Automatica 111 (2020) 108609

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Boolean Kalman filter and smoother under model uncertainty ☆

Mahdi Imani [a,*], Edward R. Dougherty [b,c], Ulisses Braga-Neto [b,c]

[a] Department of Electrical and Computer Engineering, George Washington University, Washington, DC, USA
[b] Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, USA
[c] Center for bioinformatics and Genomics Systems Engineering, TEES, College Station, TX, USA

article info

Article history:
Received 23 August 2018
Received in revised form 11 May 2019
Accepted 6 September 2019
Available online 28 September 2019

Keywords:
Partially-observed Boolean dynamical systems
Optimal Bayesian estimation
Minimum mean-square error
Boolean Kalman filter and smoother
Gene regulatory networks

abstract

Partially-observed Boolean dynamical systems (POBDS) are a general class of nonlinear state-space models that provide a rich framework for modeling many complex dynamical systems. The model consists of a hidden Boolean state process, observed through an arbitrary noisy mapping to a measurement space. The optimal minimum mean-square error (MMSE) POBDS state estimators are the Boolean Kalman Filter and Smoother. However, in many practical problems, the system parameters are not fully known and must be estimated. In this paper, for POBDS under model uncertainty, we derive an optimal Bayesian estimator for state and parameter estimation. The exact algorithms are derived for the case of discrete and finite parameter space, and for general parameter spaces, an approximate Markov-Chain Monte-Carlo (MCMC) implementation is introduced. We demonstrate the performance of the proposed methodology by means of numerical experiments with POBDS models of gene regulatory networks observed through noisy measurements.

1. Introduction

Partially-Observed Boolean dynamical systems (POBDS) are a general class of nonlinear state-space models consisting of a hidden Boolean state process observed through an arbitrary noisy mapping to a measurement space. This signal model has many applications in fields such as genomics (Kauffman, 1969), robotics (Imani & Braga-Neto, 2017a; Roli, Manfroni, Pinciroli, & Birattari, 2011), and digital communication systems (Messerschmitt, 1990). The optimal minimum mean square error (MMSE) state estimators for POBDS are the Boolean Kalman filter (BKF) (Imani, 2019; Imani & Braga-Neto, 2017a; McClenny, Imani, & Braga-Neto, 2017c) and Boolean Kalman smoother (BKS) (Imani & Braga-Neto, 2015b, 2017a), respectively. In Imani and Braga-Neto (2018c) and McClenny, Imani, and Braga-Neto (2017a), optimal state estimators for POBDS with correlated measurement noise are introduced.

Due to the structure of the multivariate Boolean lattice, the BKF and BKS have the desirable property of yielding both the optimal maximum a posteriori (MAP) and MMSE solutions for each state vector component (Imani & Braga-Neto, 2017a, 2018e; McClenny, Imani, & Braga-Neto, 2017b), which is not the case for the general multivariate MAP estimator, in general. In addition, exact algorithms are available for the computation of the BKF and BKS (Braga-Neto, 2011; Imani & Braga-Neto, 2017a), which is not the case for the optimal MMSE solution in general non-linear state-space models, in which case approximate solutions employing sequential Monte-Carlo techniques (also known as particle filters) (Doucet, De Freitas and Gordon, 2001; Doucet, Godsill, & Andrieu, 2000; Imani & Braga-Neto, 2018e; Imani, Ghoreishi, Allaire, & Braga-Neto, 2019; Kantas, Doucet, Singh, Maciejowski, Chopin, et al., 2015), the Extended Kalman filter (EKF) (Jazwinski, 1970), the Unscented Kalman filter (UKF) (Julier, Uhlmann, & Durrant-Whyte, 1995), and the Sigma-Point Kalman filter (SPKF) (Van Der Merwe, 2004) must be used. It should be noted that the exact algorithms for the computation of the optimal MMSE estimator are available in the case of a linear-Gaussian state space through the classical Kalman Filter and Smoother (Kalman, 1960), and for POBDS through BKF and BKS.

Exact calculation of the aforementioned optimal estimators requires complete information about the system model; however, in many real-world applications, the system parameters are not fully known and must be estimated. Several techniques have been developed for approximate estimation of general nonlinear non-Gaussian state space models with unknown parameters. The methods can be divided into two main categories of Maximum-Likelihood (ML) and Bayesian techniques. The class of ML techniques includes: (1) *direct gradient-based ML techniques*, where the idea is to maximize the log-likelihood function using gradient-ascent or quasi-Newton techniques (DeJong, Liesenfeld,

☆ The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Thomas Bo Schön under the direction of Editor Torsten Söderström.

* Corresponding author.
E-mail addresses: mimani@gwu.edu (M. Imani), edward@ece.tamu.edu (E.R. Dougherty), ulisses@ece.tamu.edu (U. Braga-Neto).

https://doi.org/10.1016/j.automatica.2019.108609
0005-1098/

Moura, Richard, & Dharmarajan, 2012; Ionides, Bretó, & King, 2006; Johansen, Doucet, & Davy, 2008; Malik & Pitt, 2011), (2) *Expectation–Maximization (EM) techniques* (Schön, Wills, & Ninness, 2011; Wills, Schön, Ljung, & Ninness, 2013), where the idea is to maximize the "complete" log-likelihood function, as opposed to ML-based techniques which maximize the "incomplete" log-likelihood function, using the fact that maximizing the complete log-likelihood is easier than maximizing the incomplete one. There are several particle-based Bayesian techniques for the inference of general nonlinear state-space models (Lindsten, Jordan, & Schön, 2014; Urteaga, Bugallo, & Djurić, 2016; Whiteley, Andrieu, & Doucet, 2010). An important representative is the Particle Marginal Metropolis–Hastings (PMMH) method (Andrieu, Doucet, & Holenstein, 2010). Several online particle-based techniques have also been developed for applications when fully-recursive estimation is desired (Crisan, Miguez, et al., 2018).

For POBDS under model uncertainty, maximum-likelihood (ML) and maximum a posteriori (MAP) adaptive estimators were proposed in Imani and Braga-Neto (2015a, 2017a, 2017b), respectively. These techniques are built on ML and MAP point-based estimators for unknown parameters combined with optimal MMSE state estimators for the state. The drawback of these approaches is their sensitivity to initialization and the requirement of large amount of data for good performance.

We propose in this paper instead an optimal Bayesian filter (OBF) approach to the problem of POBDS recursive estimation. The basic principle is that the unknown true model belongs to an uncertainty class of models and the OBF minimizes the expected cost over the uncertainty class. The idea has roots going back to the 1960s in control theory (Martin, 1967; Silver, 1963), but has more recently applied in a fully optimized form with intrinsically Bayesian optimal (IBR) filters, in which optimization is relative to a prior distribution (Dalton & Dougherty, 2014), and with optimal Bayesian filters, in particular, regression, where optimization is relative to a posterior distribution (Qian & Dougherty, 2016). These concepts have also been recently applied to classification in the form of optimal Bayesian classifiers (Dalton & Dougherty, 2013a) and IBR classifiers (Dalton & Dougherty, 2013b). Directly relevant to the developments in the current paper is their application in recursive linear filtering: the IBR Kalman filter (Dehghannasiri, Esfahani, & Dougherty, 2017), the optimal Bayesian Kalman filter, which uses the data to update the prior, thereby producing superior filtering to the IBR Kalman filter (Dehghannasiri, Esfahani, Qian, & Dougherty, 2018), and the optimal Bayesian Kalman smoother (Dehghannasiri & Dougherty, 2018).

Here, we extend the BKF and BKS to the cases where POBDS is under model uncertainty. The methods are optimal relative to the posterior distribution of the parameters. When the parameter space is discrete and finite, exact algorithms based on an efficient, recursive matrix-based implementation are introduced. These algorithms contain a bank of BKFs/BKSs in parallel, which is reminiscent of the multiple model adaptive estimation (MMAE) procedure for linear systems (Magill, 1965; Maybeck, 1982). These algorithms can be seen as generalizations of the regular BKF and BKS. For general parameter spaces, an approximate Markov-Chain Monte-Carlo (MCMC) implementation of the optimal Bayesian estimators is described. Via numerical examples, the performances of these filters are compared to that of the ML, MAP, and IBR estimators introduced in Dalton and Dougherty (2014) and Imani and Braga-Neto (2017a, 2017b), respectively.

The article is organized as follows. In Section 2, the POBDS signal model is introduced and its optimal MMSE estimators are briefly described. In Section 3, first the POBDS under model uncertainty is introduced, followed by the exact algorithms for computation of optimal Bayesian estimators in the case of discrete parameter space and the approximate MCMC solution for continuous parameter space. Section 4 contains numerical examples using POBDS models of gene regulatory networks observed through noisy measurements. Finally, Section 5 contains concluding remarks.

2. Partially-Observed Boolean Dynamical Systems (POBDS)

2.1. POBDS signal model

The POBDS model consists of a state model that describes the evolution of the Boolean dynamical system and an observation model that relates the state to the system output (measurements).

The state model is defined as:

$$\mathbf{X}_k = \mathbf{f}(\mathbf{X}_{k-1}, \mathbf{u}_k, \mathbf{n}_k), \quad k = 1, 2, \ldots \tag{1}$$

where $\mathbf{X}_k \in \{0,1\}^d$ represents the state of d Boolean state variables of the system at time k, $\mathbf{u}_k \in \mathrm{U}$ is the input at time step k which is assumed to be deterministic and known, and the process noise $\mathbf{n}_k$ is i.i.d. with arbitrary distribution, which is independent of $\mathbf{X}_0$. A simple example of a state process corresponds to additive Boolean input and noise, with

$$\mathbf{X}_k = \mathbf{f}(\mathbf{X}_{k-1}) \oplus \mathbf{u}_k \oplus \mathbf{n}_k, \quad k = 1, 2, \ldots \tag{2}$$

where $\mathbf{u}_k, \mathbf{n}_k \in \{0,1\}^d$ and "$\oplus$" denotes componentwise binary addition (the exclusive-or logic operation). In this case, the input and noise perturb the state by flipping the state of individual variables. However, the methodology in this paper assumes the general model in (1) where only the state must be a Boolean vector.

The state vector is observed indirectly through the general nonlinear observation model:

$$\mathbf{Y}_k = \mathbf{g}(\mathbf{X}_k, \mathbf{v}_k), \quad k = 1, 2, \ldots \tag{3}$$

where $\mathbf{Y}_k$ is a vector of (typically non-Boolean) measurements, and $\mathbf{v}_k$ is i.i.d. observation noise process with arbitrary distribution and independent of the $\mathbf{n}_k$ process. For more information, see Imani and Braga-Neto (2016a, 2016b, 2017b, 2017c, 2018a, 2018b, 2018d, 2019a, 2019b).

2.2. Optimal state estimators for POBDS

The general state estimation problem consists of finding an estimator $\hat{\mathbf{X}}_{r|k}$ of state $\mathbf{X}_r$ given the sequence of observations $\mathbf{Y}_{1:k} = (\mathbf{Y}_1, \ldots \mathbf{Y}_k)$ up to time k, where $r, k \geq 1$. As optimality criterion, we consider the mean-square error (MSE):

$$C(\mathbf{X}_r, \hat{\mathbf{X}}_{r|k}) = E\left[\|\mathbf{X}_r - \hat{\mathbf{X}}_{r|k}\|^2\right], \tag{4}$$

where $\|.\|_2$ is the L_2 norm vector. The optimal minimum mean-square error (MMSE) state estimator is

$$\hat{\mathbf{X}}^{\mathrm{MS}}_{r|k} = \operatorname*{argmin}_{\hat{\mathbf{x}}_{r|k} \in \Psi} C(\mathbf{X}_r, \hat{\mathbf{X}}_{r|k}), \tag{5}$$

where Ψ is the set of all Boolean estimators.

For a vector $\mathbf{v} \in [0,1]^d$, define the thresholding operator $\mathbf{v}^{-} \in \{0,1\}^d$ as $\overline{\mathbf{v}}(i) = 1$ if $\mathbf{v}(i) > 1/2$ and 0 otherwise, for $i = 1, \ldots d$,, respectively. It was shown in Braga-Neto (2011) and Imani and Braga-Neto (2017a) that

$$\hat{\mathbf{X}}^{\mathrm{MS}}_{r|k} = \overline{E[\mathbf{X}_r \mid \mathbf{Y}_{1:k}]}, \tag{6}$$

with optimal conditional MSE

$$C^{\mathrm{MS}}_{r|k} = C(\mathbf{X}_r, \hat{\mathbf{X}}^{\mathrm{MS}}_{r|k}) = \sum_{i=1}^{d} \min\left\{E[\mathbf{X}_r(i) \mid \mathbf{Y}_{1:k}], 1 - E[\mathbf{X}_r(i) \mid \mathbf{Y}_{1:k}]\right\}. \tag{7}$$

The contribution of each Boolean variable $\mathbf{X}_r(i)$ to $C^{\text{MS}}_{r|k}$ is $\min\{E[\mathbf{X}_r(i) \mid \mathbf{Y}_{1:k}], 1 - E[\mathbf{X}_r(i) \mid \mathbf{Y}_{1:k}]\}$. By using the identity $\min\{a, 1-a\} = 1/2 - |a - 1/2|$, for $0 \leq a \leq 1$, $C^{\text{MS}}_{r|k}$ can be written more compactly as:

$$C^{\text{MS}}_{r|k} = \frac{d}{2} - \sum_{i=1}^{d} \left| E[\mathbf{X}_r(i) \mid \mathbf{Y}_{1:k}] - \frac{1}{2} \right|. \tag{8}$$

This reveals that the maximum value for $C^{\text{MS}}_{r|k}$ is $d/2$ (in fact, the maximum contribution of each Boolean variable is $1/2$).

