Variational-Based Non-linear Bayesian Filtering with Observations Subjected to Gross Errors

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 random and gross errors. In this work, we focus on the occurrence of gross errors and propose a robust filter for their detection and mitigation during state estimation of nonlinear dynamical systems. We model the presence of gross errors 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, Gross Errors, 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 non-linear SSMs, the filtering becomes more challenging owing to the underlying functional non-linearities. 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 non-linear 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. The observations can be compromised in a variety of ways and two kinds of data corruption are common namely random and gross errors. Random errors – also known as outliers – behave as zero-mean noise, causing sudden sporadic surges in the measurements. Their occurrence can be attributed to factors like sensor degradation, communication failures, environmental influences, etc [10]. On the other hand, gross errors – also known as biases – manifest in the data statistically as nonzero-mean noise. Multiple factors in various applications lead to the appearance of gross errors. Examples include miscalibrations of sensors, other configuration aberrations like errors in sensor location or alignment, clock errors, or malfunctioning [11]. In addition, environmental inconsistencies can introduce biases in observations e.g. Ultra-wideband (UWB), Global Positioning System (GPS) or Universal Mobile Telecommunications System (UMTS) signals can get affected in non-line-of-sight (NLOS) settings in this manner. Filtering with bias compensation finds applications in diverse fields [12], [13], [14], [15]. In this work, we keep our focus only on the appearance of gross errors in the measurements 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 gross errors 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 [16]. 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) [17]. 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 [15].

The authors are with Department of Electrical Engineering, Lahore University of Management Sciences, DHA Lahore Cantt., 54792, Lahore Pakistan. (email: aamir.chughtai@lums.edu.pk; arslan.majal@lums.edu.pk; tahir@lums.edu.pk; momin.uppal@lums.edu.pk)
These kinds of suboptimal approaches are more useful in scenarios where it is safe to ignore the information regarding the evolution of gross errors. 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 the gross errors [18].

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 [19]. Such information may be available beforehand for some applications. However, it must generally be obtained from some 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 [20], biased measurements are identified simply by comparison of the standard deviation of range measurements with a detection threshold. A similar approach to use normalized residuals to detect the presence of gross errors [21]. The method proposed in [22] 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 [23], [24], [25], [21]. In addition, other methods resort to deep learning for the determination of affected measurements [26]. There are several bias compensating filtering methods that rely on these kinds of separate detectors [12], [15], [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 gross errors. These kinds of models describe gross errors 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 gross errors. 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 random errors [33], [34], [35], [36], however, generally there is a shortage of such methods that deal with gross errors. Recent attempts in this direction include the work in [37] 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 gross errors in the measurements for non-linear systems. We refer to this approach as the bias detecting and mitigating (BDM) filter. We resort to Variational inference for designing the filter with an internal bias detection mechanism. For comparison, we choose other UKF driven approaches for simulations which indicate the merits of our proposed approach.

The way we have organized this article is as follows. Section II describes our modeling choice for incorporation of gross errors 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.

II. Gross Error Modeling

The standard SSM does not consider the possibility of measurement errors in its generative structure [38]. Therefore, it needs to be modified to cater for the occurrence of gross errors. 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:

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

(1)  
(2)
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, $q_{k-1} \in \mathbb{R}^p$ and $r_k \in \mathbb{R}^m$ are white process and measurement noise vectors, $f(\cdot)$ and $h(\cdot)$ represent non-linear process and measurement dynamics respectively, $\Theta_k \in \mathbb{R}^p$ models the effect of biases in the measurements and $T_k \in \mathbb{R}^{m \times m}$ is a diagonal matrix with Bernoulli elements $T_{k_{ij}}$ 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].

$$\Theta_k = (I - T_{k-1})\tilde{\Theta}_k + T_{k-1}(\Theta_{k-1} + \Delta_k)$$

(3)

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 $\sigma^2_\Theta$ (assuming an uninformative prior) described by the respective entries of a zero mean random vector $\tilde{\Theta}_k$. The distributions of $\Delta_k$ and $\tilde{\Theta}_k$ are supposed to be white and normally distributed given as

$$\Delta_k \sim \mathcal{N}(0, \tilde{\Sigma}_k)$$

(4)

$$\tilde{\Theta}_k \sim \mathcal{N}(0, \tilde{\Sigma}_k)$$

(5)

Note that for tractability, we have modified $\tilde{\Theta}_k$ 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 $T_k$ 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 $x_k$, $z_k$ (considering only the random entries $T_{k_i}$) and $\Theta_k$ conditioned on the set of all the observations $y_{1:k}$ analytically as

$$p(x_k, z_k, \Theta_k|y_{1:k}) = \frac{p(y_{k}|x_{k}, z_{k}, \Theta_k)p(x_{k}, z_{k}, \Theta_k|y_{1:k-1})}{p(y_{k}|y_{1:k-1})}$$

