Efficient Distributed Information Fusion using Value of Information based Censoring

Beipeng Mu, *Girish Chowdhary, †Jonathan P. How‡
Laboratory for Information and Decision Systems

June 7, 2012

Abstract

In many distributed sensing applications, not all agents have valuable information at all times. Therefore, requiring all agents to communicate at all times can be resource intensive. In this work, the notion of Value of Information (VoI) is used to improve the efficiency of distributed sensing algorithms. Particularly, only agents with high VoI broadcast their measurements to the network, while others censor their measurements. New VoI realized data fusion algorithms are introduced, and an in depth analysis of the costs incurred by these algorithms and conventional distributed data fusion algorithms is presented. Numerical simulations are used to compare the performance of the VoI realized algorithms with traditional data fusion algorithms. A VoI based algorithm that adaptively adjusts the criterion for being informative is presented and shown to strike a good balance between reduced communication cost and increased accuracy.

1 INTRODUCTION

Distributed computing and sensing systems are becoming increasingly common as sensing and communication capabilities become available in compact packages. The problem of estimating parameters using a distributed set of sensors has been widely studied (see for example [1–8]). Since distributed sensing approaches are robust to single point failures, they arise in numerous domain, including networked Unmanned Aerial Vehicles, smart power grids, and distributed Sensor Networks. The key issue in distributed estimation is to ensure the scalability of the distributed sensing algorithms to systems with large number of agents, while ensuring the accuracy of the estimation.

Several authors have studied distributed inference using Bayesian techniques, a review is available in [5]. Yedidia et al. have studied Belief Networks which use acyclic graphs to represent random variables and their dependence which is often extracted from agents' local properties [9]. In these methods, nodes communicate with each

---

*Research Assistant, Department of Aeronautics and Astronautics, MIT mubp@mit.edu

†Postdoctoral Associate, Laboratory for Information and Decision Systems, MIT girishc@mit.edu

‡Richard C. Maclaurin Professor of Aeronautics and Astronautics, MIT jhow@mit.edu

other whenever new measurements are available. In Channel Filter algorithms [10] the goal is to estimate a set of common variables observed by a network of sensors. These methods require that the network be acyclic, furthermore, elimination of the acyclic constraint is complicated and costly in terms of computation and communication. Consensus is a set of algorithms in which agents reach asymptotic agreement using only locally available information (see e.g. [8,11,12]).

However, the relationship between the accuracy of the distributed estimation algorithms and the cost incurred in communication have not been deeply studied.In several real world distributed sensing situations, the measurements of all agents at all time are not equally informative for improving the global parameter estimation. Examples of such situations include situations where the observed process varies slowly (such as temperature in a building) or situations where the observed process is spatially distributed and not all agents are in a good position to take informative measurements. This situation is depicted in Figure 1, where the dark colored nodes are the only ones with valuable information, however, all the nodes in the network are communicating, resulting in wasted resources. Uney, Cetin et al., Msechu et al. and Tay et al. have studied the notion of censoring nodes in a sensing network to reduce communication cost [13–18]. In their work, a central node decides on which sensors can get good measurements and censors others to save resources. Chen et al. use a simple metric on value of information to determine whether a new piece of measurement will be broadcasted in a data association problem in a distributed sensor network [19].

The research presented here was motivated by the question of whether there are more efficient algorithms for performing distributed fusion than the standard consensus algorithms that have been proposed in the literature. In particular, we explore the notion of *Value of Information* (VoI), which quantifies how informative an agent believes its measurements are. The approach taken is to have the agents in the network check the quality of their data, and then only have *informative agents* communicate with others about their data while *uninformative ones* censor themselves, though they may be tasked with acting as relays. New methods for information fusion using VoI are introduced, and compared in depth with a consensus based algorithm. Furthermore, an adaptive algorithm that adjusts the VoI criterion is introduced to ensure a good balance between the communication cost and estimation accuracy. This paper further explores the communication cost savings afforded by a censoring based approach by providing detailed estimates of cost incurred for several distributed sensing algorithms. Furthermore, numerical simulations are employed to compare the performance and cost of distributed sensing algorithms. Our results indicate that VoI based data fusion can have significant cost savings over consensus based approaches while ensuring good estimation performance.

This paper is organized as follows. Section 2 introduces related probability, graph theory, and distributed estimation concepts. Section 3 and 4 develop VoI based data fusion algorithms. Section 5 presents an adaptive VoI realized data fusion algorithms. Results of numerical simulations are provided in 6, the paper is concluded in 7.
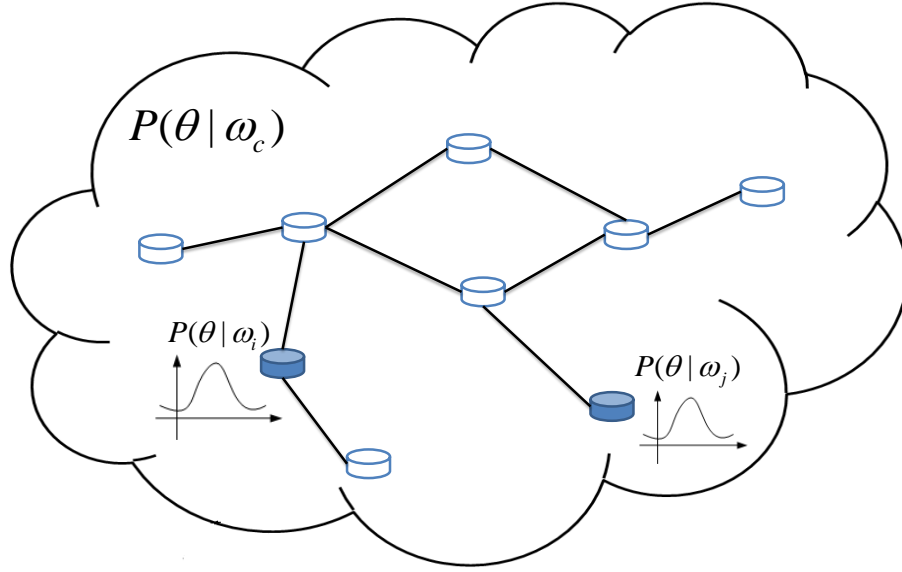
Figure 1: Cost VoI Realized Fusion

2 BACKGROUND

2.1 Bayesian Parameter Estimation

Bayesian parameter estimation algorithms (Bayesian inference algorithms) use Bayes law to update the posterior distribution of parameters of interest as new measurements arrive. Let Θ denote a random variable (RV) representation of the parameters of interest, and let θ be a realization of Θ. Let Z be a RV denoting possible measurements and let z be an actual measurement, which is a sample of Z. Let $p_\Theta(\theta)$ be the prior distribution on Θ, and $p_{\Theta|Z}(\theta|z)$ be the posterior distribution after incorporating measurement z. Then, Bayesian inference can be stated as follows (see for example [20])

$$p_{\Theta|Z}(\theta|z) = \frac{p_{Z|\Theta}(z|\theta)p_\Theta(\theta)}{\int p_{Z|\Theta}(z|\theta)p_\Theta(\theta)\,d\theta}. \tag{1}$$

In the above equation $p_{Z|\Theta}(z|\theta)$ is the likelihood function of Z given prior estimate of parameters θ.

