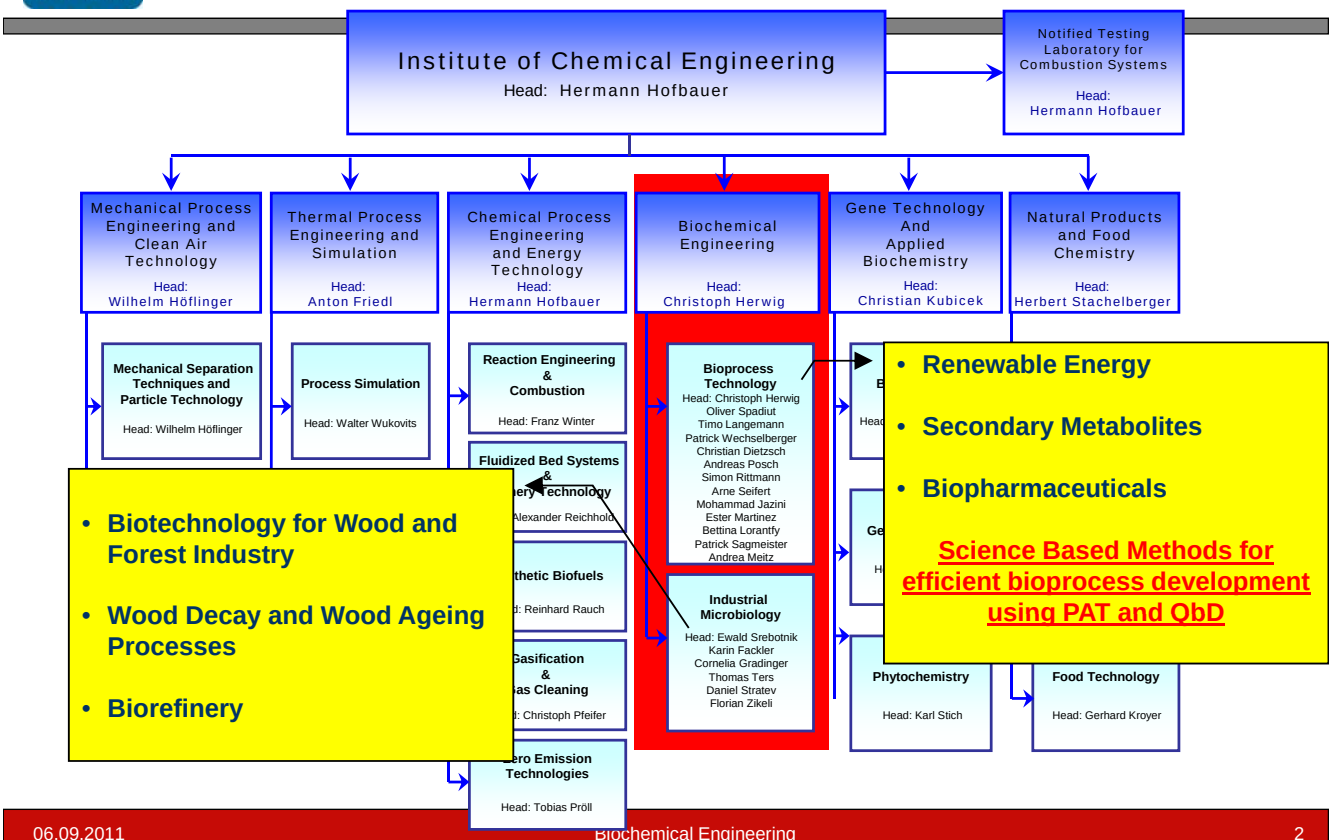
"Real-time Transformation of Data to Knowledge - preferably using soft sensors?"

1.9.2011

Christoph Herwig

Building Bridges from Engineering Sciences to Biochemistry



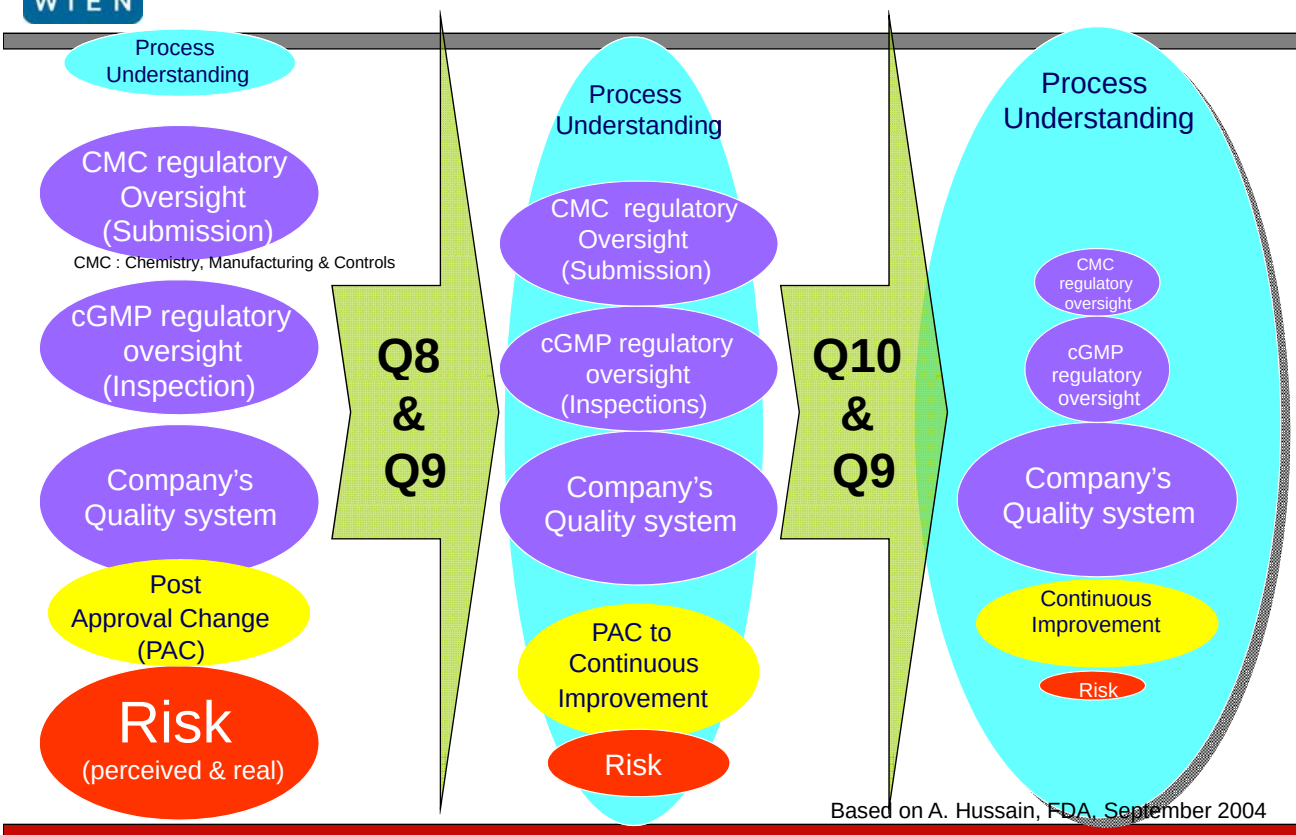
Implementation of QbD (ICH guidelines)

- Development of effective pharmaceutical quality system
  - Define **knowledge management** rules
  - Define quality **risk management** rules
  - Define **processes, resources and responsibilities** for quality assurance
  - Facilitate **continuous improvement**
- Development of production process
  - Establish **Product Understanding**
    - Critical Quality Attributes (CQA) and its claimed ranges
  - Establish **Process Understanding**
    - Critical Process Parameters (CPP) and risks for impacting CQAs
    - Sources of variability
    - Design Space for parameters to meet CQAs
    - Control Space, Normal Operating Range and Control Strategy of Process Parameters

PAT

"Quality by Design is about doing things consciously"

Potential Improvement by QbD!



Based on A. Hussain, FDA, September 2004

QbD

Process Parameters

Temperature
Stirrer Speed
Dissolved Oxygen
pH
Air Flow
Pressure
Feedrate
Nutrient concentrations
Inducer concentration
Biomass Conentration
Induction Time
Conductivity
Redox niveau
Strain
Expression casette
…

???

