A Novel Distributed State Estimation Algorithm with Consensus Strategy

Jun Liu [1], Yu Liu [1,2,*], Kai Dong [1], Ziran Ding [1] and You He [1]

[1]   Research Institute of Information Fusion, Naval Aviation University, Yantai 264001, China;
    18615042187@163.com (J.L.); 188dongkai@163.com (K.D.); DZR19931201@163.com (Z.D.);
    heyou_f@126.com (Y.H.)
[2]   School of Electronic and Information Engineering, Beihang University, Beijing 100191, China
*   Correspondence: liuyu77360132@126.com; Tel.: +86-135-8350-4046

check for
updates

Abstract: Owing to its high-fault tolerance and scalability, the consensus-based paradigm has attracted immense popularity for distributed state estimation. If a target is neither observed by a certain node nor by its neighbors, this node is naive about the target. Some existing algorithms have considered the presence of naive nodes, but it takes sufficient consensus iterations for these algorithms to achieve a satisfactory performance. In practical applications, because of constrained energy and communication resources, only a limited number of iterations are allowed and thus the performance of these algorithms will be deteriorated. By fusing the measurements as well as the prior estimates of each node and its neighbors, a local optimal estimate is obtained based on the proposed distributed local maximum a posterior (MAP) estimator. With some approximations of the cross-covariance matrices and a consensus protocol incorporated into the estimation framework, a novel distributed hybrid information weighted consensus filter (DHIWCF) is proposed. Then, theoretical analysis on the guaranteed stability of the proposed DHIWCF is performed. Finally, the effectiveness and superiority of the proposed DHIWCF is evaluated. Simulation results indicate that the proposed DHIWCF can achieve an acceptable estimation performance even with a single consensus iteration.

Keywords: sensor networks; distributed state estimation; naive node; Kalman filter; maximum a posterior estimator; consensus filter

1. Introduction

Recently, distributed state estimation has been a hot topic in the field of target tracking in sensor networks [1–19]. As a traditional method, the centralized scheme needs to simultaneously process the local measurements from all sensors in the fusion center at each time instant [3,15]. This scheme guarantees the optimality of estimates, but a lot of communication and a powerful fusion center are required to maintain the operation, which may give rise to problems when the network size is increased or the communication resources are restricted.

Unlike the centralized scheme, the distributed mechanism tries to recover the centralized performance via local communications between neighboring nodes. Specifically, each node in the network only exchanges information with its immediate neighbors to achieve a comparable performance to its centralized counterpart, which reduces communication cost and makes the network robust against possible failures of some nodes [8]. The consensus filter, which computes the average of interested values in a distributed manner, has attracted immense popularity for distributed state estimation [4–7,9–14,16–22]. Recently, in [23,24], the multiscale consensus scheme, in which the local estimated states achieve asymptotically prescribed ratios in terms of multiple scales, has been discussed and analyzed. The well-known Kalman consensus filter (KCF) [4–6,9,14] combines the local Kalman

filter with the average consensus protocol together to update the posterior state. In the update stage, each node exploits the measurement innovations as well as the prior estimates from its inclusive neighbors (including the node itself and its immediate neighbors) to correct its prior estimate. However, the prior estimates from its immediate neighbors are assigned with same weights. This may ensure consensus on the estimates from different nodes after a period of time, but the estimation accuracy is not guaranteed. It is very likely that a target is neither observed by a certain node nor observed by any of its immediate neighbors. That is, there is no measurement in the inclusive neighborhood of the node, and it is naive about the target's state. Similar to [16,22], such a node is referred as a naive node. Since a naive node contains less information about the target, it usually results in an inaccurate estimate. If a naive node is given an identical weight to the informed nodes, the final estimate will be severely contaminated, which may even cause the final estimates to diverge [9,13]. In addition, the cross covariances among different nodes are ignored in the derivation of KCF for computational and bandwidth requirements, and thus the covariance of each node is updated without regard to its neighbor's prior covariance during the consensus step. Given no naive nodes in the network, KCF is able to provide satisfactory results. However, due to limited sensing abilities or constrained communication resources, a network often consists of some naive nodes. Especially in sparse sensor networks, this phenomenon is even more serious. In such a case, KCF would result in poor estimates [3]. To solve this problem, before updating the posterior estimate, the generalized Kalman consensus filter (GKCF) performs consensus on the prior information vectors and information matrices within the inclusive neighborhood of each node [4,16,19]. As is analyzed in [4], this procedure greatly improves the estimation performance in presence of naive nodes. GKCF updates current state based on consensus on prior estimates, but the current measurements are not considered for naive nodes. Each naive node only has access to measurements of the previous time instant existing in prior estimates. Therefore, there is a delay for naive nodes to access current measurements. On the contrary, consensus on measurements algorithm (CM) performs consensus on measurements [5,25–27], which can achieve the centralized performance with infinite consensus iterations. However, the stability is not guaranteed unless the number of consensus iterations is large enough [26]. Consensus on information algorithm (CI) performs consensus on both prior estimates and measurements [26–28], which can be viewed as a generalization of the covariance intersection fusion rule to multiple iterations [29]. CI guarantees stability for any number of consensus iterations, but its estimation confidence can be degraded as a conservative rule is adopted by assuming that the correlation between estimates from different nodes are completely unknown [26,28,30].

With more consensus iterations carried out, estimates from different nodes achieve a reasonable consensus. Therefore, each node has almost completely redundant or same prior information, and hence the prior estimation errors between nodes are highly correlated. In this situation, the algorithms such as KCF, GCKF, or CM, which do not take the cross-covariance into account, are sub-optimal [16,17]. Note that the redundant information only exists in the prior estimates, which come from the converged results in the previous time instant. Using this property, the information weighted consensus filter (ICF) [18,20,21] divides the prior information of each node by N_s where N_s is the total number of nodes in the network. If each node can interact with its neighbors for infinite times, ICF will achieve the optimal estimation performance as the centralized Kalman consensus filter (CKF). In addition, ICF performs better than KCF, GKCF, CI, and CM under the same consensus iterations, which has been validated in [16,22,26]. As is pointed out in [26,30,31], the correction step by multiplying N_s may cause an overestimation of the measurement innovation for some nodes, which is often the case in sparse sensor networks. As a consequence, the estimates of some nodes may be too optimistic such that the estimation consistency will be lost, which should be avoided in recursive estimation. To address this problem, HCMCI algorithm combines the positive features of both CM and CI is proposed. It should be noted that HCMCI represents a family of different distributed algorithms dependent on the selection of scalar weights. Both CI and ICF are special cases of HCMCI. To preserve consistency of local filters as well as improve the estimation performance, the so-called normalization factor is introduced. If the

network topology is fixed, the normalization factor can be computed offline to save bandwidth. In [32], a novel distributed set-theoretic information flooding (DSIF) protocol is proposed. The DSIF protocol benefits from avoiding the reuse of information and offering the highest converging efficiency for network consensus, but it suffers from growing requirements of node-storage, more communication iterations, and higher communication load.

However, it takes sufficient consensus iterations for the algorithms discussed above to achieve an expected estimation performance. In practical applications, only a limited number of consensus iterations is allowed, and thus the performance of the afore-mentioned algorithms is corrupted. In addition, the estimation performance of the afore-mentioned algorithms depends closely on the selection of consensus weights. Inappropriate consensus weights may cause the algorithms to diverge or require more iterations to achieve consensus on the local estimates [9]. It is a common way to set the weights as a constant value as discussed in [6], which is an intuitive choice to maintain the stability of the error dynamics. However, the constant value needs the knowledge of maximum degree across the entire sensor network. Even the maximum degree is available, it remains a problem how to determine a proper constant weight to achieve the best performance while preserving the property of consistency. In addition, the initial consensus terms determined in ICF require the knowledge of the total number of nodes in the network. The global parameters, such as the maximum degree or the total number of nodes, may vary over time when the communication topology is changed, some new nodes are joined, or some existing nodes fail to communicate with others. Without the accurate knowledge of these global parameters, each node would either overestimate or underestimate the state of interest.

To deal with the problems analyzed above, a novel distributed hybrid information weighted consensus filter (DHIWCF) is proposed in this paper. Firstly, different from the previous work [4–6,16,18,22], each node assigns consensus weights to its neighbors based on their local degrees, which is fully distributed with no requirement for any knowledge of the global parameters. Secondly, the prior estimate information and measurement information at current time instant within the inclusive neighborhood are, respectively, combined together to form the local generalized prior estimate equation and the local generalized measurement equation. Then, a distributed local MAP estimator is derived with some reasonable approximations of the error covariance matrices, which achieves higher accuracy than the approaches introduce in [4–6,11,16,18,19,25–28]. Finally, the average consensus protocol with the aforementioned consensus weights is incorporated into the estimation framework, and the proposed DHIWCF is obtained. In addition, the theoretical analysis on consistency of the local estimates, stability and convergence of the estimator is performed. The comparative experiments on three different target tracking scenarios validate the effectiveness and feasibility of the proposed DHIWCF. Even with a single consensus iteration, the DHIWCF is still able to achieve acceptable estimation performance.

The remainder of this paper is organized as follows. Section 2 formulates the problem of distributed state estimation in sensor networks. The distributed local MAP estimator is derived in Section 3. Section 4 presents distributed hybrid information weighted consensus filter. The theoretical analysis on the consistency of estimates, stability and convergence of the estimator is provided in Section 5. The experimental results and analysis are considered in Section 6. The concluded remarks are given in Section 7.

Notation: $\mathbb{R}^n$ denotes the n-dimensional Euclidean space. $\|\cdot\|$ is the Euclidean norm in $\mathbb{R}^n$. For arbitrary matrix A, A^{-1} and A^{T} are respectively its inverse and transpose. $A > 0$ means A is positive definite, and tr$\{A\}$ is the shorthand for the trace of A. diag$(B_1, B_2, \ldots, B_n)$ denotes a block diagonal matrix with its main diagonal block being $B_1, B_2, \ldots, B_n$. I_n represents the $n \times n$ identity matrix. For a set C, $|C|$ means the cardinality of C. $\mathbb{E}\{\cdot\}$ is the expectation operator.

2. Problem Formulation

2.1. System Model

Consider a discrete time linear system with dynamics

$$x_{k+1} = F_k x_k + w_k \tag{1}$$

where $x_k \in \mathbb{R}^{n_x}$ represents the state vector at time instant $k \in \mathbb{Z}^+$, where $\mathbb{Z}^+$ is the set of positive integers. F_k is the state transition matrix. w_k is the Gaussian process noise with zero-mean, covariance Q_k.

The state of interest is observed by a sensor network consisting of N_s nodes in the surveillance area. The measurement model of node i is

$$z_{i,\,k} = H_{i,\,k} x_k + v_{i,k} \tag{2}$$

where $z_{i,\,k} \in \mathbb{R}^{n_z}$ is the measurement of node i at time instant k. $i = 1, 2, \ldots, N_s$ denote the sensor labels. $H_{i,\,k}$ is the measurement matrix of node i. $v_{i,k}$ is the Gaussian measurement noise with zero mean, covariance $R_{i,k}$.

Assumption 1. *The noise sequences $\{w_k\}_{k=0}^{\infty}$ and $\left\{v_{i,k}\right\}_{k=0}^{\infty}$ are mutually uncorrelated.*

Assumption 2 [16,18,22,26–28,30]**.** *If node i does not directly sense the target of interest at time instant k, then $R_{i,k}^{-1} = 0$.*

Remark 1. *Assumption 2 indicates that a node with no direct sensing ability is of infinite uncertainties about its local measurement, which guarantees consistency of the local measurements.*

2.2. Network Topology

The communication topology of the networked sensors can be represented by an undirected graph $\mathcal{G} = (\mathcal{S}, \mathcal{E})$. Here, $\mathcal{S} = \{1, 2, \ldots, N_S\}$ is the set of sensor nodes, and $\mathcal{E} \subseteq \mathcal{S} \times \mathcal{S}$ is the available communication links in the network. A communication link means that any two neighboring nodes can exchange state or measurement information with each other. A connected network means that any node in the network may directly exchange information with at least one other node. The immediate neighbors of node i is denote by $\mathcal{N}_i = \left\{j | (i, j) \in \mathcal{E}\right\}$. $d_i = |\mathcal{N}_i|$ is degree of node i, which is the number of neighboring nodes linked to node i. The inclusive neighborhood which includes node i is represented by $\mathcal{J}_i = \{i\} \cup \mathcal{N}_i$. A more comprehensible way to describe the network topology is using adjacency matrix A, where element $a_{i,j} = 1$ means that node i can exchange information with node j, and $a_{i,j} = 0$ means that there is no direct communication link between node i and node j. The immediate neighbors of node i can be easily represented by $\mathcal{N}_i = \left\{j | a_{i,j} = 1\right\}$, and The degree of node i is $d_i = \sum_{j=1}^{N_s} a_{i,j}$.

Definition 1. *If the target of interest is neither observed by node i nor observed by its neighbors $j \in \mathcal{N}_i$, then node i is referred as a naive node. It should be noted that if node i is naive about the target of interest, $R_{j,k}^{-1} = 0$ for $j \in \mathcal{J}_i$ in view of Assumption 2.*

For instance, there are 8 sensor nodes in the monitored area, and the communication topology is shown in Figure 1. Assume that only node 1 can directly observe the interested target, then nodes $\{2, 3, 4\}$ can acquire state information from node 1 by local communication. However, there are no measurements within the inclusive neighborhood of nodes $\{5, 6, 7, 8\}$, thus they are naive about the target's state.
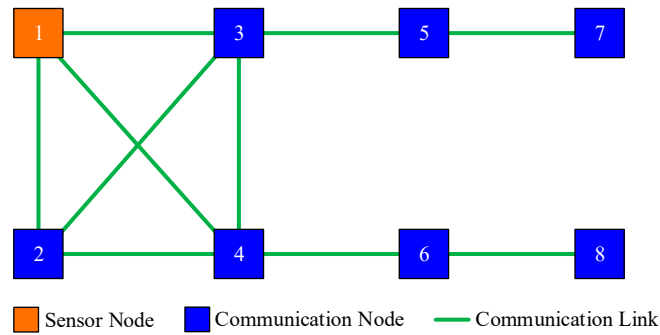
Figure 1. An illustrated sensor network with naive nodes.

From the perspective of adjacency matrix, the communication topology shown in Figure 1 can be simply represented by adjacency matrix A, where element $a_{i,j} = 1$ means the available communication links in the illustrated network. It is easy to obtain the degree of each node from A. The neighbors of each nodes are also evident in A.

$$A = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \tag{3}$$

2.3. Average Consensus

As an effective method to compute the mean value, the average consensus operates in a distributed fashion, which sheds light on the problem of distributed state estimation. Suppose the initial value of each node is α_i^0. The goal is to compute the mean $\sum_{i=1}^{N_s} \alpha_i^0 / N_s$ by local communications between neighboring nodes. At time instant k, node i sends its previous state α_i^{l-1} to its immediate neighbors $j \in \mathcal{N}_i$, and in a similar way receives the previous state α_j^{l-1} from nodes $j \in \mathcal{N}_i$. Then it updates its current state by the following fusion rule.

$$\alpha_i^l = \alpha_i^{l-1} + \sum_{j \in \mathcal{N}_i} \pi_{i,j} \left(\alpha_j^{l-1} - \alpha_i^{l-1} \right) \tag{4}$$

Here, $\pi_{i,j}$ is the consensus weight, which should satisfy certain conditions to ensure convergence to the mean of initial values [25,28,33]. A sufficient and necessary condition guaranteeing finite-time weighted average consensus has been provided in [34]. In the derivation of the proposed DHIWCF, the average consensus protocol is involved in the state update step, hence we only discuss the design of consensus weights ensuring average consensus on local estimates of all the nodes in the network.

