Hybrid Data and Decision Fusion Techniques for Model-Based Data Gathering in Wireless Sensor Networks

Lorenzo A. Rossi, Bhaskar Krishnamachari and C.-C. Jay Kuo
Department of Electrical Engineering, University of Southern California
Los Angeles, CA 90089-2564,USA
lrossi@usc.edu, bkrishna@usc.edu, cckuo@sipi.usc.edu

Abstract— **The data gathering problem in wireless sensor networks for environmental monitoring, where the physical phenomena can be modeled by partial differential equations (PDE's), is investigated. Under this context, it suffices for the sensor network to update the base station with estimates of model parameters rather than transmitting raw sensor measurements. In-network processing techniques to estimate the PDE coefficients are presented. A scheme that provides a hybrid combination of decision and data fusion is proposed to find a tradeoff between performance and energy efficiency. The role that the assumptions of PDE models can play in designing such methods is investigated.**

I. Introduction

Wireless sensor networks are used for dense and autonomous monitoring of environments, spanning from ecosystems to industrial plants or vehicles. Data gathering is a fundamental problem in such systems. Its goal is to determine the optimal way (i.e. the least energy consuming way) to transfer node measurements to a remote base station (BS). A typical assumption is that all nodes sample the sensor field uniformly in time and generate a packet for each round of measurements. All packets are delivered to the BS through various aggregation strategies [1].

The dense spatial-temporal signal sampling and transmission is not energy efficient due to strong spatial and temporal correlations of measured and transmitted data. In many real world applications, such correlations can be characterized by relatively simple physical models, *e.g.* a set of partial differential equations (PDE's). For instance, PDE's can be used to describe the diffusion of a gas in the air or of a fluid in the water.

Generally speaking, physical models can often provide fundamental insights into phenomena interesting to sensor network applications [2,3]. In particular, if the goal is data gathering, the PDE model can be used to offer a more compact data representation of the phenomenon being observed as compared with raw sensor measurements. When there are no active sources in the environment, the coefficients of the PDE, along with initial and boundary conditions, allow complete reconstruction of the time-varying scalar field associated with the phenomenon.

In this work, we assume the existence of an underlying PDE model for the physical phenomena being monitored by sensor networks. Then, the base station would be able to predict the evolution of a phenomenon using the model, without the need of a stream of raw measurements from the sensor field. Basically, it should only gather the initial and boundary conditions as well as updated PDE parameters so that the dense data field can be reconstructed from the numerical solution of the PDE problem. It is apparent that transmitting a few PDE coefficients costs much less than sending the densely sampled data field, since these coefficients have a much slower variation rate in the spatial and temporal domains.

A relevant problem in the aforementioned framework is the distributed identification of unknown PDE parameters through measured data of sensor nodes. To the best of our knowledge, the literature presents only centralized solutions to this problem [4]. In a distributed context, it may be worthwhile to have some of the nodes processing measurements from the neighbors for robust estimation, which is known as data fusion. On the other hand, the model parameters may first be estimated from nodes using their own measurements only, and then each node exchanges the estimate with its neighbors for further refinement. This is called decision fusion. Data fusion generally outperforms decision fusion in accuracy at the price of a higher communication cost. The focus of this work is to provide some hybrid data and decision fusion techniques for the estimation of parameters of PDE models in wireless sensor networks.

The rest of the paper is organized as follows. Section II gives some background on the PDE model to be used in this work. The problem formulation is presented in Section III. The proposed distributed identification scheme using hybrid data/decision fusion is presented in Section IV. Simulation results are shown in Section V.

