BAYESIAN ESTIMATION OF UNCONSTRAINED 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: Accurate estimation of state variables and model parameters is essential for efficient process operation. The Bayesian formulation of the estimation problem suggests a general solution for nonlinear systems. However, a practically feasible implementation of the solution has not been available until recently. Most existing methods have had to rely on simplifying assumptions to obtain an approximate solution. For example, extended Kalman filtering estimates the system state by linearizing the nonlinear model and assuming Gaussian distributions for all random variables. Moving horizon estimation assumes Gaussian or other fixed-shape distributions to formulate a constrained least-squares optimization problem. In this paper, Bayesian estimation is implemented by sequential Monte Carlo sampling. This approach can represent non-Gaussian distributions accurately and efficiently with minimum assumptions and computes moments by Monte Carlo integration. The features of the Monte Carlo approach are demonstrated by application to a state estimation case study of a CSTR process. The proposed method exhibits 78% improvement in estimation error and takes 95% less time than moving horizon estimation to solve the problem.

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. Due to the importance of these tasks, many methods have been developed under the names of data rectification, data reconciliation, and state and parameter estimation (Kramer and Mah, 1994; Robertson *et al.*, 1996).

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, $h_k : \Re^{n_x} \times \Re^{n_\nu} \rightarrow \Re^{n_y}$.

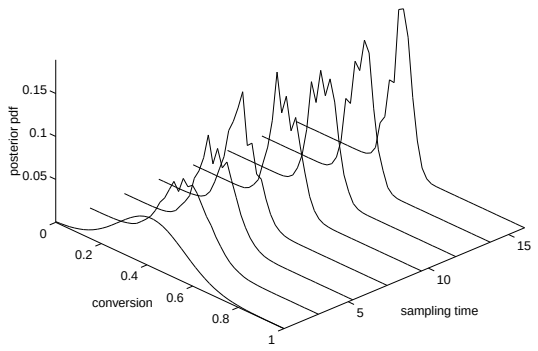

Fig. 1. Evolution of the conditional distribution of concentration in a CSTR.

Significant efforts have been focused on methods for rectification and estimation in nonlinear dynamic systems, with and without constraints (Jang *et al.*, 1986; Tjoa and Biegler, 1991; Liebman *et al.*, 1992; Robertson *et al.*, 1996; Rao and Rawlings, 2002). However, all the existing methods rely on simplifying assumptions about the nature of the model or the probability distributions of the underlying variables to obtain a tractable optimization problem. A popular assumption is that the distribution of the variables to be estimated is Gaussian or of a fixed, time-invariant shape. The crudeness of this assumption is depicted in Figure 1, which shows the conditional distribution over time for a popular continuously stirred tank reactor (CSTR) case study (Jang *et al.*, 1986; Liebman *et al.*, 1992; Robertson *et al.*, 1996). The multi-modal, skewed and time-varying nature of these distributions indicates that approximating them by Gaussian or other fixed-shape distributions can be grossly incorrect. The approximations may also fail in the presence of constraints, since constraints may require the probability of some variables to be zero in regions where the constraint is violated (Robertson *et al.*, 1996; Rao and Rawlings, 2000; Chen *et al.*, 2002; Robertson and Lee, 2002). Nevertheless, these assumptions are popular since they permit existing methods to solve a convenient problem instead of the actual estimation problem. These shortcomings of existing methods and the challenges in obtaining the Bayesian solution are well-known and have been widely recognized (Robertson *et al.*, 1996; Rao and Rawlings, 2000; Robertson and Lee, 2002).

The Bayesian formulation provides a solution to the actual estimation problem without necessitating invalid assumptions. However, until recently, the implementation of the Bayesian solution was considered impractical due to its heavy computational demand. Recently, efficient algorithms based on Monte Carlo sampling along with increasing computational ability are making Bayesian estimation feasible for real problems (Malakoff, 1999).

This paper introduces a computationally efficient approach for data rectification of nonlinear dynamic systems based on a statistically rigorous Bayesian formulation. This approach relies on sequential Monte Carlo (SMC) sampling to maximize the use of data and knowledge to obtain the Bayesian solution without relying on assumptions about the nature of the errors, model and underlying variables. The main contributions of this paper are to introduce SMC methods into process engineering, and compare their performance with currently popular methods. This work also indicates that for many nonlinear dynamic systems, Gaussian approximations are not necessarily more computationally efficient, and may be less accurate.