If it is possible for the above protocol in (4) to iterate for infinite times, the estimated state of all nodes in the network will asymptotically converge to the average value, that is, $\lim\limits_{l \to \infty} \alpha_i^l = \sum_{i=1}^{N_s} \alpha_i^0 / N_s$. In the original KCF, the consensus rate is set to be a constant value $\varepsilon \in (0, 1/d_{\max})$, where $d_{\max}$ is the maximum node degree in the network [6].

Remark 2. *A larger ε will accelerate the convergence of the protocol, but a ε equal to or more than $1/d_{max}$ will render the protocol unstable [9,16]. The constant ε treats states from different nodes with the same weights, which may slow down the convergence rate of the whole system. The choice of ε depends on d_{max}, which is not*

always available, especially in sparse sensor networks with time-varying communication topologies. In addition, there is no theoretical analysis on how to choose such a constant consensus weight.

To avoid the requirements for global parameters and speed up convergence rate, the Metropolis weights determine consensus rate between neighboring nodes based on their local node degree. As is discussed in [30], Metropolis weights enable the protocol in (4) to achieve convergence faster. The definition of Metropolis weights is

$$\pi_{ij} = \begin{cases} \left(1 + \max\{d_i, d_j\}\right)^{-1}, & \text{if } \{i, j\} \in \mathcal{E} \\ 1 - \sum_{\{i, j\} \in \mathcal{E}} \pi_{ij}, & \text{if } i = j \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

Remark 3. *The above definition in (5) indicates that a node with a larger degree will be assigned a smaller weight. All the consensus weights are computed only with the knowledge of local node degree, which is applicable to almost any kind of sensor networks. The interested reader is referred to [30] and the references therein for details. In this paper, the Metropolis weights are chosen for the proposed algorithm.*

The goal of the proposed DHIWCF is to achieve consensus on the local estimates of each node over the entire network by consensus iterations between neighboring nodes, and at the same time approach the estimation performance of CKF. If the network is fully connected, only a single iteration is enough to accomplish the estimation task. But in practical applications, a general network is often partially connected. To ensure the estimation accuracy and consensus on local estimates, it requires several iterations for the concerned information to spread throughout the entire network. However, due to constrained computation and communication resources, only a limited number of consensus iterations is available. It is urgent to design a more efficient distributed estimation scheme, which is able to achieve satisfactory estimation accuracy and consensus simultaneously with less consensus iterations.

3. Distributed Local MAP Estimation

This section starts with the centralized MAP estimator. Then we formulate the local generalized prior estimate equation based on prior estimates from the inclusive neighbors and the local generalized measurement equation based on the current measurements from the inclusive neighbors. By maximizing the local posterior probability, the local MAP estimator is derived. To implement the estimation steps in a distributed manner, approximation of the error cross covariance is required. Two special cases, where the prior errors from neighboring nodes are uncorrelated or completely identical, are considered here. The practical importance of such an approximation can be seen from the numerical examples in Section 6, which indicate that the proposed DHIWCF is effective even if the assumed cases are not fulfilled.

3.1. Global MAP Estimator

Assume $z_k = \left[z_{1,k}^T, z_{2,k}^T, \ldots, z_{N_s, k}^T \right]^T$ represents the collective measurements of the entire sensor network at time instant k. The stacked measurement matrix of all the nodes is denoted as $H_k = \left[H_{1,k}^T, H_{2,k}^T, \ldots, H_{N_s, k}^T \right]^T$. The stacked measurement noise is $v_k = \left[v_{1,k}^T, v_{2,k}^T, \ldots, v_{N_s, k}^T \right]^T$ with block diagonal covariance matrix $R_k = \text{blkdiag}\left(R_{1,k}, R_{2,k}, \ldots, R_{N_s,k} \right)$. Then the global measurement model can be formulated as

$$z_k = H_k x_k + v_k \tag{6}$$

Suppose the centralized prior estimate is $\hat{x}^c_{k|k-1}$. The corresponding prior estimation error is $e^c_{k|k-1} = \hat{x}^c_{k|k-1} - x_k$ with covariance matrix $P^c_{k|k-1} = \mathbb{E}\left\{e^c_{k|k-1}\left(e^c_{k|k-1}\right)^{\mathrm{T}}\right\}$. Let $\hat{x}^{\mathrm{MAP}}_{k|k}$ be the maximum a posterior (MAP) estimate, we have

$$\hat{x}^{\mathrm{MAP}}_{k|k} = \underset{x_k}{\mathrm{argmax}}\, p(x_k|Z_k) \tag{7}$$

where $p(z_k|Z_{k-1}) = \int p(z_k|x_k)p(x_k|Z_{k-1})\mathrm{d}x_k$ is a normalization constant. Since the process noise and measurement noise are both Gaussian, then the conditional PDF $p(z_k|x_k)$ and $p(x_k|Z_{k-1})$ are also Gaussian. The explicit form of the prior PDF $p(x_k|Z_{k-1})$ and the likelihood PDF $p(z_k|x_k)$ is formed as

$$p(x_k|Z_k) = \frac{p(z_k|x_k)p(x_k|Z_{k-1})}{p(z_k|Z_{k-1})} \tag{8}$$

where $p(z_k|Z_{k-1}) = \int p(z_k|x_k)p(x_k|Z_{k-1})\mathrm{d}x_k$ is a normalization constant. Since the process noise and measurement noise are both Gaussian, then the conditional PDF $p(z_k|x_k)$ and $p(x_k|Z_{k-1})$ are also Gaussian. The explicit form of the prior PDF $p(x_k|Z_{k-1})$ and the likelihood PDF $p(z_k|x_k)$ is formed as

$$p(x_k|Z_{k-1}) \propto \exp\left(-\frac{1}{2}\left(\hat{x}^c_{k|k-1} - x_k\right)^{\mathrm{T}}\left(P^c_{k|k-1}\right)^{-1}\left(\hat{x}^c_{k|k-1} - x_k\right)\right) \tag{9}$$

$$p(z_k|x_k) \propto \exp\left(-\frac{1}{2}(z_k - H_k x_k)^{\mathrm{T}} R_k^{-1}(z_k - H_k x_k)\right) \tag{10}$$

where $\|x\|_A^2 = x^{\mathrm{T}} A x$. Based on Gaussian product in the numerator, the criterion in (7) can be reformulated by minimizing the following cost function.

$$\hat{x}^{\mathrm{MAP}}_{k|k} = \underset{x_k}{\mathrm{argmin}}\left[(z_k - H_k x_k)^{\mathrm{T}} R_k^{-1}(z_k - H_k x_k) + \left(\hat{x}^c_{k|k-1} - x_k\right)^{\mathrm{T}}\left(P^c_{k|k-1}\right)^{-1}\left(\hat{x}^c_{k|k-1} - x_k\right)\right] \tag{11}$$

Here, the cost function in (11) is strictly convex on x_k and hence the optimal $\hat{x}^{\mathrm{MAP}}_{k|k}$ is available.

$$\hat{x}^{\mathrm{MAP}}_{k|k} = \left(\left(P^c_{k|k-1}\right)^{-1} + H_k^{\mathrm{T}} R_k^{-1} H_k\right)^{-1}\left(\left(P^c_{k|k-1}\right)^{-1}\hat{x}^c_{k|k-1} + H_k^{\mathrm{T}} R_k^{-1} z_k\right) \tag{12}$$

The corresponding posterior error covariance is

$$P^{\mathrm{MAP}}_{k|k} = \left(\left(P^c_{k|k-1}\right)^{-1} + H_k^{\mathrm{T}} R_k^{-1} H_k\right)^{-1} \tag{13}$$

The equivalent information form of the estimate in (12) and (13) can be rewritten as

$$\hat{y}^{\mathrm{MAP}}_{k|k} = \left(P^c_{k|k-1}\right)^{-1}\hat{x}^c_{k|k-1} + H_k^{\mathrm{T}} R_k^{-1} z_k \tag{14}$$

$$Y^{\mathrm{MAP}}_{k|k} = \left(P^c_{k|k-1}\right)^{-1} + H_k^{\mathrm{T}} R_k^{-1} H_k \tag{15}$$

3.2. Local MAP Estimation

Assume that each node, for instance, node i, is able to receive its neighbor's prior local estimate $\hat{x}_{j,k|k-1}$ and the corresponding covariance $P_{j,k|k-1}$, as well as its neighbor's local measurement $z_{j,k}$ and the corresponding noise covariance $R_{j,k}$ by local communication. The local generalized prior estimate, denoted by $\hat{x}'_{i,k|k-1}$, is defined as

$$\hat{x}'_{i,k|k-1} = \left[\hat{x}^{\mathrm{T}}_{i,k|k-1}, \hat{x}^{\mathrm{T}}_{j_1,k|k-1}, \ldots, \hat{x}^{\mathrm{T}}_{j_{d_i},k|k-1}\right]^{\mathrm{T}} \tag{16}$$

where $j_h \in \mathcal{N}_i(h = 1, 2, \ldots, d_i)$ denotes the index of node i's neighbors. Let $\eta_{i,k|k-1} = \hat{x}_{i,k|k-1} - x_k$ be the prior error of node i. The local collective prior error of node i with respect to its inclusive neighbors can be formulated as $\eta'_{i,k|k-1} = \left[\eta^{\mathrm{T}}_{i,k|k-1}, \eta^{\mathrm{T}}_{j_1,k|k-1}, \ldots, \eta^{\mathrm{T}}_{j_{d_i},k|k-1} \right]^{\mathrm{T}}$. The local generalized prior estimate can be expressed by

$$\hat{x}'_{i,k|k-1} = \mathcal{H}_I x_k + \eta'_{i,k|k-1} \tag{17}$$

where $\mathcal{H}_I = [\mathbf{I}_p, \mathbf{I}_p, \ldots, \mathbf{I}_p]^{\mathrm{T}}$ is the matrix stacked by $d_i + 1$ identity matrices. x_k is the true state at time instant k. The local collective prior error covariance of node i can be written as

$$\mathbf{P}'_{i,k|k-1} = \mathbb{E}\left\{ \eta'_{i,k|k-1}\left(\eta'_{i,k|k-1}\right)^{\mathrm{T}} \right\} = \begin{bmatrix} \mathbf{P}_{i,k|k-1} & \mathbf{P}_{ij_1,k|k-1} & \cdots & \mathbf{P}_{ij_{d_i},k|k-1} \\ \mathbf{P}_{j_1 i,k|k-1} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \mathbf{P}_{j_{d_i} i,k|k-1} & \cdots & \cdots & \mathbf{P}_{j_{d_i},k|k-1} \end{bmatrix} \tag{18}$$

Here, the block matrix $\mathbf{P}'_{i,k|k-1} \in \mathbb{R}^{(1+d_i)n_x \times (1+d_i)n_x}$.

Similarly, the local generalized measurement of node i with regard to its inclusive neighbors can be formulated as

$$z'_{i,\,k} = \mathbf{H}'_{i,\,k} x_k + v'_{i,k} \tag{19}$$

Here, $z'_{i,\,k} = \left[z^{\mathrm{T}}_{i,k}, z^{\mathrm{T}}_{j_1,k}, \ldots, z^{\mathrm{T}}_{j_{d_i},k} \right]^{\mathrm{T}}$ is the local generalized measurement. $\mathbf{H}'_{i,k} = \left[\mathbf{H}^{\mathrm{T}}_{i,\,k}, \mathbf{H}^{\mathrm{T}}_{j_1,\,k}, \ldots, \mathbf{H}^{\mathrm{T}}_{j_{d_i},\,k} \right]^{\mathrm{T}}$ is the local generalized measurement matrix. $v'_{i,k} = \left[v^{\mathrm{T}}_{i,\,k}, v^{\mathrm{T}}_{j_1,\,k}, \ldots, v^{\mathrm{T}}_{j_{d_i},\,k} \right]^{\mathrm{T}}$ denotes the local generalized measurement noise with covariance matrix $\mathbf{R}'_{i,k} = $ blkdiag$\left(\mathbf{R}_{i,k}, \mathbf{R}_{j_1,k}, \ldots, \mathbf{R}_{j_{d_i},k} \right)$.

Combining (17) and (19) together, one has

$$\begin{bmatrix} \hat{x}'_{i,k|k-1} \\ z'_{i,\,k} \end{bmatrix} = \begin{bmatrix} \mathcal{H}_I \\ \mathbf{H}'_{i,\,k} \end{bmatrix} x_k + \begin{bmatrix} \eta'_{i,k|k-1} \\ v'_{i,k} \end{bmatrix} \tag{20}$$

where the error covariance

$$\mathbb{E}\left\{ \begin{bmatrix} \eta'_{i,k|k-1} \\ v'_{i,k} \end{bmatrix} \begin{bmatrix} \eta'_{i,k|k-1} \\ v'_{i,k} \end{bmatrix}^{\mathrm{T}} \right\} = \text{blkdiag}\left(\mathbf{P}'_{i,k|k-1}, \mathbf{R}'_{i,k} \right) \tag{21}$$

Here the operator blkdiag$(\,\cdot\,)$ denotes the block diagonal matrix.