Product quality attributes

Enzyme activity
Titer
Purity
Stability
Batch-to-batch variability
Efficiency
Cost of product
Space-time-yield
Protein folding
Glycosylation pattern
Viability
Ease of further processing (downstream)
Potential risks for end-user
…

QbD

Process Parameters

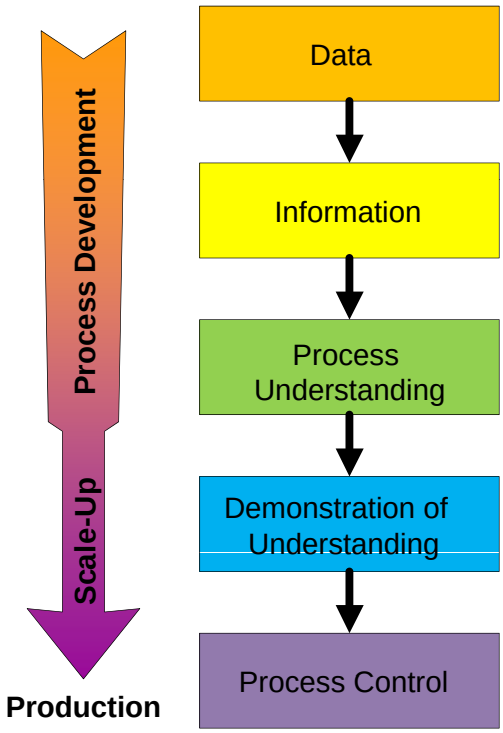
Temperature
Stirrer Speed
Dissolved Oxygen
pH
Air Flow
Pressure
Feedrate
Nutrient concentrations
Inducer concentration
Biomass Conentration
Induction Time
Conductivity
Redox niveau
Strain
Expression casette
…

???

Product quality attributes

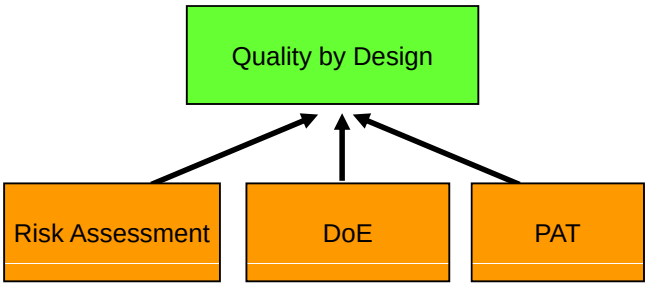
Enzyme activity
Titer
Purity
Stability
Batch-to-batch variability
Efficiency
Cost of product
Space-time-yield
Protein folding
Glycosylation pattern
Viability
Ease of further processing (downstream)
Potential risks for end-user
…

Our „how to do“ understanding



Only well quantified process parameters can be checked for correlation with Quality Attributes: -> CPP

Process Understanding Generation Tools:



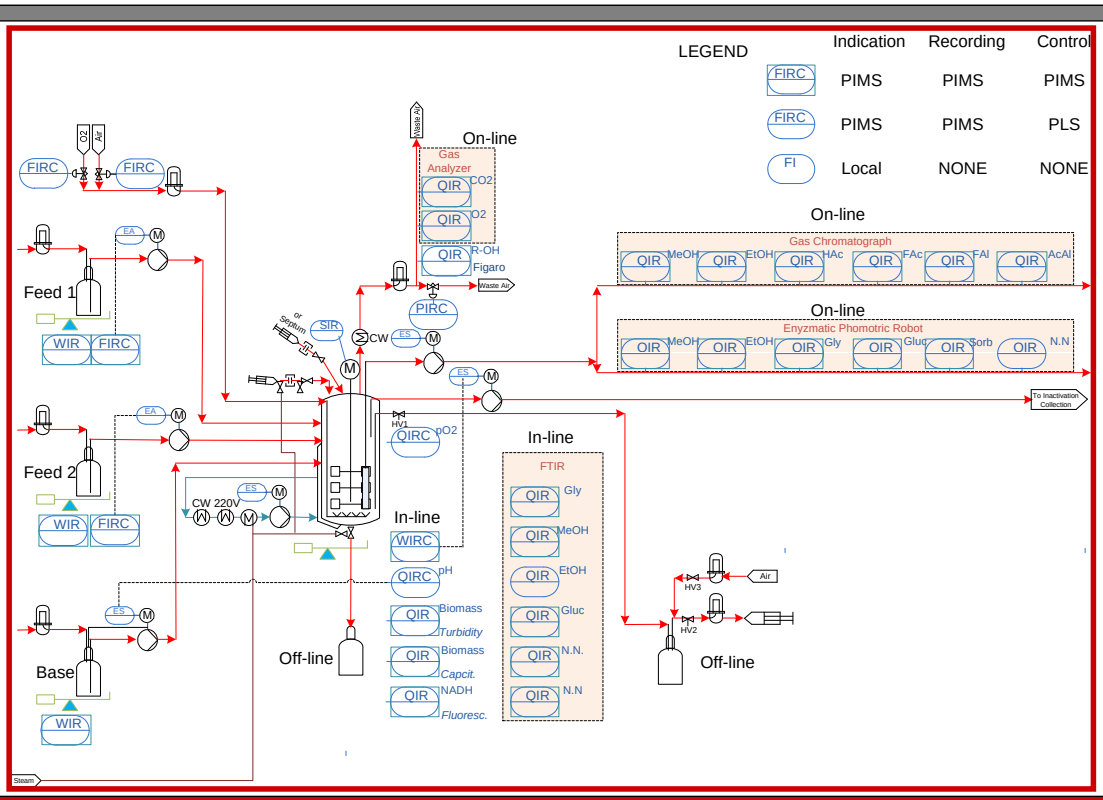
Herwig. C, Process analytical technology in biotechnology. Chemie Ingenieur Technik (2010), 82, (4), 405-414.

Presentation Goal

- Before Assessing Product Quality:
- Show Methods for **Quality Assessment** of
  - Data
  - Information
  - Knowledge
- Analyze how softsensors can help

