INNOVATIONS-BASED SAMPLING OVER SPATIALLY-CORRELATED SENSORS

Zhi Quan and Ali H. Sayed

Electrical Engineering Department, University of California at Los Angeles, CA 90095
Email: {quan, sayed}ee.ucla.edu

ABSTRACT

We consider an estimation network of many distributed sensors, where each senor takes a noisy measurement of some unknown parameter. Due to energy limitation, the network selects only a subset of sensors for data fusion as long as the distortion is tolerable. In this paper, we present a sampling framework based on linear minimum variance unbiased estimation. The framework enables the system to achieve a desired estimation fidelity level and to improve the network lifetime. Simulations illustrate the effectiveness of the proposed sampling schemes.

Index Terms— Sampling, estimation, minimum mean-squared error, and sensor networks

1. INTRODUCTION

A number of wireless sensor networks (WSNs) are being developed to estimate physical phenomena over time and space in noisy environments [1]. The associated fusion center consolidates data collected from sensors to reconstruct the state of nature, e.g., estimating a field variable given the sensor observations. Some key issues in such a setting are the fidelity at which the field variable can be estimated by the data fusion center and the cost of operating the sensor network.

Because wireless sensor devices are usually battery powered, energy efficiency is critical for sensor networks and has a direct influence on the system lifetime. It is necessary to select a group of sensors for data fusion and set the other nodes inactive (or sleeping) so as to conserve energy. Consequently, the design of sensor networks naturally requires the selection of a subset of sensors that are sufficient to meet the fidelity constraint. The goal of this paper is to propose a procedure to select a group of sensor nodes that can maximize the system lifetime while satisfying a desired fidelity.

The problem of sensor selection has been investigated for various purposes. A recent work in [2] proposed a maximum mutual information algorithm where only a *single* sensor is active at any given time and it passes its measurement to the most informative sensor which will be the next active node. In [3], the authors used a local greedy strategy to select the next most informative sensor node to reduce information entropy for target location. The problem of selecting sensors to minimize error in estimating the target position was investigated for a bounded uncertainty sensing model in [4].

In this paper, we present an energy efficient sampling framework based on linear minimum variance unbiased estimation. Each cycle, the fusion center selects a subset of sensors to estimate an unknown parameter. The proposed sampling algorithm, called innovations sensor sampling (ISS), is to select the sensors that are most informative so that the fidelity requirement can be satisfied with no more sensors than necessary. Furthermore, we show that by taking energy consumption into consideration, the energy load can be evenly distributed among sensors to achieve fairness. The performance of these two schemes are evaluated through simulations.

2. ESTIMATION IN SENSOR NETWORKS

Consider a WSN with N sensor nodes and a fusion center. Each sensor has the capability to observe a certain signal in the field and send data to the fusion center. To ensure a desired SNR level at the receiver, the measurement of sensor i should be transmitted to the fusion center at a power level p_i proportional to d_i^α, where d_i is the distance between sensor i and the fusion center and α ($2 < \alpha < 6$) is the pathloss coefficient. We assume that the wireless channels from sensors to the fusion center are scheduled by TDMA so that there is no collision or interference. From the fusion center to sensors, there is a separate broadcast channel through which the sampling decision is sent back to the sensors for activity control.

Each sensor makes observations of an unknown deterministic vector $\boldsymbol{\theta} \in \mathcal{R}^{m\times 1}$, which is distorted by a matrix $\boldsymbol{H}_i \in \mathcal{R}^{n_i\times m}$ and corrupted by additive noise, i.e.,

$$\boxed{\boldsymbol{y}_i = \boldsymbol{H}_i\boldsymbol{\theta} + \boldsymbol{v}_i \qquad i = 1, 2, \ldots, N} \tag{1}$$

Equation (1) can be represented as $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{\theta} + \boldsymbol{v}$, where $\boldsymbol{y} = \mathrm{col}\{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_N\}$, $\boldsymbol{H} = \mathrm{col}\{\boldsymbol{H}_1, \boldsymbol{H}_2, \ldots, \boldsymbol{H}_N\}$, and $\boldsymbol{v} = \mathrm{col}\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_N\}$. The measurement noise $\boldsymbol{v}$ is zero-mean Gaussian and has covariance matrix $\boldsymbol{C}$, i.e., $\boldsymbol{v} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{C})$.

