

Bayesian identification of non-linear state-space models: Part II- Error analysis

Aditya Tulsyan^{*}, Biao Huang^{*}, R. Bhushan Gopaluni^{**},
J. Fraser Forbes^{*}

^{*} *Department of Chemical and Materials Engineering, University of Alberta, Edmonton AB T6G-2G6, Canada (e-mail: {tulsyan; biao.huang; fraser.forbes}@ualberta.ca)*

^{**} *Department of Chemical and Biological Engineering, University of British Columbia, Vancouver BC V6T-1Z3, Canada, and Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge MA 02139, USA (e-mail: gopaluni@chbe.ubc.ca)*

Abstract: In the last two decades, several methods based on sequential Monte Carlo (SMC) and Markov chain Monte Carlo (MCMC) have been proposed for Bayesian identification of stochastic non-linear state-space models (SSMs). It is well known that the performance of these simulation based identification methods depends on the numerical approximations used in their design. We propose the use of posterior Cramér-Rao lower bound (PCRLB) as a mean square error (MSE) bound. Using PCRLB, a systematic procedure is developed to analyse the estimates delivered by Bayesian identification methods in terms of bias, MSE, and efficiency. The efficacy and utility of the proposed approach is illustrated through a numerical example.

1. INTRODUCTION

Bayesian identification has a long history, dating at least as far back as Peterka [1981]. Despite this, it is not commonly used in practice, except for the linear, Gaussian SSM case; wherein, Kalman filter based Bayesian estimate is routinely employed (Ninness and Henriksen [2010]). This is due to the computational complexities associated with the computation of the posterior densities, their marginals, and associated functions, such as posterior mean and variance (Juloski et al. [2005]). Recent developments in statistical methods, such as SMC and MCMC along with advances in computing technology have allowed researchers to use Bayesian methods in both on-line (Tulsyan et al. [2013a], Chen et al. [2005]) and off-line (Jang and Gopaluni [2011], Geweke and Tanizaki [2001]) identification of SSMs.

This paper is directed towards the class of Bayesian identification methods for parameter estimation in stochastic SSMs. The notation used in this paper is introduced next.

Notation: $\mathbb{N} := \{1, 2, \dots\}$; $\mathbb{R}_+ := [0, \infty)$; $\mathbb{R}^{s \times s}$ is the set of real-valued $s \times s$ matrices; $\mathcal{S}^s \subset \mathbb{R}^{s \times s}$ is the space of symmetric matrices; \mathcal{S}_+^s is the cone of symmetric positive semi-definite matrices in \mathcal{S}^s ; and \mathcal{S}_{++}^s is its interior. The partial order on \mathcal{S}^s induced by \mathcal{S}_+^s and \mathcal{S}_{++}^s are denoted by \succcurlyeq and \succ , respectively. For $A \in \mathbb{R}^{s \times s}$, $\text{Tr}[A]$ denotes its trace. For a vector $y \in \mathbb{R}^p$, $\text{diag}(y) \in \mathcal{S}^p$ is a diagonal matrix with $y \in \mathbb{R}^p$ as its entries. $|\cdot|$ is the absolute value. $\Delta_x^y \triangleq \nabla_x \nabla_y^T$ is Laplacian and $\nabla_x \triangleq [\frac{\partial}{\partial x}]$ is gradient.

2. BAYESIAN IDENTIFICATION

Let $\{X_t\}_{t \in \mathbb{N}}$ and $\{Y_t\}_{t \in \mathbb{N}}$ be $\mathcal{X} (\subseteq \mathbb{R}^n)$ and $\mathcal{Y} (\subseteq \mathbb{R}^m)$ valued stochastic processes defined on a measurable space (Ω, \mathcal{F}) . Let these stochastic processes depend on unknown

parameter vector $\theta \in \Theta$, where Θ is an open subset of \mathbb{R}^q . The discrete-time state $\{X_t\}_{t \in \mathbb{N}}$ is an unobserved process, with initial density $p_\theta(x)$ and transition density $p_\theta(x'|x)$:

$$X_0 \sim p_\theta(\cdot); \quad X_{t+1}|(X_t = x_t) \sim p_\theta(\cdot|x_t, u_t) \quad (t \in \mathbb{N}). \quad (1)$$

It is assumed that $\{Y_t\}_{t \in \mathbb{N}}$ is conditionally independent given $\{X_t\}_{t \in \mathbb{N}}$ and have a marginal density $p_\theta(y|x)$:

$$Y_t|(X_0, \dots, X_t = x_t, \dots, X_T) \sim p_\theta(\cdot|x_t) \quad (t \in \mathbb{N}). \quad (2)$$

All the densities are with respect to suitable dominating measures, such as Lebesgue measure, which are denoted generically as dx and dy . Although (1) and (2) represent a wide class of non-linear time-series models, the model form considered in this paper is given below

$$X_{t+1} = f_t(X_t, \theta_t, V_t); \quad (3a)$$

$$\theta_{t+1} = \theta_t; \quad (3b)$$

$$Y_t = g_t(X_t, \theta_t, W_t), \quad (3c)$$

where $\{\theta_t\}_{t \in \mathbb{N}} = \theta$ is a vector of unknown parameters, and $\{V_t\}_{t \in \mathbb{N}}$ and $\{W_t\}_{t \in \mathbb{N}}$ are the state and measurement noise.

Remark 1. To minimize use of notation, the input signal $\{u_t\}_{t \in \mathbb{N}}$ is not included in (3); however, all the results that appear in this paper hold with signal $\{u_t\}_{t \in \mathbb{N}}$ included. \square

For a generic sequence $\{r_t\}_{t \in \mathbb{N}}$, let $r_{i:j} \triangleq \{r_i, r_{i+1}, \dots, r_j\}$. In Bayesian identification, the problem of estimating the parameter vector $\theta \in \Theta \subseteq \mathbb{R}^q$ in (3), given a measurement sequence $\{Y_{1:t} = y_{1:t}\}_{t \in \mathbb{N}}$ is formulated as a joint state and parameter estimation problem. This is done by ascribing a prior density $\theta_0 \sim p(\theta_0)$, such that $\theta \in \text{supp } p(\theta_0)$, and computing the density $Z_t|(Y_{1:t} = y_{1:t}) \sim p(\cdot|y_{1:t})$, where $Z_t \triangleq \{X_t; \theta_t\}$ is a $\mathcal{Z} (\subseteq \mathbb{R}^{s+n+q})$ valued extended Markov process with $(Z_0 = z_0) \sim p_{\theta_0}(x_0)p(\theta_0)$, $Z_t|(Z_{t-1} = z_{t-1}) \sim p_{\theta_{t-1}}(\cdot|x_{t-1})\delta_{\theta_{t-1}}(\cdot)$. Note that a recursive method to compute $\{p(z_t|y_{1:t})\}_{t \in \mathbb{N}}$ is given by the optimal filtering

equation. Having computed $\{p(z_t|y_{1:t})\}_{t \in \mathbb{N}}$, inference on $\{\theta_t\}_{t \in \mathbb{N}}$ then relies on the marginal density $\{p(\theta_t|y_{1:t})\}_{t \in \mathbb{N}}$.

