18th International Conference on Information Fusion
Washington, DC - July 6-9, 2015

On the Use and Misuse of Bayesian filters

Tiancheng Li, Javier Prieto, Juan M. Corchado
BISITE group, Faculty of Science
University of Salamanca
37007, Salamanca, Spain
{t.c.li, javierp, corchado}@usal.es

Javier Bajo
Department of Artificial Intelligence
Technical University of Madrid
28660, Madrid, Spain
jbajo@fi.upm.es

Abstract - Since the groundbreaking work of the Kalman filter in the 1960s, considerable effort has been devoted to discrete time filters for dynamic state estimation, especially including a variety of suboptimal implementations of the Bayesian filter. The essence of the Bayesian filter is to make the (sub)optimum fusion of the observation information in time sequence based on the hidden Markov model of the state process. While admitting the success of filters in many cases, this study investigates the cases when they in fact loose to the deterministic observation-only (O_2) inference that infers the estimate by using the observation information only without modeling the state dynamics. Special attention has been paid to quantitatively analyzing when and why the Bayesian filter will underperform the O_2 inference from the information fusion perspective. Classic state space models have shown that the O_2 inference can perform better (in terms of both accuracy and computing speed) than filters in certain cases. Therefore attention is desired for the use of a filter when the model is not guaranteed to be accurate and much approximation is used.

Keywords: Bayesian statistical inference, Bayesian filter, Kalman filter, particle filter, observation-only inference.

1 Introduction

Dynamic state estimation has been a long-standing research topic concerned with the sequential process of estimating the state(s) evolving over time that is/are periodically observed by sensors. A general solution that has been most investigated in literature is based on fusing the observations with models, which assume the system as a hidden Markov model (HMM) and then a Bayesian filter can be employed. Based on the state transition model, the Bayesian posterior estimate at time $t-1$ can be propagated (generating a prior estimate) and fused with the newest observation received at time t, generating the Bayesian posterior estimate for time t, which will be further updated when new observations arrive. This online prediction-correction recursion forms the basis of Bayesian filters and has dominated the field since the groundbreaking work of the Kalman filter [1] in the 1960s.

The optimal recursive state estimator in the Bayesian sense requires the complete posterior density of the state to be determined as a function of time, which only admit closed-form solutions in the linear and Gaussian system (namely the Kalman filter) except a few special cases [2]. Therefore, considerable effort has been devoted to various discrete time filters to deal with nonlinearity models and/or non-Gaussian noise, namely suboptimal Bayesian filters. This has been accompanied by the rapid development of approximation theories and technologies and computers, which enables complicated computations. Indeed, optimal and suboptimal Bayesian filters have been demonstrated powerful and successful in many cases.

These filters perform well as long as the models are assumed accurately with few disturbances/outliers and that the approximation (if used) is insignificant. Ideally, the posterior Cramér-Rao lower bound [3, 4] can be reached if the physical world and the model simulated coincide exactly. However, this is rarely the case in real world. In general, accurate knowledge of the state dynamics model (and noise) that can be time varying (e.g. abrupt motion) and unpredictable is often missing. In this case, one has to approximate or estimate the model and noise before the use of a filter, which more or less differ from the real model and noise, leaving a difference we refer to as modeling error.

It has been well acknowledged that modeling errors (and significant disturbances/outliers) can easily cause the failure of filters, see e.g. [5-8]. Therefore, dealing with model disparities has been a fundamental problem. Within the filter framework, many strategies have been proposed to deal with the modeling error and system disturbances including "adaptive" (see e.g. [9-11]), "robust" (see e.g. [12-14]) and "direct" [8] filtering and detection and treatment of outlier [15], etc. Similar issues occur in Bayesian smoothers as well as some optimization based estimators; see [16, 17]. The situation will be much more complicated in multiple-target cases in cluttered environments, see e.g. [17-19]. We do not intend to detail these in this paper.

Despite these sophisticated Bayesian filters, it is crucial to know whether they are still effective when significant modeling errors (including disturbances) occur or when too much approximation has been used. It has been observed that simple deterministic algorithms outperform the particle filter in a type of finite-state estimation in digital communications [20], even given that the filter is properly set up. It will be shown in this paper that the so-called observation-only (O_2) inference, that infers the observed part of the state directly from the noisy observations, can significantly outperform many filters in certain cases. Of the minimal computational complexity, the performance of the O_2 inference has identified a benchmark for assessing the effectiveness of filters: If a filter cannot outperform the O_2

inference (given that it is applicable), it simply shall not be used. In other words, a filter shall only be applied for a particular estimation problem when it outperforms the O_2 inference (if applicable) on average in accuracy under the same conditions.

Simply, one filter can be better than another or some others; it does not mean, however, that the best solution for a particular problem must be a filter or a smoother as long as the comparison has not included the O_2 inference. In this paper, two primary contributions have been made.

1) The O_2 inference is addressed and is established as a benchmark to assess the effectiveness of Bayesian filters. In addition, a Monte Carlo (MC) unbiased transformation approach is proposed for realizing O_2 inference for nonlinear observation models.
2) The effectiveness of the optimal Bayesian filter is investigated from the information fusion (IF) perspective and is evaluated on a classic model. Both theoretical studies and simulation results show that, the O_2 inference can easily outperform the filters in certain cases, more than expected.

An extended and complete version of this study will be given in [21]. In the following sections, the Bayesian and O_2 inference are presented in Section 2. Section 3 investigates the effectiveness of the optimal Bayesian filter under typical Gaussian distributions quantitatively. Section 4 revisits a classic state space model (SSM) to show the misuse of many filters in certain cases. We conclude in Section 5.