The optimal MMSE filter and smoother, corresponding to $r = k$ and $r < k$ in (6), are called respectively the Boolean Kalman filter (BKF) and Boolean Kalman smoother (BKS) (Braga-Neto, 2011; Imani & Braga-Neto, 2015a, 2017a, 2018e). It is also possible to define a Boolean Kalman predictor (BKP), for the case $r > k$ (Imani & Braga-Neto, 2017a), but this estimator will be of no further interest here.

We describe below the exact algorithms for BKF and BKS computation. Let $(\mathbf{x}^1, \ldots, \mathbf{x}^{2^d})$ be an arbitrary enumeration of the possible state vectors. Define the state conditional probability distribution vector as:

$$\boldsymbol{\pi}_{r|k}(i) = P\left(\mathbf{X}_r = \mathbf{x}^i \mid \mathbf{Y}_{1:k}\right), \quad i = 1, \ldots, 2^d, \tag{9}$$

for $r, k \geq 0$. According to Eqs. (6) and (8),

$$\hat{\mathbf{X}}^{\text{MS}}_{k|k} = \overline{E[\mathbf{X}_k \mid \mathbf{Y}_{1:k}]} = \overline{A\boldsymbol{\pi}_{k|k}}, \quad k = 1, 2, \ldots \tag{10}$$

where $A = [\mathbf{x}^1 \cdots \mathbf{x}^{2^d}]$ is a matrix of size $d \times 2^d$. Notice that in (10), the posterior distribution of the Boolean state ($\boldsymbol{\pi}_{k|k}$), vector of size 2^d, is mapped to a vector of size d ($A\boldsymbol{\pi}_{k|k}$), where each element denotes the probability that the corresponding Boolean variable is 1. Meanwhile, the optimal conditional MSE can be computed as

$$C^{\text{MS}}_{k|k} = \frac{d}{2} - \sum_{i=1}^{d} \left| (A\boldsymbol{\pi}_{k|k})_i - \frac{1}{2} \right|, \quad k = 1, 2, \ldots. \tag{11}$$

The computation of $\boldsymbol{\pi}_{k|k}$ can be performed recursively. First, notice that

$$\boldsymbol{\pi}_{k|k-1} = M_k \boldsymbol{\pi}_{k-1|k-1}, \quad k = 1, 2, \ldots \tag{12}$$

where M_k is the transition matrix of the Markov state process, with entries given by:

$$(M_k)_{ij} = P(\mathbf{X}_k = \mathbf{x}^i \mid \mathbf{X}_{k-1} = \mathbf{x}^j), \; i, j = 1, \ldots, 2^d. \tag{13}$$

On the other hand,

$$\boldsymbol{\pi}_{k|k} \propto T(\mathbf{Y}_k)\, \boldsymbol{\pi}_{k|k-1}, \quad k = 1, 2, \ldots \tag{14}$$

where "$\propto$" means that the result must be normalized to add up to 1, and $T(\mathbf{Y}_k)$ is the *update matrix*, which is a diagonal matrix of size $2^d \times 2^d$ with diagonal elements:

$$\left(T_k(\mathbf{Y}_k)\right)_{ii} = p\left(\mathbf{Y}_k \mid \mathbf{X}_k = \mathbf{x}^i\right), \quad i = 1, \ldots, 2^d. \tag{15}$$

The procedure is summarized in Algorithm 1.

We describe below the optimal fixed-interval smoother, which estimates the state at each time point r in the interval $0 < r \leq k$. Define the probability distribution vector $\Delta_{r|s}$ via

$$\Delta_{r|s}(i) = p\left(\mathbf{Y}_{s+1}, \ldots, \mathbf{Y}_k \mid \mathbf{X}_r = \mathbf{x}^i\right), \quad i = 1, \ldots, 2^d, \tag{16}$$

for $r, s = 0, \ldots, k$, where $\Delta_{k|k}$ is defined to be $\mathbf{1}_{2^d \times 1}$, the vector with all components equal to 1. The vector $\Delta_{r|s}$ satisfies a backward recursion similar to the forward recursion satisfied by $\boldsymbol{\pi}_{r|k}$ (see Algorithm 2 below). It has been shown that Imani and Braga-Neto (2017a):

$$\boldsymbol{\pi}_{r|k} \propto \boldsymbol{\pi}_{r|r-1} \circ \Delta_{r|r-1}, \quad r = 1, \ldots, k, \tag{17}$$

where "$\circ$" denotes the componentwise multiplication of two vectors. According to Eqs. (6) and (8),

$$\hat{\mathbf{X}}^{\text{MS}}_{r|k} = \overline{E[\mathbf{X}_r \mid \mathbf{Y}_{1:k}]} = \overline{A\boldsymbol{\pi}_{r|k}}, \quad r = 1, \ldots, k, \tag{18}$$

with optimal conditional MSE

$$C^{\text{MS}}_{r|k} = \frac{d}{2} - \sum_{i=1}^{d} \left| (A\boldsymbol{\pi}_{r|k})_i - \frac{1}{2} \right|, \quad r = 1, \ldots, k, \tag{19}$$

The entire procedure of fixed-interval BKS is given in Algorithm 2. Notice that the BKF and BKS both provide the posterior distribution of the Boolean state as well as the point-based MMSE estimation of the state. In addition, the complexity of BKF and BKS both grows exponentially with the number of Boolean variables, due to the transition matrix involved in their process. For large systems (i.e., large d), efficient particle filter implementation of these tools is provided in Imani and Braga-Neto (2018e).

Algorithm 1 BKF: Boolean Kalman Filter

1: Initialization: $\boldsymbol{\pi}_{0|0}(i) = P\left(\mathbf{X}_0 = \mathbf{x}^i\right)$, for $i = 1, \ldots, 2^d$.
2: **for** $k = 1, 2, \ldots$ **do**
3: Prediction: $\boldsymbol{\pi}_{k|k-1} = M_k \boldsymbol{\pi}_{k-1|k-1}$.
4: Update: $\boldsymbol{\pi}_{k|k} \propto T(\mathbf{Y}_k)\, \boldsymbol{\pi}_{k|k-1}$.
5: MMSE Estimator: $\hat{\mathbf{X}}^{\text{MS}}_{k|k} = \overline{A\boldsymbol{\pi}_{k|k}}$.
6: Optimal MSE: $C^{\text{MS}}_{k|k} = \frac{d}{2} - \sum_{i=1}^{d} \left| (A\boldsymbol{\pi}_{k|k})_i - \frac{1}{2} \right|$.
7: **end for**

Algorithm 2 BKS: Fixed-Interval Boolean Kalman Smoother

1: Initialization: $\boldsymbol{\pi}_{0|0}(i) = P\left(\mathbf{X}_0 = \mathbf{x}^i\right)$, for $i = 1, \ldots, 2^d$.

Forward Probabilities:

2: **for** $r = 1, \ldots, k$ **do**
3: Prediction: $\boldsymbol{\pi}_{r|r-1} = M_r \boldsymbol{\pi}_{r-1|r-1}$.
4: Update: $\boldsymbol{\pi}_{r|r} \propto T(\mathbf{Y}_r)\, \boldsymbol{\pi}_{r|r-1}$.
5: **end for**

Backward Probabilities:

6: Initialization: $\Delta_{k|k} = \mathbf{1}_{2^d \times 1}$.
7: **for** $r = k, k-1, \ldots, 1$ **do**
8: Update: $\Delta_{r|r-1} = T_r(\mathbf{Y}_r)\, \Delta_{r|r}$.
9: Prediction: $\Delta_{r-1|r-1} = M_r^T \Delta_{r|r-1}$.
10: **end for**

MMSE Estimator Computation:

11: **for** $r = 1, \ldots, k$ **do**
12: $\boldsymbol{\pi}_{r|k} \propto \boldsymbol{\pi}_{r|r-1} \circ \Delta_{r|r-1}$.
13: $\hat{\mathbf{X}}^{\text{MS}}_{r|k} = \overline{A\boldsymbol{\pi}_{r|k}}$.
14: $C^{\text{MS}}_{r|k} = \frac{d}{2} - \sum_{i=1}^{d} \left| (A\boldsymbol{\pi}_{r|k})_i - \frac{1}{2} \right|$.
15: **end for**

3. Optimal estimators for POBDS under model uncertainty

3.1. POBDS under model uncertainty

In many practical problems, full information about the system model is not available. There might be uncertainty about the transition or observation functions or noise statistics. We assume the uncertainty is parameterized by a vector $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_l]$ of unknown parameters, where $\boldsymbol{\theta}$ takes a value in a set Θ, called the *uncertainty class*. The POBDS model can then be expressed as:

$$\mathbf{X}_k = \mathbf{f}(\mathbf{X}_k, \mathbf{u}_k, \mathbf{n}_k, \boldsymbol{\theta}), \quad k = 1, 2, \ldots$$
$$\mathbf{Y}_k = \mathbf{g}(\mathbf{X}_k, \mathbf{v}_k, \boldsymbol{\theta}), \quad k = 1, 2, \ldots$$

Direct application of the algorithms described in the previous section is not possible due to the presence of unknown parameters, and suboptimal adaptive estimators must be employed. Maximum-likelihood (ML) and maximum a posteriori (MAP) adaptive estimators for partially-known POBDS were developed in Imani and Braga-Neto (2017a, 2017b) respectively. Despite the generally good performance of these estimators in simultaneous state and parameter estimation, they are sensitive to initialization and require large sample sizes.

In this paper, we address these deficiencies by proposing optimal Bayesian Kalman filtering for recursive POBDS estimation. Consider the estimation cost associated with each realization of the parameter:

$$C_{\boldsymbol{\theta}}(\mathbf{X}_r, \hat{\mathbf{X}}_{r|k}) = E\left[\|\mathbf{X}_r - \hat{\mathbf{X}}_{r|k}\|^2 \mid \boldsymbol{\theta}\right], \tag{20}$$

The optimal-Bayesian-estimator (OBE) Boolean recursive filter is defined by

$$\hat{\mathbf{X}}_{r|k}^{\text{OBE}} = \operatorname*{argmin}_{\hat{\mathbf{x}}_{r|k} \in \Psi} E_{\boldsymbol{\theta}|\mathbf{Y}_{1:k}}\left[C_{\boldsymbol{\theta}}(\mathbf{X}_r, \hat{\mathbf{X}}_{r|k})\right]. \tag{21}$$

The OBE recursive filter provides the optimal solution relative to the posterior distribution $p(\boldsymbol{\theta} \mid \mathbf{Y}_{1:k})$. It can be shown (see the Appendix) that it is given by

$$\hat{\mathbf{X}}_{r|k}^{\text{OBE}} = \overline{E_{\boldsymbol{\theta}|\mathbf{Y}_{1:k}}[E[\mathbf{X}_r \mid \mathbf{Y}_{1:k}, \boldsymbol{\theta}]]}. \tag{22}$$

with optimal conditional MSE

$$C_{r|k}^{\text{OBE}} = \frac{d}{2} - \sum_{i=1}^{d} \left| E_{\boldsymbol{\theta}|\mathbf{Y}_{1:k}}[E[\mathbf{X}_r(i) \mid \mathbf{Y}_{1:k}, \boldsymbol{\theta}]] - \frac{1}{2} \right|. \tag{23}$$

As in the case of the optimal overall MSE, it is easy to see that the cost is upper-bounded by $d/2$.

3.2. Optimal estimator for POBDS under Discrete Model Uncertainty

Let us first consider the case in which the parameter space is discrete and finite: $\Theta = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_M\}$. In this section, we introduce algorithms for the exact computation of both the optimal Bayesian filter and smoother in this case.

3.2.1. Boolean Kalman Filter under Discrete Model Uncertainty (BKF-DMU)

Given the sequence of measurements $\mathbf{Y}_{1:k}$, define the state conditional distribution vector associated with $\boldsymbol{\theta}_m \in \Theta$ as:

$$\boldsymbol{\pi}_{r|k}^{\boldsymbol{\theta}_m}(i) = P\left(\mathbf{X}_r = \mathbf{x}^i \mid \mathbf{Y}_{1:k}, \boldsymbol{\theta}_m\right), \quad i = 1, \ldots 2^d, \tag{24}$$

for $r, k \geq 0$, where $\boldsymbol{\pi}_{0|0}^{\boldsymbol{\theta}_m} = P\left(\mathbf{X}_r = \mathbf{x}^i \mid \boldsymbol{\theta}_m\right)$ is the initial (prior) distribution of the states associated with $\boldsymbol{\theta}_m$.

Let also $M_k^{\boldsymbol{\theta}}$ be the transition matrix of the state process associated to model $\boldsymbol{\theta}$, defined as:

$$\left(M_k^{\boldsymbol{\theta}_m}\right)_{ij} = P(\mathbf{X}_k = \mathbf{x}^i \mid \mathbf{X}_{k-1} = \mathbf{x}^j, \boldsymbol{\theta}_m), \quad i, j = 1, \ldots 2^d, \tag{25}$$

for $m = 1, \ldots M$,

Additionally, given a value of the observation vector $\mathbf{Y}_k$ at time k, the *update matrix* $T_k^{\boldsymbol{\theta}}(\mathbf{Y}_k)$ associated to model $\boldsymbol{\theta}$ is a diagonal matrix of size $2^d \times 2^d$, defined as:

$$\left(T_k^{\boldsymbol{\theta}_m}(\mathbf{Y}_k)\right)_{ii} = p\left(\mathbf{Y}_k \mid \mathbf{X}_k = \mathbf{x}^i, \boldsymbol{\theta}_m\right), \quad i = 1, \ldots 2^d, \tag{26}$$

for $m = 1, \ldots M$,

The optimal Bayesian filter corresponds to the case $r = k$ in (22). Its computation requires the quantity

$$\mathbf{z}_{k|k} = E_{\boldsymbol{\theta}|\mathbf{Y}_{1:k}}\left[E[\mathbf{X}_k \mid \mathbf{Y}_{1:k}, \boldsymbol{\theta}] \mid \mathbf{Y}_{1:k}\right] = \sum_{m=1}^{M} A\, \boldsymbol{\pi}_{k|k}^{\boldsymbol{\theta}_m} p_k(\boldsymbol{\theta}_m), \tag{27}$$

where $p_k(\boldsymbol{\theta}) = P(\boldsymbol{\theta} \mid \mathbf{Y}_{1:k})$ is the posterior probability distribution of $\boldsymbol{\theta}$ at time k. From (22) and (23), the optimal Bayesian filter at time k is then given by

$$\hat{\mathbf{X}}_{k|k}^{\text{OBE}} = \overline{\mathbf{z}_{k|k}}, \tag{28}$$

and the optimal conditional MSE can be computed according to (23).