2.2 Conjugate Prior and Linear Updates

In many cases, prior knowledge about the distribution of parameters can be expressed as parameterized functions of the form $p_\Theta(\theta) = p_{\Theta|\Omega}(\theta|\omega)$, where ω is referred as the hyperparameters of the distribution [20,21]. The set of all possible hyperparameters ω

is denoted by Ω. With this notation (1) can be rewritten as

$$
\begin{aligned}
p_{\Theta|\Omega}(\theta|\omega_{post}) &= p_{\Theta|Z,\Omega}(\theta|z,\omega_{prior}) \\
&= \frac{p_{Z|\Theta}(z|\theta)p_{\Theta|\Omega}(\theta|\omega_{prior})}{\int p_{Z|\Theta}(z|\theta)p_{\Theta|\Omega}(\theta|\omega_{prior})\,d\theta}.
\end{aligned}
\tag{2}
$$

It is shown in [20,21] that if the likelihood function is in the family of exponential distributions, and if the prior distribution is conjugate to the likelihood function, the posterior distribution will have the same form as the prior. Furthermore, hyperparameters can be updated using a linear update law of the form

$$
\omega_{post} = \omega_{prior} + h(z, p(z|\theta)).
\tag{3}
$$

2.3 Distributed Estimation of Hyperparameters

Consider the situation in which distributed agents collaborate to estimate the posterior distribution of a quantity of interest by combining distributed measurements [2,4,6,10]. The posterior distribution of the quantity of interest will be referred to as the global posterior. A simple example of this situation is the estimation of temperature distribution in a building using distributed temperature sensors that can communicate. One way to perform distributed estimation is to set up a fusion center, which communicates with all the agents, and gets measurements to computes the centralized posterior [13]. However, this approach may not be robust or scalable in all scenarios. This paper focuses on distributed approaches to information fusion, that is approaches which calculate the global posterior without assuming access to a centralized fusion center.

We employ a graph theoretic representation of a communication enabled network. Particularly, the network is represented as a graph $G\langle v, E\rangle$, with $v = 1, ..., N$ denoting the set of vertexes or nodes of the network, and E denoting the set of edges $E \subset v \times v$, with the pair $(i, j) \in E$ if and only if the agents i can communicate with or otherwise sense the state of agent j. In this case, agent j is termed as a neighbor of agent i. Set of all i's neighbors is defined as agent i's neighborhood, denoted by $\mathcal{N}_i$. If the elements of the edge set (that is the pairs (i, j)) are unordered, the graph is termed as undirected. This paper focuses on the case of undirected graphs for the ease of exposition, however, the result can be extended to the directed case in a relatively straight forward manner.

Furthermore, in order to simplify analysis, this paper considers the case of a synchronized communication network, in which it is assumed that the clocks of all agents are synchronized. The result can be extended to asynchronous scenarios. In the synchronous case, time can be indexed by the integer $t \in \mathbb{R}^+$, and all variables can be indexed by t. The measurements agent i takes at t is denoted by $\mathbf{z}_i[t] \in \mathrm{R}^{m_i[t]}$. The jth ($j \in [1, m_i[t]]$) measurement node i takes at t is denoted by $z_i^j[t]$. For convenience, let $h_i[t] = h(\mathbf{z}_i[t], p(\mathbf{z}_i[t]|\theta)$, and $h_i^j[t] = h(z_i^j[t], p(z_i^j[t]|\theta)$. The global posterior can be expressed as:

$$
p(\theta|\omega_c[t]) = p(\theta|\mathbf{z}_1[1:t], \mathbf{z}_2[1:t], \cdots, \mathbf{z}_N[1:t], \omega_c[0])
\tag{4}
$$

The following assumptions put restrictions on the types of networks we consider.
Assumption 1: The network is strongly connected. That is, for every i, j in the vertex set a path exists from i to j that can be formed using pairs in the edge set.

Table 1: Brute Force Fusion

step	description	cost
1	Initialization set a global prior $\omega_i[0] = \omega[0]$:	C_{init}
2	At time $t,\forall i$ take measurements $\mathbf{z}_i[t]$ compute local update $h_i[t]$	
3	Broadcast and relay updates broadcast $h_i[t]$ $\forall j \neq i$, relay $h_i[t]$ Compute posterior $\omega_i[t] = \omega_i[t-1] + \sum_{i=1}^{N} h_i[t] = \omega[t]$	C_i $(N-1)C_i$
6	$t = t + 1$, goes to step 2	
Total cost in $[1:T]$, $C_{init} + NT\sum_{i=1}^{N} C_i$		

Assumption 2: Every agent has a unique identifying sequence that it can transmit to differentiates its measurements from others.

Assumption 3: The network topology is known.

Assumption 1 indicates that all agents can have access to each others measurement through intermediate communication. Assumption 2 guarantees that agents can tell each others measurements apart using a label. Assumption 3 can be restrictive in some scenarios, however, if agents begin by not knowing the network topology, Assumption 1 and 2 allow the agents to communicate with each other to figure out the network topology. More efficient alternatives are available, including [22–24].

Assumption 4: Relaying a message relay is much faster than recording a local measurement, processing it, and then broadcasting it.

It is also assumed that the agents begin a common global prior over the hyperparameters. Alternatively, one can introduce an initialization process during which agents agree on the global prior hyperparameter $\omega_i[0] = \omega[0]$. Global prior can be externally provided to the network, or computed by the network from local priors.

2.4 Brute Force Distributed Data Fusion

Using assumption 1 and 2, a brute force method of fusion can be envisioned by having every agent broadcast its own measurement $h_i[k]$ to its neighbors and relay updates from every neighbor to every other neighbors. Under this protocol, every agent will end up eventually with a copy of all other agents' measurements. The agents can then compute the global posterior locally by adding all measurements to its local prior hyperparameters using (3). The Brute Force algorithm is depicted in Table 1.

Cost: Let N be the number of agents in the network, and let their cost of broadcasting one message to their neighbors is C_i. In the Brute Force information fusion method, at each time step, each node needs to broadcast its own update and relay updates for all other nodes. Therefore, the total number of messages every node sends

out is N. The cost for all nodes at time t is $N \sum_i C_i$. During a time period $[0, T]$, the total cost would be $TN \sum_i C_i$.

2.5 Hyper Parameter Consensus

It is well known that the Brute Force approach of Section 2.4 is computationally in-effective. An approach often pursued in the literature is that of consensus (see for example [6], [11], [8]). In this approach at each time step t, an agent receives mes-sages from its neighbors, updates its local estimate of the parameters, and sends out the updated estimate back to its neighbors. Because every agent communicates only with its neighbors, the number of messages agents need to sent out is greatly reduced. Fraser et al. extended the consensus approach to hyperparameter estimation in [25], they termed the approach Hyper Parameter Consensus (HPC). Using assumptions 1–3 and the property of conjugate priors Fraser et al. prove the following lemma.

