

Subsurface Plume Tracking Using Sparse Wireless Sensor Networks

Vidarshana Bandara, Anura P. Jayasumana* & Ali Pezeshki

Colorado State University, Fort Collins, Colorado, USA

Email: Anura.Jayasumana@Colostate.edu

Tissa H. Illangasekare & Kevin Barnhart

Colorado School of Mines, Golden, Colorado, USA

ABSTRACT: Deployment of sensors for tracking chemical plumes such as those in the subsurface can be quite expensive. A compressed data gathering scheme for chemical plume tracking is presented. With only a fraction (about 25%) of the measurement points required to achieve a given spatial resolution this novel application of compressed sensing can track the plume within 7% accuracy compared to the case of a full sensor array. The scheme also gathers and re-distributes information of the sensors to the entire network, compressing with Discrete Wavelet Transform. The scheme can be used to disseminate global tracking information to sensors as well, with savings in communications by a factor of 5 in average, if such capability is required. The scheme is capable of interpolating randomly missing sensor points with significant accuracy. It also supports data fusion as a simple addition of coefficients requiring no changes to the message length.

1 INTRODUCTION

In a variety of situations in environmental science and engineering, some of them related to national security, it is necessary to track and monitor chemical plumes, to make predictions on their future behaviors, and to evaluate potential risk to humans and ecological environments. While it is desirable to have data on chemical concentrations collected at high spatial and temporal resolutions to facilitate reliable predictions, the cost and other logistical factors associated with installing sampling wells limit the monitoring accuracy and the resolution achievable. Specific situations dealing with tracking dissolved contaminant plumes in flowing groundwater require the collection of water samples from sparsely distributed monitoring wells. With current technology, these samples are delivered from field sites to testing laboratories to conduct chemical analysis to determine dissolved concentrations. For accurate tracking, this tedious and expensive process has to be repeated frequently. Recent advances in Wireless Sensor Network (WSNs) have the potential to alleviate this labor intensive and time consuming task of data gathering. With wireless sensor nodes (motes) that measure and transmit the concentrations in situ in sampling wells in real-time, the need for manual collection of samples can be avoided. However, installation and maintenance of a large number of WSN nodes (containing sensors interfaced with motes) required for large-scale and evolving plumes can also be expensive. Minimizing the number of sensors will allow for this real-time monitoring technology to be a viable option. Sampling at regular spatial intervals (e.g., with sensors arranged in a rectangular grid) can be challenging at field sites, wherein an unstructured deployment of motes would be more realistic. The distribution of such nodes will be determined by other factors associated with the

geography and accessibility of deployment locations (e.g., to avoid buildings and other land infrastructure features). Effectively using a random deployment of sensors to obtain satisfactory results is also of interest.

Many interesting phenomenon in chemical plume tracking, seismic activity monitoring, animal migration tracking, etc., results in data in the forms of configurations with fairly regular boundaries and smooth gradients over the sensor field. Image processing algorithms often deals with similar regular features. A number of transforms such as Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT) are known to effectively compress images with regular shapes. Moreover these transforms are realizable as linear transforms and can easily be implemented on sensor motes with limited processing capacity. The goal of this work is to use such image processing techniques to reduce the amount of information transferred back and forth on sensor networks, while improving the resolution for a given set of sensors. The technique also enables the redistribution of the state of a part or the entire network back to each node efficiently. This feature could be used to develop future smart sensing schemes that operate more intelligently.

1.1 Related work

WSNs are widely used in applications related to environmental and habitat monitoring, reconnaissance and building automation [1,2,3,4,5]. A WSN node, also called a mote, consists of a processor and a wireless transceiver interfaced to a sensing device to measure the interested phenomenon, [6].

The feasibility of using WSNs for subsurface chemical plume tracking applications has been demonstrated in [7,8]. Some of the challenges posed in

traditional data gathering as described in [9] prevent the accurate tracking of the plume. Low maintenance, miniature sensing devices such as electrical conductivity sensors [10] are placed in wells at different depths for monitoring the plumes. The WSN has to be configured for efficient operation [11], and can be coupled to numerical models to form a closed loop system that uses WSN readings to calibrate the model, while the model provides information for data collection and node activation [12].

The potential for sensor readings related to many applications to be compressible is well known. Most work exploits the local correlation of the readings. Ref. [13] provides an information theoretic derivation, based on correlations of sensor readings, on savings possible for one- and two-dimensional networks. The Hierarchical Cooperation scheme presented in [13] achieves logarithmic scaling on traffic and schedule lengths. A data compression scheme based on wavelet decomposition and reconstruction is applied hierarchically at cluster heads in [14] with the goal of reducing waste in transmitting raw data to the datacenter. Ref. [15] points out the fact that wavelets approximate missing data on sensor readings when applied on a correlated structure. Their method - Data Correlation Compression (DCC) relies on Gaussian assumption of sensor data. A scheme utilizing historical data to reduce the amount of information needed to be transmitted is presented in [16]. A localizing scheme based on correlation is used to improve the accuracy on the multi-level clustering structure. Spatio-temporal correlation can be exploited to identify redundant sensors as indicated in [17], where wavelet transform and Fourier transform are used for compressing time series on individual sensors.

1.2 Contribution

A novel approach for reducing the number of sensors used and/or improving the resolution of the measurements in plume tracking applications is presented. With a fairly small fraction of sensor readings, the scheme is capable of approximating the status of the entire network to a significant accuracy. With the proposed approach, the energy spent gathering the status of the network and then redistributing that information back to the network is a magnitude less than with the conventional approach. Fusion has no effect to the message length, thus it requires no additional bandwidth.

Section 2 discusses the theoretical background. An analysis on the communication cost is presented in Section 3. Sensor deployment is addressed in Section 4 whilst Section 5 presents the results of the work. Scheme and results are evaluated in Section 6, followed by conclusions in Section 7.

2 DISCRETE WAVELET TRANSFORM - BASED COMPRESSION

We view the measurements collected by the sensors as pixels of an image. We assume that the underlying chemical concentration image has pixels on a fine grid. However, the wireless sensors are randomly deployed and only sparsely populate this grid. Our objective is to reconstruct the image on the fine grid from the sensor measurements. To minimize communication volume, we also wish to compress the data collected by each sensor before processing. We show in this paper that a simple DWT compression provides reasonable results. We start by reviewing two-dimensional DWT.

