Data Estimation in Sensor Networks Using Physical and Statistical Methodologies

Yingshu Li   Chunyu Ai   Wiwek P. Deshmukh   Yiwei Wu
Department of Computer Science
Georgia State University, Atlanta, GA 30303
{yli, chunyuai, wdeshmukh, wyw}@cs.gsu.edu

Abstract

Wireless Sensor Networks (WSNs) are employed in many applications in order to collect data. One key challenge is to minimize energy consumption to prolong network lifetime. A scheme of making some nodes asleep and estimating their values according to the other active nodes' readings has been proved energy-efficient. For the purpose of improving the precision of estimation, we propose two powerful estimation models, Data Estimation using Physical Model (DEPM) and Data Estimation using Statistical Model (DESM). DEPM estimates the values of sleeping nodes by the physical characteristics of sensed attributes, while DESM estimates the values through the spatial and temporal correlations of the nodes. Experimental results on real sensor networks show that the proposed techniques provide accurate estimations and conserve energy efficiently.

1 Introduction

To date, data acquisition using Wireless Sensor Networks (WSNs) has became an essential ingredient of current technologies. Due to the energy limitations faced by sensor nodes, most techniques focus on *aggregating in-network data* so as to improve energy efficiency. Although this is a good way of conserving energy, it may be necessary that all sensor readings are collected at a base station. In such a scenario, *in-network aggregation* cannot be permitted as it reduces the data-set communicated to the base station. Consequently, in such situations, *data-estimation* can prove to be a very fruitful alternative to *data aggregation*. An example is the determination of light intensity distribution in a region, such as photosensitive horticulture under artificial illumination [11]. Commercial gadgets are now available [4] for such purposes, the HortiSpec [4] was developed to measure light intensity and spectral distribution in the visible and NIR range inside greenhouses. The light intensity can be measured by these sensors, and data estimation techniques employed in such environments would help conserve energy and prolong network lifetime. Commercial light sensors which use photodiodes that produce a voltage proportional to the light intensity are now available. Similarly, light intensity sensors are required to monitor maintenance of optimum lighting in animal husbandry related business [1], and also in cell-culture experiments [10] under artificial light, where also data estimation techniques would help enhance the longevity of WSNs.

Usually, users can accept approximate data. As a result, a method of data estimation is a perfect option for conserving energy. In fact, the key challenge is how to provide estimated data with high precision while consuming as little energy as possible. To address this problem, two novel data estimation methods are proposed in this paper. In our scheme, a minimal number of nodes are set to be active and the other nodes are set to sleep. All nodes serve as active working nodes by turns. The base station estimates values of sleeping nodes based on two estimation models proposed in this paper, Data Estimation using Physical Model (DEPM) and Data Estimation using Statistical Model (DESM). The scheme not only prolongs network lifetime, but also fulfills users' expectations of data precision. Notably, the DEPM model is the first one to take advantages of *physical characteristics* of monitored attributes for estimation. Experiments on a real sensor network consisting of twenty TelosB nodes [3] were conducted to evaluate the proposed models. The experimental results show that the proposed methods are energy efficient.

The rest of this paper is organized as follows. In Section 2, an overview of related works is presented. In Section 3, a formal definition of the problem is provided. In Section 4, we introduce the two estimation models, DEPM and DESM, in detail. Section 5 shows our experimental results, and Section 6 concludes this paper.

2 Related Work

Common techniques to save energy in WSNs employ various alternative strategies to minimize data acquisition and/or transmission by using a variety of approaches. In [5], Chu *et al.* have proposed a mechanism using conditional data transmission to conserve energy by seeking data collection *only* if the information sought is beyond certain bounds that are predicted using statistical models. Jain *et al.* have built dynamic procedures that employ maximum filtering of data using a technique called stochastic recursive data filtering, to conserve resources subject to meeting precision standards [9]. Primarily, these methods are aimed at reducing energy consumption by substituting data acquisition using data estimation. Another method of data estimation is based on collection of data samples for a relatively long time and calculating the autocorrelation of the vector of samples [7]. This approach aims at enabling nodes to identify patterns in the behavior of sensed processes and report only uncommon observable data to conserve energy.

