 10.7: SVR-Vorhersagen c_{P1Msvr} und $g_{P2/Xsvr}$ für den Kalibrierdatensatz

Eine gute Prädiktion für die Zielproduktkonzentration c_{P1M} wird insbesondere in den Zyklen 3 und 5 erzielt. Niedrige Konzentrationen in den Wachstumsphasen auf Glycerol sowie der exponentielle Anstieg in den Produktionsphasen können sehr gut wiedergegeben werden.

Der prinzipielle Verlauf des AOX-Gehaltes $g_{P2/Xest}$ ist auch durch die SVR darstellbar. Schwächen zeigt das Modell allerdings bei der Abbildung der Höhe und der exponentiellen Abnahme von $g_{P2/Xest}$ zu Beginn der Glycerol-Batchphasen und weist dort auch eher lineare Verläufe auf.

In Bild 10.8 sind die Vorhersagen von c_{P1M} und von $g_{P2/X}$ der SVR-Modelle für den Validierdatensatz gezeigt. Diese weisen sehr ähnliche Verläufe, wie bei der Kreuzvalidierung auf.

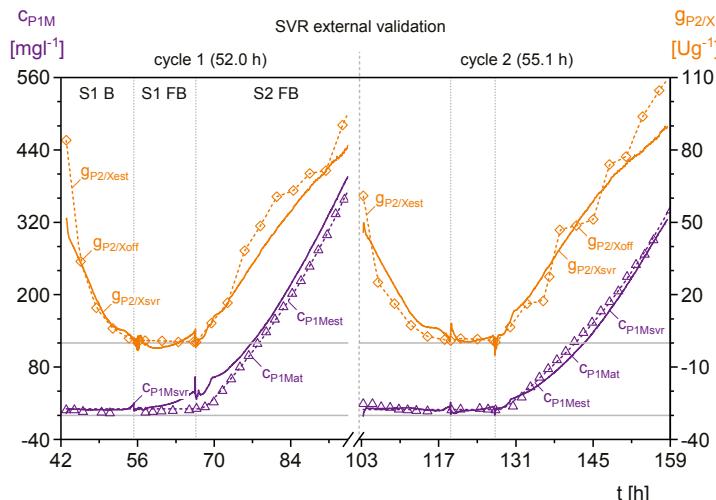


Bild 10.8: SVR-Vorhersagen c_{P1Msvr} und g_{P2Xsvr} für den Validierdatensatz

Die folgenden beiden Tabellen enthalten abschließend alle Kenngrößen der PLSR-Modelle (Tabelle 10.7) und der SVR-Modelle (Tabelle 10.8) für die untersuchten Zustandsgrößen.

Tabelle 10.7: Kenngrößen von PLSR-Modellen der Zustandsgrößen, $n_{CS} = 231$, $n_{VS} = 147$

		C_{XL}	C_{S1M}	C_{P1M}	$g_{P2/X}$
r	[\cdot]	5	6	6	6
R_x^2	[\cdot]	0,988	0,978	0,983	0,990
R_y^2	[\cdot]	0,987	0,978	0,971	0,919
R_{cv}^2	[\cdot]	0,986	0,973	0,968	0,913
RMSEcv	[div.]	$1,84 \text{ gl}^{-1}$	$2,95 \text{ gl}^{-1}$	$18,3 \text{ mg l}^{-1}$	$9,61 \text{ Ug l}^{-1}$
RMSEcv _{rel}	[%]	2,87	5,47	4,02	8,65
R_p^2	[\cdot]	0,985	0,977	0,972	0,918
RMSEP	[div.]	$2,30 \text{ gl}^{-1}$	$2,99 \text{ gl}^{-1}$	$18,1 \text{ mg l}^{-1}$	$9,52 \text{ Ug l}^{-1}$
RMSEP _{rel}	[%]	3,58	5,54	3,99	8,56

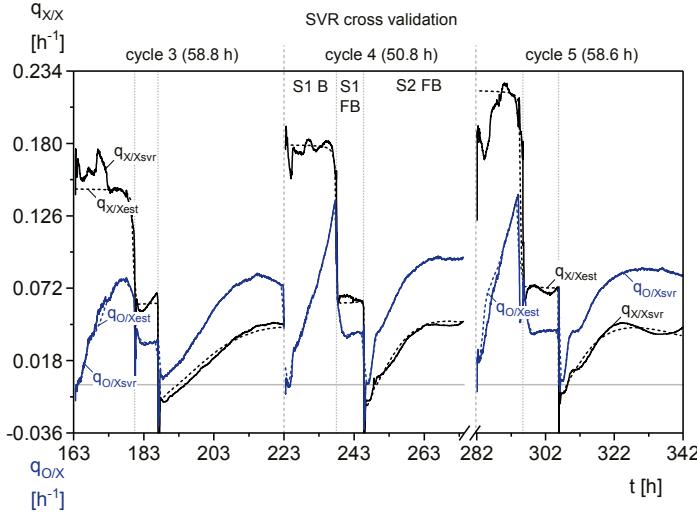
Tabelle 10.8: Kenngrößen von SVR-Modellen der Zustandsgrößen, $n_{CS} = 231$, $n_{VS} = 147$

		c_{XL}	c_{S1M}	c_{P1M}	$g_{P2/X}$
n_{sv}	[·]	104	86	112	111
C	[·]	4,22	0,962	4,64	3,48
γ	[·]	$9,08 \cdot 10^{-3}$	$1,14 \cdot 10^{-1}$	$4,65 \cdot 10^{-3}$	$1,61 \cdot 10^{-2}$
ε	[·]	$3,70 \cdot 10^{-2}$	$1,17 \cdot 10^{-2}$	$7,18 \cdot 10^{-2}$	$1,44 \cdot 10^{-1}$
R^2_{cv}	[·]	0,990	0,993	0,981	0,949
RMSEcv	[div.]	$1,59 \text{ gl}^{-1}$	$1,50 \text{ gl}^{-1}$	$14,4 \text{ mg l}^{-1}$	$7,52 \text{ Ug}^{-1}$
RMSEcv _{rel}	[%]	2,49	2,86	3,17	6,77
R^2_p	[·]	0,981	0,995	0,984	0,964
RMSEP	[div.]	$2,01 \text{ gl}^{-1}$	$1,60 \text{ gl}^{-1}$	$14,6 \text{ mg l}^{-1}$	$7,71 \text{ Ug}^{-1}$
RMSEP _{rel}	[%]	3,14	2,96	3,21	6,94

Diese Tabellen zeigen noch einmal eindrucksvoll, dass für alle vier Variablen bessere Ergebnisse mit der nichtlinearen Methode SVR erzielt werden konnten.

10.3.2 Prädiktion zellspezifischer Reaktionsraten

Die Vorhersagen mittels SVR für die Zellteilungsrate $q_{X/X}$ und die zellspezifische Sauerstoffaufnahmerate $q_{O/X}$ sind in Bild 10.9 für den Kalibrierdatensatz gezeigt.

Bild 10.9: SVR-Vorhersagen $q_{X/Xsvr}$ und $q_{O/Xsvr}$ für den Kalibrierdatensatz

Es fällt auf, dass die Vorhersage $q_{X/Xsvr}$ während der Glycerol-Batchphasen im Vergleich zu den anderen Prozessabschnitten stark schwanken, im Mittel jedoch die Höhe ihrer Referenz-

werte erreichen. Eine Ursache hierfür konnte nicht identifiziert werden. Möglich ist jedoch, dass das reale Verhalten eher diesem Verlauf entspricht, als dem der idealisierten Trainingsdaten.

Bei der externen Validierung, gezeigt in Bild 10.10, sind diese Schwankungen etwas stärker ausgeprägt. Bei der Verwendung eines unabhängigen Testsets sind jedoch höhere Fehler zu erwarten. Insgesamt wird die Zellteilungsrate zufriedenstellend abgebildet.

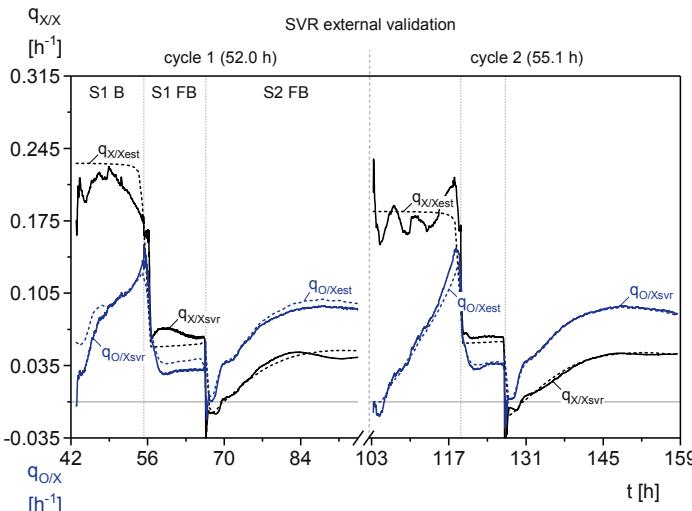


Bild 10.10: SVR-Vorhersagen $q_{X/Xsvr}$ und $q_{O2/Xsvr}$ für den Validierdatensatz

Die Sauerstoffaufnahmerate q_{O2X} kann hervorragend wiedergegeben werden. Abgesehen von Zyklus 1 des Validierdatensatzes werden sehr gute Übereinstimmungen mit den Referenzdaten erreicht.

Bild 10.11 und Bild 10.12 enthalten die entsprechenden SVR-Vorhersagen der zellspezifischen Substrataufnahmeraten $q_{S1/X}$ und der Zielproduktbildungsraten $q_{P1/X}$.