(6)

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 [39] for inference where the product of VB marginals is conveniently used to approximate the joint posterior as

$$p(x_k, z_k, \Theta_k|y_{1:k}) \approx q(x_k)q(z_k)q(\Theta_k)$$

(7)

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

$$q(x_k) \propto \exp \left( \ln p(x_k, \Theta_k|y_{1:k}) \right) q(\Theta_k)$$

(8)

$$q(z_k) \propto \exp \left( \ln p(x_k, \Theta_k|y_{1:k}) \right) q(\Theta_k)$$

(9)

$$q(\Theta_k) \propto \exp \left( \ln p(x_k, \Theta_k|y_{1:k}) \right) q(\Theta_k)$$

(10)

where $\langle \cdot \rangle_q(y_k)$ denotes the expectation of the argument with respect to a distribution $q(y_k)$. 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 $p(x_k|y_{1:k}) \approx q(x_k)$, $p(z_k|y_{1:k}) \approx q(z_k)$ and $p(\Theta_k|y_{1:k}) \approx q(\Theta_k)$ where $q(\cdot)$ 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 expressed as

$$p(x_k, z_k, \Theta_k|y_{1:k-1}) = \int \int \int p(x_k, z_k, \Theta_k|x_{k-1}, z_{k-1}, \Theta_{k-1}) p(x_{k-1}, z_{k-1}, \Theta_{k-1}|y_{1:k-1}) p(\Theta_{k-1}|y_{1:k-1}) $$

(11)

$$p(x_k, z_k, \Theta_k|y_{1:k-1}) \approx \int \int \int p(x_k|x_{k-1}) p(x_{k-1}, z_{k-1}, \Theta_{k-1}|y_{1:k-1}) p(\Theta_{k-1}|y_{1:k-1}) $$

(12)

$$p(x_k, z_k, \Theta_k|y_{1:k-1}) \approx p(z_k)p(x_k|y_{1:k-1})p(\Theta_k|y_{1:k-1})$$

(13)

with

$$p(x_k|y_{1:k-1}) = \int p(x_k|x_{k-1}) p(x_{k-1}|y_{1:k-1}) dx_{k-1}$$

(14)

$$p(\Theta_k|y_{1:k-1}) \approx \int p(\Theta_k|z_{k-1}, \Theta_{k-1}) p(z_{k-1}|y_{1:k-1})$$

(15)

$$p(\Theta_k|y_{1:k-1}) \approx p(\Theta_k|y_{1:k-1}) \delta(I_k - 1)$$

(16)

We assume that the occurrence of bias is independent for each dimension and its historical existence. Using $\delta_k$ to denote the prior probability of occurrence of bias in the $i$th observation, the distribution of $T_k$ is defined as product of independent Bernoulli distributions of each element

$$p(T_k) = \prod_{i=1}^m p(I_{k_i}) = \prod_{i=1}^m (1 - \theta_k) \delta(I_{k_i} - 1)$$

(17)

$$p(T_{k-1})p(T_k) = \prod_{i=1}^m p(I_{k-1})p(I_{k_i})$$

(18)

$$p(T_{k-1})p(T_k) = \prod_{i=1}^m (1 - \Omega_{k-1}) \delta(I_{k_i} - 1)$$

(19)
where \( \Omega_k \) denotes the posterior probability of bias occurrence in the \( i \)-th dimension. The notation \( N(x|m, \Sigma) \) 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 \( f(.) \) is assumed to be non-linear, \( p(x_k | y_{1:k-1}) \) can be approximated, using general Gaussian filtering results [40], as \( N(x_k | \hat{x}_k^-, P_k^-) \) with the parameters predicted as follows

\[
\hat{x}_k^- = (f(x_{k-1})) p(x_{k-1} | y_{1:k-1}) \tag{20}
\]

\[
P_k^- = \langle (f(x_{k-1}) - \hat{x}_k^-) (f(x_{k-1}) - \hat{x}_k^-)^T \rangle p(x_{k-1} | y_{1:k-1}) + Q_{k-1} \tag{21}
\]