2.1 DWT compression and reconstruction

Most real images are compressible in the DWT domain. The DWT successively splits an image into an approximation component, which captures the smooth part of the image, and several detail components, as shown in Fig. 1. Roughly speaking, at each level, the H_i filter is a “low-pass” filter that passes the smoother part of the image and the G_i is a “high-pass” filter that passes detail components. As the DWT branches are traversed the size of the signal decreases diadically (down-sampling by 2). Since an image is two dimensional, each transformation is applied in two dimensions, the horizontal (row-wise) and vertical (column-wise) and as we proceed through successive branches the number pixels in the DWT image is reduced by a factor of 4.

To compress the image, we discard the detail components and only keep the coarsest approximation component produced by the bottom most branch in Fig. 1.

The DWT however needs to be calculated in a distributed fashion, where each sensor computes the contribution of its own measurement to the coarse approximation term without having the knowledge of measurements from the other nodes in the network. To accomplish this, we work with point spread functions (PSF) associated with each sensor node as we now describe.

Consider the coarsest approximation branch in Fig. 1. The coarsest approximation A for this branch can be expressed in matrix form as:

$$A = \Psi_V \cdot X \cdot \Psi_H \quad (1)$$

where, X is the original image defined under the fine grid, Ψ_V is the DWT matrix in vertical direction, Ψ_H is the DWT matrix in horizontal direction and (\cdot) denotes standard matrix multiplication.

The vertical DWT matrix Ψ_V (accounting for down-sampling) at the i^{th} level for an $n \times n$ image is given by ($j=1 \dots n/2^i, k=1 \dots n/2^{i-1}$):

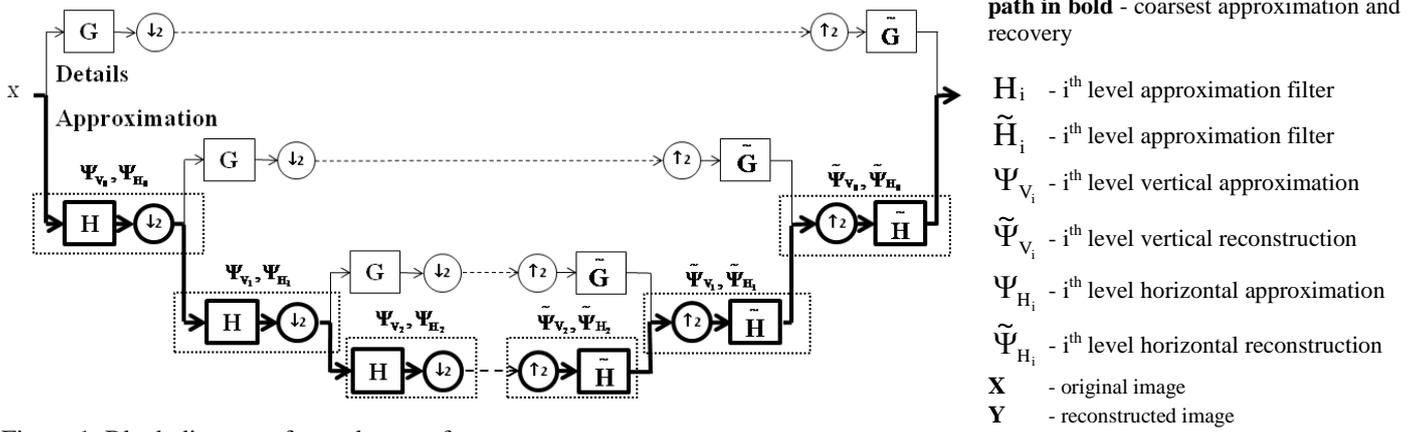


Figure 1. Block diagram of wavelet transform

$$\Psi_{V_i}(j, k) = \begin{cases} n_{j-2(k-1)} & i \leq j-2(k-1) \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

where, l is the length of the filter and h_i 's are the scaling coefficients of the filter.

The horizontal transform Ψ_{H_i} is the transpose of Ψ_{V_i} . The coarse approximation component A_i at level- i is given by:

$$A_i = \Psi_{V_i} \cdot A_{i-1} \cdot \Psi_{H_i} \quad (3)$$

where, A_0 is equal to X . If an L -level DWT the coarse approximation component A can be calculated from (1) with:

$$\Psi_V = \Psi_{V_{L-1}} \cdot \Psi_{V_{L-2}} \cdots \Psi_{V_1} \cdot \Psi_{V_0} \quad (4)$$

$$\Psi_H = \Psi_{H_0} \cdot \Psi_{H_1} \cdots \Psi_{H_{L-2}} \cdot \Psi_{H_{L-1}} \quad (5)$$

We can write the coarse term approximation to A as:

$$A = \sum_{i,j} A_{(i,j)} = \sum_{i,j} \Psi_V(\cdot, i) \cdot X(i, j) \cdot \Psi_H(j, \cdot) \quad (6)$$

where,

$$A_{(i,j)} = \Psi_V(\cdot, i) \cdot X(i, j) \cdot \Psi_H(j, \cdot) \quad (7)$$

is the contribution of the (i, j) pixel $X(i, j)$ to the coarse approximation A . The notations (\cdot, i) and (j, \cdot) respectively mean the i^{th} column and the j^{th} row of a matrix. Thus we can think of PSF of pixel (i, j) as:

$$\text{PSF}(i, j) = \Psi_V(\cdot, i) \cdot \Psi_H(j, \cdot) \quad (8)$$

If all $X(i, j)$ on the fine grid were available we could obtain the coarse approximation component A by simply transmitting PSFs scaled by the corresponding pixel value. The advantage is that each PSF can be calculated locally without knowledge of other pixels. The size of each PSF matrix is $n/2^L \times n/2^L$.

In our case sensors sparsely populate the image grid and we only have access to a small number of pixels at random locations. Nonetheless, we show that by combining the PSFs associated with these sensor locations we can still obtain a reasonable reconstruction of the chemical plume concentration.

