



Application of a Mechanistic Model as a Tool for On-line Monitoring of Pilot Scale Filamentous Fungal Fermentation Processes - The Importance of Evaporation Effects
Mechanistic model for pilot scale monitoring

Mears, Lisa; Stocks, Stuart M.; Albæk, Mads Orla; Sin, Gürkan; Gernaey, Krist V.

Published in:
Biotechnology and Bioengineering

Link to article, DOI:
[10.1002/bit.26187](https://doi.org/10.1002/bit.26187)

Publication date:
2017

Document Version
Peer reviewed version

[Link back to DTU Orbit](#)

Citation (APA):
Mears, L., Stocks, S. M., Albæk, M. O., Sin, G., & Gernaey, K. V. (2017). Application of a Mechanistic Model as a Tool for On-line Monitoring of Pilot Scale Filamentous Fungal Fermentation Processes - The Importance of Evaporation Effects: Mechanistic model for pilot scale monitoring. *Biotechnology and Bioengineering*, 114(3), 589–599. <https://doi.org/10.1002/bit.26187>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Bioprocess Engineering and Supporting Technologies **Biotechnology and Bioengineering**
DOI 10.1002/bit.26187

Application of a Mechanistic Model as a Tool for On-line Monitoring of Pilot Scale

Filamentous Fungal Fermentation Processes- The Importance of Evaporation Effects[†]

Running title: Mechanistic model for pilot scale monitoring

Lisa Mears¹, Stuart M. Stocks², Mads O. Albaek², Gürkan Sin¹, Krist V. Gernaey¹

1. CAPEC-PROCESS Research Centre, Department of Chemical and Biochemical Engineering, Technical University of Denmark, 2800 Lyngby, Denmark
2. Fermentation Pilot Plant, Novozymes A/S, 2800 Bagsværd, Denmark

Krist V. Gernaey

CAPEC-PROCESS Research Centre,

Department of Chemical and Biochemical Engineering

Technical University of Denmark

2800 Lyngby, Denmark

kvg@kt.dtu.dk

+45 45252970

[†]This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: [10.1002/bit.26187]

Additional Supporting Information may be found in the online version of this article.

This article is protected by copyright. All rights reserved

Received April 18, 2016; Revision Received August 17, 2016; Accepted September 16, 2016

This article is protected by copyright. All rights reserved

Abstract

A mechanistic model-based soft sensor is developed and validated for 550L filamentous fungus fermentations operated at Novozymes A/S. The soft sensor is comprised of a parameter estimation block based on a stoichiometric balance, coupled to a dynamic process model. The on-line parameter estimation block models the changing rates of formation of product, biomass, and water, and the rate of consumption of feed using standard, available on-line measurements. This parameter estimation block, is coupled to a mechanistic process model, which solves the current states of biomass, product, substrate, dissolved oxygen and mass, as well as other process parameters including k_{La} , viscosity and partial pressure of CO_2 . State estimation at this scale requires a robust mass model including evaporation, which is a factor not often considered at smaller scales of operation.

The model is developed using a historical dataset of eleven batches from the fermentation pilot plant (550L) at Novozymes A/S. The model is then implemented on-line in 550L fermentation processes operated at Novozymes A/S in order to validate the state estimator model on fourteen new batches utilizing a new strain. The product concentration in the validation batches was predicted with an average root mean sum of squared error (RMSSE) of 16.6%. In addition, calculation of the Janus coefficient for the validation batches shows a suitably calibrated model. The robustness of the model prediction is assessed with respect to the accuracy of the input data. Parameter estimation uncertainty is also carried out. The application of this on-line state estimator allows for on-line monitoring of pilot scale batches, including real-time estimates of multiple parameters which are not able to be monitored on-line. With successful application of a soft sensor at this scale, this allows for improved process monitoring, as well as opening up further possibilities for on-line control algorithms, utilizing these on-line model outputs. This article is protected by copyright. All rights reserved

Keywords: fermentation, modelling, monitoring, pilot scale, filamentous fungus, soft sensor

Introduction

Pilot scale fermentation activities surround the development of fermentation processes which are applicable to industrial scale. The aim is to develop understanding surrounding a strain and a process, in order to optimize the system, and facilitate the translation of this knowledge into a high productivity production process. This requires fast and efficient process development and an understanding of the system. Bioprocess monitoring and control are therefore two key areas of interest at this scale of operation.

Bioprocess monitoring

Monitoring of bioprocesses requires reliable on-line measurements, however there is a lack of on-line sensors for key parameters of interest in the field, such as substrate, product and biomass concentration (Alves-Rausch et al., 2014; Assis and Filho, 2000; Sonnleitner, 2013), which are applicable to pilot and production scale. There are challenges specific to the development of in-line sensors for industrial fermentation systems. These include the need for the probe to be robust to sterilization, and to be stable over long operation times (Alford, 2006). In addition, there is also the issue of regulation, and the need to obtain approval for changes made to the hardware used in a process operating under good manufacturing practices (GMP) (Gernaey, 2015). A practical issue is also the limited number of ports for in-line probes on the stainless steel vessels. This lack of on-line state measurement limits the ability to monitor the progression of fermentation systems.

There are many causes of batch-to-batch variation in biological systems based on physiological differences, metabolic shifts, or disturbances, for example small differences in raw materials (Villadsen et al., 2011). If on-line estimates of key performance indicators, such as product concentration, were available, it allows for better process monitoring, and also allows for implementation of advanced control. With the process analytical technology (PAT)

guidelines being introduced (FDA, 2004), there is also an additional drive towards increasing process knowledge and monitoring capabilities (Gernaey et al., 2010).

Soft sensors applied to fermentation systems

Due to the common limitations in process monitoring, there is an interest in soft sensors, which utilize on-line measured variables to predict the unknown states in real-time (Luttmann et al., 2012; Sagmeister et al., 2013). However, a report by Luttmann et al. (2012) states that there is limited application of soft sensors in industry, despite the advantage of real-time process understanding, and the fact that there is no need for investment in additional hardware. Some of the challenges associated with implementation of a soft sensor in an industrial context are cited in the report. These include data availability for model development at industrial scales, and data quality, which may contain outliers or have issues with sensor drift. Another challenge is the requirement for additional computer hardware where there is connectivity to the on-line measured data. Finally, the need for re-calibration of the model is perceived as limiting the practical applicability. These factors must be considered in the development of a soft sensor if it is to be applied in practice.

State estimator models may be developed based on first principle understanding, data-driven methods or by hybrid modelling. Data driven methods do not require an understanding of the system, and may therefore be considered a faster approach to model development. Data-driven methods may, for example, be based upon artificial neural networks (Chen et al., 2004; Linko et al., 1999), fuzzy logic (Araújo-Bravo et al., 2004; Luttmann et al., 2012), or multivariate statistical modelling approaches (Yuan et al., 2014; Zhang and Lennox, 2004). These methods have the disadvantage that they are unreliable when extrapolating outside the range of the data used to develop the model. For production scales this may be less of a problem, where process conditions are generally defined. For application at pilot scale however, this is undesirable since new processing conditions are investigated. The

development of a data-driven model also provides little insight into the process, since the model parameters have no physical meaning. Hybrid modelling approaches combine some of the benefits of data driven models, with the more robust basis of a mechanistic model. It is not necessary to have a full understanding of the process of interest and therefore the model development is faster, and the resulting model may have enhanced extrapolation capabilities (von Stosch et al., 2014).

In contrast, first principle soft sensor models are based on a fundamental understanding of the system (Sagmeister et al., 2013; Sundström and Enfors, 2008). Their development is based on existing process understanding, and the model parameters have a physical meaning. Parameter values therefore provide information about the process of interest, for example yield coefficients, which may be used to compare between strains. Soft sensor models incorporating stoichiometric balances are valuable, since they are scale independent, and based on the fundamental biochemical reactions. Since it is observed that yield coefficients change over the course of a fermentation (Golabgir et al., 2015; Jenzsch et al., 2006), this method is interesting as it allows for this adaption within a mechanistic model structure. In addition, stoichiometric balances utilize flow rates as input variables, which then avoids the need for in-line probe measurements, which may be highly dependent on their position in the vessel, and the calibration accuracy. The mechanistic model approach also has the benefit of being more generic to new processes which may require a small adaption to the model, but in general should be applicable to different strains and processes. The method is equally relevant for fungal systems such as in this work, and also bacterial systems (Sagmeister et al., 2013; Sundström and Enfors, 2008).

