Title
A.B. Editor et al. (Editors)

Greybox stochastic modelling of industrial fed-batch cultivation

Jan K. Rasmussen[a], Henrik Madsen[b], Sten B. Jørgensen[a]

[a]*CAPEC, Department of Chemical engineering, Technical University of Denmark, Building 229, DK-2800 Lyngby, Denmark*
[b]*Informatics and Mathematical Modelling, Technical University of Denmark, Building 321, DK-2800 Lyngby, Denmark*

Abstract

This paper presents application of a grey-box stochastic modelling framework for developing continuous time models for dynamic systems based on experimental data. The framework is used to develop predictive models suited for control purposes. The main idea behind the framework is to combine stochastic modelling with data to obtain information on parameter values and model (in-)validity. In case the proposed model is falsified the method can point out specific functional deficiencies which facilitates further model development. The developed model can be used for monitoring purposes as well as serve as a basis for advanced multivariable control to reject both intrabatch and interbatch disturbances. The industrial fermentation investigated in this case is production of an amylase enzyme by a filamentous fungus.
Keywords: Parameter estimation, grey-box modelling, industrial fermentation.

1. Introduction

Fed-batch processes play a very important role in chemical and biochemical industry. Fermentations are widely used in biochemical industry and are most often carried out as fed-batch processes. Present control schemes do not utilise the full potential of the production facilities and may often fail to achieve uniform product quality and optimal productivity. Application of advanced monitoring tools and multivariable control schemes can be a solution to this problem. The introduction of model based control strategies is considered difficult because suitable models are not readily available and require a significant investment in experimental work for their development.

First principles engineering models can be used for monitoring and control purposes assuming that they are sufficiently accurate. Parameter estimation in a first principles engineering model can be very time consuming and can cause problems when scaling up from laboratory to industrial fermentors. The latter phenomena can not be investigated in laboratory scale equipment which therefore makes large scale experiments desirable.

The approach taken in this paper is to combine first principle engineering models with operational data to produce predictive models suited for control purposes. The method described in this paper is grey-box stochastic modelling which consists of a set of stochastic differential equations (SDEs) describing the dynamics of the system in continuous time and a set of discrete time measurements. An important advantage using this approach compared to using deterministic models is that stochastic models can account for random variations in data and thereby provide a sound basis for hypothesis testing related to model validity based on available data.

A framework for greybox stochastic model development has already been developed (Kristensen et al., 2004) and is described in figure 1.

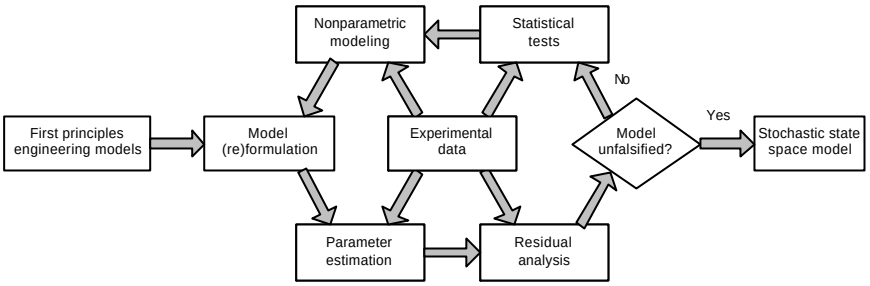

Figure 1.Grey-box stochastic modelling framework

One of the key ideas behind the grey-box stochastic modelling framework is to use all prior information for formulation of an initial first principles engineering model. Unknown parameters of the initial model are then estimated from experimental data and a residual analysis is carried out to evaluate the quality of the resulting model. The next step in the modelling cycle is the model falsification or unfalsification which aims to determine if the model is sufficiently accurate to serve its intended purpose. If the model is unfalsified the model development is completed assmuming that data are representative for the intended applications. In case of falsification the modelling cycle must be repeated by reformulating the initial model. In this case statistical tests can be used to provide indications of which parts of the model that are deficient. Nonparametric modelling can be applied to estimate which functional relationships are needed to improve the model.
If the developed model is sufficiently accurate it can be used for online monitoring of the process or serve as a software sensor for otherwise unobservable states. Furthermore it can serve as the process model for development and tuning of advanced multivariable controllers. In this paper the modelling framework is considered.