Dash *et al.* have studied using physical laws for data estimation, approximate values of land surface temperature and emissivity from passive sensor data [6]. Essentially, this work makes use of well-known laws of Physics to estimate the Land Surface Temperature realistically. While physical laws have been used in some of technologies mentioned above to govern sensor mechanisms, in the DEPM technique proposed in this work it seems for the first time that physical laws are employed in a direct energy saving strategy aimed at data estimation. Most other methods rely on some forms of temporal and/or spatial correlations and/or data filters for data estimation whereas in the present work well-established physical laws that are known to be correct with extremely high accuracy are used directly in the network plan to achieve minimization of energy consumption which is the primary challenge in WSN technology. The data estimation scheme we propose can conserve more energy than the methods mentioned above obviously, because in the methods, such as [5, 7, 9], all nodes should be always in working status; however, in our scheme nodes serve as working nodes by turns, and others go to sleep, so our scheme can prolong network lifetime significantly.

3 Problem Definition and Working Scenario

In this section, we formally define the SEAQA problem, and describe the working scenario of our estimation scheme.

In many applications, it is not necessary for users to obtain completely accurate values. A scheme that provides approximate answers might offer an opportunity to prolong network lifetime efficiently. We formally define this problem as the Sensor Energy-efficient Approximate Query Answer (SEAQA) problem.

Definition 1 (SEAQA Problem): Given a three dimensional area *A* and a set of *N* sensors *S* = $\{s_1, s_2, ..., s_N\}$, derive a working scheme *ws* for *S* such that:

1. For each sensor s_i, its returned value V_e deviates its real sensing value V as little as possible, that is, $|V_e - V|$ is minimized.

2. The energy consumptions among all the sensors are balanced.

3. The network lifetime is maximized.

The lifetime of a WSN depends on the node with the minimal remaining energy, so maintaining energy consumption balance among sensors is also a significant problem to prolong network lifetime.

To conserve energy, some nodes in a network can be in sleep mode, and only a subset of the nodes are responsible for sensing and communication. Nodes work in *rounds* as shown in Figure 1. There are three phases in each round. In the first phase, the base station gathers the remained energy information of all the nodes and selects a subset of nodes with much more remained energy to serve as *active working* nodes. In the second phase, the base station informs each node its role and working duration. Then, in the rest of this round, the selected working nodes are in charge of sensing and communication, and the other nodes go to sleep in order to conserve energy. The sensing values of the sleeping nodes are estimated by the proposed DEPM and DESM methods at the base station.

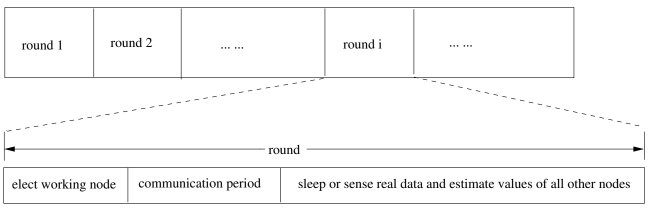

Figure 1. Working rounds.

4 Estimation Models

To address the data estimation problem, we propose two efficient methods DEPM and DESM. DEPM estimates data using physical laws and DESM accomplishes data estimation using a statistical model.

4.1 Data Estimation using Physical Model (DEPM)

In this section, we introduce DEPM which employs basic laws of Physics. This is achieved by exploiting the principle of superposition of the physical sensed parameters. Essentially, the problem is to determine a physical quantity at a *field point* when its value is the result of linear superposition produced by *multiple sources*. DEPM does not need to deploy any more sensors than a constant number, that is the number of sources, and all the *extra* sensors go to sleep. The sensing values of sleeping nodes can be predicted by DEPM model according to sensing values of the active nodes. By activating one of the sleeping nodes, one can carry out an actual sensory measurement at that location and verify the prediction of the DEPM model. After such verification, DEPM enables. As long as source properties remain invariant, the prediction of the physical quantities at an infinite number of locations within the region R and makes it redundant to deploy sensors at these new locations. DEPM thus achieves enormous energy conservation by exploiting the principle of linear superposition and solving a set of algebraic linear inhomogenous equations. DEPM provides accurate physical estimates of the physical observable at an infinite number of field points in the region without having to consume energy in activating additional sensors.