2 Bayesian filtering and O_2 inference

2.1 Bayesian statistical inference

The dynamic state estimation, also referred to as filtering, is often modeled as a HMM where the system being modeled is assumed to be a Markov process of unobserved state. The problem is generally formulated as a SSM

$$\boldsymbol{x}_t = \boldsymbol{f}_t(\boldsymbol{x}_{t-1}, \boldsymbol{u}_t) \tag{1}$$

$$\boldsymbol{y}_t = \boldsymbol{h}_t(\boldsymbol{x}_t, \boldsymbol{v}_t) \tag{2}$$

where t indicates the time instant, $\boldsymbol{x}_t$ denotes the state vector, $\boldsymbol{y}_t$ denotes the observation (also called measurement) vector, and $\boldsymbol{u}_t$ and $\boldsymbol{v}_t$ denote the noise affecting the state transition equation $\boldsymbol{f}_t(\cdot)$ and the observation equation $\boldsymbol{h}_t(\cdot)$ respectively. In particular, the state transition equation is a difference equation for the discrete time while for the continuous time it is a differential equation.

In the framework of the Bayesian statistical inference, the Bayesian posterior distribution $p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t})$ given all the historical observations $\boldsymbol{y}_{1:t} = \boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_t$ solves the filtering problem, which basically consists of predicting and correcting two steps. The predicting step combines the previous filtering distribution $p(\boldsymbol{x}_{t-1}|\boldsymbol{y}_{1:t-1})$ with the state transition $p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{y}_{1:t-1})$ as

$$p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1}) = \int p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{y}_{1:t-1}) p(\boldsymbol{x}_{t-1}|\boldsymbol{y}_{1:t-1})\, d\boldsymbol{x}_{t-1} \tag{3}$$

This forms a prior probability distribution (often called simply the prior). To note here, such predictions assume that the transforms are predictable, which is not always the case [16, 19]. Given a new observation $\boldsymbol{y}_t$, the prior distribution will be updated by Bayes' rule as follows

$$p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t}) = \frac{p(\boldsymbol{y}_t|\boldsymbol{x}_t) p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1})}{p(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1})} \tag{4}$$

where $p(\boldsymbol{y}_t|\boldsymbol{x}_t)$ is the likelihood. This gives the Bayesian posterior distribution (often called simply the posterior).

The Kalman filter, which is the closed form solution to the linear system with additive Gaussian noise (a special case of HMM), can be presented as one of the simplest dynamic Bayesian networks. For general nonlinear systems with non-Gaussian noise, approximations have to be made which can be parametric or non-parametric. In the former, the Kalman filter and its approximate extensions e.g. [2, 12, 14, 22] calculate estimates of the true values of states (in terms of both Gaussian mean and variance) recursively over time, while in the latter the particle filter e.g. [9, 15, 20, 23] and the point mass filter e.g. [24] calculate the probability density function (PDF) of the state recursively over time. All these filters need to assume a Markov process model (i.e. the state transition model $\boldsymbol{f}_t(\cdot)$) as well as the system noise $\boldsymbol{u}_t$ and $\boldsymbol{v}_t$, which are very critical for the accuracy of the filter.

However, the performance of all of these modeling-based filters will depend greatly on the matching between the physical world and the model (and parameters) assumed. The importance of the model to the performance of the filter cannot be overestimated. We will quantitatively show in Section 3 that a bad prediction (whether because of modeling error or too much approximation) will be worse than useless. Arguably, the more assumption and approximation there are, the more unreliable the filter is. In contrast, an estimator that makes fewer model assumptions will be of relatively better quality to achieve the desired results in real world as supposed. The following section presents an estimator that only requires the observation function $\boldsymbol{h}_t(\cdot)$ that is arguably the minimum requirement for the utilization of the observations, regardless the state transition $\boldsymbol{f}_t(\cdot)$, and system noise $\boldsymbol{u}_t$ and even $\boldsymbol{v}_t$.

2.2 Observation-only inference

Given the observation function $\boldsymbol{h}_t(\cdot)$, a straightforward way to estimate the state is to infer it directly from the observation(s) namely the observation-only (O_2) inference, which is independent of the state transition process. The O_2 inference can be conceptually written as follows (as long as the observation function is invertible):

$$\hat{\boldsymbol{\chi}}_t = \boldsymbol{h}_t^{-1}(\boldsymbol{y}_t, \boldsymbol{v}_t) \tag{5}$$

where $\boldsymbol{h}_t^{-1}$ is the "generalized" inverse function of $\boldsymbol{h}_t$ in the real coordinate system, $\hat{\boldsymbol{\chi}}_t$ is the O_2 inference of "the observed part of" the state $\boldsymbol{x}_t$. In a fully observed system, $\hat{\boldsymbol{\chi}}_t$ is a full-dimensional state estimate.

Obviously, the O_2 inference yields a deterministic accuracy that is only consistent with the quality of the observation as will be demonstrated in our simulation. This non-Bayesian solution is different to Bayesian statistical inference but the result is equivalent to that of the KF for the linear observation function when the observation error variance goes into zero and to that of the maximum likelihood estimation for the linear observation function with additive noise of the symmetric probability distribution.

Given the observation noise $\boldsymbol{v}_t$, it is straightforward to calculate the mean and variance of the O_2 inference for linear inversing functions as the distribution will maintain the same form after linear transformation while for nonlinear inversing function, we propose a Monte Carlo nonlinear transformation methods in section 2.2.1.

It is necessary to point out that a more general situation would include an unknown observation function $\boldsymbol{h}_t(\cdot)$, which is an indispensable requirement for the utilization of the observations in all kinds of estimators and therefore has to be identified before filtering. Here we do not include this issue for avoiding distraction from the main contribution of this paper. Nor is it our intention to comprehensively cover all the cases of inversing calculation of an arbitrary function, which is a fundamental mathematical task. We point out here that $\boldsymbol{h}_t(\cdot)$ is generally given in a few simple forms in real life problems. For example, in the tracking content, it, whether for radar, sonar or camera, is readily suitable for the O_2 inference especially when multiple sensors are used. Regarding the variety of realistic problems, the O_2 inference may still be inapplicable. In particular, there are several cases that need special treatments for calculating (5) as follows.

