A Bayesian Analysis of Compressive Sensing Data Recovery in Wireless Sensor Networks

Riccardo Masiero, Giorgio Quer, Michele Rossi and Michele Zorzi
Department of Information Engineering, University of Padova
via Gradenigo 6/B – 35131, Padova, Italy
Email: {riccardo.masiero, giorgio.quer, rossi, zorzi}@dei.unipd.it

***Abstract*—In this paper we address the task of accurately reconstructing a distributed signal through the collection of a small number of samples at a data gathering point using Compressive Sensing (CS) in conjunction with Principal Component Analysis (PCA). Our scheme compresses in a distributed way real world non-stationary signals, recovering them at the data collection point through the online estimation of their spatial/temporal correlation structures. The proposed technique is hereby characterized under the framework of Bayesian estimation, showing under which assumptions it is equivalent to optimal maximum *a posteriori* (MAP) recovery. As the main contribution of this paper, we proceed with the analysis of data collected by our indoor wireless sensor network (WSN) testbed, proving that these assumptions hold with good accuracy in the considered real world scenarios. This provides empirical evidence of the effectiveness of our approach and proves that CS is a legitimate tool for the recovery of real-world signals in WSNs.**

I. Introduction

In this paper, we look at data gathering approaches for Wireless Sensor Networks (WSNs) which are able to measure large amounts of data with high accuracy by only requiring the collection of a small fraction of the sensor readings. In the past few years, the research community has been providing interesting contributions on this topic.

In particular, Compressive Sensing (CS) [1]–[3] is a recent compression technique that takes advantage of the inherent correlation of the input data by means of quasi-random matrices. CS was originally developed for the efficient storage and compression of digital images, which show high spatial correlation. Since the pioneering work of Nowak [4], there has been a growing interest in this technique also by the networking community. In contrast to classical approaches, where the data is first compressed and then transmitted to a given data gathering point (hereafter called the sink), with CS the compression phase can be jointly executed with data transmission.

In our previous paper [5] we address the issue of designing a simple protocol based on CS for the online recovery of large data sets through the collection of a small number of readings. In detail: 1) we exploit the combination of CS with Principal Component Analysis (PCA) [6]; 2) we design a scheme which iteratively learns optimal transformations for CS through the online estimation of the monitored signal correlation structure. Hence, PCA is exploited to iteratively provide a good transformation basis that allows us to continuously reconstruct signals through the online estimation of their statistics. The effectiveness of our approach for data gathering and recovery has been proved in [5] for both synthetic and real signals.

In this paper we investigate the statistical distribution of the principal components of signals gathered from an actual Wireless Sensor Network (WSN) deployment. This analysis provides an explanation of the good results that we have obtained in [5] and proves that CS is a legitimate tool for the recovery of real-world signals in WSNs. The main contributions of this paper are:

- the inference of the statistical distribution of the principal components of real world signals;
- a Bayesian justification of the good results achieved by our monitoring framework.

The above points are tackled by means of Bayesian theory, which provides a general framework for data modeling [7], [8]. The Bayesian framework, in fact, has been addressed in the very recent literature to develop efficient and auto-tunable algorithms for CS, see [9]. However, previous work addressing CS from a Bayesian perspective has mainly been focused on the theoretical derivation of CS and its usefulness in the image processing field. With the present paper, instead, we provide empirical evidence of the effectiveness of CS in an actual WSN monitoring scenario.

The paper is structured as follows. In section II we provide a mathematical description of our data recovery framework. In section III we substantiate the optimality of our combined CS and PCA framework using tools from Bayesian theory. In section IV we analyze the principal component distribution and confirm that the assumptions under which our framework is effective hold for real world signals. Section V concludes the paper.

II. Mathematical Framework: Tools

In this section we first review basic tools from PCA and CS and we subsequently illustrate our monitoring framework, which jointly exploits these two techniques.

Principal Component Analysis [6]: the Karhunen-Loève expansion is the theoretical basis for PCA. It is a method to represent through the best M-terms approximation a generic N-dimensional signal, where $N > M$, given that we have full knowledge of its correlation structure. In practical cases, i.e., when the correlation structure of the signals is not known a

9781-4244-3941-6/09/$25.00

priori, the Karhunen-Loève expansion can be achieved thanks to PCA [6], which relies on the online estimation of the signal correlation matrix. We assume to collect measurements according to a fixed sampling rate at discrete times $k = 1, 2, \ldots, K$. In detail, let $\mathbf{x}^{(k)} \in \mathbb{R}^N$ be the vector of measurements, at a given time k, from a WSN with N nodes. $\mathbf{x}^{(k)}$ can be viewed as a single sample of a stationary vector process $\mathbf{x}$. The sample mean vector $\overline{\mathbf{x}}$ and the sample covariance matrix $\widehat{\mathbf{\Sigma}}$ of $\mathbf{x}^{(k)}$ are defined as:

$$\overline{\mathbf{x}} = \frac{1}{K}\sum_{k=1}^{K}\mathbf{x}^{(k)} \ , \ \widehat{\mathbf{\Sigma}} = \frac{1}{K}\sum_{k=1}^{K}(\mathbf{x}^{(k)} - \overline{\mathbf{x}})(\mathbf{x}^{(k)} - \overline{\mathbf{x}})^T \ .$$