Die Prädiktion der Glycerolaufnahmerate $q_{S1/X}$ zeigt in den Batchphasen ein ähnlich schwankendes Verhalten wie die Zellteilungsrate $q_{X/Xsvr}$. Der Verlauf in den Glycerol- und Methanol-Fed-Batchphasen wird, insbesondere für die Kreuzvalidierung (Bild 10.11), dagegen sehr gut getroffen. In diesen Prozessabschnitten zeigen sich die höchsten Abweichungen im Zyklus 1 der externen Validierung (Bild 10.12).

Die Vorhersage der Zielproduktbildungsraten $q_{P1/X}$ zeigt dagegen in Zyklus 2 des Validierdatensatzes den größten Fehler. Der Grund für dieses Verhalten könnte in der individuellen Erzeugung der Referenzdaten für jeden Zyklus liegen. Der prinzipielle Verlauf stimmt jedoch auch hier mit dem der estimierten Referenzwerte $q_{P1/Xest}$ überein. Insgesamt wird die Prädiktion von $q_{P1/X}$ als zufriedenstellend erachtet.

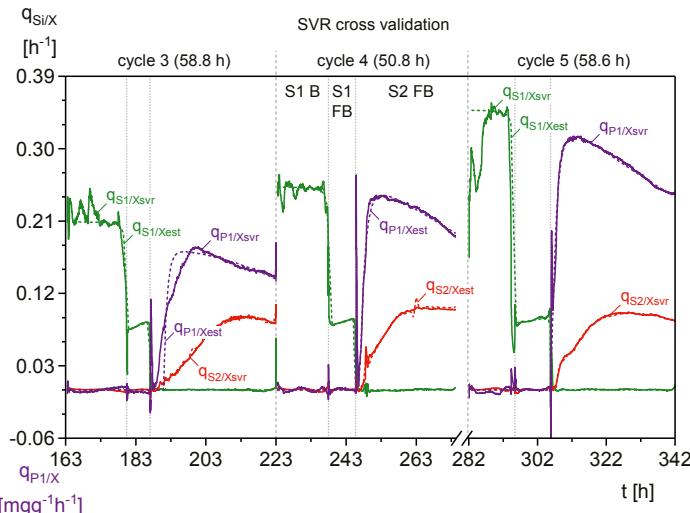


Bild 10.11: SVR-Vorhersagen $q_{Si/Xsvr}$ und $q_{P1/Xsvr}$ für den Kalibrierdatensatz

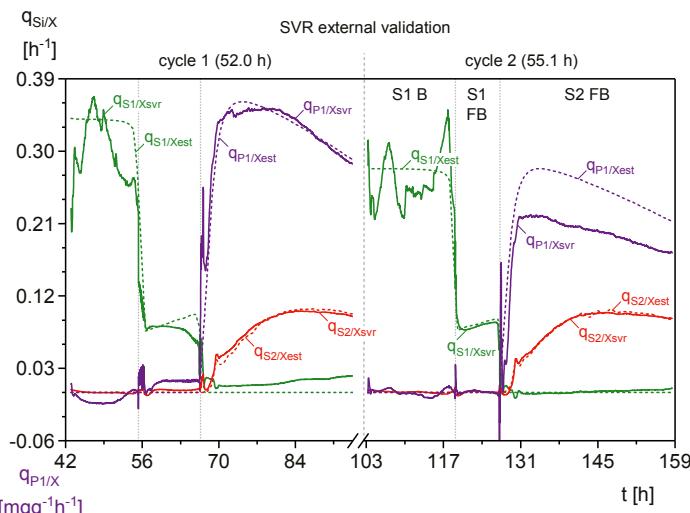


Bild 10.12: SVR-Vorhersagen $q_{Si/Xsvr}$ und $q_{P1/Xsvr}$ für den Validierdatensatz

Das beste Ergebnis konnte für die zellspezifische Methanolaufnahmerate $q_{S2/X}$ erzielt werden. Sowohl für die interne als auch für die externe Validierung wurden relative Vorhersagefehler von unter 3 % ermittelt. Außerdem unterscheidet sich das Ergebnis für die beiden Zyklen des Validierdatensatzes nicht.

In den folgenden beiden Tabellen sind die Kenngrößen der PLSR-Modelle (Tabelle 10.9) und der vorgestellten SVR-Modelle (Tabelle 10.10) für die zellspezifischen Reaktionsraten $q_{i/x}$ abschließend aufgeführt.

Auch bei der Vorhersage von $q_{i/x}$ erzielt die SVR die besseren Ergebnisse als die PLSR.

Tabelle 10.9: Kenngrößen von PLSR-Modellen für die zellspezifischen Reaktionsraten $q_{i/x}$
 $n_{CS} = 231, n_{VS} = 147$

		$q_{X/X}$	$q_{O/X}$	$q_{S1/X}$	$q_{S2/X}$	$q_{P1/X}$
r	[·]	4	5	4	4	5
R^2_X	[·]	0,962	0,981	0,959	0,961	0,968
R^2_Y	[·]	0,925	0,904	0,935	0,955	0,935
R^2_{cv}	[·]	0,915	0,890	0,931	0,953	0,930
RMSEcv	[h^{-1}]	$1,91 \cdot 10^{-2}$	$9,61 \cdot 10^{-3}$	$2,98 \cdot 10^{-2}$	$9,26 \cdot 10^{-3}$	$3,09 \cdot 10^{-5}$
RMSEcv _{rel}	[%]	8,10	6,91	8,62	8,05	9,80
R^2_P	[·]	0,936	0,809	0,961	0,959	0,954
RMSEP	[h^{-1}]	$2,17 \cdot 10^{-2}$	$1,41 \cdot 10^{-3}$	$3,14 \cdot 10^{-2}$	$9,25 \cdot 10^{-3}$	$3,79 \cdot 10^{-5}$
RMSEP _{rel}	[%]	9,20	10,1	9,03	8,04	12,0

Tabelle 10.10: Kenngrößen von SVR-Modellen für die zellspezifischen Reaktionsraten $q_{i/x}$
 $n_{CS} = 231, n_{VS} = 147$

		$q_{X/X}$	$q_{O/X}$	$q_{S1/X}$	$q_{S2/X}$	$q_{P1/X}$
n_{sv}	[·]	59	86	104	31	74
C	[·]	87,5	311	19,6	28,7	652
γ	[·]	$5,11 \cdot 10^{-2}$	$3,76 \cdot 10^{-2}$	$3,48 \cdot 10^{-1}$	$4,65 \cdot 10^{-2}$	$1,75 \cdot 10^{-2}$
ε	[·]	$7,69 \cdot 10^{-2}$	$3,45 \cdot 10^{-2}$	$1,55 \cdot 10^{-2}$	$7,69 \cdot 10^{-2}$	$3,22 \cdot 10^{-2}$
R^2_{cv}	[·]	0,979	0,973	0,962	0,994	0,964
RMSEcv	[h^{-1}]	$9,70 \cdot 10^{-3}$	$4,77 \cdot 10^{-3}$	$2,20 \cdot 10^{-2}$	$3,18 \cdot 10^{-3}$	$2,26 \cdot 10^{-2}$
RMSEcv _{rel}	[%]	4,11	3,43	6,33	2,76	7,16
R^2_P	[·]	0,973	0,935	0,966	0,996	0,962
RMSEP	[h^{-1}]	$1,32 \cdot 10^{-2}$	$9,14 \cdot 10^{-3}$	$2,57 \cdot 10^{-2}$	$3,15 \cdot 10^{-3}$	$3,22 \cdot 10^{-2}$
RMSEP _{rel}	[%]	5,58	6,57	7,38	2,74	10,2

11. Zusammenfassung

Die *Process Analytical Technology* (PAT) Initiative der U.S. Food and Drug Administration (FDA) empfiehlt den Einsatz fortschrittlicher Analysensysteme in pharmazeutischen Produktionsprozessen. Durch den Nachweis einer gesteigerten Prozesskenntnis und einer wissensbasierten Prozessführung sollen Zulassungsverfahren beschleunigt und durch den Einsatz technischer Innovationen die Produktqualität sichergestellt werden.

Als erste Aufgabe dieser Arbeit wurde die Entwicklung eines Qualitätsbewertungssystems für Hefeextrakte, basierend auf Messungen mit NIR-Spektroskopie und multivariater Klassifizierung vorgenommen. Eine solche Prüfung von Rohmaterialien stellt eine typische PAT-Anwendung dar.

Die Methode *Soft Independent Modelling of Class Analogy* (SIMCA) brachte ein sehr gutes Klassifikationsmodell hervor, dessen Fehler E_c für alle drei untersuchten Hefeextraktsorten unter 1,5 % liegt. Der Schwachpunkt hierbei war das nur in begrenztem Umfang vorhandene Probenmaterial.

Der überwiegende Teil dieser Arbeit widmet sich der Beobachtung von pharmazeutischen Kultivierungsprozessen der methylotrophen Hefe *Pichia pastoris* zur Herstellung des potentiellen Malariavakzins D1M1H.

Dabei wurden nicht direkt bestimmbare Prozessgrößen über spektroskopische Messungen und multivariate Kalibrierverfahren online quantifiziert. Diese Erweiterung des Bioprozessmonitorings erhöht das Prozessverständnis und stellt damit einen wichtigen Schritt bei der Etablierung von PAT dar.