A Inversing bias/error

The inversing will introduce biases (i.e. the expectation of the estimate is not equal to the true state) if $\boldsymbol{h}_t(\cdot)$ is nonlinear or if the noise has non-zero expectation, where the state-dependent bias highly depends on both the noise, the true state and the nonlinearity. Simply, a nonlinear conversion of a Gaussian distribution is no more Gaussian and therefore the situation can be very complicated. This has been recognized when converting polar/spherical measurements to Cartesian coordinates for the use of filters, see e.g. [25]. To a degree, the converting bias can be removed explicitly for simple inversing function and noise (e.g. Gaussian). Significantly different to filters, the O_2 inference does not assume the observation noise and therefore works with unknown and even time-varying observation noise. Hence, we omit this bias when the noise $\boldsymbol{v}_t$ is unknown by setting it to be zero. Eq. (5) is then reduced to

$$\hat{\boldsymbol{\chi}}_t = \boldsymbol{h}_t^{-1}(\boldsymbol{y}_t, \boldsymbol{0}) \tag{6}$$

This calculation is directly based on deterministic values.

If the observation noise is known, we propose to use a Monte Carlo debiasing method to remove the inversing bias as follows. This is different to the algebraic approximate methods given in [25] and the references therein and is model-free and computational easier. The idea is sampling a group of samples from the noise distribution $\boldsymbol{v}_t^{(i)} \sim \boldsymbol{v}_t, i = 1,2,\ldots,I$ and use them as noise separately in the inversing calculation of (5) as

$$\hat{\boldsymbol{\chi}}_t^{(i)} = \boldsymbol{h}_t^{-1}\left(\boldsymbol{y}_t, \boldsymbol{v}_t^{(i)}\right), i = 1,2,\ldots I \tag{7}$$

Based on this, we can easily calculate the mean and covariance of the O_2 inference, respectively as follows

$$\hat{\boldsymbol{\chi}}_t = \frac{1}{I}\sum_{i=1}^{I}\hat{\boldsymbol{\chi}}_t^{(i)} \tag{8}$$

$$\mathrm{Cov}(\hat{\boldsymbol{\chi}}_t) = \frac{1}{I-1}\sum_{i=1}^{I}\left(\hat{\boldsymbol{\chi}}_t^{(i)} - \hat{\boldsymbol{\chi}}_t\right)\left(\hat{\boldsymbol{\chi}}_t^{(i)} - \hat{\boldsymbol{\chi}}_t\right)^T \tag{9}$$

The covariance of the estimate is helpful for multi-sensor data fusion when multiple sensors are available [26, 27]. Obviously, this Monte Carlo debiasing is regardless the type of noise and the observation function. To reduce the number of samples used, deterministic sampling e.g. [28] might be used to replace the random sampling.

B Irreversibility

One of the primary challenges for the application of the O_2 inference is the irreversibility of the observation function, for which the inversing calculation is not directly applicable. This can be viewed as an underdetermined system. Under-determination occurs when the total dimensions of the observations are smaller than the total freedoms of the state that need to be estimated. In contrast, over-determination occurs when the total dimensions of the observations are more than the total freedoms of the state that need to be estimated. The over-determined system (e.g. multiple sensors are used) is beneficial, as it will provide a more accurate estimate [26, 27]. However, it is challenging to infer the state of an under-determined system, for which the observations are just too few to properly determine the state.

The under-determination is the same challenging for the filters. In practice, it should be avoided to design/use an under-determined observation system. A general solution that is worktable in practice is to improve the observability of the system by adding more sensors to make the system properly or even over determined. For an over-determined system, the O_2 inference shall use each minimum but adequate subgroup of observations to infer the estimate and finally fuse all estimates in an optimal way, where different sources of observations shall be treated equally. That is, for state $\boldsymbol{x}_t$, we seek a set of observations (e.g. corresponding to n sensors) as $\boldsymbol{y}_{i,t} = \boldsymbol{h}_{i,t}\left(\boldsymbol{x}_t, \boldsymbol{v}_{i,t}\right), i = 1,2,\ldots,n$. Denoting the covariance of $\boldsymbol{v}_{i,t}$ as $\boldsymbol{R}_{i,t}$, the nonlinear (weighted) least square estimation is given conceptually such as

$$\hat{\boldsymbol{\chi}}_t = \mathrm{argmin}_{\chi_t}\sum_{i=1}^{n}\frac{\left(\chi_t - \boldsymbol{h}_{i,t}^{-1}(y_t, v_{i,t})\right)\left(\chi_t - \boldsymbol{h}_t^{-1}(y_t, v_{i,t})\right)^T}{\boldsymbol{R}_{i,t}} \tag{10}$$

For the case of multiple or massive sensor O_2 inference and further discussions, please refer to [26, 27].

As a common irreversible case, the observation function is a non-monotonic function and its inversing calculation involves a sign problem (the sign can be viewed as an additional freedom of the state. In cases when the state is bounded in a positive or negative space only, the sign problem is not involved; otherwise it needs to be separately determined. There are two ways to determine the sign of the estimate: the first is to estimate based on the state transition function and the previous estimate, i.e. we have the sign function $\mathrm{sgn}(\hat{\boldsymbol{x}}_t) = \mathrm{sgn}(\boldsymbol{f}_t(\hat{\boldsymbol{x}}_{t-1}), 0)$. This can be taken as the default method. The second way is to use an additional estimator to estimate the sign, which is computationally more intensive. In both cases, the O_2 inference will not only utilize the observation information but also use the state transition information, which is no more strictly "observation only" and can be referred to as the O_2+ inference. Both method are demonstrated in [21]

The O_2 inference represents the probably most intuitive solution for state estimation. However, it is in fact involved directly in the triangulation, trilateration and multilateration positioning technologies based on angle of arrival, signal strength measurements and time difference of arrival techniques respectively, just to name a few. In addition, many visual tracking methods involves inferences from image data directly where no filter is used. Unfortunately, the O_2 inference has rarely been considered in the evaluation of filters/smoothers. However, we will show that the O_2 inference can easily outperform than filters.