To define and illustrate the functioning of DEPM, we use light intensity as an example monitored attribute. We consider a total of N nodes and M sources of light, where $M \ll N$. Each of the light sensors registers an amount of light intensity I_k at the node S_k ($k \in [1, N]$) in the system (Figure 2). The base station needs to determine the *power radiated by each source of light* from the measurements of light intensities measured by nodes.

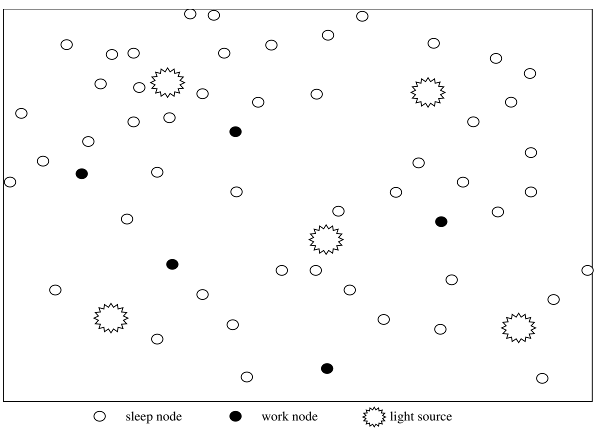

Figure 2. A light intensity monitoring WSN.

DEPM operates in two modes: *dynamic* and *static*. The dynamic mode of DEPM is employed when the power radiated by the M sources is not known and needs to be measured by active sensors. DEPM provides energy conservation by accurately estimating the power of the M sources by employing only M active sources. Toward this goal, DEPM exploits the principle of linear superposition and provides solutions by addressing a set of linear algebraic inhomogeneous equations. Energy minimization is achieved by requiring only a constant minimal number M out of the available N nodes to be activated to determine the power radiated by the M sources. These constant number of active nodes are shown as dark nodes in Figure 2. In each round, the DEPM algorithm solves the set of linear inhomogeneous equations, thus providing the accurate reliable values of the *light power P radiated by the* M light sources.

The static mode of DEPM is employed when (a) the power radiated by the sources is first determined by carrying out sensor measurements in the dynamic mode, and (b) when it is known that the power radiated by the light sources is time-invariant. In the static problem, no node at all is required to be activated, except to verify the DEPM prediction, even as the DEPM algorithm solves the inverse problem and estimates with complete accuracy the values of light intensity distribution at an infinite set of point locations in the region R. No energy is consumed in the activation of any node at all, both data acquisition energy and data transmission energy are thus fully conserved while the DEPM Static Algorithm provides light intensity solutions at an infinite number of field points as shown below.

DEPM DYNAMIC Algorithm

In each execution round, three tasks are processed:

1. At the beginning, M out of N nodes are elected to work as active nodes by using a random algorithm which ensures that energy is conserved in all of the N nodes optimally.
2. The *light intensity* I_k ($k \in [1, M]$) sensed by the M active nodes are sent to the base station.
3. The base station computes the values of *Power* P_l *radiated by each of the* l^{th} light source using the DEPM dynamic algorithm explained below.

Table 1 lists the symbols used by DEPM.

If the j^{th} light source alone is switched on, the rest of the $(M-1)$ light sources being switched off, then the *light intensity* I_k measured by the k^{th} active sensor is related to the power P_j of the j^{th} light source by the well-known *inverse square law*:

$$I_k = \frac{P_j}{4\pi d^2(l_j, s_k)}. \tag{1}$$

Table 1. Symbol Table

Symbol	Meaning
N	Number of sensor nodes.
M	Number of light sources.
P_i	Power radiated by the light source l_i.
$\{s_1, s_2, ..., s_M\}$	Set of active nodes.
I_i	Value of light intensity measured by an active node s_i.
$D(l_i, s_j)$	Distance between light source l_i and location j where the active sensor s_j is located.

Now, if all the M light sources are switched on, one requires a *minimum* of M nodes to uniquely determine the light powers of each of the M light sources. DEPM requires only a constant number M of nodes to be activated for this purpose and permits all of the remaining nodes to sleep, thereby conserving their energy. Thus, at the beginning of each round, DEPM elects M nodes as active working nodes. For the sake of balancing the energy consumptions, the active nodes are always selected randomly from the subset of sensors that have residual energy that is higher than the average *energy per sensor* in the entire network. Then, the selected active nodes sense the light intensity I_k of light at each k^{th} node and communicate the recorded values to the base station.