Denote by $\mathcal{A}_n = \{i_1, i_2, \ldots, i_n\}$ the sampling decision, i.e., the index set of selected sensors. Then, the estimator based on the selected senors is represented as $\hat{\boldsymbol{\theta}}\left(\boldsymbol{y}_{\mathcal{A}_n}\right)$.

This work was partially supported by NSF grants ECS-0401188 and ECS-0601266.

There are different data fusion functions for estimation purposes. A common approach is to restrict the estimator to be a linear function of the data and to find the linear estimator that is unbiased and has minimum error. This estimator, referred to as the linear minimum-variance unbiased estimator [5], can be determined with knowledge of only the first and second moments of the data statistics. We will derive the estimator and its associated error under a sampling decision $\mathcal{A}_n$, so it is useful to define some data structures.

Definition 1 *A partial vector of $\boldsymbol{y}$ dictated by $\mathcal{A}_n$, denoted by $\boldsymbol{y}_{\mathcal{A}_n}$, is a vector that contains entries $\boldsymbol{y}_i$ if $i \in \mathcal{A}_n$.*

For example, a partial vector of $\mathrm{col}\{\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3\}$ dictated by $\{1, 3\}$ is $\mathrm{col}\{\boldsymbol{y}_1, \boldsymbol{y}_3\}$. Similarly, we can define the partial matrix as follows.

Definition 2 *A partial matrix of the $N \times 1$ block matrix $\boldsymbol{H}$ dictated by $\mathcal{A}_n$, denoted by $\boldsymbol{H}_{\mathcal{A}_n}$, is a matrix that contains block matrices $\boldsymbol{H}_i$ if $i \in \mathcal{A}_n$; a partial matrix of the $N \times N$ block matrix $\boldsymbol{C}$ dictated by $\mathcal{A}_n$, denoted by $\boldsymbol{C}_{\mathcal{A}_n}$, is a matrix that contains block matrices $\boldsymbol{C}_{ij}$ if $i \in \mathcal{A}_n$ and $j \in \mathcal{A}_n$.*

Then, for any sampling decision $\mathcal{A}_n = \{i_1, i_2, \ldots, i_n\}$, $\boldsymbol{H}_{\mathcal{A}_n} = \mathrm{col}\{\boldsymbol{H}_{i_1}, \boldsymbol{H}_{i_2}, \ldots, \boldsymbol{H}_{i_n}\}$ and the covariance matrix $\boldsymbol{C}_{\mathcal{A}_n}$ is the partial matrix selected from $\boldsymbol{C}$ with rows and columns corresponding to $\{i_1, i_2, \ldots, i_n\}$.

Using the linear data model in (1), the minimum-variance unbiased estimator (m.v.u.e.) of $\boldsymbol{\theta}$ using $\boldsymbol{y}_{\mathcal{A}_n}$ is given by [5]

$$\hat{\boldsymbol{\theta}}\left(\boldsymbol{y}_{\mathcal{A}_n}\right) = \left(\boldsymbol{H}_{\mathcal{A}_n}^T \boldsymbol{C}_{\mathcal{A}_n}^{-1} \boldsymbol{H}_{\mathcal{A}_n}\right)^{-1} \boldsymbol{H}_{\mathcal{A}_n}^T \boldsymbol{C}_{\mathcal{A}_n}^{-1} \boldsymbol{y}_{\mathcal{A}_n} \tag{2}$$

and the resulting minimum mean-squared error (MMSE) is

$$\boldsymbol{D}\left(\mathcal{A}_n\right) = \left(\boldsymbol{H}_{\mathcal{A}_n}^T \boldsymbol{C}_{\mathcal{A}_n}^{-1} \boldsymbol{H}_{\mathcal{A}_n}\right)^{-1} \tag{3}$$

The MMSE should be less than or equal to some desired distortion, i.e., $\mathcal{A}_n$ should be selected such that $\mathrm{Tr}\left[\boldsymbol{D}\left(\mathcal{A}_n\right)\right] \leq D_0$.