This work discusses the application of a first principle soft sensor model to pilot scale (550L) filamentous fungal fermentation systems operated at Novozymes A/S. The model comprises of an on-line parameter estimation block, coupled to a dynamic model of the system. The

parameter estimation block is based on a stoichiometric balance, where the current rates of product and biomass and water formation and feed consumption are identified from available off gas measurements and ammonia addition. This parameter estimate is then used as an input to a mechanistic process model, which describes the mass transfer capabilities of the system based on the operating conditions, including stirrer speed, aeration rate, headspace pressure and temperature. The model is developed and calibrated using a historical pilot scale data set as described by Albaek et al. (2011). The model is then implemented at the fermentation pilot plant of Novozymes A/S and validated on-line using fourteen new batches at the same 550L pilot scale, but utilizing a different host strain and product. With implementation of a robust soft sensor, it is possible to incorporate the state estimate into a control structure, and open up possibilities for more advanced process control.

Materials and methods

A model is developed to describe a filamentous fungal fermentation process operated at Novozymes A/S. The process is operated as fed-batch, in pilot scale fermenters, as described by Albaek et al (2011). The model structure is shown in Figure 1.

Parameter estimation

The parameter estimation is based on stoichiometric balances using reliable on-line measurements of the carbon evolution rate (q_c), oxygen uptake rate (q_o) and ammonia addition rate (q_n), as shown in equation 1. These measurements were chosen as they are not subject to deviations between batches, as would be the case for substrate feed rate for example, which has batch-to-batch variations in concentration. A stoichiometric balance is implemented in MATLAB (MATLAB, 2013) to solve the unknown rate parameters based on equation 2, where E refers to the matrix of stoichiometric coefficients for C, H, N and O, and the subscripts refer to measured and calculated values. The concept of using stoichiometric balances to solve formation rates is well documented in literature (Fordyce et al., 1990; Van

Der Heijden et al., 1994). In this case the production of by-product is neglected, and assumed to be negligible.

$$q_g \text{CH}_2\text{O} + q_o \text{O}_2 + q_n \text{NH}_3 = q_x \text{CH}_{\text{XH}}\text{O}_{\text{XO}}\text{N}_{\text{XN}} + q_p \text{CH}_{\text{PH}}\text{O}_{\text{PO}}\text{N}_{\text{PN}} + q_c \text{CO}_2 + q_h \text{H}_2\text{O} \quad (1)$$

$$E_{\text{meas}} q_{\text{meas}} + E_{\text{cal}} q_{\text{cal}} = 0 \quad (2)$$

Dynamic model

A dynamic process model describes both the evolution of the system states, and additional physical process parameters. All states are defined on a mass basis, rather than a volume basis, and this factor is considered important to the model accuracy at this scale. Modern processes in industrial biotechnology cannot be treated as dilute systems. Fungal fed batch systems are reported to produce up to 30g/L biomass (Riley et al., 2000; Tolan and Foody, 1999), and may also produce 100g/L product, in the example of *Trichoderma reesei* producing cellulases (Cherry and Fidantsef, 2003; Schuster and Schmoll, 2010). In addition, gas hold up changes with time in a process (Hofmeister, 1988). These factors make estimates of density and volume somewhat difficult; with densities in the range 1.05 to 1.3 kg/L there is the potential for errors in concentration in the range 5 to 30%. This means that estimates of state based on volume measures can only be validated by accurate measurements of broth density or gassed hold up, the latter being far from trivial. Based on these factors it is considered that modelling of concentrations on a unit mass basis instead of volume is a more accurate approach, as it is independent of system density or gas hold up.

In equations 3 and 4, the relevant formation rates from the parameter estimation are used to solve for the current concentrations in the system, accounting also for the changing mass in the fed-batch system. This method avoids the need for fixed model yield parameters or growth rates, which are typically incorporated in unstructured fermentation models.

Similarly, the substrate concentration is calculated based on the feed added, and the estimated consumption rate from the state estimator. This requires a known substrate concentration in the feed.

The mass prediction accounts for the feed added, as well as the evaporation rates, which are significant in a pilot scale system of over 0.5m³. The rate of evaporation, F_{evap} , is calculated as shown in equation 8, where the saturation pressures at the inlet (P_{in}^*) and outlet (P_{out}^*) are calculated as in Bolton (1980), where P_{in}^* is then a function of the air humidity and air temperature (Bolton, 1980). For a given batch, the evaporation rate is therefore dependent on the air humidity and air temperature, as well as the processing conditions for aeration rate, headspace pressure and temperature. For batches where multiple samples are taken over the course of the batch, the volume of sample removed for analytical purposes should also be included in the mass balance.

In order to calculate the dissolved oxygen concentration, a log-mean concentration difference is used in order to define the driving force for oxygen transfer (Doran, 2013).

The driving force and the oxygen mass transfer coefficient, $k_L a$, define the oxygen transfer rate. There are a range of correlations available in literature for the calculation of the oxygen mass transfer coefficient (Garcia-Ochoa and Gomez, 2009). In this work the correlation shown in equation 9 is applied, as it has been shown previously in Albæk et al. (2011) to describe the pilot scale data set well. In equation 9, P_{total} accounts for power input by both mixing and power dissipation due to aeration. The apparent viscosity used in equation 9 is calculated as described in Albaek et al. (2012), where the values of the constants are also provided. The partial pressure of CO₂ is also calculated from the estimated CO₂ molar fraction in the outlet gas and the headspace pressure.

$$\frac{dX}{dt} = \frac{(q_x M r_x)}{M} - X \frac{dM}{dt} \quad (3)$$

$$\frac{dP}{dt} = \frac{(\mathbf{q}_p \text{Mr}_p)}{M} - P \frac{dM}{dt} \quad (4)$$

$$\frac{dG}{dt} = \frac{(S_f F_{\text{feed}} - \mathbf{q}_g \text{Mr}_g)}{M} - G \frac{dM}{dt} \quad (5)$$

$$\frac{dM}{dt} = F_{\text{feed}} - F_{\text{evap}} - F_{\text{sample}} + 0.001 \mathbf{q}_o \text{Mr}_o - 0.001 \mathbf{q}_c \text{Mr}_c \quad (6)$$

$$\frac{dDO}{dt} = k_L a \frac{(DO_{in}^* - DO) - (DO_{out}^* - DO)}{\log \frac{(DO_{in}^* - DO)}{(DO_{out}^* - DO)}} - \mathbf{q}_o - DO \frac{dM}{dt} \quad (7)$$

$$F_{\text{evap}} = \left[\left(\frac{P_{\text{out}}^* \text{Mr}_h}{RT} \right) \left(\frac{P_{\text{atm}}}{P_{\text{atm}} + \text{HSP}} \right) - \left(\frac{P_{\text{in}}^* \text{Mr}_h}{RT} \right) \right] Q_{\text{air}} \quad (8)$$

$$k_L a = C \left(\frac{P_{\text{total}}}{v_L} \right)^a v_g^b \mu_{\text{app}}^c \quad (9)$$

$$pp_{\text{CO}_2} = \text{HSP } CO_{2\text{out}} \quad (10)$$

Implementation algorithm

Figure 2 shows a detailed description of the online implementation algorithm. The on-line data is read directly from the OPC server every 30 seconds using MATLAB® timer objects.

The parameters are updated every 5 minutes, using the input data which is simply averaged, assuming Gaussian noise. The updated parameters are input to the dynamic model which is then also solved using ode23s in Matlab over the 5 minute window. The new state estimate is then overwritten as the initial condition for the next modelling iteration. This results in a low computational demand, by only modelling at five minute intervals. The user may plot the results at any time, and the results will be updated to within a five minute sampling interval.

For model calibration, a data set was used whereby the online data was available at one hour

intervals. This means that the same implementation algorithm was applied, except for the parameter update occurring once every hour instead of once every five minutes.