Given the above equations, let us consider the orthonormal matrix $\mathbf{U}$ whose columns are the unitary eigenvectors of $\widehat{\mathbf{\Sigma}}$, placed according to the decreasing order of the corresponding eigenvalues. It is now possible to project a given measurement $\mathbf{x}^{(k)}$ onto the vector space spanned by the columns of $\mathbf{U}$. If we define $\mathbf{s}^{(k)} \stackrel{def}{=} \mathbf{U}^T(\mathbf{x}^{(k)} - \overline{\mathbf{x}})$, by construction of the projection matrix $\mathbf{U}^T$ we have that the entries of $\mathbf{s}^{(k)}$ are ordered as follows: $s_1^{(k)} \geq s_2^{(k)} \geq \cdots \geq s_N^{(k)}$. If the instances $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \cdots, \mathbf{x}^{(K)}$ of the process $\mathbf{x}$ are temporally correlated, then there exists an $M < N$ such that for $i > M$ we have that $s_i^{(k)}$ is negligible with respect to the previous entries of $\mathbf{s}^{(k)}$, i.e., $s_i^{(k)} \ll s_j^{(k)}$ for $j \leq M$ and $i > M$. Thus, each sample $\mathbf{x}^{(k)}$ can be very well approximated in an M-dimensional space by just accounting for $M < N$ coefficients. According to the previous arguments we can write each sample $\mathbf{x}^{(k)}$ as:

$$\mathbf{x}^{(k)} = \overline{\mathbf{x}} + \mathbf{U}\mathbf{s}^{(k)} \ , \tag{1}$$

where the N-dimensional vector $\mathbf{s}^{(k)}$ can be seen as an M-sparse vector, namely, a vector with at most $M < N$ non-zero entries. Note that the set $\{\mathbf{s}^{(1)}, \mathbf{s}^{(2)}, \cdots, \mathbf{s}^{(K)}\}$ can also be viewed as a set of samples of a random vector process $\mathbf{s}$. In summary, thanks to PCA, each original point $\mathbf{x}^{(k)} \in \mathbb{R}^N$ can be transformed into a point $\mathbf{s}^{(k)}$, that can be considered M-sparse. The actual value of M, and therefore the sparseness of $\mathbf{s}$, depends on the actual level of correlation among the collected samples $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \cdots, \mathbf{x}^{(K)}$.

Compressive Sensing (CS) [10]: CS is the technique that we exploit to recover a given N-dimensional signal through the reception of a small number of samples L, which should be ideally much smaller than N.

As above, we consider signals representable through one dimensional vectors $\mathbf{x}^{(k)} \in \mathbb{R}^N$, containing the sensor readings of a WSN with N nodes. We further assume that there exists an invertible transformation matrix $\mathbf{\Psi}$ of size $N \times N$ such that

$$\mathbf{x}^{(k)} = \mathbf{\Psi}\mathbf{s}^{(k)} \tag{2}$$

and that the N-dimensional vector $\mathbf{s}^{(k)}$ is M-sparse. Assuming that $\mathbf{\Psi}$ is known, $\mathbf{x}^{(k)}$ can be recovered from $\mathbf{s}^{(k)}$ by inverting (2), i.e., $\mathbf{s}^{(k)} = \mathbf{\Psi}^{-1}\mathbf{x}^{(k)}$. Also, $\mathbf{s}^{(k)}$ can be obtained through a number L of random projections of $\mathbf{x}^{(k)}$, namely $\mathbf{y}^{(k)} \in \mathbb{R}^L$, with $M \leq L < N$, according to the following equation:

$$\mathbf{y}^{(k)} = \mathbf{\Phi}\mathbf{x}^{(k)} \ . \tag{3}$$

In our framework, $\mathbf{\Phi}$ is referred to as *routing matrix* as it captures the way in which our sensor data is gathered and transmitted to the sink. For the remainder of this paper $\mathbf{\Phi}$ will be considered as an $L \times N$ matrix with a single one in each row and at most a single one in each column (i.e., $\mathbf{y}^{(k)}$ is a sampled version of $\mathbf{x}^{(k)}$).[1] Now, using (2) and (3) we can write

$$\mathbf{y}^{(k)} = \mathbf{\Phi}\mathbf{x}^{(k)} = \mathbf{\Phi}\mathbf{\Psi}\mathbf{s}^{(k)} \stackrel{def}{=} \tilde{\mathbf{\Phi}}\mathbf{s}^{(k)} \ . \tag{4}$$

In general, this system is both ill-posed and ill-conditioned as the number of equations L is smaller than the number of variables N and small variations of the input signal can produce large variations of the output $\mathbf{y}^{(k)}$, respectively. However, if $\mathbf{s}^{(k)}$ is sparse, it has been shown that (4) can be inverted with high probability through the use of special optimization techniques [3], [12]. These allow to retrieve $\mathbf{s}^{(k)}$, whereas the original signal $\mathbf{x}^{(k)}$ is found through (2).

Joint CS and PCA [5]: in [5], our main contribution was the design of a data recovery scheme combining CS and PCA. In this scheme CS is exploited to solve the system in (4) after L data packets have been collected from our WSN testbed and PCA is the technique providing the transformation matrix $\mathbf{\Psi}$. The proposed mathematical framework is detailed in the following.