Lemma [25]: Start with the same hyperparameters on the global prior $\omega_c[0]$, let each agent i takes $m_i[t]$ unique local measurements $z_i^1[t], z_i^2[t], \cdots, z_i^{m_i}[t]]$ at t, then the fused distribution is of the same form as the global prior but with global posterior at t given by:

$$
\begin{aligned}
\omega_c[t] &= \omega_c[0] + \sum_{k=1}^{t} \sum_{i=1}^{N} h_i[k] \\
&= \sum_{k=1}^{t} \sum_{i=1}^{N} \sum_{j=1}^{m_i[t]} h_i^j[k].
\end{aligned}
\tag{5}
$$

Let $A = \{a_{ij}\}$ denote the weighted adjacency matrix for the graph of the sen-sor network [8], whose rows sum up to one. Let $\mathbf{v} = [v_1, v_2, \cdots, v_N]^T$ denote the eigenvector of corresponding to eigenvalue 1 of A, then the HPC algorithm depicted in Table 2 guarantees that the agents converge to the global posterior asymptotically.

Cost: At each time step, each agent sends out only one message containing an update of its local hyperparameters. Therefore, the cost of all nodes at time t is $\sum_i C_i$. During a time period $[0, T]$, the total cost would be $T \sum_i C_i$.

3 VALUE OF INFORMATION REALIZED INFOR-MATION FUSION

In this section communication cost-efficient methods that rely on the notion of value of information for information fusion are introduced. In several real world sensing sit-uations, the measurements of all agents at all time are not equally informative for im-proving the global parameter estimation. Examples of such situations include situations where the observed process varies slowly (such as temperature) or situations where the observed process is spatially distributed and not all agents are in a good position to take informative measurements. The Brute Force method and HPC both communicate measurements across agents without differentiating high value information. A strategy

Table 2: Hyperparameter Consensus

step	description	cost
1	Initialization set a global prior $\omega_i[0] = \omega[0]$:	C_{init}

2	At time $t, \forall i$
	    take measurements $\mathbf{z}_i[t]$
	    compute local update $h_i[t]$	

3	Compute local posterior $\omega_i[t] = \omega_i[t] + \frac{h_i[t]}{v_i}$	
4	broadcast to neighbors and update $\omega_i[t] = \sum\limits_{j \in \mathcal{N}_i} a_{ij}\omega_j[t]$	C_i
5	$t = t + 1$, goes to step 2	
Total cost in $[1:T]$, $C_{init} + T\sum_{i=1}^{N}$		

in which only high value information is transmitted may lead to significant savings in communication cost. The value of information realized information fusion methods of this section communicate information across the network only when agents decide locally that the information they have is informative for the distributed computation of the global posterior distribution.

3.1 Value of Information

We consider value of information metrics in the f-Divergence family of functions that measure the difference between two probability distributions [26]. The Kullback-Leibler (KL) divergence, also called relative entropy, is a widely used value of information metric in this family [27,28]. For two distributions P and Q, the KL divergence is defined as $D_{KL}(P||Q) = \sum_i P(i) \ln \frac{P(i)}{Q(i)}$ in discrete case, and $D_{KL}(P||Q) = \int_{-\infty}^{\infty} p(x) \frac{p(x)}{q(x)} \, dx$ in continuous case. The value of agent i's measurement at t, $\mathbf{z}_i[t]$, is quantified by how much it can change the local posterior distribution:

$$\text{VoI}_i(\omega_i[t], \mathbf{z}_i[t]) = D_{KL}(p(\theta|\mathbf{z}_i[t], \omega_i[t])||p(\theta|\omega_i[t])). \tag{6}$$

It should be noted that for general probability distributions, KL divergence can be expensive to compute since one needs to go through all possible values of the random variable. However, under the assumption of conjugate prior, the KL divergence function has a closed form, which only depends on the hyperparameters [29]:

$$\begin{aligned} \text{VoI}_i(\omega_i[t], \mathbf{z}_i[t]) &= f(\omega_i[t], \omega_i[t] + h_i[k], \\ &= f(\omega_i[t], h_i[k]). \end{aligned} \tag{7}$$

3.2 VoI Realized Information Fusion Algorithm

The algorithm is initialized by agreeing on the global prior. Upon recording a new measurement, agent updates a buffer $\hat{\mathbf{z}}_i[t]$ by adding the new measurement into it. At

every time instant t, agents compute the local posterior by adding local update $\hat{h}_i[t]$ to its current estimate of the global prior as follows

$$
\begin{aligned}
\omega_i[t] &= \omega[t] + \hat{h}_i[t], \\
\hat{h}_i[t] &= h(\hat{\mathbf{z}}_i[t], p(\hat{\mathbf{z}}_i[t]|\theta)), \\
&= h(\mathbf{z}_i[k+1:t], p(\mathbf{z}_i[k+1:t]|\theta)), \\
k &= \max\{\kappa\} \text{ s.t. } \kappa < t \text{ and } i \in \nu[\kappa].
\end{aligned}
\tag{8}
$$

The agents then compare the KL divergence between the local posterior and its current estimate of the global prior. If the value exceeds a predefined threshold V^*, the agent labels itself as informative, otherwise the agent labels itself as uninformative. The set of informative agents at any time t is denoted by $\nu(t)$, and can be represented as follows

$$
\begin{aligned}
\nu(t) &= \{i|V_i[t] > V^*\}, \\
V_i[t] &= \mathrm{VoI}_i(\omega[t], \hat{h}_i[t]).
\end{aligned}
\tag{9}
$$

The number of informative agents in the network at any time t is denoted by $|\nu(t)|$. The informative agents broadcast a message containing its local updates to their neighbors, and reinstates the measurement process by re-initializing its buffer. All agents relay every message they receive from an informative agent or a relaying agent. Since each agent has a unique identifying label, it is possible to ensure that messages are not duplicated during relay.

Therefore at the end of tth time slot agents obtain an updated estimate of the global posterior by adding relayed measurements to their previous estimates of global prior:

$$
\forall i, \omega_i[t] = \omega[t] = \omega_i[t-1] + \sum_{i \in \nu} \hat{h}_i[k].
\tag{10}
$$

The algorithm is depicted in Table 3.

3.3 Cost

At t, each agent relays updates for all the informative agents, the number of messages sent out is $|\nu[t]|$. The cost of all agents at t is $|\nu[t]| \sum_{i=1}^{N} C_i$. During period $[0, T]$, the total cost is $\sum_{t=1}^{T} |\nu[t]| \sum_{i=1}^{N} C_i$. In the worst case, $|\nu[t]| = N$ and the cost of VoI Realized Fusion is the same as Brute Force Fusion. However, with an appropriate choice of the threshold V^*, $|\nu[t]| \ll N$ most of the time, therefore VoI Realized Fusion has less communication cost.

