BAYESIAN ESTIMATION BY SEQUENTIAL MONTE CARLO SAMPLING: APPLICATION TO HIGH-DIMENSIONAL NONLINEAR DYNAMIC SYSTEMS

Wen-shiang Chen [*] **Bhavik R. Bakshi** [*]
Prem K. Goel [**] **Sridhar Ungarala** [***]

[*] *Dept. Chem. Eng., The Ohio State University*
Columbus, OH 43210, USA
[**] *Dept. Statistics, The Ohio State University*
Columbus, OH 43210, USA
[***] *Dept. Chem. Eng., Cleveland State University*
Cleveland, OH 44115, USA

Abstract: Modern industrial enterprises have invested significant resources for collecting and distributing data, with the expectation that it will enhance profitability via better decision making. Due to the complexity of these problems, existing approaches tend to make convenient, but invalid assumptions so that tractable solution may be found. For example, for estimation in nonlinear dynamic systems, extended Kalman filtering (EKF) relies on Gaussian approximation and local linearization to find a closed-form solution. Moving horizon based least-squares estimation (MHE) also relies on Gaussian approximation, but the use of nonlinear models and constraints eliminates most of the computational benefits of this approximation, but can provide more accurate estimates than EKF. Unfortunately, in most practical nonlinear dynamic systems, the posterior distributions are often far from Gaussian, and continually change their shape.

Our previous work has developed rigorous Bayesian methods for estimation in nonlinear dynamic systems with constraints. These methods rely on recent theoretical developments in Sequential Monte Carlo Sampling (SMC). It does *not* rely on assumptions about the shape of the distributions, or nature of the models. Furthermore, this approach is expected to be computationally more efficient due to its recursive formulation that does not rely on nonlinear programming. These claims have been supported via applications to relatively small scale CSTR case studies. However, there are no illustrations or theoretical proofs to indicate how SMC performs for high-dimensional systems.

This paper applies our previous work on Bayesian rectification by SMC to large scale nonlinear dynamic systems and compares the computational efficiency and accuracy with MHE and EKF. The model of the selected polymerization reactor contains eight variables, exhibit significant linear and nonlinear dynamics under different operating conditions, and requires satisfaction of process constraints. Results indicate that the accuracy and computational benefits of SMC are significant even for such high-dimensional systems.

Keywords: Bayesian estimation, Sequential Monte Carlo sampling

1. INTRODUCTION

Efficient operation of chemical and manufacturing processes relies on cleaning or rectification of measured data and estimation of unknown quantities. Data rectification and estimation form the foundation for process operation tasks such as process control, fault detection and diagnosis, real-time estimation, process monitoring, and process scale-up. Since most processes are nonlinear and subject to constraints, significant efforts have already focused on methods for estimation of nonlinear dynamic systems (Kramer and Mah, 1994; Robertson *et al.*, 1996).

In our previous work, we have developed rigorous Bayesian methods for estimation in nonlinear dynamic systems. This approach relies on recent theoretical developments at the interface of statistical physics and Bayesian statistics. Unlike existing methods, this approach of sequential Monte Carlo sampling (SMC) does not rely on assumptions about the shape of the distributions, or nature of the models. Instead, it allows the distributions to adopt any shape according to system dynamics or constraints. Furthermore, the proposed approach is expected to be computationally efficient due to its recursive formulation. Previous results have illustrated that SMC shows significant improvement in estimation accuracy over existing methods, and in computation time over moving horizon based least-squares estimation (MHE) for cases with or without constraints (Chen *et al.*, 2004; Chen *et al.*, 2003*b*). However, these illustrations were based on relatively low-dimensional examples, and were not enough to address skepticism about the feasibility and advantages of SMC based Bayesian estimation for higher dimensional problems.

Such skepticism is well-founded for data-based methods due to the "curse of dimensionality", and may also apply to sampling based methods. Traditionally, Bayesian methods have not been feasible for high-dimensional systems since these methods usually require integration in high-dimensional space over a relatively fine grid, which leads to formidable requirement of computation. This experience leads to the perception and expectation that the computation load of sampling based methods might also increase exponentially thus making them impractical for most large scale problems.