Let us assume to place the sink in the center of a wireless network with N sensor nodes. We are interested in the reconstruction of the signal at each time k based on our joint CS and PCA scheme. Note that real signals are characterized by spatial and temporal correlations that are in general non-stationary. This means that the statistics that we have to use in our solution (i.e., sample mean and covariance matrix) must be learned at runtime and might not be valid throughout the entire data collection phase. To show the effectiveness of the algorithm in [5] from a theoretical standpoint, we also make the following assumptions:

1. at each time k we have perfect knowledge of the previous K process samples, namely we perfectly know the set $\mathcal{X}^{(k)} = \{\mathbf{x}^{(k-1)}, \mathbf{x}^{(k-2)}, \cdots, \mathbf{x}^{(k-K)}\}$, referred to in what follows as training set;[2]
2. there is a strong temporal correlation between $\mathbf{x}^{(k)}$ and the set $\mathcal{X}^{(k)}$. The size K of the training set is chosen according to the temporal correlation of the observed phenomena to validate this assumption.

Using PCA, from Eq. (1) at each time k we can map our signal $\mathbf{x}^{(k)}$ into a sparse vector $\mathbf{s}^{(k)}$. The matrix $\mathbf{U}$ and the average $\overline{\mathbf{x}}$ can be thought as computed iteratively from the set $\mathcal{X}^{(k)}$, at each time sample k. Accordingly, at time k we indicate matrix $\mathbf{U}$ as $\mathbf{U}^{(k)}$ and we refer to the temporal mean and variance of $\mathcal{X}^{(k)}$ as $\overline{\mathbf{x}}^{(k)}$ and $\widehat{\mathbf{\Sigma}}^{(k)}$, respectively. Hence, we can write:

$$\mathbf{x}^{(k)} - \overline{\mathbf{x}}^{(k)} = \mathbf{U}^{(k)}\mathbf{s}^{(k)} \ . \tag{5}$$

[1] This selection of $\mathbf{\Phi}$ has two advantages: 1) the matrix is orthonormal as required by CS [11] and 2) this type of routing matrix can be obtained through realistic routing schemes.

[2] In [5] we presented a practical scheme that does not need this assumption in order to work.

Now, using equations (3) and (5), we can write:

$$\mathbf{y}^{(k)} - \mathbf{\Phi}\overline{\mathbf{x}}^{(k)} = \mathbf{\Phi}(\mathbf{x}^{(k)} - \overline{\mathbf{x}}^{(k)}) = \mathbf{\Phi}\mathbf{U}^{(k)}\mathbf{s}^{(k)} \ , \qquad (6)$$

whose form is similar to that of (4) with $\tilde{\mathbf{\Phi}} = \mathbf{\Phi}\mathbf{U}^{(k)}$. The original signal $\mathbf{x}^{(k)}$ is approximated as follows: 1) finding a good estimate[3] of $\mathbf{s}^{(k)}$, namely $\widehat{\mathbf{s}}^{(k)}$, using the techniques in [3] or [12] and 2) applying the following calculation:

$$\widehat{\mathbf{x}}^{(k)} = \overline{\mathbf{x}} + \mathbf{U}^{(k)}\widehat{\mathbf{s}}^{(k)} \ . \qquad (7)$$

III. WSN Monitoring via CS: a Bayesian Justification

In this section we justify the effectiveness of our combined CS and PCA technique from a Bayesian perspective. To this end, we refer to the general data modeling framework of [7], [8]. A good review of Bayesian estimation and fitting can also be found in [13]. According to this framework two levels of inference are involved in the data modeling task:

First level of inference. Given a set of plausible models $\{\mathcal{M}_1, \cdots, \mathcal{M}_N\}$ for the observed phenomenon, each of them depending on some set of parameters θ, we fit each model i to the collected data $\mathcal{D}$, i.e., we find the parameter set θ_{MAP} that maximizes the posterior probability density function (pdf)

$$p(\theta|\mathcal{D}, \mathcal{M}_i) = \frac{p(\mathcal{D}|\theta, \mathcal{M}_i)p(\theta|\mathcal{M}_i)}{p(\mathcal{D}|\mathcal{M}_i)} \ , \qquad (8)$$

where $p(\mathcal{D}|\theta, \mathcal{M}_i)$ and $p(\theta|\mathcal{M}_i)$ are known as the *likelihood* and the *prior* respectively, whilst the so called *evidence* $p(\mathcal{D}|\mathcal{M}_i)$ is just a normalization factor which plays a key role in the second level of inference.

Second level of inference. The Bayesian framework allows the comparison of different models and the assignment of preferences among them in the light of data. The most probable model is the one maximizing the posterior $p(\mathcal{M}_i|\mathcal{D}) \propto p(\mathcal{D}|\mathcal{M}_i)p(\mathcal{M}_i)$. Assuming that there are no reasons to assign different priors $p(\mathcal{M}_i)$, the models are ranked according to their evidence. Moreover, the evidence is proportional to the likelihood computed in θ_{MAP}, i.e., $p(\mathcal{D}|\mathcal{M}_i) \propto p(\mathcal{D}|\theta_{MAP}, \mathcal{M}_i)$ (called best fit likelihood, see [13]), where θ_{MAP} is the most probable value of the parameter set according to the observed data, i.e., the argument that maximizes (8). It can be shown, see, e.g., [14], that we can rank different models using the quantity

$$BIC(\mathcal{M}_i) \stackrel{def}{=} \ln p(\mathcal{D}|\theta_{MAP}, \mathcal{M}_i) - \frac{\ell_i}{2}\ln(T) \ , \qquad (9)$$

where ℓ_i is the number of free parameters of model $\mathcal{M}_i$ and T is the cardinality of the observed data set $\mathcal{D}$. Roughly speaking, the Bayesian Information Criterion (BIC) provides insights in the selection of the best fitting model, considering the best fit likelihood and the number of parameters. The BIC penalizes those models requiring more parameters.