DATA SOURCES

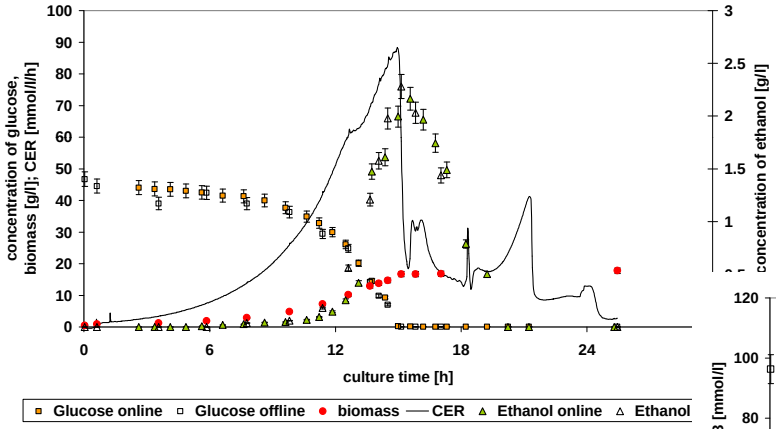
Bioreactor Process Flow Diagram

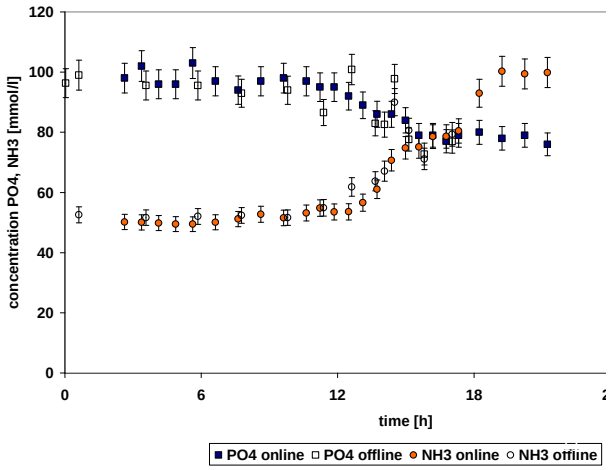


On-line Data with enzymatic photometric robot

CuBiAn Online measurement

Picha pastoris batch on Glucose

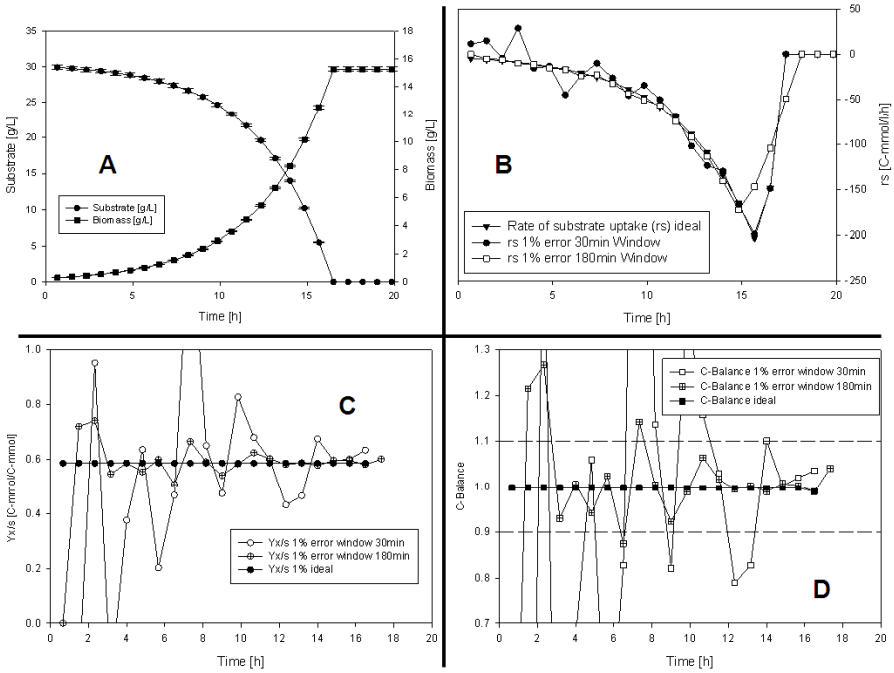




- Good on-line / off-line fit:
- sufficient for quantitative bioprocess dvelopment?

But Data are noisy and information is difficult to be extracted

- Effect of noise on data, but even worse on information

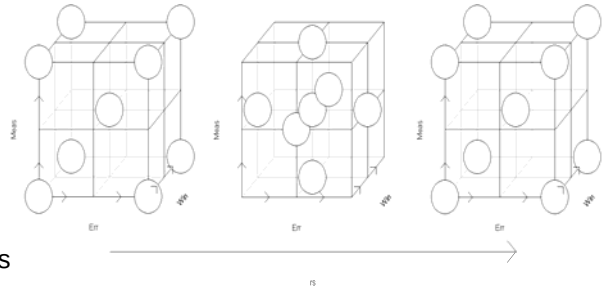


Making extractability of information quantitative

CCF-Design
Takes quadratic effects into account

Factors

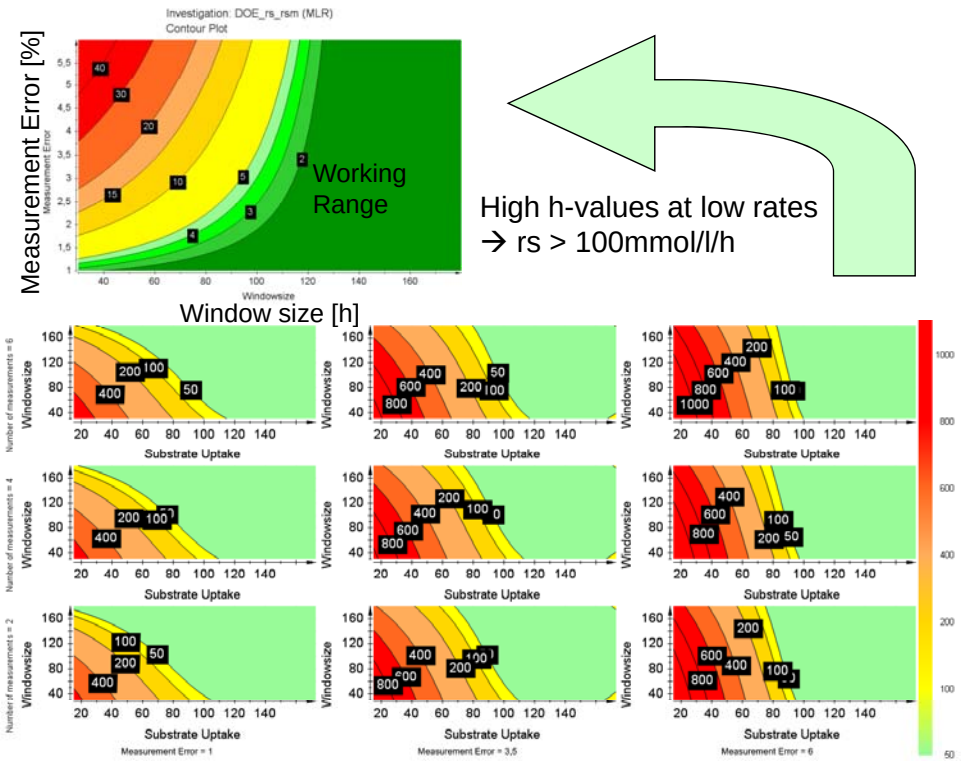
- # of measurements [h^{-1}]
- Window size [h]
- Measurement Error [%]
- Rate Substrate Uptake rs [mmol/l/h]



Response

- H-value from statistical test of sum of least squares [-]

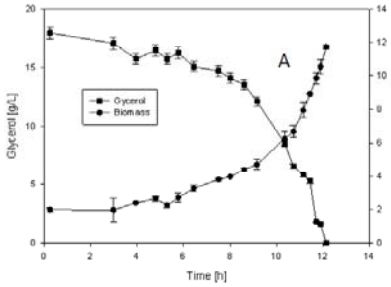
Result

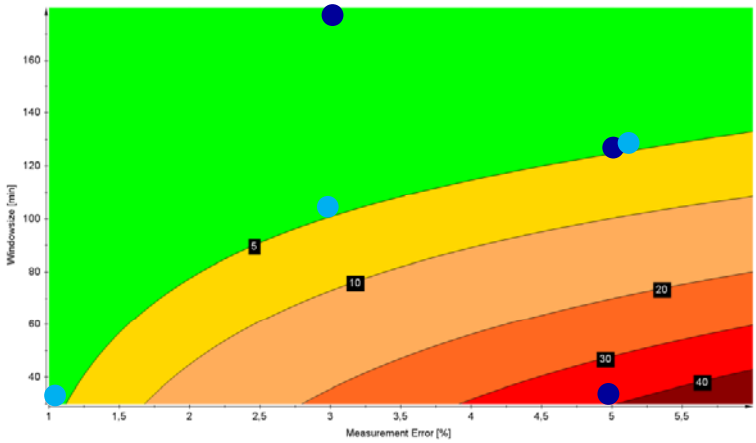


High h-values at low rates
→ rs > 100mmol/l/h

Which measurement accuracy and frequency is necessary to extract information?

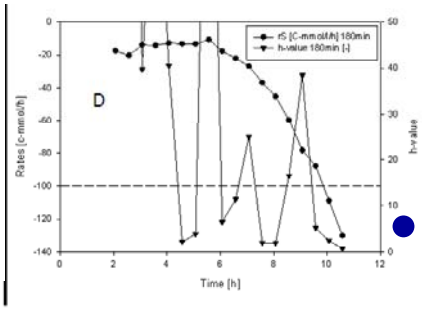
- DoE result and validation:





With
- Constant measurement frequency
- rs: >100 mmol/l/h:

1% Measurement Error / 30 min dynamics
3% Measurement Error / 100 min dynamics
5% Measurement Error / 120 min dynamics

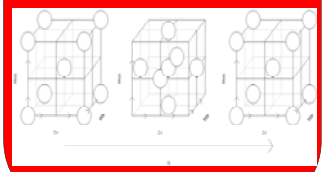


Wechselberger, P, Herwig, C, Biotechn. Progr., accepted

Integrated concept

Experimental Design → **Data** → **Information** → **Knowledge**

Design of Experiments
- Identification of CCPs/CQAs
- Risk based approach to identify non critical parameters
- Definition of analytics: What signal quality is needed for which CCPs/CQAs?