2. Process description

The process studied is a cultivation of the filamentous fungi *Aspergillus* for production of the enzyme amylase. The fermentation is initiated by transferring the contents of a seed tank to the main fermentation tank when a specified transfer criterion has been satisfied. The main fermentation tank contains an initial amount of substrate and the main fermentation process starts immediately after inoculation. The main fermentation is carried out in a batch and fedbatch phase. When the initial substrate in the main fermenter has been consumed by the microorganisms the fedbatch phase is initiated. Feed dosing is started at a low level and increased to its final value within a specified time span. The fedbatch phase continues for the rest of the fermentation duration and the majority of the enzyme is produced in this phase.
The fermentors are equipped with sensors for online measurements of different process variables but some values are only available as offline measurements which makes closed loop control more difficult and requires an accurate model for predicting the unobserved variables.

3. Model formulation

Initially a very simple first principles model for the fermentation is proposed. Measurements show that only a small amount of enzyme is formed in the batch phase

(before the substrate feed begins) and the two phases are therefore modelled seperately. Only the batch phase is considered in the model presented in this paper and it is assumed that no enzyme is formed in this phase of the process. If necessary this assumption can be changed in a later model iteration. In order to keep the model simple further assumptions have been made regarding product formation and yields. It is assumed that substrate is converted to only carbon dioxide and biomass and that the yield coefficients for conversion of substrate and uptake of oxygen are constant. The assumed yield coefficients can therefore easily be calculated from information about the initial amount of substrate, the total evolution of carbon dioxide and the total uptake of oxygen. Furthermore the specific growth rate, μ, and the oxygen mass transfer coefficient, k_La, are assumed constant.
The initial model is given by 3 types of equations. Stochastic differential equations, algebraic equations and observation equations.

Stochastic differential equations:

$$dx = \left(\mu x + \frac{x}{V} F_{evap} \right) dt + \sigma_1 dw_1 \qquad (1)$$

$$ds = -\left(\frac{1}{Y_{sx}} \mu x - \frac{s}{V} F_{evap} \right) dt + \sigma_2 dw_2 \qquad (2)$$

$$dc_{O_2} = \left(-r_{O_2} x + k_L a \left(c_{O_2}^{sat} - c_{O_2} \right) + \frac{c_{O_2}}{V} F_{evap} \right) dt + \sigma_3 dw_3 \qquad (3)$$

$$dV = -F_{evap} dt + \sigma_4 dw_4 \qquad (4)$$

The states considered in this system are: biomass (x), substrate concentration (s), oxygen concentration (cO_2) and volume (V).
w_i are four independent Wiener processes with incremental standard deviations given by σ_i.

Algebraic equations:

$$r_{O_2} = Y_{xo} \mu \quad ; \quad r_{CO_2} = Y_{xc} \mu \qquad (5)$$

Here the specific rates of oxygen consumption (rO_2) and carbon dioxide evolution (rCO_2) are given. These are modelled as being proportional to the specific growth rate, the proportional factors being the yield coefficients.

Observation equations:

$$OUR = r_{O_2} xV + e_1 \quad ; \quad e_1 \in N(0, s_1^2) \qquad (6)$$

$$CER = r_{CO_2} xV + e_2 \quad ; \quad e_2 \in N(0, s_2^2) \qquad (7)$$

$$DOT = \frac{c_{O_2}}{c_{O_2}^{sat}} \cdot 100\% + e_3 \quad ; \quad e_3 \in N(0, s_3^2) \tag{8}$$

$$Volume = V + e_4 \quad ; \quad e_4 \in N(0, s_4^2) \tag{9}$$

For the above modelling 4 variables are used from the experimental data sets. OUR is the Oxygen Utilisation Rate, CER is the Carbon dioxide Evolution rate and DOT is the Dissolved Oxygen Tension (oxygen concentration measured in percent of saturation). The volume is included directly in the model. e_i are independent white noise processes taken from a normal distribution with a mean of zero and standard deviation of s_i.

