Information Sciences 329 (2016) 670–689

Contents lists available at ScienceDirect

Information Sciences

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

Effectiveness of Bayesian filters: An information fusion perspective

Tiancheng Li [a,b,*], Juan M. Corchado [a,d], Javier Bajo [c], Shudong Sun [b], Juan F. De Paz [a]

[a] BISITE Research Group, Faculty of Science, University of Salamanca, Salamanca 37008, Spain
[b] School of Mechanical Engineering, Northwestern Polytechnical University, Xi'an 710072, China
[c] Department of Artificial Intelligence, Technical University of Madrid, Madrid 28660, Spain
[d] Osaka Institute of Technology, Asahi-ku Ohmiya, Osaka 535-8585, Japan

ARTICLE INFO

Article history:
Received 20 February 2015
Revised 14 August 2015
Accepted 19 September 2015
Available online 9 October 2015

Keywords:
Recursive estimation
Bayesian estimation
Kalman filter
Particle filter

ABSTRACT

The general solution for dynamic state estimation is to model the system as a hidden Markov process and then employ a recursive estimator of the prediction–correction format (of which the best known is the Bayesian filter) to statistically fuse the time-series observations via models. The performance of the estimator greatly depends on the quality of the statistical mode assumed. In contrast, this paper presents a modeling-free solution, referred to as the observation-only (O_2) inference, which infers the state directly from the observations. A Monte Carlo sampling approach is correspondingly proposed for unbiased nonlinear O_2 inference. With faster computational speed, the performance of the O_2 inference has identified a benchmark to assess the effectiveness of conventional recursive estimators where an estimator is defined as effective only when it outperforms on average the O_2 inference (if applicable). It has been quantitatively demonstrated, from the perspective of information fusion, that a prior "biased" information (which inevitably accompanies inaccurate modelling) can be counterproductive for a filter, resulting in an ineffective estimator. Classic state space models have shown that a variety of Kalman filters and particle filters can easily be ineffective (inferior to the O_2 inference) in certain situations, although this has been omitted somewhat in the literature.

1. Introduction

Dynamic state estimation has been a long-standing and vibrant area of research concerned with the sequential process of estimating a/multiple state(s) evolving over time based on noisy observations. It is the core of many fundamental problems including positioning, tracking, econometric forecasting, adaptive control, etc.

A "naïve" estimation solution is to infer the state directly from the noisy observations received in discrete time instants, hereafter referred to as the observation-only (O_2) inference, which will be addressed in this paper. This is a computationally fast estimation method, providing accuracy that is completely dependent on the observation noise regardless of the state process (for which there is, therefore, no need to model it).

* Corresponding author at: BISITE Research Group, Faculty of Science, University of Salamanca, Salamanca 37008, Spain. Tel.: +34 655 248188; fax: +34 923 294 514.

E-mail addresses: t.c.li@usal.es, t.c.li@mail.nwpu.edu.cn, tiancheng.li1985@gmail.com (T. Li), corchado@usal.es (J.M. Corchado), jbajo@fi.upm.es (J. Bajo), sdsun@nwpu.eud.cn (S. Sun), fcofds@usal.es (J.F. De Paz).

http://dx.doi.org/10.1016/j.ins.2015.09.041

In contrast to the straightforward O_2 inference, which provides only the point state-estimate, the prevailing solution that has been most investigated is to model the system as a hidden Markov process and employ a recursive estimator to statistically fuse the observations with models in real time. In this case, a two-step estimation paradigm must be adopted, including model identification based on data and filter design based on the identified model [41]. 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. The posterior probability density function (PDF) can be analytically computed only for linear systems with additive Gaussian noises for which the known Kalman filter [24,25] gives the optimal estimate (and some other special cases [9]). In the general case of nonlinear system or/and non-Gaussian noises, it is impossible to compute the exact form of the posterior PDF; instead, one has to resort to some form of approximation which can be parametric (e.g. Gaussian filters or Gaussian sum filters), non-parametric (e.g. Monte Carlo methods) or a mixture of both. An astonishing surge of various recursive filters/smoothers has been witnessed since [24,25].

These recursive estimators, which have the Bayesian paradigm as the theoretically most elaborated base [26], perform well as long as the models used are accurate, having few disturbances, and that the approximation (required in nonlinear systems) is insignificant. Ideally, an optimality (e.g. Cramér–Rao lower bounds, CRLB [14,51,57]) can be reached if the physical world and the assumed model coincide perfectly. However, in most practical problems, accurate knowledge of the state process model (and noises) is often missing. The model of a real process may differ from the assumed model or the best available model for that process, leaving a difference we refer to as modeling error.

It has been well acknowledged in literature since at least [13,21,22] that modeling errors (and significant disturbances) can easily cause significant performance deteriorations or even failures of filters. Therefore, dealing with modeling errors has been a fundamental problem. This, however, is not a problem for the O_2 inference as it is free of state process modeling. For recursive estimation, a large variety of strategies have been proposed to enhance the filtering performance including model assessment [11], adaptive filtering (e.g. [19,34]), robust filtering (e.g. effective characteristics [7], particularly including detection and treatment of uncertain noise [45], outlier [38], abrupt motion [36], asynchronous observations [47,40] and colored noises [53]), "direct" filtering [41], variable rate filtering [15] and finite impulse response filtering [29], just to name a few. Similar issues occur in Bayesian smoothers and predictors [1,6,18,48] as well as other recursive estimators e.g. optimization-based estimator [27,42,46]. The situation will be much more complicated in the multi-target case of cluttered environments, see e.g. [3,30,31,55]. We do not intend to detail these in this paper. However, we would like to point out that:

(1) While considerable efforts have been devoted to developing sophisticated recursive filters, the general effectiveness of these filters has remained elusive. Simply stated, it is rare to be asked whether the use of a filter will pay off when modeling errors (including outlier noise) occur or when too much approximation is triggered. This is primarily because a clear definition of the effectiveness for general filters is still missing. Such a definition would require a clear, efficient and engineer-friendly benchmark that is qualified to assess all filters in a consistent manner. The same holds for the work on smoothers and other recursive estimators.
(2) It has been demonstrated that the Bayesian inference can behave very badly if the model under consideration is erroneous e.g. [18]. More specifically simple deterministic methods outperform the Bayesian filter in a type of finite-state estimation [44] even when the model is properly set up. In any case, the quantitative analysis of the failure of filters is missing. This paper will thoroughly demonstrate that the O_2 inference can outperform recursive filters in certain situations, thus indicating that filters do not always pay off.

In this paper, two primary contributions have been made with regard to these fundamental issues.

(1) The O_2 inference is established as a benchmark, a bottom line, to assess the effectiveness of recursive estimators, including the Bayesian filter, where an estimator is defined as effective only when it can at minimum outperform the O_2 inference on average in accuracy. For a nonlinear observation function, a bias is noticed in the O_2 inference and, consequently, a Monte Carlo sampling-based debiasing approach is proposed.
(2) The effectiveness of the Bayesian filter of the prediction–correction format is quantitatively investigated from the information fusion perspective, and examples are evaluated on classic filtering models via simulation. Both theoretical studies and simulation results show that the O_2 inference can easily outperform the filters in certain situations, more so than expected. This deserves particular attention for the application of any filter.

The remainder of the paper is organized as follows. The basic idea of the Bayesian filter and the O_2 inference is given in Section 2. Section 3 investigates the effectiveness of the recursive filter from the general perspective of information fusion while Section 4 presents simulation results based on three representative problem models to demonstrate the theoretical findings. We conclude in Section 5.

2. O_2 inference for state estimation

2.1. A brief review of Bayesian filters

The dynamic state estimation, also referred to as the filtering problem, is generally modeled in the state space where the system being modeled is assumed to be a Markov process of hidden state. This can be formulated as a state space model (SSM) that is comprised of a state process equation and an observation equation as follows

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

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

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 noises affecting the state process equation $\boldsymbol{f}_t$ and the observation equation $\boldsymbol{h}_t$ respectively. In particular, (1) is a difference equation for discrete time, while for continuous time it is a differential equation. However, in this paper we will not distinguish them since the proposed O_2 inference does not need this equation.

Within the framework of the Bayesian statistical inference, the Bayes posterior distribution $p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t})$ will, given all the historical observations $\boldsymbol{y}_{1:t} = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_t\}$, solve the filtering problem, which basically consists of predicting and correcting two steps. The predicting step combines the previous filtering (a posteriori) distribution $p(\boldsymbol{x}_{t-1}|\boldsymbol{y}_{1:t-1})$ with the state process $p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{y}_{1:t-1})$ (i.e. Chapman–Kolmogorov equation) 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 predicted probability distribution (often called simply "the prior"). It should be noted that such predictions assume that the transforms are predictable, which is not always the case. Given a new observation $\boldsymbol{y}_t$, the prior will be updated/corrected by the 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 [24] (also referred to as Kalman–Bucy filter [25] in the continuous-time case) gives the closed form solution to the linear system with additive Gaussian noise, which is optimal in the sense of minimum mean squared error. For general nonlinear systems, it is necessary to resort to either parametric or non-parametric approximations. In the former case, the posterior PDF is represented by a family of functions that are fully characterized by parameters such as Gaussian filters (including KF and its approximate extensions [20,23,39,54]) or Gaussian sum filters [2]; see also [50]. A few statistical parameters such as the mean and variance are sufficient to represent a Gaussian distribution but not a general PDF, for which parameterization is either impossible or will suffer from significant approximation errors. In this case, the posterior PDF must be approximated e.g. via the most popular Monte Carlo approximation, which has different formulations including particle filters [16,27,43], point mass filters [49] and particle flaw filters [10]; see also [42]. All these filters need to carefully assume the state process function $\boldsymbol{f}_t$ and the system noises $\boldsymbol{u}_t$ and $\boldsymbol{v}_t$, which are very critical for the accuracy of the filter but are not at all easy to accurately identify, especially for uncertain and abrupt systems e.g. [4,27,42].

The performance of all of these filters depends greatly on the coincidence between the physical world and the model assumed. However, it is rare for them to coincide exactly. We will quantitatively show in Section 3 that a prior "biased" information which inevitably accompanies inaccurate modelling can be counterproductive, thereby resulting in a posterior "worse" than the O_2 inference. In contrast, the O_2 inference that does not assume the state process will not experience the same predicament and may in fact achieve better results; see the following subsection.

To note, the idea of releasing the prior information is reminiscent of the maximum likelihood estimator (MLE), another popular statistical inference approach whose primary difference from the Bayesian estimation is that no prior knowledge is used. Here, MLE is available if the observation function (including $\boldsymbol{h}_t$ and $\boldsymbol{v}_t$) is fully known. However, the likelihood function can be very complicated for nonlinear models and the maximization calculation difficult. More importantly, MLE does not guarantee unbiasedness [12]; the bias can be significant in situations when the mean of the likelihood distribution is far from the peak.