This paper demonstrates that Bayesian estimation by SMC is feasible even for high-dimensional systems via application to a polymerization reactor. The selected process model is nonlinear, dynamic and involves eight state variables. Non-negativity constraints are also imposed on some variables. The estimation accuracy and computational efficiency of EKF, MHE and SMC are studied under a variety of conditions. These studies show that SMC can be significantly faster than MHE for similar estimation accuracy. Since the posterior distributions for the selected operating conditions may be approximated quite well as Gaussian distributions, the accuracy of SMC and MHE can be comparable. The case study indicates that Bayesian estimation via SMC may not be subject to the curse of dimensionality, and can retain its theoretical rigor and practical benefits for high-dimensional problems. The theoretical basis for this observation is also discussed in this paper.

In the following sections, a review of Bayesian estimation is first introduced. After that, the algorithm for SMC is provided. Performance of the proposed approach is compared with that of existing approaches in the case study section.

2. BAYESIAN ESTIMATION FOR DYNAMIC SYSTEMS

In general, the goal of estimation may be expressed as follows. Given measurements $y_{1:k} = \{y_1, y_2, \ldots, y_k\}$, process models, and the distribution of the initial condition $p(x_0)$, determine the current state, x_k. Process models may be expressed as follows,

$$x_k = f_{k-1}(x_{k-1}, \omega_{k-1}) \tag{1}$$

$$y_k = h_k(x_k, \nu_k) \tag{2}$$

where $x_k \in \Re^{n_x}$ is the state vector and $f_k : \Re^{n_x} \times \Re^{n_\omega} \rightarrow \Re^{n_x}$ is the system equation. Measurements, $y_k \in \Re^{n_y}$, are related to the state vector through the measurement equation, $h_k : \Re^{n_x} \times \Re^{n_\nu} \rightarrow \Re^{n_y}$.

Bayesian estimation provides a rigorous way of maximizing the use of all available information and can handle all types of errors, models and constraints. All variables are considered to be stochastic, and previous belief can be combined with the current measurements via Bayes rule.

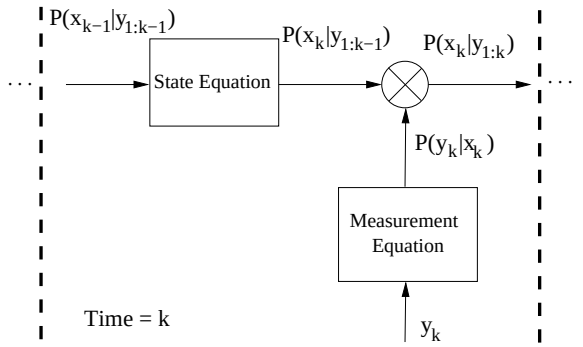

Fig. 1. Algorithm of recursive Bayesian estimation.

The recursive formulation of Bayesian estimation has been proposed decades ago and may be represented as follows (Ho and Lee, 1964),

$$p(x_k|y_{1:k}) = \frac{p(y_k|x_k)\ p(x_k|y_{1:k-1})}{p(y_k|y_{1:k-1})}, \quad (3)$$

where $p(x_k|y_{1:k})$ is the posterior. $p(x_k|y_{1:k-1})$ may be found by solving the following two equations.

$$p(x_k|y_{1:k-1}) = \int p(x_k|x_{k-1})\ p(x_{k-1}|y_{1:k-1})\ dx_{k-1} \quad (4)$$

$$p(x_k|x_{k-1}) = \int \delta(x_k - f_{k-1}(x_{k-1}, \omega_{k-1}))\ p(\omega_{k-1})\ d\omega_{k-1} \quad (5)$$

Likelihood distribution may be found as follows.

$$p(y_k|x_k) = \int \delta(y_k - h_k(x_k, \nu_k))\ p(\nu_k)\ d\nu_k \quad (6)$$

Further detail of the derivation of these equations can be found in Chen *et al.* (2004).

The algorithm for recursive Bayesian estimation may be visualized as in Figure 1. Information in previous measurements up to time $k-1$ is captured by the prior, $p(x_{k-1}|y_{1:k-1})$. Prediction of distribution of the current state is implemented by utilizing Equations (4) and (5). Information in current measurement is represented as the likelihood function based on Equation (6). The posterior can then be found by combining previous and current information by Equation (3).