4 VoI REALIZED FUSION OVER A SUBNETWORK RELAY

In the previous algorithms considered, it was required that all agents relay information. However, this approach may not be very efficient when the set of informative agents is small. If the number of agents in the informative set $\nu[t]$ increases by one, the

Table 3: VoI Realized Fusion with Full Network Relay

step	description	cost
1	Initialization set a global prior $\omega_i[0] = \omega[0]$:	C_{init}
2	At time t $\forall i$,update buffer $\hat{h}_i[t]$ all agents check whether in informative set $i \in \nu[t]$ if $V_i > V^*$	
3	broadcast and relay updates for informative set $\forall i \in \nu$, broadcast $\hat{h}_i[t]$ $\forall j$, relay $\hat{h}_i[t], i \in \nu[t]$	$\forall i, |\nu[t]| C_i$
4	update $\forall i, \omega_i[t] = \omega_i[t-1] + \sum_{j \in \nu[t]} \hat{h}_j[t])$	
5	$t = t + 1$, goes to step 2	
Total cost in $[1:T]$, $C_{init} + \sum_{t=1}^{T} |\nu[t]| \sum_{i=1}^{N} C_i$		

communication cost increases by $\sum_i C_i$. This can be a potential waste of resource, because most agents in the network need not know every single update but rather are interested only in the final estimate of the global posterior. In this section, we consider the case of VoI realized fusion over a smaller subnetwork.

4.1 Algorithm

The agents initialize similar to the algorithm of Section 3.2. In addition, at every time step the agents agree on what the informative set is. Based on the informative set and network topology, agents can locally run an algorithm to work out a connected subgraph of the original graph of the network that contains all informative agents. The set of all agents in this subgraph is denoted as $\tilde{\nu}$. This problem can be solved by variations of Shortest Path algorithms, *e.g.,* Dijkstra's algorithm [30], Floyd-Warshall algorithm [31]. The VoI Realized Fusion is now performed on the subnetwork. After the agents in the subnetwork compute the posterior distribution, they broadcast the result to others. The algorithm is depicted in Table 4.

4.2 Cost

The total cost consists of two parts, the first is the communication cost incurred when the agents determine the set of informative agents at each time step c_i, and the second is cost of hyperparameter updates C_i. In most cases $c_i \ll C_i$. At every time step t, every node needs to relay messages indicating whether every other node is in the informative set, the cost therefore is $N \sum_{i=1}^{N} c_i$. The cost of fusion in subnetwork is $|\nu[t]| \sum_{i \in \tilde{\nu}[t]} C_i$. The cost of broadcasting agreed posterior on subnetwork is $\sum_{i=1}^{N} C_i$ During period $[0, T]$, the cost would be $T \sum_{i=1}^{N} (Nc_i + C_i) + \sum_{t=1}^{T} |\nu[t]| \sum_{i \in \tilde{\nu}[t]} C_i$.

Table 4: VoI Realized Fusion with Sub Network Relay

step	description	cost
1	Initialization set a global prior $\omega_i[0] = \omega[0]$:	C_{init}
2	At time t $\forall i$,update buffer $\hat{h}_i[t]$ all agents check and agree on the informative set $\forall i, i \in \nu[t]$ if $V_i > V^*$	Nc_i
3	Compute strongly connected subnetwork $\forall i$, compute $\tilde{\nu}[t] \supseteq \nu[t]$, s.t. $\tilde{\nu}[t]$ is minimal and connected	
4	VoI Realized fusion on $\tilde{\nu}[t]$ $\forall i \in \tilde{\nu}, \omega_i[t] = \omega_i[t-1] + \sum_{j \in \tilde{\nu}[t]} \hat{h}_j[t])$	$|\nu[t]|$
5	Broadcast and relay result to non sub-net agents	$\forall i, C_i$
6	$t = t + 1$, goes to step 2	

Total cost in $[1:T]$, $C_{init} + \sum_{t=1}^{T} \sum_{j \in \tilde{\nu}[t]} |\nu[t]|C_j$

$+ T \sum_{i=1}^{T}(Nc_i + C_i)$

Therefore, it is seen that this approach avoids the full network relay, but adds an overhead on communicating the informative set and computing subnetwork. In situations where the set of informative agents is large, savings in the total cost would be realized even with this overhead. If the set of informative agents is small, the overhead can deteriorate any cost savings obtained due to communicating only over a smaller network. It is possible to reduce the overhead cost by increasing the time between deciding on the informative set.

5   Adaptive VoI Realized Fusion

The growth of cost at different stages of estimation can be unbalanced in VoI Realized fusion of Section 3.2. At the early stage of estimation, agents know little about the parameters to be estimated, therefore new measurements tend to contain more information, $|\nu[t]|$ is larger and the communication cost builds up quickly. But at later stages of the estimation process, agents have developed a good understanding of the parameters and new measurements are less informative therefore agents declare themselves informative less frequently. Consequently, the growth of the cost slows down. From the analysis of cost, it can be seen that the cost grows with increasing $|\nu[t]|$, which in turn is dependent on the VoI threshold V^*. The Adaptive VoI Realized Fusion algorithm developed in this section adjusts the VoI threshold V^* to control how fast the cost grows.

Table 5: Adaptive VoI Realized Fusion with Full Network Relay

step	description	cost
1	Initialization set a global prior $\omega_i[0] = \omega[0]$:	C_{init}
2	At time t $\forall i$, update buffer $h_i[t]$ locally check whether in informative set $i \in \nu[t]$ if $V_i > V^*$	$\forall i, NC_i$
3	broadcast and relay updates for informative set $\forall i \in \nu$, broadcast $h_i[t]$ $\forall j$, relay $h_i[t], i \in \nu$	
4	update $\forall i, \omega_i[t] = \omega_i[t-1] + \sum\limits_{j \in \nu} h_j[t])$	
5	adjust V^* according to (11) $\forall i, V^* = \gamma V^*, \gamma \in \{\gamma_1, \gamma_2\}$	
6	$t = t+1$, goes to step 2	

Total cost in $[1:T]$, $C_{init} + \sum_{t=1}^{T} |\nu[t]| \sum_{i=1}^{N} C_i$

5.1   Algorithm

The Adaptive VoI Realized Fusion algorithm is similar in most parts to the VoI Realized Fusion algorithm. The main difference is that after the calculation of the global posterior, agents adaptively adjust V^* according to (11). Note that the algorithm depends only on $|\nu(t)|$. When $|\nu[t]|$ is large, V^* is reduced to reduce communication cost, whereas when $|\nu[t]|$ is small, V^* is increased to improve the accuracy. If all agents have the same estimate of $|\nu[t]|$, their estimates of V^* will be the same. In (11) C_u^* and C_l^* denote the tunable upper bound and lower bound on $\bar\nu[t-l+1:t]$, which is the average of $|\nu(t)|$ over the interval $[t-l+1:t]$.

$$
V^* = \begin{cases} \gamma_1 V^* & \bar{v}[t-l+1:t] < C_l^* \\ V^* & C_l^* \leq \bar{v}[t-l+1:t] < C_u^* \\ \gamma_2 V^* & \bar{v}[t-l+1:t] \geq C_u^* \end{cases} \tag{11}
$$
$$\gamma_1 < 1, \gamma_2 > 1$$

5.2   Cost

The cost function for the adaptive VoI case is the same as VoI Realized Fusion: $\sum_{t=1}^{T} |\nu[t]| \sum_{i=1}^{N} C_i$. However, the cost incurred is different, since in the long run, $|\nu[t]|$ is bounded between $[C_l^*, C_u^*]$.