Equations (5)–(7) show that the considered framework does not depend on the particular topology considered; the only requirement is that the sensor nodes be ordered (e.g., based on the natural order of their IDs). Our monitoring application can therefore be seen, at each time k, as an interpolation problem: from a sampled M-dimensional vector $\mathbf{y}^{(k)} = \mathbf{\Phi}\mathbf{x}^{(k)} \in \mathbb{R}^M$, we are interested in recovering, via interpolation, the signal $\mathbf{x}^{(k)} \in \mathbb{R}^N$. Typically, this problem can be solved through a linear interpolation on a set $\mathcal{F}$ of h basis functions $\mathbf{f}_i \in \mathbb{R}^N$, i.e., $\mathcal{F} = \{\mathbf{f}_1, \cdots, \mathbf{f}_h\}$. We can assume that the interpolated function has the form:

$$\mathbf{x}^{(k)} = \overline{\mathbf{x}}^{(k)} + \sum_{i=1}^{h} s_i\mathbf{f}_i \ . \qquad (10)$$

A Bayesian approach would estimate the most probable value of $\mathbf{s}$ by maximizing the posterior pdf $p(\mathbf{s}|\mathbf{y}^{(k)}, \mathcal{F}, \mathcal{M})$, where $\mathcal{M}$ is a plausible model for the vector $\mathbf{s} = (s_1, \cdots, s_h)$. As in [13], we assume that $\mathcal{M}$ can be specified by a further parameter set α (called hyper-prior) related to $\mathbf{s}$, so that the posterior can be written as $p(\mathbf{s}|\mathbf{y}^{(k)}, \mathcal{F}, \mathcal{M}) = \int p(\mathbf{s}|\mathbf{y}^{(k)}, \alpha, \mathcal{F}, \mathcal{M})p(\alpha|\mathbf{y}^{(k)}, \mathcal{F}, \mathcal{M})\,\mathrm{d}\alpha$. If the hyper-prior can be inferred from the data and has non zero values $\hat{\alpha}$, maximizing the posterior corresponds to maximizing $p(\mathbf{s}|\mathbf{y}^{(k)}, \hat{\alpha}, \mathcal{F}, \mathcal{M})$, that as shown in [13] corresponds to maximizing the following expression

$$\begin{aligned} p(\mathbf{s}|\mathbf{y}^{(k)}, \mathcal{F}, \mathcal{M}) &\propto p(\mathbf{s}|\mathbf{y}^{(k)}, \hat{\alpha}, \mathcal{F}, \mathcal{M}) \\ &= \frac{p(\mathbf{y}^{(k)}|\mathbf{s}, \mathcal{F})p(\mathbf{s}|\hat{\alpha}, \mathcal{M})}{p(\mathbf{y}^{(k)}|\hat{\alpha}, \mathcal{F}, \mathcal{M})} \ , \end{aligned} \qquad (11)$$

where $p(\mathbf{y}^{(k)}|\mathbf{s})$ is the likelihood function, $p(\mathbf{s}|\hat{\alpha}, \mathcal{M})$ is the prior and $p(\mathbf{y}^{(k)}|\hat{\alpha}, \mathcal{F}, \mathcal{M})$ is a normalization factor. The parameters $\hat{\alpha}$ are estimated maximizing the evidence $p(\mathbf{y}^{(k)}|\alpha, \mathcal{F}, \mathcal{M})$, which is a function of α.

At each time k, our data recovery scheme can be analyzed through the above Bayesian framework thanks to the following associations: the columns of the PCA matrix $\mathbf{U}^{(k)}$ as the set of $h = N$ basis functions, i.e., $\mathcal{F} = \{\mathbf{f}_1, \cdots, \mathbf{f}_N\} = \{\mathbf{u}_1^{(k)}, \cdots, \mathbf{u}_N^{(k)}\} = \mathcal{U}^{(k)}$; the sparse vector $\mathbf{s}^{(k)}$ as the parameter vector $\mathbf{s} = (s_1, \cdots, s_N) = (s_1^{(k)}, \cdots, s_N^{(k)}) = \mathbf{s}^{(k)}$. In this perspective the interpolated function has the form (see Eq. (5))

$$\mathbf{x}^{(k)} - \overline{\mathbf{x}}^{(k)} = \sum_{i=1}^{N} s_i^{(k)}\mathbf{u}_i^{(k)} \ . \qquad (12)$$

Without loss of generality we assume that $\overline{\mathbf{x}}^{(k)} = 0$, thus the constraints on the relationship between $\mathbf{y}^{(k)}$ and $\mathbf{s}^{(k)}$ can be translated into a likelihood of the form:

$$\begin{aligned} p(\mathbf{y}^{(k)}|\mathbf{s}, \mathcal{F}) &= p(\mathbf{y}^{(k)}|\mathbf{s}^{(k)}, \mathcal{U}^{(k)}) \\ &= \delta(\mathbf{y}^{(k)}, \mathbf{\Phi}\mathbf{U}^{(k)}\mathbf{s}^{(k)}) \ , \end{aligned} \qquad (13)$$

where $\delta(x, y)$ is 1 if $x = y$ and zero otherwise. In Section II, we have seen that the vector $\mathbf{s}^{(k)}$ is required to be sparse. In order to guarantee a sparse representation of $\mathbf{s}^{(k)}$, let us consider the Laplacian ($\mathcal{L}_0$) density function, having zero mean $\mu = 0$, which is widely used in the literature [9], [12]

[3] In this paper we refer to a good estimate of $\mathbf{s}^{(k)}$ as $\widehat{\mathbf{s}}^{(k)}$ such that $\|\mathbf{s}^{(k)} - \widehat{\mathbf{s}}^{(k)}\|_2 \leq \epsilon$. Note that by keeping ϵ arbitrarily small, assumption 1 above holds.