In einem ersten Abschnitt dieses Arbeitspaketes wurden im Rahmen einer Machbarkeitsstudie die NIR-, die Raman- und die 2D-Fluoreszenzspektroskopie hinsichtlich Ihrer Eignung bei der Vorhersage von fünf Bioprozessgrößen mit der multivariaten Regressionsmethode *Partial Least Squares* (PLSR) in einem zyklischen Produktionsprozess untersucht und verglichen.

Die spektralen Messungen fanden hier zunächst offline, also nach Abschluss des Prozesses statt. Aufgrund des Verzichts auf Inline-Messung wurden dadurch einige Störquellen, wie die Begasung in Verbindung mit einer starken Durchmischung des Bioreaktors, ausgeschlossen. Weiterhin erlaubte die Offline-Methode eine Untersuchung des Einflusses der Zellen auf die PLSR-Modelle.

Diese Arbeiten brachten optimale Einstellungen für Kalibriermodelle, beispielsweise hinsichtlich der Vorverarbeitung der Spektren, des gewählten Spektralbereiches sowie der Anzahl zu berechnender PLS-Komponenten, hervor.

Gute Ergebnisse mit relativen Vorhersagefehlern $RMSE_{rel}$ von < 5 % konnten für die Zelldichte c_{XL} mit allen getesteten Spektroskopiearten erreicht werden. Die niedrigsten Fehler für Konzentrationen von Glycerol (c_{S1L}) und Ammonium (C_{AltotL}) in der Flüssigphase L wurden dagegen mit Raman-Spektroskopie erzielt. Dabei stellt die Vorhersage von c_{S1L} mit einem Fehler von 3,3 % ein besonders gutes Ergebnis dar.

Die 2D-Fluoreszenzspektroskopie ergab die besten Resultate bei der Vorhersage der Gesamtproteinconzentration $c_{ProteinM}$ in der Medienphase mit einem Fehler von 6,7 %. Auch die Prädiktion der zellspezifischen Alkoholoxidaseaktivität $g_{P2/X}$ war mit diesem Messverfahren möglich.

In einer weiteren Phase dieser Arbeit wurden die Erkenntnisse der Machbarkeitsstudie auf den Online-Betrieb am Bioreaktor mit spektroskopischer Inline-Messung in einem zweistufigen Produktionsprozess übertragen.

Die Modellbildung für c_{X_L} und c_{S1L} fokussierte sich auf die Raman-Spektroskopie. Neben der linearen Regressionsmethode PLSR wurde dabei auch das nichtlineare Verfahren *Support Vector Regression* (SVR) erprobt. Die SVR erlaubte eine Verbesserung der Vorhersage im Vergleich mit der PLSR und erreichte für beide Variablen sehr gute Ergebnisse mit Fehlern von etwa 3 %.

Während eine Übertragung der Vorhersage obiger Variablen auf den Online-Betrieb erfolgreich war, schlug die Umsetzung der Quantifizierung der Gesamtproteininkonzentration $c_{P_{tot}}$ und von $g_{P2/X}$ leider fehl. Eine wahrscheinliche Ursache liegt in dem hohen Einfluss der Zell-dichte auf die spektralen Messungen und damit auf die Leistung der Kalibriermodelle.

Die prinzipielle Eignung gebildeter Online-PLSR-Modelle für eine prozesstechnische Anwendung unter PAT-Gesichtspunkten konnte durch die Regelung der Glycerolkonzentration c_{S1L} , basierend auf Raman-Messungen, aufgezeigt werden.

Hierzu erfolgten zunächst eine theoretische Beschreibung des Regelungsproblems und die Entwicklung einer Strategie mit einer Störgrößenaufschaltung und einer adaptiven Nachfüh- rung der Reglerparameter zur Kompensation der nichtlinearen Streckendynamik.

Dies wurde mit Hilfe einer umfangreichen Prozess-EDV umgesetzt, in der die PAT-Software SIMATIC SIPAT für die Steuerung und Ausführung von Berechnungen sowohl der spektralen Regelgröße c_{S1Lram} mit der MVDA-Software SIMCA-Q als auch für die adaptive Steuerung der Zufüttertrajektorie F_{R1k} und des Reglerparameters T_{Ik} mit MATLAB® verantwortlich ist.

Sowohl im aperiodischen Grenzfall ohne Schwingungen als auch mit einer gewählten Perio- dendarauer T_p im Schwingfall war eine Regelung von c_{S1L} für einen Zeitraum von etwa drei Stun- den erfolgreich. Im weiteren Verlauf beider Experimente versagte die Glycerolmessung mit Raman-Spektroskopie, was auf eine zu starke Beeinflussung des Kalibriermodells durch die exponentiell ansteigende Zelldichte zurückzuführen ist.

Zum Abschluss dieser Arbeit wurde die Anwendung der MVDA auf nichtspektroskopische Pro- zessdaten untersucht. Ziel dabei war ebenfalls die Vorhersage nicht direkt messbarer Variab- len zur Erweiterung des Prozessmonitorings.

Hierbei kamen 15 klassische Inline- und Online-Messgrößen, wie beispielsweise Waagensig- nale oder das Ergebnis einer Abgasbilanzierung, anstelle der Spektren zum Einsatz. Multiva- riante Kalibriermodelle für neun verschiedene Zielgrößen wurden daraus mit den Verfahren PLSR und SVR erstellt.

Insbesondere mit der nichtlinearen SVR konnten sehr gute Ergebnisse erzielt werden. So lagen relative Vorhersagefehler $RMSE_{rel}$ für die Zelldichte c_{X_L} sowie die Konzentrationen von Glycerol c_{S1M} und des Zielproduktes c_{P1M} in der Medienphase bei etwa 3 %.

Auch die Vorhersage des Verlaufes zellspezifischer Reaktionsraten war auf diesem Wege mög- lich. Die geringste Abweichung zeigte die Methanolaufnahmerate $q_{S2/X}$ mit 2,7 %. Die Bil- dungsrate $q_{P1/X}$ des Zielproduktes konnte mit einem Fehler von 10,2 % immer noch zufrieden- stellend wiedergegeben werden.

12. Anhang

12.1 Kulturmedien

Als Vorkultur-, Bioreaktor- und Refreshmedium kamen die drei in Tabelle 12.1 gezeigten Varianten des Grundsatzmediums FM22 zum Einsatz. Das verwendete Medium sowie die zu Prozessbeginn eingesetzte Glycerolkonzentration c_{S1M0} sind vom Prozessschritt bzw. der Reaktoranlage abhängig und können Tabelle 12.2 entnommen werden.

Die Zugabe der Biotin- und PTM4-Stammlösungen erfolgt nach der Sterilisation des Mediums bei 121 °C für mindestens 20 min.

Tabelle 12.1: Varianten des Kulturmediums FM22, modifiziert nach [STRATTON et al., 1998]
(Grundsatzmedium ohne Kohlenstoffquelle)

Komponente	Konzentration [g l^{-1}]		
	Variante 1	Variante 2	Variante 3
KH_2PO_4	25,7	7,00	12,9
$(\text{NH}_4)_2\text{SO}_4$	5,00	5,00	2,50
K_2SO_4	8,60	2,00	4,30
$\text{CaSO}_4 \cdot 2 \text{H}_2\text{O}$	1,40	0,80	0,70
$\text{MgSO}_4 \cdot 7 \text{H}_2\text{O}$	16,4	8,00	8,20
$\text{Na}_3\text{-citrate} \cdot 2 \text{H}_2\text{O}$	5,90	5,00	2,90
Biotin-Stammlsg.	8,0 mll^{-1}	8,0 mll^{-1}	8,0 mll^{-1}
PTM4-Stammlsg.	4,0 mll^{-1}	4,0 mll^{-1}	4,0 mll^{-1}

Tabelle 12.2: Medienvarianten und Glycerolkonzentrationen c_{S1M0}

Prozessschritt	Medien-variante	Konzentration [g l^{-1}]
Vorkultur	1	30,0
Batch C10	2	54,0
Produktion C10	2	0
Startup ED10	3	47,6
Zyklus ED10	3	43,6
Zyklus C30	3	0

Die Rezepturen der Stammlösungen sind in Tabelle 12.3 gegeben. Nach dem Ansatz in der angegebenen Reihenfolge erfolgt eine Sterilfiltration mit 0,22 µm Porendurchmesser und eine Lagerung der Stammlösungen bei 4 °C.

Tabelle 12.4 enthält eine Auflistung eingesetzter Zusatzstoffe und Reservoir-Konzentrationen.

Tabelle 12.3: Spurenelement- und Vitaminstammlösungen

PTM4			Biotin		
Komponente	Konzentration		Komponente	Konzentration	
$\text{CuSO}_4 \cdot 5 \text{H}_2\text{O}$	2,0	g l^{-1}	$\text{C}_{10}\text{H}_{16}\text{N}_2\text{O}_3\text{S}$	0,20	g l^{-1}
NaI	$8,0 \cdot 10^{-2}$	g l^{-1}			
$\text{MnSO}_4 \cdot \text{H}_2\text{O}$	3,0	g l^{-1}			
$\text{Na}_2\text{MoO}_4 \cdot 2 \text{H}_2\text{O}$	0,20	g l^{-1}			
H_3BO_4	$2,0 \cdot 10^{-2}$	g l^{-1}			
CaSO_4	0,50	g l^{-1}			
$\text{CoCl}_2 \cdot 6 \text{H}_2\text{O}$	0,50	g l^{-1}			
$\text{ZnSO}_4 \cdot 7 \text{H}_2\text{O}$	7,0	g l^{-1}			
$\text{FeSO}_4 \cdot 7 \text{H}_2\text{O}$	22	g l^{-1}			
H_2SO_4 (96 %)	1,0	m l l^{-1}			

Tabelle 12.4: Eingesetzte Zusatzstoffe und Reservoirkonzentrationen

Reservoir	Symbol	Substanz	Konzentration
R1 (Glycerol)	c_{S1R1}	$\text{C}_3\text{H}_8\text{O}_3$	400 - 600 g l^{-1}
R2 (Methanol)	c_{S2R2}	CH_3OH	790 g l^{-1}
T1 (Säure)	-	H_3PO_4	2,0 M
T2 (Base)	-	NH_4OH	12,5 %
AF (Antischäum)	-	Struktol® J673	100 %

12.2 Kultivierungsbedingungen

Gewählte Kultivierungsbedingungen sind Tabelle 12.5 und Tabelle 12.6 zu entnehmen.