II. The Parabolic Equation

Consider a physical phenomenon represented by the space and time varying scalar field $x(\xi,t)$, where $0<\xi<L$ and $t>0$, that satisfies the following relationship:

$$x_t(\xi,t)=\theta_1 x_{\xi\xi}(\xi,t)+\theta_2 x_{\xi}(\xi,t)+\theta_3 x(\xi,t)+u(\xi,t), \quad (1)$$

where $x_t(.)$, $x_{\xi\xi}(.)$ and $x_{\xi}(.)$ denote the partial derivatives of $x(.)$ with respect to time, t, and space, ξ, and $u(\xi,t)$ represents a source term. The PDE (1) is called the *parabolic*

equation.[1]

A particular case of (1) is the *diffusion equation* $x_t(\xi,t) = Dx_{\xi\xi}(\xi,t)$. The scalar D is called *diffusivity*. Typically, $x(\xi,t)$ represents a concentration function. The meaning of the diffusion equation is that there is net flow of substance from the regions of higher concentration of the substance to the ones of lower concentration [5]. An interesting property of the diffusion equation is the *smoothing* effect. Thanks to the proportionality between time derivative $x_t(\xi,t)$ and local curvature $x_{\xi\xi}(\xi,t)$, the spatial profile of $x(\xi,t)$ is subject to a low pass filtering action over time.

The parabolic equation can be solved with the knowledge of the initial condition (IC), $x(0,\xi) = x_0(\xi)$, and the boundary conditions (BC's). The parameters θ_2 and θ_3 are sometime referred to as *velocity*, and *dissipation* (or *dispersion*), respectively. For some particular initial/boundary conditions and parameters, the diffusion equation can be solved analytically. Otherwise, numerical methods are required.

To solve the diffusion equation by means of numerical methods, the derivatives must be approximated by finite difference as

$$x_\xi(\xi,t)\big|_{\xi=ih} \approx \frac{x_{i+1}(t) - x_i(t)}{h}, \tag{2}$$

$$x_{\xi\xi}(\xi,t)\big|_{\xi=ih} \approx \frac{x_{i+1}(t) + x_{i-1}(t) - 2x_i(t)}{h^2}, \tag{3}$$

where h is the spatial sampling period. It can be shown that the sampling time t_s must obey the following inequality for the discrete approach to converge [5]: $t_s < \frac{h^2}{2D}$.

III. Problem Formulation

We consider a finite set of sensor nodes $S = \{s_i\}$ deployed over the one-dimensional sensor field $[0, L]$. The nodes measure discrete-time samples of the scalar field $x(\xi,t)$, continuous in time, for $t > 0$, and space, for $0 < \xi < L$. We assume that the evolution in time and space of $x(.)$ can be modeled by the following sourceless parabolic equation:

$$x_t(\xi,t) = \theta_1 x_{\xi\xi}(\xi,t) + \theta_2 x_\xi(\xi,t) + \theta_3 x(\xi,t), \tag{4}$$

with IC $x_0(\xi)$. The two boundary points, $\xi = 0$ and $\xi = L$, are also subject to some boundary conditions (BC's). Note that (4) does not include a source term. The knowledge of sources may not be available in a realistic sensor networks scenario. The topic of the joint source and parameter estimation is an open problem for future research. However, the effect of external sources can be modeled through the BC's.

It is assumed that node measurements are corrupted by additive zero mean white Gaussian noise. Furthermore, nodes $\{s_i\}$, partitioned into clusters $\{S_j\}$, know their absolute locations $\{p_i\}$ in the field and the BC's (e.g. through some data flooding mechanism). We also assume that there is reliable communication among nodes (i.e. no packets get lost) and that member nodes can talk simultaneously to the leader without interference within a cluster.

Here, we focus on the problem of identifying the parameters $\{\theta_j\}$, $j = 1, 2, 3$, at some of the nodes $\{s_i\}$ (i.e. the cluster heads). We assume that the sensor network can transmit parameter estimates, along with IC's and BC's, to the BS through some data gathering mechanisms, so that the BS can predict the measured scalar field, by numerically solving the PDE.

Besides data gathering, distributed estimation of parameters can have several other applications in sensor networks. For instance, the parameters can be estimated in return to a query from the BS to study the physical properties of a particular environment. The estimates of parameters can be used for in-network prediction of the phenomenon. Furthermore, the estimates can be used to perform coding schemes [3].