Since the *light intensity* I_k at the k^{th} node obeys the principle of superposition with respect to light reaching that node from each of the M number of light sources, we have:

$$I_k = \frac{P_1}{4\pi d^2(l_1, s_k)} + \frac{P_2}{4\pi d^2(l_2, s_k)} + ... + \frac{P_k}{4\pi d^2(l_k, s_k)} \quad (2)$$

$$I_k = \sum_{j=1}^{M} \frac{P_j}{4\pi d^2(l_j, s_k)} = \sum_{j=1}^{M} a_{jk}\ P_j,\ a_{jk} = \frac{1}{4\pi d^2(l_j, s_k)} \quad (3)$$

Equation 3 represents a family of linear algebraic equations in which the coefficients a_{jk} are known from the geometrical arrangements of the WSN and the locations of the M light sources. The inhomogeneous linear equations can be solved using well-known techniques in [8]. These equations can be written in a matrix form:

$$\alpha\ \pi = \iota,$$

where $\alpha = \begin{bmatrix} a_{1,2} & a_{1,2} & ..... & a_{1,l} & ..... & a_{1,M} \\ ..... & ..... & ..... & ..... & ..... & ..... \\ a_{l,1} & a_{l,2} & ..... & a_{l,l} & ..... & a_{l,M} \\ ..... & ..... & ..... & ..... & ..... & ..... \\ a_{M,1} & a_{M,2} & ..... & a_{M,l} & ..... & a_{M,M} \end{bmatrix}$,
$\pi = [P_1, P_2, ..., P_l, ..., P_M]^T$, $\iota = [I_1, I_2, ..., I_l, ..., I_M]^T$.

We note that the matrix α is determined entirely by geometrical arrangement and is completely independent of the light sources and the sensor properties.

The solution to this system of equations is: $\pi = \alpha^{-1}\iota$, *i.e.* $P_l = \sum_{j=1}^{M} .\alpha_{l,j}^{-1}\ \iota_j$, . where α^{-1} is the inverse of the matrix α. A condition is easily incorporated in the algorithm that when the determinant of the matrix α, $|\alpha|$, is small, a different set of active nodes must be chosen such that $|\alpha| \succ \delta$, where δ is a user-defined small quantity.

Thus, from the values of the light intensities sensed by the set of M active nodes, the DEPM dynamic algorithm determines the power radiated by each of the M light sources. So far, we can use Equation 3 to calculate the light intensity of each sleeping node easily.

4.2 Data Estimation using Statistical Model (DESM)

Another estimation mode, DESM, is introduced in this section. A WSN is logically divided into cells using a grid (as shown in Figure 3) of which each cell is called an *observing region*. In each observing region and in each round, only the node that has the maximal remained energy is chosen to be the active working node.

Usually, the collected data has strong temporal and spatial locality property. The temporal locality means that values sensed by the same sensor over a continuous time domain have strong relationships. Spatial locality means that values sensed by the sensors whose locations are nearby often are also similar. For instance, HortiSpec mentioned above shows the strong temporal and spatial locality. Aware of these two locality properties, DESM studies the temporal relationship of a sleeping node *s* from its historical data set and the spatial relationship between *s* and the current working node that is in the same grid with *s*, and gives an estimation value for *s*.

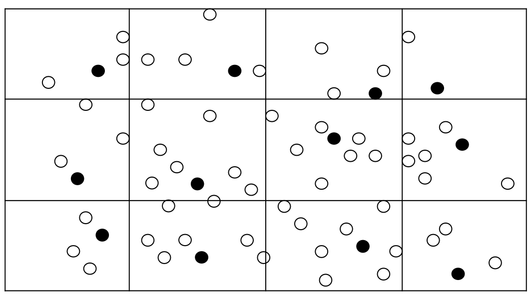

Figure 3. A Grid.

DESM needs a real historical sensing data set of each node to start. In the first several rounds, all the nodes are in working mode. The sensed values are collected and stored at the base station as a historical data set. Thereafter, only one node in each cell is active and the other nodes in the same cell go to sleep. We use the following method to estimate the sensing values of the sleeping nodes.