Tabelle 12.5: Typische Sollwerte geregelter Prozessgrößen im BIOSTAT® C10

Prozesssollwerte	Symbol		Batch	Fed-Batch	Induktion
Methanolkonzentration	c_{S2Mw}	$[\text{g l}^{-1}]$	0	0	1,0
Begasungsrate	F_{nGw}	$[\text{l min}^{-1}]$	12	12	12
Überdruck im Kopfraum	p_{Gw}	$[\text{mbar}]$	0	0	0
pH-Wert	pH_w	[-]	5,0	5,0	5,6
relativer Gelöstsauerstoffgehalt	pO_{2w}	[%]	25	25	25
Temperatur der Flüssigphase	ϑ_{Lw}	$[\text{°C}]$	30	30	22

Tabelle 12.6: Sollwerte in der zweistufigen Anlage (abweichend zum BIOSTAT® C10)

Prozesssollwerte	Symbol	Batch (ED10)	Induktion (ED10)	Induktion (C30)
Begasungsrate	F_{nGw}	[lmin ⁻¹]	10	10
Überdruck im Kopfraum	p_{Gw}	[mbar]	500	500

12.3 Reaktionskinetische Parameter

Reaktionskinetische Parameter des in dieser Arbeit verwendeten *Pichia pastoris* Stammes zur Herstellung von D1M1H sind Tabelle 12.7 zu entnehmen.

Tabelle 12.7: Zeitinvariante Reaktionsparameter für zellspezifische Reaktionsraten und Ausbeutekoeffizienten des D1M1H-Produktionsstammes nach [MARTENS, 2014]

Parameter	Substrat	
	Glycerol (i = 1)	Methanol (i = 2)
$q_{iX/X_{max}}$	[h ⁻¹]	0,240
q_{iX/X_m}	[h ⁻¹]	1,35·10 ⁻²
$q_{S_i/X_{max}}$	[h ⁻¹]	0,286
q_{S_i/X_m}	[h ⁻¹]	1,52·10 ⁻²
$y_{X/S_{i,max}}$	[\cdot]	0,840
$y_{X/S_{i,gr}}$	[\cdot]	0,887
$q_{iO/X_{max}}$	[h ⁻¹]	0,112
q_{iO/X_m}	[h ⁻¹]	9,60·10 ⁻³
$y_{iX/O_{max}}$	[\cdot]	2,14
$y_{iX/O_{gr}}$	[\cdot]	2,47

12.4 Offline-Messungen

Bestimmung des Gesamtproteingehalts (Bradford-Test)

Zur Bestimmung der Gesamtproteinkonzentration c_{PlotMoff} wurde ein Bradford-Test in Mikrotiterplatten durchgeführt. Lösungen:

A Standards: Quick Start™ Bovine γ -Globulin Standard Set (Bio-Rad)

B Reaktionslösung: Quick Start™ Bradford Dye Reagent (Bio-Rad)

Unter Verwendung von A wurde im Bereich von 12,5 bis 200 μgml^{-1} BGG in PBS eine Kalibriergerade erstellt. Es wurden 20 μl Leerprobe (PBS), Standard oder Probe (ggf. verdünnt in PBS) mit 300 μl B gemischt, wobei jeweils eine Dreifachbestimmung durchgeführt wurde.

Zellaufschluss

Zur Bestimmung des intrazellulären Gehaltes an Alkoholoxidase $g_{P2/X}$ wurde zuvor ein mechanischer Zellaufschluss unter Verwendung von Glas-Beads® (Carl Roth) mit 0,5 mm Durchmesser durchgeführt. Lösungen:

- A Aufschlusspuffer: 1 mM EDTA in PBS
- B Zellsuspension: 100 g⁻¹ Zellen in A
- C Aufschlussansatz: 1 g Glasbeads, 100 µl B in 900 µl A
- D Proteaseinhibitor: 1 mM PMSF in C, kurz vor Gebrauch ansetzen

Nach Ansetzen der jeweiligen Lösung A wurden die Zellpellets darin für 10 min bei 2.000 rpm auf einem Vibrax resuspendiert (Lösung B). Aus der Zellsuspension erfolgte eine BTM-Bestimmung sowie die Herstellung von Lösung C in Mikroreaktionsgefäßchen.

Unter Verwendung von Proteaseinhibitoren (D) wurden die Zellen für 20 min bei 2.000 rpm auf einem Vibrax aufgeschlossen. Nach Zentrifugation für 30 min bei 14.000 min⁻¹ und 4 °C konnte der Überstand für die jeweilige Analyse verwendet oder bei -20 °C gelagert werden.

AOX-Bestimmung

Nach einem mechanischen Zellaufschluss erfolgt die Bestimmung der Aktivität von Alkoholoxidase in einem enzymatischen Assay. Lösungen:

- A Kaliumphosphatpuffer: 100 mM KP_i
- B Wasserstoffperoxid: 0,003 % H₂O₂, kurz vor Gebrauch ansetzen
- C Peroxidase: 2.500 Uml⁻¹, aliquotiert bei -20 °C lagern
- D Reaktionslösung: 2,2 mM ABTS, 5 % B, 0,05 % C in A
kurz vor Gebrauch ansetzen
- E Substratlösung: 1 % Methanol in A

Es wurden jeweils 50 µl Probe (in geeigneter Verdünnung in A) in einer Mikrotiterplatte mit 200 µl D gemischt und 0,5 h bei 30 °C inkubiert. Die Reaktion wurde durch Zufügen von 50 µl E gestartet und die Absorption bei 405 nm in einem Mikroplattenphotometer (Sunrise, Tecan) über 4,5 min alle 30 s gemessen.

13. Quellenverzeichnis

- Abu-Absi**, N.R., Kenty, B.M., Cuellar, M.E., Borys, M.C., Sakhamuri, S., Strachan, D.J., Hausladen, M.C., Li, Z.J., **2011**. Real time monitoring of multiple parameters in mammalian cell culture bioreactors using an in-line Raman spectroscopy probe. *Biotechnol. Bioeng.* 108, 1215–1221. doi:10.1002/bit.23023
- Ahnert**, K., **Abel**, M., **2007**. Numerical differentiation of experimental data: Local versus global methods. *Comput. Phys. Commun.* 177, 764–774. doi:10.1016/j.cpc.2007.03.009
- Arnold**, S.A., Crowley, J., Vaidyanathan, S., Matheson, L., Mohan, P., Hall, J.W., Harvey, L.M., McNeil, B., **2000**. At-line monitoring of a submerged filamentous bacterial cultivation using near-infrared spectroscopy. *Enzyme Microb. Technol.* 27, 691–697. doi:10.1016/S0141-0229(00)00271-4
- Arvanitoyannis**, I.S., Katsota, M.N., Psarra, E.P., Soufleros, E.H., Kallithraka, S., **1999**. Application of quality control methods for assessing wine authenticity: Use of multivariate analysis (chemometrics). *Trends Food Sci. Technol.* 10, 321–336. doi:10.1016/S0924-2244(99)00053-9
- Berrueta**, L.A., Alonso-Salces, R.M., Héberger, K., **2007**. Supervised pattern recognition in food analysis. *J. Chromatogr. A, Data Anal. in Chrom.* 1158, 196–214. doi:10.1016/j.chroma.2007.05.024
- Berry**, B., Moretto, J., Matthews, T., Smelko, J., Wiltberger, K., **2015**. Cross-scale predictive modeling of CHO cell culture growth and metabolites using Raman spectroscopy and multivariate analysis. *Biotechnol. Prog.* 31, 566–577. doi:10.1002/btpr.2035
- Blanco**, M., **Villarroya**, I., **2002**. NIR spectroscopy: A rapid-response analytical tool. *TrAC Trends Anal. Chem.* 21, 240–250. doi:10.1016/S0165-9936(02)00404-1
- Böcker**, J., **1997**. Spektroskopie: Instrumentelle Analytik mit Atom- und Molekülspektrometrie, 1. Aufl. ed, LaborPraxis. Vogel, Würzburg.
- Bondi**, R.W., **Drennen**, J.K., **2011**. Quality by Design and the Importance of PAT in QbD, in: Ahuja, S., Scypinsky, S. (Eds.), *Handbook of Modern Pharmaceutical Analysis*. Academic Press. pp. 195–224.
- Borchert**, S.-O., **2015**. Entwicklung, Beobachtung und Steuerung integrierter, quasi-kontinuierlicher pharmazeutischer Produktionsprozesse mit Methoden der Multivariaten Datenverarbeitung. Dissertation, Leibniz Universität Hannover.
- Bradford**, M.M., **1976**. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72, 248–254. doi:10.1016/0003-2697(76)90527-3
- Brereton**, R.G., **2015**. Pattern recognition in chemometrics. *Chemom. Intell. Lab. Syst.* 149, Part B, 90–96. doi:10.1016/j.chemolab.2015.06.012
- Candolfi**, A., De Maesschalck, R., Jouan-Rimbaud, D., Hailey, P.A., Massart, D.L., **1999a**. The influence of data pre-processing in the pattern recognition of excipients near-infrared spectra. *J. Pharm. Biomed. Anal.* 21, 115–132. doi:10.1016/S0731-7085(99)00125-9
- Candolfi**, A., De Maesschalck, R., Massart, D.L., Hailey, P.A., Harrington, A.C.E., **1999b**. Identification of pharmaceutical excipients using NIR spectroscopy and SIMCA. *J. Pharm. Biomed. Anal.* 19, 923–935. doi:10.1016/S0731-7085(99)00125-9
- Chang**, C.-C., **Lin**, C.-J., **2011**. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol. TIST* 2, 1–27. doi:10.1145/1961189.1961199
- Chen**, Q., Zhao, J., Zhang, H., Wang, X., **2006**. Feasibility study on qualitative and quantitative analysis in tea by near infrared spectroscopy with multivariate calibration. *Anal. Chim. Acta* 572, 77–84. doi:10.1016/j.aca.2006.05.007
- Childs**, R.E., **Bardsley**, W.G., **1975**. The steady-state kinetics of peroxidase with 2, 2'-azino-di-(3-ethyl-benzthiazoline-6-sulphonic acid) as chromogen. *Biochem. J.* 145, 93–103. doi:10.1042/bj1450093