IV. Distributed Identification of PDE Parameters

In this section, we present an approach to the distributed estimation of PDE parameters. The identification of the parameters is performed at some of the cluster heads in a possibly data fusion modality. That is, a cluster head receives measurements from its member nodes and processes them adaptively. We first define a discrete-time state-space model of the sampled scalar field in Sec. IV-A. Then, we propose an approach to identification of the parameters based on the extended Kalman filter in Sec. IV-B. Finally, some criteria for switching to decision fusion are discussed in Sec. IV-C.

A. Model Discretization and Clustering

In the discretized model for the PDE (4), the scalar field $x(\xi,t)$ is sampled in space and time at points (ih, kt_s), where h and t_s are the sampling periods in space and time, respectively, and i and k are integers. The discrete samples $x_i(k)$ represent the state of the system. The choice of the sampling space h divides the sensor field in $N+1$ points where $N = \frac{L}{h}$. The sampling time t_s must be selected according to the inequality constraint presented at the end of Section (II) and thus it is usually smaller than h. The discretization in space and time of the PDE (4) is obtained by approximating the spatial and time derivatives via the finite difference approximations in (2-3). To complete the definition of the discretized model, called the *lumped system*, the noisy measurements of nodes are related to the state according to the positions of the sensor nodes with respect to the points of the state variable.

The discretization in space and time of PDE and measurements leads to the lumped model described by the following state-space equations:

$$\mathbf{x}(k+1) = \mathbf{A}(\boldsymbol{\theta})\mathbf{x}(k) + \mathbf{B}(\boldsymbol{\theta})\mathbf{u}(k), \tag{5}$$

$$\mathbf{y}_j(k) = \mathbf{C}_j\mathbf{x}(k) + \mathbf{v}(k), \tag{6}$$

[1] For simplicity we consider a uni-dimensional (1-D) case in space. The extension to a higher dimensional case is straightforward [2]. However, note that equation (1) can already model real phenomena such as the diffusion of fluid in a water channel or the propagation of heat in a metallic rod.

where j is the index of cluster S_j. The state vector $\mathbf{x}(k)$ represents the uniformly sampled version of scalar field $x(\xi,t)$ with $x_i(k) := x(ih, kt_s)$. The state matrix $\mathbf{A}(\boldsymbol{\theta})$ is a square matrix depending on the spatial derivatives of $x(.)$ on the right hand side of (4), the BC's, the sampling periods h and t_s and the set $\boldsymbol{\theta}$ of parameters to identify. If the parameters are constant over the space, $\mathbf{A}(\boldsymbol{\theta})$ is given by:

$$\mathbf{A}(\boldsymbol{\theta}) = t_s \sum_{m=1}^{3} \theta_m \mathbf{A}_i + \mathbf{I}, \tag{7}$$

where $\mathbf{I}$ is the identity matrix and matrices $\mathbf{A}_i$ depend on the discretized spatial derivatives on the right hand side of (4).

Equation (6) represents the noisy sensor measurements at cluster S_j. $\mathbf{v}(k)$ is a vector of AWG noise. Matrix $\mathbf{C}_j$ is defined on the basis of the geometrical relations between the sampling points of the scalar field $\mathbf{x}(k)$ and sensor locations. If a sensor is not located in one of the points in the set $\{\xi = ih : i = 1, 2, ...\}$, the relationship between the measurement and the state variables can be defined by either linear or polynomial interpolation.

All cluster leaders share the model defined by (5), through the knowledge of their absolute location in the field, but have different measurement equations given by (6). Note that the general expression of the system model in (5) and (6) is valid even if the coefficients θ_i vary with time and space. The parameters of the lumped model (e.g., spatial sampling h and state dimension) can be passed to the nodes through some information flooding mechanism.