Real Time Measurements
- Off-gas
- Feed rates, base consumption
- In-line sensors: FT-MIR, capacitance
- At-line: GC, Enyzmatic Photometric

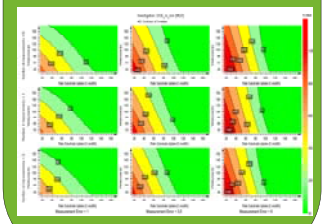
Morphology
- High throughput,
- Automated light microscopy
- x,y,z axis scanning
- Feature extraction algorithms

Data Mining
- Calculation of scale-independent specific rates and yields
- Determination of morphologic key parameters

Soft Sensors
- Kinetic Models, Metabolic Analysis Observers
- Calculation of non-measured quantities

Reconciliation
Consistency check

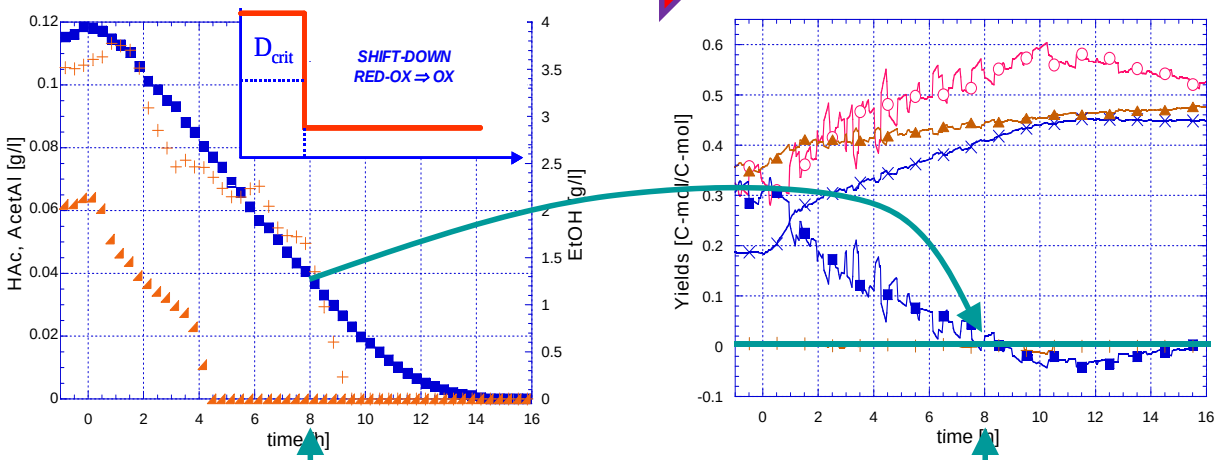
Design space
- Data exploitation routines
- Extraction of process understanding
- Establishment of design space
- Control strategies



EXTRACTION OF INFORMATION

Why is the conversion of Data to Information important?

DATA → **INFORMATION**



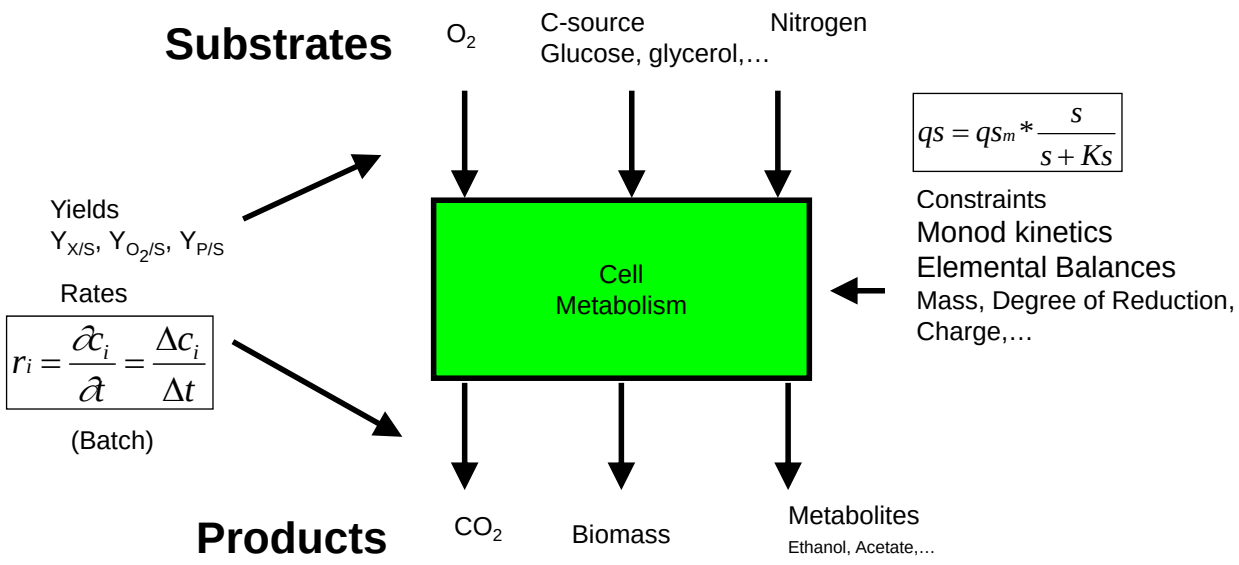
Raw Data can be misleading!

Own method to quantify consistent rates in dynamic process conditions in real time

Herwig, C, et.al.. 2001. On-line stoichiometry and identification of metabolic state in dynamic process conditions. Biotechnol. Bioeng. 75:345-354.

Black Box Model and first principle relationships

Metabolic Network



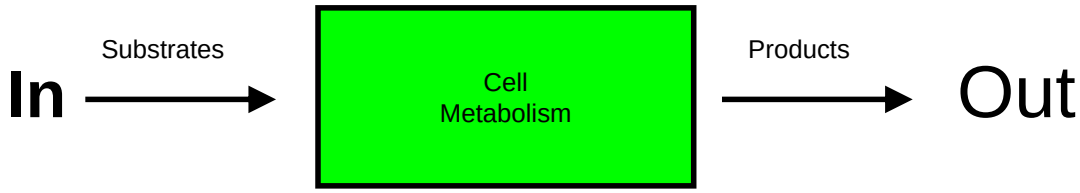
Simplified version of cell metabolism = black box

Elemental Balances as tool for consistency check and soft sensor

In = Out

C-Balance $C_{C\text{-}Source} = C_{CO_2} + C_{Biomass} + C_{Metabolite}$

N-Balance $N_{N\text{-}Source} = N_{Biomass} + N_{Metabolite}$

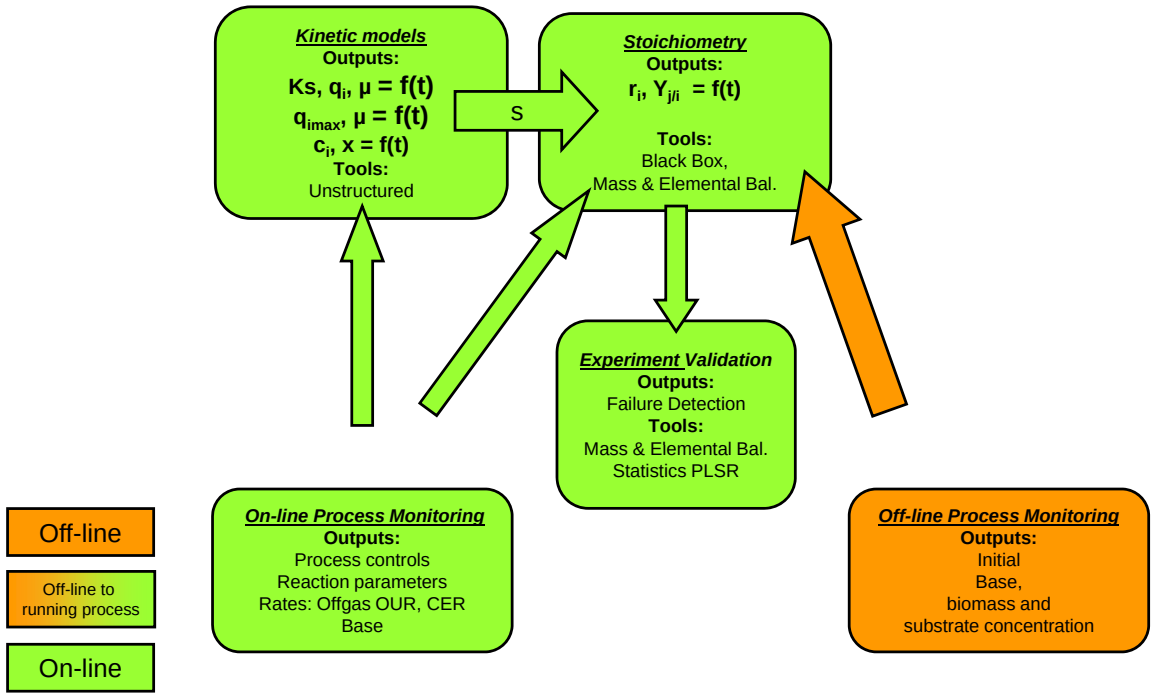


DoR-Balance $DoR_{C\text{-}Source} + DoR_{O_2} = DoR_{CO_2} + DoR_{Biomass} + DoR_{Metabolite}$