3. INNOVATIONS-BASED SAMPLING

In this section, we present a sampling framework based on innovations. At the beginning of each period, the fusion center starts with a seed node i_1, i.e., $\mathcal{A}_1 = \{i_1\}$. One may choose the seed node according to different criteria. A simple criterion is to choose the sensor that has the most energy level. The fusion center then iteratively adds one sensor into the set of selected sensors until the desired fidelity is achieved. The fusion center activates the selected sensors for estimation and powers them off at the end of the period. When the next period begins, a new group of sensors will be activated and the procedure repeats until the sensors deplete their energy.

3.1. Uncorrelated Noise

If the noises were spatially uncorrelated, i.e., $\boldsymbol{C}_{kl} = \boldsymbol{0}$ for $k \neq l$, then (3) becomes

$$\boldsymbol{D}\left(\mathcal{A}_n\right) = \left(\sum_{k=1}^{n} \boldsymbol{H}_{i_k}^T \boldsymbol{C}_{i_k i_k}^{-1} \boldsymbol{H}_{i_k}\right)^{-1} \tag{4}$$

This expression decouples the contribution of each sensor to the total MMSE value $\boldsymbol{D}_{\mathcal{A}_n}$. Each term $\boldsymbol{H}_{i_k}^T \boldsymbol{C}_{i_k i_k}^{-1} \boldsymbol{H}_{i_k}$ has the essential properties of an information measure in that it is

1. non-negative definite,
2. and additive for independent observations.

Intuitively, the more information sensor i_k has (i.e., the larger $\boldsymbol{H}_{i_k}^T \boldsymbol{C}_{i_k i_k}^{-1} \boldsymbol{H}_{i_k}$ is), the lower its contribution to the MMSE. This suggests that at each step the network should choose the most informative sensor in order to maximally reduce the MSE. Specifically, assume that nodes $\mathcal{A}_{n-1} = \{i_1, i_2, \ldots, i_{n-1}\}$ have been selected. Then, at step n we would choose the sensor i_n whose information measure $\boldsymbol{H}_{i_n}^T \boldsymbol{C}_{i_n}^{-1} \boldsymbol{H}_{i_n}$ is the largest. In this way, the resulting $\boldsymbol{D}_{\mathcal{A}_n}$ will be the smallest compared with other choices for i_n.

3.2. Correlated Noise

However, the noises are generally spatially correlated. In this case, the contributions of the individual sensors are coupled in the MMSE expression $\boldsymbol{D}_{\mathcal{A}_n}$ in (3). We thus need to develop a procedure to find the most informative sensor with respect to the previous selected ones. To achieve this goal, we start by whitening the observation data subject to the order dictated by the choice of sensors, and then obtain a set of transformed measurements with uncorrelated noises.

Suppose that we have already selected $n-1$ sensors, i.e., $\mathcal{A}_{n-1} = \{i_1, i_2, \ldots, i_{n-1}\}$. For any $i_n \notin \mathcal{A}_{n-1}$, we define its innovation [5, 6] as

$$\boldsymbol{e}_{i_n} \triangleq \boldsymbol{y}_{i_n} - \hat{\boldsymbol{y}}_{i_n|\mathcal{A}_{n-1}} \tag{5}$$

where

$$\hat{\boldsymbol{y}}_{i_n|\mathcal{A}_{n-1}} = \boldsymbol{H}_{i_n}\boldsymbol{\theta} + \hat{\boldsymbol{v}}_{i_n|\mathcal{A}_{n-1}} \tag{6}$$

with $\hat{\boldsymbol{v}}_{i_n|\mathcal{A}_{n-1}}$ being the projection of $\boldsymbol{v}_i$ onto the affine space of previous measurements, denoted by $\mathcal{L}\{\boldsymbol{y}_{i_1}, \boldsymbol{y}_{i_2}, \ldots, \boldsymbol{y}_{i_{n-1}}\}$. The quantity $\boldsymbol{e}_{i_n}$ in (5) denotes the new information contained in sensor i_n and not in any of the previous measurements $\{\boldsymbol{y}_{i_1}, \boldsymbol{y}_{i_2}, \ldots, \boldsymbol{y}_{i_{n-1}}\}$.