According to the derivation of the global maximum a posterior estimator described in Section 3.1, the updated local information matrix can be computed by

$$\begin{aligned} Y^{\text{LMAP}}_{i,k|k} &= \begin{bmatrix} \mathcal{H}_I \\ \mathbf{H}'_{i,\,k} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \mathbf{P}'_{i,k|k-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}'_{i,k} \end{bmatrix}^{-1} \begin{bmatrix} \mathcal{H}_I \\ \mathbf{H}'_{i,\,k} \end{bmatrix} \\ &= \mathcal{H}^{\mathrm{T}}_I \left(\mathbf{P}'_{i,k|k-1}\right)^{-1} \mathcal{H}_I + \left(\mathbf{H}'_{i,k}\right)^{\mathrm{T}} \left(\mathbf{R}'_{i,k}\right)^{-1} \mathbf{H}'_{i,k} \\ &= \sum_{r=1}^{1+d_i} \sum_{s=1}^{1+d_i} \left[\left(\mathbf{P}'_{i,k|k-1}\right)^{-1} \right]_{r,s} + \sum_{j \in \mathcal{J}_i} \left(\mathbf{H}_{j,\,k}\right)^{\mathrm{T}} \left(\mathbf{R}_{j,k}\right)^{-1} \mathbf{H}_{j,\,k} \end{aligned} \tag{22}$$

Similarly, the updated local information vector is

$$
\begin{aligned}
\hat{y}_{i,k|k}^{\text{LMAP}} &= \begin{bmatrix} \mathcal{H}_I \\ H'_{i,k} \end{bmatrix}^{\text{T}} \begin{bmatrix} P'_{i,k|k-1} & 0 \\ 0 & R'_{i,k} \end{bmatrix}^{-1} \begin{bmatrix} \hat{x}'_{i,k|k-1} \\ z'_{i,k} \end{bmatrix} \\
&= \mathcal{H}_I^{\text{T}} \left(P'_{i,k|k-1}\right)^{-1} \hat{x}'_{i,k|k-1} + \left(H'_{i,k}\right)^{\text{T}} \left(R'_{i,k}\right)^{-1} z'_{i,k} \\
&= \sum_{r=1}^{1+d_i} \sum_{s=1}^{1+d_i} \left[\left(P'_{i,k|k-1}\right)^{-1}\right]_{r,s} \left[\hat{x}'_{i,k|k-1}\right]_s + \sum_{j\in\mathcal{J}_i} \left(H_{j,\,k}\right)^{\text{T}} R_{j,k}^{-1} z_{j,\,k}
\end{aligned}
\tag{23}
$$

Here, $\left[\left(P'_{i,k|k-1}\right)^{-1}\right]_{r,s}$ denotes the (r,s)-th block matrix of $\left(P'_{i,k|k-1}\right)^{-1}$. Similarly, $\left[\hat{x}'_{i,k|k-1}\right]_s$ denotes the s-th block vector of $\hat{x}'_{i,k|k-1}$.

3.3. Approximation of $\left(P'_{i,k|k-1}\right)^{-1}$

It is shown in (22) and (23) that the key to acquire the local posterior estimate is to compute the inverse of the local collective prior error covariance, that is, $\left(P'_{i,k|k-1}\right)^{-1}$. However, as is shown in (18), the computation of $\left(P'_{i,k|k-1}\right)^{-1}$ requires the knowledge of cross-covariance between neighbors of node i. As is shown in [6], to compute the cross-covariance matrix $P_{ij_h,k|k-1}$, the information of the neighbors of node j_h is also required. Therefore, it is not practical to directly compute $\left(P'_{i,k|k-1}\right)^{-1}$ due to the fact that large amounts of communication among neighboring nodes are required, which may cause tremendous burden on computation and communication for the networked system. Although some work has been done in [35,36] to incorporate cross-covariance information into the estimation framework, no technique for computing the required terms are offered and predefined values are used instead [4].

Therefore, an approximation of $P'_{i,k|k-1}$ in a distributed manner is necessary. In the following derivation, two special cases are discussed. The first case is that the prior estimates from different nodes are completely uncorrelated with each other. This is true at the beginning of the estimation procedure when the prior information are initialized with random quantities. The second case is for converged priors, which is critical for the reason that with sufficient consensus iterations, the prior estimates from all nodes will converge to the identical value.

3.3.1. Case 1: Uncorrelated Priors

In this case, the prior errors from different nodes are assumed to be uncorrelated with each other, i.e., $\mathbb{E}\left\{\eta_{i,k|k-1}\eta_{j_h,k|k-1}^{\text{T}}\right\} = 0$. Hence, $P'_{i,k|k-1}$ in (18) turns into a block diagonal matrix $P'_{i,k|k-1} = \text{blkdiag}\left(P_{i,k|k-1}, P_{j_1,k|k-1}, \ldots, P_{j_{d_i},k|k-1}\right)$. The local posterior estimate in (22) and (23) can be approximated as

$$
\hat{y}_{i,k|k}^{\text{LMAP}} = \sum_{j\in\mathcal{J}_i} Y_{j,k|k-1}\hat{x}_{j,k|k-1} + \sum_{j\in\mathcal{J}_i} \left(H_{j,\,k}\right)^{\text{T}} R_{j,k}^{-1} z_{j,\,k}
\tag{24}
$$

$$
Y_{i,k|k}^{\text{LMAP}} = \sum_{j\in\mathcal{J}_i} Y_{j,k|k-1} + \sum_{j\in\mathcal{J}_i} \left(H_{j,\,k}\right)^{\text{T}} R_{j,k}^{-1} H_{j,\,k}
\tag{25}
$$

Note that after enough consensus iterations, the prior estimates of each node in the network asymptotically converges to the centralized result, i.e., $Y_{i,k|k-1} = Y_{c,k|k-1}$ and $\hat{y}_{i,k|k-1} = \hat{y}_{c,k|k-1}$. In such a case, the local prior information matrix in (25) turns into $\sum_{j\in\mathcal{J}_i} Y_{j,k|k-1} = (1+d_i)Y_{c,k|k-1}$. However, after convergence, the total amount of prior information in the network is $Y_{c,k|k-1}$. That is, the local prior information matrix in the inclusive neighborhood is overestimated by a factor $(1+d_i)$. Therefore, the approximation of $P'_{i,k|k-1}$ should be modified by multiplying a factor $(1+d_i)$, which is $P'_{i,k|k-1} =$

$(1 + d_i)\text{blkdiag}\left(\boldsymbol{P}_{i,k|k-1}, \boldsymbol{P}_{j_1,k|k-1}, \ldots, \boldsymbol{P}_{j_{d_i},k|k-1}\right)$ to avoid underestimation of the prior covariance. Hence, the results in (24) and (25) should be modified as

$$Y_{i,k|k} = \frac{1}{1 + d_i} \sum_{j \in \mathcal{J}_i} Y_{j,k|k-1} + \sum_{j \in \mathcal{J}_i} \left(\boldsymbol{H}_{j,\,k}\right)^{\text{T}} \boldsymbol{R}_{j,k}^{-1} \boldsymbol{H}_{j,\,k} \tag{26}$$

$$\hat{y}_{i,k|k} = \frac{1}{1 + d_i} \sum_{j \in \mathcal{J}_i} J_{j,k|k-1} \hat{x}_{j,k|k-1} + \sum_{j \in \mathcal{J}_i} \left(\boldsymbol{H}_{j,\,k}\right)^{\text{T}} \boldsymbol{R}_{j,k}^{-1} z_{j,\,k} \tag{27}$$

3.3.2. Case 2: Converged Priors

When the prior estimate of each node converges to the centralized result, one has

$$\sum_{r=1}^{1+d_i} \sum_{s=1}^{1+d_i} \left[\left(\boldsymbol{P'}_{i,k|k-1}\right)^{-1} \right]_{r,s} = Y_{c,k|k-1} = \frac{1}{1 + d_i} \sum_{j \in \mathcal{J}_i} Y_{c,k|k-1} \tag{28}$$

Note that for converged priors, $Y_{j,k|k-1} = Y_{c,k|k-1}$, $j \in \mathcal{J}_i$. Substituting this fact into (28), there is

$$\sum_{r=1}^{1+d_i} \sum_{s=1}^{1+d_i} \left[\left(\boldsymbol{P'}_{i,k|k-1}\right)^{-1} \right]_{r,s} = \frac{1}{1 + d_i} \sum_{j \in \mathcal{J}_i} Y_{j,k|k-1} \tag{29}$$

Therefore, the estimated results in (22) and (23) can be transformed into the weighted summation of the prior information and current measurement innovations, which are the same forms as the results shown in (26) and (27).

Remark 4. *It should be noted that the assumed cases are not always satisfied in realistic applications, but it is still of great significance for distributed filtering algorithms. The effectiveness and feasibility of such an approximation is evaluated by numerical examples in Section 6.*

4. Hybrid Information Weighted Consensus Filter

In Section 3, the prior estimate of each node is assumed to be known. Here, the prediction step is given.

$$\hat{x}_{i,k|k-1} = F_{k-1} \hat{x}_{i,k-1|k-1} \tag{30}$$

$$Y_{i,k|k-1} = \left(F_{k-1} Y_{i,\,k-1|k-1}^{-1} F_{k-1}^{\text{T}} + Q_{k-1}\right)^{-1} \tag{31}$$

For simplicity, the prediction step in (31) can be rewritten as

$$Y_{i,k|k-1} = \Psi_{k-1}\left(Y_{i,\,k-1|k-1}\right) \tag{32}$$

with

$$\Psi_{k-1}\left(Y_{i,\,k-1|k-1}\right) = \left(F_{k-1} Y_{i,\,k-1|k-1}^{-1} F_{k-1}^{\text{T}} + Q_{k-1}\right)^{-1} \tag{33}$$

The corresponding prior information vector is

$$\hat{y}_{i,k|k-1} = Y_{i,k|k-1} \hat{x}_{i,k|k-1} \tag{34}$$

With the above prediction steps and a weighted consensus protocol incorporated into the distributed local MAP estimator, a novel state estimation algorithm is obtained. Since the prior estimates and the measurement innovation are fused with different schemes, the proposed algorithm

is referred as distributed hybrid information weighted consensus filter (DHIWCF). The recursive form of DHIWCF is detailed in Algorithm 1.

Algorithm 1. DHIWCF implemented by node i at time instant k.

---

1. Obtain the local measurement $z_{i,\,k}$ with covariance matrix $R_{i,k}$.
2. Compute the measurement contribution vector and contribution matrix.

$$
\begin{cases}
\boldsymbol{u}_i = \boldsymbol{H}_{i,k}^{\mathrm{T}} \boldsymbol{R}_{i,k}^{-1} \boldsymbol{z}_{i,k} \\
\boldsymbol{U}_i = \boldsymbol{H}_{i,k}^{\mathrm{T}} \boldsymbol{R}_{i,k}^{-1} \boldsymbol{H}_{i,k}
\end{cases}
\tag{35}
$$

3. Broadcast state message $\{\boldsymbol{y}_{i,k|k-1}, \boldsymbol{Y}_{i,k|k-1}, \boldsymbol{u}_i, \boldsymbol{U}_i\}$ to its neighboring nodes $j \in \mathcal{N}_i$.
4. Receive state message $\{\boldsymbol{y}_{j,k|k-1}, \boldsymbol{Y}_{j,k|k-1}, \boldsymbol{u}_j, \boldsymbol{U}_j\}$ from its neighboring nodes $j \in \mathcal{N}_i$.
5. Compute the initial values.

$$
\begin{cases}
\hat{\boldsymbol{y}}_{i,k|k}^0 = \frac{1}{1+d_i} \sum\limits_{j \in \mathcal{J}_i} \hat{\boldsymbol{y}}_{j,k|k-1} + \sum\limits_{j \in \mathcal{J}_i} \left(\boldsymbol{H}_{j,\,k}\right)^{\mathrm{T}} \boldsymbol{R}_{j,k}^{-1} \boldsymbol{z}_{j,\,k} \\
\boldsymbol{Y}_{i,k|k}^0 = \frac{1}{1+d_i} \sum\limits_{j \in \mathcal{J}_i} \boldsymbol{Y}_{j,k|k-1} + \sum\limits_{j \in \mathcal{J}_i} \left(\boldsymbol{H}_{j,\,k}\right)^{\mathrm{T}} \boldsymbol{R}_{j,k}^{-1} \boldsymbol{H}_{j,\,k}
\end{cases}
\tag{36}
$$

6. Perform consensus.
    for $l = 1 : L$ **do**

      (1)   Send $\hat{\boldsymbol{y}}_{i,\,k|k-1}^{l-1}$ and $\boldsymbol{Y}_{i,\,k|k-1}^{l-1}$ to its neighbors $j \in \mathcal{N}_i$;
      (2)   Receive $\hat{\boldsymbol{y}}_{j,k|k-1}^{l-1}$ and $\boldsymbol{Y}_{j,k|k-1}^{l-1}$ from its neighbors $j \in \mathcal{N}_i$;
      (3)   Update its consensus state.

$$
\begin{cases}
\hat{\boldsymbol{y}}_{i,\,k|k}^l = \hat{\boldsymbol{y}}_{i,\,k|k}^{l-1} + \sum\limits_{j \in \mathcal{N}_i} \pi_{i,j}\left(\hat{\boldsymbol{y}}_{j,\,k|k}^{l-1} - \hat{\boldsymbol{y}}_{i,\,k|k}^{l-1}\right) \\
\boldsymbol{Y}_{i,\,k|k}^l = \boldsymbol{Y}_{i,\,k|k}^{l-1} + \sum\limits_{j \in \mathcal{N}_i} \pi_{i,j}\left(\boldsymbol{Y}_{j,\,k|k}^{l-1} - \boldsymbol{Y}_{i,\,k|k}^{l-1}\right)
\end{cases}
\tag{37}
$$

    end for
7. Compute the posterior estimate.
$$
\hat{\boldsymbol{x}}_{i,k|k} = \left(\boldsymbol{Y}_{i,\,k|k}^L\right)^{-1} \hat{\boldsymbol{y}}_{i,\,k|k'}^L \; \boldsymbol{Y}_{i,k|k} = \boldsymbol{Y}_{i,k|k'}^L \; \hat{\boldsymbol{y}}_{i,\,k|k} = \hat{\boldsymbol{y}}_{i,\,k|k}^L
$$
8. Prediction at time instant $k+1$
$$
\hat{\boldsymbol{x}}_{i,k+1|k} = \boldsymbol{F}_k \hat{\boldsymbol{x}}_{i,k|k'} \; \boldsymbol{Y}_{i,k+1|k} = \Psi_k\left(\boldsymbol{Y}_{i,\,k|k}\right), \hat{\boldsymbol{y}}_{i,k+1|k} = \boldsymbol{Y}_{i,k+1|k} \hat{\boldsymbol{x}}_{i,k+1|k}
$$

---

5. Performance Analysis

5.1. Consistency of Estimates

One of the most fundamental but important properties of a recursive filtering algorithm is that the estimated error statistics should be consistent with the true estimation errors. The approximated error covariance of an inconsistent filtering algorithm is too small or optimistic, which does not really indicate the uncertainty of the estimate and may result in divergence since subsequent measurements in this case are prone to be neglected [28].

Definition 2 [28,30,37,38]**.** *Consider a random vector* $\boldsymbol{x}$. *Let* $\hat{\boldsymbol{x}}$ *and* $\boldsymbol{P}$ *be, respectively, the estimate of* $\boldsymbol{x}$ *and the estimate of the corresponding error covariance. Then the pair* $(\hat{\boldsymbol{x}}, \boldsymbol{P})$ *is said to be consistent if*

$$
\mathbb{E}\left\{(\hat{\boldsymbol{x}} - \boldsymbol{x})(\hat{\boldsymbol{x}} - \boldsymbol{x})^{\mathrm{T}}\right\} \leq \boldsymbol{P}
\tag{38}
$$

It is shown in (38) that consistency requires that the true error covariance should be upper bounded (in the positive definite sense) by the approximated error covariance $\boldsymbol{P}$. In the distributed estimation

paradigm, due to the unaware reuse of the redundant data in the consensus iteration and the possible correlation between measurements from different nodes, the filter may suffer from inconsistency and divergence. In such a case, preservation of consistency is even much more important.