Charge-Balance $Charge_{Substrates} = Charge_{Products}$

SOFTSENSORS TO GAIN INFORMATION

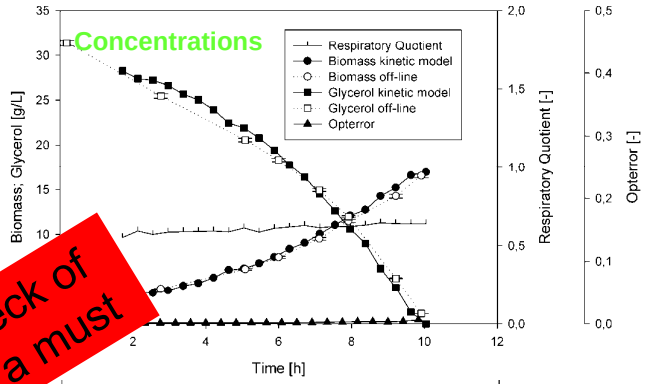
Strain Characterization in BATCH CULTURES with simple on-line continuous measurements

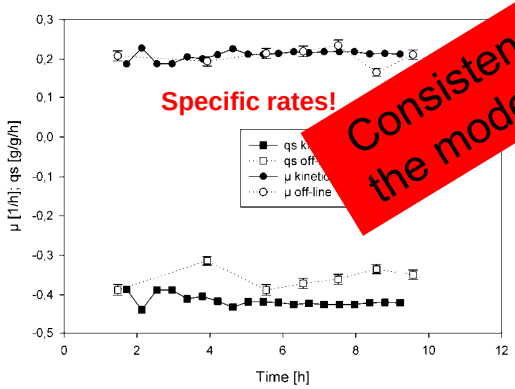


Early strain screening: On-line yields, rates and experimental validation in batch culture

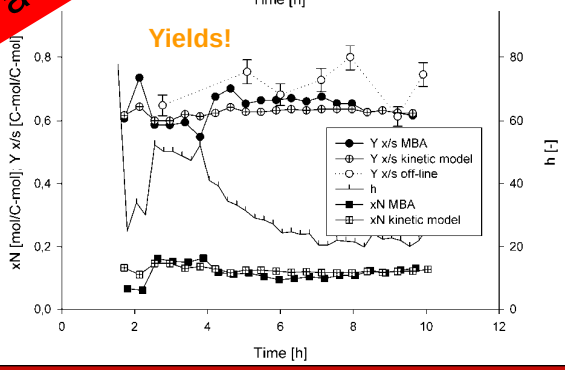
- Determined Parameters
  - μ (t), q_{smax} (t),
  - $Y_{x/s}$ (t), reconciled
  - s (t), x (t)
  - Statistically validated

P. Wechselberger, A.Seifert,C. Herwig, PAT method to gather bioprocess parameters in real-time using simple input variables and first principle relationships, 2010, Chemical Engineering Science

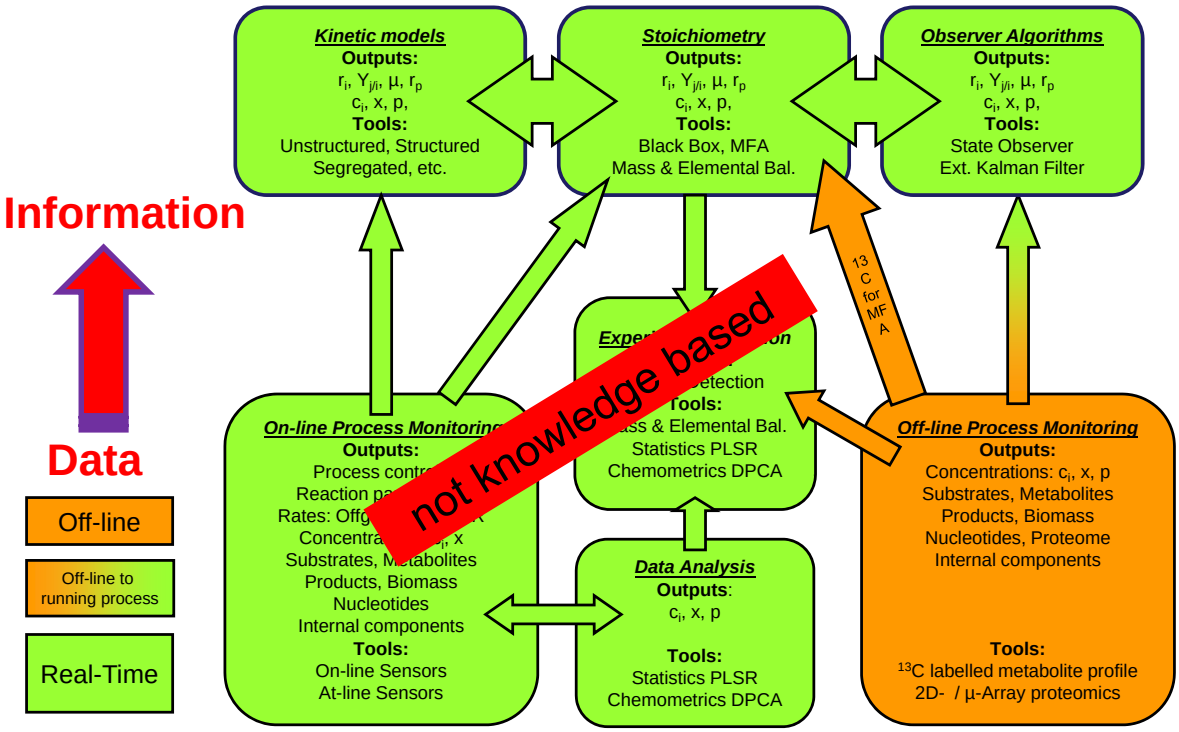




Consistency check of the model fit is a must

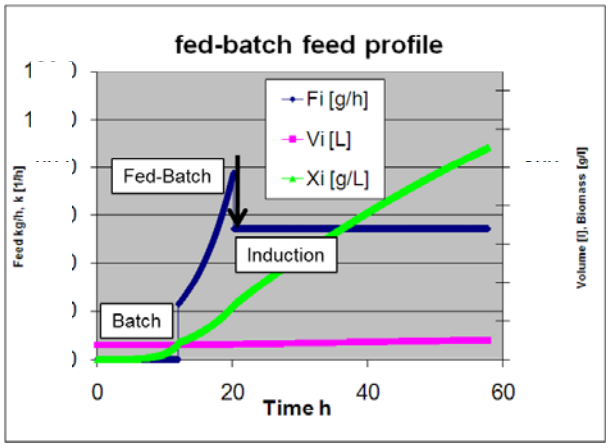


Real-time Hybrid Exploitation based on first principle relationships

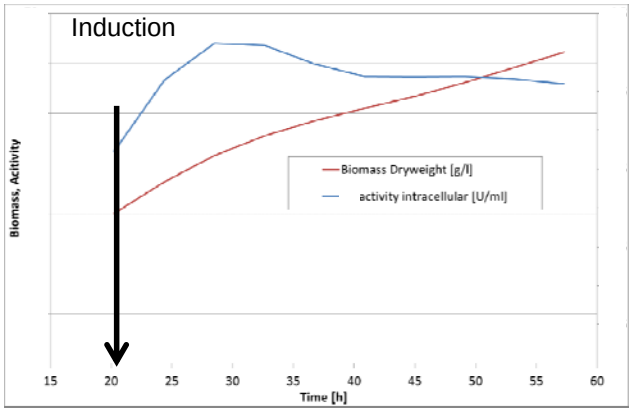


not knowledge based

Model System
recombinant *E. coli* process



Effect of feed profile on productivity / enzyme activity?

