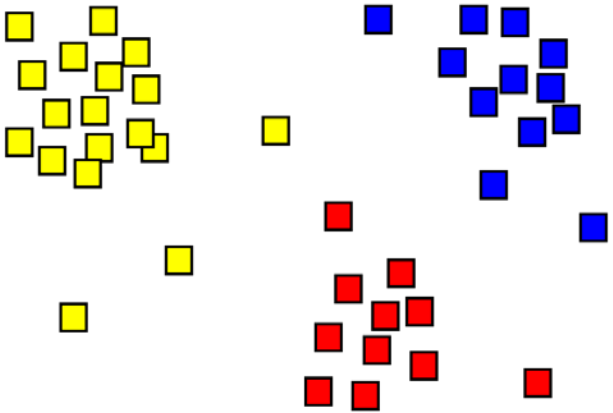


Discrimination with NIR in the biotech and pharma industry

Erik Skibsted, Principal Scientist PhD
Novo Nordisk A/S

18èmes Recontres HélioSPIR
Discrimination et classification par SPIR
Agropolis International, 27th et 28th Novembre 2017 Montpellier

novo nordisk®

Outline

- Facts about Novo Nordisk (and me)
- Applications of Near-Infrared and Discrimination Models in biotech and pharma
- Understanding powder blending with inline NIR, PCA and design of experiments
- NIR soy spectroscopy and PLS-DA model to predict titer yield in cell cultivation

Facts about Novo Nordisk

- Headquarter in Denmark
- 16 production sites on 5 continents
- Affiliates or offices in 77 countries
- Products marketed in more than 165 countries
- 41.400 employees (june 2017)
  - 19% within research and development
  - 32% in production and production administration
  - 37% in international sales and marketing
  - 12% in administration
- Five product areas
  - Diabetes care
  - Obesity and weight management
  - Haemophilia management
  - Growth hormone therapy
  - Hormone replacement therapy
- Annual sales of 111,78 billion DKr (2016) ~ 14.9 billion Euro

Who am I?

Chemical Engineer from The Technical University of Denmark (DTU)

Fluorescence Sensor Technology
- Waste water treatment
- Fermentation

PhD in near-infrared spectroscopy, chemometrics and tablet production (University of Amsterdam)

Near-Infrared applications in tablet production and control
- Power mixing
- High shear wet granulation
- Tablet analysis
- Development of statistical quality control for spectral data, splitting the spectral vector into three parts (Net Analyte Signal-SQC)

Analytical scientist and data scientist (Novo Nordisk)

Near-Infrared, mid-infrared and Raman spectroscopy
Data analysis of large non-designed data sets by chemometrics
Quality by Design implementation
- Risk assessment
- Design of Experiments

Powder mixing, tablet coating, dissolution monitoring

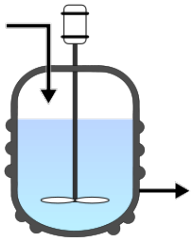
novo nordisk®

Applications of Near-Infrared and Discrimination Models in biotech and pharma

Raw materials

◦ Identification of in coming raw materials with handheld systems and spectral libraries

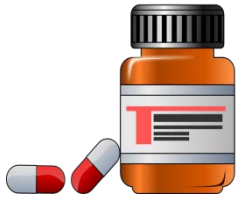
Production



◦ Process dynamics in batch fermentation
◦ Crystallisation and polymorphic transitions
◦ Steady-state monitoring of continuous processes

◦ NIR chemical imaging of tablets

Product



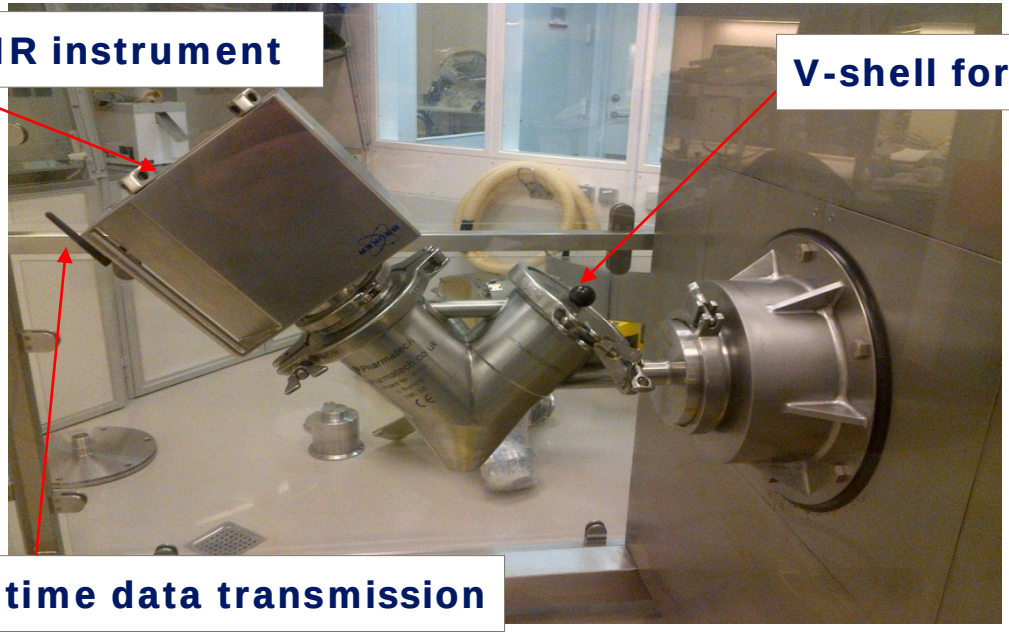
◦ Counterfeit analysis
◦ Control of products for clinical trials to avoid product mix-up
◦ Tablet friability

Understanding powder blend process using in-line NIR, principal component analysis and design of experiment methodology

novo nordisk®

In-line NIR analysis of powder blending

Battery driven NIR instrument



V-shell for powder blending

Antenna for real-time data transmission

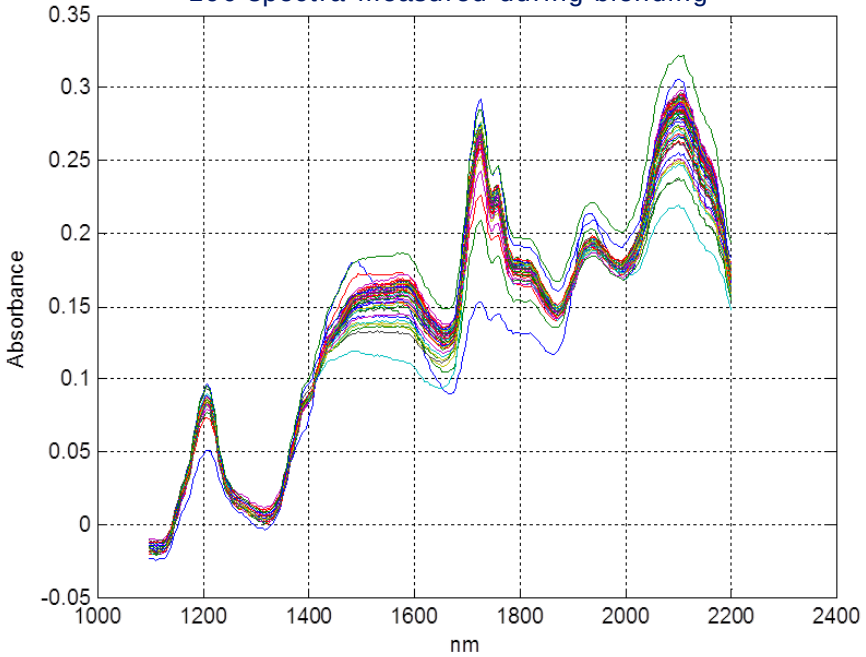
Each time the v-shell face downwards, the NIR instrument records a spectrum through a window in the top of the v-shell. The spectrum is transmitted to a PC and the concentration calculated.

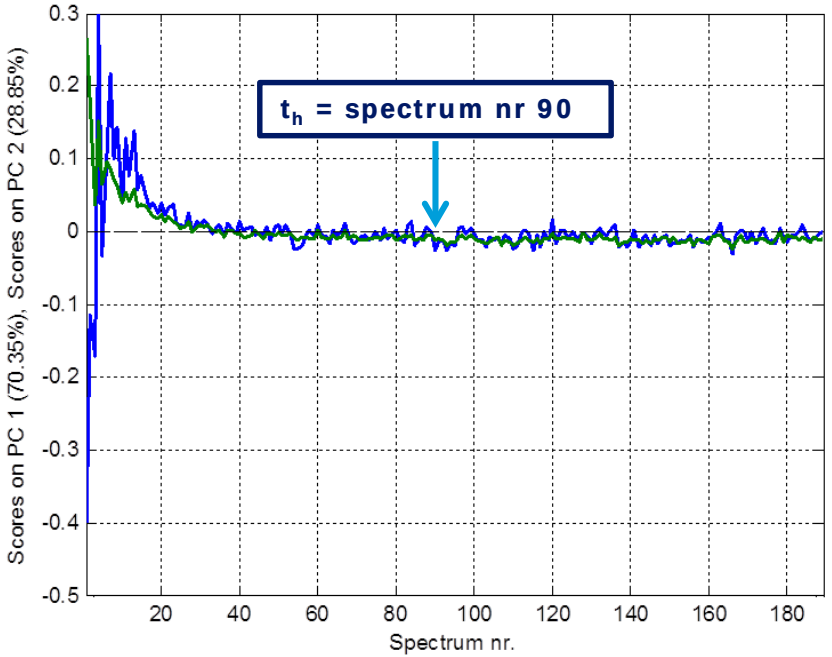
novo nordisk®

In-line NIR used to determination steady-state = time to reach homogeneity t_h

Spectra analysed by principal component analysis (PCA)
(qualitative method)