2.2. O_2 inference: concept and practice

Given the observation function $\boldsymbol{h}_t$, a straightforward way to estimate the state is to infer it directly from the noisy observation(s), namely the O_2 inference, which is independent of any prior information. It can be conceptually written as follows

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

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

To enable the calculation of (5), a necessary and sufficient condition is that the observation function $\boldsymbol{h}_t$ is reversible and the random variable $\boldsymbol{v}_t$ is specified. A more general situation would include an unknown observation function $\boldsymbol{h}_t$. It is a mandatory requirement for understanding the observations in all kinds of estimators and therefore must be identified prior to estimation if it is unknown. Here we do not include this issue so as not to distract from the main contribution of this paper. Moreover, this paper has not included missing observation [40] or delay [47] issues.

First, observation noise $\boldsymbol{v}_t$ can be complicated (e.g. colored/correlated [53]) and time-varying. More importantly, however, it may be inestimable in practice, not to mention that it is a single white noise. In practice, we may overcome this difficulty by simply omitting it, i.e. setting it to be zero. Eq. (5) is then reduced to

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

In fact, the result given by (6) is equivalent to that of the KF for the linear observation function, when the observation error variance goes to zero [17], and to that of the MLE for the linear observation function with additive noises of the symmetric probability distribution. However, this will introduce biases (i.e. the expectation of the estimate is not equal to the true state) if $\boldsymbol{h}_t$ is nonlinear or if the noise has non-zero expectation, where the bias depends on the noise, the true state and the inversing function. This has been recognized when converting polar/spherical observations to Cartesian coordinates for the use of Kalman filters, see e.g. [5]. To a certain extent, the bias can be avoided algebraically for simple inversing function and noises (such as Gaussian noises).

If the observation noise is known (e.g. through off-line training of the observation data), it is straightforward to calculate the unbiased mean and (co)variance of the O_2 inference for linear inversing function and simple white noise. For nonlinear inversing and/or non-Gaussian noise, we propose a Monte Carlo (MC) sampling method for unbiased O_2 inference. The idea is to sample a group of samples from the noise distribution $\boldsymbol{v}_t^{(i)} \sim \boldsymbol{v}_t, i = 1, 2, \ldots, I$ and use them to enable the inversing calculation of (5) as

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

Then, we can easily calculate the mean and (co)variance of the O_2 inference based on these transformed samples as follows

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

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

The MC method accommodates any type of noises and inversing function and can achieve any degree of accuracy given a sufficient number of samples, which is superior to the algebraic debiasing approaches [5] that only apply to simple noises and few particular inversing functions. To save computation, the samples can be created in a deterministic manner in case of simple noises e.g. sigma point approaches [20,23].

Second, with regard to the observation function $\boldsymbol{h}_t$, a primary challenge for the O_2 inference is the irreversibility of the function, for which inversing is not directly applicable. It can be viewed as an under-determined system which is equally challenging for a filter. In practice, the design/use of an under-determined observation system should be avoided. 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 determined or even over determined. That is, for state $\boldsymbol{x}_t$, we resort to a set of synchronous observations (e.g. corresponding to n sensors) such as

$$\begin{cases} \boldsymbol{y}_{1,t} = \boldsymbol{h}_{1,t}(\boldsymbol{x}_t, \boldsymbol{v}_{1,t}) \\ \boldsymbol{y}_{2,t} = \boldsymbol{h}_{2,t}(\boldsymbol{x}_t, \boldsymbol{v}_{2,t}) \\ \qquad \cdots \\ \boldsymbol{y}_{n,t} = \boldsymbol{h}_{n,t}(\boldsymbol{x}_t, \boldsymbol{v}_{n,t}) \end{cases} \tag{10}$$

where $\boldsymbol{y}_{i,t}$ denotes the observation received by sensor i, and $\boldsymbol{v}_{i,t}$ denotes the noise affecting the observation function $\boldsymbol{h}_{i,t}$. As mentioned, the noise can be omitted if it is unknown, but shall be taken into account if known.

The O_2 inference then works by solving (10) for χ_t (the observed part of $\boldsymbol{x}_t$). There is generally one unique solution for the properly determined system; however, in other cases, the equations can be over-determined. The over-determined system will be divided into multiple sub-systems, exactly determined, each of which infers an estimate. Finally, all estimates are fused in an optimal way, e.g. nonlinear (weighted) least square estimation. This will be addressed separately in another work. The over-determined system is beneficial, as it will provide a more accurate estimate and handle clutter and miss-detection through multi-sensor data fusion [31,32].

However, we point out here that $\boldsymbol{h}_t$ is generally given in a few simple forms for real life sensor models. As a common irreversible case, the observation function is non-monotonic and its inversing calculation involves a sign problem. In cases when the state is bounded in a positive or negative space, the sign problem can be avoided or easily solved; see Section 4.3. Otherwise it needs to be separately determined. There are two ways to determine the sign of the state apart from using more sensors. The first is to employ the state process function (if given) and the previous estimate. This can be taken as the default method. The second way is to use an additional estimator/filter to estimate the sign (if the model is fully known), which is computationally more intensive. Both will be shown in our simulation in Sections 4.1 and 4.2. In both cases, the O_2 inference will not only utilize the observation information, but also use the state process information, which can be referred to as O_2+ inference. Finally, we note that, arguably, there are still situations of very poor observability for which the O_2 inference is inapplicable.

We should clarify that the O_2 inference only estimates the dimensions of the states that have been observed. It is the same with filters where the unobserved dimensions of the state are implicitly inferred from the observed dimensions based on their physical relationship contained in the state process model. For example, with respect to time, the differentiation of position is velocity while differentiation of velocity is acceleration. Given their relationship, the unobserved dimensions can be inferred from the observed dimensions based on successive estimates in time series, for which smoothing/fitting may be necessary. This will form a key content of our future study.

To note, the O_2 inference is in fact involved in the core of the wireless triangulation, trilateration and multilateration positioning technology based on angle of arrival, signal strength and time difference of arrival respectively [8,37]. While the O_2 inference has rarely been considered in the evaluation of Bayesian filters/smoothers, we will in fact show that the O_2 inference can easily outperform filters.

2.3. Effectiveness of a recursive filter

The essence of the Bayesian statistical inference is the information fusion of the prior (that contains the history information transited via a state process model) and the observation, while the O_2 inference is a computationally faster, data-driven solution which disregards history observations. Comparably, both the strength and the weakness of the filter lie on the utilization of the state process model which is helpful only when a better estimate, namely the posterior, is obtained. In this case, we define the effectiveness of a filter as follows:

Definition 1. A filter is defined as effective for a particular estimation problem only when its estimates, at minimum, statistically outperform the O_2 inference (if applicable) in accuracy, under the same observation conditions.

According to this definition, a filter is effective only when the estimate a posteriori is statistically more accurate than the estimate directly inferred from the observation. This does not conflict with the fact that distributions, rather than point estimates, are propagated in the Bayes recursions. While the posterior CRLB [14,51,57] provides a lower bound on the mean-square error (MSE) 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. The following section will quantitatively assess the effectiveness of the recursive filter under the representative Gaussian assumption.

3. Probability of filter benefit

For simplicity, both the prior x_p and the O_2 inference x_o are assumed to be subject to Gaussian in the 1-dimensional state space, either biased or unbiased with regard to the true state x_T i.e. $p(x_o) = \mathcal{N}(m_o, \delta_o^2)$, $p(x_p) = \mathcal{N}(m_p, \delta_p^2)$. Here, we omit the reasons that cause the bias (to the prior or to the O_2 inference) and are only concerned with how the bias that once occurred will affect the filtering result in different situations. The Bayesian filter fuses $p(x_o)$ and $p(x_p)$ obtaining the posterior $p(x_f) = \mathcal{N}(m_f, \delta_f^2)$. We have the following proposition of the Kalman-rule fusion.

Proposition 1. *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.

It is very critical to note, as we will show in what follows, that $x_f \sim p(x_f)$ might not be more preferable (namely closer to the true state) as compared with $x_o \sim p(x_o)$, although the variance of the estimate is actually smaller as $\delta_f^2 \le \min\{\delta_o^2, \delta_p^2\}$. Instead, it is when and only when $|x_T - x_o| > |x_T - x_f|$, that the filter estimate x_f is more preferable than the O_2 inference x_o. Here, we define

Definition 2. The probability of filter benefit (PoFB) is defined as $\text{PoFB} = P(|x_T - x_o| > |x_T - x_f|)$.

The calculation of the PoFB can be expanded as follows

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

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

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

Eq. (13) can be rewritten in terms of expected values as

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

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

As a quantitative interpretation of Definition 1 with regard to the PoFB, we have

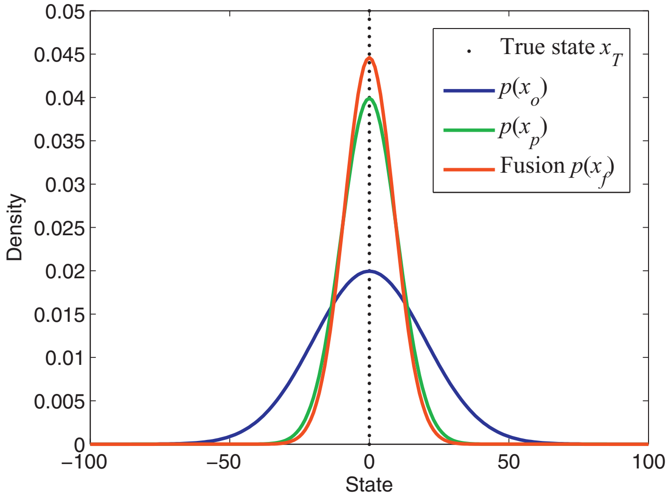

Fig. 1. Optimal fusion of two unbiased Gaussian distributions. (For interpretation of the references to color in the text, the reader is referred to the web version of this article.)

Proposition 2. *A filter is effective when and only when the PoFB* > 0.5.

If a filter is defined as ineffective ($PoFB \leq 0.5$), it means the prior is statistically useless or even counterproductive on average. In the following subsections, we will evaluate the PoFB under different situations.

3.1. Optimal fusion of two unbiased Gaussian distributions

First, we consider the case that both the prior and the O_2 inference are unbiased, i.e. $m_o = m_p = x_T$ for which we have the following remark with regard to Proposition 1.

Remark 1. Optimal fusion of two unbiased Gaussian distributions gives an unbiased Gaussian distribution.

As shown in Fig. 1, for example, the optimal fusion of the Gaussian distribution (blue) $p(x_o) = \mathcal{N}(0, 400)$ and the Gaussian distribution (green) $p(x_p) = \mathcal{N}(0, 100)$, results in the Gaussian distribution (red) $p(x_f) = \mathcal{N}(0, 80)$. More generally, for a range of different variance ratio δ_p^2/δ_o^2, the PoFB is given by the red curve shown in Fig. 3. The results show that in this situation the PoFB is always larger than 50%, indicating that the fusion will give a better estimate than the O_2 inference on average i.e. the filter is effective.

