# Sensing, Compression and Recovery for Wireless Sensor Networks: Sparse Signal Modelling

**Riccardo Masiero, Giorgio Quer, Gianluigi Pillonetto,**

**Michele Rossi, Michele Zorzi**

Department of Information Engineering, University of Padova

6/B, Via G. Gradenigo, I–35131 Padova (PD), Italy

E-mail: {masieror, quergior, giapi, rossi, zorzi} @dei.unipd.it

**Abstract**

In this paper, we propose a sparsity model that allows the use of Compressive Sensing (CS) for the online recovery of large data sets in real Wireless Sensor Network (WSN) scenarios. We advocate the joint use of CS for the recovery and of Principal Component Analysis (PCA) to capture the spatial and temporal characteristics of real signals. The statistical characteristics of the signals are thus exploited to design the sparsification matrix required by CS recovery. In this paper, we represent this framework through a Bayesian Network (BN) and we use Bayesian analysis to infer and approximate the statistical distribution of the principal components. We show that the Laplacian distribution provides an accurate representation of the statistics of the data measured from real WSN testbeds. Therefore, the joint use of CS and PCA for data recovery in real WSNs is legitimate, and is equivalent to Maximum A Posteriori (MAP) recovery.

**Index Terms**

Compressive Sensing, sparse signal, real signal, WSN, bayesian analysis, bayesian network.

Part of this work has been presented at the IEEE International Workshop on Scalable Ad Hoc and Sensor Networks Saint Petersburg, Russia, Oct. 2009.

# I. Introduction

Compressive Sensing (CS) is a convex optimization technique that has received great attention by the signal and image processing community in the past few years [1]–[3]. CS was originally developed for the efficient storage and compression of digital images, which show high spatial correlation. In the very recent literature, the Bayesian approach has been used to develop efficient and auto-tunable algorithms for CS, see [4]. However, previous work addressing CS from a Bayesian perspective mainly focused on proving fundamental results and on understanding its usefulness in the image processing field. In particular, in [5] a hierarchical Bayesian model is considered to utilize CS for the reconstruction of sparse images when the observations are obtained from linear transformations and corrupted by additive and white Gaussian noise. In [6], the authors model the components of the CS problem using a Bayesian framework to recover synthetic 1-D sparse signals and simple images with high spatial correlation.

Since the pioneering work of Nowak [7], [8], there has been a growing interest in this technique also in the networking community. Specifically, the great interest around the use of CS in Wireless Sensor Networks (WSNs) comes from the fact that the CS framework lends itself to the accurate reconstruction of large sensor fields through the collection of a small fraction of the sensor readings. However, the application of CS to data gathered from actual WSN deployments faces several problems. In particular, we can not assume that the distributed signal before compression is sparse, or equivalently, we can not assume a Laplacian prior for each element of the signal. In this case, CS can not be successfully used without accounting for a technique that effectively sparsifies the data. Additionally, and in contrast to classical approaches, where the data is first compressed and then transmitted to a given Data Collection Point (DCP), when CS is applied to WSNs it is desirable to jointly compress and transmit the data. In this way, we reduce the amount of information that is sent to the DCP, with a subsequent reduction of the energy consumed by the sensor nodes. A preliminary work with this aim is [9], where the coherence between the routing matrix and the sparsification matrix is studied to exploit CS as a reconstruction technique for WSNs. The results of this work are anyway unsatisfactory, due to the fact that only exploiting the spatial correlation of the data does not suffice to efficiently recover the signal.

In this work we address the issue of designing a technique based on CS for the online recovery of large data sets through the collection of a small number of sensor readings. Since only exploiting the spatial correlation does not suffice, for an efficient recovery we also jointly exploit the temporal correlation of the signal. In particular, through the use of Principal Component Analysis (PCA), which extracts the spatial and temporal correlation characteristics from the past recovered signal samples, we learn at the DCP the relevant statistics for CS. Thus, we exploit the learned statistics to recover the signal through CS from its sampled version that is received at the DCP. We analyze the joint use of CS and PCA with a Bayesian framework, depicting the probabilistic relations among all the variables involved in the compression, transmission and recovery process through a Bayesian Network (BN). A BN is a probabilistic graphical model [10], a very useful tool to graphically capture the probabilistic relations among a set of random variables and to study the conditional probabilities within this set. This tool has been used for networking applications in recent papers on the optimization of single-hop [11] as well as multi-hop [12] Wireless Local Area Networks (WLAN).

In order to substantiate our framework, we consider different WSN testbeds, whose data is available on-line. We analyze the statistics of the principal components of the signals gathered by these WSNs, designing a Bayesian model to approximate the statistical distribution of the principal components. An overview on the use of Bayesian theory to define a general framework for data modeling can be found in [13], [14]. Hence, we provide empirical evidence of the effectiveness of CS in actual WSNs, showing that the principal components of real signals are well approximated by a Laplacian distribution. Moreover, we identify the conditions under which the recovery through convex optimization (i.e., CS) is equivalent to Maximum A Posteriori (MAP) recovery, which occurs when each component is distributed according to a statistically independent Laplacian distribution. We conclude the paper showing example results for CS recovery when used in conjunction with a simple data gathering technique. The integration of the proposed mathematical framework into an actual Data Collection and Recovery technique for WSN signals, with an extensive performance analysis, is presented in the companion paper [15], where we compare this technique with other data collection methods that exploit the same information learned by PCA, as well as with standard data recovery schemes.

The rest of the paper is structured as follows: in Section II we present the mathematical details for the joint use of CS and PCA, and in Section III we propose the framework to exploit these two techniques and the signal model to describe the statistics of the principal components. Thus, we analyze a large number of WSN testbeds and signals in Section IV. In Section V we present the details of the probabilistic model, that exploits a two-level Bayesian inference to infer the best fitting distribution for such signals, and in Section VI we identify the optimality conditions for CS recovery, showing some example results. Section VII concludes the paper.

## II. Mathematical tools for CS recovery

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

### A. Principal Component Analysis

The Karhunen-Loève expansion is the theoretical basis for PCA. It is a method to represent through the best $M$-term 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 priori, the Karhunen-Loève expansion can be approximated thanks to PCA [16], 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

$\mathbf{U}$. Therefore, let us define $\mathbf{s}^{(k)} \stackrel{def}{=} \mathbf{U}^T(\mathbf{x}^{(k)} - \overline{\mathbf{x}})$. If the instances $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \cdots, \mathbf{x}^{(K)}$ of the process $\mathbf{x}$ are temporally correlated, then only a fraction of the elements of $\mathbf{s}^{(k)}$ can be sufficient to collect the overall energy of $\mathbf{x}^{(k)} - \overline{\mathbf{x}}$. In other words, 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 $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \cdots, \mathbf{x}^{(K)}$.

### *B. Compressive Sensing (CS)*

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 $\mathbf{s}^{(k)} \in \mathbb{R}^N$ is $M$-sparse. $\mathbf{s}^{(k)}$ is said to be $M$-sparse when it has only $M$ significant components, while the other $N - M$ are negligible with respect to the average energy per component, defined[1] as $E_s^{(k)} = \frac{1}{N}\sqrt{\langle \mathbf{s}^{(k)}, \mathbf{s}^{(k)} \rangle}$. Assuming that $\mathbf{\Psi}$ is known, $\mathbf{x}^{(k)}$ can be recovered from

[1]For any two column vectors $\mathbf{a}$ and $\mathbf{b}$ of the same length, we define $\langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a}^T\mathbf{b}$.

$\mathbf{s}^{(k)}$ 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 DCP. For the remainder of this paper $\mathbf{\Phi}$ will be considered as an $L \times N$ matrix of all zeros except for 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 $\mathbf{x}^{(k)}$).[2] Now, using (2) and (3) we can write

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

This system is ill-posed since the number of equations $L$ is smaller than the number of variables $N$. It may also be ill-conditioned, i.e., a small variation of the output $\mathbf{y}^{(k)}$ can produce a large variation of the input signal [17], [18]. However, if $\mathbf{s}^{(k)}$ is sparse, it has been shown that (4) can be inverted with high probability through the use of specialized optimization techniques [3], [19]. These allow to retrieve $\mathbf{s}^{(k)}$, from which the original signal $\mathbf{x}^{(k)}$ is found through (2).

### *C. Joint CS and PCA*

We have seen that PCA is a method to represent through the best $M$-term approximation a generic $N$-dimensional signal, where $N > M$, and we have introduced CS, a technique to recover an $N$-dimensional signal through the reception of a small number of samples $L$, with $L < N$. In this section we propose a technique that jointly exploits PCA and CS to reconstruct a signal $\mathbf{x}^{(k)}$ at each time $k$. Assume that the signal is correlated both in time and in space, but that in general it is 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 time frame in which we want to reconstruct the signal. We should also make the following assumptions, that will be justified in the next sections:

[2]This selection of $\mathbf{\Phi}$ has two advantages: 1) the matrix is orthonormal as required by CS, see [9], and 2) this type of routing matrix can be obtained through realistic routing schemes.

1. at each time $k$ we have perfect knowledge of the previous $K$ process samples, namely we perfectly know the set $\mathcal{X}^{(k)} = \left\{\mathbf{x}^{(k-1)}, \mathbf{x}^{(k-2)}, \cdots, \mathbf{x}^{(k-K)}\right\}$, referred to in what follows as training set;[3]
2. there is a strong temporal correlation between $\mathbf{x}^{(k)}$ and the set $\mathcal{X}^{(k)}$ that will be explicated in the next session via a Bayesian network. 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 of 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 covariance of $\mathcal{X}^{(k)}$ as $\overline{\mathbf{x}}^{(k)}$ and $\widehat{\mathbf{\Sigma}}^{(k)}$, respectively. Hence, we can write:

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

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

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

where with the symbol $\mathbf{\Phi}^{(k)}$ we make explicit that also the routing matrix $\mathbf{\Phi}$ can change over time. The form of Eq. (6) is similar to that of (4) with $\tilde{\mathbf{\Phi}} = \mathbf{\Phi}^{(k)}\mathbf{U}^{(k)}$. The original signal $\mathbf{x}^{(k)}$ is approximated as follows: 1) finding a good estimate[4] of $\mathbf{s}^{(k)}$, namely $\widehat{\mathbf{s}}^{(k)}$, using the techniques in [3] or [19] and 2) applying the following calculation:

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

## III. Monitoring Framework and Sparse Signal Model

In this section we describe a model to represent a broad range of environmental signals that can be gathered from a WSN. The aim is to analyze the stochastic properties of these signals, in order to select the most appropriate sampling, compression and recovery techniques to minimize the number of transmitting nodes while keeping a certain level of reconstruction accuracy, as detailed in Section VI.

[3]In [20] we presented a practical scheme that does not need this assumption in order to work.

[4]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 $\epsilon$ arbitrarily small, assumption 1 above is very accurate.