3 Filtering: worth or not?

3.1 Probability of filter benefit

The predicting and correcting steps of the Bayesian filters correspond to two sources of information about the state: the prediction based on the previous estimate and the inference from the newest observation. The predicting step employs the state transition equation (that is assumed to be a Markov process) to infer an a priori estimate while the updating step uses the observation to correct the estimate a priori. Using the observation to update the prediction, obtaining the state a posteriori, which is the core art of the Bayesian filter. Therefore, whether the filter/fusion is beneficial depends on whether a better (closer to the true state) estimate will be obtained as compared with the estimate directly inferred from the observation. For which, we define a new metric called probability of filter benefit (PoFB) to refer the probability that the posteriori estimate is more accurate, statistically, than the estimate inferred from the observation.

For simplicity, both the estimate from the prediction x_p and the inference x_o from the observation are assumed to be subject to 1D Gaussian distribution, either biased or unbiased, with regard to the true state, i.e., $p(x_o) = \mathcal{N}(m_o, \delta_o^2)$, $p(x_p) = \mathcal{N}(m_p, \delta_p^2)$. The Bayesian filter fuses $p(x_o)$ and $p(x_p)$ to get a fused distribution $p(x_f) = \mathcal{N}(m_f, \delta_f^2)$ as an estimate of the posterior distribution of the true state x_T. The Kalman filter gives the optimal fusion of two Gaussian distributions according to the covariance in the sense of minimizing the square estimate error, obtaining

$$m_f = \frac{\delta_o^2 m_p + \delta_p^2 m_o}{\delta_o^2 + \delta_p^2} \tag{11}$$

$$\delta_f^2 = \frac{\delta_o^2 \delta_p^2}{\delta_o^2 + \delta_p^2} \tag{12}$$

We refer to this as *optimal fusion* under Gaussian conditions.

We will show that $x_f \sim p(x_f)$ might not be a better estimate than $x_o \sim p(x_o)$, although the variance of the estimate is smaller as $\delta_f^2 \le \min\{\delta_o^2, \delta_p^2\}$. It is when and only when $|x_T - x_o| > |x_T - x_f|$, that the estimate x_f is better than x_o. The probability of filter/fusion benefit is defined as

$$\begin{aligned}
&\mathrm{PoFB} = \mathrm{P}\left(|x_T - x_o| > |x_T - x_f|\right) \\
&= \mathrm{P}\left((x_T - x_o)^2 > (x_T - x_f)^2\right) \\
&= P\left((2x_T - x_f - x_o)(x_f - x_o) > 0\right) \\
&= P\left(x_o < x_f < (2x_T - x_o)\right) + P\left(2x_T - x_o < x_f < x_o\right)
\end{aligned} \tag{13}$$

It is known that the cumulative distribution function of the Gaussian distribution $p(x_f) = \mathcal{N}(m_f, \delta_f^2)$is

$$\Phi_f(x) = \frac{1}{\delta_f\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(t-m_f)^2/2\delta_f^2}\, dt \tag{14}$$

Therefore, Eq. (14) can be written in terms of expected values as

$$\mathrm{PoFB} = \int_{-\infty}^{x_T} \left(\Phi_f(2x_T - x) - \Phi_f(x)\right) p(x)dx + \int_{x_T}^{\infty} \left(\Phi_f(x) - \Phi_f(2x_T - x)\right) p(x)dx \tag{15}$$

where $p(x) = \frac{1}{\delta_o\sqrt{2\pi}} e^{-(x-m_o)^2/2\delta_o^2}$.

$PoFB \le 0.5$ just means the prediction/prior is useless or even harmful for the filter. In the general setting of the Bayesian filter, the observation is assumed to be unbiased. We define the variance ratio (VR) r, the ratio of the variances of two distributions, and the bias ratio (BR) p, the ratio of the prediction bias of $p(x_p)$ over the standard deviation of the observation $p(x_o)$, respectively as

$$r = \frac{\delta_p^2}{\delta_o^2} \tag{16}$$

$$p = \frac{m_p - m_o}{\delta_o} \tag{17}$$

The PoFB in this case is highly related to VR r and BR p. Due to the symmetry of the Gaussian distribution, we only consider the case of a positive BR $p \ge 0$ and the result holds the same for a negative BR. 100,000 random samples are generated separately from distributions as $x_o \sim p(x_o)$, $x_p \sim p(x_p)$ and $x_f \sim p(x_f)$ to calculate the PoFB for different VR $r \in [0.01, 1000]$ and BR $p \in [0,10]$. In particular, $p = 0$ means that two distributions are unbiased. The PoFB results are given in Fig.1. The results show that

1) The PoFB will tend to be stable with 0.5 when r goes to infinite. In particular, for $p \ge 2$, the larger VR r is, approximately the larger the PoFB is; for $p \le 0.4$, the larger VR is, the smaller the PoFB is; for $0.4 < p < 2$, the PoFB goes up and then reduces down to 0.5 with the increasing of VR r. This agrees with the fact that a larger r corresponds to a larger δ_p^2 of $p(x_p)$ which will have a smaller effect on the fusion distribution $p(x_f)$. In the information theory, it is uninformative prior. For a very large r, the effect can be omitted, after which we have $p(x_f) \approx p(x_o)$, and $x_f \sim p(x_f)$ vs $x_o \sim p(x_o)$ is then 50-50. This demonstrates that the O_2 inference is nothing else but equivalent to the KF when the variance ratio (the observation error variance divided by the prior estimate error variance) goes to zero.