The remaining term required to approximate the predictive density recursively in (13) is \( p(\Theta_k | y_{k-1}) \). Observing (3) and (15), it is evident that \( p(\Theta_k | y_{k-1}) \) 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 \( N(\hat{\Theta}_k^-, \Sigma_k^-) \) using moment matching [40]. The parameters of the distribution are updated as

\[
\hat{\Theta}_k^- = \Omega_{k-1} \hat{\Theta}_{k-1} \tag{22}
\]

\[
\Sigma_k^- = (I - \Omega_{k-1}) \Sigma_{k-1} + \Omega_{k-1} \Sigma_{k} + \Sigma_{k-1} \odot (\text{diag}(\Omega_{k-1}))^2 \tag{23}
\]

where \( \Omega_{k-1} \) is a diagonal matrix with entries \( \Omega_{k,i} \) denoting the posterior probability of bias occurrence at time step \( k-1 \). The \( \odot \) is the Hadamard product and \( \text{diag} \) is used for vector to diagonal matrix conversion and vice versa. The reader is referred to the Appendix for detailed derivations of (22)-(23).

**Remarks:** We note the following in (22)-(23):

- \( \Omega_{k-1} \) 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 \( 1 \), \( \hat{\Theta}_{k}^- = \hat{\Theta}_{k-1}^- \) 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 \( \Omega_{k}^- \). 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 (13) for approximating the predictive density. For detailed derivations, the reader is referred to the Appendix.

Parameters of \( q(x_k) \) are updated iteratively as:

\[
\hat{x}_k^+ = \hat{x}_k^- + K_k(y_k - \Omega_k \hat{\Theta}_k^- - \mu_k) \tag{24}
\]

\[
\mu_k = \langle h(x_k) \rangle_{p(x_k | y_{1:k-1})} \tag{25}
\]

\[
P_k^+ = P_k^- - C_k K_k^\dagger \tag{26}
\]

\[
K_k = C_k S_k^{-1} \tag{27}
\]

\[
C_k = \langle \langle h(x_k) - \mu_k \rangle h(x_k) - \mu_k \rangle \rangle_{p(x_k | y_{1:k-1})} \tag{28}
\]

\[
S_k = \langle (h(x_k) - \mu_k)^2 \rangle_{p(x_k | y_{1:k-1})} + R_k \tag{29}
\]

Parameters of \( q(\mathcal{I}_k) \) are updated iteratively as:

\[
\Omega_k^i = \text{Pr}(\mathcal{I}_k = 1)/\text{Pr}(\mathcal{I}_k = 1) + \text{Pr}(\mathcal{I}_k = 0) \tag{30}
\]

where denoting \( k_1 \) as the proportionality constant

\[
\text{Pr}(\mathcal{I}_k = 0) = k_1 (1 - \theta_k^i) \exp \left( -\frac{1}{2} \frac{(\nu_k^i - \theta_k^i)^2}{R_k^i + \bar{h}_k^2} \right) \tag{31}
\]

\[
\text{Pr}(\mathcal{I}_k = 1) = k_1 \theta_k^i \exp \left( -\frac{1}{2} \frac{\bar{h}_k^2 + \theta_k^i + \hat{\Theta}_k^- - y_k^i)^2}{R_k^i} \right) \tag{32}
\]

\[
\nu_k = \langle h(x_k) \rangle_{q(x_k)} \tag{33}
\]
\[ \tilde{h}_k^2 = \langle (h'(x_k) - \nu_k^2)^2 \rangle q(\Theta_k) \]
\[ \tilde{\Theta}_k^2 = \langle (\Theta_k^2 - \tilde{\Theta}_k^2)^2 \rangle q(\Theta_k) \]

Parameters of \(q(\Theta_k)\) are updated iteratively as:
\[ \Theta_k^* = \tilde{\Theta}_k + C_k(y_k - (\nu_k + \Omega_k \tilde{\Theta}_k)) \]
\[ \Sigma_k^* = \Sigma_k - C_k P_k^T_k \]
\[ K_k = C_k S_k^{-1} \]
\[ C_k = \Sigma_k^* \Omega_k^T \]
\[ S_k = \Omega_k S_k + R_k \]
\[ \tilde{\Theta}_k^* = (\Omega_k(I - \Omega_k) \Sigma_k^{-1} + \Sigma_k^{-1})^{-1} \]
\[ \Sigma_k^* = (\Omega_k(I - \Omega_k) \Sigma_k^{-1} + \Sigma_k^{-1})^{-1} \]

### C. BDM Filter

Using the proposed approximations, in the prediction and update steps, we have devised a recursive filter referred as the BDM filter. We propose using an uninformative prior for the occurrence of bias in each dimension, i.e. \( \theta_{k_i} = 0.5 \) for all \( i \), unless real-world experiments reveal any statistics regarding their arrival. The usage of uninformative prior has been advocated in the literature for designing robust filters given no prior information about the occurrence of the considered data abnormality is available \([35, 36]\). Algorithm 1 outlines the devised BDM filter.