3.2. Optimal fusion of one biased and one unbiased Gaussian distribution

If there is only one distribution between the prior and the observation that is unbiased in a filter, it must be the observation. This is because the observation is independent of the prior but the prior is dependent on the observation. While a biased observation will surely cause a biased prior for the next time-instant, the contrary does not hold.

For the situation in which the prior is biased but the O_2 inference is unbiased, as shown in Fig. 2, we have the following remark with regard to Proposition 1.

Remark 2. Optimal fusion of one unbiased Gaussian distribution with one biased Gaussian distribution gives a biased distribution with the mean lying between the means of the original two distributions.

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 bias of $p(x_p)$ over the standard deviation of $p(x_o)$, 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 $p \geq 0$ and the result holds true for a negative BR. 100,000 random samples are generated separately from distributions such as $x_o \sim p(x_o)$ and $x_f \sim p(x_f)$ to calculate the PoFB for different VR $r \in [0.01,\ 1000]$ and different BR $p \in [0, 10]$.

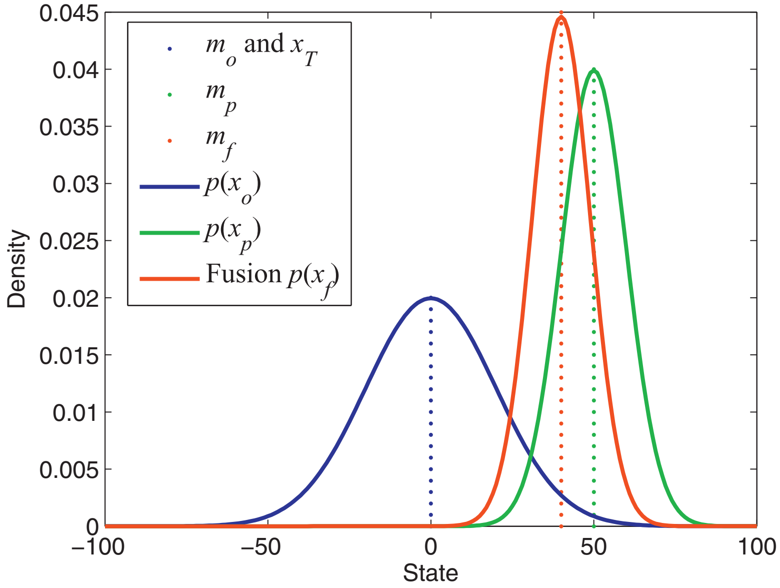

Fig. 2. Optimal fusion of one unbiased and one biased Gaussian distributions.

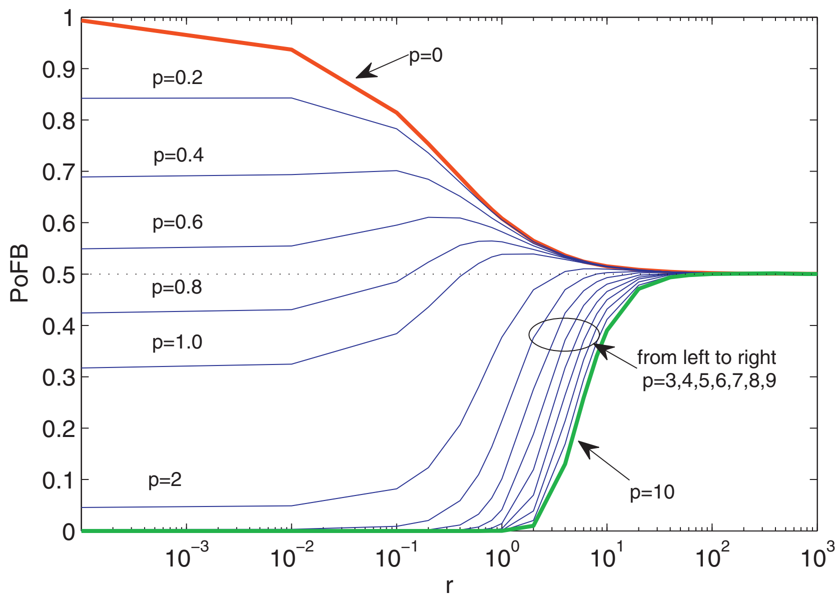

Fig. 3. PoFB for different VR r and BR p when the observation is unbiased.

In particular, $p = 0$ means that two distributions are unbiased as addressed in the preceding subsection. The results are given in Fig. 3. We have the following observations.

First, the PoFB tends to converge to 0.5 when r goes to infinite. In particular, for $p \geq 2$, the larger the VR r is, the larger the PoFB is, approximately; for $p \leq 0.4$, the larger the VR is, the smaller the PoFB is, approximately; for $0.4 < p < 2$, the PoFB goes up and then reduces down to 0.5 with the increasing of VR r. This is in line with the fact that a larger r corresponds to a relatively larger δ_p^2 of $p(x_p)$ which will have a smaller effect on the fusion distribution $p(x_f)$ in the KF. For a very large r, the effect can be omitted, after which we have $p(x_f) \approx p(x_o)$, and then PoFB = 0.5. This confirms that the O_2 inference is nothing more than the equivalent to the KF when the variance ratio (the observation error variance divided by the prior estimate error variance) goes to zero.

Secondly, when BR $p \leq 0.6$, PoFB > 0.5 . That is to say, the fusion has more than an approximately 50% possibility of obtaining a more accurate estimate than the O_2 estimate. In other words, when the bias of the biased distribution is not significant, the

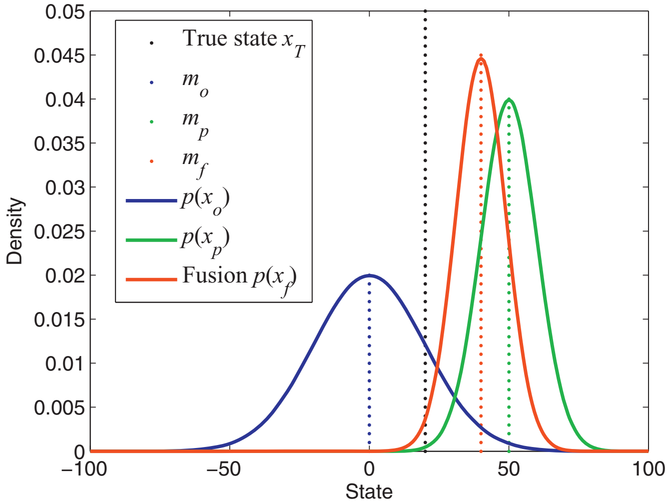

Fig. 4. Optimal fusion of two biased Gaussian distributions.

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.

Thirdly, when BR $p \geq 0.8$ (and 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 that when the bias of the prior is significant (whether because of large modeling/approximation errors or disturbances/outliers), 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, rather than the filter, is recommended.

Fig. 3 shows that for $r \to 0$, the prior $p(x_p)$ will dominate the fusion result, causing $p(x_f) \approx p(x_p)$, then the PoFB will almost fully depend on the BR p: the smaller p is, the larger the PoFB is. However, in general the prior that is affected by both the process noise and the history observation noises cannot be so accurate as compared with the underlying observation (except when it has been confirmed by the filter that a significant outlier observation occurs).

In general terms, the results indicate that if the prior is slightly biased or just unbiased, an accurate prior (of small variance) will be beneficial; otherwise it will be counterproductive for an accurate estimation.

3.3. Optimal fusion of two biased Gaussian distributions

In a more general case, both the prior and the O_2 inference can be biased, for which we have another remark.

Remark 3. Optimal fusion of two biased distributions gives an almost surely biased estimate in which the bias will be at least smaller than the larger bias of the original two.

This can be illustrated as shown in Fig. 4 although only one specific case is given where $m_o < x_T < m_p$. This is the case in which the fusion $p(x_f)$ has a comparably high possibility of obtaining a better estimate. Without loss of generality, we assume a different true state m_T which is chosen by adjusting a scaling parameter m that is defined as

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

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

For different cases scaled by parameters $m = \{-10, \ -5, -2, \ -1, -0.1, \ 0.1, \ 1, \ 2, \ 5, \ 10, \ 30\}$, the PoFB results are plotted separately in Fig. 5 which compares the O_2 inference $x_o \sim p(x_o)$ with 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 bias of the prior is larger than that of the observation), the PoFB will be smaller than 50% and the larger p is, the smaller the PoFB is.
(2) When $m \geq \ p$ (i.e. $m_o \leq m_p \leq x_T$; the bias of the prior is smaller than that of the observation), the PoFB will be larger than 50% and the larger p is, the larger the PoFB is.
(3) When $0 < m < p$ (i.e. $m_o < x_T < \ m_p$), the PoFB is complicated and depends on r, m, p (see the sub-plots for $m = 1, \ 2, \ 5$). Roughly, with the increase of $r > 1$, the PoFB will go up over 0.5 and then finally decrease to 0.5.

In summary, the filter is not very likely to outperform the O_2 inference in this case except when the observation (inference in the state space) is much worse than the prior.