190 spectra measured during blending

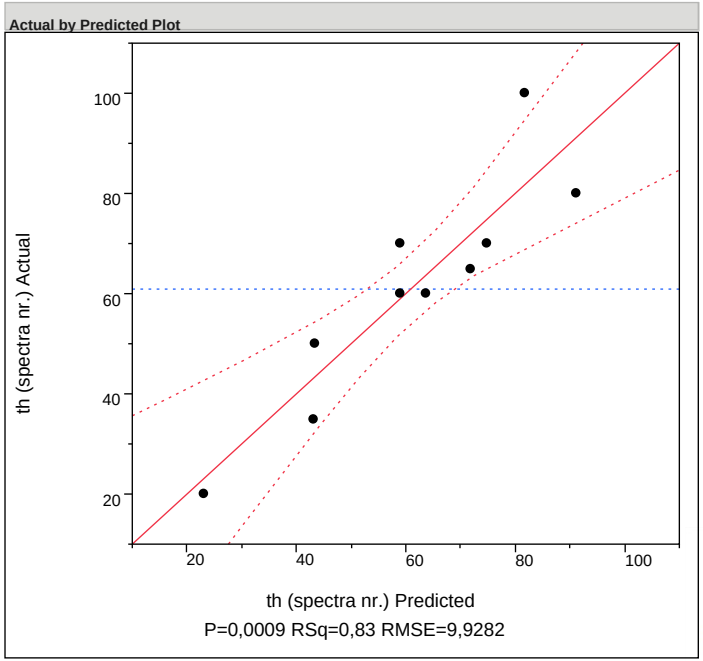




Design of Experiments (DoE) methodology used to model blend composition influence on time to reach homogeneity

$$th = a_0 + a_1 \cdot A + a_2 \cdot B + a_3 \cdot C$$

#	A (% in blend)	B (% in blend)	C (% in blend)	th (spectra nr.)
1	64.7	17.6	17.6	70
2	55.0	45.0	0.0	20
3	78.6	16.1	5.4	100
4	55.0	33.8	11.3	50
5	55.0	22.5	22.5	60
6	64.7	26.5	8.8	60
7	64.7	26.5	8.8	60
8	78.6	21.4	0.0	65
9	78.6	10.7	10.7	80
10	64.7	26.5	8.8	70
11	64.7	35.3	0.0	35



novo nordisk®

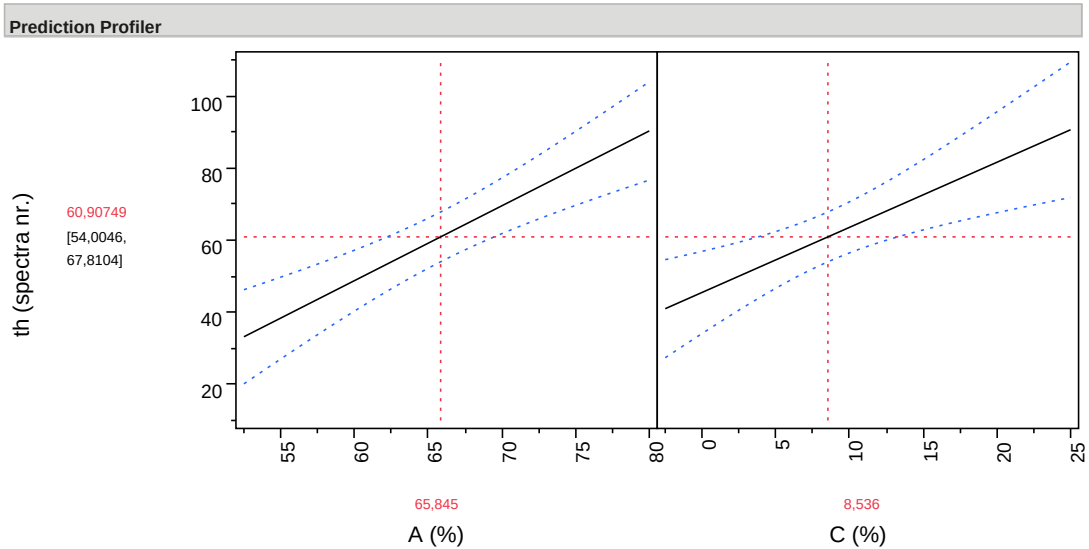
Model parameters shows which components that has the highest influence on the blend process and their quantitative influence

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	
Intercept	-91,15523	25,40519	-3,59	0,0071	^
A (% in blend)	2,0754862	0,359889	5,77	0,0004	*
C (% in blend)	1,8043966	0,460279	3,92	0,0044	*

Final model

$$th = -91.2 + 2.08 \cdot A + 1.80 \cdot C$$

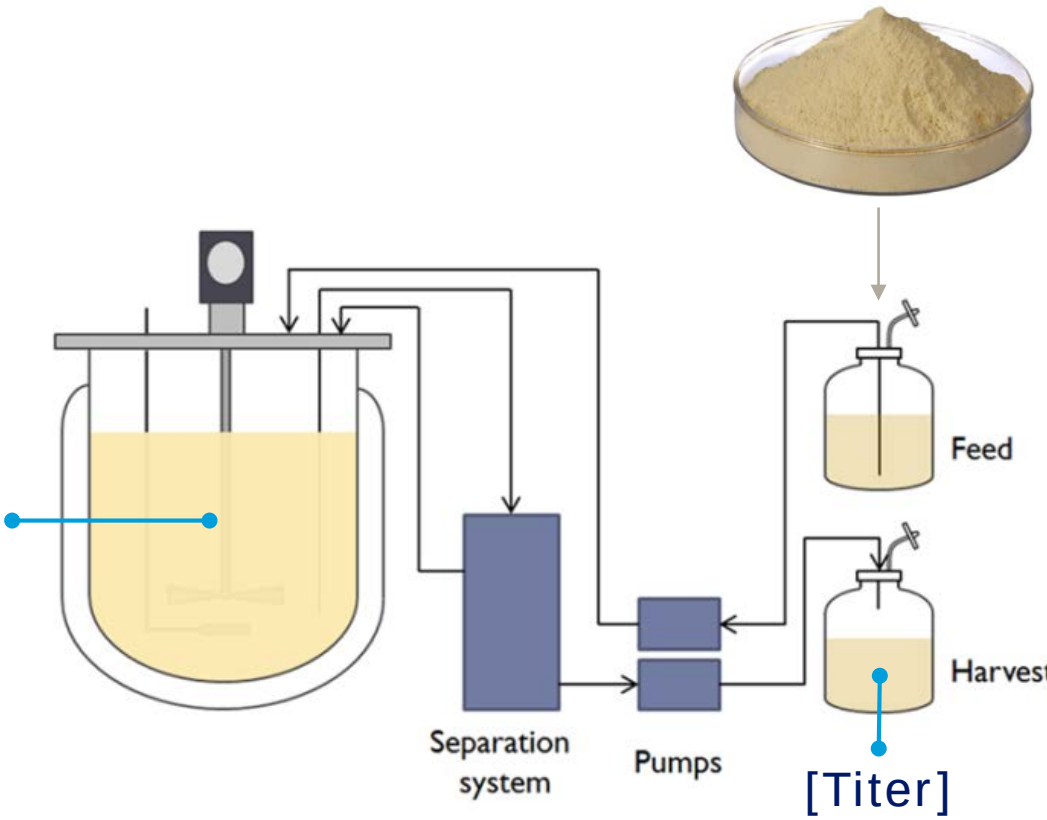


novo nordisk®

NIR classification method to screen soy hydrolysate used as feed component in a cell cultivation process

novo nordisk®

Soyhydrolysate and other feed components are used to feed a continuous cell cultivation



[Cells]
Viability
Cell diameter
[Titer]
[metabolites]

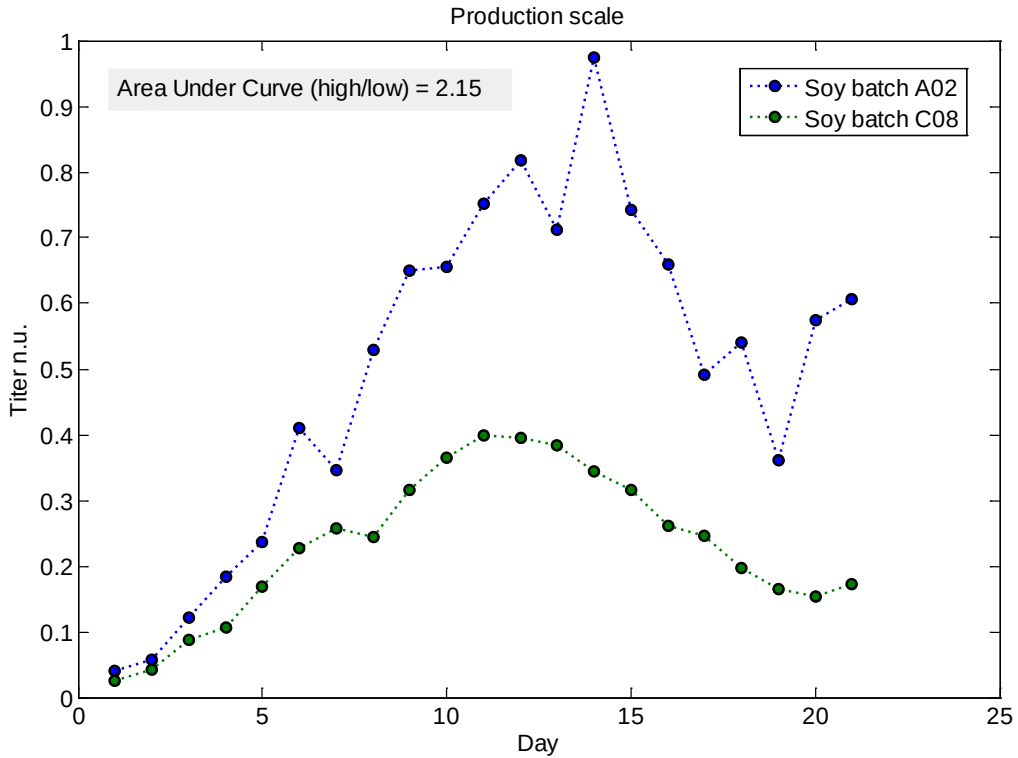
+ purification
+ formulation
+ lyophilisation

Normalised units

- Due to *intellectual property rights* are all cultivation data recalculated to normalised units (n.u.)