- Chong, I.-G., Jun, C.-H., 2005.** Performance of some variable selection methods when multicollinearity is present. *Chemom. Intell. Lab. Syst.* 78, 103–112. doi:10.1016/j.chemolab.2004.12.011
- Cocchi, M., Corbellini, M., Foca, G., Lucisano, M., Pagani, M.A., Tassi, L., Ulrici, A., 2005.** Classification of bread wheat flours in different quality categories by a wavelet-based feature selection/classification algorithm on NIR spectra. *Anal. Chim. Acta* 544, 100–107. doi:10.1016/j.aca.2005.02.075
- Cornelissen, G., 2004.** Integrierte Bioprozessentwicklung zur Herstellung pharmakologisch wirksamer Proteine mit *Pichia pastoris*. Dissertation, Leibniz Universität Hannover.
- Craven, S., Whelan, J., Glennon, B., 2014.** Glucose concentration control of a fed-batch mammalian cell bioprocess using a nonlinear model predictive controller. *J. Process Control* 24, 344–357. doi:10.1016/j.jprocont.2014.02.007
- Cregg, J., Cereghino, J., Shi, J., Higgins, D., 2000.** Recombinant protein expression in *Pichia pastoris*. *Mol. Biotechnol.* 16, 23–52. doi:10.1385/MB:16:1:23
- Crowley, J., Arnold, S.A., Wood, N., Harvey, L.M., McNeil, B., 2005.** Monitoring a high cell density recombinant *Pichia pastoris* fed-batch bioprocess using transmission and reflectance near infrared spectroscopy. *Enzyme Microb. Technol.* 36, 621–628. doi:10.1016/j.enzmictec.2003.12.016
- Desai, K., Badhe, Y., Tambe, S.S., Kulkarni, B.D., 2006.** Soft-sensor development for fed-batch bioreactors using support vector regression. *Biochem. Eng. J.* 27, 225–239. doi:10.1016/j.bej.2005.08.002
- Ellert, A., 2015.** Automatisierung und Optimierung von Herstellungsprozessen rekombinanter pharmazeutischer Proteine. Dissertation, Leibniz Universität Hannover.
- Ellis, S.B., Brust, P.F., Koutz, P.J., Waters, A.F., Harpold, M.M., Gingeras, T.R., 1985.** Isolation of alcohol oxidase and two other methanol regulatable genes from the yeast *Pichia pastoris*. *Mol. Cell. Biol.* 5, 1111–1121. doi:10.1128/MCB.5.5.1111
- EMA, 2015.** First malaria vaccine receives positive scientific opinion from EMA. URL: http://www.ema.europa.eu/docs/en_GB/document_library/Press_release/2015/07/WC550019044.pdf (abgerufen am 24.11.2016)
- Eriksson, L., Johansson, E., Kettaneh-Wold, N., Wold, S., 2001.** Multi- and megavariate data analysis: Principles and applications. MKS Umetrics, Umeå, Schweden.
- Faber, B.W., Remarque, E.J., Morgan, W.D., Kocken, C.H.M., Holder, A.A., Thomas, A.W., 2007.** Malaria Vaccine-Related Benefits of a Single Protein Comprising *Plasmodium falciparum* Apical Membrane Antigen 1 Domains I and II Fused to a Modified Form of the 19-Kilodalton C-Terminal Fragment of Merozoite Surface Protein 1. *Infect. Immun.* 75, 5947–5955. doi:10.1128/IAI.01804-06
- Faber, B.W., Younis, S., Remarque, E.J., Rodriguez Garcia, R., Riasat, V., Walraven, V., van der Werff, N., van der Eijk, M., Cavanagh, D.R., et al., 2013.** Diversity Covering AMA1-MSP_{1¹⁹} Fusion Proteins as Malaria Vaccines. *Infect. Immun.* 81, 1479–1490. doi:10.1128/IAI.01267-12
- Farrés, M., Platikanov, S., Tsakovski, S., Tauler, R., 2015.** Comparison of the variable importance in projection (VIP) and of the selectivity ratio (SR) methods for variable selection and interpretation. *J. Chemom.* 29, 528–536. doi:10.1002/cem.2736
- Fawcett, T., 2004.** ROC graphs: Notes and practical considerations for researchers. HP Laboratories, Kluwer Academic Publishers, Dordrecht.
- FDA, 2004a.** PAT - A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance. URL: <http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidancengu/ucm070305.pdf> (abgerufen am 23.03.2016)
- FDA, 2004b.** Pharmaceutical CGMPs for the 21st Century - A Risk-Based Approach. URL: <https://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/Manufacturing/QuestionsandAnswersonCurrentGoodManufacturingPracticesCGMPforDrugs/UCM176374.pdf> (abgerufen am 23.03.2016)

- Fernández-Ibañez, V., Soldado, A., Martínez-Fernández, A., de la Roza-Delgado, B., 2009.** Application of near infrared spectroscopy for rapid detection of Aflatoxin B1 in maize and barley as analytical quality assessment. *Food Chem.* 113, 629–634.
doi:10.1016/j.foodchem.2008.07.049
- Fisher, R.A., 1936.** The Use of Multiple Measurements in Taxonomic Problems. *Ann. Eugen.* 7, 179–188. doi:10.1111/j.1469-1809.1936.tb02137.x
- Fricke, J., 2014.** Optimierung eines Herstellungsprozesses artifizieller Malaria-Vakzine mittels Design of Experiments. Dissertation, *Leibniz Universität Hannover*.
- Gemperline, P.J., Webber, L.D., Cox, F.O., 1989.** Raw materials testing using soft independent modeling of class analogy analysis of near-infrared reflectance spectra. *Anal. Chem.* 138–144. doi:10.1021/ac00177a012
- Glassey, J., Gernaey, K.V., Clemens, C., Schulz, T.W., Oliveira, R., Striedner, G., Mandenius, C.-F., 2011.** Process analytical technology (PAT) for biopharmaceuticals. *Biotechnol. J.* 6, 369–377. doi:10.1002/biot.201000356
- Hall, J.W., McNeil, B., Rollins, M.J., Draper, I., Thompson, B.G., Macaloney, G., 1996.** Near-Infrared Spectroscopic Determination of Acetate, Ammonium, Biomass, and Glycerol in an Industrial *Escherichia coli* Fermentation. *Appl. Spectrosc.* 50, 102–108.
doi:10.1366/0003702963906726
- Healer, J., Murphy, V., Hodder, A.N., Masciantonio, R., Gemmill, A.W., Anders, R.F., Cowman, A.F., Batchelor, A., 2004.** Allelic polymorphisms in apical membrane antigen-1 are responsible for evasion of antibody-mediated inhibition in *Plasmodium falciparum*. *Mol. Microbiol.* 52, 159–168. doi:10.1111/j.1365-2958.2003.03974.x
- Henriques, J.G., Buziol, S., Stocker, E., Voogd, A., Menezes, J.C., 2009.** Monitoring Mammalian Cell Cultivations for Monoclonal Antibody Production Using Near-Infrared Spectroscopy, in: Rao, G. (Ed.), Optical Sensor Systems in Biotechnology, Advances in Biochemical Engineering / Biotechnology. Springer, Berlin, Heidelberg, pp. 29–72.
doi:10.1007/10_2009_11
- Higgins, D.R., Cregg, J.M., 1998.** Pichia protocols. Methods in Molecular Biology. Humana Press, Totowa.
- Hoehse, M., Alves-Rausch, J., Prediger, A., Roch, P., Grimm, C., 2015.** Near-infrared spectroscopy in upstream bioprocesses. *Pharm. Bioprocess.* 3, 153–172.
doi:10.4155/pbp.15.1
- Hotelling, H., 1951.** A Generalized t-Test and Measure of Multivariate Dispersion. Second Berkeley Symposium. 23–41.
- Huang, H., Yu, H., Xu, H., Ying, Y., 2008.** Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages: A review. *J. Food Eng.* 87, 303–313.
doi:10.1016/j.jfoodeng.2007.12.022
- ICH, 2009.** Pharmaceutical Development Q8 (R2). URL: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q8_R1/Step4/Q8_R2_Guideline.pdf (abgerufen am 15.12.2016)
- ICH, 2008.** Pharmaceutical Quality System Q10. URL: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q10/Step4/Q10_Guideline.pdf (abgerufen am 15.12.2016)
- ICH, 2005.** Quality Risk Management Q9. URL: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q9/Step4/Q9_Guideline.pdf (abgerufen am 15.12.2016)
- Jain, S., 2014.** Quality by design (QBD): A comprehensive understanding of implementation and challenges in pharmaceuticals development. *Int. J. Pharm. Pharm. Sci.* 6, 29–35.
- Janssen, F.W., Ruelius, H.W., 1968.** Alcohol oxidase, a flavoprotein from several basidiomycetes species. *Biochim. Biophys. Acta BBA - Enzymol.* 151, 330–342.
doi:10.1016/0005-2744(68)90100-9