The previous computation can be performed exactly and efficiently by running M BKFs in parallel, each tuned to a different value of the parameter, as we show next. First, notice that the posterior probabilities at time k can be computed via the following Bayesian recursion (Ghoreishi, 2019; Ghoreishi & Allaire, 2017, 2019; Ghoreishi, Friedman, & Allaire, 2019):

$$p_k(\boldsymbol{\theta}_m) \propto p(\mathbf{Y}_k \mid \mathbf{Y}_{1:k-1}, \boldsymbol{\theta}_m)\, p_{k-1}(\boldsymbol{\theta}_m), \tag{29}$$

for $m = 1, \ldots M$, with $p_0(\boldsymbol{\theta})$ as the prior probability of model $\boldsymbol{\theta} \in \Theta$ holding $\sum_{\boldsymbol{\theta} \in \Theta} p_0(\boldsymbol{\theta}) = 1$. Now,

$$\begin{aligned} &p(\mathbf{Y}_k \mid \mathbf{Y}_{1:k-1}, \boldsymbol{\theta}_m) \\ &= \sum_{j=1}^{d} p(\mathbf{Y}_k \mid \mathbf{X}_k = \mathbf{x}^j, \boldsymbol{\theta}_m)\, P(\mathbf{X}_k = \mathbf{x}^j \mid \mathbf{Y}_{1:k-1}, \boldsymbol{\theta}_m) \\ &= \sum_{j=1}^{d} \left(T_k^{\boldsymbol{\theta}_m}(\mathbf{Y}_k)\right)_{jj} \boldsymbol{\pi}_{k|k-1}^{\boldsymbol{\theta}_m}(j) = \|\boldsymbol{\beta}_k^{\boldsymbol{\theta}_m}\|_1, \end{aligned} \tag{30}$$

where $\boldsymbol{\beta}_k^{\boldsymbol{\theta}_m} = T_k^{\boldsymbol{\theta}_m}(\mathbf{Y}_k)\, \boldsymbol{\pi}_{k|k-1}^{\boldsymbol{\theta}_m}$ and $\|\mathbf{v}\|_1 = \sum_{i=1}^{2^d} \mathbf{v}(i)$. Combining (29) and (30), we obtain

$$p_k(\boldsymbol{\theta}_m) \propto \|\boldsymbol{\beta}_k^{\boldsymbol{\theta}_m}\|_1\, p_{k-1}(\boldsymbol{\theta}_m), \tag{31}$$

for $m = 1, \ldots M$, Notice that $\boldsymbol{\beta}_k^{\boldsymbol{\theta}_m}$ is the unnormalized state posterior probability distribution at time k, which is computed in the update step of the BKF tuned to parameter $\boldsymbol{\theta}_m$ (see Algorithm 1). This allows the computation of (27) and of the optimal state estimator. An optimal estimator of the parameter is also readily available based on the posterior $p_k(\boldsymbol{\theta})$. For example, the MAP estimator is

$$\hat{\boldsymbol{\theta}}_k^{\text{MAP}} = \operatorname*{argmax}_{\boldsymbol{\theta} \in \Theta} p_k(\boldsymbol{\theta}), \tag{32}$$

providing a full joint state and parameter estimation approach.

The entire procedure is presented in Algorithm 3 and represented as a diagram in Fig. 1. The BKF-DMU recursively computes the posterior distribution of the unknown parameter and state as well as their point-based estimations. The posterior distribution of the unknown parameter at time step k is denoted by $[p_k(\boldsymbol{\theta}_1), \ldots p_k(\boldsymbol{\theta}_M)]$ and the posterior distribution of the state can be computed by $\sum_{m=1}^{M} p_k(\boldsymbol{\theta}_m) \boldsymbol{\pi}_{k|k}^{\boldsymbol{\theta}_m}$. Notice that the BKF-DMU has similarity with the multiple model adaptive estimation (MMAE) procedure developed for the linear-Gaussian state-space model (Magill, 1965; Maybeck, 1982) and later extended for estimation of the nonlinear/non-Gaussian state-space model using

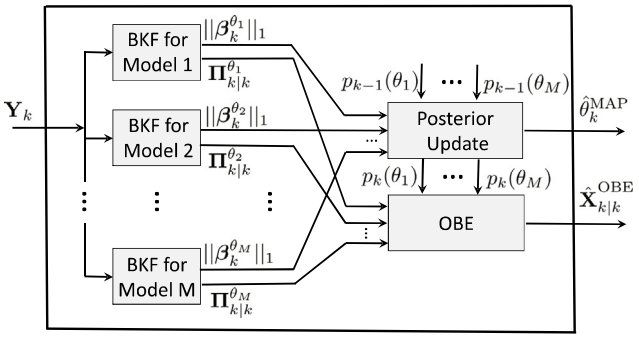

Fig.1. Schematic representation of the proposed optimal Boolean Kalman Filter under Discrete Model Uncertainty (BKF-DMU).

a particle filtering scheme (Doucet, Gordon and Krishnamurthy, 2001; Martino, Read, Elvira, & Louzada, 2017).

As with the ordinary BKF in Algorithm 1, the BKF-DMU is fully online, i.e., as new measurements arrive, the optimal Bayesian estimate of state can be computed recursively. In fact, when uncertainty about the parameter decreases to zero, the BKF-DMU reduces to the BKF. The computational complexity of BKF-DMU is of order $O(M2^{2d})$ at each time point.

Algorithm 3 BKF-DMU: Boolean Kalman Filter under Discrete Model Uncertainty

1: Initialization: $p_0(\boldsymbol{\theta}_m)$, for $m = 1, \ldots, M$.
2: Run M BKFs, each one tuned to a different $\boldsymbol{\theta}_m \in \Theta$.
3: **for** $k = 1, 2, \ldots$ **do**
4: Posterior Update: Using the outputs $||\boldsymbol{\beta}_k^{\boldsymbol{\theta}_m}||_1$ of the BKF bank, update the posterior of each parameter as:

$$p_k(\boldsymbol{\theta}_m) \propto ||\boldsymbol{\beta}_k^{\boldsymbol{\theta}_m}||_1 \, p_{k-1}(\boldsymbol{\theta}_m), \quad m = 1, \ldots, M.$$

5: Using the outputs $\boldsymbol{\Pi}_{k|k}^{\boldsymbol{\theta}}$ of the bank of BKFs, compute:

$$\mathbf{z}_{k|k} = \sum_{m=1}^{M} p_k(\boldsymbol{\theta}_m) A \boldsymbol{\Pi}_{k|k}^{\boldsymbol{\theta}_m}.$$

6: Optimal Bayesian state estimator:

$$\hat{\mathbf{X}}_{k|k}^{\text{OBE}} = \overline{\mathbf{z}_{k|k}}.$$

7: Conditional optimal MSE of state estimator:

$$C_{k|k}^{\text{OBE}} = \frac{d}{2} - \sum_{i=1}^{d} \left| \left(\mathbf{z}_{k|k} \right)_i - \frac{1}{2} \right|.$$

8: Optimal Bayesian parameter estimator:

$$\hat{\boldsymbol{\theta}}_k^{\text{MAP}} = \underset{\boldsymbol{\theta} \in \Theta}{\text{argmax}} \; p_k(\boldsymbol{\theta}),$$

9: **end for**

There are numerous differences between the BKF-DMU and the optimal Bayesian Kalman filter (OBKF) in Dehghannasiri et al. (2018). Most important is the fact that, whereas the BKF-DMU employs a bank of ordinary BKFs each possessing its own update matrix, the OBKF is solved by a set of recursive equations taking a similar form to the equations for the ordinary Kalman filter except that various statistics are replaced by effective statistics, which are related to the entire uncertainty class via the posterior distribution, and the ordinary Kalman gain matrix is replaced by the effective Kalman gain matrix. This is analogous to the structure of IBR filters in Dalton and Dougherty (2014) where for Wiener filtering the power spectra are replaced by effective power spectra, and for morphological filtering the granulometric size densities are replaced by the effective granulometric size densities. The basic idea is to replace the characteristics determining the ordinary solution with effective characteristics that are global with respect to the model uncertainty. This methodology constitutes a general paradigm for optimal signal processing under uncertainty (Dougherty, 2018).

In the present situation, there is no notion of an effective update matrix. In (26) an update matrix is formed for each parameter vector, and these are used jointly in (30). The difficulty is the discrete state space of the underlying Boolean dynamical system. In fact, the notion of an IBR operator was first introduced in Yoon, Qian, and Dougherty (2013), where the problem was to choose an operator, from a class of operators, to minimize the expected undesirable steady-state mass of an uncertain hidden Markov model, the application being to optimally perturb the logic of gene regulatory networks modeled as Boolean networks. As with the BKF-DMU, there were no effective characteristics with which to frame the solution; indeed, the optimal operator was obtained by computing the cost of each operator and choosing the one possessing minimum cost. This kind of direct minimization has many biomedical applications, such as in Mohsenizadeh, Dehghannasiri, and Dougherty (2016), where intervention consists of interrupting a subset of interactions in order to obtain more desired dynamics.

3.2.2. *Boolean Kalman Smoother under Discrete Model Uncertainty (BKS-DMU)*

The extension of the fixed-interval smoother in Algorithm 2 to the Bayesian case is considered here. Given the sequence of measurements $\mathbf{Y}_{1:k}$, we define the backward probability distribution vector associated with parameter $\boldsymbol{\theta}_m$ as:

$$\Delta_{r|s}^{\boldsymbol{\theta}_m}(i) = p(\mathbf{Y}_{s+1}, \ldots \mathbf{Y}_k \,|\, \mathbf{X}_r = \mathbf{x}^i, \boldsymbol{\theta}_m), \quad i = 1, \ldots 2^d, \tag{33}$$

for $r, s = 0, \ldots k$, where $\Delta_{k|k}^{\boldsymbol{\theta}_m}$ is defined to be $\mathbf{1}_{2^d \times 1}$.

According to (22), the optimal Bayesian state smoother requires the computation of

$$\mathbf{z}_{r|k} = E_{\boldsymbol{\theta}|\mathbf{Y}_{1:k}}[E[\mathbf{X}_r \,|\, \mathbf{Y}_{1:k}, \boldsymbol{\theta}]] = \sum_{m=1}^{M} A \boldsymbol{\Pi}_{r|k}^{\boldsymbol{\theta}_m} p_k(\boldsymbol{\theta}_m), \tag{34}$$

with the optimal Bayesian smoother given by

$$\hat{\mathbf{X}}_{r|k}^{\text{OBE}} = \overline{\mathbf{z}_{r|k}}, \tag{35}$$

with optimal conditional MSE given by (23)

$$C_{r|k}^{\text{OBE}} = \frac{d}{2} - \sum_{i=1}^{d} \left| \left(\mathbf{z}_{r|k} \right)_i - \frac{1}{2} \right|. \tag{36}$$

As in the BKF-DMU case, this computation can be performed by executing M BKSs in parallel, each tuned to a different value of the parameter. The posterior probability distribution of $\boldsymbol{\theta}$ given all observations $\mathbf{Y}_{1:k}$ can be computed by applying (31) repeatedly to obtain

$$p_k(\boldsymbol{\theta}_m) \propto p_0(\boldsymbol{\theta}_m) \prod_{r=1}^{k} ||\boldsymbol{\beta}_r^{\boldsymbol{\theta}_m}||_1, \tag{37}$$

for $m = 1, \ldots M$. The optimal MAP estimator of the parameter can be computed as in (32). The entire procedure is presented in Algorithm 4 and represented as a diagram in Fig. 2. Notice that given a time series of length k, the BKS-DMU computes the posterior distribution of the unknown parameters (i.e., $[p_r(\boldsymbol{\theta}_1), \ldots, p_r(\boldsymbol{\theta}_M)]$, $r = 1, \ldots k$), in a forward process and the smoothed distribution of the Boolean state in a backward process (i.e., $\sum_{m=1}^{M} p_r(\boldsymbol{\theta}_m) \boldsymbol{\Pi}_{r|k}^{\boldsymbol{\theta}_m}$, $r = 1, \ldots k$). The computational complexity of BKS-DMU for estimation over the whole interval is of order $O(Mk2^{2d})$.