After the posterior is available, a point estimate and its uncertainty may be obtained via,

$$E[\mathcal{L}(x_k)] = \int \mathcal{L}(x_k)\ p(x_k|y_{1:k})\ dx_k, \quad (7)$$

where $\mathcal{L}(\cdot)$ is a chosen loss or criterion function (Jazwinski, 1970). Mean, mode or median of the posterior are popular choices for the point estimate.

In general, there is no closed-form solution for Equations (4) to (6). Direct integration is computationally expensive and may not be practical for high-dimensional systems. Most estimation methods address these computational challenges by making simplifying assumptions about the nature of the model and/or posterior distributions at the cost of accuracy and computational efficiency. However, recent theoretical advances coupled with fast computation are providing the foundation of building a feasible Bayesian approach even for large scale systems. This computationally efficient algorithm is based on sequential Monte Carlo sampling and will be discussed in Section 3.

2.1 Bayesian view of existing methods

Existing methods may be described as special cases of Bayesian estimation. EKF assumes Gaussian prior and additive Gaussian noise and relies on local linearization to have closed-form solution (Jazwinski, 1970). MHE also relies on the assumption of Gaussian prior and additive Gaussian noise to have the least-squares formulation (Robertson *et al.*, 1996). MHE may be more accurate results than EKF by keeping the nonlinearity and being able to handle constraints. However, a closed-form solution is not available anymore. Instead, MHE needs to solve a constrained nonlinear programming problem over each moving window and lacks a recursive formulation. This makes MHE a computationally demanding algorithm. In the presence of constraints, MHE implicitly uses truncated Gaussian prior. Modifications via approximating truncated Gaussian distributions with combination of Gaussian or other fixed-shape distributions have also been suggested (Robertson and Lee, 2002). However, a general algorithm to decide the optimal number of Gaussian distributions for approximation is not easy to find. Further discussion of EKF and MHE with SMC can be found in Chen *et al.* (2004).

3. BAYESIAN ESTIMATION BY SEQUENTIAL MONTE CARLO SAMPLING

Monte Carlo sampling based Bayesian approach has been an active area of research for a few years (Gordon *et al.*, 1993; Doucet *et al.*, 2000; Andrieu *et al.*, 2003). Existing methods may be categorized into two groups, sequential Monte Carlo sampling (SMC) and Markov chain Monte Carlo sampling (MCMC). Both SMC and MCMC use Monte Carlo sampling for its convenience in computing the properties of distributions from available samples. MCMC employs iterative algorithm for generating samples, while SMC draws samples from an importance function and adjusts samples'

importance with weight. In this paper, SMC is favored for its computational efficiency.

The algorithm for SMC may be represented in pseudo-code as follows (Chen *et al.*, 2003*b*):

- FOR times $k = 1, 2, 3, \ldots$
  - · FOR samples $i = 1, 2, 3, \ldots, N$
    - Draw sample, $x_k(i)$ from an importance function, $\pi(x_k|x_{k-1}(i), y_k)$
    - Enforce constraints by acceptance and rejection algorithm
    - Assign a weight to $x_k(i)$, $q_k^*(i)$
  - · END FOR
  - · Normalize $q_k^*(i)$ to find $q_k(i)$
  - · Implement resampling when necessary
  - · Find $E[\phi_k(x_k)] \approx \sum_{l=1}^{N} \phi_k(x_k(l)) q_k(l)$
- END FOR

where

$$q_k^*(i) = q_{k-1}(i) \frac{p(y_k|x_k(i))\ p(x_k(i)|x_{k-1}(i))}{\pi(x_k(i)|x_{k-1}(i), y_k)} \quad (8)$$

A convenient choice of importance function is to use samples of prior as the importance function (Gordon *et al.*, 1993),

$$\pi(x_k(i)|x_{k-1}(i), y_k) = p(x_k(i)|x_{k-1}(i)) \quad (9)$$

This choice simplifies Equation (8) to

$$q_k^*(i) = q_{k-1}(i)\ p(y_k|x_k(i)) \quad (10)$$

More sophisticated choice of importance functions is expected to improve the robustness of SMC (Doucet *et al.*, 2000).

Various issues of applying SMC have been found and practical solutions are available. SMC may develop degeneracy after several time steps when most samples have insignificant importance to the distribution. Degeneracy can be easily improved by performing resampling Chen *et al.* (2004). SMC may also encounter slow initial convergence when poor initial guess is used. Slow initial convergence can be significantly improved via empirically estimating the initial state (Chen *et al.*, 2004).