Coefficients of the PSF depend on the wavelet transform and the filter selected. They can be built into the sensor nodes prior to deployment. Therefore computing the contribution of each sensor to the approximation can be done locally, which involves only scaling the PSF by the sensor reading.

Another advantage of this method is that computing the approximation of a part of or the entire sensor field becomes an addition of the contributions of each of the sensors. This allows sensors nodes to fuse their contributions to a single message conveniently and opportunely. For example, if the sensors are reporting to a base station over a tree, an intermediate node will add its coefficient matrix to the coefficient matrices it receives from its children nodes and transmits the result to its parent. Thus there will be only one transmission per link in the tree carrying all the information of the sub-tree below. Regardless of the position of a link in the tree the size of the composite PSF matrix that the link needs to communicate stays the same.

Once the base-station receives the sum of all contributions the image can be approximated by applying the inverse-DWT, as shown in Fig. 1. The synthesis filter pair (\tilde{H}, \tilde{G}) and analysis filter pair (H, G) are quadrature mirror filters, satisfying the perfect reconstruction condition [18]. Note that in the synthesis tree all detail components are zero.

However, this reconstruction can also be done at each sensor as will be shortly explained. While it is not essential for a node to know about the plume spread in the entire flow region, we envision future intelligent plume tracking systems where sensing operations within a locality may benefit by having global information on plumes. A simple extension to the scheme provides the ability to re-distribute the global information to cluster heads or individual sensors efficiently.

Suppose the coarse approximation component A is broadcasted from the base-station, then the inverse-DWT approximation can be calculated as:

$$\tilde{X} = \tilde{\Psi}_V \cdot \tilde{A} \cdot \tilde{\Psi}_H \quad (9)$$

where, \tilde{A} is the sensor network approximation to the coarse approximation component A . The matrices $\tilde{\Psi}_V$ and $\tilde{\Psi}_H$ are vertical and horizontal synthesis matrices. They are of the form (4) and (5) respectively, but constructed from elements of \tilde{H} similar to (2).

2.2 Implementation with Daubechies D4 wavelet

We present a sample implementation of a single level compression using Daubechies D4 wavelet. The coefficients of the H filter are shown in Table 1.

Table 1. Daubechies D4 scaling coefficients

Coefficient	Value
h_1	$(1 + \sqrt{3}) / 4\sqrt{2}$
h_2	$(3 + \sqrt{3}) / 4\sqrt{2}$
h_3	$(3 - \sqrt{3}) / 4\sqrt{2}$
h_4	$(1 - \sqrt{3}) / 4\sqrt{2}$

The single level vertical and horizontal DWT matrices are given by:

$$\Psi_V = \begin{pmatrix} h_1 & h_2 & h_3 & h_4 & 0 & 0 & 0 & 0 \\ 0 & 0 & h_1 & h_2 & h_3 & h_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & h_1 & h_2 & h_3 & h_4 \\ & & & & & & & \ddots \end{pmatrix} \quad (10)$$

$$\Psi_H = \Psi_V^T \quad (11)$$

The PSF to be loaded to each node is computed using (8). The PSF for a node is a few non-zero elements often appear as a single patch on a mostly empty matrix. During reporting, each node will scale the pre-loaded PSF by its measurement. Details of the compression algorithm are summarized in Fig. 2.

```
Compressed_report(reading, reports_from_children)
  initialize report := [] // empty matrix of the size
  of PSF
  for i=1:number_of_children
    report_i := report_of_child_node_i
    report += report_i // a matrix addition
  my_report := PSF*reading // a matrix scaling
  report += my_report // a matrix addition
  forward report to parent
```

Figure 2. Pseudo code for the compression algorithm

As a simple example, let us consider the 10×10 matrix of sensor readings shown in Fig 3.

0.00	0.00	1.76	2.47	2.67	2.47	1.76	0.00	0.00	0.00
0.00	0.00	2.20	2.80	2.98	2.80	2.20	0.00	0.00	0.00
0.00	1.05	2.47	3.02	3.18	3.02	2.47	1.05	0.00	0.00
0.00	1.36	2.62	3.14	3.30	3.14	2.62	1.36	0.00	0.00
0.00	1.45	2.67	3.18	3.33	3.18	2.67	1.45	0.00	0.00
0.00	1.36	2.62	3.14	3.30	3.14	2.62	1.36	0.00	0.00
0.00	1.05	2.47	3.02	3.18	3.02	2.47	1.05	0.00	0.00
0.00	0.00	2.20	2.80	2.98	2.80	2.20	0.00	0.00	0.00
0.00	0.00	1.76	2.47	2.67	2.47	1.76	0.00	0.00	0.00
0.00	0.00	0.93	1.96	2.20	1.96	0.93	0.00	0.00	0.00

Figure 3. A sample 10×10 measurement matrix

Each sensor is also assigned with a PSF calculated according to (8). For example node (7,4) would

calculate its PSF using (8) for a single level compression by multiplying the 7th column of Ψ_V and 4th row of Ψ_H to produce:

$$\text{PSF}(7,4) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -0.029 & 0.188 & 0 & 0 & 0 \\ -0.062 & 0.404 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

According to Fig. 3 the reading of the sensor node (7,4) is 3.02. Therefore the contribution of the node (7,4) is obtained by scaling PSF(7,4) by 3.02.

Similarly, all the nodes will calculate their PSF and then scale by their measurement. The approximation of the entire sensor field is obtained by summing up all the approximations generated by individual nodes.

$$A = \sum_{i=1}^{10} \sum_{j=1}^{10} A_{(i,j)} \quad (12)$$

$$A = \begin{pmatrix} 0.210 & 5.011 & 5.794 & 1.456 & -0.008 \\ 1.754 & 5.899 & 6.485 & 3.265 & -0.233 \\ 2.000 & 6.027 & 6.583 & 3.549 & -0.269 \\ 0.638 & 5.474 & 6.204 & 2.026 & -0.066 \\ -0.002 & 3.731 & 4.539 & 0.839 & 0.000 \end{pmatrix}$$

To obtain an approximation to the chemical plume image we inverse-DWT is applied on A and the resultant matrix is shown in Fig 4.