For convenience, consider the information pair $(\hat{y}, Y)$, where $\hat{y} = P^{-1}\hat{x}$ and $Y = P^{-1}\hat{x}$. The consistency defined by (38) can be rewritten as

$$Y \le \left(\mathbb{E}\left\{ \left(Y^{-1}\hat{y} - x\right)\left(Y^{-1}\hat{y} - x\right)^{\mathrm{T}} \right\} \right)^{-1} \tag{39}$$

Assumption 3. *The initialized estimate of each node, represented by* $(\hat{x}_{i,0|0}, P_{i,0|0}), i \in \mathcal{S}$, *is consistent. Equivalently, inequality* $P_{i,0|0} \ge \mathbb{E}\left\{ \left(\hat{x}_{i,0|0} - x_0\right)\left(\hat{x}_{i,0|0} - x_0\right)^{\mathrm{T}} \right\}$ *holds for* $i \in \mathcal{S}$.

Remark 5. *In general, Assumption 3 can be easily satisfied. The initial information on the state vector can be acquired in an off-line fashion before the fusion process. In the worst case where no prior information is available, each node can simply set the initialized information matrix as* $P_{i,0|0}^{-1} = \mathbf{0}$, *which implies infinite estimate uncertainty in each node at the beginning so that Assumption 3 is fulfilled.*

Assumption 4. *The system matrix* F_k *is invertible.*

Lemma 1 [28]**.** *Under Assumption 4, if two positive semidefinite matrices* Y_1 *and* Y_2 *satisfy* $Y_1 \le Y_2$, *then* $0 \le \Psi_k(Y_1) \le \Psi_k(Y_2)$. *In other words, the function* $\Psi_k(\,\cdot\,)$ *is monotonically nondecreasing for any* $k \ge 0$.

Theorem 1. *Let Assumptions 1, 2, and 3 hold. Then, for each time instant k and each node* $i \in \mathcal{S}$, *the information pair* $\left(\hat{y}_{i,k|k}, Y_{i,k|k}\right)$ *of the DHIWCF is consistent in that*

$$Y_{i,k|k} \le \left(\mathbb{E}\left\{ \left(\hat{x}_{i,k|k} - x_k\right)\left(\hat{x}_{i,k|k} - x_k\right)^{\mathrm{T}} \right\} \right)^{-1} \tag{40}$$

with $\hat{x}_{i,k|k} = Y_{i,k|k}^{-1}\hat{y}_{i,\,k|k}$.

Proof. An inductive method is utilized here to prove this theorem. It is supposed that, at time instant $k-1$

$$Y_{i,k-1|k-1} \le \left(\mathbb{E}\left\{ \left(\hat{x}_{i,k-1|k-1} - x_{k-1}\right)\left(\hat{x}_{i,k-1|k-1} - x_{k-1}\right)^{\mathrm{T}} \right\} \right)^{-1} \tag{41}$$

for any $i \in \mathcal{S}$. For brevity, the predicted information matrix in (31) can be rewritten as

$$Y_{i,k|k-1} = \Psi\left(Y_{i,\,k-1|k-1}\right) \tag{42}$$

On the basis of Lemma 1, it is immediate to see

$$\begin{aligned} Y_{i,k|k-1} &= \Psi_{k-1}\left(Y_{i,\,k-1|k-1}\right) \\ &\le \Psi_{k-1}\left(\left(\mathbb{E}\left\{ \left(\hat{x}_{i,k-1|k-1} - x_{k-1}\right)\left(\hat{x}_{i,k-1|k-1} - x_{k-1}\right)^{\mathrm{T}} \right\} \right)^{-1} \right) = \left(\mathbb{E}\left\{ \left(\hat{x}_{i,k|k-1} - x_k\right)\left(\hat{x}_{i,k|k-1} - x_k\right)^{\mathrm{T}} \right\} \right)^{-1} \end{aligned} \tag{43}$$

According to (26) and (27), the local estimation error is

$$\begin{aligned} \hat{x}_{i,k|k}^0 - x_k &= \left(Y_{i,k|k}^0\right)^{-1}\hat{y}_{i,k|k}^0 - \left(Y_{i,k|k}^0\right)^{-1}Y_{i,k|k}^0 x_k \\ &= \left(Y_{i,k|k}^0\right)^{-1}\left[\frac{1}{1+d_i} \sum_{j \in \mathcal{J}_i} Y_{j,k|k-1}\left(\hat{x}_{j,k|k-1} - x_k\right) + \sum_{j \in \mathcal{J}_i} \left(H_{j,\,k}\right)^{\mathrm{T}} R_{j,k}^{-1}\left(z_{j,\,k} - H_{j,\,k}x_k\right) \right] \\ &= \left(Y_{i,k|k}^0\right)^{-1}\left[\frac{1}{1+d_i} \sum_{j \in \mathcal{J}_i} Y_{j,k|k-1}\left(\hat{x}_{j,k|k-1} - x_k\right) + \sum_{j \in \mathcal{J}_i} \left(H_{j,\,k}\right)^{\mathrm{T}} R_{j,k}^{-1}v_{j,\,k} \right] \end{aligned} \tag{44}$$

Then,

$$\left(\mathbb{E}\left\{\left(\hat{x}_{i,k|k}^0 - x_k\right)\left(\hat{x}_{i,k|k}^0 - x_k\right)^{\mathrm{T}}\right\}\right)^{-1} = \left(\left(Y_{i,k|k}^0\right)^{-1}\left(\mathbb{E}\{\Delta_{k,i}\} + \sum_{j\in\mathcal{J}_i}\left(H_{j,\,k}\right)^{\mathrm{T}}R_{j,k}^{-1}H_{j,\,k}\right)\left(Y_{i,k|k}^0\right)^{-1}\right)^{-1} \quad (45)$$

where

$$\Delta_{k,i} = \left(\sum_{j\in\mathcal{J}_i}\frac{1}{1+d_i}Y_{j,k|k-1}\left(\hat{x}_{j,k|k-1} - x_k\right)\right)\left(\sum_{j\in\mathcal{J}_i}\frac{1}{1+d_i}Y_{j,k|k-1}\left(\hat{x}_{j,k|k-1} - x_k\right)\right)^{\mathrm{T}} \quad (46)$$

According to the consistency property of covariance intersection [29,38], it holds that

$$\mathbb{E}\{\Delta_{k,i}\} \le \sum_{j\in\mathcal{J}_i}\frac{1}{1+d_i}Y_{j,k|k-1}\mathbb{E}\left\{\left(\hat{x}_{j,k|k-1} - x_k\right)\left(\hat{x}_{j,k|k-1} - x_k\right)^{\mathrm{T}}\right\}Y_{j,k|k-1} \le \frac{1}{1+d_i}\sum_{j\in\mathcal{J}_i}Y_{j,k|k-1} \quad (47)$$

Then, exploiting (47) and (43) in (45), the following result is obtained.

$$\left(\mathbb{E}\left\{\left(\hat{x}_{i,k|k}^0 - x_k\right)\left(\hat{x}_{i,k|k}^0 - x_k\right)^{\mathrm{T}}\right\}\right)^{-1} \ge \left(\left(Y_{i,k|k}^0\right)^{-1}\left(\frac{1}{1+d_i}\sum_{j\in\mathcal{J}_i}Y_{j,k|k-1} + \sum_{j\in\mathcal{J}_i}\left(H_{j,\,k}\right)^{\mathrm{T}}R_{j,k}^{-1}H_{j,\,k}\right)\left(Y_{i,k|k}^0\right)^{-1}\right)^{-1} \ge Y_{i,k|k}^0 \quad (48)$$

Since the information pair $\left(\hat{x}_{i,k|k}^{l+1}, Y_{i,k|k}^{l+1}\right)$ is computed based on the previous information pair $\left(\hat{x}_{i,k|k}^l, Y_{i,k|k}^l\right)$ by (3), and the covariance intersection involved in (3) preserves the consistency of estimates [29,37–39], it can be concluded that $\left(\mathbb{E}\left\{\left(\hat{x}_{i,k|k}^l - x_k\right)\left(\hat{x}_{i,k|k}^l - x_k\right)^{\mathrm{T}}\right\}\right)^{-1} \ge Y_{i,k|k}^l$ indicates $\left(\mathbb{E}\left\{\left(\hat{x}_{i,k|k}^{l+1} - x_k\right)\left(\hat{x}_{i,k|k}^{l+1} - x_k\right)^{\mathrm{T}}\right\}\right)^{-1} \ge Y_{i,k|k}^{l+1}$ for any $l = 1,\dots,L$. In other words, if the estimate obtained with l consensus iterations is consistent, the estimate obtained with $l + 1$ consensus iterations is also consistent. Therefore, it is straightforward to conclude that (40) holds with $\hat{x}_{i,k|k} = \hat{x}_{i,k|k}^L$ and $Y_{i,k|k} = Y_{i,k|k}^L$. The proof is concluded since the initial estimate $\hat{x}_{i,0|0}, \forall i \in \mathcal{S}$ is consistent. □

5.2. Boundedness of Error Covariances

According to the consistency of the proposed DHIWCF in Theorem 1, it is sufficient to prove that $Y_{i,k|k}$ is lower bounded by a certain positive matrix (or equivalently, to prove $P_{i,k|k} = Y_{i,k|k}^{-1}$ is upper bounded by some constant matrix) for the proof of the boundedness of the error covariance $\mathbb{E}\left\{\left(\hat{x}_{i,k|k} - x_k\right)\left(\hat{x}_{i,k|k} - x_k\right)^{\mathrm{T}}\right\}$. To derive the bounds for the information matrix $Y_{i,k|k}$, The following assumptions are required.

Assumption 5. *The system is collectively observable. That is, the pair (F_k, H_k) is observable where $H_k = \mathrm{col}\left(H_{i,\,k}, \ i \in \mathcal{S}\right)$.*

Let $\mathbf{\Pi}$ be the consensus matrix, whose elements are the consensus weights $\pi_{i,j}$ for any $i, j \in \mathcal{S}$. Further, let $\pi_{i,j}^L$ be the (i, j)-th element of $\mathbf{\Pi}^L$, which is the L-th power of $\mathbf{\Pi}$.

Assumption 6. *The consensus matrix $\mathbf{\Pi}$ is row stochastic and primitive.*

Assumption 7. *There exist real numbers $\underline{f}, \overline{f}, \underline{h}, \overline{h} \ne 0$ and positive real numbers $\overline{p} > \underline{p} > 0, \overline{q} > \underline{q} > 0$, such that the following bounds are fulfilled for each $k \ge 0$, $i \in \mathcal{S}$.*

$$\begin{cases} \underline{f}^2 I_n \le F_k F_k^{\mathrm{T}} \le \overline{f}^2 I_n, & \underline{h}^2 I_m \le H_{i,k}\left(H_{i,k}\right)^{\mathrm{T}} \le \overline{h}^2 I_m \\ \underline{q} I_n \le Q_k \le \overline{q} I_n, & \underline{r} I_m \le R_{i,k} \le \overline{r} I_m \end{cases} \quad (49)$$

Lemma 2 [28]. *Under Assumptions 4 and 5, and the proposed DHIWCF algorithm, if there exists a positive semidefinite matrix $\breve{Y}$ such that $Y_{i,k|k} \leq \breve{Y}$, $\forall k \geq 0$, $i \in \mathcal{S}$, then there always exists a strictly positive constant $0 < \alpha < 1$ such that*

$$Y_{i,k+1|k} \geq \alpha (F_k)^{-T} Y_{i,\,k|k} F_k^{-1} \tag{50}$$

By virtue of Lemma 2, Theorem 2 which depicts the boundedness of error covariances is presented below.

Theorem 2. *Let Assumptions 4–7 hold, there exist positive definite matrices $\underline{\Omega}$ and $\overline{\Omega}$ such that*

$$0 < \underline{\Omega} \leq Y_{i,\,k|k} \leq \overline{\Omega}, \quad \forall k \geq 0, \ i \in \mathcal{S} \tag{51}$$

where $Y_{i,\,k|k}$ is the information matrix given by the proposed DHIWCF.

Proof. For simplicity, the proof is concluded for the case $L = 1$. The generalization for $L > 1$ can be directly derived in a similar way. According to the proposed DHIWCF, the information matrix for node i at time instant k can be written as

$$Y_{i,k|k} = \sum_{j \in \mathcal{S}} \sum_{r \in \mathcal{J}_j} \pi_{i,j} (1 + d_j)^{-1} Y_{r,k|k-1} + \sum_{j \in \mathcal{S}} \sum_{r \in \mathcal{J}_j} \pi_{i,j} (H_{r,\,k})^T R_{r,k}^{-1} H_{r,\,k} \tag{52}$$

In view of Assumption 6, 7 and fact that $Y_{r,k|k-1} \leq Q_{k-1}^{-1}$ by (31), one can get

$$\begin{aligned} Y_{i,k|k} &\leq Q_{k-1}^{-1} + \sum_{j \in \mathcal{S}} \sum_{r \in \mathcal{J}_j} \pi_{i,j} (H_{r,\,k})^T R_{r,k}^{-1} H_{r,\,k} \\ &\leq \left(\frac{1}{\underline{q}} + \frac{(1 + d_{\max})\overline{h}^2}{\underline{r}} \right) I_{n_x} \triangleq \overline{\Omega} \end{aligned} \tag{53}$$

Hence, the upper bound is achieved. Next a lower bound will be guaranteed under Assumption 5.

