European Symposium on Computer Aided Process Engineering - 11
R. Gani and S.B. Jørgensen (Editors)

Computer Aided Continuous Time Stochastic Process Modelling

Niels Rode Kristensen[a], Henrik Madsen[b] and Sten Bay Jørgensen[a]

[a]Computer Aided Process Engineering Center (CAPEC), Department of Chemical Engineering

[b]Section for Mathematical Statistics, Department of Mathematical Modelling

Technical University of Denmark, DTU, DK-2800 Lyngby, Denmark

A grey-box approach to process modelling that combines deterministic and stochastic modelling is advocated for identification of models for model-based control of batch and semi-batch processes. A computer-aided tool designed for supporting decision-making within the corresponding modelling cycle is presented.

1. INTRODUCTION

With the development and increasing number of possible applications of advanced model based process control schemes, e.g. model predictive control (MPC), more and more rigorous demands are placed on the quality of available dynamic process models. Model quality measures the ability of the model to predict the future evolution of the process, so in order to obtain good prediction performance, these models must be able to capture the inherently nonlinear behaviour of many process systems, such as batch and semi-batch processes. Furthermore these models must be able to provide predictions in the presence of noise, i.e. *process noise* due to approximation errors, unmodelled inputs and plant-model mismatch and *measurement noise* due to imperfect measurements. Meeting both demands with the same model is difficult, so there is a tendency in litterature to use either a *deterministic* approach or a *stochastic black-box* approach to process modelling.

The *deterministic* approach is based on using first engineering principles to derive ordinary differential equation (ODE) models. These models are well-suited for describing nonlinear behaviour, but they lack the desired predictive capabilities in the presence of noise, because they do not encompass a noise model and because unknown parameters are estimated in an output error (OE) setting, which tends to emphasize the pure simulation capabilities of the model instead of the predictive capabilities, cf. Young (1981).

The *stochastic black-box* approach, on the other hand, is based solely on using time series data for identifying a model, usually in the form of a discrete time transfer function model. These models usually have very nice predictive capabilities because of their inherent noise model and because unknown parameters are estimated in a prediction error (PE) setting, cf. Young (1981). Unfortunately these models are not equally well-suited for describing nonlinear behaviour, especially not outside the (possibly narrow) operating region, within which the time series data for identification is obtained.

In this paper an alternative *grey-box* approach to process modelling is advocated. This approach combines the deterministic approach and the stochastic black-box approach in a way that seeks to combine their respective strengths, i.e. from ODE models the intuitive appeal of their derivation from first engineering principles and their ability to describe nonlinear behaviour, and from stochastic black-box models the nice predictive capabilities and their ability to handle both process and measurement noise.

The aim of this paper is to describe the grey-box approach and outline its advantages. This is done in Section 2, where a computer aided tool that aims to support decision-making within this approach is also presented. In Section 3 a small example is given to illustrate one of the advantages of this approach and the conclusions are presented in Section 4.

2. A GREY-BOX APPROACH TO PROCESS MODELLING

A very appealing way of combining the deterministic and the stochastic approaches to process modelling is to use stochastic differential equation (SDE) models as shown by Åström (1970). The grey-box approach advocated in this paper is therefore based on SDE models in the Itô sense or, to be more specific, on the *continuous-discrete stochastic state space model*

$$dx_t = f(x_t, u_t, t, \theta)dt + \sigma(t, \theta)d\omega_t \tag{1}$$

$$y_k = h(x_k, u_k, t_k, \theta) + e_k \tag{2}$$

where $t \in \mathbb{R}$ is time, $x_t \in \mathcal{X} \subset \mathbb{R}^n$ is a vector of state variables, $u_t \in \mathcal{U} \subset \mathbb{R}^m$ is a vector of input variables and $y_k \in \mathcal{Y} \subset \mathbb{R}^l$ is a vector of measurements. $x_k = x_{t=t_k}$ and $u_k = u_{t=t_k}$. $\theta \in \Theta \subset \mathbb{R}^p$ is a vector of parameters, and $f(\cdot) \in \mathbb{R}^n$, $\sigma(\cdot) \in \mathbb{R}^{n \times q}$ and $h(\cdot) \in \mathbb{R}^l$ are nonlinear functions, ω_t is a q-dimensional standard Wiener process and $e_k \in N(0, S(t_k, \theta))$ is an l-dimensional white noise process.