Table 5.2 compares the communication cost of the algorithms discussed in this paper.

Table 6: Cost Summary

Algorithm	Total Cost
Brute Force	$C_{init} + NT \sum_{i=1}^{N} C_i$
HPC	$C_{init} + T \sum_{i=1}^{N} C_i$
Full Network VoI	$C_{init} + \sum_{t=1}^{T} |\nu[t]| \sum_{i=1}^{N} C_i$
Sub Network VoI	$C_{init} + \sum_{t=1}^{T} \sum_{j \in \tilde{\nu}[t]} |\nu[t]| C_j + T \sum_{i=1}^{T} (Nc_i + C_i)$
Adaptive VoI	$C_{init} + \sum_{t=1}^{T} |\nu[t]| \sum_{i=1}^{N} C_i$

6 NUMERICAL STUDY

The algorithms described in previous sections are compared in terms of communication cost incurred in this section using numerical simulations. The goal of the agents is to estimate the parameter γ of a Poisson distribution. The conjugate prior of Poisson distribution is Gamma distribution $p_{\Lambda|A,B}$. The likelihood function from which the measurements are drawn from is a Poisson distribution $p_{X|\Lambda,T}$.

$$
\begin{aligned}
p_{X|\Lambda,T} &= \frac{(\lambda t)^x e^{-\lambda t}}{x!} \\
p_{\Lambda|A,B} &= \frac{\beta^\alpha \lambda^{\alpha-1} e^{-\beta\lambda}}{\Gamma(\alpha)}
\end{aligned}
\tag{12}
$$

The hyperparameters are α and β. The update law is: [29]

$$
\alpha \leftarrow \alpha + x, \quad \beta \leftarrow \beta + t.
\tag{13}
$$

The total number of agents in the network is a hundred. Measurements from each agent have a bias which is normally distributed $\lambda_i \sim N(5, 2)$. At each time step, every agent take one measurement, $z_i[t] \sim Poi(\lambda_i)$. The communication cost for each message containing the updated hyperparameters is $C_i = 1$.

6.1 VoI Realized Information Fusion

Figure 2 shows the cost incurred during the simulation of the Brute Force, HPC, and VoI Realized information fusion algorithms with $V^* = 0.02, 0.1, 0.5$. As expected, the cost of the consensus based approach (HPC) is significantly better than Brute Force. At the beginning, VoI Realized Fusion with lower thresholds has higher cost, however the growth of the cost slows down quickly. The KL divergence to the centralized estimate of the posterior is compared in Figure 3. Note that the centralized estimate is the same as that of the Brute Force method. It can be seen that the HPC error is decreasing over time but is non-zero, this is in agreement with the fact that the consensus algorithm is guaranteed to converges asymptotically. The error of VoI realized fusion decreases

with decreasing threshold V^*. The lower thresholds have performance comparable to that of HPC.

6.2 VoI Realized Fusion in Sub-Net

In this section the VoI realized information fusion algorithm on a subnetwork (see Section 4) is analyzed. The cost incurred in communicating the informative set of agents is assumed to be $c_i = 0.01$. Figure 4 and Figure 5 show the communication cost and the KL divergence to centralized posterior of VoI Realized fusion with subnetwork relay ($V^* = 0.1$) and fusion with full-network relay ($V^* = 0.02, 0.1, 0.5$). The error traces of VoI and subnetwork VoI with $V^* = 0.1$ overlap, this is expected, because they share the same algorithm for calculating the informative set, and are run on the same set of data. The figure highlights the fact that the error for VoI and subnetwork fusion will be the same. The cost of subnetwork VoI is less because relay is limited within a small set. However, the reduction in cost is not significant. These results indicate that information fusion over a subnetwork may not result in significant cost savings due to the overhead in communicating the informative set.

6.3 Adaptive VoI Realized Fusion

The parameters of the adaptive algorithm of 11 are set to $C_u^* = 15$, $C_l^* = 1$, $\gamma_1 = 0.99$, $\gamma_2 = 1.01$, $l = 10$. The adaptive VoI V^* is initialized at 0.5. The cost is compared in figure 6 and the KL divergence to the centralized estimate of the posterior is compared in Figure 7. The results indicate that adaptive VoI Realized fusion strikes a good balance between the cost incurred and the estimation error. The rate of cost increase is bounded, and cost is seen to increase in almost a linear pattern. Furthermore, the error reduces over time. It can be observed that with a cost of no more than $V^* = 0.1$ of VoI Realized Fusion, this adaptive method can achieve similar error with $V^* = 0.02$ of VoI Realized Fusion. The evolution of V^* is shown in Figure 8. The adaptive algorithm (11) increases the V^* initially in response to the presence of a large number of informative agents in the network, and then reduces V^* as the number of informative agents drop.

6.4 Comparison of cost and estimation accuracy

In Figure 9 the performance of the algorithms discussed is compared in cost-error coordinates. The horizontal axis represents the final cost at the end of simulation and the vertical axis represents the average KL-divergence to centralized result in last 300 time steps. An ideal algorithm would be situated in the bottom left corner of that graph, since it would have low error and low communication cost. The consensus algorithm (HPC) is situated in the bottom right corner, with low error but high cost. VoI Realized Fusion with bigger V^* thresholds (e.g. $V^* = 0.5$) are in the left corner, with low cost (because the agents do not declare themselves as informative easily) but high error. Adaptive VoI Realized Fusion is situated closest to the lower left corner than fixed VoI algorithms.

7 CONCLUSION

In this paper we considered the problem of distributed estimation in presence of nodes with disparate Value of Information (VoI). The problem of hyperparameter estimation was formulated through a consensus framework as well as a VoI based information fusion framework. The results agree with the current state-of-the art that the communication cost when using a consensus based approach is better than a brute force information fusion approach where every agent relays every other agent's information. However, the results also indicated that an information fusion approach, in which only informative agents communicate their measurements, and others censor their measurements based on a VoI based metric outperforms the consensus algorithm in terms of communication cost. The accuracy of the solution achieved by the consensus was found to be in general better than VoI based approach. An adaptive-VoI information fusion approach was presented that adjusted the set of informative agents based on cost of information. The results indicated that the adaptive approach strikes an excellent balance between cost of communication and accuracy of estimates.