According to Lemma 2 and Assumption 7 and (31), (53), it follows from (52) that

$$Y_{i,k|k} \geq \alpha \sum_{j \in \mathcal{S}} \sum_{r \in \mathcal{J}_j} \pi_{i,j} (1 + d_j)^{-1} (F_{k-1})^{-T} Y_{r,\,k-1|k-1} F_{k-1}^{-1} + \sum_{j \in \mathcal{S}} \sum_{r \in \mathcal{J}_j} \pi_{i,j} (H_{r,\,k})^T R_{r,k}^{-1} H_{r,\,k} \tag{54}$$

where α is a positive scalar with $0 < \alpha < 1$. By recursively exploiting (52) and (54) for a certain number (denoted by $\bar{k}$) of times, there is

$$\begin{aligned} Y_{i,k|k} \geq{} & \alpha^{\bar{k}} \sum_{j \in \mathcal{S}} \sum_{r \in \mathcal{S}} \pi_{i,j}^{\bar{k}} \Xi_{j,\,r}^{\bar{k}} \left(\underbrace{F_{k-1} F_{k-2} \ldots F_{k-\bar{k}}}_{length = \bar{k}} \right)^{-T} Y_{i,\,k-1|k-1} \left(F_{k-1} F_{k-2} \ldots F_{k-\bar{k}} \right)^{-1} \\ & + \sum_{\tau=1}^{\bar{k}} \alpha^{\tau-1} \left(\underbrace{F_{k-1} F_{k-2} \ldots F_{k-\tau+1}}_{length = \tau-1} \right)^{-T} \left(\sum_{j \in \mathcal{S}} \sum_{r \in \mathcal{S}} \pi_{i,j}^{\tau} \Xi_{j,\,r}^{\tau-1} (H_{r,\,k-\tau+1})^T R_{r,k-\tau+1}^{-1} H_{r,\,k-\tau+1} \right) \left(F_{k-1} F_{k-2} \ldots F_{k-\tau+1} \right)^{-1} \end{aligned} \tag{55}$$

where $\pi_{i,j}^{\tau}$ is the (i, j)-th element of Π^{τ}. Ξ is a matrix with elements

$$\Xi_{i,j} = \begin{cases} 1/(1 + d_i), & \text{if } j \in \mathcal{J}_i \\ 0, & \text{otherwise} \end{cases} \tag{56}$$

Note that the matrix Ξ is constructed based on the network topology and is naturally stochastic. According to [40,41], as long as the undirected network is connected, similar to the definition of Π, Ξ is primitive. Therefore, there exist strictly positive integers m and n such that all the elements of Π^s and Ξ^t are positive for $s \geq m$, $t \geq n$. Let us define

$$\Omega_1 = \sum_{\tau=1}^{\bar{k}} \alpha^{\tau-1} \underbrace{\left(F_{k-1}F_{k-2}\ldots F_{k-\tau+1}\right)}_{length=\,\tau-1}^{-T} \left(\sum_{j\in\mathcal{S}}\sum_{r\in\mathcal{S}} \pi_{i,j}^{\tau}\Xi_{j,\,r}^{\tau-1}\left(H_{r,\,k-\tau+1}\right)^{T}R_{r,k-\tau+1}^{-1}H_{r,\,k-\tau+1}\right)\left(F_{k-1}F_{k-2}\ldots F_{k-\tau+1}\right)^{-1} \quad (57)$$

It should be noted that, under Assumption 5, Ω_1 is definite positive for $\bar{k} \geq \max(m, n+1)$. Therefore, for $k \geq \bar{k}$, $Y_{i,k|k} \geq \Omega_1 > 0$. Since $\bar{k}$ is finite, for $0 \leq k \leq \bar{k}-1$, there exists a constant positive definite matrix Ω_2 such that $Y_{i,k|k} \geq \Omega_2 > 0$. Hence, there exists a positive definite matrix $\underline{\Omega}$ such that $0 < \underline{\Omega} \leq Y_{i,\,k|k}$. The proof is now complete. $\square$

Remark 6. *The result shown in Theorem 2 is only dependent on collective observability. This is distinct from some algorithms that require some sort of local observability or detectability condition [5,6,8,11,25], which poses a great challenge to the sensing abilities of sensors and restricts the scope of application.*

5.3. Convergence of Estimation Errors

In line with the boundedness of $Y_{i,k|k}$ proven in Theorem 2, the convergence of local estimation errors obtained by the proposed DHIWCF is analyzed in this section. To facilitate the analysis, the following preliminary lemmas are required.

Lemma 3 [26,28,31]**.** *Given an integer $N \geq 2$, N positive definite matrices $M_1, \ldots, M_N$ and N vectors $v_1, \ldots, v_N$, the following inequality holds*

$$\left(\sum_{i=1}^{N} M_i v_i\right)^{\top}\left(\sum_{i=1}^{N} M_i\right)^{-1}\left(\sum_{i=1}^{N} M_i v_i\right) \leq \sum_{i=1}^{N} v_i^{\top} M_i v_i \quad (58)$$

Lemma 4 [26,28]**.** *Under Assumptions 4 and 5, and the proposed DHIWCF algorithm, if there exists a positive semidefinite matrix $\widetilde{Y}$ such that $Y_{i,k|k} \geq \widetilde{Y}$, $\forall k \geq 0$, $i \in \mathcal{S}$, then there always exists a strictly positive scalar $0 < \beta < 1$ such that*

$$Y_{i,k+1|k} \leq \beta(F_k)^{-T}Y_{i,\,k|k}F_k^{-1} \quad (59)$$

For the sake of simplicity, let us denote the prediction and estimation error at node i by $\widetilde{x}_{i,k|k-1} = \hat{x}_{i,k|k-1} - x_k$ and $\widetilde{x}_{i,k|k} = \hat{x}_{i,k|k} - x_k$, respectively. The collective forms are, respectively, $\widetilde{x}_{k|k-1} = \text{col}\left(\widetilde{x}_{1,k|k-1}, \ldots, \widetilde{x}_{N_s,k|k-1}\right)$ and $\widetilde{x}_{k|k} = \text{col}\left(\widetilde{x}_{1,k|k}, \ldots, \widetilde{x}_{N_s,k|k}\right)$.

Theorem 3. *Under Assumptions 4–6, the proposed DHIWCF algorithm yields an asymptotic estimate in each node of the network in that*

$$\lim_{k\to+\infty} \mathbb{E}\left\{\hat{x}_{i,k|k} - x_k\right\} = 0, \quad \forall i \in \mathcal{S} \quad (60)$$

Proof. Under Assumptions 4–6, Theorem 2 holds. Therefore, $Y_{i,k|k}$ is uniformly lower and upper bounded. Let us define the following candidate Lyapunov function

$$V_{i,k}(x) = x^{\top}Y_{i,k|k-1}x, \quad i \in \mathcal{S} \quad (61)$$

By virtue of Lemma 2, it can be concluded that there exists a positive real number $0 < \widetilde{\beta} < 1$ such that

$$Y_{i,k+1|k} \leq \widetilde{\beta}(F_k)^{-T}Y_{i,\,k|k}F_k^{-1} \tag{62}$$

Then, one has

$$\begin{aligned} V_{i,\,k+1}\left(\mathbb{E}\{\widetilde{x}_{i,k+1|k}\}\right) &= \left(\mathbb{E}\{\widetilde{x}_{i,k+1|k}\}\right)^{T}Y_{i,k+1|k}\mathbb{E}\{\widetilde{x}_{i,k+1|k}\} \\ &\leq \widetilde{\beta}\left(\mathbb{E}\{\widetilde{x}_{i,k+1|k}\}\right)^{T}(F_k)^{-T}Y_{i,\,k|k}F_k^{-1}\mathbb{E}\{\widetilde{x}_{i,k+1|k}\} \end{aligned} \tag{63}$$

Since $\mathbb{E}\{\widetilde{x}_{i,k+1|k}\} = \mathbb{E}\{F_k\hat{x}_{i,k|k} - (F_kx_k + w_k)\} = F_k\mathbb{E}\{\widetilde{x}_{i,k|k}\}$, one can obtain

$$V_{i,\,k+1}\left(\mathbb{E}\{\widetilde{x}_{i,k+1|k}\}\right) \leq \widetilde{\beta}\left(\mathbb{E}\{\widetilde{x}_{i,k|k}\}\right)^{T}Y_{i,\,k|k}\mathbb{E}\{\widetilde{x}_{i,k|k}\} \tag{64}$$

Notice that

$$Y_{i,k|k} = Y^L_{i,k|k} = \sum_{j\in\mathcal{S}} \pi^L_{i,j}Y^0_{j,\,k|k} \tag{65}$$

$$\hat{y}_{i,k|k} = \hat{y}^L_{i,k|k} = \sum_{j\in\mathcal{S}} \pi^L_{i,j}\hat{y}^0_{j,\,k|k} \tag{66}$$

Here, pre-multiplying (65) by $Y^{-1}_{i,k|k}$ and post-multiplying it by x_k yields

$$x_k = Y^{-1}_{i,k|k}\sum_{j\in\mathcal{S}} \pi^L_{i,j}Y^0_{j,\,k|k}x_k \tag{67}$$

In a similar way, pre-multiplying (66) by $Y^{-1}_{i,k|k}$ yields

$$\hat{x}_{i,k|k} = Y^{-1}_{i,k|k}\sum_{j\in\mathcal{S}} \pi^L_{i,j}\hat{y}^0_{j,\,k|k} \tag{68}$$

Therefore,

$$\mathbb{E}\{\widetilde{x}_{i,k|k}\} = \mathbb{E}\{\hat{x}_{i,k|k} - x_k\} = Y^{-1}_{i,k|k}\sum_{j\in\mathcal{S}} \pi^L_{i,j}\mathbb{E}\{\hat{y}^0_{j,\,k|k} - Y^0_{j,\,k|k}x_k\} \tag{69}$$

According to (36), there is

$$\hat{y}^0_{j,\,k|k} - Y^0_{j,\,k|k}x_k = \left(1+d_j\right)^{-1}\sum_{r\in\mathcal{J}_j} Y_{r,k|k-1}\widetilde{x}_{r,k|k-1} + \sum_{r\in\mathcal{J}_j} \left(H_{r,\,k}\right)^{T}R^{-1}_{r,k}v_{r,\,k} \tag{70}$$

Since $\mathbb{E}\{v_{r,\,k}\} = 0$, one can get

$$\mathbb{E}\{\widetilde{x}_{i,k|k}\} = Y^{-1}_{i,k|k}\sum_{j\in\mathcal{S}}\sum_{r\in\mathcal{J}_j} \pi^L_{i,j}\left(1+d_j\right)^{-1}Y_{r,k|k-1}\mathbb{E}\{\widetilde{x}_{r,k|k-1}\} \tag{71}$$

Substituting (71) into (64) yields

$$\begin{aligned} &V_{i,\,k+1}\left(\mathbb{E}\{\widetilde{x}_{i,k+1|k}\}\right) \\ &\leq \widetilde{\beta}\left[\sum_{j\in\mathcal{S}}\sum_{r\in\mathcal{J}_j} \pi^L_{i,j}\left(1+d_j\right)^{-1}Y_{r,k|k-1}\mathbb{E}\{\widetilde{x}_{r,k|k-1}\}\right]^{T}Y^{-1}_{i,\,k|k}\left[\sum_{j\in\mathcal{S}}\sum_{r\in\mathcal{J}_j} \pi^L_{i,j}\left(1+d_j\right)^{-1}Y_{r,k|k-1}\mathbb{E}\{\widetilde{x}_{r,k|k-1}\}\right] \end{aligned} \tag{72}$$

Applying the fact that $Y_{i,\,k|k} \geq \sum_{j \in \mathcal{S}} \sum_{r \in \mathcal{J}_j} \pi_{i,j}^L (1 + d_j)^{-1} Y_{r,k|k-1}$ and Lemma 3 to the right hand side of (72), one can obtain that

$$
\begin{aligned}
V_{i,\,k+1}\left(\mathbb{E}\{\widetilde{x}_{i,k+1|k}\}\right) &\leq \widetilde{\beta} \sum_{j \in \mathcal{S}} \sum_{r \in \mathcal{J}_j} \pi_{i,j}^L (1 + d_j)^{-1} \left(\mathbb{E}\{\widetilde{x}_{r,k|k-1}\}\right)^{\mathrm{T}} Y_{r,k|k-1} \mathbb{E}\{\widetilde{x}_{r,k|k-1}\} \\
&= \widetilde{\beta} \sum_{j \in \mathcal{S}} \sum_{r \in \mathcal{J}_j} \pi_{i,j}^L (1 + d_j)^{-1} V_{r,\,k}\left(\mathbb{E}\{\widetilde{x}_{r,k|k-1}\}\right)
\end{aligned}
\tag{73}
$$

Writing (73) for $i = 1, 2, \ldots, N_s$ in a collective form, it turns out that

$$
V_{k+1}\left(\mathbb{E}\{\widetilde{x}_{k+1|k}\}\right) \leq \widetilde{\beta}\, \Pi^L \Xi\, V_k\left(\mathbb{E}\{\widetilde{x}_{k|k-1}\}\right)
\tag{74}
$$

where

$$
V_k\left(\mathbb{E}\{\widetilde{x}_{k|k-1}\}\right) = \mathrm{col}\left(V_{1,\,k}\left(\mathbb{E}\{\widetilde{x}_{i,k|k-1}\}\right), \ldots, V_{N_s,\,k}\left(\mathbb{E}\{\widetilde{x}_{N_s,k|k-1}\}\right)\right)
\tag{75}
$$

and Ξ is a matrix with elements satisfying

$$
\Xi_{i,j} = \begin{cases} 1/(1 + d_i), & \text{if } j \in \mathcal{J}_i \\ 0, & \text{otherwise} \end{cases}
\tag{76}
$$

Since the consensus matrix Π and the constructed matrix Ξ are both row stochastic, thus their spectral radiuses are both 1. As a consequence, for $0 < \widetilde{\beta} < 1$, the elements of vector $V_{k+1}\left(\mathbb{E}\{\widetilde{x}_{k+1|k}\}\right)$ vanishes as k tends to infinity in that $\lim_{k \to +\infty} \mathbb{E}\{\widetilde{x}_{k+1|k}\} = \mathbf{0}$. Due to the equation $\mathbb{E}\{\widetilde{x}_{k+1|k}\} = F_k \mathbb{E}\{\widetilde{x}_{i,k|k}\}$ and Assumption 4, it is straightforward to conclude that $\lim_{k \to +\infty} \mathbb{E}\{\hat{x}_{i,k|k} - x_k\} = \mathbf{0}$ for any $i \in \mathcal{S}$. □

Remark 7. *The Lyapunov function defined in (61) plays a crucial role in the convergence proof of the proposed algorithm, which can be easily extended to stability analysis of Kalman-like consensus filters in other scenarios. The reason for the non-singularity requirements of F_k in Theorem 3 is that the proof of the Lyapunov method depends on Lemma 4, the establishment of which needs the invertibility of F_k.*

6. Experimental Results and Analysis

6.1. Simulation Setting

A target tracking scenario is adopted here to validate the effectiveness and superiority of the proposed DHIWCF. The centralized Kalman filter (CKF) is chosen as a benchmark to compare the proposed DHIWCF with following algorithms: The Kalman consensus filter (KCF), the generalized Kalman consensus filter (GKCF), the information weighted consensus filter (ICF), the consensus on information algorithm (CI), the consensus on measurements algorithm (CM), the hybrid consensus on measurement + consensus on information algorithm (HCMCI).

In the surveillance area, a target is moving with the discrete time linear model shown in (1). $x_k = [x_k, y_k, \dot{x}_k, \dot{y}_k]^{\mathrm{T}}$ is the state vector at time instant k. (x_k, y_k) and $(\dot{x}_k, \dot{y}_k)$ are, respectively, the position and velocity components of the state. The state transition matrix F_k and the process noise covariance matrix Q_k are set as follows.

$$
F_k = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad Q_k = \begin{bmatrix} 10 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}
\tag{77}
$$

The initial position of the target is randomly located at the 500×500 space. The initial speed is set to 2 units per time step, with a random direction uniformly chosen from 0 to 2π. In each simulation

run, the initial prior error covariance is $P_0 = \text{diag}(100, 100, 10, 10)$, and all nodes in the network share the same P_0. The initial prior estimate of each node is generated by adding zero-mean Gaussian noise with covariance P_0 to the true initial state. The total number of time steps is $K = 100$ unless stated otherwise. The sampling time interval is $T = 1$ s.

The target of interest is observed by a number of networked sensors with measurement model shown in (2). The measurement matrix $H_{i,\,k}$ and the measurement noise covariance $R_{i,k}$ are given below.

$$H_{i,\,k} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \ R_{i,k} = \begin{bmatrix} 100 & 0 \\ 0 & 100 \end{bmatrix} \tag{78}$$

6.2. Performance Metrics

For a fair comparison, a total number $M_c = 200$ of independent Monte Carlo runs are carried out. The consensus iterations L is set from 1 to 10. The consensus rate parameter is selected as $\varepsilon = 0.65/d_{\max}$. For the proposed DHIWCF algorithm, the Metropolis weight matrix is chosen, which is computed only with knowledge of local node degree. Following metrics are chosen to evaluate the estimation performance from different aspects.