4. Parameter estimation and results

Experimental data are taken from three batches run in pilot plant at Novozymes A/S. All batches have been run inder similar conditions and using the same fermentation recipe. Parameters and corresponding standard deviations are estimated using a maximum likelihood method implemented in a computer program called CTSM (Continous Time Stochastic Modelling). The program solves the stochastic differential equations in continous time and estimates parameters using discrete time measurements.
The initial parameter estimation shows that the uncertainties on parameters μ and k_La are very large and additionally the standard deviations of the Wiener processes in equation (1) and (3) are significant. This shows that the assumptions of constant μ and k_La does not hold. In order to reveal the time variation of the two phenomena they are introduced as states and two stochastic differential equations are therefore added to the previous four:

$$dm = 0 + s_5 dw_5 \tag{10}$$

$$dk_L a = 0 + s_6 dw_6 \tag{11}$$

The phenomena are assumed to be constant over time which is clearly not true, but the CTSM approach allows for subsequent estimation of the timewise behaviour of the phenomena and furthermore of their functional dependence of other states.

	Batch 1		Batch 2		Batch 3	
Parameter	Estimate	Std. dev.	Estimate	Std. dev.	Estimate	Std. dev.
x_0	5.71E-01	1.30E-01	9.51E-02	8.35E-02	4.19E-01	1.55E-01
$cO2_0$	4.16E-04	3.16E-06	4.18E-04	5.61E-07	4.13E-04	8.26E-07
V_0	1.17E+03	4.54E+00	1.17E+03	7.97E-01	1.16E+03	4.10E-01
μ_0	1.62E-01	3.21E-02	1.56E-01	2.97E-02	1.41E-01	3.70E-02
k_La_0	9.51E+02	1.06E+02	5.22E+02	6.76E+01	1.04E+03	8.56E+01

Table 1. Estimates from the second estimation using eq. (1-11) with corresponding standard deviations of initial states.

Table 1 gives estimates of the initial states in the model as well as uncertainty information (standard deviation). The initial substrate concentration is a known parameter and is therefore not estimated. It is seen that the initial estimates of cO_2, V

and μ are similar for the three batches. The initial estimate of biomass (x) varies a lot from batch to batch suggesting that different amounts of biomass have been transferred to the main fermentation tank. The estimate on k_La is similar for batch 1 and 3 but only aproximately half the value in batch 2.

	Batch 1		Batch 2		Batch 3	
	t-score	P(>|t|)	t-score	P(>|t|)	t-score	P(>|t|)
log s_1 *(x)*	-31.9147	0	-6.1517	0	-4.7905	0
log s_2 *(s)*	-0.029	0.9769	-0.1521	0.8792	-0.09	0.9284
log s_3 *(cO_2)*	-5.1117	0	-62.6838	0	-8.2029	0
log s_4 *(V)*	-0.0581	0.9537	0.0259	0.9794	-0.057	0.9546
log s_5 *(μ)*	-38.6969	0	-35.7205	0	-29.4522	0
log s_6 *(k_La)*	23.6974	0	6.1096	0	28.4267	0

Table 2. Estimate of significances of state noise terms for model eq. (1-11).

The estimates of the standard deviations of the state noise terms are given in table 2. For numerical reasons the logarithmic standard deviations are estimated as this gives a better numerical conditioning of the system.
The p(>|t|) value is the fraction of probability of the corresponding *t*-distribution outside the limits set by the t-score. It can be interpreted as the probability that the parameter is insignificant. If the parameter is insignificant it can be replaced by a fixed value which is smaller than the estimate or even by a zero. If the noise term is insignificant it indicates that the corresponding SDE gives a satisfactory description of the particular state variable. Any model deficiencies will be contained in the noise term. As long as the noise term is significant the corresponding SDE does not match with the experimental observations and should be modified. This is seen for parameters s_1, s_3, s_5 and s_6. This suggests that functional relationships related to x, cO_2, μ and k_La should be improved.

