Variational-Based Nonlinear Bayesian Filtering With Biased Observations

Aamir Hussain Chughtai, Arslan Majal, Muhammad Tahir, Senior Member, IEEE, and Momin Uppal, Senior Member, IEEE

Abstract—State estimation of dynamical systems is crucial for providing new decision-making and system automation information in different applications. However, the assumptions on the standard computational models for sensor measurements can be violated in practice due to different types of data abnormalities such as outliers and biases. In this work, we focus on the occurrence of measurement biases and propose a robust filter for their detection and mitigation during state estimation of nonlinear dynamical systems. We model the presence of bias in each dimension within the generative structure of the state-space models. Subsequently, employing the theory of Variational Bayes and general Gaussian filtering, we devise a recursive filter which we call the Bias Detecting and Mitigating (BDM) filter. As the error detection mechanism is embedded within the filter structure its dependence on any external detector is obviated. Simulations verify the performance gains of the proposed BDM filter compared to similar Kalman filtering-based approaches in terms of robustness to temporary and persistent bias presence.

Index Terms—State-space models, robust nonlinear filtering, approximate Bayesian inference, variational methods, parameter and state estimation, bias detection and mitigation.

I. INTRODUCTION

STATE estimation of a dynamical system plays an indispensable role in the correct functionality of a wide variety of applications such as energy management systems [1], smart grids [2], robotics [3], and intelligent transportation systems [4]. For linear state-space models (SSMs) with additive White Gaussian noise (AWGN), Kalman Filter (KF) is the optimal state estimator in terms of the minimum mean square error (MSE) criterion [5]. For nonlinear SSMs, the filtering becomes more challenging owing to the underlying functional nonlinearities. For such systems, several filtering approaches have been presented in the literature including the Extended Kalman Filter (EKF) [6], the Unscented Kalman Filter (UKF) [7], Particle Filters (PFs) [8] etc. For a survey of nonlinear state estimators, the reader is referred to [9].

The performances of standard filtering approaches rely on the assumption that the statistics of nominal noise entering the system are perfectly known apriori. However, the assumption can easily be violated in practice when the measurements are disturbed by noise not described by the known statistics. In addition to the nominal noise, the observations can further be corrupted by other abnormalities commonly referred to as gross errors which include outliers and biases [10]. Generally modeled with zero-mean statistics, outliers are sudden sporadic surges in the measurements. Their occurrence can be attributed to factors like sensor degradation, communication failures, environmental influences, etc. [11]. On the other hand, biases manifest in the data with non-zero mean noise statistics [12]. Multiple factors in various applications lead to the appearance of biased observations. Examples include miscalibrations of sensors, other configuration aberrations like errors in sensor location or alignment, clock errors, or malfunctioning [13]. In this work, we keep our focus only on the appearance of measurement biases and how these can be compensated within the filtering framework.

Since the magnitude of biases in the data, the instances of their occurrence, and the particular measurement dimensions which get affected are unknown and only partial statistics describing such corruptions are available, the problem of filtering during their presence is nontrivial. The challenge is further compounded by the functional non-linearities appearing in the SSMs. Given the significance of dealing with biases in the data during filtering, the topic has historically garnered the attention of various investigators. The approaches for devising bias-robust filters differ in the way such anomalies are neutralized in the filtering process.

Conventionally, biases are catered by assuming that the affected measurement dimensions are known in advance. Moreover, the bias evolution models are assumed to be simplistic or completely ignored during filtering. A straightforward approach is to jointly consider the state vector and bias vector for inference supposing the biases are described in a simple Markovian manner. With computational limitations at that time, earlier works attempted to reduce the processing overhead for such formulations [14]. In a similar vein, Schmidt aimed to simplify the joint state and bias estimation in the SKF formulation resulting in the celebrated Schmidt Kalman Filter (SKF) [15]. Interestingly, the bias is not estimated at each time step and only its correlations with the state are updated instead, making the SKF suboptimal even if the bias transition can perfectly be modeled [6]. Ideas similar to the SKF have also been proposed to cater for biased measurements, in terms of exploiting...
partial information e.g. positivity of biases [16]. These kinds of \textit{suboptimal} approaches are more useful in scenarios where it is safe to ignore the information regarding the evolution of measurement biases. With the advances in available processing power, the joint state and parameter estimation approach, in the KF framework, remains the standard go-to approach for catering biased observations [17].

Conventional methods are more relevant when the bias manifests in the observations consistently throughout the entire duration of system operation. However, these methods inherently assume prior knowledge regarding the affected measurement dimensions. For example, the authors do not consider bias estimation in the angle of arrival (AOA) measurements for filtering [18]. Such information may be available beforehand for some applications. However, it must generally be obtained from some \textit{detection} mechanism for the algorithms to work properly especially if the compromised dimensions vary over time.

As a result, more sophisticated schemes have been proposed by integrating the detection process with the filtering framework. To this end, following two possibilities exist: 1) use some external/separate detectors, 2) incorporate the detection/compensating process within a unified filtering framework. Both of these approaches have their merits and drawbacks. External detectors are particularly advantageous in terms of their off-the-shelf accessibility to several options. However, the performance of robust filtering is highly dependent on the functionality of these detectors and the way they integrate. On the other hand, internal detection methodologies are harder to design but obviate dependence on external algorithms.

In the literature, several kinds of external bias detectors sometimes called fault detection and identification algorithms (FDI), have been reported for several applications like tracking using UWB, GPS, and UMTS measurements. For example, in [19], biased measurements are identified simply by comparison of the standard deviation of range measurements with a detection threshold. A similar approach is to use normalized residuals to detect the presence of any bias [20]. The method proposed in [21] uses a historical record of sequential observations and performs a hypothesis test for detection. For bias detection, the use of classical statistical hypothesis tests like likelihood ratio test (LRT) and other probability ratio tests, has also been documented [20], [22], [23], [24]. In addition, other methods resort to deep learning for the determination of affected measurements [25]. There are several bias compensating filtering methods that rely on these kinds of separate detectors [16], [26], [27].

The other approach based on inherent bias detection and compensation for filtering is more challenging due to two underlying reasons. First, modeling bias inside the SSM, in a Markovian fashion, is tricky. As noted in [28] the bias transition cannot be simply modeled as a Gaussian centered at the current bias value. In [28], the authors model the bias stochastically by assuming it remains clamped to the previous value with certain predefined probability and jumps with the remaining probability within a set range represented as a uniform distribution. In [29], a similar model for describing bias is used. The use of random bias model has been proposed in [30] where the bias is represented with Gaussian density and subsequently the mean of the distribution is described in a Markovian manner. The use of Bernoulli random vectors is proposed to switch between multiple models catering for the presence and absence of biases. These kinds of models describe measurement biases effectively but make the use of KFs variants difficult since the underlying distributions for modeling do not remain Gaussian. Furthermore, the sole use of Gaussian approximations for inference no longer remains suitable. Therefore, the authors in [28], [29], [30], opt for the powerful PFs for the inference that can effectively handle arbitrary probability densities. The use of a large number of particles ranging from 15000 to 20000 is reported in these works. Therefore, the computational overhead of PF remains a concern and can be prohibitive in different applications.

Literature survey indicates the need for computationally efficient methods, with inherent detection mechanisms, to deal with measurement biases. Though PFs, with reduced processing burden, can be devised and several researchers are concerned with this direction of investigation [31], [32], an alternate is to use the Kalman filtering results for devising tractable robust filters. Recently, the use of Variational Bayes methods has gained traction in this regard and we focus on this class of approach. Though several robust methods belonging to this class exist that consider outliers in data [33], [34], [35], [36], however, generally there is a shortage of such methods that deal with biased measurements. Recent attempts in this direction include the work in [12] where the authors consider the Student’s-t-inverse-Wishart distribution to handle time-varying bias. However, only linear systems are considered in the derivation.

Given the backdrop, we present a novel robust filtering method to deal with biases in the measurements for nonlinear systems. The main contributions of this work are as follows.

- We present the bias detecting and mitigating (BDM) filter to deal with measurements bias using the Gaussian filtering framework. We resort to Variational inference for designing the filter with an internal bias detection mechanism as opposed to schemes requiring external detectors.
- For a given model we evaluate the Posterior Cramer-Rao Bound (PCRB) to determine the theoretical benchmark for the error performance.
- For two different scenarios i.e. persistent and temporary bias presence we evaluate the performance of the BDM filter as compared to different estimators in the literature indicating the advantages of the proposed method.

The way we have organized this article is as follows. Section II describes our modeling choice for incorporation of biases inside the SSM. In Section III, the derivation of the filter is provided. Subsequently, the performance evaluation results have been discussed in Section IV. Lastly, conclusive comments are given in Section V.

\textbf{Related Works}

Robust filtering theory has been extensively investigated in the literature. Researchers have been interested in various dimensions of the problem including the design and evaluation of filters robust to different effects like anomalies (outliers and biases) in the process and measurement models,
observations, system modeling errors, adversarial attacks on sensors, etc. Several robust filters have been proposed and evaluated for different applications resorting to techniques from diverse areas. In this subsection, we briefly highlight some other robust filtering methods.

The classical Wiener filter [37] has been successfully extended to its robust counterparts in the literature. The basic idea is to consider the least favorable power spectral densities (PSDs) for any specific uncertainty model assumed for the signal of interest and the noise. Subsequently, the optimal (using min-max MSE criterion) solution is sought to devise such filters [38].