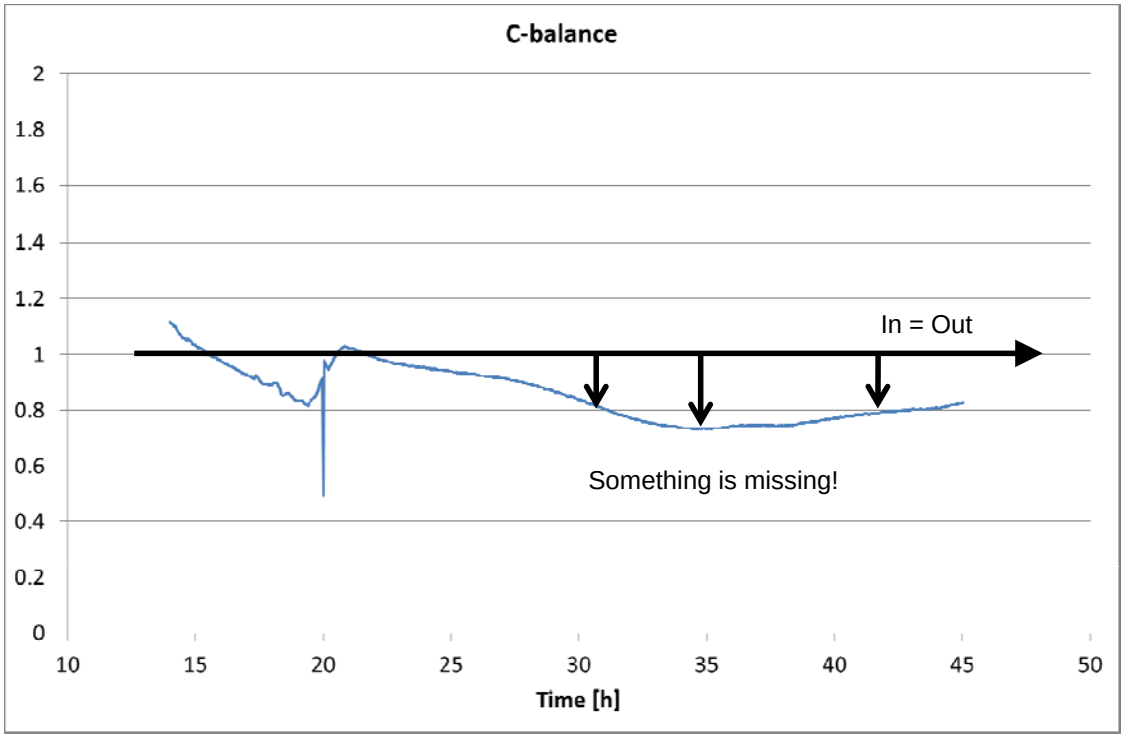


Standard feed profile

Exponential fed-batch to produce biomass

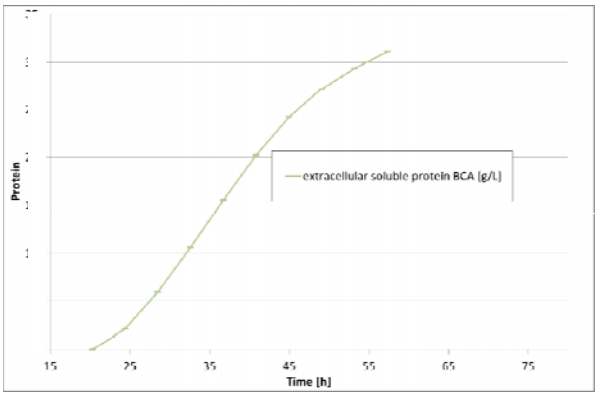
Linear feed induction phase

Check of performance information

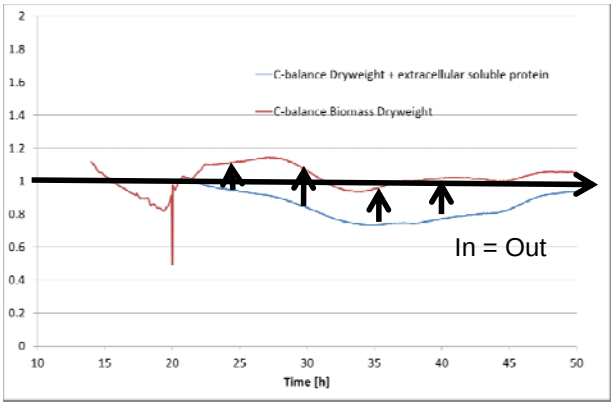


Model System
recombinant *E. coli* process

- BCA-Assay reveals: release of extracellular protein



Protein missing on dryweight biomass!



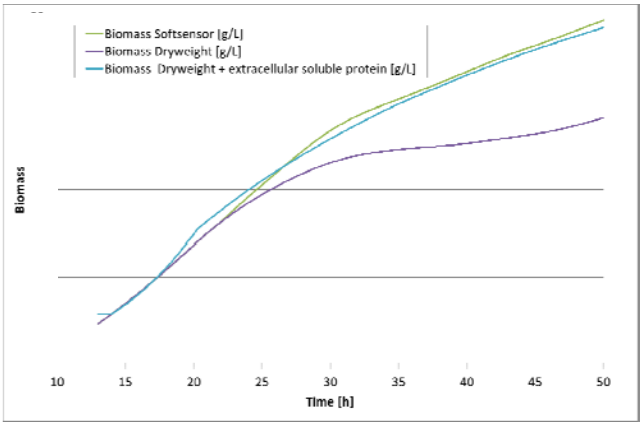
Balances – valuable tool to verify measurements!

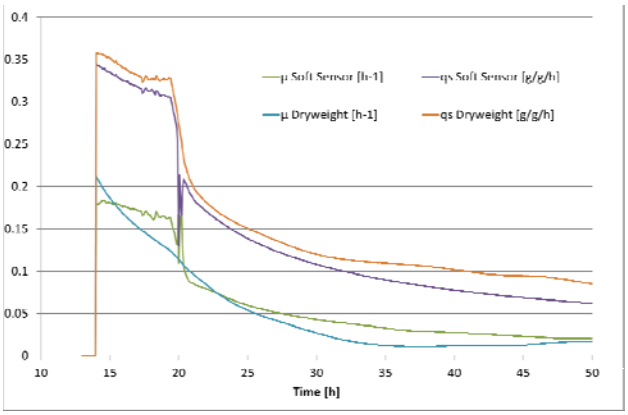
Gross errors?

Incorrect growth stochiometry?

Real-time Exploitation
Biomass and specific rate estimation

- Soft-Sensor fits Biomass + extracellular soluble protein!