B. Parameter Estimation via Kalman Filtering

We adopt the extended Kalman filter (EKF) for the task of adaptive identification of unknown parameters $\boldsymbol{\theta}$. The key idea of the EKF approach is to treat unknown parameters as additional state variables [6], by defining an *augmented system*, where the state vector is

$$\mathbf{z}(k) := \begin{bmatrix} \mathbf{x}(k) \\ \boldsymbol{\theta} \end{bmatrix}.$$

The augmented system is non linear, since parameters $\boldsymbol{\theta}$ are multiplied with state variables $\mathbf{x}(k)$. Under these circumstances, the Kalman filter is used as the state predictor to the linearized version of the augmented system and the estimates of parameters $\boldsymbol{\theta}$ are obtained because they are treated as additional state variables in the augmented system. The augmented system can be written as

$$\begin{aligned} \mathbf{z}(k+1) &= \mathbf{f}(\mathbf{z}(k), \mathbf{u}(k)) \\ &= \begin{bmatrix} \mathbf{A}(\boldsymbol{\theta})\mathbf{x}(k) + \mathbf{B}(\boldsymbol{\theta})\mathbf{u}(k) \\ \boldsymbol{\theta} \end{bmatrix}, \qquad (8) \\ \mathbf{y}_j(k) &= \mathbf{C}_j \mathbf{x}(k) + \mathbf{v}(k). \qquad (9) \end{aligned}$$

To apply Kalman filtering, the Jacobians of the state and measurement equations must be derived. From (8) and (9), we have the following:

$$\begin{aligned} \mathbf{J}(k) &:= \left. \frac{\partial \mathbf{f}(\mathbf{z}(k)\mathbf{u}(k))}{\partial \mathbf{z}} \right|_{\mathbf{z}=\hat{\mathbf{z}}}, \qquad (10) \\ \mathbf{H}_j &:= [\mathbf{C}_j \;\; \mathbf{0}]. \qquad (11) \end{aligned}$$

Then, the Kalman filter can be applied. Given the mean of the state vector, $\bar{\mathbf{z}}_0$, the initialization for the covariance matrix $\mathbf{P}$ can be found conventionally as

$$\mathbf{P}_0 = E[(\mathbf{z} - \bar{\mathbf{z}}_0)(\mathbf{z} - \bar{\mathbf{z}}_0)^T]. \tag{12}$$

At each measurement time, the filter equations are [6]:

$$\mathbf{L}(k) = \mathbf{P}(k/k-1)\mathbf{H}^T \left(\mathbf{H}\mathbf{P}(k/k-1)\mathbf{H}^T + \mathbf{R}\right)^{-1} \tag{13}$$

$$\mathbf{P}(k/k) = \mathbf{P}(k/k-1) - \mathbf{L}(k)\mathbf{H}\mathbf{P}(k/k-1) \tag{14}$$

$$\hat{\mathbf{z}}(k/k) = \hat{\mathbf{z}}(k/k-1) + \mathbf{L}(k)\left(\mathbf{y}(k) - \mathbf{C}_i\hat{\mathbf{x}}(k/k-1)\right) \tag{15}$$

where $\mathbf{R}$ is the covariance matrix of noise measured by sensors the fact that $\mathbf{H}\hat{\mathbf{z}} = \mathbf{C}\hat{\mathbf{x}}$ is applied in the derivation of (15). Then, the state estimate and the covariance are propagated to the next measurement time via

$$\begin{aligned} \mathbf{P}(k+1/k) &= \mathbf{J}(k)\mathbf{P}(k/k)\mathbf{J}^T(k) \qquad (16) \\ \hat{\mathbf{z}}(k+1/k) &= \mathbf{f}(\hat{\mathbf{z}}(k), u(k)). \qquad (17) \end{aligned}$$

C. Hybrid Data and Decision Fusion Techniques

The distributed algorithm described above implies that the member nodes in a cluster pass their measurements to the cluster head so that the head can perform the EKF. This process is called *data fusion*. The transmission of collected time series from member nodes may be expensive in terms of energy consumption due to a large number of data samples being transmitted, even if the size of a cluster is kept small. An alternative approach is to let each node in a cluster to process its own measurements and then to pass estimated parameter values to the cluster head. This is a *decision fusion* process, which is less expensive in energy consumption. The cluster head needs only to average the received estimates as

$$\hat{\boldsymbol{\theta}}_j = \frac{1}{|S_j|} \sum_{s_l \in S_j} \hat{\boldsymbol{\theta}}_{s_l}, \tag{18}$$

where $|S_j|$ is the number of member nodes s_l in the cluster S_j. However, this approach may result in poor estimates since only the measurements of one sensor are available to the EKF algorithm (i.e. y_{s_i} is a scalar) at each node.