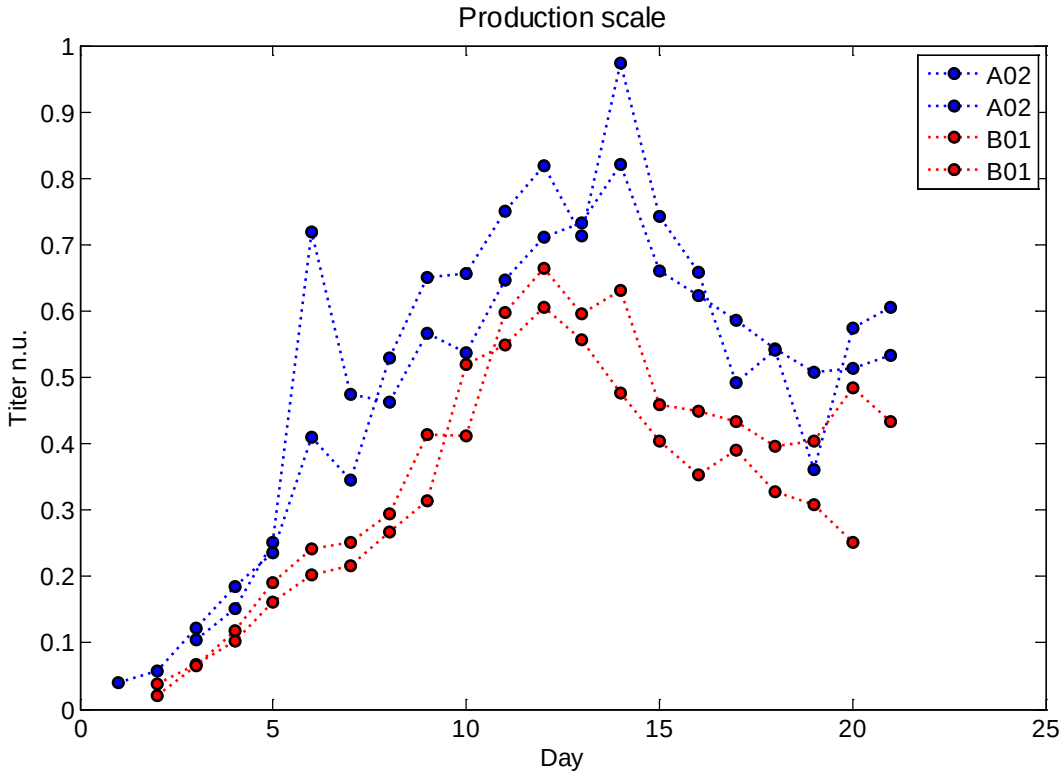
$$\text{n.u.} = \frac{\text{value}}{\max \text{value}}, \in [0;1]$$

Titer yield can vary a lot when changing soy batch. Titer quality is not affected, this is only an economic problem.



novo nordisk®

Titer results are reproducible when using the same soy

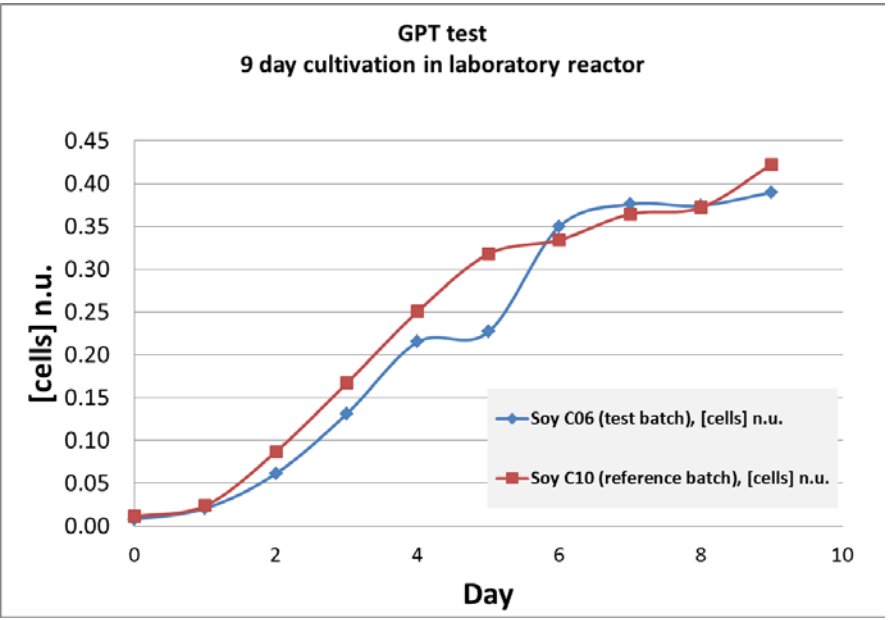


novo nordisk®

Current method to evaluate soy performance is a Growth Promotion Test (GPT) performed in a laboratory scale bioreactor

Accumulated cell concentration for day 0 to 9 (ICA_{0-9}) is compared for test and reference soy batch

If $ICA_{0-9(test)} \geq 0.82 * ICA_{0-9(reference)}$, the test batch can be used



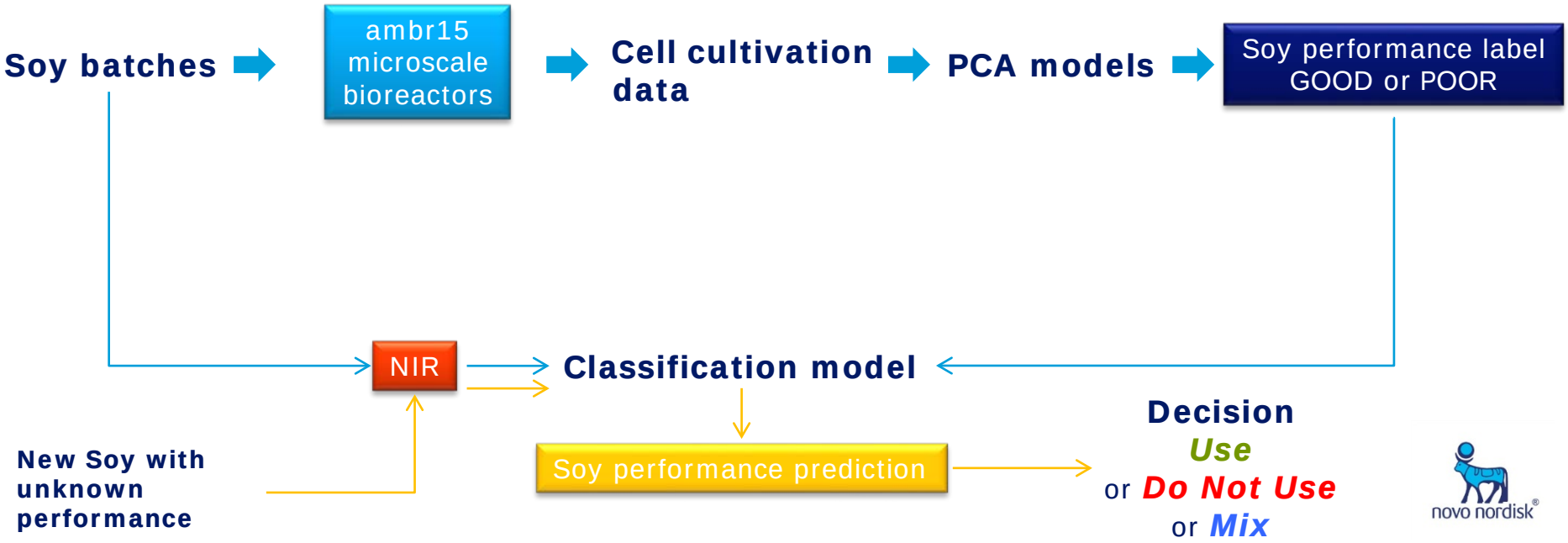
Day	Soy C06 (test batch), [cells] n.u.	Soy C10 (reference batch), [cells] n.u.
0	0.01	0.01
1	0.02	0.02
2	0.06	0.09
3	0.13	0.17
4	0.22	0.25
5	0.23	0.32
6	0.35	0.33
7	0.38	0.36
8	0.37	0.37
9	0.39	0.42
$ICA_{0-9\ (test)}$	2.15	
$ICA_{0-9\ (reference)}$		2.35
$0.82*ICA_{0-9\ (reference)}$		1.93
Is $ICA_{0-9\ (test)} \geq 0.82*ICA_{0-9\ (reference)}$?	YES →	Use test batch

Soy hydrolysate characterization by NIR

- Widely used nutrient for cell cultivations
- Complex raw material (1000+ components)
  - 60±5 % peptides/aminoacids & 20±5 % carbohydrates
- Soy hydrolysate NIR spectra has been correlated to titer yield and cell growth in mammalian cell cultivations (CHO cells) using PLS modelling
  (Jose et al. *Biotechnol. Prog.*, 2011, Vol. 27, No. 5)
- Challenges
  - Reference cultivations
    - Must be scale independent; ml → production scale 1000+ Litres
    - Biological processes (noisy data)
    - Titer analysis difficult and time consuming
    - Non-linear processes
  - NIR analysis
    - Small spectral differences between high and low yielding soy
    - Class labels must be correct
    - Risk of overfitting models

novo nordisk®

Method development and future application



Reference cultivations in ambr15

- Reference cultivations in automated microscale bioreactor system (ambr)
- 15 ml cultivation volume
- 24 small bioreactors run in parallel in one experiment
- Cultivations run for 18 days
- Cell concentration, viability, cell diameter and five metabolites were measured every day
- Titer (protein product) was measured with two different methods on day 5, 7, 10, 12, 14 and 17
- PCA models were made with SIMCA 14.1

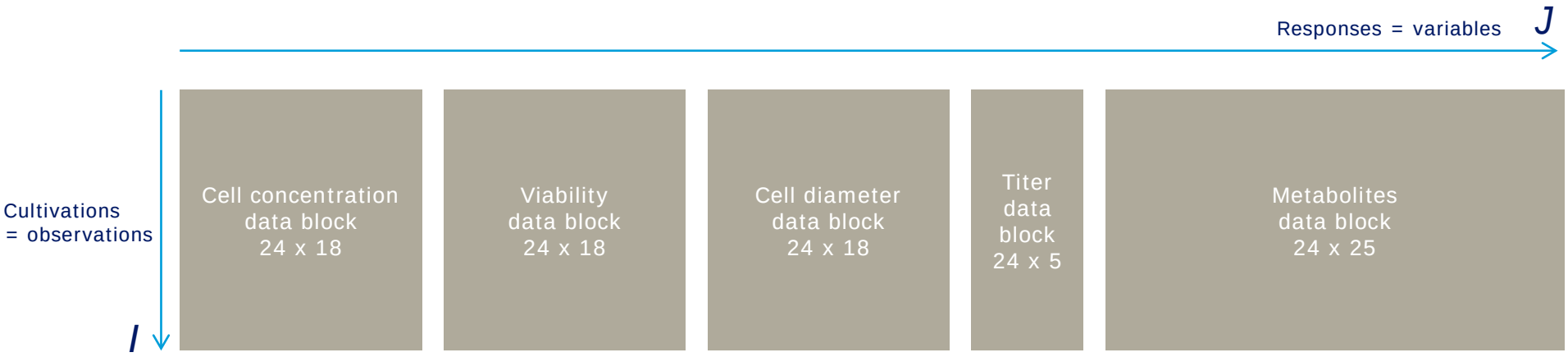
novo nordisk®

Data

	ambr15 cultivations	Laboratory scale cultivations	Production scale cultivations
No. of cultivations	9 ambr experiments x 24 cultivations = 216	18	48
Soy batches used	44 batches and 51 batch-mixtures*	14 batches	17 batches and 1 batch-mixture*
Use of data	1. Create Soy performance label 2. Calibration and validation of NIR PLS-DA model	Verification of Soy label and PLS-DA model	1. Verification of Soy label and PLS-DA model 2. Evaluation of economic potential