Now note that $\hat{\boldsymbol{v}}_{i_n|\mathcal{A}_{n-1}}$ is given by

$$\hat{\boldsymbol{v}}_{i_n|\mathcal{A}_{n-1}} = \boldsymbol{B}_{i_n|\mathcal{A}_{n-1}} \boldsymbol{C}_{\mathcal{A}_{n-1}}^{-1} \boldsymbol{v}_{\mathcal{A}_{n-1}} \tag{7}$$

where

$$\boldsymbol{B}_{i_n|\mathcal{A}_{n-1}} = \begin{pmatrix} \boldsymbol{C}_{i_n i_1} & \boldsymbol{C}_{i_n i_2} & \ldots & \boldsymbol{C}_{i_n i_{n-1}} \end{pmatrix} \tag{8}$$

Combining (1), (5), (6), and(7) gives

$$\begin{aligned} \boldsymbol{e}_{i_n} &= \boldsymbol{v}_{i_n} - \hat{\boldsymbol{v}}_{i_n|\mathcal{A}_{n-1}} \\ &= \begin{pmatrix} -\boldsymbol{B}_{i_n|\mathcal{A}_{n-1}} \boldsymbol{C}_{\mathcal{A}_{n-1}}^{-1} & \boldsymbol{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{v}_{\mathcal{A}_{n-1}} \\ \boldsymbol{v}_{i_n} \end{pmatrix} \end{aligned} \tag{9}$$

It can be verified that $\boldsymbol{e}_{i_n} \perp \boldsymbol{v}_j$ for any $j \in \mathcal{A}_{n-1}$. Thus, the matrix

$$\boldsymbol{P}_{i_n|\mathcal{A}_{n-1}} \triangleq \begin{pmatrix} -\boldsymbol{B}_{i_n|\mathcal{A}_{n-1}} \boldsymbol{C}_{\mathcal{A}_{n-1}}^{-1} & \boldsymbol{I} \end{pmatrix} \tag{10}$$

projects $\boldsymbol{v}_{i_n}$ onto a space orthogonal to $\mathcal{L}\{\boldsymbol{v}_{i_1}, \boldsymbol{v}_{i_2}, \ldots, \boldsymbol{v}_{i_{n-1}}\}$. Given a sampling decision $\mathcal{A}_n$, the corresponding innovation process $\{\boldsymbol{e}_{i_k}\}_{k=1}^n$ has the important property

$$E(\boldsymbol{e}_{i_k}\boldsymbol{e}_{i_l}^T) = \begin{cases} \mathbf{0} & i_k \neq i_l \\ \boldsymbol{Q}_{i_k} & \text{otherwise} \end{cases} \tag{11}$$

where $\boldsymbol{Q}_{i_k}$ is the covariance matrix of $\boldsymbol{e}_{i_k}$ and has the form

$$\boldsymbol{Q}_{i_k} = \boldsymbol{P}_{i_k|\mathcal{A}_{k-1}} \begin{pmatrix} \boldsymbol{C}_{\mathcal{A}_{k-1}} & \boldsymbol{B}_{i_k|\mathcal{A}_{k-1}}^T \\ \boldsymbol{B}_{i_k|\mathcal{A}_{k-1}} & \boldsymbol{C}_{i_k i_k} \end{pmatrix} \boldsymbol{P}_{i_k|\mathcal{A}_{k-1}}^T \tag{12}$$