Although computing $\theta_t|Y_{1:t} = y_{1:t} \sim p(\cdot|y_{1:t})$ appears similar to computing $X_t|Y_{1:t} = y_{1:t} \sim p(\cdot|y_{1:t})$ (under known parameter case) in the state estimation problem, calculating $\{p(\theta_t|y_{1:t})\}_{t \in \mathbb{N}}$ for (3) has proved to be a non-trivial problem (Minvielle et al. [2010], Kantas et al. [2009]). No analytical solution to $\{p(\theta_t|y_{1:t})\}_{t \in \mathbb{N}}$ is available, even for linear and Gaussian SSM, or when \mathcal{X} is a finite set (Kantas et al. [2009]). There are several simulation and numerical methods (e.g., SMC, MCMC, Kalman based filters), which allow for recursive approximation of $\{p(\theta_t|y_{1:t})\}_{t \in \mathbb{N}}$. Although tractable, the quality of these identification methods depends on the underlying numerical and statistical approximations used in their design.

Despite the widespread interest in developing advanced simulation and numerical methods for Bayesian identification of (3), there have been no elaborate study on the quality of these methods. With this background, this paper proposes the use of PCRLB as an error bound. Using PCRLB, a systematic approach to assess the quality of a Bayesian identification method, in terms of bias, MSE, and efficiency is developed. Initial results reported by the authors in Tulsyan et al. [2013b] use PCRLB for assessment of state (but not parameter) estimation algorithms. The focus of this paper is to extend the results in Tulsyan et al. [2013b] to the Bayesian parameter estimation algorithms.

3. PCRLB AS AN ERROR BOUND

The conventional Cramér-Rao lower bound (CRLB) provides a theoretical lower bound on the MSE of any maximum-likelihood (ML) based unbiased parameter estimator. An analogous extension of the CRLB to the Bayesian estimators was derived by Trees [1968], and is commonly referred to as the PCRLB inequality. The PCRLB, derived recently by Tichavský et al. [1998] for (3), provides a lower bound on the MSE associated with the joint estimation of the states and parameters from $\{p(z_t|u_{1:t}, y_{1:t})\}_{t \in \mathbb{N}}$, and is given in the next lemma.

Lemma 2. Let $\{Y_{1:t} = y_{1:t}\}_{t \in \mathbb{N}}$ be an output sequence generated from (3), then the MSE associated with the estimation of $\{Z_t\}_{t \in \mathbb{N}}$ from $\{p(z_t|y_{1:t})\}_{t \in \mathbb{N}}$ is bounded by

$$P_{t|t}^z \triangleq \mathbb{E}_{p(z_t, y_{1:t})}[(Z_t - Z_{t|t})(Z_t - Z_{t|t})^T] \succcurlyeq [J_t^z]^{-1}, \quad (4)$$

where: $Z_{t|t} := \mathbb{R}^{tm} \rightarrow \mathbb{R}^s$ is a point estimate of $\{Z_t\}_{t \in \mathbb{N}}$;

$$P_{t|t}^z \triangleq \begin{bmatrix} P_{t|t}^x & P_{t|t}^{x\theta} \\ (P_{t|t}^{x\theta})^T & P_{t|t}^\theta \end{bmatrix} \in \mathcal{S}_{++}^s, J_t^z \triangleq \begin{bmatrix} J_t^x & J_t^{x\theta} \\ (J_t^{x\theta})^T & J_t^\theta \end{bmatrix} \in \mathcal{S}_{++}^s,$$

$[J_t^z]^{-1} \triangleq \begin{bmatrix} L_t^x & L_t^{x\theta} \\ (L_t^{x\theta})^T & L_t^\theta \end{bmatrix} \in \mathcal{S}_{++}^s$ are the MSE, posterior information matrix (PIM), and PCRLB, respectively.

Proof. See Tichavský et al. [1998] for proof. \square

A recursive approach to compute $J_t^z \in \mathcal{S}_{++}^s$ was derived by Tichavský et al. [1998], and is given next. But first, we give the assumptions on the model considered in (3).

Assumption 3. $\{V_t\}_{t \in \mathbb{N}}$ and $\{W_t\}_{t \in \mathbb{N}}$ are mutually independent sequences of independent random variables known a priori in their distribution classes (e.g., Gaussian) and parametrized by a known and finite number of moments.

Assumption 4. $f_t := \mathcal{X} \times \Theta \times \mathbb{R}^m \rightarrow \mathbb{R}^n$ and $g_t := \mathcal{X} \times \Theta \times \mathbb{R}^m \rightarrow \mathbb{R}^m$ are non-linear functions, such that in the open set \mathcal{X} and Θ , $\{f_t; g_t\}$ is $\mathcal{C}^k(\mathcal{X})$ and $\mathcal{C}^k(\Theta)$, and in \mathbb{R}^n and \mathbb{R}^m , f_t is $\mathcal{C}^{k-1}(\mathbb{R}^n)$, and g_t is $\mathcal{C}^{k-1}(\mathbb{R}^m)$, where $k \geq 2$.

Assumption 5. For any random sample $(x_{t+1}, x_t, \theta_t, v_t) \in \mathcal{X} \times \mathcal{X} \times \Theta \times \mathbb{R}^n$ and $(y_t, x_t, \theta_t, w_t) \in \mathcal{Y} \times \mathcal{X} \times \Theta \times \mathbb{R}^m$ satisfying (3), $\nabla_{v_t} f_t^T(x_t, \theta_t, v_t)$ and $\nabla_{w_t} g_t^T(x_t, \theta_t, w_t)$ have rank n and m , respectively, such that using implicit function theorem, $p_\theta(x_{t+1}|x_t) = p(V_t = \tilde{f}_t(x_t, \theta_t, x_{t+1}))$ and $p_\theta(y_t|x_t) = p(W_t = \tilde{g}_t(x_t, \theta_t, y_t))$ are defined.