References

[1] D. Hall and J. Llinas, "An introduction to multisensor data fusion," *Proceeedings of the IEEE*, vol. 85, no. 1, pp. 6–23, 1997.

[2] T. Berg and H. Durrant-Whyte, "Distributed and decentralized estimation," in *Intelligent Control and Instrumentation, 1992. SICICI '92. Proceedings., Singapore International Conference on*, vol. 2, feb 1992, pp. 1118 –1123.

[3] J. Ferber, *Multi-Agent Systems : An Introduction to Distributed Artificial Intelligence*. New York: Addison-Wesley, 1999.

[4] P. K. Varshney, *Distributed Detection and Data Fusion*. Springer-Verlag New York Inc., 1997.

[5] A. Makarenko, A. Brooks, T. Kaupp, H. Durrant-Whyte, and F. Dellaert, "Decentralised data fusion: A graphical model approach," in *Information Fusion, 2009. FUSION '09. 12th International Conference on*, july 2009, pp. 545 –554.

[6] V. Gupta, "Distributed estimation and control in networked systems," Ph.D. dissertation, California Institute of Technology, 2006.

[7] R. Olfati-Saber, E. Franco, E. Frazzoli, and J. S. Shamma, "Belief Consensus and Distributed Hypothesis Testing in Sensor Networks," *Workshop of Network Embedded Sensing and Control*, vol. October, 2005.

[8] M. Egerstedt and M. Mesbahi, *Graph Theoretic Methods in Multiagent Networks*. Princeton University Press, 2010.

[9] J. Yedidia, W. Freeman, and Y. Weiss, *Exploring Artificial Intelligence in the New Millennium*. Science and Technology Books, 2003.

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

[11] R. Olfati-Saber, A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *IEEE Transactions on Automatic Control*, vol. 95, no. 1, pp. 215–233, January 2007.

[12] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Transactions on Information Theory*, vol. 52, pp. 2508 – 2530, June 2006.

[13] E. Msechu and G. Giannakis, "Distributed measurement censoring for estimation with wireless sensor networks," in *Signal Processing Advances in Wireless Communications (SPAWC), 2011 IEEE 12th International Workshop on*, june 2011, pp. 176 –180.

[14] ——, "Distributed measurement censoring for estimation with wireless sensor networks," in *Signal Processing Advances in Wireless Communications (SPAWC), 2011 IEEE 12th International Workshop on*, june 2011, pp. 176 –180.

[15] M. Cetin, L. Chen, J. Fisher III, A. Ihler, R. Moses, M. Wainwright, and A. Willsky, "Distributed fusion in sensor networks," *Signal Processing Magazine, IEEE*, vol. 23, no. 4, pp. 42–55, 2006.

[16] W. P. Tay, "Decentralized detection in resource-limited sensor network architectures," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, 2008.

[17] W. P. Tay, J. Tsitsiklis, and M. Win, "Asymptotic performance of a censoring sensor network," *Information Theory, IEEE Transactions on*, vol. 53, no. 11, pp. 4191 –4209, nov. 2007.

[18] M. UNEY, "Decentralized eestimation under ccommunication constraints," Ph.D. dissertation, MIDDLE EAST TECHNICAL UNIVERSITY, Turkey, 2009.

[19] L. Chen, M. J. Wainwright, M. Cetin, and A. S. Willsky, "Data association based on optimization in graphical models with application to sensor networks," *Mathematical and Computer Modelling*, vol. 43, no. 9–10, pp. 1114 – 1135, 2006.

[20] A. Gelman, J. Carlin, H. Stern, and D. Rubin, *Bayesian Data Analysis.* Chapman and Hall, 1995.

[21] J. Bernardo and A. F. M. Smith, *Bayesian Theory*. Wiley, 1994.

[22] A. Muhammad and A. Jadbabaie, "Decentralized computation of homology groups in networks by gossip," in *American Control Conference*, 2007.

[23] M. Nabi Abdolyousefi and M. Mesbahi, "Network identification via node knock-out," in *IEEE Conference on Decision and Control*, 2010.

[24] G. Chowdhary, M. Egerstedt, and E. N. Johnson, "Network discovery: An estimation based approach," in *American Control Conference*, 2011, submitted.

[25] C. S. Fraser, L. F. Bertuccelli, H.-L. Choi, and J. P. How, "A hyperparameter consensus method for agreement under uncertainty," *Automatica*, vol. 48, no. 2, pp. 374–380, February 2012.

[26] S. M. Ali and S. D. Silvey, "A general class of coefficients of divergence of one distribution from another," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 28, no. 1, pp. 131–142, 1966.

[27] S. Kullback, *Information Theory and Statistics*.   New York: John Wiley and Sons, 1959.

[28] S. Kullback and R. A. Leibler, "On information and sufficiency," *Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[29] F. Nielsen and R. Nock, "Entropies and cross-entropies of exponential families," in *Image Processing (ICIP), 2010 17th IEEE International Conference on*, sept. 2010, pp. 3621 –3624.

[30] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische Mathematik*, vol. 1, pp. 269–271, 1959.