2) When the bias $p \le 0.6$, PoFB > 0.5 i.e. the fusion has more than an approximately 50% possibility of obtaining a more accurate estimate than the original estimate. This indicates that when the bias of the biased distribution is not significant, the fusion will be acceptable and is still more

likely to benefit. This is the case (when the prior estimate is only slightly biased) whereby the filter is recommended.

3) When the bias $p \geq 0.8$ (and the VR $r \geq 0.1$ or a little larger), PoFB < 0.5 i.e. the fusion has less than an approximately 50% possibility of obtaining a more accurate estimate. This indicates, when the bias of the prediction is significant (whether because of large modeling/ approximation error or disturbances), the fusion will be more likely to obtain a worse result than simply inferring from the observation. This is the case whereby the O_2 inference but not the filter is recommended.

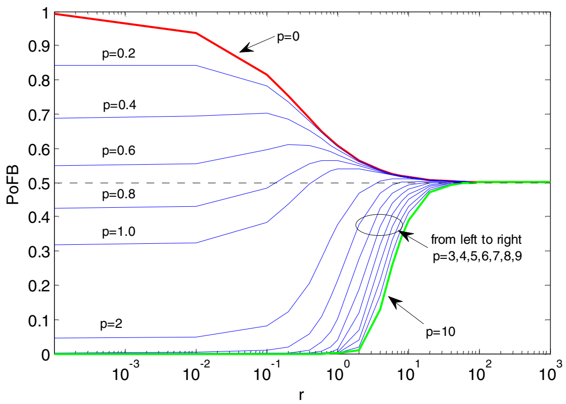

Fig.1 PoFB for different variance ratio r and bias ratio p

As shown in Fig.1, for $r \to 0$, the prediction $p(x_p)$ is relatively extremely accurate (with very small variance; but biased) and will dominate the fusion result fully, leaving us with $p(x_f) \approx p(x_p)$ then PoFB will almost fully depend on the bias of the prediction p: the smaller p is, the larger the PoFB is. However, in general the prediction of a filter that is affected by both the process noise and the observation noise cannot be so accurate (as compared with the observation).

The results also indicate that if the prediction is slightly biased or just unbiased, accurate prediction (of small variance) will be beneficial; otherwise it will be harmful for the filter (the filter is useless).

In a more general case, both the prediction and the observation can be biased. In the quantitative study, we use different true state m_T which is chosen by adjusting a scaling parameter m defined as

$$m = \frac{x_T - m_o}{\delta_o} \tag{18}$$

This parameter indicates the level of the bias of $p(x_o)$.

For $m = \{-10, -5, -2, -1, -0.1, 0.1, 1, 2, 5, 10, 30\}$, the results of (13) for the PoFB are plotted separately in Fig.2 which compares the observation-based inference $x_o \sim p(x_o)$ (the estimate obtained if no filter is employed) to the fusion $x_f \sim p(x_f)$.

The results show again that, all PoFBs will converge to 50% when r goes into infinite. Furthermore,

1) When $m \leq 0$ (i.e. $x_T \leq m_o \leq m_y$; the prediction bias is larger than that of the observation), all PoFBs will be smaller than 50% and the larger p, the smaller PoFB;

2) When $m \geq p$ (i.e. $m_o \leq m_p \leq x_T$; the prediction bias is smaller than that of the observation), all PoFBs will be larger than 50% and the larger p, the larger PoFB;

3) When $0 < m < p$ (i.e. $m_o < x_T < m_p$), the PoFB depends on r, m, p (see the sub-plots for $m = 1,2,5$): generally, with the increase of $r > 1$, the PoFB will go up over 0.5 and then decrease to 0.5 finally.

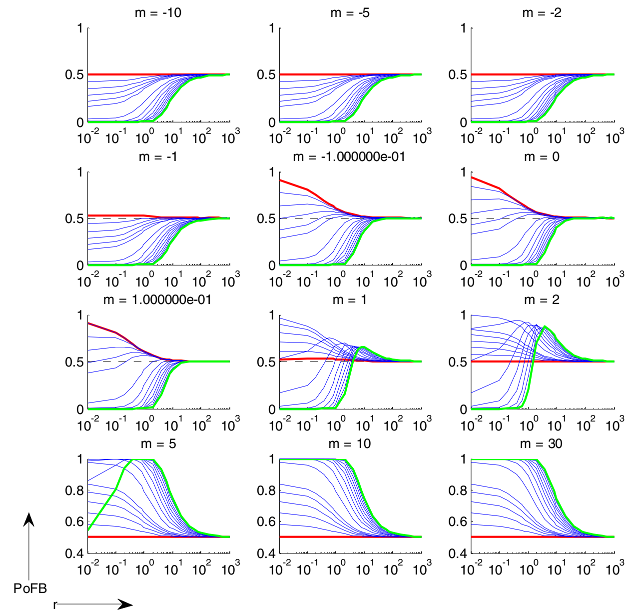

Fig. 2 PoFB for different scaling parameter m, variance ratio r and bias ratio p; the red line is for $p = 0$, the green line is for $p = 10$ while the blue lines are in between

3.2 Discussions

The above result has clearly demonstrated that prediction-observation fusion is not guaranteed to provide a benefit. It is only when 1) both the observation and the prediction are unbiased, 2) the bias of the prediction is very small while the observation is unbiased or 3) the bias of the observation is more serious than the prediction, that the fusion/filter is likely to get a more accurate estimate than the O_2 inference, namely being effective otherwise, the filter can easily be ineffective. To note, we have only considered the error on the mean of the estimate (bias) but not on the variance which is not used in the O_2 inference. If there is an error with the assumption of the variance, the performance of the filter will likely be worse (see e.g. [29]), which however does not matter the performance of the O_2 inference.