0.05	0.08	1.60	2.71	2.57	2.77	1.29	0.25	0.20	-0.12
0.08	0.15	1.83	3.08	2.89	3.08	1.49	0.38	0.25	-0.15
0.41	0.76	2.12	3.21	3.03	3.19	1.91	1.02	0.37	-0.34
0.65	1.23	2.39	3.40	3.22	3.36	2.28	1.52	0.48	-0.49
0.62	1.16	2.35	3.37	3.19	3.33	2.22	1.45	0.46	-0.47
0.65	1.23	2.39	3.39	3.21	3.35	2.28	1.53	0.48	-0.49
0.34	0.64	2.10	3.24	3.06	3.22	1.86	0.91	0.35	-0.31
0.13	0.23	1.90	3.14	2.95	3.14	1.57	0.47	0.27	-0.18
0.07	0.12	1.49	2.52	2.41	2.61	1.22	0.26	0.19	-0.12
-0.04	-0.07	1.15	2.03	1.99	2.19	0.88	-0.02	0.10	-0.05

Figure 4. The reconstructed measurement matrix

3 EXCHANGE OF SENSOR DATA

Consider a network in which the sensors are placed on an $n \times n$ grid as shown in Fig. 5(a) with a tree communication structure rooted at the center of the grid. We assume that a node is capable of communicating with its eight immediate neighbors. The levels of the tree then form co-centric squares, with the maximum depth of the tree at $n/2$. The average depth of a node from the root is $n/6$.

Communication cost is two folds: reporting and re-distributing. During reporting, sensors report their readings to the root. Then, root informs the status of the network to each of the sensors during re-distributing. Reporting costs can be alleviated by making sensors not report, if the reading is null. However, the node still needs to take part in communication to relay reports from the sub-tree descending from itself.

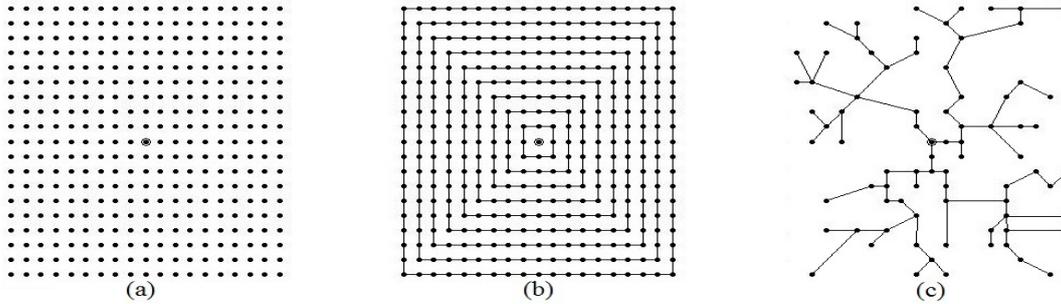


Figure 5. (a) Nodes placed in a grid with root at the center (b) Levels of nodes (c) A random node deployment with a tree communication structure

3.1 Conventional reporting

Under the conventional monitoring scheme, each node reports its measurement along with its ID or coordinates. This report has to be to the root, i.e. $n/6$ times on average. Since there are n^2 nodes in the network, the total reporting cost is $O(n^3)$. If the null readings are not transmitted, then communication cost reduces to $O(kn^2)$, where k is the number of nodes having a non-zero reading. Further, overhead in transmitting individual reports as separate packets can be saved by packing a few if not all reports received from the sub-tree to a single message along with its report. Such a fusion saves overhead cost, yet no savings are made on the amount of payload transmitted. It is to be noted that in the conventional scheme, reporting the location and the reading provides no loss of information.

3.2 Compressed reporting, fusion and recovering missing data

Compressed reporting exploits the compressibility of data. Instead of reporting the reading and location information tuple, nodes report wavelet coefficients. Further, data is fused by adding coefficient matrices. As in conventional scheme, the nodes having a null reading do not contribute to the coefficient matrix.

Each contributing node will produce a coefficient matrix, which is a small patch of non-zero elements. By putting together these patches, the approximation for the entire matrix is formed. Patches are in fact added onto the coefficient matrix - which enable an effective fusion scheme, where the message length does not change. Under conventional reporting, reading of each node was stored in the message separately. Such would lengthen message length as the message arrives at the root. But with the compressed reporting, nodes keep adding their contributions onto the existing message. Therefore the length of the message is not affected.

The size of the coefficient matrix is $m \times m$, with m equal to $n/2^L$, where L is the number of levels of compression applied. When fewer nodes have readings, instead of transmitting the entire coefficient

matrix, the patch and its location information can be transmitted to save cost.

Compressed reporting imposes a smoothing operation on the measurements. Thus it automatically approximates readings of the locations which provided no input. If a malfunctioning node feeds in an abnormally large contribution (an outlier), that would be suppressed as well.

Let us demonstrate recovering missing data points using the matrix in Fig 3 by randomly dropping some 10 measurements out of the 100. The resultant is shown in Fig 6. This doesn't correspond to a sparse sensor network, but still demonstrates how missing data points are handled. The actual chemical plume example presented in Section 5 corresponds to a truly sparse network where the sensors populate only 25% of the grid points.

0.00	0.00	1.76	2.47	2.67	2.47		0.00	0.00	0.00
0.00	0.00	2.20	2.80	2.98	2.80	2.20		0.00	0.00
0.00	1.05	2.47	3.02	3.18	3.02	2.47	1.05	0.00	0.00
		2.62		3.30	3.14	2.62	1.36	0.00	0.00
0.00	1.45	2.67	3.18	3.33	3.18	2.67	1.45	0.00	0.00
0.00	1.36	2.62	3.14	3.30	3.14	2.62		0.00	0.00
0.00		2.47	3.02	3.18	3.02	2.47	1.05	0.00	0.00
		0.00	2.20	2.80	2.98	2.80	2.20	0.00	0.00
0.00	0.00	1.76	2.47	2.67			1.76	0.00	0.00
0.00	0.00	0.93		2.20	1.96	0.93	0.00	0.00	0.00

Figure 6. A sample 10x10 measurement matrix with 10 missing values

The approximation derived from the available nodes if shown in Fig 7(a) and the reconstruction is shown in Fig 7(b). By comparing Fig 7(b) with Fig 3, it can be noted missing points are approximated quite closely compared to the range of measurements.