### 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) \([41]\) and the constrained Unscented Kalman Filter (CUKF). The CUKF is devised by modifying the CSRUKF \([15]\) 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 non-linear 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 is added as a shifted Gaussian \( N(\mu, \sigma) \) where \( \mu \geq 0 \) \([15]\).

We consider the process equation given as
\[ x_k = \begin{pmatrix} 1 & \sin(\omega_k \delta_k) & 0 & \cos(\omega_k \delta_k) & 0 \\ 0 & \cos(\omega_k \delta_k) & 1 & -\sin(\omega_k \delta_k) & 0 \\ 0 & \sin(\omega_k \delta_k) & 0 & \cos(\omega_k \delta_k) & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} x_{k-1} + q_{k-1} \]

where the state vector \( x_k = [a_k, \dot{a}_k, b_k, \dot{b}_k, \omega_k]^T \) contains the 2D position coordinates \( (a_k, b_k) \), the respective velocities \( \dot{a}_k, \dot{b}_k \), the angular velocity \( \omega_k \) of the target at time instant \( k, \delta_k \) is the sampling period, and \( q_{k-1} \sim N(0, Q_{k-1}) \). \( Q_{k-1} \) is given in terms of scaling parameters \( \eta_1 \) and \( \eta_2 \) as \([33]\)
\[ Q_{k-1} = \begin{pmatrix} \eta_1 M & 0 & 0 \\ 0 & \eta_1 M & 0 \\ 0 & 0 & \eta_2 M \end{pmatrix}, M = \begin{pmatrix} \delta_k^3/3 & \delta_k^2/2 & 0 \\ \delta_k^2/2 & \delta_k & 0 \\ 0 & 0 & 1 \end{pmatrix} \]

Range readings are supposed to be obtained from \( m \) sensors installed around a rectangular area where its \( i \)th sensor is located at \( (a^{(i)}, b^{(i)}) = (350(i-1), 350((i-1) \mod 2)) \).

\[ y_k^i = \sqrt{((a_k - a^{(i)})^2 + (b_k - b^{(i)})^2 + r_k^i + o_k^i)} \]

where \( o_k^i \) includes the possibility of presence of bias in the \( i \)th dimension. We assume that given the occurrence of bias \( o_k^i \sim N(\xi^i_k, \sigma_k^i) \) where \( \xi_k^i \) is uniformly distributed between 0 to 90 and \( \sigma_k^i = 0.4 \forall i \). Moreover, we suppose that bias occurs independently in each dimension with probability \( \lambda \) at the specified onset time index for each experimental case. Following values are assigned for evaluation: \( x_0 = [0, 10, 0, -5, \frac{2\pi}{3}]^T, R_k = 4I, \delta_k = 1, \eta_1 = 0.1, \eta_2 = 1.75 \times 10^{-4} \). 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 \( x_0 \) its error covariance as \( Q_k \) and the bias estimate is set equal...
A. Case 1: Consistent bias presence

First, we consider the case where gross errors consistently corrupt the measurements representative of real-world scenarios where the observations are systematically biased. We consider that the bias corrupts each observation according to (44) 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 $\lambda$. We can observe an intuitive trend. 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. The USKF is found to have more comparative 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 one MC run depictive of the general trend in Fig. 1 for large values of $\lambda$ for this case. The BDM-UKF deals with biased observations more effectiveness followed by the partially compensating methods namely the CUKF and the USKF. The SOR-UKF, on the other hand, loses track of the ground truth due to permanent rejection of essential information. Similarly, the standard UKF with its inability to deal with gross errors also exhibits large errors.

| $\lambda$ | UKF | USKF | BDM-UKF | SOR-UKF | CUKF |
|-----------|-----|------|---------|---------|------|
| 0.2       | 0.0469 | 0.0781 | 0.0866   | 0.0821   | 0.5762 |
| 0.4       | 0.0470 | 0.0849 | 0.1000   | 0.1199   | 1.2499 |
| 0.6       | 0.0471 | 0.0883 | 0.1121   | 0.1416   | 1.8012 |
| 0.8       | 0.0470 | 0.0903 | 0.1492   | 0.1637   | 2.2305 |

TABLE I: Average Time for 100 MC Runs with Constant Bias at Different Values of $\lambda$.