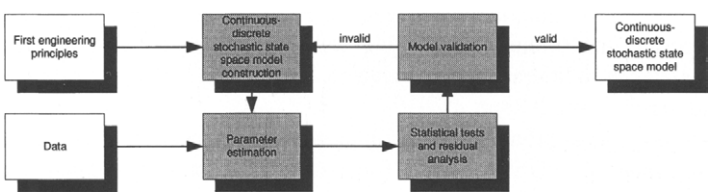

Fig. 1. *The modelling cycle for control, which constitutes the core of the grey-box approach to process modelling.*

Figure 1 shows a modelling cycle based on this model, which describes the grey-box approach, and by means of which some of its advantages can be outlined.

- The continuous time system equation (1) allows the initial structure of the model to be determined from first engineering principles in the form of an ODE model, which is intuitively appealing, since any prior physical knowledge can be included and because

the parameters of the model can easily be given a physical interpretation. Furthermore, most chemical and process systems engineers are familiar with this way of constructing a model.

- When subsequently determining unknown parameters of the model from a set of data, the continuous time system equation (1) and the discrete time measurement equation (2) make the model flexible by allowing varying sample times and missing observations.
- The model provides a separation between process and measurement noise, which along with the stochastic nature of the model allow the parameters to be estimated in a PE setting using a statistically sound method, e.g. *maximum likelihood* (ML).
- For the same reasons statistical tests and residual analysis can subsequently be applied in a systematic manner to validate the model, and if it is found that the model is not valid these tools also provide information on how to alter the model to improve its quality.

In the following the individual elements of the modelling cycle are explained in more detail.

Once a model structure has been determined from first engineering principles, unknown parameters of the model can be estimated from a set of data. Nielsen *et al.* (2000) have recently reviewed the state of the art with respect to parameter estimation in discretely observed Itô SDE's and found that only methods based on nonlinear filtering provide an approximate solution to the full problem of determining ML estimates of the parameters of the continuous-discrete stochastic state space model. Unfortunately, applying nonlinear filtering is difficult, so in order for the grey-box approach to be feasible, extended Kalman filtering (EKF) is used instead as shown in the following.

Determining ML estimates of the parameters means finding the parameters θ, including the initial conditions x_0, that maximize the likelihood function with respect to θ given a set of measurements $y_0, y_1, \ldots, y_k, \ldots, y_N$. By introducing $\mathcal{Y}_k = [y_k, y_{k-1}, \ldots, y_1, y_0]$ and $\hat{y}_{k|k-1} = E\{y_k|\mathcal{Y}_{k-1}, \theta\}$, $R_{k|k-1} = V\{y_k|\mathcal{Y}_{k-1}, \theta\}$ and $\varepsilon_k = y_k - \hat{y}_{k|k-1}$ and by assuming that the conditional probability densities are Gaussian, the likelihood function becomes

$$L(\mathcal{Y}_N|\theta) = \left(\prod_{k=1}^{N} p(y_k|\mathcal{Y}_{k-1}, \theta)\right) p(y_0|\theta) = \left(\prod_{k=1}^{N} \frac{\exp\left(-\frac{1}{2}\varepsilon_k^T R_{k|k-1}^{-1}\varepsilon_k\right)}{\sqrt{\det\left(R_{k|k-1}\right)}\left(\sqrt{2\pi}\right)^l}\right) p(y_0|\theta) \tag{3}$$

where, for given parameters θ, ε_k and $R_{k|k-1}$ can be computed by using a continuous-discrete EKF. If prior information is available in the form of an a priori probability density function $p(\theta)$ for the parameters, Bayes rule can provide an improved estimate of the parameters by forming the posterior probability density function, i.e.

$$p(\theta|\mathcal{Y}_N) = \frac{L(\mathcal{Y}_N|\theta)p(\theta)}{p(\mathcal{Y}_N)} \propto L(\mathcal{Y}_N|\theta)p(\theta) \tag{4}$$

and subsequently finding the parameters that maximize this function, i.e. by performing *maximum a posteriori* (MAP) estimation. By assuming that the prior probability density of the parameters is Gaussian, and by introducing $\mu_\theta = E\{\theta\}$, $\Sigma_\theta = V\{\theta\}$ and $\varepsilon_\theta = \theta - \mu_\theta$ the posterior probability density function becomes

$$p(\theta|\mathcal{Y}_N) \propto \left(\prod_{k=1}^{N} \frac{\exp\left(-\frac{1}{2}\varepsilon_k^T R_{k|k-1}^{-1}\varepsilon_k\right)}{\sqrt{\det\left(R_{k|k-1}\right)}\left(\sqrt{2\pi}\right)^l}\right) p(y_0|\theta) \frac{\exp\left(-\frac{1}{2}\varepsilon_\theta^T \Sigma_\theta^{-1}\varepsilon_\theta\right)}{\sqrt{\det\left(\Sigma_\theta\right)}\left(\sqrt{2\pi}\right)^p} \tag{5}$$