Similarly, the KF has numerous robust extensions apart from the ones discussed in the main introduction section. Some of its basic outlier-robust derivatives include 3σ-rejection and score function type KFs [39]. These formulations use score functions applied to the residuals of the observed and predicted measurements to minimize the effect of outliers. Another robust KF derivative is the approximate conditional mean (ACM) filter which is based on the approximation of the conditional observation density prior to updating [40]. Some other variations rely on a bank of KFs to gain robustness [41]. The cause of robust filtering has also been well served by the theory of M-estimation which has helped develop many of these methods [39]. In addition, regression-based KFs have also been proposed for achieving robustness [42]. Similarly, guaranteed cost-based methods [43], Krein space methods [44] and linearly constrained KFs (LCKFs) have also been proposed in this regard [45]. Ambiguity sets for catering model distributional uncertainties have also been employed to this end [46]. Besides different information theoretic criteria have been used to devise robust KFs [47]. KFs extensions to deal with system parametric uncertainties are also well-documented [48].

Other types of robust filtering approaches have also been reported in the literature. H∞ is a popular approach that aims to minimize the worst-case estimation error by formulating a min-max problem using a smartly chosen objective function [49]. Similarly, mixed Kalman/H∞ approaches have also been proposed leveraging the merits of both the methods [50]. The use of finite impulse response (FIR) filters can also be found in this regard [51], [52], [53]. Similarly, robust recursive estimators for SSMs resorting to sensitivity penalization-based methods have been devised [54]. The use of nonparametric techniques can also be found in the robust filtering literature as well [39].

II. BIAS MODELING

As the standard SSM does not consider the possibility of measurement biases in its generative structure [55] it needs to be modified. At the same time, the model should remain amenable for VB inference. To this end, we choose the inference model from our previous work [30], with a few modifications. For a discrete time SSM, the process and measurement equations are given as follows

\[ x_k = f(x_{k-1}) + q_{k-1} \]  
\[ y_k = h(x_k) + r_k + I_k \Theta_k \]  

where \( k \) denotes the time-index, \( x_k \in \mathbb{R}^n \) and \( y_k \in \mathbb{R}^m \) are the state and measurement vectors respectively, \( q_{k-1} \in \mathbb{R}^n \) and \( r_k \in \mathbb{R}^m \) are white process and measurement noise vectors, \( f(.) \) and \( h(.) \) represent nonlinear process and measurement dynamics respectively, \( \Theta_k \in \mathbb{R}^m \) models the effect of biases in the measurements and \( I_k \in \mathbb{R}^{m \times m} \) is a diagonal matrix with Bernoulli elements \( I_k \) used to indicate the occurrence of bias in different dimensions. We assume the following noise distributions: \( q_{k-1} \sim \mathcal{N}(0, Q_{k-1}) \) and \( r_k \sim \mathcal{N}(0, R_k) \). We assume that measurements are obtained from independent sensors making \( R_k \) diagonal. For inferential tractability, the model as originally reported is simplified by ignoring the added randomness in the bias magnitude in (2). The bias evolution is expressed as follows where the modeling rationale remains the same as originally reported in [30].

\[ \Theta_k = (I - I_{k-1}) \Theta_{k-1} + I_{k-1} (\Theta_{k-1} + \Delta_k) \]  

In (3), each entry of \( \Delta_k \) allows for any drifts/changes in the bias value over time, in the corresponding dimension, given bias was present at the previous time step. On the contrary, if no bias occurred in any given dimension, at the preceding instant, it can possibly occur with a very large variance \( \sigma_k^2 \) (assuming an uninformative prior) described by the respective entries of a zero mean random vector \( \Theta_k \). The distributions of \( \Delta_k \) and \( \Theta_k \) are supposed to be white and normally distributed given as

\[ \Delta_k \sim \mathcal{N}(0, \Sigma_k) \]  
\[ \Theta_k \sim \mathcal{N}(0, \Sigma_k) \]

Note that for tractability, we have modified \( \Theta_k \) to be normally distributed instead of obeying a uniform distribution. However, assuming a very large variance does not make a practical difference. Also note that for simplicity, we do not take any transition model for \( I_k \) and assume its elements occur independently at each instance. Remaining modeling assumptions are kept the same as originally reported.

III. RECURSIVE BAYESIAN INFERENCE

Considering the inference model in (1)–(3), the Bayes rule can be employed recursively to express the joint posterior distribution of \( x_k, \mathcal{I}_k \) (considering only the random entries \( \mathcal{I}_k \)) and \( \Theta_k \) conditioned on the set of all the observations \( y_{1:k} \) analytically as

\[ p(x_k, \mathcal{I}_k, \Theta_k | y_{1:k}) = \frac{p(y_k | x_k, \mathcal{I}_k, \Theta_k) p(x_k, \mathcal{I}_k, \Theta_k | y_{1:k-1})}{p(y_k | y_{1:k-1})} \]  

Theoretically, the joint posterior can be marginalized to obtain the expression for \( p(x_k | y_{1:k}) \). With this approach, the exact sequential Bayesian processing becomes computationally infeasible. Therefore, we adopt the VB method [56] for inference where the product of VB marginals is conveniently used to approximate the joint posterior as

\[ p(x_k, \mathcal{I}_k, \Theta_k | y_{1:k}) \approx q(x_k)q(\mathcal{I}_k)q(\Theta_k) \]
With an objective to minimize the Kullback-Leibler divergence (KLD) between the marginal product and the true posterior, the VB method leads to the following marginals

\[ q(x_k) \propto \exp\left( \ln(p(x_k, I_k, \Theta_k | y^{1:k})) q(I_k) q(\Theta_k) \right) \]  
\[ q(I_k) \propto \exp\left( \ln(p(x_k, I_k, \Theta_k | y^{1:k})) q(x_k) q(\Theta_k) \right) \]  
\[ q(\Theta_k) \propto \exp\left( \ln(p(x_k, I_k, \Theta_k | y^{1:k})) q(x_k) q(I_k) \right) \]  

where \( \langle \cdot \rangle_{q(\psi)} \) denotes the expectation of the argument with respect to a distribution \( q(\psi) \). The VB marginals can be updated iteratively until convergence, using (8)–(10) in turn. The procedure provides a convenient way to approximate the true marginals of the joint posterior by approximating these as \( p(x_k | y^{1:k}) \approx q^r(x_k) \), \( p(I_k | y^{1:k}) \approx q^r(I_k) \) and \( p(\Theta_k | y^{1:k}) \approx q^r(\Theta_k) \) where \( q^r(\cdot) \) denotes the VB marginals obtained after convergence.

A. Prediction

Assuming that at each time step the posterior is approximated with a product of marginals, the predictive density can be approximated as

\[ p(x_k, I_k, \Theta_k | y^{1:k-1}) \approx p(I_k) p(x_k | y^{1:k-1}) p(\Theta_k | y^{1:k-1}) \]  

(11)

with

\[ p(x_k | y^{1:k-1}) = \int p(x_k | x_{k-1}) p(x_{k-1} | y^{1:k-1}) dx_{k-1} \]  

(12)

\[ p(\Theta_k | y^{1:k-1}) \approx \prod_{i=1}^{m} p(\Theta_{k,i} | y^{1:k-1}) \]  

(13)

We assume that the occurrence of bias is independent for each dimension and its historical existence. Using \( \theta_{ki}^\delta \) to denote the prior probability of occurrence of bias in the \( i \)th observation, the distribution of \( I_{ki}^+ \) is defined as product of independent Bernoulli distributions of each element

\[ p(I_k) = \prod_{i=1}^{m} p(I_{ki}^+) = \prod_{i=1}^{m} (1 - \theta_{ki}) \delta(I_{ki}^+) + \theta_{ki} \delta(I_{ki} - 1) \]  

(14)

where \( \delta(\cdot) \) denotes the delta function.

To evaluate (12)–(13), suppose the following distributions for the posterior marginals at time step \( k-1 \)

\[ p(x_{k-1} | y^{1:k-1}) \approx q^r(x_{k-1}) \approx N(x_{k-1} | \hat{x}_{k-1}^+, P_{k-1}^+) \]  

(15)

\[ p(I_{k-1} | y^{1:k-1}) = \prod_{i=1}^{m} p(I_{k-1,i} | y^{1:k-1}) \]  

\[ = \prod_{i=1}^{m} \left( 1 - \Omega_{k-1,i}^+ \right) \delta(I_{k-1,i}^+) + \Omega_{k-1,i}^+ \delta(I_{k-1,i} - 1) \]  

(16)

\[ p(\Theta_{k-1} | y^{1:k-1}) \approx q^r(\Theta_{k-1}) \approx N\left(\Theta_{k-1} | \hat{\Theta}_{k-1}^+, \Sigma_{k-1}^+\right) \]  

(17)

where \( \Omega_{k-1,i}^+ \) denotes the posterior probability of bias occurrence in the \( i \)th dimension. The notation \( N(m, \Sigma) \) represents a multivariate normal distribution with mean \( m \) and covariance \( \Sigma \), evaluated at \( x \). The verification of the functional forms of the distributions and the expressions of their parameters are provided in the subsequent update step of the Bayesian filter.

