Sequential Bayesian Estimation With Censored Data for Multi-Sensor Systems

Yujiao Zheng, *Student Member, IEEE*, Ruixin Niu, *Senior Member, IEEE*, and Pramod K. Varshney, *Fellow, IEEE*

***Abstract*—In this paper, a new framework for sequential Bayesian estimation in sensor networks is proposed, which consists of two processes: censoring of measurements at local sensors and fusion of both received measurements and missing ones at the fusion center (FC). In our scheme, each local sensor maintains a Kalman filter (KF) for a linear Gaussian system or an extended Kalman filter (EKF) for a nonlinear system and the FC runs a particle filter (PF) to track the system state. Informative measurements are selected for transmission by an innovation based per-sensor censoring process executed at the sensors at each time. Though the less informative measurements are not sent to the FC, their absence still conveys some information, and the proposed scheme exploits such information from the missing messages. Numerical results show that, under the same bandwidth constraint, the proposed scheme outperforms the one that ignores missing data information and the one that selects sensors randomly for information transmission.**

***Index Terms*—Sensor censoring, missing data, particle filters, sequential Bayesian estimation, target tracking, sensor networks.**

I. Introduction

In the literature, the sequential Bayesian estimation problem has been mainly investigated for three fundamental network architectures: centralized, distributed and decentralized networks. In a centralized structure, the local sensor nodes transmit either analog [1] or quantized measurements [2]–[4] to a FC, where the sensor data are fused by a Bayesian filter to update the system state estimate in a straightforward manner. If all the analog sensor data are transmitted to the FC, the FC yields the optimal estimation performance, meaning that no other network architecture can deliver a better performance. But a centralized network requires a large amount of communication between the sensors and the FC, and it is vulnerable to the failure of the FC.

In a distributed network, each local sensor node runs a local Bayesian state estimator, and makes its own local state estimate based on its local measurements. These local estimates, or state posterior probability density functions (PDFs), are transmitted to a global FC, where they are fused to get a more accurate global state estimate. The distributed network has reduced communication requirements, since instead of transmitting raw sensor data at the sensor sampling rate, each sensor could transmit state estimates at a much lower rate. Furthermore, the distributed network is much more robust, since each local sensor node maintains its own state estimate. However, one challenging problem for fusion of estimates is that all the local estimates are dependent since all the local filters are estimating the same Markov stochastic process [1]. The problem of distributed Kalman filtering has been investigated in [1], [5]–[10]. For nonlinear filtering in distributed networks, the optimal fusion scheme was developed in [11], [12] which involves the transmission of the local state posterior PDFs to the FC and high dimensional integrals at the FC.

In a decentralized network, each sensor fuses its own local state estimate with information received from its neighboring sensors, and each local sensor communicates only with its neighbors. Due to its diffusive communication strategy, this architecture does not require specialized routing, and in general avoids bottleneck in communications. It is scalable and very robust to single point of failure. However, the implementation of the optimal fusion algorithm, the so-called channel filter [5], [13], [14], is very challenging and existing fusion algorithms in decentralized networks are typically suboptimal approaches. In decentralized networks, estimate consensus among distributed agents has drawn much attention. For linear estimation problems in decentralized networks, algorithms have been proposed to reach a consensus among all the nodes [15]–[18]. For nonlinear problems, efforts have been made to develop consensus particle filtering [19]–[22].

The framework we propose in this paper combines the advantages of both the centralized and distributed networks to achieve communication efficiency, improved estimation performance, and robustness. In this framework, each local sensor node runs its local state estimator, which facilitates censoring of its measurement so that only informative measurements are sent to the FC. Since local state estimation is performed at each local sensor, it is robust against single point of failure. Compared to the centralized network, it has reduced communication rate through sensor censoring. However, different from a typical

Manuscript received April 05, 2013; revised August 18, 2013 and January 08, 2014; accepted March 16, 2014. Date of publication April 01, 2014; date of current version April 24, 2014. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Amir Asif. This work was supported in part by U.S. Air Force Office of Scientific Research under Grant FA9550-10-1-0263, by U.S. Army Research Office under Award W911NF-09-1-0244, and by Southeastern Center for Electrical Engineering Education (SCEEE) under SCEEE Research Initiation Grant Agreement SCEEE-12-002. This work was partly presented in the Statistical Signal Processing Workshop (SSP), 2012.

Y. Zheng and P. K. Varshney are with the Department of EECS, Syracuse University, Syracuse, NY 13244 USA (e-mail: yuzheng@syr.edu; varshney@syr.edu).

R. Niu is with Department of Electrical and Computer Engineering, Virginia Commonwealth University, Richmond, VA 23284 USA (e-mail: rniu@vcu.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2014.2315163

distributed architecture but similar to a centralized architecture, only informative *raw* sensor measurements are sent to the FC in our proposed framework.

As discussed earlier, in a sensor network with a FC, the ideal scenario is for all the sensors to send their observations to the FC for sequential Bayesian estimation. However, due to bandwidth constraints or energy limitations in the network, it is usually desirable to have only a subset of sensors transmit their data at each time step. This gives rise to two interesting problems: 1) In a centralized sensor management framework, for the next time step(s), how does the FC select the subset of sensors which are the most informative based on the accumulated information up to the current time step? 2) In a distributed sensor management system, where each local sensor generates a local estimate based on its local measurements, how does each local sensor determine whether or not its current local measurement, which is already in hand, is informative enough to merit its transmission to the FC?

The first problem is a typical sensor management or sensor selection problem and a lot of effort has been devoted to it by different authors [2], [4], [23]–[31]. For linear and Gaussian filtering problems, since the Kalman filter state covariance matrices can be evaluated offline, one can determine the optimal sensor selection and scheduling strategies offline [25], [27], [28]. For nonlinear filtering problems, efficient sensor selection/management should be performed in an online manner using all the past observation information. In such problems, the informativeness of the sensors could be measured by information theoretic measures, such as entropy and mutual information [24], [29], the posterior Cramér-Rao lower bound (PCRLB) on the mean squared state estimation error [2], [4], [30], or the covariance matrix calculated by the extended Kalman filter (EKF) [31].

The second problem results in the so called censoring method in the area of distributed detection [32]–[35]. In [32], under a constraint on communication, an optimal censoring structure is proposed, through which, local sensors censor their likelihood ratios before sending them to the FC. Only the local likelihood ratio falling in the send region is sent to the FC for making the global decision. Later in [33], the fusion of decisions from censoring sensors transmitted over wireless fading channels was investigated, where optimal and suboptimal fusion rules were designed based on the knowledge of fading channels. Some practical issues on the design of censoring sensor networks including joint dependence of sensor decision rules, randomization of decision strategies, and partially known distributions of observations were further addressed in [34]. Per-sensor censoring scheme was also employed in [35], in which an ordering approach follows censoring to reduce the number of transmissions in the network, and the sensors with more informative observations transmit first. Sensor censoring has also been used to solve estimation problems [36]. The authors in [36] proposed another transmission scheme in which the sensor transmissions are ordered according to the magnitude of their measurements, and the sensors with magnitude smaller than a threshold, do not transmit.

Methods used to solve problems 1) and 2) can be categorized as data selection methods and all of them result in missing data from the viewpoint of the FC. Then, a crucial issue is whether the fact that variables are missing is related to the underlying values of the variables in the data set [37], and this would categorize missing data into three mechanisms according to [37]: i) missing completely at random (MCAR), i.e., missingness does not depend on the data values; ii) missing at random (MAR), i.e., missingness depends only on the observed components, not on the missing ones; iii) not missing at random (NMAR), i.e., missingness depends on the missing values. Obviously, the missing data issue due to the data selection methods such as censoring when solving problem 2) belongs to the third mechanism mentioned above. In this paper, we focus on missing data due to the third mechanism, namely, on NMAR. Since the missing data also convey some information, they can be exploited to obtain better inference. In fact, the information conveyed by missing data due to NMAR has been considered implicitly in the distributed detection problem [32]. The parameter estimation problem that takes into account the NMAR missing data information has been considered in [38]. Nevertheless, to the best of our knowledge, for the Bayesian sequential estimation problem in the context of data selection/sensor censoring, such kind of approach has not yet been explored. A related but different work has been reported in [39] and references therein, which exploits 'negative' sensor evidence (expected but missing sensor data) for target tracking and data fusion. Though the work in [39] is similar to ours, it is different from this paper in two major aspects: first, the missing measurements in [39] are due to the failed attempt by a radar system to detect a target, while in our work certain sensor data are missing because sensors censor their local data in a distributed manner to conserve communication bandwidth and send more informative sensor data to a FC; second, the missing information or 'negative' information in [39] is exploited in terms of fictitious measurements given by appropriate sensor models which is designed based on the background information on the sensor characteristics, while in our work, the missing information is exploited in terms of the statistics of the missingness which can be computed giving the prior knowledge on the censoring rule. Hence, the two novelties of our work are: censoring measurements at local sensors to select informative measurements in a distributed manner, and fusing both received measurements and missing ones at the FC to exploit the information conveyed by the missingness of data. Some preliminary results based on our work were presented in [40], which are extended significantly in this paper.

The main contribution of this paper is that we propose a scheme which provides better performance for target tracking in a sensor network when the bandwidth constraint and/or energy cost at local sensors is important to increase the lifetime of the network. In the proposed scheme, firstly, the local sensors censor their measurements in a distributed manner, and then the FC fuses both the received observations and missing ones. The proposed scheme is shown to be applicable to both linear and nonlinear systems, and both scalar and vector observations. Furthermore, we investigate the relationship between the censoring rule based on the innovation and the one based on the Kullback-Leibler (KL) divergence between the prior state distribution before the measurement is available and the posterior state distribution after the measurement is obtained.

For the convenience of discussion throughout this paper, we call the proposed scheme Censoring and Fusion with Missing Data (CFwMD), since in this scheme, a censoring method is employed at the sensors and the FC fuses data considering the information of missing data that are NMAR. We call the scheme which uses the same censoring method at the sensors but ignores the information about the missing data at the FC as Censoring and Fusion without Missing Data (CFoMD). The scheme, which does not use censoring at the sensor level but a probabilistic transmission strategy, which results in missing data that are MCAR, is called random-selection throughout this paper. Numerical results demonstrate that CFwMD incurs less performance loss compared to the all-send case (all sensors send their measurements to the FC) than CFoMD, while they both outperform the random-selection under the same bandwidth constraint.

The rest of this paper is organized as follows. In the next section, we formulate the problem. Then, we present the proposed CFwMD scheme for linear Gaussian systems when scalar observations are obtained at local sensors in Section III, followed by the discussion on the equivalence between the censoring rule based on the innovation and the one based on the KL divergence in Section IV. Section V discusses the framework when vector observations are available at local sensors, and Section VI generalizes the framework to nonlinear systems. We provide simulation results in Section VII and conclude this paper in Section VIII.

II. Problem Formulation

A. System Model

In this paper, we consider a sequential Bayesian estimation problem in a sensor network with N sensors. Sensors report measurements to the FC for the inference task, i.e., estimation of the system state, for example, the position and velocity of the target in the target tracking problem. Throughout this paper, the channels between local sensors and the FC are assumed to be perfect.

The state model of the system is given as follows:

$$\mathbf{x}_{k+1} = \mathbf{F}_k \mathbf{x}_k + \mathbf{u}_k \tag{1}$$

where $\mathbf{F}_k$ is the state transition matrix, $\mathbf{x}_k$ is the $d \times 1$ state vector and $\mathbf{u}_k$ is the white Gaussion process noise with zero-mean and covariance matrix $\mathbf{Q}_k$. Sensor's measurements are given by

$$\mathbf{z}_k^i = \mathbf{H}^i \mathbf{x}_k + \mathbf{n}_k^i \quad (i = 1, 2, \cdots, N) \tag{2}$$

where $\mathbf{H}^i$ is the observation matrix which maps the state space into the observation space and $\mathbf{n}_k^i$ is white Gaussian measurement noise with zero-mean and covariance $\mathbf{R}^i$. In this paper, we first discuss the case in which scalar observations are obtained at local sensors, i.e.,

$$z_k^i = {\mathbf{h}^i}^T \mathbf{x}_k + n_k^i \quad (i = 1, 2, \cdots, N) \tag{3}$$

where $\mathbf{h}^i$ is the measurement vector, the superscript T denotes vector/matrix transpose and n_k^i is white Gaussian noise with zero-mean and variance r^i.

In our CFwMD scheme, we design a censoring rule which measures the informativeness of the measurements at the sensor level, i.e., at each time step, the ith sensor first examines its measurement according to the designed censoring rule. When the measurement falls in the send region, i.e., it is informative enough, the ith sensor sends it to the FC. Otherwise, it is censored and not sent. For the Bayesian sequential estimation problem, we design the following measurement censoring rule based on the normalized innovation squared (NIS) [41]:

$${\nu_k^i}^T {s_k^i}^{-1} \nu_k^i \begin{cases} \geq \eta_k, & \text{send} \\ < \eta_k, & \text{not send} \end{cases} \tag{4}$$

where $\nu_k^i = z_k^i - {\mathbf{h}^i}^T \hat{\mathbf{x}}_{k|k-1}^i$ is the innovation [41] of the ith sensor at time k, s_k^i is the variance of ν_k^i, given by $s_k^i = r^i + {\mathbf{h}^i}^T \mathbf{P}_{k|k-1}^i \mathbf{h}^i$ in the KF update procedure [41] ($\mathbf{P}_{k|k-1}^i$ is the covariance of the state prediction at the ith sensor), and η_k is a certain threshold that is designed based on performance requirements or bandwidth constraints. Hence, the censoring rule given by (4) implicitly requires that the ith (for $i = 1, \cdots, N$) sensor should perform a KF covariance update at each time, in order to compute the variance of its innovation. Note that (4) is a reasonable way to select informative measurements. One can get an intuition by considering a special case: when sensors are identical, then $s_k^i = s_k^j (i \neq j)$, and a larger magnitude of ν_k^i can pass the censoring threshold more easily. This indicates that the measurement that gives a larger magnitude of ν_k^i is more informative, since a larger magnitude of ν_k^i means larger difference between the measurements and the prediction.