If, instead of a single set of measurements, several consecutive, but yet separate, sets of measurements, i.e. $\mathcal{Y}_{N_1}^1$, $\mathcal{Y}_{N_2}^2$, ..., $\mathcal{Y}_{N_i}^i$, ..., $\mathcal{Y}_{N_S}^S$, possibly of varying length, are available, a similar estimation method can be applied by expanding the expression for the posterior probability density function to the general form

$$p(\theta|\mathbf{Y}) \propto \left(\prod_{i=1}^{S} \left(\prod_{k=1}^{N_i} \frac{\exp\left(-\frac{1}{2}(\varepsilon_k^i)^T (R_{k|k-1}^i)^{-1} (\varepsilon_k^i)\right)}{\sqrt{\det\left(R_{k|k-1}^i\right)}\left(\sqrt{2\pi}\right)^l} \right) p(y_0^i|\theta) \right) \frac{\exp\left(-\frac{1}{2}\varepsilon_\theta^T \Sigma_\theta^{-1} \varepsilon_\theta\right)}{\sqrt{\det(\Sigma_\theta)}\left(\sqrt{2\pi}\right)^p} \tag{6}$$

where $\mathbf{Y} = [\mathcal{Y}_{N_1}^1, \mathcal{Y}_{N_2}^2, \ldots, \mathcal{Y}_{N_i}^i, \ldots, \mathcal{Y}_{N_S}^S]$. Finding the estimates of the parameters θ is now a matter of further conditioning on $\mathbf{y_0} = [y_0^1, y_0^2, \ldots, y_0^i, \ldots, y_0^S]$ and applying nonlinear optimisation to find the minimum of the negative logarithm of the resulting posterior probability density function, i.e.

$$\hat{\theta} = \arg\min_{\theta \in \Theta} -\ln\left(p(\theta|\mathbf{Y}, \mathbf{y_0})\right) \tag{7}$$

With this formulation it is possible to perform MAP estimation on several data sets of varying length, but as special cases it is also possible to perform ML estimation on several data sets (with $p(\theta)$ uniform), MAP estimation on a single data set (with $S = 1$) and ML estimation on a single data set (with $p(\theta)$ uniform and $S = 1$).

When the unknown parameters of the model have been found using one of the above estimators, statistical tests and residual analysis can be performed. First of all, since the estimators are all asymptotically Gaussian the parameter estimates and their standard deviations can be used to perform marginal t-tests for parameter significance, i.e. to test if the parameters are significantly different from zero. This is particularly important for the process noise parameters, because parameters that are significantly different from zero indicate that the model structure is not perfect, i.e. that there may be approximation errors, unmodelled inputs or plant-model mismatch.

It is an inherent assumption of the above methods for estimation of parameters that the conditional probability densities are Gaussian, and for nonlinear systems this assumption is only likely to hold when small sample times are used, so the validity of this assumption should also be tested by performing a test for Gaussianity.

Finally it is possible to test if the model is correct by performing a goodness of fit test as shown by Bak *et al.* (1999) and by performing residual analysis. For the latter purpose both standard linear methods and nonlinear methods based on nonparametric modelling are available, cf. Nielsen and Madsen (2001).

For supporting decision-making within the modelling cycle a computer aided tool, CTSM, has been developed, cf. Kristensen and Madsen (2000). Within this tool a number of program units corresponding to the individual elements of the modelling cycle have been or will be implemented, including a graphical user interface for setting up the model structure and algorithms for estimating parameters and performing statistical tests and residual analysis. Altogether these program units aid the chemical or process systems engineer when setting up models.

3. EXAMPLE

The following is an example, which illustrates an important feature of the grey-box approach - the possibility of determining whether a given model structure is correct from estimates of the

process noise parameters.