Algorithm 4 BKS-DMU: Boolean Kalman Smoother under Discrete Model Uncertainty

1: Initialization: $p_0(\boldsymbol{\theta}_m)$, for $m = 1, \ldots, M$.
2: Run M BKSs, each one tuned to a different $\boldsymbol{\theta}_m \in \Theta$.
3: Posterior Calculation: Using the outputs $\{\|\boldsymbol{\beta}_r^{\boldsymbol{\theta}_m}\|_1\}_{r=1}^k$ of the BKSs, compute

$$p_k(\boldsymbol{\theta}_m) \propto p_0(\boldsymbol{\theta}_m) \prod_{r=1}^{k} \|\boldsymbol{\beta}_r^{\boldsymbol{\theta}_m}\|_1, \quad m = 1, \ldots, M.$$

4: Optimal Bayesian parameter estimator:

$$\hat{\boldsymbol{\theta}}_k^{\text{MAP}} = \underset{\boldsymbol{\theta}\in\Theta}{\text{argmax}}\ p_k(\boldsymbol{\theta}).$$

5: **for** $r = 1, \ldots, k$ **do**
6: Using the outputs $\boldsymbol{\Pi}_{r|k}^{\boldsymbol{\theta}}$ of the bank of BKSs, compute:

$$\mathbf{z}_{r|k} = \sum_{m=1}^{M} p_k(\boldsymbol{\theta}_m)\, A\boldsymbol{\Pi}_{r|k}^{\boldsymbol{\theta}_m}.$$

7: Optimal Bayesian state estimation:

$$\hat{\mathbf{x}}_{r|k}^{\text{OBE}} = \overline{\mathbf{z}_{r|k}}.$$

8: Conditional MSE:

$$C_{r|k}^{\text{OBE}} = \frac{d}{2} - \sum_{i=1}^{d} \left| \left(\mathbf{z}_{r|k}\right)_i - \frac{1}{2} \right|.$$

9: **end for**

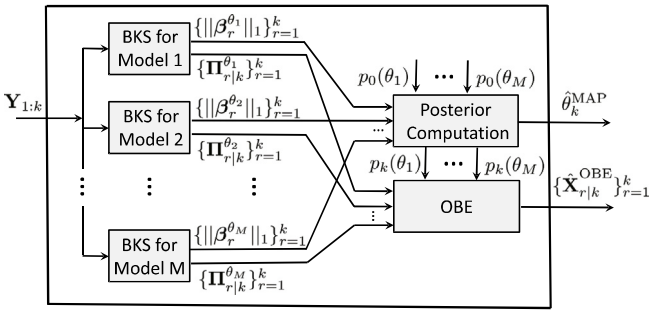

Fig. 2. Schematic representation of the proposed Boolean Kalman Smoother under Discrete Model Uncertainty (BKS-DMU).

3.3. Estimator for POBDS under continuous model uncertainty

When the parameter space Θ is continuous, e.g., $\Theta \subset \mathrm{R}^l$, the algorithms proposed in the previous section cannot be applied. This is due to the fact that the computation of the expectation in (22) requires integration over the posterior distribution of the unknown parameter. The exact computation of the posterior distribution is not possible for a general set of unknown parameters. A naive way of handling this problem is discretization of the parameter space. However, except for very simple cases, this will result in finite but very large parameter spaces, which also renders intractable the application of the previously proposed algorithms. In addition, the discretization process introduces error, which may lead to poor estimation performance. A popular framework for dealing with a continuous parameter space for a general nonlinear/non-Gaussian state-space model is the particle Markov chain Monte Carlo (PMCMC) framework (Andrieu et al., 2010). This framework relies on particle methods for approximation of the likelihood function, $p(\mathbf{Y}_{1:k} \mid \boldsymbol{\theta})$ for $\boldsymbol{\theta} \in \Theta$.

Unlike general nonlinear/non-Gaussian state-space models, the exact likelihood function can be computed for POBDS, using the BKF and BKS. Thus, we employ here the Metropolis–Hastings MCMC (Hastings, 1970) algorithm to obtain an approximate solution for POBDS under continuous model uncertainty. Let the current MCMC sample be $\boldsymbol{\theta}^{(j)}$. A candidate MCMC sample $\boldsymbol{\theta}^{\text{cand}}$ is drawn according to a symmetric proposal distribution $q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(j)})$. Then $\boldsymbol{\theta}^{(j+1)} = \boldsymbol{\theta}^{\text{cand}}$ with probability

$$\alpha = \min\left\{1, \frac{p(\mathbf{Y}_{1:k} \mid \boldsymbol{\theta}^{\text{cand}})\, p_0(\boldsymbol{\theta}^{\text{cand}})}{p(\mathbf{Y}_{1:k} \mid \boldsymbol{\theta}^{(j)})\, p_0(\boldsymbol{\theta}^{(j)})}\right\}, \tag{38}$$

with $\boldsymbol{\theta}^{(j+1)} = \boldsymbol{\theta}^{(j)}$, otherwise. In (38), $p_0(\boldsymbol{\theta})$ is the prior probability of $\boldsymbol{\theta}$ and $p(\mathbf{Y}_{1:k} \mid \boldsymbol{\theta})$ is computed by running a BKF (or the forward iteration of a BKS) tuned to parameter $\boldsymbol{\theta}$. The positivity condition $q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(j)}) > 0$ for all $\boldsymbol{\theta}^{(j)}$ guarantees an ergodic Markov chain, the steady-state distribution of which is the target distribution $p(\boldsymbol{\theta} \mid \mathbf{Y}_{1:k})$ (Gilks, Richardson, & Spiegelhalter, 1995). (a Gaussian distribution, which satisfies the positivity condition, is used in this paper as the proposal distribution.)

After discarding the first N_{Burn} "burn-in" MCMC samples, the next N_{MCMC} samples are used to approximate the posterior distribution as:

$$p_k(\boldsymbol{\theta}) \approx \frac{1}{N_{\text{MCMC}}} \sum_{j=1}^{N_{\text{MCMC}}} \delta(\boldsymbol{\theta} - \boldsymbol{\theta}^{(j)}), \tag{39}$$

where δ is the standard generalized function of calculus. Using the fact that

$$E_{\boldsymbol{\theta}|\mathbf{Y}_{1:k}}[E[\mathbf{x}_r \mid \mathbf{Y}_{1:k}, \boldsymbol{\theta}]] = \int_{\Theta} E[\mathbf{x}_r \mid \mathbf{Y}_{1:k}, \boldsymbol{\theta}]\, p_k(\boldsymbol{\theta})\, d\boldsymbol{\theta},$$

and

$$P(\mathbf{x}_r \mid \mathbf{Y}_{1:k}, \boldsymbol{\theta}) = A\boldsymbol{\Pi}_{r|k}^{\boldsymbol{\theta}},$$

obtained from the application of the BKF or BKS, using (22), the MCMC approximation can be written as:

$$E_{\boldsymbol{\theta}|\mathbf{Y}_{1:k}}[E[\mathbf{x}_r \mid \mathbf{Y}_{1:k}, \boldsymbol{\theta}]] \approx \frac{1}{N_{\text{MCMC}}} \sum_{j=1}^{N_{\text{MCMC}}} A\boldsymbol{\Pi}_{r|k}^{\boldsymbol{\theta}^{(j)}}, \tag{40}$$

and the optimal Bayesian estimator and optimal conditional MSE can be computed by applying (22) and (23), respectively. Thus, the BKF-CMU and BKS-CMU provide the approximate posterior distribution of the state and parameters along with their point-based estimations.

The entire process is summarized in Algorithm 5. The complexity of BKF-CMU is of order $O((N^{\text{Burn}} + N^{\text{MCMC}})k2^{2d})$. This computation can be intractable or very slow in the following two conditions: (1) Large Model Uncertainty: this requires selection of large MCMC samples, which linearly affects the complexity of Algorithm 5; (2) Large Systems: the complexity of Algorithm 5 grows exponentially with respect to the increase in the number of Boolean variables (i.e., d). One way of dealing with large systems is to replace the BKF and BKS with their particle filter implementation introduced in Imani and Braga-Neto (2018e). Dealing with these two conditions, which deals with scalability of the proposed methods, will be part of our future research.

Notice that several recommendations have been made for proper stopping criterion of the MCMC process. These methods mostly rely on the effective sample size which depends on the correlation between the trajectories of the Markov chain (Brooks, Gelman, Jones, & Meng, 2011). Furthermore, the burn-in sample size should ideally be selected large enough to guarantee with a high probability that the start of the MCMC procedure after the burn-in time is a sample from the posterior distribution. This selection can be challenging, and in practice needs to be selected according to the computational resources, dimension of the parameter space and the nature of the problem (for more information, see Gilks et al. (1995)).

Algorithm 5 BKF/BKS: under Continuous Model Uncertainty (BKF-CMU/BKS-CMU)

Set N_{Burn}, N_{MCMC} and $\mathbf{z}_{r|k} = \mathbf{0}$, $r = 1, \ldots, k$.,

Draw an initial sample: $\boldsymbol{\theta}^{\text{old}} \sim p_0(\boldsymbol{\theta})$.

Run a BKF tuned to $\boldsymbol{\theta}^{\text{old}}$ for computation of $L^{\boldsymbol{\theta}^{\text{old}}} = \sum_{r=1}^{k} \log ||\boldsymbol{\beta}^{\boldsymbol{\theta}^{\text{old}}}||_1$ and $\boldsymbol{\pi}_{r|k}^{\boldsymbol{\theta}^{\text{old}}}$, $r = 1, \ldots, k$.,

for $j = 1, 2, \ldots N_{\text{Burn}} + N_{\text{MCMC}}$ **do**

  Draw a sample from proposal: $\boldsymbol{\theta}^{\text{cand}} \sim q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{\text{old}})$.

  Run a BKF/BKS tuned to $\boldsymbol{\theta}^{\text{cand}}$ for computation of $L^{\boldsymbol{\theta}^{\text{cand}}} = \sum_{r=1}^{k} \log ||\boldsymbol{\beta}^{\boldsymbol{\theta}^{\text{cand}}}||_1$ and $\boldsymbol{\pi}_{r|k}^{\boldsymbol{\theta}^{\text{cand}}}$, $r = 1, \ldots, k$.,

  Acceptance rate: $\alpha = \min\left\{1, \frac{\exp(L^{\boldsymbol{\theta}^{\text{cand}}})\, p_0(\boldsymbol{\theta}^{\text{cand}})}{\exp(L^{\boldsymbol{\theta}^{\text{old}}})\, p_0(\boldsymbol{\theta}^{\text{old}})}\right\}$

  $(\boldsymbol{\theta}^{\text{new}}, L^{\boldsymbol{\theta}^{\text{new}}}) = \begin{cases} (\boldsymbol{\theta}^{\text{cand}}, L^{\boldsymbol{\theta}^{\text{cand}}}) & \text{with probability } \alpha \\ (\boldsymbol{\theta}^{\text{old}}, L^{\boldsymbol{\theta}^{\text{old}}}) & \text{otherwise} \end{cases}$

  if $j > N_{\text{Burn}}$ **then**

    $\mathbf{z}_{r|k} = \mathbf{z}_{r|k} + \frac{1}{N_{\text{MCMC}}} A \boldsymbol{\pi}_{r|k}^{\boldsymbol{\theta}^{\text{new}}}$, $r = 1, \ldots, k$.,

  end if

  $\boldsymbol{\theta}^{\text{old}} \leftarrow \boldsymbol{\theta}^{\text{new}}$, $L^{\boldsymbol{\theta}^{\text{old}}} \leftarrow L^{\boldsymbol{\theta}^{\text{new}}}$.

end for

Optimal Bayesian estimation:

$$\hat{\mathbf{x}}_{r|k}^{\text{OBE}} \approx \overline{\mathbf{z}_{r|k}}, \quad \text{for } r = 1, \ldots, k.,$$

Approximate conditional MSE:

$$C_{r|k}^{\text{OBE}} \approx \frac{d}{2} - \sum_{i=1}^{d} \left| \left(\overline{\mathbf{z}_{r|k}} \right)_i - \frac{1}{2} \right|, \quad r = 1, \ldots, k.,$$

4. Numerical results and performance analysis

In this section, we present the results of numerical experiments using a gene regulatory network model, which compare the performance of the proposed framework with four other approaches: (1) the optimal model-specific BKF/BKS (Braga-Neto, 2011; Imani & Braga-Neto, 2017a); (2) the maximum-likelihood (ML) adaptive BKF/BKS (Imani & Braga-Neto, 2017a); (3) the maximum a posteriori (MAP) adaptive BKF/BKS (Imani & Braga-Neto, 2017b); and (4) the intrinsically Bayesian robust (IBR) estimator (Dalton & Dougherty, 2014). These methods are defined next. The optimal "model-specific" BKF/BKS consists of the classical BKF/BKS relative to the underlying true parameters. These filters are therefore the baseline for performance.

The ML-BKF/ML-BKS selects the model with the largest log-likelihood and then obtains the state estimator by plug-in:

$$\hat{\boldsymbol{\theta}}_k^{\text{ML}} = \operatorname*{argmax}_{\boldsymbol{\theta} \in \Theta} \log p(\mathbf{Y}_{1:k} \mid \boldsymbol{\theta}),$$

$$\hat{\mathbf{x}}_{r|k}^{\text{ML}} = \overline{E[\mathbf{x}_r \mid \mathbf{y}_{1:k}, \hat{\boldsymbol{\theta}}_k^{\text{ML}}]}.$$

Computation of $\hat{\boldsymbol{\theta}}_k^{\text{ML}}$ involves either a bank of filters for discrete parameter spaces, or an expectation–maximization algorithm in the case of continuous parameter spaces (Imani & Braga-Neto, 2017a).

The MAP-BKF/MAP-BKS maximizes the posterior probability and then applies the plug-in approach:

$$\hat{\boldsymbol{\theta}}_k^{\text{MAP}} = \operatorname*{argmax}_{\boldsymbol{\theta} \in \Theta} p(\boldsymbol{\theta} \mid \mathbf{Y}_{1:k}),$$

$$\hat{\mathbf{x}}_{r|k}^{\text{MAP}} = \overline{E[\mathbf{x}_r \mid \mathbf{y}_{1:k}, \hat{\boldsymbol{\theta}}_k^{\text{MAP}}]}.$$

Finding $\hat{\boldsymbol{\theta}}_k^{\text{MAP}}$ relies on the computation of $p_k(\boldsymbol{\theta}) = p(\boldsymbol{\theta} \mid \mathbf{Y}_{1:k})$, described in the previous two sections.