Lemma 6. A recursive approach to compute $J_t^z \in \mathcal{S}_{++}^s$ for (3) under Assumptions 3 through 5 is given as follows

$$J_{t+1}^x = H_t^{33} - (H_t^{13})^T [J_t^x + H_t^{11}]^{-1} H_t^{13}; \quad (5a)$$

$$J_{t+1}^{x\theta} = (H_t^{23})^T - (H_t^{13})^T [J_t^x + H_t^{11}]^{-1} (J_t^{x\theta} + H_t^{12}); \quad (5b)$$

$$J_{t+1}^\theta = J_t^\theta + H_t^{22} - (J_t^{x\theta} + H_t^{12})^T [J_t^x + H_t^{11}]^{-1} \times (J_t^{x\theta} + H_t^{12}), \quad (5c)$$

where:

$$H_t^{11} = \mathbb{E}_{p(x_{0:t+1}, \theta_t, y_{1:t+1})}[-\Delta_{X_t}^{X_t} \log p_t]; \quad (6a)$$

$$H_t^{12} = \mathbb{E}_{p(x_{0:t+1}, \theta_t, y_{1:t+1})}[-\Delta_{X_t}^{\theta_t} \log p_t]; \quad (6b)$$

$$H_t^{13} = \mathbb{E}_{p(x_{0:t+1}, \theta_t, y_{1:t+1})}[-\Delta_{X_t}^{X_{t+1}} \log p_t]; \quad (6c)$$

$$H_t^{22} = \mathbb{E}_{p(x_{0:t+1}, \theta_t, y_{1:t+1})}[-\Delta_{\theta_t}^{\theta_t} \log p_t]; \quad (6d)$$

$$H_t^{23} = \mathbb{E}_{p(x_{0:t+1}, \theta_t, y_{1:t+1})}[-\Delta_{\theta_t}^{X_{t+1}} \log p_t]; \quad (6e)$$

$$H_t^{33} = \mathbb{E}_{p(x_{0:t+1}, \theta_t, y_{1:t+1})}[-\Delta_{X_{t+1}}^{X_{t+1}} \log p_t]; \quad (6f)$$

and: $p_t = p(X_{t+1}|Z_t)p(Y_{t+1}|\theta_t, X_{t+1})$; and the PIM at $t = 0$ can be computed using $J_0 = \mathbb{E}_{p(z_0)}[-\Delta_{Z_0}^{Z_0} \log p(Z_0)]$.

Proof. See Tichavský et al. [1998] for proof. \square

Since the focus here is on $\{\theta_t\}_{t \in \mathbb{N}}$ alone, a lower bound on the MSE associated with the estimation of $\{\theta_t\}_{t \in \mathbb{N}}$ is of interest to us. Using Lemmas 2 and 6, a bound on the MSE for parameter estimates can be derived, as given next.

Corollary 7. Let $P_{t|t}^z \in \mathcal{S}_{++}^s$ and $J_t^z \in \mathcal{S}_{++}^s$ be such that they satisfy (4), then the MSE associated with the estimation of $\{\theta_t\}_{t \in \mathbb{N}}$ from $\{p(\theta_t|y_{1:t})\}_{t \in \mathbb{N}}$, is bounded by

$$P_{t|t}^\theta = \mathbb{E}_{p(\theta_t, y_{1:t})}[(\theta_t - \theta_{t|t})(\theta_t - \theta_{t|t})^T] \succcurlyeq L_t^\theta, \quad (7)$$

where $\theta_{t|t} := \mathbb{R}^{tm} \rightarrow \mathbb{R}^q$ is the parameter estimate delivered by a Bayesian identification algorithm, and $L_t^\theta \in \mathcal{S}_{++}^q$ is the lower right matrix of $[J_t^z]^{-1} \in \mathcal{S}_{++}^s$ in Lemma 2.

Proof. The proof is based on the fact that the inequality in Lemma 2 guarantees that $P_{t|t}^z - [J_t^z]^{-1} \in \mathcal{S}_{++}^s$. \square

A recursive approach to compute $L_t^\theta \in \mathcal{S}_{++}^q$ is given next.

Theorem 8. Let $J_t^z \in \mathcal{S}_{++}^s$ be the PIM for $\{Z_t\}_{t \in \mathbb{N}}$, and $L_t^\theta \in \mathcal{S}_{++}^q$ be the lower bound on the MSE associated with the estimation of $\{\theta_t\}_{t \in \mathbb{N}}$ in (3), then given $J_t^z \in \mathcal{S}_{++}^s$, $L_t^\theta \in \mathcal{S}_{++}^q$ at $t \in \mathbb{N}$ can be recursively computed as follows

$$L_t^\theta = [J_t^\theta - (J_t^{x\theta})^T (J_t^x)^{-1} J_t^{x\theta}]^{-1}, \quad (8)$$

where J_t^θ , $J_t^{x\theta}$ and J_t^x are the PIMs given in Lemma 2.

Proof. The proof is based on the matrix inversion lemma (see R.B. Bapat and T.E.S. Raghavan [1997]). \square

Remark 9. Theorem 8 shows that for (3), L_t^θ is not only a function of the PIM for $\{\theta_t\}_{t \in \mathbb{N}}$, i.e., J_t^θ , but it also depends on the PIMs for $\{X_t\}_{t \in \mathbb{N}}$, i.e., $J_t^{x\theta}$ and J_t^x . \square

Remark 10. Integral in (6) with respect to $p(x_{0:t}, \theta_{t-1}, y_{1:t})$ makes L_t^θ in (8) independent of any random sample from \mathcal{X}^{t+1} , Θ , and \mathcal{Y}^t . L_t^θ in fact only depends on: the process dynamics in (3); noise characteristics of $V_t \sim p(v_t)$ and $W_t \sim p(w_t)$; and the choice of $Z_0 \sim p(z_0)$. This makes L_t^θ a system property, independent of any Bayesian identification method or any specific realization from \mathcal{X} , Θ or \mathcal{Y} . This motivates the use of PCRLB as a benchmark for error analysis of Bayesian identification algorithms. \square

Finally, using the inequality in (7), the MSE associated with the parameter estimates obtained with any Bayesian identification method can be compared against the theoretical lower bound. Our approach to systematically compare and analyse the MSE and PCRLB is discussed next.