Off-line sample analysis

The biomass concentration is measured by dry mass determination, by drying at 105°C for 48 hours. The biomass sample is washed twice with deionised water to remove soluble media components. The product concentration is determined based on a generic protein assay used at Novozymes A/S. Viscosity is measured off-line in an AR-G2 rheometer from TA instruments using a vane-and-cup geometry. The vane consists of four blades at right angles (14 mm x 42 mm), the cup had a 15 mm radius, and the gap between the vane and cup was 4000 µm. Measurements were taken in the interval of 10 to 600 1/s and the bingham plastic model was applied to describe the rheological behaviour (Bingham, 1916). The shear rate for apparent viscosity determination is found by the approach of Metzner and Otto, $k_s N$, where k_s is 11 (Metzner and Otto, 1957). The $k_L a$ is determined by the direct method (Villadsen et al., 2011) also assuming a log mean driving force, as described in equation 7.

The historic data set

The stoichiometric model parameters (XH, XO, XN, PH, PO, PN) are obtained by least square fitting to a historical data set of eleven batches, where biomass concentration, product concentration, dissolved oxygen concentration and mass are fitted. This dataset is designed as a full factorial design including two levels for three process variables, namely specific power input (1.5-15kW/m³), aeration rate (96-320NL/min) and headspace pressure (0.1-1.3 bar) as described by Albaek et al. (2011). The processing conditions affect the biomass concentration and product concentration achievable, due to oxygen mass transfer limitations. This makes this dataset ideal for calibration of the model parameters, as there is significant deviation in these states between the batches.

The on-line process

The model is then validated on-line for fourteen new batches. These batches are also operated at different stirrer speeds (200-400 rpm), aeration rate (200-400NL/min) and headspace pressure (0.7-1.3 bar) which is the reason for the different biomass and product concentrations achieved. In this newer process, media optimisation has led to the inclusion of partially soluble compounds in the media. Due to the solid content in the media, it has not been possible to quantify biomass concentration in these batches. Off-line analysis includes product concentration, off-line viscosity, and $k_L a$.

Statistical assessment of model fit

$$RMSSE (\%) = \frac{\sqrt{\frac{1}{n} \sum_i^n (y_{meas,i} - \hat{y}_i)^2}}{\frac{1}{n} \sum_i^n (y_{meas,i})} \quad (11)$$

$$J^2 = \frac{\frac{1}{n_{val}} \sum_i^{n_{val}} (y_{meas,i} - \hat{y}_i)^2}{\frac{1}{n_{cal}} \sum_i^{n_{cal}} (y_{meas,i} - \hat{y}_i)^2} \quad (12)$$

When assessing the goodness of fit for the model, the root mean sum of squared errors (RMSSE) is applied, as defined in equation 11, where $y_{meas,i}$ is one of, n , measurement points, and \hat{y} is a model prediction of the same variable. This is expressed as a percentage of the average measured value, for confidentiality reasons. In addition, to assess the validation model fit compared to the calibration model fit, the Janus coefficient, J^2 , is used, as discussed in Sin et al. (2008). In this work, the model prediction of product concentration is the primary focus, as this is a key process performance indicator which otherwise takes considerable time to obtain analytically.

Parameter estimation uncertainty

In order to conclude on the uncertainty of the fitted stoichiometric parameter values, bootstrap sampling is applied (Efron, 1979). In this method, the residuals between the model and the

data are sampled in order to create simulated data sets for fitting. The errors are sampled from the biomass concentration, the product concentration, the dissolved oxygen concentration and the mass. The method for sampling is random sampling of 100 residuals with replacement. By fitting to each of the simulated data sets, a distribution of parameter values is identified, which provides an indication of the parameter uncertainty. The parameter uncertainty in this work is provided as two standard deviations as a percentage of the mean.

Results and discussion

State estimator development based on historic dataset

The soft sensor model is applied off-line to an eleven batch pilot scale data set, as described in the Materials and Methods. The purpose is to fit the model parameters for the stoichiometric model (XH, XO, XN, PH, PO, PN). The measured on-line data for carbon evolution rate (q_c), oxygen uptake rate (q_o) and ammonia addition rate (q_n) are used as input to the parameter estimation block in order to simulate the system as would be done on-line. The parameter update occurs every hour.

Figure 3 shows the results of the dynamic model for one batch of data, where the model fit is representative of all eleven batches. This batch was chosen as it shows measured data which has previously been published (Albaek et al., 2011). There is a very good agreement between the model prediction and the measured data for all variables. The dynamics in the dissolved oxygen profile as shown in Figure 3 are due to the oxygen uptake rate. This measured oxygen uptake rate has fluctuations corresponding to the feed rate applied, however in the real system this direct impact is not seen. This suggests a limitation in the model description which should be considered when the model is applied. In addition, the final substrate concentration is seen to go negative, which may be explained by uncertainty in the initial substrate concentration in the batch phase, variations in the feed concentration between batches, as well as consumption of additional media components which are not considered in the stoichiometric balance

methods. This is a limitation in the model, as of course this is not physically possible. The consumption rate of substrate is independent of the substrate concentration, as it is entirely based on the stoichiometric balance. The substrate balance is certainly an area of future work for the model. Overall, the qualitative trends are captured well in the model, and the prediction accuracy is considered acceptable.

The product prediction for all eleven batches is shown in Figure 4. Across the eleven batches, the mean deviation between the model prediction and the data for product concentration ranges from 4.3% to 26.2%, and a summary of the results for all 11 batches is shown in Table I. In addition, Figure 5 shows the biomass concentration model prediction for the same eleven batches. The mean deviation between the model prediction and the data for biomass concentration ranges from 6.2% to 21.8%. The model results are considered robust to the different operating conditions, and the accuracy of the model prediction across the batches operated in such different physical conditions is very encouraging for future application as a process monitoring tool.

State estimator applied on-line in Novozymes A/S fermentation pilot plant

Once the model parameters have been fitted to the calibration batches, it is possible to implement the model on-line. Focussing on the product concentration as the fundamental measure of process performance, it is shown that the model predicts the current product concentration well, with an average RMSSE of 16.6% in fourteen new validation batches. Table I summarises the assessment of model fit for the validation batches. It is seen from Table I that the Janus coefficient is close to one for all batches, which suggests that the model accuracy in the validation batches is comparable to that of the calibration batches, with an average Janus coefficient of 1.5. Batches 8 and 11 show the greatest Janus coefficient, due to a significant underprediction of the product concentration at the end of the batch. The high Janus coefficient in Batch 6 may be partly due to measurement error, as it is seen one data

point does not follow the trend. The greatest RMSSE is in Batch 2. Although the absolute error of the prediction is comparable to the other batch predictions as shown by the Janus coefficient, the percentage error is greater due to the lower average product concentration. This shows the importance of utilising more than one assessment of model fit, where in this case the Janus coefficient shows that the model fit is no worse based on absolute errors.

Overall the results show that the model provides acceptable prediction accuracy of product concentration, and that the model is calibrated successfully, such that the model errors seen in the validation batches are of a similar degree to those obtained from the calibration set. An on-line measure of product concentration is a valuable monitoring parameter which allows operators to compare between batches on-line, and intervene in the case of poor performance. The ultimate goal of introducing soft sensors is to incorporate the knowledge obtained from the state estimation into an on-line control algorithm.

The complete model output is shown in Figure 7 for one of the fourteen additional batches, which was selected since the results are comparable to the results reported in the calibration batch. This full model output provides important monitoring information to operators, in particular the viscosity measurements, which are of high importance in industrial filamentous fungal processes (Olsvik and Kristiansen, 1994).

It is also important to note that the mass prediction model is accurate, utilising the evaporation term in the model. This is an important addition to a pilot scale fermentation model, which is not often discussed in literature focussed on smaller scale experimentation. The evaporation rates calculated are significant, when considering fungal fermentation processes are operated for approximately one week. At the maximum rate of 40 kg/week, this equates to roughly 10% of the final mass at this scale over a week operation. The complication of implementing an evaporation model is that it depends on the relative humidity and temperature of the incoming air, as this defines the water content incoming to the system, in addition to the air

flow rate, headspace pressure and temperature of the system, which defines the rate of water stripping. The environmental conditions are seen to vary significantly over a year period, as shown in Figure 8, and this fact must be accounted for in the model. Currently this is done by manually changing air inlet conditions, but this could become integrated as a measured parameter also, which is part of the future work for this model.