In the following sections, a Bayesian view of existing methods is first discussed. After that, a brief introduction on Monte Carlo sampling is provided. Then the detail of the implementation of the proposed approach is provided and discussed. Performance of the proposed approach is compared with that of existing approaches in the case study section.

2. BAYESIAN VIEW OF EXISTING METHODS

2.1 Background

Bayesian estimation maximizes the use of all available information and can handle all types of errors, models and constraints. In addition, Bayesian estimation finds the distribution of states, which can provide uncertainty information. For dynamic systems, recursive Bayesian estimation may be represented as follows (Ho and Lee, 1964),

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

where the posterior distribution, $p(x_k|y_{1:k})$, combines information from current measurement via the likelihood function $p(y_k|x_k)$, and past information using the prior distribution $p(x_k|y_{1:k-1})$. The denominator is a normalizing constant. Each term in Equation (3) may be obtained as follows. For the second term in the numerator,

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

where $p(x_{k-1}|y_{1:k-1})$ is the posterior of time step $k-1$. $p(x_k|x_{k-1})$ may be further manipulated as the following equation.

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

Similarly, $p(y_k|x_k)$ in Equation (3) may be found as follows,

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

In general, there is no closed-form solution for Equation (3) to Equation (6) except for linear Gaussian systems. Even when the functionality of the distributions is known, the calculation of their moments needs multi-dimensional integrations, which may be computationally expensive. Therefore, it is not surprising to see that methods based on simplification have been popular in the past. Many existing approaches may be interpreted as approximate Bayesian estimation. These methods tend to simplify the real problem so that a convenient solution may be found. In the following section, an overview of existing approaches for estimation from the view point of Bayesian estimation is provided. These methods may be categorized by how posterior distributions are propagated over time and their moments are computed. In the following sections, two main categories are discussed, Gaussian approximation and direct integration.

2.2 Gaussian approximation

Gaussian distributions are convenient since only two parameters, mean and variance, are required to describe a whole distribution. Although, the assumption of Gaussian prior is suitable in linear systems, it can be easily violated in nonlinear dynamic systems (Chen *et al.*, 2003). The assumption of Gaussian prior worsens when process constraints are enforced and results in truncated distributions (Chen *et al.*, 2002; Robertson and Lee, 2002). Even though Gaussian approximation may not be a valid assumption for nonlinear dynamic systems, approaches based on this assumption are popular for its simplicity. Two variations of Gaussian approximation namely, extended Kalman filtering (EKF) and moving-horizon estimation (MHE), are discussed here.

EKF is an extension of Kalman filtering to nonlinear dynamic systems. Kalman filtering is the optimal estimator for linear dynamic systems with Gaussian prior and additive independent and identically distributed (iid) Gaussian noise without constraints. The filter is optimal with respect to minimum variance criterion. In addition, Kalman filtering has a closed-form solution that makes estimation extremely efficient. The natural extension of Kalman filtering into nonlinear dynamic systems is to linearize nonlinear process models so that the same solution strategy for Kalman filtering can be applied. In doing so, EKF inherits all assumptions made by Kalman filtering, including Gaussian prior and noise. EKF is favored for its simplicity and efficiency, but the filter may diverge from the true state and does not necessarily satisfy process constraints. Further discussion of Kalman filtering and EKF can be found in Jazwinski (1970) and Maybeck (1979).

Efforts have been made to avoid divergence of EKF. One suggestion is to retain higher order terms of Taylor's expansion so that more accurate local linearization may be achieved. Divergence due to poor approximation may be reduced, but new complexity arises in determining the "right" highest term to keep, which may not be a trivial task.

MHE also relies on the assumption of Gaussian prior and noise so that a least-squares estimation (LSE) may be found (Robertson *et al.*, 1996). Unlike EKF, MHE can enforce constraints which is equivalent to using truncated Gaussian prior (Robertson and Lee, 2002). MHE also needs selection of a proper window size to compromise between the accuracy of batch-processed least-squares estimation and the efficiency of solving a smaller problem. Furthermore, MHE relies on constrained nonlinear programming, which is usually computationally expensive and it becomes difficult to asses its statistical properties. Even in cases where Gaussian approximation may be an acceptable assumption, the proposed Bayesian approach usually has better accuracy and tends to require less computation than approaches like MHE.