We have chosen to represent the variables involved with a Bayesian Network (BN) [10], i.e., a Directed Acyclic Graph (DAG) where nodes represent random variables and arrows represent conditional dependencies among them. From the DAG it is always possible to determine the conditional independence between two variables, applying a set of rules known as *d-separation* rules, e.g., see [21] for a detailed description about BN properties. In this section, we propose two graphical models which illustrate the perspective we adopted in Section V and Section VI, respectively:

1) Fig. 1 represents a stochastic model for the signal $\mathbf{s}$;
2) Fig. 2 is a BN which links together all the variables involved in our analysis, highlighting those required to define the monitoring framework.

In detail, with Fig. 1 we introduce a Bayesian model to describe the statistical properties of the elements of $\mathbf{s}^{(k)}$. Given the realizations of the signal $\mathbf{s}^{(k)}$ at time $k = 1, \ldots, K$, we use a Bayesian estimation method, described in Sec. V, to infer a suitable model $\mathcal{M}$ along with the best-fitting values of its parameters. In particular, for a Gaussian model the parameters to infer are the mean value $m$ of each component and the standard deviation $\sigma$, whereas for a Laplacian model are the location parameter $\mu$ and the scale parameter $\lambda$, respectively. This modeling approach is exploited in Sec. V to determine which stochastic model, chosen among a set of plausible ones, better describes the signal $\mathbf{s}^{(k)}$.

Fig. 2, instead, depicts the whole considered framework that involves the following variables for each time sample $k$: the training set $\mathcal{X}^{(k)}$, the WSN signal $\mathbf{x}^{(k)}$, its compressed version $\mathbf{y}^{(k)}$, obtained sampling $\mathbf{x}^{(k)}$ according to matrix $\mathbf{\Phi}^{(k)}$ as in Eq. (3), the invertible matrix $\mathbf{U}^{(k)}$, obtained through PCA, and the sparse representation $\mathbf{s}^{(k)}$, introduced in Eq. (1). From the results presented in Section V, it turns out that $\mathbf{s}^{(k)}$ is well approximated by a Laplacian distribution. Analyzing the DAG in Fig. 2, based on the *d-separation* rules, we can make the following observations:

- **data gathering**: the WSN signal $\mathbf{x}^{(k)}$ is independent of the stochastic sampling matrix $\mathbf{\Phi}^{(k)}$, whose nature is described in [20], but the observation of $\mathbf{y}^{(k)}$ reveals a link between these two variables;
- **PCA transformation**: this is the core of our model, that describes how the system learns the statistics of the signal of interest $\mathbf{x}^{(k)}$. According to the dynamic system framework, $\mathbf{U}^{(k)}$ can be

seen as the state of a system, since it summarizes at each instant $k$ all the past history of the system, represented by the matrix $\mathcal{X}^{(k)}$. The system input is the signal $\mathbf{s}^{(k)}$, that can be seen as a Laplacian or Gaussian innovation process. This type of priors on the signal induces estimators that use, respectively, the $L_1$ and $L_2$ norm of the signal as regularization terms. Hence, such priors are often used in the literature in view of the connection with powerful shrinkage methods such as ridge regression and LASSO, as well as for the many important features characterizing them, see Section 3.4 in [17] for a thorough discussion. Note also that the observation of the WSN signal $\mathbf{x}^{(k)}$ has a twofold effect: the former is the creation of a deterministic dependence between the PCA basis $\mathbf{U}^{(k)}$ and the sparse signal $\mathbf{s}^{(k)}$, that are otherwise independent; the latter is the separation of $\mathbf{U}^{(k)}$ and $\mathbf{s}^{(k)}$ from the data gathering variables, i.e., they become independent of $\mathbf{y}^{(k)}$ and $\mathbf{\Phi}^{(k)}$;

- **sparse signal model**: we observe that the priors assigned to the variable $\mathcal{M}$ and to the corresponding parameters $\mu$ (resp. $m$) and $\lambda$ (resp. $\sigma$) are non informative, except for the non-negativity of the variance. Here the observation of the sparse signal $\mathbf{s}^{(k)}$ separates the sparse signal model from the monitoring framework, i.e., after observing the signal $\mathbf{s}^{(k)}$, the variable $\mathcal{M}$ and the corresponding parameters $\mu$ (resp. $m$) and $\lambda$ (resp. $\sigma$) will no longer be dependent on the variables of the monitoring framework, so they can be analyzed separately as we do in Section V.

In the next section we will describe the real world signals that will be used to develop a statistical analysis on the principal component distribution and from which we obtain a set of realizations for the signal $\mathbf{s}^{(k)}$. In Sec. V we will show that the Laplacian is a good model to represent the principal components of typical WSN data. In turn, this provides a justification for using CS in WSNs, as detailed in Section VI.

## IV. Description of Considered Signals and WSNs

The ultimate aim of WSN deployments is to monitor the evolution of certain physical phenomena over time. Examples of applications that require such infrastructure include monitoring for security, health-care or scientific purposes. Many different types of signals can be sensed, processed and stored, e.g., the motion of objects and beings, the heart beats, or environmental signals like the values of

temperature and humidity, indoor or outdoor. Very often the density of sensor network deployments is very high and therefore sensor observations are strongly correlated in space. Furthermore, the physics itself of the observed signals makes consecutive observations of a sensor node to be also temporally correlated.

The spatial and temporal correlation represents a huge potential that can be exploited in the design of collaborative protocols for WSNs. In this perspective, we can think of reducing the energy consumption of the network by tuning the overall number of transmissions required to monitor the evolution of a given phenomenon over time. The appeal of the techniques presented in Section II follows from the fact that CS enables us to significantly reduce the number of samples needed to estimate a signal of interest with a certain level of quality. Clearly, the effectiveness of CS is subject to the knowledge of a transformation basis for which the observed signals result sparse.

In this section we illustrate the WSNs and the gathered signals that will be used in Section V to test, using the Bayesian framework presented in Section III, whether CS and PCA are effective for real signals, i.e., whether a real signal transformed by the PCA matrix is actually sparse.

**Networks.** In addition to our own experimental network deployed on the ground floor of the Department of Information Engineering at the University of Padova, we consider other three WSNs whose sensor reading databases are available on-line, and a further deployment called Sense&Sensitivity, whose data has been kindly provided to the authors by Dr. Thomas Watteyne of Dust Networks, Inc. A brief technical overview of each of these five experimental network scenarios follows.

**T1** WSN testbed of the Department of Information Engineering (DEI) at the University of Padova, collecting data from $68$ TmoteSky wireless sensor nodes [22], [23]. The node hardware features an IEEE 802.15.4 Chipcon wireless transceiver working at $2.4$ GHz and allowing a maximum data rate of $250$ Kbps. These sensors have a TI MSP430 micro-controller with $10$ Kbytes of RAM and $48$ Kbytes of internal FLASH;

**T2** LUCE (Lausanne Urban Canopy Experiment) WSN testbed at the Ecole Polytechnique Fédérale de Lausanne (EPFL), [24]. This measurement system exploits $100$ SensorScope weather sensors

which have been deployed across the EPFL campus. The node hardware is based on a TinyNode module equipped with a Xemics XE1205 radio transceiver operating in the 433, 868 and 915 MHz license-free ISM (Industry Scientific and Medical) frequency bands. Also these sensors have a TI MSP430 micro-controller;

**T3** St-Bernard WSN testbed at EPFL, [25]. This experimental WSN deployment is made of 23 SensorScope stations deployed at the Grand St. Bernard pass at 2400 m, between Switzerland and Italy. See point T2 for a brief description of the related hardware;

**T4** CitySense WSN testbed, developed by Harvard University and BBN Technologies, [26]. CitySense is an urban scale deployment that will consist of 100 wireless sensor nodes equipped with an ALIX 2d2 single-board computer. The transmitting interface is reconfigurable by the user and by default it operates in 802.11b/g ad hoc mode at 2.4 GHz. Nowadays this WSN deployment counts about twenty nodes;

**T5** The Sense&Sensitivity [27] testbed is a WSN of 86 nodes, which embed Texas Instrument Inc. technology: a MSP430 micro-controller and a CC1100 radio chip operating in the ISM band (from 315 to 915 MHz).

**Signals.** From the above WSNs, we gathered seven different types of signals: **S1)** temperature; **S2)** humidity; **S3-S4)** luminosity in two different ranges ($320-730$ and $320-1100$ nm, respectively); **S5)** wind direction; **S6)** voltage and **S7)** current. Concerning the signals gathered from our testbed T1, we collected measurements from all nodes every 5 minutes for 3 days. We repeated the data collection for three different measurement campaigns, choosing different days of the week. Regarding the data collection from WSNs T2–T5, we studied the raw data available on-line with the aim of identifying a portion of data that could be used as a suitable benchmark for our research purposes. This task has turned out to be really challenging due to packet losses, device failures and battery consumption that are very common and frequent with currently available technology. For the acquisition of the signals we divided the time axis in frames (or time slots) such that each of the working nodes was able to produce a new sensed data per frame. Details of the signals extracted from the records of T1–T5, and

organized in different campaigns, are reported schematically in Table I.

## V. Sparsity Analysis of Real Signal Principal Components

In this section we infer the statistical distribution of the vector random process $\mathbf{s}$ 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 in Table I.

From the theory [16] we know that signals in the PCA domain (in our case $\mathbf{s}$) 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, i.e., $p(s_1, \ldots, s_N) = \prod_{i=1}^{N} p(s_i)$. A further assumption that we make is to consider the components of $\mathbf{s}$ as stationary over the entire monitoring period.[5] The model developed following this approach leads to good results [20], 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

$$s_i = \sum_{j=1}^{N} u_{ji}(x_j - \overline{x}_j) \ , \ i = 1, \ldots, N \ , \tag{8}$$

where the r.v. $u_{ji}$ is an element of matrix $\mathbf{U}$ in Eq. (5) and the r.v. $x_j$ is an element of vector $\mathbf{x}$.

A statistical model for each $s_i$ can be determined through the Bayesian estimation procedure detailed below. Similarly to the approach adopted in [28], we rely upon two levels of inference.