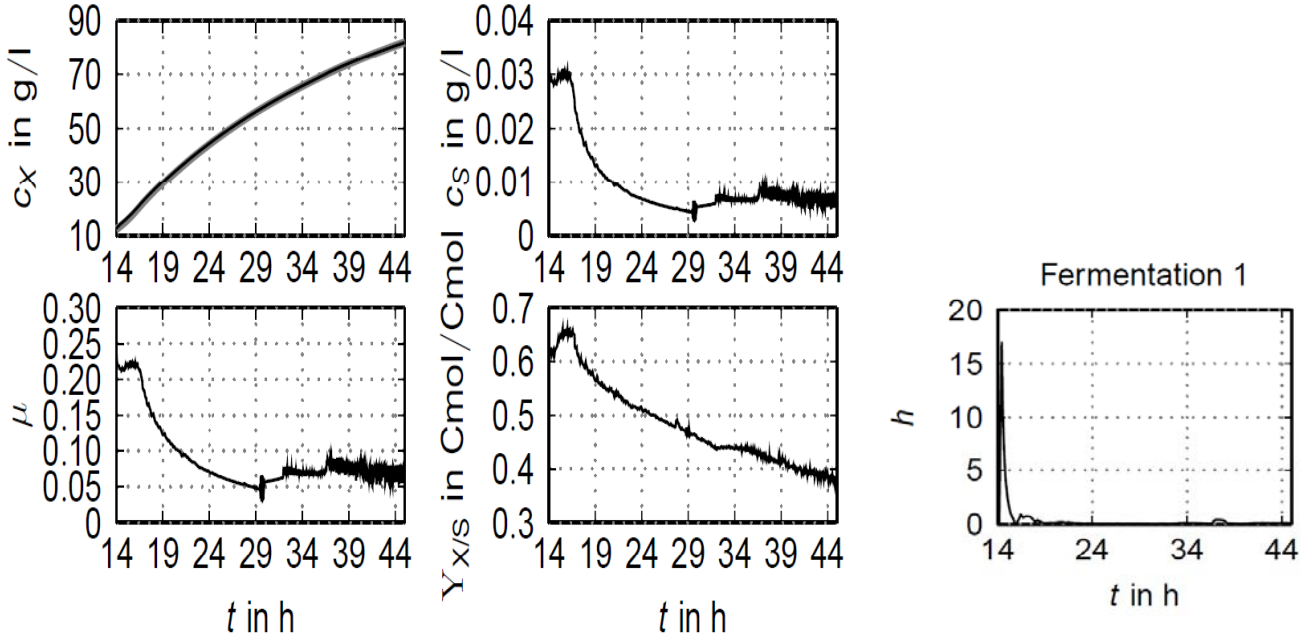




Soft-sensor specific rates well in accordance with off-line specific rates

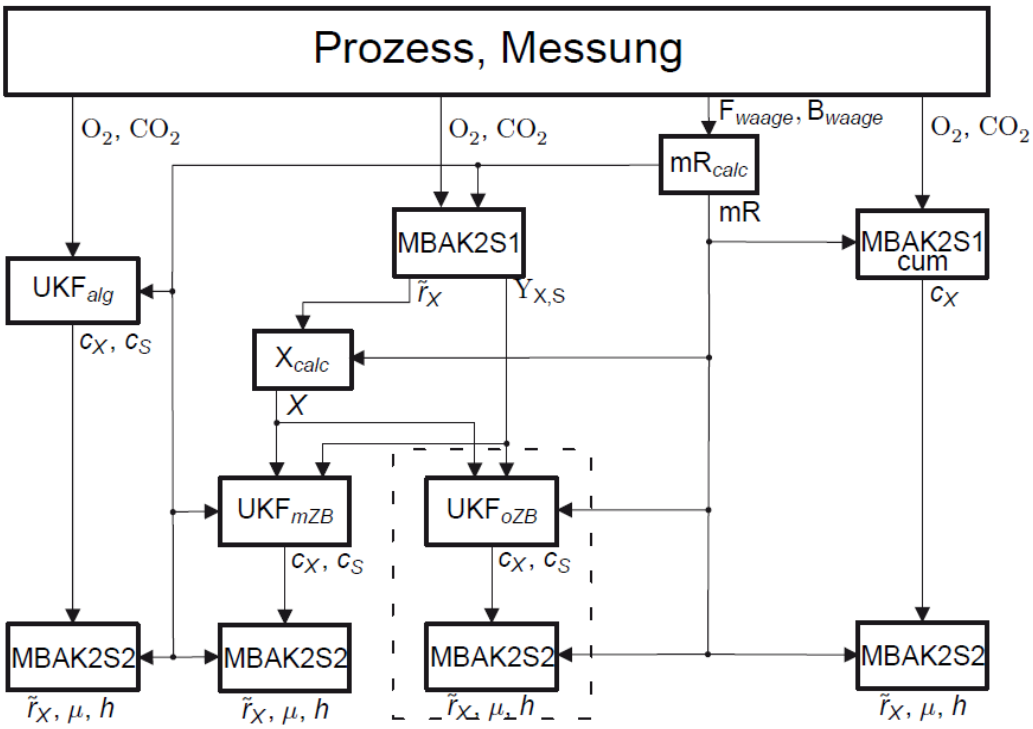
Even better!

Use of Unscented Kalman Filter



- Quality check via statistical test OK

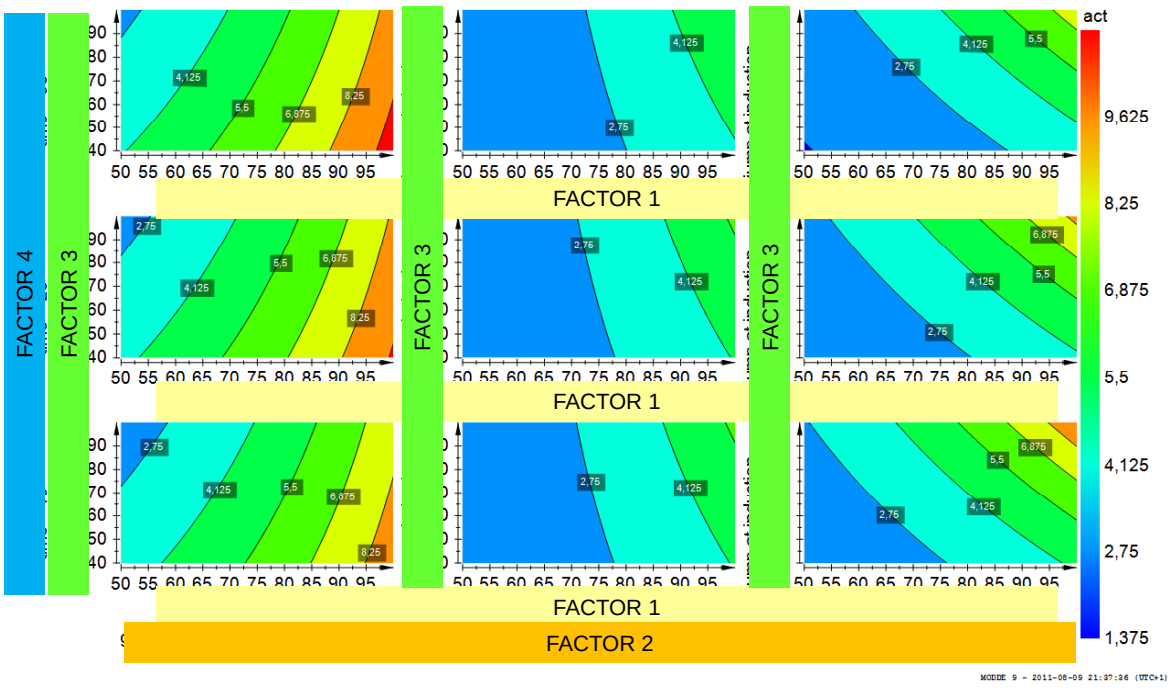
Implemented Softsensor Hybrids



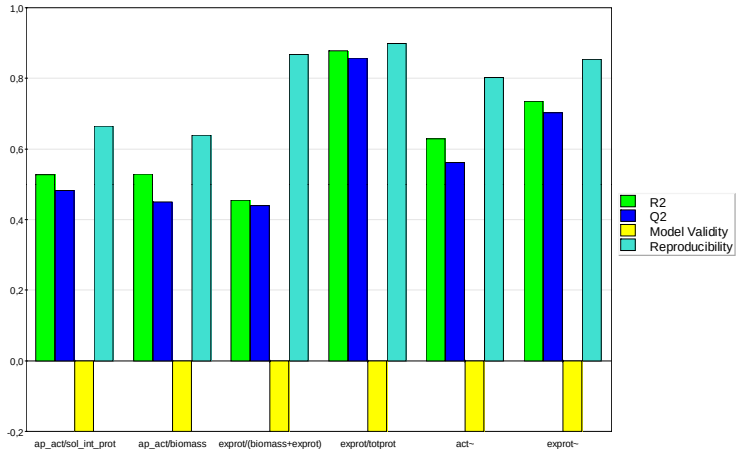
EXTRACTION OF KNOWLEDGE

DoE: Relationship of CPPs to CQA (i.e. volumetric activity) (1)

- Is the quality of this multivariate relationship OK?



Model statistics (1)

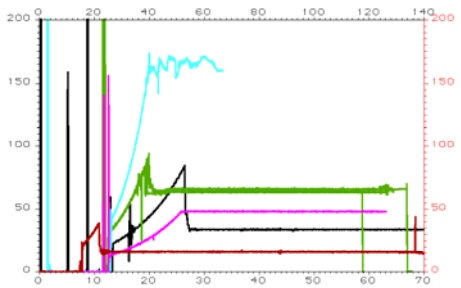


Model Validity for all models very low;
this usually means the model is not adequate.
Model error can not be accounted to the replicate error solely.

→ Search for factors which explain the variation better.