2.3 Direct Numerical Integration

Methods in this category represent the distribution of interest over a grid of points in state space. Once a suitable grid is identified, numerical integration may be used to compute the moments of the distribution. This approach can provide the exact solution if the state space is discrete and finite. In most cases, the number of states is not finite, and selecting the grid can be quite challenging since a fine grid is computationally expensive, while a coarse grid may be inaccurate. Many variations have been developed based on fixed or adaptive grids. Approaches such as cell-to-cell mapping and Hidden Markov Models may be considered to be special cases of this approach. While this approach has become more feasible with advances in computing, it is still too expensive for solving multidimensional problems.

3. MONTE CARLO SAMPLING

Monte Carlo sampling based approaches use samples to approximate a distribution as,

$$p(x) \approx \sum_{i=1}^{N} q(i)\ \delta(x - x(i)) \tag{7}$$

where $x(i)$ is the i-th sample that represents the distribution. The coefficient, $q(i)$, is the probability mass associated with each sample. $q(i)$ equals $1/N$ for $x(i)$ randomly drawn from $p(x)$. By the law of large numbers, as the number of samples goes to infinity, the approximation converges to the exact distribution.

Integration based on Monte Carlo sampling may be expressed as,

$$E[\phi(x)] = \int \phi(x)\ p(x)dx$$
$$\approx \frac{1}{N}\sum_{i=1}^{N} \phi(x(i)) \quad (8)$$

where $x(i)$ again is the i-th sample drawn from the distribution $p(x)$.

Estimation based on Equation (8) relies on samples drawn from the known distribution, $p(x)$. In real problems, $p(x)$ may not be readily available for sampling, but its value can be evaluated for a given sample of $x(i)$. This leads to the use of importance sampling.

3.1 Importance Sampling

Importance sampling relaxes the requirement of generating samples from the true distribution for estimating Equation (8). Instead, it relies on drawing samples from a convenient distribution, $\pi(x)$, called the importance function. Equation (8) may be reformulated as,

$$E[\phi(x)] = \int \phi(x)\ p(x)\ dx$$
$$= \int \frac{\phi(x)\ p(x)}{\pi(x)}\ \pi(x)\ dx$$
$$\approx \frac{1}{N}\sum_{i=1}^{N} q(i) \quad (9)$$

where $q(i) = \frac{\phi(x(i))\ p(x(i))}{\pi(x(i))}$ is the weight function. It should be noted here that $x(i)$ are samples drawn from $\pi(x)$ instead of $p(x)$. Convergence is almost guaranteed under minimal assumptions, such as, the support of $\pi(x)$, contains the support of $p(x)$ (Geweke, 1989).

For dynamic systems, Monte Carlo sampling techniques can be implemented recursively when new measurements arrive, and is termed sequential Monte Carlo sampling. The following section describes a Bayesian estimation approach based on sequential Monte Carlo sampling.

4. SEQUENTIAL MONTE CARLO SAMPLING FOR BAYESIAN ESTIMATION

The goal of Bayesian estimation is to obtain the posterior accurately and efficiently. The algorithm for recursive Bayesian estimation may be visualized as in Figure 2. Information in previous measurements up to time $k-1$ is captured by the posterior distribution, $p(x_{k-1}|y_{1:k-1})$. 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).

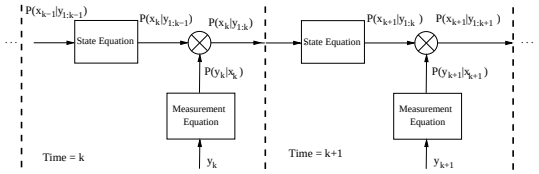

Fig. 2. Algorithm of recursive Bayesian estimation.

The application of sequential Monte Carlo sampling may be described as finding the appropriate weight for each sample so that posterior distribution may be approximated by the samples as in Equation (7). The algorithm may be represented in pseudo-code as follows (Arulampalam *et al.*, 2002):

- FOR times $k = 1,2,3,\ldots$
  - FOR samples $i = 1,2,3,\ldots,N$
    - Draw sample, $x_k(i)$ from an importance function, $\pi(x_k(i)|x_{k-1}(i),y_k)$
    - Assign a weight to $x_k(i)$, $q_k^*(i)$
  - END FOR
  - Normalize $q_k^*(i)$ to find $q_k(i)$
- END FOR

where

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

can be found based on Equations (3) to (6).

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

$$\pi(x_k(i)|x_{k-1}(i),y_k) = p(x_k(i)|x_{k-1}(i)) \quad (11)$$

This choice simplifies Equation (10) to

$$q_k^*(i) = q_{k-1}(i)\ p(y_k|x_k(i)) \quad (12)$$

More sophisticated choice of importance functions is expected to improve the robustness of SMC (Doucet *et al.*, 2000; Cheng and Druzdzel, 2000). In the next section, one practical issue in applying the proposed approach to estimation problem, known as degeneracy, is discussed.