**First level of inference.** Given a set of competitive models $\{\mathcal{M}_1, \cdots, \mathcal{M}_N\}$ for the observed phenomenon, each of them depending on the parameter vector $\boldsymbol{\theta}$, we fit each model to the collected data denoted by $\mathcal{D}$, i.e., we find the $\boldsymbol{\theta}_{\text{MAP}}$ that maximizes the a posteriori probability density function (pdf)

$$p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{M}_i) = \frac{p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{M}_i)p(\boldsymbol{\theta}|\mathcal{M}_i)}{p(\mathcal{D}|\mathcal{M}_i)} \ , \tag{9}$$

[5]Note that this model is also able to follow signals whose frequency content varies over time since the signal basis adapts to the data.

i.e.,

$$\boldsymbol{\theta}_{\text{MAP}} = \underset{\theta}{\operatorname{argmax}}\, p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{M}_i) \ , \tag{10}$$

where $p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{M}_i)$ and $p(\boldsymbol{\theta}|\mathcal{M}_i)$ 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.** According to Bayesian theory, the most probable model is the one maximizing the posterior $p(\mathcal{M}_i|\mathcal{D}) \propto p(\mathcal{D}|\mathcal{M}_i)p(\mathcal{M}_i)$. Hence, when the models $\mathcal{M}_i$ are equiprobable, they are ranked according to their evidence. In general, evaluating the evidence involves the computation of analytically intractable integrals. For this reason, we rank the different models according to a widely used approximation, the Bayesian Information Criterion (BIC) [29], that we define as:

$$\text{BIC}(\mathcal{M}_i) \stackrel{def}{=} \ln\left[p(\mathcal{D}|\boldsymbol{\theta}_{MAP}, \mathcal{M}_i)p(\boldsymbol{\theta}_{MAP}|\mathcal{M}_i)\right] - \frac{\ell_i}{2}\ln(T) \ , \tag{11}$$

where $\boldsymbol{\theta}_{MAP}$ is defined in (10), $\ell_i$ 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 insight in the selection of the best fitting model penalizing those models requiring more parameters.

According to the introduced formalism we consider $\{\mathbf{s}^{(1)}, \mathbf{s}^{(2)}, \ldots, \mathbf{s}^{(T)}\}$ as the set of collected data $\mathcal{D}$; further, the observation of the experimental data gives empirical evidence for the selection of four statistical models $\mathcal{M}_i$ and corresponding parameter vectors $\boldsymbol{\theta}$:

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

The space of models for each $\mathbf{s}_i$ is therefore described by the set $\{\mathcal{L}, \mathcal{G}, \mathcal{L}_0, \mathcal{G}_0\}$. In detail, for each signal $S1 - S7$ in the corresponding WSNs and campaigns of Table I, we collected the $T + K$ signal samples $\left\{\mathbf{x}^{(1-K)}, \ldots, \mathbf{x}^{(-1)}, \mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \ldots \mathbf{x}^{(T)}\right\}$ from which we computed $\left\{\mathbf{s}^{(1)}, \mathbf{s}^{(2)}, \ldots, \mathbf{s}^{(T)}\right\}$ according to what explained in Section II-C. Then, for each component $s_i, i = 1, \ldots, N$, and for each model $\mathcal{M}_i, i = 1, \ldots, 4$, we estimated the parameters (i.e., the most probable *a posteriori*, $MAP$) that best fit

the data according to (9)–(10). These estimations are related to the BN in Fig. 1 and since we deal with Gaussian and Laplacian distributions, they have well known and closed form solutions [14]. In detail, for each component $s_i$:

$\mathcal{M}_1$ $\quad \hat{\mu} = \mu_{1/2}(s_i)$ and $\hat{\lambda} = \frac{\sum_{k=1}^{T} \left| s_i^{(k)} - \hat{\mu} \right|}{T}$, where $\mu_{1/2}(s_i)$ is the median of the data set $\left\{ s_i^{(1)}, \ldots, s_i^{(T)} \right\}$;

$\mathcal{M}_2$ $\quad \hat{m} = \frac{\sum_{j=1}^{T} s_i^{(k)}}{T}$ and $\hat{\sigma}^2 = \frac{\sum_{k=1}^{T} \left( s_i^{(k)} - \hat{m} \right)^2}{T-1}$;

$\mathcal{M}_3$ $\quad \hat{\lambda} = \frac{\sum_{k=1}^{T} \left| s_i^{(k)} \right|}{T}$;

$\mathcal{M}_4$ $\quad \hat{\sigma}^2 = \frac{\sum_{k=1}^{T} \left( s_i^{(k)} \right)^2}{T}$.

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 (but not the first one, as explained in the following) of the temperature (S1) and the luminosity (S3), respectively. Both these signals have been observed during the data collection of the campaign A, in the WSN testbed T1 (DEI). From the graphs in Figs. 3–4 we see that the distribution of the principal components of our signals is well described by a Laplacian distribution. Formally, the best among the four considered models can be determined ranking them according to the Bayesian Information Criterion (BIC) introduced in Eq. (11). Since we assigned non informative priors to the model parameters, $p(\boldsymbol{\theta}_{\text{MAP}}|\mathcal{M}_i)$ is a constant for each $\mathcal{M}_i$ and therefore the BIC can be redefined as:

$$\text{BIC}(\mathcal{M}_i) \stackrel{def}{=} \ln p(\mathcal{D}|\boldsymbol{\theta}_{MAP}, \mathcal{M}_i) - \frac{\ell_i}{2} \ln(T) \; . \tag{12}$$

Fig. 5 shows the BIC for the aforementioned humidity signal, for all its principal components and for all the considered models. From this figure we see that the Laplacian models better fit the data for all principal components $s_i$, $i = 1, 2, \ldots, N$. The average BIC for each model, for the different signals, campaigns and WSN testbeds, is shown in Table II. 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 $T^{-\frac{\ell_i}{2}}$ (see Eq. (12)). Based on the above results, we can conclude that the Laplacian model

slightly better describes the principal components of all the real signals that we considered than the Gaussian one. Furthermore, it is worth noting that the first principal components (to be more precise, the first $K-1$ principal components[6] of the signal, where $K$ is the training set length) have different statistics from the remaining ones, in terms of both signal range dynamics and amplitude of the components. This is due to the fact that the first $K-1$ components actually map the observed signal into the training set vector space, instead the remaining ones are random projections of the signals. The former capture the "core" of the signal $\mathbf{x}$, the latter allow to recover its details which can lie outside the linear span of the training data. In our simulations we set $K = 2$, in accordance to the rationale presented in [20], so that only the first principal component shows a behavior different from the one illustrated in Figs. 3–4 as reported in Figs. 6–7. In any case, the Laplacian model still fits better the observed data compared to the Gaussian one.

## VI. Bayesian MAP Condition and CS Recovery for Real Signals

In the previous section we have seen that the Laplacian model is a good representation for the principal components of typical WSN signals. This legitimates the use of CS in WSNs when it is exploited according to the framework presented in Section II-C. To support this claim, in this section we review a Bayesian perspective that highlights the equivalence between the output of the CS reconstruction algorithm and the solution that maximizes the posterior probability in Eq. (9).

Assume a DCP is placed in the center of a WSN with $N$ sensor nodes and let our goal be to determine at each time $k$ all the $N$ sensor readings by just collecting at the DCP a small fraction of them. To this end, we exploit the joint CS and PCA scheme presented in Section II-C. 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 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 $\mathbf{x}^{(k)} \in \mathbb{R}^N$.

[6]Note that, according to Eq. (7), the matrix $\mathbf{U}^{(k)}$ is obtained from the elements of the training set $\mathcal{X}^{(k)}$ minus their mean, i.e., from the set $\left\{\mathbf{x}^{(k-1)} - \overline{\mathbf{x}}^{(k)}, \mathbf{x}^{(k-2)} - \overline{\mathbf{x}}^{(k)}, \cdots, \mathbf{x}^{(k-K)} - \overline{\mathbf{x}}^{(k)}\right\}$ which spans a vector space of dimension at most $K-1$.

Typically (e.g., see [28]) 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:

$$\mathbf{x}^{(k)} = \overline{\mathbf{x}}^{(k)} + \sum_{i=1}^{h} \theta_i \mathbf{f}_i \ . \tag{13}$$

In accordance to what explained in Section II-C, at each time $k$ we can do the following associations: the columns of the PCA matrix $\mathbf{U}^{(k)}$ as the set of $h = N$ basis functions in Eq. (13), 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 $\mathbf{s}^{(k)} = (s_1^{(k)}, \cdots, s_N^{(k)})^T$ as the parameter vector $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_N)^T$ in Eq. (13). In this perspective, the interpolated function has the form (see Eq. (5))

$$\mathbf{x}^{(k)} - \overline{\mathbf{x}}^{(k)} = \sum_{i=1}^{N} s_i^{(k)} \mathbf{u}_i^{(k)} \ . \tag{14}$$

A Bayesian approach would estimate the most probable value of $\mathbf{s}^{(k)} = (s_1^{(k)}, \cdots, s_N^{(k)})^T$ by maximizing a posterior pdf of the form $p(\mathbf{s}^{(k)}|\mathbf{y}^{(k)}, \mathcal{U}^{(k)}, \mathcal{M})$, where $\mathcal{M}$ is a plausible model for the vector $\mathbf{s}^{(k)}$. To avoid confusion, it is important to note that in this section the interpretation of all the variables involved is slightly different from the one adopted in Section V. In detail, now the vector $\mathbf{s}^{(k)}$ is seen as the parameter vector $\boldsymbol{\theta}$ in Eq. (9), whilst the vector $\mathbf{y}^{(k)}$ represents the set $\mathcal{D}$ of collected data. Moreover, the observed phenomenon $\mathbf{x}^{(k)}$ is modeled through both a set $\mathcal{U}^{(k)}$ of basis functions (i.e., the columns of the matrix $\mathbf{U}^{(k)}$) and a model $\mathcal{M}$ for the parameter vector $\mathbf{s}^{(k)}$, according to the BN in Fig. 2. In Eq. (9) we indicated with the symbol $\mathcal{M}_i$ a possible model for the observed phenomenon: here that symbol is replaced with the couple $(\mathcal{U}^{(k)}, \mathcal{M})$, where $\mathcal{M}$ directly refers to $\mathbf{s}^{(k)}$. Using the symbol $\mathcal{M}$ to indicate a model for $\mathbf{s}^{(k)}$ (even if $\mathbf{s}^{(k)}$ is now interpreted as the parameter vector $\boldsymbol{\theta}$) allows us to highlight the correspondence between the adoption of a particular model for $\mathbf{s}^{(k)}$ and the results of the study carried out in Sec. V. This correspondence will become clear in the following.

As in [28], we also assume that $\mathcal{M}$ can be specified by a further parameter set $\alpha$ (called hyper-prior) related to $\mathbf{s}^{(k)}$, so that the posterior can be written as

$$p(\mathbf{s}^{(k)}|\mathbf{y}^{(k)}, \mathcal{U}^{(k)}, \mathcal{M}) = \int p(\mathbf{s}^{(k)}|\mathbf{y}^{(k)}, \alpha, \mathcal{U}^{(k)}, \mathcal{M}) p(\alpha|\mathbf{y}^{(k)}, \mathcal{U}^{(k)}, \mathcal{M}) \, \mathrm{d}\alpha \ .$$

If the hyper-prior can be inferred from the data and has non zero values $\hat{\alpha}$, maximizing the posterior corresponds to maximizing $p(\mathbf{s}^{(k)}|\mathbf{y}^{(k)}, \hat{\alpha}, \mathcal{U}^{(k)}, \mathcal{M})$, that as shown in [28] corresponds to maximizing the following expression

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

where $p(\mathbf{y}^{(k)}|\mathbf{s}^{(k)}, \mathcal{U}^{(k)})$ is the likelihood function, $p(\mathbf{s}^{(k)}|\hat{\alpha}, \mathcal{M})$ is the prior and $p(\mathbf{y}^{(k)}|\hat{\alpha}, \mathcal{U}^{(k)}, \mathcal{M})$ is a normalization factor. The parameters $\hat{\alpha}$ are estimated maximizing the evidence $p(\mathbf{y}^{(k)}|\alpha, \mathcal{U}^{(k)}, \mathcal{M})$, which is a function of $\alpha$. Note that here the hyper-prior plays, in regard to $\mathbf{s}^{(k)}$, exactly the same role as the parameter vector $\boldsymbol{\theta}$ in the previous section, where $\mathbf{s}^{(k)}$ was interpreted as the collected data set $\mathcal{D}$ of the observed phenomenon; for example, if we choose $\mathcal{M} = \mathcal{L}_0$ for $\mathbf{s}^{(k)}$ then $\alpha = \lambda$, i.e., the hyper-prior is the scale parameter of the Laplacian prior assigned to $\mathbf{s}^{(k)}$.