The IBR-BKF/IBR-BKS is computed by solving the following minimization problem:

$$\hat{\mathbf{x}}_{r|k}^{\text{IBR}} = \operatorname*{argmin}_{\hat{\mathbf{x}}_{r|k} \in \Psi} E_{\boldsymbol{\theta}}\left[C_{\boldsymbol{\theta}}(\mathbf{x}_r, \hat{\mathbf{x}}_{r|k}) \right], \tag{41}$$

where the expectation is relative to the prior distribution $p(\boldsymbol{\theta})$. It can be shown that the solution is

$$\hat{\mathbf{x}}_{r|k}^{\text{IBR}} = \overline{E_{\boldsymbol{\theta}}[E[\mathbf{x}_r \mid \mathbf{y}_{1:k}, \boldsymbol{\theta}]]}. \tag{42}$$

Contrasting the previous two equations with (21) and (22), respectively, makes clear that the difference between OBE and IBR estimators is that the latter are optimized with respect to the prior distribution of the parameter, where as the former are based on the posterior distribution, and therefore are expected to obtain smaller conditional MSE. Notice that the computational complexities of ML-BKF, MAP-BKF and IBR-BKF are similar to the BKF-DMU, which is $O(M 2^{2d})$ for discrete parameter space of size M.

4.1. Gene regulatory network model

The POBDS state model for gene regulatory networks can be written as:

$$\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}) \oplus \mathbf{u}_k \oplus \mathbf{n}_k, \quad k = 1, 2, \ldots \tag{43}$$

where $\mathbf{x}_k \in \{0, 1\}^d$ is a vector of gene expressions, $\mathbf{f} : \{0, 1\}^d \to \{0, 1\}^d$ is called the *network function*, "$\oplus$" indicates component-wise modulo-2 addition, $\mathbf{u}_k \in \{0, 1\}^d$ is a Boolean control input, and $\mathbf{n}_k \in \{0, 1\}^d$ is Boolean "transition noise".

The network function is expressed in component form as $\mathbf{f} = (f_1, \ldots f_d)$, where each component $f_i : \{0, 1\}^d \to \{0, 1\}$ is a Boolean function that predicts the next state of gene i, for $i = 1, \ldots d$., Here, we consider the specific model for the network function (Bahadorinejad, Imani, & Braga-Neto, 2018; Ghoreishi & Imani, 2019; Hajiramezanali, Imani, Braga-Neto, Qian, & Dougherty, 2019; Imani, Dehghannasiri, Braga-Neto and Dougherty, 2018; Imani, Ghoreishi, & Braga-Neto, 2018):

$$f_i(\mathbf{x}) = \begin{cases} 1, & \sum_{j=1}^{d} a_{ij} \mathbf{x}(j) + b_i > 0, \\ 0, & \sum_{j=1}^{d} a_{ij} \mathbf{x}(j) + b_i \le 0, \end{cases} \tag{44}$$

where a_{ij} and b_i are system parameters. This model is based on pathway diagrams typically used in molecular biology (Lau, Ganguli, & Tang, 2007). Parameter a_{ij} can take three values: if $a_{ij} = +1$, there is positive regulation (activation) from gene j to gene i; if $a_{ij} = -1$, there is negative regulation (inhibition) from gene j to gene i; whereas if $a_{ij} = 0$ then gene j is not an input to gene i. Parameter b_i can take two values: $b_i = +1/2$ if gene i is positively biased, in the sense that an equal number of activation and inhibition inputs will produce activation; the reverse being the case if $b_i = -1/2$.

Clearly, if $\mathbf{u}_k(i) = 1$ or $\mathbf{n}_k(i) = 1$, the next state of gene i will be flipped. For simplicity, we assume that the process noise $\mathbf{n}_k$ has independent components, distributed as Bernoulli(p). Parameter p indicates the amount of "perturbation" to the Boolean state process, and can be assumed to be in the interval $0 < p \le 0.5$ (the case $p > 0.5$ is equivalent to taking the complement of $\mathbf{f}$

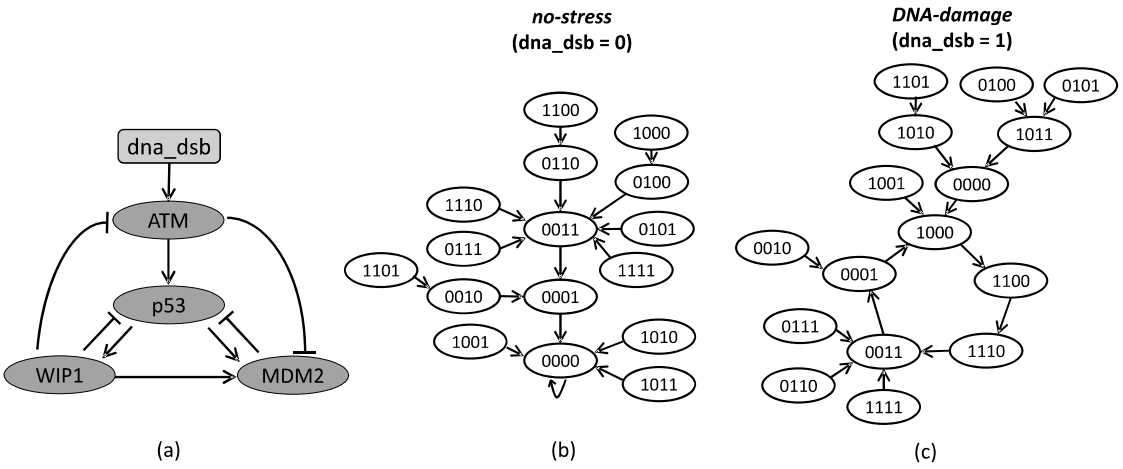

Fig. 3. Negative feedback-loop p53-MDM2 gene network. (a) Pathway diagram. State transition diagrams corresponding to (b) no-stress (dna_dsb = 0), and (c) DNA-damage (dna_dsb = 1).

with $p' = 1 - p < 0.5$). Larger values of the noise intensity p lead to more chaotic and unpredictable state transitions. On the other hand, values of p close to zero mean that the state trajectories are nearly deterministic.

Throughout this section, we assume the following Gaussian linear POBDS observation model

$$\mathbf{Y}_k = \boldsymbol{\mu} + D\mathbf{X}_k + \mathbf{v}_k, \quad k = 1, 2, \ldots \tag{45}$$

where $\mathbf{v}_k \sim N(0, \sigma^2 I)$ is an uncorrelated zero-mean Gaussian noise vector, $\boldsymbol{\mu}$ is a vector of baseline gene expressions (corresponding to the "zero" state for each gene) and D is a diagonal matrix containing differential expression values for each gene along the diagonal (these indicate by how much the "one" state of each gene is overexpressed over the "zero" state). Such a Gaussian linear model is an appropriate model for many important gene-expression measurement technologies, such as cDNA microarrays (Chen, Dougherty, & Bittner, 1997) and live cell imaging-based assays (Hua et al., 2012).

4.2. Experiments with discrete parameter spaces

The numerical experiments in this section evaluate the performance of the BKF-DMU and BKS-DMU developed in Section 3.2. The system model is assumed to be known except for some of the regulation parameters a_{ij}. This corresponds therefore to a gene network inference problem. Since each regulation parameter a_{ij} can take values in the set $\{-1, 0, 1\}$, the size of the parameter space is $M = 3^L$, where L is the number of unknown regulations.

We use the symmetric Dirichlet distribution to define priors over the parameter space. Consider a vector $[w_1, \ldots, w_{3^L}] \sim$ Dirichlet(φ), where $\varphi > 0$ is the *concentration parameter*. By definition, $w_i \geq 0$, for $i = 1, \ldots 3^L$, and $\sum_{i=1}^{3^L} w_i = 1$. The prior probability of the true model is assigned as:

$$p_0(\boldsymbol{\theta}^*) = w_l, \text{ where } l \sim \text{Cat}(w_1, \ldots, w_{3^L}) \tag{46}$$

where Cat($w_1, \ldots, w_{3^L}$) is the "categorical distribution"; this means that $p_0(\boldsymbol{\theta}^*) = w_l$ with probability w_i, for $i = 1, \ldots 3^L$. The remaining weights are randomly assigned to other values of the parameter as their prior probability. The concentration parameter φ controls how peaked the priors are around the true parameter value. our aim being to average over different sets of priors, each set with peakedness determined by φ. The larger φ is, the less peaked the prior distribution is. In the limit $\varphi \to \infty$, the prior distribution becomes uniform.

Some of the experiments below use a variation of this setup, where the weights $\{w_1, \ldots, w_{3^L}\}$ are randomly assigned to the various parameters, in order to model the case of a misspecified prior.

4.2.1. Experiments using the p53-Mdm2 network

The *p53* gene codes for the tumor suppressor protein p53 in humans, and its activation plays a critical role in cellular responses to various stress signals that might cause genome instability and thus cancer (Batchelor, Loewer, & Lahav, 2009). We use a model for the p53-MDM2 negative feedback-loop gene regulatory network, which consists of four genes: *ATM*, *p53*, *Wip1*, and *MDM2*, and the input "dna_dsb", which indicates the presence of DNA damage (double strand breaks). The pathway diagram of the network is presented in Fig. 3(a). Normal arrows represent activating regulations and blunt arrows represent suppressive regulations. Letting the state vector to be $\mathbf{X}_k = (ATM, p53, Wip1, MDM2)$, the gene interaction parameters a_{ij} can be read off Fig. 3 (a):

$$\begin{aligned} &a_{11} = 0, \quad a_{12} = 0, \quad a_{13} = -1, \quad a_{14} = 0 \\ &a_{21} = +1, \quad a_{22} = 0, \quad a_{23} = -1, \quad a_{24} = -1 \\ &a_{31} = 0, \quad a_{32} = +1, \quad a_{33} = 0, \quad a_{34} = 0 \\ &a_{41} = -1, \quad a_{42} = +1, \quad a_{43} = +1, \quad a_{44} = 0 \end{aligned}$$

The input vector is $\mathbf{u}_k = (\text{dna\_dsb}, 0, 0, 0)$, and can take one of its possible two values: DNA damage, $\mathbf{u}_k = (1, 0, 0, 0)$, or no stress, $\mathbf{u}_k = (0, 0, 0, 0)$, for $k = 1, 2, \ldots$ Under the assumption of negative regulation biases, $b_i = -1/2$, for $i = 1, \ldots d$, two state transition diagrams shown in Fig. 3 (b–c), are obtained for the system. We can see that under no-stress, "0000" is a singleton attractor state, while the other states are transient. On the other hand, under DNA damage, there is a cyclic attractor, corresponding to an oscillation of p53, along with the other proteins in its regulatory pathway. These behaviors match the known biological properties of the p53-MDM2 network (Weinberg, 2006).

The following parameter settings are used in our simulation: $p = 0.01$, $D = 20I$, $\sigma^2 = 10$, $b_i = -1/2$, $\boldsymbol{\mu}(i) = 30$, for $i = 1, 2, 3, 4$. The initial state distribution is assumed to be uniform: $P(\mathbf{X}_0 = \mathbf{x}^i) = 1/16$, for $i = 1, \ldots 16$. This leads to the following initial joint distribution:

$$\boldsymbol{\pi}_{0|0}^{\boldsymbol{\theta}_m}(i) = P(\mathbf{X}_k = \mathbf{x}^i, \boldsymbol{\theta}_m) = \frac{1}{16} \times p_0(\boldsymbol{\theta}_m), \tag{47}$$

for $i = 1, \ldots 16$, and $m = 1, \ldots 3^L$. All results are averaged over 1000 different prior distributions $p_0(\boldsymbol{\theta})$ obtained as described previously.

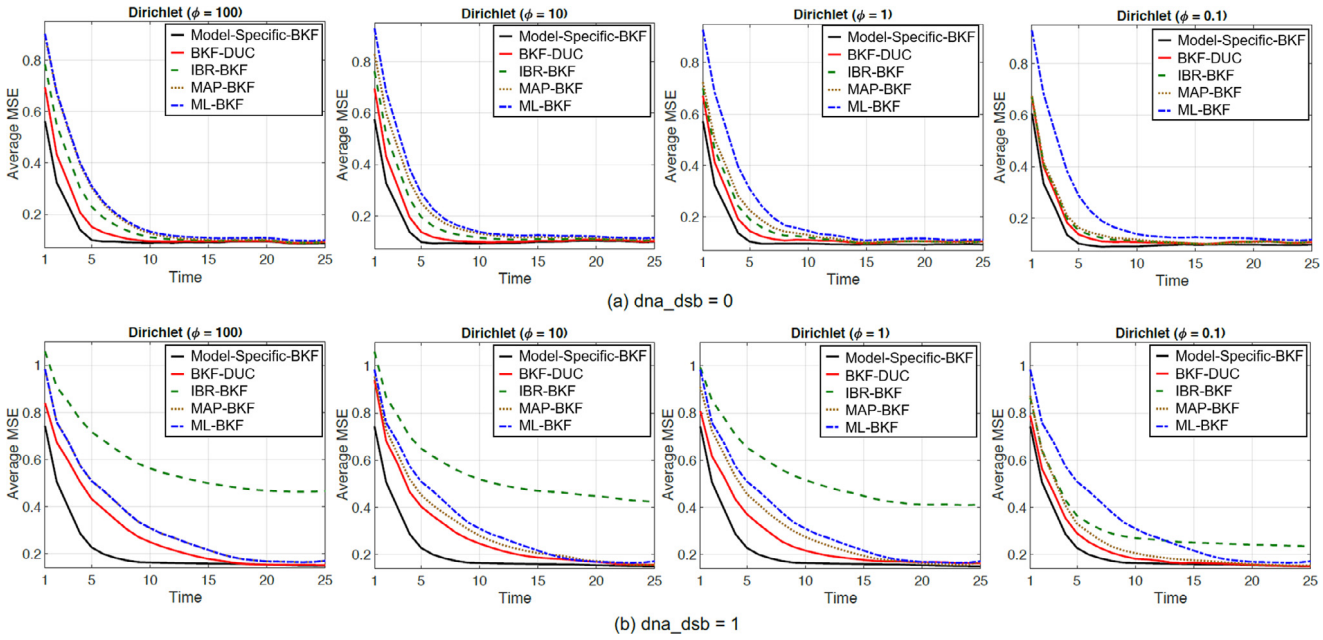

Fig. 4. Average MSE of different filters for the p53-Mdm2 network with (a) dna_dsb = 0 (no stress), and (b) dna_dsb = 1 (DNA damage).