At time k, the complete measurement vector is $\mathbf{z}_k^{1:N} \triangleq (z_k^1, z_k^2, \cdots, z_k^N) \triangleq (\mathbf{z}_k^{\text{obs}}, \mathbf{z}_k^{\text{mis}})$, where $\mathbf{z}_k^{\text{obs}}$ denotes the observed values at the FC and $\mathbf{z}_k^{\text{mis}}$ denotes the missing values. For the NMAR problem induced by (4), we define a missing-data indicator vector $\mathbf{m}_k = (m_k^1, m_k^2, \cdots, m_k^N)$ for $\mathbf{z}_k^{1:N}$, where m_k^i is the indicator variable for the ith sensor, which takes value 1 if the measurement is sent to the FC and 0 otherwise. That is,

$$m_k^i = \begin{cases} 1, & \text{if sensor } i \text{ sends } z_k^i \text{ to FC at time k;} \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

Under the assumption that the channels between the local sensors and the FC are perfect, a missing sensor measurement means that it has been censored by the corresponding sensor node. Hence, $\mathbf{m}_k$, which contains the information on missingness, is available at the FC, and the actual observed data at the FC consist of $(\mathbf{z}_k^{\text{obs}}, \mathbf{m}_k)$. In order to exploit the information conveyed by the missing data, the corresponding likelihood function of the underlying state of the system, which is denoted as $p(\mathbf{z}_k^{\text{obs}}, \mathbf{m}_k | \mathbf{x}_k)$ should be computed by the FC, and how to compute it will be considered in Section III.B.

B. Particle Filter at the FC

In the proposed CFwMD scheme, a PF is adopted at the FC. The KF is known to provide the optimal solution to the Bayesian sequential estimation problem when the system is linear and Gaussian. An EKF can provide suboptimal estimation by linearizing the nonlinear state dynamics and/or nonlinear measurement equation locally in nonlinear systems. However, even for linear and Gaussian systems, when the sensor measurements are quantized, the EKF does not perform very well [42]. The censoring process defined in (4) can be treated as a special case of measurement quantization, since if the measurement falls in the send region, a continuous value is sent; otherwise, no data are sent, which is equivalent to a quantization of the sensor data to the symbol "0". Hence, the PF is a reasonable choice at the FC for Bayesian sequential estimation.

As we know, the main idea of the PF is to represent the posterior distribution $p(\mathbf{x}_k|\mathbf{Z}_{1:k})$ by a set of particles $\{\mathbf{x}_k^l\}$ with associated weights $\{w_k^l\}$. Let N_p denote the total number of particles used in the PF. The posterior distribution can be then approximated as [43]

$$p(\mathbf{x}_k|\mathbf{Z}_{1:k}) \approx \sum_{l=1}^{N_p} w_k^l \delta\left(\mathbf{x}_k - \mathbf{x}_k^l\right) \tag{6}$$

The missing data information can be exploited by using the full likelihood $p(\mathbf{z}_k^{\text{obs}}, \mathbf{m}_k|\mathbf{x}_k)$ instead of the simple likelihood $p(\mathbf{z}_k^{\text{obs}}|\mathbf{x}_k)$ to update the weights of particles at time k. Hence, in the CFwMD scheme, after the FC has received all the measurements sent by local sensors at time k, it computes the full likelihood and uses it to update the particle weights.

C. Censoring Threshold Design

The threshold η_k in (4) is designed such that on an average, l sensors send their measurements to the FC at time k. Thus, we have

$$E\left[\sum_{i=1}^{N} m_k^i\right] = \sum_{i=1}^{N} E\left(m_k^i\right) = l$$

where

$$\begin{aligned} E\left(m_k^i\right) &= p\left(m_k^i = 1\right) \\ &\overset{(a)}{=} p\left({\nu_k^i}^T {s_k^i}^{-1} \nu_k^i \geq \eta_k\right) \end{aligned} \tag{7}$$

and (a) is due to the definition of m_k^i (5).

Since $\nu_k^i \sim \mathcal{N}(0, s_k^i)$, we have ${\nu_k^i}^T {s_k^i}^{-1} \nu_k^i \sim \chi^2_{n_{\nu_k^i}}$, the chi-square distribution with degree of freedom $n_{\nu_k^i}$, and $n_{\nu_k^i}$ is the dimension of the innovation ν_k^i. Since scalar observations are obtained at local sensors, their innovations have the same dimension n_ν, which is equal to 1. Hence, $\sum_{i=1}^N E[m_k^i] = Np({\nu_k^i}^T {s_k^i}^{-1} \nu_k^i \geq \eta_k) = l$, which implies $p({\nu_k^i}^T {s_k^i}^{-1} \nu_k^i \geq \eta_k) = l/N$. Then, we can obtain $\eta_k = \chi^2_{n_\nu}(l/N)$, where $\chi^2_{n_\nu}(l/N)$ represents the critical value such that the probability greater than it is equal to l/N. Note that η_k completely depends on the rate of transmission l/N at time k and the dimension of the innovation n_ν. Hence, once l is set to be the same value for any given time, η_k remains constant over the entire duration of tracking, and it can be computed offline and independently by local sensors and the FC without extra transmission, i.e.,

$$\eta = \chi^2_{n_\nu}(l/N) \tag{8}$$

III. Censoring and Fusion With Missing Data

A. Overview

The proposed CFwMD scheme consists of two major procedures: censoring and fusion, the former is executed at each local sensor while the latter is executed at the FC. At the initial step, local sensors and the FC compute η independently according to (8). Then, at any given time k, each local sensor updates the covariance of its innovation S_k^i following the covariance update of the standard KF, and then determines whether its measurement at the current time is informative enough or not by the proposed innovation based censoring rule (4). Only if the measurement is informative, it is sent to the FC. At the FC, after it gathers all the informative measurements from the local sensors, it fuses them to infer the target state. In this paper, it is assumed that the delays in transmitting sensor measurements to the FC are all less than the sampling interval of the sensors, so that the FC can fuse the arriving measurements in time. We also assume that the FC knows the censoring rule. Since the channels in the system have been assumed to be perfect, the only cause of a missing measurement is that it is not informative enough. Then, based on the two assumption above, the FC can compute the statistics of the missing measurements, which we propose to incorporate in the fusion procedure for better inference performance. Note that the FC maintains a particle filter to track the target. In order to fuse both the received measurements and missing ones, we propose to use the full likelihood function, the details of which will be given in the following section, to update the particle weights.

To make the CFwMD scheme more clear to the readers, we describe one cycle of the scheme in the following algorithm:

Algorithm 1: The CFwMD scheme

Initial step: Design η by (8)

At time k:

At the ith local sensor, $(i = 1, \cdots, N)$:

(A1.1) $s_k^i = r^i + {\mathbf{h}^i}^T \mathbf{P}_{k|k-1}^i \mathbf{h}^i$ (KF update)

(A1.2) Apply the censoring rule (4) to measurement z_k^i

At the FC: (PF with N_s particles, $l = 1, \cdots, N_s$)

(A1.3) $\mathbf{x}_k^l = \mathbf{F}_{k-1}\mathbf{x}_{k-1}^l + \mathbf{u}_k^l$ (Propagating particles)

(A1.4) $w_k^l \propto$ full likelihood function

(A1.5) Normalize weights and estimate the state by $\{\mathbf{x}_k^l, w_k^l\}$

(A1.6) Resampling to get $\{\mathbf{x}_k^l, N_s^{-1}\}$

B. The Full Likelihood Function

One of the critical elements of our CFwMD scheme is the full likelihood function which includes the missing data information according to the previous section. In this section, we derive the full likelihood function at time k for two cases, i.e., for a feedback system as well as for a non-feedback system, depending on whether the state prediction $\hat{\mathbf{x}}_{k|k-1}$ is a global one or a local one.

1) Feedback System: The system is called a feedback system when at the beginning of time k, certain global information, such as prediction of the target state $\hat{\mathbf{x}}_{k|k-1}$, is broadcast to the local sensors by the FC.

Proposition 1: For the linear Gaussian system (1) with measurement model (3), if censoring strategy (4) is used and the state prediction $\hat{\mathbf{x}}_{k|k-1}$ is fed back from the FC to the local sensors, then the full likelihood of the system state at time k, which is used to update the weights of particles at the FC at step (A1.4) in Algorithm 1 of the CFwMD scheme is given as

$$p\left(\mathbf{z}_k^{\text{obs}}, \mathbf{m}_k | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}\right) = \prod_{i=1}^{N} \left[p\left(z_k^i|\mathbf{x}_k\right)\right]^{m_k^i} \times \left[Q\left(\xi_{k,1}^i\right) - Q\left(\xi_{k,2}^i\right)\right]^{1-m_k^i} \tag{9}$$

where $Q(\cdot)$ is the complementary cumulative distribution function of a normal random variable with zero mean and unit variance, $\xi_{k,1}^i \triangleq \frac{-\sqrt{\eta s_k^i} - \mu_k^i}{\sqrt{r^i}}$, $\xi_{k,2}^i \triangleq \frac{\sqrt{\eta s_k^i} - \mu_k^i}{\sqrt{r^i}}$, $\mu_k^i = \mathbf{h}^{i^T}(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})$, the conditional mean of i^{th} sensor's innovation, and m_k^i is defined in (5).

Proof: At time k, given $\hat{\mathbf{x}}_{k|k-1}$, the full likelihood function is $p(\mathbf{z}_k^{\text{obs}}, \mathbf{m}_k|\mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1})$. Let N_{obs} denote the number of received observations, and N_{mis} denote the number of missing observations, then

$$\begin{aligned} p&\left(\mathbf{m}_k, \mathbf{z}_k^{\text{obs}}|\mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}\right) \\ &= \int p\left(\mathbf{m}_k, \mathbf{z}_k^{\text{obs}}, \mathbf{z}_k^{\text{mis}}|\mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}\right) d\mathbf{z}_k^{\text{mis}} \\ &= \int p\left(\mathbf{m}_k|\mathbf{z}_k^{\text{obs}}, \mathbf{z}_k^{\text{mis}}, \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}\right) \\ &\quad \cdot p\left(\mathbf{z}_k^{\text{obs}}, \mathbf{z}_k^{\text{mis}}|\mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}\right) d\mathbf{z}_k^{\text{mis}} \\ &= \int \prod_{i=1}^{N} \left[p\left(m_k^i|z_k^i, \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}\right) p\left(z_k^i|\mathbf{x}_k\right)\right] d\mathbf{z}_k^{\text{mis}} \end{aligned} \tag{10}$$

The last line in (10) is due to the fact that local sensor observations are conditionally independent.

By decomposing the product inside the integral in (10) into two parts: one related to the received observations, and the other related to the missing observations, we can obtain

$$\begin{aligned} p&\left(\mathbf{m}_k, \mathbf{z}_k^{\text{obs}}|\mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}\right) \\ &= \prod_{i=1}^{N_{\text{obs}}} \left[p\left(m_k^i = 1|z_k^i, \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}\right) p\left(z_k^i|\mathbf{x}_k\right)\right] \\ &\quad \cdot \int \prod_{j=1}^{N_{\text{mis}}} p\left(m_k^j, z_k^j|\mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}\right) d\mathbf{z}_k^{\text{mis}} \\ &= \prod_{i=1}^{N_{\text{obs}}} \left[p\left(m_k^i = 1|z_k^i, \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}\right) p\left(z_k^i|\mathbf{x}_k\right)\right] \\ &\quad \cdot \prod_{j=1}^{N_{\text{mis}}} p\left(m_k^j = 0|\mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}\right) \end{aligned} \tag{11}$$

Obviously, $p(m_k^i = 1|z_k^i, \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}) = 1$, and

$$\begin{aligned} p&\left(m_k^j = 0|\mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}\right) \\ &= p\left(\nu_k^{j^T} {s_k^j}^{-1} \nu_k^j < \eta \middle| \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}\right) \\ &\overset{(b)}{=} p\left(|\nu_k^j| < \sqrt{\eta s_k^j} \middle| \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}\right) \end{aligned} \tag{12}$$

where (b) is due to the fact that scalar observations are obtained at local sensors. Given $\mathbf{x}_k$ and $\hat{\mathbf{x}}_{k|k-1}$, ν_k^j is Gaussian with mean

$$\begin{aligned} E&\left[\nu_k^j|\mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}\right] \\ &= E\left[\mathbf{h}^{j^T}\mathbf{x}_k + n_k^j - \mathbf{h}^{j^T}\hat{\mathbf{x}}_{k|k-1}|\mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}\right] \\ &= \mathbf{h}^{j^T}(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}) \\ &\triangleq \mu_k^j \end{aligned} \tag{13}$$

and covariance

$$\text{Var}\left[\nu_k^j|\mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}\right] = r^j$$

Hence,

$$\begin{aligned} p&\left(m_k^j = 0|\mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}\right) \\ &= Q\left(\frac{-\sqrt{\eta s_k^j} - \mu_k^j}{\sqrt{r^j}}\right) - Q\left(\frac{\sqrt{\eta s_k^j} - \mu_k^j}{\sqrt{r^j}}\right) \end{aligned} \tag{14}$$

where η is given by (8). Thus, we can obtain (9) by plugging (14) in (11). ■

Remark 1: (I) We assume that the FC knows each local sensor's measurement model and it maintains a KF covariance update for each local sensor, and, therefore, the full likelihood given by (9) is completely computable at the FC without extra transmission from the local sensors. (II) It is not necessary for each sensor to run a complete KF, including the state update and the covariance update. But, at each sensor, the KF covariance update recursion is still needed to calculate its innovation covariance s_k^i, which is required to censor its measurement. (III) The threshold is designed by assuming that local state predictions are employed to calculate the innovations, but in the feedback system, the innovations are obtained by using the global state prediction $\hat{\mathbf{x}}_{k|k-1}$ fed back by the FC. This implies that the communication rate constraint specified in (7) may not be strictly satisfied in a feedback system, which can be understood by checking the definition of innovation and its covariance right below (4). One can see that, in a feedback system, since the innovation is computed by the global $\hat{\mathbf{x}}_{k|k-1}$ instead of the local estimate $\hat{\mathbf{x}}_{k|k-1}^i$, it is not strictly Gaussian with covariance s_k^i which is still computed by using local $\mathbf{P}_{k|k-1}^i$. Therefore, (7) is not strictly true which indicates that the communication rate constraint is not strictly satisfied. Nevertheless, if the FC also feeds back $\hat{\mathbf{P}}_{k|k-1}$ which is an empirical estimate by the PF, then the bandwidth constraint can be more strictly satisfied with the cost of extra transmission, which gives us Proposition 2.