(1) The position root mean squared error (PRMSE), which indicates the tracking accuracy, is defined as

$$\text{PRMSE} = \sqrt{\frac{1}{M_c} \sum_{m=1}^{M_c} \frac{1}{N_s} \sum_{i=1}^{N_s} \left[\left(\hat{x}_{i,k}^m - x_k^m \right)^2 + \left(\hat{y}_{i,k}^m - y_k^m \right)^2 \right]} \tag{79}$$

where $\left(\hat{x}_{i,k}^m,\ \hat{y}_{i,k}^m \right)$ and $\left(x_k^m,\ y_k^m \right)$ are, respectively, the estimated position and the true position in the m-th Monte Carlo run.

(2) The averaged position root mean squared error (APRMSE), which implies the overall tracking accuracy of an algorithm over all simulation runs, all time instants and all sensors, is defined as

$$\text{APRMSE} = \sqrt{\frac{1}{M_c} \sum_{m=1}^{M_c} \frac{1}{K} \sum_{k=1}^{K} \frac{1}{N_s} \sum_{i=1}^{N_s} \left[\left(\hat{x}_{i,k}^m - x_k^m \right)^2 + \left(\hat{y}_{i,k}^m - y_k^m \right)^2 \right]} \tag{80}$$

(3) The averaged consensus estimate error (ACEE), which indicates the degree of consensus among estimates from all nodes in the network, is defined as

$$\text{ACEE} = \frac{1}{M_c} \sum_{m=1}^{M_c} \frac{1}{N_s(N_s-1)} \sum_{i=1}^{N_s} \sum_{j=1}^{N_s} \sqrt{\left[\left(\hat{x}_{j,k}^m - \hat{x}_{i,k}^m \right)^2 + \left(\hat{y}_{j,k}^m - \hat{y}_{i,k}^m \right)^2 \right]} \tag{81}$$

(4) The normalized estimation error squared (NEES), which is used to check for filter consistency, is defined as

$$\varepsilon_k = \left(x_k - \hat{x}_{k|k} \right)^{\text{T}} P_{k|k}^{-1} \left(x_k - \hat{x}_{k|k} \right) \tag{82}$$

where x_k and $\hat{x}_{k|k}$ are, respectively, the true state and estimated state. $P_{k|k}$ is the posterior covariance at time instant k. Suppose that the filter is consistent, the NEES is subject to Chi-squared distribution with n_x degrees of freedom. A way to check filter consistency is by testing the average NEES over a number of M_c Monte Carlo runs, i.e.,

$$\bar{\varepsilon}_k = \frac{1}{M_c} \sum_{i=1}^{M_c} \varepsilon_k^i \tag{83}$$

Under similar assumptions $M_c \bar{\varepsilon}_k$ will be Chi-squared distributed with $M_c n_x$ degrees of freedom. Suppose the acceptance interval is $[r_1,\ r_2]$, the Chi-square test is accepted if $\bar{\varepsilon}_k \in [r_1,\ r_2]$. The filter is optimistic if the computed $\bar{\varepsilon}_k$ is much higher than r_2, while it is conservative with the computed $\bar{\varepsilon}_k$ below r_1.

(5) Computational cost. The computational cost is defined as the averaged running time over all Monte Carlo runs.

6.3. Reuslts and Analysis

In this subsection, three simulation scenarios are chosen to evaluate and compare the estimation performance of the proposed DHIWCF algorithm with respect to the afore-mentioned metrics.

6.3.1. Evaluation of the Effectiveness of the Proposed DHIWCF Algorithm

This scenario is designed to validate the effectiveness of the proposed DHIWCF algorithm. The target of interest is tracked by 8 networked sensors with a communication topology shown in Figure 1. Only node 1 can observe the target, then the node set $\{5, 6, 7, 8\}$ is naive about the target state. Figure 2 shows the estimated tracks obtained by local nodes with the proposed DHIWCF and the CKF. For simplicity, only the estimated tracks of node 1 and node 8 are plotted. To illustrate the evolution of each track, checkpoints are plotted in the same color every 20 steps. The covariance ellipses with 95% confidence at each checkpoint are plotted in dashed lines. As is shown, the true position (cross in black) is always enveloped by the corresponding ellipse in red (node 1) or blue (node 8), which validates the consistency of the local estimates. Compared with the CKF, the local estimates by the proposed DHIWCF is much more conservative. This is due to the fact that the network in Figure 1 has a weak connectivity and most nodes have poor joint observability.
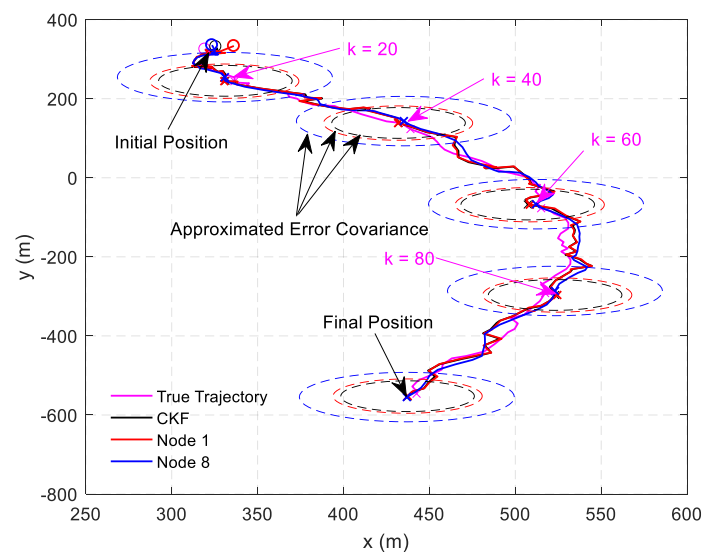
Figure 2. State estimation results: distributed hybrid information weighted consensus filter (DHIWCF) versus centralized Kalman consensus filter (CKF).

In Figure 3, the PRMSE of the compared algorithms with a single consensus iteration is given. Both KCF and CM diverge in the considered scenario, while others can effectively track the target. Later on in this section KCF and CM are not considered for their poor performance. The proposed DHIWCF is more accurate for its lower PRMSE close to the CKF. Due to limited consensus iterations, GKCF, ICF, and HCMCI obtain PRMSE higher than DHIWCF, but is much lower than CI.
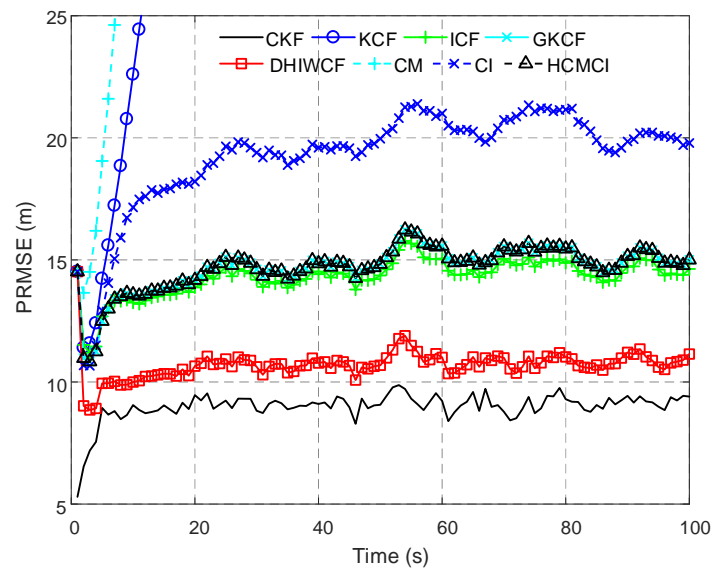
Figure 3. Position root mean squared error (PRMSE) averaged over all 8 nodes and 200 Monte Carlo runs.

Figure 4 compares the APRMSE of different algorithms. It shows that GKCF and ICF obtain APRMSE much higher than that of ICF, HCMCI and DHIWCF. Specifically, DHIWCF performs the best with limited consensus iterations $L \leq 2$. As consensus iteration increases, DHIWCF asymptotically converges to the CKF. In addition, the performance of DHIWCF is a little much better than that of ICF. In Figure 5, the average NEES of different algorithms with a single consensus iteration is compared. It is obvious that the NEES curve of ICF lies much higher than the 95% concentration regions, which indicates that ICF has poor consistency in such a scenario. The NEES curve of CI always lies below the concentration regions, and hence its estimates is much conservative. The NEES curves of GKCF, HCMCI, DHIWCF, and CKF lie either below or within the concentration regions all the time steps, which shows an enhanced consistency.
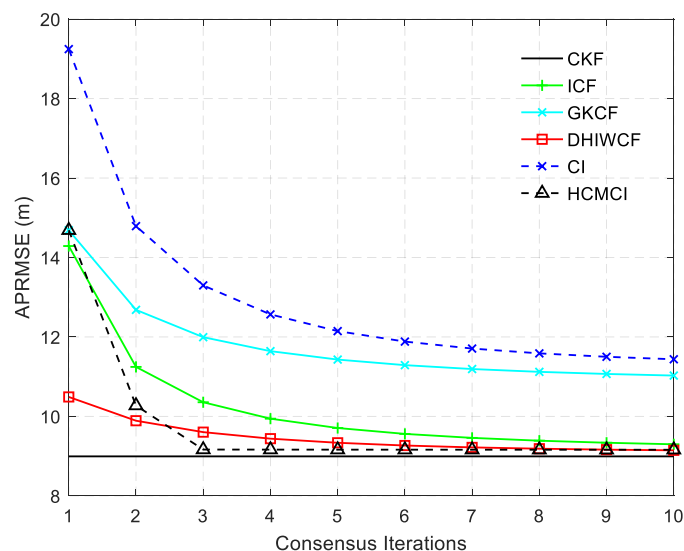
Figure 4. The averaged position root mean squared error (APRMSE) averaged over all nodes, all time steps and all Monte Carlo runs.
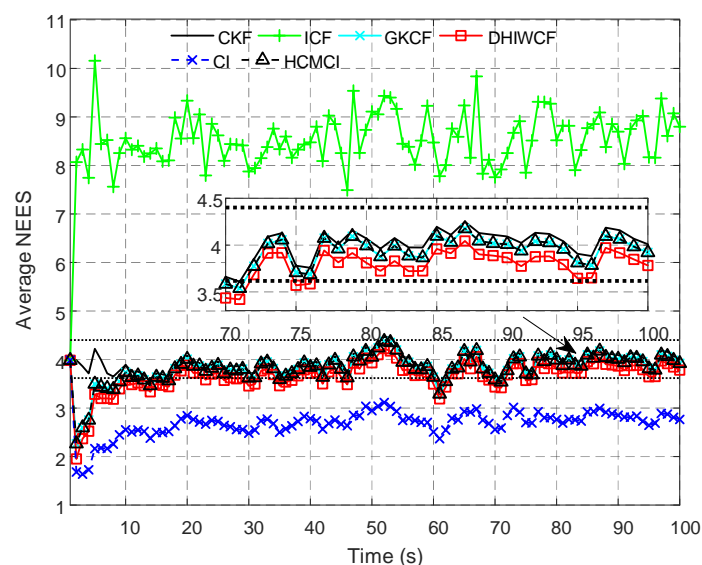
Figure 5. The average normalized estimation error squared (NEES) for the compared algorithms.

The ACEE comparison of different algorithms with a single consensus iteration is shown in Figure 6. The proposed DHIWCF performs much better with regard to consensus in that it has relatively lower ACEE than other algorithms. Figure 7 shows the computational time with different number of consensus iterations. Although HCMCI performs a little better than DHIWCF in the aspect of APRMSE as shown in Figure 4, its ACEE and computational time are much higher than that of DHIWCF. Moreover, DHIWCF is a little more time-consuming than the most efficient CI as shown in Figure 7.
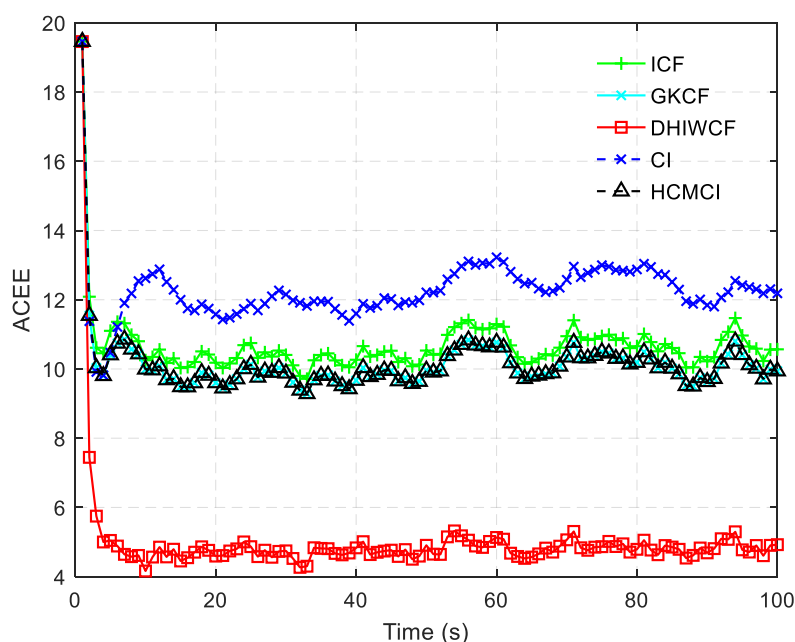
Figure 6. The averaged consensus estimate error (ACEE) for distributed algorithms.
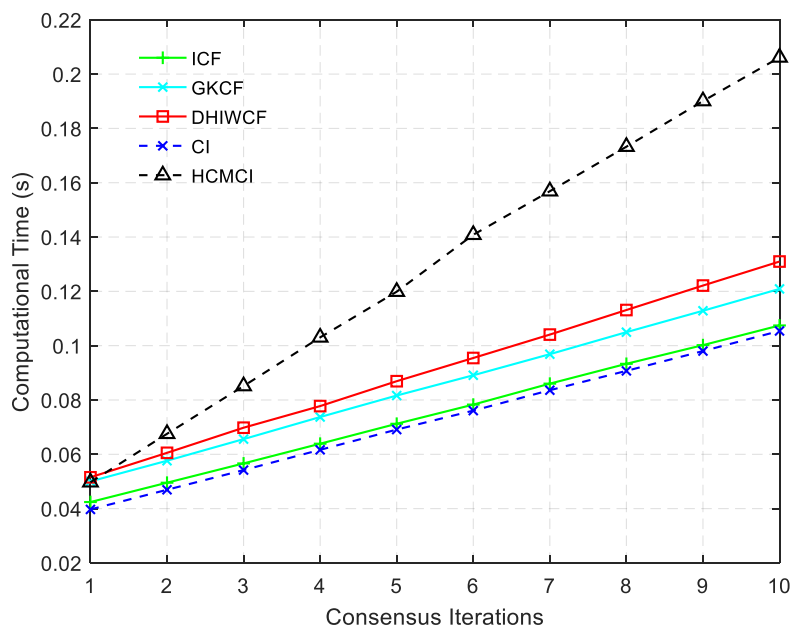
Figure 7. Computational cost of different algorithms.