In the first experiment, a_{13}, a_{23} and a_{24} are assumed to be unknown, so that the number of candidate models is $M = 3^3 = 27$. Fig. 4 displays the average MSE obtained over time by the baseline model-specific BKF, BKF-DMU, ML-BKF, MAP-BKF, and IBR-BKF. Moving from left to right in the figure, φ gets smaller so that the prior distributions become more peaked around the true parameter values.

As expected, the model-specific filter has the smallest average MSE. The BKF-DMU consistently outperforms the ML-BKF, MAP-BKF and IBR-BKF.When the prior distribution is close to uniform (i.e., for large φ), the ML-BKF and MAP-BKF display essentially the same performance, as expected (since their formulation is identical for a uniform prior), while for more peaked distributions, the MAP-BKF significantly outperforms the ML-BKF. This is due to the fact that the MAP-BKF can take advantage of the information in the peaked prior, which the ML-BKF cannot. Notice that the performance of the BKF-DMU approaches the baseline performance of the model-specific-BKF as the prior distribution becomes more peaked — this is particularly true in the case of DNA-damage.

All filters tend to display larger average MSE in the case of DNA damage. The reason for this is that under the no-stress condition, the system spends a significant amount of time in the rest state "0000", whereas under DNA damage, more states are visited due to the cyclic attractor, and the state estimation problem is more challenging.

One can observe that the IBR-BKF performs poorly for less peaked priors, and it is essentially useless in the case of DNA-damage, unless the prior is highly peaked. The reason for this behavior is that, in contrast to the BKF-DMU, the distribution over the parameter space does not get updated as more data are observed.

The parameter estimation error rate over time for the BKF-DMU, ML-BKF, and IBR-BKF is presented in Fig. 5 for two settings of the concentration parameter. The error rate is computed as the percentage of times the value of the parameter obtained by the adaptive filter differs from the true parameter over the 1000 repetitions of the experiment. The MAP-BKF is omitted since it is equivalent to the BKF-DMU for parameter estimation purposes. The IBR results are based entirely on the prior and so are constant

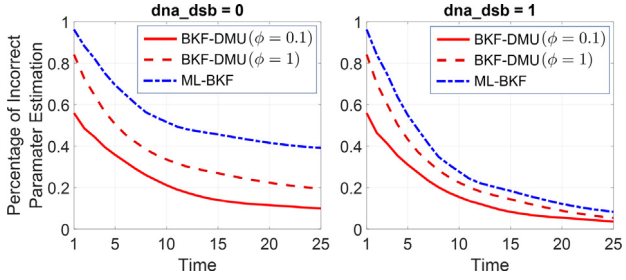

Fig. 5. Parameter estimation error rates for different filters for the p53-Mdm2 network with (a) dna_dsb = 0 (no stress), and (b) dna_dsb = 1 (DNA damage).

over time. Notice that the BKF-DMU displays smaller error rates than the other estimators at all points in time. Smaller error rates are obtained in the DNA-damage condition, as the cyclic attractor allows visiting more states and helps the estimation process. Error rates for both the BKF-DMU and IBR-BKF are smaller for more peaked priors, as expected.

An interesting fact becomes apparent at this point. As mentioned in connection with Fig. 5, parameter estimation performance is worse in the no-stress condition than under DNA-damage. However, in Fig. 4 we saw that state estimation performance is much better in the no-stress condition. The reason for this is that in the no-stress condition, most of the candidate models have similar dynamics and the same attractor "0000", whereas in the DNA-damage condition, perturbation of the model leads to different cyclic and singleton attractors, which means that the state estimation problem is more challenging under DNA-damage, despite the fact that the parameter estimation problem is easier in this case.

Next, we consider the performance of the smoothers. The system is assumed to be in the DNA-damage condition, and four regulation parameters are assumed to be unknown: a_{13}, a_{21}, a_{23}, and a_{32}. Two different values for φ are considered, corresponding to a peaked and a non-peaked prior. The results for both filters and smoothers are displayed in Fig. 6. The horizon for the smoothers were set at $k = 20$, so that corresponding smoothers and filters have the same performance at time $k = 20$. The results

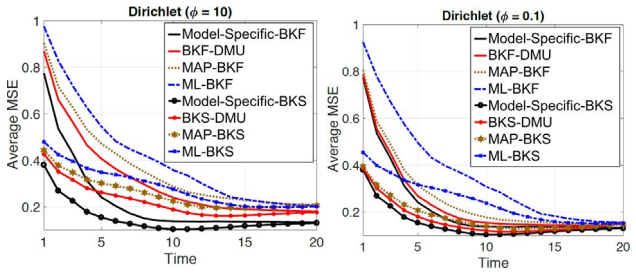

Fig. 6. Average MSE of different filters and smoothers for the p53-Mdm2 network with dna_dsb = 1 (DNA damage).

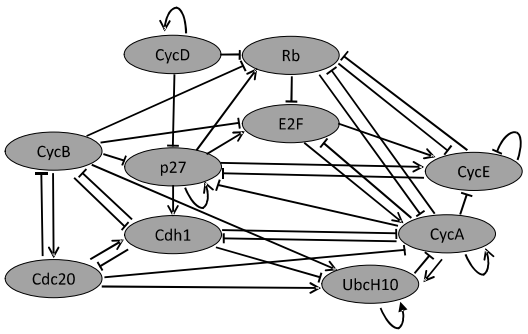

Fig. 8. Pathway diagram for the mammalian cell-cycle network.

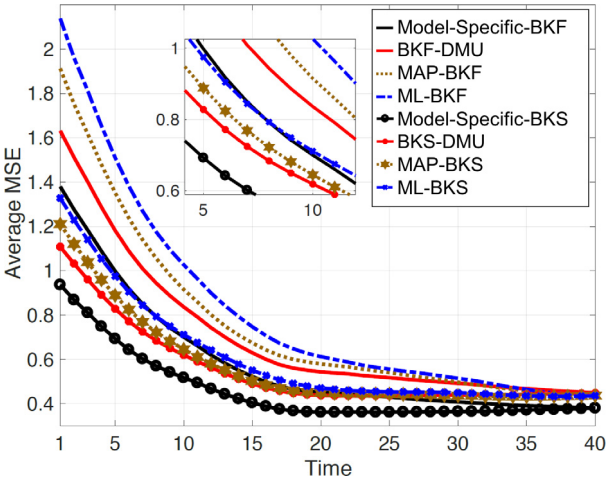

Fig. 9. Average MSE of different filters and smoothers for the mammalian cell-cycle network.

for the IBR-BKF and IBR-BKS are not shown since they were significantly worse than those of the others estimators. One can observe a significant reduction in average MSE for all smoothers in comparison to their corresponding filters, given the fact that the smoothers use more data than the filters. The BKF-DMU and BKS-DMU achieve the smallest average MSE in comparison to the other estimators. Furthermore, for the more peaked prior, the performances of the MAP-BKF, MAP-BKS, BKF-DMU and BKS-DMU all become closer to the baseline performance of the model-specific estimators.

The effect of a misspecified prior distribution is examined next. The unknown regulation parameters are assumed to be a_{13}, a_{23} and a_{24}. Here, we assume that the Dirichlet probabilities $[w_1, \ldots, w_{32}]$ are randomly assigned to different parameters, creating misspecification. Fig. 7 displays the average MSE of various filters for four distinct misspecified priors with different values for φ. We can see that the more peaked the prior is, the worse the effect of misspecification of the prior on estimator performance becomes, as expected. In fact, in the case $\varphi = 0.1$, corresponding to a very peaked prior, the average MSE of the BKF-DMU and the MAP-BKF have not converged to the baseline after 25 time steps. The BKF-DMU still performs better than the other filters, except in the case of a very peaked misspecified prior, in which case its performance is worse than that of the ML-BKF, which is immune to prior misspecification. Hence, the ML-BKF is the conservative choice if there are questions about the accuracy of prior modeling.

4.2.2. *Experiments using the mammalian cell cycle network*

Here we use a model for the well-known mammalian cell-cycle network (Fauré, Naldi, Chaouiya, & Thieffry, 2006), which is a relatively densely connected network of 10 genes, as displayed in Fig. 8. The state vector is $\mathbf{x}_k = (\text{CycD, Rb, p27, E2F, CycE, CycA, Cdc20, Cdh1, UbcH10, CycB})$. The gene interaction parameters a_{ij} can be read off Fig. 8. For example, Rb is activated by p27, and is inactivated by CycD, CycE, CycA, and CycB. These interactions can be expressed in terms of interaction parameters as $a_{21} = -1, a_{22} = 0, a_{23} = +1, a_{24} = 0, a_{25} = -1, a_{26} = -1, a_{27} = 0, a_{28} = 0, a_{29} = 0$ and $a_{210} = -1$. In all numerical experiments in this section, we assume that $p = 0.01$, $D = 20I$, $\sigma^2 = 15$, $b_i = -1/2$, $\boldsymbol{\mu}(i) = 30$, for $i = 1, \ldots, 10$. The initial state distribution is assumed to be uniform: $P(\mathbf{x}_0 = \mathbf{x}_i) = 1/2^{10}$, for $i = 1, \ldots, 2^{10}$.

We assume that the regulation parameters a_{23} and a_{42} are unknown, and all other parameters are known. Fig. 9 displays the average MSE for all filters and smoothers. The BKF-DMU and BKS-DMU achieved better results than all ML and MAP estimators. In addition, the performance of all estimators converges to the baseline as more data are available.

The effect of the noise parameters was examined in this section. Table 1 displays the average MSE per step for filters and smoothers over 1000 time series. The priors are non-peaked ($\varphi = 10$). One can observe the increase in average MSE of all estimators as process and measurement noise intensities increase and the estimation problem becomes more challenging. One can see that the BKF-DMU and BKS-DMU achieve better results than

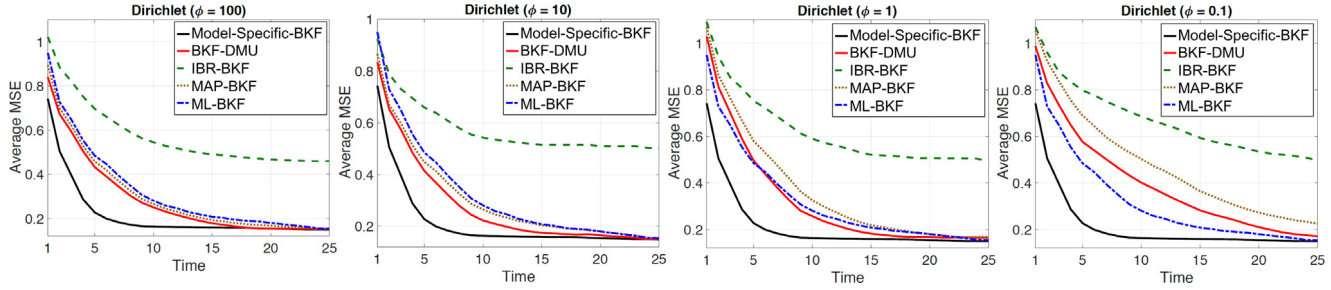

Fig. 7. Average MSE of different filters under misspecified priors for the p53-Mdm2 network with dna_dsb = 1 (DNA damage).

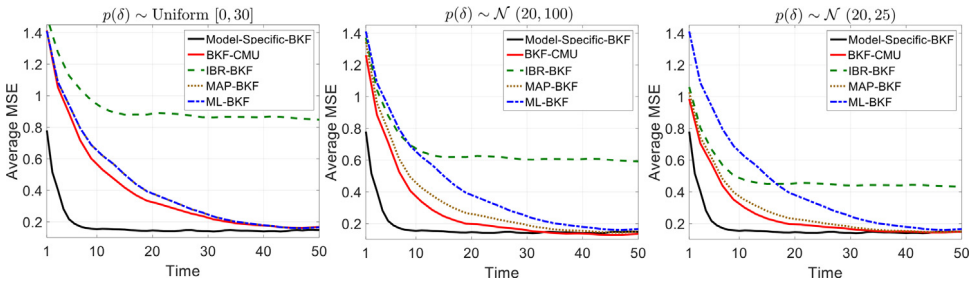

Fig. 10. Average MSE of different filters for the p53-MDM2 network with dna_dsb = 1 (DNA damage).

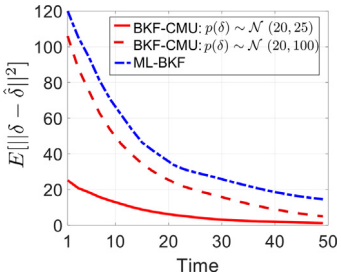

Fig. 11. The average MSE in estimation of continuous parameter δ for the p53-Mdm2 network with dna dsb = 1.

all other estimators. In addition, as expected, smoothers perform significantly better than the corresponding filters.

4.3. Experiments with continuous parameter spaces

In this section, we employ the p53-Mdm2 Boolean network to assess the performance of the BKF-CMU and BKS-CMU estimators developed in Section 3.3. All gene connections are assumed to be known, and so are all parameters of the observational model in (45) except for the matrix $D = \boldsymbol{\delta}$. The positive real-valued parameter δ must thus be estimated. In the simulation, we set $\delta = 20$. Three prior distributions are considered for δ: Uniform ($[0, 30]$), Gaussian N (20, 100) (less peaked), and Gaussian N (20, 25) (more peaked). The MCMC parameters are set to $N_{\text{Burn}} = 500$ and $N_{\text{MCMC}} = 2000$.