However, we must be aware that the fusion discussed so far maximally corresponds to one filter iteration, while in the time sequence the condition of the system varies. That is to say, r, p and m vary with time, which gives way to a situation in which at some stages when a filter is effective (the prediction obtained is good enough) while at other stages (the prediction is relatively poor) it is not. It is desirable albeit challenging to distinguish these in real-time so that an "optimal" decision is made so that the O_2 inference and filters work interactively. Here we may extend the results obtained under Gaussian filtering to a general conclusion

Remark 1 *Whether the filter is effective or not primarily depends on the quality of the prior, especially the bias of the prior (the lesser, the better); the extent of effectiveness will depend on the variance of the prior (as compared to the variance of the observation)*

It is clear that many issues can affect the quality of the prediction such as initialization errors, system disturbances, modeling errors. This, together with the approximation that has to be used in suboptimal Bayesian filters, can lead to large discrepancies between the prediction and the true states. More seriously, any bias once generated in the Markov process, whether due to mismodeling or approximation, will affect the subsequent steps as the posterior is based on all the history information. All of these indicate that the filter can easily be ineffective as compared with the O_2 inference in practice, more than expected; see our further discussion and demonstrations given in [21]. Therefore, it is fair to say a filter shall only be applied when it at minimum outperforms the O_2 inference on average in estimation accuracy. More precisely, we have several general principles as follows on the use or not of the Bayesian filter.

1) The O_2 inference is comparably more sensitive to the observation noise than filters: it enjoys small noise the same greatly as it suffers from bad noise. To note, if the observation noise is significant, neither the O_2 inference nor the filter can be good.
2) If the system is fully known that can be (close to) correctly modeled and a filter can be well initialized, affected with no or small disturbance, the filter will then work well as supposed.
3) If the state model cannot be correctly simulated (or the filter has to make great approximation) and there are relatively large disturbances from time to time, the filter will not work well; instead, it might be better to use the O_2 inference rather than a filter.
4) In case of multiple sensors, the more sensors, the better accuracy and reliability for the O_2 inference which can additionally filter clutter; see [26, 27].

We point out here that through 'fitting/smoothing' the results of the O_2 inference between successive scans by using the state transition information (if available), more accurate or further information about the state can be inferred. This forms a key part of our future work.

4 Simulations

In this simulation, we use another SSM that has also been widely employed for filter evaluation since first proposed in [23], with the state transition equation and the observation equation respectively given as follows

$$x_t = 1 + \sin(w\pi t) + \phi_1 x_{t-1} + u_t \quad (19)$$

$$y_t = \begin{cases} \phi_2 x_t^2 + v_t & t \le 30 \\ \phi_3 x_t - 2 + v_t & t > 30 \end{cases} \quad (20)$$

where x_t, y_t are respective the state and observation at time t, the scale parameters $\omega = 0.04$, $\phi_1 = 0.5$, $\phi_2 = 0.2$ and $\phi_3 = 0.5$, the process noise u_t is a Gamma $\mathcal{G}a(3,2)$ random variable and the observation noise is Gaussian $v_t \sim \mathcal{N}(0, R)$.

To carry out the O_2+ inference (default), inversing Eq. (20) after taking off the unknown noise item v_t, we have

$$\hat{x}_t = \begin{cases} \text{sgn}\sqrt{|y_t/\phi_2|} & t \le 30 \\ \frac{y_t+2}{\phi_3} & t > 30 \end{cases} \quad (21)$$

where sgn stands for $\text{sgn}(1 + \sin(w\pi t) + \phi_1 \hat{x}_{t-1})$. To note here, the state transition noise is useless.

If the observation noise v_t is available, the proposed MC debiasing strategy can be further applied for the nonlinear inversing calculation when $t \le 30$. That is, we have

$$\hat{x}_t = \text{sgn} \times \frac{1}{I}\sum_{i=1}^{I}\left(\sqrt{\left|\left(y_t - v_t^{(i)}\right)/\phi_2\right|}\right) \quad t \le 30 \quad (22)$$

where $I = 100$.

A series of filters are employed for comparison that include extended KF (EKF), Unscented KF (UKF) [28], the SIR PF, the PF that use EKF and UKF separately as the proposal [23]. We use 200 particles for the PF and we use the initial state variance as 0.75 for the EKF/UKF. The unscented transform parameter is set as $\alpha = 1, \beta = 0, \kappa = 2$ (the same as used in [23]). The true state and the initial unbiased estimate of all filters are all starting from $x_1 = 1$. Here, no initialization error is applied for all filters. Since UKF/EKF cannot be used directly for this Gamma noise, we assume equivalent variance 0.75 as alternative for them, i.e. they admit a modeling error/bias of 1.5 of the process noise, as $\mathcal{G}a(3,2)$ is of mean 1.5, variance 0.75. This corresponds to the real life situation where the processing noise is unknown and has not be estimated correctly when a filter is employed. But, the PFs use the correct model and parameters (via weighted particles for approximation). They represent respective realistic and ideal cases for filtering.

To capture the average result, 100 MC runs are performed with random re-initialization for each run, generating much large RMSE variance for the O_2 inference. Each run consists of 60 time-steps. We first set $R = 0.00001$. These are the default parameter settings in many publications. The true state and estimates given by different filters for one run are plotted in Fig.3. In particular, as shown in Fig.3, if the simulation is properly assumed that the state is known to be always positive, the sign for the estimate can just be set as positive and no separate sign estimation is needed. Then, the O_2 inference will computer faster. This, however, puts the O_2 inference in a favorable situation, otherwise the sign can cause significant problems [21].