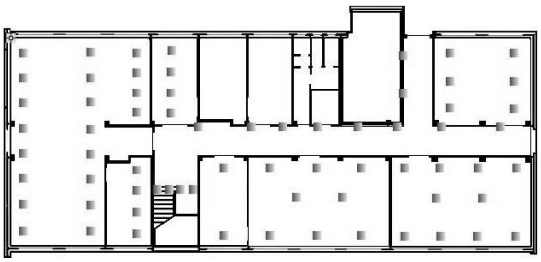

Fig. 1. Layout of the WSN testbed.

to statistically model sparse random vectors. This pdf has the form:

$$p(\mathbf{s}^{(k)}|\hat{\alpha}, \mathcal{M} = \mathcal{L}_0) = \frac{e^{-\hat{\alpha}\sum_{i=1}^{N}|s_i^{(k)}|}}{(2/\hat{\alpha})^N} \ . \tag{14}$$

In this equation, all the components of $\mathbf{s}^{(k)}$ are assumed to be independent and equally distributed. If (11) holds, we can therefore obtain the following posterior:

$$\begin{aligned} p(\mathbf{s}^{(k)}|\mathbf{y}^{(k)}, \mathcal{F}, \mathcal{M}) &= p(\mathbf{s}^{(k)}|\mathbf{y}^{(k)}, \mathcal{U}^{(k)}, \mathcal{L}_0) \\ &\simeq p(\mathbf{s}^{(k)}|\mathbf{y}^{(k)}, \hat{\alpha}, \mathcal{U}^{(k)}, \mathcal{L}_0) \\ &\propto p(\mathbf{y}^{(k)}|\mathbf{s}^{(k)}, \mathcal{U}^{(k)}) p(\mathbf{s}^{(k)}|\hat{\alpha}, \mathcal{L}_0). \end{aligned} \tag{15}$$

Using (13)–(15), maximizing the posterior corresponds to solving the problem

$$\begin{aligned} & \arg\max_{\mathbf{s}^{(k)}} p(\mathbf{s}^{(k)}|\mathbf{y}^{(k)}, \mathcal{U}^{(k)}, \mathcal{L}_0) \\ = \ & \arg\max_{\mathbf{s}^{(k)}} p(\mathbf{y}^{(k)}|\mathbf{s}^{(k)}, \mathcal{U}^{(k)}) p(\mathbf{s}^{(k)}|\hat{\alpha}, \mathcal{U}^{(k)}, \mathcal{L}_0) \\ = \ & \arg\max_{\mathbf{s}^{(k)}} \delta(\mathbf{y}^{(k)}, \mathbf{\Phi}\mathbf{U}^{(k)}\mathbf{s}^{(k)}) \frac{e^{-\hat{\alpha}\sum_{i=1}^{N}|s_i^{(k)}|}}{(2/\hat{\alpha})^N} \\ = \ & \arg\min_{\mathbf{s}^{(k)}} \sum_{i=1}^{N} |s_i^{(k)}|, \text{ given that } \mathbf{y}^{(k)} = \mathbf{\Phi}\mathbf{U}^{(k)}\mathbf{s}^{(k)} \\ = \ & \arg\min_{\mathbf{s}^{(k)}} \|\mathbf{s}^{(k)}\|_1, \text{ given that } \mathbf{y}^{(k)} = \mathbf{\Phi}\mathbf{U}^{(k)}\mathbf{s}^{(k)} \ , \end{aligned} \tag{16}$$

which is actually the problem solved by the CS reconstruction algorithms, see, e.g., [3].

In the next section we will develop a statistical analysis on the principal component distribution of real world signals that validates the use of (14), i.e., that the Laplacian is a good model to represent the principal components of typical WSN data. This provides a justification for using CS in WSNs.

IV. Principal Component Distribution of Real Signals gathered from a WSN

In our experimental campaign, we collected different realizations of five real world signals, we computed the principal components of each and we analyzed the distribution of each principal component independently. According to the notation of Section II, if $\mathbf{x}^{(k)}$ is the sampled signal at time k and $\overline{\mathbf{x}}^{(k)}$ and $\mathbf{U}^{(k)}$ are computed from the set $\mathcal{X}^{(k)}$, we obtain the vector of principal components $\mathbf{s}^{(k)}$ inverting (5):

$$\mathbf{s}^{(k)} = (\mathbf{U}^{(k)})^{-1}(\mathbf{x}^{(k)} - \overline{\mathbf{x}}^{(k)}) = (\mathbf{U}^{(k)})^{T}(\mathbf{x}^{(k)} - \overline{\mathbf{x}}^{(k)}) \ , \tag{17}$$

since $\mathbf{U}^{(k)}$ is orthonormal by construction and therefore $\mathbf{U}^{(k)}(\mathbf{U}^{(k)})^T = I_N$.