* Mixtures of soy batches were used in some cultivations

Data structure and PCA modelling of an ambr experiment with 24 cultivations



Step 1) Unit variance and block weight scaling of each data point x_{ij}

$(x_{ij} - \bar{x_j})/(s_j \cdot \sqrt{J_b})$, $\bar{x_j}$ is column average, s_j is column standard deviation, J_b is number of variables in a block

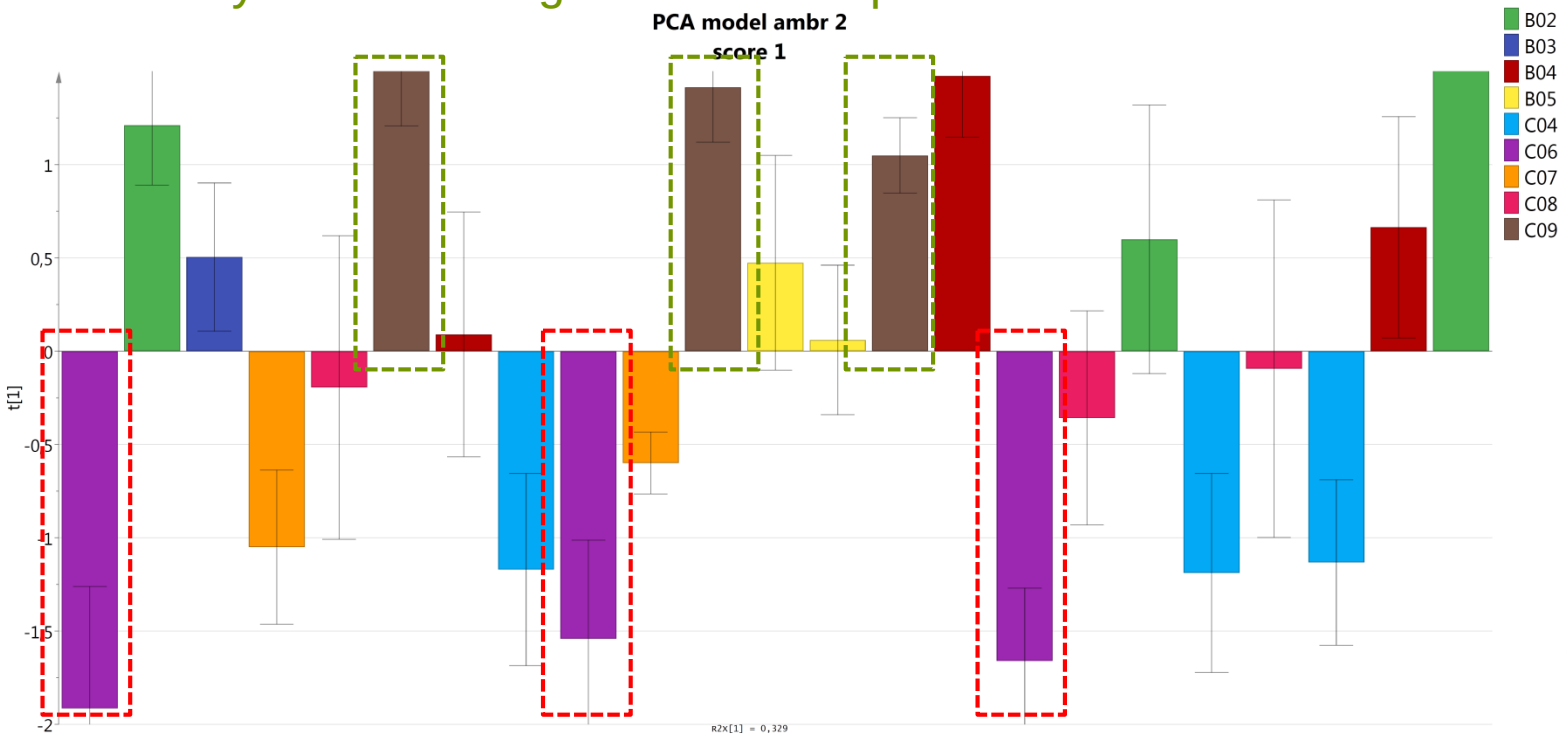
Step 2) Principal Component Analysis (PCA) of all data blocks combined

novo nordisk®

PCA model of ambr experiment 2 cultivation data

(label = soy batch)

Soy batch C09 gets a GOOD performance label



Soy batch C06 gets a POOR performance label

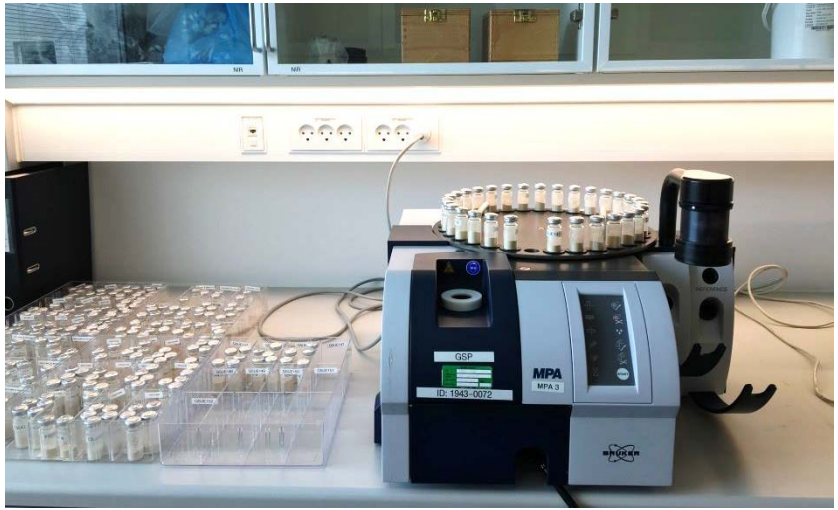
novo nordisk®

Soy batch labels based on PCA models of cultivation data – 6 of 7 soy labels were verified on laboratory scale and 7 of 8 labels were verified on production scale

Calibration and Validation, ambr		Verification, Laboratory scale		Verification, Production scale	
GOOD	POOR	GOOD	POOR	GOOD	POOR
A02	B03		B03 √	A02 √	
B02	C03	B02 √		B02 √	
C09	C04		C09 %, C04 √		C09 %
C05	C06		C06 √	C05 √	C06 √
C01	C08			C01 √	C08 √
F01	G02	F01 √		F01 √	
F02	G07	F02 √			
G03	G10				
G06	G13				
G15					
G16					
11 batches and 63 samples	9 batches and 49 samples				

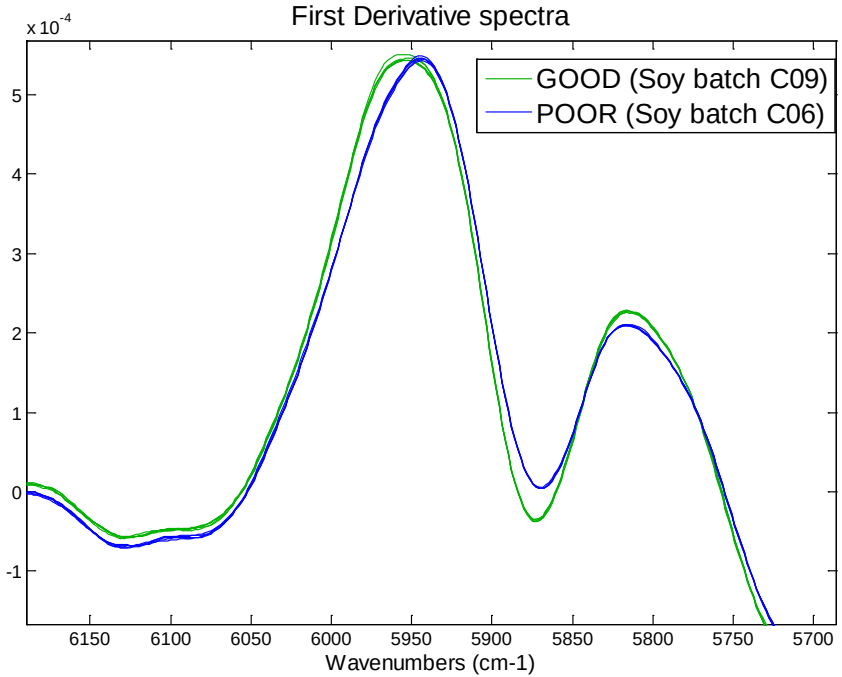
NIR analysis of soy hydrolysate

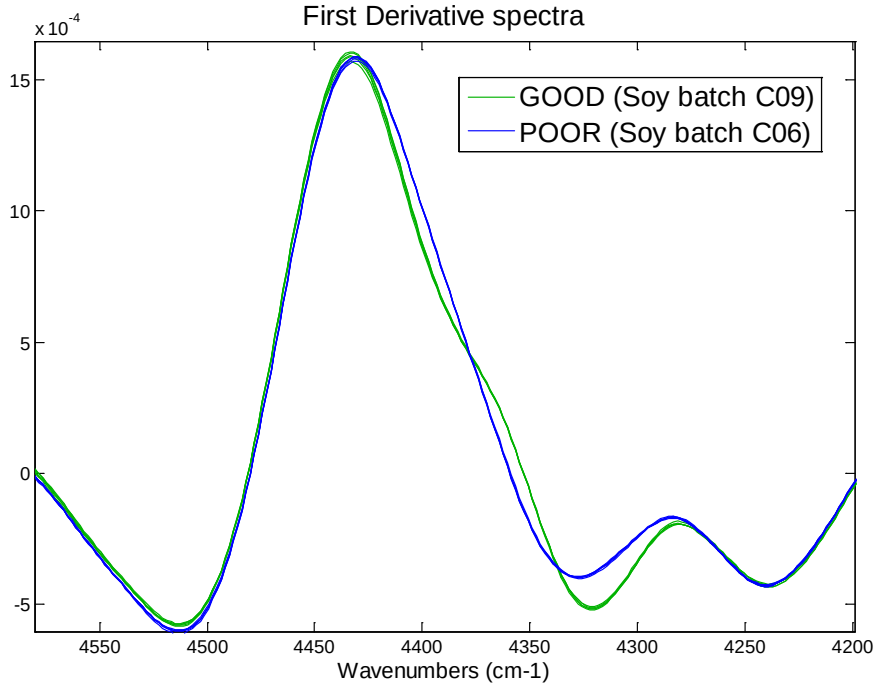
- Bruker FT-NIR MPA system with autosampler
- Reflection analysis with integrating sphere (PbS detector)
- 3800 to 12000 cm^{-1} with 4 cm^{-1} resolution and 64 scan/spectrum
- Soy hydrolysate samples kept dry in glass vials with rubber stopper and aluminium cap
- ~ Five samples per soy hydrolysate lot
- Mixture samples were prepared by weighing and mixing
- PLS-DA model were made with MatLab and PLS Toolbox 8.5.2