4. PCRLB INEQUALITY BASED ERROR ANALYSIS

A common approach to compute $\theta_{t|t} \in \mathbb{R}^q$, is to minimize $\text{Tr}[P_{t|t}^\theta] \in \mathbb{R}_+$. This ensures that $\text{Tr}[P_{t|t}^\theta - L_t^\theta] \geq 0$ is minimized. The optimal estimate that minimizes $\text{Tr}[P_{t|t}^\theta] \in \mathbb{R}_+$ is referred to as the minimum MSE (MMSE) estimate, and is the conditional mean of $\theta_t|(Y_{1:t} = y_{1:t}) \sim p(\cdot|y_{1:t})$, i.e., $\theta_{t|t} = \theta_{t|t}^* \triangleq \mathbb{E}_{p(\theta_t|y_{1:t})}[\theta_t]$ (see Trees [1968] for derivation).

Remark 11. Bayesian identification methods only approximate the true density $\{p(\theta_t|y_{1:t})\}_{t \in \mathbb{N}}$, thus in practice, the estimate delivered by identification methods may not be an MMSE estimate, i.e., $\theta_{t|t} \triangleq \mathbb{E}_{\tilde{p}(\theta_t|y_{1:t})}[\theta_t] \neq \theta_{t|t}^*$ almost surely, where $\theta_{t|t}$ is the mean of $\theta_t|(Y_{1:t} = y_{1:t}) \sim \tilde{p}(\cdot|y_{1:t})$ and $\{\tilde{p}(\theta_t|y_{1:t})\}_{t \in \mathbb{N}}$ is the approximate posterior. \square

The second-order error associated with $\{\theta_{t|t}\}_{t \in \mathbb{N}}$ is completely characterized by its MSE. A thorough assessment of any identification algorithm or that of its estimates requires clear understanding of the MSE. The next theorem shows decomposition of the MSE into its sources of errors.

Theorem 12. Let $\theta_{t|t}^* \in \mathbb{R}^q$ and $V_{t|t}^* \in \mathcal{S}_{++}^q$ be the mean and covariance of $\theta_t|(Y_{1:t} = y_{1:t}) \sim p(\cdot|y_{1:t})$ and $\theta_{t|t} \in \mathbb{R}^q$ be the mean of $\theta_t|(Y_{1:t} = y_{1:t}) \sim \tilde{p}(\cdot|y_{1:t})$ computed by a Bayesian identification method, then for $\theta_{t|t} \neq \theta_{t|t}^*$ almost surely, $P_{t|t}^\theta$ at $t \in \mathbb{N}$ can be decomposed and written as

$$P_{t|t}^\theta = \mathbb{E}_{p(y_{1:t})}[V_{t|t}^*] + \mathbb{E}_{p(y_{1:t})}[B_{t|t}^*[B_{t|t}^*]^T], \quad (9)$$

where $B_{t|t}^* \triangleq [\theta_{t|t}^* - \theta_{t|t}] \in \mathbb{R}^q$ is the conditional bias in estimating the true conditional mean $\theta_{t|t}^* \in \mathbb{R}^q$ at $t \in \mathbb{N}$.

Proof. The proof is adapted from (Tulsyan et al. [2013b]). From the definition of expectation, MSE in (7) can be written as $P_{t|t}^\theta = \mathbb{E}_{p(y_{1:t})} \mathbb{E}_{p(\theta_t|y_{1:t})}[(\theta_t - \theta_{t|t})(\theta_t - \theta_{t|t})^T]$, where we have used $p(\theta_t, y_{1:t}) = p(y_{1:t})p(\theta_t|y_{1:t})$. Adding and subtracting $\theta_{t|t}^*$ in $P_{t|t}^\theta$, followed by several algebraic manipulations yield $P_{t|t}^\theta = \mathbb{E}_{p(y_{1:t})} \mathbb{E}_{p(\theta_t|y_{1:t})}[F_{t|t}^* + G_{t|t}^* + [G_{t|t}^*]^T + B_{t|t}^*[B_{t|t}^*]^T]$, where $F_{t|t}^* = [\theta_t - \theta_{t|t}^*][\theta_t - \theta_{t|t}^*]^T$; $G_{t|t}^* = [\theta_t - \theta_{t|t}^*][\theta_t - \theta_{t|t}]^T$. Now $\mathbb{E}_{p(\theta_t|y_{1:t})}[F_{t|t}^*] = V_{t|t}^*$; $\mathbb{E}_{p(\theta_t|y_{1:t})}[G_{t|t}^*] = 0$, since $\mathbb{E}_{p(\theta_t|y_{1:t})}[\theta_t - \theta_{t|t}^*] = 0$; and $\mathbb{E}_{p(\theta_t|y_{1:t})}[B_{t|t}^*[B_{t|t}^*]^T] = [B_{t|t}^*][B_{t|t}^*]^T$, since $[B_{t|t}^*][B_{t|t}^*]^T$ is independent of $\theta_t|(Y_{1:t} = y_{1:t})$. Substituting the results into $P_{t|t}^\theta$ yields (9), which completes the proof. \square

Note that Theorem 12 is the Bayesian equivalent of the classical MSE decomposition results available for the likelihood based estimators. Using Theorem 12, bias in the Bayesian parameter estimates $\{\theta_{t|t}\}_{t \in \mathbb{N}}$ is defined next.

Definition 13. $\{\theta_{t|t}\}_{t \in \mathbb{N}} \in \mathbb{R}^q$ is unconditionally unbiased if $\mathbb{E}_{p(y_{1:t})}[B_{t|t}^*] = 0$, and conditionally unbiased if $B_{t|t}^* = 0$ almost surely. The estimate which is both conditionally and unconditionally unbiased is an unbiased estimate. \square

Bias in $\theta_{t|t} \in \mathbb{R}^q$ can be similarly defined as Definition 13. The condition under which an identification method delivers unbiased parameter estimate is discussed next.

Theorem 14. Let $\theta_{t|t} \in \mathbb{R}^q$ be the estimate of $\theta_{t|t}^* \in \mathbb{R}^q$, as computed by an identification method, where $\theta_{t|t}^* \in \mathbb{R}^q$ is the mean of $\theta_t|(Y_{1:t} = y_{1:t}) \sim p(\cdot|y_{1:t})$, and let $B_{t|t}^* \in \mathbb{R}^q$ be the corresponding conditional bias, then $B_{t|t}^* = 0$ almost surely is only a sufficient condition for $\mathbb{E}_{p(y_{1:t})}[B_{t|t}^*] = 0$, but sufficient and necessary for $\mathbb{E}_{p(y_{1:t})}[B_{t|t}^*[B_{t|t}^*]^T] = 0$.

Proof. See Billingsley [1995] for proof. \square

Remark 15. Theorem 14 shows that if the parameter estimate $\theta_{t|t} \in \mathbb{R}^q$ is unconditionally unbiased, it does not imply it is unbiased as well, but if it is conditionally unbiased, it implies $\theta_{t|t} \in \mathbb{R}^q$ is unbiased as well. \square

The MSE for an unbiased estimate $\theta_{t|t} \in \mathbb{R}^q$ is given next.