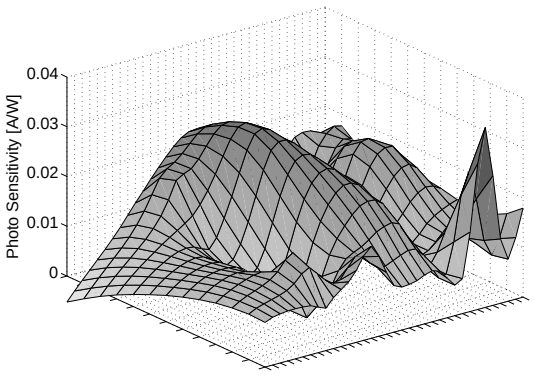

Fig. 2. Signal sample: luminosity in the range $320 - 730$ nm.

In what follows we describe the considered signals and the WSN deployment. After that, we present the analysis on the distribution of the elements of $\mathbf{s}$.

Network: we consider the WSN testbed of Fig. 1. This experimental network is deployed on the ground floor of the Department of Information Engineering at the University of Padova. The WSN consists of $N = 68$ TmoteSky wireless nodes equipped with IEEE 802.15.4 compliant radio transceivers.

Signals: From the above WSN, we gathered five different types of signals $\mathbf{x}$: S1) humidity, S2-S3) luminosity in two different ranges: $320-730$ and $320-1100$ nm, respectively, S4) temperature and S5) battery voltage. We collect measurements from all nodes every 5 minutes for at least 3 days. We repeated the data collection for three different measurement campaigns during the month of March 2009, choosing different days of the week: C1) from the 13^{th} to the 16^{th} of March, C2) from the 19^{th} to the 23^{th} and C3) from the 24^{th} to the 27^{th}. Fig. 2 shows an example signal of type S2, i.e., luminosity in the range $320 - 730$ nm.

Principal Component Analysis of Real World Signals: our aim is to infer the statistical distribution of the vector random process $\mathbf{s}^{(k)}$ from the samples $\{\mathbf{s}^{(1)}, \mathbf{s}^{(2)}, \ldots, \mathbf{s}^{(T)}\}$ which are obtained from the above WSN signals. The parameter T is the duration (number of time samples) of each monitoring campaign C1–C3.

From the theory [6] we know that signals in the PCA domain (in our case $\mathbf{s}^{(k)}$) have in general uncorrelated components. Also, in our particular case we experimentally verified that this assumption is good since $E[s_i s_j] \simeq E[s_i]E[s_j]$ for $i, j \in \{1, \ldots, N\}$ and $i \neq j$. In our analysis, we make a stronger assumption, i.e., we build our model of $\mathbf{s}$ considering statistical independence among its components,

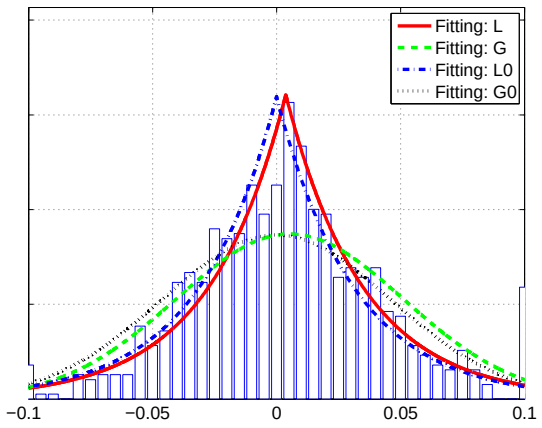

Fig. 3. Empirical distribution and model fitting for a principal component of signal S1, humidity.

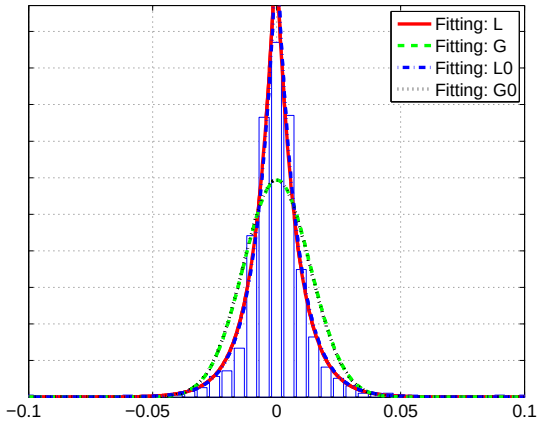

Fig. 4. Empirical distribution and model fitting for a principal component of signal S2, luminosity in the range $320-730$ nm.

i.e., $p(s_i, s_j) = p(s_i)p(s_j)$ with $i \neq j$, which allows us to consider the prior (14). A further assumption that we make is to consider the components of $\mathbf{s}$ as stationary over the entire monitoring period[4]. The model developed following this approach leads to good results [5], which allow us to validate these assumptions.

Owing to these assumptions, the problem of statistically characterizing $\mathbf{s}$ reduces to that of characterizing the random variables

$$s_i = \sum_{j=1}^{N} u_{ji}(x_j - \overline{x}_j) \ , \ i = 1, \ldots, N \ , \qquad (18)$$

where the r.v. u_{ji} is an element of matrix $\mathbf{U}$ and the r.v. x_j is an element of vector $\mathbf{x}$.