In Eq. (14), without loss of generality we can 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 (see Eq. (6)):

$$p(\mathbf{y}^{(k)}|\mathbf{s}^{(k)}, \mathcal{U}^{(k)}) = \delta(\mathbf{y}^{(k)}, \boldsymbol{\Phi}^{(k)}\mathbf{U}^{(k)}\mathbf{s}^{(k)}) , \tag{16}$$

where $\delta(x, y)$ is 1 if $x = y$ and zero otherwise. In Section V, we have seen that the statistics of vector $\mathbf{s}^{(k)}$ is well described by a Laplacian density function with location parameter $\mu$ equal to 0 ($\mathcal{L}_0$). This pdf is widely used in the literature [4], [19] to statistically model sparse random vectors and, owing to the assumption of statistical independence of the components of $\mathbf{s}^{(k)}$, we can write it in the form:

$$p(\mathbf{s}^{(k)}|\hat{\alpha}, \mathcal{M} = \mathcal{L}_0) = \frac{e^{-\hat{\alpha}\sum_{i=1}^{N}|s_i^{(k)}|}}{(2/\hat{\alpha})^N} . \tag{17}$$

In this equation, all the components of $\mathbf{s}^{(k)}$ are assumed to be equally distributed. If (15) holds, we can obtain the following posterior:

$$\begin{aligned} p(\mathbf{s}^{(k)}|\mathbf{y}^{(k)}, \mathcal{U}^{(k)}, \mathcal{L}_0) &\propto p(\mathbf{s}^{(k)}|\mathbf{y}^{(k)}, \hat{\alpha}, \mathcal{U}^{(k)}, \mathcal{L}_0) \\ &\propto p(\mathbf{y}^{(k)}|\mathbf{s}^{(k)}, \mathcal{U}^{(k)}) p(\mathbf{s}^{(k)}|\hat{\alpha}, \mathcal{L}_0). \end{aligned} \tag{18}$$

Using (16)–(18), maximizing the posterior corresponds to solving the problem

$$\begin{aligned}
& \arg\max_{\mathbf{s}^{(k)}} p(\mathbf{s}^{(k)}|\mathbf{y}^{(k)}, \mathcal{U}^{(k)}, \mathcal{L}_0) \\
= \; & \arg\max_{\mathbf{s}^{(k)}} p(\mathbf{y}^{(k)}|\mathbf{s}^{(k)}, \mathcal{U}^{(k)}) p(\mathbf{s}^{(k)}|\hat{\alpha}, \mathcal{L}_0) \\
= \; & \arg\max_{\mathbf{s}^{(k)}} \delta(\mathbf{y}^{(k)}, \mathbf{\Phi}^{(k)}\mathbf{U}^{(k)}\mathbf{s}^{(k)}) \frac{e^{-\hat{\alpha}\sum_{i=1}^{N}|s_i^{(k)}|}}{(2/\hat{\alpha})^N} \\
= \; & \arg\min_{\mathbf{s}^{(k)}} \sum_{i=1}^{N} |s_i^{(k)}|, \text{ given that } \mathbf{y}^{(k)} = \mathbf{\Phi}^{(k)}\mathbf{U}^{(k)}\mathbf{s}^{(k)} \\
= \; & \arg\min_{\mathbf{s}^{(k)}} \|\mathbf{s}^{(k)}\|_1, \text{ given that } \mathbf{y}^{(k)} = \mathbf{\Phi}^{(k)}\mathbf{U}^{(k)}\mathbf{s}^{(k)} \; , \qquad (19)
\end{aligned}$$

which is the convex optimization problem solved by the CS reconstruction algorithms (see [3] and [30]). In our approach, unlike in the classical CS problems, the sparsification matrix $\mathbf{U}^{(k)}$ is not fixed but varies over time adapting itself to the current data.

### *A. Example Results*

In light of the above results we are legitimate in using CS for WSN monitoring applications via data reconstruction. In particular, here we describe for illustrative purposes a naive solution that, despite its simplicity, provides useful insight into the advantages that can be obtained by exploiting the CS reconstruction algorithm.

We assume a data gathering technique that alternates between a *training phase* and a *monitoring phase* [20]. During the former phase, all sensors transmit their data to a central unity (e.g., an application server) which estimates the signal statistics in terms of mean and covariance. This information is subsequently used to reconstruct the WSN signal in the latter phase, where only a small fraction of nodes transmit. Specifically, we assume that the training phase lasts $K_1$ time samples or collection rounds. During this period of time, all sensors transmit their data to the application server which therefore receive $K_1$ complete vectors with $N$ measurements[7]. During the monitoring phase, which lasts $K_2$ time

[7]In case there is an error during the transmissions, so that not all $N$ measurements are received, the server will use as the training set the last complete training set received.

samples, each sensor node transmits its data with probability $p_{\text{tx}} = L/N$, where $L < N$. The value of $L$, and therefore of $p_{\text{tx}} = L/N$, can be chosen by a central entity, based on some metric to optimize, and sent to the WSN nodes according to a feedback-like mechanism. Thus, for $K_2$ time samples each sensor transmits its data with this probability (i.e., on average only $L$ of the $N$ sensors transmit at each time sample). The interleaving between training and monitoring phases can be viewed as follows:

$$\ldots, \underbrace{\mathbf{y}^{(k)}, \ldots, \mathbf{y}^{(k+K_1-1)}}_{\text{training phase}}, \underbrace{\mathbf{y}^{(k+K_1)}, \ldots, \mathbf{y}^{(k+K_1+K_2-1)}}_{\text{monitoring phase}}, \ldots$$

$$p_{\text{tx}} = 1 \qquad\qquad p_{\text{tx}} = L/N$$

With this simple implementation, the overall energy consumption of the network can be reduced by limiting the number of transmitting sensor nodes (small $p_{\text{tx}}$) during the monitoring phase. The entire signal of interest is then reconstructed at the application server using CS. As a first step, at each time $k$ we can associate to the vector $\mathbf{y}^{(k)}$ the corresponding matrix $\mathbf{\Phi}^{(k)}$, based on the IDs of the transmitting nodes, according to (3). Then we can exploit the samples of the last recorded training set to infer the reconstructed value of $\hat{\mathbf{x}}^{(k)}$. The statistics necessary to build the sparsifying matrix $\mathbf{U}^{(k)}$ are derived from the samples of the recorded training set (i.e., mean and covariance matrix), and the CS recovery problem is solved via a convex optimization problem, e.g., see [30] and [31].

Figs. 8–9 show the quality of the monitored signal reconstruction (at the application server) $vs$ the transmission probability $p_{\text{tx}}$. The results are obtained implementing the simple above mechanism combining software simulation of protocol stack operation with the real measurements described in Section IV. The x-axis of the figures represents the $p_{\text{tx}}$ adopted during the monitoring phase, while the y-axis represents the average relative reconstruction error in the whole simulation ($k = 1, \ldots, K$), defined as

$$\overline{\xi}_R = \frac{1}{K} \sum_{k=1}^{K} \xi_R^{(k)} \ , \tag{20}$$

where $\xi_R^{(k)}$ is the relative reconstruction error at time $k$, i.e.,

$$\xi_R^{(k)} = \frac{\left\| \mathbf{x}^{(k)} - \hat{\mathbf{x}}^{(k)} \right\|_2}{\left\| \mathbf{x}^{(k)} \right\|_2} \ . \tag{21}$$

In our simulations we set the length of the training phase to $K_1 = 2$ and the length of the monitoring phase to $K_2 = 4$, according to [20]. In Fig. 8 we plot the recovery performance achieved with different kinds of signals. We note that for highly correlated signals like voltage and humidity, the reconstruction error is sufficiently small, i.e., below $\overline{\xi}_R < 0.01$, for relatively small values of $p_{\text{tx}} \approx 0.6$. Instead for more unpredictable signals like luminosity and wind the error increases sharply with decreasing $p_{\text{tx}}$; in this case an error below $\overline{\xi}_R < 0.05$ is only achievable with $p_{\text{tx}} > 0.9$. In Fig. 9 we make a comparison, instead, between signals of the same kind but measured in different environments, i.e., in indoor and outdoor environments. In detail, the indoor environment here considered is the WSN testbed T1 (DEI), where the nodes have been placed on the ground floor of the Information Engineering Department of the University of Padova, whilst we consider T2 (EPFL LUCE) as a representative outdoor WSN testbed, since sensor nodes have been placed outside, throughout the EPFL campus. Still, the possibility of reducing the transmission cost, given a desired quality threshold, strongly depends on the signal statistics. In case of indoor signals (high correlation and low variation) we can have a reconstruction error below $\overline{\xi}_R < 0.01$ even with $p_{\text{tx}} \approx 0.1$, which leads to substantial savings in transmission energy. Conversely, with outdoor signals, whose lower correlation depends also on the wider extension of the WSN itself, we need $p_{\text{tx}} > 0.7$ to obtain $\overline{\xi}_R < 0.05$. Even if the possible gains depend on the actual statistics of the observed signal, the proposed approach, despite its simplicity, adapts to the monitored process and allows us to achieve important savings in all the considered cases. This opens an interesting perspective for a useful and effective exploitation of CS in WSNs.

## VII. Conclusions

In this paper we investigated the effectiveness of data recovery through joint Compressive Sensing (CS) and Principal Component Analysis (PCA) in actual WSN deployments. 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 equivalent to Maximum A Posteriori in terms of recovery performance. Hence, we have given empirical evidence of the fact that these assumptions hold for real world signals, which we gathered from different WSN deployments 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] S. Ji, Y. Xue, and L. Carin, "Bayesian Compressive Sensing," *IEEE Trans. on Signal Processing*, vol. 56, no. 6, pp. 2346–2356, Jun. 2008.

[5] N. Dobigeon, A. O. Hero, and J.-Y. Tourneret, "Hierarchical Bayesian Sparse Image Reconstruction With Application to MRFM," *IEEE Trans. on Image Processing*, vol. 18, no. 9, pp. 2059–2070, Sep. 2009.