We can now introduce a transformed measurement of sensor i_n as

$$\boldsymbol{z}_{i_n} \triangleq \boldsymbol{P}_{i_n|\mathcal{A}_{n-1}} \begin{pmatrix} \boldsymbol{y}_{\mathcal{A}_{n-1}} \\ \boldsymbol{y}_{i_n} \end{pmatrix} \tag{13}$$

i.e.,

$$\boxed{\boldsymbol{z}_{i_n} = \boldsymbol{G}_{i_n|\mathcal{A}_{n-1}}\boldsymbol{\theta} + \boldsymbol{e}_{i_n}} \tag{14}$$

where

$$\boldsymbol{G}_{i_n|\mathcal{A}_{n-1}} = \boldsymbol{P}_{i_n|\mathcal{A}_{n-1}} \begin{pmatrix} \boldsymbol{H}_{\mathcal{A}_{n-1}} \\ \boldsymbol{H}_{i_n} \end{pmatrix} \tag{15}$$

Then, the m.v.u.e. of $\boldsymbol{\theta}$ given $\{\boldsymbol{z}_{i_1}, \boldsymbol{z}_{i_2}, \ldots, \boldsymbol{z}_{i_n}\}$ coincides with the m.v.u.e. of $\boldsymbol{\theta}$ given $\{\boldsymbol{y}_{i_1}, \boldsymbol{y}_{i_2}, \ldots, \boldsymbol{y}_{i_n}\}$. The advantage of working with the transformed quantities $\{\boldsymbol{z}_{i_k}\}_{k=1}^n$ is that the noises $\{\boldsymbol{e}_{i_k}\}_{k=1}^n$ in the model (13) are now uncorrelated. Thus, for a given set of observations $\mathcal{A}_n$, the MMSE is given by

$$\begin{aligned} \boldsymbol{D}(\mathcal{A}_n) &= \left(\boldsymbol{H}_{\mathcal{A}_n}^T \boldsymbol{C}_{\mathcal{A}_n}^{-1} \boldsymbol{H}_{\mathcal{A}_n}\right)^{-1} \\ &= \left(\sum_{k=1}^{n} \boldsymbol{G}_{i_k|\mathcal{A}_{k-1}}^T \boldsymbol{Q}_{i_k}^{-1} \boldsymbol{G}_{i_k|\mathcal{A}_{k-1}}\right)^{-1} \end{aligned} \tag{16}$$

3.3. Innovations Sensor Selection (ISS)

To meet the fidelity requirement with a minimum number of measurements, the algorithm chooses a group of sensors, each of which has the most new information with respect to others. Specifically, we define a *utility matrix* as

$$\boldsymbol{U}_{i_n|\mathcal{A}_{n-1}} \triangleq \boldsymbol{G}_{i_n|\mathcal{A}_{n-1}}^T \boldsymbol{Q}_{i_n}^{-1} \boldsymbol{G}_{i_n|\mathcal{A}_{n-1}} \tag{17}$$

which, as mentioned before, has the essential properties of an information measure. Now recall that the set of sensor observations $\mathcal{A}_{n-1}$ has an MMSE $D_{\mathcal{A}_{n-1}} = \mathrm{Tr}\left[\boldsymbol{D}(\mathcal{A}_{n-1})\right]$. According to (16), the MMSE with measurements $\{\mathcal{A}_{n-1}, i_n\}$ can be written as

$$D_{\mathcal{A}_n} = \mathrm{Tr}\left[\left(\boldsymbol{D}^{-1}(\mathcal{A}_{n-1}) + \boldsymbol{U}_{i_n|\mathcal{A}_{n-1}}\right)^{-1}\right] \tag{18}$$

Then, we can define a utility function as the difference between $D_{\mathcal{A}_{n-1}}$ and $D_{\mathcal{A}_n}$, i.e.,

$$\begin{aligned} u_{i_n|\mathcal{A}_{n-1}} &\triangleq D_{\mathcal{A}_{n-1}} - D_{\mathcal{A}_n} \\ &= \mathrm{Tr}\left[\boldsymbol{D}(\mathcal{A}_{n-1}) - \left(\boldsymbol{D}^{-1}(\mathcal{A}_{n-1}) + \boldsymbol{U}_{i_n|\mathcal{A}_{n-1}}\right)^{-1}\right] \end{aligned}$$

The function $u_{i_n|\mathcal{A}_{n-1}}$ serves as a measure of the new information provided by sensor i_n's observation.