	Batch 1		Batch 2		Batch 3	
	t-score	P(>|t|)	t-score	P(>|t|)	t-score	P(>|t|)
log s_1 *(OUR)*	0.0011	0.9991	0.1948	0.8457	0.001	0.9992
log s_2 *(CER)*	0.0008	0.9993	0.1918	0.8481	0.001	0.9992
log s_3 *(DOT)*	-4.3865	0	-24.651	0	-8.6884	0
log s_4 *(V)*	23.4939	0	-2.7129	0.0072	5.5806	0

Table 3. Estimate of significances of measurement noise terms for model eq. (1-11).

The estimates of the standard deviations (logarithmic value) on the measurement noise terms are given in table 3. Again it is seen that similar conclusions can be drawn from all three batches. It is seen that the measurent equations for OUR and CER fit the experimental data very well whereas the equations for DOT and volume should be reexamined. The significant noise in the DOT measurement can be explained by poor calibration of the sensor and the uncertainty on the volume measurement is known to be large.

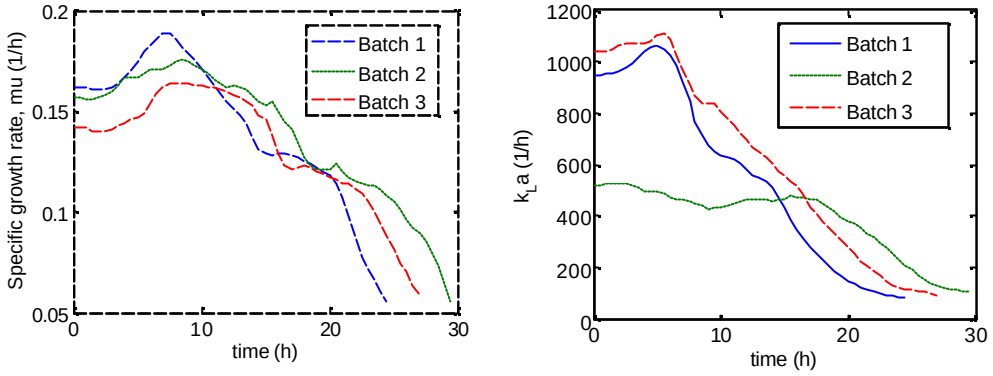

Figure 2. Estimates of specific growth rate, μ, and oxygen mass transfer coefficient , k_La, as function of time.

The specific growth rate (figure 2) is seen to exhibit a very similar behaviour in all three batches. In the beginning the growth rate increases to a maximum value around 7-10h, which can be explained by intracellular formation of proteins due to the new growth conditions in the batch phase. Thereafter a decrease appears and at a certain point (around 15-18h) the growth rate seems to level off for a few hours whereafter the decrease continues. This phenomenon is not simple to explain but is likely to be due to the cells producing biproducts. The evolution of μ as shown in the figure is only valid if the yield coefficients indeed can be assumed constant.
The k_La shows a very similar behaviour for two of the three batches. For batch 2 it is seen that the value is much lower in the beginning of the process than for the two other batches. This might be due to poor aeration or stirring conditions.

5. Discussion

Application of a grey-box stochastic modelling framework used to model a fermentation process has been illustrated in this paper. A stochastic model has been combined with experimental data to obtain information on phenomelogical dependencies and uncertainties. Data from three batches run under similar conditions has been used and similar behaviour, in eg., the specific growh rates is observed. This indicates that this modelling methodology provides a sound basis for development of a model which can capture essential process dynamics. Essential unknown phenomena, eg. formation of biproducts, must be investigated experimentally before the modelling cycle can be continued.

References

1. Kristensen, N. R., H. Madsen and S.B. Jørgensen, 2004, A Method for Systematic Improvement of Stochastic Grey-Box Models, Comp. & Chem. Eng., 28/8, 1431-1449
2. Kristensen, N. R. and H. Madsen Continuos, Time Stochastic Modelling Mathematics Guide, Informatics and Mathematical Modelling, Technical University of Denmark

Acknowledgments

The first author thanks the support from Novozymes Bioprocess Academy which made this project possible.