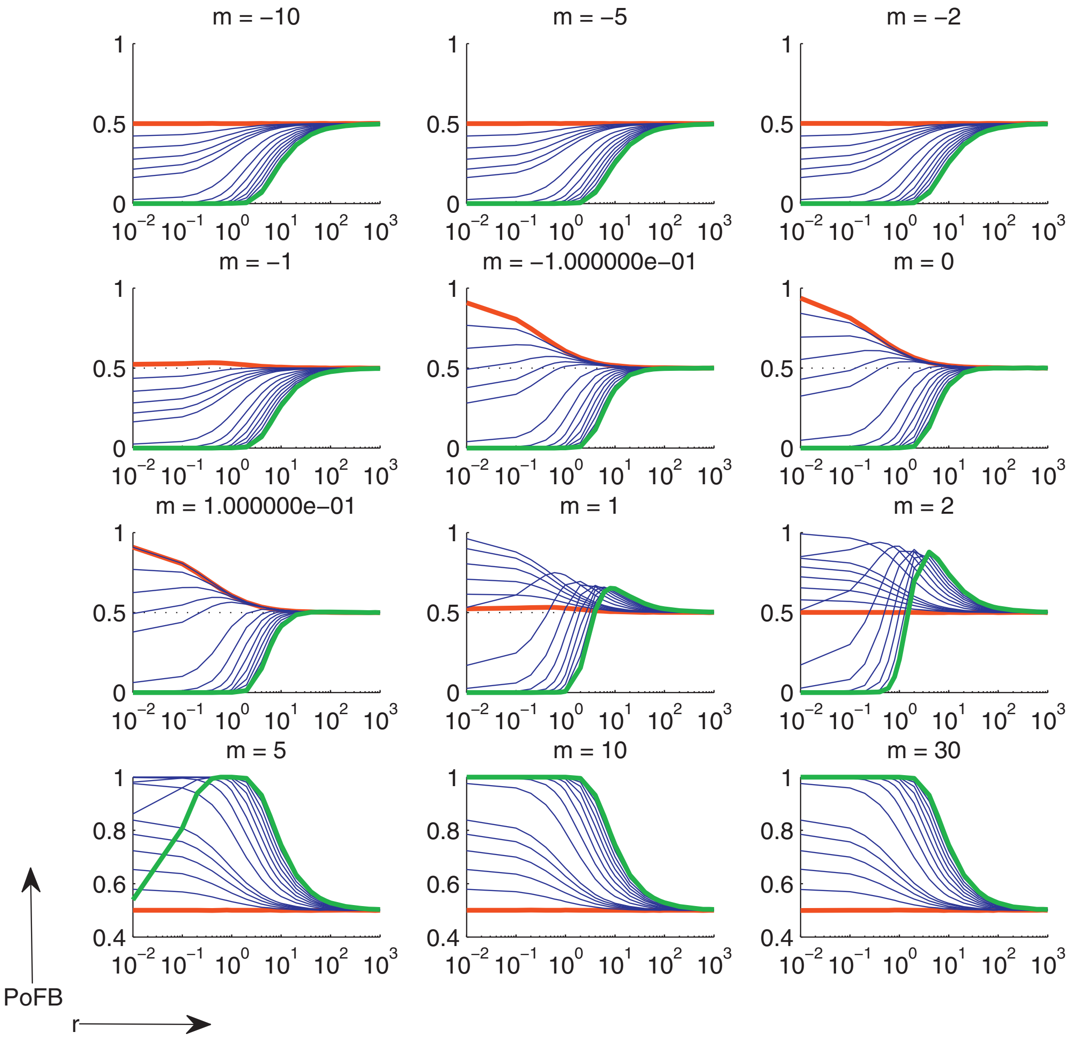

Fig. 5. PoFB for different scaling parameters m, VR r and BR p; the red line is for $p = 0$, the green line is for $p = 10$, while the blue lines are in between. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.4. Suboptimal particle Bayes fusion

The particle filter (PF) represents the posterior PDF by a set of weighted particles $\{x_t^{(n)},\ w_t^{(n)}\},\ n = 1, 2, \ldots, N$, i.e.

$$p(x_t) \approx \sum_{n=1}^{N} w_t^{(n)}\ \delta\left(x_t - x_t^{(n)}\right) \tag{19}$$

where $\delta(\cdot)$ is the Dirac delta impulse, $x_t^{(n)}$ are the possible values of the true state x_t at time t, N is the number of particles, $w_t^{(n)}$ are weights assigned to the particles and all the weights whose sum is one.

The essence of the PF is to evaluate how well each particle conforms to the dynamic model and explains the observations, using this assessment to generate a weighted particulate approximation to the filtering distribution, and hence form state estimates. The weights of the particles are reweighted over time based on the sequential importance sampling principle

$$w_t^{(n)} \propto w_{t-1}^{(n)}\ \frac{p\left(y_t|x_t^{(n)}\right)p\left(x_t^{(n)}|x_{t-1}^{(n)}\right)}{\pi\left(x_t^{(n)}\right)} \tag{20}$$

where $\pi(\cdot)$ is a proposal distribution to generate particles.

After weight updating, resampling is commonly applied to reduce the weight variance so that all particles will have equal or approximate weights [33], namely the sampling importance resampling (SIR) filter. Arguably, if the number of particles is

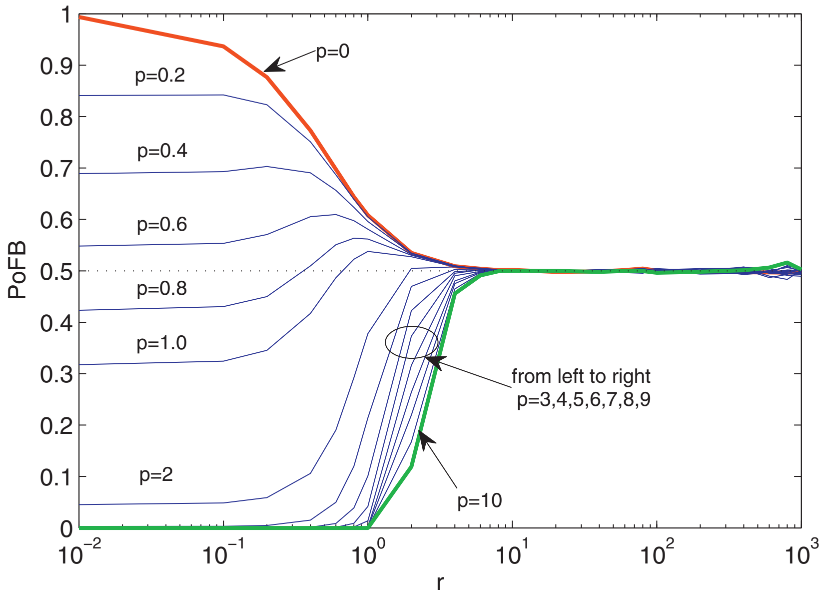

Fig. 6. PoFB of the PF for different VR r and BR p.

sufficiently large, the PF is able to achieve the Bayesian optimal performance. The particle filter does not require the underlying distribution to be Gaussian, but for simplicity, we still assume Gaussian distributions here.

Assume that the prior $p(x_p)$ is represented by a set of equally weighted particles $\{x_t^{(n)}, \frac{1}{N}\},\ n = 1, 2 \ldots, N$ (after resampling) i.e. $p(x_p)$. The likelihood distribution based on $p(x_o)$ is then used to update the weights $\{w_t^{(n)}\},\ n = 1, 2 \ldots, N$, obtaining the Bayes posterior distribution $p(x_f)$. In order to calculate the PoFB based on the distribution represented by particles, 1,000,000 random samples are generated separately from the O_2 inference $x_o \sim p(x_o)$ and the posterior $x_f \sim p(x_f)$ for different $r \in [0.01,\ 1000]$ and $p \in [0, 10]$ as defined in (16) and (17). In particular, $p = 0$ means the prior $p(x_p)$ is unbiased. The results are shown in Fig. 6, and are very similar to the optimal fusion as shown in Fig. 3. This makes sense as, theoretically, if the posterior is normally distributed, the particle filter is equivalent to the Kalman filter given an adequate number of particles.

We have the following observations that are consistent to those of the preceding sections on the filter fusion:

(1) When the bias of the prior is not significant, the fusion will be acceptable and is more likely to benefit (as compared with observation-inference). This is the case whereby the filter is effective;
(2) When the bias of the prior is significant, the fusion will be more likely to obtain worse results than the unbiased observation-inference. This is the case whereby the filter is more likely to be ineffective.

We leave here the PoFB result and the discussion of two biased distributions for the particle Bayes fusion, which are very similar to Fig. 5 and can be found in [32].

3.5. Discussions

The above study on the effectiveness of Bayesian filters has demonstrated that “prior-observation” fusion is not guaranteed to provide a benefit. Instead, it is only when (1) both the observation and the prior are ideally unbiased, (2) the bias of the prior is very small while the observation-inference is unbiased, or (3) the bias of the observation-inference is more significant than the prior, that the 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, in general the conditions of the system, namely r, p and m, vary with time giving way to a situation in which at some stages a filter is effective (the prior obtained is good) while at other stages it is not (the prior is relatively bad). In our simulations given in Section 4, we will evaluate the filters on their average performance over the entire simulation period.

We have only considered the error on the mean of the distribution (bias) but not on the (co)variance, which is also critical to the filter; sophisticated methods proposed for tuning the noise variance can be seen here [4,7,17,45,53]. If there is an error with the assumption of the (co)variance, the performance of the filter will more likely degrade; in contrast, with the O_2 inference it does not matter much (except when debiasing is applied). Here we may extend the results given under Gaussian fusion/filtering to a general albeit somewhat obvious conclusion:

Remark 4. Whether the filter is effective or not primarily depends on the quality of the prior (which in turn depends on the quality of the state process model assumed), especially the bias of the prior; the extent of the effectiveness will depend on the ratio of the variances of the prior and the observation.

As addressed, the approximation, together with filter initialization errors, system disturbances/outliers and modeling errors, can all lead to a prior error in the Bayesian recursive inference. Owning to the infinite impulse response nature of the recursive filter [28,29], any error once generated in the filter, whether due to erroneous modeling, outlier data or too much approximation, will have an effect on the subsequent estimates (priors and posteriors). That is to say, recursive estimators are potentially endangered by accumulation of approximation errors. More importantly, with the rapid development of sensors, the observation obtained can be very accurate (especially when multiple sensors are jointly used), corresponding to a large VR r as previously addressed, which will further prevent the benefit of filtering. As such, a filter shall be carefully evaluated before being applied. More precisely, we have several principles as follows before presenting our simulations.

(1) If the observation noise is significant (e.g. not zero-mean and of high variance), neither the O_2 inference nor the filter can be good; the O_2 inference comparably suffers more from the observation noise.
(2) If the system can be correctly modeled and the filter can be well initialized, being affected with no or small disturbances/outliers, then the filter will work as well as expected.
(3) If the state process model cannot be well modeled or the filter has to make significant approximation for use, the filter may lose to the O_2 inference.
(4) If multiple/massive sensors are available, the O_2 inference can benefit from multiple sensor data fusion where its modeling-free advantage will be more prominent. The more sensors, the better accuracy and reliability [30,32]. However, this may not hold for the filter.
(5) By ‘fitting/smoothing’ the results of the O_2 inference across successive scans via the state process information (if available), more accurate or further information about the state can be inferred.

4. Simulations

In this section, we will investigate the effectiveness of several known (extensions of) Kalman filters and particle filters based on two popular one-dimensional state space models, one with Gaussian state process noise and the other non-Gaussian, and a representative maneuvering target tracking case.

4.1. Filters using correct models vs. the O_2 inference

In this simulation, all of the filters will use exactly the correct state process model and known system noises, and are initialized properly with regard to the ground truth. No disturbance or outlier occurs in such a perfectly assumed model. This is the most favorable situation for the filters to achieve the best possible performance. The state process equation and the observation equation are given respectively as follows

$$x_t = \frac{x_{t-1}}{2} + \frac{25x_{t-1}}{1+x_{t-1}^2} + 8\cos(1.2(t-1)) + u_t \tag{21}$$

$$y_t = \frac{x_t^2}{20} + v_t \tag{22}$$

where the process noise u_t is Gaussian $u_t \sim \mathcal{N}(0,\ Q)$ and the observation noise is also Gaussian $v_t \sim \mathcal{N}(0,\ R)$. We firstly set $Q = 10, R = 1$ which have been the default parameter settings in many publications since [16]. We set the initial state as $x_0 \sim N(0, 1)$.