At each step n, the fusion center then selects a sensor that has the maximum utility function, i.e.,

$$i_n = \arg \max_{i \notin \mathcal{A}_{n-1}} u_{i|\mathcal{A}_{n-1}} \tag{19}$$

The algorithm successively adds a sensor that has the most new information into the set of selected sensors until the desired estimation fidelity D_0 is achieved.

3.4. Fairness Considerations

The sampling scheme should avoid the hot spot problem and evenly distribute the energy load among the nodes, so that there are no overly-used nodes that will run out of energy before the others. To achieve this goal, we propose to incorporate the residual energy at each sensor in making the sampling decision.

Denote by $p_{i_n}(t)$ the amount of energy required for sensor i_n to participate in estimation for the t-th cycle. We can define a net utility function for sensor i_n as its utility minus energy cost, i.e.,

$$f_{i_n|\mathcal{A}_{n-1}} = u_{i_n|\mathcal{A}_{n-1}} - \gamma_{i_n}(t) p_{i_n}(t) \tag{20}$$

where $\gamma_{i_n}(t)$ is the price coefficient; it can be adapted to the current residual sensor energy as

$$\gamma_{i_n}(t) = \frac{\beta}{E_{i_n}(t)} \tag{21}$$

where β is a constant and $E_{i_n}(t)$ represents the residual energy of sensor i_n. The form (21) implies that if sensor i_n has a large residual energy level, it has a low price; on the other hand, when the sensor has little energy left, selecting this sensor will result in a high cost. Instead of using the criterion in (19), we now use the net utility function $f_{i_n|\mathcal{A}_{n-1}}$.

Suppose there is a seed sensor initially, i.e., $\mathcal{A}_1 = \{i_1\}$. At each step n, the algorithm chooses a sensor that has the maximum net utility given the previous selected sensor observations $\mathcal{A}_{n-1}$, i.e.,

$$i_n = \arg \max_{i \notin \mathcal{A}_{n-1}} f_{i|\mathcal{A}_{n-1}} \tag{22}$$

The algorithm successively adds the sensor measurement into the set of selected sensors until the mean-squared error is less than or equal to the desired estimation fidelity level, i.e., $D_{\mathcal{A}_n} \leq D_0$.

4. SIMULATIONS

In this section, we experimentally evaluate the effectiveness of the proposed sampling scheme through simulations. We randomly generate $N = 100$ sensors within a unit square $[0, 1]$ by $[0, 1]$ with the fusion center located at $\left(\frac{1}{2}, \frac{1}{2}\right)$. Consider a simple linear model $y_i = \theta + v_i$. The covariance matrix

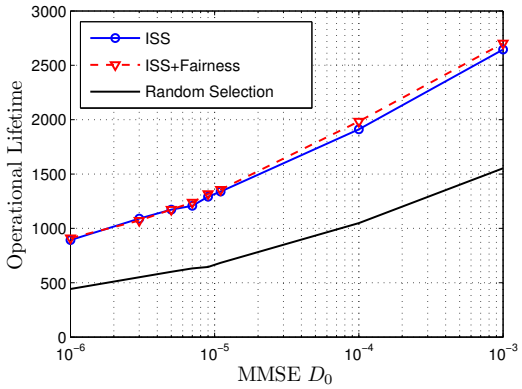

Fig. 1. Average active sensors vs. estimation fidelity.