0.305	5.351	5.603	1.045	-0.031
1.139	3.701	6.462	3.413	-0.085
1.802	6.027	6.730	2.595	-0.238
0.179	5.758	5.741	2.026	0.000
0.210	2.510	3.451	0.649	0.000

0.09	0.16	1.56	2.61	2.39	2.50	1.09	0.09	0.14	-0.10
0.10	0.19	2.06	3.44	2.98	3.01	1.33	0.11	0.16	-0.12
0.29	0.52	1.58	2.42	2.78	3.27	1.89	1.01	0.40	-0.28
0.43	0.77	1.37	1.89	2.80	3.60	2.36	1.67	0.58	-0.41
0.51	0.96	2.06	2.98	3.11	3.45	2.03	1.08	0.39	-0.37
0.61	1.16	2.51	3.64	3.34	3.42	1.91	0.83	0.30	-0.38
0.22	0.42	2.11	3.39	2.99	3.05	1.71	0.75	0.32	-0.25
-0.03	-0.07	1.93	3.39	2.80	2.77	1.54	0.63	0.32	-0.15
0.07	0.12	1.24	2.08	1.99	2.15	1.03	0.26	0.17	-0.10
0.07	0.13	0.69	1.11	1.34	1.62	0.62	-0.04	0.06	-0.04

Figure 7. (a) Approximation for the sensor readings with missing nodes (b) Reconstruction

3.3 Hybrid reporting

Compressed reporting captures the status of the (partial) network by an $m \times m$ matrix, instead of the number of nodes many tuples. However, at lower depths of the tree, where only a few nodes observe some reading, reporting the $m \times m$ matrix or the coefficients patch may be too costly. The hybrid scheme proposes to use conventional scheme until the number of nodes with a non-zero reading is below $\frac{1}{2} m \times m$. Once this threshold is reached, readings are to be transformed to the $m \times m$ coefficient matrix. Implementation of the hybrid scheme is explained in Fig 8. Until transmitting the coefficient matrix is effective than reporting raw data, conventional scheme is followed, denoted by *reporting_mode* : 0. When the list of raw data and the coordinate information grow past the threshold, the coefficient matrix is formed. Thereafter, all nodes contribute in the compressed mode.

```

Get_Mode_of_Reporting()
enum mode {0,1} // 0:raw data, 1: compressed data
message_length:=0
for i=1:number_of_children
    receive message_i
    message_length += length(message_i)
if message_length > 0.5*m^2
    reporting_mode:=1
else
    reporting_mode:=0
Construct_Message()
if reporting_mode==0
    message=[];
    for i=1:number_of_children
        receive message_i
        message:=[message; message_i];
    if measurement != 0
        message:=[message; [coordinates, measurement]];
if reporting_mode==1
    message=[];
    for i=1:number_of_children
        receive message_i
        message += message_i
        message += PSF*reading
Transmitting_Message()
if reporting_mode==1
    transmit(message)
if reporting_mode==0
    if length(message) < 0.5*m^2
        transmit(message)
    else
        msg=[];
        for i=1:length(message)
            coord=message_i(coordinate)
            value=message_i(measurement)
            msg += PSF(coord)*value
        transmit(msg)

```

Figure 8. Pseudo code for hybrid reporting scheme

3.4 Re-distribution

Future smart sensing schemes on large networks would benefit by being aware of the state of the entire network. Thus a phase where the status of the network is re-distributed back to the network is discussed here. The methodology of re-distributing status of the network is intuitive from the hybrid scheme. The status of the $n \times n$ network is com-

pressed to $m \times m$. But if the number of non-zero values in the network is below $\frac{1}{2} m \times m$, re-distribution is more effectively done with a conventional approach, where the location and reading information tuple is broadcast. Otherwise the compressed matrix delivers the status of the network more effectively.

Then at each node, the inverse transform is performed to recover the status of the entire network. By doing so each and every node becomes aware of the entire network.

3.5 Potential issues

Here we discuss issues associated with the scheme that would be of interest to certain applications.

- Reconfiguring nodes: since each node use a unique combination of Ψ_V and Ψ_H , reconfiguring can be tedious. However, using nodes programmable over the network would alleviate the effort.
- Speed of the plume: a cycle of reporting and dissemination is expected to complete while the plume is effectively stationary. If the cycle is a slow process, the picture built using the reports would be inaccurate.
- Ringing effects: this is an issue natural to lossy compression. Since the high-frequency components are discarded, a slight ringing artifact builds on the image.
- Blurred image: approximation is analogous to a low-pass filter, which smoothens the image. Thus reconstructed images would be less crispy and more blurred.
- Effects of missing contributions: although the scheme interpolates the missing locations quite accurately according to a smoother description of the plume, it draws energy from the available contributions. Thus missing contributions causes noise on the available.
- Poor alignment: the simulation results presented later assumes a worst case of a purely random deployment. A pure random deployment would have a few cluttered nodes and a few blank areas. Thus the approximation would be more biased to the cluttered locality and less towards the empty regions. However, actual deployments are not purely random and will suffer less from such effects.

3.6 Analytical results

To identify uniquely and for communication sensor nodes need $\gtrsim \log_2 n^2$ bit long address. If we assume the reading produces some b bit floating point, the cost of reporting under conventional scheme is $\approx (\log_2 n^2 + b) \cdot n/6$ per node. If only k nodes read non-zero values and report, then the total reporting cost is $\approx k \cdot (\log_2 n^2 + b) \cdot n/6$. Pure compressed reporting scheme requires transmitting an $m \times m$ matrix. Thus

the reporting cost would be $k \cdot m^2 \cdot b \cdot n/6$, where b is the length of a coefficient. Reporting cost can be saved by reporting patches instead of the entire matrix, where applicable. Moreover, the hybrid scheme would provide much savings.

Reporting is economical for certain choices of wavelets and levels, which also determine the required precision of the coefficients. Nonetheless, reporting in compressed form is essential to implement compressed re-distribution in a distributed form. As well to interpolate for the missing location, compression scheme has to be employed at reporting, irrespective of the communication cost. Re-distributing is effectively achieved for large values of k and n , i.e. for large network with a large fraction of nodes reading non-zero measurements.