A statistical model for each s_i can be determined through a Bayesian approach (see [13]). For the first level of inference, the observation of the experimental data gives empirical evidence for the selection of four statistical models:

- $\mathcal{M}_1$ a Laplacian distribution with parameters $\{\mu, \lambda\}$, that we call $\mathcal{L}$;
- $\mathcal{M}_2$ a Gaussian distribution with $\{m, \sigma^2\}$, that we call $\mathcal{G}$;
- $\mathcal{M}_3$ a Laplacian distribution with $\mu = 0$ and λ, that we call $\mathcal{L}_0$;
- $\mathcal{M}_4$ a Gaussian distribution with $m = 0$ and σ^2, that we call $\mathcal{G}_0$.

The space of models for each $\mathbf{s}_i$ is therefore described by the set $\{\mathcal{L}, \mathcal{G}, \mathcal{L}_0, \mathcal{G}_0\}$. For each signal $S1 - S5$, for each component $s_i, i = 1, \ldots, N$, and for each model $\mathcal{M}_i, i = 1, \ldots, 4$, we have estimated the parameters (i.e., the most probable *a posteriori*, MAP) that best fit the data according to (8). Since we deal with Gaussian and Laplacian distributions, these estimations have well known and closed form solutions [8]. In detail:

- $\mathcal{M}_1$ $\hat{\mu} = \mu_{1/2}(\mathbf{s})$ and $\hat{\lambda} = \frac{\sum_{j=1}^{T} |s_j - \hat{\mu}|}{T}$, where $\mu_{1/2}(\mathbf{s})$ is the median of the data;
- $\mathcal{M}_2$ $\hat{m} = \frac{\sum_{j=1}^{T} s_j}{T}$ and $\hat{\sigma}^2 = \frac{\sum_{j=1}^{T} (s_j - \hat{m})^2}{T-1}$;
- $\mathcal{M}_3$ $\hat{\lambda} = \frac{\sum_{j=1}^{T} |s_j|}{T}$;
- $\mathcal{M}_4$ $\hat{\sigma}^2 = \frac{\sum_{j=1}^{T} s_j^2}{T}$.

Figs. 3–4 show two examples of data fitting according to the aforementioned models; in these figures we plot the empirical distribution and the corresponding inferred statistical model for a generic principal component of the humidity (S2) and the luminosity (S4), respectively. From these graphs it is already clear that the distribution of the principal components of our signals is well described by a Laplacian distribution. For the second level of inference we ranked each model according to the Bayesian Information Criterion (BIC) (see Eq. (9)). Fig. 5 shows the BIC for the humidity signal of campaign C1 for all principal components and for all the considered models. From this figure we see that the Laplacian models better fit the data

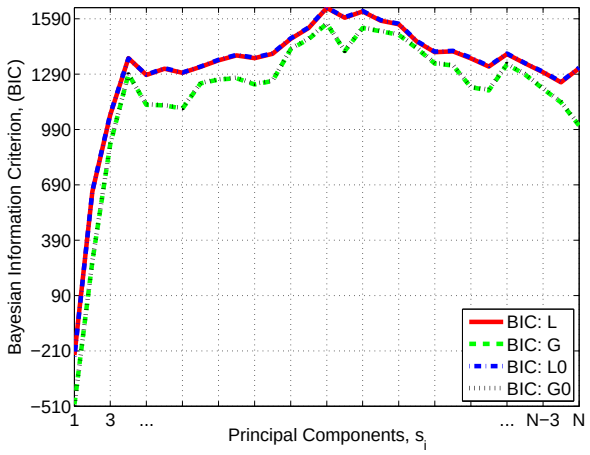

Fig. 5. Bayesian Information Criterion (BIC) per Principal Component, for each model $\mathcal{M}_1$–$\mathcal{M}_4$, campaign C1 and signal S1, humidity.

[4]Note that this assumption does not imply that also the observed process $\mathbf{x}$ is assumed to be stationary. The basis $\mathcal{U}^{(k)}$ does not represent a fixed linear transformation between $\mathbf{x}^{(k)}$ and $\mathbf{s}^{(k)}$, but it changes at each time sample k according to the statistics of $\mathbf{x}$.

Campaign C1					
	S1	S2	S3	S4	S5
$\mathcal{L}$	1320.2	2481.9	2186.3	1708.5	5374.0
$\mathcal{G}$	1191.7	2226.8	1813.0	1540.1	4139.2
$\mathcal{L}_0$	**1322.6**	**2485.0**	**2189.3**	**1711.4**	**5377.3**
$\mathcal{G}_0$	1194.1	2229.8	1815.9	1542.7	4141.5
Campaign C2					
	S1	S2	S3	S4	S5
$\mathcal{L}$	921.2	2740.8	2065.4	1483.4	6094.0
$\mathcal{G}$	463.2	1727.7	815.6	749.0	5152.4
$\mathcal{L}_0$	**924.3**	**2744.0**	**2068.8**	**1486.1**	**6097.5**
$\mathcal{G}_0$	466.2	1730.7	818.4	751.9	5155.3
Campaign C3					
	S1	S2	S3	S4	S5
$\mathcal{L}$	430.3	1207.4	851.3	773.9	3239.7
$\mathcal{G}$	272.9	737.0	301.1	585.8	2676.7
$\mathcal{L}_0$	**432.7**	**1210.4**	**854.4**	**776.5**	**3242.9**
$\mathcal{G}_0$	275.5	739.8	303.8	588.6	2679.3