Inversing (22), the (biased) O_2 inference gives

$$\hat{x}_t = \pm\sqrt{20 \times (y_t - v_t)} \tag{23}$$

Here, we explore three different ways to determine the sign of the estimate. The first uses the state process function (default solution), the second uses the PF filtering result, and the third uses the true sign (although it is in fact unknown; here we assume there is one method that could correctly estimate the sign of the true state). They correspond respectively to the following three calculations, i.e. O_2+ inferences

$$\hat{x}_t = \mathrm{sgn}\left(\frac{\hat{x}_{t-1}}{2} + \frac{25\hat{x}_{t-1}}{1+\hat{x}_{t-1}^2} + 8\cos(1.2(t-1))\right)\sqrt{20 \times y_t} \tag{24a}$$

$$\hat{x}_t = \mathrm{sgn}(\hat{x}_{t,PF})\sqrt{20 \times y_t} \tag{24b}$$

$$\hat{x}_t = \mathrm{sgn}(x_t)\sqrt{20 \times y_t} \tag{24c}$$

The above calculations set the noise v_t to be zero and the results are actually biased as stated. This is the last choice we can have if the noise v_t is unknown. If it is known, the unbiased estimation can be obtained by using the proposed MC debiasing

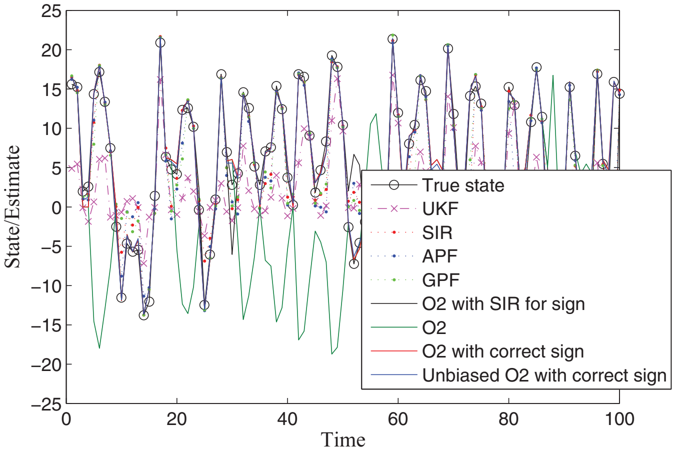

Fig. 7. True state and estimates of different estimators against time.

Table 1
Performance of different estimators (100 MC runs × 100 time-steps).

	RMSE	
	Mean	Variance
UKF	7.254	0.001
SIR (PF)	3.951	0.294
GPF	4.520	0.253
APF	4.207	0.448
O_2 inference	16.243	1.4×10^{-29}
O_2 inference with SIR for sign	4.063	0.602
O_2 inference with correct sign	1.391	5.5×10^{-32}
Unbiased O_2I with correct sign	1.229	5.8×10^{-4}

method. That is, to sample a set of samples from the noise distribution $v_t^{(i)} \sim v_t, i = 1, 2, \ldots, I$ and use them as sample noises separately in the inversing calculation, e.g. (24c), as

$$\hat{x}_t^{(i)} = \text{sgn}(x_t) \times \sqrt{20 \times \left(y_t - v_t^{(i)}\right)} \tag{24d}$$

Based on these samples, we can easily get the mean and variance of the O_2+ inferences as given in (8) and (9) respectively. Here, we set the number of noise samples $I = 100$.

For comparison, the UKF [23] (unscented KF; the unscented transform parameters are set as $\alpha = 1, \beta = 0, \kappa = 2$, which, however, are by no means considered the best choice here), auxiliary PF (APF) [43], Gaussian PF (GPF) [28] as well as the basic SIR PF that uses systematic resampling [33] have been implemented. All filters are initialized with a zero-mean random state with variance 2.

The root mean square error (RMSE) is used and is defined as follows:

$$\text{RMSE} = \sqrt{\frac{1}{M}\sum_{i=1}^{M}(x_{t,i} - \hat{x}_{t,i})^2} \tag{25}$$

where M is the number of MC runs, $x_{t,i}$ and $\hat{x}_{t,i}$ are the true state and estimate at time t of run i respectively.

To capture the average performance, 100 MC runs are executed with the same ground truth for each run. Each run consists of 100 time-steps. Firstly, when all PFs use 100 particles, the true state and estimates given by different estimators are plotted in Fig. 7, and the mean and variance of RMSE (over 100 time-steps) are given in Table 1. Secondly, for a range of a different number of particles from 20 to 500 used for the PFs, the mean RMSE and computing time of different estimators are given in Figs. 8 and 9. Finally, for a range of different observation noise variances $R \in [0.0001, 10{,}000]$, the mean RMSE of different estimators of a single run of 10,000 time-steps (where all PFs use 100 particles) are given in Fig. 10. These results show that:

(1) All the (biased and unbiased) O_2 inference approaches are extremely faster computationally than the filters, except the one using the SIR PF for estimating the sign of the state which is slowed down by the PF.
(2) Compared with others, the PFs (SIR, GPF and APF) do not make much difference with each other on this model. Specifically, a small observation variance is not always good for the PF whether GPF, APF or SIR: when R is reduced from 1 to

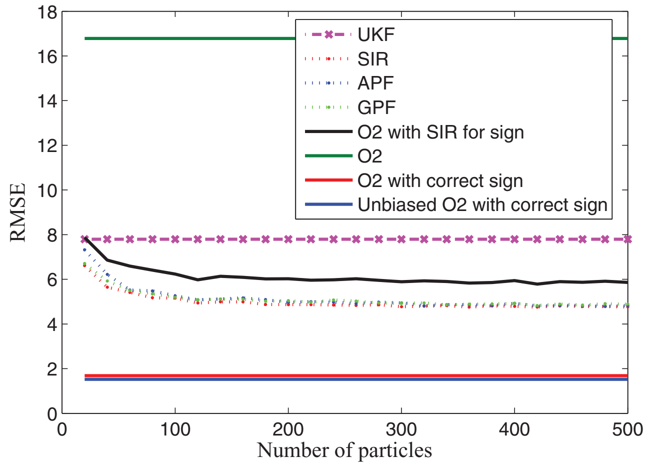

Fig. 8. RMSE of different estimators against the number of particles used in PFs.

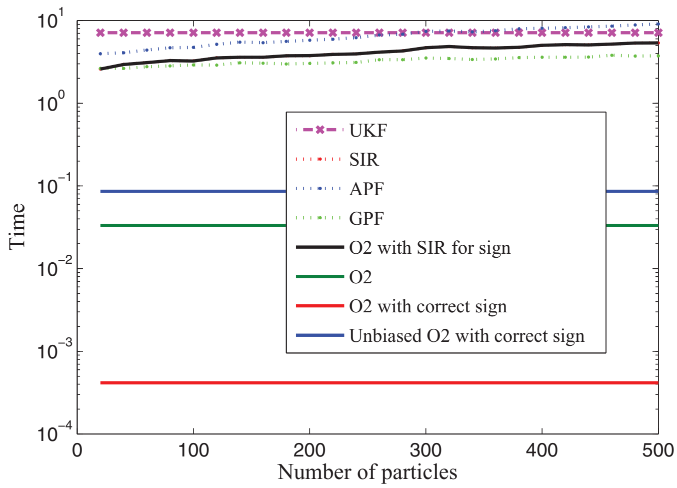

Fig. 9. Processing time of different estimators against the number of particles used in PFs.

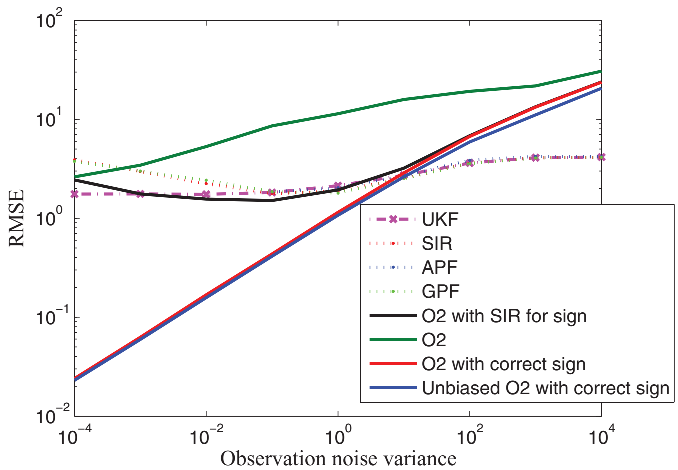

Fig. 10. RMSE of different estimators against different observation noise variances.

0.00001 or increased from 1 to 10, 000, the RMSE of the estimate increases. The best R for them is around 1. When the observation noise variance is larger than 1, it is straightforward that the larger the noise is, the worse the filters are. But, a very accurate observation (e.g. $R < 1$) corresponds to a sharp likelihood distribution, which can cause significant weight degeneracy/impoverishment, also reducing the filtering quality. This is a particular problem of PFs [35] for which considerable efforts have been devoted to designing a good proposal that coincides with the posterior distribution; see our next simulation.

(3) With the number of particles used increasing, the PFs will obtain gradually better results (up to an approximately stable level) but will also consume more time.

(4) The unbiased O_2 inference with the correct sign performs the best of all the O_2 inference approaches. But, the improvement due to the debiasing strategy is insignificant, indicating that the bias is insignificant.

Straightforwardly, we have the following findings on the effectiveness of the filters, where the threshold is approximate:

(1) With regard to the default O_2 inference, all the simulated filters are effective except the PFs for $R < 0.0005$.

(2) With regard to the O_2 inference with the SIR filter for estimating the sign, all the filters are effective for different observation noise except for $R \in [0.001, 1]$.

(3) With regard to the O_2 inference with the correct sign (biased or unbiased), all the filters are ineffective except when the observation noise is very large e.g. $R > 10$.

The results show that the (biased or unbiased) O_2 inference with the correct sign for the estimate can easily outperform the filters in terms of both small mean and variance of RMSE. It also shows that, for this model, the sign of the estimate affects the O_2 inference significantly as both (24a) and (24b) can erroneously estimate the sign of the state that switches between positive and negative, as shown in Fig. 7 (in this regard, this model is very challenging). The wrong choice of the sign of the state will significantly increase the RMSE. This is the reason the default O_2 inference performs poorly. However, it is possible to find a method to estimate the sign of the state and, therefore, efforts should be made to do so, which might be more valuable than designing a filter for this model.