Proposition 2: For the linear Gaussian system (1) with measurement model (3), if censoring strategy (4) is used and the state prediction $\hat{\mathbf{x}}_{k|k-1}$ and the related covariance $\hat{\mathbf{P}}_{k|k-1}$ are fed back from the FC to the local sensors, then the full likelihood of the system state at time k is given as

$$\begin{aligned} p&\left(\mathbf{z}_k^{\text{obs}}, \mathbf{m}_k|\mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{P}}_{k|k-1}\right) \\ &= \prod_{i=1}^{N} \left[p\left(z_k^i|\mathbf{x}_k\right)\right]^{m_k^i} \left[Q\left(\xi_{k,1}'^i\right) - Q\left(\xi_{k,2}'^i\right)\right]^{1-m_k^i} \end{aligned} \tag{15}$$

where $\xi'^{i}_{k,1} \triangleq \frac{-\sqrt{\eta s'^{i}_{k}}-\mu^{i}_{k}}{\sqrt{r^{i}}}$, $\xi'^{i}_{k,2} \triangleq \frac{\sqrt{\eta s'^{i}_{k}}-\mu^{i}_{k}}{\sqrt{r^{i}}}$, $\mu^{i}_{k} = \mathbf{h}^{iT}(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})$, the conditional mean of ith sensor's innovation, and m^{i}_{k} is defined in (5).

Proof: The result can be obtained by following similar procedures as in the proof of Proposition 1, and we skip the details for brevity. ∎

Remark 2: (I) The superscript '$\prime$' in Proposition 2 indicates that the global state prediction covariance $\hat{\mathbf{P}}_{k|k-1}$ instead of the local $\mathbf{P}^{i}_{k|k-1}$ is involved in the computation of the covariance of the innovation. (II) Since the global $\hat{\mathbf{P}}_{k|k-1}$ in the Proposition is an empirical estimate, (12) through (14) involved in the proof are approximate ones. One should keep in mind that, for the feedback system, a feedback step should be added at the beginning of the CFwMD scheme given in Algorithm 1. If only the state prediction is fed back, the remaining parts remain unchanged; if both the state prediction and related covariance are fed back, $\mathbf{P}^{i}_{k|k-1}$ at step (A1.1) should be replaced by the global state prediction covariance $\hat{\mathbf{P}}_{k|k-1}$. Thus, we do not repeat the algorithm here for brevity.

2) Non-Feedback System: In a non-feedback system, local sensors censor their measurements according to (4) using the innovations computed by their own system state prediction, which implies that each local sensor needs to run a KF. The full likelihood in the non-feedback system is derived and given as follows.

Proposition 3: For the linear Gaussian system (1) with measurement model (3), if censoring strategy (4) is used, then the full likelihood of the target state at time k is given as

$$p\left(\mathbf{m}_k, \mathbf{z}^{\mathrm{obs}}_k|\mathbf{x}_k\right) = \int p\left(\hat{\mathbf{x}}^{1:N}_{k|k-1}|\mathbf{x}_k\right)\prod_{i=1}^{N}\left[p\left(z^{i}_{k}|\mathbf{x}_k\right)\right]^{m^{i}_{k}} \cdot \left[Q\left(\tilde{\xi}^{i}_{k,1}\right) - Q\left(\tilde{\xi}^{i}_{k,2}\right)\right]^{1-m^{i}_{k}} d\hat{\mathbf{x}}^{1:N}_{k|k-1} \tag{16}$$

where $\tilde{\xi}^{i}_{k,1} \triangleq \frac{-\sqrt{\eta s^{i}_{k}}-\tilde{\mu}^{i}_{k}}{\sqrt{r^{i}}}$, $\tilde{\xi}^{i}_{k,2} \triangleq \frac{\sqrt{\eta s^{i}_{k}}-\tilde{\mu}^{i}_{k}}{\sqrt{r^{i}}}$, and $\tilde{\mu}^{i}_{k} = \mathbf{h}^{iT}(\mathbf{x}_k - \hat{\mathbf{x}}^{i}_{k|k-1})$, which is the conditional mean of ith sensor's innovation. m^{i}_{k} is defined in (5) and $p(\hat{\mathbf{x}}^{1:N}_{k|k-1}|\mathbf{x}_k) = p(\hat{\mathbf{x}}^{1}_{k|k-1}, \cdots, \hat{\mathbf{x}}^{N}_{k|k-1}|\mathbf{x}_k)$ is the joint PDF of the local sensor state predictions given the current true state, which will be given later in the paper.

Proof: Let $\hat{\mathbf{x}}^{1:N}_{k|k-1} \triangleq \left(\hat{\mathbf{x}}^{1}_{k|k-1}, \cdots, \hat{\mathbf{x}}^{N}_{k|k-1}\right)$ denote the local sensors' state predictions.

$$\begin{aligned}
&p(\mathbf{m}_k, \mathbf{z}^{\mathrm{obs}}_k|\mathbf{x}_k) \\
&= \int\int p\left(\mathbf{m}_k, \mathbf{z}^{\mathrm{obs}}_k, \mathbf{z}^{\mathrm{mis}}_k, \hat{\mathbf{x}}^{1:N}_{k|k-1}|\mathbf{x}_k\right) d\mathbf{z}^{\mathrm{mis}}_k d\hat{\mathbf{x}}^{1:N}_{k|k-1} \\
&= \int\int p\left(\mathbf{m}_k|\mathbf{z}^{\mathrm{obs}}_k, \mathbf{z}^{\mathrm{mis}}_k, \hat{\mathbf{x}}^{1:N}_{k|k-1}, \mathbf{x}_k\right) \\
&\quad \cdot p\left(\mathbf{z}^{\mathrm{obs}}_k, \mathbf{z}^{\mathrm{mis}}_k|\hat{\mathbf{x}}^{1:N}_{k|k-1}, \mathbf{x}_k\right) p\left(\hat{\mathbf{x}}^{1:N}_{k|k-1}|\mathbf{x}_k\right) d\mathbf{z}^{\mathrm{mis}}_k d\hat{\mathbf{x}}^{1:N}_{k|k-1} \\
&= \int\int \prod_{i=1}^{N}\left\{p\left(m^{i}_{k}|z^{i}_{k}, \hat{\mathbf{x}}^{i}_{k|k=1}, \mathbf{x}_k\right) p\left(z^{i}_{k}|\mathbf{x}_k\right)\right\} \\
&\quad \cdot p\left(\hat{\mathbf{x}}^{1:N}_{k|k-1}|\mathbf{x}_k\right) d\mathbf{z}^{\mathrm{mis}}_k d\hat{\mathbf{x}}^{1:N}_{k|k-1}
\end{aligned} \tag{17}$$

Similar to the feedback case, we split observed data and missing data in the inner integral, then,

$$\begin{aligned}
&p\left(\mathbf{m}_k, \mathbf{z}^{\mathrm{obs}}_k|\mathbf{x}_k\right) \\
&= \int \prod_{i=1}^{N_{\mathrm{obs}}}\left[p\left(z^{i}_{k}|\mathbf{x}_k\right) p\left(m^{i}_{k} = 1|z^{i}_{k}, \hat{\mathbf{x}}^{i}_{k|k-1}, \mathbf{x}_k\right)\right] \\
&\quad \cdot \int \prod_{j=1}^{N_{\mathrm{mis}}}\left[p\left(m^{j}_{k}, z^{j}_{k}|\hat{\mathbf{x}}^{j}_{k|k-1}, \mathbf{x}_k\right)\right] \\
&\quad \cdot p\left(\hat{\mathbf{x}}^{1:N}_{k|k-1}|\mathbf{x}_k\right) d\mathbf{z}^{\mathrm{mis}}_k d\hat{\mathbf{x}}^{1:N}_{k|k-1} \\
&= \int \prod_{i=1}^{N_{\mathrm{obs}}}\left[p\left(z^{i}_{k}|\mathbf{x}_k\right) p\left(m^{i}_{k} = 1|z^{i}_{k}, \hat{\mathbf{x}}^{i}_{k|k-1}, \mathbf{x}_k\right)\right] \\
&\quad \cdot \prod_{j=1}^{N_{\mathrm{mis}}}\left[p\left(m^{j}_{k} = 0|\hat{\mathbf{x}}^{j}_{k|k-1}, \mathbf{x}_k\right)\right] \\
&\quad \cdot p\left(\hat{\mathbf{x}}^{1:N}_{k|k-1}|\mathbf{x}_k\right) d\hat{\mathbf{x}}^{1:N}_{k|k-1}
\end{aligned} \tag{18}$$

Again, we have $p(m^{i}_{k} = 1|z^{i}_{k}, \hat{\mathbf{x}}^{i}_{k|k-1}, \mathbf{x}_k) = 1$ in (18). Now, we compute $p(m^{j}_{k} = 0|\hat{\mathbf{x}}^{j}_{k|k-1}, \mathbf{x}_k)$ in (18) by following a similar procedure as for the feedback system:

$$\begin{aligned}
&E\left[\nu\left(z^{j}_{k}, \hat{\mathbf{x}}^{j}_{k|k-1}\right)|\mathbf{x}_k, \hat{\mathbf{x}}^{j}_{k|k-1}\right] \\
&= \mathbf{h}^{jT}\left(\mathbf{x}_k - \hat{\mathbf{x}}^{j}_{k|k-1}\right) \\
&\triangleq \tilde{\mu}^{j}_{k}
\end{aligned} \tag{19}$$

and

$$\mathrm{Var}\left[\nu\left(z^{j}_{k}, \hat{\mathbf{x}}^{j}_{k|k-1}\right)|\mathbf{x}_k, \hat{\mathbf{x}}^{j}_{k|k-1}\right] = r^{j}$$

Thus,

$$\begin{aligned}
&p\left(m^{j}_{k}|\hat{\mathbf{x}}^{j}_{k|k-1}, \mathbf{x}_k\right) \\
&= Q\left(\frac{-\sqrt{\eta s^{j}_{k}} - \tilde{\mu}^{j}_{k}}{\sqrt{r^{j}}}\right) - Q\left(\frac{\sqrt{\eta s^{j}_{k}} - \tilde{\mu}^{j}_{k}}{\sqrt{r^{j}}}\right)
\end{aligned} \tag{20}$$

Hence, we can obtain (16) by using (20) in (18). ∎

Note that the joint PDF $p(\hat{\mathbf{x}}^{1}_{k|k-1}, \cdots, \hat{\mathbf{x}}^{N}_{k|k-1}|\mathbf{x}_k)$ is a multivariate normal distribution with mean $\boldsymbol{\pi}_{k|k-1}$ and covariance $\boldsymbol{\Sigma}_{k|k-1}$, where $\boldsymbol{\pi}_{k|k-1} = \tilde{\mathbf{H}}\mathbf{x}_k$, $\tilde{\mathbf{H}}$ is given by $\tilde{\mathbf{H}} = [\mathbf{I}_d \cdots \mathbf{I}_d]^T$ with dimension $Nd \times d$, and $\mathbf{I}_d$ denotes the $d \times d$ identity matrix. That is, the mean $\boldsymbol{\pi}_{k|k-1}$ is the concatenation by N true states $\mathbf{x}_k$. The N diagonal elements of the covariance $\boldsymbol{\Sigma}_{k|k-1}$ are filled with $\mathbf{P}^{i}_{k|k-1}$, the covariance of each sensor's own prediction, and the remaining terms of $\boldsymbol{\Sigma}_{k|k-1}$ are filled with $\tilde{\mathbf{P}}^{i,j}_{k|k-1}$, cross-covariance between the ith sensor's prediction and the jth sensor's. Thus,

$$\boldsymbol{\Sigma}_{k|k-1} = \begin{bmatrix} \mathbf{P}^{1}_{k|k-1} & \tilde{\mathbf{P}}^{1,2}_{k|k-1} & \cdots & \tilde{\mathbf{P}}^{1,N}_{k|k-1} \\ \tilde{\mathbf{P}}^{2,1}_{k|k-1} & \mathbf{P}^{2}_{k|k-1} & \cdots & \tilde{\mathbf{P}}^{2,N}_{k|k-1} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{\mathbf{P}}^{N,1}_{k|k-1} & \tilde{\mathbf{P}}^{N,2}_{k|k-1} & \cdots & \mathbf{P}^{N}_{k|k-1} \end{bmatrix} \tag{21}$$

For two arbitrary sensors i, j:

$$\begin{aligned}\tilde{\mathbf{P}}^{i,j}_{k|k-1} &= E\left[\left(\mathbf{x}_k - \hat{\mathbf{x}}^i_{k|k-1}\right)\left(\mathbf{x}_k - \hat{\mathbf{x}}^j_{k|k-1}\right)^T\right] \\ &= E\left[\mathbf{F}\left(\mathbf{x}_{k-1} - \hat{\mathbf{x}}^i_{k-1|k-1}\right)\left(\mathbf{x}_{k-1} - \hat{\mathbf{x}}^j_{k-1|k-1}\right)^T\mathbf{F}^T\right] \\ &\quad + E\left[\mathbf{u}_{k-1}\mathbf{u}^T_{k-1}\right] + E\left[\mathbf{F}\left(\mathbf{x}_{k-1} - \hat{\mathbf{x}}^i_{k-1|k-1}\right)\mathbf{u}^T_{k-1}\right] \\ &\quad + E\left[\mathbf{u}_{k-1}\left(\mathbf{F}\left(\mathbf{x}_{k-1} - \hat{\mathbf{x}}^j_{k-1|k-1}\right)\right)^T\right] \\ &= \mathbf{F}\mathbf{P}^{i,j}_{k-1|k-1}\mathbf{F}^T + \mathbf{Q}_{k-1} \end{aligned} \tag{22}$$

where according to [1]

$$\begin{aligned}\mathbf{P}^{i,j}_{k-1|k-1} &= \left[I - \mathbf{w}^i_{k-1}\mathbf{h}^{i^T}\right] \\ &\quad \times \left[\mathbf{F}\mathbf{P}^{i,j}_{k-2|k-2}\mathbf{F}^T + \mathbf{Q}_{k-2}\right]\left[I - \mathbf{w}^j_{k-1}\mathbf{h}^{j^T}\right]^T, \end{aligned} \tag{23}$$

and $\mathbf{w}^i_{k-1}$ is the Kalman gain at time $k-1$.