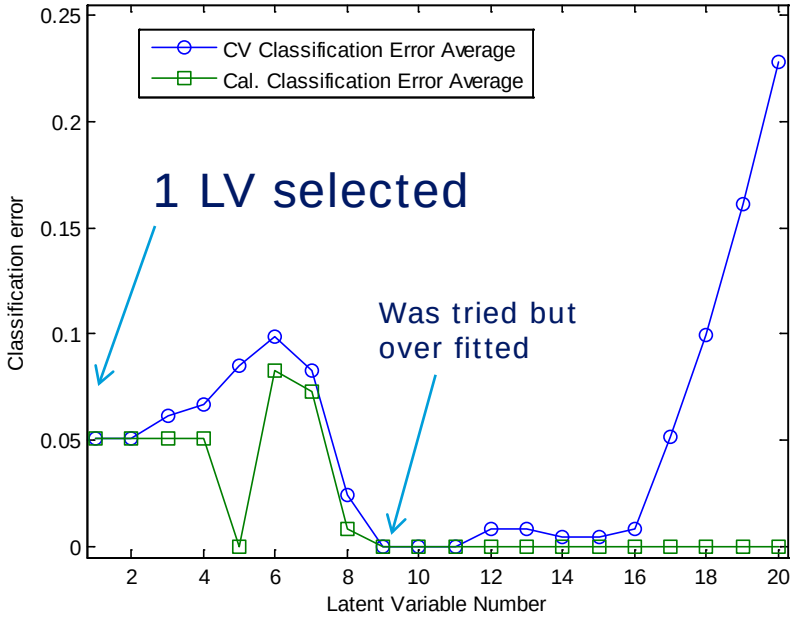
novo nordisk®

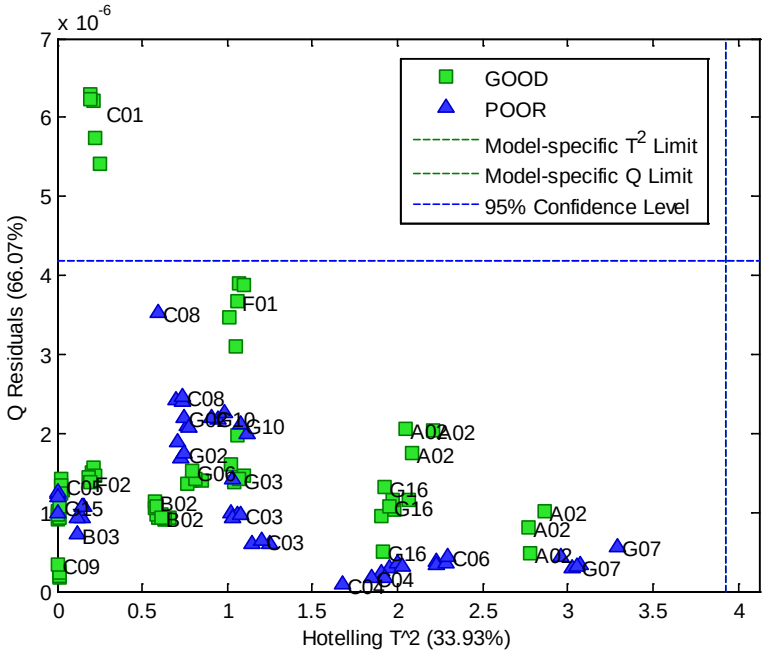
Typical spectra of GOOD and POOR soy discriminate in two wavelength regions after pre-processing



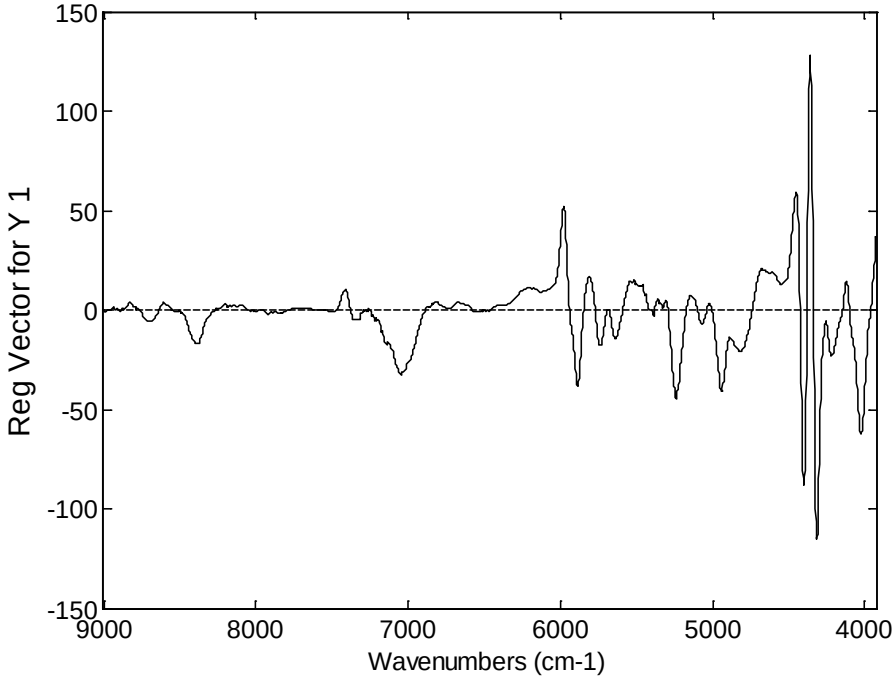


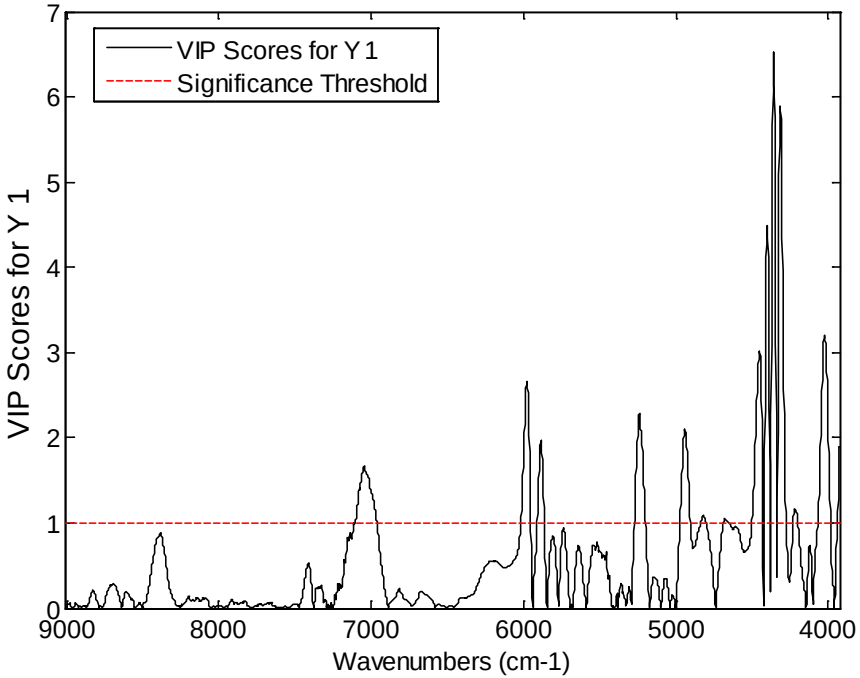
A very simple PLS-DA model with 1 latent variable (34% spectral variance explained) was developed using first derivative spectra





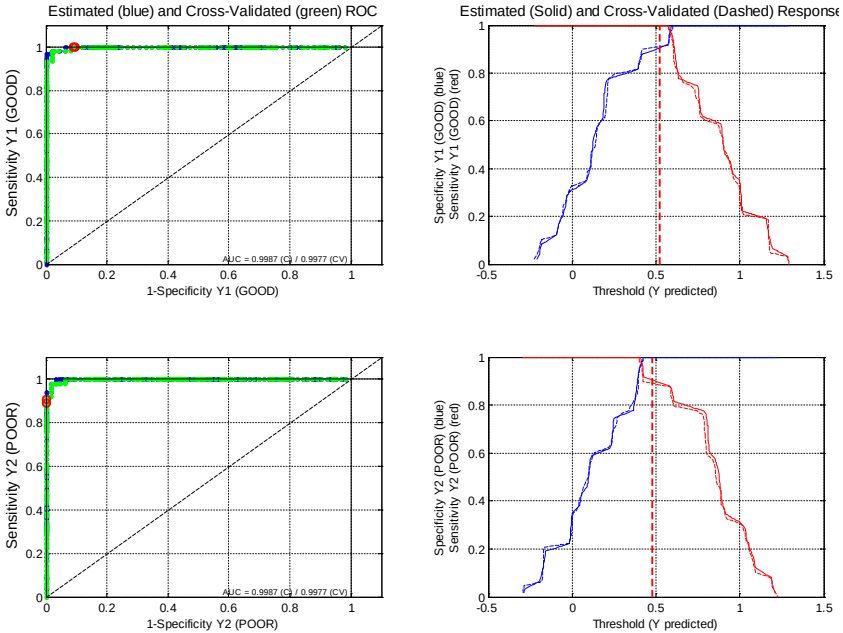
The regression vector showed clear spectral features and the VIP vector identified the important wavelengths for the classification