4.1 Degeneracy

Degeneracy is a phenomenon where the weights of most samples become insignificant after a few time steps. Therefore, computation may be wasted on samples with little or no importance to the distribution. In addition, since estimation is mainly determined by a few samples, approximation of distributions may result in spurious spikes.

Degeneracy may be reduced by choosing importance functions that minimize the variance of sample weights. $\pi(x_k(i)|x_{k-1}(i),y_k)$, has been suggested as one such importance function (Doucet *et al.*, 2000).

Cheng and Druzdzel (2000) have also suggested an adaptive algorithm for finding importance functions, which is more robust when unlikely measurements occur. This approach to updating importance functions may help in reducing degeneracy since it tends to become severe when the measurement and prediction do not match each other. Markov chain Monte Carlo (MCMC) sampling which may be interpreted as iterative process of finding importance functions may also reduce degeneracy (Andrieu *et al.*, 2003).

Degeneracy can also be reduced by resampling. Resampling involves drawing samples from the weighted sample pool according to samples' weights. Samples with insignificant weights are less likely to be resampled. Further discussion of resampling can be found in Chen *et al.* (2003).

5. CASE STUDY

A typical chemical engineering problem, an adiabatic CSTR, is studied. Governing equations for this CSTR case study are provided as follows (Jang *et al.*, 1986; Liebman *et al.*, 1992; Henson and Seborg, 1997; Robertson and Lee, 1995).

$$\frac{dC}{dt} = \frac{q}{V}(C_0 - C) - k\,C\,e^{\frac{-E_A}{T}} \tag{13}$$

$$\frac{dT}{dt} = \frac{q}{V}(T_0 - T) - \frac{\Delta H}{\rho\,C_p}\,k\,C\,e^{\frac{-E_A}{T}} - \frac{U\,A}{\rho\,C_p\,V}(T - T_c) \tag{14}$$

Operating conditions and simulation parameters can be found in Henson and Seborg (1997). Three estimation approaches are compared in this case study, including EKF, MHE and SMC. MHE is implemented with horizon width 2, while 500 samples are used at each time step for SMC.

Figure 3 displays the evolution of posterior distribution of concentration into skewed non-Gaussian distributions, approximated by SMC with 5000 samples. Figure 4 shows the multi-modal posterior distributions.

Performance of these three methods is compared based on mean-squares error (MSE) and CPU time required for estimation (in units of CPU seconds per time step). Results provided in Table 1 are based on 100 realizations of simulation, and in each realization, 1600 measurements are rectified. CPU time is based on a personal computer with Pentium 400 MHz and 128MB RAM.

The proposed approach, SMC, exhibits significant improvement over both EKF and MHE in estimation error. SMC shows 78% improvement over MHE, and confirms the expectation that without making invalid assumptions on distributions, estimation by SMC is more accurate. Although MHE has better estimation results than EKF, it requires 19 times more computational effort than SMC. Verification of this result using customized MHE is also under progress. This result indicates that methods based on Gaussian approximation need not be computationally more efficient than methods based on other distributions.

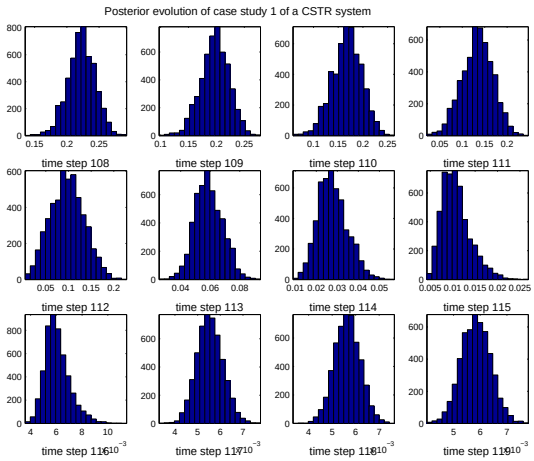

Fig. 3. Skewed posterior distributions: time step 109 and 115.

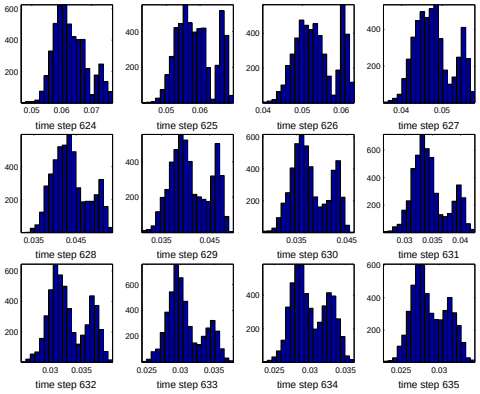

Fig. 4. Multi-modal posterior distributions.