The RMSE of different methods are plotted in Fig.4. The mean and variance of RMSE over time and the computing time of each method are given in Table 1. It shows that the O_2 method (whether biased or unbiased) has outperformed all the filters by several orders of magnitude in terms of both RMSE and computing speed, which indicates these filters are actually useless for this model for the specified parameters. The unbiased O_2 inference outperforms the biased O_2 inference slightly, indicating that the bias caused by the inversing calculation is insignificant, through really exists. According to our knowledge, such a good result exhibited by the O_2 inference has never been reported before (except when a huge number of particles are used for particle filtering, it can perform comparably), although many filters have been proposed to apply to this model.

Furthermore, for a range of different observation noise variances $R = [0.00001, 100]$, the average RMSE (mean) is given in Fig.5. It can be seen that (approximately): when $R < 0.04$, all these filters are ineffective; when $0.04 < R \leq 1$, PF is effective while others are not; when $2 < R < 40$, PF, UKF and EKF are effective while the EKPF and UKPF are not; when $40 < R$, all filters used become effective. The simulation results just indicate that these filters can be easily underperform to the simple O_2 inference on this particular model, regardless that the straightforward O_2 inference is much computationally faster than any filter. This not unique. More simulation results and discussions can be found in [21]. This is a critical fact that shall not be omitted but instead great cautions shall be paid to the effectiveness of filters: It is not always advisable to apply the Bayesian filter especially when the unknown system cannot be modelled correctly or much approximation has to be used.

Table 1 Performance of different filters and the O_2 inference

	RMSE		Computing time
	mean	variance	(s)
EKF	0.353	0.181	0.008
UKF	0.277	0.113	0.035
SIR(PF)	0.554	0.090	1.845
EKPF	0.353	0.188	3.793
UKPF	0.240	0.089	9.512
O_2 inference	0.005	1.085×10^{-5}	7.23×10^{-5}
Unbiased O_2 inference	0.005	1.083×10^{-5}	4.26×10^{-4}

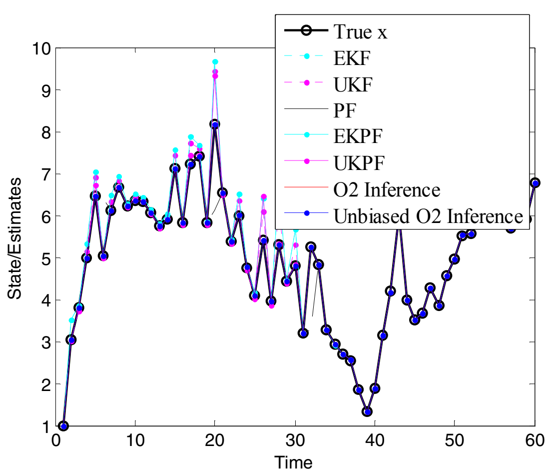

Fig.3 The true state and estimates of different estimators

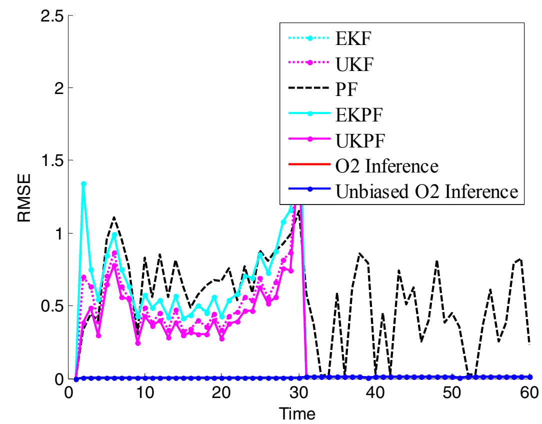

Fig.4 RMSE of different estimators of 100 MC runs

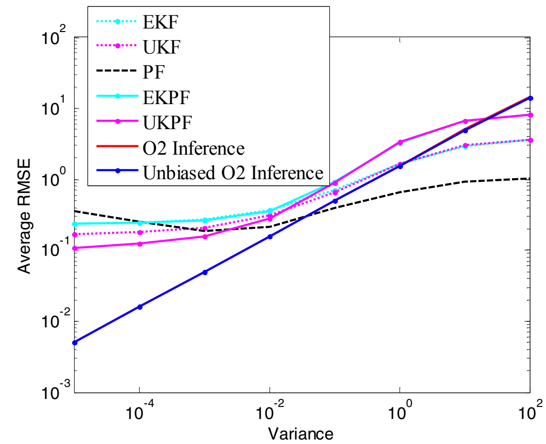

Fig.5 Average RMSE of different estimators of 60 steps×100 MC runs for different observation noise variances

5 Conclusion

This paper presents a non-Bayesian solution for dynamic state estimation, referred to as the observation-only (O_2) inference, which infers the state directly from the observations regardless of the state transition process. In addition, a Monte Carlo sampling-based debiasing approach is proposed for unbiased nonlinear O_2 inference.

Good filtering results require correct and accurate models and few system disturbances and little approximate otherwise the filter does not guarantee a benefit as compared to the O_2 inference and plus (O_2+). We have quantitatively investigated when and why the Bayesian filter does not give a more accurate estimation than the O_2 inference from the information fusion perspective, which is further demonstrated by the simulation on a typical filtering model. Arguably, the filter is not always preferable.

While the posterior CRLB provides a lower bound on the mean-square error of any "unbiased" estimator of the random parameter, the O_2 takes a more practical approach by setting a higher bound on the mean error of any "effective" estimator. Our future work includes extending the O_2 inference for state prediction and the multi-sensor O_2 inference for multi-target tracking in cluttered environment.

Acknowledgment

This work is partly supported by Ministry of Economy and Finance of Spain (No. TIN2012-36586-C03-03), FEDER funds.

References

[1] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," Journal of Basic Engineering, vol. 82, no.1, pp. 35–45, 1960.

[2] I. Arasaratnam and S. Haykin, "Cubature Kalman Filters," IEEE Transactions on automatic control, vol. 54, no. 6, pp. 1254-1269, 2009.