3.1 SMC of High-Dimensional Problems

The computational challenges faced by direct computation of Bayesian integrals (Equations (4) to (6)) are formidable for low-dimensional problems, and could become intractable for high-dimensional cases. As discussed earlier, this computational challenge has motivated the application of Monte Carlo sampling for Bayesian estimation. However, the feasibility of SMC for high-dimensional problems is not obvious since the number of samples to maintain equal density or coverage in the solution space increases exponentially with increasing dimensionality. Furthermore, since more samples are required to fill the solution space, it is possible that SMC faces more frequent and serious encounters with various practical challenges like, degeneracy or/and slow initial convergence. It is not clear whether Monte Carlo sampling based approaches suffer more from the curse of dimensionality than other competing approaches like EKF and MHE. Recent theoretical research has been addressing these and other related issues. The key relevant findings are summarized in this subsection based on the survey of Crisan and Doucet (2002).

Theoretical studies indicate that the convergence of SMC in terms of mean-squares error toward zero is almost sure when the importance weights, $\{q_k(i)\}$, are upper bounded, and when a standard resampling scheme is applied. Under these slightly restrictive conditions, the convergence rate is inversely proportional to the number of samples, and is independent of the dimensionality of the problem. Even though more studies are needed to explore the existence of a general convergence property, the number of samples required for SMC in high-dimensional systems may not have to be exponentially increasing. Thus the computation load of SMC may not have to be increased dramatically to obtain accurate estimates. Furthermore, the convergence of SMC may be independent or weakly dependent on the dimensionality.

Another benefit of SMC is that it can be readily implemented in parallel. This is because each sample can be computed independently for most of the computation algorithm. Parts of the algorithm that can not be incorporated into the parallel computation include the occasional implementation of resampling, and the inference from the posterior. Thus, a parallel implementation can retain the computational efficiency even when large number of samples become necessary.

Recent theoretical development shows that the convergence of SMC may be independent of the dimension of the problem, and is inversely proportional to the number of samples. In addition, the recursive formulation also reduces the complexity of the problem compared to a moving horizon formulation. The computational efficiency of SMC may actually be practical for high-dimensional cases, and our results confirm the expectation.

4. CASE STUDY

Start-up of a polymerization process in a continuous stirred tank reactor (CSTR) is studied (Tatiraju *et al.*, 1999). Monomer styrene, initiator, and solvent are continuously fed into the

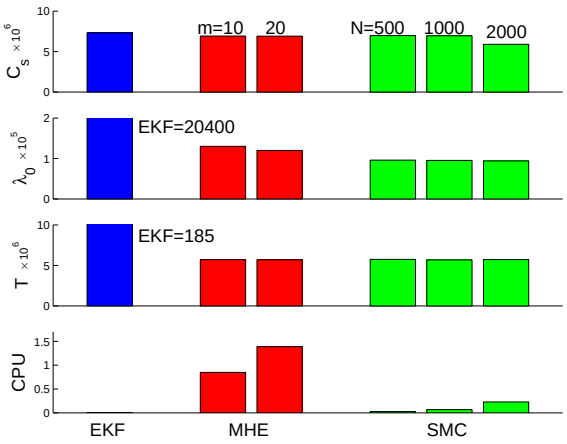

Fig. 2. Performance comparison of Polymerization case study. Top three plots illustrate MSE results, while the bottom one shows CPU load for each method. Not all results are presented for clarity.

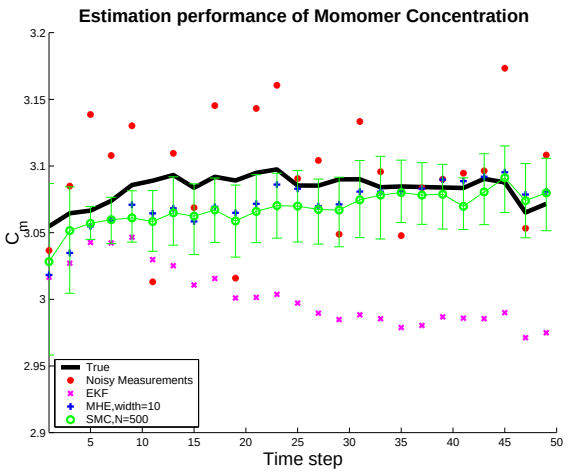

Fig. 3. Estimation results of polymerization case. Estimates provided by SMC are bounded with 95% error bar.

reactor with constant flow rates and compositions. The system has eight state variables, including concentrations of initiator, solvent, and monomer in the CSTR, temperatures in the CSTR and in the cooling device, and first three moments of the molecular weight distribution. Finite difference is used to simulate the true states. Nonlinear process model and simulation conditions and parameters can be found in Chen *et al.* (2003*a*).