Table 1. Average Mean-Squares Error and CPU time for CSTR Case Study

	EKF	MHE	SMC
MSE	0.13 ± 0.05	0.09 ± 0.04	0.02 ± 0.01
CPU	0.002 ± 0.00	0.58 ± 0.22	0.03 ± 0.01
Parameters		width$= 2$	$N = 500$

6. CONCLUSIONS

In this paper, a novel estimation approach based on a rigorous Bayesian formulation is introduced. The proposed approach uses sequential Monte Carlo sampling to propagate state information recursively. The Monte Carlo approach avoids direct numerical integration for computing the moments of state probability distributions. SMC benefits from not making invalid assumptions, such as Gaussian or other fixed-shape prior and

noise, compared with most existing approaches. SMC is shown to outperform EKF by a wide margin in accuracy. It outperforms MHE in terms of accuracy and computation time even when the distributions satisfy the MHE assumption of being Gaussian. The benefits of this proposed approach are expected to be even more significant for constrained nonlinear dynamic systems (Chen *et al.*, 2002). The proposed approach can handle all types of errors, models and constraints with the same solution strategy.

7. ACKNOWLEDGMENTS

Financial support from the National Science Foundation (CTS-9733627) is 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.

Arulampalam, M. Sanjeev, Simon Maskell, Neil Gordon and Tim Clapp (2002). A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing* **50**(2), 174–188.

Chen, Wen-shiang, Bhavik R. Bakshi, Prem K. Goel and Sridhar Ungarala (2002). Bayesian estimation of nonlinear dynamic systems - dealing with constraints and non-Gaussian errors. In: *AIChE Annual Meeting, Indianapolis, IN*. `http://www.che.eng.ohio-state.edu/~chenwe/publications/aiche2002_chen.pdf`.

Chen, Wen-shiang, Bhavik R. Bakshi, Prem K. Goel and Sridhar Ungarala (2003). Bayesian estimation of unconstrained nonlinear dynamic systems via sequential Monte Carlo sampling. Technical report. Department of Chemical Engineering, The Ohio State University. `http://www.che.eng.ohio-state.edu/~chenwe/publications/BayMC_Intro.pdf`.

Cheng, Jian and Marek J. Druzdzel (2000). Aisbn: An adaptive importance sampling algorithm for evidential reasoning in large bayesian networks. *Journal of Artificial Intelligence Research* **13**, 155–188.

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

Geweke, John (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica* **57**(6), 1317–1339.

Gordon, N. J. and 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.

Henson, Michael A. and Dale E. Seborg (1997). *Nonlinear Process Control*. Upper Saddle River, New Jersey; Prentice Hall PTR.

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.

Jang, Shi-Shang, Babu Joseph and Hiro Mukai (1986). Comparison of two approaches to on-line parameter and state estimation of nonlinear systems. *Ind. Eng. Chem. Process Des. Dev.* **25**, 809–814.

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, CACHE, Austin, TX* (D. Rippin, J. Hale and J. Davis, Eds.).

Liebman, M. J., T. F. Edgar and L. S. Lasdon (1992). Efficient data reconciliation and estimation for dynamic processes using nonlinear programming techniques. *Computers and Chemical Engineering* **16**(10/11), 963–986.

Malakoff, David (1999). Bayes offers a 'new way' to make sense of numbers. *Science* **286**, 1460–1464.

Maybeck, Peter S. (1979). *Stochastic Models, estimation and control*. Academic Press, New York.

Rao, Christopher V. and James B. Rawlings (2000). *Nonlinear moving horizon state estimation*. Nonlinear Model Predictive Control. Birkhauser.

Rao, Christopher V. and James B. Rawlings (2002). Constrained process monitoring: Moving-horizon approach. *AIChE Journal* **48**(1), 97–109.

Robertson, Douglas G. and Jay H. Lee (1995). A least squares formulation for state estimation. *J. Proc. Cont.* **5**(4), 291–299.

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.

Tjoa, I. B. and L. T. Biegler (1991). Simultaneous strategies for data reconciliation and gross error detection of nonlinear systems. *Computers and Chemical Engineering* **15**(10), 679–690.