[3] P. Tichavský, C. H. Muravchik, and A. Nehorai, "Posterior Cramér-Rao bounds for discrete-time nonlinear filtering," IEEE Trans. Signal Process., vol. 46, no. 5, pp. 1386–1396, May 1998.

[4] C. Fritsche, E. Özkan, L. Svensson, and F. Gustafsson, A fresh look at Bayesian Cramér-Rao Bounds for discrete-time nonlinear filtering, FUSION 2014, Salamanca, Spain, July 7-10, 2014.

[5] H. Heffes, "The effect of erroneous models on the Kalman filter response," IEEE Trans. Autom. Control, vol. 11, no. 3, pp. 541–543, 1966.

[6] B. Friedland, "Treatment of bias in recursive filtering," IEEE Trans. Autom. Contr. 14, 359–367, 1969.

[7] P. M. Djuric, J. Miguez, "Assessment of nonlinear dynamic models by Kolmogorov–Smirnov statistics," IEEE Trans. Signal Processing, vol. 58, no.10, pp. 5069–5079, 2010.

[8] C. Novara, F. Ruiz, and M. Milanese, "Direct Filtering: A new approach to optimal filter design for nonlinear systems," IEEE Transactions on Automatic Control, vol. 58, no. 1, pp. 86-99, 2013.

[9] T. Li, S. Sun, T. P. Sattar, "Adapting sample size in particle filters through KLD-resampling," Electronics Letters, vol.46, no.12, pp.740-742, 2013.

[10] J. Prieto, S. Mazuelas, A. Bahillo, P. Fernandez, R. M. Lorenzo, E. J. Abril, "Adaptive data fusion for wireless localization in harsh environments," IEEE Trans. Signal Processing, vol. 60, no. 4, pp. 1585–1596, 2012.

[11] T. Li, S. Sun, J. M. Corchado and M. F. Siyau, Random finite set-based Bayesian filters using magnitude-adaptive target birth intensity, FUSION 2014, Salamanca, Spain, 7-10th July 2014.

[12] M. E. Çelebi, and L. Kurz, "Robust locally optimal filters: Kalman and Bayesian estimation theory," Information Sciences, vol. 92, no.1–4, pp. 1–32, 1996.

[13] S.O.R. Moheimani, A.V. Savkin, and I.R. Petersen, "Robust filtering, prediction, smoothing, and observability of uncertain systems," IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications, vol. 45, no. 4, pp. 446-457, 1998.

[14] W. Qi, P. Zhang, and Z. Deng, "Robust weighted fusion time-varying Kalman smoothers for multisensor system with uncertain noise variances," Information Sciences, vol.282 no. 20, pp. 15–37, 2014.

[15] C.S. Maíz, J. Míguez, E. M. Molanes-López, and P. M. Djurić, "A robustified particle filtering scheme for processing time series corrupted by outliers," IEEE Trans. on Signal Processing, vol. 60, no. 9, pp. 4611-4627, 2012.

[16] S. C. Patwardhan, S. Narasimhan, P. Jagadeesan, B. Gopaluni, S. L. Shah, "Nonlinear Bayesian state estimation: A review of recent developments," Control Engineering Practice, vol.20, no.10, pp. 933-953, 2012.

[17] Y. Bar-Shalom, P. K. Willett and X. Tian, Tracking and Data Fusion: a handbook of algorithms, YBS Publishing, 2011.

[18] R. Mahler, Advances in Statistical Multisource-Multitarget Information Fusion. Artech House, 2014.

[19] O. Cappe, E. Moulines, T. Ryden, Inference in Hidden Markov Models, 2nd ed. New York: Springer, 2007.

[20] E. Punskaya, A. Doucet, and W. J. Fitzgerald, On the use and misues of particle filtering in digital commnunications, In Proc. EUSIPCO, Sept. 2002.

[21] T. Li, J. M. Corchado, J. Bajo, S. Sun and J. F. Paz, "Effectiveness of Bayesian filters: An information fusion perspective," Information Sciences, 2015.

[22] P. Stano, Z. Lendek, J. Braaksma, R. Babuska, C. de Keizer, A.J. den Dekker, "Parametric Bayesian filters for nonlinear stochastic dynamical systems: a survey," IEEE Trans. Cybern, vol.43, no.6, pp. 1607-1624, 2013.

[23] R van der Merwe, A. Doucet, N. Freitas, and W. Eric. The unscented particle filter, England: Cambridge University, Engineering Department, 2000.

[24] M. Simandl, J. Kralovec, and T. Söderstörm, Advanced point-mass method for nonlinear state estimation, Automatica, vol. 42, no.7, pp.1133–1145, 2006.

[25] S. Bordonaro, P. Willett, and Y. Bar-Shalom, "Decorrelated unbiased converted measurement Kalman filter," IEEE Trans. on Aerospace and Electronic Systems, vol. 50, no.2, pp.1431-1444, 2014.

[26] T. Li, J. M. Corchado, J. Bajo and G. Chen, Multi-target detection and estimation with the use of massive independent, identical sensors, Proc. SPIE 9469, Sensors and Systems for Space Applications VIII, 94690G, Baltimore, Maryland, US, April 20-24, 2015.

[27] T. Li, J. M. Corchado, J. Bajo and S. Sun, "Clustering for filtering: Multi-object detection and estimation using multiple/massive sensors," IEEE J. Selected Topics in Signal Process. Submitted, 2015.

[28] S. Julier, J. Uhlmann, A consistent, debiased method for converting between polar and cartesian coordinate systems, In Proceedings of AeroSense: Acquisition, Tracking and Pointing XI, Orlando, FL, USA, Oct 1997, pp. 110–121.

[29] V. A. Bavdekar, A. P.Deshpande, S. C. Patwardhan, "Identification of process and measurement noise covariances for state and parameter estimation using extended Kalman filter," Journal of Process Control, vol. 21, 585–601, 2011.