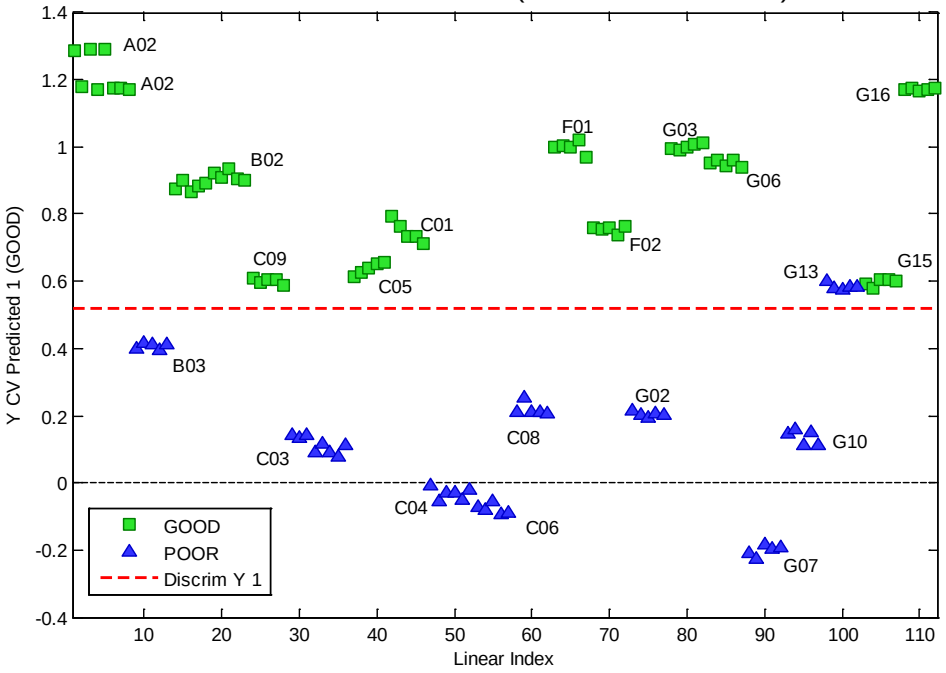


The model has a class error of 5% and a specificity of 0.898 – caused by wrong classification of batch G13 samples

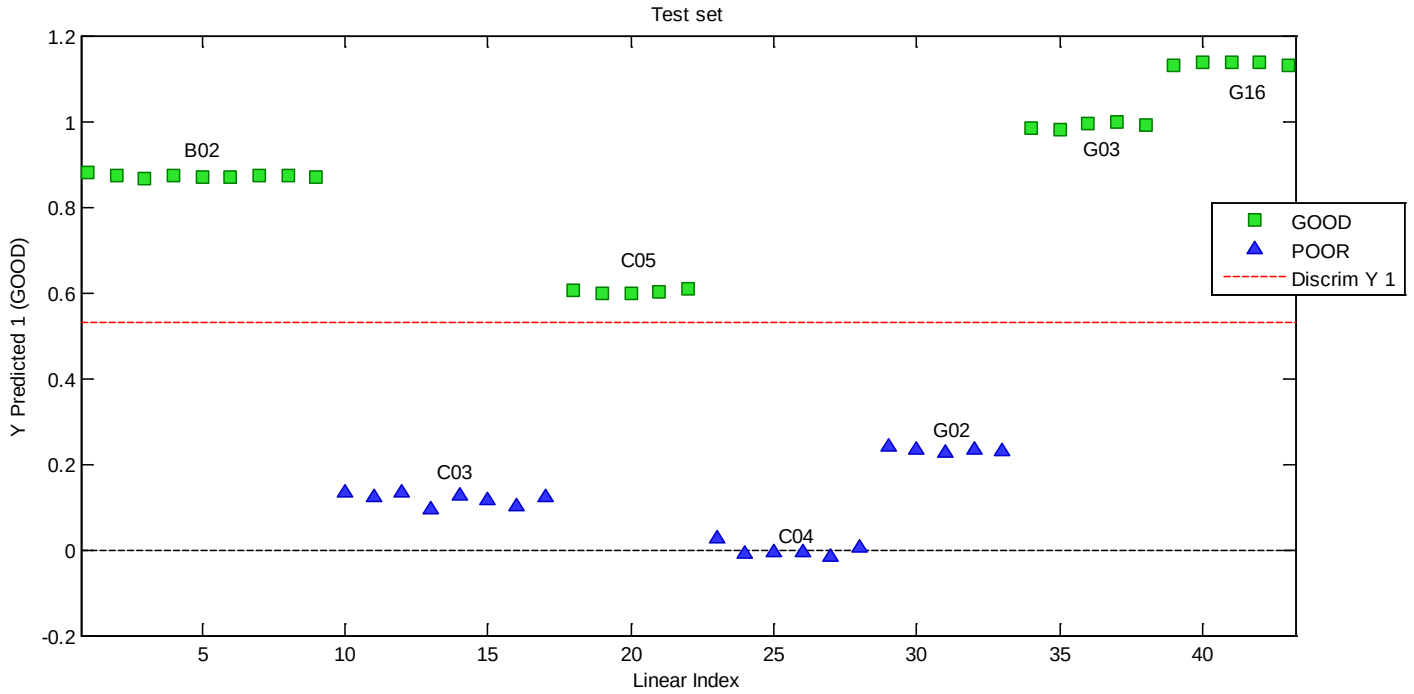
Receiver Operator Characteristics curves (ROC)



Class Prediction GOOD (cross validated result)

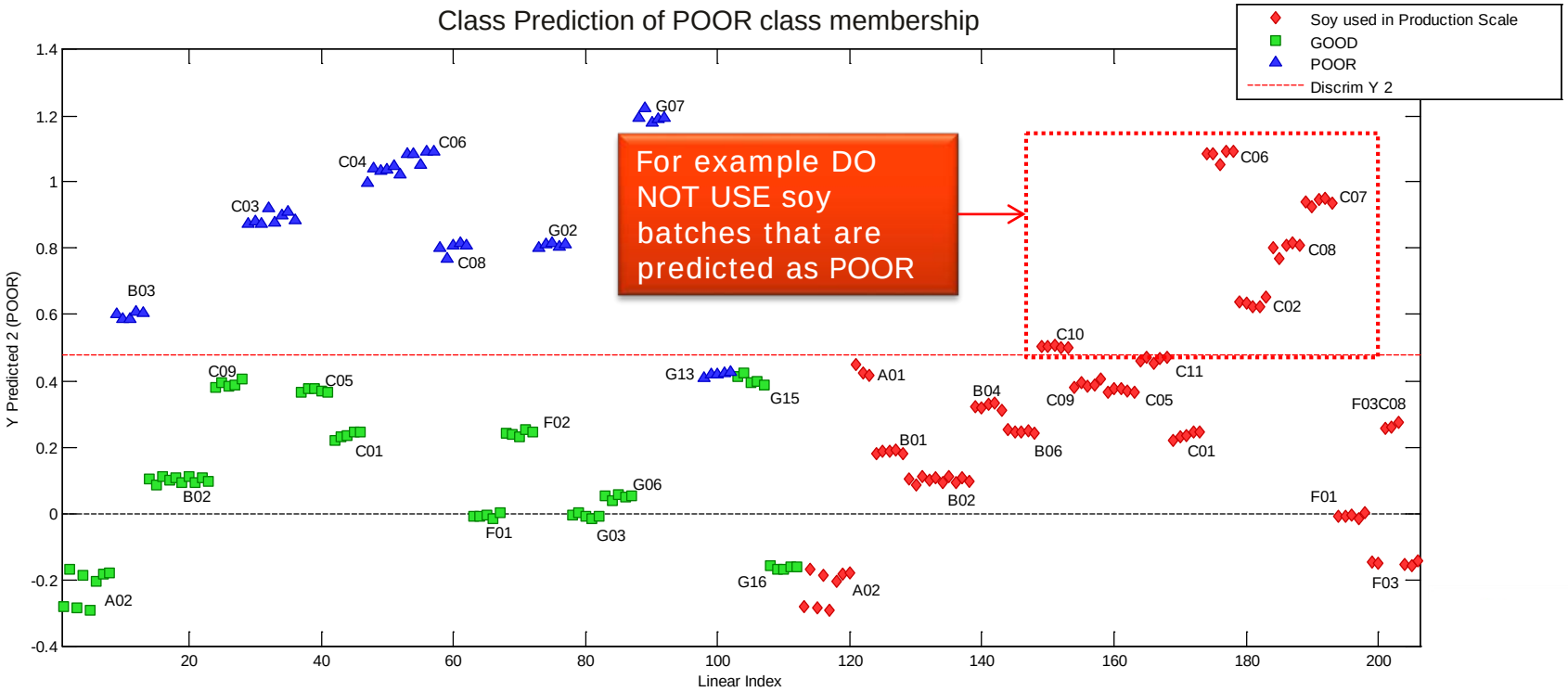


Data were split to create a test set with 24 GOOD sample spectra from four batches and 19 POOR sample spectra from three batches – 100% correct classification