Formally, let S_R = $\{S_1,\ S_2,\ \ldots,\ S_n\}$ be the nodes in an observing region *R*. For each node $S_i \in S_R$, $X_{i1}, X_{i2}, ..., X_{im}$ represent its corresponding historical value sequence. The historical value sequences of the node set S_R are stored in a matrix A_R. Suppose that S_i is the active node in S_R. Each time S_i observes a new value $X_{i(m+1)}$, the values of the other $n-1$ nodes can be estimated according to $X_{i(m+1)}$ and A_R.

$$A_R = \begin{bmatrix} X_{11} & X_{12} & ... & X_{1m} \\ X_{21} & X_{22} & ... & X_{2m} \\ ... & ... & ... & ... \\ X_{n1} & X_{n2} & ... & X_{nm} \end{bmatrix}$$

Being aware of the locality of nodes, we can use the following equation to estimate the value $\hat{X}_{j(m+1)}$ of S_j.

$$\hat{X}_{j(m+1)} = (1-\alpha)\hat{Y} + (\alpha)\hat{Z} \qquad (4)$$

where $\hat{Y}$ is the last estimated value $\hat{X}_{j(m)}$, which measures the influence of the historical sensing data on the current value at node S_j. $\hat{Z}$ is computed by $X_{i(m+1)}$, which measures the influence of the data sensed by the active node S_i on the data of node S_j. α is a weight parameter that evaluates the effects of $\hat{Y}$ and $\hat{Z}$ on the estimated value $\hat{X}_{j(m+1)}$, whose value is in $[0, 1]$.

We use the following method to compute $\hat{Z}$. Given two random variables $X_i = \{X_{i1}, X_{i2}, \ldots, X_{im}\}$ and $X_j = \{X_{j1}, X_{j2}, \ldots, X_{jm}\}$ with $X_{i(m+1)}$, the estimated value $\hat{Z}$ for $X_{j(m+1)}$ is computed as follows:

$$\hat{Z} = X_{j(m)}(1 + \frac{X_{i(m+1)} - X_{i(m)}}{X_{i(m)}}) \qquad (5)$$

Equation 5 is based on an assumption that X_i and X_j have the similar data fluctuation trend, which is usually true for the two nodes in the same cell according to the space locality of a WSN.

Intuitively, if X_i is more related with X_j, then $\hat{Z}$ is expected to impact more influence on the value of $\hat{X}_{j(m+1)}$, that is, α should be a larger value. According to the above analysis, correlation coefficient $\varphi(X_i, X_j)$ is an exact way to define α.

Given two random variables X_i and X_j, whose first *m* values are $X_{i1}, \ldots, X_{im}$ and $X_{j1}, \ldots, X_{jm}$ respectively. We can use the following formula to compute the correlation coefficient between X_i and X_j:

$$\varphi(X_i, X_j) = \frac{Cov(X_i, X_j)}{\sigma_{X_i}\sigma_{X_j}} \qquad (6)$$

In Equation 3, $Cov(X_i, X_j)$ is covariance between X_i and X_j. $Cov(X_i, X_j) = E[(X_i - EX_i)(X_j - EX_j)] = E(X_iX_j) - E(X_i)E(X_j)$.

4.3 Discussion

The Dynamic DEPM technique offers by itself as a very powerful algorithm to solve the inverse problem *when it is known that the power radiated by each of the light sources is not changing*, as is often the case, and it is nevertheless of importance to know the distribution of intensity of light at a large number of point-locations in the region illuminated by the M light sources. Such problems are of interest in very many different areas, such as in the determination of light intensity distribution in a sports arena where games are played under artificial light sources of constant power, and in horticulture experiments [4] in which cultivation of some vegetation is sought under artificial light from a set of constant-power sources. Likewise, cell-culture experiments [10] under controlled light intensity environment is another example of a situation where the Static DEPM model introduced below can be extremely valuable.