Note that (23) is recursive, and once the initialization $\mathbf{P}^{i,j}_{0|0} = \mathbf{0}$ is given, $\mathbf{P}^{i,j}_{k|k}$ at any given time step k can be computed recursively, based on which (21) can be evaluated.

We should point out that, for the non-feedback system, a KF state update should be added to step (A1.1) in Algorithm 1, but the remaining steps are kept the same.

It should be noted that the CFoMD follows the same procedure as the CFwMD, except that the full likelihood is replaced by the simple likelihood, i.e., $p(\mathbf{z}^{\text{obs}}_k|\mathbf{x}_k)$ in step (A1.4) of Algorithm 1.

IV. Censoring Based on an Information Theoretic Metric

In the previous sections, we proposed to use innovations in the censoring rule to select informative measurements. Though we have given an intuitive motivation for this choice, one may wonder about its optimality. In this section, we use an information theory based metric to measure the informativeness of measurements. A good metric which can measure whether or not a measurement z_k is informative enough is the KL divergence between the prior distribution $p(\mathbf{x}_k|z_{1:k-1})$ before the measurement is available and the posterior distribution $p(\mathbf{x}_k|z_{1:k})$ after the measurement is obtained. The censoring rule based on KL divergence can be expressed as

$$D_{\text{KL}}(p(\mathbf{x}_k|z_{1:k-1})||p(\mathbf{x}_k|z_{1:k}))\begin{cases}\geq \zeta_k & \text{send}\\ < \zeta_k & \text{not send}\end{cases} \tag{24}$$

where $D_{\text{KL}}(\cdot||\cdot)$ denotes the distance between two distributions in terms of KL divergence, which is defined as

$$D_{\text{KL}}(p(y)||q(y)) = \int p(y)\ln\frac{p(y)}{q(y)}dy$$

for distributions p and q of the continuous random variable y.

We show that under certain conditions, the proposed innovation based censoring rule is equivalent to that based on the KL divergence.

Theorem 1: For the linear Gaussian system (1) with measurement model (3), if scalar measurements are acquired, then the censoring rule based on the metric $D_{\text{KL}}(p(\mathbf{x}_k|z_{1:k-1})||p(\mathbf{x}_k|z_{1:k}))$ in (24) is equivalent to the one based on the NIS ${\nu_k}^2/s_k$.

Proof: For a linear Gaussian system, we have $p(\mathbf{x}_k|z_{1:k-1}) = \mathcal{N}(\hat{\mathbf{x}}_{k|k-1}, P_{k|k-1})$, and $p(\mathbf{x}_k|z_{1:k}) = \mathcal{N}(\hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k})$. Then, according to [44]

$$\begin{aligned} &D_{\text{KL}}(p(\mathbf{x}_k|z_{1:k-1})||p(\mathbf{x}_k|z_{1:k})) \\ &\quad = \frac{1}{2}\left\{\text{tr}\left(\mathbf{P}^{-1}_{k|k}\mathbf{P}_{k|k-1}\right) - ln\frac{|\mathbf{P}_{k|k-1}|}{|\mathbf{P}_{k|k}|} - d\right\} \\ &\quad\quad + \frac{1}{2}\left(\hat{\mathbf{x}}_{k|k} - \hat{\mathbf{x}}_{k|k-1}\right)^T\mathbf{P}^{-1}_{k|k}(\hat{\mathbf{x}}_{k|k} - \hat{\mathbf{x}}_{k|k-1}) \end{aligned} \tag{25}$$

Since $\mathbf{P}_{k|k-1}$ and $\mathbf{P}_{k|k}$ are determined offline for a linear Gaussian system, and d in (25) is the dimension of the state $\mathbf{x}_k$, they are all deterministic once the system is determined. Therefore, (24) is equivalent to

$$(\hat{\mathbf{x}}_{k|k} - \hat{\mathbf{x}}_{k|k-1})^T\mathbf{P}^{-1}_{k|k}(\hat{\mathbf{x}}_{k|k} - \hat{\mathbf{x}}_{k|k-1})\begin{cases}\geq \gamma_k & \text{send}\\ < \gamma_k & \text{not send}\end{cases} \tag{26}$$

where $\gamma_k \triangleq 2\zeta_k + d + ln\frac{|\mathbf{P}_{k|k-1}|}{|\mathbf{P}_{k|k}|} - \text{tr}(\mathbf{P}^{-1}_{k|k}\mathbf{P}_{k|k-1})$. Note that the censoring is performed at each local sensor which maintains a KF. Thus,

$$\begin{aligned}\hat{\mathbf{x}}_{k|k} &= \hat{\mathbf{x}}_{k|k-1} + \mathbf{w}_k(z_k - \mathbf{h}^T\hat{\mathbf{x}}_{k|k-1}) \\ &= \hat{\mathbf{x}}_{k|k-1} + \mathbf{w}_k\nu_k \end{aligned} \tag{27}$$

where $\mathbf{w}_k$ is the KF gain, which is a column vector if scalar measurements are obtained. Then,

$$(\hat{\mathbf{x}}_{k|k} - \hat{\mathbf{x}}_{k|k-1}) = \mathbf{w}_k\nu_k \tag{28}$$

Thus, (26) is equivalent to

$$\nu^T_k\mathbf{w}^T_k\mathbf{P}^{-1}_{k|k}\mathbf{w}_k\nu_k\begin{cases}\geq \gamma_k & \text{send}\\ < \gamma_k & \text{not send}\end{cases} \tag{29}$$

When scalar measurements are obtained, both $\mathbf{w}^T_k\mathbf{P}^{-1}_{k|k}\mathbf{w}_k$ and s_k are scalars. Hence, by comparing (29) to (4), we conclude that they are equivalent when appropriate thresholds are selected. ■

Theorem 1 indicates that the innovation based censoring rule selects more informative measurements to send, which is intuitively pleasing. The above result can be easily extended to symmetric KL divergence.

Corollary 2: For the linear Gaussian system (1) with measurement model (3), if scalar measurements are obtained, then the censoring rule based on the symmetric KL divergence

$$\begin{aligned} &D_{\text{KL}}(p(\mathbf{x}_k|z_{1:k-1})||p(\mathbf{x}_k|z_{1:k})) \\ &\quad\quad + D_{\text{KL}}(p(\mathbf{x}_k|z_{1:k})||p(\mathbf{x}_k|z_{1:k-1})) \end{aligned} \tag{30}$$

is equivalent to that based on the NIS ${\nu_k}^2/s_k$.

Proof: See Appendix A. ■

V. The Vector Observation Case

So far, our discussion was limited to the scalar observation case. When vector observations are obtained at the local sensors, i.e., the measurement model (2) is used, we still propose to use NIS based censoring rule, i.e.,

$$\boldsymbol{\nu}_k^{i\,T}\mathbf{S}_k^{i\,-1}\boldsymbol{\nu}_k^i \begin{cases} \geq \tilde{\eta}_k, & \text{send} \\ < \tilde{\eta}_k, & \text{not send} \end{cases} \tag{31}$$

Again, we use m_k^i as the indicator variable for the ith sensor, which takes the value 1 if the vector measurement of sensor i is sent to the FC and 0 otherwise.

As in the scalar measurement case, we design $\tilde{\eta}_k$ such that, at time k, there are only l sensors that are active. Without loss of generality, we assume that local sensors' innovations have the same dimension $n_{\boldsymbol{\nu}}$. If l is set to be the same value at any given time and the dimension of the innovation $n_{\boldsymbol{\nu}}$ remains unchanged over time, i.e., the measurement model (2) remains unchanged, then we still have

$$\tilde{\eta}_k = \tilde{\eta} = \chi^2_{n_{\boldsymbol{\nu}}}(l/N) \tag{32}$$

According to the discussion above, Algorithm 1 can be straightforwardly applied to the vector observation case by replacing $s_k^i = r^i + \mathbf{h}^{i^T}\mathbf{P}^i_{k|k-1}\mathbf{h}^i$ by $\mathbf{S}_k^i = \mathbf{R}^i + \mathbf{H}^i\mathbf{P}^i_{k|k-1}\mathbf{H}^{i^T}$. Then, the main concern now is to compute the corresponding full likelihood for the vector observation case which are discussed in the following sub-sections.

A. Feedback System

Proposition 4: For the linear Gaussian system (1) with vector measurement (2), when the global state estimate feedback from the FC is available, and the censoring strategy (31) is used, the full likelihood of the system state at time k is given as

$$p\left(\mathbf{z}_k^{\text{obs}}, \mathbf{m}_k|\mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}\right) = \prod_{i=1}^{N}\left[p\left(\mathbf{z}_k^i|\mathbf{x}_k\right)\right]^{m_k^i}\left[p\left(q^{i^T}q^i < \tilde{\eta}\right)\right]^{1-m_k^i} \tag{33}$$

where $q^i \sim \mathcal{N}(\mathbf{S}_k^{i\,-\frac{1}{2}}\mathbf{H}^i(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}), \mathbf{S}_k^{i\,-\frac{1}{2}}\mathbf{R}^i\mathbf{S}_k^{i\,-\frac{1}{2}})$.

Proof: Following a similar procedure as in Proposition 1, we can obtain

$$\begin{aligned} p\left(\mathbf{m}_k, \mathbf{z}_k^{\text{obs}}|\mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}\right) &= \prod_{i=1}^{N_{\text{obs}}}\left[p\left(m_k^i = 1|\mathbf{z}_k^i, \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}\right)p\left(\mathbf{z}_k^i|\mathbf{x}_k\right)\right] \\ &\quad \cdot \prod_{j=1}^{N_{\text{mis}}} p\left(m_k^j = 0|\mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}\right) \end{aligned} \tag{34}$$

where

$$p\left(m_k^j = 0|\mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}\right) = p\left(\boldsymbol{\nu}_k^{j\,T}\mathbf{S}_k^{j\,-1}\boldsymbol{\nu}_k^j < \tilde{\eta}|\mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}\right)$$

Denoting $q^j \triangleq \mathbf{S}_k^{j\,-\frac{1}{2}}\boldsymbol{\nu}_k^j$, we have

$$\begin{aligned} E[q^j|\mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}] &= \mathbf{S}_k^{j\,-\frac{1}{2}}E\left[\boldsymbol{\nu}_k^j|\mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}\right] \\ &= \mathbf{S}_k^{j\,-\frac{1}{2}}\mathbf{H}^j(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}) \end{aligned} \tag{35}$$

$$\begin{aligned} &\text{Cov}\left[q^j|\mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}\right] \\ &= E\left[q^j q^{j^T}|\mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}\right] \\ &\quad - E\left[q^j|\mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}\right]E^T\left[q^j|\mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}\right] \\ &= E\left[\mathbf{S}_k^{j\,-\frac{1}{2}}\boldsymbol{\nu}_k^j\boldsymbol{\nu}_k^{j\,T}\mathbf{S}_k^{j\,-\frac{1}{2}}|\mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}\right] \\ &\quad - E\left[q^j|\mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}\right]E^T\left[q^j|\mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}\right] \end{aligned} \tag{36}$$

Since

$$\begin{aligned} E\left[\boldsymbol{\nu}_k^j\boldsymbol{\nu}_k^{j\,T}|\mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}\right] &= \text{Cov}\left[\boldsymbol{\nu}_k^j|\mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}\right] \\ &\quad + E\left[\boldsymbol{\nu}_k^j|\mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}\right]E^T\left[\boldsymbol{\nu}_k^j|\mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}\right] \\ &= \mathbf{R}^j + \mathbf{H}^j(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})^T\mathbf{H}^{j^T} \end{aligned} \tag{37}$$

we can obtain

$$\text{Cov}[q^j|\mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}] = \mathbf{S}_k^{j\,-\frac{1}{2}}\mathbf{R}^j\mathbf{S}_k^{j\,-\frac{1}{2}} \tag{38}$$

Therefore, $q^j \sim \mathcal{N}(\mathbf{S}_k^{j\,-\frac{1}{2}}\mathbf{H}^j(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}), \mathbf{S}_k^{j\,-\frac{1}{2}}\mathbf{R}^j\mathbf{S}_k^{j\,-\frac{1}{2}})$.■

Following a similar discussion as that in Remark 1 (III), we provide the following result.

Proposition 5: For the linear Gaussian system (1) with vector measurements (2), when the global state prediction $\hat{\mathbf{x}}_{k|k-1}$ and its covariance $\hat{\mathbf{P}}_{k|k-1}$ are fed back from the FC to the sensors, the full likelihood of the system state at time k is given as

$$p\left(\mathbf{z}_k^{\text{obs}}, \mathbf{m}_k|\mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{P}}_{k|k-1}\right) = \prod_{i=1}^{N}\left[p\left(\mathbf{z}_k^i|\mathbf{x}_k\right)\right]^{m_k^i}\left[p\left(q'^{i^T}q'^i < \tilde{\eta}\right)\right]^{1-m_k^i} \tag{39}$$

where $q'^i \sim \mathcal{N}(\mathbf{S}_k'^{i\,-\frac{1}{2}}\mathbf{H}^i(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}), \mathbf{S}_k'^{i\,-\frac{1}{2}}\mathbf{R}^i\mathbf{S}_k'^{i\,-\frac{1}{2}})$, and $\mathbf{S}_k'^i$ is computed using the global state prediction covariance $\hat{\mathbf{P}}_{k|k-1}$ instead of the local one.