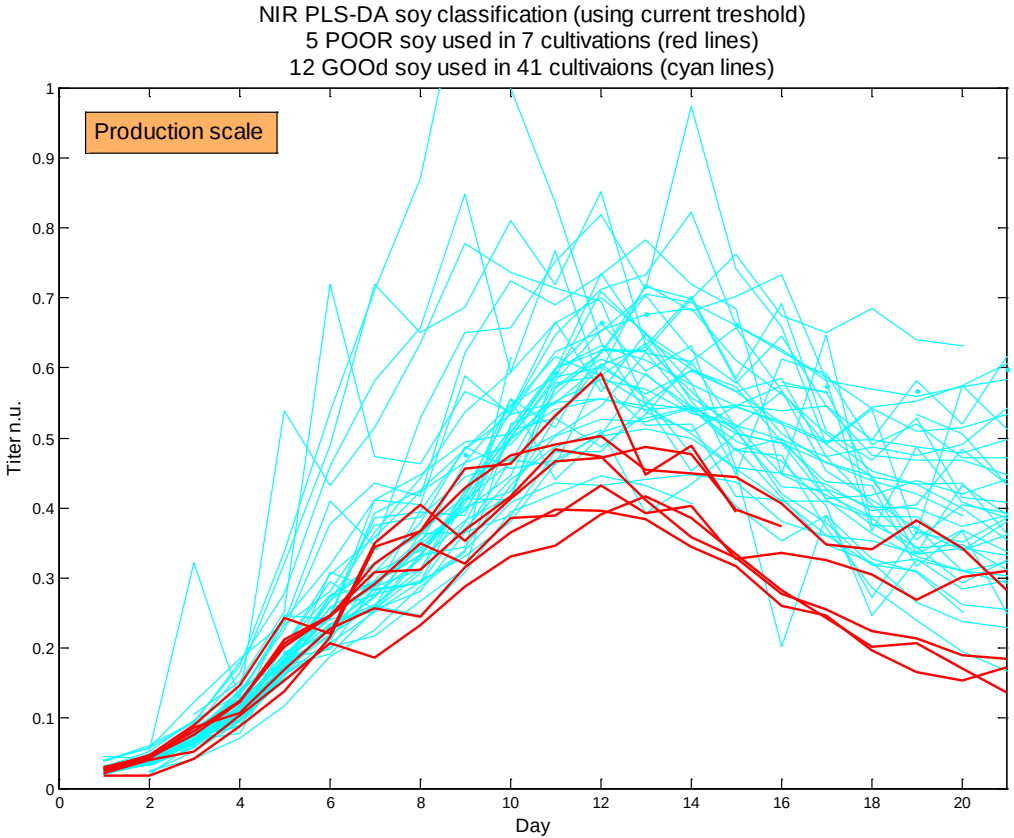


novo nordisk®

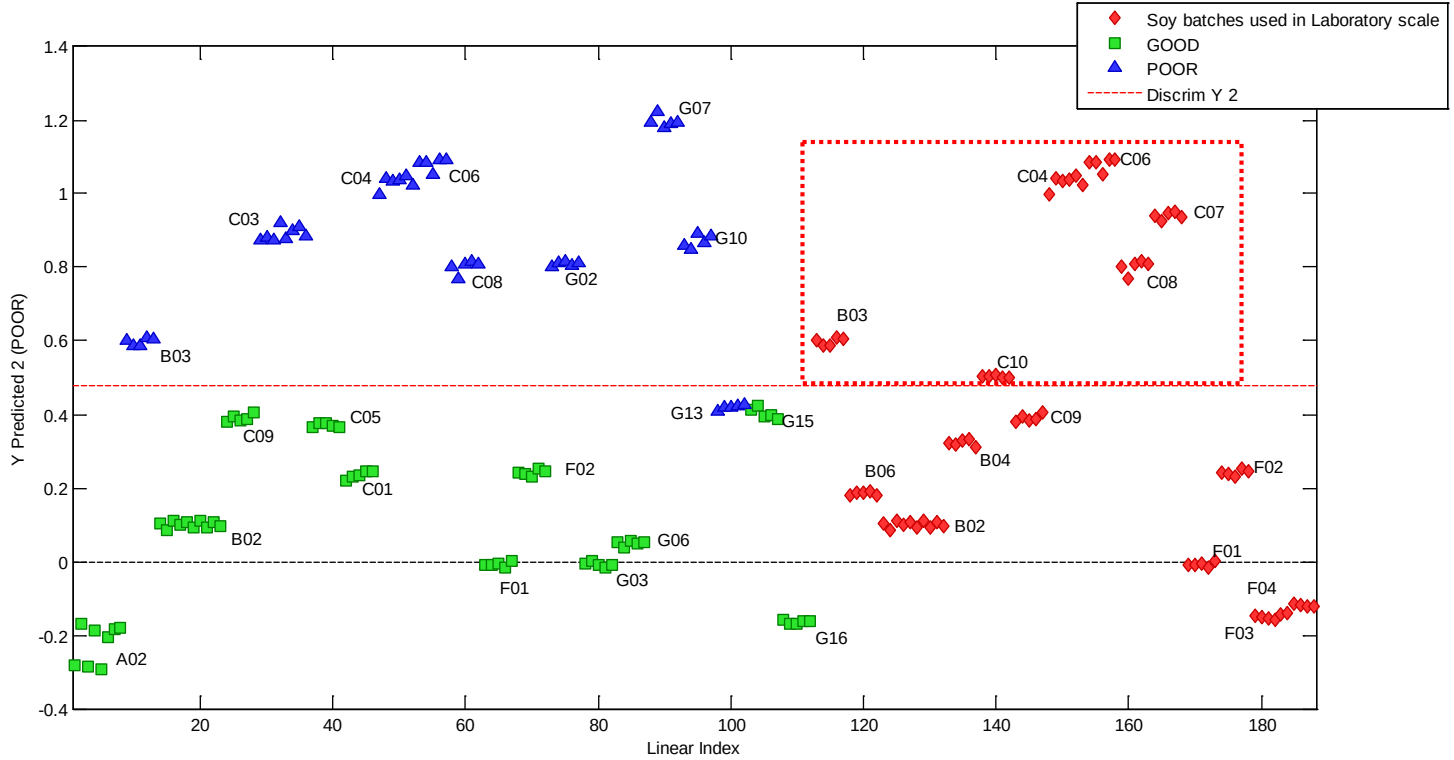
The economic potential depends on where to make the decision line



Historically seven cultivations with low titer concentration could have been avoided

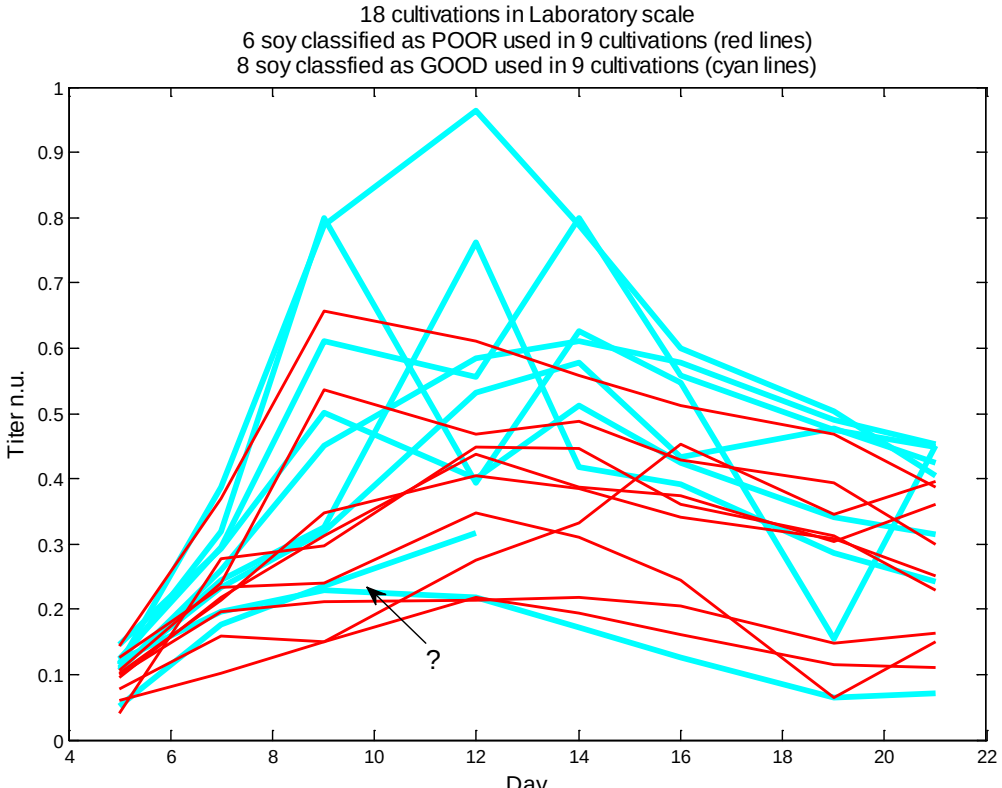


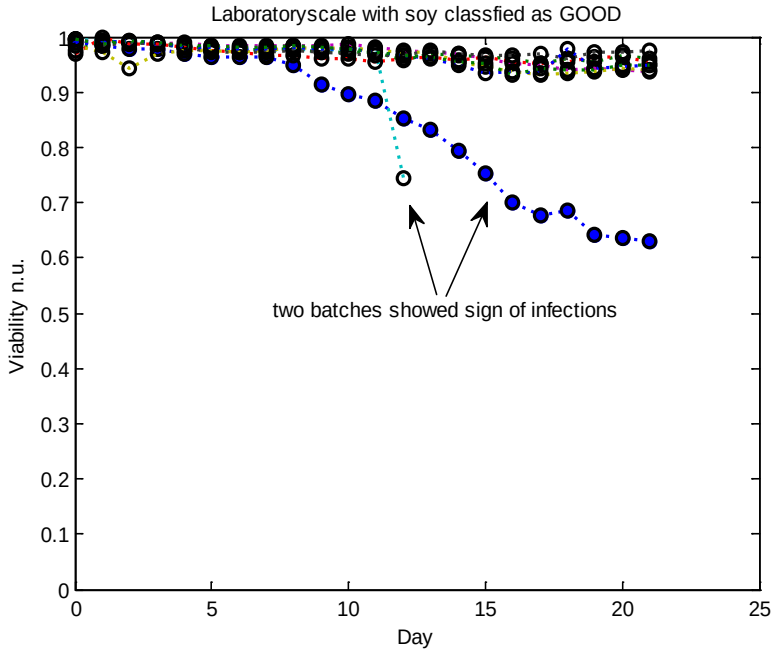
Six of the batches used in Laboratory scale were classified as POOR