It is also seen in Figure 8 that the temperature of the air and the humidity of the air are correlated, with a correlation coefficient of -0.53. With this understanding of the expected variation in inlet conditions, it is possible to simulate the range of expected evaporation rates over a year period. Latin Hypercube Sampling (Helton and Davis, 2003) is used to simulate 250 data points which represent this year of data, taking into consideration the correlation in input variables using the method of Iman and Conover (Iman and Conover, 1982; Sin et al., 2009). These 250 sample points are then simulated by Monte Carlo simulations in order to show the variation in evaporation rate, for a fixed set of processing conditions, as shown in Figure 9.

This shows the importance of accounting for evaporation in this pilot scale model, and further work will include validating the different evaporation rates over the year by applying the state estimator and assessing the mass prediction.

Model uncertainty assessment

In order to understand the uncertainty of the model prediction we assess two scenarios: The effect of measurement quality on the model prediction, and an assessment of model parameter uncertainty.

Since this modelling method relies on process measurements as input parameters it is important to understand how the accuracy of these measurements affects the model prediction. In Table II, a constant 5% deviation is applied to the input measurements, and the effect on the model output is shown, for the same batch as shown in Figure 7. Table II shows

the results for the relative change in the final model prediction compared to the base case for four scenarios; 5% increases are applied in each of the input parameters individually, as well as a simultaneous 5% error in both the CER and OUR. This is also assessed since these off gas measurements are utilizing the same equipment, and therefore it is likely that there could be an equal error in both measurements. This may for example be due to an incorrect gas flow rate measurement, which affects both the OUR and CER calculation equally. The results show that the balance between the OUR and the CER is very important for the prediction, meaning that if there is an error in both CER and OUR, then the prediction is not significantly affected, however an error in only one means that unacceptable errors are observed. This shows the importance of the respiratory quotient ($RQ=CER/OUR$) for the model prediction, especially for the biomass prediction. The ammonia flow rate affects both the biomass and product predictions by less than 10%.

In addition to the model sensitivity to the input measured data, it is also important to assess the confidence in the fitted parameter values. In this case, the bootstrap method was chosen in order to provide an indication of model parameter uncertainty. The parameter uncertainty can be analysed by the distribution of the parameter values, where in this case we consider two standard deviations as a percentage of the mean, for confidentiality. This provides an indication of the percentage uncertainty in the parameter values. In this case, when all six stoichiometric parameters are fitted simultaneously, it is found that the parameter uncertainties for parameters XH, XO, XN, PH, PO, and PN are 16.6%, 15.6%, 41.4%, 7.2%, 16.6%, and 5.9% respectively. The relatively high uncertainties are due to high correlation between the parameter values, which is expected in this simple stoichiometric model. If for example, the product stoichiometry is fixed, and only the stoichiometric parameters for the biomass are fitted, the uncertainty is below 5.5% for all parameters. The correlation matrices may be found in the supplementary material for both of these scenarios.

Conclusion

In this work, a soft sensor is developed for 550L filamentous fungal fermentations operated at Novozymes A/S. The parameter estimation uses only standard on-line measurements, of oxygen uptake rate, carbon dioxide evolution rate and ammonia flow rate, which are considered robust measurements, not subject to drift, and without the need for calibration. It is therefore considered that the model should be applicable to other strains, other scales, and other processes, and this provides an area for future work. The flexibility of the method and the simplicity of implementation make this a valuable tool for industrial application. Due to the scale of operation, it is considered important to introduce the evaporation term into the model, and define the states on a mass basis, rather than a volume basis. This accuracy in the mass prediction avoids large errors propagating to the state estimates, and aids prediction accuracy.

In order to apply more advanced optimization and control strategies to industrial fermentation processes there is a need for robust state estimators in order to identify key performance indicators in real time. The future work for this project is to further develop this monitoring tool, and work towards on-line control and optimization at pilot and production scales.

Acknowledgements

The PhD project of Lisa Mears is funded by Novozymes A/S and the Technical University of Denmark (DTU). The authors would like to thank Novozymes A/S for the experimental data used in this work, and the Novozymes fermentation pilot plant operators for running of the fermentation systems presented in this work.

Nomenclature

CO_{2out} = CO₂ molar ratio in outlet gas

CER= Carbon Evolution Rate (mol/h)

DO= Dissolved O₂ (mol O₂/kg)

F_{evap} = Evaporation rate (kg/h)

F_{feed} = Feed flowrate (L/h)

HSP= Headspace pressure (bar)

K_La = Mass transfer coefficient (h⁻¹)

M= Mass (kg)

M_r = Molecular mass (kg/mol)

N= Stirrer speed (RPM)

OUR= Oxygen Uptake Rate (mol/h)

P= Product concentration (g/kg)

P^* = Saturated vapour pressure (bar)

$ppCO_2$ = Partial pressure CO₂ (mbar)

Q_{air} = Aeration rate (m³/h)

R= Gas constant (J/mol.K)

RH= Relative humidity (%)

S_f = Feed substrate concentration (g/L)

T = Temperature (°C)

v_g = Superficial gas velocity (m/s)

X= Biomass concentration (g/kg)

Rate Subscripts

g = Glucose

o = Oxygen

n = Ammonia

x =Biomass

p = Product

c = Carbon dioxide

h = Water

Model terms

U = [F_{feed} , HSP, N, Q_{air}]

q_m = [q_o , q_c , q_n]

q_c = [q_x , q_p , q_w , q_g]

\hat{x} = [X, P, G, DO, M]

\hat{y} = [k_La , μ , F_{con}]

E = Stoichiometric matrix of coefficients

C, H, N, O

References

Albaek MO, Gernaey K V, Hansen MS, Stocks SM. 2011. Modeling enzyme production with *Aspergillus oryzae* in pilot scale vessels with different agitation, aeration, and agitator types. *Biotechnol. Bioeng.* **108**:1828–40.

Alford JS. 2006. Bioprocess control: Advances and challenges. *Comput. Chem. Eng.* **30**:1464–1475.

Alves-Rausch J, Bienert R, Grimm C, Bergmaier D. 2014. Real time in-line monitoring of large scale *Bacillus* fermentations with near-infrared spectroscopy. *J. Biotechnol.* **189**:120–8.

Araújo-Bravo MJ, Cano-Izquierdo JM, Gómez-Sánchez E, López-Nieto MJ, Dimitriadis YA, López-Coronado J. 2004. Automatization of a penicillin production process with soft sensors and an adaptive controller based on neuro fuzzy systems. *Control Eng. Pract.* **12**:1073–1090.

Assis AJ De, Filho RM. 2000. Soft sensors development for on-line bioreactor state estimation. *Comput. Chem. Eng.* **24**:1099–1103.

Bingham EC. 1916. An Investigation of the Laws of Plastic Flow. *US Bur. Stand. Bull.*:309–353.

Bolton D. 1980. The Computation of Equivalent Potential Temperature. *Mon. Weather Rev.* **108**:1046–1053.

Chen LZ, Nguang SK, Li XM, Chen XD. 2004. Soft sensors for on-line biomass measurements. *Bioprocess Biosyst. Eng.* **26**:191–5.

Cherry JR, Fidantsef AL. 2003. Directed evolution of industrial enzymes: an update. *Curr. Opin. Biotechnol.* **14**:438–443.

Doran PM. 2013. Bioprocess Engineering Principles. *Bioprocess Eng. Princ.* Elsevier.

Efron B. 1979. Bootstrap methods: Another look at the jackknife. *Ann. Stat.* **7**:1–26.

FDA. 2004. PAT — A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance.

Fordyce AP, Rawlings JB, Edgar TF. 1990. Control strategies for fermentation processes. In: *Comput. Control Ferment. Process.* Boca Raton: CRC Press, pp. 165–206.