6.3.2. Performance Comparison under Chain Topology

In this subsection, an even worse scenario is considered, where the networked sensors are connected with a chain topology as shown in Figure 8. As is illustrated, the target is observed only by node 1, and the remaining are communication nodes with no sensing abilities. Node set $\{3, 4, 5, 6, 7, 8\}$ and their immediate neighbors do not have measurement of the target, so they are naive about the target's state information. It takes at least 7 consensus iterations for node 8 to be affected by node 1. As is shown in Figure 4, the APRMSE of GKCF and CI is much higher than that of ICF, HCMCI, and DHIWCF. Here, the estimation results of GKCF and CI are not considered.
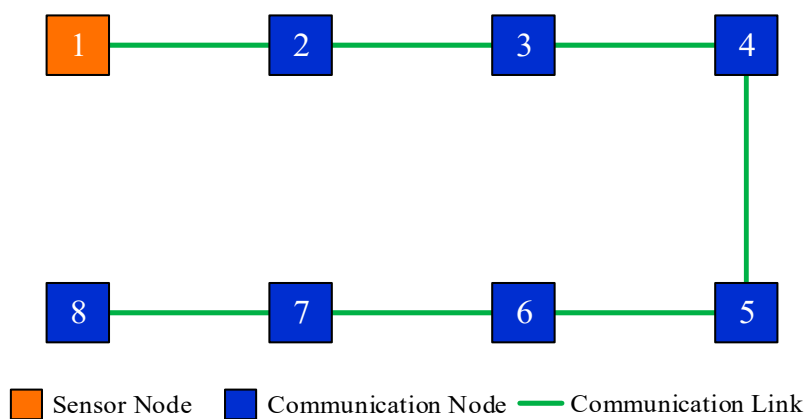

Figure 8. A sensor network with chain topology.

In Figure 9, the PRMSE averaged over all nodes and all Monte Carlo runs for different consensus iterations is given. With a single consensus iteration, DHIWCF performs much better than ICF and HCMCI. When consensus iterations increase to $L = 3$, DHIWCF is still smaller than the improved HCMCI. The result of averaged ACEE with a single consensus iteration is provided in Figure 10, which indicates that except for ICF, the remaining algorithms preserve good consistency.
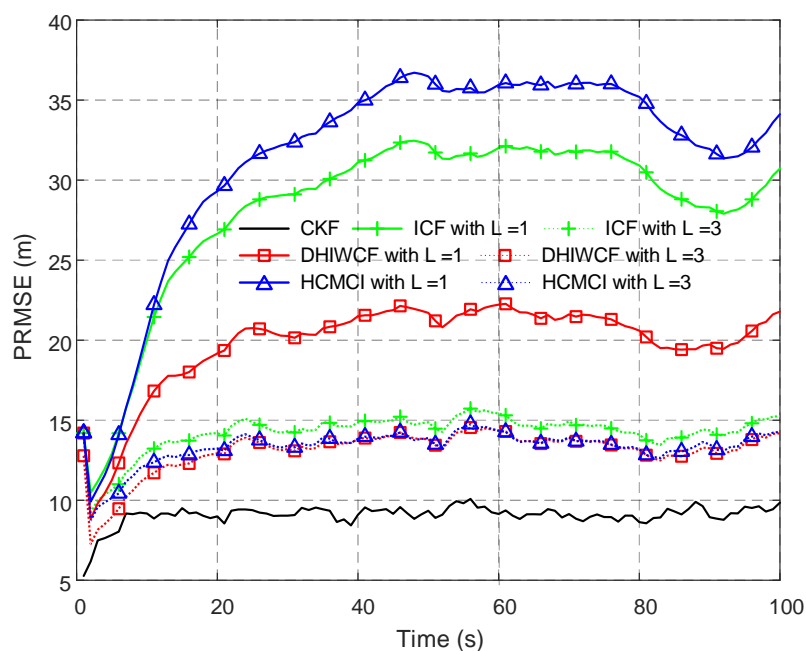
Figure 9. The averaged PRMSE for distributed algorithms with different consensus iterations.
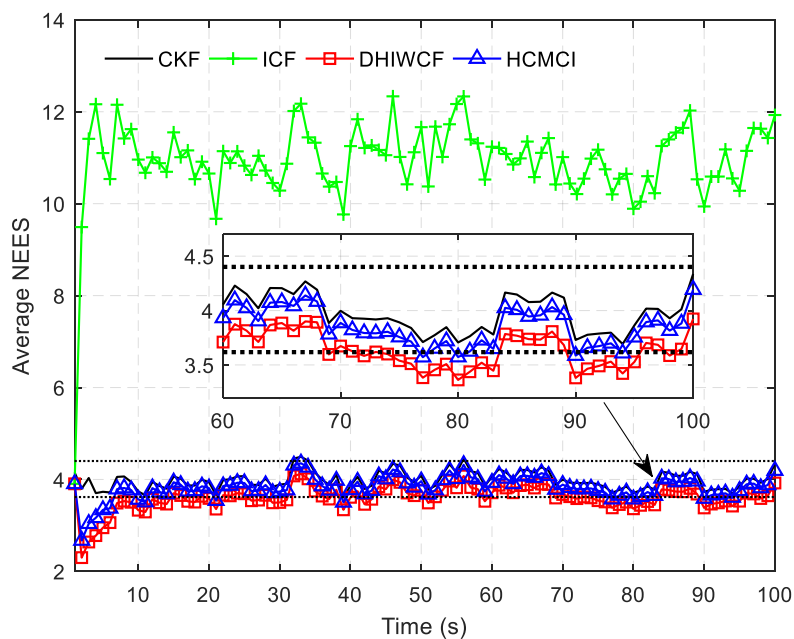

Figure 10. The averaged NEES of the compared algorithms.

The APRMSE for the algorithms under discussion is plotted in Figure 11. Similar to the result in Figure 4, DHIWCF has smaller APRMSE with $L \leq 3$, and its performance approaches the KCF as consensus iterations progress. Although HCMCI obtains APRMSE a little smaller than DHIWCF, its average ACEE is relatively higher as shown in Figure 12, especially in the case $L = 1$. Therefore, the proposed DHIWCF makes a good tradeoff between estimation accuracy and consensus on local estimates.
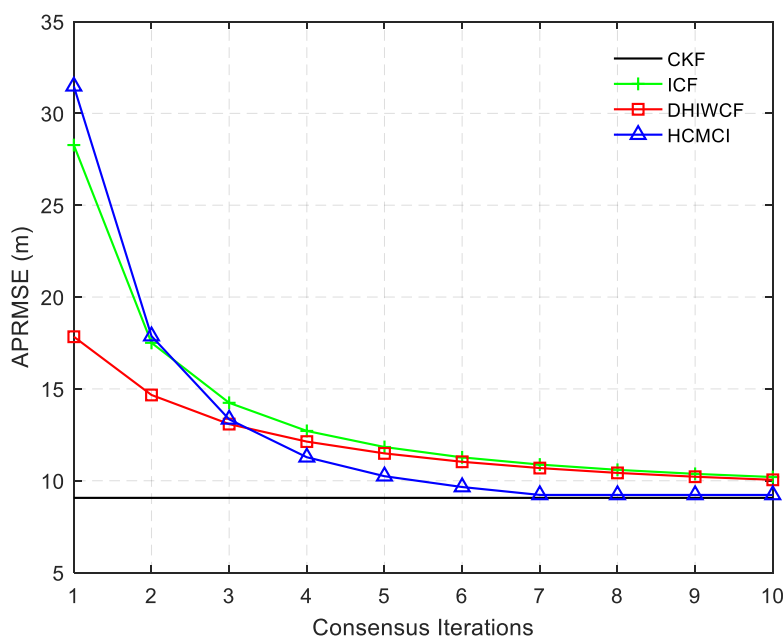
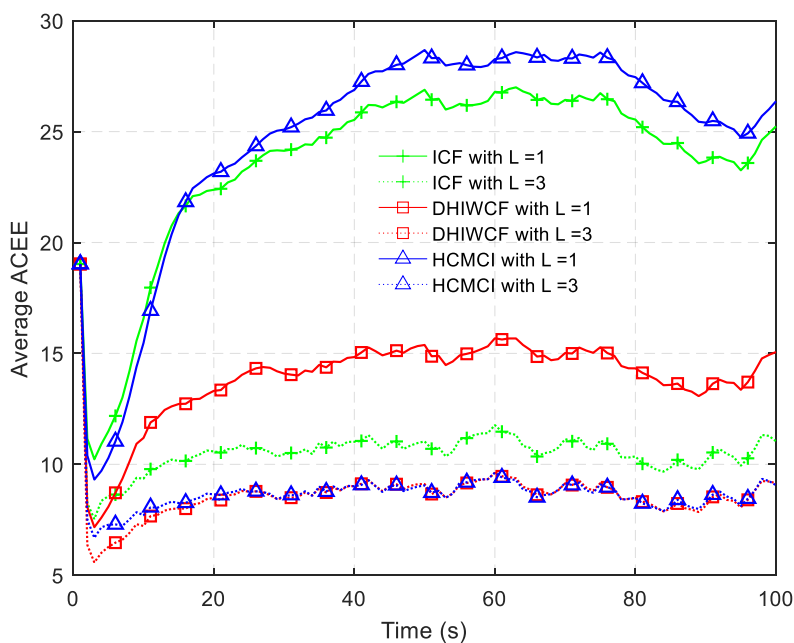**Figure 11.** APRMSE for different algorithms under discussion.

Figure 12. The average ACEE for distributed algorithms under discussion.

6.3.3. Performance Comparison in Large-Scale Sparse Sensor Networks

This experiment is designed to test the performance of DHIWCF in large-scale sparse sensor networks. Assume that the interested target is tracked by 100 sensors, which are randomly located within the 500×500 space. The communication range of each sensor is set to be $\boldsymbol{R}_c = 10 \sqrt{N_s}$. As is shown in Figure 13, there are 10 sensor nodes and 90 communication nodes in the surveillance area. The sensor nodes are able to observe the target, while the communication nodes act as relays of information among distant nodes and has no observation ability [24]. As is shown in Figure 13, most of nodes in the network are naive about the target's state, which brings great challenges to target tracking.
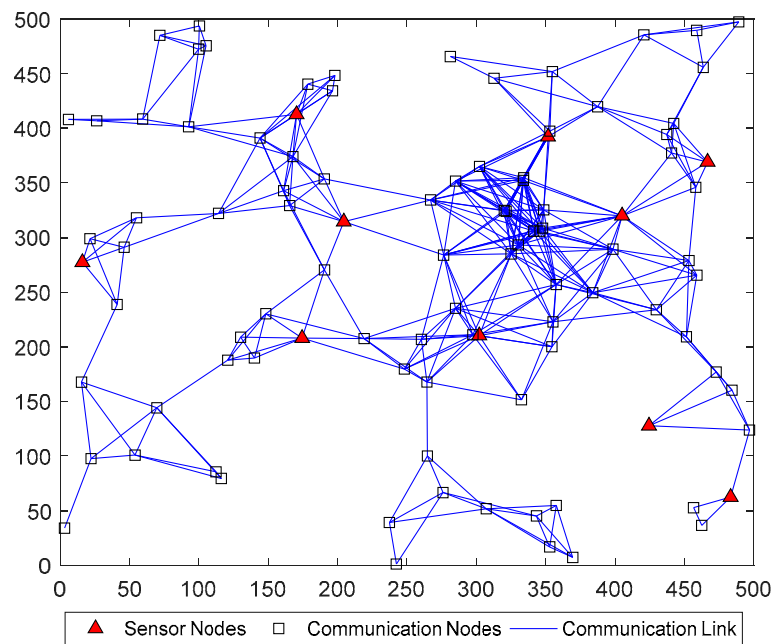
Figure 13. A large-scale sparse sensor network with 100 nodes.

In Figure 14, the estimated tracks by different algorithms with a single consensus iteration in a certain Monte Carlo run are plotted. It is intuitive to see that DHIWCF and KCF perform much better than ICF and HCMCI. Especially when the target suffers from relatively obvious process noise (for instance, the target suddenly changes its moving direction), DHIWCF recovers its estimate to the CKF more quickly. The results in Figure 15 further suggest that with limited consensus iterations, DHIWCF is able to obtain more accurate estimates than ICF and HCMCI in that it has relatively lower PRMSE compared with its counterparts. With respect to estimation consistency, it is shown in Figure 16 that the NEES curve of ICF lies higher than the concentration region, while the NEES curves of the remaining algorithms are always within or below the concentration region. Therefore, both HCMCI and DHIWCF show sound consistency on local estimates.
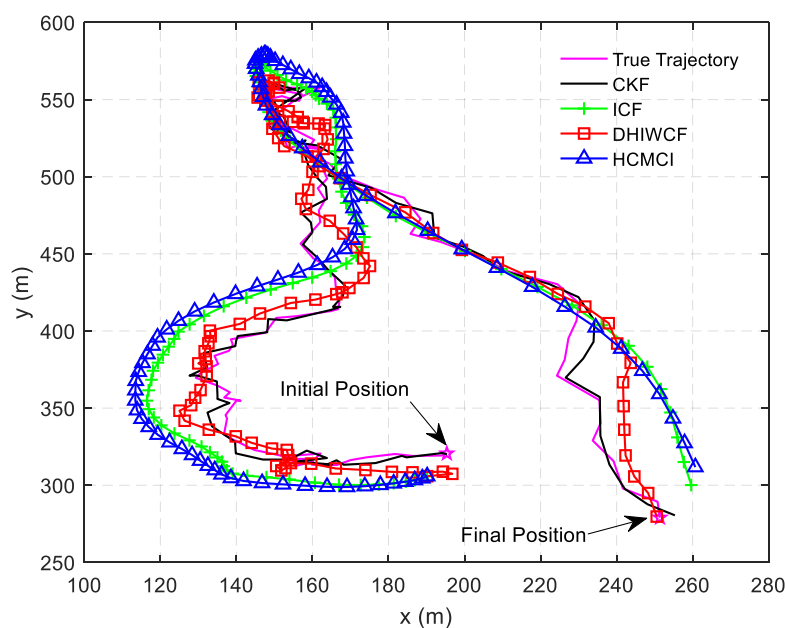

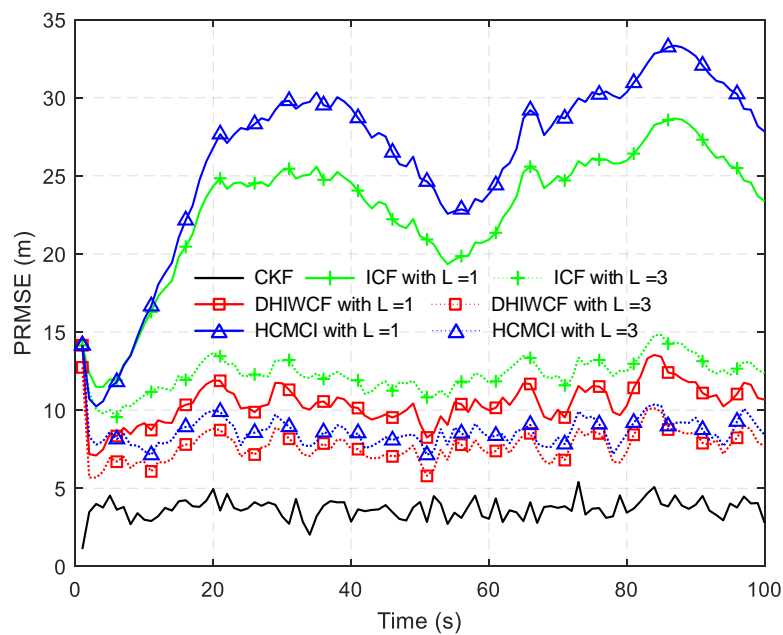**Figure 14.** The true and estimated tracks of different algorithms.