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 (44) 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 \( \lambda \). We see a similar pattern as for the previous case owing to the same rationale regarding the functionality of different 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. 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 gross errors in the measurements, 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. Moreover, owing to better utilization of the measurements, the BDM filter has lower estimation errors as compared to the methods with similar KF based approaches, the USKF and the CUKF, aiming to exploit information from the corrupted dimensions. The BDM filter is easier to employ as it avoids the use of external detectors and any 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 \( \mathcal{N}(\hat{\Theta}_k | \hat{\Theta}_k^-, \Sigma_k^-) \)

\[
\hat{\Theta}_k^- = (\Theta_k)p(\Theta_k | y_{1:k-1}) \\
\Theta_k = \langle (I - \mathcal{I}_{k-1})\hat{\Theta}_k + \mathcal{I}_{k-1}(\Theta_k + \Delta_k)p(\Theta_k | y_{1:k-1})p(\Theta_k),p(\Theta_k | y_{1:k-1})p(\Delta_k) \rangle \\
\hat{\Theta}_k = \Omega_{k-1}\hat{\Theta}_k^+ \\
\Sigma_k^- = \langle (\Theta_k - \hat{\Theta}_k^-)(\Theta_k - \hat{\Theta}_k^-)^\top)p(\Theta_k | y_{1:k-1}) \rangle \\
= \langle v_{1k}v_{1k}^\top)p(\Theta_k | y_{1:k-1})p(\Theta_k),p(\Theta_k | y_{1:k-1})p(\Theta_k | y_{1:k-1})p(\Delta_k) \rangle
\]

where

\[
v_{1k} = (I - \mathcal{I}_{k-1})\hat{\Theta}_k + \mathcal{I}_{k-1}(\Theta_k + \Delta_k) - \Omega_{k-1}\hat{\Theta}_k^+ - \mathcal{I}_{k-1}\hat{\Theta}_k^+ + \mathcal{I}_{k-1}\hat{\Theta}_k^+
\]
\[ \mathbf{v}_{1k}^\top \mathbf{v}_{1k} \] can be written as 

\[ \mathbf{v}_{1k}^\top \mathbf{v}_{1k} = ((\mathbf{I} - \mathbf{I}_{k-1})^\top \mathbf{\Theta} \mathbf{\Theta}^\top (\mathbf{I} - \mathbf{I}_{k-1}))^\top + \mathbf{I}_{k-1} \mathbf{\Delta}_k^\top \mathbf{\Delta}_k 
+ \mathbf{I}_{k-1} \mathbf{\Delta}_{k-1} \mathbf{\Delta}_{k-1}^\top 
+ \mathbf{I}_{k-1} \mathbf{\Omega}_{k-1} \mathbf{\Theta} \mathbf{\Theta}^\top \mathbf{\Omega}_{k-1}^\top 
+ \mathbf{I}_{k-1} \mathbf{\Omega}_{k-1} \mathbf{\Omega}_{k-1}^\top 
\]

(51)

\[ = \mathbf{v}_{2k}^\top \mathbf{v}_{2k} + \mathbf{v}_{3k}^\top \mathbf{v}_{3k} + \mathbf{v}_{4k}^\top \mathbf{v}_{4k} + \mathbf{v}_{5k}^\top \mathbf{v}_{5k} \]

(52)

with \( \mathbf{v}_{2k} = \left( \begin{array}{c} (1 - \mathbf{I}_{k-1}) \mathbf{\Theta}_k^\top \\
\vdots \\
(1 - \mathbf{I}_{k-1}) \mathbf{\Theta}_k^\top 
\end{array} \right) \), \( \mathbf{v}_{3k} = \left( \begin{array}{c} \mathbf{I}_{k-1} \mathbf{\Delta}_k \\
\vdots \\
\mathbf{I}_{k-1} \mathbf{\Delta}_k 
\end{array} \right) \)

\[ \mathbf{v}_{4k} = \left( \begin{array}{c} (1 - \mathbf{I}_{k-1}) \mathbf{\Theta}_k^\top \mathbf{\Theta}^\top \\
\vdots \\
(1 - \mathbf{I}_{k-1}) \mathbf{\Theta}_k^\top \mathbf{\Theta}^\top 
\end{array} \right) \), \( \mathbf{v}_{5k} = \left( \begin{array}{c} \mathbf{I}_{k-1} \mathbf{\Omega}_k^\top \\
\vdots \\
\mathbf{I}_{k-1} \mathbf{\Omega}_k^\top 
\end{array} \right) \)