[6] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Bayesian Compressive Sensing Using Laplace Priors," *IEEE Trans. on Image Processing*, vol. 19, no. 1, pp. 53–63, Jan. 2010.

[7] 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.

[8] 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.

[9] 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.

[10] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*. The MIT Press, 2009.

[11] G. Quer, H. Meenakshisundaram, B. R. Tamma, B. S. Manoj, R. Rao, and M. Zorzi, "Cognitive Network Inference through Bayesian Network Analysis," in *IEEE Globecom*, Miami, FL, US, Dec. 2010.

[12] G. Quer, H. Meenakshisundaram, B. Tamma, B. S. Manoj, R. Rao, and M. Zorzi, "Using Bayesian Networks for Cognitive Control of Multi-hop Wireless Networks," in *MILCOM 2010*, San Jose, CA, US, Nov. 2010.

[13] 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.

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

[15] G. Quer, R. Masiero, M. Rossi, and M. Zorzi, "SCoRe1: Sensing Compression and Recovery through Online Estimation for Wireless Sensor Networks," *Submitted to IEEE Trans. on Wireless Communication*, 2011.

[16] 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.

[17] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning*. Springer, 2008.

[18] M. Bertero, "Linear inverse and ill-posed problems," *Advances in Electronics and Electron Physics*, vol. 75, pp. 1–120, 1989.

[19] 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*, vol. 57, no. 1, pp. 289–301, Jan. 2009.

[20] 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.

[21] C. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[22] P. Casari, A. P. Castellani, A. Cenedese, C. Lora, M. Rossi, L. Schenato, and M. Zorzi, "The "Wireless Sensor Networks for City-Wide Ambient Intelligence (WISE-WAI)" Project," *Sensors*, vol. 9, no. 6, pp. 4056–4082, May 2009.

[23] R. Crepaldi, S. Friso, A. F. Harris III, M. Mastrogiovanni, C. Petrioli, M. Rossi, A. Zanella, and M. Zorzi, "The Design, Deployment, and Analysis of SignetLab: A Sensor Network Testbed and Interactive Management Tool," in *IEEE Tridentcom*, Orlando, FL, US, May 2007.

[24] "EPFL LUCE SensorScope WSN," Last time accessed: March 2011. [Online]. Available: http://sensorscope.epfl.ch/index.php/LUCE

[25] "EPFL Grand-St-Bernard SensorScope WSN," Last time accessed: March 2011. [Online]. Available: http://sensorscope.epfl.ch/index.php/Grand-St-Bernard_Deployment

[26] "CitySense," Last time accessed: March 2011. [Online]. Available: http://www.citysense.net

[27] T. Watteyne, D. Barthel, M. Dohler, and I. Auge-Blum, "Sense&Sensitivity: A Large-Scale Experimental Study of Reactive Gradient Routing," *Measurement Science and Technology, Special Issue on Wireless Sensor Networks: Designing for Real-world Deployment and Deployment Experiences*, vol. 21, no. 12, Oct. 2010.

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

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

[30] S. Bercker, J. Bobin, and E. J. Candés, "NESTA: a fast and accurate first order method for sparse recovery." *Submitted for publication*. [Online]. Available: http://www-stat.stanford.edu/~candes/papers/NESTA.pdf

[31] G. Quer, D. Zordan, R. Masiero, M. Zorzi, and M. Rossi, "WSN-Control: Signal Reconstruction through Compressive Sensing in Wireless Sensor Networks," in *IEEE SenseApp Workshop*, Denver, CO, US, Dec. 2010, pp. 937–944.

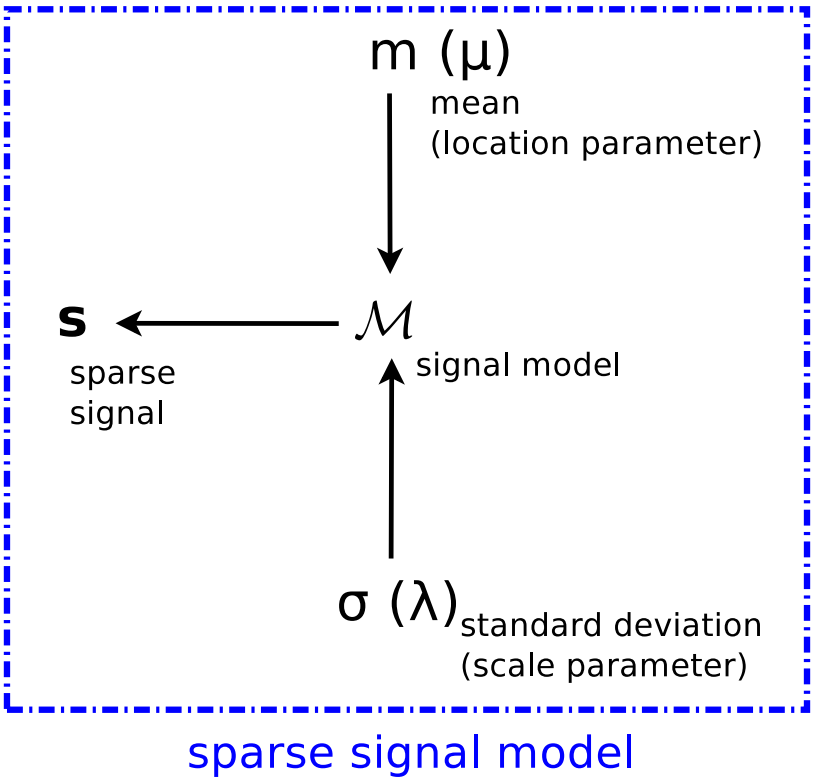


Fig. 1. Bayesian network used to model the probability distribution of the innovation signal $\mathbf{s}$.

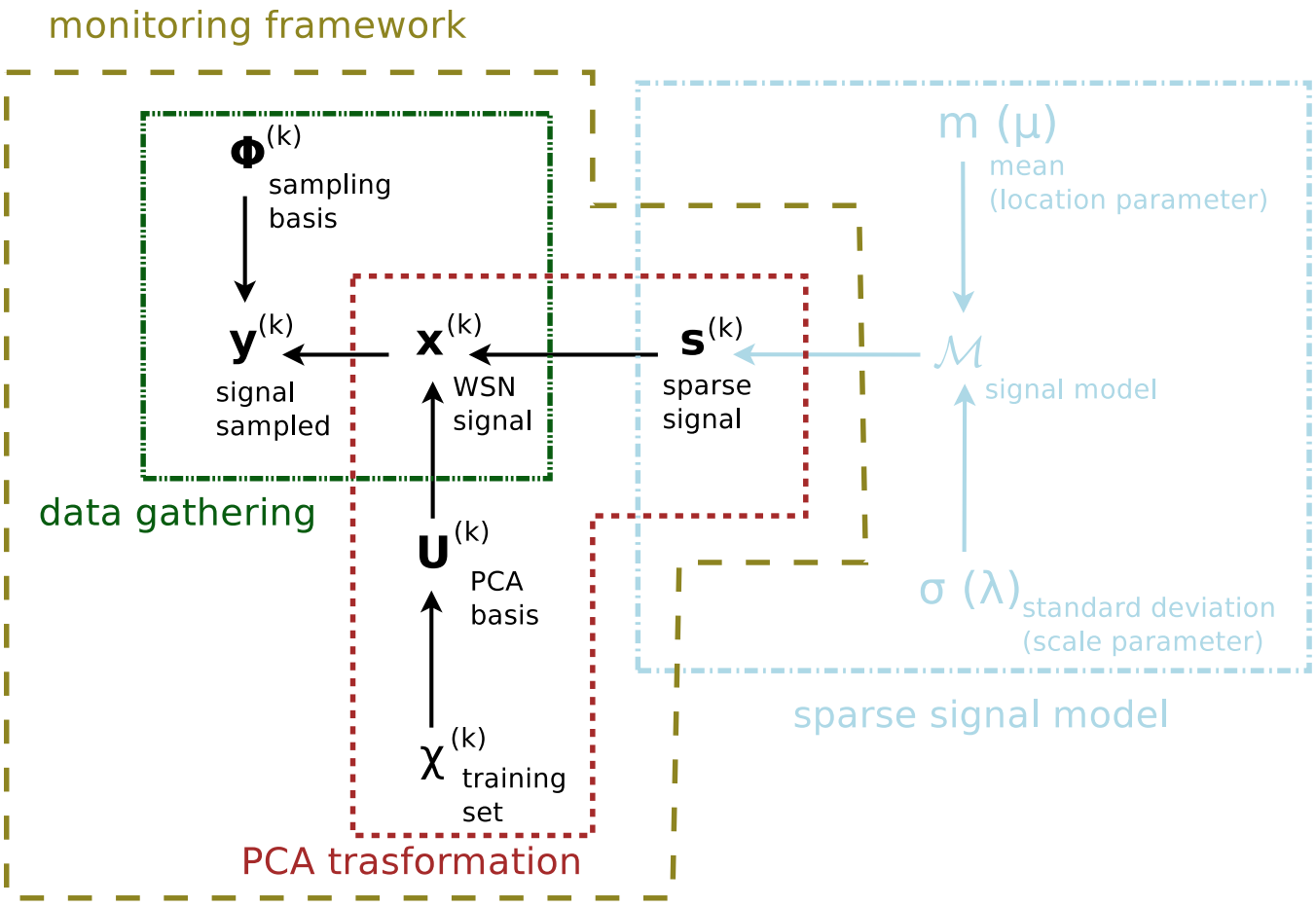


Fig. 2. Bayesian network used to model the considered real signals. In the scheme we highlight the monitoring framework at each time sample $k$.