Proof: The result can be obtained by following a similar procedure as in the proof of Proposition 4, and we skip the details for brevity. ■

B. Non-Feedback System

Proposition 6: For the linear Gaussian system (1) with vector measurements, when global estimate feedback from the FC is not available, if censoring strategy (31) is used, then the full likelihood of the system state at time k is given as

$$\begin{aligned} p\left(\mathbf{m}_k, \mathbf{z}_k^{\text{obs}}|\mathbf{x}_k\right) &= \int p\left(\hat{\mathbf{x}}^{1:N}_{k|k-1}|\mathbf{x}_k\right) \\ &\times \prod_{i=1}^{N}\left[p\left(\mathbf{z}_k^i|\mathbf{x}_k\right)\right]^{m_k^i}\left[p\left(t^{i^T}t^i < \tilde{\eta}_k\right)\right]^{1-m_k^i} d\hat{\mathbf{x}}^{1:N}_{k|k-1} \end{aligned} \tag{40}$$

where $t^i \sim \mathcal{N}(\mathbf{S}_k^{i\,-\frac{1}{2}}\mathbf{H}^i(\mathbf{x}_k - \hat{\mathbf{x}}^i_{k|k-1}), \mathbf{S}_k^{i\,-\frac{1}{2}}\mathbf{R}^i\mathbf{S}_k^{i\,-\frac{1}{2}})$.

Proof: The result can be obtained in a straightforward manner following a similar procedure as in Proposition 3 and Proposition 4. ■

VI. Censoring and Fusion With Missing Data for Nonlinear Systems

In the previous sections, we have discussed the proposed CFwMD scheme for linear Gaussian systems. To make it more general, we extend the scheme to a general nonlinear system in this section. Consider the following nonlinear state-space model

$$\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k) + \mathbf{u}_k \tag{41}$$

and measurement model for the ith sensor

$$z_k^i = h^i(\mathbf{x}_k) + n_k^i \tag{42}$$

where $\mathbf{u}_k \sim \mathcal{N}(0, \mathbf{Q})$ is the state process noise, and $n_k^i \sim \mathcal{N}(0, r^i)$ is the measurement noise. We first consider the scalar observation case. Note that, due to the nonlinearity of the system, when the CFwMD scheme is used in the considered nonlinear system, each local sensor maintains an EKF and the FC uses a particle filter to infer the target state. We should point out that the nonlinearity of the system makes it different from the linear Gaussian system in several aspects:

1) The innovation $\breve{\nu}_k^i$ is no longer exactly distributed as Gaussian with zero mean and variance s_k^i, but can be approximated as $\mathcal{N}(0, s_k^i)$.
2) Since $s_k^i = \mathbf{g}_k^i \mathbf{P}_{k|k-1}^i {\mathbf{g}_k^i}^T + r_k^i$, where $\mathbf{g}_k^i = \frac{\partial h}{\partial \mathbf{x}}|_{\hat{\mathbf{x}}_{k|k-1}^i}$, $\mathbf{P}_{k|k-1}^i = \mathbf{F}_{k-1}^i \mathbf{P}_{k-1|k-1}^i \mathbf{F}_{k-1}^{iT} + \mathbf{Q}$, and $\mathbf{F}_{k-1}^i = \frac{\partial \mathbf{f}}{\partial \mathbf{x}}|_{\hat{\mathbf{x}}_{k-1|k-1}^i}$, s_k^i cannot be evaluated offline as in the case of linear systems.

Inspired by the linear Gaussian system we have discussed earlier, we propose that, for a nonlinear system, the ith sensor again censors its measurement based on the NIS, i.e., $(\breve{\nu}_k^i)^T {s_k^i}^{-1} \breve{\nu}_k^i$ at time k, where $\breve{\nu}_k^i \triangleq z_k^i - \hat{z}_{k|k-1}^i$, and it is approximated as a Gaussian distribution with zero mean and covariance s_k^i.

The censoring threshold can also be designed by the bandwidth constraint as in linear Gaussian systems, given the approximation that $\breve{\nu}_k^i \sim \mathcal{N}(0, s_k^i)$. Following a similar procedure as in Section II.C, we have

$$\breve{\eta} = \chi_{n_{\breve{\nu}}}^2 (l/N) \tag{43}$$

where $n_{\breve{\nu}} = 1$, since scalar observations are obtained.

For the considered nonlinear system, if the global state estimate is fed back from the FC to the local sensors, the full likelihood function in the CFwMD scheme is provided in the following proposition.

Proposition 7: For a general nonlinear system given by (41)–(42), if innovation based censoring strategy is used with threshold given by (43) and the global estimate of the state $\hat{\mathbf{x}}_{k-1|k-1}$ is fed back to the local sensors, then the full likelihood of the system state at time k of the CFwMD scheme is given as

$$p\left(\mathbf{z}_k^{\text{obs}}, \mathbf{m}_k | \mathbf{x}_k, \hat{\mathbf{x}}_{k-1|k-1}\right) = \prod_{i=1}^N \left[p\left(z_k^i|\mathbf{x}_k\right)\right]^{m_k^i} \times \left[Q\left(\breve{\xi}_{k,1}^i\right) - Q\left(\breve{\xi}_{k,2}^i\right)\right]^{1-m_k^i} \tag{44}$$

where $\breve{\xi}_{k,1}^i \triangleq \frac{-\sqrt{\breve{\eta} s_k^i} - \breve{\mu}_k^i}{\sqrt{r^i}}$, $\breve{\xi}_{k,2}^i \triangleq \frac{\sqrt{\breve{\eta} s_k^i} - \breve{\mu}_k^i}{\sqrt{r^i}}$, m_k^i is defined in (5), $\breve{\mu}_k^i = h^j(\mathbf{x}_k) - h(\hat{\mathbf{x}}_{k|k-1})$, the conditional mean of ith sensor's innovation, and $\hat{\mathbf{x}}_{k|k-1} = \mathbf{f}(\hat{\mathbf{x}}_{k-1|k-1})$.

Proof: Following a procedure similar to that in Proposition 1, we can obtain (44) in a straightforward manner. ∎

Remark 3: (I) As in the linear Gaussian system, we assume that the FC knows each local sensor's measurement model and it performs an EKF covariance update for each local sensor. Note that an EKF is also maintained at each local sensor, and each local sensor computes the linearized state transition matrix F_k^i and measurement matrix (vector) $\mathbf{g}_k^i$ using the global state estimate $\hat{\mathbf{x}}_{k-1|k-1}$ fed back from the FC. Also, since the local sensors using the global feedback $\hat{\mathbf{x}}_{k-1|k-1}$ in its censoring process, and the FC maintains an EKF covariance update for each local sensor, the FC is able to compute s_k^i involved in $\breve{\xi}_{k,1}^i$ and $\breve{\xi}_{k,2}^i$ in the proposition above, and therefore, (44) is completely computable by the FC without requiring extra information from local sensors. (II) In addition to the state estimate $\hat{\mathbf{x}}_{k-1|k-1}$, the FC can also feed back the covariance $\hat{\mathbf{P}}_{k-1|k-1}$ to local sensors as in the linear Gaussian system. Note that, due to the nonlinearity, $\breve{\nu}_k^i$ is approximated as Gaussian distributed $\mathcal{N}(0, s_k^i)$. Nevertheless, if the FC also feeds back the global covariance $\hat{\mathbf{P}}_{k-1|k-1}$, and then, $\breve{\nu}_k^i$ can be approximated as $\mathcal{N}(0, s_k'^i)$ (the global $\hat{\mathbf{P}}_{k-1|k-1}$ contributes to the computation of $s_k'^i$), which is more accurate than the previous approximation.

Proposition 8: For a general nonlinear system given by (41)–(42), if innovation based censoring strategy is used with the threshold given by (43) and both the global estimate of the state $\hat{\mathbf{x}}_{k-1|k-1}$ and the related covariance $\hat{\mathbf{P}}_{k-1|k-1}$ are fed back to local sensors, then the full likelihood of the system state at time k of the CFwMD scheme is given as

$$p\left(\mathbf{z}_k^{\text{obs}}, \mathbf{m}_k | \mathbf{x}_k, \hat{\mathbf{x}}_{k-1|k-1}, \hat{\mathbf{P}}_{k-1|k-1}\right) = \prod_{i=1}^N \left[p\left(z_k^i|\mathbf{x}_k\right)\right]^{m_k^i} \left[Q\left(\breve{\xi}_{k,1}'^i\right) - Q\left(\breve{\xi}_{k,2}'^i\right)\right]^{1-m_k^i} \tag{45}$$

where $\breve{\xi}_{k,1}'^i \triangleq \frac{-\sqrt{\breve{\eta} s_k'^i} - \breve{\mu}_k^i}{\sqrt{r^i}}$, $\breve{\xi}_{k,2}'^i \triangleq \frac{\sqrt{\breve{\eta} s_k'^i} - \breve{\mu}_k^i}{\sqrt{r^i}}$, m_k^i is defined in (5), $\breve{\mu}_k^i = h^j(\mathbf{x}_k) - h(\hat{\mathbf{x}}_{k|k-1})$, the conditional mean of ith sensor's innovation, and $\hat{\mathbf{x}}_{k|k-1} = \mathbf{f}(\hat{\mathbf{x}}_{k-1|k-1})$.

Proof: Following a procedure similar to that in Proposition 1 and the discussion **Remark 1** (III), we can obtain (45) in a straightforward manner. ∎

Remark 4: (I) The global $\hat{\mathbf{P}}_{k-1|k-1}$ contributes to the computation of $s_k'^i$. (II) If vector observations are obtained by local sensors, one can follow a similar procedure as in Section V to get the corresponding full likelihood for the nonlinear system with feedback (feedback consists of state estimate with/without covariance), which is not provided here for brevity. (III) For the considered nonlinear system without feedback, one may expect to get a similar result as in Proposition 3. But, this is not true. The reason is as follows: consider the joint PDF $p(\hat{\mathbf{x}}_{k|k-1}^{1:N}|\mathbf{x}_k)$ in the nonlinear system. Let us approximate it as Gaussian with mean $\boldsymbol{\pi}_{k|k-1}$ and covariance $\check{\boldsymbol{\Sigma}}_{k|k-1}$, which has the same structure as (21). However, it can be easily found that the diagonal element $\mathbf{P}_{k|k-1}^i$ in $\check{\boldsymbol{\Sigma}}_{k|k-1}$ depends on the state estimate $\hat{\mathbf{x}}_{k-1|k-1}^i$, and the off-diagonal element $\tilde{\mathbf{P}}^{i,j}_{k|k-1}$ depends on the state estimate $\hat{\mathbf{x}}^{i}_{k-1|k-1}$ and $\hat{\mathbf{x}}^{j}_{k-1|k-1}$, which prevents us from obtaining a similar result to that in Proposition 3.

VII. Simulation Results

In this section, we show the advantage of the proposed CFwMD scheme for both linear and nonlinear systems via simulation. For linear systems, we show that, for a certain threshold, the CFwMD scheme achieves less performance loss than CFoMD, while saving the same amount of communication resources compared to the all-send case. We also show that among the three schemes, i.e., CFwMD, CFoMD and the random-selection method, the proposed CFwMD scheme performs the best, under the same bandwidth constraint. We explore the performance comparison for both feedback and non-feedback scenarios. For nonlinear systems, the advantage of the proposed CFwMD scheme over the CFoMD and random-selection schemes is shown by simulations when feedback is included in the system.

A. Linear System—The Scalar Observation Case

A one-dimensional target tracking system is considered in this scenario, with state vector $\mathbf{x}_k = [x_k \;\; \dot{x}_k]^T$, state transition matrix

$$\mathbf{F} = \begin{bmatrix} 1 & \mathcal{D} \\ 0 & 1 \end{bmatrix}$$

and observation matrix $\mathbf{h}^T = [1 \;\; 0]$, where $\mathcal{D} = 1$ second, which is the sampling interval. Without loss of generality, in this example, we use N identical sensors to track the target which moves only along the x-axis following the white noise acceleration model. The state process noise covariance is set as

$$\mathbf{Q} = \sigma^2 \begin{bmatrix} \mathcal{D}^4/4 & \mathcal{D}^3/2 \\ \mathcal{D}^3/2 & \mathcal{D}^2 \end{bmatrix}$$

where $\sigma^2 = 4$. The measurement noise variance is set as $r^i = 1$ for $i = 1, 2$. The initial state of the target $\mathbf{x}_0$ is chosen to be $[0 \;\; 10]$. We observe the target for 20 seconds, namely, we track the target over $T_S = 20$ time steps for each Monte-Carlo trial. The number of particles used in the particle filter at the FC is $N_s = 10^3$.

1) Feedback System: In this example, at the beginning of each time step in a trial, the FC broadcasts the global state prediction to local sensors. We compare the RMSEs, averaged over 5000 Monte-Carlo trials at each time, for the random-selection, CFwMD, CFoMD and all-send cases. To perform the comparison under the same bandwidth constraint, we set the censoring threshold η for the CFwMD and CFoMD schemes at the value such that the average number of active sensors is $l = 1$ at any given time, and we let each sensor send its measurement to the FC with a probability $1/N$ for the random-selection scheme.