A critical factor to consider for decision fusion based estimates is the spatial bandwidth of the sampled process. Switching too soon to decision fusion can lead into spatial aliasing, and, therefore, wrong estimates, because the spatial sampling frequency decreases. On the other hand, the smoothing effect (as described in Sec. II) will progressively decrease the bandwidth of the sampled field.

To find a good balance between the estimate quality and the communication cost, we consider a hybrid strategy that

integrates in time data and decision fusion. Some preliminary iterations are performed within each cluster via the data fusion modality. Then, the algorithm switches to the decision fusion modality. This allows to save energy, since no time series are exchanged among the nodes after the initial iterations. We propose two criteria for the switching: (1) a function defined on some diagonal terms of covariance matrix associated with parameters estimates and (2) the number of sample being sent. The algorithm switches to decision fusion if either

$$f(p_i(\boldsymbol{\theta})) < t_1, \tag{19}$$

or

$$n_{it} > t_2, \tag{20}$$

where t_1 and t_2 are two thresholds and n_{it} is the number of samples sent to the fusion point since the beginning of the iterative algorithm. The two above conditions should be also integrated with a spatial frequency test to avoid the risk of aliasing.

Condition (20) is used to avoid excessive battery consumption for participating nodes. Threshold t_2 can be decreased as the remaining energy level of nodes is decreasing to extend the lifetime of the network with a soft degradation of the performance.

The simple computation of the arithmetic average of the parameter estimates may have some drawbacks, since the estimates from some of the nodes can be less reliable than others. Our experiments show that the SNR is not the only factor influencing the quality of the estimates. The performance depends also on the rate of variability of the field. In other words, areas where the scalar field is close to be stationary cannot lead to good estimates because the PDE problem turns in an ODE problem at the equilibrium and becomes under-determined. In other words, in the case of equilibrium, there is no longer a net diffusion of concentration to observe and therefore the diffusion coefficients cannot be estimated. Hence, weighted averages given by

$$\hat{\boldsymbol{\theta}}_j = \frac{1}{|S_j|} \sum_{s_l \in S_j} w_l \hat{\boldsymbol{\theta}}_{s_l}, \tag{21}$$

should be performed instead, in order to give more relevance to the estimates coming from areas with better dynamics.

V. Simulation Results

The purpose of the first experiment is to show the effect of spatial aliasing on the estimate performance. Here we consider the parabolic equation

$$x_t(\xi,t) = \frac{1}{50} x_{\xi\xi}(\xi,t) - \frac{1}{50} x_\xi(\xi,t) - .32x(\xi,t). \tag{22}$$

The equation models a phenomenon of so called advection-diffusion with dissipation, due to non zero velocity and dispersion parameters. The IC is $x(\xi,0) = \exp(-(\xi - .3).^2/(.04))$, with homogeneous Dirichlet BC's: $x(0,t) = x(1,t) = 0$. The field $x(\xi,t)$ is shown in Figure 1. It is initially characterized by a relatively large spatial bandwidth, as it is a narrow pulse. In Figure 3, we compare the estimates of the parameters performed according to different modalities: (1) data fusion of three sensor nodes, (2) only one sensor node processing its own data and (3) data fusion with three nodes switched to one after 20 iterations. One node alone since the first iteration gives poor estimates, because of aliasing problems. On the other hand, thanks to the smoothing effect (Sec II), switching to one node after few iterations of data fusion in a cluster gives results comparable to data fusion alone, as it can be also noticed from the estimated scalar field (Fig. 2).