Garcia-Ochoa F, Gomez E. 2009. Bioreactor scale-up and oxygen transfer rate in microbial processes: an overview. *Biotechnol. Adv.* **27**:153–76.

Gernaey K V, Lantz AE, Tufvesson P, Woodley JM, Sin G. 2010. Application of mechanistic models to fermentation and biocatalysis for next-generation processes. *Trends Biotechnol.* **28**:346–54.

Gernaey K V. 2015. A Perspective on PSE in Fermentation Process Development and Operation. *Comput. Aided Chem. Eng.* **37**:123–130.

Golabgir A, Hoch T, Zhariy M, Herwig C. 2015. Observability analysis of biochemical process models as a valuable tool for the development of mechanistic soft sensors. *Biotechnol. Prog.* **31**:1703–15.

Van Der Heijden RTJM, Heijden JJ, Hellinga C, Romein B, Luyben KCAM. 1994. Linear constraint relations in biochemical reaction systems: I. Classification of the calculability and the balanceability of conversion rates. *Biotechnol. Bioeng.* **43**:3–10.

Helton JC, Davis FJ. 2003. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliab. Eng. Syst. Saf.* **81**:23–69.

Hofmeester JJM. 1988. Gas hold-up measurements in bioreactors. *Trends Biotechnol.* **6**:19–22.

Iman RL, Conover WJ. 1982. A distribution-free approach to inducing rank correlation among input variables. *Commun. Stat. - Simul. Comput.* **11**:311–334.

Jenzsch M, Simutis R, Luebbert A. 2006. Generic model control of the specific growth rate in recombinant *Escherichia coli* cultivations. *J. Biotechnol.* **122**:483–93.

Linko S, Zhu Y-H, Linko P. 1999. Applying neural networks as software sensors for enzyme engineering. *Trends Biotechnol.* **17**:155–162.

Luttmann R, Bracewell DG, Cornelissen G, Gernaey K V, Glassey J, Hass VC, Kaiser C, Preusse C, Striedner G, Mandenius C-F. 2012. Soft sensors in bioprocessing: a status report and recommendations. *Biotechnol. J.* **7**:1040–8.

MATLAB. 2013. version 8.2 (R2013b).

Metzner AB, Otto RE. 1957. Agitation of non-Newtonian fluids. *AIChE J.* **3**:3–10.

Olsvik E, Kristiansen B. 1994. Rheology of filamentous fermentations. *Biotechnol. Adv.* **12**:1–39.

Riley GL, Tucker KG, Paul GC, Thomas CR. 2000. Effect of biomass concentration and mycelial morphology on fermentation broth rheology. *Biotechnol. Bioeng.* **68**:160–72.

Sagmeister P, Wechselberger P, Jazini M, Meitz A, Langemann T, Herwig C. 2013. Soft sensor assisted dynamic bioprocess control: Efficient tools for bioprocess development. *Chem. Eng. Sci.* **96**:190–198.

Schuster A, Schmoll M. 2010. Biology and biotechnology of *Trichoderma*. *Appl. Microbiol. Biotechnol.* **87**:787–99.

Sin G, Gernaey K V, Lantz AE. 2009. Good modeling practice for PAT applications: propagation of input uncertainty and sensitivity analysis. *Biotechnol. Prog.* **25**:1043–53.

Sonnleitner B. 2013. Automated Measurement and Monitoring of Bioprocesses : Key Elements of the M 3 C Strategy. *Adv. Biochem. Eng. Biotechnol.* **132**:1–33.

von Stosch M, Oliveira R, Peres J, Feyo de Azevedo S. 2014. Hybrid semi-parametric

modeling in process systems engineering: Past, present and future. *Comput. Chem. Eng.* **60**:86–101.

Sundström H, Enfors SO. 2008. Software sensors for fermentation processes. *Bioprocess Biosyst. Eng.* **31**:145–152.

Tolan JS, Foody B. 1999. Cellulase from Submerged Fermentation. In: Tsao, GT, editor. *Recent Prog. Bioconversion Lignocellul.* Berlin, Heidelberg: Springer Berlin Heidelberg. Advances in Biochemical Engineering/Biotechnology, Vol. 65, pp. 41–67.

Villadsen J, Nielsen J, Lidén G. 2011. *Bioreaction Engineering Principles.*

Wang PR, Scharling M, Nielsen KP. 2013. 2001– 2010 Danish Design Reference Year. Copenhagen 31 p.

Yuan X, Ge Z, Song Z. 2014. Locally Weighted Kernel Principal Component Regression Model for Soft Sensing of Nonlinear Time-Variant Processes. *Ind. Eng. Chem. Res.* **53**:13736–13749.

Zhang H, Lennox B. 2004. Integrated condition monitoring and control of fed-batch fermentation processes. *J. Process Control* **14**:41–50.

List of tables

Table I: Statistical analysis of model fit of product concentration for 11 calibration batches, and 14 on-line validation batches. Root mean sum of squared errors, RMSSE (%), and the Janus coefficient, J^2 is shown.

Calibration	1	2	3	4	5	6	7	8	9	10	11	mean			
% RMSSE	21.8	6.4	6.0	26.2	10.7	8.7	10.3	11.8	24.4	4.3	14.5	13.2			
Validation	1	2	3	4	5	6	7	8	9	10	11	12	13	14	mean
% RMSSE	20.5	48.4	8.8	9.3	14.5	18.3	11.9	16.6	12.1	16.7	23.3	15.5	5.4	11.4	16.6
Janus coefficient	1.3	1.1	0.4	0.5	2.0	2.7	0.7	3.5	1.7	1.5	3.5	1.5	0.1	0.6	1.5

Table II: Sensitivity analysis of the model prediction to process measurement errors. The results are given for the batch shown in Figure 7. The table provides the relative change in the model prediction for biomass concentration and product concentration compared to when no measurement error is applied.

	+5% NH3	+5% CER	+5% OUR	+5% CER +5% OUR
Change in final biomass prediction	3.1%	63.8%	-58.4%	5.5%
Change in final product prediction	9.3%	-10.7%	8.9%	-1.3%

List of Figures

Figure 1: Model structure: On-line parameter estimation block, coupled to a dynamic process model.

Figure 2: Implementation algorithm

Figure 3: Coupled parameter estimator and dynamic model applied to pilot scale batch data. Batch data obtained from Novozymes A/S. Model prediction (grey), off-line measured data (black). Scaled axis labels for confidentiality reasons.

Figure 4: Product concentration prediction for 11 historical pilot scale batches. Batch data obtained from Novozymes A/S with different conditions for headspace pressure, aeration rate and stirrer speed. Model prediction (grey), off-line measured data (black). Scaled axis labels for confidentiality reasons.

Figure 5: Biomass concentration prediction for 11 historical pilot scale batches. Batch data obtained from Novozymes A/S with different conditions for headspace pressure, aeration rate and stirrer speed. Model prediction (grey), off-line measured data (black). Scaled axis labels for confidentiality reasons.

Figure 6: Product concentration prediction for fourteen validation pilot scale batches. Batches were performed using different operating conditions for headspace pressure, aeration rate and stirrer speed, resulting different levels of product formation. Model prediction (grey), off-line measured data (black). Scaled axis labels for confidentiality reasons. The y-axis scaling is the same as in Figure 4.

Figure 7: Coupled parameter estimator and dynamic model applied on-line in Novozymes A/S fermentation pilot plant. Model prediction (grey), off-line measured data (black). Scaled axis labels for confidentiality reasons. The y-axis scaling is the same as in Figure 3.

Figure 8: Temperature and relative humidity data obtained from the 2001-2010 Design

Reference Year for Denmark dataset from the Danish Meteorological Institute (Wang et al., 2013). The correlation coefficient is -0.53.

Figure 9: Evaporation rate model, for a pilot scale process operated at 300NL/min and 0.7 bar headspace pressure. Effect of the inlet air relative humidity and temperature is shown. The Monte Carlo simulations are applied to 250 sample points, which are generated using Latin Hypercube Sampling (Helton and Davis, 2003) taking into consideration the correlation coefficient of -0.53 which is obtained from the data in Figure 8. Coloured surface plot of the results is shown, with colours corresponding to evaporation rate (kg/week).