We need to reiterate that the sign problem does not exist (at least not so significantly) in many other problems, such as target tracking where the state of interest is most commonly bounded in a limited region (e.g. in a known view field in the coordinate system); see our third simulation. Simply, if we define the RMSE[1] on the magnitude (absolute value) of the estimate only, namely root mean square absolute error (RMSAE), as follows

$$\text{RMSAE} = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\left(|x_{t,i}| - \left|\hat{x}_{t,i}\right|\right)^2} \tag{26}$$

then, the sign is no longer a problem. The default O_2 inference will perform the same as the O_2 inference with the correct sign, which will outperform all filters as long as the observation is not too bad (the variance $R < 10$). In the sense of this matric, the effectiveness of the filters used on this problem model is not optimistic at all.

4.2. Filters using biased process noises vs. the O_2 inference

In this simulation, we use another state space model that has also been widely employed for filter evaluation since first proposed in [52], with the state process equation and the observation equation respectively given as follows

$$x_t = 1 + \sin(\omega\pi t) + \phi_1 x_{t-1} + u_t \tag{27}$$

$$y_t = \begin{cases} \phi_2 x_t^2 + v_t & t \le 30 \\ \phi_3 x_t - 2 + v_t & t > 30 \end{cases} \tag{28}$$

where 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 $v_t \sim \mathcal{N}(0, R)$. We first set $R = 0.00001$. These are the default parameter settings in many publications including [52].

To carry out the O_2 inference, inversing (28) after taking off the unknown noise item v_t, we have

$$\hat{x}_t = \begin{cases} \text{sgn}\sqrt{|y_t/\phi_2|} & t \le 30 \\ \dfrac{y_t + 2}{\phi_3} & t > 30 \end{cases} \tag{29}$$

where sgn stands for $sgn(1 + \sin(w\pi t) + \phi_1\hat{x}_{t-1})$, namely the default O_2 inference.

If the observation noise v_t is available, the proposed MC debiasing strategy can be further applied for the nonlinear transformation when $t \le 30$. That is, for $i = 1, 2, \ldots, I$ we have

$$x_t^{(i)} = \text{sgn} \times \sqrt{\left|\left(y_t - v_t^{(i)}\right)/\phi_2\right|} \quad t \le 30 \tag{30}$$

[1] This metric is inspired by prof. Petar Djurić at Stony Brook University in the first author's email conversation with him.

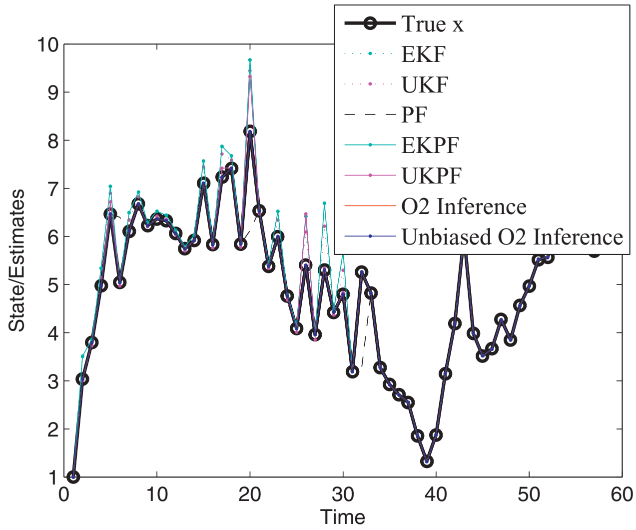

Fig. 11. The true state and estimates of different estimators against time.

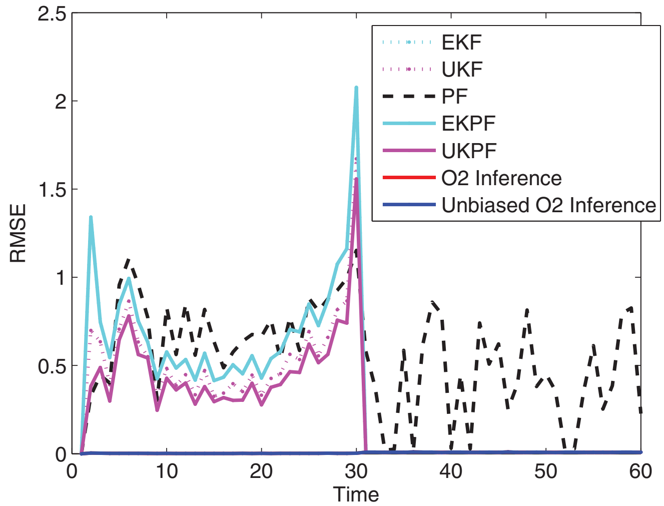

Fig. 12. RMSE against time of different estimators of 100 MC runs.

where $v_t^{(i)} \sim \mathcal{N}(0,\ R)$ and $I = 100$. For the linear transformation when $t > 30$, the O_2 inference given by (29) is unbiased and its variance can be analytically given, namely R/ϕ_3.

A series of filters are employed for comparison, including EKF (extended KF), UKF, the SIR PF, the PF that use EKF and UKF separately as the proposal. We use 200 particles for the PF and the initial state variance of 0.75 for the EKF/UKF. The unscented transform parameter is set as $\alpha = 1, \beta = 0, \kappa = 2$ (the same as used in [52]). The true state and the initial unbiased estimate of all filters are all starting from $x_1 = 1$. Since UKF/EKF cannot be used directly for this Gamma noise, we assume equivalent variance 0.75 as an alternative, i.e. they admit a modeling error of mean 1.5 of the process noise as $\mathcal{G}a(3, 2)$ is of mean 1.5, and variance 0.75. This corresponds to the practical situation where the state process noise is unknown and is incorrectly assumed when a filter is employed – the chances of this occurring are quite high in reality. While the PFs use the correct model and parameters, they admit MC approximation errors.

To capture the average result, 100 MC runs are performed with random re-initialization for each run (different to the simulation given in Section 4.1, leading to a large RMSE variance for the O_2 inference). Each run consists of 60 time-steps. The true state and estimates given by different estimators for one run are plotted in Fig. 11. In particular, if the state is known to be always positive, then the sign for the O_2 estimate can simply be set as positive and no sign estimation is needed. The O_2 inference will then compute much faster.

The RMSE of different estimators are plotted in Fig. 12. The mean and variance of RMSE over time and the computing time of each estimator are given in Table 2. It shows that the O_2 method (whether biased or unbiased) has outperformed all the

Table 2
Performance of different filters and the O_2 inference.

	RMSE		Computing time (s)
	Mean	Variance	
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}

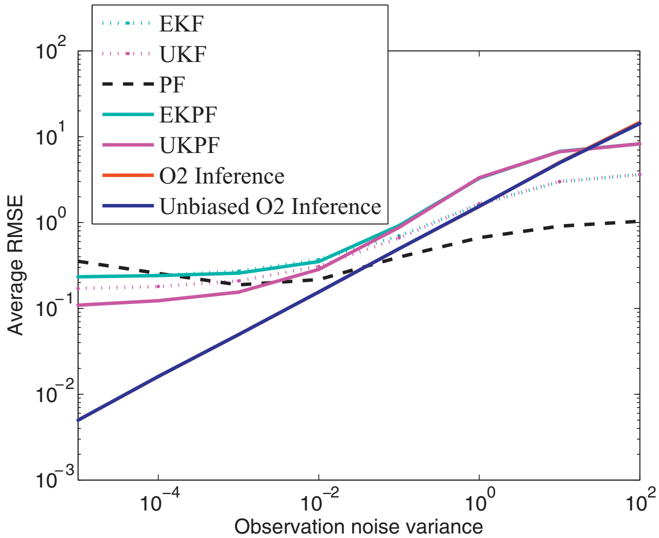

Fig. 13. Average RMSE of different estimators of 60 steps × 100 MC runs for different observation noises.

filters by several orders of magnitude in terms of both RMSE and computing speed, which indicates these filters are significantly ineffective for this model. The unbiased O_2 inference outperforms the biased O_2 inference slightly, indicating that the bias caused by the inversing calculation is, again, insignificant. To the best of our knowledge, such a good performance exhibited by the O_2 inference has never been reported before, although many filters have been proposed to apply to this simulation model.

Furthermore, for a range of different observation noise variances ranging from 0.00001 to 100 for R, the average RMSE is given in Fig. 13. It can be seen that (here thresholds are approximate): when $R < 0.04$, all these filters are ineffective; when $0.04 < R \leq 1$, PF is effective while the 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 become effective.

Both filtering models given by ((21) and (22)) and ((27) and (28)) have been widely investigated to demonstrate the advantage of one filter over others, but they have all failed to include the comparison with the O_2 inference, despite the fact that the straightforward O_2 inference is computationally much faster than any filter. More seriously, the performance of these filters will be further reduced in case of mismodeling (e.g. the state process function is not exactly known) and/or significant disturbances/outliers, leading to more negative effectiveness. We believe these two models are not unique. This is a critical fact that shall not be omitted; instead, great precautions shall be taken before the use of a filter.

4.3. *Maneuver target tracking*

In this simulation, we apply the O_2 inference for a maneuver target tracking based on the model used in [56]. However, we will not reproduce the sophisticated algorithms (including both maneuver estimator and tracking filter) that have been implemented therein. Instead, we will test the O_2 inference on exactly the same observation condition and target trajectories. The target-moving scenario is given as follows.

The state of the target is denoted as (at time t)

$$\boldsymbol{x}_t = [p_{x,t}, \dot{p}_{x,t}, p_{y,t}, \dot{p}_{y,t}]^T \tag{31}$$

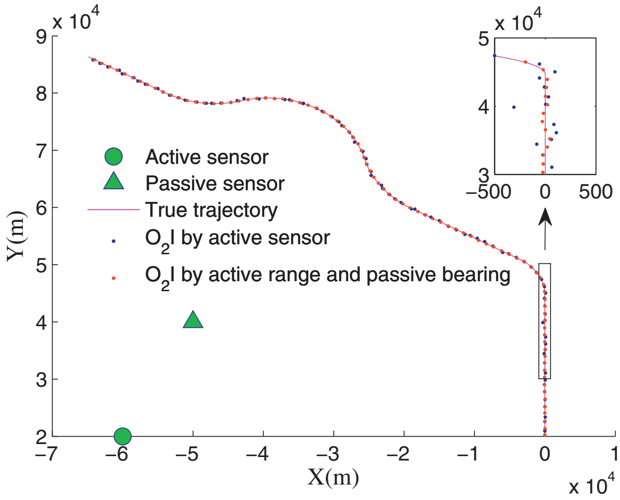

Fig. 14. The maneuver target tracking scenario and the O_2 inferences given by sensor data. (For interpretation of the references to color in the text, the reader is referred to the web version of this article.)

where $[p_{x,t},\ p_{y,t}]^T$is the position while $[\dot{p}_{x,t}, \dot{p}_{y,t}]^T$ is the velocity in the $x-y$ dimensions in Cartesian coordinates respectively. The ground truth is a target moving with a constant speed of 250 m/s (from $t = 1$ s to $t = 100s$) with initial state $\boldsymbol{x}_0 = [0, 0, 20000, 250]^T$. The maneuver over time is as follows (as shown in Fig. 14):

From $t = 100$ s to $t = 130$ s, the target turns left with 2°/s;
From $t = 130$ s to $t = 200$ s, the target moves in a straight line;
From $t = 200$ s to $t = 245$ s, the target turns right with 1°/s;
From $t = 245$ s to $t = 335$ s, the target turns left with 1°/s;
From $t = 335$ s to $t = 380$ s, the target turns right with 1°/s;
From $t = 380$ s to $t = 430$ s, the target moves in a straight line.