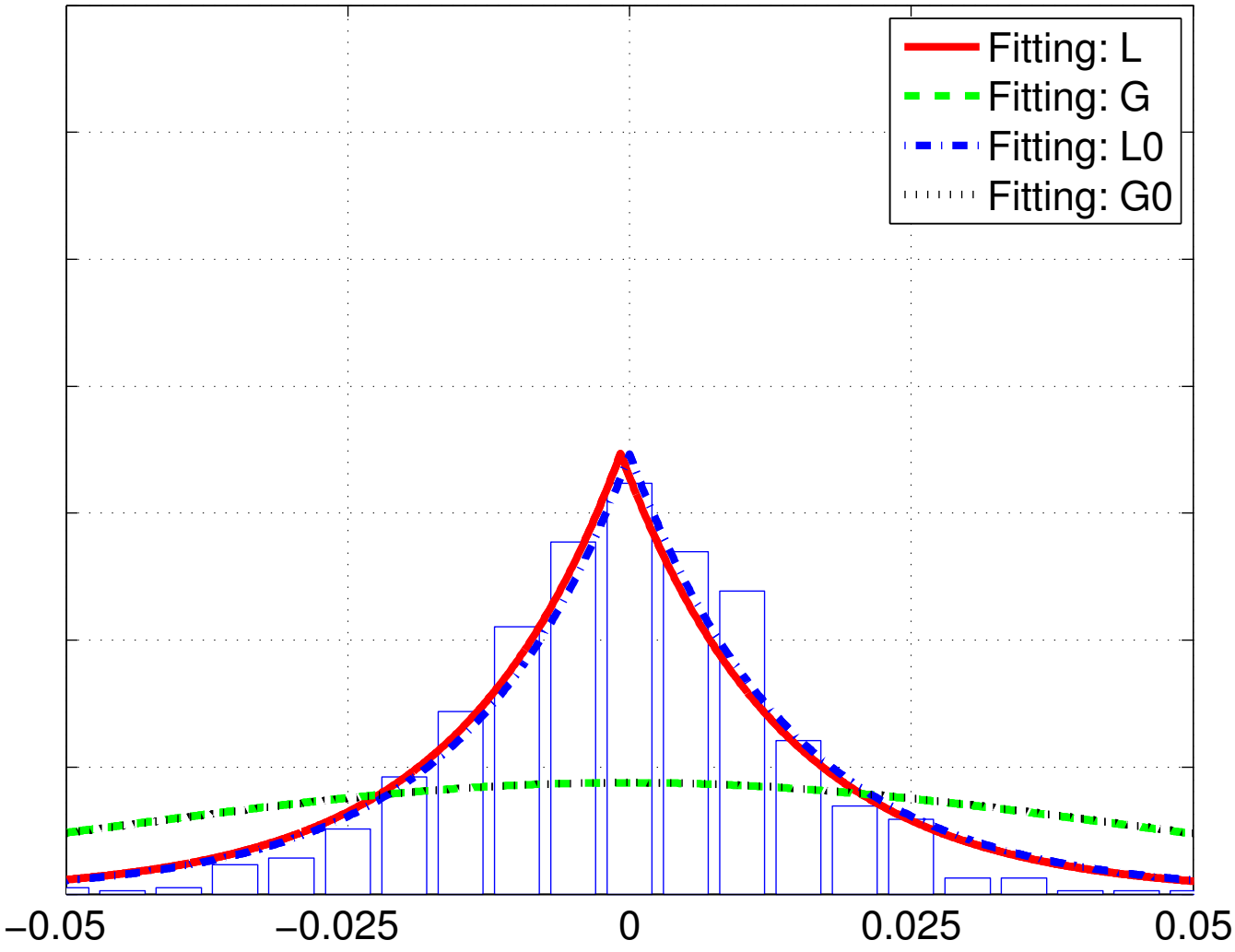


Fig. 3. Empirical distribution and model fitting for a principal component of signal S1, temperature.

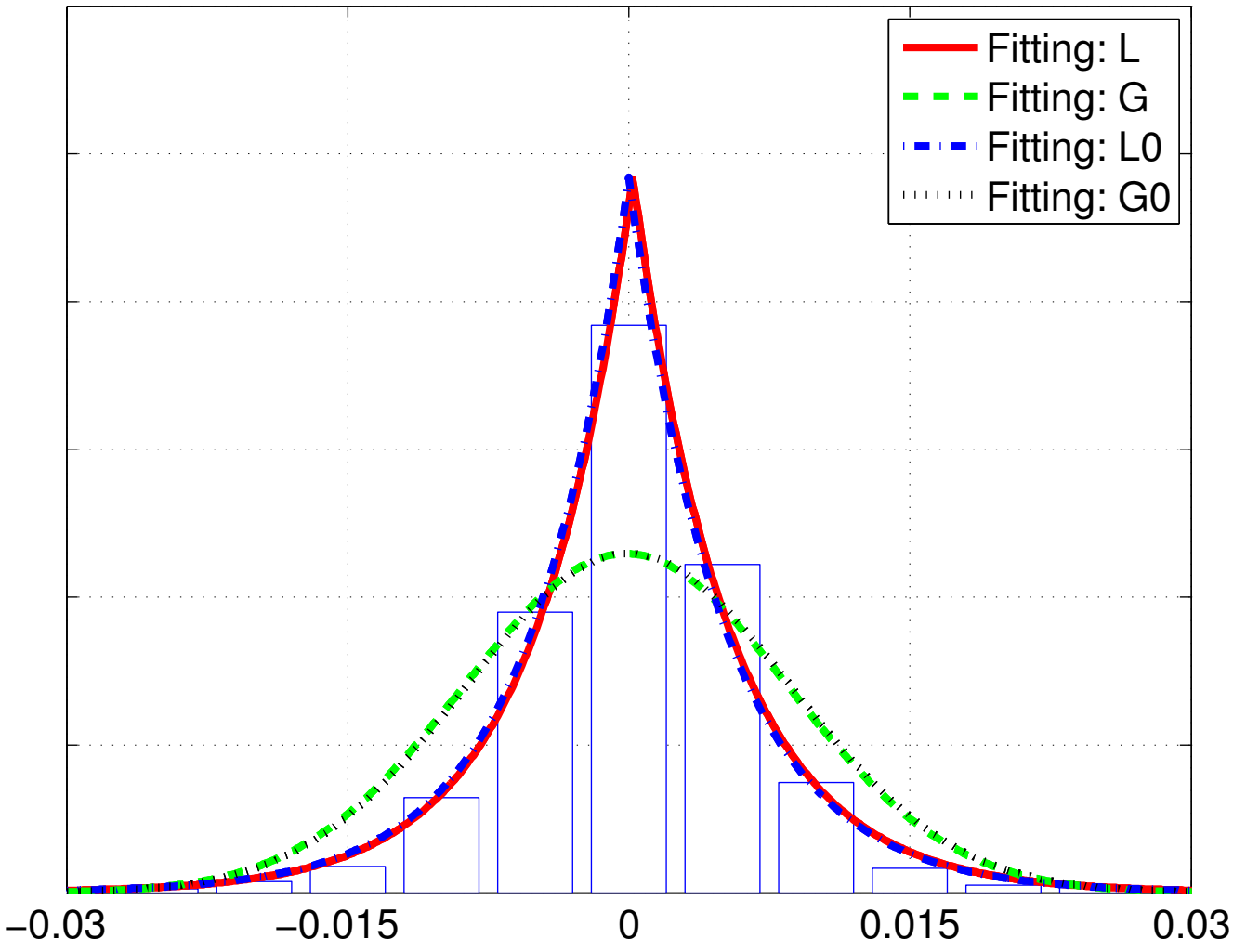


Fig. 4. Empirical distribution and model fitting for a principal component of signal S3, luminosity in the range $320 - 730$ nm.

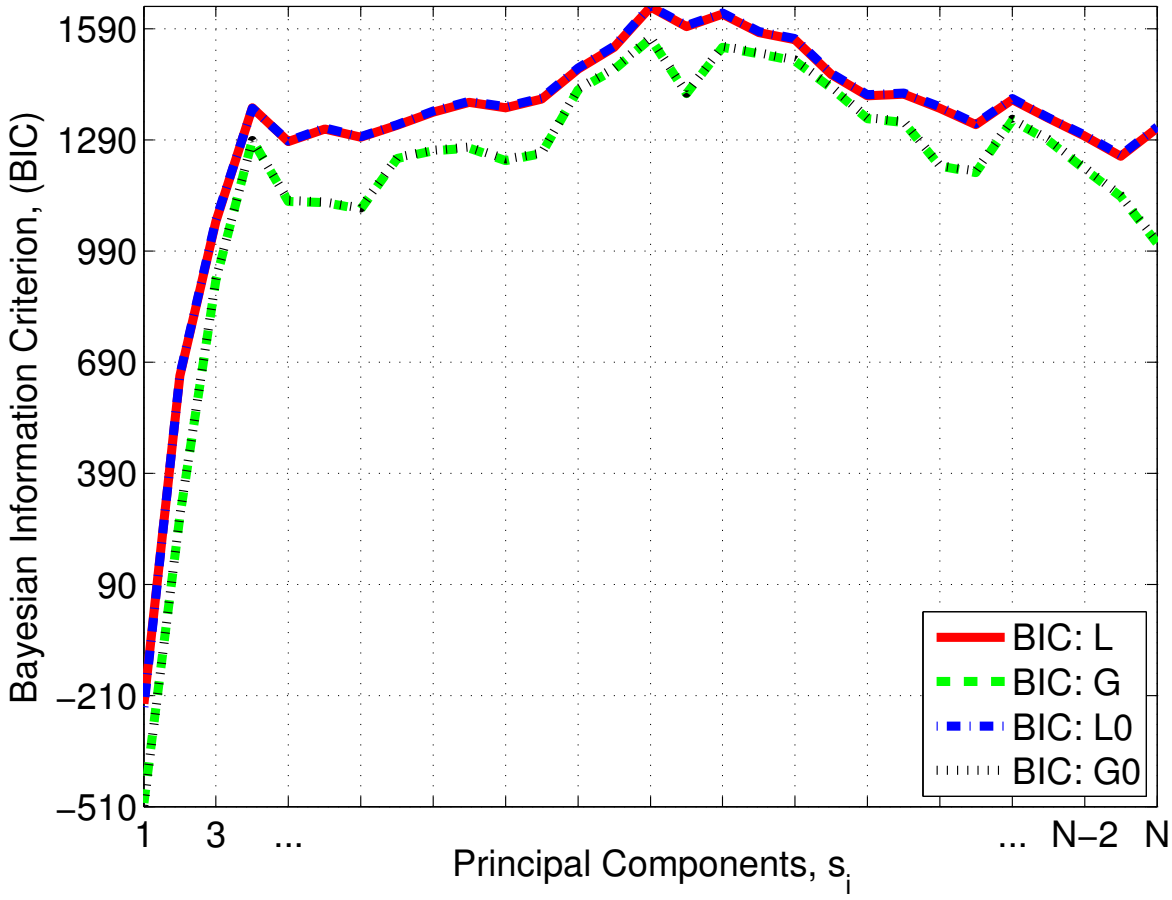


Fig. 5. Bayesian Information Criterion (BIC) per Principal Component, for each model $\mathcal{M}_1$–$\mathcal{M}_4$, WSN T1 (DEI), campaign A and signal S2, humidity.