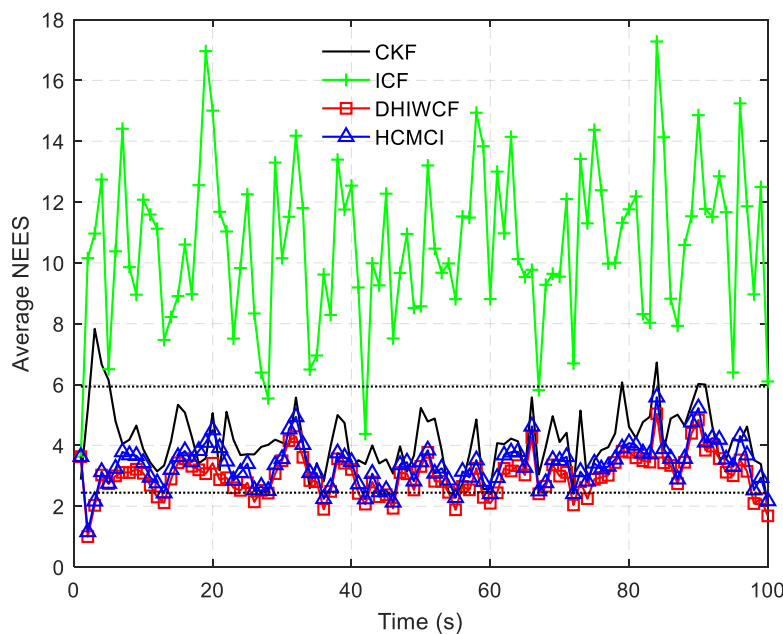**Figure 15.** PRMSE for different algorithms in a spare sensor network.

Figure 16. The averaged NEES for different algorithms in a spare sensor network.

To compare the overall performance of the distributed algorithms under discussion, Table 1 gives the APRMSE for different algorithms versus consensus iterations. Compared with ICF and HCMCI, the proposed DHIWCF has lower APRMSE. Especially in case of consensus iterations $L \leq 3$, the advantage is more obvious. This implies that DHIWCF is relatively more accurate. The computational time relative to that of CKF is investigated in Table 2, where RCT means the relative computation time. The proposed DHIWCF runs faster than HCMCI for fewer information exchanges. Although it takes less time for ICF to operate, the lower estimation accuracy and poor consistency make it not a good choice to estimate the state of interest.

Table 1. Comparison of ARMSE for different algorithms.

APRMSE(m)	CKF	ICF	DHIWCF	HCMCI
1	3.68	22.87	10.44	26.24
2	3.68	14.80	9.23	11.93
3	3.68	12.28	8.37	8.64
4	3.68	11.05	7.63	7.93
5	3.68	10.30	7.46	7.78
10	3.68	8.56	6.87	7.25

Table 2. The relative computational time for different algorithms.

RCT	CKF	ICF	DHIWCF	HCMCI
$L = 1$	1	5.26	7.92	9.49
$L = 2$	1	6.58	10.84	15.55
$L = 3$	1	7.30	13.71	21.57
$L = 4$	1	8.34	16.68	27.04
$L = 5$	1	9.39	19.55	32.25
$L = 10$	1	14.18	34.43	60.05

7. Conclusions

This paper considers the problem of distributed state estimation in presence of naive nodes with constrained communication resources. A novel distributed hybrid information weighted consensus filter, in which each node exploits not only the measurement information but also the prior estimate information from its immediate neighbors to update its local posterior estimate, is proposed. The proposed DHIWCF is able to settle the problem under consideration without any knowledge of global parameters, and preserve consistency of local estimates as well as achieve relatively high estimation accuracy and satisfactory consensus. Theoretical analysis with regard to consistency of local estimates, stability, and convergence of the estimator is also provided. The experimental results indicate that with limited consensus iterations, the proposed DHIWCF is much more accurate and reaches better consensus than the existing algorithms. In addition, DHIWCF preserves good consistency of local estimates in the experiments. Even a single consensus iteration is allowed, the proposed DHIWCF still performs much better. If more consensus iterations are available, the proposed DHIWCF would approach the performance of the centralized scheme. In the future research, a further investigation for distributed state estimation in mobile sensor networks, consensus protocol with event-triggered communication, more efficient design of consensus weights, and distributed nonlinear filtering problems and stability analysis will be taken into account.

Author Contributions: Conceptualization, Y.L. and Y.H.; Data curation, Z.D.; Formal analysis, J.L.; Funding acquisition, Y.H.; Methodology, J.L.; Project administration, Y.L.; Software, J.L., Y.L. and Z.D.; Visualization, K.D.; Writing—original draft, J.L.; Writing—review and editing, J.L., Y.L. and K.D.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Marelli, D.; Zamani, M.; Fu, M.; Ninness, B. Distributed Kalman filter in a network of linear systems. *Syst. Control Lett.* **2018**, *116*, 71–77. [CrossRef]
2. Battistelli, G.; Chisci, L.; Selvi, D. A distributed Kalman filter with event-triggered communication and guaranteed stability. *Automatica* **2018**, *93*, 75–82. [CrossRef]

3. Kalman, R.E. A New Approach to Linear Filtering and Prediction Problems. *J. Basic Eng.* **1960**, *82*, 35–45. [CrossRef]

4. Kamal, A.T.; Ding, C.; Song, B.; Farrell, J.A.; Roy-Chowdhury, A.K. A Generalized Kalman Consensus Filter for Wide-Area Video Networks. In Proceedings of the 50th IEEE Conference on Decision and Control and European Control Conference, Orlando, FL, USA, 12–15 December 2011; pp. 7863–7869.

5. Olfati-Saber, R. Distributed Kalman Filtering for Sensor Networks. In Proceedings of the 46th IEEE Conference on Decision and Control, New Orleans, LA, USA, 12–14 December 2007; pp. 5492–5498.

6. Olfati-Saber, R. Kalman-Consensus Filter: Optimality, Stability, and Performance. In Proceedings of the 48h IEEE Conference on Decision and Control (CDC) Held Jointly with 2009 28th Chinese Control Conference, Shanghai, China, 15–18 December 2009; pp. 7036–7042.

7. AminiOmam, M.; Torkamani-Azar, F.; Ghorashi, S.A. Generalised Kalman-consensus filter. *IET Signal Process* **2017**, *11*, 495–502. [CrossRef]

8. Deshmukh, R.; Kwon, C.; Hwang, I. Optimal Discrete-Time Kalman Consensus Filter. In Proceedings of the 2017 American Control Conference (ACC), Seattle, WA, USA, 24–26 May 2017; pp. 5801–5806.

9. Yao, P.; Liu, G.; Liu, Y. Average information-weighted consensus filter for target tracking in distributed sensor networks with naivety issues. *Int. J. Adapt. Control* **2018**, *5*, 681–699. [CrossRef]

10. Chong, C.; Chang, K.; Mori, S. Comparison of Optimal Distributed Estimation and Consensus Filtering. In Proceedings of the 19th International Conference on Information Fusion, Heidelberg, Germany, 5–8 July 2016; pp. 1034–1041.

11. Liu, Y.; Liu, J.; Xu, C.; Qi, L.; Sun, S.; Ding, Z. Consensus Algorithm for Distributed State Estimation in Multi-Clusters Sensor Network. In Proceedings of the 20th International Conference on Information Fusion, Xi'an, China, 10–13 July 2017; pp. 1605–1609.

12. Rastgar, F.; Rahmani, M. Consensus-based distributed robust filtering for multisensor systems with stochastic uncertainties. *IEEE Sens. J.* **2018**, *18*, 7611–7618. [CrossRef]

13. Ji, H.; Lewis, F.L.; Hou, Z.; Mikulski, D. Distributed information-weighted Kalman consensus filter for sensor networks. *Automatica* **2017**, *77*, 18–30. [CrossRef]

14. Liu, Q.; Wang, Z.; He, X.; Zhou, D.H. On Kalman-consensus filtering with random link failures over sensor networks. *IEEE Trans. Autom. Control* **2018**, *63*, 2701–2708. [CrossRef]

15. Soatti, G.; Nicoli, M.; Savazzi, S.; Spagnolini, U. Consensus-based algorithms for distributed network-state estimation and localization. *IEEE Trans. Signal Inf. Proc. Over Netw.* **2017**, *3*, 430–444. [CrossRef]

16. Kamal, A.T.; Farrell, J.A.; Roy-Chowdhury, A.K. Information weighted consensus filters and their application in distributed camera networks. *IEEE Trans. Autom. Control* **2013**, *58*, 3112–3125. [CrossRef]

17. Tang, W.; Zhang, G.; Zeng, J.; Yue, Y. Information weighted consensus-based distributed particle filter for large-scale sparse wireless sensor networks. *IET Commun.* **2014**, *8*, 3113–3121. [CrossRef]

18. Kamal, A.T. Information Weighted Consensus for Distributed Estimation in Vision Networks. Ph.D. Dissertation, University of California Riverside, Riverside, CA, USA, 2013.

19. Liu, G.; Tian, G.; Zhao, Y. Information Weighted Consensus Filtering with Improved Convergence Rate. In Proceedings of the IEEE 35th Chinese Control Conference (CCC), Chengdu, China, 27–29 July 2016; pp. 8356–8359.

20. Bin, J.; Khanh, D.P.; Erik, B.; Dan, S.; Zhonghai, W.; Genshe, C. Cooperative space object tracking using space-based optical sensors via consensus-based filters. *IEEE Trans. Aerosp. Electron. Syst.* **2016**, *52*, 1908–1936.

21. Jia, B.; Pham, K.D.; Blasch, E.; Shen, D.; Chen, G. Consensus-Based Auction Algorithm for Distributed Sensor Management in Space Object Tracking. In Proceedings of the 2017 IEEE Aerospace Conference, Big Sky, MT, USA, 4–11 March 2017; pp. 1–8.

22. Kamal, A.T.; Farrell, J.A.; Roy-Chowdhury, A.K. Information Weighted Consensus. In Proceedings of the 51st IEEE Conference on Decision and Control, Maui, HI, USA, 10–13 December 2012; pp. 2732–2737.

23. Shang, Y. Resilient consensus of switched multi-agent systems. *Syst. Control Lett.* **2018**, *122*, 12–18. [CrossRef]

24. Shang, Y. Resilient Multiscale Coordination Control against Adversarial Nodes. *Energies* **2018**, *11*, 1844. [CrossRef]

25. Olfati-Saber, R.; Fax, J.A.; Murray, R.M. Consensus and cooperation in networked multi-agent systems. *IEEE Proc.* **2007**, *95*, 215–233. [CrossRef]

26. Battistelli, G.; Chisci, L.; Mugnai, G.; Farina, A.; Graziano, A. Consensus-based linear and nonlinear filtering. *IEEE Trans. Autom. Control* **2015**, *60*, 1410–1415. [CrossRef]

27. Battistelli, G.; Chisci, L.; Mugnai, G.; Farina, A.; Graziano, A. Consensus-Based Algorithms for Distributed Filtering. In Proceedings of the IEEE 51st IEEE Conference on Decision and Control (CDC), Maui, HI, USA, 10–13 December 2012; pp. 794–799.

28. Battistelli, G.; Chisci, L. Kullback–Leibler average, consensus on probability densities, and distributed state estimation with guaranteed stability. *Automatica* **2014**, *50*, 707–718. [CrossRef]

29. Julier, S.J.; Uhlmann, J.K. A Non-Divergent Estimation Algorithm in the Presence of Unknown Correlations. In Proceedings of the 1997 American Control Conference, Albuquerque, NM, USA, 6 June 1997; pp. 2369–2373.

30. Wang, S.; Ren, W. On the convergence conditions of distributed dynamic state estimation using sensor networks: A unified framework. *IEEE Trans. Control Syst. Technol.* **2018**, *26*, 1300–1316. [CrossRef]

31. Chen, Q.; Yin, C.; Zhou, J.; Wang, Y.; Wang, X.; Chen, C. Hybrid consensus-based cubature Kalman filtering for distributed state estimation in sensor networks. *IEEE Sens. J.* **2018**, *18*, 4561–4569. [CrossRef]

32. Li, T.; Corchado, J.M.; Prieto, J. Convergence of Distributed Flooding and Its Application for Distributed Bayesian Filtering. *IEEE Trans. Signal Inf. Proc. Over Netw.* **2017**, *3*, 580–591. [CrossRef]

33. Xiao, L.; Boyd, S. Fast Linear Iterations for Distributed Averaging. In Proceedings of the 42nd IEEE International Conference on Decision and Control, Maui, HI, USA, 9–12 December 2003; Volume 5, pp. 4997–5002.

34. Yilun, S. Finite-Time Weighted Average Consensus and Generalized Consensus Over a Subset. *IEEE Access* **2016**, *4*, 2615–2620.

35. Ren, W.; Beard, R.W.; Kingston, D.B. Multi-Agent Kalman Consensus with Relative Uncertainty. In Proceedings of the American Control Conference, Portland, OR, USA, 8–10 June 2005; pp. 1865–1870.

36. Alighanbari, M.; How, J.P. Unbiased Kalman consensus algorithm. *J. Aerosp. Comput. Inf. Commun.* **2008**, *5*, 298–311. [CrossRef]

37. Motion, L. General Decentralized Data Fusion with Covariance Intersection. In *Handbook of Multisensor Data Fusion*; CRC Press: Boca Raton, FL, USA, 2001; pp. 1–40.

38. Niehsen, W. Information Fusion Based on Fast Covariance Intersection Filtering. In Proceedings of the Fifth International Conference on Information Fusion, Annapolis, MD, USA, 8–16 July 2002; pp. 901–904.

39. Li, T.; Fan, H.; García, J.; Corchado, J.M. Second-order statistics analysis and comparison between arithmetic and geometric average fusion: Application to multi-sensor target tracking. *Inf. Fusion* **2019**, *51*, 233–243. [CrossRef]

40. Calafiore, G.C.; Abrate, F. Distributed linear estimation over sensor networks. *Int. J. Control* **2009**, *82*, 868–882. [CrossRef]

41. Ren, W.; Beard, R.W. Consensus seeking in multiagent systems under dynamically changing interaction topologies. *IEEE Trans. Autom. Control* **2005**, *50*, 655–661. [CrossRef]