Since \( f(\cdot) \) is assumed to be nonlinear, \( p(x_k | y^{1:k-1}) \) can be approximated, using general Gaussian filtering results [57], as \( N(x_k | \hat{x}_k, P_k^-) \) with the parameters predicted as follows

\[ \hat{x}_k^- = \left( f(x_{k-1}) \right)_{p(x_{k-1} | y^{1:k-1})} \]  

(18)

\[ P_k^- = \left( f(x_{k-1}) - \hat{x}_k^- \right) \left( f(x_{k-1}) - \hat{x}_k^- \right)^\top_{p(x_{k-1} | y^{1:k-1})} + Q_k \]  

(19)

The remaining term required to approximate the predictive density recursively in (11) is \( p(\Theta_k | y^{1:k-1}) \). Observing (3) and (13), it is evident that \( p(\Theta_k | y^{1:k-1}) \) is a sum of \( 2^m \) Gaussian densities scaled by the probabilities of combinations of bias occurrence at previous time instance. Obviously this makes recursive inference intractable, so we propose to approximate this distribution with a single Gaussian density \( N(\Theta_k | \hat{\Theta}_k^- , \Sigma_k^-) \) using moment matching [57]. The parameters of the distribution are updated as

\[ \hat{\Theta}_k^- = \Omega_{k-1} \hat{\Theta}_k^- \]  

(20)

\[ \Sigma_k^- = (I - \Omega_{k-1}) \Sigma_k + \Omega_{k-1} \Sigma_k \]  

+ \( \Sigma_{k-1}^+ \odot (\text{diag}(\Omega_{k-1})) (\text{diag}(\Omega_{k-1}))^\top + \Omega_{k-1} (I - \Omega_{k-1}) \)  

+ \( \Omega_{k-1} (I - \Omega_{k-1}) (\text{diag}(\hat{\Theta}_k^-))^2 \)  

(21)

where \( \Omega_{k-1} \) is a diagonal matrix with entries \( \Omega_{k-1,i} \) denoting the posterior probability of bias occurrence at time step \( k-1 \). The operator \( \odot \) is the Hadamard product and diag is used for vector to diagonal matrix conversion and vice versa. The reader is referred to the Appendix for detailed derivations of (20)–(21).

Remarks: We note the following in (20)–(21):

- \( \Omega_{k-1} \) dictates how the parameters \( \hat{\Theta}_k^- \) and \( \Sigma_k^- \) are predicted.
- For the case when \( \Omega_{k-1} = I \), i.e. bias is inferred in each dimension at time step \( k-1 \) with probability 1, \( \hat{\Theta}_k^- = \Theta_k^- \) and \( \Sigma_k^- = \Sigma_{k-1}^+ + \Sigma_k \). In other words, the mean of the bias prediction (for each dimension) is retained and its covariance is predicted as sum of previous covariance and the covariance considered for the amount of drift/change in the bias.
- For the case when \( \Omega_{k-1} = 0 \), i.e. no bias is inferred in each dimension at time step \( k-1 \) with probability 1, \( \hat{\Theta}_k^- = 0 \) and \( \Sigma_k^- = \Sigma_k \). In other words, the mean of the bias prediction (for each dimension) is 0 and its covariance is predicted with very large entries.
- Similarly, the prediction mechanism can be understood if only some dimensions are inferred to be disturbed with probability 1. The bias in the particular dimensions are predicted with the mean retained and covariance updated as addition of previous covariance plus the covariance allowed for the drift/change.
- Lastly, if there is partial confidence on the occurrence of bias in any dimension at \( k-1 \), the predicted Gaussian distribution is shifted to the mean of bias estimate at \( k-1 \) scaled with a factor of \( \Omega_{k-1,i} \). In addition, the covariance gets inflated by addition of scaled elements of \( \Sigma_k \) and squared
terms of mean at $k-1$. In other words, it can be interpreted in a sense that unless there is a very high confidence of occurrence of bias at the previous time instance, the bias would be predicted with a large covariance.

B. Update

For the update step, we resort to (6)–(10) and use (11) for approximating the predictive density. For detailed derivations, the reader is referred to the Appendix.

Parameters of $q(x_k)$ are updated iteratively as

$$\dot{x}_k^+ = \dot{x}_k^- + K_k \left( y_k - \Omega_k \hat{\Theta}_k^+ - \mu_k \right)$$  \hspace{1cm} (22)

$$\mu_k = \langle h(x_k) \rangle_{p(x_k|y_{k-1})}$$  \hspace{1cm} (23)

$$P_k^+ = P_k^- - C_k K_k^T$$  \hspace{1cm} (24)

$$K_k = C_k S_k^{-1}$$  \hspace{1cm} (25)

$$C_k = \langle (x_k - \hat{x}_k^-) (h(x_k) - \mu_k) \rangle_{p(x_k|y_{k-1})}$$  \hspace{1cm} (26)

$$S_k = \langle (h(x_k) - \mu_k) (h(x_k) - \mu_k)^T \rangle_{p(x_k|y_{k-1})} + R_k$$  \hspace{1cm} (27)

Parameters of $q(I_k)$ are updated iteratively as

$$\Omega_k^+ = \text{Pr}(I_k^+ = 1) / \left( \text{Pr}(I_k^+_i = 1) + \text{Pr}(I_k^+_i = 0) \right)$$  \hspace{1cm} (28)

where denoting $k_1$ as the proportionality constant

$$\text{Pr}(I_k^+_i = 0) = k_1 (1 - \theta_k^i) \exp \left( -\frac{1}{2} \frac{(y_k^i - \nu_k^i)^2}{R_k^i} \right)$$  \hspace{1cm} (29)

$$\text{Pr}(I_k^+_i = 1) = k_1 \theta_k^i \exp \left( -\frac{1}{2} \frac{\tilde{h}_k^2 + \Theta_k^2 + (\nu_k^i + \Theta_k^i - y_k^i)^2}{R_k^i} \right)$$  \hspace{1cm} (30)

$$\nu_k = \langle h(x_k) \rangle_{q(x_k)}$$  \hspace{1cm} (31)

$$\tilde{h}_k^2 = \langle (h(x_k) - \nu_k^i)^2 \rangle_{q(x_k)}$$  \hspace{1cm} (32)

$$\Theta_k^2 = \langle (\Theta_k^i - \Theta_k^i)^2 \rangle_{q(\Theta_k)}$$  \hspace{1cm} (33)

Parameters of $q(\Theta_k)$ are updated iteratively as

$$\hat{\Theta}_k^* = \hat{\Theta}_k^- + K_k \left( y_k - \left( \nu_k + \Omega_k \hat{\Theta}_k^- \right) \right)$$  \hspace{1cm} (34)

$$\Sigma_k^* = \Sigma_k^- - C_k K_k^T$$  \hspace{1cm} (35)

$$K_k = C_k S_k^{-1}$$  \hspace{1cm} (36)

$$C_k = \Sigma_k^* \Omega_k^T$$  \hspace{1cm} (37)

$$S_k = \Omega_k \Sigma_k^* \Omega_k^T + R_k$$  \hspace{1cm} (38)

$$\hat{\Theta}_k^- = \hat{\Theta}_k^+ + \Sigma_k^* \Omega_k^T$$  \hspace{1cm} (39)

$$\Sigma_k^+ = \left( \Omega_k (I - \Omega_k) R_k^{-1} + \Sigma_k^* \right)^{-1}$$  \hspace{1cm} (40)

C. BDM Filter

Using the proposed approximations, in the prediction and update steps, we have devised a recursive filter referred as the BDM filter. Unless real world experiments reveal any prior information regarding the occurrence of bias in each dimension, we propose using an uninformative prior for $I_k^+_i$ which is commonly adopted for such cases. The Bayes-Laplace and the maximum entropy methods for obtaining uninformative prior for a parameter with finite values lead to the uniform prior distribution [58], [59]. We adopt this choice of prior for our case i.e. $\theta_k^i = 0.5 \forall i$ which has been advocated in the literature for designing robust filters [35], [36]. Algorithm 1 outlines the devised BDM filter.

### Algorithm 1: The proposed BDM filter

Initialize $\hat{x}_0^+, \hat{P}_0^+ \Theta^0_k, \Sigma_k^+, \hat{S}_k, \hat{\Sigma}_k, Q_k, R_k$;

for $k = 1, 2 \ldots K$ do

Evaluate $\hat{x}_k^+, \hat{P}_k^+$ with (18) and (19);

Evaluate $\Theta_k^+, \Sigma_k^+$ with (20) and (21);

Initialize $\theta_k^i, \Omega_k^i, \Theta_k^i$ the convergence threshold $\tau, \gamma = \tau + 1$, the iteration index $l = 1$;

Evaluate $\hat{x}_k^{+\gamma}$ and $P_k^{+\gamma}$ with (22)-(27);

while $\gamma > \tau$ do

Update $\Omega_k^i \forall i$ with (28)-(33);

Update $\Theta_k^i$ and $\Sigma_k^i$ with (34)-(40);

Update $\hat{x}_k^i$ and $P_k^i$ with (22)-(27);

Evaluate $\gamma = \| \hat{x}_k^{+(i-1)} - \hat{x}_k^{+(i-1)} \| / \| \hat{x}_k^{+(i-1)} \|$;

$l = l + 1$;

end

$\hat{x}_k^+ = \hat{x}_k^{+(i-1)}$ and $\hat{P}_k^+ = P_k^{+(i-1)}$;

$\Theta_k^- = \Theta_k^{+(i-1)}$ and $\Sigma_k^- = \Sigma_k^{+(i-1)}$;

end

IV. NUMERICAL EXPERIMENTS

To evaluate the comparative performance of the devised algorithm, numerical experiments have been conducted on an Apple MacBook Air with a 3.2 GHz M1 Processor and 8 GBs of unified RAM using Matlab R2021a.

For comparative fairness, we consider methods based on the Unscented Kalman Filter (UKF) as their basic algorithmic workhorse. The following filters have been taken into account for comparisons: the standard UKF, the selective observations rejecting UKF (SOR-UKF) [36], the Unscented Schmidt Kalman Filter (USKF) [60] and the constrained Unscented Kalman Filter (CUKF). The CUKF is devised by modifying the CSRUUKF [16] by using the standard UKF instead of SRUKF as its core algorithm.