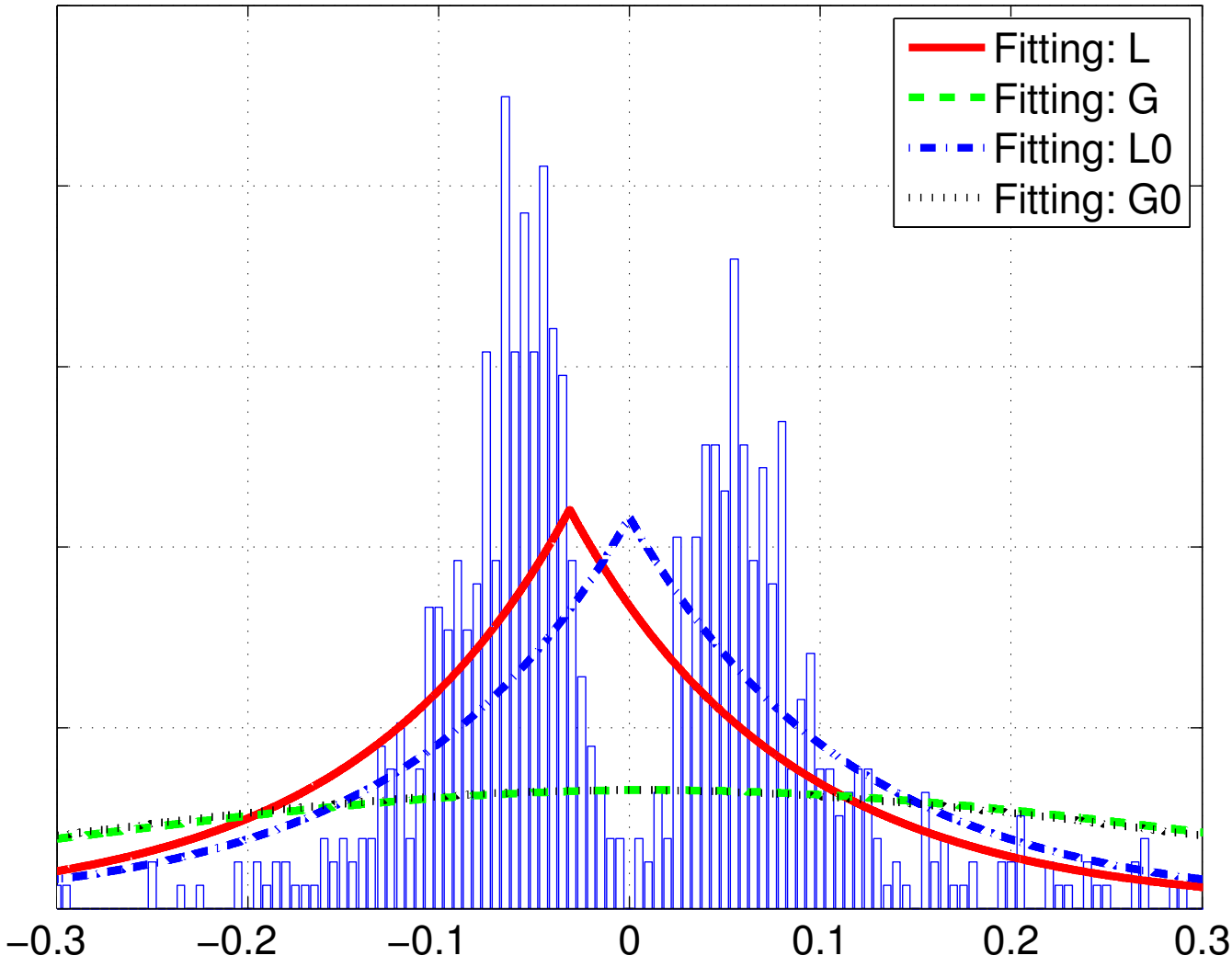


Fig. 6. Empirical distribution and model fitting for the first principal component of signal S1, temperature.

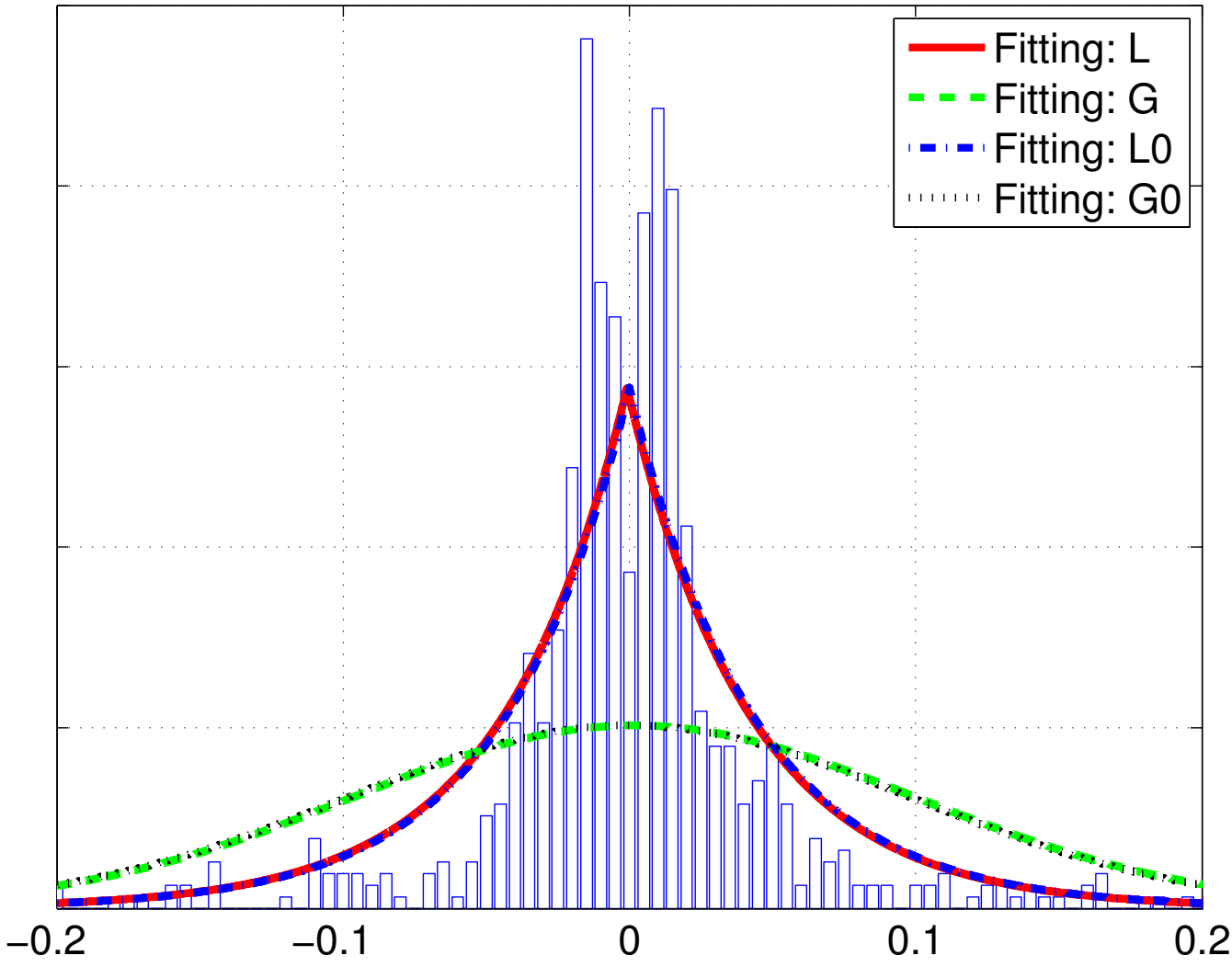


Fig. 7. Empirical distribution and model fitting for the first principal component of signal S3, luminosity in the range $320 - 730$ nm.

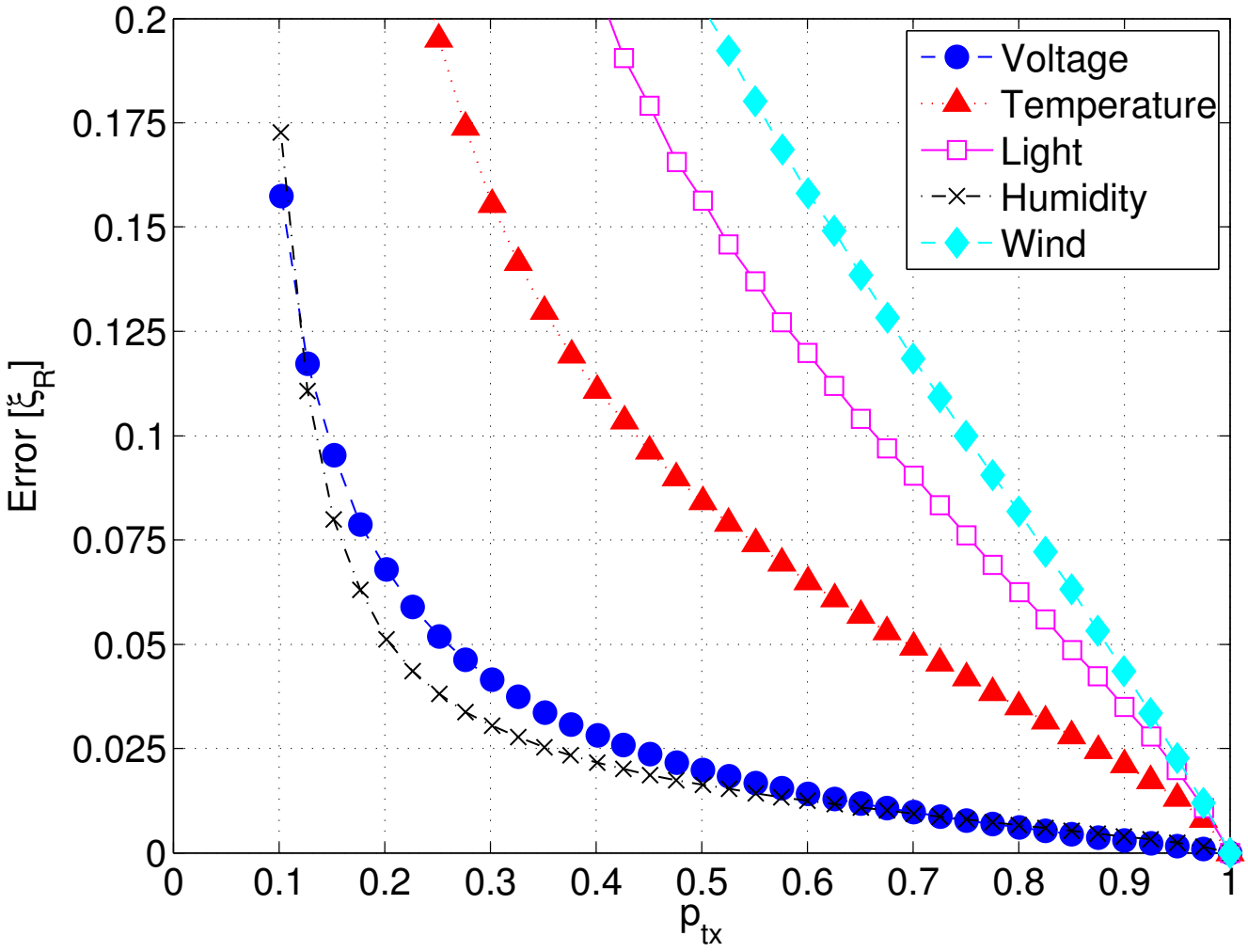


Fig. 8. Average reconstruction error for different types of signals.