In Fig. 1, there are $N = 2$ sensors. Since we set the censoring threshold η to constrain the average number of active sensors as $l = 1$ at any given time, both the CFwMD and CFoMD schemes save 50% transmissions, compared to the all-send case. However, the CFoMD incurs a larger performance loss according to

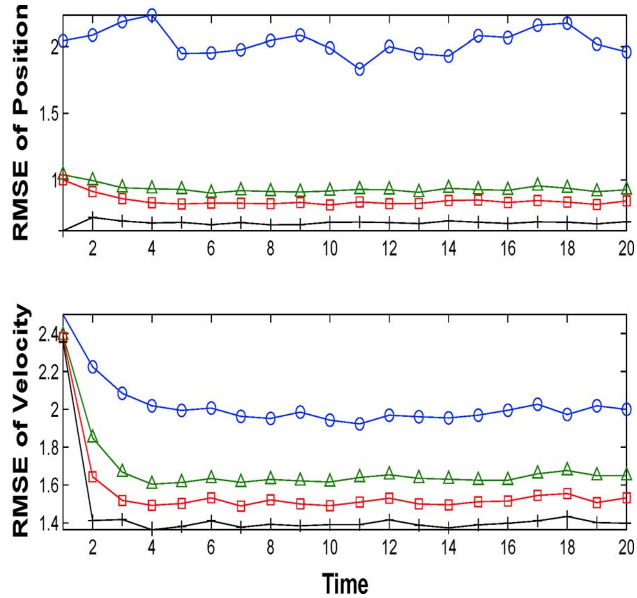


Fig. 1. RMSE comparison for the feedback system with $N = 2$. Solid line with circle: random-selection, solid line with triangle: CFoMD, solid line with square: CFwMD, solid line with plus: all-send.

Fig. 1. The reason is that the censoring process selects more informative measurements, and the missing data due to censoring process in the CFoMD is NMAR, i.e., is non-ignorable [37]. Ignoring the data as in CFoMD will certainly result in some information loss. We can also observe that there is only a small gap between the performances of CFwMD and that of the all-send case. Since in the random-selection scheme, each sensor has probability of 1/2 to send its observation, it also saves 50% transmissions on an average, compared to the all-send case. But, it performs the worst among the four schemes as expected, since the per-sensor censoring process in the CFwMD and CFoMD schemes select more informative data than random selection.

In Fig. 2, we compare the RMSEs of two feedback cases with different values of l, i.e., $l = 1, 1.2, 1.5, 2, 3$, when the total number of sensors is increased to $N = 4$. The CFwMD in the figure is the case when only the global state prediction $\hat{\mathbf{x}}_{k|k-1}$ is fed back, while the CFwMD2 is the case when both the global state prediction $\hat{\mathbf{x}}_{k|k-1}$ and its covariance $\hat{\mathbf{P}}_{k|k-1}$ are fed back at any given time k. We can observe that, when $l > 1$, the CFwMD2 performs better than the CFwMD, due to the extra feedback from the FC. However, when $l = 3$, the CFwMD2 does not provide much performance improvement. This is because, on the average three sensors' observations, the FC can provide very good estimation performance. Therefore, the extra feedback does not contribute much. On the other hand, it can be observed that the performance of the CFwMD is better than that of the CFwMD2 when $l = 1$. The reason is as follows: when $l = 1$, the probability that at a particular time none of the sensors sends data, which is $(3/4)^4 = 0.32$, is much greater than that when $l = 2$, which is $(1/2)^4 = 0.06$. If at a certain time step, no data are sent to the FC, it would be more likely that at the next time step no sensor data are sent to the FC either. This is because if no data are available for the FC to update its state estimate at time $k-1$, both $\hat{\mathbf{P}}_{k-1|k-1}$ and $\hat{\mathbf{P}}_{k|k-1}$ will increase

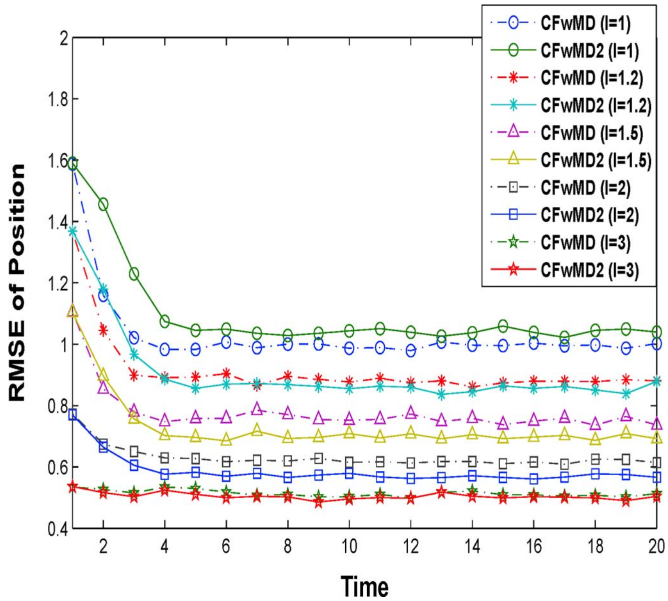


Fig. 2. RMSEs for the CFwMD with/without covariance feedback for different ls.

TABLE I
AVERAGE NUMBER OF TRANSMISSIONS (SCALAR OBSERVATION)

Theoretical l	l (CFwMD)	l (CFwMD2)
1	1.0625	1.0530
1.2	1.1709	1.2230
1.5	1.3536	1.4992
2	1.7399	2.0002
3	2.7838	3.0095

significantly. A larger $\hat{\mathbf{P}}_{k|k-1}$, which is fed back to local sensors in the CFwMD2 scheme, results in a larger s_k^i and makes it more difficult for the sensor data to pass the censoring rule defined in (4) at time k, while in CFwMD, $\mathbf{P}_{k|k-1}^i$, which completely depends on the system model, is not affected by the estimation process at all. Hence, the probability that no data are sent for several consecutive time steps is much larger for CFwMD2 when $l = 1$. This has been verified by Monte-Carlo simulations, where we observe more instances of no sensor data being sent over several consecutive time steps in the case of CFwMD2 than those in CFwMD when $l = 1$. Indeed, in Table I, one can observe that, when $l = 1$, the experimental average number of transmissions of CFwMD2 is smaller than that of CFwMD. We did not observe similar phenomena for the cases when the state process noise σ is smaller or when the observation is a vector consisting of both position and velocity observations, the latter of which will be given later in the paper. This is because $\hat{\mathbf{P}}_{k|k-1}$ is smaller in either of these two cases.

Another observation from Table I is that, for each l, the average number of transmissions of CFwMD2 is closer to the theoretical value than that of CFwMD, which justifies our expectation that the bandwidth constraint of CFwMD2 should be more strictly satisfied than CFwMD.

2) Non-Feedback System: For a non-feedback system, again the RMSEs of the four schemes, i.e., random-selection, CFwMD, CFoMD, and all-send, are compared. In Fig. 3, the results for a system with $N = 2$ sensors are presented. As in the feedback system, it is obvious that CFoMD outperforms

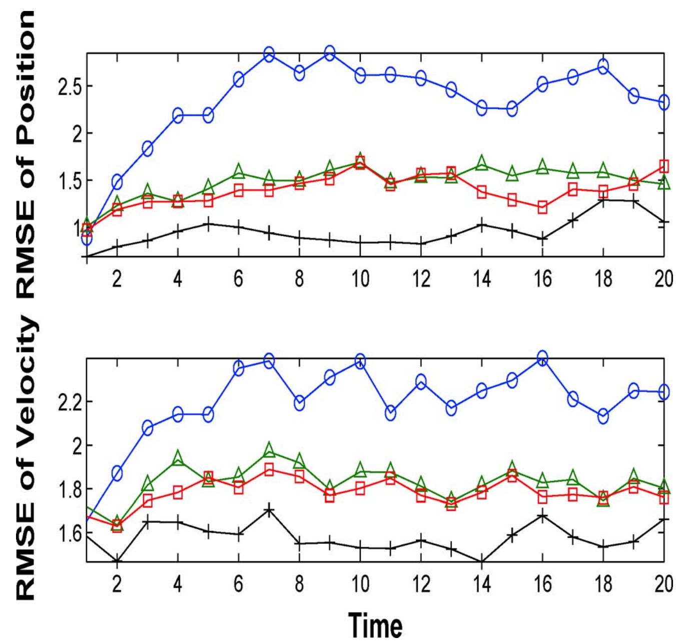


Fig. 3. RMSE comparison for the non-feedback system with $N = 2$. Solid line with circle: random-selection, solid line with triangle: CFoMD, solid line with square: CFwMD, solid line with plus: all-send.

random-selection, and CFwMD performs the best among the three schemes, i.e., random-selection, CFoMD and CFwMD. By observing Fig. 3, we can also conclude that, though the random-selection saves 50% transmissions when $N = 2$, it incurs a large loss of performance as expected.

B. Linear System—The Vector Observation Case

In this example, the same one-dimensional moving target is tracked as that in Section VII.A. But, the observation matrix is set as $H = \mathbf{I}_2$, an identity matrix with dimension 2×2. Thus, both the position and the velocity of the target can be observed by local sensors. Again, $N = 2$ identical sensors are used, and the measurement covariance is set as $\mathbf{R}^i = \mathrm{diag}[2\ \ 4]$ for $i = 1, 2$. As in Section VII.A, we design the censoring threshold $\tilde{\eta}$ such that there is only one active sensor, i.e., $l = 1$, at any given time on the average. The target is tracked for 20 seconds for each Monte-Carlo trial and 5000 Monte-Carlo trials are performed. We compare the RMSEs for the random-selection, CFwMD, CFoMD and all-send cases. In Fig. 4, the results for the feedback system with vector observations are presented. Obviously, similar conclusion as that in Section VII.A can be drawn here.

In Fig. 5, as in the scalar observation case, the position RMSEs of the CFwMD with only global state feedback and the CFwMD2 with both the global state and covariance feedback are compared for different l, and the total number of sensors is again set as $N = 4$. Obviously, CFwMD2 outperforms CFwMD for each l, which is due to the extra feedback. Similar results can be observed for the RMSE comparison of the velocity, which is omitted here for brevity. On the other hand, the experimental average number of transmission $\bar{l}$ of CFwMD2 for each l, especially when $l > 1$, provided in Table II is closer to the theoretical one than that of CFwMD, which again verifies

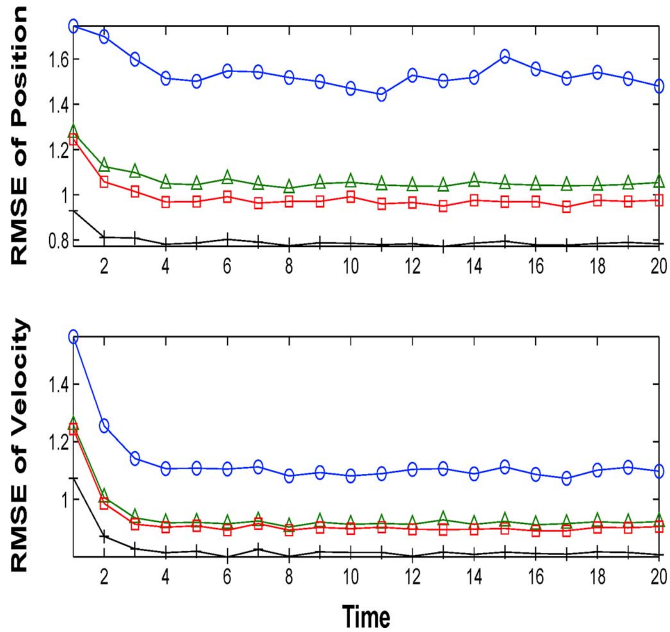


Fig. 4. RMSE comparison for the feedback system with $N = 2$ (vector observation). Solid line with circle: random-selection, solid line with triangle: CFoMD, solid line with square: CFwMD, solid line with plus: all-send.

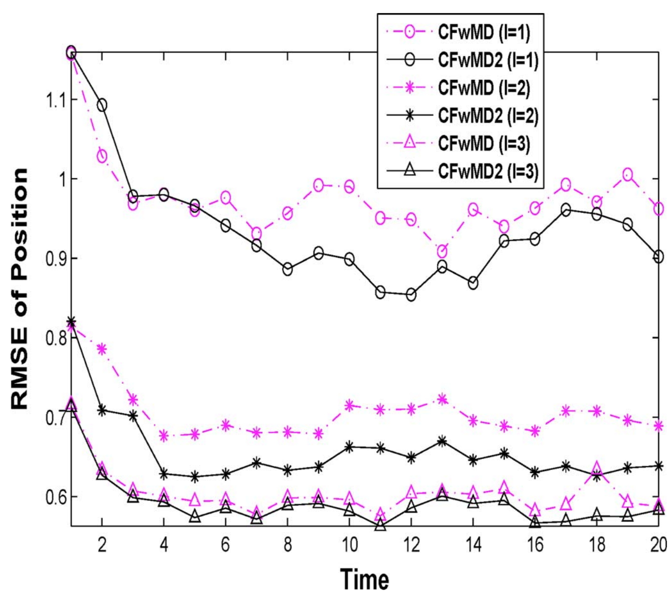


Fig. 5. RMSEs for the CFwMD with/without covariance feedback for different l s (vector observation).

TABLE II
EXPERIMENTAL AVERAGE NUMBER OF TRANSMISSIONS (VECTOR OBSERVATION)

Theoretical l	$\bar{l}$ (CFwMD)	$\bar{l}$ (CFwMD2)
1	0.9296	1.0069
2	1.7573	2.0035
3	2.8029	3.0151

that the bandwidth constraint of CFwMD2 is more strictly satisfied due to the feedback of the global covariance.

The results for the non-feedback system with vector observations are provides in Fig. 6. Obviously, we can draw similar conclusions as that in Section VII.A2.