Resultingly, \( \mathbf{\Sigma}_k^- \) can be further expressed as

\[ \mathbf{\Sigma}_k^- = \begin{pmatrix} (1 - \mathbf{I}_{k-1})^2 \mathbf{\Theta}_k^\top & \cdots & \mathbf{\Theta}_k^\top \\
\vdots & \ddots & \vdots \\
(1 - \mathbf{I}_{k-1})^2 \mathbf{\Theta}_k^\top & \cdots & \mathbf{\Theta}_k^\top 
\end{pmatrix} \mathbf{\Delta}_k^\top \mathbf{\Delta}_k 
+ \begin{pmatrix} (1 - \mathbf{I}_{k-1})^2 \mathbf{\Theta}_k^\top & \cdots & \mathbf{\Theta}_k^\top \\
\vdots & \ddots & \vdots \\
(1 - \mathbf{I}_{k-1})^2 \mathbf{\Theta}_k^\top & \cdots & \mathbf{\Theta}_k^\top 
\end{pmatrix} \mathbf{I}_{k-1} \mathbf{\Delta}_k^\top 
+ \begin{pmatrix} (1 - \mathbf{I}_{k-1})^2 \mathbf{\Theta}_k^\top & \cdots & \mathbf{\Theta}_k^\top \\
\vdots & \ddots & \vdots \\
(1 - \mathbf{I}_{k-1})^2 \mathbf{\Theta}_k^\top & \cdots & \mathbf{\Theta}_k^\top 
\end{pmatrix} \mathbf{I}_{k-1} \mathbf{\Omega}_k^\top 
+ \begin{pmatrix} (1 - \mathbf{I}_{k-1})^2 \mathbf{\Theta}_k^\top & \cdots & \mathbf{\Theta}_k^\top \\
\vdots & \ddots & \vdots \\
(1 - \mathbf{I}_{k-1})^2 \mathbf{\Theta}_k^\top & \cdots & \mathbf{\Theta}_k^\top 
\end{pmatrix} \mathbf{I}_{k-1} \mathbf{\Omega}_k^\top 
\]

where \( i \in \{1, m\} \)

(53)

\[ \mathbf{A}_{k-1} = \begin{pmatrix} (1 - \mathbf{I}_{k-1})^2 \mathbf{\Theta}_k^\top & \cdots & \mathbf{\Theta}_k^\top \\
\vdots & \ddots & \vdots \\
(1 - \mathbf{I}_{k-1})^2 \mathbf{\Theta}_k^\top & \cdots & \mathbf{\Theta}_k^\top 
\end{pmatrix} \mathbf{I}_{k-1} \mathbf{\Delta}_k^\top + \mathbf{I}_{k-1} \mathbf{\Omega}_k^\top 
\]

(54)

The parameters \( \mathbf{x}_k^{+} \) and \( \mathbf{P}_k^{+} \) can be updated using (31)-(30).
C. Derivation of \(q(I_k)\)

Using (9) and (13) we can write \(q(I_k)\) as

\[
q(I_k) \propto \exp \left( \left( \ln(p(y_k | x_k, I_k, \Theta_k))p(x_k | y_{1:k-1}) - q(x_k | q(\Theta_k)) \right) \right) \times \\
\exp \left( \left( \ln(p(y_k | x_k, I_k, \Theta_k))q(x_k_q | q(\Theta_k)) \right) \right) p(I_k)
\]

(75)

As we consider \(R_k\) to be diagonal we can write

\[
q(I_k) \propto \exp \left( \sum_i \frac{1}{2R_k^i}a_k \left( \prod_i(1-\theta_k^i)\delta(I_k^i) + \theta_k^i(I_k^i-1) \right) \right)
\]

(77)

where

\[
a_k = \langle (y_k^i - (h^i(x_k) + I_k^i\Theta_k^i))^2 \rangle_q(x_k|q(\Theta_k)) = \langle (h^i(x_k) - \nu_k^i + I_k^i(\Theta_k^i - \hat{\Theta_k}^i) + \nu_k^i + I_k^i\hat{\Theta_k}^i - y_k^i)^2 \rangle_q(x_k|q(\Theta_k))
\]

(78)

(79)

\[
= \hat{h}_k^2 + I_k^2\Theta_k^2 + (\nu_k^i + I_k^i\hat{\Theta_k}^i - y_k^i)^2
\]

(80)

Consequently \(q(I_k)\) can be expressed as follows with its parameters updated using (30)-(35).

\[
q(I_k) \propto \prod_{i=1}^m \left( 1 - \omega_k^i \right) \delta(I_k^i) + \omega_k^i \delta(I_k^i - 1)
\]

(81)

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

Using (10) and (13) we can write \(q(\Theta_k)\) as

\[
q(\Theta_k) \propto \exp \left( \left( \ln(p(y_k | x_k, I_k, \Theta_k))p(x_k | y_{1:k-1}) - q(x_k | q(\Theta_k)) \right) \right) \times \\
\exp \left( \left( \ln(p(y_k | x_k, I_k, \Theta_k))q(x_k_q | q(\Theta_k)) \right) \right) p(\Theta_k)
\]