The key advantage of the compressed reporting and re-distributing is the information of a vast network is represented using only a few coefficients. Thus less information needed to be transmitted in order to deliver the status of the network.

4 SENSOR DEPLOYMENT

When sensor nodes are placed on a regular grid, they can be matched to pixels of an image (Fig. 5(a) and Fig. 5(b)). To calculate the contribution made by each pixel (sensor) for the approximation, each sensor is fed with a corresponding PSF. Thus at reporting, each node will scale its PSF by the reading and report the resultant matrix. Further, nodes fuse readings simply by adding the contribution matrices.

When all the contribution matrices of all the sensors are added, the approximation for the entire sensor field is formed. This approximation can be then transmitted back to the sensor field, so that each of the sensors learns the status of the entire network.

When constructing the approximation, if the contribution of some of the pixels were not available, an interpolated value will be automatically assigned to those pixels. This relaxes the need of a complete grid which is attractive for many applications, and is discussed next.

4.1 Random points on a grid

A random deployment of sensors can be treated as a sparse deployment on a grid (Fig. 5(c)). As pointed out above, DWT based approximation scheme fills out the missing grid points with interpolated values based on the available grid points automatically. This allows using the same scheme even for a random deployment of sensor nodes.

As before, each node is assigned with its PSF matrix based on the grid point the sensor is located. The rest of procedure is the same. When reconstructing, an image of the size of the grid is formed, where missing grid points are assigned with an interpolated value based on the wavelet used.

4.2 Representing a more realistic scenario

Deploying sensors on an exact grid is difficult and not economical for many environmental sensing applications. A deployment exhibiting characteristics of a random deployment can be considered more realistic. The nodes can be considered to be randomly placed at points on a finer grid for computational convenience. Another issue with wireless sensor networks is the availability of the nodes. At a given time it is quite likely that a significant fraction of the nodes may either be sleeping, or even dead. Once the random deployment is treated as a sparse deployment over a fine grid, unavailability can be accounted as a much sparser deployment. Thus when resolved, measurements will be interpolated for on each point on the fine grid.

5 RESULTS

In this section we evaluate the compressed data reporting and dissemination scheme using a dataset corresponding to a subsurface plume. The dataset and numerical results are presented next.

5.1 Synthetic Plume Data Set

The data that is needed in field problems will come from a set of sensors that are installed in water quality monitoring wells. As such data set was not available, a synthetic data set using a groundwater flow (MODFLOW) transport model (MT3DMS) was generated [19,20,21,22]. Synthetic data emulating a propagating plume over a period of 3 years, collecting daily samples are used as experimental data for this work. The synthesizer software allows placing sensors and making measurements at any desired location. By placing sensors at a complete fine grid, the actual plume is recognized. Then sensors are placed at random location for the experiments. Sensor field is represented as a 64×64 pixel image. The readings are compressed using a two-level Daubechies-4 wavelet. The compressed image is 16×16 .

5.2 Numerical results

At each time interval, selected based on plume tracking application, a snapshot of the sensor field is built using the compressed reporting method described in Section 3. For our experiment, the time interval was selected to be a day. The error is defined as the difference between the calculated value and the actual value normalized to the largest reading (which is the range of the measurements), and expressed as a percentage.

Four versions of errors are defined. Given a snapshot, the mean of the errors and the maximum of the errors can be taken. Then over the entire sensing duration (3 years in our case) the mean and the maximum of above two can be taken.

Transmission cost is evaluated in terms of the number of transmissions. The experiment used double precision floating point values for both measurements and coefficient matrices. Thus the actual transmission cost is a factor of the number of transmissions made.

5.2.1 Accuracy

The proposed scheme exploits the effectiveness of lossy compression. Inevitably, some of the information is destroyed during the reporting phase. Table 2 assesses the error introduced by the approximation.

Table 2. Error of approximation

Error (%)	Mean over entire sensing period		Max over entire sensing period	
	Mean over a snapshot	Max over a snapshot	Mean over a snapshot	Max over a snapshot
	2.5	55.4	9.5	82.4

More realistic networks are represented as a sparse deployment of nodes over a grid. Their performance is comparable when a large fraction of nodes are unavailable on a grid. Table 3 summarizes error performance when 25%, 50% and 75% of the nodes are unavailable. Table 4 shows the mean and the standard deviation of the mean error over 100 random network settings. It can be noted that mean error is small and it varies very little.

Table 3. Effect of partial availability on the error

Error compared against	Dead nodes 25%			
	Mean over entire sensing period		Max over entire sensing period	
	Mean over a snapshot	Max over a snapshot	Mean over a snapshot	Max over a snapshot
Actual	3.2	65.9	10.5	96.8
Approximation	2.7	24.9	4.9	47.7
Error compared against	Dead nodes 50%			
	Mean over entire sensing period		Max over entire sensing period	
	Mean over a snapshot	Max over a snapshot	Mean over a snapshot	Max over a snapshot
Actual	4.9	76.7	13.6	98.8
Approximation	5.3	41.1	9.0	82.6
Error compared against	Dead nodes 75%			
	Mean over entire sensing period		Max over entire sensing period	
	Mean over a snapshot	Max over a snapshot	Mean over a snapshot	Max over a snapshot
Actual	7.0	87.9	18.4	103.5
Approximation	7.9	56.1	11.8	84.9

Table 4. Mean and standard deviation of the accuracy

Dead node %	Error compared against	Mean	Standard deviation
25	Actual	3.2	0.09
	Approximation	2.7	0.16
50	Actual	4.9	0.14
	Approximation	5.3	0.19
75	Actual	7.0	0.15
	Approximation	7.9	0.16

5.2.2 Communications cost savings

Compression based data gathering and re-distributing scheme saves floating point transmissions by a factor of 5 in average. When hybrid scheme is employed instead, the saving reaches a factor of 10. Figure 9 shows the total cost of using the three schemes over the sensing period.

Largest factor of the savings is accounted to the re-distributing phase as shown in Fig. 10. However, for the re-distribution to be implemented in a distributed fashion, the reporting scheme has to be implemented in either the compressed form or the hybrid form. The hybrid scheme improves the compressed scheme further by a factor of 2 in average. The performance of the hybrid scheme over the compressed scheme is presented in Fig. 11.