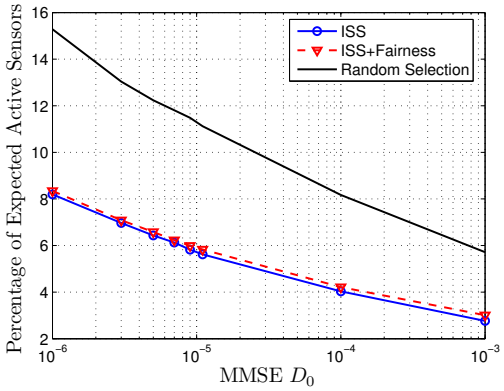

Fig. 2. Average active sensors vs. estimation fidelity.

C is randomly generated according to the spatial correlation model

$$C_{ij} = \begin{cases} \sigma_i^2 & i = j \\ \sigma_i \sigma_j \exp\left(-5\, d_{ij}^2\right) & i \neq j \end{cases} \qquad (23)$$

where $\sigma_i^2 = 0.01 + 0.99\chi^2(1)$, with $\chi^2(1)$ generated by the Chi-square distribution of degree 1 and the correlation is an exponential function of the distance between nodes. Suppose that each sensor initially has a unit of energy, and it consumes $p_i = 0.001 + d_i^{3.5}$ unit of energy if it participates in estimation; otherwise, it neither measures the signal nor transmits any information. The network keeps operating until the sampling algorithm cannot produce any feasible solution to meet the fidelity requirement D_0 due to insufficient energy in the sensors. Figures 1 and 2 respectively illustrate the operational lifetime and expected active sensors in each cycle with respect to D_0. The results show that the proposed schemes outperform a random selection algorithm, in which sensors are selected at random. In Fig. 3, it can be seen that the energy variance of the fair algorithm is considerably smaller than ISS due to the even distribution of energy load.

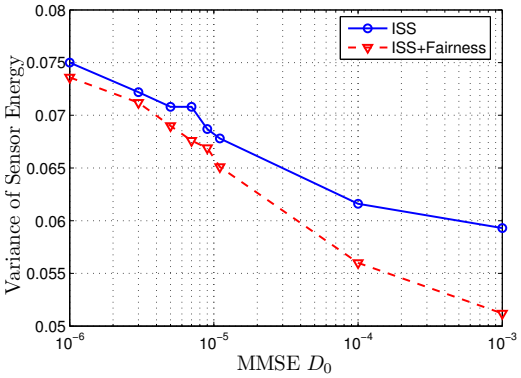

Fig. 3. The variance of residual energy at the sensors with $\beta = 100$.

5. CONCLUSION

In this paper, we proposed a framework for energy efficient sampling for estimating a field variable in sensor networks. The framework suggests that energy-efficiency can be achieved through reducing the number of active sensors. Furthermore, the energy load can be evenly distributed among all sensors in order to achieve fairness. The performance of the sampling scheme is evaluated through simulation. Some interesting extensions are worth further investigation for important practical issues such as finite-bit communication, transmission errors, distributed implementation, and reliable communication protocols.

6. REFERENCES

[1] I. F. Akyildiz, W. Su, Y. Sankarsubramaniam, and E. Cayirci, "Wireless sensor networks: A survey," *Computer Networks*, vol. 38, pp. 393–422, Mar. 2002.

[2] E. Ertin, J.W. Fisher, and L.C. Potter, "Maximum mutual information principle for dynamic sensor query problems," in *Proc. IPSN*, Palo Alto, CA, Apr. 2003, pp. 405–416.

[3] H. Wang, G. Pottie, K. Yao, and D. Estrin, "Entropy-based sensor selection heuristic for target localization," in *Proc. IPSN*, Berkeley, CA, Apr. 2004, pp. 36–45.

[4] V. Isler and R. Bajcsy, "The sensor selection problem for bounded uncertainty sensing models," in *Proc. IPSN*, Los Angeles, CA, Apr. 2005, pp. 151–158.

[5] A. H. Sayed, *Fundamentals of Adaptive Filtering*, John Wiley & Sons, NJ, 2003.

[6] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*, Prentice-Hall, Upper Saddle River, NJ, 2000.