- Jose**, G.E., Folque, F., Menezes, J.C., Werz, S., Strauss, U., Hakemeyer, C., **2011**. Predicting mab product yields from cultivation media components, using near-infrared and 2D-fluorescence spectroscopies. *Biotechnol. Prog.* 27, 1339–1346. doi:10.1002/btpr.638
- Junker**, B. h., **Wang**, H. y., **2006**. Bioprocess monitoring and computer control: Key roots of the current PAT initiative. *Biotechnol. Bioeng.* 95, 226–261. doi:10.1002/bit.21087
- Kaiser**, C., **2010**. Methoden zur Prozessüberwachung und Modellierung von Herstellungsprozessen rekombinanter Proteine mit *Escherichia coli*. Dissertation, Leibniz Universität Hannover.
- Kasprow**, R.P., Lange, A.J., Kirwan, D.J., **1998**. Correlation of Fermentation Yield with Yeast Extract Composition as Characterized by Near-Infrared Spectroscopy. *Biotechnol. Prog.* 14, 318–325. doi:10.1021/bp980001j
- Kecman**, V., **2005**. Support Vector Machines - An Introduction, in: Wang, L. (Ed.), *Support Vector Machines: Theory and Applications*. Springer, Berlin, Heidelberg. pp. 1–47. doi:10.1007/10984697_1
- Keesey**, J., **1987**. Biochemica information. *Boehringer Mannh. Biochem. Indianap.* 56–59.
- Keil**, K.M., **Tilkins**, M.L., **2013**. Screening yeastolate raw material used in insect cell culture media. *Bioprocess Int.* 11, 44–51.
- Kennard**, R.W., **Stone**, L.A., **1969**. Computer Aided Design of Experiments. *Technometrics* 11, 137. doi:10.2307/1266770
- Kennedy**, M.C., Wang, J., Zhang, Y., Miles, A.P., Chitsaz, F., Saul, A., Long, C.A., Miller, L.H., Stowers, A.W., **2002**. In Vitro Studies with Recombinant *Plasmodium falciparum* Apical Membrane Antigen 1 (AMA1): Production and Activity of an AMA1 Vaccine and Generation of a Multiallelic Response. *Infect. Immun.* 70, 6948–6960. doi:10.1128/IAI.70.12.6948-6960.2002
- Kessler**, W., **2008**. Multivariate Datenanalyse für die Pharma-, Bio- und Prozessanalytik: Ein Lehrbuch, 1. Aufl., 1. Nachdr. ed. Wiley-VCH, Weinheim.
- Kirdar**, A.O., Chen, G., Weidner, J., Rathore, A.S., **2009**. Application of near-infrared (NIR) spectroscopy for screening of raw materials used in the cell culture medium for the production of a recombinant therapeutic protein. *Biotechnol. Prog.* 26, 527–531. doi:10.1002/btpr.329
- Kocken**, C.H.M., Withers-Martinez, C., Dubbeld, M.A., Wel, A. van der, Hackett, F., Blackman, M.J., Thomas, A.W., **2002**. High-Level Expression of the Malaria Blood-Stage Vaccine Candidate *Plasmodium falciparum* Apical Membrane Antigen 1 and Induction of Antibodies That Inhibit Erythrocyte Invasion. *Infect. Immun.* 70, 4471–4476. doi:10.1128/IAI.70.8.4471-4476.2002
- Koutz**, P., Davis, G.R., Stillman, C., Barringer, K., Cregg, J., Thill, G., **1989**. Structural comparison of the *Pichia pastoris* alcohol oxidase genes. *Yeast* 5, 167–177. doi:10.1002/yea.320050306
- Lakowicz**, J.R., **2006**. Principles of fluorescence spectroscopy, 3rd ed., Springer, New York.
- Le**, H., Kabbur, S., Pollastrini, L., Sun, Z., Mills, K., Johnson, K., Karypis, G., Hu, W.-S., **2012**. Multivariate analysis of cell culture bioprocess data – Lactate consumption as process indicator. *J. Biotechnol.* 162, 210–223. doi:10.1016/j.biote.2012.08.021
- Li**, Y., **Yuan**, J., **2006**. Prediction of Key State Variables using Support Vector Machines in Bioprocesses. *Chem. Eng. Technol.* 29, 313–319. doi:10.1002/ceat.200500182
- Lindemann**, C., Marose, S., Nielsen, H.O., Schepet, T., **1998**. 2-Dimensional fluorescence spectroscopy for on-line bioprocess monitoring. *Sens. Actuators B Chem.* 51, 273–277. doi:10.1016/S0925-4005(98)00195-6
- Lögering**, K., **2015**. Automatisierung, Optimierung und Analyse integrierter Produktionsprozesse zur Herstellung rekombinanter Proteine mit *Pichia pastoris*. Dissertation, Leibniz Universität Hannover.

- Lögering**, K., Mueller, C., Voss, J.-P., Wagenfuehrer, C., Zahn, D., Bertelsen, H.-P., Scheffler, U., Luttmann, R., **2011**. An integrated scale-down plant for optimal recombinant enzyme production by *Pichia pastoris*. *Biotechnol. J.* 6, 428–436. doi:10.1002/biot.201000384
- Lottspeich**, F., **Engels**, J.W. (Eds.), **2012**. Bioanalytik, 3., [aktualisierte und erw.] Aufl. ed. Springer Spektrum, Berlin.
- Lourenço**, N.D., Lopes, J.A., Almeida, C.F., Saraguça, M.C., Pinheiro, H.M., **2012**. Bioreactor monitoring with spectroscopy and chemometrics: A review. *Anal. Bioanal. Chem.* 404, 1211–1237. doi:10.1007/s00216-012-6073-9
- Luchner**, M., Striedner, G., Cserjan-Puschmann, M., Strobl, F., Bayer, K., **2015**. Online prediction of product titer and solubility of recombinant proteins in *Escherichia coli* fed-batch cultivations. *J. Chem. Technol. Biotechnol.* 90, 283–290. doi:10.1002/jctb.4463
- Luttmann**, R., Borchert, S.-O., Mueller, C., Lögering, K., Aupert, F., Weyand, S., Kober, C., Faber, B., Cornelissen, G., **2015**. Sequential/parallel production of potential Malaria vaccines – A direct way from single batch to quasi-continuous integrated production. *J. Biotechnol.* 213, 83–96. doi:10.1016/j.jbiotec.2015.02.022
- Märk**, J., Andre, M., Karner, M., Huck, C.W., **2010**. Prospects for multivariate classification of a pharmaceutical intermediate with near-infrared spectroscopy as a process analytical technology (PAT) production control supplement. *Eur. J. Pharm. Biopharm.* 76, 320–327. doi:10.1016/j.ejpb.2010.06.015
- Martens**, S., **2014**. Analyse und Synthese von zyklischen Herstellungsprozessen artifizieller Diversity Covering (DiCo) Malariavakzinkandidaten. Dissertation, Leibniz Universität Hannover.
- Martin**, K.A., **1992**. Recent Advances in Near-Infrared Reflectance Spectroscopy. *Appl. Spectrosc. Rev.* 27, 325–383. doi:10.1080/05704929208018109
- Mercier**, S.M., Diepenbroek, B., Wijffels, R.H., Streefzahl, M., **2014**. Multivariate PAT solutions for biopharmaceutical cultivation: Current progress and limitations. *Trends Biotechnol.* 32, 329–336. doi:10.1016/j.tibtech.2014.03.008
- Mitchell**, G.H., Thomas, A.W., Margos, G., Dluzewski, A.R., Bannister, L.H., **2004**. Apical Membrane Antigen 1, a Major Malaria Vaccine Candidate, Mediates the Close Attachment of Invasive Merozoites to Host Red Blood Cells. *Infect. Immun.* 72, 154–158. doi:10.1128/IAI.72.1.154–158.2004
- Mosser**, M., Chevalot, I., Olmos, E., Blanchard, F., Kapel, R., Oriol, E., Marc, I., Marc, A., **2013**. Combination of yeast hydrolysates to improve CHO cell growth and IgG production. *Cytotechnology* 65, 629–641. doi:10.1007/s10616-012-9519-1
- Naes**, T. (Ed.), **2004**. A user-friendly guide to multivariate calibration and classification. NIR Publications, Chichester.
- Nelder**, J.A., **Mead**, R., **1965**. A simplex method for function minimization. *Comput. J.* 7, 308–313. doi:10.1093/comjnl/7.4.308
- Nikodem**, D.-P., **Davidson**, E.-A., **2000**. Identification of a novel antigenic domain of *Plasmodium falciparum* merozoite surface protein-1 that specifically binds to human erythrocytes and inhibits parasite invasion, *in vitro*. *Mol. Biochem. Parasitol.* 108, 79–91. doi:10.1016/S0166-6851(00)00206-1
- Paul**, A., Carl, P., Westad, F., Voss, J.-P., Maiwald, M., **2016**. Towards Process Spectroscopy in Complex Fermentation Samples and Mixtures. *Chem. Ing. Tech.* 88, 756–763. doi:10.1002/cite.201500118
- Perkampus**, H.-H., **1993**. Lexikon Spektroskopie, Parat. VCH, Weinheim.
- Peterson**, W., Birdsall, T., Fox, W., **1954**. The theory of signal detectability. *Trans. IRE Prof. Group Inf. Theory* 4, 171–212. doi:10.1109/TIT.1954.1057460
- Peuker**, T., **2006**. Modellgestützte Verfahren zur Beobachtung und Steuerung bioverfahrenstechnischer Prozesse. Dissertation, Leibniz Universität Hannover.