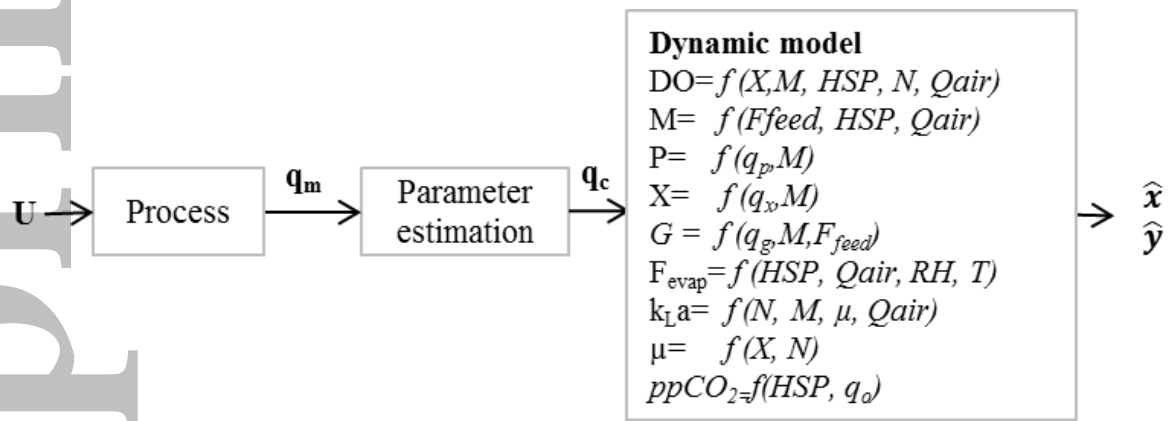


Figure 1

- Establish OPC server connection
 - Initialise MATLAB® timer object for reading and saving online data (30 seconds)
 - Initialise MATLAB® timer object for running soft sensor model (300 seconds)
 - Load batch conditions, and model parameters (HSP, Q_{air} , N , S_f , RH , T)
 - Define initial conditions for states ($DO_{t=0}$, $X_{t=0}$, $P_{t=0}$, $G_{t=0}$, $M_{t=0}$)
 - Start timers
- Every 30 seconds: Read data timer
 - Read and save OUR (mol/h)
 - Read and save CER (mol/h)
 - Read and save ammonia flowrate (mol/h)
 - Read and save feed flowrate

$\downarrow q_o, q_c, q_n, F_{feed}$

 - Every 300 seconds: Soft sensor timer
 - Take average of past 300 seconds of online data (10 data points)
 - Update parameter estimation (q_g, q_x, q_p, q_h)
 - Solve dynamic model ($DO, X, P, G, M, k_L a, F_{evap}, \mu, pp_{CO_2}$)
 - Save soft sensor predictions (DO, X, P, G, M)
 - Save current state as initial conditions for next iteration ($DO=DO_{t=0}, X=X_{t=0}, P=P_{t=0}, G=G_{t=0}, M=M_{t=0}$)
- Manually stop timers at end of batch and save dataset