Fig. 10 displays the average MSE obtained over time by the baseline model-specific BKF, MMMC-BKF, ML-BKF, MAP-BKF, and IBR-BKF. Among the different prior distributions, the best results are obtained, unsurprisingly, for the peaked Gaussian prior. Among all filters, the BKF-CMU displays the smallest average MSE. As expected, the ML-BKF and MAP-BKF perform similarly in the case of a uniform prior. In addition, the IBR-BKF performs better in the presence of a peaked prior, but its average MSE does not converge to the baseline, since it does not update the prior as data are observed.

The parameter estimation error rate over time for the BKF-CMU and ML-BKF is presented in Fig. 11 for two prior distributions. The MAP-BKF is omitted since it is equivalent to the BKF-CMU for parameter estimation purposes. Notice that the BKF-DMU displays smaller error rates than ML-BKF at all time points. Meanwhile, the error rate of BKF-CMU is lower for the tighter prior distribution.

The effect of a poorly centered prior distribution is examined in Fig. 12. Three prior distributions are considered for parameter δ: N (15, 25), N (10, 25), and N (5, 25). We can see that the further the prior from the true value, the worse the effect on estimator performance, as expected. The BKF-CMU still performs better than the other filters, except in the case of a very highly misspecified prior, in which case its performance is worse than that of the ML-BKF, which is immune to prior mis-centered prior. In fact, in the cases of N (5, 25) and N (10, 25), the average MSE of the BKF-CMU and the MAP-BKF have not converged to the baseline after 50 time steps. Hence, the ML-BKF is the conservative choice if there are questions about the accuracy of prior modeling.

5. Conclusion

This paper introduces an optimal Bayesian framework for joint state and parameter estimation for a class of partially-observed Boolean dynamical systems (POBDS) under model uncertainty. The proposed framework provides the optimal expected MSE solution relative to the posterior distribution over the parameter space. For a discrete (finite) parameter space, we introduce exact optimal filter and smoother algorithms, called the BKF-DMU and BKS-DMU respectively. These two estimators can be seen as direct extensions of the ordinary BKF and BKS for POBDS with complete information. For continuous parameter spaces, we proposed an approximate solution based on MCMC. A comprehensive set of numerical experiments using gene regulatory network models demonstrate the superior performance of the proposed estimators over a number of other approaches. Future work includes developing efficient estimators for systems with large numbers of unknown parameters.

Acknowledgment

The authors acknowledge the support of the National Science Foundation, USA, through NSF award CCF-1718924.

Appendix

The following theorem gives the solution to the minimization problem in (21).

Theorem 1. *The optimal Bayesian state estimator* $\hat{\mathbf{X}}_{r|k}^{\text{OBE}}$ *is:*

$$\hat{\mathbf{X}}_{r|k}^{\text{OBE}} = E_{\boldsymbol{\theta}|\mathbf{Y}_{1:k}}[E[\mathbf{X}_r \mid \mathbf{Y}_{1:k}, \boldsymbol{\theta}]]. \tag{48}$$

with optimal conditional MSE

$$C_{r|k}^{OBE} = E_{\boldsymbol{\theta}|\mathbf{Y}_{1:k}}\left[C_{\boldsymbol{\theta}}(\mathbf{X}_r, \hat{\mathbf{X}}_{r|k}^{\text{OBE}})\right] = \frac{d}{2} - \sum_{i=1}^{d} \left| E[\mathbf{X}_r(i) \mid \mathbf{Y}_{1:k}] - \frac{1}{2} \right|. \tag{49}$$

Proof. Given the sequence of observations $\mathbf{Y}_{1:k}$, we seek a Boolean estimator $\hat{\mathbf{X}}_{r|k}$ of the state $\mathbf{X}_r$ by solving the minimization

Table 1
Average MSE per step of different filters and smoothers.

p	σ^2	Filters					Smoothers				
		Model-Specific	BKF-DMU	ML	MAP	IBR	Model-Specific	BKS-DMU	ML	MAP	IBR
0.01	15	0.5843	0.6934	0.8118	0.7593	0.8975	0.4518	0.5495	0.6045	0.5675	0.8039
	20	0.9875	1.0916	1.2139	1.1546	1.2962	0.8561	0.9509	1.086	0.9669	1.2043
0.05	15	0.7797	0.8986	1.0875	0.9501	1.0948	0.6552	0.7504	0.8105	0.7675	1.0039
	20	1.1850	1.3001	1.4088	1.3550	1.4917	1.0577	1.1510	1.2101	1.1666	1.4104

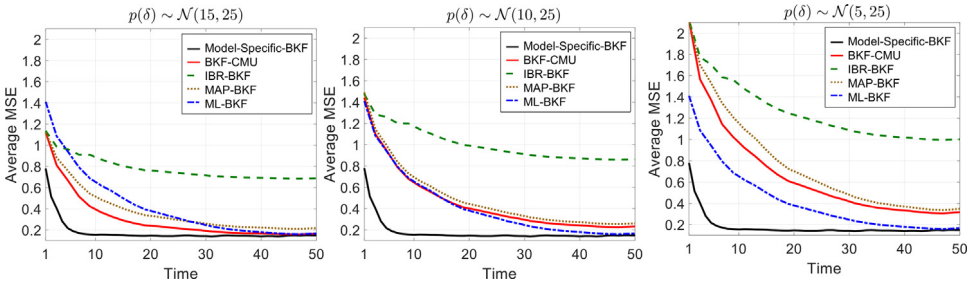

Fig. 12. Average MSE of different filters under misspecified priors for the p53-Mdm2 network with dna dsb = 1 (DNA damage).

in (21). This minimization can be expanded as:

$$\begin{aligned}\hat{\mathbf{X}}_{r|k}^{\mathrm{OBE}} &= \operatorname*{argmin}_{\hat{\mathbf{x}}_{r|k}\in\Psi} E_{\boldsymbol{\theta}|\mathbf{Y}_{1:k}}\left[C_{\boldsymbol{\theta}}\left(\mathbf{X}_r, \hat{\mathbf{X}}_{r|k}\right)\right] \\ &= \arg\min_{\hat{\mathbf{x}}_{r|k}\in\Psi} E_{\boldsymbol{\theta}|\mathbf{Y}_{1:k}}\left[E\left[\|\mathbf{X}_r - \hat{\mathbf{X}}_{r|k}\|^2 \mid \boldsymbol{\theta}\right]\right] \\ &= \arg\min_{\hat{\mathbf{x}}_{r|k}\in\Psi} E_{\boldsymbol{\theta}|\mathbf{Y}_{1:k}}\left[E\left[\|\mathbf{X}_r - \hat{\mathbf{X}}_{r|k}\|_1 \mid \boldsymbol{\theta}\right]\right] \\ &= \operatorname*{argmin}_{\hat{\mathbf{x}}_{r|k}\in\Psi} \sum_{i=1}^{d} E_{\boldsymbol{\theta}|\mathbf{Y}_{1:k}}\left[E\left[|\mathbf{X}_r(i) - \hat{\mathbf{X}}_{r|k}(i)| \mid \boldsymbol{\theta}\right]\right].\end{aligned} \tag{50}$$

Eq. (50) exploits the fact that $\|\mathbf{v}\|^2 = \|\mathbf{v}\|_1 = \sum_{i=1}^{d} \mathbf{v}(i)$ for a Boolean vector $\mathbf{v}$.

The minimization in (50) can be achieved by choosing $\hat{\mathbf{X}}_{r|k}(i)$ that minimizes $E_{\boldsymbol{\theta}|\mathbf{Y}_{1:k}}[E[|\mathbf{X}_r(i) - \hat{\mathbf{X}}_{r|k}(i)| \mid \boldsymbol{\theta}]]$ for each $i = 1, \ldots, d$. But it is easy to see that, since the state variables are Boolean, the minimizer is given by

$$\begin{aligned}\hat{\mathbf{X}}_{r|k}^{\mathrm{OBE}}(i) &= \begin{cases} 1, & \text{if } E_{\boldsymbol{\theta}|\mathbf{Y}_{1:k}}[E[\mathbf{X}_r(i) \mid \mathbf{Y}_{1:k}, \boldsymbol{\theta}] > 1/2, \\ 0, & \text{otherwise,} \end{cases} \\ &= \overline{E_{\boldsymbol{\theta}|\mathbf{Y}_{1:k}}[E[\mathbf{X}_r(i) \mid \mathbf{Y}_{1:k}, \boldsymbol{\theta}]}},\end{aligned}$$

for $i = 1, \ldots, d$., In other words,

$$\hat{\mathbf{X}}_{r|k}^{\mathrm{OBE}} = \overline{E_{\boldsymbol{\theta}|\mathbf{Y}_{1:k}}[E[\mathbf{X}_r \mid \mathbf{Y}_{1:k}, \boldsymbol{\theta}]]}\ . \tag{51}$$

The optimal conditional MSE achieved by $\hat{\mathbf{X}}_{r|k}^{\mathrm{OBE}}$ is computed as follows:

$$\begin{aligned}C_{r|k}^{\mathrm{OBE}} &= E_{\boldsymbol{\theta}|\mathbf{Y}_{1:k}}\left[C_{\boldsymbol{\theta}}(\mathbf{X}_r, \hat{\mathbf{X}}_{r|k}^{\mathrm{OBE}})\right] \\ &= \sum_{i=1}^{d} E_{\boldsymbol{\theta}|\mathbf{Y}_{1:k}}\left[E\left[|\hat{\mathbf{X}}_{r|k}^{\mathrm{OBE}}(i) - \mathbf{X}_r(i)| \mid \boldsymbol{\theta}\right]\right] \\ &= \sum_{i=1}^{d} P\left(\hat{\mathbf{X}}_{r|k}^{\mathrm{OBE}}(i) \neq \mathbf{X}_r(i) \mid \mathbf{Y}_{1:k}\right).\end{aligned} \tag{52}$$

The ith element in summation in the last line of Eq. (52) can be computed as:

$$\begin{aligned}&P\left(\hat{\mathbf{X}}_{r|k}^{\mathrm{OBE}}(i) \neq \mathbf{X}_r(i) \mid \mathbf{Y}_{1:k}\right) \\ &= \begin{cases} 1 - E_{\boldsymbol{\theta}|\mathbf{Y}_{1:k}}[E[\mathbf{X}_r(i) \mid \mathbf{Y}_{1:k}, \boldsymbol{\theta}]\,, \\ \quad \text{if } E_{\boldsymbol{\theta}|\mathbf{Y}_{1:k}}[E[\mathbf{X}_r(i) \mid \mathbf{Y}_{1:k}, \boldsymbol{\theta}] > 1/2\,, \\ E_{\boldsymbol{\theta}|\mathbf{Y}_{1:k}}[E[\mathbf{X}_r(i) \mid \mathbf{Y}_{1:k}, \boldsymbol{\theta}]\,, \quad \text{otherwise.} \end{cases}\end{aligned} \tag{53}$$

Now, replacing (53) into (52) leads to

$$\begin{aligned}C_{r|k}^{\mathrm{OBE}} &= \sum_{i=1}^{d} P\left(\hat{\mathbf{X}}_{r|k}^{\mathrm{OBE}}(i) \neq \mathbf{X}_r(i) \mid \mathbf{Y}_{1:k}\right) \\ &= \frac{d}{2} - \sum_{i=1}^{d} \left| E_{\boldsymbol{\theta}|\mathbf{Y}_{1:k}}[E[\mathbf{X}_r(i) \mid \mathbf{Y}_{1:k}, \boldsymbol{\theta}] - \frac{1}{2} \right|. \quad \square\end{aligned}$$

References

Andrieu, C., Doucet, A., & Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *72*(3), 269–342.

Bahadorinejad, A., Imani, M., & Braga-Neto, U. (2018). Adaptive particle filtering for fault detection in partially-observed Boolean dynamical systems. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, in press.

Batchelor, E., Loewer, A., & Lahav, G. (2009). The ups and downs of p53: understanding protein dynamics in single cells. *Nature Reviews Cancer*, *9*(5), 371–377.

Braga-Neto, U. (2011). Optimal state estimation for Boolean dynamical systems. In *Signals, systems and computers (ASILOMAR), 2011 conference record of the forty fifth asilomar conference on* (pp. 1050–1054). IEEE.

Brooks, S., Gelman, A., Jones, G., & Meng, X.-L. (2011). *Handbook of Markov chain Monte Carlo*. CRC press.

Chen, Y., Dougherty, E. R., & Bittner, M. L. (1997). Ratio-based decisions and the quantitative analysis of cDNA microarray images. *Journal of Biomedical Optics*, *2*(4), 364–374.

Crisan, D., Miguez, J., et al. (2018). Nested particle filters for online parameter estimation in discrete-time state-space Markov models. *Bernoulli*, *24*(4A), 3039–3086.

Dalton, L. A., & Dougherty, E. R. (2013a). Optimal classifiers with minimum expected error within a Bayesian framework — Part I: Discrete and Gaussian models. *Pattern Recognition*, *46*(5), 1301–1314.

Dalton, L. A., & Dougherty, E. R. (2013b). Optimal classifiers with minimum expected error within a Bayesian framework—Part II: properties and performance analysis. *Pattern Recognition*, *46*(5), 1288–1300.

Dalton, L. A., & Dougherty, E. R. (2014). Intrinsically optimal Bayesian robust filtering. *IEEE Transactions on Signal Processing*, *62*(3), 657–670.

Dehghannasiri, R., & Dougherty, E. R. (2018). Bayesian Robust Kalman smoother in the presence of unknown noise statistics. *EURASIP Journal on Advances in Signal Processing*, *2018*(55).

Dehghannasiri, R., Esfahani, M. S., & Dougherty, E. R. (2017). Intrinsically Bayesian robust Kalman filter: An innovation process approach. *IEEE Transactions on Signal Processing*, *65*(10), 2531–2546.