Corollary 16. Let $\theta_{t|t} \in \mathbb{R}^q$ be the estimate of the mean of $\theta_t|(Y_{1:t} = y_{1:t}) \sim p(\cdot|y_{1:t})$ computed by a Bayesian identification method, such that $B_{t|t}^* = 0$ almost surely, then the MSE associated with $\theta_{t|t} \in \mathbb{R}^q$ is $P_{t|t}^\theta = \mathbb{E}_{p(y_{1:t})}[V_{t|t}^*]$. \square

Definition 17. An identification method delivering an estimate $\theta_{t|t} \in \mathbb{R}^q$ is efficient at $t \in \mathbb{N}$ if $\text{Tr}[P_{t|t}^\theta - L_t^\theta] = 0$. \square

Theorem 18. Let $\theta_{t|t} \in \mathbb{R}^q$ be the estimate of $\theta_{t|t}^* \in \mathbb{R}^q$, as computed by an identification method, and let $B_{t|t}^* \in \mathbb{R}^s$ be the conditional bias in estimating $\theta_{t|t}^* \in \mathbb{R}^q$, then $B_{t|t}^* = 0$ almost surely is both necessary and sufficient condition for the identification method to be efficient.

Proof. For $\theta_{t|t} \in \mathbb{R}^q$ satisfying $B_{t|t}^* = 0$ almost surely, the MSE is given by $P_{t|t}^\theta = \mathbb{E}_{p(Y_{1:t})}[V_{t|t}^*]$ (see Corollary 16). Since $P_{t|t}^\theta$ only depends on $V_{t|t}^*$, which is the covariance of $\theta_t|(Y_{1:t} = y_{1:t}) \sim p(\cdot|Y_{1:t})$, $P_{t|t}^\theta$ cannot be reduced any further i.e., $P_{t|t}^\theta = L_t^\theta$. Thus from Definition 17 the identification method delivering $\theta_{t|t} \in \mathbb{R}^q$ is efficient at $t \in \mathbb{N}$. \square

Finally, the procedure to systematically assess the quality of the parameter estimates obtained with any Bayesian identification method is summarized in the next theorem.

Theorem 19. Let $L_t^\theta \in \mathcal{S}_{++}^q$ be the PCRLB on (3), and let $\theta_{t|t}^* \in \mathbb{R}^q$ and $V_{t|t}^* \in \mathcal{S}_{++}^q$ be the mean and covariance of $\theta_t|(Y_{1:t} = y_{1:t}) \sim p(\cdot|y_{1:t})$. Now if $\theta_{t|t} \in \mathbb{R}^q$ is an estimate of $\theta_{t|t}^* \in \mathbb{R}^q$, as computed by an identification method, such that $B_{t|t}^* \in \mathbb{R}^q$ is the conditional bias in estimating $\theta_{t|t}^* \in \mathbb{R}^q$, then for $P_{t|t}^\theta \in \mathcal{S}_{++}^q$ as the associated MSE, the quality of the estimate $\theta_{t|t} \in \mathbb{R}^q$ can be assessed as follows: (a) If $B_{t|t}^* = 0$ almost surely, then (7) is given by

$$P_{t|t}^\theta = \mathbb{E}_{p(y_{1:t})}[V_{t|t}^*] = L_t^\theta, \quad (10)$$

which implies the identification method is efficient, and the corresponding estimate $\theta_{t|t} \in \mathbb{R}^q$ is unbiased and MMSE. (b) If $B_{t|t}^* \neq 0$ almost surely, then (7) is given by

$$P_{t|t}^\theta = \mathbb{E}_{p(y_{1:t})}[V_{t|t}^*] + \mathbb{E}_{p(y_{1:t})}[B_{t|t}^* [B_{t|t}^*]'] \succ L_t^\theta, \quad (11)$$

which implies the identification method is not efficient, and the estimate $\theta_{t|t} \in \mathbb{R}^q$ is biased (only conditionally biased if $\mathbb{E}_{p(y_{1:t})}[B_{t|t}^*] = 0$) and not an MMSE estimate.

Proof. The proof is based on the collective developments of Section 4, and is omitted here for the sake of brevity. \square

The PCRLB inequality based error analysis tool developed in this section allows for assessment of parameter estimates obtained with Bayesian identification methods; however, obtaining a closed form solution to (7) is non-trivial for (3). Use of numerical methods is discussed next.

5. NUMERICAL METHODS

It is well known that computing the MSE and PCRLB in (7) in closed form is non-trivial for the model considered in (3) (see Tichavský et al. [1998], Bergman [2001]). This is because of the complex, high-dimensional integrals in the MSE with respect to $p(\theta_t, y_{1:t})$ (see (7)) and in the PCRLB with respect to $p(x_{0:t}, \theta_{t-1}, y_{1:t})$ (see (6a) through (6f)), which do not admit any analytical solution.

To address this issue, we use Monte Carlo (MC) sampling to numerically compute the MSE and PCRLB in (7). For the sake of brevity, the procedure for MC approximation of the PCRLB is not provided here, but can be found in Tulsyan et al. [2013c]; however, for completeness, we provide an example for computation of MC based MSE.

Example 20. Simulating samples $\{(\theta_t = \theta_t^j, Y_{1:t} = y_{1:t}^j)\}_{j=1}^M \sim p(\theta_t, y_{1:t})$, M times using (3), starting at M i.i.d. initial draws from $\{\theta_0\}_{i=1}^M \sim p(\theta_0)$ and computing the estimates $\{\theta_{t|t}^j\}_{j=1}^M$, the MSE $P_{t|t}^\theta$ at $t \in \mathbb{N}$ can be approximated as

$$\tilde{P}_{t|t}^\theta = \frac{1}{M} \sum_{j=1}^M (\theta_t^j - \theta_{t|t}^j)(\theta_t^j - \theta_{t|t}^j)^T, \quad (12)$$

where $\tilde{P}_{t|t}^\theta \in \mathcal{S}_{++}^q$ is an M -sample MC estimate of $P_{t|t}^\theta$. \square

Since (12) is based on perfect sampling, using strong law of large numbers $\tilde{P}_{t|t}^\theta \xrightarrow{a.s.} P_{t|t}^\theta$ as $M \rightarrow +\infty$, where $\xrightarrow{a.s.}$ denotes almost sure convergence (see P. Del Moral [2004]). Note that \tilde{L}_t^θ , which is an M -sample MC estimate of L_t^θ can also be similarly approximated using MC sampling. Details are omitted here, but can be found in Tulsyan et al. [2013c]. Despite the convergence proof, there are practical issues with the use of numerical methods, as given next.