To solve such problems, the DEPM *inverse* algorithm is to be used, treating the matrix π as known and determining the intensity matrix ι for any set of M point-locations whose coordinates alone determine the corresponding matrix α. Since in this model the matrix π is considered to be known (*i.e.* pre-determined), the inverse algorithm is referred to as STATIC DEPM. Thus for a predetermined matrix π, and for a set of arbitrary M number of locations whose coordinates alone determine the required matrix, the STATIC DEPM algorithm solves the matrix relation: $\alpha\pi$ = ι, thereby giving the values of the light intensities at a set of M arbitrary locations whose coordinates alone need to be provided as input for the STATIC DEPM algorithm. Not a single node needs to be activated at this set of M locations, thus providing reliable data estimation that substitutes data acquisition conserving energy of the WSN. Of course, some of the sensors can be activated to verify the prediction of the static DEPM algorithm, and this is a tremendous advantage that lends itself as a tool to check reliability of the model.

Light intensity distribution at arbitrary locations (x_k, y_k, z_k) can be achieved by data estimation using a method completely based on physical laws instead of using actual nodes being activated at these locations. The physical laws that are employed are very rigorous. Hence, as long as the physical conditions of the algorithm are satisfied, the DEPM is expected to provide very accurate predictions. To the best of our knowledge, most of the energy saving techniques employed in WSNs employ various combinations of spatial and temporal coherence and/or data filters, whereas the present method employs well-established laws of physics as direct energy conservation strategy and thus provides a very novel approach to develop energy saving mechanisms in WSNs based on physical laws. Based on the temporal and spatial correlation among adjacent nodes, DESM method can estimate the intensity of light, temperature, humidity and *etc*. We merely need to know the position information of sensor nodes, and do not require any information about the monitored objects.

How to set up a grid is a primary factor in the accuracy of estimation. In order to conserve energy as much as possible, it is desirable that the number of nodes in sleep mode is maximized. However, the accuracy of DESM would decrease in such a case. There is thus a trade-off between the estimation accuracy and energy conservation. In some applications, using clusters to separate the observation regions is more preferred. For instance, in order to monitor the temperature of a big building, in each room must be placed several sensors. Consequently, a room can be considered as an observing region, and it is no doubt that the trend of temperature fluctuation in a room is rather similar.

5 Experiment Results

In the prototype experiments that were performed, twenty TelosB [3] sensors were deployed in a laboratory of size 25'x50'. The laboratory was insulated from natural light so that the sensors would sense and record only light intensities from two artificial electrical light sources L_1 and L_2. The twenty sensors and the two light sources were geometrically distributed randomly in the room, such that there is no obstacle between sensors and the light sources.

Each sensor sensed 10 values of light intensity every 2 seconds, and sent the average value to the base station. The data that were received by the base station were stored as the tested data set. Three different experimental sets were employed to test the methodology developed. In the first set, intensity data recorded by the sensors were analyzed with both the light sources turned on. In the second data set, intensity data were collected from the sensors and the experiment was performed while the two light sources were turned on and off every 5 minutes. In the third set, data were obtained with the two light sources turned on and off randomly and frequently. Detailed experiments settings can be found at [2].

We have implemented our two methods on TelosB motes and have conducted extensive experiments. Our experiments are designed with two objectives in mind. First, we verify the proposed methods, DEPM and DESM, are able to report highly accurate answers. Second, we assess the two methods DEPM and DESM for their energy consumption.

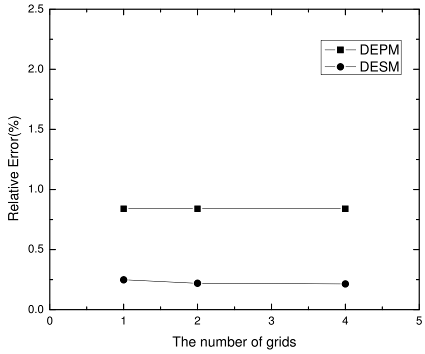

Figure 4. The relative error of estimations reported by DEPM and DESM using data set 1.

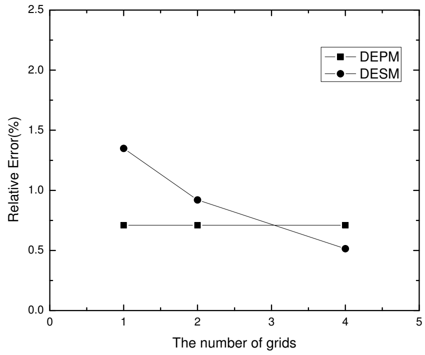

Figure 5. The relative error of estimations reported by DEPM and DESM using data set 2.