All system states are assumed to be measurable and both states and measurements are contaminated with independent and identically distributed Gaussian noise. Estimation performance compares three estimation methods, EKF, MHE and SMC, on the accuracy and efficiency of estimation with noisy measurements. Non-negative constraints on all state variables are enforced. Results provided in Figure 2 and Table 1 are based on 100 realizations of simulation by Octave on a personal computer with Pentium 2.0 GHz and 512MB RAM. 50 measurements are rectified in each realization. Accuracy is compared based on mean-squares error (MSE). For comparison purpose, point estimate is chosen for SMC. Mean is chosen simply for computational convenience. MHE is based on the codes by James Rawlings' group at University of Wisconsin. No systematic way of finding the proper window size of MHE exists, and usually the number of state variables is used as the horizon width. CPU time required is also studied (in units of CPU seconds per time step) with SMC by uncompiled code, while MHE is based on compiled code.

The proposed approach, SMC, exhibits significant efficiency in data rectification of a high-dimensional problem. SMC easily outperforms EKF in MSE since no linearization is required. SMC and MHE has competitive performance on MSE since the posterior exhibits Gaussian like features. However, SMC requires less than 7% of computation than MHE even when SMC is based on uncompiled code. This result indicates that methods based on Gaussian approximation need not be computationally more efficient than methods based on other distributions.

Bayesian estimation can readily provide uncertainty information. Figure 3 shows SMC estimate with 95% error bar. Estimating the error bounds or any moment of the posterior is straightforward and can be implemented by choosing proper loss function, $\mathcal{L}(\cdot)$, in Equation (7).

5. CONCLUSION AND DISCUSSIONS

This paper extends our previous work on Bayesian estimation by sequential Monte Carlo sampling (SMC) to high-dimensional systems. The formulation is general and has shown significantly better performance for a variety of systems including, linear and nonlinear, Gaussian and non-Gaussian, constrained and non-constrained systems. To date, most applications were to relatively small scale problems (Chen *et al.*, 2004; Chen *et al.*, 2003*b*). Application to high-dimensional systems is essential to identify and address any new challenges, and to support claims about the practicality of the proposed approach.

CPU results indicate that SMC may not suffer severely from the curse of dimensionality as commonly expected. Rectification performance on an eight state variables nonlinear dynamic systems shows that SMC requires less than 7% of computation load than MHE while providing competitive estimation accuracy. The accuracy of SMC can be much better than MHE and EKF for systems with highly non-Gaussian posterior distributions. In this example, estimates of MHE and SMC have

Table 1. Mean-Squares Error and CPU time for Polymerization Case Study. Not All Results Are Presented for Clarity.

	$MSEC_s$	$MSE\lambda_0$	$MSEC_m$	CPU	Parameter
EKF	$7.3 \times 10^{-6} \pm 6.7 \times 10^{-6}$	$2.0 \times 10^{-1} \pm 2.7 \times 10^{-3}$	$1.8 \times 10^{-4} \pm 5.5 \times 10^{-5}$	0.003	N/A
MHE	$6.9 \times 10^{-6} \pm 6.6 \times 10^{-6}$	$1.3 \times 10^{-5} \pm 9.8 \times 10^{-6}$	$5.7 \times 10^{-6} \pm 6.6 \times 10^{-6}$	0.85	$m = 10$
	$6.9 \times 10^{-6} \pm 6.6 \times 10^{-6}$	$1.2 \times 10^{-5} \pm 9.6 \times 10^{-6}$	$5.7 \times 10^{-6} \pm 6.6 \times 10^{-6}$	1.39	$m = 20$
SMC	$7.0 \times 10^{-6} \pm 6.6 \times 10^{-6}$	$9.6 \times 10^{-6} \pm 1.0 \times 10^{-5}$	$5.7 \times 10^{-6} \pm 6.5 \times 10^{-6}$	0.03	$N = 500$
	$7.0 \times 10^{-6} \pm 6.6 \times 10^{-6}$	$9.5 \times 10^{-6} \pm 1.0 \times 10^{-5}$	$5.7 \times 10^{-6} \pm 6.3 \times 10^{-6}$	0.07	$N = 1000$
	$6.9 \times 10^{-6} \pm 6.6 \times 10^{-6}$	$9.4 \times 10^{-6} \pm 1.0 \times 10^{-5}$	$5.7 \times 10^{-6} \pm 6.7 \times 10^{-6}$	0.23	$N = 2000$