Remark 21. With $M < +\infty$, the MC estimate of the MSE and PCRLB may not necessarily satisfy the positive semi definite condition $\tilde{P}_{t|t}^\theta - \tilde{L}_t^\theta \succcurlyeq 0$ for all $t \in \mathbb{N}$. \square

Remark 22. Since $M < +\infty$, the conditions in Theorem 19 are relaxed to $|B_{t|t}^*| \leq \epsilon$ and $|\mathbb{E}_{p(Y_{1:t})}[B_{t|t}^*]| \leq \alpha$, and $\epsilon \in \mathbb{R}_+^q$ and $\alpha \in \mathbb{R}_+^q$ are pre-defined tolerance levels set based on M and the required degree of accuracy. \square

Remark 23. An identification method satisfying $|B_{t|t}^*| \leq \epsilon$ is ϵ -efficient at $t \in \mathbb{N}$ and the corresponding estimate is

ϵ -unbiased and ϵ -MMSE (see Theorem 19(a)). Similarly, if the estimate only satisfies $|\mathbb{E}_{p(y_{1:t})}[B_{t|t}^*]| \leq \alpha$, then it is α -unconditionally unbiased (see Theorem 19(b)). \square

6. FINAL ALGORITHM

A systematic approach to assess the quality of a Bayesian identification method, proposed in Sections 3 through 5 is formally outlined in Algorithm 1.

Algorithm 1 Analysis of Bayesian identification methods

Module 1: Computing the lower bound

Input: Given (3), define $Z_t = \{X_t, \theta_t\}$ and assume a prior density on $\{Z_t\}_{t \in \mathbb{N}}$, such that $(Z_0 = z_0) \sim p(z_0)$

Output: Lower bound on the system in (3)

- 1: Generate M i.i.d. samples from the assumed prior density $Z_0 \sim p(\cdot)$, such that $\{(Z_0 = z_0^i)\}_{i=1}^M \sim p(z_0^i)$
- 2: **for** $t = 1$ to T **do**
- 3: Generate M random samples from the states $\{X_t = x_t^i | (Z_{t-1} = z_{t-1}^i)\}_{i=1}^M \sim p(x_t^i | z_{t-1}^i)$ using (3a)
- 4: Generate M random samples from the parameters $\{\theta_t = \theta_t^i | (Z_{t-1} = z_{t-1}^i)\}_{i=1}^M \sim p(\theta_t^i)$ using (3b). Note that in this step $\theta_t^i = \theta_0^i$ for all $1 \leq i \leq M$ (see (3b))
- 5: Generate M random samples from the measurements $\{Y_t = y_t^i | (Z_t = z_t^i)\}_{i=1}^M \sim p(y_t^i | z_t^i)$ using (3c)
- 6: Compute an M -sample MC estimate of \tilde{J}_t^z
- 7: Compute an M -sample MC estimate of \tilde{L}_t^θ
- 8: **end for**

Module 2: Computing the estimates

Input: Measurement sequences from Module 1, denoted as $\{(Y_{1:T} = y_{1:T}^i)\}_{i=1}^M$ and a Bayesian identification method, which can compute $\{p(\theta_t | y_{1:t})\}_{t \in \mathbb{N}}$ (e.g., SMC, MCMC, EKF, and UKF)

Output: Parameter estimates

- 9: **for** $i = 1$ to M **do**
- 10: **for** $t = 1$ to T **do**
- 11: Compute $p(\theta_t | y_{1:t}^i)$ using an identification method and denote density approximation by $\tilde{p}(\theta_t | y_{1:t}^i)$
- 12: Using $\tilde{p}(\theta_t | y_{1:t}^i)$, compute parameter point estimate as $\theta_{t|t}^i = \mathbb{E}_{\tilde{p}(\theta_t | y_{1:t}^i)}[\theta_t]$
- 13: **end for**
- 14: **end for**

Module 3: Analysis of Bayesian identification method

Input: Parameter sequences from Module 1, denoted by $\{(\theta_{1:T} = \theta_{1:T}^i)\}_{i=1}^M$ and their estimates from Module 2, denoted as $\{(\theta_{t|t} = \theta_{t|t}^i)\}_{i=1, t=1}^{M, T}$. Matrices $\tilde{L}_t^\theta \in \mathcal{S}_{++}^q$

and $\tilde{P}_{t|t}^\theta \in \mathcal{S}_{++}^q$ and tolerance level $\epsilon \in \mathbb{R}_+^q$ and $\alpha \in \mathbb{R}_+^q$

Output: Error analysis of identification method

- 15: **for** $t = 1$ to T **do**
 - 16: Compute an M -sample MC estimate of $\tilde{P}_{t|t}^\theta$
 - 17: Compare $\tilde{P}_{t|t}^\theta$ against \tilde{L}_t^θ
 - 18: Compute $\{B_{t|t}^{*,i}\}_{i=1}^M$ and compare against $\epsilon \in \mathbb{R}_+^q$
 - 19: Compute an M -sample MC estimate of $\mathbb{E}_{p(y_{1:t})}[B_{t|t}^*]$ and compare against $\alpha \in \mathbb{R}_+^q$
 - 20: Use Theorem 19 for error analysis
 - 21: **end for**
-

7. SIMULATION EXAMPLE

In this section we use a simulated system to assess the quality of a Bayesian identification method using the

procedure outlined in Algorithm 1. A brief introduction to the identification method considered here, is given next.

7.1 Bayesian identification: Artificial dynamics approach

Artificial dynamics approach (ADA) is a popular Bayesian identification method to compute $\{p(\theta_t|y_{1:t})\}_{t \in \mathbb{N}}$. In ADA, artificial dynamics is introduced to the otherwise static parameters, such that $\{\theta_t\}_{t \in \mathbb{N}}$ in (3b) evolves according to

$$\theta_{t+1}|\theta_t \sim \mathcal{N}(\cdot|\theta_t, Q_t^\theta), \quad (13)$$

where $\theta_{t+1}|\theta_t \sim \mathcal{N}(\cdot|\theta_t, Q_t^\theta)$ is a sequence of independent Gaussian random variable, realized independent of $\{V_t\}_{t \in \mathbb{N}}$ and $\{W_t\}_{t \in \mathbb{N}}$. By appending (3a) and (3c) with (13), methods such as SMC, EKF, UKF can be used to recursively compute $\{p(\theta_t|y_{1:t})\}_{t \in \mathbb{N}}$. A detailed review on ADA can be found in Tulsyan et al. [2013a] and Kantas et al. [2009].