Figure 2

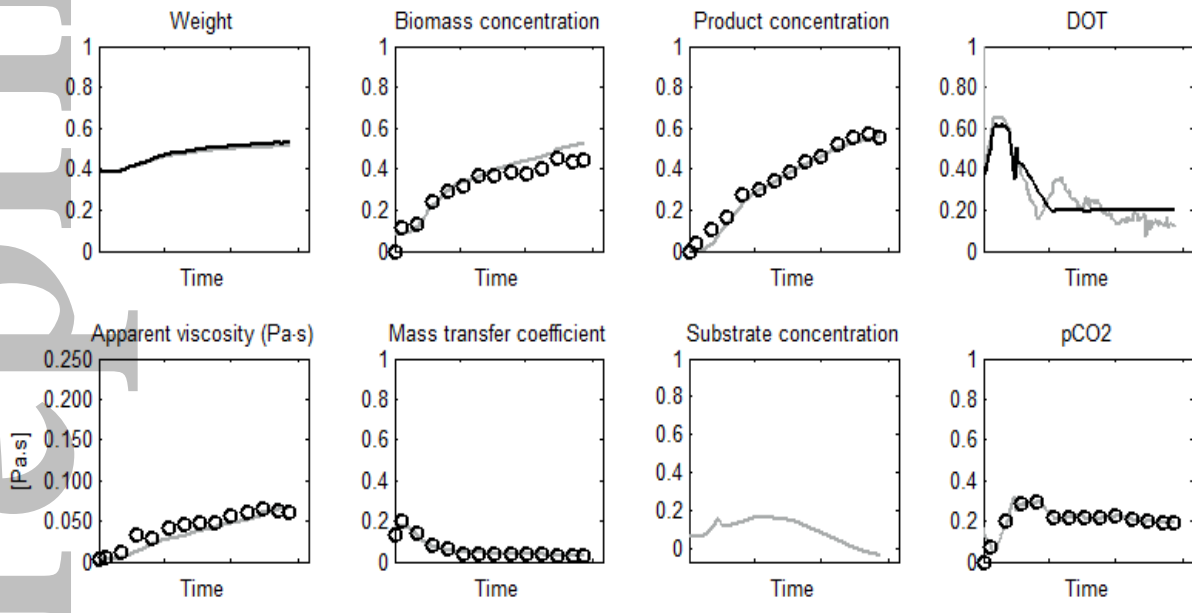


Figure 3

Product concentration

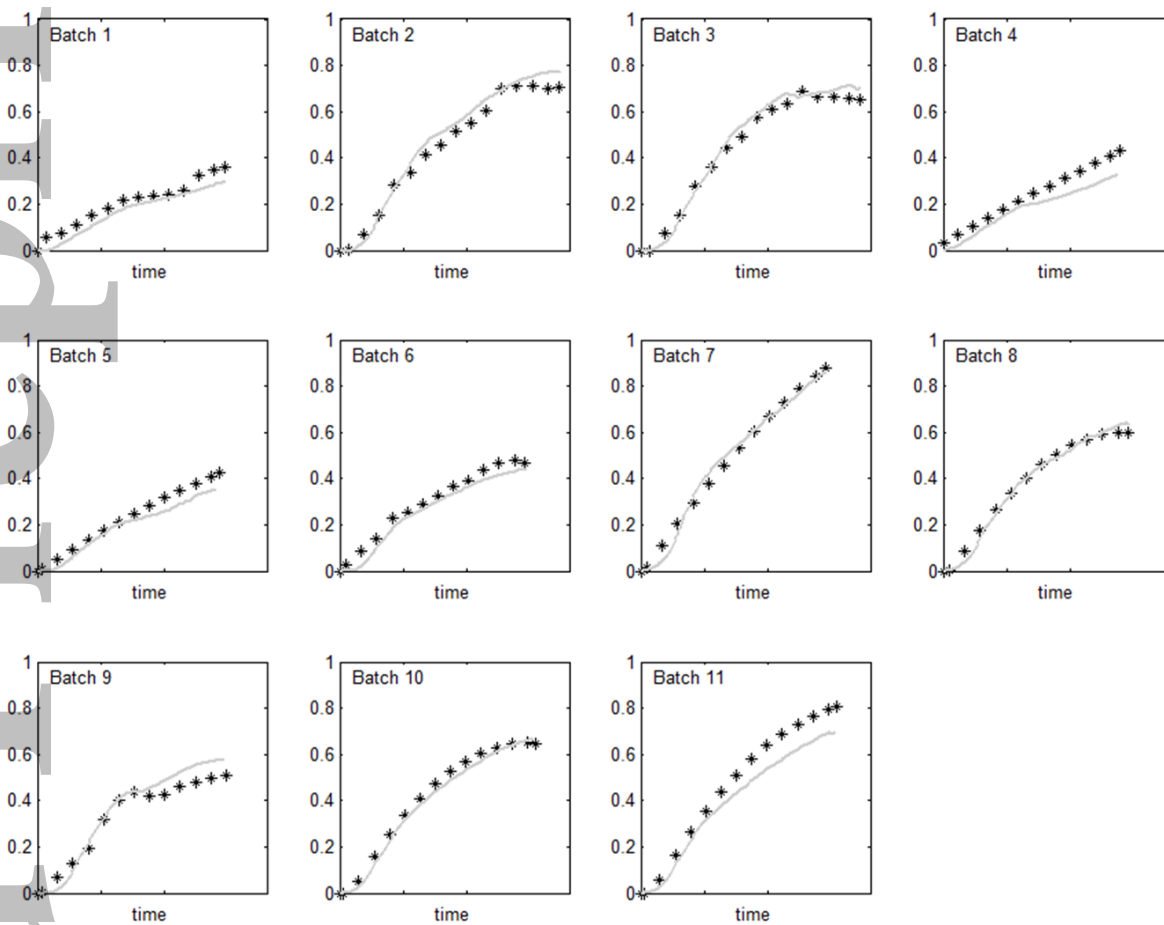


Figure 4

Biomass concentration

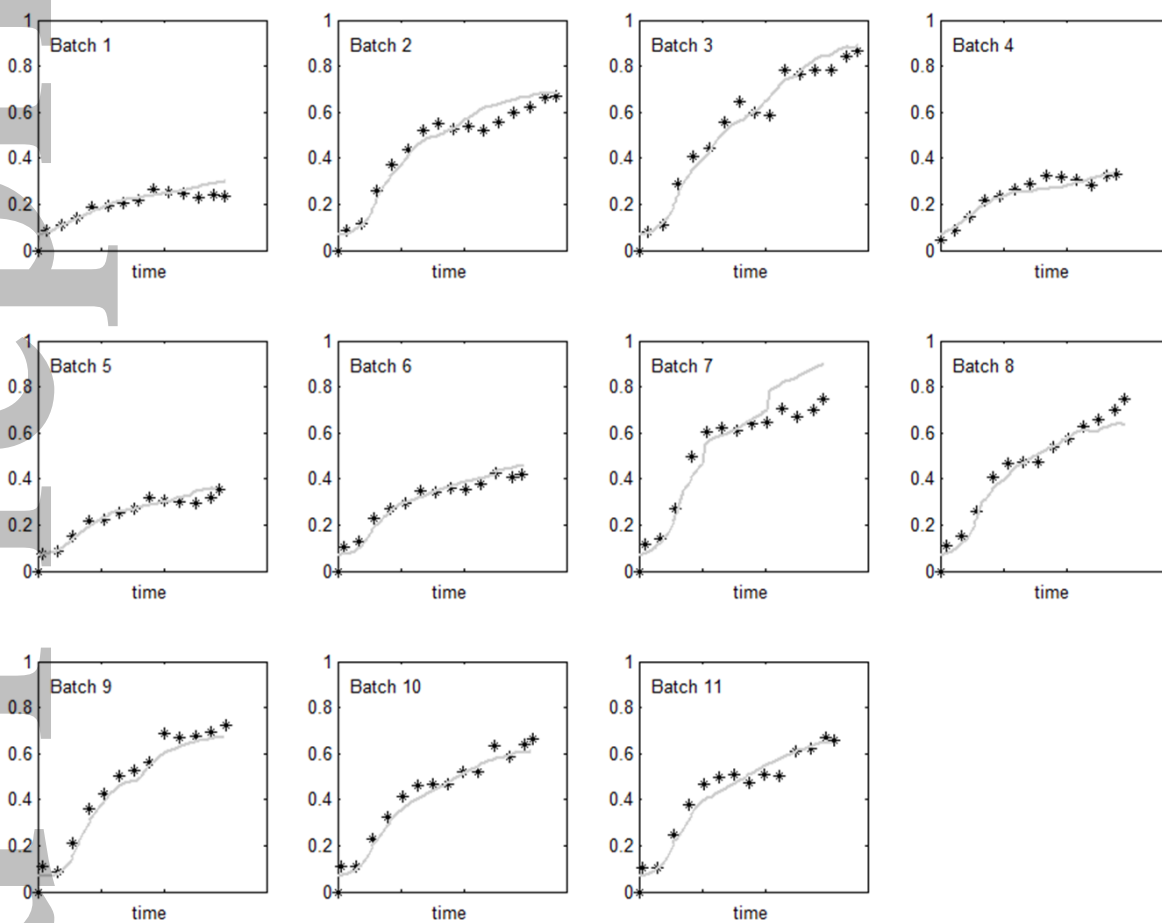


Figure 5