(82)

\[
\propto \exp \left( \left( \ln(\mathbf{N}(\mathbf{y}_k | \mathbf{h}(x_k) + I_k\Theta_k, \mathbf{R}_k))q(x_k | q(\Theta_k)) \right) \right) \\
\mathbf{N}(\Theta_k | \hat{\Theta}_k, \Sigma_k)
\]

(83)

\[
\propto \exp \left( \left( \mathbf{v}_k^\top \mathbf{R}_k^{-1} \mathbf{v}_k - \frac{1}{2} \ln(2\pi)^m | \mathbf{R}_k | \right)q(x_k | q(\Theta_k)) \right) \\
\mathbf{N}(\Theta_k | \hat{\Theta}_k, \Sigma_k)
\]

(84)

\[
\propto \exp \left( \left( \frac{1}{2} \text{tr}(\mathbf{v}_k^\top \mathbf{R}_k^{-1} \mathbf{v}_k) \right)q(x_k | q(\Theta_k)) \right) \\
\mathbf{N}(\Theta_k | \hat{\Theta}_k, \Sigma_k)
\]

(85)

\[
\propto \exp \left( \frac{1}{2} (\mathbf{v}_k^\top \mathbf{R}_k^{-1} \mathbf{v}_k) \right)q(x_k | q(\Theta_k)) \mathbf{N}(\Theta_k | \hat{\Theta}_k, \Sigma_k)
\]

(86)

where we write \(\mathbf{v}_k\) in a useful form \(\mathbf{v}_k = (I_k - \Omega_k)\Theta_k + \Omega_k\Theta_k + \mathbf{h}(x_k) - \mathbf{y}_k\) for the subsequent derivation of \(q(\Theta_k)\) as

\[
q(\Theta_k) \propto \exp \left( \left( \frac{1}{2} \text{tr}(\mathbf{B}_k^\top \mathbf{R}_k^{-1} \mathbf{B}_k + \mathbf{v}_{10k}^\top \mathbf{R}_k^{-1} \mathbf{v}_{10k} + \mathbf{v}_{11k}^\top \mathbf{R}_k^{-1} \mathbf{v}_{11k}) \right) \right) \\
\mathbf{N}(\Theta_k | \hat{\Theta}_k, \Sigma_k)
\]

(87)

where \(\mathbf{B}_k = \text{diag}(\Theta_k)\Omega_k(1-\Omega_k)\text{diag}(\Theta_k), \mathbf{v}_{10k} = \Omega_k\Theta_k\) and \(\mathbf{v}_{11k} = \nu_k - y_k\) where \(\nu_k = \langle h_k | x_k \rangle \rangle_q(x_k_q).\) Adding a constant term \(\mathbf{v}_{11k}^\top \mathbf{v}_{11k}\) to complete the square in the exponential expression and considering \(\mathbf{R}_k\) to be diagonal yields

\[
q(\Theta_k) \propto \exp \left( - \frac{1}{2} \mathbf{v}_{12k}^\top \mathbf{R}_k^{-1} \mathbf{v}_{12k} \right) \mathbf{N}(\Theta_k | \hat{\Theta}_k, \Sigma_k)
\]

(88)

\[
\propto \mathbf{N}(\Theta_k | \hat{\Theta}_k, \Sigma_k)
\]

(89)

where \(\mathbf{v}_{12k} = y_k - \nu_k - \Omega_k\Theta_k.\) Using general Gaussian filtering results \([40]\), we can update \(\hat{\Theta}_k\) and \(\Sigma_k\) using (36)-(40). The following appears as a result of the product of two multivariate Gaussian distributions \([42]\)

\[
\mathbf{N}(|x_1, \Sigma_1)\mathbf{N}(|x_2, \Sigma_2) \propto \mathbf{N}(|x_1, \Sigma_1, \Sigma_2)
\]

(90)

\[
\Sigma_c = (\Sigma_1^{-1} + \Sigma_2^{-1})^{-1}
\]

(91)

\[
m_c = \Sigma_c(\Sigma_1^{-1}m_1 + \Sigma_2^{-1}m_2)
\]

(92)

Using the above result we can update the parameters \(\hat{\Theta}_k^+\) and \(\Sigma_k^+\) using (41)-(42).

REFERENCES

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

[2] T. Joseph, B. Tyagi, and V. Kumar, “Unbiased minimum variance filterbased generator state estimation using pmu measurements for unknown generator input,” IEEE Systems Journal, vol. 13, no. 3, pp. 3176–3184, 2019.