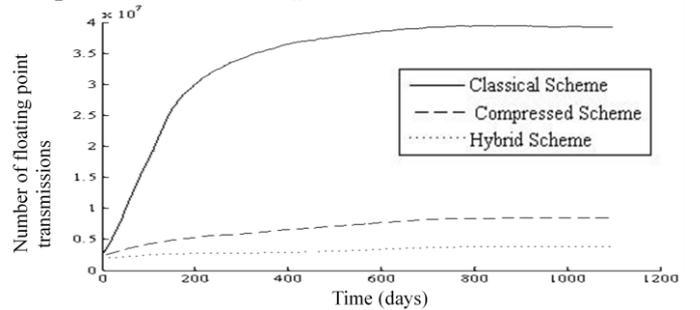


Figure 9. Cost of re-distributing sensor information over the entire network over time.

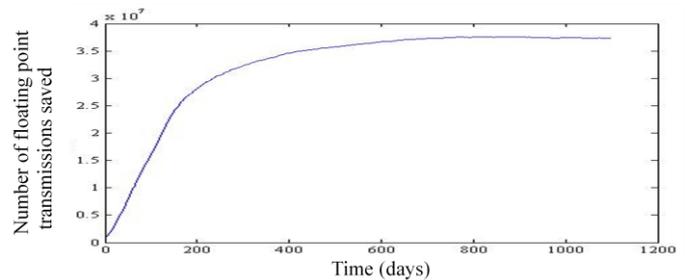


Figure 10. Amount of floating point transmissions saving by compression over time.

5.2.3 Approximating missing data

DWT coefficients automatically approximate values for the missing locations during reconstruction. Figure 12 displays the approximation capacity of the scheme. Only 25% of the sensors were activated in the sensor field. This could also be interpreted as, only 25% of the grid points actually contained sensors. Figure 12(a) shows a snapshot of the plume to be detected. But only some random 25% of the grid points indicated in Fig. 12(b) are available for measurements. The non-zero measurements provided by

the available sensors are indicated in Fig. 12(c). With coefficients for these non-zero measurements the plume is approximated as in Fig. 12(d). It is to be noted that the mean error between the approximated reconstruction using only 25% of the measurements is only 7% as shown in Table 3.

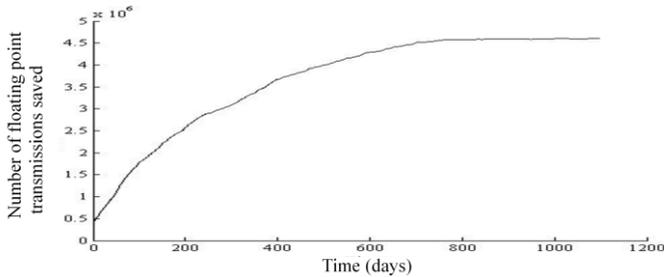


Figure 11. Amount of floating point transmissions saved using hybrid reporting instead of compressed reporting over time.

6 DISCUSSION

The goal of the presented scheme is to gather and re-distributed sensor data from each of the sensors to entire network cost effectively. The communication structure is a tree rooted at the center of the network. All the nodes observing the interested phenomenon generates a report and pass it up the tree. Thus a description of the entire network is generated at the root. Then the root sends down this information back to the network, making all the nodes aware of the entire network.

Under conventional scheme each node reports its reading and the location information, and all the nodes take part in passing this information to the root. The root collects all the information and build

giant picture of the network which is then passed down to the network. The conventional scheme does not take into account the compressibility of data. Although it preserves perfect accuracy, most applications tolerate errors to a certain degree to account for noise which is inevitable in measurements. Compressed re-distributing scheme proposed exploit the tolerance to mild loss of information. The coefficients also enable data fusion. Thus when multiple messages are to be transmitted on the same link, they can be fused to a single message saving overhead. Moreover, the fusion does not change the data length, whereas under the conventional scheme the length of the message is increased when multiple messages are packed.

Compressed scheme reduces the operations at the root. Under the conventional scheme, the root has to gather and form the giant message containing information of the entire network. In the compressed scheme the root has no more operations than a regular node in the network. It sums the coefficients and pass on to the children nodes.

Although compressed reporting alone may not be communication effective, it is essential to facilitate interpolation of missing points, improve resolution and for a distributed implementation of the dissemination scheme. So that the burden on the root is alleviated, and producing a distributed deployment of the scheme. The hybrid scheme utilizes the effective components from both the conventional and compressed schemes. It prevents forming a large coefficient matrix where data is effectively transmitted conventionally, but also applies compression later on, to utilize the advantages in both the schemes.

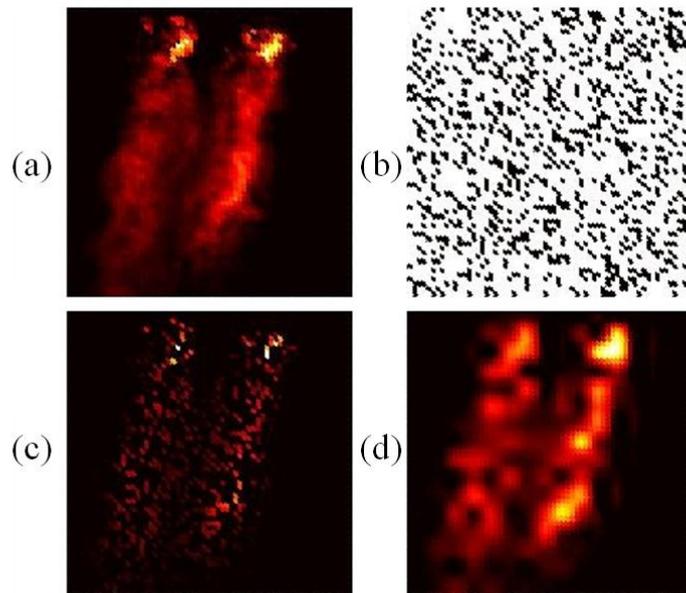


Figure 12. (a) The actual plume (b) a sample deployment of sensors (c) non-zero reading provided by 25% of sensors (d) approximate plume reconstructed.