In terms of handling data corruption, the standard UKF has no bias compensation means in its construction. By contrast, the SOR-UKF is an outlier-robust filter, with inherent data anomaly detection mechanism, which discards the observations found to be corrupted. Lastly, the USKF and CUKF both compensate for the bias partially and require an external bias detection mechanism. Bias mitigating filters with inherent detection mechanism are generally scant in the literature and most of these are based on the PFs. The USKF is a modified version of the SKF, adapted for nonlinear systems, where the bias is not exactly estimated
rather its correlations with the state are updated. The CUKF is based on two major functional components. First, it resorts to the UKF for estimation. Subsequently, it draws sigma points based on these estimates which are projected onto a region, by solving an optimization problem, supposing a constraint that the measurements can only be positively biased. In our numerical evaluation, we assume perfect detection for these two algorithms. Note that in the implementation of the USKF and CUKF, we switch to the standard UKF when no bias is detected. Also note that the proposed BDM filter has no limitations in terms of whether any bias positively or negatively disturbs the measurements. However, since the CUKF assumes a positive bias we keep this restriction in our simulations. In particular, bias in each dimension is added as a shifted Gaussian $\mathcal{N}(\mu, \sigma^2)$ where $\mu \geq 0$ [16].

For performance evaluation, we resort to a target tracking problem in a wireless network where the range measurements are typically biased [61]. The observations get biased owing to the transitions between line-of-sight (LOS) and non-line-of-sight (NLOS) conditions [62]. Such tracking systems find applications in emergency services, fleet management, intelligent transportation, etc. [63].

We consider the process equation for the target assuming an unknown turning rate as [33]

$$x_k = f(x_{k-1}) + q_{k-1} \tag{41}$$

where the state vector $x_k = [a_k, b_k, \omega_k]^{T}$ contains the 2D position coordinates $(a_k, b_k)$, the respective velocities $(\dot{a}_k, \dot{b}_k)$, the angular velocity $\omega$ of the target at time instant $k$, $\zeta$ is the sampling period, and $q_{k-1} \sim \mathcal{N}(0, Q_{k-1})$. $Q_{k-1}$ is given in terms of scaling parameters $\eta_1$ and $\eta_2$ as [33]

$$Q_{k-1} = \begin{pmatrix} \eta_1 M & 0 \\ 0 & \eta_2 M \end{pmatrix}$$

where $M = \begin{pmatrix} \zeta_t^3 / 3 & \zeta_t^2 / 2 \\ \zeta_t^2 / 2 & \zeta_t \end{pmatrix}$

Range readings are obtained from $m$ sensors installed around a rectangular area where the $i$th sensor is located at $(a^i, b^i)$, $b^i = 350 \mod 2)$. The nominal measurement equation can therefore be expressed as

$$y_k = h(x_k) + r_k \tag{43}$$

with

$$h^i(x_k) = \sqrt{(a_k - a^i)^2 + (b_k - b^i)^2} \tag{44}$$

For the duration of bias presence the following observation equations, based on the random bias model [64], are assumed

$$y_k = h(x_k) + r_k + \mathcal{J}_k(\dot{o}_k + \Delta o_k) \tag{45}$$

$$y_{k+1} = h(x_{k+1}) + r_{k+1} + \mathcal{J}_k(\dot{o}_k + \Delta o_{k+1}) \tag{46}$$

where $o_k \in \mathbb{R}^m$ models the effect of statistically independent biases in the measurements and $\mathcal{J}_k \in \mathbb{R}^{m \times m}$ is a diagonal matrix with independent Bernoulli elements with parameter $\lambda$. Our measurement model for evaluation is inspired by the LOS and NLOS models in [62], [63] where shifted Gaussian distributions are used to model the biased measurements. In effect, bias with magnitude $a^i_k$ with some additional uncertainty $\Delta o_k^i \sim \mathcal{N}(0, \Sigma^i)$ affects the $i$th dimension at time step $k$ if $\mathcal{J}_k^i = 1$. Since the exact magnitudes of biases are not generally known apriori and can occur randomly in a given range we assume $a^i_k$ to be uniformly distributed i.e. $a^i_k \sim U(0, \lambda^i)$. This is in contrast to the approaches in [62], [63] where the magnitudes of biases are assumed to be known perfectly. Similarly at time step $k + 1$, the bias sustains with the same magnitude of the previous time step $k$ in the $i$th dimension i.e. $o^i_{k+1}$ with some uncertainty $\Delta o^i_{k+1}$ if the $i$th dimension at the previous time step was affected i.e. $\mathcal{J}_k^i = 1$. Note that the evaluation model allows us to compare different methods under extreme conditions varying from the case of no biased dimension to the case where every observation can possibly get biased.

For evaluation we assign the following values to different parameters: $\Sigma^i = 0.4, \lambda^i = 90, x_0 = [0, 10, 0, -5, 3 \pi / 180]^T$, $R_k = 4I, \zeta = 1, \eta_1 = 0.1, \eta_2 = 1.75 \times 10^{-4}$. We assume $m = 4$, since for higher-dimensional problems even rejection-based methods like the SOR-UKF can have acceptable performance for a larger probability of errors owing to the redundancy of useful information in other uncorrupted dimensions. However, this does not limit the applicability of the proposed method for large $m$. Our point is to emphasize how properly utilizing information from an affected dimension, a characteristic of analytical redundancy approaches, is more useful in contrast to completely rejecting the information, a characteristic of hardware redundancy approaches [30]. In addition, we evaluate the relative performance of the proposed filter with similar analytical redundancy approaches.

For each method, the UKF parameters are set as $\alpha = 1, \beta = 2$ and $\kappa = 0$. We consider $T = 400$ time steps and carry out 100 Monte Carlo (MC) simulations for each case. All the filters are initialized with a state estimate equal to $x_0$ and its error covariance as $Q_0$ and the bias estimate is set equal to zero initially. The other parameters supposed for the BDM-UKF are: $\Sigma_0 = 0.001I$, $\Sigma_k = 1000R_k$, $\Sigma_k = 0.1R_k$ and $\theta_k = 0.5 \forall i$ and $\tau = 10^{-4}$. In CUKF the auxiliary parameters are assigned values as: $\epsilon = 1, \sigma = 2$. Lastly, the USKF is initialized with a zero mean bias with a variance of 500 having zero cross-covariance with the state. Other parameters for all the filters are kept the same as originally reported.

Performance bounds

To benchmark the relative performance of the considered algorithms, we evaluate the PCRB for the target tracking model
in (41)–(45). The bound remains valid where system dynamics can be modeled as (41) and where the measurement model can be considered to switch between (43) and (45)–(46) since we use a generalized model in our derivation. We evaluate the PCRB assuming perfect apriori knowledge of the occurrence time, duration of the bias and λ, resorting to the approach presented in [65], [66] for a generalized nonlinear dynamical system, with white process and measurement noise, given as follows

\[ x_{k+1} = f(x_k, q_k) \] (47)
\[ y_k = h(x_k, r_k, y_{k-1}, \ldots, y_{k-z}) \] (48)

where \( z \) is a positive integer. The PCRB matrix for the estimation error of \( x_k \) can be written as

\[ \text{PCRB}_k \triangleq J_k^{-1} \] (49)

where \( J_k \) can be expressed recursively as

\[ J_{k+1} = D_{k}^{22} - D_{k}^{21} (J_k + D_{k}^{11})^{-1} D_{k}^{12} \] (50)

with

\[ D_{k}^{11} = \langle -\Delta \phi_{x_k} \log p(x_{k+1} \mid x_k) \rangle p(x_{k+1}, x_k) \] (51)
\[ D_{k}^{12} = \langle -\Delta \phi_{x_k} \log p(x_{k+1} \mid x_k) \rangle p(x_{k+1}, x_k) \] (52)
\[ D_{k}^{21} = \langle -\Delta \phi_{x_k} \log p(x_{k+1} \mid x_k) \rangle p(x_{k+1}, x_k) = [D_{k}^{22}]^T \] (53)
\[ D_{k}^{22} = D_{k}^{22}(1) + D_{k}^{22}(2) \] (54)

\[ D_{k}^{22}(1) = \langle -\Delta \phi_{x_k} \log p(x_{k+1} \mid x_k) \rangle p(x_{k+1}, x_k) \] (55)
\[ D_{k}^{22}(2) = \langle -\Delta \phi_{x_k} \log p(x_{k+1} \mid x_k) \rangle p(x_{k+1}, x_k) \]

where

\[ \Delta \phi = \nabla \phi \nabla_{\phi}^T \] (56)
\[ \nabla_{\phi} = \left[ \frac{\partial}{\partial \theta_1}, \ldots, \frac{\partial}{\partial \theta_r} \right]^T \] (57)

The bound is valid given the existence of derivatives and expectations terms in (49)–(57) for an asymptotically unbiased estimator [65]. Using results from [65], [66] we obtain

\[ D_{k}^{11} = [\nabla \phi_{x_k}^T (x_k)] Q_k^{-1} [\nabla \phi_{x_k}^T (x_k)]^T \] (58)
\[ D_{k}^{12} = -\nabla \phi_{x_k}^T (x_k) Q_k^{-1} \] (59)
\[ D_{k}^{22}(1) = Q_k^{-1} \] (60)

For the period where the nominal (43) remains applicable, \( D_{k}^{22}(2) \) can be expressed as [66]

\[ D_{k}^{22}(2) = -Q_k^{-1} \left[ \tilde{F}_k \right]_{j=1}^p (x_k) \] (61)

where \( \tilde{F}_k \) and \( \tilde{H}_k \) are the Jacobians of \( f(x_k) \) and \( h(x_k) \) respectively.