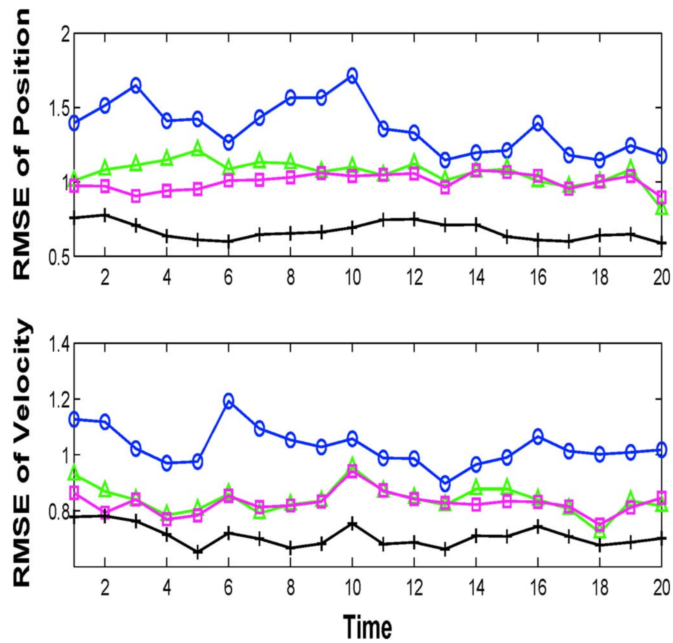


Fig. 6. RMSE comparison for the non-feedback system with $N = 2$ (vector observation). Solid line with circle: random-selection, solid line with triangle: CFoMD, solid line with square: CFwMD, solid line with plus: all-send.

We should point out that simulation approach has been used to compute the probability $p({q^i}^T q^i < \tilde{\eta}_k)$ to get the full likelihood function (33) when using the CFwMD scheme for a feedback system. That is, we first draw N_q samples from the normal distribution $\mathcal{N}(\mathbf{S}_k^{-\frac{1}{2}}\mathbf{H}(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}), \mathbf{S}_k^{-\frac{1}{2}}\mathbf{R}\mathbf{S}_k^{-\frac{1}{2}})$, and then count the number of samples which satisfy the condition $q^T q < \tilde{\eta}_k$, denoted as n_q. Then, the probability can be approximated by n_q/N_q. The same approach is also used to compute the probability $p({t^i}^T t^i < \tilde{\eta}_k)$ involved in (40) for a non-feedback system.

C. Nonlinear System

In this experiment, we assume $N = 9$ sensors are grid deployed in a $b^2 = 20\ \mathrm{m} \times 20\ \mathrm{m}$ surveillance area, and an acoustic or an electromagnetic source is moving in this region, as shown in Fig. 7. Target motion is defined by the white noise acceleration model (1) with state vector $\mathbf{x} = [x\ \dot{x}\ y\ \dot{y}]$, where the state transition matrix F and the state noise covariance $\mathbf{Q}$ are given as follows

$$\mathbf{F} = \begin{bmatrix} 1 & \mathcal{D} & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \mathcal{D} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{Q} = \sigma^2 \begin{bmatrix} \mathcal{D}^4/4 & \mathcal{D}^3/2 & 0 & 0 \\ \mathcal{D}^3/2 & \mathcal{D}^2 & 0 & 0 \\ 0 & 0 & \mathcal{D}^4/4 & \mathcal{D}^3/2 \\ 0 & 0 & \mathcal{D}^3/2 & \mathcal{D}^2 \end{bmatrix}$$

At time k, the signal power received at the ith sensor is given as $z_k^i = \sqrt{\frac{P_0}{1+\alpha(d_k^i)^n}} + n_k^i$, where P_0 denotes the signal power of the target, d_k^i is the distance between the target and the ith sensor at time k, α and n are model parameters, and n_k^i is Gaussian noise with zero mean and variance r^i. Without loss

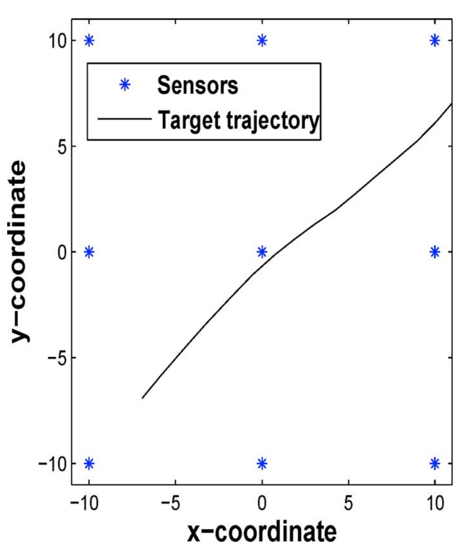


Fig. 7. Target trajectory and sensor deployment in the ROI.

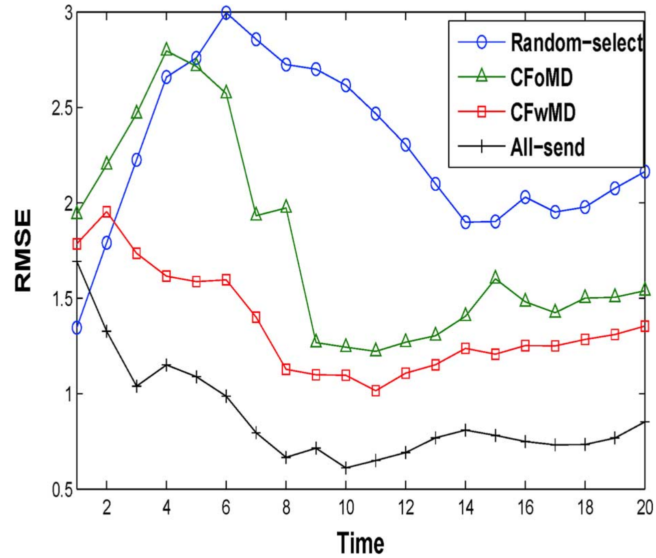


Fig. 8. RMSE comparison for the nonlinear system with feedback.

of generality, local sensors are set up with the same measurement noise variance $r^i = 1 (i = 1, \ldots, N)$ in this example. We set $P_0 = 10^3$, $\alpha = 1$, and $n = 2$. The target's initial state $\mathbf{x}_0$ is assumed to be Gaussian with mean $\boldsymbol{\mu}_0 = [-8\ 2\ -8\ 2]$ and covariance $\boldsymbol{\Sigma}_0 = \mathrm{diag}[9\ 4\ 9\ 4]$ (i.e., a poor prior on the initial state). The state process noise parameter σ^2 is set as 0.1, indicating that the target trajectory has relatively large uncertainty. Measurements are assumed to be taken at regular intervals of $\mathcal{D} = 0.5$ seconds and the tracking length is 10 seconds, namely, we track the target over $T_S = 20$ time steps for each Monte-Carlo trial. 200 Monte-Carlo trials are performed in this experiment. The number of particles used in the particle filter at the FC is $N_s = 10^4$.

As in linear systems, the RMSEs, averaged over the Monte-Carlo trials at each time, for the random-selection, CFwMD, CFoMD and all-send cases are compared. The average number of transmission at any given time in this experiment is constrained as $l = 2$.

In Fig. 8, the RMSE comparison results are shown. Note that only state estimate is fed back in this figure. It can be observed that the proposed CFwMD outperforms CFoMD and random-selection under the same bandwidth constraint. On the other hand, compared to the all-send case, CFwMD loses not much performance but saves 78% transmission. One may observe that RMSEs increase with time at later time steps in Fig. 8. This is because the target is moving out of the region of interest (ROI) monitored by the sensors, so there is less and less information available for the estimator.

In Fig. 9, the RMSEs of the four schemes, namely, the random-selection, CFoMD, CFwMD, and all-send, are plotted as a function of the average number of transmissions at any time step. One can observe that, when the allowed number of transmissions is small, the proposed CFwMD has significant advantage over both CFoMD and random-selection. It incurs a little bit performance loss compared to the all-send case. As we increase the allowed number of transmissions, the RMSEs of the four schemes approach each other, especially when l is close

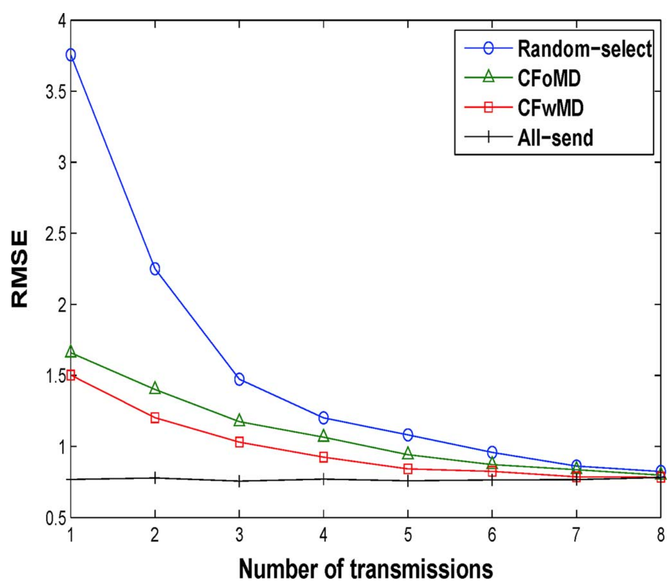


Fig. 9. RMSEs as a function of the average number of transmission at each time.

to the total number of sensors $N = 9$ in the network. This is intuitively reasonable, since when the number of transmissions is large enough, the received observations can already provide enough information for good inference performance, and then either the censoring procedure or the information conveyed by the missing data cannot improve the performance much.

For the nonlinear system, we are also interested in the performance comparison between the two feedback scenarios: 1) only global state estimate feedback is available; 2) the feedback consists of both the global state estimate and its covariance, and the results are provided in Fig. 10 for $l = 2, 4, 6$ (the total number of sensors in the ROI is $N = 9$). It can be observed that, as in the linear Gaussian system, CFwMD2 performs better than CFwMD as time goes along for each l, since extra global information is fed back to local sensors by the FC. Again, the experimental average number of transmissions over 200 Monte-carlo

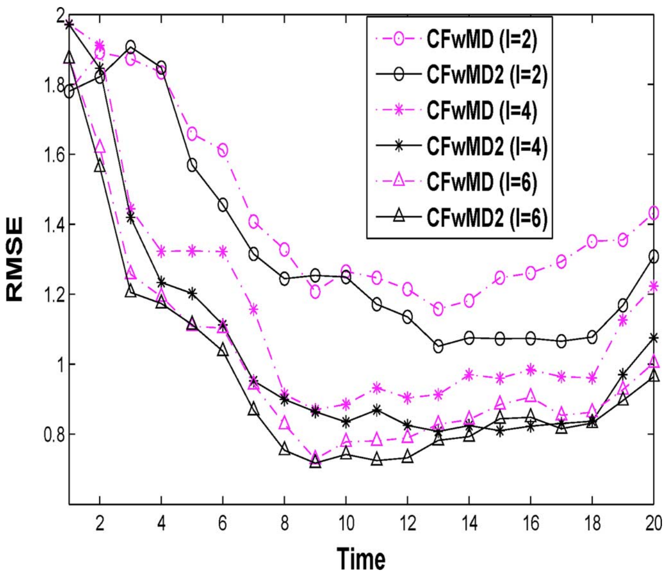


Fig. 10. RMSEs for the CFwMD with/without covariance feedback for different l s in a nonlinear system.

TABLE III
EXPERIMENTAL AVERAGE NUMBER OF TRANSMISSIONS (NONLINEAR SYSTEM)

Theoretical l	$\bar{l}$ (CFwMD)	$\bar{l}$ (CFwMD2)
2	1.5978	1.8970
4	3.3480	3.8775
6	5.4835	5.9533

trials provided in Table III indicates that the bandwidth constraint of CFwMD2 is more strictly satisfied than CFwMD.

D. Discussion

It should be noted that the models used in the simulations have relatively low dimension and the network size is rather small. However, such scenarios are frequently used in the target tracking literature [41], [1], [4]. Therefore, we think that they are appropriate to illustrate the effectiveness of the proposed algorithm. We would like to point out that the proposed methodology can also be applied to moderately high dimensional systems without requiring large computation effort if feedback is available from the fusion center to local sensors. This is clear if one checks (9), (15), (33) and (39) for linear systems, and (44) and (45) for nonlinear systems. For a non-feedback system, if the dimensionality of the dynamic system is high and/or the number of sensors is large, the proposed methodology involves computationally intensive multiple integrals in (16) and (40). However, if the fusion center is very powerful, the proposed methodology can still be applicable relying on efficient numerical integration approaches, such as those based on Monte Carlo integration techniques [45]. Note that in this paper, we have implicitly assumed that identical dynamical model is observed at each sensor. However, this may not be true in some realistic scenarios such as very large-scale dynamical systems [46], [47], and this will be addressed in future work.

VIII. CONCLUSION

In this paper, we have proposed a new scheme to solve linear Bayesian sequential estimation problems by combining the censoring procedure at local sensors and the fusion procedure which fuses both received observations and missing ones, due to the censoring process, at the FC. Both scalar observation and vector observation cases have been discussed in the paper. In addition, for scalar observation case, it has been shown that the proposed innovation based censoring rule is equivalent to that based on the KL divergence between the prior state PDF and the posterior state PDF. Then, we extended the proposed CFwMD to a general nonlinear filtering problem when feedback is available. Numerical results show that, for both linear and nonlinear filtering problems we considered in this paper, CFwMD achieves less performance loss than the CFoMD, while both save the same amount of transmissions, compared to the all-send case. In addition, under the same bandwidth constraint, the proposed CFwMD is shown to perform the best among the three schemes, i.e., CFwMD, CFoMD and random-selection. Future work will theoretically analyze the performance of the proposed CFwMD scheme. In the current work, the channels between the local sensors and the FC are assumed to be perfect. Then, taking a fading channel into consideration is another interesting future work.