One active sensor located at $[S_{x,1}, S_{y,1}] = [-60{,}000, 20{,}000]$ and one passive sensor located at $[S_{x,2}, S_{y,2}] = [-50{,}000, 40{,}000]$ are used for observation. For the active sensor, the observation is a noisy range and bearing vector, given by

$$\boldsymbol{z}_t = \begin{bmatrix} r_t \\ \theta_t \end{bmatrix} = \begin{bmatrix} \sqrt{(p_{x,t} - S_{x,1})^2 + (p_{y,t} - S_{y,1})^2} \\ \arctan\left(\dfrac{p_{x,t} - S_{x,1}}{p_{y,t} - S_{y,1}}\right) \end{bmatrix} + \begin{bmatrix} v_{r,t} \\ v_{\theta,t} \end{bmatrix} \tag{32}$$

where the observation noise is Gaussian and uncorrelated between rang and bearing, which can be written as $v_{r,t} \sim N(\cdot; 0, \sigma_r^2)$, $v_{\theta,t} \sim N(\cdot; 0, \sigma_\theta^2)$ with $\sigma_r = 20$ m, $\sigma_\theta = 5$ mrad.

For the passive sensor, the observation is bearing-only, as given by

$$\beta_t = \mathrm{actan}\left(\frac{p_{x,t} - S_{x,2}}{p_{y,t} - S_{y,2}}\right) + v_{\beta,t} \tag{33}$$

where the observation noise is Gaussian, $v_{(\beta,t)}\ \sim N(\cdot;\ 0,\ \sigma_\beta^2),\ \sigma_\beta = 1$ mrad.

The observation of the active sensor is made every 5 s while the observation of the passive sensor is made every 1 s. First, we use only the active sensor for O_2 inference, which is a properly-determined observation system. Every 5 s, one estimate will be obtained, as was achieved with [56]. If the observation noises are unknown, the O_2 inference can be realized by inversing (32) after removing v_t, we have

$$\begin{bmatrix} \hat{p}_{x,t} \\ \hat{p}_{y,t} \end{bmatrix} = +/- \begin{bmatrix} \tan(\theta_t)\sqrt{\frac{r_t^2}{1+\theta_t^2}} \\ \sqrt{\frac{r_t^2}{1+\theta_t^2}} \end{bmatrix} + \begin{bmatrix} S_{x,1} \\ S_{y,1} \end{bmatrix} \tag{34}$$

where the inversing calculation of the arctan function involves a sign problem, denoted by "+/-" in (34) which can be easily determined in this scenario: when the bearing observation is smaller than $\pi/2$, the sign is positive "+"otherwise it is negative"−".

From (34), we have the O_2 inference results given by the active sensor plotted in the blue dots in Fig. 14. If the observation noises are known, the proposed MC debiasing strategy shall be applied by sampling a set of noise samples $v_{r,t}^{(i)} \sim v_{r,t}, v_{\theta,t}^{(i)} \sim$

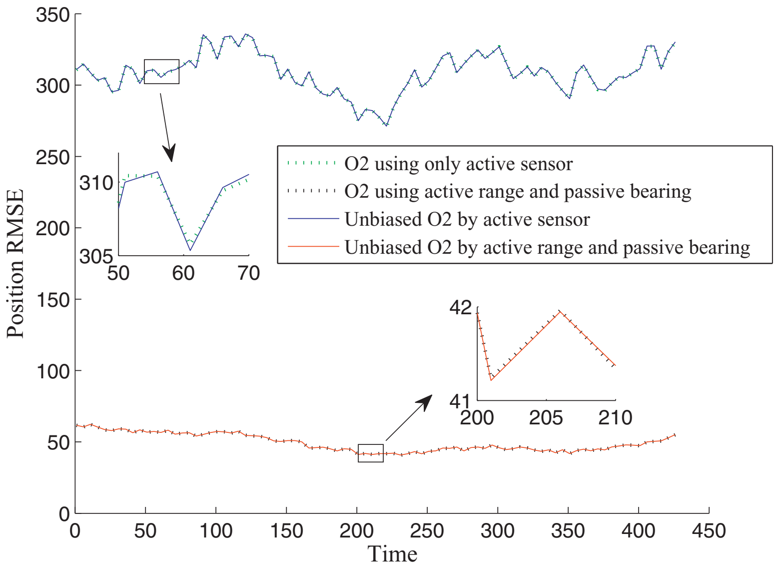

Fig. 15. Position RMSE of the (slightly biased) O_2 inference using different observations.

$v_{\theta,t}, i = 1, 2, \ldots, I$, generating correspondingly a set of estimates ($i = 1, 2, \ldots, I$) as follows:

$$\begin{bmatrix}\hat{p}_{x,t}^{(i)}\\ \hat{p}_{y,t}^{(i)}\end{bmatrix} = +/- \begin{bmatrix} \tan\left(\theta_t - v_{\theta,t}^{(i)}\right)\sqrt{\frac{\left(r_t - v_{r,t}^{(i)}\right)^2}{1+\left(\theta_t - v_{\theta,t}^{(i)}\right)^2}} \\ \sqrt{\frac{\left(r_t-v_{r,t}^{(i)}\right)^2}{1+\left(\theta_t - v_{\theta,t}^{(i)}\right)^2}}\end{bmatrix} + \begin{bmatrix}S_{x,1}\\ S_{y,1}\end{bmatrix} \tag{35}$$

Based on these sample estimates, we can easily get the mean and variance of the unbiased O_2 inferences as given in (8) and (9) for the x and y positions respectively.

Second, since the passive sensor has a higher quality (smaller observation error $\sigma_\beta < \sigma_\theta$ and faster scanning frequency) than the bearing observation obtained by the active sensor, we use the range observation received by the active sensor (discarding its bearing observation information) and part of the bearing observation received by the passive senor to carry out the O_2 inference (iteration in very 5 s), i.e. solving the following equations for $[\hat{p}_{x,t}, \hat{p}_{y,t}]^T$:

$$\begin{bmatrix} r_t \\ \beta_t \end{bmatrix} = \begin{bmatrix} \sqrt{(\hat{p}_{x,t} - S_{x,1})^2 + (\hat{p}_{y,t} - S_{y,1})^2} \\ \arctan\left(\frac{\hat{p}_{x,t} - S_{x,2}}{\hat{p}_{y,t} - S_{y,2}}\right) \end{bmatrix} \tag{36}$$

The O_2 inference results given by (36) are plotted by red dots in Fig. 14. The proposed MC debiasing solution can also be applied here for unbiased O_2 inference.

To note, one can also utilize only the bearing observation from two sensors for O_2 inference. But since two sensors will have very similar bearing observations around the circle curve part of the trajectory, as shown in Fig. 14, it will lead to very poor observability (a bad triangulation). Therefore, this sensor combination has not been included in our implementation. Nor have we considered the utilization of the full information from all sensors.

The position RMSE of four implementations of the O_2 inference over 1000 Monte Carlo trials is given in Fig. 15. As shown, the performance of the unbiased O_2 inference is very close to that of the O_2 inference without using the MC debiasing, or does not even improve it. This is because the target is far from the sensors, the bias caused by the nonlinear transformation is very small, and debiasing is not so necessary in this model. The mean RMSE and the variances of four different O_2 inferences are given in Table 3. It is necessary to note that the simple combination of the range observation of the active sensor, and in part the bearing observation of the passive sensor in this model, obtains an acceptable accuracy (on average, position RMSE < 50 m) that is very close to the filtering result given by the linear minimum mean square error estimator (LMMSE) in [56] with the assistance of the centralized interacting multiple model (IMM) estimator. As addressed, if the motion model of this target changes so frequently (namely abrupt motion) that it cannot be estimated by the sophisticated IMM properly, the performance of the LMMSE will greatly degrade and can easily become worse than our O_2 inference. Comparably, the O_2 inference is free of this problem and enjoys an extremely faster computational speed than the others. It therefore provides a benchmark performance for assessing the joint effectiveness of the maneuver estimator and filter used on this problem.

Table 3
Average performance of the O_2 inference using different sensor data.

	Position RMSE	
	Mean	Variance
O_2 inference (by using active sensor)	308.56	195.07
O_2 inference (active range and passive bearing)	49.43	37.84
Unbiased O_2 inference (by using active sensor)	308.29	195.32
Unbiased O_2 inference (active range and passive bearing)	49.40	37.96

In this simulation, all procedures, including the observation generation and both implementations of the O_2 inference (excluding debiasing) for 1000 MC runs, cost less than 1.4 s in total in the Matlab. It is several orders of magnitude faster than the IMM-LMMSE tracker. In fact, given that the target moves smoothly most of the time, a 'fitting/smoothing' procedure can be employed on the sensor data for gaining better O_2 inference as well as providing velocity estimation. We will investigate this possibility in our future work.

We would like to reiterate that one filter can be better than another or various others; it does not mean, however, that the best solution for a particular estimation problem must be a filter, especially when little is known about the background. Therefore, when we design a new recursive estimator or use an existing one, it is indeed necessary to compare the result to the O_2 inference to know whether the estimator is helpful, as it makes no sense to use an estimator that costs more computationally but estimates worse. This is precisely the goal of this paper. We are not criticizing any particular estimator but are instead highlighting the notion that greater attention should be paid to the models we use.

5. Conclusions

The observation-only (O_2) inference is a straightforward, and probably the simplest, solution for dynamic state estimation. We have elaborated this method systematically and proposed a Monte Carlo sampling solution for unbiased nonlinear implementation. 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 including the Bayesian filter, where an estimator is defined as ineffective for any particular problem if it does not outperform the O_2 inference (if applicable) on average in the estimate accuracy. In particular, the effectiveness of Kalman and particle filters, behind which the core idea is information fusion of the observation with the prior, is quantitatively analyzed in detail. It is shown that the Bayesian filter does not guarantee a benefit if the prior is biased, although it will more likely benefit if it is unbiased. Simulations on classic state space models have demonstrated our theoretical findings. These seemingly innocuous facts are crucial and must be considered with great caution whenever a new filter is designed or an existing one is used.

Further work will include inferring further and more accurate state-information from the O_2 inference by utilizing any available certain or uncertain state process information (e.g. via smoothing and fitting), and extending the O_2 inference to accommodate complicated sensor models and cluttered environments.