[31] R. W. Floyd, "Algorithm 97: Shortest path," *Commun. ACM*, vol. 5, pp. 345–, June 1962. [Online]. Available: http://doi.acm.org/10.1145/367766.368168
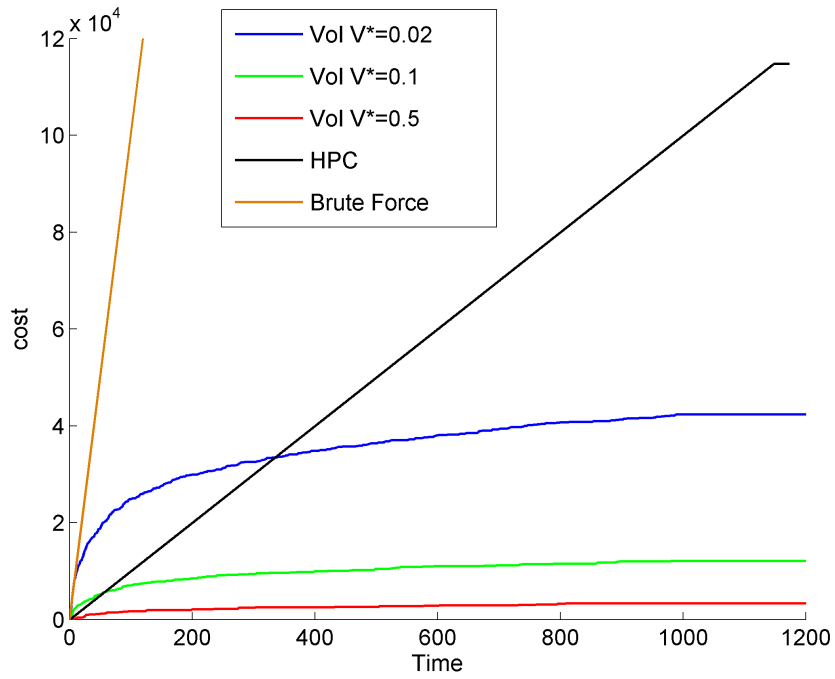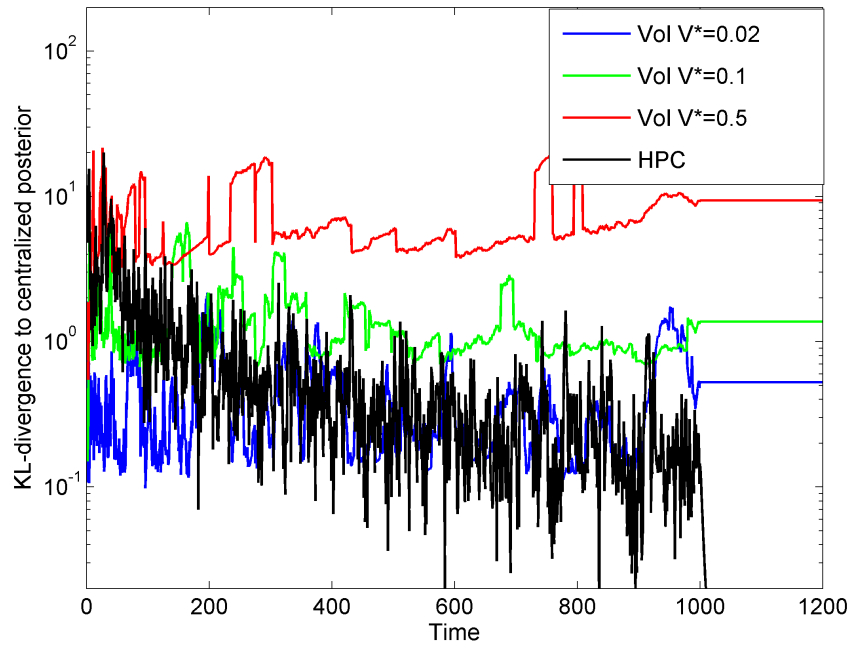
Figure 2: Cost VoI Realized Fusion

Figure 3: KL-divergence to centralized posterior, VoI Realized Fusion
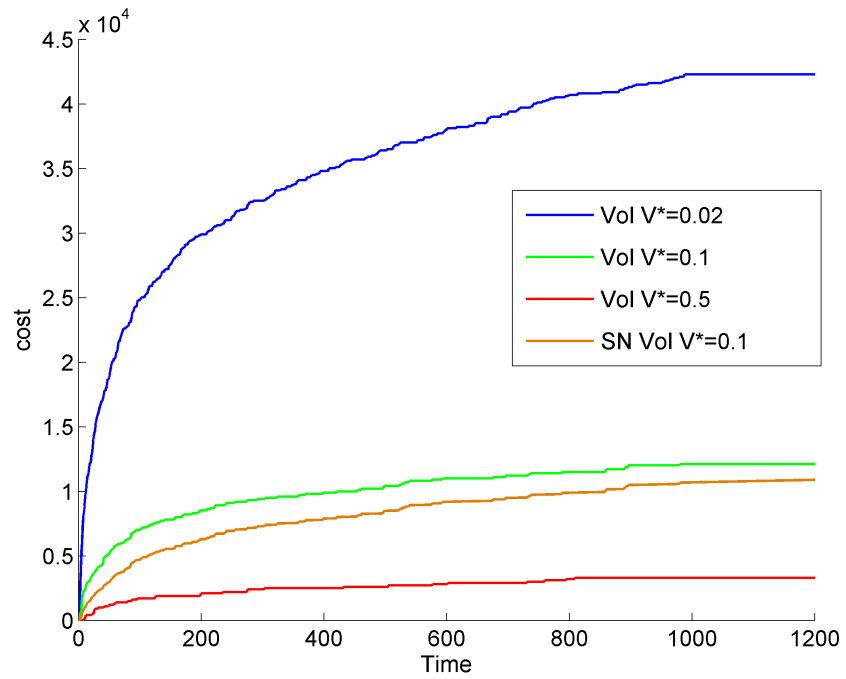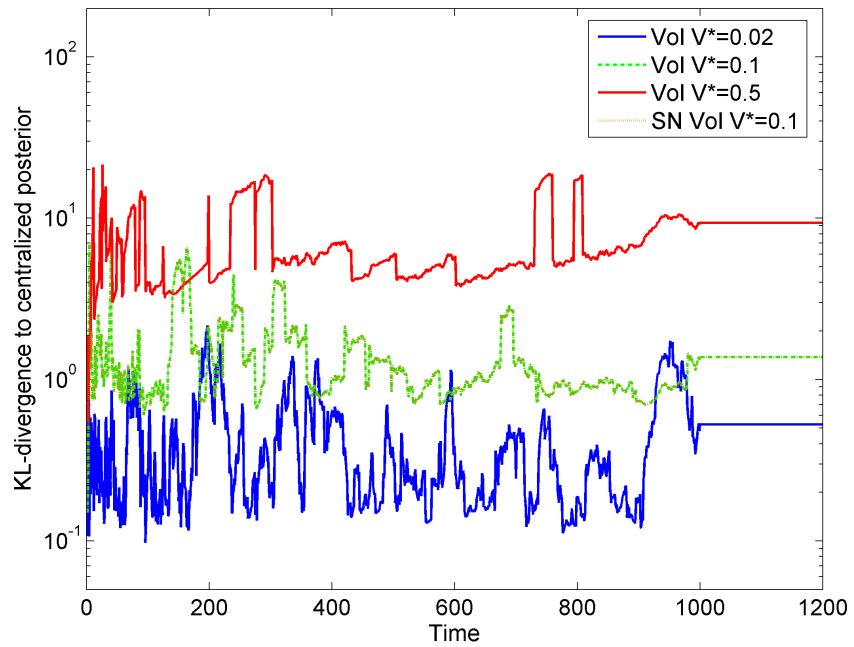
Figure 4: VoI Realized Fusion Sub-Net