- Platt, J.C., 1999.** Fast Training of Support Vector Machines Using Sequential Minimal Optimization, in: Schölkopf, B., Burges, C.J.C., Smola, A.J. (Eds.), *Advances in Kernel Methods - Support Vector Learning*. MIT Press, Cambridge, USA. pp. 41–64.
- Pomerantsev, A.L., Rodionova, O.Y., 2012.** Process analytical technology: A critical view of the chemometriicians. *J. Chemom.* 26, 299–310. doi:10.1002/cem.2445
- Rathore, A.S., 2014.** QbD/PAT for bioprocessing: Moving from theory to implementation. *Curr. Opin. Chem. Eng.* 6, 1–8. doi:10.1016/j.coche.2014.05.006
- Rathore, A.S., Bhushan, N., Hadpe, S., 2011.** Chemometrics applications in biotech processes: A review. *Biotechnol. Prog.* 27, 307–315. doi:10.1002/btpr.561
- Rathore, A.S., Winkle, H., 2009.** Quality by design for biopharmaceuticals. *Nat. Biotechnol.* 27, 26–34. doi:10.1038/nbt0109-26
- Remarque, E.J., Faber, B.W., Kocken, C.H.M., Thomas, A.W., 2008.** A Diversity-Covering Approach to Immunization with *Plasmodium falciparum* Apical Membrane Antigen 1 Induces Broader Allelic Recognition and Growth Inhibition Responses in Rabbits. *Infect. Immun.* 76, 2660–2670. doi:10.1128/IAI.00170-08
- Rhee, J.I., Kang, T.-H., 2007.** On-line process monitoring and chemometric modeling with 2D fluorescence spectra obtained in recombinant *E. coli* fermentations. *Process Biochem.* 42, 1124–1134. doi:10.1016/j.procbio.2007.05.007
- Riley, M.R., Rhiel, M., Zhou, X., Arnold, M.A., Murhammer, D.W., et al., 1997.** Simultaneous measurement of glucose and glutamine in insect cell culture media by near infrared spectroscopy. *Biotechnol. Bioeng.* 55, 11–15. doi:10.1002/(SICI)1097-0290(19970705)55:1<11::AID-BIT2>3.0.CO;2-#
- Romanos, M., 1995.** Advances in the use of *Pichia pastoris* for high-level gene expression. *Curr. Opin. Biotechnol.* 6, 527–533. doi:10.1016/0958-1669(95)80087-5
- Ross, S.M., 2006.** Statistik für Ingenieure und Naturwissenschaftler, 3. Aufl. ed. Elsevier, Spektrum Akad. Verl., München.
- Savitzky, A., Golay, M.J., 1964.** Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* 36, 1627–1639. doi:10.1021/ac60214a047
- Siesler, H.W. (Ed.), 2002.** Near-infrared spectroscopy: Principles, instruments, applications. Wiley-VCH, Weinheim.
- Simon, L.L., Pataki, H., Marosi, G., Meemken, F., Hungerbühler, K., Baiker, A., Tummala, S., Glennon, B., Kuentz, M., et al., 2015.** Assessment of Recent Process Analytical Technology (PAT) Trends: A Multiauthor Review. *Org. Process Res. Dev.* 19, 3–62. doi:10.1021/op500261y
- Skibsted, E., Engelsen, S.B., 2010.** Spectroscopy for Process Analytical Technology (PAT), in: Lindon, J.C., Tranter, G.E., Koppenaal, D.W. (Eds.), *Encyclopedia of Spectroscopy and Spectrometry* (Third Edition). Elsevier, pp. 2651–2661. doi:10.1016/b978-0-12-374413-5.00026-9
- Skrabal, P.M., 2009.** Spektroskopie: Eine methodenübergreifende Darstellung vom UV- bis zum NMR-Bereich, UTB Chemie, vdf Hochschulverlag, Zürich.
- Small, G.W., 2006.** Chemometrics and near-infrared spectroscopy: Avoiding the pitfalls. *TrAC Trends Anal. Chem.*, Use and abuse of chemometrics 25, 1057–1066. doi:10.1016/j.trac.2006.09.004
- Stratton, J., Chiruvolu, V., Meagher, M., 1998.** High Cell-Density Fermentation, in: Higgins, D.R., Cregg, J.M. (Eds.), *Pichia Protocols, Methods in Molecular Biology*. Humana Press, Totowa, USA. pp. 107–120.
- Streefzahl, M., Martens, D.E., Beuvery, E.C., Wijffels, R.H., 2013.** Process Analytical Technology (PAT) tools for the cultivation step in biopharmaceutical production. *Eng. Life Sci.* 13, 212–223. doi:10.1002/elsc.201200025

- Surribas**, A., Geissler, D., Gierse, A., Scheper, T., Hitzmann, B., Montesinos, J.L., Valero, F., **2006**. State variables monitoring by in situ multi-wavelength fluorescence spectroscopy in heterologous protein production by *Pichia pastoris*. *J. Biotechnol.* 124, 412–419. doi:10.1016/j.biotec.2006.01.002
- Tschopp**, J.F., Brust, P.F., Cregg, J.M., Stillman, C.A., Gingeras, T.R., **1987**. Expression of the lacZ gene from two methanol-regulated promoters in *Pichia pastoris*. *Nucleic Acids Res.* 15, 3859–3876. doi:10.1093/nar/15.9.3859
- Valero**, F., **2013**. Bioprocess Engineering of *Pichia pastoris*, an Exciting Host Eukaryotic Cell Expression System, in: Ogawa, T. (Ed.), Protein Engineering - Technology and Application. *InTech*. doi:10.5772/56407
- Vandenabeele**, P., **2013**. Practical Raman spectroscopy: An introduction. *Wiley*, Chichester, UK. doi:10.1002/9781119961284
- Vapnik**, V., **Chervonenkis**, A., **1974**. Theory of Pattern Recognition. *Nauka*, Moskau, USSR.
- Vapnik**, V., **2000**. The Nature of Statistical Learning Theory. *Springer*, New York, USA. doi:10.1007/978-1-4757-3264-1
- Wang**, J., Yu, T., Jin, C., **2006**. On-line Estimation of Biomass in Fermentation Process Using Support Vector Machine. *Chin. J. Chem. Eng.* 14, 383–388. doi:10.1016/S1004-9541(06)60087-6
- Whelan**, J., Craven, S., Glennon, B., **2012**. In situ Raman spectroscopy for simultaneous monitoring of multiple process parameters in mammalian cell culture bioreactors. *Biotechnol. Prog.* 28, 1355–1362. doi:10.1002/btpr.1590
- WHO**, **2015**. World Malaria Report 2015. URL: http://apps.who.int/iris/bitstream/10665/200018/1/9789241565158_eng.pdf (abgerufen am 24.11.2016)
- Wilmes**, J., **2012**. Establishment and optimization of a Varioskan® Flash Multimode Reader as screening tool for bacterial and yeast prolongation. Master thesis, unveröffentlicht, Hochschule für Angewandte Wissenschaften Hamburg.
- Wold**, H., **1975**. Path Models with latent variables: The NIPALS approach. *Academic Press*, 307–357.
- Wold**, H., **1966**. Nonlinear Estimation by Iterative Least Squares Procedures, in: David, F.N. (Ed.) Research Papers in Statistics, *Wiley*, New York. pp. 411–444.
- Wold**, S., **1976**. Pattern recognition by means of disjoint principal components models. *Pattern Recognit.* 8, 127–139. doi:10.1016/0031-3203(76)90014-5
- Wold**, S., Esbensen, K., Geladi, P., **1987**. Principal component analysis. *Chemom. Intell. Lab. Syst.* 2, 37–52. doi:10.1016/0169-7439(87)80084-9
- Wold**, S., Johansson, E., Coch, M., **1993**. PLS - Partial Least Squares Projection to Latent Structures. In: Kubinyi, H. (Ed.): 3D QSAR in Drug Design: Volume 1: Theory Methods and Applications. *Kluwer Academic Publishers*, Dordrecht. pp. 523–550.
- Wold**, S., **Sjöström**, M., **1977**. SIMCA: A method for analyzing chemical data in terms of similarity and analogy, in: Kowalski, B.R. (Ed.) Chemometrics: Theory and Application. ACS Symposium Series, *American Chemical Society*. pp. 243–282. doi:10.1021/bk-1977-0052.ch012
- Woodcock**, T., O'Donnell, C., Downey, G., **2008**. Review: Better quality food and beverages: the role of near infrared spectroscopy. *J. Infrared Spectrosc.* 16, 1–29. doi:10.1255/jnirs.758
- Yu**, L.X., Amidon, G., Khan, M.A., Hoag, S.W., Polli, J., Raju, G.K., Woodcock, J., **2014**. Understanding Pharmaceutical Quality by Design. *AAPS J.* 16, 771–783. doi:10.1208/s12248-014-9598-3

Veröffentlichungen des Autors

Die nachfolgende Liste enthält die Veröffentlichungen des Autors, die während seines Studiums und der nachfolgenden experimentellen Forschung zur Doktorarbeit entstanden sind. In einigen sind Teile dieser Arbeit vorab veröffentlicht worden.