Computation requirement at sensors nodes are commendable as well. Compression and decompression require matrix multiplication, which is an $O(n^2)$ floating point operation. Fusion requires matrix addition which is $O(n)$ floating point operation. The PSF needed for each node is proposed to be preloaded to each node. The hybrid scheme requires a list of potential PSFs of its children which can also be preloaded.

7 CONCLUSIONS

The scheme estimated the state of the entire network within a 7% error bound using only 25% of the measurements, and demonstrated a communication savings by factor of 10 when applied for the plume data. Thus the scheme is capable of improving resolution of the measurements made, and also to reduce the number of sensors to be used to achieve a given error bound.

Hybrid scheme exploits the effective components from conventional and the compressed reporting schemes and cuts down the communication cost by a magnitude. Computation and memory requirements needed for all the operation in the feasible range for most common place sensor nodes.

8 REFERENCES

1. K. Romer, F. Mattern, "The Design Space of Wireless Sensor Networks," *Wireless Communications, IEEE*, vol.11, no.6, pp. 54- 61, Dec. 2004.
2. F. Shu, M. N. Halgamuge and W. Chen, "Building Automation System Using Wireless Sensor Networks: Radio Characteristics and Energy Efficient Communication Protocols", *Special Issue on Sensor Network for Building Monitoring: From Theory to Real Application, Electronic Journal of Structural Engineering, EJSE*, pp 66-73, 2009.
3. M. N. Halgamuge, T. K. Chan and P. Mendis, "Experimental Study of Link Quality Distribution in Sensor Network Deployment for Building Environment", *Special Issue on Sensor Network for Building Monitoring: From Theory to Real Application, Electronic Journal of Structural Engineering, EJSE*, pp 28-34, 2009.
4. L.M.R. Peralta, L.M.P.L. Brito, and B.A.T. Gouveia "The WISE-MUSE Project: Environmental Monitoring and Controlling of Museums based on Wireless Sensor Networks", *Special Issue on Sensor Network for Building Monitoring: From Theory to Real Application, Electronic Journal of Structural Engineering, EJSE*, pp 46-57, 2009.
5. S. Alahakoon, D.M.G. Preethichandra, and E.M. Ekanayake, "Sensor Network Applications in Structures - A Survey", *Special Issue on Sensor Network for Building Monitoring: From Theory to Real Application, Electronic Journal of Structural Engineering, EJSE*, pp 1-10, 2009.
6. I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: A survey", *IEEE Transactions on Wireless Communications, Computer Networks*, 2002, 38, pp.393-422.
7. L. Porta, T. H. Illangasekare, P. Loden, Q. Han, and A. P. Jayasumana, "Continuous Plume Monitoring Using Wireless Sensors: Proof of Concept in Intermediate Scale Tank", *Journal of Environmental Engineering*, September 2009, volume 135, issue 9, pp.831-838.
8. P. Loden, Q. Han, L. Porta, T. Illangasekare, A. P. Jayasumana, "A Wireless Sensor System for Validation of Real-time Automatic Calibration of Groundwater Transport Models," *The Journal of Systems and Software, Elsevier*, 82 (2009) 1859–1868.
9. R.W. Puls, and C.J. Paul, "Multi-layer sampling in conventional monitoring wells for improved estimation of vertical contaminant distributions and mass", *Journal of Contaminant Hydrology*, 25, 1997, pp.85-111.
10. A. Kaya, and H.Y. Fang, "Identification of contaminated soils by dielectric constant and electrical conductivity", *Journal of Environmental Engineering*, 123(2), 1997, pp.169-177.
11. A. P. Jayasumana, Q. Han and T. Illangasekare, "Virtual Sensor Networks - A Resource Efficient Approach for Concurrent Applications," *Proc. 4th International Conference on Information Technology: New Generations (ITNG 2007)*, Las Vegas, NV, April 2007.
12. Q. Han, A. P. Jayasumana, T. Illangasekare & T. Sakaki, "A Wireless Sensor Network Based Closed-Loop System for Subsurface Contaminant Plume Monitoring," *Proc. 22nd IEEE International Parallel and Distributed Processing Symposium, Miami, FL, April, 2008*.
13. T. Elbatt, "On the trade-offs of cooperative data compression in wireless sensor networks with spatial correlations", *IEEE Transactions on Wireless Communications*, May 2009, vol.8, no.5, pp.2546-2557.
14. J. Zhuo, C. Meng, and M. Zou, "A Task Scheduling Algorithm of Single Processor Parallel Test System", *Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, (SNPD 2007) July 30 2007-Aug. 1 2007*, vol.1, pp.627-632.
15. H. Jin, Q. Ren, J. Li, and Yong Shi, "An Approximate Query Processing Method Based on Data Correlation in Sensor Networks", *International Conference on Embedded Software and Systems (ICCESS '08)*, 29-31 July 2008, pp13-18.
16. A. Deligiannakis, Y. Kotidis, and N. Roussopoulos, "Dissemination of compressed historical information in sensor networks", (Oct. 2007), *The VLDB Journal* 16, issue 4, pp 439-461.
17. A. Guitton, A. Skordylis, and N. Trigoni, "Correlation-based data dissemination in traffic monitoring sensor networks", *IEEE Wireless Communications and Networking Conference (WCNC 2007)*, Hong Kong, 11–15 March 2007.
18. M. Vetterli, and J. Kovacevic, "Wavelets and subband coding", Prentice Hall, 1995.
19. K. Barnhart, "Re-conceptualizing groundwater transport models 2 in the wireless sensor network data context", *PhD Dissertation, Environmental Sciences and Engineering, Colorado School of Mines*, 2010.
20. K. Barnhart, I. Urteaga, Q. Han, A. Jayasumana, and T. Illangasekare, "On integrating groundwater transport models with wireless sensor networks," *Ground Water*, 48(5), pp771-780.
21. C. Zheng, and P. Wang. "MT3DMS: A modular 3D multispecies transport model for simulation of advection, dispersion, and chemical reactions of contaminants in groundwater systems," *US Army Engineer Research and Development Center Contract Report SERDP-99-1*, 1999.
22. A. Harbaugh, E. Banta, M. Hill, and M. McDonald. "the USGS modular ground-water model user guide to modularization concepts and the groundwater flow process," *USGS Open File Report 00-92*. Reston, Virginia: USGS, 2000.