For the duration of bias occurrence, \( D_{k}^{22}(2) \) can be evaluated using (55), equivalently written as follows using a technique similar to [67]

\[ D_{k}^{22}(2) = \left\langle \frac{\left[ \nabla \phi_{x_k+1} \right]_{j=1}^p (y_{k+1} \mid x_{k+1}, y_k, x_k) \left[ \nabla \phi_{x_k+1} \right]_{j=1}^p (y_{k+1} \mid x_{k+1}, y_k, x_k)}{p(y_{k+1} \mid x_{k+1}, y_k, x_k)^2} \right\rangle_{p(y_{k+1} \mid x_{k+1}, \ldots)} \] (62)

First consider (45) to evaluate \( D_{k}^{22}(2) \) at the instance of bias occurrence for which \( p(y_{k+1} \mid x_{k+1}, y_k, x_k) \) can be approximated using Monte Carlo method as

\[ p(y_{k+1} \mid x_{k+1}, y_k, x_k) = p(y_{k+1} \mid x_{k+1}) \] (63)

\[ = \int p(y_{k+1} \mid x_{k+1}, b_{k+1}) p(b_{k+1}) db_{k+1} \] (64)

\[ = \int \mathcal{N}(y_{k+1}, h(x_{k+1}) + b_{k+1}, R_k) p(b_{k+1}) db_{k+1} \] (65)

\[ \approx \frac{1}{N_{mc1}} \sum_i \mathcal{N}\left(y_{k+1}, h(x_{k+1}) + b_{k+1}^{(i)} \right) \] (66)

where \( b_{k+1}^{(i)}, i = 1, \ldots, N_{mc1} \), are independent and identically distributed (i.i.d.) samples such that \( b_{k+1}^{(i)} \sim p(b_{k+1}) \).

Accordingly, \( D_{k}^{22}(2) \) can be approximated as

\[ D_{k}^{22}(2) = \left\langle \frac{\left[ \nabla \phi_{x_k+1} \right]_{j=1}^p \left[ \nabla \phi_{x_k+1} \right]_{j=1}^p (y_{k+1} \mid x_{k+1})}{\left[ \sum_i \exp(\phi_{j}^{(i)} \mid x_{k+1}, y_k, x_k) \right]^2} \right\rangle_{p(y_{k+1} \mid x_{k+1})} \] (67)

where \( \phi_{j}^{(i)} = -0.5 \left(y_{k+1} - (h(x_{k+1}) + b_{k+1}^{(i)})^T R_{k+1}^{-1} (y_{k+1} - (h(x_{k+1}) + b_{k+1}^{(i)})) \right) \). We can further write

\[ D_{k}^{22}(2) = \left\langle \frac{\left[ \sum_i \exp(\phi_{j}^{(i)} \mid x_{k+1}, y_k, x_k) \right]_{j=1}^p \left[ \sum_i \exp(\phi_{j}^{(i)} \mid x_{k+1}, y_k, x_k) \right]_{j=1}^p \left[ \nabla \phi_{x_k+1} \right]_{j=1}^p (y_{k+1} \mid x_{k+1})}{\left[ \sum_i \exp(\phi_{j}^{(i)} \mid x_{k+1}, y_k, x_k) \right]^2} \right\rangle_{p(y_{k+1} \mid x_{k+1})} \] (68)

where \( \nabla \phi_{x_k+1} \left[ \phi_{j}^{(i)} \right] = \tilde{H}_{k+1} R_{k+1}^{-1} (y_{k+1} - h(x_{k+1}) - b_{k+1}^{(i)}) \). Resultingly, \( D_{k}^{22}(2) \) is approximated as

\[ D_{k}^{22}(2) \approx \frac{1}{N_{mc2}} \sum_{j=1}^{N_{mc2}} \sum_i \left[ \sum_i \exp(\phi_{j}^{(i)} \mid x_{k+1}, y_k, x_k) \right]_{j=1}^p \left[ \sum_i \exp(\phi_{j}^{(i)} \mid x_{k+1}, y_k, x_k) \right]_{j=1}^p \left[ \nabla \phi_{x_k+1} \right]_{j=1}^p (y_{k+1} \mid x_{k+1}) \] (69)

with \( \phi_{j}^{(i)} = -0.5 \left(y_{j}^{(i)} \mid k+1 - (h(x_{j}^{(i)} \mid k+1) + b_{k+1}^{(i)})^T R_{k+1}^{-1} (y_{j}^{(i)} \mid k+1) - (h(x_{j}^{(i)} \mid k+1) + b_{k+1}^{(i)})) \). For \( j = 1, \ldots, N_{mc2} \), are i.i.d. samples such that \( (y_{j}^{(i)} \mid k+1) \sim p(y_{j}^{(i)} \mid x_{k+1}, y_k, x_k) \).

Lastly, to evaluate \( D_{k}^{22}(2) \) for the bias persistence period we consider (46) and the difference of (45)–(46). Therefore, \( p(y_{k+1} \mid x_{k+1}, y_k, x_k) \) can be approximated for this case using Monte Carlo method as

\[ p(y_{k+1} \mid x_{k+1}, y_k, x_k) = \int p(y_{k+1} \mid x_{k+1}, y_k, x_k, \mathcal{J}_k) p(\mathcal{J}_k) d\mathcal{J}_k \] (70)
The observations are systematically biased. We consider that the measurements representative of real-world scenarios where one of the dimensions is corrupted unlike the CUKF which compensating methods namely the CUKF and the USKF. The CUKF and the USKF generally perform better than the standard UKF which exhibits the worst estimation quality since it has no means for bias detection and compensation. For the data rejection-based method i.e. the SOR-UKF, we see that for lower probabilities of existence of bias discounting the observations performs satisfactorily. However, for higher values of $\lambda$ the rejection scheme does not work well since consistently rejecting large number of the measurements leads to loss of essential information. The CUKF and the USKF generally perform better than the SOR-UKF. As compared to the other robust filters the USKF is found to have more error at lower values of $\lambda$ since it does not treat each dimension selectively. It rather uses the entire vector of measurements for compensation even if one of the dimensions is corrupted unlike the CUKF which offers a selective treatment. Importantly, it can be observed that BDM-UKF results in the least error among all the methods.

Fig. 2 shows the state RMSE of the algorithms over time for an example MC run for Case 1.

![Fig. 2. Tracking performance of algorithms over time for an example MC run for Case 1.](image-url)

A. Case 1: Consistent Bias Presence

First, we consider the case of biases consistently appearing in the measurements representative of real-world scenarios where the observations are systematically biased. We consider that the bias corrupts each observation according to (45)–(46) and bias occurs in each dimension from the start of the simulation with probability $\lambda$ and sustains for the complete run time.

Fig. 1 shows the distribution of the root mean squared error (RMSE) calculated over 100 MC runs for the algorithms under consideration at different values of $\lambda$. The lower bounds based on the diagonal entries of $\text{PCRB}_k$ depict the benchmark performance for different methods theoretically achievable considering different values of $\lambda$. For evaluation of $\text{PCRB}_k$ we assume $N_{mc1} = N_{mc2} = N_{mc3} = N_{mc4} = 100$. We can observe an intuitive trend in the relative performance of different methods. The standard UKF exhibits the worst estimation quality since it has no means for bias detection and compensation. For the data rejection-based method i.e. the SOR-UKF, we see that for lower probabilities of existence of bias discounting the observations performs satisfactorily. However, for higher values of $\lambda$ the rejection scheme does not work well since consistently rejecting large number of the measurements leads to loss of essential information. The CUKF and the USKF generally perform better than the SOR-UKF. As compared to the other robust filters the USKF is found to have more error at lower values of $\lambda$ since it does not treat each dimension selectively. It rather uses the entire vector of measurements for compensation even if one of the dimensions is corrupted unlike the CUKF which offers a selective treatment. Importantly, it can be observed that BDM-UKF results in the least error among all the methods.

![Fig. 1. Box plots of state RMSE for Case 1 with increasing $\lambda$.](image-url)

\[
\begin{align*}
    &\lambda = 0.2 \\
    &\lambda = 0.4 \\
    &\lambda = 0.6 \\
    &\lambda = 0.8
\end{align*}
\]

where $\mathcal{J}_k^{(i)}$, $i = 1, \ldots, N_{mc3}$, are i.i.d. samples such that $\mathcal{J}_k^{(i)} \sim p(\mathcal{J}_k)$. Resultingly, $D_{k}^{22}(2)$ is approximated as

\[
D_{k}^{22}(2) = \left( \frac{\sum_{i} \exp(\theta(i)) \left[ \nabla_{x_{k+1}} \theta(i) \right] \left[ \nabla_{x_{k+1}} \theta(i) \right]^\top}{\sum_{i} \exp(\theta(i))^2} p(y_{k+1}, x_{k+1}, \cdots) \right)
\]

where $\nabla_{x_{k+1}} \theta(i) = \nabla_{y_{k+1}} y_{k+1} - \nabla_{x_{k+1}} \theta(i)$.

![Fig. 2. Tracking performance of algorithms over time for an example MC run for Case 1.](image-url)

\[
D_{k}^{22}(2) \approx \frac{1}{N_{mc4}} \sum_{j} \left[ \exp(\theta(i,j)) \nabla_{x_{k+1}} \theta(i,j) \right] \left[ \nabla_{x_{k+1}} \theta(i,j) \right]^\top
\]

with $\theta(i,j) = -0.5 \left( y_{k+1} - \theta(i,j) \right) + \mathcal{J}_k^{(i)} \left( y_{k+1} - \theta(i,j) \right)^\top$

\[
\begin{align*}
    &\text{where } \mathcal{J}_k^{(i)}, i = 1, \ldots, N_{mc3}, \text{ are i.i.d. samples such that } \mathcal{J}_k^{(i)} \sim p(\mathcal{J}_k). \\
    &\text{Resultingly, } D_{k}^{22}(2) \text{ is approximated as }
\end{align*}
\]
Lastly, we evaluate the computational overhead of each algorithm for Case 1. The mean processing time, considering 100 MC runs, for each method has been summarized in Table I. We find the standard UKF to be the most economical and does not exhibit any significant change with different values of $\lambda$. The USKF takes more time since it involves updating the state and bias correlation terms. We observe a rise in the processing time of the USKF with increasing $\lambda$ since the USKF is invoked more frequently than the standard UKF we utilize when no bias is detected. The BDM-UKF and the SOR-UKF, both having an inbuilt detection mechanism, have a similar processing burden. Lastly, we find the CUKF to be the most computationally expensive algorithm. This can be owed to the presence of a convex quadratically constrained quadratic program (QCQP) which we solve using the MATLAB optimization toolbox.