Acknowledgments

The authors acknowledge the insights of Prof./Dr. Yu-Chi (Larry) Ho, Huimin Chen, Xiao-Rong Li, Miodrag Bolicć, Petar Djuricć, Mahendra Mallick, Quan Pan, etc. on this work and the language checking by Deanna Garcia. This work is partly supported by European Commission: FP7-PEOPLE-2012-IRSES (ref. 318878) and MSCA-RISE-2014 (ref. 641794) and National Natural Science Foundation of China (No. 51475383) and Tiancheng Li's work has been supported by the Excellent Doctorate Foundation of Northwestern Polytechnical University and the Postdoctoral Fellowship of the University of Salamanca.

References

[1] J. Ala-Luhtala, S. Särkkä, R. Piché, Gaussian filtering and variational approximations for Bayesian smoothing in continuous-discrete stochastic dynamic systems, Signal Process. 111 (2015) 124–136.
[2] D.L. Alspach, Gaussian sum approximations in nonlinear filtering and control, Inform. Sci. 7 (1974) 271–290.
[3] Y. Bar-Shalom, P.K. Willett, X. Tian, Tracking and Data Fusion: A Handbook of Algorithms, YBS Publishing, 2011.
[4] 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, J. Process Control 21 (2011) 585–601.
[5] S. Bordonaro, P. Willett, Y. Bar-Shalom, Decorrelated unbiased converted measurement Kalman filter, IEEE Trans. Aerosp. Electron. Syst. 50 (2) (2014) 1431–1444.
[6] F.S. Cattivelli, A.H. Sayed, Diffusion strategies for distributed Kalman filtering and smoothing, IEEE Trans. Autom. Control 55 (9) (2010) 2069–2084.
[7] L.A. Dalton, E.R. Dougherty, Intrinsically optimal Bayesian robust filtering, IEEE Trans. Signal Process. 62 (3) (2014) 657–670.
[8] D. Dardari, P. Closas, P.M. Djurić, Indoor tracking: theory, methods, and technologies, IEEE Trans. Veh. Technol. 64 (4) (2015) 1263–1278.
[9] F. Daum, Exact finite-dimensional nonlinear filters, IEEE Trans. Autom. Control 31 (7) (1986) 616–622.
[10] F. Daum, J. Huang, A. Noushin, Exact particle flow for nonlinear filters, in: Proceedings of the SPIE 7697, Signal Processing, Sensor Fusion, and Target Recognition XIX, 769704, April 27, 2010.
[11] P.M. Djuric, J. Miguez, Assessment of nonlinear dynamic models by Kolmogorov–Smirnov statistics, IEEE Trans. Signal Process. 58 (10) (2010) 5069–5079.
[12] J.A. Fessler, Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): applications to tomography, IEEE Trans. Image Process. 5 (3) (1996) 493–506.

[13] B. Friedland, Treatment of bias in recursive filtering, IEEE Trans. Autom. Control 14 (1969) 359–367.
[14] C. Fritsche, E. Ozkan, L. Svensson, F. Gustafsson, A fresh look at Bayesian Cramér–Rao bounds for discrete-time nonlinear filtering, in: 17th International Conference on Information Fusion, Salamanca, Spain, July 2014, pp. 7–10.
[15] S. Godsill, J. Vermaak, W. Ng, J.F. Li, Models and algorithms for tracking of maneuvering objects using variable rate particle filters, Proc. IEEE 95 (5) (2007) 925–952.
[16] N. Gordon, D. Salmond, A. Smith, Novel approach to nonlinear/non-Gaussian Bayesian state estimation, IEE Proc. F Radar Signal Process. 140 (2) (1993) 107–113.
[17] W. Greg, B. Gary, An Introduction to the Kalman Filter, University of North Carolina, Department of Computer Science, North Carolina, USA, 2006.
[18] P.D. Grünwald, T. van Ommen. Inconsistency of Bayesian inference for misspecified linear models, and a proposal for repairing it, 2014, arXiv:1412.3730.
[19] F. Gustafsson, Adaptive Filtering and Change Detection, Wiley, 2000.
[20] F. Gustafsson, G. Hendeby, Some relations between extended and unscented Kalman filters, IEEE Trans. Signal Process. 60 (2) (2012) 545–555.
[21] H. Heffes, The effect of erroneous models on the Kalman filter response, IEEE Trans. Autom. Control 11 (3) (1966) 541–543.
[22] A.H. Jazwinski, Stochastic Processes and Filtering Theory, Academic Press Inc., London, 1970.
[23] 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, October 1997, pp. 110–121.
[24] R.E. Kalman, A new approach to linear filtering and prediction problems, J. Basic Eng. 82 (1) (1960) 35–45.
[25] R.E. Kalman, R.S. Bucy, New results in linear filtering and prediction theory, J. Basic Eng. 83 (1) (1961) 95–108.
[26] M. Kárný, Recursive estimation of high-order Markov chains: approximation by finite mixtures, Inform. Sci. 326 (2016) 188–201.
[27] M.M. Kogan, Optimal estimation and filtration under unknown covariances of random factors, Autom. Remote Control 75 (11) (2014) 1964–1981.
[28] J.H. Kotecha, P. Djuric, Gaussian particle filtering, IEEE Trans. Signal Process. 51 (10) (2003) 2592–2601.
[29] W.H. Kwon, P.S. Kim, S. Han, A receding horizon unbiased FIR filter for discrete-time state space models, Automatica 38 (3) (2002) 545–551.
[30] T. Li, J.M. Corchado, J. Bajo, G. Chen, Multi-target detection and estimation with the use of massive independent, identical sensors, in: Proceedings of SPIE, vol. 9469-15, Baltimore, Maryland, US, April 20–24, 2015.
[31] T. Li, J.M. Corchado, J. Bajo, S. Sun, Multi-source data clustering, in: The 18th International Conference on Information Fusion, Washington DC, US, July 6–9, 2015.
[32] T. Li, J.M. Corchado, J. Bajo, S. Sun, J.F. de Paz. Do we always need a filter? 2014, arXiv:1408.4636.
[33] T. Li, M. Bolic, P. Djuric, Resampling methods for particle filtering: classification, implementation, and strategies, IEEE Signal Process. Mag. 32 (3) (2015) 70–86.
[34] T. Li, S. Sun, J.M. Corchado, M.F. Siyau, Random finite set-based Bayesian filters using magnitude-adaptive target birth intensity, in: The 17th International Conference on Information Fusion, Salamanca, Spain, July 7–10, 2014.
[35] T. Li, S. Sun, T.P. Sattar, J.M. Corchado, Fight sample degeneracy and impoverishment in particle filters: a review of intelligent approaches, Expert Syst. Appl. 41 (8) (2014) 3944–3954.
[36] M.K. Lim, C.S. Chan, D. Monekosso, P. Remagnino, Refined particle swarm intelligence method for abrupt motion tracking, Inform. Sci. 283 (2014) 267–287.
[37] H. Liu, H. Darabi, P. Banerjee, J. Liu, Survey of wireless indoor positioning techniques and systems, IEEE Trans. Syst. Man Cybern. C, Appl. Rev. 37 (6) (2007) 1067–1080.
[38] C.S. Maíz, J. Míguez, E.M. Molanes-López, P.M. Djurić, A robustified particle filtering scheme for processing time series corrupted by outliers, IEEE Trans. Signal Process. 60 (9) (2012) 4611–4627.
[39] M. Morelande, A. Garcia-Fernandez, Analysis of Kalman filter approximations for nonlinear measurements, IEEE Trans. Signal Process. 61 (22) (2013) 5477–5484.
[40] M.M. Naushad Ali, M. Abdullah-Al-Wadud, S.-L. Lee, Multiple object tracking with partial occlusion handling using salient feature points, Inform. Sci. 278 (2014) 448–465.
[41] C. Novara, F. Ruiz, M. Milanese, Direct filtering: a new approach to optimal filter design for nonlinear systems, IEEE Trans. Autom. Control 58 (1) (2013) 86–99.
[42] S.C. Patwardhan, S. Narasimhan, P. Jagadeesan, B. Gopaluni, S.L. Shah, Nonlinear Bayesian state estimation: a review of recent developments, Control Eng. Practice 20 (10) (2012) 933–953.
[43] M.K. Pitt, N. Shephard, Filtering via simulation: auxiliary particle filters, J. Am. Stat. Assoc. 94 (446) (1999) 590–591.
[44] E. Punskaya, A. Doucet, W.J. Fitzgerald, On the use and misuse of particle filtering in digital communications, in: Proceedings of the EUSIPCO, Toulouse, France, September 3–6, 2002.
[45] W. Qi, P. Zhang, Z. Deng, Robust weighted fusion time-varying Kalman smoothers for multisensor system with uncertain noise variances, Inform. Sci. 282 (20) (2014) 15–37.
[46] C.V. Rao, J.B. Rawlings, D.Q. Mayne, Constrained state estimation for nonlinear discrete-time systems: stability and moving horizon approximations, IEEE Trans. Autom. Control 48 (2) (2003) 246–258.
[47] H. Rezaei, R.M. Esfanjani, M.H. Sedaaghi, Improved robust finite-horizon Kalman filtering for uncertain networked time-varying systems, Inform. Sci. 293 (1) (2015) 263–274.
[48] S. Särkkä, Bayesian Filtering and Smoothing, Cambridge University Press, 2013.
[49] M. Simandl, J. Kralovec, T. Söderstörm, Advanced point-mass method for nonlinear state estimation, Automatica 42 (7) (2006) 1133–1145.
[50] 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. 43 (6) (2013) 1607–1624.
[51] P. Tichavsky, C.H. Muravchik, A. Nehorai, Posterior Cramér–Rao bounds for discrete-time nonlinear filtering, IEEE Trans. Signal Process. 46 (5) (1998) 1386–1396.
[52] R. Van der Merwe, A. Doucet, N. Freitas, W. Eric, The unscented particle filter, Cambridge University, Engineering Department, England, 2000 Technical report.
[53] X. Wang, Y. Liang, Q. Pan, C. Zhao, F. Yang, Nonlinear Gaussian smoothers with colored measurement noise, IEEE Trans. Autom. Control 60 (3) (2015) 870–876.
[54] Y. Wu, D. Hu, M. Wu, A numerical integration perspective on Gaussian filters, IEEE Trans. Signal Process. 54 (8) (2006) 2910–2921.
[55] S. Yildirim, L. Jiang, S.S. Singh, T.A. Dean, Calibrating the Gaussian multi-target tracking model, Stat. Comput. 25 (3) (2015) 595–608.
[56] T. Yuan, Y. Bar-shalom, X. Tian, Heterogeneous track-to-track fusion, J. Adv. Inform. Fusion 6 (2) (2011) 131–149.
[57] L. Zuo, R. Niu, P.K. Varshney, Conditional posterior Cramér–Rao lower bounds for nonlinear sequential Bayesian estimation, IEEE Trans. Signal Process. 59 (1) (2011) 1–14.