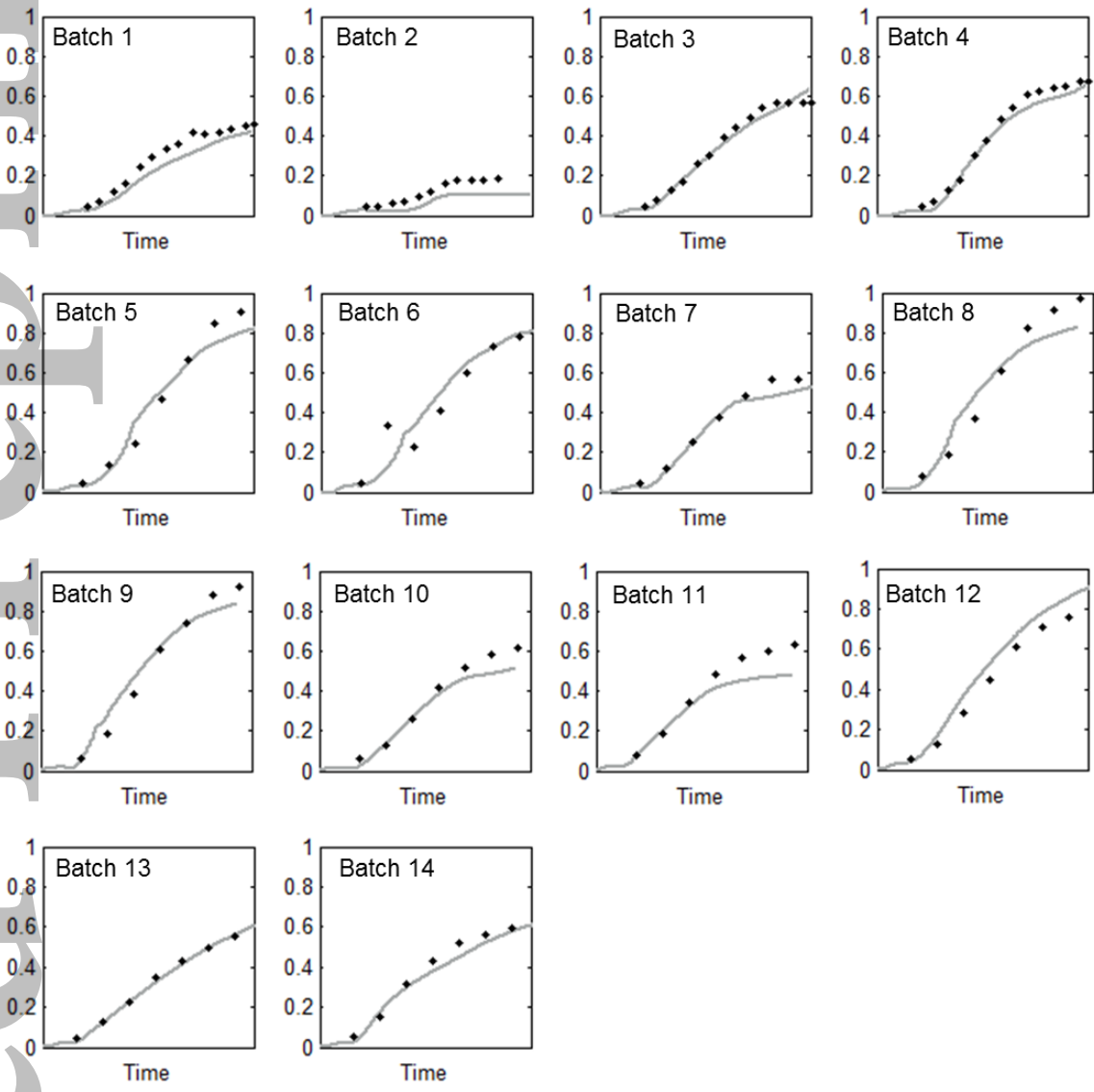


Figure 6

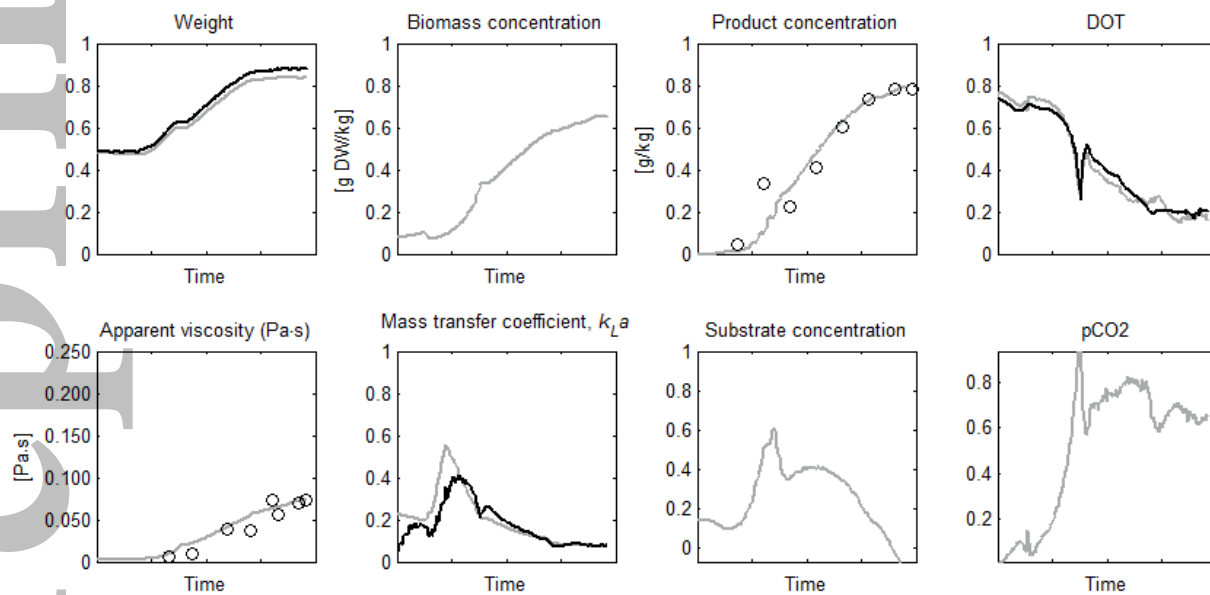


Figure 7

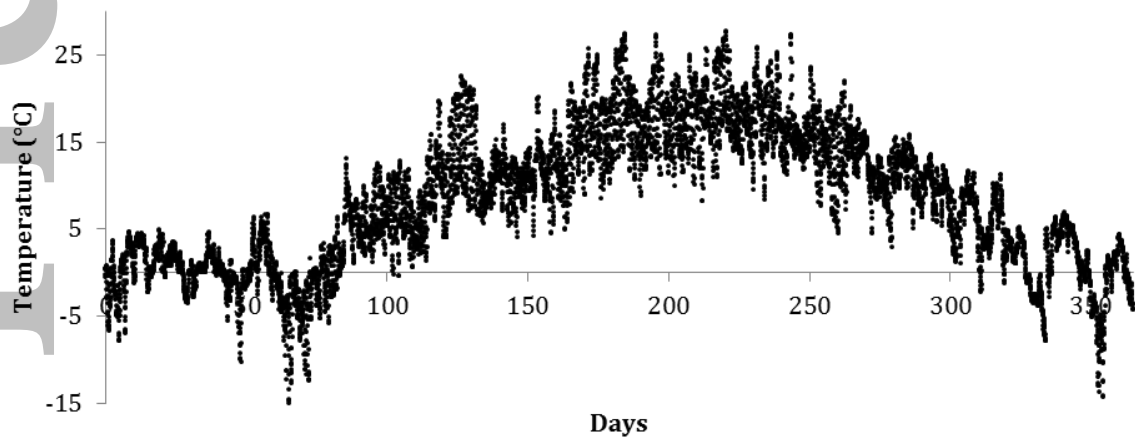
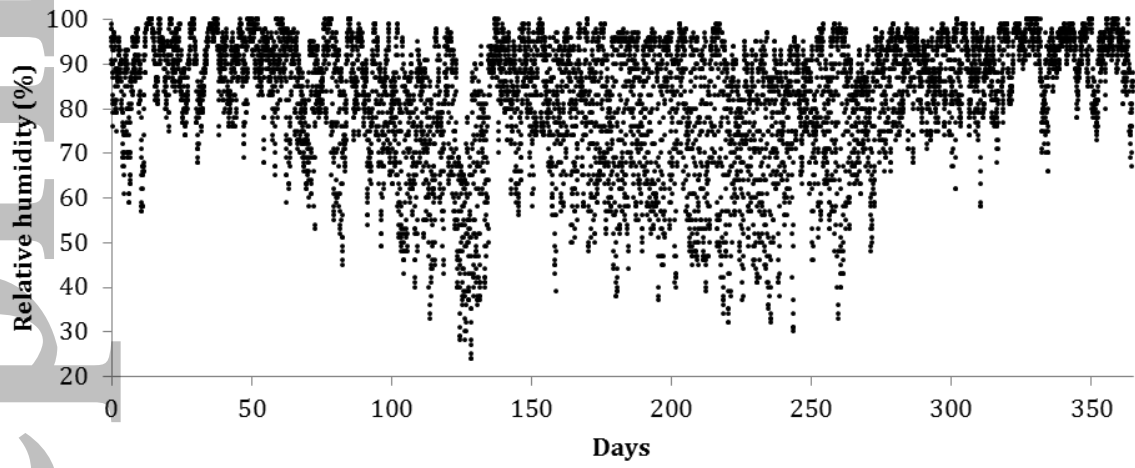


Figure 8

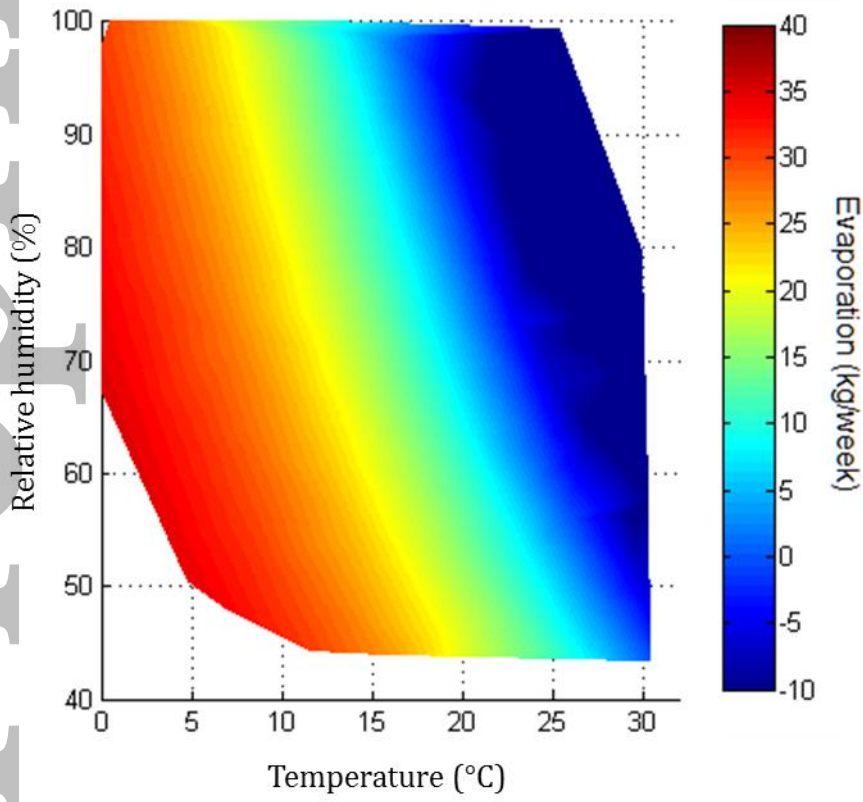


Figure 9