Dehghannasiri, R., Esfahani, M. S., Qian, X., & Dougherty, E. R. (2018). Optimal Bayesian Kalman filtering with prior update. *IEEE Transactions on Signal Processing*, *66*(8).

DeJong, D. N., Liesenfeld, R., Moura, G. V., Richard, J.-F., & Dharmarajan, H. (2012). Efficient likelihood evaluation of state-space representations. *Review of Economic Studies*, *80*(2), 538–567.

Doucet, A., De Freitas, N., & Gordon, N. (2001). *Sequential Monte Carlo methods in practice*. Springer-Verlag.

Doucet, A., Godsill, S., & Andrieu, C. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, *10*(3), 197–208.

Doucet, A., Gordon, N. J., & Krishnamurthy, V. (2001). Particle filters for state estimation of jump Markov linear systems. *IEEE Transactions on Signal Processing*, *49*(3), 613–624.

Dougherty, E. R. (2018). *Optimal signal processing under uncertainty*. SPIE Press.

Fauré, A., Naldi, A., Chaouiya, C., & Thieffry, D. (2006). Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle. *Bioinformatics*, *22*(14), e124–e131.

Ghoreishi, S. F. (2019). *Bayesian Optimization in Multi-Information Source and Large-Scale Systems* (Ph.D. thesis), College Station, TX: Texas A&M University, PhD thesis.

Ghoreishi, S., & Allaire, D. (2017). Adaptive uncertainty propagation for coupled multidisciplinary systems. *American Institute of Aeronautics and Astronautics*, 3940–3950.

Ghoreishi, S. F., & Allaire, D. (2019). Multi-information source constrained Bayesian optimization. *Structural and Multidisciplinary Optimization*, *59*(3), 977–991.

Ghoreishi, S. F., Friedman, S., & Allaire, D. L. (2019). Adaptive dimensionality reduction for fast sequential optimization with Gaussian processes. *Journal of Mechanical Design*, *141*(7), 071404.

Ghoreishi, S. F., & Imani, M. (2019). Offline fault detection in gene regulatory networks using next-generation sequencing data. In *2019 53rd asilomar conference on signals, systems and computers*. IEEE, in press.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. (1995). *Markov chain Monte Carlo in practice*. Chapman and Hall/CRC.

Hajiramezanali, E., Imani, M., Braga-Neto, U., Qian, X., & Dougherty, E. R. (2019). Scalable optimal Bayesian classification of single-cell trajectories under regulatory model uncertainty. *BMC Genomics*, *20*(6), 435.

Hastings, W. K. (1970). Monte Carlo Sampling methods using Markov chains and their applications. *Biometrika*, *57*(1), 97–109.

Hua, J., Sima, C., Cypert, M., Gooden, G. C., Shack, S., Alla, L., et al. (2012). Dynamical analysis of drug efficacy and mechanism of action using GFP reporters. *Journal of Biological Systems*, *20*(04), 403–422.

Imani, M. (2019). *Estimation, inference and learning in nonlinear state-space models* (Ph.D. thesis), College Station, TX: Texas A&M University, PhD thesis.

Imani, M., & Braga-Neto, U. (2015a). Optimal gene regulatory network inference using the Boolean Kalman filter and multiple model adaptive estimation. In *2015 49th asilomar conference on signals, systems and computers* (pp. 423–427). IEEE.

Imani, M., & Braga-Neto, U. (2015b). Optimal state estimation for Boolean dynamical systems using a Boolean Kalman smoother. In *2015 IEEE global conference on signal and information processing (GlobalSIP)* (pp. 972–976). IEEE.

Imani, M., & Braga-Neto, U. (2016a). Point-based value iteration for partially-observed Boolean dynamical systems with finite observation space. In *Decision and control (CDC), 2016 IEEE 55th conference on* (pp. 4208–4213). IEEE.

Imani, M., & Braga-Neto, U. (2016b). State-feedback control of partially-observed Boolean dynamical systems using RNA-seq time series data. In *American control conference (ACC), 2016* (pp. 227–232). IEEE.

Imani, M., & Braga-Neto, U. M. (2017a). Maximum-likelihood adaptive filter for partially observed Boolean dynamical systems. *IEEE Transactions on Signal Processing*, *65*(2), 359–371.

Imani, M., & Braga-Neto, U. (2017b). Multiple model adaptive controller for partially-observed Boolean dynamical systems. In *Proceedings of the 2017 American control conference (ACC 2017), Seattle, WA* (pp. 1103–1108). IEEE.

Imani, M., & Braga-Neto, U. (2017c). Optimal finite-horizon sensor selection for Boolean Kalman filter. In *2017 51th asilomar conference on signals, systems and computers* (pp. 1481–1485). IEEE.

Imani, M., & Braga-Neto, U. M. (2018a). Control of gene regulatory networks with noisy measurements and uncertain inputs. *IEEE Transactions on Control of Network Systems*, *5*(2), 760–769.

Imani, M., & Braga-Neto, U. M. (2018b). Finite-horizon LQR controller for partially-observed Boolean dynamical systems. *Automatica*, *95*, 172–179.

Imani, M., & Braga-Neto, U. (2018c). Gene regulatory network state estimation from arbitrary correlated measurements. *EURASIP Journal on Advances in Signal Processing*, *2018*(1), 22.

Imani, M., & Braga-Neto, U. (2018d). Optimal control of gene regulatory networks with unknown cost function. In *Proceedings of the 2018 American control conference (ACC 2018)* (pp. 3939–3944). IEEE.

Imani, M., & Braga-Neto, U. M. (2018e). Particle filters for partially-observed Boolean dynamical systems. *Automatica*, *87*, 238–250.

Imani, M., & Braga-Neto, U. (2019a). Control of gene regulatory networks using Bayesian inverse reinforcement learning. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *16*(4), 1250–1261.

Imani, M., & Braga-Neto, U. M. (2019b). Point-based methodology to monitor and control gene regulatory networks via noisy measurements. *IEEE Transactions on Control Systems Technology*, *27*, 1023–1035.

Imani, M., Dehghannasiri, R., Braga-Neto, U. M., & Dougherty, E. R. (2018). Sequential experimental design for optimal structural intervention in gene regulatory networks based on the mean objective cost of uncertainty. *Cancer Informatics*, *17*.

Imani, M., Ghoreishi, S. F., Allaire, D., & Braga-Neto, U. M. (2019). MFBO-SSM: Multi-fidelity Bayesian optimization for fast inference in state-space models, In: AAAI, Vol. 33 (pp. 7858–7865).

Imani, M., Ghoreishi, S. F., & Braga-Neto, U. M. (2018). Bayesian control of large MDPs with unknown dynamics in data-poor environments, In: *Advances in neural information processing systems*, (pp. 8156–8166).

Ionides, E. L., Bretó, C., & King, A. A. (2006). Inference for nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, *103*(49), 18438–18443.

Jazwinski, A. H. (1970). *Stochastic processes and filtering theory, Vol. 64*. Academic Press.

Johansen, A. M., Doucet, A., & Davy, M. (2008). Particle methods for maximum likelihood estimation in latent variable models. *Statistics and Computing*, *18*(1), 47–57.

Julier, S. J., Uhlmann, J. K., & Durrant-Whyte, H. F. (1995). A new approach for filtering nonlinear systems. In *American control conference, proceedings of the 1995, Vol. 3* (pp. 1628–1632). IEEE.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, *82*(1), 35–45.

Kantas, N., Doucet, A., Singh, S. S., Maciejowski, J., Chopin, N., et al. (2015). On particle methods for parameter estimation in state-space models. *Statistical Science*, *30*(3), 328–351.

Kauffman, S. (1969). Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology*, *22*, 437–467.

Lau, K.-Y., Ganguli, S., & Tang, C. (2007). Function constrains network architecture and dynamics: A case study on the yeast cell cycle Boolean network. *Physical Review E*, *75*(5), 051907.

Lindsten, F., Jordan, M. I., & Schön, T. B. (2014). Particle Gibbs with ancestor sampling. *Journal of Machine Learning Research (JMLR)*, *15*(1), 2145–2184.

Magill, D. (1965). Optimal adaptive estimation of sampled stochastic processes. *IEEE Transactions on Automatic Control*, *10*(4), 434–439.

Malik, S., & Pitt, M. K. (2011). Particle filters for continuous likelihood evaluation and maximisation. *Journal of Econometrics*, *165*(2), 190–209.

Martin, J. J. (1967). *Bayesian decision problems and Markov chains*. Wiley.

Martino, L., Read, J., Elvira, V., & Louzada, F. (2017). Cooperative parallel particle filters for online model selection and applications to urban mobility. *Digital Signal Processing*, *60*, 172–185.

Maybeck, P. S. (1982). *Stochastic models, estimation, and control, Vol. 3*. Academic press.

McClenny, L. D., Imani, M., & Braga-Neto, U. M. (2017a). Boolean Kalman filter with correlated observation noise. In *Acoustics, speech and signal processing (ICASSP), 2017 IEEE international conference on* (pp. 866–870). IEEE.

McClenny, L. D., Imani, M., & Braga-Neto, U. M. (2017b). Boolfilter: an R package for estimation and identification of partially-observed Boolean dynamical systems. *BMC Bioinformatics*, *18*(1), 519.

McClenny, L. D., Imani, M., & Braga-Neto, U. (2017c). BoolFilter package vignette. *The Comprehensive R Archive Network (CRAN)*.

Messerschmitt, D. (1990). Synchronization in digital system design. *IEEE Journal on Selected Areas in Communications*, *8*(8), 1404–1419.

Mohsenizadeh, D., Dehghannasiri, R., & Dougherty, E. (2016). Optimal objective-based experimental design for uncertain dynamical gene networks with experimental error. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*.

Qian, X., & Dougherty, E. R. (2016). Bayesian Regression with network prior: Optimal Bayesian filtering perspective. *IEEE Transactions on Signal Processing*, *64*(23), 6243–6253.

Roli, A., Manfroni, M., Pinciroli, C., & Birattari, M. (2011). On the design of Boolean network robots. In *Applications of evolutionary computation* (pp. 43–52). Springer.

Schön, T. B., Wills, A., & Ninness, B. (2011). System identification of nonlinear state-space models. *Automatica*, *47*(1), 39–49.

Silver, E. A. (1963). Markovian decision processes with uncertain transition probabilities or rewards, tech. rep., Technical report, Defense Technical Information Center.

Urteaga, I., Bugallo, M. F., & Djurić, P. M. (2016). Sequential Monte Carlo methods under model uncertainty. In *Statistical signal processing workshop (SSP), 2016 IEEE* (pp. 1–5). IEEE.

Van Der Merwe, R. (2004). *Sigma-point Kalman filters for probabilistic inference in dynamic state-space models* (Ph.D. thesis), Oregon Health & Science University.

Weinberg, R. (2006). The biology of Cancer. *Garland Science*.

Whiteley, N., Andrieu, C., & Doucet, A. (2010). Efficient Bayesian inference for switching state-space models using discrete particle Markov chain Monte Carlo methods, arXiv preprint arXiv:1011.2437.

Wills, A., Schön, T. B., Ljung, L., & Ninness, B. (2013). Identification of Hammerstein–Wiener models. *Automatica*, *49*(1), 70–81.

Yoon, B.-J., Qian, X., & Dougherty, E. R. (2013). Quantifying the objective cost of uncertainty in complex dynamical systems. *IEEE Transactions on Signal Processing*, *61*(9), 2256–2266.

Mahdi Imani received his Ph.D. degree in Electrical and Computer Engineering from Texas A&M University, College Station, TX in 2019. He is an Assistant Professor in the Department of Electrical and Computer Engineering at George Washington University, Washington, DC, USA. His research interests include Machine Learning, Bayesian Statistics and Decision Theory, with a wide range of applications from computational biology to cyber–physical systems. He is the recipient of the Association of Former Students Distinguished Graduate Student Award for Excellence in Research-Doctoral in 2019, the Best PhD Student Award in ECE department at Texas A&M University in 2015, and a single finalist nominee of ECE department for the Outstanding Graduate Student Award in college of engineering at Texas A&M University in 2018. He is also recipient of the best paper finalist award from the 49th Asilomar Conference on Signals, Systems, and Computers, 2015.

Edward R. Dougherty received the M.Sc. degree in computer science from Stevens Institute of Technology, the Ph.D. degree in mathematics from Rutgers University, and has been awarded the Doctor Honoris Causa by the Tampere University of Technology, Finland. He is a Distinguished Professor in the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, USA, where he holds the Robert M. Kennedy '26 Chair in electrical engineering and is Scientific Director of the Center for Bioinformatics and Genomic Systems Engineering. He is the author of 16 books, the editor of 5 others, and author of more than 300 journal papers. Dr. Dougherty is a Fellow of SPIE, has received the SPIE President's Award, and served as the Editor of the SPIE/IS&T Journal of Electronic Imaging. At Texas A&M University, he received the Association of Former Students Distinguished Achievement Award in Research, and was named Fellow of the Texas Engineering Experiment Station, and Halliburton Professor of the Dwight Look College of Engineering.

Ulisses M. Braga-Neto received the Ph.D. degree in electrical and computer engineering from The Johns Hopkins University, Baltimore, MD, USA. He is an Associate Professor in the Department of Electrical and Computer Engineering and a member of the Center for Bioinformatics and Genomic Systems Engineering, Texas A&M University, College Station, TX, USA. He has held postdoctoral positions at the University of Texas M.D. Anderson Cancer Center, Houston, TX, and at the Oswaldo Cruz Foundation, Recife, Brazil. His research interests include pattern recognition and statistical signal processing. He is the author of the textbook Error Estimation for Pattern Recognition (IEEE-Wiley, 2015) and has received the NSF CAREER Award for his work in this area.