[3] T. D. Barfoot, State estimation for robotics. Cambridge University Press, 2017.

[4] H. Ahmed and M. Tahir, “Accurate attitude estimation of a moving land vehicle using low-cost imu sensors,” IEEE Transactions on Intelligent Transportation Systems, vol. 18, no. 7, pp. 1723–1739, 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. John Wiley & Sons, 2014.

[7] S. J. Julier and K. J. Uhlmann, “New extension of the kalman filter to nonlinear systems,” in Signal processing, sensor fusion, and target recognition VI, vol. 3068. International Society for Optics and Photonics, 1997, pp. 182–193.

[8] N. J. Gordon, D. J. Salmond, and A. F. Smith, “Novel approach to nonlinear/non-gaussian bayesian state estimation,” in IEE Proceedings F-radar and signal processing, vol. 140, no. 2. IET, 1993, pp. 107–113.

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

[10] A. H. Chughtai, U. Akram, M. Tahir, and M. Uppal, “Dynamic state estimation in the presence of sensor outliers using map-based ekf,” IEEE Sensors Letters, vol. 4, no. 4, pp. 1–4, 2020.

[11] D. Belfadel, Y. Bar-Shalom, and P. Willett, “Single space based sensor bias estimation using a single target of opportunity,” IEEE Transactions on Aerospace and Electronic Systems, vol. 56, no. 3, pp. 1676–1684, 2019.

[12] J. Zhu and S. S. Kia, “Bias compensation for uwb ranging for pedestrian geolocation applications,” IEEE Sensors Letters, vol. 3, no. 9, pp. 1–4, 2019.

[13] S. Zhao, S. R. Duncan, and D. A. Howey, “observability analysis and state estimation of lithium-ion batteries in the presence of sensor biases,” IEEE Transactions on Control Systems Technology, vol. 25, no. 1, pp. 326–333, 2017.

[14] Z. Zhu, Z. Meng, Z. Zhang, J. Chen, and Y. Dai, “Robust particle filter for state estimation using measurements with different types of gross errors,” ISA transactions, vol. 69, pp. 281–295, 2017.
A. H. Chughtai, M. Tahir, and M. Uppal, “Outlier-robust filtering for Y.-T. Chan, W.-Y. Tsui, H.-C. So, and P. Chung Ching, “Time-of-M. F. Abdel-Hafez, “Detection of bias in GPS satellites’ measurements: C. Jiang, J. Shen, S. Chen, Y. Chen, D. Liu, and Y. Bo, “UWB NLOS/LOS A. H. Chughtai, M. Tahir, and M. Uppal, “A robust Bayesian approach J. González, J.-L. Blanco, C. Galindo, A. Ortiz-de Galisteo, J.-A. Fernández-Madrigal, F. A. Moreno, and J. L. Martínez, “Mobile robot local-ization based on ultra-wideband ranging: A particle filter approach,” Robotics and autonomous systems, vol. 57, no. 5, pp. 496–507, 2009. A. H. Chughtai, M. Tahir, and M. Uppal, “A robust Bayesian approach for online filtering in the presence of contaminated observations,” IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1–15, 2020. N. de Freitas, “Rao-blackwellised particle filtering for fault diagnosis,” in Proceedings, IEEE Aerospace Conference, vol. 4, 2002, pp. 4–4. A. Banerjee and P. Burlina, “Efficient particle filtering via sparse kernel density estimation,” IEEE Transactions on Image Processing, vol. 19, no. 9, pp. 2480–2490, 2010. H. Wang, H. Li, J. Fang, and H. Wang, “Robust Gaussian Kalman filter with outlier detection,” IEEE Signal Processing Letters, vol. 25, no. 8, pp. 1236–1240, 2018. R. Piché, S. Särkkä, and J. Hartikainen, “Recursive outlier-robust filtering and smoothing for nonlinear systems using the multivariate Student-t distribution,” in 2012 IEEE International Workshop on Machine Learning for Signal Processing, 2012, pp. 1–6. A. Nakabayashi and G. Ueno, “Nonlinear filtering method using a switching error model for outlier-contaminated observations,” IEEE Transactions on Automatic Control, vol. 65, no. 7, pp. 3150–3156, 2020. A. H. Chughtai, M. Tahir, and M. Uppal, “Outlier-resistant filtering for nonlinear systems with selective observations rejection,” arXiv preprint arXiv:2106.05706, 2021. Y. Huang, G. Jia, B. Chen, and Y. Zhang, “A new robust kalman filter with adaptive estimate of time-varying measurement bias,” IEEE Signal Processing Letters, vol. 27, pp. 700–704, 2020.