Even though ADA is the most widely used approach amongst the class of Bayesian identification methods, there are several standing limitations of this approach as summarized in Kantas et al. [2009] (a) the dynamics of $\{\theta_t\}_{t \in \mathbb{N}}$ in (13) is related to the artificial noise covariance Q_t^θ , which is often difficult to tune; and (b) adding dynamics to $\{\theta_t\}_{t \in \mathbb{N}}$ modifies the original problem, which means, it is hard to quantify the bias introduced in the estimates.

For the former problem, the authors in see Tulsyan et al. [2013a] proposed an optimal rule to automatically tune Q_t^θ for all $t \in \mathbb{N}$; however, for the later problem, we will see how the tools developed in this paper can be used to assess the quality of ADA based Bayesian identification methods.

7.2 Simulation setup

Consider the following univariate, non stationary, non-linear stochastic SSM (Tulsyan et al. [2013c])

$$X_{t+1} = aX_t + \frac{X_t}{b + X_t^2} + u_t + V_t, \quad V_t \sim \mathcal{N}(0, Q_t), \quad (14a)$$

$$Y_t = cX_t + dX_t^2 + W_t, \quad W_t \sim \mathcal{N}(0, R_t), \quad (14b)$$

where $\theta \triangleq [a \ b \ c \ d]$ is a vector of unknown static model parameters. The noise covariances are constant, and selected as $Q_t = 10^{-3}$ and $R_t = 10^{-3}$ for all $t \in [1, T]$, where $T = 300$. $\{u_t\}_{t \in [1, T]}$ is a sequence of optimal input (see Tulsyan et al. [2013c]). For Bayesian identification of θ , we define $\{\theta_t = \theta_{t-1}\}_{t \in [1, T]} = \theta$ as a stochastic process, such that $Z_t = \{X_t, \theta_t\}$ is a \mathcal{Z} valued extended Markov process with $Z_0 \sim \mathcal{N}(z_m, z_c)$, where $z_m = [1 \ 0.7 \ 0.6 \ 0.5 \ 0.4]$, $z_c = \text{diag}(0.01, 0.01, 0.01, 0.01, 0.01)$. Starting at $t = 0$, we are interested in assessing the ADA based SMC identification method proposed in Tulsyan et al. [2013a].

7.3 Results

Using $M = 1000$ MC simulations, we compute the PCRLB for (14) using Module 1 of Algorithm 1. Figure 1 gives the diagonal entries of $\{\tilde{L}_t^\theta\}_{t \in [1, T]}$. Note that amongst the four PCRLBs, the PCRLB for b is the highest for all $t \in [1, T]$. This suggest estimation difficulties with parameter b . This result is not surprising, since (14) is non-linear in parameter b ; however, the overall decaying trend of PCRLBs in Figure 1 suggests that starting with $\theta_0 \sim p(\theta_0)$, theoretically, it is possible for a Bayesian identification method to

reduce the MSE associated with the parameter estimates.

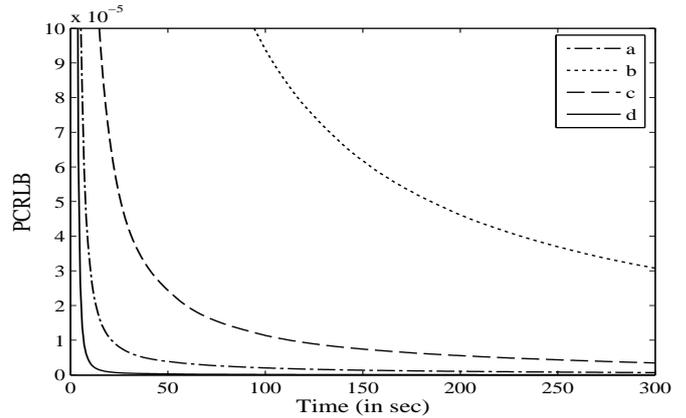


Fig. 1. PCRLB for parameters as a function of time. Note that the vertical axis has been appropriately scaled for clarity.

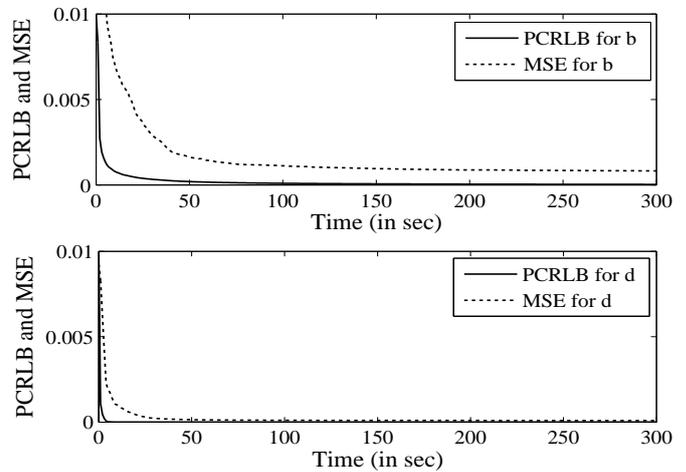


Fig. 2. Comparing PCRLB against the MSE computed using the ADA based SMC identification method. Graphs are shown only for parameters b and d . Results for a and c are similar.

Figure 2 compares the PCRLB against the MSE for b and d , computed using Module 2 of Algorithm 1. Despite the MC approximations involved (see Remark 21), the MSE is greater than the PCRLB at all sampling time instants (see Figure 2). It is instructive to highlight that at $T = 300$, the MSE associated with the estimation of d is about 89% less than that for b , which validates the claim made earlier about estimation difficulties with parameter b . It is also important to point that for the ADA based SMC method, $\text{Tr}[\tilde{P}_{t|t}^\theta - \tilde{L}_t^\theta] \neq 0$ for all $t \in [1, T]$. Since the ADA based SMC method fails to satisfy the condition in Definition 17, it is not efficient, and therefore requires error analysis.