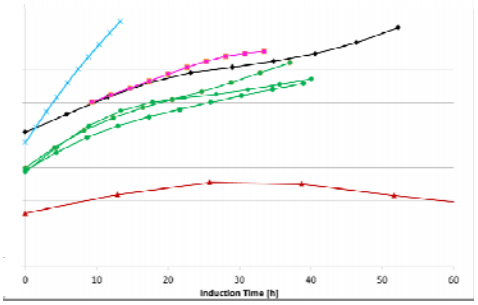
Use of specific entities for knowledge extraction

Feed-Profiles

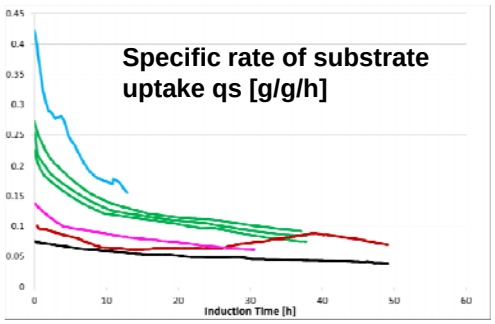


+

Biomass [g/L]



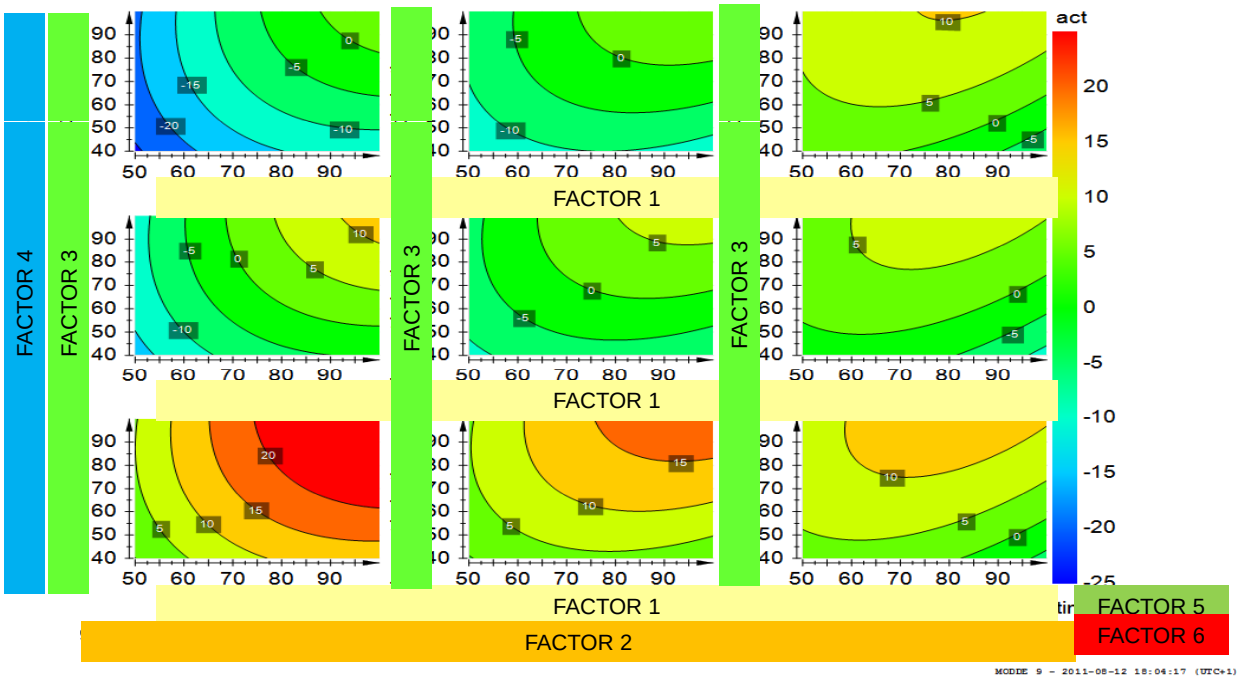
=



- Combine raw data to scalable entities

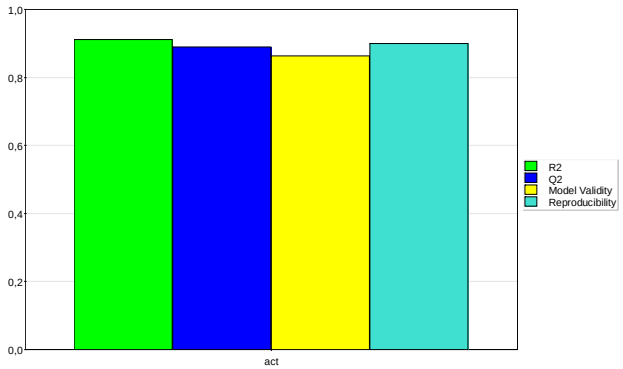
DoE: Relationship of CPPs to CQA (i.e. volumetric activity) (2)

- Addition of two further factors results in different regression pattern



Model statistics (2)

- Additional factors identified by analyzing specific parameters
- Much better model statistics with new model, which can explain the variability of the data



- Reduce number of model factors when using specific parameters directly in the DoE!

PAT in it‘s original sense

Science based not knowlegde based

Little to no a priori knowlegde
Stoichiometric relationsships

First principle relationships:
Elemental Balances

Simple, well establish measurement instruments
Gravimetric measurement, gas analysis
Very accurate measurements

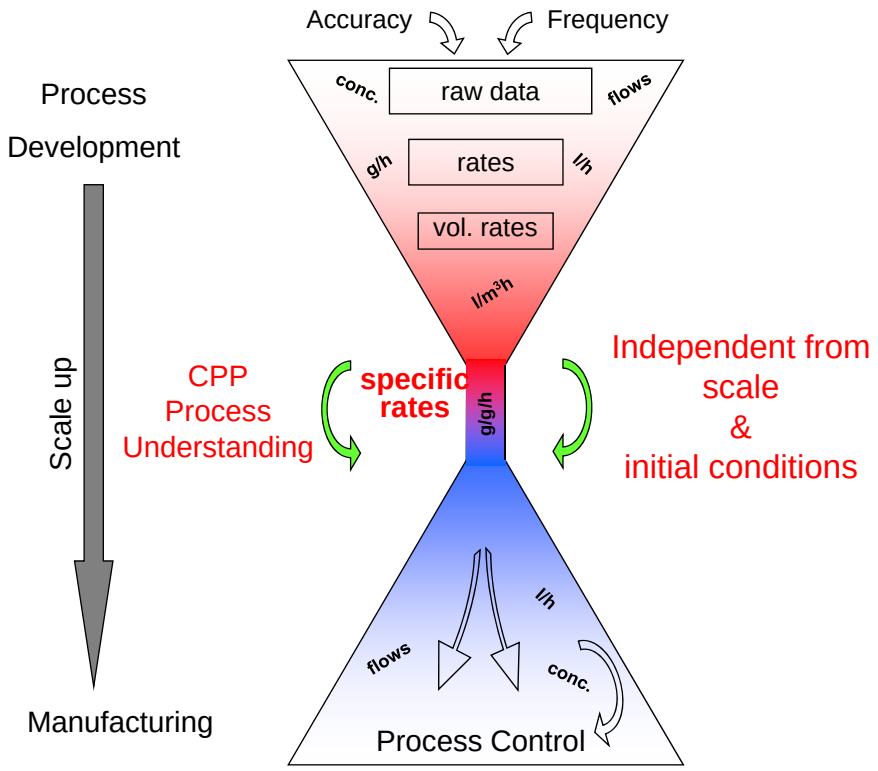
real-time

Real-time validity check
Statistical test

Process variables
Rates, Yields, Coefficients

Process understanding
Strain characterisation

Strategy to transfer data from Process Development to Manufacturing



Used Model Systems and status of method imlementation

- Suspended Microbial Cultures
  - Defined media
  - High Density
  - Complex media
- Strict Anaerobic Cultures
- Adherent Microbial Cultures
  - Influence of Morphology
- Suspended Mammalian Cultures
  - Validate methods off-line
  - Run methods on-line

Take Home Messages (1)

- Data are sufficiently available and need robust methods for **automated extraction of information**
  - **Real time extraction of knowledge?**
- **No knowledge without quality assessment** of data, information and knowledge
- Presented tools need only few prior knowledge: applicable to process development tasks
- Process development based on specific rates allows scalability and enhanced knowledge extraction capability
- Use of softsensors allow to use these tools in early quantitative strain screening
- Generic approach inside of QbD

Thank you for your attention!

Questions?

Member of

Univ.Prof. Dr. Christoph Herwig
Vienna University of Technology
Research Division Biochemical Engineering
Gumpendorferstrasse 1a/ 166 A-1060 Wien
Austria
emailto: christoph.herwig@tuwien.ac.at
Tel (Office): +43 1 58801 166400
Tel (Mobile): +43 676 47 37 217
Fax: +43 1 58801 166980
URL : http://www.vt.tuwien.ac.at/division/index.php?division_id=6