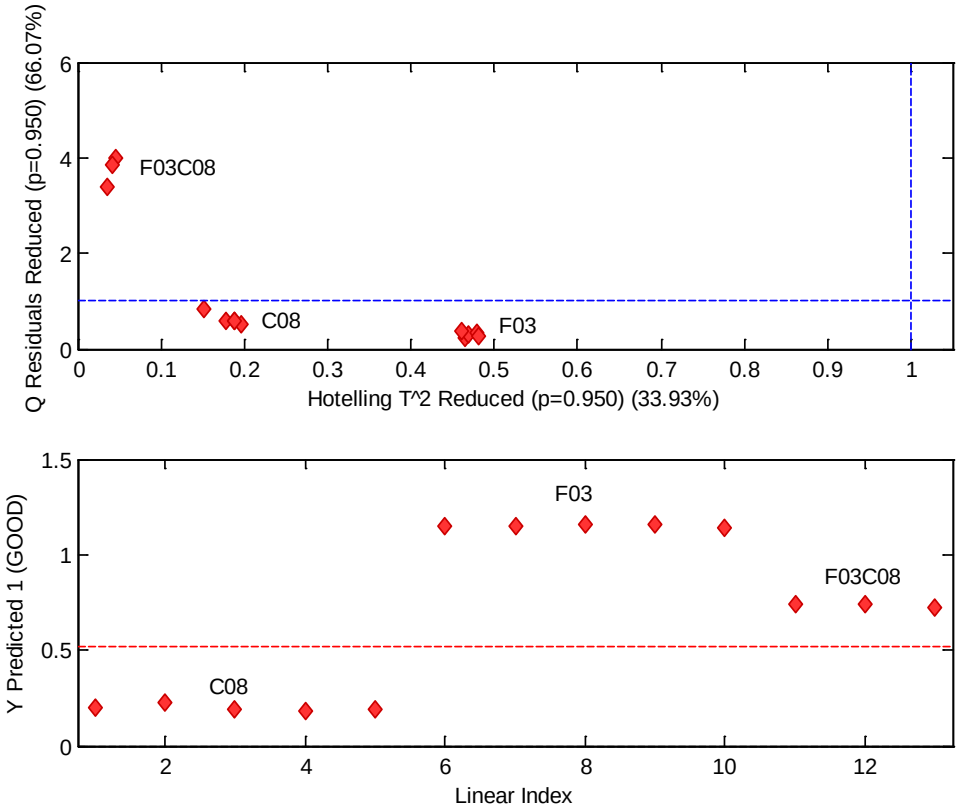
novo nordisk®

There were more titer in laboratory batches using GOOD soy compared to POOR soy. The two low producing cultivations with GOOD soy showed clear signs of infections which explains their low titer concentration!

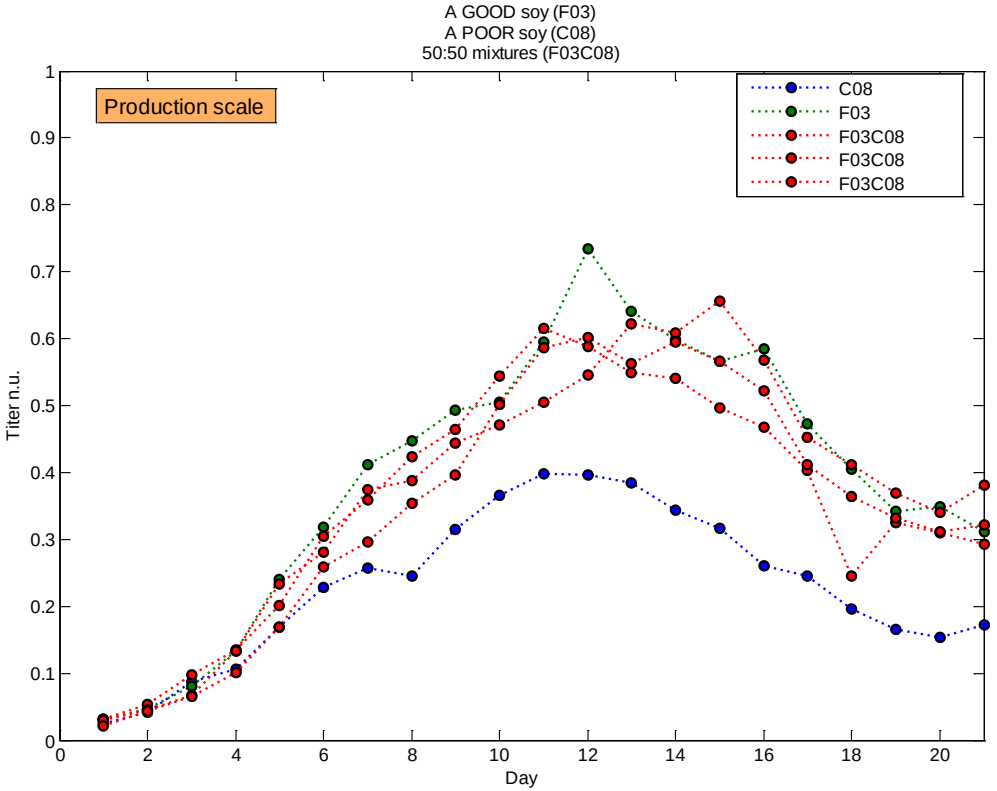




50:50 mixtures of a GOOD and POOR soy have been used in Production scale batches

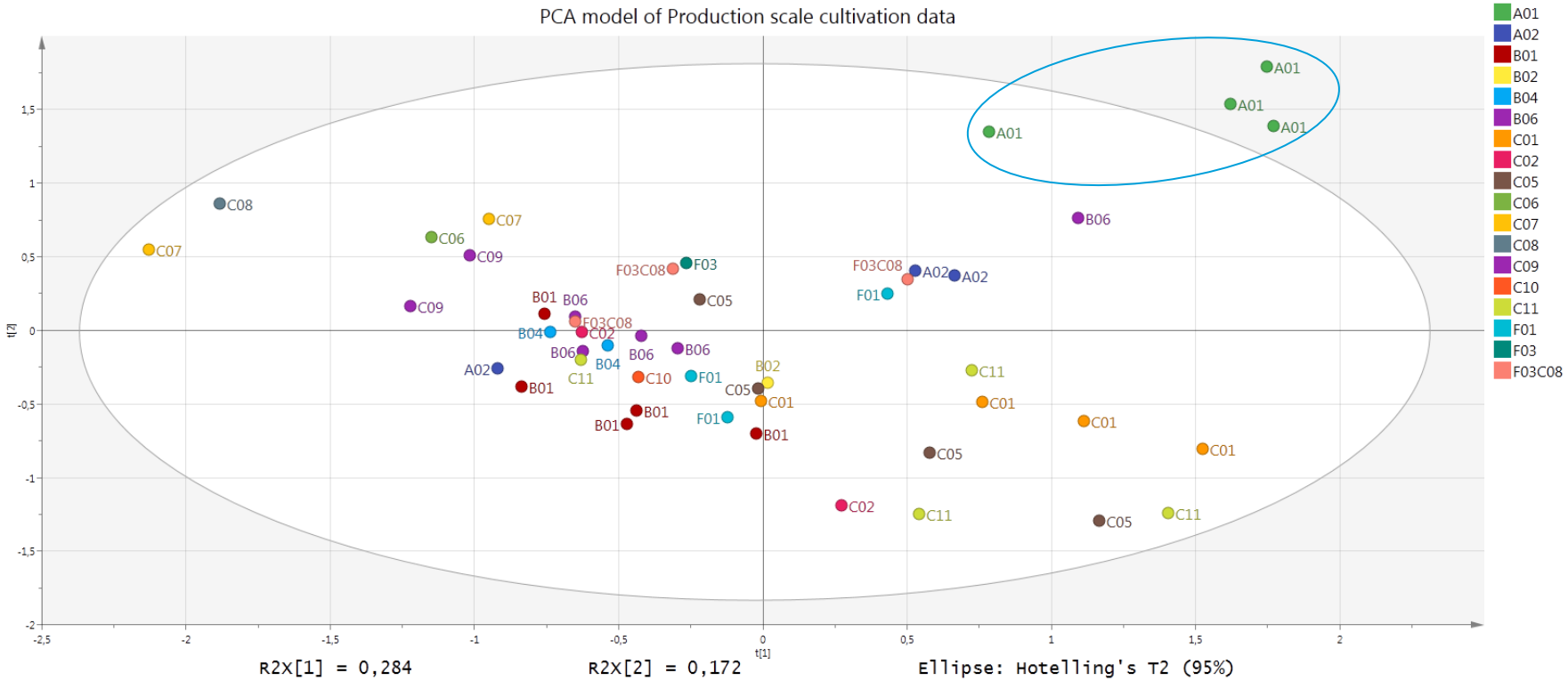


50:50 mixtures of a GOOD and POOR soy have been used in Production scale batches



novo nordisk®

An unique soy batch was observed in a PCA model of Production scale cultivation data



Regression vs Classification modelling for cell cultivation systems and raw material impact

Model type	PROS	CONS
Regression	◦ Described in literature for soy impact on titer concentration ◦ Y=titer	◦ Tried early in this project and failed! ◦ Need low-noise titer data ◦ Scale dependent ◦ Cultivation is non-linear and it is difficult to identify the titer data that shows the effect of raw material quality ◦ Missing data affects the model quality ◦ Production staff will take a titer prediction very literary
Classification	◦ Can use many response variables to establish Class label ◦ Robust to noise and missing data ◦ Scale independent ◦ E. Szabó et al., *J. Near Infrared Spectrosc.* 24, 373–380 (2016) [soy quality after heat treatment]	◦ Must select number of classes

Summary

- In this application Classification were better than Regression model
  - Scale independent model
  - Can handle a lot of noise/missing data in responses
- Simple (1 latent variable) and robust model developed
- Important to look on all available response variables, to achieve “correct” label e.g. by PCA
- Start with simple linear classification model i.e. 2-class using samples that have a *clear* labelling (provides *direction* for model)
- Potentially many applications in biopharmaceutical industry where **X** data (e.g. spectroscopy) has low noise and response **Y** (e.g. biological processes) is noisy and there are missing data

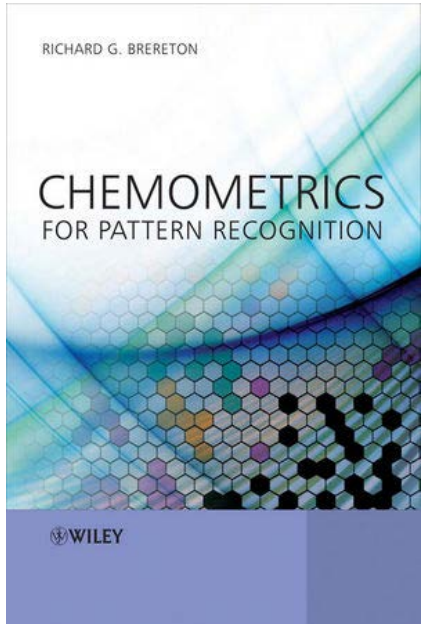
Acknowledgements

- Senior scientist Lars Poulsen (ambr)
- Statistician David Meisch (database)
- Principal Scientist Hanne Vierø Tøttrup (production scale)
- Senior pilot scientist Leif Kongerslev (production scale)
- Technician Johnni Damm (laboratory scale)
- Project Director Henrik Nordstrøm Ferré (project management)

novo nordisk®

Merci beaucoup pour votre attention….

Avez-vous des questions?



novo nordisk®