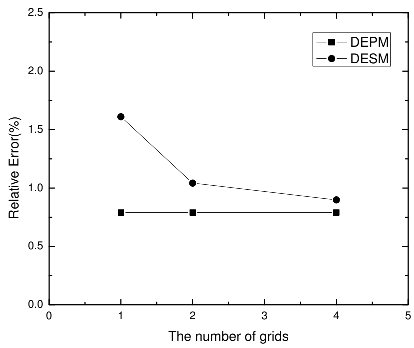

Figure 6. The relative error of estimations reported by DEPM and DESM using data set 3.

5.1 Estimation accuracy

In this section, we report the results of DEPM and DESM and assess the accuracy of the data estimates reported by the proposed methods. The average relative error is used as the metric to measure the accuracy. We conducted the experiments with the three data sets that mentioned above.

As is shown in Figures 4, 5, and 6, with the number of grids increases, the relative error of answers reported by DESM goes down. This is because the more the number of grids is, the less the number of sensors in each grid, while the stronger the space locality of sensors in one grid has. We could also see that DESM is more accurate for data set 1 but gets worse for the data sets 2 and 3. The reason is that the light intensity values of data set 1 are stable while vary in the data set 2 and 3.

The accuracy of DEPM method has no relationship with the number of grids as shown in Figures 4, 5 and 6. DEPM outperforms DESM on data set 2 and 3, but performs worse on data set 1. Furthermore, obviously DEPM is not affected by different data sets. It can keep good and stable quality on all three data sets.

5.2 Energy consumption

Table 2. System Parameters and Setting.

Parameter	Setting
Number of sensor nodes	20
Message size	8 bytes
Transmission distance	50m
Energy Cost for Sending a Message	19.2uJ
Energy Cost for Receiving a Message	3.2uJ
Energy Cost for Sensing a Light Intensity	100nJ
Energy Cost in Sleeping Mode	0.016mW
Initial Energy Budget at Each Sensor Node	1J

In this section, we assess the energy consumption. We use the energy model in [12] as presented in Table 2. For the DESM and DEPM methods, one execution round is set to 10 minutes. For DESM, all the sensors collect data in the first two rounds of every two hours considered as the history data. After the history data collecting phase, in each round one node with the most remained energy works as an active node in every grid, and others go to sleep in the rest time. For DEPM, in each round two nodes with the most remained energy work as active nodes, and others go to sleep.

We compare the proposed models with a naive model where all the sensors are set active. As shown in Figure 7, the average energy consumption of both DEPM and DESM is much less than that of the naive model. After running about 10 rounds, the accumulative average energy consumption of DEPM is about 50% of that of the naive model. So DESM and DEPM can prolong network lifetime effectively. DEPM is more effective than DESM, because DESM has the collecting history data phase, consuming more energy. DESM-1 (DESM-2, DESM-4) represents the DESM method with one grid (2 grids, 4 grids correspondingly) in the WSN. From Figure 7, we can see that the more grids for DESM, the more energy consumption. The reason is that each grid has an active sensor in DESM. The more number of grids means the more sensors are set to active in the same time, hence the amount of energy consumed is more.

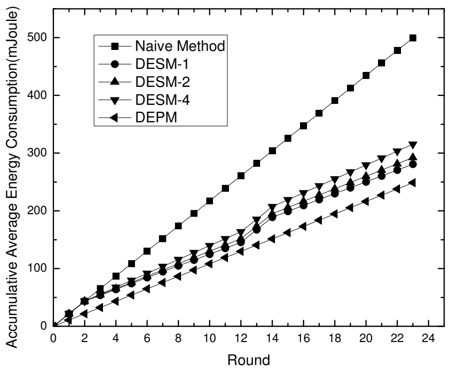

Figure 7. The energy consumptions of different models.

We show the remained energy of DEPM with different constraints in Figure 8. It is defined as the coefficient of variation of all the nodes' remained energy. The average energy consumption of all the nodes is almost half of initial energy. When we use larger δ, the estimation accuracy is better, but it makes some nodes impossible to be active, so the balance of energy consumptions is worse.