Figure 5: KL-divergence to centralized posterior, VoI Realized Fusion Sub-Net
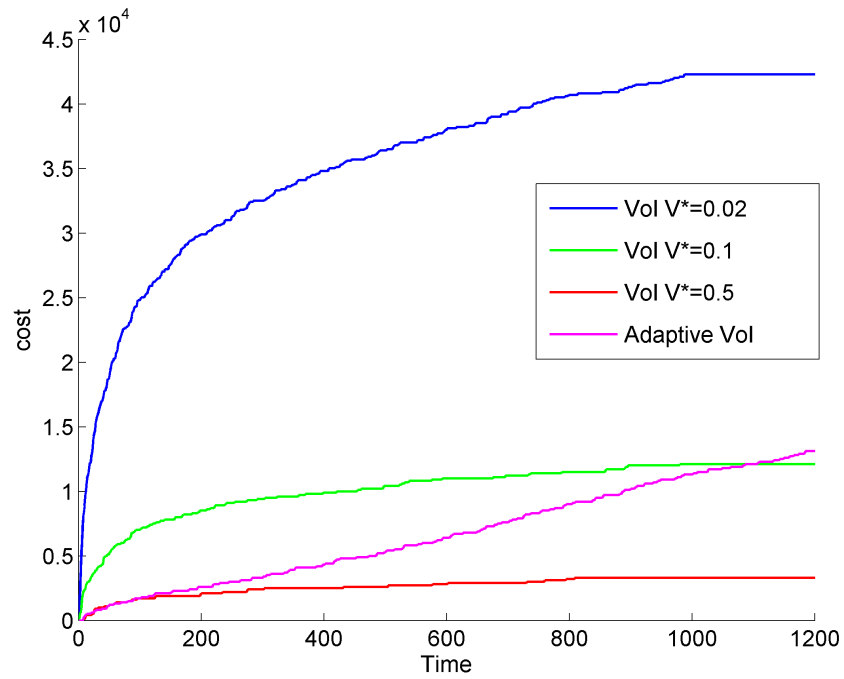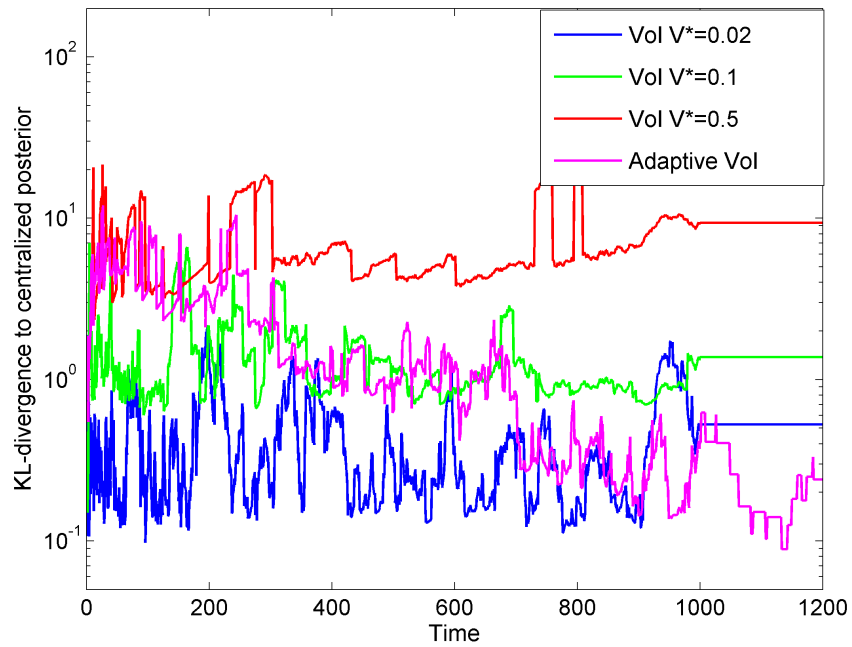
Figure 6: Adaptive VoI Realized Fusion

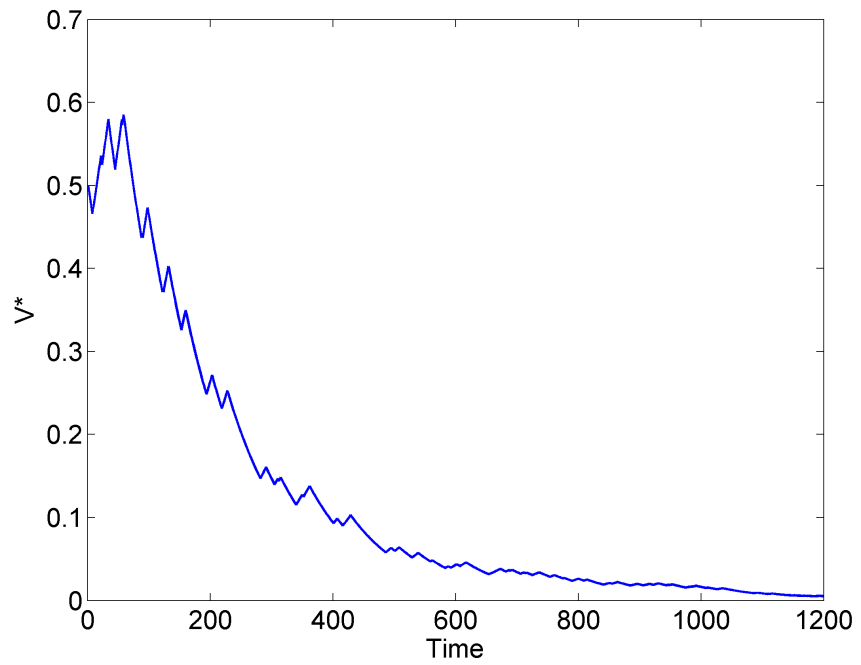Figure 7: KL-divergence to centralized posterior, Adaptive VoI Realized Fusion

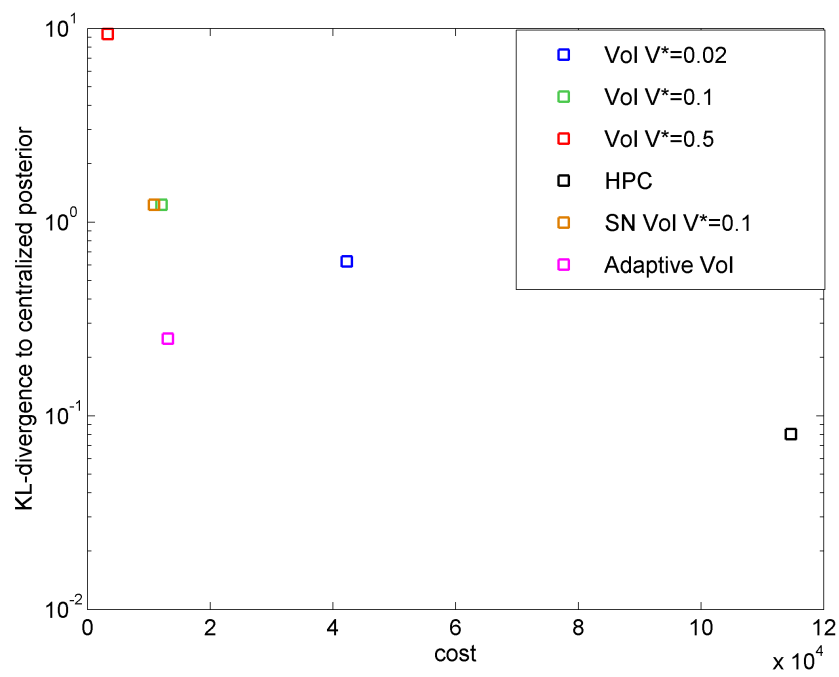Figure 8: Change of VoI threshold V^* in Adaptive VoI Realized Fusion

Figure 9: KL-divergence to centralized posterior vs Cost