APPENDIX A
PROOF OF COROLLARY 2

Proof: When symmetric KL divergence is used, the metric to select more informative data in (26) is changed to

$$\Lambda_k \triangleq (\hat{\mathbf{x}}_{k|k} - \hat{\mathbf{x}}_{k|k-1})^T \left(\mathbf{P}_{k|k}^{-1} + \mathbf{P}_{k|k-1}^{-1}\right) (\hat{\mathbf{x}}_{k|k} - \hat{\mathbf{x}}_{k|k-1}) \tag{46}$$

Following the same manipulation on Λ_k as in the proof of Theorem 1, we can obtain

$$\Lambda_k = \nu_k^T \mathbf{w}_k^T \left(\mathbf{P}_{k|k}^{-1} + \mathbf{P}_{k|k-1}^{-1}\right) \mathbf{w}_k \nu_k \tag{47}$$

Again, since scalar measurements are obtained, $\mathbf{w}_k^T(\mathbf{P}_{k|k}^{-1} + \mathbf{P}_{k|k-1}^{-1})\mathbf{w}_k$ is a scalar, so is s_k. Therefore, we have

$$\mathbf{w}_k^T \left(\mathbf{P}_{k|k}^{-1} + \mathbf{P}_{k|k-1}^{-1}\right) \mathbf{w}_k \propto s_k^{-1} \tag{48}$$

■

REFERENCES

[1] Y. Bar-Shalom, P. K. Willett, and X. Tian, *Tracking and Data Fusion: A Handbook of Algorithms*. Storrs, CT, USA: YBS Publishing, 2011.
[2] L. Zuo, R. Niu, and P. K. Varshney, “A sensor selection approach for target tracking in sensor networks with quantized measurements,” presented at the Int. Conf. Acoust., Speech, Signal Process. (ICASSP), Las Vegas, NV, USA, Mar. 2008.
[3] O. Ozdemir, R. Niu, and P. K. Varshney, “Tracking in wireless sensor networks using particle filtering: Physical layer considerations,” *IEEE Trans. Signal Process.*, vol. 57, no. 5, pp. 1987–1999, May 2009.

[4] E. Masazade, R. Niu, and P. K. Varshney, "Dynamic bit allocation for object tracking in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 60, pp. 5048–5063, Oct. 2012.

[5] S. Grime and H. Durrant-Whyte, "Data fusion in decentralized sensor networks," *IFAC Control Eng. Practice*, vol. 2, no. 5, pp. 849–863, 1994.

[6] S. J. Julier and J. K. Uhlmann, "A non-divergent estimation algorithm in the presence of unknown correlations," in *Proc. Amer. Control Conf.*, 1997, pp. 2369–2373.

[7] O. E. Drummond, "On track and tracklet fusion filtering," in *Proc. SPIE Conf. Signal Data Process. Small Targets*, Aug. 2002, vol. 4728, pp. 176–195.

[8] H. Chen, T. Kirubarajan, and Y. Bar-Shalom, "Performance limits of track-to-track fusion versus centralized estimation: Theory and application," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 2, pp. 386–400, Apr. 2003.

[9] Y. M. Zhu, Z. S. You, J. Zhou, K. S. Zhang, and X. R. Li, "The optimality for the distributed Kalman filtering fusion with feedback," *Automatica*, vol. 37, pp. 1489–1493, Sep. 2001.

[10] F. Govaers and W. Koch, "An exact solution to track-to-track-fusion at arbitrary communication rates," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 48, no. 3, pp. 2718–2729, Jul. 2012.

[11] D. Castanon and D. Teneketzis, "Distributed estimation algorithms for nonlinear systems," *IEEE Trans. Autom. Control*, vol. 30, pp. 418–425, May 1985.

[12] C. Y. Chong, S. Mori, and K. C. Chang, "Distributed multitarget multisensor tracking," in *Multitarget-Multisensor Tracking: Advanced Applications*, Y. Bar-Shalom, Ed. Norwood, MA, USA: Artech House, 1990.

[13] F. Bourgault and H. F. Durrant-Whyte, "Communication in general decentralized filters and the coordinated search strategy," in *Proc. 2004 7th Int. Conf. Inf. Fusion*, 2004, pp. 723–770.

[14] L. L. Ong, T. Bailey, H. Durrant-Whyte, and B. Upcroft, "Decentralised particle filtering for multiple target tracking in wireless sensor networks," presented at the 2008 11th Int. Conf. Inf. Fusion, Cologne, Germany, Jun. 2008, pp. 1–8.

[15] S. Kar and J. M. F. Moura, "Gossip and distributed Kalman filtering: Weak consensus under weak detectability," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1766–1784, Apr. 2011.

[16] R. Olfati-Saber, "Kalman-consensus filter: Optimality, stability, and performance," in *Proc. Joint 48th IEEE Conf. Decision Control/28th Chin. Control Conf.*, Shanghai, China, Dec. 2009, pp. 7036–7042.

[17] R. Carli, A. Chiuso, L. Schenato, and S. Zampieri, "Distributed Kalman filtering based on consensus strategies," *IEEE J. Sel. Areas Commun.*, vol. 26, pp. 622–633, Apr. 2008.

[18] T. C. Aysal, M. E. Yildiz, A. D. Sarwate, and A. Scaglione, "Broadcast gossip algorithms for consensus," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2748–2761, Jul. 2009.

[19] H. Liu, H. So, F. Chan, and K. Lui, "Distributed particle filtering for target tracking in sensor networks," *Progr. Electromagn. Res. C*, vol. 11, pp. 171–182, 2009.

[20] S. Farahmand, S. I. Roumeliotis, and G. B. Giannakis, "Set-membership constrained particle filter: Distributed adaptation for sensor networks," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4122–4138, Sep. 2011.

[21] O. Hlinka, O. Sluciak, F. Hlawatsch, P. M. Djuric, and M. Rupp, "Likelihood consensus and its application to distributed particle filtering," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 4334–4349, Aug. 2012.

[22] A. Mohammadi and A. Asif, "Distributed particle filter implementation with intermittent/irregular consensus convergence," *IEEE Trans. Signal Process.*, vol. 61, no. 10, pp. 2572–2587, May 2013.

[23] S. Joshi and S. Boyd, "Sensor selection via convex optimization," *IEEE Trans. Signal Process.*, vol. 57, pp. 451–462, Feb. 2009.

[24] J. Williams, J. Fisher, and A. Willsky, "Approximate dynamic programming for communication-constrained sensor network management," *IEEE Trans. Signal Process.*, vol. 55, pp. 4300–4311, Aug. 2007.

[25] Y. Mo, R. Ambrosino, and B. Sinopoli, "Sensor selection strategies for state estimation in energy constrained wireless sensor networks," *Automatica*, vol. 47, pp. 1330–1338, Jul. 2011.

[26] H. Zhang, J. Moura, and B. Krogh, "Dynamic field estimation using wireless sensor networks: Tradeoffs between estimation error and communication cost," *IEEE Trans. Signal Process.*, vol. 57, pp. 2383–2395, June 2009.

[27] M. P. Vitus, W. Zhang, A. Abate, J. Hu, and C. J. Tomlin, "On efficient sensor scheduling for linear dynamical systems," *Automatica*, vol. 48, pp. 2482–2493, Oct. 2012.

[28] V. Gupta, T. Chung, B. Hassibi, and R. M. Murray, "On a stochastic sensor selection algorithm with applications in sensor scheduling and sensor coverage," *Automatica*, vol. 42, pp. 251–260, Feb. 2006.

[29] F. Zhao, J. Shin, and J. Reich, "Information-driven dynamic sensor collaboration," *IEEE Trans. Signal Process.*, vol. 19, no. 1, pp. 61–72, May 2002.

[30] L. Zuo, R. Niu, and P. K. Varshney, "Posterior CRLB based sensor selection for target tracking in sensor networks," presented at the Int. Conf. Acoust., Speech, Signal Process. (ICASSP), Honolulu, HI, USA, Apr. 2007.

[31] E. Masazade, M. Fardad, and P. K. Varshney, "Sparsity-promoting extended Kalman filtering for target tracking in wireless sensor networks," *IEEE Signal Process. Lett.*, vol. 19, pp. 845–848, Jun. 2012.

[32] C. Rago, P. Willett, and Y. Bar-Shalom, "Censoring sensors: A low-communication-rate scheme for distributed detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 32, no. 1, pp. 554–568, 1996.

[33] R. Jiang and B. Chen, "Fusion of censored decisions in wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 2668–2673, Nov. 2005.

[34] S. Appadwedula, V. V. Veeravalli, and D. L. Jones, "Decentralized detection with censoring sensors," *IEEE Trans. Signal Process.*, vol. 56, pp. 1362–1373, Apr. 2008.

[35] R. S. Blum and B. M. Sadler, "Energy efficient signal detection in sensor networks using ordered transmission," *IEEE Trans. Signal Process.*, vol. 56, pp. 3229–3235, Jul. 2008.

[36] X. Chen, R. Blum, and B. M. Sadler, "A new scheme for energy-efficient estimation in a sensor network," presented at the 43rd Annu. Conf. Inf. Sci. Syst., Baltimore, MD, USA, Mar. 2009.

[37] R. J. Little and D. B. Rubin, *Statistical Analysis With Missing Data*. New York, NY, USA: Wiley, 2002.

[38] E. J. Msechu and G. B. Giannakis, "Sensor-centric data reduction for estimation with WSNs via censoring and quantization," *IEEE Trans. Signal Process.*, vol. 60, pp. 400–414, Jan. 2012.

[39] W. Koch, "On exploiting 'negative' sensor evidence for target tracking and sensor data fusion," in *Information Fusion, Special Issue on the Seventh International Conference on Information Fusion—Part II*, Jan. 2007, vol. 8, no. 10, pp. 28–39.

[40] Y. Zheng, R. Niu, and P. K. Varshney, "Sequential Bayesian estimation with censored data," presented at the Statist. Signal Process. Workshop (SSP), Ann Arbor, MI, USA, Aug. 2012.

[41] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation With Applications to Tracking and Navigation*. New York, NY, USA: Wiley, 2001.

[42] Y. Ruan, P. Willett, A. Marrs, F. Palmieri, and S. Marano, "Practical fusion of quantized measurements via particle filtering," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 44, no. 1, pp. 15–29, Jan. 2008.

[43] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174–188, Feb. 2002.

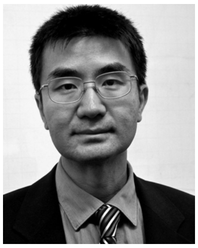
[44] S. J. Roberts and W. D. Penny, "Variational Bayes for generalized autoregressive models," *IEEE Trans. Signal Process.*, vol. 50, pp. 2245–2257, Sep. 2002.

[45] W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, and M. Metcalf, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 1992.

[46] U. A. Khan and J. M. F. Moura, "Distributing the Kalman filter for large-scale systems," *IEEE Trans. Signal Process.*, vol. 56, pp. 4919–4935, Oct. 2008.

[47] A. Mohammadi and A. Asif, "Distributed particle filtering for large scale dynamical systems," presented at the IEEE 13th Int.l Multitopic Conf. (INMIC), Islamabad, Pakistan, Dec. 2009, pp. 1–5.

Yujiao Zheng (S'13) received her B.S. degree in electronic engineering and information science from University of Science and Technology of China (USTC) in 2008, M.S. degree in electrical engineering and Ph.D. degree in electrical and computer engineering from Syracuse University, Syracuse, NY, in 2011 and 2014 respectively. Her research interests are in the areas of statistical signal processing with its application in target tracking, sensor management, and compressive sensing. She received the Best Student Paper Award at the Thirteenth International Conference on Information Fusion in 2010.



Ruixin Niu (M'04–SM'11) received his B.S. degree from Xian Jiaotong University, Xian, China, in 1994, M.S. degree from the Institute of Electronics, Chinese Academy of Sciences, Beijing, in 1997, and Ph.D. degree from the University of Connecticut, Storrs, in 2001, all in electrical engineering. He is currently an assistant professor with the Department of Electrical and Computer Engineering, Virginia Commonwealth University (VCU), Richmond. Before joining VCU he was a research assistant professor with Syracuse University, Syracuse, NY. His research interests are in the areas of statistical signal processing and its applications, including detection, estimation, information fusion, sensor networks, communications, and compressive sensing. Dr. Niu received the Best Paper Award at the Seventh International Conference on Information Fusion in 2004. He is a coauthor of the paper that won the Best Student Paper Award at the Thirteenth International Conference on Information Fusion in 2010. He is an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the Associate Administrative Editor of the *Journal of Advances in Information Fusion*. He was an Associate Editor of the *International Journal of Distributed Sensor Networks* between 2010 and 2012.

Pramod K. Varshney (S'72–M'77–SM'82–F'97) was born in Allahabad, India, on July 1, 1952. He received the B.S. degree in electrical engineering and computer science (with highest honors), and the M.S. and Ph.D. degrees in electrical engineering from the University of Illinois at Urbana-Champaign in 1972, 1974, and 1976 respectively.

From 1972 to 1976, he held teaching and research assistantships with the University of Illinois. Since 1976, he has been with Syracuse University, Syracuse, NY, where he is currently a Distinguished Professor of Electrical Engineering and Computer Science and the Director of CASE: Center for Advanced Systems and Engineering. He served as the Associate Chair of the department from 1993 to 1996. He is also an Adjunct Professor of Radiology at Upstate Medical University, Syracuse. His current research interests are in distributed sensor networks and data fusion, detection and estimation theory, wireless communications, image processing, radar signal processing, and remote sensing. He has published extensively. He is the author of Distributed Detection and Data Fusion (New York: Springer-Verlag, 1997). He has served as a consultant to several major companies.

Dr. Varshney was a James Scholar, a Bronze Tablet Senior, and a Fellow while at the University of Illinois. He is a member of Tau Beta Pi and is the recipient of the 1981 ASEE Dow Outstanding Young Faculty Award. He was elected to the grade of Fellow of the IEEE in 1997 for his contributions in the area of distributed detection and data fusion. He was the Guest Editor of the Special Issue on Data Fusion of the IEEE PROCEEDINGS January 1997. In 2000, he received the Third Millennium Medal from the IEEE and Chancellors Citation for exceptional academic achievement at Syracuse University. He is the recipient of the IEEE 2012 Judith A. Resnik Award. He is on the Editorial Boards of the *Journal on Advances in Information Fusion* and *IEEE Signal Processing Magazine*. He was the President of International Society of Information Fusion during 2001.