### B. Case 2: Momentary Bias Presence

We also consider the case where bias randomly appears for a short duration, characterizing practical scenarios e.g. where ambient effects disturb the data signals briefly. We again use (45)–(46) for each observation contamination and suppose bias occurs in each dimension with probability $\lambda$ at instant $t = 100$ and sustains till $t = 130$ before disappearing.

Fig. 3 depicts the spread of the RMSEs for 100 MC runs for each algorithm with varying $\lambda$ along with lower bounds based on the PCRB using 100 MC samples for all the calculations. The plotting scale in this case allows us to clearly see a rise in the lower bound with increasing values of $\lambda$ which has also been observed for the previous case. This can be attributed to more chances of occurrence of biased observations with increasing $\lambda$ leading to increased estimation uncertainty. In terms of the relative performance of different methods we see a similar pattern as for the previous case, owing to the same rationale regarding the functionality of the methods. The UKF generally has the largest RMSE, followed by the USKF, the SOR-UKF and the CUKF. The momentary appearance of bias does not degrade the performance of SOR-UKF as compared to the last case, except for $\lambda = 0.8$ where it mostly diverges. SUKF is found to be relatively less effective at lower values of $\lambda$ due to its non-selective nature. Lastly, the BDM-UKF results in the lowest RMSE. Fig. 4 shows the state RMSE of the algorithms over time for one MC run representative of the general trend in Fig. 3 for large values of $\lambda$ for this case. We can observe similar results as in the preceding case.

Lastly, the mean processing overhead of each algorithm for this case is presented in Table II. We can see that the UKF takes approximately similar times as for Case 1. The overhead of remaining methods is reduced since the bias duration has now decreased. The order in which different algorithms appear in terms of the relative computational expense remains the same as observed in the previous case following from the same reasoning.  

### V. Conclusion

The performances of standard filtering approaches degrade when the measurements are disturbed by noise with unknown statistics. Focusing on the presence of measurement biases, we
devise the BDM filter with inherent error detection and mitigation functionality. Performance evaluation reveals the efficacy of the BDM in dealing with both persistently and temporarily present biases. We find the BDM filter more accurate compared to rejection-based KF methods i.e. SOR-UKF and lower than the CUKF which requires an external optimization solver. The gains come at the expense of increased computational overhead which is comparatively higher compared to the UKF and USKF. However, it is comparable to SOR-UKF and lower than the CUKF which requires an additional optimization solver.

**APPENDIX**

A. Predicting Parameters of $\mathcal{N}(\Theta^i_k, \Sigma^i_k)$

\[
\hat{\Theta}^i_k = \langle \Theta^i_k | p(\Theta^i_k|y_{1:k-1}) \rangle = \Omega^i_{k-1} \hat{\Theta}^i_{k-1}
\]

\[
\Sigma^i_k = \langle v_{1k} v_{1k}^\top | p(I_{k-1}|y_{1:k-1}) \rangle p(\Theta^i_k|y_{1:k-1},p(\Delta_{k}))
\]

where

\[
v_{1k} = (I - I_{k-1}) \tilde{\Theta}_k + I_{k-1}(\Theta^i_{k-1} + \Delta_k) - \Omega^i_{k-1} \hat{\Theta}^i_{k-1}
\]

\[
- I_{k-1}\hat{\Theta}^i_{k-1} + I_{k-1}\hat{\Theta}^i_{k-1}
\]

Resultingly $\Sigma^i_k$ can be further expressed as

\[
\Sigma^i_k = \begin{pmatrix}
(1 - I_{k-1})^2 \tilde{\Theta}^i_{k-1} & \cdots & \Pi_i (1 - I_{k-1}) \hat{\Theta}^i_{k-1} \\
\vdots & \ddots & \vdots \\
(1 - I_{k-1})^2 \tilde{\Theta}^i_{k-1} & \cdots & \Pi_i (1 - I_{k-1}) \hat{\Theta}^i_{k-1}
\end{pmatrix}
\]

with $v_{2k} = \begin{pmatrix}
(1 - I_{k-1})\tilde{\Theta}^i_{k-1} \\
\vdots \\
(1 - I_{k-1})\tilde{\Theta}^i_{k-1}
\end{pmatrix}$, $v_{3k} = \begin{pmatrix}
I_{k-1}\Delta^i_{k-1} \\
\vdots \\
I_{k-1}\Delta^i_{k-1}
\end{pmatrix}$

\[
v_{4k} = \begin{pmatrix}
I^i_{k-1}(\tilde{\Theta}^i_{k-1} + \hat{\Theta}^i_{k-1}) \\
\vdots \\
I^i_{k-1}(\tilde{\Theta}^i_{k-1} + \hat{\Theta}^i_{k-1})
\end{pmatrix},
\]

\[
v_{5k} = \begin{pmatrix}
I^i_{k-1}(\tilde{\Theta}^i_{k-1} + \hat{\Theta}^i_{k-1}) \\
\vdots \\
I^i_{k-1}(\tilde{\Theta}^i_{k-1} + \hat{\Theta}^i_{k-1})
\end{pmatrix}
\]

\[
\text{Resultingly } \Sigma^i_k \text{ can be further expressed as}
\]

\[
= \text{diag} \langle (1 - I_{k-1})^2 | p(I_{k-1}|y_{1:k-1}) \rangle \text{ diag } \langle \hat{\Theta}^i_{k-1} | p(\Theta^i_k) \rangle
\]

\[
+ \text{diag} \langle (1 - I_{k-1})^2 | p(I_{k-1}|y_{1:k-1}) \rangle \text{ diag } \langle \hat{\Theta}^i_{k-1} | p(\Theta^i_k) \rangle
\]

\[
\text{where } i \in \{1, m\}
\]

\[
\langle \Theta^i_k | p(\Theta^i_k) \rangle = \langle \Theta^i_{k-1} - \hat{\Theta}^i_{k-1} | p(\Theta^i_k) \rangle = 0
\]

\[
\text{Since } \langle \Theta^i_k | p(\Theta^i_k) \rangle = \langle \Theta^i_{k-1} - \hat{\Theta}^i_{k-1} | p(\Theta^i_k) \rangle = 0
\]

\[
\text{other terms in the expression of } \Sigma^i_k \text{ disappear and we can write}
\]

\[
\Sigma^i_k = \langle (I - I_{k-1}) \tilde{\Theta}_k + I_{k-1}(\Theta^i_{k-1} + \Delta_k) - \Omega^i_{k-1} \hat{\Theta}^i_{k-1}
\]

\[
- I_{k-1}\hat{\Theta}^i_{k-1} + I_{k-1}\hat{\Theta}^i_{k-1}
\]

\[
(\Theta^i_{k-1} - \hat{\Theta}^i_{k-1})(\Theta^i_{k-1} - \hat{\Theta}^i_{k-1})^\top I_{k-1}
\]

\[
+ (I_{k-1} - \Omega^i_{k-1}) \hat{\Theta}^i_{k-1} \hat{\Theta}^i_{k-1} + (I_{k-1} - \Omega^i_{k-1})^\top
\]

\[
(\Theta^i_{k-1} - \hat{\Theta}^i_{k-1})^\top I_{k-1}
\]

\[
+ (I_{k-1} - \Omega^i_{k-1}) \hat{\Theta}^i_{k-1} \hat{\Theta}^i_{k-1} + (I_{k-1} - \Omega^i_{k-1})^\top
\]

\[
(\Theta^i_{k-1} - \hat{\Theta}^i_{k-1})^\top I_{k-1}
\]

\[
\text{We can further write}
\]

\[
\Sigma^i_k = \langle v_{2k} v_{2k}^\top + v_{3k} v_{3k}^\top + v_{4k} v_{4k}^\top + v_{5k} v_{5k}^\top | p(I_{k-1}|y_{1:k-1}) \rangle p(\Theta^i_k|y_{1:k-1}, p(\Delta_{k}))
\]

\[
\text{with } v_{2k} = \begin{pmatrix}
(1 - I_{k-1})\tilde{\Theta}^i_{k-1} \\
\vdots \\
(1 - I_{k-1})\tilde{\Theta}^i_{k-1}
\end{pmatrix},
\]

\[
v_{3k} = \begin{pmatrix}
I_{k-1}\Delta^i_{k-1} \\
\vdots \\
I_{k-1}\Delta^i_{k-1}
\end{pmatrix}
\]

\[
v_{4k} = \begin{pmatrix}
I^i_{k-1}(\tilde{\Theta}^i_{k-1} + \hat{\Theta}^i_{k-1}) \\
\vdots \\
I^i_{k-1}(\tilde{\Theta}^i_{k-1} + \hat{\Theta}^i_{k-1})
\end{pmatrix},
\]

\[
v_{5k} = \begin{pmatrix}
I^i_{k-1}(\tilde{\Theta}^i_{k-1} + \hat{\Theta}^i_{k-1}) \\
\vdots \\
I^i_{k-1}(\tilde{\Theta}^i_{k-1} + \hat{\Theta}^i_{k-1})
\end{pmatrix}
\]

\[
\text{and}
\]

\[
\text{for } i = j
\]

\[
A^i_{k-1} = \langle (1 - I_{k-1})^2 | p(I_{k-1}|y_{1:k-1}) \rangle \langle (\Theta^i_{k-1} - \hat{\Theta}^i_{k-1})^2 | p(\Theta^i_k) \rangle
\]

\[
= \Omega^i_{k-1} \Sigma^i_{k-1}
\]
for $i \neq j$

$$A_{k-1}^{ij} = \langle (x_{k-1}^i)^p(x_{k-1}^{i-1}) (I_{k-1}^j) p(x_{k-1}^{i-1}) \rangle_{\Omega_{k-1}}$$

$$= \Omega_{k-1}^{i-1} \Sigma_{k-1}^{i-1}$$

(91)

\[ \therefore \] \[ A_{k-1} = \Sigma_{k-1}^{+} \otimes (\text{diag}(\Omega_{k-1}) \text{diag}(\Omega_{k-1})^\top + \Omega_{k-1}(I - \Omega_{k-1})) \]