Neither DESM nor DEPM is a perfect method in all situations. Table 3 provides a guide for choosing the proper estimation model in different situations. For instance, if users prefer high accuracy, it is better to choose DESM or DEPM with large δ. If users expect long network lifetime, DEPM is a good choice.

6 Conclusions

To conserve energy in WSNs, we propose a scheme where only a few nodes are chosen to work, and the other nodes are set to sleep. The data sensed by sleeping nodes are estimated by two efficient methods, DESM and DEPM.

Table 3. Model Selection Guideline.

Which model should be used?	DESM	DEPM (large δ)	DEPM (small δ)
Lowest total energy consumption	No	Yes	Yes
Accuracy	Yes	Yes	No
Accuracy and balanced energy consumptions	Yes	No	No
Sensing variable light intensity sources	No	Yes	Yes
Accuracy at the cost of balanced energy consumptions	N/A	Yes	No
Balanced energy consumptions at the cost of accuracy	N/A	No	Yes

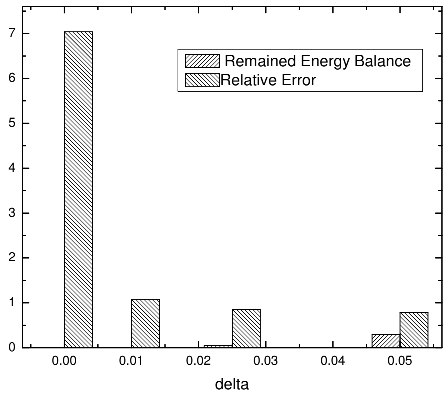

Figure 8. The remained energy of DEPM.

DESM is a statistical method that takes advantages of the time and space locality of a WSN. DEPM technique utilizes the physical laws to estimate data. Since the principle of superposition on which DEPM is based on is a very common principle in physics and is obeyed by a large number of physical quantities, the presented techniques have a very wide applicability. Furthermore, DEPM model enables us to deploy a rather small constant number of sensors out of a large number of available sensors, thereby conserving much energy. Experimental results show that the proposed methods are efficient and effective.

References

[1] Commonwealth of massachusetts. proposed regulations and public hearing notice. Available: http://www.mass.gov/agr/animalhealth/petshops/. [Accessed November 03, 2007].

[2] Computer science of georgia state university. experimental demonstration of the physical methodology. Available: http://www.cs.gsu.edu/yli/datasets/sensor.html. [Accessed November 03, 2007].

[3] Crossbow wireless sensor networks. telosb. Available: http://www.xbow.com/Products/productdetails.aspx?sid=252. [Accessed November 03, 2007].

[4] Horticulture measurements. [online]. Available: http://www.avantes.com/Applications/HortiSpec.htm. [Accessed November 03, 2007].

[5] D. Chu, A. Deshpande, J. M. Hellerstein, and W. Hong. Approximate data collection in sensor networks using probabilistic models. In *ICDE*, page 48, 2006.

[6] P. Dash. Land surface temperature and emissivity estimation from passive sensor data: Theory and practice–current trends. *International Journal of Remote Sensing*, 23:2563–2594, 2002.

[7] D. de O. Cunha, R. P. Laufer, I. M. Moraes, M. D. Bicudo, P. B. Velloso, and O. C. M. B. Duarte. Bio-inspired field estimation with wireless sensor networks. Technical report, Universidade Federal do Rio de Janeiro. http://www.gta.ufrj.br/ftp/gta/TechReports/CLMB05a.pdf, accessed November 03, 2007.

[8] G.B.Arfken and H.J.Weber. *Mathematical Methods for Physicists*. Academic Press, 2003.

[9] A. Jain, E. Y. Chang, and Y. Wang. Adaptive stream resource management using kalman filters. In *SIGMOD Conference*, pages 11–22, 2004.

[10] G. Richter. Blue light control of the level of two plastid mrnas in cultured plant cells. *Plant Molecular Biology*, 3:271–276, 1984.

[11] E. Runckle and R. Heins. Manipulating the light environment to control flowering and morphogenesis or herbaceous plants. In *International Symposium on Artificial Lighting in Horticulture*, June 2006.

[12] X.Tang and J.Xu. Extending network lifetime for precision constrained data aggregation in wireless sensor networks. In *Proc. IEEE INFOCOM*, Apr. 2006.