TABLE I
BAYESIAN INFORMATION CRITERION (BIC) AVERAGED OVER ALL PRINCIPAL COMPONENTS, FOR EACH MODEL $\mathcal{M}_1$–$\mathcal{M}_4$, EXPERIMENTAL CAMPAIGNS C1–C3 AND SIGNALS S1–S5.

for all principal components $s_i, i = 1, 2, \ldots, N$. The average BIC for each model, for the different signals and the three campaigns C1–C3, is shown in Table I. The values of this table are computed averaging over the N principal components. From these results we see that model $\mathcal{L}_0$ provides the best statistical description of the experimental data. In fact, the BIC metric is higher for Laplacian models in all cases; furthermore, $\mathcal{L}_0$ has a higher evidence with respect to $\mathcal{L}$, since it implies the utilization of a single parameter. As previously mentioned, the over-parameterization of the model is penalized according to the factor $T^{\frac{-\ell}{2}}$ (see Eq. (9)). Based on the above results we can make the following observations:

1 In our monitoring framework which jointly exploits CS and PCA (Section II), the principal components of real world signals are Laplacian distributed with good approximation, therefore it is legitimate to use the prior (14);

2 it is possible to determine, for each principal component, an optimal parameter $\hat{\lambda}$ that differs from zero, and therefore we can exploit (15) with $\hat{\alpha} \leftarrow \hat{\lambda}$;

2.1 in case all principal components are equally distributed with parameter $\hat{\lambda}$, we have that, as demonstrated in Section III, CS obtains the best recovery performance, i.e., at each time k, it finds the $\mathbf{s}^{(k)}$ that maximizes the posterior (15);

2.2 in case the principal components are Laplacian distributed with different parameters and these can all be estimated from the data, following a rationale similar to that of Section III we can say that the best recovery performance can be obtained using CS reweighted [15]. We observe that this is often the case in practice, as confirmed by our empirical measurements. We shall note, however, that the correct online estimation of the different parameters is not straightforward. Nevertheless, standard CS still provides very good reconstruction performance as shown in [5].

V. CONCLUSIONS

In this paper we investigated the effectiveness of data recovery through joint Compressive Sensing (CS) and Principal Component Analysis (PCA) in Wireless Sensor Networks (WSNs). At first, we framed our recovery scheme into the context of Bayesian theory proving that, under certain assumptions on the signal statistics, the use of CS is legitimate, and is in fact optimal in terms of recovery performance. Hence, as the main contribution of the paper we have shown that these assumptions hold for real world data, which we gathered from an actual WSN deployment and processed according to our monitoring framework. This allows us to conclude that the use of CS not only is legitimate in our recovery scheme but also makes it possible to obtain very good performance for the considered data sets.

REFERENCES

[1] D. Donoho, "Compressed sensing," *IEEE Trans. on Information Theory*, vol. 52, no. 4, pp. 4036–4048, Apr. 2006.

[2] E. Candès and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. on Information Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[3] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. on Information Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[4] J. Haupt, W. Bajwa, M. Rabbat, and R. Nowak, "Compressive Sensing for Networked Data: a Different Approach to Decentralized Compression," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 92–101, Mar. 2008.

[5] R. Masiero, G. Quer, D. Munaretto, M. Rossi, J. Widmer, and M. Zorzi, "Data Acquisition through joint Compressive Sensing and Principal Component Analysis," in *IEEE Globecom 2009*, Honolulu, Hawaii, USA, Nov.-Dec. 2009.

[6] C. R. Rao, "The Use and Interpretation of Principal Component Analysis in Applied Research," *Sankhya: The Indian Journal of Statistics*, vol. 26, pp. 329–358, 1964.

[7] S. Gull, "The Use and Interpretation of Principal Component Analysis in Applied Research," *Maximum Entropy and Bayesian Methods in Science and Engineering*, vol. 1, pp. 53–74, 1988.

[8] J. Skilling, *Maximum Entropy and Bayesian Methods*. Kluwer academic, 1989.

[9] S. Ji, Y. Xue, and L. Carin, "Bayesian Compressive Sensing," *IEEE Trans. on Signal Processing*, vol. 56, no. 6, pp. 2346–2356, Jun. 2008.

[10] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," *IEEE Trans. on Information Theory*, vol. 52, no. 9, pp. 4036–4048, Sep. 2006.

[11] G. Quer, R. Masiero, D. Munaretto, M. Rossi, J. Widmer, and M. Zorzi, "On the Interplay Between Routing and Signal Representation for Compressive Sensing in Wireless Sensor Networks," in *Information Theory and Applications Workshop (ITA 2009)*, San Diego, CA, US, Feb. 2009.

[12] H. Mohimani, M. Babaie-Zadeh, and C. Jutten, "A fast approach for overcomplete sparse decomposition based on smoothed L0 norm," *IEEE Trans. on Signal Processing*, 2009, Accepted for publication.

[13] D. J. MacKay, "Bayesian Interpolation," *Neural Computation Journal*, vol. 4, no. 3, pp. 415–447, May 1992.

[14] G. Schwarz, "Estimating the Dimension of a Model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.

[15] E. Candès, M. B. Wakin, and S. Boyd, "Enhancing Sparsity by Reweighted ℓ_1 Minimization," *Journal of Fourier Analysis and Applications*, vol. 14, pp. 877–905, Dec. 2008.