(92)

\[ \Sigma_k^{-} = \left( I - \Omega_{k-1} \right) \Sigma_k + \Omega_{k-1} \Sigma_k \]

$$+ \Sigma_k^{-1} \otimes (\text{diag}(\Omega_{k-1}) \text{diag}(\Omega_{k-1})^\top + \Omega_{k-1}(I - \Omega_{k-1}))$$

$$+ \Omega_{k-1}(I - \Omega_{k-1}) \left( \text{diag} \left( \Omega_{k-1}^+ \right) \right)^2$$

(93)

\[ B. \ Derivation \ of \ q(x_k) \]

Using (8) and (11) we can write $q(x_k)$ as

$$q(x_k) \propto \text{exp} \left( \langle \ln(p(y_k|x_k, I_k, \Theta_k)) p(I_k) p(x_k|y_{1:k-1}) \rangle_{q(I_k)} \right)$$

$$p(\Theta_k|y_{1:k}) q(\Theta_k)$$

$$\propto \text{exp} \left( \langle \ln(p(y_k|x_k, I_k, \Theta_k)) q(\Theta_k) \rangle_{q(I_k)} \right)$$

$$p(\Theta_k|y_{1:k}) q(\Theta_k)$$

$$\propto \text{exp} \left( \langle \ln(N(y_k|h(x_k) + I_k \Theta_k, R_k)) \rangle_{q(I_k)} q(\Theta_k) \right)$$

$$\propto \text{exp} \left( \left( -\frac{1}{2} v_{6k}^\top R_k^{-1} v_{6k} - \frac{1}{2} \ln(2\pi)^m |R_k| \right) \right)$$

$$\propto \text{exp} \left( -\frac{1}{2} \left( \text{tr} (v_{6k}^\top \Sigma_k R_k^{-1}) \right) \right)$$

$$\propto \text{exp} \left( -\frac{1}{2} \text{tr} \left( (v_{7k}^\top v_{7k} - v_{7k}^\top v_{8k} + v_{8k}^\top v_{7k}) R_k^{-1} \right) \right)$$

$$\propto \text{exp} \left( -\frac{1}{2} \text{tr} \left( (v_{9k}^\top v_{9k}) R_k^{-1} \right) \right)$$

$$\propto \text{exp} \left( -\frac{1}{2} \text{tr} \left( (v_{10k}^\top v_{10k}) R_k^{-1} \right) \right)$$

$$\propto \text{exp} \left( -\frac{1}{2} \text{tr} \left( (B_k + v_{10k}^\top v_{10k}^\top + v_{11k}^\top v_{11k}^\top) R_k^{-1} \right) \right)$$

(100)

The parameters $x_k^+$ and $P_k^+$ can be updated using (28)–(33).

\[ C. \ Derivation \ of \ q(I_k) \]

Using (9) and (11) we can write $q(I_k)$ as

$$q(I_k) \propto \text{exp} \left( \langle \ln(p(y_k|x_k, I_k, \Theta_k)) p(I_k) p(x_k|y_{1:k-1}) \rangle_{q(x_k)} q(\Theta_k) \right)$$

$$\propto \text{exp} \left( \langle \ln(p(y_k|x_k, I_k, \Theta_k)) q(x_k) q(\Theta_k) \rangle \right)$$

$$\propto \text{exp} \left( \langle \ln(p(y_k|x_k, I_k, \Theta_k)) q(x_k) q(\Theta_k) \rangle \right) \text{p}(I_k)$$

(101)

As we consider $R_k$ to be diagonal we can write

$$q(I_k) \propto \text{exp} \left( \sum_i \frac{1}{2} a_i(a_i - h^i(x_k) + I_k \Theta_i^i)^2 \right)$$

(102)

$$\text{exp} \left( \left( h^i(x_k) - v_i^i + I_k \Theta_i^i \right)^2 \right)$$

(103)

Consequently $q(I_k)$ can be expressed as follows with its parameters updated using (28)–(33).

$$q(I_k) \propto \prod_{i=1}^m \left( 1 - \Omega_i^i \right) \delta(I_k^i) + \Pi_k^i \delta(I_k^i - 1)$$

(104)

\[ D. \ Derivation \ of \ q(\Theta_k) \]

Using (10) and (11) we can write $q(\Theta_k)$ as

$$q(\Theta_k) \propto \text{exp} \left( \langle \ln(p(y_k|x_k, I_k, \Theta_k)) p(I_k) p(x_k|y_{1:k-1}) \rangle_{q(I_k)} \right)$$

$$\propto \text{exp} \left( \langle \ln(p(y_k|x_k, I_k, \Theta_k)) q(I_k) \rangle \right)$$

$$\propto \text{exp} \left( \langle \ln(N(y_k|h(x_k) + I_k \Theta_k, R_k)) \rangle_{q(I_k)} \right)$$

$$\propto \text{exp} \left( \left( -\frac{1}{2} v_{6k}^\top R_k^{-1} v_{6k} - \frac{1}{2} \ln(2\pi)^m |R_k| \right) \right)$$

$$\propto \text{exp} \left( -\frac{1}{2} \text{tr} \left( (v_{7k}^\top v_{7k} - v_{7k}^\top v_{8k} + v_{8k}^\top v_{7k}) R_k^{-1} \right) \right)$$

$$\propto \text{exp} \left( -\frac{1}{2} \text{tr} \left( (v_{9k}^\top v_{9k}) R_k^{-1} \right) \right)$$

$$\propto \text{exp} \left( -\frac{1}{2} \text{tr} \left( (B_k + v_{10k}^\top v_{10k}^\top + v_{11k}^\top v_{11k}^\top) R_k^{-1} \right) \right)$$

$$\propto \text{exp} \left( -\frac{1}{2} \text{tr} \left( (B_k + v_{10k}^\top v_{10k}^\top + v_{11k}^\top v_{11k}^\top) R_k^{-1} \right) \right)$$

(105)

$$\propto \text{exp} \left( \left( h^i(x_k) - v_i^i + I_k \Theta_i^i \right)^2 \right)$$

$$\text{exp} \left( \left( h^i(x_k) - v_i^i + I_k \Theta_i^i \right)^2 \right)$$

(106)

$$\text{exp} \left( \left( h^i(x_k) - v_i^i + I_k \Theta_i^i - y_k^i \right)^2 \right)$$

Consequently $q(I_k)$ can be expressed as follows with its parameters updated using (28)–(33).

$$q(I_k) \propto \prod_{i=1}^m \left( 1 - \Omega_i^i \right) \delta(I_k^i) + \Omega_i^i \delta(I_k^i - 1)$$

(107)

$$\text{exp} \left( \left( h^i(x_k) - v_i^i + I_k \Theta_i^i - y_k^i \right)^2 \right)$$

(108)

$$\text{exp} \left( \left( h^i(x_k) - v_i^i + I_k \Theta_i^i - y_k^i \right)^2 \right)$$

(109)

$$\text{exp} \left( \left( h^i(x_k) - v_i^i + I_k \Theta_i^i - y_k^i \right)^2 \right)$$

(110)

$$\text{exp} \left( \left( h^i(x_k) - v_i^i + I_k \Theta_i^i - y_k^i \right)^2 \right)$$

(111)

$$\text{exp} \left( \left( h^i(x_k) - v_i^i + I_k \Theta_i^i - y_k^i \right)^2 \right)$$

(112)

$$\text{exp} \left( \left( h^i(x_k) - v_i^i + I_k \Theta_i^i - y_k^i \right)^2 \right)$$

(113)

$$\text{exp} \left( \left( h^i(x_k) - v_i^i + I_k \Theta_i^i - y_k^i \right)^2 \right)$$

(114)
\[ q(\Theta_k) \propto \exp \left( -\frac{1}{2} \left( y_k - \hat{y}_k - \left( \hat{\Theta}_k \Sigma_r^{-1} \hat{\Theta}_k^\top - \Omega_k \right) \right)^\top \left( y_k - \hat{y}_k - \left( \hat{\Theta}_k \Sigma_r^{-1} \hat{\Theta}_k^\top - \Omega_k \right) \right) \right) \]

where \( \hat{y}_k = \mathbf{h}(x_k) \), \( \hat{\Theta}_k \) is the Kalman filter estimate, \( \Sigma_r \) is the residual covariance, and \( \Omega_k \) is the innovation covariance. Adding a constant term \( \nu_k = y_k - e_k \) where \( e_k = (\mathbf{h}(x_k) - \hat{y}_k) \), yields the expression for \( q(\Theta_k) \) in diagonal form.

References:

[1] T. Chen, Y. Cao, L. Sun, X. Qing, and J. Zhang, “A distributed robust power system state estimation approach using t-distribution noise model,” IEEE Trans. Syst., vol. 15, no. 1, pp. 1066–1076, Mar. 2021.

[2] T. Joseph, B. Tyagi, and V. Kumar, “Unbiased minimum variance filter-based generator state estimation using PMU measurements for unknown generator input,” IEEE Syst. J., vol. 13, no. 3, pp. 3176–3184, Sep. 2019.

[3] T. D. Barfoot, State Estimation for Robotics. Cambridge, U.K.: Cambridge Univ. Press, 2017.

[4] H. Ahmed and M. Tahir, “Accurate attitude estimation of a moving land vehicle using low-cost MEMS IMU sensors,” IEEE Trans. Intel. Transp. Syst., vol. 18, no. 7, pp. 1723–1739, Jul. 2017.