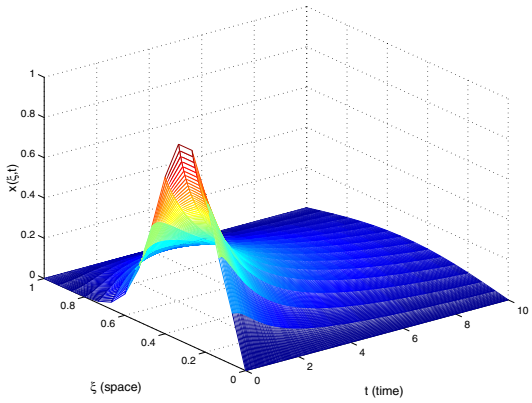

Fig. 1. Scalar field described by the parabolic equation (22). An advection phenomenon can be observed: the concentration is moving toward one of the boundaries.

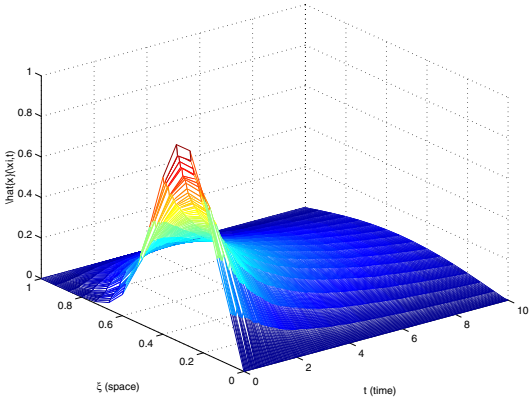

Fig. 2. Estimate of scalar field via hybrid method.

In the second experiment, we want to study how the quality of the estimate varies with respect to the noise and to the location of the nodes sampling the field. In this setup, a single node estimates the diffusivity parameter from its own noisy readings. We measure how the estimation error varies with the location of the node in the space and the level of the noise. The measured scalar field is modeled by the diffusion equation

$$x_t(\xi,t) = \frac{1}{\pi^2} x_{\xi\xi}(\xi,t),$$

defined in the interval $0 < \xi < 1$ and $t > 0$, with initial condition $x(\xi,0) = 2\cos(\frac{\pi}{2}\xi)$ and Dirichlet boundary conditions: $x(0,t) = 2$, $x(1,t) = 0$.

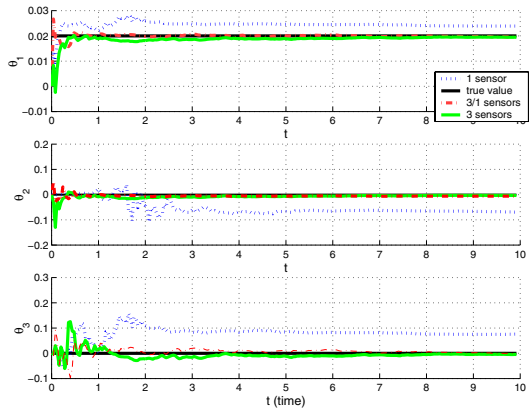

Fig. 3. Estimates of coefficients via data fusion of three sensors, one sensor and then three switched to one sensors.

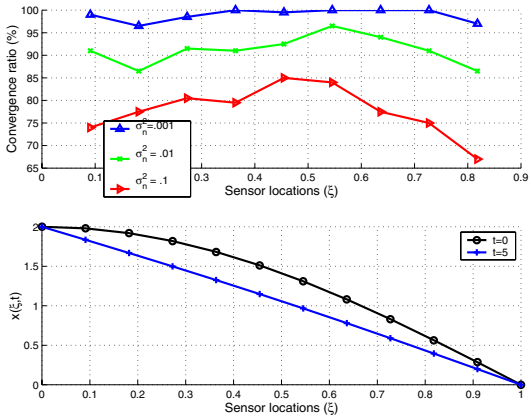

Fig. 4. Percentage of converging iterations versus sensor location and noise level (top). Profile of the scalar field at the initial state and at the steady state (bottom).