Figures 3 and 4 give the conditional and unconditional bias with ADA based SMC method. The results are obtained using Module 3 of Algorithm 1. Based on an assumed tolerance level $\epsilon = [0.01; 0.01; 0.01; 0.01]$ and $\alpha = [0.001; 0.001; 0.001; 0.001]$, in the interval $t = [1, 50]$, less than 70% of the simulations are within the specified ϵ limit (see Figure 3). Thus from Theorem 19(b), for $t = [1, 50]$, the ADA based SMC method is not even ϵ -efficient, and fails to yield ϵ -unbiased (except for d , which is

REFERENCES

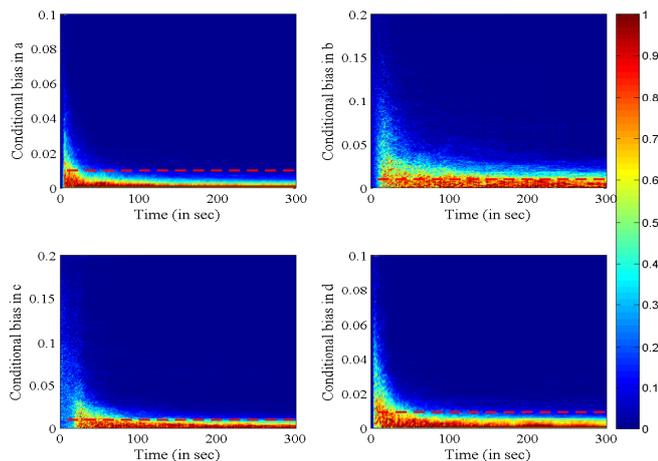


Fig. 3. Conditional bias in parameter estimates with ADA based SMC method. The broken red line is the ϵ value.

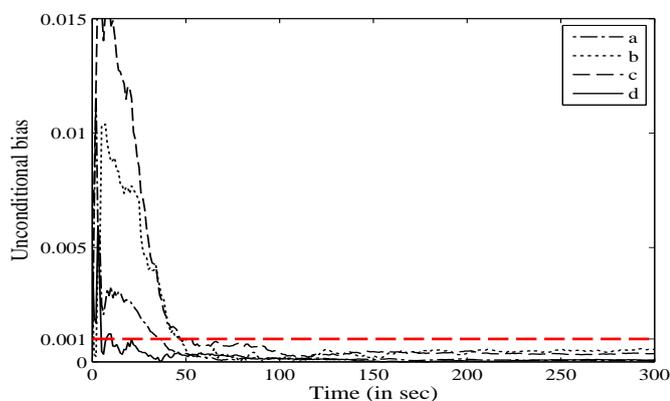


Fig. 4. Unconditional bias in parameter estimates with ADA based SMC method. The broken red line is the α value.

α -unconditionally unbiased, see Figure 4) or ϵ -MMSE estimates. Another interesting interval is $t = [100, T]$; wherein, more than 70% of the simulations are within the specified ϵ limit (except for parameter b , where only 60% of simulations are within ϵ , see Figure 3). Thus from Theorem 19(a), the ADA based SMC method is ϵ -efficient for all the parameters, except for b , and the resulting estimates are ϵ -unbiased and ϵ -MMSE; whereas, for b , the estimates are not MMSE, but are α -unconditionally unbiased.

In summary, the results suggest that for model given in (14), the ADA based SMC method at $t = T$ yields ϵ -unbiased, ϵ -MMSE estimates for all the parameters, except for parameter b , which is only α -unconditionally unbiased.

8. CONCLUSIONS

A PCRLB based approach is proposed for error analysis in Bayesian identification methods of non-linear SSMs. Using the proposed tool it was illustrated how the quality of the parameter estimates obtained using artificial dynamics approach, which is a popular Bayesian identification method can be assessed in terms of bias, MSE and efficiency.

ACKNOWLEDGEMENT

This work was supported by the Natural Sciences and Engineering Research Council (NSERC), Canada.

N. Bergman. *Sequential Monte Carlo Methods in Practice*, chapter Posterior Cramér-Rao bounds for sequential estimation. Springer-Verlag, New York, 2001.

P. Billingsley. *Probability and Measure*. John Wiley & Sons, New York, 1995.

T. Chen, J. Morris, and E. Martin. Particle filters for state and parameter estimation in batch processes. *Journal of Process Control*, 15(6):665–673, 2005.

J. Geweke and H. Tanizaki. Bayesian estimation of state-space models using the Metropolis-Hastings algorithm within Gibbs sampling. *Computational Statistics & Data Analysis*, 37(2):151–170, 2001.

S.S. Jang and R.B. Gopaluni. Parameter estimation in nonlinear chemical and biological processes with unmeasured variables from small data sets. *Chemical Engineering Science*, 66(12):2774–2787, 2011.

A.L. Juloski, S. Weiland, and W.P.M.H. Heemels. A Bayesian approach to identification of hybrid systems. *IEEE Transactions on Automatic Control*, 50(10):1520–1533, 2005.

N. Kantas, A. Doucet, S.S. Singh, and J. Maciejowski. An overview of sequential Monte-Carlo methods for parameter estimation in general state-space models. In *Proceedings of the 15th IFAC Symposium on System Identification*, pages 774–785, Saint-Malo, France, 2009.

P. Minvielle, A. Doucet, A. Marrs, and S. Maskell. A Bayesian approach to joint tracking and identification of geometric shapes in video sequences. *Image and Vision Computing*, 28(1):111–123, 2010.

B. Ninness and S. Henriksen. Bayesian system identification via Markov chain Monte Carlo techniques. *Automatica*, 46(1):40–51, 2010.

P. Del Moral. *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Springer-Verlag, New York, 2004.

V. Peterka. Bayesian system identification. *Automatica*, 17(1):41–53, 1981.

R.B. Bapat and T.E.S. Raghavan. *Nonnegative Matrices and Applications*. Cambridge University Press, Cambridge, 1997.

P. Tichavský, C. Muravchik, and A. Nehorai. Posterior Cramér-Rao bounds for discrete-time nonlinear filtering. *IEEE Transactions on Signal Processing*, 46(5):1386–1396, 1998.

H.L.V. Trees. *Detection, Estimation and Modulation Theory-Part I*, chapter Classical detection and estimation theory. Wiley, New York, 1968.

A. Tulsyan, B. Huang, R.B. Gopaluni, and J.F. Forbes. On simultaneous on-line state and parameter estimation in non-linear state-space models. *Journal of Process Control*, 23(4):516–526, 2013a.

A. Tulsyan, B. Huang, R.B. Gopaluni, and J.F. Forbes. Performance assessment, diagnosis, and optimal selection of non-linear state filters. *Journal of Process Control, In Review*, 2013b.

A. Tulsyan, S.R. Khare, B. Huang, R.B. Gopaluni, and J.F. Forbes. Bayesian identification of non-linear state space models: Part I-Input design. In *Proceedings of the 10th International Symposium on Dynamics and Control of Process Systems*, Mumbai, India, 2013c.