comparable accuracy which results from the Gaussian like priors of the system under the operating condition studied. SMC has shown significant improvement over MHE in situations when non-Gaussian priors exist with or without constraints (Chen *et al.*, 2004; Chen *et al.*, 2003*b*). Other operating conditions with non-Gaussian distributions are currently being explored.

The results in this paper compare point estimates of SMC with other methods. Since details about the posterior distribution at each time point are available in SMC, choice of the loss function can have a significant effect on the accuracy of the point estimate. The use of the mean value, as in current work may not be appropriate, particularly for multi-modal distributions. Such a point estimate may not utilize all the benefits of Bayesian estimation, and its use for performance comparison with other methods that rely on Gaussian approximation may be biased against SMC. Research on appropriate loss functions for non-Gaussian distributions is on-going for a better and fairer portrayal of the benefits of SMC-based Bayesian estimation.

6. ACKNOWLEDGMENTS

Special thanks to Eric L. Haseltine and Dr. James Rawlings at the University of Wisconsin for providing MHE code and assistance in running the program. Financial support from the National Science Foundation (CTS-9733627 and CTS-0321911) is also gratefully acknowledged.

REFERENCES

Andrieu, Christophe, Nando de Freitas, Arnaud Doucet and Michael Jordan (2003). An introduction to MCMC for machine learning. *Machine Learning* **50**(1-2), 5–43.

Chen, Wen-shiang, Bhavik R. Bakshi, Prem K. Goel and Sridhar Ungarala (2003*a*). Bayesian estimation by sequential Monte Carlo sampling: Application to large-scale nonlinear dynamic systems. Technical report. Department of Chemical Engineering, The Ohio State University.

Chen, Wen-shiang, Bhavik R. Bakshi, Prem K. Goel and Sridhar Ungarala (2003*b*). Bayesian estimation of nonlinear dynamic systems - dealing with constraints. Technical report. Department of Chemical Engineering, The Ohio State University.

Chen, Wen-shiang, Bhavik R. Bakshi, Prem K. Goel and Sridhar Ungarala (2004). Bayesian estimation of unconstrained nonlinear dynamic systems via sequential Monte Carlo sampling. *Industrial & Engineering Chemistry Research*. Submitted for publication.

Crisan, Dan and Arnaud Doucet (2002). A survey of convergence results on particle filtering methods for practitioners. *IEEE Transactions on Signal Processing* **50**(3), 736–746.

Doucet, Arnaud, Simon Godsill and Christophe Andrieu (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing* **10**, 197–208.

Gordon, N. J., D. J. Salmond and A. F. M. Smith (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings-F* **140**(2), 107–113.

Ho, Y. C. and R. C. K. Lee (1964). A Bayesian approach to problems in stochastic estimation and control. *IEEE Transactions on Automatic Control* pp. 333–339.

Jazwinski, Andrew H. (1970). *Stochastic Processes and filtering theory*. Academic Press. New York.

Kramer, M. A. and R. S. H. Mah (1994). Model-based monitoring. In: *Proceedings of the International Conference on Foundations of Computer Aided Process Operations* (D. Rippin, J. Hale and J. Davis, Eds.). CACHE. Austin, TX. pp. 45–68.

Robertson, Douglas G. and Jay H. Lee (2002). On the use of constraints in least squares estimation and control. *Automatica* **38**, 1113–1123.

Robertson, Douglas G., Jay H. Lee and James B. Rawlings (1996). A moving horizon-based approach for least-squares estimation. *AIChE Journal* **42**(8), 2209–2224.

Tatiraju, Srinivas, Masoud Soroush and Babatunde Ogunnaike (1999). Multirate nonlinear state estimation with application to a polymerization reactor. *AIChE Journal* **45**(4), 769–780.