A single node s in position p is estimating the diffusivity parameter with sampling time $t_s = 0.004$. Note that the sampling space of the lumped model (5)-(6) is set to $1/11$. We assume that the node knows the BC's. The position p of s varies uniformly on the ξ axis. Different noise levels, σ_n^2, are considered. The average percent error is measured over 200 Monte Carlo trials and for $p = k/11, k \in \{1, 2, ..., 10\}$. We notice that the algorithm does not always converge. The divergent iterations are excluded from the computation of the mean error.

The percentage of convergent trials is shown in the upper part of Figure 4 while the mean error w.r.t. the sensor location is shown in Figure 5 for two different levels of noise. The profiles of the scalar field at the initial time and at the steady state are also shown in Figure 4. It can be noticed that the performance is worst closer to the boundaries and it is not simply proportional to the SNR. The latter parameter is monotonically decreasing when approaching the point $\xi = 1$, but the estimation performance seems to depend also on the rate of temporal variability of the sample scalar field, as mentioned in Subsec. IV-C.

For the same experimental setup, but with three nodes involved in the process, we compare the performance of data fusion with two different modalities of decision fusion: consisting respectively in the arithmetical and weighted averages of the estimates of the three nodes. Table V compares the errors in estimating the diffusivity. The weights are defined proportional to the distance of the nodes to the boundaries.

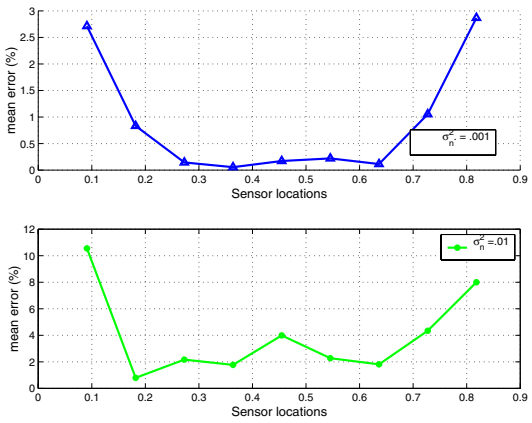

Fig. 5. Percent mean estimation of parameter error versus sensor location and noise level.

TABLE I

Comparison between data and decision fusion

	Data fusion	Decision fusion	Weighted decision f.
Mean err.	.44%	1.85%	1.78%
Std	3.57 %	7.98 %	2.61%

VI. Conclusion

This work addressed the problem of the distributed identification of parameters of PDE models for environmental monitoring in wireless sensor networks. In a source free environment, PDE models allow a compact representation of physical phenomena which is based on initial conditions, BC's and PDE coefficients. This representation is suitable for data gathering applications. Here we proposed a hybrid data and decision fusion approach to the estimation of PDE parameters. We focused our attention on the constraints and guidelines that come out by assuming PDE models for the phenomena being monitored. Future research will study more in depth the relationships between node locations and estimation performance.

References

[1] K. Kalpakis, K. Dasgupta, and P. Namjoshi, "Maximum lifetime data gathering and aggreagation in wireless sensor networks," in *IEEE International conference on networking*, Aug. 2002, pp. 685–696.

[2] L. A. Rossi, B. Krishnamachari, and C.-C. Jay Kuo, "Monitoring of diffusion processes with pde models in wireless sensor networks," in *Defense and Security Symposium 2004*, April 12-16 2004.

[3] B. Beferull-Lozano, R. L. Konsbruck, and M. Vetterli, "Rate-distortion problem for physics based distributed sensing," in *Information Processing in Sensor Networks (IPSN)*, April 26-27 2004.

[4] Y. Orlov and J. Bentsman, "Adaptive distributed parameter systems identification with enforceable identifiability conditions and reduced-order spatial differentiation," *IEEE Trans. on Automatic Control*, vol. 45, no. 2, pp. 36–50, February 2000.

[5] R. Ghez, *Diffusion Phenomena*, Kluwer Academic/Plenum Publishers, 2nd edition, 2001.

[6] J. M. Mendel, *Lessons in Estimation Theory for Signal Processing, Communications and Control*, Prentice Hall, 1995.