Zum aktuellen Zeitpunkt befindet sich eine Publikation mit dem Titel

"Advanced monitoring and control of pharmaceutical production processes with *Pichia pastoris* by using Raman spectroscopy and multivariate calibration methods"

im Begutachtungsverfahren beim Journal *Engineering in Life Sciences*.

Publikationen

Lögering, K., Mueller, C., Voss, J.-P., Wagenfuehrer, C., Zahn, D., Bertelsen, H.-P., Scheffler, U., Luttmann, R., **2011**. An integrated scale-down plant for optimal recombinant enzyme production by *Pichia pastoris*. *Biotechnol. J.* 6, 428–436.
doi:10.1002/biot.201000384

Paul, A., Carl, P., Westad, F., Voss, J.-P., Maiwald, M., **2016**. Towards Process Spectroscopy in Complex Fermentation Samples and Mixtures. *Chem. Ing. Tech.* 88, 756–763.
doi:10.1002/cite.201500118

Tagungsbeiträge

Lögering, K., Müller, C., Voß, J.-P., Fricke, J., Luttmann, R., **2010**: Automation and Optimization of Recombinant Enzyme Production in an Integrated Bioprocess with *Pichia pastoris*. Vortrag auf der *BioCat2010 – 5th International Congress on Biocatalysis*, 29.08. – 02.09.2010, Hamburg.

Lögering, K., Müller, C., Voß, J.-P., Fricke, J., Luttmann, R., **2010**: Development of a Self Optimizing Production Process of Recombinant Enzymes with *Pichia pastoris* in an Integrated Scale-down Production Plant. Vortrag auf der *ESBES 2010 – 8th European Symposium on Biochemical Engineering Science*, 05. – 08.09.2010, Bologna, Italien.

Martens, S., Borchert, S.-O., Paul, J., Voss, J.-P., Faber, B.W., Luttmann, R., Cornelissen, G., **2012**: Analysis and Optimization of a Sequential Malaria Vaccine Production Process with In-Situ Product Removal, Poster auf der *Vaccine Technology IV*, 20. – 25.05.2012, Albufeira, Portugal.

Martens, S., Voss, J.-P., Thiesing, D., Faber, B. W., Cornelissen, G., **2012**: Investigation in Expression Dynamics during Sequential Recombinant Production with *Pichia pastoris*. Vortrag auf der *ESBES/ISPPP*, 23. - 26.09.2012, Istanbul, Türkei.

Voss, J.-P., Martens, S., Thiesing, D., Mittelheuser, N.E., Cornelissen, G., Faber, B.W., Luttmann, R., **2014**: Protein Expression Dynamics and External Cell Interactions during Recursive Changes between Glycerol and Methanol in Long Term Recombinant Protein Production Runs with *Pichia pastoris*. Vortrag auf der *Pichia Protein Expression Conference 2014*, 02. - 05.03.2014, San Diego, USA.

Voss, J.-P., Mittelheuser, N., Maiwald, M., Luttmann, R., **2014**: Off-line Analysis and On-line Monitoring of Biopharmaceutical Production Processes with Spectroscopic Methods and Multivariate Data Analysis. Poster auf dem *Sartorius PAT & QbD Forum 2014*, 17. - 19.02.2014, Göttingen.

Voss, J.-P., Mittelheuser, N.E., Maiwald, M., Luttmann, R., **2014**: Off-line Analysis and On-line Monitoring of Biopharmaceutical Production Processes With Spectroscopic Methods and Multivariate Data Analysis. Vortrag auf der *EuroPACT 2014*, 06. - 09.05.2014, Barcelona, Spanien.

- Luttmann**, R., Borchert, S.-O., Voss, J.-P., Kober, C., Weyand, S., Mittelheuser, N., **2014:** From PAT via QbD and Golden Batch to Quality Control – Applications of MVDA in Pharmaceutical Production. Vortrag auf der *ESBES 2014 - 10th European Symposium on Biochemical Engineering Sciences*, 07. – 10.09.2014, Frankreich.
- Voss**, J.-P., Mittelheuser, N., Lemke, R., Martens, S., Luttmann, R., **2015:** On-line Monitoring and Control of a Production Process with *Pichia pastoris* Based on Spectroscopic Measurements and MVDA. Poster auf der *DECHEMA working party on Biochemical Engineering*, 13. – 15.05.2015, Hamburg.
- Luttmann**, R., Borchert, S.O., Voss, J.-P., Cornelissen, G., **2015:** Monitoring and control of reproducibility in quasi-continuous integrated production processes of Active Pharmaceutical Ingredients. Vortrag auf der ECI-Konferenz *Integrated Continuous Biomanufacturing II*, 01. – 05.11.2015, Berkeley, USA.

Lebenslauf des Autors

Persönliche Angaben

Name, Vorname Voß, Jan-Patrick

Geburtsdatum 26. April 1984

Geburtsort Hamburg

Wissenschaftlicher Werdegang

11/2012 - 07/2017 **Promotion**

Forschungs- und Transferzentrum für Bioprozess- und Analysentechnik der *Hochschule für Angewandte Wissenschaften Hamburg*
in Kooperation mit dem Institut für Technische Chemie der
Gottfried Wilhelm Leibniz Universität Hannover
Mündliche Promotionsprüfung am 31.07.2017 (Note 1,0)

03/2010 - 10/2012 **Masterstudiengang Biotechnology**

Hochschule für Angewandte Wissenschaften Hamburg
Abschluss als M.Sc. in Biotechnology am 19.10.2012 (Note 1,2)

03/2005 - 04/2010 **Studium der Biotechnologie**

Hochschule für Angewandte Wissenschaften Hamburg
Abschluss als Dipl.-Ing. (FH) am 28.04.2010 (Note 1,7)

Schulbildung

08/1994 - 06/2003 Abitur (Note: 2,4)

Gymnasium Lohbrügge, Hamburg

Berufliche Tätigkeiten

11/2012 - 07/2016 **Wissenschaftlicher Mitarbeiter**

Hochschule für Angewandte Wissenschaften Hamburg

09/2012 - 07/2016 **Lehrbeauftragter** (Umfang ca. 4 Wochenstunden)

Hochschule für Angewandte Wissenschaften Hamburg

Online-Buchshop für Ingenieure

■■■ VDI nachrichten

Online-Shops



Fachliteratur und mehr -
jetzt bequem online recher-
chieren & bestellen unter:
www.vdi-nachrichten.com/
Der-Shop-im-Ueberblick



Täglich aktualisiert:
Neuerscheinungen
VDI-Schriftenreihen



BUCHSHOP

Im Buchshop von vdi-nachrichten.com finden Ingenieure und Techniker ein speziell auf sie zugeschnittenes, umfassendes Literaturangebot.

Mit der komfortablen Schnellsuche werden Sie in den VDI-Schriftenreihen und im Verzeichnis lieferbarer Bücher unter 1.000.000 Titeln garantiert fündig.

Im Buchshop stehen für Sie bereit:

VDI-Berichte und die Reihe **Kunststofftechnik**:

Berichte nationaler und internationaler technischer Fachtagungen der VDI-Fachgliederungen

Fortschritt-Berichte VDI:

Dissertationen, Habilitationen und Forschungsberichte aus sämtlichen ingenieurwissenschaftlichen Fachrich- tungen

Newsletter „Neuerscheinungen“:

Kostenfreie Infos zu aktuellen Titeln der VDI-Schriften- reihen bequem per E-Mail

Autoren-Service:

Umfassende Betreuung bei der Veröffentlichung Ihrer Arbeit in der Reihe Fortschritt-Berichte VDI

Buch- und Medien-Service:

Beschaffung aller am Markt verfügbaren Zeitschriften, Zeitungen, Fortsetzungsreihen, Handbücher, Technische Regelwerke, elektronische Medien und vieles mehr – einzeln oder im Abo und mit weltweitem Lieferservice

VDI nachrichten

BUCHSHOP www.vdi-nachrichten.com/Der-Shop-im-Ueberblick

Die Reihen der Fortschritt-Berichte VDI:

- 1 Konstruktionstechnik/Maschinenelemente
- 2 Fertigungstechnik
- 3 Verfahrenstechnik
- 4 Bauingenieurwesen
- 5 Grund- und Werkstoffe/Kunststoffe
- 6 Energietechnik
- 7 Strömungstechnik
- 8 Mess-, Steuerungs- und Regelungstechnik
- 9 Elektronik/Mikro- und Nanotechnik
- 10 Informatik/Kommunikation
- 11 Schwingungstechnik
- 12 Verkehrstechnik/Fahrzeugtechnik
- 13 Fördertechnik/Logistik
- 14 Landtechnik/Lebensmitteltechnik
- 15 Umwelttechnik
- 16 Technik und Wirtschaft
- 17 Biotechnik/Medizintechnik
- 18 Mechanik/Bruchmechanik
- 19 Wärmetechnik/Kältetechnik
- 20 Rechnerunterstützte Verfahren (CAD, CAM, CAE, CAQ, CIM ...)
- 21 Elektrotechnik
- 22 Mensch-Maschine-Systeme
- 23 Technische Gebäudeausrüstung

ISBN 978-3-18-329317-9