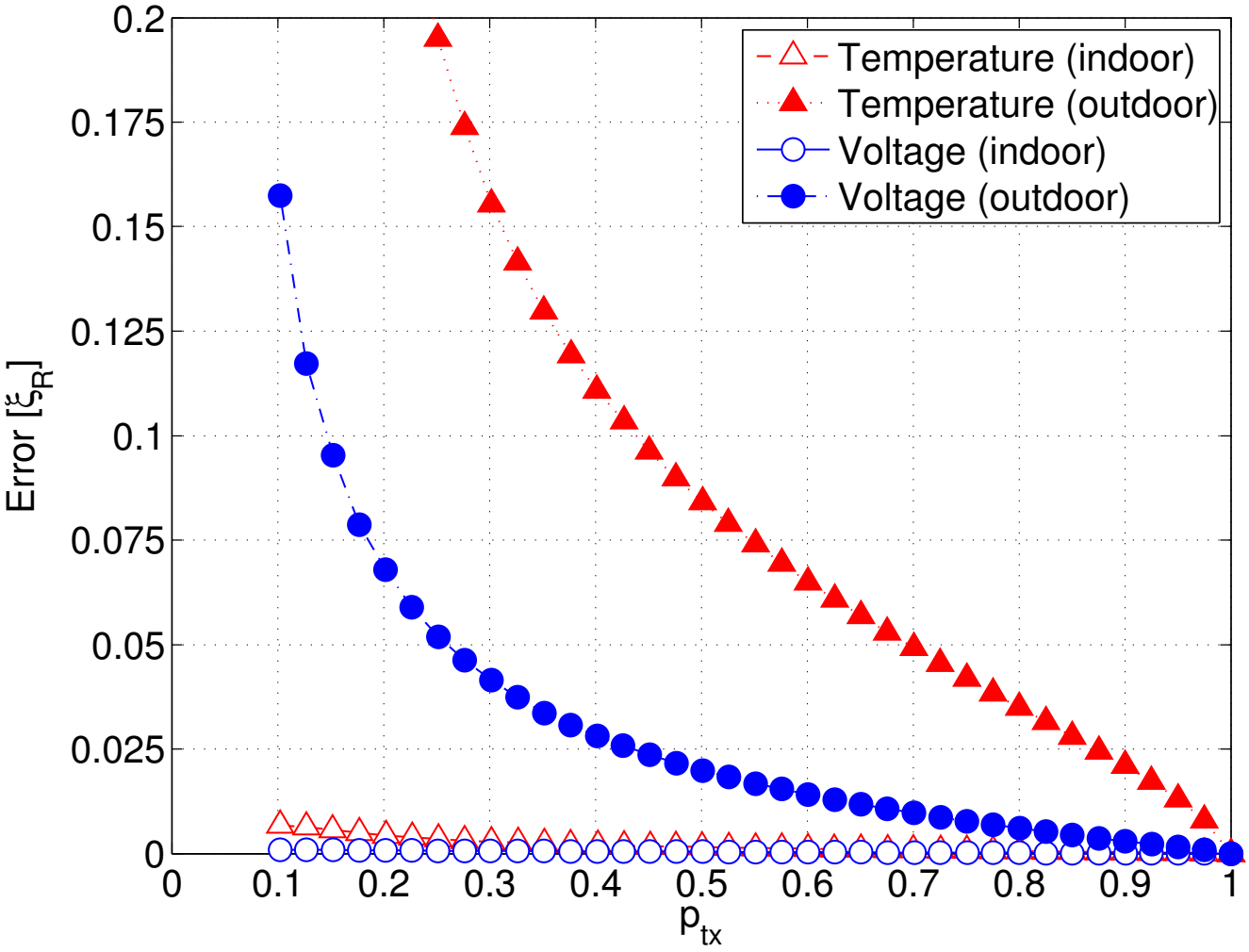


Fig. 9. Average reconstruction error for two types of signals. Comparison between signals gathered in indoor and outdoor environments, respectively.

| WSN Testbed T1 (DEI) | | | | | | |
|---|---|---|---|---|---|---|
| | **# of nodes** | **frame length** | **# of frames** | **starting time (G.M.T)** | **stopping time (G.M.T)** | **signals** |
| **Campaign A** | 37 | 5 min | 783 | 13/03/2009, 09:05:22 | 16/03/2009, 18:20:28 | S1 S2 S3 S4 S6 |
| **Campaign B** | 45 | 5 min | 756 | 19/03/2009, 10:00:34 | 22/03/2009, 17:02:54 | S1 S2 S3 S4 S6 |
| **Campaign C** | 31 | 5 min | 571 | 24/03/2009, 11:05:10 | 26/03/2009, 10:15:42 | S1 S2 S3 S4 S6 |
| **WSN Testbed T2 (EPFL LUCE)** | | | | | | |
| | **# of nodes** | **frame length** | **# of frames** | **starting time (G.M.T)** | **stopping time (G.M.T)** | **signals** |
| **Campaign A** | 85 | 5 min | 865 | 12/01/2007, 15:09:26 | 15/01/2007, 15:13:26 | S1 S2 S5 |
| **Campaign B** | 72 | 5 min | 841 | 06/05/2007, 16:09:26 | 09/05/2007, 14:13:26 | S1 S2 S5 |
| **Campaign C** | 83 | 30 min | 772 | 02/02/2007, 17:09:26 | 18/02/2009, 19:09:26 | S6 S7 |
| **WSN Testbed T3 (EPFL St Bernard)** | | | | | | |
| | **# of nodes** | **frame length** | **# of frames** | **starting time (G.M.T)** | **stopping time (G.M.T)** | **signals** |
| **Campaign A** | 23 | 5 min | 742 | 03/10/2007, 12:35:37 | 06/10/2007, 02:35:37 | S1 S2 S5 |
| **Campaign B** | 22 | 5 min | 756 | 19/10/2007, 12:35:37 | 22/10/2007, 03:35:37 | S1 S2 S5 |
| **Campaign C** | 22 | 30 min | 778 | 02/10/2007, 07:06:05 | 19/10/2007, 12:06:50 | S6 S7 |
| **WSN Testbed T4 (CitySense)** | | | | | | |
| | **# of nodes** | **frame length** | **# of frames** | **starting time (G.M.T)** | **stopping time (G.M.T)** | **signals** |
| **Campaign A** | 8 | 60 min | 887 | 14/10/2009, 14:01:57 | 21/11/2009, 00:01:57 | S1 |
| **Campaign B** | 8 | 60 min | 888 | 14/10/2009, 13:00:01 | 21/11/2009, 00:00:01 | S5 |
| **WSN Testbed T5 (Sense&Sensitivity)** | | | | | | |
| | **# of nodes** | **frame length** | **# of frames** | **starting time (G.M.T)** | **stopping time (G.M.T)** | **signals** |
| **Campaign A** | 77 | 15 min | 65 | 26/08/2008, 14:46:46 | 27/08/2008, 07:31:07 | S1 S3 S4 S6 |

TABLE I
DETAILS OF THE CONSIDERED WSN AND GATHERED SIGNALS.

| **WSN Testbed T1 (DEI)** | | | | | | | |
|---|---|---|---|---|---|---|---|
| | S1 (Temperature) | S2 (Humidity) | S3 (Light) | S4 (IR) | S5 (Wind) | S6 (Voltage) | S7 (Current) |
| $\mathcal{L}$ | 1382.8 | 1059.8 | 2191.7 | 1760.9 | - | 4656.9 | - |
| $\mathcal{G}$ | 1042.1 | 804.9 | 1690 | 1154.5 | - | 3814.1 | - |
| $\mathcal{L}_0$ | **1385.5** | **1062.4** | **2194.9** | **1764.1** | - | **4660.1** | - |
| $\mathcal{G}_0$ | 1044.9 | 807.60 | 5078.3 | 1157.4 | - | 3816.9 | - |
| **WSN Testbed T2 (EPFL LUCE)** | | | | | | | |
| | S1 (Temperature) | S2 (Humidity) | S3 (Light) | S4 (IR) | S5 (Wind) | S6 (Voltage) | S7 (Current) |
| $\mathcal{L}$ | -36.1 | -992.3 | - | - | -3694.9 | 1854.1 | -972.8 |
| $\mathcal{G}$ | -195.3 | -1163.7 | - | - | -4026.5 | 1191.4 | -1520.3 |
| $\mathcal{L}_0$ | **-33.3** | **-989.5** | - | - | **-3691.5** | **1856.3** | **-969.6** |
| $\mathcal{G}_0$ | -192.5 | -1160.9 | - | - | -4023.6 | 1194.2 | -1517.3 |
| **WSN Testbed T3 (EPFL St Bernard)** | | | | | | | |
| | S1 (Temperature) | S2 (Humidity) | S3 (Light) | S4 (IR) | S5 (Wind) | S6 (Voltage) | S7 (Current) |
| $\mathcal{L}$ | -82.3 | -1473 | - | - | -3700.2 | 1617.8 | -1557.5 |
| $\mathcal{G}$ | -487.4 | -1700.7 | - | - | -3850.3 | 1087.9 | -1877.2 |
| $\mathcal{L}_0$ | **-79.3** | **-1469.9** | - | - | **-3697.3** | **1619** | **-1554.2** |
| $\mathcal{G}_0$ | -484.7 | -1697.8 | - | - | -3847.5 | 1090.7 | -1874.2 |
| **WSN Testbed T4 (CitySense)** | | | | | | | |
| | S1 (Temperature) | S2 (Humidity) | S3 (Light) | S4 (IR) | S5 (Wind) | S6 (Voltage) | S7 (Current) |
| $\mathcal{L}$ | -858.1 | - | - | - | -4309.5 | - | - |
| $\mathcal{G}$ | -1094.6 | - | - | - | -4384.2 | - | - |
| $\mathcal{L}_0$ | **-856.8** | - | - | - | **-4306.4** | - | - |
| $\mathcal{G}_0$ | -1091.9 | - | - | - | -4381.2 | - | - |
| **WSN Testbed T5 (Sense&Sensitivity)** | | | | | | | |
| | S1 (Temperature) | S2 (Humidity) | S3 (Light) | S4 (IR) | S5 (Wind) | S6 (Voltage) | S7 (Current) |
| $\mathcal{L}$ | -127.7 | - | -196.2 | -184.4 | - | 110 | - |
| $\mathcal{G}$ | -176.1 | - | -232.1 | -227.5 | - | 70.2 | - |
| $\mathcal{L}_0$ | **-125.7** | - | **-194.2** | **-182.3** | - | **111.9** | - |
| $\mathcal{G}_0$ | -174.7 | - | -230.6 | -225.8 | - | 71.8 | - |

TABLE II

BAYESIAN INFORMATION CRITERION (BIC) AVERAGED OVER ALL PRINCIPAL COMPONENTS AND RELATIVE CAMPAIGNS, FOR EACH MODEL $\mathcal{M}_1$–$\mathcal{M}_4$, FOR EACH TESTBED T1–T5 AND EACH CORRESPONDING PROVIDED SIGNAL AMONG S1–S7.