[5] R. E. Kalman, “A new approach to linear filtering and prediction problems,” 1960.

[6] M. S. Grewal and A. P. Andrews, Kalman Filtering: Theory and Practice With MATLAB. Hoboken, NJ, USA: Wiley, 2014.

[7] S. J. Julier and J. K. Uhlmann, “New extension of the Kalman filter to nonlinear systems,” Proc. SPIE, vol. 3068, 1997, pp. 182–193.

[8] N. J. Gordon, D. J. Salmond, and A. F. Smith, “Novel approach to nonlinear/non-gaussian bayesian state estimation,” IEEE Proc. F-Radar Signal Process., IET, vol. 140, no. 2, pp. 107–113, 1993.

[9] M. Norgaard, N. K. Poulsen, and O. Ravn, “New developments in state estimation for nonlinear systems,” Automatica, vol. 36, no. 11, pp. 1627–1638, 2000.

[10] S. Narasimhan and C. Jordache, Data Reconciliation and Gross Error Detection: An Intelligent Use of Process Data. Amsterdam, The Netherlands: Elsevier, 1999.

[11] A. H. Chughtai, U. Akram, M. Tahir, and M. Uppal, “Dynamic state estimation in the presence of sensor outliers using MAP-based EKF,” IEEE Sens. Lett., vol. 4, no. 4, pp. 1–4, Apr. 2020.

[12] Y. Huang, G. Jia, B. Chen, and Y. Zhang, “A new robust Kalman filter with adaptive estimate of time-varying measurement bias,” IEEE Signal Process. Lett., vol. 27, pp. 700–704, 2020.

[13] D. Belfadil, Y. Bar-Shalom, and P. Willett, “Single space based sensor bias estimation using a single target of opportunity,” IEEE Trans. Aerosp. Electron. Syst., vol. 56, no. 3, pp. 1676–1684, Jun. 2020.

[14] B. Friedland, “Treatment of bias in recursive filtering,” IEEE Trans. Autom. Control, vol. 14, no. 4, pp. 359–367, Aug. 1969.

[15] S. F. Schmidt, “Application of State-Space Methods to Navigation Problems,” in Advances in Control Systems, vol. 3. Amsterdam, The Netherlands: Elsevier, 1966, pp. 293–340.

[16] S. Yousefi, X.-W. Chang, and B. Champagne, “Mobile localization in non-line-of-sight using constrained square-root unscented Kalman filter,” IEEE Trans. Veh. Technol., vol. 64, no. 5, pp. 2071–2083, May 2015.

[17] X. Zhang, W. Xue, and H.-T. Fang, “On extended state based Kalman filter for nonlinear time-varying uncertain systems with measurement bias,” Control Theory Technol., vol. 19, no. 1, pp. 142–152, 2021.

[18] M. Najar, J. Huerta, J. Vidal, and J. Castro, “Mobile location with bias tracking in non-line-of-sight,” in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., 2004, vol. 3, pp. iii–956.

[19] S.-S. Woo, H.-R. You, and J.-S. Koh, “The NLOS mitigation technique for position location using IS-95 CDMA networks,” in Proc. Veh. Technol. Conf. Fall VTC2000. 52nd Veh. Technol. Conf., 2000, vol. 6, pp. 2556–2560.

[20] G. Hu, Z. Zhang, A. Armaou, and Z. Yan, “Robust extended Kalman filter based state estimation for nonlinear dynamic processes with measurements corrupted by gross errors,” J. Taiwan Inst. Chem. Engineers, vol. 106, pp. 29–33, 2020.

[21] M. Wylie and J. Holtzman, “The non-line-of-sight problem in mobile location estimation,” in Proc. 5th Int. Conf. Universal Pers. Commun., 1996, vol. 2, pp. 827–831.

[22] J. Borras, P. Hatrack, and N. Mandayam, “Decision theoretic framework for NLOS identification,” in Proc. 48th IEEE Veh. Technol. Conf. Pathway Glob. Wireless Revolution, 1998, vol. 2, pp. 1583–1587.

[23] G. Andrey and G. Julie, “An unscented Kalman filter based maximum likelihood ratio for NLOS bias detection in UMTS localization,” in Proc. 16th Eur. Signal Process. Conf., 2008, pp. 1–5.

[24] M. F. Abdel-Hafez, “Detection of bias in GPS satellites’ measurements: A probability ratio test formulation,” IEEE Trans. Control Syst. Technol., vol. 22, no. 3, pp. 1166–1173, May 2014.

[25] C. Jiang, J. Shen, S. Chen, Y. Chen, D. Liu, and Y. Bo, “UWB NLOS/LOS classification using deep learning method,” IEEE Commun. Lett., vol. 24, no. 10, pp. 2226–2230, Oct. 2020.

[26] J. Zhu and S. S. Kia, “Bias compensation for UWB ranging for pedes- trian geolocation applications,” IEEE Sens. Lett., vol. 3, no. 9, pp. 1–4, Sep. 2019.

[27] Y.-T. Chan, W.-Y. Tsai, H.-C. So, and P. chung Ching, “Time-of-arrival based localization under NLOS conditions,” IEEE Trans. Veh. Technol., vol. 55, no. 1, pp. 17–24, Jan. 2006.

[28] D. B. Jourdan, J. J. Deyst, M. Z. Win, and N. Roy, “Monte Carlo localization in dense multipath environments using UWB ranging,” in Proc. IEEE Int. Conf. Ultra-Wideband, 2005, pp. 314–319.

[29] J. González et al., “Mobile robot localization based on ultra-wide-band ranging: A particle filter approach,” Robot. Auton. Syst., vol. 57, no. 5, pp. 496–507, 2009.

[30] A. H. Chughtai, M. Tahir, and M. Uppal, “A robust bayesian approach for online filtering in the presence of contaminated observations,” IEEE Trans. Trans. Instrum. Meas., vol. 70, pp. 1–15, 2020.

[31] N. de Freitas, “ Rao-blackwellised particle filtering for fault diagnosis,” in Proc. IEEE Aero. Conf., 2002, vol. 4, pp. 4–4.

[32] A. Banerjee and P. Burlina, “Efficient particle filtering via sparse kernel density estimation,” IEEE Trans. Image Process., vol. 19, no. 9, pp. 2480–2490, Sep. 2010.

[33] H. Wang, H. Li, J. Fang, and H. Wang, “Robust gaussian Kalman filter with outlier detection,” IEEE Signal Process. Lett., vol. 25, no. 8, pp. 1236–1240, Aug. 2018.

[34] R. Piché, S. Särkkä, and J. Hartikainen, “Recurrent outlier-robust filtering and smoothing for nonlinear systems using the multivariate student-t distribution,” in Proc. IEEE Int. Workshop Mach. Learn. Signal Process., 2012, pp. 1–6.

[35] A. Nakabayashi and G. Ueno, “Nonlinear filtering method using a switching error model for outlier-contaminated observations,” IEEE Trans. Auto. Control, vol. 65, no. 7, pp. 3150–3156, Jul. 2020.

[36] A. H. Chughtai, M. Tahir, and M. Uppal, “Outlier-robust filtering for nonlinear systems with selective observations rejection,” IEEE Sensors J., vol. 22, no. 7, pp. 6887–6897, Sep. 2022.

[37] N. Wiener, N. Wiener, C. Mathematician, N. Wiener, N. Wiener, and C. Mathematician, Extrapolation, Interpolation, and Smoothing of Stationary Time Series: With Engineering Applications, vol. 113. Cambridge, MA, USA: MIT Press, 1949, no. 21.
Aamir Hussain Chughtai received the B.Sc. degree in electrical engineering from the University of Engineering and Technology, Lahore, Pakistan, in 2011, and the M.Sc. degree in electrical engineering in 2018 from the Lahore University of Management Sciences, Lahore, Pakistan, where he is currently working toward the Ph.D. degree in electrical engineering. His research interests include robust parameter and state estimation during data corruption, statistical learning and inference problems, with applications in navigation, localization, cyber-physical and intelligent systems.

Muhammad Tahir (Senior Member, IEEE) received the M.S. and Ph.D. degrees in electronics and communication engineering from Politecnico di Torino, Turin, Italy, in 2009 and 2013, respectively. During 2013, he was a Research Associate with the Navigation Signal Analysis and Simulation Group, Istituto Superiore Mario Boella, Italy. In June 2014, he joined Lahore University of Management Sciences, Lahore, Pakistan, where he is currently an Associate Professor with the Department of Electrical Engineering. His research interests include finding connections between traditional model-based signal processing and modern data-based machine learning algorithms and methods, such as robust Bayesian estimation and filtering, developing low-dimensional models for large scale machine learning problems, scalable optimization methods, and exploring innovative data-fusion architectures for real-world sensory data streams.

Momin Uppal (Senior Member, IEEE) received the B.S. degree in electronic engineering with highest distinction from the GI Khan Institute of Engineering Sciences, Topi, Pakistan, in 2002, and the M.S. and Ph.D. degrees in electrical engineering from Texas A&M University, College Station, TX, USA, in 2010 and 2006, respectively. He is currently a tenured Associate Professor and the Chair of electrical engineering with Lahore University of Management Sciences (LUMS), Lahore, Pakistan. His research interests include urban computing, machine learning and signal processing, backscatter communications, and environmental sensing using radio signals. He was the recipient of the HECS Best University Teacher Award 2021 (National award), The LUMS Vice Chancellors Teaching Excellence Award 2021 (University-Wide Award), and the LUMS Department of Electrical Engineering Distinguished Teaching Award 2019. He is also the Principal Investigator of a multidisciplinary PKR 210 million grant from the Higher Education Commission of Pakistan dealing with the use of technology and data for addressing urban issues.

CHUGHTAI et al.: V ARIATIONAL-BASED NONLINEAR BAYESIAN FILTERING WITH BIASED OBSERVATIONS 5307