The process considered is a simple fed-batch fermentation process described by an unstructured model, i.e.

$$\begin{pmatrix} dX \\ dS \\ dV \end{pmatrix} = \left(\begin{pmatrix} \mu(S)X \\ -\frac{\mu(S)X}{0.5} \\ 0 \end{pmatrix} + \begin{pmatrix} -\frac{X}{V} \\ \frac{(10-S)}{V} \\ 1 \end{pmatrix} F \right) dt + \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{bmatrix} d\omega_t \ , \ t \in [0, 3.8] \tag{8}$$

$$\begin{pmatrix} y_k^X \\ y_k^S \\ y_k^V \end{pmatrix} = \begin{pmatrix} X_k \\ S_k \\ V_k \end{pmatrix} + \begin{pmatrix} e_k^X \\ e_k^S \\ e_k^V \end{pmatrix} \tag{9}$$

where X and S are the concentrations of biomass and substrate, V is the volume of the fermenter and F is the feed flow rate, and finally $e_k^X \in N(0, 0.01)$, $e_k^S \in N(0, 0.001)$ and $e_k^V \in N(0, 0.01)$. For the growth rate $\mu(S)$ two different cases are considered, namely

- A correct model structure with $\mu(S) = \mu_{max} \frac{S}{K_1 S^2 + S + 0.5}$.
- An incorrect model structure with $\mu(S) = \tilde{\mu}_{max} \frac{S}{S + \tilde{K}_1}$.

corresponding to biomass growth with Monod kinetics and with and without substrate inhibition respectively.

Using the true parameter values in Table 1, 10 sets of simulation data (100 samples each with a sample time of 0.038) have been generated by perturbing the feed flow rate along an analytically determined optimal trajectory, and all the results mentioned in the following correspond to ML estimation of μ_{max} (or $\tilde{\mu}_{max}$), K_1 (or $\tilde{K}_1$), σ_1^2, σ_2^2, σ_3^2, $\sigma_{e^X}^2$, $\sigma_{e^S}^2$, $\sigma_{e^V}^2$ and the initial conditions using all 10 data sets.

Table 1

True and estimated values of the parameters of the fermentation process model. Upper part: Case 1 - correct structure of $\mu(S)$. Lower part: Case 2 - incorrect structure of $\mu(S)$.

Parameter	True value	Estimated value	Standard deviation	Significant
μ_{max}	1	1.021	0.0044	YES
K_1	0.03	0.03005	0.00139	YES
σ_1^2	0	4.026e-4	1.270e-4	NO
σ_2^2	0	1.365e-5	1.391e-5	NO
σ_3^2	0	3.100e-4	1.298e-4	NO
$\tilde{\mu}_{max}$	-	0.7661	0.0066	YES
$\tilde{K}_1$	-	0.01066	0.00007	YES
σ_1^2	0	0.05687	0.00369	YES
σ_2^2	0	0.08714	0.00935	YES
σ_3^2	0	0.002089	0.000167	YES

With the correct model structure, the parameter estimates in the upper part of Table 1 are obtained. The estimates of μ_{max} and K_1 are very accurate, and the estimates and standard deviations of σ_1^2, σ_2^2 and σ_3^2 indicate that these parameters are not significantly different from zero.

This is subsequently confirmed by performing t-tests, and this indicates that the model structure is indeed correct.

With the incorrect model structure, on the other hand, the parameter estimates in the lower part of Table 1 are obtained. Now σ_1^2, σ_2^2 and σ_3^2 are all significantly different from zero, indicating approximation errors, unmodelled inputs or, as in this case, plant-model mismatch.

4. CONCLUSION

A grey-box approach to process modelling that combines deterministic and stochastic modelling is advocated for identification of models for model-based control of batch and semi-batch processes, and a computer-aided tool designed for supporting decision-making in the corresponding modelling cycle has been presented.

The grey-box approach is based on flexible and statistically sound continuous-discrete stochastic state space models, which have the same appeal as ODE models with respect to their derivation from first engineering principles. One of the most important advantages of the approach is its built-in features for performing model validation by means of statistical tests and residual analysis, e.g. that the significance of the parameters of the process noise term may provide information about the validity of a proposed nominal model.

REFERENCES

Bak, Jakob, Henrik Madsen and Henrik Aalborg Nielsen (1999). Goodness of fit of stochastic differential equations. In: *Symposium i Anvendt Statistik* (Peter Linde and Anders Holm, Eds.).

Kristensen, Niels Rode and Henrik Madsen (2000). *CTSM - Continuous Time Stochastic Modeling - Version 1.0.* IMM, DTU, Lyngby, Denmark.

Nielsen, Henrik Aalborg and Henrik Madsen (2001). A generalization of some classical time series tools. *Computational Statistics and Data Analysis.* To appear.

Nielsen, Jan Nygaard, Henrik Madsen and Peter C. Young (2000). Parameter estimation in stochastic differential equations: An overview. *Annual Reviews in Control* **24**, 83–94.

Young, Peter C. (1981). Parameter estimation for continuous-time models - a survey. *Automatica* **17**(1), 23–39.

Åström, Karl Johan (1970). *Introduction to Stochastic Control Theory.* Academic Press, New York, USA.